JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 13 (2025) ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Viewpoints

Should Digital Interventions for HIV Self-Testing Be Regulated with World Health Organization Prequalification? (e60276)
Alex Fischer, Samanta Lalla-Edward, Vinodh Edward, Musaed Abrahams, Luke Shankland, John de Wit. . . . 18

Evolving Digital Health Technologies: Aligning With and Enhancing the National Institute for Health and Care Excellence Evidence Standards Framework (e67435)
Shayan Bahadori, Peter Buckle, Tayana Soukup Ascensao, Saira Ghafur, Patrick Kierkegaard. . . . 25

What to Consider When Developing Multidomain Mobile Health Interventions for Lifestyle Management (e63573)
Manuel Weber, Renato Mattli, Anja Raab, Anja Frei, Karin Haas, Thimo Marcin, Albrecht Vorster, Kai-Uwe Schmitt. . . . 33

The Challenge of Measuring Exercise: Advancing Metrological Barriers in Wearable Sensing (e79347)
Jennifer Corso, Evan Peikon. . . . 44

Reviews

Wearable Technologies in Head and Neck Oncology: Scoping Review (e72372)
Matthew Shammas-Toma, Gianluca Sampieri, Michael Xie, Aishwaria Maxwell, Alex Esemezie, Quynh Pham, Joseph Cafazzo, Philip Wong, C Tsai, David Goldstein, John de Almeida, Ervin Sejdic, Christopher Yao. . . . 61

Digital Health Interventions to Reduce Cancer-Related Fatigue Among Adolescents and Young Adults: Scoping Review (e68834)
Shanshan Jiang, Xiaoyu Yang, Xinying Yu. . . . 73

Digital Health Interventions for Military Members, Veterans, and Public Safety Personnel: Scoping Review (e65149)
Rashell Allen, Myrah Malik, Carley Aquin, Lucijana Herceg, Suzette Brémault-Phillips, Phillip Sevigny. . . . 86

User Experience in mHealth Research: Bibliometric Analysis of Trends and Developments (2007–2023) (e75909)
Bashaer Alkhwaiter, Monira Aloud, Nora Almezeini. . . . 104

The Effect and Safety of App-Based Interventions for Populations With Osteoarthritis: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e71193)
He Zhang, Hongchi Wang, Cheng Zhang, Ruili Zhao, Qian Lv, Luyao Zhang, Haoran Zhang, Shiyan Yan. . . . 136

XSL•FO
RenderX

Behavior Change Resources Used in Mobile App–Based Interventions Addressing Weight, Behavioral, and Metabolic Outcomes in Adults With Overweight and Obesity: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e63313)
Sijia Li, You Zhou, Ying Tang, Haoming Ma, Yuying Zhang, Aoqi Wang, Xingyi Tang, Runyuan Pei, Meihua Piao. 153

Impact of Smartphone Apps on Reperfusion Times and Clinical Outcomes in Acute ST-Segment Elevation Myocardial Infarction: Systematic Review and Meta-Analysis (e66605)
William Gibson, Dawoud Al Kindi, Elie Akl, Kshitij Dandona, Jean-Philippe Pelletier, Nicolo Piazza, Ali Zgheib, Giuseppe Martucci, Marco Spaziano. 177

Comparative Effectiveness of Digital Health Technologies in Tuberculosis Treatment: Systematic Review and Network Meta-Analysis of Randomized Controlled Trials (e75424)
Qinglin Cheng, Ping Chen, Ruoqi Dai, Qingjun Jia, Xuexin Bai, Qiancheng Cao, Qingchun Li, Yifei Wu, Yinyan Huang. 194

Integrating Mobile Health App Data Into Electronic Medical or Health Record Systems and Its Impact on Health Care Delivery and Patient Health Outcomes: Scoping Review (e66650)
Jialing Lin, Shona Bates, Luke Allen, Michael Wright, Limin Mao, Michael Kidd. 209

Effectiveness of Digital Health Interventions on Sedentary Behavior Among Patients With Chronic Diseases: Systematic Review and Meta-Analysis (e59943)
Yan Zhang, Fei Ngai, Qingling Yang, Yao Xie. 220

Adherence to eHealth Interventions Among Patients With Heart Failure: Scoping Review (e63409)
Arno Gingele, Bianca Steiner, Bettina Zippel-Schultz, Hans-Peter Brunner-La Rocca. 235

Efficacy of Mobile App–Based Dietary Interventions Among Cancer Survivors: Systematic Review and Meta-Analysis (e65505)
Krista Chung, Naomi Takemura, Wendy Lam, Mandy Ho, Antoinette Lee, Wynnie Chan, Daniel Fong. 250

Experience of Using Electronic Inhaler Monitoring Devices for Patients With Chronic Obstructive Pulmonary Disease or Asthma: Systematic Review of Qualitative Studies (e57645)
Jilong Duan, Xia Chen, Di Fan, Haikun Jiang, Xue Zhang, Wenyue Zhang, Zhiping Liu, Hongyan Lu. 272

Digital Intervention in Children With Developmental Language Disorder: Systematic Review (e59992)
Zhaowen Zhou, Cheng Deng, Dongling Yin, Qiaoxue Yang, Zhuoming Chen. 284

Effectiveness of Telemedicine on Wound-Related and Patient-Reported Outcomes in Patients With Chronic Wounds: Systematic Review and Meta-Analysis (e58553)
Xiaoyan Zhang, Zhanghui Guo, Yu Duan, Chao Sun, Jiayin Luo. 296

The Use of Artificial Intelligence and Wearable Inertial Measurement Units in Medicine: Systematic Review (e60521)
Ricardo Smits Serena, Florian Hinterwimmer, Rainer Burgkart, Rudiger von Eisenhart-Rothe, Daniel Rueckert. 325

Applying AI in the Context of the Association Between Device-Based Assessment of Physical Activity and Mental Health: Systematic Review (e59660)
Simon Woll, Dennis Birkenmaier, Gergely Biri, Rebecca Nissen, Luisa Lutz, Marc Schroth, Ulrich Ebner-Priemer, Marco Giurgiu. 344

The Role of Environmental Factors in Technology-Assisted Physical Activity Intervention Studies Among Older Adults: Scoping Review (e59570)
Carl-Philipp Jansen, Désirée Nijland, Jean-Michel Oppert, Veysel Alcan, Kirsi Keskinen, Emmi Matikainen-Tervola, Zada Pajalic, Merja Rantakokko, Signe Tomsone, Essi-Mari Tuomola, Erja Portegijs, Erik Timmermans. 362

XSL•FO
RenderX

Effects of Remote Web-Based Interventions on the Physiological and Psychological States of Patients With Cancer: Systematic Review With Meta-Analysis (e71196)
Lv Tian, Yixuan Wen, Jingmiao Li, Jiexin Guan, Tao Li, Jun Fan. . . . 387

Recent Advancements in Wearable Hydration-Monitoring Technologies: Scoping Review of Sensors, Trends, and Future Directions (e60569)
Nazim Belabbaci, Raphael Anaadumba, Mohammad Alam. . . . 409

Mobile Health Intervention Tools Promoting HIV Pre-Exposure Prophylaxis Among Adolescent Girls and Young Women in Sub-Saharan Africa: Scoping Review (e60819)
Alex Fischer, Homaira Hanif, Jacob Stocks, Aimee Rochelle, Karen Dominguez, Eliana Armora Langoni, H Reyes, Gustavo Doncel, Kathryn Muessig. . . . 451

Use of Wearable Sensors to Assess Fall Risk in Neurological Disorders: Systematic Review (e67265)
Mirjam Bonanno, Augusto Ielo, Paolo De Pasquale, Antonio Celesti, Alessandro De Nunzio, Angelo Quartarone, Rocco Calabrò. . . . 470

Engagement With Conversational Agent–Enabled Interventions in Cardiometabolic Disease Self-Management: Systematic Review (e67913)
Nick Kashyap, Ann Sebastian, Chris Lynch, Paul Jansons, Ralph Maddison, Tilman Dingler, Brian Oldenburg. . . . 495

Effects of Remote Patient Monitoring on Health Care Utilization in Patients With Noncommunicable Diseases: Systematic Review and Meta-Analysis (e68464)
Geir Smedslund, Nina Østerås, Christine Hestevik. . . . 509

AI-Driven Real-Time Monitoring of Cardiovascular Conditions With Wearable Devices: Scoping Review (e73846)
Ali Abedi, Anshul Verma, Dherya Jain, Jathushan Kaetheeswaran, Cynthia Chui, Milad Lankarany, Shehroz Khan. . . . 534

Use and Application of mHealth Technologies in Perioperative Surgical Care: Narrative Review (e52206)
Maria Sierra, Henry To, Wan Song. . . . 975

Mobile Apps and Wearable Devices for Cardiovascular Health: Narrative Review (e65782)
Gauri Chauhan, Patrick Vavken, Christine Jacob. . . . 1906

Participant Compliance With Ecological Momentary Assessment in Movement Behavior Research Among Adolescents and Emerging Adults: Systematic Review (e52887)
Shirlene Wang, Chih-Hsiang Yang, Denver Brown, Alan Cheng, Matthew Kwan. . . . 2995

Original Papers

Natural Language Processing Chatbot–Based Interventions for Improvement of Diet, Physical Activity, and Tobacco Smoking Behaviors: Systematic Review (e66403)
Jing Chen, Run-Ze Hu, Yu-Xuan Zhuang, Jia-Qi Zhang, Rui Shan, Yang Yang, Zheng Liu. . . . 311

Perspectives and Experiences of Family Caregivers Using Supportive Mobile Apps in Dementia Care: Meta-Synthesis of Qualitative Research (e65983)
Haifei Shen, Yi Han, Wen Shi, Jiangxuan Yu, Xueqi Shan, Hongyao Wang, Junjie Wang. . . . 430

Effects of a Smartphone-Based Positive Reflection Diary on Work Engagement Among Japanese Workers: Randomized Controlled Trial (e55664)
Masahito Tokita, Shuichiro Kobayashi, Daisuke Miyanaka, Nobuyuki Takate, Hiroyuki Nakano, Kazuki Takeuchi, Akihito Shimazu. . . . 551

Automatic Image Recognition Meal Reporting Among Young Adults: Randomized Controlled Trial (e60070)
Prasan Sahoo, Sherry Chiu, Yu-Sheng Lin, Chien-Hung Chen, Denisa Irianti, Hsin-Yun Chen, Mekhla Sarkar, Ying-Chieh Liu. . . . 568

XSL•FO
RenderX

Exploring the Utility of Digital Voice Assistants for Primary Care Patients, Including Those With Physical and Visual Disabilities: Cross-Sectional Study (e66185)
Maya Rajan, Allison Furgal, Reema Kadri, Omar Arman, Kate Panzer, Donna Wicker, Michael McKee, Melissa Plegue, Alexandria Degner, Lorraine Buis. . . . 582

Objectively and Subjectively Measured Physical Activity and Their Associations With Cardiometabolic Risk in the UK Biobank: Retrospective Cohort Study (e54820)
Charlyne Bürki, Caiwei Tian, Kenneth Westerman, Chirag Patel. . . . 598

Using Wear Time for the Analysis of Consumer-Grade Wearables' Data: Case Study Using Fitbit Data (e46149)
Loubna Baroudi, Ronald Zernicke, Muneesh Tewari, Noelle Carlozzi, Sung Choi, Stephen Cain. . . . 613

An Actor-Partner Interdependence Mediation Model for Assessing the Association Between Health Literacy and mHealth Use Intention in Dyads of Patients With Chronic Heart Failure and Their Caregivers: Cross-Sectional Study (e63805)
Xiaorong Jin, Yimei Zhang, Min Zhou, Qian Mei, Yangjuan Bai, Qiulan Hu, Wei Wei, Xiong Zhang, Fang Ma. . . . 625

Participant Adherence in Repeated-Dose Clinical Studies Using Video-Based Observation: Retrospective Data Analysis (e65668)
Jihong Song, Sungpil Han, Suein Choi, Jonghyuk Lim, Byeong Oh, Dongoh Shin, Seunghoon Han. . . . 641

Enhancing Cardiopulmonary Resuscitation Quality Using a Smartwatch: Neural Network Approach for Algorithm Development and Validation (e57469)
Gaurav Rao, David Savage, Gabrielle Erickson, Nathan Kyryluk, Pawan Lingras, Vijay Mago. . . . 655

Relationship Between Activity Tracker Metrics and the Physical Activity Index and Their Association With Cardiometabolic Phenotypes, Subclinical Atherosclerosis, and Cardiac Remodeling: Cross-Sectional Study (e71213)
Weiting Huang, Mark Wong, Enver Loh, Tracy Koh, Alex Tan, Xiayan Shen, Onur Varli, Siew Kong, Calvin Chin, Swee Tan, Jonathan Yap, Eddie Ng, Khung Yeo. . . . 670

Within- and Between-Individual Compliance in Mobile Health: Joint Modeling Approach to Nonrandom Missingness in an Intensive Longitudinal Observational Study (e65350)
Young Cho, Sy-Miin Chow, Jixin Li, Wei-Lin Wang, Shirlene Wang, Linying Ji, Vernon Chinchilli, Stephen Intille, Genevieve Dunton. . . . 685

Evaluating the Clinical Effectiveness and Patient Experience of a Large Language Model–Based Digital Tool for Home-Based Blood Pressure Management: Mixed Methods Study (e68361)
Alan Jelic, Igor Sesto, Luka Rotkvic, Luka Pavlovic, Nikola Erceg, Nina Sesto, Zeljko Kraljevic, Joshua Au Yeung, Amos Folarin, Richard Dobson, Petroula Laiou. . . . 703

Adolescent Engagement With a Multicomponent mHealth Tool: Identifying Usage Patterns, Determinants, and Health Behavior Change in an Intervention Trial (e59041)
Carmen Peuters, Ann DeSmet, Laura Maenhout, Greet Cardon, Dries Debeer, Geert Crombez. . . . 712

Exploring Fit in a Mobile Health Intervention for Alcohol Use Disorder: Qualitative Study (e65218)
Nora Jacobson, Linda Park, Alice Pulvermacher, Samantha Voelker, Mallory Herzog, Andrew Quanbeck. . . . 730

Exploring the Barriers and Facilitators to Implementing a Smartphone App for Physicians to Improve the Management of Acute Myocardial Infarctions: Multicenter, Mixed Methods, Observational Study (e60173)
Katelyn Cullen, Hassan Mir, Madhu Natarajan, Marija Corovic, Karen Mosleh, Jacob Crawshaw, Mathew Mercuri, Hassan Masoom, JD Schwalm. . . . 744

XSL•FO
RenderX

Efficacy, Safety, and Cost-Effectiveness of “Internet + Pharmacy Care” Via the Alfalfa App in Warfarin Therapy Management After Cardiac Valve Replacement: Randomized Controlled Trial (e53586)
Yiyi Qian, Weizhao Chen, Bin Zhou, Jiangya Li, Yuanyuan Guo, Zhiying Weng, Jinhua Zhang. 762

Preferences for Mobile Apps That Aim to Modify Alcohol Use: Thematic Content Analysis of User Reviews (e63148)
Megan Kirouac, Christina Gillezeau. 774

Insights Into How mHealth Applications Could Be Introduced Into Standard Hypertension Care in Germany: Qualitative Study With German Cardiologists and General Practitioners (e56666)
Susann May, Frances Seifert, Dunja Bruch, Martin Heinze, Sebastian Spethmann, Felix Muehlensiepen. 785

Users' Perceptions and Trust in AI in Direct-to-Consumer mHealth: Qualitative Interview Study (e64715)
Katie Ryan, Justin Hogg, Max Kasun, Jane Kim. 799

Population-Level Dissemination of a Smoking Cessation Smartphone App: Quasi-Experimental Comparison of Values-Based Messages in Social Media Advertisements (e71619)
Jonathan Bricker, Margarita Santiago-Torres, Kristin Mull, Brianna Sullivan, Ravi Mehrotra. 816

Leveraging Smartphone Mobility Data to Understand HIV Risk Among Rural South African Young Adults: Feasibility Study (e67519)
Thulile Mathenjwa, Elphas Okango, Khai Tram, Maxime Inghels, Diego Cuadros, Hae-Young Kim, Fiona Walsh, Till Barnighausen, Adrian Dobra, Frank Tanser. 827

Adaptation of an mHealth Solution for the Nutritional Management of Diabetes in a Low- and Middle-Income Country: Pre-Post Mixed Methods Pilot Study (e58029)
Cecilia Anza-Ramirez, Karen Bonilla-Aguilar, David Beran, Jean-Luc Mando, Lorena Saavedra-Garcia, Monica Angulo-Barranca, Marco Taboada Garcia, Leonardo Albitres-Flores, Alejandro Loayza, Jessica Zafra-Tanaka, Olivia Heller, María Lazo-Porras, Montserrat Castellsague Pique Perolini. 840

Impact of a Mobile Money–Based Conditional Cash Transfer Intervention on Health Care Utilization in Southern Madagascar: Mixed-Methods Study (e60811)
Mara Franke, Anne Neumann, Kim Nordmann, Daniela Suleymanova, Onja Ravololohanitra, Julius Emmrich, Samuel Knauss. 859

SEARCH Study: Text Messages and Automated Phone Reminders for HPV Vaccination in Uganda: Randomized Controlled Trial (e63527)
Sabrina Kitaka, Joseph Rujumba, Sarah Zalwango, Betsy Pfeffer, Lubega Kizza, Juliane Nattimba, Ashley Stephens, Nicolette Nabukeera-Barungi, Chelsea Wynn, Juliet Babirye, John Mukisa, Ezekiel Mupere, Melissa Stockwell. 874

Effectiveness of a Mobile Health Intervention (DOT Selfie) in Increasing Treatment Adherence Monitoring and Support for Patients With Tuberculosis in Uganda: Randomized Controlled Trial (e57991)
Juliet Sekandi, Esther Buregyeya, Sarah Zalwango, Damalie Nakkonde, Patrick Kaggwa, Trang Quach, David Asiimwe, Lynn Atuyambe, Kevin Dobbin. 884

Effectiveness of the Better Operative Software Outcomes Tool for Monitoring Cataract Surgery Outcomes in Inhambane Province, Mozambique: Prospective Pilot Diagnostic Accuracy Study (e64500)
Mónica Lecumberri, Dórcia Mandlate, Vasco Cote, Montserrat Martín-Baranera, Carlos Moser. 902

The Effectiveness of an Artificial Intelligence–Based Gamified Intervention for Improving Maternal Health Outcomes Among Refugees and Underserved Women in Lebanon: Community Interventional Trial (e65599)
Shadi Saleh, Nour El Arnaout, Nadine Sabra, Asmaa El Dakdouki, Zahraa Chamseddine, Randa Hamadeh, Abed Shanaa, Mohamad Alameddine. 912

The Impact of an Adaptive mHealth Intervention on Improving Patient-Provider Health Care Communication: Secondary Analysis of the DIAMANTE Trial (e64296)
Lynn Leng, Marvyn Arévalo Avalos, Adrian Aguilera, Courtney Lyles. 930

XSL•FO
RenderX

Validity, Accuracy, and Safety Assessment of an Aerobic Interval Training Using an App-Based Prehabilitation Program (PROTEGO MAXIMA Trial) Before Major Surgery: Prospective, Interventional Pilot Study (e55298)
Sara Faqar Uz Zaman, Svenja Sliwinski, Lisa Mohr-Wetzel, Julia Dreilich, Natalie Filmann, Charlotte Detemble, Dora Zmuc, Felix Chun, Wojciech Derwich, Waldemar Schreiner, Wolf Bechstein, Johannes Fleckenstein, Andreas Schnitzbauer. 938

Integrating a Multimodal Digital Device for Continuous Perioperative Monitoring in Patients With Lung Cancer Undergoing Thoracic Surgery: Development and Usability Study (e69512)
Runchen Wang, Jianqi Zheng, Wenwei Guo, Haiqi Huang, Qixia Wang, Yihong Li, Manwan Lin, Linchong Huang, Qing Zhang, Kaishen Chen, Zhiming Ye, Hongsheng Deng, Yu Jiang, Yuechun Lin, Yi Feng, Ying Huang, Ying Chen, Jianxing He, Hengrui Liang. 956

Obtaining Patient-Reported Outcome Data via a Home Patient Monitoring App: Development, Implementation, and Validation of Novel Interface Terminology (e65504)
Lucia Sacchi, Giordano Lanzola, Silvana Quaglini, Nicole Veggiotti, Silvia Panzarasa, Valentina Tibollo, Matteo Terzaghi, Itske Fraterman, Savannah Glaser, Manuel Ottaviano, Vadzim Khadakou, Vitali Hisko, Mor Peleg, Sofie Wilgenhof, Henk Mallo, Alexandra Kogan, David Glasspool, Stephanie Medlock, Laura Del Campo, Matteo Gabetta, Mimma Rizzo, Laura Locati, Paola Gabanelli, Sara Demurtas, Andrea Premoli, Szymon Wilk. . .
9 9 2

Dynamic Assessment of Fine Motor Control and Vocalization in Parkinson Disease Through a Smartphone App: Cross-Sectional Study of Time-Severity Interaction Effects (e69028)
Youho Myong, Seo Yun, Kyudong Park, Byung-Mo Oh, Han Seo. 1008

Associations Between Daily Symptoms and Pain Flares in Rheumatoid Arthritis: Case-Crossover mHealth Study (e64889)
Ting-Chen Hsu, Belay Yimer, Pauline Whelan, Christopher Armitage, Katie Druce, John McBeth. 1021

A Digital Software Support Platform for Hyperthyroidism Management in South Korea: Markov Simulation Model-Based Cost-Effectiveness Analysis (e56738)
Jung Kim, Jaeyong Shin, Man Kim, Jae Moon. 1036

Mobile App–Assisted Self-Monitoring of Blood Glucose in Type 2 Diabetes in Ningbo, China: 12-Month Retrospective Cohort Study (e65919)
Xujia Ma, Kaushik Chattopadhyay, Miao Xu, Li Li, Jialin Li. 1050

Impact of a 24-Week Mobile App–Based Human Coaching Program on Body Composition and Lipid Metabolism in Breast Cancer Survivors With Overweight or Obesity: Single-Arm Prospective Cohort Study (e64846)
Eun-Gyeong Lee, Dong-Eun Lee, Jeongseon Kim, Jaihong Han, Seeyoun Lee, Han-Sung Kang, Eun Lee, Heejung Chae, Sung Sim, Keun Lee, Jungeun Lee, Hyun Lee, Ji Yoo, Gyung Wie, So-Youn Jung. 1061

Providing 2 Types of mHealth Interventions to Support Self-Management Among People Living With HIV: Randomized Clinical Trial (e60905)
Gwang Kim, Layoung Kim, Seoyoung Baek, Sooyoung Kwon, Ji Kim, Jun Choi, Jae-Phil Choi. 1078

Preferences for Mobile App Features to Support People Living With Chronic Heart Diseases: Discrete Choice Study (e58556)
Sumudu Hewage, Sameera Senanayake, David Brain, Michelle Allen, Steven McPhail, William Parsonage, Tomos Walters, Sanjeewa Kularatna. 1095

mHealth-Augmented Care for Reducing Depression Symptom Severity Among Patients With Chronic Pain: Exploratory, Retrospective Cohort Study (e52764)
Dan Holley, Amanda Brooks, Matthew Hartz, Sudhir Rao, Thomas Zaubler. 1113

XSL•FO
RenderX

Evaluating the Effectiveness of a Mobile App for Breast Cancer Self-Management on Self-Efficacy: Nonrandomized Intervention Trial (e63989)
Sun Kim, Da Kim, Yoonsung Jang, Min Kim, Eun-Seung Yu, Doug Han, Hee Kim. 1124

A Smartphone App Self-Management Program for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial of Clinical Outcomes (e56318)
Lisa Glynn, Eddie Moloney, Stephen Lane, Emma McNally, Carol Buckley, Margaret McCann, Catherine McCabe. 1147

Clinical Impact of Personalized Physician's Education and Remote Feedback Via a Digital Platform on Glycemic Control: Pilot Randomized Controlled Trial (e67151)
Jin Yu, Joonyub Lee, Yeoree Yang, Eun Lee, Seung-Hwan Lee, Jae-Hyoung Cho. 1172

The Effect of a Mobile App (eMOM) on Self-Discovery and Psychological Factors in Persons With Gestational Diabetes: Mixed Methods Study (e60855)
Sini Määttänen, Saila Koivusalo, Hanna Ylinen, Seppo Heinonen, Mikko Kytö. 1188

A Culturally and Linguistically Tailored Intervention to Improve Diabetes-Related Outcomes in Chinese Americans With Type 2 Diabetes: Pilot Randomized Controlled Trial (e78036)
Jing Liu, Jiepin Cao, Yun Shi, Mary Sevick, Nadia Islam, Naumi Feldman, Huilin Li, Chan Wang, Yanan Zhao, Kosuke Tamura, Natalie Levy, Nan Jiang, Ziqiang Zhu, Yulin Wang, Jia Hong, Lu Hu. 1207

Effects of a Transtheoretical Model–Based mHealth Intervention on Transition Readiness in Adolescents With Epilepsy: Quasi-Experimental Study (e70085)
Qing Xia, Shuangzi Li, Ting Wang, Mingping Fan, Jie Xia, Lingling Xie, Huaying Yin. 1221

Transdiagnostic Cognitive Control Training for Patients Waiting for Outpatient Psychotherapy: Randomized Clinical Trial (e65867)
Maximilian Blomberg, Lisa Oberender, Ernst Koster, Timo Brockmeyer. 1237

Smartphone-Based Approach-Avoidance Bias Modification Training for Depression: Randomized Clinical Trial (e69033)
Maximilian Blomberg, Hilmar Zech, Maximilian Kluge, Nico Böhmert, Helmut Platte, Timo Brockmeyer. 1257

How Engagement Changes Over Time in a Digital Eating Disorder App: Observational Study (e68824)
Rachael Flatt, Laura Thornton, Jenna Tregarthen, Stuart Argue, Cynthia Bulik. 1271

Hybrid Health IT and Telehealth–Delivered Behavioral Weight Loss Services for Primary Care Patients With Cardiovascular Risk Factors: Intervention Component Design and Pragmatic Randomized Feasibility Trial (e58722)
Ronald Ackermann, Kenzie Cameron, David Liss, Nancy Dolan, Cassandra Aikman, Amy Carson, Sterling Harris, Kathryn Doyle, Andrew Cooper, Brian Hitsman. 1285

Evaluating a Mobile Digital Therapeutic for Vasomotor and Behavioral Health Symptoms Among Women in Midlife: Randomized Controlled Trial (e58204)
Jennifer Duffecy, Arfa Rehman, Scott Gorman, Yong Huang, Heide Klumpp. 1298

An App-Based Cognitive Behavioral Therapy Program Tailored for College Students: Randomized Controlled Trial (e50006)
Min Kim, Yeon-kug Moon, Kyong-Mee Chung. 1315

Assessing the Efficacy of the INTELLECT Cognitive Behavioral Therapy Mobile App for Anxiety and Depressive Symptoms Among At-Risk Japanese Employees: Randomized Controlled Trial (e60871)
Kengo Yokomitsu, Riki Oimatsu, Sean Toh, Oliver Sündermann. 1333

XSL•FO
RenderX

Effect of a Cognitive Behavioral Therapy–Based AI Chatbot on Depression and Loneliness in Chinese University Students: Randomized Controlled Trial With Financial Stress Moderation (e63806)
Yahui Wang, Xuhong Li, Qiaochu Zhang, Dannii Yeung, Yihan Wu. . . . 1347

Effectiveness and Implementation Outcomes of an mHealth App Aimed at Promoting Physical Activity and Improving Psychological Distress in the Workplace Setting: Cluster-Level Nonrandomized Controlled Trial (e70473)
Kazuhiro Watanabe, Mitsuhiro Sato, Shoichi Okusa, Akizumi Tsutsumi. . . . 1366

Feasibility and Effects of a Gait Assistance and Gait Resistance Training Program Using a Walking-Assist Wearable Robot for Community-Dwelling Older Adults: Single-Group, Pre-, and Posttest Study (e58142)
Eunhee Cho, Sinwoo Hwang, Seok-Jae Heo, Bokman Lim, Jewoo Lee, Younbaek Lee. . . . 1381

Behavioral Markers in Older Adults During COVID-19 Confinement: Secondary Analysis of In-Home Sensor Data (e56678)
Knoo Lee, Noah Marchal, Erin L Robinson, Kimberly Powell. . . . 1393

Comparative Effectiveness of Wearable Devices and Built-In Step Counters in Reducing Metabolic Syndrome Risk in South Korea: Population-Based Cohort Study (e64527)
Kyung-In Joung, Sook An, Joon Bang, Kwang Kim. . . . 1403

Assessment of Environmental, Sociocultural, and Physiological Influences on Women's Toileting Decisions and Behaviors Using "Where I Go": Pilot Study of a Mobile App (e56533)
Abigail Smith, Elizabeth Mueller, Cora Lewis, Alayne Markland, Caroline Smerdon, Ariana Smith, Siobhan Sutcliffe, Jean Wyman, Lisa Low, Janis Miller, The Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium. . . . 1420

An Explanation Interface for Healthy Food Recommendations in a Real-Life Workplace Deployment: User-Centered Design Study (e51271)
Robin De Croon, Daniela Segovia-Lizano, Paul Finglas, Vero Vanden Abeele, Katrien Verbert. . . . 1436

The Effects of Self-Monitoring Using a Smartwatch and Smartphone App on Stress Awareness, Self-Efficacy, and Well-Being–Related Outcomes in Police Officers: Longitudinal Mixed Design Study (e60708)
Herman de Vries, Roos Delahaij, Marianne van Zwieten, Helen Verhoef, Wim Kamphuis. . . . 1449

A Mobile App–Based Gratitude Intervention's Effect on Mental Well-Being in University Students: Randomized Controlled Trial (e53850)
Chloë Fuller, Silvia Marin-Dragu, Ravishankar Iyer, Sandra Meier. . . . 1464

Smoking Cessation Smartphone App for Nondaily Smoking With Telephone Onboarding: Proof-of-Concept Randomized Controlled Trial (e53971)
Bettina Hoeppner, Kaitlyn Siegel, Allison Futter, Diadora Finley-Abboud, Alivia Williamson, Christopher Kahler, Elyse Park, Susanne Hoeppner.
1 4 7 5

Proximal Effects of a Just-in-Time Adaptive Intervention for Smoking Cessation With Wearable Sensors: Microrandomized Trial (e55379)
Christine Vinci, Steve Sutton, Min-Jeong Yang, Sarah Jones, Santosh Kumar, David Wetter. . . . 1496

Development of a Mobile Intervention for Procrastination Augmented With a Semigenerative Chatbot for University Students: Pilot Randomized Controlled Trial (e53133)
Seonmi Lee, Jaehyun Jeong, Myungsung Kim, Sangil Lee, Sung-Phil Kim, Dooyoung Jung. . . . 1510

Exploring the Potential of a Digital Intervention to Enhance Couple Relationships (the Paired App): Mixed Methods Evaluation (e55433)
Catherine Aicken, Jacqui Gabb, Salvatore Di Martino, Tom Witney, Mathijs Lucassen. . . . 1532

XSL•FO
RenderX

Assessment of an App-Based Sleep Program to Improve Sleep Outcomes in a Clinical Insomnia Population: Randomized Controlled Trial (e68665)
Walter Staiano, Christine Callahan, Michelle Davis, Leah Tanner, Chelsea Coe, Sarah Kunkle, Ulrich Kirk. . . . 1550

Promoting Hand Hygiene During the COVID-19 Pandemic: Randomized Controlled Trial of the Optimized Soapp+ App (e57191)
Dario Baretta, Carole Rüttimann, Melanie Amrein, Jennifer Inauen. . . . 1564

The Impact of Technology-Enabled Medical Nutrition Therapy on Weight Loss in Adults With Overweight and Obesity: Retrospective Observational Study (e70228)
Emily Hu, Tommy Kelley, Ajay Haryani. . . . 1580

Outdoor Exercise Facility–Based Integrative Mobile Health Intervention to Support Physical Activity, Mental Well-Being, and Exercise Self-Efficacy Among Older Adults With Prefrailty and Frailty in Hong Kong: Pilot Feasibility Randomized Controlled Trial Study (e69259)
Janet Lee, Arnold Wong, Peter Ng, S Fu, Kenneth Fong, Andy Cheng, Karen Lee, Rui Sun, Hao Zhang, Rong Xiao. . . . 1592

The Association of Previous Day Carbohydrate Consumption With Fasted, Exhaled Carbon Dioxide in Lumen Users: Retrospective Real-World Study (e64604)
Shlomo Yeshurun, Tomer Cramer, Daniel Souroujon, Merav Mor. . . . 1613

Efficacy of a Self-Guided Transdiagnostic Intervention for Adults With Anxiety and Depression: Randomized Controlled Trial (e79759)
Walter Staiano, Christine Callahan, Michelle Davis, Leah Tanner, Sarah Kunkle, Jenna Glover, Jon Kole, Neeru Bakshi, Marco Romagnoli, Ulrich Kirk. . . . 1623

Requirements and Concerns of Individuals Remitted From Depression for an Early Relapse Detection mHealth App: Focus Group Study (e67141)
Tina Coenen, Matthias Maerevoet, Stephanie Chen, Mathias De Brouwer, Sofie Van Hoecke, Ernst Koster, Mariek Vanden Abeele, Klaas Bombeke. . . . 1636

Pilot Study of a Mobile, Virtual Reality–Based Digital Therapeutic for Smoking Cessation: Randomized Controlled Trial (e66411)
Yeong-Seon Jo, Arom Pyeon, Min-Kyung Hu, SungMin Kim, In-Young Choi, Dai-Jin Kim, Ji-Won Chun. . . . 1656

Unraveling the Factors Associated With Digital Health Intervention Uptake: Cross-Sectional Study (e63896)
Ilona Ruotsalainen, Mikko Valtanen, Riikka Kärsämä, Adil Umer, Hilkka Liedes, Suvi Parikka, Annamari Lundqvist, Kirsikka Aittola, Suvi Koivunen, Jussi Pihlajamäki, Anna-Leena Vuorinen, Jaana Lindström. . . . 1677

Effectiveness of Internet-Based Cognitive Behavioral Therapy for Depressive Symptoms During Pregnancy by Using Real-World Data: Retrospective Cohort Study (e73512)
Asuka Takae, Natsu Sasaki, Kotaro Imamura, Daisuke Nishi. . . . 1696

Text Messaging Versus Postal Reminders to Improve Participation in a Colorectal Cancer Screening Program: Randomized Controlled Trial (e64243)
Nuria Vives, Gemma Binefa, Noemie Travier, Albert Farre, Jon Panera, Berta Casas, Carmen Vidal, Gemma Ibáñez-Sanz, Montse Garcia, M-TICS research group. . . . 1706

Evaluation of a Pilot mHealth Intervention to Engage Primary Care Clients at an Urban Clinic Serving Marginalized Populations: Mixed-Methods Cohort Study (e68561)
Lauren Harrison, Christina Fulton, Antonio Marante Changir, Adedayo Ajidahun, Cassidy Tam, Wendy Zhang, Tian Shen, Rolando Barrios, Julio Montaner, Kate Salters, Richard Lester, Surita Parashar, David Moore. . . . 1715

Implementation and Evaluation of a Virtual Transitional Care Intervention Using Automated Text Messaging and Virtual Visits After Emergency Department Discharges: Retrospective Cohort Study (e77973)
Grace Lee, Courtenay Bruce, Tariq Nisar, Brendan Holderread, Sarah Pletcher, Ngoc Nguyen. . . . 1726

XSL•FO
RenderX

Effectiveness of an mHealth Intervention With Short Text Messages to Promote Treatment Adherence Among HIV-Positive Mexican Adults: Randomized Controlled Trial (e57540)
Luis Del Moral Trinidad, Jaime Andrade Villanueva, Pedro Martínez Ayala, Rodolfo Cabrera Silva, Melva Herrera Godina, Luz González-Hernández. . . . 1740

Factors Influencing Engagement in a Digital Substance Use Prevention Program: Qualitative Study Using the Capability, Opportunity, Motivation Model and Theoretical Domains Framework (e64861)
Nikolai Kiselev, Rebecca Schaffner, Zsofia Csepregi, Andreas Wenger, Raquel Paz Castro, Severin Haug. . . . 1749

WhatsApp-Based Coaching Program to Support Smoking and Vaping Cessation Among Young People: Pre-Post Study on Acceptance and Preliminary Efficacy (e65301)
Severin Haug, Lisa Caputo, Andreas Wenger, Nikolai Kiselev, Olivia Studhalter, Michael Schaub. . . . 1763

mHealth-Based Diabetes Prevention Program for Chinese Mothers With Abdominal Obesity: Randomized Controlled Trial (e47837)
Qinyuan Huang, Qinyi Zhong, Yanjing Zeng, Yimeng Li, James Wiley, Man Wang, Jyu-Lin Chen, Jia Guo. . . . 1778

A Tailored mHealth Intervention for Improving Antenatal Care Seeking and Its Determinants Among Pregnant Adolescent Girls and Young Women in South Africa: Pilot Randomized Controlled Trial (e59144)
Ronel Sewpaul, Ken Resnicow, Rik Crutzen, Natisha Dukhi, Priscilla Reddy. . . . 1794

Cerina—Cognitive Behavioral Therapy–Based Mobile App for Managing Generalized Anxiety Disorder Symptoms Among University Students: Results From a Pilot Feasibility Randomized Controlled Trial (e70691)
Ozlem Eylem-van Bergeijk, Tony Robinson, Matthew Manktelow, Michail Olympios, Siobhan Poulter, Prasannajeet Mane, Maria Panagioti, Joan Condell, Gerard Leavey. . . . 1816

Participant Evaluation of Blockchain-Enhanced Women's Health Research Apps: Mixed Methods Experimental Study (e65747)
Madelena Ng, Jodi Halpern, Olivia Shane, Tina Teng, Michael Nguy n, Casey Alt, Anaïs Leite, Sean Moss-Pultz, Courtney Lyles, Coye Cheshire. . . . 1836

Investigating the Quality of Mobile Apps for Drug-Drug Interaction Management Using the Mobile App Rating Scale and K-Means Clustering: Systematic Search of App Stores (e65927)
Ayush Bhattacharya, Jose Florez-Arango. . . . 1850

Evaluation of Mobile Intermittent Fasting Applications in Chinese App Stores: Quality Evaluations and Content Analysis (e66339)
Laihao Fang, Cheng Huang, Bing Lin, Kuanlin Lei, Jiazhen Zhou, Xiaoni Zhong, Yanbing Liu, Jiaxiu Liu. . . . 1868

Mobile Apps for HIV and Sexually Transmitted Infection Prevention in Canada, Mexico, and the United States: Environmental Scan (e72009)
Higinio Fernández-Sánchez, Javier Salazar-Alberto, Jhan Fernández-Delgado, Annalynn Galvin, Michael Mugavero, Carlos Rodriguez-Diaz, Diane Santa Maria. . . . 1884

Smartphone Apps for Cardiovascular and Mental Health Care: Digital Cross-Sectional Analysis (e63642)
Manjot Singh, Julian Herpertz, Noy Alon, Sarah Perret, John Torous, Daniel Kramer. . . . 1898

Trustworthiness of Web-Based Pharmacy Apps in Pakistan Based on the Mobile App Rating Scale: Content Analysis and Quality Evaluation (e59884)
Anum Sattar, Hina Rehman, Safila Naveed, Sumaira Khadim, Nargis Khan, Ahmad Kazi, Wajid Syed, Mahmood Al-Rawi, Shazia Jamshed. .
1 9 2 0

XSL•FO
RenderX

mHealth Apps Available in Italy to Support Health Care Professionals in Antimicrobial Stewardship Implementation: Systematic Search in App Stores and Content Analysis (e51122)
Giuseppa Russo, Annachiara Petrazzuolo, Marino Trivisani, Giuseppe Virone, Elena Mazzolini, Davide Pecori, Assunta Sartor, Sergio Intini, Stefano Celotto, Rossana Roncato, Roberto Cocconi, Luca Arnoldo, Laura Brunelli. . . . 1937

Review and Comparative Evaluation of Mobile Apps for Cardiovascular Risk Estimation: Usability Evaluation Using mHealth App Usability Questionnaire (e56466)
Adrijana Svenšek, Lucija Gosak, Mateja Lorber, Gregor Štiglic, Nino Fija ko. . . . 1950

Artificial Intelligence–Based Mobile Phone Apps for Child Mental Health: Comprehensive Review and Content Analysis (e58597)
Fan Yang, Jianan Wei, Xuejun Zhao, Ruopeng An. . . . 1964

Users' Needs for Mental Health Apps: Quality Evaluation Using the User Version of the Mobile Application Rating Scale (e64622)
Siyeon Ko, Hyekyung Woo. . . . 1976

Cognitive Training Mobile Apps for Older Adults With Cognitive Impairment: App Store Search and Quality Evaluation (e69637)
Leyi Wu, Jiajuan Pan, Chuwen Dou, An Gu, An Huang, Hong Tao, Xiaoyan Wang, Chen Zhang, Lina Wang. . . . 1990

Grouping Digital Health Apps Based on Their Quality and User Ratings Using K-Medoids Clustering: Cross-Sectional Study (e57279)
Maciej Marek Zych, Raymond Bond, Maurice Mulvenna, Lu Bai, Jorge Martinez-Carracedo, Simon Leigh. . . . 2004

Identifying and Evaluating Mobile and Web Apps for Patients to Manage Hidradenitis Suppurativa: Systematic Search in App Stores and Content Analysis (e69030)
Caroline Glatzel, Tassilo Dege, Bernadette Glatzel, Matthias Goebeler, Dagmar Presser, Astrid Schmieder. . . . 2018

Application of Behavior Change Techniques and Rated Quality of Smoking Cessation Apps in China: Content Analysis (e56296)
Qiumian Hong, Shuochi Wei, Hazizi Duoliken, Lefan Jin, Ning Zhang. . . . 2027

Digital Health Resilience and Well-Being Interventions for Military Members, Veterans, and Public Safety Personnel: Environmental Scan and Quality Review (e64098)
Rashell Allen, Myrah Malik, Carley Aquin, Lucijana Herceg, Suzette Brémault-Phillips, Phillip Sevigny. . . . 2039

Evaluating Characteristics and Quality of Mental Health Apps Available in App Stores for Indian Users: Systematic App Search and Review (e79238)
Seema Mehrotra, Ravikesh Tripathi, Pramita Sengupta, Abhishek Karishiddimath, Angelina Francis, Pratiksha Sharma, Paulomi Sudhir, T Srikanth, Girish Rao, Rajesh Sagar. . . . 2057

Selection of Behavior Change Techniques for Asthma Medication Adherence Apps: Evidence-Based Design Study (e49348)
Alison Wright, Jeremy Holland, Iain Simpson, Samantha Walker, Naomi Bennett-Steele, John Weinman. . . . 2076

Effectiveness of a Digital Peer-Supported App Intervention in Promoting Smoking Cessations: Nonrandomized Controlled Trial (e68638)
Shota Yoshihara, Kayoko Takahashi, Chiaki Uemura, Shin Murakami, Daichi Harada, Hiroshi Yamato. . . . 2088

Effects of Information Length and Implementation Intentions on Adherence to Weight Management Strategies: Experimental Study (e65260)
Khaleda Ahmadyar, Joanna Szypula, Angeliki Bogosian, Katy Tapper. . . . 2099

XSL•FO
RenderX

User Archetypes of a Well-Being–Promoting Mobile App Among Adults: Cross-Sectional Study and Cluster Analysis of Usage Patterns (e68982)
Hanna Rekola, Tommi Tolmunen, Elina Mattila, Juho Strömmer, Timo Lakka, Helena Länsimies, Tomi Mäki-Opas. 2113

Using Social Media to Recruit a Diverse Sample of Participants for a Mobile Health (mHealth) Intervention to Increase Physical Activity: Exploratory Study (e56329)
Laura Pathak, Rosa Hernandez-Ramos, Karina Rosales, Jose Miramontes-Gomez, Faviola Garcia, Vivian Yip, Suchitra Sudarshan, Anupama Cemballi, Courtney Lyles, Adrian Aguilera. 2131

Combined Use of Digital and Analog Physical Therapy in Patients With Musculoskeletal Disorders and Indicators of Chronicity: German Claims Data Analysis (e63935)
Silke Frey, Annika Schmitz, Udo Schneider, Linda Kerkemeyer, Birgitta Weltermann. 2145

Sociodemographic Differences in Logins and Engagement With the Electronic Health Coach Messaging Feature of a Mobile App to Support Opioid and Stimulant Use Recovery: Results From a 1-Month Observational Study (e54753)
Lindsey Filiatreau, Hannah Szlyk, Alex Ramsey, Erin Kasson, Xiao Li, Zhuoran Zhang, Patricia Cavazos-Rehg. 2159

Prevention Needs and Target Behavior Preferences in an App-Based Addiction Prevention Program for German Vocational School Students: Cluster Randomized Controlled Trial (e59573)
Diana Guertler, Elaine Kraft, Dominic Bläsing, Anne Möhring, Christian Meyer, Hannah Schmidt, Florian Rehbein, Merten Neumann, Arne Dreißigacker, Anja Bischof, Gallus Bischof, Svenja Sürig, Lisa Hohls, Susanne Wurm, Stefan Borgwardt, Severin Haug, Hans-Jürgen Rumpf.
2 1 7 1

Parental and Demographic Predictors of Engagement in an mHealth Intervention: Observational Study From the Let's Grow Trial (e60478)
Johanna Sandborg, Brittany Markides, Savannah Simmons, Katherine Downing, Jan Nicholson, Liliana Orellana, Harriet Koorts, Valerie Carson, Jo Salmon, Kylie Hesketh. 2190

Health Motivation as a Predictor of mHealth Engagement Across BMI: Cross-Sectional Survey (e71625)
Shao-Hsuan Chang, Lung-Kun Yeh, Daishi Chen, Kae-Kuen Hu, Mei-Ching Yu, Ching-Mao Chang. 2206

Interpretable Machine Learning Models for Analyzing Determinants Affecting the Use of mHealth Apps Among Family Caregivers of Patients With Stroke in Chinese Communities: Cross-Sectional Survey Study (e73903)
Yun Du, Jun-Ying Fan, Guang-Zhi Liu, Zi-Yue Yang, Yang Lei, Yu-Fang Guo. 2224

Clinical Efficacy of Multimodal Exercise Telerehabilitation Based on AI for Chronic Nonspecific Low Back Pain: Randomized Controlled Trial (e56176)
Chongwu Xiao, Yijin Zhao, Gege Li, Zhuodong Zhang, Siyu Liu, Weichao Fan, Jinjing Hu, Qiuru Yao, Chengduan Yang, Jihua Zou, Qing Zeng, Guozhi Huang. 2240

Community-Based Intelligent Blood Glucose Management for Older Adults With Type 2 Diabetes Based on the Health Belief Model: Randomized Controlled Trial (e60227)
Anqi Zhang, Jinsong Wang, Xiaojuan Wan, Ziyi Zhang, Shuhan Zhao, Shuo Bai, Yamin Miao, Shuang Yang, Xue Jiang. 2254

Impact of 12-Month mHealth Home Telemonitoring on Clinical Outcomes in Older Individuals With Hypertension and Type 2 Diabetes: Multicenter Randomized Controlled Trial (e59733)
Matic Mihevc, Majda Mori Lukan i , rt Zavrnik, Tina Virti Poto nik, Nina Ruži Gorenjec, Marija Petek Šter, Zalika Klemenc-Ketiš, Antonija Poplas Susi 2266

XSL•FO
RenderX

Barriers and Facilitators in Implementing a Telemonitoring Application for Patients With Chronic Kidney Disease and Health Professionals: Ancillary Implementation Study of the NeLLY (New Health e-Link in the Lyon Region) Stepped-Wedge Randomized Controlled Trial (e50014)
Marion Delvallée, Abdallah Guerraoui, Lucas Tchetgnia, Jean-Pierre Grangier, Nassira Amamra, Anne-Laure Camarroque, Julie Haesebaert, Agnès Caillette-Beaudoin. . . . 2283

Impact of mHealth on Postoperative Quality of Life, Self-Management, and Dysfunction in Patients With Oral and Maxillofacial Tumors: Nonrandomized Controlled Trial (e59926)
Yufei Li, Yueping Wang, Yifan Wu, Hong Yu, Hua Yao, Yuqun Wang, Yulan Yin, Lan Wang, Lili Hou. . . . 2299

Outcomes of an App-Based Intervention to Target Naming Among Individuals With Poststroke Aphasia: Virtual Randomized Controlled Trial (e67711)
Esther Kim, Laura Laird, Carlee Wilson, Steven Stewart, Philip Mildner, Sebastian Möller, Raimund Schatz, Robert Spang, Jan-Niklas Voigt-Antons, Elizabeth Rochon. . . . 2312

Smartphone App–Guided Pulmonary Rehabilitation in Chronic Respiratory Diseases: Randomized Controlled Trial (e64884)
Chiwook Chung, Deog Kim, Jung-Kyu Lee, Eun Heo, Hee Kwon, Dongbum Kim, Woo Kim, Sei Lee. . . . 2325

Effectiveness of Smartwatch Device on Adherence to Home-Based Cardiac Rehabilitation in Patients With Coronary Heart Disease: Randomized Controlled Trial (e70848)
Sisi Zhang, Yuehui Wang, Jiahui Wu, Changsheng Ma, Xiaoping Meng. . . . 2337

Effect of Lifestyle Modification Through Web-Based Telerehabilitation Monitoring Combined With Supervised Sensorimotor Training After Total Knee Arthroplasty: Randomized Controlled Trial (e64643)
Samreen Sadiq, Rabiya Noor, Rizwan Akram. . . . 2351

Innovating Care for Postmenopausal Women Using a Digital Approach for Pelvic Floor Dysfunctions: Prospective Longitudinal Cohort Study (e68242)
Ana Pereira, Dora Janela, Anabela Areias, Maria Molinos, Xin Tong, Virgílio Bento, Vijay Yanamadala, Jennesa Atherton, Fernando Dias Correia, Fabíola Costa. . . . 2369

The Effectiveness of the Be Prepared mHealth App on Recovery of Physical Functioning After Major Elective Surgery: Multicenter Randomized Controlled Trial (e58703)
Miriam van der Velde, Marike van der Leeden, Edwin Geleijn, Cindy Veenhof, Karin Valkenet, Be Prepared Group. . . . 2385

Efficacy of a Digital Postoperative Rehabilitation Intervention in Patients With Primary Liver Cancer: Randomized Controlled Trial (e59228)
Kaitao Yu, Baobing Yin, Ying Zhu, Hongdao Meng, Wenwei Zhu, Lu Lu, Junqiao Wang, Shugeng Chen, Jun Ni, Yifang Lin, Jie Jia. . . . 2404

Clinical, Psychological, Physiological, and Technical Parameters and Their Relationship With Digital Tool Use During Cardiac Rehabilitation: Comparison and Correlation Study (e57413)
Fabian Wiesmüller, David Haag, Mahdi Sareban, Karl Mayr, Norbert Mürzl, Michael Porodko, Christoph Puelacher, Lisa-Marie Moser, Marco Philippi, Heimo Traninger, Stefan Höfer, Josef Niebauer, Günter Schreier, Dieter Hayn. . . . 2417

Effects of a Computer Vision–Based Exercise Application for People With Knee Osteoarthritis: Randomized Controlled Trial (e63022)
Dian Zhu, Jianan Zhao, Tong Wu, Beiyao Zhu, Mingxuan Wang, Ting Han. . . . 2429

Wearable Devices in Remote Cardiac Rehabilitation With and Without Weekly Online Coaching for Patients With Coronary Artery Disease: Randomized Controlled Trial (e63797)
Ryota Nishio, Tomotaka Dohi, Miho Yokoyama, Taisuke Nakade, Norihito Takahashi, Yuichi Chikata, Hirohisa Endo, Hiroki Nishiyama, Iwao Okai, Hiroshi Iwata, Shinya Okazaki, Katsumi Miyauchi, Hiroyuki Daida, Tohru Minamino. . . . 2443

Digital Therapeutics–Based Cardio-Oncology Rehabilitation for Lung Cancer Survivors: Randomized Controlled Trial (e60115)
Guangqi Li, Xueyan Zhou, Junyue Deng, Jiao Wang, Ping Ai, Jingyuan Zeng, Xuelei Ma, Hu Liao. . . . 2460

XSL•FO
RenderX

Recovery of Patient-Reported Outcome Measures vs Gait Parameters Obtained by Instrumented Insoles After Tibial and Malleolar Fractures: Prospective Longitudinal Observational Study (e71022)
Elke Warmerdam, Marianne Huebner, Caroline Stoll, Andrey Lange, Bergita Ganse. 2473

Efficacy of a Mobile Multidisciplinary Digital Therapeutics App for Patellofemoral Pain: Randomized Controlled Trial (e69627)
Sanghee Lee, Chan Yoon, Chi-Hyun Choi, Tae Park, Sang Yang, Ha Cha, Tae Kim, Jae Park, Moon Chang, Chong Chang. 2486

Completion and Compliance Rates for an Intensive mHealth Study Design to Promote Self-Awareness and Self-Care Among Care Partners of Individuals With Traumatic Brain Injury: Secondary Analysis of a Randomized Controlled Trial (e73772)
Noelle Carlozzi, Jonathan Troost, Wendy Lombard, Jennifer Miner, Christopher Graves, Sung Choi, Zhenke Wu, Srijan Sen, Angelle Sander.
2 4 9 9

Current Practice and Expert Perspectives on Cultural Adaptations of Digital Health Interventions: Qualitative Study (e59965)
Vasileios Nittas, Sarah Chavez, Paola Daniore. 2519

Patterns of Engagement With the mHealth Component of a Sexual and Reproductive Health Risk Reduction Intervention for Young People With Depression: Latent Trajectory Analysis (e70219)
Lydia Shrier, Carly Milliren, Brittany Ciriello, Madison O'Connell, Sion Harris. 2533

Integrating CEN ISO/TS 82304-2 in the Catalan Health App Assessment Framework: Comparative Case Study (e67858)
Berta Llebot Casajuana, Petra Hoogendoorn, Maria Villalobos-Quesada, Carme Pratdepàdua Bufill. 2545

Leveraging Social Media and Crowdsourcing to Recruit and Retain Military Veterans With Posttraumatic Stress Disorder or Experience of Harmful Gambling for mHealth Interventions: Descriptive Study (e73706)
Conor Heath, Jess Williams, Daniel Leightley, Dominic Murphy, Simon Dymond. 2564

Cluster-Based Predictive Modeling of User Ratings for Physical Activity Apps Using Mobile App Rating Scale (MARS) Dimensions: Model Development and Validation (e70987)
Ayush Bhattacharya, Jose Florez-Arango. 2575

Effects of Heat Adaptation Behaviors on Resting Heart Rate Response to Summer Temperatures in Older Adults: Wearable Device Panel Study (e67721)
Chi-Hsien Chen, Feipei Lai, Yu-Lin Chen, Yue Guo. 2593

Reliability and Accuracy of the Fitbit Charge 4 Photoplethysmography Heart Rate Sensor in Ecological Conditions: Validation Study (e54871)
Maxime Ceugniez, Hervé Devanne, Eric Hermand. 2616

Trade-Offs Between Simplifying Inertial Measurement Unit–Based Movement Recordings and the Attainability of Different Levels of Analyses: Systematic Assessment of Method Variations (e58078)
Manu Airaksinen, Okko Räsänen, Sampsa Vanhatalo. 2626

A New Wearable System for Personal Air Pollution Exposure Estimation: Pilot Observational Study (e60426)
Sara Bernasconi, Alessandra Angelucci, Andrea Rossi, Andrea Aliverti. 2642

Exploring the Value of Continuous Plantar Temperature Monitoring for Diabetic Foot Health Management: Observational, Prospective Cohort Study (e73187)
Maryam Hajizadeh, Emily Matijevich, Emily Bray, Evan Minty, Brock Liden. 2659

XSL•FO
RenderX

Temporal Trajectories in Sleep, Temperature Trends, Cardiorespiratory, and Activity Metrics Measured via Oura Ring During Pregnancy: Large-Scale Observational Analysis (e80213)
Rebecca Adaimi, Nina Thigpen, Alicia Clausel, Neta Gotlieb, Ketan Patel, Massimiliano de Zambotti. . . . 2672

Multimodal Sleep Measurement and Alignment Analysis in Outpatients With Major Depressive Episode: Observational Study (e82465)
Afrooz Mahir, Nguyen Luong, Ilya Baryshnikov, Annasofia Martikkala, Erkki Isometsä, Talayeh Aledavood. . . . 2689

A Multiple Technology–Based Physical Activity Intervention for Latina Adolescents: Results From the Chicas Fuertes Randomized Controlled Trial (e71623)
Jacob Carson, Emily Greenstadt, Brittany Olivera, Shira Dunsiger, Michelle Zive, Michael Higgins, Job Godino, Bess Marcus, Dawn Meyer, Britta Larsen. . . . 2711

Minimal Clinically Important Difference of Average Daily Steps Measured Through a Consumer Smartwatch in People With Mild-to-Moderate Parkinson Disease: Cross-Sectional Study (e64213)
Edoardo Bianchini, Marika Alborghetti, Silvia Galli, Clint Hansen, Alessandro Zampogna, Antonio Suppa, Marco Salvetti, Francesco Pontieri, Domiziana Rinaldi, Nicolas Vuillerme. . . . 2724

Comparison of ActiGraph CentrePoint Insight Watch Placement on Dominant and Nondominant Wrists in Young Adults in Free-Living Conditions: Observational Validation Study (e63033)
Daehyoung Lee, Haley Voermans-Dean, Jung Lee, Jong Shin, Gregory Dominick. . . . 2743

Boosting Physical Activity Among Individuals With Low Engagement Through Double-Point Incentives in a Community-Based mHealth Intervention: Retrospective Observational Study (e66227)
Jungin Joo, Mangyeong Lee, Junghee Yoon, Hyeonjin Cho, Govind Warrier, Johannes Thrul, Juhee Cho. . . . 2755

Accelerometry-Assessed Physical Activity and Circadian Rhythm to Detect Clinical Disability Status in Multiple Sclerosis: Cross-Sectional Study (e57599)
Nicole Bou Rjeily, Muraleetharan Sanjayan, Pratim Guha Niyogi, Blake Dewey, Alexandra Zambriczki Lee, Christy Hulett, Gabriella Dagher, Chen Hu, Rafal Mazur, Elena Kenney, Erin Brennan, Anna DuVal, Peter Calabresi, Vadim Zipunnikov, Kathryn Fitzgerald, Ellen Mowry. . . . 2768

Effects of Integrating Wearable Activity Trackers With a Home-Based Multicomponent Exercise Intervention on Fall-Related Parameters and Physical Function in Older Adults: Randomized Controlled Trial (e64458)
Yejin Kim, Kyung Park, Hye-Mi Noh. . . . 2785

Consumer Wearable Usage to Collect Health Data Among Adults Living in Germany: Nationwide Observational Survey Study (e59199)
Kristin Manz, Susanne Krug, Charlotte Kühnelt, Johannes Lemcke, Ilter Öztürk, Julika Loss. . . . 2800

College Community–Based Physical Activity Support at a Public University During the COVID-19 Pandemic: Retrospective Longitudinal Analysis of Intra- Versus Interpersonal Components for Uptake and Outcome Association (e51707)
Garrett Ash, Selene Mak, Adrian Haughton, Madilyn Augustine, Phillip Bodurtha, Robert Axtell, Beatrice Borsari, Jason Liu, Shaoke Lou, Xin Xin, Lisa Fucito, Sangchoon Jeon, Matthew Stults-Kolehmainen, Mark Gerstein. . . . 2812

Effect of Mobile Augmented Reality Counseling on Improving Shared Decision-Making in Thoracic Surgery: Randomized Clinical Crossover Trial (e79632)
Ying-Shian Chen, Yi-Chen Hsu, Worachate Romalee, Ding-Han Wang, Jennifer Lai, Tsai-Wang Huang, Kuan Lin. . . . 2831

Preferences and Willingness to Pay for Health App Assessments Among Health Care Stakeholders: Discrete Choice Experiment (e57474)
Anna-Lena Frey, Simon Leigh, Carla Toro, Carme Pratdepàdua Bufill, Charles McCay, Tatjana Pren a Trupec, Giuseppe D'Avenio, Menno Kok, Antanas Montvila, Philipp Goedecker, Petra Hoogendoorn. . . . 2843

XSL•FO
RenderX

Economic Evaluation of a Multicomponent mHealth Intervention for Stroke Management in Rural China: Cluster-Randomized Trial With 6-Year Follow-Up (e75326)
Bolu Yang, Enying Gong, Xingxing Chen, Jie Tan, Nicholas Peoples, Yuhan Li, Jiayu Cai, Yan Li, Brian Oldenburg, Chen Chen, Dejin Dong, Xiaochen Zhang, Eric Finkelstein, Lei Si, Lijing Yan. . . . 2862

Utility of Digital Phenotyping Based on Wrist Wearables and Smartphones in Psychosis: Observational Study (e56185)
Zixu Yang, Creighton Heaukulani, Amelia Sim, Thisum Buddhika, Nur Abdul Rashid, Xuancong Wang, Shushan Zheng, Yue Quek, Sutapa Basu, Kok Lee, Charmaine Tang, Swapna Verma, Robert Morris, Jimmy Lee. . . . 2896

Measuring Psychological Well-Being and Behaviors Using Smartphone-Based Digital Phenotyping: An Intensive Longitudinal Observational mHealth Pilot Study Embedded in a Prospective Cohort of Women (e71375)
Li Yi, Claudia Trudel-Fitzgerald, Cindy Hu, Grete Wilt, Jorge Chavarro, Jukka-Pekka Onnela, Francine Grodstein, Laura Kubzansky, Peter James. . . . 2909

Exploring Age-Related Patterns in Smartphone Keystroke Dynamics Considering Temporal Variability: Cross-Sectional Study With AI-Based Analysis (e80094)
Junhyung Moon, Yu Huh, Hee Cho, Chaeyeon Kim, Hyeongrae Lee, Dong Jang, Baek Cho. . . . 2926

Bidirectional Associations Between Physical Activity, Sedentary Behavior, and Daily Symptoms in Patients With Chronic Obstructive Pulmonary Disease: Longitudinal Observational Study (e65653)
Banchia Palmen, Zjala Ebadi, Maarten van Herck, Yvonne Goërtz, Qichen Deng, Melissa Thong, Chris Burtin, Jeannette Peters, Roy Sprooten, Erik Bischoff, Emiel Wouters, Mirjam Sprangers, Jan Vercoulen, Sarah Houben-Wilke, Anouk Vaes, Daisy Janssen, Martijn Spruit. . . . 2939

Impacts of Environmental Distractions and Interruptions on Unsupervised Digital Cognitive Assessments in Older Adults: Cognitive Ecological Momentary Assessment Study (e71578)
Matthew Welhaf, Hannah Wilks, Andrew Aschenbrenner, Samhita Katteri, John Morris, Jason Hassenstab. . . . 2955

Understanding the Relationship Between Ecological Momentary Assessment Methods, Sensed Behavior, and Responsiveness: Cross-Study Analysis (e57018)
Diane Cook, Aiden Walker, Bryan Minor, Catherine Luna, Sarah Tomaszewski Farias, Lisa Wiese, Raven Weaver, Maureen Schmitter-Edgecombe. . . . 2968

Feature Selection for Physical Activity Prediction Using Ecological Momentary Assessments to Personalize Intervention Timing: Longitudinal Observational Study (e57255)
Devender Kumar, David Haag, Jens Blechert, Josef Niebauer, Jan Smeddinck. . . . 2983

Assessing Social Interaction and Loneliness and Their Association With Frailty Among Older Adults With Subjective Cognitive Decline or Mild Cognitive Impairment: Ecological Momentary Assessment Approach (e64853)
Bada Kang, Dahye Hong, Seolah Yoon, Chaeeun Kang, Jennifer Kim. . . . 3009

Health-Promoting Effects and Everyday Experiences With a Mental Health App Using Ecological Momentary Assessments and AI-Based Ecological Momentary Interventions Among Young People: Qualitative Interview and Focus Group Study (e65106)
Selina Hiller, Christian Götzl, Christian Rauschenberg, Janik Fechtelpeter, Georgia Koppe, Eva Wierzba, Julia Sauter, Sina Dietrich, Daniel Durstewitz, Ulrich Reininghaus, Silvia Krumm. . . . 3030

Understanding Longitudinal Ecological Momentary Assessment Completion: Results From 12 Months of Burst Sampling in the TIME Study (e67117)
Tyler Prochnow, Wei-Lin Wang, Shirlene Wang, Jixin Li, Alexander Rothman, Stephen Intille, Donald Hedeker, Genevieve Dunton. . . . 3047

XSL•FO
RenderX

Health Care Workers' Experience With a Psychological Self-Monitoring App During the COVID-19 Pandemic: Mixed Methods Study (e70412)
Lydia Khaldoun, Francois Bellemare, Christine Genest, Nicolas Bergeron, Steve Geoffrion. . . . 3063

Letters to the Editor

Clarifying Blood Indices in Patients With Ovarian Cancer (e70895)
Chengwei Lian, Yu Fan, Jinke Li. . . . 2877

Authors' Reply: Clarifying Blood Indices in Patients With Ovarian Cancer (e74931)
Jiahua Zhang. . . . 2879

Authors' Reply: Methodological Considerations in Evaluating Mental Health Apps: Ensuring Reliability and Patient Safety (e85910)
Seema Mehrotra, Ravikesh Tripathi, Pramita Sengupta, Abhishek Karishiddimath, Angelina Francis, Pratiksha Sharma, Paulomi Sudhir, T Srikanth, Girish Rao, Rajesh Sagar. . . . 2881

Methodological Considerations in Evaluating Mental Health Apps: Ensuring Reliability and Patient Safety (e85329)
Harikrishnan Balakrishna. . . . 2883

Enhancing Self-Management Support Apps for Spinal Cord Injury: The Missing Role of Caregivers (e72037)
Cheila Pessoa. . . . 2885

Corrigenda and Addendas

Correction: Clinical Efficacy of Multimodal Exercise Telerehabilitation Based on AI for Chronic Nonspecific Low Back Pain: Randomized Controlled Trial (e78188)
Chongwu Xiao, Yijin Zhao, Gege Li, Zhuodong Zhang, Siyu Liu, Weichao Fan, Jinjing Hu, Qiuru Yao, Chengduan Yang, Jihua Zou, Qing Zeng, Guozhi Huang. . . . 2887

Correction: Use and Engagement With Low-Intensity Cognitive Behavioral Therapy Techniques Used Within an App to Support Worry Management: Content Analysis of Log Data (e76573)
Paul Farrand, Patrick Raue, Earlise Ward, Dean Repper, Jonathan Baker, Patricia Areán. . . . 2889

Correction: Participant Adherence in Repeated-Dose Clinical Studies Using Video-Based Observation: Retrospective Data Analysis (e79507)
Jihong Song, Sungpil Han, Suein Choi, Jonghyuk Lim, Byeong Oh, Dongoh Shin, Seunghoon Han. . . . 2891

Correction: The Role of Environmental Factors in Technology-Assisted Physical Activity Intervention Studies Among Older Adults: Scoping Review (e83822)
Carl-Philipp Jansen, Désirée Nijland, Jean-Michel Oppert, Veysel Alcan, Kirsi Keskinen, Emmi Matikainen-Tervola, Zada Pajalic, Merja Rantakokko, Signe Tomsone, Essi-Mari Tuomola, Erja Portegijs, Erik Timmermans. . . . 2894

Research Letter

Smartphone-Based Care Platform Versus Traditional Care in Primary Knee Arthroplasty in the United States: Cost Analysis (e46047)
Jess Lonner, Ashwini Naidu-Helm, David Van Andel, Mike Anderson, Richard Ditto, Roberta Redfern, Jared Foran. . . . 3085

XSL•FO
RenderX

Should Digital Interventions for HIV Self-Testing Be Regulated with World Health Organization Prequalification?

Alex Emilio Fischer[1,2], BSc, MPH, PhD; Samanta Tresha Lalla-Edward[3], BSc, MA, PhD; Vinodh Edward[1,4], BSc, BTech, MTech, DTech; Musaed Abrahams[2], MBchB; Luke Shankland[2], BA, MBA; John de Wit[1], MSc, PhD

[1]Department of Interdisciplinary Social Science, Public Health, Utrecht University, Heidelberglaan 1 3584 CS, Utrecht, The Netherlands

[2]Aviro Health, Cape Town, South Africa

[3]Ezintsha, Faculty of Health Sciences, University of the Witswatersrand, Johannesburg, South Africa

[4]The Aurum Institute, Johannesburg, South Africa

Corresponding Author:
Alex Emilio Fischer, BSc, MPH, PhD
Department of Interdisciplinary Social Science, Public Health, Utrecht University, Heidelberglaan 1 3584 CS, Utrecht, The Netherlands

Abstract

HIV self-testing (HIVST) allows people to test for HIV outside traditional health facilities, but this presents challenges around pre- and posttest counseling, reporting results, and linking to care. Digital interventions for HIVST, a type of Software as a Medical Device (SaMD), have been shown to address these challenges, but there is currently no standardized system for regulating or approving these interventions. The World Health Organization Prequalification Program (WHOPQ) is an international regulatory body that approves vaccines, medications, and in vitro diagnostics (IVDs) for low- and middle-income countries that do not have the capacity to do their own approvals. This paper explores whether WHOPQ could be used to prequalify digital interventions for HIVST. Over half the World Health Organization (WHO) member states have national regulatory bodies for medical devices, but low- and middle-income countries, especially in Africa, do not have the capacity to regulate medical devices, let alone SaMD. This gap parallels the gap in vaccine regulation that initially led to the development of WHOPQ, and while sophisticated artificial intelligence (AI)–enabled SaMD are being developed, digital interventions for HIVST could be used as a low-risk test case for prequalification of SaMD. The WHOPQ already has a strong history with HIV; over half the WHOPQ funding is from HIV-related funders and half of all prequalified medicines and IVDs are for treatment and diagnosis of HIV; however, only 2% are manufactured in Africa. If digital interventions for HIVST become prequalified, this could improve interoperability and ensure quality, accelerating their adoption at scale. However, care must be taken to remove barriers for African developers and ensure that prequalification does not delay access to people testing for HIV.

(JMIR Mhealth Uhealth 2025;13:e60276) doi:10.2196/60276

KEYWORDS

HIV self-screening; digital health; Software as a Medical Device; WHO; vaccine regulation; vaccination; Africa; accessibility; HIV testing; LMIC; World Health Organization; low- and middle-income countries

Introduction

HIV self-testing (HIVST) allows an individual to conveniently learn their HIV status by collecting and testing their own specimen (oral swab or blood drop from a finger prick) in the privacy of their own home [1]. In 2012, the first HIVST became commercially available when the Food and Drug Administration (FDA) approved OraQuick for over-the-counter sales in the United States [2]. Since then, a growing body of evidence has supported using HIVST to complement traditional HIV testing programs toward achieving the Joint United Nations Program on HIV/AIDS (UNAIDS) 95-95-95 goals, where 95% of people living with HIV know their status, 95% of those who know their status are receiving antiretroviral therapy, and 95% of those on antiretroviral therapy are virally suppressed, by 2030 [3]. In 2015, only 2 countries integrated HIVST into their national HIV strategy, but after the World Health Organization (WHO) recommended HIVST in the 2016 Guidelines on HIV Self-Testing and Partner Notification [1], and officially prequalified OraQuick in 2017 [4], the number jumped to over 40 [5]. Five years later, there are almost 100 countries that endorse HIVST in their national HIV strategy, as well as 6 prequalified HIVST kits, including one for just US$1 [6,7]. As HIVST gained global popularity, the shift away from health care facilities was recognized to present some challenges for counseling, reporting results, and linkage to care [8,9]. In response, digital interventions for HIVST have emerged as a novel approach for closing these gaps [10-13]. Like other digital health platforms, digital interventions for HIVST use digital technology (apps, social media, messaging platforms, or the

XSL•FO
RenderX

Internet) to assist people through an HIVST journey [14]. These digital interventions have been used for demand generation to promote and distribute HIVSTs, but they are also being used to address the challenges experienced with HIVST. Evidence shows that digital interventions for HIVST can deliver standardized counseling, allow users to report results and link them to appropriate care [13].

Before the COVID-19 pandemic accelerated the familiarization of digital health [15,16], the WHO was already drafting its *Global Strategy on Digital Health 2020 - 2025* [14]. Subsequently, in 2022, UNAIDS and WHO published a policy brief on *Virtual Interventions in Response to HIV, Sexually Transmitted Infections and Viral Hepatitis* [17]. These documents reaffirm a strong commitment to digital health interventions, including for HIVST, and introduce guiding principles and implementation frameworks. However, these documents do not outline a plan for regulating or prequalifying digital health interventions, unlike for HIVST kits [4].

Nearly 60% of the WHO member states have national regulatory bodies that regulate medical devices, and many of these countries have started to regulate and approve digital health interventions as a category called Software as a Medical Device (SaMD) [18]. A digital health intervention is considered a SaMD if it provides recommendations for the prevention, diagnosis, or treatment of a disease to health care workers, patients, or caregivers [19]. Most low- and middle-income countries (LMICs) do not have the capacity to regulate SaMD, as they are still trying to develop regulatory processes for traditional medical devices. Of the 15 countries represented in the College of Surgeons of East, Central, and Southern Africa, 7 have no regulatory processes, 7 are currently in development, and only 1, South Africa, has an established regulatory process for medical devices [18]. However, despite the established procedure, SaMD in South Africa does not yet require approval, although digital health interventions are considered a Class C IVD medical device, and the South African Health Products Regulatory Authority has committed to regulating all medical devices in the future [20].

These local regulatory authorities are not yet in a position to regulate digital interventions for HIVSTs, so prequalification at an international level may be more suitable. This viewpoint article provides background on the WHO prequalification process and discusses reasons for and challenges related to the WHO prequalification of digital interventions for HIVST.

WHO Prequalification

The WHO Prequalification Program (WHOPQ) is an international regulatory body that oversees the quality and safety of various medications and medical devices. While many high-income countries have their own regulatory bodies, such as the FDA in the United States or the Medicines and Healthcare products Regulatory Agency in the United Kingdom [19], LMICs do not have the capacity to ensure the quality of these goods, so the WHOPQ was established to serve these countries by identifying products that the WHO deems fit [21]. There are currently over 1700 products on the WHOPQ list, which are categorized into 4 streams: vaccines, medicines, in vitro diagnostics (IVDs), and vector control [22].

The WHOPQ was launched in 1987 to standardize and scale the WHO's Expanded Programme on Immunization, which had been used since 1974 to determine whether the vaccines for tuberculosis (TB) (Bacillus Calmette–Guérin [BCG]) and Yellow Fever purchased by United Nations Children's Fund (UNICEF) met suitable quality, safety, and efficacy standards [21,23]. In 2001, the medicines stream was introduced in response to the need for affordable, quality generic medications for the treatment of HIV, which allowed the WHO to begin prequalifying generic antiretroviral drugs (ARVs). Most of the generic ARVs were produced in India, and many procurement agencies did not have confidence in the oversight provided by the Indian drug regulatory authorities, as there were no originator equivalents from highly regulated countries to compare them with [24].

Over the next decade, the prequalification program expanded to include generic medications for TB and malaria, followed by generics from the essential medicines list for reproductive health, pandemic influenza variants, and diarrhea. In 2013, the medicines stream extended from finished pharmaceutical products to include the prequalification of the active pharmaceutical ingredients used to make medications [25]. The IVD stream was introduced in 2011 for the prequalification of HIV tests and expanded in 2013 to include male circumcision devices. In 2014, the IVD stream was redesigned to include tests for malaria, cholera, viral hepatitis, and syphilis. Subsequently, in 2016, HIVSTs were added to the IVD stream. The vector control stream was also added in 2016 [25].

For a product to become WHO prequalified, it must go through various assessments. The process varies somewhat across streams, but always starts with an invitation for expressions of interest for a specific product category released by the WHO. Manufacturers are invited to submit a cover letter, product dossier, and product samples (including final packaging and instructions for use) to undergo evaluation against internationally accepted regulatory standards for safety, quality, and efficacy [21]. The evaluation also includes a presubmission screening for eligibility and inspections of the manufacturing and clinical trial sites, where applicable. If successful, the entire WHOPQ process can take up to one and a half years [26], costing the manufacturer up to US$31,000 in fees, depending on the product stream [27].

The gap in regulatory processes for SaMD between high- and low-income countries [18] parallels the gap in regulation surrounding vaccines and medications that initially led to the development of the WHOPQ [21]. Since 2017, there have been no new streams added to WHOPQ. However, it has been suggested that the WHOPQ process could be expanded to include digital health technologies in the future [28]. This suggestion corresponds with the proposed actions of the WHO *Global Strategy on Digital Health 2020 - 2025*, which includes the development of a WHO framework for assessing and regulating digital health technologies [14].

XSL•FO
RenderX

Reasons for Prequalification of Digital Interventions for HIVST

WHOPQ and HIV

The SaMD space is large and diverse, with an estimated 350,000 digital health apps ranging from simple educational chatbots to advanced AI algorithms that use machine learning for early cancer diagnoses [29]. There are approximately 50 digital interventions for HIVST for which a peer-reviewed publication is available [13], and they do not provide a diagnosis but are intended as a screening tool to educate users and link them to care for confirmatory testing and appropriate treatment [10]. Due to their small number and built-in screening-confirmation control, digital interventions for HIVST could provide a low-risk test case for the WHOPQ to develop preliminary prequalification guidelines for SaMD.

The WHOPQ also has a strong history in the domain of HIV. Although the program started with the prequalification of vaccines, expansion into new product streams has been guided by the need to diagnose and treat HIV [23]. In 2002, the first medicine prequalified by the program was Ritonavir, an ARV, and the 10 other medications prequalified in the first year of the WHOPQ were all for treating HIV. A decade later, when the prequalification of IVDs was added to the program, the first prequalified IVD was for malaria, then the following 6 were all for the diagnosis of HIV. Currently, almost half the prequalified medicines (46%; 297/646) are for the treatment of HIV and more than half of the IVDs (57%; 62/109) are for HIV diagnosis [4].

Since 2006, the primary WHOPQ funder has been UNITAID, whose mission is to increase access to treatment for HIV, TB, and malaria, especially in LMICs. In 2017, UNITAID accounted for over half the WHOPQ budget, with supplementary funding from other HIV-related donors, including the Bill and Melinda Gates Foundation and The Global Fund to Fight HIV, TB, and Malaria [4]. With a proven track record of introducing new product streams by prequalifying solutions for HIV treatment and diagnosis, and the majority of WHOPQ funding coming from HIV-related donors, the prequalification of digital interventions for HIVST could be taken up by WHOPQ as a way of exploring the introduction of a new SaMD stream.

Improved Data Quality and Interoperability

The International Medical Device Regulators Forum (IMDRF) is a voluntary group of national medical device regulators, chaired by the FDA. IMDRF's objective is to define and standardize medical device registration, and in 2013 they formed the SaMD Working Group [18]. This working group initially defined SaMD using standardized vocabulary and then published a framework for risk categorization, a quality management system, and the clinical evaluation of SaMD [18].

These working group documents are supplemented by guidance from the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), independently producing guidance documents for medical devices and software. ISO 13,485 and ISO 14,971 define the requirements for quality and risk management for SaMD, while ISO 82,304 is directly concerned with health apps. Furthermore, IEC 62,304 and IEC 62,366 address the software lifecycle and usability engineering processes [30,31]. SaMD manufacturers must also consider the Fast Healthcare Interoperability Resources (FHIR) regulations, which outline the ideal rules for collecting and exchanging health care data electronically, and the European Union's General Data Protection Regulation (GDPR), which ensures data security and privacy [32,33].

Despite this abundance of guidelines and frameworks, not all regulatory bodies adopt and follow them with the same rigor [34]. A study of 9 countries found that most countries had centralized regulatory authorities that use a framework to evaluate health apps [35]. However, these countries had varying oversight of the usability, accountability, and interoperability of the apps [35].

Interoperability is especially important for digital interventions for HIVST, as there is a need to record results for linkage to care and epidemiological tracing. Furthermore, there is a need to standardize how to capture information from emerging populations of HIV testers, including individuals on PrEP and people living with HIV who continue to self-test. However, these indicators are not included or defined in the WHO document recommending HIV self-testing [36]. By prequalifying digital interventions for HIVST, the WHOPQ could consolidate data collection and de-silo interoperability issues.

Adoption at Scale

Although digital health interventions, including those for HIVST, have shown the ability to improve health outcomes and patient experiences, they are not yet incorporated into traditional budgeting practices that often only consider medications, medical devices, and services [37]. The lack of funding and reimbursement mechanisms makes it difficult for purchasing bodies to select and justify the prioritization of these digital health interventions at scale. By taking an evidence-based approach, some regulatory bodies such as the UK's National Institute for Health and Care Excellence (NICE) have introduced assessment frameworks specifically for digital health technologies. If an intervention can prove that it is safe and secure, then it may be formally endorsed by NICE. Although an endorsement does not guarantee national adoption, it is valued as a reliable independent assessment by decision-making bodies, when determining which digital health interventions to prioritize for funding [37].

Regarding digital interventions for HIVST, anything that will accelerate their widespread adoption will create a positive impact. With faster access, these promising interventions would guide more people through HIV testing and link more people living with HIV to life-saving ARV treatment. Early access to ARV treatment increases the life expectancy and quality of life of people living with HIV. In addition, there are epidemiological benefits at the population level, as ARV treatment decreases HIV viral load, and when the viral load is undetectable, HIV becomes untransmittable [38]. Furthermore, newly diagnosed people living with HIV change their sexual behavior after learning their status, leading to increased condom use and fewer sexual partners [39].

XSL•FO
RenderX

Many UN agencies, nongovernment organizations (NGOs), and donors, such as the Global Fund, now require WHOPQ for bulk purchasing of medicines and IVDs. Although an endorsement by national regulatory bodies would be beneficial, prequalifying digital interventions for HIVST would likely have a much larger impact on their widespread adoption at scale [40]. The widespread adoption of HIVST has also paved the way for the self-testing of other sexually transmitted infections, such as hepatitis and syphilis [41]. These other sexually transmitted infection self-tests follow the same procedures as HIVST, so digital interventions for HIVST could be easily adapted and applied to promote these tests as well.

Concerns Regarding Prequalification of Digital Interventions for HIVST

More Important Digital Health Priorities

The regulation of SaMD is on the WHO roadmap, and while digital interventions for HIVST could offer a low-risk pilot, there may be more impactful digital interventions to start with. In 2019, the WHO held its first Digital Health Technical Advisory Group meeting, which discussed how a prequalification process could be applied to data analytics, specifically the predictive algorithms used in AI-enabled SaMD [28]. It would be challenging to prequalify a learning algorithm that is constantly changing and adapting. However, these algorithms are already in use and, in some cases, have been shown to outperform their human counterparts [42]. AI-enabled SaMD is effective in spotting cancerous tumors through image-based AI algorithms, performing insulin dose calculations for diabetes through AI-enabled decision trees, and predicting disease patterns through algorithms developed during COVID-19 [42,43]. AI is also being piloted in digital interventions for HIVST, with computer vision technology to interpret test results, instead of untrained human users [44].

Despite these benefits, there are also ethical, privacy, and liability risks related to AI-enabled SaMD. An example of an ethical risk is racial bias. One study revealed that in a risk scoring algorithm, Black patients were significantly sicker than their White counterparts with the same score. If race were corrected for in the algorithm, the number of Black people eligible for additional care would increase from 17.7% to 46.5% [45]. In this example, incorrect assumptions in the training data led to bias. In addition, there are also privacy concerns regarding ownership and security of the data used to train the AI models. Furthermore, there are liability implications, especially with emerging algorithms used for diagnosis and treatment in clinical environments, where they may present a substantial risk to people's health and well-being [46]. Despite these considerable risks, many AI-enabled algorithms are already being trained and used in health care facilities without regulation. Prequalifying these SaMD by the WHOPQ may present a critical opportunity to positively impact patient outcomes.

Existing Barriers to WHOPQ

With almost 26 million people living with HIV, Africa accounts for over two-thirds of the global disease burden of HIV [47]. Despite such a high disease burden, less than 2% (5/297) of the prequalified medications for HIV were from African manufacturers; Lamivudine/Zidovudine (Universal Corporation Ltd., Kenya and Aspen Pharmacare Ltd.), Lamivudine/Zidovudine+Nevirapine (Aspen Pharmacare Ltd.) and Ceftriaxone, 500 mg and 1000 mg (Egyptian International Pharmaceutical Industries Company). Regarding HIV diagnostics, none of the 62 prequalified IVDs were from African manufacturers; however, one device, the Mylan HIV self-test (Atomo Diagnostics), has a secondary manufacturing site in South Africa [4].

The cost of research and development for medications and IVDs is a crucial barrier to entry that keeps 98% of these items from being manufactured in Africa [4]; however, the obstacles to developing digital interventions are much lower. There are over 7000 tech start-ups in Africa, and in 2021, these African start-ups raised over USD 2 billion [48,49]. Most of these start-ups develop fintech and e-commerce platforms, with less than 6% of the total funding going to digital health start-ups [50]. For these digital health start-ups, the cost of USD 31,000 for WHOPQ may be prohibitive, especially considering that the process could keep their product out of the market for at least one and a half years before the prequalification is completed [24,25].

In 2024, the Africa Centers for Disease Control and Prevention (CDC) launched the Africa HealthTech Marketplace as a way to empower local innovators by showcasing a variety of locally designed digital health solutions [51]. The marketplace is intended to improve access to local digital health interventions by categorizing products and resources from trusted sources, although it does not provide regulatory approvals. While this platform is an important step forward, there is no evidence yet to suggest that NGOs will base purchasing decisions on the marketplace, like they do with WHOPQ.

Conclusions

Access to the market of SaMD is regulated nationally by more than half of the WHO member states. However, despite the IMDRF, ISO, and EIC frameworks, there is no standardized international regulatory board for digital health interventions. The WHOPQ provides regulatory approvals for vaccines, medication, and IVD, which could also be used to prequalify SaMD. National regulatory bodies, such as NICE in the United Kingdom, can endorse digital health apps, while the CDC's new Africa HealthTech Marketplace can showcase local innovations; however, neither of these options is as influential as WHOPQ, which many funders now require for bulk purchases.

Digital interventions for HIVST could be considered for a low-risk pilot case, based on the WHOPQ's history with lifesaving HIV interventions and the need to standardize data and interoperability issues. If prequalification of digital interventions for HIVST were to become available, care should be taken to ensure that the process does not inequitably introduce barriers for digital health companies from LMICs, including in Africa, and that the prequalification does not unnecessarily delay access to people testing for HIV.

XSL•FO
RenderX

Conflicts of Interest

AEF, LS, and MA are shareholders in Aviro Health, a company that develops digital interventions for HIV self-testing. All other authors declare no conflicts of interest.

References

1. Guidelines on HIV Self-Testing and Partner Notification: Supplement to Consolidated Guidelines on HIV Testing Services: World Health Organization; 2016.
2. Roehr B. FDA approves “instant” HIV home test. BMJ 2012 Jul 5;345:e4636. [doi: 10.1136/bmj.e4636]
3. Johnson CC, Kennedy C, Fonner V, et al. Examining the effects of HIV self-testing compared to standard HIV testing services: a systematic review and meta-analysis. J Int AIDS Soc 2017 May 15;20(1):21594. [doi: 10.7448/IAS.20.1.21594] [Medline: 28530049]
4. WHO Prequalification. 2023. URL: https://extranet.who.int/pqweb/ [accessed 2023-01-07]
5. Number of countries adopting HIV self-testing policies rises sharply. Unitaid. 2017 Jul 25. URL: https://unitaid.org/news-blog/number-countries-adopting-hiv-self-testing-policies-rises-sharply/ [accessed 2025-08-11]
6. WHO list of prequalified in vitro diagnostic products. World Health Organization. 2022. URL: https://extranet.who.int/prequal/vitro-diagnostics/prequalified/in-vitro-diagnostics [accessed 2025-07-31]
7. New US$1 price for HIV self-tests. World Health Organization. 2022 Jul 27. URL: https://www.who.int/news/item/27-07-2022-new-1-dollar-price-for-hiv-self-tests [accessed 2022-11-08]
8. Stevens DR, Vrana CJ, Dlin RE, Korte JE. A global review of HIV self-testing: themes and implications. AIDS Behav 2018 Feb;22(2):497-512. [doi: 10.1007/s10461-017-1707-8] [Medline: 28155039]
9. Venter F, Majam M, Jankelowitz L, et al. South African HIV self-testing policy and guidance considerations. South Afr J HIV Med 2017;18(1):775. [doi: 10.4102/sajhivmed.v18i1.775] [Medline: 29568643]
10. Fischer AE, Abrahams M, Shankland L, Lalla-Edward ST, Edward VA, De Wit J. The evolution of HIV self-testing and the introduction of digital interventions to improve HIV self-testing. Front Reprod Health 2023;5:1121478. [doi: 10.3389/frph.2023.1121478] [Medline: 36864847]
11. Gous N, Fischer AE, Rhagnath N, Phatsoane M, Majam M, Lalla-Edward ST. Evaluation of a mobile application to support HIV self-testing in Johannesburg, South Africa. South Afr J HIV Med 2020;21(1):a1088. [doi: 10.4102/sajhivmed.v21i1.1088]
12. Fischer AE, Phatsoane M, Majam M, et al. Uptake of the Ithaka mobile application in Johannesburg, South Africa, for human immunodeficiency virus self-testing result reporting. South Afr J HIV Med 2021;22(1):a1197. [doi: 10.4102/sajhivmed.v22i1.1197]
13. McGuire M, de Waal A, Karellis A, et al. HIV self-testing with digital supports as the new paradigm: a systematic review of global evidence. EClinicalMedicine 2021 Sep;39:101059. [doi: 10.1016/j.eclinm.2021.101059] [Medline: 34430835]
14. (WHO) Global Strategy on Digital Health 2020-2025: World Health Organization; 2021.
15. Fischer AE, Van Tonder T, Gumede SB, Lalla-Edward ST. Changes in perceptions and use of mobile technology and health communication in South Africa during the COVID-19 lockdown: cross-sectional survey study. JMIR Form Res 2021 May 17;5(5):e25273. [doi: 10.2196/25273] [Medline: 33956640]
16. D’Anza B, Pronovost PJ. Digital Health: unlocking value in a post-pandemic world. Popul Health Manag 2022 Feb;25(1):11-22. [doi: 10.1089/pop.2021.0031] [Medline: 34042532]
17. Virtual Interventions in Response to HIV, Sexually Transmitted Infections and Viral Hepatitis: World Health Organization; 2022.
18. Hubner S, Maloney C, Phillips SD, et al. The evolving landscape of medical device regulation in East, Central, and Southern Africa. Glob Health Sci Pract 2021 Mar 31;9(1):136-148. [doi: 10.9745/GHSP-D-20-00578] [Medline: 33764886]
19. Software as a Medical Device (SaMD). International Medical Device Regulators Forum. 2023. URL: https://www.imdrf.org/working-groups/software-medical-device-samd [accessed 2023-03-10]
20. South African Health Products Regulatory Authority, Classification of Medical Devices and IVDs: South African Health Products Regulatory Authority; 2019.
21. Coyne PE. The World Health Organization Prequalification Programme-playing an essential role in assuring quality medical products. Int Health 2019 Mar 1;11(2):79-80. [doi: 10.1093/inthealth/ihy095] [Medline: 30551179]
22. Department of Regulation of Medicines and Other Health Technologies. Impact Assessment of WHO Prequalification and Systems Supporting Activities: World Health Organization; 2019.
23. Dellepiane N, Wood D. Twenty-five years of the WHO vaccines prequalification programme (1987-2012): lessons learned and future perspectives. Vaccine (Auckl) 2015 Jan 1;33(1):52-61. [doi: 10.1016/j.vaccine.2013.11.066] [Medline: 24300593]
24. ’t Hoen EFM, Hogerzeil HV, Quick JD, Sillo HB. A quiet revolution in global public health: The World Health Organization’s Prequalification of Medicines Programme. J Public Health Policy 2014 May;35(2):137-161. [doi: 10.1057/jphp.2013.53] [Medline: 24430804]
25. Blaschke TF, Lumpkin M, Hartman D. The World Health Organization prequalification program and clinical pharmacology in 2030. Clin Pharmacol Ther 2020 Jan;107(1):68-71. [doi: 10.1002/cpt.1680] [Medline: 31773722]

XSL•FO
RenderX

26. Dey R. Procedure and Outcome of Prequalification of Medicines Program by World Health Organization (WHO). Biomed J Sci Tech Res 2020;28(1) [FREE Full text] [doi: 10.26717/BJSTR.2020.28.004607]
27. Prequalification fees: WHO prequalification of in vitro diagnostics. World Health Organization. 2018. URL: https://iris.who.int/bitstream/handle/10665/259171/WHO-EMP-RHT-PQT-2017.04-eng.pdf?sequence=1&ua=1 [accessed 2025-07-31]
28. Shaffer L. WHO wants to bring order to health data. Nat Med 2020 Jan;26(1):2-3. [doi: 10.1038/s41591-019-0717-7] [Medline: 31932787]
29. Digital health trends 2021: innovation, evidence, regulation, and adoption. : IQVIA Institute for Human Data Science; 2021.
30. Odaibo S. Risk management of AI/ML Software as a Medical Device (SaMD): on ISO 14971 and related standards and guidances. arXiv. Preprint posted online on Sep 11, 2021. [doi: 10.48550/arXiv.2109.07905]
31. ISO/TS 82304-2:2021 health software — part 2: health and wellness apps — quality and reliability. International Organization for Standardization. 2021. URL: https://www.iso.org/standard/78182.html [accessed 2023-03-10]
32. Welcome to FHIR. FHIR. 2023. URL: https://build.fhir.org/index.html [accessed 2023-03-09]
33. Data Protection Act 2018. Legislation.gov.uk. 2018. URL: https://www.legislation.gov.uk/ukpga/2018/12/part/2 [accessed 2022-10-24]
34. Hermon R, Williams PP, Mccauley PV. Software as a Medical Device (SaMD): useful or useless term? Presented at: Hawaii International Conference on System Sciences; Jan 5-8, 2021; Kauai, Hawaii, USA. [doi: 10.24251/HICSS.2021.451]
35. Essén A, Stern AD, Haase CB, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):31. [doi: 10.1038/s41746-022-00573-1] [Medline: 35304561]
36. WHO recommends HIV self-testing as an approach to HIV testing services. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/WHO-CDS-HIV-19.36 [accessed 2025-08-11]
37. Recognising the Value of Digital Health Apps: An Assessment of Five European Healthcare Systems: MedTech Europe; 2021.
38. Bor J, Fischer C, Modi M, et al. Changing knowledge and attitudes towards HIV Treatment-as-Prevention and "Undetectable = Untransmittable": a systematic review. AIDS Behav 2021 Dec;25(12):4209-4224. [doi: 10.1007/s10461-021-03296-8] [Medline: 34036459]
39. Tiwari R, Wang J, Han H, et al. Sexual behaviour change following HIV testing services: a systematic review and meta-analysis. J Int AIDS Soc 2020 Nov;23(11):e25635. [doi: 10.1002/jia2.25635] [Medline: 33161636]
40. Huang Y, Pan K, Peng D, Stergachis A. A qualitative assessment of the challenges of WHO prequalification for anti-malarial drugs in China. Malar J 2018 Apr 3;17(1):149. [doi: 10.1186/s12936-018-2303-8] [Medline: 29615066]
41. Zhang Y, Tapa J, Johnson CC, et al. HIV, hepatitis, and syphilis self-testing among adolescents and young adults: a systematic review and meta-analysis. J Infect Public Health 2025 Jun;18(6):102764. [doi: 10.1016/j.jiph.2025.102764] [Medline: 40157332]
42. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98. [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
43. Nimri R, Battelino T, Laffel LM, et al. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. Nat Med 2020 Sep;26(9):1380-1384. [doi: 10.1038/s41591-020-1045-7] [Medline: 32908282]
44. Roche SD, Ekwunife OI, Mendonca R, et al. Measuring the performance of computer vision artificial intelligence to interpret images of HIV self-testing results. Front Public Health 2024;12:1334881. [doi: 10.3389/fpubh.2024.1334881] [Medline: 38384878]
45. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
46. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 1;27(3):491-497. [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
47. HIV/AIDS. World Health Organization African Region. 2018. URL: https://www.afro.who.int/health-topics/hivaids [accessed 2023-03-12]
48. Estimated number of startups in selected African countries in 2022. Statista. 2023. URL: https://www.statista.com/statistics/1290679/number-of-startups-in-africa-by-country/#:~:text=The%20number%20of%20startups%20in,in%20the%20same%20year%2C%20respectively [accessed 2023-03-12]
49. Dushime A. These four countries are leading Africa's start-up scene — here's why. World Economic Forum. 2022. URL: https://www.weforum.org/agenda/2022/08/africa-start-up-nigeria-egypt-kenya-south-africa/ [accessed 2023-03-12]
50. Jackson T. Fintech dominant for African startup funding but overall share fell in 2022. Disrupt Africa. URL: https://disrupt-africa.com/2023/02/15/fintech-dominant-for-african-startup-funding-but-overall-share-fell-in-2022/ [accessed 2023-03-12]
51. A new digital health platform for Africa. Africa CDC. 2024 Nov 10. URL: https://africacdc.org/news-item/a-new-digital-health-platform-for-africa/ [accessed 2023-04-23]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
ARV: antiretroviral drug
CDC: Centers for Disease Control and Prevention
FDA: Food and Drug Administration
FHIR: Fast Healthcare Interoperability Resources
GDPR: General Data Protection Regulation
HIVST: HIV self-testing
IEC: International Electrotechnical Commission
IMDRF: International Medical Device Regulators Forum
ISO: International Organization for Standardization
IVD: in vitro diagnostics
NGO: nongovernment organization
NICE: National Institute for Health and Care Excellence
SaMD: Software as a Medical Device
TB: tuberculosis
UNAIDS: the Joint United Nations Programme on HIV/AIDS
UNICEF: United Nations Children's Fund
WHO: World Health Organization
WHOPQ: World Health Organization Prequalification Program

Edited by L Buis; submitted 06.05.24; peer-reviewed by J Darrow, N Tieosapjaroen, O Ezeigwe, OS Igunma, S Rasania; revised version received 28.04.25; accepted 22.05.25; published 22.08.25.

Please cite as:
Fischer AE, Lalla-Edward ST, Edward V, Abrahams M, Shankland L, de Wit J
Should Digital Interventions for HIV Self-Testing Be Regulated with World Health Organization Prequalification?
JMIR Mhealth Uhealth 2025;13:e60276
URL: https://mhealth.jmir.org/2025/1/e60276
doi:10.2196/60276

© Alex Emilio Fischer, Samanta Tresha Lalla-Edward, Vinodh Edward, Musaed Abrahams, Luke Shankland, John de Wit. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evolving Digital Health Technologies: Aligning With and Enhancing the National Institute for Health and Care Excellence Evidence Standards Framework

Shayan Bahadori, BEng, MSc, PhD; Peter Buckle, PhD; Tayana Soukup Ascensao, PhD; Saira Ghafur, MD; Patrick Kierkegaard, PhD

Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, United Kingdom

Corresponding Author:
Shayan Bahadori, BEng, MSc, PhD
Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, United Kingdom

Abstract

The rapid advancement of artificial intelligence (AI)–driven diagnostics and wearable health technologies is transforming health care delivery by enabling real-time health monitoring and early disease detection. These innovations are catalyzing a shift toward personalized medicine, with interventions tailored to individual patient profiles with unprecedented precision. This paper examines the current National Institute for Health and Care Excellence (NICE) evidence standards framework (ESF) for digital health technologies (DHTs) and evaluates the challenges associated with integrating DHTs into existing health and care systems. A comprehensive review of the NICE ESF guidelines was conducted, alongside an evaluation of their applicability to emerging AI and wearable technologies. Key limitations and barriers were identified, with particular focus on the framework's responsiveness to technologies that evolve through machine learning and real-world data integration. Our findings indicate that while the NICE ESF provides a structured approach for evaluating DHTs, it lacks the adaptability required for rapidly evolving innovations. The framework does not sufficiently incorporate real-world evidence or support continuous learning models, which are critical for the safe and effective deployment of AI-based diagnostics and wearables. To remain effective and relevant, the NICE ESF should transition to a dynamic, adaptive model co-designed with industry stakeholders. By embedding real-world evidence-based strategies and promoting transparency, efficiency, and collaborative innovation, the updated framework would better facilitate the integration of AI-driven diagnostics and wearables into health care systems, ultimately enhancing patient outcomes and optimizing health care delivery.

(JMIR Mhealth Uhealth 2025;13:e67435) doi:10.2196/67435

KEYWORDS

mHealth; mobile app; app; smartphone; digital health; digital technology; digital intervention; wearable device; medical device; machine learning; artificial intelligence; algorithm; model; analytics; implementation; mobile health

Introduction

The swift evolution of artificial intelligence (AI)–driven diagnostics and wearable health technologies is pushing the boundaries of health care delivery, enabling early disease detection, real-time monitoring, and personalized interventions [1,2]. Although these innovations offer significant clinical and operational benefits, they also challenge existing regulatory mechanisms [3,4]. The National Institute for Health and Care Excellence (NICE) evidence standards framework (ESF) [5], designed to assess the value of digital health technologies (DHTs), struggles to accommodate tools that rely on continuous learning algorithms, real-world data, and data interoperability. A pertinent example is AliveCor's KardiaMobile, an AI-powered wearable electrocardiogram device [6]. This technology highlights several areas where the ESF may fall short, namely, its dynamic learning algorithms, which require continual updates; its dependence on real-world data, which the current framework is not fully equipped to evaluate; and its reliance on data interoperability, which is often hindered by fragmented systems [7]. For instance, despite clinical interest in its potential to support atrial fibrillation diagnosis, the technology faced challenges in aligning with traditional trial methodologies and producing economic evidence acceptable under current NICE criteria [7]. Additionally, concerns around algorithmic bias and the traditionally protracted evaluation timelines further complicate its integration [8]. These characteristics expose a critical misalignment. Although the ESF provides a robust foundation, it is grounded in static, predefined evaluation methodologies that may not sufficiently accommodate the iterative and adaptive nature of modern digital health tools, that is, technologies which are continuously refined and improved through ongoing cycles of data collection, analysis, and algorithmic updates. As such, a transition toward a more agile and evidence-responsive framework is needed to

XSL•FO
RenderX

ensure timely, safe, and effective adoption of innovations like KardiaMobile within health care systems.

This paper posits that the NICE ESF must undergo strategic updates to remain relevant and effective in the face of these rapidly advancing technologies. By critically examining the challenges and opportunities presented by AI-based diagnostics and wearable devices, this analysis will propose specific adaptations to the NICE ESF, ensuring that it can support the integration of these technologies into health care systems, ultimately enhancing patient outcomes and optimizing health care efficiency.

This paper presents a viewpoint grounded in existing literature, expert knowledge, and informal stakeholder engagement. No new data were collected for this work. The sources consulted included peer-reviewed articles, policy documents, and gray literature relevant to the topic. Stakeholder perspectives were drawn from prior collaborations, professional experiences, and publicly available materials, rather than formal interviews or surveys.

The NICE Framework

The NICE developed the ESF for DHTs to provide clear and consistent guidelines for evaluating the efficacy, safety, and economic value of new digital health innovations [5]. The ESF categorizes DHTs into different tiers based on their function and associated risk level, thereby ensuring that the evidence requirements are proportionate to the technology's impact. This tiered approach ranges from low-risk wellness apps to high-risk digital therapeutics and diagnostic aids [5].

The NICE ESF includes four main components:

1. Evidence for effectiveness, which includes clinical and nonclinical evidence to demonstrate that the technology works as intended and delivers the anticipated health benefits [5].
2. Evidence for economic impact, requiring technologies to show they provide value for money through cost-effectiveness analysis and budget impact assessments [5].
3. Regulatory compliance, data privacy, and security, ensuring that technologies meet regulatory standards and adhere to strict data privacy and security protocols, while also being interoperable with existing health systems [5].
4. Safety and performance standards, ensuring technologies meet safety requirements to protect patients and users, with ongoing monitoring to track performance and adverse events [5].

Although the NICE ESF provides a robust framework for evaluating DHTs, it faces significant challenges in addressing the complexities of rapidly evolving technologies such as AI-based diagnostics, wearable health monitors, and telemedicine platforms. For example, AI algorithms can continuously learn and improve from new data, which necessitates a different approach to evidence generation compared to static digital technologies [9]. Wearables generate real-time health data that can offer insights into patient behavior and health status outside clinical settings, presenting challenges in data integration and analysis [10]. The current NICE ESF, with its reliance on clinical trials and controlled studies, may struggle to adapt swiftly enough to these technological advancements, potentially delaying the evaluation and adoption of cutting-edge digital health innovations.

Furthermore, while this article focuses on the NICE ESF due to its relevance in the UK context, it is worth noting that other international approaches to evaluating DHTs exist. For example, the US Food and Drug Administration's Digital Health Software Precertification Program [11] emphasizes streamlined review processes for trusted developers, while the European Medicines Agency [12] has explored adaptive pathways to support innovation under conditions of uncertainty. Compared to these models, the NICE ESF offers a detailed stratification of evidence requirements according to risk and functionality, providing a transparent and structured pathway that aligns well with the needs of both developers and decision-makers within the UK health system. However, the existence of multiple international frameworks underscores the importance of regulatory adaptability and cross-jurisdictional learning in digital health innovation.

The Innovation Ecosystem

The advancement of DHTs is propelled by a complex and dynamic innovation ecosystem, characterized by deep collaboration across academia, industry, government, regulatory bodies, and patient and public involvement [13]. This ecosystem is fundamental to the development and dissemination of transformative technologies, such as AI-driven diagnostics and wearable health devices, which promise to significantly enhance health care delivery and patient outcomes [13].

Central to this ecosystem are start-ups and established companies, which, with substantial backing by venture capital investment and public funding, lead the innovation pipeline [13]. For example, 52North Health Limited [14] recently secured £3.6 million (US $4.8 million) in funding from venture capital firms and the National Institute for Health and Care Research to enhance its innovative cancer patient support platform. This investment enabled 52North Health to expand its personalized medication management solutions, which include tailored reminders, educational content, and real-time health analytics [15]. These investments not only validate the potential of digital health innovations but also highlight the critical importance of interdisciplinary research and development that incorporates the expertise of computer science, engineering, biology, and health care [16].

However, despite the rapid growth of the innovation ecosystem, the NICE ESF remains out of step with the fast pace of technological advancement. Innovators, particularly smaller start-ups, often encounter substantial barriers when attempting to satisfy the stringent evidence requirements mandated by NICE [17]. The traditional emphasis on extensive clinical trials can be prohibitive, especially for technologies that evolve rapidly or rely on continuous data integration and machine learning [17]. Furthermore, without well-established economic evidence, even highly promising technologies may face significant barriers to market entry and long-term adoption

XSL•FO
RenderX

[17,18]. Therefore, early engagement with economic modeling and real-world evidence generation, defined as data derived from routine clinical practice outside traditional clinical trials, and technological advancements that inform regular updates to the framework can help innovators prioritize their efforts and improve their chances of achieving reimbursement and scaling effectively [19].

Furthermore, both public and private sectors are increasingly prioritizing high-need, high-cost populations, where DHTs can have the most significant impact [16]. This prioritization amplifies the urgency for a regulatory framework that is both rigorous and flexible, capable of adapting to the unique demands of cutting-edge innovations. To prevent the stifling of innovation, the NICE ESF must evolve to streamline evidence generation processes and foster more collaborative relationships between regulators and innovators. Such adaptations will be essential to accelerating the adoption of promising technologies, ensuring that they reach patients and health care systems without unnecessary delay.

Challenges and Barriers to Integration

Despite the transformative potential of DHTs, integrating them into the health and care system presents a multifaceted challenge that requires strategic consideration and intervention.

Resistance to change is a major barrier within health care organizations. Clinicians accustomed to established protocols may be hesitant to adopt unfamiliar tools due to concerns over utility, job security, or disruption to workflow [20]. Addressing this requires not only rigorous evidence of clinical value but also engaging providers in co-design and implementation [21]. Structured workshops and pilot programs can help embed these technologies in real-world settings, fostering trust and usability [21].

Cost and infrastructure constraints also impede adoption. Implementing and maintaining advanced technologies can be financially burdensome, particularly in underfunded systems [20,22]. Limited technical infrastructure and expertise, especially in low-resource settings, further restrict integration [23].

Data security and privacy present additional barriers. The sensitive nature of health data demands robust governance frameworks that comply with regulation while maintaining usability. Overly restrictive protocols can hinder clinical workflows, so a balance between security and accessibility is essential [23]. The integration of DHTs into routine care raises significant concerns around the treatment of personal data and patient privacy [18]. These technologies often rely on continuous data collection from wearable sensors and mobile apps, which can include sensitive physiological, behavioral, and location-based information. Ensuring compliance with the UK General Data Protection Regulation and the Data Protection Act 2018 is essential, particularly regarding data minimization, lawful processing, and patient consent [24]. Inadequate safeguards may lead to breaches of trust and reduced uptake among users [24,25]. Privacy-by-design principles must be embedded into the development life cycle of digital tools, incorporating features such as data anonymization, secure encryption protocols, and user-controlled data sharing. In addition, emerging technical solutions like federated learning offer promising pathways to analyze health data without requiring centralized storage, thereby reducing privacy risks [26]. Regulators and developers must collaborate to establish transparent, accountable data governance frameworks that protect individual rights while enabling meaningful innovation.

Addressing the Problems

To address these barriers, a more dynamic approach to the NICE ESF is essential. Establishing a bidirectional feedback mechanism, where real-world evidence and technological advancements inform regular updates to the framework, is crucial [20,27]. This mechanism would allow for the continuous refinement of the ESF as new data and insights emerge from the use of DHTs in clinical practice. Practical approaches for integrating real-world evidence could include the use of prospective observational studies, pragmatic clinical trials, and the systematic analysis of electronic health records and patient-reported outcomes [27]. Additionally, incorporating methodologies such as federated learning [28] (a technique that enables AI models to be trained across multiple decentralized data sources without sharing raw data) and adaptive trial designs [29] (clinical studies that allow for modifications to trial procedures based on interim results) would enable real-time evaluation of AI algorithms without compromising data privacy. For example, integrating machine learning algorithms into the NICE framework could enable real-time adjustments to evidence standards based on live clinical data, ensuring the framework remains relevant and adaptive [20,30].

Moreover, a supportive policy environment, characterized by government incentives, grants, and favorable regulations, would encourage health organizations to invest in new technologies. Additionally, fostering a culture of innovation within health and care institutions can help overcome resistance [30]. Training programs and continuous professional development can equip staff with the skills and confidence needed to use new technologies effectively [31].

Engaging with health care providers, patients, policy makers, and technology developers through interviews, focus groups, and workshops can provide valuable insights into their needs and concerns [32]. This collaborative approach not only helps in designing user-friendly technologies but also ensures that the solutions are practical and widely accepted [32]. However, a significant challenge with such engagement arises from the conflicting priorities among stakeholders. For instance, technology developers and companies may prioritize the rapid adoption of innovations to gain a competitive market advantage, often pushing for faster regulatory approval or market entry. In contrast, health care providers and regulatory bodies typically emphasize safety, reliability, and long-term efficacy, demanding thorough testing and validation. This tension can lead to conflicts of interest, where the push for speed may compromise the rigor required for ensuring patient safety and data privacy [33].

Regular consultations and iterative feedback sessions with these stakeholders can ensure that the NICE ESF evolves in a way

XSL•FO
RenderX

that supports rather than stifles innovation, ultimately enabling the seamless integration of transformative technologies into health care systems.

Aligning Regulation With Innovation in Digital Health

The integration of DHTs into contemporary health care systems is contingent upon alignment between industry practices and the regulatory standards set forth by the NICE ESF [34]. This serves as a pivotal instrument in ensuring that these technologies are evaluated rigorously for safety, efficacy, and economic value [34]. However, the existing framework poses substantial challenges, particularly for small-to-medium enterprises and start-ups, which often struggle to meet the resource-intensive demands of evidence generation and submission. The disparities in resource availability between established industry giants and emerging innovators exacerbate these challenges, potentially stifling the introduction of groundbreaking technologies that could transform patient care.

Government intervention is critical in addressing these disparities and fostering an environment conducive to innovation. As highlighted in a recent report [35], it is imperative for the National Health Service to embrace digital innovations. Darzi [35] argued that greater investment in digital health care solutions could enhance efficiency, reduce the burden on staff, and improve patient outcomes. He advocated for the expanded use of data and technology to support diagnostics, treatment plans, and remote care. Effective governmental support should extend beyond policy frameworks to include targeted funding initiatives, regulatory flexibility, and infrastructure development aimed at supporting small-to-medium enterprises and start-ups. Hence, the active involvement of key stakeholders, health care providers, patients, technology developers, and policy makers is indispensable. Conducting in-depth interviews, facilitating focus group discussions, and engaging stakeholders through workshops and surveys provide a comprehensive understanding of the barriers and enablers. The government must ensure that these diverse perspectives are systematically incorporated into the regulatory process, thereby fostering a more inclusive and responsive health care innovation ecosystem.

Such strategic engagement with stakeholders should be underpinned by a rigorous methodology, incorporating qualitative research methods such as in-depth interviews, focus groups, and Delphi panels [36]. Analyzing the findings from these activities can elicit rich, nuanced insights into the barriers and enablers of DHT adoption, which can then be synthesized into evidence-based implementation strategies [37]. However, such engagement is not without challenges. Conflicting priorities, such as differing risk tolerances, budget constraints, and timelines between regulators, industry innovators, and clinical stakeholders, can impede consensus-building [36]. To navigate these tensions, structured facilitation techniques, such as modified Delphi processes, stakeholder mapping, and consensus workshops, can be used to ensure all voices are heard while aligning objectives [36]. A relevant example is the National Health Service AI Lab's Skunkworks program, which uses agile, multistakeholder co-design sprints to identify clinical problems, rapidly prototype solutions, and harmonize feedback from clinicians, patients, and developers [38]. This model demonstrates the value of structured, iterative engagement in achieving cross-sector alignment and fostering adoption readiness. The insights derived from stakeholder engagement should culminate in a detailed report outlining pragmatic strategies for the government to facilitate technology integration [39]. Such strategies should not only identify common challenges but also propose scalable solutions tailored to the unique needs of different health care environments. These insights must then be integrated into a comprehensive, living document that guides government action and supports the continuous evolution of the health care system.

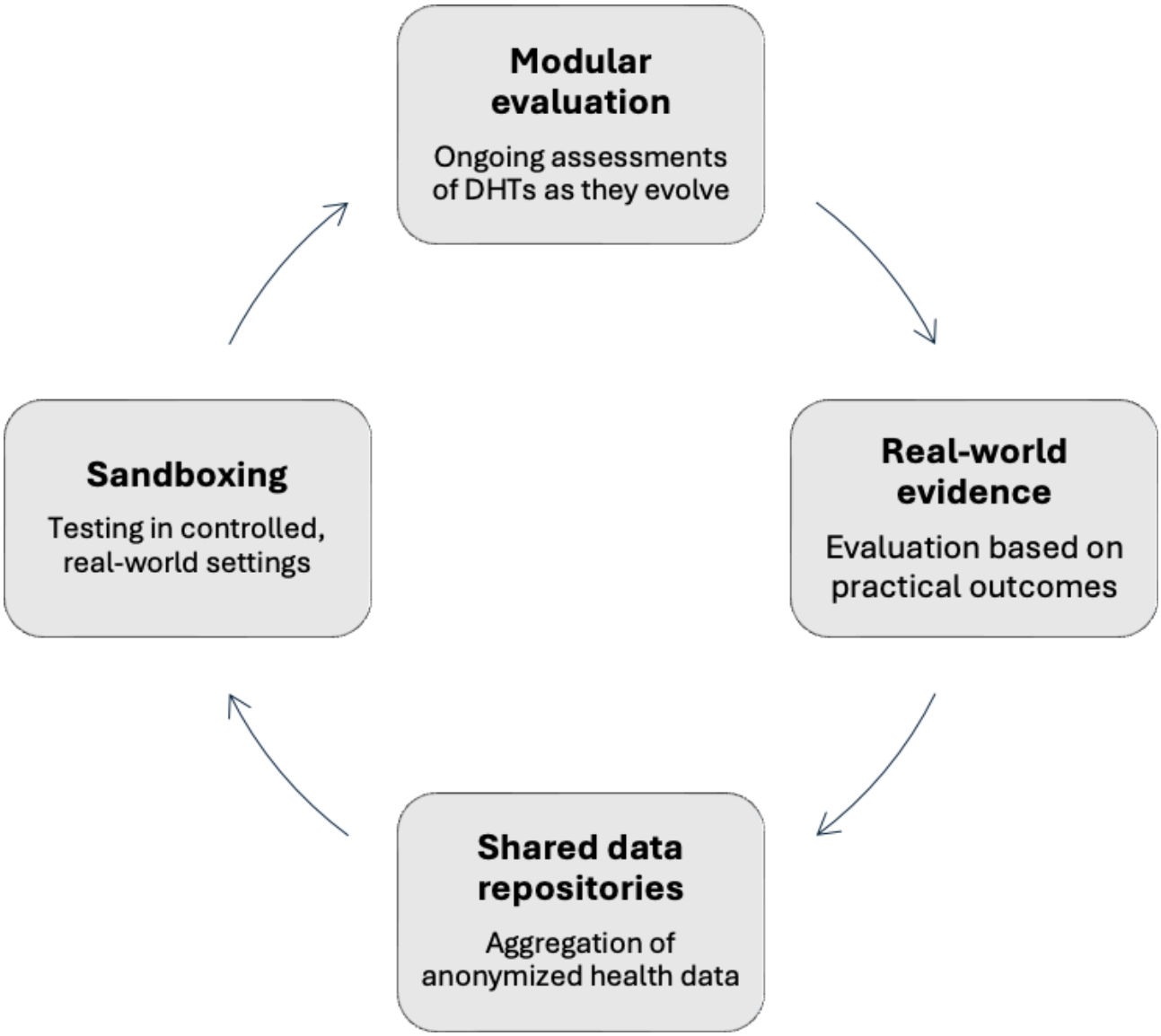
Recommendations for the Evolution of the NICE ESF

To address the challenges identified, we propose that the NICE ESF undergo a comprehensive transformation into a dynamic, iterative framework that evolves in real time with advancements in technology (Figure 1). One key innovation is the introduction of modular evaluation processes that would allow the NICE framework to assess different aspects of DHTs as they evolve. For instance, an early-stage prototype of an AI tool or wearable device could undergo an initial review based on preliminary data, while its efficacy and safety are continuously assessed and updated as real-world data accumulate. This iterative approach, supported by ongoing monitoring, would ensure that the NICE ESF remains adaptive to the pace of technological development.

Furthermore, to improve the speed of evaluation and reduce the burden on emerging innovators, we propose the establishment of a "sandboxing" environment within the regulatory framework. This would involve testing DHTs in controlled, real-world settings, where they can be iteratively refined based on real-time feedback from patients, clinicians, and developers. A sandbox model would facilitate rapid experimentation, helping to identify and address potential issues early on while also enabling faster, more flexible approval processes for DHTs. This method has been successfully applied in financial technology sectors (eg, the Financial Conduct Authority's Regulatory Sandbox), demonstrating its effectiveness in fostering innovation while maintaining consumer protection.

XSL•FO
RenderX

Figure 1. Proposed modification to the NICE ESF. DHT: digital health technology; NICE ESF: National Institute for Health and Care Excellence evidence standards framework.



Additionally, shared data repositories should be developed as collaborative platforms where anonymized health data from a range of sources (including wearable devices, health records, and patient surveys) can be aggregated. These repositories would provide developers with access to diverse datasets that can be used to train machine learning models while ensuring compliance with data protection regulations. By pooling data, stakeholders can create more robust AI-driven solutions that are reflective of real-world clinical environments and demographics. This also opens the door for collaborative data governance, where regulators, developers, and clinicians work together to define ethical and security standards for data sharing, ensuring transparency and trust across the health care ecosystem.

Finally, the development of real-world evidence strategies co-designed with industry partners would ensure that technologies are evaluated based on the actual outcomes they produce in practice, rather than just theoretical models or preclinical trials. This collaboration would result in evidence generation that is not only more relevant but also timelier, facilitating the rapid adoption of promising innovations. By integrating such strategies into the NICE ESF, we would enhance the framework's ability to provide continuous oversight while fostering a more flexible, innovation-friendly regulatory environment.

In conclusion, to facilitate the effective integration of DHTs, the NICE ESF must transition from a static framework to a dynamic, adaptive model. This evolution should include modular evaluation processes, sandboxing environments for real-world testing, shared data repositories, and co-designed real-world evidence strategies. By embedding flexibility and collaboration into the framework, stakeholders can ensure that the NICE ESF continues to support innovation while maintaining rigorous standards for safety and efficacy. These adjustments will not only streamline the adoption process but also enhance patient outcomes, ensuring that health care systems can keep pace with the rapid advancements in digital health technology.

XSL•FO
RenderX

Acknowledgments

Infrastructure support for this research was provided by the National Institute for Health and Care Research London HealthTech Research Centre in In-Vitro Diagnostics.

Conflicts of Interest

None declared.

References

1. Sharma A, Singh A, Gupta V, Arya S. Advancements and future prospects of wearable sensing technology for healthcare applications. Sens Diagn 2022 May 19;1(3):387-404. [doi: 10.1039/D2SD00005A]
2. Mukherjee S, Suleman S, Pilloton R, Narang J, Rani K. State of the art in smart portable, wearable, ingestible and implantable devices for health status monitoring and disease management. Sensors (Basel) 2022 Jun 1;22(11):4228. [doi: 10.3390/s22114228] [Medline: 35684847]
3. Aslam N, Zhou H, Urbach EK, et al. Quantum sensors for biomedical applications. Nat Rev Phys 2023;5(3):157-169. [doi: 10.1038/s42254-023-00558-3] [Medline: 36776813]
4. Ortiz BL, Gupta V, Kumar R, et al. Data preprocessing techniques for AI and machine learning readiness: scoping review of wearable sensor data in cancer care. JMIR mHealth uHealth 2024 Sep 27;12(1):e59587. [doi: 10.2196/59587] [Medline: 38626290]
5. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2024. URL: https://www.nice.org.uk/what-nice-does/digital-health/evidence-standards-framework-esf-for-digital-health-technologies [accessed 2025-08-07]
6. KardiaMobile. AliveCor. 2020. URL: https://store.alivecor.co.uk/products/kardiamobile [accessed 2025-08-07]
7. Savickas V, Stewart AJ, Short VJ, et al. Screening for atrial fibrillation in care homes using pulse palpation and the AliveCor Kardia Mobile device: a comparative cross-sectional pilot study. Int J Clin Pharm 2024 Apr;46(2):529-535. [doi: 10.1007/s11096-023-01672-z] [Medline: 38151689]
8. Chattopadhyay R, Vassiliou V, Chousou A, Tsampasian V, Pugh P. 108 Does alivecor kardiamobile represent a plausible method to detect atrial fibrillation in patients with embolic stroke of undetermined source - a pilot study. . 2023 p. A121. [doi: 10.1136/heartjnl-2023-BCS.108]
9. Feng J, Phillips RV, Malenica I, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. NPJ Digit Med 2022 May 31;5(1):66. [doi: 10.1038/s41746-022-00611-y] [Medline: 35641814]
10. Canali S, Schiaffonati V, Aliverti A. Challenges and recommendations for wearable devices in digital health: data quality, interoperability, health equity, fairness. PLOS Digit Health 2022 Oct;1(10):e0000104. [doi: 10.1371/journal.pdig.0000104] [Medline: 36812619]
11. Developing the software precertification program: summary of learnings and ongoing activities. US Food and Drug Administration. 2020. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-precert-program?utm_campaign=3108&dpid=3998&dqi=1&dqCampaignId=302 [accessed 2025-08-07]
12. European Medicines Agency. Final report on the adaptive pathways pilot. Hainan Lecheng Real World Research Institute. 2016. URL: https://hnrws.cn/uploads/file/20220317/17/6069ada2e13a927907b0918d59d9f311.pdf [accessed 2025-08-07]
13. Roland Sinker C, Rutter KA, Steve Bates O, et al. The innovation ecosystem programme–how the UK can lead the way globally in health gains and life sciences powered growth. NHS England. 2024 Nov 24. URL: https://www.england.nhs.uk/long-read/the-innovation-ecosystem-programme [accessed 2025-08-07]
14. 52North. 2024. URL: https://52north.health/? [accessed 2025-08-07]
15. Medtech. Meet the start-up: 52north health. 2024 URL: https://siliconcanals.com/uk-medtech-52north-bags-5-7m/
16. Abernethy A, Adams L, Barrett M, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022;2022. [doi: 10.31478/202206e] [Medline: 36177208]
17. Tomada L. Start-ups and the proposed EU AI Act: bridges or barriers in the path from invention to innovation? J Intell Prop Info Tech Elec Com L 2022;13:53 [FREE Full text]
18. Senna PP, Ferreira LMDF, Barros AC, Bonnín Roca JB, Magalhães V. Prioritizing barriers for the adoption of Industry 4.0 technologies. Comput Ind Eng 2022 Sep;171:108428. [doi: 10.1016/j.cie.2022.108428]
19. Ghafur S, Fontana G, Holmes H, Malcolm R, Varghese A. Generating economic evidence for digital health solutions. Prova Health. 2024. URL: https://www.provahealth.com/white-papers-list/health-economics-for-digital-health-solutions [accessed 2025-08-07]
20. Mumtaz H, Riaz MH, Wajid H, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health 2023;5:1203945. [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]

XSL•FO
RenderX

21. Peters S, Guccione L, Francis J, et al. Evaluation of research co-design in health: a systematic overview of reviews and development of a framework. Implement Sci 2024 Sep 11;19(1):63. [doi: 10.1186/s13012-024-01394-4] [Medline: 39261956]
22. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. Int J Environ Res Public Health 2023 Feb 15;20(4):3407. [doi: 10.3390/ijerph20043407] [Medline: 36834105]
23. Bente BE, Van Dongen A, Verdaasdonk R, van Gemert-Pijnen L. eHealth implementation in Europe: a scoping review on legal, ethical, financial, and technological aspects. Front Digit Health 2024;6:1332707. [doi: 10.3389/fdgth.2024.1332707] [Medline: 38524249]
24. Özçağdavul M. General Data Protection Regulation compliance and privacy protection in wearable health devices: challenges and solutions. Artuklu Health 2024(10):29-37. [doi: 10.58252/artukluhealth.1566573]
25. Sivakumar CLV, Mone V, Abdumukhtor R. Addressing privacy concerns with wearable health monitoring technology. WIREs Data Min Knowl 2024 May;14(3):e1535. [doi: 10.1002/widm.1535]
26. Li N, Xu M, Li Q, et al. A review of security issues and solutions for precision health in Internet-of-Medical-Things systems. Security and Safety 2023;2:2022010. [doi: 10.1051/sands/2022010]
27. Unsworth H, Dillon B, Collinson L, et al. The NICE Evidence Standards Framework for digital health and care technologies - developing and maintaining an innovative evidence framework with global impact. Digit Health 2021;7:20552076211018617. [doi: 10.1177/20552076211018617] [Medline: 34249371]
28. Pei J, Liu W, Li J, Wang L, Liu C. A review of federated learning methods in heterogeneous scenarios. IEEE Trans Consumer Electron 2024;70(3):5983-5999. [doi: 10.1109/TCE.2024.3385440]
29. Ben-Eltriki M, Rafiq A, Paul A, et al. Adaptive designs in clinical trials: a systematic review-part I. BMC Med Res Methodol 2024 Oct 4;24(1):229. [doi: 10.1186/s12874-024-02272-9] [Medline: 39367313]
30. Wubineh BZ, Deriba FG, Woldeyohannis MM. Exploring the opportunities and challenges of implementing artificial intelligence in healthcare: a systematic literature review. Urol Oncol 2024 Mar;42(3):48-56. [doi: 10.1016/j.urolonc.2023.11.019] [Medline: 38101991]
31. Salzmann-Erikson M. Integrating technology in aged care: challenges, opportunities, and a nursing lens. Contemp Nurse 2023 Dec;59(6):413-415. [doi: 10.1080/10376178.2023.2291119] [Medline: 38096232]
32. Soukup T, Zamora-Talaya B, Bahadori S, et al. Defining the value proposition in diagnostic technology: challenges and opportunities for its understanding and development - a review with a multiperspective reflective analysis. Front Med (Lausanne) 2025;12:1498618. [doi: 10.3389/fmed.2025.1498618] [Medline: 40051729]
33. Nejstgaard CH, Bero L, Hróbjartsson A, et al. Association between conflicts of interest and favourable recommendations in clinical guidelines, advisory committee reports, opinion pieces, and narrative reviews: systematic review. BMJ 2020 Dec 9;371:m4234. [doi: 10.1136/bmj.m4234] [Medline: 33298430]
34. Gomis-Pastor M, Berdún J, Borrás-Santos A, et al. Clinical validation of digital healthcare solutions: state of the art, challenges and opportunities. Healthcare (Basel) 2024 May 22;12(11):1057. [doi: 10.3390/healthcare12111057] [Medline: 38891132]
35. Darzi A. Independent investigation of the national health service in england (darzi review). GOV.UK. 2024. URL: https://www.gov.uk/government/publications/independent-investigation-of-the-nhs-in-england
36. Ahmed SK. The pillars of trustworthiness in qualitative research. J Med Surg Public Health 2024 Apr;2:100051. [doi: 10.1016/j.glmedi.2024.100051]
37. Waltz TJ, Powell BJ, Fernández ME, Abadie B, Damschroder LJ. Choosing implementation strategies to address contextual barriers: diversity in recommendations and future directions. Implement Sci 2019 Apr 29;14(1):42. [doi: 10.1186/s13012-019-0892-4] [Medline: 31036028]
38. AI knowledge repository. NHS England. 2023. URL: https://digital.nhs.uk/services/ai-knowledge-repository [accessed 2025-08-07]
39. Watson AK, Hernandez BF, Kolodny-Goetz J, et al. Using implementation mapping to build organizational readiness. Front Public Health 2022;10:904652. [doi: 10.3389/fpubh.2022.904652] [Medline: 35646781]

Abbreviations

AI: artificial intelligence
DHT: digital health technology
ESF: evidence standards framework
NICE: National Institute for Health and Care Excellence

XSL•FO
RenderX

Edited by L Buis; submitted 11.10.24; peer-reviewed by AK Jameil, S Ciccarelli; revised version received 24.06.25; accepted 21.07.25; published 22.08.25.

Please cite as:
Bahadori S, Buckle P, Soukup Ascensao T, Ghafur S, Kierkegaard P
Evolving Digital Health Technologies: Aligning With and Enhancing the National Institute for Health and Care Excellence Evidence Standards Framework
JMIR Mhealth Uhealth 2025;13:e67435
URL: https://mhealth.jmir.org/2025/1/e67435
doi:10.2196/67435

© Shayan Bahadori, Peter Buckle, Tayana Soukup Ascensao, Saira Ghafur, Patrick Kierkegaard. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

What to Consider When Developing Multidomain Mobile Health Interventions for Lifestyle Management

Manuel Weber[1,2,3*], PhD; Renato Mattli[1*], PhD; Anja M Raab[1,3], PhD; Anja Frei[2,3], PhD; Karin Haas[3,4], PhD; Thimo Marcin[5], PhD; Albrecht Vorster[6,7], PhD; Kai-Uwe Schmitt[1,3], PhD

[1]Academic-Practice-Partnership Between School of Health Professions at Bern University of Applied Sciences and University Hospital of Bern, Bern University of Applied Sciences, Murtenstrasse 10, Bern, Switzerland

[2]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

[3]Swiss School of Public Health, Zurich, Switzerland

[4]Institute on Ageing, School of Health Professions, Bern University of Applied Sciences, Bern, Switzerland

[5]Rehabilitation and Sports Medicine, Berner Reha Zentrum, Insel Group, University Hospital of Bern, University of Bern, Bern, Switzerland

[6]Swiss Sleep House Bern, Department of Neurology, University Hospital of Bern, University of Bern, Bern, Switzerland

[7]Interdisciplinary Sleep-Wake-Epilepsy-Center, University Hospital of Bern, University of Bern, Bern, Switzerland

[*]these authors contributed equally

Corresponding Author:
Renato Mattli, PhD
Academic-Practice-Partnership Between School of Health Professions at Bern University of Applied Sciences and University Hospital of Bern, Bern University of Applied Sciences, Murtenstrasse 10, Bern, Switzerland

Abstract

Mobile health (mHealth) interventions can transform health care delivery and improve public health. At the same time, the evidence on lifestyle interventions continues to grow. They show promising results in preventing and treating noncommunicable diseases and enhancing health-related quality of life. These factors highlight the potential of multidomain mHealth interventions for lifestyle management. This viewpoint paper focuses on drawing valuable lessons from past experiences and providing guidance to developers of mHealth interventions for lifestyle management. We underscore the critical role of sharing practical insights to advance innovation in the field of mHealth interventions. We used an iterative consensus process to derive lessons learned, identify challenges, and reflect on possible actions. Our insights are based on our experience in developing 2 smartphone-based lifestyle interventions. Challenges and corresponding options in the following areas are presented: target population (preferences, personalization, and delivery), user involvement and testing, human support, and multidomain interventions (interdisciplinarity, flexibility, and core team). The development of multidomain mHealth interventions for lifestyle management requires a participatory and iterative approach involving relevant stakeholders (including end users) so that the right people get the right content at the right time. Additionally, it is crucial to consider established frameworks, guidelines, and regulations; allocate appropriate resources; and form a core team committed to the project's aims and open to working in an interdisciplinary team.

(JMIR Mhealth Uhealth 2025;13:e63573) doi:10.2196/63573

KEYWORDS

app; health-related quality of life; lifestyle medicine; mobile phone; nutrition; physical activity; prevention; public health; self-management; sleep; stress management.

Introduction

mHealth

Globally rising health care costs, the shortage of health care professionals in many countries, and technological progress give mobile health (mHealth) interventions great potential to contribute to public health. In 2024, 4 out of 5 individuals worldwide owned a mobile phone, and about 68% were using the internet, representing an increase of 15 percentage points since 2019 [1]. The World Health Organization defines "mHealth or mobile health as medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [2]. mHealth is considered a component or subset of eHealth. mHealth research has evolved from its early inclination to support systems for health services (eg, clinical decision support systems and drug information sources) to more diversified purposes, including preventive care, health promotion, diagnosis, treatment, and monitoring [3].

mHealth interventions have the potential to improve health-related quality of life (HRQoL) at a favorable cost-effectiveness ratio [4,5]. mHealth technologies can enable continuous health monitoring at both individual and population levels, promote healthy behaviors to prevent or mitigate health

XSL•FO
RenderX

issues, support self-management of chronic diseases, and provide interventions accessible anytime and anywhere [6]. Despite these opportunities, a recent scoping review [7] identified several barriers associated with the use of mHealth apps. Key issues include problems with validity, usability concerns, technical problems, and data privacy and security concerns. Another systematic review further specified key patient concerns about mHealth apps, including unauthorized access, data breaches, third-party sharing, and lack of transparency [8]. It also identified facilitators, including user education and ensuring data confidentiality and privacy, to enhance trust and adoption. Specifically, for mHealth apps targeting lifestyle behaviors and mental health, maintaining user engagement over time remains a major challenge. Many users stop using these apps after a short period due to technical issues, privacy concerns, poor user experience, and evolving user needs [9]. By improving app design, enhancing usability, personalizing content, and implementing effective user engagement strategies, mHealth interventions may achieve greater long-term adoption and impact.

Evidence on the effectiveness and efficacy of mHealth interventions is growing. In an umbrella review of systematic reviews, positive results of mHealth interventions were reported for chronic disease management, improving chronic pulmonary disease symptoms and heart failure symptoms, reducing deaths and hospitalization, and improving quality of life [10]. In a further umbrella review with meta-analysis, eHealth and mHealth interventions were found to significantly improve physical activity, reduce sedentary behavior, enhance dietary habits, and promote better sleep quality across diverse age and health populations [11]. These findings highlight their potential as scalable and accessible strategies for public health.

Lifestyle and Health

Lifestyle behaviors emerge globally as strong determinants of morbidity and mortality [9-18]. A recent study showed that adherence to low-risk lifestyle factors was associated with an increased life expectancy at age 30 years [15]. Additionally, different patterns of healthy lifestyles appear to be associated with an increase in the number of years lived without major chronic disease [16]. Furthermore, adopting healthy lifestyles could mitigate the genetic risk of a shorter lifespan [17]. This remains crucial even in later life, as a healthy lifestyle not only reduces the risk of mortality but also contributes to extending life expectancy among older adults [18].

According to the American College of Lifestyle Medicine, lifestyle medicine is a rapidly growing discipline that follows a biopsychosocial approach and rests on 6 pillars: nutrition, physical activity, stress management, restorative sleep, social connection, and avoidance of risky substances [18]. Rooted in these principles, lifestyle interventions center on the individual and take a holistic view of their daily life [19]. Lifestyle interventions have demonstrated substantial potential to enhance HRQoL [20] and extend life expectancy [21]. In recent years, research has increasingly supported these interventions for the prevention and management of noncommunicable diseases (NCDs), such as cardiovascular diseases, respiratory diseases, diabetes, and cancers. Adhering to healthy lifestyle behaviors is an effective [22] and cost-effective strategy for preventing NCDs [23]. The prevention of NCDs could reduce global mortality rates and alleviate sources of morbidity, disability, health care expenditures, and productivity losses [24]. Additionally, lifestyle behaviors are essential factors in the self-management of various NCDs such as cancer [25], cardiovascular diseases [26,27], chronic pain [28], and mental disorders [29].

Development of mHealth Lifestyle Interventions

Several frameworks, such as IDEAS (Integrate, Design, Assess, and Share) [30] and the updated Medical Research Council framework [31], provide valuable guidance for developing multidomain mHealth interventions. IDEAS represents a step-by-step process to guide the development and evaluation of more effective digital interventions. The updated Medical Research Council framework is designed for developing and evaluating complex interventions. However, these frameworks often operate at a higher, overarching level, with individual steps remaining broad and lacking specificity at a micro level. Hence, it is essential for researchers and health professionals to transparently present their development processes and share insights gained [32]. Current reporting on mHealth interventions and their underlying development processes tends to be insufficient for understanding and replicating those interventions [33].

Aims

This paper aims to provide guidance to developers of mHealth interventions for lifestyle management. Our focus lies on multidomain lifestyle interventions and their development processes. Specifically, we derive valuable lessons learned, address challenges, and explore opportunities based on our experience in developing 2 multidomain mHealth interventions for lifestyle management. Our focus extends beyond the immediate development process, emphasizing the importance of reflecting on past experiences to inform future digital health initiatives. We recognize the dynamic nature of technology and advocate for adaptability and evolution based on practical experience. Furthermore, we underscore the critical role of sharing practical experiences within the field of mHealth interventions, contributing to its advancement and innovation.

Methods

To derive lessons learned, identify challenges, and reflect on possible actions, we used an iterative consensus process. Our derived lessons are complementary to the ones of a previous viewpoint paper from 2021 [34] that discusses 5 lessons learned from randomized controlled trials with mHealth interventions. Based on available literature, frameworks, guidelines, and our experience, we identified distinct overarching categories of options. We organized and prioritized them through multiple rounds of discussion, using both inductive and deductive coding strategies, along with mapping techniques. Two authors (MW and RM) categorized the topics, and any discrepancies were resolved through face-to-face discussions and the inclusion of 2 further authors (AMR and KUS) to reach a final consensus. The insights shared in this paper draw from our practical

XSL•FO
RenderX

experience in developing 2 smartphone-based interventions, which are detailed in the subsequent sections.

App-Based mHealth Project Descriptions

Throughout the development processes, we maintained ongoing discussions within the project team about the progress and findings of each step. Our advisory board consisted of experienced health professionals from various fields, including physiotherapy, exercise and sports science, nutrition and dietetics, psychology, nursing, pulmonology, oncology, medical informatics, gerontology, and software development.

DELIA

The DELIA (*Digital Lifestyle Intervention for Older Adults*) project aimed to develop and evaluate a digital lifestyle intervention for community-dwelling older adults aged 65 years and older [35]. The project was primarily guided by the IDEAS framework and followed an iterative participatory approach following these key steps:

1. Cross-sectional survey study (phase 1 of IDEAS): At first, a needs assessment for digital lifestyle interventions in Swiss community-dwelling older adults aged 60 years and older was conducted [36]. The goal was to investigate the needs, requirements, and preferences for digital interventions to promote physical activity and additional lifestyle-related content in the target population. A total of 922 respondents with a mean age of 72 years (SD 6.4 years; range 60 - 98 years) completed the web-based survey.
2. Semistructured interviews and brainstorming sessions (phases 2 - 4 of IDEAS): Interviews with experienced health professionals and researchers from different lifestyle medicine disciplines including sleep, nutrition and dietetics, physiotherapy, exercise and sports science, mindfulness, psychology, gerontology, and software development were carried out to specify target behaviors, identify behavioral strategies, and generate ideas. All experts had experience working with older adults. Their input helped tailor our lifestyle intervention specifically for older adults.
3. Prototype (phase 5 of IDEAS): Based on the previous steps, we developed an initial mock-up of the lifestyle app.
4. Mock-up testing (phase 6 of IDEAS): The mock-up was tested with 6 patients from the patient advisory board of the University Hospital of Bern. They provided valuable feedback on the app's interface, navigation, and usability.
5. Minimum viable product (phase 7 of IDEAS): Based on the previous steps, we built a fully functional minimum viable product with the most essential features.
6. Pilot testing (phase 8 of IDEAS): Once all app features were implemented, we tested its functionality with project team members during a 1-week testing phase. Subsequently, we performed a pilot test with 6 older adults. Their 2-week testing experience yielded valuable insights that we used to further improve the app. Afterward, we conducted a crossover interventional study (N=108) using a mixed methods sequential explanatory approach to investigate the use and implementation of the mHealth app.

QUALUCA

The QUALUCA (*Quality of Life in Lung Cancer Survivors*) project evaluates a digital lifestyle intervention developed for lung cancer survivors. Likewise, this project is primarily guided by the IDEAS framework, whereas the development followed an iterative participatory approach. Key steps included the following:

1. Interdisciplinary workshop (phases 1 - 4 of IDEAS): We organized a web-based workshop (4 hours) involving experienced health professionals who work with non–small cell lung cancer (NSCLC) survivors. The goal was to gather insights and experiences from professionals across different disciplines. Their inputs helped tailor our lifestyle intervention specifically for NSCLC survivors by specifying target behaviors, identifying behavioral strategies, and generating ideas.
2. Semistructured interviews and brainstorming sessions (phase 2 of IDEAS): Short interviews with 4 lung cancer survivors who were undergoing inpatient rehabilitation at that time were conducted. These interviews allowed us to better understand their needs and perspectives toward a digital aftercare program.
3. Prototype (phase 5 of IDEAS): Same as DELIA.
4. Mock-up testing (phase 6 of IDEAS): Same as DELIA.
5. Minimum viable product (phase 7 of IDEAS): Same as DELIA.
6. Pilot testing (phase 8 of IDEAS): We performed a pilot test with 3 lung cancer survivors contacted through a national lung cancer survivors' patient organization. As with DELIA, valuable insights gained from their 2-week testing experience were used to make further improvements to the app.
7. Evaluate efficacy (phase 9 of IDEAS): Currently, we are conducting a randomized controlled trial to primarily assess the efficacy of our digital lifestyle intervention on HRQoL in NSCLC survivors. We are also assessing explanatory outcomes within the intervention group, such as app usability, feasibility, appropriateness, acceptability of the intervention, participant experiences and satisfaction, and app usage data. Detailed information can be found in the published study protocol [37].

Results and Discussion

In the development of mHealth interventions for lifestyle management, various challenges may arise that complicate the development process and impede progress. Simultaneously, there are also potential courses of action and facilitating factors to address these challenges and ensure smooth processes. The following subsections describe challenges and corresponding options related to the distinct categories identified through the iterative consensus process.

Challenges and Options Related to Target Population

User preferences and personalization are essential in shaping mHealth interventions. These factors can influence technical implementation and content design. Therefore, user preferences and personalization should be carefully considered in relation to both aspects.

XSL•FO
RenderX

Preferences

An understanding of the target population's characteristics and preferences, including the intervention's format and layout, is fundamental in the development of mHealth interventions. However, the diverse preferences within target populations present significant challenges [38,39]. Variations in gender, socioeconomic status, cultural background, education, and digital as well as health literacy require prudent and well-considered approaches.

We encourage researchers to conduct initial investigations aimed at understanding the characteristics of their specific target population as proposed in the IDEAS framework. While national statistics and insights from other studies can provide a glimpse into overall trends, a nuanced comprehension of the distinctive attributes of the intended population is essential [40]. To foster engagement, developers must remain mindful of both users and their context. For instance, young people often gravitate toward language and designs that are tailored to their age group [41].

Practical Insights

In our projects (DELIA and QUALUCA), both apps allowed participants to adjust the font size, which seems obvious but is crucial for older users. Furthermore, a rather neutral design was chosen with colors high in contrast based on the experience of our interdisciplinary project team with similar target groups. In addition, participants had to specify the frequency of reminders they wanted to receive from the app. The framework MOLD-US [42] was used, which synthesizes barriers influencing mHealth usability for older adults, to inform our decisions in both projects.

We have observed that although the majority of the QUALUCA target population (lung cancer survivors) constitutes a subgroup of DELIA (older adults), they exhibit distinct preferences. For instance, app usage data revealed that DELIA participants often engaged in more activities than the program prescribed. In contrast, preliminary app usage data for QUALUCA participants indicate that the number of prescribed activities tended to be at the upper limit.

Our projects were conducted in Switzerland, where strict data protection regulations apply. Although individual preferences may exist in regard to data protection and privacy, we did not receive any feedback on this topic from participants in our projects. This may be partly because the projects were conducted within the framework of a research study in collaboration with health professionals, potentially fostering a higher level of inherent trust among participants.

Personalization

Personalization is a "process that changes the functionality, interface, content or distinctiveness of a system to increase its personal relevance" [43]. Research indicates that personalization plays a key role in enhancing user engagement and adherence to mHealth solutions [44]. Additionally, it can positively impact the effectiveness of mHealth interventions [44,45]. Personalization also influences the adoption and utilization of mHealth technology. Rather than adopting a generic "one-size-fits-all" approach, personalized strategies tailor health care services to address the unique needs of each subgroup and individual. In the context of mHealth, personalization can manifest in various ways, including customized messaging, personalized treatment plans, and individualized data analysis [46]. The main purposes of personalized mHealth solutions are disease self-management and promotion of lifestyle behaviors [43].

A metaethnographic review [41] investigating patients' perceptions of mHealth apps revealed a prominent emphasis on the necessity for greater personalization of app content, as highlighted by numerous participants across these studies. Patients felt that since they are the ones using these apps, they should be able to make them suit their needs better. Therefore, it is important that these apps have different features that let users customize them to fit what they need.

Tailored mHealth interventions appear to be more effective in physical activity promotion among adults than nontailored interventions [47]. However, future research is needed to address the following areas to fully realize the potential benefits of personalization in mHealth: developing design frameworks to increase users' motivation and engagement, creating guidelines for integrating personalization into mHealth solutions, establishing reporting standards, studying the effects of personalization techniques, and exploring how to adapt accessibility for different user groups [43].

Practical Insights

Every individual has their own lifestyle and daily routines with different factors having different levels of importance. In contrast, there are general lifestyle recommendations for different target groups. Based on our experience, personalization is a key aspect to ensure that every individual gets the right content at the right time. For instance, considering the wide variation in initial fitness levels, offering various difficulty levels is essential. Our apps assigned participants to different levels or tracks in the background based on the initial data participants provided upon installation. After the first week and again after the sixth week, participants were automatically asked by the app whether the training intensity was sufficient. Depending on their response, the level was maintained or adjusted by algorithms in the background. Such options for personalization can give users a sense of self-determination.

Delivery

When considering content delivery, various decisions must be made. For example, determining the format of the content: Should it be delivered through videos featuring subtitles and instructors, audio files with voice-overs in different languages and dialects, or simply text? Preferences vary among individuals; while some may prefer visual aids, others may find audio instructions more convenient. It is also important to choose a linguistic style that is appropriate for each subgroup of the target population [48]. Many users may not be familiar with technical terms, so using too much jargon can confuse them and make the intervention unnecessarily complex.

Besides, integrating gamification elements, such as interactive features, may be a promising approach to enhancing user engagement and motivation [49]. Research indicates that gamification may be particularly effective in healthy and

XSL•FO
RenderX

younger adults [50]. However, the level of gamification should be aligned with the preferences and cultural norms of the target population.

Practical Insights

Although confronted with financial and time constraints, there was some content delivered in different forms (text and audio) to give the participants the option to choose according to their individual preferences. However, only limited interactive elements could be incorporated into our projects. Overall, the decision was intentionally made to keep the app simple and minimalistic, considering our target populations. Introducing too many elements could overwhelm individuals, especially those with lower digital literacy. Furthermore, several experts from various fields related to the target population reviewed the content repeatedly and provided valuable inputs. This approach facilitated the accumulation of diverse perspectives and the harnessing of expertise from each discipline involved.

Challenges and Options Related to User Involvement and Testing

Participatory approaches, such as co-design or cocreation, foster user-centered design in mHealth interventions. These methods involve an iterative process in which targeted end users and other relevant stakeholders collaborate with researchers [51]. Ideally, shared decision-making should be maintained throughout the project life cycle. In this context, co-designed mHealth interventions may be more effective than traditional approaches, where interventions are primarily designed by researchers and clinicians [52].

Initiating testing as early as possible is crucial. Using methods, such as paper mock-ups, allows for preliminary assessment even before the product reaches its final stages. This early testing phase aids in identifying potential issues and gathering initial feedback, which can inform subsequent iterations. However, remaining receptive to user feedback throughout the entire development process is vital. Being open to feedback at any stage enables developers to address emerging concerns promptly and make necessary adjustments, thereby enhancing the overall user experience.

Additionally, careful consideration should be given to the selection of participants for testing. By including diverse participants with varying backgrounds, preferences, and levels of expertise, developers can gain comprehensive insights and ensure that the product meets the needs of its intended users.

Practical Insights

We began involving various stakeholders early on, such as health professionals and software developers, along with potential end users. This approach helped us better characterize and understand the perspective of the target population. Additionally, throughout the development process, we remained open to feedback at every stage. Even when certain adjustments were no longer feasible due to limited resources, they still contributed valuable insights for future projects. From a more technical point of view, we recommend thoroughly balancing between having an early prototype and specifying the technical implementation.

Challenges and Options Related to Human Support

In mHealth interventions, human support (digital person-to-person support or face-to-face) can play a role in overcoming various challenges. One key challenge is ensuring that users have access to adequate support when navigating complex health technologies. This includes addressing technical issues, clarifying instructions, and providing guidance on how to effectively use the intervention. Additionally, maintaining user engagement and motivation over time can be challenging, requiring personalized interactions and encouragement from human support personnel. Furthermore, cultural and linguistic diversity among users may necessitate tailored support approaches to ensure inclusivity and effectiveness. Despite these challenges, leveraging human support in mHealth interventions offers opportunities to enhance user satisfaction, improve intervention outcomes, and foster a sense of trust and connection between users and the intervention program [53].

The question of how much human support is truly necessary poses a significant consideration. While providing ample personal support is essential for ensuring users' adherence [44,54], excessive support can potentially hinder user autonomy and scalability of the intervention. Striking the right balance is crucial, tailoring support levels to the specific needs and capabilities of the target users. Understanding users' proficiency with technology, health literacy, and individual preferences is key in determining the appropriate level of support. Moreover, using scalable support models, such as tiered support systems or automated assistance (eg, chatbots), can help optimize resource allocation while still meeting users' support needs effectively. Ultimately, achieving the optimal balance between human support and user autonomy is pivotal in maximizing the impact and sustainability of mHealth interventions.

Practical Insights

In the DELIA study, there was no individualized support provided at the beginning of the intervention. Instead, an introductory email was sent to the participants including all the required information to install and use the mHealth intervention. Furthermore, a project-specific email address was provided, and participants were also able to contact the study team directly through the mHealth app. Although some participants required some support for installing the app, the chosen path for implementation was considered appropriate for the target population of this project.

In the QUALUCA study, we conduct a comprehensive onboarding session (90 minutes) at the onset of the intervention. During this web-based onboarding, the app is installed together with the participants. Additionally, the app's features are explained, and safety-related aspects are discussed. The onboarding aims to ensure that participants use the app effectively and derive maximum benefit from it. Furthermore, it is designed to minimize the need for support during the 12 weeks of intervention. Participants who have already completed the study highly appreciated the onboarding appointment, as it allowed them to address ambiguities and questions directly. Ambiguities and questions could be addressed directly. Additionally, participants have a dedicated contact person available to assist them with any challenges they may encounter.

XSL•FO
RenderX

This can build trust in the intervention and the study, as well as enhance overall motivation and adherence. Furthermore, addressing questions raised by participants during onboarding helps gain a better understanding of the target group and how they use the app.

Challenges and Options Related to Multidomain Interventions

Interdisciplinarity

Self-management and lifestyle interventions often use holistic and interdisciplinary approaches. Not only are the topics covered in such interventions interdisciplinary, there are also different theoretical frameworks, pedagogical concepts, and technical solutions that need to be considered [55]. This challenging task asks for a team effort combining different disciplines. To successfully connect people from different disciplines, there needs to be a common understanding of the underlying principles and aim of the intervention. Therefore, we recommend allocating some specific time to build such a common understanding and fundament for future collaboration at the beginning of each project. Although such effort costs time and money, it can pay off during the remainder of the project. Furthermore, we recommend involving various health professionals and other relevant stakeholders through focus group meetings or similar methods to gain insights from different perspectives and to understand the broad context [56].

Practical Insights

Highly experienced health professionals and scientists from various backgrounds were involved in the iterative design of our interventions, contributing to both the development of themes and the functionalities of the app. Disciplines covered included sleep, nutrition and dietetics, physiotherapy, exercise and sports science, mindfulness, psychology, and medical informatics. The concept of lifestyle medicine and lifestyle interventions was thoroughly introduced to the entire project team at the beginning of the intervention. We also highlighted the primary aim of the interventions, that is, increasing participants' HRQoL. We strived to embrace diverse perspectives and foster a shared understanding, particularly crucial in bridging the communication gap between software developers and health professionals.

Flexibility

The development of multidomain lifestyle interventions demands a high degree of flexibility from the involved professionals. Frequently, professionals must step out of their comfort zones to develop solutions that meet the needs of the target populations while contributing to the overarching project goal. When combining several lifestyle domains in parallel, the time requirement of an intervention increases. This must be offset by the needs and capabilities of the participants and requires professionals to take a step away from their traditional single-domain intervention. However, this calls for project members who see the value in the other lifestyle domains and are willing to change perspectives.

Practical Insights

The degree of flexibility of the professionals and project partners involved was carefully considered when setting up the project organization. For example, the overarching project goal was communicated at the outset and the commitment of each professional was a mandatory requirement for collaboration. This can be further illustrated using the example of mindfulness exercises. According to various traditional views, mindfulness exercises should last at least 45 minutes. However, recent research has shown the effectiveness of shorter exercises [57]. Such shorter mindfulness exercises were key for our interventions as the participants were exposed in parallel to content and exercises from other lifestyle domains. Therefore, it was important for us to involve project partners who were open to shorter mindfulness exercises.

Core Team

The development of mHealth interventions has 2 core aspects: content and technical implementation. Normally, professionals are experienced in one or the other aspect. This requires each core team member, independent from the background, to relinquish control at least partially. This is also related to another key necessity: mutual trust between core team members. We recommend planning enough time for the technical specification and the development of the content of the mHealth intervention. Additionally, it is important to allocate enough time to build trust in interdisciplinary collaboration, especially in new collaborations where team members often do not know each other. Staffing such projects with open-minded personnel is crucial for success. A high degree of openness is required, for example, when assumptions turn out to be incorrect. Furthermore, high adaptability is needed to be able to immediately respond to unexpected feedback or bugs that occur during the development process.

Practical Insights

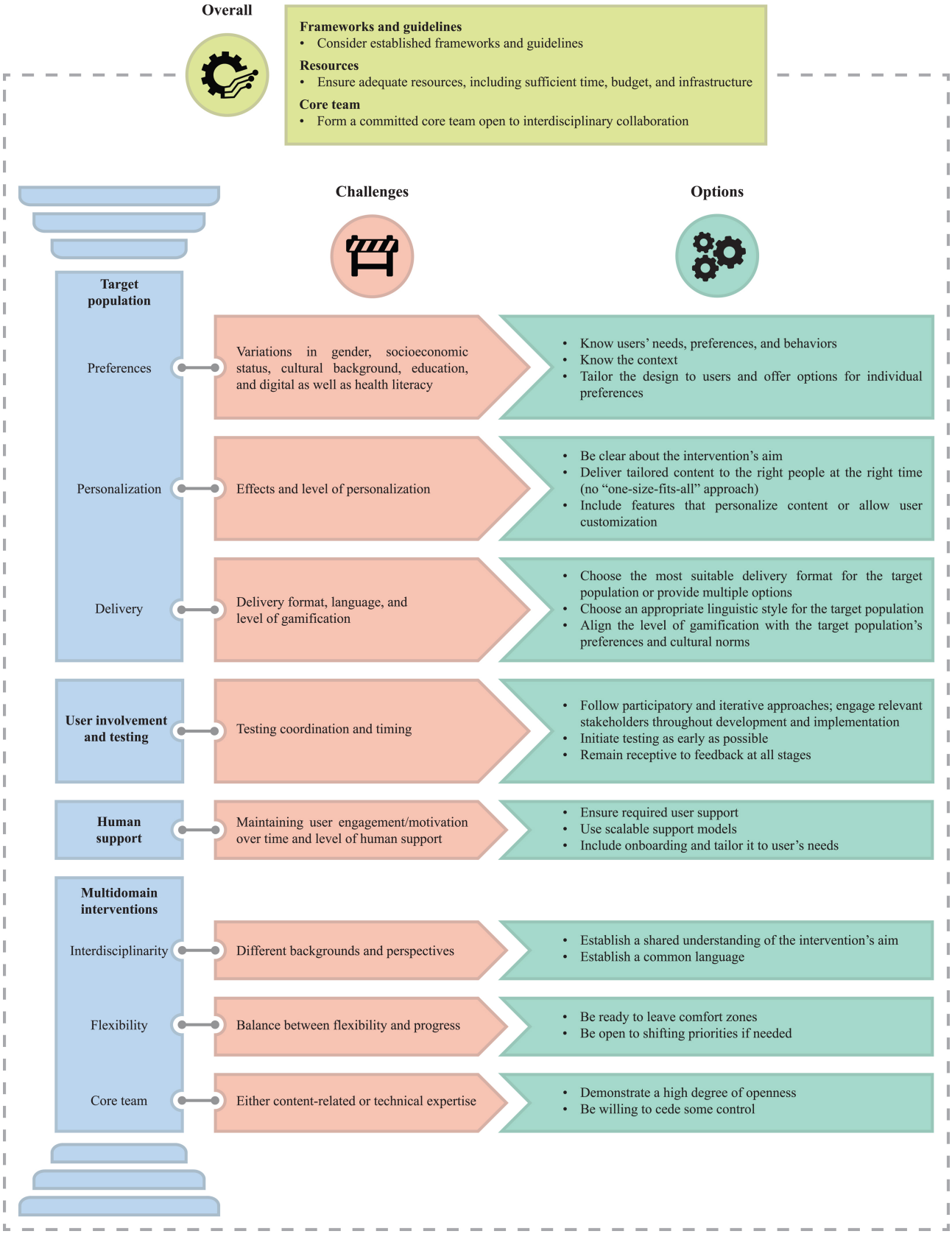
Although we prespecified the technical implementation in detail, we encountered unexpected new features and unforeseen bugs that required rapid adaptation. The challenges of executing research projects under tight time and financial constraints often leave certain aspects open-ended, requiring continuous problem-solving and iteration. To navigate these uncertainties effectively, we prioritized assembling a core team of open-minded, adaptable, and multidisciplinary professionals. Their diverse expertise allowed us to approach challenges from multiple perspectives, enabling faster troubleshooting and innovative solutions.

Summary

Figure 1 shows a summary of the key considerations outlined in this viewpoint paper.

XSL•FO
RenderX

Figure 1. Key considerations in the development of multidomain mHealth interventions.



Conclusions

In this viewpoint paper, we discuss the valuable insights gained from the development of multidomain mHealth interventions for lifestyle management. Our investigation encompasses various aspects, including considerations related to the target population, user involvement and testing, human support, and multidomain interventions. Understanding the diverse needs and preferences of the target population is crucial. Factors such as gender, socioeconomic status, knowledge and education, and digital

XSL•FO
RenderX

literacy, as well as health literacy, considerably impact user engagement and adherence. Successful multidomain lifestyle interventions deliver the right content at the right time to the right people and, therefore, include features that tailor the content toward the individual's needs or let users customize them to fit what they need. Furthermore, these interventions require collaboration across multiple disciplines and the involvement of key stakeholders, including end users. Project leaders should establish a common understanding of the intervention's aim and a common language between the project members early on. Stepping out of comfort zones allows for innovative solutions. Valuing other disciplines rather than centering solely on one's own expertise is essential. Partially relinquishing control, maintaining flexibility, and embracing openness to revising assumptions lead to better outcomes. In summary, the development of multidomain mHealth interventions for lifestyle management demands a holistic and interdisciplinary approach to address challenges and leverage options in the development process.

Acknowledgments

The authors sincerely thank their software developer, Peter Bruins from *Skyscraper Software GmbH*, for the technical implementation of their interventions.

Authors' Contributions

MW and RM wrote the first draft of the manuscript; AMR, AF, KH, TM, AV, and KUS provided critical review; and all authors read, contributed to, and approved the final version of this manuscript.

Conflicts of Interest

None declared.

References

1. Measuring digital development: facts and figures 2024. International Telecommunication Union (ITU)—Telecommunication Development Sector. 2024. URL: https://www.itu.int/hub/publication/D-IND-ICT_MDD-2024-4/ [accessed 2025-7-30]
2. mHealth: new horizons for health through mobile technologies: second global survey on ehealth. World Health Organization (WHO)—Global Observatory for eHealth. 2011. URL: https://iris.who.int/handle/10665/44607 [accessed 2024-04-01]
3. Ali EE, Chew L, Yap KL. Evolution and current status of mhealth research: a systematic review. BMJ Innov 2016 Jan;2(1):33-40. [doi: 10.1136/bmjinnov-2015-000096]
4. Rinaldi G, Hijazi A, Haghparast-Bidgoli H. Cost and cost-effectiveness of mHealth interventions for the prevention and control of type 2 diabetes mellitus: a systematic review. Diabetes Res Clin Pract 2020 Apr;162(108084):108084. [doi: 10.1016/j.diabres.2020.108084] [Medline: 32061819]
5. Carrandi A, Hu Y, Karger S, et al. Systematic review on the cost and cost-effectiveness of mHealth interventions supporting women during pregnancy. Women Birth 2023 Feb;36(1):3-10. [doi: 10.1016/j.wombi.2022.03.007] [Medline: 35339412]
6. Kumar S, Nilsen WJ, Abernethy A, et al. Mobile health technology evaluation. Am J Prev Med 2013 Aug;45(2):228-236. [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
7. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808. [doi: 10.2196/43808] [Medline: 37171838]
8. Alhammad N, Alajlani M, Abd-Alrazaq A, Epiphaniou G, Arvanitis T. Patients' perspectives on the data confidentiality, privacy, and security of mHealth apps: systematic review. J Med Internet Res 2024 May 31;26:e50715. [doi: 10.2196/50715] [Medline: 38820572]
9. Kidman PG, Curtis RG, Watson A, Maher CA. When and why adults abandon lifestyle behavior and mental health mobile apps: scoping review. J Med Internet Res 2024 Dec 18;26:e56897. [doi: 10.2196/56897] [Medline: 39693620]
10. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
11. Singh B, Ahmed M, Staiano AE, et al. A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. NPJ Digit Med 2024 Jul 5;7(1):179. [doi: 10.1038/s41746-024-01172-y] [Medline: 38969775]
12. Ford ES, Zhao G, Tsai J, Li C. Low-risk lifestyle behaviors and all-cause mortality: findings from the National Health and Nutrition Examination Survey III Mortality Study. Am J Public Health 2011 Oct;101(10):1922-1929. [doi: 10.2105/AJPH.2011.300167] [Medline: 21852630]
13. Kvaavik E, Batty GD, Ursin G, Huxley R, Gale CR. Influence of individual and combined health behaviors on total and cause-specific mortality in men and women: the United Kingdom health and lifestyle survey. Arch Intern Med 2010 Apr 26;170(8):711-718. [doi: 10.1001/archinternmed.2010.76] [Medline: 20421558]
14. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. Prev Med 2012 Sep;55(3):163-170. [doi: 10.1016/j.ypmed.2012.06.017] [Medline: 22735042]

XSL•FO
RenderX

15. Sun Q, Yu D, Fan J, et al. Healthy lifestyle and life expectancy at age 30 years in the Chinese population: an observational study. Lancet Public Health 2022 Dec;7(12):e994-e1004. [doi: 10.1016/S2468-2667(22)00110-4] [Medline: 35926549]
16. Nyberg ST, Singh-Manoux A, Pentti J, et al. Association of healthy lifestyle with years lived without major chronic diseases. JAMA Intern Med 2020 May 1;180(5):760-768. [doi: 10.1001/jamainternmed.2020.0618] [Medline: 32250383]
17. Bian Z, Wang L, Fan R, et al. Genetic predisposition, modifiable lifestyles, and their joint effects on human lifespan: evidence from multiple cohort studies. BMJ Evid Based Med 2024 Jul 23;29(4):255-263. [doi: 10.1136/bmjebm-2023-112583] [Medline: 38684374]
18. Frates B, Bonnet JP, Joseph R, Peterson JA. Lifestyle Medicine Handbook: An Introduction to the Power of Healthy Habits, 2nd edition: Healthy Learning; 2020.
19. Lippman D, Stump M, Veazey E, et al. Foundations of lifestyle medicine and its evolution. Mayo Clin Proc Innov Qual Outcomes 2024 Feb;8(1):97-111. [doi: 10.1016/j.mayocpiqo.2023.11.004] [Medline: 38304165]
20. Amiri S, Mahmood N, Junaidi S, Khan MA. Lifestyle interventions improving health-related quality of life: a systematic review and meta-analysis of randomized control trials. J Educ Health Promot 2024;13(193):193. [doi: 10.4103/jehp.jehp_1156_23] [Medline: 39268447]
21. Li Y, Pan A, Wang DD, et al. Impact of healthy lifestyle factors on life expectancies in the US population. Circulation 2018 Jul 24;138(4):345-355. [doi: 10.1161/CIRCULATIONAHA.117.032047] [Medline: 29712712]
22. Ford ES, Bergmann MM, Kröger J, Schienkiewitz A, Weikert C, Boeing H. Healthy living is the best revenge: findings from the European Prospective Investigation Into Cancer and Nutrition-Potsdam study. Arch Intern Med 2009 Aug 10;169(15):1355-1362. [doi: 10.1001/archinternmed.2009.237] [Medline: 19667296]
23. Tackling NCDs: “best buys” and other recommended interventions for the prevention and control of noncommunicable diseases. World Health Organization. 2017. URL: https://iris.who.int/handle/10665/259232 [accessed 2024-02-29]
24. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet 2014 Jul 5;384(9937):45-52. [doi: 10.1016/S0140-6736(14)60648-6] [Medline: 24996589]
25. Ligibel J. Lifestyle factors in cancer survivorship. J Clin Oncol 2012 Oct 20;30(30):3697-3704. [doi: 10.1200/JCO.2012.42.0638] [Medline: 23008316]
26. Doughty KN, Del Pilar NX, Audette A, Katz DL. Lifestyle medicine and the management of cardiovascular disease. Curr Cardiol Rep 2017 Oct 4;19(11):116. [doi: 10.1007/s11886-017-0925-z] [Medline: 28980137]
27. Katz DL, Frates EP, Bonnet JP, Gupta SK, Vartiainen E, Carmona RH. Lifestyle as medicine: the case for a true health initiative. Am J Health Promot 2018 Jul;32(6):1452-1458. [doi: 10.1177/0890117117705949] [Medline: 28523941]
28. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021 May 29;397(10289):2082-2097. [doi: 10.1016/S0140-6736(21)00393-7] [Medline: 34062143]
29. Firth J, Solmi M, Wootton RE, et al. A meta-review of “lifestyle psychiatry”: the role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. World Psychiatry 2020 Oct;19(3):360-380. [doi: 10.1002/wps.20773] [Medline: 32931092]
30. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317. [doi: 10.2196/jmir.5927] [Medline: 27986647]
31. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
32. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17 Suppl 1(11-21):11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
33. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-management education through mhealth: review of strategies and structures. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e10771. [doi: 10.2196/10771] [Medline: 30341042]
34. Pach D, Rogge AA, Wang J, Witt CM. Five lessons learned from randomized controlled trials on mobile health interventions: consensus procedure on practical recommendations for sustainable research. JMIR Mhealth Uhealth 2021 Feb 8;9(2):e20630. [doi: 10.2196/20630] [Medline: 33555263]
35. Mattli R, Weber M, Raab AM, Haas K, Vorster A, Schmitt KU. Digital multidomain lifestyle intervention for community-dwelling older adults: a mixed methods evaluation. Int J Public Health 2025;70:1608014. [doi: 10.3389/ijph.2025.1608014] [Medline: 40160302]
36. Weber M, Schmitt KU, Frei A, Puhan MA, Raab AM. Needs assessment in community-dwelling older adults toward digital interventions to promote physical activity: cross-sectional survey study. Digit Health 2023;9:20552076231203785. [doi: 10.1177/20552076231203785] [Medline: 37799500]
37. Weber M, Raab AM, Schmitt KU, et al. Efficacy of a digital lifestyle intervention on health-related QUAlity of life in non-small cell LUng CAncer survivors following inpatient rehabilitation: protocol of the QUALUCA Swiss multicentre randomised controlled trial. BMJ Open 2024 Mar 7;14(3):e081397. [doi: 10.1136/bmjopen-2023-081397] [Medline: 38453202]

XSL•FO
RenderX

38. Giebel GD, Abels C, Plescher F, et al. Problems and barriers related to the use of mhealth apps from the perspective of patients: focus group and interview study. J Med Internet Res 2024 Apr 23;26:e49982. [doi: 10.2196/49982] [Medline: 38652508]
39. Xie Z, Or CK. Consumers' preferences for purchasing mhealth apps: discrete choice experiment. JMIR Mhealth Uhealth 2023 Sep 13;11:e25908. [doi: 10.2196/25908] [Medline: 37707310]
40. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167. [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]
41. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mhealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817] [Medline: 31293246]
42. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114(66-75):66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
43. Rivera-Romero O, Gabarron E, Ropero J, Denecke K. Designing personalised mHealth solutions: an overview. J Biomed Inform 2023 Oct;146(104500):104500. [doi: 10.1016/j.jbi.2023.104500] [Medline: 37722446]
44. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
45. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 9;22(12):e21687. [doi: 10.2196/21687] [Medline: 33295292]
46. Gosetto L, Ehrler F, Falquet G. Personalization dimensions for mHealth to improve behavior change: a scoping review. Stud Health Technol Inform 2020 Nov 23;275(77-81):77-81. [doi: 10.3233/SHTI200698] [Medline: 33227744]
47. Davis A, Sweigart R, Ellis R. A systematic review of tailored mHealth interventions for physical activity promotion among adults. Transl Behav Med 2020 Oct 12;10(5):1221-1232. [doi: 10.1093/tbm/ibz190] [Medline: 33044542]
48. Amiri P, Nadri H, Bahaadinbeigy K. Facilitators and barriers of mHealth interventions during the COVID-19 pandemic: systematic review. BMC Health Serv Res 2023 Oct 28;23(1):1176. [doi: 10.1186/s12913-023-10171-w] [Medline: 37898755]
49. Schmidt-Kraepelin M, Toussaint PA, Thiebes S, Hamari J, Sunyaev A. Archetypes of gamification: analysis of mHealth apps. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e19280. [doi: 10.2196/19280] [Medline: 33074155]
50. Lee C, Lee K, Lee D. Mobile healthcare applications and gamification for sustained health maintenance. Sustainability 2017;9(5):772. [doi: 10.3390/su9050772]
51. Leask CF, Sandlund M, Skelton DA, et al. Framework, principles and recommendations for utilising participatory methodologies in the co-creation and evaluation of public health interventions. Res Involv Engagem 2019;5(1):2. [doi: 10.1186/s40900-018-0136-9] [Medline: 30652027]
52. Eyles H, Jull A, Dobson R, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016 Sep;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
53. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144. [doi: 10.2196/jmir.8480] [Medline: 29643048]
54. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30. [doi: 10.2196/jmir.1602] [Medline: 21393123]
55. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
56. Araújo-Soares V, Hankonen N, Presseau J, Rodrigues A, Sniehotta FF. Developing behavior change interventions for self-management in chronic illness: an integrative overview. Eur Psychol 2019;24(1):7-25. [doi: 10.1027/1016-9040/a000330] [Medline: 31496632]
57. Howarth A, Smith JG, Perkins-Porras L, Ussher M. Effects of brief mindfulness-based interventions on health-related outcomes: a systematic review. Mindfulness (N Y) 2019 Oct;10(10):1957-1968. [doi: 10.1007/s12671-019-01163-1]

Abbreviations

DELIA: Digital Lifestyle Intervention for older Adults
HRQoL: health-related quality of life
IDEAS: Integrate, Design, Assess, and Share
mHealth: mobile health
NCD: noncommunicable disease
NSCLC: non–small cell lung cancer
QUALUCA: Quality of Life in Lung Cancer Survivors

XSL•FO
RenderX

Edited by L Buis; submitted 25.06.24; peer-reviewed by MN Shalaby, N Holliday; revised version received 13.03.25; accepted 12.06.25; published 25.08.25.

Please cite as:
Weber M, Mattli R, Raab AM, Frei A, Haas K, Marcin T, Vorster A, Schmitt KU
What to Consider When Developing Multidomain Mobile Health Interventions for Lifestyle Management
JMIR Mhealth Uhealth 2025;13:e63573
URL: https://mhealth.jmir.org/2025/1/e63573
doi:10.2196/63573

© Manuel Weber, Renato Mattli, Anja M Raab, Anja Frei, Karin Haas, Thimo Marcin, Albrecht Vorster, Kai-Uwe Schmitt. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

The Challenge of Measuring Exercise: Advancing Metrological Barriers in Wearable Sensing

Jennifer L Corso[1], MA; Evan Peikon[1], MS

NNOXX Inc, Seattle, WA, United States

Corresponding Author:
Jennifer L Corso, MA
NNOXX Inc
600 1st Ave Ste 330 PMB 92856
Seattle, WA, 98104
United States
Phone: 1 425 371 5509
Email: jennifer.corso@nnoxx.com

Abstract

Regular physical activity offers extensive health benefits, yet current consumer wearables struggle to accurately quantify these effects at an individualized level. Sensor performance often falls short due to susceptibility to interferences, nonstandardized validation, and reliance on indirect estimations. Further, sensors often cannot capture or account for disparities in measurement types, populations, and physiological or anatomical characteristics, nor can they account for how different exercise modalities affect results on a personalized scale. There is a drive for developers to refine the impact of how we measure the benefits of exercise, improving the usefulness of data through advanced optical modeling and spectroscopic applications. This review critically examines the shortcomings of prevailing noninvasive measurements and techniques used in common, commercially available fitness trackers and describes why it is difficult to quantify the effects of exercise as an individualized, quality-based metric. Next, we discuss newer sensing applications that attempt to curtail known limitations, some of which may unveil novel biometric insights through differentiated approaches, bridging gaps not only in technological advancement but also in physiological metrology. In conclusion, we believe that new sensing techniques should explore solutions beyond population-based statistics and aim to provide an individualized understanding of a person's response to exercise, while also reducing disparities in personalized health monitoring. The results could lead to a more effective understanding of exercise efficacy and its impact on performance management and clinical outcomes.

(JMIR Mhealth Uhealth 2025;13:e79347) doi:10.2196/79347

KEYWORDS

photoplethysmography; PPG; muscle oxygen saturation; activity tracker; near-infrared spectroscopy; NIRS; wearable technology; remote patient monitoring; exercise

Introduction

Background

The phrase "sitting is the new smoking" has gained prominence as a compelling call to action, urging individuals to reconsider the implications of inactivity on their overall health and longevity. Among US adults, the obesity rate has exceeded 41% of the population according to the most recent reports provided by the Centers for Disease Control and Prevention, and heart disease, largely preventable via exercise, has been the leading cause of death since 1950 [1,2]. Despite the growing body of evidence highlighting the negative impacts of a sedentary lifestyle on lifelong health outcomes, many Americans still struggle to incorporate regular physical activity into their routines. Studies consistently show that engaging in regular exercise significantly boosts cardiovascular and musculoskeletal health, ultimately leading to reductions in all-cause mortality [3-5]. Further, large cohort studies indicate that participation in sporting activities may reduce all-cause mortality by nearly 40% [6], which further stresses the positive implications of fitness on overall longevity.

A 2023 meta-analysis published in the *British Journal of Sports Medicine* examined dose-response associations to physical activity in more than 30 million participants across 94 cohorts [7]. The study found that moderate physical activity significantly reduces chronic disease risk and suggested that even half of the recommended 150 minutes of weekly exercise could prevent 1 in 10 premature deaths and improve overall health outcomes. Exercise continually stands out as one of the most effective

XSL•FO
RenderX

interventions for improving the quality of life [8], and the addition of activity trackers has been shown to improve exercise-related outcomes in both clinical and healthy populations [9]. The use of consumer wearables is linked to improved body composition and overall fitness through increased amounts of physical activity when compared to nonusers.

The cardiovascular benefits conferred by engaging in physical activity are robustly documented [10]. Regular exercise is linked to a lower risk of developing atherosclerotic heart disease and hypertension, largely due to enhanced vascular reactivity and an increased expression of endothelial nitric oxide synthase (eNOS) [11]. Enhanced eNOS activity boosts overall nitric oxide synthesis, contributing to lower vascular resistance and reduced arterial blood pressure, consequently mitigating risks associated with all-cause mortality [12].

The hematological benefits of exercise are equally significant and dose dependent. Regular physical activity leads to increased plasma volume, erythrocyte mass, and erythropoietin synthesis, which are beneficial for endurance performance and lead to improved aerobic capacity [13]. The current gold standard for assessing aerobic capacity and overall cardiovascular fitness is through the measurement of an individual's maximum volume of oxygen consumed (VO_{2max}) using a graded exercise test and gas analysis. A higher VO_{2max} indicates a greater physiological capacity to uptake and use oxygen during aerobic exercise and correlates to a lower risk of cardiovascular disease.

It is abundantly clear that the acute and chronic health benefits of exercise are research supported and plenty. Physical activity significantly influences tissue perfusion, oxygen delivery, and consumption, as well as muscular strength, cognitive function, and overall longevity [14,15]. However, quantifying the benefits on a personalized scale poses substantial challenges outside of controlled laboratory settings.

Review Objectives

Currently, there is no standardized, user-friendly method to assess the *benefit* of exercise that is tailored to individual physiological dynamics. This creates a gap in the practical understanding of the effective impact of physical activity on an individual's health, as the biometrics used today are also limited. Therefore, we cannot easily answer the following question: What is the *efficacy* of exercise for an individual? At present, the general public is unable to quantify how well they acutely respond or adapt, thereby limiting any actionable modifications to maximize both acute and chronic exercise quality, objectively.

The aim of this viewpoint is to provide an overview of the current state and limitations of exercise metrology, followed by potential solutions to measuring exercise efficacy on an individualized scale. First, we review the limitations of current optical sensors, applications, and associated physiological biometrics commonly used in the noninvasive measurement of exercise performance. Next, we discuss various technological developments that attempt to unveil new information in relation to measuring exercise efficacy, with the ultimate goal of generating personalized exercise physiology data to improve biometric insights. We then conclude with how we believe the measurement science should move forward.

Measurement Challenges in Exercise and Health

Emerging technologies, particularly wearable devices and artificial intelligence/machine learning (AI/ML)–powered algorithms, are beginning to bridge these gaps by enhancing access to one's personalized health data. Integrated platforms deliver insights into disease risk and behavioral outcomes related to physical activity, albeit within certain limitations, such as data quality and interoperability [16,17]. For elite athletes, even marginal improvements in the quality of exercise data can have substantial effects on performance outcomes. However, technological enhancements can sometimes teeter on ethical lines: Do the data provided simply enhance training improvement or push natural adaptations outside the bounds of the current understanding of sport and medicine [18]? Can technologies do this accurately and lend more confidence to outcomes, or do the data raise more questions and create controversy? How are we contextualizing the information based on the known performance attributes of current sensors? For clinical patients, the current generation of noninvasive devices fails to meet the stringent reliability standards required to improve health outcomes, thereby making medical therapeutics a market relatively void of effective user-friendly wearables [19,20].

Limitations of Photoplethysmography in Biometric Assessment

Despite advancements in wearable technologies, significant limitations exist when using such devices for assessing the benefits of exercise for specific applications, such as for professional athletics and aging adults. Many devices use light-emitting diode (LED) photodiodes at specific wavelengths coupled with photoplethysmography (PPG) to estimate health metrics, such as pulse rate (eg, heart rate [HR]). Attempts have been made to expand the use of the PPG signal to monitor more complex variables, such as cardiac output (CO) and blood pressure [21]. Figure 1 illustrates the general principle of using PPG to assess biometric variables by analyzing a variety of waveform features [22].

PPG, which optically measures changes in blood volume over time, is susceptible to a variety of interferences. PPG primarily reflects changes in peripheral blood volume, which is influenced by factors such as temperature, sympathetic nervous system activity, and certain medications. These factors can alter peripheral vascular tone, affecting the PPG waveform and potentially leading to inaccuracies in biometric estimations. Further, the optical properties of living tissue vary by tissue type, blood and water content in the tissue, collagen, melanin, and any connective tissue fiber development [23]. It is widely known that biases in the data stem from interferences, such as melanin, and the US Food and Drug Administration (FDA) has proposed new guidance to reduce performance discrepancies specifically in pulse oximetry [24,25].

XSL•FO
RenderX

Figure 1. (a) Principle of using PPG and waveform features used to infer physiological parameters. (b) Systolic and diastolic phases. PPG: photoplethysmography; PPI: pulse-pulse interval; PPIdV/dt: maximum dV/dt of adjacent pulse-pulse interval; PPIonset: adjacent pulse-pulse onset interval; PPIsystolic: adjacent pulse-pulse interval; PWx: pulse width at x% of the systolic amplitude. Reproduced with permission [22].

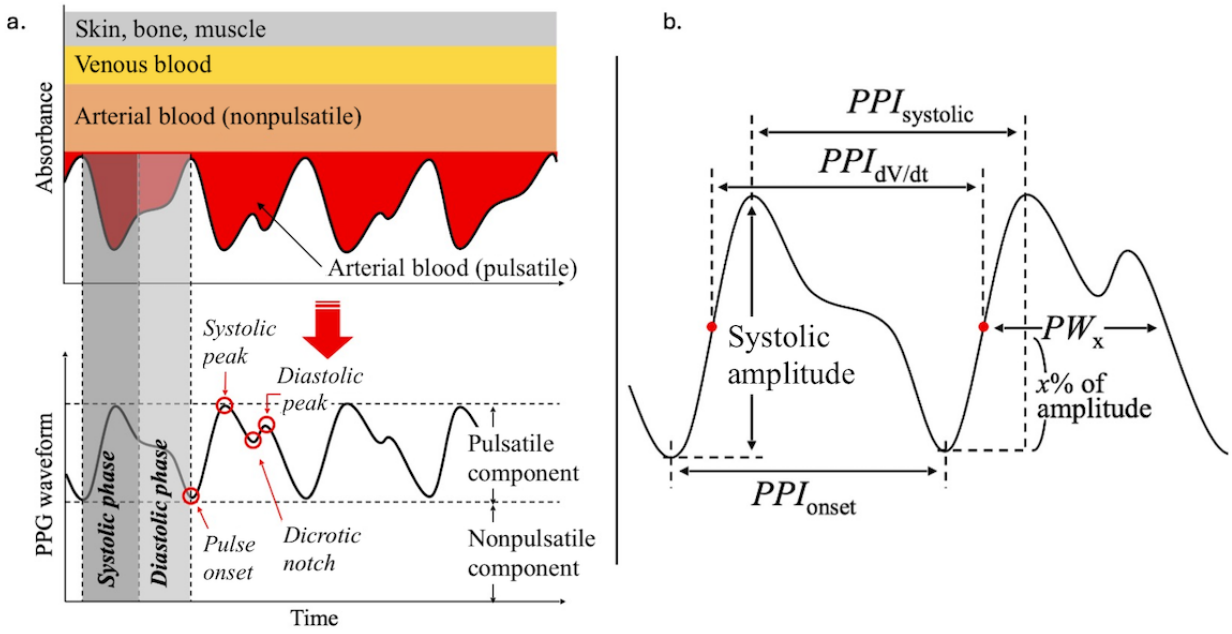
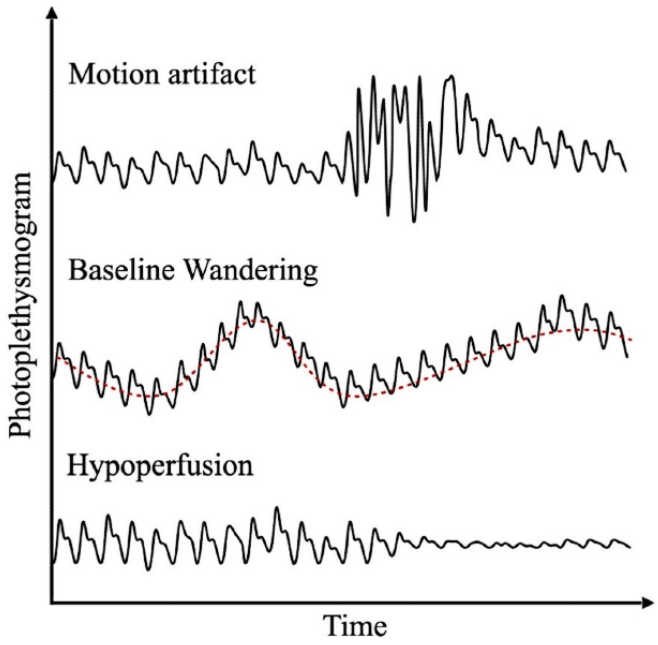


Figure 2 provides three examples illustrating how the PPG signal may distort over time [22]. Generally, this happens when interferences, such as those factors explained previously, influence the signal and degrade the ability to interpret the targeted signal features clearly. Motion artifacts and signal interference are paramount and can compromise the signal-to-noise ratio of the PPG measurement, thereby limiting physiological insights [26]. Signal corruption stemming from motion can be estimated and rectified using applications such as quadrature reference signals or other correlational filters [27]. Other interferences include baseline drift, baseline wandering, and stochastic noise. Further, body location, skin-to-sensor pressure at the measurement location and breathing characteristics also significantly impact waveform characteristics, including the mean amplitude, dicrotic notch time, and reflection index [28]. Because quality and reliability are crucial features of data applicability, variability among sensors makes it difficult to measure these attributes and to ascertain their impact on outcomes. A lack of stringent validation standards for PPG-based devices strongly limits their analytical utility and reliability.

Figure 2. PPG signal distortion due to motion artifact, baseline wandering, and hypoperfusion. PPG: photoplethysmography. Reproduced with permission [22].



XSL•FO
RenderX

These limitations may necessitate the use of auxiliary hardware for more accurate data collection. This is the case in arrhythmia detection, which has propelled itself into modern consumer wearables. Additionally, PPG cannot adequately capture certain vascular dynamics, such as reflective effects on the vessel expansion interrogated after transmission of the waveform, making estimates for parameters such as blood pressure challenging and leading to low application validation in clinical populations [29,30].

In general, device limitations impact the usability of a variety of commercially available and common health metrics. The data are only as useful as the metrological capabilities.

Step Counts and Health Outcomes

The recommendation of taking 10,000 steps per day originated from the Japanese company Yamasa, which launched a pedometer called *Manpo-Kei* around 1965, shortly after the Tokyo Olympics. The pedometer name translates to "10,000 steps meter." The campaign promoted this number with the catchy slogan "Let's walk 10,000 steps a day," which coincided with the rise of walking clubs in Japan at that time. The intention was to encourage people to be more active.

The current recommendation of 10,000 steps per day set forth by the American Heart Association (AHA) was built upon the original Japanese campaign. The aim of the AHA was to promote activity using an accessible benchmark, with a goal to reduce chronic disease risk. The step count quantity has been challenged by recent research. Though not an exhaustive list, several recent studies indicate that lower step counts may still be effective for improving cardiovascular and overall health, thereby prompting a re-evaluation of the necessary number of steps for optimal health outcomes and improvements in cardiovascular fitness [4,31-34]. A range of 8000-9000 steps per day may protect against diseases such as hypertension, diabetes, sleep apnea, major depressive disorder, and obesity [32]. Even lower counts, roughly 4000 steps, may lower mortality rates significantly [4,33].

Steps are considered objective, assuming the device measuring them has a relatively low margin of error and the exerciser is using the device as indicated by the manufacturer. If these attributes are not maintained, steps can be an inaccurate depiction of exercise quantification, as gait is often misinterpreted by electronic or electromechanical pedometers and inertial measurement units (IMUs)—accelerometers and gyroscopes used to measure the body's angular rate, force, and orientation. Over- or underestimating a true value is a significant limitation, with errors sometimes surpassing 10%, depending on body placement and validation criteria [35,36].

Often, if an individual is walking with a device located on the wrist and the arm is static, for example, while pushing a stroller or shopping cart, the pedometer does not register some if not all of the steps accrued. This is due to low or no cyclic motion registered by the device sensor. Further, metrological inconsistencies may conflict with Global Positioning System (GPS)–derived data. For example, if 2000 steps are roughly equivalent to 1 mile when using an average stride length, then an over- or underestimation may show a user traveled farther or less via GPS, which may be confusing.

The intensity of physical activity associated with a specific step count is often inadequately assessed by many wearable devices unless additional parameters, such as HR, are integrated. Standard step-counting metrics can fail to capture the differences in activity intensity; for instance, a person may accumulate the same number of steps through both leisurely walking and brisk walking, yet the cardiovascular and metabolic benefits vary significantly [35].

Furthermore, vigorous activities of daily living (ADLs), such as lifting heavy objects or engaging in high-intensity interval training, are typically not reflected in step counts. These forms of activity can yield significant health benefits, particularly for populations such as older adults or individuals with chronic conditions [32,37]. Other types of activity are necessary for people with disabilities that limit their ability to walk, such as strength training.

Heart Rate as a Proxy for Exercise Intensity

HR is frequently used as a proxy for gauging exercise intensity and is typically assessed via PPG in consumer wearables. However, this methodology is not without inherent limitations: PPG's reliance on pulse wave analysis is vulnerable to motion artifacts, variations in the quality of skin contact, and signal loss during irregular movement, all which can severely compromise measurement accuracy [38]. Improvements in validation testing criteria, denoising, and waveform peak identification are being enhanced with the incorporation of trained deep neural networks (DNNs) [39,40].

Delays in adaptive algorithms limit the accuracy of HR to nonambulatory conditions and, to a lesser extent, during steady-state exercise. Many algorithms use sliding windows (eg, 8- or 12-second processing windows) to estimate HR across a predesignated timespan. This approach introduces lag, preventing real-time detection and tracking of rapid HR changes, such as during interval training [41]. The error increases dramatically, and data are often unusable in these instances; in addition, the last reportable value is carried forward until the error decreases below a predetermined threshold of acceptable accuracy. In general, HR estimates are only reliable during rest or in steady-state conditions. However, inaccuracies may still be present during sleep, when HR may drop below 50 beats per minute in trained adults or those with underlying conditions.

Lastly, HR is not a direct measure of the benefits of exercise in all cases, such as during strength training. The estimate is not equally validated across all forms of exercise or in contraindicated groups [42]. Additionally, users often lack the requisite knowledge to interpret maximum HR and its implications for training regimens, thereby leading to potential errors in exercise intensity assessments [43].

VO_2 and VO_{2max} Measurements

Simply, the volume of oxygen consumed (VO_2) is the volumetric difference between oxygen inspired and oxygen expired over a measured amount of time (1 minute). VO_{2max} is the maximum amount of oxygen one's body can consume at absolute maximal

XSL•FO
RenderX

exertion. VO_{2max} is used as a gauge to indicate an individual's overall cardiovascular fitness. Commercially available wearables often rely on HR data to estimate VO_2 and VO_{2max}, but this approach can yield significant discrepancies from actual values, with variations reaching as high as 20% [44]. The estimate is only as good as the HR metric, which is estimated from a pulsatile waveform.

Accurate assessment of VO_{2max} typically necessitates maximal exertion testing conducted under laboratory conditions, which is rarely achievable in nonclinical settings. For accurate estimation, VO_{2max} testing must take participants to volitional fatigue, validated by physiological markers, such as blood lactate concentration, respiratory exchange ratio (RER), and individual rate of perceived exertion (RPE) [45]. Most individuals deriving a VO_{2max} estimate from commercial wearables do not achieve a true value because of the absence of rigorous testing protocols. In general, users may reach a "peak" value (VO_2peak), at best.

A user must know their maximum HR and resting HR. There are several formulas that can be used to calculate maximum HR, some more relative to an individual than others and often relying on an age to predict the value. Other methods use a heart rate ratio method (HRRM) in tandem with maximum HR. Because maximum HR is an estimate and linearly related to age, simple formulas may not account for the cardiorespiratory fitness of an individual [46,47]. Differing VO_{2max} values may be acquired for a given person, depending on the device being used, the test being given, and the user experience level, coupled with the different maximum HR equations. Further, many wearables overestimate the resting HR, particularly for endurance athletes, and they may even be outside the validated algorithm range, with many devices having no validation below 40-45 beats per minute (bpm). Moreover, factors such as dehydration and ambient temperature can further distort HR-VO_2 relationships, further complicating accuracy [48,49].

Metabolic Equivalents and Energy Expenditure

Metabolic equivalents (METs) serve as a useful framework for quantifying the energy expenditure (EE) of a given physical activity. One (1) MET is defined as a VO_2 rate of 3.5 mL/kg/minute, equal to an adult's resting metabolic rate [45].

METs are used as a relative way to compare the EE of a given activity to someone's resting metabolic rate over a given unit of time. Activities with higher MET values can be used to assign a specific intensity of exercise, help personal trainers and practitioners plan an exercise routine, and also estimate the number of calories burned during exercise.

For instance, brisk walking registers at approximately 3.5-4.0 METs, while running at 7 mph is around 7.5 METs [50]. However, discrepancies arise when using standard MET values for individuals of varying body weights, as these calculations are based on an average body weight of 70 kg (154 lb). Consequently, comparisons across individuals with different body compositions and adiposity may yield inaccurate assessments [51].

Further, not all METs are equal across sport or exercise type, particularly if full body movement is not involved in the activity [52]. Often, occupational activities and ADLs that elevate HR are not counted in the overall time spent at a given MET.

Moreover, the public's understanding of METs is often restricted to clinical or academic contexts, complicating their practical application. The complexity of these metrics can deter individuals from leveraging them effectively in their personal fitness regimens [53]. Generally, people don't want to perform math to understand how many kilocalories are burned at a given VO_2 and MET value. With accessibility of simpler metrics, they often default to using HR to understand how hard they are working.

Cardiac Output Measurement

The CO (Q) reflects the heart's ability to meet metabolic demands, providing insight into cardiovascular function. It is particularly useful in clinical contexts such as heart failure management [54]. It is calculated by multiplying HR by stroke volume and reflects the volume of blood ejected by the heart per minute. It can be estimated using the Fick principle during exercise testing. It is typically measured using thermodilution, which is invasive and requires an arterial catheter, and can also be estimated with other methods, such as echocardiography and esophageal Doppler.

Recent advancements in technology have enabled the development of devices that purport to measure the CO using PPG, with the signal most often acquired from the fingertip. PPG and electrocardiographs (ECGs) have been used in tandem to aid in determining stroke volume by measuring pulse transit times from the ECG's pulse rate and the pulse measurement estimated by the PPG device. Stroke volume is estimated as a function of the slope transit time and by analyzing the pulse contour of the primary peak in the waveform, though the accuracy of this estimation can be influenced by changes in vascular tone and afterload [55,56]. Limitations exist around PPG's accuracy in determining stroke volume alone [57] and also inferring physiological measurements, such as total peripheral resistance. Some of the PPG applications exhibit error margins approaching 40% when compared to thermodilution and echo Doppler, including those in use cases where patients are surgical in nature [58-61]. Current developments use robust ML methodologies in an attempt to satisfy performance accuracy closer to thermodilution in assessing hemodynamics [62]. Please see Table 1 for an overview of the previously discussed measurements.

XSL•FO
RenderX

Table 1. Summary of noninvasive physiological measurement methods and limitations used widely in the noninvasive assessment of exercise performance.

Measurement	Method	Validation reference standard	Measurement limitations
Steps	Accelerometry (eg, IMU[a])	ActiGraph; manual step counting with video recording	• Unable to account for intensity or activities not reflected by steps (eg, resistance training) [32,35,36] • Misinterpretation of gait by accelerometer or misalignment with GPSb-derived data [63] • Error dependent on anatomical location (1% to >10%) [35,36] • More or less beneficial based on risk profile of individual [4,32,33]
HR[c]	PPG[d] (eg, waveform analysis, peak detection, calculation of bpm[e])	ECG[f], chest strap HR monitor	• Susceptible to motion artifact limiting accuracy during motion [38] • Susceptible to sliding windows limiting accuracy during intensity changes [41] • Not validated or equal across all forms of exercise or cardiac rhythms [42]
VO_2[g], VO_{2max}[h]	Typically inferred from HR using PPG	Laboratory metabolic testing using gas analysis	• Limited by accuracy and calculation of HR and maximum HR, as well as factors that affect HR [44-47,49]
MET[i] and EE[j]	Derived from VO_2 (typically inferred from HR if delivered by a PPG-based wearable)	1 MET, which is a VO_2 of 3.5 mL/kg/minute standardized to 70 kg body mass	• Inequality of MET values across different body compositions [51] • Inequality of MET values across exercise types if full body movement is not involved or during ADLsk [52] • Not well understood by all users/prone to misapplication [53]
CO[l] (Q)	PPG (eg, estimation of stroke volume and HR) with or without other methods (eg, ECG, impedance cardiography)	Transpulmonary thermodilution, transthoracic or esophageal echo Doppler	• High error potential due to physiologic influences (eg, afterload, total peripheral resistance, vascular tone) [55,58-62] • Limitations in interpretation of stroke volume, which may require additional/coordinated inputs (eg, PPG+ECG) [55,57,64]

[a]IMU: inertial measurement unit.

[b]GPS: Global Positioning System.

[c]HR: heart rate.

[d]PPG: photoplethysmography.

[e]bpm: beats per minute.

[f]ECG: electrocardiograph.

[g]VO_2: volume of oxygen consumed.

[h]VO_{2max}: maximum volume of oxygen consumed.

[i]MET: metabolic equivalent.

[j]EE: energy expenditure.

[k]DL: activity of daily living.

[l]CO: cardiac output.

Overcoming Barriers to Effectively Measure Exercise

To make general strides in measuring exercise, two interconnected points need to be addressed, as new technology ideally gleans new biometric insights:

- Some biometric limitations exist because of technological methods. What additional sensing capabilities are available to help improve current measurements and improve the overall measurement of exercise efficacy?
- From newer sensing technologies, what additional metrics can be derived that may lead to a better understanding of an individual's physiology, therefore lending an improved representation of exercise effectiveness, efficiency, and overall efficacy of application?

Advancements in Sensing Capabilities

Noninvasive spectroscopic methods are still fairly limited to the interrogation of the dermis and the relevant substructures and chemistries contained within. Several microneedle studies analyzing dermal blister fluid have refined the understanding of what is quantitatively represented in the skin's interstitial fluid, both somatically (eg, from the blood) and locally (eg, produced by regional cells) [65]. Measuring these putative biomarkers noninvasively poses unique hurdles, as concentrations are typically low and therefore difficult to quantify with acceptable accuracy and precision [66]. The challenge can be exemplified by the multidecade pursuit to measure glucose concentrations in the skin without relying on minimally invasive techniques [67].

Sensor capabilities are expanding, from material selection to hybrid hardware and software integration. Newer optical

XSL•FO
RenderX

approaches may hold promise for measuring individualized biomarkers and chemistries and provide continuous data streaming, some in real time, with little to no significant lag. Further, combining technologies may offer a way to obtain improved sensing performance compared to stand-alone applications. Novel methods, such as those incorporating concepts from link budgeting in telecommunications with bio-optics and AI, may produce sensors less prone to signal interferences and motion artifacts, allowing for cleaner and more accurate data collection during movement [68]. Advances in the engineering and use of novel materials and structures, including metamaterials, may enhance spectroscopic methods by confining light to subwavelength scales, improving integration and sensitivity across a broad optical spectrum [69].

Major advancements in PPG applications focus on dynamic, reconfigurable sensors to optimize signal quality using deep learning algorithms and neural networks [70]. Deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), may improve the analytical performance of PPG-based devices by enhancing analysis of the shape and other characteristics of the waveform. Such algorithms can learn complex patterns and may access biometrics that would be immensely helpful in clinical scenarios [71,72].

Imaging photoplethysmography (iPPG) and remote photoplethysmography (rPPG) methods aim to provide biometric analysis using a contactless approach [73,74]. Off-body cameras are used to assess cardiovascular-based indicators via video by detecting fluctuations in skin blood volume between diastole and systole. Skin color analysis is performed, often of the face. In addition to deep learning and CNNs, spatial-spectral-temporal fusion (hybrid red-green-blue [RGB] camera and near-infrared [NIR] facial video) and spatial-temporal attention networks are a few methods that are applied to increase the accuracy biometric interpolation of video recordings [75-78].

Hybrid sensing is becoming increasingly popular, where PPG is also combined and packaged with additional hardware, such as electrical bioimpedance (BIA). Samsung's new BioActive sensor is an example of a PPG-based noninvasive wearable that has not only expanded the number and type of LEDs but also cointegrated ECG and BIA in an attempt to improve metrological performance from data collected on the dorsal wrist [79]. The cointegration is not novel; however, the sensor includes blue, yellow, violet, and ultraviolet wavelengths, alongside an increase in green, red, and infrared (IR) LEDs, with a claim to improve performance compared to previous versions, in some instances nearly 30%. Other PPG applications related to sport science include interpretation of heat stress and overexertion through the monitoring of HR and heart rate variability (HRV), though the accuracy of these applications can be improved, with errors upward of 0.5 C [80].

Wearable microelectromechanical systems (MEMS) have been playing an increasing role in fitness-based sensing because of their ability to capture IR radiation emitted from the body. Broadband thermal MEMS are a major contributor in noninvasive thermal imaging and thermography, being able to detect fairly small changes in surface body temperature [81,82]. Generally, the sensors contain thermopile elements or microbolometers that operate across a broad IR spectrum, typically in the long-wave infrared (LWIR) range of approximately 8-14 µm. Thermopiles generate a voltage output proportional to the incident IR radiation, while microbolometers exhibit a change in electrical resistance as a function of absorbed thermal energy. When fabricated as focal plane arrays (FPAs), these MEMS-based detectors provide significantly higher spatial resolution compared to single-point thermal sensors, enabling detailed thermal mapping of complex surfaces [83]. Data translate to several performance-related use cases, such as estimating EE and body heat mapping, capturing information related to the physiological processes of both skeletal muscle heat creation and thermoregulation [81]. These advances may lead to potential improvements in understanding training load through entropy analysis, as well as recovery and fatigue assessment [84-86].

A significant amount of research and development has been focused on improving the bandwidth, energy efficiency, and scalability of biosensors. Advancements in short-wave infrared (SWIR) spectroscopy allows for the detection of absorption peaks from 900 to 2500 nm (depending on the source of reference), opening up a plethora of new technological opportunities in biometric monitoring [87-90]. The SWIR region contains absorption peaks for O-H bonds (1430 and 1940 nm), lipid-associated C-H bonds (1210, 1730, and 1760 nm), and collagen (1200 and 1500 nm) [88]. Through an expanded wavelength region, SWIR sensors can target new metrics and enhance the performance of current biomarkers, including hydration, body (dermal) temperature, albumin, glucose, lactate, ethanol, and others [90-92]. Several large players have expanded their R&D in silicon photonic biosensors, including Apple, building off many of the advancements previously made by technology pioneers specializing in the development of laser-based SWIR spectroscopy and photonics-based health sensors, including Rockley Photonics [93,94]. Depending on the overall optical solution, combining silicon photonics with microelectronics improves overall sensing capabilities, with an improved signal-to-noise ratio, lower propagation loss, a smaller overall package size, greater power handling, and overall enhanced performance. Combining SWIR diffuse reflectance spectroscopy with LED-based PPG may further refine and expand biomarker monitoring.

Unlike applications that target specific absorption peaks, applications using broadband-light spectroscopy (BLS) and white-light spectroscopy emit a wide spectrum of light to create a molecular "fingerprint" with relatively high specificity and sensitivity [95]. A label-free approach allows for the use of broadband light without tagged fluro- or chromophores to collect biometrics in a wearable form factor [96]. Some applications are combined with Raman-based techniques to reveal new bands in a given molecular fingerprint region [97]. Others use differentiated InGaAs photodetectors to obtain a broader range of light and a greater signal-to-noise ratio compared to silicon-based or traditional InGaAs solutions [98,99]. Broadband applications may use LEDs or laser diodes and be combined with additional noninvasive methods, such as laser Doppler and continuous-wave near-infrared spectroscopy (CW-NIRS) [100].

XSL•FO
RenderX

Further, emerging research indicates that boron nitride (BN) nanosheet-based photodetectors may offer ultrabroadband sensitivity—potentially spanning from deep ultraviolet to midinfrared—along with high thermal stability, spectral sensitivity, and self-powering capabilities, making them candidates for future applications in breath analysis and glucose sensing [101].

The application of deep optics using CW-NIRS may further improve wearable applications by defining specificity of tissue type through the use of multiple path lengths and algorithmically solving for absorption and scattering coefficients, which accounts for the heterogeneity of layered tissue from superficial to deep [68,102]. This allows the assessment of hematological variables with reliable performance when compared to invasive methods or common benchtop options, including frequency domain near-infrared spectroscopy (FD-NIRS) [103,104]. Unlike FD-NIRS, the CW-NIRS approach eliminates the need for modulated light sources and phase-sensitive detection systems, while maintaining improved signal quality. Recent recommendations aim to improve metrological best practices (data processing and interpretation) if NIRS is used to monitor cerebral and muscle oxygenation during exercise, particularly to account for noise generated by extracerebral tissue layers [105].

Albeit a snapshot of wearable sensing applications, all of these optical approaches are enhanced by the integration of advances in software and computing methods, complex large language models (LLMs), sophisticated and DNNs, and other ML methods including AI, which help further improve on-device processing and compensate for source and sensor limitations. Please see Table 2 for an overview of the previously discussed sensing advancements in exercise metrology.

XSL•FO
RenderX

Table 2. Summary of developing noninvasive sensing methods and emerging biometrics related to exercise physiology.

Sensing method	Sensing advancements	Emerging biometrics
PPG[a]	• Hybrid sensing [77,79] • Expanded wavelengths/photodiode options [79] • MLb and computational advancements (eg, LLMsc, DNNsd, spatial-temporal applications) [70-72,75-78] • Contactless methods (rPPGe, iPPGf) [73,74]	• Heat stress/overexertion [80] • Remote analysis of cardiovascular and respiratory vitals (eg, HRg, respiratoryrate) [73,74] • Apnea, hypopnea [74]
Broadband MEMS[h] thermography	• High spatial resolution and improved thermal mapping [83]	• Entropy analysis, EEi, body heat mapping/thermoregulatory analysis, training load analysis, recovery analysis [81,84-86]
SWIR[j] spectroscopy	• Expanded absorption peak range (900-2500 nm) [88-90] • Improved signal-to-noise ratio, power handling, package size (chip) [94]	• SWIR (diffuse reflectance spectroscopy): ethanol, hydration, dermal temperature, glucose, lactate, lipids [89-92]
BLS[k]/white-light spectroscopy	• Label-free application [96] • Improved InGaAs photodetector solutions [98,99] • Ultrabroadband sensitivity, improved thermal stability with nanosheet photodetectors [101] • Hybrid sensing [97,100]	• Multi-omic and molecular fingerprinting, hydration, breath analysis, glucose [97,98,101]
FD-NIRS[l] and CW-NIRS[m]	• Enhanced signal quality and tissue specificity [104,106] • Deep-tissue analysis [68,102,107]	• Local hematological and oxygenation metrics (eg, hemoglobin, SmO2o), metabolic oxygen kinetics, cerebral oxygenation, bioactive nitric oxide (S-nitrosothiols), blood flow, injury analysis, recovery [102-104,107-111]

[a]PPG: photoplethysmography.

[b]ML: machine learning.

[c]LLM: large language model.

[d]DNN: deep neural network.

[e]rPPG: remote photoplethysmography.

[f]iPPG: imaging photoplethysmography.

[g]HR: heart rate.

[h]MEMS: electromechanical systems.

[i]EE: energy expenditure.

[j]SWIR: short-wave infrared.

[k]BLS: broadband-light spectroscopy.

[l]FD-NIRS: frequency domain near-infrared spectroscopy.

[m]CW-NIRS: continuous-wave near-infrared spectroscopy.

[o]SmO_2: muscle oxygen saturation.

Expanding Physiological Insights

Newer wearables, software applications, and ML tools are generating a variety of indirect metrics to better reflect human performance and fitness status. Many still use PPG-derived measurements with advanced software applications to obtain estimates of aerobic and anaerobic thresholds, the lactate threshold, VLa_{max} (lactate maximum), critical power, critical speed, and more. Although these metrics can provide valuable insights, most commercially available devices do not measure anything truly new and simply deliver an inferred and estimated biometric (eg, strain, recovery, or stress, derived from already-measured information, such as HR or HRV, or a combination that is undisclosed). The degree of accuracy of these inferences relies directly on both device placement and the performance of the underlying inputs, which are sometimes trade secrets. Overall, these metrics do not represent the *quality* or *significance* of exercise.

It is difficult to quantify how effective or efficient an acute session of exercise is for an individual. Often, population statistics are applied in an attempt to assess individualized outcomes relevant to performance or health, without attributing the nuanced limitations of either the technology (ie, an optical model or computational factors) or the body (ie, genetic limitations on oxidative capacity) or even disease. How does one know whether a type or intensity/duration of exercise is working in the short term? How does chronic training impact quality or efficiency? Body composition, improved strength, or decreases in the resting HR are routinely used to assess improvements (or the benefits of exercise) [45,112]. These attributes take weeks or months to denote relevant changes in

XSL•FO
RenderX

physiology. It is not easily possible to assess the benefits of exercise during a workout, nor is it possible to assess recovery at a local level during acute periods between workouts.

Given the limited ability of current wearables to leverage individualized biostatistical data, a key question arises: *Which* biometrics paired with *which* sensing technology, either singlehandedly or in composite, yields the best indicators of exercise quality and efficacy? A holy grail in cardiovascular sports medicine is to measure the impact of activity on blood flow, not only to working muscle, but also to the myocardium, cerebral tissue, and nonworking tissues [113]. These data would provide a link between systemic and local metabolism, including oxygen kinetics, fuel homeostasis, and deviations during non-steady-state activity [114]. Further, blood flow information may provide unprecedented insights into vascular health, athletic injury, and recovery. Advancements in ultrasound techniques, such as superb microvascular imaging and laser Doppler, have opened a window to measuring tissue perfusion and directional blood flow at the 100-200 µm level in lesions, organs, and skeletal muscle [100,115-117] but are not field-practical and cannot be used easily during activity.

Insights gleaned from newly attainable, hematologically derived variables may allow for the measurement of exercise more effectively, and several NIRS techniques are making this realistic [102,104,107-109,118,119]. Obtaining local muscle oxygen saturation (SmO_2), oxygen kinetics and oxidative metabolic activity, bioactive nitric oxide (*S*-nitrosothiols), local blood flow and perfusion, and other hemoglobin or flow-derived indicators during activity would be a monumental feat in cardiovascular assessment. Research supports the enhancement of key performance indicators with such biometrics, including equivalent or improved evaluation of lactate and ventilatory thresholds, as well as critical power in multiple sports [110,111,120,121]. Combined with information from hybrid-sensors, such as hydration and pulse rate, these data may be powerful, with substantial outcomes in denoting increments of value in both professional athletics and clinical care [20].

Conclusion

The array of insights offered by modern wearables holds promise for understanding individual physical limitations, yet it often falls short in providing accurate and meaningful assessments of how well exercise is doing for someone. As technology continues to evolve, a more nuanced understanding of the relationship between exercise and health made possible through improved metrological techniques will be essential for fostering better lifestyle outcomes among individuals.

From the sensing advancements discussed previously, emerging measurements are being unveiled with physiological legitimacy and utility, including networked biomarkers, new relationships between biomarkers, and cause-and-effect chains between biomarkers. Progressing beyond PPG, developers can begin to offer alternative solutions to navigating improvements in physiological insights. With enough reliable data, we may begin to comprehend how measurable concepts, such as metabolomic fingerprinting and biological aging, may indicate where exercise may holistically slow the progression of disease and impact cellular senescence. Though this does not enhance individualized measurement of physiological performance, it does lend to long-term understanding of the influence of exercise on physiology from a real-time, continuous output.

The ultimate goal is to understand how individualized information related to exercise quality and efficacy may help patients, athletes, and practitioners reach therapeutic and training goals without relying on misapplied population-based statistics. Future research should focus on developing methods that bridge the gap between laboratory findings and real-world applicability through the use of real-time, wearable monitoring, which will ultimately empower users to make informed health choices based on truly personalized data.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

Data Availability

No new data were created or analyzed in this study. Data sharing is not applicable to this paper.

Authors' Contributions

Conceptualization, investigation, writing—original draft preparation, and figure preparation were managed by JLC, and writing—review and editing was managed by JLC and EP. All authors have read and agreed to the published version of the manuscript. Generative artificial intelligence was not used in any portion of manuscript preparation.

Conflicts of Interest

JLC and EP are employees of NNOXX Inc and have patents related to the optical design, measurement, and applications.

References

1. Stierman B, Afful J, Carroll MD, Chen T, Davy O, Fink S, et al. National Health and Nutrition Examination Survey 2017-March 2020 prepandemic data files—development of files and prevalence estimates for selected health outcomes. Natl Health Stat Report 2021 Jun 14(158):1-21 [FREE Full text] [doi: 10.15620/cdc:106273] [Medline: 39380201]

XSL•FO
RenderX

2. Heron M, Anderson R. Changes in the leading cause of death: recent patterns in heart disease and cancer mortality. NCHS Data Brief 2016 Aug(254):1-8 [FREE Full text] [Medline: 27598767]
3. Leroux A, Cui E, Smirnova E, Muschelli J, Schrack J, Crainiceanu C. NHANES 2011–2014: objective physical activity is the strongest predictor of all-cause mortality. Med Sci Sports Exerc 2024 Oct 1;56(10):1926-1934 [FREE Full text] [doi: 10.1249/mss.0000000000003497] [Medline: 38949152]
4. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 01;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
5. Saint-Maurice PF, Troiano RP, Bassett DR, Graubard BI, Carlson SA, Shiroma EJ, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020 Mar 24;323(12):1151-1160 [FREE Full text] [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
6. Khan KM, Thompson AM, Blair SN, Sallis JF, Powell KE, Bull FC, et al. Sport and exercise as contributors to the health of nations. Lancet 2012 Jul 07;380(9836):59-64 [FREE Full text] [doi: 10.1016/S0140-6736(12)60865-4] [Medline: 22770457]
7. Garcia L, Pearce M, Abbas A, Mok A, Strain T, Ali S, et al. Non-occupational physical activity and risk of cardiovascular disease, cancer and mortality outcomes: a dose-response meta-analysis of large prospective studies. Br J Sports Med 2023 Aug 28;57(15):979-989 [FREE Full text] [doi: 10.1136/bjsports-2022-105669] [Medline: 36854652]
8. Mitchell T, Barlow CE. Review of the Role of Exercise in Improving Quality of Life in Healthy Individuals and in Those With Chronic Diseases. Current Sports Medicine Reports 2011;10(4):211-216 [FREE Full text] [doi: 10.1249/jsr.0b013e318223cc9e]
9. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626 [FREE Full text] [doi: 10.1016/s2589-7500(22)00111-x]
10. Zheng H, Orsini N, Amin J, Wolk A, Nguyen VTT, Ehrlich F. Quantifying the dose-response of walking in reducing coronary heart disease risk: meta-analysis. Eur J Epidemiol 2009 Mar 22;24(4):181-192 [FREE Full text] [doi: 10.1007/s10654-009-9328-9] [Medline: 19306107]
11. Tanabe T, Maeda S, Miyauchi T, Iemitsu M, Takanashi M, Irukayama‐Tomobe Y, et al. Exercise training improves ageing‐induced decrease in eNOS expression of the aorta. Acta Physiol Scand 2003 Apr 25;178(1):3-10 [FREE Full text] [doi: 10.1046/j.1365-201x.2003.01100.x]
12. Naseem KM. The role of nitric oxide in cardiovascular diseases. Mol Aspects Med 2005 Apr;26(1-2):33-65 [FREE Full text] [doi: 10.1016/j.mam.2004.09.003] [Medline: 15722114]
13. Montero D, Breenfeldt-Andersen A, Oberholzer L, Haider T, Goetze J, Meinild-Lundby A, et al. Erythropoiesis with endurance training: dynamics and mechanisms. Am J Physiol Regul Integr Comp Physiol 2017 Jun 01;312(6):R894-R902 [FREE Full text] [doi: 10.1152/ajpregu.00012.2017] [Medline: 28381454]
14. United States Department of Health and Human Services. Physical Activity Guidelines for Americans, 2nd edition. Washington, DC: United States Department of Health and Human Services; 2018.
15. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
16. Canali S, Schiaffonati V, Aliverti A. Challenges and recommendations for wearable devices in digital health: data quality, interoperability, health equity, fairness. PLOS Digit Health 2022 Oct 13;1(10):e0000104 [FREE Full text] [doi: 10.1371/journal.pdig.0000104] [Medline: 36812619]
17. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 03;313(5):459-460 [FREE Full text] [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
18. Loland S. Technology in sport: three ideal-typical views and their implications. In: The Ethics of Sports Technologies and Human Enhancemen. London, UK: Routledge; 2017:163-172.
19. Patel V, Orchanian-Cheff A, Wu R. Evaluating the validity and utility of wearable technology for continuously monitoring patients in a hospital setting: systematic review. JMIR Mhealth Uhealth 2021 Aug 18;9(8):e17411 [FREE Full text] [doi: 10.2196/17411] [Medline: 34406121]
20. Corso JL. Overcoming wearable market limitations through bioactive nitric oxide monitoring. NNOXX Inc. 2024. URL: https://tinyurl.com/466ncdmz [accessed 2025-11-27]
21. Elgendi M, Haugg F, Fletcher RR, Allen J, Shin H, Alian A, et al. Recommendations for evaluating photoplethysmography-based algorithms for blood pressure assessment. Commun Med (Lond) 2024 Jul 12;4(1):140 [FREE Full text] [doi: 10.1038/s43856-024-00555-2] [Medline: 38997447]
22. Park J, Seok HS, Kim S, Shin H. Photoplethysmogram analysis and applications: an integrative review. Front Physiol 2022 Mar 1;12:808451 [FREE Full text] [doi: 10.3389/fphys.2021.808451] [Medline: 35300400]
23. Jacques SL. Optical properties of biological tissues: a review. Phys. Med. Biol 2013 Jun 07;58(11):R37-R61 [FREE Full text] [doi: 10.1088/0031-9155/58/11/r37]
24. Lipnick MS, Ehie O, Igaga EN, Bicker P. Pulse oximetry and skin pigmentation—new guidance from the FDA. JAMA 2025 Apr 22;333(16):1393-1395 [FREE Full text] [doi: 10.1001/jama.2025.1959] [Medline: 40042945]

XSL•FO
RenderX

25. Rathod M, Ross HJ, Franklin D. Improving the accuracy and equity of pulse oximeters: collaborative recommendations. JACC Adv 2022 Oct;1(4):100118 [FREE Full text] [doi: 10.1016/j.jacadv.2022.100118] [Medline: 38939706]
26. Wijshoff RWCGR, Mischi M, Aarts RM. Reduction of periodic motion artifacts in photoplethysmography. IEEE Trans Biomed Eng 2017 Jan;64(1):196-207 [FREE Full text] [doi: 10.1109/tbme.2016.2553060]
27. Qi Y, Zhang A, Ma Y, Wang H, Li J. Interference source-based quality assessment method for postauricular photoplethysmography signals. Biomed Signal Process Control 2023 Jul;84:104751 [FREE Full text] [doi: 10.1016/j.bspc.2023.104751]
28. Hartmann V, Liu H, Chen F, Qiu Q, Hughes S, Zheng D. Quantitative comparison of photoplethysmographic waveform characteristics: effect of measurement site. Front Physiol 2019;10:198 [FREE Full text] [doi: 10.3389/fphys.2019.00198] [Medline: 30890959]
29. Dias FM, Cardenas DAC, Toledo MAF, Oliveira FAC, Ribeiro E, Krieger JE, et al. Exploring the limitations of blood pressure estimation using the photoplethysmography signal. Physiol Meas 2025 Apr 22;46(4):045007 [FREE Full text] [doi: 10.1088/1361-6579/adcb86] [Medline: 40209759]
30. Qin C, Wang X, Xu G, Ma X. Advances in cuffless continuous blood pressure monitoring technology based on PPG signals. Biomed Res Int 2022 Oct;2022(1):8094351 [FREE Full text] [doi: 10.1155/2022/8094351] [Medline: 36217389]
31. Del Pozo Cruz B, Ahmadi MN, Lee I, Stamatakis E. Prospective associations of daily step counts and intensity with cancer and cardiovascular disease incidence and mortality and all-cause mortality. JAMA Intern Med 2022 Nov 01;182(11):1139-1148 [FREE Full text] [doi: 10.1001/jamainternmed.2022.4000] [Medline: 36094529]
32. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al. Association of step counts over time with the risk of chronic disease in the All of Us research program. Nat Med 2022 Nov 10;28(11):2301-2308 [FREE Full text] [doi: 10.1038/s41591-022-02012-w] [Medline: 36216933]
33. Banach M, Lewek J, Surma S, Penson P, Sahebkar A, Martin S, et al. The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis. Eur J Prev Cardiol 2023 Dec 21;30(18):1975-1985 [FREE Full text] [doi: 10.1093/eurjpc/zwad229] [Medline: 37555441]
34. Stens NA, Bakker EA, Mañas A, Buffart LM, Ortega FB, Lee D, et al. Relationship of daily step counts to all-cause mortality and cardiovascular events. J Am Coll Cardiol 2023 Oct 10;82(15):1483-1494 [FREE Full text] [doi: 10.1016/j.jacc.2023.07.029] [Medline: 37676198]
35. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul 22;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
36. Pan J, Wei S. Accuracy and reliability of accelerometer-based pedometers in step counts during walking, running, and stair climbing in different locations of attachment. Sci Rep 2024 Nov 12;14(1):27761 [FREE Full text] [doi: 10.1038/s41598-024-78684-w] [Medline: 39532924]
37. McLeod K, Jones M, Thom J, Parmenter B. Resistance training and high-intensity interval training improve cardiometabolic health in high risk older adults: a systematic review and meta-anaylsis. Int J Sports Med 2022 Mar;43(3):206-218 [FREE Full text] [doi: 10.1055/a-1560-6183] [Medline: 34320660]
38. Prieto-Avalos G, Cruz-Ramos NA, Alor-Hernández G, Sánchez-Cervantes JL, Rodríguez-Mazahua L, Guarneros-Nolasco LR. Wearable devices for physical monitoring of heart: a review. Biosensors (Basel) 2022 May 02;12(5):292 [FREE Full text] [doi: 10.3390/bios12050292] [Medline: 35624593]
39. Mühlen JM, Stang J, Lykke Skovgaard E, Judice PB, Molina-Garcia P, Johnston W, et al. Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE Network. Br J Sports Med 2021 Jul 04;55(14):767-779 [FREE Full text] [doi: 10.1136/bjsports-2020-103148] [Medline: 33397674]
40. Chang X, Li G, Xing G, Hao T, Tu L. DeepHeart: accurate heart rate estimation from PPG signals based on deep learning. 2019 Presented at: IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS); November 4-7, 2019; Monterey, CA p. 371-379. [doi: 10.1109/MASS.2019.00051]
41. Kong Y, Chon KH. Heart rate tracking using a wearable photoplethysmographic sensor during treadmill exercise. IEEE Access 2019;7:152421-152428 [FREE Full text] [doi: 10.1109/access.2019.2948107]
42. Quinn R, Leader N, Lebovic G, Chow C, Dorian P. Accuracy of wearable heart rate monitors during exercise in sinus rhythm and atrial fibrillation. J Am Coll Cardiol 2024 Mar 26;83(12):1177-1179 [FREE Full text] [doi: 10.1016/j.jacc.2024.01.024] [Medline: 38508851]
43. Mann T, Lamberts RP, Lambert MI. Methods of prescribing relative exercise intensity: physiological and practical considerations. Sports Med 2013 Jul;43(7):613-625 [FREE Full text] [doi: 10.1007/s40279-013-0045-x] [Medline: 23620244]
44. Caserman P, Yum S, Göbel S, Reif A, Matura S. Assessing the accuracy of smartwatch-based estimation of maximum oxygen uptake using the Apple watch series 7: validation study. JMIR Biomed Eng 2024 Jul 31;9:e59459 [FREE Full text] [doi: 10.2196/59459] [Medline: 39083800]
45. Powers S, Howley E. Exercise Physiology: Theory and Application to Fitness and Performance. 10th ed. New York, NY: McGraw-Hill Education; 2017.

XSL•FO
RenderX

46. Ozemek C, Whaley MH, Finch WH, Kaminsky LA. Maximal heart rate declines linearly with age independent of cardiorespiratory fitness levels. Eur J Sport Sci 2017 Jun;17(5):563-570 [FREE Full text] [doi: 10.1080/17461391.2016.1275042] [Medline: 28099086]
47. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol 2001 Jan;37(1):153-156 [FREE Full text] [doi: 10.1016/s0735-1097(00)01054-8] [Medline: 11153730]
48. Smith D, Fernhall B. Advanced Cardiovascular Exercise Physiology. 1st ed. Champaign, IL: Human Kinetics; 2010.
49. Astrand P, Rodahl K, Dahl H, Stromme S. Textbook of Work Physiology: Physiological Bases of Exercise. 4th ed. Champaign, IL: Human Kinetics; 2003.
50. Ainsworth B, Haskell W, Herrmann S, Meckes N, Bassett D, Tudor-Locke C, et al. 2011 Compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581 [FREE Full text] [doi: 10.1249/mss.0b013e31821ece12]
51. Tompuri TT. Metabolic equivalents of task are confounded by adiposity, which disturbs objective measurement of physical activity. Front Physiol 2015 Aug 11;6:226 [FREE Full text] [doi: 10.3389/fphys.2015.00226] [Medline: 26321958]
52. Holtermann A, Stamatakis E. Do all daily metabolic equivalent task units (METs) bring the same health benefits? Br J Sports Med 2019 Aug;53(16):991-992 [FREE Full text] [doi: 10.1136/bjsports-2017-098693] [Medline: 29794085]
53. Sharkey B, Gaskill S. Fitness & Health. Champaign, IL: Human Kinetics; 2013.
54. Hassan M, Wagdy K, Kharabish A, Selwanos PP, Nabil A, Elguindy A, et al. Validation of noninvasive measurement of cardiac output using inert gas rebreathing in a cohort of patients with heart failure and reduced ejection fraction. Circ Heart Failure 2017 Mar;10(3):e003592 [FREE Full text] [doi: 10.1161/circheartfailure.116.003592]
55. Colquhoun D, Dunn LK, McMurry T, Thiele RH. The relationship between the area of peripherally-derived pressure volume loops and systemic vascular resistance. J Clin Monit Comput 2013 Dec;27(6):689-696 [FREE Full text] [doi: 10.1007/s10877-013-9493-y] [Medline: 23881417]
56. Lee J, Sohn J, Park J, Yang S, Lee S, Kim HC. Novel blood pressure and pulse pressure estimation based on pulse transit time and stroke volume approximation. Biomed Eng Online 2018 Jun 18;17(1):81 [FREE Full text] [doi: 10.1186/s12938-018-0510-8] [Medline: 29914491]
57. Liu S, Wang J, Chen W, Pan K, Su C. Classification of photoplethysmographic signal quality with fuzzy neural network for improvement of stroke volume measurement. Appl Sci 2020 Feb 21;10(4):1476 [FREE Full text] [doi: 10.3390/app10041476]
58. Chen G, Meng L, Alexander B, Tran NP, Kain ZN, Cannesson M. Comparison of noninvasive cardiac output measurements using the Nexfin monitoring device and the esophageal Doppler. J Clin Anesth 2012 Jun;24(4):275-283 https://pubmed.ncbi.nlm.nih.gov/22608581/. [doi: 10.1016/j.jclinane.2011.08.014] [Medline: 22608581]
59. van der Spoel AG, Voogel AJ, Folkers A, Boer C, Bouwman RA. Comparison of noninvasive continuous arterial waveform analysis (Nexfin) with transthoracic Doppler echocardiography for monitoring of cardiac output. J Clin Anesth 2012 Jun;24(4):304-309 [FREE Full text] [doi: 10.1016/j.jclinane.2011.09.008] [Medline: 22608585]
60. Hofhuizen C, Lansdorp B, van der Hoeven JG, Scheffer G, Lemson J. Validation of noninvasive pulse contour cardiac output using finger arterial pressure in cardiac surgery patients requiring fluid therapy. J Crit Care 2014 Feb;29(1):161-165 [FREE Full text] [doi: 10.1016/j.jcrc.2013.09.005] [Medline: 24475498]
61. Broch O, Renner J, Gruenewald M, Meybohm P, Schöttler J, Caliebe A, et al. A comparison of the Nexfin® and transcardiopulmonary thermodilution to estimate cardiac output during coronary artery surgery. Anaesthesia 2012 Apr 11;67(4):377-383 [FREE Full text] [doi: 10.1111/j.1365-2044.2011.07018.x] [Medline: 22324797]
62. Callejas Pastor CA, Oh C, Hong B, Ku Y. Machine learning-based cardiac output estimation using photoplethysmography in off-pump coronary artery bypass surgery. J Clin Med 2024 Nov 26;13(23):7145 [FREE Full text] [doi: 10.3390/jcm13237145] [Medline: 39685605]
63. Bramble DM, Lieberman DE. Endurance running and the evolution of Homo. Nature 2004 Nov 18;432(7015):345-352 [FREE Full text] [doi: 10.1038/nature03052] [Medline: 15549097]
64. Calderon C, Borras C, Sierra D, Niño C, Conde C. Stroke volume and cardiac output estimation using pulse wave analysis on pulse oximetry and ECG data. 2013 Presented at: Pan American Health Care Exchanges (PAHCE); April 29-May 4, 2013; Medellin, Colombia URL: https://ieeexplore.ieee.org/document/6568266 [doi: 10.1109/pahce.2013.6568266]
65. Samant PP, Niedzwiecki MM, Raviele N, Tran V, Mena-Lapaix J, Walker DI, et al. Sampling interstitial fluid from human skin using a microneedle patch. Sci Transl Med 2020 Nov 25;12(571):eaaw0285 [FREE Full text] [doi: 10.1126/scitranslmed.aaw0285] [Medline: 33239384]
66. Ahmad Tarar A, Mohammad U, K Srivastava S. Wearable skin sensors and their challenges: a review of transdermal, optical, and mechanical sensors. Biosensors (Basel) 2020 May 28;10(6):56 [FREE Full text] [doi: 10.3390/bios10060056] [Medline: 32481598]
67. Smith J. The Pursuit of Noninvasive Glucose: Hunting the Deceitful Turkey. 9th ed. Online (self-published): https://www.researchgate.net/publication/376199071_The_Pursuit_of_Noninvasive_Glucose_9th_Edition; 2023.
68. Peikon E, Saul J, Coles B. Spectrometry systems and methods. International Patent U.S. Patent Application No. 63/477,989; WO2024/145419A1. 2024 Jul 4. URL: https://patentscope.wipo.int/search/en/WO2024145419 [accessed 2025-11-28]

XSL•FO
RenderX

69. Li D, Wu X, Chen Z, Liu T, Mu X. Surface-enhanced spectroscopy technology based on metamaterials. Microsyst Nanoeng 2025 Apr 03;11(1):60 [FREE Full text] [doi: 10.1038/s41378-025-00905-7] [Medline: 40180920]
70. Ngoc-Thang B, Tien Nguyen TM, Truong TT, Nguyen BL, Nguyen TT. A dynamic reconfigurable wearable device to acquire high quality PPG signal and robust heart rate estimate based on deep learning algorithm for smart healthcare system. Biosens Bioelectron X 2022 Dec;12:100223 [FREE Full text] [doi: 10.1016/j.biosx.2022.100223]
71. Baker S, Xiang W, Atkinson I. Determining respiratory rate from photoplethysmogram and electrocardiogram signals using respiratory quality indices and neural networks. PLoS One 2021 Apr 8;16(4):e0249843 [FREE Full text] [doi: 10.1371/journal.pone.0249843] [Medline: 33831075]
72. Kim JW, Choi S. Normalization of photoplethysmography using deep neural networks for individual and group comparison. Sci Rep 2022 Feb 24;12(1):3133 [FREE Full text] [doi: 10.1038/s41598-022-07107-5] [Medline: 35210522]
73. Saikevičius L, Raudonis V, Kozlovskaja-Gumbrienė A, Šakalytė G. Advancements in remote photoplethysmography. Electronics 2025 Mar 03;14(5):1015 [FREE Full text] [doi: 10.3390/electronics14051015]
74. Nardelli M, Bailón R. Advances in wearable photoplethysmography applications in health monitoring. Sensors (Basel) 2023 Aug 10;23(16):7064 [FREE Full text] [doi: 10.3390/s23167064] [Medline: 37631601]
75. Firmansyah RA, Prabowo YA, Suheta T, Muharom S. Implementation of 1D convolutional neural network for improvement remote photoplethysmography measurement. Indones J Electr Eng Comput Sci 2023 Mar 01;29(3):1326 [FREE Full text] [doi: 10.11591/ijeecs.v29.i3.pp1326-1335]
76. Przybyło J. A deep learning approach for remote heart rate estimation. Biomedical Signal Processing and Control 2022 Apr;74:103457 [FREE Full text] [doi: 10.1016/j.bspc.2021.103457]
77. Kado S, Monno Y, Yoshizaki K, Tanaka M, Okutomi M. Spatial-spectral-temporal fusion for remote heart rate estimation. IEEE Sens J 2020 Oct 1;20(19):11688-11697 [FREE Full text] [doi: 10.1109/jsen.2020.2997785]
78. Hu M, Guo D, Jiang M, Qian F, Wang X, Ren F. rPPG-based heart rate estimation using spatial-temporal attention network. IEEE Trans Cogn Dev Syst 2022 Dec;14(4):1630-1641 [FREE Full text] [doi: 10.1109/tcds.2021.3131197]
79. Unlocking new possibilities for preventative wellness with new Galaxy watch and BioActive sensor. Samsung Newsroom. 2024. URL: https://news.samsung.com/us/unlocking-new-possibilities-for-preventative-wellness-with-new-galaxy-watch-and-bioactive-sensor [accessed 2025-08-18]
80. Richard N, Cheung S, Claydon V, Koehle M, Coté AT. Accuracy and precision of the SlateSafety BandV2 and CORE devices in estimating resting and moderate hyperthermic exercise temperature in eumenorrheic females. Int J Sports Physiol Perform 2025 Aug 01;20(8):1068-1078 [FREE Full text] [doi: 10.1123/ijspp.2025-0034] [Medline: 40537134]
81. Perpetuini D, Formenti D, Cardone D, Trecroci A, Rossi A, Di Credico A, et al. Can data-driven supervised machine learning approaches applied to infrared thermal imaging data estimate muscular activity and fatigue? Sensors (Basel) 2023 Jan 11;23(2):832 [FREE Full text] [doi: 10.3390/s23020832] [Medline: 36679631]
82. Gardner J, Varadan V, Awadelkarim O. Microsensors, MEMS, and Smart Devices. Reprinted. Chichester Weinheim: Wiley; 2005.
83. Wang B, Zhao X, Dong T, Li B, Zhang F, Su J, et al. Convergence of thermistor materials and focal plane arrays in uncooled microbolometers: trends and perspectives. Nanomaterials (Basel) 2025 Aug 27;15(17):1316 [FREE Full text] [doi: 10.3390/nano15171316] [Medline: 40937994]
84. Hu C, Du N, Liu Z, Song Y. Can infrared thermal imaging reflect exercise load? An incremental cycling exercise study. Bioengineering (Basel) 2025 Mar 11;12(3):280 [FREE Full text] [doi: 10.3390/bioengineering12030280] [Medline: 40150744]
85. Hadžić V, Širok B, Malneršič A, Čoh M. Can infrared thermography be used to monitor fatigue during exercise? A case study. J Sport Health Sci 2019 Jan;8(1):89-92 [FREE Full text] [doi: 10.1016/j.jshs.2015.08.002] [Medline: 30719388]
86. Bogomilsky S, Hoffer O, Shalmon G, Scheinowitz M. Preliminary study of thermal density distribution and entropy analysis during cycling exercise stress test using infrared thermography. Sci Rep 2022 Aug 18;12(1):14018 [FREE Full text] [doi: 10.1038/s41598-022-18233-5] [Medline: 35982085]
87. Salimi M, Roshanfar M, Tabatabaei N, Mosadegh B. Machine learning-assisted short-wave infrared (SWIR) techniques for biomedical applications: towards personalized medicine. J Pers Med 2023 Dec 26;14(1):33 [FREE Full text] [doi: 10.3390/jpm14010033] [Medline: 38248734]
88. Mironov S, Hwang CD, Nemzek J, Li J, Ranganathan K, Butts JT, et al. Short-wave infrared light imaging measures tissue moisture and distinguishes superficial from deep burns. Wound Repair Regen 2020 Mar;28(2):185-193 [FREE Full text] [doi: 10.1111/wrr.12779] [Medline: 31675450]
89. VerSteeg B, Newhouse T, Zhang Y, Driscoll J, Nelson D, McCann D, et al. Silicon photonics based ultra-compact short-wave infrared laser spectrometer platform for non-invasive wearable sensors. 2023 Presented at: Next-Generation Spectroscopic Technologies XV; April 30-May 4, 2023; Orlando, FL URL: https://tinyurl.com/33kx5zr9 [doi: 10.1117/12.2664085]
90. Brolis sensor technology. Brolis. URL: https://brolis-sensor.com/ [accessed 2025-09-10]
91. Corso J, Paul M, Laxminarayan S, Merritt S. Non-invasive real-time biosensing of physiologic hydration status via ultra-compact laser-based short-wave infrared spectrophotometry. 2023 Presented at: Next-Generation Spectroscopic Technologies XV; April 30-May 4, 2023; Orlando, FL URL: https://tinyurl.com/bp55u7jn [doi: 10.1117/12.2664051]

XSL•FO
RenderX

92. Driscoll J, Perea P, Kauffman A, Zilkie A, Steeg B. Pioneering silicon photonics for wearable sensors. 2023 Presented at: Optical Fiber Communication Conference (OFC) 2023; March 5-9, 2023; San Diego, CA URL: https://www.researchgate.net/publication/370887647_Pioneering_Silicon_Photonics_for_Wearable_Sensors [doi: 10.1364/ofc.2023.th1a.6]
93. Zilkie A. Silicon photonics and AI for advanced wearable health sensors. 2022 Presented at: Smart Photonic and Optoelectronic Integrated Circuits 2022; January 22-27, 2022; San Francisco, CA URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/0/PC120050/Silicon-photonics-and-AI-for-advanced-wearable-health-sensors/10.1117/12.2612761.short [doi: 10.1117/12.2612761]
94. Zilkie A. Multi-micron silicon photonics platform for imaging and sensing. 2020 Presented at: Advanced Photon Counting Techniques XIV; April 27-May 8, 2020; Online (United States) URL: https://tinyurl.com/yeytxjdr [doi: 10.1117/12.2571184]
95. Eissa T, Leonardo C, Kepesidis KV, Fleischmann F, Linkohr B, Meyer D, et al. Plasma infrared fingerprinting with machine learning enables single-measurement multi-phenotype health screening. Cell Rep Med 2024 Jul 16;5(7):101625 [FREE Full text] [doi: 10.1016/j.xcrm.2024.101625] [Medline: 38944038]
96. Dhanabalan SS, Sriram S, Walia S, Avaninathan SR, Carrasco MF, Bhaskaran M. Wearable label-free optical biodetectors: progress and perspectives. Adv Photonics Res 2020 Dec 18;2(2):2000076 [FREE Full text] [doi: 10.1002/adpr.202000076]
97. Murakami Y, Ando M, Imamura A, Oketani R, Leproux P, Honjoh S, et al. Molecular fingerprinting of mouse brain using ultrabroadband coherent anti-Stokes Raman scattering (CARS) microspectroscopy empowered by multivariate curve resolution-alternating least squares (MCR-ALS). Chem Biomed Imaging 2024 Oct 28;2(10):689-697 [FREE Full text] [doi: 10.1021/cbmi.4c00034] [Medline: 39483635]
98. NibbleBeat 2.0α ChipSenseTM evaluation device. MantiSpectra. 2025. URL: https://mantispectra.com/wp-content/uploads/2025/09/NibbleBeat-2.0%CE%B1-Datasheet-2508.pdf [accessed 2025-11-28]
99. van KA, van ED, Li C, Petruzzella M, Hakkel K, Ou F, et al. High-performance photodetector arrays for near-infrared spectral sensing. APL Photonics 2023 Apr 10;8(4):041302 [FREE Full text] [doi: 10.1063/5.0136921]
100. Scholz O, Nowicki C, Warmerdam E, Rother S, Ganse B. New sensor options for smart fracture implants and wearable devices: laser-Doppler and white-light spectroscopy allow monitoring of bone regeneration via perfusion measurement. Biosens Bioelectron 2025 Jul 15;280:117442 [FREE Full text] [doi: 10.1016/j.bios.2025.117442] [Medline: 40199098]
101. Feng PX, Pacheco Cabrera E, Chu J, Zhou B, Flores SY, Peng X, et al. Self-powered photodetectors with ultra-broad spectral response and thermal stability for broadband, energy efficient wearable sensing and optoelectronics. Molecules 2025 Jul 08;30(14):2897 [FREE Full text] [doi: 10.3390/molecules30142897] [Medline: 40733163]
102. Hamaoka T, McCully K, Niwayama M, Chance B. The use of muscle near-infrared spectroscopy in sport, health and medical sciences: recent developments. Philos Trans A Math Phys Eng Sci 2011 Nov 28;369(1955):4591-4604 [FREE Full text] [doi: 10.1098/rsta.2011.0298] [Medline: 22006908]
103. Fantini S, Sassaroli A. Frequency-domain techniques for cerebral and functional near-infrared spectroscopy. Front Neurosci 2020 Apr 7;14:300 [FREE Full text] [doi: 10.3389/fnins.2020.00300] [Medline: 32317921]
104. Peikon E, Corso JL, Otic N, Kierul O, Franceschini MA, Robinson M. Continuous monitoring of muscle oxygenation in endurance athletes during incremental cycling: experimental validation of a wearable continuous-wave NIRS sensor using frequency-domain near-infrared spectroscopy. Bioengineering (Basel) 2025 Oct 24;12(11):1153 [FREE Full text] [doi: 10.3390/bioengineering12111153] [Medline: 41301108]
105. Quaresima V, Ferrari M, Scholkmann F. Best practices for simultaneous measurement of NIRS-based cerebral and muscle oximetry during exercise. J Sport Health Sci 2024 Oct 17;14:100997 [FREE Full text] [doi: 10.1016/j.jshs.2024.100997] [Medline: 39424058]
106. Kleiser S, Nasseri N, Andresen B, Greisen G, Wolf M. Comparison of tissue oximeters on a liquid phantom with adjustable optical properties. Biomed Opt Express 2016 Jul 11;7(8):2973-2992 [FREE Full text] [doi: 10.1364/boe.7.002973]
107. Feldmann A, Schmitz R, Erlacher D. Near-infrared spectroscopy-derived muscle oxygen saturation on a 0% to 100% scale: reliability and validity of the Moxy Monitor. J Biomed Opt 2019 Nov 1;24(11):1-11 [FREE Full text] [doi: 10.1117/1.jbo.24.11.115001] [Medline: 31741352]
108. Stamler J, Peikon E. Noninvasive measurement of endogenous S-nitrosothiols. U.S. Patent Publication No. 2023/0050352A1; International Publication Number WO2023/01886A1, 16 February 2024; Australian Patent Publication No. AU2022326479B2, 6 March 2024; Hong Kong Patent No. HK40112867A. 2024. URL: https://patents.google.com/patent/HK40112876A/en [accessed 2025-01-28]
109. Cooper C, Springett R, Panagiotopoulou A, Penrice J. Near infrared spectroscopy of nitrosyl haemoglobin--relevance to in vivo detection of nitric oxide. Biochem Soc Trans 1996 Aug;24(3):448S [FREE Full text] [doi: 10.1042/bst024448s] [Medline: 8878992]
110. Batterson PM, Kirby BS, Hasselmann G, Feldmann A. Muscle oxygen saturation rates coincide with lactate-based exercise thresholds. Eur J Appl Physiol 2023 Oct 01;123(10):2249-2258 [FREE Full text] [doi: 10.1007/s00421-023-05238-9] [Medline: 37261552]
111. Feldmann A, Erlacher D. Critical oxygenation: can muscle oxygenation inform us about critical power? Med Hypotheses 2021 May;150:110575 [FREE Full text] [doi: 10.1016/j.mehy.2021.110575] [Medline: 33857860]
112. Ozemek C, Bonikowske A, Christle J, Gallo P. ACSM's Guidelines for Exercise Testing and Prescription, 12 ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2025.

XSL•FO
RenderX

113. Allen BW, Stamler JS, Piantadosi CA. Hemoglobin, nitric oxide and molecular mechanisms of hypoxic vasodilation. Trends Mol Med 2009 Oct;15(10):452-460 [FREE Full text] [doi: 10.1016/j.molmed.2009.08.002] [Medline: 19781996]
114. Casey DP, Curry TB, Joyner MJ. Measuring muscle blood flow: a key link between systemic and regional metabolism. Curr Opin Clin Nutr Metab Care 2008 Sep;11(5):580-586 [FREE Full text] [doi: 10.1097/MCO.0b013e32830b5b34] [Medline: 18685453]
115. Akkoc O, Caliskan E. The impact of strenght electromyostimulation training on muscle stiffness and blood flow: an exploration using shear wave elastography and superb microvascular imaging. Univers J Educ Res 2019 Jan;7(1):80-86 [FREE Full text] [doi: 10.13189/ujer.2019.070111]
116. Sato W, Suto Y, Yamanaka T, Watanabe H. An advanced ultrasound application used to assess peripheral vascular diseases: superb microvascular imaging. J Echocardiogr 2021 Sep 15;19(3):150-157 [FREE Full text] [doi: 10.1007/s12574-021-00527-8] [Medline: 33856650]
117. Fu Z, Zhang J, Lu Y, Wang S, Mo X, He Y, et al. Clinical applications of superb microvascular imaging in the superficial tissues and organs: a systematic review. Acad Radiol 2021 May;28(5):694-703 [FREE Full text] [doi: 10.1016/j.acra.2020.03.032] [Medline: 32418782]
118. Reynolds JD, Jenkins T, Matto F, Nazemian R, Farhan O, Morris N, et al. Pharmacologic targeting of red blood cells to improve tissue oxygenation. Clin Pharmacol Ther 2018 Sep 17;104(3):553-563 [FREE Full text] [doi: 10.1002/cpt.979] [Medline: 29238951]
119. Premont RT, Reynolds JD, Zhang R, Stamler JS. Role of nitric oxide carried by hemoglobin in cardiovascular physiology: developments on a three-gas respiratory cycle. Circ Res 2020 Jan 03;126(1):129-158 [FREE Full text] [doi: 10.1161/CIRCRESAHA.119.315626] [Medline: 31590598]
120. Perrey S, Quaresima V, Ferrari M. Muscle oximetry in sports science: an updated systematic review. Sports Med 2024 Apr;54(4):975-996 [FREE Full text] [doi: 10.1007/s40279-023-01987-x] [Medline: 38345731]
121. Feldmann A, Ammann L, Gächter F, Zibung M, Erlacher D. Muscle oxygen saturation breakpoints reflect ventilatory thresholds in both cycling and running. J Hum Kinet 2022 Aug;83:87-97 [FREE Full text] [doi: 10.2478/hukin-2022-0054] [Medline: 36157967]

Abbreviations

ADL: activity of daily living
AHA: American Heart Association
AI: artificial intelligence
BIA: bioimpedance
BLS: broadband-light spectroscopy
bpm: beats per minute
CNN: convolutional neural network
CO: cardiac output
CW-NIRS: continuous-wave near-infrared spectroscopy
DNN: deep neural network
ECG: electrocardiograph
EE: energy expenditure
eNOS: endothelial nitric oxide synthase
FD-NIRS: frequency domain near-infrared spectroscopy
GPS: Global Positioning System
HR: heart rate
HRV: heart rate variability
IMU: inertial measurement unit
iPPG: imaging photoplethysmography
IR: infrared
LED: light-emitting diode
LLM: large language model
MEMS: electromechanical systems
MET: metabolic equivalent
ML: machine learning
NIR: near infrared
NIRS: near-infrared spectroscopy
PPG: photoplethysmography
PPI: pulse-pulse interval
rPPG: remote photoplethysmography
SmO2: muscle oxygen saturation

XSL•FO
RenderX

SWIR: short-wave infrared
VO_2: volume of oxygen consumed
VO_{2max}: maximum volume of oxygen consumed

Edited by A Stone; submitted 19.06.25; peer-reviewed by WNJM Colier, M Liu; comments to author 06.08.25; accepted 25.11.25; published 10.12.25.

Please cite as:
Corso JL, Peikon E
The Challenge of Measuring Exercise: Advancing Metrological Barriers in Wearable Sensing
JMIR Mhealth Uhealth 2025;13:e79347
URL: https://mhealth.jmir.org/2025/1/e79347
doi:10.2196/79347
PMID:

©Jennifer L Corso, Evan Peikon. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Wearable Technologies in Head and Neck Oncology: Scoping Review

Matthew Shammas-Toma[1*], BSc; Gianluca Sampieri[2*], MD, MSc; Michael Xie[3], MD; Aishwaria Maxwell[1], MSc; Alex Esemezie[3], BSc; Quynh Pham[4], PhD; Joseph A Cafazzo[4], PhD; Philip Wong[5], MD, MSc; C Jillian Tsai[5], PhD; David P Goldstein[3], MD, MSc; John R de Almeida[3], MD, MSc; Ervin Sejdic[6,7], PhD; Christopher M K L Yao[3], MD

[1]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Department of Otolaryngology - Head and Neck Surgery, University of Toronto, Toronto, ON, Canada

[3]Department of Otolaryngology - Head and Neck Surgery, Princess Margaret Cancer Center, University Health Network, 610 University Avenue, Toronto, ON, Canada

[4]Toronto General Hospital, Center for Digital Therapeutics, University Health Network, Toronto, ON, Canada

[5]Department of Radiation Oncology, Princess Margaret Cancer Center, University Health Network, Toronto, ON, Canada

[6]Edward S. Rogers Department of Electrical and Computer Engineering, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON, Canada

[7]North York General Hospital, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Christopher M K L Yao, MD
Department of Otolaryngology - Head and Neck Surgery, Princess Margaret Cancer Center, University Health Network, 610 University Avenue, Toronto, ON, Canada

Abstract

Background: Head and neck cancer (HNC) survivors face profound functional and quality-of-life deficits due to disease- and treatment-related sequelae, ranging from mild fatigue to debilitating dysphagia. Wearable technology, by monitoring biometric data such as step counts or providing swallowing biofeedback, offers a unique method for tracking and monitoring the negative effects of HNC.

Objective: The aim of this study was to explore the current applications of wearable technology in HNC.

Methods: A scoping review was conducted following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. A search strategy was built, and a literature search was performed across 5 databases. The initial search yielded 5256 studies, which underwent a 2-phase screening process: title and abstract review followed by full-text review. Inclusion criteria included peer-reviewed, English-language articles published between January 2002 and April 2024 that used wearable technology in HNC care. After full-text review, 9 studies met the inclusion criteria. Data were manually extracted and synthesized narratively.

Results: The included studies examined 3 main types of wearable devices: radioactivity (2 studies), physical activity (4 studies), and throat physiology monitors (3 studies). Radioactivity monitors detected residual radioactivity and thyroidal radioiodine uptake. They demonstrated potential to reduce radioactivity exposure risk and personalize radiation doses for patients with thyroid cancer. Physical activity monitors tracked step counts, heart rate, and sleep habits. Low step counts were significantly associated with increased anxiety, radiation-related toxicity, hospital admission rates, and feeding tube placement. One study also linked poor sleep patterns to declines in quality of life. Throat physiology monitors measured pharyngeal electromyography data as well as extrinsic laryngeal muscle movements. Throat sensors achieved high accuracy in classifying swallowing events and translating muscle movements into speech. While earliest in the development continuum, they are promising tools for swallowing and vocal rehabilitation therapy. Barriers to wearable adoption included wearable discomfort, technical difficulties, and patient withdrawal due to treatment side effects. As the definition of wearable adherence varied widely, we propose that future studies report wearable adherence as “percentage of prescribed wear time achieved” to facilitate cross-study comparisons.

Conclusions: Wearable technology may enhance treatment monitoring, prognostication, and rehabilitation in head and neck oncology. Radioactivity and physical activity monitors provide actionable insights for clinical decision-making, while throat physiology monitors offer innovative solutions for speech and swallowing therapy. However, challenges such as device adherence, data integration, and patient comfort must be addressed to realize their full potential. Future research should prioritize larger, longitudinal studies, standardized adherence metrics, and consider the integration of artificial intelligence to refine predictive

XSL•FO
RenderX

capabilities. By overcoming these barriers, wearable technology could transform survivorship care, improving functional outcomes and quality of life for patients with HNC.

(JMIR Mhealth Uhealth 2025;13:e72372) doi:10.2196/72372

KEYWORDS

wearable technology; head and neck cancer; digital health; therapeutics; vocal rehabilitation; otorhinolaryngologic diseases

Introduction

Head and neck cancer (HNC) had the eighth highest incidence of any cancer in the world, accounting for over 700,000 new cases globally in 2022 [1]. HNCs arise in the upper aerodigestive tract and can affect patients' ability to interact with the world, including breathing, swallowing, and speaking. Furthermore, treatment toxicities can lead to significant changes to patients' physical and psychosocial quality of life (QOL) [2]. While visits with clinicians may detect recurrence, they are unable to assess the full spectrum of needs for patients with HNC [3,4].

In recent years, wearable technology has emerged as a promising tool to address some of these challenges. These are portable devices equipped with sensor technologies to enable real-time tracking of health data including but not limited to: vital sign monitoring, daily step counts, sleep quality, and more. Wearable technologies have been used for safety monitoring, early and real-time diagnosis, chronic disease optimization, and patient engagement [5]. Within oncology, wearable technology has demonstrated benefit in prognostication, treatment monitoring, and rehabilitation. Examples in the literature include improving physical activity among patients with breast cancer [6], and predicting readmission after peritoneal cancer surgery [7].

Wearable data can improve performance status assessment to guide treatment selection and flag patients at heightened risk of treatment-related toxicities [8-10]. These insights are particularly valuable for HNC survivors, who often experience functional decline driven by fatigue, xerostomia, hoarseness, dysphagia, and shoulder weakness [11,12]. For example, by quantifying early dysphagia events, wearable sensors may prompt timely swallowing therapy or dietary interventions. Similarly, step counters can flag declining physical activity in patients struggling with fatigue, and shoulder-mounted sensors may support physiotherapy in those with accessory nerve weakness. These applications highlight the head and neck as a practical site for wearable technology. Herein, our aim is to explore the current applications of wearable technology in head and neck oncology.

Methods

This scoping review was conducted according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) 2020 guidelines [13]. A health science researcher (AM) conducted a literature search on PubMed, Embase, OVID Medline, CINAHL Plus, and the Cochrane Database of Systematic Reviews for studies published between January 2002 and April 2024, focusing on recent advances in head and neck oncology [14]. A combination of controlled vocabularies and free-text keywords was used (Table S1-S6 in Multimedia Appendix 1). Two reviewers (MS and AM) conducted a snowballing search to find additional citations.

Inclusion criteria consisted of peer-reviewed, English-language journal articles and full-length conference papers that used wearable technology applicable to HNC care. More specifically, we sought to review the use of wearable technology in all aspects of HNC care (ie, prevention, diagnosis, treatment, and surveillance). Wearable studies involving other cancer types were included if they presented data specific to patients with HNC. Reviews and meta-analyses were excluded, but their reference lists were manually searched for key articles.

All title, abstract, and full-text screening was completed using Covidence (Covidence systematic review software; Veritas Health Innovation). Search titles and abstracts were independently screened by 2 reviewers (MS and AM). Full articles were retrieved and independently reviewed. Disagreements were resolved by consensus and in consultation with a third reviewer (CY).

A narrative synthesis was conducted to synthesize the different wearable technologies reported in the HNC literature. All data were charted manually. Data were charted using the following categories: authors, study year, geographic location, journal, study design, HNC type, sample size, participant mean age, participant sex, follow-up duration, HNC treatment, timing of wearable use relative to HNC treatment, wearable type, wearable parameters, study objective, study outcomes, wearable clinical impact, and explicit barriers to wearable use. The barriers to use were based on the listed reasons for participant withdrawal and poor adherence. Data were extracted by 2 reviewers (MS and GS) and independently verified with a third reviewer (CY).

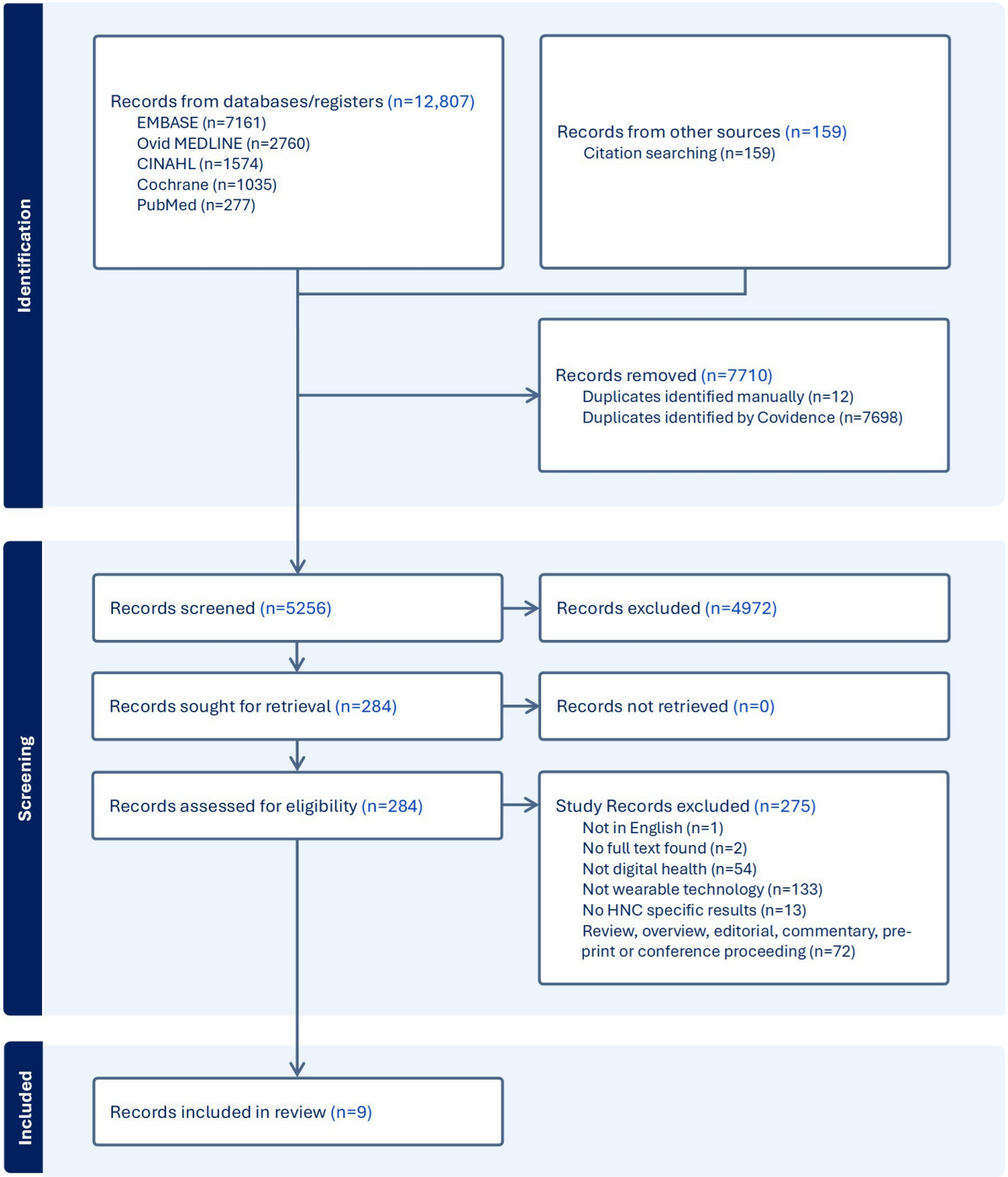
Results

Study Selection

Initial database search yielded 5256 studies (Figure 1). After title and abstract screening, 284 studies met criteria for full-text review. The reduction from an initial pool of 5256 studies to 284 studies is primarily attributable to a broad search strategy which included all forms of digital health across head and neck oncology from 2002 onward (Table S1-S6 in Multimedia Appendix 1). However, the vast majority of studies did not include wearable devices and wearables were not studied in HNC prior to 2018. After a full-text review, 271 studies were excluded. We excluded studies that did not end up using wearable technology (n=185), were reviews (n=72), were not applicable to HNC care (n=13), had no full text available (n=2), and were not reported in the English language (n=1). After exclusion, a total of 9 studies met the inclusion criteria for this review [15-23].

XSL•FO
RenderX

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart depicting study identification and screening. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart illustrating the study selection process detailing databases used as well as studies identified, screened, assessed for eligibility, and included in the final review.



Study Characteristics

Table 1 illustrates the characteristics of the 9 studies published from 2018 to 2024 [15-23]. Of the 9 articles included in this review, there were 3 observational prospective cohort studies [18-20], 3 feasibility studies [16,17,21], 2 validation studies [22,23], and 1 case-control study [15]. Five studies [18-22] examined wearables for various HNC types while 4 studies [15-17,23] examined wearables for specific HNC types. The specific HNC types included well-differentiated thyroid carcinoma, as well as oral cavity, oropharyngeal, and laryngeal cancer.

XSL•FO
RenderX

Table . Wearable study characteristics.

Source	Study design	HNC[a] type	Sample size (n)	Mean age (range)	Sex, n (%)	Follow-up duration	Treatment	Treatment phase
Constantinescu 2018, Canada, Dysphagia [15]	Case-control study	Oral cancer Oropharyngeal cancer	20	57.85 (41-65)	Male: 14 (70) Female: 6 (30)	N/A[b]	Surgery, RT,[c] or CRT[d]	During and post
Gallicchio 2020, Italy, Journal of Translational Engineering in Health and Medicine [16]	Feasibility study	Differentiated thyroid carcinoma	20	41 (33 - 49)	Male: 8 (40) Female: 12 (60)	2.3 days	Radionuclide therapy	During
Santhanam 2021, USA, Nature Scientific Reports [17]	Feasibility study	Papillary and Hurthle cell thyroid carcinoma	2	46 (37 - 55)	Male: 2 (100) Female: 0 (0)	NA	Radionuclide therapy	During and post
Boeke 2022, Germany, Nature Scientific Reports [18]	Observational prospective cohort	Various HNCs	19	69 (68-75)	Male: 12 (63) Female: 7 (37)	46 days	RT or CRT	During
Sher 2022, USA, Technical Innovations and Patient Support in Radiation Oncology [19]	Observational prospective cohort	Various HNCs	43	61 (54 - 69)	Male: 37 (86) Female: 6 (14)	72 days	RT or CRT	During
Ohri 2023, USA, JCO Clinical Care Informatics [20]	Observational prospective cohort	Various HNCs	29	61 (45 - 78)	Male: 24 (83) Female: 5 (17)	40 days	RT or CRT	During
Holländer-Mieritz 2023, Denmark, JCO Clinical Care Informatics [21]	Feasibility study	Various HNCs	10	62 (50 - 75)	Male: 8 (80) Female: 2 (20)	60 days	RT or CRT	During
Xu 2023, China, Cancers [22]	Validation study	Various HNCs	N/A	N/A	N/A	N/A	N/A	N/A
Che 2024, USA, Nature Communications [23]	Validation study	Laryngeal cancer	N/A	N/A	N/A	N/A	N/A	N/A

[a]HNC: head and neck cancer.

[b]Not applicable.

[c]RT: radiotherapy.

[d]CRT: chemoradiotherapy.

The median sample size was 20 with a range of 2 - 43. The overall mean participant age among the studies was 57 (SD 8.58) years. Studies predominantly included male participants with a mean of 74% (SD 17.84). Patients were followed with wearables for a median of 21 days (IQR 5-40). Seven studies [15-21] investigated wearables in a clinical setting while 2 studies [22,23] focused on the development of new wearables. Of the 7 studies [15-21], 5 investigated wearables during patient treatment [16-20] and 2 investigated wearables during treatment and posttreatment [15,21] (Table 1).

Types of Wearable Devices Used in HNC Care

Table 2 summarizes the different types of wearable devices used in HNC care across the 9 studies [15-23]. There were 3 throat physiology monitor studies [15,22,23], 2 radioactivity monitor studies [16,17], and 4 physical activity monitor studies

XSL•FO
RenderX

[18-21]. Two radioactivity monitor studies followed patients during radionuclide therapy [16,17] (Table 1). One study aimed to detect residual radioactivity [16], whereas the second study aimed to detect thyroidal radioiodine uptake (Table 2) [17].

All 4 physical activity monitor studies were implemented around patients' radiotherapy or chemoradiotherapy treatment courses [18-21] (Table 1). There have been no studies around the time of surgery. These studies monitored step counts, with 2 tracking heart rate and 1 tracking sleeping habits (Table 2) [18-21].

Three studies assessed externally applied throat physiology monitors. Xu et al [22] developed a general throat sensor that incorporates a 3-axis accelerometer and surface electromyography data. Che et al [23] developed a throat sensor that measures extrinsic laryngeal muscle activity to predict intended speech through machine learning. Finally, Constantinescu et al [15] studied the performance of a throat physiology monitor that detects patient swallowing using submental surface electromyography data to track home-based swallowing therapy (Table 2).

XSL•FO
RenderX

Table . Wearable technology parameters, outcomes, and response rate.

Source	Wearable technology	Parameters measured	Purpose	Outcomes	Clinical impact	Barriers to use
Constantinescu 2018 [15]	Throat physiology monitor	Submental sEMG[a]	Remote detection of swallowing for home-based swallowing therapy	92.7% sensitive for swallowing detection in patients with HNC[b]; Swallowing duration, frequency, and power were significantly lower for patients with HNC	Automated feedback during home-based swallowing therapy	Surgical reconstruction of submental area; History of stroke
Gallicchio 2020 [16]	Radioactivity monitor	Residual radioactivity	Detect residual radioactivity	Remote monitor correlates with home monitor (r=0.96)[c]	Operational efficiency and reduced exposure risk for medical staff	Not specified
Santhanam 2021 [17]	Radioactivity monitor	Thyroid radioiodine uptake	Detect thyroidal radioiodine uptake	Remote monitor correlates with gold-standard gamma probes (r=0.87)[c]	Potential to increase the accuracy of radioiodine therapy dosing	Not specified
Boeke 2022 [18]	Physical activity monitor	Step count	Relationship between daily step counts with QOL[d] and treatment-related toxicities	Step count decline during RT[e] and cetuximab treatment[c]; Step counts increase post-RT[c]; Step count decline correlated with QOL decline	Low step count may predict QOL and treatment-related toxicities	Not specified
Sher 2022 [19]	Physical activity monitor	Step count, heart rate, sleep habits	Relationship between step counts, heart rate, and sleep with QOL, hospital admissions, and change in pain medication	Step count decline precedes hospital admission and QOL decline by 1 week[c]; Less sleep precedes QOL decline by 1 week[c]	Step count and sleep predict hospital admission and QOL decline by 1 week	Severe treatment-related side effects Wearable discomfort
Ohri 2023 [20]	Physical activity monitor	Step count	Relationship between daily step counts with hospitalization and PROs[f]	Low step counts were associated with increased anxiety and sadness[c]; High average daily step counts were associated with a reduced risk of hospitalization and reduced risk of feeding tube placement[c]	Low step count predicts patient mood and risk of hospitalization	Severe treatment-related side effects Wearable discomfort Permanent data loss
Holländer-Mieritz 2023 [21]	Physical activity monitor	Step count, heart rate	Describe the variations in step count and heart rate during a RT/CRT[g] course	One patient had declining heart rate correlated with decreased step count; The second patient had stable heart rates and step counts	Conclusions were limited by poor adherence	Patient technical difficulties Demanding to charge the smartwatch

XSL•FO
RenderX

Source	Wearable technology	Parameters measured	Purpose	Outcomes	Clinical impact	Barriers to use
Xu 2023 [22]	Throat physiology monitor	3-axis accelerometer sEMG	Remote diagnosis and treatment evaluation	92% accuracy in predicting coughing, drinking, swallowing, and talking states	Swallowing therapy and postsurgical rehabilitation	NA
Che 2024 [23]	Throat physiology monitor	Laryngeal muscle movements	Speech for patients with dysfunctional or absent vocal cords	94.68% accuracy in translating laryngeal muscle movements into voice signals; No change in wearable temperature or sound pressure level after 40 minutes use	Voice restoration for patients with dysfunctional or absent vocal cords	NA

[a]sEMG: surface electromyography.

[b]HNC: head and neck cancer.

[c]statistical significance.

[d]QOL: quality of life.

[e]RT: radiotherapy.

[f]PRO: patient-reported outcome.

[g]CRT: chemoradiotherapy.

Effectiveness of Wearable Devices in HNC Care

Wearable radioactivity monitors correlated well with conventional measurements of residual radioactivity and thyroid radioactivity reuptake [16,17]. In the study by Gallicchio et al [16], a wearable radioactivity monitor measured residual radioactivity following radioactive iodine treatment in patients with thyroid cancer. Compared to a conventional installed room device, their wearable radioactivity monitor was found to significantly reduce staff radioactivity exposure levels and promote safe discharge home for patients. In the study by Santhanam et al [17], a collarbone radioactivity monitor provided more detailed information on thyroid radioiodine uptake compared to standard gamma probes, including higher sampling frequency, fractional uptake distribution over time, and differences in uptake lateralization.

Physical activity monitors provided evidence for tracking step counts. Across 3 studies, low step counts were associated with several clinical outcomes [18-20]. In the study by Boeke et al [18], step counts decreased with increasing radiotherapy toxicity in all but 1 patient. Sher et al [19] identified that step count decline was significantly associated with a reduction in QOL and hospital admission 1 week later in 25 patients actively undergoing radiotherapy. Moreover, this was the only study in which the wearable device monitored sleep patterns using movement and heart rate tracking, whereby fewer total minutes of recorded sleep were associated with lower QOL scores. In a third study by Ohri et al [20], high daily step count was significantly associated with a reduced risk of feeding tube placement and hospitalization, while low step counts were significantly associated with anxiety and depression.

Xu et al [22] and Che et al [23] both developed stretchable external throat sensors. The wearable by Xu et al [22] could measure respiratory rate and heart rate as well as predict the patient's state (vocalization, coughing, drinking, swallowing, or talking) using a machine learning algorithm with a limited set of classified outcomes. The external throat sensor investigated by Che et al [23] offers a novel method for laryngeal speech generation by detecting extralaryngeal muscle activity and using a machine learning algorithm to predict intended speech. Collectively, while originally developed to classify throat movement and predict speech, these throat sensor platforms hold significant clinical promise. Potential applications include supporting swallowing therapy, evaluating aspiration risk, assessing rehabilitation adherence, quantifying aphonia burden, and facilitating voice restoration.

Constantinescu et al [15] used a throat sensor to predict patient swallowing events for remote swallowing therapy using a classification algorithm. The device achieved an acceptable sensitivity and positive predictive value for swallowing detection (Table 2). Moreover, a portable swallowing detector can provide patients with live biofeedback and physicians with real-time updates, acting as a powerful adjunctive tool for monitoring at-home swallowing therapy [15].

Barriers to Use of Wearable Devices in HNC Care

Among the 9 studies [15-23], 4 directly assessed wearable feasibility and adherence [18-21]. Boeke et al [18] identified a higher rate of adherence at 92%, whereas the remaining 3 studies did not reach predetermined feasibility targets (80% - 90%) for adherence [19,20], or found a low rate of adherence (31%) [21]. However, the definition of adherence varied.

Notably, Sher et al [19] was the only study to examine wearable adherence during sleep. They found that patients wore wearables 60.5% of all potential sleep time, and only 1 patient was 100% adherent during sleep.

XSL•FO
RenderX

The common themes for participant study withdrawal were wearable discomfort, patient technical difficulties, and noncompliance due to treatment-related side effects (Table 2). Patient technical difficulties and reasons for wearable discomfort were not elaborated further in any study. With regard to treatment-related side effects, Sher et al [19] found that malnutrition and failure to thrive moderately correlated with participant study withdrawal.

Discussion

Principal Findings

Wearable technologies have emerged as an effective tool for health care delivery. In oncology, these technologies have been geared toward the phases of prognostication, treatment monitoring, and rehabilitation planning [24]. HNC treatment can have a profound impact on a patient's physical function, receptive and expressive communication, swallowing, and overall QOL. As such, there are many existing and novel applications of wearable technology in this patient population. Herein, we identified 9 studies that explored 3 areas of wearable technology within HNC: radioactivity, physical activity, and throat physiology monitors [15-23]. Their clinical applications include improving radionuclide therapy delivery, prognostication of clinical outcomes, as well as speech and swallowing rehabilitation.

In this review, wearable devices allowed for accurate assessment of residual radioactivity and fractional uptake over time and localization after radioactive iodine treatment [16,17]. Step counts and physical activity monitoring were found to be associated with chemoradiotherapy-related toxicity, reduced QOL, mood, increased rates of hospital admission, and feeding tube placement [18-20]. Similarly, wearable device studies in other disease sites have also identified step counts to be associated with frailty, increased hospital readmission rates, treatment-related toxicity, prolonged length of stay in hospital, and death [25-28]. Regarding wearable adherence, participant adherence ranged from 31% to 92% [18-21]. However, these studies differ in their definitions of adherence and compliance between studies. This highlights the need for establishing standard definitions for wear time, as suggested by Beauchamp et al [9] and Huang et al [29]. We recommend that future wearable studies define and report adherence as "percentage of prescribed wear time achieved," as shown in the equation below.

Throat physiology monitors, while earlier in the development continuum than other wearable monitors, represent an example where novel sensor technologies allow for unique wearable technology applications in HNC. For one, muscle activity sensors applied externally to the neck can measure extrinsic laryngeal muscle activity, which is then processed by a pretrained machine learning algorithm to predict anticipated words, enabling laryngeal speech [23]. While the study itself had a limited set of trained outputs, it establishes the groundwork for a novel voice rehabilitation approach. Similarly, Constantinescu et al [15] developed a wireless external throat sensor that rests on the submental area and provides patients with biofeedback during swallowing therapy exercises. While these technologies offer significant promise, further studies demonstrating clinical utility and feasibility are needed. In particular, throat sensor wearable adherence has not been studied. This represents an opportunity for future study, especially by assessing whether device type influences wearable adherence. Ongoing development of these technologies that focus on scalability and cost-minimization would further the potential benefit to patients with HNC worldwide.

$$\text{Adherence \%} = \frac{\text{actual wear time (h)}}{\text{prescribed wear time (h)}} \times 100\%$$

Barriers

Barriers to integrating wearable technology in patients with HNC can be framed as technical factors, patient-related factors, and provider-related factors. Technical factors include the resources needed to troubleshoot devices, short battery life, and poor data management [20,21]. Patient-related factors include an initial device learning curve, wearable discomfort, and withdrawal from use due to severe chemotherapy- or radiotherapy-related side effects [19,20]. While provider-related factors were not explicitly discussed in the studies we identified, it is known that implementation from health care practitioners and stakeholders is often key to introducing these novel technologies [30].

An important component of provider-driven implementation is the integration of wearable data with existing electronic medical records. A 2024 wearable feasibility study involving 20 cancer patients demonstrated that Apple Watch data from 95% of participants was successfully transmitted to an electronic medical record and reviewed by an oncologist [31]. By automating data transfer, the undue burden to physicians and patients was reduced. This proof of concept lays the groundwork for scalable, low-burden integration of wearable data into oncology practice.

Data Security and Privacy

Management of wearable device health data storage, device to database information transfer, and privacy varied between studies when reported. Like other personal health information, health data from wearable devices should be held to the same standards of confidentiality. Five studies used commercial devices [16,18-21]. Of these, 2 used in-house software and direct data transfer [16,18], 2 used commercial application programming interfaces with personal or study-provided smartphones for data storage and subsequent data syncing [19,20], and 1 study used an independent industry application programming interface with secure cloud servers for storage and data transfer [21]. Remaining studies used in-house developed devices and software.

While the technical methods of data storage and transfer vary widely, a consistent ethical standard must be applied. Two foundational principles emerge: informed consent and the right to withdraw. Patients should be fully informed about the nature of data collection, how and where their information is stored, and with whom it may be shared. Furthermore, patients must retain the ability to withdraw their data without penalty. These ethical obligations are particularly pressing in studies using third-party commercial devices, where data ownership and access are often unclear.

XSL•FO
RenderX

A cross-sectional survey of commercial wearable device users identified that over 50% of patients understood that wearable device health information is confidential in nature; however, their understanding of personal data handling and the expected privacy standards was lacking [32]. In this review, many of these wearables used third-party commercial devices and software. As wearable devices are implemented within research and clinical settings, the importance of educating patients on wearable device health data privacy and ensuring secure data storage and transfer must be clarified.

Future Directions

A notable gap in the HNC wearable literature is the investigation of physical activity monitors during surgery. By contrast, physical activity monitoring in colorectal cancer has been studied in the preoperative, perioperative, and postoperative phases [33]. Surgery represents the main treatment for many patients with HNC, with surgery often being lengthy, complicated, and requiring a prolonged period of recovery. In turn, this represents a unique opportunity whereby physical activity monitors may aid in surgical risk assessment, acute postoperative monitoring, and beyond [34].

Another critical limitation is the small sample sizes across the included studies, which ranged from 2 to 43 participants. As a result, the current findings should be interpreted as hypothesis-generating rather than sufficient to inform clinical decision-making. Large randomized controlled trials are needed to validate early findings and establish the clinical utility of wearables in HNC care.

A third gap in the current literature is the lack of longitudinal studies. Most existing studies were limited to the acute radiotherapy phase, with follow-up periods under 3 months (Table 1). Longitudinal studies could help establish predictive thresholds for late-onset dysphagia and hospital readmission while also tracking rehabilitation over time. These studies can help incorporate wearables into long-term survivorship care.

The integration of artificial intelligence (AI) with wearable technology data has significant promise [35]. Recent studies demonstrate that machine learning algorithms can leverage wearable technology to train risk stratification models [36], monitor tolerance and response to chemotherapeutics [37], detect biomarkers [38], as well as adhere to treatment regimens [39]. In this review, 2 studies integrated AI with external laryngeal sensor data to classify speech and swallowing states to facilitate laryngeal rehabilitation [22,23]. AI is likely to further enhance the capabilities offered by wearable technologies.

Study Limitations

Our scoping review identified 9 studies which reflected various applications including physical activity monitors, radioactivity monitoring, and novel throat sensors [15-23]. While their inherent differences limit our ability for further analyses, this limitation reflects the current state of research in wearable technology for HNC and highlights the need for more studies in this field. Second, our review only included studies published in English, which could have led to the exclusion of relevant studies published in other languages.

Conclusion

This scoping review provides the first comprehensive overview of wearable technology applications in HNC. The analysis of 9 studies reveals a diverse range of wearable applications and identifies potential future directions [15-23]. These technologies demonstrate potential benefit for risk stratification, treatment monitoring, and augmenting rehabilitation for patients with HNC. With continued research in wearable technologies, care must also be taken to define measures of adherence, understand clinical integration barriers, and establish standards for patient health data and privacy with wearable devices in research and clinical settings.

Acknowledgments

The authors have no acknowledgments to declare. During the preparation of this work, the author(s) used ChatGPT 4.0 to proofread the manuscript for grammatical errors. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

MS-T and GS contributed to the conceptualization, data curation, formal analysis, investigation, visualization, and writing of the original draft, as well as reviewing and editing the manuscript. MX contributed to data curation and writing of the original draft and reviewed and edited the manuscript. AM was involved in methodology and investigation and contributed to the original draft and manuscript review and editing. AE contributed to formal analysis, software development, and manuscript review and editing. QP, JAC, PW, CJT, DPG, JRA, and ES were responsible for data curation and contributed to reviewing and editing the manuscript. CMKLY contributed to conceptualization, methodology, resource provision, supervision, and writing of the original draft, as well as reviewing and editing the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 22 KB - mhealth_v13i1e72372_app1.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 16 KB - mhealth_v13i1e72372_app2.docx]

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263. [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Taylor KJ, Amdal CD, Bjordal K, et al. Serious long-term effects of head and neck cancer from the survivors' point of view. Healthcare (Basel) 2023 Mar 21;11(6):906. [doi: 10.3390/healthcare11060906] [Medline: 36981562]
3. Paterson C, Toohey K, Bacon R, Kavanagh PS, Roberts C. What are the unmet supportive care needs of people affected by cancer: an umbrella systematic review. Semin Oncol Nurs 2023 Jun;39(3):151353. [doi: 10.1016/j.soncn.2022.151353] [Medline: 36435657]
4. Hart NH, Crawford-Williams F, Crichton M, et al. Unmet supportive care needs of people with advanced cancer and their caregivers: a systematic scoping review. Crit Rev Oncol Hematol 2022 Aug;176:103728. [doi: 10.1016/j.critrevonc.2022.103728] [Medline: 35662585]
5. Lu L, Zhang J, Xie Y, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 9;8(11):e18907. [doi: 10.2196/18907] [Medline: 33164904]
6. Teo NR, Siew LED, Ang WHD, Lau Y. Wearable-technology-assisted interventions for breast-cancer survivors: a meta-analysis and meta-regression. Semin Oncol Nurs 2023 Jun;39(3):151403. [doi: 10.1016/j.soncn.2023.151403] [Medline: 36894449]
7. Wolk S, Linke S, Bogner A, et al. Use of activity tracking in major visceral surgery-the enhanced perioperative mobilization trial: a randomized controlled trial. J Gastrointest Surg 2019 Jun;23(6):1218-1226. [doi: 10.1007/s11605-018-3998-0] [Medline: 30298422]
8. Beg MS, Gupta A, Stewart T, Rethorst CD. Promise of wearable physical activity monitors in oncology practice. J Oncol Pract 2017 Feb;13(2):82-89. [doi: 10.1200/JOP.2016.016857] [Medline: 28387544]
9. Beauchamp UL, Pappot H, Holländer-Mieritz C. The use of wearables in clinical trials during cancer treatment: systematic review. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e22006. [doi: 10.2196/22006] [Medline: 33174852]
10. Izmailova E, Huang C, Cantor M, Ellis R, Ohri N. Daily step counts to predict hospitalizations during concurrent chemoradiotherapy for solid tumors. J Clin Oncol 2019 Sep 20;37(27_suppl):293-293. [doi: 10.1200/JCO.2019.37.27_suppl.293]
11. Hunter KU, Jolly S. Clinical review of physical activity and functional considerations in head and neck cancer patients. Support Care Cancer 2013 May;21(5):1475-1479. [doi: 10.1007/s00520-013-1736-4] [Medline: 23417564]
12. Midgley AW, Lowe D, Levy AR, Mepani V, Rogers SN. Exercise program design considerations for head and neck cancer survivors. Eur Arch Otorhinolaryngol 2018 Jan;275(1):169-179. [doi: 10.1007/s00405-017-4760-z] [Medline: 29058083]
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
14. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 5;21(9):e14017. [doi: 10.2196/14017] [Medline: 31489843]
15. Constantinescu G, Kuffel K, Aalto D, Hodgetts W, Rieger J. Evaluation of an automated Swallow-Detection Algorithm using visual biofeedback in healthy adults and head and neck cancer survivors. Dysphagia 2018 Jun;33(3):345-357. [doi: 10.1007/s00455-017-9859-2] [Medline: 29098398]
16. Gallicchio R, Scapicchio D, Nardelli A, et al. Assessment of residual radioactivity by a comprehensive wireless, wearable device in thyroid cancer patients undergoing radionuclide therapy and comparison with the results of a home device: a feasibility study. IEEE J Transl Eng Health Med 2021;9:2700306. [doi: 10.1109/JTEHM.2020.3042118] [Medline: 33329944]
17. Santhanam P, Solnes L, Nath T, et al. Real-time quantitation of thyroidal radioiodine uptake in thyroid disease with monitoring by a collar detection device. Sci Rep 2021 Sep 16;11(1):18479. [doi: 10.1038/s41598-021-97408-y] [Medline: 34531443]

XSL•FO
RenderX

18. Boeke S, Hauth F, Fischer SG, et al. Acceptance of physical activity monitoring in cancer patients during radiotherapy, the GIROfit phase 2 pilot trial. Tech Innov Patient Support Radiat Oncol 2022 Jun;22:16-21. [doi: 10.1016/j.tipsro.2022.03.004] [Medline: 35402741]
19. Sher DJ, Radpour S, Shah JL, et al. Pilot study of a wearable activity monitor during head and neck radiotherapy to predict clinical outcomes. JCO Clin Cancer Inform 2022 Mar;6:e2100179. [doi: 10.1200/CCI.21.00179] [Medline: 35333597]
20. Ohri N, Bar-Ad V, Fernandez C, et al. Remote activity monitoring and electronic patient-reported outcomes collection during radiotherapy for head and neck cancer: a pilot study. JCO Clin Cancer Inform 2023 Apr;7:e2200132. [doi: 10.1200/CCI.22.00132] [Medline: 37071027]
21. Holländer-Mieritz C, Steen-Olsen EB, Kristensen CA, Johansen C, Vogelius IR, Pappot H. Feasibility of using wearables for home monitoring during radiotherapy for head and neck cancer-results from the OncoWatch 1.0 Study. Cancers (Basel) 2023 Jan 9;15(2):422. [doi: 10.3390/cancers15020422] [Medline: 36672370]
22. Xu H, Zheng W, Zhang Y, et al. A fully integrated, standalone stretchable device platform with in-sensor adaptive machine learning for rehabilitation. Nat Commun 2023 Nov 27;14(1):7769. [doi: 10.1038/s41467-023-43664-7] [Medline: 38012169]
23. Che Z, Wan X, Xu J, Duan C, Zheng T, Chen J. Speaking without vocal folds using a machine-learning-assisted wearable sensing-actuation system. Nat Commun 2024 Mar 12;15(1):1873. [doi: 10.1038/s41467-024-45915-7] [Medline: 38472193]
24. Chow R, Drkulec H, Im JHB, et al. The use of wearable devices in oncology patients: a systematic review. Oncologist 2024 Apr 4;29(4):e419-e430. [doi: 10.1093/oncolo/oyad305] [Medline: 37971410]
25. Daskivich TJ, Houman J, Lopez M, et al. Association of wearable activity monitors with assessment of daily ambulation and length of stay among patients undergoing major surgery. JAMA Netw Open 2019 Feb 1;2(2):e187673. [doi: 10.1001/jamanetworkopen.2018.7673] [Medline: 30707226]
26. Low CA, Bovbjerg DH, Ahrendt S, et al. Fitbit step counts during inpatient recovery from cancer surgery as a predictor of readmission. Ann Behav Med 2018 Jan 5;52(1):88-92. [doi: 10.1093/abm/kax022] [Medline: 29538623]
27. Lin FP, Bloomer PM, Grubbs RK, et al. Low daily step count is associated with a high risk of hospital admission and death in community-dwelling patients with cirrhosis. Clin Gastroenterol Hepatol 2022 Aug;20(8):1813-1820. [doi: 10.1016/j.cgh.2022.03.012] [Medline: 35331941]
28. Rice H, Hill K, Fowler R, Watson C, Waterer G, Harrold M. Reduced step count and clinical frailty in hospitalized adults with community-acquired pneumonia. Respir Care 2020 Apr;65(4):455-463. [doi: 10.4187/respcare.06992] [Medline: 31575707]
29. Huang Y, Upadhyay U, Dhar E, Kuo LJ, Syed-Abdul S. A scoping review to assess adherence to and clinical outcomes of wearable devices in the cancer population. Cancers (Basel) 2022 Sep 13;14(18):4437. [doi: 10.3390/cancers14184437] [Medline: 36139602]
30. Dzimitrowicz HE, Blakely LJ, Jones LW, LeBlanc TW. Bridging new technology into clinical practice with mobile apps, electronic patient-reported outcomes, and wearables. Am Soc Clin Oncol Educ Book 2022 Apr;42:1-6. [doi: 10.1200/EDBK_350550] [Medline: 35522912]
31. Weng JK, Virk R, Kaiser K, et al. Automated, real-time integration of biometric data from wearable devices with electronic medical records: a feasibility study. JCO Clin Cancer Inform 2024 Sep;8:e2400040. [doi: 10.1200/CCI.24.00040] [Medline: 39348612]
32. Cilliers L. Wearable devices in healthcare: privacy and information security issues. Health Inf Manag 2020;49(2-3):150-156. [doi: 10.1177/1833358319851684] [Medline: 31146589]
33. Kavallieros K, Karakozis L, Hayward R, Giannas E, Selvaggi L, Kontovounisios C. Wearable devices in colorectal surgery: a scoping review. Cancers (Basel) 2024 Jun 22;16(13):2303. [doi: 10.3390/cancers16132303] [Medline: 39001367]
34. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 2019 Mar;69(2):127-157. [doi: 10.3322/caac.21552] [Medline: 30720861]
35. Birla M, Rajan R, Roy PG, Gupta I, Malik PS. Integrating artificial intelligence-driven wearable technology in oncology decision-making: a narrative review. Oncology (Williston Park, NY) 2025;103(1):69-82. [doi: 10.1159/000540494] [Medline: 39072365]
36. Torrente M, Sousa PA, Hernández R, et al. An artificial intelligence-based tool for data analysis and prognosis in cancer patients: results from the clarify study. Cancers (Basel) 2022 Aug 22;14(16):4041. [doi: 10.3390/cancers14164041] [Medline: 36011034]
37. Cay G, Sada YH, Dehghan Rouzi M, et al. Harnessing physical activity monitoring and digital biomarkers of frailty from pendant based wearables to predict chemotherapy resilience in veterans with cancer. Sci Rep 2024 Jan 31;14(1):2612. [doi: 10.1038/s41598-024-53025-z] [Medline: 38297103]
38. Tu J, Min J, Song Y, et al. A wireless patch for the monitoring of C-reactive protein in sweat. Nat Biomed Eng 2023 Oct;7(10):1293-1306. [doi: 10.1038/s41551-023-01059-5] [Medline: 37349389]
39. Weeks WA, Dua A, Hutchison J, et al. A low-power, low-cost ingestible and wearable sensing platform to measure medication adherence and physiological signals. Presented at: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 18-21, 2018; Honolulu, HI p. 5549-5553. [doi: 10.1109/EMBC.2018.8513593]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
HNC: head and neck cancer
PRIMSA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
QOL: quality of life

Edited by L Buis; submitted 08.02.25; peer-reviewed by OS Igunma, T Awofala, Y Oyende; revised version received 23.06.25; accepted 21.07.25; published 10.10.25.

Please cite as:
Shammas-Toma M, Sampieri G, Xie M, Maxwell A, Esemezie A, Pham Q, Cafazzo JA, Wong P, Tsai CJ, Goldstein DP, de Almeida JR, Sejdic E, Yao CMKL
Wearable Technologies in Head and Neck Oncology: Scoping Review
JMIR Mhealth Uhealth 2025;13:e72372
URL: https://mhealth.jmir.org/2025/1/e72372
doi:10.2196/72372

©Matthew Shammas-Toma, Gianluca Sampieri, Michael Xie, Aishwaria Maxwell, Alex Esemezie, Quynh Pham, Joseph A Cafazzo, Philip Wong, C Jillian Tsai, David P Goldstein, John R de Almeida, Ervin Sejdic, Christopher M K L Yao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Interventions to Reduce Cancer-Related Fatigue Among Adolescents and Young Adults: Scoping Review

Shanshan Jiang[1]; Xiaoyu Yang[2]; Xinying Yu[1]

[1]Department of Pediatrics, Shengjing Hospital of China Medical University, 36 Sanhao Street, Heping District, Shenyang, China

[2]Department of Oncology, Shengjing Hospital of China Medical University, Shenyang, China

Corresponding Author:
Xinying Yu
Department of Pediatrics, Shengjing Hospital of China Medical University, 36 Sanhao Street, Heping District, Shenyang, China

Abstract

Background: Cancer-related fatigue is a common and significant symptom experienced by patients with cancer and survivors across all age groups, profoundly impacting their quality of life. Adolescents and young adults often encounter substantial academic, career, and personal demands, which pose unique challenges in managing this symptom and may have a more profound overall impact on their lives. While digital health interventions show considerable promise in managing cancer-related fatigue, few reviews have specifically addressed their use among adolescents and young adults.

Objective: This scoping review aimed to identify and assess the types and effectiveness of digital health interventions in managing cancer-related fatigue among adolescents and young adults.

Methods: A comprehensive literature search was conducted using the keywords “digital health,” “adolescent,” “young adult,” “fatigue,” and “neoplasms” across 6 databases: PubMed, CINAHL, PsycINFO, Embase, Cochrane Library, and Web of Science. The search included English-language publications from the inception of each database to August 2024. Two researchers independently screened the studies based on predetermined inclusion and exclusion criteria.

Results: A total of 2965 articles were retrieved during the initial search, of which 10 (0.34%) satisfied the inclusion criteria of this review. The 10 included studies comprised 5 (50%) randomized controlled trials, 2 (20%) quasi-experimental studies, 2 (20%) mixed methods studies, and 1 (10%) cohort study. On the basis of the functions and forms of digital health interventions, the interventions included in this review were categorized into the following 5 types: dynamic health monitoring and feedback, automated online guidance and feedback, live remote coaching and instruction, gamified interventions, and robot-assisted interventions. Multiple studies (7/10, 70%) demonstrated that digital health interventions are effective in reducing cancer-related fatigue in adolescents and young adults and show potential in improving physical function and emotional well-being in this population.

Conclusions: Digital health interventions overcome the time and spatial limitations of traditional treatments and provide holistic support across physical, psychological, and social domains. They hold significant potential to alleviate cancer-related fatigue in adolescents and young adults. Future research should integrate various fatigue measurement scales and conduct large-scale studies and long-term follow-ups to capture a more comprehensive range of fatigue experiences, validate these findings, and enhance the effectiveness of digital health interventions.

(JMIR Mhealth Uhealth 2025;13:e68834) doi:10.2196/68834

KEYWORDS

digital health; adolescent; young adult; cancer-related fatigue; scoping review

Introduction

Background

Cancer-related fatigue (CRF) is a widespread symptom that significantly diminishes quality of life. It is characterized by persistent physical, emotional, and cognitive fatigue that is not proportional to recent activities, profoundly impacting daily functioning [1]. To accurately evaluate this complex symptom, various scales have been developed in the literature, including unidimensional single-item scales, unidimensional multi-item scales, and multidimensional scales specifically designed for patients with cancer. Among these, the most widely used and extensively validated scales are the Functional Assessment of Cancer Therapy–Fatigue, the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (fatigue subscale), and the Fatigue Questionnaire [2,3]. Epidemiological studies indicate that approximately 65% of patients with cancer are affected by CRF, with more than two-thirds of patients reporting that fatigue persists for at least 6 months and is moderate to severe. One-third of patients report that fatigue continues for several years after treatment [4].

XSL•FO
RenderX

Specifically, the prevalence of CRF is highest in patients undergoing cancer treatment before discharge and decreases within 6 months following the completion of acute treatment [5], and the incidence of fatigue is higher in patients with leukemia and lymphoma than in those with central nervous system tumors and solid tumors [6]. CRF is prevalent among patients with cancer and survivors across all age groups, but its severity differs among age groups [7]. Adolescents and young adults, defined as individuals aged 15 to 39 years, face the highest CRF severity [8]. They must cope with not only the physical and psychological demands of cancer treatment but also the pressures associated with major life stages, such as academic responsibilities, career initiation, social connections, and family dynamics [9-11]. These compounded pressures heighten both mental and physical stress, worsening CRF symptoms. The management of CRF primarily depends on nonpharmacological approaches, including modifications to activity levels, lifestyle improvements, and the promotion of mental health [12]. This process often requires comprehensive support, including treatment adherence, health behavior modification, self-management skills, physician-patient communication, and support from family and social networks. CRF requires long-term management. However, traditional CRF management is constrained by factors such as the availability of health care professionals, time, financial resources, environments, and facilities [13]. Furthermore, adolescents and young adults frequently lack sufficient understanding of their cancer treatment history and the risks associated with late effects. This, in combination with their busy schedules, often results in the neglect of CRF treatment [14]. Thus, there is an urgent need for effective and sustainable interventions for adolescents and young adults to assist them in effectively managing CRF.

As the application and popularity of digital health interventions (DHIs) continue to grow, numerous studies have demonstrated that DHIs have potential in supporting the symptom management of patients with cancer [15,16]. *Digital health technologies* refer to the application of digital technology to the health sector, with examples including telemedicine; robotics; and mobile health supported by smartphones, wearable devices, mobile apps, and various monitoring technologies [17]. Studies have shown that the application of these technologies combined with interventions in CRF management is of great importance. Telemedicine services allow patients to consult with physicians via video, phone, or other digital platforms, obtaining personalized guidance on fatigue management [18]. This reduces the time and financial burden associated with seeking medical care and enhances patient accessibility. Robotics technology can offer precise movement guidance and support throughout the patient's rehabilitation, assisting patients in more effectively strengthening muscles and restoring mobility, thereby helping alleviate CRF symptoms [19]. Mobile health technologies, using various mobile devices and apps, allow users to continuously monitor their health status in real time—such as heart rate, activity levels, and sleep quality—and offer personalized health recommendations based on data analysis, thereby assisting users in achieving their health goals and alleviating CRF [20]. For patients experiencing mild fatigue symptoms, self-monitoring and adjustment of health behaviors using DHIs can effectively alleviate fatigue. For those with more severe symptoms, DHIs can offer more comprehensive personalized interventions and support, aiding in alleviating CRF symptoms.

Objectives

Overall, DHIs can provide personalized, real-time, and flexible CRF management, which may more effectively address the challenges faced by adolescents and young adults in managing CRF, thereby enhancing treatment efficacy. However, the development of DHIs targeting CRF remains in its infancy, especially compared to those designed for other cancer-related conditions. In addition, some existing studies include adolescents and young adults and adults in mixed samples, and this may increase data heterogeneity [21]. Therefore, it is essential to synthesize the current evidence to assess the effectiveness of DHIs. A scoping review is especially useful for synthesizing evidence in emerging research fields. When the research question is unclear and specific hypotheses cannot be addressed through traditional systematic reviews, the scoping review method helps organize key concepts in the relevant research field, identify gaps in the existing literature, and systematically integrate emerging evidence [22]. To our knowledge, there have been no high-quality systematic or scoping reviews specifically focused on DHIs for CRF management in adolescents and young adults. This scoping review aims to summarize the current evidence on the types of DHIs for CRF management in adolescent and young adult patients and their effectiveness. In this context, this study holds dual significance. First, systematic reviews on the effectiveness of DHIs in CRF management within the adolescent and young adult group are insufficient, and this study aims to evaluate the specific effect of DHIs in CRF management within this group. Second, in light of the rapid advancements in digital health technologies, the adoption of a scoping review methodology effectively reduces the knowledge translation cycle, offering timely, evidence-based support to researchers and assisting clinicians in formulating empirically grounded CRF intervention strategies.

Methods

We followed the 5-stage scoping review framework by Arksey and O'Malley [23] and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines (Checklist 1) [24].

Search Strategy

Researchers SJ and X Yang conducted a comprehensive search across 6 databases—PubMed, CINAHL, PsycINFO, Embase, Cochrane Library, and Web of Science—as well as a manual search of relevant references. The search spanned each database from its inception to August 2024 and was restricted to studies published in English. The search strategy involved a combination of MeSH (Medical Subject Heading) terms and keywords, including but not limited to ("digital health" OR "telemedicine") AND ("adolescent" OR "young adult") AND ("fatigue" OR "tired") AND ("neoplasms" OR "carcinoma"). Details of the search strategy are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Inclusion and Exclusion Criteria

Two researchers independently screened studies based on predefined inclusion and exclusion criteria. The inclusion criteria were as follows: (1) participants were patients with cancer or survivors aged 15 to 39 years, (2) the study used digital health technology as an intervention, (3) the study included fatigue-related outcomes, (4) the study was published in English, and (5) the full text was available. The exclusion criteria were reviews, protocols without results, case studies, reports, letters, conference proceedings, and abstract-only articles.

Study Selection and Data Extraction

All relevant studies were imported into EndNote X9 (reference management software; Clarivate Analytics) to remove duplicates. Two researchers independently conducted an initial screening of titles and abstracts to evaluate whether they met the predetermined inclusion and exclusion criteria. Following this, they conducted a detailed full-text review to confirm each study's eligibility. Any disagreements about study inclusion were resolved through consultation with a third researcher. The 2 authors independently extracted data from the included studies. They entered the data into a predesigned data extraction form that captured the following information: author, publication year, country, study design, participant ages, sample size, and cancer type. Any discrepancies were resolved through discussion or the involvement of a third researcher.

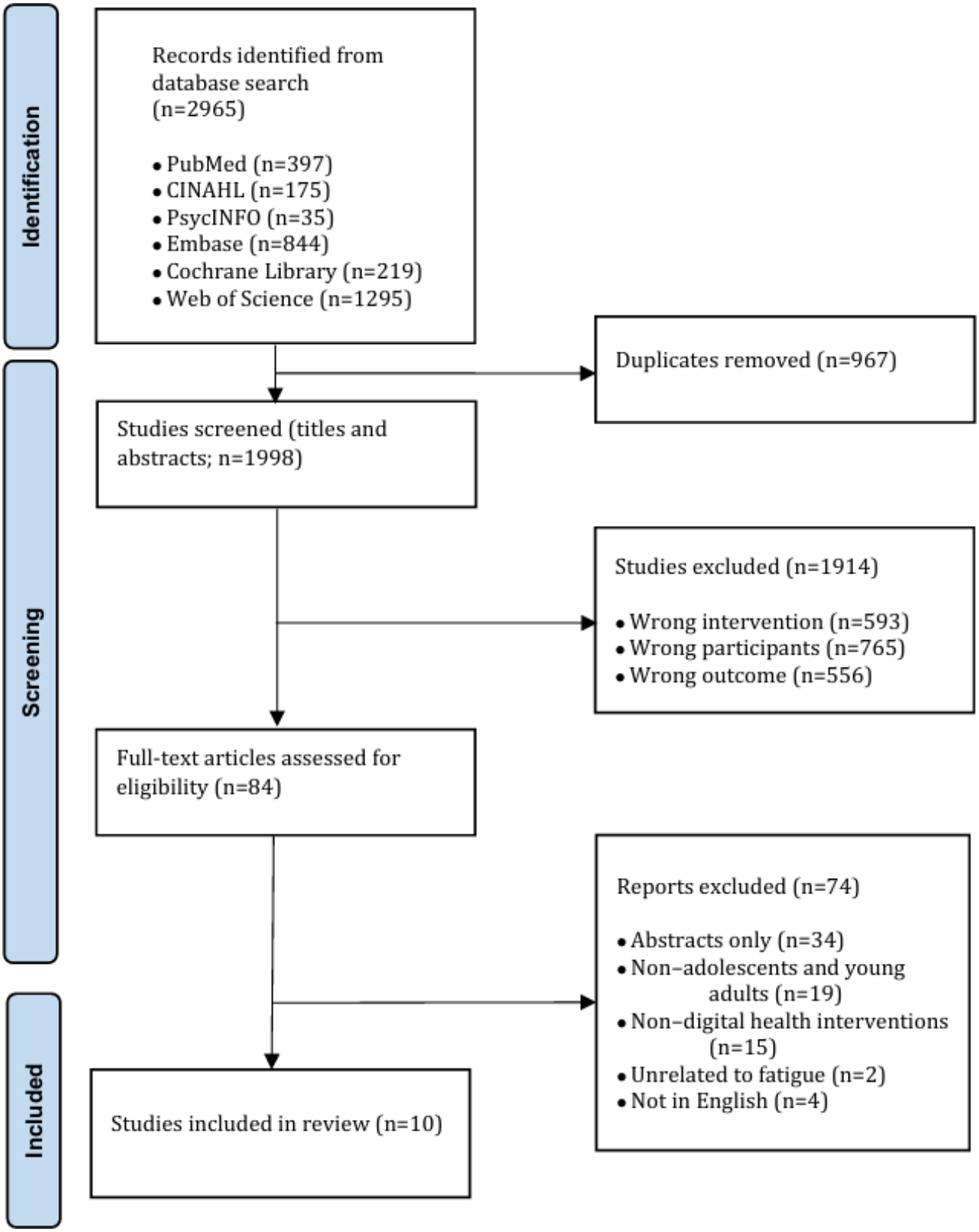
Results

Search Results

A total of 2965 articles were identified through the literature search. Of these 2965 articles, after removing 967 (32.61%) duplicates, an additional 1914 (64.55%) were excluded based on title and abstract screening as they did not meet the inclusion criteria. This left 84 articles for full-text review, of which 74 (88%) were excluded due to inaccessible full texts, non-English language, incompatible study population, or unsuitable interventions. Ultimately, 10 studies were included in the data synthesis. Figure 1 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the retrieved studies, the level of screening, and the included studies.

XSL•FO
RenderX

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



Characteristics of the Selected Articles

The included studies were published in the years 2012 (1/10, 10%) [25], 2018 (1/10, 10%) [26], 2020 (2/10, 20%) [27,28], 2022 (1/10, 10%) [29], 2023 (3/10, 30%) [30-32], and 2024 (2/10, 20%) [33,34]. The 10 studies included 5 (50%) randomized controlled trials [25,28,29,31,33], 2 (20%) quasi-experimental studies [27,30], 2 (20%) mixed methods studies [32,34], and 1 (10%) cohort study [26]. Of the 10 studies, 5 (50%) were conducted in the United States [25-29]; 2 (20%) were conducted in Canada [32,33]; and 1 (10%) each was conducted in Australia [30], Turkey [31], and the Netherlands [34]. A detailed summary of the characteristics of the included studies is provided in Multimedia Appendix 2 [25-34].

XSL•FO
RenderX

Summary of DHIs for Adolescents and Young Adults

Overview

Multimedia Appendix 2 provides a detailed description of the DHIs for adolescents and young adults. On the basis of the functions and forms of DHIs, we categorized them into the following 5 types: dynamic health monitoring and feedback (4/10, 40% of the studies) [26,28,29,33], automated online guidance and feedback (2/10, 20%) [25,27], live remote coaching and instruction (2/10, 20%) [32,34], gamified interventions (1/10, 10%) [31], and robot-assisted interventions (1/10, 10%) [30]. These DHIs were delivered through various technology platforms, including internet-based platforms (3/10, 30%), wearable devices with apps (3/10, 30%), wearable devices with internet-based programs (1/10, 10%), telehealth platforms (2/10, 20%), and robotic systems (1/10, 10%).

Dynamic Health Monitoring and Feedback

In total, 40% (4/10) of the studies explored the application of dynamic health monitoring and feedback technologies among adolescent and young adult patients with cancer. In this review, *dynamic health monitoring and feedback* refers to tracking and recording individual health data using wearable devices and assisting patients in adjusting their health behaviors through feedback provided by mobile apps.

In the study by Yurkiewicz et al [26], the intervention was based on Fitbit and iPad technology. Each participant received a Fitbit and an iPad Air preloaded with Headspace (a meditation app to support relaxation and mental health management), Re-Mission 2 (an interactive cancer-fighting game aimed at boosting morale), and the National Comprehensive Cancer Network Guidelines for Adolescent and Young Adult Oncology (informational sheets on treatment plans and recommendations to help participants understand their treatment). The Fitbit tracked steps, sleep, and calories burned. Participants could use these devices flexibly, either at the hospital or at home, and had access to an anonymous online community for social support. They were also allowed to download additional apps to suit their individual needs.

Devine et al [28] implemented an intervention called FitSurvivor grounded in social cognitive theory that emphasized the interactive influence of personal, environmental, and behavioral factors on individual behavior. The main objectives of the intervention were to enhance self-efficacy, goal setting, and self-monitoring; promote proper exercise posture; increase knowledge of cancer-related and general health behaviors; and provide social support. The intervention consisted of an initial 8-week face-to-face group course followed by a transition to a 4-week self-management program supported by a custom-built mobile app and Fitbit. Coaches assisted participants in setting specific exercise goals and provided weekly progress reviews. The app prompted participants to set goals upon their initial log-in and offered an exercise log feature, whereas the Fitbit enabled self-monitoring of steps and activity duration.

Tock et al [33] implemented the Lymfit intervention, which was designed based on self-determination theory, with the aim of enhancing patients' motivation to engage in physical activity (PA) by fulfilling the needs of autonomy, competence, and relatedness. The Lymfit intervention was conducted remotely. Participants first received a Fitbit and resistance bands by mail and were guided through device setup via videoconference. All participants were then added to a virtual Lymfit lounge within the Fitbit app, which provided a peer support group. On the basis of baseline assessments, a kinesiologist developed a personalized 12-week exercise plan for each participant, with biweekly video follow-ups to adjust the plan and offer support. The Fitbit tracked daily PA and provided real-time feedback to assist participants in self-monitoring and maintaining motivation. This intervention combined personalized exercise guidance with virtual peer support, using regular plan adjustments to encourage sustained health management and improve participants' exercise motivation and PA levels.

In the study conducted by Johnson et al [29], the intervention was similarly grounded in self-determination theory and was designed with a focus on the 3 core elements of autonomy, competence, and relatedness. The intervention included a wrist-worn PA tracking device (Fitbit), a Facebook group, step count goal setting, and a "buddy" system. Participants wore a Fitbit Flex activity tracker, set personalized goals, joined a Facebook group, participated in a PA "buddy" system, and received regular SMS text message reminders. The Fitbit tracked daily steps, and participants monitored their progress toward their goals through a mobile app. The Facebook group served as a social platform where participants could share their PA experiences, engage in discussions, and take part in group activities. In addition, participants could select an adult friend or family member as a "buddy" to provide encouragement and share exercise goals and achievements. They also received a daily motivational SMS text message related to PA at 3 PM Pacific Standard Time.

Automated Online Guidance and Feedback

A total of 20% (2/10) of the studies examined the use of automated online guidance and feedback for adolescent and young adult patients with cancer. In this review, *automated online guidance and feedback* refers to intervention services delivered through an internet platform. The system generates personalized health guidance and feedback automatically based on the patient's individual information and needs.

In the study conducted by Rabin et al [25], the intervention was grounded in the transtheoretical model and social cognitive theory, facilitating PA by integrating the stages of behavior change with the interaction among the individual, behavior, and environment. Participants in the intervention group were given 12 weeks of access to the *Step Into Motion* website, where researchers guided them in setting weekly PA goals, logging activities, and completing questionnaires to generate personalized feedback. The website provided a customized PA manual, feedback reports, additional resources, a frequently asked question section, links to the benefits and precautions of PA for young survivors of cancer, and access to an online discussion forum.

In the study conducted by Zhou and Recklitis [27], the SHUTi-AYA automated intervention program was used. Grounded in cognitive behavioral therapy for insomnia, the program focuses on the interplay among cognition, emotions,

XSL•FO
RenderX

and behaviors in the individual. It targets changing poor sleep habits, negative cognitive patterns, and emotional responses, thus improving sleep quality. SHUTi-AYA is an online, self-guided program that requires no external support. It offers personalized goal setting, graph-based feedback on insomnia symptoms, patient case studies, expert video explanations, and other interactive features, with content and visuals specifically tailored for adolescent and young adult survivors of cancer. After receiving personalized log-in credentials, participants completed 6 intervention sessions over an 8-week period, covering topics such as sleep restriction, stimulus control, cognitive therapy, sleep hygiene, and relapse prevention.

Live Remote Coaching and Instruction

In total, 20% (2/10) of the studies explored the use of live remote coaching and instruction among adolescent and young adult patients with cancer. In this review, *live remote coaching and instruction* refers to interactive courses and guidance delivered by professionals (eg, nurses, coaches, or therapists) via an online platform.

In the study by Wurz et al [32], the intervention was an 8-week online yoga program delivered via Zoom tailored for young people affected by cancer and their supporters. The program, developed from the evidence-based Yoga Thrive (a yoga program for patients with cancer) and the research team's expertise, incorporated systematic reviews, virtual class preferences, and participant feedback. Each week, a 60-minute session was offered, including physical postures, breathing exercises, and meditation, complemented by self-guided resources, social support, and reflection prompts to encourage behavior change.

In the study conducted by Bouwman et al [34], the intervention, called REVIVER, combined cognitive behavioral therapy with motivational interviewing. It aimed to help participants identify and alter negative thought patterns, enhance self-awareness and motivation, and foster behavior change. The intervention was designed to assist child, adolescent, and young adult survivors of cancer in managing the direct or indirect late effects of cancer. The intervention sessions were facilitated by trained nurses via video coaching. Over 3 months, participants completed an initial interview and 3 to 6 video counseling sessions, followed by a reflection meeting 6 months later.

Gamified Interventions

A total of 10% (1/10) of the studies investigated the use of gamified interventions for adolescent and young adult patients with cancer. In this review, *gamified interventions* refer to interventions incorporating game elements into the health intervention process.

In the study by Uluhan and Akçay Didişen [31], participants used the Re-Mission video game, which was designed to enhance the physical and mental well-being of patients with cancer. The game aimed to increase adolescents' understanding of the disease process, expand their knowledge of their condition, demonstrate potential physical changes during treatment, and enhance their understanding of both the effects and side effects of medication, all within a virtual environment. Participants received a game disc, with each session lasting 10 to 15 minutes, followed by a 1-hour break, with a maximum daily gameplay limit of 1 hour.

Robot-Assisted Interventions

A total of 10% (1/10) of the studies examined robot-assisted interventions among adolescent and young adult patients with cancer. In this review, *robot-assisted interventions* refer to those involving rehabilitation training conducted using robotic devices.

In the study by Atkinson et al [30], participants trained using either the Lokomat or ArmeoSpring rehabilitation robots focusing on gait or upper-limb function according to clinical needs, with only 1 type of training conducted per cycle. Training was conducted twice weekly over a 6-week period, starting with 20-minute sessions and gradually increasing to 40 minutes. Robot settings—such as reducing body weight support, increasing walking speed, decreasing guidance force, lowering upper-limb support, reducing motion control assistance, and increasing task speed—were customized based on participants' abilities and goals. Trainers made progressive adjustments to the settings in line with guidelines from the American College of Sports Medicine and Exercise and Sports Science Australia. Heart rate and blood oxygen saturation were monitored during training using a Nellcor forehead peripheral oxygen saturation sensor, and training intensity was adjusted in response to participant feedback.

CRF Outcomes

Of the 10 studies included in this review, 7 (70%) demonstrated a significant improvement in fatigue symptoms following the intervention. This section presents an analysis and summary of the results for each intervention type. Detailed data and further results are available in Multimedia Appendix 2.

In the dynamic health monitoring and feedback category, the study by Yurkiewicz et al [26] found a significant improvement in the *energy/fatigue* dimension of the RAND 36-Item Short Form Health Survey before and after the intervention, with scores increasing from 37.74 to 54.75 (P<.001). In the study by Devine et al [28], which used the Pediatric Quality of Life Inventory multidimensional fatigue scale, although the percentage increase generally favored the intervention group, there were no statistically significant group × time differences across fatigue outcomes after the intervention or at 6 months. In the study by Tock et al [33], the patient-reported outcomes measurement information system scale was used, and the results of the analysis of covariance model indicated a significant reduction in fatigue scores in the intervention group (P=.002). Johnson et al [29], through Fatigue Symptom Inventory scoring, found a significant difference in fatigue interference between the intervention and control groups (P=.03).

In the automated online guidance and feedback category, the study by Rabin et al [25] showed that the intervention group experienced a reduction in Profile of Mood States fatigue scores over 12 weeks. Although regression analysis did not reveal significant differences between the groups (P=.06), the findings suggested an alleviation of fatigue symptoms. Zhou and Recklitis [27] observed that, at 16 weeks after the intervention, participants' scores on the Pediatric Quality of Life Inventory

XSL•FO
RenderX

multidimensional fatigue scale were 70.9 (SD 22.2) on average, with an effect size of 1.2, corresponding to a large effect.

In the live remote coaching and instruction category, the study by Wurz et al [32] evaluated the improvement in fatigue symptoms using the Functional Assessment of Chronic Illness Therapy–Fatigue Scale. Although the *P* value was .10 and did not reach statistical significance, the effect size of 0.094 indicated a small to moderate improvement. Bouwman et al [34] assessed fatigue using the Checklist Individual Strength scale and found a significant reduction in fatigue symptoms after the intervention, with stable results maintained at the 6-month follow-up.

In the gamified intervention category, the study by Uluhan and Akçay Didişen [31] used the scale for the assessment of fatigue in pediatric oncology patients aged 13 to 18 years. After undergoing the Re-Mission video game intervention, the experimental group exhibited a significant reduction in fatigue scores at both the 1- and 3-month assessments compared to the control group ($P<.001$), indicating a significant improvement in fatigue level.

In the robot-assisted intervention category, the study by Atkinson et al [30] assessed fatigue using the Functional Assessment of Chronic Illness Therapy–Fatigue Scale. The results demonstrated a significant reduction in fatigue levels following Lokomat robotic training, with an effect size of $r=0.53$ and a *P* value of .04. For the ArmeoSpring robotic intervention, due to the limited sample size (n=3), only pre- and postintervention effect sizes were calculated, and inferential analysis was not conducted. However, preliminary findings suggested a large effect on fatigue improvement, with an effect size of $r>0.5$.

Physical Function and Emotional Well-Being

A total of 9 studies reported outcomes related to physical function, of which 6 (67%) demonstrated significant improvements [26,27,29,31-34], whereas 3 (33%) did not show significant changes [25,28,30]. The studies that reported significant improvement involved the DHI categories of dynamic health monitoring and feedback, automated online guidance and feedback, live remote coaching and instruction, and gamified interventions. In contrast, the studies showing no significant improvement primarily focused on dynamic health monitoring and feedback, automated online guidance and feedback, and robot-assisted interventions. In terms of emotional well-being, 6 studies reported related outcomes, of which 4 (67%) demonstrated significant improvements [26,27,31,32], whereas 2 (33%) did not show significant changes [25,28]. The studies with significant improvements involved the DHI categories of dynamic health monitoring and feedback, automated online guidance and feedback, live remote coaching and instruction, and gamified interventions. The studies that showed no significant improvement primarily focused on dynamic health monitoring and feedback and automated online guidance and feedback.

Discussion

Principal Findings

This scoping review comprehensively examined the effectiveness of various DHIs in reducing CRF among adolescents and young adults, including dynamic health monitoring and feedback, automated online guidance and feedback, live remote coaching and instruction, gamified interventions, and robot-assisted interventions. Multiple studies demonstrated significant changes in participants' fatigue scores following the interventions, with effects that persisted beyond the intervention period, indicating that DHIs are notably effective in alleviating CRF. In addition, DHIs also demonstrated potential in improving physical function and emotional health in this population. DHIs are well suited to the behavioral characteristics and preferences of adolescents and young adults, offering flexible, multidimensional support for CRF management and highlighting distinct advantages in this population.

Dynamic health monitoring and feedback refers to the use of apps and wearable devices to collect health data in real time and provide feedback. This enables patients to promptly identify potential health concerns and dynamically adjust their lifestyle and rehabilitation plans based on data-driven feedback [35,36]. For adolescents and young adults, the strength of these programs lies in their capacity to offer comprehensive support and high flexibility, aligning with their daily life rhythms. They can be easily incorporated into school, work, or social activities. Wearable devices continuously collect health data, whereas the apps process and analyze these data. This dynamic health monitoring and feedback assists adolescents and young adults in effectively managing their health through features such as setting personalized goals, tracking progress, sending regular reminders, and facilitating social interactions. This approach enables adolescents and young adults to track health changes continuously amid their busy lives and receive emotional support and encouragement, thus enhancing self-management motivation; improving exercise adherence; fostering healthy behavior habits; and, ultimately, alleviating CRF symptoms.

However, some studies did not show improvements in CRF [28,29]. This may be due to factors such as the limitations of the intervention content, insufficient intervention design, and underuse of social support. First, some intervention programs primarily focused on strength training and health education, but improving fatigue typically requires a more comprehensive approach [37]. For instance, incorporating aerobic training and psychological support can better address both the physiological and psychological aspects of fatigue, thereby enhancing the effectiveness of the intervention. Second, in interventions involving reminder systems, excessively frequent reminders may cause participants to focus too much on their fatigue, increasing their sensitivity and amplifying its negative impact on quality of life. In addition, although some programs included a social support component, the underuse of this support by participants led to suboptimal outcomes in this area, weakening the overall potential of the intervention to alleviate fatigue [35]. Therefore, future dynamic health monitoring and feedback

XSL•FO
RenderX

programs should optimize their content by integrating more aerobic training and psychological support to comprehensively target the physiological and psychological factors of fatigue. Furthermore, reminder systems should be personalized to avoid the negative effects of excessive reminders. The design of social support modules should also be refined to enhance the quality of social interactions, thereby improving participant engagement and the sustainability of the intervention, and more effectively alleviating CRF symptoms in adolescents and young adults.

Automated online guidance and feedback interventions provide adolescents and young adults with personalized, automated health guidance via web platforms or applications. Using algorithms and preset programs, these interventions automatically generate health advice and feedback based on participants' input data, ensuring that continuous support is accessible anytime and anywhere [36,38]. For busy adolescents and young adults, automated systems can rapidly respond to their needs, delivering timely and personalized recommendations through data-driven feedback, thereby reducing the delays typically associated with manual interventions. This supports participants in adhering more effectively to their rehabilitation plans, thereby facilitating effective CRF management and improving overall health management outcomes. Regarding interventions that did not lead to significant improvements, a potential reason could be that participants were unable to fully engage due to time constraints, which impacted the overall effectiveness of fatigue management [25]. To address this, automated online guidance and feedback interventions could incorporate the concept of microexercises. This approach would enable participants to perform short, efficient sessions within their busy schedules, thereby overcoming the common barrier of time constraints associated with longer workouts. In addition, the platform could integrate features such as health tracking, dietary advice, and stress management tools, further supporting comprehensive fatigue management. This would ensure that participants can maintain continuous fatigue management even when extended exercise sessions are not possible.

Live remote coaching and instruction interventions provide personalized health guidance and support to adolescents and young adults through online tools such as videoconferencing. A key advantage of these interventions is their ability to overcome geographical and temporal constraints, making them especially suitable for this population's complex schedules and varied lifestyle needs. This model allows participants to receive high-quality guidance without the need to travel to specific locations, significantly enhancing the intervention's convenience and flexibility and facilitating the smoother integration of rehabilitation into daily life [39]. In addition, live remote coaching and instruction offers adolescents and young adults the option to conduct rehabilitation at home or in other familiar settings, effectively reducing psychological stress and anxiety. This is particularly beneficial for managing CRF in adolescents and young adults as it minimizes the added fatigue associated with travel or unfamiliar environments, ultimately optimizing rehabilitation outcomes [40].

Gamified interventions effectively enhance the motivation and engagement of adolescents and young adults through interactive games. Given that this age group is particularly drawn to novel and engaging activities, the interactive and enjoyable nature of games can stimulate their motivation for PA, alleviating fatigue through enjoyable experiences and positive feedback, thereby optimizing overall fatigue management [41]. A key advantage of gamified interventions lies in their ability to integrate healthy behaviors within the game context, allowing participants to engage in health-related activities in a natural manner. In-game challenges and achievement systems provide adolescents and young adults with a sense of progress and accomplishment, which plays a crucial role in alleviating fatigue. Furthermore, gamification elements help structure health activities, enabling participants to consistently follow a well-organized plan, leading to more effective fatigue management [42].

Robot-assisted interventions provide more stable and reliable treatment than human-assisted methods through precise control and dynamically adjustable intensity. This high level of control accurately simulates natural body movements, ensuring optimal effectiveness in each session and allowing for personalized adjustments based on individual needs, thus enhancing both effectiveness and adaptability [43]. Moreover, robot-assisted technology is highly interactive and appeals to adolescents and young adults, who tend to show strong interest in emerging technologies and interactive experiences. The advanced, high-technology aspect of robot-assisted interventions can boost adolescents and young adults' motivation to participate, making the intervention process more engaging and sustainable. This technology-driven experience helps robot-assisted technology integrate more effectively into the treatment process, facilitating better management of CRF and improving intervention outcomes.

In the studies included in this review, those that did not yield significant results predominantly involved posttreatment patients with cancer. Thus, the effectiveness of DHIs in managing CRF among patients with cancer may be influenced by the treatment phase, and the selection of the type of DHI should be adaptable depending on the treatment phase. During the course of treatment, patients frequently experience severe fatigue directly associated with the treatment [44]. As a result, dynamic health monitoring and feedback–based DHIs are especially crucial. These technologies enable real-time monitoring of the patient's physical condition (eg, heart rate, sleep, and activity levels) and provide immediate feedback, assisting patients in better managing treatment-related discomfort. Furthermore, gamified interventions are highly appropriate for patients during treatment as their engaging and interactive characteristics can capture patients' interest and motivate them to participate more actively in PAs.

However, during the later phase of treatment, particularly in the recovery phase following cancer treatment, patients' needs evolve [45]. At this stage, fatigue typically becomes a chronic and persistent issue, making automated online guidance and feedback or live remote coaching and instruction–type DHIs more appropriate. These technologies allow patients to receive personalized health recommendations and interventions tailored to their symptoms, helping them learn how to manage fatigue in daily life, restore physical strength, and make adjustments to their diet and lifestyle while alleviating the difficulties arising from limited medical resources or the inconvenience of

XSL•FO
RenderX

accessing care during treatment. In addition, gamified interventions can play a positive role in the later phase of treatment. By promoting patients' adherence to PA through engaging interventions, these approaches help alleviate fatigue.

For patients with severe CRF who need long-term rehabilitation, robot-assisted rehabilitation offers an effective intervention. Robotic technology enables patients to undergo personalized physical rehabilitation training, enhancing physical functions such as balance, coordination, and muscle strength; restoring motor function [46]; and alleviating CRF symptoms.

Furthermore, research has indicated that DHIs show potential in improving both the physical function and emotional well-being of adolescent and young adult patients with cancer. These DHIs promote exercise adherence and motivation among adolescent and young adult patients; enhance their PA levels; and improve physical function through engaging exercise approaches, personalized exercise protocols, and real-time monitoring feedback. However, some studies did not show significant improvements in physical function, which may be attributed to the nature of the intervention technology and its usability experience. Research has shown that patients' active participation and self-efficacy are strongly associated with improvements in physical function [47]. However, the design of certain DHIs may have inadequately considered patient interaction, resulting in insufficient active patient participation, which in turn impairs the effectiveness of the intervention. Therefore, future DHIs should focus on the design of interactive modules to promote patients' autonomy and engagement, thus enhancing the effectiveness of the intervention. With respect to emotional well-being, DHIs help patients manage psychological stress during treatment by providing virtual social support and personalized guidance and interaction, thereby reducing anxiety and depression and fostering mental health recovery. However, some studies did not demonstrate significant emotional improvements, which may result from these interventions primarily focusing on physical health and behavior changes while the emotional support component remains relatively underdeveloped. When emotional support is insufficiently prioritized, the intensity of the intervention may be inadequate, leading to patients lacking continuous emotional support, which in turn impedes improvements in emotional well-being. Therefore, future DHIs should enhance the emotional support component to ensure that they not only address physical health but also provide robust emotional support and psychological guidance, thus amplifying the intervention's positive effects on emotional well-being.

In summary, DHIs that integrate digital health technologies with various intervention forms have demonstrated significant benefits in managing CRF in the adolescent and young adult population. Digital tools, due to their flexibility, personalization, engaging nature, and interactivity, not only transcend the limitations of time and space to cater to the diverse needs of adolescent and young adult patients across various living environments but also more effectively enhance CRF management by fostering greater patient engagement and long-term adherence. Simultaneously, by incorporating various intervention strategies, including behavioral interventions, emotional regulation, social interaction, and exercise motivation, they offer a comprehensive and integrated support system. This comprehensive intervention model transcends the time and spatial limitations of traditional treatments and optimizes CRF management by integrating multidimensional interventions that encompass physical, psychological, and social aspects. This highlights the significant potential of combining digital health technologies with multidimensional interventions, offering new perspectives on managing cancer-related symptoms and furthering the application of DHIs in symptom management.

Future Research

Although this review discussed the potential effects of various DHIs in reducing CRF among adolescents and young adults, several unresolved questions require further investigation. First, future research should prioritize larger sample sizes and use more rigorous study designs such as randomized controlled trials to enhance the external validity and statistical significance of the findings. Second, studies should explore the long-term effects of DHIs. Currently, most research assesses only short-term outcomes, whereas CRF, as a chronic symptom, requires sustained management to ensure optimal patient quality of life and health outcomes. Therefore, long-term follow-up studies are needed to evaluate the lasting effects of these interventions on fatigue reduction, quality of life, and recovery. In addition, the studies included in this review used various scales for fatigue assessment, underscoring the multidimensional nature of fatigue as a symptom. Future research may need to explore integrating multiple scales to capture the full spectrum of fatigue.

Limitations

The studies included in this review involved a diverse range of cancer types, with large variability in sample sizes across cancer types, thus preventing exploration of the relationship between specific cancer types and the choice of DHIs. In addition, this review included only 10 studies, of which 5 (50%) had small sample sizes, which may affect the generalizability of the results. Moreover, while the scoping review approach provides a broad overview, it does not assess the quality of included studies, which may affect the overall strength of the evidence.

Conclusions

This scoping review highlights the positive impact of DHIs in alleviating CRF among adolescents and young adults. Through diverse technological methods, DHIs offer a more engaging and interactive approach, effectively enhancing adherence and self-management skills in adolescent and young adult patients while also facilitating access to psychological and social support. DHIs demonstrate significant potential for delivering personalized, patient-centered cancer care, providing adolescent and young adult patients with cancer with new avenues to manage CRF and offering compelling evidence for the role of DHIs in symptom management.

XSL•FO
RenderX

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

Authors' Contributions

SJ: conceptualization, methodology, software, formal analysis, writing—original draft, and data curation. X Yang: software, formal analysis, writing—review and editing, and data curation. X Yu: conceptualization, writing—review and editing, supervision, project administration, funding acquisition, and resources.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 30 KB - mhealth_v13i1e68834_app1.docx]

Multimedia Appendix 2
Description of digital health interventions and related outcomes.
[DOCX File, 29 KB - mhealth_v13i1e68834_app2.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 172 KB - mhealth_v13i1e68834_app3.pdf]

References

1. Stone PC, Minton O. Cancer-related fatigue. Eur J Cancer 2008 May;44(8):1097-1104. [doi: 10.1016/j.ejca.2008.02.037] [Medline: 18381237]
2. D'Silva F, Javeth A, Singh P. Cancer-related fatigue - clinical evaluation scales and interventions: a systematic review. Indian J Palliat Care 2022;28(1):88-98. [doi: 10.25259/IJPC_455_20] [Medline: 35673374]
3. Minton O, Stone P. A systematic review of the scales used for the measurement of cancer-related fatigue (CRF). Ann Oncol 2009 Jan;20(1):17-25. [doi: 10.1093/annonc/mdn537] [Medline: 18678767]
4. Fabi A, Bhargava R, Fatigoni S, et al. Cancer-related fatigue: ESMO clinical practice guidelines for diagnosis and treatment. Ann Oncol 2020 Jun;31(6):713-723. [doi: 10.1016/j.annonc.2020.02.016] [Medline: 32173483]
5. Horneber M, Fischer I, Dimeo F, Rüffer JU, Weis J. Cancer-related fatigue: epidemiology, pathogenesis, diagnosis, and treatment. Dtsch Arztebl Int 2012 Mar;109(9):161-171. [doi: 10.3238/arztebl.2012.0161] [Medline: 22461866]
6. Dobrozsi S, Yan K, Hoffmann R, Panepinto J. Patient-reported health status during pediatric cancer treatment. Pediatr Blood Cancer 2017 Apr;64(4):e26295. [doi: 10.1002/pbc.26295]
7. Ma Y, He B, Jiang M, et al. Prevalence and risk factors of cancer-related fatigue: a systematic review and meta-analysis. Int J Nurs Stud 2020 Nov;111:103707. [doi: 10.1016/j.ijnurstu.2020.103707] [Medline: 32920423]
8. Nowe E, Stöbel-Richter Y, Sender A, Leuteritz K, Friedrich M, Geue K. Cancer-related fatigue in adolescents and young adults: a systematic review of the literature. Crit Rev Oncol Hematol 2017 Oct;118(63-69):63-69. [doi: 10.1016/j.critrevonc.2017.08.004] [Medline: 28917270]
9. Cosma A, Stevens G, Martin G, et al. Cross-national time trends in adolescent mental well-being from 2002 to 2018 and the explanatory role of schoolwork pressure. J Adolesc Health 2020 Jun;66(6S):S50-S58. [doi: 10.1016/j.jadohealth.2020.02.010] [Medline: 32446609]
10. Law PCF, Too LS, Butterworth P, Witt K, Reavley N, Milner AJ. A systematic review on the effect of work-related stressors on mental health of young workers. Int Arch Occup Environ Health 2020 Jul;93(5):611-622. [doi: 10.1007/s00420-020-01516-7] [Medline: 31932956]
11. Braun I, Friedrich M, Morgenstern L, et al. Changes, challenges and support in work, education and finances of adolescent and young adult (AYA) cancer survivors: a qualitative study. Eur J Oncol Nurs 2023 Jun;64:102329. [doi: 10.1016/j.ejon.2023.102329] [Medline: 37172338]
12. Agbejule OA, Hart NH, Ekberg S, Crichton M, Chan RJ. Self-management support for cancer-related fatigue: a systematic review. Int J Nurs Stud 2022 May;129:104206. [doi: 10.1016/j.ijnurstu.2022.104206] [Medline: 35279591]
13. Xu A, Wang Y, Wu X. Effectiveness of eHealth based self-management to improve cancer-related fatigue, self-efficacy and quality of life in cancer patients: systematic review and meta-analysis. J Adv Nurs 2019 Dec;75(12):3434-3447. [doi: 10.1111/jan.14197] [Medline: 31566769]
14. Devine KA, Viola AS, Coups EJ, Wu YP. Digital health interventions for adolescent and young adult cancer survivors. JCO Clin Cancer Inform 2018 Dec;2:1-15. [doi: 10.1200/CCI.17.00138] [Medline: 30652583]

XSL•FO
RenderX

15. Darley A, Coughlan B, Maguire R, McCann L, Furlong E. A bridge from uncertainty to understanding: the meaning of symptom management digital health technology during cancer treatment. Digit Health 2023;9:20552076231152163. [doi: 10.1177/20552076231152163] [Medline: 36714543]
16. Pan LC, Wu XR, Lu Y, et al. Artificial intelligence empowered digital health technologies in cancer survivorship care: a scoping review. Asia Pac J Oncol Nurs 2022 Dec;9(12):100127. [doi: 10.1016/j.apjon.2022.100127] [Medline: 36176267]
17. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144. [doi: 10.2196/34144] [Medline: 35311682]
18. Chang P, Engle J. Telemedicine and virtual interventions in cancer rehabilitation: practical application, complications and future potentials. Curr Oncol Rep 2024 Dec;26(12):1600-1605. [doi: 10.1007/s11912-024-01612-8] [Medline: 39503989]
19. Høyer E, Opheim A, Jørgensen V. Implementing the exoskeleton Ekso GTTM for gait rehabilitation in a stroke unit - feasibility, functional benefits and patient experiences. Disabil Rehabil Assist Technol 2022 May;17(4):473-479. [doi: 10.1080/17483107.2020.1800110] [Medline: 32838594]
20. Bandani-Susan B, Montazeri A, Haghighizadeh MH, Araban M. The effect of mobile health educational intervention on body image and fatigue in breast cancer survivors: a randomized controlled trial. Ir J Med Sci 2022 Aug;191(4):1599-1605. [doi: 10.1007/s11845-021-02738-5] [Medline: 34370166]
21. Ester M, Eisele M, Wurz A, McDonough MH, McNeely M, Culos-Reed SN. Current evidence and directions for future research in eHealth physical activity interventions for adults affected by cancer: systematic review. JMIR Cancer 2021 Sep 20;7(3):e28852. [doi: 10.2196/28852] [Medline: 34542415]
22. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143. [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
24. Westphaln KK, Regoeczi W, Masotya M, et al. From Arksey and O'Malley and beyond: customizations to enhance a team-based, mixed approach to scoping review methodology. MethodsX 2021;8:101375. [doi: 10.1016/j.mex.2021.101375] [Medline: 34430271]
25. Rabin C, Dunsiger S, Ness KK, Marcus BH. Internet-based physical activity intervention targeting young adult cancer survivors. J Adolesc Young Adult Oncol 2011 Dec;1(4):188-194. [doi: 10.1089/jayao.2011.0040] [Medline: 23610737]
26. Yurkiewicz IR, Simon P, Liedtke M, Dahl G, Dunn T. Effect of Fitbit and iPad wearable technology in health-related quality of life in adolescent and young adult cancer patients. J Adolesc Young Adult Oncol 2018 Oct;7(5):579-583. [doi: 10.1089/jayao.2018.0022] [Medline: 29924668]
27. Zhou ES, Recklitis CJ. Internet-delivered insomnia intervention improves sleep and quality of life for adolescent and young adult cancer survivors. Pediatr Blood Cancer 2020 Sep;67(9):e28506. [doi: 10.1002/pbc.28506] [Medline: 32568460]
28. Devine KA, Viola A, Levonyan-Radloff K, et al. Feasibility of FitSurvivor: a technology-enhanced group-based fitness intervention for adolescent and young adult survivors of childhood cancer. Pediatr Blood Cancer 2020 Sep;67(9):e28530. [doi: 10.1002/pbc.28530] [Medline: 32589339]
29. Johnson AM, Baker KS, Haviland MJ, et al. A pilot randomized controlled trial of a Fitbit- and Facebook-based physical activity intervention for young adult cancer survivors. J Adolesc Young Adult Oncol 2022 Aug;11(4):379-388. [doi: 10.1089/jayao.2021.0056] [Medline: 34677081]
30. Atkinson M, Tully A, Maher CA, Innes-Wong C, Russo RN, Osborn MP. Safety, feasibility and efficacy of Lokomat® and Armeo®Spring training in deconditioned paediatric, adolescent and young adult cancer patients. Cancers (Basel) 2023 Feb 16;15(4):1250. [doi: 10.3390/cancers15041250] [Medline: 36831593]
31. Uluhan C, Akçay Didişen N. Effects of the re-mission video game on fatigue and quality of life levels of adolescents diagnosed with cancer: a randomized controlled trial. Psychol Health Med 2024 Mar;29(3):603-614. [doi: 10.1080/13548506.2023.2289471] [Medline: 38112283]
32. Wurz A, McLaughlin E, Hughes K, et al. Exploring feasibility, perceptions of acceptability, and potential benefits of an 8-week yoga intervention delivered by videoconference for young adults affected by cancer: a single-arm hybrid effectiveness-implementation pilot study. Pilot Feasibility Stud 2023 Mar 10;9(1):37. [doi: 10.1186/s40814-023-01244-y] [Medline: 36899410]
33. Tock WL, Johnson NA, Andersen RE, et al. Pilot randomized controlled trial of Lymfit: a theory-guided exercise intervention for young adults with lymphoma. Healthcare (Basel) 2024 May 28;12(11):1101. [doi: 10.3390/healthcare12111101] [Medline: 38891177]
34. Bouwman E, Stollman I, Wilbers J, et al. Feasibility and potential effectiveness of nurse-led video-coaching interventions for childhood, adolescent, and young adult cancer survivors: the REVIVER study. BMC Cancer 2024 Jun 11;24(1):722. [doi: 10.1186/s12885-024-12430-3] [Medline: 38862904]
35. Dun L, Xian-Yi W, Si-Ting H. Effects of cognitive training and social support on cancer-related fatigue and quality of life in colorectal cancer survivors: a systematic review and meta-analysis. Integr Cancer Ther 2022;21:15347354221081271. [doi: 10.1177/15347354221081271] [Medline: 35225053]

XSL•FO
RenderX

36. Provoost S, Ruwaard J, van Breda W, Riper H, Bosse T. Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: an exploratory study. Front Psychol 2019;10:1065. [doi: 10.3389/fpsyg.2019.01065] [Medline: 31156504]
37. Brownstein CG, Twomey R, Temesi J, et al. Physiological and psychosocial correlates of cancer-related fatigue. J Cancer Surviv 2022 Dec;16(6):1339-1354. [doi: 10.1007/s11764-021-01115-6] [Medline: 34609702]
38. Mooney KH, Beck SL, Wong B, et al. Automated home monitoring and management of patient-reported symptoms during chemotherapy: results of the symptom care at home RCT. Cancer Med 2017 Mar;6(3):537-546. [doi: 10.1002/cam4.1002] [Medline: 28135050]
39. Dong X, Yi X, Ding M, et al. A longitudinal study of a multicomponent exercise intervention with remote guidance among breast cancer patients. Int J Environ Res Public Health 2020 May 14;17(10):3425. [doi: 10.3390/ijerph17103425] [Medline: 32423038]
40. O'Neill L, Brennan L, Sheill G, Connolly D, Guinan E, Hussey J. Moving forward with telehealth in cancer rehabilitation: patient perspectives from a mixed methods study. JMIR Cancer 2023 Nov 9;9:e46077. [doi: 10.2196/46077] [Medline: 37943595]
41. Rosipal NC, Mingle L, Smith J, Morris GS. Assessment of voluntary exercise behavior and active video gaming among adolescent and young adult patients during hematopoietic stem cell transplantation. J Pediatr Oncol Nurs 2013;30(1):24-33. [doi: 10.1177/1043454212461071] [Medline: 23160792]
42. Kim S, Wilson P, Abraham O. Investigating the use of serious games for cancer control among children and adolescents: scoping review. JMIR Serious Games 2024 Jul 10;12:e58724. [doi: 10.2196/58724] [Medline: 38985502]
43. Yeh SW, Lin LF, Tam KW, Tsai CP, Hong CH, Kuan YC. Efficacy of robot-assisted gait training in multiple sclerosis: a systematic review and meta-analysis. Mult Scler Relat Disord 2020 Mar 3;41:102034. [doi: 10.1016/j.msard.2020.102034] [Medline: 32200337]
44. Kang YE, Yoon JH, Park NH, Ahn YC, Lee EJ, Son CG. Prevalence of cancer-related fatigue based on severity: a systematic review and meta-analysis. Sci Rep 2023 Aug 7;13(1):12815. [doi: 10.1038/s41598-023-39046-0] [Medline: 37550326]
45. Tan CJ, Yip SYC, Chan RJ, Chew L, Chan A. Investigating how cancer-related symptoms influence work outcomes among cancer survivors: a systematic review. J Cancer Surviv 2022 Oct;16(5):1065-1078. [doi: 10.1007/s11764-021-01097-5] [Medline: 34424498]
46. Mugnosso M, Marini F, Holmes M, Morasso P, Zenzeri J. Muscle fatigue assessment during robot-mediated movements. J Neuroeng Rehabil 2018 Dec 17;15(1):119. [doi: 10.1186/s12984-018-0463-y] [Medline: 30558608]
47. Shao JH, Yu KH, Chen SH. Effectiveness of a self-management program for joint protection and physical activity in patients with rheumatoid arthritis: a randomized controlled trial. Int J Nurs Stud 2021 Apr;116:103752. [doi: 10.1016/j.ijnurstu.2020.103752] [Medline: 32928503]

Abbreviations

CRF: cancer-related fatigue
DHI: digital health intervention
MeSH: Medical Subject Heading
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by L Buis; submitted 15.11.24; peer-reviewed by KL Du, S Chantziara, WV Harten; revised version received 27.09.25; accepted 29.09.25; published 21.10.25.

Please cite as:
Jiang S, Yang X, Yu X
Digital Health Interventions to Reduce Cancer-Related Fatigue Among Adolescents and Young Adults: Scoping Review
JMIR Mhealth Uhealth 2025;13:e68834
URL: https://mhealth.jmir.org/2025/1/e68834
doi:10.2196/68834

© Shanshan Jiang, Xiaoyu Yang, Xinying Yu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Interventions for Military Members, Veterans, and Public Safety Personnel: Scoping Review

Rashell R Allen[1,2], MEd; Myrah A Malik[2,3], MSc; Carley Aquin[2,4], BA; Lucijana Herceg[2,4], MEd; Suzette Brémault-Phillips[2,3], PhD; Phillip R Sevigny[1,2,4], PhD, RPsych

[1]School and Clinical Child Psychology, Faculty of Education, University of Alberta, Edmonton, AB, Canada

[2]Heroes in Mind, Advocacy and Research Consortium, Faculty of Rehabilitation Medicine, University of Alberta, 8205 114 St NW, Edmonton, AB, Canada

[3]Department of Occupational Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[4]Counselling Psychology, Faculty of Education, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Rashell R Allen, MEd
School and Clinical Child Psychology, Faculty of Education, University of Alberta, Edmonton, AB, Canada

Abstract

Background: Accessible mental health support is essential for military members (MMs), veterans, and public safety personnel (PSP) who are at an increased risk of mental health challenges. Unique barriers to care, however, often leave these populations going untreated. Mental health treatment delivered via apps or websites (ie, digital mental health interventions [DMHIs]) offers an accessible alternative to in-person therapy.

Objective: We aimed to synthesize the current literature on apps and web-based programs focused on enhancing MMs', PSPs', and veterans' resilience or well-being. A multidimensional well-being model, I-COPPE (interpersonal, community, occupational, physical, psychological, economic, and overall well-being), was used as a framework guiding the scoping review.

Methods: A search of 6 databases was conducted using key terms related to (1) population, (2) resilience and well-being constructs, and (3) web- or mobile-based programs. At all levels of screening, at least 2 researchers (RRA, MAM, and CA) reviewed each paper independently. Data were extracted and recorded to include relevant study characteristics including program name and description, target population, number of participants, therapeutic approach, results, limitations, and I-COPPE dimension supported. A narrative synthesis was performed to summarize the eligible studies.

Results: In total, 44 papers were included in the study and 39 unique resilience or well-being apps or web-based programs identified for MMs, PSP, or veterans. The programs largely focused on veteran populations (28/44, 64%). In total, 51% (20/39) of programs relied on cognitive behavioral approaches and most aimed to support posttraumatic stress disorder–related symptoms. In consideration of the I-COPPE model, a majority supported psychological well-being, followed by interpersonal and physical well-being. Most apps were believed to support more than 1 domain of well-being. The main methodologies used in the literature to evaluate digital mental health interventions include randomized controlled trials, secondary analyses, and pilot randomized controlled trials with evaluations of feasibility, acceptability, satisfaction, or qualitative feedback. Generalizability of findings was commonly limited by attrition rates and small sample sizes.

Conclusions: DMHIs for MMs, PSP, and veterans appear promising due to their accessibility and scalability. More research is needed, however, to determine whether DMHIs are an effective alternative to in-person mental health care. The current review contributes to the literature by compiling evidence of DMHIs and the domains of well-being supported by, and the therapeutic orientation of, these programs. Our review revealed that more research is needed to determine the effectiveness and efficacy of DMHIs offered to these populations.

(JMIR Mhealth Uhealth 2025;13:e65149) doi:10.2196/65149

KEYWORDS

public safety personnel; veterans; military members; mobile apps; web-based programs; resilience; well-being; community; psychological; PTSD; scoping review; PRISMA

XSL•FO
RenderX

Introduction

Background

Military members (MMs) and public safety personnel (PSP; ie, police officers, paramedics, firefighters, correctional officers, peacekeepers, and emergency medical technicians [1]) stand ready and resilient to protect civilians at a moment's notice. These individuals, including veterans who have completed their military service, have been grouped in this paper due to similarities across their public service work. These individuals, as a result of their work, typically turn toward danger when most turn away. Although there may be potential worldview and cultural and day-to-day task differences across these populations, they have been grouped due to their increased risk of exposure to potentially psychologically traumatic events (PPTEs) as a result of their work [2-4]. Such events may include direct or indirect exposure to death (accidental or violent), transportation accidents, assaults, natural disasters, and military conflict [2,5-7]. PPTE exposure is associated with an increased risk for mental health concerns, including but not limited to posttraumatic stress disorder (PTSD), generalized anxiety disorder, and depression [2,8,9]. The occupational hazards of PPTEs and other occupational stressors such as family separations, relocations, and the long, unpredictable work hours often experienced have implications for well-being and family life [10,11]. Such professional expectations combined with occupational stressors can have negative implications for well-being and thus impact their readiness for occupational duties.

A systematic review (k=11) revealed that approximately 29% (55,336/189,021) of MMs who reported a recent mental health concern accessed or sought out services [12], with similar rates observed for veterans [13]. Barriers to service access for PSP, MMs, and veterans include fear of experiencing worsening symptoms, stigma and confidentiality breaches, and avoidance often characteristic of PTSD, as well as practical barriers related to availability of treatment, remote and shift work, transportation, and relocation [13-17]. In addition, the COVID-19 pandemic highlighted the need for greater investment in technology-delivered mental health interventions, referred to as digital mental health interventions (DMHIs) [18-20]. The measures taken for infection control (ie, enforced restrictions, such as social distancing) subsequently reduced access to protective factors (eg, social support networks), contributed to worsened mental health, and interrupted the delivery of in-person treatment [21,22]. Such barriers necessitate specialized mental health services for PSP, MMs, and veterans that are private, accessible, secure, and effective.

Given the accessibility of DMHIs, there are several advantages for their use by MMs, PSP, and veterans. DMHIs offer portable, low-cost, accessible, convenient, personalized, scalable, and self-guided solutions to enhancing well-being [23,24]. Moreover, DMHIs are amenable to inconspicuous use and customization to user preferences and symptomatology [25]. There has been a shift toward DMHIs [26], a promising method for delivering patient-centered health interventions based on people's present needs [27,28]. Notably, however, in a scoping review of apps for MMs and PSP, only 50% (11/22) were subject to a randomized controlled trial (RCT) [29]. Additional limitations of the apps were noted, such as small sample sizes, limited generalizability of findings, uncontrolled evaluations, and limited follow-up periods [29]. Tam-Seto et al [30], in another scoping review of mental health apps for MMs and veterans, have also reported a lack of research evidence. Based on their review, Tam-Seto et al [30] reported that these populations show a general willingness to use apps, and they suggest that apps may be a supplement to traditional treatment and could be a desirable option for those who fear stigmatization. At this time, although there are some advantages noted for DMHIs, there appears to be insufficient empirical support [31-33]; however, a more recent systematic review of meta-analyses for the general population showed more promising results [34]. For example, Goldberg et al [34] looked at mobile apps, interventions with components of smartphones, and app use with additional equipment and found 14 meta-analyses that met criteria for their review. Of the 14 meta-analyses, 8 showed highly suggestive effect sizes, 4 suggestive, 14 weak, and 8 non-significant [34].

While DMHIs offer many advantages, apps and web-based programs tend to focus solely on enhancing psychological resilience, as opposed to viewing resilience as a multifaceted and dynamic construct [35,36]. Researchers recognize that resilience interventions may lead to positive spillover effects or *collateral change* [37,38]. Such spillover effects, or the ways in which systems influence the resilience of other distinct systems, can be accounted for by a multisystemic resilience perspective [39,40]. Therefore, we define resilience as a system's process of adaptation (eg, individuals, groups, and organizations) following adversity or risk exposure. Resilience is a dynamic process influenced by socioecological system interactions and available resources, which function as protective factors and lead to improved outcomes when coping with and overcoming adversity [41-43].

Given the importance of well-being in our ability to adapt to stress [44], researchers have started to endorse models wherein improved well-being leads to greater resilience [44,45]. Traditionally, 2 major dimensions of well-being include eudaimonic well-being (behaving virtuously or *functioning well*) and hedonic well-being (*feeling good*) [46]. A multidimensional and multilevel perspective of well-being is the degree to which individual, dyadic, community, and organizations' needs are satisfied [47,48]. Prilleltensky [49] proposed 7 dimensions of well-being: interpersonal, community, occupational, physical, psychological, economic, and overall well-being (I-COPPE). Given that resilience is multiply determined [40,44], well-being interventions may promote resilience through the enhancement of resources, capacities, and coping skills [37,40]. Consideration of multidimensional well-being and its effects on resilience responds to the call for resilience researchers to expand their focus from remediation of mental distress to positive adaptation and protective processes across multiple levels of influence [50].

A multidimensional focus on well-being is conducive to a *whole-person* approach, which recognizes individuals' diverse needs that may influence resilience trajectories. As a result, interventions should aim to strengthen protective factors and

XSL•FO
RenderX

bolster the development and use of resources to promote resilient responding [45,51-53]. Conversely, the loss of cardinal resources (eg, shelter, safety, and physical well-being) may limit the degree to which individuals benefit from interventions [51,54,55]. While a number of variables are consistently linked to individual resilience trajectories, no particular variable exerts a dominating influence [35], as such interventions should strive to strengthen protective networks [53].

Purpose

Two relevant scoping reviews have been published recently. First, Voth et al [29] completed a review of resilience-based apps and programs for MMs and PSP, with 32 studies meeting inclusion criteria. Second, Tam-Seto et al [30] completed a scoping review of mental health mobile apps for MMs and veterans, with 35 papers meeting inclusion criteria. Since the publication of scoping reviews by both Voth et al [29] and Tam-Seto et al [30], there has been a greater reliance on DMHIs as a result of the COVID-19 pandemic [56,57], necessitating an updated review. Our scoping review differs from that of Voth et al [29] and contributes to the literature by (1) including both well-being- and resilience-based DMHIs, (2) including 34 new search terms, (3) using a multidimensional well-being framework to describe the DMHIs and potential resilience spillover effects, and (4) including 3 years of novel research published after their search, which also captures new DMHI research during and post–COVID-19 pandemic. Our scoping review also differs from that of Tam-Seto et al [30], who focused exclusively on apps for MMs and veteran populations, by including PSP populations and web-based programs in our search. No study to date has reviewed well-being and resilience DMHIs for MMs, PSP, and veterans. The findings will be synthesized with respect to the I-COPPE model [58], given the proposed upstream effects of well-being interventions on resilience trajectories [37]. We aim to explore the quality of and to highlight the benefits, gaps, and limitations of peer-reviewed literature on independently used DMHIs (ie, without mental health professional, peer, or researcher support or guidance).

Methods

Design of the Review

We initially based our search on similar concepts to Voth et al [29]; however, using an iterative process, we added 34 new search terms, including *front-line worker, veteran, well-being, emotion regulation, internet or online intervention or program, mental health app, and mobile app* (see Multimedia Appendix 1 for full list of search terms). The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) was used to inform decision-making and paper identification [59]. No registered review protocol exists for the current project; however, the information provided may be used to replicate the current findings.

Eligibility Criteria

Literature included in the search encompassed studies aimed at capturing app- and web-based resilience and well-being interventions for MMs, PSP, and veterans (see Textbox 1 for eligibility criteria). Eligible studies evaluated app or web-based resilience and well-being interventions for MMs, PSP, and veterans. The authors defined a web-based program as an *interactive online platform* that does not involve an invitation or log-in code.

Textbox 1. Eligibility criteria for papers included in the study.

Inclusion criteria

- Self-directed app or web-based intervention meant to improve well-being or resilience.
- The study population included military members (MMs), veterans, public safety personnel (PSP), and students or individuals in training for these professions. Studies focusing on MMs, veterans, and PSP family units were also eligible for inclusion.
- Primary analysis (ie, quantitative analysis and qualitative), including studies evaluating app or program effects, acceptability, usability, feasibility, or change over time.
- English papers published from 2000 to December 13, 2022.
- The study included an app-based or web-based program related to and intended to improve resilience or well-being (ie, substance use cessation, sleep coaching, yoga, social, mental health, physical health, and mindfulness).

Exclusion criteria

- The intervention included a guided (synchronous) support component for the intervention (eg, in-person or virtual check-ins [or sessions], therapist interaction, virtual reality, or in-person meetings or discussions).
- App or programs targeting MMs, PSP, and veteran family members, without also focusing on the MMs, PSP, or veteran. App or programs intended only for service provider use with clientele.
- Study is not a primary analysis (eg, single-subject studies, review papers, meta-analyses, research proposals or protocols, and abstract or conference submissions).
- Non-English papers published before 2000 and after February 27, 2024.
- Intervention involved additional technology related to the intervention (eg, heart rate variability monitors and wearable technology).

XSL•FO
RenderX

Inclusion Criteria

To capture the increase in apps and websites available and technological innovations during this time, this study included literature published from 2000 to February 27, 2024. In addition, studies meeting inclusion criteria examined DMHIs through primary analyses and focused on PSP, MMs, or veteran populations.

Exclusion Criteria

Research on synchronous interventions was excluded, given our aim is to evaluate asynchronous apps and programs. More specifically, we aimed to evaluate DMHIs that were not potentially mediated by therapist, peer, psychologist, or researcher support and guidance. Therefore, any DMHI study that included peer, therapist, psychologist, or researcher support was excluded to allow us to review the literature that solely looked at DMHI usability, acceptability, feasibility, effectiveness, or efficacy that was not potentially mediated by another support person. Studies that included 2 treatment conditions, 1 with additional support and 1 without additional support, were included in this study. DMHIs requiring additional technology (beyond a mobile or computer device) were also excluded for the same reason. Finally, studies on DMHIs aimed at supporting MMs, PSP, or veteran families that did not include a focus on the MMs, PSP, or veterans were excluded.

Search Strategy and Information Sources

To identify relevant literature, a search (using a Boolean format) was conducted using key terms based on three concepts: (1) population (eg, military, veteran, or PSP), (2) resilience and well-being–related constructs (eg, hardiness or grit), and (3) web- or mobile-based programs (eg, game or apps). The final search was conducted on February 2024 of the following databases: Academic Search Complete, CINAHL, APA PsycINFO, Embase, SocINDEX, and MEDLINE.

Eligibility Assessment and Study Selection

Three researchers (RRA, MAM, and CA) were involved in the eligibility assessment and study selection, with at least 2 reviewing each article independently. The researchers initially met to discuss and review the eligibility criteria and conducted a prescreening with 12 papers to assess the comprehensiveness and overall agreement of the search criteria. There was disagreement on 1 paper, which was then discussed in-depth, leading to further clarification of eligibility criteria. At all levels of screening (eg, title, abstract, and full-text review), a minimum of 2 researchers reviewed each paper independently. All abstracts with discrepant ratings were included for further review during full-text screening. During the full-text review, conflicts were reviewed by a third reviewer (RRA, MAM, or CA) and discussed in a research team meeting. Following this, final eliminations and decisions were made. Any discrepancies in agreements were recorded for the purpose of calculating interrater reliability.

Charting the Data

Data were extracted and recorded by a minimum of 2 researchers (RRA, MAM, and CA) for each study to include relevant study characteristics: app or program name and description, target population and number of participants, facets of resilience they purport to support, methods used to support resilience, results, gaps in service or limitations, and I-COPPE dimension supported. Since scoping reviews are meant to be descriptive as opposed to evaluative [60], eligible studies were not assessed in terms of quality. Instead, a narrative synthesis was performed by 3 researchers (RRA, MAM, and CA) to summarize the eligible studies. Specifically, the findings from multiple studies were summarized and explained descriptively. After 2 researchers (RRA, MAM, or CA) charted the data independently, a third reviewer (RRA, MAM, or CA) synthesized the data and identified and rectified any discrepancies between the previous researchers to minimize potential errors. The researchers then met to discuss the resulting data synthesis and core content themes that emerged based on the data extraction. The I-COPPE model [58] provided a theoretical framework for the synthesis, whereas the cascading resilience model [37] represents the underlying theory illuminating how resilience may be supported in these interventions.

Results

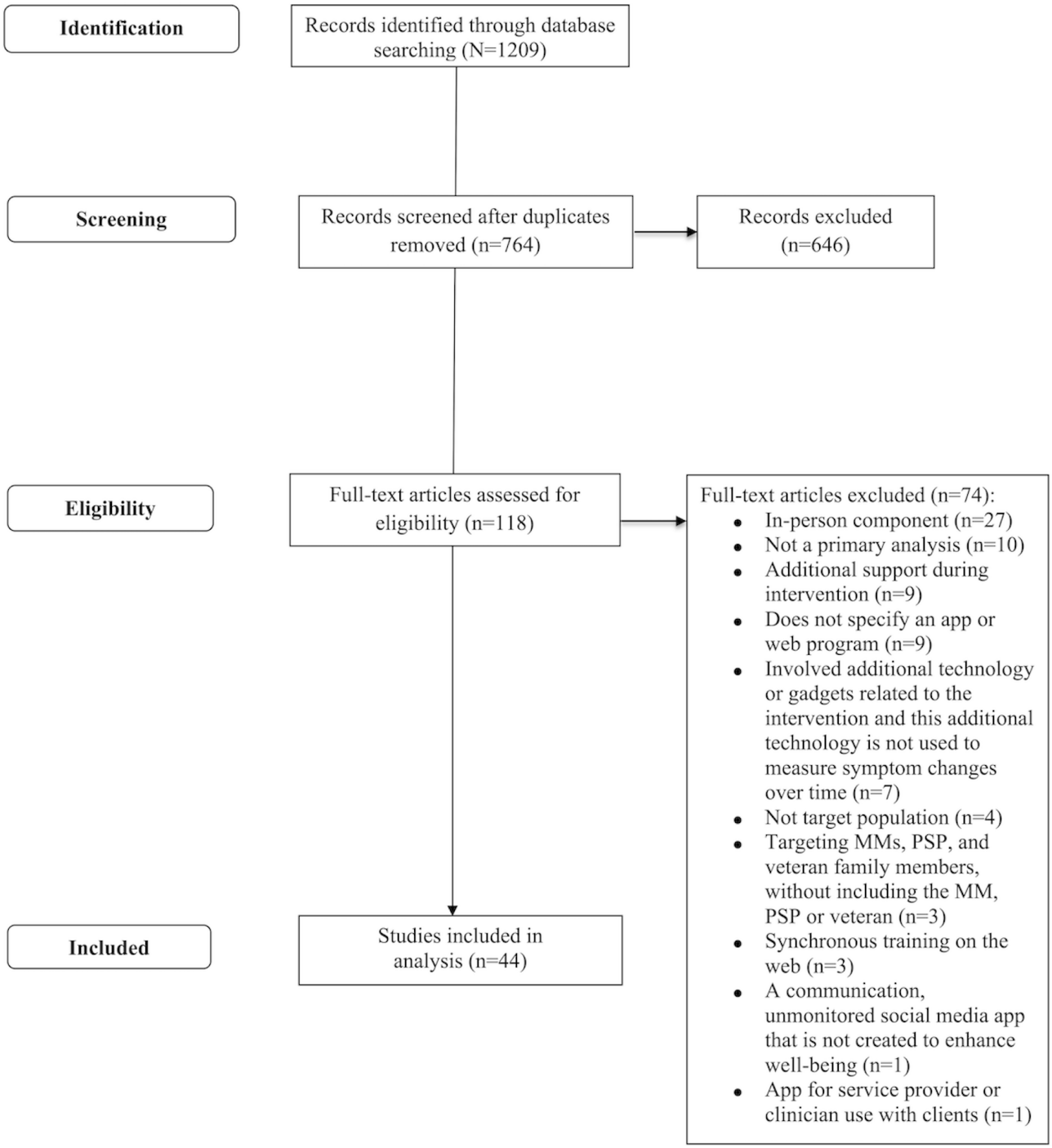
Study Selection

Details of the screening process can be found in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (Figure 1). A total of 1209 papers were identified, with 434 duplicates, resulting in 775 papers. Three researchers completed initial reviews of titles and abstracts with a range of agreement (Cohen κ) from moderate (κ=0.48) to substantial (κ=0.79). All papers with disagreement were included in the full-text review to allow for a more comprehensive evaluation of the discrepancies. In the end, 118 papers met criteria for full-text review.

Agreement for full-text reviews was fair (κ=0.20) to substantial (κ=0.79). One researcher (MAM) contacted 8 authors because the DMHI name was not specified in reviewed papers. Five studies were excluded due to no response. Three authors responded [61-63], and upon further review, 2 studies met criteria [61,62] whereas 1 [63] did not because the program included various unnamed resources from the Veterans Affairs website. Finally, 1 study [64] was included on the merit that the program would indirectly help veterans through enhancing support strategies provided by their spouse or partner. In the end, 44 papers were included in the study.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. PSP: public safety personnel.



Narrative Synthesis

Characteristics of Studies and Populations

Of the 44 papers reviewed, a majority (35/44, 80%) were written by authors residing in the United States and 73% (32/44) were published from 2020 onward, with only 7% (3/44) of papers published prior to 2013 (Table 1).

Most programs focused on veteran populations (28/44, 64%), followed by PSP (8/44, 18%) and MMs (5/44, 11%). Across populations, substance use and PTSD symptomatology were the most common presenting concerns targeted by the programs (Table 2).

XSL•FO
RenderX

Table . Summary of study location and publication year (N=44).

Paper information	Values, n
Location of study	1
Canada	1
United States	35
United Kingdom	3
Switzerland	1
Germany and Switzerland[a]	1
Netherlands	1
Australia	2
Publication year	
2022 to February 2024	8
2020 - 2022	13
2017 - 2019	11
2014 - 2016	9
2013 or earlier	3

[a]The study was conducted in both Germany and Switzerland.

XSL•FO
RenderX

Table . Summary of key paper information (N=44).

Paper criteria	Values, n
Population and presenting concerns	
Family, friend, or partner of veterans	1
Family member of veteran	1
With PTSD[a]	1
Military members	2
With behavioral health issues	2
Military members and veterans	2
With PTSD symptoms	2
Public safety personnel	8
Firefighters	2
Medical or emergency personnel (ie, police and firefighters)	2
Ambulance workers and firefighters	1
Health care professionals	1
With PTSD symptoms	1
Police officers or students	1
Student paramedics	1
Veterans	28
With PTSD symptoms	5
With PTSD symptoms and substance use	5
With substance use	5
With suicidal ideation	3
With chronic musculoskeletal pain	2
With lower back pain	2
With mental illness (borderline personality disorder, PTSD, bipolar disorder, or depression)	2
With history of chronic mental illness and psychosis symptoms	1
With history of mild traumatic brain injury	1
With insomnia symptoms	1
With substance use and sleep problems	1
Veterans and their nonveteran partners	1
DMHI[b] as stand-alone or supplementary to therapeutic care	
Supplementary to therapeutic care	5
Not supplementary to therapeutic care	39

[a]PTSD: posttraumatic stress disorder.

[b]DMHI: digital mental health intervention.

Characteristics of DMHIs

While Voth et al [29] reviewed a total of 32 papers and 22 apps, this study evaluated 44 papers and 39 DMHIs (see Table S1 in Multimedia Appendix 2 for a list of the DMHIs and their characteristics [23,45,61,62,64-103]). Out of the programs evaluated, 16 were assessed as web-based programs, 22 were examined as apps, and 1 was evaluated as both an app and a web-based program. In total, 20 programs relied on cognitive behavioral approaches (ie, cognitive behavioral therapy, exposure therapy, and stress inoculation). Ten of the apps used either psychoeducation or self-management, and 5 programs used motivational interviewing or applied behavior change theory. Five programs emphasized social skills or engagement, several programs focused on regulation strategies (eg, relaxation, stress management, and emotion regulation), and 5 focused on mindfulness as a therapeutic approach. Few programs emphasized social cognitive theory and self-monitoring. The

XSL•FO
RenderX

least common therapeutic approaches were SMART goals, sleep restriction therapy, and massage, as well as biopsychosocial, patient-centered, and strengths-focused approaches. Four apps did not specify their therapeutic approach.

Of all 44 papers included in the scoping review, 30% (13/44) were RCTs and 20% (9/44) pilot RCTs (most with evaluations of feasibility, acceptability, satisfaction, or qualitative feedback). Other common research designs included secondary analyses of studies already included (6/44, 14%), quasi-experimental (4/44, 9%), multiphase (eg, app development, feedback, and feasibility; 3/44, 7%), pilot quasi-experimental (2/44, 5%), and feasibility or acceptability (2/44, 5%; Table 3). Other methodological approaches in the DMHI literature for MMs, PSP, and veterans included pilot study for user satisfaction, perceived helpfulness, and usage patterns, mixed methods, and proof of concept.

Table . Summary of study methodologies (N=44).

Paper criteria	Values, n
Study design	
RCT[a]	13
Pilot study	11
RCT and feasibility, acceptability, satisfaction, or qualitative	6
RCT	3
Quasi-experimental	2
User satisfaction, perceived helpfulness, and usage patterns	1
Secondary analysis of study already included	6
Quasi-experimental design	4
Multiphase (eg, app development, feedback, and feasibility)	3
Feasibility or acceptability trial	2
Mixed methods	1
Proof of concept	1
Qualitative	1
Repeated-measures design	1

[a]RCT: randomized controlled trial.

I-COPPE Domains

All papers were examined and synthesized with consideration of the I-COPPE model [58] of well-being. The majority of apps were judged to support more than 1 domain of well-being (see Table S1 in Multimedia Appendix 3 for summary). Programs conceptualized within the psychological well-being domain (97%, 38/39) featured themes of mental health symptom management, treatment modalities such as cognitive behavioral therapy and dialectical behavioral therapy, self-efficacy, and motivation for change. Of the 39 programs included, the most commonly targeted psychological proximal and distal factors were symptoms of PTSD (22/39, 56% of programs); depression (14/39, 36%); resilience (12/39, 31%); self-efficacy, coping, or emotion regulation (10/39, 26%); and quality of life (8/39, 21%). Papers that were grouped within the interpersonal well-being dimension (21/39, 54%) tended to emphasize the development of social support networks for support in mitigating physical and mental health symptoms. A common interpersonal proximal and distal factor targeted and evaluated was communication (4/39, 10%). Apps and programs that fell within the physical well-being domain (21/39, 54%) emphasized reduction of physical symptoms related to sleep difficulties, substance use, PTSD (ie, elevated heart rate), and concussion. Management of psychosis, substance abuse, and pain were categorized within this domain, given their associated physical health implications. Studies categorized under the overall well-being dimension (3/39, 8%) included holistic measures of quality of life and life satisfaction. For community well-being (2/39, 5%), these papers and programs focused on promoting participants' community engagement and support. Papers supporting occupational well-being (1/39, 3%) concerned occupational support, guided communication, reducing stigma, and greater support services and resources. The economic domain was not captured within the current literature.

Results of DMHI Programs

Overall, program use typically led to reported improvements in these areas; however, few reported statistically significant changes as compared with treatment-as-usual or control groups (see Table S1 in Multimedia Appendix 4 for summary of study methodologies and results [23,45,61,62,64-103]). The programs that reported positive efficacy results via RCT were (1) Concussion Coach for decreasing postconcussive symptoms [65], (2) Family Foundations for decreasing parent depression [61], (3) Mission Reconnect for improving sleep quality and response to stressful experience [66], (4) Resilience@Work for facilitating coping behaviors (ie, optimism, use of instrumental and emotional support, and active coping) [67], (5) Support Coach for increasing psychological resilience [68], (6) Thinking

XSL•FO
RenderX

Forward for decline in perceived alcohol consumption [69], (7) VetChange for improving drinking behaviors and PTSD symptoms [70], and (8) Virtual Hope Box for facilitating coping with unpleasant thoughts or emotions [71].

Treatment as Stand-Alone or With Support

Some mobile apps and web programs differed in whether they were examined as stand-alone interventions, or whether participants received additional support. Most DMHIs were self-guided. Studies on T2 Mood Tracker [72], Virtual Hope Box [71,73,74], PTSD Coach [75], and Information about Drinking in Ex-serving personnel [23] were supplemented by clinician support, feedback, or clinical monitoring. Other nonclinician supplementary care was provided by peers [76], research staff [77], or mental health specialists [78]. Four papers did not assess intervention effectiveness based on group membership (ie, support vs no support) [23,72,74,78]. For PTSD Coach [75], both groups (ie, those with or without additional support) experienced significant reductions in PTSD symptoms; however, no significant group differences emerged. Those in the clinician-supported group were more likely to attend an additional PTSD session and accept a treatment referral [75]. For Virtual Hope Box, no differences emerged between treatment groups (treatment-as-usual with or without Virtual Hope Box) were reported for enlisting support from friends and family, suicidal ideation, reasons for living, coping self-efficacy, and suicidal ideation [71,73]. For support offered via check-in versus self-managed groups, no statistically significant group differences emerged [77]. Finally, in the feasibility pilot study of Thinking Forward, Possemato et al [76] found no significant difference between peer support and self-managed groups for alcohol use, PTSD symptoms, resiliency, social quality of life, coping, and psychological quality of life.

Mediators and Moderators

A number of moderators and mediators were considered across the studies to assess their impact on post–program outcomes, engagement, and program effectiveness. Several moderators were statistically significant, including parent gender [61], PTSD symptoms [79], combat exposure [79], self-efficacy [65], and interpersonal problems [80]. More specifically, there was a dosage effect for father-reported parenting undermining, but not for mothers [61]. Brief et al [79] found that there was a sharper decline in drinking behaviors for individuals with higher levels of baseline PTSD symptoms and baseline combat exposure. Belanger et al [65] discovered a greater probability for reduction in PTSD symptoms and psychological distress when self-efficacy was increased. Finally, Polizzi et al [80] found that individuals who reported higher interpersonal problems as a result of their drinking at baseline demonstrated greater PTSD symptoms at baseline and exhibited greater reduction in PTSD symptoms postintervention than those with moderate and low interpersonal problems. In terms of mediators, Williams et al [81] reported significant mediators for 2 apps. For Drinker's Check-Up, perceived norms of same-age peers for quantity of drinks and number of drinking occasions were mediators negatively and significantly impacting participants' alcohol use behaviors [81]. For Alcohol Savvy, perceived norms (of same-age peers) regarding number of drinking occasions were a mediator negatively and significantly impacting participants' alcohol use behaviors [81].

Gaps and Limitations

From the 38 studies, there appeared to be recurring limitations for the research reviewed. Two of the most commonly cited limitations were difficulties with attrition (attrition rates ranging from 0% to 82%) and small sample size, limiting the generalizability of findings. In addition, the majority of studies used self-report outcome measures and mentioned concerns with the psychometrics of the measures used. Of the 14 of 42 (32%) papers that mentioned or directly measured resilience, only 5 provided an operational definition of resilience. Features of study design were reported as a limitation by 57% (22/39) of studies with concerns regarding lack of random assignment, blinding, control group, and control for confounding variables. A number of studies (11/39, 29%) also made a call for future research to increase the duration of intervention and follow-up period. Finally, many studies reported concerns regarding app features, app accessibility, and study design.

Discussion

Principal Findings

This study reviewed 44 papers and 39 DMHIs. We identified 9 of the same apps compared with the review by Voth et al [29] and 30 new apps and web-based programs compared with their search. Out of the 39 DMHIs, 16 are web-based programs, 22 are apps, and 1 is available as an app and web-based program. Notably, 21 new papers have been published since 2020, highlighting the growth of apps and web-based programs for these populations and justifying the need for this updated review. The majority of the studies took place in the United States and were published between 2020 and 2024. Most studies recruited veteran populations, suggesting a need for apps and programs for MMs and PSP populations.

I-COPPE and Cascading Resilience

The DMHIs included in this study were organized by the authors based on domains of well-being perceived to be supported (ie, the I-COPPE model). In total, 97% (38/39) of the DMHIs supported psychological well-being. This is not surprising, given that resilience and well-being are often conceptualized as the lack of mental health symptomatology or diagnosis [104]. Although psychological well-being is defined as satisfaction with one's emotional life [58], this literature tended to focus on decreasing distal outcomes (eg, reducing symptoms) as opposed to promoting proximal factors (eg, self-efficacy or coping to remediate symptoms). Targeting proximal processes may instigate positive cascades across systems [37]. For example, Pavlacic et al [105] found that MMs and veteran populations with higher coping behaviors and self-efficacy were more likely to exhibit a resilience response, emphasizing that targeting proximal factors may be an effective intervention approach.

It is also notable that there were a high number of apps and web-based programs that supported physical (21/39, 54%) and interpersonal (21/39, 54%) well-being. Both well-being domains are key for functioning well, feeling good, and exhibiting a resilience response [105,106]. Pavlacic et al [105] found that

XSL•FO
RenderX

MMs and veterans who reported alcohol problems, poorer physical health, smoking, and presence of a sleep disorder were less likely to exhibit a resilience trajectory and were more likely to exhibit symptomatology. Conversely, veterans with lower self-reported physical health difficulties were more likely to exhibit a resilient response [107]. These results suggest that physical well-being is closely related to resilience and an imperative area to target in a holistic approach to intervention. In terms of interpersonal well-being, MMs and veterans who reported decreased societal exclusion at home and increased social support were more likely to exhibit a resilience response [105]. In a similar study with veterans, Pietrzak and Cook [107] found that individuals with positive reports of social connectedness and social engagement were more likely to respond in a resilient way. Therefore, by targeting physical and interpersonal well-being, there are likely to be positive well-being and resilient responses, as well as potential positive spillover effects across other domains of well-being.

A limitation of the DMHI research, from a whole-person multisystemic resilience framework, is that community, occupational, and economic well-being were each emphasized by 4 or fewer programs. Clearly, greater consideration for community, economic, and organizational well-being is needed [48,58], especially as enhanced well-being across these domains may lead to increased likelihood of a resilience response [58] and positive spillover effects, such as improved resilience across systems [40,108]. Individuals with higher levels of protective psychosocial characteristics related to community well-being and overall well-being (eg, purpose in life), and positive perceptions of the military's effect on one's life (ie, occupational well-being) are more likely to exhibit a resilient response when faced with a number of PPTEs [107]. Economic well-being is posited to be closely related to satisfaction with one's financial situation and is closely related to physical and psychological health [58]. Therefore, by supporting one's economic well-being, there are likely spillover effects into psychological and physical well-being, which likely also have positive impacts on one's ability to positively engage in work, with the community, and with important people in their lives.

Critical Analysis of DMHIs: Strengths and Weaknesses

Most studies (40/44, 91%) specified a therapeutic orientation, with 51% (20/44) purporting a cognitive behavioral framework. Problematically, 9% (4/44) of studies did not provide a therapeutic approach or framework for their intervention. It is important to frame an intervention within a therapeutic framework or approach as this acts as a road map to understand presenting problems and how potential treatments may present solutions to these problems [109]. Not including a therapeutic framework may introduce an array of pitfalls in terms of implementation and evaluating program effectiveness [109]. The most common presenting problems in the populations were PTSD symptoms and substance use behaviors, and the most commonly targeted proximal and distal factors were PTSD, depression, and resilience. Most DMHIs were evaluated in terms of symptom reduction, and few were evaluated in terms of their ability to improve protective factors, indicative of the proclivity toward a diagnostic approach. This emphasis on pathology over resilience-supporting processes is problematic [110]. By evaluating processes, researchers may have an improved understanding of the extent to which DMHIs support well-being and subsequent resilience cascades. Future research would benefit from a process-based orientation and longitudinal methodologies to investigate the process of DMHIs on individual resilience trajectories [110-113]. A longitudinal approach can provide insights into healthy trajectories [112] and illuminate whether and how certain protective factors impact resilience trajectories [104].

An area for growth in the DMHI research is the need for RCTs. These rigorously controlled studies are necessary as they evaluate the impact of interventions with high levels of internal validity and determine the efficacy of the intervention [114,115]. In the current review, only 30% (13/44) of the studies were RCTs and 20% (9/44) were pilot RCTs (most with evaluations of feasibility, acceptability, satisfaction, or qualitative feedback). The pilot RCTs are promising such that they may lead to full efficacy trials, and the initial results illuminate some positive results. More specifically, there were 8 studies that reported positive efficacy results, including Concussion Coach (decreased postconcussive symptoms), Family Foundations (decreased parent depression), Mission Reconnect (improved sleep quality and response to stressful experience), Resilience@Work (improved coping behaviors), Support Coach (increased psychological resilience), Thinking Forward (decreased perceived alcohol consumption), VetChange (improved drinking behaviors and PTSD symptoms), and Virtual Hope Box (improved coping with unpleasant thoughts or emotions). Despite methodological strengths, RCTs' prioritization of internal validity often comes at the expense of external validity (ie, generalizability to real-world clinical settings) [114]. Our review found that 14% (6/44) of studies prioritized external validity and evaluated their intervention via quasi-experimental methods. Further research in this area can provide valuable insights into real-world effectiveness and application of resilience and well-being DMHIs for MMs, PSP, and veterans. Given the high use of apps and programs, it is crucial to continue evaluating their efficacy and effectiveness, particularly prior to public use [32,115].

In addition, only 14% (6/44) of the studies were focused on feasibility or acceptability. Attention to these feasibility and acceptability issues is critical as these populations often do not engage with in-person mental health services [12,13] and experience a number of barriers to accessing services [13-17]. In addition, the high attrition rates, a notable limitation across the DMHI literature, emphasize the need for further investigation into acceptability, feasibility, and satisfaction with resilience and well-being DMHIs. For example, van Stolk-Cook et al [82] found that PTSD Family Coach 1.0 users opened their apps an average of 2.38 (week 1), 0.45 (week 2), 0.14 (week 3), and 0.22 (week 4) times across their study, and Parkes et al [83] found that participants used their app for a duration of 20.7 seconds (median). These examples are common throughout the DMHI research and provide valuable insights and a vital starting point for ongoing research.

A majority of the included research had participants use the app or program without additional support, but there were 9 interventions supplemented by clinician support, peer support,

XSL•FO
RenderX

research staff support, mental health specialist support, feedback, or clinical monitoring [23,71-77]. There were no significant differences between groups on outcome measures across the 5 studies that compared clinician support versus no support [75], with or without treatment-as-usual [71-73], check-in versus self-managed groups [77], and peer support versus self-managed groups [76]. The only exception, however, is that Possemato et al [75] reported that those in the clinician-supported group were more likely to attend an additional PTSD session and accept a treatment referral. Therefore, these studies provide evidence that DMHIs may be just as effective without support; however, this may not be the case when compared with other treatment modalities. For example, Liu et al [32] found that self-guided resilience interventions appear to have little or no effect on increasing resilience, whereas there is a small, meaningful effect for in-person, remote, individual, and group interventions. Therefore, further research is needed to evaluate the effectiveness and efficacy of DMHIs compared with additional support and control conditions.

A potential weakness is that this literature is in its infancy for evaluating potential factors impacting intervention effectiveness. Many of these evaluations had small sample sizes and high attrition rates, posing challenges for detecting true mediation or moderation [116,117] and limiting both the internal and external validity of the results [118]. Although this was the case for many studies, 4 authors were able to successfully evaluate and find evidence for moderator variables, including parent gender [61], PTSD symptoms [79], combat exposure [79], self-efficacy [65], and interpersonal problems [80]. One study reported three significant mediators, including perceived norms of same-age peers (1) number of drinking occasions, (2) number of drinks, (3) and number of drinking occasions [81]. These variables, along with other potential moderators and mediators, should continue to be evaluated. This is an essential step in intervention research as it is necessary to explore potential factors that underlie or influence the association between an intervention and a desired outcome [119].

Implications and Key Takeaways

This scoping review aimed to look at the breadth of literature for accessible and available DMHIs for MMs, PSP, and veterans, who are individuals who are regularly underserved and often do not access or have access to formal support. A key gap illuminated in this review is the lack of DMHIs developed and validated for MMs and PSP populations specifically. In addition, the vast majority of DMHIs focused on psychological well-being, while only a few focused on community, economic, and occupational well-being: important domains for one's overall health, well-being, and resilience. Since 2020, there has been an increase in published literature on DMHIs for these populations; however, many DMHIs are available to these underserved populations before they have been empirically validated. For example, of the 44 papers included in this review, 13 were examined via RCT, and 4 via quasi-experimental design; therefore, additional evaluations of DMHIs for MMs, PSP, and veteran populations must be completed to examine the internal and external validity of these interventions.

There were 8 DMHIs for these populations that showed positive results via RCT and, therefore, may be recommended for these populations, including Concussion Coach, Family Foundations, Mission Reconnect, Resilience@Work, Support Coach, Thinking Forward, VetChange, and Virtual Hope Box. Although in clinical practice, DMHIs may be recommended as adjunct support between sessions, based on this review, it appears that some DMHIs are supported regardless of additional support, including PTSD Coach, Virtual Hope Box, and Thinking Forward. An additional consideration for future research and DMHI development includes the moderators that were found to have a significant impact on MMs, PSP, and veteran outcomes, such as baseline PTSD symptoms, combat exposure, self-efficacy, and interpersonal problems, as these factors may influence the effectiveness of the DMHI.

Limitations and Strengths of the Review

Some methodological limitations exist in the current scoping review. Only English studies were included, and a gray literature search and snowball sampling were not used, potentially impacting the scope of this review. A potential limitation across this literature is defining *web-based programs*. Efforts by the research team to differentiate web-based programs from web-based courses, module-based courses or programs, chat rooms, Facebook groups, YouTube channels, and resource banks lead to determination of web-based programs as *interactive online platforms* that do not involve an invitation or log-in code provided by the DMHI developer. It would be beneficial, in the future, to have a clear consensus on the definition and further differentiate the aforementioned web-based resources.

This scoping review also has several strengths. The project used the PRISMA-ScR to ensure the study's quality and minimize potential bias and included all search information to facilitate replication. In addition, a detailed, rigorous, and extensive search was conducted of 6 databases with additional search terms expanding upon previous reviews. At least 2 independent reviewers were engaged in the review, minimizing risk of bias. The study's inclusion and exclusion criteria were established upfront and refined throughout. The scoping review also contributes to the literature by summarizing key information surrounding the quality of the evidence base for DMHIs for MMs, PSP, and veterans. This type of information is paramount as it is key to informing these populations and clinicians in terms of the best available DMHIs based on the empirical evidence, which is key to evidence-based practice and clinical decision-making [92]. This type of review and information is also key to providing app developers and researchers with information for next steps in terms of development, implementation, and research. Finally, the current project applied a whole-person approach by evaluating programs through the I-COPPE model [58] and cascading resilience framework [37]. These frameworks prioritize multidimensional and systems-based understanding of one's well-being, allowing researchers and clinicians to focus on protective networks, rather than deficit-based frameworks.

Future research can build on the current project. First, it is important for future research to review DMHIs that involve synchronous or professional, peer, or researcher support. Such

XSL•FO
RenderX

research may glean important information about the efficacy or effectiveness of resilience and well-being with guidance and support. In addition, it may clarify whether independent use of DMHIs is comparable with supported use of DMHIs. Second, future research should consider the acceptability of DMHIs across these populations. Although the current review grouped these populations (MMs, PSP, and veterans), we cannot speak to whether a DMHI developed for one population generalizes or is as effective for the other populations. Therefore, further research is needed to explore the acceptability of interventions across these populations and to explore ways in which DMHIs can be contextualized appropriately for each population.

Conclusions

DMHIs have the potential to promote resilience and well-being in PSP, MMs, and veterans. This scoping review summarizes the state of the literature and surveys available DMHI programs and apps. Several themes related to DMHI literature were identified, together with intervention characteristics including the target populations, purpose, therapeutic approach, targeted symptoms, therapeutic modalities, evidence and methodologies used, and domains of well-being supported by identified programs and apps. More rigorous research is needed, however, to examine the effectiveness and efficacy, acceptability, feasibility, and satisfaction of DMHIs for MMs, PSP, and veterans. While the determination of DMHIs as an evidence-based alternative to in-person mental health care requires more research, the accessibility, customization, and scalability of DMHIs make this mode of delivery promising.

Acknowledgments

This research has been funded by the Government of Alberta, Mental Health and Addiction COVID-19 Community Funding Grant.

Authors' Contributions

RRA maintained a leadership role in developing the current project and collaborated with the remaining authors in the conceptualization of the manuscript. RRA took the lead, and MAM, CA, and LH assisted in writing the first draft of the manuscript. RRA, MAM, and CA were involved in the data search, review, and eligibility assessment. SB-P and PRS took leadership roles in reviewing and editing the manuscript. RRA amalgamated edit suggestions to write the final draft. All authors significantly contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string example.
[DOCX File, 18 KB - mhealth_v13i1e65149_app1.docx]

Multimedia Appendix 2
App and web-based program therapeutic approaches and modes of delivery.
[DOCX File, 28 KB - mhealth_v13i1e65149_app2.docx]

Multimedia Appendix 3
Well-being domains supported by each program.
[DOCX File, 24 KB - mhealth_v13i1e65149_app3.docx]

Multimedia Appendix 4
Summary of app and web-based program results.
[DOCX File, 49 KB - mhealth_v13i1e65149_app4.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 521 KB - mhealth_v13i1e65149_app5.pdf]

References

1. Standing Committee on Public Safety and National Security. Healthy minds, safe communities: supporting our public safety officers through a national strategy for operational stress injuries. : Parliament of Canada; 2016 URL: https://www.ourcommons.ca/Content/Committee/421/SECU/Reports/RP8457704/securp05/securp05-e.pdf

XSL•FO
RenderX

2. Carleton RN, Afifi TO, Taillieu T, et al. Exposures to potentially traumatic events among public safety personnel in Canada. Can J Behav Sci 2019;51(1):37-52. [doi: 10.1037/cbs0000115]
3. Kilpatrick DG, Resnick HS, Milanak ME, Miller MW, Keyes KM, Friedman MJ. National estimates of exposure to traumatic events and PTSD prevalence using DSM-IV and DSM-5 criteria. J Trauma Stress 2013;26(5):537-547. [doi: 10.1002/jts.21848]
4. Easterbrook B, Brown A, Millman H, et al. Original qualitative research—the mental health experience of treatment-seeking military members and public safety personnel: a qualitative investigation of trauma and non-trauma-related concerns. Public Health Agency Can 2022(42):252-260. [doi: 10.24095/hpcdp.42.6.03]
5. Born JA, Zamorski MA. Contribution of traumatic deployment experiences to the burden of mental health problems in Canadian Armed Forces personnel: exploration of population attributable fractions. Soc Psychiatry Psychiatr Epidemiol 2019 Feb;54(2):145-156. [doi: 10.1007/s00127-018-1562-6]
6. Thompson JM, Dursun S, VanTil L, et al. Group identity, difficult adjustment to civilian life, and suicidal ideation in Canadian Armed Forces Veterans: life after service studies 2016. J Mil Veteran Fam Health 2019 Sep 1;5(2):100-114. [doi: 10.3138/jmvfh.2018-0038]
7. Carleton RN, Afifi TO, Turner S, et al. Mental disorder symptoms among public safety personnel in Canada. Can J Psychiatry 2018 Jan;63(1):54-64. [doi: 10.1177/0706743717723825]
8. Milligan-Saville J, Choi I, Deady M, et al. The impact of trauma exposure on the development of PTSD and psychological distress in a volunteer fire service. Psychiatry Res 2018 Dec;270:1110-1115. [doi: 10.1016/j.psychres.2018.06.058] [Medline: 30190167]
9. Na PJ, Schnurr PP, Pietrzak RH. Mental health of U.S. combat veterans by war era: results from the National Health and Resilience in Veterans Study. J Psychiatr Res 2023 Feb;158:36-40. [doi: 10.1016/j.jpsychires.2022.12.019] [Medline: 36565542]
10. Carleton RN, Afifi TO, Taillieu T, et al. Assessing the relative impact of diverse stressors among public safety personnel. Int J Environ Res Public Health 2020 Feb 14;17(4):1234. [doi: 10.3390/ijerph17041234] [Medline: 32075062]
11. Clever M, Segal DR. The demographics of military children and families. Future Child 2013 Sep;23(2):13-39. [doi: 10.1353/foc.2013.0018]
12. Hom MA, Stanley IH, Schneider ME, Joiner TE. A systematic review of help-seeking and mental health service utilization among military service members. Clin Psychol Rev 2017 Apr;53:59-78. [doi: 10.1016/j.cpr.2017.01.008]
13. Kline AC, Panza KE, Nichter B, et al. Mental health care use among U.S. military veterans: results from the 2019–2020 National Health and Resilience in Veterans Study. Psychiatr Serv 2022 Jun 1;73(6):628-635. [doi: 10.1176/appi.ps.202100112]
14. Blais RK, Tsai J, Southwick SM, Pietrzak RH. Barriers and facilitators related to mental health care use among older veterans in the United States. Psychiatr Serv 2015 May 1;66(5):500-506. [doi: 10.1176/appi.ps.201300469]
15. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL. Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N Engl J Med 2004 Jul 1;351(1):13-22. [doi: 10.1056/NEJMoa040603] [Medline: 15229303]
16. Stecker T, Shiner B, Watts BV, Jones M, Conner KR. Treatment-seeking barriers for veterans of the Iraq and Afghanistan conflicts who screen positive for PTSD. Psychiatr Serv 2013;64:280-283. [doi: 10.1176/appi.ps.001372012] [Medline: 23450385]
17. Pietrzak RH, Johnson DC, Goldstein MB, Malley JC, Southwick SM. Perceived stigma and barriers to mental health care utilization among OEF-OIF veterans. Psychiatr Serv 2009 Aug;60(8):1118-1122. [doi: 10.1176/ps.2009.60.8.1118] [Medline: 19648201]
18. Gaebel W, Stricker J. E-mental health options in the COVID-19 pandemic and beyond. Psychiatry Clin Neurosci 2020;74(8):441-442. [doi: 10.1111/pcn.13079]
19. Khanna R, Forbes M. Telepsychiatry as a public health imperative: slowing COVID-19. Aust N Z J Psychiatry 2020 Jul;54(7):758. [doi: 10.1177/0004867420924480]
20. Moreno C, Wykes T, Galderisi S, et al. How mental health care should change as a consequence of the COVID-19 pandemic. Lancet Psychiatry 2020 Sep;7(9):813-824. [doi: 10.1016/S2215-0366(20)30307-2]
21. Serafini G, Parmigiani B, Amerio A, Aguglia A, Sher L, Amore M. The psychological impact of COVID-19 on the mental health in the general population. QJM 2020 Aug 1;113(8):531-537. [doi: 10.1093/qjmed/hcaa201]
22. Gruber J, Prinstein MJ, Clark LA, et al. Mental health and clinical psychological science in the time of COVID-19: challenges, opportunities, and a call to action. Am Psychol 2021 Apr;76(3):409-426. [doi: 10.1037/amp0000707] [Medline: 32772538]
23. Leightley D, Puddephatt JA, Jones N, et al. A smartphone app and personalized text messaging framework (InDEx) to monitor and reduce alcohol use in ex-serving personnel: development and feasibility study. JMIR Mhealth Uhealth 2018 Sep 11;6(9):e10074. [doi: 10.2196/10074] [Medline: 30206054]
24. Cucciare MA, Coleman J, Reid PV. Exploring the viability of telehealth for providing clinical services to populations with high rates of mental health concerns. J Telemed Telecare 2018;24(7):465-470. [doi: 10.1177/1357633X17732581]
25. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]

XSL•FO
RenderX

26. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247. [doi: 10.2196/jmir.2791] [Medline: 24240579]
27. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101. [doi: 10.2196/mhealth.4924] [Medline: 26537656]
28. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146. [doi: 10.2196/jmir.3077] [Medline: 24905070]
29. Voth M, Chisholm S, Sollid H, Jones C, Smith-MacDonald L, Brémault-Phillips S. Efficacy, effectiveness, and quality of resilience-building mobile health apps for military, veteran, and public safety personnel populations: scoping literature review and app evaluation. JMIR Mhealth Uhealth 2022 Jan 19;10(1):e26453. [doi: 10.2196/26453] [Medline: 35044307]
30. Tam-Seto L, Wood VM, Linden B, Stuart H. A scoping review of mental health mobile apps for use by the military community. Mhealth 2018;4:57. [doi: 10.21037/mhealth.2018.12.01] [Medline: 30701175]
31. Khoshrounejad F, Hamednia M, Mehrjerd A, et al. Telehealth-based services during the COVID-19 pandemic: a systematic review of features and challenges. Front Public Health 2021;9:711762. [doi: 10.3389/fpubh.2021.711762] [Medline: 34350154]
32. Liu JJW, Ein N, Gervasio J, Battaion M, Fung K. The pursuit of resilience: a meta-analysis and systematic review of resilience-promoting interventions. J Happiness Stud 2022 Apr;23(4):1771-1791. [doi: 10.1007/s10902-021-00452-8]
33. Skorburg JA, Yam J. Is there an app for that?: ethical issues in the digital mental health response to COVID-19. AJOB Neurosci 2022;13(3):177-190. [doi: 10.1080/21507740.2021.1918284] [Medline: 33989127]
34. Goldberg SB, Lam SU, Simonsson O, Torous J, Sun S. Mobile phone-based interventions for mental health: a systematic meta-review of 14 meta-analyses of randomized controlled trials. PLOS Digit Health 2022;1(1):e0000002. [doi: 10.1371/journal.pdig.0000002] [Medline: 35224559]
35. Bonanno GA, Romero SA, Klein SI. The temporal elements of psychological resilience: an integrative framework for the study of individuals, families, and communities. Psychol Inq 2015 Apr 3;26(2):139-169. [doi: 10.1080/1047840X.2015.992677]
36. Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales. Health Qual Life Outcomes 2011 Feb 4;9(1):8. [doi: 10.1186/1477-7525-9-8] [Medline: 21294858]
37. Doty JL, Davis L, Arditti JA. Cascading resilience: leverage points in promoting parent and child well‐being. J Fam Theory Rev 2017 Mar;9(1):111-126. [doi: 10.1111/jftr.12175]
38. Patterson GR, Forgatch MS, Degarmo DS. Cascading effects following intervention. Dev Psychopathol 2010 Nov;22(4):949-970. [doi: 10.1017/S0954579410000568] [Medline: 20883592]
39. Walsh F. Family resilience: a dynamic systemic framework. In: Multisystemic Resilience: Adaptation and Transformation in Contexts of Change: Oxford University Press; 2021:255-270. [doi: 10.1093/oso/9780190095888.003.0015]
40. Ungar M. Modeling multisystemic resilience: connecting biological, psychological, social, and ecological adaptation in contexts of adversity. In: Ungar M, editor. Multisystemic Resilience: Adaptation and Transformation in Contexts of Change: Oxford University Press; 2021:6-31.
41. Ungar M. The social ecology of resilience: addressing contextual and cultural ambiguity of a nascent construct. Am J Orthopsychiatry 2011 Jan;81(1):1-17. [doi: 10.1111/j.1939-0025.2010.01067.x] [Medline: 21219271]
42. Ungar M. Resilience, trauma, context, and culture. Trauma Violence Abuse 2013 Jul;14(3):255-266. [doi: 10.1177/1524838013487805]
43. Luthar SS, Cicchetti D, Becker B. The construct of resilience: a critical evaluation and guidelines for future work. Child Dev 2000;71(3):543-562. [doi: 10.1111/1467-8624.00164] [Medline: 10953923]
44. de Vries LP, Baselmans BML, Luykx JJ, et al. Genetic evidence for a large overlap and potential bidirectional causal effects between resilience and well-being. Neurobiol Stress 2021 May;14:100315. [doi: 10.1016/j.ynstr.2021.100315] [Medline: 33816719]
45. Wild J, Smith KV, Thompson E, Béar F, Lommen MJJ, Ehlers A. A prospective study of pre-trauma risk factors for post-traumatic stress disorder and depression. Psychol Med 2016 Sep;46(12):2571-2582. [doi: 10.1017/S0033291716000532] [Medline: 27348599]
46. Keyes CLM, Annas J. Feeling good and functioning well: distinctive concepts in ancient philosophy and contemporary science. J Positive Psychol 2009 May 1;4(3):197-201. [doi: 10.1080/17439760902844228]
47. Keyes CLM. The mental health continuum: from languishing to flourishing in life. J Health Soc Behav 2002 Jun;43(2):207-222. [doi: 10.2307/3090197] [Medline: 12096700]
48. Prilleltensky I. Wellness as fairness. Am J Community Psychol 2012 Mar;49(1-2):1-21. [doi: 10.1007/s10464-011-9448-8] [Medline: 21643926]
49. Prilleltensky I. Promoting well-being: time for a paradigm shift in health and human services. Scand J Public Health 2005 Oct;33(Suppl 66):53-60. [doi: 10.1080/14034950510033381]
50. Luthar SS, Cicchetti D. The construct of resilience: implications for interventions and social policies. Dev Psychopathol 2000 Dec;12(4):857-885. [doi: 10.1017/S0954579400004156]

XSL•FO
RenderX

51. Bonanno GA, Brewin CR, Kaniasty K, Greca AML. Weighing the costs of disaster: consequences, risks, and resilience in individuals, families, and communities. Psychol Sci Public Interest 2010 Jan;11(1):1-49. [doi: 10.1177/1529100610387086] [Medline: 26168411]
52. Bonanno GA, Galea S, Bucciarelli A, Vlahov D. What predicts psychological resilience after disaster? The role of demographics, resources, and life stress. J Consult Clin Psychol 2007 Oct;75(5):671-682. [doi: 10.1037/0022-006X.75.5.671] [Medline: 17907849]
53. Stainton A, Chisholm K, Kaiser N, et al. Resilience as a multimodal dynamic process. Early Interv Psychiatry 2019 Aug;13(4):725-732. [doi: 10.1111/eip.12726] [Medline: 30126047]
54. McNally RJ, Bryant RA, Ehlers A. Does early psychological intervention promote recovery from posttraumatic stress? Psychol Sci Public Interest 2003 Nov;4(2):45-79. [doi: 10.1111/1529-1006.01421] [Medline: 26151755]
55. Litz BT, Gray MJ, Bryant RA, Adler AB. Early intervention for trauma: current status and future directions. Clin Psychol Sci Pract 2002;9(2):112-134. [doi: 10.1093/clipsy.9.2.112]
56. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 4;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
57. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr;395(10231):1180-1181. [doi: 10.1016/S0140-6736(20)30818-7]
58. Prilleltensky I, Dietz S, Prilleltensky O, et al. Assessing multidimensional well-being: development and validation of the I COPPE Scale. J Community Psychol 2015 Mar;43(2):199-226. [doi: 10.1002/jcop.21674]
59. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
60. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
61. Feinberg ME, Boring J, Le Y, et al. Supporting military family resilience at the transition to parenthood: a randomized pilot trial of an online version of Family Foundations. Fam Relat 2020 Feb;69(1):109-124. [doi: 10.1111/fare.12415]
62. Vaughan AD, Stoliker BE, Anderson GS. Building personal resilience in primary care paramedic students, and subsequent skill decay. Australas J Paramed 2020 Jan;17:1-8. [doi: 10.33151/ajp.17.803]
63. Kroenke K, Baye F, Lourens SG, et al. Automated self-management (ASM) vs. ASM-enhanced collaborative care for chronic pain and mood symptoms: the CAMMPS randomized clinical trial. J Gen Intern Med 2019 Sep;34(9):1806-1814. [doi: 10.1007/s11606-019-05121-4] [Medline: 31228055]
64. Albright G, Goldman R, Shockley KM, McDevitt F, Akabas S. Using an avatar-based simulation to train families to motivate veterans with post-deployment stress to seek help at the VA. Games Health J 2012 Feb;1(1):21-28. [doi: 10.1089/g4h.2011.0003] [Medline: 26196429]
65. Belanger HG, Toyinbo P, Barrett B, King E, Sayer NA. Concussion coach for postconcussive symptoms: a randomized, controlled trial of a smartphone application with Afghanistan and Iraq war Veterans. Clin Neuropsychol 2022;36(8):2093-2119. [doi: 10.1080/13854046.2021] [Medline: 34184976]
66. Kahn JR, Collinge W, Soltysik R. Post-9/11 veterans and their partners improve mental health outcomes with a self-directed mobile and web-based wellness training program: a randomized controlled trial. J Med Internet Res 2016 Sep 27;18(9):e255. [doi: 10.2196/jmir.5800] [Medline: 27678169]
67. Joyce S, Shand F, Lal TJ, Mott B, Bryant RA, Harvey SB. Resilience@Work Mindfulness Program: results from a cluster randomized controlled trial with first responders. J Med Internet Res 2019 Feb 19;21(2):e12894. [doi: 10.2196/12894] [Medline: 30777846]
68. van der Meer CAI, Bakker A, van Zuiden M, Lok A, Olff M. Help in hand after traumatic events: a randomized controlled trial in health care professionals on the efficacy, usability, and user satisfaction of a self-help app to reduce trauma-related symptoms. Eur J Psychotraumatol 2020 Dec 31;11(1):1717155. [doi: 10.1080/20008198.2020.1717155]
69. Acosta MC, Possemato K, Maisto SA, et al. Web-delivered CBT reduces heavy drinking in OEF-OIF veterans in primary care with symptomatic substance use and PTSD. Behav Ther 2017 Mar;48(2):262-276. [doi: 10.1016/j.beth.2016.09.001] [Medline: 28270335]
70. Brief DJ, Rubin A, Keane TM, et al. Web intervention for OEF/OIF veterans with problem drinking and PTSD symptoms: a randomized clinical trial. J Consult Clin Psychol 2013;81(5):890-900. [doi: 10.1037/a0033697]
71. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A virtual hope box: randomized controlled trial of a smartphone app for emotional regulation and coping with distress. Psychiatr Serv 2017 Apr 1;68(4):330-336. [doi: 10.1176/appi.ps.201600283] [Medline: 27842473]
72. Bush NE, Ouellette G, Kinn J. Utility of the T2 Mood Tracker mobile application among army warrior transition unit service members. Mil Med 2014 Dec;179(12):1453-1457. [doi: 10.7205/MILMED-D-14-00271] [Medline: 25469967]
73. Denneson LM, Smolenski DJ, Bauer BW, Dobscha SK, Bush NE. The mediating role of coping self-efficacy in Hope Box use and suicidal ideation severity. Arch Suicide Res 2019;23(2):234-246. [doi: 10.1080/13811118.2018.1456383] [Medline: 29624123]

XSL•FO
RenderX

74. Bush NE, Dobscha SK, Crumpton R, et al. A Virtual Hope Box smartphone app as an accessory to therapy: proof-of-concept in a clinical sample of veterans. Suicide Life Threat Behav 2015 Feb;45(1):1-9. [doi: 10.1111/sltb.12103] [Medline: 24828126]
75. Possemato K, Kuhn E, Johnson E, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
76. Possemato K, Johnson EM, Emery JB, et al. A pilot study comparing peer supported web-based CBT to self-managed web CBT for primary care veterans with PTSD and hazardous alcohol use. Psychiatr Rehabil J 2019 Sep;42(3):305-313. [doi: 10.1037/prj0000334] [Medline: 30489140]
77. McLean C, Davis CA, Miller M, Ruzek J, Neri E. The effects of an exposure-based mobile app on symptoms of posttraumatic stress disorder in veterans: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Nov 4;10(11):e38951. [doi: 10.2196/38951] [Medline: 36331540]
78. Buck B, Nguyen J, Porter S, Ben-Zeev D, Reger GM. FOCUS mHealth intervention for veterans with serious mental illness in an outpatient department of veterans affairs setting: feasibility, acceptability, and usability study. JMIR Ment Health 2022 Jan 28;9(1):e26049. [doi: 10.2196/26049] [Medline: 35089151]
79. Brief DJ, Solhan M, Rybin D, et al. Web-based alcohol intervention for veterans: PTSD, combat exposure, and alcohol outcomes. Psychol Trauma 2018 Mar;10(2):154-162. [doi: 10.1037/tra0000281] [Medline: 28569525]
80. Polizzi CP, Sistad RE, Livingston NA, et al. Alcohol-related problems as moderators of PTSD symptom change during use of a web-based intervention for hazardous drinking and PTSD. J Stud Alcohol Drugs 2024 Jan;85(1):51-61. [doi: 10.15288/jsad.23-00094] [Medline: 37796630]
81. Williams J, Herman-Stahl M, Calvin SL, Pemberton M, Bradshaw M. Mediating mechanisms of a military Web-based alcohol intervention. Drug Alcohol Depend 2009 Mar 1;100(3):248-257. [doi: 10.1016/j.drugalcdep.2008.10.007] [Medline: 19081206]
82. van Stolk-Cooke K, Wielgosz J, Hallenbeck HW, et al. The PTSD Family Coach app in veteran family members: pilot randomized controlled trial. JMIR Form Res 2023;7:e42053. [doi: 10.2196/42053]
83. Parkes S, Croak B, Brooks SK, et al. Evaluating a smartphone app (MeT4VeT) to support the mental health of UK armed forces veterans: feasibility randomized controlled trial. JMIR Ment Health 2023 Aug 28;10:e46508. [doi: 10.2196/46508] [Medline: 37639295]
84. Babson KA, Ramo DE, Baldini L, Vandrey R, Bonn-Miller MO. Mobile app-delivered cognitive behavioral therapy for insomnia: feasibility and initial efficacy among veterans with cannabis use disorders. JMIR Res Protoc 2015 Jul 17;4(3):e87. [doi: 10.2196/resprot.3852] [Medline: 26187404]
85. Bush NE, Prins A, Laraway S, O'Brien K, Ruzek J, Ciulla RP. A pilot evaluation of the AfterDeployment.org online posttraumatic stress workshop for military service members and veterans. Psychol Trauma 2014;6(2):109-119. [doi: 10.1037/a0032179]
86. Chen JI, Smolenski DJ, Dobscha SK, Bush NE, Denneson LM. Correlates of mental health smartphone application use among patients with suicidal ideation. J Technol Hum Serv 2018 Oct 2;36(4):191-207. [doi: 10.1080/15228835.2018.1502119]
87. Coifman KG, Disabato DD, Seah THS, et al. Boosting positive mood in medical and emergency personnel during the COVID-19 pandemic: preliminary evidence of efficacy, feasibility and acceptability of a novel online ambulatory intervention. Occup Environ Med 2021 Apr 26;doi:oemed-2021-107427. [doi: 10.1136/oemed-2021-107427] [Medline: 33903280]
88. Davis JP, Pedersen ER, Borsari B, et al. Development of a mobile mindfulness smartphone app for post-traumatic stress disorder and alcohol use problems for veterans: beta test results and study protocol for a pilot randomized controlled trial. Contemp Clin Trials 2023 Jun;129:107181. [doi: 10.1016/j.cct.2023.107181] [Medline: 37059261]
89. Dillon KH, Hertzberg JA, Mosher TM, et al. Development and refinement of the mobile anger reduction intervention for veterans with posttraumatic stress disorder. Psychol Trauma 2024 Dec;16(Suppl 3):S658-S667. [doi: 10.1037/tra0001491] [Medline: 37166919]
90. Engel CC, Litz B, Magruder KM, et al. Delivery of self training and education for stressful situations (DESTRESS-PC): a randomized trial of nurse assisted online self-management for PTSD in primary care. Gen Hosp Psychiatry 2015;37(4):323-328. [doi: 10.1016/j.genhosppsych.2015.04.007] [Medline: 25929985]
91. Heyen JM, Weigl N, Müller M, et al. Multimodule web-based COVID-19 anxiety and stress resilience training (COAST): single-cohort feasibility study with first responders. JMIR Form Res 2021 Jun 7;5(6):e28055. [doi: 10.2196/28055] [Medline: 33999835]
92. Higgins DM, Buta E, Williams DA, et al. Internet-based pain self-management for veterans: feasibility and preliminary efficacy of the Pain EASE Program. Pain Pract 2020 Apr;20(4):357-370. [doi: 10.1111/papr.12861] [Medline: 31778281]
93. Hofmann L, Glaesmer H, Przyrembel M, Wagner B. An evaluation of a suicide prevention e-learning program for police officers (COPS): improvement in knowledge and competence. Front Psychiatry 2021;12:770277. [doi: 10.3389/fpsyt.2021.770277] [Medline: 34966307]
94. Ivory RA. Attenuating Chronic Pain & Trauma Among Naval Special Warfare Veterans Using Psychoeducation. Doctoral Dissertation.: University of Delaware; 2023.

XSL•FO
RenderX

95. Johnson SS, Levesque DA, Broderick LE, Bailey DG, Kerns RD. Pain self-management for veterans: development and pilot test of a stage-based mobile-optimized intervention. JMIR Med Inform 2017 Oct 17;5(4):e40. [doi: 10.2196/medinform.7117] [Medline: 29042341]
96. Joyce S, Shand F, Tighe J, Laurent SJ, Bryant RA, Harvey SB. Road to resilience: a systematic review and meta-analysis of resilience training programmes and interventions. BMJ Open 2018 Jun 14;8(6):e017858. [doi: 10.1136/bmjopen-2017-017858] [Medline: 29903782]
97. Kuhn E, Miller KE, Puran D, et al. A pilot randomized controlled trial of the Insomnia Coach mobile app to assess its feasibility, acceptability, and potential efficacy. Behav Ther 2022 May;53(3):440-457. [doi: 10.1016/j.beth.2021.11.003] [Medline: 35473648]
98. Kuhn E, Greene C, Hoffman J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]
99. Livingston NA, Mahoney CT, Ameral V, et al. Changes in alcohol use, PTSD hyperarousal symptoms, and intervention dropout following veterans' use of VetChange. Addict Behav 2020 Aug;107:106401. [doi: 10.1016/j.addbeh.2020.106401] [Medline: 32272356]
100. Newberger NG, Yeager S, Livingston NA, et al. Life satisfaction following treatment-related reductions in alcohol use and PTSD symptoms: results from VetChange. Psychol Trauma 2023 Nov;15(8):1299-1306. [doi: 10.1037/tra0001234] [Medline: 35254847]
101. Possemato K, Acosta MC, Fuentes J, et al. A web-based self-management program for recent combat veterans with PTSD and substance misuse: program development and veteran feedback. Cogn Behav Pract 2015 Aug 1;22(3):345-358. [doi: 10.1016/j.cbpra.2014.03.005] [Medline: 26120269]
102. Roy MJ, Costanzo ME, Highland KB, Olsen C, Clayborne D, Law W. An app a day keeps the doctor away: guided education and training via smartphones in subthreshold post traumatic stress disorder. Cyberpsychol Behav Soc Netw 2017 Aug;20(8):470-478. [doi: 10.1089/cyber.2017.0221] [Medline: 28737954]
103. Solar C, Halat AM, MacLean RR, et al. Predictors of engagement in an internet-based cognitive behavioral therapy program for veterans with chronic low back pain. Transl Behav Med 2021 Jun 17;11(6):1274-1282. [doi: 10.1093/tbm/ibaa098] [Medline: 33098304]
104. Forbes S, Fikretoglu D. Building resilience: the conceptual basis and research evidence for resilience training programs. Rev Gen Psychol 2018 Dec;22(4):452-468. [doi: 10.1037/gpr0000152]
105. Pavlacic JM, Buchanan EM, McCaslin SE, Schulenberg SE, Young JN. A systematic review of posttraumatic stress and resilience trajectories: identifying predictors for future treatment of veterans and service members. Prof Psychol Res Pract 2022;53(3):266-275. [doi: 10.1037/pro0000451]
106. Lyubomirsky S. The How of Happiness: A Scientific Approach to Getting the Life You Want: Penguin Press; 2007.
107. Pietrzak RH, Cook JM. Psychological resilience in older U.S. veterans: results from the national health and resilience in veterans study. Depress Anxiety 2013 May;30(5):432-443. [doi: 10.1002/da.22083] [Medline: 23468170]
108. Masten AS. Risk and resilience in development. In: Zelazo PD, editor. The Oxford Handbook of Developmental Psychology, Vol 2: Self and Other: Oxford University Press; 2013:579-608. [doi: 10.1093/oxfordhb/9780199958474.013.0023]
109. Halbur DA, Halbur KV. Developing Your Theoretical Orientation in Counselling and Psychotherapy: Pearson; 2019.
110. Masten AS, Cicchetti D. Resilience in development: progress and transformation. In: Cicchetti D, editor. Developmental Psychopathology: Wiley; 2016:1-63. [doi: 10.1002/9781119125556.devpsy406]
111. Bonanno GA, Diminich ED. Annual research review: positive adjustment to adversity--trajectories of minimal-impact resilience and emergent resilience. J Child Psychol Psychiatry 2013 Apr;54(4):378-401. [doi: 10.1111/jcpp.12021] [Medline: 23215790]
112. Kalisch R, Baker DG, Basten U, et al. The resilience framework as a strategy to combat stress-related disorders. Nat Hum Behav 2017 Nov;1(11):784-790. [doi: 10.1038/s41562-017-0200-8] [Medline: 31024125]
113. Pietrzak RH, Southwick SM. Psychological resilience in OEF–OIF veterans: application of a novel classification approach and examination of demographic and psychosocial correlates. J Affect Disord 2011 Oct;133(3):560-568. [doi: 10.1016/j.jad.2011.04.028]
114. Dozois DJA. Psychological treatments: putting evidence into practice and practice into evidence. Can Psychol 2022;54(1):1-11. [doi: 10.1037/a0031125]
115. Dozois DJA, Mikail SF, Alden LE, et al. The CPA presidential task force on evidence-based practice of psychological treatments. Can Psychol 2014;55(3):153-160. [doi: 10.1037/a0035767]
116. Memon MA, Cheah JH, Ramayah T, Ting H, Chuah F, Cham TH. Moderation analysis: issues and guidelines. J Appl Struct Equation Model 2019;3(1):i-xi. [doi: 10.47263/JASEM.3(1)01]
117. Fritz MS, Mackinnon DP. Required sample size to detect the mediated effect. Psychol Sci 2007 Mar;18(3):233-239. [doi: 10.1111/j.1467-9280.2007.01882.x] [Medline: 17444920]
118. Faber J, Fonseca LM. How sample size influences research outcomes. Dental Press J Orthod 2014;19(4):27-29. [doi: 10.1590/2176-9451.19.4.027-029.ebo] [Medline: 25279518]
119. Breitborde NJK, Srihari VH, Pollard JM, Addington DN, Woods SW. Mediators and moderators in early intervention research. Early Interv Psychiatry 2010 May;4(2):143-152. [doi: 10.1111/j.1751-7893.2010.00177.x] [Medline: 20536970]

XSL•FO
RenderX

Abbreviations

DMHI: digital mental health intervention
I-COPPE: interpersonal, community, occupational, physical, psychological, economic, and overall well-being
MM: military member
PPTE: potentially psychologically traumatic event
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PSP: public safety personnel
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial

Edited by L Buis; submitted 06.08.24; peer-reviewed by A Eisingerich, K Drude; revised version received 09.05.25; accepted 22.06.25; published 28.10.25.

Please cite as:
Allen RR, Malik MA, Aquin C, Herceg L, Brémault-Phillips S, Sevigny PR
Digital Health Interventions for Military Members, Veterans, and Public Safety Personnel: Scoping Review
JMIR Mhealth Uhealth 2025;13:e65149
URL: https://mhealth.jmir.org/2025/1/e65149
doi:10.2196/65149

© Rashell R Allen, Myrah A Malik, Carley Aquin, Lucijana Herceg, Suzette Brémault-Phillips, Phillip R Sevigny. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

User Experience in mHealth Research: Bibliometric Analysis of Trends and Developments (2007–2023)

Bashaer Alkhwaiter[1,2*]; Monira Aloud[1*]; Nora Almezeini[1*]

[1]Department of Management Information Systems, College of Business Administration, King Saud University, Building No. 3, Riyadh, Saudi Arabia

[2]Department of E-Commerce, College of Administrative and Financial Sciences, Saudi Electronic University, Riyadh, Saudi Arabia

[*]all authors contributed equally

Corresponding Author:
Bashaer Alkhwaiter
Department of Management Information Systems, College of Business Administration, King Saud University, Building No. 3, Riyadh, Saudi Arabia

Abstract

Background: The significance of mobile health (mHealth) apps transforms traditional health care delivery and enables individuals to actively manage their health. The success and effectiveness of mHealth apps heavily depend on the user experience and satisfaction. Previous studies have examined mHealth adoption through systematic literature reviews, focusing on mental health, chronic disease management, fitness, and public health responses to crises like the COVID-19 pandemic. However, the state of research, the key trends, themes, and gaps in the user experience and satisfaction with mHealth apps remain unexplored.

Objective: This study aimed to investigate the state of research on user experience in mHealth apps through a bibliometric analysis. Furthermore, the study aims to systematically identify research trends and themes by extending the analysis of the science mapping technique, co-word analysis, and bibliographic coupling.

Methods: The bibliographic data corpus was collected from Scopus and Web of Science and systematically analyzed using bibliometric performance analysis and science mapping techniques. The methodology incorporates various data processing and visualization tools, including VOS Viewer, OriginLab, and SiteSpace. Then, a comprehensive review metric, combining the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework and a 4-step approach from data collection to interpretation is used.

Results: The bibliographic analysis spans 16 years and includes 814 unique publications authored by 4870 researchers from 81 countries and 1948 organizations, published across 351 high-impact journals and prominent conferences. The analysis of research trends identifies 2 key trends: the differentiation in keyword usage for user experience and user satisfaction, and the research methodologies used within the domain. Furthermore, 5 research themes were identified exploring critical aspects of technology use, user engagement, and clinical integration. Although all 5 themes overlap, each theme focuses on distinct elements that help delineate their contributions to the overall understanding of mHealth apps: technological evaluation (Theme 1), design features for engagement (Theme 2), patient usability (Theme 3), long-term engagement factors (Theme 4), and clinical integration (Theme 5).

Conclusions: This study offers a fundamental understanding of the bibliographic landscape of research on user experience and satisfaction with mHealth apps. By identifying major research clusters, influential works, and emerging topics, this analysis provides evidence-based guidance for researchers, developers, and health informatics practitioners. Furthermore, based on the research trends findings, future research should prioritize expanding the scope of user experience (UX) evaluation by incorporating diverse user populations, longitudinal studies, and emerging technologies such as artificial intelligence and personalized interventions. Integrating insights from interdisciplinary perspectives such as human-computer interaction, behavioral sciences, and health care informatics, the understanding of user needs and app effectiveness can be enhanced. A more standardized framework for assessing UX in mHealth apps is also recommended to facilitate comparability across studies and improve app design to maximize user engagement and health outcomes.

(JMIR Mhealth Uhealth 2025;13:e75909) doi:10.2196/75909

KEYWORDS

bibliometric analysis; mHealth; origin lab; sitespace; systematic literature review; user experience; user satisfaction; VOS Viewer

XSL•FO
RenderX

Introduction

With the advancement of mobile technology and increased global connectivity, mobile health (mHealth) apps have become a significant tool in health care [1,2]. Modern access to medical information through mHealth apps allows users to monitor their biometrics and manage chronic illnesses effectively [3-5]. These mHealth apps are reshaping traditional health care delivery and empowering individuals to take an active role in managing their health [1,2,6,7]. In addressing health care access gaps, mHealth apps are vital, particularly in remote, underserved areas and during global crises such as the COVID-19 pandemic [8,9]. The importance of research in mHealth apps has only increased in recent years, as they present unique solutions for health and fitness, chronic disease management, and mental health support, making health care more accessible to diverse populations [10,11]. However, as mHealth apps become integrated into daily life, user experience and satisfaction have emerged as key factors in their effectiveness, with research revealing that engagement, usability, and personalization directly influence users' sustained interaction with these apps [6,8,12-22].

Research in mHealth can be derived from 2 motivations: business opportunities and human impact. The digital health market is experiencing significant growth because of the increasing use of health apps for fitness tracking and chronic disease management. In 2023, global app downloads were projected to reach 257 billion, highlighting the pervasive role of mobile apps in everyday life [23]. Mobile app revenues are also expected to exceed US $613 billion by 2025, indicating substantial financial prospects [23]. Specifically, in the context of this research, health and fitness apps were projected to achieve 3.76 billion downloads in 2022, with medical apps reaching 305.4 million downloads [24,25]. These figures underscore the strong global demand for mobile health solutions. Concurrently, the human impact is equally influential, with health and wellness becoming priorities for individuals and governments alike. Initiatives are increasingly focused on promoting patient empowerment and improving health care accessibility, as underscored by World Health Organization's 2022 framework, which emphasizes the importance of active patient engagement in health decisions [26].

Understanding how and why people adopt mHealth solutions is of critical significance. On an individual level, effective digital tools may promote healthier lifestyles, increase adherence to treatment plans, enhance patient education, and ultimately improve health outcomes [3-5]. On a societal level, the large-scale adoption of mHealth apps and wearables can ease the burden on health care systems, reduce costs, and improve health equity by delivering accessible and personalized care [1,2,6,7]. However, to realize these benefits, it is essential to investigate the factors that shape user experience and satisfaction with mHealth apps [6,8,12-22]. Research in this area aims to ensure effective engagement, address issues of interoperability and integration with clinical workflows, and develop strategies that sustain long-term usage and trust. However, to build on insights from technology development, health, behavioral science, and user experience design, research in this area faces a critical gap in fully mapping the intellectual structure, theoretical foundations, and interdisciplinary perspectives shaping this evolving body of knowledge.

Previous research has tried to bridge this gap by reviewing the literature influencing various mHealth apps and wearables. Using a systematic literature review method to examine mHealth adoption, user experience, and satisfaction, tailored to the contexts of mental health [15,19,27-39] and chronic disease management [6,40-49], fitness apps [8,50-52], and public health responses to crises like the COVID-19 pandemic [53-57]. Most reviews targeted mHealth for specific demographic groups, including women [58], older adults [59], and children [60,61]. Some reviews have adopted a broader scope, addressing general mHealth topics over extended periods [62-67]. However, the studies incorporated in these reviews often involved small sample sizes (2‐178 studies), which may limit the representativeness and generalizability of the conclusions. A larger number of reviewed studies (n=365) were included in the study by Hu et al [68], the only bibliometric analysis identified, while offering a distinct perspective on the impact of research in this area, health wearables were the main focus. Collectively, these studies underscore the importance of considering user satisfaction, engagement, the usability of mHealth, and the impact of content and design on user experience. This suggests that future digital health solutions must adapt to user needs and preferences to maximize their impact, requiring ongoing collaboration between researchers, designers, and end users to refine and optimize these mHealth innovations. These insights are crucial for overcoming current limitations and enhancing the efficacy of digital solutions in health care.

Despite this helpful effort in reviewing related literature, a clear intellectual structure has not yet been offered. To address this gap, a more comprehensive review is required to investigate the bibliographic contributions to user experience and satisfaction with mHealth solutions. We included both user experience and user satisfaction as keywords to capture their combined and distinct usage within the same body of literature. Therefore, this bibliographic analysis explores the state of research on user experience and satisfaction with mHealth apps (UXS-mHealth apps). Aiming to exhibit the intellectual structure of related research in this area, this study identifies the leading publications, authors, countries, organizations, and sources that contribute to understanding mHealth apps and impact research development in UXS-mHealth apps. Furthermore, this study explores the research trends within thematic research areas related to UXS-mHealth apps via a systematic qualitative review. By integrating these quantitative and qualitative review methods, this review offers a data-driven perspective that complements traditional literature reviews and enables a more holistic understanding of the research on UXS-mHealth apps. These findings will contribute to the academic discourse and provide practical insights for developers, designers, and policymakers who aim to improve the effectiveness and user-friendliness of mHealth apps.

XSL•FO
RenderX

Methods

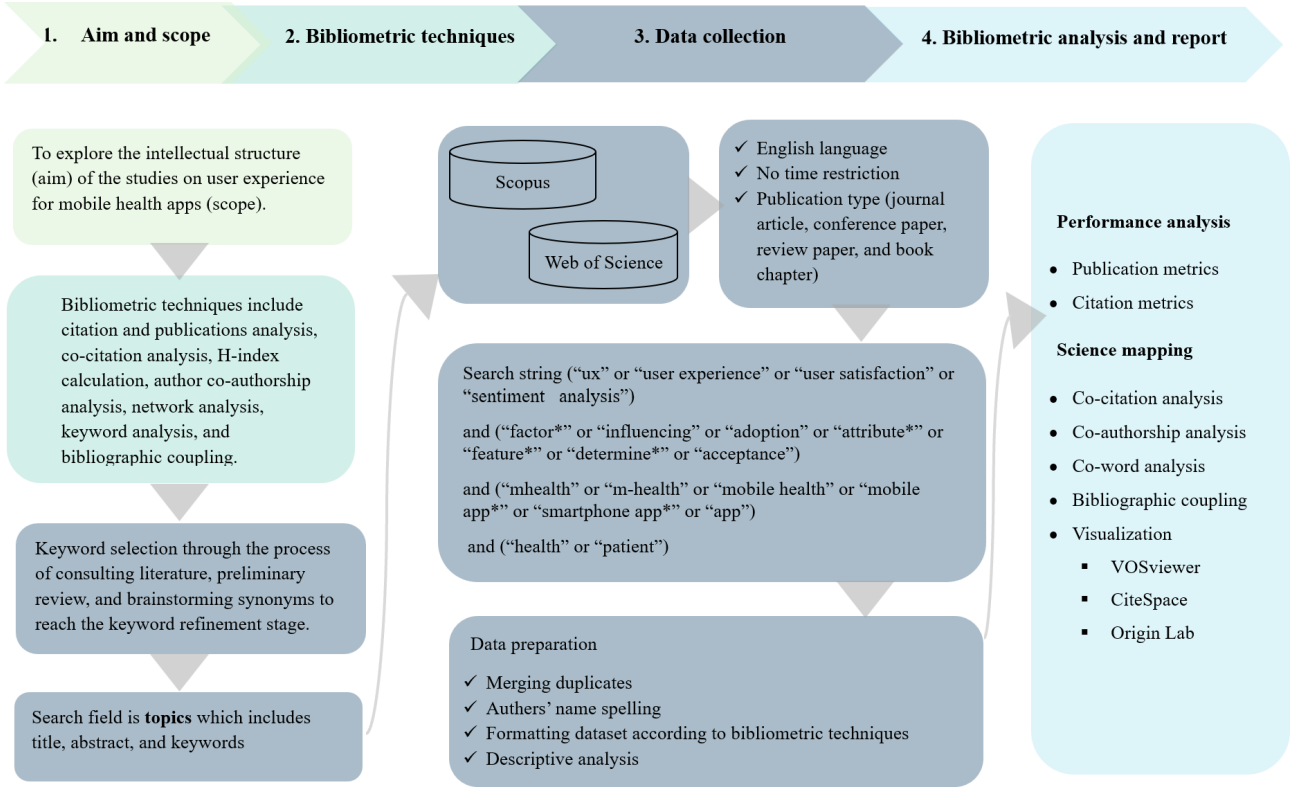
Overview

This study adopts a bibliometric analysis in conjunction with a systematic literature review (SLR). The bibliometric analysis is undertaken to map the current state of research on UXS-mHealth apps and to identify the most influential publications and authors in the field. Building on these findings, the SLR offers an in-depth examination of literature, identifying the evolution of research themes, highlighting emerging topics, and revealing gaps that warrant further investigation.

Bibliographic Analysis

The bibliometric analysis follows the procedural guidance established in the study by Donthu et al [69]. Figure 1 illustrates the bibliometric analysis process, including defining study aims, selecting techniques, data collection, and conducting analysis. This framework ensures transparency and reproducibility when analyzing UXS-mHealth apps.

Figure 1. The 4-step bibliometric analysis search design used in studying the research structure of user experience and satisfaction with mobile health (mHealth) apps.



Data Sources and Search Strategy

The data sources used in this analysis are Scopus and Web of Science. The focus on retrieving bibliographic data for publications indexed in the established databases, Scopus and Web of Science, ensures the inclusion of high-quality and widely recognized sources to explore the field development. Scopus and Web of Science are widely considered the premier databases for conducting bibliometric research because of their comprehensive coverage, rigorous indexing criteria, and multidisciplinary scope [69]. Both platforms offer robust citation tracking capabilities, essential for a thorough bibliometric assessment, enabling the identification of influential papers and the evaluation of research impact [70]. Furthermore, rigorous selection criteria and quality control mechanisms ensure the inclusion of reputable journals and high-quality research [71]. Including both databases eliminates bias and provides access to various scholarly articles, conference papers, and book chapters, ensuring researchers can capture a broad and representative sample of relevant literature [72]. In addition, Web of Science and Scopus provide advanced analytical tools for examining citation networks, author metrics, and journal impact, which are critical for detailed bibliometric studies [73]. Other databases may not offer the same level of comprehensive coverage or advanced analytical features, potentially limiting the scope and depth of bibliometric analyses [10]. Thus, selecting these 2 databases addresses the study's aim because their inclusion of citation data allows for sophisticated analyses of research impact, trends, and collaborations, making them indispensable resources for bibliometric studies [69,71].

From both platforms, the bibliographic data were extracted for each publication. This included authorship, publication year, journal name, affiliation, document type, volume, issue, pages, citations, and sponsorship. This data was gathered using the export functions embedded in Scopus and Web of Science with full record selection, complemented by manual checks to verify accuracy.

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria in this study ensure that the research captures a comprehensive and representative literature sample designed based on the study's objectives. These criteria involve several dimensions: publication type, language,

XSL•FO
RenderX

data source, keywords, and search field. The publication type inclusion criteria include journal articles, conference papers, review papers, and book chapters. Only publications indexed in Scopus and Web of Science were considered, and all search results were limited to publications published in English, without restriction on publication date. The search was done in June 2024. The selection is restricted to full-text articles to facilitate comprehensive analysis and ensure complete bibliographic and citation data availability.

The selection of relevant keywords began with a comprehensive review of the literature, focusing on titles, keywords, and abstracts of highly cited and pertinent articles. After analyzing over 20 articles, including [3,4,10,15,18,74,75], it became evident that many commonly used keywords were already present in the initial search string. To enhance the search's alignment with the study's focus, additional keywords were brainstormed by identifying core concepts, such as user experience, user satisfaction, and mHealth apps, followed by expanding these concepts with synonyms and related terms. This iterative process ensured a comprehensive and targeted search string that effectively captured the essential aspects of the research topic.

The search string is categorized into 4 groups: A, B, C, and D. Group A includes user experience and user satisfaction, whereas those concerning the identification of influences, factors, and determinants are included in Group B. Groups C and D focus on keywords specific to the mHealth domain and health and patient contexts, respectively. The advanced search tools in Scopus and Web of Science use the Boolean operator AND to connect the groups, and OR was used within each group to enhance the specificity and relevance of the search results. These criteria ensure that the bibliometric analysis is focused, relevant, and based on high-quality sources. After a preliminary review of the relevance of the abstract, some keywords were excluded. The exclusion of keywords from each group is illustrated in the following:

- Group A: "UX" OR "user experience" OR "user satisfaction" OR "sentiment analysis." The keyword "Ranking" was excluded as it predominantly related to ranking methodologies rather than user experience or satisfaction.
- Group B: "factor*" OR "influencing" OR "adoption" OR "attribute*" OR "feature*" OR "determine*" OR "acceptance." The keywords "use," "usage," and "perceived value" were excluded because they predominantly led to publications on the clinical use of mHealth systems and perceived value in clinical contexts, which were outside the review's scope.
- Group C: "mhealth" OR "m-health" OR "mobile health" OR "mobile app*" OR "smartphone app*" OR "app." The keywords "app*" and "system*" or "digital platform*" were excluded as they primarily retrieved publications on electronic clinical systems or web-based platforms, which were not relevant to mHealth apps.
- Group D: "health" OR "patient."

Data Preparation

Data preparation is crucial for aligning the data corpus with the bibliometric techniques chosen, as recommended by [69]. The data cleansing process includes merging duplicate publications within and between databases, checking for variations in authors' name spellings, and processing author affiliations. A manual review is conducted to reconcile discrepancies such as differences in initials, name order, or the inclusion of middle names. In addition, author identifiers are used to ensure the precise identification and merging of publications by the same author, regardless of variations and the indexed platform. Furthermore, the preparation process involves verifying the publication type, where publications initially categorized as review papers, such as those reviewing apps, are reclassified as journal or conference papers. This reclassification ensures that publications are accurately categorized according to content and format. The tools used for data preparation include Microsoft Excel for manual checks, and CiteSpace (6.4.R1-64-bit-Advanced) and Zotero for merging duplicates. This approach of preparing data will effectively address the study's aim of exploring the state of research on user experience in mHealth apps by enabling the filtration of journals, publication type trends, time trends, subject areas, author productivity, and citation metrics.

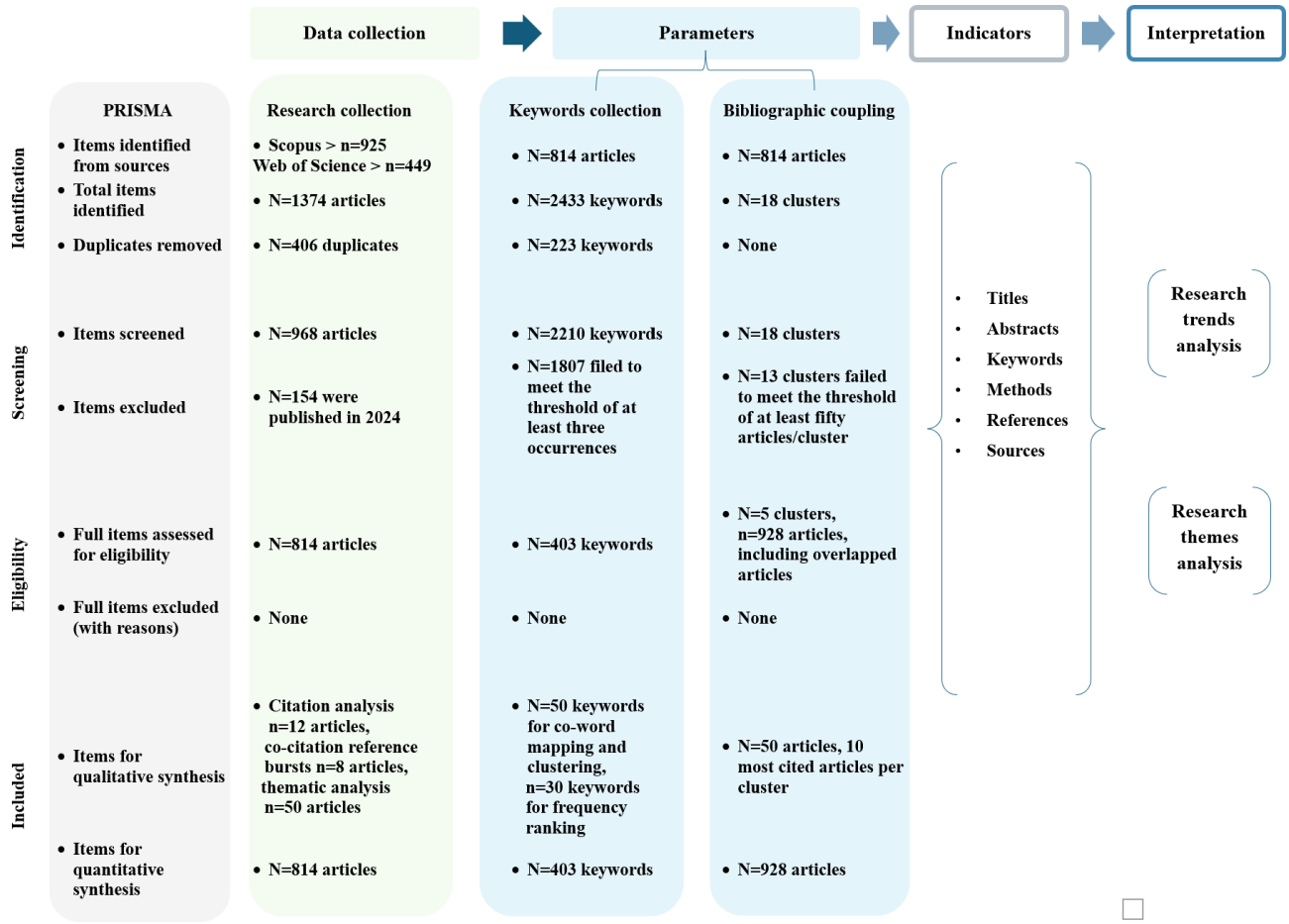
Systematic Literature Review

The bibliometrics is extended for further exploration of co-word analysis and bibliographic coupling by using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework [76] with a 4-step approach from data collection to interpretation [77].

In Figure 2, the PRISMA flow diagram is represented vertically and outlines the vertical systematic filtration process involving identification, screening, eligibility, and inclusion phases. This framework is applied to the horizontal process of data collection, which involves defining parameters and indicators and the interpretation of results. This process is used to identify meaningful conceptual linkages among the collected data. For each publication, indicators such as titles, abstracts, keywords, research methods, research type, and sources are systematically extracted to characterize and categorize the literature. The PRISMA checklist can be found in Checklist 1.

XSL•FO
RenderX

Figure 2. Methodological workflow for mHealth literature synthesis (2007 - 2023) using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-guided selection and the data collection to interpretation process.



Results

The bibliographic dataset encompasses 1376 search results from Scopus and Web of Science, including unique and overlapping results across the two databases. Of the 1376 results, 42% (407/1376) are duplicates, while Web of Science contributes 4.4% (43/1376) of unique results, and Scopus contributes (53.6%, 519/1376). The distribution of sources and different document types is illustrated in Multimedia Appendix 1.

Bibliographic Performance Analysis

Annual Trend Analysis

To investigate the intellectual framework of research on UXS-mHealth apps, the annual trend analysis provides valuable insights into the progression of research activity over time. The 2 most commonly used metrics are the annual total number of publications (TP) and total citations (TC); publications reflect research productivity, whereas citations measure research impact [69]. By examining these metrics, it becomes evident how the field has grown, contracted, or responded to influential events over time [69,78].

This study's annual analysis of the 814 publications from 2007 to 2023 reveals distinct patterns that illustrate the growth of the UXS-mHealth apps domain, as shown in Figure 3. The dataset's first publication year is 2007, although it covers an unrestricted time span. This event may indicate a connection to technological advancement in 2007, synchronized with Apple's launch of the first interactive smartphone. The launch of the iPhone marked a significant shift in mobile technology, making smartphones more accessible and elevating the importance of mobile apps [79,80]. This technological milestone is evident in the initiation of academic output, as scholars began exploring the broader implications of smartphone technology across various fields.

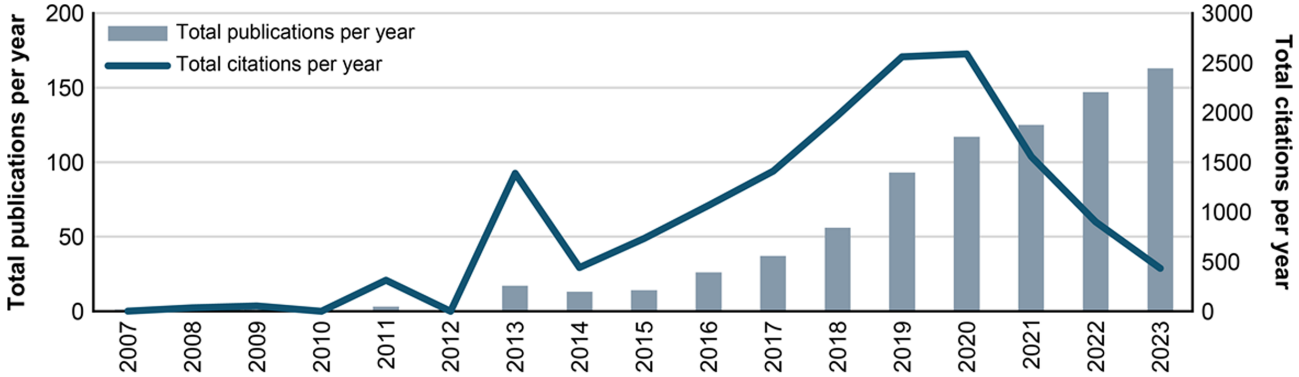
From 2007 to 2012, publications primarily focused on mobile user interactions and apps. In 2013, the field experienced a significant peak, signaling a surge of interest or noteworthy advancements that captured researchers' attention. This might have been due to the rising competition between iOS and Android operating systems, as Google released the Android Play Store around this time [79,80]. However, this was followed by a decline in 2014, which may suggest a temporary shift in focus, challenges in the field, or the resolution of particular research questions [78]. During this decline, most publications concentrated on fundamental concepts of mobile app acceptance and usability. However, an innovation breakthrough, such as the announcement of the Apple Watch in 2014, designed to track health and fitness metrics, likely shifted research interest [81].

From 2015 onward, a steady increase in publications indicates sustained and growing interest in UXS-mHealth apps. The number of research efforts based on annual TP increased from 17 in 2013 to 163 in 2023. This consistent rise in TP from 2015 to 2023 could be attributed to technological advancements,

XSL•FO
RenderX

methodology, or emerging mobile apps that have evolved research efforts [69]. These trends indicate that the field is progressing and expanding its impact as it matures, attracting contributions from a growing number of researchers.

Figure 3. The annual analysis of publications on UXS-mHealth apps from 2007 to 2023 shows the productivity of publications and the research impact based on the annual total publications (TP) and total citations (TC).



The citation trend depicted in Figure 3 shows an annual increase in citations, except for two declines in 2014 and between 2021 and 2023. There is also a slight increase in citations during COVID-19, followed by a decrease. The heightened attention during this period led to a rapid increase in research dissemination, which resulted in many papers being cited quickly [82]. The decline in citations for recent publications between 2021 and 2023 is often due to the time required for new publications to gain visibility within the academic community [83-85]. The citation count for new papers is still evolving and cannot be fairly compared to older publications that have had more time to accumulate citations [83-85].

Overall, the gradual evolution of the field underscores the developing stages of research into UXS-mHealth apps. This analysis provides valuable insights into the development of this area, revealing both promising opportunities for future research and aspects that warrant further exploration. As the field advances, understanding these stages can guide researchers in addressing emerging challenges and leveraging new technologies, ultimately enhancing user experience on mobile app outcomes.

Leading Countries Contributions

The analysis of regional distribution and collaborative networks for UXS-mHealth apps publications from 2007 to 2023 reveals several key trends that enhance our understanding of geographical productivity and impact. Countries contributing to the research were collected from the authors' affiliations. Table 1 presents the 20 top contributing countries based on the TP, TC, citations per publication (TC/TP), and the measure of influence h-index (h). Additional visualization for countries, ranked based on the TP and citations, is presented in Multimedia Appendix 2. The distribution of these outputs highlights how different countries have shaped research within this domain.

The United States leads in most metrics, collaborating primarily with Ireland, Switzerland, the United Kingdom, Australia, and others. In terms of impact, the United States (h=30) is the most influential, followed by the United Kingdom and Australia (h=20). The United States accounts for 30.47% (248/814) of the TP, followed closely by the United Kingdom, Australia, and China. However, a different pattern emerges when analyzing the citations per publication (TC/TP): Saudi Arabia, Portugal, and the United Kingdom surpass other nations, coming to the forefront. Notably, Saudi Arabia and Portugal stand out due to citations in publications despite having fewer publications. Saudi Arabia, in particular, demonstrates a unique position with relatively few publications but a strong influence on the field. This impact is further emphasized by Saudi Arabia's collaborative strength, which ranks fifth internationally. This suggests that strategic international collaborations can enhance a country's influence and citation impact, even with lower research output.

The chord diagram in Figure 4 visualizes the collaboration between the top 20 countries contributing to the UXS-mHealth apps research. The diagram was produced using the OriginLab software package for visualization purposes. The size of each segment corresponds to the total number of research publications produced by each country. The more significant segments indicate higher output, indicating the most prolific countries in the research area. The collaborative links between countries are depicted as lines or chords, connecting the segments. These lines symbolize coauthorships or joint research projects between nations. Thicker lines reflect more highly cited collaborations, indicating that the collaborative efforts between certain countries significantly influence the research landscape.

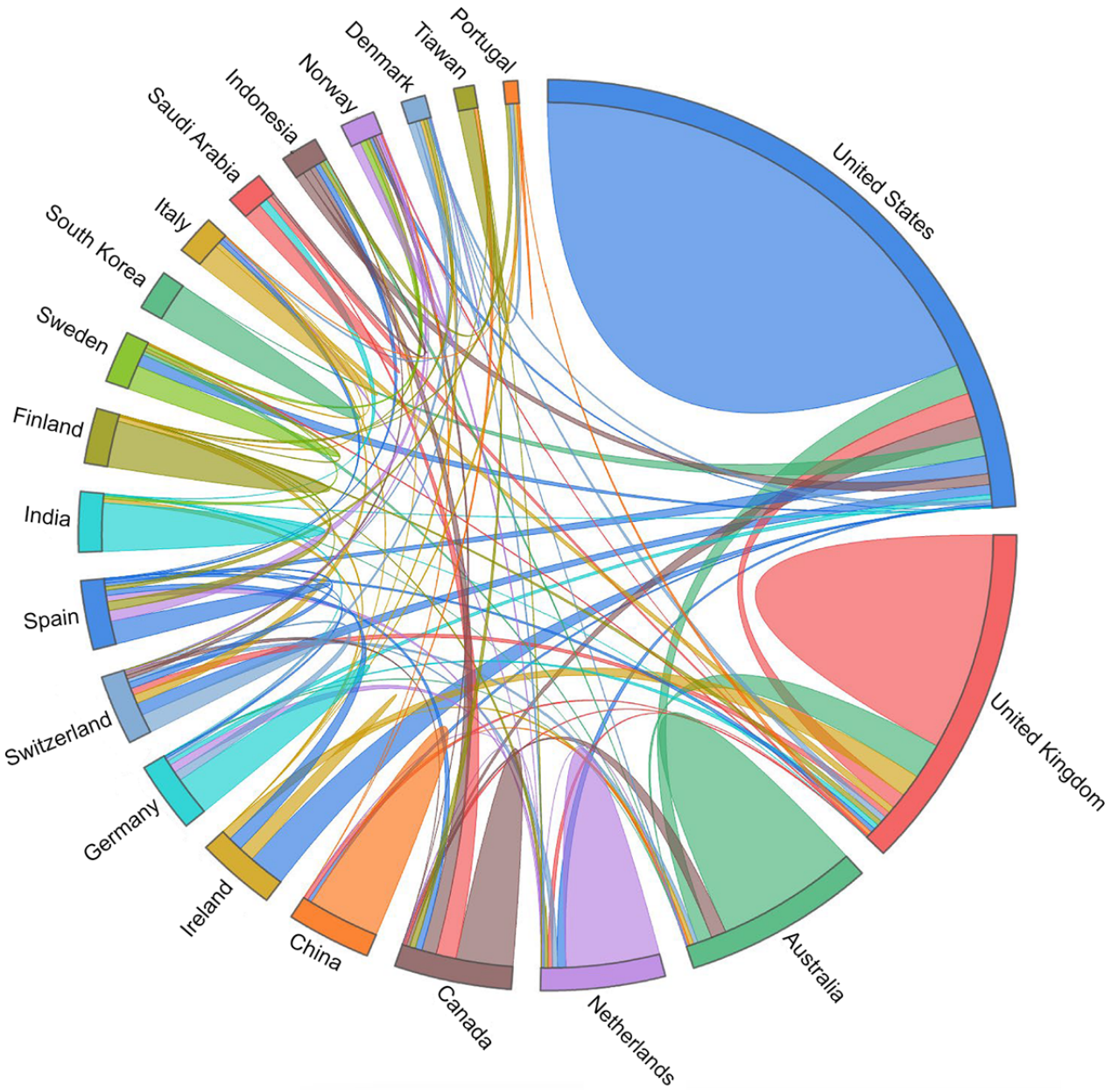
XSL•FO
RenderX

Table . Contribution of the top 20 countries, out of 81, to research on UXS-mHealth apps from 2007 to 2023. The country ranking is based on two bibliometric indicators: total publications (TP), total citations (TC), and the measure of influence h-Index.

Country	TP	TC	(TC/TP)	h-Index
United States	248	4275	17.24	32
United Kingdom	103	1971	19.14	20
Australia	87	1357	15.60	20
China	75	558	7.44	13
Germany	64	809	12.64	14
Netherlands	60	593	9.88	14
Canada	56	542	9.68	13
Spain	46	405	8.80	13
India	45	262	5.82	9
Italy	37	428	11.57	8
Switzerland	34	137	4.03	8
Indonesia	29	226	7.79	5
South Korea	27	120	4.44	8
Denmark	24	115	4.79	4
Ireland	23	314	13.65	8
Portugal	21	487	23.19	4
Sweden	21	223	10.62	8
Saudi Arabia	20	464	23.20	7
Norway	19	310	16.32	6
Taiwan	19	237	12.47	6

XSL•FO
RenderX

Figure 4. A chord diagram illustrating the top 20 countries' publication contributions and collaboration networks in user experience and satisfaction with mobile health apps, with the thickness of the connecting chords representing the number of citations.



Furthermore, the chord diagram represents the proportion of international collaborations relative to local, independent work for each country. This feature visualizes the proportion of research involving international cooperation relative to domestic efforts. Therefore, collaboration patterns, in countries such as China, India, Finland, South Korea, and Taiwan, exhibit a more localized approach to collaboration. While these nations are productive in their research output, their collaborative networks remain primarily confined to their local regions. These countries' focus on local connections might restrict their influence in a broader global context. The country cooperation network shown in Figure 4 is obtained using the Origin tool through affiliation analysis. This visualization offers more profound insights into a country's research volume and impact and the balance between local and global collaboration.

These findings underscore the importance of productivity and collaboration in shaping regional contributions and global influence within a specific research domain. These leading countries have contributed to UXS-mHealth apps research and have advanced the domain through local and international partnerships.

Leading Research Entities Contributions

Analyzing research entities independently of country affiliation in UXS-mHealth apps publications reveals distinct patterns compared to country-level analysis. Although country-based assessments often emphasize geopolitical trends, research entity analysis provides a more granular view of institutional characteristics. This analysis encompasses 1948 research entities, including academic institutes, hospitals, medical centers, and private organizations. Such diversity suggests the multidisciplinary nature of the field, where theoretical research,

XSL•FO
RenderX

practical health care applications, and commercial development contribute to the growth of UXS-mHealth app publications. Table 2 ranks the top 30 research entities based on bibliometric indicators, including TP and citations received. The clustered bar chart compares these metrics for each research entity contributing to UXS-mHealth apps research, which is offered in Multimedia Appendix 3.

XSL•FO
RenderX

Table . Top 30 research entities, out of 1948, contributed to research on UXS-mHealth apps[a] from 2007 to 2023. The ranking of the research entities is based on 2 bibliometric indicators: the total publications (TP) and the number of citations received (TC).

Rank	Research entity	Country	TP	TC	TC/TP
1	University of Toronto	Canada	30	373	12.43
2	University College London	The United Kingdom	27	362	13.41
3	University of Sydney	Australia	18	352	19.56
4	University of Melbourne	Australia	17	224	13.18
5	Deakin University	Australia	17	121	7.12
6	University of Utah	The United States	16	453	28.31
7	University of Queensland	Australia	16	234	14.63
8	University of California	The United States	16	160	10.00
9	University of Copenhagen	Denmark	15	113	7.53
10	National University of Singapore	Singapore	15	56	3.73
11	The University of Sydney	Australia	13	100	7.69
12	Sacred Heart University	Italy	13	84	6.46
13	Karolinska Institute	Sweden	12	137	11.42
14	University of Michigan	The United States	12	102	8.50
15	IRCCS Policlinico Gemelli	Italy	12	84	7.00
16	University of California System	The United States	11	462	42.00
17	University Road	Ireland	11	224	20.36
18	University of Pennsylvania	The United States	11	158	14.36
19	University of Washington	The United States	11	142	12.91
20	King’s College London	The United Kingdom	11	126	11.45
21	Zhejiang University	China	11	44	4.00
22	Northwestern University	The United States	10	756	75.60
23	University of California San Francisco	The United States	10	206	20.60
24	University of British Columbia	Canada	10	180	18.00
25	King Saud University	Saudi Arabia	10	170	17.00
26	Monash University	Australia	10	119	11.90
27	Bina Nusantara University	Indonesia	10	31	3.10
28	Imperial College London	The United Kingdom	9	668	74.22
29	Johns Hopkins University	The United States	9	641	71.22

XSL•FO
RenderX

Rank	Research entity	Country	TP	TC	TC/TP
30	University of Pitts-burgh	The United States	9	250	27.78

[a]UXS-mHealth apps: user experience and satisfaction with mobile health apps.

The University of Toronto in Canada leads the list with 30 publications, followed by University College London with 27. Several Australian institutions, including the University of Sydney (n=18), the University of Melbourne (n=17), and Deakin University (n=17), also feature prominently. Notably, research entities from the United States account for 33.33% of the top contributors, whereas Australian institutions contribute 20%. Although Canada has fewer institutions in the top 30, the University of Toronto leads research production in UXS-mHealth apps.

This analysis also uncovers distinct patterns in specialization and influence among the entities. While producing fewer publications, certain institutions demonstrate high citation impact, suggesting groundbreaking contributions to specific subfields. For example, Northwestern University in the United States leads in citations with 756, followed by Imperial College London in the United Kingdom with 668 citations, and Johns Hopkins University in the United States with 641. This highlights that some institutions achieve significant academic impact even with fewer TP, underscoring their role as thought leaders within the field.

Distribution of Literature Sources

The analysis of the leading sources for UXS-mHealth apps reveals a total of 351 sources, comprising 260 journals and 91 conferences, demonstrating the extensive research activity in this area. Table 3 depicts that the most prolific journals account for 56.26% (458/814) of the TP in the dataset. Notably, among the most influential sources of knowledge, the *Journal of Medical Internet Research* and its specialized series emerge as critical contributors. Although *JMIR mHealth and uHealth* leads the list with a total of 74 publications and 2512 citations, the journals' list, including *JMIR Formative Research*, *JMIR Human Factors*, *JMIR Mental Health*, *JMIR Diabetes*, *JMIR Cardio*, and *JMIR Pediatrics and Parenting*, plays a pivotal role in advancing research related to UXS-mHealth apps.

XSL•FO
RenderX

Table . Top 30 scientific sources contributing to UXS-mHealth apps[a] publications from 2007 to 2023, showcasing the number of publications (TP) and impact based on total citations received (TC).

Ranking	Source title	TP	TC
1	*JMIR mHealth and uHealth*	74	2512
2	*JMIR Formative Research*	60	234
3	*Journal of Medical Internet Research*	52	2493
4	*JMIR Human Factors*	28	406
5	*Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*	25	58
6	*JMIR Research Protocols*	18	165
7	*Digital Health*	17	56
8	*ACM International Conference Proceeding Series*	15	44
9	*International Journal of Medical Informatics*	13	646
10	*International Journal of Environmental Research and Public Health*	12	139
11	*Frontiers in Public Health*	12	86
12	*Telemedicine and e-Health*	11	432
13	*BMC Medical Informatics and Decision Making*	10	171
14	*Studies in Health Technology and Informatics*	10	166
15	*PLoS ONE*	8	398
16	*JMIR Mental Health*	8	153
17	*Journal of Diabetes Science and Technology*	8	126
18	*Internet Interventions*	8	72
19	*BMJ Open*	8	66
20	*Communications in Computer and Information Science*	7	14
21	*IEEE Access*	6	167
22	*JMIR Diabetes*	6	57
23	*JMIR Cardio*	6	38
24	*Advances in Intelligent Systems and Computing*	6	5
25	*BMC Public Health*	5	158
26	*Healthcare (Switzerland)*	5	41
27	*Pervasive Health: Pervasive Computing Technologies for Healthcare*	5	36
28	*International Journal of Human-Computer Interaction*	5	31
29	*JMIR Pediatrics and Parenting*	5	30
30	*Conference on Human Factors in Computing Systems*	5	29

[a]UXS-mHealth apps: user experience and satisfaction with mobile health apps.

XSL•FO
RenderX

Leading Authors' Contributions

The analysis of the leading authors' contributions to UXS-mHealth apps highlights a total of 4523 authors who have collectively contributed to the field's body of knowledge. Table 4 illustrates that the most prolific authors have a concentrated research output. Mohamed Abdelrazek leads the productivity metric with eight publications, followed closely by Liam Glynn, Gearóid ÓLaighin, Leo R. Quinlan, and Richard Harte from the National University of Ireland, Galway, each with 5 publications. Notably, authors affiliated with the National University of Ireland, Galway, play a pivotal role in shaping the discourse surrounding UXS-mHealth apps. A total of 5 of their researchers rank among the top 10 prolific authors, underscoring the institution's prominence in this research area.

While these authors hold the second spot in publication volume, they achieved the highest impact, amassing 304 citations, which signifies the broad influence and high relevance of their work in the field. In addition, authors from Finland, specifically from the VTT Technical Research Center of Finland, the University of Jyväskylä, and Tampere University, have made significant contributions. Their influence, based on the TC they received, underscores the quality and innovation of their findings in advancing the field.

XSL•FO
RenderX

Table . The 20 most prolific authors contributed to UXS-mHealth apps[a] publications from 2007 to 2023, summarizing the number of publications (TP), and impact based on total citations received (TC), country, and affiliation.

Author name	TP	TC	Country	Affiliation
Abdelrazek, Mohamed	8	57	Australia	Deakin University
Glynn, Liam	5	304	Ireland	School of Medicine, National University of Ireland Galway
ÓLaighin, Gearóid	5	304	Ireland	School of Engineering & Informatics and the CÚRAM SFI Center for Research in Medical Devices, National University of Ireland Galway
Quinlan, Leo R.	5	304	Ireland	School of Medicine, and the CÚRAM SFI Center for Research in Medical Devices, National University of Ireland Galway,
Harte, Richard	5	304	Ireland	Human Movement Laboratory and the CÚRAM SFI Center for Research in Medical Devices, National University of Ireland Galway
Mattila, Elina	5	299	Finland	VTT Technical Research Center of Finland
Grundy, John	5	52	Australia	Monash University
Scharf, Thomas	4	289	Ireland	Irish Center for Social Gerontology at the National University of Ireland, Galway
Lappalainen, Raimo	4	219	Finland	University of Jyväskylä
Ahtinen, Aino	4	215	Finland	Tampere University
King, Dominic	4	193	The United Kingdom	Imperial College London
Denecke, Kerstin	4	110	Switzerland	Bern University of Applied Sciences
Schueller, Stephen M.	4	46	The United States	University of California
Bonti, Alessio	4	16	The United States	IBM Research
Lei, Jianbo	4	13	China	Peking University
Wang, Tong	4	13	Japan	Ritsumeikan University
Rodríguez–Molinero, Alejandro	3	275	Spain	Consorci Sanitari del Garraf (CSG)
Orji, Rita	3	263	Canada	Dalhousie University
Alqahtani, Felwah	3	263	Canada	Dalhousie University
Kaipainen, Kirsikka	3	219	Finland	Tampere University

[a]UXS-mHealth apps: user experience and satisfaction with mobile health apps.

Bibliographic Science Mapping

Citation Analysis

Exploring the annual most cited references is a citation analysis metric. The summary of the most cited publications per year can be found in Multimedia Appendix 4. The exclusion of data from 2007 to 2010 in this table is due to the relatively low number of publications during those years, which makes them less relevant for analyzing annual trends compared to the surge from 2011 onward.

Co-Citation Analysis

The co-citation analysis is useful for revealing the most influential publications in a field [69]. This study encompasses 51,612 cited references from 814 publications between 2007

XSL•FO
RenderX

and 2023 on UXS-mHealth apps. We used the CiteSpace (6.4.R1-64-bit-Advanced) co-citation tool default settings to identify and visualize the most co-cited references, highlighting seminal works in the UXS-mHealth apps domain.

The co-citation cluster and the strongest citation bursts are presented in Multimedia Appendix 5. The major co-citation clusters, labeled in gray, where the cluster number indicates the cluster volume. The smaller the number, the larger the volume. Clusters such as 1, 9, and others that are unrepresented in the network suggest that they are large but isolated outside the main network structure. The color of each cluster reflects the average publication year, with greenish clusters representing earlier works from 2013 to 2016 and yellow clusters signifying more recent publications from 2018 to 2020. The number of nodes in each cluster corresponds to the volume of publications; larger nodes indicate greater connectivity. Notably, the largest node in each cluster represents a highly influential publication, demonstrating its prominence within that area. In addition, the analysis highlights the top 8 publications with the strongest citation bursts. This indicates periods of intense scholarly attention and influence within the research domain. Notably, the study "Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps" by [86] leads the list, reflecting its significant impact and influence in the field.

The synthesized findings from the 8 studies provide a comprehensive understanding of various aspects of the use, design, and evaluation of mHealth apps, focusing on quality assessment, user engagement, usability, and retention. Each study offers unique insights; however, in combination, they present a cohesive overview of key factors that influence the adoption and effectiveness of mHealth technologies.

One of the primary areas of focus is the quality assessment and rating of mHealth apps. Researchers have developed comprehensive tools to address the limitations of traditional star-based rating systems. For example, the Mobile App Rating Scale (MARS) was introduced as a 23-item tool that evaluates apps across 5 categories: engagement, functionality, esthetics, information quality, and subjective quality. The proposed tool demonstrated excellent psychometric properties, making it a reliable framework for app classification and quality assessment [86]. Complementing this effort, the Mobile Health App Usability Questionnaire (MAUQ) was created to address usability gaps in existing tools, such as the System Usability Scale (SUS) and the Post-Study System Usability Questionnaire (PSSUQ) [87]. The MAUQ was tailored for interactive and standalone apps and demonstrated strong reliability and validity. The MARS and MAUQ are foundational instruments for assessing app quality and usability, offering structured approaches for researchers and developers [86,87].

The publications on user engagement emphasize the crucial role that effective design and functionality play in maintaining app usage. The study explored how users interact with health apps, revealing that initial engagement often declines after reaching certain milestones. Factors such as app design, ease of use, and data management significantly influence this decline [74]. To address this problem, "effective engagement" was proposed, indicating that sufficient interaction with an app is necessary to achieve the desired health outcomes. Tailored interventions that consider individual user needs and contexts are recommended to enhance engagement [88]. A focused study on mental health apps identified barriers to engagement, including a lack of user-centered design and privacy concerns. The findings emphasized the necessity for a user-centric approach to improve sustained engagement and adoption [89].

Usability is critical in influencing user satisfaction and the long-term use of mHealth apps. A comprehensive review of usability testing methods highlighted a predominance of traditional techniques, such as questionnaires, interviews, and task completion protocols [90]. Although questionnaires provide broad usability insights, qualitative methods like "Think-Aloud" protocols are more effective in pinpointing specific usability issues [90]. In addition, there remains a significant underuse of automated testing methods, such as eye-tracking, which can provide objective usability insights [90]. These findings align with the emphasis on developing robust usability instruments like MAUQ, which can adapt to different app types and user contexts [87]. These studies indicate that combining qualitative and quantitative methods is essential for a holistic understanding of usability, and further research should explore integrating automated testing tools to streamline the evaluation process.

Postadoption behavior and retention are critical for the long-term success of health apps. Researchers used models, such as the Post-Acceptance Model and Technology Acceptance Model (TAM), to examine factors influencing the continued use of health apps. They found that perceived usefulness, ease of use, confirmation, and satisfaction significantly predict the intention to continue using the app [91]. Consistent with these findings, an analysis of real-world usage patterns of mental health apps revealed that only a small proportion of users maintain long-term engagement [92]. Certain app features, such as mindfulness and peer support, were associated with higher retention rates than other functionalities, such as trackers or breathing exercises [92]. These results suggest that long-term retention is driven by the perceived value and satisfaction derived from specific app features rather than the overall app experience [91,92].

Coauthorship Analysis

The coauthorship analysis aimed to explore the social structure and collaborative networks among researchers contributing to the field of study [69]. VOSviewer (version 1.6.20) was used for this analysis, and the "authors" were selected as the unit of analysis. The UXS-mHealth research dataset comprised 814 publications from 2007 to 2023, with contributions from 4870 unique authors. A sensitivity analysis was conducted on the coauthorship network, focusing on 3 attributes: documents, citations, and total link strength. Based on this analysis, a threshold of three documents per author was established to ensure meaningful representation, and 158 authors met the criteria. To enhance the coauthorship network visualization, the coauthorship network illustrates collaborations among the top 100 authors.

The resulting coauthorship network depicts 19 distinct clusters, each distinguished by a unique color. Larger node sizes indicate the most significant collaborative relationships based on the total number of citations received. The coauthorship

XSL•FO
RenderX

collaborations among the top 5 cited clusters reveal knowledge-sharing patterns and intellectual partnerships within the UXS-mHealth research. Furthermore, using the CiteSpace tool, the co-citation analysis for authors reveals the most cited authors, their citation strength, and citation bursts over the years. The coauthorship networks, the coauthorship theme summary, and the coauthor citation bursts can be found in Multimedia Appendix 6.

Co-Word Analysis

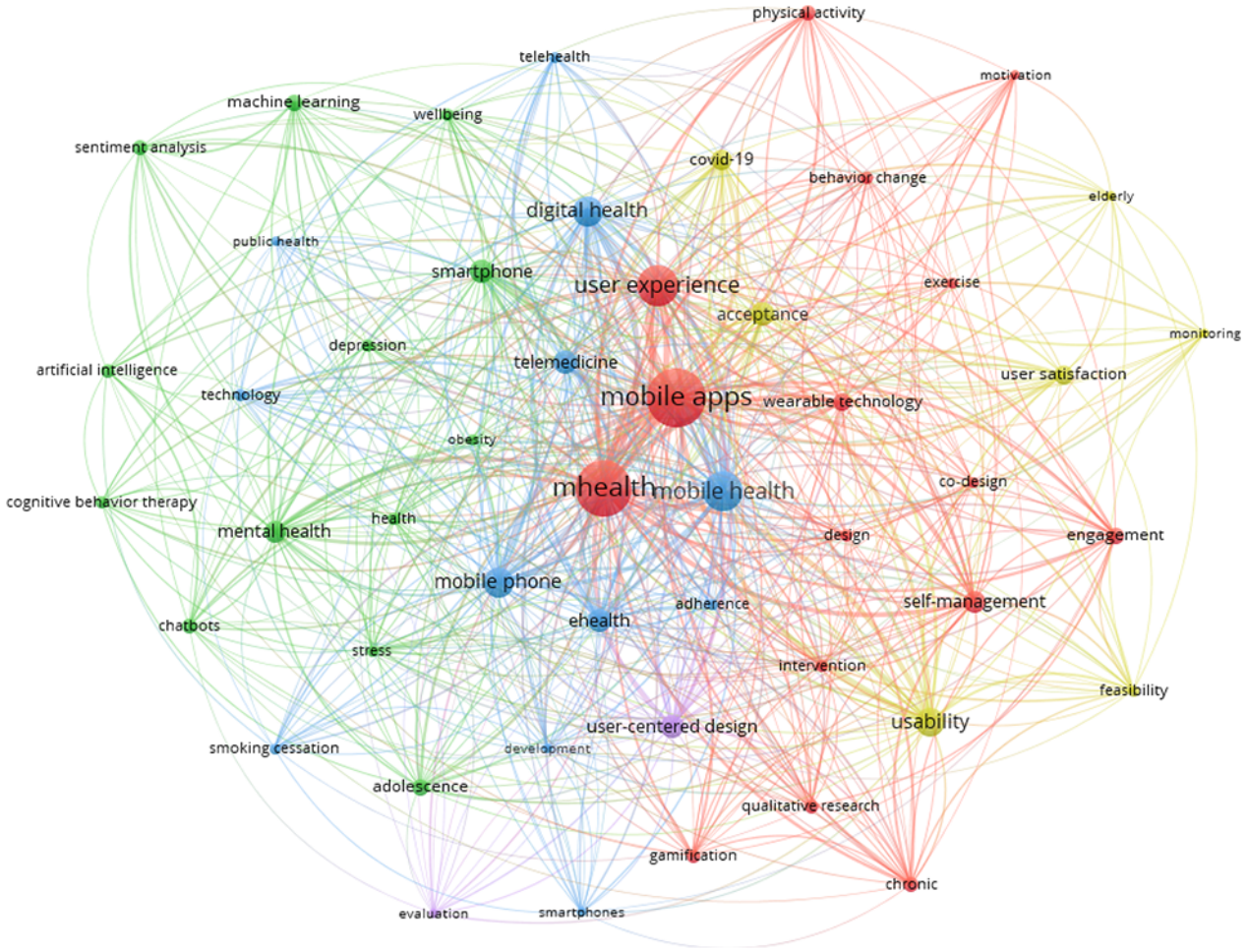
Co-word analysis investigates the relationships between topics by focusing on the content of publications rather than their bibliographical details [69]. This approach allows for a deeper understanding of existing or potential connections among themes within the research. Co-word analysis examines existing or future relationships among topics because it targets publications' content rather than bibliographical units [69]. In this study, co-word analysis is used to examine 2210 author keywords used in 814 UXS-mHealth apps publications from 2007 to 2023.

The tool used to create the co-word network is VOSviewer (version 1.6.20) for author keywords as a unit for co-occurrence analysis. The keywords that meet a threshold of at least three occurrences are 403. The 50 most frequently occurring keywords are selected for network visualization. The network visualization captures their relationships, with node size representing frequency and link thickness indicating co-occurrence frequency within six distinct clusters depicted in Figure 5. In addition, Figure 6 visualizes the author keywords co-occurrence network timeline, illustrating the dynamic evolution of research themes in UXS-mHealth apps over time. For frequency analysis, Multimedia Appendix 7 includes the rank for the 30 most frequent keywords.

This timeline shows when certain keywords began gaining prominence and whether their popularity has grown or diminished. For instance, keywords such as artificial intelligence, machine learning, chatbots, and feasibility may exhibit a steep rise in frequency from 2021, indicating emerging topics that are receiving increased attention from the mHealth research community. In contrast, other keywords, such as telemedicine, may show a decline, suggesting that interest in related topics has waned. For clustering analysis, the network of keywords can be further broken down into thematic clusters. These clusters reveal how topics are grouped and provide insights into the thematic structure of the research field. Figure 5 depicts how the red and blue clusters are tightly connected, indicating specialized areas of study, while others display overlapping themes, such as the purple cluster, suggesting interdisciplinary efforts. Each cluster represents a specific content hotspot within the domain, providing a snapshot of the diverse areas being explored. Overall, the co-word analysis reveals that the six identified clusters provide insights into how digital health technologies are conceptualized, developed, and implemented to address different health needs.

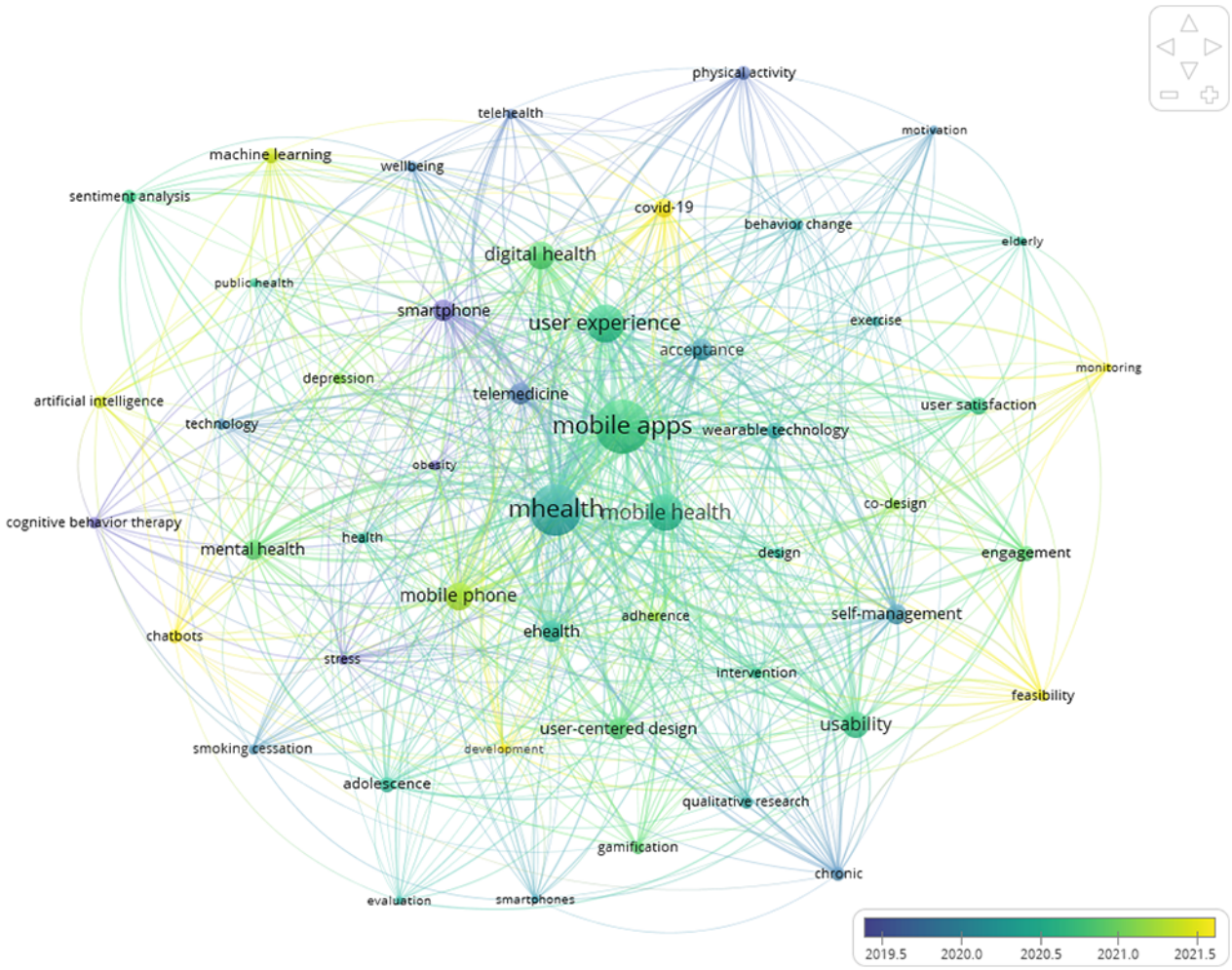
XSL•FO
RenderX

Figure 5. The co-word analysis of author keywords for user experience and satisfaction with mobile health apps from 2007 to 2023 identified the top 50 keywords co-occurring within 6 distinct clusters.



XSL•FO
RenderX

Figure 6. The co-word analysis for the author keyword on user experience and satisfaction with mobile health apps is represented in the timeline network for the top 50 keyword co-occurrences.



Bibliographic Coupling

Bibliographic coupling is a science mapping technique to analyze the relationships among citing publications, assuming 2 publications can be similar if they share references, enabling the grouping of publications into thematic clusters [69]. In this study, the bibliographic coupling analysis conducted on 814 publications for the UXS-mHealth apps from 2007 to 2023 reveals 18 distinct clusters, each representing a group of publications sharing commonly cited references as depicted in Figure 7. The tool used to create the bibliographic coupling network is VOSviewer (version 1.6.20), which is used to analyze bibliographic coupling at the document level, for the full counting method, and for a minimum of zero citations of a document. Normalization was performed using the association strength method. The layout was generated with default values (attraction=2 and repulsion=1). Clustering was conducted with a resolution of 1.00, a minimum cluster size of 1, and with the option to merge small clusters enabled. Among these are the largest clusters (341 publications in red color), followed by the green cluster (242 publications), the blue cluster (218 publications) grouped under the label Patient Engagement, Usability, and Design, the yellow cluster (71 publications), and the purple cluster (56 publications).

The tool used to create network visualization reveals cluster overlap, indicating potential thematic intersections or conceptual similarities across different research streams. This overlapping may arise due to shared foundational literature, similar methodological approaches, or overlapping research topics examined from different disciplinary perspectives. However, the size and structure of these clusters indicate varying degrees of thematic similarity among the grouped publications, suggesting the concentration of research activity in specific thematic areas.

These 5 significant clusters are further analyzed by reviewing the titles, abstracts, and keywords of the top 10 most cited studies in each cluster to uncover the thematic development within each research group. Figure 8 shows the bibliographic coupling timeline network based on the publication year, where clusters 2 and 4 include more recent publications. While Table 5 summarizes clusters of information.

XSL•FO
RenderX

Figure 7. The bibliographic coupling network of 813 publications on user experience and satisfaction with mobile health apps from 2007 to 2023 identifies 18 clusters. Each color represents a cluster, while the node's size represents the total number of citations a publication received. The order of the most significant clusters based on total publications is (red, green, blue, yellow, and purple).

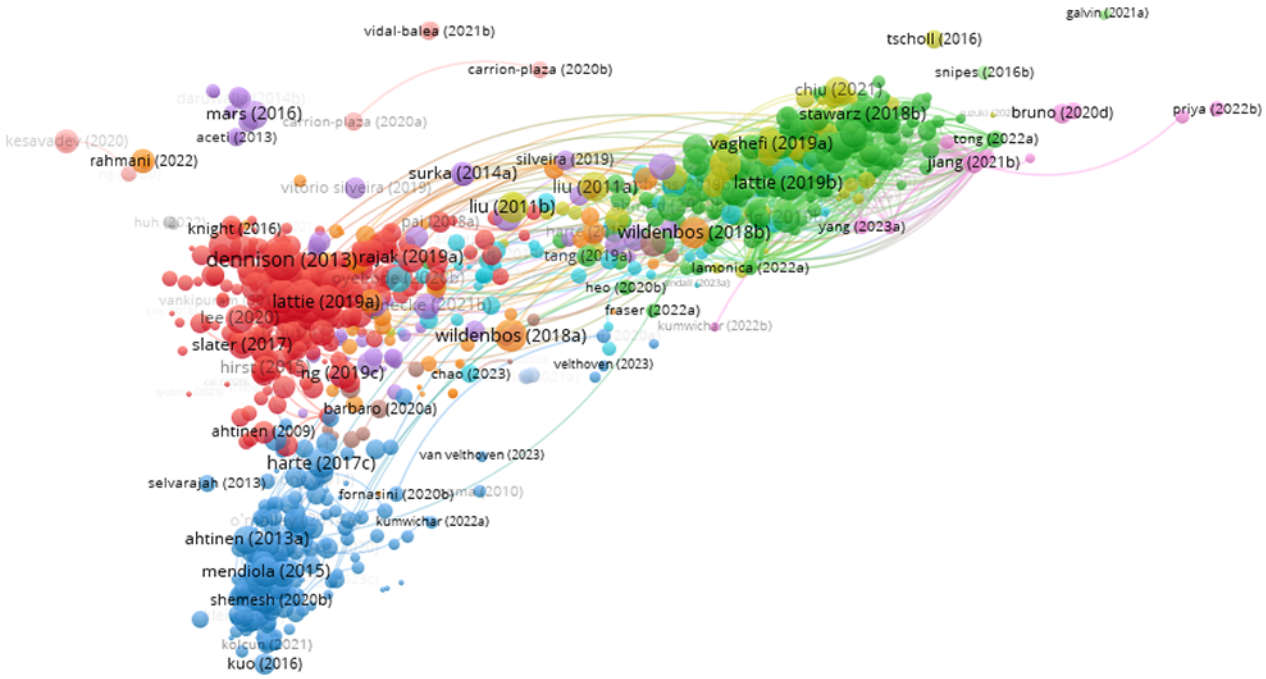
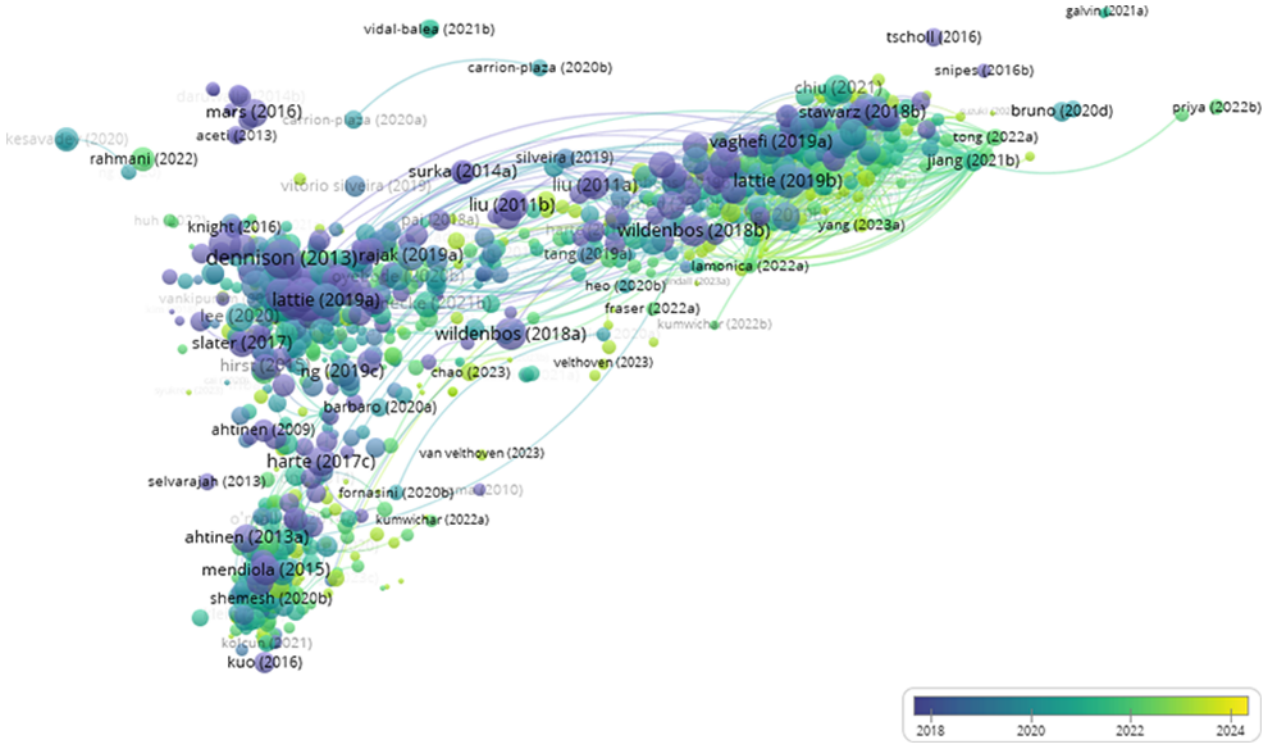


Figure 8. The bibliographic coupling timeline network of 813 publications on user experience and satisfaction with mobile health apps from 2007 to 2023, showing clusters 2 and 4, includes more recent publications than others.



XSL•FO
RenderX

Table . The descriptive summary of the most significant clusters includes total publications (TP), cluster time (years), the top citation and publication year, and the suggested theme labels.

Cluster	Theme	TP	Years	Top citation (year)
1	Evaluating technological elements in using mHealth[a] apps and wearables	341	2008 - 2023	623 (2013)
2	Key factors influencing user engagement and experience or app design features enhancing user engagement	242	2013 - 2023	355 (2019)
3	Evaluation of mHealth app effectiveness and acceptance for patient self-management	218	2010 - 2023	190 (2017)
4	Determinants of mHealth user experience and long-term engagement	75	2011 - 2023	230 (2015)
5	Barriers and enablers of mHealth integration in clinical practice	53	2011 - 2023	131 (2016)

[a]mHealth: mobile health.

Systematic Literature Review

Through a systematic procedure, keyword and thematic analyses were performed. The keyword analysis included the analysis of the co-occurrence of 403 keywords at least 3 times. Keyword analysis provides insights into the most frequently used terms in titles, abstracts, and keywords, thereby offering a high-level view of recurring topics and areas of interest [69]. In addition, thematic analysis based on bibliographic coupling includes 928 articles to further explore the underlying themes and patterns within the research, revealing the conceptual frameworks and areas of scholarly focus [69]. By combining these methods, the following section discusses evolving research trends, thematic analyses, dominant theories, and areas requiring further exploration.

Discussion

Research Trend Analysis

A total of 2 research trends in UXS-mHealth apps are presented: differentiation in using the keywords “user experience” and “user satisfaction” and use of research methodologies.

User Experience and User Satisfaction

User experience and user satisfaction were among the most frequently occurring keywords, reflecting the initial search scope of this study. User experience appeared 218 times with 47 links and a total link strength of 564, whereas user satisfaction had 49 occurrences, 22 links, and a total link strength of 104 (see Figure 5 and Multimedia Appendix 7). These keywords belonged to different clusters despite their conceptual similarities, indicating potential research gaps. Furthermore, user experience focused on themes such as behavioral change, engagement, and interventions, emphasizing sustained interaction via design features. Associated keywords such as “behavior change,” “gamification,” and “mobile apps” highlighted their role in influencing long-term behaviors and improving health outcomes. In contrast, user satisfaction was related to the short-term evaluations of functionality and usability, with keywords such as “acceptance,” “usability,” and “monitoring” reflecting a focus on meeting immediate user expectations. User satisfaction is more transactional and often measured via usability testing or acceptance in specific contexts such as “COVID-19” or “elderly” populations. The weak co-occurrence link between these concepts indicated a gap in the research on mHealth apps and that the efforts used to enhance user experience were not systematically aligned with strategies to improve user satisfaction. This disconnect indicated an opportunity for further research to explore how better design can drive both user engagement and user satisfaction.

Theory-Driven Versus Data-Driven Research

The analysis of keywords and themes for the research on UXS-mHealth apps from 2017 to 2013 revealed that 32.2% (262/813) of publications included the terms “theor*,” “model,” or “survey” in their title, keywords, or abstract, as summarized in Table S1 in Multimedia Appendix 8. Among these studies, the TAM, health belief model (HBM), and Unified Theory of Acceptance and Use of Technology (UTAUT) were most frequently used in 38.9% (102/262), 11.5% (30/262), and 9.5% (25/262) of studies, respectively. The top 10 most cited studies [16,17,74,93-99] that applied the theoretical framework were also reviewed to explore their shared research theme.

These studies primarily focused on the use of mHealth technologies for health monitoring, self-management, and behavior change, particularly for chronic conditions and mental health. They assessed the usability, functionality, and effectiveness of mHealth apps for engaging users in self-care, improving health outcomes, and promoting sustained usage. They applied behavioral frameworks to explore factors influencing user experience, including technical functionality, engagement features, ease of use, and user satisfaction, aiming to inform the design and optimization of health apps that effectively meet user needs. A shared goal of these studies was

XSL•FO
RenderX

to leverage theoretical grounds for evaluating mHealth technologies to create scalable, accessible, and evidence-based mHealth interventions that enhance health monitoring and behavior change.

The timeline network in Figure 6 shows that the keywords "artificial intelligence," "machine learning," and "sentiment analysis" lead the recent keywords list. Thematic analysis also revealed the same results, where data-driven research was conducted using machine learning, artificial intelligence, and sentiment analysis. The analysis of titles, abstracts, and keywords of the data corpus revealed that research on UXS-mHealth apps began incorporating machine learning in 2017, sentiment analysis in 2018, and artificial intelligence in 2019, as indicated by the initial publication years see Table S2 in Multimedia Appendix 8). The terms "machine learning," "sentiment analysis," and "artificial intelligence" were used in 4.18% (34/814), 5.17% (42/814), and 3.08% (25/814) of studies, respectively. The most frequently cited studies related to the terms "machine learning," "sentiment analysis," and "artificial intelligence" are [13 ,100], [101,102], and [103,104], respectively.

To understand the trends in these studies, we reviewed the most cited studies that explored machine learning, sentiment analysis, or artificial intelligence to address various challenges with mHealth apps, particularly those focused on mental health, physical activity, and chronic disease management. In particular, we focused on studies that evaluated and enhanced health-related apps and wearable devices to improve user experience and treatment outcomes. For instance, machine learning and thematic analysis were used to identify barriers and enablers that impacted the effectiveness of mental health apps [105]. Chatbots apply cognitive behavioral therapy principles to support emotional regulation and behavior change in adolescents managing obesity and prediabetes. The research on activity trackers examined predictors of sustained use and behavior modification, highlighting technical issues and user experience as critical determinants. Similarly, the development of automated cough detection systems underscored the potential for artificial intelligence–powered tools to facilitate continuous, objective health monitoring. These studies highlighted the potential of analyzing user feedback and leveraging sentiment analysis and artificial intelligence to enhance health outcomes. By making mHealth apps more engaging, accessible, and responsive to user needs, these approaches can help develop more scalable, cost-effective solutions that address mHealth challenges.

The theory-driven versus data-driven research trend revealed that future studies should continue using a mixed-method approach that combines the strengths of both paradigms. Integrating theoretical and qualitative approaches with data-driven, quantitative methodologies enables a multidimensional analysis of UXS in mHealth applications, fostering a holistic understanding of their design and impact. Furthermore, theoretical insights contribute depth and context, whereas data-driven results offer generalizability and objectivity. This combination can enhance the robustness of findings and facilitate impactful, multifaceted conclusions that bridge gaps across diverse perspectives in the UXS-mHealth apps field while potentially increasing relevance and applicability in future research.

The observed trends also highlight the need for a new methodological triangulation approach that combines the strengths of both paradigms. By integrating quantitative insights with data-driven, qualitative analysis, particularly via text analytics and machine learning algorithms on user reviews, researchers can comprehensively understand the challenges associated with UXS in mHealth apps. Quantitative perspectives provide structured insights, whereas qualitative data from user reviews add depth and context to the results, uncovering patterns and sentiments that might otherwise remain hidden. This methodological integration will strengthen the robustness of findings and support the development of impactful and layered conclusions, thereby connecting diverse perspectives within UXS-mHealth research. Such an approach promises to enhance the relevance and practical applicability of future studies in this field.

Research Thematic Analysis

The 5 largest thematic clusters obtained from the bibliographic analysis included in Table 5 are discussed in the following section. The studies that overlapped across themes were assigned to all relevant themes. The keyword analysis for each cluster is included in Multimedia Appendix 9.

Theme 1: Evaluating Technological Elements in Using mHealth Apps and Wearables

This theme was the largest cluster identified during bibliographic coupling and focused on reviewing technological elements in mHealth apps and wearables designed for monitoring health behavior change, offering mental health support, and physical activity tracking. This cluster evaluated the key factors influencing user engagement, usability, and long-term use of mHealth apps, particularly those designed for mental health, fitness, and self-management purposes. A significant number of studies within this cluster conducted SLRs to confirm findings from existing research and assess the effectiveness of mHealth apps and user satisfaction.

Various methods were employed to explore users' perspectives on mHealth apps and wearables related to behavior change. Qualitative approaches dominated in these studies, and systematic reviews [12,15,19,50,106], interviews [75], and focus groups [107] were the most frequently used techniques. This cluster also comprised studies that used quantitative methods such as surveys [98] and randomized controlled trials [106].

Theoretical frameworks further enriched the analysis, providing a foundation for understanding user interaction with mHealth apps. The research was structured using a combination of TAM, the health information technology acceptance model (HITAM) [74], and UTAUT [98], as well as a combination of the analytical hierarchy process (AHP) and Fuzzy Technique for Order Preference by Similarity to Ideal Solution (Fuzzy TOPSIS) [16].

The corresponding findings were published in several prominent journals in health technology, medical informatics, and social sciences, including *JMIR mHealth and uHealth*, *Journal of*

XSL•FO
RenderX

Medical Internet Research, *PLoS ONE*, *Telemedicine and e-Health*, *Journal of Cardiovascular Nursing*, *Technology in Society*, and *Psychiatric Services*.

Key factors affecting the success of mHealth apps and wearables emerged from several influential studies in this cluster. For instance, app accuracy, security, and efforts required for use were the significant determinants of user engagement [107]. In addition, performance expectancy, hedonic motivations, and price value were the predictors of continued use [98]. The ability of the apps and wearables to track behavior and goals, along with the option to receive advice and information quality, was highly valued by users [16,107]. Other important factors, such as interface design, navigation, notifications, and actionable recommendations, were critical in influencing user engagement [75].

Despite the opportunities created by advances in mHealth technologies, studies in this cluster highlighted several challenges [12,15,74,107]. The majority of studies agreed on the existence of a significant research gap regarding the validation of elements that promoted long-term app use and engagement [12,15,19,50,74,106]. These studies also highlighted the variability in evaluating mHealth apps and wearables, necessitating the development of standardized measures to assess the effectiveness factors on UXS, usability, and engagement [15,16,50].

Future recommendations include exploring hybrid trial designs, developing user-centered designs, and addressing the long-term efficacy and scalability of mHealth interventions [55,57,58]. Overall, these studies highlighted the opportunities and challenges associated with mHealth apps, particularly in sustaining user engagement, improving usability, and ensuring the long-term effectiveness of health behavior change interventions [54,55,58]. Promising future research directions include the development of more robust frameworks for evaluating the long-term impact of these technologies on health outcomes and user behavior. The integration of emerging technologies such as artificial intelligence and machine learning into mHealth solutions presents new opportunities for enhancing personalization and improving sustained engagement.

Theme 2: Key Factors Influencing User Engagement and User Experience or App Design Features Enhancing User Engagement

Thematic synthesis of studies highlighted shared objectives, methodologies, and findings that focused on improving mHealth across various domains, particularly mental health and chronic disease management. A recurring aim in these studies was to explore the effectiveness, user engagement, and design features of mHealth apps, focusing on enhancing user experience (UX) and sustained use. For instance, mental health apps of [15,17-19], medication adherence [12], physical activity apps [13], and chronic disease management [6,14] were focused on in the cited studies. The publication sources used for this cluster were drawn from leading academic journals and conferences that offered diverse and valuable perspectives on mHealth research. These included *JMIR mHealth and uHealth*, *Diagnostics*, *The International Journal of Engineering and Technology (UAE)*, and the 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud).

Most studies in this cluster predominantly used SLR to evaluate various types of mHealth apps, ranging from mental health and medication adherence to chronic disease management [6,12,15,19]. mHealth apps were reviewed to thematically analyze their features [12,16], and qualitative thematic analysis was performed to analyze app user reviews and extract engagement features [18]. Quantitative data [13] collected via a questionnaire and an activity tracker were analyzed using random forest—a machine learning technique—to identify the determinants of long-term engagement. Theoretical frameworks were only represented by Jeffrey et al [14] using semistructured interviews based on the TAM and HITAM to assess the factors that influenced the use of diabetes apps and [16]; they also used the AHP and fuzzy TOPSIS framework to evaluate mHealth apps.

Findings across these studies consistently revealed that factors such as usability, personalization, and user satisfaction were critical for the success of mHealth apps. Rajak et al and Hermsen et al [13,16] emphasized the significance of evaluating mHealth apps using multicriteria decision-making frameworks and machine learning techniques to account for diverse user needs and preferences. Their studies collectively advocated for ongoing research to further refine app functionalities, with a particular focus on their real-world implementation and obtaining continuous user feedback to enhance the overall effectiveness and engagement of mHealth interventions [18,108].

Across these studies, common concepts emerged regarding the importance of specific app features such as reminders and educational content to support user adherence [12]. The role of user satisfaction, ease of use, and user interface design was emphasized by Rajak et al and Alqahtani et al [16,18], who found that customizable, user-friendly apps with varied content and functionalities considerably enhance engagement. In contrast, poor usability and lack of personalization were frequently cited reasons for app abandonment. Adaptable design features tailored to diverse user needs were reported as a crucial factor for maintaining a positive UX [17]. In the context of chronic disease management, users favored apps with educational content, monitoring, and tracking functionalities [6,14]; the critical determinants of sustained engagement included age, goal setting, and UX [13]. A key finding across the studies was that design features such as personalization, reinforcement, and compelling message presentation were vital for maintaining user engagement [8,18].

Although previous studies have identified key design features in mHealth apps, more rigorous, evidence-based research is required to validate the effectiveness of these features. These studies [12,15,19] emphasized the importance of establishing standardized metrics for evaluating user engagement and app efficacy to improve overall user satisfaction. Stawarz et al [17] advocated for developing evidence-based approaches to improve UX in mental health apps. In contrast, Hermsen et al [13] underscored the critical need to focus on technical performance and UX in activity tracking technologies. Alessa et al [6] called

XSL•FO
RenderX

for large-scale clinical studies to assess the impact of mHealth apps on chronic disease management, particularly regarding their potential to reduce blood pressure. In addition, Wei et al [8] highlighted the need for more robust quantitative assessments to better understand the correlation between app design features and sustained user engagement.

In summary, future research should prioritize longitudinal studies that examine the impact of personalized and adaptive design features on user retention, health outcomes, and behavioral changes across diverse user demographics. In addition, interdisciplinary research that integrates fields such as behavioral science, human-computer interaction, and health care must be conducted to develop a comprehensive framework for mHealth app design. By investigating the potential of continuously incorporating adaptive machine learning algorithms to tailor app features to individual user preferences and needs, critical insights into optimizing UX can be obtained. Furthermore, research exploring the scalability of mHealth apps in underrepresented or underserved populations can further contribute to the development of inclusive, effective digital health solutions.

Theme 3: Evaluation of mHealth App Effectiveness and Acceptance for Patient Self-Management

The analysis of studies focusing on the usefulness and effectiveness of mobile apps for patient management or chronic disease monitoring revealed a diverse set of themes and influencing factors. Researchers primarily used methodologies such as qualitative interviews, systematic and narrative reviews, thematic analysis, and quantitative techniques such as questionnaires, regression analyses, and randomized control trials to evaluate mHealth apps in the literature reviewed.

These studies were published in health research journals, each with a different perspective. *JMIR mHealth and uHealth* focus on patient-centric mHealth apps and technology usability. Journal of Medical Internet Research offers a broad, interdisciplinary approach to eHealth, emphasizing theoretical and clinical apps. *JMIR Research Protocols* are methodologically oriented, highlighting research transparency and design considerations. *JAMA Internal Medicine* provides a clinical and health policy perspective, prioritizing robust evidence for clinical effectiveness and health outcomes. Each journal contributes uniquely to understanding mobile health interventions, reflecting their specific editorial focus and research priorities.

The key themes identified from the reviewed studies can be broadly categorized into technical, health-related, and usability factors. Mendiola et al identified five core factors affecting the performance of mHealth apps: plan or order, data export, usability, cost, and tracking features [4]. Similarly, design implications that focused on enhancing everyday life, providing user flexibility, and supporting self-improvement were emphasized [5]. Wei et al synthesized findings from 35 studies and categorized critical success factors into personalization, reinforcement, communication, navigation, credibility, message presentation, and interface esthetics [8]. The standardized software usability measurement inventory that uncovers five user satisfaction dimensions, that is, efficiency, effectiveness, helpfulness, controllability, and learnability, was applied [9]. Other studies expanded the thematic scope to include technical (eg, ease of use and data-related issues), health-related (eg, integration into the patient journey and insurance status), and social factors (eg, demographic and cultural aspects) [62]. Crafoord et al [109] explored patient engagement via symptom reporting and highlighted the need for continuous feedback and comprehensive information to support patients throughout their treatment journey.

Several studies mentioned patient feedback and continuous support [63]; however, only a few studies have addressed how mHealth apps fully integrate with patients' broader healthcare journeys, particularly regarding long-term adherence and coordination with healthcare providers. The influence of demographic and cultural factors [62] remains underexplored, particularly in their effect on the usability and acceptance of mHealth apps across diverse populations. Future research should address such gaps by focusing on the influence of cultural and demographic diversity on app design and user satisfaction, ensuring personalization in mHealth innovations.

Mendiola et al and Wei et al [4,8] explored technical aspects such as data export and navigation; however, data privacy and security remain unexplored. Moreover, the influence of these factors on user trust and satisfaction, which are critical concerns in the modern healthcare technology landscape, has not been sufficiently studied. This gap should be addressed in future research by focusing on trust and data security concerns, which are becoming increasingly critical as healthcare apps become more complex and data-driven. Understanding how to build and maintain user trust while ensuring robust data security measures will be essential for successfully adopting and using these apps [6].

Theme 4: Determinants of mHealth UX and Long-Term Engagement

The studies within this cluster collectively explored various factors that influenced adoption, UX, and long-term engagement with mHealth apps. These determinants shape the perception of users regarding the utility of mHealth apps and their intention to continue using them. Common concepts identified include user satisfaction, perceived usefulness, personalization, technical functionality, and social influence. A critical finding across several studies was the importance of user satisfaction in driving the continuous use of mHealth apps. Liu et al [10] demonstrated the impact of unique smartphone features, such as context awareness and portability, on enhancing user satisfaction by improving app functionality. Similarly, Vaghefi and Tulu [75] underscored that UX is pivotal for maintaining user engagement, particularly in interface design and navigation. Birkmeyer et al [110] highlighted how personalization, social networking, and app design contribute to user satisfaction, directly correlating with continuance intentions and word-of-mouth recommendations.

Furthermore, perceived usefulness was a recurring determinant of whether users will continue to engage with mHealth apps. Factors such as performance expectancy and the confirmation of expectations consistently emerged as drivers of sustained app use. Chiu et al [94] emphasized that emotional and temporal

XSL•FO
RenderX

investments in an app contribute to a sense of commitment and encourage its continuous usage among users. Chiu et al and Yuan et al [94,98] found that users' beliefs regarding the utility of health apps considerably impacted their behavioral intentions. Personalization was another central theme that influenced long-term engagement with mHealth apps in their studies. Biduski et al and Lappalainen et al [20,111] highlighted how personalized features, such as goal setting, tailored feedback, and user-system fit, are crucial for retaining users over time. They also reported that customizing app functionalities to meet individual needs fostered a sense of relevance and enhances user engagement. Al Ayubi et al [112] suggested that offering diverse physical activities within mHealth apps increased user motivation by aligning the content with their specific health objectives.

Several studies identified technical challenges and usability issues as barriers to long-term app use. Biduski et al [20] emphasized that software glitches and repetitive data-entry processes diminished user motivation. Ehn et al [113] observed similar challenges, noting that older adults struggled with the complexity of activity monitors that was exacerbated by a nonintuitive design. These usability issues must be addressed by simplifying interfaces and improving technical reliability to maintain user engagement, particularly among less tech-savvy populations.

The social dimension of mHealth app use still plays a role in shaping user behavior. Al Ayubi et al [112] demonstrated that social networking features motivated users to engage in physical activity through peer support and competition. However, Yuan et al [98] found that social influence posed minimal impact on the continued use of health apps, suggesting that different social norms may affect users differently. Future research can investigate these nuances to better understand the contribution of social factors to sustained app usage.

The studies in this cluster employed diverse research methodologies to explore various factors that influenced user engagement, satisfaction, and long-term use of mHealth apps. Quantitative surveys, grounded in theoretical models such as UTAUT2 and expectation–confirmation model, were used to analyze factors such as performance expectancy, perceived usefulness, and continuance intention; these surveys provided generalizable insights into user behavior [94,98]. In-depth qualitative interviews were also conducted to capture detailed user experiences, motivations, and barriers [75,113], which allowed for a more nuanced understanding of individual behaviors. Several studies [110] employed mixed-method approaches combining quantitative data with qualitative insights to offer a more holistic perspective on the use of mHealth apps. Longitudinal studies [75] tracked user engagement over time, providing valuable insights into the evolution of app usage. Biduski et al [20] introduced an innovative assessment method using in-app embedded conversation-based questionnaires, enabling real-time data collection on UX throughout different stages of app interaction. The methodologies utilized across the studies in this cluster reflected the complexity of investigating user engagement with mHealth apps. Quantitative surveys provided generalizable findings on user behavior, whereas qualitative interviews offered deeper insights into individual experiences. Mixed method approaches and longitudinal studies added further depth by tracking the evolution of user engagement. Together, these methodologies provided a comprehensive understanding of the determinants influencing long-term success of mHealth apps.

The studies in this cluster were published in various journals and cover different aspects of mHealth research, including technical and behavioral areas. For instance, the *Journal of Systems and Software* focuses on the technical foundations of mHealth app development, emphasizing innovations in software design and systems engineering. Telemedicine and eHealth specialize in integrating telecommunication technologies with health care delivery, particularly in underserved areas. *JMIR mHealth and uHealth*—one of the leading journals in this field—frequently publishes studies on app usability, user engagement, and health outcomes. Other notable journals, such as *Information Technology & People*, explore the socio-technical dynamics of mHealth apps. *Computers in Human Behavior* focuses on the impact of technology on user behavior.

Theme 5: Barriers and Enablers of mHealth Integration in Clinical Practice

Thematic analysis of 10 most cited publications in this cluster highlighted several critical factors affecting the successful adoption and integration of mHealth apps in clinical settings. Usability and design emerged as consistent themes, with studies identifying poor user interfaces and complexity as significant barriers to the adoption of mHealth apps across various contexts, from chronic disease management to surgical wound monitoring. Multiple studies highlighted the need for seamless integration into existing clinical workflows and identified challenges in workflow disruption, duplication of tasks, and time constraints faced by healthcare professionals. Concerns over privacy and security also featured prominently, where secure data transmission and regulatory compliance appeared as crucial factors for user confidence. Future work in this area should devise guidelines, improve app design, and provide tailored training modules to address the diverse needs of both healthcare providers and patients while ensuring that mHealth apps are adaptable, reliable, and accessible across different technological and cultural contexts.

The analysis further revealed similarities and differences across the research methods and samples used in these studies and the types of journals in which they were published. Research methods varied from literature reviews [114,115] and qualitative evaluations [116] to usability testing [117], mixed methods studies [118,119], focused groups [120], and surveys [121-123]. The samples also differed considerably across studies. Most studies in this cluster focused on involving users or physicians in the development and testing of mHealth apps, with an emphasis on creating user-centered designs [116,118-123].

The focus areas of journals in which these studies were published also differed. *JMIR mHealth and uHealth* and the *Journal of Medical Internet Research* frequently publish studies focusing on mHealth technologies, patient engagement, and telehealth innovations. *BMC Medical Informatics and Decision Making* and the *International Journal of Medical Informatics* cater to broader medical informatics and decision-making

XSL•FO
RenderX

processes in clinical settings. *Studies in Health Technology and Informatics* are more focused on advancing health technology, which might offer a more technology-centered lens than patient-focused journals such as *Healthcare*. These differences in journal focus reflect the varied perspectives on mHealth adoption—from usability and design to workflow integration and patient-provider relationships, emphasizing the need for interdisciplinary research approaches in advancing mHealth technologies.

Limitations

The bibliometric analysis provides a structured overview of the research landscape, and the SLR offers deeper insights into research trends and knowledge gaps, guiding future research directions by combining breadth with depth. However, their limitations are particularly notable in dynamic and interdisciplinary fields like mHealth apps, where technology and knowledge evolve rapidly. Bibliometric analysis may struggle to capture emerging studies, while SLRs are restricted by the preselected sample of articles defined by inclusion and exclusion criteria, which can overlook relevant contributions outside those boundaries. These constraints highlight the importance of interpreting results with caution and updating analyses regularly to ensure relevance. Furthermore, while most of the studies might originate from countries with well-developed digital health infrastructures, they reflect the current global distribution of research and investment. This may limit the generalizability of our findings to underresourced settings, where barriers such as limited internet access, lower digital literacy, and weaker health system capacities are more prevalent. Further research from developing countries is needed to provide a more comprehensive and balanced understanding of digital health implementation.

Conclusion

This study represents the first bibliometric analysis of user experience and satisfaction with mHealth apps (UXS-mHealth apps). Using a comprehensive data corpus sourced from Scopus and Web of Science, we systematically analyzed the literature through bibliometric performance analysis and science mapping techniques. Our methodology used a variety of data processing and visualization tools, including VOS Viewer, Excel, OriginLab, and SiteSpace, to identify influential publications, authors, countries, and research sources.

The analysis highlights the evolution and emergence of key research topics in this domain by examining annual research trends, co-citation analysis, co-word analysis, and bibliographic coupling. Over a 16-year period (2007 - 2023), we analyzed 814 unique publications authored by 4870 researchers from 81 countries and 1948 organizations, published across 351 high-impact journals and prominent conferences. This study provides a foundational understanding of the bibliographic landscape in UXS-mHealth apps by identifying major research clusters, influential works, and emerging topics, providing evidence-based guidance for researchers, developers, and health informatics practitioners.

Furthermore, research trends and themes in UXS-mHealth were systematically reviewed using co-word analysis and bibliographic coupling. This approach provided a comprehensive overview of the existing literature by addressing the limitations of previous studies, which often focused on specific applications, limited time frames, or restricted database coverage. Results revealed key thematic areas, emerging trends, and gaps in mHealth UX research, highlighting the interdisciplinary nature of the field and the need for broader methodological integration. The analysis of research trends revealed 2 trends: the differentiation in using user experience and user satisfaction keywords and the research methodology used in the domain, highlighting the need to incorporate user-generated data into mixed-method research.

Thematic analysis was performed across five most significant themes: (Theme 1) evaluating technological elements in using mHealth apps and wearables, (Theme 2) key factors influencing user engagement and experience or app design features enhancing user engagement, (Theme 3) evaluation of mHealth app effectiveness and acceptance for patient self-management, (Theme 4) determinants of mHealth UX and long-term engagement, and (Theme 5) barriers and enablers of mHealth integration in clinical practice.

The research trends and 5 themes identified in this review explored critical aspects of technology use, user engagement, and clinical integration. While all five themes overlapped, particularly themes 1, 2, and 4, each of them focuses on distinct elements that help delineate their contributions to the overall understanding of mHealth apps: technological evaluation (Theme 1), design features for engagement (Theme 2), patient usability (Theme 3), long-term engagement factors (Theme 4), and clinical integration (Theme 5). By analyzing the identified trends and themes, valuable insights into future research on UXS-mHealth can be gained.

The findings suggest the importance of prioritizing user-centered design and implementing a more engaging experience for mHealth apps. Furthermore, future research should prioritize expanding the scope of UX evaluation by incorporating diverse user populations, longitudinal studies, and emerging technologies such as artificial intelligence and personalized interventions. By integrating insights from interdisciplinary perspectives such as human-computer interaction, behavioral sciences, and health care informatics, the understanding of user needs and app effectiveness can be enhanced. A more standardized framework for assessing UX in mHealth apps is also a gap that needs to be addressed to facilitate comparability across studies and improve app design to maximize user engagement and health outcomes. While this analysis provides valuable insights into the fundamental features of the field, it does not explore the underlying causes or offer predictive capabilities. Future research should integrate these findings with inferential or exploratory analyses to uncover deeper insights.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the Saudi Electronic University for its support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of database sources and proportion of publication types in the bibliometric analysis.
[PDF File, 99 KB - mhealth_v13i1e75909_app1.pdf]

Multimedia Appendix 2
Chart of the leading countries' contributions.
[PDF File, 167 KB - mhealth_v13i1e75909_app2.pdf]

Multimedia Appendix 3
Chart of the leading research entities' contributions.
[PDF File, 301 KB - mhealth_v13i1e75909_app3.pdf]

Multimedia Appendix 4
Citation analysis summary showing the annual most cited publications.
[PDF File, 105 KB - mhealth_v13i1e75909_app4.pdf]

Multimedia Appendix 5
The co-citation network and citation burst.
[PDF File, 222 KB - mhealth_v13i1e75909_app5.pdf]

Multimedia Appendix 6
The coauthorship analysis figures and the authors' collaboration summary.
[PDF File, 353 KB - mhealth_v13i1e75909_app6.pdf]

Multimedia Appendix 7
Co-words frequency analysis.
[PDF File, 278 KB - mhealth_v13i1e75909_app7.pdf]

Multimedia Appendix 8
Theory-driven versus data-driven research on UXS-mHealth apps.
[PDF File, 140 KB - mhealth_v13i1e75909_app8.pdf]

Multimedia Appendix 9
Keyword analysis for the 5 largest clusters.
[PDF File, 845 KB - mhealth_v13i1e75909_app9.pdf]

Checklist 1
PRISMA checklist.
[PDF File, 200 KB - mhealth_v13i1e75909_app10.pdf]

References

1. Istepanian RSH, Al-Anzi T. m-Health 2.0: New perspectives on mobile health, machine learning and big data analytics. Methods 2018 Dec 1;151:34-40. [doi: 10.1016/j.ymeth.2018.05.015] [Medline: 29890285]
2. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
3. Harte R, Glynn L, Rodríguez-Molinero A, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8. [doi: 10.2196/humanfactors.5443] [Medline: 28302594]

XSL•FO
RenderX

4. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40. [doi: 10.2196/mhealth.4283] [Medline: 25972309]
5. Ahtinen A, Mattila E, Välkkynen P, et al. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR Mhealth Uhealth 2013 Jul 10;1(2):e11. [doi: 10.2196/mhealth.2596] [Medline: 25100683]
6. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. JMIR Mhealth Uhealth 2018 Jul 23;6(7):e10723. [doi: 10.2196/10723] [Medline: 30037787]
7. Jaana M, Sherrard H, Paré G. A prospective evaluation of telemonitoring use by seniors with chronic heart failure: Adoption, self-care, and empowerment. Health Informatics J 2019 Dec;25(4):1800-1814. [doi: 10.1177/1460458218799458] [Medline: 30247080]
8. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 9;22(12):e21687. [doi: 10.2196/21687] [Medline: 33295292]
9. O'Malley G, Dowdall G, Burls A, Perry IJ, Curran N. Exploring the usability of a mobile app for adolescent obesity management. JMIR Mhealth Uhealth 2014 Jun 27;2(2):e29. [doi: 10.2196/mhealth.3262] [Medline: 25098237]
10. Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: A developer's perspective. Journal of Systems and Software 2011 Nov;84(11):2022-2033. [doi: 10.1016/j.jss.2011.06.049]
11. Istepanian RS, Woodward B. M-Health: Fundamentals and Applications, 1st edition: John Wiley & Sons; 2016.
12. Ahmed I, Ahmad NS, Ali S, et al. Medication adherence apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62. [doi: 10.2196/mhealth.6432] [Medline: 29549075]
13. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164. [doi: 10.2196/mhealth.7311] [Medline: 29084709]
14. Jeffrey B, Bagala M, Creighton A, et al. Mobile phone applications and their use in the self-management of Type 2 Diabetes Mellitus: a qualitative study among app users and non-app users. Diabetol Metab Syndr 2019;11:84. [doi: 10.1186/s13098-019-0480-4] [Medline: 31636719]
15. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 1;70(7):538-544. [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
16. Rajak M, Shaw K. Evaluation and selection of mobile health (mHealth) applications using AHP and fuzzy TOPSIS. Technol Soc 2019 Nov;59:101186. [doi: 10.1016/j.techsoc.2019.101186]
17. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 6;20(6):e10120. [doi: 10.2196/10120] [Medline: 29875087]
18. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066. [doi: 10.1177/1460458219896492] [Medline: 31920160]
19. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869. [doi: 10.2196/12869] [Medline: 31333198]
20. Biduski D, Bellei EA, Rodriguez JPM, Zaina LAM, De Marchi ACB. Assessing long-term user experience on a mobile health application through an in-app embedded conversation-based questionnaire. Comput Human Behav 2020 Mar;104:106169. [doi: 10.1016/j.chb.2019.106169]
21. Wang T, Wang W, Feng J, Fan X, Guo J, Lei J. A novel user-generated content-driven and Kano model focused framework to explore the impact mechanism of continuance intention to use mobile APPs. Comput Human Behav 2024 Aug;157:108252. [doi: 10.1016/j.chb.2024.108252]
22. Yin S, Cai X, Wang Z, Zhang Y, Luo S, Ma J. Impact of gamification elements on user satisfaction in health and fitness applications: A comprehensive approach based on the Kano model. Comput Human Behav 2022 Mar;128:107106. [doi: 10.1016/j.chb.2021.107106]
23. Statista. Market insights advertising & media download - app - worldwide. 2024. URL: https://www.statista.com/outlook/amo/app/worldwide [accessed 2025-10-08]
24. Statista. Market insights digital app - health & fitness - worldwide. URL: https://www.statista.com/outlook/dmo/app/health-fitness/worldwide [accessed 2025-10-08]
25. Statista. Market insights advertising & media app - medical -worldwide. URL: https://www.statista.com/outlook/amo/app/medical/worldwide [accessed 2025-10-08]
26. Papanicolas I, Rajan D, Karanikolos M, Soucat A, Figueras J. Health system performance assessment: a framework for policy analysis HPSN 57. WHO. URL: https://www.who.int/publications/i/item/9789240042476 [accessed 2025-10-08]
27. Lemon C, Huckvale K, Carswell K, Torous J. A narrative review of methods for applying user experience in the design and assessment of mental health smartphone interventions. Int J Technol Assess Health Care 2020;36(1):64-70. [doi: 10.1017/S0266462319003507] [Medline: 31973787]

XSL•FO
RenderX

28. Cliffe B, Tingley J, Greenhalgh I, Stallard P. mHealth interventions for self-harm: scoping review. J Med Internet Res 2021 Apr 30;23(4):e25140. [doi: 10.2196/25140] [Medline: 33929329]
29. Kolcun JPG, Ryu WHA, Traynelis VC. Systematic review of telemedicine in spine surgery. J Neurosurg Spine 2021 Feb 1;34(2):161-170. [doi: 10.3171/2020.6.SPINE20863] [Medline: 33126219]
30. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digit Med 2023 Dec 19;6(1):236. [doi: 10.1038/s41746-023-00979-5] [Medline: 38114588]
31. Malloy J, Partridge SR, Kemper JA, Braakhuis A, Roy R. Co-design of digital health interventions with young people: A scoping review. Digit Health 2023;9:20552076231219117. [doi: 10.1177/20552076231219117] [Medline: 38107986]
32. Wu W, Graziano T, Salner A, et al. Acceptability, effectiveness, and roles of mHealth applications in supporting cancer pain self-management: integrative review. JMIR Mhealth Uhealth 2024 Jul 18;12(1):e53652. [doi: 10.2196/53652] [Medline: 39024567]
33. Osborne EL, Ainsworth B, Hooper N, Atkinson MJ. Experiences of using digital mindfulness-based interventions: rapid scoping review and thematic synthesis. J Med Internet Res 2023 Sep 28;25:e44220. [doi: 10.2196/44220] [Medline: 37768709]
34. Opie JE, Vuong A, Welsh ET, et al. Outcomes of best-practice guided digital mental health interventions for youth and young adults with emerging symptoms: part I. a systematic review of socioemotional outcomes and recommendations. Clin Child Fam Psychol Rev 2024 Jun;27(2):424-475. [doi: 10.1007/s10567-024-00469-4] [Medline: 38489101]
35. Gryglewicz K, Orr VL, McNeil MJ, et al. Translating suicide safety planning components into the design of mHealth app features: systematic review. JMIR Ment Health 2024 Mar 28;11(1):e52763. [doi: 10.2196/52763] [Medline: 38546711]
36. Luo M, Yue Y, Du N, Xiao Y, Chen C, Huan Z. Needs for mobile and internet-based psychological intervention in patients with self-injury and suicide-related behaviors: a qualitative systematic review. BMC Psychiatry 2024;24(1):26. [doi: 10.1186/s12888-023-05477-2]
37. Chiu YH, Lee YF, Lin HL, Cheng LC. Exploring the role of mobile apps for insomnia in depression: systematic review. J Med Internet Res 2024 Oct 18;26:e51110. [doi: 10.2196/51110] [Medline: 39423009]
38. Fajnerova I, Hejtmánek L, Sedlák M, Jablonská M, Francová A, Stopková P. The journey from nonimmersive to immersive multiuser applications in mental health care: systematic review. J Med Internet Res 2024 Nov 7;26:e60441. [doi: 10.2196/60441] [Medline: 39509153]
39. Greenway FT, Weal M, Palmer-Cooper EC. Hybrid mHealth care: Patient perspectives of blended treatments for psychosis. A systematic review. Schizophr Res 2024 Dec;274:1-10. [doi: 10.1016/j.schres.2024.08.017] [Medline: 39244945]
40. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. J Am Med Inform Assoc 2020 Jun 1;27(6):939-945. [doi: 10.1093/jamia/ocaa029] [Medline: 32357368]
41. Fang YE, Zhang Z, Wang R, et al. Effectiveness of eHealth smoking cessation interventions: systematic review and meta-analysis. J Med Internet Res 2023 Jul 28;25:e45111. [doi: 10.2196/45111] [Medline: 37505802]
42. Giannattasio A, Sestino A, Baima G. Crafting a healthier future: exploring the nexus of product design, digital innovations and dynamic marketing for obesity prevention. A literature review. BFJ 2024 Jun 17;126(7):2668-2685. [doi: 10.1108/BFJ-10-2023-0897]
43. O'Neill M, Houghton C, Crilly G, Dowling M. A qualitative evidence synthesis of users' experience of mobile health applications in the self-management of type 2 diabetes. Chronic Illn 2022 Mar;18(1):22-45. [doi: 10.1177/1742395320983877] [Medline: 33583207]
44. Pernencar C, Saboia I, Dias JC. How far can conversational agents contribute to IBD patient health care-a review of the literature. Front Public Health 2022;10:862432. [doi: 10.3389/fpubh.2022.862432] [Medline: 35844879]
45. Rincon E, Monteiro-Guerra F, Rivera-Romero O, Dorronzoro-Zubiete E, Sanchez-Bocanegra CL, Gabarron E. Mobile phone apps for quality of life and well-being assessment in breast and prostate cancer patients: systematic review. JMIR Mhealth Uhealth 2017 Dec 4;5(12):e187. [doi: 10.2196/mhealth.8741] [Medline: 29203459]
46. Safiee L, Rough DJ, Whitford H. Barriers to and facilitators of using eHealth to support gestational diabetes mellitus self-management: systematic literature review of perceptions of health care professionals and women with gestational diabetes mellitus. J Med Internet Res 2022 Oct 27;24(10):e39689. [doi: 10.2196/39689] [Medline: 36301613]
47. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017 Dec 12;19(12):e406. [doi: 10.2196/jmir.8888] [Medline: 29233804]
48. Suresh Kumar S, Connolly P, Maier A. Considering user experience and behavioral approaches in the design of mHealth interventions for atrial fibrillation: systematic review. J Med Internet Res 2024 Oct 4;26:e54405. [doi: 10.2196/54405] [Medline: 39365991]
49. Woodley SJ, Moller B, Clark AR, et al. Digital technologies for women's pelvic floor muscle training to manage urinary incontinence across their life course: scoping review. JMIR Mhealth Uhealth 2023 Jul 5;11:e44929. [doi: 10.2196/44929] [Medline: 37405818]

XSL•FO
RenderX

50. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329. [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
51. Moorthy P, Weinert L, Schüttler C, et al. Attributes, methods, and frameworks used to evaluate wearables and their companion mHealth apps: scoping review. JMIR Mhealth Uhealth 2024 Apr 5;12(1):e52179. [doi: 10.2196/52179] [Medline: 38578671]
52. Gabarron E, Larbi D, Rivera-Romero O, Denecke K. Human factors in AI-driven digital solutions for increasing physical activity: scoping review. JMIR Hum Factors 2024 Jul 3;11:e55964. [doi: 10.2196/55964] [Medline: 38959064]
53. Tindall LN, Xavier NA. Innovations in diabetes device training: a scoping review. Endocr Pract 2023 Oct;29(10):803-810. [doi: 10.1016/j.eprac.2023.05.012] [Medline: 37290557]
54. Yeo VA, Mi Y, Kwok KT. Factors affecting adoption of digital contact tracing during the COVID-19 pandemic: a literature review. J Public Health Emerg 2022;6:23-23. [doi: 10.21037/jphe-22-11]
55. Zang S, Zhang X, Xing Y, Chen J, Lin L, Hou Z. Applications of social media and digital technologies in COVID-19 vaccination: scoping review. J Med Internet Res 2023 Feb 10;25:e40057. [doi: 10.2196/40057] [Medline: 36649235]
56. Rosic T, Petrina N, Baysari M, Ritchie A, Poon SK. Patient and clinician use characteristics and perceptions of pulse oximeter use: A scoping review. Int J Med Inform 2022 Jun;162:104735. [doi: 10.1016/j.ijmedinf.2022.104735] [Medline: 35325661]
57. Oyibo K, Sahu KS, Oetomo A, Morita PP. Factors influencing the adoption of contact tracing applications: systematic review and recommendations. Front Digit Health 2022;4:862466. [doi: 10.3389/fdgth.2022.862466] [Medline: 35592459]
58. Tang K, Gerling K, Chen W, Geurts L. Information and communication systems to tackle barriers to breastfeeding: systematic search and review. J Med Internet Res 2019 Sep 27;21(9):e13947. [doi: 10.2196/13947] [Medline: 31573903]
59. Rienzo A, Cubillos C. Playability and player experience in digital games for elderly: a systematic literature review. Sensors (Basel) 2020 Jul 16;20(14):3958. [doi: 10.3390/s20143958] [Medline: 32708679]
60. Simon JDHP, Hooijman IS, Van Gorp M, et al. Digital health tools for pain monitoring in pediatric oncology: a scoping review and qualitative assessment of barriers and facilitators of implementation. Support Care Cancer 2023 Feb 21;31(3):175. [doi: 10.1007/s00520-023-07629-2] [Medline: 36802278]
61. Schulze A, Lindemann AK, Brand F, et al. Mobile apps aimed at preventing and handling unintentional injuries in children aged <7 years: systematic review. Interact J Med Res 2023 Sep 6;12:e45258. [doi: 10.2196/45258] [Medline: 37672312]
62. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
63. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e14376. [doi: 10.2196/14376] [Medline: 32319965]
64. Wang T, Giunti G, Melles M, Goossens R. Digital patient experience: umbrella systematic review. J Med Internet Res 2022 Aug 4;24(8):e37952. [doi: 10.2196/37952] [Medline: 35925651]
65. Marquard J. Human factors and organizational issues in health informatics: innovations and opportunities. Yearb Med Inform 2021 Aug;30(1):91-99. [doi: 10.1055/s-0041-1726511] [Medline: 34479382]
66. Hongsanun W, Insuk S. Quality assessment criteria for mobile health apps: a systematic review. Walailak J Sci & Tech 2020;17(8):745-759. [doi: 10.48048/wjst.2020.6482]
67. Yingta N, Nocera JA, Rehman IU, Brew O. A systematic review of usefulness design goals of occupational mhealth apps for healthcare workers. Presented at: 34th British HCI Conference; Jul 20-21, 2021. [doi: 10.14236/ewic/HCI2021.22]
68. Hu L, Chen Y, Cao E, Hu W. User experience & usability of wearable health device: a bibliometric analysis of 2014–2023. International Journal of Human–Computer Interaction 2025 Apr 18;41(8):5100-5119. [doi: 10.1080/10447318.2024.2357905]
69. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. J Bus Res 2021 Sep;133:285-296. [doi: 10.1016/j.jbusres.2021.04.070]
70. Bornmann L, Leydesdorff L. Scientometrics in a changing research landscape: bibliometrics has become an integral part of research quality evaluation and has been changing the practice of research. EMBO Rep 2014 Dec;15(12):1228-1232. [doi: 10.15252/embr.201439608] [Medline: 25389037]
71. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008 Feb;22(2):338-342. [doi: 10.1096/fj.07-9492LSF] [Medline: 17884971]
72. Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 2016 Jan;106(1):213-228. [doi: 10.1007/s11192-015-1765-5]
73. Moed HF. Citation Analysis in Research Evaluation: Springer Science & Business Media, Vol. 9.
74. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS ONE 2016;11(5):e0156164. [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
75. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983. [doi: 10.2196/12983] [Medline: 31469081]
76. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341. [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]

XSL•FO
RenderX

77. Leavy P. Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches: Guilford publications; 2022.
78. Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Asso for Info Science & Tech 2015 Nov;66(11):2215-2222. [doi: 10.1002/asi.23329]
79. Moss S. From “brick” to smartphone: the evolution of the mobile phone. MRS Bull 2021 Mar;46(3):287-288. [doi: 10.1557/s43577-021-00067-7]
80. Stocchi L, Pourazad N, Michaelidou N, Tanusondjaja A, Harrigan P. Marketing research on Mobile apps: past, present and future. J of the Acad Mark Sci 2022 Mar;50(2):195-225. [doi: 10.1007/s11747-021-00815-w]
81. Lupton D. Apps as artefacts: towards a critical perspective on mobile health and medical apps. Societies (Basel) 2014;4(4):606-622. [doi: 10.3390/soc4040606]
82. Aviv-Reuven S, Rosenfeld A. Publication patterns’ changes due to the COVID-19 pandemic: a longitudinal and short-term scientometric analysis. Scientometrics 2021;126(8):6761-6784. [doi: 10.1007/s11192-021-04059-x] [Medline: 34188333]
83. Bornmann L, Daniel HD. What do citation counts measure? A review of studies on citing behavior. Journal of Documentation 2008 Jan 18;64(1):45-80. [doi: 10.1108/00220410810844150]
84. Redner S. How popular is your paper? An empirical study of the citation distribution. Eur Phys J B 1998 Aug;4(2):131-134. [doi: 10.1007/s100510050359]
85. Zeng A, Shen Z, Zhou J, et al. The science of science: From the perspective of complex systems. Phys Rep 2017 Nov;714-715:1-73. [doi: 10.1016/j.physrep.2017.10.001]
86. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
87. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
88. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
89. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Mental Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891]
90. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
91. Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016 Mar;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]
92. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
93. Bucci S, Barrowclough C, Ainsworth J, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080. [doi: 10.1093/schbul/sby032] [Medline: 29566206]
94. Chiu W, Cho H, Chi CG. Consumers’ continuance intention to use fitness and health apps: an integration of the expectation–confirmation model and investment model. ITP 2021 May 24;34(3):978-998. [doi: 10.1108/ITP-09-2019-0463]
95. Laurie J, Blandford A. Making time for mindfulness. Int J Med Inform 2016 Dec;96:38-50. [doi: 10.1016/j.ijmedinf.2016.02.010] [Medline: 26965526]
96. Lee SM, Lee D. Healthcare wearable devices: an analysis of key factors for continuous use intention. Serv Bus 2020 Dec;14(4):503-531. [doi: 10.1007/s11628-020-00428-3]
97. Swendeman D, Ramanathan N, Baetscher L, et al. Smartphone self-monitoring to support self-management among people living with HIV: perceived benefits and theory of change from a mixed-methods randomized pilot study. J Acquir Immune Defic Syndr 2015 May 1;69 Suppl 1(0 1):S80-S91. [doi: 10.1097/QAI.0000000000000570] [Medline: 25867783]
98. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users’ perception of health and fitness apps with the UTAUT2 model. Telemed J E Health 2015 Sep;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
99. Alsswey A, Al-Samarraie H. Elderly users’ acceptance of mHealth user interface (UI) design-based culture: the moderator role of age. J Multimodal User Interfaces 2020 Mar;14(1):49-59. [doi: 10.1007/s12193-019-00307-w]
100. Cesnakova L, Boza V, Dubec P, et al. Continuous sound collection using smartphones and machine learning to measure cough. Digit Biomark 2020;3(3):166-175. [doi: 10.1159/000504666]
101. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. IEEE Trans Emerg Topics Comput 2020;9(3):1170-1182. [doi: 10.1109/TETC.2020.2974478]
102. Oyebode O, Alqahtani F, Orji R. Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews. IEEE Access 2020;8:111141-111158. [doi: 10.1109/ACCESS.2020.3002176]
103. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]

XSL•FO
RenderX

104. Ta V, Griffith C, Boatfield C, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020 Mar 6;22(3):e16235. [doi: 10.2196/16235] [Medline: 32141837]
105. Pal S, Biswas B, Gupta R, Kumar A, Gupta S. Exploring the factors that affect user experience in mobile-health applications: A text-mining and machine-learning approach. J Bus Res 2023 Feb;156:113484. [doi: 10.1016/j.jbusres.2022.113484] [Medline: 36475057]
106. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110. [doi: 10.2196/jmir.9157] [Medline: 29567635]
107. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86. [doi: 10.2196/jmir.2583] [Medline: 23598614]
108. Ahtinen A, Mattila E, Vaatanen A, et al. User experiences of mobile wellness applications in health promotion: user study of wellness diary, mobile coach and selfrelax. Presented at: 3d International ICST Conference on Pervasive Computing Technologies for Healthcare; Apr 1-3, 2009. [doi: 10.4108/ICST.PERVASIVEHEALTH2009.6007]
109. Crafoord MT, Fjell M, Sundberg K, Nilsson M, Langius-Eklöf A. Engagement in an interactive app for symptom self-management during treatment in patients with breast or prostate cancer: mixed methods study. J Med Internet Res 2020 Aug 10;22(8):e17058. [doi: 10.2196/17058] [Medline: 32663140]
110. Birkmeyer S, Wirtz BW, Langer PF. Determinants of mHealth success: An empirical investigation of the user perspective. Int J Inf Manage 2021 Aug;59:102351. [doi: 10.1016/j.ijinfomgt.2021.102351] [Medline: 36540093]
111. Lappalainen P, Kaipainen K, Lappalainen R, et al. Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. JMIR Res Protoc 2013 Jan 9;2(1):e1. [doi: 10.2196/resprot.2389] [Medline: 23611946]
112. Al Ayubi SU, Parmanto B, Branch R, Ding D. A persuasive and social mHealth application for physical activity: a usability and feasibility study. JMIR Mhealth Uhealth 2014 May 22;2(2):e25. [doi: 10.2196/mhealth.2902] [Medline: 25099928]
113. Ehn M, Eriksson LC, Åkerberg N, Johansson AC. Activity monitors as support for older persons' physical activity in daily life: qualitative study of the users' experiences. JMIR Mhealth Uhealth 2018 Feb 1;6(2):e34. [doi: 10.2196/mhealth.8345] [Medline: 29391342]
114. Mars M, Scott RE. WhatsApp in clinical practice: a literature review. The Promise of New Technologies in an Age of New Health Challenges 2016:82-90. [doi: 10.3233/978-1-61499-712-2-82]
115. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85. [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
116. Melnick ER, Hess EP, Guo G, et al. Patient-centered decision support: formative usability evaluation of integrated clinical decision support with a patient decision aid for minor head injury in the emergency department. J Med Internet Res 2017 May 19;19(5):e174. [doi: 10.2196/jmir.7846] [Medline: 28526667]
117. Yen PY, Walker DM, Smith JMG, Zhou MP, Menser TL, McAlearney AS. Usability evaluation of a commercial inpatient portal. Int J Med Inform 2018 Feb;110:10-18. [doi: 10.1016/j.ijmedinf.2017.11.007] [Medline: 29331248]
118. Silveira DV, Marcolino MS, Machado EL, et al. Development and evaluation of a mobile decision support system for hypertension management in the primary care setting in Brazil: mixed-methods field study on usability, feasibility, and utility. JMIR Mhealth Uhealth 2019 Mar 25;7(3):e9869. [doi: 10.2196/mhealth.9869] [Medline: 30907740]
119. Sobrinho A, da Silva LD, Perkusich A, Pinheiro ME, Cunha P. Design and evaluation of a mobile application to assist the self-monitoring of the chronic kidney disease in developing countries. BMC Med Inform Decis Mak 2018 Jan 12;18(1):7. [doi: 10.1186/s12911-018-0587-9] [Medline: 29329530]
120. Surka S, Edirippulige S, Steyn K, Gaziano T, Puoane T, Levitt N. Evaluating the use of mobile phone technology to enhance cardiovascular disease screening by community health workers. Int J Med Inform 2014 Sep;83(9):648-654. [doi: 10.1016/j.ijmedinf.2014.06.008] [Medline: 25002305]
121. Daruwalla ZJ, Wong KL, Thambiah J. The application of telemedicine in orthopedic surgery in singapore: a pilot study on a secure, mobile telehealth application and messaging platform. JMIR Mhealth Uhealth 2014 Jun 5;2(2):e28. [doi: 10.2196/mhealth.3303] [Medline: 25100283]
122. Gunter R, Fernandes-Taylor S, Mahnke A, et al. Evaluating patient usability of an image-based mobile health platform for postoperative wound monitoring. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e113. [doi: 10.2196/mhealth.6023] [Medline: 27683059]
123. McClure JB, Hartzler AL, Catz SL. Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers. JMIR mHealth uHealth 2016 Feb 12;4(1):e17. [doi: 10.2196/mhealth.5181] [Medline: 26872940]

Abbreviations

AHP: Analytical Hierarchy Process
HITAM: Health Information Technology Acceptance Model

XSL•FO
RenderX

MARS: Mobile App Rating Scale
MAUQ: Mobile Health App Usability Questionnaire
mHealth: mobile health
PRISMA: Preferred Reporting Items For Systematic Reviews and Meta-Analyses
PSSUQ: Post-Study System Usability Questionnaire
SLR: systematic literature review
SUS: System Usability Scale
TAM: Technology Acceptance Model
TC: total citations
TP: total publications
UTAUT: Unified Theory of Acceptance and Use of Technology
UX: user experience
UXS: User Experience And Satisfaction

Edited by NM Ali; submitted 13.04.25; peer-reviewed by G Lu, J Farias; revised version received 01.09.25; accepted 12.09.25; published 10.11.25.

Please cite as:
Alkhwaiter B, Aloud M, Almezeini N
User Experience in mHealth Research: Bibliometric Analysis of Trends and Developments (2007–2023)
JMIR Mhealth Uhealth 2025;13:e75909
URL: https://mhealth.jmir.org/2025/1/e75909
doi:10.2196/75909

© Bashaer Alkhwaiter, Monira Aloud, Nora Almezeini. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Effect and Safety of App-Based Interventions for Populations With Osteoarthritis: Systematic Review and Meta-Analysis of Randomized Controlled Trials

He Zhang[1], MSc; Hongchi Wang[2], MSc; Cheng Zhang[2], MD; Ruili Zhao[2], MD; Qian Lv[1], MSc; Luyao Zhang[1], MSc; Haoran Zhang[3], PhD; Shiyan Yan[2], PhD

[1]School of Management, Beijing University of Chinese Medicine, Beijing, China

[2]School of Acupuncture-Moxibustion and Tuina, Beijing University of Chinese Medicine, 11 Beisanhuan East Road, Chaoyang District, Beijing, China

[3]School of Preschool Education, Beijing Youth Politics College, Beijing, China

Corresponding Author:
Shiyan Yan, PhD
School of Acupuncture-Moxibustion and Tuina, Beijing University of Chinese Medicine, 11 Beisanhuan East Road, Chaoyang District, Beijing, China

Abstract

Background: Interventions based on apps are becoming increasingly popular for the treatment of osteoarthritis (OA), but research on the potential moderators of treatment efficacy is lacking.

Objective: The aim of this study was to examine the treatment efficacy, cost-effectiveness, and safety associated with app-based interventions for populations with OA and identify the potential factors associated with better treatment outcomes.

Methods: PubMed, Web of Science, Embase, and the Cochrane Library were searched from their inception to September 19, 2024, for randomized controlled trials on app-based interventions for patients with OA that report efficacy or health economic outcomes. The quality of each included study was assessed using the revised Cochrane Risk of Bias Assessment Tool (ROB 2.0). The primary outcome measure is the change in pain intensity before and after treatment. Secondary outcomes included function, quality of life, adverse events, and self-management. If I^2 was >50%, a random-effects model was applied. In addition to preplanned subgroup analyses based on OA type, intervention duration, risk of bias, age, and type of app-based intervention, exploratory post hoc subgroup analyses were conducted on variables related to the population, intervention characteristics, and study design features.

Results: The review includes 14 studies, comprising 12 randomized controlled trials (RCTs) and 2 health economics studies. The RCTs involved a total of 1410 participants, whose mean age ranged from 54 to 67 years. Compared with controls, app-based interventions led to a reduction in pain and improvement in physical function (standardized mean difference [SMD]=−0.36; 95% CI: −0.58 to −0.14; P<.001; I^2=72% and SMD 0.39; 95% CI 0.16 to 0.62; P<.001; I^2=67%; respectively), but showed no significance for quality of life and adverse events (SMD 0.23; 95% CI −0.04 to 0.50; P=.10; I^2=68% and odds ratio [OR]=1.33; 95% CI 0.84 to 2.12; P=.23; I^2=7%; respectively). The cost of the intervention group was lower than that of the control group. Subgroup analysis revealed a significant difference between those aged 60 years and older and those younger than 60 years (SMD −0.29; 95% CI −0.51 to −0.06 and SMD −0.84; 95% CI −1.25 to −0.43). The study also reported a high level of satisfaction and compliance rate, with all scores of the System Usability Scale exceeding 70 points, and this score is considered acceptable.

Conclusions: This study showed that app-based interventions were safe and effective for patients with OA, which might provide a cost-effective option, especially in resource-limited settings. Age is a critical factor for optimizing treatment benefits, especially when considering individual needs.

Trial Registration: PROSPERO CRD42024559569; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024559569

(JMIR Mhealth Uhealth 2025;13:e71193) doi:10.2196/71193

KEYWORDS

osteoarthritis; mobile applications; mobile health; physical therapy; meta-analysis; app

XSL•FO
RenderX

Introduction

Background

Osteoarthritis (OA) is the most common degenerative joint disorder, affecting approximately 528 million people worldwide and affecting one or multiple diarthrodial joints [1,2]. The knee, hip, and hand are the joints most frequently affected by this disease. OA imposes substantial pain and disability on patients, which reduce their quality of life (QoL) and lead to high economic costs [3-6]. With the aging population and increasing obesity rates, the prevalence of OA is expected to continue rising globally. In addition to the health burden [7], the economic burden of OA is also anticipated to increase significantly. The prevalence and alarming economic burden highlight the importance of implementing more applicable evidence-based practices specifically targeting these conditions.

Amid the COVID-19 pandemic and the challenges posed by rapid internet development, there has been a growing demand for mobile health care, with increasing interest from patients and health care providers. The World Health Organization (WHO) and Global Observatory for eHealth (GOe) [8] define mobile health (mHealth) as using appropriate digital technologies for public health. Compared with traditional care models, mHealth offers advantages such as user flexibility, allowing patients to receive remote care without having to travel. Additionally, it provides evidence-based frontline care and low-cost patient education and also holds significant potential for promoting the sustainable development of health care services. The mHealth strategies are usually used to manage chronic diseases, including diabetes, depression, asthma, and cardiovascular diseases [9-12]. Before promoting the clinical application of mHealth technology, it is crucial to fully verify its clinical efficacy and economic benefits. Previous studies have demonstrated that digital health interventions are cost-effective [13-16]. Molina-Garcia et al have confirmed the cost-effectiveness advantage of telemedicine in the management of musculoskeletal diseases through a systematic review [13]. The study shows that remote rehabilitation is US $89.55 cheaper per person than conventional treatment. Fatoye et al evaluated the economic efficiency of 2 digital health interventions for osteoarthritis, but both studies focused on telephone-based health guidance services [14]. However, synchronous interactive remote interventions such as telephone consultations not only require medical staff to have solid communication skills but also demand a significant amount of their time and energy. In contrast, application-related interventions can reach a much broader audience; by providing systematic rehabilitation education content and disease-related information, they can effectively enhance users' willingness to actively read and improve their understanding of disease self-management [17]. The results of a 3-arm study also confirm this advantage: compared with telephone-based interventions, mobile apps are more economically advantageous and have more significant intervention effects [18]. As far as we know, there is currently no study that has carried out a systematic review and meta-analysis on the economic benefits of app-based interventions in patients with OA, and this research gap needs to be filled urgently. In OA, the primary role of mHealth apps is to provide patients with personalized health management and treatment support through smartphone apps or other mobile devices [19]. However, systematic reviews and meta-analyses of app-based interventions in OA are still relatively insufficient, and among the limited studies available, the results regarding their effects on improving patients' QoL are somewhat inconsistent [20,21]. A systematic review and meta-analysis published in 2022 found that mobile apps supporting therapeutic exercise or tailored physical activity programs for musculoskeletal pain conditions may help reduce pain intensity and improve physical function [22]. This review included only studies on mobile apps acting as standalone digital therapeutics. Additionally, due to the restriction on publication years, a systematic review and meta-analysis on knee or hip OA published in 2020 included 12 studies, of which only 2 involved app-based interventions, and recent randomized controlled trials (RCTs) have not yet been reviewed [23]. Currently, most studies focus on the efficacy of app-based interventions [13,20-22,24], with a lack of evidence regarding their safety and the key factors that may influence treatment efficacy. In conclusion, conducting this meta-analysis is essential for resolving the discrepancies in previous studies and providing solid evidence to guide clinical practice and the development of effective digital interventions for OA.

This review aims to consolidate existing evidence to assess the efficacy, safety, and cost-effectiveness of app-based interventions for patients with OA and to identify potential factors associated with better treatment outcomes.

Research Questions

What is the effectiveness, safety, and cost-effectiveness of app-based interventions for patients with OA, and which patient characteristics are associated with better outcomes?

Methods

This systematic review and meta-analysis was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [25]. This study was registered with PROSPERO on June 18, 2024 (CRD42024559569).

Search Strategy

We systematically searched electronic databases, including PubMed, Web of Science, Embase, and the Cochrane Library, covering the period from database inception to September 19, 2024. The search strategy involved combining Medical Subject Headings (MeSH) terms and free text terms related to "osteoarthritis," "applications," and "randomized controlled trials." (Multimedia Appendix 1). Additionally, we manually searched the references of relevant systematic reviews and meta-analyses.

Inclusion and Exclusion Criteria

Under the PICOS (Population, Intervention, Comparison, Outcomes, and Study type) framework—a structured, evidence-based tool widely used in clinical research and systematic reviews to clarify study scope by defining key components— the inclusion criteria for our study were as

XSL•FO
RenderX

follows: (1) Population: adults aged 18 years and older with OA, confirmed by X-ray, clinical diagnosis, or self-report. (2) Intervention: app-based interventions must comply with the functional classification standards for mHealth apps, established by the National Institute for Health and Care Excellence (NICE) and the National Health Service (NHS) [26]. These classifications include the following: supporting clinical diagnosis and/or decision-making; improving clinical outcomes through behavior change and enhancing patient adherence to treatment; and acting as standalone digital therapeutics and primarily delivering disease-related education. (3) Control group: conventional care, no intervention, or other nonspecific controls. (4) Outcomes: clinical effectiveness measures, economic indicators, or safety indicators. (5) Study design: randomized controlled trial (RCT).

Exclusion criteria were as follows: (1) Research on wearable devices that record only patient information. (2) Use of WeChat, Facebook, Twitter, Tencent Meeting, Zoom, or similar platforms to conduct video or message intervention research. (3) Use of websites for intervention research. (4) Studies using only phone calls and text messages for intervention. (5) Apps designed exclusively for medical staff.

Study Selection and Data Extraction

Two independent reviewers (LZ and QL) screened titles and abstracts using EndNote software to exclude irrelevant studies. They further assessed the full texts of the remaining articles. In case of any disagreements, a third reviewer (SY) was consulted for discussion and resolution.

Data extraction was performed independently by the two reviewers (HZ and HW) using a predesigned form. The following data were recorded: basic information of study (eg, publication year, country, and authors), participant baseline characteristics (eg, number, age, gender ratio, BMI, and education level), app-based intervention features, and health economic data.

Assessment of Risk of Bias

The quality of each included study was assessed using the revised Cochrane Risk of Bias Assessment Tool (ROB 2.0) [27], evaluating 5 domains of bias: (1) bias arising from the randomization process; (2) bias due to deviations from intended interventions; (3) bias due to missing outcome data; (4) bias in the measurement of outcomes; and (5) bias in the selection of reported results. Two independent reviewers conducted the assessments, and any disagreements were resolved through discussion until a consensus was reached. The overall risk of bias was categorized as high, some concerns, or low. When all 5 domains are assessed as having a low risk of bias, the overall assessment is categorized as low risk of bias. If at least 1 domain is identified as having a possible risk of bias and no domain is rated as having a high risk of bias, the overall assessment is categorized as some concerns. If any domain is rated as having a high risk of bias or multiple domains are identified as having a possible risk of bias, the overall assessment is categorized as high risk of bias.

Statistical Analyses

For continuous variables, standard mean differences (SMDs) and 95% CIs were used to estimate the effect size with Hedge's g. For the meta-analysis of adverse events, odds ratios (ORs) and 95% CIs were calculated. Heterogeneity was assessed using the I^2 statistic; I^2 values were calculated for the degree of heterogeneity [28]. If I^2>50%, a random-effects model was applied. Sensitivity analyses were conducted by excluding studies with a high risk of bias and those with a sample size of less than 50.

In addition to preplanned subgroup analyses based on OA type, intervention duration, risk of bias, age, and type of app-based intervention, exploratory post hoc subgroup analyses were conducted on variables related to the population, intervention characteristics, and study design features. Publication bias was assessed using funnel plots and Egger regression test [29]. A P value <.05 was considered statistically significant. All analyses were performed using Excel (Microsoft) and Stata MP version 14.

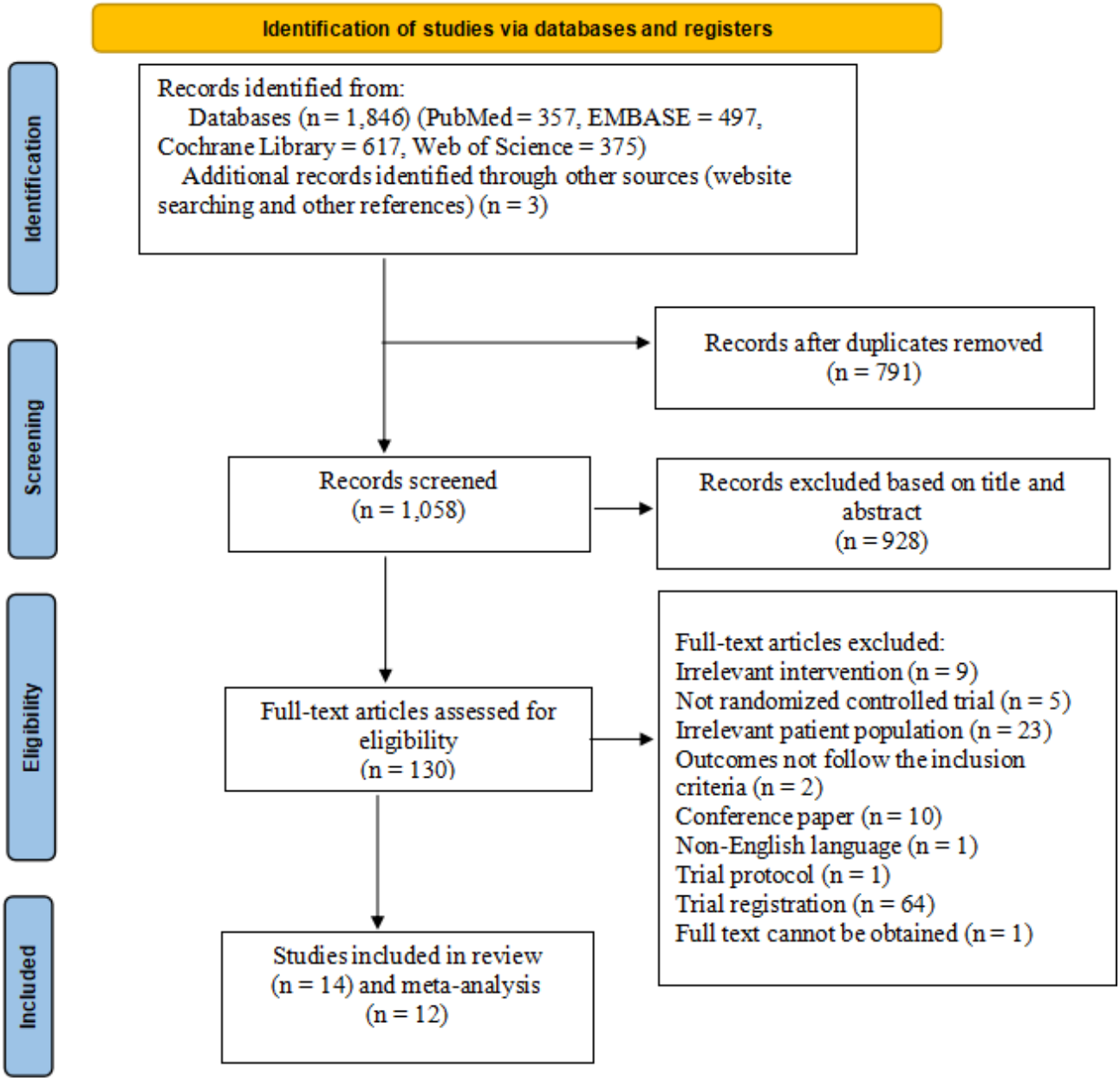
Results

Study Selection

A total of 1849 potentially relevant articles were initially retrieved. After removing 791 duplicates, 1058 articles were screened based on title and abstract. Of these, 928 were excluded, and an additional 130 were excluded after a full-text review. Ultimately, 14 studies were included in the review (Figure 1).

XSL•FO
RenderX

Figure 1. The flowchart of literature screening.



Study Characteristics

The study characteristics are summarized in Table 1. The review includes 14 articles published between 2017 and 2024, comprising 12 RCTs and 2 health economics studies. Of these, 10 were conducted in Western countries and 4 in Asian countries. The mean age of the participants ranged from 54 to 67 years. The RCTs involved a total of 1410 participants, with 702 (49.8%) in the intervention group and 708 (50.2%) in the control group. These studies include the following types of OA, with 8 exclusively addressing knee OA, 1 focusing on hand OA, and 3 including both knee and hip OA. Interventions based on apps lasted between 4 and 12 weeks and included features such as exercise, education, reminders, goal setting, and feedback. Four studies involved the use of wearable devices. In the control groups, 7 studies used usual care, 1 had no intervention, 1 included standard-of-care plus a wearable activity monitor, 1 used a delayed intervention, and 2 used paper-based self-guided materials.

XSL•FO
RenderX

Table . Study characteristics.

Study	Country	Conditions	Exp[a]/Con[b] Sample size (F[c]%)	Mean age (SD) years	Interventions	Control	Duration (weeks)	Outcome(s)
Dieter et al [30]	Germany	Knee OA	30 (40%) /31 (58%)	62.9 (8.5)	The reflex app can be classified as a fully automated, digital health app including a training app and 2 accelerometers to monitor joint movement.	Usual care	12	Pain intensity, physical function, QoL[d], HRQoL[e], exercise-specific self-efficacy, isometric maximum force measurement, and postural control
Weber et al [31]	Germany	Knee OA & Hip OA	32 (65.6%) /28 (57.1%)	61.9 (7.2)	App-based Join2Move program includes exercise, physical activity, and education program.	Usual care	12	Pain intensity, physical function, usability, satisfaction, strength, ROM[f], physical activity, self-management
Rodríguez Sánchez-Laulhé et al [32]	Spain	Hand OA	34 (73.5%) /40 (62.5%)	63.3 (8.2)	App-based care and home exercise program includes recommendations, joint protection material, general information about the disease, videos with demonstrations, and "follow-up diary."	Usual care	12	Pain intensity, physical function, morning stiffness, and strength
Lee et al [33]	Korea	Knee OA	15 (100%)/16 (100%)	67.0 (4.4)	Exercises can be performed at home via the app exact movement and repetitions plus a remote rehabilitation medical device exoRehab.	Without any intervention	8	Pain intensity and strength
Thiengwittayaporn et al [34]	Thailand	Knee OA	42 (85.7%) /40 (62.5%)	62.6 (8.3)	"Love your knee" app modules for education and assessment for the stage of disease.	Handout exercise guidance	4	Pain intensity, physical function, ability, ROM[f], symptoms, sports, recreation activities, and QoL[d]

XSL•FO
RenderX

Study	Country	Conditions	Exp[a]/Con[b] Sample size (F[c]%)	Mean age (SD) years	Interventions	Control	Duration (weeks)	Outcome(s)
Alasfour and Almarwani [35]	Saudi Arabia	Knee OA	20 (100%) /20 (100%)	54.4 (4.3)	The "My Dear Knee" app provides a guide for exercise performance, alerts, and a monitoring system.	Paper group	6	Pain intensity, physical function, and adherence
Chitkar et al [36]	Iran	Knee OA	31 (100%) / 29 (100%)	58.3 (6.9)	Patients received all the educational content through the mobile app.	Routine cares	8	Pain intensity, physical function, stiffness, and HRQoL[e]
Gohir et al [37]	UK	Knee OA	48 (70.8%)/57 (64.9%)	66.7 (9.2)	iOS (Apple) or Google Play (Alphabet) app provided the intervention group with daily exercises and informative texts.	Usual care	6	Pain intensity, physical function, stiffness, and MSK-HQ[g]
Pelle et al [4]	Netherlands	Knee OA & Hip OA	214 (68.7%) /213 (74.6%)	62.1 (7.4)	Dr. Bart app includes education, physical activity, vitality, and nutrition.	Usual care	12	Pain intensity, physical function, number of secondary health care consultations, symptoms, HRQoL[e], and self-management
Li et al [38]	Canada	Knee OA	26 (88.5%)/25 (76.0%)	64.9 (8.4)	FitViz is a new Fitbit-compatible web-based app, participants could view their physical activity goal attainment on it plus the use of a Fitbit Flex-2 wristband.	Delay group	12	Pain intensity, physical function, mean daily MVPA[h] time, symptoms, ADL[i], sports/recreation, QoL[d], self-management, and motivation
Kloek et al [39]	Netherlands	Knee OA & Hip OA	109 (67.9%) /99(67.7%)	63.1(8.7)	e-Exercise is a combination of about 5 face-to-face sessions with a physical therapist and an online application focusing on exercises and information.	Usual physical therapy	12	Pain intensity, physical activity, symptoms, QoL[d], self-perceived effect, self-efficacy, and tiredness

XSL•FO
RenderX

Study	Country	Conditions	Exp[a]/Con[b] Sample size (F[c]%)	Mean age (SD) years	Interventions	Control	Duration (weeks)	Outcome(s)
Skrepnik et al [40]	USA	Knee OA	107 (55.1%)/104 (45.2%)	62.6 (9.4)	The OA GO app provided motivational messages and requested that the patient enter pain and mood data on a once-daily basis plus unblinded wearable activity monitor.	Standard of care plus wearable activity monitor	12	Pain intensity, adherence, mobility, QoL[d], safety, and tolerability
Pelle et al [41]	Netherlands	Knee OA & Hip OA	214 (68.7%) /213 (74.6%)	62.1 (7.4)	Dr. Bart app includes education, physical activity, vitality, and nutrition.	Usual care	12	Pain intensity, physical function, number of secondary health care consultations, symptoms, HRQoL[e], self-management, costs, and QALYs[j]
Kloek et al [42]	Netherlands	Knee OA & Hip OA	109(67.9%) / 99 (67.7%)	63.1 (8.7)	e-Exercise is a combination of about 5 face-to-face sessions with a physical therapist and an online app focusing on exercises, and information.	Usual physical therapy	12	Physical function, cost, physical activity, QALYs[j]

[a]Exp: experimental group.
[b]Con: control group.
[c]F: female.
[d]QoL: quality of life.
[e]HRQoL: health-related quality of life.
[f]ROM: range of motion.
[g]MSK-HQ: Musculoskeletal Health Questionnaire.
[h]MVPA: moderate-to-vigorous physical activity.
[i]ADL: activities of daily living.
[j]QALYS: quality-adjusted life years.

Risk of Bias Assessment

The results of the ROB2 assessment are shown in Multimedia Appendix 2. Six studies were considered to have a high risk of bias, 2 were considered to have some concerns, and 4 were considered to have low risk of bias. Overall, the risk of bias was high due to a significant number of participant dropouts, baseline differences between groups, and failure to adhere to published protocols.

Pain

All 12 RCTs included pain intensity. Pain intensity was measured using 5 different scales: 6 studies used the KOOS/HOOS scale [4,30,31,34,38,39], 3 used the NRS scale [35,37,40], 1 used the AUSCAN scale [32], 1 used the VAS scale [31], and 1 used the WOMAC scale [36]. For patients with OA, the pooled effect of app-based interventions in reducing pain was significant (SMD −0.36; 95% CI −0.58 to −0.14; P<.001; I^2=72%) (Figure 2 [4,32-42]).

XSL•FO
RenderX

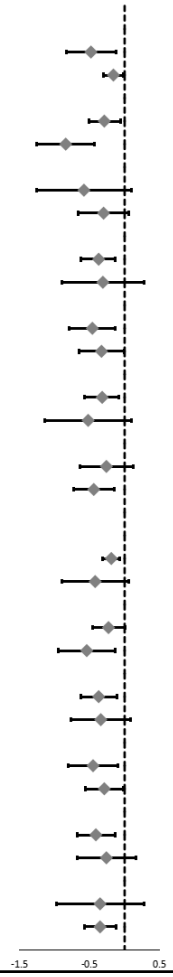
Figure 2. Results of meta-analyses [4,32-42] (A) pain; (B) function; and (C) QoL. Exp: experimental group; Con: control group; QoL: quality of life.

XSL•FO
RenderX

In subgroup analysis, regarding population characteristics, there were no significant differences between only knee OA group and other conditions group (SMD −0.48; 95% CI −0.83 to −0.12 vs SMD −0.16; 95% CI −0.30 to −0.02; *Q*=2.72; *P*=.1). However, there was a significant difference in pain reduction effects based on age, with notable differences between the 60 years and older age group and younger than 60 years age group (SMD −0.29; 95% CI −0.51 to −0.06 vs SMD −0.84; 95% CI −1.25 to −0.43; *Q*=5.37; *P*=.02). Intervention effects were not significant in the group with a higher proportion of individuals with higher education (SMD −0.58; 95% CI −1.26 to 0.10) and the group with a lower proportion of higher education (SMD −0.30; 95% CI −0.66 to 0.06). The difference in effects between these two education level groups did not reach statistical significance (*Q*=0.52; *P*=.47). Similarly, Western ethnicity, developed countries, and BMI also showed no significant group difference (Figure 3 and Multimedia Appendix 3).

Figure 3. Subgroup analyses. OA: osteoarthritis; BMI: body mass index.

Characteristics	No.of trials	No.of patients	SMD(95%CI)	Heterogeneity		Between-group tests	
				Q	I²	Q	P
Population Characteristics							
OA type						2.72	0.10
Knee OA	8	641	−0.48 [−0.83, −0.12]	30.32	77%		
Other conditions	4	769	−0.16 [−0.30, −0.02]	0.59	0%		
Mean age, year						5.37	0.02
≥60	10	1310	−0.29 [−0.51, −0.06]	31.20	71%		
<60	2	100	−0.84 [−1.25, −0.43]	0.32	0%		
The proportion of high level education						0.52	0.47
≥1/2	3	548	−0.58 [−1.26, 0.10]	16.07	88%		
<1/2	9	441	−0.30 [−0.66, 0.06]	11.89	66%		
Western ethnicity						0.03	0.85
Yes	8	1197	−0.37 [−0.62, −0.13]	25.94	73%		
No	4	213	−0.31 [−0.90, 0.28]	12.93	77%		
BMI, kg/m²						0.32	0.57
≥28	4	407	−0.46 [−0.79, −0.14]	6.54	54%		
<28	7	922	−0.33 [−0.65, −0.01]	27.09	78%		
Developed Country						0.34	0.56
Yes	9	1228	−0.32 [−0.57, −0.08]	29.61	73%		
No	3	182	−0.52 [−1.14, 0.10]	8.26	76%		
Proportion of female						0.54	0.46
≥70%	6	324	−0.26 [−0.64, 0.12]	14.09	65%		
<70%	6	1086	−0.44 [−0.73, −0.15]	24.66	80%		
Study design and Intervention-level							
Intervention duration, week						0.17	0.68
>8	7	1092	−0.31 [−0.55, −0.07]	19.37	69%		
≤8	5	318	−0.42 [−0.90, 0.06]	16.71	76%		
Risk of Bias						1.70	0.19
high	6	905	−0.23 [−0.46, 0.01]	12.20	59%		
some concern or low	6	505	−0.54 [−0.94, −0.13]	21.89	77%		
App involves educational modules						0.01	0.91
Yes	6	934	−0.37 [−0.63, −0.11]	15.59	68%		
No/Unclear	6	476	−0.34 [−0.77, 0.09]	22.93	78%		
App involves a supervisory reminder function						0.46	0.50
Yes	5	790	−0.45 [−0.81, −0.09]	17.93	78%		
No/Unclear	7	967	−0.29 [−0.56, −0.02]	20.55	71%		
Customization						0.34	0.56
Yes	8	1143	−0.41 [−0.68, −0.14]	29.78	76%		
No/Unclear	4	267	−0.26 [−0.68, 0.16]	4.72	37%		
Wearable electronic devices						0.00	1.00
Yes	4	354	−0.35 [−0.97, 0.28]	18.94	84%		
No/Unclear	8	1056	−0.35 [−0.57, −0.12]	18.99	63%		



Interventions with a duration of >8 weeks showed significant effects, while those with a duration of ≤8 weeks did not (SMD −0.31; 95% CI −0.55 to −0.07 vs SMD −0.42; 95% CI −0.90 to 0.06). The presence of educational modules in the app was associated with significant intervention effects, whereas the absence of educational modules was not (SMD −0.37; 95% CI −0.63 to −0.11 vs SMD −0.34; 95% CI −0.77 to 0.09). Similarly, interventions with customized app content yielded significant effects (SMD −0.35; 95% CI −0.97 to −0.28), and significant effects were also observed in studies without wearable devices (SMD −0.35; 95% CI −0.57 to −0.12). However, the differences in effectiveness across these factors did not reach statistical significance (Figure 3).

Function

Ten studies included physical functioning as an outcome, measured using 4 different scales: 6 studies used the HOOS/KOOS scale [4,30,31,38,39], 3 used the WOMAC scale [35-37], and 1 used the AUSCAN scale [32]. The pooled effect showed that the intervention group improved physical functioning more than the control group (SMD 0.39; 95% CI 0.16 to 0.62; *P*<.001; I^2=67%) (Figure 2 [4,32-42]).

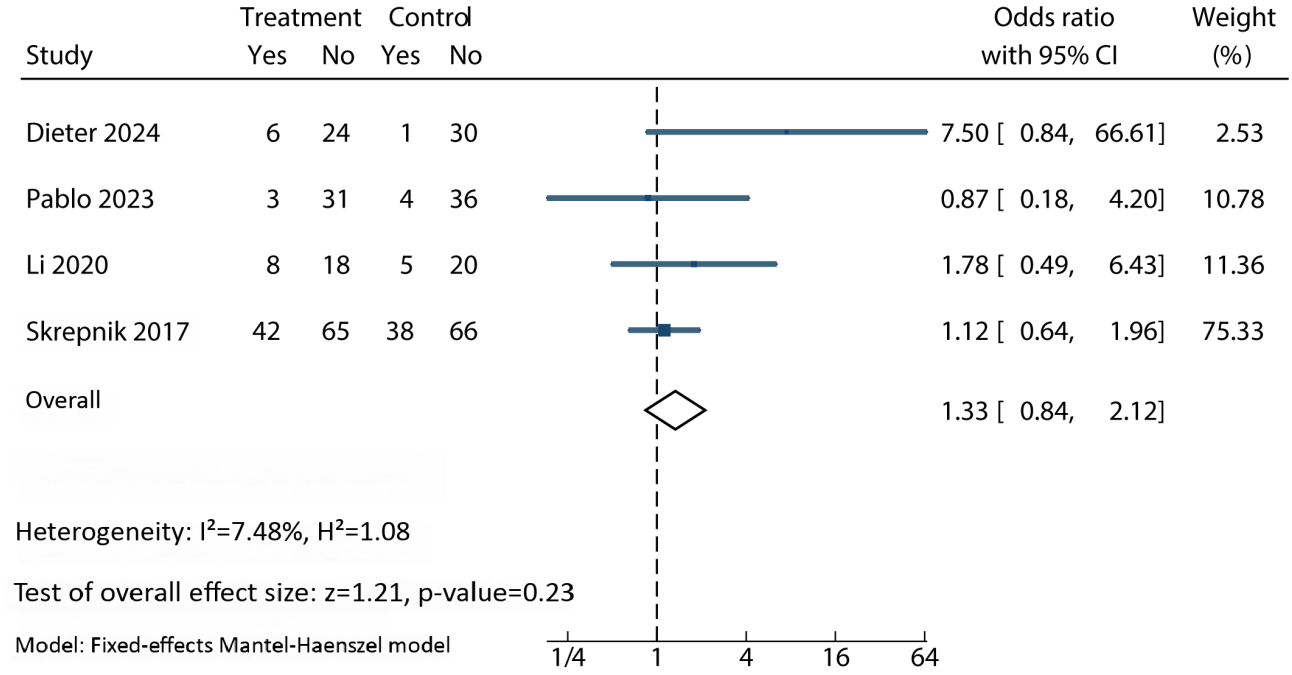
Safety

Of the 12 studies included in this meta-analysis, 2 (16.7%) explicitly reported that no adverse events had occurred [31,37], while 6 (50%) made no mention of adverse events [4,33-36,39]. Adverse events were reported in 4 studies [30,32,38,40], totaling

XSL•FO
RenderX

107 events. There were 59 adverse events among 197 participants allocated to the app intervention group. For participants allocated to usual care (n=200), 48 adverse events were reported. Of the 59 reported events, the most commonly reported adverse events among patients were muscle pain, arthralgia, upper respiratory tract infection, and falling while being physically active. There was no significant difference between groups (OR=1.33; 95% CI 0.84 to 2.12; *P*=.23; I^2=7%) (Figure 4 [30,32,38,40]).

Figure 4. Forest plot for the efficacy of an app-based intervention on safety [30,32,38,40].

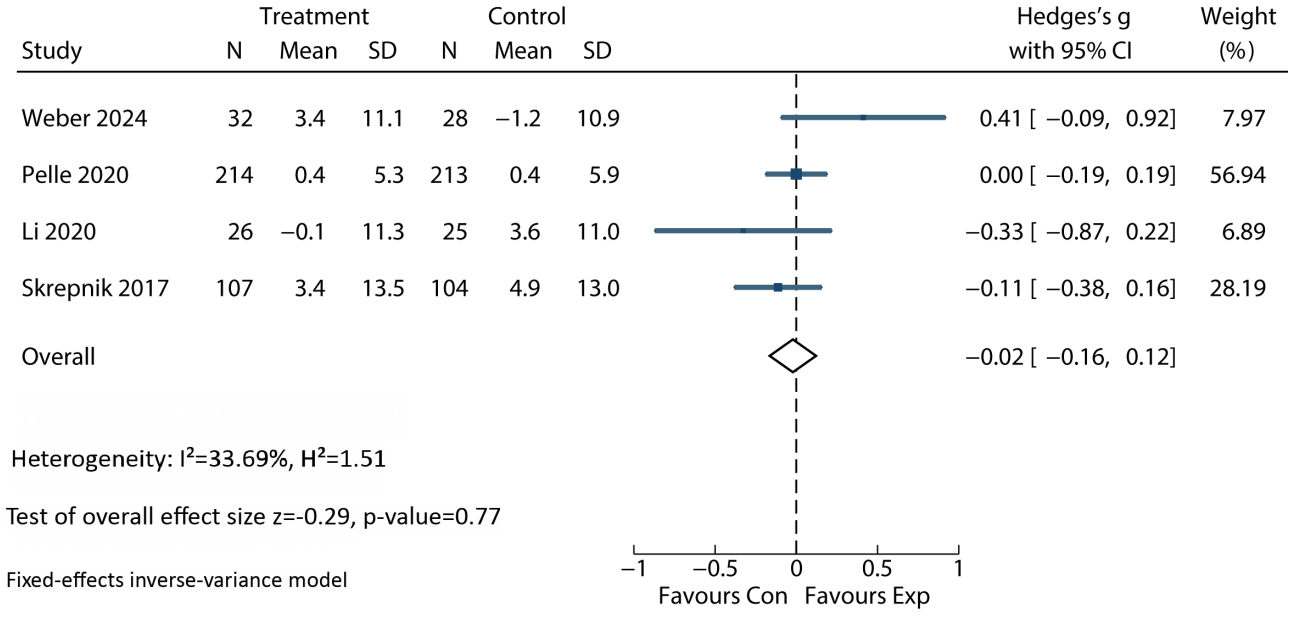


Other Outcomes

QoL and Self-Management

Six studies assessed QoL as an outcome [4,30,31,34,38,39], revealing no significant difference between the intervention and control groups (SMD 0.12; 95% CI −0.01 to 0.25; *P*=.08; I^2=68%)(Figure 2 [4,32-42]). Additionally, 4 studies evaluated self-management [4,31,38,40], finding no significant difference between groups (SMD −0.02; 95% CI: −0.16 to 0.12; *P*=.77; I^2=33.69%) (Figure 5 [31,32,38,40]).

Figure 5. Forest plot for the efficacy of an app-based intervention on self-management [31,32,38,40].



XSL•FO
RenderX

Health Economics of App-Based Intervention

Two studies included app-based intervention cost analyses [41,42]. One study reported no significant cost difference between the intervention and control groups (MD −1371.00; 95% CI −451.00 to 1240.00) [41]. Another study performed cost-effectiveness and cost-utility analyses [42], revealing no significant difference in quality-adjusted life years between groups (MD 0.00; 95% CI −0.00 to 0.01) and a cost saving of €22 (95% CI −36.00 to −3.00).

Adherence, Satisfaction, and Usability

Five studies reported on compliance, which was defined as the proportion of exercise sessions completed relative to the total number of scheduled sessions, or the completion of specified modules or requirements within the app [4,30,35,38-40]. Two studies reported on satisfaction, with one using the ZUF-8 scale (0 - 32) and the other using a percentage-based reporting method [31,40]. Three studies used the System Usability Scale (SUS) (0 - 100) to assess usability.

Sensitivity Analyses

We excluded RCTs with a high risk of bias and those with a sample size smaller than 50. The sensitivity analysis for pain intensity results robustly supports the intervention group (SMD −0.54; 95% CI −0.94 to −0.13; *P*<.001 and SMD −0.38; 95% CI −0.61 to −0.16; *P*<.001; respectively) (Multimedia Appendices 4 and 5).

Assessment of Publication Bias

The funnel plot and Egger test (*P*=.29) indicate no significant publication bias for pain intensity (Multimedia Appendix 6).

Discussion

Principal Findings and Comparison With Previous Works

This systematic review and meta-analysis evaluated the effectiveness, safety, and economics of app-based interventions for patients with osteoarthritis—findings that are crucial for the development and promotion of app-based OA intervention measures. Our results show that exercise and care delivered via app-based interventions provide greater efficacy and benefits compared to control interventions, and age is a key factor in optimizing treatment outcomes.

To our knowledge, this systematic review and meta-analysis is the first comprehensive evaluation of app-based interventions for patients with OA, assessing their efficacy, safety, and economic impact, and it is also the most comprehensive study to explore the patient characteristics and research intervention factors associated with better outcomes. Overall, this review shows that app-based interventions result in significant improvements in pain intensity and physical function compared with interventions without app support. However, no significant advantages were shown in QoL and self-management. Furthermore, we did not find an increased risk of adverse events when app intervention was compared with other treatments.

Our meta-analysis results show that app-based interventions can significantly reduce pain. This finding is consistent with the results of a systematic review and meta-analysis focusing on musculoskeletal diseases within a narrower range of mHealth interventions [22], as well as with a systematic review and meta-analysis on a more extensive digital intervention for patients with OA [43].

Our study demonstrated that age of participants was a significant factor influencing the effectiveness of app-based interventions. We observed notable reductions in pain intensity for participants across age groups, especially in the subgroup of those younger than 60 years. This result is consistent with the meta-analysis by Goh et al [44]. This might be related to the prevalence of age-related comorbidities in older patients or differences in app intervention acceptance and preferences. Lim et al's study on consumer preferences for mHealth apps found that younger individuals (18 - 54 y) are more inclined to use these apps [45]. However, the studies included in this review primarily involved participants aged between their 50s and 60s. This suggests that the findings may not be generalizable to younger or older populations, highlighting the need for further research to evaluate the efficacy of app-based interventions across a broader age range. Additionally, it was noted that higher-income individuals were more likely to use apps, and in the subgroup analysis of whether the study was conducted in a developed country, better outcomes were observed in developed countries, although the differences were not statistically significant. This finding appears to support the conclusion. Then, it is noteworthy that studies on apps with supervisory reminder features, and those with unclear/no reminder features did not show a difference in the effectiveness of app-based interventions. Intervention groups demonstrated better pain relief compared with the control group, indicating that app-based interventions have a positive effect on pain reduction overall, regardless of the presence of reminder features. However, this might be due to some studies not specifying whether the app had reminder features, leading to a substantial amount of error by grouping them as having unclear/no reminders. While differences in effectiveness were observed regarding whether the app included an educational component and whether it was customized, these differences were not statistically significant. This lack of significance might be attributed to small sample sizes and considerable heterogeneity among studies. Interventions using apps with educational components and customization showed better outcomes, though not statistically significant, suggesting potential advantages in practical applications [46]. This warrants further investigation in future research.

As for other results, the meta-analysis results indicate that app-based interventions offer benefits in improving physical function. This is consistent with previous research [25,46]. This consistency may be related to the core functions of app-based interventions—most intervention-oriented apps help patients develop regular rehabilitation habits through standardized rehabilitation training guidance, such as joint range of motion exercises and muscle-strengthening movements, or phased goal management, thereby improving physical function. This also indirectly confirms the stability of apps as rehabilitation auxiliary tools in improving physical function.

The results of this study showed no significant difference in the number of adverse events between the app intervention group

XSL•FO
RenderX

and the control group, which may be attributed to the fact that only a few of the included studies reported adverse events. The reported adverse events mainly focused on musculoskeletal pain and falls after exercise. However, it is noteworthy that the greatest potential risk of using digital health technologies lies in the privacy and security of patients' personal data. Regrettably, none of the studies included in this research reported on this aspect. In view of this, future studies should attach great importance to the protection of patients' data privacy and incorporate it into the monitoring and evaluation system of adverse events, so as to comprehensively assess the safety and reliability of digital health interventions.

Only 2 studies assessed the cost differences between app-based interventions and control groups. One study showed cost-saving benefits [41], while the other did not [42]. However, prior economic evaluations of digital health interventions for musculoskeletal diseases have found that these interventions are cost-effective compared with control groups [14]. Similar positive cost-effectiveness has also been observed for people with mental disorders and those with type 2 diabetes mellitus [15,16,47,48]. Therefore, researchers should place greater emphasis on the health economic outcomes of app-based interventions in future studies evaluating their effects on OA.

However, there was no significant improvement in QoL or self-management abilities compared with the control group. This might be due to the multidimensional nature of QoL, which is not solely influenced by disease-specific factors [49]. Only 3 studies reported self-management outcomes, with significant variability. Previous research has indicated that being an engaged and active participant in one's own care is associated with better health outcomes [50]. The NICE guidance committee suggests that patients should actively participate in and manage their care [51]. Future research could further explore how app-based interventions can effectively enhance patients' self-management skills and health engagement, with a focus on developing personalized apps and their potential role in long-term health improvement.

Additionally, high satisfaction and adherence rates were reported, aligning with previous findings in OA and other conditions [52]. Particularly in terms of usability, all 3 studies reported SUS scores exceeding 70 points, with scores above this threshold considered acceptable. These results suggest the potential of app-based interventions as relatively low-cost alternatives to traditional treatments in settings with limited resources or funding.

The mHealth apps in this study provide exercise and physical activity support, as well as personalized and tailored health information and education. The exercise interventions in this study include aerobic exercise, muscle strength training, and joint range-of-motion training. Previous systematic reviews and meta-analyses have confirmed that exercise therapy can significantly reduce pain and improve joint function, and the conclusions of this study are consistent with these findings [53,54]. As a flexible and efficient rehabilitation solution, mobile or remote exercise interventions based on the app can completely break through the time and location constraints of traditional treatment—patients do not need to make frequent trips to and from the hospital and can arrange rehabilitation training flexibly according to their own schedules. In terms of practical effects, it can not only significantly improve the overall efficiency of exercise therapy but also greatly reduce patients' waiting time. Economic evaluations have found that physical therapy supported by remote apps can effectively reduce costs, and its therapeutic effect is equivalent to that of the traditional offline model [55].

The ROB assessment highlights that the lack of blinding among participants and excessive missing data are significant sources of bias in the studies. Implementing blinding is relatively challenging due to the nature of app-based interventions. Additionally, the autonomy associated with home exercise and app usage may result in many participants being unable to adhere to the interventions. The broad scope of OA, coupled with the inclusion of various types of app-based interventions and control groups in this study, has led to considerable heterogeneity within the population. These factors may affect the reliability and applicability of the study findings.

Limitations

This study has several limitations. First, the number of included studies was relatively small, which was due to the stringent inclusion criteria applied. Second, while only 2 studies included economic evaluations, which necessitates caution in interpreting the economic findings, existing international research has established that app-based interventions are cost-effective in the management of other diseases [13,56]. It is recommended that prospective studies be conducted in the future to rigorously evaluate its health economics. Third, only English language studies were included. Focusing solely on English language publications may result in the omission of relevant studies conducted in other languages, potentially leading to an incomplete understanding of the effectiveness of app-based interventions across different cultural contexts and health care systems. Consequently, the generalizability of the conclusions may be constrained. Fourth, in an age-based subgroup analysis, it should be noted that only 2 studies were included in this subgroup with a small sample size. The age range of the population included in this paper is relatively narrow. Therefore, future studies should enroll participants from different age groups and explore the associations between age and potential moderating factors of treatment efficacy. Additionally, the results of such small sample subgroups in the subgroup analysis should also be interpreted with caution. Finally, the interventions evaluated had relatively short duration, leaving their long-term efficacy unclear.

Conclusions

These systematic review and meta-analysis confirm that app-based interventions are an effective and safe treatment option for patients with OA. Our study also underscores the importance of considering individual characteristics, particularly age, as a key factor influencing the effectiveness of app-based interventions. Younger patients tended to show greater improvements, which suggests that age-tailored approaches could enhance the efficacy of these digital solutions. These findings are expected to serve as a reference for developers and researchers in the rapidly evolving field of mHealth, offering

XSL•FO
RenderX

guidance on the design, promotion, and implementation of app-based interventions for OA.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1-6, Checklist 1.

Authors' Contributions

Author Contributions: He Zhang and SY had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. He Zhang and HW contributed equally in as co–first authors. SY and Haoran Zhang are co-corresponding authors. Concept and design: All authors. Acquisition, analysis, or interpretation of data: He Zhang, HW, QL, LZ. Drafting of the manuscript: He Zhang. Critical review of the manuscript for important intellectual content: SY, Haoran Zhang, Statistical analysis: He Zhang, HW. Obtained funding: SY. Administrative, technical, or material support: SY. Supervision: SY, Haoran Zhang.

This research was supported by the National Natural Science Foundation of China (Grant No. 82174234).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the literature search and the number of citations found.
[DOCX File, 21 KB - mhealth_v13i1e71193_app1.docx]

Multimedia Appendix 2
Risk of bias of included studies.
[DOCX File, 67 KB - mhealth_v13i1e71193_app2.docx]

Multimedia Appendix 3
Characteristics of subgroup.
[DOCX File, 22 KB - mhealth_v13i1e71193_app3.docx]

Multimedia Appendix 4
Results of sensitivity analyses: excluded RCTs with a high risk of bias.
[DOCX File, 176 KB - mhealth_v13i1e71193_app4.docx]

Multimedia Appendix 5
Results of sensitivity analyses: excluded RCTs with a sample size smaller than 50.
[DOCX File, 254 KB - mhealth_v13i1e71193_app5.docx]

Multimedia Appendix 6
Funnel plot assessing publication bias.
[DOCX File, 97 KB - mhealth_v13i1e71193_app6.docx]

Checklist 1
PRISMA checklist.
[PDF File, 101 KB - mhealth_v13i1e71193_app7.pdf]

References

1. Martel-Pelletier J, Barr AJ, Cicuttini FM, et al. Osteoarthritis. Nat Rev Dis Primers 2016 Oct 13;2(1):16072. [doi: 10.1038/nrdp.2016.72] [Medline: 27734845]
2. World health organization. Osteoarthritis. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/osteoarthritis [accessed 2024-11-25]
3. Bijlsma JWJ, Berenbaum F, Lafeber F. Osteoarthritis: an update with relevance for clinical practice. Lancet 2011 Jun 18;377(9783):2115-2126. [doi: 10.1016/S0140-6736(11)60243-2] [Medline: 21684382]

XSL•FO
RenderX

4. Pelle T, Bevers K, van der Palen J, van den Hoogen FHJ, van den Ende CHM. Effect of the Dr. Bart application on healthcare use and clinical outcomes in people with osteoarthritis of the knee and/or hip in the Netherlands; a randomized controlled trial. Osteoarthritis Cartilage 2020 Apr;28(4):418-427. [doi: 10.1016/j.joca.2020.02.831] [Medline: 32119972]
5. McConaghy K, Klika AK, Apte SS, Erdemir A, Derwin K, Piuzzi NS. A call to action for musculoskeletal research funding: the growing economic and disease burden of musculoskeletal conditions in the United States is not reflected in musculoskeletal research funding. J Bone Joint Surg Am 2023 Mar 15;105(6):492-498. [doi: 10.2106/JBJS.22.00693] [Medline: 36574617]
6. Ackerman IN, Pratt C, Gorelik A, Liew D. Projected burden of osteoarthritis and rheumatoid arthritis in Australia: a population-level analysis. Arthritis Care Res (Hoboken) 2018 Jun;70(6):877-883. [doi: 10.1002/acr.23414] [Medline: 28898565]
7. Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jul;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
8. World health organization. Global Observatory for eHealth World Health Organization. 2021. URL: https://www.who.int/observatories/global-observatory-for-ehealth [accessed 2024-11-25]
9. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12(12):CD007459. [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
10. Bae H, Shin H, Ji HG, Kwon JS, Kim H, Hur JW. App-based interventions for moderate to severe depression: a systematic review and meta-analysis. JAMA Netw Open 2023 Nov 1;6(11):e2344120. [doi: 10.1001/jamanetworkopen.2023.44120] [Medline: 37983028]
11. Bonoto BC, de Araújo VE, Godói IP, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 1;5(3):e4. [doi: 10.2196/mhealth.6309] [Medline: 28249834]
12. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480. [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
13. Molina-Garcia P, Mora-Traverso M, Prieto-Moreno R, Díaz-Vásquez A, Antony B, Ariza-Vega P. Effectiveness and cost-effectiveness of telerehabilitation for musculoskeletal disorders: a systematic review and meta-analysis. Ann Phys Rehabil Med 2024 Feb;67(1):101791. [doi: 10.1016/j.rehab.2023.101791] [Medline: 38128150]
14. Fatoye F, Gebrye T, Mbada C, Useh U. Economic evaluations of digital health interventions for the management of musculoskeletal disorders: systematic review and meta-analysis. J Med Internet Res 2023 Jul 6;25:e41113. [doi: 10.2196/41113] [Medline: 37410542]
15. Rohrbach PJ, Dingemans AE, Evers C, et al. Cost-effectiveness of internet interventions compared with treatment as usual for people with mental disorders: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2023 Jan 5;25:e38204. [doi: 10.2196/38204] [Medline: 36602854]
16. Rinaldi G, Hijazi A, Haghparast-Bidgoli H. Cost and cost-effectiveness of mHealth interventions for the prevention and control of type 2 diabetes mellitus: A systematic review. Diabetes Res Clin Pract 2020 Apr;162:108084. [doi: 10.1016/j.diabres.2020.108084] [Medline: 32061819]
17. Jeon E, Park HA. Experiences of patients with a diabetes self-care app developed based on the information-motivation-behavioral skills model: before-and-after study. JMIR Diabetes 2019 Apr 18;4(2):e11590. [doi: 10.2196/11590] [Medline: 30998218]
18. Chiu CJ, Yu YC, Du YF, et al. Comparing a social and communication app, telephone intervention, and usual care for diabetes self-management: 3-arm quasiexperimental evaluation study. JMIR Mhealth Uhealth 2020 Jun 2;8(6):e14024. [doi: 10.2196/14024] [Medline: 32484448]
19. Agnew JMR, Hanratty CE, McVeigh JG, Nugent C, Kerr DP. An investigation into the use of mHealth in musculoskeletal physiotherapy: scoping review. JMIR Rehabil Assist Technol 2022 Mar 11;9(1):e33609. [doi: 10.2196/33609] [Medline: 35275089]
20. Fan I, Govil D, King MG, Scholes MJ, Semciw AI. How effective are exercises delivered digitally (requiring internet), amongst patients with hip or knee osteoarthritis? A systematic review and meta-analysis. Osteoarthr Cartil 2024 Mar;32(3):254-265. [doi: 10.1016/j.joca.2023.11.011] [Medline: 38030118]
21. Schäfer AGM, Zalpour C, von Piekartz H, Hall TM, Paelke V. The efficacy of electronic health-supported home exercise interventions for patients with osteoarthritis of the knee: systematic review. J Med Internet Res 2018 Apr 26;20(4):e152. [doi: 10.2196/jmir.9465] [Medline: 29699963]
22. Thompson D, Rattu S, Tower J, Egerton T, Francis J, Merolli M. Mobile app use to support therapeutic exercise for musculoskeletal pain conditions may help improve pain intensity and self-reported physical function: a systematic review. J Physiother 2023 Jan;69(1):23-34. [doi: 10.1016/j.jphys.2022.11.012] [Medline: 36528508]
23. Chen T, Or CK, Chen J. Effects of technology-supported exercise programs on the knee pain, physical function, and quality of life of individuals with knee osteoarthritis and/or chronic knee pain: A systematic review and meta-analysis of randomized controlled trials. J Am Med Inform Assoc 2021 Feb 15;28(2):414-423. [doi: 10.1093/jamia/ocaa282] [Medline: 33236109]

XSL•FO
RenderX

24. Latif-Zade T, Tucci B, Verbovetskaya D, et al. Systematic review shows tele-rehabilitation might achieve comparable results to office-based rehabilitation for decreasing pain in patients with knee osteoarthritis. Medicina (Kaunas) 2021 Jul 28;57(8):764. [doi: 10.3390/medicina57080764] [Medline: 34440970]
25. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
26. National institute for health and care excellence. Evidence standards framework for digital health technologies. 2018. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2024-11-28]
27. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
29. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634. [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
30. Dieter V, Janssen P, Krauss I. Efficacy of the mHealth-based exercise intervention reflex for patients with knee osteoarthritis: pilot randomized controlled trial. JMIR Mhealth Uhealth 2024 Sep 9;12:e54356. [doi: 10.2196/54356] [Medline: 39250181]
31. Weber F, Kloek C, Stuhrmann S, Blum Y, Grüneberg C, Veenhof C. Usability and preliminary effectiveness of an app-based physical activity and education program for people with hip or knee osteoarthritis - a pilot randomized controlled trial. Arthritis Res Ther 2024 Apr 10;26(1):83. [doi: 10.1186/s13075-024-03291-z] [Medline: 38600607]
32. Rodríguez Sánchez-Laulhé P, Biscarri-Carbonero Á, Suero-Pineda A, et al. The effects of a mobile app-delivered intervention in people with symptomatic hand osteoarthritis: a pragmatic randomized controlled trial. Eur J Phys Rehabil Med 2023 Feb;59(1):54-64. [doi: 10.23736/S1973-9087.22.07744-9] [Medline: 36633498]
33. Lee EL, Jang MH, Lee BJ, et al. Home-based remote rehabilitation leads to superior outcomes for older women with knee osteoarthritis: a randomized controlled trial. J Am Med Dir Assoc 2023 Oct;24(10):1555-1561. [doi: 10.1016/j.jamda.2023.08.013] [Medline: 37699531]
34. Thiengwittayaporn S, Wattanapreechanon P, Sakon P, et al. Development of a mobile application to improve exercise accuracy and quality of life in knee osteoarthritis patients: a randomized controlled trial. Arch Orthop Trauma Surg 2023 Feb;143(2):729-738. [doi: 10.1007/s00402-021-04149-8] [Medline: 34453570]
35. Alasfour M, Almarwani M. The effect of innovative smartphone application on adherence to a home-based exercise programs for female older adults with knee osteoarthritis in Saudi Arabia: a randomized controlled trial. Disabil Rehabil 2022 Jun;44(11):2420-2427. [doi: 10.1080/09638288.2020.1836268] [Medline: 33103499]
36. Chitkar SSA, Hakkak HRM, Saadati H, Hosseini SH, Jafari Y, Ganji R. Correction to: the effect of mobile-app-based instruction on the physical function of female patients with knee osteoarthritis: a parallel randomized controlled trial. BMC Womens Health 2021 Oct 5;21(1):349. [doi: 10.1186/s12905-021-01491-2] [Medline: 34610846]
37. Gohir SA, Eek F, Kelly A, Abhishek A, Valdes AM. Effectiveness of internet-based exercises aimed at treating knee osteoarthritis: the iBEAT-OA randomized clinical trial. JAMA Netw Open 2021 Feb 1;4(2):e210012. [doi: 10.1001/jamanetworkopen.2021.0012] [Medline: 33620447]
38. Li LC, Feehan LM, Xie H, et al. Effects of a 12-week multifaceted wearable-based program for people with knee osteoarthritis: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 3;8(7):e19116. [doi: 10.2196/19116] [Medline: 32618578]
39. Kloek CJJ, Bossen D, Spreeuwenberg PM, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a blended physical therapist intervention in people with hip osteoarthritis, knee osteoarthritis, or both: a cluster-randomized controlled trial. Phys Ther 2018 Jul 1;98(7):560-570. [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
40. Skrepnik N, Spitzer A, Altman R, Hoekstra J, Stewart J, Toselli R. Assessing the impact of a novel smartphone application compared with standard follow-up on mobility of patients with knee osteoarthritis following treatment with Hylan G-F 20: a randomized controlled trial. JMIR Mhealth Uhealth 2017 May 9;5(5):e64. [doi: 10.2196/mhealth.7179] [Medline: 28487266]
41. Pelle T, Bevers K, van den Hoogen F, van der Palen J, van den Ende C. Economic evaluation of the Dr. Bart application in individuals with knee and/or hip osteoarthritis. Arthritis Care Res (Hoboken) 2022 Jun;74(6):945-954. [doi: 10.1002/acr.24608] [Medline: 33768675]
42. Kloek CJJ, van Dongen JM, de Bakker DH, Bossen D, Dekker J, Veenhof C. Cost-effectiveness of a blended physiotherapy intervention compared to usual physiotherapy in patients with hip and/or knee osteoarthritis: a cluster randomized controlled trial. BMC Public Health 2018 Aug 31;18(1):1082. [doi: 10.1186/s12889-018-5975-7] [Medline: 30170586]
43. Safari R, Jackson J, Sheffield D. Digital self-management interventions for people with osteoarthritis: systematic review with meta-analysis. J Med Internet Res 2020 Jul 20;22(7):e15365. [doi: 10.2196/15365] [Medline: 32706657]
44. Goh SL, Persson MSM, Stocks J, et al. Efficacy and potential determinants of exercise therapy in knee and hip osteoarthritis: a systematic review and meta-analysis. Ann Phys Rehabil Med 2019 Sep;62(5):356-365. [doi: 10.1016/j.rehab.2019.04.006] [Medline: 31121333]
45. Lim D, Norman R, Robinson S. Consumer preference to utilise a mobile health app: a stated preference experiment. PLoS ONE 2020;15(2):e0229546. [doi: 10.1371/journal.pone.0229546] [Medline: 32084250]

XSL•FO
RenderX

46. Gay C, Chabaud A, Guilley E, Coudeyre E. Educating patients about the benefits of physical activity and exercise for their hip and knee osteoarthritis. Systematic literature review. Ann Phys Rehabil Med 2016 Jun;59(3):174-183. [doi: 10.1016/j.rehab.2016.02.005] [Medline: 27053003]
47. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
48. Naslund JA, Mitchell LM, Joshi U, Nagda D, Lu C. Economic evaluation and costs of telepsychiatry programmes: a systematic review. J Telemed Telecare 2022 Jun;28(5):311-330. [doi: 10.1177/1357633X20938919] [Medline: 32746762]
49. Carr AJ, Higginson IJ. Are quality of life measures patient centred? BMJ 2001 Jun 2;322(7298):1357-1360. [doi: 10.1136/bmj.322.7298.1357] [Medline: 11387189]
50. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026. [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
51. Wood G, Neilson J, Cottrell E, Hoole SP, Guideline Committee. Osteoarthritis in people over 16: diagnosis and management-updated summary of NICE guidance. BMJ 2023 Jan 24;380:24. [doi: 10.1136/bmj.p24] [Medline: 36693668]
52. Bunting JW, Withers TM, Heneghan NR, Greaves CJ. Digital interventions for promoting exercise adherence in chronic musculoskeletal pain: a systematic review and meta-analysis. Physiotherapy 2021 Jun;111:23-30. [doi: 10.1016/j.physio.2020.08.001] [Medline: 33316865]
53. Jiang Y, Tan Y, Cheng L, Wang J. Effects of three types of resistance training on knee osteoarthritis: a systematic review and network meta-analysis. PLoS ONE 2024;19(12):e0309950. [doi: 10.1371/journal.pone.0309950] [Medline: 39636953]
54. Wang X, Chen Z, Liang Y, et al. Effects of exercise on balance function in people with knee osteoarthritis: a systematic review and meta-analysis of randomized controlled trials. Healthcare (Basel) 2025 Jun 1;13(11):1312. [doi: 10.3390/healthcare13111312] [Medline: 40508925]
55. Hayes AJ, Withers HG, Glinsky JV, et al. Remotely delivered physiotherapy for musculoskeletal conditions is cost saving for the health system and patients: economic evaluation of the REFORM randomised trial. J Physiother 2025 Jul;71(3):179-184. [doi: 10.1016/j.jphys.2025.06.006] [Medline: 40579312]
56. Morrish N, Stallard P, Whittle K, et al. Cost-effectiveness of adding a smartphone app (BlueIce) to the mental health care of adolescents who repeatedly self-harm. Psychiatry Res 2024 Dec;342:116186. [doi: 10.1016/j.psychres.2024.116186] [Medline: 39293280]

Abbreviation

GOe: Global Observatory for eHealth
MeSH: medical subject headings
mHealth: mobile health
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
OA: osteoarthritis
OR: odds ratio
PICOS: Population, Intervention, Comparison, Outcomes, and Study type
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
QoL: quality of life
RCT: randomized controlled trial
ROB: risk of bias
SMD: standardized mean difference
WHO: World Health Organization

Edited by L Buis; submitted 12.01.25; peer-reviewed by YS Jiang, Y Duan; revised version received 21.07.25; accepted 25.08.25; published 22.09.25.

Please cite as:
Zhang H, Wang H, Zhang C, Zhao R, Lv Q, Zhang L, Zhang H, Yan S
The Effect and Safety of App-Based Interventions for Populations With Osteoarthritis: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2025;13:e71193
URL: https://mhealth.jmir.org/2025/1/e71193
doi:10.2196/71193

XSL•FO
RenderX

© He Zhang, Hongchi Wang, Cheng Zhang, Ruili Zhao, Qian Lv, Luyao Zhang, Haoran Zhang, Shiyan Yan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Behavior Change Resources Used in Mobile App–Based Interventions Addressing Weight, Behavioral, and Metabolic Outcomes in Adults With Overweight and Obesity: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Sijia Li[1*], MMed; You Zhou[1*], MMed; Ying Tang[2], MSN; Haoming Ma[1], PhD; Yuying Zhang[1], MMed; Aoqi Wang[1], MMed; Xingyi Tang[1], MMed; Runyuan Pei[1], BSN; Meihua Piao[1], PhD

[1]Chinese Academy of Medical Sciences, Peking Union Medical College School of Nursing, No 33 Ba Da Chu Road, Shijingshan District, Beijing, China

[2]School of Nursing, Evidence-Based Nursing Center, Lanzhou University, Lanzhou, China

[*]these authors contributed equally

Corresponding Author:
Meihua Piao, PhD
Chinese Academy of Medical Sciences, Peking Union Medical College School of Nursing, No 33 Ba Da Chu Road, Shijingshan District, Beijing, China

Abstract

Background: Overweight and obesity have become a public health issue. Lifestyle modifications delivered through mobile devices, especially mobile phones, present an opportunity to support weight loss efforts. However, evidence regarding the effects of mobile apps on other outcomes, such as blood pressure and physical activity (PA), remains limited. Recent studies on this topic require a systematic review and updating, and the active elements that promote behavior change remain unclear.

Objective: The meta-analysis aimed to explore the effects of mobile phone apps on weight-related outcomes (weight, BMI, waist circumference [WC], fat mass, fat mass percentage), behavioral outcomes (moderate-to-vigorous physical activity [MVPA], energy intake), and metabolic outcomes (systolic blood pressure [SBP], diastolic blood pressure [DBP], triglycerides, hemoglobin A_{1c} [HbA_{1c}]) among adults with overweight and obesity. Behavior change techniques (BCTs), the smallest replicable intervention elements, were also identified to clarify the components used in current studies, along with associated resources, including facilitating, boosting, and nudging. In addition, factors influencing the effectiveness of these interventions were explored.

Methods: Six databases (PubMed, Embase, CENTRAL, Web of Science, PsycINFO, and CINAHL) were searched for relevant randomized controlled trials (RCTs) published in English from inception to May 20, 2024. Two independent authors conducted study selection, data extraction, and quality assessment. The effect size of interventions was calculated using the mean difference (MD), and a random-effects model was applied for data analysis. Subgroup and sensitivity analyses were conducted to explore potential influencing factors and identify possible sources of heterogeneity.

Results: A total of 29 studies were included. The results indicated that mobile phone app interventions significantly reduced weight (MD=−1.45 kg, 95% CI −2.01 to −0.89; *P*<.001), BMI (MD=−0.35 kg/m^2, 95% CI −0.57 to −0.13; *P*=.002), WC (MD=−1.98 cm, 95% CI −3.42 to −0.55; *P*=.007), fat mass (MD=−1.32 kg, 95% CI −1.94 to −0.69; *P*<.001), DBP (MD=−1.76 mm Hg, 95% CI −3.47 to −0.04; *P*=.04), and HbA_{1c} (MD=−0.13%, 95% CI −0.22 to −0.04; *P*=.005). However, nonsignificant effects were observed for other outcomes. The most frequently used BCTs included 2.3 “self-monitoring of behavior” (n=25), 4.1 “instruction on how to perform the behavior” (n=24), 2.2 “feedback on behavior” (n=20), 1.1 “goal setting (behavior)” (n=19), and 1.4 “action planning” (n=15). Fifty-nine percent of included studies used 3 resource types (ie, facilitating, boosting, and nudging). Subgroup analyses identified combined diet and PA interventions, medium-term intervention duration, and the use of ≥8 BCTs as potential reference interventions for improving outcomes.

Conclusions: This meta-analysis demonstrates that mobile phone app interventions significantly reduce weight, BMI, WC, fat mass, DBP, and HbA_{1c} in adults with overweight and obesity. However, future studies should explore ways to optimize app interventions by incorporating behavior change strategies and resources to further enhance their overall effectiveness.

(JMIR Mhealth Uhealth 2025;13:e63313) doi:10.2196/63313

XSL•FO
RenderX

KEYWORDS

mobile app; overweight; obesity; mHealth; eHealth; physical activity; diet; app; application; behavior change; behavior; exercise; systematic review; review; meta-analysis; weight; BMI; body fat; systolic blood pressure; diastolic blood pressure; blood pressure; metabolic; obese adults

Introduction

Overweight and obesity are defined as excessive fat accumulation that can negatively affect health. People with a BMI between 25 and 30 kg/m^2 are classified as overweight, while those with a BMI over 30 kg/m^2 are classified as obese [1]. Overweight and obesity have become major public health issues. In 2016, more than 1.9 billion adults were overweight, with approximately 650 million experiencing obesity [2], and this number is estimated to affect half of the global population by 2030 [3]. Obesity is associated with an increased incidence of chronic conditions such as type 2 diabetes and hypertension [2], which in turn reduces disease-free years [4], quality of life [5], and life expectancy [6]. In addition, obesity imposes a substantial economic burden on nations [7]. In 2014, the global economic impact of obesity was projected to be US $2.0 trillion, equivalent to 2.8% of the global gross domestic product (GDP) [3]. Compared with the healthy-weight population, individuals with obesity incur 36% higher average annual health care expenses, including 105% higher prescription costs and 39% higher primary care costs [8]. Given the health consequences and economic burden associated with obesity, effective interventions are of great importance.

The fundamental approach to addressing overweight and obesity remains a multicomponent behavioral intervention [9]. However, several barriers exist in implementing lifestyle interventions aimed at weight loss [10]. A previous systematic review [11] has suggested that mobile health (mHealth) apps hold promise for health behavior change. mHealth is defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [12]. In addition, with advancements in technology, the number of mobile phone users has been steadily increasing, and mHealth interventions have the potential to engage large populations at relatively low costs, enhancing the feasibility and accessibility of public health interventions [13]. Therefore, lifestyle modifications delivered via mobile devices, especially mobile phones, present an opportunity to help people lose weight.

Several meta-analyses have evaluated the effectiveness of mobile phone apps on weight loss in adults [14-16]. However, these reviews have some limitations. First, most reviews [14-18] have primarily focused on weight and BMI, while other important outcomes, such as fat mass, blood pressure, and physical activity (PA), have also been reported in original studies. These additional outcomes hold clinical significance and should be included to provide a more comprehensive understanding of the overall effectiveness of mobile app–based interventions. This would offer valuable insights for helping adults with overweight and obesity prevent obesity-related comorbidities. Second, with the rapid development of technology in recent years, numerous new randomized controlled trials (RCTs) have emerged [19-23], necessitating updated evidence. Third, although behavioral interventions are the cornerstone of weight loss, the specific components used in these interventions have not been fully examined from the perspective of behavior change [15,16].

Behavior change techniques (BCTs) are the smallest replicable intervention elements designed to modify or redirect the causal mechanisms that regulate behavior [24,25]. BCTs are widely used to guide and understand intervention design, and identifying the resources associated with these BCTs is crucial for understanding their functional mechanisms [25]. To address this, Michaelsen et al [26] proposed a behavior change resource model (BCRM) centered on an individual's resources to elucidate the resources used to form these BCTs. This model categorizes BCTs into 3 types: facilitating, boosting, and nudging, which correspond to external resource provision, reflective resource build-up, and affective resource use, respectively. To the best of our knowledge, no review has yet identified the resources used in mobile app–based behavior change interventions.

Therefore, we aimed to conduct a systematic review and meta-analysis to explore the effects of mobile phone apps on weight-related outcomes (weight, BMI, waist circumference [WC], fat mass, and fat mass percentage), behavioral outcomes (moderate-to-vigorous physical activity [MVPA] and energy intake), and metabolic outcomes (systolic blood pressure [SBP], diastolic blood pressure [DBP], triglycerides, and hemoglobin A_{1c} [HbA_{1c}]) among adults with overweight and obesity. In addition, we identified the BCTs and behavior change resources (BCRs) used in these interventions. We also explored factors influencing the effectiveness of these interventions.

Methods

Overview

This meta-analysis was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [27] (the PRISMA checklist is provided in Checklist 1) and was registered on the International Prospective Register of Systematic Reviews (PROSPERO registration number: CRD42024513381).

Search Strategy

Six databases (ie, PubMed, Embase, CENTRAL, Web of Science, PsycINFO, and CINAHL) were independently searched by 2 authors (SJL and YZ) for relevant RCTs published in English from inception to May 20, 2024. The search strategy was developed by one author (SJL) and confirmed by another author (YZ) according to the PICOS (Participants, Interventions, Comparisons, Outcomes, and Study Design) framework. The key search terms and MeSH (Medical Subject Headings) included obesity, overweight, body weight, mobile apps, mHealth, and apps. The full search strategy is provided in

XSL•FO
RenderX

Multimedia Appendix 1. In addition, we manually searched the references of previously published relevant reviews.

Study Eligibility Criteria

Studies included in this meta-analysis had to meet the following criteria: (1) participants: adults with overweight and obesity aged ≥18 years; (2) intervention: mobile phone apps used as the primary component; (3) comparisons: usual care without mobile phone app interventions or no intervention; (4) outcomes: weight-related outcomes (ie, weight, BMI, WC, fat mass, and fat mass percentage), behavioral outcomes (ie, MVPA and energy intake), and metabolic outcomes (ie, SBP, DBP, triglycerides, and HbA_{1c}); (5) study design: RCTs published in English.

Studies were excluded if they met the following criteria: (1) participants diagnosed with diseases other than obesity; (2) conference articles, letters, reviews, commentaries, or protocols; and (3) unavailable full texts or incomplete relevant data.

Study Selection and Data Extraction

All titles and abstracts of the studies were downloaded and imported into EndNote X9, with duplicates automatically removed. Two investigators (SJL and YZ) independently screened the titles and abstracts based on the eligibility criteria. Subsequently, the full texts of potentially relevant studies were downloaded and reviewed to select the included articles. Reasons for exclusion were recorded. Any disagreements were resolved through consultation with a third author (HMM).

Data regarding paper characteristics (author, year, and country), study design, study population (eg, average age, sex distribution, and baseline BMI), intervention content (eg, intervention duration and brief description of the intervention), comparison content (eg, brief introduction to the comparison), and outcomes were extracted into a predesigned Microsoft Excel by one reviewer (SJL) and checked by a second reviewer (YZ). We did not make any assumptions about missing or unclear information when extracting data to avoid introducing misleading information. Other statistics (eg, 95% CI or SEs) were converted to SD if not available (Cochrane Handbook version 6.4, Chapter 5, 5.7) [28]. If the units of the outcome measures were not uniform, such as triglycerides measured in mg/dL, we converted them to mmol/L to ensure consistency in units. We used Michie BCT Taxonomy (BCTTv1) to identify the BCTs present in the included studies (Multimedia Appendix 2) [24]. Intervention and control descriptors in the original RCTs that aligned with BCT definitions were marked with a "✓." The initial identification of BCTs was performed by 1 reviewer (SJL) and checked and confirmed by another reviewer (YT). Subsequently, we mapped the BCTs used in each study to the corresponding BCRs.

Quality Assessment

Cochrane risk-of-bias version 2 (ROB 2) tool was used to assess the methodological quality of the included RCTs by 2 reviewers (SJL and YZ) independently [29]. This tool contains 5 domains: the randomization process, deviations from the intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Each domain was rated as either low, some concerns, or high risk of bias. An overall judgment was then assigned to each study. A third investigator (HMM) was responsible for resolving any discrepancies during the evaluation process.

Two independent authors (SJL and YZ) used the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) framework to assess the overall quality of evidence. The framework evaluates 5 domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. Based on these evaluations, the quality of evidence was classified as high, moderate, low, or very low. Any disagreements during this process were resolved by a third reviewer (HMM).

Statistical Analysis

The meta-analysis was performed to evaluate the effectiveness of mobile app interventions on weight loss among adults with overweight and obesity, with forest plots used to visually display the results of individual studies and the overall effect size. Outcomes were pooled in the meta-analysis if change scores from baseline were available, and the number of participants was recorded. The heterogeneity of the data was assessed using the chi-square test and the I^2 statistic, with 25%, 50%, and 75% representing low, moderate, and high heterogeneity, respectively [30]. Review Manager software (version 5.3; Cochrane Collaboration) was used to analyze data. The mean difference (MD) with a 95% CI was used to express the effectiveness of interventions involving mobile apps. A random-effects model was applied for data analysis to obtain more conservative results.

Subgroup analyses were conducted to explore factors, including intervention contents, presence of theory, intervention duration, number of BCTs, and number of BCRs, which may influence the effects of the intervention on weight and help identify possible sources of heterogeneity. Sensitivity analyses were also performed to assess the robustness of the results in R (version 4.4.2; R Foundation for Statistical Computing). Publication bias was evaluated by visual evaluation of funnel plot asymmetry and quantified by Egger test in Stata software (version 15; StataCorp). Publication bias assessment was performed separately for outcome measures included in meta-analyses that incorporated more than 10 original studies (Cochrane Handbook version 6.4, Chapter 13, 13.3) [28]. In all analyses, $P<.05$ was considered statistically significant.

Results

Study Selection

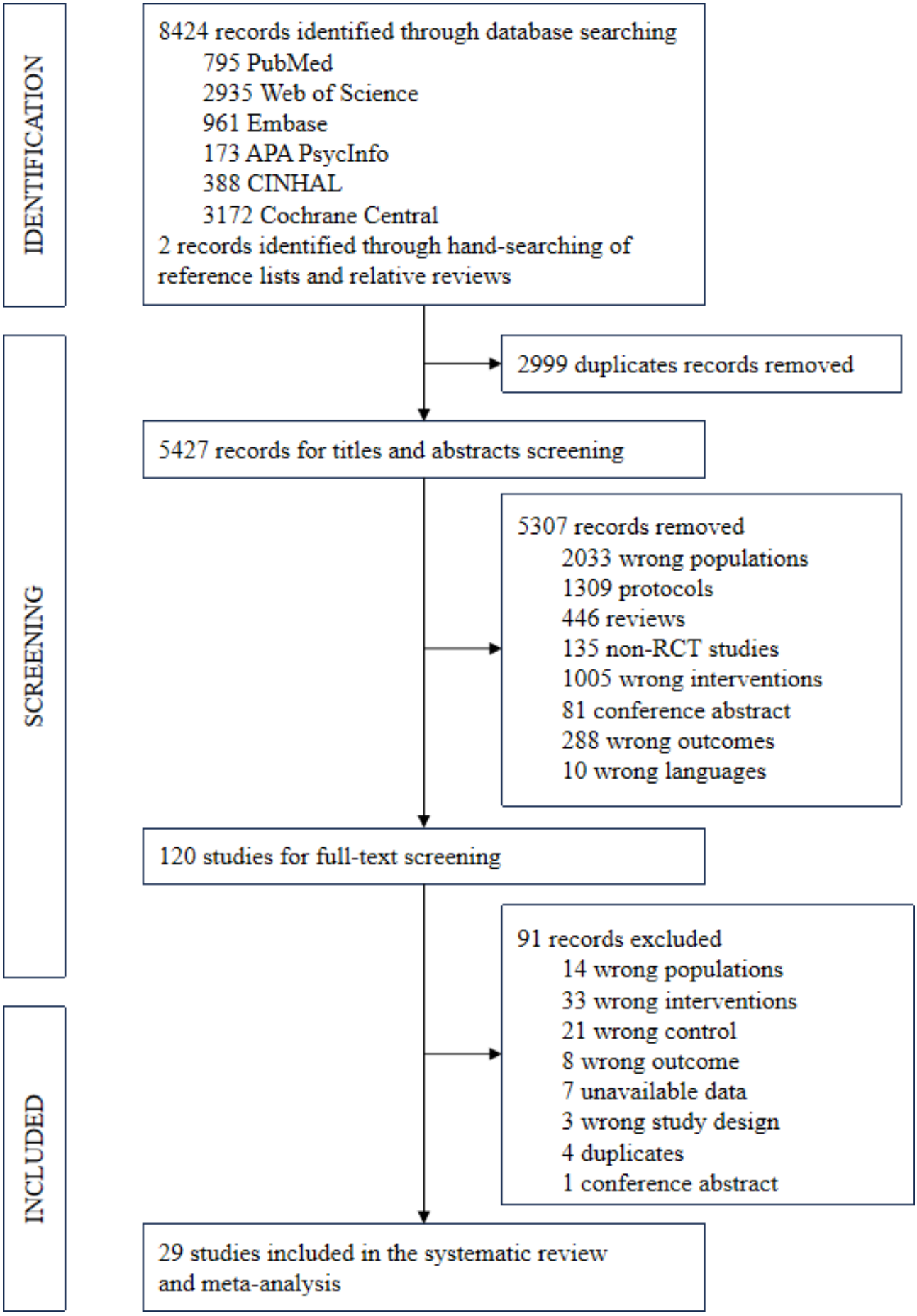
The literature search across 6 databases generated 8424 articles, and an additional 2 records were added through hand-searching relevant reviews. After removing duplicates, titles and abstracts of 5427 studies were screened. Subsequently, 120 full texts were reviewed. However, 91 studies were excluded based on the eligibility criteria for the following reasons: wrong populations (n=14), wrong interventions (n=33), wrong control (n=21), wrong outcome (n=8), unavailable data (n=7), wrong study design (n=3), duplicates (n=4), and conference abstract (n=1). A detailed list of exclusions with corresponding reasons can be found in Multimedia Appendix 3. Finally, a total of 29 studies [19-23,31-54] were included in the systematic review

XSL•FO
RenderX

and meta-analysis. Details are presented in the PRISMA flowchart in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. RCT: randomized controlled trial.



XSL•FO
RenderX

Study Characteristics

The details of the study characteristics are listed in Multimedia Appendix 4. The 29 studies were conducted in Spain [21,31,32], France [33], the United Kingdom [34-37], Korea [19,38,39], India [20], Australia [40-42], China [43], Japan [22], the United States of America [23,44-51], Belgium [52], New Zealand [53], and Germany [54]. Sample sizes ranged from 20 to 650 across the included studies, with the average age ranging from 22 to 55 years. Behavioral theories (ie, social cognitive theory, self-regulatory theory, control theory, operant conditioning, ecological theory, social network theory, transtheoretical model, habit formation theory, social support theory, and behavior self-management theory) were used in the design of the mobile app interventions. The intervention duration ranges from 2 months to 24 months. The intervention contents shared some common features, such as goal setting, social support, self-monitoring, and the use of credible sources. Usual care and no intervention were the primary forms of control groups. Outcome measures included weight-related outcomes (weight, BMI, WC, fat mass, and fat mass percentage), behavioral outcomes (MVPA and energy intake), and metabolic outcomes (SBP, DBP, triglycerides, and HbA_{1c}).

In the 29 included RCTs, 34 BCTs from 13 categories were identified in the intervention group, while 20 BCTs from 9 categories were identified in the control group. The number of BCTs identified per study ranged from 2 to 16 (median 8, mean 8.69, SD 3.79) in the intervention group and from 0 to 14 (median 2) in the control group (see Multimedia Appendix 5). The most frequently identified BCTs in the intervention group were: 2.3 “self-monitoring of behavior” (n=25), 4.1 “instruction on how to perform the behavior” (n=24), 2.2 “feedback on behavior” (n=20), 1.1 “goal setting (behavior)” (n=19), and 1.4 “action planning” (n=15). In the control group, the most frequently used BCT was 4.1 “instruction on how to perform the behavior” (n=18). Details are shown in Table 1.

XSL•FO
RenderX

Table . Frequency of behavior change techniques used in the intervention and control group in each study.

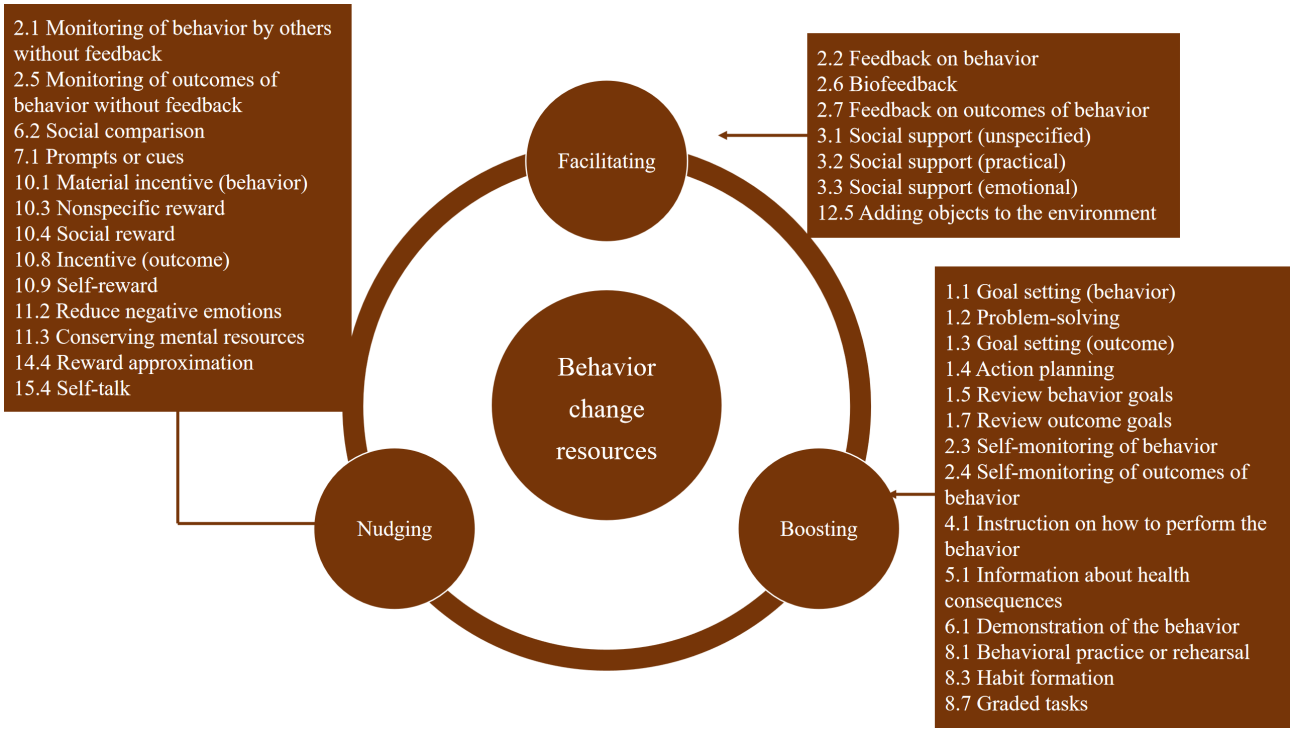
BCT[a] Taxonomy	Intervention group	Control group
1.1 Goal setting (behavior)	19 (65.5)	8 (27.6)
1.2 Problem solving	7 (24.1)	2 (6.9)
1.3 Goal setting (outcome)	8 (27.6)	3 (10.3)
1.4 Action planning	15 (51.7)	6 (20.7)
1.5 Review behavior goals	5 (17.2)	1 (3.4)
1.7 Review outcome goals	4 (13.8)	1 (3.4)
2.1 Monitoring of behavior by others without feedback	2 (6.9)	0 (0.0)
2.2 Feedback on behavior	20 (70.0)	2 (6.9)
2.3 Self-monitoring of behavior	25 (86.2)	7 (24.1)
2.4 Self-monitoring of outcomes of behavior	14 (48.3)	4 (13.8)
2.5 Monitoring of outcomes of behavior without feedback	1 (3.4)	0 (0.0)
2.6 Biofeedback	3 (10.3)	0 (0.0)
2.7 Feedback on outcomes of behavior	13 (44.8)	2 (6.9)
3.1 Social support (unspecified)	16 (55.2%)	4 (13.8%)
3.2 Social support (practical)	6 (20.7)	3 (10.3)
3.3 Social support (emotional)	5 (17.2)	1 (3.4)
4.1 Instruction on how to perform the behavior	24 (82.8)	18 (62.1)
5.1 Information about health consequences	7 (24.1)	5 (17.2)
6.1 Demonstration of the behavior	3 (10.3)	1 (3.4)
6.2 Social comparison	3 (10.3)	0 (0.0)
7.1 Prompts/cues	13 (44.8)	2 (6.9)
8.1 Behavioral practice/rehearsal	2 (6.9)	0 (0.0)
8.3 Habit formation	2 (6.9)	0 (0.0)
8.7 Graded tasks	2 (6.9)	0 (0.0)
10.1 Material incentive (behavior)	3 (10.3)	0 (0.0)
10.3 Nonspecific reward	4 (13.8)	0 (0.0)
10.4 Social reward	5 (17.2)	0 (0.0)
10.8 Incentive (outcome)	1 (3.4)	0 (0.0)
10.9 Self-reward	1 (3.4)	0 (0.0)
11.2 Reduce negative emotions	1 (3.4)	0 (0.0)
11.3 Conserving mental resources	4 (13.8)	1 (3.4)
12.5 Adding objects to the environment	12 (41.4)	2 (6.9)
14.4 Reward approximation	1 (3.4)	0 (0.0)
15.4 Self-talk	1 (3.4)	0 (0.0)

[a]BCT: behavior change technique

The identified BCTs were categorized into 3 resource types: facilitating, boosting, and nudging (see Figure 2 and Multimedia Appendix 6). The specifics of BCRs identified in each study are presented in Table 2. Among the 29 included RCTs, 24 studies [19-22,31-34,36-38,40-44,46-52,54] used facilitating resources, 24 [19-22,31-38,40-42,44,47-54] used boosting resources, and 19 [21,22,32-38,40-42,44,45,47-50,54] used nudging resources. Fifty-nine percent of the included studies used all 3 types of resources, while only 4 studies [43,45,46,53] used 1 resource type.

XSL•FO
RenderX

Figure 2. Behavior change resources were identified in each study.



XSL•FO
RenderX

Table . Behavior change resources (BCRs) identified in each study.

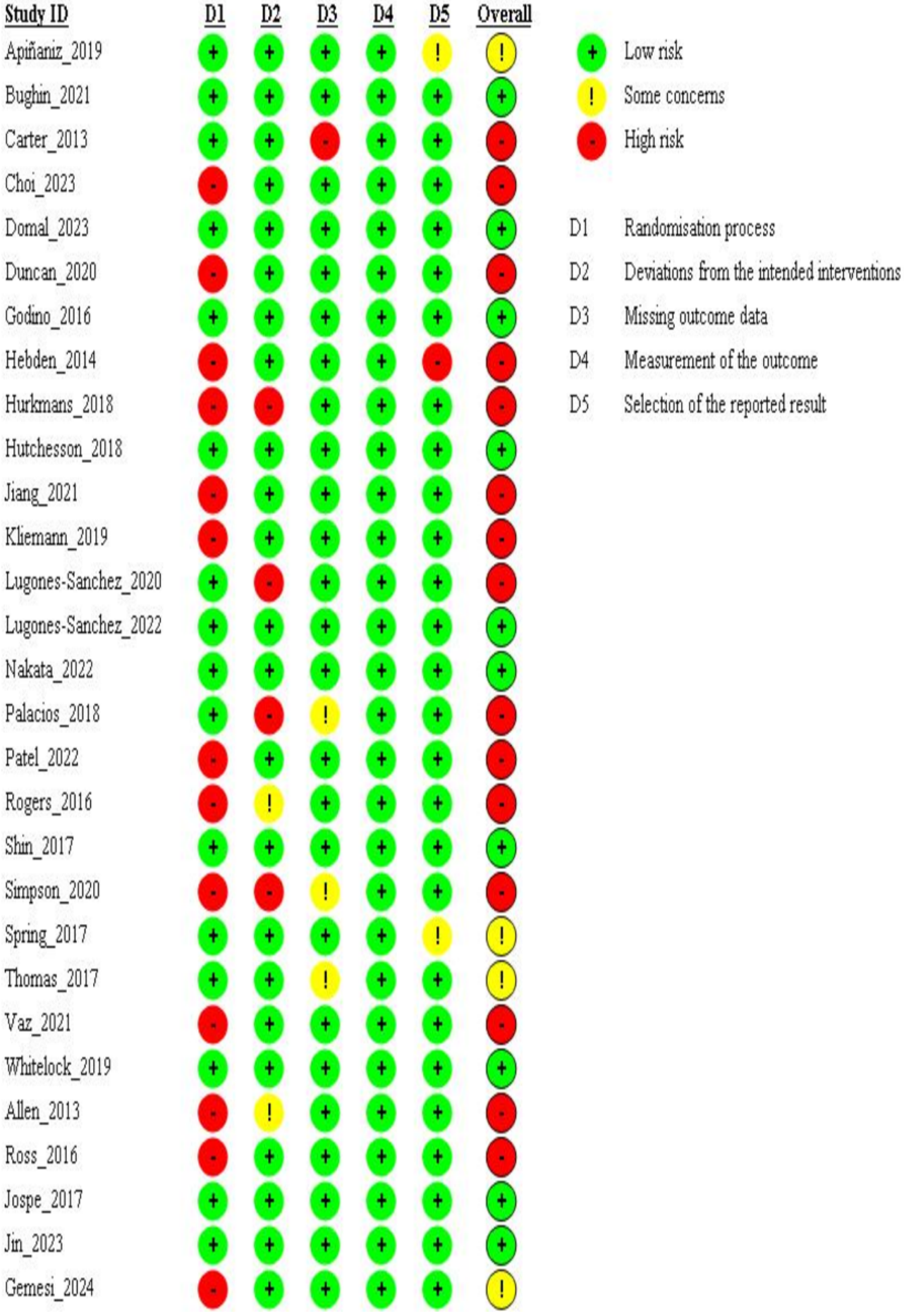
Studies	Facilitating	Boosting	Nudging	Total
Apiñaniz et al [31]	✓	✓		2
Bughin et al [33]	✓	✓	✓	3
Carter et al [34]	✓	✓	✓	3
Choi et al [19]	✓	✓		2
Domal et al [20]	✓	✓		2
Duncan et al [40]	✓	✓	✓	3
Godino et al [44]	✓	✓	✓	3
Hebden et al [41]	✓	✓	✓	3
Hurkmans et al [52]	✓	✓		2
Hutchesson et al [42]	✓	✓	✓	3
Jiang et al [43]	✓			1
Kliemann et al [35]		✓	✓	2
Lugones-Sanchez et al [21]	✓	✓	✓	3
Lugones-Sanchez et al [32]	✓	✓	✓	3
Nakata et al [22]	✓	✓	✓	3
Palacios et al [45]			✓	1
Patel et al [23]				0
Rogers et al [46]	✓			1
Shin et al [38]	✓	✓	✓	3
Simpson et al [36]	✓	✓	✓	3
Spring et al [47]	✓	✓	✓	3
Thomas et al [48]	✓	✓	✓	3
Vaz et al [49]	✓	✓	✓	3
Whitelock et al [37]	✓	✓	✓	3
Allen et al [50]	✓	✓	✓	3
Ross et al [51]	✓	✓		2
Jospe et al [53]		✓		1
Jin et al [39]				0
Gemesi et al [54]	✓	✓	✓	3
Total number of each BCR	24	24	19	67

Risk of Bias and GRADE Assessment

Of the 29 included studies, 10 [20-22,33,37-39,42,44,53] were judged to have low risk of bias, 4 [31,47,48,54] as having some concerns, and 15 [19,23,32,34-36,40,41,43,45,46,49-52] as having high risk of bias (Figure 3). While all studies reported the use of randomization, 13 studies [19,23,35,36,40,41,43,46,49-52,54] failed to explain the allocation concealment process, which was considered a high risk in this domain. Other significant sources of bias included high dropout rates with unreported reasons and the absence of appropriate analysis to assess the impact of missing outcome data.

XSL•FO
RenderX

Figure 3. Risk of bias assessment of 29 included studies [19-23,31-54].



XSL•FO
RenderX

Of the 11 outcomes included, the overall quality of evidence was assessed to be very low to moderate according to the GRADE assessment (Table 3). The risk of bias was considered serious based on the results of ROB 2 assessments. Indirectness was not considered serious, as the outcomes were directly measured. No publication bias was detected for any of the outcomes. Inconsistency ranged from not serious to very serious based on the heterogeneity values, and imprecision ranged from not serious to serious, as determined by the 95% CI.

Table . Grading of Recommendation, Assessment, Development, and Evaluation assessment.

Outcome	Studies, n/N (%)	Quality assessment					Patients (I[a]/C[b]), n/N	Effect, mean difference (95% CI)	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias			
Weight	28/29 (97)	Serious	Very Serious	Not serious	Not serious	Undetected	1743/1680	−1.45 (−2.01 to −0.89)	Very Low
BMI	18/29 (62)	Serious	Serious	Not serious	Not serious	Undetected	1241/1216	−0.35 (−0.57 to −0.13)	Low
WC[c]	12/29 (41)	Serious	Very serious	Not serious	Serious	Undetected	897/877	−1.98 (−3.42 to −0.55)	Very Low
Fat mass	10/29 (34)	Serious	Serious	Not serious	Not serious	Undetected	537/521	−1.32 (−1.94 to −0.69)	Low
Fat mass percentage	11/29 (38)	Serious	Very serious	Not serious	Not serious	Undetected	630/632	−0.40 (−1.00 to 0.19)	Very Low
MVPA[d]	6/29 (21)	Serious	Not serious	Not serious	Serious	Undetected	319/248	−0.69 (−5.67 to 4.28)	Low
Energy intake	6/29 (21)	Serious	Very serious	Not serious	Serious	Undetected	619/573	−62.72 (−181.62 to 56.18)	Very Low
SBP[e]	9/29 (31)	Serious	Very serious	Not serious	Serious	Undetected	493/502	−0.14 (−2.66 to 2.37)	Very Low
DBP[f]	8/29 (28)	Serious	Very serious	Not serious	Not serious	Undetected	460/469	−1.76 (−3.47 to −0.04)	Very Low
Triglycerides	6/29 (21)	Serious	Not serious	Not serious	Not serious	Undetected	294/293	0.06 (−0.19 to 0.31)	Moderate
HbA_{1c}[g]	5/29 (17)	Serious	Very serious	Not serious	Not serious	Undetected	316/275	−0.13 (−0.22 to −0.04)	Very Low

[a]I: intervention group.

[b]C: control group.

[c]WC: waist circumference.

[d]MVPA: moderate-to-vigorous physical activity.

[e]SBP: systolic blood pressure.

[f]DBP: diastolic blood pressure.

[g]HbA_{1c}: hemoglobin A_{1c}.

Effects of Mobile Phone App

The review examined the effects of mobile phone app interventions on weight-related outcomes (weight, BMI, WC, fat mass, and fat mass percentage), behavioral outcomes (MVPA and energy intake), and metabolic outcomes (SBP, DBP, triglycerides, and HbA_{1c}). A summary of the meta-analyses' results was shown in Table 4.

XSL•FO
RenderX

Table . A summary of meta-analyses results on each outcome.

Outcomes	Studies, n/N (%)	Sample size (intervention group), n	Sample size (control group), n	Mean difference (95% CI)	*P* value	I^2 (%)
Weight-related outcomes						
Weight (kg)	28/29 (97)	1743	1680	−1.45 (−2.01 to −0.89)	<.001	56
BMI (kg/m^2)	18/29 (62)	1241	1216	−0.35 (−0.57 to −0.13)	.002	43
WC[a] (cm)	12/29 (41)	897	877	−1.98 (−3.42 to −0.55)	.007	82
Fat mass (kg)	10/29 (34)	537	521	−1.32 (−1.94 to −0.69)	<.001	31
Fat mass percentage (%)	11/29 (38)	630	632	−0.40 (−1.00 to 0.19)	.18	60
Behavioral outcomes						
MVPA[b] (mins/day)	6/29 (21)	319	248	−0.69 (−5.67 to 4.28)	.78	0
Energy intake (kcal/day)	6/29 (21)	619	573	−62.72 (−181.62 to 56.18)	.30	62
Metabolic outcomes						
SBP[c] (mm Hg)	9/29 (31)	493	502	−0.14 (−2.66 to 2.37)	.91	66
DBP[d] (mm Hg)	8/29 (28)	460	469	−1.76 (−3.47 to −0.04)	.04	61
Triglycerides (mmol/L)	6/29 (21)	294	293	0.06 (−0.19 to 0.31)	.64	9
HbA_{1c}[e] (%)	5/29 (17)	316	275	−0.13 (−0.22 to -0.04)	.005	82

[a]WC: waist circumference.

[b]MVPA: moderate-to-vigorous physical activity.

[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

[e]HbA_{1c}: hemoglobin A_{1c}.

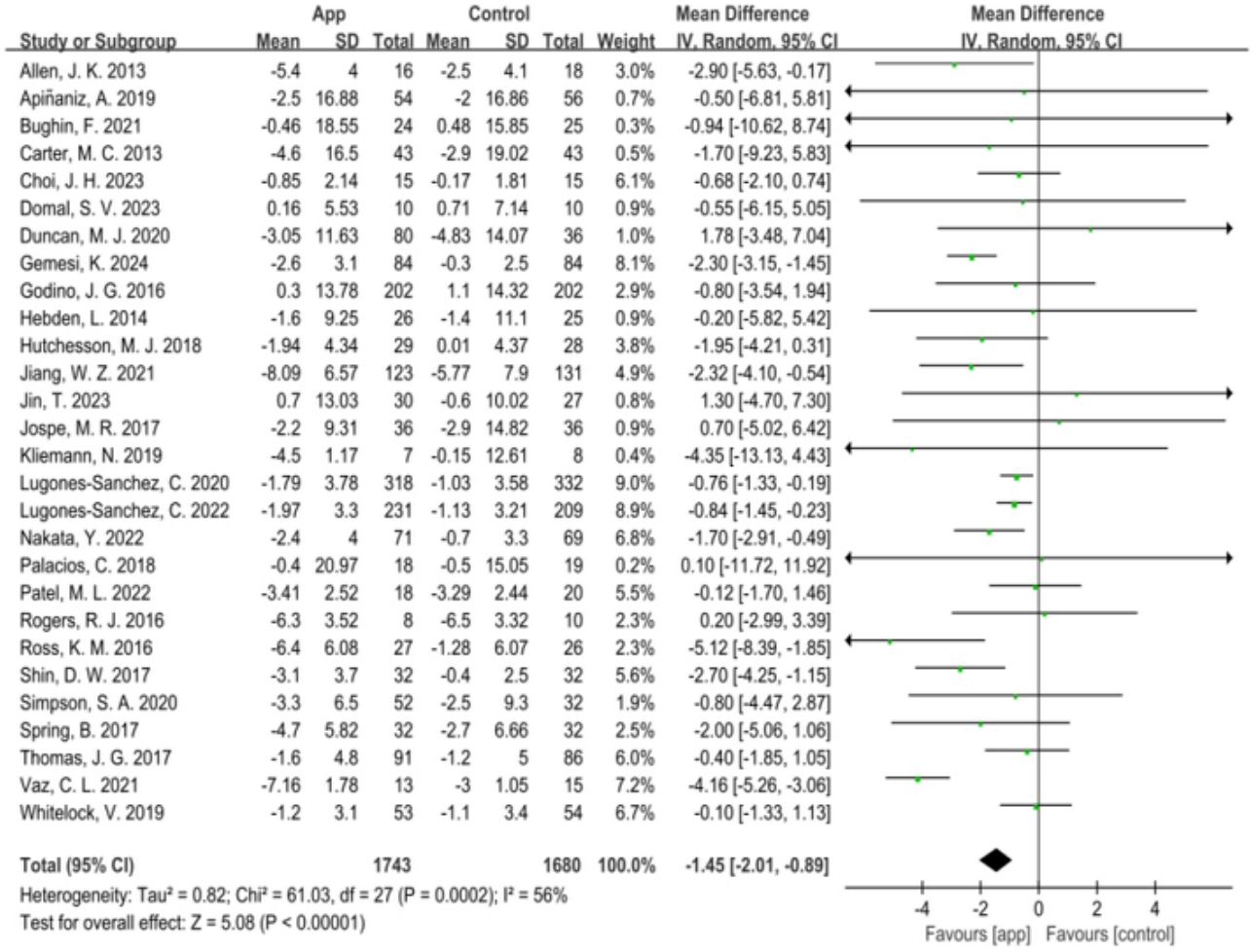
Weight-Related Outcomes

Weight

Twenty-nine RCTs [19-23,31-54] measured weight as an outcome, but 1 study [52] was excluded from the meta-analysis due to unavailable outcome data. Therefore, the outcomes of 28 studies involving 1743 participants in the intervention group and 1680 participants in the control group were pooled. The results indicated that mobile phone app interventions significantly reduced weight (MD=−1.45 kg, 95% CI −2.01 to −0.89; *P*<.001), with modest heterogeneity (X^2=61.03, df=27, I^2=56%), as determined by the random-effects model (Figure 4).

XSL•FO
RenderX

Figure 4. Meta-analysis of weight [19-23,31-51,53,54].



BMI

Eighteen studies [19-21,32-34,36,38,39,41-46,50,52,53] assessing BMI were included in the meta-analysis. The mobile phone app intervention resulted in a significant reduction in BMI (MD=−0.35 kg/m^2, 95% CI −0.57 to −0.13; *P*=.002), with moderate heterogeneity (X^2=29.91, df=17, I^2=43%; Multimedia Appendix 7).

WC

Twelve studies [20,21,38-40,42-44,46,49,50,53] with 897 participants in the intervention group were included in the meta-analysis to examine the effects of mobile phone app intervention on WC. The results showed that WC was significantly reduced (MD=−1.98 cm, 95% CI −3.42 to −0.55; *P*=.007), with high heterogeneity (X^2=62.49, df=11, I^2=82%; Multimedia Appendix 8).

Fat Mass

The effect on fat mass was measured by 11 RCTs [19,20,32,33,38,39,42,43,46,53,54], with a total of 1058 participants pooled for the meta-analysis. Fat mass was significantly reduced (MD=−1.32 kg, 95% CI −1.94 to −0.69; *P*<.001) due to the mobile app intervention. Heterogeneity was low across the included studies (X^2=13.11, df=9, I^2=31%; Multimedia Appendix 9).

Fat Mass Percentage

Eleven RCTs [19,20,32-34,37,38,42,43,46,54] were included to examine the effects on fat mass percentage. The mobile app intervention did not lead to a statistically significant reduction in fat mass percentage (MD=−0.40, *P*=.18, 95% CI −1.00 to 0.19), and the heterogeneity was moderate (X^2=25.08, df=10, I^2=60%; Multimedia Appendix 10).

Behavioral Outcomes (MVPA and Energy Intake)

The mobile phone app intervention decreased MVPA, but this change was not statistically significant (MD=−0.69, 95% CI −5.67 to 4.28; *P*=.78; Multimedia Appendix 11). No heterogeneity was observed. When pooling the data on energy intake in the meta-analysis, the results indicated a nonsignificant effect size (MD=−62.72 kcal/day, 95% CI −181.62 to 56.18; *P*=.30) with moderate heterogeneity (X^2=13.33, df=5, I^2=62%; Multimedia Appendix 12).

Metabolic Outcomes (SBP, DBP, Triglycerides, and HbA_{1c})

SBP was nonsignificantly reduced (MD=−0.14 mm Hg, 95% CI −2.66 to 2.37; *P*=.91). The mobile phone app intervention significantly decreased DBP (MD=−1.76 mm Hg, 95% CI −3.47 to −0.04; *P*=.04). It led to a nonsignificant increase in triglyceride levels (MD=0.06 mmol/L, 95% CI −0.19 to 0.31; *P*=.64). In addition, there was a significant decrease in HbA_{1c}

XSL•FO
RenderX

levels (MD=−0.13%, 95% CI −0.22 to −0.04; *P*=.005), with high heterogeneity (X^2=22.79, df=4, I^2=82%; Figures S7-S10 in Multimedia Appendices 13-16).

Subgroup and Sensitivity Analysis of Weight

Five subgroup analyses were conducted to explore potential factors influencing the effects of the apps, intervention content (single intervention [diet or PA] versus combined intervention [diet+PA]), theory presence (theory-based versus nontheory-based), intervention duration (short-term [≤3 months] versus medium-term [6 months] versus long-term [>6 months]), number of BCRs used (3 resources vs <3 resources), and number of BCTs used (<median [8] vs ≥median [8]).

Among these, the combined intervention demonstrated a greater effect size (MD=−1.82, 95% CI −2.48 to −1.16 kg) compared with the single intervention group (MD=−0.24, 95% CI −1.00 to 0.53 kg), with statistical significance. Regarding intervention duration, the medium-term (6 mo) group exhibited the largest effects (MD=−2.50, 95% CI −3.65 to −1.35 kg), significantly outperforming the other two groups. In addition, the number of BCTs used ≥8 had a greater effect size (MD=−1.83, 95% CI −2.56 to −1.09 kg vs MD=−0.61, 95% CI −1.28 to −0.07 kg; *P*=.02; Table 5).

Table . Subgroup analyses.

Variables and subgroups	Number of studies	MD[a] (95% CI)	I^2(%)	*P* value of subgroup difference
Intervention contents				.002
Single intervention (diet or PA)	9	−0.24 (−1.00 to 0.53)	0	
Combined intervention (diet+PA[b])	19	−1.82 (−2.48 to –1.16)	64	
Theory				.34
Theory-based	8	−0.97 (−1.96 to 0.02)	0	
Nontheory-based	20	−1.54 (−2.20 to –0.89)	65	
Intervention duration				.04
Short-term (≤3 months）	14	−1.12 (−1.61 to −0.62)	36	
Medium-term (6 months)	10	−2.50 (−3.65 to −1.35)	42	
Long-term (＞6 months)	4	−0.47 (−1.65 to 0.71)	0	
BCRs[c]				.44
=3 resources	17	−1.56 (−2.23 to –0.88)	67	
<3 resources	11	−1.10 (−2.06 to –0.13)	14	
BCTs[d]				.02
<8	12	−0.61 (−1.28 to 0.07)	0	
≥8	16	−1.83 (−2.56 to –1.09)	70	

[a]MD: mean difference.

[b]PA: physical activity.

[c]BCR: behavior change resources.

[d]BCT: behavior change techniques.

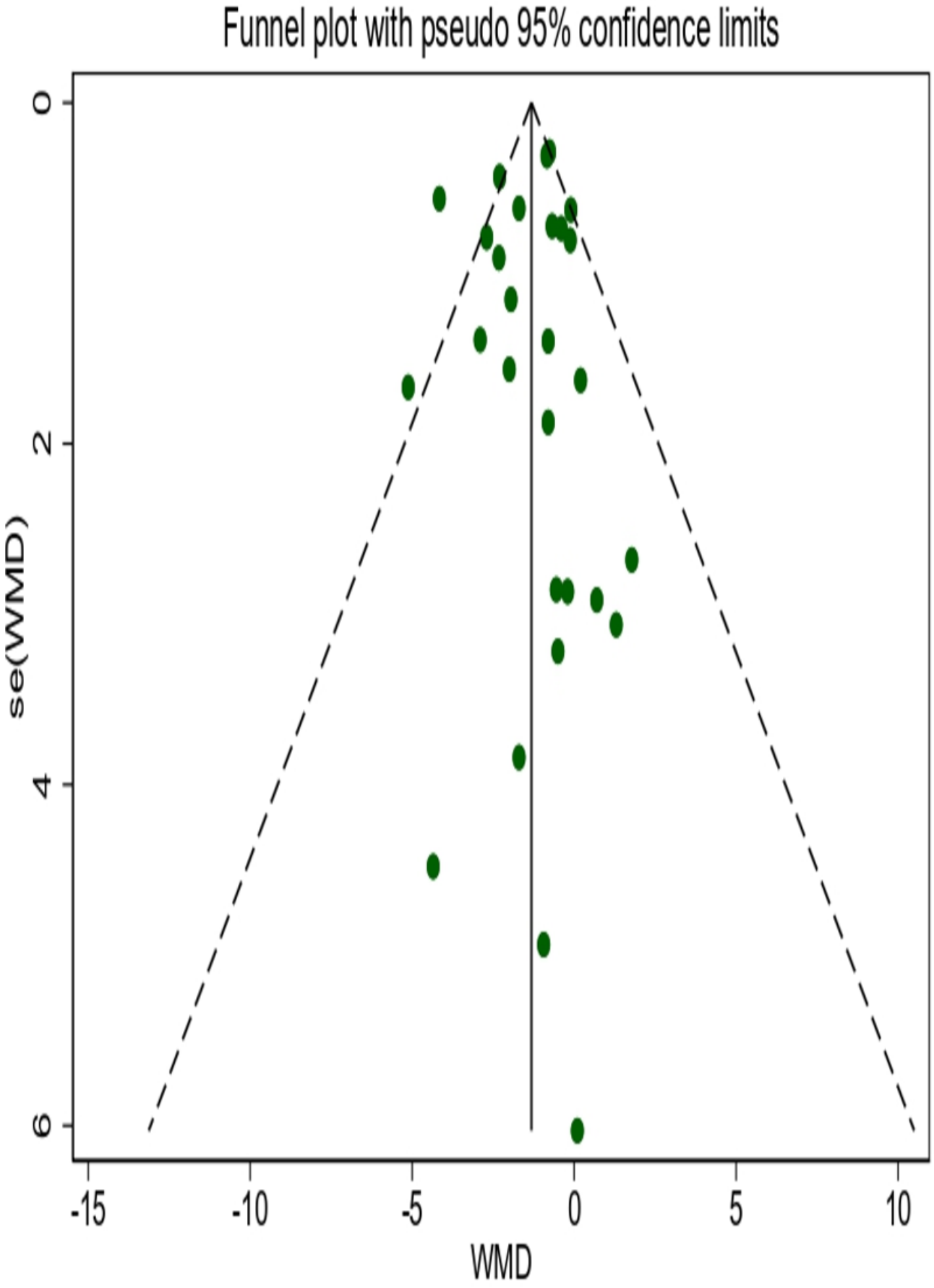
Sensitivity analyses were performed using the leave-one-out and high-risk study exclusion methods. Notably, after excluding one study from the meta-analysis, the weight reduction effect size changed from −1.45 kg (95% CI −2.01 to −0.89) to −1.20 kg (95% CI −1.61 to −0.79), indicating that this study had a relatively significant influence on the overall effect size [49]. The heterogeneity also decreased, from 56% to 23%. In addition, excluding high-risk studies resulted in a weight reduction of −1.24 kg (95% CI −1.84 to −0.83) and an I^2 of 36%. High-risk studies may overestimate the effect size. However, the sensitivity analyses demonstrate that the results are robust, as no statistically significant changes were observed. This specific study and high-risk studies may account for the heterogeneity to some extent.

Publication Bias

Funnel plot and Egger test for publication bias were conducted when a minimum of 10 studies were incorporated into the meta-analysis. No significant publication bias was detected for weight (Egger test, t_{27}=−0.28; *P*=.78), BMI (Egger test, t_{17}=0.63; *P*=.54), WC (Egger test, t_{11}=0.10; *P*=.92), fat mass (Egger test, t_9=0.54; *P*=.61), or fat mass percentage (Egger test, t_{10}=−0.69; *P*=.51). The funnel plots were displayed in Figure 5 and Figure S11-S14 in Multimedia Appendices 17-20.

XSL•FO
RenderX

Figure 5. Funnel plot of weight. WMD: weighted mean difference.



XSL•FO
RenderX

Discussion

Principal Findings

A total of 29 RCTs [19-23,31-54] were included in this systematic review and meta-analysis, examining 11 outcomes. The results of this review indicated that mobile app interventions significantly reduced weight, BMI, WC, fat mass, DBP, and HbA_{1c}. However, nonsignificant effects were observed for other outcomes, including fat mass percentage, behavioral outcomes (ie, MVPA and energy intake), and metabolic outcomes (ie, SBP and triglycerides). BCTs and BCRs used in the included studies were also identified, with most studies (59%) using 3 resource types (ie, facilitating, boosting, and nudging). In addition, intervention content, duration, and the number of BCTs used were identified as influencing factors through subgroup analyses.

Interpretation of Findings

Effects of Mobile Phone App on Weight-Related Outcomes

The meta-analysis suggested that interventions delivered via mobile phone apps significantly reduced weight among adults with overweight and obesity, which is consistent with findings from previous meta-analyses [14,16]. Notably, in our meta-analysis, the effect size for weight loss was 1.45 kg, while meta-analyses conducted by Chew et al [16] and Antoun et al [14] reported weight loss values of 2.55 kg and 1.99 kg, respectively. The main reason for this difference may be that the two aforementioned reviews included interventions involving both mobile phone apps and nonmobile interventions, whereas our meta-analysis focused on studies using interventions delivered solely by mobile apps or applications combined with nonmobile interventions [55]. This suggests that, compared with app interventions alone, the combination of app interventions with other strategies, such as health coaching, is more effective. Overall, mobile app–based interventions are a promising and cost-effective way of supporting weight loss.

According to the results of the subgroup analyses, the combined intervention group (diet+PA) showed better outcomes than the single intervention group (diet or PA), which is consistent with the results of previous studies [56,57]. These findings suggest that future researchers should design weight-loss apps targeting both diet and PA, rather than focusing on a single behavior change. Intervention duration is also a significant factor influencing the effects, with the greatest reductions observed in the medium term. This can be explained by the phenomenon where the effects of mobile apps show initial improvements, peak at 6 months, and then gradually decrease due to a decline in behavior change adherence over time. Therefore, the sustainability of weight-loss effects through the effective use of behavior change strategies is crucial for successful weight loss.

High BMI is associated with various conditions, including cardiovascular disease (CVD), certain cancers, and breathing problems [58]. BMI was also significantly reduced in our meta-analysis. This finding contrasts with the results of a previous meta-analysis [16], which may be attributed to the discrepancy in the number of included original studies (18 vs 3). Therefore, our meta-analysis may provide a more comprehensive insight.

WC, a reliable indicator of not only total body fat but also abdominal visceral fat [59], is an accurate predictor of disease [60]. It exhibits superior efficacy as a marker for all-cause mortality, the incidence of cardiovascular disease [61], and metabolic syndrome [62] compared with BMI [63-65], which solely represents total body fat mass. This association between WC and these health outcomes can be explained by the cellular mechanisms linking metabolic syndrome to a proinflammatory state, with visceral obesity playing a central role. In addition, visceral obesity disrupts the normal physiological balance of adipokines, insulin resistance, and endothelial dysfunction, creating a proatherogenic state. Therefore, visceral obesity is closely associated with the incidence of cardiovascular disease [66]. Given the importance of WC in clinical practice, we consider it a critical outcome. However, among the 29 included studies, only 12 (41%) RCTs reported the effects of mobile apps on WC. We recommend incorporating WC as an outcome in future original studies.

After combining data from 12 RCTs in the meta-analysis, WC was significantly reduced through the use of weight loss apps among individuals with overweight and obesity, consistent with a review performed by Cai et al [15], but inconsistent with a previous meta-analysis by Chew et al [67]. This discrepancy may be explained by the fact that only four RCTs were included in Chew et al's review [67] to examine the effects on WC, 2 of which focused on the older population. This may have reduced the effects of mobile phone apps on WC, as basal metabolic rate decreases almost linearly with age [68], making it more difficult for older adults to lose weight.

Our meta-analysis also identified fat mass and fat mass percentage as primary outcomes, revealing a significant reduction in fat mass but not in fat mass percentage. This finding suggests that body components other than fat, such as lean muscle mass and bone mass, may have also decreased during the weight loss period. However, these components are crucial for body functions and overall health. Therefore, future research should focus on strategies aimed specifically at fat reduction, such as individualized fat-loss plans and physical fitness training.

Effects of Mobile Phone App on Behavioral and Metabolic Outcomes

Lifestyle modification, including increasing physical activity and restricting calories, is considered the cornerstone of obesity management [69,70]. Based on the effectiveness of weight loss apps observed in our meta-analysis, it is worth discussing the underlying reasons behind these lifestyle interventions. Surprisingly, an increase in MVPA was not observed, which may be attributed to several factors. Many weight loss apps are designed to help users self-monitor their PA rather than actively encourage PA [15,21,32]. Our review found that using mobile phone apps can reduce energy intake by approximately 100 kcal/day, which may partly explain the effectiveness of weight loss. However, this reduction was not statistically significant. We recommend that behavior change strategies be more

XSL•FO
RenderX

thoroughly considered in the design of future mobile app–based interventions.

In our meta-analysis, we explored the effects on four outcomes (ie, SBP, DBP, triglycerides, and HbA_{1c}) and observed significant reductions in DBP and HbA_{1c}. However, these reductions did not result in clinically significant benefits. This phenomenon may be explained by two factors. First, reductions in blood pressure, triglycerides, and HbA_{1c} are influenced by various factors (eg, pharmacological treatment, diet, and PA) and are not solely dependent on weight management. Weight loss mobile apps are not specifically designed to lower blood pressure, triglycerides, or HbA_{1c}. As a result, the intervention strategies within the weight loss app may not be sufficient to meet the requirements for reducing blood pressure, lowering triglycerides, and improving HbA_{1c} levels. Given that individuals with overweight or obesity are at higher risk of developing metabolic syndromes [62], the app could be redesigned to better serve the needs of this high-risk population.

Second, while weight reduction was statistically significant in our review, the extent of weight loss may be insufficient to achieve a clinically meaningful reduction in metabolic outcomes [71,72]. A previous study [73] demonstrated that a 5% weight loss significantly decreases plasma concentrations (eg, glucose, insulin, triglycerides, and leptin) of certain risk factors for cardiometabolic disease. A systematic review [74] also indicated that for every 10 kg reduction in weight, there is a corresponding decrease of 6.0 mm Hg in SBP and 4.6 mm Hg in DBP levels. However, due to the limited number of RCTs investigating the effects on metabolic outcomes, our meta-analysis is based on a relatively small number of original studies. Given the importance of clinical significance in weight loss, more RCTs that include assessments of metabolic outcomes are needed to clarify the relationship between weight loss and metabolic outcomes.

Interpretation of Findings of BCTs and BCRs

To our knowledge, this is the first review to identify BCRs used in mobile app interventions for adults with overweight and obesity. The classification standards were based on the BCRM proposed by Michaelsen in 2022 [26]. Our findings indicate that 59% of studies used 3 resource types in their intervention designs, 21% used 2 types, 14% used 1 type, and 2 studies did not use any resources. Standardization of these classification standards is needed for the future, and further exploration of optimal resource combinations is warranted. Based on the results of the subgroup analyses, the group using ≥8 BCTs showed better outcomes than the group using <8 BCTs, suggesting that a higher number of BCTs was associated with better effect sizes to some extent. Future studies should explore more effective combinations of BCTs to enhance their effects.

Strengths and Limitations

Several strengths can be identified in this review. First, a thorough examination of the effects of mobile phone apps on weight loss among adults with overweight and obesity was conducted. Our meta-analysis includes 11 outcomes based on findings from 29 studies. Weight-related outcomes reflect the direct effects of mobile apps, metabolic outcomes provide evidence of the association between weight reduction and health benefits, and behavioral outcomes inform researchers about the extent to which current interventions target diet and PA, offering guidance for future intervention design. This comprehensive set of outcomes offers substantial insights into the current effectiveness and status of mobile apps in weight loss interventions. Second, only RCTs were included in this review, as they are considered the gold standard in clinical research. This selection criterion ensures that our findings are based on high-quality studies, minimizing bias and enhancing the reliability of the results. Third, the BCTs used in the included RCTs were identified. Systematically reviewing the use of BCTs contributes to the standardization of intervention strategies and enhances comparability across studies, especially given the considerable methodological variability in current weight loss interventions. This provides a reference for researchers to replicate, refine, or design new intervention approaches. Fourth, mapping BCTs to BCRs based on the BCRM helps clarify the resources required to implement specific BCTs during intervention design. This approach also supports more tailored and practical intervention strategies by taking into account the resources available to patients, thereby promoting interventions that are both targeted and feasible.

However, there are also some limitations in this study. First, we only included adults with overweight and obesity without pre-existing medical conditions, which may limit the generalizability of the results to broader populations, such as individuals with diabetes or CVD. However, adults with overweight and obesity without medical conditions represent a critical population from the perspective of preventive medicine, as they are at high risk of developing chronic diseases. Early interventions targeting this population group could potentially reduce the burden of future disease onset. Moreover, there are notable differences in intervention responsiveness and health needs between adults with overweight and obesity with and without medical conditions. By including only participants without comorbidities, this study allows for a more focused analysis of lifestyle-based mobile app interventions, without the influence of complex confounding factors such as medication use or the bidirectional interactions between obesity and chronic diseases. This approach allows for a clearer assessment of behavior change outcomes, especially in the identification of BCTs and the associated BCRs. Second, we identified the BCTs and BCRs used in the RCT interventions included in this review. Although interrater agreement in coding the BCTs and resources was high, the coding process inherently involves a degree of subjectivity, which may influence the results related to BCTs and BCRs. Furthermore, given the relatively limited app of the BCRM in previous research, standardized procedures for mapping BCTs to corresponding resources have not yet been established. The mapping criteria proposed in this study, therefore, require further testing and validation in future research. Finally, the effectiveness of weight loss was influenced not only by the quantity of BCTs and resources used but also by their quality, including factors such as the intensity, frequency, and tailoring of interventions to individual needs. As a result, no meta-regression was conducted to link specific BCTs or BCRs to their effectiveness. We advocate for the

XSL•FO
RenderX

conducting of more RCTs to better clarify the specific effects of individual BCT or BCR.

Conclusions

This meta-analysis presents findings that demonstrate that mobile phone app interventions significantly reduce weight, BMI, WC, fat mass, DBP, and HbA_{1c} among adults with overweight and obesity. These mobile phone app interventions are cost-effective and can be applied to a large population. However, current mobile apps have not achieved clinically significant weight loss. Future studies should focus on optimizing app interventions by incorporating more effective behavior change strategies and resources to enhance their overall effectiveness.

Acknowledgments

This study was funded by the Non-Profit Central Research Institute Fund of Chinese Academy of Medical Sciences (grant 2023-RC320-01).

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request. Generative AI was used to proofread this manuscript to enhance its readability and clarity. The use of AI tools was aimed at improving the overall flow and presentation of the text without altering the original meaning or content.

Authors' Contributions

SJL contributed to conceptualization, methodology, original draft preparation, and review and editing of the manuscript. Y Zhou was involved in conceptualization, methodology, and data curation. YT contributed to software development, data curation, and formal analysis. Haoming Ma was responsible for methodology, software development, and data curation. Y Zhang contributed to the review and editing of the manuscript. AW was involved in conceptualization. XT contributed to methodology and software development. RP contributed to methodology and software development. MHP was involved in conceptualization, methodology, formal analysis, supervision, and the review and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 16 KB - mhealth_v13i1e63313_app1.docx]

Multimedia Appendix 2
Taxonomy of behavioral change techniques.
[PDF File, 189 KB - mhealth_v13i1e63313_app2.pdf]

Multimedia Appendix 3
Reasons for exclusion of studies.
[DOCX File, 35 KB - mhealth_v13i1e63313_app3.docx]

Multimedia Appendix 4
Characteristics of the included studies.
[DOCX File, 93 KB - mhealth_v13i1e63313_app4.docx]

Multimedia Appendix 5
Behavioral change techniques identified in each study.
[DOCX File, 48 KB - mhealth_v13i1e63313_app5.docx]

Multimedia Appendix 6
Mapping behavioral change techniques to behavioral change resources.
[DOCX File, 18 KB - mhealth_v13i1e63313_app6.docx]

Multimedia Appendix 7
Meta-analysis of BMI.

XSL•FO
RenderX

[PDF File, 822 KB - mhealth_v13i1e63313_app7.pdf]

Multimedia Appendix 8
Meta-analysis of waist circumference.
[PDF File, 608 KB - mhealth_v13i1e63313_app8.pdf]

Multimedia Appendix 9
Meta-analysis of fat mass.
[PDF File, 546 KB - mhealth_v13i1e63313_app9.pdf]

Multimedia Appendix 10
Meta-analysis of fat mass percentage.
[PDF File, 588 KB - mhealth_v13i1e63313_app10.pdf]

Multimedia Appendix 11
Meta-analysis of moderate-to-vigorous physical activity.
[PDF File, 417 KB - mhealth_v13i1e63313_app11.pdf]

Multimedia Appendix 12
Meta-analysis of energy intake.
[PDF File, 493 KB - mhealth_v13i1e63313_app12.pdf]

Multimedia Appendix 13
Meta-analysis of systolic blood pressure.
[PDF File, 519 KB - mhealth_v13i1e63313_app13.pdf]

Multimedia Appendix 14
Meta-analysis of diastolic blood pressure.
[PDF File, 477 KB - mhealth_v13i1e63313_app14.pdf]

Multimedia Appendix 15
Meta-analysis of triglycerides.
[PDF File, 395 KB - mhealth_v13i1e63313_app15.pdf]

Multimedia Appendix 16
Meta-analysis of hemoglobin A_{1c}.
[PDF File, 380 KB - mhealth_v13i1e63313_app16.pdf]

Multimedia Appendix 17
Funnel plot of BMI.
[PDF File, 58 KB - mhealth_v13i1e63313_app17.pdf]

Multimedia Appendix 18
Funnel plot of fat mass percentage.
[PDF File, 58 KB - mhealth_v13i1e63313_app18.pdf]

Multimedia Appendix 19
Funnel plot of waist circumference.
[PDF File, 58 KB - mhealth_v13i1e63313_app19.pdf]

Multimedia Appendix 20
Funnel plot of fat mass.
[PDF File, 58 KB - mhealth_v13i1e63313_app20.pdf]

Checklist 1

XSL•FO
RenderX

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 160 KB - mhealth_v13i1e63313_app21.pdf]

References

1. Obesity and overweight. WHO. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2025-07-02]
2. GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 6;377(1):13-27. [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
3. Richard Dobbs CS, Thompson F. Overcoming Obesity: An Initial Economic Analysis: McKinsey Global Institute; 2014.
4. Nyberg ST, Batty GD, Pentti J, et al. Obesity and loss of disease-free years owing to major non-communicable diseases: a multicohort study. Lancet Public Health 2018 Oct;3(10):e490-e497. [doi: 10.1016/S2468-2667(18)30139-7] [Medline: 30177479]
5. Stephenson J, Smith CM, Kearns B, Haywood A, Bissell P. The association between obesity and quality of life: a retrospective analysis of a large-scale population-based cohort study. BMC Public Health 2021 Nov 3;21(1):1990. [doi: 10.1186/s12889-021-12009-8] [Medline: 34732156]
6. Vidra N, Trias-Llimós S, Janssen F. Impact of obesity on life expectancy among different European countries: secondary analysis of population-level data over the 1975-2012 period. BMJ Open 2019 Jul 31;9(7):e028086. [doi: 10.1136/bmjopen-2018-028086] [Medline: 31371290]
7. Tremmel M, Gerdtham UG, Nilsson PM, Saha S. Economic burden of obesity: a systematic literature review. Int J Environ Res Public Health 2017 Apr 19;14(4):28422077. [doi: 10.3390/ijerph14040435] [Medline: 28422077]
8. Hammond RA, Levine R. The economic impact of obesity in the United States. Diabetes Metab Syndr Obes 2010 Aug 30;3(285-95):285-295. [doi: 10.2147/DMSOTT.S7384] [Medline: 21437097]
9. Lagerros YT, Rössner S. Obesity management: what brings success? Therap Adv Gastroenterol 2013 Jan;6(1):77-88. [doi: 10.1177/1756283X12459413] [Medline: 23320052]
10. de Jong M, Jansen N, van Middelkoop M. A systematic review of patient barriers and facilitators for implementing lifestyle interventions targeting weight loss in primary care. Obes Rev 2023 Aug;24(8):e13571. [doi: 10.1111/obr.13571] [Medline: 37226636]
11. Arroyo AC, Zawadzki MJ. The implementation of behavior change techniques in mHealth apps for sleep: systematic review. JMIR Mhealth Uhealth 2022 Apr 4;10(4):e33527. [doi: 10.2196/33527] [Medline: 35377327]
12. MHealth: new horizons for health through mobile technologie. WHO. 2011. URL: https://www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technologie [accessed 2025-07-02]
13. Yang Q, Van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR Mhealth Uhealth 2019 Apr 3;7(4):e11244. [doi: 10.2196/11244] [Medline: 30942695]
14. Antoun J, Itani H, Alarab N, Elsehmawy A. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 8;10(4):e35479. [doi: 10.2196/35479] [Medline: 35394443]
15. Cai X, Qiu S, Luo D, Wang L, Lu Y, Li M. Mobile application interventions and weight loss in type 2 diabetes: a meta-analysis. Obesity (Silver Spring) 2020 Mar;28(3):502-509. [doi: 10.1002/oby.22715] [Medline: 31965748]
16. Chew HSJ, Rajasegaran NN, Chin YH, Chew WSN, Kim KM. Effectiveness of combined health coaching and self-monitoring apps on weight-related outcomes in people with overweight and obesity: systematic review and meta-analysis. J Med Internet Res 2023 Apr 18;25:e42432. [doi: 10.2196/42432] [Medline: 37071452]
17. Hutchesson MJ, Rollo ME, Krukowski R, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
18. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
19. Choi JH, Kim SW, Seo J, et al. Effects of a mobile-health exercise intervention on body composition, vascular function, and autonomic nervous system function in obese women: a randomized controlled trial. J Multidiscip Healthc 2023;16(1601-15):1601-1615. [doi: 10.2147/JMDH.S406905] [Medline: 37313274]
20. Domal SV, Chandrasekaran B, Palanisamy HP. Influence of smartphone-based physical activity intervention on executive functions and cardiometabolic disease risk in obese young adults: a pilot randomised controlled trial. J Diabetes Metab Disord 2023 Jun;22(1):619-628. [doi: 10.1007/s40200-023-01182-9] [Medline: 37255799]
21. Lugones-Sanchez C, Recio-Rodriguez JI, Agudo-Conde C, et al. Long-term effectiveness of a smartphone app combined with a smart band on weight loss, physical activity, and caloric intake in a population with overweight and obesity (evident 3 study): randomized controlled trial. J Med Internet Res 2022 Feb 1;24(2):e30416. [doi: 10.2196/30416] [Medline: 35103609]
22. Nakata Y, Sasai H, Gosho M, et al. A smartphone healthcare application, CALO mama Plus, to promote weight loss: a randomized controlled trial. Nutrients 2022 Nov 2;14(21):4608. [doi: 10.3390/nu14214608] [Medline: 36364870]

XSL•FO
RenderX

23. Patel ML, Cleare AE, Smith CM, Rosas LG, King AC. Detailed versus simplified dietary self-monitoring in a digital weight loss intervention among racial and ethnic minority adults: fully remote, randomized pilot study. JMIR Form Res 2022 Dec 13;6(12):e42191. [doi: 10.2196/42191] [Medline: 36512404]
24. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
25. Carey RN, Connell LE, Johnston M, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2019 Jul 17;53(8):693-707. [doi: 10.1093/abm/kay078] [Medline: 30304386]
26. Michaelsen MM, Esch T. Functional mechanisms of health behavior change techniques: a conceptual review. Front Psychol 2022;13(725644):725644. [doi: 10.3389/fpsyg.2022.725644] [Medline: 35369223]
27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
28. Chandler J, Cumpston M, Li T, Page MJ. In: Welch VA, editor. Cochrane Handbook for Systematic Reviews of Interventions Version 2023, Vol. 6:4.
29. Chandler J, Cumpston M, Li T, Page MJ, Welch VA. In: Higgins JT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 65: Cochrane; 2024.
30. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
31. Apiñaniz A, Cobos-Campos R, Sáez de Lafuente-Moríñigo A, et al. Effectiveness of randomized controlled trial of a mobile app to promote healthy lifestyle in obese and overweight patients. Fam Pract 2019 Nov 18;36(6):699-705. [doi: 10.1093/fampra/cmz020] [Medline: 31093681]
32. Lugones-Sanchez C, Sanchez-Calavera MA, Repiso-Gento I, et al. Effectiveness of an mHealth intervention combining a smartphone app and smart band on body composition in an overweight and obese population: randomized controlled trial (EVIDENT 3 study). JMIR Mhealth Uhealth 2020 Nov 26;8(11):e21771. [doi: 10.2196/21771] [Medline: 33242020]
33. Bughin F, Bui G, Ayoub B, et al. Impact of a mobile telerehabilitation solution on metabolic health outcomes and rehabilitation adherence in patients with obesity: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 6;9(12):e28242. [doi: 10.2196/28242] [Medline: 34874887]
34. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32. [doi: 10.2196/jmir.2283] [Medline: 23587561]
35. Kliemann N, Croker H, Johnson F, Beeken RJ. Development of the top tips habit-based weight loss app and preliminary indications of its usage, effectiveness, and acceptability: mixed-methods pilot study. JMIR Mhealth Uhealth 2019 May 10;7(5):e12326. [doi: 10.2196/12326] [Medline: 31094352]
36. Simpson SA, Matthews L, Pugmire J, et al. An app-, web- and social support-based weight loss intervention for adults with obesity: the “HelpMeDoIt!” feasibility randomised controlled trial. Pilot Feasibility Stud 2020;6(1):133. [doi: 10.1186/s40814-020-00656-4] [Medline: 32968544]
37. Whitelock V, Kersbergen I, Higgs S, Aveyard P, Halford JCG, Robinson E. A smartphone based attentive eating intervention for energy intake and weight loss: results from a randomised controlled trial. BMC Public Health 2019 May 21;19(1):611. [doi: 10.1186/s12889-019-6923-x] [Medline: 31113400]
38. Shin DW, Yun JM, Shin JH, et al. Enhancing physical activity and reducing obesity through smartcare and financial incentives: A pilot randomized trial. Obesity (Silver Spring) 2017 Feb;25(2):302-310. [doi: 10.1002/oby.21731] [Medline: 28063226]
39. Jin T, Kang G, Song S, et al. The effects of dietary self-monitoring intervention on anthropometric and metabolic changes via a mobile application or paper-based diary: a randomized trial. Nutr Res Pract 2023 Dec;17(6):1238-1254. [doi: 10.4162/nrp.2023.17.6.1238] [Medline: 38053827]
40. Duncan MJ, Fenton S, Brown WJ, et al. Efficacy of a multi-component m-Health weight-loss intervention in overweight and obese adults: a randomised controlled trial. Int J Environ Res Public Health 2020 Aug 26;17(17):6200. [doi: 10.3390/ijerph17176200] [Medline: 32859100]
41. Hebden L, Cook A, van der Ploeg HP, King L, Bauman A, Allman-Farinelli M. A mobile health intervention for weight management among young adults: a pilot randomised controlled trial. J Hum Nutr Diet 2014 Aug;27(4):322-332. [doi: 10.1111/jhn.12155] [Medline: 23992038]
42. Hutchesson MJ, Callister R, Morgan PJ, et al. A targeted and tailored eHealth weight loss program for young women: the be positive be healthe randomized controlled trial. Healthcare (Basel) 2018 May 2;6(2):39. [doi: 10.3390/healthcare6020039] [Medline: 29724054]
43. Jiang W, Huang S, Ma S, et al. Effectiveness of companion-intensive multi-aspect weight management in Chinese adults with obesity: a 6-month multicenter randomized clinical trial. Nutr Metab (Lond) 2021 Feb 3;18(1):17. [doi: 10.1186/s12986-020-00511-6] [Medline: 33536048]

XSL•FO
RenderX

44. Godino JG, Merchant G, Norman GJ, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. Lancet Diabetes Endocrinol 2016 Sep;4(9):747-755. [doi: 10.1016/S2213-8587(16)30105-X] [Medline: 27426247]
45. Palacios C, Torres M, López D, Trak-Fellermeier MA, Coccia C, Pérez CM. Effectiveness of the nutritional app "MyNutriCart" on food choices related to purchase and dietary behavior: a pilot randomized controlled trial. Nutrients 2018 Dec 12;10(12):1967. [doi: 10.3390/nu10121967] [Medline: 30545125]
46. Rogers RJ, Lang W, Barone Gibbs B, et al. Applying a technology-based system for weight loss in adults with obesity. Obes Sci Pract 2016 Mar;2(1):3-12. [doi: 10.1002/osp4.18] [Medline: 27812375]
47. Spring B, Pellegrini CA, Pfammatter A, et al. Effects of an abbreviated obesity intervention supported by mobile technology: The ENGAGED randomized clinical trial. Obesity (Silver Spring) 2017 Jul;25(7):1191-1198. [doi: 10.1002/oby.21842] [Medline: 28494136]
48. Thomas JG, Raynor HA, Bond DS, et al. Weight loss in Weight Watchers Online with and without an activity tracking device compared to control: A randomized trial. Obesity (Silver Spring) 2017 Jun;25(6):1014-1021. [doi: 10.1002/oby.21846] [Medline: 28437597]
49. Vaz CL, Carnes N, Pousti B, Zhao H, Williams KJ. A randomized controlled trial of an innovative, user-friendly, interactive smartphone app-based lifestyle intervention for weight loss. Obes Sci Pract 2021 Oct;7(5):555-568. [doi: 10.1002/osp4.503] [Medline: 34631134]
50. Allen JK, Stephens J, Dennison Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013(151597):151597. [doi: 10.1155/2013/151597] [Medline: 24392223]
51. Ross KM, Wing RR. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: A randomized pilot study. Obesity (Silver Spring) 2016 Aug;24(8):1653-1659. [doi: 10.1002/oby.21536] [Medline: 27367614]
52. Hurkmans E, Matthys C, Bogaerts A, Scheys L, Devloo K, Seghers J. Face-to-face versus mobile versus blended weight loss program: randomized clinical trial. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e14. [doi: 10.2196/mhealth.7713] [Medline: 29326093]
53. Jospe MR, Roy M, Brown RC, et al. The effect of different types of monitoring strategies on weight loss: a randomized controlled trial. Obesity (Silver Spring) 2017 Sep;25(9):1490-1498. [doi: 10.1002/oby.21898] [Medline: 28703448]
54. Gemesi K, Winkler S, Schmidt-Tesch S, Schederecker F, Hauner H, Holzapfel C. Efficacy of an app-based multimodal lifestyle intervention on body weight in persons with obesity: results from a randomized controlled trial. Int J Obes (Lond) 2024 Jan;48(1):118-126. [doi: 10.1038/s41366-023-01415-0] [Medline: 38017117]
55. Schoeppe S, Alley S, Van Lippevelde W, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec;13(1):27927218. [doi: 10.1186/s12966-016-0454-y]
56. Elliot CA, Hamlin MJ. Combined diet and physical activity is better than diet or physical activity alone at improving health outcomes for patients in New Zealand's primary care intervention. BMC Public Health 2018 Feb 8;18(1):230. [doi: 10.1186/s12889-018-5152-z] [Medline: 29422040]
57. Avenell A, Brown TJ, McGee MA, et al. What interventions should we add to weight reducing diets in adults with obesity? A systematic review of randomized controlled trials of adding drug therapy, exercise, behaviour therapy or combinations of these interventions. J Human Nutrition Diet 2004 Aug;17(4):293-316. [doi: 10.1111/j.1365-277X.2004.00530.x]
58. Haase CL, Lopes S, Olsen AH, Satylganova A, Schnecke V, McEwan P. Weight loss and risk reduction of obesity-related outcomes in 0.5 million people: evidence from a UK primary care database. Int J Obes (Lond) 2021 Jun;45(6):1249-1258. [doi: 10.1038/s41366-021-00788-4] [Medline: 33658682]
59. Bouchard C. BMI, fat mass, abdominal adiposity and visceral fat: where is the "beef"? Int J Obes (Lond) 2007 Oct;31(10):1552-1553. [doi: 10.1038/sj.ijo.0803653] [Medline: 17549092]
60. Lofgren I, Herron K, Zern T, et al. Waist circumference is a better predictor than body mass index of coronary heart disease risk in overweight premenopausal women. J Nutr 2004 May;134(5):1071-1076. [doi: 10.1093/jn/134.5.1071] [Medline: 15113947]
61. Lee CMY, Huxley RR, Wildman RP, Woodward M. Indices of abdominal obesity are better discriminators of cardiovascular risk factors than BMI: a meta-analysis. J Clin Epidemiol 2008 Jul;61(7):646-653. [doi: 10.1016/j.jclinepi.2007.08.012] [Medline: 18359190]
62. Fox CS, Massaro JM, Hoffmann U, et al. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. Circulation 2007 Jul 3;116(1):39-48. [doi: 10.1161/CIRCULATIONAHA.106.675355] [Medline: 17576866]
63. Pischon T, Boeing H, Hoffmann K, et al. General and abdominal adiposity and risk of death in Europe. N Engl J Med 2008 Nov 13;359(20):2105-2120. [doi: 10.1056/NEJMoa0801891] [Medline: 19005195]
64. Zhang C, Rexrode KM, van Dam RM, Li TY, Hu FB. Abdominal obesity and the risk of all-cause, cardiovascular, and cancer mortality: sixteen years of follow-up in US women. Circulation 2008 Apr 1;117(13):1658-1667. [doi: 10.1161/CIRCULATIONAHA.107.739714] [Medline: 18362231]

XSL•FO
RenderX

65. van den Brandt PA, Goldbohm RA. Nutrition in the prevention of gastrointestinal cancer. Best Pract Res Clin Gastroenterol 2006;20(3):589-603. [doi: 10.1016/j.bpg.2006.04.001] [Medline: 16782531]
66. Ritchie SA, Connell JMC. The link between abdominal obesity, metabolic syndrome and cardiovascular disease. Nutr Metab Cardiovasc Dis 2007 May;17(4):319-326. [doi: 10.1016/j.numecd.2006.07.005] [Medline: 17110092]
67. Chew HSJ, Koh WL, Ng J, Tan KK. Sustainability of weight loss through smartphone apps: systematic review and meta-analysis on anthropometric, metabolic, and dietary outcomes. J Med Internet Res 2022 Sep 21;24(9):e40141. [doi: 10.2196/40141] [Medline: 36129739]
68. Zampino M, AlGhatrif M, Kuo PL, Simonsick EM, Ferrucci L. Longitudinal changes in resting metabolic rates with aging are accelerated by diseases. Nutrients 2020 Oct 7;12(10):3061. [doi: 10.3390/nu12103061] [Medline: 33036360]
69. Jacob JJ, Isaac R. Behavioral therapy for management of obesity. Indian J Endocrinol Metab 2012 Jan;16(1):28-32. [doi: 10.4103/2230-8210.91180] [Medline: 22276250]
70. Strasser B. Physical activity in obesity and metabolic syndrome. Ann N Y Acad Sci 2013 Apr;1281(1):141-159. [doi: 10.1111/j.1749-6632.2012.06785.x] [Medline: 23167451]
71. Magkos F, Fraterrigo G, Yoshino J, et al. Effects of moderate and subsequent progressive weight loss on metabolic function and adipose tissue biology in humans with obesity. Cell Metab 2016 Apr 12;23(4):591-601. [doi: 10.1016/j.cmet.2016.02.005] [Medline: 26916363]
72. Kompaniyets L, Freedman DS, Belay B, et al. Probability of 5% or greater weight loss or BMI reduction to healthy weight among adults with overweight or obesity. JAMA Netw Open 2023 Aug 1;6(8):e2327358. [doi: 10.1001/jamanetworkopen.2023.27358] [Medline: 37548978]
73. Ryan DH, Yockey SR. Weight loss and improvement in comorbidity: differences at 5%, 10%, 15%, and over. Curr Obes Rep 2017 Jun;6(2):187-194. [doi: 10.1007/s13679-017-0262-y] [Medline: 28455679]
74. Aucott L, Poobalan A, Smith WCS, Avenell A, Jung R, Broom J. Effects of weight loss in overweight/obese individuals and long-term hypertension outcomes: a systematic review. Hypertension 2005 Jun;45(6):1035-1041. [doi: 10.1161/01.HYP.0000165680.59733.d4] [Medline: 15897373]

Abbreviations

BCR: behavior change resource
BCRM: behavior change resource model
BCT: Behavior change techniques
CVD: cardiovascular disease
DBP: diastolic blood pressure
GDP: gross domestic product
GRADE framework: Grading of Recommendations, Assessment, Development, and Evaluation
HbA_{1c}: hemoglobin A_{1c}
MD: mean difference
MeSH: Medical Subject Headings
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PICOS : Participants, Interventions, Comparisons, Outcomes, and Study Design framework
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: Prospective Register of Systematic Reviews
RCT: randomized controlled trial
ROB 2: Cochrane risk-of-bias version 2
SBP: systolic blood pressure
WC: waist circumference

Edited by L Buis; submitted 17.06.24; peer-reviewed by A Batrakoulis, J Xia, N Noshadi; revised version received 22.05.25; accepted 30.05.25; published 19.08.25.

Please cite as:
Li S, Zhou Y, Tang Y, Ma H, Zhang Y, Wang A, Tang X, Pei R, Piao M
Behavior Change Resources Used in Mobile App–Based Interventions Addressing Weight, Behavioral, and Metabolic Outcomes in Adults With Overweight and Obesity: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2025;13:e63313
URL: https://mhealth.jmir.org/2025/1/e63313
doi:10.2196/63313

XSL•FO
RenderX

© Sijia Li, You Zhou, Ying Tang, Haoming Ma, Yuying Zhang, Aoqi Wang, Xingyi Tang, Runyuan Pei, Meihua Piao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Smartphone Apps on Reperfusion Times and Clinical Outcomes in Acute ST-Segment Elevation Myocardial Infarction: Systematic Review and Meta-Analysis

William Gibson, MBChB, MSc; Dawoud Al Kindi, MD; Elie Akl, MD; Kshitij Badal Dandona, MD; Jean-Philippe Pelletier, MD; Nicolo Piazza, MD, PhD; Ali Zgheib, MD; Giuseppe Martucci, MD; Marco Spaziano, MD

Department of Cardiology, Royal Victoria Hospital, McGill University Health Centre, 1001 Decarie Blvd, Montreal, QC, Canada

Corresponding Author:
William Gibson, MBChB, MSc
Department of Cardiology, Royal Victoria Hospital, McGill University Health Centre, 1001 Decarie Blvd, Montreal, QC, Canada

Abstract

Background: Smartphone- and tablet-based apps have been increasingly used in the management of acute ST-segment elevation myocardial infarction (STEMI), with the goal of enhancing care efficiency. These apps facilitate improved team coordination through a single platform, enabling secure sharing of clinical data, arrival times, and enabling data storage and processing capabilities. The potential of these technologies to reduce reperfusion times and improve both clinical and process outcomes, compared to traditional communication methods, is promising.

Objective: This study aimed to evaluate the effectiveness of smartphone-based interventions in reducing door-to-balloon (D2B) time, first medical contact-to-balloon (FMC2B) time, mortality, and false activation rates in STEMI care pathways.

Methods: This review followed the PRISMA guidelines and the PICO framework for eligibility criteria. Studies were included if they compared smartphone- or tablet-enabled interventions with usual care for STEMI management, focusing on D2B time, FMC2B time, short-term mortality, and false activation rates. A systematic literature search was conducted across MEDLINE, Embase, and Google Scholar for studies published between 2008 and 2024. Studies using purpose-built software or commercially available instant messaging apps that enabled digital ECG transfer and real-time communication between providers were included. The protocol was registered with PROSPERO (CRD42023481024). Data synthesis was performed using SPSS (IBM) with random-effects meta-analysis for continuous and binary outcomes.

Results: A total of 903 articles were identified after removing duplicates, and 21 studies, involving 3267 patients, were included. Studies varied in design: 14 being retrospective and 7 prospective, conducted across 12 countries. Thirteen studies evaluated dedicated apps, and 8 used instant messaging platforms such as WhatsApp (Meta Platforms Inc) and WeChat (Tencent Holdings Ltd). The primary outcome, D2B time, showed a significant reduction in the intervention group (mean difference –19.11 mins, 95% CI –26.22 to −12.00; P<.01), with substantial heterogeneity (I^2=89%). A similar reduction in FMC2B time was observed (mean difference −19.85 minutes, 95% CI −29.45 to −20.25; P=.01). Subgroup analysis indicated a more pronounced reduction in D2B time in low-income countries compared to high-income countries. There were no significant differences regarding short-term mortality (risk difference −0.03, 95% CI −0.07 to 0.01; P=.10). False activation rates were evaluated in 7 studies, with varying results, but no pooled analysis was feasible due to differences in definitions and study design. The health care setting (ie, low- or high-income countries) was the most significant factor contributing to the observed heterogeneity in the meta-regression analysis.

Conclusions: Smartphone-based interventions significantly reduce reperfusion times in STEMI care pathways. Digital technology can improve the efficiency of STEMI management, particularly with lower-resource health care infrastructure. Future studies should explore the impact on long-term outcomes and investigate regional differences in treatment effects.

(JMIR Mhealth Uhealth 2025;13:e66605) doi:10.2196/66605

KEYWORDS

door-to-balloon time; ST-elevation myocardial infarction; STEMI; cardiology; cardiac; cardiologists; Telemedicine; Telehealth; tele-medicine; tele-health; virtual care; virtual health; virtual medicine; mHealth; mobile health; mobile applications; mobile apps; apps; smartphones; applications; digital; digital health; digital technology; digital interventions

XSL•FO
RenderX

Introduction

ST-segment elevation myocardial infarction (STEMI) is a time-sensitive medical emergency in which prompt diagnosis and management is integral to minimizing total ischemic time and reducing morbidity and mortality [1]. Ischemic time is influenced by a number of elements, and delays may result from patient, emergency services, or systemic factors, which can further vary depending on the mode of patient presentation (ie, direct presentation to a PCI-capable ED, to a non-PCI center ED, or via emergency medical services [EMS].)" A number of time goals to achieving reperfusion are recommended by societal best practice guidelines internationally [2], and there is evidence to suggest an association between shorter door-to-balloon (D2B) times and improved survival outcomes. One pooled analysis including 300,000 patients from 32 prospective cohort studies demonstrated an association between longer D2B times (>90 mins) and increased short-term mortality [3]. Furthermore, shorter D2B times have been associated with lower hospital readmission rates and greater economic efficiency.

Achieving reperfusion within these targets requires streamlined patient transition, achieved by adequate coordination between the multidisciplinary systems involved in the chain of care and promoted by effective communication. Furthermore, coordinated efforts to collect, access, and interpret data efficiently are integral to improving the quality of these systems on an ongoing basis through continuous clinical audit. While patient delay is multifactorial and necessitates public health measures to enhance awareness of symptoms to reduce the time to the “first call for help,” system delays can be amenable to improvements by organizational means [4]. Advancements in telemedicine technologies have improved the efficiency of acute STEMI care in recent years, primarily through the transmission of prehospital electrocardiograms, leading to earlier activation of the catheterization laboratory (cath lab) and mobilization of the STEMI team. For example, prehospital electrocardiogram (ECG) transmission was associated with a 40% relative reduction in time to treatment in one meta-analysis, including 11 nonrandomized studies [5], while this effect was mirrored in a more recent analysis which further demonstrated an association between the implementation of a telemedicine strategy in acute STEMI and a reduction in mortality compared with usual care [6]. Notably, these reviews primarily included outdated telecommunication methods such as telephone and fax. However, a recent meta-analysis of nonrandomized studies, explored the impact of exclusively digital prehospital ECG transmission methods on reperfusion times and mortality and noted a significant reduction in door-to-device times (−33.3 mins) and in all-cause mortality (hazard ratio 0.53, 95% CI 0.40 - 0.69), highlighting the potential benefits of digital technology in this setting.

Nowadays, mobile devices are ubiquitous and an essential tool for health care practitioners in their day-to-day practice, with accompanying technological features including Wi-Fi, GPS, Bluetooth pairing, cameras, microphones, sensors, and cloud storage [7,8]. These capabilities enhance the potential use of mobile health (mHealth) apps to enhance care coordination in acute STEMI care. A number of apps have emerged in this domain, using tailor-built platforms as well as commercially available, nondedicated instant messaging apps (eg WhatsApp [Meta Platforms Inc] [9]. In comparison to instant messaging platforms, dedicated apps provide a number of added functions beyond a communication channel alone. First, they allow for notification of each individual STEMI team members’ acknowledgment of the incoming case which can foster more rapid and effective mobilisation of the team. They may further enable GPS and live time tracking of the patient relative to reperfusion targets, with integrated push alerts in situations where time to reperfusion may be too long to favor pPCI, and fibrinolytic therapy may be warranted instead. Moreover, they may provide a single interface for the entry of automated time stamped data which can be accessed easily, promoting improved efficiency of clinical auditing. Of further relevance, these technologies can be tailored to comply with data regulations ensuring confidentiality of storage and transmission of protected health information between health care providers [10-12]. At the time of writing, and to the best of our knowledge, there have been no reviews published exploring the effects of these technologies on outcomes in acute STEMI. This systematic review and meta-analysis aims to (1) evaluate the impact of smartphone-based STEMI coordination apps on reperfusion time targets and mortality in acute care settings, (2) identify the key features of available apps that may contribute to improved clinical efficiency and patient outcomes, and (3) compare the effects of dedicated versus nondedicated platforms, assessing their potential advantages and limitations in streamlining STEMI care. By synthesizing the current evidence, this review seeks to provide insights into the role of mHealth technologies in optimizing acute STEMI pathways and informing future advancements in digital health care solutions.

Methods

Selection and Eligibility Criteria

The review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines; the prespecified eligibility criteria and search strategy were structured using the PICO (population, people, patient, or problem), intervention, comparison and outcome template as displayed in Table 1. The study protocol was registered with PROSPERO (the International Prospective Register of Systematic Reviews; CRD42023481024.) Studies were included if they had a comparative study design and compared a strategy of smartphone or tablet-enabled technology as outlined in the eligibility criteria combined with usual care versus traditional methods of communication (eg, via telephone; review of STEMI ECG in person by STEMI team) for the diagnosis and management of patients with STEMI. The coprimary outcome was the difference in reperfusion times including door-to-balloon time and first medical contact-to-balloon (FMC2B) time; D2B time was defined as the time from arrival at the pPCI center to the deployment of a balloon within the infarct-related coronary artery, while FMC2B time refers to the time elapsed from first contact with a health care professional in the chain of care to the initial balloon inflation. Prespecified secondary outcomes of interest were the differences in short-term mortality and false activation rates

XSL•FO
RenderX

between groups, in addition to the proportion of patients meeting guideline directed time targets, that is, D2B time ≤90 minutes or FMC2B time ≤120 minutes.

Table . PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework for study eligibility.

Criteria		Description
Population		• Patients >18 years of age treated for suspected acute ST-segment elevation myocardial infarction in prehospital or hospital setting eligible to receive primary intervention (percutaneous coronary intervention or lysis where applicable).
Intervention		
	Inclusion criteria	• Smartphone or tablet-enabled mobile apps. • Used by health care provider. • Cloud-based ECG[a] transmission. • Real-time communication between relevant health care providers or stakeholders involved in the chain of care in acute STEMI[b].
	Exclusion criteria	• Patient-centered app usage. • Chronic care or postmyocardial infarction management. • Nonmobile or cellular-enabled telemedicine. • Posthoc communication (ie, not enabled in real time). • Focus on qualitative or technical process–related outcomes.
Comparator		• Patients treated for suspected acute STEMI using traditional communication methods (ie, nonmobile data sources or tablet app–based communication) between relevant stakeholders.
Outcomes		• Reperfusion times (door to balloon and first medical contact to balloon), proportion of patients meeting targets (eg, D2B[c] <90 min and FMC2B[d] <120 min), false activation rate, and clinical events (eg, mortality and heart failure).

[a]ECG: electrocardiogram.

[b]STEMI: ST-segment elevation myocardial infarction.

[c]D2B: door to balloon.

[d]FMC2B: first medical contact to balloon time.

The types of digital technology studied in this review include purpose-built software applications and commercially available instant messaging apps into STEMI care pathways, granted the latter fulfilled the prespecified eligibility checklist, namely, digital ECG transfer and real time communication between providers at the point of care. In cases wherein duplicate studies were identified, those with the largest sample size and follow up duration were solely included.

A total of 2 investigators independently screened potential titles and abstracts and those with potential eligibility by either reviewer were selected for full-text review. Disagreements were subsequently resolved by a third reviewer. Efforts were made to contact corresponding authors as required where relevant data was not included in the published report. The article was subsequently excluded if a satisfactory response or the relevant data was not obtained. Attempts were made to contact all authors in relation to missing data and study design clarification via email to which 4 replies were received.

Data Sources and Search Strategy

A systematic literature search was conducted across multiple databases, including MEDLINE, Embase and Google Scholar, to identify relevant studies published between January 1st, 2008, and February 1st, 2025. Textbox 1 details the structured search strategy used. The search was restricted to English-language studies, with exclusion of the term “rehab” to focus on studies evaluating acute care. To ensure comprehensive coverage, simultaneous searches of identified app names (eg, STENOA

XSL•FO
RenderX

[STENOA Inc] and SCUNA [Mehergen) were performed using title-word searches.

Textbox 1. Database search strategy.

Search statement 1: “mhealth”.tw OR “mobile health”.tw OR “app”.tw OR “smartphone”.tw OR “smart phone”.tw OR “mobile phone”.tw OR “tablet”.tw OR “mobile application”.tw OR “m-health”.tw OR “mobile device”.tw OR “mobile app”.tw OR “mobile apps”.tw OR “cloud*”.tw

AND

Search statement 2: “STEMI”.tw OR “ST elevation myocardial infarction”.tw OR “door to balloon”.tw. OR “door-to-balloon”.tw OR “myocardial infarction”.tw OR “acute coronary syndrome”.tw OR “ACS”.tw

NOT “rehab*”.tw

Data Extraction and Quality Assessment

A total of 2 investigators (WG and DAK) extracted data from each study according to a standardized protocol. Any disagreements were resolved by the third reviewer (MS). The methodological quality of each study included was assessed using the Newcastle-Ottawa Scale and 1 author (WG) assessed the quality of each study while a second author (DAK) reviewed these assessments in conjunction with the studies to support article inclusion. A third reviewer (MS) resolved any discordance.

Quantitative Data Synthesis

All analyses were performed using the meta-analysis function within SPSS (IBM Corp) comparing app-based care with usual care with respect to D2B times, FMC2B, and the proportion of patients meeting reperfusion targets (<90 mins and<120 mins) with continuous outcome measurements (unstandardized mean difference in mins) and odds ratio (OR and 95% CI, accordingly). The risk difference was used for the secondary outcome of mortality. When data concerning the mean and variance were not available, efforts to obtain these were made by contacting the authors. Failing this, the mean and variance were estimated by using the median, IQR, sample size and/or reported CIs [13]. In cases where the SDs for the mean in both control and intervention groups were not given, these were derived from the mean difference and corresponding 95% CI and/or the *P* value, where available [14]. In situations where more than 1 control group was encountered, the group which bore more similarity to the intervention group was selected. For example, if the intervention group concerned patients transferred from a non-PCI capable hospital, compared to those transferred by EMS or direct ED presentations with whom no app was used, the EMS transfers were chosen as the control group for analysis. Owing to expected clinical heterogeneity, a random-effects model was used for pooling the results of included studies. Mean differences were calculated for continuous outcomes; pooled relative risks and pooled risk difference for binary outcomes, and calculated 95% CIs and 2-sided *P* values for each outcome. Cochran Q test, chi-square tests, and the I^2 statistic were used to investigate statistical heterogeneity of the treatment effect among the included studies and values greater than 50% were considered to indicate high heterogeneity. A *P* value of<.05 (2-sided) was deemed statistically significant. We specified the following possible explanations for heterogeneity: study quality, specificities of the telemedicine intervention (dedicated or nondedicated app), and health care setting (ie, low- and middle-income countries vs high-income countries), owing to expected differences in health care system organization that might influence overall reperfusion times. Sensitivity analyses were predefined to evaluate the robustness of the primary outcome results. In sensitivity analyses, the effect size was examined by omitting studies individually and also excluding simultaneously studies with extreme results (highest and lowest impact). Subgroup analyses were performed to assess whether the treatment effect differed with regards to the type of technology used (dedicated vs nondedicated platform) and study quality (low vs medium to high risk of bias). Publication bias was investigated by constructing a funnel plot and plot symmetry was determined both visually and formally using Egger test.

Ethical Considerations

This study was conducted using only aggregated, study-level data extracted from previously published sources. No individual patient-level data were accessed or analyzed, and no identifiable personal health information was used. As such, ethics approval and informed consent were not required.

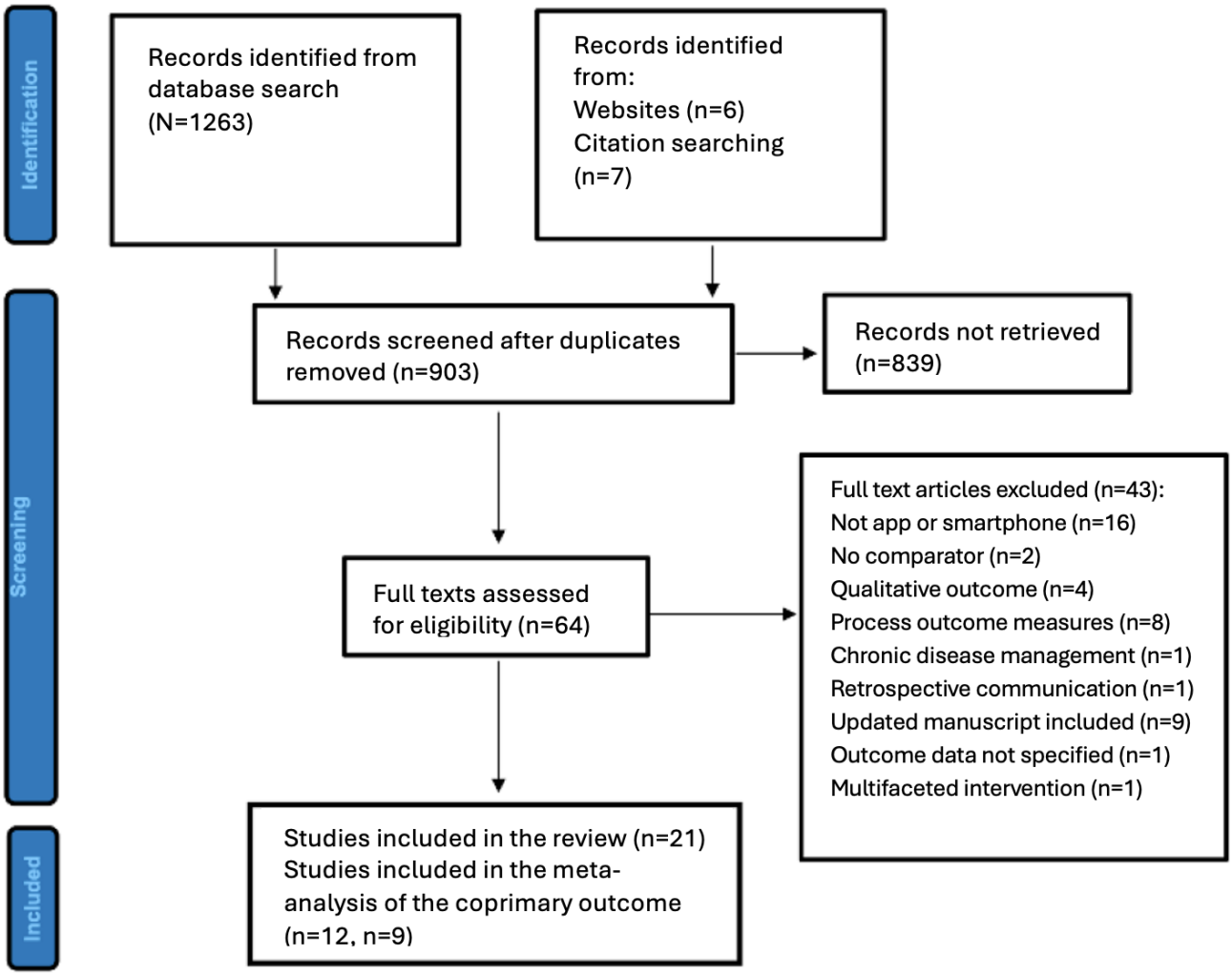
Results

Overview

The search yielded 1263 results and 360 duplicates were removed. Of the remaining 903 studies, 890 were derived from databases (MEDLINE and Embase), while citations of evaluated studies and web searching (Google Scholar) identified apps provided a further 13 records for screening. Figure 1. displays the PRISMA flowchart depicting the final article inclusion. Following title and abstract screening, 839 records were removed, and the remaining 64 studies underwent full text assessment for inclusion based on the eligibility criteria. Exclusion of 43 studies was carried out with the reasons outlined (Figure 1). Thus, the final number of studies included in the review was 21, involving 3267 patients.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of articles identified during the search process.



Characteristics of Included Studies

The main characteristics of each included study are shown in Table 2. The selected studies were published between 2013 and 2024. A total of 18 studies were published in peer-reviewed journals and 3 were published conference abstracts [15-17]. All studies were nonrandomized in design: 14 were retrospective, involving a historical comparator group of patients treated before the implementation of a telemedicine intervention; the remaining 7 studies were prospective, quasi-experimental, and included a concurrent parallel control group in which the technology was not used [9,18-23]. The total number of participants in each study ranged from 46 to 428. The studies were conducted across 10 countries: Japan (n=4), United States (n=3), Taiwan (n=3), Canada (n=2), Australia (n=2), Argentina (n=1), Slovakia (n=1), Turkey (n=1), Egypt (n=1), and China (n=1). About 13 studies evaluated purpose-built apps, including 18 distinct platforms, while instant messaging apps (WhatsApp [n=3], LINE [n=2], Band [n=1], and WeChat (Tencent Holdings Ltd; n=2) were implemented in 8 studies [9,19,21,22,24-27]. The platforms described were commercially available from mainstream app stores. In all cases, the software was used to transmit ECGs to a dedicated chat group involving STEMI care providers, the referring practitioner, and in certain cases, the emergency physicians. The app was used in patients transferred by EMSs from the field in 14 studies, for interhospital transfer from non-PCI capable centers in 8 studies, and for patients presenting directly to the emergency department in 5 studies. In 9 studies, a combination of the type of presentation to the pPCI center was addressed. All but one [24] of the studies involving nondedicated platforms evaluated interhospital transfers of patients from non-PCI capable centers to a pPCI center.

XSL•FO
RenderX

Table . Characteristics and primary outcomes of included studies.

Study	Year	Country	Participants		Presentation			App users			Out-comes	
			(I/C)[a]	App	EMS[b]	IHT[c]	ED[d]	EMS	ED	pPCI[e]	D2B[f] (mins)	FMC2B[g] (mins)
Aboal et al [18]	2024	Spain	98/129	ODISEA (Biomedical Research Institute of Girona)	✓	✓	—	✓	—	✓	—	102 (SD 36)[h] vs 117 (SD 45)[h]
Abrahim et al [28]	2023	USA	132/296	e-Bridge (General Devices LLC)	✓	—	✓	✓	✓	✓	61.2 (SD 24.3) vs 68.4 (SD 24.6)[h]	94.9 (SD 26.5) vs 100.2 (SD 28.0)[b]
Aldajani et al [29]	2024	Canada	31/84	STENOA (Stenoa Inc)	✓	—	✓	—	✓	✓	—	85.3 (SD 25.5) vs 95.5 (SD 52.2)[b]
Anai et al [15].	2024	Japan	113/130	SCUNA (Mehergen Group Holdings, Inc)	✓	—	—	✓	—	✓	60 (SD 25.1) vs 75 (SD 20)[h]	—
Arinaga et. al. [30]	2022	Japan	23/25	SCUNA	✓	—	—	✓	—	✓	49 (41 - 85.5) vs 59 (38-67)[i]	72 (60.5 - 107) vs 80 (63-92)[i]
Astarcioglu et al [9]	2015	Turkey	53/55	WhatsApp (Meta Platforms Inc)	—	✓	—	—	✓	✓	—	109 (SD 31) vs 130 (SD 46)[h]
Bendary et al [19]	2019	Egypt	100/100	WhatsApp	✓	✓	—	—	✓	✓	120 (SD 35) vs 157 (SD 30)[h]	—
Bladin et al [20]	2022	Australia	171/76	Pulsara (Pulsara, Inc)	✓	—	✓	✓	✓	✓	59.37 (SD 41.33) vs 77.3 (SD 61.6)[i]	—
Brouillette et al [16]	2023	Canada	31/84	STENOA	✓	—	✓	—	✓	✓	—	—
Chao et al [24]	2017	Taiwan	44/40	LINE (LY Corporation)	—	—	✓	—	✓	✓	90.4 (SD 9.8) vs 119.3 (SD 16.3)[i]	—
Cordero et al [17]	2013	Spain	177	MBEAT (ImaxDI)	✓	—	—	✓	—	✓	—	128 vs 187[h] (p<0.01)

XSL•FO
RenderX

Study	Year	Country	Participants		Presentation			App users			Out-comes	
Dickson et al [31]	2023	USA	69/36	Pulsara (Pulsara, Inc)	✓	—	✓	✓	✓	✓	71 (SD 49) vs 91 (SD 49)[h]	—
Ibanez et al [21]	2019	Argentina	43/62	WhatsApp	✓	✓	—	✓	✓	✓	—	132 (80 - 150) vs 166 (135-210)[h]
Kang et al [25]	2023	China	100/98	WeChat (Tencent Holdings Ltd)	—	✓	—	—	✓	✓	65.9 (SD 13.86) vs 98.21 (SD 21.02)[h]	98.6 (SD 18.7) vs 128.28 (SD 20.22)[h]
Kini et al [32]	2024	USA	17/43	STEMI-CathAid (Icahn School of Medicine at Mount Sinai)	—	✓	—	✓	✓	✓	—	86 (SD 16) vs 120 (SD 48)[h]
Kohashi et al [33]	2023	Japan	77/160	SCUNA	✓	—	—	✓	—	✓	67.7 (SD 23.6) vs 78.0 (SD 29.3)[h]	—
Liu et al [26]	2020	Taiwan	70/70	WeChat (Tencent Holdings Ltd)	—	✓	—	✓	✓	✓	60 vs 95.5[h] (p<0.001)	132 vs 171[h] (p<0.001)
Park et al [22]	2016	Korea	50/64	BAND (Naver Corporation/Camp Mobile)	—	✓	—	—	✓	✓	47.5 (40-56) vs 56.5 (47-69.5)[h]	102.5 (89-139) vs 129.5 (98-159)[h]
Studencan et al [34]	2018	Slovakia	178/67	STEMI (STEMI Global)	✓	—	✓	—	✓	—	—	
Yu et al [27]	2019	Taiwan	51/89	LINE (LY Corporation)	—	✓	—	—	✓	✓	52.6 (SD 42.2) vs 78.4 (SD 50.6)[h]	—
Yufu et al [23]	2019	Japan	17/29	SCUNA	✓	—	—	✓	—	✓	70 (SD 26) vs 96 (SD 24)[h]	—

[a]I/C: intervention/control.

[b]EMS: emergency medical service.

[c]IHT: interhospital transfer.

[d]ED: emergency department.

[e]pPCI: primary percutaneous coronary intervention.

[f]D2B: door to balloon.

[g]FMC2B: first medical contact to balloon.

[h]*P*<.05 (significant).

[i]*P*≥.05 (not significant); values in parentheses: median and IQR; values with (SD): mean (SD).

XSL•FO
RenderX

Assessment of Study Quality

Table 3 outlines the quality assessment as graded by Newcastle-Ottawa Scale. A total of 11 studies were deemed to be at low risk of bias, while 10 were considered medium to high risk of same with a median score of 6.5 (IQR 5 - 7) being observed.

XSL•FO
RenderX

Table . Newcastle-Ottawa assessment scale of study quality.

Author	Year	Selection				Compara-bility	Outcomes			Points	RoB
		1	2	3	4	1	1	2	3		
Aboal et al [18]	2024	*	—	*	*	*	*	*	—	6	Low
Abrahim et al [28]	2024	*	*	*	*	—	*	*	—	6	M-H
Aldajani et al [29]	2024	*	*	*	*	*	*	*	—	7	Low
Anai et al [15]	2024	*	—	*	*	—	*	—	—	4	M-H
Arinaga et. al. [30]	2023	*	—	*	*	**	—	*	—	7	Low
Astar-cioglu et al [9]	2015	*	*	*	*	*	—	—	—	5	M-H
Bendary et al [19]	2019	*	—	*	*	**	*	*	—	7	Low
Bladin et al [20]	2023	*	—	—	*	*	*	*	—	5	M-H
Brouil-lette et al [16]	2022	*	*	*	*	*	*	*	—	7	Low
Chao et al [24]	2017	*	*	*	*	*	*	*	—	7	Low
Cordero et al [17]	2014	-	*	*	*	—	—	—	—	3	M-H
Dickson et al [31]	2014	*	*	*	*	—	**	*	—	7	M-H
Ibanez et al [21]	2019	*	*	*	*	**	—	*	—	7	Low
Kang et al [25]	2024	-	*	*	*	*	—	*	—	5	M-H
Kini et al [32]	2024	*	—	*	*	—	*	*	—	5	M-H
Kohashi et al [33]	2023	*	*	*	*	**	*	*	*	9	Low
Liu et al [26]	2020	*	*	*	*	*	*	*	—	7	Low
Park et al [22]	2016	*	—	*	*	—	—	—	—	3	M-H
Studencan et al [34]	2018	*	—	*	*	*	*	*	—	6	Low
Yu et al [27]	2019	*	—	*	—	*	—	—	*	4	M-H
Yufu et al [23]	2019	*	*	*	*	**	—	*	*	8	Low

[a]RoB: risk of bias.

[b]M-H: medium to high.

XSL•FO
RenderX

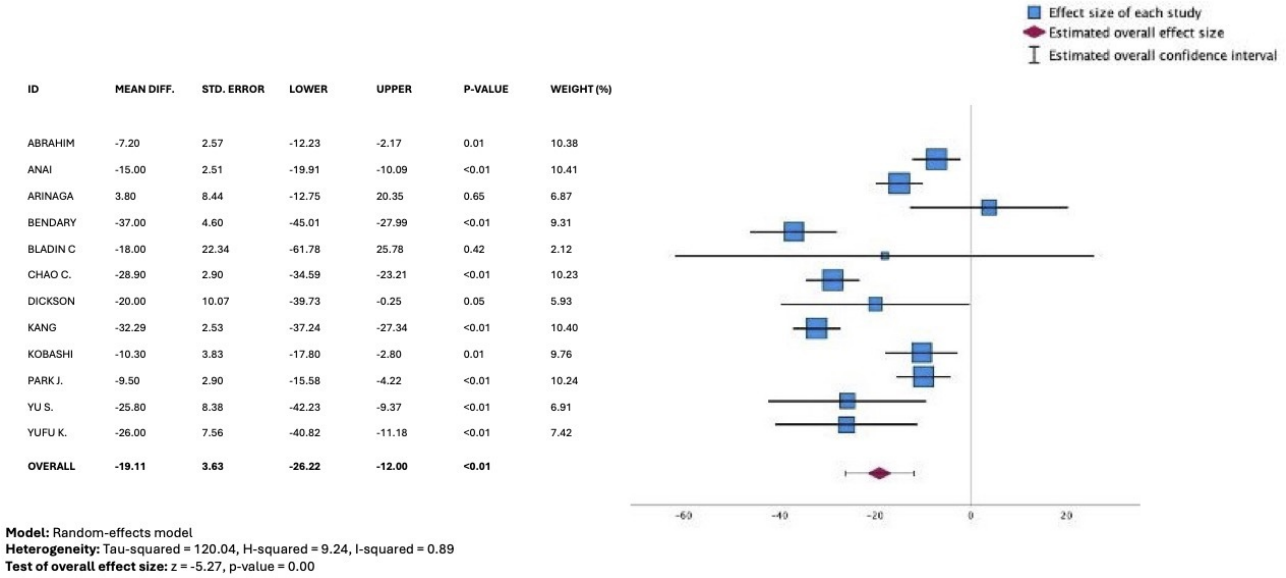
Outcomes

Primary Outcome

With respect to the primary outcome, all but one of the included studies evaluated the association of a digital telemedicine strategy with reperfusion times, including door to balloon, FMC2B, or total ischemic time. The remaining study investigated "false activations" [16] exclusively within a similar patient cohort from another included study [29]. Time target definitions varied across studies and were standardized according to our prespecified definitions as follows: door-to-device time was used in one study in reference to "the interval from the initial presentation at the referring hospital until the first balloon inflation or use of another device to achieve reperfusion" [32] and thus, we labeled this as the FMC2B time. Similarly, "first medical contact to wire" [25] or "device" [22,29] was considered the FMC2B, as was "ECG diagnosis to wire crossing." Regarding meta-analysis, complete data including both the mean and SD for the treatment and control groups was documented in 13 studies. The individual SDs were not reported in 2 studies [15,31] and were therefore estimated from the mean difference and provided *P* value using validated methods [14]. Furthermore, the median value alone were reported in 5 studies; one author provided the necessary data upon request [33] and mean and SD values were estimated, accordingly for the remaining 4 [20-22,30]. A further study was excluded from quantitative data synthesis [17] as it did not report the number of participants in both the treatment and control groups.

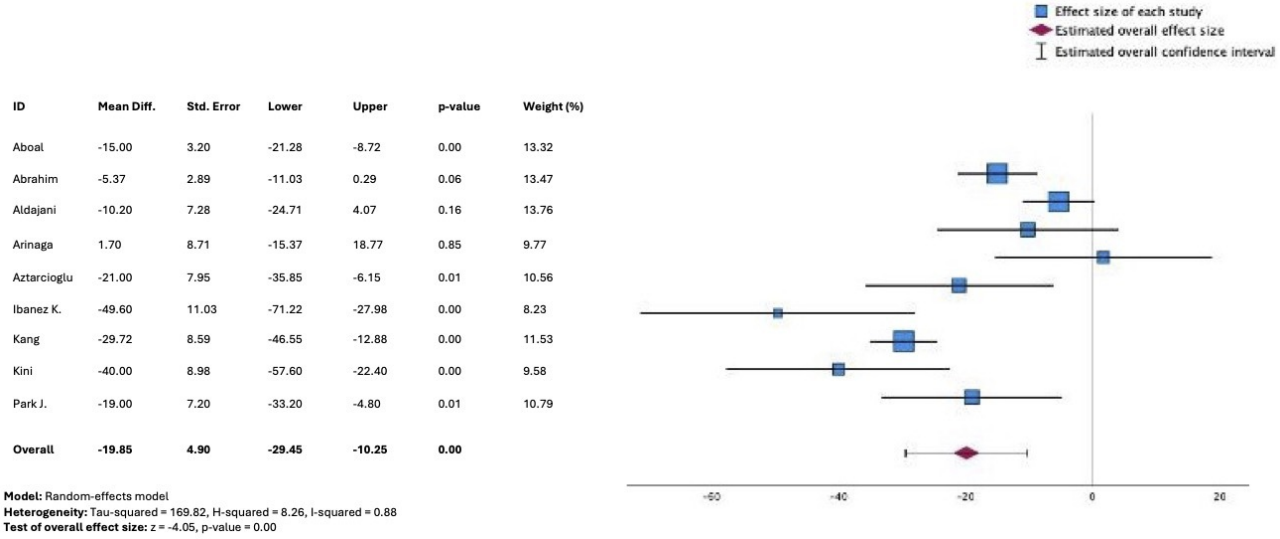
Differences in D2B times versus usual care were reported in 13 studies with sufficient data available from 12 studies to conduct meta-analysis, involving 1977 patients. There was a significant difference in D2B times observed, favoring a smartphone app–based intervention versus usual care (unstandardized mean difference, −19.11 mins, 95% CI −26.22 to −12.00); *P*<.01) (Figure 2 [15,19,20,22-25,27,28,30,31,33]), while substantial statistical heterogeneity was evident (I^2=89%). FMC2B times were reported in 11 studies overall, with 9 included in the meta-analysis (n=1888). A similar treatment effect was evident, favoring the intervention group (unstandardized mean difference, −19.85 mins, 95% CI −29.45 to –20.25; *P*=.01; Figure 3 [9,11,21,22,25,28-30,32]), with similar degrees of statistical heterogeneity found (I^2=88%).

Figure 2. Forest plot from meta-analysis of unstandardized mean difference in door-to-balloon times (minutes) comparing telemedicine strategies versus usual care [15,19,20,22-25,27,28,30,31,33].



XSL•FO
RenderX

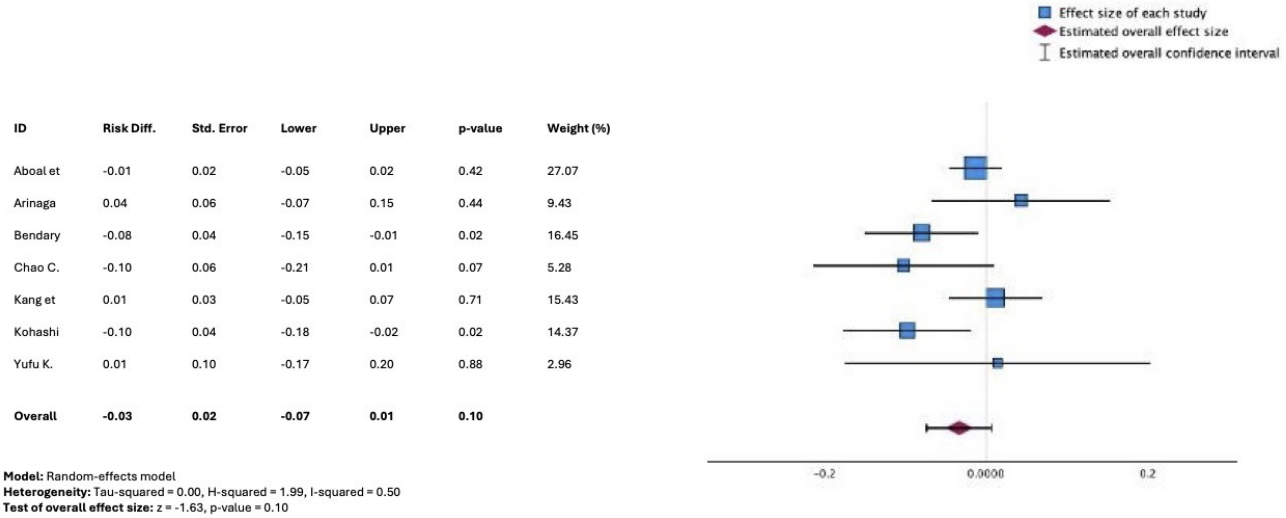
Figure 3. Forest plot from meta-analysis of unstandardized mean difference in first medical contact-to-balloon times (minutes) comparing telemedicine strategies versus usual care [9,11,21,22,25,28-30,32].



Secondary Outcomes

Mortality was evaluated in 8 studies, including 1711 patients, half of these reported in-hospital mortality [18,19,23,25], 30-day mortality in 3 [24,30,33], and one-year mortality in the remaining study [15]. For meta-analysis, the study involving one-year mortality was excluded, resulting in a total of 7 publications included in the pooled analysis. There was no significant effect observed with respect to short-term mortality associated with a smartphone-based strategy versus usual care (Risk difference −0.03, 95% CI −0.07 to 0.01); *P*=.10; Figure 4 [11,19,23-25,30,33]).

Figure 4. Forest plot from meta-analysis of risk difference in in- hospital or 30-day mortality comparing telemedicine strategies versus usual care [11,19,23-25,30,33].



Furthermore, 7 studies evaluated the impact of an app-based coordination pathways on false or inappropriate activations of the cardiac cath lab. Due to substantial variability among studies in defining false activations and comparator groups, however, meta-analysis was not feasible. Astarcioglu et al reported false positive STEMI activations in 5 out of 55 cases in the intervention group with WhatsApp versus none in the control group [9]. Brouillette et al [16] defined false activations as cancellations of cath lab activations prior to any procedure and demonstrated a reduction in false activations with use of the STENOA platform, decreasing from 9.4% to 6.4%, a 31% relative reduction (*P*<.05). Furthermore, the use of STEMICATHAID (Icahn School of Medicine at Mount Sinai) facilitated preactivation screening, rejecting 45 out of 111 activations prior to cath lab mobilization, however, no false activation data was available for comparison in the control period [32]. Dickson et al [31] observed improved resource use with a reduction in PCI procedures following implementation of the Pulsara (Pulsara, Inc) app (from 75% to 64%), representing an 11% absolute reduction. Meanwhile, there were fewer misdiagnoses of acute coronary syndrome in the group using the ODISEA (Biomedical Research Institute of Girona) app compared to controls (9.5% vs 17.1%, *P*=.004) [18]. In addition, another included study noted fewer inappropriate activations in the arm assigned to the SCUNA mobile ECG transmission system (13/36 vs 20/45) [30].

XSL•FO
RenderX

Finally, the proportion of patients meeting D2B or FMC2B times of <90 and <120 minutes were reported in 6 [9,20,23,24,31,34] and 4 studies [18,22,32,34], respectively. A pooled analysis of these outcomes combined demonstrated a higher likelihood of patients treated in the intervention arm of meeting guideline directed reperfusion times compared to those assigned to usual care (OR 3.18, 95% CI 1.59 - 6.35; P<.001; I^2=0.77; Figure S1 in Multimedia Appendix 1).

Publication Bias

Publication bias was assessed using funnel plots for the primary outcomes of D2B and FMC2B times. Visual inspection of the funnel plots revealed asymmetry for D2B time, suggesting potential publication bias, whereas minimal asymmetry was observed for FMC2B time. This was further supported by Egger test, which indicated significant publication bias for D2B time (P=.01) but not for FMC2B time (P=.5; Figures S2 and S3 in Multimedia Appendix 2 and Multimedia Appendix 3).

Sensitivity and Subgroup Analysis

Sensitivity analysis using the leave-one-out method, where individual studies were sequentially excluded, demonstrated no significant impact on the overall treatment effect for D2B or FMC2B time, confirming the robustness of the findings. Subgroup analysis based on geographical setting (low- and middle-income countries vs high-income countries) revealed a greater reduction in D2B time in studies conducted in low-income settings (mean difference −31.48 mins, 95% CI −24.86 to −28.11) compared to those in high-income settings (mean difference −11.26 mins, 95% CI −15.18 to −7.34; Figure S4 in Multimedia Appendix 4). However, this trend was not as pronounced for FMC2B time (low-income −31.02 mins, 95% CI −42.94 to −19.11 vs high-income −13.86 mins, 95% CI −23.64 to −4.07). Further subgroup analyses stratified according to the use of dedicated versus nondedicated mobile apps, risk of bias (high vs low), and study sample size (above vs below the median) showed no statistically significant differences between subgroups.

To explore potential sources of heterogeneity, meta-regression analysis was performed. For D2B time, geographical setting accounted for nearly all the observed heterogeneity (R^2=100%), suggesting a strong influence of regional health care disparities. (Figure S2 in Multimedia Appendix 3). In contrast, for FMC2B time, geographical setting explained only a moderate proportion of heterogeneity (R^2=31.4%). Other factors, including sample size, use of dedicated versus nondedicated apps, and study quality, had minimal influence on heterogeneity for both D2B and FMC2B times.

Discussion

Principal Findings

In this comprehensive systematic review and meta-analysis, we synthesized evidence from 21 studies, including 3267 patients, which evaluated smartphone apps implemented within acute STEMI care pathways. The included studies assessed the impact of these digital platforms on key reperfusion metrics and clinical outcomes. Our findings suggest that an average reduction in reperfusion times of up to 20 minutes might be achieved with the integration of smartphone apps in STEMI care pathways, with the potential for a larger proportion of patients achieving guideline directed target times (D2B <90 mins and FMC2B <120 mins) versus usual care. mHealth technologies have emerged as transformative tools in acute care settings, with up to 40% of emergency-focused mHealth solutions prioritizing rapid communication and coordination functions [8]. Immediate and accurate sharing of patient clinical and treatment information with the receiving hospital is essential in acute STEMI and a single communication channel via smartphone technology used both within the hospital and by prehospital personnel strengthens the integration of treatment advances that cover the community and medical care settings alike [20].

To our knowledge, this is the first meta-analysis specifically examining the effectiveness of dedicated smartphone-based apps within STEMI care networks. Previous systematic reviews and meta-analyses have demonstrated improvements in reperfusion times and mortality with telemedicine-based strategies, as well as digital transmission of the prehospital electrocardiogram in acute STEMI care [6,35]. However, none have exclusively addressed interventions incorporating the broader functionalities provided by smartphone apps, such as coordinated alerts, real-time geolocation, automated data collection, and integrated communication systems among multidisciplinary STEMI teams, as we have described. Since telemedicine interventions are rarely effective in isolation and instead function best as part of a multifaceted strategy integrating prehospital and in-hospital services, it stands to reason that smartphone apps offering these multiple functionalities could further streamline workflows and enhance the efficiency of STEMI care processes.

Our review further identified multiple studies reporting a significant reduction in inappropriate cath lab activations following the implementation of an app-based strategy. Inappropriate activation of the cath lab represents a substantial burden to health care systems with disruption of clinical workflows, prompting unnecessary mobilization of on-call teams, while potentially causing unnecessary harm to patients [36]. The capacity to integrate real-time ECG assessment with structured case adjudication, supported by comprehensive clinical data, may enhance diagnostic accuracy and communication among care teams, thereby reducing rates of false-positive activations.

Notably, we found no significant difference in short-term mortality between the intervention and control arms which is in contrast with the findings of a recent meta-analysis of studies evaluating the impact of digital prehospital ECG transmission in acute STEMI [35]. Therein, the authors observed a significant reduction in all-cause mortality (hazard ratio 0.53, 0.40 - 0.69), which accompanied a greater reduction in mean door and FMC2B times than we found (mean difference −33.31 mins, 95% CI −50.47 to −16.16) However, the substantial weighting of included studies (>50%) which reported long-term mortality [37], in addition to the inclusion of larger studies with higher event rates, somewhat limits comparability to our findings. Furthermore, differences in baseline and procedural characteristics or study design may have contributed to the

XSL•FO
RenderX

differences observed. They further noted substantial statistical heterogeneity (>99%) with respect to reperfusion times while demonstrating a significantly higher impact on subgroup analysis between interventions that integrated both prehospital digital ECG transmission and early cath lab activation versus ECG alone. This may be reflective of the earlier study period of their included studies, during which time STEMI care networks were comparatively rudimentary to those evaluated in our analysis, with a greater potential to yield marginal benefits from organizational improvements. Likewise, our meta-regression analysis revealed that geographical location (ie, low- vs high-income countries) was the strongest factor contributing to the observed heterogeneity in terms of D2B times. Similarly, health care systems across low-income countries may have fewer care coordination processes in place and may therefore expect to achieve a greater benefit from quality improvement interventions. Since all studies from low- and middle-income countries used nondedicated apps, it remains unclear whether the additional functionalities of dedicated software could lead to further improvements, and future research is needed to evaluate their use in these settings.

On the other hand, the influence of geographical location was less pronounced in terms of FMC2B times. A possible explanation for this disparity is that system-level factors—such as prehospital care infrastructure and EMSs efficiency—may have a greater impact on FMC2B time, whereas in-hospital workflow efficiency primarily affects D2B time. In low-income settings, delays in prehospital activation and interfacility transport [38] may have contributed to the persistence of FMC2B disparities despite significant reductions in D2B time. However, most included studies did not report data on average transport times or distances to pPCI centers, preventing meta-regression analysis from assessing whether these factors contributed to the observed heterogeneity.

Nearly half of the included studies were assessed as having moderate to high risk of bias, primarily attributed to insufficient reporting or adjustment for baseline characteristic disparities. However, subgroup analyses comparing studies with a high and low risk of bias revealed no statistically significant differences in treatment effects across outcomes, suggesting minimal influence of bias on pooled results.

We also identified a number of important distinctions between dedicated and nondedicated apps in the context of acute STEMI care (Table 4). Concerning the point of care, the functions of instant messaging apps are largely confined to the facilitation of rapid ECG transmission and clinical particulars via short messaging. On the other hand, dedicated apps allow for the seamless input of various data points to designated fields while enabling custom notifications and alert systems, which can confirm the acknowledgment of each team member's participation and readiness, automatically signal the presence of high-risk clinical features and with GPS integration, can prompt consideration of alternative treatment strategies if expected patient arrival times exceed guideline-directed targets. Moreover, the majority of dedicated apps identified boast functions beyond the point of care with automated storage of quality assurance data which can streamline clinical auditing and ultimately lessen the significant administrative burden associated with such processes. Furthermore, dedicated platforms can be structured appropriately to ensure that the handling of protected health information data is strictly compliant with local data security regulations. This is especially important for patients, as research indicates they place high value on robust data privacy and security measures in mHealth apps. Therefore, dedicated platforms that adhere to regulatory frameworks are essential to ensure compliance, build trust, and overcome adoption barriers [39]. In contrast, the use of nondedicated apps in this context raises concerns about their appropriateness.

Table . Key functions enabled by different mobile app types.

Features	Nondedicated	Dedicated
Cloud-based ECG[a] transmission	✓	✓
Real time communication	✓	✓
Automatic notification of provider readiness		✓
Real time treatment target tracking[b]		✓
Automated data recording and reporting		✓
GPS location enabled		✓
Security and compliance		✓

[a]ECG: electrocardiogram.

[b]Automatic notification when estimated first medical contact-to-balloon time exceeds target with recommendation to alter treatment course ie, fibrinolytic therapy.

Limitations

All of the included studies were observational in design and mostly comprised of relatively small sample sizes. Therefore, many studies were underpowered to establish definitive conclusions regarding the impact of smartphone-based interventions on critical clinical outcomes, such as mortality. Nevertheless, the clinical implications of our findings remain important, given existing evidence linking longer D2B times to increased short-term mortality. A pooled analysis involving over 70,000 patients, for example, demonstrated significantly higher short-term mortality among patients with D2B times exceeding 90 minutes (OR 1.39, 95% CI 1.19 - 1.62) [40]. Similarly, an individual patient data analysis comprising more

XSL•FO
RenderX

than 4000 patients found consistent results, even after adjusting for multiple potential confounders [3]. A further limitation of this study is the reliance on statistical imputation for a number of outcome variables, as only the median and IQR values were provided in the original reports. This approach may have introduced bias and limited the precision of our analysis, owing to uncertainties regarding the underlying distribution of the data [41]. In addition, the lack of distributional data in most studies in terms of different patient presentation modes (ie, direct ED, EMS, or interhospital transfer) poses a limitation. The impact of the app on outcomes across these modes is not yet clear, and further investigation is needed to determine whether the use of smartphone apps has a differential effect based on the mode of presentation.

Conclusion

Smartphone apps are playing an increasingly transformative role in emergency care, offering the potential to enhance processes and, ultimately, improve clinical outcomes. By consolidating critical functionalities into a single platform, these apps address the systemic fragmentation between pre-hospital and in-hospital care, facilitating more coordinated workflows. Our analysis highlights a reduction in reperfusion times with the integration of these systems; however, larger prospective studies are needed to further explore and validate these findings.

Data Availability

The data supporting the findings of this systematic review are available within the article and its supplementary materials. All relevant data are included in the manuscript and the supplementary files. In addition, the datasets generated and/or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Achieving door to balloon time of <90 minutes or first medical contact to balloon time of <120 minutes in patients submitted to intervention versus usual care.
[PNG File, 99 KB - mhealth_v13i1e66605_app1.png]

Multimedia Appendix 2
Funnel plot assessing publication bias for the primary endpoint of door to balloon times.
[PNG File, 27 KB - mhealth_v13i1e66605_app2.png]

Multimedia Appendix 3
Funnel plot assessing publication bias for publication bias with respect to the outcome of first medical contact to balloon times.
[PNG File, 27 KB - mhealth_v13i1e66605_app3.png]

Multimedia Appendix 4
Subgroup analysis of primary endpoint door to balloon times stratified according to health care setting (1: low- or middle-income countries versus 2: high-income countries).
[PNG File, 43 KB - mhealth_v13i1e66605_app4.png]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.
[PDF File, 77 KB - mhealth_v13i1e66605_app5.pdf]

References

1. O'Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013 Jan 29;61(4):e78-e140. [doi: 10.1016/j.jacc.2012.11.019] [Medline: 23256914]
2. Byrne RA, Rossello X, Coughlan JJ, et al. 2023 ESC Guidelines for the management of acute coronary syndromes. G Ital Cardiol (Rome) 2024 Feb;25(2 Suppl 2):e1-e112. [doi: 10.1714/4191.41785] [Medline: 38291910]
3. Foo CY, Bonsu KO, Nallamothu BK, et al. Coronary intervention door-to-balloon time and outcomes in ST-elevation myocardial infarction: a meta-analysis. Heart 2018 Aug;104(16):1362-1369. [doi: 10.1136/heartjnl-2017-312517] [Medline: 29437704]
4. Bradley EH, Herrin J, Wang Y, et al. Strategies for reducing the door-to-balloon time in acute myocardial infarction. N Engl J Med 2006 Nov 30;355(22):2308-2320. [doi: 10.1056/NEJMsa063117] [Medline: 17101617]

XSL•FO
RenderX

5. Brunetti ND, De Gennaro L, Correale M, et al. Pre-hospital electrocardiogram triage with telemedicine near halves time to treatment in STEMI: a meta-analysis and meta-regression analysis of non-randomized studies. Int J Cardiol 2017 Apr;232:5-11. [doi: 10.1016/j.ijcard.2017.01.055]
6. Marcolino MS, Maia LM, Oliveira JAQ, et al. Impact of telemedicine interventions on mortality in patients with acute myocardial infarction: a systematic review and meta-analysis. Heart 2019 Oct;105(19):1479-1486. [doi: 10.1136/heartjnl-2018-314539] [Medline: 31253696]
7. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 5;43:117-134. [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
8. Plaza Roncero A, Marques G, Sainz-De-Abajo B, et al. Mobile Health Apps for Medical Emergencies: Systematic Review. JMIR Mhealth Uhealth 2020 Dec 11;8(12):e18513. [doi: 10.2196/18513] [Medline: 33306037]
9. Astarcioglu MA, Sen T, Kilit C, et al. Time-to-reperfusion in STEMI undergoing interhospital transfer using smartphone and WhatsApp messenger. Am J Emerg Med 2015 Oct;33(10):1382-1384. [doi: 10.1016/j.ajem.2015.07.029] [Medline: 26299691]
10. Krishnamoorthy P, Vengrenyuk A, Wasielewski B, et al. Mobile application to optimize care for ST-segment elevation myocardial infarction patients in a large healthcare system, STEMIcathAID: rationale and design. Eur Heart J Digit Health 2021 Jun;2(2):189-201. [doi: 10.1093/ehjdh/ztab010] [Medline: 36712391]
11. Aboal J, Ramos R, Boada I, et al. Use of new technologies in the transfer of patients with STEMI. Results from the pilot test ODISEA APP. RECICE 2023 Feb 7;5(1):66-75. [doi: 10.24875/RECICE.M22000334]
12. Mir H, Cullen KJ, Mosleh K, et al. Smartphone app for prehospital ECG transmission in ST-elevation myocardial infarction activation: protocol for a mixed methods study. JMIR Res Protoc 2024 Sep 6;13:e55506. [doi: 10.2196/55506] [Medline: 39240681]
13. Wan X, Wang W, Liu J, et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135. [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
14. Chandler J, Cumpston M, Li T, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 64: Cochrane; 2023.
15. Anai M, Unoki T, Inamori T, et al. Impact of a prehospital mobile cloud ECG transmission system on door-to-balloon time, myocardial damage and mid-term all-cause mortality in patients with ST-elevation myocardial infarction. Eur Heart J 2024 Oct 28;45(Supplement_1). [doi: 10.1093/eurheartj/ehae666.1727]
16. Brouillette F, Aldajani A, Chaabo O, et al. Impact of a care coordination platform on false activations of cardiac catheterization laboratory: a retrospective cohort study. J Am Coll Cardiol 2023;82(17_Supplement) [FREE Full text] [doi: 10.1016/j.jacc.2023.09.080]
17. Cordero M, Garcia Gonzalez MJ, Gonzalez Cabeza E, et al. Pre-hospital electrocardiogram recording and transmission is associated with process improvements in treatment of ST elevation acute myocardial infarction. Eur Heart J Acute Cardiovasc Care 2013;2(1):27.
18. Aboal J, Ramos R, Martín C, et al. Evaluation of the ODISEA APP for improving a STEMI regional network. Int J Cardiol 2024 Sep 1;410:132217. [doi: 10.1016/j.ijcard.2024.132217] [Medline: 38830543]
19. Bendary A, Mansour AM, Kabil HK, et al. Time-to-reperfusion in ST-segment elevation myocardial infarction undergoing inter-hospital transfer using WhatsApp smartphone application. Eur Heart J 2020 Nov 1;41(Supplement_2). [doi: 10.1093/ehjci/ehaa946.3471]
20. Bladin CF, Bagot KL, Vu M, et al. Real-world, feasibility study to investigate the use of a multidisciplinary app (Pulsara) to improve prehospital communication and timelines for acute stroke/STEMI care. BMJ Open 2022 Jul 18;12(7):e052332. [doi: 10.1136/bmjopen-2021-052332] [Medline: 35851025]
21. Llempén Ibañez E, López R, Palavecino J, et al. Evaluation of strategies to reduce reperfusion times in ST-segment elevation acute myocardial infarction. Mobile health experience (mHealth). Rev Argent Cardiol 2019 Jun;87(4):305-308. [doi: 10.7775/rac.v87.i4.14399]
22. Park JJ, Yoon CH, Suh JW, et al. Reduction of ischemic time for transferred STEMI patients using a smartphone social network system. J Am Coll Cardiol 2016 Sep;68(13):1490-1492. [doi: 10.1016/j.jacc.2016.07.733]
23. Yufu K, Shimomura T, Fujinami M, et al. Impact of mobile cloud electrocardiography system on door-to-balloon time in patients with acute coronary syndrome in Oita prefecture. Circ Rep 2019;1(6):241-247. [doi: 10.1253/circrep.CR-19-0020]
24. Chao CC, Chen YC, Shih CM, et al. Smartphone transmission of electrocardiography images to reduce time of cardiac catheterization laboratory activation. J Chin Med Assoc 2018 Jun;81(6):505-510. [doi: 10.1016/j.jcma.2017.11.009] [Medline: 29287703]
25. Kang G, Zhang H, Zhou J, et al. The effectiveness of the smartphone-based WeChat platform on reducing time to diagnosis and treatment of ST-segment elevation myocardial infarction. Rev Cardiovasc Med 2023 Dec;24(12):374. [doi: 10.31083/j.rcm2412374] [Medline: 39077096]
26. Liu H, Wang W, Chen H, et al. Can WeChat group-based intervention reduce reperfusion time in patients with ST-segment myocardial infarction? A controlled before and after study. J Telemed Telecare 2020 Dec;26(10):627-637. [doi: 10.1177/1357633X19856473] [Medline: 31238784]

XSL•FO
RenderX

27. Yu SH, Shih HM, Chang SS, et al. Social media communication shorten door-to-balloon time in patients with ST-elevation myocardial infarction. Medicine (Baltimore) 2019 Mar;98(10):e14791. [doi: 10.1097/MD.0000000000014791] [Medline: 30855493]
28. Abrahim C, Capatina A, Kalyan-Sundaram A, et al. Reducing door-to-balloon time using EMS-initiated app-based communication. J Soc Cardiovasc Angiogr Interv 2024 Sep;3(9):102199. [doi: 10.1016/j.jscai.2024.102199] [Medline: 39575209]
29. Aldajani A, Chaabo O, Brouillette F, et al. Impact of a mobile, cloud-based care-coordination platform on door-to-balloon time in patients with STEMI: initial results. CJC Open 2024 Sep;6(9):1094-1098. [doi: 10.1016/j.cjco.2024.05.011] [Medline: 39525824]
30. Arinaga T, Suematsu Y, Nakamura A, et al. The effectiveness of mobile cloud 12-lead electrocardiogram transmission system in patients with ST-segment elevation myocardial infarction. Medicina (Kaunas) 2022 Feb 6;58(2):247. [doi: 10.3390/medicina58020247] [Medline: 35208570]
31. Dickson R, Nedelcut A, Seupaul R, et al. STOP STEMI©-a novel medical application to improve the coordination of STEMI care: a brief report on door-to-balloon times after initiating the application. Crit Pathw Cardiol 2014 Sep;13(3):85-88. [doi: 10.1097/HPC.0000000000000019] [Medline: 25062390]
32. Kini AS, Garcia H, Springer B, et al. A mobile application for STEMI care optimization: pilot implementation project report. Int J Cardiol 2024 Nov 15;415:132447. [doi: 10.1016/j.ijcard.2024.132447] [Medline: 39147281]
33. Kohashi K, Nakano M, Isshiki T, et al. Clinical efficacy of pre-hospital electrocardiogram transmission in patients undergoing primary percutaneous coronary intervention for ST-segment elevation myocardial infarction. Int Heart J 2023 Jul 29;64(4):535-542. [doi: 10.1536/ihj.22-633] [Medline: 37460322]
34. Studencan M, Alusik D, Plachy L, et al. Significant benefits of new communication technology for time delay management in STEMI patients. PLoS ONE ;13(11):e0205832. [doi: 10.1371/journal.pone.0205832]
35. Moxham RN, d'Entremont MA, Mir H, et al. Effect of prehospital digital electrocardiogram transmission on revascularization delays and mortality in ST-elevation myocardial infarction patients: systematic review and meta-analysis. CJC Open 2024 Oct;6(10):1199-1206. [doi: 10.1016/j.cjco.2024.06.012] [Medline: 39525334]
36. Larson DM, Menssen KM, Sharkey SW, et al. "False-positive" cardiac catheterization laboratory activation among patients with suspected ST-segment elevation myocardial infarction. JAMA 2007 Dec 19;298(23):2754-2760. [doi: 10.1001/jama.298.23.2754] [Medline: 18165668]
37. Sejersten M, Sillesen M, Hansen PR, et al. Effect on treatment delay of prehospital teletransmission of 12-lead electrocardiogram to a cardiologist for immediate triage and direct referral of patients with ST-segment elevation acute myocardial infarction to primary percutaneous coronary intervention. Am J Cardiol 2008 Apr 1;101(7):941-946. [doi: 10.1016/j.amjcard.2007.11.038] [Medline: 18359312]
38. Bhattarai HK, Bhusal S, Barone-Adesi F, et al. Prehospital emergency care in low- and middle-income countries: a systematic review. Prehosp Disaster Med 2023 Aug;38(4):495-512. [doi: 10.1017/S1049023X23006088] [Medline: 37492946]
39. Alhammad N, Alajlani M, Abd-Alrazaq A, et al. Patients' perspectives on the data confidentiality, privacy, and security of mHealth apps: systematic review. J Med Internet Res 2024 May 31;26:e50715. [doi: 10.2196/50715] [Medline: 38820572]
40. Menees DS, Peterson ED, Wang Y, et al. Door-to-balloon time and mortality among patients undergoing primary PCI. N Engl J Med 2013 Sep 5;369(10):901-909. [doi: 10.1056/NEJMoa1208200] [Medline: 24004117]
41. Spineli LM, Kalyvas C. Comparison of exclusion, imputation and modelling of missing binary outcome data in frequentist network meta-analysis. BMC Med Res Methodol 2020 Feb 28;20(1):48. [doi: 10.1186/s12874-020-00929-9] [Medline: 32111167]

Abbreviations

Cath lab: catheterization laboratory
D2B: door to balloon
ECG: electrocardiogram
ED: emergency department
EMS: emergency medical service
FMC2B: first medical contact to balloon time
mHealth: mobile health
ODISEA: Biomedical Research Institute of Girona
OR: odds ratio
PCI: percutaneous coronary intervention
PICO: populations, people, patient, or problem, intervention, comparison, and outcome
pPCI: primary percutaneous coronary intervention
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
STEMI: ST-segment elevation myocardial infarction

XSL•FO
RenderX

Edited by L Buis; submitted 18.09.24; peer-reviewed by J Aboal, J Zhang; revised version received 13.03.25; accepted 21.04.25; published 25.08.25.

Please cite as:
Gibson W, Al Kindi D, Akl E, Dandona KB, Pelletier JP, Piazza N, Zgheib A, Martucci G, Spaziano M
Impact of Smartphone Apps on Reperfusion Times and Clinical Outcomes in Acute ST-Segment Elevation Myocardial Infarction: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2025;13:e66605
URL: https://mhealth.jmir.org/2025/1/e66605
doi:10.2196/66605

© William Gibson, Dawoud Al Kindi, Elie Akl, Kshitij Badal Dandona, Jean-Philippe Pelletier, Nicolo Piazza, Ali Zgheib, Giuseppe Martucci, Marco Spaziano. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparative Effectiveness of Digital Health Technologies in Tuberculosis Treatment: Systematic Review and Network Meta-Analysis of Randomized Controlled Trials

Qinglin Cheng[1,2*], PhD; Ping Chen[3*], MD; Ruoqi Dai[1*], MD; Qingjun Jia[1], MD; Xuexin Bai[1], MD; Qiancheng Cao[2], MD; Qingchun Li[1], MD; Yifei Wu[1], MD; Yinyan Huang[1], MD

[1]Department of Tuberculosis Control and Prevention, Hangzhou Center for Disease Control and Prevention (Hangzhou Health Supervision Institution), 568 Mingshi Road, Hangzhou, China

[2]Department of Public Health, School of Public Health and Nursing, Hangzhou Normal University, Hangzhou, China

[3]Central Laboratory, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Ruoqi Dai, MD
Department of Tuberculosis Control and Prevention, Hangzhou Center for Disease Control and Prevention (Hangzhou Health Supervision Institution), 568 Mingshi Road, Hangzhou, China

Abstract

Background: Tuberculosis (TB) treatment remains a critical global health challenge, as traditional standard of care (SoC) approaches face limitations in accessibility and efficacy. While digital health technologies (DHTs) offer promising solutions to address these gaps, limited evidence exists on their comparative effectiveness.

Objective: This study systematically evaluates and compares the impact of diverse DHTs on improving TB treatment outcomes and adherence, aiming to identify optimal strategies across different patient populations.

Methods: A systematic search was conducted across PubMed, Cochrane Library, Embase, and Web of Science from database inception through February 28, 2025, with no language restrictions. Eligible studies included randomized controlled trials comparing DHTs with SoC for TB treatment. The primary outcome was treatment success, defined as completion or cure. A random-effects network meta-analysis was performed, calculating odds ratios (OR) and 95% credibility intervals (CrI) to assess treatment effects. Surface under the cumulative ranking curve (SUCRA) values were used to rank intervention effectiveness. This study is registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42025601199).

Results: From 2420 screened studies, 27 randomized controlled trials involving 23,283 patients and eight DHT interventions were included. The network meta-analysis revealed that digital health platforms showed marginal improvements in treatment success (OR=3.44; 95% CrI 0.95-11.67; SUCRA=0.913; *P*=.05). Compared with SoC, video directly observed treatment (VDOT) significantly improved treatment success (OR=2.39; 95% CrI 1.18-4.75; SUCRA=0.848; *P*=.01). Medication event reminder monitors significantly enhanced treatment adherence (OR=3.13; 95% CrI 1.55-7.05; SUCRA=0.891; *P*=.003).

Conclusions: Results underscore the significant potential of DHTs to improve TB treatment management. VDOT emerged as the most effective intervention for enhancing treatment success, while medication event reminder monitors demonstrated efficacy in sustaining adherence. Digital health platforms showed promise but require additional validation. Caution is warranted due to potential heterogeneity across studies, which may affect generalizability. This research offers actionable insights for stakeholders aiming to optimize TB management through strategic DHT integration.

(JMIR Mhealth Uhealth 2025;13:e75424) doi:10.2196/75424

KEYWORDS

digital health technologies; tuberculosis treatment; network meta-analysis; randomized controlled trials; mobile application; tuberculosis intervention

Introduction

Tuberculosis (TB) remains a formidable global health challenge, as it is a highly infectious disease caused by Mycobacterium TB that primarily targets the lungs. Annually, TB is responsible for approximately 10.6 million new infections and 1.6 million deaths worldwide [1]. Although TB is treatable, the treatment regimen is often protracted, typically spanning several months or longer, depending on the severity of the disease and the presence of drug resistance [2]. Interruptions in TB treatment can lead to disease progression, the emergence of drug-resistant

XSL•FO
RenderX

strains, and increased mortality, ultimately escalating health care costs [3]. Critical to achieving the World Health Organization (WHO)'s "End TB Strategy" and reducing the global TB burden is ensuring patients complete the full treatment course [4].

Standard of care (SoC), a cornerstone of the treatment for TB, is strongly recommended by the WHO as a critical treatment strategy [5,6]. SoC involves a health care worker or designated individual directly observing the patient as they take each dose of medication [6]. Despite its widespread adoption, some studies have suggested that SoC may not significantly improve treatment outcomes compared with self-administered therapy [6,7]. This has underscored the urgent need for innovative approaches to TB patient management. Recent advancements in digital health technologies (DHTs) have opened new avenues for enhancing TB treatment adherence, offering promising alternatives to traditional SoC [8]. Several randomized controlled trials (RCTs) have demonstrated the efficacy of various DHTs, including video directly observed treatment (VDOT) [9], wirelessly observed therapy (WOT) [10], and SMS reminders [7], in improving therapeutic outcomes for TB patients. However, systematically evaluating the comparative effectiveness of DHTs presents significant challenges due to the small sample sizes of individual studies, the heterogeneity of patient populations, and the diverse array of DHTs used [11]. A recent meta-analysis directly compared the impact of 6 types of DHTs against DOT, but the relative efficacy among different DHTs remains unclear [8]. Furthermore, as DHTs continue to evolve and proliferate at an unprecedented pace, ongoing assessment of their efficacy is essential. Network meta-analysis (NMA) offers a robust framework for simultaneously evaluating multiple interventions, even in the absence of direct comparative evidence from head-to-head trials [12]. While network meta-analysis has been widely applied across various disciplines, there is a notable gap in the literature synthesizing evidence on the efficacy of DHTs in TB treatment.

Given this context, a comprehensive reevaluation of the role of DHTs in TB treatment is imperative. To address this need, we hypothesize that certain DHTs will demonstrate superior efficacy over traditional SoC in enhancing TB treatment adherence and success rates, while significant heterogeneity may exist among different DHTs. By leveraging NMA, we anticipate identifying the most effective digital interventions, thereby informing global TB management strategies and advancing the WHO's "End TB Strategy."

Methods

Overview

This study was registered in PROSPERO (International Prospective Register of Systematic Reviews; registration number: CRD42025601199) and was meticulously designed and executed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, including its extension statement for network meta-analyses [13,14]. The PRISMA checklist can be found in Checklist 1.

Data Sources and Search Strategy

The review team [DRQ, CQL, CP, JQJ, BXX, CQC, LQC, WYF, and HYY] conducted a comprehensive literature search across PubMed, Cochrane Library, Embase, Google Scholar, and Web of Science from database inception through February 28, 2025, without language restrictions. Search strategies were developed using Medical Subject Headings (MeSH) and previous study protocols [8,15]. For PubMed and Cochrane Library, searches incorporated DHT- and TB-specific keywords. Given higher baseline retrieval volumes in Embase, Google Scholar, and Web of Science, RCT-related keywords were explicitly included to refine results. Full search syntax is detailed in Table S1 in Multimedia Appendix 1. Although the Cochrane RCTs filter wasn't applied as a discrete element, our strategy inherently embedded RCTs-specific terms and MeSH headings in PubMed or Cochrane Library searches, achieving comparable precision. Explicit RCTs keywords were used for the remaining databases to optimize specificity. In addition, we reviewed reference lists from systematic reviews and meta-analyses, incorporating eligible articles as supplementary studies. This approach ensures a rigorous and comprehensive synthesis of the available evidence, providing valuable insights into the comparative effectiveness of DHTs in TB treatment.

Intervention Definition

DHTs for TB treatment and adherence include: (1) messaging systems-SMS (1-way, automated reminders without response required) and SMS (2-way, interactive messaging enabling patient-provider communication); (2) direct contact methods-phone calls (personalized provider follow-ups) and labels (medication packaging with toll-free reporting numbers); (3) advanced monitoring devices-medication event reminder monitors (MERM; smart pillboxes with tracking capabilities), VDOT (remote video medication observation), and WOT (ingestion sensor systems for biometric verification); and (4) comprehensive solutions-digital health platforms (DHP; integrated platforms combining messaging, education, and adherence tracking). These technologies range from basic reminder systems to sophisticated verification tools, all aiming to improve TB treatment and adherence through various levels of patient engagement and monitoring. Given the diverse landscape of DHT interventions, we systematically categorized current applications into distinct groups following comprehensive analysis (Table S2 in Multimedia Appendix 1).

Selection Criteria

All literature records were downloaded and imported into Zotero 7.0 for deduplication. Three independent reviewers [CQL, DRQ, and CP] screened titles and abstracts, followed by a full-text review, to identify eligible studies. Any discrepancies were resolved through consensus discussions within the review team.

Studies were considered eligible for inclusion if they met all the following criteria: (1) parallel-arm RCTs or cluster RCTs; (2) participants aged 15 years or older diagnosed with TB, pulmonary TB, or latent TB infection requiring fixed-dose anti-TB treatment regimens; (3) the intervention group received treatment regimens incorporating DHTs, while the control group received SoC; (4) reported at least one outcome of interest (eg,

XSL•FO
RenderX

treatment success, treatment completion, cure, or treatment adherence). Studies were excluded if they met any of the following criteria: (1) DHTs were applied in TB screening or diagnosis rather than treatment; (2) the study design was cross-sectional, cohort, or quasi-experimental; (3) the study was a secondary analysis of previously published RCTs; (4) treatment adherence was reported solely based on dose administration rather than at the individual patient level. In addition, we sought published trial results related to each study protocol, where available. We also reviewed reference lists from previously published systematic reviews and meta-analyses, and any eligible articles identified were included as supplementary records in our research.

Data Extraction

Three reviewers [CQL, DRQ, and CP] independently reviewed the main text and supplementary materials to extract relevant data from eligible studies. Any discrepancies in the pooled dataset were resolved through discussion within the review team. The extracted data included the following: (1) first author; (2) publication year; (3) country or countries where the studies were conducted; (4) study design; (5) participant details (age, sample size, and classifications of TB infection); (6) characteristics of interventions and controls (types of DHTs and detailed descriptions); and (7) treatment outcomes (classification and number of events). For multiarm RCTs, data were extracted separately for each allocation group. In cases where studies reported both intention-to-treat (ITT) and per-protocol analyses, the ITT analysis was prioritized as the primary source of data for this research [16]. Building on previously published meta-analyses and incorporated literature, we conducted a comprehensive and systematic synthesis of DHTs, categorizing them into 8 distinct groups: SMS reminders (further divided into one-way and two-way SMS reminders), phone call, MERM, VDOT, WOT, Labels, and DHP, as detailed in Table S2 in Multimedia Appendix 1.

Risk of Bias and Certainty of Evidence

Three independent investigators [CQL, DRQ, and CP] conducted risk of bias assessments for primary outcomes across all included studies, using the Cochrane Risk of Bias tool for randomized trials [17]. The assessment encompassed 6 critical domains: (1) selection bias, (2) performance bias, (3) detection bias, (4) attrition bias, (5) reporting bias, and (6) other potential sources of bias. Each study was systematically classified as having either low, high, or unclear risk of bias within each domain. Any discrepancies in the reviewers' assessments were resolved through consensus-based discussions among the research team. Comprehensive details regarding the quality assessment process and results are documented in Table S3 in Multimedia Appendix 1. For the primary outcomes, we evaluated the risk of bias arising from missing evidence across all possible pairwise comparisons within the intervention network using the Risk Of Bias due to Missing Evidence in Network meta-analysis (ROB-MEN) tool. Our systematic review and subsequent analyses incorporated all eligible studies identified through the search process, regardless of whether they reported the outcome of interest.

The certainty of evidence was assessed using the CINeMA framework, which represents an adaptation of the Grading of Recommendations Assessment, Development and Evaluation approach specifically designed for NMA. We established an equivalence range of −0.5 to 0.5 for effect sizes, considering values within this range to indicate clinically insignificant differences between interventions. Within the CINeMA framework, the ROB-MEN tool was specifically applied to assess the reporting bias domain.

Outcome Measures

The 2 primary outcomes were treatment success and treatment adherence for TB. Treatment success was measured as a composite outcome comprising 2 distinct patient groups: (1) patients with bacteriologically confirmed TB who achieved successful eradication of Mycobacterium TB (cure), and (2) patients without bacteriological confirmation who completed the full course of anti-TB treatment as prescribed, regardless of bacteriological cure status [18]. Treatment adherence was assessed at the individual level and categorized as either good or poor based on the percentage of missed medication doses [19].

Data Synthesis and Analysis

Assessment of the Transitivity Assumption

We first constructed network evidence plots for primary outcomes using the "multinma" package in R software (version 4.4.2, R Foundation for Statistical Computing). To evaluate the transitivity assumption, we conducted comparative analyses of key clinical and demographic characteristics across different study designs. In our NMA, we initially sought to assess potential effect modifiers including age, gender, diagnosis, comorbidities, and country income levels to evaluate their impact across treatment comparisons and validate the NMA exchangeability assumption. However, due to limited reported data in the included studies, we were unable to extract these key covariates. Consequently, our moderator analyses were restricted to 4 available study-level characteristics: (1) sample size, (2) risk of bias assessment, (3) publication year, and (4) participant diagnostic criteria.

Network Meta-Analysis

We conducted our NMA using STATA statistical software (version 17, Stata Corporation) and R software (version 4.4.2). Further methodological details are provided in Multimedia Appendix 2. Statistical significance was determined using a 2-sided *P* value threshold of <.05. To evaluate the differential effects of various DHTs, we conducted a random-effects network meta-analysis within a Bayesian framework for each outcome. Previous treatment assumptions were incorporated to generate posterior probability distributions of treatment efficacy. Markov chain Monte Carlo sampling was performed using "JAGS," with 4 chains, a thinning interval of 20, 10,000 simulation iterations, and a 5000-iteration burn-in period. We used noninformative priors for all parameters to ensure that the results are driven primarily by the data. For treatment effects, we used a normal distribution with a mean of 0 and a large variance (eg, N[0, 1000]). For heterogeneity parameters (eg, between-study standard deviation), we used a uniform

XSL•FO
RenderX

distribution (eg, U[0, 5]). Convergence was assessed using trace plots, density plots, and Gelman-Rubin statistics, with smooth plots and values close to 1 indicating robust convergence. Odds ratios (OR) with 95% credibility intervals (CrI) were calculated. Treatment efficacy rankings were determined using the surface under the cumulative ranking area (SUCRA) and associated probabilities. The NMA was performed using the "GeMTC" and "rjags" packages in R software (version 4.4.2). Forest plots and league tables were used to visualize relative effects compared with control groups or between DHT groups, with statistical significance set at a 2-sided *P* value threshold of .05.

The Consistency in NMA

In an intervention network, consistency pertains to the congruence between direct and indirect evidence for the same comparisons. To assess the presence of incoherence, we took 2 approaches. First, we compared direct and indirect evidence within each closed loop of nodes. Second, within the framework of a "design by treatment interaction model," we compared the goodness of fit of a NMA model assuming consistency with one that allows for incoherence [20]. We used the Stata commands "mvmeta" and "ifplot" from the Stata network suite to carry out these analyses. To address inconsistency in the network meta-analysis, we used a node-splitting approach in the R package "GeMTC," with a *P* value <.05 indicating significant inconsistency between direct and indirect comparisons. Heterogeneity was assessed using the I^2 statistic for treatment pairs and network consistency, with an I^2 >50% indicating substantial heterogeneity and warranting a random-effects model. Ranking probabilities were computed after generating a heterogeneity matrix.

Publication Bias

Publication bias was evaluated using comparison-adjusted funnel plots and Begg's test. When a primary outcome encompassed more than 10 studies, we used the ROB-MEN tool to evaluate potential publication bias [21]. This comprehensive assessment included both a statistical examination of funnel plot asymmetry and an in-depth analysis of underlying causes for any observed asymmetry. To further enhance the precision of our analysis, we generated contour-enhanced funnel plots for pairwise comparisons involving over 10 studies, facilitating the differentiation between publication bias and other sources of asymmetry [22]. Furthermore, upon detecting indications of small study bias, we applied Duval's "Trim and Fill" method bilaterally to the primary outcomes, enabling us to quantify the potential magnitude of the small study effect. This rigorous, multifaceted approach ensured a robust evaluation of potential biases in our meta-analysis [22].

Sensitivity Analysis

To ensure the robustness of our findings, we adopted a rigorous sensitivity analysis framework. First, we proactively excluded trials with a high risk of bias and removed studies investigating derivatives of DHTs. To address inherent inconsistencies within the NMA, we used the node-splitting method via the R package "GeMTC," establishing a *P* value threshold of <.05 to identify statistically significant discrepancies between direct and indirect evidence. Heterogeneity was evaluated using the I^2 statistic for both pairwise comparisons and network-wide consistency, with an I^2 value exceeding 50% indicating substantial heterogeneity and prompting the adoption of random-effects models. The NMA was performed by calculating ranking probabilities following the generation of a heterogeneity matrix.

To validate associations and identify potential modifiers, we conducted a meta-analysis of DHTs' overall efficacy using random-effects models, complemented by subgroup analyses stratified by sample size, risk of bias, publication year, participant diagnosis, and multicenter status. Small-study effects and publication bias were assessed using comparison-adjusted funnel plots and Begg's test, implemented in STATA. This multilayered approach ensured methodological transparency and strengthened the reliability of our conclusions.

Results

Characteristics and Quality of Studies

Our systematic search retrieved a total of 3433 studies from database searches. This included 1250 studies from PubMed, 162 from Embase, 773 from Web of Science, 146 from the Cochrane Library, and 1102 from Google Scholar. In addition, 89 records were identified from references in previous reviews and meta-analyses. After eliminating 857 duplicate articles, 2727 records underwent title and abstract screening, leaving 248 articles for further full-text eligibility evaluation. Ultimately, 27 RCTs (consisting of 25 published articles, 1 preprint, and 1 letter) [19,23-48] involving 23,283 participants were selected for NMA, as illustrated in Figure 1. The characteristics of the included studies [19,23-48] and patients are presented in Table S8 in Multimedia Appendix 1.

Among the studies, 18,230 (85.5%) of the 21,314 participants in 18 studies reported treatment success; 2495 (46.7%) of the 5337 participants in 11 studies reported treatment completion; 1910 (52.8%) of the 3613 participants in 5 studies reported a cure, and 5648 (73.2%) of the 7717 participants in 8 studies showed good treatment adherence. Regarding DHTs, 7 studies reported one-way SMS reminders [25-28,35,40,41]; 5 studies reported 2-way SMS reminders [19,23,24,29,40]; 6 studies reported VDOT [31,33,34,44,45,48]; 8 studies reported MERM [19,36,38,39,42,43,46,47]; 2 studies reported Labels [37,47]; 1 study reported DHP [30]; and 1 study reported WOT [32].

All 27 RCTs had their study quality estimated using the Cochrane Risk-of-Bias Tool. Most of them were assessed as being of good quality with a low risk of bias (Figure S1 in Multimedia Appendix 3). Performance bias was a significant concern because it was impossible to blind patients and health care providers to the assigned DHTs.

XSL•FO
RenderX

Figure 1. PRISMA flowchart.

XSL•FO
RenderX

Certainty of Evidence

A comprehensive assessment of evidence certainty for all outcomes is presented in Tables S4-S7 in Multimedia Appendix 1. Our GRADE evaluation systematically examines the robustness of evidence across all intervention comparisons, detailing the factors influencing certainty ratings including risk of bias, inconsistency, indirectness, imprecision, and publication bias. As shown in Tables S4-S7 in Multimedia Appendix 1, the comprehensive assessment demonstrated favorable certainty of evidence for all TB outcomes.

Assessment of the Transitivity Assumption

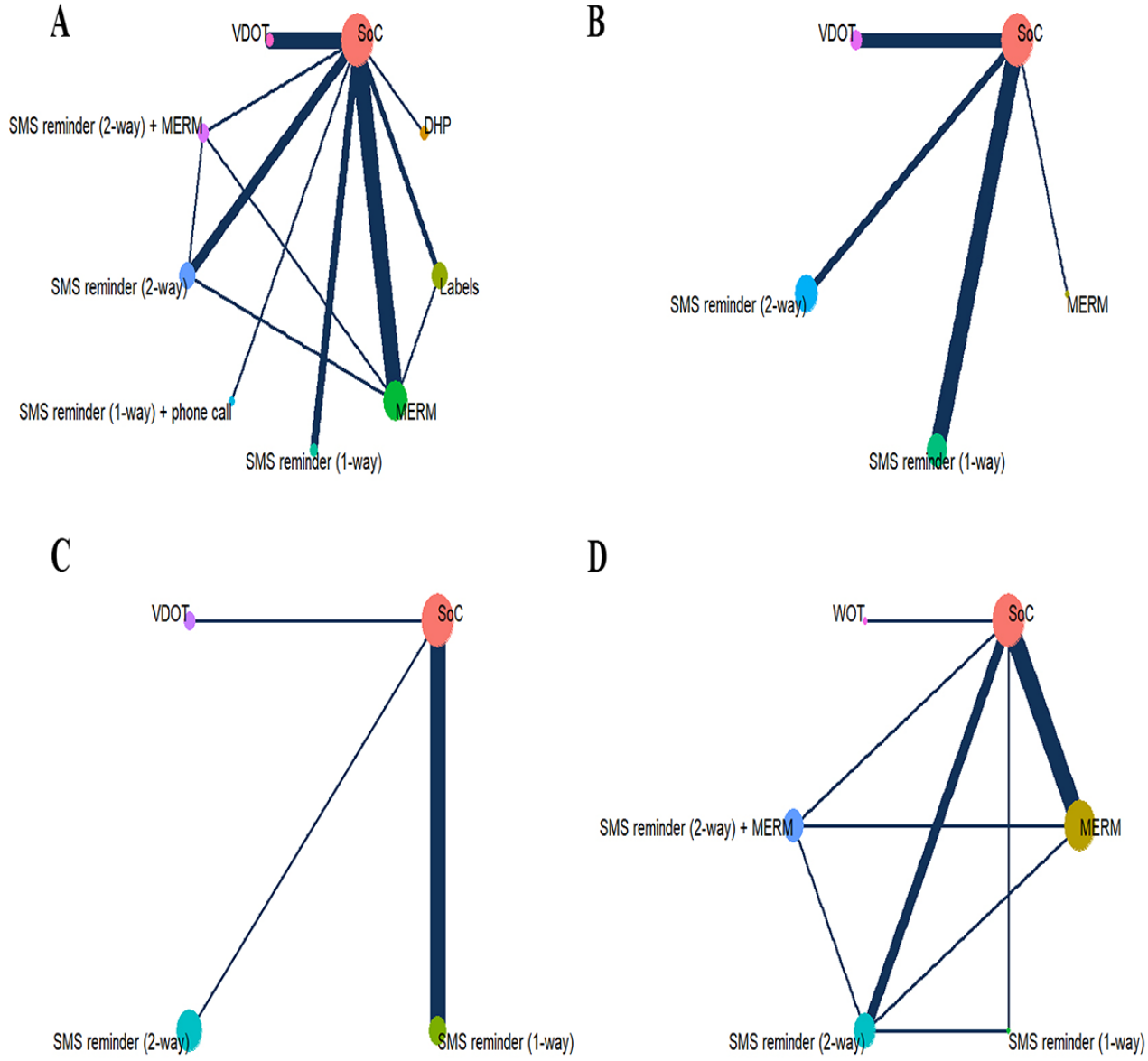
Subgroup analyses revealed that diagnostic classification (TB, OR=2.34, 95% CrI 1.37-4.01) emerged as the strongest effect modifier, with methodological factors (study size [<1000 participants, OR=1.85, 95% CrI 1.14-2.98], publication year [2013 - 2020, OR=1.60, 95% CrI 1.03-2.47]) demonstrating secondary influence, and risk-of-bias thresholds showing negligible impact (Figure S9 in Multimedia Appendix 3).

Network Meta-Analysis

Figure 2 depicts the network evidence plots for treatment success and adherence. When analyzing the network with the SoC as the control, well-connected nodes were observed. A total of 8 distinct DHTs were evaluated for their impact on treatment success. In contrast, only 4 DHTs were assessed for treatment completion and 3 for cure. MERM was the most frequently used intervention in RCTs reporting treatment success and treatment adherence. Figure S2 in Multimedia Appendix 3 displays the trace and density plots, while Figure S3 in Multimedia Appendix 3 presents the Gelman-Rubin diagnostic plots. However, in the analysis of treatment adherence, the multiple chains in the trace plot of WOT did not overlap, and the Gelman-Rubin shrinkage factor deviated from 1 (Figure S2 (D) and Figure S3 (D) in Multimedia Appendix 3). Given the poor convergence of WOT and the absence of events in the corresponding control group, this article was excluded. As a result, 5614 (73.3%) of the 7656 participants in 7 studies exhibited good treatment adherence. The NMA for primary outcomes remained unaffected.

XSL•FO
RenderX

Figure 2. The figure represents (A) network evidence map for tuberculosis (TB) treatment success, (B) TB treatment completion, (C) TB cure, and (D) TB treatment adherence. Each node represents an individual treatment, and larger nodes indicate a greater number of participants. Edges connect pairs of treatments that have been directly compared. And the thickness of an edge reflects the amount of available evidence for comparison. DHP: digital health platform; MERM: medication event reminder monitor; SMS: short messaging service; SoC: standard of care; VDOT: video directly observed treatment.



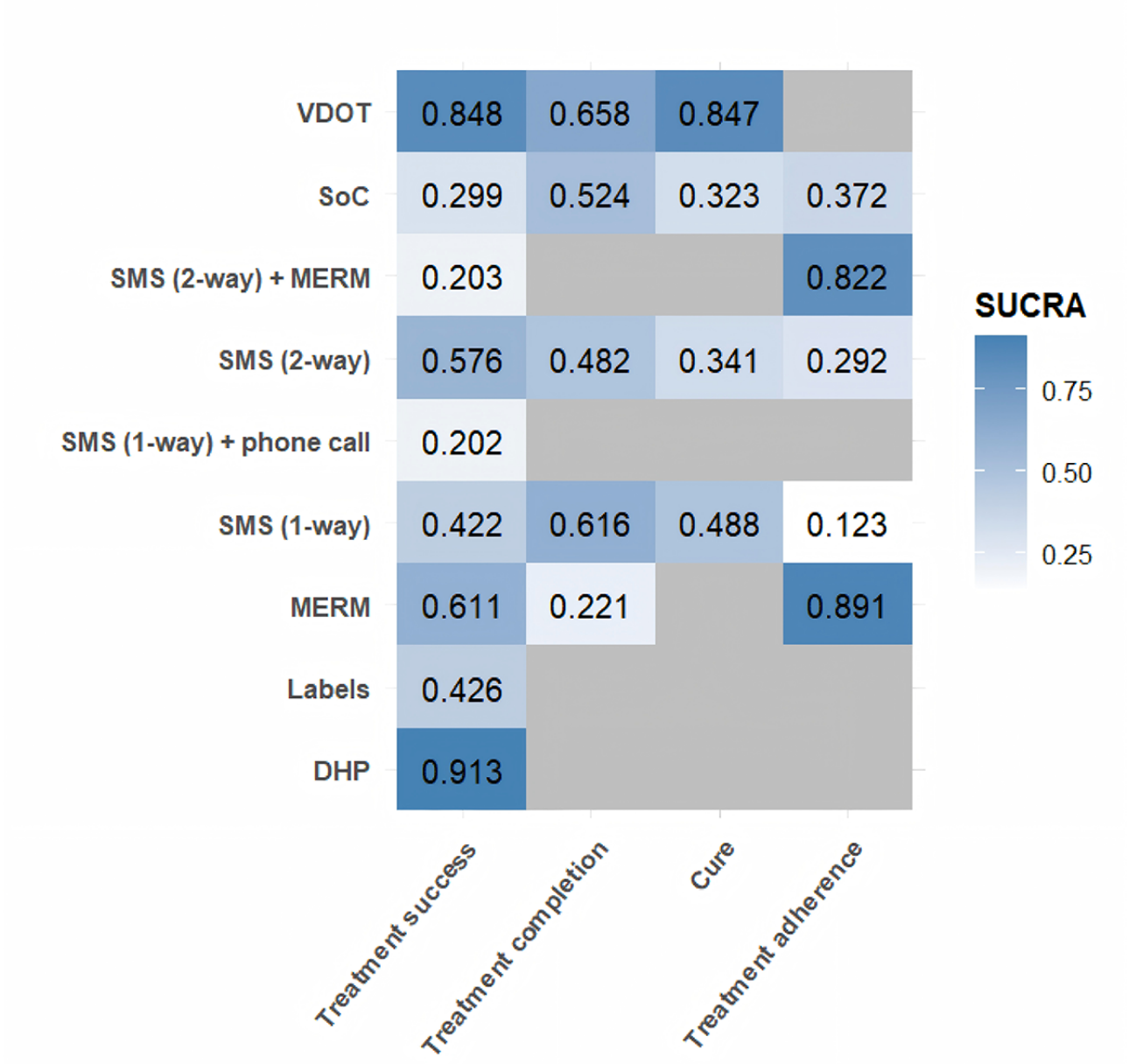
The forest plots of the NMA are presented in Figure S4 in Multimedia Appendix 3. In the NMA, compared with SoC, the adoption of VDOT was significantly associated with increased treatment success (OR=2.39; 95% CrI 1.18-4.75; *P*=.01), while the construction of DHP was marginally associated with increased treatment success (OR=3.44; 95% CrI 0.95-11.67; *P*=.05). When compared with SoC, the application of MERM significantly enhanced treatment adherence (OR=3.13; 95% CrI 1.55-7.05; *P*=.003). There were no statistically significant differences in treatment completion or cure between DHTs and SoC. The league table comparing intervention groups is presented in Table S9 in Multimedia Appendix 1. Our analysis revealed no statistically significant differences in TB treatment success among DHT interventions, with the exception of VDOT versus SoC. Similarly, treatment completion showed no significant differences between the MERM and other interventions, including 1-way SMS, two-way SMS, VDOT, and SoC. Comparable TB cure rates were observed across one-way SMS, 2-way SMS, VDOT, and SoC groups. Notably, MERM demonstrated superior treatment adherence compared with both 1-way and 2-way SMS (all *P*<.001).

The ranking probability bar plots for each outcome are presented in Figure S5 in Multimedia Appendix 3, along with the heatmap of SUCRA values visualized in Figure 3. For treatment success, DHP had the highest rank probability and a SUCRA value of 0.913, followed by VDOT with a SUCRA value of 0.848. Furthermore, VDOT had the highest SUCRA values of 0.658

XSL•FO
RenderX

and 0.847 for treatment completion and cure, respectively. Regarding treatment adherence, MERM and the combination of 2-way SMS plus MERM achieved SUCRA values of 0.891 and 0.822, ranking first and second, respectively.

Figure 3. Heatmap of surface under the cumulative ranking area values for different digital health technologies across 4 study outcomes, with a deeper steel blue in a block indicating a higher surface under the cumulative ranking area value and a gray area denoting a missing value. SUCRA: surface under the cumulative ranking area; DHTs: digital health technologies; DHP: digital health platform; MERM: medication event reminder monitor; SMS: short messaging service; SoC: standard of care; VDOT: video directly observed treatment.



The NMA demonstrated excellent model convergence across all outcomes, as evidenced by the deviance information criterion (DIC) values: treatment success (DIC=76.43), treatment completion (DIC=43.64), cure (DIC=17.51), and treatment adherence (DIC=32.25). The posterior mean deviance (Dbar) and effective parameters (pD) showed proportional relationships, with treatment success exhibiting the highest model complexity (Dbar=40.92, pD=35.51), while cure showed the most parsimonious fit (Dbar=8.76, pD=8.75). The narrow margins between Dbar and pD values across outcomes (range 0.01 - 5.42) indicate stable parameter estimates without overfitting (Table S11 in Multimedia Appendix 1).

Evaluation of Heterogeneity and Inconsistency

Table S10 in Multimedia Appendix 1 presents heterogeneity analysis results. Overall heterogeneity remained consistently high across DHT interventions and outcomes (I^2=0 for TB cure). Treatment completion and cure networks lacked sufficient comparisons for inconsistency assessment, while node-splitting tests revealed no significant inconsistencies in treatment success and adherence networks (Figure S6 in Multimedia Appendix 3).

XSL•FO
RenderX

Assessment of Small Study Effects

No significant publication bias was detected in comparison-adjusted funnel plots (Begg's test all $P>.05$; Figure S7 in Multimedia Appendix 3). The NMA demonstrated consistent direct-indirect evidence agreement, confirming network consistency.

Sensitivity Analyses

Subgroup analyses evaluated DHTs' efficacy and heterogeneity sources across outcomes. Figure S8 in Multimedia Appendix 3 presents pooled efficacy estimates, demonstrating significant treatment success improvement with DHTs (OR=1.47, 95% CrI=1.09 - 1.98; $P<.001$; random-effects model). While subgroup results showed general consistency across factors, heterogeneity for treatment success emerged from sample size, publication year, participant diagnosis, and multicenter design; treatment adherence heterogeneity derived from sample size, participant diagnosis, and multicenter studies (Figure S9 in Multimedia Appendix 3).

Discussion

Principal Findings

This NMA synthesizes evidence from 27 randomized controlled trials (n=23,283) evaluating 8 DHTs across four critical TB treatment outcomes. Key findings demonstrate: (1) VDOT significantly enhances treatment success, aligning with WHO recommendations for visual adherence verification in resource-limited settings; (2) MERM substantially improves treatment adherence; and (3) DHP shows promising but nonsignificant treatment success improvement (OR=3.44, $P=.05$), warranting further investigation despite top-ranking SUCRA performance (0.913). Cumulative ranking analysis identifies DHP, VDOT, and one-way SMS as optimal for treatment success, while MERM (particularly combined with 2-way SMS) excels for adherence. To our knowledge, this represents the first NMA of DHT efficacy in TB management, strengthened by an expanded literature search that enhances conclusion reliability compared with previous meta-analyses.

Interpreting Our Findings in the Context of the Existing Literature

Previous meta-analyses have shown the benefit of certain interventions like VDOT, MERM, and SMS in TB treatment management, but no definitive differences between them have emerged [49]. VDOT serves as a modernized alternative to conventional DOT, enabling health care providers to monitor medication intake through real-time or recorded video consultations [50]. Our study revealed that VDOT outperformed the SoC in improving TB treatment success rates. This enhancement appears to reflect superior treatment completion and cure rates, as evidenced by VDOT's higher SUCRA values compared with SoC. While our analysis did not incorporate individual-level adherence data for VDOT, a quasi-experimental study reported that VDOT implementation led to a 4% increase in attended TB treatment sessions and a 100% improvement in health care worker efficiency for patient management [51]. Furthermore, an RCT established VDOT' non-inferiority to traditional DOT in terms of dose completion rates [52]. Supporting evidence from a systematic review indicated that VDOT implementation substantially reduced clinician supervision time while significantly improving patient satisfaction with treatment [53]. Taken together, these findings—including our own results—provide compelling evidence for VDOT' substantial potential to optimize TB treatment outcomes across multiple performance metrics. MERM uses digital tools—such as smartphone apps and smart pillboxes—to prompt patients to take medications on time and track adherence, a strategy long proposed to improve TB treatment compliance [54]. Our analysis found that MERM significantly enhanced adherence compared with SoC controls, though with only marginal or no differences in TB treatment success—despite a higher SUCRA score for success. This inconsistency across outcomes suggests real-world implementation challenges. A qualitative study noted potential inaccuracies in MERM-recorded adherence due to discrepancies between pillbox opening and actual medication intake [55]. In addition, a cost-effectiveness analysis found MERM more expensive yet less efficacious than SoC [56]. Among the reviewed studies, MERM improved treatment success in small-scale RCTs in Peru and Tibet, China [38,43]. However, larger cluster RCTs in China and South Africa showed no statistically significant effects [19,42,46]. While MERM holds promise for TB treatment, its effectiveness may vary by region, population, or technology type. Further research is needed to clarify its impact, and caution is warranted when extrapolating results across different settings. The global expansion of smartphone accessibility has established SMS as a common tool for TB medication adherence reminders [57]. SMS interventions can be classified by communication modality: one-way (unidirectional messages) or two-way (interactive exchanges between patients and health care providers) [58]. Our analysis found no significant difference in treatment outcomes between either SMS modality and SoC. Similarly, combining SMS with phone call or MERM showed limited efficacy. While Kibu et al [40] reported that SMS improved adherence to antiretroviral therapy, their study found no such benefit for TB treatment. Current evidence does not conclusively support SMS as a standalone intervention for improving TB treatment outcomes.

Beyond the discussed, several other approaches show promise for future TB treatment despite limited current evidence. DHP-integrating adherence monitoring, health education, and motivational support demonstrated potential in our study [30]. While DHPs are well-established in growth hormone therapy [59] and mental health care [60], their application to TB treatment requires further validation. We initially included a WOT study but excluded it from analysis due to zero-outcome efficacy calculations [32]. Although this exclusion didn't affect our primary findings, WOT showed notable adherence benefits that may improve treatment efficacy. Medication labels (eg, 99 DOTS in India) represent a low-cost, privacy-preserving DHTs alternative to DOT in resource-limited settings [37,47]. However, our analysis found no significant advantage over SoC in treatment outcomes. Collectively, these DHTs warrant further research to establish their efficacy in TB management.

XSL•FO
RenderX

Enhanced Analysis and Synthesis

Our NMA reveals significant heterogeneity in the comparative effectiveness of DHTs for TB management. Three key insights emerge: (1) MERM demonstrated superior treatment adherence and success, likely due to its real-time monitoring capabilities and automated alerts that preempt treatment interruptions [61]; (2) VDOT showed high TB cure but incurred greater implementation costs, particularly in resource-limited settings where internet infrastructure remains suboptimal [62]; and (3) SMS-based interventions had divergent outcomes: 2-way SMS improved completion rates by enabling patient-provider dialog, whereas one-way SMS showed limited impact, suggesting passive reminders alone are insufficient to address structural barriers like poverty or stigma [63]. This heterogeneity underscores that DHTs' effectiveness is context-dependent, mediated by technological infrastructure, patient literacy, and health care system integration [64].

This comprehensive NMA establishes DHTs as transformative tools for TB control through 4 key contributions: (1) direct clinical applicability via evidence-based comparison of intervention efficacy; (2) identification of high-potential technologies including VDOT, mobile health (mHealth) solutions, and smart pillboxes (MERM) for optimizing adherence and outcomes; (3) methodological rigor through simultaneous multi-intervention assessment via NMA, providing superior decision-making insights; and (4) PRISMA-compliant design with explicit eligibility criteria, rigorous risk-of-bias evaluation, and statistical mitigation of heterogeneity and bias, ensuring robust conclusions.

Patient and Public Involvement

Effective TB management through DHTs fundamentally requires embedding patient and public involvement (PPI) throughout design, implementation, and evaluation. Robust empirical evidence confirms that early, sustained PPI enhances DHTs' acceptability, usability, and clinical efficacy by aligning innovations with end-user needs and contextual realities [65]. Qualitative syntheses establish four critical engagement pillars—personal agency and motivation, life values, recruitment strategies, and technology quality—where neglect drives inequitable adoption and attrition. Multicultural studies further identify barriers including limited digital literacy, distrust in developers, and inaccessible design, which disproportionately burden vulnerable populations; conversely, PPI-driven co-design directly addresses these challenges through gamified interfaces, culturally adapted incentives, and community-led training [66]. Regulatory authorities such as the Food and Drug Administration now prioritize PPI frameworks, recognizing patient-generated health data and lived-experience insights as indispensable for real-world validation [67]. Although our NMA did not incorporate PPI—a limitation common to efficacy-focused randomized controlled trials—future TB DHT research must integrate participatory methodologies. Consistent with Birnbaum et al's [68] assertion that "end-user engagement in each development phase is essential for scalable impact", we advocate adopting the Canadian Institutes of Health Research framework, mandating inclusive recruitment, shared decision-making, and bidirectional learning to ensure DHTs advance health equity rather than exacerbate disparities.

Methodological Rigor and Evidence Certainty

Our NMA demonstrated robust convergence (DIC range: 17.51 - 76.43) and transitivity, with diagnostic classification identified as the primary effect modifier (OR=2.34 for TB vs other diagnoses). Heterogeneity remained substantial across outcomes (I^2>0), attributable to clinical and methodological variations (Figure S9 in Multimedia Appendix 3). Nevertheless, GRADE assessments indicated moderate-to-high certainty for key comparisons (Tables S4–S7 in Multimedia Appendix 1), supported by network consistency (P>.05 for inconsistency) and minimal publication bias (Begg's test P>.05). Notably, performance bias was inherent across studies due to the impracticality of blinding DHT assignments—a limitation intrinsic to behavioral interventions. Sensitivity analyses confirmed result stability, though heterogeneity patterns suggested influences from study size, publication era, and multicenter designs, highlighting opportunities for standardization in future trials.

Clinical Significance and Implications

This NMA establishes that targeted DHTs significantly improve critical TB treatment outcomes versus SoC. VDOT substantially enhances treatment success, emerging as the most effective evidence-based strategy for elevating cure rates. Concurrently, MERM demonstrates superior efficacy in boosting treatment adherence, positioning them as optimal for adherence-focused programs. Implementation should prioritize VDOT in settings requiring higher treatment success and MERM where adherence is paramount, supported by their leading SUCRA rankings (VDOT or DHP for success and MERM for adherence). Critical translational considerations include: (1) unavoidable performance bias inherent to unblinded DHT interventions, necessitating real-world effectiveness studies; (2) significant treatment effect modification by diagnostic classification, mandating subgroup-specific implementation; and (3) inconclusive efficacy of DHP despite promising success rates, requiring confirmatory studies before scale-up. Notably, no DHTs significantly improved treatment completion or cure rates beyond SoC, revealing persistent management gaps. While moderate-certainty evidence supports VDOT and MERM benefits, heterogeneity in adherence outcomes and insufficient cure or completion data underscore the imperative for standardized outcome measurement in future trials.

Study Limitations

While recognizing their potential benefits, it is crucial to critically examine the inherent limitations within the existing research and outline directions for future investigations. Despite our efforts to comprehensively assess the impact of DHTs in TB management, several limitations and potential biases warrant careful consideration when interpreting the findings. First, despite a comprehensive literature search, the analysis was limited by the relatively small number of included studies. Some network nodes contained few trials, warranting cautious interpretation of results. Second, our conservative outcome selection excluded several short-term (eg, mortality, loss to

XSL•FO
RenderX

follow-up, and adverse reactions) and long-term (eg, relapse) endpoints. Future evaluations of DHT efficacy should consider combined adherence-treatment outcome measures. Third, the analysis focused on pooled treatment effects without examining individual-level effect modifiers. The primary reason is the aggregate nature of the data available from the included RCTs. Most RCTs report outcomes at the group level rather than providing individual patient data, which is a common challenge in systematic reviews and meta-analyses. Without access to individual patient data, it is methodologically impossible to conduct a robust analysis of individual-level effect modifiers. This limitation is inherent to the type of evidence synthesized in this NMA. Fourth, substantial heterogeneity for certain interventions limits result reliability and generalizability. Notwithstanding these limitations, this represents the first network meta-analysis evaluating multiple DHTs' impact on TB treatment outcomes using RCT evidence. Larger-scale RCTs are needed to validate and expand upon these findings.

Conclusions

DHTs offer transformative potential for TB control and advancing global elimination targets. This NMA synthesizes comparative effectiveness evidence across DHT interventions, providing actionable guidance for TB management strategies. Key findings demonstrate VDOT significantly outperforms SoC in achieving treatment success, establishing it as the optimal DHT option. DHP also has potential to enhance treatment success, and MERM is key to maintaining treatment adherence. Although VDOT and mHealth are effective, we must recognize the limitations of current evidence. Conducting rigorous future research is crucial to fully realize their potential.

Acknowledgments

We would like to acknowledge all the researchers who published the articles that contributed to this network meta-analysis. This work is supported by the Basic Public Welfare Research Project of Zhejiang Province (grant numbers: LGF21H260007 and LGF21H190002), the Medical Science and Technology Project of Zhejiang Province (grant numbers: 2020PY064, 2020KY238, 2021KY951, and 2021PY065), the Health Science and Technology Project of Hangzhou Municipality (grant number: 0020190783), and the Hangzhou City Science and Technology Bureau Project (grant numbers: 20220919Y059). No generative AI tools were used in the writing or preparation of this manuscript, consistent with JMIR's policies on AI usage.

Data Availability

We have included all relevant data related to the study in the article or provided them as supplementary materials.

Authors' Contributions

CQL, DRQ, and CP were responsible for the topic and design of this review. CQL obtained funding, and DRQ drafted the manuscript. JQJ, BXX, CQC, LQC, WYF and HYY performed the study selection and data extraction. CQL, DRQ, and CP supervised the project. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables that provide detailed information on the methodology and results of the systematic review.
[DOCX File, 102 KB - mhealth_v13i1e75424_app1.docx]

Multimedia Appendix 2
Computational code for network meta-analysis in digital health research.
[DOCX File, 32 KB - mhealth_v13i1e75424_app2.docx]

Multimedia Appendix 3
Supplementary figures for network meta-analysis of digital health interventions in tuberculosis treatment.
[DOCX File, 6212 KB - mhealth_v13i1e75424_app3.docx]

Checklist 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.
[DOCX File, 37 KB - mhealth_v13i1e75424_app4.docx]

References

XSL•FO
RenderX

1. Goletti D, Meintjes G, Andrade BB, Zumla A, Shan Lee S. Insights from the 2024 WHO Global Tuberculosis Report - More Comprehensive Action, Innovation, and Investments required for achieving WHO End TB goals. Int J Infect Dis 2025 Jan;150. [doi: 10.1016/j.ijid.2024.107325] [Medline: 39631498]
2. Motta I, Boeree M, Chesov D, et al. Recent advances in the treatment of tuberculosis. Clin Microbiol Infect 2024 Sep;30(9):1107-1114. [doi: 10.1016/j.cmi.2023.07.013] [Medline: 37482332]
3. Kimani E, Muhula S, Kiptai T, Orwa J, Odero T, Gachuno O. Factors influencing TB treatment interruption and treatment outcomes among patients in Kiambu County, 2016-2019. PLoS ONE 2021;16(4):e0248820. [doi: 10.1371/journal.pone.0248820] [Medline: 33822794]
4. Basnyat B, Caws M. Prevent TB to end TB. BMJ 2023 Sep 11;382:e077431. [doi: 10.1136/bmj-2023-077431] [Medline: 37696539]
5. Nyang'wa BT, Berry C, Kazounis E, et al. Short oral regimens for pulmonary rifampicin-resistant tuberculosis (TB-PRACTECAL). Lancet Respir Med 2024;12(2):117-128. [doi: 10.1016/S2213-2600(23)00423-8] [Medline: 38072006]
6. Migliori GB, Sotgiu G, Rosales-Klintz S, et al. ERS/ECDC statement: European Union standards for tuberculosis care, 2017 update. Eur Respir J 2018 May;51(5):1702678. [doi: 10.1183/13993003.02678-2017] [Medline: 29678945]
7. Migliori GB, Sotgiu G. Assessing tuberculosis management. Lancet Infect Dis 2015;15(11):1249-1251. [doi: 10.1016/S1473-3099(15)00095-X]
8. Guzman K, Crowder R, Leddy A, et al. Acceptability and feasibility of digital adherence technologies for drug-susceptible tuberculosis treatment supervision: A meta-analysis of implementation feedback. PLOS Digit Health 2023 Aug;2(8):e0000322. [doi: 10.1371/journal.pdig.0000322] [Medline: 37582066]
9. Lee S, Rajaguru V, Baek JS, Shin J, Park Y. Digital health interventions to enhance tuberculosis treatment adherence: scoping review. JMIR Mhealth Uhealth 2023 Dec 4;11:e49741. [doi: 10.2196/49741] [Medline: 38054471]
10. Ridho A, Alfian SD, van Boven JFM, et al. Digital health technologies to improve medication adherence and treatment outcomes in patients with tuberculosis: systematic review of randomized controlled trials. J Med Internet Res 2022 Feb 23;24(2):e33062. [doi: 10.2196/33062] [Medline: 35195534]
11. Lee Y, Raviglione MC, Flahault A. Use of digital technology to enhance tuberculosis control: scoping review. J Med Internet Res 2020 Feb 13;22(2):e15727. [doi: 10.2196/15727] [Medline: 32053111]
12. Yang A, Pechlivanoglou P, Aoyama K. Interpreting and assessing confidence in network meta-analysis results: an introduction for clinicians. J Anesth 2022 Aug;36(4):524-531. [doi: 10.1007/s00540-022-03072-5] [Medline: 35641661]
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
14. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses. Ann Intern Med 2015;162(11):777-784. [doi: 10.7326/M14-2385] [Medline: 26030634]
15. Zeng G, Niu J, Zhu K, et al. Effects of non-pharmacological interventions on depressive and anxiety symptoms in pregnant women: a systematic review and network meta-analysis. EClinicalMedicine 2025 Jan;79:103011. [doi: 10.1016/j.eclinm.2024.103011] [Medline: 39802308]
16. Molero-Calafell J, Burón A, Castells X, Porta M. Intention to treat and per protocol analyses: differences and similarities. J Clin Epidemiol 2024 Sep;173:111457. [doi: 10.1016/j.jclinepi.2024.111457] [Medline: 38977160]
17. Minozzi S, Cinquini M, Gianola S, et al. The revised Cochrane risk of bias tool for randomized trials (RoB 2). J Clin Epidemiol 2020;126:37-44. [doi: 10.1016/j.jclinepi.2020.06.015] [Medline: 32585346]
18. Maze MJ, Nyakunga G, Sakasaka PA, et al. Etiologic investigation of patients diagnosed with bacteriologically unconfirmed tuberculosis in Tanzania. Am J Trop Med Hyg 2023 Oct 4;109(4):733-739. [doi: 10.4269/ajtmh.23-0023] [Medline: 37604470]
19. Liu X, Lewis JJ, Zhang H, et al. Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. PLoS Med 2015 Sep;12(9):e1001876. [doi: 10.1371/journal.pmed.1001876] [Medline: 26372470]
20. Günhan BK, Friede T, Held L. A design-by-treatment interaction model for network meta-analysis and meta-regression with integrated nested Laplace approximations. Res Synth Methods 2018 Jun;9(2):179-194. [doi: 10.1002/jrsm.1285] [Medline: 29193801]
21. Chiocchia V, Nikolakopoulou A, Higgins JPT, et al. ROB-MEN: a tool to assess risk of bias due to missing evidence in network meta-analysis. BMC Med 2021;19(1):304. [doi: 10.1186/s12916-021-02154-7] [Medline: 34724920]
22. Masoumi S, Shahraz S. Meta-analysis using Python: a hands-on tutorial. BMC Med Res Methodol 2022;22(1):193. [doi: 10.1186/s12874-022-01662-1] [Medline: 35764948]
23. Iribarren S, Beck S, Pearce PF, et al. TextTB: a mixed method pilot study. Tuberc Res Treat 2013:349394. [doi: 10.1155/2013/349394] [Medline: 24349766]
24. Mohammed S, Glennerster R, Khan AJ. Impact of a daily SMS medication reminder system on tuberculosis treatment outcomes: a randomized controlled trial. PLoS ONE 2016;11(11):e0162944. [doi: 10.1371/journal.pone.0162944] [Medline: 27802283]
25. Fang XH, Guan SY, Tang L, et al. Effect of short message service on management of pulmonary tuberculosis patients in Anhui Province. Med Sci Monit 2017;23:2465-2469. [doi: 10.12659/MSM.901310] [Medline: 28534455]

XSL•FO
RenderX

26. Farooq RJ, Irfan M, Hussain SF, et al. The role of mobile SMS-reminders in improving drugs compliance in patients receiving anti-TB treatment. J Postgrad Med Inst 2017;31(2):156-162 [FREE Full text]
27. Bediang G, Stoll B, Geissbuhler A, et al. SMS reminders to improve adherence and cure of tuberculosis patients in Cameroon. BMC Public Health 2018;18:583. [doi: 10.1186/s12889-018-5502-x] [Medline: 29716554]
28. Belknap R, Holland D, Feng PJ, et al. Self-administered versus directly observed once-weekly isoniazid and rifapentine treatment of latent tuberculosis infection. Ann Intern Med 2017 Nov 21;167(10):689-697. [doi: 10.7326/M17-1150] [Medline: 29114781]
29. Johnston JC, van der Kop ML, Smillie K, et al. The effect of text messaging on latent tuberculosis treatment adherence: a randomised controlled trial. Eur Respir J 2018 Feb;51(2):1701488. [doi: 10.1183/13993003.01488-2017] [Medline: 29437940]
30. Yoeli E, Rathauser J, Banaei S, et al. Digital health support in treatment for tuberculosis. N Engl J Med 2019;381(10):986-987. [doi: 10.1056/NEJMc1905460] [Medline: 31483976]
31. Story A, Aldridge RW, Smith CM, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis. Lancet 2019;393(10177):1216-1224. [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30876743]
32. Browne SH, Umlauf A, Tucker JA, et al. Wirelessly observed therapy compared to directly observed therapy. PLoS Med 2019;16(10). [doi: 10.1371/journal.pmed.1002891] [Medline: 31613889]
33. Guo P, Qiao W, Wu J, et al. Telemedicine technologies and tuberculosis management. Telemed J E Health 2020;26(9):1150-1156. [doi: 10.1089/tmj.2019.0180] [Medline: 32004090]
34. Ravenscroft L, Kettle S, Persian R, et al. Video-observed therapy and medication adherence for tuberculosis patients. Eur Respir J 2020;56(3):2000493. [doi: 10.1183/13993003.00493-2020] [Medline: 32616594]
35. Khachadourian V, Truzyan N, Harutyunyan A, et al. People-centred care versus clinic-based DOT for continuation phase TB treatment in Armenia. BMC Pulm Med 2020;20(1):105. [doi: 10.1186/s12890-020-1148-4] [Medline: 32264889]
36. Ratchakit-Nedsuwan R, Sawangjit R, Thawornwan U, et al. Ensuring tuberculosis treatment adherence with a mobile-based CARE-call system in Thailand. Infect Dis (Lond) 2020;52(2):121-129. [doi: 10.1080/23744235.2019.1688860] [Medline: 31709858]
37. Cattamanchi A, Crowder R, Kityamuwesi A, et al. Digital adherence technology for tuberculosis treatment supervision. PLoS Med 2021;18(6):e1003628. [doi: 10.1371/journal.pmed.1003628] [Medline: 34161318]
38. Acosta J, Flores P. A randomised controlled trial to evaluate a medication monitoring system for TB treatment. Int J Tuberc Lung Dis 2022;26(1):44-49. [doi: 10.5588/ijtld.21.0463] [Medline: 34969429]
39. Manyazewal T, Woldeamanuel Y, Holland DP, et al. Effectiveness of a digital medication event reminder and monitor device for patients with tuberculosis (SELFTB). BMC Med 2022;20(1):310. [doi: 10.1186/s12916-022-02515-w] [Medline: 36064554]
40. Kibu OD, Kop ML, Muhula S, et al. Treatment adherence among HIV and TB patients using mobile phone text messages. J Trop Med 2022:5456782. [doi: 10.1155/2022/5456782] [Medline: 35350451]
41. Louwagie G, Okuyemi KS, Ayo-Yusuf OA. Effect of a brief motivational interview and text message intervention targeting tobacco smoking. BMJ Open 2022;12(5):e056496. [doi: 10.1136/bmjopen-2021-056496] [Medline: 35534077]
42. Liu X, Blaschke T, Thomas B, et al. Digital adherence technologies to improve tuberculosis treatment outcomes in China. Lancet Glob Health 2023;11(5):e693-e703. [doi: 10.1016/S2214-109X(23)00049-0] [Medline: 37061312]
43. Wei X, Zhang H, Liu Y, et al. Effectiveness of a comprehensive package based on electronic medication monitors in Tibet. Lancet 2024;403(10425):913-923. [doi: 10.1016/S0140-6736(23)02270-5] [Medline: 38309280]
44. Kumwichar P, Chongsuvivatwong V, Prappre T. Tuberculosis treatment compliance under smartphone-based video-observed therapy. JMIR Mhealth Uhealth 2024;12:e54321. [doi: 10.2196/53411] [Medline: 38830205]
45. Garfein RS, Liu L, Cuevas-Mota J, et al. Asynchronous video directly observed therapy to monitor short-course latent tuberculosis infection treatment. Open Forum Infect Dis 2024;11(3):fae030. [doi: 10.1093/ofid/ofae030] [Medline: 38434625]
46. Charalambous S, Dhasmana D, Mackworth-Young C, et al. Treatment adherence and clinical outcomes amongst people with drug-susceptible tuberculosis using medication monitor. EClinicalMedicine 2024;75:102434. [doi: 10.1016/j.eclinm.2024.102745] [Medline: 39170937]
47. Tadesse AW, Manyazewal T, Woldeamanuel Y, et al. Cluster-randomized trial of digital adherence technologies and differentiated care in ethiopia. medRxiv. Preprint posted online on 2024. [doi: 10.1101/2024.01.15.24301322]
48. Sekandi JN, Kasiita R, Zalwango S, et al. Effectiveness of a mobile health intervention (DOT Selfie) in Uganda. JMIR Mhealth Uhealth 2025;13:e57991. [doi: 10.2196/57991] [Medline: 39715573]
49. Langley I, de Groot LM, van den Hof S, et al. Time trend analysis of tuberculosis treatment while using digital adherence technologies. Trop Med Infect Dis 2022;7(5):65. [doi: 10.3390/tropicalmed7050065] [Medline: 35622694]
50. Chen EC, Owaisi R, Goldschmidt L, Maimets IK, Daftary A. Patient perceptions of video directly observed therapy for tuberculosis: a systematic review. J Clin Tuberc Other Mycobact Dis 2024 May;35:100406. [doi: 10.1016/j.jctube.2023.100406] [Medline: 38380432]
51. Chuck C, Robinson E, Macaraig M, et al. Enhancing management of tuberculosis treatment with video directly observed therapy in New York City. Int J Tuberc Lung Dis 2016;20(5):588-593. [doi: 10.5588/ijtld.15.0728] [Medline: 27084809]

XSL•FO
RenderX

52. Burzynski J, Mangan JM, Lam CK, et al. In-person vs electronic directly observed therapy for tuberculosis treatment adherence: a randomized noninferiority trial. JAMA Netw Open 2022 Jan 4;5(1):e2144210. [doi: 10.1001/jamanetworkopen.2021.44210] [Medline: 35050357]
53. Areas Lisboa Netto T, et al. Video-observed therapy (VOT) vs directly observed therapy (DOT) for tuberculosis treatment. PLoS Negl Trop Dis 2024;18(10). [doi: 10.1371/journal.pntd.0012565] [Medline: 39405342]
54. Velen K, Moyo S, Shroufi A, et al. The effect of medication event reminder monitoring on treatment adherence of TB patients. Int J Tuberc Lung Dis 2023;27(4):322-328. [doi: 10.5588/ijtld.22.0503] [Medline: 37035980]
55. Drabarek D, Anh NT, Nhung NV, et al. Implementation of medication event reminder monitors in rural Viet Nam. PLoS ONE 2019;14(12). [doi: 10.1371/journal.pone.0219891] [Medline: 31794578]
56. Sweeney S, Fielding K, Liu X, et al. Unit costs and cost-effectiveness of a device to improve TB treatment adherence in China. IJTLD Open 2024 Jul;1(7):299-305. [doi: 10.5588/ijtldopen.23.0451] [Medline: 39035432]
57. Portz J, Moore S, Bull S. Evolutionary trends in the adoption, adaptation, and abandonment of mobile health technologies: viewpoint based on 25 years of research. J Med Internet Res 2024 Sep 27;26:e62790. [doi: 10.2196/62790] [Medline: 39331463]
58. Daniels J, van der Merwe LLA, Portle S, Bongo C, Nadkarni S, Petrus Peters R. Tailoring a skills-based serostatus disclosure intervention for transgender women in South Africa: protocol for a usability and feasibility study. JMIR Res Protoc 2025 Mar 26;14:e52121. [doi: 10.2196/52121] [Medline: 40138676]
59. Núñez-Benjumea FJ, Gómez-González A, Martínez-Miranda J, et al. Description of a digital health platform for emotional and self-management support. Stud Health Technol Inform 2022;289:290-292. [doi: 10.3233/SHTI220904] [Medline: 35612102]
60. Zaidan AM. The leading global health challenges in the artificial intelligence era. Front Public Health 2023;11:1328918. [doi: 10.3389/fpubh.2023.1328918] [Medline: 38089037]
61. Falzon D, Timimi H, Kurosinski P, et al. Digital health for the End TB Strategy: developing priority products and making them work. Eur Respir J 2016 Jul;48(1):29-45. [doi: 10.1183/13993003.00424-2016] [Medline: 27230443]
62. Nsengiyumva NP, Mappin-Kasirer B, Oxlade O, et al. Evaluating the potential costs and impact of digital health technologies for tuberculosis treatment support. Eur Respir J 2018 Nov;52(5):1801363. [doi: 10.1183/13993003.01363-2018] [Medline: 30166325]
63. Shringarpure K, Gurumurthy M, Sagili KD, et al. Patient adherence to tuberculosis treatment in the Indian subcontinent: systematic review and meta-synthesis of qualitative research. BMJ Open 2023 May 4;13(5):e063926. [doi: 10.1136/bmjopen-2022-063926] [Medline: 37142319]
64. Veronese V, Miller C, Ogundahunsi O, et al. A new online toolkit to support implementation research to enhance the use of digital innovations to End TB. PLOS Digit Health 2023 Feb;2(2):e0000182. [doi: 10.1371/journal.pdig.0000182] [Medline: 36812644]
65. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120. [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
66. Barony Sanchez RH, Bergeron-Drolet LA, Sasseville M, Gagnon MP. Engaging patients and citizens in digital health technology development through the virtual space. Front Med Technol 2022;4:958571. [doi: 10.3389/fmedt.2022.958571] [Medline: 36506474]
67. Izmailova ES, AbuAsal B, Hassan HE, Saha A, Stephenson D. Digital technologies: Innovations that transform the face of drug development. Clin Transl Sci 2023 Aug;16(8):1323-1330. [doi: 10.1111/cts.13533] [Medline: 37157935]
68. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756. [doi: 10.1111/acem.12692] [Medline: 25997375]

Abbreviations

CrI: credibility intervals
DHP: digital health platforms
DHT: digital health technology
DIC: deviance information criterion
ITT: intention-to-treat
MERM: medication event reminder monitors
NMA: network meta-analysis
OR: odds ratios
PPI: patient and public involvement
PRISMA: preferred reporting items for systematic reviews and meta-analyses
RCTs: randomized controlled trials
ROB-MEN: risk of bias due to missing evidence in network meta-analysis
SoC: standard of care

XSL•FO
RenderX

SUCRA: surface under the cumulative ranking area
TB: tuberculosis
VDOT: video directly observed treatment
WHO: World Health Organization
WOT: wirelessly observed therapy

Edited by L Buis; submitted 07.04.25; peer-reviewed by I Farhat, ME Heidari; revised version received 16.06.25; accepted 01.08.25; published 16.09.25.

Please cite as:
Cheng Q, Chen P, Dai R, Jia Q, Bai X, Cao Q, Li Q, Wu Y, Huang Y
Comparative Effectiveness of Digital Health Technologies in Tuberculosis Treatment: Systematic Review and Network Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2025;13:e75424
URL: https://mhealth.jmir.org/2025/1/e75424
doi:10.2196/75424

© Qinglin Cheng, Ping Chen, Ruoqi Dai, Qingjun Jia, Xuexin Bai, Qiancheng Cao, Qingchun Li, Yifei Wu, Yinyan Huang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Integrating Mobile Health App Data Into Electronic Medical or Health Record Systems and Its Impact on Health Care Delivery and Patient Health Outcomes: Scoping Review

Jialing Lin[1], PhD; Shona Marie Bates[1,2], PhD; Luke N Allen[1,3], PhD; Michael Wright[1,4,5], PhD; Limin Mao[1,6], PhD; Michael Kidd[1,3,7,8], AO

[1]International Centre for Future Health Systems, University of New South Wales, Level 2, AGSM Building, UNSW Sydney, Kensington, Australia

[2]School of Population Health, Faculty of Medicine and Health, University of New South Wales (Sydney), Kensington, Australia

[3]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[4]Avant Mutual, Sydney, Australia

[5]Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, Australia

[6]Centre for Social Research in Health, University of New South Wales, Kensington, Australia

[7]College of Health & Medicine, Australian National University, Canberra, Australia

[8]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Jialing Lin, PhD
International Centre for Future Health Systems, University of New South Wales, Level 2, AGSM Building, UNSW Sydney, Kensington, Australia

Abstract

Background: Mobile health (mHealth) apps are increasingly being used to capture patient health data, provide information, and guide self-management, with reported improvements in health care service delivery and outcomes. However, the impact of integrating mHealth app data into electronic medical record or electronic health record (EMR/EHR) systems remains underexplored.

Objective: This study aims to identify what is known about the impact of integrating mHealth app data into EMR/EHR systems on health care delivery and patient outcomes.

Methods: A scoping review was conducted to identify original studies that investigated the integration of patient-facing mHealth app data into EMR/EHR systems and the impact on health care outcomes. The PubMed, Embase, Web of Science, Cochrane Library, CINAHL, ProQuest, and PsycINFO databases were searched for papers published between January 2014 and July 2024. Two authors independently screened and extracted data on study characteristics, mHealth app features, details of integration with EMR/EHR systems, and effects on health care delivery and patient outcomes.

Results: Nineteen studies with 113,135 participants were included. Among these, 6 were randomized clinical trial studies, 8 were conducted in the United States, 12 occurred in hospital settings, 15 involved adult participants, and 6 targeted diabetes management. Main features of the apps and EMR/EHR systems can be categorized into tracking or recording health data (n=19), app data integrated into EMR/EHR systems (n=19), app data summarized or presented on EMR/EHR interface (n=19), communication with the health care team (n=12), reminders or alerts (n=10), synchronization with other apps or devices (n=8), educational information (n=4), and using existing portal credentials to app access (n=2). Most studies reported benefits of integrating the app and EMR/EHR, such as enhanced patient education and self-management (n=5), real-time data recorded and shared with clinicians (n=4), support for clinical decision-making (n=3), improved communication between patients and clinicians (n=7), and improved patient outcomes (n=13). Challenges identified included high drop-off rates in app usage (n=3), limited accessibility due to device restrictions (n=3), incompatibility between mHealth apps and EMR/EHR systems (n=3), increased clinical workload in response to additional information (n=3), data accuracy issues due to network connectivity (n=1), and data security concerns (n=1).

Conclusions: Evidence suggests that the effective integration of mHealth app data into EMR/EHR systems can enhance both clinicians' health care delivery and patients' health outcomes. However, current literature is limited, and future opportunities remain to examine the impact on long-term outcomes, such as mortality, readmissions, and costs, and assess the scalability and sustainability of integration among more broader health conditions and disabilities across diverse health care settings.

(JMIR Mhealth Uhealth 2025;13:e66650) doi:10.2196/66650

XSL•FO
RenderX

KEYWORDS

mHealth; digital health; mobile phones; electronic medical records; integration; usability; implementation; scoping review

Introduction

Background

Mobile health (mHealth) apps, defined as mobile devices used to deliver health care, are increasingly being used by patients to manage chronic disease, monitor and record vital signs, self-diagnose, take medication, track general health and wellness, access electronic patient portals, and access educational resources and self-management tools [1,2]. These apps offer personalized health data tracking and can provide valuable insights for both patients and health care providers. For example, some apps designed for chronic disease management, such as those for diabetes, hypertension, cardiovascular diseases, asthma, and other chronic conditions, have had positive impacts on lifestyle modifications, including weight reduction through healthier eating and regular exercise [3]. Additionally, apps focused on monitoring specific aspects of health, such as child nutrition, mental health, and suicide crises, have shown improved health outcomes, including growth in undernourished children, weight loss in overweight or obese individuals, reduced mental health symptoms, and the prevention of suicide [4-6].

Over the past decade, significant progress has been made in developing interoperability frameworks that have the capacity to connect mHealth solutions with electronic medical record (EMR, defined as a digital version of a patient's medical chart within a single health care facility) or electronic health record (EHR, defined as a comprehensive digital record of a patient's health that can be shared across multiple health care providers and facilities) systems [7-9]. While these technical advancements are essential, integration in health care goes beyond interoperability. It also includes clinical (eg, shared care planning and coordination), functional (eg, harmonized workflows and service delivery processes), and organizational dimensions (eg, collaboration across different service providers and governance structures) [10]. Incorporating data from patients' mHealth apps into EMR/EHR systems may enhance the continuity of care by capturing health information between patient visits [11]. This broader integration can provide a more comprehensive view of a patient's health, streamline documentation processes of patient history, and reduce the burden of manual data entry for health care providers, thus reducing the time spent on documentation during office visits [12,13]. Furthermore, real-time data access allows for better-informed decision-making and more comprehensive patient care [14,15].

Research on the impact of mHealth-EMR/EHR integration on patient outcomes has yielded mixed results. For example, 1 trial recruited 269 people with diabetes and allocated them into 3 groups: a usual care group that did not use an mHealth app, a mobile self-care group that used an mHealth app allowing users to enter their blood glucose data and receive educational information, and a mobile intensive care group that used an app integrated into EMR/EHR systems, enabling physicians to provide personalized feedback. Both the mobile intensive care and mobile self-care groups experienced significant reductions in hemoglobin A_{1c} levels at 12 weeks compared with the usual care group (−1.04% vs −0.86% vs −0.49%; P=.02) [12]. Similarly, a trial involving 68 people with overweight or obesity reported significant weight loss (mean difference 1.4 kg, 95% CI 0.9 - 1.9; P<.001) and a decline in triglyceride levels (mean difference 2.6 mmol/L, 95% CI 17.6 - 75.8; P=.002) among app users compared with nonusers [16]. This app automatically collected physical activity data from a wearable tracker and allowed users to log additional health data, such as daily meals, sleep, weight, and blood pressure, which were summarized and displayed as graphs on the EMR/EHR interface.

However, there is incredible diversity in app function, design, and purpose, and some studies have reported no significant clinical outcomes. For example, a cross-sectional study found no significant difference in the number of visits to rheumatologists, or the time between appointments, among people with rheumatoid arthritis who used an app to receive reminders to record their pain, fatigue, and disease symptoms, compared with a control group (P>.05) [17]. Additionally, a pilot study involving people with diabetes reported no significant change in hemoglobin A_{1c} values between app users, with their glucometer data directly transferred from the app to an EMR/EHR system, and nonusers (P=.08) [18].

The inconsistent findings across studies underscore the need for a review of existing evidence on the integration of mHealth and EMR/EHR and health outcomes, in order to obtain insights and develop practical recommendations for health care policy makers, administrators, and providers, guiding the effective implementation of mHealth app data integration with EMR/EHR systems.

Objectives

This scoping review aimed to synthesize current evidence on how patient-facing mHealth app data can be integrated into EMR/EHR systems and affect health care delivery and patient health outcomes. The hypothesis was that effectively linking mHealth apps with EMR/EHR systems could improve outcomes.

Methods

This scoping review adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) statement [19].

Study Identification

The scoping review identified original studies that assessed the integration of mHealth app data into EMR/EHR systems and impact on health care delivery and outcomes from January 1, 2014, to July 30, 2024, from the following databases: PubMed, Embase, Web of Science, Cochrane Library, CINAHL, ProQuest, and PsycINFO. Search terms including keywords and subject headings were related to mHealth app (eg, "mobile app," "mHealth app," "phone app," "digital app," "eHealth

XSL•FO
RenderX

app," and "health app") and EMR/EHR (eg, "electronic health record," "electronic medical record," "electronic health information," and "electronic medical information") (see search string details in Multimedia Appendix 1). Additionally, the bibliographies of all included studies were manually checked for additional relevant studies. The retrieved studies were then uploaded into Covidence (Veritas Health Innovation), a systematic review management software program.

Study Selection and Data Extraction

Two authors (JL and SB) independently reviewed the studies by screening titles or abstracts and screening full texts, and they performed data extraction. Disagreements were resolved by consensus. Studies were included if they met the following criteria: the apps were patient-facing, patient-related health data were recorded in the apps, app data were integrated into EMR/EHR systems, effects on health care delivery and patient health outcomes were assessed, and the studies were peer-reviewed original studies and published in English. Specifically, this review included both third-party apps (developed outside the EMR/EHR ecosystem) and native EMR/EHR-integrated apps (eg, Epic's offerings). Native EMR/EHR-integrated apps were included because they interact with the EMR/EHR system and serve as comprehensive patient management tools, extending beyond standard EMR/EHR functionality. In addition, studies were excluded if (1) the study design was ineligible: reviews, summaries, research protocols, opinion papers, conference abstracts, correspondence, commentaries, or editorials; (2) the intervention was ineligible: the mHealth app was not patient-faced, did not record health data (eg, appointments or reminders only), consisted of mixed or multiple interventions where the effect of mHealth app use cannot be extracted or identified, or where there was only a mention of integration of app data with EMR/EHR systems; or (3) the outcomes were ineligible: where there was no mention of any health outcomes related to the impact of mHealth app data and their integration with EMR/EHR systems.

Original publications were obtained for papers considered in scope. The following data were extracted into an Excel spreadsheet (version 2408): author, year of publication, study country, study setting, study design, sample size, participant characteristics (eg, age, health condition, etc), app features (eg, app type, health data recording, education, reminder, communication, etc), details of app data integrated into EMR/EHR systems (eg, clinical workflow and efficiency, data interoperability and communication, data quality, system integration, adoption, change management, etc), and effects on health care delivery and patient health outcomes.

Data Synthesis and Analysis

The results were synthesized and presented descriptively. Key features of mHealth apps and EMR/EHR systems were summarized and visualized as a bar chart using R version 4.4.0 (R Core Team 2017, Vienna, Austria).

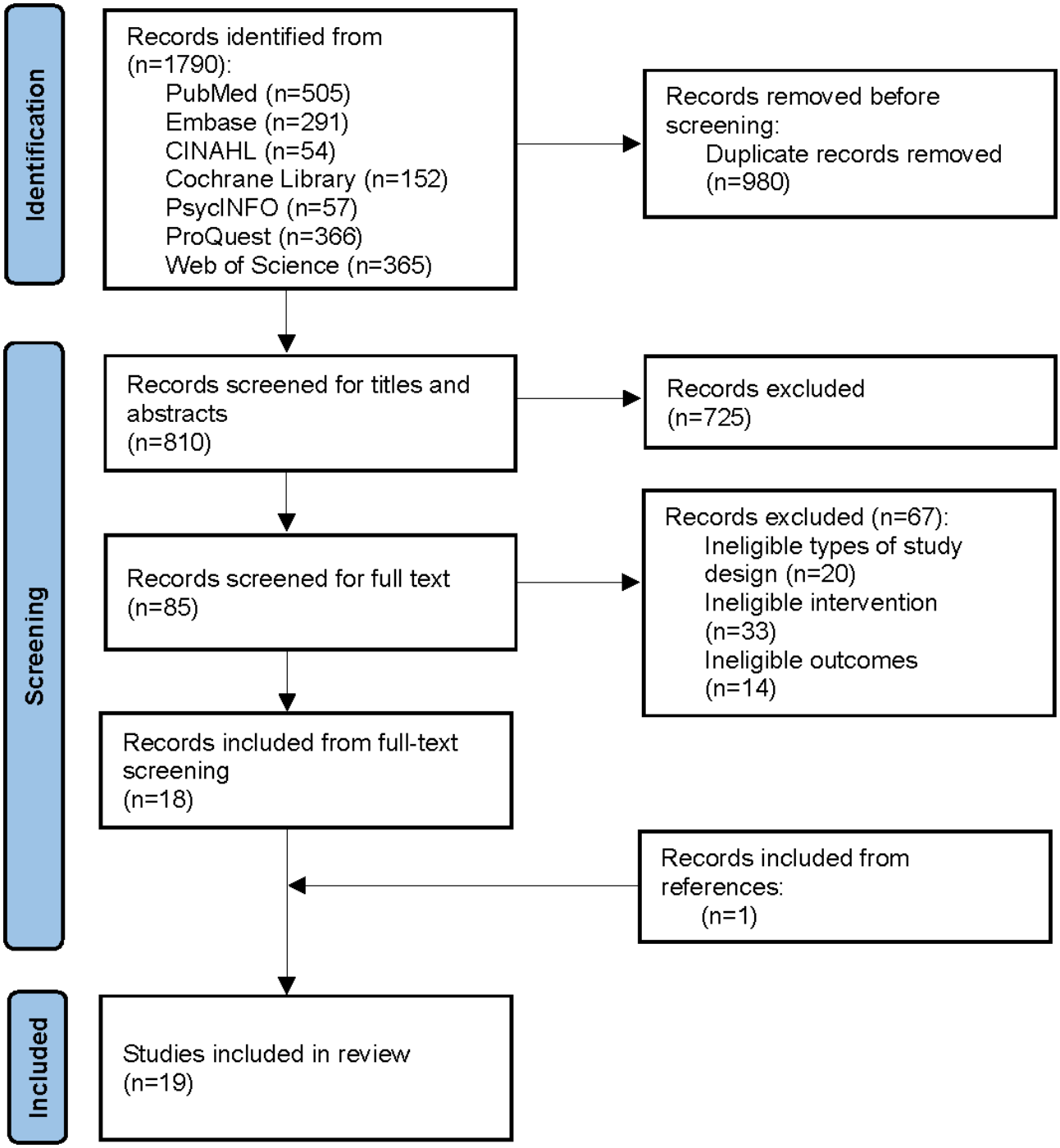
Results

Study Selection

A total of 1790 studies were identified from 7 databases (Figure 1). After removing duplicates, 810 were screened based on and titles and abstracts, resulting in the exclusion of 725 studies that did not meet the inclusion criteria. The remaining 85 studies underwent a full-text review, and 18 studies were deemed eligible. One additional study was included from the bibliography search of the in-scope studies, leaving a total of 19 studies to be included in this review [12-18,20-31].

XSL•FO
RenderX

Figure 1. Flow diagram of study selection.



Study Characteristics

The 19 studies involved 113,135 participants from 5 countries. Of these, 6 were randomized clinical trials, 6 were cohort studies, 2 were cross-sectional studies, 2 were mixed methods studies, 1 was quasi-experimental study, 1 was case series study, and 1 was qualitative study (Table 1). Approximately two-thirds of studies were based out of hospital settings (n=12). Most studies focused on the adult population (n=15). The 3 most commonly targeted health conditions were diabetes (n=6), cancer (n=3), and obstructive sleep apnea (n=2). Details of the study characteristics are shown in Multimedia Appendix 2.

XSL•FO
RenderX

Table . Summary of study characteristics.

Study characteristics	Studies, n (references)
Study design	
Randomized control trial	6 [12,14,16,27-29]
Cohort	6 [13,20,23,25,26,31]
Quasi-experimental	1 [15]
Case series	1 [24]
Cross-sectional	2 [17,21]
Qualitative study	1 [30]
Mixed methods	2 [18,22]
Study location	
United States	8 [13,17,18,22-25,30]
United Kingdom	1 [27]
Italy	1 [20]
South Korea	7 [12,14,16,26,28,29,31]
China	2 [15,21]
Study setting	
Hospital	12 [12-15,20,21,23,26-29,31]
Community	7 [16-18,22,24,25,30]
Study population age type	
Adult	15 [12,14-17,20-25,27-29,31]
Children	4 [13,18,26,30]
Study population health diagnosis	
Diabetes	6 [12,13,18,21,30,31]
Hypertension	1 [25]
Heart failure	1 [15]
Cancer	3 [20,24,28]
Overweight/obesity	1 [16]
Rheumatoid arthritis	1 [17]
Chronic insomnia disorder	1 [23]
Obstructive sleep apnea	2 [14,29]
Severe mental illness	1 [27]
Migraine	1 [22]
Epilepsy	1 [26]
Total	19

Features of mHealth Apps and Their Integrated EMR/EHR Systems

There were 17 unique apps reported in the 19 studies. One app was used in 2 studies [16,29], and another app was used in the other 2 studies [13,30]. The main features of the apps and EMR/EHR systems can be categorized as tracking or recording health data (n=19), app data integrated into EMR/EHR systems (n=19), app data being summarized or presented on EMR/EHR interface (n=19), communicating with the health care team (n=12), providing reminders or alerts (n=10), synchronizing with other apps or devices such as “wearables” (n=8), offering educational resources (n=4), and using existing portal credentials to app access (n=2) (Multimedia Appendix 3). All 17 apps included at least one of these features. Most of the apps were third-party apps developed outside the EMR/EHR ecosystem (n=17), while 2 were native EMR/EHR-integrated apps (ie, Epic’s modules) (n=2) [22,24]. Details of the characteristics of the apps and systems can be found in Multimedia Appendices 2, 4, and 5.

XSL•FO
RenderX

Study Outcomes

Patient outcomes that were explored in the included studies related to blood glucose (n=4) [12,18,21,31], blood pressure (n=2) [15,25], blood triglyceride (n=1) [16], pain intensity (n=2) [20,28], sleep patterns (n=3) [14,23,29], body weight (n=4) [14-16,29], diet (n=1) [15], respiratory distress (n=1) [29], oxygen desaturation (n=1) [29], reactions to treatment (n=1) [28], medication adherence rates (n=2) [15,26], doctor visits (n=1) [17], patient satisfaction (n=2) [15,22], patient health awareness (n=2) [13,15], patient health knowledge (n=1) [26], and patient app usage (n=3) [24,27,28]. Other outcomes reported related to health care providers, such as provider knowledge (n=1) [30], provider access to patient records (n=3) [27,29,30], communication with patients (n=3) [15,25,30], efficiency of patient data interpretation (n=5) [13,25,27,29,30], clinical decision-making (n=5) [15,22,25,27,29], and provider workflow (n=6) [13,15,17,22,25,31].

Benefits of mHealth App Data Integrated Into EMR/EHR Systems on Outcomes

Most studies reported beneficial effects of mHealth app data integration into EMR/EHR systems (n=13) [12-16,20-23,25-27,29,30]. Specifically, 5 studies reported that the integration enhanced patient education and self-management [12,14,16,21,29]. These studies showed benefits in apps targeting diabetes management [12,21], obstructive sleep apnea [14,29], and weight loss [16] through behavioral interventions and lifestyle changes.

Four studies showed benefits of real-time data recording and sharing with clinicians [13,15,27,30]. Apps with features that enable patients to record or upload health data were found to enhance and sustain patient engagement in their health care. Additionally, 3 studies reported benefits in supporting clinical decision-making [13,15,30], where mHealth apps were directly integrated with devices such as glucometers, blood pressure monitors, and wearable trackers. This enabled automatic data transmission to the app and EMR/EHR systems, allowing health care teams to promptly assess patient health status and make informed therapeutic decisions.

Seven studies reported improved communication between patients and clinicians [12,13,15,20,25,27,30], facilitated by built-in messaging features. Enhanced communication was associated with better-coordinated care, more timely interventions, and improved patient outcomes. Furthermore, EMR/EHR-integrated app-based systems with built-in reminders or alerts were found to support clinicians in managing patient care, enabling timely detection and intervention for changes in patient health status, ultimately leading to better health outcomes [17,21,26,27]. Overall, 13 studies reported better patient health outcomes [12-16,20-23,25-27,29], including reductions in blood glucose levels, blood pressure measurements, chemotherapy-related adverse events, body weights, and seizure frequency.

Challenges of mHealth App Data Integrated Into EMR/EHR Systems on Outcomes

However, some challenges were identified, including high drop-off rates in app usage due to non–user-friendly interfaces or device breakdowns (n=3) [23,24,26], limited accessibility due to device restrictions (n=3) [16,27,30], data accuracy issues related to network connectivity (n=1) [27], data security concerns (n=1) [18], compatibility problems between mHealth apps and EMR/EHR systems (n=2) [13,27], and increased clinical workload due to receiving additional data (n=3) [15,17,31].

High drop-off rates in app usage were attributed to difficulties in using the apps, the requirement for manual data entry, and device malfunctions, all of which affected compliance and reduced user satisfaction [23,24,26]. Additionally, patient age was reported as a barrier to adopting digital health care and may impact their willingness to use apps in 1 clinical trial conducted in South Korea [12]; this reinforces the importance of consumer involvement to understand their needs and preferences.

Data accuracy remains a concern when it relies heavily on patient self-reported data. A cohort study, involving 99 people with epilepsy and caregivers in a South Korea hospital, found that the accuracy of using apps to measure medication adherence varied significantly depending on patient behavior and engagement with the app when it required users to manually enter data [26].

Data confidentiality concerns also pose a barrier to app adoption and the completeness of data captured. A mixed methods study conducted in a US diabetes clinic, which found no difference in blood glucose levels between people with diabetes who used the app and those who did not, reported that some patients and caregivers expressed concerns about the confidentiality of their data recorded in the app [18]. Ensuring that patient data shared through mHealth apps and EMR/EHR systems remain confidential and adhere to relevant privacy regulations is pivotal.

Broader considerations are needed regarding data integration across different systems. For example, a clinical trial conducted in UK mental health clinics found that app data uploads, which require a wireless network, can be disrupted if users are temporarily in areas without network coverage, leading to mistimed data recording [27].

Compatibility issues between apps and different EMR/EHR platforms present another significant challenge. Two studies found that the use of multiple nonstandardized EMR/EHR platforms within a single health system acted as a barrier to the widespread rollout of integrated solutions across entire health care systems [13,27]. This lack of standardization made it difficult to streamline integration between a single app and different EMR/EHR platforms. A cohort study conducted in a US medical center found that integrating app data into EMR/EHR systems that rely on specific devices, such as Apple Health, requires the use of Apple devices and can limit patient accessibility and usage [13]. Additionally, some apps were available only for 1 operating system, such as Android or iOS (Apple), restricting access for users with mobile devices running different operating systems [16,27,30].

Increased clinical workload and insufficient training for health care providers also emerged as barriers. A cross-sectional study of patients with rheumatoid arthritis in a large US medical center compared 150 app users whose data were integrated into the

XSL•FO
RenderX

EMR/EHR system with 150 nonusers [17]. The study reported insufficient training on using EMR/EHR systems with integrated app data, while clinicians' fully booked schedules hindered their ability to incorporate patient-generated data into clinical workflows and make informed decisions. This may explain the lack of significant differences in health outcomes between app users and nonusers [17]. Furthermore, 2 studies found that an inadequate clinician workforce can hinder the timely provision of feedback by the health care team [15,31], further restricting the potential benefits of mHealth app-EMR integration. Additional details are reported in Multimedia Appendix 6.

Discussion

Principal Results

This review presents the first comprehensive synthesis of the literature examining integrating mHealth app data into EMR/EHR systems and its impact on health care outcomes. The 19 studies reported 17 unique patient-facing mHealth apps from 5 countries. The primary features of these apps and EMR/EHR systems include tracking or recording patient health data, integrating app data into EMR/EHR systems, facilitating communication with health care teams, providing reminders or alerts, synchronizing with other apps or devices, displaying summarized app data on EMR/EHR interfaces, using existing portal credentials to app access, and offering educational resources. The findings reveal both benefits and challenges to mHealth app-EMR/EHR integration across different clinical settings. Most studies reported benefits of integration, including enhanced patient education and self-management, real-time data recording and sharing with clinicians, improved support for clinical decision-making, strengthened communication between patients and clinicians, and better patient health outcomes. However, several challenges were identified, including high drop-off rates in app usage, limited accessibility due to device restrictions, incompatibility between mHealth apps and EMR/EHR systems, increased clinical workload due to additional data integration, data accuracy concerns related to network connectivity, and data security issues.

Implications

The study findings underscore the importance of addressing technical challenges and optimizing management processes to facilitate the successful integration of mHealth apps into EMR/EHR systems, ultimately leading to improved health care outcomes. Specifically, having patient health information such as blood glucose levels, blood pressure measurements, chemotherapy-related adverse events, body weights, and seizure records captured and automatically incorporated into EMR/EHR systems can significantly improve data accuracy and enhance the efficiency of provider workflows [13,17,22,23,27-30]. This also highlights that apps capture or log data specific to clinical needs, potentially requiring a level of customization. Easy access to these patient health data, along with system-generated summaries, reduces the time clinicians spend verbally discussing symptoms and signs with patients and caregivers [12,13,15,20,25,27,30]. This allows more time to be dedicated to discussing disease management strategies. Additionally, the ability to review health data before patient encounters enables clinicians to prioritize those most in need of attention, thereby improving both efficiency and the quality of patient care [13,28-30].

Patient involvement in the development and implementation of mHealth apps and its integration into EMR/EHR systems ensures that mHealth apps and the integration are tailored to their needs, preferences, and concerns, thereby enhancing usability, engagement, and trust [18,23,24,26,31]. The integration of mHealth app data into EMR/EHR systems provides significant benefits in several key areas [12,14,16,21,29]. It facilitates increased patient education by delivering timely and relevant information about health conditions and treatments. This educational feature empowers patients to take an active role in managing their health. mHealth apps also support self-management by allowing patients to track their health metrics, set goals, and receive personalized feedback. Enhanced communication with health care providers through these apps and EMR/EHR systems further improves patient engagement and adherence to treatment plans. As a result, patients are more likely to follow prescribed treatments, make necessary lifestyle changes, and achieve critical clinical outcomes, leading to an overall improvement in quality of life.

Additionally, the integration of mHealth app data with EMR/EHR systems has demonstrated notable benefits in improving patient access to health care services during the COVID-19 pandemic [13]. By seamlessly sharing patient data with health care providers, this integration facilitated remote consultations, continuous monitoring, and timely interventions, which were crucial in managing patient health during a period of widespread disruption [32].

Literature Gap and Future Directions

The number of studies included in this review (n=19) is relatively small when compared with the vast array of mHealth apps currently in use. This suggests that research into the integration of mHealth apps with EMRs/EHRs and its impact on health care delivery and patient health outcomes is still in the early stages. Further comprehensive research is required to draw definitive conclusions in this area. Most studies examined third-party apps originally developed independently of the EMR/EHR ecosystem, with only a few focusing on native EMR/EHR-integrated apps (eg, Epic's modules) [22,24]. The small number of studies on native EMR/EHR-integrated apps resulted in no substantial differences in impact being observed between these 2 types of apps in this review. This finding highlights the need for further research to explicitly investigate how third-party apps and native EMR/EHR modules interact with EMR/EHR systems differently and their respective effects on health care outcomes.

As for observational period, most studies have a short-term period, less than 1 year [12-16,18,20,21,23,25-27,29,30], indicating the need for longer study periods to confirm the persistence of effects. The cross-sectional design and small sample sizes of some studies limit the ability to establish cause-and-effect relationships between mHealth app usage and outcomes [14,16,17,21,30]. Additionally, most studies were conducted in hospital settings [12-15,20,21,23,26-29,31], while chronic diseases, such as diabetes and hypertension, are often

XSL•FO
RenderX

managed in primary care settings [33]. Therefore, further studies are necessary to assess the integration and impact of mHealth apps in primary care settings. The focus on people with specific single chronic diseases [12-18,20-25,27-31] may limit the generalizability of research findings. Further research is needed to expand the scope to other health conditions, multimorbidity, or even disability, as these factors are often linked to poorer health outcomes and higher health care system costs [34,35].

Moreover, participants in most studies had good literacy [12-18,20-31] or were restricted to those proficient in the language [13,17,18,22-25,27,30]; this may reduce health equity for individuals with lower literacy levels or for people from linguistically diverse backgrounds. Given that people with lower literacy levels and those from culturally and linguistically diverse backgrounds often experience more disadvantage in accessing health care services [36,37] and are more likely to experience a poor outcome [38,39], health system reforms involving mHealth app integration with EMR/EHR systems should consider this and ensure that they have equitable access to health care. Efforts can be made to make mHealth services easily accessible to low-literacy individuals and to enhance individual health literacy through educational programs [40]. For people with low literacy, remote-monitoring devices that require little or no knowledge of mHealth app use from patients or carers may be a better option to capture their health data [41].

Finally, some studies excluded people who did not have a smartphone compatible with the app, who had moderate to severe cognitive impairment who owned a mobile phone, or who did not have internet access with their mobile phone [18,20,22,23,26,27], which may limit equitable access to mHealth apps. This exclusion could result in a digital divide, particularly affecting those without access to smartphones or other digital devices, such as older adults, people with lower socioeconomic status, or individuals in underserved communities. Expanding research to consider these populations is essential to ensure more inclusive health care reforms.

Limitations

There are some limitations. First, this scoping review exclusively focused on academic literature published in English. The exclusion of non-English publications may result in potential bias, as relevant studies in other languages could have been overlooked. Second, most included studies were conducted in high-income countries. This may limit the generalizability of the findings to low- and middle-income countries, where health care systems, digital infrastructure, and patient populations may differ significantly.

Conclusions

In summary, this scoping review provides the first comprehensive synthesis of published literature on the integration of mHealth app data into EMR/EHR systems and its impact on health care delivery and patient health outcomes. As the use of mHealth apps continues to rise, and people with diverse health conditions require increasingly personalized care, it is crucial that health care policy makers and administrators are equipped with evidence-based knowledge and guidance. Current literature remains limited, with included studies primarily focusing on populations with specific health conditions, short-term outcomes, and single hospital settings. Significant opportunities remain to build evidence based on the effectiveness of mHealth-EMR/EHR integration in improving long-term patient outcomes, such as rehospitalization, mortality, and cost, and assessing the scalability and sustainability of integration among more broader health conditions and disabilities across diverse health care settings.

Acknowledgments

The authors acknowledge funding support from The Ian Potter Foundation and the University of New South Wales.

Authors' Contributions

JL, SB, LA, and MK contributed to the conceptualization of the study. JL and SB carried out the screening and data extraction. JL conducted the formal analysis and wrote the original draft. MK supervised the study. All authors contributed to reviewing and editing the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string.
[DOCX File, 31 KB - mhealth_v13i1e66650_app1.docx]

Multimedia Appendix 2
Study characteristics, mobile health apps, and their integration to electronic medical record or electronic health record systems.
[DOCX File, 41 KB - mhealth_v13i1e66650_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Number of studies for each feature of mobile health apps and its integration into electronic medical record or electronic health record systems (note: detailed description of the features can be found in Multimedia Appendix 5.
[PNG File, 38 KB - mhealth_v13i1e66650_app3.png]

Multimedia Appendix 4
Description of each feature of mobile health apps and their integration into electronic medical record or electronic health record systems.
[DOCX File, 32 KB - mhealth_v13i1e66650_app4.docx]

Multimedia Appendix 5
Features of mobile health apps and their integration into electronic medical record or electronic health record systems.
[DOCX File, 34 KB - mhealth_v13i1e66650_app5.docx]

Multimedia Appendix 6
Study findings, benefits, and challenges for the use of mobile health apps and their integration into electronic medical record or electronic health record systems on the health care delivery and patient health outcomes.
[DOCX File, 44 KB - mhealth_v13i1e66650_app6.docx]

Checklist 1
PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File, 87 KB - mhealth_v13i1e66650_app7.docx]

References

1. Nacinovich M. Defining mHealth. J Commun Healthc 2011 Apr;4(1):1-3. [doi: 10.1179/175380611X12950033990296]
2. Ritchie J, Welch B. Categorization of third-party apps in electronic health record app marketplaces: systematic search and analysis. JMIR Med Inform 2020 May 29;8(5):e16980. [doi: 10.2196/16980] [Medline: 32469324]
3. Debon R, Coleone JD, Bellei EA, De Marchi ACB. Mobile health applications for chronic diseases: a systematic review of features for lifestyle improvement. Diabetes Metab Syndr 2019;13(4):2507-2512. [doi: 10.1016/j.dsx.2019.07.016] [Medline: 31405669]
4. Kustiawan TC, Nadhiroh SR, Ramli R, Butryee C. Use of mobile app to monitoring growth outcome of children: a systematic literature review. Digit Health 2022;8:20552076221138641. [doi: 10.1177/20552076221138641] [Medline: 36386243]
5. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006]
6. Sarubbi S, Rogante E, Erbuto D, et al. The effectiveness of mobile apps for monitoring and management of suicide crisis: a systematic review of the literature. J Clin Med 2022 Sep 23;11(19):5616. [doi: 10.3390/jcm11195616] [Medline: 36233484]
7. Mandl KD, Mandel JC, Murphy SN, et al. The SMART Platform: early experience enabling substitutable applications for electronic health records. J Am Med Inform Assoc 2012;19(4):597-603. [doi: 10.1136/amiajnl-2011-000622] [Medline: 22427539]
8. Ndlovu K, Mars M, Scott RE. Interoperability frameworks linking mHealth applications to electronic record systems. BMC Health Serv Res 2021 May 13;21(1):459. [doi: 10.1186/s12913-021-06473-6] [Medline: 33985495]
9. Garets D, Davis M. Electronic medical records vs electronic health records: yes, there is a difference. 2006. URL: http://sirroteinfo.ezyro.com/wp-content/uploads/2020/06/WP_EMR_EHR.pdf?i=1 [accessed 2025-05-26]
10. Cavanagh R, Gerson SM, Gleason A, Mackey R, Ciulla R. Competencies needed for behavioral health professionals to integrate digital health technologies into clinical care: a rapid review. J Technol Behav Sci 2022 Feb 17;8(4):1-14. [doi: 10.1007/s41347-022-00242-w] [Medline: 35194561]
11. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018]
12. Lee EY, Cha SA, Yun JS, et al. Efficacy of personalized diabetes self-care using an electronic medical record-integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial. J Med Internet Res 2022 Jul 28;24(7):e37430. [doi: 10.2196/37430] [Medline: 35900817]
13. Crossen S, Romero C, Reggiardo A, Michel J, Glaser N. Feasibility and impact of remote glucose monitoring among patients with newly diagnosed type 1 diabetes: single-center pilot study. JMIR Diabetes 2022 Jan 17;7(1):e33639. [doi: 10.2196/33639] [Medline: 35037887]
14. Kim JW, Ryu B, Cho S, et al. Impact of personal health records and wearables on health outcomes and patient response: three-arm randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 4;7(1):e12070. [doi: 10.2196/12070] [Medline: 30609978]

XSL•FO
RenderX

15. Guo X, Gu X, Jiang J, et al. A hospital-community-family-based telehealth program for patients with chronic heart failure: single-arm, prospective feasibility study. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13229. [doi: 10.2196/13229] [Medline: 31833835]
16. Ryu B, Kim N, Heo E, et al. Impact of an Electronic health record-integrated personal health record on patient participation in health care: development and randomized controlled trial of MyHealthKeeper. J Med Internet Res 2017 Dec 7;19(12):e401. [doi: 10.2196/jmir.8867] [Medline: 29217503]
17. Solomon DH, Altwies H, Santacroce L, et al. A mobile health application integrated in the electronic health record for rheumatoid arthritis patient-reported outcomes: a controlled interrupted time-series analysis of impact on visit efficiency. Arthritis Rheumatol 2024 May;76(5):677-683. [doi: 10.1002/art.42774] [Medline: 38087859]
18. Weatherly J, Kishnani S, Aye T. Challenges with patient adoption of automated integration of blood glucose meter data in the electronic health record. Diabetes Technol Ther 2019 Nov;21(11):671-674. [doi: 10.1089/dia.2019.0178] [Medline: 31335195]
19. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
20. Masiero M, Filipponi C, Fragale E, et al. Support for chronic pain management for breast cancer survivors through novel digital health ecosystems: pilot usability study of the PainRELife mobile app. JMIR Form Res 2024 Feb 2;8:e51021. [doi: 10.2196/51021] [Medline: 38306176]
21. Huang J, Chan SC, Ko S, et al. Associations between adoption of eHealth management module and optimal control of HbA1c in diabetes patients. NPJ Digit Med 2023 Apr 13;6(1):67. [doi: 10.1038/s41746-023-00807-w] [Medline: 37055503]
22. Young NP, Ridgeway JL, Haddad TC, et al. Feasibility and usability of a mobile app-based interactive care plan for migraine in a community neurology practice: development and pilot implementation study. JMIR Form Res 2023 Oct 5;7:e48372. [doi: 10.2196/48372] [Medline: 37796560]
23. Morgenthaler TI, Kolla BP, Anderson SE, et al. Development and acceptability of a mobile health application integrated with the electronic heath record for treatment of chronic insomnia disorder. J Clin Sleep Med 2022 Dec 1;18(12):2785-2792. [doi: 10.5664/jcsm.10218] [Medline: 35959953]
24. Stan DL, Inselman JW, Ridgeway JL, et al. Pilot implementation to assess the feasibility and care team impact of an app-based interactive care plan to remotely monitor breast cancer survivors. J Cancer Surviv 2022 Feb;16(1):13-23. [doi: 10.1007/s11764-021-01136-1] [Medline: 35107791]
25. Agnihothri S, Cui L, Rajan B, Banerjee A, Ramanujan R. Mobile health application usage and quality of care at a hypertension clinic: an observational cohort study. J Hypertens 2021 Nov 1;39(11):2265-2271. [doi: 10.1097/HJH.0000000000002909] [Medline: 34074997]
26. Choi SA, Lim K, Baek H, et al. Impact of mobile health application on data collection and self-management of epilepsy. Epilepsy Behav 2021 Jun;119:107982. [doi: 10.1016/j.yebeh.2021.107982] [Medline: 33946011]
27. Lewis S, Ainsworth J, Sanders C, et al. Smartphone-enhanced symptom management in psychosis: open, randomized controlled trial. J Med Internet Res 2020 Aug 13;22(8):e17019. [doi: 10.2196/17019] [Medline: 32788150]
28. Bae WK, Kwon J, Lee HW, et al. Feasibility and accessibility of electronic patient-reported outcome measures using a smartphone during routine chemotherapy: a pilot study. Support Care Cancer 2018 Nov;26(11):3721-3728. [doi: 10.1007/s00520-018-4232-z] [Medline: 29732483]
29. Cho SW, Wee JH, Yoo S, et al. Effect of lifestyle modification using a smartphone application on obesity with obstructive sleep apnea: a short-term, randomized controlled study. Clin Exp Otorhinolaryngol 2018 Sep;11(3):192-198. [doi: 10.21053/ceo.2017.01284] [Medline: 29374961]
30. Kumar RB, Goren ND, Stark DE, Wall DP, Longhurst CA. Automated integration of continuous glucose monitor data in the electronic health record using consumer technology. J Am Med Inform Assoc 2016 May;23(3):532-537. [doi: 10.1093/jamia/ocv206] [Medline: 27018263]
31. Kim HS, Choi W, Baek EK, et al. Efficacy of the smartphone-based glucose management application stratified by user satisfaction. Diabetes Metab J 2014 Jun;38(3):204-210. [doi: 10.4093/dmj.2014.38.3.204] [Medline: 25003074]
32. Silva CRDV, Lopes RH, de Goes Bay O Jr, et al. Digital health opportunities to improve primary health care in the context of COVID-19: scoping review. JMIR Hum Factors 2022 May 31;9(2):e35380. [doi: 10.2196/35380] [Medline: 35319466]
33. Rothman AA, Wagner EH. Chronic illness management: what is the role of primary care? Ann Intern Med 2003 Feb 4;138(3):256-261. [doi: 10.7326/0003-4819-138-3-200302040-00034] [Medline: 12558376]
34. Palladino R, Tayu Lee J, Ashworth M, Triassi M, Millett C. Associations between multimorbidity, healthcare utilisation and health status: evidence from 16 European countries. Age Ageing 2016 May;45(3):431-435. [doi: 10.1093/ageing/afw044] [Medline: 27013499]
35. McPhail SM. Multimorbidity in chronic disease: impact on health care resources and costs. Risk Manag Healthc Policy 2016;9:143-156. [doi: 10.2147/RMHP.S97248] [Medline: 27462182]
36. Easton P, Entwistle VA, Williams B. Health in the "hidden population" of people with low literacy. A systematic review of the literature. BMC Public Health 2010 Aug 5;10(1):459. [doi: 10.1186/1471-2458-10-459] [Medline: 20687946]

XSL•FO
RenderX

37. Khatri RB, Assefa Y. Access to health services among culturally and linguistically diverse populations in the Australian universal health care system: issues and challenges. BMC Public Health 2022 May 3;22(1):880. [doi: 10.1186/s12889-022-13256-z] [Medline: 35505307]
38. DeWalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes. J Gen Intern Med 2004 Dec;19(12):1228-1239. [doi: 10.1111/j.1525-1497.2004.40153.x]
39. Priest N, Baxter J, Hayes L. Social and emotional outcomes of Australian children from Indigenous and culturally and linguistically diverse backgrounds. Aust N Z J Public Health 2012 Apr;36(2):183-190. [doi: 10.1111/j.1753-6405.2011.00803.x]
40. Kemp E, Trigg J, Beatty L, et al. Health literacy, digital health literacy and the implementation of digital health technologies in cancer care: the need for a strategic approach. Health Prom J Aust 2021 Feb;32(S1):104-114. [doi: 10.1002/hpja.387]
41. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The future of wearable technologies and remote monitoring in health care. Am Soc Clin Oncol Educ Book 2019 Jan;39:115-121. [doi: 10.1200/EDBK_238919] [Medline: 31099626]

Abbreviations

EHR: electronic health record
EMR: electronic medical record
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

Edited by L Buis; submitted 30.09.24; peer-reviewed by K Bochtler, R Rudin; revised version received 28.04.25; accepted 28.04.25; published 23.06.25.

Please cite as:
Lin J, Bates SM, Allen LN, Wright M, Mao L, Kidd M
Integrating Mobile Health App Data Into Electronic Medical or Health Record Systems and Its Impact on Health Care Delivery and Patient Health Outcomes: Scoping Review
JMIR Mhealth Uhealth 2025;13:e66650
URL: https://mhealth.jmir.org/2025/1/e66650
doi:10.2196/66650

© Jialing Lin, Shona Marie Bates, Luke N Allen, Michael Wright, Limin Mao, Michael Kidd. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Digital Health Interventions on Sedentary Behavior Among Patients With Chronic Diseases: Systematic Review and Meta-Analysis

Yan Zhang[1,2], MS, RN; Fei Wan Ngai[2], PhD; Qingling Yang[2], MM; Yao Jie Xie[2,3], PhD

[1]Department of Cardiology, Affiliated Hospital of Qingdao University, Qingdao, China

[2]School of Nursing, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, Hong Kong, China

[3]Research Centre for Chinese Medicine Innovation, The Hong Kong Polytechnic University, Hong Kong, China

Corresponding Author:
Yao Jie Xie, PhD
School of Nursing, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, Hong Kong, China

Abstract

Background: Individuals with chronic diseases commonly engage in a sedentary lifestyle, which may exacerbate poor disease progression and increase the burden of care. Digital health interventions have been broadly used in promoting healthy lifestyles in recent decades, while their effectiveness on sedentary behavior (SB) remains inconsistent and inconclusive.

Objective: This review aimed to evaluate the effectiveness of digital health interventions in reducing SB among patients with chronic diseases.

Methods: PubMed, Embase, Scopus, Web of Science, CINAHL Complete, Cochrane Library, and ACM Digital Library were searched for randomized controlled trials published from January 2000 to October 2023. Two researchers independently screened studies and evaluated study quality. The revised Cochrane risk-of-bias tool was used to assess the risk of bias. Mean differences (MDs) were calculated for intervention effect comparison.

Results: Twenty-six trials were selected and 3800 participants were included. The mean age was 57.32 (SD 9.91) years. The typical chronic diseases reported in the studies included obesity (n=6), arthritis (n=5), coronary artery disease (n=4), cancer (n=4), type 2 diabetes mellitus (n=3), metabolic syndrome (n=2), and stroke (n=2). Phone, web, and activity trackers were 3 digital technologies adopted in the interventions and they were used in combination in most studies (18/26, 69.2%). The functions included facilitating self-monitoring of SB, reminding interruption of long undisturbed sitting, and promoting goal attainment. Approaches targeting SB reduction included standing (n=6), walking (n=9), light physical activity (n=5), moderate to vigorous physical activity (n=4), screen time limitation (n=2), and contextual-related activities based on patients' preference (n=4). The majority (80.8%) of studies had a low to moderate risk of bias. Meta-analysis revealed significant decreases in overall sitting time (MD −30.80; 95% CI −49.79 to−11.82; I^2=65%; P=.001), pre-post sitting time changes (MD −50.28; 95% CI −92.99 to −7.57; I^2=92%; P=.02), and SB proportions (MD −4.65%; 95% CI −7.02 to −2.28; I^2=20%; P<.001) after digital health interventions, compared with nondigital interventions such as usual care, wait-list, or other active controls, with a small effect size (Cohen d=−0.27 to −0.47). No significant differences in the length of sedentary bouts and breaks were found. Subgroup analyses showed that studies with objective SB measurements and those younger than 65 years had significant reductions in sitting time.

Conclusions: Digital health interventions significantly reduced the SB among patients with chronic illness. More research with rigorous design to promote a long-term decrease in sitting time, differentiate primary and compensatory SB reductions, and explore the underlying mechanisms is needed.

Trial Registration: PROSPERO CRD42023477958; https://tinyurl.com/mr34te36

(JMIR Mhealth Uhealth 2025;13:e59943) doi:10.2196/59943

KEYWORDS

digital health intervention; sedentary behavior; physical activity; chronic diseases; PRISMA

Introduction

Chronic diseases, including cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes, contributed to 71% of all causes of death, imposing an enormous and growing global burden on health care systems and financial expenditure [1]. Individuals diagnosed with chronic diseases required consistent engagement in physical activity (PA) to prevent disease

XSL•FO
RenderX

progression and enhance long-term quality of life [2]. However, those patients were always engaged in substantially long sedentary lifestyle because of limited physical function, fatigue, and insufficient exercise endurance.

Sedentary behavior (SB), defined by the Sedentary Behavior Research Network and the American Heart Association, means any waking behaviors characterized by an energy expenditure of ≤1.5 metabolic equivalents (METs) while in a sitting or reclining posture [3,4]. Patients with chronic diseases continue to demonstrate a high level of SB. For instance, a study investigating 131 patients with coronary artery disease (CAD) demonstrated an average sitting time of 10.4 hours per day [5]. Patients with acute coronary syndrome continued to exhibit prolonged sitting of approximately 9 hours per day even after participating in cardiac rehabilitation [6]. Another study indicated that patients with rheumatoid arthritis (RA) spent 71% - 92% of their waking hours in SB [7]. Among 2497 individuals with type 2 diabetes mellitus (T2DM), they allocated an average of 64% waking hours in SB [8]. Patients with stroke also reported 9.22 hours per day sitting time in the first week after discharge in a prospective cohort study [9]. Similarly, a national survey among 741 survivors from cancer found that the patients spent at least 8 hours per day in sitting [10]. Such long-time SB accelerated the negative progression of diseases in those patients [11-13].

The detrimental effects of prolonged SB on chronic disease progression were attributed to various underlying mechanisms, including increased central arterial stiffness [14], impaired vascular function and structure, reduced antegrade blood flow and shear rate, increased insulin resistance, and heightened oxidative stress and apoptosis [15,16]. Those functional deterioration resulted in suboptimal control of cardiometabolic indicators, subsequently increased occurrence of complications, and elevated risk of hospital readmission and mortality [4,17-19]. Therefore, reducing SB, typically in patients with chronic diseases, is a public health priority.

In recent decades, digital health, or the use of digital technologies for health, has gained significant prominence in addressing various health needs [20]. In 2019, the World Health Organization (WHO) announced the Global Strategy on Digital Health 2020 - 2025, which emphasized the adoption of digital health in health care [21]. Digital health interventions have great potential for behavioral change without on-site visits, no time constraints, and few travel costs, which facilitate flexible, efficient, and cost-effective interventions [22]. Currently, many researchers have applied digital health interventions for SB reduction. However, the results of previous studies were still inconsistent and inconclusive. For example, a meta-analysis indicated that wearable activity trackers were associated with a significant reduction of 35.46 minutes per day in sedentary time among hospitalized patients, but this finding was derived from only 2 studies included in this review [23]. Additionally, 2 other systematic reviews reported conflicting results regarding the effectiveness of smartphone-based interventions in interrupting SB in older adults, whereas this evidence was also obtained from a limited number of studies (n≤2), and the definition of SB was different [24,25].

Given the rapid development in digital technologies and variation of intervention patterns, as well as the importance of improving SB among patients with chronic diseases, it is crucial to update and comprehensively summarize the evidence to incorporate digital solutions for this vulnerable population. Thereby, this systematic review and meta-analysis aimed to search and synthesize available evidence regarding the effectiveness of digital health interventions in improving SB among patients with chronic diseases.

Methods

This systematic review and meta-analysis was registered in the PROSPERO (no. CRD42023477958). The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement was followed in conducting the review (Checklist 1) [26].

Inclusion and Exclusion Criteria

The PICOS (Population, Intervention, Control, Outcome, and Study) framework was used to select eligible studies.

Population

The population included adults diagnosed with chronic diseases in terms of cardiovascular disease, chronic respiratory disease, cancer, diabetes, metabolic syndrome (MetS), obesity, RA, stroke, and other chronic conditions defined by the WHO [27].

Intervention

Digital health intervention is a discrete functionality of digital technology that is applied to achieve health objectives according to WHO guideline [20]. Therefore, interventions delivered using any type of digital technologies, consisting of mobile phones, web, software apps, wearable trackers, emails, or other digital technologies, were included. Studies that exclusively used wearable trackers to measure SB were excluded. The interventions should directly target SB reduction or PA levels increase, which may indirectly result in compensatory changes in SB [28].

Control

This included any comparison without digital technologies, including usual care, wait-list, active control, or blank control.

Outcome

The definition of SB from the Sedentary Behavior Research Network was adopted [29]. Any SB-related outcomes, including overall sitting time, pre-post sitting time changes, SB proportion, sedentary bouts, and breaks of prolonged sitting, which were assessed at baseline and end point using objective or subjective measures, either as primary or secondary outcomes, were involved.

Study

The study types included randomized controlled trials (RCTs). Pilot RCTs, feasibility studies, protocol papers, brief reports of RCTs, and studies with a sample size <10 were excluded.

XSL•FO
RenderX

Search Methods

The search strategy was developed by 1 author (YZ) and then reviewed and finalized by 2 experts (YJX and FWN) (Multimedia Appendix 1). Two keywords were first defined to develop MeSH (Medical Subject Headings) terms in the search strategy: SB and digital technology. To find patients with any potential chronic illness, no keyword, in particular, for any chronic disease was used in the search strategy. We conducted a systematic search to retrieve all papers focusing on SB interventions using digital health technologies. Two independent authors (YZ and QLY) manually screened these papers and identified those specifically targeting patients with chronic diseases, as defined by the *International Classification of Diseases, Eleventh Revision* (*ICD-11*). At last, a comprehensive search was carried out in PubMed, Embase, Scopus, Web of Science, CINAHL Complete, Cochrane Library, and ACM Digital Library. The papers in English published from 2000 onward were included, which aligned with the first release of the WHO's document on the approach to digital health strategies [21].

Study Selection and Data Extraction

Duplicate studies were identified and removed using the NoteExpress software (Beijing Aegean Software). The remaining studies were screened independently by 2 authors (YZ and QLY) in a sequential manner of title, abstract, and full-text screening. Conflicts were settled by consulting a third author (YJX). A standardized data extraction form was used to collect the data on study characteristics (authors, publication years, participants' demographic details, and dropout rates), intervention details (content, type of digital technology, and duration of the intervention), and outcomes (SB-related outcomes and other primary and secondary outcomes). The data extraction process was conducted independently by 2 authors (YZ and QLY). Any discrepancies in data interpretation were discussed and adjudicated by a third author (YJX).

Study Quality Assessment

The risk of bias was assessed using the "Revised Cochrane risk-of-bias tool for randomized trials" (RoB2) [30]. Five domains with signaling questions were evaluated, including the randomization process, the effect of assignment and adherence to the intervention, missing outcome data, outcome measurement, and the selection of reported results. Each criterion was assessed and categorized as "low risk," "some concern," or "high risk" of bias. An algorithm based on these 5 domains was used to determine the overall bias. Two independent reviewers (YZ and QLY) evaluated the selected papers, and any disagreements were resolved by a third party (YJX).

Statistical Analysis

The data analysis was conducted using Review Manager 5.3. Twenty-two studies used objective activity trackers to quantify SB-related outcomes, enabling the use of mean differences (MDs) to effectively highlight absolute differences in the mean values of these outcomes [30,31]. Cohen *d* was used to assess the effect size, with $d>0.8$ indicating a large effect, 0.5 - 0.8 a medium effect, and 0.2 - 0.5 a small effect [32]. For studies with multiple assessments of SB outcomes in 1 study, the final assessment point was selected. Objective SB outcomes (eg, daily sedentary time measured by accelerometer) were prioritized to minimize measurement heterogeneity. The heterogeneity among studies was assessed using the I^2, categorized as low ($I^2<25\%$), medium ($25\% \leq I^2<75\%$), or high ($I^2 \geq 75\%$) [33]. Forest plots were used to visually compare the estimated effects and CIs. Funnel plots were used to evaluate publication bias when there were at least 10 studies in the meta-analysis [30]. Egger test was conducted using Stata (version 12; Stata Corp). Sensitivity analyses with leave-one-out principle were performed to assess the robustness of the main results. In cases where quantitative synthesis was not appropriate, a narrative synthesis was conducted.

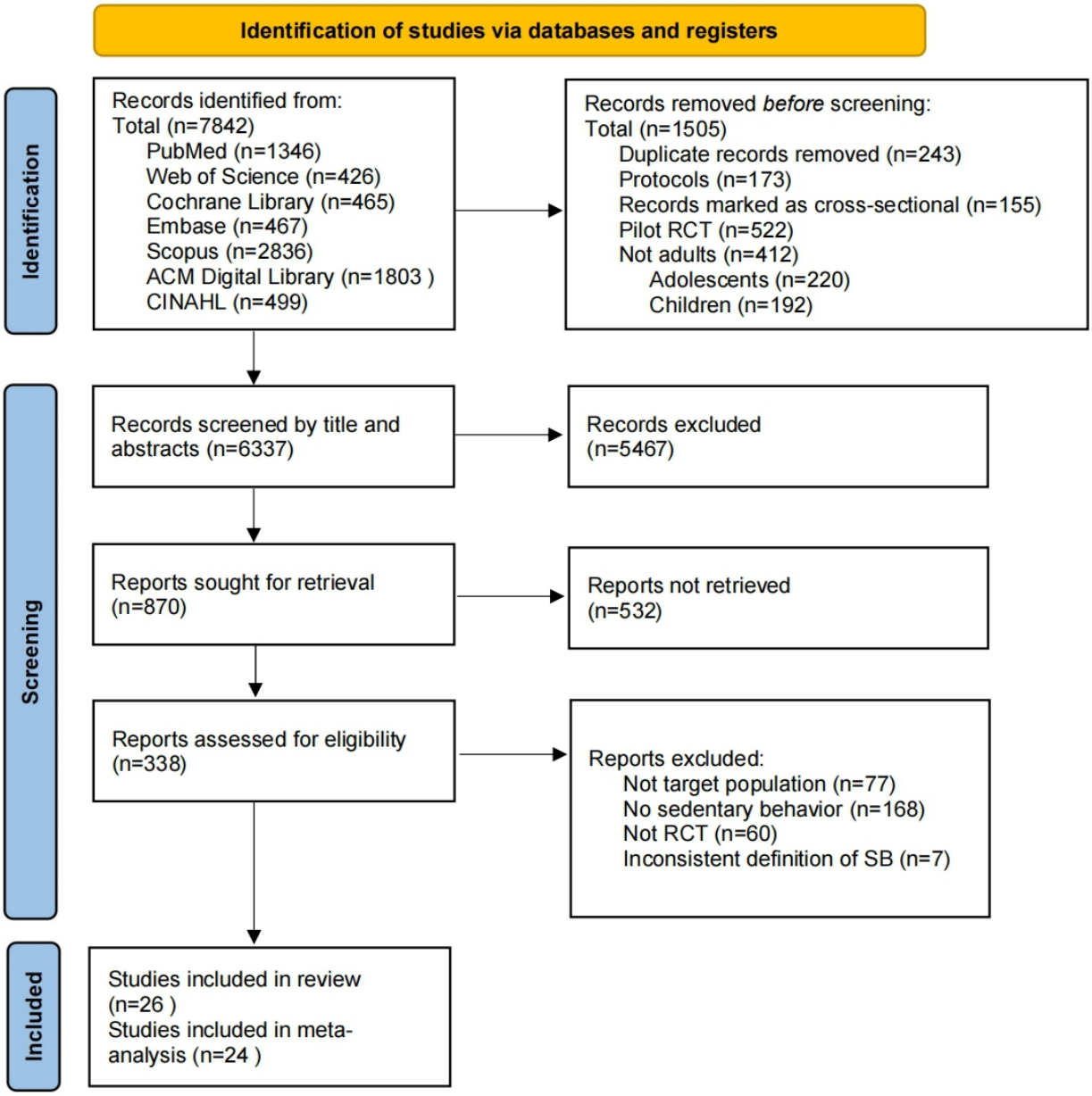
Results

Study Selection

A total of 7842 studies were identified after the systematic search. After removing duplicates, 6337 papers were screened by title and abstract. Ultimately, 338 papers were retrieved and assessed for eligibility through full text review. Of these, 312 papers were excluded due to reasons such as nontarget population, inappropriate study design, or inconsistent definitions of SB, such as TV watching, physical inactivity, or steps. The excluded studies are listed in Multimedia Appendix 1. Figure 1 shows the selection process. Among the included studies, 4 studies were based on 2 separately registered RCTs with different sample sizes or follow-up duration [34-37], and all these studies were then included. While for the other 3 RCTs based on the same study [38-40], only 1 was selected as representative [38], because it closely matched the registered content, also published earlier, and directly focused on SB interventions. In total, 26 studies were identified. Details of the selected papers are summarized in Table S1 in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the literature search and selection process. RCT: randomized controlled trial; SB: sedentary behavior.



Study Characteristics

Study Settings, Design, and Population Characteristics

All the included studies were published in English between 2011 and 2023, with the majority in Western countries (16/26, 61.5%). The Netherlands [41-44] and Australia [45-48] had the same and highest number of studies (n=4). There were 2 three-arm RCTs [49,50] and 1 four-arm RCT [28], and all others were 2-arm RCTs. The sample size varied from 51 to 622, with a total of 3800 participants involved. The average age of the participants was 57.32 years (SD 9.91). There were more women (n=2042) than men (n=1758), while 2 studies specifically focused on women [46,51]. The chronic diseases included obesity (n=6) [28,38,41,42,50,52], RA (n=5) [36,37,44,53,54], CAD (n=4) [43,48,55,56], cancer (n=4) [45-47,51], T2DM (n=3) [49,57,58], MetS (n=2) [59,60], and stroke (n=2) [34,35]. All the included studies used digital-assisted interventions using various technologies at different levels. Most control groups were assigned to usual care (n=13), maintained their usual lifestyle (n=5), or were placed on a wait-list control (n=4).

SB-Related Outcomes

Sixteen studies designated SB as the primary outcome [28,34-37,42,43,45-47,49,51,52,58-60], while the remaining 10 RCTs identified PA as the primary outcome [44,53-57,60-63]. SB-related outcomes included the overall sitting time (n=16), pre-post sitting time changes (n=8), SB proportion among all activities (n=4), sedentary bouts (n=7), and breaks of sedentary time (n=6). The overall sitting time (minutes per day) was the volume of sitting time accrued per day, which was the most common indicator in SB assessment [64]. The pre-post sitting time changes (minutes per day) were calculated as the overall sitting time after intervention minus the baseline sitting time, also called sedentary time reduction [43]. SB proportion (%) was the percentage of time spent

XSL•FO
RenderX

performing the activities that require ≤1.5 MET of tasks, after the subtraction of sleeping time [34,35]. Essentially, it also referred to the SB percentage among all activities, including SB (MET ≤1.5), light physical activity (LPA) (1.5 < MET ≤3), and moderate to vigorous physical activity (MVPA) (MET >3.0) [28].

Sedentary bout was defined as a continuous period of uninterrupted sedentary time. However, there seemed to be a lack of consensus on the criteria used to determine the sedentary bouts. Six studies used various thresholds for defining sedentary bouts, including 30 minutes [43], 20 minutes [46,47,53,54], and 10 minutes [49]. And 1 study reported the average duration of all sedentary bouts [48]. Six studies presented data on breaks of sedentary time (frequency of interrupted sedentary time per day), using the number of breaks [38,48,49,52], and changes in the number of breaks from baseline to end point [36,37]. However, only 1 study reported how to determine the "break" clearly, stating that a break was defined as a transition from SB (<99 counts per minute, measured by accelerometer; ActiGraph) to any other level of PA (>99 counts per minute) between 2 sedentary bouts (>10 consecutive minutes) [49].

In terms of measurement of SB, most of the researchers (22/26, 84.6%) used objective activity trackers. Four studies adopted subjective methods, including the International Physical Activity Questionnaire [56,58], the Activity Questionnaire for Adolescents and Adults [42], and telephone interviews [45]. Only 2 studies used both an accelerometer and a questionnaire [51,58].

Characteristics of Digital Health Interventions

The intervention primarily targeting SB encompassed screen time limitation (n=2) [34,35], standing (n=6) [36,38,43,46,52,59], walking (n=9) [36,43,45-47,49,51,52,60], LPA (n=5) [37,38,45,51,59], and MVPA (n=5) [28,38,42,47,60]. The majority of these trials (11/16, 68.8%) adopted the combination of these approaches. Regarding SB intervention, the context in which the SB occurred included leisure time, household tasks, transportation, and occupational settings [65]. Due to this context-related characteristic of SB, 4 RCTs explicitly proposed that participants could select the methods and set goals of interrupting SB based on their preference, including using standing desks, standing during phone calls, taking stairs instead of elevators, home-cleaning activities, or other kinds of LPA [36,38,43,59]. Six studies reported the compensatory SB reduction by increasing PA level (without addressing SB directly) [28,44,50,55-57], which included LPA/MVPA [50,55,57], cardiac rehabilitation [56], and free choice of any kinds of exercise, such as gardening, cycling [44], endurance, and strength training [28]. Motivational counseling or health coaching was also frequently used (8/16, 50%) to assist participants in increasing awareness of the detrimental effects of SB, establishing and revising goals, or receiving personalized prescriptions based on their preferences [36,37,45,46,51,52,58,59].

More than one-third of the studies (10/26, 38.5%) adopted a theoretic framework in intervention design and implementation. Three of them applied the Social Cognitive Theory [36,56,60], incorporating pivotal constructs such as self-efficacy and outcome expectations. This theory emphasized optimization of cognitive, behavioral aspects (eg, self-efficacy and skill mastery), and environmental influences (eg, peer support and barriers to action) to facilitate behavioral modifications. Two studies applied the Self-regulation Theory, which reinforced an individual's attitude and behavioral intention to change. Moreover, the discrepancy between a person's goal and the actual situation also served as a catalyst for behavioral transformations [42,49]. Other studies were based on the principle of Self-efficacy Theory [37], Theory of Planned Behavior [42], Acceptance Commitment Therapy [45], intervention mapping adaption framework [43], or Cognitive Behavioral Therapy [58]. In summary, these theories were used to facilitate the reduction of SB or enhancement of PA by influencing individuals' attitudes, beliefs, motivations, and cognitive processes. The intervention duration ranged from a minimum of 6 weeks to a maximum of 24 months, with a median duration of 6 months. Three months accounted for the highest proportion (7/26, 26.9%).

Digital Technologies Used in the Interventions

Three kinds of digital technologies were adopted in the intervention, including phone (phone calls, text messages, and apps), web (email and web-based messages), and activity trackers (accelerometer, pedometer, and heart rate monitor). Two studies used all 3 types of digital technologies [53,54], while 16 RCTs used 2 kinds [28,38,41,43,44,46-52,56,57,59,60] and 8 studies used 1 type such as phone calls (n=5), text messages (n=2), and web (n=1) [34-37,42,45,55,58]. Phone-based digital health interventions were the most prevalent (22/26, 84.6%), including phone calls and short text messages, which were applied to remind participants to adhere to the interventions. Two studies used common phone-based apps, such as WhatsApp [57] and WeChat [56], for liaison with participants, while others developed tailored ones, such as Mijn Actieplan [49] and Heathesteps [60], to support self-monitoring, reminding, and goal attaining. One study stated that they used a virtual coach in the smartphone app to provide personalized prescriptions and help participants maintain their goals [60]. However, details were limited even in the protocol [66].

Web-based interventions accounted for 26.92% (7/26) in all studies, and more than half of the interventions (4/7, 57.10%) were in combination with phones to facilitate the goal attainment process or disseminate supportive information [49,54,56,60]. It also served as a management platform for researchers to access participants' SB status in a real-time manner, which enabled timely feedback, information sharing, and encouragement to attain the designed goals [55].

Half of the studies (13/26, 50%) used activity trackers, including wearable accelerometer (wrist, hip, or pocket-worn) [38,41,43,46-48,52-54], pedometer [51], and heart rate monitor [51], for data collection, self-monitoring, and safety assurance. Two of them implemented alerts in the accelerometer using vibrotactile feedback [43] or visual "move" display with an audible beep [46], which directly and efficiently reminded patients about their unconscious long-time sitting. Six studies synchronized activity trackers with phone-based apps via Bluetooth or GPS automatically, enabling researchers or health

XSL•FO
RenderX

professionals to provide timely feedback or help participants to self-monitor their behavior, such as SB or PA [38,43,46,48,53,59].

Risk of Bias

The 80.8% (21/26) of the studies had a low to moderate risk of bias [28,35-38,41,43-47,50,51,53-60]. Domains assessed as having a high risk of bias were attributed to the missing outcome data. Three papers had no description of the allocation concealment [28,50,58]. One study had unbalanced baseline data regarding marital status and employment status (P<.05) [57]. One study reported longer sitting time in the intervention group than in the control group (9.8 vs 8.8 hours per day) at baseline [37]. No study adopted blinding because it was not possible due to the nature of the intervention. The majority of the studies used the intention-to-treat analysis, while 3 studies applied the on-protocol analysis [28,50,59]. For the domain of missing outcome data, only 8 of the studies reported a data availability of ≥95% [37,38,43,46,47,53,54,58]. Most of the studies were in line with existing published protocols, except for 3 papers. Two of them decreased sample size owing to the impact of COVID-19 [34,48], and 1 modified the recruitment strategy as there were not enough individuals from the target population [49]. Details regarding the risk of bias for each study and the summary could be found in Multimedia Appendix 3.

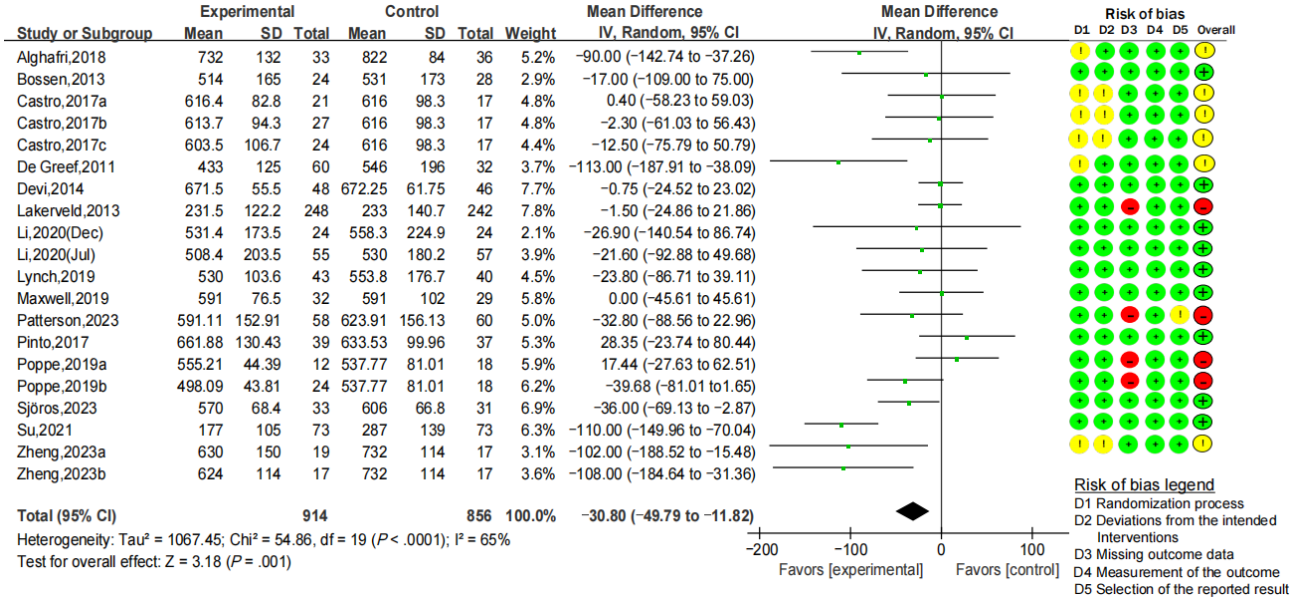
Meta-Analysis

Among the 26 included studies, 24 RCTs were included in the meta-analysis. Rest of the 2 studies just provided limited information and were therefore used only for narrative analysis [59,60].

Effects of Digital Health Interventions on Reducing Overall Sitting Time

Sixteen RCTs reported the overall sitting time (minutes per day), including 914 participants in the intervention group and 856 in the control group. The overall sitting time was significantly decreased with a pooled MD of 30.80 minutes per day (95% CI −49.79 to −11.82; I^2=65%; P=.001) after the interventions (Figure 2), with a small effect size (Cohen d=−0.27; 95% CI −0.44 to −0.11; Z=3.25; P=.001). The study weight, MD of overall sitting time, and mean age of the participants were shown in a grouped bubble plot (Multimedia Appendix 4).

Figure 2. Forest plot for pooled mean difference of overall sitting time (minutes per day) [28,38,42,44,46-51,53-58].



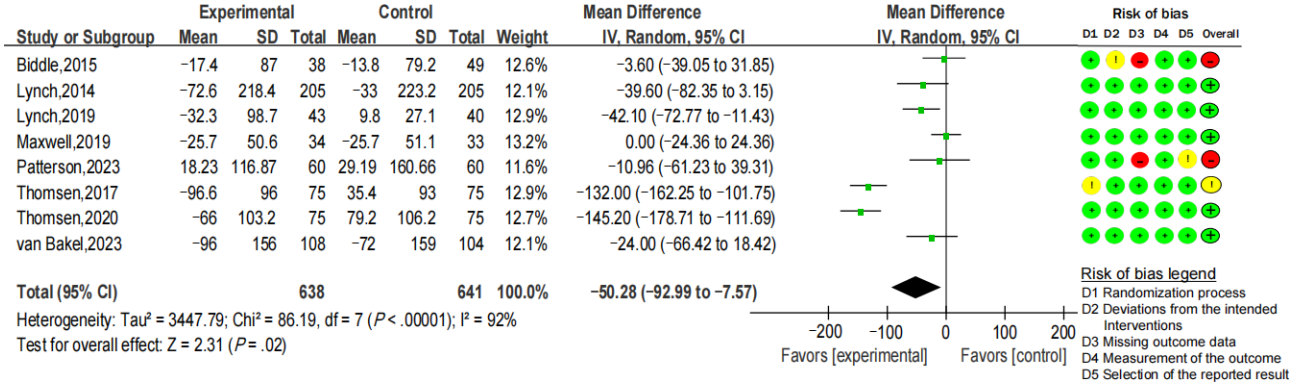
Subgroup analyses were conducted based on disease classification, age group, theoretical framework, primary or compensatory SB reduction, intervention duration, and objective or subjective measurement of SB. There was no significant reduction in overall sitting time (minutes per day) among patients with specific kinds of chronic diseases, including obesity, RA, CAD, cancer, T2DM, MetS, and stroke. As for age groups, only those aged 65 years and younger showed a noticeable decline in sitting time (MD −35.74 minutes per day; 95% CI −57.20 to −14.28; I^2=66%; P=.001). Both theory-based and non–theory-based interventions had a significant effect on reducing sitting time, while the theory-based interventions demonstrated a greater reduction (MD −52.48 vs −17.20 minutes per day). A meta-analysis was conducted separately for studies primarily focusing on SB or PA (compensatory SB reduction). The results showed that compensatory SB exhibited a significant decrease (MD −47.58 minutes per day; 95% CI −84.94 to −10.22; I^2=77%; P=.01). Regarding the duration of interventions, both the short duration (≤3 months) and the long duration (>3 months) significantly decreased sitting time (MD −55.89 minutes per day; MD −18.86 minutes per day). One study reported an average sedentary time of 18.8 hours based on objective measurement, which was clinically unreasonable and might include sleeping time [58]. Therefore, the subjective data on sedentary time was extracted in this study. The results showed that only those objectively measured sedentary time reduced 22.20 minutes per day significantly (95% CI −39.07 to −5.33; I^2=42%; P=.01). Details of the subgroup analyses are shown in Multimedia Appendix 5.

XSL•FO
RenderX

Pre-Post Sitting Time Changes After Digital Health Interventions

As depicted in Figure 3, 8 RCTs reported pre-post sitting time changes, involving 638 participants in the intervention group and 641 participants in the control group. The results showed significant changes in sitting time, with an average decrease of 50.28 minutes per day (95% CI −92.99 to −7.57; I^2=92%; *P*=.02), and a small effect size (Cohen *d*=−0.47, 95% CI −0.86 to −0.09; *Z*=2.40; *P*=.02).

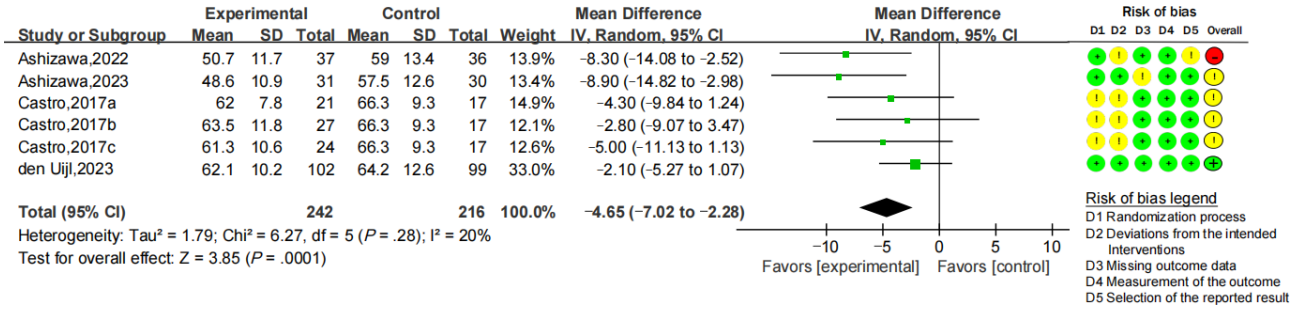
Figure 3. Forest plot for the mean difference of sitting time changes (minutes per day) after intervention [36,37,43,45-48,52].



SB Proportion of All Activities After Interventions

Four RCTs reported SB proportion results. The forest plot showed a pooled MD of 4.65% reduction in favor of the digital health interventions (95% CI −7.02 to −2.28; I^2=20%; *P*=.0001) (Figure 4), with a small effect size (Cohen *d*=−0.40; 95% CI −0.61 to −0.20; *Z*=3.91; *P*<.001).

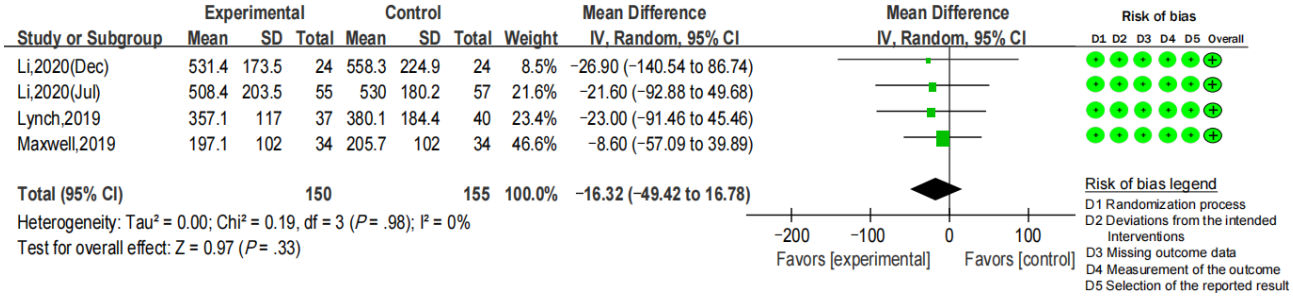
Figure 4. Forest plot for the mean difference of sedentary behavior proportion among all activities (%) [28,34,35,41].



Effects of Digital Health Interventions on Sedentary Bouts and Breaks of Sedentary Time

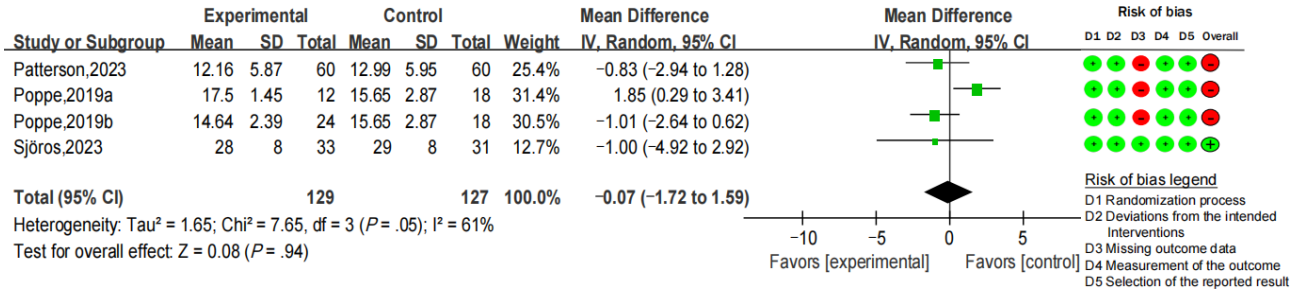
There was no significant decline observed in sedentary bouts >20 minutes (MD −16.32 minutes per day; 95% CI −49.42 to 16.78; I^2=0; *P*=.33) (Figure 5). Among 4 studies that reported the breaks of sedentary time [38,48,49,52], 1 of them presented unreasonable data regarding the breaks without explanation [52]. Therefore, only 3 RCTs were included in the meta-analysis. There was also no significant difference between the intervention and control groups (MD −0.07; 95% CI −1.72 to 1.59; I^2=61%; *P*=.94) (Figure 6).

Figure 5. Forest plot for the mean difference of sedentary bouts (>20 minutes per day) [46,47,53,54].



XSL•FO
RenderX

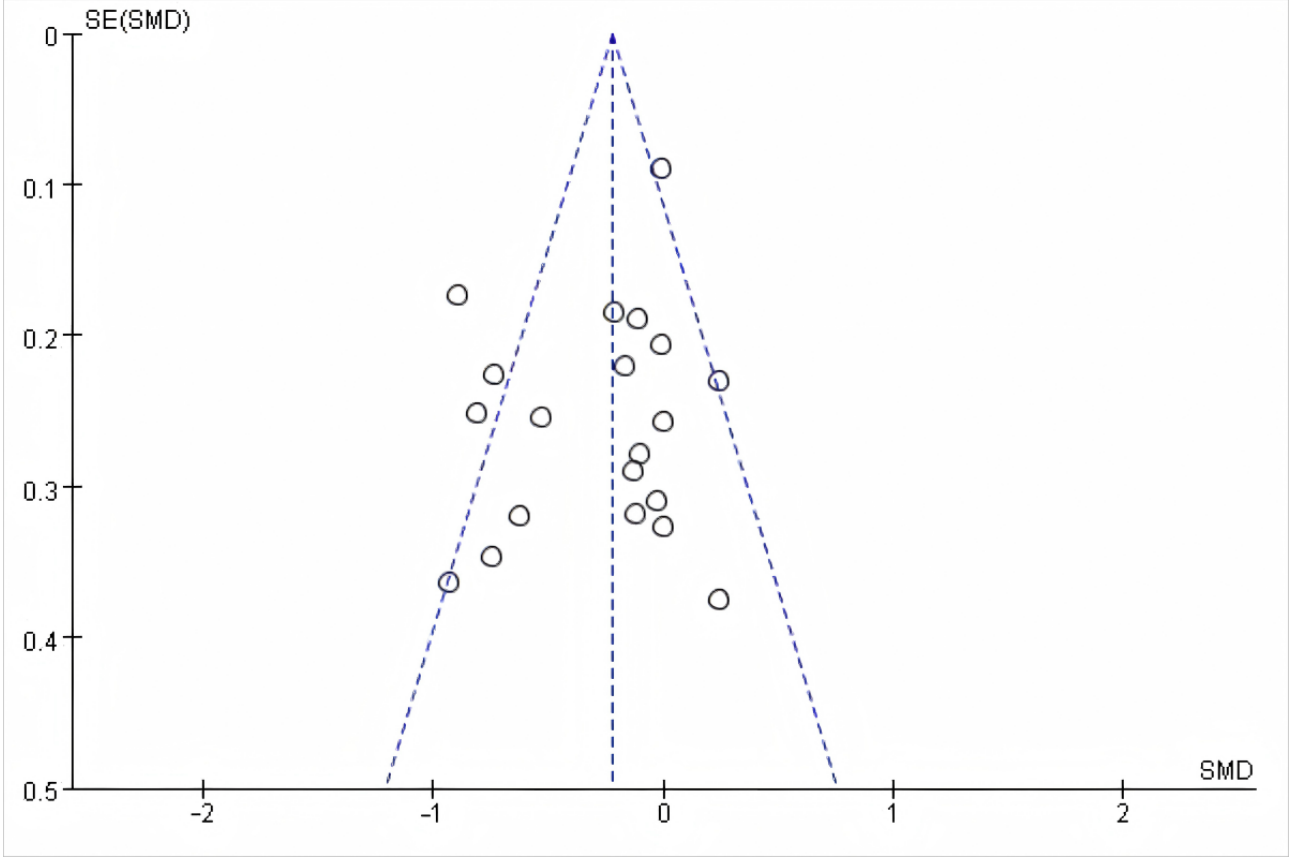
Figure 6. Forest plot for the mean difference of breaks of sedentary time (numbers per day) [38,48,49].



Publication Bias and Sensitivity Analysis

There was no serious publication bias indicated by the symmetry of the funnel plot for overall sitting time (Figure 7). Egger's estimated bias coefficient was −1.22 (*P*=.24) and supported that there was no significant evidence of publication bias. Leave-one-out sensitivity analysis showed consistent results in terms of overall sitting time, pre-post sitting time changes, SB proportion, sedentary bouts, and breaks of sedentary time, implying the robustness of the key results.

Figure 7. Funnel plot overall sitting time (minutes per day), SMD: standard mean difference.



Discussion

Principal Findings

The findings of this review supported the effectiveness of digital health interventions in modifying SB, leading to a reduction in overall sitting time, pre-post sitting time changes, and SB proportion among patients with chronic diseases. The use of digital technology, including mobile phone, web, and activity trackers, facilitated real-time self-monitoring, personalized feedback, and automatic reminders, and elevated motivation and engagement in reducing SB.

Previous researches have shown that SB, characterized by low energy expenditure, was associated with an elevated risk of all-cause mortality in a clear dose-response pattern [12]. This correlation was particularly prominent among individuals who led sedentary lifestyles, notably in those with chronic diseases [67]. The advancement of digital technology could improve the efficacy of the intervention implementation [68]. In this review, the overall sitting time decreased by 30.80 minutes per day, which was less than the 42.28 minutes per day reduction in SB observed in a previous meta-analysis evaluating digital technology–enhanced interventions on SB among healthy adults in a workplace setting [69]. This difference may be explained by different populations (healthy vs unhealthy populations) and settings. Gardner et al [70] supported this explanation by noting that workplace SB was more receptive to routinization than nonworkplace SB. Another systematic review and meta-analysis

XSL•FO
RenderX

evaluated the effectiveness of interventions using wearable activity trackers to enhance PA and reduce SB in hospitalized patients, indicating an average reduction of 35.46 minutes per day in sitting time [23]. However, a different meta-analysis including 48 papers revealed that wearable activity trackers effectively improved conscious exercise behavior, such as steps and MVPA in daily life, but they were not effective in modifying habitual behaviors, such as SB [71]. Variations in the target population and the diverse array of digital technologies used may contribute to the discrepancies in results across the referenced studies and this study.

There were no confirmed results regarding the effectiveness of digital health interventions on SB among specific chronic disease due to the limited number of available studies (n≤4). This notion was also highlighted in a scoping review that evaluated the efficacy of mHealth interventions in addressing PA and SB among survivors from cancer. The evidence on SB was inconclusive, primarily attributed to the limited number of available studies [72]. In another systematic review and meta-analysis examining mHealth app interventions in reducing sedentary time among older adults, the findings did not achieve statistical significance, possibly due to the limited sample size [25]. Therefore, more studies targeting different kinds of chronic diseases are warranted. The results were more promising for those aged 65 years and younger, which was reasonable that younger patients may be more adaptable and easier to incorporate new technologies into interventions. However, individuals aged older than 65 years were more likely to experience multiple chronic diseases and complex health conditions, making the maintenance of an active lifestyle particularly crucial for them. Hence, future study needs to explore effective strategies to enhance the efficacy of digital health interventions aimed at reducing SB among older adults, such as improving their digital health literacy [73].

Theory-based approaches showed a higher reduction in overall sitting time than no-theory–based interventions. Social Cognitive Theory was the most extensively applied framework. This theory focused on individuals' perceptions, beliefs, and expectations regarding SB [74], aiming to raise awareness of the harmful consequences of SB, emphasize the benefits of any level of activity, and promote changes in SB [75]. It was also reasonable to observe a greater SB reduction in the short term than in the long term (55.89 minutes per day vs 18.86 minutes per day). This suggested that maintaining SB reduction was still challenging due to the diminishing novelty of technology-mediated interventions [69] and additional efforts were needed to stimulate long-term SB reduction. For instance, individualized cognitive-behavioral telephone supports were proven effective in sustaining the reduction of SB for at least half a year [58]. The involvement of participants' respective centers in digital health interventions may escalate trust, facilitate information sharing, and encourage seeking support and feedback as necessary for sustained engagement over an extended duration [57]. In addition, improving awareness of SB-related health issues, addressing long sitting time patterns [58], and replacing SB with LPA together with the individual in daily life perspective [36] may prolong the long-term impact of reducing SB. Only the objective sitting time reduction was significant, while the subjective measurements varied a lot and had high heterogeneity (I^2=96%).

SB and PA were 2 of the 3 main categories of 24-hour movement and nonmovement behaviors and were interchangeable with each other [29]. For approaches targeting SB reduction, standing, walking, LPA, and MVPA were the most common substitutes. For studies focused on PA increase solely, we also found a compensatory SB reduction significantly (47.58 minutes per day). However, previous research mentioned that individuals may engage in increased SB after exercise training, which could be attributed to the perception that they earned the right to be sedentary due to their earlier PA [76]. The difference may be explained by different target populations, and further studies are necessary to conclude the difference and interactive mechanism between primary and compensatory SB reduction.

Sedentary bout, an important aspect of SB, showed no significant alteration after digital health interventions. Among the original trials that emphasized the reduction of sedentary bouts or set an upper limit time on sedentary time [47,53,54], only 1 study demonstrated a significant decrease in sedentary bouts (MD −42 minutes per day; 95% CI −83 to −2; *P*=.04) [46]. Notably, sedentary bouts ≥30 minutes were linked to a higher risk of all-cause mortality compared with shorter bouts [77-79]. Furthermore, among individuals with similar total sitting time, frequent interruptions of sitting were associated with lower cardiovascular risks [80]. Given the insufficient focus on breaks in SB (n=3) in the included studies, further research in this area is also warranted.

Strengths and Limitations

The comprehensive search strategy was solely based on RCTs and most of the selected trials had a low risk of bias. We prioritized the use of objective measures of SB in the meta-analysis, making the results of the meta-analysis more reliable. Moreover, we performed subgroup analyses based on whether the interventions targeted SB directly or indirectly to distinguish primary SB reduction from compensatory changes of SB after PA interventions.

There were also certain limitations. First, this review selected only those papers published in English, which may exist a language bias. Second, the limited number of RCTs included in subgroup analyses posed a challenge in identifying potential sources of bias, particularly when focusing on specific types of chronic diseases. Third, the substantial heterogeneity observed in the contents of the interventions and the joint application of digital technologies constrained the ability to interpret the individual contribution of each approach or technology. Therefore, caution is warranted when interpreting the results, and further research that compares different delivery methods of digital health interventions is necessary to provide more robust evidence. While it is not feasible to conduct a meta-analysis to compare different technologies, the findings emphasized the potential effectiveness of different combinations of digital technologies in reducing SB. In addition, a few publications were identified as study protocols or pilot studies,

XSL•FO
RenderX

necessitating their exclusion with the intent to update at a later time.

Clinical Implications and Future Research

In clinical practice, nurses could leverage phone, web, or activity tracker–based interventions to offer support to sedentary patients with chronic diseases, which may effectively reduce their sitting time and then facilitate better disease management.

The digital health interventions with rigorous design targeting SB reduction remain limited, particularly for certain type of chronic illnesses. Additionally, it is recommended to place a greater emphasis on addressing sedentary bouts and breaks, rather than solely focusing on reducing overall sitting time. Meanwhile, this review reveals a reduction in compensatory SB reduction following PA interventions, which has not been widely studied before. Further research is needed to understand the underlying mechanisms and their implications for behavior modification.

Conclusions

Digital health interventions based on phone, web, and activity trackers were promising in reducing sitting time among patients with chronic diseases. Screen time limitation, standing, walking, LPA, and MVPA may be used to substitute SB. Future research focuses on sedentary bouts and breaks with rigorous design are necessary, particularly among patients with specific chronic diseases.

Acknowledgments

This research was funded by the General Research Fund, Research Grants Council of the University Grants Committee of Hong Kong (grant no. 15100822) and the Affiliated Hospital of Qingdao University (grant no. QDFY+X2024206). The funding body had no role in the study's design, data collection, data analysis, interpretation, or manuscript writing.

Data Availability

The datasets in this study are available from the included original studies, which can be extracted accordingly.

Authors' Contributions

YZ participated in conceptualization, methodology, formal analysis, writing—original draft, writing—review and editing, and project administration. FWN participated in conceptualization, supervision, and writing—review and editing. QLY contributed to methodology and formal analysis. YJX participated in conceptualization, methodology, validation, supervision, and writing—review and editing. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and excluded studies.
[DOCX File, 31 KB - mhealth_v13i1e59943_app1.docx]

Multimedia Appendix 2
Characteristics of the included randomized controlled trials.
[DOCX File, 27 KB - mhealth_v13i1e59943_app2.docx]

Multimedia Appendix 3
Risk of bias in individual studies and summary of risk bias.
[DOCX File, 200 KB - mhealth_v13i1e59943_app3.docx]

Multimedia Appendix 4
Grouped bubble plot of the 16 studies.
[DOC File, 61 KB - mhealth_v13i1e59943_app4.doc]

Multimedia Appendix 5
Results of the subgroup analyses.
[DOCX File, 2509 KB - mhealth_v13i1e59943_app5.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 38 KB - mhealth_v13i1e59943_app6.docx]

XSL•FO
RenderX

References

1. Bennett JE, Stevens GA, Mathers CD, et al. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. The Lancet 2018 Sep;392(10152):1072-1088. [doi: 10.1016/S0140-6736(18)31992-5] [Medline: 29784147]
2. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019 Apr 1;40(127-146):127-146. [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]
3. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms “sedentary” and “sedentary behaviours”. Appl Physiol Nutr Metab 2012 Jun;37(3):540-542. [doi: 10.1139/h2012-024] [Medline: 22540258]
4. Young DR, Hivert MF, Alhassan S, et al. Sedentary behavior and cardiovascular morbidity and mortality: a science advisory from the American Heart Association. Circulation 2016 Sep 27;134(13):e262-e279. [doi: 10.1161/CIR.0000000000000440] [Medline: 27528691]
5. Bakker EA, van Bakel BMA, Aengevaeren WRM, et al. Sedentary behaviour in cardiovascular disease patients: risk group identification and the impact of cardiac rehabilitation. Int J Cardiol 2021 Mar 1;326(194-201):194-201. [doi: 10.1016/j.ijcard.2020.11.014] [Medline: 33186667]
6. Ter Hoeve N, Sunamura M, van Geffen ME, et al. Changes in physical activity and sedentary behavior during cardiac rehabilitation. Arch Phys Med Rehabil 2017 Dec;98(12):2378-2384. [doi: 10.1016/j.apmr.2017.05.008] [Medline: 28623143]
7. Prioreschi A, Hodkinson B, Avidon I, Tikly M, McVeigh JA. The clinical utility of accelerometry in patients with rheumatoid arthritis. Rheumatology (Oxford) 2013 Sep;52(9):1721-1727. [doi: 10.1093/rheumatology/ket216] [Medline: 23804220]
8. van der Berg JD, Stehouwer CDA, Bosma H, et al. Associations of total amount and patterns of sedentary behaviour with type 2 diabetes and the metabolic syndrome: the Maastricht Study. Diabetologia 2016 Apr;59(4):709-718. [doi: 10.1007/s00125-015-3861-8] [Medline: 26831300]
9. Wondergem R, Pisters MF, Heijmans MW, et al. Movement behavior remains stable in stroke survivors within the first two months after returning home. PLoS One 2020;15(3):e0229587. [doi: 10.1371/journal.pone.0229587] [Medline: 32214319]
10. Kim RB, Phillips A, Herrick K, et al. Physical activity and sedentary behavior of cancer survivors and non-cancer individuals: results from a national survey. PLoS One 2013;8(3):e57598. [doi: 10.1371/journal.pone.0057598] [Medline: 23483916]
11. Dempsey PC, Matthews CE, Dashti SG, et al. Sedentary behavior and chronic disease: mechanisms and future directions. J Phys Act Health 2020 Jan 1;17(1):52-61. [doi: 10.1123/jpah.2019-0377] [Medline: 31794961]
12. Patterson R, McNamara E, Tainio M, et al. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. Eur J Epidemiol 2018 Sep;33(9):811-829. [doi: 10.1007/s10654-018-0380-1] [Medline: 29589226]
13. Wu J, Fu Y, Chen D, et al. Sedentary behavior patterns and the risk of non-communicable diseases and all-cause mortality: a systematic review and meta-analysis. Int J Nurs Stud 2023 Oct;146:104563. [doi: 10.1016/j.ijnurstu.2023.104563] [Medline: 37523952]
14. Huynh QL, Blizzard CL, Sharman JE, Magnussen CG, Dwyer T, Venn AJ. The cross-sectional association of sitting time with carotid artery stiffness in young adults. BMJ Open 2014 Mar 6;4(3):e004384. [doi: 10.1136/bmjopen-2013-004384] [Medline: 24604484]
15. Carter S, Hartman Y, Holder S, Thijssen DH, Hopkins ND. Sedentary behavior and cardiovascular disease risk: mediating mechanisms. Exerc Sport Sci Rev 2017 Apr;45(2):80-86. [doi: 10.1249/JES.0000000000000106] [Medline: 28118158]
16. Pinto AJ, Bergouignan A, Dempsey PC, et al. Physiology of sedentary behavior. Physiol Rev 2023 Oct 1;103(4):2561-2622. [doi: 10.1152/physrev.00022.2022] [Medline: 37326297]
17. Chau JY, Grunseit AC, Chey T, et al. Daily sitting time and all-cause mortality: a meta-analysis. PLoS One 2013;8(11):e80000. [doi: 10.1371/journal.pone.0080000] [Medline: 24236168]
18. Biswas A, Oh PI, Faulkner GE, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
19. Wilmot EG, Edwardson CL, Achana FA, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Diabetologia 2012 Nov;55(11):2895-2905. [doi: 10.1007/s00125-012-2677-z] [Medline: 22890825]
20. World Health Organization. Recommendations on digital interventions for health system strengthening. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2023-10-01]
21. Global strategy on digital health 2020-2025. World Health Organization. 2021 Aug 18. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf
22. Blasiak A, Sapanel Y, Leitman D, et al. Omnichannel communication to boost patient engagement and behavioral change with digital health interventions. J Med Internet Res 2022 Nov 16;24(11):e41463. [doi: 10.2196/41463] [Medline: 36383427]
23. Szeto K, Arnold J, Singh B, Gower B, Simpson CEM, Maher C. Interventions using wearable activity trackers to improve patient physical activity and other outcomes in adults who are hospitalized: a systematic review and meta-analysis. JAMA Netw Open 2023 Jun 1;6(6):e2318478. [doi: 10.1001/jamanetworkopen.2023.18478] [Medline: 37318806]

XSL•FO
RenderX

24. Daryabeygi-Khotbehsara R, Shariful Islam SM, Dunstan D, McVicar J, Abdelrazek M, Maddison R. Smartphone-based interventions to reduce sedentary behavior and promote physical activity using integrated dynamic models: systematic review. J Med Internet Res 2021 Sep 13;23(9):e26315. [doi: 10.2196/26315] [Medline: 34515637]
25. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343. [doi: 10.2196/14343] [Medline: 31778121]
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535. [doi: 10.1136/bmj.b2535] [Medline: 19622551]
27. Noncommunicable diseases. World Health Organization. 2024 Dec 23. URL: https://www.who.int/health-topics/noncommunicable-diseases#tab=tab_1 [accessed 2023-10-01]
28. Castro EA, Júdice PB, Silva AM, Teixeira PJ, Benito PJ. Sedentary behavior and compensatory mechanisms in response to different doses of exercise-a randomized controlled trial in overweight and obese adults. Eur J Clin Nutr 2017 Dec;71(12):1393-1398. [doi: 10.1038/ejcn.2017.84] [Medline: 28561039]
29. Tremblay MS, Aubert S, Barnes JD, et al. Sedentary behavior research network (SBRN)—terminology consensus project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75. [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
30. Higgins J, Thomas J. Cochrane handbook for systematic reviews of interventions version 64. 2023 Aug 1. URL: https://training.cochrane.org/handbook/current [accessed 2023-12-06]
31. Dettori JR, Norvell DC, Chapman JR. Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. Global Spine J 2022 Sep;12(7):1624-1626. [doi: 10.1177/21925682221110527] [Medline: 35723546]
32. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 1988. URL: https://www.utstat.toronto.edu/brunner/oldclass/378f16/readings/CohenPower.pdf [accessed 2025-06-11]
33. Higgins JPTS. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557]
34. Ashizawa R, Honda H, Take K, et al. Approaches to promote reduction in sedentary behavior in patients with minor ischemic stroke: a randomized controlled trial. Arch Phys Med Rehabil 2022 Feb;103(2):255-262. [doi: 10.1016/j.apmr.2021.08.019] [Medline: 34562434]
35. Ashizawa R, Honda H, Take K, Yoshizawa K, Kameyama Y, Yoshimoto Y. Effects on sedentary behaviour of an approach to reduce sedentary behaviour in patients with minor ischaemic stroke: a randomised controlled trial. Clin Rehabil 2023 Apr;37(4):545-556. [doi: 10.1177/02692155221135412] [Medline: 36357967]
36. Thomsen T, Aadahl M, Beyer N, et al. Sustained long‐term efficacy of motivational counseling and text message reminders on daily sitting time in patients with rheumatoid arthritis: long‐term follow‐up of a randomized, parallel‐group trial. Arthritis Care Res (Hoboken) 2020 Nov;72(11):1560-1570. [doi: 10.1002/acr.24060] [Medline: 31507095]
37. Thomsen T, Aadahl M, Beyer N, et al. The efficacy of motivational counselling and SMS reminders on daily sitting time in patients with rheumatoid arthritis: a randomised controlled trial. Ann Rheum Dis 2017 Sep;76(9):1603-1606. [doi: 10.1136/annrheumdis-2016-210953] [Medline: 28584189]
38. Sjöros T, Laine S, Garthwaite T, et al. Reducing sedentary time and whole-body insulin sensitivity in metabolic syndrome: a 6-month randomized controlled trial. Med Sci Sports Exerc 2023 Mar 1;55(3):342-353. [doi: 10.1249/MSS.0000000000003054] [Medline: 36251378]
39. Norha J, Sjöros T, Garthwaite T, et al. Effects of reducing sedentary behavior on cardiorespiratory fitness in adults with metabolic syndrome: a 6-month RCT. Scand J Med Sci Sports 2023 Aug;33(8):1452-1461. [doi: 10.1111/sms.14371] [Medline: 37073456]
40. Sjöros T, Laine S, Garthwaite T, et al. The effects of a 6-month intervention aimed to reduce sedentary time on skeletal muscle insulin sensitivity: a randomized controlled trial. Am J Physiol Endocrinol Metab 2023 Aug 1;325(2):E152-E162. [doi: 10.1152/ajpendo.00018.2023] [Medline: 37378623]
41. den Uijl I, van den Berg-Emons RJG, Sunamura M, et al. Effects of a dedicated cardiac rehabilitation program for patients with obesity on body weight, physical activity, sedentary behavior, and physical fitness: the OPTICARE XL randomized controlled trial. Phys Ther 2023 Sep 1;103(9):pzad055. [doi: 10.1093/ptj/pzad055] [Medline: 37265452]
42. Lakerveld J, Bot SDM, van der Ploeg HP, Nijpels G. The effects of a lifestyle intervention on leisure-time sedentary behaviors in adults at risk: the Hoorn Prevention Study, a randomized controlled trial. Prev Med 2013 Oct;57(4):351-356. [doi: 10.1016/j.ypmed.2013.06.011] [Medline: 23777672]
43. van Bakel BMA, Kroesen SH, Bakker EA, et al. Effectiveness of an intervention to reduce sedentary behaviour as a personalised secondary prevention strategy for patients with coronary artery disease: main outcomes of the SIT LESS randomised clinical trial. Int J Behav Nutr Phys Act 2023 Feb 14;20(1):17. [doi: 10.1186/s12966-023-01419-z] [Medline: 36788615]
44. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257. [doi: 10.2196/jmir.2662] [Medline: 24269911]

XSL•FO
RenderX

45. Lynch BM, Courneya KS, Sethi P, Patrao TA, Hawkes AL. A randomized controlled trial of a multiple health behavior change intervention delivered to colorectal cancer survivors: effects on sedentary behavior. Cancer 2014 Sep 1;120(17):2665-2672. [doi: 10.1002/cncr.28773] [Medline: 24816611]
46. Lynch BM, Nguyen NH, Moore MM, et al. A randomized controlled trial of a wearable technology‐based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: the ACTIVATE Trial. Cancer 2019 Aug 15;125(16):2846-2855. [doi: 10.1002/cncr.32143] [Medline: 31012970]
47. Maxwell-Smith C, Hince D, Cohen PA, et al. A randomized controlled trial of WATAAP to promote physical activity in colorectal and endometrial cancer survivors. Psychooncology 2019 Jul;28(7):1420-1429. [doi: 10.1002/pon.5090] [Medline: 30980691]
48. Patterson K, Davey R, Keegan R, et al. Testing the effect of a smartphone app on hospital admissions and sedentary behavior in cardiac rehabilitation participants: ToDo-CR randomized controlled trial. JMIR Mhealth Uhealth 2023 Oct 3;11:e48229. [doi: 10.2196/48229] [Medline: 37788043]
49. Poppe L, De Bourdeaudhuij I, Verloigne M, et al. Efficacy of a self-regulation–based electronic and mobile health intervention targeting an active lifestyle in adults having type 2 diabetes and in adults aged 50 years or older: two randomized controlled trials. J Med Internet Res 2019 Aug 2;21(8):e13363. [doi: 10.2196/13363] [Medline: 31376274]
50. Zheng C, Gill JMR, Sun FH, et al. Effects of increasing light versus moderate-to-vigorous physical activity on cardiometabolic health in Chinese adults with obesity. J Sports Sci 2023 Sep;41(16):1547-1557. [doi: 10.1080/02640414.2023.2282278] [Medline: 37963176]
51. Pinto B, Dunsiger S, Stein K. Does a peer-led exercise intervention affect sedentary behavior among breast cancer survivors? Psychooncology 2017 Nov;26(11):1907-1913. [doi: 10.1002/pon.4255] [Medline: 27531024]
52. Biddle SJH, Edwardson CL, Wilmot EG, et al. A randomised controlled trial to reduce sedentary time in young adults at risk of type 2 diabetes mellitus: Project STAND (sedentary time and diabetes). PLoS ONE 2015;10(12):e0143398. [doi: 10.1371/journal.pone.0143398] [Medline: 26623654]
53. Li LC, Feehan LM, Xie H, et al. Effects of a 12-week multifaceted wearable-based program for people with knee osteoarthritis: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 3;8(7):e19116. [doi: 10.2196/19116] [Medline: 32618578]
54. Li LC, Feehan LM, Xie H, et al. Efficacy of a physical activity counseling program with use of a wearable tracker in people with inflammatory arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2020 Dec;72(12):1755-1765. [doi: 10.1002/acr.24199] [Medline: 32248626]
55. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186. [doi: 10.2196/jmir.3340] [Medline: 25217464]
56. Su JJ, Yu DSF. Effects of a nurse-led eHealth cardiac rehabilitation programme on health outcomes of patients with coronary heart disease: a randomised controlled trial. Int J Nurs Stud 2021 Oct;122:104040. [doi: 10.1016/j.ijnurstu.2021.104040] [Medline: 34333211]
57. Alghafri TS, Alharthi SM, Al-Farsi Y, et al. "MOVEdiabetes": a cluster randomized controlled trial to increase physical activity in adults with type 2 diabetes in primary health in Oman. BMJ Open Diabetes Res Care 2018;6(1):e000605. [doi: 10.1136/bmjdrc-2018-000605] [Medline: 30487976]
58. De Greef KP, Deforche BI, Ruige JB, et al. The effects of a pedometer-based behavioral modification program with telephone support on physical activity and sedentary behavior in type 2 diabetes patients. Patient Educ Couns 2011 Aug;84(2):275-279. [doi: 10.1016/j.pec.2010.07.010] [Medline: 20732776]
59. Garthwaite T, Sjöros T, Laine S, et al. Effects of reduced sedentary time on cardiometabolic health in adults with metabolic syndrome: a three-month randomized controlled trial. J Sci Med Sport 2022 Jul;25(7):579-585. [doi: 10.1016/j.jsams.2022.04.002] [Medline: 35487860]
60. Gill DP, Blunt W, Boa Sorte Silva NC, Stiller-Moldovan C, Zou GY, Petrella RJ. The HealtheSteps. BMC Public Health 2019 Jun 28;19(1):841. [doi: 10.1186/s12889-019-7141-2] [Medline: 31253112]
61. Ter Hoeve N, Sunamura M, Stam HJ, et al. Effects of two behavioral cardiac rehabilitation interventions on physical activity: a randomized controlled trial. Int J Cardiol 2018 Mar 15;255(221-228):221-228. [doi: 10.1016/j.ijcard.2017.12.015] [Medline: 29425564]
62. Ormel HL, van der Schoot GGF, Westerink NDL, Sluiter WJ, Gietema JA, Walenkamp AME. Self-monitoring physical activity with a smartphone application in cancer patients: a randomized feasibility study (SMART-trial). Support Care Cancer 2018 Nov;26(11):3915-3923. [doi: 10.1007/s00520-018-4263-5] [Medline: 29785635]
63. Valle CG, Diamond MA, Heiling HM, et al. Effect of an mHealth intervention on physical activity outcomes among young adult cancer survivors: the IMPACT randomized controlled trial. Cancer 2023 Feb 1;129(3):461-472. [doi: 10.1002/cncr.34556] [Medline: 36444676]
64. Brocklebank LA, Falconer CL, Page AS, Perry R, Cooper AR. Accelerometer-measured sedentary time and cardiometabolic biomarkers: a systematic review. Prev Med 2015 Jul;76:92-102. [doi: 10.1016/j.ypmed.2015.04.013] [Medline: 25913420]
65. Priscila M, Vera Z, Élvio RG, Bruce J, Adilson M. Sedentary behaviour: definition, determinants, impacts on health, and current recommendations. In: Adilson M, Élvio RG, editors. Sedentary Behaviour: IntechOpen; 2021:1. [doi: 10.5772/intechopen.100250]

XSL•FO
RenderX

66. Gill DP, Blunt W, Bartol C, et al. HealtheSteps. BMC Public Health 2017 Feb 7;17(1):173. [doi: 10.1186/s12889-017-4047-8] [Medline: 28173782]
67. Katzmarzyk PT, Ross R, Blair SN, Després JP. Should we target increased physical activity or less sedentary behavior in the battle against cardiovascular disease risk development? Atherosclerosis 2020 Oct;311:107-115. [doi: 10.1016/j.atherosclerosis.2020.07.010] [Medline: 32773106]
68. Jandoo T. WHO guidance for digital health: what it means for researchers. Digit Health 2020;6:2055207619898984. [doi: 10.1177/2055207619898984] [Medline: 31949918]
69. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105. [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
70. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJH. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2016;10(1):89-112. [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
71. Li C, Chen X, Bi X. Wearable activity trackers for promoting physical activity: a systematic meta-analytic review. Int J Med Inform 2021 Aug;152:104487. [doi: 10.1016/j.ijmedinf.2021.104487] [Medline: 34020170]
72. Schepens Niemiec SL, Cariño B, Chatfield AJ, Quan K. mHealth-supported interventions with potential to address sedentary behavior in older adults: a scoping review. J Aging Phys Act 2022 Dec 1;30(6):1085-1100. [doi: 10.1123/japa.2021-0338] [Medline: 35354668]
73. Dong Q, Liu T, Liu R, Yang H, Liu C. Effectiveness of digital health literacy interventions in older adults: single-arm meta-analysis. J Med Internet Res 2023 Jun 28;25:e48166. [doi: 10.2196/48166] [Medline: 37379077]
74. Maddison R, Rawstorn JC, Stewart RAH, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan;105(2):122-129. [doi: 10.1136/heartjnl-2018-313189]
75. Motl RW, Sasaki JE, Cederberg KL, Jeng B. Social-cognitive theory variables as correlates of sedentary behavior in multiple sclerosis: preliminary evidence. Disabil Health J 2019 Oct;12(4):622-627. [doi: 10.1016/j.dhjo.2019.05.002] [Medline: 31130491]
76. Martin A, Fitzsimons C, Jepson R, et al. Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis. Br J Sports Med 2015 Aug;49(16):1056-1063. [doi: 10.1136/bjsports-2014-094524] [Medline: 25907181]
77. Buffey AJ, Herring MP, Langley CK, Donnelly AE, Carson BP. The acute effects of interrupting prolonged sitting time in adults with standing and light-intensity walking on biomarkers of cardiometabolic health in adults: a systematic review and meta-analysis. Sports Med 2022 Aug;52(8):1765-1787. [doi: 10.1007/s40279-022-01649-4] [Medline: 35147898]
78. Dempsey PC, Friedenreich CM, Leitzmann MF, et al. Global public health guidelines on physical activity and sedentary behavior for people living with chronic conditions: a call to action. J Phys Act Health 2021 Jan 1;18(1):76-85. [doi: 10.1123/jpah.2020-0525] [Medline: 33276323]
79. Dempsey PC, Larsen RN, Dunstan DW, Owen N, Kingwell BA. Sitting less and moving more: implications for hypertension. Hypertension 2018 Nov;72(5):1037-1046. [doi: 10.1161/HYPERTENSIONAHA.118.11190] [Medline: 30354827]
80. Hibbing PR, Bellettiere J, Carlson JA. Sedentary profiles: a new perspective on accumulation patterns in sedentary behavior. Med Sci Sports Exerc 2022 Apr 1;54(4):696-706. [doi: 10.1249/MSS.0000000000002830] [Medline: 34772905]

Abbreviations

CAD: coronary artery disease
LPA: light physical activity
MD: mean difference
MeSH: Medical Subject Headings
MET: metabolic equivalent
MetS: metabolic syndrome
MVPA: moderate to vigorous physical activity
PA: physical activity
PICOS: Population, Intervention, Control, Outcome, and Study
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RA: rheumatoid arthritis
RCT: randomized controlled trial
SB: sedentary behavior
T2DM: type 2 diabetes mellitus
WHO: World Health Organization

XSL•FO
RenderX

Edited by L Buis; submitted 26.04.24; peer-reviewed by H Baumann, J Sanders; revised version received 11.12.24; accepted 15.04.25; published 24.06.25.

Please cite as:
Zhang Y, Ngai FW, Yang Q, Xie YJ
Effectiveness of Digital Health Interventions on Sedentary Behavior Among Patients With Chronic Diseases: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2025;13:e59943
URL: https://mhealth.jmir.org/2025/1/e59943
doi:10.2196/59943

© Yan Zhang, Fei Wan Ngai, Qingling Yang, Yao Jie Xie. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adherence to eHealth Interventions Among Patients With Heart Failure: Scoping Review

Arno Joachim Gingele[1], MD, PhD; Bianca Steiner[2], PhD; Bettina Zippel-Schultz[2], PhD; Hans-Peter Brunner-La Rocca[1], MD

[1]Department of Cardiology, Maastricht University Medical Center, P Debyelaan 25, 6229 HX, Maastricht, The Netherlands

[2]German Foundation for the Chronically Ill, Berlin, Germany

Corresponding Author:
Arno Joachim Gingele, MD, PhD
Department of Cardiology, Maastricht University Medical Center, P Debyelaan 25, 6229 HX, Maastricht, The Netherlands

Abstract

Background: Heart failure (HF) is a significant global health challenge, requiring innovative management strategies like eHealth. However, the success of eHealth in managing HF heavily relies on patient adherence, an area currently not sufficiently investigated despite its critical role in ensuring the effectiveness of this approach.

Objective: This review was initiated to gather evidence on adherence to eHealth devices among patients with HF. The goal was to survey the current state of adherence, pinpoint factors that promote successful engagement, and identify gaps needing further research.

Methods: A scoping review was conducted to gather quantitative data on eHealth engagement from relevant clinical HF studies indexed in PubMed, CINAHL, and PsycINFO up to February 2025. Descriptive characteristics of the publications were extracted, and generalized mixed model analyses were used to identify eHealth characteristics affecting patient adherence.

Results: Our analysis included 70 studies, primarily using noninvasive eHealth interventions with wearables (n=51), followed by wearables only (n=8), noninvasive eHealth interventions without wearables (n=6), invasive devices (n=3), and telephone support (n=2). The median number of patients per study was 49 (IQR 20 - 139), and the median follow-up duration was 180 (IQR 84 - 360) days. Variability in reporting and definitions of eHealth adherence was noted. In total, 20 studies assessed adherence trends, with 13 noting a decline, 6 observing no change, and 1 reporting an increase over time. Factors influencing adherence were explored in 29 studies; 7 indicated higher adherence with increasing patient age, 2 showed a negative correlation, and 9 detected no age-related differences. No gender differences were found in the 10 publications that reported on gender, and 9 studies found no association between adherence and the New York Heart Association classification, while 1 noted higher adherence in patients with more severe symptoms. In 35 (50%) studies, adherence was quantified as the percentage of mean days the intervention was used, yielding a median adherence rate of 78% (IQR 61% - 86%; range 31% - 98%). No significant correlations were found between adherence rates and the number of eHealth device users, type of intervention, follow-up duration, number of parameters monitored, or data collection frequency.

Conclusions: Reporting and definitions of patient adherence in HF studies are incomplete and inconsistent. Trends indicate a decrease in eHealth use over time. Customizing devices to meet patient needs may help mitigate this issue. Future research should offer a more detailed description of adherence to pinpoint factors that enhance patient adherence with eHealth technologies.

(JMIR Mhealth Uhealth 2025;13:e63409) doi:10.2196/63409

KEYWORDS

heart failure; eHealth; patient adherence; review; quantitative data; mixed model; wearables; invasive devices; telephone support

Introduction

Background

Heart failure (HF) is a critical global health issue marked by high morbidity and mortality, challenging health care systems with its complex management needs and significant resource use [1,2]. The aging population is expected to drive a 46% increase in HF prevalence over the next decade [3], necessitating innovative management strategies to address the growing burden [4].

eHealth, defined as the use of information and communication technologies to support health services or information, offers the potential to revolutionize HF care by enhancing access to health care services, improving patient outcomes, and reducing the necessity for hospital readmissions [5]. Evidence suggests that eHealth interventions like web-based remote patient management systems or mobile apps can significantly improve

XSL•FO
RenderX

morbidity and mortality outcomes in patients with HF, presenting a viable pathway to mitigate the burden of this condition [6]. For instance, the Optimization of the Ambulatory Monitoring for Patients With Heart Failure by Tele-cardiology trial demonstrated that telemonitoring reduced HF hospitalizations in patients with more advanced HF [7].

However, the effectiveness and potential benefits of eHealth in HF management are contingent upon patient adherence to the technologies used [8]. Adherence in eHealth has been defined by Donkin et al [9] as "the degree to which the user followed the program as it was designed" and is influenced by various technological (eg, front-end design) and individual (eg, digital literacy) factors [10]. Despite the acknowledged importance of adherence, there is a paucity of research exploring this aspect, particularly in the context of the diverse range of eHealth devices available, each with its own set of functionalities [6]. Existing reviews on this topic are scarce and often focus on specific subdomains of eHealth, such as mobile health (mHealth) [11], or examine adherence in chronic diseases more broadly rather than addressing HF specifically [12]. Understanding the factors that influence adherence to eHealth in HF is essential for the development and implementation of effective interventions [13].

Consequently, our study embarked on a scoping review to collate evidence on adherence to eHealth devices within the patient population with HF. The aim was to provide an overview of the current state of adherence, identify the determinants of successful engagement, and highlight areas requiring further investigation. This endeavor is critical for informing future strategies to enhance the utility and impact of eHealth solutions in managing HF, ultimately contributing to better patient outcomes and more sustainable health care systems.

Objectives

The primary objective of our literature analysis was to present an overview of eHealth adherence in HF studies, describing trends in adherence and identifying patient characteristics associated with improved adherence. The secondary objective was to identify eHealth characteristics linked to better adherence by assessing the relationship between adherence and these characteristics. Such insights could inform the design of future eHealth interventions by incorporating characteristics that promote sustained adherence.

Methods

Overview

The scoping review conducted to synthesize evidence regarding engagement with eHealth followed the guidelines outlined in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist, as detailed in Checklist 1 [14].

Information Sources

The search for clinical studies was conducted through the PubMed, CINAHL, and PsycINFO databases, encompassing literature published up to February 2025.

Search Strategy

The keywords used in our search were "heart failure," "adherence," and "eHealth," combined with the MeSH (Medical Subject Headings) terms "Heart Failure," "Patient Compliance," and "Telemedicine" (see Multimedia Appendix 1 for the exact searching strategy, including search terms used for PubMed, PsycINFO, and CINAHL). Our search strategy was guided by the population, intervention, comparison, and outcome framework.

- Population: Studies involving patients with HF were included due to the significant impact of HF on health care systems.
- Intervention: Studies that involved eHealth interventions were included, as they have the potential to improve HF care by providing more accessible and continuous care.
- Comparison: Clinical studies, with or without a comparison group, were included. This broad inclusion criterion ensured that a wide range of evidence was captured.
- Outcome: Studies reporting on patient adherence were included, as improved adherence is crucial for the success of eHealth interventions.

Study Selection

Studies qualified for inclusion if they were peer-reviewed and investigated the impact of eHealth interventions on patients aged 18 years or older diagnosed with HF across its entire spectrum (HF with reduced ejection fraction, HF with mildly reduced ejection fraction, or HF with preserved ejection fraction). Additionally, the publications needed to report quantitative system use data to enable objective measurement of patient adherence to the eHealth device. The paper had to be written in English. Study protocols, systematic reviews, qualitative studies, meta-analyses, and conference abstracts were excluded.

Data Collection

Sourced studies were assessed for full-text eligibility by a dual independent review performed by AJG and HPB-LR. In case of disagreement, a third reviewer (BZ-S) was consulted for the final decision. Data extraction was executed by AJG using a structured extraction form to ensure consistency and completeness. For each study, the following study characteristics were extracted: title of the study, name of the first author, year of publication, number of patients included in the eHealth intervention group, type of eHealth intervention (noninvasive eHealth without wearables, noninvasive eHealth with wearables, invasive eHealth, telephone support, and wearables only), description of the eHealth intervention implemented, duration of intervention, overview of adherence metrics, and any factors associated with adherence as reported by the authors. To enhance accuracy, the extracted data were cross-checked for completeness and correctness, and any uncertainties were resolved through discussion with the coauthors.

Statistical Analysis

Descriptive statistics were used to describe the study characteristics of included studies, trends in adherence, and patient characteristics associated with adherence. To assess the relationship between adherence and eHealth characteristics (type

XSL•FO
RenderX

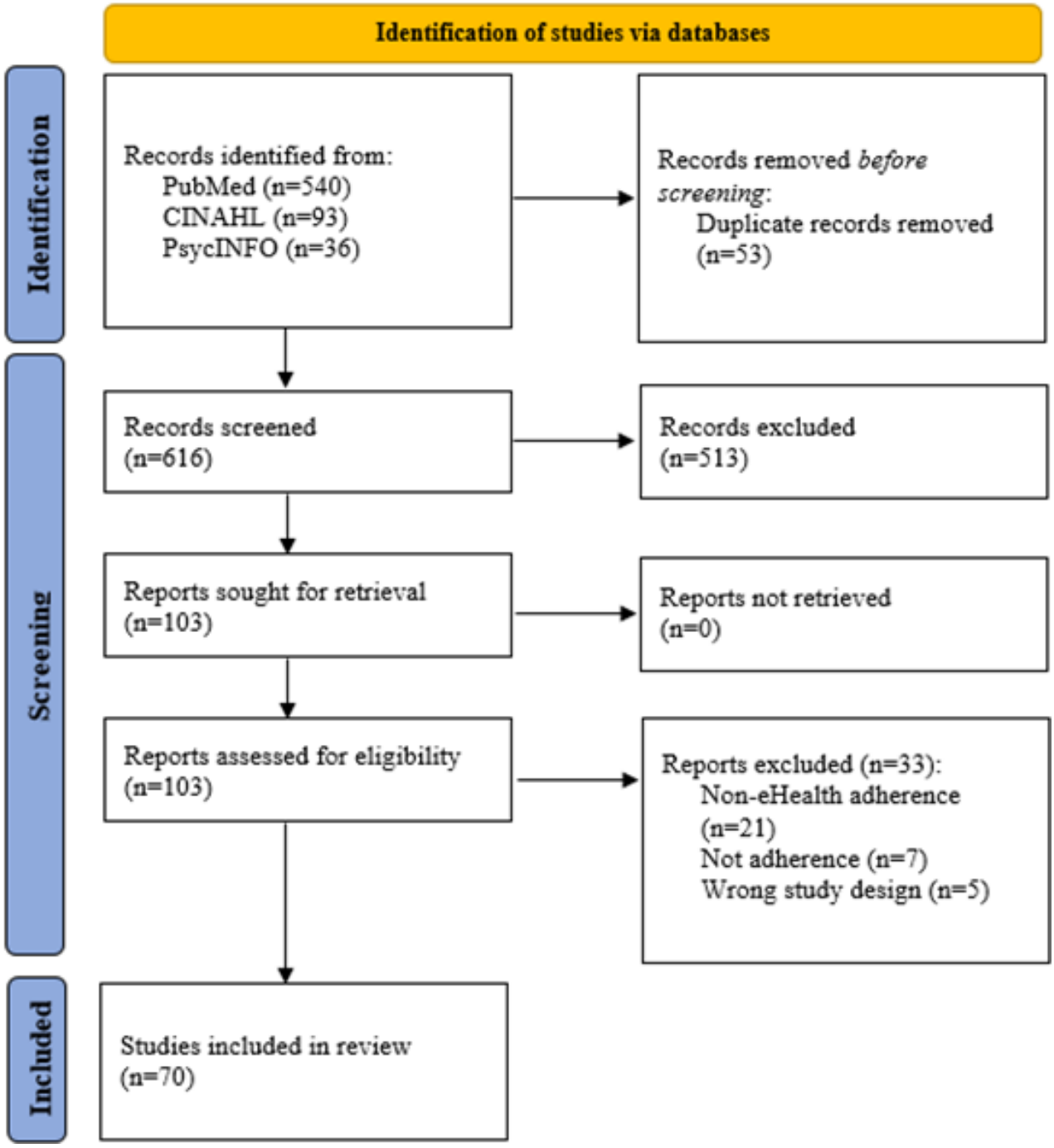
of eHealth intervention, duration of intervention, number of patients with eHealth, number of parameters collected by eHealth, and frequency of data collection), a generalized mixed model analysis was performed. As it was expected that definitions and reporting of adherence might vary significantly between studies, adherence was defined for this analysis as the mean percentage of days the intervention was used. Studies that did not report adherence in this manner or did not provide sufficient data to calculate it were excluded from the inferential statistics. A *P* value of .05 or less was considered to be statistically significant. All analyses were performed in SPSS (version 28; IBM Corp).

Results

Description of Study Characteristics

A total of 669 titles were initially identified (Figure 1). Subsequently, title and abstract screening was conducted, resulting in 103 records being deemed potentially relevant. Following a detailed assessment of the full texts for eligibility, 70 papers were ultimately included in the review.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for study selection.



XSL•FO
RenderX

Among the 70 studies included, the majority used noninvasive eHealth devices with wearables (n=51), followed by wearables only (n=8), noninvasive eHealth devices without wearables (n=6), invasive eHealth devices (n=3), and telephone support (n=2). Among the studies reporting on wearables, the most commonly used devices were weight scales (n=44), blood pressure monitors (n=32), activity trackers (n=11), and electrocardiogram recorders (n=8). The most frequent combination was a weight scale and a blood pressure monitor (n=31). Median number of patients per study using an eHealth device was 49 (IQR 20 - 139, range 9 - 3449). Median duration of follow-up was 180 (IQR 84 - 360, range 7 - 1023) days.

Description of eHealth Adherence Across Studies

The reporting of eHealth adherence exhibited substantial variation across studies, including differences in the definitions used. For example, Koehler et al [15] reported adherence as the number of patients with at least 70% of daily data transfers and no break for >30 days (except during hospitalizations), whereas Kastner et al [16] described adherence as a number of days a complete set of patient data was sent. Additionally, adherence documentation was frequently concise and lacked detail. The diversity of devices used and the types of data collected (eg, vital parameters and symptoms) complicated the pooling of results (see Multimedia Appendix 2 [7,15-83] for detailed study characteristics of all included studies).

In total, 20 (29%) studies explored trends in adherence: 13 reported a decline in adherence over the course of the follow-up, 6 found no change, and 1 study noted improved adherence throughout the study period. For example, Ware et al [17] reported a decline in adherence to a noninvasive eHealth intervention with wearables, defined as the percentage of days all patient data were transmitted, from 81.2% in the first month to 63.1% after 12 months.

Association Between Adherence and Patient Characteristics

In 29 (41%) studies, factors influencing eHealth adherence were reported (Table 1). Among these, 7 studies identified a positive association between patient age and adherence, indicating that older patients were more likely to adhere to eHealth device protocols. Conversely, 2 studies reported a negative association between patient age and adherence levels, while 9 studies observed no association. Regarding the impact of gender on adherence, all 10 studies examining this variable found no significant association between gender and eHealth adherence. Furthermore, 9 of 10 studies reported no association between the New York Heart Association (NYHA) functional classification and eHealth adherence, and 1 study found increased adherence among patients exhibiting more severe symptoms. Besides age, NYHA classification, and gender, most studies reported on living status (n=7), educational level (n=5), and race (n=4). Additionally, 31 factors were mentioned once, including loss of interest and mental overload.

XSL•FO
RenderX

Table . Factors associated with adherence.

Author (year of publication)	Study characteristics	Intervention	Adherence	Factors reported (direction of association, level of significance)
Yoon et al (2024) [18]	• n[a]=39 • Country: Korea • Duration: 1 month	App with Bluetooth-enabled wearables	• App: 80% (mean number of days with logins)	• Age (−[b], s[c])
Kagiyama et al (2024) [19]	• n=77 • Country: Japan • Duration: 6 months (median)	Digital tablet with Bluetooth-enabled wearables	• Overall: 75% (patients with ≥70% of days with self-measurement)	• Age (+[d], s) • QoL[e] measured with MLHFQ[f] (−, s) • *Gender[g], BMI, residence type, past HF[h] hospitalization, NYHA[i], QoL measured with EQ-5D[j]*
Brons et al (2023) [20]	• n=150 • Country: The Netherlands • Duration: 12 months	e-Vita HF web-based platform with wearables	• Weight: 74% (patients with 3 or more data transmissions a week for at least 42 weeks within 1 year) • Weight, blood pressure, and heart rate combined: 67%	• HF hospitalizations (−, s) • *Age, gender, education level, living alone, BMI, smoking, comorbidities, NYHA, LVEF[k], duration of HF, self-care*
Prescher et al (2023) [21]	• n=670 • Country: Germany • Duration: 12 months	Digital tablet with Bluetooth-enabled wearables	• Overall: 89.1% (days all parameters collected) • Blood pressure monitor: 93.71%; weight scale: 92.4% • Electrocardiogram: 92.1%	• Age (+, s) • Cardiovascular hospitalization (−, s) • Rural areas (+, s) • ICD[l] (−, s) • *Gender, DM[m], NT-proBNP[n], living alone, CRT[o], LVEF, NYHA*
Ziacchi et al (2023) [22]	• n=138 • Country: Italy • Duration: 12 months	MyTriage app	• Overall: 49.3% (mean)	• Age (+, s) • NYHA (+, s) • School qualification (−, s) • Assistance by caregiver in app use (+,s)
Barbaric et al (2022) [23]	• n=20 • Country: Canada • Duration: 1 month	Medly voice app	• Overall: 73%	• Age (+) • Confidence in using technology (+)
Sabatier (2022) [24]	• n=659 • Country: France • Duration: 4 months	Web-based telehealth platform	• Overall: 65.7% (mean of anticipated monitoring sessions)	• Visual analog scale moral score (+, s) • *Comorbidities, NYHA, LVEF, treatment adherence*
Apergi et al (2021) [25]	• n=47 • Country: United States • Duration: 3 months	Amazon's Alexa+ or Avatar (via tablet)	• Alexa+: mean use of 35.3 times • Avatar: mean use of 37.8 times	• Age (+, s) • Number of HF medications (−, s) • *Race, household income, or confidence in using technology*
Ploux et al (2021) [26]	• n=51 • Country: France • Duration: 2 months	CareLine Solutions app with wearables	• Before lockdown: 84% • During lockdown: 87%	• Age (+) • Confidence in using technology (+)

XSL•FO
RenderX

Author (year of publication)	Study characteristics	Intervention	Adherence	Factors reported (direction of association, level of significance)
Guzman-Clark et al (2021) [27]	• n=3449 • Country: United States • Duration: 12 months	Telehealth website with wearables	• Overall: 57.1% (mean number of days patients logged in)	• Age (−, s) • White race (−, s) • Severity of illness (−, s)
Haynes et al (2021) [28]	• n=292 • Country: United States • Duration: 6 months	Monitor with Bluetooth-enabled wearables	• Overall: 49% (days weight was transmitted)	• *Emergency department visits*
Zisis et al (2021) [29]	• n=10 • Country: Australia • Duration: 2 months	HF app (via tablet)	• Overall: 20% of participants completed ≥70% of the full program, 80% did not engage at all	• *Age, education level, QoL, HF knowledge, self-care, anxiety, cognition*
Radhakrishnan et al (2021) [30]	• n=15 • Country: United States • Duration: 3 months	Heart Health Mountain app with digital gaming	• Weight: 80% (patients transmitted weight data for 5 or more days a week)	• Use of digital gaming (+, s)
Sohn et al (2020) [31]	• n=20 • Country: United States • Duration: 6 months	FitBit activity tracker with BodyTrace scale and smart pill bottles	• FitBit: 79.1% (median number of hours activity tracker was worn) • Bathroom scale: 59.7% (days scale was used) • Smart pill bottle: 2.8% (days bottle was used)	• SCHFI[p] confidence subscale scores (−, s) (FitBit) • SAQ[q] scores (−, s) (FitBit) • SCHFI confidence subscale scores (−, s) (Body scale) • *Gender, age, race, NYHA, ejection fraction, education, annual income*
Ding et al (2020) [32]	• n=91 • Country: Australia • Duration: 6 months	Bluetooth-enabled weight scale	• Weight: 74% (patients with 4 or more days with data transmission a week)	• Being away from home (−) • Technical issues (−) • Hospitalizations and emergency department presentations (−) • Being unwell, falling, surgery, and chemotherapy (−)
Haynes et al (2020) [33]	• n=538 • Country: United States • Duration: 6 months	Telehealth device with Bluetooth-enabled wearables	• Weight: 53.3% (study days with weight transmission)	• Weekend (−) • Winter months (−)
Haynes et al (2020) [34]	• n=12 • Country: United States • Duration: 34 months (median)	CardioMEMS device (sensor in the pulmonary artery), CardioMEMS pillow	• Overall: 77.6%	• Participants self-identified as committed, capable, eager to follow instructions, and generally do things well (+)
Hovland-Tånneryd et al (2019) [35]	• n=82 • Country: Sweden • Duration: 6 months	OPTILOGG device (app with wearables)	• Overall: 94% (median number of days patients used the eHealth tool)	• Living in rural areas (+) • *Age, gender*
Ware et al (2019) [17]	• n=232 • Country: Canada • Duration: 12 months	Medly smartphone app with wearables	• Overall: 73.6% (days patients took all readings)	• Time device is used (−) • Age (+) • *NYHA, gender*

XSL•FO
RenderX

Author (year of publication)	Study characteristics	Intervention	Adherence	Factors reported (direction of association, level of significance)
Rosen et al (2017) [36]	• n=48 • Country: United States • Duration: 6 months	Telehealth platform (via tablet) with Bluetooth-enabled wearables	• Overall: 96% (median number of days, data were transmitted)	• *Gender, race, age, living situation, depression, cognitive ability, risk for readmission*
Siebermair et al (2015) [37]	• n=159 • Country: Germany • Duration: 21 months (mean)	CareLink system (handheld telemetry wand for data transmission of implanted devices)	• Initial data transmission: 76.1% (patients completing transmission)	• Loss of interest (−) • Mental overload (−) • Lack of support by the responsible general practitioner (−) • Concerns about privacy of relevant health-related device data (−) • Doubts regarding the telemonitoring concept (−) • Technical problems (−)
Prescher et al (2014) [38]	• n=354 • Country: Germany • Duration: 26 months (median)	App with wearables	• Overall: 88.9% (patients transmitting at least 1 vital parameter a day)	• *Age, gender, NYHA, LVEF, depression, diabetes, NT-proBNP, BMI, living alone, duration of HF*
Guzman-Clark et al (2013) [39]	• n=248 • Country: United States • Duration: 3 months	Health Buddy (portable telemonitoring device)	• Overall: 54.9 days (mean number of days with response per study days)	• Primary care (+) • Comorbidity burden (−, s) • *Age, marital status, gender, caregiver available, income, insurance*
Seto et al (2012) [84]	• n=50 • Country: Canada • Duration: 6 months	App with Bluetooth-enabled wearables	• Overall: 70% (patients completing at least 80% of their possible daily readings)	• Planned absence (+) • Equipment failure (+)
Morak et al (2011) [40]	• n=21 • Country: Austria • Duration: 1 week	App with Near Field Communication–enabled wearables	• Overall: 82.2% (at an expected rate of 13 datasets per patient)	• Technical problems (−)
Mortara et al (2009) [41]	• n=301 • Country: Italy • Duration: 12 months	Telephone support with wearables	• Vital signs: 81% (requested data that were transmitted); cardiorespiratory recordings: 92%	• *NYHA, age*
Piette et al (2008) [42]	• n=52 • Country: United States • Duration: 3 months (mean)	Touch-tone telephone support	• Overall: 92% (successful assessment attempts)	• *Age, severity of HF, education*
Clark et al (2007) [43]	• n=79 • Country: Australia • Duration: 12 months	Telephone support	• Overall: 65.8% (patients submitting data once a month or more)	• *Gender, age, living status, weight, NYHA, rural area, comorbidities, HF medication*
de Lusignan et al (2001) [44]	• n=10 • Country: United Kingdom • Duration: 12 months (mean)	Video consultation equipment with wearables	• Weight: 74% (mean number of days data were transmitted) • Blood pressure and pulse watch: 90%	• Planned absence and equipment failure (−) (blood pressure and pulse watch)

XSL•FO
RenderX

[a]n=number of patients with eHealth intervention.
[b]"–" indicates negative association with adherence.
[c]s=association statistically significant.
[d]"+" indicates positive association with adherence.
[e]QoL: quality of life.
[f]MLHFQ: Minnesota Living with Heart Failure Questionnaire.
[g]Factors in italics format indicate no association with adherence.
[h]HF: heart failure.
[i]NYHA: New York Heart Association.
[j]EQ-5D: EuroQol 5 dimensions questionnaire.
[k]LVEF: left ventricular ejection fraction.
[l]ICD: implantable cardioverter defibrillator.
[m]DM: diabetes mellitus.
[n]NT-proBNP: N-terminal pro-brain natriuretic peptide.
[o]CRT: cardiac resynchronization therapy.
[p]SCHFI: Self-Care of Heart Failure Index.
[q]SAQ: Seattle Angina Questionnaire.

Association Between Adherence and eHealth Characteristics

In 35 (50%) studies, adherence was either reported directly or could be inferred as the mean percentage of days the intervention was used (see Multimedia Appendix 2 for detailed study characteristics of all included studies). For these studies, median adherence was 78% (IQR 61% - 86%; range 31% - 98%). No significant correlations were found between adherence rates and predefined eHealth characteristics (Table 2). Telephone support was not detailed in the included studies, and therefore, was not incorporated into the analysis. In the other 35 (50%) studies, adherence was reported in a different manner. For instance, these studies might report the number of patients using the application or solely providing a percentage without further definitions. Therefore, they were excluded from this analysis.

Table . Correlation of eHealth characteristics and adherence.

eHealth characteristics	Wald chi-square (*df*=1)	*P* value
Frequency of data collection	1.057	.30
Number of patients with eHealth	0.070	.79
Duration of intervention	0.002	.97
Number of parameters collected by eHealth	0.306	.58
Type of eHealth intervention (reference category: noninvasive eHealth with wearables)		
Noninvasive eHealth without wearables	0.332	.57
Invasive eHealth	0.053	.82
Wearables only	0.376	.54

Discussion

Principal Findings

The reporting of adherence of patients with HF with eHealth interventions across the included studies was limited and lacked uniformity, which hindered the identification of influencing factors. Despite this, adherence to eHealth interventions was generally high, though there was a large variation between studies. In addition, it tended to decrease over time.

As HF is a chronic disease, its treatment requires lifelong therapy. We observed a trend toward decreased eHealth use over time, a finding corroborated by Nelson et al [85], who documented a similar decline in eHealth use among patients with diabetes mellitus. This pattern is also evident in pharmacological HF therapy, where patient adherence decreases over time [86]. A potential explanation is that patient interest may wane, as the duration of an intervention extends. This issue deserves more attention in general, as our results indicated that loss of interest was reported as a factor influencing eHealth adherence in only one of the included studies. To bolster long-term adherence, it might be crucial to tailor eHealth devices to meet the specific needs of patients [87]. The information provided by the device should be suitable for the patient's cognitive abilities and cultural background. Additionally, the design and user interface should align with the patient's digital experience and interests. As patients will be using the device for years, updating patient education and self-care information in various formats could enhance patient engagement as well. Moreover, adherence to self-care behaviors might vary depending on the specific task. Physical activity is likely the most feasible to track, as it can be measured passively using a smartwatch or activity tracker, whereas behaviors such as weight monitoring or symptom reporting require active patient input,

XSL•FO
RenderX

potentially leading to lower adherence. As evidence suggests that changes in functional status can serve as an early indicator of impending decompensation [88], integrating this parameter more routinely into HF care could enhance disease management.

A significant obstacle to the implementation of eHealth solutions in health care is the concern that patients may be overwhelmed by the additional workload imposed by the devices, potentially leading to the absence or discontinuation of use [89]. Still, no clear trends indicating lower adherence for devices requiring more interaction (such as those monitoring more parameters or collecting data more frequently) were observed in our study. Moreover, none of the other predefined eHealth characteristics were associated with adherence. This could suggest that the eHealth device itself is not the primary driver of nonadherence. However, our findings must be interpreted with caution, as the analysis lacks statistical power due to the small sample size of included studies. Additionally, the heterogeneity of the included devices further limits the robustness of our conclusions. This needs further clarification, as our results indicated that mental overload was reported in only one of the included studies as a factor influencing patient adherence. Previous studies have highlighted the ease of use of technology as a crucial facilitator for eHealth implementation [90]. Thus, patient adherence may depend less on the workload and more on the device's accessibility [91], and future devices should focus on an appropriate front-end design to optimize patient engagement. Clearly, these findings must be interpreted with caution, given the limited evidence available.

In addition to patients' ability to use eHealth devices, it is essential that health care professionals also trust these technologies. A major concern among health care providers regarding the implementation of eHealth is the apprehension that older patients or those with more severe diseases may not be able to use the technology effectively [92]. However, this perception is not corroborated by our findings, as most studies that reported on these parameters did not find a negative association between adherence and either age or NYHA classification. In fact, adherence even improved with increasing age in a substantial number of studies. Health care professionals play a crucial role in providing the necessary clinical information to tailor the device to the patient's needs and should be actively involved in the implementation of eHealth devices in HF management. Additionally, these findings might also encourage health care policy makers to extend eHealth implementation to broader patient populations, rather than limiting it to specific patient characteristics, while emphasizing the need to improve long-term commitment to these devices.

Effective patient engagement is essential for the successful implementation of eHealth solutions in HF management and alleviating the HF burden on health care systems. Yet, clinical studies exploring the impact of eHealth frequently provide minimal information on patient engagement, underscoring a gap in current research methodologies. In addition, the identification of factors influencing adherence is crucial for successful implementation. This should include patients who refuse to use eHealth devices. Future studies should provide a more comprehensive description of patient adherence, including information on the frequency and intensity of eHealth use, as described in the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [93], or the World Health Organization mHealth evidence reporting and assessment checklist [94]. This will ensure better comparability across studies and enhance understanding of adherence behaviors. Furthermore, more research has to be conducted to study suited strategies for enhancing patient engagement. For example, financial incentives have been shown to improve patient adherence to eHealth devices in HF [45]. A similar effect is hypothesized for gamification, the use of game design elements in nongaming settings, although evidence supporting this remains limited [46]. To advance research in this field, HF specialists should be more actively involved in the development and implementation of eHealth solutions. This requires greater interdisciplinary collaboration, including engagement with patients to assess their needs and with software developers to ensure that eHealth tools align with both patient and clinical requirements. Additionally, educating health care professionals about eHealth possibilities can be highly valuable in fostering adoption and effective use.

Limitations

We exclusively used quantitative data to characterize adherence behaviors. While qualitative data from interviews or focus groups could offer deeper insights, our objective was to present trends in adherence as objectively as possible. To enable a meaningful aggregation of the results and quantify the relationships between adherence and eHealth characteristics, outcome measures needed to be comparable. Therefore, we had to exclude nearly 50% (n=35) of the included studies from the inferential statistical analysis, resulting in a small sample size. Due to this small sample size, our study might lack sufficient power to detect a relationship between adherence and eHealth characteristics, thereby increasing the risk of type II errors. The inclusion of studies was restricted to English-language publications to ensure consistency in reporting and interpretation. However, this may have introduced language bias, potentially leading to the exclusion of relevant findings from non-English sources. Nonetheless, as most high-impact papers are published in English, we are confident that the risk of missing substantial data is minimal. Comparing adherence across different device types could have provided valuable guidance for clinical practice. However, this analysis was not feasible due to inconsistencies in adherence reporting, varying adherence definitions across studies, and the small sample sizes within each device category.

Comparison With Prior Work

Recent reviews on patient adherence have primarily focused on chronic diseases broadly, with HF often being underrepresented [12]. Furthermore, a recent review by Madujibeya et al [11] specifically addressing adherence to eHealth in HF only included mHealth while neglecting other eHealth categories. Consequently, we offer a more targeted and comprehensive review of adherence data concerning eHealth in HF. Still, their results are comparable to ours, as they also found a lack of consistency in reporting patient adherence and observed trends toward decreased use of eHealth over time.

XSL•FO
RenderX

Conclusions

Reporting and definition of patient adherence in HF studies lack completeness and consistency. Overall, trends toward a decrease of eHealth use over time could be identified, and tailoring the devices to meet the patients' needs might overcome this problem. Future studies should provide a more comprehensive description and definition of adherence to identify factors improving patient compliance with eHealth.

Acknowledgments

This work was supported by INTERREG-North-West Europe (grant NWE 702). The funder had no role in the study design, data collection, analysis, interpretation, writing of the manuscript, or the decision to submit for publication.

Conflicts of Interest

HPB-LR collaborates with the following digital health companies in European Union–funded projects: Sananet, NL; Orbital Global, UK; Exploris, CH; Nurogames, DE; and Huawei, FI. HPB-LR is the coordinator of the INTERREG-funded Patient Self-Care Using eHealth in Chronic Heart Failure (PASSION-HF) consortium. All other authors declare that they have no conflicts of interest.

Multimedia Appendix 1
Searching strategy.
[DOCX File, 12 KB - mhealth_v13i1e63409_app1.docx]

Multimedia Appendix 2
Overview of included studies.
[DOCX File, 37 KB - mhealth_v13i1e63409_app2.docx]

Checklist 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File, 562 KB - mhealth_v13i1e63409_app3.pdf]

References

1. McMurray JJV, Pfeffer MA. Heart failure. Lancet 2005;365(9474):1877-1889. [doi: 10.1016/S0140-6736(05)66621-4] [Medline: 15924986]
2. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287. [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
3. Virani SS, Alonso A, Aparicio HJ, et al. Heart Disease and Stroke Statistics-2021 Update: a report from the American Heart Association. Circulation 2021 Feb 23;143(8):e254-e743. [doi: 10.1161/CIR.0000000000000950] [Medline: 33501848]
4. Conrad N, Judge A, Tran J, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. Lancet 2018 Feb 10;391(10120):572-580. [doi: 10.1016/S0140-6736(17)32520-5] [Medline: 29174292]
5. Liu S, Li J, Qu Z, Li R, Liu J. The effectiveness of eHealth self-management interventions in patients with chronic heart failure: protocol for a systematic review and meta-analysis. PLoS ONE 2022;17(9):e0268446. [doi: 10.1371/journal.pone.0268446]
6. Scholte NTB, Gürgöze MT, Aydin D, et al. Telemonitoring for heart failure: a meta-analysis. Eur Heart J 2023 Aug 14;44(31):2911-2926. [doi: 10.1093/eurheartj/ehad280] [Medline: 37216272]
7. Galinier M, Roubille F, Berdague P, et al. Telemonitoring versus standard care in heart failure: a randomised multicentre trial. Eur J Heart Fail 2020 Jun;22(6):985-994. [doi: 10.1002/ejhf.1906] [Medline: 32438483]
8. Berry R, Keeling P. Compliance with telemonitoring in heart failure. Are study findings representative of reality?: A narrative literature review. Telemed J E Health 2022 Apr;28(4):467-480. [doi: 10.1089/tmj.2021.0225] [Medline: 34255565]
9. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
10. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 6;19(12):e402. [doi: 10.2196/jmir.8578] [Medline: 29212630]
11. Madujibeya I, Lennie T, Aroh A, Chung ML, Moser D. Measures of engagement with mHealth interventions in patients with heart failure: scoping review. JMIR Mhealth Uhealth 2022 Aug 22;10(8):e35657. [doi: 10.2196/35657] [Medline: 35994345]

XSL•FO
RenderX

12. Pham Q, Graham G, Carrion C, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941. [doi: 10.2196/11941] [Medline: 30664463]
13. Maeder A, Poultney N, Morgan G, Lippiatt R. Patient compliance in home-based self-care telehealth projects. J Telemed Telecare 2015 Dec;21(8):439-442. [doi: 10.1177/1357633X15612382] [Medline: 26556057]
14. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
15. Koehler F, Winkler S, Schieber M, et al. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation 2011 May 3;123(17):1873-1880. [doi: 10.1161/CIRCULATIONAHA.111.018473] [Medline: 21444883]
16. Kastner P, Morak J, Modre R, et al. Innovative telemonitoring system for cardiology: from science to routine operation. Appl Clin Inform 2010;1(2):165-176. [doi: 10.4338/ACI-2009-12-RA-0021] [Medline: 23616835]
17. Ware P, Dorai M, Ross HJ, et al. Patient adherence to a mobile phone-based heart failure telemonitoring program: a longitudinal mixed-methods study. JMIR Mhealth Uhealth 2019 Feb 26;7(2):e13259. [doi: 10.2196/13259] [Medline: 30806625]
18. Yoon M, Lee S, Choi JY, et al. Effectiveness of a smartphone app-based intervention with Bluetooth-connected monitoring devices and a feedback system in heart failure (SMART-HF Trial): randomized controlled trial. J Med Internet Res 2024 Apr 29;26:e52075. [doi: 10.2196/52075] [Medline: 38683665]
19. Kagiyama N, Kasai T, Murata N, et al. Feasibility of self-measurement telemonitoring using a handheld heart sound recorder in patients with heart failure—SELPH multicenter pilot study. J Cardiol 2024 Oct;84(4):266-273. [doi: 10.1016/j.jjcc.2024.04.009]
20. Brons M, Ten Klooster I, van Gemert-Pijnen L, et al. Patterns in the use of heart failure telemonitoring: post hoc analysis of the e-Vita heart failure trial. JMIR Cardio 2023 Jan 31;7:e41248. [doi: 10.2196/41248] [Medline: 36719715]
21. Prescher S, Winkler S, Riehle L, et al. Patient reported experience and adherence to remote patient management in chronic heart failure patients: a posthoc analysis of the TIM-HF2 trial. Eur J Cardiovasc Nurs 2023 Apr 12;22(3):245-253. [doi: 10.1093/eurjcn/zvac080] [Medline: 36062451]
22. Ziacchi M, Molon G, Giudici V, et al. Integration of a smartphone HF-dedicated app in the remote monitoring of heart failure patients with cardiac implantable electronic devices: patient access, acceptance, and adherence to use. J Clin Med 2023 Aug 25;12(17):5528. [doi: 10.3390/jcm12175528] [Medline: 37685593]
23. Barbaric A, Munteanu C, Ross H, Cafazzo JA. A voice app design for heart failure self-management: proof-of-concept implementation study. JMIR Form Res 2022 Dec 21;6(12):e40021. [doi: 10.2196/40021] [Medline: 36542435]
24. Sabatier R, Legallois D, Jodar M, et al. Impact of patient engagement in a French telemonitoring programme for heart failure on hospitalization and mortality. ESC Heart Fail 2022 Oct;9(5):2886-2898. [doi: 10.1002/ehf2.13978] [Medline: 35715956]
25. Apergi LA, Bjarnadottir MV, Baras JS, et al. Voice interface technology adoption by patients with heart failure: pilot comparison study. JMIR Mhealth Uhealth 2021 Apr 1;9(4):e24646. [doi: 10.2196/24646] [Medline: 33792556]
26. Ploux S, Strik M, Abu-Alrub S, et al. Remote monitoring of patients with heart failure during the first national lockdown for COVID-19 in France. Eur Heart J Digit Health 2021 Sep;2(3):487-493. [doi: 10.1093/ehjdh/ztab044] [Medline: 37115642]
27. Guzman-Clark J, Farmer MM, Wakefield BJ, et al. Why patients stop using their home telehealth technologies over time: predictors of discontinuation in Veterans with heart failure. Nurs Outlook 2021;69(2):159-166. [doi: 10.1016/j.outlook.2020.11.004] [Medline: 33279151]
28. Haynes SC, Tancredi DJ, Tong K, et al. The effect of rehospitalization and emergency department visits on subsequent adherence to weight telemonitoring. J Cardiovasc Nurs 2021;36(5):482-488. [doi: 10.1097/JCN.0000000000000689] [Medline: 32398500]
29. Zisis G, Carrington MJ, Oldenburg B, et al. An m-Health intervention to improve education, self-management, and outcomes in patients admitted for acute decompensated heart failure: barriers to effective implementation. Eur Heart J Digit Health 2021 Dec;2(4):649-657. [doi: 10.1093/ehjdh/ztab085] [Medline: 36713108]
30. Radhakrishnan K, Julien C, Baranowski T, et al. Feasibility of a sensor-controlled digital game for heart failure self-management: randomized controlled trial. JMIR Serious Games 2021 Nov 8;9(4):e29044. [doi: 10.2196/29044] [Medline: 34747701]
31. Sohn A, Speier W, Lan E, et al. Integrating remote monitoring into heart failure patients' care regimen: a pilot study. PLoS ONE 2020;15(11):e0242210. [doi: 10.1371/journal.pone.0242210] [Medline: 33211733]
32. Ding H, Jayasena R, Chen SH, et al. The effects of telemonitoring on patient compliance with self-management recommendations and outcomes of the innovative telemonitoring enhanced care program for chronic heart failure: randomized controlled trial. J Med Internet Res 2020 Jul 8;22(7):e17559. [doi: 10.2196/17559] [Medline: 32673222]
33. Haynes SC, Tancredi DJ, Tong K, et al. Association of adherence to weight telemonitoring with health care use and death: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2020 Jul 1;3(7):e2010174. [doi: 10.1001/jamanetworkopen.2020.10174] [Medline: 32648924]

XSL•FO
RenderX

34. Haynes SC, Pallin R, Tong K, Henderson S, Romano PS. Understanding adherence to the CardioMEMS pulmonary artery pressure monitoring system for heart failure: a qualitative study. Heart Lung 2020;49(3):251-258. [doi: 10.1016/j.hrtlng.2020.01.011] [Medline: 32087985]
35. Hovland-Tånneryd A, Melin M, Hägglund E, Hagerman I, Persson HE. From randomised controlled trial to real world implementation of a novel home-based heart failure tool: pooled and comparative analyses of two clinical controlled trials. Open Heart 2019;6(1):e000954. [doi: 10.1136/openhrt-2018-000954] [Medline: 31217992]
36. Rosen D, McCall JD, Primack BA. Telehealth protocol to prevent readmission among high-risk patients with congestive heart failure. Am J Med 2017 Nov;130(11):1326-1330. [doi: 10.1016/j.amjmed.2017.07.007] [Medline: 28756266]
37. Siebermair J, Clauss S, Martens E, et al. Remote monitoring of implantable cardioverter-defibrillators. Problems and implications using a telemonitoring system. Herz 2015 Apr;40 Suppl 2:110-118. [doi: 10.1007/s00059-014-4080-7] [Medline: 24848864]
38. Prescher S, Deckwart O, Koehler K, et al. Will telemonitoring be adopted by patients with chronic heart failure? Dtsch Med Wochenschr 2014 Apr;139(16):829-834. [doi: 10.1055/s-0034-1369849] [Medline: 24722932]
39. Guzman-Clark JRS, van Servellen G, Chang B, Mentes J, Hahn TJ. Predictors and outcomes of early adherence to the use of a home telehealth device by older veterans with heart failure. Telemed J E Health 2013 Mar;19(3):217-223. [doi: 10.1089/tmj.2012.0096]
40. Morak J, Kumpusch H, Hayn D, et al. Near Field Communication-based telemonitoring with integrated ECG recordings. Appl Clin Inform 2011;2(4):481-498. [doi: 10.4338/ACI-2010-12-RA-0078] [Medline: 23616890]
41. Mortara A, Pinna GD, Johnson P, et al. Home telemonitoring in heart failure patients: the HHH study (Home or Hospital in Heart Failure). Eur J Heart Fail 2009 Mar;11(3):312-318. [doi: 10.1093/eurjhf/hfp022] [Medline: 19228800]
42. Piette JD, Gregor MA, Share D, et al. Improving heart failure self-management support by actively engaging out-of-home caregivers: results of a feasibility study. Congest Heart Fail 2008;14(1):12-18. [doi: 10.1111/j.1751-7133.2008.07474.x] [Medline: 18256564]
43. Clark RA, Yallop JJ, Piterman L, et al. Adherence, adaptation and acceptance of elderly chronic heart failure patients to receiving healthcare via telephone-monitoring. Eur J Heart Fail 2007 Nov;9(11):1104-1111. [doi: 10.1016/j.ejheart.2007.07.018] [Medline: 17942364]
44. de Lusignan S, Wells S, Johnson P, Meredith K, Leatham E. Compliance and effectiveness of 1 year's home telemonitoring. The report of a pilot study of patients with chronic heart failure. Eur J Heart Fail 2001 Dec;3(6):723-730. [doi: 10.1016/s1388-9842(01)00190-8] [Medline: 11738225]
45. Mohapatra S, Issa M, Ivezic V, et al. Increasing adherence and collecting symptom-specific biometric signals in remote monitoring of heart failure patients: a randomized controlled trial. J Am Med Inform Assoc 2025 Jan 1;32(1):181-192. [doi: 10.1093/jamia/ocae221] [Medline: 39172649]
46. Shakoor A, Mohansingh C, van der Boon RMA, Brugts JJ, Schaap J. Gamification and its potential for better engagement in the management of heart failure or quality of care registries: a viewpoint. Curr Heart Fail Rep 2024 Dec 5;22(1):4. [doi: 10.1007/s11897-024-00692-8] [Medline: 39636506]
47. Athilingam P, Jenkins B, Johansson M, Labrador M. A mobile health intervention to improve self-care in patients with heart failure: pilot randomized control trial. JMIR Cardio 2017 Aug 11;1(2):e3. [doi: 10.2196/cardio.7848] [Medline: 31758759]
48. Auton A, Zaman S, Padayachee Y, et al. Smartphone-based remote monitoring for chronic heart failure: mixed methods analysis of user experience from patient and nurse perspectives. JMIR Nurs 2023 Jun 6;6:e44630. [doi: 10.2196/44630] [Medline: 37279054]
49. Blockhaus C, Guelker JE, Feyen L, Bufe A, Seyfarth M, Shin DI. Telemonitoring potential of wearable cardioverter-defibrillators during the follow-up of patients with heart failure. J Cardiovasc Dev Dis 2022 Jun 1;9(6):175. [doi: 10.3390/jcdd9060175] [Medline: 35735804]
50. Blomqvist A, Bäck M, Klompstra L, Strömberg A, Jaarsma T. Usability and feasibility analysis of an mHealth-tool for supporting physical activity in people with heart failure. BMC Med Inform Decis Mak 2024 Feb 12;24(1):44. [doi: 10.1186/s12911-024-02452-z] [Medline: 38347499]
51. Blomqvist A, Bäck M, Klompstra L, Strömberg A, Jaarsma T. Testing the recruitment frequency, implementation fidelity, and feasibility of outcomes of the heart failure activity coach study (HEALTHY): pilot randomized controlled trial. JMIR Form Res 2025 Jan 8;9:e62910. [doi: 10.2196/62910] [Medline: 39778202]
52. Boyne JJJ, Vrijhoef HJM, Spreeuwenberg M, et al. Effects of tailored telemonitoring on heart failure patients' knowledge, self-care, self-efficacy and adherence: a randomized controlled trial. Eur J Cardiovasc Nurs 2014 Jun;13(3):243-252. [doi: 10.1177/1474515113487464] [Medline: 23630403]
53. Carter J, Swack N, Isselbacher E, Donelan K, Thorndike AN. Feasibility and acceptability of a combined digital platform and community health worker intervention for patients with heart failure: single-arm pilot study. JMIR Cardio 2023 Oct 2;7:e47818. [doi: 10.2196/47818] [Medline: 37698975]
54. Chausiaux O, Williams G, Keyser M, et al. Insights from implementing an AI-enabled peripheral edema monitor in UK-based community settings for therapeutic decision-making. Heart Lung 2024 Nov;68:385. [doi: 10.1016/j.hrtlng.2024.08.006]

XSL•FO
RenderX

55. Dorsch MP, Farris KB, Rowell BE, Hummel SL, Koelling TM. The effects of the ManageHF4Life mobile app on patients with chronic heart failure: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 7;9(12):e26185. [doi: 10.2196/26185] [Medline: 34878990]
56. Eilat-Tsanani S, Golovner M, Marcus O, et al. Evaluation of telehealth service for patients with congestive heart failure in the north of Israel. Eur J Cardiovasc Nurs 2016 Apr;15(3):e78-e84. [doi: 10.1177/1474515115602677] [Medline: 26311654]
57. Evans J, Papadopoulos A, Silvers CT, et al. Remote health monitoring for older adults and those with heart failure: adherence and system usability. Telemed J E Health 2016 Jun;22(6):480-488. [doi: 10.1089/tmj.2015.0140] [Medline: 26540369]
58. Gardner CL, Flanagan MC, Franklin C, et al. Electronic physiologic and subjective data acquisition in home-dwelling heart failure patients: an assessment of patient use and perception of usability. Int J Med Inform 2016 Sep;93:42-48. [doi: 10.1016/j.ijmedinf.2016.06.001] [Medline: 27435946]
59. Guo X, Gu X, Jiang J, et al. A hospital-community-family-based telehealth program for patients with chronic heart failure: single-arm, prospective feasibility study. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13229. [doi: 10.2196/13229] [Medline: 31833835]
60. Ho K, Novak Lauscher H, Cordeiro J, et al. Testing the feasibility of sensor-based home health monitoring (TEC4Home) to support the convalescence of patients with heart failure: pre-post study. JMIR Form Res 2021 Jun 3;5(6):e24509. [doi: 10.2196/24509] [Medline: 34081015]
61. Johnston W, Keogh A, Dickson J, et al. Human-centered design of a digital health tool to promote effective self-care in patients with heart failure: mixed methods study. JMIR Form Res 2022 May 10;6(5):e34257. [doi: 10.2196/34257] [Medline: 35536632]
62. Keogh A, Brennan C, Johnston W, et al. Six-month pilot testing of a digital health tool to support effective self-care in people with heart failure: mixed methods study. JMIR Form Res 2024 Mar 1;8:e52442. [doi: 10.2196/52442] [Medline: 38427410]
63. Kitsiou S, Gerber BS, Buchholz SW, Kansal MM, Sun J, Pressler SJ. Patient-centered mHealth intervention to improve self-care in patients with chronic heart failure: phase 1 randomized controlled trial. J Med Internet Res 2025 Jan 15;27:e55586. [doi: 10.2196/55586] [Medline: 39813671]
64. Ledwidge MT, O'Hanlon R, Lalor L, et al. Can individualized weight monitoring using the HeartPhone algorithm improve sensitivity for clinical deterioration of heart failure? Eur J Heart Fail 2013 Apr;15(4):447-455. [doi: 10.1093/eurjhf/hfs186] [Medline: 23204211]
65. Lieback A, Proff J, Wessel K, Fleck E, Götze S. Remote monitoring of heart failure patients using implantable cardiac pacing devices and external sensors: results of the Insight-HF study. Clin Res Cardiol 2012 Feb;101(2):101-107. [doi: 10.1007/s00392-011-0369-1] [Medline: 21971974]
66. Lloyd T, Buck H, Foy A, et al. The Penn State Heart Assistant: a pilot study of a web-based intervention to improve self-care of heart failure patients. Health Informatics J 2019 Jun;25(2):292-303. [doi: 10.1177/1460458217704247] [Medline: 28504048]
67. Marier-Tétrault E, Bebawi E, Béchard S, et al. Remote patient monitoring and digital therapeutics enhancing the continuum of care in heart failure: nonrandomized pilot study. JMIR Form Res 2024 Nov 6;8:e53444. [doi: 10.2196/53444] [Medline: 39504548]
68. Morgan JM, Kitt S, Gill J, et al. Remote management of heart failure using implantable electronic devices. Eur Heart J 2017 Aug 7;38(30):2352-2360. [doi: 10.1093/eurheartj/ehx227] [Medline: 28575235]
69. Murugappan T, Raut R. Enhancing home monitoring of patients with heart failure: primary care experience with an autonomous foot-measuring device. Home Health Care Manag Pract 2024 Feb;36(1):31-39. [doi: 10.1177/10848223231176843]
70. Nederend M, Zandstra TE, Kiès P, et al. Potential of eHealth smart technology in optimization and monitoring of heart failure treatment in adults with systemic right ventricular failure. Eur Heart J Digit Health 2021 Jun;2(2):215-223. [doi: 10.1093/ehjdh/ztab028] [Medline: 36712397]
71. Nouryan CN, Morahan S, Pecinka K, et al. Home telemonitoring of community-dwelling heart failure patients after home care discharge. Telemed J E Health 2019 Jun;25(6):447-454. [doi: 10.1089/tmj.2018.0099] [Medline: 30036166]
72. Petrie S, Simard A, Innes E, et al. Bringing care close to home: remote management of heart failure in partnership with Indigenous communities in Northern Ontario, Canada. CJC Open 2024 Dec;6(12):1423-1433. [doi: 10.1016/j.cjco.2024.08.011] [Medline: 39735946]
73. Rao VN, Kaltenbach LA, Granger BB, et al. The association of digital health application use with heart failure care and outcomes: insights from CONNECT-HF. J Card Fail 2022 Oct;28(10):1487-1496. [doi: 10.1016/j.cardfail.2022.07.050] [Medline: 35905867]
74. Skov Schacksen C, Dyrvig AK, Henneberg NC, et al. Patient-reported outcomes from patients with heart failure participating in the future patient telerehabilitation program: data from the intervention arm of a randomized controlled trial. JMIR Cardio 2021 Jul 2;5(2):e26544. [doi: 10.2196/26544] [Medline: 34255642]
75. Scherrenberg M, Leenen JP, van der Velde AE, et al. Bringing the hospital to home: patient-reported outcome measures of a digital health-supported home hospitalisation platform to support hospital care at home for heart failure patients. Digit Health 2023;9:20552076231152178. [doi: 10.1177/20552076231152178] [Medline: 36762022]

XSL•FO
RenderX

76. Schmaderer MS, Struwe L, Loecker C, et al. Feasibility, acceptability, and intervention description of a mobile health intervention in patients with heart failure. J Cardiovasc Nurs 2023;38(5):481-491. [doi: 10.1097/JCN.0000000000000955] [Medline: 36288470]
77. Smeets CJ, Storms V, Vandervoort PM, et al. A novel intelligent two-way communication system for remote heart failure medication uptitration (the CardioCoach Study): randomized controlled feasibility trial. JMIR Cardio 2018 Apr 4;2(1):e8. [doi: 10.2196/cardio.9153] [Medline: 31758773]
78. Triantafyllidis A, Velardo C, Chantler T, et al. A personalised mobile-based home monitoring system for heart failure: the SUPPORT-HF Study. Int J Med Inform 2015 Oct;84(10):743-753. [doi: 10.1016/j.ijmedinf.2015.05.003] [Medline: 26037921]
79. Vuorinen AL, Leppänen J, Kaijanranta H, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014 Dec 11;16(12):e282. [doi: 10.2196/jmir.3651] [Medline: 25498992]
80. Wakefield BJ, Holman JE, Ray A, et al. Outcomes of a home telehealth intervention for patients with heart failure. J Telemed Telecare 2009;15(1):46-50. [doi: 10.1258/jtt.2008.080701] [Medline: 19139220]
81. Wei KS, Ibrahim NE, Kumar AA, et al. Habits Heart app for patient engagement in heart failure management: pilot feasibility randomized trial. JMIR Mhealth Uhealth 2021 Jan 20;9(1):e19465. [doi: 10.2196/19465] [Medline: 33470941]
82. Yokota T, Fukushima A, Tsuchihashi-Makaya M, et al. The AppCare-HF randomized clinical trial: a feasibility study of a novel self-care support mobile app for individuals with chronic heart failure. Eur Heart J Digit Health 2023 Aug;4(4):325-336. [doi: 10.1093/ehjdh/ztad032] [Medline: 37538146]
83. Zan S, Agboola S, Moore SA, Parks KA, Kvedar JC, Jethwani K. Patient engagement with a mobile web-based telemonitoring system for heart failure self-management: a pilot study. JMIR Mhealth Uhealth 2015 Apr 1;3(2):e33. [doi: 10.2196/mhealth.3789] [Medline: 25842282]
84. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31. [doi: 10.2196/jmir.1909] [Medline: 22356799]
85. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of user engagement with mobile- and web-delivered self-care interventions for adults with T2DM: a review of the literature. Curr Diab Rep 2016 Jul;16(7):66. [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
86. Juarez DT, Williams AE, Chen C, et al. Factors affecting medication adherence trajectories for patients with heart failure. Am J Manag Care 2015 Mar 1;21(3):e197-e205. [Medline: 26014307]
87. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251. [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
88. Gingele AJ, Brandts L, Vossen K, Knackstedt C, Boyne J, Brunner-La Rocca HP. Prognostic value of signs and symptoms in heart failure patients using remote telemonitoring. J Telemed Telecare 2024 Jan;30(1):180-185. [doi: 10.1177/1357633X211039404] [Medline: 34516318]
89. Mileski M, Kruse CS, Catalani J, Haderer T. Adopting telemedicine for the self-management of hypertension: systematic review. JMIR Med Inform 2017 Oct 24;5(4):e41. [doi: 10.2196/medinform.6603] [Medline: 29066424]
90. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197. [doi: 10.2196/14197] [Medline: 31755869]
91. Barbaric A, Munteanu C, Ross H, Cafazzo JA. Design of a patient voice app experience for heart failure management: usability study. JMIR Form Res 2022 Dec 6;6(12):e41628. [doi: 10.2196/41628] [Medline: 36472895]
92. Cohen Rodrigues TR, de Buisonjé DR, Keesman M, et al. Facilitators of and barriers to lifestyle support and eHealth solutions: interview study among health care professionals working in cardiac care. J Med Internet Res 2021 Oct 15;23(10):e25646. [doi: 10.2196/25646] [Medline: 34652280]
93. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
94. Agarwal S, LeFevre AE, Lee J, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
HF: heart failure
MeSH: Medical Subject Headings

XSL•FO
RenderX

mHealth: mobile health
NYHA: New York Heart Association
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by M Dorsch; submitted 19.06.24; peer-reviewed by A Svenšek, C Moss, G Petrova, K Radhakrishnan; revised version received 25.03.25; accepted 25.04.25; published 27.06.25.

Please cite as:
Gingele AJ, Steiner B, Zippel-Schultz B, Brunner-La Rocca HP
Adherence to eHealth Interventions Among Patients With Heart Failure: Scoping Review
JMIR Mhealth Uhealth 2025;13:e63409
URL: https://mhealth.jmir.org/2025/1/e63409
doi:10.2196/63409

© Arno Joachim Gingele, Bianca Steiner, Bettina Zippel-Schultz, Hans-Peter Brunner-La Rocca. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy of Mobile App–Based Dietary Interventions Among Cancer Survivors: Systematic Review and Meta-Analysis

Krista Ching Wai Chung[1], BNurs; Naomi Takemura[2], PhD; Wendy Wing Tak Lam[3], PhD; Mandy Man Ho[1], PhD; Antoinette Marie Lee[4], PhD; Wynnie Yuen Yee Chan[5], PhD; Daniel Yee Tak Fong[1], PhD

[1]School of Nursing, University of Hong Kong, 5/F, HKUMed Academic Building, 3 Sassoon Road, Pokfulam, Hong Kong, China (Hong Kong)

[2]School of Nursing, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[3]School of Public Health, University of Hong Kong, Hong Kong, China (Hong Kong)

[4]Department of Psychology, University of Hong Kong, Hong Kong, China (Hong Kong)

[5]School of Professional and Continuing Education, University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Daniel Yee Tak Fong, PhD
School of Nursing, University of Hong Kong, 5/F, HKUMed Academic Building, 3 Sassoon Road, Pokfulam, Hong Kong, China (Hong Kong)

Abstract

Background: The World Health Organization recommends that cancer survivors maintain a healthy diet and weight control to prevent cancer recurrence. Albeit a growing interest in using mobile apps for health promotion, there is a need for comprehensive evidence on the effects of mobile apps, particularly on dietary behaviors.

Objective: This study aims to evaluate the efficacy, feasibility, and acceptability of mobile app–based dietary interventions among cancer survivors and explore the potential mobile app features worth incorporating.

Methods: In this systematic review and meta-analysis, we searched Embase, Cochrane Library, PubMed, and Web of Science from inception to September 2023 without language restriction. We identified studies that used mobile apps for dietary interventions as a major module for cancer survivors. In addition, 2 independent reviewers screened the studies, extracted data, and assessed methodological quality using Cochrane's risk of bias tools for randomized trials (RoB 2) and nonrandomized studies (ROBINS-I). A meta-analysis was conducted on body weight, BMI, nutritional outcomes, and quality of life using random-effects models.

Results: Of the 2621 records identified, 22 studies involving 1204 cancer survivors were included. Notably, existing trials involved only breast and gastrointestinal cancer survivors. Preliminary evidence suggested that mobile app–based dietary interventions demonstrated a beneficial effect on energy intake (Hedges g=1.00, 95% CI 0.96-1.03) and weight changes (Hedges g=−0.43, 95% CI −0.45 to −0.41); as well as a potential to improve protein intake and quality of life among gastrointestinal cancer survivors. The usability, quality, and satisfaction of app use as measured by standardized questionnaires, including the System Usability Scale, the Mobile Application Rating Scale, and the Questionnaire for User Interface Satisfaction, were positive. While feedback messages and dietary goal setting were considered facilitators of mobile app use, concerns regarding the time required for app use and limited food logging options were raised.

Conclusions: Our review found the preliminary efficacy, feasibility, and acceptability of mobile app–based dietary interventions for cancer survivors. However, study heterogeneity should be recognized. More trials are warranted to confirm the effectiveness of these interventions and explore any differences based on cancer types, staging, treatment statuses, the mode of communication with dietitians, and the engagement of family or caregivers. Existing mobile apps could maintain important features such as feedback messages and dietary goal setting while considering the incorporation of artificial intelligence-powered food recognition in food logging and cancer-specific dietary recommendations.

Trial Registration: PROSPERO CRD42023465641; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023465641

(JMIR Mhealth Uhealth 2025;13:e65505) doi:10.2196/65505

KEYWORDS

mobile app; mobile phone; mHealth; cancer survivorship; dietary intervention; digital health intervention; behavior change; mobile health

XSL•FO
RenderX

Introduction

Background

Cancer poses a significant global health concern, with a projected 77% increase in new cases by 2050, surpassing 35 million [1]. Early detection and treatment have resulted in a growing population of cancer survivors [2]. However, the journey of cancer survivorship presents various challenges, including side effects related to cancer treatments, pain, fear of cancer recurrence, financial concerns, impaired sleep, and cognitive functioning [2,3]. Furthermore, lifestyle behaviors such as unhealthy diet, sedentary behaviors, smoking, and alcohol consumption were associated with increased risks of cancer recurrence and overall mortality [4-7].

To address these challenges and reduce the risk of cancer recurrence, the World Health Organization and the World Cancer Research Fund International recommend cancer survivors adopt lifestyle behavioral modifications. These include maintaining a healthy weight, exercising regularly, having a healthy diet, and restricting the consumption of fast foods, red and processed meat, and alcohol [8,9]. Among these recommendations, dietary control for weight reduction and diet quality to ensure adequate nutritional intake are particularly important.

There has been mounting evidence suggesting the relationships between obesity and carcinogenesis. A meta-analysis of 203 observational studies, involving more than 6.3 million participants, concluded that obese patients with breast, colorectal, or prostate cancer experienced higher cancer mortality and relapse rates [10]. While the underlying mechanisms of carcinogenesis in obese individuals remain uncertain, it has been proposed that hyperinsulinemia, elevated BMI, and the overproduction and over-secretion of estrogen and adipokines by adipose tissues may trigger carcinogenesis [11]. Thus, implementing effective dietary control programs for reducing weight in obese cancer patients may help to mitigate mortality and recurrence risks.

In addition, diet quality plays a vital role, especially considering that the etiology of malnutrition among cancer survivors could be different from that in the general population. Specifically, while insufficient nutritional intake may generally be associated with anorexia or problems with oral intake, malnutrition among cancer survivors could also stem from catabolic metabolic derangements [12]. Compliance with dietary guidelines or evidence-based recommendations for better diet quality has been associated with improved survival outcomes among breast and colorectal cancer patients [13]. Therefore, all cancer survivors should undergo regular screening for malnutrition in accordance with the European Society for Clinical Nutrition and Metabolism guideline [12], and nutritional counseling should be provided throughout cancer survivorship [14]. It is important to prioritize adequate energy, protein, and fluid intake over weight management and dietary fiber intake for cancer survivors experiencing treatment-related symptoms and difficulties complying with general recommendations [14].

Traditionally, dietary interventions for cancer survivors have been delivered through face-to-face and telephone formats. A meta-analysis of 25 randomized controlled trials (RCTs) published in 2019 demonstrated the positive impact of traditional dietary interventions on fruit and vegetable intake as well as diet quality among cancer survivors who had completed active treatments [15]; and a network meta-analysis of 98 RCTs published in 2021 concluded that traditional dietary interventions, delivered alone or combined with exercise, were associated with greater reductions in BMI, waist circumferences, and weight compared to standard care among early-stage overweight and obese cancer survivors [16]. While traditional delivery of dietary interventions appears to be effective, it may limit the accessibility, adherence, and engagement of the intervention. The emergence of mobile health (mHealth) technology has revolutionized the provision of remote care for cancer survivors, enabling dietary intervention by alternative methods such as mobile apps, websites, and emails. In addition to synchronous care, which care is provided and received simultaneously, these platforms can facilitate the asynchronous delivery of dietary interventions that allow users to receive care at their convenience. With the growing interest in dietary interventions delivered by mobile apps in the recent decade, several studies had applied mobile apps with diverse features and examined their impact on a variety of outcomes, including anthropometric changes, dietary patterns, nutritional status, and quality of life (QoL) among cancer survivors. Although the existing systematic reviews by Gong et al [17] and Wang et al [18] evaluated the effects of mHealth apps on anthropometric changes, fruit and vegetable consumption, QoL, and fat intake among cancer survivors, they only included 2 studies with dietary interventions being delivered by mobile apps and were limited to English studies [17,18]. In addition, the systematic review by Wang et al [18] searched only 1 electronic database, potentially limiting the comprehensiveness of the findings. To our knowledge, there is no systematic review or meta-analysis specifically evaluating the impact of dietary interventions delivered by mobile apps, highlighting the need to synthesize the existing evidence on this body of research.

Objective

Therefore, this study aims to perform a systematic review and meta-analysis to evaluate the effects of mobile app–based dietary interventions on anthropometric changes, nutritional outcomes, and QoL among cancer survivors. This review also aimed to explore the feasibility, acceptability, and potential features of mobile apps worth incorporating.

Methods

This systematic review and meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [19] (see Checklist 1). The study protocol was registered in the PROSPERO registry (CRD42023465641).

Search Strategy

A comprehensive literature search was conducted from inception to September 19, 2023, in EMBASE, Cochrane Library, PUBMED, and Web of Science with no language restrictions.

XSL•FO
RenderX

The search terms included (cancer OR oncology OR tumour OR tumor OR malignan*) AND (diet* OR nutrition* OR behavior* OR behavior*) AND (mobile app* OR mhealth OR smartphone app* OR mobile-assisted OR technology-supported OR app OR e-health). The search strategy for each database is listed in Multimedia Appendix 1.

Eligibility Criteria

A study was eligible if it (1) evaluated a mobile health app with dietary intervention as the major component in either an intervention or a control group and (2) included patients diagnosed with cancer, regardless of cancer type, stage, and treatment status.

Study Selection and Data Extraction

A total of 2 reviewers (KCWC and NT) independently screened the titles and abstracts of all retrieved records. After agreeing on a list of potentially eligible studies, they independently read the full texts to confirm the eligibility. Full texts of studies not written in English were translated into English using Google Translate. All inconsistencies were resolved by discussion with a third reviewer (DYTF). Reference lists of the identified studies were also read to determine if there were more relevant studies. Corresponding authors of the included studies were contacted when information on the study methods or results was unclear.

The outcomes of interest were body weight, BMI, QoL, dietary patterns, and nutritional status. Among the eligible studies, information regarding study design, settings, sample characteristics, descriptions of intervention and control, follow-up duration, attrition rates, outcomes of interest at baseline and follow-up time points, and acceptability in terms of quality, satisfaction, and usability of app use were extracted independently by 2 reviewers using a standardized form (KCWC and NT).

Quality Assessment

The quality of RCTs was assessed using the Cochrane Collaboration's Tool for Assessing Risk of Bias (RoB) 2.0, whereas that of single-arm trials and quasi-experimental studies was evaluated using Risk of Bias in Nonrandomized Studies (ROBINS-I) [20]. The evaluation was conducted independently by 2 reviewers (KCWC and NT). They also independently assessed the certainty of evidence of each meta-analysis result by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [21].

Data Synthesis and Analysis

Outcomes of interest, common features, feasibility, and acceptability outcomes on which the included studies evaluated the mobile apps were first identified. To obtain the overall effect size of the outcomes of interest, their postintervention means and SDs were extracted to estimate the standardized mean difference (SMD), specifically Hedges *g*, with 95% CIs. Hedges *g* was calculated to standardize the study intervention effects of different studies based on a uniform scale [20]. According to the Cochrane Handbook, SMDs of 0.2, 0.5, and 0.8 represent small, moderate, and large effects, respectively [20]. For each outcome of interest, meta-analysis was performed when at least 2 RCTs measured the same outcome of interest. Thus, only RCTs with at least two of them reporting the same outcome of interest were included in the quantitative synthesis. Random-effects models were adopted, given the heterogeneous mobile app features and intervention durations. Heterogeneity across studies was then assessed based on I^2 statistics [20]. Depending on the magnitude and direction of effects and the strength of evidence for heterogeneity, approximately an I^2 statistic between 0% and 40% indicates low heterogeneity; between 30% and 60% conveys moderate heterogeneity; between 50% and 90% represents substantial heterogeneity; and 75% and 100% implies considerable heterogeneity [20]. Due to the limited number of available trials, funnel plots to assess publication bias and sensitivity analyses could not be performed. All meta-analyses were conducted using RStudio (version 4.3.1; Posit). All significance tests were 2-tailed, and a P value of <.05 was considered statistically significant.

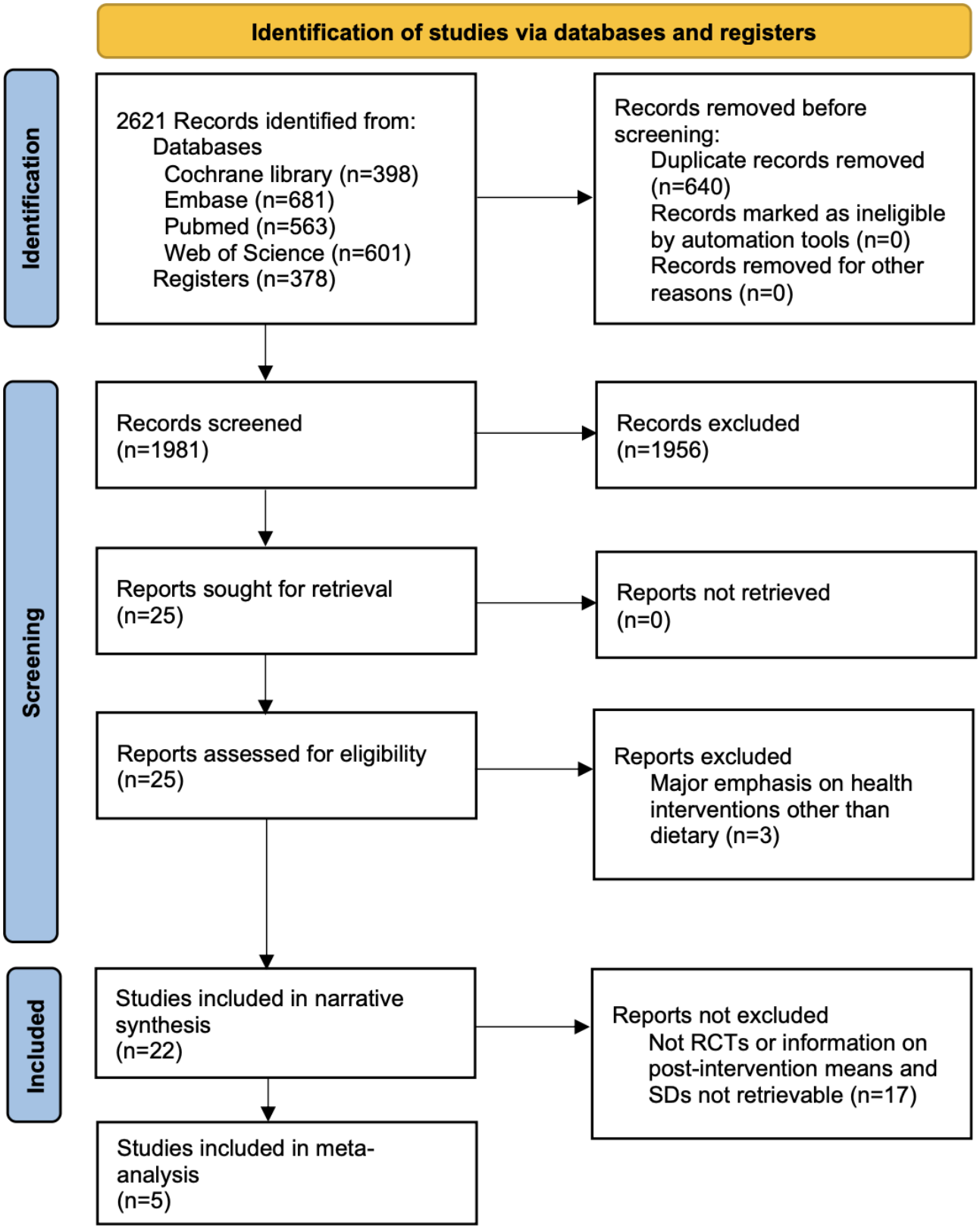
Results

Study Selection

Figure 1 shows the PRISMA flow diagram. The search identified 2621 records from 4 databases. No additional record was found from citation searching. After removing 640 duplicate records and screening the titles and abstracts, 25 studies remained. Their full texts were retrieved and reviewed. Ultimately, 22 studies, comprising 2 in Korean and the remaining in English, met the predetermined eligibility criteria and were included in the review.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for included studies. RCT: randomized controlled trial.



Study and Participants' Characteristics

Of the 22 included studies, 8 were RCTs, 3 were quasi-experimental studies, and 11 were single-arm trials. Their study characteristics are summarized in Table 1 and in Multimedia Appendix 2. The studies were published between 2015 and 2023 and conducted in South Korea (n=8), the United States (n=7), China (n=3), Australia (n=1), Germany (n=1), Iran (n=1), and Spain (n=1). The sample size ranged from 16 to 127, amounting to a total of 1204 participants. The participants' mean age ranged between 15 and 68 years. A total of 9 (45%) studies considered only breast cancer [22-30], whereas 1 study

XSL•FO
RenderX

(5%) involved breast and/or endometrial cancer [31], followed by 8 (36%) on gastrointestinal (GI) cancer [32-39], 2 (9%) on leukemia or lymphoma [40,41], and 1 (5%) on lung cancer [42]. There was 1 study that considered a mix of different cancer types [43]. A total of 10 (45%) studies involved cancer survivors at nonmetastatic stages (0-III) [23,24,26-32,38], and 11 (50%) studies included cancer survivors who had either completed [22,23,25,27-30,38,41] or no longer required active cancer treatments [31,40].

XSL•FO
RenderX

Table . Study characteristics.

Author (year)	Country	Study design	Sample size, n	IG[a], n	CG[b], n	Females, n (%)	Age (years), mean (SD)	Cancer type, staging, and treatment status	Intervention	Control	Intervention + follow-up duration (weeks)
Allicock et al (2021) [22]	United States	RCT[c]	22	13	9	22 (100)	52 (9)	Breast, NR[d], completed (excluding Herceptin and endocrine therapies).	Ecological momentary assessments by Project CHAT[e] app + waist-worn accelerometers + newsletter about healthy diet and physical activity.	Waist-worn accelerometers + newsletter.	4 + 8
Baik et al (2020) [23]	United States	RCT	80	40	40	80 (100)	53 (11)	Breast, 0-III, completed (excluding endocrine therapy).	Psychoeducation and self-management program by My Guide app + telecoaching.	Healthy lifestyle promotion by My Health app + telecoaching.	6 + 8
Cairo et al (2020) [24]	United States	Quasi-experimental study	127	66	61	127 (100)	51[a] (8) and 57[b] (10)	Breast, 0-III, NR.	Coach-guided dietary interventions by Vida app + printed survivorship care plan.	Printed survivorship care plan + self-guided "toolkit".	24 + 48
Cheng et al (2020) [32]	China	Single-arm trial	20	20	NA[f]	2 (10)	62 (7)	Esophageal, I-III, Post-esophagectomy.	Dietary, exercise, and mental support by WeChat app in addition to standard postoperative care.	NA	12 + 12
Choi et al (2020) [25]	South Korea	RCT	50	25	25	50 (100)	49 (8)[a] and 50 (10)[b]	Breast, NR, completed.	Nutritional management program delivered by efilcare R app.	Continued daily life as usual.	4 + 4

XSL•FO
RenderX

Author (year)	Country	Study design	Sample size, n	IG[a], n	CG[b], n	Females, n (%)	Age (years), mean (SD)	Cancer type, staging, and treatment status	Intervention	Control	Intervention + follow-up duration (weeks)
Chow et al (2021) [40]	United States	RCT	41	24	17	20 (49)	45 (20-55)[g]	Hematologic, NR, in remission but not receiving.	Diet and physical activity promotion by Healthwatch360 app and Fitbit app + Fitbit Flex wristband + telephone-based consultation.	Given access to Fitbit tracker and HealthWatch360 app, but did not receive usage reminders or guidance.	16 + 24
Fuemmeler et al (2020) [41]	United States	Single-arm trial	16[h]	16	NA	7 (40)	15 (2)	Hematologic, NR, completed.	Healthy diet and physical activity promotion by Mila Blooms app + pedometer + graphic novel + one brief motivational phone interview + newsletters and educational materials.	NA	8 + 8
Huggins et al (2022) [33]	Australia	3-arm RCT	111	38[i] and 36[j]	37	37 (33)	67 (10)[j] and 68 (10)[i]	Upper gastrointestinal (esophageal, gastric, and pancreatic), all clinical staging, just received emergent surgery and pending curative cancer treatments.	Telephone group: nutritional assessment and education by phone + telephone reviews; myPace app group: nutritional assessment and education + electronic reviews.	Usual care	18 + 48

XSL•FO
RenderX

Author (year)	Country	Study design	Sample size, n	IG[a], n	CG[b], n	Females, n (%)	Age (years), mean (SD)	Cancer type, staging, and treatment status	Intervention	Control	Intervention + follow-up duration (weeks)
Jiang et al (2023) [34]	China	RCT	24	12	12	8 (33)	55 (10)	Gastric, I-IV, post-discharged after gastrectomy.	Dietary intervention by iNutrition applet + telephone dietary consultations.	Usual care	12 + 12
Keum et al (2021) [35]	South Korea	RCT	40	20	20	15 (38)	49 (8)[a] and 50 (10)[b]	Pancreatic, all clinical staging, receiving chemotherapy.	Dietary intervention to promote caloric intake by Noom app	No access to Noom	12 + 12
Lim et al (2023) [26]	South Korea	Single-arm trial	29	29	NA	29 (100)	43 (8)	Breast, 0-IIB, just received surgical treatment and pending chemo-, radio-, or hormonal therapy.	Tailored post-surgery self-management by Breast Cancer by Second Doctor app + smart band.	NA	48 + 48
Lozano-Lozano et al (2019) [27]	Spain	Single-arm trial	80	80	NA	80 (100)	52 (9)	Breast, I-IIIA, completed	Energy balance-focused dietary interventions by BENECA[k] app	NA	8 + 8
McCarroll et al (2015) [31]	United States	Single-arm trial	50	50	NA	50 (100)	58 (10)	Breast and/or endometrial, I-II, not requiring treatment with no sign of recurrence.	Weight loss and dietary modifications by "beta" health care provider version LoseIt! app + weight tracking by Bluetooth scale.	NA	4 + 4

XSL•FO
RenderX

Author (year)	Country	Study design	Sample size, n	IG[a], n	CG[b], n	Females, n (%)	Age (years), mean (SD)	Cancer type, staging, and treatment status	Intervention	Control	Intervention + follow-up duration (weeks)
Orlemann et al (2018) [43]	Germany	Quasi-experimental study	24	12	12	24 (62)	NR	Mixed (67% GI cancer), NR, and NR.	Food record by OncoFood app + paper nutritional documentation + nutrition counseling.	Usual care	4 + 4
Park et al (2019) [28]	South Korea	Quasi-experimental study	71	36	35	71 (100)	52 (4)[a] and 53 (6)[b]	Breast, 0 - 3A, completed active treatments and receiving outpatient follow-up.	Promotion of healthy behaviors (eg, diet and exercising) and peer support by MyFitnessPal and RunKeeper apps + face-to-face health education and peer support group meetings.	Usual care	12 + 24
Salmani et al (2022) [36]	Iran	Single-arm trial	17	17	NA	7 (41)	57 (17)	Colorectal, NR, receiving.	Cancer self-management by the "Colorectal Cancer Along" app.	NA	2 + 2
Seo et al (2021) [29]	South Korea	Single-arm trial	20	20	NA	20 (100)	51 (7)	Breast, I-III, completed.	Promotion of healthy lifestyle (eg, diet and exercising) by "Health for You" app.	NA	2+ 2

XSL•FO
RenderX

Author (year)	Country	Study design	Sample size, n	IG[a], n	CG[b], n	Females, n (%)	Age (years), mean (SD)	Cancer type, staging, and treatment status	Intervention	Control	Intervention + follow-up duration (weeks)
Soh et al (2018) [37]	South Korea	Single-arm trial	203	203	NA	75 (37)	NR[l]	Gastric or colon, NR, received surgery or receiving chemotherapy (50.2% receiving palliative care).	Support for nutrition, QoL[m], and rehabilitation by multidisciplinary mobile care system, Life Manager app + anthropometric measurements.	NA	12 + 12
Stubbins et al (2018) [30]	The United States	Single-arm trial	33	33	NA	33 (100)	57 (9)	Breast, I-III, completed.	Daily record of diet and physical activity + real-time communication with dietitians by MOCHA[n] app.	NA	4 + 4
Wang et al (2022) [38]	China	RCT	60	30	30	25 (45)	69 (7)[a] and 68 (9)[b]	Colorectal, I-III, completed.	Dietary consultation and adjustment of dietary plans by dietitians by WeChat app-based video calls + nutritional status assessments during two home visits.	Routine care (telephone follow-up assessments on disease-related symptoms + lifestyle advice + mental rehabilitation).	24 + 24

XSL•FO
RenderX

Author (year)	Country	Study design	Sample size, n	IG[a], n	CG[b], n	Females, n (%)	Age (years), mean (SD)	Cancer type, staging, and treatment status	Intervention	Control	Intervention + follow-up duration (weeks)
Yang et al (2021) [39]	South Korea	Single-arm trial	36	36	NA	0 (0)[o]	59[g,p]	Esophageal, all clinical, receiving neoadjuvant concurrent chemoradiotherapy.	Health coaching interventions to prevent malnutrition and excessive muscle loss by Noom app.	Retrospective review by Yoon et al (2020) among subjects who underwent operation and received usual care without Noom use.	8 + 8
Yang et al (2022) [42]	South Korea	Single-arm trial	50	50	NA	22 (44)	58 (12)	Lung, I-IV, receiving outpatient chemotherapy treatments or paying routine outpatient visits after lung resection surgery.	Tailored diet and exercise program by Smart After-Care app + self-monitoring devices (eg, sphygmomanometer, finger pulse oximeter, and digital spirometer) + telephone health counseling.	NA	12 + 12

[a]IG: intervention group.

[b]CG: control group.

[c]RCT: randomized controlled trial.

[d]NR: not reported.

[e]CHAT: Creating Healthy Actions through Technology.

[f]NA: not available.

[g]Age reported as median (IQR).

[h]Parent-adolescent dyads.

[i]Telephone intervention group.

[j]myPace app intervention group.

[k]BENECA: The Energy Balance on Cancer.

[l]35.5% in their 50s.

[m]QoL: quality of life.

[n]MOCHA: Methodist Hospital Cancer Health Application.

[o]All participants were male.

[p]IQR not reported.

XSL•FO
RenderX

Characteristics of Dietary Interventions and Features of Mobile Apps

A total of 8 (36%) studies used mobile apps as the sole delivery means of dietary interventions [25,27,29,30,33,35,36,39]. Other studies used additional delivery means such as telephone-based consultations [23,24,34,40,42] and face-to-face health educational classes plus support group meetings [28]. Among all the included studies, 10 studies (45%) had app users maintain regular interaction with dietitians [25,30,33-35,38,43] or coaches [24,39,41]. Both the intervention period and follow-up duration of the 22 studies ranged between 1 to 12 months, with a median of 2.5 months.

Among the 22 included studies, 2 applied the WeChat app (a messaging and social media app; Tencent), 2 used the Noom app (a weight management and behavior change app), and the remaining 18 studies used different mobile apps, summing up a total of 20 different mobile apps. These mobile apps incorporated different intervention components. A total of 5 (23%) studies used mobile apps that exclusively focused on diet [25,33,34,38,43]. However, 10 (45%) studies adopted apps with additional components on physical activity [22,27,28,30,31,35,39-42], while the remaining 7 (32%) studies additionally incorporated other health-related components, including cancer symptom management, mental health support, medication, and pain management [23,24,26,29,32,36,37]. A total of 8 (36%) studies incorporated a theoretical framework to develop or deliver interventions, which included the Social Cognitive Theory (n=4) [22,27,28,31], the Self-Determination Theory (n=2) [40,41], the Health Action Process Approach theory (n=1) [34], and the Supportive Accountability Framework (n=1) [33].

A total of 6 common features of the 20 mobile apps are identified and summarized (see Table 2). The most common feature was self-diet monitoring [22,24-37,39-41,43], followed by dietary goals setting by health care professionals, research staff, or mobile app users [24,27-31,33,34,36,37,40,41,43], feedback messages (messages based on self-diet monitoring data [22,24,25,27,31,32,35,41], motivational messages [22,25,31], and replies to mobile app users' questions [30,32,34,35,38,39]), self-body weight monitoring [24,29,31,33-35,37,39,43], personalized dietary management planning either developed by dietitians, research staff, or co-developed by dietitians and mobile app users [25,33,34,38,42], and social platforms to facilitate peer support [28,32,34,41].

Table . Summary of adopted mobile app features.

Mobile app features	Studies, n (%)
Self diet monitoring	19 (86)
Dietary goals setting	13 (59)
Feedback messages	12 (55)
Self body weight monitoring	9 (41)
Personalized dietary management planning	5 (23)
Social platform	4 (18)

Methodological Quality

The risk of bias in all RCTs and nonrandomized trials is summarized in Multimedia Appendix 3. All RCTs were rated as having "some concerns" [22,23,25,33,34,38,40] due to the approach of measuring self-reported outcomes. Among the nonrandomized trials, 3 studies were identified as having serious bias [24,26,28], while the risk of bias in 1 study could not be determined due to insufficient information [43]. One study had a serious bias due to confounding and selection bias, as it did not control for the prognostic factor age that showed significant baseline difference, and its recruitment of curative-intent cancer survivors may have included participants with a higher level of health consciousness [24]. Another study had a serious bias due to substantial missing data, as it had 42% of participants dropping out before the start of the intervention and a high attrition rate of 61% [28]. The other study had serious measurement bias due to the use of non-validated instruments [26].

Usability, Quality, and Satisfaction of Mobile App Use

A total of 9 of the 20 mobile apps were evaluated for their usability, quality, or satisfaction, which involved breast, GI, and lung cancer survivors. Their evaluation results are summarized in Multimedia Appendix 4. In addition, 5 of the 9 apps were evaluated using standardized questionnaires, including the 10-item System Usability Scale (SUS; n=2) [30,34], the 23-item Mobile Application Rating Scale (MARS) and its 26-item user version (n=2) [27,29], and the 27-item Questionnaire for User Interface Satisfaction (QUIS; n=1) [36]. The other 4 apps were rated using nonstandardized questionnaires assessing satisfaction and usability [25,26,37,42].

Only 2 studies evaluated the usability of their apps using the SUS [30,34], whereas another study rated its app using a self-developed scale [26]. With a mean total SUS score above 70 considered acceptable and with good adjective ratings [44], the MOCHA (Methodist Hospital Cancer Health Application) app was rated with a mean total SUS score of 77.4 (SD not reported) among stage I to III breast cancer survivors [30], and the iNutrition applet was rated with a mean total SUS score of 77.27 (SD 10.69) among postgastrectomy cancer survivors [34]. The Second Doctor app was evaluated based on a self-developed 14-item usability scale and received a mean total usability score of 80.2 out of 100 among breast cancer survivors [26].

Only 2 studies examined the quality of their apps using the MARS and the user version of the MARS (uMARS) [27,29].

XSL•FO
RenderX

Both scales rate app quality based on domains including engagement, functionality, aesthetics, information, app subjective quality, and perceived app use impact, with individual items rated on a 5-point scale, in which "1" indicates inadequate and "5" represents excellent [45,46]. The BENECA and "Health for You" apps received a mean total MARS and uMARS score of 3.71 (SD 0.47) and 3.60 (SD 0.69), respectively, among stage I to III breast cancer survivors [27,29].

Regarding satisfaction, only 1 app was rated using the QUIS [36], while 4 other apps were evaluated by nonstandardized self-developed questionnaires [25,26,37,42]. The "Colorectal Cancer Along" app was assessed on user satisfaction using QUIS on a 10-point Likert scale, which assesses specific domains, including overall reaction to the app, screen design and layout, terminology and application information, learnability, and app features [36]. The higher the score, the greater the level of satisfaction with the user interface [36]. Specifically, the QUIS score of this app for the "overall reaction to the app" domain was 7.94 (SD 1.38) among colorectal cancer survivors receiving active cancer treatments [36]. Among the other 4 apps, which involved breast, GI, and lung cancer survivors, a higher score on the non-standardized questionnaires indicates a higher level of app use satisfaction [25,26,37,42]. The mean total satisfaction scores of the efilcare R app (LifeSemantics) and the Life Manager app (LifeSemantics) were between 3.93 and 4.2 out of 5 [25,37], and that of the Second Doctor app was 22.4 out of 30 [26]. While the mean total score was not reported, 88% of the Smart After-Care app users rated app use satisfaction as "very good" or "good" [42].

Facilitators and Barriers to App Use

A total of 5 mobile apps were evaluated qualitatively [24,26,27,40,43]. Common facilitators of app use included the provision of educational information [27,40], feedback [27,40], motivational messages [40], and setting and tracking of nutritional goals [43]. On the other hand, barriers to app use included concerns about the amount of time committed to app use [24,26,43], limited food diary entry options for assessing food intake [26,27], perceived ineffectiveness of action plans for inducing behavioral changes [40], challenges in adopting the app's advice in real-life situations [40], and the lack of concrete meal plan to guide dietary choices [26].

Effect on Anthropometric Measurements

A total of 5 RCTs assessed body weight and/or BMI, which involved breast and GI cancer survivors receiving dietary interventions that did not specify body weight goals [22,25,33,34,38]. The corresponding pooled effects are shown in Figures 2-5. For breast cancer survivors, their mobile app–based dietary interventions neither showed statistically significant effects on weight changes (Hedges g=0.20, 95% CI −2.24 to 2.64; P=.49; I^2=0%; see Figure 2) nor on BMI (Hedges g=−0.23, 95% CI −6.15 to 5.69; P=.71; I^2=69%; see Figure 3) compared to the control groups (see Table 1). For GI cancer survivors, meta-analysis showed a significant effect on weight changes (Hedges g=−0.43, 95% CI −0.45 to −0.41; P=.003; I^2=0%; see Figure 4), but not on BMI (Hedges g=0.03, 95% CI −5.10 to 5.16; P=.96; I^2=65%; see Figure 5) compared to control groups (see Table 1). The certainty of evidence for body weight using the GRADE approach was low for breast cancer and high for GI cancer survivors, respectively (see Multimedia Appendix 5). The certainty of evidence for BMI based on the GRADE approach was very low for both breast and GI cancer survivors (see Multimedia Appendix 5).

Figure 2. Meta-analysis results for body weight among breast cancer survivors [22,25]. SMD: standardized mean difference.

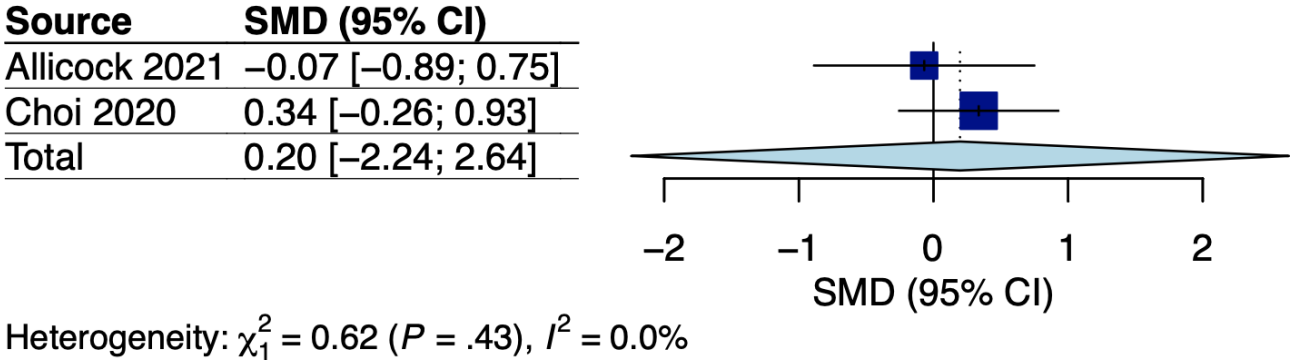
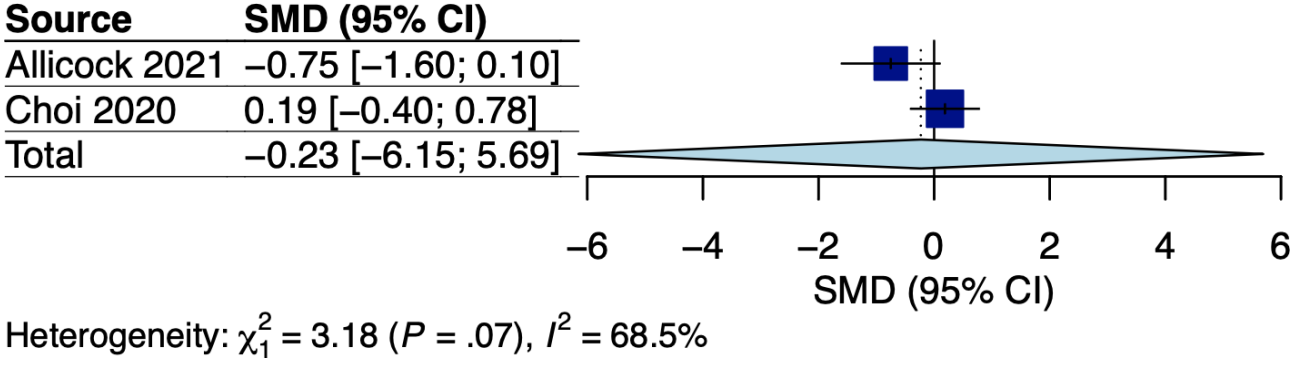


Figure 3. Meta-analysis results for BMI among breast cancer survivors [22,25]. SMD: standardized mean difference.



XSL•FO
RenderX

Figure 4. Meta-analysis results for body weight among gastrointestinal (GI) cancer survivors [33,34]. SMD: standardized mean difference.

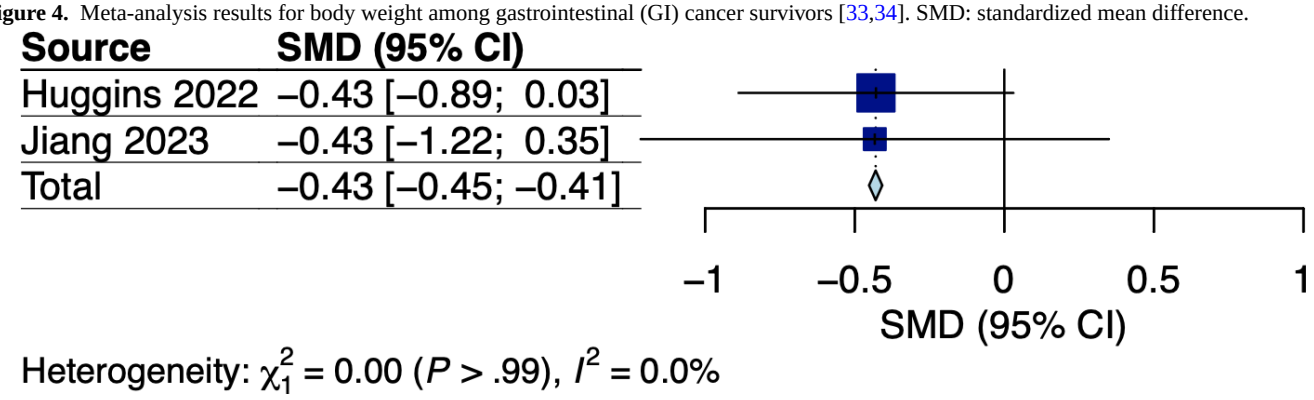
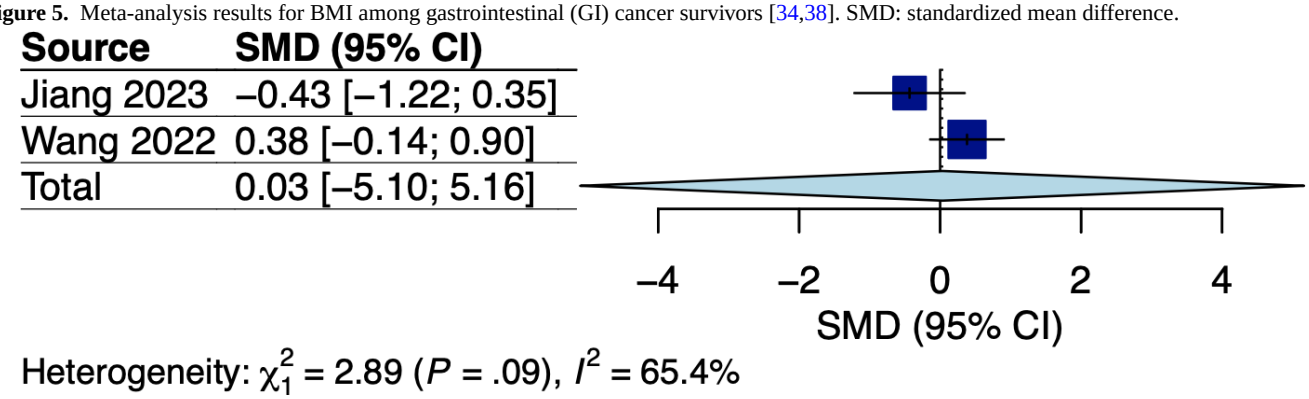


Figure 5. Meta-analysis results for BMI among gastrointestinal (GI) cancer survivors [34,38]. SMD: standardized mean difference.



Among these 5 RCTs, only 1 study reported a baseline mean BMI that exceeded the overweight or obese threshold [22]. Another study did not report the baseline mean BMI [33]. The remaining 3 studies reported baseline mean BMIs within the normal range [25,34,38], although one of these studies only provided BMI at follow-up, without body weight information [38]. Subgroup analyses of patients with a normal baseline BMI showed insignificant effects compared to control groups for both body weight (Hedges g=0.002, 95% CI −4.84 to 4.85; P≥.99; I^2=55%) and BMI (Hedges g=0.14, 95% CI −0.80 to 1.08; P=.59; I^2=27%).

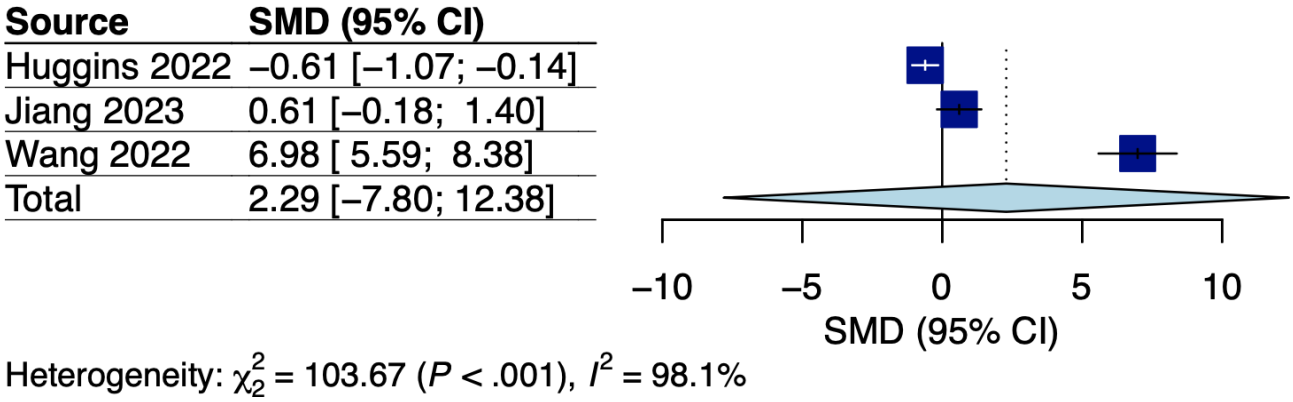
For nonrandomized trials, 3 studies conducted among overweight or obese breast cancer survivors [24,27,31] and another study involving esophageal cancer survivors [32] reported significantly reduced weight and BMI.

Effect on Quality of Life

A total of 4 RCTs that involved GI cancer survivors and 3 single-arm trials which involved breast, esophageal, and lung cancer survivors reported QoL using the global health status domain of the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) [27,32-35,38,42]. Since the mean global health status scores in the RCT by Keum and colleagues [35] were not reported, only 3 RCTs were included in our pooled analysis, with the corresponding pooled effect shown in Figure 6. Mobile app–based dietary interventions did not show an effect on QoL outcomes with substantial heterogeneity (Hedges g=2.29, 95% CI −7.80 to 12.38; P=.43; I^2=98%; see Figure 6). However, meta-regression or subgroup analysis was not possible due to the limited available trials. The certainty of evidence using the GRADE approach was very low (see Multimedia Appendix 5). Among the single-arm trials, significantly improved within-group global health status scores were reported among breast and lung cancer survivors [27,42], whereas no improvement was reported among esophageal cancer survivors [32].

XSL•FO
RenderX

Figure 6. Meta-analysis results for quality of life among gastrointestinal (GI) cancer survivors [33,34,38]. SMD: standardized mean difference.



Effect on Dietary Intake and Nutritional Status

Energy and Protein Intake

A total of 2 RCTs measured caloric and protein intake based on 24-hour dietary recall questionnaires among colorectal and gastric cancer survivors receiving dietary interventions tailored to individuals' nutritional requirements [34,38], with the respective pooled effects shown in Figures 7 and 8. Meta-analysis showed significant effects of mobile app–based interventions on energy intake (Hedges *g*=1.00, 95% CI 0.96-1.03; *P*=.002; I^2=0%; see Figure 7). Nonetheless, mobile app–based interventions had no significant effect on protein intake with moderate heterogeneity (Hedges *g*=1.29, 95% CI −3.79 to 6.37; *P*=.19; I^2=59%; see Figure 8). The certainties of evidence for energy and protein intake using the GRADE approach were high and moderate, respectively (see Multimedia Appendix 5).

Figure 7. Meta-analysis results for energy intake among gastrointestinal (GI) cancer survivors [34,38]. SMD: standardized mean difference.

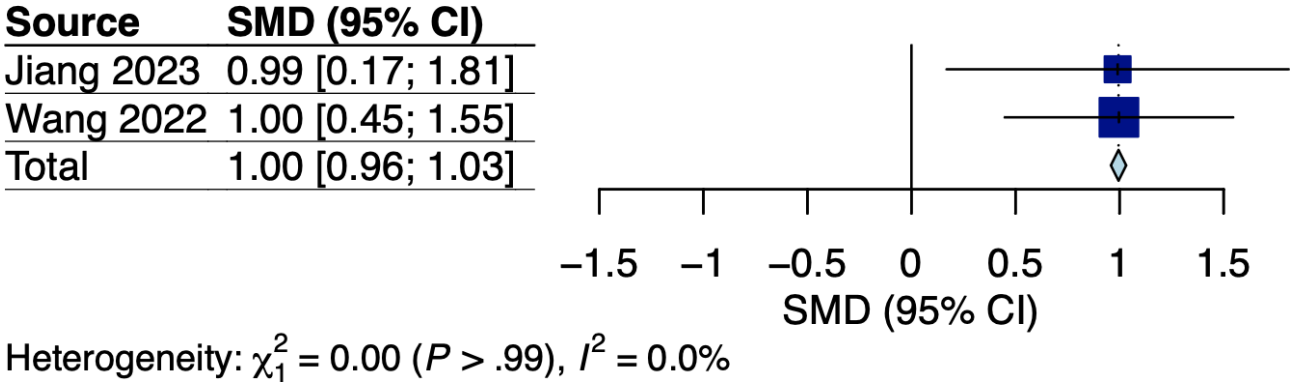
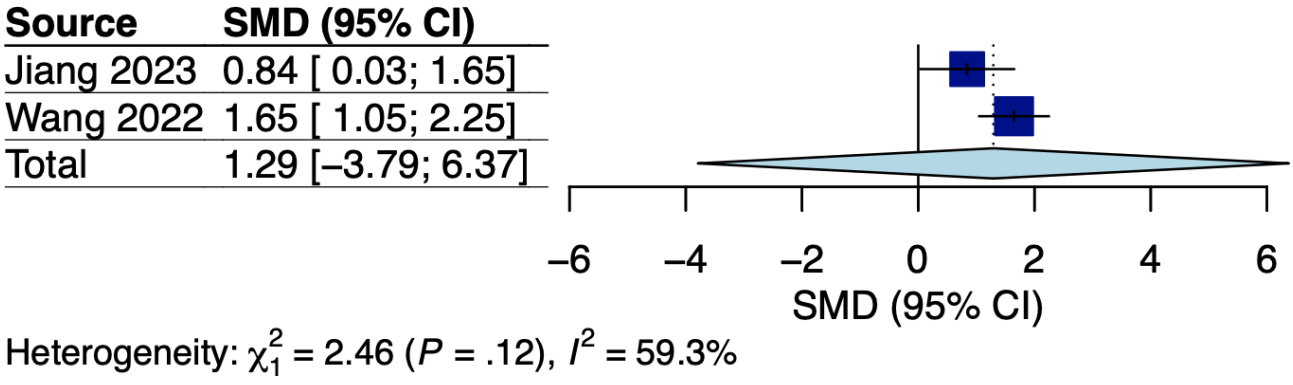


Figure 8. Meta-analysis results for protein intake among gastrointestinal (GI) cancer survivors [34,38]. SMD: standardized mean difference.

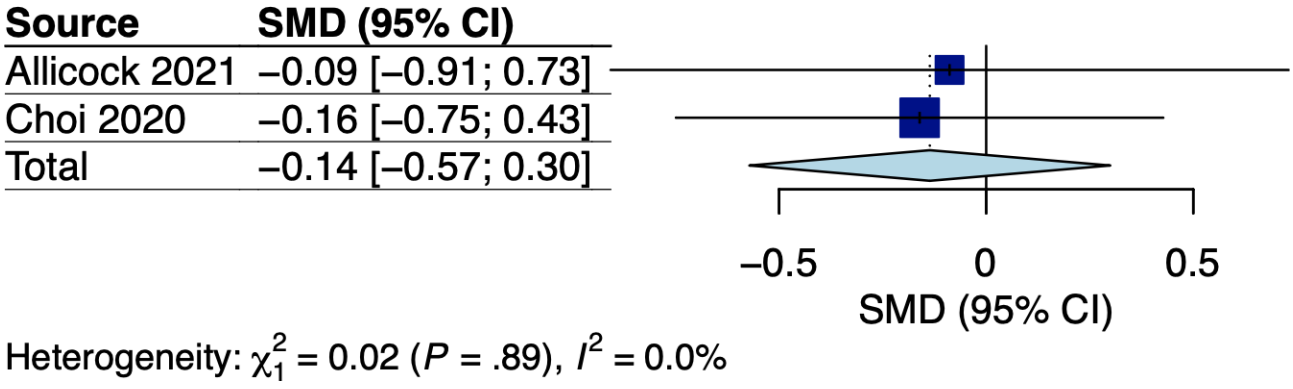


Fast Food Intake

A total of 2 RCTs measured fast food intake using 24-hour recall questionnaires and food frequency questionnaires among breast cancer survivors [22,25], with the respective pooled effect shown in Figure 9. Mobile app–based dietary interventions had no significant effect on fast food intake (Hedges *g*=−0.14, 95% CI −0.57 to 0.30; *P*=.16; I^2=0%; see Figure 9). The certainty of evidence using the GRADE approach was low (see Multimedia Appendix 5).

XSL•FO
RenderX

Figure 9. Meta-analysis results for fast food intake among breast cancer survivors [22,25]. SMD: standardized mean difference.



Vegetable and Fruit Intake

Only 2 of the 4 studies that assessed vegetable and/or fruit intake [22,25,28,41] based on 24-hour recall questionnaires and food frequency questionnaires reported significant between-group differences [25,28]. In the RCT by Choi et al [25], breast cancer mobile app users consumed significantly more weekly vegetables and fruit than the control group (see Table 1). Another quasi-experimental study by Park et al [28] also showed greater vegetable consumption among breast cancer survivors compared to the control group (see Table 1).

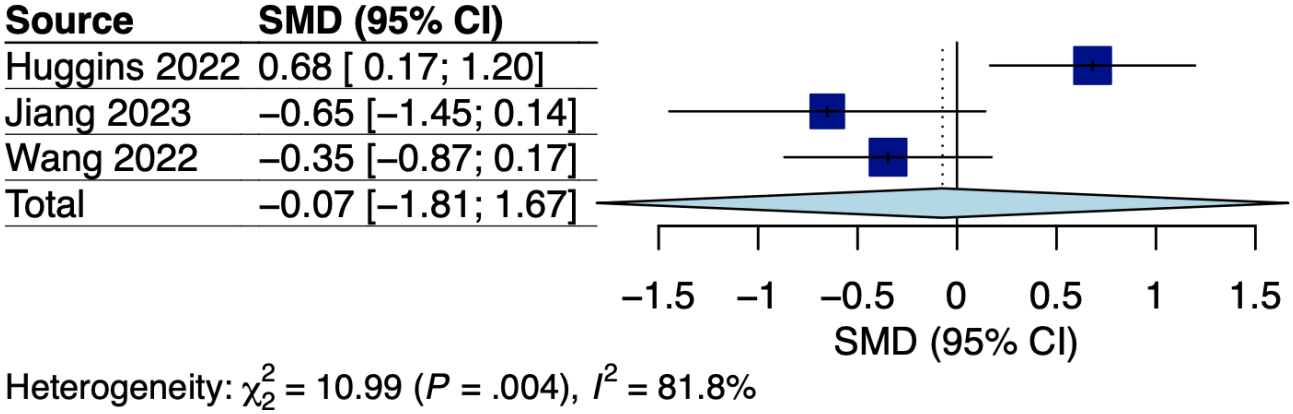
Sugar Intake

An RCT and a single-arm trial measured sugar intake using food frequency questionnaires among hematologic cancer survivors [40,41]. Neither significant between-group changes in sugar intake nor significant within-group changes in sugary drink consumption were reported [40,41].

Nutritional Status

A total of 4 RCTs [33-35,38] and 1 single-arm trial [47] used the Scored Patient-Generated Subjective Global Assessment (PG-SGA) and the prognostic nutritional index (PNI) to assess nutritional status. Given Keum et al [35] did not report the mean PG-SGA scores, we included only 3 RCTs in the pooled analysis, with the corresponding pooled effect shown in Figure 10. Meta-analysis did not show a significant effect on nutritional status among GI cancer survivors with substantial heterogeneity (Hedges g=−0.07, 95% CI −1.81 to 1.67; P=.87; I^2=82%; see Figure 10). However, meta-regression or subgroup analysis was not possible due to the limited number of studies. The certainty of evidence using the GRADE approach was very low (see Multimedia Appendix 5). Although the single-arm trial by Yang et al [39] showed a reduced PNI among esophageal cancer survivors receiving adjuvant chemoradiotherapy, the reduction was significantly less than that compared to those receiving postoperative usual care [47].

Figure 10. Meta-analysis results for nutritional status among gastrointestinal (GI) cancer survivors [33,34,38]. SMD: standardized mean difference.



Discussion

Principal Findings

To our best knowledge, this is the first systematic review and meta-analysis to synthesize the preliminary evidence on the effects of mobile app–based dietary interventions on anthropometric changes, nutritional outcomes, and QoL among cancer survivors. However, existing trials primarily focused on breast and GI cancer survivors. Based on these existing findings, mobile app–based dietary interventions showed potential in improving energy intake and reducing body weight among GI cancer survivors.

Our meta-analysis demonstrated that mobile app–based dietary interventions might vastly improve energy intake, specifically for GI cancer survivors. Notably, our findings on energy intake differ from a previous meta-analysis by Gong et al [17], which found no significant effects of dietary interventions on energy intake among breast, bladder, and endometrial cancer survivors using phone calls, websites, and emails. This suggests that mobile apps can be an effective alternative to other mHealth means, such as websites and emails, for improving energy intake among GI cancer survivors. Nonetheless, it is worth noting the

XSL•FO
RenderX

paucity of available trials precludes definitive conclusions about whether our observed improved energy intake stems from the nature of the dietary intervention. This underscores the need for more future trials to confirm the relative contributions of dietary intervention methodology to energy intake. Furthermore, considering the effects observed in our analysis were assessed over relatively short follow-up durations (mean 4.5 months), future RCTs should evaluate the long-term impact of mobile app–based intervention on energy intake.

Notably, the WeChat app in the study by Wang et al [38] and the iNutrition applet in the study by Jiang et al [34] contributed to the pooled effect on energy intake and demonstrated high feasibility of app use among GI cancer survivors. Specifically, both studies reported attrition rates of less than 10% in the intervention groups, and the mobile app usage experience of iNutrition was rated with a good SUS score greater than 70. In contrast, another study by Huggins and colleagues [33] that also involved GI cancer survivors reported a substantially higher attrition rate (69%) in the intervention group. The differences in attrition rates may be attributed to the adoption of different dietary intervention delivery modes. Huggins and colleagues [33] solely relied on an asynchronous mode, which may have limited the immediate addressing of participants' concerns regarding nutrition impact symptom management. Conversely, Wang et al [38] and Jiang et al [34] used both synchronous and asynchronous modes, allowing participants to choose their preferred delivery mode based on their preferred timing of mobile app use and frequency of dietitian contact via the app. Therefore, future studies targeting GI cancer survivors could explore whether such hybrid delivery modes could better address variability in participants' needs to promote retention.

Although mobile app–based dietary interventions yielded nonsignificant effects on protein intake (Hedges g=1.3) and QoL (Hedges g=2.29), their effect sizes were relatively larger compared to other outcomes. Among the included studies, Wang and colleagues [38] showed substantially greater significant effects on protein intake and QoL than the other studies. Several reasons may explain this superior effect. First, there are differences in the clinical staging of the target population. Wang and colleagues [38] recruited stage I to III colorectal cancer survivors, while Huggins et al [33] and Jiang et al [34] included upper GI cancer patients across all clinical stages. With the inclusion of metastatic stage upper GI cancer survivors, who are more likely to experience cachexia [48], sole dietary management through mobile apps may not be sufficient to improve protein intake and QoL. Multimodal management, including pharmacotherapy, physical activity, and mental support, in addition to dietary interventions, is recommended for cachexia management according to the European Society of Medical Oncology guidelines [49]. Second, there are also differences in participants' treatment statuses. Given nutrition impact symptoms were associated with poorer QoL among oncology patients [50], the upper GI cancer survivors in the studies by Huggins et al [33] and Jiang et al [34] were either receiving active cancer treatments or had just received gastrectomy. They may be suffering from worse nutrition impact symptoms compared to the colorectal cancer survivors in the study by Wang et al [38] who had completed active cancer treatments. Third, psychological and social functioning are essential determinants of cancer survivors' perceived QoL [51], as measured by the global health status score of the EORTC QLQ-C30. Wang and colleagues [38] could have provided better psychological support through monthly dietitian-led dietary consultation video calls, enabling both nonverbal and verbal communication, compared to telephone and text messages, which are restricted to verbal communication. To leverage the potential of dietary mobile app–based interventions on protein intake and QoL, further trials should investigate whether cancer staging, treatment status, and the mode of communication with dietitians influence their effectiveness.

Our meta-analysis found that dietary apps may help reduce weight among GI cancer survivors. Notably, the 2 studies included, conducted by Jiang et al [34] and Huggins et al [33], did not specify their weight change goals. However, a recent systematic review has indicated that GI cancer was associated with overweight and obesity but not underweight [52]. Therefore, the weight reduction observed through the use of dietary apps can be considered beneficial. Nonetheless, additional trials with clearly defined weight loss goals are desirable to strengthen the evidence supporting the effectiveness of dietary apps in promoting favorable weight loss among GI cancer survivors.

Mobile app–based dietary interventions yielded insignificant effects on body weight in breast cancer survivors and insignificant impact on nutritional status in GI cancer survivors due to the positive and negative effects in the included studies. The mixed effects on body weight could be due to the involvement of breast cancer survivors with normal BMI and obesity, as weight loss is only indicated for those with obesity at baseline. On the other hand, the mixed effects on nutritional status could be attributed to the involvement of GI cancer survivors with different cancer treatment statuses and staging. Notably, although one of the included studies by Jiang and colleagues [34] recruited upper GI cancer survivors who may be susceptible to gastrectomy-related side effects, it still demonstrated a superior effect compared to other included studies. This may be due to the engagement of family members and caregivers during the receiving of mobile app–based dietary interventions, given both are considered an essential source of self-efficacy [34] and could have facilitated participants' implementation of dietary management plans. Future trials may consider caregivers' or family members' involvement in supporting mobile app–based dietary interventions.

Only five of the 20 mobile apps reviewed in our study were evaluated on their usability, quality, and satisfaction using standard questionnaires such as the SUS, MARS, and QUIS. All users of these 5 apps, who were breast, gastric, and colorectal cancer survivors with varying treatment statuses and clinical staging, rated these apps highly in terms of usability, quality, and satisfaction. While the QUIS primarily focuses on the acceptability of computer interfaces and the SUS is not exclusively designed for mobile apps, the user version of the MARS offers a tailored evaluation of mobile apps, including engagement, functionality, aesthetics, and information quality [46]. It can be administered to individuals without expertise in mHealth [53], and also assesses the users' perceived impact on

XSL•FO
RenderX

awareness, knowledge, attitudes, and intentions toward behavioral health changes by mobile apps [46]. This aligns with the constructs of "affective attitude," "perceived effectiveness," "intention," and "experience" based on the theoretical framework of acceptability for healthcare interventions [54]. Future trials could consider using the user version of the MARS for a more comprehensive and customized evaluation of mobile apps.

In over half of the studies included in our review, mobile apps incorporated feedback messages [27,40] and dietary goal setting [43], which are considered facilitators of mobile app usage. However, qualitative feedback from participants highlighted concerns related to the time required for app usage [24,26,43] and limited options for logging food entries [26,27]. To address these issues, future mobile apps could consider integrating artificial intelligence (AI)-powered food item recognition during the food logging process. This enhancement could streamline and facilitate the entry of food items, potentially reducing the time spent using the app. In addition, this approach can minimize potential recall bias associated with self-reported food intake data collected through food frequency questionnaires. Another qualitative feedback reported the significant influence of participants' knowledge on their intention to use mobile apps [34]. Considering the lack of cancer-specific dietary recommendations in the mobile apps examined in our review, future trials could consider incorporating education on the summary of evidence regarding the association between dietary patterns and cancer risks provided by the World Cancer Research Fund International and the American Institute for Cancer Research.

Strengths and Limitations

A strength of this study is that the systematic search of this review was not limited by language, and there were no restrictions on the cancer types, treatment statuses, prognoses, weight statuses, and age of our target population. However, several limitations should be acknowledged. First, most of the included studies comprised a small number of participants and were quasi-experimental in nature. To enhance the robustness of our findings, we conducted meta-analyses only on outcomes of interest using postintervention means and SDs from RCTs. Future trials with larger sample sizes are needed to further validate the efficacy of these interventions. Second, despite the high study heterogeneity among studies, conducting subgroup analyses or meta-regressions to examine relevant moderators by study, intervention, and participants' characteristics was not feasible due to the limited number of available trials. In addition, the diverse combinations of app features across studies hindered the identification of specific features that were more effective in promoting the desired outcomes. Future research could isolate specific app features to explore their individual contributions. Third, the studies included in the meta-analyses did not specify weight change goals, leaving uncertainty about whether the observed weight changes were in a favorable direction. However, we attempted to stratify the intervention effect by cancer type and by weight status as determined by the mean baseline BMI. All results on weight and BMI remained the same. Future studies should specify weight change goals to better determine the direction of favorable change. Fourth, the meta-analysis included only a limited number of studies. It would be beneficial to have more research, particularly on effects that are currently considered insignificant but may have potential, such as protein intake. Fifth, the studies incorporated into the meta-analysis focused solely on survivors of breast or GI cancer, which means our findings may not be generalized to survivors of other types of cancer.

Conclusions

In conclusion, this systematic review and meta-analysis provide preliminary evidence for the potential efficacy, feasibility, and acceptability of dietary interventions delivered through mobile apps for cancer survivors. While the existing evidence needs to be strengthened, dietary mobile apps can have a positive impact on energy intake and weight changes among GI cancer survivors. The usability, quality, or satisfaction of all mobile apps evaluated by standardized questionnaires was rated highly by breast, gastric, or colorectal cancer survivors with varying treatment statuses and clinical staging. To optimize the effectiveness of mobile apps for cancer survivors, future mobile apps could retain essential features such as feedback messages and dietary goal setting while incorporating new features such as AI-powered food item recognition during food logging and cancer-specific dietary recommendations. Further trials are necessary to determine whether the effectiveness of intervention varies based on cancer types, staging, treatment status, the mode of communication with dietitians, and the engagement of family or caregivers.

Authors' Contributions

DYTF and NT conceptualized the study. KCWC performed data curation and analysis and prepared the original draft of the manuscript. All authors reviewed, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 28 KB - mhealth_v13i1e65505_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

Additional study characteristics.
[DOCX File, 44 KB - mhealth_v13i1e65505_app2.docx]

Multimedia Appendix 3
Risk of bias assessments.
[DOCX File, 448 KB - mhealth_v13i1e65505_app3.docx]

Multimedia Appendix 4
Summary of usability, quality, or satisfaction of mobile app use.
[DOCX File, 32 KB - mhealth_v13i1e65505_app4.docx]

Multimedia Appendix 5
Grading of Recommendations Assessment, Development, and Evaluation (GRADE) assessment.
[DOCX File, 40 KB - mhealth_v13i1e65505_app5.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File, 36 KB - mhealth_v13i1e65505_app6.docx]

References

1. Global cancer burden growing, amidst mounting need for services. World Health Organization. 2024. URL: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services [accessed 2024-08-14]
2. Emery J, Butow P, Lai-Kwon J, Nekhlyudov L, Rynderman M, Jefford M. Management of common clinical problems experienced by survivors of cancer. Lancet 2022 Apr 16;399(10334):1537-1550. [doi: 10.1016/S0140-6736(22)00242-2] [Medline: 35430021]
3. Harrison JD, Young JM, Price MA, Butow PN, Solomon MJ. What are the unmet supportive care needs of people with cancer? A systematic review. Support Care Cancer 2009 Aug;17(8):1117-1128. [doi: 10.1007/s00520-009-0615-5] [Medline: 19319577]
4. Swain CTV, Nguyen NH, Eagles T, et al. Postdiagnosis sedentary behavior and health outcomes in cancer survivors: a systematic review and meta-analysis. Cancer 2020 Feb 15;126(4):861-869. [doi: 10.1002/cncr.32578] [Medline: 31714596]
5. Tao L, Wang R, Gao YT, Yuan JM. Impact of postdiagnosis smoking on long-term survival of cancer patients: the Shanghai cohort study. Cancer Epidemiol Biomarkers Prev 2013 Dec;22(12):2404-2411. [doi: 10.1158/1055-9965.EPI-13-0805-T] [Medline: 24319070]
6. Schwedhelm C, Boeing H, Hoffmann G, Aleksandrova K, Schwingshackl L. Effect of diet on mortality and cancer recurrence among cancer survivors: a systematic review and meta-analysis of cohort studies. Nutr Rev 2016 Dec;74(12):737-748. [doi: 10.1093/nutrit/nuw045] [Medline: 27864535]
7. Cancer care settings and smoking cessation. US Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/tobacco/patient-care/care-settings/cancer/index.htm [accessed 2024-08-14]
8. Cancer prevention recommendations. World Cancer Research Fund International. URL: https://www.wcrf.org/diet-activity-and-cancer/cancer-prevention-recommendations/ [accessed 2024-08-14]
9. I have been diagnosed with cancer. Will eating a healthy diet be beneficial to me? International Agency for Research on Cancer. URL: https://cancer-code-europe.iarc.fr/index.php/en/ecac-12-ways/diet-recommendation/66-healthy-diet-beneficial [accessed 2024-08-14]
10. Petrelli F, Cortellini A, Indini A, et al. Association of obesity with survival outcomes in patients with cancer: a systematic review and meta-analysis. JAMA Netw Open 2021 Mar 1;4(3):e213520. [doi: 10.1001/jamanetworkopen.2021.3520] [Medline: 33779745]
11. Pati S, Irfan W, Jameel A, Ahmed S, Shahid RK. Obesity and cancer: a current overview of epidemiology, pathogenesis, outcomes, and management. Cancers (Basel) 2023 Jan 12;15(2):485. [doi: 10.3390/cancers15020485] [Medline: 36672434]
12. Arends J, Bachmann P, Baracos V, et al. ESPEN guidelines on nutrition in cancer patients. Clin Nutr 2017 Feb;36(1):11-48. [doi: 10.1016/j.clnu.2016.07.015] [Medline: 27637832]
13. Castro-Espin C, Agudo A. The role of diet in prognosis among cancer survivors: a systematic review and meta-analysis of dietary patterns and diet interventions. Nutrients 2022 Jan 14;14(2):348. [doi: 10.3390/nu14020348] [Medline: 35057525]
14. Greenlee H, Santiago-Torres M, McMillen KK, Ueland K, Haase AM. Helping patients eat better during and beyond cancer treatment: continued nutrition management throughout care to address diet, malnutrition, and obesity in cancer. Cancer J 2019;25(5):320-328. [doi: 10.1097/PPO.0000000000000405] [Medline: 31567459]
15. Burden S, Jones DJ, Sremanakova J, et al. Dietary interventions for adult cancer survivors. Cochrane Database Syst Rev 2019 Nov 22;2019(11):CD011287. [doi: 10.1002/14651858.CD011287.pub2] [Medline: 31755089]

XSL•FO
RenderX

16. LeVasseur N, Cheng W, Mazzarello S, et al. Optimising weight-loss interventions in cancer patients-a systematic review and network meta-analysis. PLoS ONE 2021;16(2):e0245794. [doi: 10.1371/journal.pone.0245794] [Medline: 33539414]
17. Gong Y, Jiang X, Chen X, et al. Effectiveness of mHealth diet interventions in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Asia Pac J Oncol Nurs 2023 Mar;10(3):100196. [doi: 10.1016/j.apjon.2023.100196] [Medline: 37124242]
18. Wang L, Langlais CS, Kenfield SA, et al. mHealth interventions to promote a healthy diet and physical activity among cancer survivors: a systematic review of randomized controlled trials. Cancers (Basel) 2022 Aug 6;14(15):3816. [doi: 10.3390/cancers14153816] [Medline: 35954479]
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. J Clin Epidemiol 2021 Jun;134:178-189. [doi: 10.1016/j.jclinepi.2021.03.001] [Medline: 33789819]
20. Higgins JPT, Thomas J, Chandler J, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 64: John Wiley & Sons; 2023. URL: https://training.cochrane.org/handbook [accessed 2024-08-15]
21. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
22. Allicock M, Kendzor D, Sedory A, et al. A pilot and feasibility mobile health intervention to support healthy behaviors in African American breast cancer survivors. J Racial Ethn Health Disparities 2021 Feb;8(1):157-165. [doi: 10.1007/s40615-020-00767-x] [Medline: 32385847]
23. Baik SH, Oswald LB, Buscemi J, et al. Patterns of use of smartphone-based interventions among Latina breast cancer survivors: secondary analysis of a pilot randomized controlled trial. JMIR Cancer 2020 Dec 8;6(2):e17538. [doi: 10.2196/17538] [Medline: 33289669]
24. Cairo J, Williams L, Bray L, Goetzke K, Perez AC. Evaluation of a mobile health intervention to improve wellness outcomes for breast cancer survivors. J Patient Cent Res Rev 2020;7(4):313-322. [doi: 10.17294/2330-0698.1733] [Medline: 33163551]
25. Choi JH, Park SJ, Kwon H, Lee HJ. Application and evaluation of mobile nutrition management service for breast cancer patients. J Nutr Health 2020;53(1):83. [doi: 10.4163/jnh.2020.53.1.83]
26. Lim JY, Kim Y, Yeo SM, Chae BJ, Yu J, Hwang JH. Feasibility and usability of a personalized mHealth app for self-management in the first year following breast cancer surgery. Health Informatics J 2023;29(1):14604582231156476. [doi: 10.1177/14604582231156476] [Medline: 36772832]
27. Lozano-Lozano M, Cantarero-Villanueva I, Martin-Martin L, et al. A mobile system to improve quality of life via energy balance in breast cancer survivors (BENECA mhealth): prospective test-retest quasiexperimental feasibility study. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14136. [doi: 10.2196/14136] [Medline: 31237570]
28. Park G, Lee H, Kim HY. The effects of utilizing smartphone application peer support on health behavior and body mass index among breast cancer survivors. J Korean Acad Nurs 2019 Oct;49(5):550-561. [doi: 10.4040/jkan.2019.49.5.550] [Medline: 31672949]
29. Seo SJ, Nho JH, Park Y. The development of a lifestyle modification mobile application, “Health for You” for overweight and obese breast cancer survivors in Korea. Korean J Women Health Nurs 2021 Sep 30;27(3):243-255. [doi: 10.4069/kjwhn.2021.09.14] [Medline: 36311982]
30. Stubbins R, He T, Yu X, et al. A behavior-modification, clinical-grade mobile application to improve breast cancer survivors' accountability and health outcomes. JCO Clin Cancer Inform 2018 Dec;2(2):1-11. [doi: 10.1200/CCI.18.00054] [Medline: 30652617]
31. McCarroll ML, Armbruster S, Pohle-Krauza RJ, et al. Feasibility of a lifestyle intervention for overweight/obese endometrial and breast cancer survivors using an interactive mobile application. Gynecol Oncol 2015 Jun;137(3):508-515. [doi: 10.1016/j.ygyno.2014.12.025] [Medline: 25681782]
32. Cheng C, Ho RTH, Guo Y, et al. Development and feasibility of a mobile health-supported comprehensive intervention model (CIMmH) for improving the quality of life of patients with esophageal cancer after esophagectomy: prospective, single-arm, nonrandomized pilot study. J Med Internet Res 2020 Aug 18;22(8):e18946. [doi: 10.2196/18946] [Medline: 32808933]
33. Huggins CE, Hanna L, Furness K, et al. Effect of early and intensive telephone or electronic nutrition counselling delivered to people with upper gastrointestinal cancer on quality of life: a three-arm randomised controlled trial. Nutrients 2022 Aug 7;14(15):3234. [doi: 10.3390/nu14153234] [Medline: 35956410]
34. Jiang X, Chen J, Yuan X, et al. Feasibility of an individualized mHealth nutrition (iNutrition) intervention for post-discharged gastric cancer patients following gastrectomy: a randomized controlled pilot trial. Nutrients 2023 Apr 13;15(8):1883. [doi: 10.3390/nu15081883] [Medline: 37111102]
35. Keum J, Chung MJ, Kim Y, et al. Usefulness of smartphone apps for improving nutritional status of pancreatic cancer patients: randomized controlled trial. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e21088. [doi: 10.2196/21088] [Medline: 34463630]
36. Salmani H, Nahvijou A, Sheikhtaheri A. Smartphone-based application for self-management of patients with colorectal cancer: development and usability evaluation. Support Care Cancer 2022 Apr;30(4):3249-3258. [doi: 10.1007/s00520-021-06754-0] [Medline: 34984548]

XSL•FO
RenderX

37. Soh JY, Cha WC, Chang DK, et al. Development and validation of a multidisciplinary mobile care system for patients with advanced gastrointestinal cancer: interventional observation study. JMIR Mhealth Uhealth 2018 May 7;6(5):e115. [doi: 10.2196/mhealth.9363] [Medline: 29735478]
38. Wang X, Zeng H, Li L, et al. Personalized nutrition intervention improves nutritional status and quality of life of colorectal cancer survivors in the community: a randomized controlled trial. Nutrition 2022;103-104:111835. [doi: 10.1016/j.nut.2022.111835] [Medline: 36183482]
39. Yang K, Oh D, Noh JM, et al. Feasibility of an interactive health coaching mobile app to prevent malnutrition and muscle loss in esophageal cancer patients receiving neoadjuvant concurrent chemoradiotherapy: prospective pilot study. J Med Internet Res 2021 Aug 27;23(8):e28695. [doi: 10.2196/28695] [Medline: 34448714]
40. Chow EJ, Doody DR, Di C, et al. Feasibility of a behavioral intervention using mobile health applications to reduce cardiovascular risk factors in cancer survivors: a pilot randomized controlled trial. J Cancer Surviv 2021 Aug;15(4):554-563. [doi: 10.1007/s11764-020-00949-w] [Medline: 33037989]
41. Fuemmeler BF, Holzwarth E, Sheng Y, et al. Mila Blooms: a mobile phone application and behavioral intervention for promoting physical activity and a healthy diet among adolescent survivors of childhood cancer. Games Health J 2020 Aug;9(4):279-289. [doi: 10.1089/g4h.2019.0060] [Medline: 32391734]
42. Yang HC, Chung SH, Yoo JS, Park B, Kim MS, Lee JM. Evaluation of a smart after-care program for patients with lung cancer: a prospective, single-arm pilot study. J Chest Surg 2022 Apr 5;55(2):108-117. [doi: 10.5090/jcs.21.138] [Medline: 35135903]
43. Orlemann T, Reljic D, Zenker B, et al. A novel mobile phone app (OncoFood) to record and optimize the dietary behavior of oncologic patients: pilot study. JMIR Cancer 2018 Nov 20;4(2):e10703. [doi: 10.2196/10703] [Medline: 30459139]
44. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
45. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
46. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
47. Yoon HG, Oh D, Ahn YC, et al. Prognostic impact of sarcopenia and skeletal muscle loss during neoadjuvant chemoradiotherapy in esophageal cancer. Cancers (Basel) 2020 Apr 10;12(4):925. [doi: 10.3390/cancers12040925] [Medline: 32290037]
48. Poulia KA, Sarantis P, Antoniadou D, et al. Pancreatic cancer and cachexia-metabolic mechanisms and novel insights. Nutrients 2020 May 26;12(6):1543. [doi: 10.3390/nu12061543] [Medline: 32466362]
49. Arends J, Strasser F, Gonella S, et al. Cancer cachexia in adult patients: ESMO Clinical Practice Guidelines☆. ESMO Open 2021 Jun;6(3):100092. [doi: 10.1016/j.esmoop.2021.100092] [Medline: 34144781]
50. Tong H, Isenring E, Yates P. The prevalence of nutrition impact symptoms and their relationship to quality of life and clinical outcomes in medical oncology patients. Support Care Cancer 2009 Jan;17(1):83-90. [doi: 10.1007/s00520-008-0472-7] [Medline: 18551322]
51. Jitender S, Mahajan R, Rathore V, Choudhary R. Quality of life of cancer patients. J Exp Ther Oncol 2018 May;12(3):217-221. [Medline: 29790313]
52. Azizi N, Zangiabadian M, Seifi G, et al. Gastric cancer risk in association with underweight, overweight, and obesity: a systematic review and meta-analysis. Cancers (Basel) 2023 May 16;15(10):2778. [doi: 10.3390/cancers15102778] [Medline: 37345115]
53. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022 Jan 27;22(1):22. [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]
54. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88. [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]

Abbreviations

AI: artificial intelligence
GI: gastrointestinal
GRADE: Grading of Recommendations Assessment, Development and Evaluation
MARS: Mobile Application Rating Scale
mHealth: mobile health
PG-SGA: Scored Patient-Generated Subjective Global Assessment
PNI: prognostic nutritional index
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

QLQ-C30: European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire
QoL: quality of life
QUIS: Questionnaire for User Interface Satisfaction
RCT: randomized controlled trial
RoB: Risk of bias
ROBINS-I: Risk of Bias in Nonrandomized Studies
SMD: standardized mean difference
uMARS: user version of the Mobile Application Rating Scale

Edited by L Buis; submitted 18.08.24; peer-reviewed by PH Lin, W Demark-Wahnefried; revised version received 24.04.25; accepted 04.06.25; published 31.07.25.

Please cite as:
Chung KCW, Takemura N, Lam WWT, Ho MM, Lee AM, Chan WYY, Fong DYT
Efficacy of Mobile App–Based Dietary Interventions Among Cancer Survivors: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2025;13:e65505
URL: https://mhealth.jmir.org/2025/1/e65505
doi:10.2196/65505

© Krista Ching Wai Chung, Naomi Takemura, Wendy Wing Tak Lam, Mandy Man Ho, Antoinette Marie Lee, Wynnie Yuen Yee Chan, Daniel Yee Tak Fong. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Experience of Using Electronic Inhaler Monitoring Devices for Patients With Chronic Obstructive Pulmonary Disease or Asthma: Systematic Review of Qualitative Studies

Jilong Duan[1*], MSc; Xia Chen[1*], MSN; Di Fan[1], MSc; Haikun Jiang[1], MSc; Xue Zhang[1], MSc; Wenyue Zhang[1], PhD; Zhiping Liu[1], MSc; Hongyan Lu[2], PhD

[1]Department of Nursing, Ningxia Medical University, Yinchuan, China

[2]Department of Nursing, General Hospital of Ningxia Medical University, No. 804 Shengli Street, Xingqing District, Yinchuan, China

[*]these authors contributed equally

Corresponding Author:
Hongyan Lu, PhD
Department of Nursing, General Hospital of Ningxia Medical University, No. 804 Shengli Street, Xingqing District, Yinchuan, China

Abstract

Background: Electronic inhaler monitoring devices (EIMDs) can enhance medication adherence in patients with chronic obstructive pulmonary disease (COPD) and asthma, yet patient perceptions and experiences with these devices vary widely. A systematic qualitative synthesis is required to comprehensively understand patient perspectives on EIMDs, to lay the foundation for developing strategies to improve patient compliance.

Objective: This study aims to systematically evaluate qualitative studies on the experiences of patients with COPD and asthma using EIMDs, providing insights to support their clinical application and improve patient engagement.

Methods: This review synthesized qualitative data from reports found through a systematic search of PubMed, Web of Science, CINAHL, Embase, Cochrane Library, and PsycInfo from January 1983 to July 2024. The reports assessed patient experiences with EIMDs for COPD and asthma. The quality of the included reports was appraised using the Critical Appraisal Skills Program criteria developed by the Centre for Evidence-Based Medicine, University of Oxford, UK.

Results: A total of 7 reports were included, encompassing data from 44 patients with COPD and 146 with asthma. Findings were organized into 9 sub-themes and 3 themes: positive experiences with EIMDs (usability and easy acceptance, enhanced self-management); stresses and challenges of using these devices (negative emotional stress, device trust issues, social difficulties, economic burdens, and technical challenges); and patient expectations from these devices (expectations related to device construction and function and external support).

Conclusions: Patients have positive experiences using electronic monitoring devices for inhalation devices but also face various social, psychological, and technical challenges. Health care workers should consider patient experiences with EIMDs to tailor these devices to patient needs, ultimately enhancing device acceptance and adherence. Further research should focus on increasing EIMDs convenience and usability for patients with COPD and asthma.

Trial Registration: PROSPERO CRD42023480463; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023480463

(JMIR Mhealth Uhealth 2025;13:e57645) doi:10.2196/57645

KEYWORDS

asthma; drug inhalation; electronic monitoring devices; chronic obstructive pulmonary disease; qualitative research; systematic review

Introduction

Respiratory diseases pose a global health challenge, with chronic obstructive pulmonary disease (COPD) and asthma among the most prevalent types [1]. These conditions result in significant morbidity and mortality worldwide, imposing a substantial and growing health burden [2,3].

While severe respiratory diseases are often incurable, inhaled medications can prevent acute exacerbations [4]. Inhalation therapy, acting directly on the lungs, offers advantages such as rapid onset, low dosage requirements, and minimal side effects [5], making it the primary treatment approach for respiratory conditions like COPD and asthma [6,7].

The efficacy of inhalation therapy relies on the correct use of inhalers [8,9], yet many patients with COPD and asthma struggle

XSL•FO
RenderX

with proper technique. Studies report that 4% - 94% of patients do not use inhalation devices correctly [10], which limits the drug's effectiveness and can lead to poor disease control [11].

Historically, health care providers often use checklists to assess the effectiveness of a patient's inhalation technique [12]. Although checklists are cost-effective and easy to administer [13], they depend on the skill and knowledge of health care professionals [14,15]. Plaza et al's [16] questionnaire survey of 1514 practicing physicians and Giner et al's [17] cross-sectional study of 1496 nurses both found that only approximately 14% of them possessed adequate knowledge about inhalation therapy, suggesting widespread problems with the use of inhalation equipment and a lack of health literacy related to inhalation techniques among health care professionals. This gap in knowledge hinders the ability of medical staff to consistently evaluate the effectiveness of inhalation therapy through standardized assessment tools, thereby compromising the reliability of these evaluations.

A solution to the above problem is electronic monitoring [18]. The International Healthcare Membership Organization defines eHealth as the practice of health care where health care professionals, with the assistance of information engineers, using electronic information or communication technologies to provide health care services and information to patients. eHealth includes forms of applications such as telemedicine, mobile health (mHealth), electronic testing devices, and social media software [19]. Electronic inhaler monitoring devices (EIMDs), also known as smart inhalers, can objectively monitor the adherence and inhalation technical ability of patients with COPD or asthma through built-in sensors, external mobile apps, and other software or hardware [20], as well as provide objective data feedback to patients and health care professionals [21].

The use of EIMDs can be efficacious in improving medication adherence and inhalation techniques, but patient perceptions and experiences while using these devices differ [22]. A meta-analysis by Garin et al [23], found that compared with traditional care, EIMDs can significantly improve medication compliance and inhalation techniques in patients with COPD and asthma. However, each study included in the meta-analysis differed in its approach to intervention. In addition, several published qualitative studies have reported on the experiences of people with COPD or asthma while using EIMDs [24-26], including patient usability and acceptability of the devices, technical challenges in their application by patients, and their costs. However, significant gaps remain in existing qualitative research findings [25]. Diverse cultural, health care, and educational backgrounds may lead to unique patient experiences with EIMDs across different countries [25,27]. Consequently, isolated qualitative studies cannot provide a comprehensive understanding of patient perspectives on these devices for COPD and asthma. Incorporating patient insights may help enhance the functional development of EIMDs [24].

This systematic review aimed to synthesize qualitative studies to create a more comprehensive understanding of the patient experience with EIMDs for COPD and asthma.

Methods

Design

This systematic review used Thomas and Harden's thematic synthesis approach to identify key themes from qualitative data [25]. This method enables the development of new insights based on previous findings. The review protocol is registered with PROSPERO (International Prospective Register of Systematic Reviews) (CRD42023480463).

Study Selection Criteria

The inclusion criteria were based on the PICoS (Population, Phenomenon of interest, Context, Types of studies) principle. Population: our study population was adult patients with COPD or asthma aged≥18 years. Phenomenon of interest: this systematic review will explore the experiences and perspectives of patients with COPD or asthma who have received EIMDs for inhalation devices. Context: this systematic review considers patient experiences and perspectives with inhaler EIMDs for COPD or asthma, regardless of whether they live at home, in the hospital, in the community, or in other health care settings, as well as regardless of their cultural backgrounds. Types of studies: this systematic review considers all types of qualitative research as well as the qualitative component of mixed studies. We excluded reports that did not provide patient citations and those that did not provide full texts (eg, conference abstracts), as well as study protocols that had not yet been conducted. In addition, we excluded non-English language reports. Specific inclusion and exclusion criteria are shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Review inclusion and exclusion criteria.

Inclusion Criteria

- Reports in which the study population comprised patients with chronic obstructive pulmonary disease (COPD) or asthma.
- Reports in which the patients are≥18 years.
- Reports on patients who have had experience or feelings of using electronic inhaler monitoring devices (EIMDs).
- Qualitative research or the qualitative part of mixed research.
- Reports published in English.

Exclusion Criteria

- Unpublished articles that have not been peer-reviewed.
- In order to better evaluate the quality of the authors' interpretation and analysis of the data, reports that did not report citations from patients were excluded.
- If the content of an article involved patients with COPD or asthma but did not analyze the data of these patients, the article will be excluded.
- Conference abstracts, quantitative studies, literature reviews, and reports that have not yet been conducted will also be excluded.

Search Strategy

A comprehensive search was conducted in 6 databases—PubMed, Web of Science, CINAHL, Embase, Cochrane Library, and PsycInfo—to identify qualitative studies on the experiences of patients with COPD or asthma using EIMDs. Electronic monitoring equipment for inhalation devices was first reported in 1983 [28]. To maximize the inclusion of relevant reports, this search covered publications from January 1983 through July 2024. The search terms were formulated according to the PICoS principles and included key terms such as “Pulmonary Disease, Chronic Obstructive/Asthma”, “Electronic/Monitoring/Sensing/passive monitoring/inhaler monitoring/electronic medication monitor/electronic medication/monitoring sensors/medication monitoring”, “Nebulizers and Vaporizers/Inhalers/Inhalator/Inhalation Device/Administration, Inhalation/Drug Administration, Respiratory/Drug Administration”, and “Inhalation interview*/experience*/qualitative”. The search strategy incorporated trade names of existing EIMDs to ensure comprehensive coverage, as recommended by Garin [23] and Kikidis [29]. Details of the search strategy for each database are provided in Multimedia Appendix 1.

Study Screening and Data Extraction

A total of 2 researchers (DF, ZPL), trained in evidence-based nursing, independently conducted the screening and data extraction processes, adhering to the established inclusion and exclusion criteria. In cases of disagreement, a third researcher (WYZ) was consulted to reach a consensus. Initially, all retrieved reports were imported into EndNote X9 (Clarivate) for deduplication. Titles and abstracts were reviewed to exclude irrelevant studies, followed by a full-text review to confirm the final set of included reports. Data extracted from each report included author details, publication date, country, research method, study population, primary findings, and patient quotations. When reports included perspectives from broader health care interest groups (eg, physicians, nurses, pharmacists, or respiratory therapists), only patient-specific data were extracted for analysis.

Assessment of Methodological Quality

The methodological quality of included reports was assessed independently by 2 researchers (XC, HKJ) using the Critical Appraisal Skills Programme (CASP) checklist developed by the Centre for Evidence-Based Medicine at the University of Oxford, UK [30]. This tool, focusing on evaluating the validity, utility, and reliability of qualitative research, comprises 10 items evaluated with “yes,” “no,” or “unclear” responses. The CASP checklist enabled the identification of strengths and limitations in each report.

Data Analysis

Extracted data were imported into NVivo 11.0 software (QSR International) for thematic synthesis, following the method recommended by Thomas and Harden [31]. This approach is conducive to the development of theoretical and conceptual insights applicable to clinical research planning [32]. For the thematic synthesis, data were merged from the included reports, enabling researchers to identify salient themes from each primary report. The synthesis was conducted in 3 stages: first, the 2 researchers (JLD, XZ) independently coded the extracted data from each report line by line. In the second phase, these initial codes were used to construct “descriptive” themes. Finally, in the third phase, the descriptive themes were iteratively examined, aggregated, and generalized to further form “analytical” themes. Any disagreements between the researchers during this process were resolved through discussion with the third researcher (HYL) to reach a consensus and finalize the findings.

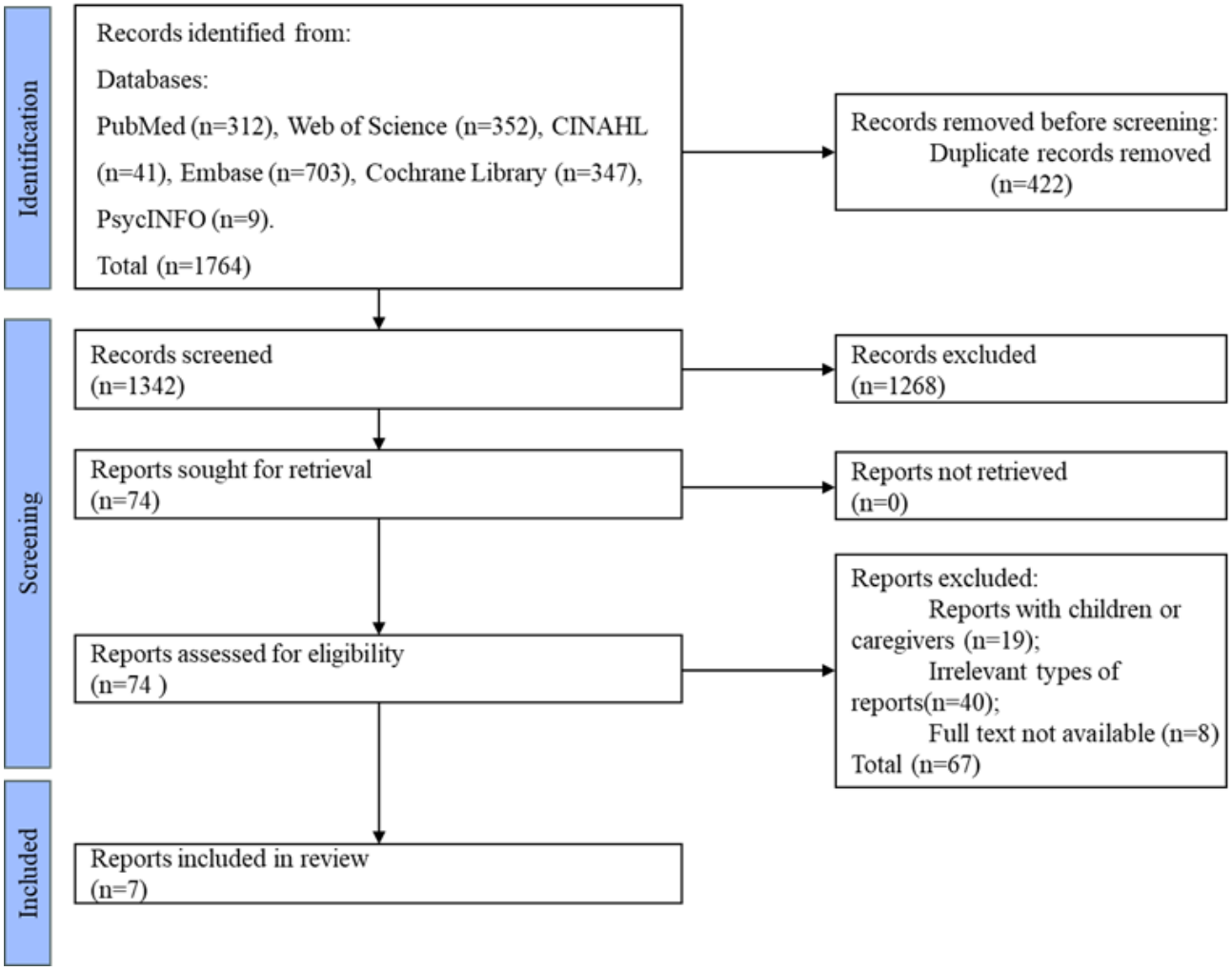
Results

Search Results

The initial database search yielded 1764 articles. After removing 422 duplicates, 1342 articles remained. Following a review of titles and abstracts, 74 potentially relevant articles were selected for further assessment. Full-text analysis led to the final inclusion of 7 articles, all in English. No articles were excluded based on methodological quality assessment. The screening process is illustrated in Figure 1 (Checklist 1).

XSL•FO
RenderX

Figure 1. Reports screening flowchart.



Study Characteristics

In total, 7 reports met the inclusion criteria, comprising 4 qualitative studies (57%) and 3 mixed-method studies (43%). Among them 4 of the included reports originated from the United Kingdom (57%), while the remaining 3 were conducted in the Netherlands (14%), Portugal (14%), and Australia (14%). The study population included 44 patients with COPD and 146 patients with asthma. Data were primarily collected through semistructured interviews (n=5), with two reports using focus groups. Additional demographic details are presented in Table 1.

XSL•FO
RenderX

Table . Basic characteristics of the included reports (n=7).

Author Year Country	Study design	Aim	Participants' details (Sample size Gender Mean age or age range)	Data collection methods Data analysis methods	Main results
Van et al [24] 2023 Netherlands	Qualitative study	Identification of expected facilitators and barriers related to the implementation of smart inhalers	• Patients with asthma (n=9) • 9 female • Mean age 34.7 (13.3)	• Focus group • Inductive approach	5 themes: • Perceived benefits; usability • Feasibility • Payment and reimbursement • Data security and ownership
Hesso et al [33] 2023 England	Mixed Study	Understanding patients' perceptions and acceptability of EIMDs	• Patients with COPD (n=12); Patients with asthma (n=6) • 10 female; 8 male • Mean (SD) age: 64.5 (20.3) years	• Semistructured interviews • Inductive or deductive approaches	4 themes: • Acceptability of EIMDs technology • Patients misconceptions about the use of EIMDs • Acceptability of personalized EIMDs feedback • Positive perceptions of tailored consultations
Hui et al [34] 2022 England	Mixed Study	Understanding patient preferences for documenting asthma and the difficulties encountered in connecting EIMDs to the system to record data	• Patients with asthma (n=8) • 4 female; 4 male • Age range: 26 - 65 years	• Semistructured interviews • Framework analysis	4 themes: • performance expectations; • effort expectations; • social impacts; • facilitating conditions
Adejumo et al [25] 2022 England	Qualitative study	Understanding patients' perceptions and experiences with EIMDs	• Patients with asthma (n=28) • 19 female; 9 male • Mean age IQR: 46.7 (33.5, 54.2) years	• Semistructured interviews • Framework approach	5 themes: • Participants' experiences of asthma • Participants' experiences of asthma treatment • Participants' experiences of involvement in research and use of EIMDs • Future applications of EIMDs - potential improvements and uses • Future applications of EIMDs - desirability, ethics and wider implications
Jácome et al [35] 2021 Portugal	Mixed Study	Understanding the shortcomings of EIMDs and their adapted applications and suggestions for improvement	• Patients with asthma (n=77) • Unclear • Unclear	• Semistructured interviews • Thematic qualitative analysis	4 themes: • Drug-related characteristics • Gamification and social network • Symptom monitoring and physician communication • Others

XSL•FO
RenderX

Author Year Country	Study design	Aim	Participants' details (Sample size Gender Mean age or age range)	Data collection methods Data analysis methods	Main results
Foster et al [26] 2017 Australia	Qualitative study	To explore patients' perceptions of barriers, facilitators, usefulness and impact of using the monitor and its reminders	• Patients with asthma (n=18) • 9 female; 9 male • Age range: 18-68 years • Mean age: 39 years	• Semistructured interviews • Thematic approach	3 themes: • Feasibility and acceptability • Utility and behavioral impact of reminders • Sustainability
Kayyali et al [36] 2016 England	Qualitative study	Understanding the perceptions of people with COPD regarding holistic telemedicine systems and monitoring of inhalation techniques	• Patients with COPD (n=32) • 15 female; 17 male • Unclear	• Focus group • Inductive or deductive approaches	7 themes: • Fragmented care • Poor medication adherence • Many reasons why patients are not adherent to their medications • Poor mental health of patients • Limited health care resources available to patients • Reported satisfaction • Others

Methodological Quality of the Included Reports

All included reports met the quality standards of the Critical Appraisal Skills Programme (CASP) checklist, with no items marked as "no" or "unclear". The results of the quality assessment are detailed in Multimedia Appendix 2.

Synthesis of Research Results

Through a line-by-line coding of the original qualitative data, 41 initial codes were generated. After repeated analysis and comparison, these codes were synthesized into 27 findings, which were organized into nine sub-themes and three overarching themes: (1) positive experiences with EIMDs for patients with COPD or asthma; (2) stresses and challenges associated with using these devices; and (3) patient expectations regarding EIMDs. A summary of the synthesis results is provided in Textbox 2.

XSL•FO
RenderX

Textbox 2. Thematic analysis and patient quotes from the included reports.

1. Positive experiences with EIMDs for COPD or asthma patients

- Strong usability and easy acceptance

"It is an easy tool to use" [24]. "It is easy to use" [26]."Recording any inhaler use and reading peak flows would not require much effort for me" [34]; "Both my family doctor and I can see the data that is being recorded" [25]; "The app includes a demonstration on how to use the inhaler with videos" [35].

"I liked it just the way it was" [26]; "It (referring to the chart) is just easier to read" [33]; "It just fits in my pocket" [26]; "I would pay 200- - 400 Australian Dollars for long-time use" [26].

- Improve self-management ability

"With the data, I have a better understanding of my true adherence" [35]; "The feedback data told me I was using the wrong inhalation device" [33]; "It made me realize even more that now I have to use it correctly every morning and evening" [33]. "When data is shared, we can discuss how it is going. Is it possible to change the dose of inhaled medication?" [24]. "I can control it myself now" [26].

2. Stresses and challenges of using EIMDs in patients with COPD or asthma

- Negative emotional stress

"I would like the data generated to be stored on the NHS for security or anonymously stored on the manufacturer's servers and subject to NHS regulation" [25]; "I think sometimes healthcare professionals receive a lot of useless information or even wrong data" [24]. "I don''t like being monitored by other people" [24]. "I don''t need it to remind me because I know I''m going to take it every day and I''d rather do it in my own time" [26].

- Lack of trust in EIMDs

"I don''t think the data it monitors is accurate" [33]; "The default on the system is not the dose I always use" [34]. "It's best to chat with my doctor to clarify questions and schedule appointments" [35]; "I think contacting the nurse is easiest in general practice" [24].

- Social difficulties

"The buttons on the monitor are too small", "The monitor is unattractive and monochromatic, I don''t want to use it in public anymore" [26]; "the reminder goes off and if you go out you have to find a place to hide in a corner to take it out" [26]; "People in the house usually laugh and say 'you''re dropping your stuff', I know they do it to help but I''ll get there" [26].

- Economic burden

"If it's very expensive, I don''t think I''ll use it because my asthma medication is at a price" [24];

- Device technical challenges

"I need to record some data manually sometimes" [34]. "We face some challenges in setting up the inhaler" [34,36]; "I had two breakdowns while using it and I didn''t know what to do" [26]. "The app could be smaller" [35]. "Functions like some logins could be simplified" [35].

3. Expectations from EIMDs for patients with COPD or asthma

- Expectations of device construction and function

"Some of the devices are big and not portable" [25]; "If the device could track someone's movements, what the weather was like, so that you might get a better understanding of asthma symptoms". "But if it takes more than ten minutes every day then I don''t have enough time" [24].

- Expectations of external support for COPD or asthma patients

"The monitoring device didn''t give much advice. Honestly, I think the doctor's advice was more helpful because she seemed to understand the problem better" [26]. "My fiancé is very, very, very worried about my asthma ..... I take my medication diligently to alleviate his worries" [26]. "I just wish this could be generalized to more patients with asthma" [33].

Positive Experiences With EIMDs for Patients With COPD or Asthma

Studies indicated that patients with COPD or asthma found EIMDs simple, easy to use, and practical [24,33,34]. Some patients appreciated features like medication reminders [26,33,34], data recording for inhalation techniques [34], data visualization and translation of the collected data, and feedback provision to health care providers; the educational component of accompanying mobile apps was also valued [25,33,35]. High user acceptability was also reported [26], with patients finding data charts clear and easy to interpret [26,33] and display sizes satisfactory [25,26]. Therefore, many patients expressed a willingness to pay for these devices [26].

The devices were found to improve self-management capabilities of patients with COPD or asthma, as they improved medication adherence and allowed monitoring of inhalation technique errors [33], thereby preventing misuse or double dosing and increasing patient awareness of inhalation medication use habits [26,33,35].

XSL•FO
RenderX

In addition, EIMDs facilitated shared decision-making between patients and health care professionals [34], empowering patients to self-manage their health and reduce the burden on health care systems [24,26].

Stresses and Challenges Associated With EIMDs Use

Patients with COPD or asthma also reported negative emotional stresses with these devices. Around 3 reports reported concerns among patients about data security and loss or incorrect data received by health care professionals [24,25,34]. In addition, some patients felt that certain features of the devices created a feeling of being watched and controlled [24-26].

Trust in using EIMDs was an issue, which resulted in nonadherence to their use, with some patients questioning the accuracy of device-generated measurements and perceived inaccuracies in the displayed information [33,34]. Many patients trusted their health care providers more than their devices [24,35].

Social challenges were also noted. Certain physical characteristics of the devices, such as size, color, and reminder tones, were perceived as barriers to social interaction, as they could attract unwanted attention in public or cause discomfort in relationships with roommates or partners [25,26].

The cost of EIMDs was another obstacle, as the devices were perceived to impose a financial burden on patients. One participant noted, “If it’s very expensive, I don’t think I’ll use it because my asthma medication is at a price” [24].

Technical issues were another challenge. Some patients reported that device functionality was insufficiently developed, leading to inaccuracies and unstable data recording [34]. Operational difficulties arose due to the complex functions of the devices, as mobile apps required substantial storage space and were complicated to navigate [26,34-36]. Incompatibility between monitoring devices and various inhalation devices, as well as between different monitoring device applications, led to patients feeling bored [24].

Expectations for EIMDs

Patients expressed a desire for more compact, portable, and user-friendly devices that were quick and easy to use. They wanted the devices to offer additional functionality while being time-saving [24,25].

Patients also desired external support guidance, such as family members, doctors, and nurses, when using the devices. Many felt that EIMDs are not a substitute for face-to-face interactions with health care professionals and preferred to rely on specialist nurses and general practitioners for providing support [24,26]. Support from family caregivers was also valued while using EIMDs [26], patients expressed a wish to make the devices easier for other patients to learn to use them [33].

Discussion

Principal Findings

This systematic review reveals diverse experiences and perceptions among patients with COPD and asthma regarding electronic inhaler monitoring devices. While many patients report positive experiences, finding these devices helpful in supporting medication adherence and inhalation techniques, they also face notable challenges and concerns. Patients expressed a desire to improve the structural design and functionality of the EIMDs to enhance their ease of use.

Several studies underscore patients’ positive experiences with EIMDs, aligning with the findings in this review [26,37]. Patients generally perceive these devices as versatile and easy to use, offering a range of functional support that enhances their medication adherence and self-management skills. Many view the potential of these devices to aid inhalation techniques positively, expecting that effective use of such features will optimize therapeutic outcomes.

However, the review also indicates that patients often lack a full understanding of these devices’ functionalities, leading to apprehensions about data monitoring, collection, and transmission processes. These concerns are compounded by uncertainties about data security and accuracy, which can erode patient confidence in the devices. As noted by Howard et al [37], features such as timing reminders and device appearance can evoke negative emotional responses, making some patients feel monitored or controlled, thereby reducing social acceptance. Similar themes emerged in this review, as social relationship stressors and technical barriers impacted patients’ perceptions of device usefulness and ease of use. Literature suggests that perceived usefulness and ease of use are pivotal in shaping patient attitudes toward new technologies, which in turn influence user behaviors [38].

Therefore, medical staff should develop standardized protocols for equipment inspection and patient orientation before use. Health care professionals must also remain attentive to patients’ emotional responses, identifying negative reactions promptly and intervening with tailored support to improve patient experiences. By addressing the causes of patient discomfort, medical staff can mitigate barriers to device adoption and enhance patient satisfaction.

Emerging evidence highlights the influence of mHealth technology on patients’ real-life experiences [39]. Despite advancements, current electronic inhaler monitoring devices still lack functionality and compatibility with various inhalers and mobile apps. For example, current devices lack environmental monitoring capabilities, such as air quality feedback, that could provide added value [28,40]. The clinical integration of diverse device-linked mobile apps remains challenging, as does the collection of comprehensive patient data [41]. Design failures are likely when developers overlook patient needs [42]. Thus, clinical staff should regularly collect patient feedback on their expectations for device design and functionality, sharing this input with manufacturers and researchers to drive design and functional upgrades of EIMDs [43].

The future development of electronic inhaler monitoring devices would benefit from a collaborative, interdisciplinary approach. A team inclusive of stakeholders from clinical practice, design, and patient advocacy could apply participatory design methods to develop low-cost, user-friendly devices that meet diverse patient needs, improving both accessibility and acceptance.

XSL•FO
RenderX

Previous research [44] also suggests that family caregivers are supportive of patients adopting innovative mHealth technologies. Patients often rely on family caregivers and peers for assistance with such devices, consistent with our findings. Patients express a desire for family support during device use, underscoring the importance of engaging family members actively in the management of electronic inhaler monitoring devices. Family involvement not only provides oversight but can also promote adherence to prescribed treatment regimens. However, mobile monitoring devices may inadvertently reduce opportunities for direct patient-provider interactions, potentially affecting patient satisfaction and device usage [45]. Some patients believe that electronic monitoring devices cannot fully substitute for in-person communication with health care providers [26]. This review also indicates that patients require ongoing guidance from medical professionals to navigate device use effectively.

To support patient engagement, researchers should develop workflows that integrate electronic inhaler monitoring devices into clinical practice, clearly defining roles and responsibilities for all relevant stakeholders. Establishing a personalized management model that involves patients and their families may ultimately enhance adherence and improve clinical outcomes.

Limitations

While this thematic review offers insight into variations across individual reports and provides a nuanced understanding of specific issues, certain limitations are unavoidable. The constraints of this systematic review primarily stem from the search strategy and inclusion criteria. Notably, we did not search the websites of companies that manufacture electronic monitoring devices for inhalation, focusing solely on peer-reviewed literature and excluding grey literature. In addition, reports not published in English were excluded, and reports were omitted if they focused exclusively on mobile medical apps without integrating electronic inhaler monitoring devices. Although the reports included represent 4 countries, with over half based in the United Kingdom, introducing potential bias due to limited geographical diversity.

Conclusion

Through the qualitative synthesis of reports on the experiences and perceptions of COPD and asthma patients using electronic inhaler monitoring devices, this review highlights both positive experiences and significant challenges that impact patient acceptability of these devices. Moving forward, device manufacturers should prioritize equipment and software upgrades that reflect patient expectations and needs. Expanding research on these devices in diverse respiratory patient populations will be essential to ultimately enhance device acceptability and improve patient outcomes.

Acknowledgments

We would like to thank all the study participants for their valuable contributions. We would also like to express our gratitude to the college and hospital for supporting the study.

This work was funded by the Key Research and Development Projects of Ningxia Hui Autonomous Region (No. 2021BEG03116).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOC File, 407 KB - mhealth_v13i1e57645_app1.doc]

Multimedia Appendix 2
Quality assessment form for included studies.
[DOC File, 37 KB - mhealth_v13i1e57645_app2.doc]

Checklist 1
PRISMA checklist
[DOCX File, 32 KB - mhealth_v13i1e57645_app3.docx]

References

1. Poureslami I, FitzGerald JM, Tregobov N, et al. Health literacy in asthma and chronic obstructive pulmonary disease (COPD) care: a narrative review and future directions. Respir Res 2022 Dec 19;23(1):361. [doi: 10.1186/s12931-022-02290-5] [Medline: 36529734]
2. Salim H, Ramdzan SN, Ghazali SS, et al. A systematic review of interventions addressing limited health literacy to improve asthma self-management. J Glob Health 2020 Jun;10(1):010427. [doi: 10.7189/jogh.10.010428] [Medline: 32566166]

XSL•FO
RenderX

3. Wang C, Xu J, Yang L, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet 2018 Apr 28;391(10131):1706-1717. [doi: 10.1016/S0140-6736(18)30841-9] [Medline: 29650248]
4. Rose S, Paul C, Boyes A, et al. Stigma-related experiences in non-communicable respiratory diseases: a systematic review. Chron Respir Dis 2017 Aug;14(3):199-216. [doi: 10.1177/1479972316680847] [Medline: 28111991]
5. Levy ML, Bacharier LB, Bateman E, et al. Key recommendations for primary care from the 2022 global initiative for asthma (GINA) update. NPJ Prim Care Respir Med 2023 Feb 8;33(1):7. [doi: 10.1038/s41533-023-00330-1] [Medline: 36754956]
6. Cataldo D, Hanon S, Peché RV, et al. How to choose the right inhaler using a patient-centric approach? Adv Ther 2022 Mar;39(3):1149-1163. [doi: 10.1007/s12325-021-02034-9] [Medline: 35080761]
7. Pires N, Pinto P, Marçal N, et al. Pharmacological treatment of COPD - new evidence. Pulmonology 2019;25(2):90-96. [doi: 10.1016/j.pulmoe.2018.10.005] [Medline: 30502972]
8. van der Kolk A, Lammers N, Brusse-Keizer M, et al. Comparison of inhalation technique with the Diskus and Autohaler in asthmatic children at home. ERJ Open Res 2021 Apr;7(2):00215-2019. [doi: 10.1183/23120541.00215-2019] [Medline: 33898617]
9. Hanania NA, Braman S, Adams SG, et al. The role of inhalation delivery devices in COPD: perspectives of patients and health care providers. Chronic Obstr Pulm Dis 2018 Apr 1;5(2):111-123. [doi: 10.15326/jcopdf.5.2.2017.0168] [Medline: 30374449]
10. Schnoor K, Versluis A, Bakema R, et al. A pharmacy-based eHealth intervention promoting correct use of medication in patients with asthma and COPD: nonrandomized pre-post study. J Med Internet Res 2022 Jun 8;24(6):e32396. [doi: 10.2196/32396] [Medline: 35675120]
11. Harb HS, Ibrahim Laz N, Rabea H. Determinants of incorrect inhaler technique in chronic obstructive pulmonary disease patients. Int J Clin Pract 2021 Jun;75(6):e14073. [doi: 10.1111/ijcp.14073] [Medline: 33559260]
12. Taylor TE, Zigel Y, Egan C, et al. Objective assessment of patient inhaler user technique using an audio-based classification approach. Sci Rep 2018 Feb 1;8(1):2164. [doi: 10.1038/s41598-018-20523-w] [Medline: 29391489]
13. Nassar RI, Basheti IA, Saini B. Exploring validated self-reported instruments to assess adherence to medications used: a review comparing existing instruments. Patient Prefer Adherence 2022;16:503-513. [doi: 10.2147/PPA.S352161] [Medline: 35237029]
14. Nassar RI, Saini B, Obeidat NM, et al. Assessing adherence to medications: is there a difference between a subjective method and an objective method, or between using them concurrently? Pharm Pract (Granada) 2022;20(4):2723. [doi: 10.18549/PharmPract.2022.4.2723] [Medline: 36793904]
15. Anghel LA, Farcas AM, Oprean RN. An overview of the common methods used to measure treatment adherence. Med Pharm Rep 2019 Apr;92(2):117-122. [doi: 10.15386/mpr-1201] [Medline: 31086837]
16. Plaza V, Sanchis J, Roura P, et al. Physicians' knowledge of inhaler devices and inhalation techniques remains poor in Spain. J Aerosol Med Pulm Drug Deliv 2012 Feb;25(1):16-22. [doi: 10.1089/jamp.2011.0895] [Medline: 22047451]
17. Giner J, Roura P, Hernández C, et al. Knowledge and attitudes of nurses in Spain about inhaled therapy: results of a national survey. J Aerosol Med Pulm Drug Deliv 2016 Feb;29(1):86-93. [doi: 10.1089/jamp.2014.1198] [Medline: 26168021]
18. Almutairi M, Marriott JF, Mansur A. Effect of monitoring adherence to regular inhaled corticosteroid (ICS) alone or in combination with a long-acting β2-agonist (LABA) using electronic methods on asthma outcomes: a narrative systematic review. BMJ Open 2023 Aug 16;13(8):e074127. [doi: 10.1136/bmjopen-2023-074127] [Medline: 37586865]
19. Free C, Phillips G, Watson L, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363. [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
20. McCrossan P, O'Donoghue D, McElnay JC, et al. The use of remote video directly observed therapy to improve both inhaler technique and adherence to asthma medications. Front Public Health 2022;10:965629. [doi: 10.3389/fpubh.2022.965629] [Medline: 36276358]
21. Dhruve H, Jackson DJ. Assessing adherence to inhaled therapies in asthma and the emergence of electronic monitoring devices. Eur Respir Rev 2022 Jun 30;31(164):210271. [doi: 10.1183/16000617.0271-2021] [Medline: 35613744]
22. Alshabani K, Attaway AA, Smith MJ, et al. Electronic inhaler monitoring and healthcare utilization in chronic obstructive pulmonary disease. J Telemed Telecare 2020;26(7-8):495-503. [doi: 10.1177/1357633X19850404] [Medline: 31096842]
23. Garin N, Zarate-Tamames B, Gras-Martin L, et al. Clinical impact of electronic monitoring devices of inhalers in adults with asthma or COPD: a systematic review and meta-analysis. Pharmaceuticals (Basel) 2023 Mar 8;16(3):414. [doi: 10.3390/ph16030414] [Medline: 36986513]
24. van de Hei SJ, Stoker N, Flokstra-de Blok BMJ, et al. Anticipated barriers and facilitators for implementing smart inhalers in asthma medication adherence management. NPJ Prim Care Respir Med 2023 May 20;33(1):22. [doi: 10.1038/s41533-023-00343-w] [Medline: 37208358]
25. Adejumo I, Patel M, McKeever TM, et al. Qualitative study of user perspectives and experiences of digital inhaler technology. NPJ Prim Care Respir Med 2022 Dec 22;32(1):57. [doi: 10.1038/s41533-022-00320-9] [Medline: 36550136]

XSL•FO
RenderX

26. Foster JM, Reddel HK, Usherwood T, et al. Patient-perceived acceptability and behaviour change benefits of inhaler reminders and adherence feedback: a qualitative study. Respir Med 2017 Aug;129(39-45):39-45. [doi: 10.1016/j.rmed.2017.05.013] [Medline: 28732834]
27. Tong A, Flemming K, McInnes E, et al. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012 Nov 27;12:181. [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
28. Toelle BG, Marks GB, Dunn SM. Validation of the inhaler adherence questionnaire. BMC Psychol 2020 Sep 3;8(1):95. [doi: 10.1186/s40359-020-00461-x] [Medline: 32883369]
29. Kikidis D, Konstantinos V, Tzovaras D, et al. The digital asthma patient: the history and future of inhaler based health monitoring devices. J Aerosol Med Pulm Drug Deliv 2016 Jun;29(3):219-232. [doi: 10.1089/jamp.2015.1267] [Medline: 26919553]
30. CASP checklist: for qualitative research. CASP Checklist. 2018. URL: https://casp-uk.net/casp-checklists/CASP-checklist-qualitative-2024.pdf [accessed 2025-05-10]
31. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8(45):45. [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
32. Boehm LM, Jones AC, Selim AA, et al. Delirium-related distress in the ICU: a qualitative meta-synthesis of patient and family perspectives and experiences. Int J Nurs Stud 2021 Oct;122:104030. [doi: 10.1016/j.ijnurstu.2021.104030] [Medline: 34343884]
33. Hesso I, Nabhani-Gebara S, Kayyali R. Objective assessment of adherence and inhaler technique among asthma and COPD patients in London: a study in community pharmacies using an electronic monitoring device. Pharmacy (Basel) 2023 Jun 4;11(3):94. [doi: 10.3390/pharmacy11030094] [Medline: 37368420]
34. Hui CY, McKinstry B, Mclean S, et al. Assessing the technical feasibility of a flexible, integrated Internet-of-things connected for asthma (C4A) system to support self-management: a mixed method study exploring patients and healthcare professionals perspectives. JAMIA Open 2022 Dec;5(4):ooac110. [doi: 10.1093/jamiaopen/ooac110] [Medline: 36601366]
35. Jácome C, Almeida R, Pereira AM, et al. Feasibility and acceptability of an asthma app to monitor medication adherence: mixed methods study. JMIR mHealth uHealth 2021 May 25;9(5):e26442. [doi: 10.2196/26442] [Medline: 34032576]
36. Kayyali R, Savickas V, Spruit MA, et al. Qualitative investigation into a wearable system for chronic obstructive pulmonary disease: the stakeholders' perspective. BMJ Open 2016 Aug 31;6(8):e011657. [doi: 10.1136/bmjopen-2016-011657] [Medline: 27580831]
37. Howard S, Lang A, Sharples S, et al. See I told you I was taking it! - Attitudes of adolescents with asthma towards a device monitoring their inhaler use: implications for future design. Appl Ergon 2017 Jan;58:224-237. [doi: 10.1016/j.apergo.2016.06.018] [Medline: 27633217]
38. Gücin N, Berk Ö. Technology acceptance in health care: an integrative review of predictive factors and intervention programs. Procedia Soc Behav Sci 2015 Jul;195:1698-1704. [doi: 10.1016/j.sbspro.2015.06.263]
39. Shareef MA, Kumar V, Kumar U. Predicting mobile health adoption behaviour: a demand side perspective. J Cust Behav 2014 Oct 31;13(3):187-205. [doi: 10.1362/147539214X14103453768697]
40. Leonardsen ACL, Hardeland C, Helgesen AK, et al. Patient experiences with technology enabled care across healthcare settings- a systematic review. BMC Health Serv Res 2020 Aug 24;20(1):779. [doi: 10.1186/s12913-020-05633-4] [Medline: 32838784]
41. Jansen EM, van de Hei SJ, Dierick BJH, et al. Global burden of medication non-adherence in chronic obstructive pulmonary disease (COPD) and asthma: a narrative review of the clinical and economic case for smart inhalers. J Thorac Dis 2021 Jun;13(6):3846-3864. [doi: 10.21037/jtd-20-2360] [Medline: 34277075]
42. Grocott P, Weir H, Ram MB. A model of user engagement in medical device development. Int J Health Care Qual Assur 2007;20(6):484-493. [doi: 10.1108/09526860710819422] [Medline: 18030966]
43. Chan AHY, Harrison J, Black PN, et al. Using electronic monitoring devices to measure inhaler adherence: a practical guide for clinicians. J Allergy Clin Immunol Pract 2015;3(3):335-349. [doi: 10.1016/j.jaip.2015.01.024] [Medline: 25840665]
44. Thilo FJ, Hahn S, Halfens RJ, et al. Facilitating the use of personal safety alerting device with older adults: the views, experiences and roles of relatives and health care professionals. Geriatr Nurs (Lond) 2021 Jul;42(4):935-942. [doi: 10.1016/j.gerinurse.2021.05.005]
45. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 1;20(5):e10235. [doi: 10.2196/10235] [Medline: 29716883]

Abbreviations

CASP: Critical Appraisal Skills Programme
COPD: chronic obstructive pulmonary disease
EIMD: electronic inhaler monitoring device
ENTREQ: Enhancing Transparency in Reporting the Synthesis of Qualitative Research
mHealth: mobile health
NHS: National Health Service

XSL•FO
RenderX

PICoS: Population, Phenomenon of interest, Context, Types of studies
PROSPERO: International Prospective Register of Systematic Reviews

Edited by L Buis; submitted 23.02.24; peer-reviewed by H Holmen, R Jeminiwa; revised version received 21.02.25; accepted 18.03.25; published 16.05.25.

Please cite as:
Duan J, Chen X, Fan D, Jiang H, Zhang X, Zhang W, Liu Z, Lu H
Experience of Using Electronic Inhaler Monitoring Devices for Patients With Chronic Obstructive Pulmonary Disease or Asthma: Systematic Review of Qualitative Studies
JMIR Mhealth Uhealth 2025;13:e57645
URL: https://mhealth.jmir.org/2025/1/e57645
doi:10.2196/57645

© Jilong Duan, Xia Chen, Di Fan, Haikun Jiang, Xue Zhang, Wenyue Zhang, Zhiping Liu, Hongyan Lu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Intervention in Children With Developmental Language Disorder: Systematic Review

Zhaowen Zhou[1*], BSc, MSc, OTD; Cheng Deng[2*], BPH; Dongling Yin[2], MBBS; Qiaoxue Yang[2], BSc; Zhuoming Chen[1], MBBS, MM, MD

[1]Department of Rehabilitation Medicine, The First Affiliated Hospital of Jinan University, 613 Huangpu Avenue West, Guangzhou, China

[2]Department of Children's Health Care, Zhongshan Torch Development Zone People's Hospital, Zhongshan, China

[*]these authors contributed equally

Corresponding Author:
Zhuoming Chen, MBBS, MM, MD
Department of Rehabilitation Medicine, The First Affiliated Hospital of Jinan University, 613 Huangpu Avenue West, Guangzhou, China

Abstract

Background: Developmental language disorder (DLD) is one of the most common neurodevelopmental disorders. Effective intervention is primarily important for improving the language and communication skills of children with DLD, and strengthening these skills ensures quality of life and prevents negative effects in adulthood. Digital interventions have the potential to complement conventional language intervention, reducing the workload for therapists and increasing accessibility to language training in homes or schools.

Objective: This systematic review aimed to explore the language domain that is most frequently targeted by digital intervention in children with DLD.

Methods: The study protocol was registered in the International Prospective Register for Systematic Reviews (PROSPERO) and was ascribed the CRD42023477946 registration code. The initial search was conducted on May 2023 from 4 databases: "PubMed," "Scopus," "PsycInfo," and "IEEE Xplore," following a method adapted from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Inclusion criteria include studies recruiting patients diagnosed with DLD; studies that reported digital interventions based on apps, video games, augmented reality, or any other type of software based on language outcomes; and English language studies. Reviews, letters, conference proceedings, abstracts, editorials, and studies not published in English were removed. The titles and abstracts of the identified records were initially screened and selected by 2 independent and blinded reviewers. Data extraction and quality assessment were performed by 3 independent reviewers.

Results: Overall, 13 studies were included; 961 children with DLD underwent a digital intervention. The mean age ranged from 3.47 (SD 0.17) to 11.19 (SD 1.12) years. A total of 8 were randomized controlled trials, and 5 were quasi-experimental studies. Targeting domains of digital intervention were phonological skills (n=5), general language function (n=3), grammar (n=3), and vocabulary (n=2).

Conclusions: This systematic review indicates that phonological skills are the most frequently targeted language domain by digital interventions in children with DLD. Given the limited number and the heterogeneity of the studies included, it is still unclear whether digital intervention was effective in improving different language skills in children with DLD. There was less evidence supporting its effectiveness in expressive language skills, which indicates a need to update expressive language digital training programs in the future. Further higher-level evidence, such as randomized controlled trial studies in this area, is needed to direct the development of digital programs.

Trial Registration: PROSPERO CRD42023477946; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023477946

(JMIR Mhealth Uhealth 2025;13:e59992) doi:10.2196/59992

KEYWORDS

developmental language disorder; digital intervention; systematic review; language rehabilitation; phonological skills

Introduction

Developmental language disorder (DLD) is one of the most prevalent neurodevelopmental disorders that has profound and lasting effects on individual development [1]. It occurs in approximately 7.6%-8.5% of preschool children around the world [2,3]. According to the *International Classification of Diseases, Eleventh Revision*, DLD is characterized by persistent deficits in the acquisition, understanding, production, or use of language (spoken or signed), that arise during the developmental

XSL•FO
RenderX

period, typically during early childhood, and cause significant limitations in the individual's ability to communicate. The individual's ability to understand, produce, or use language is markedly below what would be expected given the individual's age. Language deficits are not explained by another neurodevelopmental disorder, or sensory impairment, or neurological condition, including the effects of brain injury or infection [4]. The term DLD has been suggested to replace earlier terms, such as specific language impairment, language impairment, language disorder, and primary spoken language disorder. However, for clarity purposes, in this review, we will use DLD, although due to its recent introduction, the studies included in this work used specific language impairment as a diagnostic label.

Language development is a critical domain of children's overall development. Language skills are an important means of communication, which comprises the ability to send and receive information through oral and written language [5]. DLD can affect different aspects of language processing, such as the form of language (phonetic, phonological, morphological, morpho-syntactic, and syntactic processing); content (semantic-lexical and phrasal processing); and use (pragmatic and discursive processing) [6]. Individuals with DLD may not only experience difficulties in communication but also face challenges from other different domains, such as reduced engagement in playing and academic learning [7]. DLD may also increase the risk of poor health-related quality of life [8], including mental health problems such as anxiety and depression [9,10]. In addition, DLD has an impact on the development of cognitive functioning [11,12], sensorimotor functioning [11,13], or behavioral functioning [14]. DLD can coexist with many other conditions, such as learning disorders, delayed motor milestones, auditory processing disorders, attention-deficit/hyperactivity disorder, lack of coordination, and other motor disorders [15,16]. The impact of DLD may persist to adulthood and have long-term effects on cognitive function, community function, interpersonal relationships, and employment [17-19]. Therefore, effective interventions during childhood are essential to improve the quality of life of individuals with DLD and to prevent further negative effects in adulthood.

Convention interventions for DLD are usually conducted one by one in person by speech and language pathologists (SLPs). During treatment sessions, SLPs will design challenging interventions for children using interesting toys and activities to stimulate children's language skills and teach language learning strategies. However, that requires SLPs to be well-trained and experienced. Limited by the number of therapists, the geographic distance between the SLP and the recipient of services, and economic conditions, in-person training is not always available to children with DLD [20].

Digital interventions have the potential to be used as an adjunct to conventional language intervention, which could reduce therapists' workload and increase children's accessibility to language training in settings such as homes or schools, enabling them to practice under the supervision of caretakers or teachers [21]. Digital intervention provides repeated training of particular skills easier; it motivates children by giving automated timely feedback in a game-like format [22]; digital intervention can also be programmed to respond adaptively to the child's level of performance so that training is focused on materials that are just beyond current competence; every response made by the child can be detailly recorded for further analysis. In the economic aspect, digital intervention also has the potential to increase accessibility to care, reduce patients' travel and costs, and develop culturally appropriate services, especially for different language speakers [23].

There is a wide range of digital interventions for children with DLD that are aimed at different language domains. This review aimed to systematically analyze which language domains are most frequently supported by digital interventions for children with DLD.

Methods

Study Identification

This systematic literature review was performed according to the methodology described in the Cochrane Handbook for Systematic Reviews [24] and was reported based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement for reporting systematic reviews [25]. Additionally, the study protocol was registered in the International Prospective Register for Systematic Reviews (PROSPERO) and was ascribed to the CRD42023477946 registration code.

The initial search was conducted on May 2023 by searching the databases "PubMed," "Scopus," "PsycInfo," and "IEEE Xplore" using the following search terms: (("developmental language disorder" OR "language delay" OR "speech delay" OR "language impairment*" OR "speech impairment*" OR "language disorder*" OR "speech disorder*" OR "language difficult*" OR "speech difficult*") AND ("Child*" OR "Preschool") AND ("computer-based" OR "computer assisted therapy" OR "computer games" OR "software" OR "websites" OR "computer*" OR "digital*" OR "electronic" OR "gaming" OR "internet*" OR "video game*" OR "online" OR "on-line" OR "web*")).

No limitations in the search strategy were applied to the publication date, study design, or language. References of considered studies were also searched to identify any further relevant data.

The records identified by the search were uploaded to "Rayyan" [26] to organize the study selection more efficiently. The titles and abstracts of the identified records were initially screened and selected by 2 independent and blinded reviewers (ZZ and CD) based on their pertinence to the review topic. Conflicts and disagreements were resolved by consensus.

The following set of predefined inclusion criteria were then individually applied to the selected studies in their full-text version: (1) studies recruiting patients diagnosed with DLD; (2) studies that reported digital interventions based on apps, video games, augmented reality, or any other type of software based on language outcomes; and (3) English language studies.

XSL•FO
RenderX

Reviews, letters, conference proceedings, abstracts, editorials, case reports, and case series were excluded. Studies not published in English were removed. Studies in which the targeted populations were children with cognitive delay, deafness, autism spectrum disorders, genetic syndromes (Down syndrome and Klinefelter syndrome), neurological deficits, pervasive developmental disorders, traumatic brain injuries, primary disorders (sensory, neurological, and psychiatric), children with dysphonia, dysarthria, dysrhythmias or stuttering, dyslalia or specific speech articulation disorder, and bilingualism were excluded.

Data Extraction and Quality Assessment

Data extraction and quality assessment were performed by 3 independent reviewers (ZZ, CD, and DY). The extracted data and results of the quality assessment are reported in Multimedia Appendix 1 [22,27-38]. The quality of retrieved studies was assessed by the Joanna Briggs Institute (JBI) appraisal tool [39]. The JBI checklist for randomized controlled trials (RCTs) [40] and for quasi-experimental studies [41] was used according to the study design of included studies. JBI appraisal tool evaluates internal validity and statistical conclusion validity of the targeted studies. For RCTs, questions related to internal validity were further separated to assess bias related to selection and allocation, administration of intervention or exposure, assessment, detection and measurement of outcome, and participant retention, and for quasi-experimental studies, domains of internal validity include the bias related to temporal precedence, selection and allocation, confounding factors, administration of intervention or exposure, assessment, detection, and measurement of the outcome, and participant retention.

The purpose of this appraisal is to assess the methodological quality of studies and to determine the extent to which a study has addressed the possibility of bias in its design, implementation, and analysis. Studies with percentage scores of <49% were classified as weak, studies between 50% and 79% were moderate, and studies ≥80% were classified as high in quality with reduced bias. This classification has been used in prior systematic reviews [42].

Treatment Fidelity Assessment

A 40-item checklist with definitions of intervention fidelity domains and relevant examples was used to assess the treatment fidelity [43]. The authors used the checklist to assess the presence of strategies identified in the 5 specific domains: treatment design, training providers, delivery, receipt, and enactment. Intervention fidelity information was categorized and coded as “present (1),” “not present but should be present (0),” or “not applicable (NA).” Authors coded studies independently and discussed and reconciled any discrepancies.

Results

Overview

The following records were identified via electronic databases: PubMed (n=1563); Scopus (n=2200), PsycInfo (n=111), and IEEE Xplore (n=549), which yielded 4423 unique records. One additional record was identified via manual searching. The flow diagram in Figure 1 shows the records remaining at each stage and the reasons for exclusion of studies reviewed at the full-text stage. One full-text study could not be sourced for review and was excluded at the full-text stage. A total of 13 studies met the inclusion criteria for this study.

Of the 13 studies that met the inclusion criteria, 8 studies were RCTs [22,28-30,33-35,37], and 5 studies were quasi-experimental studies [27,31,32,36,38]. According to the quality assessment, 6 studies were of high quality, and 7 studies were of moderate quality out of the 13 studies. The oldest study was conducted in 2005, and the most recent study was conducted in 2022.

Overall, 961 children diagnosed with DLD were included, 63.89% (n=460) were male and 36.11% (n=260) were female (gender proportion not reported in 5 studies). Mean age ranges from 3.47 (SD 0.17) to 11.19 (SD 1.12) years old.

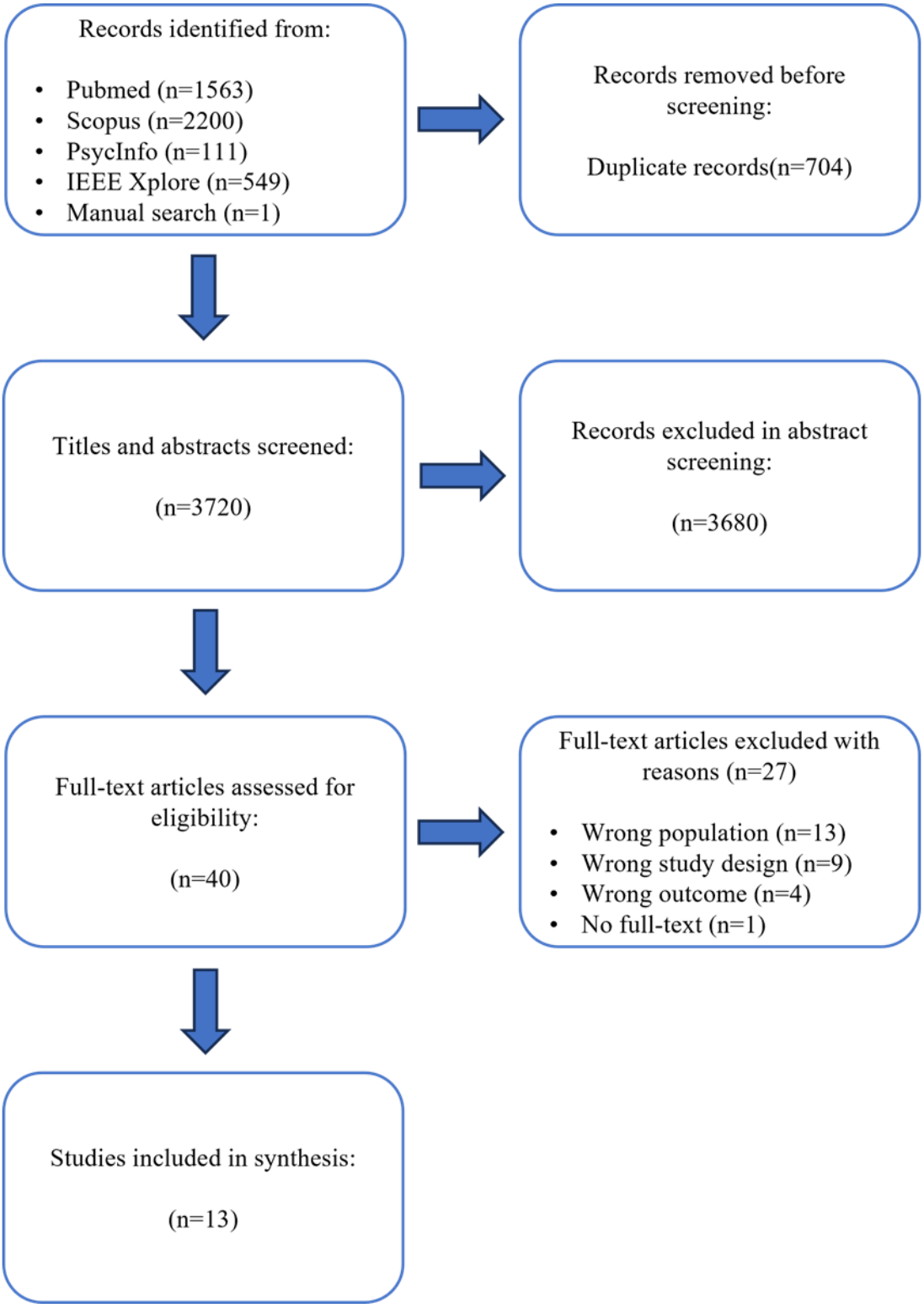
The digital intervention programs used included: Fast ForWord (FFW; n=6) [27], Reading Doctor (n=1) [28], 作業TipOn (n=1) [29], My Sentence Builder (n=1) [37], Jingyun Rehab Platform (n=1) [38], My PlayHome (n=1) [22], Dr. Neuronowski (n=1) [30], and a program developed by Heikkilä et al [31] and was run with Presentation software (Neurobehavioral systems; n=1). Among the 8 programs mentioned above, 7 of them are software for computers, tablets, or mobile phones, and Jingyun Rehab Platform is a website for computers, tablets, and mobile phones. A total of 4 of the programs were developed in English, 2 of them were developed in Chinese (作業TipOn and Jingyun Rehab Platform), 1 was in Finnish, and 1 was in Polish.

Of the 13 studies included, 5 studies had a follow-up of 5 weeks to 6 months after the end of the intervention. A total of 4 of these studies showed that the intervention effects of digital intervention on language functioning in children with DLD were maintained or improved at follow-up.

Phonological skills were the most frequently targeted skill in digital intervention for DLD (5 out of 13). General language function and grammar were each reported in 3 studies. Vocabulary was reported in 2 studies.

XSL•FO
RenderX

Figure 1. PRISMA flowchart describing the inclusion process of the studies. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Phonological Skills (n=5)

The effectiveness of digital intervention in phonological skills for children with DLD was reported in 5 studies. These studies were targeted at phonological skills, including phoneme awareness, letter-sound aptitude, and early decoding ability. The programs used to conduct phonological skills training in 5 studies were varied, including the Reading Doctor iPad app, 作業TipOn (mobile app), FFW, and Dr. Neuronowski. The duration of the digital intervention was 4 to 9 weeks. The main design of the game to train receptive phonological skills was based on improving basic auditory processing skills, either through nonspeech stimuli such as sounds and tones or through speech stimuli such as syllables and words. In these games, the sound or an audiovisual video of the targeted letter, syllables, or words was provided, and children were required to point to

XSL•FO
RenderX

the best-matched picture among the distractors or type the correct spelling of the word. Significant improvement in receptive phonological skills was found in 4 studies [28-31] compared with the control group. Results showed that audiovisual speech might be more effective than auditory speech in training phonological skills in children. However, in the study targeted spelling [27], which is a higher level of phonological skills, trained groups did not differ from the untrained control group in spelling. In the studies targeting phonological skills, researchers also aimed to improve participants' word learning and reading [27-29] through phonemic awareness training. Bishop et al [27] suggested that digital intervention did not improve word and nonword reading compared to the untrained group, whereas Carson [28] found a significant improvement in word reading skills in the digital intervention group compared to the control group. In addition, Chen and Lin [29] found that the total number of games played and the accuracy in games were significantly correlated with the performance of the prepost gain in the word definition task, which asked children to say the meaning of a vocabulary word they heard.

General (n=3)

There were 3 studies [32-34] that examined interventions targeting more than one domain, including phonological skills, vocabulary, and grammar. The same digital program and intervention procedures were applied to both studies. Children participated in the FFW program for 1 hour and 40 minutes (100 min: five games of 20 min each) each weekday for 6 weeks (30 d). The intervention ended when they reached the dismissal criterion established by Scientific Learning Corporation of 90% completion on 5 of 7 exercises or until they exhibited plateaus in performance for 10 days before the 6-week end date. FFW included games that targeted discrimination of tones (viz, Circus Sequence), detection of individual phoneme changes (viz, Old McDonald's Flying Farm), matching phonemes to a target (viz, Phoneme Identification), identifying matched syllable pairs (viz, Phonic Match), discriminating between minimal pair words (viz, Phonic Words), recalling commands (viz, Block Commander), and comprehending grammatical morphemes and complex sentence structures (viz, Language Comprehension Builder). RCTs conducted by Cohen et al [33] and Gillam et al [34] revealed that though participants gained improvement from FFW, there was no additional effect for computer intervention compared to conventional intervention. However, Loeb et al [32] found no significant improvement in phonological skills and reading.

Grammar (n=3)

A total of 3 studies included grammar as the primary focus of digital intervention for DLD, of which, 2 studies examined the effectiveness of FFW, and 1 study investigated My Sentence. Both Hsu and Bishop [35] and Bishop et al [36] used FFW to implement receptive training of grammatical skills. Children heard a spoken sentence and then moved or activated objects on the computer screen to match the spoken sentence. Children in these two studies both showed greater accuracy in the training program, but showed no significant improvement on standardized tests of receptive grammar. Washington et al [37] used My Sentence to address expressive grammar deficits in DLD. Results showed that computer-based training significantly outperformed conventional intervention, but no significant differences in treatment gains were found between computer-based training and noncomputer-based training.

Vocabulary (n=2)

Among the 13 studies included, 2 studies targeted vocabulary learning. Yi et al [38] and Zwitserlood et al [22] focus on the receptive vocabulary of children with DLD, using pictures of web-based platforms or software games. During the vocabulary training, Zwitserlood et al [22] designed play activities with the digital game, and exposed children to the targeted words with symbolic play sessions, narrative, natural interaction, and language facilitation strategies. Yi et al [38] presented the sound of words or syllables and pictures to the children and instructed them to point at the picture that matched the word best. The intervention periods of intervention for vocabulary skills were approximately 1-3 months. Both studies showed that digital interventions were beneficial for children's vocabulary skills.

Fidelity of the Studies

Table 1 presents the treatment fidelity for the included studies, and Multimedia Appendix 2 [22,27-38] presents the scores of fidelity for each study.

There are 2 domains in which the treatment fidelity scores were above 50%, including the "Treatment design" domain at 75% (SD 0.15%) and the "Enactment" domain at 58% (SD 0.18%). The lowest mean proportion of adherence to strategies was found in the "Receipt" domain, where, on average, only 4% (SD 0.09%) of strategies were reported among the studies. Finally, the mean proportion of adherence to strategies in the "Training providers" and "Delivery" domains was 16% (SD 0.17%) and 19% (SD 0.15%), respectively. The mean proportion of adherence to treatment fidelity strategies included across all 5 domains for all studies was 43% (SD 0.10%).

Based on the fidelity scoring by Borrelli [43], where 50% constitutes low-fidelity scoring, 69% (9/13) of the interventions scored a low treatment fidelity across all 5 domains. In total, 31% (4/13) of the studies, including Chen and Lin [29], Loeb et al [32], Gillam et al [34], and Zwitserlood et al [22], scored >0.50 in the overall treatment fidelity (range 0.50-0.58).

XSL•FO
RenderX

Table . Treatment fidelity for the included studies.

Fidelity domains	Treatment design	Training providers	Delivery	Receipt	Enactment	Overall fidelity proportion per study
Bishop et al [27]	0.63	0.14	0.11	0.00	0.50	0.34
Carson [28]	0.63	0.00	0.00	0.25	1.00	0.34
Chen and Lin [29]	0.88	0.14	0.44	0.00	0.50	0.53
Dacewicz et al [30]	0.88	0.14	0.22	0.00	0.50	0.47
Heikkilä et al [31]	0.88	0.29	0.00	0.00	0.50	0.45
Loeb et al [32]	0.88	0.29	0.22	0.00	0.50	0.50
Cohen et al [33]	0.88	0.00	0.33	0.00	0.50	0.47
Gillam et al [34]	0.88	0.29	0.33	0.00	0.50	0.53
Hsu and Bishop [35]	0.56	0.00	0.00	0.00	0.50	0.26
Bishop et al [36]	0.56	0.00	0.00	0.25	1.00	0.32
Washington et al [37]	0.81	0.29	0.11	0.00	0.50	0.45
Yi et al [38]	0.50	0.00	0.33	0.00	0.50	0.32
Zwitserlood et al [22]	0.88	0.57	0.33	0.00	0.50	0.58
Proportion, mean (SD)	0.75 (0.15)	0.16 (0.17)	0.19 (0.15)	0.04 (0.09)	0.58 (0.18)	0.43 (0.10)

Discussion

Principal Findings

This systematic review aimed to explore the language domain that is most frequently targeted by digital intervention in children with DLD. Phonological skills were the most researched skills among all included studies, especially receptive phonological skills. Phonological awareness is the skill of being aware of sounds in spoken language. It ranges from dividing words into syllables, recognizing and producing rhymes, and matching words with the same beginning sound, to more complex abilities such as separating and processing individual sounds or phonemes, which are causally related to early word decoding skills [44]. Studies showed that digital intervention is helpful in improving phonological skills. Digital intervention programs have the advantage of providing grading acoustically modified speech signals and image or video resources [32]. Hierarchically auditory and visual cueing is crucial in phonological training [45]. Digital interventions are designed to adapt the level of difficulty and cues of the training according to the success of the child on multiple trials, which might be the reason digital interventions are effective.

Word learning ability has also been proven to be improved through phonological skills digital training, while the conclusion is still controversial. Word learning involves processes such as encoding, working memory, lexical access, and mapping of phonological form to meaning [46,47]. Phonological skills, especially phonemic awareness skills, were proven to be the strongest predictor of children's word reading ability [48]. Children with DLD might experience impaired mapping of phonological form to meaning, given their less sensitivity to the pronunciation of the words and the information within words during the word learning phase [49]. Digital intervention, featuring visual-auditory stimuli and a game format, has the potential to better engage participants' attention during the intervention. However, Bishop et al [27] found that participants in the digital intervention group were not better improved in word learning, which might be caused by the short intervention period (4 wk). In addition, it is suggested that children with language impairment and poor reading skills need a more comprehensive approach to improve their reading skills that extends beyond an emphasis on phonemic awareness [32].

General language and speech function was the second most concerned topic in digital interventions. Results showed that only half of the children with DLD gained progress after training. Cohen et al [33] indicated that children with DLD had already received intensive specialist therapy and educational support, while digital interventions were not sufficient in and of themselves to confer additional therapeutic benefits. The varied results between individuals, combined with the generally inconsistent patterns of performance on certain games, create uncertainty as to which elements of FFW are producing improvements [50]. A meta-analysis conducted by Strong et al [51] also indicated that there was no significant effect of FFW on any outcome measure in comparison to active or untreated control groups. Therefore, further studies are needed to evaluate the effectiveness of specific training games or formats of digital interventions.

Studies targeting grammar deficits in DLD showed greater accuracy after intervention but no significant improvement on standardized tests of receptive grammar, or no significant

XSL•FO
RenderX

differences in treatment gains between computer-based training and noncomputer-based training. In grammar training games, accuracy could be improved by simply rote learning the meaning of the whole sentence, or by memorizing the correct answers [35]. Therefore, even if children with DLD have difficulties in analyzing the sentence structure, they can still behave well in the training. For this reason, when targeting higher language functions such as grammar, more diverse materials, and learning patterns are needed to promote the generalization of functions. Since the number of studies included in this review was limited, further studies with follow-up periods were needed to examine the effectiveness of digital intervention in receptive and expressive grammar.

In studies targeting vocabulary skills, digital intervention was of promising effectiveness. Children with DLD showed significant improvement after approximately 2-3 months of intervention. Vocabulary training might involve cognitive functions and enhance more global skills [52,53]. Yi et al [38] suggested that children's language development can promote growth hormone secretion in children, stimulate brain development, promote children's intellectual development, and have a positive role in improving children's language communication skills. Therefore, it is recommended that children with DLD receive vocabulary training.

Although digital interventions for DLD were proven effective in several studies, there was evidence showing that DLD did not additionally benefit from digital intervention. Several factors might need to be considered when delivering digital treatment for DLD.

The first factor to be considered is the targeting domain of intervention. In the domains discussed above, digital interventions are demonstrated to be practical and effective in improving basic language and speech skills, such as phonological skills and vocabulary skills. Training of these skills requires lots of repetition and hierarchical audiovisual input, which can be easily achieved by digital techniques [32]. Phonological skills and vocabulary skills are also the prerequisites for learning higher language skills such as grammar and reading [54,55]. However, digital interventions targeting higher language functions, such as grammar, which might involve more cognitive functions [56], were required to be more diversified and more interesting.

Expressive language tended to receive less attention in digital interventions compared to receptive language. Only one study included in this review focused on expressive language. Law et al [57] reported that in-person speech and language therapy interventions were effective for expressive phonological and expressive vocabulary difficulties. However, in computer-based training for expressive language skills, in-time feedback is not provided automatically, which might lower children's motivation to participate. With the development of computerized language analysis techniques, in-time feedback from the digital program might make expressive language evaluation and training more effective [58,59].

People and places to deliver the intervention might affect the effectiveness of digital interventions. In the included studies, digital interventions were delivered in clinical settings, schools, or at home, and people who carried out digital interventions were speech therapists, school teachers, parents, or a combination of two or all three. Digital language training facilitated by trained therapists in clinical settings was shown to be more effective. The reason might be that in home environments, children are used to playing the game on their own and tend to ignore the instruction during intervention [22]. In addition, interventions delivered in school might take away time from their regular classes, which, to some extent, could be detrimental to their language development [27]. Yi et al [38] applied the intervention pattern with therapists deciding on the training topics and sending them to the patient's account, and parents led the training at home. This kind of cooperation proved to be beneficial for children during COVID-19, or for children living in remote areas.

The third factor is duration and intensity. The duration of the included studies ranges from approximately 4-12 weeks, and intensity ranges from approximately 15-100 minutes per day. The literature regarding dosage was unclear in optimal intensity, frequency, and duration to maximize efficacy. It seems that interventions carried out for 6 weeks or longer are more effective than those lasting less than 6 weeks. Further investigation of the optimal dosage is necessary. Yi et al [38] and Chen and Lin [29] suggested that the number of training sessions was directly proportional to the performance of children with DLD in language training tasks.

However, the optimal duration and intensity of digital intervention that can meet both treatment needs and avoid the negative effects caused by excessive screen time remain to be resolved. Frizelle et al [60] conducted a systematic review and narrative synthesis and suggested that frequent, short sessions (2/3× per wk, approximately 2 min) and less frequent, long sessions (1× per wk, approximately 20 min) have yielded the best outcomes when composite language measures have been used. For the dosage form, explicit instruction was more beneficial compared with implicit instruction [61]. Variability in input, elicited production, and gestural and other visual supports was beneficial to language development [61]. The amount of screen time was cumulatively and negatively linked to the children's lexical and general language abilities [62,63]. In addition, the early onset of screen exposure had negative effects on language development [64]. It is suggested that no more than 2 hours of screen time per day has minimal negative effects on development [65]. Better-quality screen use such as educational programs and coviewing was associated with stronger child language skills [63].

The usability and practicality of the digital intervention should also be taken into consideration. Digital intervention for children with DLD uses game-based elements, making the intervention engaging for children. Tablets or mobile phones, with touch screens, provided children with multisensory and direct interaction [22]. Personalization and progress tracking are other advantages of digital intervention. The web-based intervention platform used by Yi et al [38] would automatically pop up the training results, reaction times, and training plan suggestions after each training and automatically adjust the training plan. In addition, digital intervention can offer teachers and families a time-efficient and cost-effective alternative to close learning

XSL•FO
RenderX

gaps and avoid the implications of growing disparities in skills due to lengthy waiting times [28]. However, digital intervention might be too distracting and restrict the use of language-facilitating techniques and language interaction between the SLP [22].

Intervention fidelity is an important aspect of designing and implementing intervention effectiveness studies [66]. Most studies have poor treatment fidelity in the areas of "training providers," "delivery," and "receipt." There was a lack of training for the intervention providers, such as parents or school staff. The reason may be related to the characteristics of the digital intervention. In part of the studies, the intervention process is already set in the digital intervention program, which reduces the professional requirements for the providers of the intervention. However, if the treatment protocols were not set in advance in the software, and nonmedical professionals are allowed to use the digital intervention program freely, it is difficult to ensure the effectiveness of the digital intervention for DLD.

Limitations

This study had some limitations. First, the number of studies regarding the effectiveness of digital intervention in children with DLD in each language domain was insufficient, and most of them have no follow-up period, making it difficult to draw a convincing conclusion. In addition, influential factors such as dosage varied among included studies, and due to the heterogeneity of included studies, the validity of the review analysis might be limited. Besides, this review included research on various research designs, different populations, and interventions. The inclusion of studies with different research designs increases the number of studies included, thereby broadening the perspective on this issue. However, at the same time, the risk of bias is also increasing. In addition, the quality of the included studies was varied, and more rigorous studies were needed in the future. Regarding the intervention domain, more studies focused on the effects of digital interventions for receptive language skills training than for expressive language skills training. Therefore, results showing that receptive language skills training is more effective need to be further examined. Regarding the targeting population, further research could examine the efficacy of digital interventions in children with DLD with different predominant language impairments separately to explore the optimal population for digital interventions.

Conclusions

This systematic review indicates that phonological skills are the most frequently targeted language domain by digital interventions in children with DLD. Given the limited number and the heterogeneity of the studies included, it is still unclear whether digital intervention was effective in improving different language skills in children with DLD. There is less evidence supporting that digital interventions are effective for expressive language skills. Due to the lack of normality of children's language development and speech analysis technology, digital programs are unable to respond to children's errors in a timely manner. Addressing these issues is critical to ensure that mobile health technology is effective in expressive language interventions for children with DLD. Further higher-level evidence, such as RCT studies in this area, is also needed to direct future updates to the digital programs. Digital intervention could be complementary to regular in-person language interventions.

Acknowledgments

The authors appreciate the librarian at Jinan University for their technical support. This review was supported by the Guangzhou Municipal Basic and Applied Basic Research Program (grant 2024A04J4181), the Fundamental Research Funds for the Central Universities (grant 21623324), the Zhongshan City Social Welfare and Basic Research Project (grant 2021B118), the Science and Technology Projects in Guangzhou (grant 202201020082), the Key Realm R&D Program of Guangdong Province (grant 2019B030335001), the Guangdong Medical Science and Technology Research Foundation of China (grant A2023353), and the Administration of Traditional Chinese Medicine of Guangdong Province (grant 20242015).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description and main outcomes of digital intervention studies for children with developmental language disorder.
[DOCX File, 30 KB - mhealth_v13i1e59992_app1.docx]

Multimedia Appendix 2
Fidelity scores of the included studies.
[XLSX File, 13 KB - mhealth_v13i1e59992_app2.xlsx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 86 KB - mhealth_v13i1e59992_app3.pdf]

XSL•FO
RenderX

References

1. Conti-Ramsden G, Durkin K, Toseeb U, Botting N, Pickles A. Education and employment outcomes of young adults with a history of developmental language disorder. Int J Lang Commun Disord 2018 Mar;53(2):237-255. [doi: 10.1111/1460-6984.12338] [Medline: 29139196]
2. Norbury CF, Gooch D, Wray C, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. J Child Psychol Psychiatry 2016 Nov;57(11):1247-1257. [doi: 10.1111/jcpp.12573] [Medline: 27184709]
3. Wu S, Zhao J, de Villiers J, et al. Prevalence, co-occurring difficulties, and risk factors of developmental language disorder: first evidence for Mandarin-speaking children in a population-based study. Lancet Reg Health West Pac 2023 May;34(100713):100713. [doi: 10.1016/j.lanwpc.2023.100713] [Medline: 37283967]
4. Developmental language disorder. WHO-FIC Foundation. 2022. URL: https://icd.who.int/dev11/f/en#/http%3a%2f%2fid.who.int%2ficd%2fentity%2f862918022 [accessed 2024-12-18]
5. Zupan B, Hutchings SM, Everitt LE, Gupta C. Language disorder and internalizing mental health problems in youth offenders: a systematic review. Int J Lang Commun Disord 2022 Nov;57(6):1207-1228. [doi: 10.1111/1460-6984.12759] [Medline: 35841339]
6. Rinaldi S, Caselli MC, Cofelice V, et al. Efficacy of the treatment of developmental language disorder: a systematic review. Brain Sci 2021 Mar 23;11(3):407. [doi: 10.3390/brainsci11030407] [Medline: 33806938]
7. Nicola K, Watter P. The comparison of perceived health-related quality of life between Australian children with severe specific language impairment to age and gender-matched peers. BMC Pediatr 2018 Feb 14;18(1):62. [doi: 10.1186/s12887-018-1058-2] [Medline: 29444654]
8. Le HND, Mensah F, Eadie P, et al. Health-related quality of life of children with low language from early childhood to adolescence: results from an Australian longitudinal population-based study. J Child Psychol Psychiatry 2021 Mar;62(3):349-356. [doi: 10.1111/jcpp.13277] [Medline: 32488955]
9. Lee YC, Chen VCH, Yang YH, et al. Association between emotional disorders and speech and language impairments: a national population-based study. Child Psychiatry Hum Dev 2020 Jun;51(3):355-365. [doi: 10.1007/s10578-019-00947-9] [Medline: 31802296]
10. Toseeb U, Oginni OA, Dale PS. Developmental language disorder and psychopathology: disentangling shared genetic and environmental influences. J Learn Disabil 2022;55(3):185-199. [doi: 10.1177/00222194211019961] [Medline: 34112015]
11. Ottosson S, Schachinger Lorentzon U, Kadesjö B, Gillberg C, Miniscalco C. Neurodevelopmental problems and quality of life in 6-year-olds with a history of developmental language disorder. Acta Paediatr 2022 Jan;111(1):115-122. [doi: 10.1111/apa.16104] [Medline: 34516681]
12. Norbury C, Griffiths S, Goh SKY, Boyes M, Hill E, Viding E. Developmental language disorder: a hidden condition. Lancet Child Adolesc Health 2024 Jul;8(7):473-475. [doi: 10.1016/S2352-4642(24)00016-6] [Medline: 38330984]
13. Tseng YT, Hsu HJ. Not only motor skill performance but also haptic function is impaired in children with developmental language disorder. Res Dev Disabil 2023 Mar;134(104412):104412. [doi: 10.1016/j.ridd.2022.104412] [Medline: 36638673]
14. Goh SKY, Griffiths S, Norbury CF, SCALES Team. Sources of variability in the prospective relation of language to social, emotional, and behavior problem symptoms: Implications for developmental language disorder. J Abnorm Psychol 2021 Aug;130(6):676-689. [doi: 10.1037/abn0000691] [Medline: 34553962]
15. Nitin R, Shaw DM, Rocha DB, et al. Association of developmental language disorder with comorbid developmental conditions using algorithmic phenotyping. JAMA Netw Open 2022 Dec 1;5(12):e2248060. [doi: 10.1001/jamanetworkopen.2022.48060] [Medline: 36580336]
16. Diepeveen FB, van Dommelen P, Oudesluys-Murphy AM, Verkerk PH. Children with specific language impairment are more likely to reach motor milestones late. Child Care Health Dev 2018 Nov;44(6):857-862. [doi: 10.1111/cch.12614] [Medline: 30155913]
17. McGregor KK, Ohlmann N, Eden N, Arbisi-Kelm T, Young A. Abilities and disabilities among children with developmental language disorder. Lang Speech Hear Serv Sch 2023 Jul 5;54(3):927-951. [doi: 10.1044/2023_LSHSS-22-00070] [Medline: 37159846]
18. Chilosi AM, Brovedani P, Cipriani P, Casalini C. Sex differences in early language delay and in developmental language disorder. J Neurosci Res 2023 May;101(5):654-667. [doi: 10.1002/jnr.24976] [Medline: 34822733]
19. Clegg J, Hollis C, Mawhood L, Rutter M. Developmental language disorders—a follow-up in later adult life. Cognitive, language and psychosocial outcomes. J Child Psychol Psychiatry 2005 Feb;46(2):128-149. [doi: 10.1111/j.1469-7610.2004.00342.x] [Medline: 15679523]
20. Tohidast SA, Mansuri B, Bagheri R, Azimi H. Provision of speech-language pathology services for the treatment of speech and language disorders in children during the COVID-19 pandemic: problems, concerns, and solutions. Int J Pediatr Otorhinolaryngol 2020 Nov;138:110262. [doi: 10.1016/j.ijporl.2020.110262] [Medline: 32705994]
21. Saeedi S, Bouraghi H, Seifpanahi MS, Ghazisaeedi M. Application of digital games for speech therapy in children: a systematic review of features and challenges. J Healthc Eng 2022;2022:4814945. [doi: 10.1155/2022/4814945] [Medline: 35509705]

XSL•FO
RenderX

22. Zwitserlood R, Harmsel MT, Schulting J, Wiefferink K, Gerrits E. To game or not to game? Efficacy of using tablet games in vocabulary intervention for children with DLD. Appl Sci (Basel) 2022;12(3):1643. [doi: 10.3390/app12031643]
23. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020 Oct 19;22(10):e17298. [doi: 10.2196/17298] [Medline: 33074157]
24. Higgins PTJ, Thomas J, Chandler J, Cumpston M, Li T, et al. Cochrane Handbook for Systematic Reviews of Interventions, 1st edition: Wiley; 2019. [doi: 10.1002/9781119536604]
25. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34. [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
26. Tifratene K, Robert P, Metelkina A, Pradier C, Dartigues JF. Progression of mild cognitive impairment to dementia due to AD in clinical settings. Neurology (ECronicon) 2015 Jul 28;85(4):331-338. [doi: 10.1212/WNL.0000000000001788] [Medline: 26136516]
27. Bishop D, Adams CV, Lehtonen A, Rosen S. Effectiveness of computerised spelling training in children with language impairments: a comparison of modified and unmodified speech input. J Res Reading 2005 May;28(2):144-157. [doi: 10.1111/j.1467-9817.2005.00259.x]
28. Carson KL. Can an app a day keep illiteracy away? Piloting the efficacy of Reading Doctor apps for preschoolers with developmental language disorder. Int J Speech Lang Pathol 2020 Aug;22(4):454-465. [doi: 10.1080/17549507.2019.1667438] [Medline: 31594401]
29. Chen Y, Lin WJ. Efficacy of an integrated intervention with vocabulary and phonetic training for Mandarin-speaking children with developmental language disorders. Child Lang Teach Ther 2022 Oct;38(3):288-302. [doi: 10.1177/02656590221101180]
30. Dacewicz A, Szymaszek A, Nowak K, Szelag E. Training-induced changes in rapid auditory processing in children with specific language impairment: electrophysiological indicators. Front Hum Neurosci 2018;12(310):310. [doi: 10.3389/fnhum.2018.00310] [Medline: 30131683]
31. Heikkilä J, Lonka E, Meronen A, et al. The effect of audiovisual speech training on the phonological skills of children with specific language impairment (SLI). Child Lang Teach Ther 2018 Oct;34(3):269-287. [doi: 10.1177/0265659018793697]
32. Loeb DF, Gillam RB, Hoffman L, Brandel J, Marquis J. The effects of Fast ForWord Language on the phonemic awareness and reading skills of school-age children with language impairments and poor reading skills. Am J Speech Lang Pathol 2009 Nov;18(4):376-387. [doi: 10.1044/1058-0360(2009/08-0067)] [Medline: 19564439]
33. Cohen W, Hodson A, O'Hare A, et al. Effects of computer-based intervention through acoustically modified speech (Fast ForWord) in severe mixed receptive-expressive language impairment: outcomes from a randomized controlled trial. J Speech Lang Hear Res 2005 Jun;48(3):715-729. [doi: 10.1044/1092-4388(2005/049)] [Medline: 16197283]
34. Gillam RB, Loeb DF, Hoffman LM, et al. The efficacy of Fast ForWord Language intervention in school-age children with language impairment: a randomized controlled trial. J Speech Lang Hear Res 2008 Feb;51(1):97-119. [doi: 10.1044/1092-4388(2008/007)] [Medline: 18230858]
35. Hsu HJ, Bishop DVM. Training understanding of reversible sentences: a study comparing language-impaired children with age-matched and grammar-matched controls. PeerJ 2014;2:e656. [doi: 10.7717/peerj.656] [Medline: 25392757]
36. Bishop DVM, Adams CV, Rosen S. Resistance of grammatical impairment to computerized comprehension training in children with specific and non-specific language impairments. Int J Lang Commun Disord 2006;41(1):19-40. [doi: 10.1080/13682820500144000] [Medline: 16272001]
37. Washington KN, Warr-Leeper G, Thomas-Stonell N. Exploring the outcomes of a novel computer-assisted treatment program targeting expressive-grammar deficits in preschoolers with SLI. J Commun Disord 2011;44(3):315-330. [doi: 10.1016/j.jcomdis.2011.01.002] [Medline: 21288539]
38. Yi A, Chen Z, Ling W, et al. Effectiveness of cloud-based rehabilitation in children with developmental language disorder during the COVID-19 pandemic: a prospective cohort study. Medicine (Baltimore) 2022 Aug 19;101(33):e30056. [doi: 10.1097/MD.0000000000030056] [Medline: 35984120]
39. Barker TH, Stone JC, Sears K, et al. Revising the JBI quantitative critical appraisal tools to improve their applicability: an overview of methods and the development process. JBI Evid Synth 2023 Mar 1;21(3):478-493. [doi: 10.11124/JBIES-22-00125] [Medline: 36121230]
40. Barker TH, Stone JC, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evid Synth 2023 Mar 1;21(3):494-506. [doi: 10.11124/JBIES-22-00430] [Medline: 36727247]
41. Barker TH, Habibi N, Aromataris E, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. JBI Evid Synth 2024 Mar 1;22(3):378-388. [doi: 10.11124/JBIES-23-00268] [Medline: 38287725]
42. Roberts F, Cooper K. Effectiveness of high fidelity simulation versus low fidelity simulation on practical/clinical skill development in pre-registration physiotherapy students: a systematic review. JBI Database System Rev Implement Rep 2019 Jun;17(6):1229-1255. [doi: 10.11124/JBISRIR-2017-003931] [Medline: 30964770]
43. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dent 2011;71(s1):S52-S63. [doi: 10.1111/j.1752-7325.2011.00233.x] [Medline: 21499543]

XSL•FO
RenderX

44. Schuele CM, Boudreau D. Phonological awareness intervention: beyond the basics. Lang Speech Hear Serv Sch 2008 Jan;39(1):3-20. [doi: 10.1044/0161-1461(2008/002)] [Medline: 18162644]
45. Gijbels L, Yeatman JD, Lalonde K, Lee AKC. Audiovisual speech processing in relationship to phonological and vocabulary skills in first graders. J Speech Lang Hear Res 2021 Dec 13;64(12):5022-5040. [doi: 10.1044/2021_JSLHR-21-00196] [Medline: 34735292]
46. Bloom P. Précis of how children learn the meanings of words. Behav Brain Sci 2001 Dec;24(6):1095-1103. [doi: 10.1017/s0140525x01000139] [Medline: 12412326]
47. Xu F, Tenenbaum JB. Word learning as Bayesian inference. Psychol Rev 2007 Apr;114(2):245-272. [doi: 10.1037/0033-295X.114.2.245] [Medline: 17500627]
48. Melby-Lervåg M, Lyster SAH, Hulme C. Phonological skills and their role in learning to read: a meta-analytic review. Psychol Bull 2012 Mar;138(2):322-352. [doi: 10.1037/a0026744] [Medline: 22250824]
49. Moav-Scheff R, Yifat R, Banai K. Phonological memory and word learning deficits in children with specific language impairment: a role for perceptual context? Res Dev Disabil 2015;45-46:384-399. [doi: 10.1016/j.ridd.2015.08.010] [Medline: 26301906]
50. Deppeler JM, Taranto AM, Bench J. Language and auditory processing changes following Fast ForWord. Aust N.Z J Audiol 2004 Nov 1;26(2):94-109. [doi: 10.1375/audi.26.2.94.58276]
51. Strong GK, Torgerson CJ, Torgerson D, Hulme C. A systematic meta-analytic review of evidence for the effectiveness of the “Fast ForWord” language intervention program. J Child Psychol Psychiatry 2011 Mar;52(3):224-235. [doi: 10.1111/j.1469-7610.2010.02329.x] [Medline: 20950285]
52. Michaud M, Roy-Charland A, Richard J, Nazair A, Perron M. The role of vocabulary skills in a storybook-based intervention to stimulate emotion comprehension in preschoolers. J Genet Psychol 2021 Nov 2;182(6):471-487. [doi: 10.1080/00221325.2021.1969884]
53. Griffiths S, Kievit RA, Norbury C. Mutualistic coupling of vocabulary and non-verbal reasoning in children with and without language disorder. Dev Sci 2022 May;25(3):e13208. [doi: 10.1111/desc.13208] [Medline: 34862694]
54. McCarthy KM, Skoruppa K. Language-specific phonological skills and the relationship with reading accuracy in Sylheti-English sequential bilinguals. Child Dev 2023 Mar;94(2):e85-e102. [doi: 10.1111/cdev.13880] [Medline: 36515489]
55. Kehoe M, Poulin-Dubois D, Friend M. Within- and cross-language relations between phonological memory, vocabulary, and grammar in bilingual children. J Speech Lang Hear Res 2021 Dec 13;64(12):4918-4948. [doi: 10.1044/2021_JSLHR-21-00176] [Medline: 34731575]
56. Gervain J. Typical language development. Handb Clin Neurol 2020;173(171-83):171-183. [doi: 10.1016/B978-0-444-64150-2.00016-2] [Medline: 32958172]
57. Law J, Garrett Z, Nye C. Speech and language therapy interventions for children with primary speech and language delay or disorder. Cochrane Database Syst Rev 2003;2003(3):CD004110. [doi: 10.1002/14651858.CD004110] [Medline: 12918003]
58. Finestack LH, Rohwer B, Hilliard L, Abbeduto L. Using computerized language analysis to evaluate grammatical skills. Lang Speech Hear Serv Sch 2020 Apr 7;51(2):184-204. [doi: 10.1044/2019_LSHSS-19-00032] [Medline: 32255745]
59. Pezold MJ, Imgrund CM, Storkel HL. Using computer programs for language sample analysis. Lang Speech Hear Serv Sch 2020 Jan 8;51(1):103-114. [doi: 10.1044/2019_LSHSS-18-0148] [Medline: 31697609]
60. Frizelle P, Tolonen AK, Tulip J, Murphy CA, Saldana D, McKean C. The influence of quantitative intervention dosage on oral language outcomes for children with developmental language disorder: a systematic review and narrative synthesis. Lang Speech Hear Serv Sch 2021 Apr 20;52(2):738-754. [doi: 10.1044/2020_LSHSS-20-00058] [Medline: 33465314]
61. Frizelle P, Tolonen AK, Tulip J, Murphy CA, Saldana D, McKean C. The Impact of intervention dose form on oral language outcomes for children with developmental language disorder. J Speech Lang Hear Res 2021 Aug 9;64(8):3253-3288. [doi: 10.1044/2021_JSLHR-20-00734] [Medline: 34213951]
62. Mustonen R, Torppa R, Stolt S. Screen time of preschool-aged children and their mothers, and children’s language development. Children (Basel) 2022 Oct 18;9(10):36291513. [doi: 10.3390/children9101577] [Medline: 36291513]
63. Madigan S, McArthur BA, Anhorn C, Eirich R, Christakis DA. Associations between screen use and child language skills: a systematic review and meta-analysis. JAMA Pediatr 2020 Jul 1;174(7):665-675. [doi: 10.1001/jamapediatrics.2020.0327] [Medline: 32202633]
64. Karani NF, Sher J, Mophosho M. The influence of screen time on children’s language development: a scoping review. S Afr J Commun Disord 2022 Feb 9;69(1):e1-e7. [doi: 10.4102/sajcd.v69i1.825] [Medline: 35144436]
65. Dy ABC, Dy ABC, Santos SK. Measuring effects of screen time on the development of children in the Philippines: a cross-sectional study. BMC Public Health 2023 Jun 28;23(1):1261. [doi: 10.1186/s12889-023-16188-4] [Medline: 37380949]
66. Murphy SL, Gutman SA. Intervention fidelity: a necessary aspect of intervention effectiveness studies. Am J Occup Ther 2012;66(4):387-388. [doi: 10.5014/ajot.2010.005405] [Medline: 22742685]

Abbreviations

DLD: developmental language disorder
FFW: Fast ForWord

XSL•FO
RenderX

JBI: Joanna Briggs Institute
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized control trial
SLP: speech and language pathologist

Edited by L Buis; submitted 28.04.24; peer-reviewed by JK Kraljevic, S Harding; revised version received 21.01.25; accepted 15.04.25; published 23.05.25.

Please cite as:
Zhou Z, Deng C, Yin D, Yang Q, Chen Z
Digital Intervention in Children With Developmental Language Disorder: Systematic Review
JMIR Mhealth Uhealth 2025;13:e59992
URL: https://mhealth.jmir.org/2025/1/e59992
doi:10.2196/59992

© Zhaowen Zhou, Cheng Deng, Dongling Yin, Qiaoxue Yang, Zhuoming Chen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Telemedicine on Wound-Related and Patient-Reported Outcomes in Patients With Chronic Wounds: Systematic Review and Meta-Analysis

Xiaoyan Zhang[1], PhD; Zhanghui Guo[2], MSN; Yu Duan[2], MSN; Chao Sun[3], MSN; Jiayin Luo[1], MSN

[1]Department of Vascular Surgery, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, 1 Dahua Road, Dongdan, Dongcheng District, Beijing, China

[2]School of Nursing, Beijing University of Chinese Medicine, Beijing, China

[3]Department of Nursing, Beijing Hospital, National Center of Gerontology; Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China

Corresponding Author:
Jiayin Luo, MSN
Department of Vascular Surgery, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, 1 Dahua Road, Dongdan, Dongcheng District, Beijing, China

Abstract

Background: Telemedicine may provide new vitality and opportunities to the field of wound care and has been advocated as being a potential and feasible strategy for chronic wound management.

Objective: This systematic review and meta-analysis aimed to assess the effectiveness of telemedicine on wound-related outcomes and patient-reported outcomes in patients with chronic wounds.

Methods: A comprehensive search of 9 databases, including PubMed, Embase, PsycINFO, the Cochrane Library, CINAHL, Web of Science, the China National Knowledge Infrastructure database, the Wanfang database, and the VIP database, was performed to identify eligible randomized controlled trials that investigated the effectiveness of telemedicine for patients with chronic wounds. The primary outcome was wound healing, including healing score, healing time, and healing rate. The quality of the included studies was examined via the Cochrane risk-of-bias tool. Data synthesis was conducted via Review Manager (version 5.4; the Cochrane Collaboration). Due to anticipated heterogeneity, a random-effects meta-analysis was used. Effect estimates are presented as risk ratio (RR) or standard mean differences (SMDs) with 95% CI. The quality of the evidence was assessed via the Grading of Recommendations, Assessment, Development, and Evaluation approach.

Results: A total of 22 randomized controlled trials involving 2397 participants met the inclusion criteria. This review demonstrated that telemedicine significantly improved the healing score (SMD −1.46, 95% CI −2.27 to −0.66; P<.001; I^2=95%; P<.001), healing time (SMD −0.47, 95% CI −0.92 to 0.02; P=.04; I^2=85%; P<.001), amputation rate (RR 0.52, 95% CI 0.31-0.88; P=.02; I^2=23%; P=.28), pain (SMD−0.62, 95% CI −0.90 to −0.34; P<.001; I^2=0%; P=.32), and quality of life (SMD 1.90, 95% CI 0.32-3.48; P=.02; I^2=98%; P<.001). Although the meta-analysis results indicated that telemedicine enhanced the healing rate (RR 1.16, 95% CI 1.02-1.33; P=.03; I^2=50%; P=.03), potential publication bias was detected (Egger test, bias=1.801; SE 0.367; P<.001). Upon imputing the missing studies using the trim-and-fill method, the recalculated pooled RR was adjusted, resulting in a new estimate of RR 1.06 (95% CI 0.98-1.15; P=.16). In addition, no significant differences were found in mortality, depression, anxiety, or patient satisfaction.

Conclusions: There is some evidence that telemedicine contributes to improvements in the healing score, healing time, amputation rate, pain, and quality of life of patients with chronic wounds. Nevertheless, further high-quality studies are essential to examine the impact of telemedicine on healing rate and patient-reported outcomes in patients with chronic wounds.

Trial Registration: PROSPERO CRD42023462475; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=462475

(JMIR Mhealth Uhealth 2025;13:e58553) doi:10.2196/58553

KEYWORDS

telemedicine; chronic wound; systematic review; wound healing; meta-analysis; patient-reported outcome; randomized controlled trial

XSL•FO
RenderX

Introduction

Background

A chronic wound is defined as a wound that fails to achieve anatomical and functional integrity through a normal and orderly repair process within 4 weeks [1,2]. Chronic wounds pose a significant global challenge and include conditions such as venous ulcers, arterial ulcers, diabetic foot ulcers, and pressure injuries. The global incidence of chronic wounds is approximately 2.21 per 1000 people [3]. According to previous studies, the treatment of chronic wounds accounts for between 3% and 5.5% of all medical expenses in high-income countries [4]. The cost of hospitalization for diabetes-related amputations ranges from US $12,851 to US $16,267 per admitted patient [5]. Additionally, the cost for leg ulcers ranges from US $44,900 to US $23,700 per patient for primary and outpatient care [5]. Moreover, patients with chronic wounds frequently experience discomfort, infection, anxiety, despair, and even the possibility of amputation. These factors profoundly impact the physical and mental well-being of patients as well as their quality of life [6]. As the aging population grows and the prevalence of chronic diseases (such as diabetes) continues to increase, the incidence and economic burden associated with chronic wounds are rapidly escalating [7,8].

Patients with chronic wounds typically experience complex conditions, thus necessitating long hospital stays averaging 13 days [9]. Due to constraints such as bed turnover rates and hospitalization costs, these patients often do not fully recover in hospital settings and must continue receiving care at home. This transition poses challenges, especially when primary care nurses and family caregivers lack adequate knowledge and resources for effective wound management. Over half of nurses have insufficient knowledge of treatment methods and wound care evaluation, and only 36.74% (n=313) can easily access essential information [10]. Family members, who spend approximately 10 hours daily on caregiving, are overwhelmed by the demands of home-based care [11]. This gap in continuous, effective management has led to frequent hospital readmissions, thus resulting in worsened patient despair and hopelessness. Therefore, addressing these issues via technological innovations in chronic wound management is critical.

The adoption of telemedicine has markedly increased since the COVID-19 pandemic, as evidenced by significant policy efforts being enacted throughout the world. In 2022, the World Health Organization released its Consolidated Telemedicine Implementation Guide [12], whereas the Registered Nurses Association of Ontario introduced the Clinical Practice in a Digital Health Environment guideline [13]. Similarly, China has advanced its digital health care through key policies such as the “14th Five-Year Plan” for Digital Economy Development (Guofa [2021] No 29) [14] and the “National Nursing Development Plan (2021 - 2025)” (Guofa [2022] No 15) [15]. As a result, telemedicine is now a vital tool for both patients and a wide range of health care professionals. Telemedicine, which enables the distribution of health care services and information via electronic and telecommunications technologies, supports long-distance patient care, education, and monitoring [16,17]. Its applications are diverse and encompass diabetes management [18], exercise regimens for coronary heart disease [19], and stroke care [20]; moreover, it holds significant potential for managing chronic wounds. By facilitating remote diagnosis and treatment, telemedicine offers continuous care, particularly to patients in remote locations, thereby optimizing medical resource use [21]. It also allows health care providers to deliver targeted chronic wound care training and supervision via remote platforms, thus improving patient self-management and ensuring precise wound care [22,23].

Objective

Given these advantages, telemedicine may provide new vitality and opportunities for the management of chronic wounds. Researchers have begun to verify its effectiveness and safety in the remote care of patients with chronic wounds. However, the sample size of a single study is small, and there are differences in the types of chronic wounds, forms of intervention, intervention time, number of types of telemedicine, and whether telemedicine is associated with a face-to-face component. The results of a systematic review [24] revealed that there was no significant difference in the efficacy or safety of telemedicine compared with traditional treatment methods; however, this study still had several limitations. First, this review searched only 3 databases and included only a few randomized controlled trials (RCTs), although more RCTs have been published in recent years [25-32]. Furthermore, this review reported only wound-related outcomes and did not summarize evidence related to patient-reported outcomes (PROs), such as pain, depression, anxiety, and quality of life. However, PROs play a crucial role in treating chronic wounds [33]. These outcomes can help researchers to understand patients’ attitudes and needs as well as comprehensively evaluate the role of interventions in improving patients’ overall health status [34]. Therefore, we aimed to conduct a systematic review and meta-analysis of RCTs to assess the effectiveness of telemedicine on wound-related outcomes and PROs in patients with chronic wounds.

Methods

This study followed the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) statement and checklist [35] (Checklist 1). The protocol of this review was registered in PROSPERO (CRD42023462475).

Data Sources and Search Strategy

A 3-step search strategy was used. The first search was conducted in PubMed and the China National Knowledge Infrastructure database. The used search terms were located by analyzing keywords related to chronic wounds, telemedicine, and study design within titles and abstracts, as well as the index terms that were used to describe the papers. The second search was performed by using all of the identified search terms across 9 databases from their inception to the present time. These databases included PubMed, Embase, PsycINFO, the Cochrane Library, CINAHL, Web of Science, the China National Knowledge Infrastructure database, the Wanfang database, and the VIP database. The search strategy for each database was

XSL•FO
RenderX

discussed by all of the reviewers with literature search expertise (see Multimedia Appendix 1 for the full search strategy). The third search was performed by reviewing the reference lists of the identified papers for additional relevant studies. Publications in English or Chinese were included, with no restrictions on the year of publication.

Study Selection

The rationale for the selection of inclusion or exclusion criteria was based on the population, intervention, control, outcome, study design framework [36]: (1) population: individuals of all ages with chronic wounds, (2) intervention: telemedicine as the main intervention, (3) control: any control condition (eg, usual care, treatment as usual, or nontelemedicine interventions), (4) outcome: reported wound-related outcomes (eg, wound healing, amputation rate, mortality, and economic evaluation) or PROs (eg, pain, quality of life, depression, anxiety, and patient treatment and care satisfaction). Wound healing was the primary outcome, others were the secondary outcomes. While our registered protocol initially included sleep as a PRO, this outcome was subsequently excluded, as it was not central to our primary research questions, and none of the included studies reported sleep-related outcomes; and (5) study design: RCTs. The exclusion criteria were as follows: (1) studies identified as protocols and (2) studies not published in English or Chinese. Study selection was conducted in 2 distinct stages. In the first stage, 2 reviewers (ZG and YD) independently screened all titles and abstracts against the predefined inclusion and exclusion criteria. In the second stage, full-text papers that potentially met the inclusion criteria were retrieved and independently evaluated by 1 reviewer (ZG) who assessed all papers, while the second reviewer (YD) independently assessed a randomly selected 30% (n=22) subset of full texts. If disagreements were identified in 2 or more papers within this 30% (n=22) subset, the second reviewer would then assess the remaining 70% (n=22) of papers to ensure methodological rigor. Throughout both stages, all disagreements between the 2 reviewers were documented and resolved through discussion. When consensus could not be reached, a third reviewer (JL) was consulted to make the final determination regarding study inclusion.

Data Extraction

The data from the included studies were extracted by ZG and then cross-checked by YD. Any discrepancies were resolved by a third reviewer (CS). For included RCTs, we extracted the following data as recommended in the Cochrane Handbook for Systematic Reviews of Interventions [37]: (1) general information: author, country, publication language, publication year, and journal citation; (2) participants: inclusion and exclusion criteria, sample size, baseline characteristics, and setting; (3) interventions: details of telemedicine according to the TIDieR (Template for the Intervention Description and Replication)-telehealth checklist [38]; (4) risk of bias in trials; (5) follow-up: length of follow-up, the reason for and number of dropouts and withdrawals, and method of analysis; and (6) outcome measures: mean and SD for continuous outcomes and the number of events for dichotomous outcomes. In instances where the data were missing, unclear, or incomplete, attempts were made to contact the authors via email for further clarification. In cases where the SD was not reported with means and the necessary information was not obtained from the trial authors, it was imputed based on the information provided, such as the SE, 95% CI, or *P* values, following the guidelines in the *Cochrane Handbook*. Alternatively, if the SD for the missing outcome was not available, it was assumed to be the average of the SDs from trials where this information was reported.

Risk of Bias

The methodological quality of the included studies was independently assessed by 2 reviewers (XZ and JL) via version 2 of the Cochrane tool to assess the risk of bias in randomized trials [39]. It included 5 domains: “randomization process,” “deviations from intended interventions,” “missing outcome data,” “measurement of the outcome,” and “selection of the reported result.” Any disagreements were resolved by a third reviewer (JL).

Statistical Analysis

Pooled analyses were performed using Review Manager (version 5.4; The Cochrane Collaboration). Given the anticipated high levels of heterogeneity due to the diverse forms of telemedicine, the meta-analyses were based on the random - effects model to obtain the most conservative result. Where available, analyses were based on intention-to-treat data from the individual trials. Furthermore, when possible, end-of-treatment scores (rather than change-from-baseline scores) were extracted for the continuous outcomes. Dichotomous data are presented as the risk ratio (RR) and 95% CI, whereas continuous variables are expressed as the mean difference and 95% CI. Standardized mean differences (SMDs) were calculated for continuous outcomes that were measured or reported differently. Narrative synthesis was performed when the quantitative synthesis was judged to be unsuitable.

Heterogeneity was assessed via the Cochran *Q* test and I^2 statistics. If substantial heterogeneity was found (I^2 of 50% or more and a *P* value of *Q* test<.10), the possible causes by undertaking a prespecified subgroup analysis were reported and investigated. Planned subgroup analyses (where applicable) were performed to compare effect estimates between studies based on the following criteria: (1) the average age of participants (ie, 18 - 60 or ≥60 years), (2) study duration (ie,≤6, 6 - 12, or>12 months), (3) number of types of telemedicine (ie, 1 or ≥2), (4) types of chronic wounds (ie, mixed chronic wounds, diabetic foot ulcers, stress injuries, or static ulcers), (5) whether telemedicine was used in conjunction with a face-to-face component (ie, yes or no), and (6) whether telemedicine included a communication component (ie, yes or no). Sensitivity analyses were performed for primary outcomes to determine whether the conclusions were robust to arbitrary decisions made regarding eligibility and analysis. When substantial heterogeneity existed, sensitivity analysis was also conducted to further investigate the potential sources of heterogeneity. Potential publication bias was evaluated by using funnel plots and the Egger test using R statistical software (version 4.4.2; R Foundation for Statistical Computing) when sufficient studies were available. If publication bias was detected, a trim-and-fill method was used

XSL•FO
RenderX

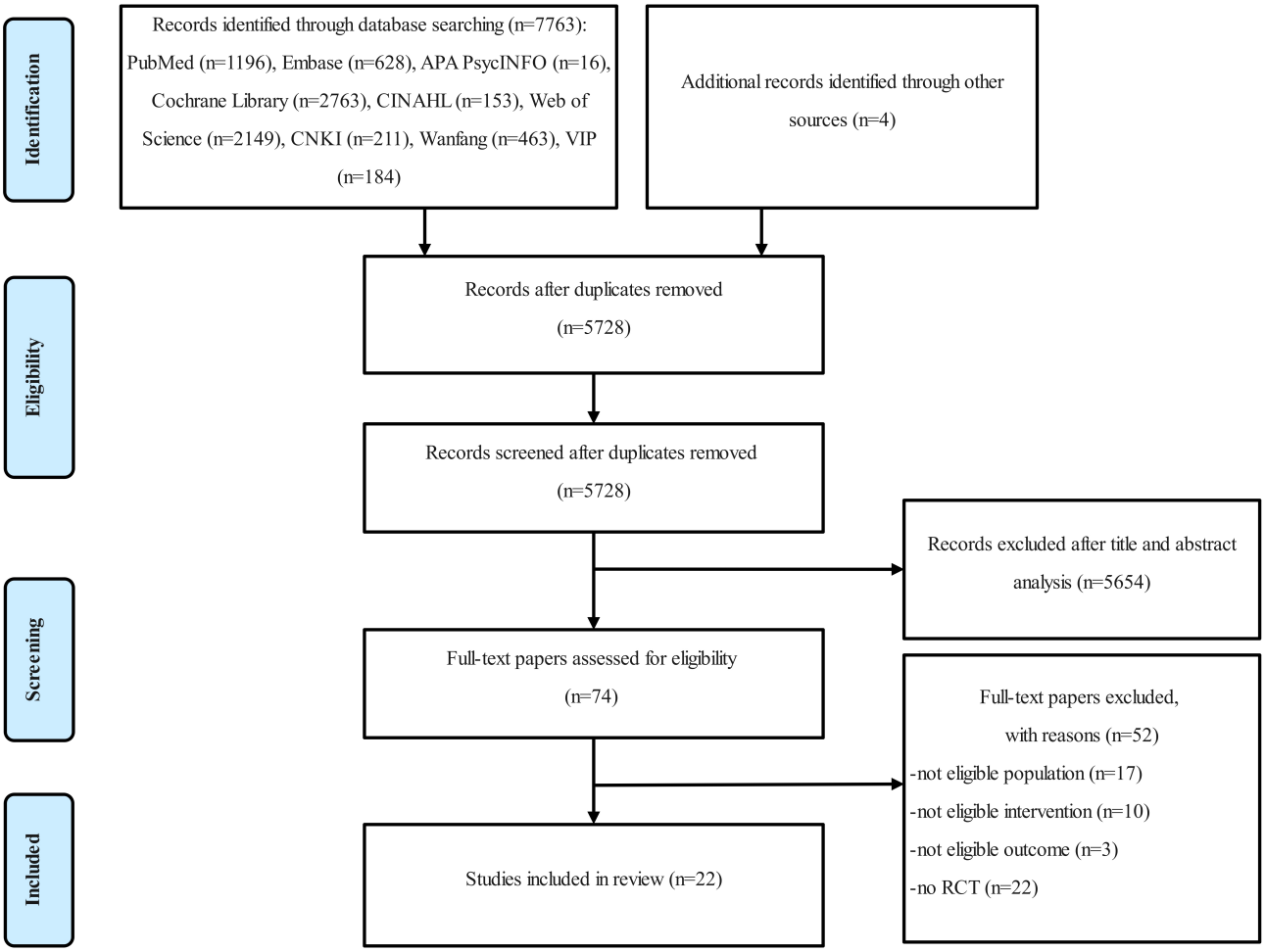
to adjust for publication bias. Two reviewers (XZ and JYL) assessed the certainty of evidence for hospitalization using the Grading of Recommendations, Assessment, Development, and Evaluation approach. Any disagreements were resolved by the third reviewer (JL).

Results

Overview

A total of 7767 records were retrieved from the initial database search and reference lists of key papers. Ultimately, 22 RCTs [25-32,40-53] involving 2397 patients met the inclusion criteria. Figure 1 displays the PRISMA flow diagram.

Figure 1. Flow diagram of the study screening process. CNKI: China National Knowledge Infrastructure; RCT: randomized controlled trial.



Study Characteristics

The basic characteristics of the included studies are presented in Multimedia Appendix 2 [25-32,40-53]. These studies were published between 2004 and 2023, with 12 studies conducted in China [25-28,31,32,40,41,44,46,47,50], 2 in the United States [51,52], 2 in Australia [43,53], 2 in Norway [29,42], 1 in Denmark [45], 1 in Canada [48], 1 in France [30], and 1 in the United Kingdom [49]. Among the studies, 11 [25,28,30,31,40,41,49-53] focused on mixed chronic wounds, 5 [26,27,29,42,45] focused on diabetic foot ulcers, 5 [32,43,44,46,48] focused on stress injuries, and 1 [47] focused on static ulcers. Various telemedicine methods were used across the studies, including mobile apps (eg, wound management systems, WeChat, and Tencent QQ), remote robots, web-based interventions, emails, phone calls, and multimedia messaging services (see Multimedia Appendix 2 for details). A total of 10 studies [25,28,30,31,40,43,49,50,52,53] included 1 telemedicine component, and the remaining 12 studies [26,27,29,32,41,42,44-48,51] used 2 or more telemedicine components. In addition, 19 studies [25-32,40-48,51,52] integrated telemedicine with a communication component, and 13 studies [25-28,32,40-42,44-46,48,52] involved telemedicine in conjunction with a face-to-face component. The intervention duration varied among the included studies and ranged from 2 weeks to 1 year. In total, 15 studies [25,27,28,30-32,40,41,43,44,46,47,49,50,52] had a short duration (≤6 months), 6 studies [26,29,42,45,48,53] had a medium duration (>6 to ≤12 months), and 1 study [51] had a long duration (>12 months). A total of 15 interventions [25,29,30,32,40,42-45,48-53] were delivered individually, 6 [27,28,31,41,46,47] were delivered in a mixed manner, and 1 [26] was delivered in a group. Most of the telemedicine procedures involved tailoring interventions via individualized assessments, individualized guidance, individualized question answering, or individualized strategies. The comparison conditions included usual care, standard care, traditional care, and nontelemedicine components. The number of participants

XSL•FO
RenderX

in the included studies ranged from 26 to 374, and their mean age ranged from 35 to 83 years. A total of 13 studies [28-32,40,42,44-46,48,49,53] included a slightly older participant population, with an average age of ≥60 years observed in both groups. All of the reported outcomes, including wound-related outcomes and PROs, were measured after treatment.

Risk of Bias

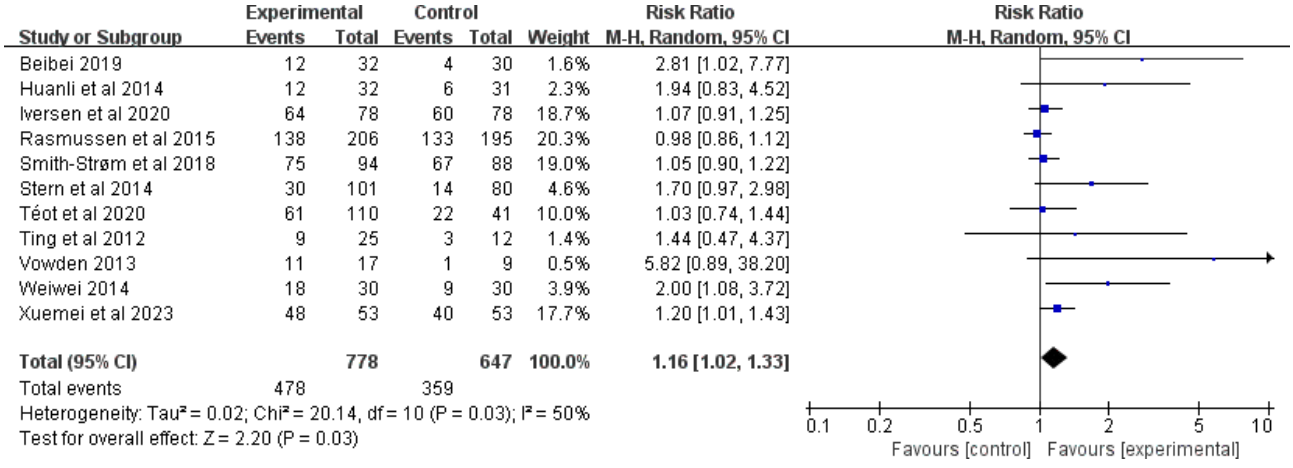
A total of 21 studies [25-32,40-49,51-53] were categorized as having "some concerns" in terms of the overall risk of bias, whereas 1 study [50] was deemed to have a "high risk." All of the studies were designated as "randomized," with 13 studies [25,26,28,29,31,40-44,47-49] using a computer-generated random number generator or a random number table. In contrast, other studies [27,30,32,45,46,51,52] only mentioned random allocation and did not clarify the allocation method. Given the nature of the telemedicine, participants' awareness of their assigned interventions during the trial was plausible, and a lack of information about whether researchers blinded the methods resulted in a high judgment of "some concerns." All of the studies [25-32,40-53] included the outcome data of the vast majority of the subjects, and the analysis results were robust. Thus, all of the studies were assessed as having a low risk of bias. However, no studies provided sufficient information to determine whether the blinding of outcome assessments was achieved, thus leading to the assumption that the assessors may have been aware of the interventions received by the study participants. Thus, all of the studies [25-32,40-53] had "some concerns" regarding the measurements of the outcomes. Six studies [29,30,42,43,45,48] registered study protocols, reported prespecified outcomes, and were assessed to be at low risk of reporting bias. A total of 16 studies [25-28,31,32,40,41,44,46,47,49-53] did not have a protocol available, thus resulting in "some concerns" about the risk of bias in the selection of the reported results (see Multimedia Appendix 3 [25-32,40-53] for more details).

Effects of Telemedicine

Wound Healing

A total of 13 studies [25,29,30,41,42,45-51,53] involving 1631 participants reported the wound healing rate. Among these, 2 studies [51,53] were excluded from the quantitative synthesis because of the absence of an exact number of healed wounds or an uneven distribution of wound severity among groups. After pooling, telemedicine significantly increased the healing rate (RR 1.16, 95% CI 1.02-1.33; *P*=.03), with substantial heterogeneity observed between the studies (I^2=50%; *P*=.03; Figure 2). The quality of the evidence was rated as being low because of imprecision and publication bias (see Multimedia Appendix 4 for more details). Subgroup analyses were further performed (see Multimedia Appendix 5 for more details). Short-term telemedicine [25,30,41,46,47,49,50] significantly increased the healing rate (≤6 months; RR 1.47, 95% CI 1.07-2.00; *P*=.02), whereas moderate-term [29,42,45] and long-term [48] telemedicine did not statistically increase the healing rate. Regarding the types of chronic wounds, combining telemedicine for pressure injuries [46,48] (RR 1.77, 95% CI 1.11-2.82; *P*=.02) or venous ulcers [47] (RR 2.00, 95% CI 1.08-3.72; *P*=.03) had a greater effect on the healing rate than diabetic foot ulcers [29,42,45] or mixed chronic wounds [25,30,41,49,50]. The combination of telemedicine in conjunction with a face-to-face component [25,41,42,45-48] had a greater effect on increasing the healing rate (RR 1.22, 95% CI 1.02-1.47; *P*=.03) than the absence of face-to-face components [29,30,49,50].

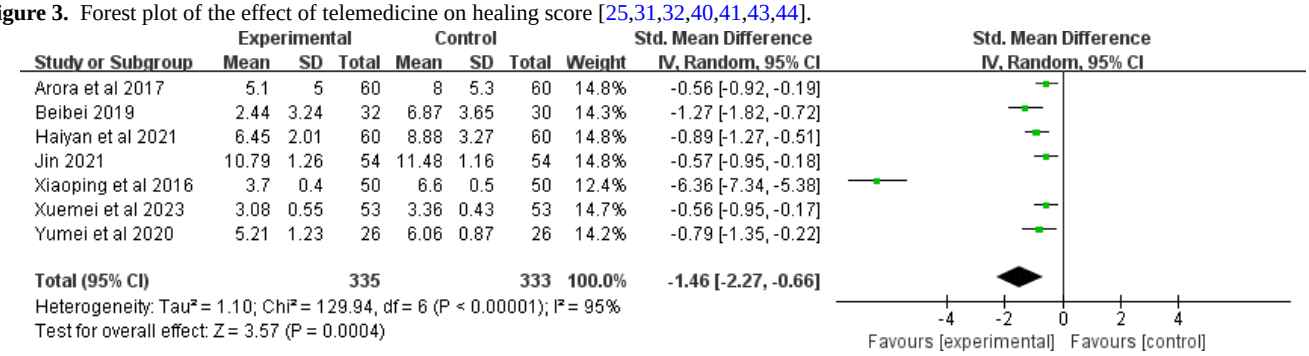
Figure 2. Forest plot of the effect of telemedicine on wound healing rate [25,29,30,41,42,45-50].



A total of 8 studies [25,31,32,40,41,43,44,50] comprising 705 participants reported the healing score; 7 studies [25,31,32,40,41,43,44] used the pressure ulcer scale for healing, and 1 study [50] used a self-designed questionnaire. The study [50] that reported positive results was excluded from the quantitative synthesis due to the use of a different scoring approach. After pooling, telemedicine significantly improved the healing score (SMD −1.46, 95% CI −2.27 to −0.66; *P*<.001), with substantial heterogeneity observed between the studies (I^2=95%; *P*<.001; Figure 3). The quality of the evidence was rated as being moderate because of inconsistency (see Multimedia Appendix 4 for more details). Subgroup analyses were further performed (see Multimedia Appendix 5 for more details). All of the subgroup analyses by average age, type of chronic wound, number of types of telemedicine, and whether the telemedicine was associated with a face-to-face component revealed statistically significant differences.

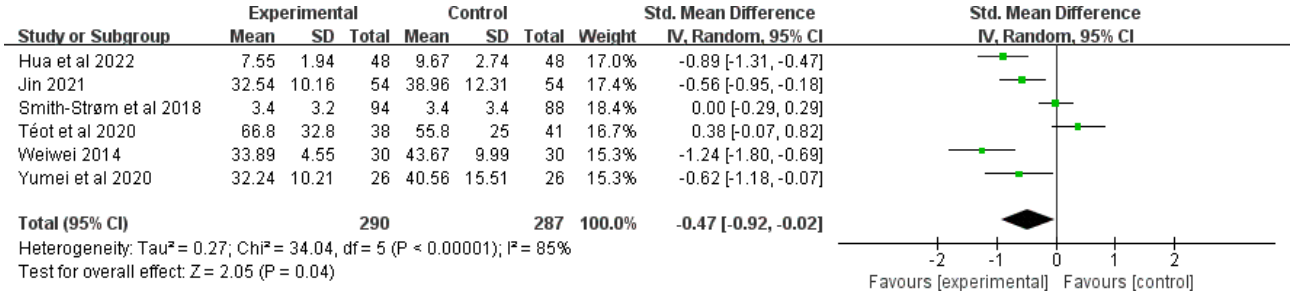
XSL•FO
RenderX

Figure 3. Forest plot of the effect of telemedicine on healing score [25,31,32,40,41,43,44].



A total of 8 studies [28,30,31,40,42,43,47,51] with 810 participants reported the healing time. Among these, 2 studies [43,51] were excluded from the quantitative synthesis because of the absence of an exact healing time or uneven distribution of the severity of the wounds among the groups. After pooling, the evidence suggested that telemedicine decreased the healing time (SMD −0.47, 95% CI −0.92 to −0.02; *P*=.04), although substantial heterogeneity existed among the studies (I^2=85%; *P*<.001; Figure 4). The quality of the evidence was rated as being low because of inconsistency and imprecision (see Multimedia Appendix 4 for more details). Subgroup analyses were further performed (see Multimedia Appendix 5 for more details). Combining telemedicine for venous ulcers [47] (SMD −1.24, 95% CI −1.80 to −0.69; *P*<.001) had a greater effect size on the healing rate than those for diabetic foot ulcers [42] or mixed chronic wounds [28,30,31,40]. Combining telemedicine in conjunction with a face-to-face component [28,40,42,47] indicated a greater effect size for increasing healing time (SMD −0.66, 95% CI −1.25 to −0.08; *P*=.03) than those with no face-to-face components [30,31].

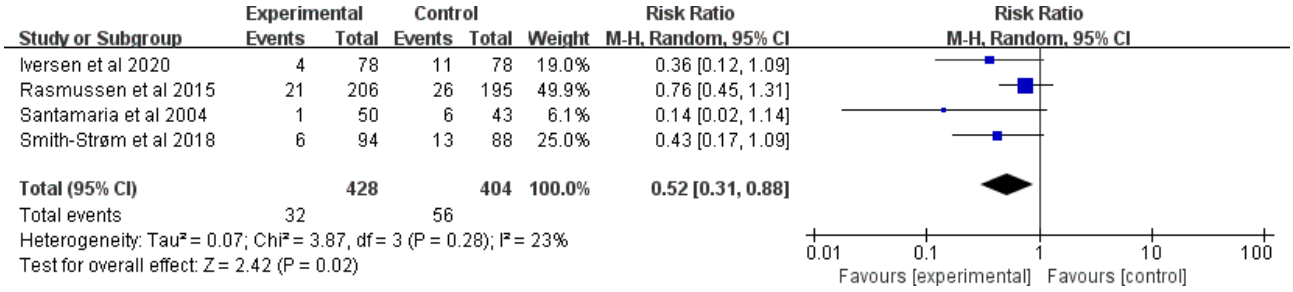
Figure 4. Forest plot of the effect of telemedicine on healing time [28,30,31,40,42,47].



Amputation Rate

A total of 4 studies [29,42,45,53] comprising 832 participants reported the amputation rate. After pooling, telemedicine significantly decreased the amputation rate (RR 0.52, 95% CI 0.31-0.88; *P*=.02) with small and nonsignificant heterogeneity between studies (I^2=23%; *P*=.28; Figure 5). The quality of the evidence was rated as moderate due to the imprecision (see Multimedia Appendix 4 for more details). Subgroup analyses were further performed (see Multimedia Appendix 5 for more details). Combining studies [29,42,45] with more telemedicine types had greater effects on the amputation rate (RR 0.60, 95% CI 0.39-0.93; *P*=.02) than those with 1 telemedicine type [53]. Concerning the types of chronic wounds, combining telemedicine for diabetic foot ulcers [29,42,45] (RR 0.60, 95% CI 0.39-0.93; *P*=.02) had a greater effect size on amputation rate than mixed chronic wounds [53]. Furthermore, combining telemedicine with communication components [29,42,45] had a greater effect size on decreasing the amputation rate (RR 0.60, 95% CI 0.39-0.93; *P*=.02) than telemedicine without communication components [53].

Figure 5. Forest plot of the effect of telemedicine on amputation rate [29,42,45,53].



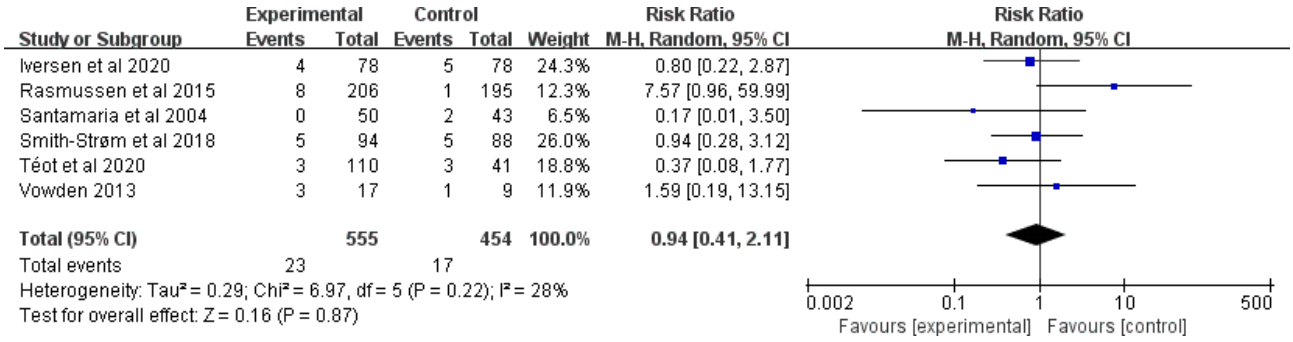
Mortality

A total of 6 studies [29,30,42,45,49,53] involving 1009 participants reported the outcome of mortality. After pooling, the evidence suggested that there was no statistically significant difference in the outcome of mortality (RR 0.94, 95% CI 0.41-2.11; *P*=.87), with small and nonsignificant heterogeneity observed between the studies (I^2=28%; *P*=.22; Figure 6). The

XSL•FO
RenderX

quality of the evidence was rated as being moderate due to imprecision (see Multimedia Appendix 4 for more details). Subgroup analyses were further performed (see Multimedia Appendix 5 for more details). Subgroup analysis by the study duration, number of types of telemedicine, type of chronic wounds, whether telemedicine was associated with a face-to-face component, and whether telemedicine included communication components revealed no statistically significant subgroup differences.

Figure 6. Forest plot of the effect of telemedicine on mortality [29,30,42,45,49,53].



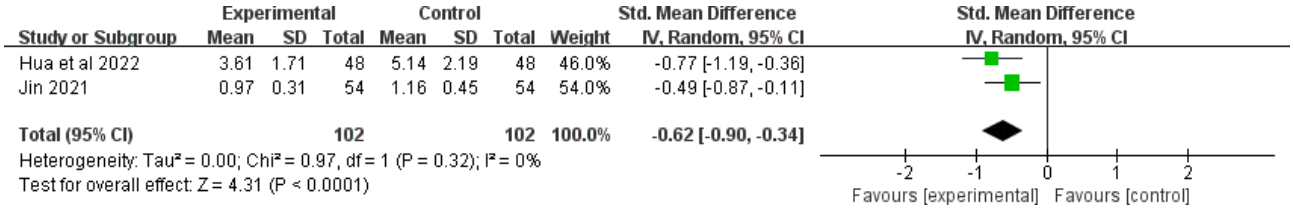
Economic Evaluation

A total of 6 studies [30,43,45,48,51,53] mentioned a cost analysis for the telemedicine intervention versus the control intervention. Santamaria et al [53] reported total cost savings of US $191,935 in the telemedicine group compared with the control group; however, the cost of ongoing management and support for amputees was not included in the analysis. Téot et al [30] demonstrated reduced costs associated with transport for wound management. Rasmussen et al [45,54] indicated that telemedicine follow-up was US $2300 less per patient than standard care, although this difference was not statistically significant (*P*=.42). Arora et al [43,55] reported that the mean total cost per participant in the intervention group was US $2460 compared with US $2391 for the control group. The per-participant cost of delivering the intervention was US $119. Another study [48] revealed that the economic evaluation demonstrated a mean reduction in direct care costs of US $650 per patient compared with the control group. Additionally, 1 study [51] revealed a higher mean total cost per patient due to greater and more severe ulcers in the telemedicine group.

Pain

A total of 4 studies [28,31,48,50] involving 465 participants reported pain outcomes. Among these, 3 studies used the visual analog scale, whereas 1 study used a self-designed questionnaire. Two studies [48,50] were not included in the quantitative synthesis because of the absence of exact data on pain or different scoring approaches. Ting et al [50] reported that telemedicine significantly improved pain compared with the control group (*P*=.004). Stern et al [48] reported that the mean visual analog scale score for wound-specific pain was estimated to be 0.39 units higher; however, the difference was not statistically significant (*P*=.42). Furthermore, after pooling, a significant difference in the outcome of pain was detected between the 2 groups (SMD −0.62, 95% CI −0.90 to −0.34; *P*<.001; Figure 7). The quality of the evidence was rated as being moderate because of imprecision (see Multimedia Appendix 4 for more details).

Figure 7. Forest plot of the effect of telemedicine on pain [28,31].



Quality of Life

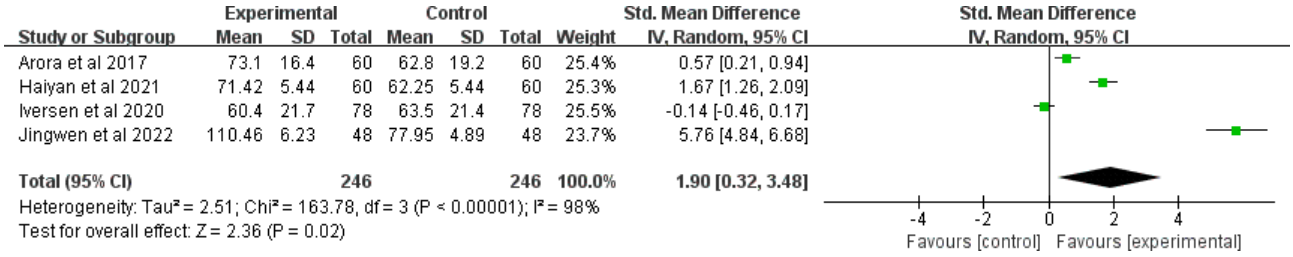
A total of 8 studies [26-29,32,40,41,43] comprising 784 participants reported quality of life outcomes. Among these, 4 studies [27,29,32,43] used the Short Form 36-item Health Survey, whereas the remaining 4 used the European Quality of Life-5 dimensions Questionnaire [29,43], the Diabetes-Specific Quality of Life Scale [27], and the World Health Organization Quality of Life-Brief Version [32]. Four studies [27,29,32,43] indicated that telemedicine significantly increased the overall quality of life (SMD 1.90, 95% CI 0.32-3.48; *P*=.02; Figure 8). The quality of the evidence was rated as being moderate because of inconsistency (see Multimedia Appendix 4 for more details).

Additionally, the remaining 4 studies [26,28,40,41] reported Short Form 36-item Health Survey scores in each dimension. After pooling, telemedicine significantly increased physical functioning (3 RCTs, SMD 7.70, 95% CI 3.91-11.49; *P*<.001), role-physical (4 RCTs, SMD 2.46, 95% CI 0.46-4.46; *P*=.02), bodily pain (4 RCTs, SMD 3.87, 95% CI 1.34-6.40; *P*=.003), general health (4 RCTs, SMD 6.68, 95% CI 2.55-10.81; *P*=.002), vitality (4 RCTs, SMD 9.54, 95% CI 4.29-14.79; *P*=.001), social functioning (4 RCTs, SMD −0.80, 95% CI −2.11 to 0.51; *P*=.23), role-emotional (3 RCTs, SMD 4.61, 95% CI 1.19-8.04; *P*=.008), and mental health (4 RCTs, SMD 4.32, 95% CI 2.69-5.96; *P*<.001). However, no significant difference

XSL•FO
RenderX

was detected in the outcome of the social functioning dimension of the 2 groups.

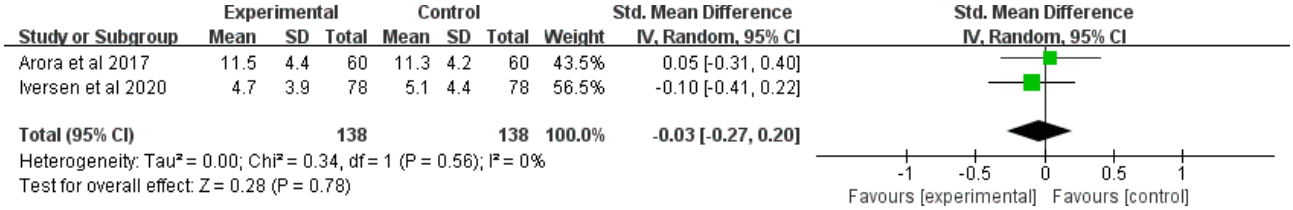
Figure 8. Forest plot of the effect of telemedicine on quality of life [27,29,32,43].



Depression

A total of 2 studies [29,43] involving 276 participants reported the outcome of depression, as measured via the hospital anxiety and depression scale. After pooling, no significant difference was detected in the outcome of depression between the 2 groups (SMD −0.03, 95% CI −0.27 to 0.20; *P*=.78; Figure 9). The quality of the evidence was rated as being moderate due to imprecision (see Multimedia Appendix 4 for more details).

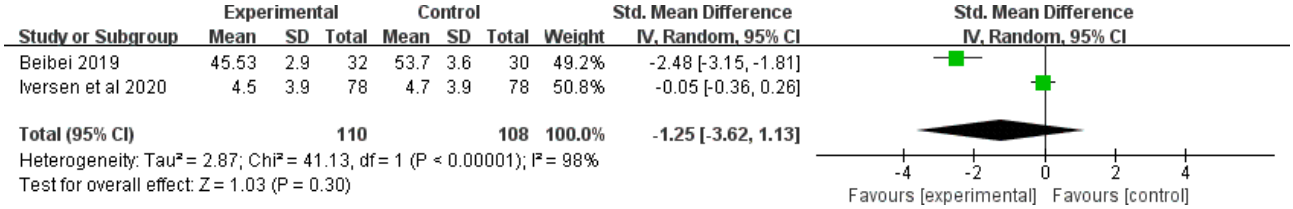
Figure 9. Forest plot of the effect of telemedicine on depression [29,43].



Anxiety

A total of 2 studies [29,41] with 218 participants reported outcomes of anxiety. Among these, 1 study [41] used the self-rating anxiety scale, and 1 study [29] used the hospital anxiety and depression scale. After pooling, no significant difference in the outcome of anxiety was detected between the 2 groups (SMD −1.25, 95% CI −3.62 to 1.13; *P*=.30; Figure 10). The quality of the evidence was rated as being low because of inconsistency and imprecision (see Multimedia Appendix 4 for more details).

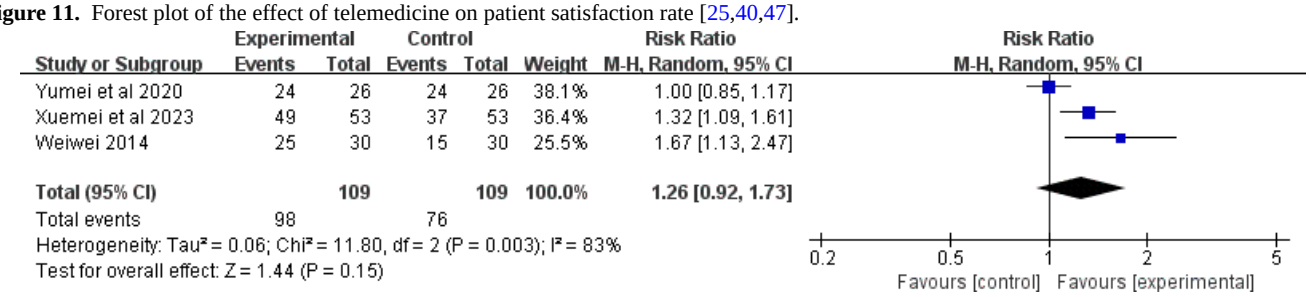
Figure 10. Forest plot of the effect of telemedicine on anxiety [29,41].



Patient Satisfaction

A total of 3 studies [25,40,47] involving 174 participants reported the outcome of the patient satisfaction rate. After pooling, no significant difference was detected in the patient satisfaction rate between the 2 groups (RR 1.26, 95% CI 0.92-1.73; *P*=.15; Figure 11). The quality of the evidence was rated as being low due to inconsistency and imprecision (see Multimedia Appendix 4 for more details).

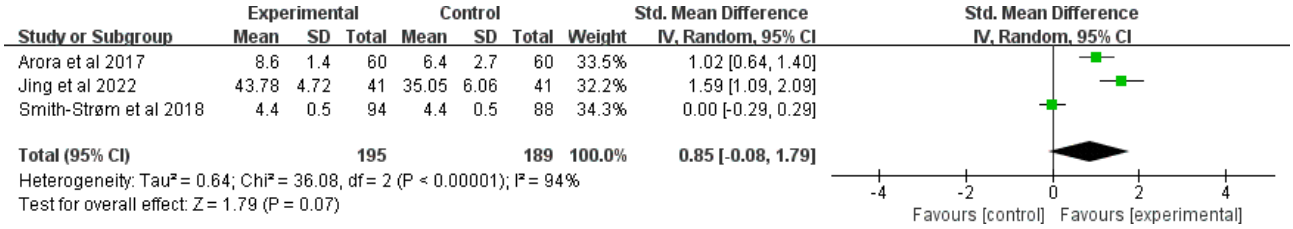
Figure 11. Forest plot of the effect of telemedicine on patient satisfaction rate [25,40,47].



A total of 3 studies [26,42,43] comprising 174 participants reported the outcome of patient satisfaction scores. After pooling, no significant difference was detected in patient satisfaction scores between the 2 groups (SMD 0.85, 95% CI −0.08 to 1.79; *P*=.07; Figure 12). The quality of the evidence was rated as being low due to inconsistency and imprecision (see Multimedia Appendix 4 for more details). Additionally, 1 study [52] reported that telemedicine serves as a valuable aid in increasing the rate of satisfaction, with care decisions ultimately being made during direct consultations (*P*<.01).

XSL•FO
RenderX

Figure 12. Forest plot of the effect of telemedicine on patient satisfaction score [26,42,43].



Sensitivity Analysis

A total of 6 sensitivity analyses were conducted to explore possible sources of heterogeneity and assess the robustness of the results. The removal of 1 study [49] with a small sample explained part of the substantial heterogeneity in the healing rate (I^2=49%; *P*=.04) but did not alter the significance or direction of the effect (SMD 1.14, 95% CI 1.01-1.29; *P*=.04). The removal of 1 study [44] that used a different measurement explained the most substantial heterogeneity in healing sores (I^2=25%; *P*=.25), but the significance or direction of the effect remained unchanged (SMD −0.73, 95% CI −0.93 to −0.53; *P*<.001). The removal of 1 study [30] that delivered an intervention through a dedicated cloud-based software program also explained part of the substantial heterogeneity in healing time (I^2=82%; *P*=.001). However, the significance or direction of the effect did not change (SMD −0.64, 95% CI −1.08 to −0.19; *P*=.005). The removal of 1 study [27] that delivered the intervention through an app explained a small amount of substantial heterogeneity in the quality of life (I^2=96%; *P*<.001), and the significance and direction of the effect changed (SMD 0.69, 95% CI −0.32 to 1.71; *P*=.18). The removal of 1 study [40] with patients with an older average age explained the most substantial heterogeneity in the patient satisfaction rate (I^2=14%; *P*=.28), resulting in a change in the significance and direction of the effect (SMD 1.40, 95% CI 1.15-1.71; *P*=.001). Similarly, the removal of 1 study [42] with patients with an older average age also explained part of the substantial heterogeneity in patient satisfaction scores (I^2=69%; *P*=.07), leading to a change in the significance and direction of the effect (SMD 1.28, 95% CI 0.72-1.84; *P*<.001).

Publication Bias

The funnel plots (n>10) for the outcomes of the healing rate were asymmetrical (Egger test, bias=1.801; SE 0.367; *P*<.001), thus indicating the presence of a potential bias (see Multimedia Appendix 6 funnel plot for more details). The trim-and-fill test revealed 5 missing studies (see Multimedia Appendix 7 funnel plot for more details). The addition of the missing studies to the left part of the funnel plot changed the effect size (RR 1.06, 95% CI 0.98-1.15; *P*=.16). Moreover, no significant publication bias was found for the outcome of healing score, healing time, and mortality (n>5; Egger test, all *P*>.05).

Discussion

Main Findings

This review aimed to evaluate the effectiveness of telemedicine in chronic wound management. A total of 22 RCTs involving 2397 participants met the inclusion criteria. Overall, the use of telemedicine significantly improved the healing time, healing score, amputation rate, pain, and quality of life of patients with chronic wounds. Subgroup analyses revealed significant effects of short-term telemedicine, telemedicine for pressure injuries or venous ulcers, and telemedicine in conjunction with a face-to-face component on improving the healing rate. Subgroup analyses revealed a significant decrease in healing time with telemedicine for venous ulcers or telemedicine in conjunction with a face-to-face component. Subgroup analyses also revealed significant effects of telemedicine with ≥2 telemedicine types, telemedicine for diabetic foot ulcers, or telemedicine in combination with communication components in improving the amputation rate. In conclusion, telemedicine appears to be an effective method for chronic wound management.

Interpretation of Results

An important finding of this review is that telemedicine can significantly improve chronic wound healing, particularly in terms of reducing healing time (SMD −0.47, 95% CI −0.92 to −0.02; *P*=.04) and improving the healing score (SMD −1.46, 95% CI −2.27 to −0.66; *P*=.001). This effect is likely attributable to telemedicine's ability to provide real-time, efficient remote monitoring, diagnosis, and treatment, which ensures timely and professional medical intervention, thereby optimizing wound management and accelerating the healing process. However, these findings contrast with those of a previous study [56], in which telemedicine was shown to have no significant effect on healing time. One explanation for this result is that the previous meta-analysis included only 4 RCTs with 254 participants, whereas our result was based on 6 RCTs with 577 participants. A greater sample size would increase confidence in our findings. Furthermore, our study demonstrated that the healing score in the telemedicine group was significantly higher than in the control group. The healing score encompasses critical factors such as wound area, fluid volume, and tissue type, which are integral to the improvement of wound healing time.

Moreover, the evidence in this review suggested that the healing rate of chronic wounds in the telemedicine group was 1.16 times greater than that in the control group (RR 1.16, 95% CI 1.01-1.33; *P*=.03). This result contrasts with the outcome of a prior meta-analysis conducted by Chen et al [24] (RR 1.21, 95%CI 0.96-1.53; *P*=.11). However, potential publication bias was observed in our study (Egger test, bias=1.801; SE 0.367; *P*<.001). After imputing missing studies by using the trim-and-fill method, the recalculated pooled RR was changed (RR 1.06, 95% CI 0.98-1.15; *P*=.16). Therefore, this result needs to be interpreted with caution and should be verified by more studies in the future. Additionally, this review revealed that the impact of telemedicine on healing rates may differ according

XSL•FO
RenderX

to the intervention duration and the type of chronic wound. Patients with chronic wounds may benefit from the implementation of telemedicine in the short term, thus indicating that the effect of telemedicine on wound healing rates is more pronounced in the early stages. Concerning types of chronic wounds, combining telemedicine for pressure injuries or venous ulcers had a greater effect on the healing rate than telemedicine for diabetic foot ulcers or mixed chronic wounds. However, due to the relatively small number of studies included in these subgroups, these findings also should be interpreted with caution.

This review demonstrated that telemedicine significantly reduced the amputation rate compared to the control interventions, thus corroborating findings from another meta-analysis [24]. Telemedicine real-time remote monitoring and treatment technologies enable more accurate and timely interventions for chronic wounds, thus reducing the risk of amputation due to treatment delays or errors. Furthermore, the subgroup analysis suggested that ≥2 types of telemedicine had a greater effect on the amputation rate than 1 type of telemedicine. This enhanced efficacy likely results from the comprehensive and precise management strategies enabled by the integration of multiple technologies. Subgroup analysis indicated that combining telemedicine for diabetic foot ulcers had a greater effect on the amputation rate than telemedicine for mixed chronic wounds. This finding can be explained by the fact that patients with diabetes have a 15- to 20-fold greater risk of amputation than patients without diabetes [57-59]. Moreover, the subgroup analysis revealed that telemedicine including communication components had a greater effect on decreasing the amputation rate than telemedicine with no communication components. This is supported by a Cochrane review [60], which highlighted the benefits of interactive telemedicine in improving health care outcomes and professional practices.

Notably, this review indicates that, compared with telemedicine alone, combining telemedicine with face-to-face components significantly enhances healing rates and reduces healing time. Although telemedicine provides convenient and timely assessments, adding face-to-face interactions allows wound therapists to perform more comprehensive and detailed assessments. Direct observation and tactile examinations during these interactions offer additional insights into the wound's condition, the health of the surrounding tissues, and the patient's overall health, thus enabling more accurate treatment planning. Moreover, face-to-face interactions facilitate direct communication between patients and therapists, thus enhancing patients' understanding of and trust in the treatment plan, which can correspondingly improve compliance. These findings suggest that, in practical terms, a combination of remote and face-to-face management can be effective for chronic wound management.

Concerning the PROs in this review, telemedicine significantly decreased pain compared with the control group. Telemedicine overcomes geographical limitations and facilitates patients' access to continuous care from a professional medical team and personalized pain management solutions, thus effectively alleviating pain symptoms. Additionally, this review demonstrated evidence that telemedicine significantly improved the quality of life compared with that of the control group. Numerous studies [61,62] have illustrated the potential of telemedicine to increase patients' quality of life. Regrettably, the robustness of the results of this study was compromised in the sensitivity analysis. The removal of 1 study [27] that delivered interventions through an app explained a small amount of the substantial heterogeneity in the quality of life, and the significance and direction of the effect changed. Similarly, the findings regarding whether telemedicine can enhance patient satisfaction are also not robust. Hence, future research is needed to investigate the impacts of telemedicine on the quality of life and satisfaction of patients with chronic wounds.

Strengths and Limitations

This review has notable strengths. First, our conclusions were grounded in the best available evidence because we included only RCTs. Second, the study's robustness was enhanced by subgroup and sensitivity analyses. Third, we included both wound-related outcomes and PROs to provide a comprehensive assessment of patients with chronic wounds. However, several limitations warrant acknowledgment. First, the methodological quality of the included studies varied, with some exhibiting relatively low quality, which may have influenced the meta-analysis results. Second, publication bias may be present, although we refrained from using a funnel plot for most outcomes due to the insufficient number of studies. Nevertheless, we mitigated potential publication bias by using a comprehensive search strategy. Third, only a randomly selected 30% (n=22) of full-text papers underwent dual independent assessment. While this approach could potentially introduce subjective interpretation risk, several factors mitigate this concern: we implemented a rigorous protocol with clear criteria; the dual review of the randomly selected 30% (n=22) sample showed complete agreement between reviewers, suggesting consistent application of inclusion or exclusion criteria. Fourth, language bias may have occurred due to limitations in the literature retrieval based on language. Fifth, we were unable to draw definitive conclusions regarding the potential of telemedicine to reduce costs, depression, and anxiety in patients with chronic wounds due to inconsistent results, the inability to pool data for meta-analysis, and the limited number of available studies. Despite these uncertainties, the inconclusive findings regarding these outcomes offer crucial insights that can guide future research in selecting appropriate outcome indicators. We encourage further investigations to comprehensively explore the effectiveness of telemedicine in enhancing PROs in individuals with chronic wounds.

Conclusions

This review provides evidence supporting the significant enhancement of the healing time, healing score, amputation rate, pain, and quality of life of patients with chronic wounds through the application of telemedicine compared to the control group. Moreover, subgroup analyses revealed significant effects of short-term telemedicine, telemedicine for pressure injuries or venous ulcers, and telemedicine in conjunction with a face-to-face component on improving the healing rate. Subgroup analyses revealed a significant decrease in healing time with

XSL•FO
RenderX

telemedicine for venous ulcers or telemedicine in conjunction with a face-to-face component. Subgroup analyses also revealed significant effects of telemedicine with ≥2 telemedicine types, telemedicine for diabetic foot ulcers, or telemedicine in combination with communication components in improving the amputation rate. Hence, telemedicine appears to be an effective method for chronic wound management, with future studies needed to explore their impact on healing rate and PROs in patients with chronic wounds.

Acknowledgments

This study was funded by the Central High-Level Hospital Clinical Research Funding-Talent Cultivation Special (BJ-2023 - 133), the National High Level Hospital Clinical Research Funding (BJ-2024 - 198), and the Fundamental Research Funds for the Central Universities (3332024098). The authors confirm that no generative artificial intelligence tools were used in the writing or preparation of this manuscript.

Authors' Contributions

XZ and JL were responsible for conceptualization. XZ completed database searching. XZ, ZG, YD, CS, and JL completed paper screening, data extraction, and critical appraisal. XZ, ZG, YD, CS, and JL contributed to data interpretation. XZ, ZG, YD, CS, and JL wrote the original draft. JL provided advice. All authors contributed to reviewing and editing the final manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOC File, 78 KB - mhealth_v13i1e58553_app1.doc]

Multimedia Appendix 2
Characteristics of included studies.
[DOC File, 68 KB - mhealth_v13i1e58553_app2.doc]

Multimedia Appendix 3
Risk of bias of the included studies.
[DOC File, 34 KB - mhealth_v13i1e58553_app3.doc]

Multimedia Appendix 4
Grading of Recommendations, Assessment, Development, and Evaluation form of evidence certainty for telemedicine versus control group.
[DOC File, 51 KB - mhealth_v13i1e58553_app4.doc]

Multimedia Appendix 5
Subgroup analyses of effectiveness of telemedicine on wound-related and patient self-reported outcomes in patients with chronic wounds.
[DOC File, 171 KB - mhealth_v13i1e58553_app5.doc]

Multimedia Appendix 6
Funnel plot (effectiveness of telemedicine on healing rate).
[DOC File, 36 KB - mhealth_v13i1e58553_app6.doc]

Multimedia Appendix 7
Funnel plot after trim-and-fill method (effectiveness of telemedicine on healing rate).
[DOC File, 37 KB - mhealth_v13i1e58553_app7.doc]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) checklist.
[DOC File, 78 KB - mhealth_v13i1e58553_app8.doc]

XSL•FO
RenderX

References

1. Graves N, Phillips CJ, Harding K. A narrative review of the epidemiology and economics of chronic wounds. Br J Dermatol 2022 Aug;187(2):141-148. [doi: 10.1111/bjd.20692] [Medline: 34549421]
2. Han G, Ceilley R. Chronic wound healing: a review of current management and treatments. Adv Ther 2017 Mar;34(3):599-610. [doi: 10.1007/s12325-017-0478-y] [Medline: 28108895]
3. Martinengo L, Olsson M, Bajpai R, et al. Prevalence of chronic wounds in the general population: systematic review and meta-analysis of observational studies. Ann Epidemiol 2019 Jan;29:8-15. [doi: 10.1016/j.annepidem.2018.10.005] [Medline: 30497932]
4. Iversen AKS, Lichtenberg M, Fritz BG, et al. The chronic wound characterisation study and biobank: a study protocol for a prospective observational cohort investigation of bacterial community composition, inflammatory responses and wound-healing trajectories in non-healing wounds. BMJ Open 2024 Oct 17;14(10):e084081. [doi: 10.1136/bmjopen-2024-084081] [Medline: 39419618]
5. Olsson M, Järbrink K, Divakar U, et al. The humanistic and economic burden of chronic wounds: a systematic review. Wound Repair Regen 2019 Jan;27(1):114-125. [doi: 10.1111/wrr.12683] [Medline: 30362646]
6. Zhu X, Olsson MM, Bajpai R, Järbrink K, Tang WE, Car J. Health-related quality of life and chronic wound characteristics among patients with chronic wounds treated in primary care: a cross-sectional study in Singapore. Int Wound J 2022 Aug;19(5):1121-1132. [doi: 10.1111/iwj.13708] [Medline: 34854234]
7. Verdes M, Mace K, Margetts L, Cartmell S. Status and challenges of electrical stimulation use in chronic wound healing. Curr Opin Biotechnol 2022 Jun;75:102710. [doi: 10.1016/j.copbio.2022.102710] [Medline: 35398709]
8. Bowers S, Franco E. Chronic wounds: evaluation and management. Am Fam Physician 2020 Feb 1;101(3):159-166. [Medline: 32003952]
9. Yao Z, Niu J, Cheng B. Prevalence of chronic skin wounds and their risk factors in an inpatient hospital setting in Northern China. Adv Skin Wound Care 2020 Sep;33(9):1-10. [doi: 10.1097/01.ASW.0000694164.34068.82] [Medline: 32740027]
10. Wanwan P. Investigation and analysis on the present situation of nursing knowledge and training needs of nursing staff in primary hospital [Thesis]. : Yangtze University; 2021.
11. Lommi M, Raffaele B, Tolentino Diaz MY, Montini G, Puleio C, Porcelli B. Nursing outcomes in wound care management: a mixed method study. Nurs Open 2023 Apr;10(4):2249-2263. [doi: 10.1002/nop2.1477] [Medline: 36478413]
12. Consolidated telemedicine implementation guide. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240059184 [accessed 2024-02-12]
13. Clinical practice in a digital health environment. Registered Nurses' Association of Ontario. 2024. URL: https://rnao.ca/bpg/guidelines/clinical-practice-digital-health-environment [accessed 2024-08-12]
14. Notice of the State Council on issuing the 14th five-year plan for the development of the digital economy. The State Council. 2022. URL: https://www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm [accessed 2024-02-10]
15. National Nursing Development Plan (2021-2025). National Health Commission of China. 2022. URL: https://www.gov.cn/gongbao/content/2022/content_5705846.htm [accessed 2024-02-11]
16. Mercadal-Orfila G, Herrera-Pérez S, Piqué N, et al. Implementing systematic patient-reported measures for chronic conditions through the Naveta Value-Based Telemedicine Initiative: observational retrospective multicenter study. JMIR Mhealth Uhealth 2024 Jun 28;12:e56196. [doi: 10.2196/56196] [Medline: 38545697]
17. Dinh A, Yin AL, Estrin D, Greenwald P, Fortenko A. Augmented reality in real-time telemedicine and telementoring: scoping review. JMIR Mhealth Uhealth 2023 Apr 18;11:e45464. [doi: 10.2196/45464] [Medline: 37071458]
18. Bassi M, Strati MF, Parodi S, et al. Patient satisfaction of telemedicine in pediatric and young adult type 1 diabetes patients during Covid-19 pandemic. Front Public Health 2022;10:857561. [doi: 10.3389/fpubh.2022.857561]
19. Avila A, Claes J, Buys R, Azzawi M, Vanhees L, Cornelissen V. Home-based exercise with telemonitoring guidance in patients with coronary artery disease: does it improve long-term physical fitness? Eur J Prev Cardiol 2020 Mar;27(4):367-377. [doi: 10.1177/2047487319892201] [Medline: 31787026]
20. Sarpourian F, Ahmadi Marzaleh M, Fatemi Aghda SA, Zare Z. Application of telemedicine in the ambulance for stroke patients: a systematic review. Prehosp Disaster med 2023 Dec;38(6):774-779. [doi: 10.1017/S1049023X23006519] [Medline: 37877359]
21. Høyland SA, Holte KA, Islam K, et al. A cross-sector systematic review and synthesis of knowledge on telemedicine interventions in chronic wound management—implications from a system perspective. Int Wound J 2023 May;20(5):1712-1724. [doi: 10.1111/iwj.13986] [Medline: 36261052]
22. Zhang J, Mihai C, Tüshaus L, Scebba G, Distler O, Karlen W. Wound image quality from a mobile health tool for home-based chronic wound management with real-time quality feedback: randomized feasibility study. JMIR Mhealth Uhealth 2021 Jul 30;9(7):e26149. [doi: 10.2196/26149] [Medline: 34328440]
23. Dege T, Glatzel B, Borst V, et al. Patient-centered chronic wound care mobile apps: systematic identification, analysis, and assessment. JMIR Mhealth Uhealth 2024 Mar 21;12:e51592. [doi: 10.2196/51592] [Medline: 38533818]
24. Chen L, Cheng L, Gao W, Chen D, Wang C, Ran X. Telemedicine in chronic wound management: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jun 25;8(6):e15574. [doi: 10.2196/15574] [Medline: 32584259]

XSL•FO
RenderX

25. Xuemei Y, Mengmiao P, Haiyan L. The application impact of the internet + network nurse integrated service model in chronic wound nursing. J Clin Nurs Pract 2023 Jan;9(1):145-147. [doi: 10.11997/nitcwm.202301043]
26. Jing C, Shanshan J, Aiping W, Jinshan S. Application of 5A nursing model based on internet platform in the nursing of patients with 0 grade diabetic foot. J Qilu Nurs 2022;28(7):12-16. [doi: 10.3969/j.issn.1006-7256.2022.07.004]
27. Jingwen L, Xiaochun Y, Li Z. The utilization of the internet + hospital-community-family smart nursing model in home care for patients with diabetic foot enhances professionalism. Chin Gen Pract Nurs 2022 Dec;20(36):5174-5177. [doi: 10.12104/j.issn.1674-4748.2022.36.036]
28. Hua S, Ruijiao S, Yang Y, Hua B. Practice of internet information platform in home nursing service of difficult healing wounds. Clin Res 2022 Sep;30(10):185-188. [doi: 10.12385/j.issn.2096-1278(2022)10-0185-04]
29. Iversen MM, Igland J, Smith-Strøm H, et al. Effect of a telemedicine intervention for diabetes-related foot ulcers on health, well-being and quality of life: secondary outcomes from a cluster randomized controlled trial (DiaFOTo). BMC Endocr Disord 2020 Oct 21;20(1):157. [doi: 10.1186/s12902-020-00637-x] [Medline: 33087074]
30. Téot L, Geri C, Lano J, Cabrol M, Linet C, Mercier G. Complex wound healing outcomes for outpatients receiving care via telemedicine, home health, or wound clinic: a randomized controlled trial. Int J Low Extrem Wounds 2020 Jun;19(2):197-204. [doi: 10.1177/1534734619894485] [Medline: 31852312]
31. Jin X. Application effect of internet + collaborative nursing service model in nursing home chronic wounds of the elderly. Chin Modern Med 2021;28(21). [doi: 10.3969/j.issn.1674-4721.2021.21.073]
32. Haiyan Z, Haihong Q, Huaxian Z, Fengying R, Ying M, Jie Z. Application effect of nursing intervention based on telemedicine technology in patients with stress injury in community. Clin Med Res Pract 2021;6(6):120-122. [doi: 10.19347/j.cnki.2096-1413.202106043]
33. Squitieri L, Tsangaris E, Klassen AF, et al. Patient-reported experience measures are essential to improving quality of care for chronic wounds: an international qualitative study. Int Wound J 2020 Aug;17(4):1052-1061. [doi: 10.1111/iwj.13374] [Medline: 32320141]
34. Klassen AF, van Haren E, van Alphen TC, et al. International study to develop the WOUND-Q patient-reported outcome measure for all types of chronic wounds. Int Wound J 2021 Aug;18(4):487-509. [doi: 10.1111/iwj.13549] [Medline: 33694326]
35. Knobloch K, Yoon U, Vogt PM. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and publication bias. J Craniomaxillofac Surg 2011 Mar;39(2):91-92. [doi: 10.1016/j.jcms.2010.11.001] [Medline: 21145753]
36. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7:16. [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
37. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition: John Wiley & Sons; 2019.
38. Rhon DI, Fritz JM, Kerns RD, et al. TIDieR-telehealth: precision in reporting of telehealth interventions used in clinical trials—unique considerations for the Template for the Intervention Description and Replication (TIDieR) checklist. BMC Med Res Methodol 2022 Jun 2;22(1):161. [doi: 10.1186/s12874-022-01640-7] [Medline: 35655144]
39. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
40. Yumei L, Wenxian Z, Miaosang X, Aihua C, Chunmei Z. Application of internet + network contract service model in nursing care of chronic wound patients at home. Hospital Manage Forum 2020 May;37(5):109-112. [doi: 10.3969/j.issn.1671-9069.2020.05.035]
41. Beibei Y. Observation on effect of internet + collaborative care model in the care of patients with chronic wounds at county medical community. : Zhejiang University of Traditional Chinese Medicine; 2019. [doi: 10.27465/d.cnki.gzzyc.2019.000069]
42. Smith-Strøm H, Igland J, Østbye T, et al. The effect of telemedicine follow-up care on diabetes-related foot ulcers: a cluster-randomized controlled noninferiority trial. Diabetes Care 2018 Jan;41(1):96-103. [doi: 10.2337/dc17-1025] [Medline: 29187423]
43. Arora M, Harvey LA, Glinsky JV, et al. Telephone-based management of pressure ulcers in people with spinal cord injury in low- and middle-income countries: a randomised controlled trial. Spinal Cord 2017 Feb;55(2):141-147. [doi: 10.1038/sc.2016.163] [Medline: 27995939]
44. Xiaoping L, Lihong W, Tong R, Shihua L, Hongying Z. Application effect of "Hospital - community" Joint Pressure Sore Care Group in stroke patients with pressure sores on continuity of care after discharged. Chin Gen Pract Med 2016 Jun;19(16):1964-1968. [doi: 10.3969/j.issn.1007-9572.2016.16.025]
45. Rasmussen BSB, Froekjaer J, Bjerregaard MR, et al. A randomized controlled trial comparing telemedical and standard outpatient monitoring of diabetic foot ulcers. Diabetes Care 2015 Sep;38(9):1723-1729. [doi: 10.2337/dc15-0332] [Medline: 26116717]
46. Huanli Z. Evaluation of the application of remote nursing consultation in community pressure ulcer nursing. Jilin Med J 2014 Sep;35(26):5918-5919. [doi: 10.3969/j.issn.1004-0412.2014.26.140]
47. Weiwei D. Application of continuity nursing in patients with leg venous ulcer. : Central South University; 2014.

XSL•FO
RenderX

48. Stern A, Mitsakakis N, Paulden M, et al. Pressure ulcer multidisciplinary teams via telemedicine: a pragmatic cluster randomized stepped wedge trial in long term care. BMC Health Serv Res 2014 Feb 24;14:83. [doi: 10.1186/1472-6963-14-83] [Medline: 24559218]
49. Vowden K, Vowden P. A pilot study on the potential of remote support to enhance wound care for nursing-home patients. J Wound Care 2013 Sep;22(9):481-488. [doi: 10.12968/jowc.2013.22.9.481] [Medline: 24005782]
50. Ting C, Zhengying C, Dongqiong W, Xihan L. Application of tele-wound treatment platform to patients with chronic lower limb ulcers in remote areas. J Nurs Sci 2012 Dec;27(24):22-24. [doi: 10.7748/ns2012.11.27.10.24.p9841]
51. Terry M, Halstead LS, O'Hare P, et al. Feasibility study of home care wound management using telemedicine. Adv Skin Wound Care 2009;22(8):358-364. [doi: 10.1097/01.ASW.0000358638.38161.6b]
52. Dobke MK, Bhavsar D, Gosman A, De Neve J, De Neve B. Pilot trial of telemedicine as a decision aid for patients with chronic wounds. Telemed J E Health 2008 Apr;14(3):245-249. [doi: 10.1089/tmj.2007.0038] [Medline: 18570547]
53. Santamaria N, Carville K, Ellis I, Prentice J. The effectiveness of digital imaging and remote expert wound consultation on healing rates in chronic lower leg ulcers in the Kimberley region of Western Australia. Primary Intention 2004 May;2(12):62-70 [FREE Full text]
54. Fasterholdt I, Gerstrøm M, Rasmussen BSB, Yderstræde KB, Kidholm K, Pedersen KM. Cost-effectiveness of telemonitoring of diabetic foot ulcer patients. Health Informatics J 2018 Sep;24(3):245-258. [doi: 10.1177/1460458216663026]
55. Arora M, Harvey LA, Glinsky JV, et al. Cost-effectiveness analysis of telephone-based support for the management of pressure ulcers in people with spinal cord injury in India and Bangladesh. Spinal Cord 2017 Dec;55(12):1071-1078. [doi: 10.1038/sc.2017.87]
56. Huang Z, Wu S, Yu T, Hu A. Efficacy of telemedicine for patients with chronic wounds: a meta-analysis of randomized controlled trials. Adv Wound Care 2021 Feb 1;10(2):103-112. [doi: 10.1089/wound.2020.1169]
57. Pscherer S, Dippel FW, Lauterbach S, Kostev K. Amputation rate and risk factors in type 2 patients with diabetic foot syndrome under real-life conditions in Germany. Prim Care Diabetes 2012 Oct;6(3):241-246. [doi: 10.1016/j.pcd.2012.02.004]
58. Crawford F, Mccowan C, Dimitrov BD, et al. The risk of foot ulceration in people with diabetes screened in community settings: findings from a cohort study. QJM 2011 May;104(5):403-410. [doi: 10.1093/qjmed/hcq227]
59. Bild DE, Selby JV, Sinnock P, Browner WS, Braveman P, Showstack JA. Lower-extremity amputation in people with diabetes: epidemiology and prevention. Diabetes Care 1989 Jan 1;12(1):24-31. [doi: 10.2337/diacare.12.1.24]
60. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 7;2015(9):CD002098. [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
61. Zhang QL, Lei YQ, Liu JF, Cao H, Chen Q. Using telemedicine to improve the quality of life of parents of infants with CHD surgery after discharge. Int J Qual Health Care 2021 Sep 25;33(3):mzab133. [doi: 10.1093/intqhc/mzab133] [Medline: 34490460]
62. Shih YH, Wang JY, Chou PH, Lin KH. The effects of treatment via telemedicine interventions for patients with depression on depressive symptoms and quality of life: a systematic review and meta-ranalysis. Ann Med 2023 Dec;55(1):1092-1101. [doi: 10.1080/07853890.2023.2187078] [Medline: 36920229]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
PRO: patient-reported outcome
RCT: randomized controlled trial
RR: risk ratio
SMD: standard mean difference
TIDieR: Template for the Intervention Description and Replication

Edited by L Buis, TDA Cardoso; submitted 18.03.24; peer-reviewed by A Ortega, JI Alvarado-Sanchez, M Afshari; revised version received 09.04.25; accepted 14.04.25; published 10.06.25.

Please cite as:
Zhang X, Guo Z, Duan Y, Sun C, Luo J
Effectiveness of Telemedicine on Wound-Related and Patient-Reported Outcomes in Patients With Chronic Wounds: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2025;13:e58553
URL: https://mhealth.jmir.org/2025/1/e58553
doi:10.2196/58553

XSL•FO
RenderX

© Xiaoyan Zhang, Zhanghui Guo, Yu Duan, Chao Sun, Jiayin Luo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Natural Language Processing Chatbot–Based Interventions for Improvement of Diet, Physical Activity, and Tobacco Smoking Behaviors: Systematic Review

Jing Chen[1*], PhD; Run-Ze Hu[2*]; Yu-Xuan Zhuang[2]; Jia-Qi Zhang[2]; Rui Shan[1], MD; Yang Yang[1], MD; Zheng Liu[1,2], PhD

[1]Department of Maternal and Child Health, School of Public Health, Peking University, No. 38, Xueyuan Road, Haidian District, Beijing, China

[2]Department of Preventive Medicine, School of Public Health, Peking University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zheng Liu, PhD
Department of Maternal and Child Health, School of Public Health, Peking University, No. 38, Xueyuan Road, Haidian District, Beijing, China

Abstract

Background: The rapid development of artificial intelligence technology has enabled chatbots to increasingly promote health-related behaviors, addressing the high demand for human resources in traditional interventions. Several systematic reviews have been conducted in this area. However, the existing reviews have not focused on the rigorously designed randomized trials of the state-of-the-art chatbots (interacting with users through unconstrained natural language), thus calling for an updated review.

Objective: We aimed to explore the effects of natural language processing (NLP) chatbot–based interventions on improving diet, physical activity, and tobacco smoking behaviors in the general population and to evaluate the chatbot use behaviors during the implementation process.

Methods: We comprehensively searched 12 databases or registers for eligible studies published from January 1, 2010, until July 16, 2024, and obtained a total of 6301 studies. We included randomized controlled trials (RCTs) that used NLP-chatbots to promote diet, physical activity, or tobacco smoking behaviors among adults or children. Due to considerable heterogeneity across the included studies, we adopted the synthesis without meta-analysis guidelines and summarized the effectiveness of NLP chatbot–based interventions. We used the new evidence-mapping method (bubble plot) to visualize the results. We also described the results related to the changes in diet, physical activity, or tobacco smoking behaviors (eg, change of BMI and stage of change). To evaluate the implementation process of the intervention, we summarized users' interaction with NLP-chatbots and their feelings (eg, satisfaction) about NLP-chatbot use. Additionally, we assessed the risk of bias of studies using the RoB 2.0 (Risk of Bias; The Cochrane Collaboration) tools.

Results: We finally included 7 RCTs. Concerning dietary and physical activity behaviors, the effectiveness of NLP chatbot–based interventions was inconsistent among adults, while no evidence of effect was observed among children. Concerning tobacco smoking behaviors, the included studies showed consistent evidence of improving this behavior among adults. Regarding the risk of bias of the changes in diet, physical activity, and tobacco smoking behaviors, 2 of 3, 2 of 4, and 1 of 2 studies had a high risk of bias, respectively, while the remaining had a low risk of bias. Concerning the interactions with NLP-chatbots, studies showed an overall high percentage of general interaction between users and NLP-chatbots, but not a satisfactorily high percentage of interactions specific to health behaviors. Concerning feelings about NLP-chatbot use, users showed a positive impression of NLP-chatbot use, feeling it was useful, credible, and financially feasible.

Conclusions: NLP chatbot–based interventions were beneficial for adults' tobacco smoking behaviors, but no such evidence was found on diet or physical activity behaviors among adults or children. More RCTs with larger samples and lower risk of bias are urgently needed to enhance our findings in the future.

(JMIR Mhealth Uhealth 2025;13:e66403) doi:10.2196/66403

KEYWORDS

NLP-chatbot; intervention; diet; physical activity; tobacco smoking; RCTs; natural language processing; randomized controlled trial

XSL•FO
RenderX

Introduction

Worldwide, physical inactivity, unhealthy diet, and tobacco use are the 3 major behavioral risk factors responsible for chronic, noncommunicable diseases [1]. The prevalence of these 3 factors was high in the population. Concerning physical inactivity, a significant proportion of populations fail to meet recommended activity levels [2]. Specifically, 81% of school-aged children (11 - 17 y) did not achieve the minimum requirement of 60 minutes of moderate-to-vigorous physical activity daily, while 27.5% of adults fell short of the weekly recommendation of either 150 minutes of moderate-intensity activity, 75 minutes of vigorous-intensity activity, or an equivalent combination of both [2,3]. Concerning unhealthy diets, the vegetable supply was insufficient to meet the recommendations in 61% of the countries [4]. Specifically, most African and South American populations, as well as a part of Asian and North American populations, did not have sufficient (200 - 250 g per day) vegetable intake [5]. Additionally, most African and parts of Asian populations did not have sufficient (200 g per day) fruit intake [6]. Concerning tobacco smoking, the global smoking prevalence rate among people aged 15 years and older was 16.7% in 2022, despite significant past achievements in tobacco smoking control [7]. These high prevalences also posed challenges to global actions for better human health. Physical inactivity and unhealthy diet have paralleled the rising prevalence of overweight or obesity (BMI≥25 kg/m²) from 38% in 2020 to 46% in 2030 according to the published trends from 1975 to 2016 [8], making it challenging to reverse the obesity epidemic and achieve the World Health Organization's target of "no increase on obesity levels" by 2025 (based on 2010 levels). Tobacco smoking remains the most prevalent form of tobacco use. The persistently high tobacco smoking prevalence poses a significant challenge to achieving the World Health Organization's target of a 30% reduction in tobacco use from 2010 to 2025 [9].

To achieve a higher level of human health, actions are urgently needed to reverse these unhealthy behavioral risk factors. However, most conventional behavioral interventions cannot be faithfully implemented at the population scale due to the high demand for human resources. Thanks to the rapid development of artificial intelligence (AI) technology, chatbots have become a viable alternative to delivering resource-intensive, conventional behavioral interventions. Chatbots, also called intelligent dialogue systems or conversational agents, are machine agents that are designed to converse with humans using natural language through text or voice interactions, which can be classified into constrained and unconstrained ones. The former refers to those that only interact with the user through selection questions with fixed predefined options, while the latter can engage in free human-like dialogue with users and interact with users through unconstrained natural language. Notably, the recent rapid development of natural language processing (NLP) has led to obvious advancements in the capabilities from constrained ones to unconstrained ones. Therefore, it is urgent to summarize the development potential of chatbots based on NLP in improving individuals' health behavior.

There is evidence that chatbots have great potential to persuade, support, and promote individuals to change health-related behaviors and can be used to improve diet, physical activity, and tobacco smoking behaviors [10-12]. Previous systematic reviews involving qualitative summary or quantitative description through meta-analysis showed that chatbot-based interventions can improve physical activity, increase fruit and vegetable consumption, and enhance the individuals' intention to quit smoking [13-15]. However, few previous reviews have strictly distinguished between constrained and unconstrained chatbots [13,16-20]. Besides, concerning study design, reviews have included both randomized and nonrandomized studies, resulting in a variable strength of evidence [14,21,22]. The existing reviews were also limited in the search strategies and risk of bias assessment [23].

Bearing these research gaps in mind, we aimed to systematically review the existing randomized trials of the present topic and rigorously evaluate the evidence quality to (1) explore the effects of NLP chatbot–based interventions on improving diet, physical activity, and tobacco smoking behaviors in the general population; and (2) evaluate the NLP-chatbot use behaviors during the implementation process. Findings from this study would pave the way for the improvement of the NLP chatbot–based intervention for health behavior change.

Methods

Study Design

We conducted the systematic review following the guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [24]. The protocol was registered in the PROSPERO (International Prospective Register of Systematic Reviews) on December 21, 2023 (CRD42023492013).

Search Strategy

We conducted an extensive electronic search to identify all randomized controlled trials (RCTs) that reported outcomes measured by changes in behaviors in diet, physical activity, or tobacco smoking after the NLP chatbot–based intervention.

Eleven databases (PubMed, Embase, ACM Digital Library, Web of Science Core Collection, PsycINFO, IEEE, CINAHL Plus with Full Text, Cochrane Library, Scopus, Information Science & Technology Abstracts, and China National Knowledge Infrastructure) and one register (ClinicalTrials.gov) were systematically searched until July 16, 2024. The search strategy used a combination of keywords and Medical Subject Headings terms from the following 5 aspects: chatbot (eg, chatbot, AI agent, or conversational agent), diet (eg, diet or nutrition), physical activity (eg, physical activity, exercise, or sedentary behavior), tobacco smoking (eg, smoking cessation or smoking habit), and lifestyle (eg, weight control or lifestyle). The Boolean operator "OR" was used in each aspect, and between the last 4 aspects, and "AND" was used to combine "chatbot" with the other 4 aspects. The detailed search strategy for each database is reported in Multimedia Appendix 1. There were no restrictions on publication status or language. Due to the rapid development of conversational agents in recent

XSL•FO
RenderX

decades, we limited the included studies to those published in or after 2010.

Study Selection

Table 1 presents a summary of the inclusion and exclusion criteria of the study characteristics based on the PICOS framework (eg, populations or participants, interventions and comparators, outcomes, and study designs or types). Two reviewers (JC and YXZ) independently conducted a 2-stage study selection process: initially screening titles and abstracts, followed by a thorough examination of the full texts in the second stage. Discrepancies were resolved through consultation with a third review author (ZL).

Table . Summary of inclusion and exclusion criteria.

PICOS[a]	Inclusion criteria	Exclusion criteria
Population	• General population	None
Interventions	• NLP[b] chatbot–based interventions that enable natural language conversations with users for improving diet, physical activity, or tobacco smoking behaviors	Chatbots that cannot understand or generate natural language conversations
Comparators	• Active controls[c] or no-intervention controls[d]	None
Outcomes	• Primary outcomes[e]: changes in behaviors in diet, physical activity, or tobacco smoking • Secondary outcomes: the following 2 aspects are included: • Effectiveness-related outcomes: change of obesity-related outcomes (eg, BMI and weight), stage of change of diet, physical activity or tobacco smoking behavior (eg, precontemplation, contemplation, action, or maintenance), or change of mediators[f] (eg, motivation) • Implementation of NLP chatbot–based interventions: chatbot use behaviors (eg, duration, frequency, or interaction), facilitators and barriers to chatbot use (eg, satisfaction, appreciation, or feelings or acceptability), and safety issues[g] (eg, adverse events)	Studies that report only chatbot infrastructure or algorithm designs, or that did not report any of the primary outcomes
Study designs	• Individual-randomized or cluster-randomized parallel-group controlled trials	Observational studies, nonrandomized controlled studies

[a]PICOS: Population, Interventions, Comparators, Outcomes, and Study Designs.

[b]NLP: natural language processing.

[c]Active controls refer to providing participants with intervention measures other than the chatbot.

[d]No-intervention group includes those groups that have not implemented any special intervention measures or the wait-list control groups that have not yet received intervention in this study.

[e]Primary outcomes are measured by various means. For example, daily physical activity was recorded through a portable accelerometer, the intake of various types of food was investigated through a questionnaire, and the concentration of carbon monoxide in exhaled gas was measured to detect smoking behavior.

[f]Potentially mediate the effect of natural language processing–chatbot-based interventions for the improvement of diet, physical activity, and tobacco smoking behavior.

[g]Safety issues refer to unintended adverse events and the privacy protection of participants.

Data Extraction

Information extracted from the studies included participant characteristics (mean age, sex, country, race, income, education, baseline BMI score, frequency of smoking, stage of behavior change, recruitment setting, or sample size), study design, NLP-chatbot (referred to as the “chatbot” in the following sections) characteristics (theoretical framework, media or technology, dialogue initiative, input modality, output modality, or task-oriented), intervention measures (chatbot only or not, duration, or length and frequency), and results (primary outcomes and second outcomes) along with barriers and facilitators to the use of chatbot. Two researchers (YXZ and RZH) independently extracted data from the included studies and consulted with the other 2 researchers (JC and ZL) when discrepancies arose.

XSL•FO
RenderX

Data Synthesis

The primary outcome was the difference in changes in physical activity, diet, and tobacco smoking from baseline to follow-up between the intervention (NLP-chatbot) and control groups. Due to heterogeneity in outcome measurement and varied indicators across studies (eg, daily active time, sedentary behavior time, or average number of daily steps), we refrained from conducting a meta-analysis. Following the Cochrane Handbook (Chapter 12: synthesizing and presenting findings using other methods) [25], we qualitatively summarized the effectiveness of NLP chatbot–based interventions instead. We evaluated the intervention effectiveness estimates for each study individually. Categories of intervention effectiveness were differentiated into 4 groups based on the proportion of effective indicators in the primary outcome: (1) if all indicators were effective, it was considered effective (E); (2) if the effective ratio exceeded 50%, it was considered to be probably effective (PE); (3) if the effective ratio was less than 50%, it was considered not to be probably effective (PNE); (4) if the effective ratio was 0, it was considered not effective (NE). An indicator was considered effective when there was a statistically significant difference (P value <.05) in prepost changes between the intervention and control groups.

The following are the specific methods for judging whether the primary outcome indicators were effective. For each primary outcome, when studies reported P values for between-group comparisons (intervention vs control), these P values were prioritized in our analysis. For studies that did not report between-group P values directly, we calculated them using independent samples t-tests. We used the original data provided in the papers, including sample sizes, prepost mean differences, and SE for both intervention and control groups. When change score data were unavailable, between-group differences in postintervention values were analyzed instead, with the assumption of baseline equivalence justified by randomized study design. For studies reporting both baseline data and results at multiple follow-up time points, measurements across all follow-up time points were integrated using an equally weighted average method. The composite result was then compared against the baseline data for analysis. For the 3-arm intervention studies included, the intervention groups were both NLP chatbot–based interventions with slight differences in form. Therefore, we integrated the 2 intervention groups into 1 group and compared it with the control group. For the studies based on families, we considered adults and children separately. We examined whether each primary outcome indicator was effective and then summarized the possible effectiveness of the intervention for adults and children, respectively.

To present the results intuitively, we adopted a new evidence-mapping method to summarize the findings [26,27]. Specifically, we used bubble plots to visualize the whole patterns of results, considering the following factors in an intergrade manner: the outcomes (diet, physical activity, or smoking behaviors), sample size, population (adults and children, adults only, or children only), effectiveness of intervention, and risk of bias of studies.

For our secondary outcomes, we provided a descriptive summary of the effectiveness-related outcomes and the implementation of the NLP chatbot–based intervention. We used appropriate metrics, such as proportion, to quantify the frequency and duration of chatbot use and the number of interaction turns per dialogue in the included studies. We also focused on the acceptability of chatbots and therefore summarized the feelings of participants in the corresponding studies using them.

Assessment of the Outcome Quality and Evidence Certainty

Risk of bias assessment was performed exclusively for primary outcomes. For individual RCTs, we used the RoB 2.0 (Risk of Bias) tool [28], while for cluster RCTs, we used the RoB 2.0 tool for cluster-randomized trials [29]. The assessment for outcomes from individual RCTs contains the following five domains: (1) bias in the randomization process; (2) bias in deviation from intended interventions; (3) bias in missing outcome data; (4) bias in outcome measurement; and (5) bias in the selection of the reported result. Each domain was rated as having a high, low, or some concerns about the risk of bias. The assessment for outcomes from cluster RCTs was mostly consistent with that for individual RCTs, except for the first domain specified into (1) bias in the randomization process and (2) bias arising from the timing of identification or recruitment of participants.

Based on the biases identified in the aforementioned 5 aspects, we further assessed the overall bias risk of each primary outcome and rated them as: low risk of bias, some concerns, or high risk of bias. We then generated a bias risk assessment plot to present these results. Two researchers (YXZ and RZH) independently conducted the assessment of the included studies' bias, while a third author (JC) facilitated discussions to achieve consensus on discrepancies.

Ethical Considerations

Human subject ethics review approvals or exemptions: our study was reviewed and approved by the Peking University Institutional Review Board (IRB00001052-22091).

Informed consent: the original studies included in this review have obtained informed consent. Therefore, informed consent does not need to be obtained again for this review (secondary analysis).

Privacy and confidentiality: the data used in this study had been anonymized.

Compensation details: not applicable.

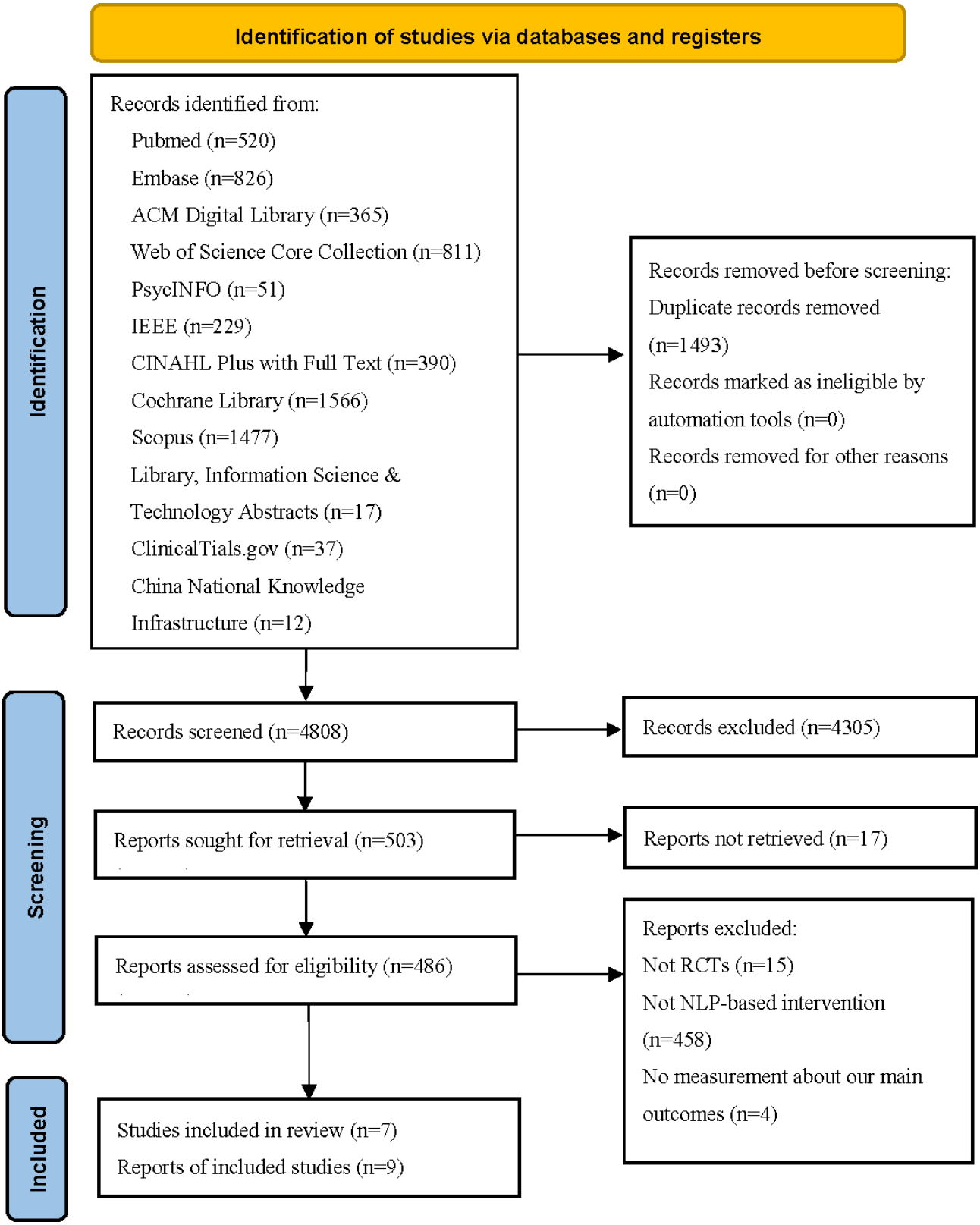
Results

Literature Search

Figure 1 shows the flow of study selection. Our search yielded 4808 records after excluding duplicates. After screening the titles and abstracts, we assessed the full-text articles of the remaining 486 records. A total of 7 studies were finally included.

XSL•FO
RenderX

Figure 1. Flowchart of the study selection process. NLP: natural language processing; RCT: randomized controlled trial.



Characteristics of Included Studies

Table 2 shows the primary characteristics of the 7 included studies, which were conducted in America (n=2) [30,31], Northern Ireland (n=1) [32], Spain (n=1) [33], Dutch (n=1) [34], China (n=1) [35], and Saudi Arabia (n=1) [36] from 2013 to 2022. Two studies focused only on physical activity [31,34], 1 study focused only on diet [36], 2 studies focused only on smoking [33,35], and 2 studies focused on both physical activity and diet [30,32]. In 5 studies, the participants were only adults [31,33-36], while the other 2 studies, which were family-based, contained both adults and children [30,32].

XSL•FO
RenderX

Table . Characteristics of participants among the 7 included studies.

ID	Study	Mean age (SD) years or range	Sex (female), (%)	Country, % (n/N)	BMI	Tobacco use	Baseline sample size	Attrition and rate	Follow-up sample size
1	Wright et al (2013) [30]	Children 10.3 (1.1), parents 40 (9.1)	Children 42, parents 96	USA (100)	Children 25.7 (2.1), parents 34 (6.7)	—[a]	Families n=50, children n=50, parents n=50	Attrition n=7, rate=14%	Families n=43, children n=43, parents n=43
2	Hassoon et al (2021) [31]	62.1 (9.8)	90	USA (100)	32.9 (5.0)	—	Adults n=42	Attrition n=0, rate =0%	Adults n=42
3	Carlin et al (2021) [32]	Phase 1: adults 40.5 (5.4), children 9.1 (2.0); phase 2: adults 38.9 (5.2), children 7.9 (2.0)	Phase 1: adults 10 (91), children 9 (56); phase 2: adults 11 (73), children 8 (44)	Western Trust area of Northern Ireland (100)	Phase 1: children —, adults 35.0 (6.4); phase 2: children—, adults 29.1 (4.9)	—	Phase 1: families n=11, parents n=11, children n=16; phase 2: families n=15, parents n=15, children n=18	Phase 1: attrition n=3, rate =27.3%; phase 2: attrition n=0, rate =0	Phase 1: families n=8, parents and children —; phase 2: families n=15, parents n=15, children n=18
4	Olano-Espinosa et al (2022) [33]	49.8 (10.82)	59.30	93.8 (481/513) were Spanish	—	10.1% (52/513) of patients reported moderate or high dependence on nicotine with Heavy Smoking Index values of 4 - 6 points and average consumption of 16.5 cigarettes/day (SD 7.75).	Adults n=513	Attrition n=281, rate =54.8%	Adults n=232
5	Friederichs et al (2014) [34]	Baseline 42.9 (14.5), follow up 45.3 (14.2)	Female: baseline 60.4 (578/958), follow-up 57.8 (289/500)	Dutch (100)	—	—	Adults n=958; attrition n=458, rate =47.8%	n=958; attrition n=458, rate =47.8%	Adults n=500
6	Wang et al (2018) [35]	Intervention group 32.8, control group 33.1	40.4	China (100)	—	—	Adults n=401	Attrition n=114, rate =28.4%	Adults n=287
7	Alghamdi and Alnanih (2021) [36]	—	—	Saudi Arabia (100)	—	—	n=60	Attrition n=0, rate =0%	n=60

[a]Not applicable.

As shown in Tables 3 and 4, among the included studies, 5 were 2-arm individual RCTs [30,32,33,35,36], while 2 used a 3-arm design [31,34]. The control groups of the 5 two-arm studies varied, including 4 no-intervention groups and 1 active control group. In the active control group, the participants received information passively, and the process did not involve chatbots or interactions with other participants. In both 3-arm RCTs, 2 groups received distinct NLP-chatbot interventions, while the third served as a control. Specifically speaking, in 1 study [34], 1 intervention group used a motivational interviewing chatbot with an avatar for web-based physical activity guidance, while the other used a simpler avatar-based chatbot; the control group received no intervention. Another study [31] compared 2 AI coaching methods: voice-assisted delivery via a smart speaker (MyCoach) and text-based delivery (SmartText), with the active control group receiving standard cancer education materials.

XSL•FO
RenderX

Table . Characteristics of the study designs of the 7 included studies.

ID	Study	Arm	Method of randomized parallel controlled trials	Classification of chatbot				
				Media or platform of technology	Dialogue initiative (user, system, or mixed)	Input modality	Output modality	Task-oriented (yes or no)
1	Wright et al (2013) [30]	2	Cluster-randomized	Telephone	Mixed	Spoken	Spoken	Yes
2	Hassoon et al (2021) [31]	3	Individual-randomized	Smart speaker	Mixed	Spoken	Spoken	Yes
3	Carlin et al (2021) [32]	2	Cluster-randomized	A smart speaker (Echo Dot)	Mixed	Spoken	Spoken	Yes
4	Olano-Espinosa et al (2022) [33]	2	Individual-randomized	Telegram, a widely used messaging app	Mixed (bidirectional)	Written	Written	Yes
5	Friederichs et al (2014) [34]	3	Individual-randomized	Website	N/A[a]	Written	Written	Yes
6	Wang et al (2018) [35]	2	Individual-randomized	Software (combined with WeChat)	Mixed	Written	Written	Yes
7	Alghamdi and Alnanih (2021) [36]	2	Individual-randomized	App (WhatsApp, social network)	N/A	Written	Written	Yes

[a]N/A: not applicable.

XSL•FO
RenderX

Table . Characteristics of the interventions of the 7 included studies.

ID	Study	Intervention type	Length, frequency	Measures (brief)			Chatbot only in intervention measures
				Intervention group		Control group	
1	Wright et al (2013) [30]	PA[a] and diet	12 wk, twice a week	HEAT[b] (telephone calls twice a week delivered by an automated IVR[c] system)	—[d]	No intervention (wait-list control)	Yes
2	Hassoon et al (2021) [31]	PA	4 wk, —[d]	Voice-assisted AI[e] coaching delivered by smart speaker (MyCoach)	Autonomous AI coaching delivered by text (SmartText)	Received written information	Yes
3	Carlin et al (2021) [32]	PA and diet	12 wk, —	Receive an intelligent personal assistant	—	Continue as usual	No (phase 1: SWEET[f] program)
4	Olano-Espinosa et al (2022) [33]	Smoking behavior	6 mo, —	Chatbot: Dejal@bot	—	Usual clinical practice	Yes
5	Friederichs et al (2014) [34]	PA	—, 1-time	A web-based PA intervention based on MI[g] with an avatar (AVATAR)	A content-identical intervention without an avatar (TEXT)	No intervention	Yes
6	Wang et al (2018) [35]	Smoking behavior	2 mo, —	Conversational agents (in a WeChat group talk with each other and conversational agent server by announcements, sharing, reminders, and responses)	—	Active control (in a WeChat group, but only received smoking cessation information and tips without social support or interactions with other participants)	No
7	Alghamdi and Alnanih (2021) [36]	Diet	90 d, —	Proposed chatbot	—	No intervention	Yes

[a]PA: physical activity.

[b]HEAT: healthy eating and activity today.

[c]IVR: interactive voice response.

[d]Not applicable.

[e]AI: artificial intelligence.

[f]SWEET: safe wellbeing eating and exercise together.

[g]MI: motivational interviewing.

More than half of the studies (n=4) used chatbots with written input modality [33-36], while the remaining 3 studies used those with spoken input modality [30-32]. All 7 studies used chatbots only using simplex output modality, spoken or written. Three studies mentioned the adoption of theoretical frameworks including “Motivation Interview,” “Social Cognitive Theory,” and “The Chronic-Disease Extended Model” [30,34,36]. The duration of interventions ranged from 1 time to 6 months.

Effectiveness of NLP Chatbot–Based Interventions on the Changes of Behaviors in Diet, Physical Activity, or Tobacco Smoking

Concerning diet behaviors, a study (n=54) conducted only among adults indicated that NLP chatbot–based intervention was effective [36]. Two studies (n=43 and n=22, respectively) conducted in family units suggested that NLP chatbot–based intervention did not effect children’s diet behaviors, while the results were probably not effective and not effective respectively for adults’ diet behaviors [30,32].

XSL•FO
RenderX

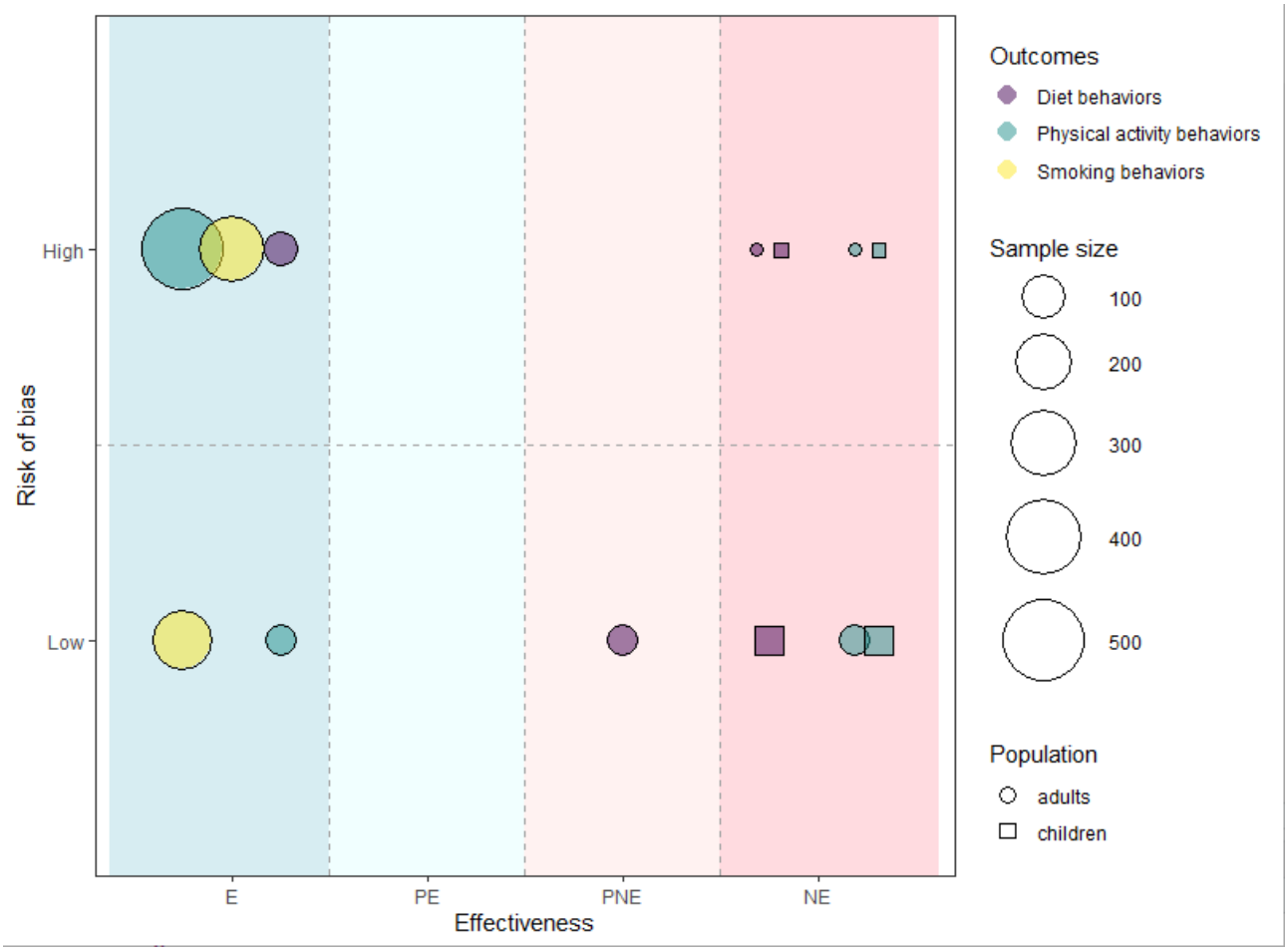
Concerning physical activity, both studies (n=500 and n=42, respectively) showed positive results from the intervention [31,34]. Specifically, NLP chatbot–based intervention groups showed an increased average number of daily steps or number of weekly days with at least 30 minutes of moderate physical activity and daily steps than the control groups. However, 2 other small-sample studies (n=43 and n=22, respectively) conducted on both adults and children did not show evidence of intervention effectiveness [30,32].

Concerning smoking behavior, 2 studies (n=232 and n=287, respectively) showed that NLP chatbot–based intervention can improve smoking behavior in adults [33,35].

Please see the specific assessment of results in 3 primary outcomes, as well as secondary outcomes of 7 studies in Table S1 in Multimedia Appendix 2.

Figure 2 shows the bubble plot displaying the effectiveness of NLP chatbot–based interventions on the changes in diet, physical activity, and smoking behaviors, with the bubble color, size, and shape representing the outcomes, sample size, and population, respectively. To aid in understanding, we have included an example of a bubble chart for reference. The pale green bubble in the upper left corner of Figure 2 corresponds to a study involving 500 adults. This study falls into the final effectiveness category of E (effective) due to its finding that NLP-chatbots had impacts on the changes in all of the physical activity behaviors. However, it is important to note that the outcome of this study indicated a high risk of bias.

Figure 2. Bubble chart of the effectiveness of NLP chatbot–based interventions. E: effective; NE: not effective; NLP: natural language processing; PE: possibly effective; PNE: possibly not effective.



Effectiveness of NLP Chatbot–Based Interventions on the Effectiveness-Related (Secondary) Outcomes

One study [33] compared the quality of life between the NLP chatbot–based intervention group and the control group at baseline and follow-up, and no differences were observed at either time.

Another [30] study compared changes in BMI (for adults and children), BMI percentile (for children), and BMI *z* scores (for children) between the NLP chatbot–based intervention group and the control group from baseline to follow-up, but no significant differences were observed.

Additionally, a study [36] assessed participants' perceived difficulty in adhering to the dietary treatment plan. At baseline, 16.67% of the intervention group and 20% of the control group reported challenges in committing to the dietary treatment plan. After the NLP chatbot–based intervention, a significantly higher proportion of the intervention group (46.67%) reported no adherence difficulties, whereas only 6.67% of the control group reported no committing difficulties. The study further proposed

XSL•FO
RenderX

a four-stage model of patient adaptation to chronic diseases: (1) be conscious of the need, (2) be ready to deal with the disease, (3) feel confident in dealing with the disease, and (4) stick to the plan. According to this change phase, this study investigated the proportion of different change stages among the participants in the NPL-chatbot intervention group and the control group. It was found that in the intervention group, the proportion of participants in phase 1 before the intervention was 30%, and there was no one in phase 1 after the intervention, while the proportion of those in phase 4 was 43.33%. In the control group, the proportion of participants in phase 1 before the intervention was 30%, and there were 26.67% of participants in phase 1 after the intervention, while no one was in phase 4.

Implementation of NLP Chatbot–Based Interventions

For this review, 4 included studies assessed the implementation of NLP chatbot–based interventions. A study conducted a 2-phase family-based intervention including both adults and children [32], aiming to (1) assess the acceptability of chatbot technology for promoting and maintaining physical activity and other health-related behaviors in families attending a community-based obesity prevention project, and (2) further assess the acceptability of chatbot intervention in potentially eligible families (not restricted to those attending the project) at the second phase. The intervention time for the 2 stages was the same, lasting 12 weeks. The study found lower interaction times in phase 1 than in phase 2 (65 vs 312 times) but a higher relevant interaction rate in phase 1 (42% vs 11%). The relevant interaction rate was the rate of interactions related to diet, physical activity, or well-being. This indicated that it is possible to actively interact with chatbots in populations without a strong need for health interventions, but it also suggested that encouraging households to purposefully use the device for health-related interactions is a challenge. Another study conducted in America with a sample size of 50 dyads of parent and child aiming to improve family diet behaviors, found that a high percentage of children (81%) and parents (76%) interacted with chatbots at least once [30]. The mean number of calls for parents and children was 9.1 (SD 5.2) and 9.0 (SD 5.7).

A study aiming to improve smoking secession found that the interaction time of the chatbot intervention group was much higher than that of the control group of usual clinical practice (121 minutes vs 21.2 minutes, $P<.001$), and the number of interactions was also much higher, too (45.56 vs 2.92, $P<.001$) [33]. Within the intervention group, those who successfully quit smoking interacted with chatbots much more frequently than those who did not successfully quit smoking. Another study focusing on helping people quit smoking through group chats found that, compared to only receiving smoking cessation information and tips in group chats without social support or interactions with other participants and chatbots, significant increments of active conversations (455/341, increased by 33%) and the number of messages (248/1328, increased by 87%) were found when the chatbot was involved in the group chats [35].

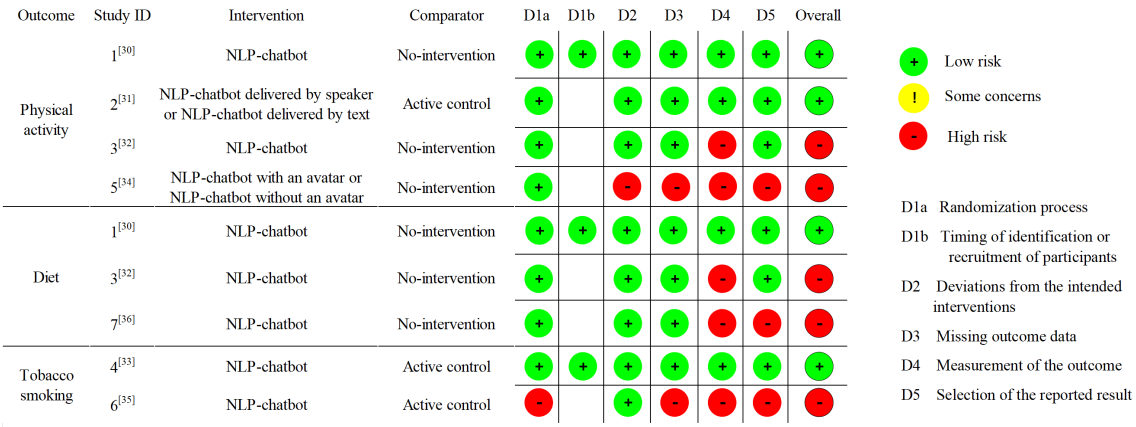
Two included studies measured participants' feelings about chatbot use, and both showed positive results. A web-based research conducted in Dutch with a large sample size of 958 participants measured participants' appreciation for the intervention from entertainment, trustworthiness, and overall appreciation score domains, and found that participants had a very positive impression of chatbot use in all 3 domains [34]. A study conducted through phone calls between participants and chatbots had similar findings that more than 75% of those who had made calls with the chatbot to gain assistance with diet and physical activity agreed that chatbots were useful, credible, financially feasible, and really helped them eat healthy foods [30].

Two included studies reported the results regarding safety issues. A study stated that for the privacy protection of the research subjects, the search history information of the NPL-chatbot used in the research was strictly kept confidential, and during the intervention period, this device was not used for any other purposes [32]. A study emphasized that no adverse events occurred during the intervention period [31].

Assessment of Risk of Bias

Figure 3 shows the result of the assessment of the risk of bias. For the changes in diet behaviors, 2 trials had a high risk of bias [32,36], and 1 trial had a low risk of bias [30]. For the changes in physical activity, 2 trials [32,34] had a high risk of bias, while the remaining 2 trials [30,31] had a low risk of bias. For the changes in tobacco smoking behavior, 1 trial [35] was judged to be at a high risk of bias, while 1 trial had a low risk of bias [33].

Figure 3. Risk of bias (RoB2 and ROB2_cluster tool) quality assessment [30,36]. NLP: natural language processing; RoB: Risk of Bias.



XSL•FO
RenderX

Discussion

Principal Findings

Our study was the first systematic review specifically dedicated to RCTs using NLP-chatbots for health behavior interventions related to physical activity, diet, and tobacco smoking. This focus strengthened the evidence for the results in a relatively strong manner and with the novelty prominent.

The results of the studies on dietary behavior included in our review were inconsistent. Consequently, it was ultimately impossible to clearly determine the impact of chatbots on dietary behavior. It differed slightly from those in previous reviews. One meta-analysis [14] found that chatbot intervention had a significant impact on increasing the intake of fruits and vegetables. Another review [19] showed that participants in the intervention group showed a higher self-reported willingness to reduce consumption of red and processed meat within 2 weeks compared to the control group. These 2 reviews included both RCTs and quasi-experimental studies and did not limit the types of constrained or unconstrained chatbots, differing from the eligibility criteria used in our review. Different inclusion criteria might lead to inconsistent results about the effectiveness of chatbot intervention on the changes in diet behaviors.

Among the 4 studies incorporated into this review, 2 studies discovered that chatbots exerted a positive influence on physical activity. In contrast, the other 2 studies did not observe such an effect, presenting complex and inconclusive results. Some reviews [14,19,22] have reported the positive impacts of chatbots on physical activity behaviors. However, a review [37] specifically focusing on teenage participants aged 10 to 19 years revealed limited evidence regarding the feasibility of chatbots in promoting such behaviors. This review [37] also indicated that in only 40% (2/5) of the studies, the subjects were satisfied with the application of chatbots in interventions, suggesting that there is insufficient evidence for the acceptability of chatbots. The low satisfaction level might lead to a difference in the research results by affecting individual compliance with a chatbot.

The findings of this systematic review showed that NLP chatbot–based intervention had a positive impact on the alteration of adults' smoking behaviors. Regarding the effective role of chatbots in smoking cessation, a systematic review using meta-analysis indicated that at the 6-month follow-up, participants (aged 15 years and older) who received chatbot-based interventions were significantly more likely to quit smoking than those in the control group [15], supporting the findings derived from our review.

Our review revealed that participants had a very positive impression of chatbot use, feeling chatbots are useful, credible, and financially feasible. Similarly, a review indicated that participants emphasized numerous positive aspects of chatbots, especially their unique personalities and the capacity to offer empathetic and emotional support. However, several limitations were also pointed out. For instance, chatbots often had trouble fully understanding users, their responses were repetitive, and they lacked interactivity [18]. Moreover, another review showed that among the 5 studies included, only 2 were content with the application of chatbots in the intervention [37]. We assume that the user experience of participants is closely related to the performance of specific chatbots, such as their language comprehension, interaction, empathy, and persuasion abilities. NLP chatbots that are capable of free dialogue have an advantage.

We conducted a comprehensive literature search of preprinted, unpublished, and published records based on the preregistered study protocol. The inclusion criteria of this review were clear and strict (RCT only or NLP-chatbot only) which reached a high hierarchy of evidence. We analyzed 7 RCTs focusing on diet, physical activity, and tobacco smoking behaviors to explore the effectiveness of chatbot intervention. In this review, in addition to indicators of chatbot-related behavior change, we also paid attention to multiple secondary outcomes to explore an individual's acceptability of chatbot intervention. No change was made to methods when compared to our study protocol.

However, the following limitations need to be noted when interpreting the findings. Few included studies measured user acceptance of chatbot intervention or the promoting or hindering factors of them to use chatbots. This impeded us from further studying the specific mechanism between the population's use of chatbots and the change of related behaviors. There is little literature on privacy issues when using chatbots, which we thought was a serious and significant issue to consider. Besides, about half of the included studies had an elevated overall risk of bias.

Based on this review, we had some suggestions for future research. Chatbots are just a form and carrier of intervention, essentially requiring the support of various behavioral change theories. Future research would better incorporate appropriate theoretical frameworks, such as motivational interviewing theory [38] and the transtheoretical model [39] when designing chatbots to better achieve the goal of promoting health. Most of the included studies focused on a single chatbot intervention through voice conversation and message exchange in the intervention group, limiting the potential applications of chatbots. Future studies can integrate chatbots with other interventions to augment compliance with other interventions. Approximately half of the included studies did not describe the acceptability of NLP chatbot–based interventions. From the perspective of research implementation, future research needs to pay more attention to process evaluation, such as frequency and time of chatbot use. In terms of research outcome, future research needs to evaluate both outcome indicators and mediators, such as knowledge, motivation, and intention, to explore the deep reasons for the behavior change of participants after the chatbot intervention.

Conclusion

Our results indicated that NLP-chatbots were promising in reducing tobacco smoking among adults, while their effects on the changes in dietary and physical activity behaviors remained inconclusive. Future research can be improved in aspects such as increasing the theoretical support for interventions and monitoring the interaction between users and NLP-chatbots.

XSL•FO
RenderX

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (Nos.82373694), Young Elite Scientists Sponsorship Program by CAST (China Association for Science and Technology; 2023QNRC001), Beijing Education Sciences Planning Program during the 14th Five-Year Plan (No.BECA23111), and the Fundamental Research Funds for the Central Universities (No.BMU2021YJ030).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms for databases.
[PDF File, 142 KB - mhealth_v13i1e66403_app1.pdf]

Multimedia Appendix 2
Table S1: characteristics and outcomes.
66403-1122796-1-SP.xlsx

Checklist 1
PRISMA checklist. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[PDF File, 72 KB - mhealth_v13i1e66403_app2.pdf]

References

1. Noncommunicable diseases. World Health Organization. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2025-05-08]
2. World Health Organization. Global status report on physical activity 2022. : Geneva: World Health Organization; 2022 URL: https://www.who.int/teams/health-promotion/physical-activity/global-status-report-on-physical-activity-2022 [accessed 2025-05-13]
3. Global recommendations on physical activity for health. : World Health Organization; 2010 URL: https://www.who.int/publications/i/item/9789241599979 [accessed 2025-05-08]
4. Kalmpourtzidou A, Eilander A, Talsma EF. Global vegetable intake and supply compared to recommendations: a systematic review. Nutrients 2020 May 27;12(6):1558. [doi: 10.3390/nu12061558] [Medline: 32471188]
5. Average per capita vegetable intake vs minimum recommended guidelines. Our World in Data. 2023. URL: https://ourworldindata.org/grapher/average-per-capita-vegetable-intake-vs-minimum-recommended-guidelines [accessed 2025-05-08]
6. Fruit consumption per capita. Our World in Data. 2023. URL: https://ourworldindata.org/grapher/fruit-consumption-per-capita [accessed 2025-05-08]
7. World Health Organization. WHO global report on trends in prevalence of tobacco use 2000–2030. : Geneva: World Health Organization; 2024 URL: https://www.who.int/publications/i/item/9789240088283 [accessed 2025-05-08]
8. World obesity atlas. World Obesity Federation. 2023. URL: https://data.worldobesity.org/publications/ [accessed 2025-05-08]
9. World Health Organization. Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020.: World Health Organization; 2013 URL: https://www.who.int/publications/i/item/9789241506236 [accessed 2025-05-08]
10. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020 Nov 3;22(11):e20251. [doi: 10.2196/20251] [Medline: 33141095]
11. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. JMIR mHealth uHealth 2020 Jul 10;8(7):e17558. [doi: 10.2196/17558] [Medline: 32673246]
12. Kocielnik R, Xiao L, Avrahami D, et al. Reflection companion: a conversational system for engaging users in reflection on physical activity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(2):70. [doi: 10.1145/3214273]
13. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845. [doi: 10.2196/22845] [Medline: 32996892]
14. Singh B, Olds T, Brinsley J, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. NPJ Digit Med 2023 Jun 23;6(1):118. [doi: 10.1038/s41746-023-00856-1] [Medline: 37353578]

XSL•FO
RenderX

15. Bendotti H, Lawler S, Chan GCK, Gartner C, Ireland D, Marshall HM. Conversational artificial intelligence interventions to support smoking cessation: a systematic review and meta-analysis. Digit Health 2023;9:20552076231211634. [doi: 10.1177/20552076231211634] [Medline: 37928336]
16. Kramer LL, Ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing embodied conversational agents for coaching people in a healthy lifestyle: scoping review. J Med Internet Res 2020 Feb 6;22(2):e14058. [doi: 10.2196/14058] [Medline: 32022693]
17. Li Y, Liang S, Zhu B, et al. Feasibility and effectiveness of artificial intelligence-driven conversational agents in healthcare interventions: a systematic review of randomized controlled trials. Int J Nurs Stud 2023 Jul;143:104494. [doi: 10.1016/j.ijnurstu.2023.104494] [Medline: 37146391]
18. Milne-Ives M, de Cock C, Lim E, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346. [doi: 10.2196/20346] [Medline: 33090118]
19. Oh YJ, Zhang J, Fang ML, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act 2021 Dec 11;18(1):160. [doi: 10.1186/s12966-021-01224-6] [Medline: 34895247]
20. Chew HSJ. The use of artificial intelligence-based conversational agents (chatbots) for weight loss: scoping review and practical recommendations. JMIR Med Inform 2022 Apr 13;10(4):e32578. [doi: 10.2196/32578] [Medline: 35416791]
21. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258. [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
22. Luo TC, Aguilera A, Lyles CR, Figueroa CA. Promoting physical activity through conversational agents: mixed methods systematic review. J Med Internet Res 2021 Sep 14;23(9):e25486. [doi: 10.2196/25486] [Medline: 34519653]
23. Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions 6.4. Cochrane. 2023. URL: www.training.cochrane.org/handbook [accessed 2025-05-08]
24. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341. [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
25. Brennan SE, Mckenzie JE. Chapter 12: synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 65: Cochrane; 2024.
26. Snilstveit B, Vojtkova M, Bhavsar A, Stevenson J, Gaarder M. Evidence & gap maps: a tool for promoting evidence informed policy and strategic research agendas. J Clin Epidemiol 2016 Nov;79:120-129. [doi: 10.1016/j.jclinepi.2016.05.015] [Medline: 27387966]
27. Miake-Lye IM, Hempel S, Shanman R, Shekelle PG. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. Syst Rev 2016 Feb 10;5:28. [doi: 10.1186/s13643-016-0204-x] [Medline: 26864942]
28. Higgins JPT, Page MJ, Elbers RG, Sterne JAC. Chapter 8: assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 6.4: Cochrane; 2023. URL: www.training.cochrane.org/handbook [accessed 2025-05-08]
29. Higgins JPT, Eldridge S, Campbell MK, et al. Revised cochrane risk of bias tool for randomized trials (rob 2) for cluster-randomized trials. riskofbias.info. 2021. URL: https://www.riskofbias.info/welcome/rob-2-0-tool/rob-2-for-cluster-randomized-trials [accessed 2025-05-08]
30. Wright JA, Phillips BD, Watson BL, Newby PK, Norman GJ, Adams WG. Randomized trial of a family-based, automated, conversational obesity treatment program for underserved populations. Obesity (Silver Spring) 2013 Sep;21(9):E369-E378. [doi: 10.1002/oby.20388] [Medline: 23512915]
31. Hassoon A, Baig Y, Naiman DQ, et al. Randomized trial of two artificial intelligence coaching interventions to increase physical activity in cancer survivors. NPJ Digit Med 2021 Dec 9;4(1):168. [doi: 10.1038/s41746-021-00539-9] [Medline: 34887491]
32. Carlin A, Logue C, Flynn J, Murphy MH, Gallagher AM. Development and feasibility of a family-based health behavior intervention using intelligent personal assistants: randomized controlled trial. JMIR Form Res 2021 Jan 28;5(1):e17501. [doi: 10.2196/17501] [Medline: 33507155]
33. Olano-Espinosa E, Avila-Tomas JF, Minue-Lorenzo C, et al. Effectiveness of a conversational chatbot (Dejal@bot) for the adult population to quit smoking: pragmatic, multicenter, controlled, randomized clinical trial in primary care. JMIR mHealth uHealth 2022 Jun 27;10(6):e34273. [doi: 10.2196/34273] [Medline: 35759328]
34. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a web-based physical activity intervention with an avatar: randomized controlled trial. J Med Internet Res 2014 Feb 13;16(2):e48. [doi: 10.2196/jmir.2974] [Medline: 24550153]
35. Wang H, Zhang Q, Ip M, Fai Lau JT. Social media–based conversational agents for health management and interventions. Computer (Long Beach Calif) 2018;51(8):26-33. [doi: 10.1109/MC.2018.3191249]
36. Alghamdi E, Alnanih R. Chatbot design for a healthy life to celiac patients: a study according to a new behavior change model. IJACSA 2021;12(10):12. [doi: 10.14569/IJACSA.2021.0121077]

XSL•FO
RenderX

37. Han R, Todd A, Wardak S, Partridge SR, Raeside R. Feasibility and acceptability of chatbots for nutrition and physical activity health promotion among adolescents: systematic scoping review with adolescent consultation. JMIR Hum Factors 2023 May 5;10:e43227. [doi: 10.2196/43227] [Medline: 37145858]
38. Bischof G, Bischof A, Rumpf HJ. Motivational interviewing: an evidence-based approach for use in medical practice. Dtsch Arztebl Int 2021 Feb 19;118(7):109-115. [doi: 10.3238/arztebl.m2021.0014] [Medline: 33835006]
39. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]

Abbreviations

AI: artificial intelligence
E: effective
NE: not effective
NLP: natural language processing
PE: probably effective
PNE: not to be probably effective
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RoB: Risk of Bias

Edited by Z Ni; submitted 12.09.24; peer-reviewed by A Jafarizadeh, KM O'Brien, M Chatzimina; revised version received 02.04.25; accepted 09.04.25; published 11.06.25.

Please cite as:
Chen J, Hu RZ, Zhuang YX, Zhang JQ, Shan R, Yang Y, Liu Z
Natural Language Processing Chatbot–Based Interventions for Improvement of Diet, Physical Activity, and Tobacco Smoking Behaviors: Systematic Review
JMIR Mhealth Uhealth 2025;13:e66403
URL: https://mhealth.jmir.org/2025/1/e66403
doi:10.2196/66403

©Jing Chen, Run-Ze Hu, Yu-Xuan Zhuang, Jia-Qi Zhang, Rui Shan, Yang Yang, Zheng Liu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Use of Artificial Intelligence and Wearable Inertial Measurement Units in Medicine: Systematic Review

Ricardo Smits Serena[1,2], MSc; Florian Hinterwimmer[1,2], Dr; Rainer Burgkart[1], Prof Dr Med; Rudiger von Eisenhart-Rothe[1], Prof Dr Med; Daniel Rueckert[2], Dr

[1]Department of Orthopaedics and Sports Orthopaedics, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany

[2]Institute for AI and Informatics in Medicine, Technical University of Munich, Munich, Germany

Corresponding Author:
Ricardo Smits Serena, MSc
Department of Orthopaedics and Sports Orthopaedics
Klinikum rechts der Isar
Technical University of Munich
Ismaninger Strasse 22
Munich, 81675
Germany
Phone: 49 8941402271
Email: ricardo.smits@tum.de

Abstract

Background: Artificial intelligence (AI) has already revolutionized the analysis of image, text, and tabular data, bringing significant advances across many medical sectors. Now, by combining with wearable inertial measurement units (IMUs), AI could transform health care again by opening new opportunities in patient care and medical research.

Objective: This systematic review aims to evaluate the integration of AI models with wearable IMUs in health care, identifying current applications, challenges, and future opportunities. The focus will be on the types of models used, the characteristics of the datasets, and the potential for expanding and enhancing the use of this technology to improve patient care and advance medical research.

Methods: This study examines this synergy of AI models and IMU data by using a systematic methodology, following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, to explore 3 core questions: (1) Which medical fields are most actively researching AI and IMU data? (2) Which models are being used in the analysis of IMU data within these medical fields? (3) What are the characteristics of the datasets used for in this fields?

Results: The median dataset size is of 50 participants, which poses significant limitations for AI models given their dependency on large datasets for effective training and generalization. Furthermore, our analysis reveals the current dominance of machine learning models in 76% on the surveyed studies, suggesting a preference for traditional models like linear regression, support vector machine, and random forest, but also indicating significant growth potential for deep learning models in this area. Impressively, 93% of the studies used supervised learning, revealing an underuse of unsupervised learning, and indicating an important area for future exploration on discovering hidden patterns and insights without predefined labels or outcomes. In addition, there was a preference for conducting studies in clinical settings (77%), rather than in real-life scenarios, a choice that, along with the underapplication of the full potential of wearable IMUs, is recognized as a limitation in terms of practical applicability. Furthermore, the focus of 65% of the studies on neurological issues suggests an opportunity to broaden research scope to other clinical areas such as musculoskeletal applications, where AI could have significant impacts.

Conclusions: In conclusion, the review calls for a collaborative effort to address the highlighted challenges, including improvements in data collection, increasing dataset sizes, a move that inherently pushes the field toward the adoption of more complex deep learning models, and the expansion of the application of AI models on IMU data methodologies across various medical fields. This approach aims to enhance the reliability, generalizability, and clinical applicability of research findings, ultimately improving patient outcomes and advancing medical research.

(JMIR Mhealth Uhealth 2025;13:e60521) doi:10.2196/60521

XSL•FO
RenderX

KEYWORDS

artificial intelligence; accelerometer; gyroscope; IMUs; time series data; wearable; systematic review; patient care; machine learning; data collection

Introduction

The integration of advanced computational methods such as artificial intelligence (AI) in medicine represents a significant leap in the pursuit of more accurate, efficient, and personalized health care [1]. The increasing popularity of wearable devices has led to a surge in the collection of various physiological signals, including accelerometer data from wristbands, smartwatches, and other sensors [2]. While these wearables offer valuable insights into our daily activities, inertial measurement units (IMUs) stand out for their unique ability to capture 3-dimensional motion data, including acceleration, angular velocity, and orientation. Furthermore, IMUs are also highly accessible and widely available, making them an attractive choice for researchers and clinicians alike. This combination of precision and accessibility has led to a surge in the adoption of IMUs across various medical fields, from neurology to emergency medicine, where they are increasingly being used to track complex motor behaviors, such as those seen in neurological disorders or injuries.

The ability of AI models to extract meaningful patterns from complex, multidimensional movement data offers unprecedented opportunities in diagnostics, patient monitoring, and treatment efficacy assessment. However, despite the potential benefits, there are many challenges in applying these technologies in a medical context [1]. These include issues related to data preprocessing, model selection, dataset size, and study setting (dataset bias), which can significantly impact the validity and reliability of the results. In addition, integrating these new technologies into existing clinical workflows poses significant logistical and ethical challenges, including data privacy concerns, the need for extensive validation, and the training of health care professionals to effectively use these tools.

This systematic review aims to provide a comprehensive overview of the current state of AI applications in processing IMU data for medical purposes. By examining a variety of studies across different medical specialties, this review seeks to understand the common methodologies, tools, and challenges faced in this rapidly evolving field. In doing so, it aims to identify gaps in the current knowledge base, suggest areas for future research, and provide insights into the practical implications of these technologies in clinical settings. Ultimately, the goal is to foster a deeper understanding of how AI algorithms and IMUs can be harnessed to improve patient outcomes and streamline health care delivery.

Methods

Search Strategy

The design of this systematic review was centered around 3 fundamental research questions, all pertaining to the application of AI models in analyzing IMU data within various medical fields. The first question aimed to identify which medical fields are most actively incorporating such techniques with IMU data. The second question focused on determining the specific models that are used in the analysis of IMU data. Finally, the third question tries to understand the characteristics of the datasets used for training. To address these questions, this review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [3] to ensure a methodical approach to data collection and analysis, with the PRISMA checklist provided in Multimedia Appendix 1.

The literature search was conducted on November 15, 2023, using 2 major scientific databases: PubMed and Web of Science. PubMed yielded 327 articles with the search string: (“Artificial Intelligence”[Mesh]) AND (health OR medical OR clinical OR patient) AND (wearabl* OR smartphone OR smartwatch) AND (magnetometer OR accelerom* OR gyroscope OR “imu” OR “imus” OR “Inertial measurement unit”) AND (health OR medical OR clinical OR patient) NOT (Review[Publication Type] OR Systematic Review[Publication Type]).

Web of Science yielded 853 articles with the search string: (“Artificial Intelligence” OR “Machine Learning” OR “Deep Learning”) AND (health OR medical OR clinical OR patient) AND (wearabl* OR smartphone OR smartwatch) AND (magnetometer OR accelerom* OR gyroscope OR “imu” OR “imus” OR “Inertial measurement unit”) AND (health OR medical OR clinical OR patient).

Study Selection

The retrieved articles were screened by 2 Computer Scientists (RSS and FH) based on their titles and abstracts to assess their relevance to the research questions (inclusion and exclusion criteria outlined in Textbox 1).

XSL•FO
RenderX

Textbox 1. Guidelines for selecting studies from the literature research.

Inclusion criteria

- Study uses 1 or more inertial measurement units.
- Study targets a clinical condition.
- Study focuses on machine learning or deep learning algorithms for data analysis.
- Study published in English.
- Studies with available full-text articles.
- Studies with more than 20 participants.

Exclusion criteria

- Review articles.
- Studies not related to medical applications.
- Studies not using inertial measurement unit data.
- Study not mentioning details on the machine learning or deep learning techniques used.

Data Extraction

In each included study, relevant data were extracted, concentrating on several key aspects. This included the medical field of application, the specific purpose of using machine learning/deep learning (ML/DL), and the ML/DL methods. In addition, the size of the study dataset and the study setting were used. The synthesized data from this extraction process provided a comprehensive overview, highlighting the current state of ML/DL applications in processing IMU data for medical purposes. Based on research suggesting that participant diversity is required for good generalization [4], we consider dataset size to be the number of participants.

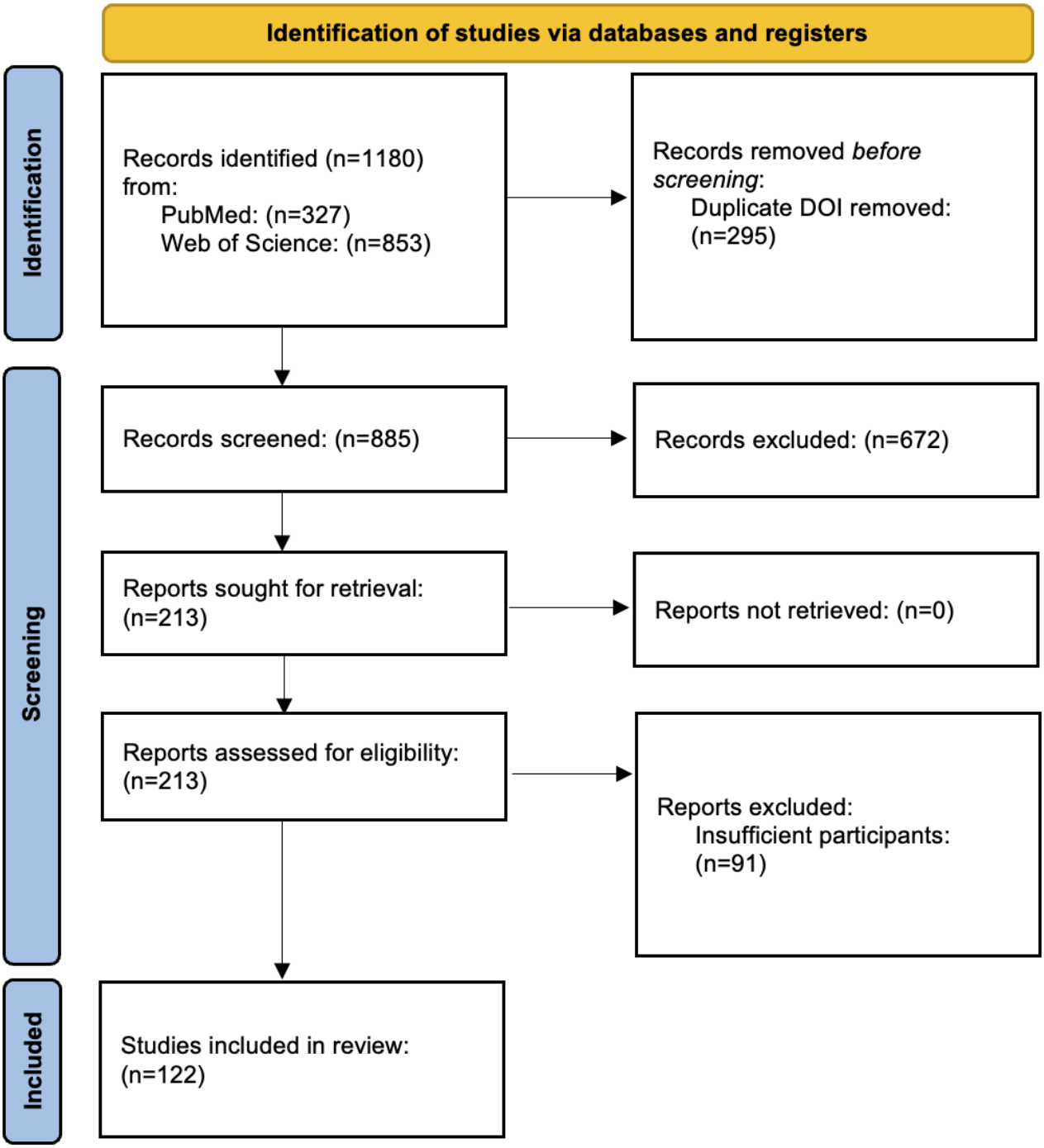
Results

Overview

This systematic review initially identified 1180 papers through a comprehensive search. After the screening process, 122 papers from the years 2008 to 2023 were selected for detailed analysis. The PRISMA diagram illustrating this selection process is presented in Figure 1, and the distribution of these papers over the years is found in Figure 2. The results are organised based on predefined research questions, providing a detailed overview of the current landscape, prevalent trends, and challenges in this domain.

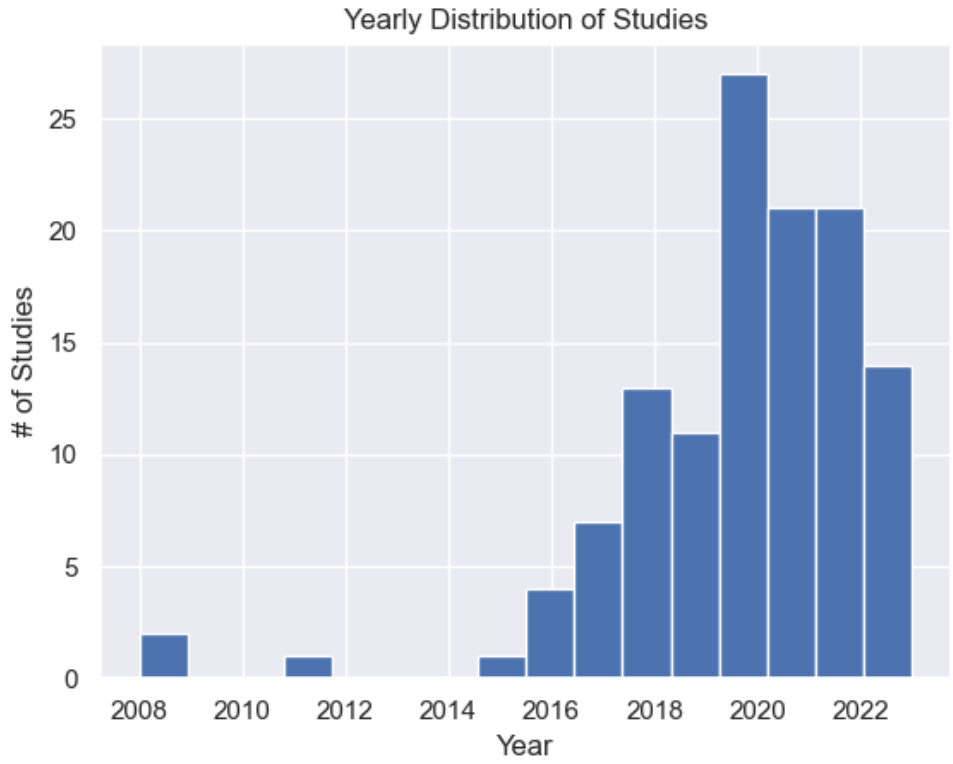
XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the filtering procedure applied to bibliographic search results.



XSL•FO
RenderX

Figure 2. Histogram of annual publication frequency of the final 122 studies.



Distribution of Papers

Neurology stood out as the most represented field, with a substantial count of 79 (64.7%) papers, underscoring its significant role in the application of IMU data. Geriatric medicine followed, with 12 (9.8%) papers, reflecting an interest in this area for elderly care. Pulmonary and cardiology medicine, with 10 (8.1%) papers, indicated their importance in chronic disease management using these technologies. Musculoskeletal and orthopedic care was another notable field, represented by 8 (6.5%) papers, emphasizing the role of AI in diagnosing and treating musculoskeletal disorders. Pediatric medicine and psychiatry also featured in the analysis, with 4 (3.2%) and 2 (1.6%) papers respectively, highlighting some areas of opportunity. In addition, fields such as Emergency medicine, general and preventive medicine, internal medicine, oncology, otolaryngology, sleep medicine, and vestibular medicine each appearing one time, in total 7, demonstrating the wide-ranging applications of IMU data across various medical disciplines.

In analyzing the applications of AI models on medical IMU data, disease monitoring emerged as the primary focus, with 38 (31.1%) papers dedicated to this aspect. This indicates a strong trend toward using ML models for ongoing patient care and management. Disease detection and disease severity assessment were also prominent, with 19 (15.5%) and 24 (19.6%) papers respectively, highlighting the interest early diagnosis and evaluation of disease progression. Functional and mobility assessment, covered in 13 (10.6%) papers, pointed to the importance of AI in evaluating patients' physical abilities. Rehabilitation and recovery monitoring, with 10 (8.1%) papers, further emphasized the use of AI in patient recovery processes. Risk prediction and preventive analysis, represented by 18 (14.7%) papers, showed a proactive approach in using AI for forecasting health risks and preventing diseases (Table 1).

XSL•FO
RenderX

Table 1. List of studies' references ordered by medical field and machine learning purpose.

Medical field and AI[a] purpose	Studies	Total, n
Neurology		
Disease monitoring	[5-32]	28
Disease severity assessment	[33-50]	18
Disease detection	[51-62]	12
Functional and mobility assessment	[63-70]	8
Risk prediction and preventive analysis	[71-77]	7
Rehabilitation and recovery monitoring	[78-83]	6
Geriatrics		
Risk prediction and preventive analysis	[84-91]	8
Functional and mobility assessment	[92,93]	2
Disease monitoring	[94]	1
Disease severity assessment	[95]	1
Cardiorespiratory		
Disease detection	[96-101]	6
Disease monitoring	[102-105]	4
Rehabilitation and recovery monitoring	[106]	1
Orthopedics		
Disease monitoring	[107,108]	2
Disease severity assessment	[53,109]	2
Rehabilitation and recovery monitoring	[110,111]	2
Functional and mobility assessment	[112]	1
Risk prediction and preventive analysis	[113]	1
Pediatrics		
Functional and mobility assessment	[114,115]	2
Disease monitoring	[116]	1
Disease severity assessment	[117]	1
Psychiatry		
Disease monitoring	[118,119]	2
Emergency Medicine		
Risk prediction and preventive analysis	[120]	1
Preventive medicine		
Rehabilitation and recovery monitoring	[121]	1
Oncology		
Risk prediction and preventive analysis	[122]	1
Otolaryngology		
Disease detection	[123]	1
Sleep medicine		
Disease severity assessment	[124]	1
Vestibular medicine		
Disease severity assessment	[125]	1

XSL•FO
RenderX

Machine Learning and Deep Learning Models Used in Inertial Measurement Unit Data Analysis

In total, 326 models were identified across the 122 reviewed papers, averaging 2.67 tested models per paper. The analysis of the AI models used for analyzing IMU data revealed distinct trends in the usage of both ML and DL approaches. The results are presented in 2 subsections detailing the prevalence and variety of models observed in the reviewed papers. Performance metrics were not included in the analysis because studies often used different metrics, making direct comparison impossible. Table 2 provides a list of all studies sorted by the AI model used. If a study employed more than one model, it appears multiple times. A full description of the model names is provided in Multimedia Appendix 2.

XSL•FO
RenderX

Table 2. List of studies sorted by artificial intelligence model used in the study.

Name	Studies	Total, n
Support vector machine	[6,7,9-13,15,17,20,21,25-27,31,33,36,38-43,45-48,52,56-58,62,65,66,68,72-75,81,87,88,93,94,96,98,101,104-106,110,111,113,115,119,121,123,125,126]	61
Random forest	[7,9,11,16,17,25,29,38-40,44,45,47,48,51,55,56,58,61,65-68,73,74,77,80,82,87,91,94,95,97,98,104,105,111,114,115,117,121,123,126]	43
k-nearest neighbors	[10,20,22,27,33,36,40,41,43,44,47,52,56,58,59,62,65,73,81,87,91,95,111,117,123]	25
Convolutional neural network	[7-9,17,18,23,29,30,39,53,54,63,64,78,85,94,103,107,120,124]	20
Linear regression	[5,27,35,38,39,44,47,51,52,56,82,90,91,100,104,111,121]	17
Decision tree	[10,20,33,38-41,58,72,73,81,97,98,111,117,123,126]	17
Naive bayes	[10,25,33,36,39,43,47,52,56,58,69,72,74,88]	14
Multilayer perceptron	[39,44,86,92,95,99,109,113,117,126]	10
Long short-term memory	[14,29,34,71,89,94,116,122,124]	9
Neural network	[27,32,72,88,97,105,108,125]	8
XGBoost	[7,94,98,105,108]	5
Support vector regressor	[45,49,91,109]	4
Bagged trees	[11,56,62,81]	4
CNN-LSTM[a]	[9,17,34,76]	4
AdaBoost	[9,17,58,104]	4
Linear discriminant analysis	[47,52,81,126]	4
Gaussian process regression	[69,83,109]	3
Lasso	[60,91,123]	3
Discriminant analysis	[40,41,43]	3
Gradient boosted tree	[14,34,121]	3
Light gradient boosted machine	[55,111]	2
Principal component analysis	[13,70]	2
Inception time	[19,65]	2
Gaussian mixture model	[24,28]	2
Boosted decision tree	[92,115]	2
Hidden Markov model	[28,45]	2
Denoising autoencoder	[9,17]	2
K-means	[113,118]	2
Deep neural network	[29,35]	2
Fully convolutional network	[65,107]	2
Residual neural network	[94,107]	2
Tree ensemble	[33,43]	2
Extra trees	[111]	1
Gradient boosting	[95]	1
Seq2Seq	[50]	1
Gaussian process	[117]	1
Bayesian regression	[91]	1
Ridge regression	[91]	1
ASRF[b]	[50]	1
Boosted ensemble decision trees	[112]	1

XSL•FO
RenderX

Name	Studies	Total, n
NN-DTW[c]	[82]	1
Recurrent neural network	[116]	1
Cluster tree	[52]	1
Generative kernel	[52]	1
Gated recurrent unit	[64]	1
t-SNE[d]	[106]	1
MLSTM-FCN[e]	[107]	1
MALSTM-FCN[f]	[107]	1
LRCF[g]	[84]	1
Nonnegative matrix factorization	[70]	1
Multidimensional scaling	[70]	1
Transformer	[29]	1
Baum-Welch	[28]	1
Random forest classifier	[119]	1
Logistic model trees	[25]	1
J48[h]	[25]	1
Neural network clustering	[22]	1
Gaussian naive bayes	[73]	1
Feedforward neural network	[45]	1
Multilayer Gaussian process	[37]	1
Radial basis function network	[66]	1
M5 prime algorithm	[44]	1
Blending ensemble learning	[120]	1
Generative adversarial network	[23]	1
Adaptive mixtures of local experts	[92]	1
Bayesian decision support	[92]	1
Sparse component analysis	[118]	1
Extreme learning machine	[79]	1
Naive bayes classifier	[68]	1
Structural equation modeling	[56]	1
Bayesian ridge regression	[109]	1
Gradient boosting algorithm	[73]	1
CRNN[i]	[102]	1

[a]CNN-LSTM: convolutional neural network followed by a long short-term memory layer.

[b]ASRF: action segment refinement framework.

[c]NN-DTW: neural network with dynamic time warping.

[d]t-SNE: t-distributed stochastic neighbor embedding.

[e]MLSTM-FCN: multidimensional long short-term memory followed by a fully convolutional network.

[f]MALSTM-FCN: multidimensional attention long short-term memory followed by a fully convolutional network.

[g]LRCF: local rank convolutional framework.

[h]J48: Java implementation of the C4.5 algorithm for decision trees.

[i]CRNN: convolutional recurrent neural network.

XSL•FO
RenderX

A significant majority of the models used, 232 (71.1%) in total, were supervised ML models, showcasing their widespread adoption in IMU data analysis. In contrast, unsupervised ML models were less commonly used, with only 15 (4.6%) papers reporting their application. Among these ML models, 5 methods stood out in terms of frequency, support vector machine was the most prevalent, then random forest, k-nearest neighbors, and decision trees followed and all had an average publication year of 2020. Only logistic regression had a slightly earlier average publication year of 2019.

These average publication years suggest that support vector machine, random forest, k-nearest neighbors, and decision trees, even if they are considered as old models, have been predominantly used in more recent studies. In contrast, the slightly earlier average year for logistic regression may reflect a decrease of interest on this model.

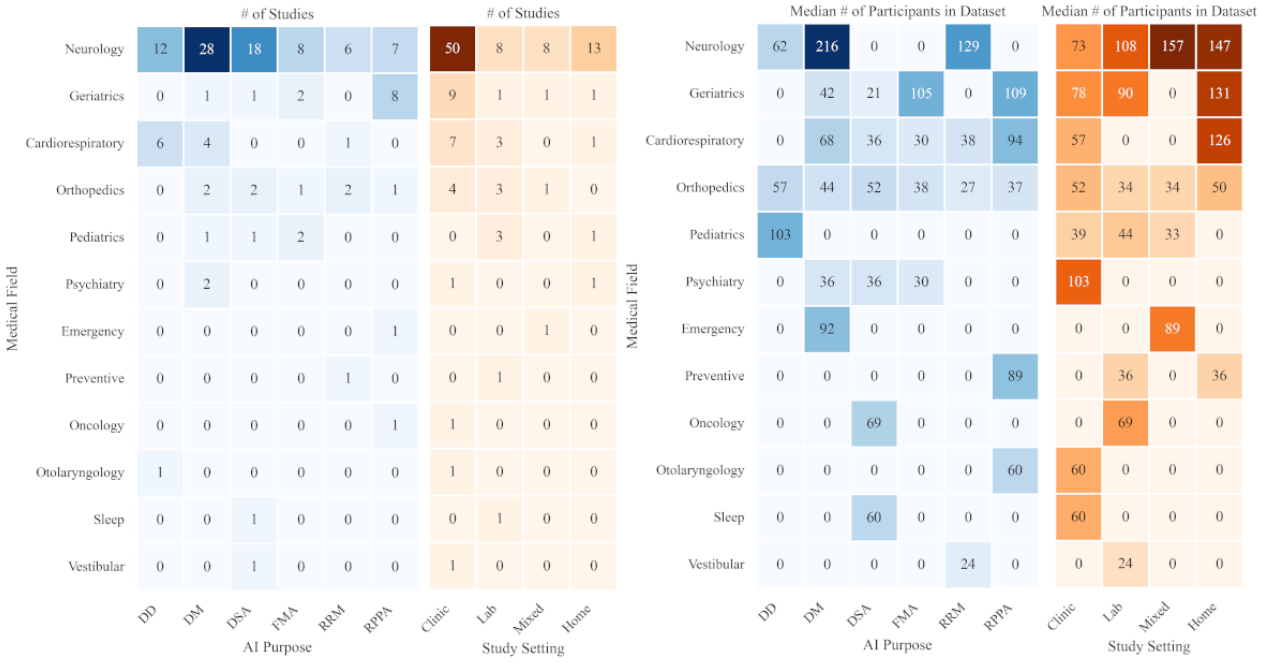
In the realm of DL, 72 (22%) models were supervised. Again, unsupervised DL models were considerably less common, with only 7 (2.1%) models used. The top 5 DL methods identified were convolutional neural networks (CNN), leading with an average publication year of 2020, multilayer perceptron (MLP) averaging at 2019, long short-term memory (LSTM) networks, with an average publication year of 2021, neural networks (NN), also averaging 2019, and CNN-LSTM hybrid models averaging at 2020.

These average publication years suggest that LSTM networks have seen increased adoption in the most recent studies (2021) over other models, likely due to their strength in handling sequential and time-series data inherent in IMU analyses. The consistent average years for CNN and CNN-LSTM models (2020) indicate sustained interest and applicability in feature extraction and sequence modeling. In contrast, the slightly earlier average years for MLP and general NNs (2019) may reflect earlier exploration phases or a shift toward more specialized architectures in recent research.

Dataset Size

The median dataset sizes varied significantly across different medical fields. The field of general and preventive medicine used the smallest median dataset size at 24, followed by pediatric medicine with 36 (IQR 32.5-36.5). Musculoskeletal and orthopedic care, neurology, and sleep medicine had median sizes of 38.5 (IQR 29.8-59.5), 45 (IQR 30-78.5), and 69, respectively. Vestibular medicine and sleep medicine both showed a median dataset size of 60. Larger median dataset sizes were observed in emergency medicine at 89, geriatric medicine at 91.5 (IQR 66.5-147.75), psychiatry at 91.5 (IQR 74.3-108.8), and pulmonary and cardiology medicine at 93.5 (IQR 42.5-130). The largest median dataset size was noted in otolaryngology, with 103 participants (Figure 3).

Figure 3. The left heatmap quantifies the volume of studies, while the right heatmap displays the median number of participants within datasets employed across these studies. The blue sides illustrate the diversity of machine learning applications—Disease Detection (DD), Disease Monitoring (DM), Disease Severity Assessment (DSA), Functional and Mobility Assessment (FMA), Rehabilitation and Recovery Monitoring (RRM), and Risk Prediction and Preventive Analysis (RPPA)—across various medical specialties. The orange sides visualize the data collection settings for ML training datasets, encompassing clinical, laboratory, home, and hybrid environments. The median number of participants was used to reduce the impact of the outliers with very big datasets.



When examining the dataset sizes based on the purpose of ML application, varying trends were observed. Rehabilitation and recovery monitoring had the smallest median dataset size at 31.5 (IQR 24-42.3). This was followed by functional and mobility assessment at 38 (IQR 32-53), disease severity assessment at 49 (IQR 30-80.5), disease monitoring at 49.5 (IQR 29.3-103) and disease detection at 58 (IQR 44-86.5). The largest median dataset size, at 74 (IQR 52-109.5), was for risk prediction and preventive analysis.

XSL•FO
RenderX

The study settings also influenced the median dataset sizes. In clinic and real-life mixed settings, the median dataset size was 35 (IQR 30-69.5). Laboratory settings showed a similar median size of 35.5 (IQR 29.75-81.75). The clinic setting alone had a median size of 54.5 (IQR 38.25-83). Finally, the real-life setting presented the largest median dataset size of 60 (IQR 24-147).

Overall, the median dataset size across all studies was 50.5 (IQR 33.25-87.25), indicating a moderate scale of data used in training AI algorithms in these diverse settings and purposes." Under discussion section it is also 50.5 (IQR 33.25-87.25).

Discussion

Principal Findings

One of the most significant findings is the overall median dataset size of 50.5 (IQR 54) participants, and it raises significant concerns about the practical effectiveness of ML and, more so, DL methodologies in this context. These algorithms perform better with large datasets as they not only train more effectively but also ensure the generalizability of their outcomes across diverse populations and conditions [127]. The relatively small size of datasets in current research could limit the robustness and generalizability of ML and DL technologies, rather than their accuracy. Furthermore, even though there is no exact definition of what constitutes a “small” or “large” dataset, it is understood that the number must be large enough to represent the population adequately in order to generalize effectively [4], highlighting the need for a sufficiently large dataset.

Given the importance of participant variability for achieving good generalization, these findings about dataset size, or dataset variability, also highlight the importance of differentiating dataset length and dataset variability in the context of IMU data use of ML in health care. Dataset length refers to the acquisition of a broad number of data points from a participant (eg, 2 participants and 1000 datapoints per participant), which could potentially lead to overfitting due to the emphasis on big multidimensional arrays of features such as temporal patterns, spatial movements, or biometric signals. This focus might inadvertently prioritize the complexity of the dataset over its generalizability, suggesting that it could be less beneficial, and possibly even detrimental, compared to dataset variability which focuses on increasing the dataset by adding more participants and not more data per participant (eg, 50 participants and 40 datapoints per participant). Therefore, with a balance between dataset length and dataset variability, researchers could better leverage ML and DL methodologies to their full potential in this field.

The challenges highlighted above underline the need for combined efforts to expand the scope and scale of data collection in future studies. To assemble larger and more diverse datasets, researchers should consider multi-institutional collaborations or using new technologies for data collection and sharing. Furthermore, even with successful collection of wearable IMU data or data from other wearable devices, a significant obstacle remains in integrating this data into hospital systems or electronic health records [128]. The current lack of infrastructure and protocols for efficiently transferring and storing wearable device data within medical records systems not only impedes its immediate use in clinical decision making but also makes it challenging to gather data for future research and build large datasets for long-term patient monitoring and analysis. There’s a critical need for developing standardized protocols and systems capable of handling the varied data types produced by wearable technologies to ensure this valuable information is integrated into patient care workflows and health records management seamlessly. Finally, advancements in ML and DL techniques that are adept at working with smaller datasets, or that improve data augmentation practices, could play a crucial role in navigating these challenges. Such innovations could provide a pathway to maximize the use of wearable technologies in improving patient outcomes and propelling medical research forward, despite the current limitations posed by dataset sizes and integration issues.

Another finding indicates a substantial reliance on supervised machine learning methods, observed in 71.1% of the studies, which may also be attributed to the limited number of publicly available labeled datasets in this area. This highlights the reliance on labeled data for model training and validation. However, the labor-intensive process of data labeling, especially with complex medical datasets, poses significant challenges [129]. In this context, unsupervised learning methods are a compelling alternative, offering the advantage of discovering hidden patterns and correlations in data without the need for prelabeled instances. The application of unsupervised learning techniques can mitigate the bottleneck of data labeling, enabling researchers to leverage larger datasets more efficiently. This approach would not only facilitate a deeper understanding of the underlying data structures but also aid in identifying novel biomarkers and disease subtypes.

On the other hand, DL contains enormous potential for analyzing IMU medical data, particularly when fueled by larger datasets. DL’s capacity to model complex, nonlinear relationships, and its prowess in feature extraction from high-dimensional data make it exceptionally suited for IMU data analysis. The challenge of assembling substantial, well-annotated datasets in the medical domain, however, limits the full exploitation of deep learning’s capabilities. The availability of larger datasets, potentially facilitated by unsupervised learning and other innovative data gathering and processing techniques, would not only address this limitation but also propel forward the development of more sophisticated and accurate DL models. By enabling access to more extensive and diverse data, the research community could accelerate the advancement of deep learning methodologies, leading to breakthroughs in predictive accuracy, diagnostic reliability, and ultimately, patient outcomes.

Another interesting finding is that, on average, each study uses 2.67 ML or DL techniques. This statistic highlights the field’s ongoing experimental nature and its search for direction, as researchers actively explore a variety of algorithms and model architectures to identify the most effective methods for processing IMU data in medical applications. In addition, we found that for classification tasks the metrics used to evaluate these methods vary widely among studies. Some use accuracy, others use F_1-score, and some rely on the area under the curve.

XSL•FO
RenderX

This diversity in metrics makes it difficult to compare results directly and to ascertain which ML/DL models are most effective for medical research involving IMU data. The absence of standardized evaluation metrics is a significant barrier to effectively comparing and synthesizing results from various studies, thereby underlining the importance of establishing reproducibility and enabling scalable research methods.

To address these challenges, it is advised that the research community works together to develop standardized evaluation metrics. This collaborative effort could be enhanced by actively including discussions and presentations on these topics in existing conferences and congresses. By integrating these studies into broader scientific gatherings, it would promote consensus-building among researchers and clinicians, fostering the exchange of insights and best practices. Ultimately, the creation of baseline results and standardized evaluation metrics would streamline research efforts and markedly improve the reliability, generalizability, and clinical applicability of the findings.

The predominance of machine learning studies analyzing IMU data within controlled environments such as clinics and laboratories, which represents 77% of the research reviewed, highlights a critical focus on obtaining high-quality, precise data for specific disease studies. While beneficial for eliminating external variables that might affect the data's integrity, this approach contrasts with the unpredictable nature of daily life where wearable IMUs are most impactful. The real value of wearable IMUs lies in their potential to continuously monitor health conditions in natural settings, capturing nuanced data on disease manifestation and progression. However, the transition from controlled environments to real-life application faces significant challenges, primarily due to technological limitations such as inadequate battery life and insufficient data storage capacity in wearable devices. For example, the UK Biobank, which holds the largest dataset of accelerometer data from patients, used the Axivity AX3 (Axivity) to collect data at high frequency but was only able to record up to 7 days of data before the battery was depleted or the memory was full, and only recording accelerometer data and no gyroscope. This limitation restricts the potential for capturing valuable longer-term trends, such as detecting subtle changes in activity patterns or sleep disturbances that might occur over a longer period, which could be critical for early diagnosis of conditions like Parkinson disease or monitoring gradual recovery after surgery. These goes alongside difficulties ensuring patient compliance outside of clinical settings, where frequent charging and maintenance of the devices may disrupt consistent monitoring.

Addressing these challenges requires advancements in wearable technology to enhance device usability and reliability in everyday settings, alongside strategies to improve patient engagement and adherence to usage protocols. Enhancing battery life and expanding storage without compromising device comfort, coupled with the development of intuitive interfaces, could significantly boost patient compliance and data collection quality. Furthermore, leveraging machine learning algorithms to manage data and power more efficiently, and implementing patient education and motivation strategies, could bridge the gap between controlled studies and the dynamic context of real-world scenarios. By overcoming these technological and logistical problems, researchers can unlock deeper insights into health conditions through wearable IMUs, paving the way for more personalized and effective health care solutions that reflect the complexity of real-life patient experiences.

Finally, the fact that 64.7% of the studies focused on neurological conditions underscores a notable trend in the scientific community, particularly regarding the use of IMUs for analyzing movement disorders related to Parkinson disease. This focus is likely because neurological disorders exhibit distinct movement disorders, which can be readily detected and analyzed by Inertial IMUs and ML algorithms, leveraging their capacity to identify clear patterns in movement data. However, this concentration on neurology suggests an unintentional narrowness in application, potentially overlooking the vast possibilities in other medical fields.

Expanding the application of IMUs and ML/DL to include other medical areas, like musculoskeletal care, represents a promising avenue to harness the full potential of these technologies. To rectify this imbalance and encourage a broader application spectrum, targeted funding initiatives and interdisciplinary research collaborations could be established. These efforts would incentivize exploration into how IMUs and ML/DL can revolutionize diagnostics, treatment planning, and outcome monitoring. By diversifying research directions, the medical community can ensure a more equitable distribution of technological advancements across various domains, maximizing the impact of these cutting-edge tools in improving patient care and treatment outcomes.

Limitations

Despite its strengths, our systematic review has some limitations. While our strict inclusion criteria ensured high-quality research, it may have excluded some relevant studies. Furthermore, using an inclusion criterion of "study targets a clinical condition" will remove useful studies like sleep analysis or human activity recognition, but that is not the scope of this review. In addition, the variability in study quality is a common challenge in systematic reviews, and we prioritized datasets with a minimum number of sample size to increase confidence in our conclusions.

Conclusions

In conclusion, this systematic review has provided valuable insights into the current state of ML and DL applications on IMU data in the medical field. The key findings highlight several critical aspects: the relatively small median dataset size of 50.5 participants, the dominance of supervised ML methods, the exploratory nature of method selection in research, the preference for clinical settings over real-life data collection, and the focus on neurological problems. These findings suggest that while the integration of ML and DL in medical research is progressing, there is still considerable room for growth and optimization. The challenges identified, such as limited dataset sizes and the need for more real-world data, must be addressed to fully realize the potential of these technologies in health care.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Revues and Meta-Analyses) main and abstract checklists.
[PDF File (Adobe PDF File), 195 KB - mhealth_v13i1e60521_app1.pdf]

Multimedia Appendix 2
Description for all Artificial Intelligence Models reviewed.
[DOCX File , 19 KB - mhealth_v13i1e60521_app2.docx]

References

1. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022;28(1):31-38. [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
2. Wang C, He T, Zhou H, Zhang Z, Lee C. Artificial intelligence enhanced sensors - enabling technologies to next-generation healthcare and biomedical platform. Bioelectron Med 2023;9(1):17 [FREE Full text] [doi: 10.1186/s42234-023-00118-1] [Medline: 37528436]
3. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
4. Jimale AO, Mohd Noor MH. Subject variability in sensor-based activity recognition. J Ambient Intell Human Comput 2021;14(4):3261-3274. [doi: 10.1007/s12652-021-03465-6]
5. Gupta AS, Patel S, Premasiri A, Vieira F. At-home wearables and machine learning sensitively capture disease progression in amyotrophic lateral sclerosis. Nat Commun 2023;14(1):5080 [FREE Full text] [doi: 10.1038/s41467-023-40917-3] [Medline: 37604821]
6. Zhang Y, Wang H, Yao Y, Liu J, Sun X, Gu D. Walking stability in patients with benign paroxysmal positional vertigo: an objective assessment using wearable accelerometers and machine learning. J Neuroeng Rehabil 2021;18(1):56 [FREE Full text] [doi: 10.1186/s12984-021-00854-y] [Medline: 33789693]
7. Hope AJ, Vashisth U, Parker MJ, Ralston AB, Roper JM, Ralston JD. Phybrata sensors and machine learning for enhanced neurophysiological diagnosis and treatment. Sensors (Basel) 2021;21(21):7417 [FREE Full text] [doi: 10.3390/s21217417] [Medline: 34770729]
8. Tang J, El Atrache R, Yu S, Asif U, Jackson M, Roy S, et al. Seizure detection using wearable sensors and machine learning: Setting a benchmark. Epilepsia 2021;62(8):1807-1819 [FREE Full text] [doi: 10.1111/epi.16967] [Medline: 34268728]
9. Sigcha L, Costa N, Pavón I, Costa S, Arezes P, López JM, et al. Deep learning approaches for detecting freezing of gait in Parkinson's disease patients through on-body acceleration sensors. Sensors (Basel) 2020;20(7):1895 [FREE Full text] [doi: 10.3390/s20071895] [Medline: 32235373]
10. Aich S, Pradhan PM, Chakraborty S, Kim H, Kim H, Lee H, et al. Design of a machine learning-assisted wearable accelerometer-based automated system for studying the effect of dopaminergic medicine on gait characteristics of Parkinson's patients. J Healthc Eng 2020;2020:1823268 [FREE Full text] [doi: 10.1155/2020/1823268] [Medline: 32148741]
11. Abujrida H, Agu E, Pahlavan K. Machine learning-based motor assessment of Parkinson's disease using postural sway, gait and lifestyle features on crowdsourced smartphone data. Biomed Phys Eng Express 2020;6(3):035005. [doi: 10.1088/2057-1976/ab39a8] [Medline: 33438650]
12. Reches T, Dagan M, Herman T, Gazit E, Gouskova N, Giladi N, et al. Using wearable sensors and machine learning to automatically detect freezing of gait during a fog-provoking test. Sensors (Basel) 2020;20(16):4474 [FREE Full text] [doi: 10.3390/s20164474] [Medline: 32785163]
13. Johansson D, Thomas I, Ericsson A, Johansson A, Medvedev A, Memedi M, et al. Evaluation of a sensor algorithm for motor state rating in Parkinson's disease. Parkinsonism Relat Disord 2019;64:112-117. [doi: 10.1016/j.parkreldis.2019.03.022] [Medline: 30935826]
14. Hssayeni MD, Jimenez-Shahed J, Burack MA, Ghoraani B. Wearable sensors for estimation of Parkinsonian tremor severity during free body movements. Sensors (Basel) 2019;19(19):4215 [FREE Full text] [doi: 10.3390/s19194215] [Medline: 31569335]
15. Gubbi J, Kusmakar S, Rao AS, Yan B, OBrien T, Palaniswami M. Automatic detection and classification of convulsive psychogenic nonepileptic seizures using a wearable device. IEEE J Biomed Health Inform 2016;20(4):1061-1072. [doi: 10.1109/JBHI.2015.2446539] [Medline: 26087511]
16. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C. A smartphone-based tool for assessing Parkinsonian hand tremor. IEEE J Biomed Health Inform 2015;19(6):1835-1842. [doi: 10.1109/JBHI.2015.2471093] [Medline: 26302523]
17. Xia Y, Yao Z, Lu Y, Zhang D, Cheng N. A machine learning approach to detecting of freezing of gait in parkinson's disease patients. j med imaging hlth inform 2018;8(4):647-654. [doi: 10.1166/jmihi.2018.2379]

XSL•FO
RenderX

18. Camps J, Samà A, Martín M, Rodríguez-Martín D, Pérez-López C, Moreno Arostegui JM, et al. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. Knowl.-Based Syst 2018;139:119-131. [doi: 10.1016/j.knosys.2017.10.017]
19. Dimoudis D, Tsolakis N, Magga-Nteve C, Meditskos G, Vrochidis S, Kompatsiaris I. InSEption: a robust mechanism for predicting FoG episodes in PD patients. Electronics 2023;12(9):2088. [doi: 10.3390/electronics12092088]
20. Borzì L, Fornara S, Amato F, Olmo G, Artusi CA, Lopiano L. Smartphone-based evaluation of postural stability in Parkinson's disease patients during quiet stance. Electronics ETF 2020;9(6):919. [doi: 10.3390/electronics9060919]
21. Onorati F, Regalia G, Caborni C, Migliorini M, Bender D, Poh M, et al. Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors. Epilepsia 2017;58(11):1870-1879 [FREE Full text] [doi: 10.1111/epi.13899] [Medline: 28980315]
22. Channa A, Ifrim R, Popescu D, Popescu N. A-WEAR bracelet for detection of hand tremor and bradykinesia in Parkinson's patients. Sensors (Basel) 2021;21(3):981 [FREE Full text] [doi: 10.3390/s21030981] [Medline: 33540570]
23. Ramesh V, Bilal E. Detecting motor symptom fluctuations in Parkinson's disease with generative adversarial networks. NPJ Digit Med 2022;5(1):138. [doi: 10.1038/s41746-022-00674-x] [Medline: 36085350]
24. Williamson JR, Telfer B, Mullany R, Friedl KE. Detecting Parkinson's disease from wrist-worn accelerometry in the U.K. biobank. Sensors (Basel) 2021;21(6):2047 [FREE Full text] [doi: 10.3390/s21062047] [Medline: 33799420]
25. Zia S, Khan AN, Zaidi KS, Ali SE. Detection of generalized tonic clonic seizures and falls in unconstraint environment using smartphone accelerometer. IEEE Access 2021;9:39432-39443. [doi: 10.1109/ACCESS.2021.3063765]
26. Nguyen A, Roth N, Ghassemi NH, Hannink J, Seel T, Klucken J, et al. Development and clinical validation of inertial sensor-based gait-clustering methods in Parkinson's disease. J Neuroeng Rehabil 2019;16(1):77 [FREE Full text] [doi: 10.1186/s12984-019-0548-2] [Medline: 31242915]
27. Borzì L, Olmo G, Artusi CA, Lopiano L. Detection of freezing of gait in people with Parkinson's disease using smartphones. : IEEE; 2020 Presented at: IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC); July 13-17, 2020; Madrid, Spain. [doi: 10.1109/COMPSAC48688.2020.0-186]
28. Roth N, Küderle A, Ullrich M, Gladow T, Marxreiter F, Klucken J, et al. Hidden markov model based stride segmentation on unsupervised free-living gait data in Parkinson's disease patients. J Neuroeng Rehabil 2021;18(1):93 [FREE Full text] [doi: 10.1186/s12984-021-00883-7] [Medline: 34082762]
29. Sigcha L, Borzì L, Pavón I, Costa N, Costa S, Arezes P, et al. Improvement of performance in freezing of gait detection in parkinson's disease using transformer networks and a single waist-worn triaxial accelerometer. Eng. Appl. Artif. Intell 2022;116:105482. [doi: 10.1016/j.engappai.2022.105482]
30. Shi B, Tay A, Au WL, Tan DML, Chia NSY, Yen S. Detection of freezing of gait using convolutional neural networks and data from lower limb motion sensors. IEEE Trans Biomed Eng 2022;69(7):2256-2267. [doi: 10.1109/TBME.2022.3140258] [Medline: 34986092]
31. Cruz NE, Solarte J, Gonzalez-Vargas A. Automated epileptic seizure detection system based on a wearable prototype and cloud computing to assist people with epilepsy. In: Figueroa-García J, Villegas J, Orozco-Arroyave J, Maya Duque P, editors. Applied Computer Sciences in Engineering. WEA 2018. Communications in Computer and Information Science, vol 916. Switzerland: Springer, Cham; 2018.
32. Chitnis T, Glanz BI, Gonzalez C, Healy BC, Saraceno TJ, Sattarnezhad N, et al. Quantifying neurologic disease using biosensor measurements in-clinic and in free-living settings in multiple sclerosis. NPJ Digit Med 2019;2:123 [FREE Full text] [doi: 10.1038/s41746-019-0197-7] [Medline: 31840094]
33. Ma C, Zhang P, Wang J, Zhang J, Pan L, Li X, et al. Objective quantification of the severity of postural tremor based on kinematic parameters: a multi-sensory fusion study. Comput Methods Programs Biomed 2022;219:106741. [doi: 10.1016/j.cmpb.2022.106741] [Medline: 35338882]
34. Hssayeni MD, Jimenez-Shahed J, Burack MA, Ghoraani B. Ensemble deep model for continuous estimation of Unified Parkinson's disease rating scale III. Biomed Eng Online 2021;20(1):32 [FREE Full text] [doi: 10.1186/s12938-021-00872-w] [Medline: 33789666]
35. Park DJ, Lee JW, Lee MJ, Ahn SJ, Kim J, Kim GL, et al. Evaluation for Parkinsonian bradykinesia by deep learning modeling of kinematic parameters. J Neural Transm (Vienna) 2021;128(2):181-189. [doi: 10.1007/s00702-021-02301-7] [Medline: 33507401]
36. Ricci M, Di Lazzaro G, Pisani A, Mercuri NB, Giannini F, Saggio G. Assessment of motor impairments in early untreated parkinson's disease patients: the wearable electronics impact. IEEE J Biomed Health Inform 2020;24(1):120-130. [doi: 10.1109/JBHI.2019.2903627] [Medline: 30843855]
37. Lang M, Pfister FMJ, Frohner J, Abedinpour K, Pichler D, Fietzek U, et al. A multi-layer gaussian process for motor symptom estimation in people with parkinson's disease. IEEE Trans Biomed Eng 2019;66(11):3038-3049. [doi: 10.1109/TBME.2019.2900002] [Medline: 30794163]
38. Thomas I, Westin J, Alam M, Bergquist F, Nyholm D, Senek M, et al. A treatment-response index from wearable sensors for quantifying parkinson's disease motor states. IEEE J Biomed Health Inform 2018;22(5):1341-1349. [doi: 10.1109/JBHI.2017.2777926] [Medline: 29989996]

XSL•FO
RenderX

39. Kim HB, Lee WW, Kim A, Lee HJ, Park HY, Jeon HS, et al. Wrist sensor-based tremor severity quantification in Parkinson's disease using convolutional neural network. Comput Biol Med 2018;95:140-146. [doi: 10.1016/j.compbiomed.2018.02.007] [Medline: 29500984]
40. Jeon H, Lee W, Park H, Lee H, Kim S, Kim H, et al. Automatic classification of tremor severity in Parkinson's disease using a wearable device. Sensors (Basel) 2017;17(9):2067 [FREE Full text] [doi: 10.3390/s17092067] [Medline: 28891942]
41. Jeon H, Lee W, Park H, Lee HJ, Kim SK, Kim HB, et al. High-accuracy automatic classification of Parkinsonian tremor severity using machine learning method. Physiol Meas 2017;38(11):1980-1999. [doi: 10.1088/1361-6579/aa8e1f] [Medline: 28933707]
42. Pastorino M, Cancela J, Arredondo MT, Pansera M, Pastor-Sanz L, Villagra F, et al. Assessment of bradykinesia in Parkinson's disease patients through a multi-parametric system. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:1810-1813. [doi: 10.1109/IEMBS.2011.6090516] [Medline: 22254680]
43. Ma C, Li D, Pan L, Li X, Yin C, Li A, et al. Quantitative assessment of essential tremor based on machine learning methods using wearable device. Biomed Signal Process Control 2022;71:103244. [doi: 10.1016/j.bspc.2021.103244]
44. Wan S, Liang Y, Zhang Y, Guizani M. Deep multi-layer perceptron classifier for behavior analysis to estimate parkinson‘s disease severity using smartphones. IEEE Access 2018;6:36825-36833. [doi: 10.1109/ACCESS.2018.2851382]
45. Wang C, Peng L, Hou Z, Li Y, Tan Y, Hao H. A hierarchical architecture for multisymptom assessment of early parkinson's disease via wearable sensors. IEEE Trans. Cogn. Dev. Syst 2022;14(4):1553-1563. [doi: 10.1109/tcds.2021.3123157]
46. Grigoriou S, Espa E, Odin P, Timpka J, von Grothusen G, Jakobsson A, et al. Comparison of dyskinesia profiles after L-DOPA dose challenges with or without dopamine agonist coadministration. Neuropharmacology 2023;237:109630 [FREE Full text] [doi: 10.1016/j.neuropharm.2023.109630] [Medline: 37315840]
47. Rehman RZU, Zhou Y, Del Din S, Alcock L, Hansen C, Guan Y, et al. Gait analysis with wearables can accurately classify fallers from non-fallers: a step toward better management of neurological disorders. Sensors (Basel) 2020;20(23):6992 [FREE Full text] [doi: 10.3390/s20236992] [Medline: 33297395]
48. Lin S, Gao C, Li H, Huang P, Ling Y, Chen Z, et al. Wearable sensor-based gait analysis to discriminate early Parkinson's disease from essential tremor. J Neurol 2023;270(4):2283-2301 [FREE Full text] [doi: 10.1007/s00415-023-11577-6] [Medline: 36725698]
49. Borzì L, Mazzetta I, Zampogna A, Suppa A, Irrera F, Olmo G. Predicting axial impairment in Parkinson's disease through a single inertial sensor. Sensors (Basel) 2022;22(2):412 [FREE Full text] [doi: 10.3390/s22020412] [Medline: 35062375]
50. Parnandi A, Kaku A, Venkatesan A, Pandit N, Fokas E, Yu B, et al. Data-driven quantitation of movement abnormality after stroke. Bioengineering (Basel) 2023;10(6):648 [FREE Full text] [doi: 10.3390/bioengineering10060648] [Medline: 37370579]
51. Balachandar A, Algarni M, Oliveira L, Marsili L, Merola A, Sturchio A, et al. Are smartphones and machine learning enough to diagnose tremor? J Neurol 2022;269(11):6104-6115. [doi: 10.1007/s00415-022-11293-7] [Medline: 35861853]
52. Juutinen M, Wang C, Zhu J, Haladjian J, Ruokolainen J, Puustinen J, et al. Parkinson's disease detection from 20-step walking tests using inertial sensors of a smartphone: machine learning approach based on an observational case-control study. PLoS One 2020;15(7):e0236258 [FREE Full text] [doi: 10.1371/journal.pone.0236258] [Medline: 32701955]
53. Papadopoulos A, Iakovakis D, Klingelhoefer L, Bostantjopoulou S, Chaudhuri KR, Kyritsis K, et al. Unobtrusive detection of Parkinson's disease from multi-modal and in-the-wild sensor data using deep learning techniques. Sci Rep 2020;10(1):21370 [FREE Full text] [doi: 10.1038/s41598-020-78418-8] [Medline: 33288807]
54. Bringas S, Salomón S, Duque R, Lage C, Montaña JL. Alzheimer's disease stage identification using deep learning models. J Biomed Inform 2020;109:103514 [FREE Full text] [doi: 10.1016/j.jbi.2020.103514] [Medline: 32711124]
55. Shcherbak A, Kovalenko E, Somov A. Detection and classification of early stages of parkinson's disease through wearable sensors and machine learning. IEEE Trans. Instrum. Meas 2023;72:1-9. [doi: 10.1109/tim.2023.3284944]
56. Talitckii A, Anikina A, Kovalenko E, Shcherbak A, Mayora O, Zimniakova O, et al. Defining optimal exercises for efficient detection of parkinson‘s disease using machine learning and wearable sensors. IEEE Trans. Instrum. Meas 2021;70:1-10. [doi: 10.1109/tim.2021.3097857]
57. Milosevic M, Van de Vel A, Bonroy B, Ceulemans B, Lagae L, Vanrumste B, et al. Automated detection of tonic-clonic seizures using 3-D accelerometry and surface electromyography in pediatric patients. IEEE J Biomed Health Inform 2016;20(5):1333-1341. [doi: 10.1109/JBHI.2015.2462079] [Medline: 26241981]
58. Rastegari E, Ali H, Marmelat V. Detection of Parkinson's disease using wrist accelerometer data and passive monitoring. Sensors (Basel) 2022;22(23):9122 [FREE Full text] [doi: 10.3390/s22239122] [Medline: 36501823]
59. Belić M, Radivojević Z, Bobić V, Kostić V, Đurić-Jovičić M. Quick computer aided differential diagnostics based on repetitive finger tapping in Parkinson's disease and atypical parkinsonisms. Heliyon 2023;9(4):e14824 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e14824] [Medline: 37077676]
60. Schalkamp A, Peall KJ, Harrison NA, Sandor C. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. Nat Med 2023;29(8):2048-2056. [doi: 10.1038/s41591-023-02440-2] [Medline: 37400639]
61. Kovalenko E, Shcherbak A, Somov A, Bril E, Zimniakova O, Semenov M, et al. Detecting the parkinson’s disease through the simultaneous analysis of data from wearable sensors and video. IEEE Sensors J 2022;22(16):16430-16439. [doi: 10.1109/jsen.2022.3191864]

XSL•FO
RenderX

62. Skaramagkas V, Andrikopoulos G, Kefalopoulou Z, Polychronopoulos P. Towards differential diagnosis of essential and parkinson's tremor via machine learning. : IEEE; 2020 Presented at: 28th Mediterranean Conference on Control and Automation (MED); September 15-18, 2020; Saint-Raphaël, France. [doi: 10.1109/MED48518.2020.9182922]
63. Shin H. Deep convolutional neural network-based hemiplegic gait detection using an inertial sensor located freely in a pocket. Sensors (Basel) 2022;22(5):1920 [FREE Full text] [doi: 10.3390/s22051920] [Medline: 35271066]
64. Kim Y, Joa K, Jeong H, Lee S. Wearable IMU-based human activity recognition algorithm for clinical balance assessment using 1D-CNN and GRU ensemble model. Sensors (Basel) 2021;21(22):7628 [FREE Full text] [doi: 10.3390/s21227628] [Medline: 34833704]
65. Wasselius J, Lyckegård Finn E, Persson E, Ericson P, Brogårdh C, Lindgren AG, et al. Detection of unilateral arm paresis after stroke by wearable accelerometers and machine learning. Sensors (Basel) 2021;21(23):7784 [FREE Full text] [doi: 10.3390/s21237784] [Medline: 34883800]
66. Park E, Chang H, Nam HS. Use of machine learning classifiers and sensor data to detect neurological deficit in stroke patients. J Med Internet Res 2017;19(4):e120 [FREE Full text] [doi: 10.2196/jmir.7092] [Medline: 28420599]
67. Ibrahim AA, Küderle A, Gaßner H, Klucken J, Eskofier BM, Kluge F. Inertial sensor-based gait parameters reflect patient-reported fatigue in multiple sclerosis. J Neuroeng Rehabil 2020;17(1):165 [FREE Full text] [doi: 10.1186/s12984-020-00798-9] [Medline: 33339530]
68. da Rosa Tavares JE, Ullrich M, Roth N, Kluge F, Eskofier BM, Gaßner H, et al. uTUG: an unsupervised timed up and go test for parkinson's disease. Biomed Signal Process Control 2023;81:104394. [doi: 10.1016/j.bspc.2022.104394]
69. Atrsaei A, Dadashi F, Mariani B, Gonzenbach R, Aminian K. Toward a remote assessment of walking bout and speed: application in patients with multiple sclerosis. IEEE J Biomed Health Inform 2021;25(11):4217-4228. [doi: 10.1109/JBHI.2021.3076707] [Medline: 33914688]
70. Li Y, Zhang X, Gong Y, Cheng Y, Gao X, Chen X. Motor function evaluation of hemiplegic upper-extremities using data fusion from wearable inertial and surface EMG sensors. Sensors (Basel) 2017;17(3):582 [FREE Full text] [doi: 10.3390/s17030582] [Medline: 28335394]
71. Meyer BM, Tulipani LJ, Gurchiek RD, Allen DA, Adamowicz L, Larie D, et al. Wearables and deep learning classify fall risk from gait in multiple sclerosis. IEEE J Biomed Health Inform 2021;25(5):1824-1831 [FREE Full text] [doi: 10.1109/JBHI.2020.3025049] [Medline: 32946403]
72. Aich S, Pradhan PM, Park J, Sethi N, Vathsa VSS, Kim H. A validation study of freezing of gait (FoG) detection and machine-learning-based fog prediction using estimated gait characteristics with a wearable accelerometer. Sensors (Basel) 2018;18(10):3287 [FREE Full text] [doi: 10.3390/s18103287] [Medline: 30274340]
73. Jeon Y, Kang J, Kim BC, Lee KH, Song J, Gwak J. Early alzheimer's disease diagnosis using wearable sensors and multilevel gait assessment: a machine learning ensemble approach. IEEE Sensors J 2023;99:1. [doi: 10.1109/jsen.2023.3259034]
74. Ullrich M, Roth N, Kuderle A, Richer R, Gladow T, Gasner H, et al. Fall risk prediction in Parkinson's disease using real-world inertial sensor gait data. IEEE J Biomed Health Inform 2023;27(1):319-328. [doi: 10.1109/JBHI.2022.3215921] [Medline: 36260566]
75. Igene L, Alim A, Imtiaz MH, Schuckers S. A machine learning model for early prediction of parkinson's disease from wearable sensors. : IEEE; 2023 Presented at: IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC); March 08-11, 2023; Las Vegas, NV, USA. [doi: 10.1109/CCWC57344.2023.10099230]
76. Thakur D, Biswas S. Attention-based deep learning framework for hemiplegic gait prediction with smartphone sensors. IEEE Sensors J 2022;22(12):11979-11988. [doi: 10.1109/jsen.2022.3172603]
77. O'Brien MK, Shin SY, Khazanchi R, Fanton M, Lieber RL, Ghaffari R, et al. Wearable sensors improve prediction of post-stroke walking function following inpatient rehabilitation. IEEE J Transl Eng Health Med 2022;10:2100711 [FREE Full text] [doi: 10.1109/JTEHM.2022.3208585] [Medline: 36304845]
78. Chae SH, Kim Y, Lee K, Park H. Development and clinical evaluation of a web-based upper limb home rehabilitation system using a smartwatch and machine learning model for chronic stroke survivors: prospective comparative study. JMIR Mhealth Uhealth 2020;8(7):e17216 [FREE Full text] [doi: 10.2196/17216] [Medline: 32480361]
79. Yu L, Xiong D, Guo L, Wang J. A remote quantitative Fugl-Meyer assessment framework for stroke patients based on wearable sensor networks. Comput Methods Programs Biomed 2016;128:100-110. [doi: 10.1016/j.cmpb.2016.02.012] [Medline: 27040835]
80. Ernesto C, Parisi F, Adans-Dester C, O'Brien A, Vergara-Diaz G, Black-Schaffer R. Wearable technology and machine learning to monitor upper-limb use in brain injury survivors. 2022 Presented at: IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); 2022 November 17-19; Arlington, VA, USA.
81. Baraka A, Shaban H, Abou El-Nasr M, Attallah O. Wearable accelerometer and sEMG-based upper limb BSN for tele-rehabilitation. Appl. Sci 2019;9(14):2795. [doi: 10.3390/app9142795]
82. Lee SI, Adans-Dester CP, Grimaldi M, Dowling AV, Horak PC, Black-Schaffer RM, et al. Enabling stroke rehabilitation in home and community settings: a wearable sensor-based approach for upper-limb motor training. IEEE J Transl Eng Health Med 2018;6:2100411 [FREE Full text] [doi: 10.1109/JTEHM.2018.2829208] [Medline: 29795772]

XSL•FO
RenderX

83. Lee SI, Adans-Dester CP, OBrien AT, Vergara-Diaz GP, Black-Schaffer R, Zafonte R, et al. Predicting and monitoring upper-limb rehabilitation outcomes using clinical and wearable sensor data in brain injury survivors. IEEE Trans Biomed Eng 2021;68(6):1871-1881 [FREE Full text] [doi: 10.1109/TBME.2020.3027853] [Medline: 32997621]
84. Greene BR, McManus K, Ader LGM, Caulfield B. Unsupervised assessment of balance and falls risk using a smartphone and machine learning. Sensors (Basel) 2021;21(14):4770 [FREE Full text] [doi: 10.3390/s21144770] [Medline: 34300509]
85. Buisseret F, Catinus L, Grenard R, Jojczyk L, Fievez D, Barvaux V, et al. Timed up and go and six-minute walking tests with wearable inertial sensor: one step further for the prediction of the risk of fall in elderly nursing home people. Sensors (Basel) 2020;20(11):3207 [FREE Full text] [doi: 10.3390/s20113207] [Medline: 32516995]
86. Giansanti D, Maccioni G, Cesinaro S, Benvenuti F, Macellari V. Assessment of fall-risk by means of a neural network based on parameters assessed by a wearable device during posturography. Med Eng Phys 2008;30(3):367-372. [doi: 10.1016/j.medengphy.2007.04.006] [Medline: 17560825]
87. Drover D, Howcroft J, Kofman J, Lemaire E. Faller classification in older adults using wearable sensors based on turn and straight-walking accelerometer-based features. Sensors (Basel) 2017;17(6):1321 [FREE Full text] [doi: 10.3390/s17061321] [Medline: 28590432]
88. Howcroft J, Kofman J, Lemaire ED. Prospective fall-risk prediction models for older adults based on wearable sensors. IEEE Trans Neural Syst Rehabil Eng 2017;25(10):1812-1820. [doi: 10.1109/TNSRE.2017.2687100] [Medline: 28358689]
89. Jung D, Kim J, Kim M, Won CW, Mun K. Classifying the risk of cognitive impairment using sequential gait characteristics and long short-term memory networks. IEEE J Biomed Health Inform 2021;25(10):4029-4040. [doi: 10.1109/JBHI.2021.3073372] [Medline: 33857005]
90. Marschollek M, Wolf KH, Gietzelt M, Nemitz G, Meyer zu Schwabedissen H, Haux R. Assessing elderly persons' fall risk using spectral analysis on accelerometric data--a clinical evaluation study. Annu Int Conf IEEE Eng Med Biol Soc 2008;2008:3682-3685. [doi: 10.1109/IEMBS.2008.4650008] [Medline: 19163511]
91. Lin B, Zhang Z, Peng C, Lin C, Lin C, Lin B. Automatic assessment system based on IMUs and machine learning for predicting berg balance scale. IEEE Sensors J 2022;22(20):19919-19930. [doi: 10.1109/jsen.2022.3200986]
92. Hellmers S, Izadpanah B, Dasenbrock L, Diekmann R, Bauer J, Hein A, et al. Towards an automated unsupervised mobility assessment for older people based on inertial TUG measurements. Sensors (Basel) 2018;18(10):3310 [FREE Full text] [doi: 10.3390/s18103310] [Medline: 30279374]
93. KIM YW, CHO WH, JOA KL, JUNG HY, LEE S. A new auto-scoring algorithm for balance assessment with wearable IMU device based on nonlinear model. J. Mech. Med. Biol 2020;20(10):2040011. [doi: 10.1142/s0219519420400114]
94. Kim JK, Bae MN, Lee K, Kim JC, Hong SG. Explainable artificial intelligence and wearable sensor-based gait analysis to identify patients with osteopenia and sarcopenia in daily life. Biosensors (Basel) 2022;12(3):167 [FREE Full text] [doi: 10.3390/bios12030167] [Medline: 35323437]
95. Kroll F, Löffler S, Becker I, Hofstedt P. Automatic detection of timed-up and go tests with IMU sensor data. 2022 Presented at: Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 1: BIOSTEC; Jan 01, 2022; Setúbal, Portugal. [doi: 10.5220/0010775300003123]
96. Lahdenoja O, Hurnanen T, Iftikhar Z, Nieminen S, Knuutila T, Saraste A, et al. Atrial fibrillation detection via accelerometer and gyroscope of a smartphone. IEEE J Biomed Health Inform 2018;22(1):108-118. [doi: 10.1109/JBHI.2017.2688473] [Medline: 28391210]
97. Yang C, Aranoff ND, Green P, Tavassolian N. Classification of aortic stenosis using time-frequency features from chest cardio-mechanical signals. IEEE Trans Biomed Eng 2020;67(6):1672-1683. [doi: 10.1109/TBME.2019.2942741] [Medline: 31545706]
98. Shokouhmand A, Aranoff ND, Driggin E, Green P, Tavassolian N. Efficient detection of aortic stenosis using morphological characteristics of cardiomechanical signals and heart rate variability parameters. Sci Rep 2021;11(1):23817 [FREE Full text] [doi: 10.1038/s41598-021-03441-2] [Medline: 34893693]
99. Benedetti D, Olcese U, Bruno S, Barsotti M, Maestri Tassoni M, Bonanni E, et al. Obstructive sleep apnoea syndrome screening through wrist-worn smartbands: a machine-learning approach. Nat Sci Sleep 2022;14:941-956 [FREE Full text] [doi: 10.2147/NSS.S352335] [Medline: 35611177]
100. Koivisto T, Lahdenoja O, Hurnanen T, Vasankari T, Jaakkola S, Kiviniemi T, et al. Mechanocardiography in the detection of acute ST elevation myocardial infarction: the MECHANO-STEMI study. Sensors (Basel) 2022;22(12):4384 [FREE Full text] [doi: 10.3390/s22124384] [Medline: 35746166]
101. Shah AJ, Isakadze N, Levantsevych O, Vest A, Clifford G, Nemati S. Detecting heart failure using wearables: a pilot study. Physiol Meas 2020;41(4):044001 [FREE Full text] [doi: 10.1088/1361-6579/ab7f93] [Medline: 32163936]
102. Mehrang S, Jafari Tadi M, Knuutila T, Jaakkola J, Jaakkola S, Kiviniemi T, et al. End-to-end sensor fusion and classification of atrial fibrillation using deep neural networks and smartphone mechanocardiography. Physiol Meas 2022;43(5). [doi: 10.1088/1361-6579/ac66ba] [Medline: 35413698]
103. Lee H, Whang M. Heart rate estimated from body movements at six degrees of freedom by convolutional neural networks. Sensors (Basel) 2018;18(5):1392 [FREE Full text] [doi: 10.3390/s18051392] [Medline: 29724006]
104. Chatterjee S, Rahman MM, Nemati E, Nathan V, Vatanparvar K, Kuang J. mLung++: automated characterization of abnormal lung sounds in pulmonary patients using multimodal mobile sensors. : ACM Digital Library; 2019 Presented at:

XSL•FO
RenderX

Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers; September 09,2019; London, United Kingdom. [doi: 10.1145/3341162.3344850]
105. Mehrang S, Airaksinen J, Koivisto T, Pankaala M, Tadi MJ, Hurnanen T, et al. Reliability of self-applied smartphone mechanocardiography for atrial fibrillation detection. IEEE Access 2019;7:146801-146812. [doi: 10.1109/access.2019.2946117]
106. De Cannière H, Corradi F, Smeets CJP, Schoutteten M, Varon C, Van Hoof C, et al. Wearable monitoring and interpretable machine learning can objectively track progression in patients during cardiac rehabilitation. Sensors (Basel) 2020;20(12):3601 [FREE Full text] [doi: 10.3390/s20123601] [Medline: 32604829]
107. Sharifi Renani M, Myers CA, Zandie R, Mahoor MH, Davidson BS, Clary CW. Deep learning in gait parameter prediction for OA and TKA patients wearing IMU sensors. Sensors (Basel) 2020;20(19):5553 [FREE Full text] [doi: 10.3390/s20195553] [Medline: 32998329]
108. Wang C, Chan PPK, Lam BMF, Wang S, Zhang JH, Chan ZYS, et al. Real-time estimation of knee adduction moment for gait retraining in patients with knee osteoarthritis. IEEE Trans Neural Syst Rehabil Eng 2020;28(4):888-894. [doi: 10.1109/TNSRE.2020.2978537] [Medline: 32149643]
109. Chen C, Liu K, Lu T, Chang C, Chan C, Tsao Y. Wearable-based pain assessment in patients with adhesive capsulitis using machine learning. 2023 Presented at: 11th International IEEE/EMBS Conference on Neural Engineering (NER); April 24-27, 2023; Baltimore, MD, USA. [doi: 10.1109/NER52421.2023.10123790]
110. Teufl W, Taetz B, Miezal M, Lorenz M, Pietschmann J, Jöllenbeck T, et al. Towards an inertial sensor-based wearable feedback system for patients after total hip arthroplasty: validity and applicability for gait classification with gait kinematics-based features. Sensors (Basel) 2019;19(22):5006 [FREE Full text] [doi: 10.3390/s19225006] [Medline: 31744141]
111. Kyritsis AI, Willems G, Deriaz M, Konstantas D. Gait recognition with smart devices assisting postoperative rehabilitation in a clinical setting. 2018 Presented at: First International Conference on Artificial Intelligence for Industries (AI4I); September 26-28, 2018; Laguna Hills, CA, USA. [doi: 10.1109/AI4I.2018.8665677]
112. Nazarahari M, Chan KM, Rouhani H. A novel instrumented shoulder functional test using wearable sensors in patients with brachial plexus injury. J Shoulder Elbow Surg 2021;30(8):e493-e502. [doi: 10.1016/j.jse.2020.10.025] [Medline: 33246080]
113. Abdollahi M, Ashouri S, Abedi M, Azadeh-Fard N, Parnianpour M, Khalaf K, et al. Using a motion sensor to categorize nonspecific low back pain patients: a machine learning approach. Sensors (Basel) 2020;20(12):3600 [FREE Full text] [doi: 10.3390/s20123600] [Medline: 32604794]
114. Ahmadi MN, O'Neil ME, Baque E, Boyd RN, Trost SG. Machine learning to quantify physical activity in children with cerebral palsy: comparison of group, group-personalized, and fully-personalized activity classification models. Sensors (Basel) 2020;20(14):3976 [FREE Full text] [doi: 10.3390/s20143976] [Medline: 32708963]
115. Ahmadi M, O'Neil M, Fragala-Pinkham M, Lennon N, Trost S. Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy. J Neuroeng Rehabil 2018;15(1):105 [FREE Full text] [doi: 10.1186/s12984-018-0456-x] [Medline: 30442154]
116. Muñoz-Organero M, Powell L, Heller B, Harpin V, Parker J. Using recurrent neural networks to compare movement patterns in ADHD and normally developing children based on acceleration signals from the wrist and ankle. Sensors (Basel) 2019;19(13):2935 [FREE Full text] [doi: 10.3390/s19132935] [Medline: 31277297]
117. Soangra R, Wen Y, Yang H, Grant-Beuttler M. Classifying toe walking gait patterns among children diagnosed with idiopathic toe walking using wearable sensors and machine learning algorithms. IEEE Access 2022;10:77054-77067. [doi: 10.1109/access.2022.3192136]
118. Ko Y, Kuo P, Wang C, Chen Y, Chuang P, Li S, et al. Quantification analysis of sleep based on smartwatch sensors for Parkinson's disease. Biosensors (Basel) 2022;12(2):74 [FREE Full text] [doi: 10.3390/bios12020074] [Medline: 35200335]
119. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
120. Jeon S, Lee Y, Son SH. Cascade windows-based multi-stream convolutional neural networks framework for early detecting in-sleep stroke using wristbands. IEEE Access 2023;11:84944-84956. [doi: 10.1109/access.2023.3301872]
121. Fozoonmayeh D, Le HV, Wittfoth E, Geng C, Ha N, Wang J, et al. A scalable smartwatch-based medication intake detection system using distributed machine learning. J Med Syst 2020;44(4):76. [doi: 10.1007/s10916-019-1518-8] [Medline: 32112271]
122. Yang TY, Kuo P, Huang Y, Lin H, Malwade S, Lu L, et al. Deep-learning approach to predict survival outcomes using wearable actigraphy device among end-stage cancer patients. Front Public Health 2021;9:730150 [FREE Full text] [doi: 10.3389/fpubh.2021.730150] [Medline: 34957004]
123. Aggelides X, Bardoutsos A, Nikoletseas S, Papadopoulos N, Raptopoulos C, Tzamalis P. A gesture recognition approach to classifying allergic rhinitis gestures using wrist-worn devices a multidisciplinary case study. : IEEE; 2020 Presented at: 16th International Conference on Distributed Computing in Sensor Systems (DCOSS); May 25-27, 2020; Marina del Rey, CA, USA. [doi: 10.1109/DCOSS49796.2020.00015]

XSL•FO
RenderX

124. Hafezi M, Montazeri N, Saha S, Zhu K, Gavrilovic B, Yadollahi A, et al. Sleep apnea severity estimation from tracheal movements using a deep learning model. IEEE Access 2020;8:22641-22649. [doi: 10.1109/access.2020.2969227]
125. Nguyen TQ, Young JH, Rodriguez A, Zupancic S, Lie DYC. Differentiation of patients with balance insufficiency (vestibular hypofunction) versus normal subjects using a low-cost small wireless wearable gait sensor. Biosensors (Basel) 2019;9(1):29 [FREE Full text] [doi: 10.3390/bios9010029] [Medline: 30813585]
126. Kim J, Park G, Lee S, Nam Y. Analysis of machine learning-based assessment for elbow spasticity using inertial sensors. Sensors (Basel) 2020;20(6):1622 [FREE Full text] [doi: 10.3390/s20061622] [Medline: 32183281]
127. Sun C, Shrivastava A, Singh S, Gupta A. Revisiting unreasonable effectiveness of data in deep learning Era. arXiv:1707.02968 2017.
128. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
129. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging data for machine learning. Radiology 2020;295(1):4-15 [FREE Full text] [doi: 10.1148/radiol.2020192224] [Medline: 32068507]

Abbreviations

AI: artificial intelligence
CNN: convolutional neural network
DL: deep learning
IMU: inertial measurement unit
LSTM: long short-term memory
ML: machine learning
MLP: multilayer perceptron
NN: neural network
PRISMA: Preferred Reporting Items for Systematic Revues and Meta-Analyses

Edited by D Cox; submitted 14.05.24; peer-reviewed by C Sakal, G Vos; comments to author 18.08.24; revised version received 20.10.24; accepted 12.11.24; published 29.01.25.

Please cite as:
Smits Serena R, Hinterwimmer F, Burgkart R, von Eisenhart-Rothe R, Rueckert D
The Use of Artificial Intelligence and Wearable Inertial Measurement Units in Medicine: Systematic Review
JMIR Mhealth Uhealth 2025;13:e60521
URL: https://mhealth.jmir.org/2025/1/e60521
doi:10.2196/60521
PMID:

©Ricardo Smits Serena, Florian Hinterwimmer, Rainer Burgkart, Rudiger von Eisenhart-Rothe, Daniel Rueckert. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Applying AI in the Context of the Association Between Device-Based Assessment of Physical Activity and Mental Health: Systematic Review

Simon Woll[1], MSc; Dennis Birkenmaier[2], MSc; Gergely Biri[2], MSc; Rebecca Nissen[1], MSc; Luisa Lutz[1], BA; Marc Schroth[2], MSc; Ulrich W Ebner-Priemer[1,3,4], Prof Dr; Marco Giurgiu[1], Dr

[1]Mental mHealth Lab, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

[2]Department of Embedded Systems and Sensors Engineering, Research Center for Information Technology, Karlsruhe, Germany

[3]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[4]German Center for Mental Health, Mannheim, Germany

Corresponding Author:
Simon Woll, MSc
Mental mHealth Lab
Institute of Sports and Sports Science
Karlsruhe Institute of Technology
Hertzstrasse 16
Karlsruhe, 76187
Germany
Phone: 49 721 608 ext 41974
Email: simon.woll@kit.edu

Abstract

Background: Wearable technology is used by consumers worldwide for continuous activity monitoring in daily life but more recently also for classifying or predicting mental health parameters like stress or depression levels. Previous studies identified, based on traditional approaches, that physical activity is a relevant factor in the prevention or management of mental health. However, upcoming artificial intelligence methods have not yet been fully established in the research field of physical activity and mental health.

Objective: This systematic review aims to provide a comprehensive overview of studies that integrated passive monitoring of physical activity data measured via wearable technology in machine learning algorithms for the detection, prediction, or classification of mental health states and traits.

Methods: We conducted a review of studies processing wearable data to gain insights into mental health parameters. Eligibility criteria were (1) the study uses wearables or smartphones to acquire physical behavior and optionally other sensor measurement data, (2) the study must use machine learning to process the acquired data, and (3) the study had to be published in a peer-reviewed English language journal. Studies were identified via a systematic search in 5 electronic databases.

Results: Of 11,057 unique search results, 49 published papers between 2016 and 2023 were included. Most studies examined the connection between wearable sensor data and stress (n=15, 31%) or depression (n=14, 29%). In total, 71% (n=35) of the studies had less than 100 participants, and 47% (n=23) had less than 14 days of data recording. More than half of the studies (n=27, 55%) used step count as movement measurement, and 44% (n=21) used raw accelerometer values. The quality of the studies was assessed, scoring between 0 and 18 points in 9 categories (maximum 2 points per category). On average, studies were rated 6.47 (SD 3.1) points.

Conclusions: The use of wearable technology for the detection, prediction, or classification of mental health states and traits is promising and offers a variety of applications across different settings and target groups. However, based on the current state of literature, the application of artificial intelligence cannot realize its full potential mostly due to a lack of methodological shortcomings and data availability. Future research endeavors may focus on the following suggestions to improve the quality of new applications in this context: first, by using raw data instead of already preprocessed data. Second, by using only relevant data based on empirical evidence. In particular, crafting optimal feature sets rather than using many individual detached features and consultation with in-field professionals. Third, by validating and replicating the existing approaches (ie, applying the model

XSL•FO
RenderX

to unseen data). Fourth, depending on the research aim (ie, generalization vs personalization) maximizing the sample size or the duration over which data are collected.

(JMIR Mhealth Uhealth 2025;13:e59660) doi:10.2196/59660

KEYWORDS

machine learning; mental health; wearables; physical behavior; artificial intelligence; mobile phone; smartphone

Introduction

Background

Mental disorders such as depression, anxiety, or bipolar disorder remain among the top 10 leading causes of burden worldwide [1], with an estimated number of 970 million people in the world living with a mental disorder and a steady prevalence rate of around 13% [2]. The economic burden is estimated at US $5 trillion [3]. Moreover, global changes such as the COVID-19 pandemic may further boost the prevalence rate of mental health disorders [4]. Thus, the prevention and management of mental health are urgent topics. Researchers identified that social determinants of health represent the most modifiable set of targets for nonpharmacological intervention to prevent the onset of mental health disorders. One of these modifiable lifestyle behaviors that can play an evident role in the prevention and management of mental health is being physically active and performing regular exercise in daily life [5]. Earlier meta-analysis indicated that regular aerobic exercise results in moderate increases in self-reported affect [5,6]. A further review showed that the association between physical activity and mental health in young people is evident with small to moderate effects [7].

Previous studies applied traditional assessment methods (eg, self-reports and questionnaires) and analytical approaches (eg, regression analyses and 1-tailed *t* tests) and were able to unveil associations between physical activity and mental health outcomes retrospectively. A methodological aspect of earlier studies that display weaker research designs is the application of self-reports for the assessment of physical activity. Prince et al [8] compared self-reports and direct measures of physical activity across 148 studies and concluded that the measurement method may have a significant impact on the observed levels of physical activity. In particular, for example, due to retrospective biases, self-report measures of physical activity were both higher and lower than directly measured levels of physical activity. In line with technological developments, previous studies intensively used accelerometers as a passive monitoring device for capturing physical activity. The purpose of their application is variable, such as an observational tool in surveillance studies [9]; a motivational tool in interventions [10]; a way to better understand the underlying mechanisms of health, treatment, and recovery (ie, understanding the role of physical behavior in it); or as a diagnostic tool in clinical settings [11]. Using a device-based assessment of physical activity, researchers replicated the positive impact of physical activity on various mental health outcomes. For example, a meta-analysis found that higher levels of physical activity may offer prevention against the onset of depression [12], stress-related disorders [13], and psychotic disorders [14]. Furthermore, studies also indicated preliminary results that being physically inactive in terms of spending higher time in sedentary behavior is associated with an increased risk of various mental disorders [15-17]. In addition, not only observational studies but also results of randomized controlled trials have shown that physical activity interventions may reduce mental health symptoms among individuals affected by depression, stress-related disorders, and schizophrenia [18]. Most of those studies applied traditional statistical approaches, for example, analyzing the association between sedentary behavior and mood dimensions via multilevel modeling [19]. However, those approaches are limited in predicting, classifying, or detecting future conditions.

Nowadays, the ability to collect raw accelerometer data without great effort (eg, via passive continuous longitudinal measurements from wearables) and in large quantities allows the application of sophisticated artificial intelligence algorithms for the detection, classification, or prediction of mental health states and traits. In this context, the widely used term “machine learning” refers to a range of mathematical techniques that leverage the computational power to identify meaningful patterns within a large dataset [20]. The application of machine learning algorithms comprises a series of iterative processes and phases from learning and training within a single original dataset over testing in a new dataset. The use of artificial intelligence can be seen as a new methodological area with increasing awareness in health science but not yet established in clinical research and practice. Contrary to traditional approaches, which account for patterns in an original dataset, machine learning algorithms evaluate how accurately they can predict patterns and relationships in new datasets [20]. Furthermore, machine learning algorithms emphasize multidimensionality by capturing the complex interactions between multiple potential predictors and the outcome of interest. The application of machine learning can serve different purposes—some models are trained on groups to provide general predictions, classifications, or detections applicable to all individuals (ie, generalizability), while others are personalized, trained on an individual’s data to generate tailored predictions specific to that person [21].

Comparing traditional retrospective and modern analytical methods (eg, machine learning), we see a gap in both the possibilities and the performance. Passive measurement methods in the use of wearables have become established in research studies and make it possible to record enormous high-resolution datasets in everyday life. However, it is still unclear how the potential can be optimally used on the basis of new evaluation approaches. Machine learning enables new strategies to analyze huge amounts of data compared to traditional methods (eg, retrospective vs real-time analysis of data, analysis of new

XSL•FO
RenderX

datasets with already existing models without new training in less time, and possible forecasting of future trends and events). To get an overview of the status quo, we summarized studies that integrated passive monitoring of physical activity data measured via wearable technology in machine learning algorithms for the detection, prediction, or classification of mental health states and traits.

Objective

Although the application of machine learning approaches in health science increased over the last years [22], we are not aware of a previous work that integrated passive monitoring of physical activity data measured via wearable technology in machine learning algorithms for the detection, prediction, or classification of mental health states and traits. Thus, the aim of the review is (1) to provide a comprehensive overview of studies that combined machine learning approaches based on accelerometer measures with mental health outcomes; (2) to summarize the results of the selected studies, the applied machine learning methods, and the used wearable technologies; and (3) to rank and discuss quality aspects of the studies.

Methods

Overview

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines [23] and was registered in the PROSPERO international prospective register of systematic reviews (CRD42023436926). The PRISMA checklist can be found in Multimedia Appendix 1.

Search Strategy and Study Selection

To identify relevant publications, we used a search string that included terms for (1) machine learning, (2) physical activity, and (3) wearables (Multimedia Appendix 2). Publications were searched from 1970 to March 2023 using the following databases: EBSCOhost, IEEE Xplore, PubMed, Scopus, and Web of Science.

All papers were imported to a reference manager, Citavi library (Citavi, version 6.14; Swiss Academic Software GmbH). After merging all duplicates first electronically and afterward manually, the study selection process included 3 screening phases for eligibility. In the first phase, a minimum of 2 reviewers (SW, MG, and RN) independently screened the titles of the publications. Papers were only excluded if all reviewers categorized a paper as not eligible for review purposes. In the second phase, a minimum of 2 reviewers independently screened the publications' abstracts (SW, MG, and RN) to determine whether a full-text review was warranted. Only those papers moved to the next phase if a minimum of 1 reviewer (SW, MG, or RN) categorized it as eligible. In the third phase, the full texts of the remaining papers were assessed for eligibility by 4 reviewers (SW, MG, RN, and LL). Each paper was screened independently by at least 2 reviewers (SW, MG, RN, or LL). Discrepancies in screening were resolved by a third review and if unanimously a fourth review. If there was still no agreement, the papers were discussed until a consensus was reached. There was direct agreement (after 2 ratings) in 52% (73/140) of the studies, a third rating was necessary in 29% (41/140) of the studies, and a fourth rating was needed in 19% (26/140) of the studies. After the fourth rating, 3 studies needed further discussion to decide whether to consider the papers in the review. Reviewers were not blinded to author or journal information.

Inclusion and Exclusion Criteria

We included peer-reviewed, English-language publications that met the following criteria: first, studies were conducted under real-life scenarios. Second, raw accelerometer measures were collected with wearables or smartphones. Third, machine learning methods were used to process the data. Fourth, input for the machine learning algorithm must include physical activity measurements. Fifth, the outcome must be mental health–related. All other papers were excluded, for example, technical descriptions and protocols, studies with synthetically generated data, studies with animals, or not mental health–related such as pure human activity recognition.

Data Extraction

In total, 3 authors (SW, RN, and LL) independently extracted data. For every study, one author (SW, RN, or LL) extracted data from the paper, and another author (SW, RN, or LL) complemented the extracted data with additional information. Occurring discrepancies were discussed until a consensus was reached. The following study details were extracted: author, year, location, population information (sample size, mean age of participants, percentage of male population, ethnicity, and special population group), study protocol (measurement period, type, and environment), study purpose (mental health outcome), used data (wearable, wearing position, physical activity recording, and additional sensors used), machine learning (algorithms, software, feature description, and classifier performance), study conclusion, funding, and conflict of interest information.

Quality Assessment and Feasibility

Since no quality assessment tool was found that covered all quality aspects of interest, we modified a previously used version [24] based on a combination of the appraisal tool for cross-sectional studies and the Newcastle-Ottawa Scale for longitudinal studies. We adopted the mentioned quality assessment tools and modified them to fit our specific purposes by complementing them with our own categories and items. A total of 9 categories were selected, and each category scored between 0 and 2 points, resulting in a maximum achievable score of 18 points. In particular, we defined scoring rules to get points (Multimedia Appendix 3). For example, in the category "missing data management and dataset balancing" studies received 0 points if no information was provided, 1 point if the handling of missing data is described, and a further point if an imbalanced dataset problem is addressed. Finally, we split the points into 3 classes: low (0-6 points), moderate (7-12 points), and high (13-18 points).

XSL•FO
RenderX

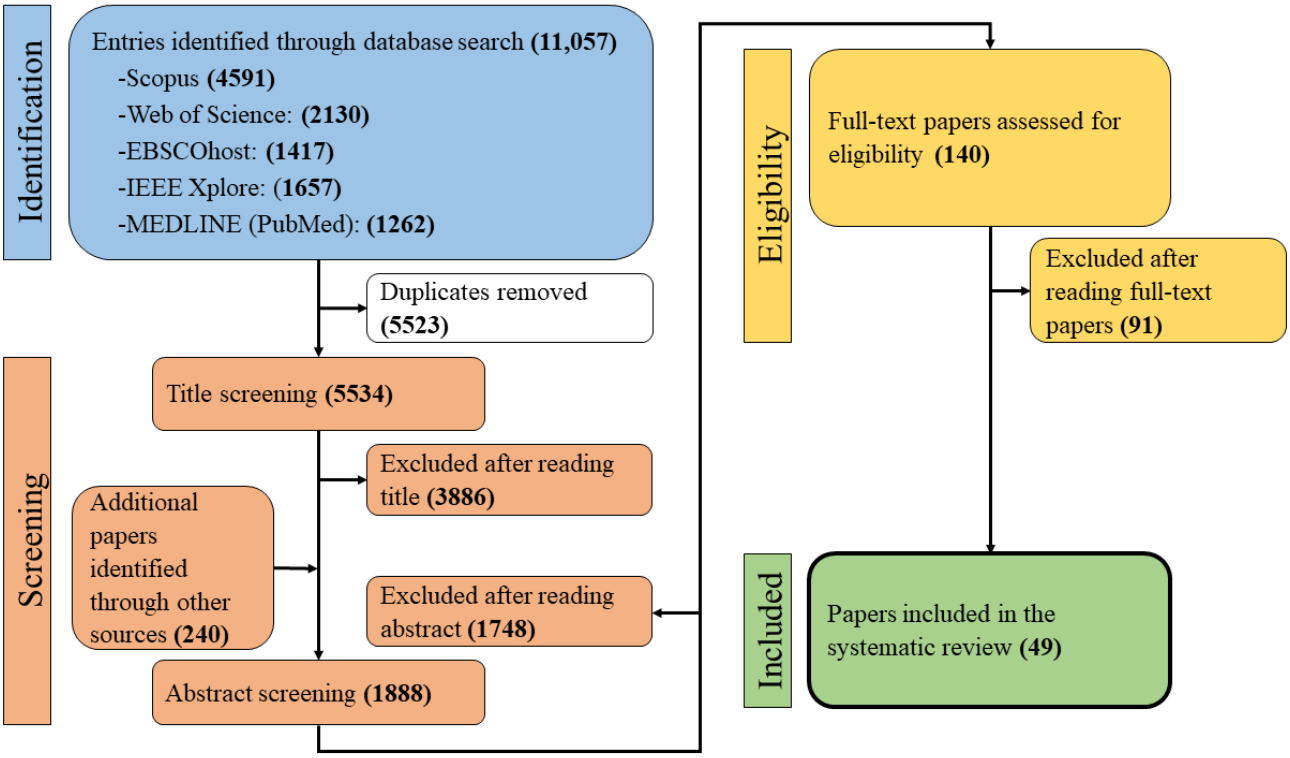
Results

Overview

Out of 11,057 records, we removed 5523 duplicates. The remaining 5534 papers were screened by title, and a further 3886 papers were excluded. We included another 240 papers from different sources (eg, citation screening). After screening all abstracts, 140 papers remained. After reading all full texts, 49 publications were eligible for the current systematic review. The process can be seen in Figure 1.

All 49 selected studies with corresponding extracted data points are listed in Multimedia Appendix 4 [25-73]. A brief overview of the studies can be found in Table 1. Numbers reported must not always add up to 49, as some studies belong to more than 1 subcategory.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



XSL•FO
RenderX

Table 1. Overview over all included studies and extracted data.

Category and subcategory	Values (N=49), n (%)
Publication year	
2016-2019	16 (33)
2020-2023	33 (66)
Study location	
Africa	0 (0)
Asia	20 (41)
Europe	12 (24)
North America	16 (33)
Australia or Oceania	0 (0)
South America	1 (0)
Sample size	
≤20	12 (24)
21-50	11 (22)
51-100	12 (24)
101-1000	12 (24)
>1000	2 (0)
Study duration (days)	
≤1	8 (16)
2-7	10 (20)
8-14	9 (18)
15-31	6 (12)
32-90	8 (16)
>90	5 (10)
NR[a]	3 (1)
Study setting	
Free living	34 (69)
Laboratory	7 (14)
Clinical setting	5 (10)
Mental health outcome	
Depression	14 (29)
Emotions	5 (10)
Fatigue	2 (0)
Mood	5 (10)
State of mind	2 (0)
Stress	15 (31)
Other	9 (18)
Sensor class	
Commercial wearable	22 (45)
Scientific sensor carrier	18 (37)
Smartphone	13 (27)
Sensor position	
Wrist	33 (67)

XSL•FO
RenderX

Category and subcategory	Values (N=49), n (%)
Chest	3 (1)
Other	5 (10)
NR	10 (20)
Machine learning used	
Decision tree	8 (16)
K-nearest neighbor	8 (16)
Linear regression	15 (31)
Long short-term memory	8 (16)
Multilayer perceptron	6 (12)
Neural network	17 (35)
Random forest	28 (57)
Support vector machine	18 (37)
Extreme gradient boosting	18 (37)
Other algorithms	15 (31)
NR	1 (0)

[a]NR: not reported.

Participant and Study Characteristics

The paper's publication dates ranged from 2016 to 2023, with nearly half of them being published after 2020 (n=22, 45%). The studies are from a total of 21 different countries, leading in the list is the United States with a total of 15 studies followed by the Republic of Korea with 5 studies. The number of participants ranged from 2 to 4612, with a mean number of 232.68 (SD 310.86). Almost half of the studies (n=23, 47%) recruited less or equal to 50 participants. Only 14 studies had more than 100 participants. Only 29 studies reported information about the age and sex of the study participants. Almost all of these studies (n=26) focused on adults aged between 20 and 50 years, whereas 2 studies examined older people with a mean age of 70 (SD 10) years, and 1 study examined teenagers between the age of 12 and 20 years. In 21% (n=6) of these studies, less than one-third were male participants. In total, 17 studies had a sex distribution between one-third and two-thirds of male participants. In 6 studies, more than two-thirds of the participants were male. Convenient samples such as college or university students were recruited in 13 studies, and 8 studies recruited patients with a specific illness or condition and a distinct control group. Almost all studies were observational studies (n=45, 92%), and the remaining studies were experiments. The duration of the studies varied from less than 1 day to 365 days, with a mean of 48.04 (SD 54.10) days. A total of 27 studies had a duration of 14 or fewer days, with 6 studies having less than 1 day. Only 6 studies had 3 months or more of data recording. Most studies (n=34, 69%) were carried out in free-living conditions, while 11 reported a clinical or laboratory environment.

Mental Health

The focus of the studies had to be connected to a mental health state or trait. In total, we identified 14 different outcomes. A total of 31% (n=15) of the studies had stress as their main focus, while 29% (n=14) examined depression. Further, we identified 5 studies each on mood and emotion outcomes. The other studies investigated topics like the state of mind, posttraumatic stress disorder, fatigue, or different anxiety disorders. Almost all studies used an initial onboarding phase with different questionnaires or ecological momentary assessment procedures. Across all studies, 14 studies were concerned with classification into different groups (eg, classifying a serious mental illness participant group from a control group or classifying participants into different severe depression levels), while 16 studies focused on detecting individual characteristics (eg, different emotions). The remaining 19 studies focused on predicting future conditions (eg, predicting the presence and severity of a depressive state or predicting a future stress rating).

Physical Activity Assessment Via Wearables

All of the included studies collected sensor information from either a wearable or a smartphone or in 4 cases both. Of all studies, 22 used commercial-grade wearables (eg, Fitbit fitness tracker), 18 studies used scientific sensors (eg, ActiGraph) to record sensor measurements, and 13 studies used smartphones. The most used wearing position was the wrist (n=33, 67%), followed by 8 studies that used other wearing positions such as chest, hip, thigh, or finger-worn sensors. In total, 10 studies had no specific wearing position because they used smartphone recordings as used in everyday life. Physical activity was always measured with triaxial accelerometer sensor measurements. Only 5 studies used exclusively accelerometer sensor values as input, whereas all other studies used a minimum of 1 additional sensor. In particular, 17 studies used heart rate recordings, 4 studies used electrocardiography and 2 studies used heart rate variability recordings. Furthermore, 12 studies used location recordings, primarily acquired via GPS, and 10 studies used electrodermal activity, also mentioned as galvanic skin response or electrodermal response measurement recordings. Gyroscope

XSL•FO
RenderX

as an additional body movement sensor was included in 9 studies.

According to the selection of raw measurements or preprocessed features, 27 of the studies used steps as physical activity features, 21 used raw accelerometer values, and 13 studies did not specify how they measured activity or used activity classification. Further, 7 studies used energy expenditure (eg, metabolic equivalent of task or calories) as a measure of physical activity.

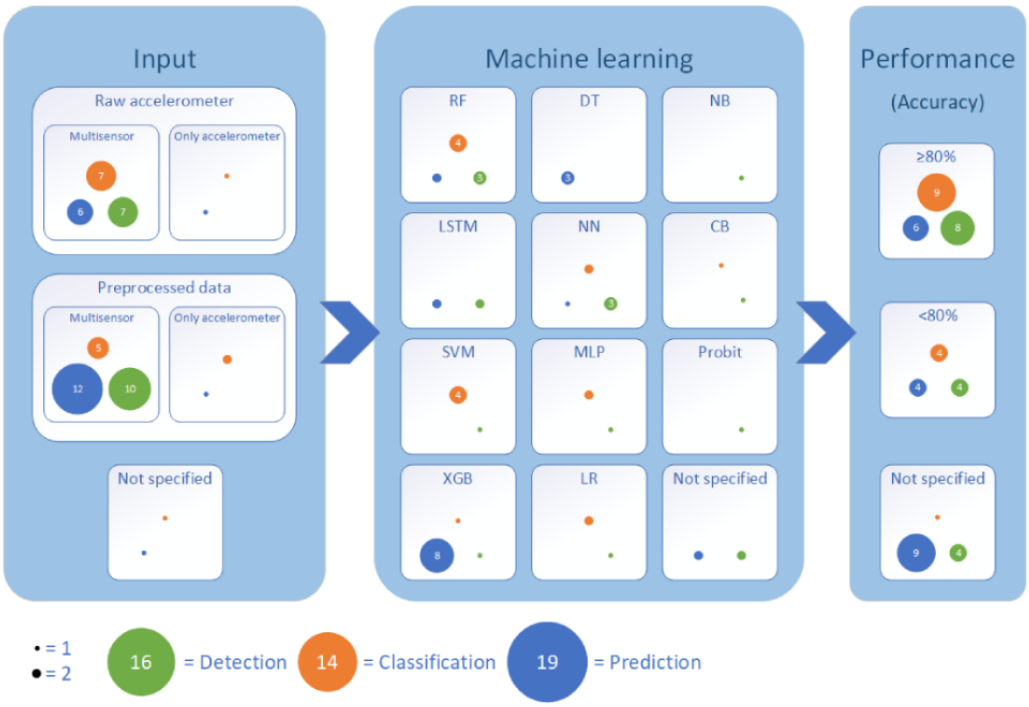
Machine Learning

The features extracted from the recorded sensor measurements and used by the evaluated machine learning algorithms were reported by 63% (n=31) of all studies. Only 27 studies reported a feature evaluation. The top 10 features were widely spread, and the consensus regarding what are the best 10 features for a specific use case was very low. Most studies explored more than 1 machine learning algorithm and compared the monitored performance. More than half of all studies (n=28, 57%) analyzed the use of random forest, followed by support vector machines (n=18, 37%), different neural networks (eg, deep neural network, convolutional neural network, and recurrent neural network; n=17, 35%), linear regression (n=15, 31%), and extreme gradient boosting (XGB; n=13, 27%). Overall, 39 studies reported their best-performing classifier led by XGB (n=10, 26%), followed by random forest (n=9, 23%) and support vector machine (n=5, 13%). When it comes to performance values, 39 studies reported the performance achieved by their used machine learning algorithm. The top reported performance value was accuracy (n=35, 90%), followed by sensitivity (n=25, 64%), precision (n=20, 51%), and *F* value (n=18, 46%).

In total, 22 studies reported the used programming language or used packages. In particular, 12 studies used Python (Python Software Foundation) as a programming language, followed by 7 studies that used R (R Foundation for Statistical Computing). Only 39 studies reported some performance values, and 15 of them reported only 1 or 2 performance values (mostly only accuracy or area under the receiver operating characteristic curve values). The most mentioned limitations were a small sample size (n=20, 41%), lack of diversity in the sample (n=14, 29%), or a low number of data points (n=10, 20%).

Figure 2 presents a summary of the input, machine learning approaches, and performance differentiated by the study aims, namely, detection, prediction, or classification of mental health states or traits. The most significant visual aspects are described in the following. Obviously, only 5 studies used accelerometer measures as input signals, whereas most studies combined accelerometer measures with additional sensor signals such as gyroscope or heart rate. Nearly half of the studies that focused on the prediction of mental health outcomes applied XGB as a machine learning algorithm. In contrast, studies that focused on detection reported a variety of 11 different approaches. Nearly half of the studies that focused on the prediction of mental health outcomes and used XGB as a machine learning model did not report the accuracy. Nearly two-thirds of all studies that focused on the classification reported overall high accuracy (≥80%). A more specific view on the different mental health outcomes revealed, for example, that raw data were more frequently used for the classification of depression, and the classification accuracy is quite promising. In comparison, when it comes to stress, the distribution of input data, machine learning, and accuracy performance is more balanced. For more details on the distribution of outcomes and specific details about the 2 most studied mental health outcomes, depression and stress, see Multimedia Appendix 5.

Figure 2. Overview of the input, machine learning approaches, and performance differentiated by the study aims. Notably, some studies appear multiple times due to different input configuration. CB: cat boost; DT: decision tree; LR: linear regression; LSTM: long short-term memory; MLP: multilayer perceptron; NB: naïve Bayes; NN: neural network; RF: random forest; SVM: support vector machine; XGB: extreme gradient boosting.

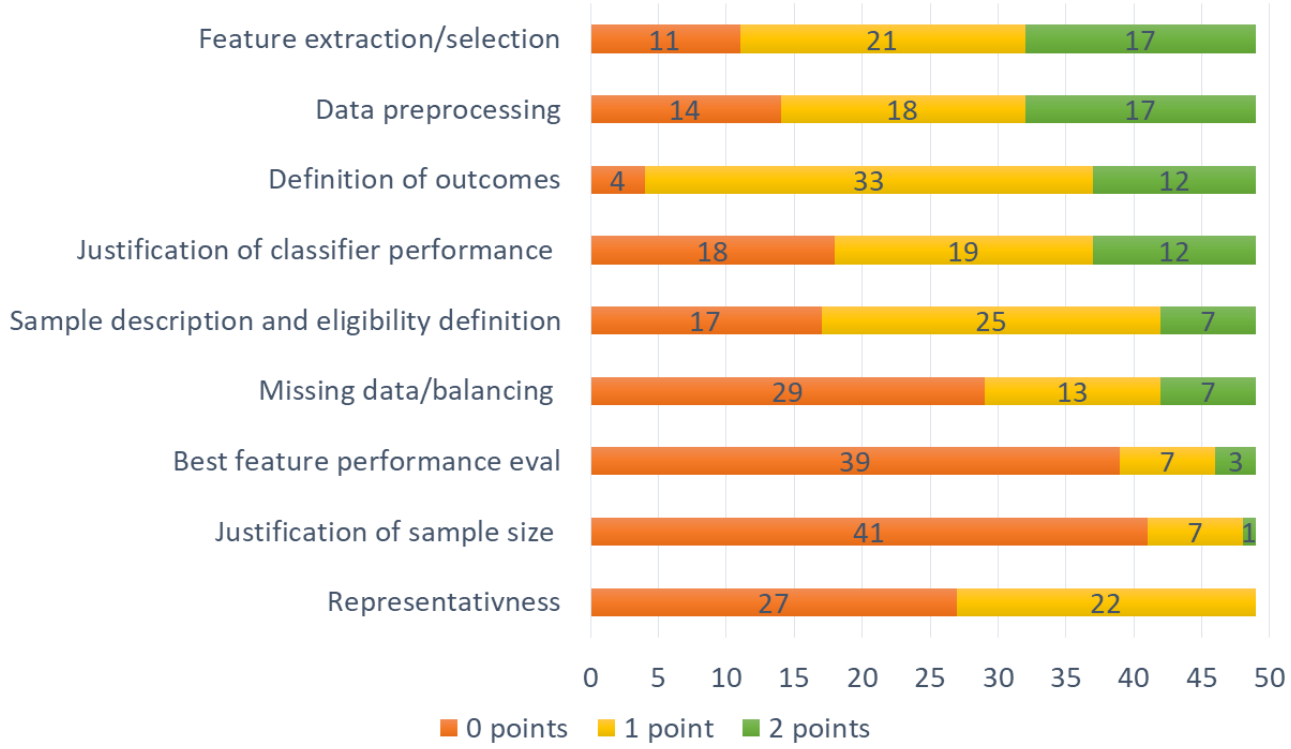


XSL•FO
RenderX

Study Quality

We rated all 49 studies in the 9 categories and classified them into the specified quality class. We classified 21 studies as low quality, 27 studies with moderate quality, and only 1 study with high quality. From another perspective, only 10 studies reached half of the points or more. On the positive side, almost all studies defined the outcome to achieve, but 4 studies did lack even a description of what they wanted to achieve. In the "representativeness" category, no study got 2 points mostly because they did not provide their implementation or data or did not have an independent validation (eg, testing the implemented model with another independent dataset). In the category "justification of sample size," only a total of 9 points (of possible 98 points) was scored. Reporting the sample size is important, as small sample sizes could lead to inaccurate results and therefore to drawing false conclusions [74]. On average, the papers were rated with 6.47 (SD 3.1) points; therefore, the overall quality of the reviewed studies can be rated moderate to low. Figure 3 shows the distribution of points in the individual categories.

Figure 3. Quality of the papers sorted by categories.



Discussion

Principal Findings

In our review, we included a total of 49 studies that applied machine learning methods based on accelerometer measures via wearable technology to predict, detect, or classify mental health states or traits. According to the aims of our review, we can summarize the following main results: first, studies focused on different mental health objectives (prediction, classification, or detection of affect or mood, anxiety, depression, anorexia nervosa, emotions, fatigue, posttraumatic stress disorder, state of mind, and stress). The 2 most frequently studied outcomes were stress with 15 studies and depression with 14 studies. Second, we identified heterogeneity in the used machine learning methods. In particular, across all studies, 30 different machine learning algorithms were used and integrated human acceleration data either from smartphones or wearables. Third, the top 10 studies with the highest accuracy and 8 of the top 10 studies with the highest precision used raw accelerometer measurements instead of already preprocessed data. Fourth, the quality of the studies can be judged as low to moderate. Although the studies pursued different purposes and followed different methodological approaches, obvious limitations become apparent that merit further consideration.

Raw Measurements and Features

As an inclusion criterion for the selection of studies, we set the requirement that either raw accelerometer measurements or at least preprocessed parameters from raw accelerometer measurements were included as an input signal to the machine learning algorithm. In total, 22 studies used raw accelerometer measurements, whereas 27 studies used preprocessed data such as steps or dimensions of energy consumption. According to Fedor et al [22], raw measurement describes the data recorded by wearables or smartphones (eg, accelerometer, gyroscope, location, and heart rate) and can provide direct insights about general mental well-being. For example, an approximation of stress detection can be associated with heart rate or heart rate variability [75].

When focusing on those studies that reported machine learning performance values, the top 10 studies that reported high accuracy used raw accelerometer measurements as input. Moreover, focusing on the performance classifier precision, 8 of the top 10 studies also used raw accelerometer measurements.

XSL•FO
RenderX

When looking at sensitivity and *F* value, 6 of the top 10 studies used again raw accelerometer measurements. This suggests the assumption that using raw accelerometer measurements instead of preprocessed data provides better overall performance. Thus, whenever possible, to ensure the objectivity of the input data, we encourage future research efforts to develop machine learning algorithms based on raw accelerometer measurements rather than preprocessed data. In other words, if algorithms are based on preprocessed data, decisions have already been made that may influence the results [76]. There is no consensus about how the raw data should be processed optimally and harmoniously. In particular, different possibilities for processing raw accelerometer data into metrics or parameters are presented in the literature, for example, counts [77], movement acceleration intensity [78], Euclidian norm minus one [79], or mean amplitude deviation [80]. Furthermore, 20 of all included studies integrated data from commercial-grade wearables such as Fitbit or Garmin. This brings up further challenges, as researchers often do not have access to raw accelerometer measurements of consumer wearables, and they do not have access to the "black-boxed" algorithms either [76]. Optimally, researchers should have access to raw accelerometer measurements or high-resolution feature data that are derived from the raw data. Furthermore, at the very least, comprehensive details of algorithms as well as the version of the device and software should be made transparent. In this context, transparency is a requirement for reproducibility, comparability, stacking of datasets, or creating norms. In line with the mentioned challenges, a significant number of wearables are not validated through high-quality protocols [81,82].

Nearly all studies (n=39) integrated additional measurements next to accelerometer measurements such as heart rate, location, electrodermal activity, or ambient light for the application of the machine learning algorithm. The idea behind this approach might be to add supplementary information to increase the chance to identify patterns. However, we suggest adding only meaningful data to the algorithm based on empirical evidence [22]. Besides the aim to maximize the most accurate prediction, detection, or classification of mental health states or traits based on several raw measurements, it might also be interesting to identify the isolated impact of, for example, raw accelerometer or heart rate measurements on mental health outcomes. We expect an increase in studies applying a machine learning approach in the future and thus the opportunity to tackle these issues.

Although we prefer raw measurements of sensors, these are mostly not expressive enough to gain insights into specific behavior. To gain meaningful information from raw measurements, extracting the so-called low-level features is necessary. For example, wearables worn on the wrist collect linear acceleration via an accelerometer sensor. Through applying feature extraction algorithms to the raw acceleration measurements, they can be transformed into, for example, step count [83] or movement intensity [80].

In total, 18 studies did not report the used features, which hinders the possibility of replicating the application of the described machine learning model. Of the 31 studies that reported the used features, only 23 mentioned their feature selection process, which includes feature evaluation as well as describing the importance of and influence of a feature on the outcome. Feature evaluation is essential to gain insights into which features impact the outcome and what may be correlated to the studied mental health condition [84]. We strongly recommend reporting the used feature selection process and the final feature subset used in the model, especially to transparently explain how the results are achieved.

Additionally, not the most influential features should be selected but the ones highly correlated with the studied outcome. Therefore, consultation with domain experts is highly recommended to ensure that the calculated features contain the necessary information for the identification, classification, or prediction of the targeted mental health condition. Experts in the field recommend that when performing machine learning analyses, 2 assumptions are made, namely, first, the desired outputs of the data can be generated, given the input data; and second, the available data contain the necessary information to learn the desired output [22,84].

In general, the use of low-level features is essential for the application of machine learning, as it joins raw measurements into more meaningful values. A further step is the aggregation of calculated feature values into the so-called high-level features or high-level information [22]. This high-level information can combine multiple low-level features to reflect different behavioral patterns (eg, physical activity or sleep) [22,85]. We found that none of the reviewed studies integrated high-level information, which can be crucial to explain certain behaviors. For example, some studies used features like "hours of sleep" for depression detection or severity classification, but looking at only hours of sleep could not be enough. In particular, Riemann et al [86] provided an overview of the complex relationship between sleep and depression, indicating that several sleep parameters such as sleep stages or sleep efficiency are also related to depression. Therefore, combining multiple features into high-level features might be necessary to get a clearer picture. Another example is the correlation between stress and physical activity. Stress might be associated with reduced physical activity [87]. However, reduced physical activity can have various reasons besides stress (eg, illness or injury or bad weather conditions). In contrast, Smets et al [88] showed that a combination of different physiological signals (eg, electrodermal activity or pulse rate variability) and contextual information (eg, temperature and location) can also be used to predict stress. Therefore, combining both approaches might improve stress detection.

We recommend using different combinations of features depending on the targeted objective. Future studies should focus on crafting optimal feature sets for distinct detection, classification, or prediction of the various mental health states and traits rather than using many individual detached features.

Machine Learning

While classic programming transforms input data via previously defined rules into information-enriched output data, machine learning tries to find these rules (model) that describe the correlation or connection between input data and given output [89]. The more data we have to train our rule set, the better and

XSL•FO
RenderX

more precise the machine learning model can predict, classify, or detect the outcome of unknown input data. Furthermore, an additional effect of a larger dataset for training is a higher generalizability of the model. Our included studies varied widely in terms of aim and rationale. However, some methodological and machine learning–specific issues affect all studies and have an impact on the quality that merits further consideration.

First, when collecting data in studies, 2 main dimensions can increase the size of the dataset: the sample size and the duration over which data are collected. A longer duration helps collect more data points per person and increase the potential to personalize results (ie, within-subject approach), and larger sample sizes help build a more diverse dataset (ie, between-subject approach) [90]. Nevertheless, both dimensions are important to obtain meaningful output when using machine learning. Depending on the use case, however, 1 dimension might be crucial.

For detection or classification purposes, a broad and diverse sample is important for generalization [91]. However, only 2 of 13 studies addressing detection and only 4 of 14 studies focusing on classification had a sample size greater than 100. Overall, almost half of all studies (n=23, 47%) recruited less or equal to 50 participants. Only 14 studies recruited more than 100 participants. A further problem occurring in 11 of 27 studies is the acquisition of participants from a particular group of people (eg, students), which also does not play in favor of generalization.

The duration of data acquisition can play a role as well in detection or classification. Fedor et al [22] show that a recording of 16 days of physical activity and depression level might not be enough to detect an association between physical activity and depression, whereas a duration of 6 weeks may show an association. When predicting trends or forecasting events ideally, a long duration for data collection is envisaged. As seen in the aforementioned example of Fedor et al [22], trends or correlations might not be visible in short-duration datasets. In total, 9 of 19 studies examining prediction or forecasting reported a study duration of 30 days or longer, and 7 of 19 studies only reported 14 or fewer days down to even 1 day. This can mean that correlations did not lie within the analyzed period and dampened the overall performance results reported.

All included studies implemented a minimum of 1 supervised machine learning algorithm for their application. When implementing machine learning models with supervised learning, splitting your primary dataset into different partitions for training and testing is common [84]. The performance measure of the developed model is thus only valid for that particular primary dataset. For use in the clinical area, further testing is needed to ensure that the measured accuracy holds when applying the model to unseen data. Therefore, large and precisely labeled datasets from diverse populations are required [22]. None of the reviewed studies used external supplementary datasets to validate their implemented models and presented performance values. Additionally, none of the studies used different wearables—particularly studies using commercial wearables—that generate the same sensor measurements and features to validate their results. We recommend not only replicating and validating the developed models with independent datasets but also incorporating other wearables that produce the same measurements and features as the original device.

As seen in the analyzed studies, machine learning enables the detection, classification, and prediction of mental health conditions from either raw acceleration measurement or preprocessed physical activity data (eg, steps). In most studies, additional supplementary sensors are used to find possible patterns in the measured data typical for various mental health conditions. Analyzing the influence of the different sensor measurements on the outcome can help strengthen the developed models. However, only a few of the studies compared the capabilities of their machine learning models with different subsets of the recorded measurements. As a future task, not only implementing machine learning models from already collected data but also testing them in real-time and real-world scenarios optimally with live sensor measurements can help collect more insights on how to improve and reinforce the models and how to interpret the outcome [85,92].

The Future: Clinical State

The use of machine learning is becoming increasingly popular when it comes to analyzing large datasets. Machine learning is also being used more and more frequently in sensitive areas, for example, when analyzing health-related data (eg, physical activity and sleep) [93,94]. Considerable progress has been made in some areas of application, leading to high-level information that already enables more valuable insights. Nevertheless, the advancements made are not sophisticated enough to be considered clinically relevant [22,95,96]. For validation and evaluation of developed algorithms, large datasets are needed, which are not available yet. Additionally, large studies are necessary to validate the developed applications and their reported outcomes.

Another factor influencing the outcome of the developed machine learning model is the respective context. For example, a drop-off in physical activity of a person can be a symptom of depressive disorder but can also be the result of a physical injury or bad weather conditions. Only a few studies included contextual information in the form of features in their implementation and training of their model. Here, next to passive monitoring of biosignals or geolocation, the collection of self-reported information via ecological momentary assessment is a valuable addition for deeper insights into contextual and latent factors [97,98].

For some mental health conditions, we identified only one study, such as, for posttraumatic stress disorder [25], social anxiety disorder [26], or anorexia nervosa [27]. To finally conclude, whether the mentioned conditions can be meaningfully recognized with machine learning models via physical activity and optional supplementary sensor measurements is doubtful. Therefore, consultation with in-field professionals is inevitable to assess the feasibility and to evaluate the results. Further, there are not enough studies of some of the investigated mental health conditions to be able to make a reliable statement about the significance of the results.

XSL•FO
RenderX

It should also be considered whether the mental health condition under investigation can be studied with a generalized model at all or if a personalized model needs to be developed for each individual [99]. Measuring the same value for 2 different people could mean different things. Even within 1 person, the meaning of the measurement can change over time. Therefore, huge datasets with diverse populations are necessary for development before these models are included in clinical practice.

As we get closer and closer to implementing new machine learning applications in clinical practice, studies should also focus on the cost-effectiveness of their developed solutions compared to traditional approaches. The initial costs of implementing machine learning applications are high due to the enormous amount of high-quality data needed to train and test meaningful models. After initial training, the costs decrease, as base models need less data for fine-tuning for, for example, personally tailored treatment. First studies have shown that the implementation of machine learning algorithms might be cost-effective not only when it comes to predictive analytics for early disease detection or prediction but also in the areas of remote health care and monitoring or in the management of personalized treatment (eg, maintaining medication plans) [100,101].

In conclusion, the developments in the field of mental health that use machine learning for detection, classification, or prediction are not sophisticated enough for use in the clinical area. In the near future, first applications will assist diagnosis and treatment of selected mental health conditions, but automatically estimating the full clinical state of a patient will take more time and studies [22,102]. All assessed studies explore small parts of the complex field of mental health care and can be considered the first steps toward more digital and intelligent health care systems. These studies can be seen as the first "proof-of-concept" applications in the area of digital health care applications. Despite all this, there are already initial endeavors to use machine learning applications in clinical practice. One example is analyzing electronic health records with the help of machine learning for improving disease phenotyping for differential diagnosis of major depression or predictive modeling of depression and anxiety [103-105]. In another example, machine learning is used to track the clinical state via speech analysis [106]. However, there also remain not only technical challenges but also questions about health equity, data privacy, and security and a new role of health care professionals [107].

Even though the full potential of artificial intelligence has not been reached so far, some examples demonstrate what can be achieved in the foreseeable future. Masud et al [28] developed a system to monitor daily live activities for estimating the depression score (16-item Quick Inventory of Depressive Symptomatology) as well as the depression severity level. The system has the potential to identify depression and classify the severity by assessing day-to-day activities. This enables passive monitoring and can help provide immediate health care when depression is detected. Another study by Rozet et al [29] developed models for predicting individual stress ratings. Different approaches were analyzed, more generalized nomothetic approaches and more individualized ideographic approaches. It was demonstrated that, initially, the ideographic model performed better, but with ongoing data collection, the nomothetic model surpasses it, highlighting the importance of more personalized models because of the heterogeneity of the importance of the predictors on stress. We are expecting an increase of studies that apply machine learning techniques in the area of lifestyle factors (eg, physical activity) and mental health in the future. With the increase in studies and, above all, the development of homogeneous study protocols and methods for reporting machine learning performance, meaningful meta-analyses can be conducted. Ultimately, individual participant data meta-analyses can be carried out in the future.

Strengths and Limitations

The main strength of this study is that it focuses on several mental health conditions examined with machine learning and the possibility to detect, classify, or predict aspects of it based on wearable signals. This provides a broad overview of the possibilities of generating valuable information with the measurements of wearables in the field of mental health. Of course, there are several limitations. First, some relevant studies may have been missed because they did not appear in our search due to missing key terms in the specified search criteria. Additionally, only English-language search results were considered. Additionally, due to the fast development of machine learning applications in research, new studies might have been published after the initial search. Second, study quality was assessed using a combination of the appraisal tool for cross-sectional studies and the Newcastle-Ottawa Scale for longitudinal studies already used by De Angel et al [24] but further complemented with own categories in line with core principles, recommendations, and expert statements [24,84,108,109]. Third, due to the large heterogeneity of included studies (eg, different outcomes, different study designs, and different reported performance values), our systematic review provides no quantitative synthesis of the results in the form of a meta-analysis, as this would not allow any meaningful interpretation.

Future Directions and Conclusions

Our review indicated that the application of machine learning in the context of the association between physical activity and mental health arose over the last couple of years, thus showing the relevance in research. The detection, classification, and prediction of mental health states and traits using machine learning based on wearable signals have shown to be a promising approach. Not only are wearables widely used and easy to operate in today's society, but also they can provide valuable insights about nontrivial aspects like physical activity or sleep. Using machine learning models is a promising next step to gain even more high-level information. Although there are serious pitfalls, which need to be mastered before, these applications can be included in clinical practice. First, researchers should use huge amounts of diverse datasets to improve and enhance the capabilities of their machine learning models. Second, independent evaluation and validation are inevitable to ensure maximum stability and to strengthen the meaningfulness. Third, researchers should consult a domain professional for the development as well as evaluation of the application. Fourth, the research community should agree on

XSL•FO
RenderX

well-formulated guidelines for reporting the analysis of the developed application to guarantee a maximum of transparency. Fifth, researchers should train the machine learning models whenever possible with data that are preprocessed as little as possible.

This review of state-of-the-art applications for assessing mental health states shows that the use of wearable data with the help of machine learning can indeed provide valuable information, but various aspects of the development and reporting process need to improve especially in such a sensitive field as mental health.

Acknowledgments

This study was supported by the Federal Ministry of Education and Research (16DKWN014A).

Authors' Contributions

SW and MG were responsible for the topic and design of this review. MG obtained funding, and SW drafted the manuscript. SW, MG, LL, and RN performed the study selection and data extraction. UWE-P supervised the project. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 69 KB - mhealth_v13i1e59660_app1.pdf]

Multimedia Appendix 2
Search terms used in EBSCOhost, IEEE Xplore, Scopus, PubMed, and Web of Science databases.
[DOCX File , 24 KB - mhealth_v13i1e59660_app2.docx]

Multimedia Appendix 3
Quality assessment categories and scoring details.
[DOCX File , 23 KB - mhealth_v13i1e59660_app3.docx]

Multimedia Appendix 4
Overview of all included studies and respective extracted information.
[DOCX File , 39 KB - mhealth_v13i1e59660_app4.docx]

Multimedia Appendix 5
Overview of input, machine learning, and outcome or performance.
[DOCX File , 1399 KB - mhealth_v13i1e59660_app5.docx]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
2. World mental health report: transforming mental health for all. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2025-02-17]
3. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. EClinicalMedicine 2022;54:101675 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101675] [Medline: 36193171]
4. Hossain MM, Tasnim S, Sultana A, Faizah F, Mazumder H, Zou L, et al. Epidemiology of mental health problems in COVID-19: a review. F1000Res 2020;9:636 [FREE Full text] [doi: 10.12688/f1000research.24457.1] [Medline: 33093946]
5. Reed J, Buck S. The effect of regular aerobic exercise on positive-activated affect: a meta-analysis. Psychol Sport Exerc 2009;10(6):581-594. [doi: 10.1016/j.psychsport.2009.05.009]
6. Reed J, Ones DS. The effect of acute aerobic exercise on positive activated affect: a meta-analysis. Psychol Sport Exerc 2006;7(5):477-514. [doi: 10.1016/j.psychsport.2005.11.003]
7. Biddle SJH, Asare M. Physical activity and mental health in children and adolescents: a review of reviews. Br J Sports Med 2011;45(11):886-895. [doi: 10.1136/bjsports-2011-090185] [Medline: 21807669]

XSL•FO
RenderX

8. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
9. Crowley P, Ikeda E, Islam SMS, Kildedal R, Schade Jacobsen S, Roslyng Larsen J, et al. The surveillance of physical activity, sedentary behavior, and sleep: protocol for the development and feasibility evaluation of a novel measurement system. JMIR Res Protoc 2022;11(6):e35697 [FREE Full text] [doi: 10.2196/35697] [Medline: 35666571]
10. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
11. Lee IM, Keadle SK, Matthews CE. Fitness trackers to guide advice on activity prescription. JAMA 2023;330(18):1733-1734. [doi: 10.1001/jama.2023.19332] [Medline: 37812433]
12. Schuch FB, Vancampfort D, Firth J, Rosenbaum S, Ward PB, Silva ES, et al. Physical activity and incident depression: a meta-analysis of prospective cohort studies. Am J Psychiatry 2018;175(7):631-648. [doi: 10.1176/appi.ajp.2018.17111194] [Medline: 29690792]
13. Schuch FB, Stubbs B, Meyer J, Heissel A, Zech P, Vancampfort D, et al. Physical activity protects from incident anxiety: a meta-analysis of prospective cohort studies. Depress Anxiety 2019;36(9):846-858. [doi: 10.1002/da.22915] [Medline: 31209958]
14. Brokmeier LL, Firth J, Vancampfort D, Smith L, Deenik J, Rosenbaum S, et al. Does physical activity reduce the risk of psychosis? A systematic review and meta-analysis of prospective studies. Psychiatry Res 2020;284:112675. [doi: 10.1016/j.psychres.2019.112675] [Medline: 31757637]
15. Del Pozo Cruz B, Alfonso-Rosa RM, McGregor D, Chastin SF, Palarea-Albaladejo J, Del Pozo Cruz J. Sedentary behaviour is associated with depression symptoms: compositional data analysis from a representative sample of 3233 US adults and older adults assessed with accelerometers. J Affect Disord 2020;265:59-62. [doi: 10.1016/j.jad.2020.01.023] [Medline: 31959584]
16. Hamer M, Coombs N, Stamatakis E. Associations between objectively assessed and self-reported sedentary time with mental health in adults: an analysis of data from the health survey for England. BMJ Open 2014;4(3):e004580 [FREE Full text] [doi: 10.1136/bmjopen-2013-004580] [Medline: 24650807]
17. Stanczykiewicz B, Banik A, Knoll N, Keller J, Hohl DH, Rosińczuk J, et al. Sedentary behaviors and anxiety among children, adolescents and adults: a systematic review and meta-analysis. BMC Public Health 2019;19(1):459 [FREE Full text] [doi: 10.1186/s12889-019-6715-3] [Medline: 31039760]
18. Ashdown-Franks G, Firth J, Carney R, Carvalho AF, Hallgren M, Koyanagi A, et al. Exercise as medicine for mental and substance use disorders: a meta-review of the benefits for neuropsychiatric and cognitive outcomes. Sports Med 2020;50(1):151-170. [doi: 10.1007/s40279-019-01187-6] [Medline: 31541410]
19. Giurgiu M, Koch ED, Ottenbacher J, Plotnikoff RC, Ebner-Priemer UW, Reichert M. Sedentary behavior in everyday life relates negatively to mood: an ambulatory assessment study. Scand J Med Sci Sports 2019;29(9):1340-1351. [doi: 10.1111/sms.13448] [Medline: 31044456]
20. Coutanche MN, Hallion, LS. Machine learning for clinical psychology and clinical neuroscience. In: Wright GC, Hallquist MN, editors. The Cambridge Handbook of Research Methods in Clinical Psychology. Cambridge: Cambridge University Press; Mar 23, 2019:467-482.
21. Tyler NS, Mosquera-Lopez C, Young GM, El Youssef J, Castle JR, Jacobs PG. Quantifying the impact of physical activity on future glucose trends using machine learning. iScience 2022;25(3):103888 [FREE Full text] [doi: 10.1016/j.isci.2022.103888] [Medline: 35252806]
22. Fedor S, Lewis R, Pedrelli P, Mischoulon D, Curtiss J, Picard RW. Wearable technology in clinical practice for depressive disorder. N Engl J Med 2023;389(26):2457-2466. [doi: 10.1056/nejmra2215898]
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
24. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022;5(1):3 [FREE Full text] [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
25. Sadeghi M, McDonald AD, Sasangohar F. Posttraumatic stress disorder hyperarousal event detection using smartwatch physiological and activity data. PLoS One 2022;17(5):e0267749 [FREE Full text] [doi: 10.1371/journal.pone.0267749] [Medline: 35584096]
26. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J Med Internet Res 2020;22(5):e16875 [FREE Full text] [doi: 10.2196/16875] [Medline: 32348284]
27. Lekkas D, Gyorda JA, Jacobson NC. A machine learning investigation into the temporal dynamics of physical activity-mediated emotional regulation in adolescents with anorexia nervosa and healthy controls. Eur Eat Disord Rev 2023;31(1):147-165 [FREE Full text] [doi: 10.1002/erv.2949] [Medline: 36005065]
28. Masud MT, Rahman N, Alam A, Griffiths MD, Alamin M. Non-pervasive monitoring of daily-life behavior to access depressive symptom severity via smartphone technology. : IEEE; 2020 Nov 02 Presented at: 2020 IEEE Region 10

XSL•FO
RenderX

Symposium (TENSYMP); June 7, 2020; Dhaka URL: https://ieeexplore.ieee.org/abstract/document/9230984 [doi: 10.1109/tensymp50017.2020.9230984]
29. Rozet A, Kronish IM, Schwartz JE, Davidson KW. Using machine learning to derive just-in-time and personalized predictors of stress: observational study bridging the gap between nomothetic and ideographic approaches. J Med Internet Res 2019;21(4):e12910 [FREE Full text] [doi: 10.2196/12910] [Medline: 31025942]
30. Ahmed A, Ramesh J, Ganguly S, Aburukba R, Sagahyroon A, Aloul F. Investigating the feasibility of assessing depression severity and valence-arousal with wearable sensors using discrete wavelet transforms and machine learning. Information 2022;13(9):406. [doi: 10.3390/info13090406]
31. Opoku Asare K, Moshe I, Terhorst Y, Vega J, Hosio S, Baumeister H, et al. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: a longitudinal data analysis. Pervasive Mob Comput 2022;83:101621. [doi: 10.1016/j.pmcj.2022.101621]
32. Booth BM, Vrzakova H, Mattingly SM, Martinez GJ, Faust L, D'Mello SK, et al. Toward robust stress prediction in the age of wearables: modeling perceived stress in a longitudinal study with information workers. IEEE Trans Affect Comput 2022;13(4):2201-2217. [doi: 10.1109/TAFFC.2022.3188006]
33. Can YS. Stressed or just running? Differentiation of mental stress and physical activity by using machine learning. Turk J Electr Eng Comput Sci 2022;30(1):312-327. [doi: 10.3906/elk-2102-138]
34. Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study. Sensors (Basel) 2019;19(8):1849 [FREE Full text] [doi: 10.3390/s19081849] [Medline: 31003456]
35. Chakraborty S, Aich S, Joo MI, Sain M, Kim HC. A multichannel convolutional neural network architecture for the detection of the state of mind using physiological signals from wearable devices. J Healthc Eng 2019;2019:5397814 [FREE Full text] [doi: 10.1155/2019/5397814] [Medline: 31687119]
36. Choi J, Ko I, Han S. Depression level classification using machine learning classifiers based on actigraphy data. IEEE Access 2021;9:116622-116646. [doi: 10.1109/ACCESS.2021.3105393]
37. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data. JMIR Mhealth Uhealth 2019;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]
38. Fukazawa Y, Ito T, Okimura T, Yamashita Y, Maeda T, Ota J. Predicting anxiety state using smartphone-based passive sensing. J Biomed Inform 2019;93:103151 [FREE Full text] [doi: 10.1016/j.jbi.2019.103151] [Medline: 30880254]
39. Garcia-Ceja E, Riegler M, Jakobsen P, Torresen J, Nordgreen T, Oedegaard KJ. Motor activity based classification of depression in unipolar and bipolar patients. 2018 Presented at: IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS); June 18-21, 2018; Karlstad, Sweden. [doi: 10.1109/CBMS.2018.00062]
40. Gwak M, Woo E, Sarrafzadeh M. The role of accelerometer and gyroscope sensors in identification of mild cognitive impairment. 2018 Presented at: IEEE Global Conference on Signal and Information Processing; November 26-29, 2018; Anaheim, CA, United States. [doi: 10.1109/GlobalSIP.2018.8646622]
41. Hart A, Reis D, Prestele E, Jacobson NC. Using smartphone sensor paradata and personalized machine learning models to infer participants' well-being: ecological momentary assessment. J Med Internet Res 2022;24(4):e34015 [FREE Full text] [doi: 10.2196/34015] [Medline: 35482397]
42. Hashmi MA, Riaz Q, Zeeshan M, Shahzad M, Fraz MM. Motion reveal emotions: identifying emotions from human walk using chest mounted smartphone. IEEE Sens J 2020;20(22):13511-13522. [doi: 10.1109/JSEN.2020.3004399]
43. Husom EJ, Dautov R, Nedisan AA, Gonidis F, Papatzelos S, Malamas N. Machine learning for fatigue detection using Fitbit fitness trackers. 2022 Presented at: 10th International Conference on Sport Sciences Research and Technology Support; October 26-27, 2022; Valletta, Malta p. 41-52 URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146197563&partnerID=40&md5=5519bd244c7ed404a9f5393c5a8bae97
44. Jacobson NC, Feng B. Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life. Transl Psychiatry 2022;12(1):336 [FREE Full text] [doi: 10.1038/s41398-022-02038-1] [Medline: 35977932]
45. Jakobsen P, Garcia-Ceja E, Riegler M, Stabell LA, Nordgreen T, Torresen J, et al. Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. PLoS One 2020;15(8):e0231995 [FREE Full text] [doi: 10.1371/journal.pone.0231995] [Medline: 32833958]
46. Kilimci ZH, Güven A, Uysal M, Akyokus S. Mood detection from physical and neurophysical data using deep learning models. Complexity 2019;2019(1):1-15. [doi: 10.1155/2019/6434578]
47. Kim H, Lee S, Lee SE, Hong S, Kang HJ, Kim N. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 2019;7(10):e14149 [FREE Full text] [doi: 10.2196/14149] [Medline: 31621642]
48. Lietz R, Harraghy M, Brady J, Calderon D, Cloud J, Makedon F. A wearable system for unobtrusive mood detection. 2019 Presented at: PETRA '19: Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments; June 5-7, 2019; Rhodes, Greece p. 329-330. [doi: 10.1145/3316782.3322743]

XSL•FO
RenderX

49. Luo H, Lee PA, Clay I, Jaggi M, De Luca V. Assessment of fatigue using wearable sensors: a pilot study. Digit Biomark 2020;4(Suppl 1):59-72 [FREE Full text] [doi: 10.1159/000512166] [Medline: 33442581]
50. Magal N, Rab SL, Goldstein P, Simon L, Jiryis T, Admon R. Predicting chronic stress among healthy females using daily-life physiological and lifestyle features from wearable sensors. Chronic Stress (Thousand Oaks) 2022;6:24705470221100987 [FREE Full text] [doi: 10.1177/24705470221100987] [Medline: 35911618]
51. Masud MT, Mamun MA, Thapa K, Lee DH, Griffiths MD, Yang SH. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. J Biomed Inform 2020;103:103371 [FREE Full text] [doi: 10.1016/j.jbi.2019.103371] [Medline: 31935462]
52. Maxhuni A, Hernandez-Leal P, Morales EF, Sucar LE, Osmani V, Mayora O. Unobtrusive stress assessment using smartphones. IEEE Trans Mobile Comput 2021;20(6):2313-2325. [doi: 10.1109/TMC.2020.2974834]
53. Narziev N, Goh H, Toshnazarov K, Lee SA, Chung KM, Noh Y. STDD: short-term depression detection with passive sensing. Sensors (Basel) 2020;20(5):1396 [FREE Full text] [doi: 10.3390/s20051396] [Medline: 32143358]
54. Olsen AF, Torresen J. Smartphone accelerometer data used for detecting human emotions. 2016 Presented at: 3rd International Conference on Systems and Informatics (ICSAI); Shanghai, China; November 19-21, 2016. [doi: 10.1109/ICSAI.2016.7810990]
55. Padmaja B, Prasad VVR, Sunitha KVN, Reddy NCS, Anil CH. Detectstress: a novel stress detection system based on smartphone and wireless physical activity tracker. 2019 Presented at: 1st International Conference on Artificial Intelligence and Cognitive Computing, AICC 2018; February 2-3, 2018; Hyderabad, India p. 67-80 URL: https://tinyurl.com/yc3jk6th
56. Padmaja B, Prasad VVR, Sunitha KVN. Machine learning approach for stress detection using wireless physical activity tracker. Int J Mach Learn Comput 2018;8(1):33-38. [doi: 10.18178/ijmlc.2018.8.1.659]
57. Quiroz JC, Yong MH, Geangu E. Emotion-recognition using smart watch accelerometer data: preliminary findings. 2017 Presented at: UbiComp '17: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers; September 11-15, 2017; Maui, Hawaii. [doi: 10.1145/3123024.3125614]
58. Ross MK, Tulabandhula T, Bennett CC, Baek EG, Kim D, Hussain F, et al. A novel approach to clustering accelerometer data for application in passive predictions of changes in depression severity. Sensors (Basel) 2023;23(3):1585 [FREE Full text] [doi: 10.3390/s23031585] [Medline: 36772625]
59. Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. JMIR Mhealth Uhealth 2021;9(10):e24872 [FREE Full text] [doi: 10.2196/24872] [Medline: 34694233]
60. Saito T, Suzuki H, Kishi A. Predictive modeling of mental illness onset using wearable devices and medical examination data: machine learning approach. Front Digit Health 2022;4:861808 [FREE Full text] [doi: 10.3389/fdgth.2022.861808] [Medline: 35493532]
61. Sanchez W, Martinez A, Hernandez Y, Estrada H, Gonzalez-Mendoza M. A predictive model for stress recognition in desk jobs. J Ambient Intell Human Comput 2018;14(1):17-29. [doi: 10.1007/s12652-018-1149-9]
62. Sa-Nguannarm P, Elbasani E, Kim JD. Human activity recognition for analyzing stress behavior based on Bi-LSTM. Technol Health Care 2023;31(5):1997-2007. [doi: 10.3233/THC-235002] [Medline: 36872815]
63. Sevil M, Rashid M, Hajizadeh I, Askari MR, Hobbs N, Brandt R, et al. Discrimination of simultaneous psychological and physical stressors using wristband biosignals. Comput Methods Programs Biomed 2021;199:105898 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105898] [Medline: 33360529]
64. Shah RV, Grennan G, Zafar-Khan M, Alim F, Dey S, Ramanathan D, et al. Personalized machine learning of depressed mood using wearables. Transl Psychiatry 2021;11(1):338 [FREE Full text] [doi: 10.1038/s41398-021-01445-0] [Medline: 34103481]
65. Spulber S, Elberling F, Svensson J, Tiger M, Ceccatelli S, Lundberg J. Patterns of activity correlate with symptom severity in major depressive disorder patients. Transl Psychiatry 2022;12(1):226 [FREE Full text] [doi: 10.1038/s41398-022-01989-9] [Medline: 35654778]
66. Sultana M, Al-Jefri M, Lee J. Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: exploratory study. JMIR Mhealth Uhealth 2020;8(9):e17818 [FREE Full text] [doi: 10.2196/17818] [Medline: 32990638]
67. Sükei E, Norbury A, Perez-Rodriguez MM, Olmos PM, Artés A. Predicting emotional states using behavioral markers derived from passively sensed data: data-driven machine learning approach. JMIR Mhealth Uhealth 2021;9(3):e24465 [FREE Full text] [doi: 10.2196/24465] [Medline: 33749612]
68. Tazawa Y, Liang KC, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. Heliyon 2020;6(2):e03274 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03274] [Medline: 32055728]
69. Umematsu T, Sano A, Taylor S, Tsujikawa M, Picard RW. Forecasting stress, mood, and health from daytime physiology in office workers and students. Annu Int Conf IEEE Eng Med Biol Soc 2020;2020:5953-5957. [doi: 10.1109/EMBC44109.2020.9176706] [Medline: 33019329]

XSL•FO
RenderX

70. Umematsu T, Sano A, Picard RW. Daytime data and LSTM can forecast tomorrow's stress, health, and happiness. 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany.
71. Umematsu T, Sano A, Taylor S, Picard RW. Improving students' daily life stress forecasting using LSTM neural networks. 2021 Presented at: IEEE International Conference on Biomedical and Health Informatics (BHI); May 19-22, 2019; Chicago, IL, United States URL: https://ieeexplore.ieee.org/document/8834624 [doi: 10.1109/BHI.2019.8834624ISBN:2641-3604]
72. Yan S, Hosseinmardi H, Kao HT, Narayanan S, Lerman K, Ferrara E. Affect estimation with wearable sensors. J Healthc Inform Res 2020;4(3):261-294 [FREE Full text] [doi: 10.1007/s41666-019-00066-z] [Medline: 35415445]
73. Zebin T, Peek N, Casson AJ. Physical activity based classification of serious mental illness group participants in the UK biobank using ensemble dense neural networks. Annu Int Conf IEEE Eng Med Biol Soc 2019;2019:1251-1254. [doi: 10.1109/EMBC.2019.8857532] [Medline: 31946119]
74. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016;6(12):e011458 [FREE Full text] [doi: 10.1136/bmjopen-2016-011458] [Medline: 27932337]
75. Vrijkotte TGM, van Doornen LJP, de Geus EJC. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. Hypertension 2000;35(4):880-886 [FREE Full text] [doi: 10.1161/01.hyp.35.4.880] [Medline: 10775555]
76. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
77. Yang CC, Hsu YL. A review of accelerometry-based wearable motion detectors for physical activity monitoring. Sensors (Basel) 2010;10(8):7772-7788. [doi: 10.3390/s100807772] [Medline: 22163626]
78. Van Someren EJ, Lazeron RH, Vonk BF, Mirmiran M, Swaab DF. Gravitational artefact in frequency spectra of movement acceleration: implications for actigraphy in young and elderly subjects. J Neurosci Methods 1996;65(1):55-62. [doi: 10.1016/0165-0270(95)00146-8] [Medline: 8815309]
79. van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 2013;8(4):e61691 [FREE Full text] [doi: 10.1371/journal.pone.0061691] [Medline: 23626718]
80. Vähä-Ypyä H, Vasankari T, Husu P, Suni J, Sievänen H. A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer. Clin Physiol Funct Imaging 2015;35(1):64-70. [doi: 10.1111/cpf.12127] [Medline: 24393233]
81. Giurgiu M, Timm I, Becker M, Schmidt S, Wunsch K, Nissen R, et al. Quality evaluation of free-living validation studies for the assessment of 24-hour physical behavior in adults via wearables: systematic review. JMIR Mhealth Uhealth 2022;10(6):e36377 [FREE Full text] [doi: 10.2196/36377] [Medline: 35679106]
82. Giurgiu M, Ketelhut S, Kubica C, Nissen R, Doster AK, Thron M, et al. Assessment of 24-hour physical behaviour in adults via wearables: a systematic review of validation studies under laboratory conditions. Int J Behav Nutr Phys Act 2023;20(1):68 [FREE Full text] [doi: 10.1186/s12966-023-01473-7] [Medline: 37291598]
83. Zhao N. Full-featured pedometer design realized with 3-axis digital accelerometer. Analog Dialogue. 2010. URL: https://www.analog.com/media/en/analog-dialogue/volume-44/number-2/articles/pedometer-design-3-axis-digital-acceler.pdf [accessed 2025-02-12]
84. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. Circ Cardiovasc Qual Outcomes 2020;13(10):e006556 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.120.006556] [Medline: 33079589]
85. Farrahi V, Rostami M. Machine learning in physical activity, sedentary, and sleep behavior research. JASSB 2024;3(1). [doi: 10.1186/s44167-024-00045-9]
86. Riemann D, Berger M, Voderholzer U. Sleep and depression—results from psychobiological studies: an overview. Biol Psychol 2001;57(1-3):67-103. [doi: 10.1016/s0301-0511(01)00090-4] [Medline: 11454435]
87. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Sports Med 2014;44(1):81-121 [FREE Full text] [doi: 10.1007/s40279-013-0090-5] [Medline: 24030837]
88. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1:67 [FREE Full text] [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]
89. Zhou ZH. Machine Learning. Singapore: Springer; 2021.
90. Simkus J. Between-subjects vs. within-subjects study design. Simply Psychology. 2023. URL: https://www.simplypsychology.org/between-subjects-vs-within-subjects-design.html [accessed 2024-03-14]
91. Cho G, Yim J, Choi Y, Ko J, Lee SH. Review of machine learning algorithms for diagnosing mental illness. Psychiatry Investig 2019;16(4):262-269 [FREE Full text] [doi: 10.30773/pi.2018.12.21.2] [Medline: 30947496]
92. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]

XSL•FO
RenderX

93. Alanazi A. Using machine learning for healthcare challenges and opportunities. Inform Med Unlocked 2022;30:100924. [doi: 10.1016/j.imu.2022.100924]
94. Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics 2021;22(4):291-300 [FREE Full text] [doi: 10.2174/1389202922666210705124359] [Medline: 35273459]
95. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 2018;14(1):91-118. [doi: 10.1146/annurev-clinpsy-032816-045037] [Medline: 29401044]
96. Elyoseph Z, Levkovich I, Shinan-Altman S. Assessing prognosis in depression: comparing perspectives of AI models, mental health professionals and the general public. Fam Med Community Health 2024;12(Suppl 1):e002583 [FREE Full text] [doi: 10.1136/fmch-2023-002583] [Medline: 38199604]
97. Ebner-Priemer UW, Trull TJ. Ambulatory assessment. Eur Psychol 2009;14(2):109-119. [doi: 10.1027/1016-9040.14.2.109]
98. Reichert M, Giurgiu M, Koch ED, Wieland LM, Lautenbach S, Neubauer AB, et al. Ambulatory assessment for physical activity research: state of the science, best practices and future directions. Psychol Sport Exerc 2020;50:101742 [FREE Full text] [doi: 10.1016/j.psychsport.2020.101742] [Medline: 32831643]
99. Aalbers G, Hendrickson AT, Vanden Abeele MM, Keijsers L. Smartphone-tracked digital markers of momentary subjective stress in college students: idiographic machine learning analysis. JMIR Mhealth Uhealth 2023;11:e37469 [FREE Full text] [doi: 10.2196/37469] [Medline: 36951924]
100. Yadav V. AI and economics of mental health: analyzing how AI can be used to improve the cost-effectiveness of mental health treatments and interventions. J Sci Eng Res 2021;8(7):274-284. [doi: 10.5281/zenodo.13600238]
101. de Vos J, Visser LA, de Beer AA, Fornasa M, Thoral PJ, Elbers PWG, et al. The potential cost-effectiveness of a machine learning tool that can prevent untimely intensive care unit discharge. Value Health 2022;25(3):359-367 [FREE Full text] [doi: 10.1016/j.jval.2021.06.018] [Medline: 35227446]
102. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
103. Chen ZS, Kulkarni P, Galatzer-Levy IR, Bigio B, Nasca C, Zhang Y. Modern views of machine learning for precision psychiatry. Patterns (N Y) 2022;3(11):100602 [FREE Full text] [doi: 10.1016/j.patter.2022.100602] [Medline: 36419447]
104. Ettore E, Müller P, Hinze J, Riemenschneider M, Benoit M, Giordana B, et al. Digital phenotyping for differential diagnosis of major depressive episode: narrative review. JMIR Ment Health 2023;10:e37225 [FREE Full text] [doi: 10.2196/37225] [Medline: 36689265]
105. Nemesure MD, Heinz MV, Huang R, Jacobson NC. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep 2021;11(1):1980 [FREE Full text] [doi: 10.1038/s41598-021-81368-4] [Medline: 33479383]
106. Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ, et al. Clinical state tracking in serious mental illness through computational analysis of speech. PLoS One 2020;15(1):e0225695 [FREE Full text] [doi: 10.1371/journal.pone.0225695] [Medline: 31940347]
107. Singhal S, Cooke DL, Villareal RI, Stoddard JJ, Lin CT, Dempsey AG. Machine learning for mental health: applications, challenges, and the clinician's role. Curr Psychiatry Rep 2024;26(12):694-702. [doi: 10.1007/s11920-024-01561-w] [Medline: 39523249]
108. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
109. Seghier ML. Ten simple rules for reporting machine learning methods implementation and evaluation on biomedical data. Int J Imaging Syst Tech 2021;32(1):5-11. [doi: 10.1002/ima.22674]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
XGB: extreme gradient boosting

XSL•FO
RenderX

Edited by L Buis; submitted 18.04.24; peer-reviewed by H Yin, A Aminbeidokhti, M Salvi; comments to author 17.10.24; revised version received 29.11.24; accepted 06.02.25; published 06.03.25.

Please cite as:
Woll S, Birkenmaier D, Biri G, Nissen R, Lutz L, Schroth M, Ebner-Priemer UW, Giurgiu M
Applying AI in the Context of the Association Between Device-Based Assessment of Physical Activity and Mental Health: Systematic Review
JMIR Mhealth Uhealth 2025;13:e59660
URL: https://mhealth.jmir.org/2025/1/e59660
doi:10.2196/59660
PMID:40053765

©Simon Woll, Dennis Birkenmaier, Gergely Biri, Rebecca Nissen, Luisa Lutz, Marc Schroth, Ulrich W Ebner-Priemer, Marco Giurgiu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Role of Environmental Factors in Technology-Assisted Physical Activity Intervention Studies Among Older Adults: Scoping Review

Carl-Philipp Jansen[1,2], PhD; Désirée Nijland[3], MSc; Jean-Michel Oppert[4], PhD; Veysel Alcan[5], PhD; Kirsi E Keskinen[6], PhD; Emmi Matikainen-Tervola[7], MSc; Zada Pajalic[8], PhD; Merja Rantakokko[6,9], PhD; Signe Tomsone[10], PhD; Essi-Mari Tuomola[6], PhD; Erja Portegijs[3], PhD; Erik J Timmermans[11], PhD

[1]Geriatric Center, Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany

[2]Clinic for Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany

[3]Department of Human Movement Sciences, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[4]Department of Nutrition, Pitié-Salpêtrière Hospital (AP-HP), Sorbonne University, Paris, France

[5]Department of Electrical and Electronics Engineering, Engineering Faculty, Tarsus University, Tarsus, Turkey

[6]Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

[7]Institute of Rehabilitation, JAMK University of Applied Sciences, Jyväskylä, Finland

[8]Campus Drammen, Faculty of Health and Social Sciences, University of South-Eastern Norway, Oslo, Norway

[9]The Wellbeing Services County of Central Finland, Jyväskylä, Finland

[10]Department of Rehabilitation, Faculty of Health and Sport Sciences, Rīga Stradiņš University, Riga, Latvia

[11]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Corresponding Author:
Carl-Philipp Jansen, PhD
Geriatric Center
Medical Faculty Heidelberg
Heidelberg University
Rohrbacherstrasse 149
Heidelberg, 69126
Germany
Phone: 49 6221 319 1760
Email: Carl-Philipp.Jansen@med.uni-heidelberg.de

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2025/1/e83822

Abstract

Background: The rapidly emerging integration of both technological applications and environmental factors in physical activity (PA) interventions among older adults highlights the need for an overarching investigation.

Objective: This scoping review compiled the current literature and aimed to provide an overview of the role of physical, social, socioeconomic, and systemic environmental factors in technology-assisted PA interventions for older adults.

Methods: We systematically searched 6 common databases up to September 16, 2024, for original longitudinal studies (with at least one preintervention measurement and one postintervention measurement) that reported on the role of environmental factors in technology-assisted PA interventions for independently living, community-dwelling older adults. In a stepwise process, data on study characteristics (step 1), environmental factors and their role in the included studies (step 2), and intervention outcomes and effects by type of environmental factor (step 3) were summarized.

Results: A total of 8020 articles were screened, and 25 (0.31%) were included. Most studies were conducted in Europe (11/25, 44%), followed by North America (5/25, 20%), Asia (5/25, 20%), and Oceania (4/25, 16%). Social environmental factors were most often considered (19/25, 76%), followed by factors from the physical (8/25, 32%), socioeconomic (1/25, 4%), and systemic environment (1/25, 4%). Environmental factors were used as the outcome (8/25, 32%), setting variable (7/25, 28%), moderator

XSL•FO
RenderX

or facilitator (8/25, 32%), and intervention component (3/25, 12%). In most studies (19/25, 76%), the intervention had a beneficial effect on the outcome of interest, and the included environmental factor played a supportive role in achieving this effect. In some studies, no effect (3/25, 12%), mixed effects (2/25, 8%), or adverse effects (1/25, 4%) of the interventions were reported.

Conclusions: This is the first comprehensive description of how environmental factors interact with technology-assisted interventions to increase or optimize PA in older adults. It was found that the investigation of environmental factors in this field is at an early stage. Environmental factors were found to play a supportive role in achieving beneficial effects of technology-assisted PA interventions, but the findings were based on a heterogeneous empirical platform. Still, certain aspects such as the application of virtual reality environments and social (or peer) comparison have shown significant potential that remains to be leveraged. A better understanding of intervention results and support in tailoring intervention programs can be provided through the inclusion of environmental aspects in technology-assisted PA interventions for older adults.

(JMIR Mhealth Uhealth 2025;13:e59570) doi:10.2196/59570

KEYWORDS

environmental factors; intervention; older adults; physical activity; technology; PRISMA

Introduction

Background

Physical activity (PA) in older adults is crucial for the prevention of major chronic noncommunicable diseases and for the improvement or maintenance of mobility, independence, and quality of life [1-3]. Although the health benefits of PA are well established, the prevalence of insufficient PA among older adults is high and points to a considerable scope and need for improvement [4-7]. Therefore, interventions are needed to encourage older adults to initiate and maintain regular PA [8].

In the past 2 decades, PA interventions have increasingly been incorporating technological applications because these may help increase motivation and adherence among participants, (remotely) measure and monitor (changes in) intervention outcomes, and provide feedback about this to participants [9-11]. Previous reviews and meta-analyses have shown that technological applications such as websites, mobile or wearable devices, smartphone apps, and virtual reality have been reported to support PA in the older adult population [9,11-16].

In parallel, the importance of applying socioecological approaches to advancing the understanding of PA determinants, including those from different environmental domains, has increasingly been acknowledged [8,17]. Previous research has emphasized that a variety of physical, social, socioeconomic, and systemic environmental factors play a crucial role in facilitating or hindering PA in older adults [18-20]. Therefore, environmental factors have also been considered increasingly in PA interventions among older adults [8,17]. For instance, one study indicated that a PA intervention was more effective in maintaining or increasing older adults' PA when implemented in more walkable neighborhoods that are characterized by higher levels of residential density, land use mix, and intersection density [21]. In addition to characteristics of the physical environment (eg, walkability) [8,21], previous studies have also indicated that aspects of the social (eg, receiving social support) [22], socioeconomic (eg, area-level income) [23], and systemic (eg, ethnicity and climate) [24,25] environment are important to consider when implementing PA interventions among (older) adults. However, environmental factors have rarely been investigated regarding their impact on adherence and effectiveness of technology-assisted PA interventions [10,26]. The rapidly emerging integration of both technological applications and environmental factors in PA interventions in the recent past justifies and at the same time highlights the need for an overarching investigation.

Objectives

Against this background, this scoping review compiled the current literature and provided an overview of the role of physical, social, socioeconomic, and systemic environmental factors in technology-assisted PA interventions for older adults. In this way, we provided new knowledge on the specific environmental factors that have been considered in such previous interventions and investigated whether these factors are associated with adherence to and outcomes of interventions. Furthermore, we aimed to increase insights into how environmental factors may modify outcomes, how they are affected by technology-assisted PA interventions, or how they might be part of underlying mechanisms of this type of interventions. As a consequence of the work at hand, health care professionals, policy makers, and researchers may be enabled to better design effective technology-assisted PA interventions for the target group of older adults.

Methods

This scoping review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews; Multimedia Appendix 1) guidelines [27], and the protocol was registered on the Open Science Framework platform in March 2023 [28].

Literature Search

A search query was carried out on March 16, 2023, and updated on September 16, 2024. It was run in 6 bibliographic databases: CINAHL, Embase, MEDLINE, PsycINFO, Scopus, and Web of Science. The search algorithm was built using search terms based on definitions and synonyms of intervention types; technologies; the target group (ie, older adults); and the physical, social, socioeconomic, and systemic environment and its attributes. The search was not limited to a specific time frame. A detailed search strategy for each bibliographic database can be found in Multimedia Appendix 2. The reference lists of the

XSL•FO
RenderX

included articles were screened to identify additional eligible papers.

Inclusion and Exclusion Criteria

In this scoping review, only published articles in the form of original, longitudinal intervention studies with at least one preintervention measurement and one postintervention measurement and with any health-, function-, or behavior-related outcome were included. Of those articles, only papers were selected that included (1) independently living, community-dwelling older adults aged ≥60 years (ie, study sample average age of ≥60 years with a minimum individual age of 50 years) regardless of their health status; (2) an intervention period involving PA (components); (3) a PA intervention with or without a control setting that was based on or aided or strengthened by technology applications or technological components; (4) an assessment of physical, social, socioeconomic, or systemic environmental factors or a specific comparison of groups with different environmental conditions; and (5) a report of associations of environmental factors with adherence to or outcomes of the PA intervention or whether environmental factors were included as an outcome. The language restriction was set to English, Dutch, German, French, Finnish, Latvian, Norwegian, Swedish, and Turkish because these languages were spoken by the review team. Articles that were literature reviews, study protocols, conference proceedings, or abstracts only were excluded.

Study Selection

After removing duplicate records in EndNote (version 20.0; Clarivate Analytics), the title and abstract of each record were independently screened by 2 reviewers out of the pool of reviewers (ie, all authors of this scoping review). To increase consistency between reviewers, the procedure was discussed within the group of reviewers, who first screened a set of 30 records as a pilot test and discussed results before initiating the full-text screening. Subsequently, the full text of each potentially relevant paper was independently assessed for eligibility by 2 reviewers. Any disagreement between reviewers was resolved through discussion or, if no consensus could be reached, through discussion with a third reviewer. The screening of titles, abstracts, and full texts was conducted using the Rayyan software (Qatar Computing Research Institute), a noncommercial web-based application [29].

Data Extraction

The data extraction from each eligible paper was performed by 2 reviewers per manuscript using a predefined standardized extraction form in Microsoft Excel (Microsoft Corp). Both reviewers then compared their results and harmonized their findings. Any data extraction issues identified by the reviewers were resolved through group discussion. For each eligible paper, information was extracted on (1) study characteristics (ie, full reference, country in which the study was conducted, and main study objectives); (2) methodological aspects of the study (ie, sample size, study sample characteristics, and study design); (3) intervention components (eg, PA component, duration and frequency of PA or exercise sessions, and supervision of sessions); (4) technology components (ie, technology devices used); (5) environmental factors (ie, type of environmental factor included, such as physical, social, socioeconomic, or systemic), an assessment tool for environmental factors, specific environmental characteristics assessed, and the role of an environmental factor (eg, outcome, a feature of the study design, or a factor with a modifying effect on the intervention outcome); (6) study outcomes (ie, primary and secondary outcomes, outcome measurements, and main study findings); and (7) main study findings in relation to the environmental factors assessed. The reviewers did not contact the authors of the eligible articles to collect unreported data or additional details.

Data Analysis

In this scoping review, data from the included studies were analyzed through a stepwise process. First, study characteristics were described using descriptive statistics. Second, the environmental factors and their role in technology-assisted PA interventions were described per environmental domain. Third, intervention outcomes were described by each type of environmental factor, and the actual effects of or on environmental factors were summarized.

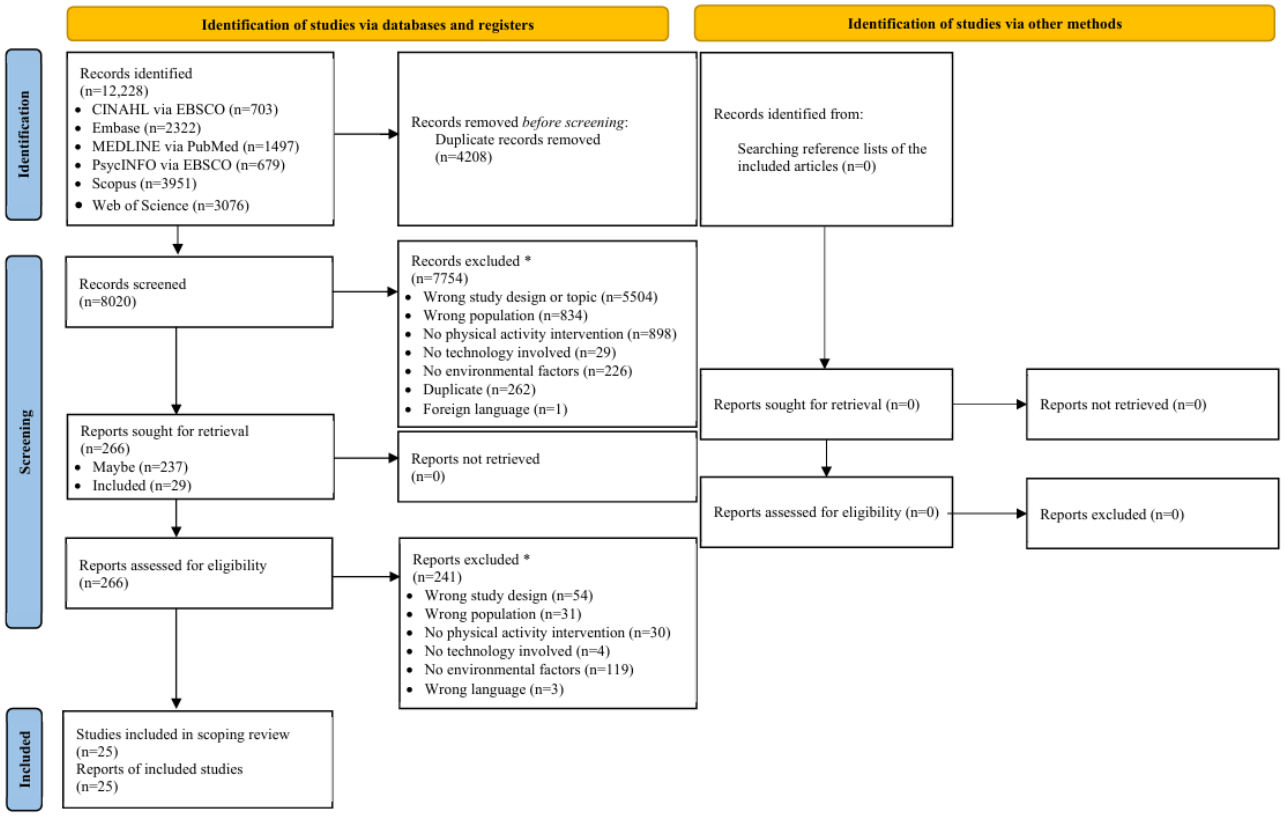
Results

Literature Search

A total of 12,228 articles were identified from CINAHL (n=703, 5.75%), Embase (n=2322, 18.99%), MEDLINE (n=1497, 12.24%), PsycINFO (n=679, 5.55%), Scopus (n=3951, 32.31%), and Web of Science (n=3076, 25.16%). After removing 34.41% (4208/12,228) of duplicates, 65.59% (8020/12,228) of the articles were included in the title and abstract screening phase. After this phase, of the 8020 included articles, 266 (3.32%) were screened for eligibility in the full-text screening phase. In total, 25 articles met the inclusion criteria and were included in this scoping review (Figure 1) [30-54].

XSL•FO
RenderX

Figure 1. Flow chart of study inclusion. The exclusion of reports was conducted based on the order of the reasons listed in this figure, and only 1 reason for exclusion was recorded.



Study Characteristics

An overview of the characteristics of all 25 included studies is presented in Table 1. The papers were published between 2011 and 2024, and most articles (15/25, 60%) were published between 2019 and 2024. Of the included studies, most (11/25, 44%) were conducted in Europe [33,36,37,39,41,44,45,47,50-52], followed by North America (5/25, 20%) [32,42,46,49,53], Asia (5/25, 20%) [30,34,35,43,48], and Oceania (4/25, 16%) [31,38,40,54]. The 2 study designs that were most often used across the studies were pretest-posttest (12/25, 48%) [30,32,33,36,41,43,44,46,48,50,52,54] and randomized (controlled) trials (8/25, 32%) [31,34,36-39,41,46]. The sample sizes in the included studies ranged from 1 to 409 (mean 85.7, SD 98.9 participants), and the percentage of women in the study samples ranged from 0% to 100%. Across all the included studies, the lowest and highest reported age of an individual was 50 and 99 years, respectively. The PA components that were most often considered in the technology-assisted PA interventions were (treadmill) walking (8/25, 32%) [30,36,40,42,45,49,53,54] and balance and coordination exercises (8/25, 32%) [34,35,38,44,48,51-53], followed by stretching and flexibility (7/25, 28%) [31,33,34,40,43,50,51] and functional training (5/25, 20%) [33,34,36,51,54]. Of the 25 included studies, 18 (72%) reported supervised PA interventions [34-45,47-49,51,53,54], 4 (16%) included a nonsupervised PA intervention [31,48,52,54], and 3 (12%) did not provide information regarding the supervision of the technology-assisted PA intervention [30,32,33]. The duration of these PA interventions, the frequency and duration of individual intervention sessions, and the technology device used varied across the studies (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of the study design, the study sample, and the technology-assisted physical activity interventions in the included studies (N=25).

Study, year, and country	Main study objective	Study design	Sample size, n (% women)	Age (y)	Physical activity component	Technology device	Supervised intervention and duration of the intervention	Session details
Abe et al [30], 2023, Japan	To examine the effects of the Kikoeru app on social connectedness, subjective health, loneliness, and setting a target number of steps	Pretest-posttest without a control group	7 (0)	Mean 74 (SD 4; range 65-78)	Walking	Smartphone	• NR[a] • 60 d	Daily use of the app
Alley et al [31], 2024, Australia	To examine the moderating effect of social support on the effectiveness of a web-based, computer-tailored physical activity intervention for older adults	RCT[b]	243 (78.6)	Mean 68.84 (SD 3.85; range 65-98)	Physical activity advice, stretching and flexibility, and strength exercises	Computer and Fitbit device	• No • 12 wk	Average time on the website during the whole intervention period ranged from 126.89 to 140.24 minutes between groups
Anderson-Hanley et al [32], 2011, United States	To examine the effect of virtual social facilitation (avatars) and competitiveness on exercise effort in exergaming older adults	Pretest-posttest without a control group	14 (92.9)	Total sample: NR (range 60-99); low-competitiveness group: mean 80.7 (SD 12.3; range NR); high-competitiveness group: mean 75.6 (SD 13.5; range NR)	Cycling	Computer	• NR • 1 mo	2 to 3 rides per week
Boekhout et al [33], 2019, the Netherlands	To examine (1) which individual characteristics predict differences in preference between printed and web-based delivery and (2) which user characteristics and delivery aspects predicted attrition	Pretest-posttest without a control group	409 (64.5)	Total sample: NR; printed delivery group: mean 79.2 (SD 7.6); web-based delivery group: mean 73.3 (SD 6.6)	Functional training and stretching and flexibility	Computer	• NR • 3 mo	NR
Chen et al [34], 2019, China	To examine the effects of a home-based exercise intervention to reduce knee osteoarthritis symptoms and improve physical functioning in older adults	Quasi-experimental study with a control group	171 (84.4)	Mean 68.9 (SD 7.4)	Functional training, stretching and flexibility, and balance and coordination exercises	Phone	• Yes • 12 wk	4 sessions of 2 hours; at least 3 sessions of 30-40 minutes per week

XSL•FO
RenderX

Study, year, and country	Main study objective	Study design	Sample size, n (% women)	Age (y)	Physical activity component	Technology device	Supervised intervention and duration of the intervention	Session details
Choi and Lee [35], 2019, South Korea	To examine the effects of virtual kayak paddling exercises using real-world video recordings on postural control, muscle performance, and cognitive function in older adults with mild cognitive impairment	RCT	60 (85)	Total sample: NR (range 69-85); intervention group: mean 77.3 (SD 4.4; rage NR); control group: mean 75.4 (SD 4.0; range NR)	Interval training and balance and coordination exercises	Video projected on a screen	• Yes • 6 wk	2 sessions of 60 minutes per week
Domingos et al [36], 2022, Portugal	To examine the acceptability and safety of delivering dual-task programs in an online group format with people with Parkinson disease in early to late stages of the disease	Pretest-posttest with a control group	15 (60)	Mean 69.4 (SD 9.3; range NR)	Walking and functional training	NR	• Yes • 16 wk	2 sessions of 60 minutes
Dommes and Cavallo [37], 2012, France	To examine the effects of a training method combined with behavioral and educational interventions on street-crossing decisions by providing practice on a simulator	RCT	40 (57)	Mean 72.2 (SD 5.3; range NR)	Overall physical activity	Simulation laboratory	• Yes • 1 wk	2 sessions
Duque et al [38], 2013, Australia	To examine the effect of a virtual reality system to assess balance and provide a training system for balance in a population of community-dwelling older participants with a known history of falls	RCT	60 (total sample: NR; intervention group: 63; control group: 61)	Total sample: NR; intervention group: mean 79 (SD 10; range NR); control group: mean 75 (SD 8; range NR)	Balance and coordination exercises	Virtual reality system combining input from a force platform and virtual reality glasses containing a head tracker	• Yes • 6 wk	2 sessions of 30 minutes per week
Haeger et al [39], 2021, Germany	To examine the effects of a multicomponent approach on walking parameters and assess transfer effects on aspects of cognition, motivation, and control beliefs	RCT	34 (44.1)	Mean 75.0 (SD 3.7; range NR)	Various smartphone-based activities	Smartphone	• Yes • NR	2 sessions per week
Jansons et al [40], 2017, Australia		RCT	105 (64)	Total sample: NR; group at home: mean 66 (SD 13; range NR); group at the gym: mean 68 (SD 11; range NR)	Walking, running, weight training, stretching and flexibility, and cycling	Phone	• Yes • 12 mo	3 sessions of 60 minutes per week

XSL•FO
RenderX

Study, year, and country	Main study objective	Study design	Sample size, n (% women)	Age (y)	Physical activity component	Technology device	Supervised intervention and duration of the intervention	Session details
	To compare the effects on outcome measures of gym-based exercise versus home-based exercise with telephone follow-up among adults with chronic conditions who had completed a short-term exercise program supervised by a health professional							
Jurkeviciute et al [41], 2020, Italy and Sweden	To identify contextual factors that determine similarities and differences in the value of an eHealth intervention between 2 contexts	Pretest-posttest without a control group	Italy: 53 (51); Sweden: 54 (56)	Italy: mean 77.6 (SD 5.3); Sweden: mean 74.8 (SD 5.9)	The Otago fall prevention program	Tablet	• Yes • 6 mo	NR
King et al [42], 2020, United States	To examine whether counseling by a computer-based virtual advisor was not worse than counseling by trained human advisors for increasing 12-month walking levels among inactive older adults	Single-blind, cluster-randomized noninferiority parallel trial	245 (78.8)	Mean 62.3 (SD 8.4; range 50-87)	Walking	Computer	• Yes • 12 mo	Weekly sessions in the first 2 months, and twice-per-month sessions for the remaining 10 months
Masumoto et al [43], 2017, Japan	To quantitatively measure and visualize face-to-face interactions among older adults in an exercise program and examine relationships among interactional variables; personality; and interest in community involvement, including interactions with the local community	Pretest-posttest without a control group	27 (63)	Mean 73.5 (SD NR; range NR)	Stretching and flexibility and yoga	Karaoke-on-demand system with images projected on a screen	• Yes • 2 mo	4 sessions of 90 minutes
Nikitina et al [44], 2018, Russia	To examine the feasibility of home-based online group training under different group cohesion settings and its effects on adherence and well-being among older adults; in addition, to assess the effects of a technology-supported intervention on subjective well-being and loneliness	Pretest-posttest without a control group	60 (total sample: NR; pilot study 1: 95; pilot study 2: 100)		Weight training and balance and coordination exercises	Tablet and activity monitor	• Yes • 8 wk	3 sessions of 20-40 minutes per week

XSL•FO
RenderX

Study, year, and country	Main study objective	Study design	Sample size, n (% women)	Age (y)	Physical activity component	Technology device	Supervised intervention and duration of the intervention	Session details
				Total sample: NR (range 59-83); pilot study 1, individual group: mean 65 (SD 6; range NR); pilot study 1, interaction group: mean 68 (SD 8; range NR); pilot study 2, individual group: mean 69 (SD 7; NR); pilot study 2, interaction group: mean 68 (SD 6; NR)				
O'Brien et al [45], 2021, Ireland	To investigate the experiences and attitudes of older adults following a community-led walking program using activity trackers	Qualitative study	11 (100)	NR (range 60-80)	Walking	Wearable device	• Yes • 6 wk	Biweekly sessions
Pauly et al [46], 2019, Canada	To examine associations between portable ICT[c] use and changes in physical activity, loneliness, and executive functioning in older adults	Pretest-posttest without a control group	92 (64)	Mean 67.7 (SD 8.7; range 51-85)	Reporting daily physical activity	Tablet	• No • 5 wk	NR
Pischke et al [47], 2022, Germany	To compare the acceptance and effectiveness of 2 interventions for physical activity promotion among initially inactive community-dwelling older adults	Cross-over randomized trial	242 (66.2)	Mean 68.7 (SD 5.4; range 60-82)	Physical activity exercises, recommendations, and brochures	Smartphone	• Yes • 10 wk	1 session of 90 minutes per week
Qiu et al [48], 2023, China	To examine the effects of the Social Balance Ball exergame on intergenerational interactions and assess which factors affect intergenerational interactions in social balance training games	Pretest-posttest	36 (18 older adults; 72.2)	Older adults: mean 64.9 (SD 3.5; range NR)	Balance exercises	Balance board, Social Balance ball, television, and computer	• Yes • NR	NR
Richards et al [49], 2018, Canada		Single-case study	1 (0)	62	Treadmill walking	Virtual reality–coupled treadmill system	• Yes • 3 wk	9 sessions in 3 weeks

XSL•FO
RenderX

Study, year, and country	Main study objective	Study design	Sample size, n (% women)	Age (y)	Physical activity component	Technology device	Supervised intervention and duration of the intervention	Session details
	To show that virtual reality technology can be coupled with a self-paced treadmill to further improve walking competency in individuals after stroke							
Scase et al [50], 2017, United Kingdom	To design a gamified environment through which applications could be delivered to promote cognition, exercise, social interaction, and healthy eating and examine adherence to this technology solution through an intervention in which older people were asked to play serious games	Pretest-posttest without a control group	18 (78.6)	Total sample: NR; focus group 1: mean 77.0 (SD 7.5); focus group 2: mean 74.6 (SD 5.5); focus group 3: mean 78.5 (SD 1.9)	Stretching and flexibility	Tablet	• No • 7 wk	5 sessions per week; the average duration of the sessions was 28 minutes
Thiel et al [51], 2022, Germany	To examine the feasibility and effects of an intervention on combining smartphone-assisted group activities in the neighborhood with training in physical and cognitive skills on the social participation and connectedness of older adults	Noncontrolled proof-of-concept study	39 (85)	Total sample: mean 73.1 (SD 6.8; range NR); cycle group 1: mean 71.9 (SD 7.1; range NR); cycle group 2: mean 73.6 (SD 6.3; range NR)	Aerobic training, functional training, balance and coordination exercises, and stretching and flexibility	Smartphone	• Yes • 6 mo	1 mandatory session of 90 minutes per week plus additional nonmandatory activities
van Het Reve et al [52], 2014, Switzerland	To compare 3 different home-based training programs and their effects on measures of gait quality while considering adherence to the training program	Pretest-posttest preclinical exploratory trial	44 (NR)	Mean 75 (SD 6; NR)	Strength exercises and balance and coordination exercises	Tablet	• No • NR	2 sessions of resistance training and 5 sessions of 3 balance exercises
VanRavenstein and Davis [53], 2018, United States	To increase older adults' daily physical activity with the aim of decreasing chronic disease morbidity, disability, falls, and social isolation	Qualitative study	21 (90)	Total sample: NR (range 57-85); Garden Vistas group: mean 72.8 (SD 9.7; range 58-83); Garden North group: mean 72.3 (SD 7.9; range 57-85)	Walking, step climbing, and balance and coordination exercises	Wearable activity monitor	• Yes • 12 wk	2 sessions per week
Wilczynska et al [54], 2021, Australia		Pretest-posttest	59 (95)	Mean 62.3 (SD 11.6; range 50-82)	Walking, jogging, outdoor exercises, functional training, and aerobic training	Smartphone	• Yes • 20 wk	Sessions of 90 minutes

XSL•FO
RenderX

Study, year, and country	Main study objective	Study design	Sample size, n (% women)	Age (y)	Physical activity component	Technology device	Supervised intervention and duration of the intervention	Session details
	To conduct a pilot evaluation of the Ecofit intervention using a scalable implementation model among inactive older adults residing in an Australian rural community; to examine the preliminary effectiveness and feasibility of the Ecofit intervention in a "real-world" setting							

[a]NR not reported.

[b]RCT: randomized controlled trial.

[c]ICT: information and communications technology.

Environmental Factors and Their Role in Technology-Assisted PA Interventions

An overview of the environmental factors and their role in technology-assisted PA interventions for older adults is presented in Table 2. Of the 25 included studies, 3 (12%) included aspects of multiple environmental domains [41,47,51]. The environmental domain that was most often considered in the included studies was the social environment (19/25, 76%) [30-34,36,40-48,50-53], followed by the physical environment (8/25, 32%) [35,37-39,47,49,51,54], the socioeconomic environment (1/25, 4%) [41], and the systemic environment (1/25, 4%) [41].

Several specific social environmental factors were considered in the various technology-assisted PA interventions across the studies, including social connectedness [30,48,51] (including loneliness [30,33,46]), social interaction [34,43,50,52,53], social support [31,33,36,44,45,47], and delivery aspects (eg, home environment vs gym environment or the involvement of virtual advisors vs human advisors) [32,40,42]. Furthermore, a variety of specific physical environmental factors were also considered, including aspects of the street-crossing environment (eg, traffic speed) [37,49] and the neighborhood built environment (eg, walking paths; presence of benches; and hot spots, ie, highly frequented and meaningful nearby places, such as weekly markets and parks) [38,39,47,49,51,54]. A range of social (eg, household composition), socioeconomic (eg, costs of treatments), and systemic (eg, local preferences on the quality of patient care) environmental factors were considered in one study [41]. In total, 16% (4/25) of the studies were conducted in a simulated or virtual reality environment and included aspects of the physical environment [35,37,38,49]. The tools used to assess the environmental factors differed across the studies (Table 2).

The specific environmental factors fulfilled various roles in the technology-assisted PA interventions in the included studies. A total of 32% (8/25) of the studies included an environmental factor as an outcome. Of these 8 studies, 7 (88%) focused on a social environmental factor [30,34,36,43,45,46,48] and 1 (12%) focused on a physical environmental factor [39]. In 28% (7/25) of the studies, social [40,42,48,51] and physical [35,37,38,47] environmental factors were used as a comparator of the intervention; that is, different environmental backgrounds were compared, such as a virtual program and a face-to-face program, without direct measurement of environmental factors. In another 32% (8/25) of the studies, environmental factors were used as study results or factors influencing the intervention [31,42,44,46,50,51,53,54]. In 12% (3/25) of the studies, social [32,33] and physical [49] environmental factors were used as components of the intervention.

XSL•FO
RenderX

Table 2. An overview of technology-assisted physical activity interventions, the role of environmental factors, and outcomes in the included studies (N=25).

Study	Physical activity component and specific technology components	Environmental factor category, role of environmental factor, and specific environmental factor	Assessment tool of environmental factor	Outcomes	Outcome measurements	Main study results	Findings on environmental factors
Abe et al [30], 2023	• Walking • Kikoeru app on smartphone	• Social • Outcome • Social connectedness and loneliness	Qualitative interviews on social connectedness through app use	Social connectedness related to interactions through the app, subjective health, loneliness, and setting a target number of steps	Qualitative interviews on social connectedness and step goals; subjective health was measured using a VAS[a], and loneliness was measured using the Ando-Osada-Kodama Loneliness Scale	Social connectedness benefits were reported, loneliness decreased for 4/7 of the participants and remained stable for 3/7, target number of steps increased in 3/7 of the participants and remained stable in 4/7, and the intervention improved subjective health.	Participants experienced social connectedness and reduced loneliness after the intervention.
Alley et al [31], 2024	• Physical activity advice, stretching and flexibility, and strength exercises • Web-based program with 6 modules of computer-tailored physical activity advice and a Fitbit device	• Social • Moderator • Social support	Abbreviated Duke Social Support Index	Moderate to vigorous physical activity, engagement, and acceptability	ActiGraph GT9X wrist-worn accelerometer, website data and Google Analytics, and 9-item questionnaire	In participants with lower social support, both tailoring-only and Fitbit+tailoring participants increased their moderate to vigorous physical activity from baseline to the postintervention time point, whereas the control group decreased their physical activity. In comparison, all participants with higher social support regardless of group decreased their moderate to vigorous physical activity per day from baseline to the postintervention time point.	Among participants with lower social support, the Fitbit+tailoring participants but not the tailoring-only participants increased their moderate to vigorous physical activity more than the controls. Among participants with higher social support, no differences in moderate to vigorous physical activity changes were observed between groups. No significant (interaction) effects of social support and group were found on engagement and acceptability.
Anderson-Hanley et al [32], 2011	• Cycling • Cybercycle and computer	• Social • Social aspect was introduced as an intervention component • Virtual social facilitation	NR[b]	Pedaling effort	Cycling exercise effort (watts) captured in 10-second intervals by cybercycle sensors	Significant group (high vs low competitiveness) × time (before to after the introduction of the virtual avatar competitors) interaction; the virtual avatar increased exercise effort among high-competitiveness exercisers	Virtual social facilitation through introduction of avatar competitors increased exercise effort among more competitive exercisers.

XSL•FO
RenderX

Study	Physical activity component and specific technology components	Environmental factor category, role of environmental factor, and specific environmental factor	Assessment tool of environmental factor	Outcomes	Outcome measurements	Main study results	Findings on environmental factors
Boekhout et al [33], 2019	• Functional training and stretching and flexibility • Active Plus65 intervention—web based (website) and print based (delivered by mail)	• Social • Social aspect is part of the intervention components • Social support for physical activity and loneliness	Social support for physical activity—self-report (2 questions); loneliness—6-item De Jong Gierveld Loneliness Scale	Delivery mode preference and attrition	Delivery mode and attrition	Attrition differed significantly between the delivery modes—50% in the printed delivery mode and 71% in the web-based delivery mode.	Age and degree of loneliness were significant predictors of delivery mode preference. When adjusting for psychosocial variables, loneliness became nonsignificant, and social support for physical activity then emerged as a significant predictor with participants in the web-based delivery group who had higher levels of social support than those in the printed delivery group.
Chen et al [34], 2019	• Functional training, stretching and flexibility, and balance and coordination exercises • Telephone support	• Social • Secondary outcome • Social interaction	AIMS2-SF[c] society dimension	Primary: pain intensity and joint stiffness related to knee osteoarthritis; secondary: muscle strength of the lower limbs, balance, walking ability, and quality of life	Western Ontario and McMaster Universities Osteoarthritis Index	Pain and stiffness had a significantly stronger decrease in the intervention group than in the control group.	Quality of life and the society domain improved in the intervention group but not in the control group.
Choi and Lee [35], 2019	• Interval training and balance and coordination exercises • Video projected on a screen	• Physical • Virtual environment was compared to normal home exercise • Virtual kayak paddling and home environment	NR	Primary: static and dynamic postural balance; secondary: arm curl test, handgrip strength, and cognitive function	Static balance: 1-leg stance test and the Good Balance System; dynamic balance: Timed Up and Go Test, functional reach test, Berg Balance Scale, and Four Square Step Test	Significant improvement in balance components, motor capacity and function, and cognitive function in the virtual kayak paddling exercise group compared to the control group	Virtual kayak paddling was beneficial for balance, cognition, and muscle performance.
Domingos et al [36], 2022	• Walking and functional training • Online program (PD3 Move)	• Social • Outcome • Support from family members or caregivers during the sessions	Structured phone interview		Exit questionnaire sent via email to participants	Attendance rate and satisfaction were high.	Receiving support from family members or caregivers during the sessions was identified as a facilitator by participants.

XSL•FO
RenderX

Study	Physical activity component and specific technology components	Environmental factor category, role of environmental factor, and specific environmental factor	Assessment tool of environmental factor	Outcomes	Outcome measurements	Main study results	Findings on environmental factors
				Primary: attendance rate and satisfaction with the program; secondary: willingness to attend future online classes, perceived benefits of the program, feedback on format and delivery, and perceived difficulties and facilitators			
Dommes and Cavallo [37], 2012	• Overall physical activity • Simulation tool in simulation laboratory adapted to street-crossing situation	• Physical • Setting of the experiment • Street-crossing environment	NR	Street-crossing behavior	8 measures describing street-crossing behavior (eg, median accepted time gap between vehicles and collision)	Intervention group showed improvement in street-crossing decisions in (1-week) posttest assessment compared to controls. Differences disappeared 6 months after training.	On both postintervention tests (ie, 1 week and 6 months), the intervention and control groups still made more unsafe decisions when the car was approaching at a high speed and missed more crossing opportunities when a car was approaching at a low speed.
Duque et al [38], 2013	• Balance and coordination exercises • Balance training protocol with virtual reality system combining input from a force platform and virtual reality glasses containing a head tracker	• Physical • Setting of the experiment • Virtual training versus usual care (including optional Otago participation)	NR	Postural control, falls, fear of falling, gait, serum measurement, depression, and nutrition status	Posturography, retrospective questionnaire on falls, Survey of Activities and Fear of Falling in the Elderly, GAITRite assessment, venous blood, Geriatric Depression Scale, and Mini-Nutritional Assessment	Balance parameters were significantly improved in the BRU[d] training group. This effect was also associated with a significant reduction in falls and lower levels of fear of falling.	Significantly higher reduction in falls and fear of falling, improvement in balance parameters, and higher adherence to virtual training in the virtual training group compared to the control group.
Haeger et al [39], 2021	• Various smartphone-based activities • App on smartphone	• Physical • Outcome • Walking path and maximum distance from home	uFall smartphone app	Functional mobility, cognition outcomes, motivation, activity-related outcomes, and personality outcomes	6-Minute Walk Test, Instrumented Timed Up and Go Test, self-concordance and personality, System Usability Scale, Stroop test, and task-switching paradigm	No significant effects on any of the outcomes	No significant effect on Global Position System–based measures

XSL•FO
RenderX

Study	Physical activity component and specific technology components	Environmental factor category, role of environmental factor, and specific environmental factor	Assessment tool of environmental factor	Outcomes	Outcome measurements	Main study results	Findings on environmental factors
Jansons et al [40], 2017	• Walking, running, weight training, stretching and flexibility, and cycling • Telephone support	• Social • Comparison of different environments in which the intervention was carried out • Home environment and gym environment	NR	Primary: quality of life; secondary: productivity, social activity, depression and anxiety, motor capacity, physical activity, and attendance to community-based fitness center	Primary: EQ-5D; secondary: Health and Labour Questionnaire, Friendship Scale, Hospital Anxiety and Depression Scale, Phone-FITT[e], 6-Minute Walk Test, BMI, and 15-second sit-to-stand test	There were no significant differences between study groups in quality of life across the 12-month intervention period. The gym group showed slightly fewer symptoms of depression over the 12-month period than the home group.	No change was found in social isolation.
Jurkeviciute et al [41], 2020	• The Otago fall prevention program • Web-based portal on a tablet	• Social, socioeconomic, and systemic • Contextual factors moderating the intervention outcomes • Lifestyle habits of the population (eg, if they were living alone or with family), hourly rates of staff for delivering the intervention, organizational setup of the intervention, and local preferences on the quality of patient care	Semistructured interviews (patients and health care professionals) and monetary data from health care and technology providers	Cognitive performance, anxiety, perceived health care satisfaction, and monetary and nonmonetary benefits and sacrifices	Mini-Mental State Examination and the clock-drawing test, EQ-5D-5L, and VAS; other data from semistructured interviews and monetary data from health care and technology providers	In Sweden, patients improved cognitive performance, experienced a reduction in anxiety, and perceived their health as better, and both patients and health care professionals were satisfied with care. There were increased costs and higher workload for health care professionals. The intervention was not cost-efficient. In Italy, patients were satisfied with care, and the health care professionals felt empowered and had an acceptable workload. The intervention was cost-effective. There were no improvements in clinical efficacy and quality of life.	In total, 6 factors that influence eHealth interventions were identified: process of delivery, organizational structure and professionals involved, cost of different treatments, hourly rates of staff for delivering the intervention, lifestyle habits of the population, and local preferences on the quality of patient care.
King et al [42], 2020	• Walking • Program on computer	• Social • Comparison of different modes of delivery • Virtual advisor and human advisor	NR	Primary: total walking time; secondary: moderate to vigorous physical activity, sedentary behavior, BMI, resting blood pressure and heart rate, and well-being	CHAMPS[f] questionnaire and Vitality Plus Scale	The 12-month change in walking was more pronounced in the virtual advisor cohort compared to the human advisor cohort. There were improvements in both arms regarding clinical risk factors, sedentary behavior, and well-being.	The virtual advisor produced significant 12-month walking increases for older, low-income Latino adults that were no worse than the significant improvements achieved with human advisors.

XSL•FO
RenderX

Study	Physical activity component and specific technology components	Environmental factor category, role of environmental factor, and specific environmental factor	Assessment tool of environmental factor	Outcomes	Outcome measurements	Main study results	Findings on environmental factors
Masumoto et al [43], 2017	• Stretching and flexibility and yoga • The DK Elder System, with a karaoke-on-demand system with images projected on a screen	• Social • Primary and secondary outcomes • Communication networks of participants in the exercise program, time of interactions, number of persons interacted with, interaction among inhabitants, and community involvement	Business Microscope for the primary outcome and environmental factors survey for the secondary outcome	Primary: communication networks of participants in the exercise program, time of interactions, and number of persons interacted with; secondary: interaction among inhabitants and community involvement	Business Microscope (ie, name tag–type wearable sensor node with a built-in infrared signal transmitter and receiver to collect data on the face-to-face interactions of participants)	Network density in the initial session was low but increased as the number of sessions increased. Density in the third session was greater than in the final session (ie, increasing the number of sessions does not necessarily lead to promotion of more face-to-face interactions).	Significant enhancement of interest in interacting with local community residents. Marginally significant enhancement of interest in community involvement, communication time, and number of communication partners.
Nikitina et al [44], 2018	• Weight training and balance and coordination exercises • App; Gymcentral app program; Otago exercise program on a tablet and activity monitor	• Social • Moderator • Social support	Medical Outcomes Study Social Support Survey	Primary: usability and acceptance of the program; secondary: adherence to the program	Usability: System Usability Scale; acceptance: questionnaire	Online group exercising was proven feasible among healthy, independently living older adults in Russia.	Physical training performed in a virtual environment positively affected life satisfaction but not loneliness. High-cohesion groups were preferable for group exercise, and social support was a good predictor for adherence in the low-cohesion condition.
O'Brien et al [45], 2021	• Walking • Program and wearable device	• Social • Outcome • Views on the social dimension of the program	Interview	Program as a source of motivation, user experiences with the technology, and views on the social dimension of the program	Qualitative interview	The program and activity trackers were useful in maintaining motivation to stay active. Social support was considered a useful component.	Social support was considered a useful component.
Pauly et al [46], 2019	• Reporting daily physical activity • App (iDialogPad) on tablet	• Social • Social loneliness as an outcome and social function as a moderator • Social loneliness and social function	Social loneliness: revised UCLA[g] Loneliness Scale; social function: list of ICT[h] functions including social function	Changes in physical activity, loneliness, and executive functioning	International Physical Activity Questionnaire, revised UCLA Loneliness Scale, and Trail Making Test—part B		More frequent use of the social component was associated with more social loneliness of the participants.

XSL•FO
RenderX

Study	Physical activity component and specific technology components	Environmental factor category, role of environmental factor, and specific environmental factor	Assessment tool of environmental factor	Outcomes	Outcome measurements	Main study results	Findings on environmental factors
						No change in physical activity over ≥6 months; time spent sitting decreased. More frequent exercise was associated with more moderate physical activity intensity and less sitting. More frequent use of the social component was associated with more social loneliness.	
Pischke et al [47], 2022	• Physical activity exercises, recommendations, and brochures • App on smartphone and Fitbit device	• Physical and social • Background variable • Physical activity neighborhood environment, neighborhood environment, walking environment, and social support for engaging in physical activity	Physical activity neighborhood environment scale, neighborhood scales, walking environment, activity-related support from family and friends (modified), and activity-related social support	Primary: moderate to vigorous physical activity and sedentary behavior; secondary: subjective health and technology commitment, use, and experience	Triaxial accelerometers (ActiGraph GT3X+), SF-12[i] (1 item), and self-generated items	Moderate to vigorous physical activity increased between baseline and T1 (if unadjusted) and decreased between baseline and T2 regardless of the intervention group. A total of 18.6% of the participants met physical activity recommendations at baseline, 16.4% met physical activity recommendations at T1, and 20.3% met physical activity recommendations at T2. For sedentary behavior, there were no significant differences or effects at T1 or T2. Intervention acceptance was high.	NR
Qiu et al [48], 2023	• Balance exercises • Exergame, balance board, Social Balance Ball, television, and computer	• Social • Outcome • Social networks—social presence and connectedness	Networked Minds Social Presence Inventory, Inclusion of Other in the Self Scale, and Lubben Social Network Scale	Mode of exergames (social presence), role of participants (older or younger), and gender	Mode of exergames and demographic information	Higher levels of social interaction and positive feelings in player modes with human interaction compared to the virtual player mode without human interaction	Social interaction within player modes was associated with more positive experiences as opposed to modes in which no interaction was possible.
Richards et al [49], 2018	• Treadmill walking • Virtual reality–coupled treadmill system		Pictures, virtual reality training settings, and Assessment of Life Habits scale mobility score	Walking competence			

XSL•FO
RenderX

Study	Physical activity component and specific technology components	Environmental factor category, role of environmental factor, and specific environmental factor	Assessment tool of environmental factor	Outcomes	Outcome measurements	Main study results	Findings on environmental factors
		• Physical • Part of the intervention • Street crossing, corridor walking, park stroll, terrain changes, and moving obstacles			5-Meter Walk Test, 6-Minute Walk Test, Berg Balance Scale, Activities-Specific Balance Confidence Scale, Assessment of Life Habits scale, and personal appraisal	Control protocol training and virtual reality training resulted in a similar progression through the training sessions of total time walked on the treadmill. Virtual reality training led to additional increase in gait speed and 6-Minute Walk Test distance as well as improved balance self-efficacy and anticipatory locomotor adjustments.	Virtual reality training was superior to the control protocol training for improving motor capacity, balance self-efficacy, and anticipatory locomotor adjustments.
Scase et al [50], 2017	• Stretching and flexibility • App (DOREMI) on tablet	• Social • The social interaction element enhanced well-being • Social interaction	Thematic analysis for social aspects	Adherence to the intervention	Number of sessions and mean session length	Significant group differences in engagement in game sessions related to different social arrangements. Gamified environment can help engage with computer-based applications. Social community factors influenced long-term adherence.	Social community factors influenced adherence to the intervention. Bonding and sense of community between participants supported engagement.
Thiel et al [51], 2022	• Aerobic training, functional training, balance and coordination exercises, and stretching and flexibility • App (designed for the Quartier Agil program) on smartphone	• Physical and social • Facilitator • Hot spots (highly frequented and meaningful sites) in the neighborhood environment	Hot spots were identified in group discussions; social participation facet of the World Health Organization Quality of Life Instrument–Older Adults	Physical function, balance, leg strength, and physical activity level	6-Minute Walk Test, Berg Balance Scale, isometric leg strength, and sensor-based moderate to vigorous physical activity	Combined physical and cognitive training supported by technical devices (smartphones) appears feasible.	The hot spots can be considered a facilitator of the intervention.
van Het Reve et al [52], 2014	• Strength exercises and balance and coordination exercises • Active Lifestyle app on tablet	• Social • Social aspect in one group compared to other intervention modalities • Social interaction	Number of dispatched messages to a bulletin board (within the social group only)	Gait analysis, physical performance, fear of falling, and adherence	Walking analysis (GAITRite), Short Physical Performance Battery, Falls Efficacy Scale International, and compliance recordings	Tablet groups showed significant improvements in gait parameters and adherence compared to the brochure group but not in physical performance.	Program adherence was highest in the group with a social aspect compared to other intervention modalities.

XSL•FO
RenderX

Study	Physical activity component and specific technology components	Environmental factor category, role of environmental factor, and specific environmental factor	Assessment tool of environmental factor	Outcomes	Outcome measurements	Main study results	Findings on environmental factors
VanRavenstein and Davis [53], 2018	• Walking, step climbing, and balance and coordination exercises • Telehealth, exercise program (Otago), and Fitbit device	• Social • Moderator of intervention effect • Social isolation and depression	Qualitative analysis to form categories about socialization	Mobility	Self-Efficacy for Exercise Scale, 30-second sit-to-stand test, Mini-Balance Evaluation Systems test, Berg Balance Scale, and 2-Minute Walk Test	Successful implementation of telehealth physical therapy–led intervention to increase physical activity. Social isolation and depression need to be addressed to encourage successful aging in place.	Physical activity and socialization are critical to older adults who are aging in place. Mental health needs to be considered when attempting to engage older adults in group activities.
Wilczynska et al [54], 2021	• Walking, jogging, outdoor exercises, functional training, and aerobic training • Ecofit app on smartphone	• Physical • Facilitator of the intervention • Outdoor built environmental characteristics (ie, railings, stairs, benches, and parks)	No assessment; built environment was used in the intervention.	Aerobic fitness, functional mobility, and upper- and lower-body muscular fitness	6-Minute Walk Test, Timed Up and Go Test, arm curl test, and chair stand test	Significant improvements in aerobic fitness, functional mobility, and upper- and lower-body muscular fitness at 6 and 20 weeks.	The Ecofit program makes use of simple infrastructure (ie, railings, stairs, and benches) and can be adapted to outdoor locations.

[a]VAS: visual analog scale.

[b]NR: not reported.

[c]AIMS2-SF: Arthritis Impact Measurement Scales 2–Short Form.

[d]BRU: balance rehabilitation unit.

[e]Phone-FITT: brief physical activity interview for older adults.

[f]CHAMPS: Community Health Activities Model Program for Seniors.

[g]UCLA: University of California, Los Angeles.

[h]ICT: information and communications technology.

[i]SF-12: 12-item Short Form Health Survey.

Intervention Outcomes by Environmental Factor

As described in Table 2, the technology-assisted PA interventions in most studies (19/25, 76%) were evaluated positively [30,32-38,41-43,45,48-54]. The included environmental factors played a supportive role in achieving this effect in several of these studies in which a between-group comparison was made [33-38,42]. In some studies, no effects (3/25, 12%) [31,39,40], mixed effects (2/25, 8%) [44,47], or adverse effects (1/25, 4%) of the interventions or environmental factors (eg, moderators) were reported [46].

The studies that considered social environmental factors in the technology-assisted PA interventions (12/25, 48%) focused on a variety of outcomes, including PA and exercise [30-32,46]; mental health outcomes [30,32,46]; quality of life [34,40]; physical performance [52,53]; physical health outcomes [42]; social interaction network [43]; and aspects such as delivery, acceptability, usability, adherence, attrition, satisfaction, experiences, and motivation [31,33,36,44,45,48,50,52]. The studies that considered physical environmental factors (7/25, 28%) also focused on various outcomes, including cognition [35,39], physical performance [35,38,39,49,51,54], adherence [38], and street-crossing decision-making [37]. The 12% (3/25) of the studies that considered aspects of multiple environmental domains also focused on various outcomes, including PA [47,51]; mental health [41,47]; cognition [41]; physical performance [51]; costs [41]; and aspects such as satisfaction, commitment, experience, and acceptability [47].

In several studies (5/25, 20%), the technology-assisted PA interventions improved social environmental factors, such as social connectedness [30], social interaction networks [43], and other societal factors [34]. In one study, the technology-assisted PA intervention did not change social isolation in older adults [40]. In the study by Pauly et al [46], more frequent use of the social component of an app was associated with more loneliness in older adults. In various studies (7/25, 28%), social interaction and support were found to increase the exercise effort in more competitive but not in less competitive older adults [32], as well as adherence [36,50,52,53] and the positive experiences [48] of older adults in technology-assisted PA interventions. In

XSL•FO
RenderX

contrast, in the study by Alley et al [31], older adults with lower baseline levels of social support increased their moderate to vigorous PA more than those with higher levels of social support.

In various studies (5/25, 20%), physical environmental factors were considered in a virtual or simulated environment [35,37,38,49]. Compared to usual home exercises, virtual kayak paddling was beneficial for balance, muscle performance, and cognition [35]. In a simulated street-crossing situation, the intervention group improved street-crossing decisions compared to controls, but these differences disappeared 6 months after training [37]. On postintervention tests, the intervention group and the control group both made more unsafe decisions when a car approached at a high speed and missed more crossing opportunities when a car was approaching at a low speed compared to the preintervention test [37]. In the study by Duque et al [38], there were significantly fewer falls and lower levels of fear of falling in older adults who received balance training in a virtual environment compared to a control group that received usual care. The study by Richards et al [49] showed that virtual reality training was superior to the control protocol training for improving motor capacity, balance self-efficacy, and anticipatory locomotor adjustments.

One study showed that a technology-assisted PA intervention that made use of features of the outdoor built environment (community hot spots such as parks and markets) improved physical function in older adults [51]. Through the definition of hot spots in the neighborhood, older adults were encouraged to carry out certain activities with their peers when meeting there. However, other technology-assisted PA interventions that considered such physical environmental factors did not significantly affect functional mobility [39] or only showed limited effectiveness on PA outcomes [47]. The study by Jurkeviciute et al [41] indicated that several social, socioeconomic, and systemic environmental factors influenced the effects of an eHealth intervention focusing on cognitive performance, mental health outcomes, satisfaction, and costs.

Discussion

Principal Findings

Overview

With this work, we provide an overview of the role of environmental factors in technology-assisted PA interventions for the target group of older adults. It has been suggested previously that environmental aspects play a crucial role in modulating human behavior, especially regarding PA [55]. Emerging evidence has shown that PA should not be seen as an isolated entity but—apart from being an intrinsically motivated behavior—also as an output and reaction to environmental circumstances and surroundings [56]. These surroundings can have many shapes as they can be differentiated in many ways. In this work, we chose to partition the potentially numerous environmental influences into physical, social, socioeconomic, and systemic environmental factors. This provided the opportunity for a comprehensive description of a variety of aspects that may (or may not) interact with technology-assisted interventions to increase or optimize PA in older adults. The fact that 25 studies from almost all continents were identified highlights that researchers worldwide are aware of the relevance of environmental factors in such settings. However, this number of studies is not high, which indicates that this topic is still emerging. As expected, these studies were highly variable in terms of design, sample size, and intervention characteristics.

Social Environment

The social environment was the most frequently addressed environmental domain as it was considered in 76% (19/25) of the studies [30-34,36,40-48,50-53]. A total of 37% (7/19) of these studies used aspects of the social environment as an outcome [30,34,36,43,45,46,48]. The main reason for this probably lies in the (partially empirically substantiated) expectation that PA is related to increased social participation [20] and reduced loneliness [57]. However, whether this is the case would need further empirical substantiation. Previous research has shown that PA interventions per se do not necessarily come with the advantage of additional social benefits [58], but technological solutions might add benefits related to (online) connectedness and communication options. Mixed results on social environmental factors were found, showing benefits for social connectedness [30], social interaction networks [43], and other societal factors [44] but not for social isolation [40,46]. An important social aspect seems to be social comparison or role models, which may have been the reason for increased exercise effort [32] and adherence [36,50,52]. However, it has to be acknowledged that social comparison was not assessed as such in these studies. Nonetheless, it may be that, through seeing peers performing well or better than oneself, people tend to put more effort into their exercise and PA behavior [59]. Furthermore, social support has been shown to be important for adherence and maintaining motivation [44,45] and to be a significant predictor of delivery preference (ie, people who receive higher social support for PA were more likely to prefer a web-based delivery than a printed delivery) [33]. However, there were also results showing that lower social support at baseline may come with more room for benefits in moderate to vigorous PA when applying tailored online exercise advice [31]. The study by King et al [42] suggests that delivery aspects can also affect intervention outcomes. In this study, counseling by virtual advisors significantly increased walking in low-income, Latino older adults, which was comparable to the significant improvements achieved with human advisors [43].

Physical Environment

The second most considered environmental domain was the physical environment (8/25, 32% of the studies) [35,37-39,47,49,51,54]. In 50% (4/8) of these studies, physical environmental factors were considered in a virtual or simulated environment [35,37,38,49]. It is an important finding that physical and cognitive capacities such as balance, muscle performance, and cognition can be enhanced when exercising under these virtual or simulated environmental conditions [35,37,38,49]. These effects even translated to a reduction in the number of falls and fear of falling in one study compared

XSL•FO
RenderX

to controls who received usual care [38]. There seems to be a huge potential and growing empirical foundation for using augmented and virtual reality–based exercise applications [60]. The constant development and sophistication of these technologies in terms of physical experience and aspects of (online) communication highlight the manifold potential for future application in research and beyond. However, there are several barriers to the use of such technologies when targeting older adults that may hamper their success [61]. Beyond virtual environments, physical environmental factors from the "real world" were also considered, mainly concerning the neighborhood built environment (eg, walking paths, hot spots, and benches) [38,39,47,49,51,54].

Socioeconomic and Systemic Environment

An investigation of socioeconomic and systemic environmental factors was carried out in the study by Jurkeviciute et al [41], where both aspects were considered as moderators of the intervention effect. These authors found several factors that influence eHealth intervention effects, including the process of delivery, organizational structure and professionals involved, treatment costs including hourly rates of staff for delivering the intervention, lifestyle habits of the population, and local preferences regarding quality of patient care [41]. While these findings are from one study only and have to be interpreted with caution, they do underline that intervention effects often are a product of their setting and the professionals (eg, trainers and therapists) and participants involved. More research is needed on the role of socioeconomic and systemic environmental factors in technology-assisted PA interventions to draw more concrete conclusions on this issue.

Implications for Future Research, Practice, and Policy

On the basis of our findings, it becomes apparent that there is a significant potential for a better understanding of intervention results and better tailoring of intervention programs when including environmental aspects in research endeavors. Environmental factors demonstrated multifaceted effects on intervention outcomes, albeit sometimes contradictory. A thorough understanding of the underlying mechanisms is still lacking as the use and applications of these environmental factors remain scarce. Future research should elucidate the causal pathways through which environmental factors exert their effects, considering potential mediators and moderators that may influence intervention outcomes. In particular, social aspects related to intervention delivery and group dynamics seem to have the potential to reveal important mechanisms that could positively enhance technology-assisted PA interventions for older adults. The same applies to the physical environment. We found that physical and even cognitive capacity may benefit largely from exercise in virtual or simulated environments, showing the large potential for augmented and virtual reality exercises. As this is a field growing at high speed, further work should be carried out to define more specific methodologies and target outcomes especially for the population of older adults. Considering that young and middle-aged adults are using these kinds of technologies more and more, the potential for interventions targeted at PA will be enormous. As such technologies become increasingly accessible and user-friendly, their integration into PA interventions holds promise for enhancing engagement and adherence and obtaining positive outcomes among older adults. As older adults are not left out of these developments and show more acceptance of such technologies [62], intervention development can be expected to be more inclusive of virtual environments in the future.

Strengths and Limitations

We followed a well-defined methodology for this scoping review. Although we conducted a detailed search in various established databases, we might have missed potentially relevant papers (eg, from gray literature or conference proceedings not yet published as full papers). It also has to be acknowledged that there is no clear framework for categorizing environmental factors into specific domains (ie, social, physical, socioeconomic, and systemic). The framework we decided to use was based on thorough discussions within the group of researchers involved. We chose not to evaluate the quality of the studies included in this scoping review as the objective of this work was to explore the role of environmental factors in technology-assisted PA interventions among older adults and not to assess the quality of the studies. As a result, the study findings need to be interpreted with caution, especially those of studies without randomization procedures or controls. The comprehensive summary of methods in Table 1 provides all the information in this regard. As technology-assisted PA interventions can be expected to grow in number and application in the following years, evaluating the quality of the evidence is undoubtedly warranted at some point in the future.

Conclusions

In conclusion, the results of this scoping review addressing the fast-growing domain of technology-assisted PA interventions for older adults show that the role of environmental factors is still emerging in this field. The studies predominantly focused on social environmental factors, followed by physical environmental factors. Studies that integrate socioeconomic and systemic environmental factors in technology-assisted PA interventions were scarce. Important findings were that the included environmental factors played a supportive role in achieving beneficial effects of technology-assisted PA interventions. The studies reviewed exhibited heterogeneity in how environmental factors were incorporated—some studies incorporated them as integral components of the experimental design, whereas in other studies, they served as effect modifiers or outcomes. Drawing from the results, there is a significant potential for a better understanding of intervention outcomes and better tailoring of intervention programs when systematically including environmental aspects in technology-assisted PA interventions for older adults.

XSL•FO
RenderX

Acknowledgments

This publication is based upon work from COST Action CA20104 – Network on evidence-based physical activity in old age (PhsAgeNet), supported by COST (European Cooperation in Science and Technology). Michael Brach (Institute of Sport and Exercise Sciences, University of Münster, Münster, Germany) enabled collaboration as leader of this European Cooperation in Science and Technology Action. EJT is supported by a Dutch Research Council Gravitation grant (Exposome-NL; 024.004.017). The funding bodies played no role in the study design; data collection, analysis, and interpretation; manuscript preparation; or decision to submit the manuscript for publication.

Authors' Contributions

CPJ, EP, and EJT coordinated the study. All authors conceptualized the study. All authors contributed to the title and abstract screening phase of this scoping review. CPJ, VA, EM-T, ZP, MR, ST, and EJT contributed to the full-text screening phase of this scoping review. CPJ, DN, JMO, KEK, EM-T, MR, E-MT, and EJT contributed to the data extraction phase of this scoping review. CPJ and EJT drafted the manuscript; all authors are responsible for the overall content of the manuscript. DN, JMO, VA, KEK, EM-T, ZP, MR, ST, E-MT, and EP provided feedback on the manuscript and helped with interpreting the data. All authors approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 532 KB - mhealth_v13i1e59570_app1.pdf]

Multimedia Appendix 2
Detailed search strategy for each bibliographic database.
[DOCX File , 42 KB - mhealth_v13i1e59570_app2.docx]

References

1. Lear SA, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, et al. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. Lancet 2017 Dec 16;390(10113):2643-2654. [doi: 10.1016/S0140-6736(17)31634-3] [Medline: 28943267]
2. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
3. Pahor M, Guralnik JM, Ambrosius WT, Blair S, Bonds DE, Church TS, et al. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. JAMA 2014 Jun 18;311(23):2387-2396 [FREE Full text] [doi: 10.1001/jama.2014.5616] [Medline: 24866862]
4. Gomes M, Figueiredo D, Teixeira L, Poveda V, Paúl C, Santos-Silva A, et al. Physical inactivity among older adults across Europe based on the SHARE database. Age Ageing 2017 Jan 20;46(1):71-77 [FREE Full text] [doi: 10.1093/ageing/afw165] [Medline: 28181637]
5. Kohl HW3, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, et al. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul 21;380(9838):294-305 [FREE Full text] [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
6. Timmermans EJ, Hoogendijk EO, Broese van Groenou MI, Comijs HC, van Schoor NM, Thomése FC, et al. Trends across 20 years in multiple indicators of functioning among older adults in the Netherlands. Eur J Public Health 2019 Dec 01;29(6):1096-1102 [FREE Full text] [doi: 10.1093/eurpub/ckz065] [Medline: 31008512]
7. Timmermans EJ, Visser M, Wagtendonk AJ, Noordzij JM, Lakerveld J. Associations of changes in neighbourhood walkability with changes in walking activity in older adults: a fixed effects analysis. BMC Public Health 2021 Jul 06;21(1):1323 [FREE Full text] [doi: 10.1186/s12889-021-11368-6] [Medline: 34225681]
8. McCormack GR, Patterson M, Frehlich L, Lorenzetti DL. The association between the built environment and intervention-facilitated physical activity: a narrative systematic review. Int J Behav Nutr Phys Act 2022 Jul 14;19(1):86 [FREE Full text] [doi: 10.1186/s12966-022-01326-9] [Medline: 35836196]
9. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012 Nov 21;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
10. Nigg CR. Technology's influence on physical activity and exercise science: the present and the future. Psychol Sport Exerc 2003;4(1):57-65 [FREE Full text] [doi: 10.1016/S1469-0292(02)00017-1]

XSL•FO
RenderX

11. Western MJ, Armstrong ME, Islam I, Morgan K, Jones UF, Kelson MJ. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021 Nov 09;18(1):148 [FREE Full text] [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
12. Dermody G, Whitehead L, Wilson G, Glass C. The role of virtual reality in improving health outcomes for community-dwelling older adults: systematic review. J Med Internet Res 2020 Jun 01;22(6):e17331 [FREE Full text] [doi: 10.2196/17331] [Medline: 32478662]
13. Sohaib Aslam A, van Luenen S, Aslam S, van Bodegom D, Chavannes NH. A systematic review on the use of mHealth to increase physical activity in older people. Clin eHealth 2020;3:31-39. [doi: 10.1016/j.ceh.2020.04.002]
14. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
15. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
16. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
17. Bauman AE, Sallis JF, Dzewaltowski DA, Owen N. Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. Am J Prev Med 2002 Aug;23(2 Suppl):5-14. [doi: 10.1016/s0749-3797(02)00469-5] [Medline: 12133733]
18. Barnett DW, Barnett A, Nathan A, van Cauwenberg J, Cerin E, Council on Environment and Physical Activity (CEPA) – Older Adults Working Group. Built environmental correlates of older adults' total physical activity and walking: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 07;14(1):103. [doi: 10.1186/s12966-017-0558-z] [Medline: 28784183]
19. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. Annu Rev Public Health 2006 Apr 01;27(1):297-322. [doi: 10.1146/annurev.publhealth.27.021405.102100] [Medline: 16533119]
20. Lindsay Smith G, Banting L, Eime R, O'Sullivan G, van Uffelen JG. The association between social support and physical activity in older adults: a systematic review. Int J Behav Nutr Phys Act 2017 Apr 27;14(1):56 [FREE Full text] [doi: 10.1186/s12966-017-0509-8] [Medline: 28449673]
21. Colom A, Mavoa S, Ruiz M, Wärnberg J, Muncunill J, Konieczna J, et al. Neighbourhood walkability and physical activity: moderating role of a physical activity intervention in overweight and obese older adults with metabolic syndrome. Age Ageing 2021 May 05;50(3):963-968 [FREE Full text] [doi: 10.1093/ageing/afaa246] [Medline: 33219673]
22. Huang WY, Huang H, Wu CE. Physical activity and social support to promote a health-promoting lifestyle in older adults: an intervention study. Int J Environ Res Public Health 2022 Nov 03;19(21):14382 [FREE Full text] [doi: 10.3390/ijerph192114382] [Medline: 36361256]
23. Cleland CL, Tully MA, Kee F, Cupples ME. The effectiveness of physical activity interventions in socio-economically disadvantaged communities: a systematic review. Prev Med 2012 Jun;54(6):371-380. [doi: 10.1016/j.ypmed.2012.04.004] [Medline: 22521997]
24. Bock C, Jarczok MN, Litaker D. Community-based efforts to promote physical activity: a systematic review of interventions considering mode of delivery, study quality and population subgroups. J Sci Med Sport 2014 May;17(3):276-282. [doi: 10.1016/j.jsams.2013.04.009] [Medline: 23693030]
25. Welch WA, Spring B, Phillips SM, Siddique J. Moderating effects of weather-related factors on a physical activity intervention. Am J Prev Med 2018 May;54(5):e83-e89 [FREE Full text] [doi: 10.1016/j.amepre.2018.01.025] [Medline: 29551330]
26. Taylor J, Walsh S, Kwok W, Pinheiro MB, de Oliveira JS, Hassett L, et al. A scoping review of physical activity interventions for older adults. Int J Behav Nutr Phys Act 2021 Jun 30;18(1):82 [FREE Full text] [doi: 10.1186/s12966-021-01140-9] [Medline: 34193157]
27. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Timmermans E. The role of environmental factors in technology-assisted physical activity intervention studies in older adults: a scoping review. Open Science Framework Registries. 2023 Mar 20. URL: https://osf.io/r7qvz [accessed 2025-03-04]
29. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
30. Abe H, Kamishima T, Ojima R, Onishi R, Hirano M. Social and physical effects of a pedometer and communication application among older men: a mixed-methods, pre/post pilot study. Inf Commun Soc 2021 Jul 14;26(3):459-478 [FREE Full text] [doi: 10.1080/1369118x.2021.1954231]

XSL•FO
RenderX

31. Alley SJ, Schoeppe S, Moore H, To QG, van Uffelen J, Parker F, et al. The moderating effect of social support on the effectiveness of a web-based, computer-tailored physical activity intervention for older adults. J Health Psychol 2024 Apr 15:13591053241241840 [FREE Full text] [doi: 10.1177/13591053241241840] [Medline: 38618999]
32. Anderson-Hanley C, Snyder AL, Nimon JP, Arciero PJ. Social facilitation in virtual reality-enhanced exercise: competitiveness moderates exercise effort of older adults. Clin Interv Aging 2011;6:275-280 [FREE Full text] [doi: 10.2147/CIA.S25337] [Medline: 22087067]
33. Boekhout JM, Peels DA, Berendsen BA, Bolman C, Lechner L. A web-based and print-delivered computer-tailored physical activity intervention for older adults: pretest-posttest intervention study comparing delivery mode preference and attrition. J Med Internet Res 2019 Aug 28;21(8):e13416 [FREE Full text] [doi: 10.2196/13416] [Medline: 31464186]
34. Chen H, Zheng X, Huang H, Liu C, Wan Q, Shang S. The effects of a home-based exercise intervention on elderly patients with knee osteoarthritis: a quasi-experimental study. BMC Musculoskelet Disord 2019 Apr 09;20(1):160 [FREE Full text] [doi: 10.1186/s12891-019-2521-4] [Medline: 30967131]
35. Choi W, Lee S. The effects of virtual kayak paddling exercise on postural balance, muscle performance, and cognitive function in older adults with mild cognitive impairment: a randomized controlled trial. J Aging Phys Act 2019 Dec 01;27(4):861-870 [FREE Full text] [doi: 10.1123/japa.2018-0020] [Medline: 31185775]
36. Domingos J, Dean J, Fernandes JB, Godinho C. An online dual-task cognitive and motor exercise program for individuals with Parkinson disease (PD3 move program): acceptability study. JMIR Aging 2022 Dec 22;5(4):e40325 [FREE Full text] [doi: 10.2196/40325] [Medline: 36548037]
37. Dommes A, Cavallo V. Can simulator-based training improve street-crossing safety for elderly pedestrians? Transp Res F Traffic Psychol Behav 2012 Mar;15(2):206-218 [FREE Full text] [doi: 10.1016/j.trf.2011.12.004]
38. Duque G, Boersma D, Loza-Diaz G, Hassan S, Suarez H, Geisinger D, et al. Effects of balance training using a virtual-reality system in older fallers. Clin Interv Aging 2013;8:257-263 [FREE Full text] [doi: 10.2147/CIA.S41453] [Medline: 23467506]
39. Haeger M, Bock O, Zijlstra W. [Smartphone-based health promotion in old age : an explorative multi-component approach to improving health in old age]. Z Gerontol Geriatr 2021 Mar;54(2):146-151. [doi: 10.1007/s00391-020-01700-x] [Medline: 32052186]
40. Jansons P, Robins L, O'Brien L, Haines T. Gym-based exercise and home-based exercise with telephone support have similar outcomes when used as maintenance programs in adults with chronic health conditions: a randomised trial. J Physiother 2017 Jul;63(3):154-160 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.018] [Medline: 28655559]
41. Jurkeviciute M, van Velsen L, Eriksson H, Lifvergren S, Trimarchi PD, Andin U, et al. Identifying the value of an eHealth intervention aimed at cognitive impairments: observational study in different contexts and service models. J Med Internet Res 2020 Oct 08;22(10):e17720 [FREE Full text] [doi: 10.2196/17720] [Medline: 33064089]
42. King AC, Campero MI, Sheats JL, Castro Sweet CM, Hauser ME, Garcia D, et al. Effects of counseling by peer human advisors vs computers to increase walking in underserved populations: the COMPASS randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1481-1490 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4143] [Medline: 32986075]
43. Masumoto K, Yaguchi T, Matsuda H, Tani H, Tozuka K, Kondo N, et al. Measurement and visualization of face-to-face interaction among community-dwelling older adults using wearable sensors. Geriatr Gerontol Int 2017 Oct;17(10):1752-1758. [doi: 10.1111/ggi.12933] [Medline: 28112476]
44. Nikitina S, Didino D, Baez M, Casati F. Feasibility of virtual tablet-based group exercise among older adults in Siberia: findings from two pilot trials. JMIR Mhealth Uhealth 2018 Feb 27;6(2):e40 [FREE Full text] [doi: 10.2196/mhealth.7531] [Medline: 29487045]
45. O'Brien J, Mason A, Cassarino M, Chan J, Setti A. Older women's experiences of a community-led walking programme using activity trackers. Int J Environ Res Public Health 2021 Sep 17;18(18):9818 [FREE Full text] [doi: 10.3390/ijerph18189818] [Medline: 34574741]
46. Pauly T, Lay JC, Kozik P, Graf P, Mahmood A, Hoppmann CA. Technology, physical activity, loneliness, and cognitive functioning in old age. GeroPsych 2019 Sep;32(3):111-123 [FREE Full text] [doi: 10.1024/1662-9647/a000208]
47. Pischke CR, Voelcker-Rehage C, Ratz T, Peters M, Buck C, Meyer J, et al. Web-based versus print-based physical activity intervention for community-dwelling older adults: crossover randomized trial. JMIR Mhealth Uhealth 2022 Mar 23;10(3):e32212 [FREE Full text] [doi: 10.2196/32212] [Medline: 35319484]
48. Qiu S, Kaisar E, Ding R, Han T, Hu J, Rauterberg M. Social balance ball: designing and evaluating an exergame that promotes social interaction between older and younger players. Int J Hum Comput Interact 2023 Feb 16;40(11):2838-2861 [FREE Full text] [doi: 10.1080/10447318.2023.2175157]
49. Richards CL, Malouin F, Lamontagne A, McFadyen BJ, Dumas F, Comeau F, et al. Gait training after stroke on a self-paced treadmill with and without virtual environment scenarios: a proof-of-principle study. Physiother Can 2018;70(3):221-230 [FREE Full text] [doi: 10.3138/ptc.2016-97] [Medline: 30275647]
50. Scase M, Marandure B, Hancox J, Kreiner K, Hanke S, Kropf J. Development of and adherence to a computer-based gamified environment designed to promote health and wellbeing in older people with mild cognitive impairment. Stud Health Technol Inform 2017;236:348-355. [Medline: 28508817]

XSL•FO
RenderX

51. Thiel C, Günther L, Osterhoff A, Sommer S, Grüneberg C. Feasibility of smartphone-supported, combined physical and cognitive activities in the neighbourhood for stimulating social participation of the elderly. BMC Geriatr 2022 Jul 30;22(1):629 [FREE Full text] [doi: 10.1186/s12877-022-03303-0] [Medline: 35907804]
52. van Het Reve E, Silveira P, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: part 2 of a phase II preclinical exploratory trial. J Med Internet Res 2014 Jun 25;16(6):e159 [FREE Full text] [doi: 10.2196/jmir.3055] [Medline: 24966165]
53. VanRavenstein K, Davis BH. When more than exercise is needed to increase chances of aging in place: qualitative analysis of a telehealth physical activity program to improve mobility in low-income older adults. JMIR Aging 2018 Dec 21;1(2):e11955 [FREE Full text] [doi: 10.2196/11955] [Medline: 31518250]
54. Wilczynska M, Jansson AK, Lubans DR, Smith JJ, Robards SL, Plotnikoff RC. Physical activity intervention for rural middle-aged and older Australian adults: a pilot implementation study of the ecofit program delivered in a real-world setting. Pilot Feasibility Stud 2021 Mar 24;7(1):81 [FREE Full text] [doi: 10.1186/s40814-021-00823-1] [Medline: 33757587]
55. Sallis JF, Spoon C, Cavill N, Engelberg JK, Gebel K, Parker M, et al. Co-benefits of designing communities for active living: an exploration of literature. Int J Behav Nutr Phys Act 2015 Feb 28;12:30 [FREE Full text] [doi: 10.1186/s12966-015-0188-2] [Medline: 25886356]
56. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. Am J Prev Med 2009 Apr;36(4 Suppl):S99-123.e12 [FREE Full text] [doi: 10.1016/j.amepre.2009.01.005] [Medline: 19285216]
57. Pels F, Kleinert J. Loneliness and physical activity: a systematic review. Int Rev Sport Exerc Psychol 2016 Jun 07;9(1):231-260 [FREE Full text] [doi: 10.1080/1750984x.2016.1177849]
58. Tcymbal A, Abu-Omar K, Hartung V, Bußkamp A, Comito C, Rossmann C, et al. Interventions simultaneously promoting social participation and physical activity in community living older adults: a systematic review. Front Public Health 2022;10:1048496 [FREE Full text] [doi: 10.3389/fpubh.2022.1048496] [Medline: 36568739]
59. Reicherzer L, Kramer-Gmeiner F, Labudek S, Jansen CP, Nerz C, Nystrand MJ, et al. Group or individual lifestyle-integrated functional exercise (LiFE)? A qualitative analysis of acceptability. BMC Geriatr 2021 Feb 01;21(1):93 [FREE Full text] [doi: 10.1186/s12877-020-01991-0] [Medline: 33522904]
60. Ren Y, Lin C, Zhou Q, Yingyuan Z, Wang G, Lu A. Effectiveness of virtual reality games in improving physical function, balance and reducing falls in balance-impaired older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr 2023 May;108:104924. [doi: 10.1016/j.archger.2023.104924] [Medline: 36680968]
61. Harris MT, Blocker KA, Rogers WA. Older adults and smart technology: facilitators and barriers to use. Front Comput Sci 2022 May 4;4:835927 [FREE Full text] [doi: 10.3389/fcomp.2022.835927]
62. Jokisch MR, Schmidt LI, Doh M. Acceptance of digital health services among older adults: findings on perceived usefulness, self-efficacy, privacy concerns, ICT knowledge, and support seeking. Front Public Health 2022;10:1073756. [doi: 10.3389/fpubh.2022.1073756] [Medline: 36582385]

Abbreviations

PA: physical activity
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by L Buis; submitted 16.04.24; peer-reviewed by SQ Yoong, R Dubbeldam; comments to author 05.09.24; revised version received 07.10.24; accepted 22.01.25; published 13.03.25.

Please cite as:
Jansen CP, Nijland D, Oppert JM, Alcan V, Keskinen KE, Matikainen-Tervola E, Pajalic Z, Rantakokko M, Tomsone S, Tuomola EM, Portegijs E, Timmermans EJ
The Role of Environmental Factors in Technology-Assisted Physical Activity Intervention Studies Among Older Adults: Scoping Review
JMIR Mhealth Uhealth 2025;13:e59570
URL: https://mhealth.jmir.org/2025/1/e59570
doi:10.2196/59570
PMID:

©Carl-Philipp Jansen, Désirée Nijland, Jean-Michel Oppert, Veysel Alcan, Kirsi E Keskinen, Emmi Matikainen-Tervola, Zada Pajalic, Merja Rantakokko, Signe Tomsone, Essi-Mari Tuomola, Erja Portegijs, Erik J Timmermans. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,

XSL•FO
RenderX

distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effects of Remote Web-Based Interventions on the Physiological and Psychological States of Patients With Cancer: Systematic Review With Meta-Analysis

Lv Tian[1], PhD; Yixuan Wen[1], MS; Jingmiao Li[1], MS; Jiexin Guan[1], MS; Tao Li[1], MS; Jun Fan[1,2], PhD

[1]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China

[2]Shimmer Center, Tianfu Jiangxi Laboratory, Chengdu, China

Corresponding Author:
Jun Fan, PhD
Institute of Fundamental and Frontier Sciences
University of Electronic Science and Technology of China
No. 2006, Xiyuan Avenue, High-tech Zone (West District)
Chengdu, 611731
China
Phone: 86 17390940726
Email: fan.jun@uestc.edu.cn

Abstract

Background: Patients with cancer may experience physiological and psychological adverse reactions, such as fatigue, pain, anxiety, and depression, which seriously affect their quality of life. Research has shown that remote interventions based on apps or miniprograms may help improve the physiological and mental health of patients with cancer. However, due to the limited number of relevant studies, the impact of web-based interventions in cancer management remains unclear.

Objective: We aimed to determine the efficacy of interventions based on apps, miniprograms, or other web-based tools on the physiological (body pain and fatigue) and psychological (anxiety and depression) states and the quality of life of patients with cancer.

Methods: We conducted electronic literature searches in PubMed, Scopus, Web of Science, the Cochrane Library, CINAHL, and EMBASE databases. The search period spanned from the inception of each database to October 15, 2024. Two researchers independently conducted literature retrieval and data extraction. The risk of bias was assessed with the Cochrane risk-of-bias tool, and the quality of evidence was assessed according to the Grading of Recommendations Assessment, Development, and Evaluation (GRADE). All statistical analyses were performed using Review Manager version 5.4.

Results: A total of 36 randomized controlled trials were included. The remote web-based interventions significantly improved the pain intensity (n=14, 39% studies; standardized mean difference [SMD] –0.39, 95% CI –0.64 to –0.14; I^2=82%; GRADE rating=low) and fatigue status (n=11, 31% studies; SMD –0.52, 95% CI –0.95 to –0.09; I^2=95%; GRADE rating=low) in patients with cancer. Regarding psychology, the results indicated that the remote web-based interventions significantly improved the anxiety (n=14, 39% studies; SMD –0.60, 95% CI –0.90 to –0.30; I^2=91%; GRADE rating=low) and depressive state (n=10, 28% studies; SMD –0.36, 95% CI –0.58 to –0.14; I^2=81%; GRADE rating=low) of patients with cancer. For quality of life, the results showed that the remote web-based interventions had a significant positive impact on the quality of life of patients with cancer (n=26, 72% studies; SMD 0.63, 95% CI 0.39-0.87; I^2=92%; GRADE rating=low).

Conclusions: The remote web-based interventions were effective in reducing the intensity of physiological pain, relieving fatigue, improving depression and anxiety, and improving the quality of life of patients with cancer. However, due to the low certainty of evidence, more rigorous randomized controlled trials are needed to validate these findings further.

Trial Registration: PROSPERO CRD42024611768; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024611768

(JMIR Mhealth Uhealth 2025;13:e71196) doi:10.2196/71196

KEYWORDS

meta-analysis; cancer patients; pain; depression; anxiety; quality of life; remote web-based intervention; PRISMA

XSL•FO
RenderX

Introduction

Cancer is a highly prevalent disease globally. It is estimated that by 2025, more than 20 million people will be affected by cancer [1,2]. Cancer also poses a severe global health challenge, subjecting patients to immense physiological and psychological stress [3-5]. Although there are various treatment methods, such as surgery, radiotherapy, chemotherapy, targeted therapy, and immunotherapy, cancer remains an incurable disease that requires continuous care [6]. Unfortunately, these treatments are often accompanied by side effects, such as pain, fatigue, and sleep disorders [7]. There is a close and complex inherent relationship between the physiological condition and psychological state of patients with cancer [8-12]. Physiologically, the discomfort symptoms experienced by patients, such as pain, fatigue, and nausea, can trigger a chain reaction psychologically, leading to the emergence of negative emotions like anxiety and depression, as well as a significant increase in stress levels, thus seriously affecting the patient's psychological states. At the same time, psychological distress and negative emotions can further exacerbate physiological discomfort symptoms, impede treatment compliance, and consequently have a negative impact on the overall quality of life of patients. The traditional model for managing side effects in patients with cancer mainly relies on face-to-face medical consultations, but this model may lack comprehensiveness or reliability [13]. For example, face-to-face medical consultations are restricted by geographic locations and medical resources. Patients in remote areas find it difficult to obtain timely and high-quality medical consultation services, and adverse reactions cannot be effectively managed. Time constraints make it challenging for physicians to comprehensively collect information on patients' adverse reactions in their daily lives. Meanwhile, patients may forget some symptoms due to nervousness, which affect their comprehensive assessment of their conditions.

With the rapid development of digital information technology, many health management miniprograms that can be directly used on supporting platforms, such as WeChat and Alipay, have been developed. These miniprograms rely on the app software ecosystem and are characterized by lightweight design and convenience. Therefore, remote web-based interventions based on apps or miniprograms have become a promising treatment solution [2]. Remote web-based intervention is a medical support model based on digital technology, enabling symptom management through specialized applications, such as health management apps for patients with cancer, medical miniprograms, and web-based platforms such as professional medical websites. The related applications for remote web-based interventions have become promising tools for supporting patients' self-management and improving health outcomes due to their acceptability, convenience, accessibility, personalization, and cost control [14]. This intervention model breaks through the spatiotemporal limitations of traditional health care, providing comprehensive services for patients with cancer, including health education, psychological support, symptom monitoring, and rehabilitation guidance. For example, research has shown that mobile health (mHealth) interventions can improve medication adherence, self-management participation, and health outcomes [15,16]. Through remote web-based interventions, patients can obtain various forms of support at home and receive medical services more cost-effectively [17,18].

In recent years, the management of side effects in patients with cancer through remote web-based interventions has received increasing attention. However, in current research, the efficacy of remote web-based interventions on the adverse reactions of patients with cancer is inconsistent. Regarding the management of physiological pain in patients with cancer, some studies have shown that remote web-based interventions can improve the intensity of physiological pain [19-22], yet other studies have pointed out that remote web-based interventions did not significantly improve the physiological pain of patients with cancer [23-25]. Some even indicated that remote web-based interventions could exacerbate the physiological pain of patients with cancer [26]. Regarding improving the psychological state of patients with cancer, some studies suggest that remote web-based interventions can improve the psychological state of patients with cancer [27,28], while others draw different conclusions [29-31]. In addition, remote web-based interventions have different impacts on the quality of life of patients with cancer [20,28,32,33]. However, research on remote web-based interventions in the physiological and mental management of patients with cancer remains limited. There is a lack of comprehensive and systematic reviews to reflect the impact of remote web-based interventions on patients with cancer. Therefore, based on these conflicting research results, we conducted a meta-analysis to clarify the clinical efficacy of remote web-based interventions on the physiological (pain and fatigue) and psychological (anxiety and depression) states and the quality of life of patients with cancer.

Methods

Study Design

This systematic review and meta-analysis was registered in PROSPERO (CRD42024611768), and the study was reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [34].

Search Strategy

We conducted electronic literature searches in PubMed, Scopus, Web of Science, Cochrane Library, CINAHL, and EMBASE from the inception of each database to October 15, 2024. The search terms consisted of subject headings and free terms. The full search strategy and other relevant information are provided in Table S1 in Multimedia Appendix 2. Meanwhile, manual searches were carried out according to reference lists of relevant articles.

Eligibility Criteria

Literature searches were limited to published English-language articles. Eligible studies had to meet the inclusion criteria provided in Textbox 1.

Studies meeting the inclusion criteria provided in Textbox 1 were included in this review.

XSL•FO
RenderX

Textbox 1. Inclusion criteria.

- Population: the target population had to be patients with cancer, regardless of the type of cancer.
- Intervention: the intervention group adopted remote web-based interventions based on apps, miniprograms, or other web-based platforms.
- Comparison: the intervention in the control group was routine care or other non–web-based interventions.
- Outcomes: at least one outcome indicator had to be included—pain, anxiety, depression, fatigue, and quality of life (mean and SD).
- Study design: the type of studies included had to be randomized controlled trials.

Study Selection

Search results from all databases were imported into EndNote X9 (Clarivate) software to remove duplicate studies and manage the research. Two independent reviewers (LT and YW) screened the titles and abstracts of the studies according to the inclusion criteria. After identifying potentially eligible studies, the full texts were screened again to determine the final included studies. Then, the 2 reviewers (LT and YW) discussed and reached a consensus. A third reviewer (JF) was available to assist and mediate throughout the process.

Data Extraction

Data extraction followed the PRISMA statement to ensure a systematic approach. Two reviewers (LT and YW) independently extracted data using a pretested data extraction form, and a third author (JF) cross-checked the accuracy of the results. The data extracted included publication information (author, research year, and country); study design; characteristics of the study participants (sample size, age, sex, and type of cancer); duration of intervention; intervention methods in the intervention group and the control group, respectively; primary outcomes and assessment scales (methods); and main results. The primary outcome data were extracted as mean and SDs.

Quality Assessment and Certainty of Evidence

Two authors (LT and YW) independently evaluated the risk of bias, methodological quality, and certainty of evidence of the included studies. In case of inconsistencies, discussions were held with a third reviewer (JF) until a consensus was reached. The Revised Cochrane Risk-of-Bias Tool was used to assess the risk of bias from the following aspects: randomization process, deviation from the intended interventions, missing outcome data, outcome measurement, and selection of reported results [35]. The overall quality of evidence for each outcome was rated according to the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) guidelines [36]. Considering that only randomized controlled trials (RCTs) were included, each outcome was initially assigned a high quality. Two reviewers (LT and YW) used the GRADE system to evaluate the quality of evidence, and potential disagreements were resolved through discussions with a third reviewer (JF). Table S2 in Multimedia Appendix 2 presents the GRADE downgrading details.

Data Synthesis and Analysis

Due to the expected heterogeneity among the studies, we used a random-effects model to conduct a pooled estimate of the postintervention effects. Using the inverse-variance method, we calculated the standardized mean differences (SMDs) and 95% CIs for continuous variables. The effect sizes were interpreted as small (0.2), medium (0.5), or large (0.8). Statistical heterogeneity was evaluated using I^2, which was classified as low (I^2<25%), moderate (I^2=25%-50%), high (I^2=50%-75%), and considerably high (I^2>75%). Publication bias was assessed by calculating Begg and Egger test values, and P<.05 was considered publication bias [37,38]. A leave-one-out sensitivity analysis was performed to explore the robustness of the pooled effect [39]. We performed all statistical analyses using Review Manager (version 5.4; Cochrane) and Stata (version 17; StataCorp LLC).

Results

Compliance With the Registered Protocol

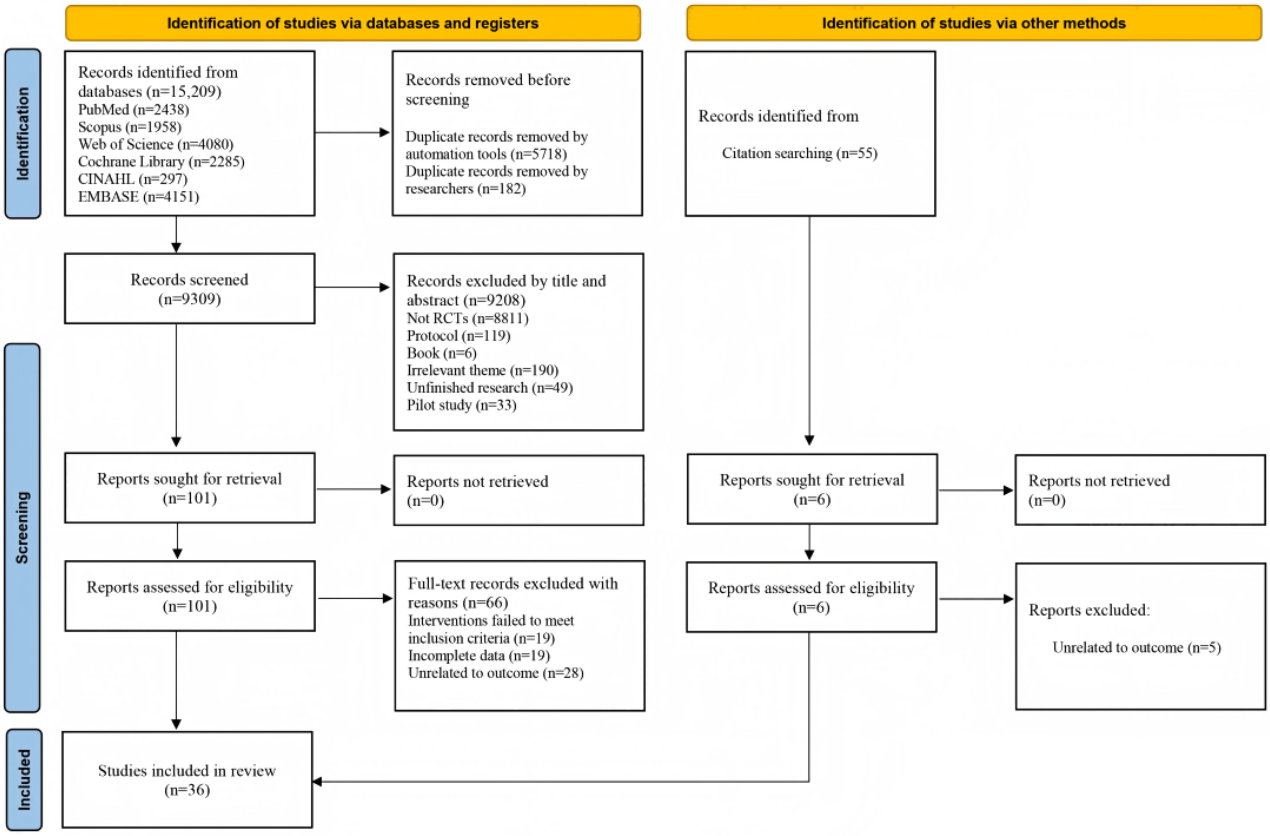
In the initial stage of the research design, our focus was on the impact of remote web-based interventions on physiological states (pain and fatigue) and the quality of life of patients with cancer. When we extracted the relevant data from the included studies, we found that there were controversies regarding the efficacy of remote web-based interventions on the anxiety and depression of patients with cancer. Although anxiety and depression were not listed as outcome indicators in the initial registration protocol, they are closely related to the research topic. Considering that anxiety and depression are common indicators of the psychological state of patients with cancer and have a significant impact on their lives, anxiety and depression cannot be ignored when evaluating the effect of remote web-based interventions. Therefore, to enrich the research content and enhance the clinical significance of the results, compared with the registered protocol, we have added outcome indicators of anxiety and depression. There were no other inconsistencies with the preregistration protocol.

Study Selection

Figure 1 illustrates the flow of this study and the reasons for exclusion. A total of 15,209 records were identified from the databases. After removing duplicate literature, 9309 (61.2%) potentially eligible studies were determined. After screening the titles and abstracts, we retrieved 101 studies for a more in-depth review. Eventually, 35 (34.6%) RCTs met the inclusion criteria. Among the 66 excluded studies, 19 (29%) had intervention methods that did not meet the inclusion criteria, 19 (29%) lacked complete data, and 28 (42%) did not contain relevant outcome indicators. In addition, we tracked and supplemented relevant citations. Finally, 36 studies met the inclusion criteria for our meta-analysis [15,19-33,40-59].

XSL•FO
RenderX

Figure 1. PRISMA flowchart for study selection. RCT: randomized controlled trial.



Study Characteristics

A total of 36 RCTs were included in this study [15,19-33,40-59]. The study sites involved countries such as China, Iran, the United States, Spain, the Netherlands, Japan, the United Kingdom, Australia, and Germany. The sample size of the studies varied from 29 to 417, with 4521 patients with cancer included. The age range of patients with cancer was wide (>18 years). The patients with cancer covered various types of cancers, including non–small cell lung cancer, breast cancer, urinary system cancers, digestive system cancers, acute leukemia, esophageal cancer, ovarian cancer, cervical cancer, uterine cancer, vaginal cancer, and mixed cancer types. The apps, miniprograms, or other web-based platforms used in the intervention group included WeChat, WhatsApp, ePAL, Microsoft Teams, Zoom, KLIK Pijnmonitor, Welby My Carte ONC, ASyMS, MyBreastHealth, Kanker Nazorg Wijzer, Breast Cancer Support Zone program, the miniprogram “Electronic Intestinal Safety,” e-CUIDATE system, Breast Cancer Self-Management Support app, ENCOURAGE project, Steps Toward Improving Diet and Exercise website, the website named “Together,” Shafayar, Mika, and Untire mobile app. The intervention duration ranged from 1 week to 18 months, and most of the studies focused on 3 to 12 months. The outcome assessment indicators included anxiety, depression, pain, fatigue, and quality of life. More information about the main results of each study is provided in Table 1.

XSL•FO
RenderX

Table 1. Characteristics of the included studies (randomized controlled trials).

Study and year	Country	Participants				Intervention time	Intervention methods		Main outcomes (assessments)	Main results	
		Sample size, n	Age (y), mean (SD)	Sex, n (%)	Types of cancer		Remote web-based intervention group[a]	Control group		Remote web-based intervention group, mean (SD)	Control group, mean (SD)
Sui et al [47], 2020	China	200	IG[b]: 61.37 (11.21); CG[c]: 62.35 (9.98)	Male: 164 (82); female: 36 (18)	Non–small cell lung cancer	12 mo	WeChat	Usual care	Anxiety (HADS[d]); depression (HADS); quality of life (EORTC QLQ-C30[e])	Anxiety: 5 (2.84); depression: 5.22 (2.77); quality of life: 74.44 (12.06)	Anxiety: 6.69 (4.01); depression: 6.55 (3.42); quality of life: 70.26 (17.29)
Peng et al [27], 2020	China	302	IG: 55.6 (6.8); CG: 56.3 (7)	Male: 144 (47.7); female: 158 (52.3)	Not specified	3 d	WeChat	Usual care	Pain (NRS[f]); anxiety (GAD-7[g]); depression (PHQ-9[h]); quality of life (not specified)	Pain: 2.5 (1.2); anxiety: 9.8 (2.2); depression: 10.1 (2.8); quality of life: 46.3 (5.5)	Pain: 2.8 (1.1); anxiety: 12.4 (3.1); depression: 11.2 (2.2); quality of life: 43.4 (5.2)
Bandani-Susan [57], 2022	Iran	38	IG: 46.79 (12.28); CG: 45.89 (7.64)	Male: 0 (0); female: 38 (100)	Breast cancer	7 wk	WhatsApp	Usual care	Fatigue (CFS[i])	Fatigue: 26.52 (10.32)	Fatigue: 28.16 (9.82)
Xu et al [44], 2021	China	126	IG: 48.74 (6.88); CG: 47.13 (7.58)	Male: 0 (0); female: 126 (100)	Breast cancer	9 mo	WeChat	Usual care	Quality of life (FACT-B[j])	Quality of life: 84.39 (8.54)	Quality of life: 72.11 (8.85)
Hao et al [52], 2022	China	80	IG: 64.8 (12.58); CG: 63.25 (17.16)	Not specified	Urinary system cancers (prostate, testicular, penile, kidney, and ureter, bladder)	3 mo	WeChat	Usual care	Quality of life (EORTC QLQ-C30)	Quality of life: 74.63 (4.25)	Quality of life: 67.20 (3.26)
Zhao et al [21], 2024	China	42	51-75[k]	Male: 23 (55); female: 19 (45)	Multiple cancers	4 wk	WeChat	Usual care	Pain (NRS)	Pain: 1.86 (0.79)	Pain: 2.76 (1)
Wang et al [28], 2022	China	72	≥18[k]	Male: 40 (55); female: 32 (45)	Acute leukemia	12 wk	WeChat	Usual care	Anxiety (HADS); depression (HADS); quality of life (FACT-Leu[l]); fatigue (BFI-C[m])	Anxiety: 11.00 (1.72); depression: 12.88 (2.21); quality of life: 65.34 (6.6); fatigue: 2.51 (0.54)	Anxiety: 16.29 (2.16); depression: 16.29 (2.98); quality of life: 71.55 (9.2); fatigue: 4.62 (1.34)

XSL•FO
RenderX

Study and year	Country	Participants				Intervention time	Intervention methods		Main outcomes (assessments)	Main results	
		Sample size, n	Age (y), mean (SD)	Sex, n (%)	Types of cancer		Remote web-based intervention group[a]	Control group		Remote web-based intervention group, mean (SD)	Control group, mean (SD)
Wang et al [22], 2024	China	125	IG: 40.29 (19.73); CG: 40.53 (22)	Male: 0 (0); female: 125 (100)	Breast cancer	6 mo (24 wk)	WeChat	Usual care	Pain (NRS); quality of life (FACT-B); fatigue (NRS)	Pain: 1.72 (1.76); quality of life: 108.69 (13.58); fatigue: 2.36 (1.46)	Pain: 2.86 (2.12); quality of life: 93.28 (13.01); fatigue: 1.97 (2.39)
Zheng et al [43], 2022	China	150	Trial 1—IG: 57.48 (9.29) and CG: 58.46 (9.96); Trial 2—IG: 59.50 (10.67) and CG: 58.46 (9.96)	Male: 123 (82); female: 27 (18)	Digestive system cancers (colon, rectum, stomach, liver, and esophagus)	4 wk	WeChat	Usual care	Anxiety (HADS); depression (HADS)	Trial 1—anxiety: 2.84 (2.61) and depression: 2.86 (2.8); Trial 2—anxiety: 2.68 (2.55) and depression: 3.32 (3.05)	Trial 1—anxiety: 4.5 (4.06) and depression: 4.58 (4.39); Trial 2—anxiety: 4.50 (4.06) and depression: 4.58 (4.39)
Yu et al [20], 2022	China	168	≥18[k]	Male: 139 (83); female: 29 (17)	Esophageal cancer	6 mo	Telephone and internet	Usual care	Pain (EORTC QLQ-C30); quality of life (EORTC QLQ-C30); fatigue (EORTC QLQ-C30)	Pain: 6.74 (12.74); quality of life: 81.93 (12.58); fatigue: 19.1 (11.24)	Pain: 14.84 (21.91); quality of life: 65.55 (16.76); fatigue: 37.67 (19.58)
Somers et al [24], 2016	United States	30	IG: 60 (11)	Male: 15 (50); female: 15 (50)	Cancers (breast, prostate, lung, and colorectal)	4 sessions over an unspecified period (designed to be weekly)	iPad (Skype)	Usual care	Pain (BPI[n])	Pain: 3.32 (1.89)	Pain: 2.05 (1.95)
Kamdar et al [31], 2024	United States	112	IG: 54.6 (10.9); CG: 50.7 (12.4)	Male: 57 (51); female: 55 (49)	Multiple cancers	8 wk	ePAL	Usual care	Pain (BPI); anxiety (GAD-7)	Pain: 3.16 (1.78); anxiety: 6.71 (5.07)	Pain: 4.28 (1.99); anxiety: 5.03(3.67)
Bilmiç et al [26], 2023	Turkey	110	≥18[k]	Male: 37 (34); female: 73 (66)	Multiple cancers	4 wk	Microsoft Teams, Zoom, and WhatsApp	Usual care	Pain (BPI); anxiety (ESAS[o]); depression (ESAS); fatigue (ESAS)	Pain: 3.88 (3); anxiety: 4.52 (3.1); depression: 4.04 (3.4); fatigue: 4.77 (2.68)	Pain: 2.04 (2.26); anxiety: 4.2 (3.21); depression: 3.09 (3.05); fatigue: 4.67 (2.85)

XSL•FO
RenderX

Study and year	Country	Participants				Intervention time	Intervention methods		Main outcomes (assessments)	Main results	
		Sample size, n	Age (y), mean (SD)	Sex, n (%)	Types of cancer		Remote web-based intervention group[a]	Control group		Remote web-based intervention group, mean (SD)	Control group, mean (SD)
Martínez-Miranda et al [19], 2024	Spain	49	IG: 49.21 (5.91); CG: 50 (8.04)	Male: 0 (0); female: 49 (100)	Breast cancer	12 wk	WhatsApp and email	Usual care	Pain (BPI); quality of life (FACT-B)	Pain: 2.39 (1.72); quality of life: 83.62 (14.71)	Pain: 4.85 (1.53); quality of life: 65.5 (13.73)
Simon et al [49], 2024	Netherlands	158	IG: 7.5 (5.1); CG: 7.5 (5.4)	Male: 79 (50); female: 79 (50)	Hematology-oncology, neuro-oncology, and solid tumors	18 mo	KLIK Pijnmonitor	Usual care	Pain (BPI)	Pain: 3.2 (2.1)	Pain: 3.9 (2.8)
Okuyama et al [50], 2024	Japan	125	IG: 63.9 (7.8); CG: 62.7 (7.4)	Male: 1 (0.8); female: 124 (99.2)	Breast cancer	3 mo	Welby My Carte ONC	Usual care	Quality of life (FACT-B)	Quality of life: 106.8 (18.4)	Quality of life: 105.2 (18)
Kearney et al [51], 2009	United Kingdom	112	IG: 55.1 (10.6); CG: 56.9 (10.5)	Male: 26 (23); female 86 (77)	Cancers (breast, lung, and colorectal cancer)	12-16 wk	ASyMS[p]	Usual care	Fatigue (combination scale[q])	Fatigue: 1.54 (1.11)	Fatigue: 1.82 (1.09)
Visser et al [46], 2018	Netherlands	109	IG: 55.8 (8.3); CG: 57.9 (8.8)	Male: 0 (0); female: 109 (100)	Breast cancer	3 mo	Tablet-based web app	Usual care	Quality of life (EORTC-QLQC30)	Quality of life: 75.28 (16.74)	Quality of life: 73.33 (15.43)
Zhu et al [40], 2018	China	114	IG: 46.2 (8.5); CG: 47.2 (8.3)	Male: 0 (0); female: 114 (100)	Breast cancer	12 wk	The app-based BCS[r] program	Usual care	Anxiety (HADS); depression (HADS); quality of life (FACT-B)	Anxiety: 9.93 (2.72); depression: 10.28 (2.46); quality of life: 92.87 (21.39)	Anxiety: 10.28 (2.46); depression: 12.58 (2.15); quality of life: 84.09 (15.99)
Zhou et al [42], 2020	China	111	IG: 49.84 (8.85); CG: 49.98 (9.84)	Male: 0 (0); female: 111 (100)	Breast cancer	6 mo	WeChat	Usual care	Pain (NRS); quality of life (FACT-B); fatigue (NRS)	Pain: 1.91 (2.07); quality of life: 106 (11.73); fatigue: 2.33 (2.09)	Pain: 2.71 (1.8); quality of life: 98.42 (16.59); fatigue: 1.75 (1.87)
Xia [45], 2020	China	155	18-70[k]	Male: 83 (54); female: 72 (46)	Low rectal cancer	3 mo	WeChat, blog, telephone, and QQ, among others	Usual care	Pain (Stoma Care Self-Efficacy Scale); anxiety (STAI[s]); quality of life (Stoma Care Self-Efficacy Scale); fatigue (Stoma Care Self-Efficacy Scale)	Pain: 26.59 (6.48); anxiety: 27.21 (6.88); quality of life: 47.59 (9.26); fatigue: 39.50 (3.26)	Pain: 33.75 (7.82); anxiety: 38.02 (6.33); quality of life: 45.84 (8.2); fatigue: 51.86 (4.56)

XSL•FO
RenderX

Study and year	Country	Participants				Intervention time	Intervention methods		Main outcomes (assessments)	Main results	
		Sample size, n	Age (y), mean (SD)	Sex, n (%)	Types of cancer		Remote web-based intervention group[a]	Control group		Remote web-based intervention group, mean (SD)	Control group, mean (SD)
Chen et al [33], 2024	China	91	IG: 48.09 (8.25); CG: 50.59 (6.10)	Male: 0 (0); female: 91 (100)	Breast cancer	7 wk	WeChat	Usual care	Anxiety (HADS); depression (HADS); quality of life (EORTC-QLQC30)	Anxiety: 7.59 (2.69); depression: 7.5 (2.54); quality of life: 79.15 (5.45)	Anxiety: 10.11 (1.94); depression: 11.11 (2.54); quality of life: 70.97 (7.35)
Balci and Faydali [58], 2024	Turkey	81	IG: 47.04 (10.21); CG: 54.24 (12.92)	Male: 0 (0); female: 81 (100)	Breast cancer	1 wk	App (MyBreast-Health)	Usual care	Quality of life (WHOQOL-BREF[t])	Quality of life: 72.26 (14.12)	Quality of life: 42.50 (14.38)
Çınar et al [56], 2021	Turkey	64	IG: 45.9 (8.3); CG: 45.5 (9.8)	Male: 0 (0); female: 64 (100)	Breast cancer	12 wk	Mobile app	Usual care	Quality of life (FACT-ES QLS[u])	Quality of life: 137.58 (16.94)	Quality of life: 123.7 (20)
Willems et al [30], 2017	Netherlands	409	IG: 56.26 (10.84); CG: 56.28 (11.45)	Male: 77 (18.8); female: 332 (81.2)	Multiple cancers	6 mo	Kanker nazorg wijzer (cancer aftercare guide)	Usual care	Anxiety (HADS); depression (HADS); quality of life (EORTC-QLQC30); fatigue (CIS[v])	Anxiety: 4.39 (3.89); depression: 2.82 (3.06); quality of life: 77.44 (18.17); fatigue: 55.9 (26.72)	Anxiety: 4.66 (3.97); depression: 3.53 (3.68); quality of life: 75.6 (20.58); fatigue: 61.77 (28.15)
Ghanbari et al [53], 2021	Iran	82	IG: 46.9 (9.83); CG: 46 (8.8)	Male: 0 (0); female: 82 (100)	Breast cancer	4 wk	Breast Cancer Support zone	Usual care	Anxiety (STAI)	Anxiety: 90.66 (13.84)	Anxiety: 106.92 (15.94)
Zhou et al [41], 2024	China	120	IG: 62.72 (7.91); CG: 61.78 (11.8)	Male: 79 (66); female: 41 (34)	Rectal cancer	3 mo	"e-bowel safety" applet	Usual care	Quality of life (EORTC-QLQC30)	Quality of life: 83.41 (2.46)	Quality of life: 78.71 (2.72)
Galiano-Castillo et al [54], 2016	Spain	76	IG: 47.4 (9.6); CG: 49.2 (7.9)	Male: 0 (0); female: 76 (100)	Breast cancer	8 wk	e-CUIDATE system	Usual care	Pain (BPI); quality of life (EORTC-QLQC30)	Pain: 2.53 (2.16); quality of life: 73.50 (20.76)	Pain: 4.12 (2.13); quality of life: 50.45 (20.78)
Hou et al [15], 2020	China	112	>20[k]	Male: 0 (0); female: 112 (100)	Breast cancer	3 mo	BCSMS[w] app	Usual care	Quality of life (EORTC-QLQC30)	Quality of life: 79.13 (15.31)	Quality of life: 79.49 (12.41)

XSL•FO
RenderX

Study and year	Country	Participants				Intervention time	Intervention methods		Main outcomes (assessments)	Main results	
		Sample size, n	Age (y), mean (SD)	Sex, n (%)	Types of cancer		Remote web-based intervention group[a]	Control group		Remote web-based intervention group, mean (SD)	Control group, mean (SD)
Admiraal et al [59], 2017	Netherlands	125	>18[k]	Male: 0 (0); female: 125 (100)	Breast cancer	12 wk	ENCOURAGE program	Usual care	Quality of life (EORTC-QLQC30)	Quality of life: 74.33 (19.42)	Quality of life: 73.41 (17.67)
Dong et al [23], 2019	China	50	IG: 48 (5.54); CG: 51.63 (7.49)	Male: 0 (0); female: 50 (100)	Breast cancer	12 wk	Phone step-recording app, social media apps	Usual care	Pain (SF-36[x]); quality of life (SF-36)	Pain: 77.31 (13.43); quality of life: 73.38 (18.16)	Pain: 79.58 (16.28); quality of life: 63.08 (18.9)
Frensham et al [25], 2018	Australia	91	IG: 65.2 (9.3); CG: 66.1 (9.4)	Male: 44 (48); female: 47 (52)	Multiple cancers	13 wk	STRIDE[y] website	Usual care	Pain (SF-36v2[z]); quality of life (SF–36v2)	Pain: 63.3 (18.5); quality of life: 69.5 (16.5)	Pain: 61.6 (20.3); quality of life: 67.4 (15.7)
Petzel et al [29], 2018	United States	29	IG: 59.6 (10); CG: 55.5 (8.4)	Male: 0 (0); female: 29 (100)	Ovarian cancer	2 mo	A website named "Together"	Usual care	Anxiety (HADS); depression (HADS)	Anxiety: 7.0 (3.6); depression: 4.9 (3.9)	Anxiety: 7.8 (2.7); depression: 3.2 (2.9)
Doosti et al [55], 2024	Iran	68	IG: 49.41 (8.8); CG: 50.26 (13.64)	Male: 0 (0); female: 68 (100)	Cancers (ovarian, cervical, uterus, and vaginal)	8 wk	Shafayar	Usual care	Anxiety (STAI); quality of life (EORTC-QLQC30)	Anxiety: 47.23 (12.38); quality of life: 73.78 (16.79)	Anxiety: 57.41 (12.72); quality of life: 62.41 (16.62)
Springer et al [32], 2024	Germany	218	IG: 55 (11); CG: 58 (10)	Male: 132 (61); female: 86 (39)	Multiple cancers	12 wk	Mika	Usual care	Anxiety (HADS); depression (HADS); quality of life (CGI-I[aa]); fatigue (FACIT-F[ab])	Results 1 (ITT[ac])—anxiety: 7.7 (4.3); depression: 6.4 (4.5); quality of life: 3.8 (1.1); fatigue: 32.2 (10.8) and Results 2 (PP[ad])—anxiety: 7 (4.3); depression: 5.6 (4.4); quality of life: 3.7 (1.2); fatigue: 34.2 (11.4)	Results 1 (ITT)—anxiety: 8.2 (4.1); depression: 7.5 (4.9); quality of life: 3.9 1.4); fatigue: 31.2 (13) and Results 2 (PP)—anxiety: 7.8 (4); depression: 7.1 (4.7); quality of life: 4.1 (1.3); fatigue: 32.9 (12.8)

XSL•FO
RenderX

Study and year	Country	Participants				Intervention time	Intervention methods		Main outcomes (assessments)	Main results	
		Sample size, n	Age (y), mean (SD)	Sex, n (%)	Types of cancer		Remote web-based intervention group[a]	Control group		Remote web-based intervention group, mean (SD)	Control group, mean (SD)
Spahrkäs et al [48], 2020	Australia, Canada, United Kingdom, and United States	417	IG: 56.7 (9.99); CG: 56.2 (9.42)	N/A[ae]	Multiple cancers	12 wk	Untire mobile app	Usual care	Quality of life (EORTC QLQ-30); fatigue (FSI[af])	Quality of life: 4.44 (1.27); fatigue: 5.81 (1.74)	Quality of life: 4.45 (1.14); fatigue: 6.2 (1.57)

[a]Remote web-based intervention group: Intervention methods were based on remote methods, such as relevant apps, applets, and websites, among others.

[b]IG: intervention group.

[c]CG: control group.

[d]HADS: Hospital Anxiety and Depression Scale.

[e]EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.

[f]NRS: Numerical Rating Scale.

[g]GAD-7: Generalized Anxiety Disorder-7.

[h]PHQ-9: Patient Health Questionnaire-9.

[i]CFS: Cancer Fatigue Scale.

[j]FACT-B: Functional Assessment of Cancer Therapy—Breast.

[k]Data reported as range.

[l]FACT-Leu: the Chinese version of the Functional Assessment of Cancer Therapy.

[m]BFI-C: the Chinese version of the Brief Fatigue Inventory-Leukemia.

[n]BPI: Brief Pain Inventory.

[o]ESAS: Edmonton Symptom Assessment Scale.

[p]ASyMS: Advanced Symptom Management System.

[q]Combination Scale: The scale is a symptom questionnaire that integrates the Common Terminology Criteria for Toxicity of Adverse Events grading system and the Chemotherapy Symptom Assessment Scale.

[r]BCS: breast cancer electronic support.

[s]STAI: State-Trait Anxiety Inventory.

[t]WHOQOL-BREF: World Health Organization Quality of Life Scale-Brief Version.

[u]FACT-ES QLS: Functional Assessment of Cancer Therapy—Endocrine Symptoms Quality of Life Scale.

[v]CIS: Checklist Individual Strength.

[w]BCSMS: Breast Cancer Self-Management Support.

[x]SF-36: 36-item Short Form Health Survey.

[y]STRIDE: Steps Toward Improving Diet and Exercise.

[z]SF-36v2: Short Form (36) Health Survey, version 2.

[aa]CGI-I: Clinical Global Impression—Improvement Scale.

[ab]FACIT-F: Functional Assessment of Chronic Illness Therapy—Fatigue.

[ac]ITT: intention-to-treat.

[ad]PP: per protocol.

[ae]N/A: not applicable.

[af]FSI: Fatigue Symptom Inventory.

Risk of Bias

As shown in Figure S1 in Multimedia Appendix 2 [15,19-33,40-59], a total of 25 (69%) studies were assessed as having an overall low risk [15,20-23,25-27,32,33,40-45,47,48,50,52,54-56,58,59], 7 (19%) studies were assessed as having some concerns [19,28-31,49,51], and 4 (11%) studies were assessed as having a high risk [24,46,53,57]. Except for 1 (3%) study [46], which was assessed as having a high risk, the remaining 35 (97%) studies reported an adequate randomization process and were evaluated as having a low risk of bias. Regarding deviation from the intended interventions, 35 (97%) studies were considered to have a low risk of bias,

XSL•FO
RenderX

and 1 (3%) study had some concerns [49]. Regarding missing outcome data, 28 (78%) studies were rated as having a low risk of bias, 6 (17%) had some concerns [19,28-31,51], and another 2 (5%) were assessed as having a high risk of bias [24,57]. The outcome measurement risks of 2 (5%) studies were rated as having a high risk of bias [53,57], and the remaining 34 (94%) studies were rated as having a low risk of bias. Regarding the selection of reported results, 36 (100%) studies had a low risk of bias. More detailed information on the risk-of-bias assessment is provided in Table S3 in Multimedia Appendix 2 [15,19-33,40-59].

Outcomes From Included Studies

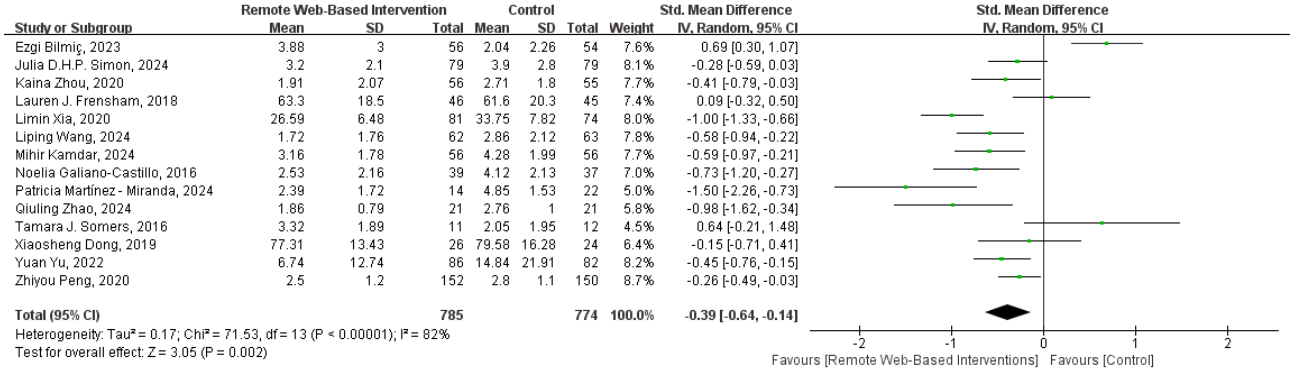
Overview

By including 36 eligible RCTs, we conducted a systematic review and meta-analysis to examine the efficacy of remote web-based interventions on pain (n=14, 39% studies [19-27,31,42,45,49,54]), fatigue (n=11, 31% studies [20,22,26,28,30,32,42,45,48,51,57]), anxiety (n=14, 39% studies [26-33,40,43,45,47,53,55]), depression (n=10, 28% studies [26-30,32,33,40,43,47]), and quality of life (n=26, 72% studies [15,19,20,22,23,25,27,28,30,32,33,40-42,44-48,50,52,54-56,58,59]) in patients with cancer. Physiologically, the meta-analysis results showed that remote web-based interventions could significantly reduce the pain intensity and relieve the fatigue state of patients with cancer. Regarding psychological state, remote web-based interventions had a substantial effect on the anxiety and depression states of patients with cancer. Overall, remote web-based interventions significantly improved the quality of life of patients with cancer. The specific results of the statistical analysis are described in subsequent sections.

Pain

As shown in Figure 2 [19-27,31,42,45,49,54], the results of the meta-analysis of 14 (39%) studies indicated that, compared with routine care, remote web-based interventions significantly reduced the pain intensity in patients with cancer (SMD –0.39, 95% CI –0.64 to –0.14; I^2=82%; GRADE rating=low).

Figure 2. Forest plot for the efficacy of a remote web-based intervention on pain.



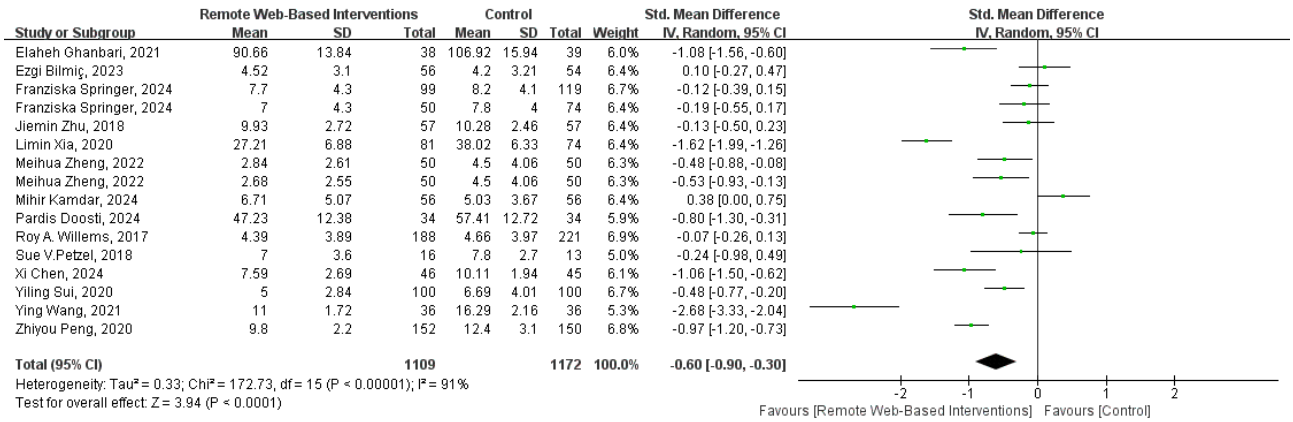
Fatigue

The results of the meta-analysis of 11 (31%) studies (Figure S2 in Multimedia Appendix 2 [20,22,26,28,30,32,42,45,48,51,57]) showed that remote web-based interventions had a significant effect on the fatigue state of patients with cancer (n=11, 31% studies; SMD –0.52, 95% CI –0.95 to –0.09; I^2=95%; GRADE rating=low).

Anxiety

As shown in Figure 3 [26-33,40,43,45,47,53,55], the results of the meta-analysis of 14 (39%) studies indicated that remote web-based interventions significantly improved the anxiety state of patients with cancer (SMD –0.60, 95% CI=–0.90 to –0.30; I^2=91%; GRADE rating=low).

Figure 3. Forest plot for the efficacy of a remote web-based intervention on anxiety.

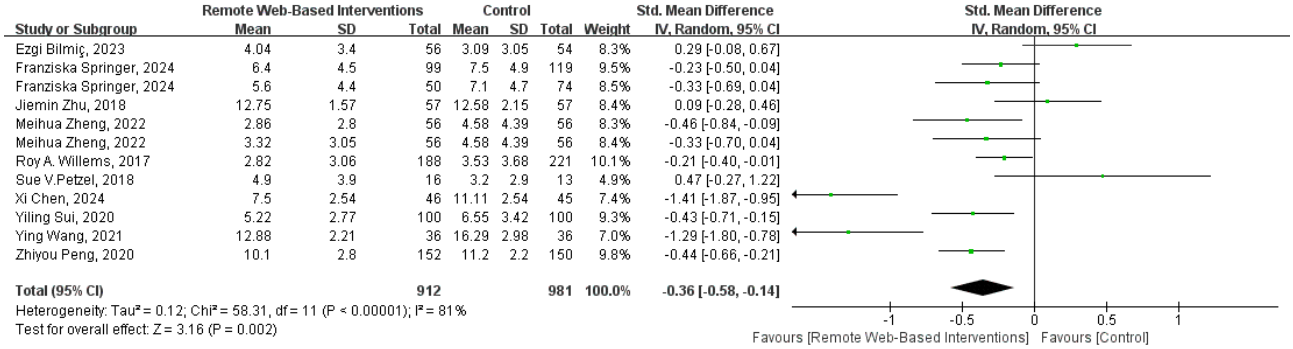


XSL•FO
RenderX

Depression

The results of the meta-analysis of 10 (28%) studies (Figure 4 [26-30,32,33,40,43,47]) showed that remote web-based interventions could effectively improve the depressive state of patients with cancer (SMD –0.36, 95% CI=–0.58 to –0.14; I^2=81%; GRADE rating=low).

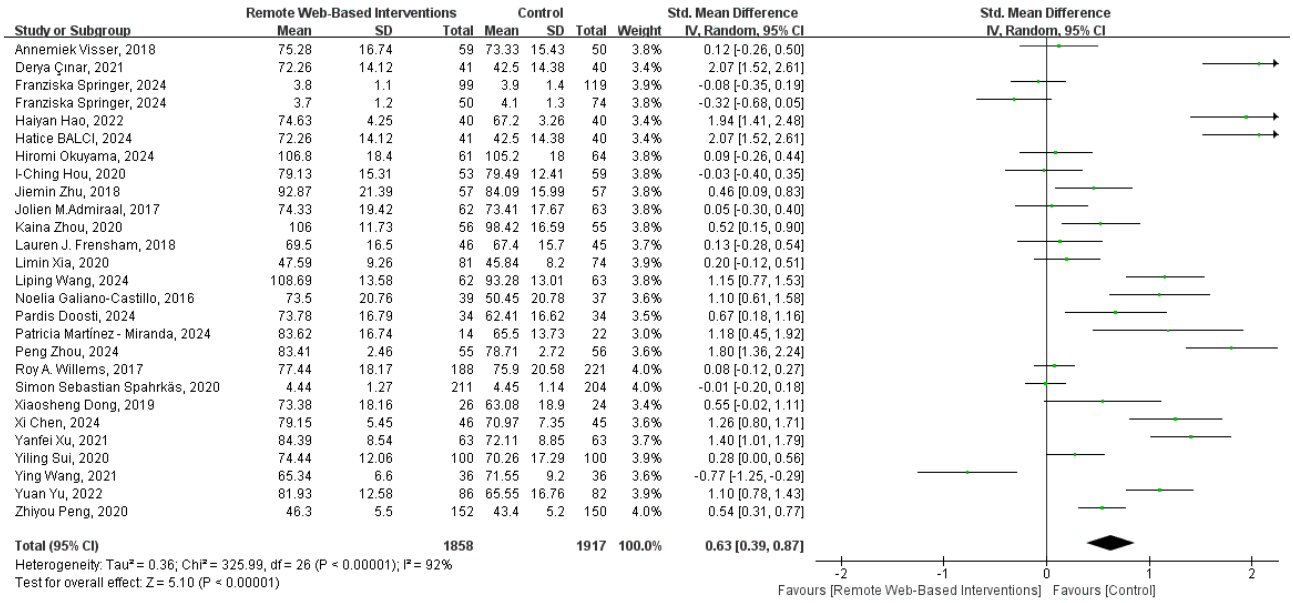
Figure 4. Forest plot for the efficacy of a remote web-based intervention on depression.



Quality of Life

As shown in Figure 5 [15,19,20,22,23,25,27,28,30,32,33,40-42,44-48,50,52,54-56,58,59], the results of the meta-analysis of 26 studies indicated that, compared with routine care, remote web-based interventions had a significant positive impact on the quality of life of patients with cancer (SMD 0.63, 95% CI 0.39-0.87; I^2=92%; GRADE rating=low).

Figure 5. Forest plot for the efficacy of a remote web-based intervention on quality of life.



Subgroup Analysis

Overview

As shown in Table 2, we conducted subgroup analyses on outcome indicators, such as pain, fatigue, anxiety, depression, and quality of life (region of the country, whether WeChat-based intervention was used, and duration of the intervention). Detailed subgroup analyses of fatigue, anxiety, depression, and quality of life are presented in Tables S4-S7 in Multimedia Appendix 2.

XSL•FO
RenderX

Table 2. Subgroup analysis—pain.

Subgroup categories and names	Pooled effect size, SMD[a] (95% CI)	Heterogeneity (I^2), %
Region		
Asia (n=7)	–0.53 (–0.76 to –0.30)	65
America (n=2)	–0.04 (–1.23 to 1.16)	85
Europe (n=3)	–0.75 (–1.36 to –0.15)	79
App		
WeChat (n=4)	–0.40 (–0.57 to –0.24)	46
Non-WeChat (n=10)	–0.33 (–0.69 to 0.03)	86
Intervention time		
<3 mo (n=5)	–0.35 (–0.86 to 0.16)	89
>3 mo (n=8)	–0.46 (–0.71 to –0.22)	68

[a]SMD: standardized mean difference.

Pain Subgroup by Country

In the Asian region, the meta-analysis results of 7 (19%) studies [20-23,27,42,45] showed that the pooled effect size SMD was –0.53 (95% CI –0.76 to –0.30), with a heterogeneity of 65%. In the American region, the results of 2 (6%) studies [24,31] showed that the pooled effect size SMD was –0.04 (95% CI –1.23 to 1.16), with a heterogeneity of 85%. In the European region, the results of 3 (8%) studies [19,49,54] showed that the pooled effect size SMD was –0.75 (95% CI –1.36 to –0.15), with a heterogeneity of 79%, suggesting that remote web-based interventions might have a pain-relieving effect in the European region.

Pain Subgroup by App

In the WeChat group, the meta-analysis of 4 (11%) studies [21,22,27,42] showed a pooled effect size SMD of –0.40 (95% CI –0.57 to –0.24), with a heterogeneity of 46%. This result indicated that the intervention using WeChat had a certain impact on pain, and the heterogeneity among studies was moderate. In the non-WeChat group, the meta-analysis of 10 (28%) studies [19,20,23-26,31,45,49,54] showed a pooled effect size SMD of –0.33 (95% CI –0.69 to 0.03), with a heterogeneity of 86%.

Pain Subgroup by Intervention Duration

In the subgroup with an intervention duration of less than 3 months, the meta-analysis of 5 (14%) studies [21,26,27,31,54] showed a pooled effect size SMD of –0.35 (95% CI –0.86 to 0.16), with a heterogeneity of 89%. In the subgroup with an intervention duration of more than 3 months, the results of 8 (22%) studies [19,20,22,23,25,42,45,49] showed a pooled effect size SMD of –0.46 (95% CI –0.71 to –0.22), with a heterogeneity of 68%, suggesting that a longer-term intervention had a certain pain-relieving effect.

Sensitivity Analysis

The sensitivity analysis for the meta-analysis results of pain (Table 3), anxiety (Table S8 in Multimedia Appendix 2 [26-33,40,43,45,47,53,55]), depression (Table S9 in Multimedia Appendix 2 [26-28,30,32,33,40,43,47]), and quality of life (Table S10 in Multimedia Appendix 2 [15,19,20,22,23,25,27,28,30,32,33,40-42,44-48,50,52,54-56,58,59]) showed that when any single study was excluded, the pooled estimates of the main analysis were robust. However, for the meta-analysis result of fatigue, the sensitivity analysis indicated that after excluding the studies by Xia [45] and Wang et al [28], respectively, the results of the overall pooled analysis changed from statistically significant (SMD –0.52, 95% CI –0.95 to –0.09; I^2=95%) to nonstatistically significant (SMD –0.28, 95% CI –0.57 to –0.01; I^2=91% and SMD –0.39, 95% CI –0.81 to –0.03; I^2=91%), suggesting that the main analysis was not robust in the pooled estimates (Table S11 in Multimedia Appendix 2 [20,22,26,28,30,32,42,45,48,51,57]).

XSL•FO
RenderX

Table 3. Sensitivity analysis (leave-one-out method)—pain.

Study removed	Pooled effect size after removing this study, SMD[a] (95% CI)	Heterogeneity (I^2), %
Bilmiç et al [26]	–0.48 (–0.68 to –0.27)	70
Simon et al [49]	–0.40 (–0.67 to –0.12)	83
Zhou et al [42]	–0.39 (–0.66 to –0.12)	83
Frensham et al [25]	–0.43 (–0.69 to –0.17)	82
Xia [45]	–0.33 (–0.58 to –0.09)	79
Wang et al [22]	–0.37 (–0.64 to –0.10)	83
Kamdar et al [31]	–0.37 (–0.64 to –0.10)	83
Galiano-Castillo et al [54]	–0.36 (–0.63 to –0.10)	83
Martínez-Miranda et al [19]	–0.33 (–0.58 to –0.09)	81
Zhao et al [21]	–0.35 (–0.61 to –0.10)	82
Somers et al [24]	–0.44 (–0.69 to –0.19)	82
Dong et al [23]	–0.41 (–0.67 to –0.14)	83
Yu et al [20]	–0.38 (–0.66 to –0.11)	83
Peng et al [27]	–0.40 (–0.69 to –0.12)	83

[a]SMD: standardized mean difference.

Publication Bias

As shown in Table 4, the results of Begg and Egger tests (P >.05) indicate no publication bias in the pooled analyses of pain, fatigue, anxiety, depression, and quality of life.

Table 4. Begg and Egger test values.

Clinical manifestations	Egger test		Begg test	
	Student t-statistic	P value	Standard Normal Z-statistic	P value
Pain	–0.29	.78	0.22	.83
Fatigue	–1.26	.23	0.75	.45
Anxiety	–1.48	.16	1.4	.16
Depression	–0.51	.62	0.34	.73
Quality of life	1.23	.25	0.187	.06

Discussion

Overview

In recent years, the health management of patients with cancer has shifted from single-disease treatment to a holistic focus on both physiological and psychological states. Patients with cancer not only have to cope with physiological symptoms (pain and fatigue) brought about by chemotherapy, radiotherapy, and other treatments, but they are also often burdened with psychosocial issues, including anxiety, depression, posttraumatic stress disorder, and fear of recurrence [60-63]. Although traditional face-to-face interventions can partially alleviate these problems, their accessibility and sustainability are limited due to the uneven distribution of medical resources, patients' limited mobility, or remote areas [64,65]. Remote web-based interventions, with their advantages of high accessibility, great flexibility, and low cost, have gradually become an important means for patients with cancer to manage their symptoms by themselves [14]. However, although previous studies have suggested that remote web-based interventions may improve physiological and psychological states and quality of life in patients with cancer, the effectiveness remains controversial [19-25,27-33]. To this end, this systematic review and meta-analysis aimed to summarize the relevant evidence regarding the physiological and psychological impacts of remote web-based interventions on patients with cancer. A total of 36 studies were included to determine the clinical efficacy of remote web-based interventions on physiological pain, fatigue, anxiety, depression, and quality of life in patients with cancer.

The Effect of Remote Web-Based Interventions on Pain

This study, through meta-analysis, found that remote web-based interventions significantly reduced the pain intensity in patients with cancer (n=14, 39% studies; SMD –0.39, 95% CI –0.64 to

XSL•FO
RenderX

–0.14; I^2=82%; GRADE rating=low). This result is consistent with the results of some previous studies. For example, in the review by Villinger et al [66], it was mentioned that some apps showed certain positive effects in cancer pain management. Through the real-time chat consultation function, patients can report pain and receive suggestions, which is helpful for pain management. The review by Wu et al [67] also pointed out that mHealth apps have certain potential in monitoring and managing cancer pain. After some patients used pain-related apps, their physiological pain intensity improved. Through remote monitoring and digital platforms, patients with cancer can report symptoms (pain) in real time, enabling health care providers to adjust treatment plans promptly [68]. Meanwhile, the health education module enhances the understanding of relevant symptoms in patients with cancer and improves their self-management [69]. In this study, the pooled effect value of pain showed high heterogeneity, with an overall heterogeneity of 82%. Further subgroup analyses revealed the impacts of different regions, app types, and intervention durations on heterogeneity: the heterogeneity in the Asian subgroup was 65%, 85% in the American subgroup, and 79% in the European subgroup. Regarding the type of intervention app, the heterogeneity in the WeChat subgroup was 46%, while it was 86% in the non-WeChat subgroup. In addition, the heterogeneity in the subgroup with an intervention duration of less than 3 months was 89%, and it decreased to 68% in the subgroup with an intervention duration of more than 3 months. The above results indicate that differences in intervention apps may be one of the main sources of heterogeneity. In contrast, factors such as region and intervention duration are related to heterogeneity but are not the main sources. The sensitivity analysis showed that when any single study was excluded, the main analysis was robust in the pooled estimates, indicating that the study's results regarding pain have certain reliability.

The Effect of Remote Web-Based Interventions on Fatigue

Studies show that, in addition to symptom management and educational empowerment [68,69], remote web-based interventions can also customize exercise programs (Tai Chi and aerobic training) for patients with cancer to improve their cardiopulmonary function and muscle strength and alleviate fatigue [70]. Our results indicated that remote web-based interventions significantly affected the fatigue state of patients with cancer (n=11, 31% studies, SMD –0.52, 95% CI –0.95 to –0.09; I^2=95%; GRADE rating=low). However, the sensitivity analysis showed that after excluding the studies by Xia [45] and Wang et al [28], respectively, the statistical significance of the pooled analysis results changed. This result suggests that the main analysis was not robust in the pooled estimates. These 2 studies might have been unique regarding intervention measures, sample characteristics, or fatigue assessment methods, significantly influencing the overall results. The results of the sensitivity analysis indicate that there is still a high degree of uncertainty in current research on the impact of remote web-based interventions on fatigue state of patients with cancer, and this result should be interpreted with caution.

The Effect of Remote Web-Based Interventions on Anxiety

This study found that remote web-based interventions significantly improved the anxiety state of patients with cancer. This result is consistent with some of the research results mentioned in the review by Villinger et al [66]. Some apps may help relieve patients' anxiety by providing information support and social interaction functions. Remote web-based intervention can continuously provide psychological support for patients with cancer [71] and comprehensive, personalized support through multidisciplinary collaboration [64]. In addition, the web-based community established by remote web-based interventions can reduce loneliness among patients with cancer. Sharing experiences among patients can enhance their sense of hope [72], thereby improving their psychological states. The pooled effect value of anxiety showed a very high degree of heterogeneity, with an overall heterogeneity as high as 91%. Subgroup analyses further revealed the potential sources of heterogeneity: the heterogeneities of the Asian subgroup, the American subgroup, and the European subgroup were 88%, 54%, and 0%, respectively, indicating significant differences among regions. Regarding the app type, the heterogeneity of the WeChat subgroup was 90%, and that of the non-WeChat subgroup was 74%. Regarding the intervention duration, the heterogeneity of the subgroup with a duration of less than 3 months was 87%, while that of the subgroup with a duration of more than 3 months was 95%. These findings suggest that the high heterogeneity may stem from regional differences and other unknown factors, and the intervention duration and the type of intervention app are not the main sources of high heterogeneity. The sensitivity analysis showed that when any single study was excluded, the main analysis was robust in the pooled estimates, indicating that the results of this study have a certain degree of reliability.

The Effect of Remote Web-Based Interventions on Depression

This study demonstrated that remote web-based interventions could significantly improve the depressive state of patients with cancer. In the review by Samadbeik et al [73], it was pointed out that mHealth interventions can help improve patients' psychological states. For example, providing psychological support, self-management education, and other content may positively impact depressive symptoms, which is consistent with the results of our study. The overall heterogeneity of the pooled effect value for depression was 81%. In the subgroup analysis, the heterogeneity of the European subgroup was 0%, and the heterogeneities of the other subgroups were all greater than 50%. This result indicates that the region may be one of the main sources of high heterogeneity. At the same time, the intervention duration and the type of intervention app are not the main sources of high heterogeneity. The results of the sensitivity analysis suggest that the results of this study have a certain degree of reliability.

The Effect of Remote Web-Based Interventions on Quality of Life

Several studies have shown that digital tools can establish a closed-loop intervention model based on the biopsychosocial

XSL•FO
RenderX

medical model through mechanisms such as psychosocial support, social connection and peer support, symptom management and educational empowerment, and multidisciplinary collaborative comprehensive interventions [64,68,69,71]. This intervention model can synergistically improve patients' physiological and psychological states, enhancing their quality of life. For example, some app interventions mentioned by Villinger et al [66] showed positive effects in aspects such as cancer-related symptom management, thereby improving patients' quality of life. Samadbeik et al [73] also pointed out that mHealth interventions could improve patients' quality of life by enhancing self-management and promoting healthy lifestyles. The results of our study also showed that remote web-based interventions had a significant positive impact on the quality of life of patients with cancer. The overall heterogeneity of the pooled effect value for quality of life and the heterogeneities of all subgroups exceeded 50%, which meant that region, the type of intervention app, and intervention duration were not the main sources of high heterogeneity. The results of Begg and Egger tests indicate no publication bias in our study. In addition, the sensitivity analysis results showed that our study was robust.

Comparison With the Published Systematic Reviews and Meta-Analyses

Previously, 4 systematic reviews [66,67,73,74] and 1 meta-analysis [75] have investigated the effects of apps, miniprograms, or other web-based platforms on symptom management in patients with cancer. We compared our study with the aforementioned articles. We found that Zou et al [74] conducted electronic database searches for studies published until December 2021 and included 20 studies (RCTs=14; quasi-experimental nonrandomized trials=4; and pre-post study=2). Villinger et al [66] carried out a comprehensive and systematic literature search in 5 electronic databases in January 2022 and finally included 17 studies (RCTs=5; nonrandomized trials=3; randomized trial study protocols=3; quasi-experimental design study protocols=1; and pre-post studies=5). Li et al [75] conducted electronic database searches for studies published until April 14, 2021, and included 28 studies (all were RCTs). Wu et al [67] conducted electronic database searches for studies published from 2013 to 2023 and included 20 studies (RCTs=4; quasi-experiments=3; prospective observational cohort studies=4; cross-sectional design=1; qualitative design=3; and mixed methods design=5). Samadbeik et al [73] conducted electronic database searches for studies published between January 2010 and February 2021 and included 28 studies (interventional studies=19, including RCTs; feasibility pilot studies=5; descriptive studies=1; qualitative studies=1; developmental studies=1; nonrandomized controlled studies=1; and mixed methods studies=1). In contrast, we conducted electronic database searches for studies published until October 15, 2024, and included 36 studies (RCTs=36). Compared with the studies by Zou et al [74], Villinger et al [66], Li et al [75], Wu et al [67], and Samadbeik et al [73], our study included newer, more voluminous studies, and all of them were RCTs. Regarding statistical methods, Zou et al [74], Villinger et al [66], Wu et al [67], and Samadbeik et al [73] reported results through comprehensive narrative descriptions. In contrast, we conducted meta-analyses on outcome indicators and reported the results based on them. Furthermore, our study conducted subgroup analysis, sensitivity analysis, and publication bias detection. Although the study by Li et al [75] also reported results using meta-analysis, compared with ours, the outcome indicators in their study were rather single and lacked comprehensiveness. Regarding the conclusions, the studies by Zou et al [74], Villinger et al [66], and Samadbeik et al [73] showed that it was impossible to determine the efficacy of web-based interventions based on apps and miniprograms on pain, fatigue, anxiety, depression, and quality of life in patients with cancer. The study by Li et al [75] indicated that web-based interventions could promote the quality of life of patients with cancer. Wu et al [67] demonstrated that mHealth apps were effective and acceptable in supporting the self-management of cancer pain and could provide patients with multimodal interventions to monitor, track, and manage pain. However, we conducted meta-analyses and quantitatively reported our conclusion: For patients with cancer, remote web-based interventions could alleviate pain, relieve fatigue, improve psychological states, such as anxiety and depression, and enhance their quality of life. In addition, we explored the potential mechanisms of remote web-based interventions to improve the physiological and psychological states of patients with cancer, which enriches the content of our research and theoretical basis. Finally, our conclusions are drawn based on subgroup analysis, sensitivity analysis, and publication bias detection. Therefore, our conclusions are more robust and persuasive.

Limitations and Strengths

Limitations

Several limitations need to be considered in this study. First, the heterogeneity of our conclusions was relatively high. Although subgroup analyses were conducted, the sources of heterogeneity were still not well explained. Second, the impact of confounding factors on our research results has not been further clarified. Third, the sensitivity analysis results indicated that the meta-analysis results regarding the impact of remote web-based interventions on fatigue state in patients with cancer were not robust, and we had to interpret this result with caution. Fourth, due to the differences in the follow-up time points of the included literature, our study only explored the short-term effects of remote web-based interventions on the physiological and mental impacts of patients with cancer. It did not investigate the medium- and long-term clinical efficacies. Finally, we only included studies published in English, which might lead to cultural and language biases, thus limiting the general applicability of the meta-analysis results.

Strengths

Despite these limitations, our meta-analysis has some significant strengths. First, this systematic review and meta-analysis only included RCTs, strengthening the evidence supporting the physiological and mental impacts of remote web-based interventions on patients with cancer. Second, the results of the sensitivity analyses showed that our main findings were robust and reliable, and the Begg and Egger tests did not reveal publication bias. Third, the literature we included was more

XSL•FO
RenderX

up-to-date, larger in quantity, and comprehensive. Fourth, compared with previous narrative reviews, we conducted meta-analyses. Finally, compared with meta-analysis studies with single outcome indicators, we explored more outcome indicators, and our methods and conclusions were more persuasive.

Implications for Clinical Practice

The results of this study provided valuable references for clinical practice, indicating that remote web-based interventions can serve as a beneficial supplement to the routine care of patients with cancer. Clinical medical staff should fully recognize the potential of remote web-based interventions in improving the symptoms and quality of life of patients with cancer. On the basis of patients' specific needs and conditions, they should reasonably select and recommend appropriate remote web-based intervention platforms or tools to provide patients with personalized physiological and mental support.

Conclusions

Remote web-based interventions have significant effects in reducing pain, relieving fatigue, improving anxiety, alleviating depression, and improving the quality of life of patients with cancer. However, we could not draw firm conclusions about these findings due to the evidence's low overall certainty. In the future, more high-quality, large-sample RCTs are needed to validate the effectiveness of remote web-based interventions further.

Acknowledgments

This work was supported by the Sichuan Local Science and Technology Development Project guided by a central government grant (2024ZYD0163).

Data Availability

All data generated or analyzed during this study are included in this paper and Multimedia Appendix 2 or available from the corresponding author upon reasonable request.

Authors' Contributions

LT and JF contributed to the review protocol. LT, YW, JL, JG, TL, and JF designed the search strategy and selected studies. LT, YW, and JF extracted the data. LT, YW, JL, JG, and TL analyzed the data. LT drafted the manuscript. All authors contributed to the drafting of the review, revised the manuscript critically for important content, approved the final version of the paper, and had access to all the data in the study and could take responsibility for the integrity of the data and the accuracy of the data analysis. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA Checklist.
[PDF File (Adobe PDF File), 82 KB - mhealth_v13i1e71196_app1.pdf]

Multimedia Appendix 2
Supplementary materials.
[DOCX File , 258 KB - mhealth_v13i1e71196_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies (randomized controlled trials) (continued from Table 1).
[DOCX File , 20 KB - mhealth_v13i1e71196_app3.docx]

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015 Mar 01;136(5):E359-E386. [doi: 10.1002/ijc.29210] [Medline: 25220842]
2. Alshawwa S, Assiri R. mHealth interventions for cancer care and support a systematic literature review. Syst Rev Pharm 2020;11(9):725-744. [doi: 10.31838/srp.2020.9.104]
3. Clark MM, Rummans TA, Atherton PJ, Cheville AL, Johnson ME, Frost MH, et al. Randomized controlled trial of maintaining quality of life during radiotherapy for advanced cancer. Cancer 2013 Feb 15;119(4):880-887 [FREE Full text] [doi: 10.1002/cncr.27776] [Medline: 22930253]

XSL•FO
RenderX

4. Solberg Nes L, Liu H, Patten CA, Rausch SM, Sloan JA, Garces YI, et al. Physical activity level and quality of life in long term lung cancer survivors. Lung Cancer 2012 Sep;77(3):611-616 [FREE Full text] [doi: 10.1016/j.lungcan.2012.05.096] [Medline: 22681871]
5. Stanton AL. What happens now? Psychosocial care for cancer survivors after medical treatment completion. J Clin Oncol 2012 Apr 10;30(11):1215-1220. [doi: 10.1200/JCO.2011.39.7406] [Medline: 22412133]
6. Mehdizadeh H, Asadi F, Mehrvar A, Nazemi E, Emami H. Smartphone apps to help children and adolescents with cancer and their families: a scoping review. Acta Oncol 2019 Jul;58(7):1003-1014. [doi: 10.1080/0284186X.2019.1588474] [Medline: 30915872]
7. Akechi T, Okuyama T, Endo C, Sagawa R, Uchida M, Nakaguchi T, et al. Patient's perceived need and psychological distress and/or quality of life in ambulatory breast cancer patients in Japan. Psychooncology 2011 May;20(5):497-505. [doi: 10.1002/pon.1757] [Medline: 20878850]
8. Servaes P, Verhagen C, Bleijenberg G. Fatigue in cancer patients during and after treatment. Eur J Cancer 2002 Jan;38(1):27-43. [doi: 10.1016/S0959-8049(01)00332-X]
9. Stepanski EJ, Walker MS, Schwartzberg LS, Blakely LJ, Ong JC, Houts AC. The relation of trouble sleeping, depressed mood, pain, and fatigue in patients with cancer. J Clin Sleep Med 2009 Apr 15;5(2):132-136 [FREE Full text] [Medline: 19968046]
10. Kuhnt S, Ernst J, Singer S, Rüffer JU, Kortmann RD, Stolzenburg JU, et al. Fatigue in cancer survivors--prevalence and correlates. Onkologie 2009 Jun;32(6):312-317. [doi: 10.1159/000215943] [Medline: 19521117]
11. Carpenter JS, Elam JL, Ridner SH, Carney PH, Cherry GJ, Cucullu HL. Sleep, fatigue, and depressive symptoms in breast cancer survivors and matched healthy women experiencing hot flashes. Oncol Nurs Forum 2004 May;31(3):591-598. [doi: 10.1188/04.onf.591-598] [Medline: 15146224]
12. Donovan KA, Jacobsen PB. Fatigue, depression, and insomnia: evidence for a symptom cluster in cancer. Semin Oncol Nurs 2007 May;23(2):127-135. [doi: 10.1016/j.soncn.2007.01.004] [Medline: 17512440]
13. McPherson CJ, Higginson IJ, Hearn J. Effective methods of giving information in cancer: a systematic literature review of randomized controlled trials. J Public Health 2001 Sep 01;23(3):227-234. [doi: 10.1093/pubmed/23.3.227]
14. Birkhoff SD, Smeltzer SC. Perceptions of smartphone user-centered mobile health tracking apps across various chronic illness populations: an integrative review. J Nurs Scholarsh 2017 Jul;49(4):371-378. [doi: 10.1111/jnu.12298] [Medline: 28605151]
15. Hou IC, Lin HY, Shen SH, Chang KJ, Tai HC, Tsai AJ, et al. Quality of life of women after a first diagnosis of breast cancer using a self-management support mHealth app in Taiwan: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 04;8(3):e17084 [FREE Full text] [doi: 10.2196/17084] [Medline: 32130181]
16. Singleton AC, Raeside R, Hyun KK, Partridge SR, Di Tanna GL, Hafiz N, et al. Electronic health interventions for patients with breast cancer: systematic review and meta-analyses. J Clin Oncol 2022 Jul 10;40(20):2257-2270 [FREE Full text] [doi: 10.1200/JCO.21.01171] [Medline: 35500200]
17. Heapy AA, Higgins DM, Cervone D, Wandner L, Fenton BT, Kerns RD. A systematic review of technology-assisted self-management interventions for chronic pain: looking across treatment modalities. Clin J Pain 2015 Jun;31(6):470-492. [doi: 10.1097/AJP.0000000000000185] [Medline: 25411862]
18. Vismara LA, McCormick C, Young GS, Nadhan A, Monlux K. Preliminary findings of a telehealth approach to parent training in autism. J Autism Dev Disord 2013 Dec;43(12):2953-2969. [doi: 10.1007/s10803-013-1841-8] [Medline: 23677382]
19. Martínez-Miranda P, Jiménez-Rejano JJ, Muñoz-Fernández MJ, García-Muñoz C, Casuso-Holgado MJ. Effectiveness of an interactive online group intervention based on pain neuroscience education and graded exposure to movement in breast cancer survivors with chronic pain: a randomised controlled trial. Support Care Cancer 2024 Oct 07;32(10):705. [doi: 10.1007/s00520-024-08887-4] [Medline: 39373766]
20. Yu Y, Li M, Kang R, Liu X, Wang N, Zhu Q, et al. The effectiveness of telephone and internet-based supportive care for patients with esophageal cancer on enhanced recovery after surgery in China: a randomized controlled trial. Asia Pac J Oncol Nurs 2022 Apr;9(4):217-228 [FREE Full text] [doi: 10.1016/j.apjon.2022.02.007] [Medline: 35571631]
21. Zhao Q, Qiu X, Liu W, Nian Z, Chen T, Chen J, et al. Application of a WeChat Mini Program to provide pharmaceutical care for cancer pain patients: a randomized controlled trial. Digit Health 2024;10:20552076241255654 [FREE Full text] [doi: 10.1177/20552076241255654] [Medline: 38766359]
22. Wang LP, Yao LH, Wang D, Lim GH, Almaghrabi MY, Shen H, et al. A WeChat-based nursing intervention program improves the postoperative rehabilitation of breast cancer patients: results from a randomized controlled trial. Gland Surg 2024 Mar 27;13(3):383-394 [FREE Full text] [doi: 10.21037/gs-24-40] [Medline: 38601277]
23. Dong X, Yi X, Gao D, Gao Z, Huang S, Chao M, et al. The effects of the combined exercise intervention based on internet and social media software (CEIBISMS) on quality of life, muscle strength and cardiorespiratory capacity in Chinese postoperative breast cancer patients: a randomized controlled trial. Health Qual Life Outcomes 2019 Jun 26;17(1):109 [FREE Full text] [doi: 10.1186/s12955-019-1183-0] [Medline: 31242926]

XSL•FO
RenderX

24. Somers TJ, Kelleher SA, Westbrook KW, Kimmick GG, Shelby RA, Abernethy AP, et al. A small randomized controlled pilot trial comparing mobile and traditional pain coping skills training protocols for cancer patients with pain. Pain Res Treat 2016;2016:2473629 [FREE Full text] [doi: 10.1155/2016/2473629] [Medline: 27891252]
25. Frensham LJ, Parfitt G, Dollman J. Effect of a 12-week online walking intervention on health and quality of life in cancer survivors: a quasi-randomized controlled trial. Int J Environ Res Public Health 2018 Sep 21;15(10):2081 [FREE Full text] [doi: 10.3390/ijerph15102081] [Medline: 30248943]
26. Bilmiç E, Selçukbiricik F, Bagcivan G. The effectiveness of online pain management education on the patient related barriers to cancer pain management: a randomized controlled trial. Eur J Oncol Nurs 2023 Dec;67:102422. [doi: 10.1016/j.ejon.2023.102422] [Medline: 37812994]
27. Peng Z, Li L, Chen Y, Feng Z, Fang X. WeChat app-based reinforced education improves the quality of opioid titration treatment of cancer-related pain in outpatients: a randomized control study. BMC Cancer 2020 Sep 04;20(1):852 [FREE Full text] [doi: 10.1186/s12885-020-07270-w] [Medline: 32887560]
28. Wang Y, Yang L, Xu W, Wang C, Chen J, Wu Y, et al. Effects of a WeChat-based multimodal psychoeducational intervention on psychological well-being and quality of life in acute leukaemia patients in China: a randomised controlled trial. J Cancer Surviv 2022 Dec;16(6):1461-1477. [doi: 10.1007/s11764-021-01124-5] [Medline: 34705196]
29. Petzel SV, Isaksson Vogel R, Cragg J, McClellan M, Chan D, Jacko JA, et al. Effects of web-based instruction and patient preferences on patient-reported outcomes and learning for women with advanced ovarian cancer: a randomized controlled trial. J Psychosoc Oncol 2018;36(4):503-519 [FREE Full text] [doi: 10.1080/07347332.2018.1457125] [Medline: 29791275]
30. Willems RA, Bolman CA, Mesters I, Kanera IM, Beaulen AA, Lechner L. Short-term effectiveness of a web-based tailored intervention for cancer survivors on quality of life, anxiety, depression, and fatigue: randomized controlled trial. Psychooncology 2017 Feb;26(2):222-230. [doi: 10.1002/pon.4113] [Medline: 26988800]
31. Kamdar M, Jethwani K, Centi AJ, Agboola S, Fischer N, Traeger L, et al. A digital therapeutic application (ePAL) to manage pain in patients with advanced cancer: a randomized controlled trial. J Pain Symptom Manage 2024 Sep;68(3):261-271. [doi: 10.1016/j.jpainsymman.2024.05.033] [Medline: 38866116]
32. Springer F, Maier A, Friedrich M, Raue JS, Finke G, Lordick F, et al. Digital therapeutic (Mika) targeting distress in patients with cancer: results from a nationwide waitlist randomized controlled trial. J Med Internet Res 2024 Apr 25;26:e51949 [FREE Full text] [doi: 10.2196/51949] [Medline: 38663007]
33. Chen X, Qin Y, Chaimongkol N. Effectiveness of a phone-based support program on self-care self-efficacy, psychological distress, and quality of life among women newly diagnosed with breast cancer: a randomized controlled trial. Eur J Oncol Nurs 2024 Aug;71:102643. [doi: 10.1016/j.ejon.2024.102643] [Medline: 38889503]
34. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
35. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
36. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
37. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994 Dec;50(4):1088-1101. [Medline: 7786990]
38. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
39. Patsopoulos NA, Evangelou E, Ioannidis JP. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. Int J Epidemiol 2008 Oct;37(5):1148-1157 [FREE Full text] [doi: 10.1093/ije/dyn065] [Medline: 18424475]
40. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]
41. Zhou P, Li H, Pang X, Wang T, Wang Y, He H, et al. Effect of a mobile health-based remote interaction management intervention on the quality of life and self-management behavior of patients with low anterior resection syndrome: randomized controlled trial. J Med Internet Res 2024 Aug 13;26:e53909 [FREE Full text] [doi: 10.2196/53909] [Medline: 39137413]
42. Zhou K, Wang W, Zhao W, Li L, Zhang M, Guo P, et al. Benefits of a WeChat-based multimodal nursing program on early rehabilitation in postoperative women with breast cancer: a clinical randomized controlled trial. Int J Nurs Stud 2020 Jun;106:103565. [doi: 10.1016/j.ijnurstu.2020.103565] [Medline: 32272281]
43. Zheng M, Zhang X, Xiao H. Effects of a WeChat-based life review program for patients with digestive system cancer: 3-arm parallel randomized controlled trial. J Med Internet Res 2022 Aug 25;24(8):e36000 [FREE Full text] [doi: 10.2196/36000] [Medline: 36006665]

XSL•FO
RenderX

44. Xu YF, Fei Xu X, Song K, Qiu C, Zhang X, Er-re He Mam DL, et al. Effects of extended care based on the WeChat platform on self-efficacy and quality of life of postoperative breast cancer patients. Indian J Pharm Sci 2021;83(1):23-30 [FREE Full text] [doi: 10.36468/pharmaceutical-sciences.spl.165]
45. Xia L. The effects of continuous care model of information-based hospital-family integration on colostomy patients: a randomized controlled trial. J Cancer Educ 2020 Apr 26;35(2):301-311. [doi: 10.1007/s13187-018-1465-y] [Medline: 30685831]
46. Visser A, Prins JB, Jansen L, Radema SA, Schlooz MS, van Dalen T, et al. Group medical consultations (GMCs) and tablet-based online support group sessions in the follow-up of breast cancer: a multicenter randomized controlled trial. Breast 2018 Aug;40:181-188. [doi: 10.1016/j.breast.2018.05.012] [Medline: 29906741]
47. Sui Y, Wang T, Wang X. The impact of WeChat app-based education and rehabilitation program on anxiety, depression, quality of life, loss of follow-up and survival in non-small cell lung cancer patients who underwent surgical resection. Eur J Oncol Nurs 2020 Apr;45:101707. [doi: 10.1016/j.ejon.2019.101707] [Medline: 32078926]
48. Spahrkäs SS, Looijmans A, Sanderman R, Hagedoorn M. Beating cancer-related fatigue with the Untire mobile app: results from a waiting-list randomized controlled trial. Psychooncology 2020 Nov;29(11):1823-1834 [FREE Full text] [doi: 10.1002/pon.5492] [Medline: 33393199]
49. Simon JD, Schepers SA, van Gorp M, Michiels EM, Fiocco M, Grootenhuis MA, et al. Pain monitoring app leads to less pain in children with cancer at home: results of a randomized controlled trial. Cancer 2024 Jul 01;130(13):2339-2350. [doi: 10.1002/cncr.35100] [Medline: 37947136]
50. Okuyama H, Takada F, Taira N, Nakamura S. A randomized trial of the impact of symptom monitoring using an electronic patient-reported outcome app on health-related quality of life in postmenopausal breast cancer patients receiving adjuvant endocrine therapy. Breast Cancer 2024 Sep;31(5):787-797. [doi: 10.1007/s12282-024-01592-4] [Medline: 38796818]
51. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444 [FREE Full text] [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
52. Hao H, Yang X, Zhu H, Wang Z, Zhang H, Huang C. Effect of the whole seamless connection of nursing from WeChat interactive platform on stigma and quality of life in patients with urinary system cancer. Digit Health 2022;8:20552076221102772 [FREE Full text] [doi: 10.1177/20552076221102772] [Medline: 35651732]
53. Ghanbari E, Yektatalab S, Mehrabi M. Effects of psychoeducational interventions using mobile apps and mobile-based online group discussions on anxiety and self-esteem in women with breast cancer: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 18;9(5):e19262 [FREE Full text] [doi: 10.2196/19262] [Medline: 34003138]
54. Galiano-Castillo N, Cantarero-Villanueva I, Fernández-Lao C, Ariza-García A, Díaz-Rodríguez L, Del-Moral-Ávila R, et al. Telehealth system: a randomized controlled trial evaluating the impact of an internet-based exercise intervention on quality of life, pain, muscle strength, and fatigue in breast cancer survivors. Cancer 2016 Oct 15;122(20):3166-3174 [FREE Full text] [doi: 10.1002/cncr.30172] [Medline: 27332968]
55. Doosti P, Etemadifar S, Aliakbari F. The impact of a continuous care model utilizing a smartphone application on quality of life and anxiety levels among gynecologic cancer patients: a randomized controlled trial. BMC Nurs 2024 Oct 01;23(1):706 [FREE Full text] [doi: 10.1186/s12912-024-02391-0] [Medline: 39354541]
56. Çınar D, Karadakovan A, Erdoğan AP. Effect of mobile phone app-based training on the quality of life for women with breast cancer. Eur J Oncol Nurs 2021 Jun;52:101960. [doi: 10.1016/j.ejon.2021.101960] [Medline: 33882446]
57. Bandani-Susan B, Montazeri A, Haghighizadeh MH, Araban M. The effect of mobile health educational intervention on body image and fatigue in breast cancer survivors: a randomized controlled trial. Ir J Med Sci 2022 Aug 09;191(4):1599-1605. [doi: 10.1007/s11845-021-02738-5] [Medline: 34370166]
58. Balci H, Faydali S. The effect of education performed using mobile application on supportive care needs and quality of life in women with breast cancer: randomized controlled trial. Semin Oncol Nurs 2024 Aug;40(4):151684. [doi: 10.1016/j.soncn.2024.151684] [Medline: 38969573]
59. Admiraal JM, van der Velden AW, Geerling JI, Burgerhof JG, Bouma G, Walenkamp AM, et al. Web-based tailored psychoeducation for breast cancer patients at the onset of the survivorship phase: a multicenter randomized controlled trial. J Pain Symptom Manage 2017 Oct;54(4):466-475 [FREE Full text] [doi: 10.1016/j.jpainsymman.2017.07.009] [Medline: 28711750]
60. Kaphle M, Bajracharya D, Regmi N, Aryal D, Karki R. Depression and anxiety among cancer patients visiting a tertiary care cancer hospital. World J Psychiatry 2024 Feb 19;14(2):287-295 [FREE Full text] [doi: 10.5498/wjp.v14.i2.287] [Medline: 38464775]
61. Levesque JV, Farnsworth C, Luckey R, Hart R, Hegarty S. Fear, worry and sadness: an exploratory study of psychological wellbeing in men caring for their partner with ovarian cancer. Support Care Cancer 2022 Jan;30(1):825-833. [doi: 10.1007/s00520-021-06488-z] [Medline: 34389907]
62. Song L, Su Z, He Y, Pang Y, Zhou Y, Wang Y, et al. Association between anxiety, depression, and symptom burden in patients with advanced colorectal cancer: a multicenter cross-sectional study. Cancer Med 2024 Jun;13(11):e7330 [FREE Full text] [doi: 10.1002/cam4.7330] [Medline: 38845478]

XSL•FO
RenderX

63. Carlson LE. Psychosocial and integrative oncology: interventions across the disease trajectory. Annu Rev Psychol 2023 Jan 18;74(1):457-487 [FREE Full text] [doi: 10.1146/annurev-psych-032620-031757] [Medline: 36104001]
64. Bergerot CD, Bergerot PG, Philip EJ, Ferrari R, Peixoto RM, Crane TE, et al. Enhancing cancer supportive care: integrating psychosocial support, nutrition, and physical activity using telehealth solutions. JCO Glob Oncol 2024 Oct;10:e2400333. [doi: 10.1200/GO-24-00333] [Medline: 39509661]
65. DeGuzman PB, Vogel DL, Bernacchi V, Scudder MA, Jameson MJ. Self-reliance, social norms, and self-stigma as barriers to psychosocial help-seeking among rural cancer survivors with cancer-related distress: qualitative interview study. JMIR Form Res 2022 May 19;6(5):e33262 [FREE Full text] [doi: 10.2196/33262] [Medline: 35588367]
66. Villinger K, Berli C, Scholz U. App-based interventions to improve cancer outcomes rely on informational support from professionals: a systematic review. Health Psychol Rev 2024 Dec;18(4):767-789 [FREE Full text] [doi: 10.1080/17437199.2024.2349617] [Medline: 38755755]
67. Wu W, Graziano T, Salner A, Chen MH, Judge MP, Cong X, et al. Acceptability, effectiveness, and roles of mHealth applications in supporting cancer pain self-management: integrative review. JMIR Mhealth Uhealth 2024 Jul 18;12:e53652 [FREE Full text] [doi: 10.2196/53652] [Medline: 39024567]
68. Caputo MP, Rodriguez CS, Padhya TA, Mifsud MJ. Telehealth interventions in head and neck cancer patients: a systematic review. Cancer Nurs 2023;46(5):E320-E327. [doi: 10.1097/NCC.0000000000001130] [Medline: 37607382]
69. Chen M, Gong J, Li Q. The application of eHealth in cancer survivorship care: a review of web-based dyadic interventions for post-treatment cancer survivors and caregivers. Asia Pac J Oncol Nurs 2022 Oct;9(10):100109 [FREE Full text] [doi: 10.1016/j.apjon.2022.100109] [Medline: 36118622]
70. Peng Y, Zhang K, Wang L, Peng Y, Liu Z, Liu L, et al. Effect of a telehealth-based exercise intervention on the physical activity of patients with breast cancer: a systematic review and meta-analysis. Asia Pac J Oncol Nurs 2022 Dec;9(12):100117 [FREE Full text] [doi: 10.1016/j.apjon.2022.100117] [Medline: 36276881]
71. Ma Z, Shi Y, Yao S, Lu N, Cheng F. Effectiveness of telemedicine-based psychosocial intervention for breast cancer patients: a systematic review and meta-analysis. Support Care Cancer 2023 Sep 28;31(10):595. [doi: 10.1007/s00520-023-08052-3] [Medline: 37768402]
72. Zhang Y, Li J, Hu X. The effectiveness of dignity therapy on hope, quality of life, anxiety, and depression in cancer patients: a meta-analysis of randomized controlled trials. Int J Nurs Stud 2022 Aug;132:104273. [doi: 10.1016/j.ijnurstu.2022.104273] [Medline: 35635908]
73. Samadbeik M, Garavand A, Aslani N, Sajedimehr N, Fatehi F. Mobile health interventions for cancer patient education: a scoping review. Int J Med Inform 2023 Nov;179:105214. [doi: 10.1016/j.ijmedinf.2023.105214] [Medline: 37729837]
74. Zou P, Huang A, Luo Y, Tchakerian N, Zhang H, Zhang C. Effects of using WeChat/WhatsApp on physical and psychosocial health outcomes among oncology patients: a systematic review. Health Informatics J 2023 Mar 22;29(1):14604582231164697 [FREE Full text] [doi: 10.1177/14604582231164697] [Medline: 36949645]
75. Li J, Liu Y, Jiang J, Peng X, Hu X. Effect of telehealth interventions on quality of life in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2021 Oct;122:103970. [doi: 10.1016/j.ijnurstu.2021.103970] [Medline: 34303269]

Abbreviations

GRADE: Grading of Recommendations Assessment, Development, and Evaluation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by L Buis; submitted 12.01.25; peer-reviewed by W Bloch; comments to author 14.04.25; revised version received 18.04.25; accepted 29.04.25; published 12.06.25.

Please cite as:
Tian L, Wen Y, Li J, Guan J, Li T, Fan J
Effects of Remote Web-Based Interventions on the Physiological and Psychological States of Patients With Cancer: Systematic Review With Meta-Analysis
JMIR Mhealth Uhealth 2025;13:e71196
URL: https://mhealth.jmir.org/2025/1/e71196
doi:10.2196/71196
PMID:

XSL•FO
RenderX

©Lv Tian, Yixuan Wen, Jingmiao Li, Jiexin Guan, Tao Li, Jun Fan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Recent Advancements in Wearable Hydration-Monitoring Technologies: Scoping Review of Sensors, Trends, and Future Directions

Nazim A. Belabbaci[1*], BS, MEng, MS; Raphael Anaadumba[1*], BS, MS; Mohammad Arif Ul Alam[1,2,3], BSc, PhD

[1]Miner School of Computer & Information Sciences, University of Massachusetts Lowell, Lowell, MA, United States

[2]National Institute on Aging, NIH, Bethesda, MD, United States

[3]UMass Chan Medical School, Worcester County, MA, United States

[*]these authors contributed equally

Corresponding Author:
Nazim A. Belabbaci, BS, MEng, MS
Miner School of Computer & Information Sciences
University of Massachusetts Lowell
220 Pawtucket St
Lowell, MA, 01854
United States
Phone: 1 978 934 4000
Email: bellabaci.nazim@gmail.com

Abstract

Background: Monitoring hydration is crucial for maintaining health and preventing dehydration. Despite the potential of wearable devices for continuous hydration monitoring, health research hasn't fully explored this application, and clear design guidelines are absent. This scoping review aimed to address this gap by analyzing current research trends and assessing the potential impact of wearable technologies for hydration monitoring.

Objective: This review comprehensively examined recent advancements in wearable hydration-monitoring technologies, focusing on their capabilities, limitations, and research and prototype designs. It explored various sensors and technologies used to track hydration, compared their advantages and disadvantages, identified trends in wearable hydration-monitoring devices, evaluated their accuracy and reliability against established benchmarks, and reviewed commercially available products to bridge research findings and practical applications.

Methods: Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, we systematically searched PubMed, IEEE Xplore, and Google Scholar for studies (2014-2024) on noninvasive wearable devices that used physiological biomarkers. Validation with human participants or comparisons with gold standards was required. Data extraction covered study characteristics, sensor technologies, validation methods, and commercial product analysis. In addition to academic research papers, gray literature was included through a Google Scholar search to investigate commercial products in the field of hydration monitoring. This approach ensured a broader perspective on technological advancements and market trends.

Results: The review synthesized 63 articles selected from 156 included for full-text analysis. The literature was categorized based on sensor types, including electrical, optical, thermal, microwave, and multimodal sensors. Most studies (47/63, 75%) examined the effects of hydration on physiological parameters, with some (16/63, 25%) focusing on hydration status during physical activity or in specific environmental conditions. Commercially available products from 8 companies were also evaluated for their technological features, functionalities, and applications. The dominance of electrical sensors in research was highlighted due to their ease of use and integration into wearable devices. While fewer in number, optical sensors demonstrated higher precision and provided molecular-level insights. The emergence of multimodal sensors suggests a trend toward combining technologies to improve accuracy, as reflected by their increasing publication share. Other sensors, such as thermal and microwave-based sensors, occupied specialized niches. The growing acceptance of optical-based wearables in the market reflects their cost-to-precision effectiveness.

Conclusions: Wearable hydration-monitoring devices provide real-time assessments of hydration status, but challenges remain in ensuring their reliability, accuracy, and applicability across diverse populations and conditions. Future directions for research

XSL•FO
RenderX

include standardized protocols, extensive clinical trials, sensor miniaturization, and enhanced wearability. Multimodal systems that integrate various sensors with artificial intelligence–driven analysis hold promise for personalized hydration management. This review offers detailed insights into the strengths and challenges of sensor technologies, paving the way for practical skin hydration-monitoring solutions.

(JMIR Mhealth Uhealth 2025;13:e60569) doi:10.2196/60569

KEYWORDS

wearable technology; hydration monitoring; sensors; trends; artificial intelligence

Introduction

Background

Water is a fundamental component of the human body, accounting for 75% of body weight in infants and 55% in older adults [1]. It plays a crucial role in the functioning of various organs. Maintaining proper hydration is essential for human health as it ensures that the body's physiological processes function optimally. When dehydration occurs, a complex array of physiological mechanisms is triggered to preserve homeostasis. Dehydration reduces blood volume and increases plasma osmolality, which stimulates the release of hormones such as antidiuretic hormone and aldosterone. These hormones work to retain water and sodium, aiming to restore fluid balance. However, these compensatory mechanisms, while crucial, can only temporarily sustain function. Prolonged or severe dehydration can lead to impaired cellular function, disrupted thermoregulation, and compromised cardiovascular stability [2].

Conversely, overhydration disrupts the body's electrolyte balance, most notably causing hyponatremia, a condition characterized by dangerously low sodium levels in the blood. This imbalance can lead to neurological symptoms, including confusion; seizures; and, in extreme cases, brain swelling. Specific medical conditions such as hydrocephalus can further complicate the effects of overhydration, necessitating careful management of fluid intake to prevent exacerbation of intracranial pressure [3]. The negative impacts of dehydration are especially pronounced in populations at elevated risk, such as athletes who lose significant amounts of fluid during intense physical activity and older individuals whose ability to sense thirst may diminish with age. On the other hand, overhydration, though less common, also presents significant health risks, particularly for those with underlying conditions such as heart failure or kidney disease. These populations are vulnerable to both dehydration and overhydration, making accurate and timely monitoring essential [2,4]. Understanding these biophysical and biochemical responses to hydration status is key to designing effective monitoring systems. The sensors discussed in this review are designed to detect changes in hydration by leveraging these underlying physiological mechanisms, offering real-time insights into an individual's hydration status.

Given the complexities and the urgent need for noninvasive, real-time hydration-monitoring systems, as highlighted by numerous studies [5], recent advancements in and the widespread adoption of wearable technology present a promising solution. Over the past few years, not only have wearables become integral to daily life for many [6-9], but their potential for scientific data collection and evaluation has also significantly increased. This is especially evident in health and fitness monitoring, where wearables are being used for various metrics [9-11].

In the realm of wearable technology, a diverse range of devices, from sophisticated smartwatches [12,13] to innovative smart clothing [14], is revolutionizing how we monitor health parameters. One particularly notable advancement is their ability to monitor hydration levels, an often overlooked but essential aspect of health. These wearables use cutting-edge methods such as analyzing electrolyte variations in sweat or using electrical and optical sensors for cellular-level hydration assessments. This technique exploits the optical properties of tissues to provide a more nuanced understanding of hydration status.

The implications of these technological strides are profound. For the first time, continuous, noninvasive hydration monitoring is within reach, offering a potential safeguard for at-risk populations. This marks a new era in preventive health care, where the risks associated with hydration can be proactively mitigated, thereby enhancing the quality of life. However, this promising future is not without its challenges. Recent research in hydration monitoring, while encouraging, reveals significant gaps. A major shortfall is the lack of robust clinical studies, particularly those focused on specific diseases or clinical conditions [15-17]. This limitation raises questions about the applicability and accuracy of these technologies in real-world scenarios as the absence of targeted research hinders our understanding of their effectiveness in addressing particular medical challenges. Furthermore, most wearable technologies are still in their developmental phase [17-19], creating uncertainty about their reliability and effectiveness in practical applications. In addition, the sample sizes in existing experiments often lack statistical significance, primarily due to challenges in recruiting participants and the logistical complexities of monitoring hydration in real-world settings. These limitations diminish the reliability of the findings, making it difficult to draw definitive conclusions. Moreover, much of the research is narrowly focused on specific population groups [20], which may not fully represent the diverse needs of the broader population. This lack of inclusivity in research design could skew the results and limit the universal applicability of wearable technologies.

Objectives

In this review, we aimed to fill a notable gap in the current literature by providing an in-depth exploration of the latest developments in wearable and portable hydration-tracking

XSL•FO
RenderX

solutions. We focused on populations at elevated risk of dehydration, such as athletes, military personnel operating in extreme environments, individuals involved in infant and maternal health, and older adults—and detailed their particular susceptibility [21] to dehydration given that body weight losses exceeding 2% can lead to heart-related injury [22,23]. We will delve into the medical and operational causes and consequences of dehydration in these groups, examining both established and emerging monitoring techniques. This analysis is informed by recent advancements in wireless body sensor networks, as highlighted in studies such as those by Hao and Foster [24] and Hooper et al [25], and pays special attention to technologies that enable real-time monitoring, emphasizing their critical role in timely health intervention and preventive care. The rationale behind this scoping review was to comprehensively explore the landscape of wearable hydration-monitoring technologies, addressing key questions to enhance understanding and inform future research and development. Specifically, the objectives were as follows: (1) to examine the various types of sensors and technologies used for hydration monitoring, providing insights into their functionalities, advantages, and limitations; (2) to assess the trade-offs between different monitoring methods, highlighting the optimal use cases for each sensor type and category; (3) to identify trends and emerging techniques in wearable hydration monitoring, offering a forward-looking view of technology adoption and innovation; (4) to evaluate the precision and reliability of wearable devices compared to gold-standard methods, examining their performance across diverse populations and scenarios; and (5) to map commercial products in the market, analyzing their technological features, functionalities, and practical applications, bridging the gap between academic research and real-world implementations.

Methods

Overview

This review concentrated on wearable or portable devices, seeking to provide a detailed overview of recent advancements in hydration-monitoring technologies, including their capabilities and features. In addition, it aimed to assess the overall direction of development, research, and prototype design in this area. This scoping review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [26].

Eligibility Criteria

We focused on studies from 2014 to 2024 to capture the latest advancements in wearable hydration monitoring. The past decade has seen rapid progress in sensor miniaturization, real-time physiological monitoring, and artificial intelligence (AI)–driven analytics, making earlier studies less relevant to current technological capabilities. Studies were included if they met the following criteria: (1) focus on wearable or portable hydration-monitoring systems to assess real-world usability; (2) use of noninvasive or minimally invasive physiological biomarkers to ensure practical application; (3) direct addressing of body hydration, targeting biomedical applications; and (4) reliance on physiological biomarkers rather than methods such as body motion tracking or monitoring fluid intake or extraction through smart bottles or toilets as they do not assess direct hydration status.

Exclusion criteria included studies without human participant validation or relevant performance evaluations, ensuring the reliability of the findings. Other exclusion criteria included research not published in peer-reviewed journals or reputable conference proceedings to maintain study quality and non–English-language publications due to resource constraints and replication potential.

Information Sources and Search

The last literature search was conducted in January 2024. The search strategies were drafted and further refined through team discussion. The search incorporated keyword variations such as “hydration monitoring” and “wearable devices,” used advanced search functions in PubMed and IEEE Xplore’s Title and Abstract fields, and further extended to Google Scholar to ensure comprehensive literature coverage. The final search string—“(hydration monitoring OR hydration sensor) AND (wearable OR device)”—targeted literature published after 2014 and yielded the most complete results. Additional studies were identified through reference tracking of relevant papers to cover any potential missed papers. Other terms did not produce relevant results in either database and, therefore, were excluded. The final search string was designed to be broad and inclusive, allowing for the identification of numerous articles considered for inclusion in this study. The results duplicates were removed and then exported into a Mendeley library (Elsevier).

Selection of Sources of Evidence

A total of 551 articles were identified initially; PubMed returned 408 (74%) articles, IEEE Xplore returned 107 (19.4%) articles, and articles from additional sources were also included (n=37, 6.7% papers that were identified from Google Scholar search and the references of the included papers that did not appear in the initial search results). After duplicate records (5/551, 0.9%) were identified and removed using Mendeley’s automated duplicate detection, followed by manual verification to resolve mismatches across databases, of the 551 initial articles, 546 (99.1%) remained. Screening based on titles and abstracts led to the inclusion of 156 studies for full-text analysis. To ensure a rigorous selection process, 2 reviewers screened the titles and abstracts of all studies for relevance. To enhance consistency, all reviewers initially screened the same set of publications, discussed the results, and refined the screening and data extraction criteria before beginning the full review. Studies that met the inclusion criteria proceeded to full-text review, conducted sequentially by 2 reviewers working in pairs.

If a study’s title or abstract did not provide enough information to determine eligibility, particularly regarding the inclusion of validation procedures, a thorough examination of the full text was conducted. Disagreements on study selection and data extraction were resolved through discussion and consensus. When necessary, a third reviewer provided a final review to resolve conflicts and ensure consistency in the selection process.

XSL•FO
RenderX

Data Extraction and Charting Process

A data extraction and charting process was jointly developed by 2 reviewers to determine which variables to extract, including study characteristics, device specifications, sensor technology, validation methods, performance metrics, key findings, and limitations. Using these variables, structured summaries were extracted to synthesize the evidence and effectively answer the research questions and objectives. The 2 reviewers conducted data extraction using Mendeley, annotating each paper and summarizing key extracted data. To maintain consistency, a calibration exercise was conducted at the beginning of the extraction process where all reviewers screened a sample of publications, discussed discrepancies, and refined data extraction criteria. The charting form was continuously refined through an iterative process, with regular discussions among reviewers to update variables and improve clarity.

Commercial Product Search

In addition to academic research papers, we conducted a Google Scholar search to investigate commercial products in the field of hydration monitoring. We used keyword variations from the database searches, such as “hydration monitoring wearable device” and “hydration sensor technology.” Through online research, we identified several companies. Detailed evaluations were undertaken to compare the technological features, functionalities, and applications of these commercial products. This analysis offered valuable insights into practical implementations and advancements in the field, bridging the gap between academic research and real-world applications.

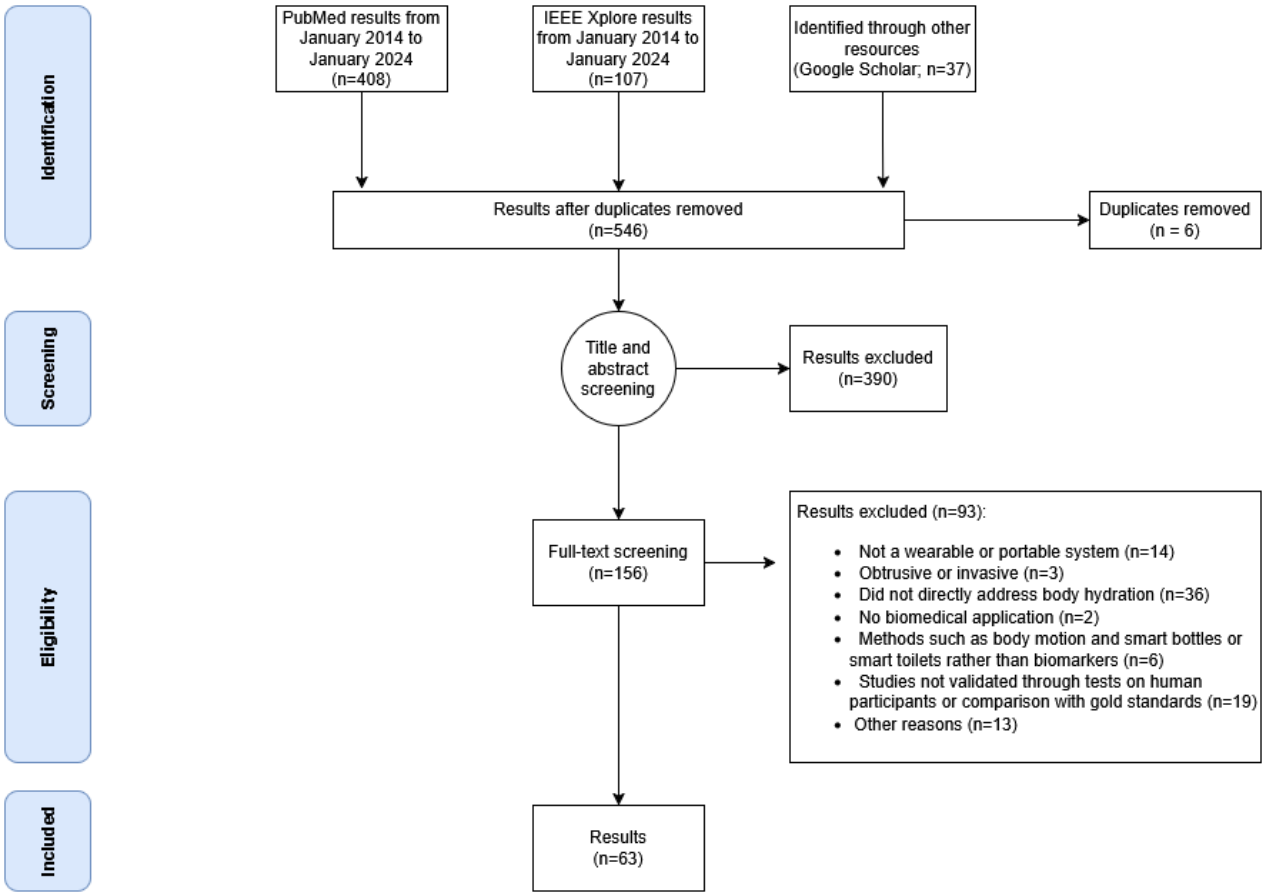
In our commercial product search, we differentiated between products designed for medical purposes and those intended for general hydration monitoring. Medical-grade devices were identified based on criteria such as Food and Drug Administration (FDA) approval, clinical validation, and their use in medical settings for patient care. These products are typically characterized by higher precision, sensitivity, and adherence to regulatory standards. In contrast, products aimed at general hydration monitoring are often marketed for consumer use, such as fitness and wellness tracking, and may not have undergone the same rigorous testing or obtained the necessary certifications for medical applications. This distinction was crucial in our analysis as it directly impacted the reliability and potential use cases of the devices evaluated. Specifically, we identified and compared 8 companies based on factors such as validation, FDA approval, and technology used.

Results

Overview

During the initial screening process, 546 articles were identified. After applying the inclusion and exclusion criteria, of the 546 articles, 156 (28.6%) were selected for full-text analysis. Following a thorough review, 63 articles were retained for synthesis and analysis (Figure 1).

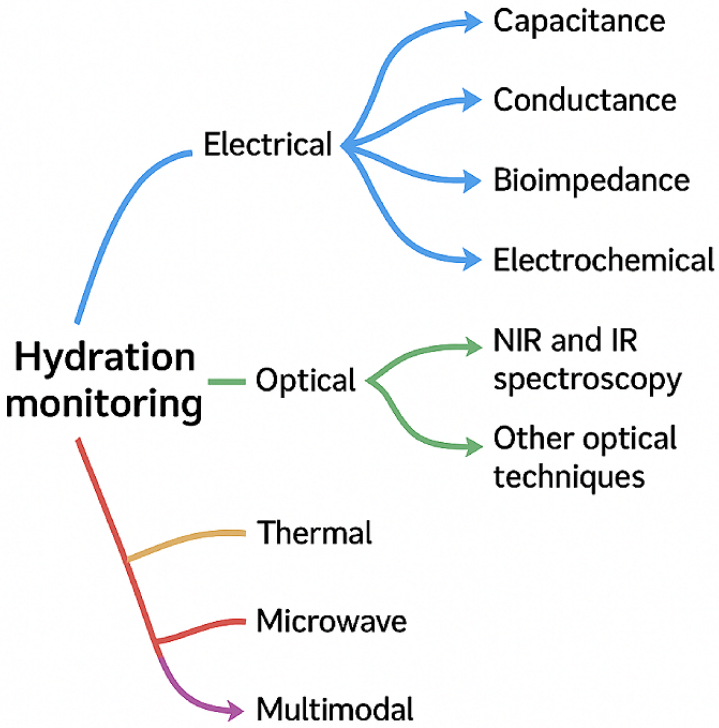
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the selection of studies on wearable and portable hydration-monitoring systems.



XSL•FO
RenderX

To organize the selected literature and facilitate a comprehensive analysis, a classification taxonomy was developed based on the types of sensors used for hydration monitoring in wearable technologies. This taxonomy categorized the papers into the following groups: electrical sensors, optical sensors, thermal sensors, microwave sensors, multimodal sensors, and commercial products (Figure 2). This classification enabled a structured review of the literature, highlighting the different sensor technologies used in hydration-monitoring devices.

Figure 2. Hydration-monitoring systems categorized based on the type of sensors used. IR: infrared; NIR: near infrared.



Electrical-Based Sensors

Overview

Electrical-based sensors for hydration monitoring use various techniques to measure changes in the electrical properties of the skin. They leverage the body's physiological changes in response to fluid imbalance. These sensors measure the electrical properties of the skin and underlying tissues, which vary according to the body's hydration state. The main methods used include capacitance, conductance, bioelectrical impedance analysis (BIA), and electrochemical analysis, each with its unique principles, advantages, and challenges.

Capacitance Method

This category includes sensors that monitor hydration by evaluating the skin's capacitance. Capacitance sensors measure the skin's ability to store an electrical charge, which fluctuates with hydration levels. When the body is hydrated, the water content in the skin is higher, increasing its dielectric constant and capacitance. As dehydration occurs, the water content decreases, lowering the skin's capacitance, as illustrated in Figure S1 in Multimedia Appendix 1 [16,27-30].

Yao et al [31] introduced a wearable skin hydration sensor (SHS) featuring silver nanowire electrodes. This sensor, with its flexible and stretchable nature, maintains consistent performance despite changes in humidity. It was calibrated against the MoistureMeterD commercial skin hydration measurement system [32] and packaged into a flexible wristband with an ultralow-power microcontroller unit (MCU) and Bluetooth capability. Future research should assess its adaptability to different skin layers and further develop its potential for widespread use.

Another study [33] presented a "skin hydration sensor patch" that integrates capacitive measurement with near-field communication (NFC) data transmission to smartphones. The patch demonstrated a high correlation with Corneometer results on various skin sites. A subsequent study with moderately dry skin in the face and forearm suggested upgrading the sensor probe calibration to enhance sensitivity. Future directions could include correlating changes in facial skin signs with varying hydration levels using AI-driven systems that analyze selfies.

In another study [34], capacitive textile sensors were developed for continuous skin hydration monitoring, with the goal of integrating them into everyday clothing. This study emphasized the sensors' potential for real-time health assessment and improved patient comfort. However, challenges related to sensor absorption, long-term performance, and measurement pressure were noted, suggesting that future research should address these issues to enhance the reliability and applicability of textile sensors.

Conductance Method

This section explores papers that focused on hydration monitoring through sensors that measure skin conductance. Conductance-based hydration monitoring assesses skin moisture

XSL•FO
RenderX

levels by measuring the ease of electrical current flow through the skin, with higher conductance indicating greater hydration.

Clarys et al [35] conducted a comparative analysis to quantify the differences between conductance-based and capacitance-based methods for hydration assessment. They examined 2 instruments, the Corneometer CM 825 (capacitance based) and the Skicon-200 EX (conductance based). Their in vitro and in vivo evaluations revealed that conductance measurements are influenced by electrolytes, whereas capacitance measurements are not. Capacitance allows for probing deeper skin layers (up to 45 µm) compared to conductance (up to 15 µm), and both methods showed strong correlations between water content and measurement accuracy.

Lu et al [15] introduced a portable device for diagnosing dehydration and its potential link to chronic kidney disease through saliva conductivity analysis. The device, consisting of an MCU, an analogue-to-digital converter, and a liquid crystal display, measures the electrical signals from saliva samples. Miniaturized electrodes reduce the sample contact area and the amount of saliva required for testing. However, significant variations in saliva conductivity due to factors such as diet, hydration status, and individual physiology present challenges to the device's accuracy and reliability.

Rizwan et al [16] proposed using galvanic skin response (GSR) as a noninvasive marker for hydration, collecting and analyzing GSR data from individuals in different hydration states and body postures (Figure S2 in Multimedia Appendix 1 [16,27-30]). The study demonstrated advancements in noninvasive hydration monitoring but suggested that further research is needed to refine algorithms and address variability in GSR measurements.

Madhvapathy et al [36] presented a versatile and cost-effective patch based on flexible printed circuit boards (PCBs) for skin hydration monitoring and dermatological diagnosis. The platform uses a thermistor to measure skin conductivity and diffusivity, enabling the measurement of volumetric water content up to approximately 1 mm in depth. The sensor communicates wirelessly with smartphones using NFC technology. However, limitations include insufficient knowledge about epidermis thickness, which cannot be directly inferred from thermal conductivity data.

Wang et al [37] introduced a wearable sweat sensor platform designed for the simultaneous measurement of sweat rate and total electrolyte concentration without calibration. The platform features a fluidic controlled design that minimizes dilution effects and detects sweat. Practical tests demonstrated its capability to measure regional sweat rate, sweat loss, and electrolyte concentration, particularly for high-sweat situations such as sports. However, the platform may face challenges with low flow rates, impacting real-time detection and introducing potential bias. Addressing this limitation could improve its accuracy and applicability in scenarios with varying sweat flow rates.

Responding to limitations in existing hydration-tracking methods, Liaqat et al [38] developed a novel system based on machine learning and deep learning techniques to accurately estimate hydration levels using GSR data. This hybrid model improved accuracy, making it valuable for applications in sports, health, and well-being.

BIA Sensors

BIA sensors differ from conductance-based sensors by measuring impedance, which includes both conductance and capacitance, as illustrated in Figure S3 in Multimedia Appendix 1 [16,27-30]. The working principle involves an initial rapid impedance change caused by electrolyte solution filling skin voids, followed by a gradual change associated with water gradient equalization and structural changes in the skin barrier [39].

Chua [40] introduced the i-Health Watch, an integrated wearable device for continuous monitoring of heart rate (HR), hydration level, and blood glucose concentration using bioimpedance. The compact prototype incorporates 2 pairs of silver electrodes with conductive gel, enabling real-time tracking. The watch's display provides continuous readings and issues on-screen alerts for abnormalities. The novelty lies in a parallel signal processing approach, deriving 3 physiological readings from a single bioimpedance feed, which streamlines the device. Initial testing on a single human participant demonstrated the potential for future integration into a smart telemonitoring health system.

Agcayazi et al [41] designed a wearable bioimpedance analyzer for infant hydration monitoring. The system uses an Analog Devices AD5933 impedance analyzer combined with an RFduino for Bluetooth communication with a smartphone. Validation on 2 adult participants using electrodes placed on specific body parts demonstrated a relative decrease in bioimpedance and an increase in hydration as participants consumed water progressively. However, the system requires initial calibration based on an individual's height and weight, which may limit its applicability in certain scenarios.

Yang and Rosa [42] introduced a small, batteryless wearable device for assessing skin hydration levels. The device combines sunlight exposure computation with skin impedance and temperature measurement. The goal was to incorporate it into a bracelet or ring for simultaneous UV index, skin hydration, and temperature measurement, which current UV trackers do not offer. Despite its simplicity, affordability, and energy efficiency, the device proves useful for detecting skin impedance variations due to water loss from sunlight exposure. Testing during the summer is crucial to establish correlations between skin impedance changes and higher UV indexes.

Sunny et al [43] developed a cost-effective bioimpedance sensor that uses 4 gel electrodes on the skin surface to mimic skin properties and detect hydration changes. Accuracy was validated across a range of skin tones and types. The sensor shows promise for detecting hydration changes in both phantom and human skin experiments. Future research could focus on enhancing the sensor's performance and usability in portable or wearable modes and standardizing experimental setups.

Leonov et al [44] introduced a sensor with significant sensitivity to hydration changes, detecting sweat loss with a limit of 55 mL (approximately 1.86 oz) in the worst-case scenario. Although the device showed sensitivity to hydration changes of approximately 700 mL (approximately 23.67 oz), more

XSL•FO
RenderX

refined protocols and device improvements are necessary to enhance reliability for practical daily-life applications.

Matsukawa et al [27] presented a method for continuous monitoring of skin hydration using nanomesh electrodes attached to the left ventral forearm. These electrodes adhere comfortably to the skin and allow for water vapor permeability during extended wear, enabling impedance measurements without pressure application. Impedance results correlated with hydration levels measured using a Corneometer. However, intense sweating may cause short-circuiting between electrodes due to the lack of an encapsulation layer. The research plan addresses this challenge by adding an encapsulation layer.

Veeralingam et al [45] developed a novel "body hydration analysis system" to enhance running performance by assessing skin hydration levels through impedance variations. The system uses the open-source QueSSence board and an AI-driven k-nearest neighbor algorithm for efficient data analysis. On the basis of a semiconductor compound with high electronic properties, the sensor demonstrated high conductivity, ideal for wearable applications. However, further validation and integration of these technologies into user-friendly wearable devices for real-time monitoring during athletic activities are necessary.

Valentin et al [21] created a wearable MCU-based BIA device for precise hydration measurements with low power consumption. Electrodes were placed below the right shoulder blade and on the bottom left of the abdomen. After an initial measurement, participants consumed 200 mL (approximately 6.76 oz) of water, and a second measurement was taken after 15 minutes. The experiment resulted in an impedance recording accuracy of <2%, suggesting potential for future clinical applications. Further research should address limitations such as sample size and long-term monitoring to validate the method for clinical use.

Songkakul et al [18] introduced a miniaturized wearable bioimpedance spectroscopy system for continuous tissue hydration monitoring, focusing on electrode optimization and advanced data processing. The system comprises Bluetooth connectivity and an analogue front-end circuit integrated with conformable, flexible, and stretchable silver nanowire electrodes. The system operates with a lithium battery providing 18 hours of functionality. The 4-electrode configuration uses 2 electrodes for current source and sink, whereas voltage is measured across the other 2 electrodes. Further research is needed to validate the system in vivo and improve its accuracy and usability in real-world applications.

AlDisi et al [17] developed and tested 2 wrist-worn interdigitated electrode designs for assessing hydration levels using BIA. The study demonstrated that interdigitated electrode designs could accurately measure hydration without complex models, addressing classification challenges and impedance reading influences. The experiments involved 6 participants classified into severe dehydration, mild dehydration, and hydration states. Although promising for hydration assessment without complex calibration, further clinical studies with a larger sample size are necessary to establish robust classification criteria and address variations between body sizes and testing sessions.

Thomas et al [19] introduced a molybdenum diselenide and polyvinyl alcohol–based wearable platform for pulse rate monitoring and skin hydration sensing. The molybdenum diselenide and polyvinyl alcohol compound behaves like a combination of resistance and capacitance in parallel, with the skin acting as resistance when the sensor is applied. Experiments involving moisturizer application and changes in relative humidity demonstrated the sensor's responsiveness and accuracy in detecting skin hydration levels. While the fabrication process is described as simple and cost-effective, it still requires specialized techniques such as hydrothermal synthesis, which may limit scalability and accessibility for widespread adoption.

Jang et al [46] presented a textile-based wearable sensor designed for stable and reliable monitoring of skin impedance changes related to hydration levels. The sensor can detect different hydration levels across various body parts while maintaining stable skin contact. Time-dependent monitoring of skin hydration on the hand demonstrated the sensor's ability to detect changes over time, with reliable results and fewer SDs than ground-truth sensors. Future research could focus on designing more intricate electrodes to improve sensing performance or adjust for different skin colors and types.

Tonello et al [20] described a multisensory wearable that integrates flexible sensors on a bracelet-shaped substrate for temperature, body impedance, and skin hydration monitoring. The device measures total impedance from 3 sensors: a resistance temperature detector, a BIA sensor, and a hydration status sensor. This combination enables continuous, noninvasive hydration level measurement through temperature, local hydration monitoring, and global body composition analysis. Preliminary in vivo validation on 2 volunteers demonstrated the sensors' capability to detect different hydration conditions. Future work will focus on refining sensor design and conducting reproducibility and stability tests.

SkinUp [47] measured skin moisture and oil levels using impedance, validated against the Corneometer CM 825, offering a portable solution for skin hydration measurements across various regions, including the forearm, cheeks, and forehead.

Electrochemical Analysis

Electrochemical analysis uses various techniques to monitor electrolyte concentrations [48]. These include sweat sensors that use ion-selective electrodes (made of field-effect transistors) with microfluidics and low-noise electronics for precise electrolyte level monitoring [49,50]. Other approaches include waterproof, epidermal microfluidic devices for capturing and analyzing sweat underwater, as well as chemiresistor-based sensors for monitoring sweat dynamics [51,52].

In their study, Culver et al [48] focused on monitoring electrolytes in sweat using a wearable sweat sensor. The objective was to evaluate the sensor during a simulated special operations field event conducted at the US Air Force Academy. While the sensor showed promise for hydration monitoring, challenges arose from limited sweat availability for detection. The study explored sensor placement on the lower back, a region known to produce more sodium-rich sweat. The device incorporates ion-selective electrodes with microfluidics and

XSL•FO
RenderX

low-noise electronics to ensure precise and reliable monitoring of sodium ion and potassium ion levels in sweat [49]. Challenges include improving sensor shelf life and stability and reducing its footprint.

Reeder et al [51] introduced waterproof, epidermal microfluidic devices for capturing and analyzing sweat underwater. Designed for athletics and fitness, these devices allow for real-time monitoring of fluid loss and electrolyte concentrations in aquatic environments. Field trials have demonstrated quantitative in situ measurements of sweat chloride concentration, local sweat loss, and skin temperature during physical activity. The devices are made of soft, waterproof materials and are designed for seamless operation in extreme environments. Future research aims to explore the physiology of aquatic sweating and validate the device's utility in various settings.

Parrilla et al [52] introduced a wearable paper-based chemiresistor for monitoring sweat dynamics, focusing on sweat rate and loss. This sensor measures volume of aqueous solution at the microliter scale by detecting resistance changes along a conductive paper substrate. The study outlined the sensor's analytical performance, sensing mechanism, and effectiveness in monitoring sweat loss during exercise. Challenges include optimizing sensitivity and reliability across different conditions.

Lafaye et al [50] presented a real-time multi-sensing wearable platform for continuous sweat biomonitoring. One prototype uses an ion-sensitive electrode soft-sensing patch, whereas the other uses silicon-based sensors and paper microfluidics. Both platforms integrate multisensory arrays for measuring sodium, potassium, and pH in sweat. Initial results from prototypes placed on athletes during exercise have demonstrated the platform's efficacy. Body water loss (BWL) percentage was used as a quantitative indicator of the participants' hydration status. Machine learning algorithms predict BWL from biomarkers such as HR and sweat sodium concentration. Future work includes expanding experiments with more participants and correlating sweat sensor recordings with other biomarkers.

Yang et al [53] highlighted a cost-effective wearable sweat sensor using screen printing technology for real-time monitoring of potassium ion and sodium ion concentrations in human sweat. A 10-day continuous monitoring experiment revealed a close relationship between potassium ion and sodium ion concentrations and hydration status. The fabricated sensors demonstrated sensitivity, linearity, repeatability, resistance to interference, and mechanical deformation resistance, making them suitable for sweat-sensing applications.

Comparative Overview of Sensor Techniques

These various techniques for monitoring skin hydration levels each have distinct strengths and limitations. The selection of an appropriate sensor depends on the specific application and user preferences, considering factors such as precision, comfort, ease of use, and the duration of wear. Table 1 summarizes the findings of electrical-based sensors used for hydration tracking.

Table 1. Comparative analysis of skin hydration–monitoring techniques.

Sensor type	Strengths	Limitations	Studies
Capacitance	Extremely sensitive to skin moisture variation and simple to use	Potentially less precise in measurement	• Yao et al [31] • Flament et al [33] • Balaban and Blecha [34]
Conductance	Balanced accuracy and user-friendly	May not be as comprehensive in assessment	• Lu et al [15] • Rizwan et al [16] • Madhvapathy et al [36] • Liaqat et al [38]
Bioimpedance	Comprehensive assessment and combines multiple metrics	Limited long-term wearability and requires continuous contact	• Chua [40] • Agcayazi et al [41] • Sunny et al [43] • Leonov et al [44] • Matsukawa et al [27] • Veeralingam et al [45] • Valentin et al [21] • AlDisi et al [17] • Jang et al [46]
Electrochemical	High sensitivity and specificity and detects specific ions	Complex setup, requires careful calibration, and potential limitations in practicality	• Culver et al [48] • Reeder et al [51] • Parrilla et al [52] • Yang et al [53]

Optical-Based Sensors

Overview

Typically, noninvasive biomedical measurements involve directing light of a specific wavelength onto the skin to collect data. A sensor then detects the light that is either reflected, absorbed, or refracted. These data are used to quantify biomedical information. The wavelength of the light is critical in determining the depth of its penetration into the skin when transmitting an optical signal (Figure S4 in Multimedia Appendix 1 [16,27-30]). This aspect primarily distinguishes the techniques reviewed in this section. Moreover, these methods

XSL•FO
RenderX

allow researchers to assess the physiological condition of internal organs through the skin [54].

Near-Infrared Spectroscopy

In recent years, near-infrared spectroscopy (NIRS) has become a powerful tool for comprehensive skin hydration analysis within the optical range of 750 to 2500 nm. It is selected for its ability to capture absorption bands of water, proteins, lipids, and other skin constituents, enabling precise measurement of hydration levels.

In 2013, Qassem and Kyriacou [55] used NIRS to investigate the effects of short-term skin contact with water and subsequent moisturizer application on human skin. The study involved in vivo measurements within the range of 9,002,100 nm, highlighting variations in peak values regarding water overtone and combination bands, providing insights into distinctions between skin types and moisturizer use patterns.

Visser et al [56] developed a portable prototype using an Arduino Uno32 MCU and 2 infrared LEDs emitting light at 1300 and 1480 nm. These wavelengths were chosen for their different water absorption coefficients, with 1300 nm selected for maximum skin penetration depth and 1480 nm selected for sensitivity to water concentrations. The study noted greater variability in hydration status among adult patients than among infants, emphasizing the need for universal calibration and validation protocols. Further research is required to ensure the sensor's effectiveness across diverse demographic groups.

In another study, Visser et al [57] investigated the use of optical sensors for assessing infant dehydration. An infrared spectrometry sensor using infrared LEDs at 1300 and 1480 nm was used to maximize skin penetration depth and sensitivity to water concentrations. Conducted on 10 infants with acute gastroenteritis, the study demonstrated promising results, showing high specificity and sensitivity of the infrared spectrometry sensor in dehydration assessment. Challenges such as infant movement artifacts and data irregularities were mitigated using supplementary algorithms and repeated measurements. A larger population sample is needed before clinical deployment.

Mamouei et al [58] developed a portable sensor for skin hydration monitoring targeting dermal water content. The multiwavelength optical sensor is designed for continuous and nonintrusive monitoring using 4 LEDs emitting light at 940, 970, 1200, and 1450 nm, 3 of which correspond to water absorption peaks in the near-infrared region. Validation involved benchmarking accuracy against a high-end broadband spectrophotometer. Future work will focus on establishing appropriate algorithms for sensor calibration in the absence of reference gravimetric data, addressing limitations, and enhancing the sensor's accuracy and usability.

Benavides et al [59] introduced a novel wearable medical device for continuous UV exposure and hydration monitoring. The device uses infrared spectroscopy, with an LED emitting light at 940 nm and a photodiode detecting unabsorbed light to measure water absorption in the skin. This enables real-time hydration assessments through a paired mobile app. Table 2 summarizes the key biomarkers and their wavelength ranges as reported in NIRS studies.

Table 2. Near-infrared spectroscopy biomarkers, wavelength range, and studies.

Biomarker	Wavelength range	Studies
Water content	1300 and 1480 nm	Visser et al [56,57]
Water absorption peaks	940, 970, 1200, and 1450 nm	Mamouei et al [58]
Water content	940 nm	Benavides et al [59]
Water content and sweat	940 and 1450 nm	Volkova et al [60]

Other Optical-Based Techniques

Other optical sensors are increasingly being used for noninvasive hydration monitoring. These sensors use various techniques such as spectroscopy across a wider range of wavelengths and photonics to analyze biomarkers in sweat or tissue.

Curto et al [61] introduced a self-contained wearable designed for real-time analysis of pH levels in sweat. The platform incorporates a surface-mount LED (smLED) and a photodiode controlled by a LilyPad Arduino MCU. Detection occurs through the photodiode and smLED positioned above and below the sensing area, respectively. Replacing standard LEDs with smLEDs reduced the bulkiness of the system. Challenges include refining sensitivity, specificity, and accuracy, as well as expanding the range of detectable analytes.

Ozana et al [62] introduced a wearable optical sensor for noninvasive glucose concentration detection and dehydration level assessment. The sensor uses 2 optical techniques: remote vibration source extraction and polarization rotation of light. It includes 4 LEDs emitting wavelengths from 600 to 1150 nm and a green laser (532 nm), along with a camera connected to a computer in a braceletlike setup. Dehydration measurements were taken in a 50 °C chamber. Challenges include implementing a motion cancellation mechanism and developing a robust automatic calibration process.

Perkov et al [63] investigated optoacoustic monitoring for assessing water content in tissue. Known for its optical contrast, ultrasound resolution, and significant penetration depth, the optoacoustic technique shows promise for skin hydration assessment. The analysis was conducted across wavelengths from 1370 to 1650 nm to optimize water content assessment in skin tissues. Challenges include optimizing wavelengths to fully exploit the potential of optoacoustic monitoring. A miniaturized complementary metal oxide semiconductor spectrometer in the 650,900-nm wavelength range designed for continuous monitoring of dermal skin hydration is presented in Figure S5 in Multimedia Appendix 1 [16,27-30]. Fabricated using a

XSL•FO
RenderX

monolithically integrated filter process, the spectrometer ensures cost-effectiveness, low power consumption, and mass production capability. In the same study, Perkov et al [63] also highlighted specific wavelengths, such as the 970-nm water absorption and 930-nm fat absorption peaks, showing the potential of the device for wearable skin biomarker measurements and its versatility beyond hydration to applications such as HR and subdermal markers such as glucose or lactate detection [28]. However, the study was limited by a small sample size and lacked extensive validation across diverse populations and clinical conditions.

Sandys et al [64] conducted a pilot observational study to evaluate the Sixty wearable hydration monitor in patients undergoing hemodialysis. The device uses diffuse reflectance spectroscopy and machine learning to detect subdermal fluid levels. Photonic sensors and wavelengths from 530 to 950 nm were used to determine fluid status based on reflected light. The study compared the device's accuracy with bioimpedance measurements during dialysis sessions and overnight for 3 weeks. Successful validation could lead to integration into a comprehensive algorithm for managing fluid overload in interdialytic periods.

Bohman et al [65] presented a silicon photonics-based spectrophotometer for monitoring temperature and water content in tissue-simulating phantoms using semiconductor lasers. The spectrophotometer demonstrated its utility in tracking temperature and water content, showing excellent agreement with reference values. However, variables such as the complexity of human tissue could affect precision in real-world scenarios.

Summary of Optical Techniques and Biomarkers

Table 3 summarizes the various optical techniques and the biomarkers used for hydration tracking.

Table 3. Optical techniques and biomarkers for hydration tracking.

Biomarkers	Technique used	Studies
pH levels in sweat	smLED[a] at visible light (300-700 nm) and LEDs within the 600-1150 nm range	• Curto et al [61] • Ozana et al [62]
Temporal changes in reflected speckle patterns were analyzed to assess skin water content	Optoacoustic monitoring (1370-1650 nm)	• Perkov et al [63] • Van Beers et al [28]
Localized PWC[b] dermal fluid status	CMOS[c] spectroscopy (650-900 nm)	• Sandys et al [64]
Temperature and water content	Photonic sensor (530-950 nm)	• Bohman et al [65]

[a]smLED: surface-mount LED.

[b]PWC: percentage of water content.

[c]CMOS: complementary metal oxide semiconductor.

Thermal Sensors

Thermal methods for assessing hydration involve measuring the skin's thermal properties, such as temperature and thermal transport, which can indicate hydration levels and blood flow dynamics. Madhvapathy et al [66] described the development of soft, skinlike thermal depth sensors to track hydration levels. The sensor measured the thermal properties of human skin at depths up to 6 mm (approximately 0.24 in) beneath the skin's surface, addressing limitations of current methods confined to superficial layers in clinical environments.

Krishnan et al [67] introduced a wireless, battery-free sensor system for measurement of the skin's thermal properties. The device comprises a wireless power-harvesting system, NFC-based data transmission, analogue signal conditioning on a flexible PCB, and a stretchable sensor with a temperature coefficient of resistance for precise temperature measurements. Human trials demonstrate continuous monitoring of skin thermal properties over a week without interruption. Challenges involve ensuring the accuracy and reliability of the data collected. Kwon et al [68] introduced a wireless, soft electronics platform designed for rapid measurements of hydration levels. Recent alternatives using thermal measurements with soft wireless devices have limitations such as a restricted operating range (1 cm) and sensitivity to environmental fluctuations. In response, this study presented innovative technologies to overcome these drawbacks, enabling high-speed, robust, automated measurements of thermal transport properties. The sensor module was controlled by a Bluetooth Low Energy (BLE) system on a chip with a smartphone interface.

Shin et al [69] presented a novel wireless and soft SHS designed for rapid and accurate diagnostics of dermatological health. The SHS integrates a BLE system on a chip within a flexible PCB. Pilot trials involving >200 patients in a dermatology clinic highlighted the practical applicability conducted at 3 skin locations (forehead, lower leg, and lower arm) using both the SHS and a commercial system (Delfin). The SHS operates without applied pressure, adheres gently to soft and curved skin regions, and provides rapid and objective measurements. Future research could improve this sensor by tackling potential limitations, including scalability, long-term reliability, and user comfort.

Microwave-Based Sensors

Microwave-based sensors use microwave or electromagnetic signals to measure various hydration-related parameters. Butterworth et al [70] presented a wearable wristband for hydration monitoring using noncontact dielectric spectroscopy in the microwave range (2-6 GHz). This technology leveraged minute variations in wrist hydration, offering a unique approach

XSL•FO
RenderX

to tracking overall hydration status. Potential areas for improvement include further validation studies involving a larger and more diverse sample of participants. The paper by Wang et al [71] presented a wearable radio frequency device for noninvasive real-time hydration monitoring, correlating the received signal strength indicator with BWL percentage. The device comprised near-field antenna sensor nodes, a radio frequency front end, and digital processing units. Challenges involve further validation and refining the accuracy and usability of the device.

Schiavoni et al [72] investigated a time-domain reflectometry (TDR)–based wearable skin hydration–sensing system, an electromagnetic technique showing accuracy, cost-effectiveness, and portability for potential medical applications. Challenges may involve ensuring reliability and connecting TDR measurements with physiological parameters.

In another study, the system by Schiavoni et al [73] combined TDR and frequency-domain data extraction in microwave reflectometry, developing calibration curves linking skin dielectric permittivity to frequency-domain responses, showing promise for real-time skin hydration monitoring. Challenges may include system validation, portability, and integration into wearables.

Besler and Fear [74] investigated microwave-based hydration assessment with fasting volunteers during Ramadan using a time-of-flight permittivity estimation technique to measure hydration changes throughout the day. Challenges include improving sensitivity and precision in detecting subtle hydration changes.

Cataldo et al [75] developed a flexible wearable for skin hydration sensing. The sensor underwent redesign for enhanced wearability, sensitivity, and patient comfort, incorporating a flexible rubber substrate with a Kapton layer. Future efforts will focus on system improvement and systematic characterization, including solutions for Bluetooth data acquisition and local processing. Their goal is to further miniaturize the system for full wearability by removing the liquid crystal display screen.

Bing et al [76] introduced a small, planar resonant loop sensor for water content monitoring based on electromagnetic resonance. Experiments on human hydration processes align well with simulations using documented skin permittivity properties. The wearable sensor, integrated on the human forearm, offers discrete and continuous measurements. It is crucial to note that the sensor's sensitivity is influenced by dielectric property changes due to water content variations.

Multimodal Sensors

This section explores research on multisensory systems designed to provide a comprehensive assessment of hydration levels by integrating various sensor types and combining various measurement techniques. Krishnan et al [77] introduced multimodal sensors for precise, quantitative in vivo monitoring of hydration levels in near-surface skin regions. The mathematical model they developed encompasses temperature, thermal conductivity, thermal diffusivity, volumetric heat capacity, and electrical impedance using a simple analysis algorithm. Challenges include evaluating the device's performance in real-world scenarios with varying temperature and humidity.

Salvo et al [78] developed a wearable sensor for real-time sweat rate monitoring using the open-chamber method based on Fick's first law of diffusion. The cylindrical chamber within a 3D-printed adapter calculated the sweat rate by measuring water vapor flow from the skin. The sensor, tested on 13 participants during a cycling test, consisted of an Arduino Pro Mini and a PCB with a humidity and temperature sensor (SHT25) and was compared to a DermaLab commercial device. Further testing in uncontrolled environments is needed to address factors such as air movement and humidity, which may affect the evaporation rate.

Bandodkar et al [29] introduced a wearable sweat-sensing platform combining battery-free, wireless electronic detection with integrated colorimetric assays (Figure S6 in Multimedia Appendix 1 [16,27-30]).

Lapadula et al [79] presented a system for analyzing body hydration to enhance running performance. It combined BIA as a benchmark with data from a Garmin Vivoactive device and a custom mobile app. The Garmin device recorded parameters such as HR, speed, altitude, calories, distance, and steps, whereas saliva analysis using cyclic voltammetry was used to extract hydration-related features. Challenges include validating the measurements and integrating these technologies into user-friendly wearable hydration monitoring for athletic activities.

Kamran et al [23] developed a data-driven method to measure hydration status using wearable sensors and normal orthostatic movements. Logistic regression models were trained to estimate dehydration status based on HR responses to postural movements, showing that shorter orthostatic tests achieved comparable accuracy to clinical tests. The study suggests that the sensor can accurately estimate mild dehydration in athletes. Challenges include expanding the range of movements and addressing study limitations such as the controlled exercise environment and specific postural movements used.

Sabry et al [80] used machine learning for on-device dehydration monitoring by integrating data from various wearable sensors such as accelerometers, magnetometers, gyroscopes, GSR, photoplethysmography, temperature, and barometric pressure. The focus on developing models suitable for memory-constrained wearable devices highlights their practical potential. Challenges include ensuring model accuracy while minimizing power consumption.

Rodin et al [81] assessed the real-world performance of a wearable body hydration sensor that integrates photoplethysmography and galvanic biosensors. The study involved 240 participants performing treadmill exercises over 90 minutes with intermittent rest periods to evaluate the sensor's accuracy in monitoring water mass loss due to perspiration. The sensor, attached to a smartwatch, showed strong agreement with the gold-standard method (body mass change measured using a medical balance). The system uses proprietary algorithms to estimate sweat volume and a galvanic contact system to estimate

XSL•FO
RenderX

total BWL, showing promising potential as an accurate wearable hydration monitor in practical real-world scenarios.

Wang et al [82] conducted a study predicting hydration status based on single-participant experiments involving 32 moderate-intensity exercise sessions with and without fluid intake. A total of 4 noninvasive physiological and sweat biomarkers—HR, core temperature, sweat sodium concentration, and whole-body sweat rate—were measured during exercise. Machine learning models were used to determine BWL percentage as an indicator of dehydration. The models revealed that whole-body sweat rate and HR achieved the highest accuracy, with sweat sodium concentration from the arms showing the best prediction accuracy. Challenges include understanding hydration's impact on biomarker relationships with BWL percentage and highlight the need for future extended studies with multiple participants and intensity variations.

Commercial Products and Comparisons

Overview

This section explores innovative hydration-monitoring solutions from leading companies. These products use advanced technologies, offering personalized insights through wearable patches, wristbands, and electronic biosensors. Table 4 provides an overview of various hydration-monitoring products, each with unique features and technologies.

Table 4. Overview of wearable hydration-monitoring system products.

Category	Name	Features	Validated	Commercial	Company	Technology
Wristband	Bioptx [83]	Hydration, temperature, HR[a], HR variability, respiratory rate, and blood saturation monitoring	No	No	Rockley	IR[b] laser technology (36 wavelengths), PPG[c] (green, red, and IR), and accelerometer
Wristband	Sixty [84]	Hydration, HR, activity levels, calories burned, and sleep tracking	No	No	Sixty	Optical spectrometry; 3 LEDs shine green, red, and IR light
Wristband	LVL [85]	Hydration, activity, sleep, mood, HR, and calorie monitoring	No	No	BSX Athletics	Red light technology
Wearable band	hDrop Gen 2 [86]	Hydration and body temperature monitoring	No	Preorder	hDrop Technologies	Electrode tracks sweat loss and rate and sodium and potassium levels
Smartwatch	Geca sensor [87]	Hydration monitoring	Yes	Preorder	Hydrostasis	Optical spectroscopy and detection of fluid concentration in the skin
Smartwatch	Aura Strap [88]	HR, blood saturation, and hydration monitoring	No	No	Apple	Electrode measures electrolytes in sweat to monitor hydration
Smartwatch	Gx Sweat Patch [89]	Hydration, sweat, electrolyte content, and body temperature monitoring	Yes	Yes	Epicore Biosystems	A thin microfluidic substrate on the skin that captures sweat
Smartwatch	Hydration Biosensor [90]	Hydration monitoring	Yes	Yes	Nix	Electrochemical biosensors use electrodes to detect biomarkers in sweat [90]

[a]HR: heart rate.

[b]IR: infrared.

[c]PPG: photoplethysmography.

Rockley Bioptx

Bioptx is a noninvasive hydration-monitoring wristband using silicon photonics-based sensors (Figure S7 in Multimedia Appendix 1), which detect changes in water concentration in the body. Rockley's platform is aimed at athletes, health-conscious individuals, and those managing hydration-related conditions [91].

Sixty

Sixty offers a hydration-monitoring solution with simple design and LED indicators. Users can choose between continuous monitoring using a wrist or arm strap or occasional checks using the stand-alone monitor. This flexibility caters to both athletes needing constant monitoring and individuals seeking occasional hydration updates. The device can detect fluid loss as low as 30 to 40 mL (approximately 1.35 oz), providing timely reminders before the sensation of thirst, typically felt after losing 300 to 400 mL (approximately 13.53 oz) of fluid. Sixty plans to gather extensive data to ensure accuracy across different phenotypes [91,92] (Figure S8 in Multimedia Appendix 1 [16,27-30]).

LVL Wristband

The LVL wearable hydration monitor uses red light technology, which penetrates deeper into the body than conventional fitness trackers that use green light sensors. Designed for individuals seeking comprehensive health monitoring, LVL calculates fluid requirements based on hydration status and sweat rates. The device emphasizes the connection between hydration and

XSL•FO
RenderX

improved sleep quality and offers compatibility with various devices through BLE technology [85] (Figure S9 in Multimedia Appendix 1 [16,27-30]).

hDrop Gen 2

hDrop Gen 2 from hDrop Technologies provides rapid, direct skin contact readings for sweat rate, electrolyte concentration, and estimated sodium and potassium levels. Primarily aimed at athletes, this technology uses electrodes to offer personalized recommendations for fluid and electrolyte replenishment, improving hydration management during workouts [93] (Figure S10 in Multimedia Appendix 1 [16,27-30]).

Hydrostasis Geca Sensor

The Geca sensor is an optical-based wristband that underwent rigorous evaluation, including comparisons with known hydration and dehydration events and urine-specific gravity. Consumer pilot studies show the sensor's potential to accurately measure hydration changes. It demonstrates greater sensitivity in predicting hydration levels associated with minimal total body weight changes ranging from 0.5% to 1%, outperforming conventional salivary osmolality tests and standard physical assessments of dehydration [94] (Figure S11 in Multimedia Appendix 1 [16,27-30]).

Aura Strap

The Aura Strap integrates smart functionality into an Apple Watch band to track hydration levels using BIA. It also provides insights into body composition, including body fat, muscle mass, minerals, and body water, when paired with user information such as age, gender, and weight. Measurements are taken by placing the sensor on a specific part of the hand, with data transmitted to the Apple Watch and stored in the dedicated app. However, it only offers spot measurements rather than continuous monitoring, and improvements in size and ease of use are needed [95].

Epicore Gx Sweat Patch

Epicore Biosystems' Gx Sweat Patch is a wearable technology aimed at athletes seeking personalized hydration insights and recovery guidance. The skinlike patch pairs with the Gatorade Gx iOS app, using microchannels to capture sweat data in real time. The app provides athletes with personalized recommendations based on their sweat profiles, catering to their unique hydration and refueling needs. A paper was published detailing their wearable microfluidic-based system for real-time analysis of sweat rate and sweat chloride concentration, validated through studies involving 312 athletes [89,96] (Figure S13 in Multimedia Appendix 1 [16,27-30]).

Nix Hydration Biosensor

The Nix Hydration Biosensor is a lightweight, compact electronic device designed for durability and reusability. It attaches to an individual sweat patch and continuously transmits data to a smartphone, watch, or bicycle computer, with a battery life of up to 36 hours, making it suitable for long training sessions. The Nix app offers real-time data for monitoring fluid and electrolyte loss, historical data to understand sweat profiles, and predictive data that calculate future hydration needs based on weather forecasts and past performance [90].

Discussion

Principal Findings

Overview

The studies revealed that many wearable devices are either in the prototype phase or tailored to specific user groups. Table 5 provides a concise overview of the distribution of publications across various hydration-tracking techniques, categorizing them into electrical, optical, thermal, microwave, and multimodal methods. This categorization sheds light on the current landscape of research in the field [83].

Table 5. Number of publications by sensor category (N=63).

Sensor category	Publications, n (%)
Electrical	31 (49)
Optical	13 (21)
Thermal	4 (6)
Microwave	7 (11)
Multimodal	8 (13)

Electrical-Based Sensors

Electrical-based methods, particularly BIA, are currently leading the field, with 49% (31/63) of the publications underscoring their widespread use and popularity. These methods are favored because they directly correlate the skin's electrical properties (such as impedance, capacitance, and conductance) with hydration levels. This relationship allows for the efficient measurement of hydration states through noninvasive, rapid, and quantitative assessment. In addition, BIA systems stand out due to several practical advantages, including their accuracy, user-friendliness, portability, and affordability. These characteristics make them particularly suitable for applications in dermatology, cosmetics, and material testing, as well as for clinical and field-based studies requiring real-time monitoring [17,68].

Despite these advantages, electrical-based methods, including BIA, face significant limitations. One major drawback is their limited ability to assess hydration in deeper dermal layers. While these methods excel at evaluating hydration within the outermost layer, the stratum corneum, they may not provide comprehensive insights into fluid distribution deeper in the skin, such as in the epidermis or dermis. This limitation can be critical in contexts in which deeper hydration dynamics are integral to

XSL•FO
RenderX

understanding skin health or the effectiveness of topical treatments. In addition, these systems are highly sensitive to external factors, including environmental conditions [97] (eg, humidity and temperature) and physiological influences (eg, skin temperature, sweat, and lipid content). For instance, higher ambient humidity or temperature can artificially increase skin conductance by softening the stratum corneum or inducing sweating, leading to overestimated hydration levels. In fluctuating or extreme environmental conditions, these sensitivities may compromise the accuracy and reliability of hydration measurements unless properly controlled or calibrated [98].

Efforts have been directed at refining BIA for more robust performance, such as using multi-frequency impedance measurements to capture both extracellular and total water content, integrating environmental monitoring and machine learning for calibration, and advancing hybrid systems that combine BIA with complementary methods. These innovations aim to mitigate environmental and depth measurement limitations while enhancing the method's sensitivity and validity. As a result, BIA remains a leading focus in skin hydration assessment research, offering high utility and adaptability and the potential for further performance optimization in diverse settings [17,68,69].

Electrochemical methods, which monitor hydration by measuring sodium ion and potassium ion levels, offer higher accuracy due to their specificity in monitoring electrolyte levels. However, they face challenges such as ion selectivity issues and electrode conditioning problems [99]. In addition, variability in fluid extraction processes across individuals adds complexity to achieving consistent measurements.

Each subcategory of electrical-based sensors offers distinct advantages and limitations. Electrical sensors are ideal for quick, noninvasive assessments, particularly in sports and fitness settings in which surface hydration measurements suffice. In contrast, electrochemical sensors are more suited for applications requiring precise electrolyte monitoring, such as medical diagnostics and sports science.

Optical-Based Sensors

Optical-based hydration tracking techniques were assessed in 21% (13/63) of the publications. While not as widespread as electrical methods, they provide distinct advantages, particularly their ability to deliver detailed molecular information about skin hydration. The smaller number of publications in this area may be due to the complexity and cost of advanced optical sensors. Recent advancements in LED-based and semiconductor photonic sensors have addressed some of these challenges by offering smaller sizes and lower costs than traditional laser-based laboratory equipment. These developments in semiconductor miniaturization and photonics technology present promising opportunities for hydration monitoring.

While infrared spectroscopy and NIRS focus on water absorption at key wavelengths (notably, 1300 nm for maximum skin penetration and 1480 nm for sensitivity to water concentrations), some studies (3/63, 5%) extended this range to include 940 nm to encompass sweat properties. Additional frequencies can lead to improved accuracy. However, the challenge remains in refining sensor sensitivity and specificity. Achieving consistent accuracy across different skin types, conditions, and environments is a significant hurdle, and variability in individual skin properties and external factors such as lighting conditions can impact measurement reliability [100].

Current optical methods primarily assess surface or near-surface hydration levels, limiting their ability to capture deeper skin layers or the body's overall hydration status. Developing techniques with deeper penetration capabilities while ensuring safety and ease of use could enhance the effectiveness of optical hydration tracking.

Leveraging AI and machine learning algorithms to analyze the vast amounts of data collected by these sensors could further improve their accuracy, sensitivity, and ability to adapt to individual variations and environmental changes. Optical sensors are particularly well suited for applications requiring detailed surface hydration analysis, such as skincare and cosmetics, where precision is paramount.

Thermal and Microwave-Based Sensors

Other methods such as thermal and microwave-based sensors were identified, comprising 17% (11/63) of the publications. While these methods represent a smaller portion of the research landscape, they cater to specific applications and niches. Thermal sensors offer a unique perspective on skin hydration, probing the skin's thermal properties for insights. They have seen notable advancements, particularly in the development of sensors capable of measuring thermal properties beneath the skin's surface. In addition, some sensors measure skin temperature and thermal transport for evaluating skin hydration and blood flow dynamics. While promising, thermal sensors face challenges in terms of precision and sensitivity. Calibration is crucial, and factors such as ambient temperature can influence the results.

Microwave-based systems leverage electromagnetic waves' interaction with water molecules to assess skin moisture. These sensors are noninvasive and can penetrate deeper layers of the skin. However, such systems face several challenges that could limit their practical application. One of the main drawbacks is the difficulty in miniaturizing these systems to fit comfortably and unobtrusively into wearable devices. The complexity and size of the components required for microwave sensing can make it challenging to develop compact, wearable devices that are convenient for everyday use. Another significant limitation is the power consumption of microwave sensors. High power requirements can lead to shorter battery life in wearable devices, reducing their practicality for continuous, long-term monitoring In addition, microwave sensors can be susceptible to interference from external electromagnetic sources, which can affect the accuracy and reliability of hydration measurements. Future improvement opportunities in this field should look to achieve greater miniaturization and integration into wearable devices, addressing power consumption issues and making continuous hydration monitoring more feasible. Furthermore, developing advanced algorithms for signal processing and interference rejection could enhance the accuracy and reliability of

XSL•FO
RenderX

microwave-based hydration tracking even in environments with significant electromagnetic noise.

Thermal sensors are best suited for applications such as medical diagnostics, where detailed thermal assessments of hydration are needed. Microwave sensors are more appropriate for applications demanding deep tissue analysis, although practical implementation challenges persist.

Multimodal Sensors

A total of 13% (8/63) of the publications explored the use of multimodal sensors for hydration monitoring. This emerging trend suggests that researchers are recognizing the potential benefits of combining multiple sensor types to enhance accuracy and reliability. Multimodal sensors could address limitations associated with single-mode sensors, providing a more comprehensive understanding of hydration dynamics by addressing the limitations of individual sensors. As technology continues to advance, we can expect to see further exploration of this approach.

However, integrating multiple sensor technologies into a single wearable device introduces a set of challenges and limitations. One primary drawback is the complexity involved in sensor integration and the management of data from diverse sources. Ensuring that sensors operate harmoniously without interfering with each other's signals requires careful design and calibration. In addition, the increased complexity might lead to higher power consumption, larger device size, and potentially higher costs, which could limit the practicality and accessibility of multimodal hydration monitors for everyday use.

Another significant challenge is optimizing data fusion algorithms to combine data from multiple sensors meaningfully. This is crucial for interpreting complex data streams but requires algorithms that can accurately process relevant data from each sensor considering context and potential errors or noise.

Multimodal sensors are suitable for applications requiring comprehensive hydration analysis, such as health care and elite sports, where precise monitoring is crucial.

Commercial Products and Practical Implementations

Our analysis of commercial hydration-monitoring products revealed a growing market with a diverse range of offerings. Companies are effectively translating research into practical products for consumers, athletes, and health care practitioners. Products such as the Gx Sweat Patch and Nix Hydration Biosensor demonstrate the potential for providing actionable insights into hydration management. Optical-based techniques, which are cost-effective and practical for real-time monitoring, are gaining traction in the market due to their versatility and convenience.

While research provides the foundation for hydration-monitoring technologies, translating findings into commercially viable products presents several challenges. Controlled research settings often differ significantly from real-world conditions, where factors such as skin properties, sweat composition, and environmental variations can impact sensor accuracy. In addition, commercial products must balance accuracy with user convenience, power efficiency, and cost-effectiveness to ensure practicality in everyday use. Regulatory approval processes such as FDA certification for medical-grade devices further influence how research findings can be integrated into consumer products. The ability to refine prototype technologies into scalable, market-ready solutions depends on advancements in sensor miniaturization, robust validation across diverse populations, and improved data interpretation algorithms to enhance reliability under variable conditions

Ensuring the successful transfer of research advancements into commercial hydration-monitoring solutions requires overcoming key technical and practical challenges. Future efforts should focus on optimizing sensor performance in diverse real-world settings, improving long-term stability and accuracy, and addressing user adoption factors such as device comfort and accessibility. In addition, integrating hydration monitoring into multifunctional wearable systems—such as smartwatches and fitness trackers—can enhance usability and increase adoption. Continued refinement of these technologies will determine their effectiveness and scalability, ultimately shaping the future of hydration-monitoring solutions in both consumer and professional markets.

Limitations

This scoping review has some limitations. The inclusion of studies with varying validation methods makes direct comparisons challenging as methodological differences obstruct the establishment of standardized benchmarks for device accuracy and reliability. In addition, the lack of long-term evaluations limits insights into the real-world applicability of both academic research findings and commercial products. While the commercial product search provides practical insights into market trends and user-centered innovations, reliance on gray literature introduces potential bias. Many commercial products, including some mentioned in this review, lack rigorous scientific validation, raising concerns about the reliability and reproducibility of their claims [85,86,88]. Finally, the limited number of large-scale clinical trials addressing device effectiveness significantly restricts the ability to draw robust conclusions about the applicability of these technologies across diverse populations. This is one of the most significant criticisms of current hydration tracking research as small-scale studies and laboratory-controlled settings fail to capture variations in hydration status markers caused by differences in age, physiology, activity levels, environmental conditions, and health status. Addressing these gaps through standardized validation protocols, long-term real-world testing, and scaled clinical trials is essential to strengthen confidence in these technologies both for research advancement and consumer trust in emerging products.

Conclusions

This scoping review provides a comprehensive analysis of recent advancements in wearable hydration-monitoring technologies, offering insights into their capabilities, limitations, and research trends. This review highlights the growing interest in noninvasive, portable hydration-monitoring systems using physiological biomarkers, with a predominant focus on electrical-based sensors such as BIA due to their ease of use and accuracy. Optical-based sensors, while less common, offer

XSL•FO
RenderX

molecular-level insights and are gaining traction with advancements in photonic technology. Thermal and microwave-based sensors cater to specific applications but face challenges regarding precision and miniaturization. The emergence of multimodal sensors suggests a growing trend toward combining technologies for enhanced accuracy and reliability. Commercially available hydration-monitoring products demonstrate a growing market with diverse offerings, primarily dominated by optical-based solutions due to their cost-effectiveness and the diverse number of biomarkers that they give access to.

Future research should focus on enhancing sensor accuracy, miniaturization, and wearability; standardizing measurement protocols; and refining data interpretation algorithms. Multimodal systems and AI-driven analysis show promise for personalized hydration management, with broad applications in health care and beyond. Overall, wearable hydration-monitoring devices offer significant potential for real-time hydration assessment. Overcoming challenges such as reliability, accuracy, and applicability across diverse populations will be crucial for their widespread adoption and impact in health care and performance optimization.

Acknowledgments

This study was funded by the National Science Foundation Division of Computer and Network Systems (2230180).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures illustrating key experimental setups and device schematics referenced in this review.
[DOCX File , 1438 KB - mhealth_v13i1e60569_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 86 KB - mhealth_v13i1e60569_app2.docx]

References

1. Popkin BM, D'Anci KE, Rosenberg IH. Water, hydration, and health. Nutr Rev 2010 Aug;68(8):439-458 [FREE Full text] [doi: 10.1111/j.1753-4887.2010.00304.x] [Medline: 20646222]
2. Garrett DC, Rae N, Fletcher JR, Zarnke S, Thorson S, Hogan DB, et al. Engineering approaches to assessing hydration status. IEEE Rev Biomed Eng 2017 Nov 20;11:233-248. [doi: 10.1109/rbme.2017.2776041]
3. Voroneanu L, Gavrilovici C, Covic A. Overhydration, underhydration, and total body sodium: a tricky "ménage a trois" in dialysis patients. Semin Dial 2018 Jan;31(1):21-25. [doi: 10.1111/sdi.12649] [Medline: 28967233]
4. Gray M, Birkenfeld JS, Butterworth I. Noninvasive monitoring to detect dehydration: are we there yet? Annu Rev Biomed Eng 2023 Jun 08;25:23-49 [FREE Full text] [doi: 10.1146/annurev-bioeng-062117-121028] [Medline: 36854261]
5. Gidado IM, Qassem M, Triantis IF, Kyriacou PA. Review of advances in the measurement of skin hydration based on sensing of optical and electrical tissue properties. Sensors (Basel) 2022 Sep 21;22(19):7151 [FREE Full text] [doi: 10.3390/s22197151] [Medline: 36236250]
6. Chong KP, Guo JZ, Deng X, Woo BK. Consumer perceptions of wearable technology devices: retrospective review and analysis. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e17544 [FREE Full text] [doi: 10.2196/17544] [Medline: 32310148]
7. Pobiruchin M, Suleder J, Zowalla R, Wiesner M. Accuracy and adoption of wearable technology used by active citizens: a marathon event field study. JMIR Mhealth Uhealth 2017 Feb 28;5(2):e24 [FREE Full text] [doi: 10.2196/mhealth.6395] [Medline: 28246070]
8. Chang CC. Exploring the usage intentions of wearable medical devices: a demonstration study. Interact J Med Res 2020 Sep 18;9(3):e19776 [FREE Full text] [doi: 10.2196/19776] [Medline: 32945778]
9. Bube B, Zanón BB, Lara Palma AM, Klocke H. Wearable devices in diving: scoping review. JMIR Mhealth Uhealth 2022 Sep 06;10(9):e35727 [FREE Full text] [doi: 10.2196/35727] [Medline: 36066926]
10. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
11. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
12. Masoumian Hosseini M, Masoumian Hosseini ST, Qayumi K, Hosseinzadeh S, Sajadi Tabar SS. Smartwatches in healthcare medicine: assistance and monitoring; a scoping review. BMC Med Inform Decis Mak 2023 Nov 03;23(1):248 [FREE Full text] [doi: 10.1186/s12911-023-02350-w] [Medline: 37924029]

XSL•FO
RenderX

13. Seshadri DR, Li RT, Voos JE, Rowbottom JR, Alfes CM, Zorman CA, et al. Wearable sensors for monitoring the physiological and biochemical profile of the athlete. NPJ Digit Med 2019 Jul 22;2:72 [FREE Full text] [doi: 10.1038/s41746-019-0150-9] [Medline: 31341957]
14. Ahsan M, Teay SH, Sayem AS, Albarbar A. Smart clothing framework for health monitoring applications. Signals 2022 Mar 02;3(1):113-145. [doi: 10.3390/signals3010009]
15. Lu YP, Huang JW, Lee IN, Weng RC, Lin MY, Yang JT, et al. A portable system to monitor saliva conductivity for dehydration diagnosis and kidney healthcare. Sci Rep 2019 Oct 14;9(1):14771 [FREE Full text] [doi: 10.1038/s41598-019-51463-8] [Medline: 31611585]
16. Rizwan A, Abu Ali N, Zoha A, Ozturk M, Alomainy A, Imran MA, et al. Non-invasive hydration level estimation in human body using galvanic skin response. IEEE Sensors J 2020 Jan 10;20(9):4891-4900. [doi: 10.1109/JSEN.2020.2965892]
17. AlDisi R, Bader Q, Bermak A. Hydration assessment using the bio-impedance analysis method. Sensors (Basel) 2022 Aug 24;22(17):6350 [FREE Full text] [doi: 10.3390/s22176350] [Medline: 36080808]
18. Songkakul T, Wu S, Ahmmed P, Reynolds WD, Zhu Y, Bozkurt A. Wearable bioimpedance hydration monitoring system using conformable AgNW electrodes. In: Proceedings of the IEEE Sensors. 2021 Presented at: SENSORS 2021; October 31-November 3, 2021; Sydney, Australia. [doi: 10.1109/sensors47087.2021.9639469]
19. Thomas M, Veeralingam S, Badhulika S. MoSe2/PVA-based wearable multi-functional platform for pulse rate monitoring, skin hydration sensor, and human gesture recognition utilizing electrophysiological signals. J Appl Phys 2022 Dec 14;132(22):224303. [doi: 10.1063/5.0123238]
20. Tonello S, Zacchini A, Galli A, Narduzzi C, Golparvar A, Meimandi A. In-Vivo validation of smart device for on body hydration monitoring. In: Proceedings of the IEEE International Workshop on Metrology for Industry 4.0 & IoT. 2023 Presented at: MetroInd4.0&IoT 2023; June 6-8, 2023; Brescia, Italy. [doi: 10.1109/metroind4.0iot57462.2023.10180167]
21. Valentin H, Olivier F, Patrick P, Nicolas I, Lionel R, Gaëlle L. Wearable bio-impedance analysis for hydration monitoring in medical applications. In: Proceedings of the Symposium on Design, Test, Integration & Packaging of MEMS and MOEMS. 2021 Presented at: DTIP 2021; August 25-27, 2021; Paris, France. [doi: 10.1109/dtip54218.2021.9568666]
22. Bhide A, Ganguly A, Parupudi T, Ramasamy M, Muthukumar S, Prasad S. Next-generation continuous metabolite sensing toward emerging sensor needs. ACS Omega 2021 Feb 22;6(9):6031-6040 [FREE Full text] [doi: 10.1021/acsomega.0c06209] [Medline: 33718694]
23. Kamran F, Le VC, Frischknecht A, Wiens J, Sienko KH. Noninvasive estimation of hydration status in athletes using wearable sensors and a data-driven approach based on orthostatic changes. Sensors (Basel) 2021 Jun 29;21(13):4469 [FREE Full text] [doi: 10.3390/s21134469] [Medline: 34210068]
24. Hao Y, Foster R. Wireless body sensor networks for health-monitoring applications. Physiol Meas 2008 Nov;29(11):R27-R56. [doi: 10.1088/0967-3334/29/11/R01] [Medline: 18843167]
25. Hooper L, Bunn DK, Downing A, Jimoh FO, Groves J, Free C, et al. Which frail older people are dehydrated? The UK DRIE study. J Gerontol A Biol Sci Med Sci 2016 Oct;71(10):1341-1347 [FREE Full text] [doi: 10.1093/gerona/glv205] [Medline: 26553658]
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
27. Matsukawa R, Miyamoto A, Yokota T, Someya T. Skin impedance measurements with nanomesh electrodes for monitoring skin hydration. Adv Healthc Mater 2020 Nov 20;9(22):e2001322. [doi: 10.1002/adhm.202001322] [Medline: 33084247]
28. Van Beers R, Jacobs M, Borgoo A, Hornikx M, Janssens S, van der Tempel W, et al. Accurate measurement of SpO2 and dermal skin hydration using a wearable miniaturized spectrometer. In: Proceedings of the Volume 11956, Biophotonics in Exercise Science, Sports Medicine, Health Monitoring Technologies, and Wearables III. 2022 Presented at: SPIE BiOS 2022; January 22-23, 2022; San Francisco, CA. [doi: 10.1117/12.2607346]
29. Bandodkar AJ, Gutruf P, Choi J, Lee K, Sekine Y, Reeder JT, et al. Battery-free, skin-interfaced microfluidic/electronic systems for simultaneous electrochemical, colorimetric, and volumetric analysis of sweat. Sci Adv 2019 Jan 04;5(1):eaav3294 [FREE Full text] [doi: 10.1126/sciadv.aav3294] [Medline: 30746477]
30. Malnati C, Fehr D, Spano F, Bonmarin M. Modeling stratum corneum swelling for the optimization of electrode-based skin hydration sensors. Sensors (Basel) 2021 Jun 09;21(12):3986 [FREE Full text] [doi: 10.3390/s21123986] [Medline: 34207803]
31. Yao S, Myers A, Malhotra A, Lin F, Bozkurt A, Muth JF, et al. A wearable hydration sensor with conformal nanowire electrodes. Adv Healthc Mater 2017 Mar;6(6). [doi: 10.1002/adhm.201601159] [Medline: 28128888]
32. MoistureMeterD compact – local measurements of tissue water. Delfin Technologies. URL: https://delfintech.com/products/moisturemeterd-compact/ [accessed 2023-12-11]
33. Flament F, Galliano A, Abric A, Matoschitz CM, Bammer M, Kampus M, et al. Skin moisture assessment using hydration sensor patches coupled with smartphones via near field communication (NFC). A pilot study with the first generation of patches that allow self-recordings of skin hydration. Skin Res Technol 2021 Sep;27(5):959-965. [doi: 10.1111/srt.13049] [Medline: 33998713]

XSL•FO
RenderX

34. Balaban J, Blecha T. Textile sensor for skin hydration measurement. In: Proceedings of the 45th International Spring Seminar on Electronics Technology. 2022 Presented at: ISSE 2022; May 11-15, 2022; Vienna, Austria. [doi: 10.1109/isse54558.2022.9812794]
35. Clarys P, Clijsen R, Taeymans J, Barel AO. Hydration measurements of the stratum corneum: comparison between the capacitance method (digital version of the Corneometer CM 825®) and the impedance method (Skicon-200EX®). Skin Res Technol 2012 Aug;18(3):316-323. [doi: 10.1111/j.1600-0846.2011.00573.x] [Medline: 22092664]
36. Madhvapathy SR, Wang H, Kong J, Zhang M, Lee JY, Park JB, et al. Reliable, low-cost, fully integrated hydration sensors for monitoring and diagnosis of inflammatory skin diseases in any environment. Sci Adv 2020 Dec 04;6(49):eabd7146 [FREE Full text] [doi: 10.1126/sciadv.abd7146] [Medline: 33277260]
37. Wang S, Liu M, Yang X, Lu Q, Xiong Z, Li L, et al. An unconventional vertical fluidic-controlled wearable platform for synchronously detecting sweat rate and electrolyte concentration. Biosens Bioelectron 2022 Aug 15;210:114351. [doi: 10.1016/j.bios.2022.114351] [Medline: 35569269]
38. Liaqat S, Dashtipour K, Rizwan A, Usman M, Shah SA, Arshad K, et al. Personalized wearable electrodermal sensing-based human skin hydration level detection for sports, health and wellbeing. Sci Rep 2022 Mar 08;12(1):3715 [FREE Full text] [doi: 10.1038/s41598-022-07754-8] [Medline: 35260675]
39. Morin M, Ruzgas T, Svedenhag P, Anderson CD, Ollmar S, Engblom J, et al. Skin hydration dynamics investigated by electrical impedance techniques in vivo and in vitro. Sci Rep 2020 Oct 14;10(1):17218 [FREE Full text] [doi: 10.1038/s41598-020-73684-y] [Medline: 33057021]
40. Chua MC. Design and development of an integrated health (i-health) monitoring watch. In: Proceedings of the IEEE/SICE International Symposium on System Integration. 2016 Presented at: SII 2016; December 13-15, 2016; Sapporo, Japan. [doi: 10.1109/sii.2016.7844014]
41. Agcayazi T, Hong GJ, Maione B, Woodard E. Wearable infant hydration monitor. In: Proceedings of the IEEE Virtual Conference on Applications of Commercial Sensors. 2016 Presented at: VCACS 2016; June 15-January 15, 2017; Online. [doi: 10.1109/vcacs.2016.7888786]
42. Yang GZ, Rosa BM. A wearable and battery-less device for assessing skin hydration level under direct sunlight exposure with ultraviolet index calculation. In: Proceedings of the 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks. 2018 Presented at: BSN 2018; March 4-7, 2018; Las Vegas, NV. [doi: 10.1109/bsn.2018.8329693]
43. Sunny AI, Rahman M, Koutsoupidou M, Cano-Garcia H, Thanou M, Rafique W. Feasibility experiments to detect skin hydration using a bio-impedance sensor. In: Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2019 Presented at: EMBC 2019; July 23-27, 2019; Berlin, Germany. [doi: 10.1109/embc.2019.8857470]
44. Leonov V, Lee S, Londergan A, Martin RA, De Raedt W, Van Hoof C. Bioimpedance method for human body hydration assessment. In: Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2019 Presented at: EMBC 2019; July 23-27, 2019; Berlin, Germany. [doi: 10.1109/embc.2019.8857207]
45. Veeralingam S, Khandelwal S, Badhulika S. AI/ML-enabled 2-D - RuS2 nanomaterial-based multifunctional, low cost, wearable sensor platform for non-invasive point of care diagnostics. IEEE Sensors J 2020 Aug 1;20(15):8437-8444. [doi: 10.1109/jsen.2020.2984807]
46. Jang M, Kim HD, Koo HJ, So JH. Textile-based wearable sensor for skin hydration monitoring. Sensors (Basel) 2022 Sep 15;22(18):6985 [FREE Full text] [doi: 10.3390/s22186985] [Medline: 36146334]
47. Westermann TV, Viana VR, Berto Junior C, Detoni da Silva CB, Carvalho EL, Pupe CG. Measurement of skin hydration with a portable device (SkinUp Beauty Device) and comparison with the Corneometer. Skin Res Technol 2020 Jul 19;26(4):571-576. [doi: 10.1111/srt.12833] [Medline: 31957168]
48. Culver DJ, Colon AB, Washington DR, Appleton MG, Strang A, Alizadeh A. Field test of wearable sensors for hydration monitoring. In: Proceedings of the 2019 Systems and Information Engineering Design Symposium. 2019 Presented at: SIEDS 2019; April 26, 2019; Charlottesville, VA. [doi: 10.1109/sieds.2019.8735637]
49. Alizadeh A, Burns A, Lenigk R, Gettings R, Ashe J, Porter A, et al. A wearable patch for continuous monitoring of sweat electrolytes during exertion. Lab Chip 2018 Aug 21;18(17):2632-2641. [doi: 10.1039/c8lc00510a] [Medline: 30063233]
50. Lafaye C, Rovira M, Demuru S, Wang S, Kim J, Paul Kunnel B. Real-time smart multisensing wearable platform for monitoring sweat biomarkers during exercise. In: Proceedings of the 2022 IEEE Biomedical Circuits and Systems Conference. 2022 Presented at: BioCAS 2022; October 13-15, 2022; Taipei, Taiwan. [doi: 10.1109/biocas54905.2022.9948565]
51. Reeder JT, Choi J, Xue Y, Gutruf P, Hanson J, Liu M, et al. Waterproof, electronics-enabled, epidermal microfluidic devices for sweat collection, biomarker analysis, and thermography in aquatic settings. Sci Adv 2019 Jan 25;5(1):eaau6356 [FREE Full text] [doi: 10.1126/sciadv.aau6356] [Medline: 30746456]
52. Parrilla M, Guinovart T, Ferré J, Blondeau P, Andrade FJ. A wearable paper-based sweat sensor for human perspiration monitoring. Adv Healthc Mater 2019 Aug 10;8(16):e1900342. [doi: 10.1002/adhm.201900342] [Medline: 31293084]
53. Yang M, Sun N, Lai X, Li Y, Zhao X, Wu J, et al. Screen-printed wearable sweat sensor for cost-effective assessment of human hydration status through potassium and sodium ion detection. Micromachines (Basel) 2023 Jul 26;14(8):1497 [FREE Full text] [doi: 10.3390/mi14081497] [Medline: 37630034]

XSL•FO
RenderX

54. Ahmad Tarar A, Mohammad U, K Srivastava S. Wearable skin sensors and their challenges: a review of transdermal, optical, and mechanical sensors. Biosensors (Basel) 2020 May 28;10(6):56 [FREE Full text] [doi: 10.3390/bios10060056] [Medline: 32481598]
55. Qassem M, Kyriacou PA. In vivo optical investigation of short term skin water contact and moisturizer application using NIR spectroscopy. In: Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2013 Presented at: EMBC 2013; July 3-7, 2013; Osaka, Japan. [doi: 10.1109/embc.2013.6610020]
56. E C, Scheffer C, Dellimore K, Kieser E, Smith J. Development of a diagnostic dehydration screening sensor based on infrared spectrometry. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2015 Presented at: EMBC 2015; August 25-29, 2015; Milan, Italy. [doi: 10.1109/embc.2015.7318599]
57. Visser C, Kieser E, Dellimore K, van den Heever D, Smith J. Investigation of the feasibility of non-invasive optical sensors for the quantitative assessment of dehydration. Med Eng Phys 2017 Oct;48:181-187. [doi: 10.1016/j.medengphy.2017.06.036] [Medline: 28734875]
58. Mamouei M, Chatterjee S, Razban M, Qassem M, Kyriacou PA. Design and analysis of a continuous and non-invasive multi-wavelength optical sensor for measurement of dermal water content. Sensors (Basel) 2021 Mar 19;21(6):2162 [FREE Full text] [doi: 10.3390/s21062162] [Medline: 33808821]
59. Benavides NR, Rutkey HE, DiDomenico CM, Wong T, Martel-Foley J, Yu CH. A wearable mobile-app controlled device for continuous monitoring of UV exposure and hydration levels. In: Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference. 2021 Presented at: CCWC 2021; January 27-30, 2021; Virtual Event. [doi: 10.1109/ccwc51732.2021.9376109]
60. Volkova E, Perchik A, Pavlov K, Nikolaev E, Ayuev A, Park J, et al. Multispectral sensor fusion in SmartWatch for in situ continuous monitoring of human skin hydration and body sweat loss. Sci Rep 2023 Aug 17;13(1):13371 [FREE Full text] [doi: 10.1038/s41598-023-40339-7] [Medline: 37591885]
61. Curto VF, Coyle S, Byrne R, Diamond D, Benito-Lopez F. Real-time sweat analysis: concept and development of an autonomous wearable micro-fluidic platform. Procedia Eng 2011;25:1561-1564. [doi: 10.1016/j.proeng.2011.12.386]
62. Ozana N, Arbel N, Beiderman Y, Mico V, Sanz M, Garcia J, et al. Improved noncontact optical sensor for detection of glucose concentration and indication of dehydration level. Biomed Opt Express 2014 May 22;5(6):1926-1940. [doi: 10.1364/boe.5.001926]
63. Perkov SA, Gorin DA, Esenaliev RO. Optoacoustic monitoring of water content in tissue phantoms and human skin. J Biophotonics 2021 Mar 24;14(3):e202000363. [doi: 10.1002/jbio.202000363] [Medline: 33205631]
64. Sandys V, Edwards C, McAleese P, O'Hare E, O'Seaghdha C. Protocol of a pilot-scale, single-arm, observational study to assess the utility and acceptability of a wearable hydration monitor in haemodialysis patients. Pilot Feasibility Stud 2022 Jan 24;8(1):17 [FREE Full text] [doi: 10.1186/s40814-022-00976-7] [Medline: 35073985]
65. Bohman AB, Nguyen JQ, Parthasarathy S, Arnold MA. Temperature and hydration monitoring in tissue-simulating phantoms using novel silicon-photonics-based spectrophotometer. In: Proceedings of the Volume 12375, Biophotonics in Exercise Science, Sports Medicine, Health Monitoring Technologies, and Wearables IV. 2023 Presented at: SPIE BiOS 2023; January 28-February 2, 2023; San Francisco, CA. [doi: 10.1117/12.2649262]
66. Madhvapathy SR, Ma Y, Patel M, Krishnan S, Wei C, Li Y, et al. Epidermal thermal depth sensors: epidermal electronic systems for measuring the thermal properties of human skin at depths of up to several millimeters. Adv Funct Mater 2018 Aug 20;28(34):1870242. [doi: 10.1002/adfm.201870242]
67. Krishnan SR, Su CJ, Xie Z, Patel M, Madhvapathy SR, Xu Y, et al. Wireless, battery-free epidermal electronics for continuous, quantitative, multimodal thermal characterization of skin. Small 2018 Nov 07;14(47):e1803192. [doi: 10.1002/smll.201803192] [Medline: 30369049]
68. Kwon K, Wang H, Lim J, Chun KS, Jang H, Yoo I, et al. Wireless, soft electronics for rapid, multisensor measurements of hydration levels in healthy and diseased skin. Proc Natl Acad Sci U S A 2021 Feb 02;118(5):e2020398118 [FREE Full text] [doi: 10.1073/pnas.2020398118] [Medline: 33468630]
69. Shin J, Wang H, Kwon K, Ostojich D, Christiansen Z, Berkovich J, et al. Wireless, soft sensors of skin hydration with designs optimized for rapid, accurate diagnostics of dermatological health. Adv Healthc Mater 2023 Feb 18;12(4):e2202021. [doi: 10.1002/adhm.202202021] [Medline: 36337006]
70. Butterworth I, Serallés J, Mendoza CS, Giancardo L, Daniel L. A wearable physiological hydration monitoring wristband through multi-path non-contact dielectric spectroscopy in the microwave range. In: Proceedings of the IEEE MTT-S 2015 International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications. 2015 Presented at: IMWS-BIO 2015; September 21-23, 2015; Taipei, Taiwan. [doi: 10.1109/imws-bio.2015.7303776]
71. Wang J, Zilic Z, Shu Y. Evaluation of an RF wearable device for non-invasive real-time hydration monitoring. In: Proceedings of the 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks. 2017 Presented at: BSN 2017; May 09-12, 2017; Eindhoven, The Netherlands. [doi: 10.1109/bsn.2017.7936015]
72. Schiavoni R, Monti G, Piuzzi E, Tarricone L, Tedesco A, De Benedetto E, et al. Feasibility of a wearable reflectometric system for sensing skin hydration. Sensors (Basel) 2020 May 16;20(10):2833 [FREE Full text] [doi: 10.3390/s20102833] [Medline: 32429375]

XSL•FO
RenderX

73. Schiavoni R, Monti G, Tedesco A, Tarricone L, Piuzzi E, de Benedetto E. Microwave wearable system for sensing skin hydration. In: Proceedings of the 2021 IEEE International Instrumentation and Measurement Technology Conference. 2021 Presented at: I2MTC 2021; May 17-20, 2021; Glasgow, UK. [doi: 10.1109/i2mtc50364.2021.9460018]
74. Besler BC, Fear EC. Microwave hydration monitoring: system assessment using fasting volunteers. Sensors (Basel) 2021 Oct 20;21(21):6949 [FREE Full text] [doi: 10.3390/s21216949] [Medline: 34770259]
75. Cataldo A, De Benedetto E, Schiavoni R, Monti G, Tedesco A, Masciullo A, et al. Portable microwave reflectometry system for skin sensing. IEEE Trans Instrum Meas 2022;71:1-8. [doi: 10.1109/tim.2022.3154804]
76. Bing S, Chawang K, Chiao JC. A radio-frequency planar resonant loop for noninvasive monitoring of water content. In: Proceedings of the IEEE Sensors. 2022 Presented at: SENSORS 2022; October 30-November 02, 2022; Dallas, TX. [doi: 10.1109/sensors52175.2022.9967154]
77. Krishnan S, Shi Y, Webb RC, Ma Y, Bastien P, Crawford KE, et al. Multimodal epidermal devices for hydration monitoring. Microsyst Nanoeng 2017 Jun 05;3(1):17014 [FREE Full text] [doi: 10.1038/micronano.2017.14] [Medline: 31057861]
78. Salvo P, Pingitore A, Barbini A, Di Francesco F. A wearable sweat rate sensor to monitor the athletes' performance during training. Sci Sports 2018 Apr;33(2):e51-e58. [doi: 10.1016/j.scispo.2017.03.009]
79. Lapadula V, Sabatini A, Zompanti A, Buscaglione S, Lanaro D, Merone M. A body hydration analysis system to improve running performance. In: Proceedings of the IEEE International Workshop on Metrology for Industry 4.0 & IoT. 2020 Presented at: MetroInd4.0IoT 2020; June 03-05, 2020; Roma, Italy. [doi: 10.1109/metroind4.0iot48571.2020.9138240]
80. Sabry F, Eltaras T, Labda W, Hamza F, Alzoubi K, Malluhi Q. Towards on-device dehydration monitoring using machine learning from wearable device's data. Sensors (Basel) 2022 Feb 28;22(5):1887 [FREE Full text] [doi: 10.3390/s22051887] [Medline: 35271034]
81. Rodin D, Shapiro Y, Pinhasov A, Kreinin A, Kirby M. An accurate wearable hydration sensor: real-world evaluation of practical use. PLoS One 2022 Aug 24;17(8):e0272646 [FREE Full text] [doi: 10.1371/journal.pone.0272646] [Medline: 36001536]
82. Wang S, Lafaye C, Saubade M, Besson C, Margarit-Taule JM, Gremeaux V, et al. Predicting hydration status using machine learning models from physiological and sweat biomarkers during endurance exercise: a single case study. IEEE J Biomed Health Inform 2022 Sep;26(9):4725-4732. [doi: 10.1109/jbhi.2022.3186150]
83. Rockley press and publications. Rockley. URL: https://rockleyphotonics.com/ [accessed 2023-07-14]
84. Sixty homepage. Sixty. URL: https://sixty.ie/ [accessed 2023-07-14]
85. LVL – the first wearable hydration monitor. Kickstarter. URL: https://www.kickstarter.com/projects/lactate-threshold/lvl-the-first-wearable-hydration-monitor [accessed 2023-07-03]
86. hDrop - electrolytes, hydration, and temperature tracker. Kickstarter. URL: https://www.kickstarter.com/projects/hdrop/hdrop-real-time-hydration-wearable-device-monitor [accessed 2023-07-06]
87. Hydrostasis Geca Watch. Hydrostasis. URL: https://www.hydrostasis.com/#intro [accessed 2023-07-06]
88. One of the next Health Features that may be coming to Apple Watch relates to Tracking Hydration Levels of a user. Patently Apple. 2021. URL: https://www.patentlyapple.com/2021/08/one-of-the-next-health-features-that-may-be-coming-to-apple-watch-relates-to-tracking-hydration-levels-of-a-user.html [accessed 2023-05-06]
89. Ghaffari R, Aranyosi AJ, Lee SP, Model JB, Baker LB. The Gx Sweat Patch for personalized hydration management. Nat Rev Bioeng 2023 Jan 19;1(1):5-7. [doi: 10.1038/s44222-022-00005-5]
90. Nix Biosensors homepage. Nix Biosensors. URL: https://nixbiosensors.com/ [accessed 2023-09-03]
91. The quest for hydration management: simplifying the complexities of measuring and reporting hydration. Rockley® Photonics. 2022 May. URL: https://rockleyphotonics.com/wp-content/uploads/2022/05/Rockley-Photonics-The-Quest-for-Hydration-Management.pdf [accessed 2023-11-03]
92. Mills MF. Sixty's smart band knows you're not drinking enough water. Wearable. 2019 Aug 27. URL: https://www.wareable.com/health-and-wellbeing/sixty-hydration-tracker-wearable-7523 [accessed 2023-09-03]
93. The science of sweat and hydration. hDrop Technologies. URL: https://hdroptech.com/hydration-sweat-testing-science/?srsltid=AfmBOoqvwjsf6AdKiaGdv6W2TZeNXuSnPOBAkZ0s1SINXx63EKnZ36zT [accessed 2023-07-03]
94. Hoogenhout M, Ciliberto M, Shirman M, Chen D. Predicting hydration changes in healthy adults with a wrist-worn wearable. Hydrostasis. 2022. URL: https://cdn.prod.website-files.com/6126438043b0680b8f07f6b6/624b76f43c9b32ba0640c5a4_Consumer-Pilot-blurb.pdf [accessed 2023-07-03]
95. AURA Devices. AURA. URL: https://auradevices.io/?srsltid=AfmBOopYk8821QyOa1b5BHZ-SktC2WdGqPaVpBUPrx5dn80Jz3NoK11q [accessed 2023-08-03]
96. Gx Sweat Patch homepage. Gx Sweat Patch. URL: https://www.epicorebiosystems.com/our-solutions/gx-sweat-patch [accessed 2024-09-06]
97. Abe Y, Nishizawa M. Electrical aspects of skin as a pathway to engineering skin devices. APL Bioeng 2021 Dec;5(4):041509 [FREE Full text] [doi: 10.1063/5.0064529] [Medline: 34849444]
98. Cravello B, Ferri A. Relationships between skin properties and environmental parameters. Skin Res Technol 2008 May 27;14(2):180-186. [doi: 10.1111/j.1600-0846.2007.00275.x] [Medline: 18412560]

XSL•FO
RenderX

99. Sohail M, De Marco R. ELECTRODES | Ion-selective electrodes. In: Reedijk J, editor. Elsevier Reference Collection in Chemistry, Molecular Sciences and Chemical Engineering. Amsterdam, The Netherlands: Elsevier; 2013.
100. Vavrinsky E, Esfahani NE, Hausner M, Kuzma A, Rezo V, Donoval M, et al. The current state of optical sensors in medical wearables. Biosensors (Basel) 2022 Apr 06;12(4):217 [FREE Full text] [doi: 10.3390/bios12040217] [Medline: 35448277]

Abbreviations

AI: artificial intelligence
BIA: bioelectrical impedance analysis
BLE: Bluetooth Low Energy
BWL: body water loss
FDA: Food and Drug Administration
GSR: galvanic skin response
HR: heart rate
MCU: microcontroller unit
NFC: near-field communication
NIRS: near-infrared spectroscopy
PCB: printed circuit board
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SHS: skin hydration sensor
smLED: surface-mount light-emitting diode
TDR: time-domain reflectometry

Edited by L Buis; submitted 15.05.24; peer-reviewed by E Bai, YZ Wu; comments to author 17.07.24; revised version received 10.09.24; accepted 12.03.25; published 13.06.25.

Please cite as:
Belabbaci NA, Anaadumba R, Alam MAU
Recent Advancements in Wearable Hydration-Monitoring Technologies: Scoping Review of Sensors, Trends, and Future Directions
JMIR Mhealth Uhealth 2025;13:e60569
URL: https://mhealth.jmir.org/2025/1/e60569
doi:10.2196/60569
PMID:

©Nazim A. Belabbaci, Raphael Anaadumba, Mohammad Arif Ul Alam. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perspectives and Experiences of Family Caregivers Using Supportive Mobile Apps in Dementia Care: Meta-Synthesis of Qualitative Research

Haifei Shen[1], BSN; Yi Han[1], BSN; Wen Shi[1], BSN; Jiangxuan Yu[1], MSN; Xueqi Shan[1], MSN; Hongyao Wang[1], BSN; Junjie Wang[1], MSN

School of Nursing, Zhejiang Chinese Medical University, Hangzhou, China

Corresponding Author:
Junjie Wang, MSN
School of Nursing
Zhejiang Chinese Medical University
548 Binwen Road, Binjiang District
Hangzhou, 310053
China
Phone: 86 0571 86633390
Email: wjjie2000@163.com

Abstract

Background: Supportive mobile apps are effective tools for family caregivers of persons with dementia to obtain online information and psychological support. Nevertheless, details about the experiences of family caregivers of persons with dementia using mobile apps are limited.

Objective: This study aimed to synthesize the perspectives and experiences of family caregivers of persons with dementia regarding supportive mobile apps.

Methods: We conducted a synthesis of qualitative research and searched 7 English-language databases and 4 Chinese-language databases. We included qualitative studies (peer-reviewed studies and gray literature) written in English and Chinese on the perspectives and experiences of family caregivers of persons with dementia regarding supportive mobile apps published from database establishment to March 2025. Two researchers independently screened the literature and used the JBI Critical Appraisal Checklist for Qualitative Research to conduct quality assessments on the final included studies. Themes were integrated using the 3-stage thematic synthesis approach by Thomas and Harden.

Results: A preliminary search yielded 4772 studies, of which 12 (0.25%) met the criteria. The included studies were from 7 different countries or regions, of which the only low- or middle-income country was Brazil. The studies involved a total of 232 family caregivers, most of whom were older adults and female. The integration of extracted content resulted in 4 themes: dynamic changes in value perception—complex attitudes toward mobile app adoption; from tools to partners—a technology-empowered multidimensional support system for family caregivers; external and internal barriers—challenges in family caregivers' use of mobile apps; and person-centered design—future directions for improving mobile apps.

Conclusions: This study found that family caregivers' attitudes toward using supportive mobile apps are influenced by their perceived value of mobile apps and their caregiving burden. In addition, such supportive mobile apps serve as valuable tools for family caregivers to enhance their caregiving abilities and efficiency, alleviate the burden of care, improve negative emotions, foster social connections, and promote self-care. Future mobile app design needs to address obstacles such as design flaws, family caregivers' lack of technological literacy, time constraints, concerns about privacy breaches, and other device-related issues, with particular attention to the ease of use of mobile apps. Meanwhile, developers need to commit to designing personalized and multifunctional mobile apps as well as promote online collaboration among members of the care network. Overall, our study offers an important reference for developing person-centered supportive mobile apps for family caregivers of persons with dementia.

Trial Registration: PROSPERO CRD42024510905; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024510905

(JMIR Mhealth Uhealth 2025;13:e65983) doi:10.2196/65983

XSL•FO
RenderX

KEYWORDS

dementia; caregivers; mobile app; qualitative research; meta-synthesis; artificial intelligence; AI

Introduction

Background

The phenomenon of an aging global population has significantly impacted the prevalence of dementia [1]. According to the World Alzheimer Report (2023), the global population of persons with dementia is anticipated to increase from 55 million in 2019 to 139 million by 2050 [2]. Dementia is a chronic, long-term, progressive disease [3]. As dementia progresses, persons with dementia gradually become incapable of caring for themselves and develop cognitive and behavioral abnormalities, requiring substantial assistance from caregivers for many daily activities in the later stages [3]. The World Health Organization declared dementia a public health priority and committed to the care of persons with dementia and the assistance of their carers in 2017 [4]. Persons with dementia are mainly taken care of by family members, which includes spouses, children, other relatives, and friends [5]. Family caregivers of persons with dementia face higher physical and psychological challenges than typical caregivers because of their long-term exposure to various stresses [6], which indirectly reduces the quality of care [7] and can lead to social issues such as intrafamily conflicts, patient abuse, and even caregiver suicide [8]. At the same time, family caregivers lack knowledge and skills related to the progression and management of dementia [9], leading to low confidence in caregiving. Inappropriate care behavior could even trigger behavioral and psychological symptoms of dementia in persons with dementia [10]. Therefore, family caregivers urgently need knowledge, skills, and emotional support.

Acceptance and commitment therapy, cognitive therapy, psychoeducation, and other interventions could effectively alleviate family caregivers' stress and provide support for them [11]. However, most existing interventions are face-to-face, which can be labor intensive and time-consuming and only help a restricted number of people. Furthermore, due to limited health care and caregiving resources [12], there is an urgent need for innovative alternative approaches and methods to promote convenient and efficient dementia care and self-management for family caregivers.

Supportive measures based on mobile health (mHealth) may be an effective alternative. The World Health Organization defined mHealth as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, PDAs, and other wireless devices" [13]. Mobile apps are one of the main forms of mHealth and generally refer to app software downloaded on smart mobile terminals such as smartphones [14]. These mobile apps, increasingly used to support family caregivers of persons with dementia, accounted for 18 out of 26 (69%) chronic illness caregiver apps in a review, highlighting their potential [15]. In our study, supportive mobile apps can be defined as a type of mobile app designed for family caregivers of persons with dementia whose functions include knowledge transfer, assistance with caregiving tasks, and provision of social and psychological support, aiming to help reduce caregiving stress and negative emotions in family caregivers as well as improve the quality of caregiving [16-18]. Compared to traditional interventions, mobile apps overcome time and space limitations, enabling around-the-clock access to resources, which is especially beneficial for remote users seeking services [19]. In contrast to other web-based applications, they allow for offline use, access to built-in components, or connection to external devices [15]. These mobile apps are critical for family caregivers, especially when dealing with the most challenging behavioral and psychological symptoms of dementia of persons with dementia. They not only provide care knowledge and social service information, reducing information retrieval burdens [18], but also include monitoring functions to help assess the status of persons with dementia. In addition, around-the-clock access and online communication ensure that family caregivers can obtain timely advice for managing challenges [20]. Studies have shown that mobile apps significantly alleviate family caregivers' caregiving burden and depressive symptoms [21,22] while enhancing their caregiving competence and quality of life [23]. With technological advancements, this support has increasingly become more precise. For instance, the Olera.care platform constructs user profiles through needs questionnaires and algorithm models to achieve precise content delivery and educational resources [24]. Lukkien et al [25] developed a dementia care knowledge base using natural language processing technology and implemented intelligent question-and-answer support through chatbots. However, most existing technological innovations remain limited to web-based applications, with mobile app implementations and related research still insufficient. Future studies should explore mobile implementations to enhance support for family caregivers of persons with dementia.

Although mobile apps demonstrate considerable potential in supporting family caregivers, research findings remain inconsistent. Several randomized controlled trials and quasi-experimental studies have shown that, compared to control groups, family caregivers using supportive mobile apps exhibit no significant improvements in caregiving burden [26], caregiving competence [21], or depressive symptoms [26]. A study by Rodriguez et al [20] revealed a decline in family caregivers' intention to use a mobile app (Brain CareNotes) after the 6-month follow-up, indicating challenges in sustaining long-term engagement. Additional studies highlight further obstacles, including low user satisfaction, underuse of specific functions [21], and concerns regarding security and privacy [27]. An evaluation of mobile apps for persons with dementia and their family caregivers also revealed quality deficiencies and insufficient functionality design to meet family caregivers' actual needs effectively [28]. These inconsistent results all indicate the need to fully consider family caregivers' complex experiences, individualized needs, and specific caregiving contexts in the design of mobile apps. Relying solely on quantitative indicators may be insufficient to fully explain the reasons behind these inconsistencies.

XSL•FO
RenderX

Objectives

Qualitative methods uniquely uncover user experiences, explain anomalies in quantitative results, and identify overlooked issues [29]. Thus, capturing users' emotional feedback and experiences is vital for optimizing mobile apps. While qualitative studies examine family caregivers' perceptions of mobile apps, individual studies cannot capture the entirety of their experiences. For instance, Chapman et al [30] showed that the community function strongly supports family caregivers, whereas Dorell et al [31] found that not all users prefer group interactions. Moreover, existing systematic reviews mainly focus on quantitative research evidence to explore intervention effects on family caregivers [23,32,33]. Although a few reviews have included qualitative studies, the number of these studies was limited, primarily focusing on assessing mobile apps' functionality [18]. Another systematic review provided a comprehensive analysis of the features, security, and usability of mobile apps in dementia care [27]. However, the reviewed mobile apps predominantly focused on supporting persons with dementia. There is still a need for further exploration of the support that family caregivers receive from mobile apps [27]. Meta-synthesis can integrate multiple perspectives to analyze problems, resulting in a more comprehensive and in-depth understanding of complex phenomena [34]. No previous study has synthesized qualitative research on family caregivers' experiences, preferences, and challenges regarding supportive mobile apps. Given the growing mHealth reliance in caregiving, addressing this gap is urgent. Consequently, this study aimed to conduct a meta-synthesis of existing qualitative research to deeply analyze common experiences and individual differences in family caregivers' use of supportive mobile apps. The results will facilitate the creation and improvement of future relevant mobile apps that are grounded in person-centered principles to alleviate the psychological and physical stress of family caregivers and enhance the quality of care.

Methods

Study Design

This meta-synthesis adhered to the ENTREQ (Enhancing Transparency in Reporting the Synthesis of Qualitative Research) [35] and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [36] guidelines. The specific details can be found in Multimedia Appendix 1. The research protocol has been registered in PROSPERO, with registration CRD42024510905.

Search Strategy

A total of 7 English-language databases (the Cochrane Library, Embase, CINAHL, PubMed, PsycINFO, Scopus, and Web of Science Core Collection) and 4 Chinese-language databases (China National Knowledge Infrastructure, Wanfang Data, Weipu, and SinoMed) were searched from database establishment to March 2025. First, primary keywords were established based on existing knowledge, and then preliminary searches were conducted in the English-language database PubMed and the Chinese-language database China National Knowledge Infrastructure. We carefully reviewed the titles, abstracts, and keywords of relevant papers to establish comprehensive and specific search strategies tailored to each database. Second, we searched each database according to its search strategy. Third, references to the literature [30,31,37-46] were reviewed to identify additional studies. The specific search strategies are provided in Multimedia Appendix 2.

Eligibility Criteria and Study Selection

Studies were screened according to the inclusion and exclusion criteria shown in Textbox 1. In total, 2 researchers (HS and YH), who were trained in evidence-based methodology, independently screened the studies, extracted data, and cross-checked their work. If there were any disagreements, they instantly consulted a third researcher (WS) for evaluation. First, the search results were imported into EndNote (version 21; Clarivate Analytics), and duplicate references were deleted. Second, a preliminary screening of studies was conducted by reading the titles and abstracts to identify works for further scrutiny. Finally, a full-text review of the studies selected was conducted to confirm their final inclusion in the review.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Population: family caregivers of persons with dementia, such as spouses and offspring (if the study population included both persons with dementia and their family caregivers, only studies involving mobile apps with explicitly defined functions dedicated solely to family caregivers were included)
- Phenomena of interest: studies that explored the attitudes, expectations, or experiences of family caregivers regarding the use of supportive mobile apps (the mobile apps must explicitly include at least one function directly related to caregiving support)
- Context: family caregivers use supportive mobile apps in any environment, including in the home, community, hospital office, and others
- Study design and publication type: all types of qualitative studies and the qualitative components of mixed methods research

Exclusion criteria

- Population: formal caregivers (eg, nurses)
- Phenomena of interest: digital tools such as websites, telephone services, and SMS text messaging services and no specific functions in the mobile apps designed to support family caregivers of persons with dementia
- Study design and publication type: non–English- and non–Chinese-language papers, conference papers, and reviews; studies with no qualitative content or for which we were unable to extract valid qualitative content; and no access to the full text

Appraisal of Methodological Quality

In total, 2 researchers (HS and YH) independently assessed the quality of the studies based on the JBI Critical Appraisal Checklist for Qualitative Research [47]. The assessment consists of 10 items, which can be rated as "yes," "no," "unclear," or "not applicable." Disagreements were resolved by involving a third person (WS).

Data Extraction

The data extracted included the author, publication year, country or region, research method, data collection method, study population, phenomena of interest, main findings, participant quotations, and relevant author interpretations in the papers. Two researchers (HS and YH) independently extracted the data; a third researcher (WS) checked the data. Participant quotations and relevant author interpretations from the papers were further entered into Microsoft Excel (Microsoft Corp) for coding.

Data Synthesis

This study applied the thematic synthesis method proposed by Thomas and Harden [48] to the qualitative synthesis. This process has three stages: (1) coding text, (2) organizing codes into descriptive themes, and (3) generating analytical themes. Initially, 2 researchers (HS and YH) conducted line-by-line coding of textual data concerning family caregivers' perceptions of and experiences with mobile apps. We systematically compared qualitative descriptions across the studies to identify commonalities and variations, ensuring that the generated codes both captured core research content and preserved original contextual meanings. Discrepancies in coding were resolved through discussion with a third researcher (WS) until consensus was reached. Before completing this stage, we also reviewed all texts associated with the codes to ensure consistency in interpretation. Following this, researcher 1 (HS) continuously compared the differences and logical relationships among the codes to inductively and iteratively generate descriptive themes, which were then evaluated and finalized by 2 team members (YH and WS). In the final stage, researcher 1 (HS) designated analytical themes based on descriptive themes to transcend the findings of individual studies and produce a higher level of conceptual understanding. The broader team subsequently reviewed this process to ensure its rigor. Microsoft Excel was used for data organization.

Reciprocal translation and constant comparison permeated the thematic synthesis. For instance, in study 1, a family caregiver saying that they "perceived the app STAV as a source of stability that helped her to stay afloat in her daily life as a caregiver" was translated as "the app provides stability in daily lives." A participant in study 2 stated, "Yes, but it felt good with this app, I did not use it much, but it was there somehow then, and it felt good to have it..." which was translated as "the existence of the app brings positive feelings." In study 3, a family caregiver mentioned, "I feel that I have my guardian angel, my companion, my security guard. We feel more relaxed and safer," which was translated as "the app creates a sense of security and companionship." Although there were some differences in the translation results of these 3 studies, the concepts of "stability, positive feelings, security, and companionship" all pointed toward the positive impact of mobile apps on family caregivers' psychological well-being, with a core focus on their role in providing psychological stability. Considering the overall context, these subtle differences did not affect the consistency of the overall concept. Therefore, we unified these 3 statements under the code "mobile app as a source of psychological stability."

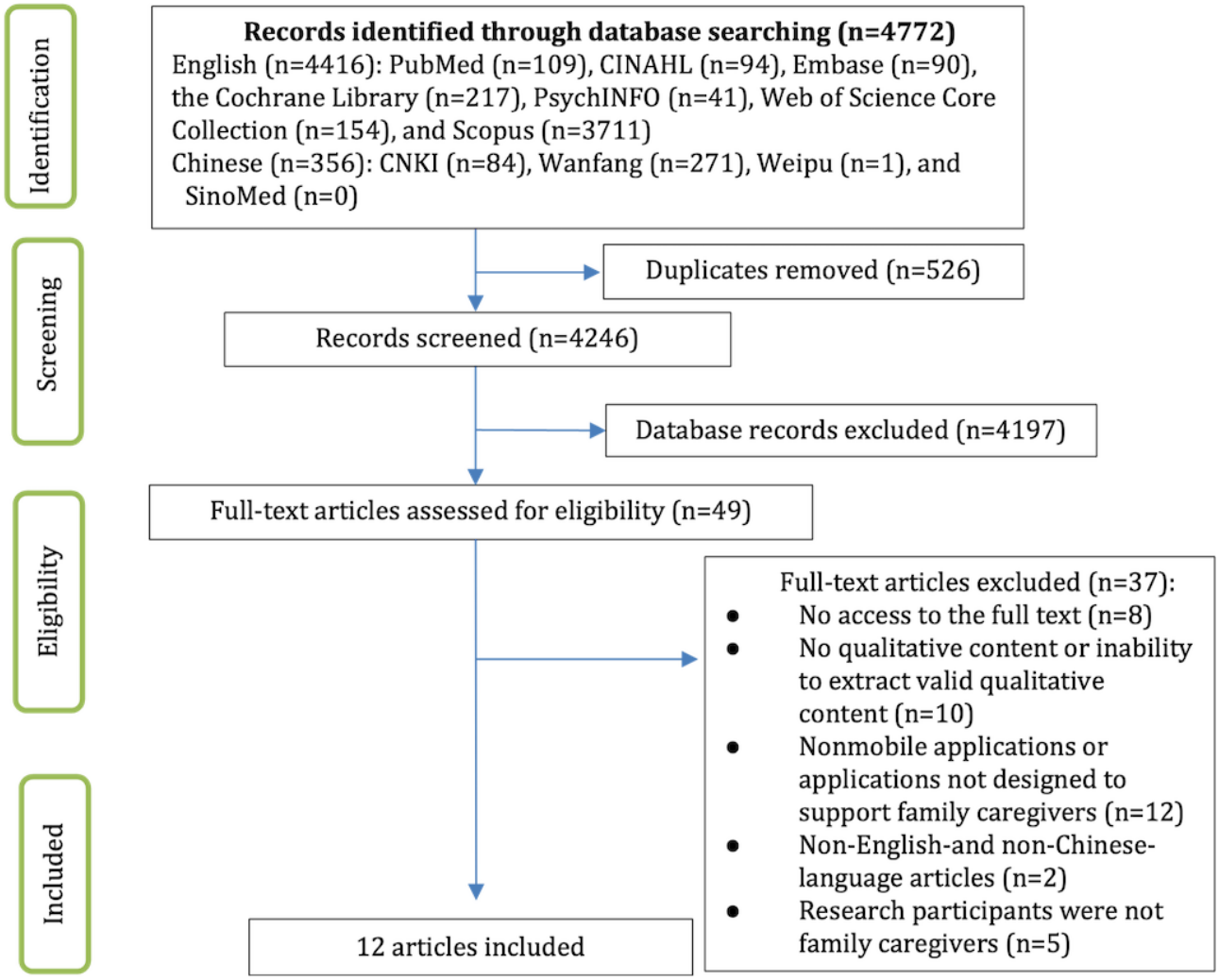
Results

Search Results

A comprehensive search of published qualitative research yielded 4772 studies. After the removal of duplicates and an examination of the titles and abstracts, of the 4772 studies, 49 (1.03%) were selected for full-text review. A total of 12 (0.25%) studies that met the requirements were ultimately included. Upon reviewing the references of the 12 studies, no more relevant studies were included. Figure 1 shows the screening procedure.

XSL•FO
RenderX

Figure 1. PRISMA flow diagram showing the study selection process. CNKI: China National Knowledge Infrastructure.



Quality Appraisal

Table 1 shows the quality assessment of the 12 included studies. None of the studies fully met the standards of the Joanna Briggs Institute, but 17% (2/12) studies met 9 items, 17% (2/12) studies met 8 items, 58% (7/12) studies met 7 items, and 8% (1/12) study met 6 items. The problematic items were that only 25% (3/12) reported the methodology, 33% (4/12) reported the researchers' circumstances, and none reported the researchers' influence on the research.

XSL•FO
RenderX

Table 1. Quality assessment of the included studies based on the JBI Critical Appraisal Tool.

Study	JBI Critical Appraisal Tool item									
	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]
Hussein et al [37], 2024	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Kagwa et al [38], 2024	Unclear	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Llaneza et al [39], 2024	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Castillo et al [40], 2024	Unclear	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Chapman et al [30], 2023	Unclear	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Mishra et al [41], 2023	Unclear	Yes	Yes	Unclear	Yes	No	No	Yes	Yes	Yes
Dorell et al [31], 2022	Unclear	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Kagwa et al [42], 2022	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Leung et al [43], 2022	Unclear	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Désormeaux-Moreau et al [44], 2021	Unclear	Yes	Yes	Yes	Yes	Yes	No	Unclear	Yes	Yes
Brites et al [45], 2020	Unclear	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Ruggiano et al [46], 2019	Unclear	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes

[a]Is there congruity between the stated philosophical perspective and the research methodology?

[b]Is there congruity between the research methodology and the research question or objectives?

[c]Is there congruity between the research methodology and the methods used to collect the data?

[d]Is there congruity between the research methodology and the representation and analysis of the data?

[e]Is there congruity between the research methodology and the interpretation of the results?

[f]Is there a statement locating the researcher culturally or theoretically?

[g]Is the influence of the researcher on the research, and vice versa, addressed?

[h]Are participants, and their voices, adequately represented?

[i]Is the research ethical according to current criteria or, for recent studies, and is there evidence of ethics approval by an appropriate body?

[j]Do the conclusions drawn in the research report flow from the analysis or interpretation of the data?

Characteristics of the Included Studies

Table 2 reports the characteristics of the qualitative components of all studies. The studies were conducted in 7 countries or regions, namely, the United States (4/12, 33%); Sweden (3/12, 25%); the United Kingdom (1/12, 8%); Australia (1/12, 8%); Brazil (1/12, 8%); Canada (1/12, 8%); and Hong Kong, China (1/12, 8%). A total of 67% (8/12) of the studies reported the types of mobile apps and their long-term use by family caregivers. In total, 33% (4/12) of the studies involved family caregivers who had only learned about and experienced mobile apps on the day of the interview or a few days before. A total of 25% (3/12) of the studies were qualitative reports on the same mobile app. The functions involved assessment, information and external resources, tracking and monitoring, communication platforms, reminders, stress management, care coordination, and digital painting. A total of 232 family caregivers were included. In total, 17% (2/12) of the studies, which were mixed methods studies, did not report the demographic characteristics of the participants. On the basis of the available data, the remaining studies primarily involved female participants (n=115), spouses (n=62), and offspring (n=52). A total of 33% (4/12) of the studies used mixed methods, and the rest (8/12, 67%) were qualitative studies. The methods included semistructured interviews (6/12, 50%), focus groups (5/12, 42%), and open-ended questions (1/12, 8%).

XSL•FO
RenderX

Table 2. Characteristics of the included studies.

Study	Study design and method of data collection and analysis	Participants	Phenomenon of interest	Type of app and use	Main findings (themes)
Hussein et al [37], 2024, Australia	Exploratory descriptive qualitative study; focus group; inductive thematic analysis	N=6 (female: 83%; spouses: 33%; offspring: 33%); average age 61.3 (range 30-85) y	To identify potential barriers to and facilitators of the implementation of PainChek by family caregivers of persons with dementia	PainChek helps caregivers generate pain assessments for persons with dementia through facial expressions or nonfacial input methods; presented with a brief video on PainChek and how to use it during the focus group	Three themes and 13 subthemes: (1) user-related factors (age of attending physician and family carer, physical limitations, skills and knowledge of mHealth[a] apps, previous experience with mHealth apps, family caregivers' knowledge of pain in dementia, and willingness to use PainChek), (2) intervention-related factors (PainChek; cost, technical issues, misinterpretations of how PainChek works, and consequences of PainChek use), and (3) contextual factors (collaboration and trust; health care team's acceptance, influence of family, pain communication, documentation, and discussion)
Kagwa et al [38], 2024, Sweden	Qualitative descriptive study; semistructured interviews; thematic analysis	Community-based social care professionals: n=11; family caregivers: n=19 (female: 84%; spouses: 68%; offspring: 16%); age ranging from 53 to 85 y	To explore how social care professionals and family caregivers of persons with dementia living at home experience providing and receiving support through a tailor-made mobile app named STAV[b]	STAV functions include chat, mindfulness exercises, web links to relevant sites, own contact list, and personal diary; use of the app for 8 wk	Three themes and 9 subthemes: (1) accessibility to support—bridging the gap (offering flexibility and enabling privacy), (2) engaging from a distance (varied levels of engagement, engagement with technology, navigating support, and forming a connection), and (3) limitations of the support (time lag in communication, limited functions in complex situations, and necessity of digital literacy)
Llaneza et al [39], 2024, United States	Qualitative descriptive study; semistructured interviews; inductive and deductive approaches	N=15 (female: 93%; spouses: 33%; offspring: 53%); average age 61.86 (range 34-80) y	To explore the experiences of family caregivers of older adults with cognitive impairment using a mindfulness therapy mobile app, determining facilitators and barriers	Mindfulness coach app; use of the app for 8 wk	Six themes: (1) convenience (flexibility to fit app use into individual schedules), (2) barriers (life and caregiving responsibilities interfere with daily app use), (3) perceived helpfulness (mindfulness coach as a source of knowledge and skill), (4) useful features (unique aspects of the app that increased mindfulness training), (5) suggested app improvements (future directions of app development for caregivers), and (6) mindfulness transfer (use of mindfulness skills in daily routines without opening the app)

XSL•FO
RenderX

Study	Study design and method of data collection and analysis	Participants	Phenomenon of interest	Type of app and use	Main findings (themes)
Castillo et al [40], 2024, United Kingdom	Mixed methods study; semistructured interviews; thematic analysis	N=35; demographic characteristics not reported	To explore the perceptions and unmet needs of family caregivers of persons with dementia regarding the use of supportive apps	App 1: Dementia Talk—provides information about dementia, tracks observed behaviors, and offers stress-related management; app 2: CLEAR[c] Dementia Care app provides information about dementia and allows to record observed behaviors; use of the app for 2 wk	Four themes and 15 subthemes: (1) positive aspects (informative, visual appeal and layout, and facilitated monitoring of behaviors and symptoms), (2) negative aspects (difficulty navigating the app, lack of customization, no added value, and presentation of the information), (3) content and feature considerations (connections to other supports, record keeping and monitoring, information and tips, strategies to manage stress, and coordination of care), and (4) user experience considerations (ease of use, layout, and cost)
Chapman et al [30], 2023, United States	Mixed methods study; in-depth, semistructured interviews; a priori qualitative data coding system to extract relevant interview themes	N=10 (female: 40%; spouses: 80%; offspring: 20%); average age 66.2 y	To gauge the usability of a new, multicomponent (mHealth) tablet app for family caregivers of persons with dementia	The CARE[d]-well app functions include assessments, psychoeducation, community, managing care, video hub, and goals; use of the app at least 4 times per wk for 1 mo	Three themes: (1) technical, (2) content and specific sections feedback, and (3) general feedback
Mishra et al [41], 2023, United States	Mixed methods study; open-ended questions; videoconferencing or in-person meetings; inductive approach	N=14 (female: 85%); average age 68 y	To explore the perceptions of family caregivers of persons with dementia regarding the acceptability and feasibility of the care coordination platform Care4AD	Care4AD app functions include coordinating care, monitoring physical activity, and scheduling reminders; demonstration of the mobile app to family caregivers before filling out the open-ended questionnaire	Five themes: (1) perceived ease of use, (2) perceived usefulness, (3) attitude toward use, (4) concerns about data privacy and sharing, and (5) technology anxiety
Dorell et al [31], 2022, Sweden	Qualitative descriptive study; semistructured interviews; thematic analysis	N=12 (female: 58%; spouses: 92%; offspring: 8%); age was not reported	To describe the experiences of family caregivers of persons with dementia who received professional support through a mobile app and its use	STAV functions include chat, mindfulness exercises, web links to relevant sites, own contact list, and personal diary; use of the app for 4 to 16 wk	Six themes: (1) filling a gap, (2) right time and right place, (3) foundation for inner calm, (4) better introduction and overcoming technical barriers, (5) relevant information in one place, and (6) way of offloading
Kagwa et al [42], 2022, Sweden	Qualitative explorative study; semistructured interviews; content analysis	N=12 (female: 58%; spouses: 92%; offspring: 8%); age was not reported	To explore how family caregivers of persons with dementia living at home, as consumers of health care services, cocreate value in a multi-stakeholder context through a tailor-made mHealth app	STAV; the functions include chat, mindfulness exercises, web links to relevant sites, own contact list, and personal diary; use of the app for 8 wk	Eight themes: (1) cooperating, (2) collating information, (3) combining complementary therapies, (4) colearning, (5) changing ways of doing things, (6) connecting, (7) coproduction, and (8) cerebral activities

XSL•FO
RenderX

Study	Study design and method of data collection and analysis	Participants	Phenomenon of interest	Type of app and use	Main findings (themes)
Leung et al [43], 2022, Hong Kong, China	Mixed methods; semistructured focus group interview; content analysis	N=14; demographic characteristics not reported	To assess the feasibility and acceptability of having family caregivers of persons with dementia draw electronic paintings using a mobile app and assess the preliminary effect of the intervention on their well-being	e-Painting app functions included picture sharing, painting, chat room, announcements, and self-assessment; use of the app for 8 wk	Five themes: (1) preferences with regard to features and expected training, (2) satisfaction and enjoyment in the use of the e-painting app, (3) the app as a channel to vent emotions, (4) the app made them feel connected, and (5) combating the challenges of caregiving
Désormeaux-Moreau et al [44], 2021, Canada	Qualitative study; focus group interview; content analysis	N=4 (female: 100%; spouses: 50%; offspring: 50%); average age 68.25 (range 58-78) y	To explore whether family caregivers view mobile apps as relevant to meeting their needs and as useful in managing disruptive behaviors and document the types of mobile apps that are of interest and appeal to most caregivers	8 apps; these apps' functions included providing dementia-related information and strategies for managing disruptive behaviors, reminders, live assistance, sharing, and scenario simulation; familiarization with the apps for 1 wk before the focus group	Two themes: (1) relevance of the mobile apps and (2) perceived usefulness of the mobile apps
Brites et al [45], 2020, Brazil	Qualitative study; 3 focus groups; thematic analysis	N=20 (female: 90%; spouses: 15%; offspring: 75%); average age 67.00 y	To analyze the perceptions of family caregivers and health professionals of a mobile app used for the caring and social support of persons with dementia	SMAI[e] functions included medication reminders, alarms, the ability to send reports and images, a GPS, and a chat with medical staff; 3 focus groups were organized after 6, 12, and 18 months of app use	Nine themes: (1) communication, (2) medication management, (3) feelings of caregivers, (4) impact of dementia on the lives of caregivers, (5) caregivers' illness, (6) care strategies, (7) caregiver safety, (8) communication with the health team, and (9) changes in daily care
Ruggiano et al [46], 2019, United States	Qualitative study; focus group interview; thematic analysis	N=36 (female: 72%; spouses: 28%; offspring:44%); average age 65.7 (range 42-89) y	To learn more about family caregivers' perspectives on how technology can improve dementia care through beta-test interviews for a family caregiver app	Care IT functions included education, tracking and monitoring, self-assessment, and linking of caregivers to dementia resources; demonstration of the functionality of the app to family caregivers before the interview began	Three themes and 3 subthemes: (1) current technology use, (2) usefulness of technology (linking members of the care network, facilitating meaningful interactions, and personalized technologies), and (3) ease of use

[a]mHealth: mobile health.

[b]STAV: STöd till AnhörigVårdare: Support to family caregivers.

[c]CLEAR: Congnition; Life Story & Personality; Emotional & Physical Wellbeing; Activity Environment, and Relationships.

[d]CARE: Caregiver Assessment, Resources, and Education.

[e]SMAI: Sistema Móvel de Assistência ao Idoso: Mobile Care System for Older Adults.

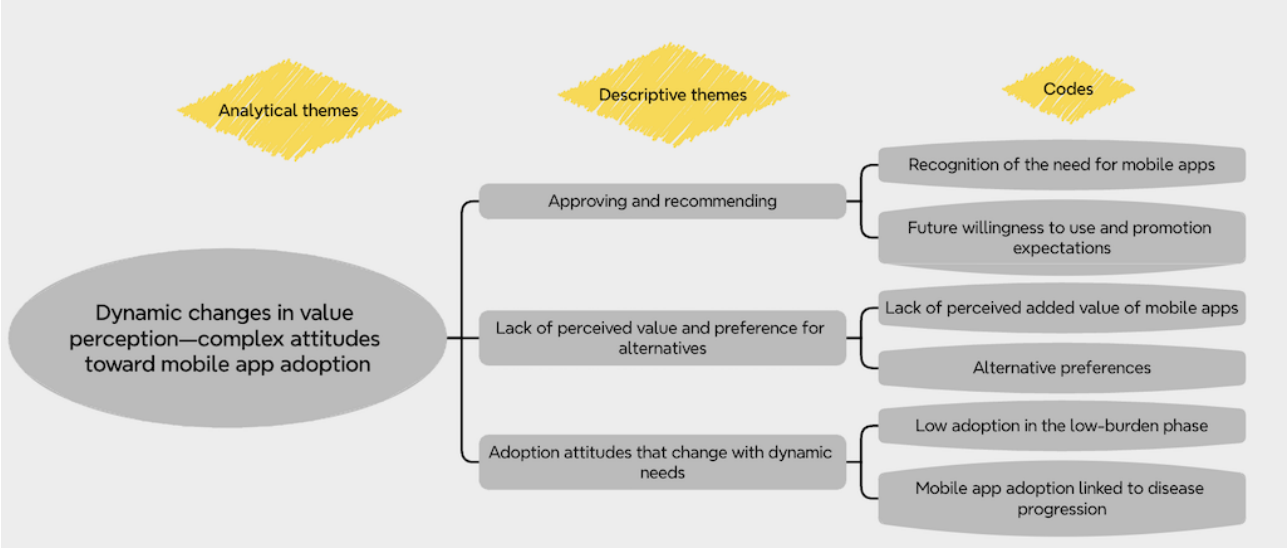
Main Findings of the Meta-Synthesis

Overview

The thematic synthesis ultimately yielded 38 codes, 12 descriptive themes, and 4 analytical themes. Figure 2 illustrates part of the thematic synthesis process, and the specific details of the synthesis can be found in Multimedia Appendix 3. Table 3 shows the specific themes and supporting quotes.

XSL•FO
RenderX

Figure 2. Part of the thematic synthesis process.



XSL•FO
RenderX

Table 3. Specific themes and supporting quotes.

Analytical themes and descriptive themes	Quotes
Dynamic changes in value perception—complex attitudes toward mobile app adoption	
Approving and recommending	• "I think the idea of it [app] is brilliant. There's definitely a need among caregivers...I was searching for something like this, and I just couldn't find it before this." [30] • "The family caregivers also said that they would recommend the app to others if it was available...Most caregivers had a positive response to the mobile app." [31]
Lack of perceived value and preference for alternatives	• "An app may not be a sufficient format to deliver resources: 'I would have rather had a book with that information than an app to read about strategies and on how to deal with some behavioral issues.'" [40] • "Some felt like the mindfulness feature was not suitable for them or achieved relaxation through other activities...I sit by the sewing machine and there is my mindfulness. There I can relax." [38]
Adoption attitudes that change with dynamic needs	• "It was too early in his [the person with dementia] disease to get such an access [the app], really...it would be more beneficial at a later stage of dementia when the condition had deteriorated." [42]
From tools to partners—a technology-empowered multidimensional support system for family caregivers	
Achieved a convenient and efficient user experience	• "Once I got to the actual program itself, it's easy to use. I mean, it's very clear...I found it pretty easy to maneuver, you know, to get to the different areas." [30] • "Wherever you are, if you have it on your phone, you can just stop and you know, actually gather your thoughts, no matter what the situation is." [39] • "We were unable to resolve things with her doctor, because the appointment was scheduled for two or three months in the future. I sent messages and photos, then received the response, within the timescale." [45]
Facilitated the optimization of care processes	• "Participants reported that the information presented in the app was comprehensive and would be helpful...I just found it very supportive and informative because I've had no experience on this." [40] • "It was good to have the notes so as not to forget points of discussion. It was also used as a timeline for the progress of the disease as it is difficult to notice small changes in a person with dementia when together all the time." [31] • "I used to pay more attention to coughing and urination, now I pay attention to the smell of her urine. One day my husband asked me: do you smell my mother's urine? I said: yes, of course, we need to see if it smells strong, because it could be a urinary tract infection. These two observations were excellent for me." [45]
Increased social connections and caregivers felt supported	• "Great support, I felt a support, from the health care, from you. When I was home alone with the problems, I had to ask and could get answers pretty quickly." [31] • "The community section of the application probably turned out to be the most substantial for me, in that there were other caregivers sharing their insight and ideas about their life with their loved ones, and the issues they were having." [30]
Regulated negative emotions	• "To focus in on my breathing really helped to reduce stress. It helped to kind of eliminate any extraneous background noises or background thoughts." [39] • "My emotions have become stable these days...Those paintings...make a person more positive, and make the emotions better." [43]
Brought intrinsic changes in self-care	• "I feel that I have my guardian angel, my companion, my security guard. The difference is the security. We feel more relaxed and safe." [45] • "The family caregiver further expressed that it had taken her a while to realize that she could use STAV for her personal needs instead of solely focusing on the needs of the PWD whom she cared for." [38] • "I haven't been using it recently because I sort of absorbed the habit into my life in a way, a way that I didn't necessarily need the app...the app sort of was like a starting point or a jump-off point for me to incorporate more of that part of thinking and action in my life." [39]
External and internal barriers—challenges in family caregivers' use of mobile apps	
External environment and technology factors	• "I couldn't really stop what I was doing and go do the app." [39] • "Participants were less likely to use mobile apps that required more steps to find needed information...one participant did not like Care4Dementia because there was 'too much research.'" [44] • "However, some caregivers experienced that they did not get answers from the nurse in time or it was too sporadic, so they stopped using the chat." [31]

XSL•FO
RenderX

Analytical themes and descriptive themes	Quotes
Intrinsic abilities or psychological factors	• "Both the social care professionals and the family caregivers expressed experiencing challenges for the family caregivers with the technicalities of using the app STAV, which they mostly related to age and a lack of experience. The lack of experience caused some family caregivers to disengage from the support features in STAV." [38] • "A major topic raised was privacy concerns which reduced the engagement level of some family caregivers with STAV...I do not want people to be able to Google me and come knocking on the door, maybe then start asking a lot of things or maybe pursue [the person with dementia]." [42]
Person-centered design—future directions for improving mobile apps	
Prioritizing the ease of use of the apps	• "All caregivers indicated that technologies targeting caregivers need to be easy to use in order to accommodate varying computer literacy...the more intuitive technologies are, the more likely they are for caregivers to adopt: When the first screen comes up, it ought to be said, 'Am I a caregiver?' Bam! Yes. 'What am I looking for as a caregiver?' Bam! And it ought to take me there. Bam! Right there. If I'm looking for my parent's health. Bam! Take me there! Am I looking for some other help? Bam! Take me there! All of these [features], that's great." [46] • "The participants suggested that having some hands-on practice before using the app at home would be beneficial." [43]
Personalization and function adaptation	• "If there are four stages [of dementia], you know, breaking it down to what resources would be helpful at this stage or at this point." [40] • "Participants shared their desire for more options within the app and to have it be more personalized either to their experience or schedule...the choices were limited, and I would have liked to see more options." [39] • "A participant suggested incorporating a 'live chat built into an app that has some resources attached to it' and noted that this feature has the potential to build a network of people they can reach out (to)." [40]

Dynamic Changes in Value Perception: Complex Attitudes Toward Mobile App Adoption

Overview

Differences in the perceived value of mobile apps contributed to the complex attitudes among family caregivers toward their use. While most family caregivers recognized the value of mobile apps and were willing to use and recommend them to others, some exhibited insufficient perception of their value, resulting in low motivation to adopt these mobile apps. However, this value perception changed dynamically in response to the burden and progression of the disease.

Approving and Recommending

A total of 42% (5/12) of the studies showed positive attitudes among family caregivers toward using mobile apps. They expressed a strong need for this tool and recognized the value that it could bring, believing that it was exactly what they had been searching for a long time [30,37,41,42]. Consequently, there was an expressed desire by family caregivers to continue using mobile apps and an indication that they would recommend these apps to others [30,31].

Lack of Perceived Value and Preference for Alternatives

Some family caregivers felt that the functionality provided by the mobile apps failed to deliver sufficient added value, which led them to question the apps' necessity [30,39,40,42,44]. This perspective may specifically pertain to mobile apps with limited functionality. For instance, "CLEAR Dementia Care" and "Care4Dementia" were 2 mobile apps that offered minimal functions, containing only information or tracking functions. Family caregivers noted that having these mobile apps was less beneficial than purchasing a book as they did not provide any additional information [40,44]. Family caregivers from Sweden stated that they did not use the diary and mindfulness functions because they had their preferred alternatives to use [40].

Adoption Attitudes That Change With Dynamic Needs

The acceptance of mobile apps among family caregivers varied according to their current caregiving burden and needs. Some family caregivers considered themselves already experienced in caregiving [42]; those with formal caregiver assistance believed that they did not need to join support programs [41], whereas family caregivers of early-stage persons with dementia felt that it was too early to seek such support [38]. Nonetheless, as the disease progressed and care demands intensified significantly, they required these mobile apps to receive more support [31,42].

From Tools to Partners: A Technology-Empowered Multidimensional Support System for Family Caregivers

Overview

The facilitation of multidimensional support functionalities enabled mobile apps to integrate into family caregivers' lives, assisting in enhancing caregiving efficiency and capabilities, alleviating caregiving burdens, providing emotional support, fostering social connections, and promoting self-care. This caused family caregivers to shift from reacting to challenges to proactively using resources, regulating emotions, and learning continuously. Thus, mobile apps became more than "problem-solving tools"—they were now "long-term partners empowering caregiving practices."

Achieved a Convenient and Efficient User Experience

A total of 67% (8/12) of the studies highlighted the mobile apps' facility and flexibility. "The Dementia Talk" was considered

XSL•FO
RenderX

visually appealing due to its multicolored layout [40], the electronic drawing platform was regarded as enjoyable [43], and the information provided in STAV was considered efficient [42]. In addition, family caregivers praised the straightforward navigation and found the technical operations to be easy and user-friendly [30,38,40,43]. More importantly, as the mobile apps were not limited by time and space, they provided greater flexibility, not only addressing physical distance constraints but also better accommodating the temporal preferences of family caregivers [31,38,39], particularly during the COVID-19 pandemic [31]. Such convenience allowed family caregivers to quickly access health care resources remotely without the need to schedule appointments or wait, thereby saving time that would otherwise be spent seeing a physician in person [38,45].

Facilitated the Optimization of Care Processes

In total, 75% (9/12) of the studies highlighted the positive impact of functionalities designed to optimize caregiving processes. Most mobile apps provided family caregivers with the opportunity to access educational resources and coping strategies, such as behavioral management recommendations and disease knowledge [30,31,40,42,44]. This promoted their understanding of the disease, especially for novice family caregivers. Furthermore, mobile apps with tracking or monitoring functions could serve as a central information repository, replacing traditional paperwork and assisting caregivers in storing, recording, and tracking the health status, daily activities, and medical information of the person with dementia to capture subtle changes in disease progression [38,40-42,45]. The aforementioned functions fostered proactive coping skills. Family caregivers could apply the coping strategies to translate them into specific actions for addressing unexpected real-life situations [30]. In addition, the monitoring function helped them develop daily observation habits, thereby proactively preventing various risks [45] and aiding in reviewing or reflecting on interventions after disruptive behaviors [44]. Other functions such as medication reminders [37,45], coordination of care [41], and pain assessments [37] were also considered able to increase the efficiency of the care provided.

Increased Social Connections and Caregivers Felt Supported

The group chat and community functions in the mobile apps held significant value as they helped family caregivers establish more social connections, alleviate feelings of loneliness, and strengthen their sense of belonging while solving problems. Through online communication with professionals, timely information exchange between family caregivers and the health team was promoted, allowing for personalized advice on addressing specific issues related to dementia, which instilled a sense of support [31,38,42,45]. STAV's chat function enabled professionals to start group conversations, helping encourage family caregivers to communicate, especially when they were hesitant to speak up [31]. The chat function in STAV also maintained continuous professional interactions by ensuring a fixed contact person [38]. In addition, the mobile apps fostered connections between family caregivers who had similar experiences, enabling them to share caregiving challenges and experiences, thus receiving mutual support and assistance [30,31,43,46].

Regulated Negative Emotions

This subtheme primarily focused on the mindfulness function, an electronic painting platform, and the diary function. Family caregivers reported that mindfulness practices helped calm their minds during stressful moments, alleviating anxiety through focused breathing or meditation [31,38,39,42]. The electronic painting platform served as an effective outlet for expressing negative emotions [43]; the act of drawing not only diverted attention and allowed for a temporary escape from worries but also enabled stress relief through creative expression. The diary provided family caregivers with a private space to express their concerns, helping them reduce their psychological burdens [38,44].

Psychological Empowerment and Self-Care

When family caregivers used mobile apps for ≥8 weeks, some of them exhibited effects on psychological empowerment and self-care. After long-term use, the mobile app gradually transcended its role as a tool and became a source of mental stability for family caregivers. Accessing the apps made family caregivers feel reassured [38,42,45]. In addition, the mobile apps promoted more self-care. Some family caregivers using STAV realized that mobile apps could be used not only to care for persons with dementia but also to meet their own needs [38]. The mindfulness function was mentioned the most in promoting self-care. Mindfulness reminded family caregivers to actively pay attention to their own health and take time to relax while fulfilling their caregiving responsibilities. This enabled family caregivers to cultivate the ability to be aware of their state and gradually establish a present-centered mindset [31,39,42]. Meanwhile, the mindfulness function also served as a stepping stone for self-care, prompting family caregivers to proactively seek extended resources and transition from passive reliance to active daily application [39].

External and Internal Barriers: Challenges in Family Caregivers' Use of Mobile Apps

Overview

The barriers arising from both external and internal sources collectively revealed the challenges that family caregivers face in using mobile apps. The explicit obstacles included the limited time and energy of family caregivers, physical conditions, design flaws, and device-related issues. Conversely, the implicit obstacles pertained to family caregivers' intrinsic abilities and psychological perceptions, which influenced their choices.

External Environment and Technology Factors

From the perspective of family caregivers' external environment, their health status [37,42], unforeseen life events, and a severe lack of time and energy due to caregiving responsibilities hindered the ability to consistently use mobile apps [30,31,39]. Undeniably, mobile apps' design flaws, such as a complex interface, confusing navigation, and malfunctioning menu functions, undermined the effectiveness of immediate support [30,31,42,44]. There were also shortcomings in functional design. Family caregivers believed that some functions, such as the goal section and video functions, were unnecessary [30]. In addition, certain functions offered limited options [39,40]. For example, while family caregivers wanted to use monitoring

XSL•FO
RenderX

functions to record issues, they could only select from preset options [40]. Family caregivers using STAV reported chat barriers such as limited interpersonal contact in chat rooms, single chats usually solving only 1 issue, and delayed responses causing chat termination [38]. After the intervention ended, they failed to maintain contact with health care providers through the chat function [42]. In addition, they expressed concerns about device-related issues such as compatibility, size, and costs [37,40,44].

Intrinsic Abilities or Psychological Factors

The limitation of mobile app use due to family caregivers' intrinsic abilities or psychological factors was reported in 42% (5/12) of the studies. From the perspective of intrinsic abilities, a lack of technical experience posed a challenge, particularly for older family caregivers unfamiliar with technology, leading them to choose to stop using mobile apps [37,38,42]. In terms of psychological factors, emotions, the desire to focus, interest, and trust reduced family caregivers' use of mindfulness functions [31,39]. Moreover, mobile apps' privacy issues caused concern for family caregivers as they feared exposing their personal information online and found it difficult to discuss personal issues without compromising their privacy [31,38,42].

Person-Centered Design: Future Directions for Improving Mobile Apps

Overview

Future improvements in mobile apps should be centered on user needs. By enhancing technological adaptability and personalized design, we can create more user-friendly and efficient collaborative mobile apps, allowing them to become truly tailored assistants for individuals.

Prioritizing the Ease of Use of the Apps

Owing to issues such as low digital literacy and the heavy caregiving burden, family caregivers in 67% (8/12) of the studies emphasized that the ease of use of mobile apps contributed to their use. Hence, the design needs to be user-friendly, with a simple and intuitive interface, such as large fonts, a clear layout, and fewer steps to follow [31,40-42,44]. It should also incorporate multilingual support, particularly in family caregivers' native languages, to enhance understanding [44]. In addition, using multimedia formats such as graphics and images can improve information engagement [40,42]. Providing training and guidance before use is imperative [30,31,39,40,43], such as one-on-one walk-throughs [30] and built-in video tutorials [40] for technical assistance.

Personalization and Function Adaptation

In total, 50% (6/12) of the studies reported family caregivers' demand for personalized and multifunctional content in mobile apps. Regarding personalization, information about dementia care should be personalized according to the different types and stages of the disease [39,40,42,46]. At the same time, the functionalities of the mobile apps could be customized to meet different needs [39,40,46]. For those mobile apps with single or fewer functions, it would be helpful to integrate more comprehensive functions to support dementia care, such as monitoring and recording functions, reminder functions, and calendar functions [31,39,40,42,46]. Furthermore, family caregivers hoped that self-care functions such as emotional assessment, background music, and mindfulness relaxation training would be included in mobile apps because they were necessary to take care of the whole person [40,43]. Of course, mobile apps should be committed to facilitating multi-role collaboration, such as allowing family members to share caregiving information and coordinate caregiving tasks [40].

Discussion

Principal Findings

Our study systematically reviewed and integrated 12 qualitative studies to examine the perspectives and experiences of family caregivers of persons with dementia regarding the use of supportive mobile apps. Ultimately, 4 analytical themes were generated. We found that family caregivers' attitudes toward using supportive mobile apps were complex and may change due to their dynamic perceptions of the mobile apps' value. At the same time, mobile apps empowered family caregivers through multidimensional support functions, helping improve care efficiency and capacity, alleviate caregiving burdens, provide emotional support, facilitate social connections, and promote self-care. Despite significant differences in the functionalities of various mobile apps, we found that they have considerable potential to evolve from simple assistive tools into long-term partners as they can integrate into family caregivers' daily lives and facilitate proactive changes in them. Nevertheless, external and internal barriers still affect family caregivers' use of these apps. External factors primarily included family caregivers' lack of time and energy, poor physical condition, design and function flaws, the devices' memory capabilities, compatibility, and costs. Internal factors included family caregivers' poor technological literacy, psychological resistance, and privacy concerns. In the future, more focus should be placed on person-centered improvements, especially on the ease of use of mobile apps. Mobile apps should address family caregivers' needs for personalization and multifunctionality, specifically by adapting content to disease stages and types while offering diverse support functions. Meanwhile, they should facilitate collaboration between family members and care teams to simplify caregiving tasks. According to the evidence obtained in this review, we discuss the results in this section and provide relevant recommendations for developing supportive mobile apps.

Family caregivers' attitudes toward supportive mobile apps can be explained through the technology acceptance model [49]. This model suggests that users' attitudes toward and intentions to use technology are driven by perceived usefulness and ease of use. Most family caregivers positively accepted and recommended mobile apps as they perceived practical value in alleviating caregiving burdens and stress. However, some family caregivers did not see mobile apps' added value or had alternative solutions, resulting in low perceived usefulness. Their acceptance of mobile apps also varied with caregiving burdens. Family caregivers of persons with dementia in the early stages or with rich experience may not use mobile apps due to perceived limited value. However, as burdens grow,

XSL•FO
RenderX

mobile apps' advantages in professional support and self-care increase caregivers' willingness to use them. This aligns with previous research, which highlights the crucial role of perceived usefulness in caregiver acceptance while indicating poorer acceptance among family caregivers who experience relatively low stress burdens, including caregivers in early-stage dementia care or those with substantial caregiving experience [50,51]. This dynamic change implies that mobile apps should be designed to suit family caregivers' evolving needs as mistimed assistance may stress them [52]. Although our study did not point out family caregivers' urgent functional needs at each stage, other research supplements this. For instance, Whitlatch and Orsulic-Jeras [53] detailed family caregivers' needs for information, education, and sociopsychological support at different stages. Just-in-time adaptive interventions are an emerging health intervention approach that emphasizes dynamically adjusting support based on individuals' changing states and contexts to provide help at the most needed times [54]. On the basis of this approach, the "CareQOL" app can dynamically push intervention measures according to family caregivers' self-reported data, effectively reducing their stress [55]. Therefore, in the future, developers should explore real-time monitoring of family caregivers' stress levels through artificial intelligence (AI) algorithms, automatically triggering matching functions to optimize mobile app design.

Findings from our study indicate that mobile apps have multidimensional functions and can provide support for family caregivers in a stressful environment. This positive impact is consistent with the stress coping model by Folkman and Lazarus [56]. Specifically, mobile apps can promote the stress adaptation of family caregivers by regulating stressors, facilitating cognitive appraisals, and providing coping processes. First, mobile apps' user-friendly design and functions that can optimize the caregiving process help family caregivers deal with stressors such as time and space constraints and difficulties in information retrieval. This boosts their sense of stress control and coping timeliness, ultimately improving their caregiving abilities and efficiency. Second, the use of mobile apps promotes family caregivers' cognitive restructuring of stress. Emotional regulation tools, especially mindfulness functions, allow family caregivers to cognitively reassess their emotions in the face of stress, weakening the threatening perception of stress and, thus, adjusting their responses to stress and mental states [57]. A certain period of use enables mobile apps to transcend their instrumental role and become psychological stabilizers, fostering a greater mindset of self-care. As previous research has indicated, when technology goes beyond meeting simple needs and ensures the satisfaction of psychological needs, users truly experience well-being [58]. Furthermore, due to the demands of caregiving, family caregivers often experience social isolation or difficulties in seeking help [59], and mobile apps can help build a social support network through professional collaboration and peer assistance. This not only provides dementia solutions but also enhances a sense of belonging and alleviates emotional isolation. Various support functions operate in coordination to alleviate family caregivers' stress through different coping mechanisms, underscoring the value of developing mobile apps with comprehensive functions to foster holistic care. These findings highlight future improvement directions in mobile apps to meet family caregivers' emotional needs and guide their self-growth. However, due to the limitations of this study, the long-term effects of mobile apps require further verification. For instance, after an intervention, family caregivers reported not maintaining contact with professionals through the chat function of the app [42]. This is in line with the results of the study by Cristancho-Lacroix et al [50]. In their online education program, almost no family caregivers continued to use the platform after the research was terminated [50]. The support provided by mobile apps may require the establishment of long-term trust as short-term impacts are often less pronounced [60]. Therefore, it is recommended that future research establish a tracking mechanism for ≥6 months or enhance user engagement through reminders, supervision, and other methods.

The barriers identified in our study have also been found in previous research [15,28,61]. Notably, many family caregivers are aged >60 years with limited technological literacy, making the use of mobile apps more challenging. The complex design of mobile apps and incompatible functions worsen their fear of technology. In addition, they are often pressed for time and, thus, have a strong need for simple, clear, and user-friendly interfaces. Furthermore, inadequate privacy measures make them cautious about mobile apps as unnecessary disclosure of sensitive information online can harm their dignity [62], and their ability to protect their online privacy appears to be lower than that of younger individuals [63]. Unfortunately, most apps developed for family caregivers of persons with dementia have very limited privacy and security measures, further hindering their use [27]. Surprisingly, our findings did not identify network barriers, likely because the included studies predominantly focused on countries with well-developed network infrastructures. However, this aspect should not be overlooked as not all mobile apps supported offline use in our included studies. Functions such as online communication and data synchronization still require internet connectivity. Rural family caregivers in another study reported a lack of internet connection as a significant barrier [19]. Optimizing the design of mobile apps and addressing a range of family caregivers' concerns requires joint efforts from developers and policy makers. Therefore, we propose the following recommendations. Developers should adopt an age-friendly design by providing large fonts, adding images, enabling voice interactions, and supporting multiple languages. Simplifying navigation processes is essential, and this can be achieved by providing step-by-step guidance and video tutorials to reduce operational difficulties. In addition, it is necessary to fully leverage the advantages of mobile apps that support offline functionality and develop core offline functions as key components to ensure that basic services can be maintained in scenarios without network access or with unstable connectivity. From a technical security perspective, it is recommended to offer privacy setting options that allow family caregivers to determine the visibility of information and enable an anonymous mode. Furthermore, the mobile app should also safeguard data security through encryption technology and provide users with privacy protocols. Importantly, users should be involved during critical development stages such as needs analysis and testing to ensure that the design truly meets their requirements [64]. For policy makers, establishing a long-term support mechanism is crucial. On the one hand, special funds

XSL•FO
RenderX

should be set up to support the development of mobile apps and related free training programs [65]. Training should include digital literacy enhancement courses and time management training courses to enhance family caregivers' digital literacy and time management efficiency. On the other hand, technology subsidy policies should be implemented, including reimbursement of mobile app download or use fees, data flow subsidies, and other substantial support. In addition, public service platform resources should be incorporated to help create a unified service entrance in mobile apps for family caregivers, aggregating support resources such as healthcare, social assistance, legal, and financial services.

Throughout the dementia care process, the experience, capabilities, and needs of family caregivers are unique. The absence of personalized content in mobile apps may lead to ineffective care outcomes [66]. Rodriguez et al [20] developed "Brain CareNotes," which enabled users to implement personalized settings, thereby helping users from ethnic minority groups adapt to their cultural background. The digital platform developed by Fan et al [24] integrated AI and large language models to customize and deliver the most pertinent information to users, thereby augmenting user engagement and satisfaction. This implies that a variety of tools and technologies, such as AI, customized settings, and customized questionnaires [21], can be used to assist mobile apps in achieving personalization. In addition, there are still shortcomings in the customization of functions within current mobile apps. For example, there are many general stress relief mobile apps but none specifically tailored to older family caregivers and their unique experiences and sources of stress [67]. In addition, there is a lack of integrated multifunctional mobile apps [16]. This means that family caregivers must use multiple apps to satisfy all their needs, evidently increasing their burden. Therefore, they need a comprehensive mobile app that integrates multiple functions and offers customization. Meanwhile, mobile apps should play an increased role in connecting members of the care network and promoting social service coordination. In our study, family caregivers expressed a desire for a mobile app that could provide a platform for real-time information sharing with stakeholders [40]. Previous research has developed electronic health information systems that enable information sharing among family members, community caregivers, and clinical nursing teams [68]. This sharing typically helps reduce information gaps and alleviates the stress associated with coordinating health care. However, these programs are often based on medical systems or have limited functionality [68,69] and are rarely integrated into supportive mobile apps for family caregivers of persons with dementia. Hence, this study suggests building a social network within the mobile app to improve online collaboration among primary caregivers, professionals, and other stakeholders.

Limitations

This study has some limitations. Only English- and Chinese-language studies were included in the analysis. The exclusion of non–English- and non–Chinese-language and unpublished studies could have left out important information. The mobile apps included in the studies varied in functionality, and the family caregivers' understanding of these functionalities also varied. Therefore, their perceptions and experiences may be biased accordingly. Moreover, none of the included studies conducted long-term interventions lasting >6 months. This resulted in a lack of sufficient evidence for us to investigate the long-term effects of mobile apps. Most of the studies (11/12, 92%) were carried out in high-income countries, so the results might not be applicable to low- and middle-income countries. Moreover, most of the included studies (9/12, 75%) lacked theoretical methodological elaboration. Thus, future research should include additional qualitative studies with clear methodologies to produce high-quality qualitative syntheses.

Conclusions

By integrating a large amount of qualitative data, this study illustrated the complex attitudes of family caregivers of persons with dementia toward mobile apps. Through the empowerment from multidimensional supportive mobile apps, family caregivers can effectively enhance their caregiving capabilities and efficiency, alleviate caregiving burdens, reduce negative emotions, foster social connections, and promote self-focus. This continuous and multidimensional support implies that mobile apps have the potential to become trusted partners for family caregivers throughout the long-term caregiving process, assisting them in addressing care challenges at different stages. However, it is crucial to address both external and internal challenges, such as design flaws, family caregivers' lack of technological literacy, time constraints, concerns about privacy breaches, and other device-related issues. Future improvement directions should prioritize the ease of use of technology and the development of personalized, multifunctional mobile apps, as well as promoting online collaboration among members of the care network.

Acknowledgments

This study was funded by the Medical Science and Technology Project of Zhejiang Province (project 2022KY219).

Authors' Contributions

HS, JW, JW, and XS contributed to the conception and design of the study. HS, YH, and WS conducted the literature screening and data extraction. HS, HY, WS, and HW conducted the data analysis and data synthesis. HS wrote the main manuscript text. All authors revised the manuscript. JW supervised the research. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
PRISMA checklist and ENTREQ checklist.
[DOCX File , 35 KB - mhealth_v13i1e65983_app1.docx]

Multimedia Appendix 2
Search strategies.
[DOCX File , 31 KB - mhealth_v13i1e65983_app2.docx]

Multimedia Appendix 3
The thematic synthesis process.
[DOCX File , 19 KB - mhealth_v13i1e65983_app3.docx]

References

1. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013 Jan;9(1):63-75.e2. [doi: 10.1016/j.jalz.2012.11.007] [Medline: 23305823]
2. World Alzheimer report 2023. Alzheimer's Disease International. URL: https://www.alzint.org/resource/world-alzheimer-report-2023/ [accessed 2024-03-20]
3. Connors MH, Seeher K, Teixeira-Pinto A, Woodward M, Ames D, Brodaty H. Dementia and caregiver burden: a three-year longitudinal study. Int J Geriatr Psychiatry 2020 Feb 10;35(2):250-258. [doi: 10.1002/gps.5244] [Medline: 31821606]
4. Chandrasekaran N, Thulasingam M, Sahu SK, Sood V, Menon V. Effects of web-based training for carers in old age homes using WHO-iSupport for dementia, as a training tool - a mixed-method study. Asian J Psychiatr 2022 Dec;78:103308. [doi: 10.1016/j.ajp.2022.103308] [Medline: 36368147]
5. Sriram V, Jenkinson C, Peters M. Informal carers' experience and outcomes of assistive technology use in dementia care in the community: a systematic review protocol. Syst Rev 2019 Jul 03;8(1):158 [FREE Full text] [doi: 10.1186/s13643-019-1081-x] [Medline: 31269995]
6. Sheehan OC, Haley WE, Howard VJ, Huang J, Rhodes JD, Roth DL. Stress, burden, and well-being in dementia and nondementia caregivers: insights from the caregiving transitions study. Gerontologist 2021 Jul 13;61(5):670-679 [FREE Full text] [doi: 10.1093/geront/gnaa108] [Medline: 32816014]
7. Teahan Á, Lafferty A, Cullinan J, Fealy G, O'Shea E. An analysis of carer burden among family carers of people with and without dementia in Ireland. Int Psychogeriatr 2021 Apr 02;33(4):347-358 [FREE Full text] [doi: 10.1017/S1041610220000769] [Medline: 32484125]
8. Kim B, Noh GO, Kim K. Behavioural and psychological symptoms of dementia in patients with Alzheimer's disease and family caregiver burden: a path analysis. BMC Geriatr 2021 Mar 05;21(1):160 [FREE Full text] [doi: 10.1186/s12877-021-02109-w] [Medline: 33663416]
9. Bressan V, Visintini C, Palese A. What do family caregivers of people with dementia need? A mixed-method systematic review. Health Soc Care Community 2020 Nov 16;28(6):1942-1960. [doi: 10.1111/hsc.13048] [Medline: 32542963]
10. Leung DK, Wong KK, Spector A, Wong GH. Exploring dementia family carers' self-initiated strategies in managing behavioural and psychological symptoms in dementia: a qualitative study. BMJ Open 2021 Aug 25;11(8):e048761 [FREE Full text] [doi: 10.1136/bmjopen-2021-048761] [Medline: 34433600]
11. Sun Y, Ji M, Leng M, Li X, Zhang X, Wang Z. Comparative efficacy of 11 non-pharmacological interventions on depression, anxiety, quality of life, and caregiver burden for informal caregivers of people with dementia: a systematic review and network meta-analysis. Int J Nurs Stud 2022 May;129:104204. [doi: 10.1016/j.ijnurstu.2022.104204] [Medline: 35247788]
12. Novais T, Dauphinot V, Krolak-Salmon P, Mouchoux C. How to explore the needs of informal caregivers of individuals with cognitive impairment in Alzheimer's disease or related diseases? A systematic review of quantitative and qualitative studies. BMC Geriatr 2017 Apr 17;17(1):86 [FREE Full text] [doi: 10.1186/s12877-017-0481-9] [Medline: 28415968]
13. Bradway M, Carrion C, Vallespin B, Saadatfard O, Puigdomènech E, Espallargues M, et al. mHealth assessment: conceptualization of a global framework. JMIR Mhealth Uhealth 2017 May 02;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]
14. Ahmad NA, Mat Ludin AF, Shahar S, Mohd Noah SA, Mohd Tohit N. Willingness, perceived barriers and motivators in adopting mobile applications for health-related interventions among older adults: a scoping review. BMJ Open 2022 Mar 09;12(3):e054561 [FREE Full text] [doi: 10.1136/bmjopen-2021-054561] [Medline: 35264349]
15. Guessi Margarido M, Shah A, Seto E. Smartphone applications for informal caregivers of chronically ill patients: a scoping review. NPJ Digit Med 2022 Mar 21;5(1):33 [FREE Full text] [doi: 10.1038/s41746-022-00567-z] [Medline: 35314766]
16. Kim E, Baskys A, Law AV, Roosan MR, Li Y, Roosan D. Scoping review: the empowerment of Alzheimer's disease caregivers with mHealth applications. NPJ Digit Med 2021 Sep 07;4(1):131 [FREE Full text] [doi: 10.1038/s41746-021-00506-4] [Medline: 34493819]

XSL•FO
RenderX

17. Faieta J, Sheehan J, DiGiovine C. Mhealth interventions to improve health and quality of life related outcomes for informal dementia caregivers: a scoping review. Assist Technol 2022 May 04;34(3):362-374. [doi: 10.1080/10400435.2020.1829174] [Medline: 33006526]
18. Rathnayake S, Moyle W, Jones C, Calleja P. mHealth applications as an educational and supportive resource for family carers of people with dementia: an integrative review. Dementia (London) 2019;18(7-8):3091-3112. [doi: 10.1177/1471301218768903] [Medline: 29631492]
19. Wilding C, Davis H, Rasekaba T, Hamiduzzaman M, Royals K, Greenhill J, et al. Volunteers' support of carers of rural people living with dementia to use a custom-built application. Int J Environ Res Public Health 2021 Sep 20;18(18):9909 [FREE Full text] [doi: 10.3390/ijerph18189909] [Medline: 34574832]
20. Rodriguez MJ, Kercher VM, Jordan EJ, Savoy A, Hill JR, Werner N, et al. Technology caregiver intervention for Alzheimer's disease (I-CARE): feasibility and preliminary efficacy of Brain CareNotes. J Am Geriatr Soc 2023 Dec 14;71(12):3836-3847. [doi: 10.1111/jgs.18591] [Medline: 37706540]
21. Núñez-Naveira L, Alonso-Búa B, de Labra C, Gregersen R, Maibom K, Mojs E, et al. UnderstAID, an ICT platform to help informal caregivers of people with dementia: a pilot randomized controlled study. Biomed Res Int 2016;2016:5726465 [FREE Full text] [doi: 10.1155/2016/5726465] [Medline: 28116300]
22. Goto Y, Suematsu M, Imaizumi T, Suzuki Y. Preliminary study of the effect of the web application on caregiver burden in dementia and behavioural and psychological symptoms of dementia. Nagoya J Med Sci 2024 Aug;86(3):383-391. [doi: 10.18999/nagjms.86.3.383] [Medline: 39355354]
23. Shin Y, Kim SK, Kim Y, Go Y. Effects of app-based mobile interventions for dementia family caregivers: a systematic review and meta-analysis. Dement Geriatr Cogn Disord 2022 May 24;51(3):203-213. [doi: 10.1159/000524780] [Medline: 35609526]
24. Fan Q, Hoang MN, DuBose L, Ory MG, Vennatt J, Salha D, et al. The Olera.care digital caregiving assistance platform for dementia caregivers: preliminary evaluation study. JMIR Aging 2024 Apr 17;7:e55132 [FREE Full text] [doi: 10.2196/55132] [Medline: 38630527]
25. Lukkien DR, Ipakchian Askari S, Stolwijk NE, Hofstede BM, Nap HH, Boon WP, et al. Making co-design more responsible: case study on the development of an AI-based decision support system in dementia care. JMIR Hum Factors 2024 Jul 31;11:e55961 [FREE Full text] [doi: 10.2196/55961] [Medline: 39083768]
26. Hong QX, Wang WF, Yang YH, Tung YC, Dai HJ, Hsu WC, et al. The effectiveness of virtual passport, an app-based intervention, for dementia care. Front Psychiatry 2024 Sep 26;15:1457923 [FREE Full text] [doi: 10.3389/fpsyt.2024.1457923] [Medline: 39391088]
27. Ye B, Chu CH, Bayat S, Babineau J, How T, Mihailidis A. Researched apps used in dementia care for people living with dementia and their informal caregivers: systematic review on app features, security, and usability. J Med Internet Res 2023 Oct 12;25:e46188 [FREE Full text] [doi: 10.2196/46188] [Medline: 37824187]
28. Werner NE, Brown JC, Loganathar P, Holden RJ. Quality of mobile apps for care partners of people with Alzheimer disease and related dementias: mobile app rating scale evaluation. JMIR Mhealth Uhealth 2022 Mar 29;10(3):e33863 [FREE Full text] [doi: 10.2196/33863] [Medline: 35348467]
29. Atkins S, Lewin S, Smith H, Engel M, Fretheim A, Volmink J. Conducting a meta-ethnography of qualitative literature: lessons learnt. BMC Med Res Methodol 2008 Apr 16;8:21 [FREE Full text] [doi: 10.1186/1471-2288-8-21] [Medline: 18416812]
30. Chapman KR, Maynard T, Sadhu S, Mankodiya K, Uebelacker L, Davis JD, et al. Beta test of a multicomponent mobile health application for dementia caregivers. J Technol Behav Sci 2023 Jul 07;8(4):422-431. [doi: 10.1007/S41347-023-00328-z]
31. Dorell Å, Konradsen H, Kallström AP, Kabir ZN. "A friend during troubled times": experiences of family caregivers to persons with dementia when receiving professional support via a mobile app. PLoS One 2022 Aug 2;17(8):e0271972 [FREE Full text] [doi: 10.1371/journal.pone.0271972] [Medline: 35917295]
32. de-Moraes-Ribeiro FE, Moreno-Cámara S, da-Silva-Domingues H, Palomino-Moral P, Del-Pino-Casado R. Effectiveness of internet-based or mobile app interventions for family caregivers of older adults with dementia: a systematic review. Healthcare (Basel) 2024 Jul 27;12(15):1494 [FREE Full text] [doi: 10.3390/healthcare12151494] [Medline: 39120195]
33. Leng M, Zhao Y, Xiao H, Li C, Wang Z. Internet-based supportive interventions for family caregivers of people with dementia: systematic review and meta-analysis. J Med Internet Res 2020 Sep 09;22(9):e19468 [FREE Full text] [doi: 10.2196/19468] [Medline: 32902388]
34. Walsh D, Downe S. Meta-synthesis method for qualitative research: a literature review. J Adv Nurs 2005 Apr;50(2):204-211. [doi: 10.1111/j.1365-2648.2005.03380.x] [Medline: 15788085]
35. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012 Nov 27;12:181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
36. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]

XSL•FO
RenderX

37. Hussein A, Moullin J, Stafford A, Hughes J, Quested E. Identification of barriers and facilitators to the use of a new pain assessment mobile health application by family carers of people with dementia: a qualitative study. Ageing Int 2024 Oct 30;50(1):22. [doi: 10.1007/S12126-024-09577-6]
38. Kagwa AS, Dorell Å, Konradsen H, Vikström S, Kabir ZN. Providing and receiving support through a tailor-made mobile app: a qualitative study on experience of professionals and family caregivers to persons with dementia. BMC Geriatr 2024 Jun 25;24(1):554 [FREE Full text] [doi: 10.1186/s12877-024-05151-6] [Medline: 38918728]
39. Llaneza DH, Llaneza AJ, Kozlov E. Perceived benefits and barriers of mHealth mindfulness use for caregivers of older adults with cognitive impairment: a qualitative exploration. Clin Gerontol 2024 Dec 11;47(1):66-77. [doi: 10.1080/07317115.2022.2154728] [Medline: 36502294]
40. Castillo LI, Tran V, Hadjistavropoulos T. Are mobile apps meeting the needs of caregivers of people living with dementia? An evaluation of existing apps for caregivers. Aging Ment Health 2024 Apr 12;28(4):577-586 [FREE Full text] [doi: 10.1080/13607863.2023.2177832] [Medline: 36775643]
41. Mishra RK, Park C, Momin AS, Rafaei NE, Kunik M, York MK, et al. Care4AD: a technology-driven platform for care coordination and management: acceptability study in dementia. Gerontology 2023 Sep 12;69(2):227-238 [FREE Full text] [doi: 10.1159/000526219] [Medline: 36096091]
42. Kagwa AS, Konradsen H, Kabir ZN. Value co-creation with family caregivers to people with dementia through a tailor-made mHealth application: a qualitative study. BMC Health Serv Res 2022 Nov 16;22(1):1362 [FREE Full text] [doi: 10.1186/s12913-022-08704-w] [Medline: 36384534]
43. Leung AY, Cheung T, Fong TK, Zhao IY, Kabir ZN. The use of an electronic painting platform by family caregivers of persons with dementia: a feasibility and acceptability study. Healthcare (Basel) 2022 May 09;10(5):870 [FREE Full text] [doi: 10.3390/healthcare10050870] [Medline: 35628007]
44. Désormeaux-Moreau M, Michel C, Vallières M, Racine M, Poulin-Paquet M, Lacasse D, et al. Mobile apps to support family caregivers of people with Alzheimer disease and related dementias in managing disruptive behaviors: qualitative study with users embedded in a scoping review. JMIR Aging 2021 Apr 16;4(2):e21808 [FREE Full text] [doi: 10.2196/21808] [Medline: 33861207]
45. Brites AD, Caldas CP, Motta LB, Carmo DS, Sztajnberg A, Stutzel MC. Mobile care system for older adults: perceptions about its use in caring for people with dementia. Rev Bras Geriatr Gerontol 2020;23(1):e190277. [doi: 10.1590/1981-22562020023.190277]
46. Ruggiano N, Brown EL, Shaw S, Geldmacher D, Clarke P, Hristidis V, et al. The potential of information technology to navigate caregiving systems: perspectives from dementia caregivers. J Gerontol Soc Work 2019 Nov 13;62(4):432-450. [doi: 10.1080/01634372.2018.1546786] [Medline: 30422754]
47. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc 2015 Sep;13(3):179-187. [doi: 10.1097/XEB.0000000000000062] [Medline: 26262565]
48. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
49. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
50. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. J Med Internet Res 2015 May 12;17(5):e117 [FREE Full text] [doi: 10.2196/jmir.3717] [Medline: 25967983]
51. Mendez KJ, Budhathoki C, Labrique AB, Sadak T, Tanner EK, Han HR. Factors associated with intention to adopt mHealth apps among dementia caregivers with a chronic condition: cross-sectional, correlational study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e27926 [FREE Full text] [doi: 10.2196/27926] [Medline: 34463637]
52. Orsulic-Jeras S, Whitlatch CJ, Szabo SM, Shelton EG, Johnson J. The SHARE program for dementia: implementation of an early-stage dyadic care-planning intervention. Dementia (London) 2019 Jan 13;18(1):360-379. [doi: 10.1177/1471301216673455] [Medline: 27738110]
53. Whitlatch CJ, Orsulic-Jeras S. Meeting the informational, educational, and psychosocial support needs of persons living with dementia and their family caregivers. Gerontologist 2018 Jan 18;58(suppl_1):S58-S73. [doi: 10.1093/geront/gnx162] [Medline: 29361068]
54. Spruijt-Metz D, Wen CK, O'Reilly G, Li M, Lee S, Emken BA, et al. Innovations in the use of interactive technology to support weight management. Curr Obes Rep 2015 Dec;4(4):510-519 [FREE Full text] [doi: 10.1007/s13679-015-0183-6] [Medline: 26364308]
55. Wang J, Wu Z, Choi SW, Sen S, Yan X, Miner JA, et al. The dosing of mobile-based just-in-time adaptive self-management prompts for caregivers: preliminary findings from a pilot microrandomized study. JMIR Form Res 2023 Sep 14;7:e43099 [FREE Full text] [doi: 10.2196/43099] [Medline: 37707948]
56. Folkman S, Lazarus RS. Coping as a mediator of emotion. J Pers Soc Psychol 1988 Mar;54(3):466-475. [Medline: 3361419]

XSL•FO
RenderX

57. Bossi F, Zaninotto F, D'Arcangelo S, Lattanzi N, Malizia AP, Ricciardi E. Mindfulness-based online intervention increases well-being and decreases stress after COVID-19 lockdown. Sci Rep 2022 Apr 20;12(1):6483 [FREE Full text] [doi: 10.1038/s41598-022-10361-2] [Medline: 35444198]
58. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797 [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
59. Tyrrell M, Hillerås P, Skovdahl K, Fossum B, Religa D. Voices of spouses living with partners with neuropsychiatric symptoms related to dementia. Dementia (London) 2019 Apr 06;18(3):903-919. [doi: 10.1177/1471301217693867] [Medline: 28385034]
60. Hong YA, Shen K, Han H, Ta Park V, Bagchi P, Lu HK, et al. A WeChat-based intervention, wellness enhancement for caregivers (WECARE), for Chinese American dementia caregivers: pilot assessment of feasibility, acceptability, and preliminary efficacy. JMIR Aging 2023 Apr 05;6:e42972 [FREE Full text] [doi: 10.2196/42972] [Medline: 37018042]
61. Wang C, Qi H. Influencing factors of acceptance and use behavior of mobile health application users: systematic review. Healthcare (Basel) 2021 Mar 22;9(3):357 [FREE Full text] [doi: 10.3390/healthcare9030357] [Medline: 33809828]
62. Price 2nd WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2019 Jan;25(1):37-43 [FREE Full text] [doi: 10.1038/s41591-018-0272-7] [Medline: 30617331]
63. Friedman AB, Pathmanabhan C, Glicksman A, Demiris G, Cappola AR, McCoy MS. Addressing online health privacy risks for older adults: a perspective on ethical considerations and recommendations. Gerontol Geriatr Med 2022 Apr 21;8:23337214221095705 [FREE Full text] [doi: 10.1177/23337214221095705] [Medline: 35493968]
64. Guisado-Fernández E, Giunti G, Mackey LM, Blake C, Caulfield BM. Factors influencing the adoption of smart health technologies for people with dementia and their informal caregivers: scoping review and design framework. JMIR Aging 2019 Apr 30;2(1):e12192 [FREE Full text] [doi: 10.2196/12192] [Medline: 31518262]
65. Quinn CC, Staub S, Barr E, Gruber-Baldini A. Mobile support for older adults and their caregivers: dyad usability study. JMIR Aging 2019 May 23;2(1):e12276 [FREE Full text] [doi: 10.2196/12276] [Medline: 31518271]
66. Wasilewski MB, Stinson JN, Cameron JI. Web-based health interventions for family caregivers of elderly individuals: a scoping review. Int J Med Inform 2017 Jul;103:109-138. [doi: 10.1016/j.ijmedinf.2017.04.009] [Medline: 28550996]
67. Grossman MR, Zak DK, Zelinski EM. Mobile apps for caregivers of older adults: quantitative content analysis. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e162 [FREE Full text] [doi: 10.2196/mhealth.9345] [Medline: 30061093]
68. Fabius CD, Wec A, Saylor MA, Smith JM, Gallo JJ, Wolff JL. "Caregiving is teamwork…" information sharing in home care for older adults with disabilities living in the community. Geriatr Nurs 2023 Nov;54:171-177. [doi: 10.1016/j.gerinurse.2023.09.001] [Medline: 37788565]
69. Nadash P, Silverstein NM, Porell F. The dementia care coordination program: engaging health systems in caregiver supports. Dementia (London) 2019 May 17;18(4):1273-1285. [doi: 10.1177/1471301217697466] [Medline: 29149791]

Abbreviations

AI: artificial intelligence
ENTREQ: Enhancing Transparency in Reporting the Synthesis of Qualitative Research
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 31.08.24; peer-reviewed by S Zhi, HR Choi, S Qu; comments to author 26.02.25; revised version received 22.04.25; accepted 14.05.25; published 18.06.25.

Please cite as:
Shen H, Han Y, Shi W, Yu J, Shan X, Wang H, Wang J
Perspectives and Experiences of Family Caregivers Using Supportive Mobile Apps in Dementia Care: Meta-Synthesis of Qualitative Research
JMIR Mhealth Uhealth 2025;13:e65983
URL: https://mhealth.jmir.org/2025/1/e65983
doi:10.2196/65983
PMID:

©Haifei Shen, Yi Han, Wen Shi, Jiangxuan Yu, Xueqi Shan, Hongyao Wang, Junjie Wang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Mobile Health Intervention Tools Promoting HIV Pre-Exposure Prophylaxis Among Adolescent Girls and Young Women in Sub-Saharan Africa: Scoping Review

Alex Emilio Fischer[1,2*], MPH, PhD; Homaira Hanif[3], PhD; Jacob B Stocks[1,2], MSc; Aimee E Rochelle[1,2], MPH; Karen Dominguez[3], MPH; Eliana Gabriela Armora Langoni[4], MSPH; H Luz McNaughton Reyes[4], MPH, PhD; Gustavo F Doncel[3], MD, PhD; Kathryn E Muessig[1,2*], PhD

[1]College of Nursing, Florida State University, Tallahassee, FL, United States

[2]Institute on Digital Health and Innovation, Florida State University, Tallahassee, FL, United States

[3]CONRAD, Eastern Virginia Medical School, Macon & Joan Brock Virginia Health Sciences at Old Dominion University, Norfolk, VA, United States

[4]Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[*]these authors contributed equally

Corresponding Author:
Alex Emilio Fischer, MPH, PhD
College of Nursing
Florida State University
98 Varsity Way
Tallahassee, FL,
United States
Phone: 234 816 812 3216
Email: fischermhealth@gmail.com

Abstract

Background: In 2022, 3100 adolescent girls and young women in sub-Saharan Africa experienced new HIV infections each week. HIV pre-exposure prophylaxis (PrEP) is effective at preventing HIV but has limited uptake and persistence. Mobile health (mHealth) interventions can improve medication adherence; however, their utility to improve PrEP adherence among adolescent girls and young women is not well established.

Objective: This scoping review synthesizes evidence supporting mHealth for PrEP among adolescent girls and young women in sub-Saharan Africa and identifies strategies for further evaluation.

Methods: We searched PubMed and Google Scholar databases, expert referrals, and reference lists using the following eligibility criteria: (1) original research study or protocol; (2) English language; (3) publication between January 1, 2012, and August 31, 2023; (4) inclusion of adolescent girls and young women; (5) conducted in sub-Saharan Africa; and (6) use of mHealth tools to promote PrEP uptake, adherence, or persistence. Titles and abstracts were screened by 2 independent researchers. Full-text manuscripts were reviewed against all eligibility criteria to determine the final included studies. The characteristics and results of the included studies were abstracted and synthesized by mHealth tool type.

Results: The search identified 482 unique citations. Title and abstract review removed 380 citations primarily for not including adolescent girls and young women or being conducted outside sub-Saharan Africa. The remaining 102 articles underwent full-text review, yielding 31 eligible publications reporting on 21 unique studies. The most common mHealth tool was SMS text message (n=11), followed by app (n=9), telehealth (n=3), website (n=4), and video (n=1). Few publications evaluated effectiveness, and the results were mixed. One study found that SMS text message reminders improved PrEP adherence, and another concluded that SMS text message reminders did not show a significant impact. Two studies found that differentiated service delivery, which included mHealth components, improved PrEP uptake or persistence; however, the findings could not be attributed solely to the mHealth components. Lastly, 1 website was shown to improve PrEP persistence. Several earlier-stage studies focused on values and preferences toward mHealth without reporting the impact on PrEP.

Conclusions: We found few rigorously evaluated mHealth interventions for supporting PrEP among adolescent girls and young women, preventing the ability to draw conclusions on its effectiveness. Studies documented high usability and acceptability but limited assessment of the impact on health outcomes. Secondary uses of mHealth were found for data collection and components

XSL•FO
RenderX

of the standard of care. There is substantial room for growth in the innovative use of mHealth to support PrEP among adolescent girls and young women. Consideration of the strengths and limitations of mHealth tools in the local setting, review of past lessons learned, and intentional measurement of mHealth exposure and use could help advance this growing field.

(JMIR Mhealth Uhealth 2025;13:e60819) doi:10.2196/60819

KEYWORDS

HIV; pre-exposure prophylaxis; adolescent girls and young women; mHealth; app; intervention

Introduction

Background

Since 2010, new global HIV infections have decreased by 39%, but the occurrence of 1.3 million new infections in 2022 still exceeds the 2020 Joint United Nations Programme on HIV and AIDS (UNAIDS) target of 500,000 [1]. The decrease in new infections coincides with the global scale-up of preventative services in high-prevalence countries; however, more work is needed to reach the ambitious target of 370,000 new infections by 2025, especially among key populations [2,3]. Over half of all new HIV infections are in sub-Saharan Africa, and 62% of those are among adolescent girls and young women [1].

Adolescent girls and young women in sub-Saharan Africa experience factors across multiple levels that increase their likelihood of acquiring HIV, including biological, socioeconomic, cultural, and system-level factors [4,5]. Further, economic vulnerability in this population contributes to power imbalance in relationships, influences partnerships and condom use, and increases the likelihood of exposure to HIV through age-disparate relationships and transactional sex [4,6,7].

When HIV pre-exposure prophylaxis (PrEP) is taken consistently as a daily oral pill, it is effective at preventing HIV infections among men who have sex with men, serodiscordant couples, and people who inject drugs; however, significant differences in effectiveness exist across geographic locations and populations [8-10]. Women have had less robust prevention results, with inconsistent adherence and discontinuation contributing to these outcomes, alongside biological differences that require higher adherence to reach protective drug levels in vaginal tissue [9,11,12].

Although PrEP uptake among adolescent girls and young women in sub-Saharan Africa continues to increase, few countries have a PrEP-to-need ratio over 1 (having more individuals on PrEP than new HIV infections) [13]. Barriers to PrEP use (uptake, adherence, and persistence) for adolescent girls and young women in sub-Saharan Africa include side effects, stigma, gender-based violence, lack of social support, low risk perception, medical distrust, and the need to regularly access clinics for prescriptions and refills [14-17]. Health systems also face financial and capacity challenges with the introduction of PrEP into an already stressed system [18,19].

Mobile health (mHealth) tools have the potential to mitigate a number of individual-level barriers to PrEP use, without adding substantial strain to health care workers and systems [20,21]. A broad range of mHealth interventions have been shown to increase PrEP adherence among sexual and gender minority men in high-income countries [22,23] and to increase HIV antiretroviral therapy (ART) adherence among adolescent girls and young women in sub-Saharan Africa [24]. mHealth interventions could also be adopted to support women in sub-Saharan Africa to uptake, adhere to, and sustain PrEP, and a detailed understanding of the state of the field could guide the expansion of this work.

Previous Reviews

We identified 3 recent reviews on general PrEP promotion in low- and middle-income countries (LMICs) [18,19,25]. A 2021 synthesis by Irungu et al [18] (study publication years unspecified) reviewed multilevel interventions for PrEP use among adolescent girls and young women. A 2022 scoping review by Ekwunife et al [19] focused on LMICs and reviewed 11 publications (years 2010 to 2021) that investigated 6 interventions to increase PrEP uptake and continuation among adolescent girls and young women engaged in sex work or transactional sex. Lastly, a 2022 systematic review by Goldstein et al [25] (publication years 2000 to 2021) investigated the use of mHealth interventions for adolescents and young adults (irrespective of gender) in LMICs across the entire HIV cascade. Among 9 identified studies for PrEP adherence, 5 were among adolescent girls and young women in sub-Saharan Africa [25].

Collectively, these reviews identified 9 unique interventions conducted in sub-Saharan Africa that included adolescent girls and young women, and 6 of these had some mHealth component. Four interventions with mHealth tools for adolescent girls and young women in sub-Saharan Africa were common to all reviews: 3 involving SMS text message reminders (HPTN 082, Monitoring PrEP among Young Adult Women [MPYA], and Prevention Options for Women Evaluation Research [POWER]) and 1 involving a phone call to deliver drug-level feedback (3Ps [Perception, Partners, Pills] for Prevention). One review [18] also included an ongoing study comparing SMS text message reminders to WhatsApp support groups (PrEP Sequential Multiple Assignment Randomized Trial [SMART]) and 1 [25] included Mobile Solutions for Women's and Children's Health (mWACh), a 2-way SMS text message system for PrEP adherence and continuation support among postpartum women. Overall, the reviews found limited data on the effectiveness of current interventions (including mHealth tools) to increase PrEP uptake and adherence among adolescent girls and young women in LMICs [18,19], and the strength of evidence was limited as most publications were uncontrolled single-arm studies with small sample sizes [25]. Overall, mHealth tools were deemed acceptable and feasible. No studies associated the use of mHealth tools with improved PrEP uptake. Moreover, 1 intervention (POWER) suggested improvement in clinic visit attendance, 2 interventions (POWER and mWACh) found improvements in PrEP continuation, and none found

XSL•FO
RenderX

associations with PrEP adherence. The barriers identified were smartphone access issues, fatigue with SMS text messaging, and the cost of phone data [18].

Our review expands prior work by focusing on the specific geographic context of sub-Saharan Africa and examining mHealth intervention types and components for adolescent girls and young women in greater detail. Many of the studies identified in prior reviews were in pilot and acceptability phases or early trial implementation, and additional impact results and new formative studies have become available over the past few years.

To the best of our knowledge, no prior reviews have focused specifically on mHealth for PrEP among adolescent girls and young women in sub-Saharan Africa, which is the population most affected by the ongoing HIV epidemic. This article aims to address this gap through a scoping review of the literature. Scoping reviews are generally more exploratory than systematic reviews and are designed to synthesize the current evidence and gaps surrounding a particular topic. By describing the current landscape and body of evidence, this review aims to identify promising strategies and key gaps, which can direct future research toward the development and implementation of interventions and the evaluation of new mHealth modalities for supporting HIV prevention in adolescent girls and young women.

Methods

This scoping review followed a common methodological framework for scoping studies [26] and was modeled on the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [27]; however, no review protocol was registered or published.

The review included English language articles that featured one or more mHealth tools to facilitate HIV PrEP, published from January 1, 2012 (the first year a randomized controlled trial [RCT] showed that PrEP could effectively prevent HIV in heterosexual men and women in Africa [28]) until August 31, 2023, when the article search was completed. Articles were eligible if the study population (or a well-defined subpopulation) included women between 14 and 25 years of age in sub-Saharan Africa. Articles were ineligible if the digital tools were developed exclusively for use by health care workers (versus for the PrEP end user) or if the tool was used primarily for research or data collection (versus with the goal of improving end-user PrEP outcomes).

Searches for eligible articles were conducted on PubMed and Google Scholar, with the reference lists of all eligible literature subsequently searched to identify additional articles. Reviews, perspectives, and policy articles were not included as our aim was to synthesize completed and ongoing original research. Grey literature was searched via expert referrals and conference proceedings, so that formative research, pilot studies, and protocol publications could be included for a holistic review. Protocol articles were included as they describe original research that is planned or underway, to capture the directions the field is heading. This information is particularly relevant for encouraging innovative ideas to address current research gaps. While these publications do not directly provide findings on the usability or effectiveness of mHealth tools, they provide important information on the state of the science and future directions in the field.

Advanced search features were used to query the abovementioned databases for groups of keywords encompassing mHealth, adolescent girls and young women, and PrEP. SMS text message, text, mHealth, digital, WhatsApp, smartphone, app, telehealth, and multimedia messaging service (MMS) were used to denote mHealth; women, girls, and adolescents were used to denote adolescent girls and young women; and PrEP and pre-exposure prophylaxis were used to denote PrEP. All articles were recorded in a shared spreadsheet, and then, duplicate articles were removed. The remaining articles were independently screened at the title and abstract levels by 2 co-authors (KEM and AEF), and articles that did not meet the eligibility criteria cited above were subsequently removed. Two co-authors (HH and HLMR) reviewed the list of eligible articles after title and abstract review and were further consulted during the full text eligibility assessment. All co-authors reviewed and approved the final list of included articles. The complete manuscripts of the remaining articles were then assessed against the eligibility criteria. The articles that remained eligible were used in this scoping review.

Once all eligible articles were identified, the final charting table described each article with its citation (author, title, journal, and publication year), geographic location, population, study name, mHealth tool, phase (study type), sample size, behaviors targeted, study outcomes, and findings. Once all data were charted, an overview of the studies and characteristics was synthesized in writing, grouped by the type of mHealth tool used (SMS text message, app, telehealth, video, and other digital tools).

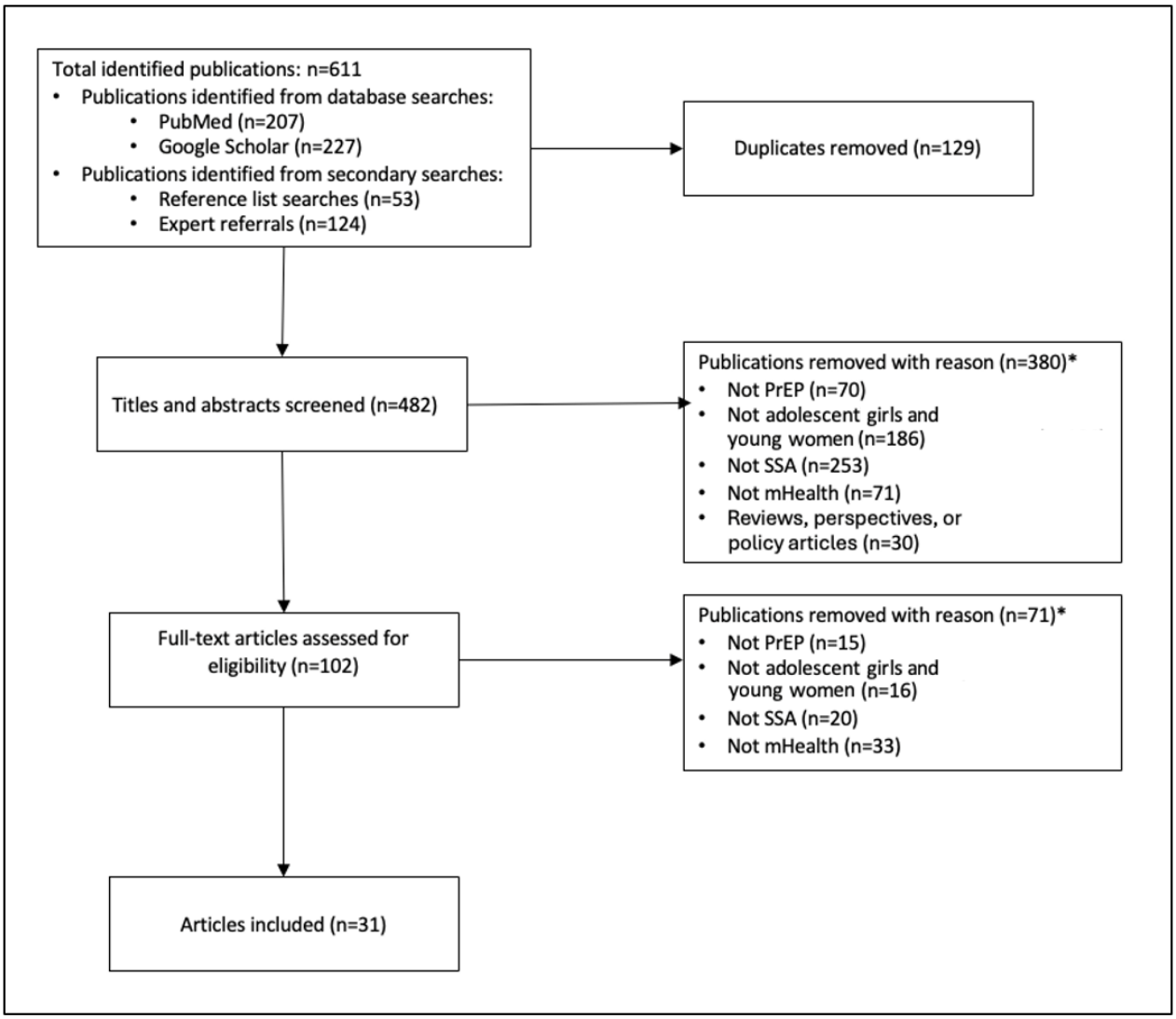
Results

Article Selection

A total of 611 publications were initially identified through database searches of PubMed and Google Scholar, as well as secondary searches through reference lists and expert referrals. Of these, 129 duplicate records were removed, and the remaining 482 publications underwent title and abstract screening, where 102 articles were selected for full-text screening against the eligibility criteria (Figure 1). Although not mutually exclusive, the most common reasons for exclusion across the 451 excluded articles were not including a study population involving adolescent girls and young women (n=202) and not conducting the study in sub-Saharan Africa (n=273). The remaining 31 articles met the inclusion criteria. They contained information on 21 unique mHealth interventions for PrEP and targeted adolescent girls and young women in sub-Saharan Africa. Several publications used SMS text messages or phone calls for data collection or reminders, as part of the standard of care; however, there was no theory of change associated with PrEP for these interventions, and they were therefore not included in this review.

XSL•FO
RenderX

Figure 1. Flow diagram of article inclusion. mHealth: mobile health; PrEP: pre-exposure prophylaxis; SSA: sub-Saharan Africa. *Excluded publications may have been removed for more than one reason.



Publication Characteristics

Table 1 presents the descriptive characteristics of the included studies. Of the 31 publications, 11 (35%) evaluated outcomes, with study designs that included RCT, pilot study (not powered to assess efficacy), pre/post evaluation, and descriptive study. Moreover, 13 publications (42%) presented formative results, including feasibility, acceptability, qualitative, and modeling studies, and 7 publications (23%) were protocols. For protocol publications, the anticipated characteristics have been included, where appropriate. No publications were published before 2018. Since 2018, there have been at least four publications per year, with 2023 having the highest number of publications (10/31, 32%). Most publications (27/31, 87%) reported sample sizes, which ranged from 50 to 7593; however, samples were generally small, with only 6 out of the 31 samples (19%) being above 500. There were 21 different interventions represented by the 31 publications, with 5 publications resulting from the MPYA study, and 2 each investigating the 3Ps for Prevention study; Determined, Resilient, Empowered, AIDS-free and Safe (DREAMS) study; HPTN 082 study; Jichunge study; mWACh study; and Safer Conception Intervention for Partners (SCIP) study.

XSL•FO
RenderX

Table 1. Publication characteristics.

Article, author and year	Study name	Study type	Country	Population	Sample size, n	mHealth[a] tool
Celum et al [29], 2020	3Ps for Prevention	Outcomes	South Africa	Adolescent girls and young women	200	Telehealth
Morton et al [30], 2020	3Ps for Prevention	Formative	South Africa	Adolescent girls and young women	320	Video
Hartmann et al [31], 2023	CHARISMA[b]	Formative	South Africa	Includes adolescent girls and young women	159	Website
Medina-Marino et al [32], 2021	Community PrEP[c]	Protocol	South Africa	Adolescent girls and young women	480	SMS text message, app
Lavoie et al [33], 2023	DREAMS[d]	Formative	Botswana	Adolescent girls and young women	131	App
Barnabee et al [34], 2023	DREAMS	Outcomes	Namibia	Adolescent girls and young women	7593	SMS text message, telehealth
Kiptinness et al [35], 2023	ePrEP	Protocol	Kenya	Includes adolescent girls and young women	500	Website
Celum et al [12], 2021	HPTN 082	Outcomes	South Africa, Zimbabwe	Adolescent girls and young women	451	SMS text message
Velloza et al [36], 2020	HPTN 082	Formative	South Africa, Zimbabwe	Adolescent girls and young women	67	SMS text message
Mauka et al [37], 2021	Jichunge	Formative	Tanzania	Female sex workers	60	App
Mbotwa et al [38], 2022	Jichunge	Formative	Tanzania	Female sex workers	470	App
Kreniske et al [39], 2023	Kirabo	Protocol	Uganda	Includes adolescent girls and young women	200	SMS text message
Busang et al [40], 2023	Let's Talk	Protocol	South Africa	Includes adolescent girls and young women	2800	SMS text message, telehealth
Moorhouse et al [41], 2019	Manicaland	Formative	Zimbabwe	Adolescent girls and young women	__[e]	SMS text message
Hanass-Hancock et al [42], 2023	Masibambane–Ladies Chat	Formative	South Africa	Adolescent girls and young women	100	App
Ngure et al [43], 2021	MPYA[f]	Formative	Kenya	Adolescent girls and young women	50	SMS text message
Haberer et al [44], 2021	MPYA	Outcomes	Kenya	Adolescent girls and young women	348	SMS text message
Haberer et al [45], 2021	MPYA	Outcomes	Kenya	Adolescent girls and young women	348	SMS text message
Haberer et al [16], 2022	MPYA	Outcomes	Kenya	Adolescent girls and young women	348	SMS text message
Ogello et al [46], 2023	MPYA	Formative	Kenya	Adolescent girls and young women	54	SMS text message
Cassidy et al [47], 2022	MSF[g] PrEP Pilot	Outcomes	South Africa	Adolescent girls and young women	164	App
Mogaka et al [48], 2023	mWACh[h]	Protocol	Kenya	Includes adolescent girls and young women	—	SMS text message
Pintye et al [49], 2020	mWACh	Outcomes	Kenya	Adolescent girls and young women	190	SMS text message
Celum et al [50], 2023	MyPrEP	Outcomes	South Africa	Adolescent girls and young women	386	Website
Thomas et al [51], 2019	NCT0356557	Protocol	Zimbabwe	Adolescent girls and young women	1200	App

XSL•FO
RenderX

Article, author and year	Study name	Study type	Country	Population	Sample size, n	mHealth[a] tool
Celum et al [52], 2022	POWER[i]	Outcomes	Kenya, South Africa	Adolescent girls and young women	2550	SMS text message, app, telehealth
Schwartz et al [53], 2018	PrEP Communications Accelerator	Formative	Kenya	Adolescent girls and young women	530	Website
Velloza et al [54], 2022	PrEP SMART[j]	Protocol	South Africa	Adolescent girls and young women	500	SMS text message, app
Heffron et al [55], 2019	SCIP[k]	Formative	Kenya	HSCs[l]	74 couples	SMS text message
Velloza et al [56], 2019	SCIP	Formative	Kenya	HSCs	74 couples	SMS text message, app
Matambanadzo et al [57], 2021	Sisters with a Voice	Outcomes	Zimbabwe	Female sex workers	—	WhatsApp, website

[a]mHealth: mobile health.

[b]CHARISMA: Community Health Clinic Model for Agency in Relationships and Safer Microbicide Adherence.

[c]PrEP: pre-exposure prophylaxis.

[d]DREAMS: Determined, Resilient, Empowered, AIDS-free and Safe.

[e]Not applicable.

[f]MPYA: Monitoring PrEP among Young Adult Women.

[g]MSF: Me´decins Sans Frontie` res.

[h]mWACh: Mobile Solutions for Women's and Children's Health.

[i]POWER: Prevention Options for Women Evaluation Research.

[j]SMART: Sequential Multiple Assignment Randomized Trial.

[k]SCIP: Safer Conception Intervention for Partners.

[l]HSC: heterosexual couples.

Regarding mHealth tools, the studies employed a total of 28 mHealth interventions, as 7 studies investigated more than one mHealth tool (Community PrEP [32], POWER [52], PrEP SMART [54], SCIP [55,56], Sisters with a Voice [57], 3Ps for Prevention [29,30], and DREAMS [33,34]). Of the 28 intervention tools, 11 (39%) included SMS text message tools, 9 (32%) included apps, 3 (11%) included telehealth, 4 (14%) included websites, and 1 (4%) included videos delivered through mobile devices.

Of the 21 unique intervention studies, 19 involved individual locations across 7 countries and 2 (HPTN 082 and POWER) involved locations in 2 countries. Of the 23 locations, 9 (39%) were in South Africa, 6 (26%) in Kenya, 4 (17%) in Zimbabwe, and 1 (4%) each in Uganda, Botswana, Namibia, and Tanzania. Adolescent girls and young women were the sole study population in 62% (13/21) of samples and were part of the population in 24% (5/21) of samples, while female sex workers and heterosexual serodiscordant couples (both populations include adolescent girls and young women) were the sole study population in 10% (2/21) of samples and were part of the population in 5% (1/21) of samples. The key findings for each type of mHealth tool are presented below. When an intervention employed more than one tool, the mHealth intervention component is discussed in its respective section.

SMS Text Message

Eighteen publications on SMS text message tools for PrEP in adolescent girls and young women reported findings from 11 different studies (Multimedia Appendix 1). The MPYA study involved an RCT of 348 adolescent girls and young women, where the intervention group received personalized daily 1-way SMS text message reminders for PrEP via an automated platform and the control group received no reminders [16,44]. Participants could opt for SMS text message reminders to be sent only after missed doses and could personalize the messages, choose message content, and update these choices during any clinic visit. In this study, adherence was measured via pharmacy refills and a real-time electronic monitor (Wisepill). SMS text message reminders were not associated with greater PrEP adherence over 24 months (47/173, 27.0% with SMS; 46/174, 26.6% without SMS) [16,44]. Among a subset of MPYA participants, in-depth exit interviews revealed that SMS text message reminders were initially highly acceptable, as they were perceived as helping form a daily adherence habit; however, over time, participants reported growing concerns about privacy and SMS text message fatigue [46]. Other barriers included phone loss, poor network connectivity, and lack of electricity [46]. Another qualitative study assessed the perceived risk of HIV and sexual behaviors among MPYA participants [43]. This publication suggested that HIV risk perception was dynamic among participants and may have influenced their PrEP use over time; however, it did not report additional findings related to participants' receipt of SMS text messages in the study [43].

The mWACh study in Kenya was a 2-group, quasiexperimental, pre-post evaluation of a 2-way SMS text message intervention with 190 pregnant adolescent girls and young women [49]. The

XSL•FO
RenderX

intervention was adapted from the existing mWACh 2-way SMS text message platform by introducing theory-based, PrEP-tailored weekly messages, which participants could respond to, and participants could ask questions, which would be replied to by nurses [49]. Compared to participants who did not receive the intervention, participants who received the SMS text message intervention were more likely to return for their first PrEP follow-up (101/190, 53.2% vs 67/166, 40.4%) and to have high PrEP adherence (<1 missed pill per week) at 1 month after starting PrEP and receiving the SMS text messages (82/190, 43.2% vs 37/166, 22.3%) [49]. Promising results from the pre-post evaluation led to an RCT study to evaluate the mWACh SMS text message intervention. Enrollment for the RCT began in March 2022, and the results are expected in 2025 [48].

In Namibia, 7593 adolescent girls and young women participated in DREAMS and received PrEP services via 3 models, based on geographic location: *community-concierge*, *community-fixed*, and *hybrid community-clinic* models [34]. All 3 models had the same PrEP education, screening, and initiation, but differed in follow-up, monitoring, and support. In the *hybrid community-clinic* model, adolescent girls and young women were referred to their public health facility for PrEP refills and clinical monitoring, while the *community-fixed* model had nurses provide refills, monitoring, and follow-up services via scheduled visits at fixed community-based locations. The *community-concierge* model provided refills and monitoring at community-based locations, tailored to participants' real-time preferences, and offered mHealth support that included visit and refill reminders via SMS text messages and a nurse-managed phone that adolescent girls and young women could message with questions. During the study, 2047 adolescent girls and young women were expected to receive a PrEP refill, but only 254 out of 2047 (12.4%) persisted with PrEP for at least 1 month. Adolescent girls and young women in the *community-concierge* model were 8.7 times (95% CI 5.44-13.9) more likely to persist compared to the *hybrid community-clinic* model; however, there were no specific findings related to the uptake or effect of the SMS text message components of the *community-concierge* model [34].

The SCIP study with 74 heterosexual serodiscordant couples (148 participants) from Kenya evaluated a tablet-based counseling app that provided HIV prevention counseling for the HIV-negative partner, adherence counseling for the partner living with HIV, and information about peak fertility timing [55]. Content for the app was informed by results from an SMS text message survey that collected fertility, PrEP, and ART adherence data. An uptake and usability pilot found that 78.0% (13,181/16,905) of daily surveys were fully answered, and 81.0% (68/84) of HIV negative partners had high PrEP adherence in the months preceding pregnancy [55]. An accompanying feasibility and acceptability study confirmed high SMS text message response rates; however, there was no evaluation of the SMS text message survey's ability to influence PrEP, and findings from the tablet-based counseling app [56] will be discussed in the App section.

For HPTN 082 in South Africa and Zimbabwe, the intervention being tested involved enhanced adherence support with feedback on tenofovir levels from dried blood spots taken twice during the first 3 months on PrEP. The control condition included 2-way SMS text messages as part of the standard adherence support, along with in-person counseling and in-person peer adherence clubs. From the RCT with 451 adolescent girls and young women [12], at month 6, there was no statistical difference in PrEP adherence between the intervention and control arms (38/179, 21.2% vs 40/184, 21.7%; *P*=.76), and the authors speculated that this may have been due to a strong effect from the standard support measures (including SMS text messaging), which may have been comparable to the benefits of drug-level feedback. During the qualitative study [36], adolescent girls and young women identified stigma and disclosure as initial concerns, but peer adherence clubs empowered many to disclose PrEP and open up about stigma. The adherence clubs, however, were not feasible for all to attend, so participants suggested offering adherence club content digitally (via WhatsApp) to reduce barriers associated with in-person clubs.

The POWER study was a prospective implementation study of PrEP uptake and continuation among 2550 adolescent girls and young women from Kenya and South Africa [52]. Participants received HIV prevention counseling and were offered PrEP. Among the cohort, 94.0% (2397/2550) initiated PrEP, with 31.3% (749/2397) refilling PrEP at the 1-month follow-up. SMS text messages, WhatsApp messages, and phone calls were used to remind participants of quarterly follow-up visits for up to 36 months. Among the participants who were reachable for the 6-month follow-up, 19.8% (128/646) had continued PrEP use, and the authors suggested that the lack of travel and study participation reimbursement may have contributed to this decline [52]. No evaluation of mHealth components was reported.

Upcoming research includes the Manicaland RCT study in Zimbabwe, a 6-month intervention delivering 1-way, tailored, personalized SMS text messages that aim to "nudge" participants toward PrEP adherence by increasing adherence self-efficacy and awareness of HIV risk [41]. Another protocol describes an ongoing study, PrEP SMART, with adolescent girls and young women in South Africa that initially randomizes participants to 1 of 2 mHealth-enhanced interventions to improve PrEP adherence: standard-of-care PrEP counseling complemented by weekly 2-way SMS text messages (modeled after HPTN 082) and standard-of-care PrEP counseling accompanied by WhatsApp adherence groups (described below) [54]. At month 3, nonresponders will continue to receive their original intervention plus either drug-level feedback counseling or intensive problem-solving counseling (no additional mHealth components at rerandomization) [54]. Another protocol describes the Community PrEP Study, which is ongoing in Eastern Cape, South Africa, with a 3-armed RCT comparing the effectiveness for PrEP adherence of the following: in-person health clubs plus WhatsApp adherence groups, one-to-one in-person counseling sessions, and a control condition that includes SMS text message and phone call reminders for study visits [32].

Another protocol is for the Kirabo study in Uganda, where the information, motivation, and behavior model is used to develop a 2-way SMS intervention to increase HIV testing and link users

XSL•FO
RenderX

to mental health counseling [39]. The RCT, with a target of 200 youth (including adolescent girls and young women), began in August 2023, and a secondary outcome of the study is the intervention's impact on PrEP adoption [39]. The final protocol is for Let's Talk in South Africa, which is a 2-armed, stepped-wedge trial involving 2800 participants, with a primary outcome of PrEP uptake for HIV-negative participants [40]. The control arm will receive the current standard of care, while the intervention arm will receive social mobilization and tailored support, provided by peer navigators and adolescent and young adult–friendly nurse-led mobile clinics. All participants who start PrEP in the intervention cluster are offered peer-navigator support as well as neutral text message reminders. The 45-month study is ongoing, and data collection is expected to be complete in 2024 [40].

App

Twelve publications described 9 different app-based interventions across 10 different studies, including 2 publications on Jichunge, a smartphone app for iOS and Android aimed at increasing PrEP adherence among female sex workers and men who have sex with men in Tanzania (Multimedia Appendix 2). One publication described the participatory design approach to the Jichunge app [37]. An information system research framework was employed so that participants from the target demographics (including 15 female sex workers in the design phase and 20 female sex workers in the testing phase) could inform app content and the accompanying web-based administrative portal. After the design and testing cycles, the final prototype included the following functionalities to promote PrEP adherence: gamification (awarding levels and points for completing tasks in the app), drug registration on the in-app medication tracker, medication time reminders, notifications, communication with peers, communication with health care workers, discussion forum, educational material, and quizzes [37]. Jichunge was then evaluated in a quasiexperimental study with 470 female sex workers, where end-user data from the app and qualitative interviews were used to evaluate "optimal" app use (defined as using 3 app functions in the first 30 days) [38]. Overall, 46.0% (216/470) of women met this definition of optimal app use, with substantially higher use among women over 35 years of age and those with prior social media experience [38]. The effectiveness of the app on adherence or health outcomes was not assessed.

The Me´decins Sans Frontie`res (MSF) PrEP Pilot in South Africa was an observational cohort of 224 adolescent girls and young women, which investigated PrEP uptake and adherence with an intervention that was guided by motivational interviews and the "stages of change" theory [47]. The intervention included in-person counseling sessions (covering PrEP use, HIV risk assessment and reduction, and sexual and reproductive health information); "PrEP Diva" gatherings, which replaced traditional "sexual and reproductive health clubs;" and a WhatsApp group where participants could engage with study staff and each other via text messages. PrEP was initiated by 73.2% (164/224) of participants. High adherence was defined as ≥700 fmol/punch (the equivalent of taking at least four pills per week), and at the 1-month follow-up, 17.7% (29/164) had high adherence; however, this dropped to 12.2% (20/164) at month 18. Forgetting to take or travel with pills was the main barrier to adherence, and over the final 12 months, encouragement from others declined as a facilitator of adherence [47].

Sisters with a Voice was an observational study involving 6539 female sex workers in Zimbabwe that investigated PrEP uptake before and after the scale-up of differentiated service delivery (DSD) models in response to COVID-19 lockdowns [57]. The DSD models included digital components like online appointment scheduling, WhatsApp, and telehealth, as well as nondigital components like peer demand generation to increase uptake, community-based delivery, and multi-month medication dispensing. For WhatsApp, this platform allowed continuous virtual engagement from peer educators, which enabled users to report adverse effects and receive adherence support. From January to March 2020, there was no DSD, and PrEP uptake rates were less than 25.0% (315 initiations or fewer out of 1260) per month. DSD was scaled up in April 2020, and in May 2020, PrEP uptake began increasing each month, peaking at 51.0% (1360 initiations out of 2667) in September 2020. PrEP uptake data were not disaggregated by use of or exposure to the various DSD components, limiting the ability to draw conclusions regarding the impact of the digital health tools [57].

The development and pilot testing of Masibambane (Let us Unite)–Ladies Chat was based on the information, motivation, and behavioral skills model; social learning theory; and gender theory to promote PrEP uptake and knowledge among adolescent girls and young women in South Africa. A publication describes the formative development and pilot testing of the intervention, which was designed based on 9 existing group-based activities (traditionally performed in person) and adapted to be delivered using a hybrid approach primarily on WhatsApp, which included asynchronous chat and live sessions [42]. Lessons learned from the pilot included a need to work around the impact of technology access (availability of free data at night and electricity outages ["load shedding"]), enhancing participant engagement through staff modeling engagement and by using mixed communication styles (verbal, written, and emoji), and introducing breaks during long sessions. No further evaluation or outcome information was provided.

While the abovementioned DREAMS program in Namibia evaluated DSD models, the DREAMS Botswana program was a separate, formative study that assessed the feasibility and acceptability of providing peer-delivered mHealth interventions (eg, WhatsApp groups) to support PrEP use among 131 adolescent girls and young women [33]. Feasibility was high, as 89.3% (117/131) of participants owned a mobile phone and 96.9% (127/131) used WhatsApp in the last year. Moreover, 84.7% (111/131) stated they would be interested in a peer-delivered mHealth intervention, 78.6% (103/131) preferred SMS text message or phone call reminders for PrEP refills, and the average acceptability of the intervention measure was 3.8 out of 5 [33].

The SCIP study, which has been introduced above, used a tablet-based Android app for counseling on fertility timing, PrEP, and ART adherence to complement the SMS text message

XSL•FO
RenderX

survey [55]. The feasibility and acceptability study confirmed that the SMS text message and MEMS Cap data could be successfully integrated into the tablet-based app, with SMS text message data integrated via an mSurvey application programming interface (Kenyan mobile data collection company) and the MEMS Cap data downloaded directly to the tablet during clinic visits. The app was not available to participants on their personal smartphones; however, the tablet-based app was available during clinic visits to complement in-person counseling on fertility timing, PrEP, and ART adherence. No evaluation was presented of the app's ability to influence PrEP adherence or maintenance [56].

In the POWER study, which has been introduced above, WhatsApp was offered as a platform to remind participants of quarterly follow-up visits for up to 36 months, similar to the SMS text message intervention. No further description of the WhatsApp intervention or evaluation of the mHealth components was reported [52].

The remaining 3 studies on apps have been published as study protocols for RCTs (no results available to date). The PrEP SMART study, which has been described above, includes WhatsApp adherence groups with 25-50 participants each to facilitate social and emotional support [54]. The adherence groups are facilitated by a staff member and a PrEP-experienced peer ("PrEP ambassador") and include educational and fun activities. The Community PrEP Study, which has been described above, includes 1 intervention arm that consists of health clubs with 8 to 10 participants who attend in-person group counseling sessions and have access to a club-specific WhatsApp group [32]. The third protocol is for an RCT in Zimbabwe with 1200 adolescent girls and young women (NCT03565575) from July 2018 through August 2020 [51]. This pilot RCT evaluated the ability of interactive digital tablet-based quizzes and counseling sessions to increase PrEP uptake compared to a control arm with no intervention. Though data collection ended in 2020, no published findings were identified during our literature search, and no information was obtained after attempting to contact the corresponding author.

Telehealth

Three studies included telehealth-based intervention components (Multimedia Appendix 3). The Sisters with a Voice observational study of DSD models to increase PrEP uptake, presented in the App section, had additional components, including telehealth and websites for appointment scheduling. The telehealth scale-up included the ability for participants to report adverse results by phone and to receive telehealth follow-ups for side effects and adherence counseling. As noted above, while the overall intervention package increased PrEP uptake, the effectiveness of or exposure to individual components was not evaluated [57].

As described in the SMS Text Message section, participants in DREAMS Namibia received PrEP services via 3 models: *community-concierge*, *community-fixed*, and *hybrid community-clinic* models. In the *community-concierge* model, which was 8.7 times more likely to experience PrEP persistence (than the hybrid model), the mHealth components (visit and refill reminders, and a nurse-managed phone with questions) could also be accessed by phone; however, no specific findings were reported related to the uptake or effect of the telehealth components [34]. The Let's Talk protocol, which has been described above, included additional mHealth components beyond the neutral text reminders [40]. Participants in the peer-navigator intervention group were also offered the choice to receive reminders by phone, instead of text, and a nurse-managed phone hotline was available 24/7 for participants seeking psychological support or clinical guidance. Three additional studies that had other mHealth components (3Ps for Prevention [29], POWER [52], and Community PrEP Study [32]) also used phone calls for clinic or study visit reminders; however, this activity was not evaluated as a PrEP-focused mHealth intervention.

Website

Four studies investigated a website or mobisite as part of the intervention, including MyPrEP, a client-facing mobisite with a decision support tool designed to improve users' risk perceptions, PrEP knowledge, informed decision-making, and motivation, in order to facilitate PrEP delivery (Multimedia Appendix 3). The study involved an RCT conducted in South Africa with 386 adolescent girls and young women, where the control group was directed to a general health website, while the intervention group was directed to the MyPrEP decision support tool mobisite. The decision support tool, which was originally developed for an electronic tablet and then reformatted for smartphone web access, presented available HIV prevention options in an impartial manner, supporting users to make informed decisions, including whether to initiate PrEP. PrEP initiation was similar across both groups, with 97.1% (167/172) initiating in the MyPrEP group and 93.9% (170/181) in the control group (odds ratio 1.79, 95% CI 0.79-1.53); however, PrEP persistence at 1 month (determined by a documented PrEP refill) was 19.2% (32/167) in the MyPrEP group and only 10.0% (17/170) in the control group (odds ratio 1.97, 95% CI 1.08-3.69) [50].

One formative study, PrEP Communications Accelerator, described the development of a website [58] to help programs generate demand for PrEP in sub-Saharan Africa by providing a customized communications plan for targeted audiences (including adolescent girls and young women and female sex workers) and according to the health care setting. Each tailored demand generation strategy suggested evidence-based media channels and communication tactics, but the outputs have not yet been evaluated for usability, acceptability, or effectiveness [53].

Another formative study investigated CHARISMA (Community Health Clinic Model for Agency in Relationships and Safer Microbicide Adherence), a mobisite designed to mitigate relationship barriers to PrEP, like intimate partner violence [31]. By using human-centered design and workshops, the mobisite was adapted from existing counselor-administered HIV prevention and relationship counseling interventions. The mobisite was further refined by beta-testing with 81 women, followed by a real-world feasibility and acceptability pilot with 159 women. In the pilot test, adolescent girls and young women rated the website 4.5 stars out of 5, found the website useful

XSL•FO
RenderX

(4.5 out of 5) and safe (4.5 out of 5), and had minimal confidentiality concerns (1.8 out of 5). Based on these findings, the authors suggested that future research should evaluate health impacts, including PrEP uptake and continuation; however, no supplementary protocols were identified [31].

The ePrEP study involves a protocol for an upcoming single-arm pilot in Kenya that will deliver PrEP via a private online pharmacy retailer, MYDAWA [35]. Potential participants can order an HIV self-test from MYDAWA (250 KES [approximately US $2]) and have it delivered to a location of their choice. They can then upload a picture of their self-test results and attend a telemedicine consult with a MYDAWA clinician to screen and receive counseling for PrEP. Eligible participants will be prescribed a free 30-day PrEP supply, and it will be delivered for 99 KES (approximately US $1). The study will measure PrEP initiation and continuation, postexposure prophylaxis-to-PrEP transition, and implementation outcomes like feasibility, acceptability, and costs. The study is anticipating at least 500 participants over 18 months; however, no expected date of completion has been provided [35].

Video

An earlier phase of the previously described 3Ps for Prevention study investigated demand generation, which included a qualitative component that examined video and information brochures [30]. A 90-second video was cast with local youth and shot by a professional filmmaker, based on a script created by a marketing firm that conducted focus group discussions and interviews with local adolescent girls and young women on PrEP. Participants were recruited in their own homes by study staff going door-to-door and were shown the video if they agreed to participate. A total of 320 adolescent girls and young women watched the PrEP demand-generation video, and among 300 replies, most (220/300, 73.3%) said they wanted to learn more about PrEP after watching the video. However, without a control group that did not watch the video, this study design does not contextualize the results. PrEP interest was found to be associated with having sex with a primary partner in the last 30 days and being in a sexual partnership for over 6 months [30] (Multimedia Appendix 3).

Summary of Key Findings

Of the 18 publications on SMS text messages, there were only 2 impact evaluations on the ability of SMS text message interventions to increase PrEP adherence, and they presented inconsistent findings. The mWACh 2-group, quasiexperimental, pre-post evaluation found that SMS text message interventions increased high PrEP adherence after 1 month [49], while the MPYA RCT found that SMS text message reminders did not increase PrEP adherence over the control condition after 24 months [16,44]. A third descriptive study, DREAMS Namibia, evaluated 3 different service models and found that the *community-concierge* model, which included SMS text message reminders and a nurse-operated phone line that could receive SMS text message questions, had 8.7 times higher PrEP persistence at 1 month than the hybrid model [34]. User preference for SMS text messages was noted in the 2 Kenyan studies that reported high SMS survey completion rates; however, there were 2 studies that reported low completion rates. Among the studies that reported high completion rates, 2 reported high SMS text message acceptability and 1 indicated that SMS text message reminders might be helpful in forming daily adherence habits. The anonymity of SMS text messages may reduce social desirability bias, and SMS text messaging as a data collection tool introduces the ability to automate data cleaning and analysis. Three studies reported barriers to SMS text messaging, which included privacy, SMS text message fatigue, and lack of a phone, network, or electricity (Table 2).

For apps, there were no explicit impact evaluations; however, WhatsApp-based interventions were associated with high PrEP uptake in the MSF PrEP pilot [47], and WhatsApp was included as one of the interventions that was scaled up in the Sisters with a Voice study that doubled PrEP uptake in 5 months [57]. In 1 study, 96.9% (127/131) of participants had used WhatsApp in the past year [33], and WhatsApp was also identified as the most accessible platform in another study [42] and suggested as a platform for digital adherence clubs in another, although with some privacy concerns [36]. Although WhatsApp can be used as a text messaging service similar to SMS text messaging, it operates on data, not airtime, which makes it more affordable. Furthermore, it has strong privacy protection through encryption and offers the ability to participate in group chats. Some participants suggested a hybrid approach with asynchronous chat and live WhatsApp sessions that used mixed communication styles (verbal, written, and emojis). The main barriers to app use were data costs and electricity outages.

This review identified a small subset of literature regarding telehealth interventions, websites, and the use of videos to support PrEP for adolescent girls and young women in sub-Saharan Africa. Outside of the forthcoming ePrEP and Let's Talk studies, the data are very limited. Telehealth (for appointment scheduling, reporting results, and follow-up consultation) was one of the interventions that was scaled-up in the Sisters with a Voice study and was a component of the *community-concierge* model (phone reminders and a nurse-operated phone line that participants could call with questions or concerns) of the DREAMS Namibia study; however, the effects of the individual mHealth components were not presented in either of these studies. Phone calls were also used in several studies as part of the standard of care for appointment reminders and data collection; however, these were not hypothesized or evaluated as interventions to improve PrEP use. Participants reported barriers associated with phone ownership and data costs.

The MyPrEP RCT found that participants who used the DCT mobisite, designed to improve risk perception and support informed decision making, were twice as likely to persist with PrEP at 1 month than the control group. No other studies evaluated the effectiveness of websites; however, the CHARISMA study did show that a mobisite to mitigate relationship barriers associated with PrEP use had high feasibility and acceptability among adolescent girls and young women. For videos, there was no empirical evidence to support their effectiveness at changing PrEP uptake; however, 73% of participants who watched the 3Ps for Prevention video requested more PrEP information after viewing the video.

XSL•FO
RenderX

Table 2. Summary of key findings.

mHealth[a] intervention	Facilitators	Barriers	Evidence for mHealth	Evidence against mHealth[b]
SMS text message	• SMS text message reminders were initially highly acceptable and helped form a daily adherence habit (MPYA[c]) [16,44] • SMS text message surveys had a high completion rate (MPYA and SCIP[d]) [16,44]	• Privacy, SMS text message fatigue, phone loss, poor network connectivity, and lack of electricity (MPYA) [16,44]	• Pre-post evaluation found that the 2-way SMS text message intervention group was more likely to continue PrEP[e] than the group without SMS text message (mWACh[f]) [49] • Descriptive study found that a model including SMS text message reminders and asking nurses questions via SMS text messages had an 8.7× higher PrEP persistence at 1 month (DREAMS[g] Namibia) [34]	• RCT[h] found that 1-way SMS text message reminders were not effective in promoting PrEP adherence (MPYA) [16,44]
App	• Digital adherence clubs (WhatsApp) were suggested to reduce barriers of in-person clubs (HPTN 082) [36] • WhatsApp as the most accessible platform (Masibambane–Ladies Chat) [42] • 89.3% (117/131) owned a smartphone, 96.9% (127/131) had used WhatsApp in the last year, and 84.7% (111/131) were interested in peer-delivered mHealth (DREAMS Botswana) [33] • SMS text message and MEMS Cap data integrated into a tablet-based app (SCIP) [55]	• Must work around the impact of technology access like availability of free data at night and electricity outages (load shedding) (Masibambane–Ladies Chat) [42]	• WhatsApp included in scaled-up DSD[i] that doubled PrEP uptake in 5 months (Sisters with a Voice) [57] • WhatsApp groups associated with high PrEP uptake of 73.2% (164/224) (MSF[j] PrEP pilot) [47] • Optimal app use was 46.0% (216/470) in a pilot (Jichunga) [38]	—[k]
Telehealth	• Phone calls for data collection (3Ps for Prevention study) [29]	—	• Telehealth included in scaled-up DSD that doubled PrEP uptake in 5 months (Sisters with a Voice) [57] • Descriptive study found that a model including phone reminders and asking nurses questions via phone had an 8.7× higher PrEP persistence at 1 month (DREAMS Namibia) [34]	—
Website	• Hybrid approach with a chat box and live sessions preferred (Masibambane–Ladies Chat) [42] • Mixed communication styles (verbal, written, and emoji) preferred (Masibambane–Ladies Chat) [42] • Pilot study of mobisite to mitigate relationship barriers to PrEP showed high feasibility and acceptability (CHARISMA[l]) [31]	—	• RCT found that the mobisite decision support tool group was twice as likely to persist with PrEP at 1 month than the control group (MyPrEP) [50]	—
Video	—	—	• After watching a video, 73.3% (220/300) wanted more PrEP info (3Ps for Prevention study) [29]	—

XSL•FO
RenderX

[a]mHealth: mobile health.

[b]No clear controlled evidence against mHealth was found in the search.

[c]MPYA: Monitoring PrEP among Young Adult Women.

[d]SCIP: Safer Conception Intervention for Partners.

[e]PrEP: pre-exposure prophylaxis.

[f]mWACh: Mobile Solutions for Women's and Children's Health.

[g]DREAMS: Determined, Resilient, Empowered, AIDS-free, Mentored and Safe.

[h]RCT: randomized controlled trial.

[i]DSD: differentiated service delivery.

[j]MSF: Me´decins Sans Frontie`res.

[k]Not available or not applicable.

[l]CHARISMA: Community Health Clinic Model for Agency in Relationships and Safer Microbicide Adherence.

Discussion

Overview

This scoping review found few robust evaluations of mHealth tools that successfully supported PrEP use (uptake, adherence, and persistence) among adolescent girls and young women in sub-Saharan Africa. While previous reviews mainly synthesized information on SMS text message interventions across 6 studies (HPTN 082, 3Ps for Prevention, POWER, PrEP SMART, MPYA, and mWACh) [18,19,25], this review expanded the sample to 21 studies, which included apps, telehealth, videos, websites, and additional SMS text message interventions. The studies presented in this review showed generally high usability and acceptability of mHealth interventions. Additional rigorous study designs and evaluations are needed to measure engagement with mHealth interventions and assess the effectiveness of mHealth interventions to increase PrEP uptake, adherence, and sustained use over time.

Use of mHealth Tools for Research, Clinical Care, and Program Implementation

Despite the limited conclusive evidence from intervention trials supporting mHealth interventions for PrEP adherence, prevention programs for adolescent girls and young women in sub-Saharan Africa are still adopting and benefiting from these interventions to support program implementation. Numerous studies not included in this review demonstrated the feasibility and acceptability of mHealth tools for the collection of PrEP adherence and sexual behavior data among adolescent girls and young women in these settings [29,55,59-62]. In some studies, mHealth tools were already employed as part of the standard of care or DSD, including 2-way SMS text message adherence and appointment reminders, online appointment scheduling, and telehealth options [12,29,32,57].

In 2023, the World Bank published *Framework for Economic Evaluation of Digital Health Interventions*, which recommends that economic evaluations of digital health interventions include an impact inventory to capture secondary benefits in addition to their anticipated health outcomes [63]. These benefits can include improved user experience, improved health care worker experience, improvements to the health system, value of data generated, and enhanced interoperability across systems. mHealth interventions from this review revealed a number of these secondary benefits that could be captured in an impact inventory, such as high usability and acceptance, the ability to capture data with less social desirability bias [61], and integration into other systems [56].

Adapting Promising mHealth Approaches Across Geographic Settings and Populations

During the literature search, only 31 publications met the inclusion criteria, while 202 were excluded for predominantly investigating men who have sex with men in various countries, including some in sub-Saharan Africa. This uneven distribution mirrors the general promotion of PrEP in the global North and among men who have sex with men [64]. Indeed, prior reviews of mHealth interventions for adolescent girls and young women in LMICs found none conducted outside of sub-Saharan Africa [18,19,25]. Our additional searches identified a number of studies for adolescent girls and young women in the United States and 1 in Brazil; however, almost all reported on formative phases [65-67].

The larger body of published research, including impact evaluations, among sexual and gender minority men has begun to show the effectiveness of mHealth interventions on PrEP for men who have sex with men [20,21]. Lessons learned from these settings could inform mHealth intervention adaptations to accelerate the process from development to implementation. In 2019, an RCT in Chicago, United States showed that a 2-way SMS intervention increased PrEP adherence by 15% over 36 weeks [23], and in 2021, a pilot study in Boston, United States revealed that a culturally and developmentally tailored PrEP adherence app nearly doubled perfect adherence rates after 6 weeks of app use, compared to a control [22]. Although all participants were men who have sex with men, the information, motivation, and behavioral skills theory of behavior change behind the medication reminders and supportive messages designed to improve PrEP knowledge and adherence may also resonate with adolescent girls and young women in sub-Saharan Africa [42]. Tailored mHealth interventions have the potential to be effective across different demographics, especially since both men who have sex with men in the United States and adolescent girls and young women in sub-Saharan Africa have expressed similar barriers to PrEP, including fear of stigma and discrimination, privacy, and long clinic wait times [25,68].

XSL•FO
RenderX

Considerations for Ongoing Development of mHealth PrEP Tools for Adolescent Girls and Young Women

Mobile penetration continues to rise in sub-Saharan Africa, and delivering information via these platforms may involve lower costs as compared to other modalities like facility-based counseling or mobile outreach clinics. In 2021, there were over 500 million unique mobile subscribers in the region, and that number is expected to reach 613 million by 2025, as an increasing number of tech-savvy youth reach the age where they can begin to purchase mobile subscriptions [69]. The COVID-19 pandemic caused a global shift toward the acceptance of and reliance on digital interventions, including those for health care [70,71]. Furthermore, in 2022, the WHO and UNAIDS released a policy brief titled "Virtual interventions in response to HIV, sexually transmitted infections and viral hepatitis," which suggested that virtual interventions should be introduced to complement existing processes by streamlining the exchange of information and improving access to care [72].

User preferences from this review include simple wording in a variety of languages as part of a hybrid mHealth approach with an affinity for WhatsApp, which is in line with the incorporation of mHealth tools as a component of decentralized DSD [73]. Implementers must be aware of use barriers, such as SMS text message fatigue, lack of data, privacy concerns, unstable network coverage, and power outages [42,46]. A hybrid approach with curated content may help reduce SMS text message fatigue, while data vouchers can be offered to supplement mHealth use. Furthermore, the widespread popularity of WhatsApp likely reflects the many network providers in sub-Saharan Africa that offer low-cost or free unlimited WhatsApp use with most plans, and the built-in encryption may address privacy concerns [74].

Knowledge Gaps and Future Directions

There are knowledge gaps that still exist, including the lack of well-controlled studies determining the effectiveness of mHealth tools for increasing uptake, adherence, and continuation of oral PrEP regimens in adolescent girls and young women living in sub-Saharan Africa. Additional gaps include the lack of "best practice" approaches to translate evidence-based interventions into mHealth delivery and the inconsistent use of behavioral theory of change frameworks. If future methodologies begin to address these gaps, future results should provide a deeper understanding of how variations in impact may be attributed to differences in the types of mHealth tools used, the content of intervention messages, and the broader implementation science approaches employed. As new models of PrEP delivery, such as next-generation long-acting PrEP [75] and pharmacy distribution [76], expand and the landscape shifts toward decentralized service delivery, mHealth interventions should continue to support their implementation and improve adherence. Future research should ideally be designed with consideration of implementation science methodologies and frameworks to ensure that the effectiveness of mHealth interventions for PrEP adherence is being accurately captured while also measuring economic impact and auxiliary benefits [19,63]. In a recent review of the implementation of mHealth interventions for HIV or sexually transmitted infection prevention in LMICs, researchers applied the taxonomy of implementation outcomes by Proctor et al [77] (acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability) and found few studies that measured these outcomes with a bias focused on acceptability, appropriateness, and feasibility and limited to no measurement of adoption, fidelity, cost, penetration, and sustainability [78]. Moreover, even fewer studies have used or reported on implementation strategies (eg, audit and feedback, identify and prepare champions, and conduct ongoing training [79]) for enhancing how digital HIV prevention and care interventions are delivered and supported (J Brasileiro, unpublished data, November 2024). Without more systematic application and measurement of implementation strategies and outcomes, the field will continue to struggle in the interpretation of mixed and conflicting results from mHealth PrEP interventions.

Forthcoming interventions and studies should also be cognizant of barriers to mHealth among adolescent girls and young women in sub-Saharan Africa, including privacy, SMS text message fatigue, shared phone ownership, network and electricity interruptions, and data costs. Greater understanding of medical distrust and other perceptions of PrEP among adolescent girls and young women in this context should also be taken into consideration.

Limitations

This scoping review had several limitations. Although this review focused on adolescent girls and young women, publications on heterosexual couples and female sex workers (adolescent girls and young women make up a significant proportion of each population) were also included, although these populations may have slightly different drivers toward PrEP than adolescent girls and young women independently. Our literature search identified a number of mHealth interventions, such as B-wise [80], a website with PrEP information for adolescent girls and young women in South Africa, and Tutu Tester [81], a mobile clinic that used mHealth to connect with HIV testers in the community [82,83]; however, these interventions were not included, as their impacts from mHealth were not published in peer-reviewed journals or as conference abstracts. A lack of robust evaluation of mHealth interventions reflects the current funding environment in sub-Saharan Africa, which is more focused on implementation and demonstration projects, and the limitation of the field of digital health, with many tools struggling to get past development and pilot testing to large scale trials owing to the nature of the technology sector and incompatible health research funding models [84]. Several studies in this review evaluated the values and preferences of mHealth or used mHealth as part of the implementation strategy (data collection, standard of care, or DSD), and very few studies used impact evaluations like RCTs or pre-post evaluations. Furthermore, most of the studies that evaluated impact were pilot studies or had small samples that had not yet been scaled up. With quasiexperimental designs and underpowered samples, it is unclear whether insignificant findings were due to ineffective mHealth interventions or weak implementation strategies.

XSL•FO
RenderX

Conclusion

Findings from this scoping review suggest that, to date, there is insufficient evidence supporting the effectiveness of mHealth interventions for increasing PrEP uptake, adherence, and sustained use among adolescent girls and young women in sub-Saharan Africa, despite high usability and acceptability across different platforms. Despite the lack of an evidence base, several studies and interventions have already adopted mHealth tools for data collection or as part of their standard of care, suggesting that they may be providing secondary benefits. Furthermore, digital technologies have become more accessible, and their use by young people has grown exponentially. Implementation science frameworks and methodologies, such as RCTs with statistically significant sample sizes, should be considered in future research to accurately capture the effectiveness of mHealth interventions for behavior changes leading to improved PrEP use. This should allow for the identification and scale-up of impactful interventions while also providing direction as to how these mHealth tools may be tailored to potentially improve different areas of care.

Acknowledgments

This review was made possible by the generous support of the American people through the United States Agency for International Development (USAID) with funds from the US President's Emergency Plan for AIDS Relief (PEPFAR), provided through cooperative agreement number 7200AA20CA00030 to CONRAD and Eastern Virginia Medical School (EVMS) and subawarded to the University of North Carolina at Chapel Hill (UNC) and Florida State University (FSU), and with additional support from Gilead Sciences, a member of the Global Development Alliance with USAID. The contents of this manuscript are the sole responsibility of the authors and do not necessarily reflect the views of USAID, Gilead, EVMS, UNC, FSU, or the United States Government.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of SMS text message publications.
[XLSX File (Microsoft Excel File), 21 KB - mhealth_v13i1e60819_app1.xlsx]

Multimedia Appendix 2
Summary of app publications.
[XLSX File (Microsoft Excel File), 18 KB - mhealth_v13i1e60819_app2.xlsx]

Multimedia Appendix 3
Summary of other publications.
[XLSX File (Microsoft Excel File), 37 KB - mhealth_v13i1e60819_app3.xlsx]

Multimedia Appendix 4
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 226 KB - mhealth_v13i1e60819_app4.pdf]

References

1. UNAIDS Fact Sheet. UNAIDS. URL: https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf [accessed 2023-08-25]
2. Kempton J, Hill A, Levi JA, Heath K, Pozniak A. Most new HIV infections, vertical transmissions and AIDS-related deaths occur in lower-prevalence countries. Journal of Virus Eradication 2019 Apr;5(2):92-101. [doi: 10.1016/s2055-6640(20)30058-3]
3. HIV prevention 2025 road map — Getting on track to end AIDS as a public health threat by 2030. UNAIDS. 2022. URL: https://www.unaids.org/en/resources/documents/2022/prevention-2025-roadmap [accessed 2023-08-25]
4. Murewanhema G, Musuka G, Moyo P, Moyo E, Dzinamarira T. HIV and adolescent girls and young women in sub-Saharan Africa: A call for expedited action to reduce new infections. IJID Reg 2022 Dec;5:30-32 [FREE Full text] [doi: 10.1016/j.ijregi.2022.08.009] [Medline: 36147901]
5. Lewis L, Kharsany ABM, Humphries H, Maughan-Brown B, Beckett S, Govender K, et al. HIV incidence and associated risk factors in adolescent girls and young women in South Africa: A population-based cohort study. PLoS One 2022 Dec 21;17(12):e0279289 [FREE Full text] [doi: 10.1371/journal.pone.0279289] [Medline: 36542645]
6. George G, Beckett S, Reddy T, Govender K, Cawood C, Khanyile D, et al. Determining HIV risk for Adolescent Girls and Young Women (AGYW) in relationships with "Blessers" and age-disparate partners: a cross-sectional survey in four districts

XSL•FO
RenderX

in South Africa. BMC Public Health 2022 May 14;22(1):973 [FREE Full text] [doi: 10.1186/s12889-022-13394-4] [Medline: 35568839]
7. Bajunirwe F, Semakula D, Izudi J. Risk of HIV infection among adolescent girls and young women in age-disparate relationships in sub-Saharan Africa. AIDS 2020 Aug 01;34(10):1539-1548. [doi: 10.1097/QAD.0000000000002582] [Medline: 32443063]
8. Chou R, Spencer H, Bougatsos C, Blazina I, Ahmed A, Selph S. Preexposure prophylaxis for the prevention of HIV: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2023 Aug 22;330(8):746-763. [doi: 10.1001/jama.2023.9865] [Medline: 37606667]
9. O Murchu E, Marshall L, Teljeur C, Harrington P, Hayes C, Moran P, et al. Oral pre-exposure prophylaxis (PrEP) to prevent HIV: a systematic review and meta-analysis of clinical effectiveness, safety, adherence and risk compensation in all populations. BMJ Open 2022 May 11;12(5):e048478 [FREE Full text] [doi: 10.1136/bmjopen-2020-048478] [Medline: 35545381]
10. Fonner VA, Dalglish SL, Kennedy CE, Baggaley R, O'Reilly KR, Koechlin FM, et al. Effectiveness and safety of oral HIV preexposure prophylaxis for all populations. AIDS 2016 Jul 31;30(12):1973-1983 [FREE Full text] [doi: 10.1097/QAD.0000000000001145] [Medline: 27149090]
11. Montgomery E, Roberts S, Reddy K, Tolley EE, Hartmann M, Mathebula F, et al. The CHARISMA randomized controlled trial: a relationship-focused counseling intervention integrated within oral PrEP delivery for HIV prevention among women in Johannesburg, South Africa. J Acquir Immune Defic Syndr 2022 Aug 01;90(4):425-433. [doi: 10.1097/QAI.0000000000002991] [Medline: 35416797]
12. Celum C, Hosek S, Tsholwana M, Kassim S, Mukaka S, Dye BJ, et al. PrEP uptake, persistence, adherence, and effect of retrospective drug level feedback on PrEP adherence among young women in southern Africa: Results from HPTN 082, a randomized controlled trial. PLoS Med 2021 Jun 18;18(6):e1003670 [FREE Full text] [doi: 10.1371/journal.pmed.1003670] [Medline: 34143779]
13. Patel P, Sato K, Bhandari N, Kanagasabai U, Schelar E, Cooney C, PEPFAR PrEP Stakeholders. From policy to practice: uptake of pre-exposure prophylaxis among adolescent girls and young women in United States President's Emergency Plan for AIDS Relief-supported countries, 2017-2020. AIDS 2022 Jun 15;36(Suppl 1):S15-S26. [doi: 10.1097/QAD.0000000000003103] [Medline: 35766572]
14. Irungu EM, Baeten JM. PrEP rollout in Africa: status and opportunity. Nat Med 2020 May 11;26(5):655-664. [doi: 10.1038/s41591-020-0872-x] [Medline: 32405065]
15. Beesham I, Dovel K, Mashele N, Bekker L, Gorbach P, Coates TJ, et al. Barriers to oral HIV pre-exposure prophylaxis (PrEP) adherence among pregnant and post-partum women from Cape Town, South Africa. AIDS Behav 2022 Sep 22;26(9):3079-3087 [FREE Full text] [doi: 10.1007/s10461-022-03652-2] [Medline: 35316471]
16. Haberer J, Mugo N, Bukusi E, Ngure K, Kiptinness C, Oware K, et al. Understanding pre-exposure prophylaxis adherence in young women in Kenya. J Acquir Immune Defic Syndr 2022 Mar 01;89(3):251-260 [FREE Full text] [doi: 10.1097/QAI.0000000000002876] [Medline: 35147580]
17. van der Straten A, Stadler J, Montgomery E, Hartmann M, Magazi B, Mathebula F, et al. Women's experiences with oral and vaginal pre-exposure prophylaxis: the VOICE-C qualitative study in Johannesburg, South Africa. PLoS One 2014 Feb 21;9(2):e89118 [FREE Full text] [doi: 10.1371/journal.pone.0089118] [Medline: 24586534]
18. Irungu E, Khoza N, Velloza J. Multi-level interventions to promote oral pre-exposure prophylaxis use among adolescent girls and young women: a review of recent research. Curr HIV/AIDS Rep 2021 Dec 31;18(6):490-499 [FREE Full text] [doi: 10.1007/s11904-021-00576-9] [Medline: 34719745]
19. Ekwunife OI, Ejie IL, Okelu V, Mita C, Durosinmi-Eti O, Powell A, et al. Interventions to increase the uptake and continuation of pre-exposure prophylaxis (PrEP) by adolescent girls and young women at high risk of HIV in low-income and middle-income countries: a scoping review. BMJ Glob Health 2022 Dec 08;7(12):e009474 [FREE Full text] [doi: 10.1136/bmjgh-2022-009474] [Medline: 36593640]
20. Lee JJ, Li Verdugo J, Xiao AY, Vo K. Digital interventions to enhance PrEP uptake and adherence through stigma reduction. Curr HIV/AIDS Rep 2023 Dec 07;20(6):458-469. [doi: 10.1007/s11904-023-00685-7] [Medline: 38057685]
21. Garrison LE, Haberer JE. Pre-exposure prophylaxis uptake, adherence, and persistence: a narrative review of interventions in the U.S. Am J Prev Med 2021 Nov;61(5 Suppl 1):S73-S86 [FREE Full text] [doi: 10.1016/j.amepre.2021.04.036] [Medline: 34686294]
22. Weitzman PF, Zhou Y, Kogelman L, Rodarte S, Vicente SR, Levkoff SE. mHealth for pre-exposure prophylaxis adherence by young adult men who have sex with men. Mhealth 2021 Jul;7:44-44 [FREE Full text] [doi: 10.21037/mhealth-20-51] [Medline: 34345621]
23. Liu AY, Vittinghoff E, von Felten P, Rivet Amico K, Anderson PL, Lester R, et al. Randomized controlled trial of a mobile health intervention to promote retention and adherence to preexposure prophylaxis among young people at risk for human immunodeficiency virus: the EPIC Study. Clin Infect Dis 2019 May 30;68(12):2010-2017 [FREE Full text] [doi: 10.1093/cid/ciy810] [Medline: 30239620]
24. Mills EJ, Lester R, Thorlund K, Lorenzi M, Muldoon K, Kanters S, et al. Interventions to promote adherence to antiretroviral therapy in Africa: a network meta-analysis. The Lancet HIV 2014 Dec;1(3):e104-e111. [doi: 10.1016/s2352-3018(14)00003-4]

XSL•FO
RenderX

25. Goldstein M, Archary M, Adong J, Haberer JE, Kuhns LM, Kurth A, et al. Systematic review of mHealth interventions for adolescent and young adult HIV prevention and the adolescent HIV continuum of care in low to middle income countries. AIDS Behav 2023 May 02;27(Suppl 1):94-115 [FREE Full text] [doi: 10.1007/s10461-022-03840-0] [Medline: 36322217]
26. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
27. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, Partners PrEP Study Team. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med 2012 Aug 02;367(5):399-410 [FREE Full text] [doi: 10.1056/NEJMoa1108524] [Medline: 22784037]
29. Celum CL, Gill K, Morton JF, Stein G, Myers L, Thomas KK, et al. Incentives conditioned on tenofovir levels to support PrEP adherence among young South African women: a randomized trial. J Int AIDS Soc 2020 Nov 28;23(11):e25636 [FREE Full text] [doi: 10.1002/jia2.25636] [Medline: 33247553]
30. Morton JF, Myers L, Gill K, Bekker L, Stein G, Thomas KK, et al. Evaluation of a behavior-centered design strategy for creating demand for oral PrEP among young women in Cape Town, South Africa. Gates Open Res 2020 Jul 3;4:29 [FREE Full text] [doi: 10.12688/gatesopenres.13103.2] [Medline: 32411946]
31. Hartmann M, Roberts ST, Triplett N, Tenza S, Maboa O, Mampuru L, et al. Development of a relationship counselling website to identify and mitigate risk of intimate partner violence in the context of women's PrEP use. PLOS Digit Health 2023 Aug 14;2(8):e0000329 [FREE Full text] [doi: 10.1371/journal.pdig.0000329] [Medline: 37578954]
32. Medina-Marino A, Bezuidenhout D, Hosek S, Barnabas RV, Atujuna M, Bezuidenhout C, et al. The Community PrEP Study: a randomized control trial leveraging community-based platforms to improve access and adherence to pre-exposure prophylaxis to prevent HIV among adolescent girls and young women in South Africa-study protocol. Trials 2021 Jul 26;22(1):489 [FREE Full text] [doi: 10.1186/s13063-021-05402-6] [Medline: 34311754]
33. Lavoie MC, Okui L, Blanco N, Stoebenau K, Magidson JF, Gokatweng G, et al. Feasibility and acceptability of peer-delivered interventions using mHealth for PrEP services among adolescent girls and young women in DREAMS program in Botswana. Glob Health Action 2023 Dec 31;16(1):2231256 [FREE Full text] [doi: 10.1080/16549716.2023.2231256] [Medline: 37462118]
34. Barnabee G, Billah I, Ndeikemona L, Silas L, Ensminger A, MacLachlan E, et al. PrEP uptake and early persistence among adolescent girls and young women receiving services via community and hybrid community-clinic models in Namibia. PLoS One 2023 Aug 30;18(8):e0289353 [FREE Full text] [doi: 10.1371/journal.pone.0289353] [Medline: 37647257]
35. Kiptinness C, Naik P, Thuo N, Malen RC, Dettinger JC, Pintye J, et al. Online HIV prophylaxis delivery: Protocol for the ePrEP Kenya pilot study. Front Public Health 2023 Feb 24;11:1054559 [FREE Full text] [doi: 10.3389/fpubh.2023.1054559] [Medline: 36908449]
36. Velloza J, Khoza N, Scorgie F, Chitukuta M, Mutero P, Mutiti K, HPTN 082 study group. The influence of HIV-related stigma on PrEP disclosure and adherence among adolescent girls and young women in HPTN 082: a qualitative study. J Int AIDS Soc 2020 Mar 06;23(3):e25463 [FREE Full text] [doi: 10.1002/jia2.25463] [Medline: 32144874]
37. Mauka W, Mbotwa C, Moen K, Lichtwarck HO, Haaland I, Kazaura M, et al. Development of a mobile health application for HIV prevention among at-risk populations in urban settings in East Africa: a participatory design approach. JMIR Form Res 2021 Oct 07;5(10):e23204 [FREE Full text] [doi: 10.2196/23204] [Medline: 34617904]
38. Mbotwa C, Kazaura M, Moen K, Leshabari M, Metta E, Leyna G, et al. Predictors of mHealth use in promoting adherence to pre-exposure prophylaxis among female sex workers: an evaluation of the Jichunge intervention in Dar es Salaam, Tanzania. BMC Health Serv Res 2022 Jul 04;22(1):859 [FREE Full text] [doi: 10.1186/s12913-022-08245-2] [Medline: 35787285]
39. Kreniske P, Namuyaba OI, Kasumba R, Namatovu P, Ssewamala F, Wingood G, et al. Mobile phone technology for preventing HIV and related youth health problems, sexual health, mental health, and substance use problems in southwest Uganda (Youth Health SMS): Protocol for a pilot randomized controlled trial. JMIR Res Protoc 2023 Dec 19;12:e49352 [FREE Full text] [doi: 10.2196/49352] [Medline: 38113102]
40. Busang J, Zuma T, Herbst C, Okesola N, Chimbindi N, Dreyer J, et al. Thetha Nami ngithethe nawe (Let's Talk): a stepped-wedge cluster randomised trial of social mobilisation by peer navigators into community-based sexual health and HIV care, including pre-exposure prophylaxis (PrEP), to reduce sexually transmissible HIV amongst young people in rural KwaZulu-Natal, South Africa. BMC Public Health 2023 Aug 15;23(1):1553 [FREE Full text] [doi: 10.1186/s12889-023-16262-x] [Medline: 37582746]
41. Moorhouse L, Schaefer R, Thomas R, Nyamukapa C, Skovdal M, Hallett TB, et al. Application of the HIV prevention cascade to identify, develop and evaluate interventions to improve use of prevention methods: examples from a study in east Zimbabwe. J Int AIDS Soc 2019 Jul 22;22 Suppl 4(Suppl Suppl 4):e25309 [FREE Full text] [doi: 10.1002/jia2.25309] [Medline: 31328375]

XSL•FO
RenderX

42. Hanass-Hancock J, Bhengu N, Exner T, Magusthwa S, Harrison A, Dolezal C, et al. Masibambane-Ladies Chat: Developing an online gender-enhanced PrEP information-motivation workshop for young South African women. AIDS Educ Prev 2023 Feb;35(1):14-35 [FREE Full text] [doi: 10.1521/aeap.2023.35.1.14] [Medline: 36735226]
43. Ngure K, Thuo N, Ogello V, Kiptinness C, Kamolloh K, Burns BFO, et al. Dynamic perceived HIV risk and sexual behaviors among young women enrolled in a PrEP trial in Kenya: a qualitative study. Front Reprod Health 2021 Aug 12;3:637869 [FREE Full text] [doi: 10.3389/frph.2021.637869] [Medline: 36304002]
44. Haberer JE, Bukusi EA, Mugo NR, Pyra M, Kiptinness C, Oware K, MPYA Study Team. Effect of SMS reminders on PrEP adherence in young Kenyan women (MPYA study): a randomised controlled trial. The Lancet HIV 2021 Mar;8(3):e130-e137. [doi: 10.1016/s2352-3018(20)30307-6]
45. Haberer J, Mugo N, Bukusi E, Pyra M, Oware K, Kiptinness C, et al. OA07.01: SMS reminders did not improve PrEP adherence in a randomized controlled trial among young Kenyan women. J Int AIDS Soc 2021 Feb;24 Suppl 1(Suppl 1):e25659 [FREE Full text] [doi: 10.1002/jia2.25659] [Medline: 33605533]
46. Ogello V, Ngure K, Thuo N, Burns B, Rono B, Oware K, et al. "Yes, I'm reminded, but it doesn't mean I'm taking them": Experiences with short message service reminder use in real-time monitoring of HIV PrEP among young women in Kenya. AIDS Behav 2023 Jan 30;27(1):65-74. [doi: 10.1007/s10461-022-03744-z] [Medline: 35907142]
47. Cassidy T, Ntuli N, Kilani C, Malabi N, Rorwana B, Mutseyekwa T, et al. Delivering PrEP to young women in a low-income setting in South Africa: lessons for providing both convenience and support. AIDS Behav 2022 Jan 14;26(1):147-159 [FREE Full text] [doi: 10.1007/s10461-021-03366-x] [Medline: 34259963]
48. Mogaka JN, Otieno FA, Akim E, Beima-Sofie K, Dettinger J, Gomez L, et al. A text messaging-based support intervention to enhance pre-exposure prophylaxis for HIV prevention adherence during pregnancy and breastfeeding: protocol for a randomized controlled trial. JMIR Res Protoc 2023 Jan 30;12:e41170 [FREE Full text] [doi: 10.2196/41170] [Medline: 36716092]
49. Pintye J, Rogers Z, Kinuthia J, Mugwanya KK, Abuna F, Lagat H, et al. Two-way short message service (SMS) communication may increase pre-exposure prophylaxis continuation and adherence among pregnant and postpartum women in Kenya. Glob Health Sci Pract 2020 Mar 05;8(1):55-67. [doi: 10.9745/ghsp-d-19-00347]
50. Celum C, Seidman D, Travill D, Dehlendorf C, Gumede S, Zewdie K, et al. A decision support tool has similar high PrEP uptake and increases early PrEP persistence in adolescent girls and young women in South Africa: results from a randomized controlled trial. J Int AIDS Soc 2023 Aug 27;26(8):e26154 [FREE Full text] [doi: 10.1002/jia2.26154] [Medline: 37634942]
51. Thomas R, Skovdal M, Galizzi MM, Schaefer R, Moorhouse L, Nyamukapa C, et al. Improving risk perception and uptake of pre-exposure prophylaxis (PrEP) through interactive feedback-based counselling with and without community engagement in young women in Manicaland, East Zimbabwe: study protocol for a pilot randomized trial. Trials 2019 Dec 02;20(1):668 [FREE Full text] [doi: 10.1186/s13063-019-3791-8] [Medline: 31791405]
52. Celum CL, Bukusi EA, Bekker LG, Delany-Moretlwe S, Kidoguchi L, Omollo V, POWER Study Team. PrEP use and HIV seroconversion rates in adolescent girls and young women from Kenya and South Africa: the POWER demonstration project. J Int AIDS Soc 2022 Jul 13;25(7):e25962 [FREE Full text] [doi: 10.1002/jia2.25962] [Medline: 35822945]
53. Schwartz K, Ferrigno B, Vining S, Gomez A, Briedenhann E, Gardiner E, et al. PrEP Communications Accelerator: a digital demand creation tool for sub-Saharan Africa. Sex. Health 2018;15(6):570. [doi: 10.1071/sh18064]
54. Velloza J, Poovan N, Ndlovu N, Khoza N, Morton JF, Omony J, et al. Adaptive HIV pre-exposure prophylaxis adherence interventions for young South African women: Study protocol for a sequential multiple assignment randomized trial. PLoS One 2022 Apr 13;17(4):e0266665 [FREE Full text] [doi: 10.1371/journal.pone.0266665] [Medline: 35417485]
55. Heffron R, Ngure K, Velloza J, Kiptinness C, Quame-Amalgo J, Oluch L, et al. Implementation of a comprehensive safer conception intervention for HIV-serodiscordant couples in Kenya: uptake, use and effectiveness. J Int AIDS Soc 2019 Apr;22(4):e25261 [FREE Full text] [doi: 10.1002/jia2.25261] [Medline: 30957420]
56. Velloza J, Ngure K, Kiptinness C, Quame-Amaglo J, Thuo N, Dew K, et al. A clinic-based tablet application to support safer conception among HIV serodiscordant couples in Kenya: feasibility and acceptability study. Mhealth 2019 Feb;5:4-4 [FREE Full text] [doi: 10.21037/mhealth.2019.01.04] [Medline: 30976596]
57. Matambanadzo P, Busza J, Mafaune H, Chinyanganya L, Machingura F, Ncube G, et al. "It went through the roof": an observation study exploring the rise in PrEP uptake among Zimbabwean female sex workers in response to adaptations during Covid-19. J Int AIDS Soc 2021 Oct 28;24 Suppl 6(Suppl 6):e25813 [FREE Full text] [doi: 10.1002/jia2.25813] [Medline: 34713613]
58. PrEP Communications Accelerator. URL: https://accelerator.prepwatch.org/ [accessed 2023-10-12]
59. Mutua G, Sanders E, Mugo P, Anzala O, Haberer JE, Bangsberg D, et al. Safety and adherence to intermittent pre-exposure prophylaxis (PrEP) for HIV-1 in African men who have sex with men and female sex workers. PLoS One 2012 Apr 12;7(4):e33103 [FREE Full text] [doi: 10.1371/journal.pone.0033103] [Medline: 22511916]
60. Mallayasamy S, Chaturvedula A, Fossler MJ, Sale M, Goti V, Bumpus NN, et al. Tenofovir plasma concentration from preexposure prophylaxis at the time of potential HIV exposure: a population pharmacokinetic modeling and simulation study involving serodiscordant couples in East Africa. Antimicrob Agents Chemother 2019 Aug;63(8):19. [doi: 10.1128/aac.00446-19]

XSL•FO
RenderX

61. Curran K, Mugo NR, Kurth A, Ngure K, Heffron R, Donnell D, et al. Daily short message service surveys to measure sexual behavior and pre-exposure prophylaxis use among Kenyan men and women. AIDS Behav 2013 Nov;17(9):2977-2985 [FREE Full text] [doi: 10.1007/s10461-013-0510-4] [Medline: 23695519]
62. Dietrich JJ, Benadé GL, Mulaudzi M, Kagee A, Hornschuh S, Makhale LM, et al. "You are on the right track with the app:" Qualitative analysis of mobile phone use and user feedback regarding mobile phone sexual risk assessments for HIV prevention research. Front Digit Health 2021 Mar 22;3:576514 [FREE Full text] [doi: 10.3389/fdgth.2021.576514] [Medline: 34713080]
63. A Framework for the Economic Evaluation of Digital Health Interventions. World Bank Group. URL: https://documents1.worldbank.org/curated/en/099446504122313917/pdf/IDU0f639726d0f11404a3509af8054677649dcd6.pdf [accessed 2023-10-16]
64. Kecojevic A, Basch C, Basch C, Kernan W. Pre-exposure prophylaxis YouTube videos: content evaluation. JMIR Public Health Surveill 2018 Feb 16;4(1):e19 [FREE Full text] [doi: 10.2196/publichealth.7733] [Medline: 29467119]
65. Magno L, Soares F, Zucchi EM, Eustórgio M, Grangeiro A, Ferraz D, PrEP1519 Study Group. Reaching out to adolescents at high risk of HIV infection in Brazil: demand creation strategies for PrEP and other HIV combination prevention methods. Arch Sex Behav 2023 Feb 25;52(2):703-719 [FREE Full text] [doi: 10.1007/s10508-022-02371-y] [Medline: 35876978]
66. Darien K, Lee S, Knowles K, Wood S, Langer MD, Lazar N, et al. Health information from web search engines and virtual assistants about pre-exposure prophylaxis for HIV prevention in adolescents and young adults: content analysis. JMIR Pediatr Parent 2023 Jul 18;6:e41806 [FREE Full text] [doi: 10.2196/41806] [Medline: 37463044]
67. Gagnon KW, Coulter RW, Egan JE, Ho K, Hawk M. Associations between sexual history documentation in electronic health records and referral to pre-exposure prophylaxis navigator on prescription of pre-exposure prophylaxis at a multi-site federally qualified health center. AIDS Patient Care STDS 2023 Aug 01;37(8):403-415. [doi: 10.1089/apc.2023.0068] [Medline: 37566534]
68. Shrestha R, Wickersham JA, Khati A, Azwa I, Ni Z, Kamarulzaman A, et al. Clinic-integrated mobile health intervention ("JomPrEP" App) to improve uptake of HIV testing and pre-exposure prophylaxis among men who have sex with men in Malaysia: protocol for an intervention development and multiphase trial. JMIR Res Protoc 2022 Dec 21;11(12):e43318 [FREE Full text] [doi: 10.2196/43318] [Medline: 36542425]
69. The Mobile Economy Sub-Saharan Africa. GSMA. 2022. URL: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/wp-content/uploads/2022/10/The-Mobile-Economy-Sub-Saharan-Africa-2022.pdf [accessed 2024-05-11]
70. D'Anza B, Pronovost PJ. Digital health: unlocking value in a post-pandemic world. Popul Health Manag 2022 Feb 01;25(1):11-22. [doi: 10.1089/pop.2021.0031] [Medline: 34042532]
71. Fischer AE, Van Tonder T, Gumede SB, Lalla-Edward ST. Changes in perceptions and use of mobile technology and health communication in South Africa during the COVID-19 lockdown: cross-sectional survey study. JMIR Form Res 2021 May 17;5(5):e25273 [FREE Full text] [doi: 10.2196/25273] [Medline: 33956640]
72. Policy Brief: Virtual Interventions in Response to HIV, Sexually Transmitted Infections and Viral Hepatitis. UNAIDS. URL: https://www.unaids.org/sites/default/files/media_asset/policy-brief_virtual-interventions_en.pdf [accessed 2023-10-16]
73. Mataboge P, Nzenze S, Mthimkhulu N, Mazibuko M, Kutywayo A, Butler V, et al. Planning for decentralized, simplified prEP: Learnings from potential end users in Ga-Rankuwa, gauteng, South Africa. Front Reprod Health 2022 Jan 9;4:1081049 [FREE Full text] [doi: 10.3389/frph.2022.1081049] [Medline: 36699142]
74. Pindayi B. Social Media Uses and Effects: The Case of WhatsApp in Africa. In: Nelson O, Ojebuyi B, Salawu A, editors. Impacts of the Media on African Socio-Economic Development. Hershey, PA: IGI Global; 2017:34-51.
75. Agrahari V, Anderson SM, Peet MM, Wong AP, Singh ON, Doncel GF, et al. Long-acting HIV pre-exposure prophylaxis (PrEP) approaches: recent advances, emerging technologies, and development challenges. Expert Opin Drug Deliv 2022 Oct 25;19(10):1365-1380 [FREE Full text] [doi: 10.1080/17425247.2022.2135699] [Medline: 36252277]
76. Kennedy CE, Yeh PT, Atkins K, Ferguson L, Baggaley R, Narasimhan M. PrEP distribution in pharmacies: a systematic review. BMJ Open 2022 Feb 21;12(2):e054121 [FREE Full text] [doi: 10.1136/bmjopen-2021-054121] [Medline: 35190430]
77. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar 19;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
78. Nwaozuru U, Obiezu-Umeh C, Shato T, Uzoaru F, Mason S, Carter V, et al. Mobile health interventions for HIV/STI prevention among youth in low- and middle-income countries (LMICs): a systematic review of studies reporting implementation outcomes. Implement Sci Commun 2021 Nov 06;2(1):126 [FREE Full text] [doi: 10.1186/s43058-021-00230-w] [Medline: 34742357]
79. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
80. B-wise Health. URL: https://www.bwisehealth.com/ [accessed 2024-02-13]
81. Desmond Tutu Health Foundation. URL: https://desmondtutuhealthfoundation.org.za/ [accessed 2024-02-13]

XSL•FO
RenderX

82. Rousseau E, Julies RF, Madubela N, Kassim S. Novel platforms for biomedical HIV prevention delivery to key populations - community mobile clinics, peer-supported, pharmacy-led PrEP delivery, and the use of telemedicine. Curr HIV/AIDS Rep 2021 Dec 27;18(6):500-507 [FREE Full text] [doi: 10.1007/s11904-021-00578-7] [Medline: 34708316]
83. Rousseau E, Bekker L, Julies RF, Celum C, Morton J, Johnson R, et al. A community-based mobile clinic model delivering PrEP for HIV prevention to adolescent girls and young women in Cape Town, South Africa. BMC Health Serv Res 2021 Aug 28;21(1):888 [FREE Full text] [doi: 10.1186/s12913-021-06920-4] [Medline: 34454505]
84. Hightow-Weidman LB, Muessig KE. New media challenges and opportunities. Sex Transm Infect 2017 Aug 25;93(5):309-310. [doi: 10.1136/sextrans-2016-052673] [Medline: 27888207]

Abbreviations

ART: antiretroviral therapy
CHARISMA: Community Health Clinic Model for Agency in Relationships and Safer Microbicide Adherence
DREAMS: Determined, Resilient, Empowered, AIDS-free and Safe
DSD: differentiated service delivery
LMIC: low- and middle-income country
mHealth: mobile health
MPYA: Monitoring PrEP among Young Adult Women
MSF: Me´decins Sans Frontie` res
mWACh: Mobile Solutions for Women's and Children's Health
POWER: Prevention Options for Women Evaluation Research
PrEP: pre-exposure prophylaxis
RCT: randomized controlled trial
SCIP: Safer Conception Intervention for Partners
SMART: Sequential Multiple Assignment Randomized Trial
UNAIDS: Joint United Nations Programme on HIV and AIDS

Edited by L Buis; submitted 23.05.24; peer-reviewed by M Hartmann, M Wang; comments to author 23.10.24; revised version received 12.12.24; accepted 21.05.25; published 20.06.25.

Please cite as:
Fischer AE, Hanif H, Stocks JB, Rochelle AE, Dominguez K, Armora Langoni EG, Reyes HLM, Doncel GF, Muessig KE
Mobile Health Intervention Tools Promoting HIV Pre-Exposure Prophylaxis Among Adolescent Girls and Young Women in Sub-Saharan Africa: Scoping Review
JMIR Mhealth Uhealth 2025;13:e60819
URL: https://mhealth.jmir.org/2025/1/e60819
doi:10.2196/60819
PMID:40540732

©Alex Emilio Fischer, Homaira Hanif, Jacob B Stocks, Aimee E Rochelle, Karen Dominguez, Eliana Gabriela Armora Langoni, H Luz McNaughton Reyes, Gustavo F Doncel, Kathryn E Muessig. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Use of Wearable Sensors to Assess Fall Risk in Neurological Disorders: Systematic Review

Mirjam Bonanno[1], PT, MSc, DO; Augusto Ielo[1], MSc; Paolo De Pasquale[1], PhD; Antonio Celesti[2], PhD; Alessandro Marco De Nunzio[3], PhD; Angelo Quartarone[1], MD, PhD; Rocco Salvatore Calabrò[1], MD, PhD

[1]IRCCS Centro Neurolesi Bonino-Pulejo, Messina, Italy

[2]Department of Mathematics, Computer Science, Physics and Earth Science, University of Messina, Messina, Italy

[3]Department of Health, LUNEX University of Applied Sciences, Differdange, Luxembourg

Corresponding Author:
Augusto Ielo, MSc
IRCCS Centro Neurolesi Bonino-Pulejo
Strada Statale 113 - C.da Casazza
Messina, 98124
Italy
Phone: 39 090 6012 ext 8100
Email: augusto.ielo@irccsme.it

Abstract

Background: Assessing fall risk, especially in individuals with neurological disorders, is essential to prevent hospitalization, hypomobility, and reduced functional independence. Wearable sensors are increasingly used in neurorehabilitation, as they enable unsupervised fall risk assessment by providing continuous monitoring during daily functional tasks, thereby offering a reflection of the individual's real-world fall risk.

Objective: We systematically reviewed the literature on reliable biomechanical gait parameters detected with wearable sensors to assess fall risk in neurological disorders, focusing on patients with Parkinson disease, multiple sclerosis, stroke, or Alzheimer disease. In addition, we examined the latest advancements in wearable sensor technology, including best practices for device placement as well as data processing and analysis.

Methods: We conducted a comprehensive systematic search for relevant peer-reviewed articles published up to April 18, 2025, using PubMed, Web of Science, Embase, and IEEE Xplore, which are the most used databases in the fields of health and bioengineering.

Results: The 19 included studies involved 2630 patients with neurological disorders, including 226 (8.59%) with multiple sclerosis (n=7, 37% studies), 2305 (87.64%) with Parkinson disease (n=8, 53% studies), 51 (1.94%) with stroke (n=3, 16% studies), and 48 (1.83%) with Alzheimer disease or cognitive impairment (n=1, 5% study).

Conclusions: This review highlights the role of wearable technologies in assessing fall risk in patients with neurological disorders. Although the included studies showed variation in methods and a focus on technology over clinical context, the lack of standardization reflects ongoing advancements, which may be seen as a strength.

Trial Registration: PROSPERO CRD42023463944; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023463944

(JMIR Mhealth Uhealth 2025;13:e67265) doi:10.2196/67265

KEYWORDS

fall risk assessment; wearable sensors; neurological disorders; neurorehabilitation; artificial intelligence; AI

Introduction

Background

Neurological disorders, such as Parkinson disease (PD), multiple sclerosis (MS), Alzheimer disease (AD), and stroke, are significant and growing global health concerns due to their high prevalence. A recent meta-analysis reported a global pooled prevalence of PD of 1.51 cases per 1000 individuals, with higher rates observed among male individuals and a significant increase over the past 4 decades [1]. Similarly, MS currently affects approximately 2.8 million people globally, with rising prevalence and incidence trends observed worldwide [2]. In Europe, the prevalence of AD has been estimated at 5.05% (95%

XSL•FO
RenderX

CI 4.73%-5.39%) [3]. The incidence of clinically diagnosed AD per 1000 person-years increases markedly with age: 3.4 for individuals aged 65 to 74 years, 13.8 for those aged 75 to 84 years, and 35.8 for those aged ≥85 years [4,5]. Regarding cerebrovascular impairments, stroke continues to be a leading cause of death and disability, with >12 million new cases per year and >100 million survivors of stroke worldwide [6].

These conditions often lead to alterations in gait and balance, increasing the risk of falls. This risk is especially pronounced in older adults and individuals with neurological impairments. Approximately one-third of adults aged >65 years experience a fall each year [7], and this rate rises to nearly 50% among those aged >80 years [8]. Consequently, falls in this patient population considered fragile can worsen quality of life due to hospitalization, hypomobility, and reduced functional independence [9]. Fall-related injuries not only entail substantial medical costs but also determine patients' mortality risk. Among neurological disorders, PD, MS, cerebrovascular accidents, and AD are among the most common, causing significant motor disability. It is noteworthy that these patients with neurological disorders are considered recurrent fallers due to their gait features. In particular, it is estimated that 39% of patients with PD experience falls, while 40% to 60% of patients with MS experience recurrent falls [10]. Similarly, patients with stroke are at high risk for falls, with fall frequencies ranging from 27% to 39% in rehabilitation hospitals [10]. Furthermore, older adults with AD are more than twice as likely to experience a fall compared to cognitively healthy peers, with rates ranging from 60% to 80% [11]. However, each neurological disorder is characterized by considerable heterogeneity in gait disturbances due to differences in underlying etiology [12]; for example, neurodegenerative disorders (such as PD, MS, and AD) are associated with distinct gait patterns, characterized by reduced gait speed, shorter step length, and increased muscle weakness [12]. In addition, ambulation in patients with PD is characterized by shorter steps, loss of dissociated arm and trunk movements during gait, and postural instability [13]. By contrast, patients with MS manifest pathological gait features related to spasticity, ataxia, muscle weakness, and sensory and proprioceptive deficits, all of which contribute to progressive disability [14]. AD independently increases fall risk due to its associated impairments in judgment, gait, visuospatial perception, and hazard recognition [15]. Differently from neurodegenerative disorders, patients with stroke are characterized by a marked spatiotemporal asymmetry between steps, which substantially increases fall risk [16]. For these reasons, gait and balance assessment in patients with neurological disorders is essential in all phases of the rehabilitation process.

To this aim, clinical scales (ie, the timed up and go [TUG] test, the Tinetti Scale, the Falls Efficacy Scale–International, and the Berg Balance Scale) as well as innovative technologies, including wearable sensors and nonwearable devices, are used. Specifically, wearable sensors are growing in popularity in both clinical and research contexts due to their suitable features [17]: because of their small size, affordability, and user-friendly nature, wearable sensors represent an optimal solution for health monitoring [18,19]. Unlike traditional laboratory-based monitoring systems, wearable sensors can facilitate the continuous monitoring of motor activity by capturing physiological data during everyday routines [20]. In addition, fall risk assessments with laboratory-based systems are generally performed in supervised conditions, in which the behavior adopted by participants may not be representative of the one adopted in everyday life—during the experimental tasks, the participants might be attempting their "best effort" [21]. Generally, falls occur in unpredictable situations during everyday activities [22]. In this context, wearable sensors enable unsupervised fall risk assessment by providing continuous monitoring during daily functional tasks, thereby offering a reflection of the individual's real-world fall risk [20,21,23].

Thus, interest in wearable sensors has been increasing with regard to monitoring fall risk among older adults [21]. Among the prevalent wearable technologies used for health monitoring are inertial measurement units (IMUs), which gather data from accelerometers and gyroscopes. IMUs can be strategically positioned on the body to capture motion data for subsequent analysis, integration, and interpretation [24]. Gait characteristics that can be analyzed include stride length, stance-to-swing phase ratio, and various other spatiotemporal gait parameters [21,25,26]. In addition, data from inertial sensors collected during walking can be used to investigate upper body dynamics through nonlinear analysis techniques, such as recurrence quantification analysis [27]. Some researchers have used inertial sensor data to detect balance alterations by analyzing sway speed in the anterior-posterior and medial-lateral directions of the body's center of mass projection path in patients with AD [28].

Other wearable sensors include insole-based devices, which have been developed for fall risk assessment and detection. These devices are embedded in the shoe sole to capture data on plantar pressure distribution [21]. The features extracted from these devices are used to objectively quantify the risk of falling in individuals with neurological disorders. However, the selection of the most reliable biomechanical gait parameters related to fall risk remains an open question [21]. In addition, the information acquired from wearable sensors, along with data processing, data fusion, and analysis techniques such as machine learning (ML) models or manual data analysis, is used to extract meaningful insights from the investigated parameters [29].

To this aim, some studies have investigated the role of deep learning (DL) in the context of neurorehabilitation. The DL-based approach can be used to extract data on stride-specific gait parameters from IMUs [30]. This approach could be useful in the field of telemedicine as a potential in-home gait monitoring tool, given that alterations in gait function can serve as a biomarker for predicting fall risk.

Objectives

In this context, our first objective was to systematically review the literature on the most reliable biomechanical gait parameters, as detected by wearable sensors, for evaluating fall risk in individuals with neurological disorders. Therefore, we selected studies that involved patients with PD, MS, AD, or stroke, as these conditions collectively represent a significant proportion of neurological disorders worldwide and are associated with

XSL•FO
RenderX

substantial gait impairment and a high risk of falls. Our second objective was to investigate existing literature on wearable sensor technology, including best practices for device placement as well as data processing and analysis.

Methods

Overview

We conducted a comprehensive systematic review to explore the existing evidence on fall risk assessment in patients with neurological disorders. We summarized the results of all published studies in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [31]. The PRISMA checklist is available in Multimedia Appendix 1. The protocol has been registered with PROSPERO (CRD42023463944).

Population, Exposure, Comparator, Outcomes Question

We used the population, exposure, comparator, outcomes model to define the search strategy [32] (Table 1). The population, exposure, comparator, outcomes framework was applied to address the following research question: "what are the most reliable biomechanical gait parameters detected with wearable sensors to assess fall risk in neurological disorders?"

Table 1. Population, exposure, comparator, outcomes (PECO) framework for included studies.

PECO elements	Definition	Search terms
Population	Patients with neurological disorders, including cerebrovascular diseases, Parkinson disease, Alzheimer disease, and multiple sclerosis	"stroke" OR "Parkinson's disease" OR "multiple sclerosis" OR "Alzheimer"
Exposure	Fall risk assessment using wearable sensors or gait analysis using wearable sensors	"wearable"
Comparator	Studies including a comparison with healthy controls	—[a]
Outcomes	Biomechanical gait parameters related to fall risk	"fall"

[a]Not applicable.

Search Strategy and Eligibility Criteria

We conducted a systematic search for all peer-reviewed articles published up to April 18, 2025, using PubMed, Web of Science, Embase, and IEEE Xplore, which are the most used databases in the fields of health and bioengineering. The full search queries are provided in Multimedia Appendix 2. We included studies involving adults (aged ≥18 y) with stroke, PD, MS, or AD. Specifically, the inclusion criteria were as follows: (1) adult patients with cerebrovascular impairments, PD, MS, or AD; (2) fall risk assessment using wearable sensors; (3) articles written in English; and (4) published in a peer-reviewed journal. We excluded articles describing theoretical models, methodological approaches, algorithms, and basic technical descriptions. We also excluded animal studies, conference proceedings and reviews, studies involving children, and case reports. The list of articles was refined for relevance, revised, and summarized, with the key topics identified from the summaries based on the inclusion and exclusion criteria. Considering the limited literature available, we included the following study designs in the qualitative synthesis: (1) randomized controlled trials, (2) observational studies, (3) cross-sectional studies, (4) case-control studies, and (5) cohort studies. All search results were imported into Rayyan (Rayyan Systems Inc) [33], a web application designed to facilitate the systematic review process, and screened by 2 blinded reviewers (AI and MB). Each reviewer independently screened the articles to ensure consistent application of the inclusion and exclusion criteria. After screening titles and abstracts, blinding was removed. In cases of disagreement, the 2 reviewers discussed the inclusion or exclusion of the undecided articles to reach consensus.

Data Extraction and Analysis

After full-text selection, data from the included studies were extracted and organized in a spreadsheet. Information was summarized using the following fields: assigned ID number, study title, year of publication or presentation and first author, study aims and design, study duration, recruitment method and setting, inclusion and exclusion criteria, use of a control group, use of wearable sensors, informed consent, conflicts of interest and funding, type of intervention or evaluation and control, number of participants, baseline characteristics, intervention setting, types of biomechanical parameters related to fall risk, results, and key conclusions. Data extraction was performed independently and in a blinded manner by 2 reviewers. In cases of disagreement, a third reviewer was consulted to resolve discrepancies.

Risk of Bias

The risk of bias in the included studies was assessed using the National Heart, Lung, and Blood Institute Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies [34]. This tool comprises 14 questions, with each item rated as "yes," "no," "cannot determine," "not reported," or "not applicable." On the basis of the responses, the overall quality of each study was categorized as good, fair, or poor. The risk of bias was assessed independently and in a blinded manner by 2 reviewers. In cases of disagreement, a third reviewer was consulted to resolve discrepancies.

Results

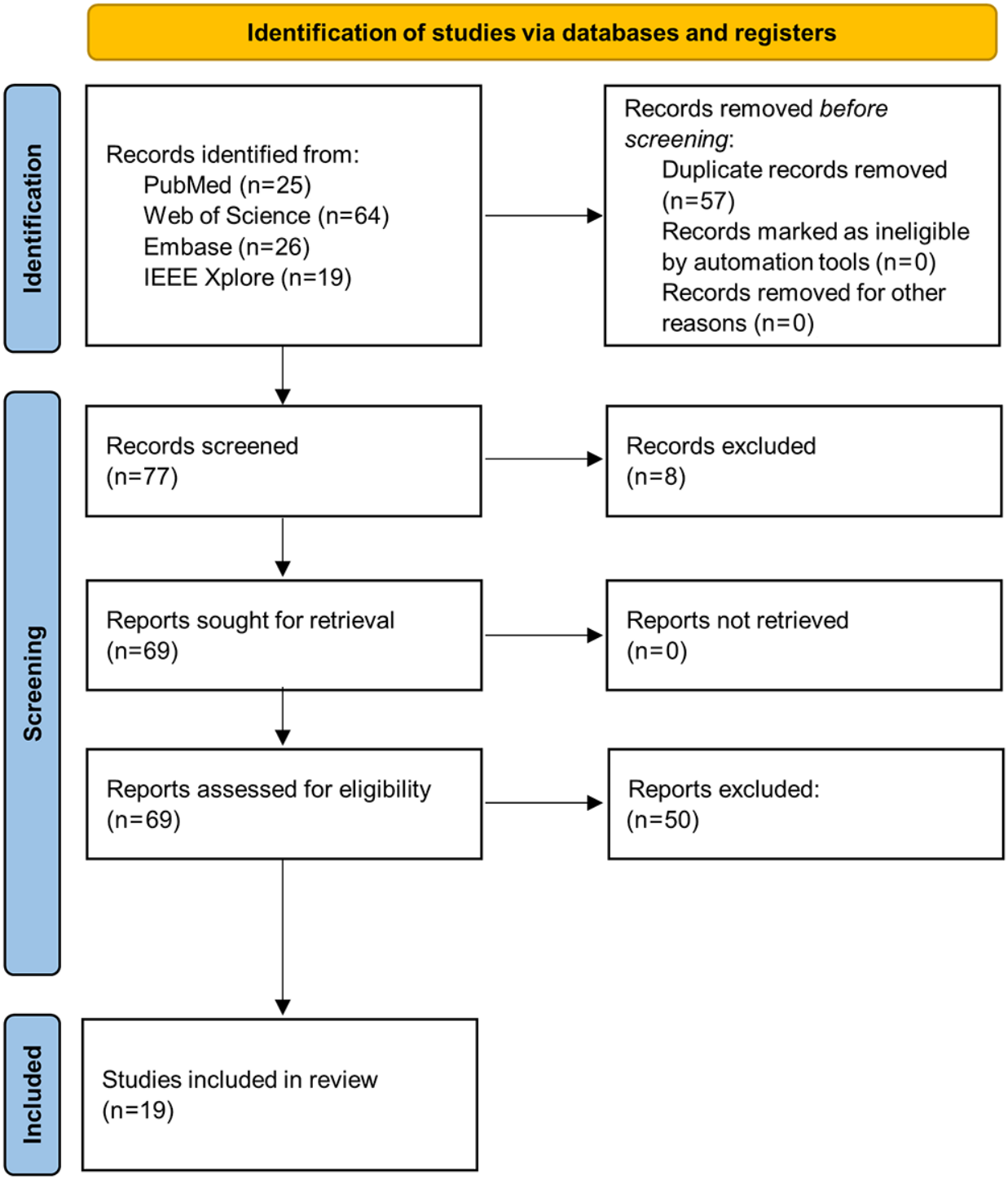
Overview

A total of 134 articles were identified through database search (PubMed: n=25, 18.7%; Web of Science: n=64, 47.8%; Embase:

XSL•FO
RenderX

n=26, 19.4; and IEEE Xplore: n=19, 14.2%). After removing duplicates (57/134, 42.5%), we screened the remaining articles (77/134, 57.5%). Of these 77 articles, 69 (90%) met the inclusion criteria, and 8 (10%) were excluded. After full-text analysis, of the 69 articles, 50 (72%) were excluded, and 19 (28%) studies involving 5859 participants were included in the review synthesis (Figure 1). The included studies were published between 2016 and 2025.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing the number of studies identified, screened, assessed for eligibility, and included in the final analysis.

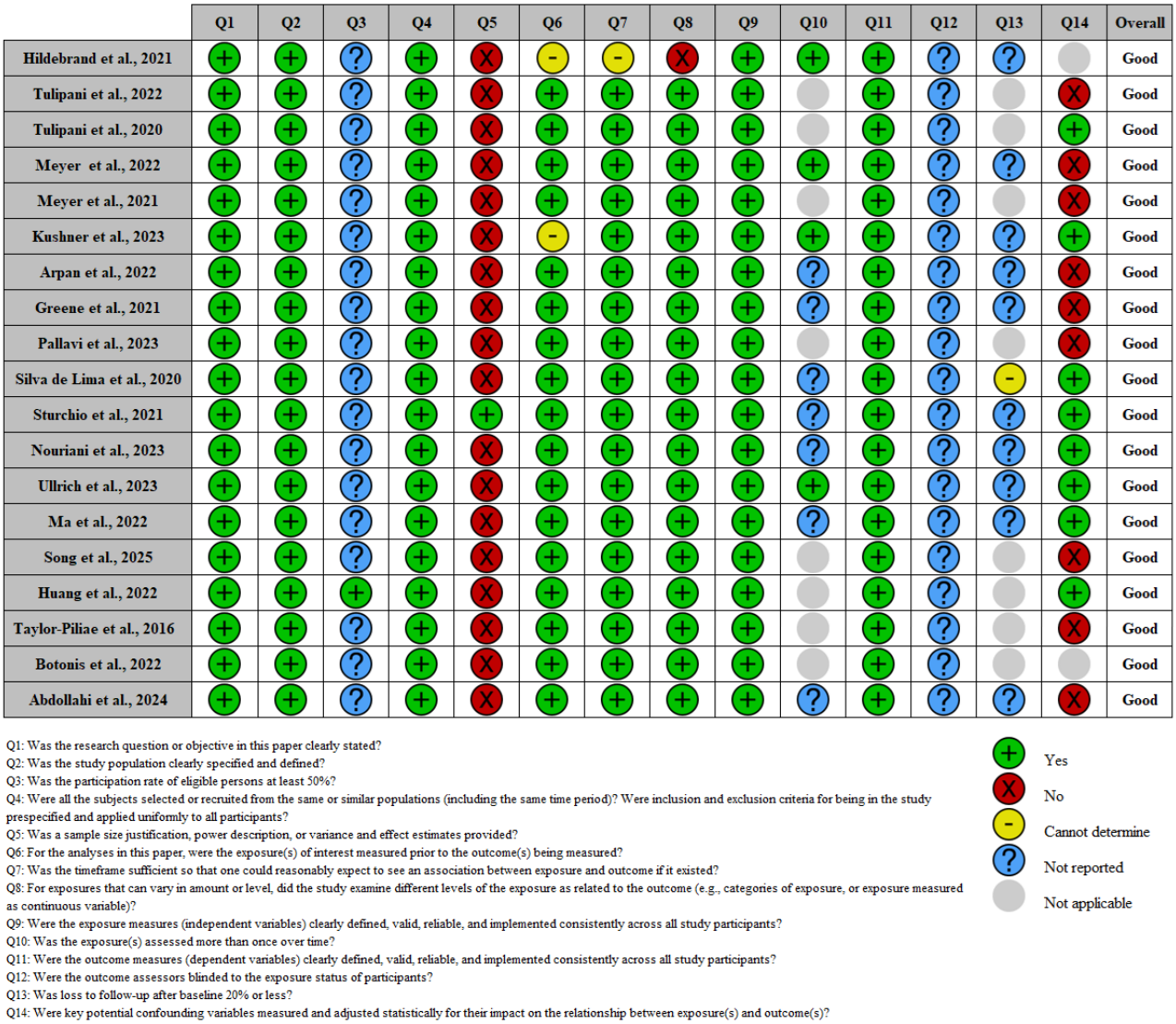


Quality of the Studies and Risk-of-Bias Assessment

The risk of bias in the included studies was assessed using the National Heart, Lung, and Blood Institute Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies, which comprises 14 questions (Figure 2)

XSL•FO
RenderX

Figure 2. Risk-of-bias assessment for included studies [23,35-52].



All included studies were rated by the reviewers as having overall good quality (Figure 2), and there was homogeneity in the methodologies and procedures used to assess fall risk in neurological disorders. However, concerns were identified in some domains of the quality assessment tool, specifically items 3, 5, 12, and 13; for example, participation rates of eligible individuals were documented in only 5% (1/19) of the studies [50], a sample-size justification or power analysis was provided only in the study by Sturchio et al [39], repeat assessment of the exposure was described in only 21% (4/19) of the studies [48,49,51,52], and 79% (15/19) of the studies relied on a single measurement or used a cross-sectional design. Moreover, procedures for blinding outcome assessors to exposure status were not reported, and information sufficient to determine loss to follow-up was not provided in the prospective cohort studies. However, these reporting limitations did not affect the overall quality classification assigned to each study.

Study Population

The included studies involved 2630 patients with neurological disorders, including 226 (8.59%) with MS (n=7, 37% studies), 2305 (87.64%) with PD (n=10, 53% studies), 51 (1.94%) with stroke (n=3, 16% studies), and 48 (1.83%) with AD or cognitive impairment (n=1, 5% study). An additional 3229 participants were healthy controls or older adults. Regarding disease severity, the studies conducted in patients with MS had a mean Expanded Disability Status Scale score of 3.4; and in patients with PD, the mean score on the Unified Parkinson's Disease Rating Scale, part III (motor examination), was 26.6, while the mean Hoehn and Yahr stage was 2.58 (based on the studies that reported these metrics). Regarding patients with stroke, 2 (11%) of the 19 studies [36,38] included participants who had experienced a cerebrovascular accident at least 5 years earlier, while 1 (5%) included patients who had experienced a stroke at least 6 months before the day of the experiment [40] (Table 2).

XSL•FO
RenderX

Table 2. Descriptive summary of the selected studies.

Neurological disorders and studies	Study design	Sample size, n (etiology)	Demographics	Study aim	Major findings
MS[a]					
Hildebrand et al [51], 2021	Prospective cohort study	25 (MS)	Female: 17; male: 8; EDSS[b]: median 6.0 (IQR 5.5-6.0); RR[c]: 11; SP[d]: 8; PP[e]: 6	The study compared the sensitivity and false discovery rates of 3 fall detection methods: prospective paper fall calendars, real-time self-reporting, and automated detection using a body-worn device	Fall calendars likely underestimate fall frequency by approximately 40%, whereas the automated detector misses very few falls but likely overestimates the number of falls by approximately 1 per day
Tulipani et al [23], 2022	Cross-sectional methodological case-control study	37 (MS—nonfallers: 17; fallers: 21)	EDSS—nonfallers: mean 2.0 (SD 0.8); fallers: mean 3.3 (SD 1.4)	Participants completed the 30CST[f] in both supervised (laboratory) and unsupervised (home) settings while wearing a thigh-mounted triaxial accelerometer; in the unsupervised setting, they also performed 30CSTs every 2 h and rated their balance confidence and fatigue over 48 h	Short periods of instrumented, unsupervised monitoring, in addition to regular clinical assessments, could enhance the accuracy of fall risk prediction
Tulipani et al [41], 2020	Cross-sectional methodological case-control study	38 (MS—nonfallers: 17; fallers: 21)	EDSS—nonfallers: mean 2.3 (SD 1.2); fallers: mean 3.4 (SD 1.2)	Participants completed self-report measures and performed the 30CST, with triaxial acceleration data collected from sensors on the thigh and chest; the study assessed the relationship between accelerometer metrics and clinical measures and developed a logistic regression model to classify fall status	Accelerometer-derived metrics were linked to clinically relevant measures of disease severity, fatigue, and balance confidence during balance-challenging tasks, suggesting that inertial sensors could improve functional assessments and fall risk identification, with their simplicity aiding community-based monitoring
Meyer et al [48], 2022	Cross-sectional study with follow-up	38 (MS—nonfallers: 17; fallers: 21)	Male: 12; female: 26; EDSS—nonfallers: mean 2.3 (SD 1.0); fallers: mean 3.3 (SD 1.4)	Participants were categorized as fallers or nonfallers based on self-reported falls, underwent clinical assessments, completed balance and fatigue surveys, performed laboratory-based mobility tests while wearing MC10 BioStamp sensors to collect accelerometer and EMG[g] data, and continued to wear the sensors at home for 48 h to monitor acceleration	Deep learning models performed better than feature-based models when analyzing home data, particularly when using all walking bouts for deep learning and shorter bouts for feature-based models in individual bout evaluations; longer free-living walking bouts showed more pronounced differences between fallers and nonfallers compared to shorter bouts, and aggregating data from all free-living walking bouts resulted in the highest accuracy for fall risk classification
Meyer et al [37], 2021	Cross-sectional methodological case-control study	37 (MS—nonfallers: 19; fallers: 18)	Male: 11; female: 26; EDSS—mean 2.99 (SD 1.47)	The study distinguished fallers from nonfallers based on a 1-min walking task; participants were evaluated using wearable sensors to collect accelerometer data	The BiLSTM[h] deep learning model achieved the highest performance with an AUC[i] of 0.88, suggesting that deep learning using gait acceleration data is effective for classifying fall status
Kushner et al [52], 2023	Prospective cohort study	25 (MS)	Male: 8; female: 17; EDSS: median 6.0 (IQR 5.5-6.0); RR: 11; SP: 8; PP: 6	Participants were monitored over 8 wk using an inertial sensor equipped with a triaxial accelerometer and time-of-flight radio transmitter, communicating with beacons placed throughout the home, to assess associations between home locations and movement behaviors before falls compared to periods without falls	Movement metrics obtained from wearable sensors and smart-home tracking systems were correlated with fall risk, with increased pauses during walking and more complex, longer movement paths indicating higher fall risk

XSL•FO
RenderX

Neurological disorders and studies	Study design	Sample size, n (etiology)	Demographics	Study aim	Major findings
Arpan et al [44], 2022	Prospective cohort study	26 (MS—nonfallers: 13; fallers: 13)	EDSS—nonfallers: mean 4.3 (SD 0.23); fallers: mean 4.2 (SD 0.18)	The study computed the AUC values for each gait measure that differentiated fallers from nonfallers and then ranked values from highest to lowest to assess their effectiveness in predicting falls based on instrumented measures of mobility using a univariate model	Objective monitoring of gait and turning in daily life identified those at risk of future falls, with the pitch at toe-off identified as the most significant predictor, suggesting potential benefits from interventions targeting plantarflexion muscle strength, range of motion, and proprioceptive input
PD[j]					
Greene et al [46], 2021	Cross-sectional study with follow-up	1057 (PD [more impaired]: 15; PD [less impaired]: 26; controls: 1015)	Participants with PD (more impaired)—male: 10; female: 5; participants with PD (less impaired)—male: 17; female: 9; UPDRS-III[k]—participants with PD (more impaired): mean 15.1 (SD 9.6); participants with PD (less impaired): mean 22.56 (SD 10.25); healthy controls—male: 344; female: 671	The study analyzed data using 2 statistical approaches to predict fall counts: a previously reported fall risk assessment algorithm as well as elastic net and ensemble regression models	The study found a robust correlation between fall counts and a previously established inertial sensor–based fall risk estimate, as well as significant associations between fall counts and various individual gait and mobility parameters, suggesting that falls predicted from inertial sensor data during a basic walking task could serve as a promising digital biomarker for PD warranting further validation in clinical settings
Pallavi et al [35], 2023	Cross-sectional study	28 (PD: 14; healthy controls: 14)	Participants with PD—male: 11; female: 3; UPDRS: mean 35.5 (SD 19.4); H&Y[l]: mean 2.2 (SD 0.7); healthy controls—male: 11; female: 3	Participants were asked to walk overground on a 10-m pathway, both without turns and with 180° turns, under varying task conditions of increasing complexity (single-, dual-, and multiple-task conditions)	The knee flexion and gait-related indices strongly aligned with clinical measures of fear of falling, especially in the group with PD, providing valuable predictive information for clinicians
Silva de Lima et al [42], 2020	Prospective cohort study	4126 (PD: 2063; healthy controls: 2063)	Participants with PD—male: 996; female: 1067; healthy controls—male: 992; female: 1071	The study used a PERS[m], which consisted of a device—worn as a necklace—with multiple embedded sensors for automatic fall detection and manual fall reporting; fall events were reported either by a button push or automatically detected by the fall detector	Wearable sensors were used to collect fall data, revealing that PD nearly doubles fall incidence in everyday life, underscoring PD as a significant fall risk; in addition, monitoring fall events in >4000 participants using a wearable sensor connected to a PERS suggests the potential of body-worn sensors for continuous home monitoring
Sturchio et al [39], 2021	Prospective cohort study	26 (PD and OH[n])	Male: 19; female: 7; UPDRS-III: mean 40.1 (SD 13.4)	The study estimated the association of PD and OH with fall risk association with fall risk, capturing fall frequency through a diary over 6 mo	Kinematic measures, rather than clinical assessments, predicted falls in PD with OH, suggesting that orthostatic mean arterial pressure of ≤75 mm Hg may indicate a hemodynamic threshold associated with increased fall risk, advocating for proactive corrective interventions
Nouriani et al [43], 2023	Cross-sectional algorithm validation with prospective cohort	21 (PD: 11; healthy controls: 10)	UPDRS pull test: mean 0.56 (SD 0.96)	The study validated the activity recognition algorithm through video footage of participants with PD wearing sensors at home	The findings suggest that incorporating near-fall detection into home monitoring data significantly enhances the performance of fall prediction algorithm

XSL•FO
RenderX

Neurological disorders and studies	Study design	Sample size, n (etiology)	Demographics	Study aim	Major findings
Ullrich et al [49], 2023	Prospective cohort study	35 (PD—nonfallers: 25; fallers: 10)	Male: 26; female: 9; UPDRS—nonfallers: mean 12.9 (SD 6.2); fallers: mean 21.7 (SD 9.1); H&Y—nonfallers: mean 2.3 (SD 0.5); fallers: mean 2.9 (SD 0.5)	The study compared various data aggregation methods and machine learning models for predicting fall risk using gait parameters obtained from continuous real-world recordings and unsupervised gait tests	The findings indicate that combining 2 wk of real-world gait data provides the most accurate prediction of fall risk, surpassing predictions based solely on unsupervised gait tests (balanced accuracy of 68%), thereby advancing understanding in fall risk prediction
Ma et al [45], 2022	Prospective cohort study	51 (PD)	Male: 33; female: 18; UPDRS-III: mean 33.6 (SD 13.5); H&Y: mean 2.4 (SD 0.8)	Using wearable sensors, the study objectively evaluated gait characteristics in participants with PD who experienced falls; it also explored the relationship between spatiotemporal gait parameters, gait variability, and falls over a 6-mo follow-up period	Increased gait variability is a notable characteristic among individuals with PD who experience falls and proves more sensitive in identifying those at elevated risk of falls than spatiotemporal parameters
Song et al [47], 2025	Cross-sectional case-control study	111 (PD: 29; healthy controls: 18 [dataset III]; PD: 35; healthy controls: 29 [dataset IV])	Participants with PD—male: 20; female: 9; age: mean 71.1 (SD 8.1) y; healthy controls—male: 10; female 8; age: mean 71.6 (SD 6.7) y (dataset III); participants with PD—male: 22; female: 13; age: mean 61.6 (SD 8.9) y; healthy controls—male: 18; female 11; age: mean 57.9 (SD 7.0) y (dataset IV)	The study identified an abnormal gait pattern termed recessive weak foot, characterized by a discontinuous high-risk gait on the weak foot side, observed through weak foot feature space; to address this, the authors proposed a trainable threshold method to discriminate individuals with this pattern, thereby enhancing the model's generalization performance; the authors conducted feasibility and ablation studies on 2 self-established datasets and tested the compatibility on 2 published gait-related datasets of participants with PD	Guided by a customized index and the optimized adaptive thresholds, the method used by the authors effectively screened out individuals with recessive weak foot; specifically, after fine adaptation, the individual-specific models could achieve accuracies of 87.5% and 73.6% on an enhanced dataset; compared to the baseline, the proposed 2-stage model demonstrated improved performance, with an accuracy of 85.4% and sensitivity of 87.5%; in the dataset of participants with PD, the authors' method mitigated potential overfitting from low feature dimensions, increasing accuracy by 4.7%
AD[o]					
Huang et al [50], 2022	Cross-sectional study	103 (older adults with CI[p], including those with aMCI[q] and mild AD: 48; cognitively healthy older adults: 55)	Participants with CI—male: 21; female: 27; age: mean 65.7 (SD 5.2) y; cognitively healthy participants—male: 15; female: 40; age: mean 67.1 (SD 4.1) y	The study primarily examined the association between memory deficit and increased fall risk; it also explored the underlying neuroanatomical linkage of this association in older adults with aMCI and mild AD	The authors' findings suggested that memory deficit was associated with increased fall risk in older adults with aMCI and mild AD; the association might be mediated by the atrophy of the medial temporal, frontal, and parietal lobes; in addition, increased fall risk, tested by timed up and go time, heel stride angles, and stride speed, might be objective and convenient kinematics markers for dynamic monitoring of both memory function and fall risk
Stroke					
Taylor-Piliae et al [36], 2016	Cross-sectional feasibility study	20 (stroke: 10 [ischemic stroke: 6; hemorrhagic stroke: 4]; healthy controls: 10)	Participants with stroke: female: 7; male: 3; healthy controls: female: 9; male: 1; poststroke duration: mean 42 (SD 25) mo	The study assessed the feasibility of using a kinematic motion sensor (PAMSys) to monitor fall risk and gait in community-dwelling survivors of stroke by evaluating the acceptability of wearing the sensor for 48 h, identifying fall risk indicators and gait parameters, and comparing these metrics with data from age-matched nonfrail controls	Survivors of stroke found 48 h of continuous PAMSys monitoring highly acceptable, suggesting that in-home wearable technology could be valuable for monitoring fall risk and gait, potentially aiding in recovery efforts

XSL•FO
RenderX

Neurological disorders and studies	Study design	Sample size, n (etiology)	Demographics	Study aim	Major findings
Botonis et al [38], 2022	Cross-sectional study	35 (stroke: 20; healthy controls: 15)	Participants with stroke—male: 10; female: 10; post-stroke duration: mean 7.47 (SD 4.58) y; healthy controls—male: 8; female: 7	The study investigated whether population-specific training data and modeling parameters were necessary to predetect falls in individuals with chronic stroke	The findings underscore the significance of population-specific sensitivity, use of nonfall data, and optimal lead time for machine learning–based preimpact fall detection tailored for individuals with stroke, suggesting a need for the inclusion of individuals with neurological impairments in model development for effective fall detection in other populations considered to be at high risk
Abdollahi et al [40], 2024	Prospective cohort study	21 (stroke)	Male: 11; female: 10; age: mean 66 (SD 10) y	The study aimed to develop machine learning models using inertial sensors to objectively classify fall risk in survivors of stroke	The random forest model achieved 91% accuracy using dual-task balance sway and timed up and go walk time features; single thorax sensor models performed similarly to multisensor models; balance and timed up and go best predicted fall risk

[a]MS: multiple sclerosis.

[b]EDSS: Expanded Disability Status Scale.

[c]RR: relapsing-remitting.

[d]SP: secondary progressive.

[e]PP: primary progressive.

[f]30CST: 30-second chair stand test.

[g]EMG: electromyography.

[h]BiLSTM: bidirectional long short-term memory.

[i]AUC: area under the receiver operating characteristic curve.

[j]PD: Parkinson disease.

[k]UPDRS-III: Unified Parkinson's Disease Rating Scale, part III (motor examination).

[l]H&Y: Hoehn and Yahr.

[m]PERS: personal emergency response system.

[n]OH: orthostatic hypotension.

[o]AD: Alzheimer disease.

[p]CI: cognitive impairment.

[q]aMCI: amnestic mild cognitive impairment.

Classification of the Results

We classified our results according to the type of motor task (supervised vs unsupervised), the biomechanical parameters extracted, and the technological equipment used as well as the artificial intelligence (AI) approach applied in the included studies.

Motor Task for Fall Risk Detection

Of the 19 included studies, 12 (63%) assessed fall risk through supervised motor tasks, administered in clinical or laboratory settings, while the remaining 7 (37%) evaluated fall risk through unsupervised gait tasks recorded in real-world environments.

Supervised Motor Tasks

Generally, supervised motor tasks consisted of instrumented versions of the TUG test [39,41,45,46,48], the 30-second chair stand test (30CST) [23,41,48], and simulated daily activities [37,39]. However, Tulipani et al [23] compared supervised 30CST with unsupervised 30CST in a home-based setting. The authors reported a substantial statistical difference between the 2 types of motor tasks, suggesting that the laboratory setting could influence the fall risk rate. Some of the studies (5/19, 26%) [37,39,41,45,48] administered a specific assessment protocol to quantify the risk of falls in patients with neurological disorders, while the studies by Pallavi et al [35] and Greene et al [46] quantified fall risk using solely TUG or gait tasks. Similarly, the study by Song et al [47] analyzed different datasets in which the only motor task consisted of walking along a 20-meter corridor for at least 2 minutes under observation. Interestingly, Sturchio et al [39] conducted an instrumented assessment of activities of daily living (ADLs) in a standardized home-like environment, in addition to clinical tests (eg, the TUG test, the 2-minute walk test, and other standard measures; Table 2). The TUG test was also used to assess gait function in patients with AD [50]. Similarly, Meyer et al [37], acquired sensor data from simulated daily activities in addition to several standard functional assessments. In the study by Abdollahi et al [40], patients with stroke completed clinical tasks, including balance tests (eyes open and closed), the TUG test, the 10-meter walk test, and the sit-to-stand test. Each was also performed as

XSL•FO
RenderX

a motor-cognitive dual task, involving backward counting during movement. Finally, Botonis et al [38] evaluated fall risk in patients with stroke through balance response reactivity in a laboratory setting.

Unsupervised Motor Tasks

Of the 19 included studies, 7 (37%) [36,42-44,49,51,52] assessed fall risk using unsupervised recordings in a real-world environment. In general, these assessments were conducted during ADLs [36,43,51] or gait tasks [44,49,52]. Notably, Silva de Lima et al [42] detected fall events automatically using a wearable sensor, without monitoring ADLs or gait.

Fall Risk Biomechanical Parameters

We analyzed features extracted from wearable sensors related to the quantification of supervised or unsupervised motor tasks in patients with MS, PD, or stroke. Several of the studies (10/19, 53%) [35,36,39-41,45,47-49,52] extracted specific gait parameters from wearable sensors. These parameters consisted mainly of spatiotemporal parameters of gait [35,45,49], anterior-posterior and medial-lateral acceleration [41], lower limb joint angles, and cadence [35]. In particular, Arpan et al [44] extracted specific gait parameters to differentiate fallers from nonfallers. These biomechanical parameters reflected gait quality (eg, gait speed, stride length, cadence, and swing) as well as turning function [44]. In addition, Song et al [47] extracted 44 gait features mainly from plantar pressure data, focusing on the weaker foot. These features were used to identify a novel gait dysfunction, termed recessive weak foot, which was characterized by discontinuous high-risk gait patterns. Furthermore, Abdollahi et al [40] and Huang et al [50] extracted spatiotemporal gait parameters derived from the clinical tests (the TUG test and the 10-meter walk test). Only 2 (11%) of the 19 studies [36,39] extracted balance features in addition to gait parameters. In particular, Sturchio et al [39] evaluated oscillations in the center of pressure and center of mass, the frequency of oscillations, and jerkiness in patients with PD. The authors also evaluated upper limb range of motion during motor tasks. Similarly, Arpan et al [44] evaluated upper body kinematics alongside gait quality through spatiotemporal parameters and additionally calculated specific parameters during gait turning. Gait turning and postural transfers were also analyzed by Greene et al [46], while Tulipani et al [23,41] focused on specific biomechanical parameters during the 30CST (Table 3).

As shown in Table 3, the majority of the included studies (18/19, 95%) analyzed gait parameters, such as spatiotemporal features. These features were mostly considered in patients with PD or MS. In addition, some of the studies (7/19, 37%) extracted biomechanical features from clinical tests, such as 30CST and TUG (eg, TUG time, turning activity, 30CST average time, and repetitions). Concerning patients with stroke, it was difficult to define a common pattern of extracted biomechanical features because only 3 studies were available; however, both considered fall characteristics and parameters related to postural changes, similar to those examined in studies involving patients with PD or MS [43,52].

XSL•FO
RenderX

Table 3. Biomechanical features—analyzed to quantify fall risk in patients with multiple sclerosis, Parkinson disease, Alzheimer disease, or stroke—extracted from wearable sensors in the included studies.

Neurological disorders and studies	Motor task	Parameters analyzed
Multiple sclerosis		
Hildebrand et al [51], 2021	Unsupervised tasks during daily life activities	• Acceleration signals from triaxial accelerometers mounted on the waist • TUGa measurements for assessing balance and mobility in participants • Location tracking using GPS and time-of-flight sensors to associate falls with specific movements or environments

Tulipani et al [23], 2022	30CST[b]—the participant was encouraged to complete as many full stands as possible within 30 s	• Metrics for sit-to-stand and stand-to-sit transitions:
		  • Sit-to-stand time (average, maximum, and minimum time across 30CST repetitions)
		  • Stand-to-sit time (average, maximum, and minimum time across 30CST repetitions)
		  • Total number of 30CST repetitions

Tulipani et al [41], 2020	Participants performed a series of functional assessments, including 1 trial each of the 30CST, T25W[c], and TUG, as well as a 30-s static standing trial with instructions to maintain a tall posture with their feet facing forward, wearing inertial sensors on the anterior right thigh and chest	• Metrics for sit-to-stand and stand-to-sit transitions:
		  • Mean and coefficient of variation of sit-stand time and stand-sit time
		  • Peak cranial-caudal, anterior-posterior, and medial-lateral acceleration during transitions

Meyer et al [48], 2022	Supervised tasks completed in the following order: TUG; T25W; 30CST; lying-to-standing transition; 3 separate 2-minute standing tests (tandem standing, feet shoulder width apart with eyes open, and feet shoulder width apart with eyes closed); 1-min hallway walk at a self-selected pace, including 1 turn; 30-s normal standing; 30-s upright sitting; 30-s slouch sitting; and 30 s each lying on the back, left side, right side, and prone	• Gait parameters extracted:
		  • Stance time, swing time, stride time, and variability measures (coefficient of variation of stride and duty factor)
		  • Nonlinear measures (entropy ratio and Lyapunov exponent)
		  • Root mean square of the anterior-posterior acceleration
		  • Medial-lateral frequency dispersion

Meyer et al [37], 2021	Data from these sensors were recorded during a variety of simulated daily activities and several standard functional assessments	• Spatiotemporal parameters of gait stride
Kushner et al [52], 2023	Unsupervised tasks during daily life activities (eg, gait quality and turning)	• Gait initiation time • Spatiotemporal parameters of gait (eg, average step length, average walking speed, and percentage of time spent turning while walking) • Time spent moving • Movement length • An entropy-based metric quantifying movement complexity through transitions between rooms • Mean walking path length inside the home

Arpan et al [44], 2022	Unsupervised tasks during daily life activities (eg, gait quality and turning)	• Instrumented measures of mobility as predictors of future falls:
		  • Pitch at toe-off
		  • Gait speed
		  • Stride length
		  • Double support
		  • Swing
		  • Pitch at initial contact
		  • Turn angle

Parkinson disease		

XSL•FO
RenderX

Neurological disorders and studies	Motor task	Parameters analyzed
Greene et al [46], 2021	Supervised TUG test, instrumented with inertial sensors	• 71 different calculated parameters quantifying gait, mobility, turning, and transfers, along with a statistical fall risk estimate and frailty estimate based on inertial sensor data: • TUG time • Gait velocity • Fall risk estimate • Frailty estimate • Mobility impairment scores
Pallavi et al [35], 2023	Gait tasks: single, dual, and multiple	• Gait parameters: • Knee flexion during heel-strike event • Knee flexion during toe-off event • Cadence • Double limb support time
Silva de Lima et al [42], 2020	Unsupervised; fall events were collected either automatically using the wearable fall detector or were registered by a button push on the same device	• Fall event (manually detected) • Fall event (automatically detected) • Changes in height • Changes in orientation • Impact during fall
Sturchio et al [39], 2021	Participants performed the following tasks: lying-to-standing test (standing up without assistance after 10 min of supine resting and keeping the upright position for 5 min), tests of gait and postural stability (TUG, 2MWT[d], and sway [eyes opened and eyes closed]), and ADLs[e] conducted in a standardized home-like environment	• Balance parameters: • Oscillations in the CoPf • Oscillations in the CoMg • Jerk sway (jerkiness) • Root mean square sway (magnitude of acceleration) • CFh sway (frequency of oscillation) • Gait parameters: • Turn duration • Total duration • Peak turn velocity • Waist sway during TUG • Gait speed • Stride length • Number of steps during turning • Range of motion of upper limbs • Cadence
Nouriani et al [43], 2023	Unsupervised tasks during daily life activities	• IMUi values: • Time spent lying down per day (lie-down frequency, ie, lying duration/total time) • Total number of ambulatory bouts at home • Frequency of sitting at home (eg, sitting duration/total time) • Walking frequency at home • Peak acceleration of the chest at home
Ullrich et al [49], 2023	Unconstrained real-world gait analysis and unsupervised gait tasks	• Spatiotemporal gait parameters, including the following: • Stride time • Stance time • Swing time • Stride length • Gait speed • Initial contact foot angle • Final contact foot angle • Maximum foot lift
Ma et al [45], 2022	Participants were required to complete a 7-meter TUG test twice (stand up from an armless chair, walk at own comfortable pace for 7 meters, turn 180°, then walk back and sit down)	

XSL•FO
RenderX

Neurological disorders and studies	Motor task	Parameters analyzed

		• Spatiotemporal gait parameters, including the following:
		  • Stride length
		  • Gait cycle time
		  • Gait phase (swing and double support percentages)
		  • Range of motion of the trunk in the sagittal plane
		  • Gait variability measures

Song et al [47], 2025	Participants were asked to walk for at least 2 min consecutively at their normal gait and speed along a 20-meter-long corridor	• 44 CoP features based on recessive weak foot were extracted:
		  • Weak and single foot features: mean and SD of medial-lateral anterior-posterior CoP, mean and SD of resultant distance, total excursions, and confidence circle area
		  • Symmetry-based features:
		  • Gait asymmetry, similarity, and divergence
		  • Temporal consistency–based features:
		  • Gait inconsistency, sequential similarity, and sequential Jensen-Shannon divergence

Alzheimer disease		
Huang et al [50], 2022	Supervised TUG test	• TUG test time • Heel-strike angles • Stride speed • Cadence
Stroke		

XSL•FO
RenderX

Neurological disorders and studies	Motor task	Parameters analyzed

Taylor-Piliae et al [36], 2016	Unsupervised tasks during daily life activities	• Parameters related to postural transitions (eg, sit-to-stand, standing-to-sitting, lying-to-sitting, and sitting-to-lying tasks):
		  • Duration (s)
		  • Number of unsuccessful attempts
		• Gait parameters (detected based on the peaks in the vertical accelerometer data, following predefined conditions):
		  • Number of steps taken
		  • Speed of walking (meters per s)
		  • Duration of walking activities (percentage of total activity)

Botonis et al [38], 2022	Participants were instructed to respond to loss of balance using any natural technique, such as using an arm to catch themselves or taking multiple steps, to encourage realistic behavior	• Preimpact data:
		  • Fall impact and acceptable intervention time
		• Statistical features from raw IMU signals, including the following:
		  • Minimum
		  • Median
		  • Maximum
		  • IQR
		  • SD
		  • Skew
		  • Kurtosis

Abdollahi et al [40], 2024	Balance test (stand as still as possible for 30 s with eyes open or closed on a firm surface), TUG, 10MWT[j], STS[k], and motor-cognitive dual tasks (repetition of the aforementioned tasks while counting backward from 200 by 10s)	• Biomechanical parameters extracted during clinical tests:
		  • Walk time
		  • Step cadence
		  • Mean and SD of swing total, single support, and stride duration
		  • Thorax, thigh, and pelvis linear acceleration
		  • Thorax and pelvis angle velocity
		  • TUG and STS time
		  • TUG: walk toward cone, turn around the cone, walk toward chair, turn and sit, steps toward cone, steps toward chair, cadence toward cone, and cadence toward chair

[a]TUG: timed up and go.
[b]30CST: 30-second chair stand test.
[c]T25W: timed 25-foot walk.
[d]2MWT: 2-minute walk test.
[e]ADL: activity of daily living.
[f]CoP: center of pressure.
[g]CoM: center of mass.
[h]CF: centroidal frequency.
[i]IMU: inertial measurement unit.
[j]10MWT: 10-meter walk test.
[k]STS: sit-to-stand.

Technological Equipment, Data Processing, and Analysis Approaches

This subsection describes the technological equipment and data processing methods used across the included studies. Of the 19 studies, 7 (37%) [23,37,41,44,48,51,52] focused on fall risk assessment in patients with MS using wearable sensors. Of these 7 studies, 4 (57%) used the MC10 BioStamp system, which integrates triaxial accelerometers, gyroscopes, and electromyography (EMG) sensors to capture a wide range of motion data [23,37,41,48]. Tulipani et al [41] examined accelerometer metrics from thigh and chest sensors and applied logistic regression models to distinguish fallers from nonfallers, achieving an area under the receiver operating characteristic curve (AUC) of 0.78 and an accuracy of 74%. In the study by Tulipani et al [23], a single sensor placed on the thigh was used to record data during tests, and the authors used receiver operating characteristic analysis to optimize classification

XSL•FO
RenderX

thresholds. This improved predictive performance in unsupervised settings, with an AUC of 0.79 and an accuracy of 78.4%.

Meyer et al [37] used BioStamp in conjunction with APDM Opal wearable sensors. Data were acquired from triaxial accelerometer sensors secured to the sternum (below the clavicle) and to the anterior portion of the right thigh. In addition, the authors recorded data from accelerometers and angular rate gyroscopes secured to the lower sternum, lower back (at the belt line), and anterior right and left shanks. The authors demonstrated that a bidirectional long short-term memory (LSTM) deep neural network could identify patients with MS who had recently fallen, achieving an AUC of 0.88 and an accuracy of 86%, based on data collected from 2 wearable sensors during a 1-minute task [37].

In the study by Meyer et al [48], accelerometer and EMG data were collected from the right and left tibialis anterior, while accelerometer and angular rate gyroscope data were collected from the chest and lower back as well as bilaterally from the anterior thighs, proximal lateral shank, and dorsal aspect of the feet. EMG data were collected to allow the investigation of foot drop. The models tested in this study showed that short walking bouts were the best predictor of fall risk in free-living conditions (AUC=0.63). By contrast, medium and long walking bouts had lower predictive performance, with AUCs of 0.52 and 0.54, respectively. DL models, especially LSTM models that used up to 22 strides, outperformed traditional feature-based models, with the LSTM model achieving an AUC of 0.76, showing its robustness in discerning fall risk from complex gait patterns. Hildebrand et al [51] and Kushner et al [52] used an automated fall detection system that integrated a body-worn IMU (placed on the waist) with a button for self-reporting falls and an ML algorithm trained to evaluate falls in real-world environments in patients with MS [53]. Specifically, the algorithm used an autoencoder that identifies individuals at increased risk of falling using reconstruction errors in accelerometer signals, followed by a hyper-ensemble of balanced random forests trained on acceleration and movement features. Hildebrand et al [51] compared 3 methods for detecting falls in patients with MS: traditional paper fall calendars, real-time self-reporting using a body-worn device, and automated fall detection using the same device. The results showed that paper calendars had a sensitivity of 61.4% and a false discovery rate of 6.7%, indicating that they likely underestimated the number of falls. By contrast, the automated detection system had a high sensitivity of 92.1% but also a high false discovery rate of 91.9%, indicating that it likely overestimated fall occurrences. The study highlighted the challenges of balancing high sensitivity with minimizing false positives in real-world applications. Kushner et al [52] focused on extracting movement complexity metrics from wearable sensors. The study used a k-nearest neighbors ML algorithm for room detection and applied complex statistical methods to estimate movement-based fall risk. The findings indicated that movement complexity, measured by the entropy of room transitions, was notably higher in the periods leading up to falls, suggesting that increased movement complexity is a strong predictor of fall risk. Arpan et al [44] used a different set of wearable technologies to passively monitor gait and turning behaviors in daily life to predict falls in patients with MS. This setup involved triaxial accelerometers, gyroscopes, and magnetometers (Opal sensors) placed on the top of the foot and integrated into instrumented socks, along with a sensor worn on the lower lumbar region. The data collected from these sensors were processed using an unscented Kalman filter technique to fuse accelerometer and gyroscope data and identify walking bouts and turning behaviors. The raw data were further processed using commercial gait analysis algorithms in Mobility Lab to derive spatiotemporal measures of gait and turning. The study found that reduced plantarflexion during toe-off, reflected by the pitch angle of the foot during the push-off phase of walking, was the most significant predictor of falls, achieving an AUC of 0.86, indicating strong predictive power [44].

Of the 19 studies, 8 (42%) [35,39,42,43,45-47,49] focused on integrating advanced wearable sensors and complex data analysis methods to monitor and predict fall risks in patients with PD. Silva de Lima et al [42] used a wearable sensor worn as a necklace (Philips Lifeline FD100) to monitor falls in patients with PD in their home environments. Although the study did not predict falls, it revealed a significantly higher incidence of falls in patients with PD compared to controls, demonstrating the pronounced fall risk in this population [42]. Song et al [47] analyzed PD gait data from intelligent footwear with 16 plantar pressure sensors embedded in each insole. The authors proposed a 2-stage model that first identifies individuals with PD with irregular gait patterns, referred to as recessive weak foot, using an adaptive threshold based on the distribution difference index. In the second stage, separate personalized ML models (such as support vector machines, logistic regression, and decision trees) are trained for individuals with recessive and dominant weak foot patterns, enabling a more accurate and individualized assessment of fall risk.

Of the 19 studies, 3 (16%) used compact gait analysis systems to collect comprehensive kinematic data. Ma et al [45] and Sturchio et al [39] used the Mobility Lab system, which includes 6 wearable sensors (accelerometers, gyroscopes, and magnetometers) placed on the feet, wrists, sternum, and lumbar region. Ullrich et al [49] used the Mobile GaitLab system, with sensors positioned on the instep of both the left and right shoes. Similarly, Huang et al [50] used the JiBuEn gait analysis system, which consists of wearable sensors embedded in specially designed shoes. Ma et al [45] focused on gait variability as a predictive marker for falls in patients with PD, capturing metrics such as stride length, gait cycle time, and range of motion of the trunk in the sagittal plane. The authors found that variability in the trunk range of motion during walking was a significant independent risk factor for falls in patients with PD, achieving a prediction accuracy with an AUC of 0.75. The overall logistic regression model, which included multiple gait parameters, predicted falls with a higher accuracy with an AUC of 0.84 [45]. Sturchio et al [39] focused on fall risk in patients with PD with orthostatic hypotension and found that waist sway, jerkiness, and postural sway were more predictive of falls than traditional clinical assessments, achieving an AUC of ≥0.81 for predicting fall risk. Ullrich et al [49] focused on fall risk prediction in patients with PD by analyzing real-world gait data. The study used ML algorithms, notably random forest classifiers, to

XSL•FO
RenderX

analyze the aggregated gait data for fall risk prediction. The highest predictive performance was achieved using participant-wise aggregated data, resulting in a balanced accuracy of 74%, with a sensitivity of 60% and a specificity of 88% [49].

Greene et al [46] explored the utility of wearable sensors during the TUG test to develop a digital biomarker for predicting fall counts. The study applied 2 statistical approaches to predict fall counts: a fall risk assessment algorithm reported in previous studies by the same team [54,55] and an ensemble model combining elastic net and regression models using Poisson regression. The research findings suggest a strong association between the fall counts predicted from the inertial sensor data and actual falls [46]. Nouriani et al [43] validated an activity recognition algorithm and developed novel behavioral biomarkers to assess fall risk in patients with movement disorders, focusing on PD and normal pressure hydrocephalus. The study used 5 IMUs strategically placed on the chest, upper legs, and lower legs of participants, capable of measuring acceleration, angular rates, and orientation. The core of the study's methodology involved a DL-based activity recognition architecture using a convolutional neural network combined with LSTM cells. This advanced ML approach allowed for the accurate detection of complex activities and postural transitions that are critical indicators of fall risk. The algorithm's performance was validated against video [43], achieving high sensitivity (>95%) in detecting activities and an 80% sensitivity for near falls.

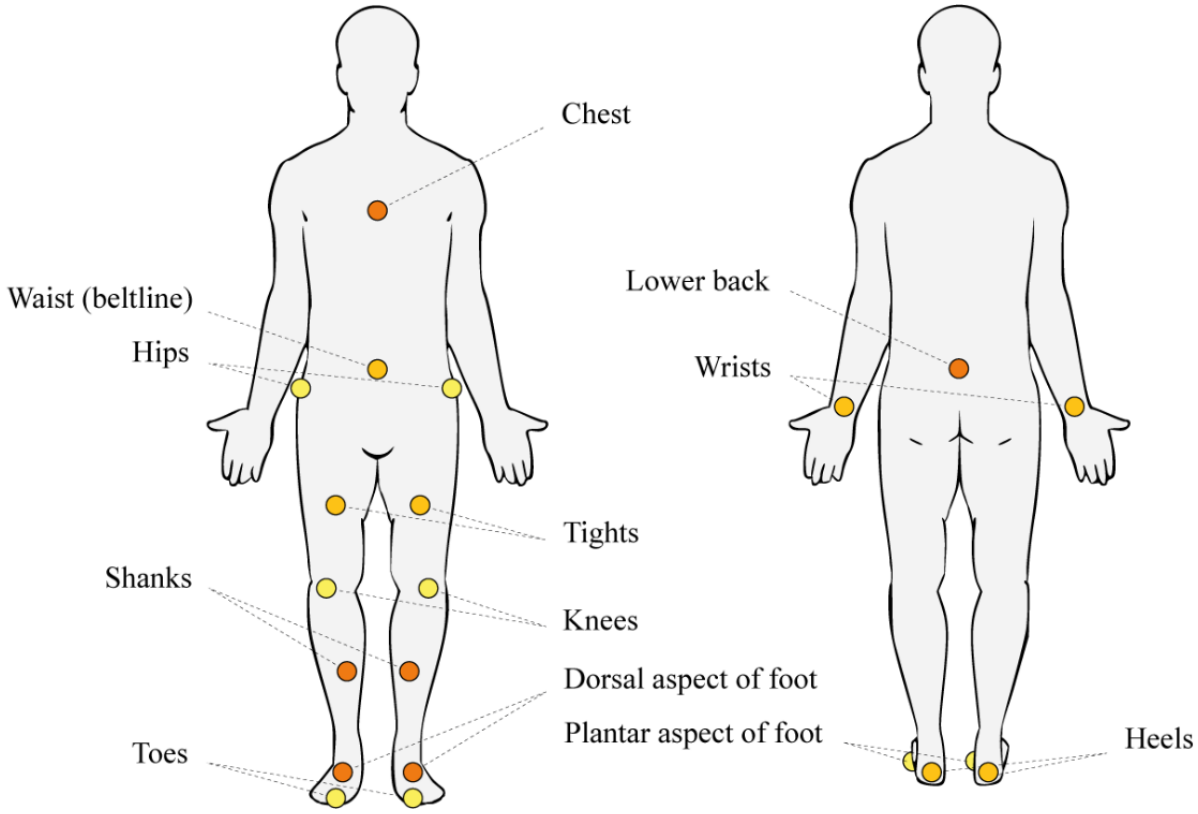
Finally, Pallavi et al [35] used a novel wearable sensor (SmartWalk) to evaluate knee flexion and gait indices during walking tasks under different cognitive load conditions. The study found that variability in knee flexion increased significantly under dual-task conditions, which were closely associated with a fear of falling. The findings indicate significant increases in the variability of knee flexion during walking tasks, particularly under dual- and multiple-task conditions, which pose higher cognitive demands and thus exacerbate gait and postural disturbances in patients with PD. A strong relationship between measured indices and clinical assessments of fear of falling was found using statistical analyses, including k-means clustering and correlation assessments [35].

Of the 19 studies, 3 (16%) [36,38,40] focused on fall assessment in survivors of stroke. Botonis et al [38] investigated the efficacy of wearable airbag technology paired with ML models to prevent falls. The device included 3 IMU sensors (on the hips and lower back) and used adaptive boosting classifiers to differentiate fallers from nonfallers. Two models were developed: one trained on survivors of stroke and the other on control data. The stroke-specific model significantly outperformed the control model in detecting true falls, especially in classifying complex fall movements in anterior-posterior directions. In addition, training the model with daily activity data enhanced its accuracy in fall classification [38]. Taylor-Piliae et al [36] explored the feasibility of using a wearable sensor technology (PAMSys) for objective fall risk and gait monitoring in community-dwelling survivors of stroke. The triaxial accelerometer was worn in a midsternal pocket of a custom T-shirt, continuously recording movement data focused on postural transitions, such as the sit-to-stand test. Data on steps, speed, and movement transitions were recorded over 48 hours and compared with controls. The analysis showed that survivors of stroke exhibited worse fall risk indicators, taking longer to change posture and having more failed sit-to-stand attempts than controls [36]. In another study [40], 8 IMUs (placed on feet, shanks, thighs, pelvis, and thorax) per participant were used. The authors analyzed the gait kinematic data extracted from IMUs with ML algorithms. In particular, they used classification algorithms, including random forest, support vector machine, and logistic regression. Among these, the random forest achieved 91% accuracy using just 2 features: thorax mediolateral acceleration during dual-task balance and TUG walk time.

The distribution of sensors reveals a concentration on the lower limbs and trunk (Figure 3).

XSL•FO
RenderX

Figure 3. Sensor locations used across the included studies. Marker colors indicate the frequency of use—yellow: n≤2; light orange: n=3-5; and dark orange: n≥6.



Discussion

Principal Findings

In this systematic review, we emphasize the important role of wearable technologies in performing objective and quantitative evaluations of fall risk among patients with neurological disorders. While the studies included were generally well conducted, we noted considerable variability in sensor placement, the types of wearable sensors, motor tasks, and assessment methodologies. In addition, the lack of a standardized approach to fall risk assessment using wearable sensors likely reflects the field's ongoing technological evolution. To the best of our knowledge, this is the first systematic review analyzing the biomechanical features detected from wearable sensors for fall risk assessment in patients with PD, MS, AD, or stroke. Other studies [21] have investigated the role of inertial sensor–based and insole-based wearable sensors to evaluate fall risk in older adults. In line with our results, these studies identified several biomechanical gait parameters, including spatiotemporal parameters, balance features (center of pressure trajectory), and ADLs that were registered to detect fall risk. However, in our systematic review, we found some differences in the assessment of the risk of falling related mainly to the characteristics of the neurological disorders considered. Furthermore, we noticed differences between the studies concerning technological equipment.

Moreover, another systematic review [56] examined the use of wearable sensors to detect freezing of gait and falls in patients with PD. Consistent with our findings, the authors identified the most common sensor placements, such as the thigh, ankle, and lower back, which align with those reported in our review (Figure 3). However, our study presents some key differences: we included a broader range of neurological disorders (eg, MS, AD, and stroke), focused specifically on fall detection and fall risk assessment, and incorporated more recent literature.

Pathological and Biomechanical Gait Characteristics Leading to Falls

Falls are very common in individuals with neurological disorders and are influenced by various aspects and features of gait. According to Zhou et al [57], gait variables related to pace (eg, stride and velocity), rhythm (eg, stride duration as well as stance and swing time), variability (eg, SD of ankle dorsiflexion at heel strike), and spatial parameters (eg, stride length, plantar flexion at toe-off, and ankle dorsiflexion at heel strike) can help classify individuals with neurological disorders as fallers. However, neurological disorders are heterogeneous, and each condition is associated with specific gait features that could lead to an increased fall risk; for instance, in patients with MS, although the location, number, and size of MS lesions vary among individuals, certain common characteristics in gait deficits associated with MS have been observed. Patients with MS [58] may exhibit shorter steps, reduced cadence, prolonged double support time, and increased swing phase time. Moreover, some studies have emphasized that patients with MS have reduced joint motion at the hip and ankle and increased joint motion at the knee, with a hyperextension that is more pronounced during midstance [59]. Falls in this patient population can occur early after disease onset; and some factors increase the risk, such as poor visual acuity, impaired postural control, and altered proprioception. However, the studies

XSL•FO
RenderX

included in this systematic review analyzed only specific biomechanical metrics related to clinical tests (eg, 30CST) [23,41] or analyzed spatiotemporal features of gait and turning angle [37,44,52]. According to Tulipani et al [41], the 30CST could be a useful test to discriminate "fallers" and "nonfallers" in patients with MS, as it can detect balance impairments and muscle fatigue, both of which are associated with falls. Another important aspect is that remote fall risk monitoring using wearable sensors has the potential advantage of augmenting clinical visits to gain a broader perspective of patient function, symptom fluctuation, and fall risk. In patients with PD, a critical factor that is associated with falls is freezing of gait. However, it is not the only risk factor involved in falls. People with PD can experience bradykinesia, muscle weakness, and body rigidity, all of which contribute to slower gait and are closely linked to fall rates [60]. According to Pelicioni et al [61], reduced quadriceps strength, poor balance and postural stability, and reduced mobility (as indicated by slower TUG test performance) could explain the increased prevalence of balance-related falls. These results are in line with those reported in the included studies; for example, Greene et al [46] found strong associations between fall counts and several individual gait and mobility parameters, including gait variability, during the TUG test. Gait variability refers to fluctuations in gait patterns, and in patients with PD, it reflects a loss of consistency in the ability to produce a steady gait rhythm, thus presenting a likelihood of a fall. In addition, Ma et al [45] found that patients with PD who experienced falls exhibited greater range of motion of the trunk in the sagittal plane than nonfallers. This aspect was identified by the authors as an independent risk factor for falls in patients with PD; and it could be used to predict falls, especially when it is combined with age, gender, and other gait parameters. From these results, it seems that kinematic gait parameters are accurate in detecting fall risk, as also confirmed by Sturchio et al [39]. The authors found that kinematic data extracted from wearable sensors showed greater predictive accuracy for falls than the Hoehn and Yahr scale, which is considered the most robust clinical predictor of falls in individuals with PD. Regarding patients with AD, Huang et al [50] reported that impaired memory is strongly associated with an increased fall risk in individuals with MCI and AD. Specifically, the authors identified TUG time, stride speed, and heel-strike angles as effective parameters for assessing fall risk in this population. However, our review identified only a single study addressing this critical aspect, highlighting a clear gap in the literature. Future research should further investigate this area to determine the most relevant biomechanical and cognitive features for fall risk assessment in patients with AD. Differently from neurodegenerative disorders (eg, MS, PD, and AD), in patients with stroke, both balance and gait deficits related to falls were observed. According to Weerdesteyn et al [62], stroke-related balance deficits include postural instability and lack of coordination in response to both self-induced and external balance perturbations. Gait deficits related to fall risk include reduced propulsion at push-off, decreased hip and knee flexion during the swing phase, and reduced stability during the stance phase. However, the selected studies [36,38] did not report specific biomechanical features related to falls in patients with stroke. In future studies, it could be essential to select and extract specific biomechanical features related to falls in this patient population.

Technological Equipment: Considerations and Limitations

Fall risk assessment and fall detection have been investigated using wearable technologies. Among these, accelerometers, IMUs, and insole-based systems were mostly used in the clinical context; for instance, triaxial accelerometers and inertial sensors were widely used in the included studies [24]. They measure body movement by detecting speed changes and calculating the resultant acceleration from the displacement of a mass element. In rehabilitation research, the most commonly used types of accelerometers include strain gauge, capacitive, piezoresistive, and piezoelectric sensors. These accelerometers typically have 1 to 3 sensing axes, enabling motion detection in 1D, 2D, or 3D. From our literature analysis [23,37,41,44,51,52], it emerged that triaxial accelerometers were combined with gyroscopes and EMG sensors to capture a wide range of motion data. In particular, gyroscopes calculate changes in angular motion by detecting Coriolis forces, which are proportional to the rate of angular rotation of the limb. Integrating accelerometer and EMG data from the lower limb provides more insightful findings, especially in the context of neurorehabilitation; for example, Meyer et al [48] collected EMG, accelerometer, and gyroscope data that allowed the investigation of foot drop, which has been identified as one of the most frequent causes of falls in individuals with MS. Interestingly, Tulipani et al [41] found significant associations between accelerometer-derived metrics and clinical assessments, notably in balance confidence and fatigue. These findings indicate that wearable sensors can offer a real-time, objective measure of fall risk that aligns well with traditional clinical evaluations. Moreover, we found that the studies on fall risk assessment in PD (10/19, 53%) primarily focused on integrating advanced sensor technologies and complex data analysis methods to monitor, predict, and manage fall risks in patients with PD. Some of these studies (2/19, 11%) used a full-body wearable motion sensor system consisting of 6 sensors (each including a triaxial accelerometer, a gyroscope, and a magnetometer) to detect kinematic gait data. By contrast, Silva de Lima [63] developed a triaxial accelerometer combined with a barometer, which automatically detected falls and allowed users to manually report falls via a button press. This dual functionality enabled the collection of accurate real-time data on fall events.

According to a previous review by Ferreira et al [64], studies involving wearable sensors for fall risk assessment mainly used 1 accelerometer, which underlines the importance of the use of acceleration data in interpreting scores from standard clinical scales. Data from the wearable sensors were used to detect prefall and unbalanced situations, enabling the identification of fall risk events. This approach aims to reduce short-term fall risk by allowing real-time daily monitoring of participants and providing immediate feedback when a fall risk event is detected.

In this context, some of the studies (7/19, 37%) performed the fall risk assessment in an unsupervised setting. Compared to laboratory settings, real-time fall risk prediction during ADLs is more suitable for detecting fall risk events. This approach

XSL•FO
RenderX

allows continuous monitoring and enables alert systems to notify participants when fall risk events are detected. However, Meyer et al [48] found that the best performance of the ML model used to predict fall risk was achieved with laboratory data, indicating that controlled environments yield more reliable data for assessing fall risk. Supporting this, some studies have found that specific biomechanical features of gait (ie, speed) can vary depending on whether they are analyzed in the home or laboratory setting; for instance, Carcreff et al [65] found that children with cerebral palsy exhibited lower gait speeds in daily life compared to laboratory settings, potentially due to walking barefoot during laboratory-based assessments. Other studies have also observed discrepancies between supervised and unsupervised gait speeds in older adults. Takayanagi et al [66] identified a weak correlation between daily life and laboratory gait speeds, with daily speeds being significantly lower. De la Cámara et al [67] noted that clinical gait speed measurements might not accurately reflect habitual gait speeds and are influenced by physical, mental, and cognitive health. These studies, which used IMUs, faced limitations such as short walking distances in laboratory settings (2.44-10 meters), which do not adequately represent real-world gait. Even in PD, laboratory assessments do not always mirror daily activity accurately. Toosizadeh et al [68] found no significant correlation between laboratory and home gait parameters, possibly due to small sample sizes and methodological differences. Corrà et al [69] found that certain laboratory tests can better represent gait speed measured at home. However, the ability to remotely monitor patients at home is beneficial for clinicians because it allows them to assess patients' motor capacity in their domestic environment. In this sense, both in-laboratory and daily living assessments using wearable sensors can provide complementary insights into PD, enhancing clinical evaluations and patient management.

The sensitivity of supervised versus unsupervised assessments for detecting motor fluctuations in PD remains unclear. Therefore, to enhance clinical care and design personalized interventions, a better understanding of gait disabilities is needed, particularly through comprehensive daily life performance data from IMUs.

Although wearable sensors have shown considerable potential in monitoring and detecting fall risk, it is important to acknowledge their limitations and the challenges associated with their use. One critical point is that predicting a fall is not equivalent to preventing it. Current systems are largely reactive, aiming to detect gait abnormalities or balance impairments before a fall occurs, but they do not directly intervene to prevent the event itself. This distinction is crucial, especially in the context of a rapidly aging global population, where fall-related injuries are expected to increase substantially. Therefore, future research should focus not only on improving prediction accuracy but also on integrating real-time intervention mechanisms (eg, alerts, assistive devices, or environmental modifications) that could help prevent falls in a practical setting [21].

Another key challenge lies in ergonomics and user acceptance of wearable technologies. Many current devices lack design considerations for comfort and aesthetics, which can lead to discomfort, skin irritation, or reluctance to use the device, particularly among older adults. These limitations negatively affect long-term adherence and reduce the ecological validity of continuous monitoring solutions. Addressing this issue requires a multidisciplinary approach involving engineers, designers, and end users to codevelop devices that are lightweight and easy to wear. In parallel, user-centered studies should investigate how different sensor placements, materials, and feedback systems influence usability and compliance. By overcoming these ergonomic barriers, wearable technologies can move closer to becoming a truly accessible and effective tool for fall risk management in real-life settings [70].

It is also important to note that data loss and signal noise can affect the reliability of sensor-derived features, especially in real-world settings. To mitigate these issues, careful attention must be given to preprocessing strategies and to the design of data acquisition protocols that minimize artefacts and signal interruptions. Moreover, differences in device types, sensor configurations, and data processing pipelines across studies make cross-study comparisons challenging. Standardization efforts and robust preprocessing techniques are essential to enhance the reproducibility and generalizability of the findings.

ML Algorithms and Data Processing

In the studies included in this systematic review, a range of AI techniques were applied to analyze data collected via wearable sensors for assessing fall risk in patients with neurological disorders. DL models were prominently featured [37,43,48]. These models are particularly proficient at processing sequential data and capturing temporal dependencies essential for interpreting complex gait patterns over time. This choice reflects a growing trend of using advanced AI methods to detect the complex patterns of movement disorders that are often overlooked by standard clinical assessments. Besides DL, traditional ML techniques also played a significant role; for instance, feature-based models such as support vector machines or random forests were used to classify patients based on the extracted biomechanical features, such as stride length and turning behavior [49,51,52]. Another crucial aspect identified in the included studies was the preprocessing of sensor data using advanced data fusion and signal processing techniques. These methods prepare the raw data for more effective AI analysis by enhancing data quality and extracting meaningful features. Arpan et al [44] used unscented Kalman filter techniques to integrate data from multiple sensors, facilitating a more accurate AI-based prediction of fall events [44]. Some of the included studies (3/19, 16%) explored the implementation of AI models that operate in real time to provide immediate insights or predictive analytics. This approach is particularly useful in a home monitoring context where immediate feedback can prevent falls [63].

Looking ahead, the integration of AI in fall risk assessment could expand in several directions. Tailoring AI models to individual patients' data over time to adapt to changes in their condition or progression of their neurological disease could improve ongoing monitoring and therapeutic accuracy. By incorporating methodologies that adjust to patient-specific needs and conditions, AI can facilitate more dynamic interventions tailored to the evolving profiles of patients. In a recent study,

XSL•FO
RenderX

Espinoza Bernal et al [71] used personalized ML algorithms to monitor and adjust individual rehabilitation strategies for patients with stroke. Using data from wearable sensors during a rehabilitation camp, these algorithms provided high classification accuracy and insights into changes in patient mobility. This capability not only demonstrates the potential of AI to adapt to and predict the needs of patients based on real-time data but also highlights its value in managing conditions that affect mobility and stability, enhancing the precision of therapy and potentially improving patient outcomes [71]. Furthermore, another recent study described neural network models specifically designed to capture the dynamic properties of gait, considering both short-term and long-term dependencies in gait cycles [72]. This approach allowed for the identification of individual-specific gait signatures, which are particularly useful in understanding and accommodating the unique gait dynamics of survivors of stroke, who often exhibit significant variability in their walking patterns due to neurological impairments [72]. The personalization of AI models to adapt to individual patient profiles over time promises to enhance the precision of interventions, although as AI models evolve, there is also a growing need for explainable AI that can make AI decisions more transparent and understandable to clinicians. This will be crucial for building trust and facilitating the integration of AI tools into everyday clinical practice. Future advancements should aim to not only refine the predictive accuracy of fall risk but also ensure that the outcome data are accessible and interpretable for all users involved in the care process. A recent study demonstrated the use of wearable sensors coupled with explainable AI techniques to provide detailed insights into muscle activity changes, aiding clinicians in understanding the predictive outcomes of AI models [73].

Clinical Considerations and Future Perspectives

Detecting fall risk in individuals with neurological disorders has a primary role in the rehabilitation context. In this review, we identified several biomechanical parameters related to fall risk, as measured by wearable sensors. Data extracted from sensors can provide valuable insights into motor deficits associated with fall risk and guide the development of personalized fall prevention programs. Of note, conventional clinical scales may be not able to capture specific fall-related movements detected by wearable sensors [74]. This discrepancy highlights the importance of body-worn sensors, combining daily life sensor data with clinical assessment outcomes to enhance fall prediction accuracy. Therefore, daily life data recording should be preferred over in-laboratory data collection, as individuals with neurological disorders are more susceptible to the effects of test instructions or external distractions in controlled settings. However, lower activity levels in individuals with neurological disorders may affect the quality of the data collected. Moreover, both the location and method of sensor attachment are crucial when collecting sensor-based data from individuals with neurological disorders. Using multiple sensors might provide more detailed information but is less practical for this patient group. In addition, researchers or trained staff need to be present to ensure the correct placement and wearing of the sensors. From a practical standpoint, the sensor's location should be chosen carefully to minimize discomfort during the assessment and avoid interfering with natural movement. Technical considerations, such as battery life, data transmission, and storage capacity, must also be factored in when selecting an appropriate sensor for research or clinical practice.

Future studies should incorporate additional fall risk factors (ie, cognitive status, comorbidities, and sensory functions) that could improve the accuracy of fall risk prediction. They should also consider using multiple sensors to capture more detailed biomechanical information. Innovative technologies such as wearable sensors hold promise for enhancing motion analysis and fall risk prediction due to their increasing simplicity, speed, and ease of interpretation compared to traditional methods. However, there remains a scarcity of studies with practical clinical applications for gait analysis in populations with neurological disorders, particularly in patients with stroke or AD. This gap may be attributed to practical challenges, such as the lack of widely available specialized devices in hospital settings and economic constraints. In addition, the absence of direct collaboration between clinicians and bioengineering professionals contributes to this issue. Increased collaboration between physiotherapists, kinesiologists, and bioengineers is essential for sharing practical insights that inform the development of specific motor training programs based on functional assessments. Technological advancements could address these challenges by promoting the use of smart tools, such as smartphones, which are becoming more prevalent in clinical settings, as supported by some studies [13,75,76]. In parallel, advancements in wearable sensor technology are being complemented by the rapid evolution of AI, particularly through its integration with big data, cloud computing, and Internet of Things platforms. Wearable sensors stand to benefit significantly from these advancements, as AI algorithms enable the efficient processing of large and complex datasets; the extraction of informative features; and the generation of accurate, data-driven decisions. These capabilities extend to a range of functions, including gesture recognition, movement analysis, object detection, and continuous health monitoring [77]. Moreover, the integration of multimodal sensors to collect various data types such as motion, physiological, and environmental signals offers promising opportunities for enhancing fall risk assessment. While traditional data fusion methods often struggle with the complexity and heterogeneity of such data, DL provides powerful tools to automatically extract and model relationships across multiple sensor modalities [78].

Limitations

This systematic review has some limitations that need to be acknowledged. One limitation is the absence of quantitative analysis. In particular, we found considerable heterogeneity among the included studies in terms of sensors used, sensor placement, and biomechanical features extracted; therefore, conducting a quantitative analysis was not feasible. This limitation underscores the need for caution in generalizing the findings and emphasizes the importance of interpreting the results within the context of individual study characteristics. The selected studies also presented other limitations, including small sample sizes, lack of a control group or normative data, and lack of long-term follow-up evaluations. Future large-scale longitudinal studies are warranted to further explore and validate

XSL•FO
RenderX

key aspects of fall risk assessment and fall detection in patients with neurological disorders.

Despite these limitations, our review provides a comprehensive qualitative synthesis that systematically summarized and interpreted the findings of individual studies to elucidate common themes, patterns, and discrepancies in the literature. As such, it offers valuable insights into the biomechanical features and wearable sensors used to perform fall risk assessment, identifying key implications for clinical practice and considerations for future investigation.

Conclusions

In this systematic review, we highlighted the critical role of wearable technologies in providing an objective and quantitative assessment of fall risk in patients with neurological disorders. Although the included studies were well executed, we identified significant heterogeneity in sensor placement, types of wearable systems, motor tasks, and evaluation methods. In addition, we observed a tendency among the included studies to primarily focus on technological and engineering aspects, often at the expense of the clinical context. However, the absence of a standardized fall risk assessment using wearable systems is likely attributable to ongoing technological advancements, which can be considered a strength.

Authors' Contributions

AI and MB conceptualized the study. MB, AI, and PDP were responsible for methodology. AI and MB were responsible for software. All authors contributed to validation and visualization. AI was responsible for formal analysis. MB was responsible for investigation. RSC and AQ were responsible for resources. AC and AMDN were responsible for data curation. MB and AI wrote the original draft. RSC reviewed and edited the manuscript. RSC, AC, and AMDN supervised the study. RSC and AMDN were responsible for project administration. AQ was responsible for funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 25 KB - mhealth_v13i1e67265_app1.docx]

Multimedia Appendix 2
Search queries.
[DOCX File , 15 KB - mhealth_v13i1e67265_app2.docx]

References

1. Zhu J, Cui Y, Zhang J, Yan R, Su D, Zhao D, et al. Temporal trends in the prevalence of Parkinson's disease from 1980 to 2023: a systematic review and meta-analysis. Lancet Healthy Longev 2024 Jul;5(7):e464-e479 [FREE Full text] [doi: 10.1016/S2666-7568(24)00094-1] [Medline: 38945129]
2. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS, third edition. Mult Scler 2020 Dec 11;26(14):1816-1821 [FREE Full text] [doi: 10.1177/1352458520970841] [Medline: 33174475]
3. Niu H, Álvarez-Álvarez I, Guillén-Grima F, Aguinaga-Ontoso I. Prevalence and incidence of Alzheimer's disease in Europe: a meta-analysis. Neurología (English Edition) 2017 Oct;32(8):523-532. [doi: 10.1016/j.nrleng.2016.02.009]
4. Tahami Monfared AA, Byrnes MJ, White LA, Zhang Q. Alzheimer's disease: epidemiology and clinical progression. Neurol Ther 2022 Jun 14;11(2):553-569 [FREE Full text] [doi: 10.1007/s40120-022-00338-8] [Medline: 35286590]
5. Niu H, Álvarez-Álvarez I, Guillén-Grima F, Aguinaga-Ontoso I. Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis. Neurologia 2017 Oct;32(8):523-532 [FREE Full text] [doi: 10.1016/j.nrl.2016.02.016] [Medline: 27130306]
6. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820 [FREE Full text] [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
7. Shankar KN, Li A. Older adult falls in emergency medicine, 2023 update. Clin Geriatr Med 2023 Nov;39(4):503-518. [doi: 10.1016/j.cger.2023.05.010] [Medline: 37798062]
8. Zhou J, Liu B, Ye H, Duan JP. A prospective cohort study on the association between new falls and balancing ability among older adults over 80 years who are independent. Exp Gerontol 2023 Sep;180:112259 [FREE Full text] [doi: 10.1016/j.exger.2023.112259] [Medline: 37473970]

XSL•FO
RenderX

9. Ehrhardt A, Hostettler P, Widmer L, Reuter K, Petersen JA, Straumann D, et al. Fall-related functional impairments in patients with neurological gait disorder. Sci Rep 2020 Dec 03;10(1):21120 [FREE Full text] [doi: 10.1038/s41598-020-77973-4] [Medline: 33273488]
10. Stolze H, Klebe S, Zechlin C, Baecker C, Friege L, Deuschl G. Falls in frequent neurological diseases--prevalence, risk factors and aetiology. J Neurol 2004 Jan;251(1):79-84. [doi: 10.1007/s00415-004-0276-8] [Medline: 14999493]
11. Kehrer-Dunlap AL, Keleman AA, Bollinger RM, Stark SL. Falls and Alzheimer disease. Adv Geriatr Med Res 2024;6(1):33 [FREE Full text] [doi: 10.20900/agmr.20240001] [Medline: 38549879]
12. Bonanno M, De Nunzio AM, Quartarone A, Militi A, Petralito F, Calabrò RS. Gait analysis in neurorehabilitation: from research to clinical practice. Bioengineering (Basel) 2023 Jun 30;10(7):785 [FREE Full text] [doi: 10.3390/bioengineering10070785] [Medline: 37508812]
13. Pistacchi M, Gioulis M, Sanson F, De Giovannini E, Filippi G, Rossetto F, et al. Gait analysis and clinical correlations in early Parkinson's disease. Funct Neurol 2017;32(1):28-34 [FREE Full text] [doi: 10.11138/fneur/2017.32.1.028] [Medline: 28380321]
14. Coca-Tapia M, Cuesta-Gómez A, Molina-Rueda F, Carratalá-Tejada M. Gait pattern in people with multiple sclerosis: a systematic review. Diagnostics (Basel) 2021 Mar 24;11(4):584 [FREE Full text] [doi: 10.3390/diagnostics11040584] [Medline: 33805095]
15. Jensen LE, Padilla R. Effectiveness of interventions to prevent falls in people with Alzheimer's disease and related dementias. Am J Occup Ther 2011;65(5):532-540. [doi: 10.5014/ajot.2011.002626] [Medline: 22026321]
16. Mohan DM, Khandoker AH, Wasti SA, Ismail Ibrahim Ismail Alali S, Jelinek HF, Khalaf K. Assessment methods of post-stroke gait: a scoping review of technology-driven approaches to gait characterization and analysis. Front Neurol 2021;12:650024 [FREE Full text] [doi: 10.3389/fneur.2021.650024] [Medline: 34168608]
17. Roos LG, Slavich GM. Wearable technologies for health research: opportunities, limitations, and practical and conceptual considerations. Brain Behav Immun 2023 Oct;113:444-452 [FREE Full text] [doi: 10.1016/j.bbi.2023.08.008] [Medline: 37557962]
18. van Schaik JE, Dominici N. Motion tracking in developmental research: methods, considerations, and applications. Prog Brain Res 2020;254:89-111. [doi: 10.1016/bs.pbr.2020.06.007] [Medline: 32859295]
19. De Pasquale P, Bonanno M, Mojdehdehbaher S, Quartarone A, Calabrò RS. The use of head-mounted display systems for upper limb kinematic analysis in post-stroke patients: a perspective review on benefits, challenges and other solutions. Bioengineering (Basel) 2024 May 24;11(6):538 [FREE Full text] [doi: 10.3390/bioengineering11060538] [Medline: 38927774]
20. Teixeira E, Fonseca H, Diniz-Sousa F, Veras L, Boppre G, Oliveira J, et al. Wearable devices for physical activity and healthcare monitoring in elderly people: a critical review. Geriatrics (Basel) 2021 Apr 07;6(2):38 [FREE Full text] [doi: 10.3390/geriatrics6020038] [Medline: 33917104]
21. Subramaniam S, Faisal AI, Deen MJ. Wearable sensor systems for fall risk assessment: a review. Front Digit Health 2022;4:921506 [FREE Full text] [doi: 10.3389/fdgth.2022.921506] [Medline: 35911615]
22. Gillespie LD, Robertson MC, Gillespie WJ, Sherrington C, Gates S, Clemson LM, et al. Interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2012 Sep 12;2012(9):CD007146 [FREE Full text] [doi: 10.1002/14651858.CD007146.pub3] [Medline: 22972103]
23. Tulipani LJ, Meyer B, Allen D, Solomon AJ, McGinnis RS. Evaluation of unsupervised 30-second chair stand test performance assessed by wearable sensors to predict fall status in multiple sclerosis. Gait Posture 2022 May;94:19-25 [FREE Full text] [doi: 10.1016/j.gaitpost.2022.02.016] [Medline: 35220031]
24. Gu C, Lin W, He X, Zhang L, Zhang M. IMU-based motion capture system for rehabilitation applications: a systematic review. Biomim Intell Robot 2023 Jun;3(2):100097. [doi: 10.1016/j.birob.2023.100097]
25. van Schooten KS, Pijnappels M, Rispens SM, Elders PJ, Lips P, van Dieën JH. Ambulatory fall-risk assessment: amount and quality of daily-life gait predict falls in older adults. J Gerontol A Biol Sci Med Sci 2015 May;70(5):608-615. [doi: 10.1093/gerona/glu225] [Medline: 25568095]
26. Howcroft J, Lemaire ED, Kofman J. Wearable-sensor-based classification models of faller status in older adults. PLoS One 2016 Apr 7;11(4):e0153240 [FREE Full text] [doi: 10.1371/journal.pone.0153240] [Medline: 27054878]
27. Sylos Labini F, Meli A, Ivanenko YP, Tufarelli D. Recurrence quantification analysis of gait in normal and hypovestibular subjects. Gait Posture 2012 Jan;35(1):48-55. [doi: 10.1016/j.gaitpost.2011.08.004] [Medline: 21900012]
28. Hsu YL, Chung PC, Wang WH, Pai MC, Wang CY, Lin CW, et al. Gait and balance analysis for patients with Alzheimer's disease using an inertial-sensor-based wearable instrument. IEEE J Biomed Health Inform 2014 Nov;18(6):1822-1830. [doi: 10.1109/jbhi.2014.2325413]
29. Ravi D, Wong C, Lo B, Yang GZ. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. IEEE J Biomed Health Inform 2017 Jan;21(1):56-64. [doi: 10.1109/jbhi.2016.2633287]
30. Romijnders R, Warmerdam E, Hansen C, Schmidt G, Maetzler W. A deep learning approach for gait event detection from a single shank-worn IMU: validation in healthy and neurological cohorts. Sensors (Basel) 2022 May 19;22(10):3859 [FREE Full text] [doi: 10.3390/s22103859] [Medline: 35632266]

XSL•FO
RenderX

31. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
32. Eriksen MB, Frandsen TF. The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. J Med Libr Assoc 2018 Oct;106(4):420-431 [FREE Full text] [doi: 10.5195/jmla.2018.345] [Medline: 30271283]
33. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
34. Study quality assessment tools. National Heart Lung and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2024-04-17]
35. Pallavi P, Jariwala N, Patel N, Kanetkar M, Diwan S, Lahiri U. Design of SmartWalk for estimating implication of pathway with turn and task condition on postural and gait indices: relevance to fear of fall. IEEE Trans Neural Syst Rehabil Eng 2023;31:406-415. [doi: 10.1109/TNSRE.2022.3225156] [Medline: 36441883]
36. Taylor-Piliae RE, Mohler MJ, Najafi B, Coull BM. Objective fall risk detection in stroke survivors using wearable sensor technology: a feasibility study. Top Stroke Rehabil 2016 Dec;23(6):393-399. [doi: 10.1179/1074935715Z.00000000059] [Medline: 26382725]
37. Meyer BM, Tulipani LJ, Gurchiek RD, Allen DA, Adamowicz L, Larie D, et al. Wearables and deep learning classify fall risk from gait in multiple sclerosis. IEEE J Biomed Health Inform 2021 May;25(5):1824-1831 [FREE Full text] [doi: 10.1109/JBHI.2020.3025049] [Medline: 32946403]
38. Botonis OK, Harari Y, Embry KR, Mummidisetty CK, Riopelle D, Giffhorn M, et al. Wearable airbag technology and machine learned models to mitigate falls after stroke. J Neuroeng Rehabil 2022 Jun 17;19(1):60 [FREE Full text] [doi: 10.1186/s12984-022-01040-4] [Medline: 35715823]
39. Sturchio A, Dwivedi AK, Marsili L, Hadley A, Sobrero G, Heldman D, et al. Kinematic but not clinical measures predict falls in Parkinson-related orthostatic hypotension. J Neurol 2021 Mar 26;268(3):1006-1015 [FREE Full text] [doi: 10.1007/s00415-020-10240-8] [Medline: 32979099]
40. Abdollahi M, Rashedi E, Jahangiri S, Kuber PM, Azadeh-Fard N, Dombovy M. Fall risk assessment in stroke survivors: a machine learning model using detailed motion data from common clinical tests and motor-cognitive dual-tasking. Sensors (Basel) 2024 Jan 26;24(3):812 [FREE Full text] [doi: 10.3390/s24030812] [Medline: 38339529]
41. Tulipani LJ, Meyer B, Larie D, Solomon AJ, McGinnis RS. Metrics extracted from a single wearable sensor during sit-stand transitions relate to mobility impairment and fall risk in people with multiple sclerosis. Gait Posture 2020 Jul;80:361-366 [FREE Full text] [doi: 10.1016/j.gaitpost.2020.06.014] [Medline: 32615409]
42. Silva de Lima AL, Smits T, Darweesh SK, Valenti G, Milosevic M, Pijl M, et al. Home-based monitoring of falls using wearable sensors in Parkinson's disease. Mov Disord 2020 Jan;35(1):109-115 [FREE Full text] [doi: 10.1002/mds.27830] [Medline: 31449705]
43. Nouriani A, Jonason A, Sabal LT, Hanson JT, Jean JN, Lisko T, et al. Real world validation of activity recognition algorithm and development of novel behavioral biomarkers of falls in aged control and movement disorder patients. Front Aging Neurosci 2023;15:1117802 [FREE Full text] [doi: 10.3389/fnagi.2023.1117802] [Medline: 36909945]
44. Arpan I, Shah VV, McNames J, Harker G, Carlson-Kuhta P, Spain R, et al. Fall prediction based on instrumented measures of gait and turning in daily life in people with multiple sclerosis. Sensors (Basel) 2022 Aug 09;22(16):5940 [FREE Full text] [doi: 10.3390/s22165940] [Medline: 36015700]
45. Ma L, Mi TM, Jia Q, Han C, Chhetri JK, Chan P. Gait variability is sensitive to detect Parkinson's disease patients at high fall risk. Int J Neurosci 2022 Sep;132(9):888-893. [doi: 10.1080/00207454.2020.1849189] [Medline: 33256488]
46. Greene BR, Premoli I, McManus K, McGrath D, Caulfield B. Predicting fall counts using wearable sensors: a novel digital biomarker for Parkinson’s disease. Sensors (Basel) 2021 Dec 22;22(1):54 [FREE Full text] [doi: 10.3390/s22010054] [Medline: 35009599]
47. Song Z, Ou J, Wu S, Shu L, Fu Q, Xu X. Wearable fall risk assessment by discriminating recessive weak foot individual. J Neuroeng Rehabil 2025 Mar 20;22(1):64 [FREE Full text] [doi: 10.1186/s12984-025-01599-8] [Medline: 40114179]
48. Meyer BM, Tulipani LJ, Gurchiek RD, Allen DA, Solomon AJ, Cheney N, et al. Open-source dataset reveals relationship between walking bout duration and fall risk classification performance in persons with multiple sclerosis. PLOS Digit Health 2022 Oct 18;1(10):e0000120 [FREE Full text] [doi: 10.1371/journal.pdig.0000120] [Medline: 36812538]
49. Ullrich M, Roth N, Kuderle A, Richer R, Gladow T, Gasner H, et al. Fall risk prediction in Parkinson’s disease using real-world inertial sensor gait data. IEEE J Biomed Health Inform 2023 Jan;27(1):319-328. [doi: 10.1109/JBHI.2022.3215921] [Medline: 36260566]
50. Huang S, Zhou X, Liu Y, Luo J, Lv Z, Shang P, et al. High fall risk associated with memory deficit and brain lobes atrophy among elderly with amnestic mild cognitive impairment and mild Alzheimer’s disease. Front Neurosci 2022;16:896437 [FREE Full text] [doi: 10.3389/fnins.2022.896437] [Medline: 35757554]
51. Hildebrand A, Jacobs PG, Folsom JG, Mosquera-Lopez C, Wan E, Cameron MH. Comparing fall detection methods in people with multiple sclerosis: a prospective observational cohort study. Mult Scler Relat Disord 2021 Nov;56:103270. [doi: 10.1016/j.msard.2021.103270] [Medline: 34562766]

XSL•FO
RenderX

52. Kushner T, Mosquera-Lopez C, Hildebrand A, Cameron MH, Jacobs PG. Risky movement: assessing fall risk in people with multiple sclerosis with wearable sensors and beacon-based smart-home monitoring. Mult Scler Relat Disord 2023 Nov;79:105019. [doi: 10.1016/j.msard.2023.105019] [Medline: 37801954]
53. Mosquera-Lopez C, Wan E, Shastry M, Folsom J, Leitschuh J, Condon J, et al. Automated detection of real-world falls: modeled from people with multiple sclerosis. IEEE J Biomed Health Inform 2021 Jun;25(6):1975-1984. [doi: 10.1109/jbhi.2020.3041035]
54. Greene BR, McManus K, Redmond SJ, Caulfield B, Quinn CC. Digital assessment of falls risk, frailty, and mobility impairment using wearable sensors. NPJ Digit Med 2019;2:125 [FREE Full text] [doi: 10.1038/s41746-019-0204-z] [Medline: 31840096]
55. Greene BR, Redmond SJ, Caulfield B. Fall risk assessment through automatic combination of clinical fall risk factors and body-worn sensor data. IEEE J Biomed Health Inform 2017 May;21(3):725-731. [doi: 10.1109/JBHI.2016.2539098] [Medline: 28113482]
56. Huang T, Li M, Huang J. Recent trends in wearable device used to detect freezing of gait and falls in people with Parkinson's disease: a systematic review. Front Aging Neurosci 2023;15:1119956 [FREE Full text] [doi: 10.3389/fnagi.2023.1119956] [Medline: 36875701]
57. Zhou Y, Zia Ur Rehman R, Hansen C, Maetzler W, Del Din S, Rochester L, et al. Classification of neurological patients to identify fallers based on spatial-temporal gait characteristics measured by a wearable device. Sensors (Basel) 2020 Jul 23;20(15):4098 [FREE Full text] [doi: 10.3390/s20154098] [Medline: 32717848]
58. Cameron MH, Nilsagard Y. Balance, gait, and falls in multiple sclerosis. Handb Clin Neurol 2018;159:237-250. [doi: 10.1016/B978-0-444-63916-5.00015-X] [Medline: 30482317]
59. Cameron MH, Nilsagard Y. Balance, gait, and falls in multiple sclerosis. Handb Clin Neurol 2018;159:237-250. [doi: 10.1016/B978-0-444-63916-5.00015-X] [Medline: 30482317]
60. Camicioli R, Morris ME, Pieruccini-Faria F, Montero-Odasso M, Son S, Buzaglo D, et al. Prevention of falls in Parkinson's disease: guidelines and gaps. Mov Disord Clin Pract 2023 Oct;10(10):1459-1469 [FREE Full text] [doi: 10.1002/mdc3.13860] [Medline: 37868930]
61. Pelicioni PH, Menant JC, Latt MD, Lord SR. Falls in Parkinson's disease subtypes: risk factors, locations and circumstances. Int J Environ Res Public Health 2019 Jun 23;16(12):2216 [FREE Full text] [doi: 10.3390/ijerph16122216] [Medline: 31234571]
62. Weerdesteyn V, de Niet M, van Duijnhoven HJ, Geurts AC. Falls in individuals with stroke. J Rehabil Res Dev 2008;45(8):1195-1213. [Medline: 19235120]
63. Silva de Lima AL, Evers LJ, Hahn T, Bataille L, Hamilton JL, Little MA, et al. Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review. J Neurol 2017 Aug;264(8):1642-1654 [FREE Full text] [doi: 10.1007/s00415-017-8424-0] [Medline: 28251357]
64. Ferreira RN, Ribeiro NF, Santos CP. Fall risk assessment using wearable sensors: a narrative review. Sensors (Basel) 2022 Jan 27;22(3):984 [FREE Full text] [doi: 10.3390/s22030984] [Medline: 35161731]
65. Carcreff L, Gerber CN, Paraschiv-Ionescu A, De Coulon G, Aminian K, Newman C, et al. Walking speed of children and adolescents with cerebral palsy: laboratory versus daily life. Front Bioeng Biotechnol 2020;8:812 [FREE Full text] [doi: 10.3389/fbioe.2020.00812] [Medline: 32766230]
66. Takayanagi N, Sudo M, Yamashiro Y, Lee S, Kobayashi Y, Niki Y, et al. Relationship between daily and in-laboratory gait speed among healthy community-dwelling older adults. Sci Rep 2019 Mar 05;9(1):3496 [FREE Full text] [doi: 10.1038/s41598-019-39695-0] [Medline: 30837520]
67. De la Cámara M, Higueras-Fresnillo S, Sadarangani KP, Esteban-Cornejo I, Martinez-Gomez D, Veiga Ó. Clinical and ambulatory gait speed in older adults: associations with several physical, mental, and cognitive health outcomes. Phys Ther 2020 Apr 17;100(4):718-727. [doi: 10.1093/ptj/pzz186] [Medline: 31899490]
68. Toosizadeh N, Mohler J, Lei H, Parvaneh S, Sherman S, Najafi B. Motor performance assessment in Parkinson's disease: association between objective in-clinic, objective in-home, and subjective/semi-objective measures. PLoS One 2015;10(4):e0124763 [FREE Full text] [doi: 10.1371/journal.pone.0124763] [Medline: 25909898]
69. Corrà MF, Atrsaei A, Sardoreira A, Hansen C, Aminian K, Correia M, et al. Comparison of laboratory and daily-life gait speed assessment during ON and OFF states in Parkinson's disease. Sensors (Basel) 2021 Jun 09;21(12):3974 [FREE Full text] [doi: 10.3390/s21123974] [Medline: 34207565]
70. Li Y, Liu P, Fang Y, Wu X, Xie Y, Xu Z, et al. A decade of progress in wearable sensors for fall detection (2015-2024): a network-based visualization review. Sensors (Basel) 2025 Mar 31;25(7):2205 [FREE Full text] [doi: 10.3390/s25072205] [Medline: 40218718]
71. Espinoza Bernal VC, Hiremath SV, Wolf B, Riley B, Mendonca RJ, Johnson MJ. Classifying and tracking rehabilitation interventions through machine-learning algorithms in individuals with stroke. J Rehabil Assist Technol Eng 2021 Oct 07;8:20556683211044640 [FREE Full text] [doi: 10.1177/20556683211044640] [Medline: 34646574]
72. Winner TS, Rosenberg MC, Jain K, Kesar TM, Ting LH, Berman GJ. Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics. PLoS Comput Biol 2023 Oct 27;19(10):e1011556 [FREE Full text] [doi: 10.1371/journal.pcbi.1011556] [Medline: 37889927]

XSL•FO
RenderX

73. Hussain I, Jany R. Interpreting stroke-impaired electromyography patterns through explainable artificial intelligence. Sensors (Basel) 2024 Feb 21;24(5):1392 [FREE Full text] [doi: 10.3390/s24051392] [Medline: 38474928]
74. De Pasquale P, Bonanno M, De Marchis C, Pergolizzi L, Lombardo Facciale A, Paladina G, et al. Beyond clinical scales: an observational study on instrumental gait analysis and biomechanical patterns in patients with Parkinson's disease. Front Bioeng Biotechnol 2025;13:1541240 [FREE Full text] [doi: 10.3389/fbioe.2025.1541240] [Medline: 40370594]
75. Alberto S, Cabral S, Proença J, Pona-Ferreira F, Leitão M, Bouça-Machado R, et al. Validation of quantitative gait analysis systems for Parkinson's disease for use in supervised and unsupervised environments. BMC Neurol 2021 Aug 28;21(1):331 [FREE Full text] [doi: 10.1186/s12883-021-02354-x] [Medline: 34454453]
76. Marković V, Stanković I, Radovanović S, Petrović I, Ječmenica Lukić M, Dragašević Mišković N, et al. Gait alterations in Parkinson's disease at the stage of hemiparkinsonism-a longitudinal study. PLoS One 2022 Jul 21;17(7):e0269886 [FREE Full text] [doi: 10.1371/journal.pone.0269886] [Medline: 35862311]
77. Kong L, Song J, Fang Z, Chen H, Fan C, Zhang Z, et al. Artificial intelligence-assisted wearable electronics for human-machine interfaces. Device 2025 Mar;3(3):100707. [doi: 10.1016/j.device.2025.100707]
78. Tang Q, Liang J, Zhu F. A comparative review on multi-modal sensors fusion based on deep learning. Signal Process 2023 Dec;213:109165. [doi: 10.1016/j.sigpro.2023.109165]

Abbreviations

30CST: 30-second chair stand test
ADL: activity of daily living
AI: artificial intelligence
AUC: area under the receiver operating characteristic curve
DL: deep learning
EMG: electromyography
IMU: inertial measurement unit
LSTM: long short-term memory
ML: machine learning
MS: multiple sclerosis
PD: Parkinson disease
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TUG: timed up and go

Edited by Z Liang; submitted 07.10.24; peer-reviewed by R Valentim, W Qi, D Rubeca; comments to author 16.04.25; revised version received 14.05.25; accepted 20.05.25; published 18.08.25.

Please cite as:
Bonanno M, Ielo A, De Pasquale P, Celesti A, De Nunzio AM, Quartarone A, Calabrò RS
Use of Wearable Sensors to Assess Fall Risk in Neurological Disorders: Systematic Review
JMIR Mhealth Uhealth 2025;13:e67265
URL: https://mhealth.jmir.org/2025/1/e67265
doi:10.2196/67265
PMID:40824690

©Mirjam Bonanno, Augusto Ielo, Paolo De Pasquale, Antonio Celesti, Alessandro Marco De Nunzio, Angelo Quartarone, Rocco Salvatore Calabrò. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Engagement With Conversational Agent–Enabled Interventions in Cardiometabolic Disease Self-Management: Systematic Review

Nick Kashyap[1,2,3], BSc, MSc; Ann Tresa Sebastian[4], MSc; Chris Lynch[1,3], BSc, MSc, PhD; Paul Jansons[4,5], PhD; Ralph Maddison[4], Prof Dr; Tilman Dingler[6], PhD; Brian Oldenburg[1,2,3], Prof Dr

[1]School of Psychology and Public Health, La Trobe University, Melbourne, Australia

[2]Baker Department of Cardiovascular Research, Translation and Implementation, La Trobe University, Melbourne, Australia

[3]Non-communicable Disease and Implementation Science laboratory, Baker Heart and Diabetes Institute, Melbourne, Australia

[4]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Melbourne, Australia

[5]Department of Medicine, School of Clinical Sciences at Monash Health, Monash University, Melbourne, Australia

[6]Delft University of Technology, Delft, The Netherlands

Corresponding Author:
Nick Kashyap, BSc, MSc
School of Psychology and Public Health
La Trobe University
1 Kingsbury Drive, Bundoora
Melbourne, 3086
Australia
Phone: 61 1300 528 762
Email: Nick.Kashyap@baker.edu.au

Abstract

Background: Well-designed conversational agents can improve health care capacity to meet the dynamic and complex needs of people self-managing cardiometabolic diseases (CMD). However, a lack of empirical evidence on conversational agent–enabled intervention design features and their impact on engagement make it challenging to comprehensively evaluate effectiveness. This review synthesizes evidence on conversational agent–enabled intervention design features and how they impact on engagement to inform the development of more engaging conversational agent–enabled interventions that effectively help people with CMD to self-manage their condition.

Objective: The aim of the study is to synthesize evidence pertaining to conversational agent–enabled intervention design features and their impact on engagement of people self-managing CMD.

Methods: Searches were conducted in Ovid (MEDLINE), Web of Science, and Scopus databases. Inclusion criteria were primary research studies reporting on conversational agent–enabled interventions that included measures of engagement and included adults with CMD. Data extraction captured perspectives of people with CMD on various design features of conversational agent–enabled interventions.

Results: Of 1366 studies identified for screening, 20 were included in the review. In total, 18 of these were qualitative or quasi-experimental evaluations of conversational agent–enabled intervention prototypes. Five domains of design features that impact user engagement with conversational agent–enabled interventions emerged: communication style, functionality, accessibility, visual appearance, and personality.

Conclusions: Across all 5 domains, integrating redundancy and anthropomorphism were identified as effective strategies for improving engagement by increasing user autonomy and investment. Future research should adopt design strategies that are inclusive and adaptive to the diverse needs of users and aligned with the unique considerations relevant to conversational agent–enabled interventions.

Trial Registration: PROSPERO CRD42023431579; https://tinyurl.com/3srmzw8f

International Registered Report Identifier (IRRID): RR2-10.2196/52973

(JMIR Mhealth Uhealth 2025;13:e67913) doi:10.2196/67913

XSL•FO
RenderX

KEYWORDS

user experience; cardiovascular disease; diabetes; chronic disease; chatbot; digital assistant; qualitative; digital health

Introduction

Supporting effective self-management is crucial for reducing both the health and economic impacts of cardiometabolic diseases (CMD) but is challenged by the dynamic and complex nature of CMDs and limitations faced by health care providers [1-3]. Encompassing a set of interrelated conditions that affect the cardiovascular system and metabolic health, such as hypertension and diabetes, CMDs are a significant health concern [4]. Globally, CMDs cause approximately 20 million deaths each year, with morbidity and mortality rates continuing to increase [5-7]. It has been estimated that nearly 80% of CMD deaths and a significant proportion of associated health care costs could be prevented with optimal self-management [2,8]. Self-management describes the ability of patients to monitor their symptoms and circumstances and understand how these factors affect health [9]. Due to the individualized and chronic course of CMDs, sustaining effective self-management requires people to continuously adjust to changing symptoms and personal circumstances [10-12]. For example, people may forget to take medications or understand changes in their medication schedule. Interventions supporting people self-managing CMDs must be continuously customized to align with the dynamic and complex experience of each individual [13]. However, limitations in time, resources, expertise, or confidence often hinder the capacity of health care providers to deliver this level of care [3,12].

Conversational agents, technologies that interact with people using natural language, can overcome some of the limitations faced by health care providers [14-16]. Conversational agents are typically accessed through app-based platforms, websites, or SMS text messaging services and can interact with the user via text, speech, or visually as embodied avatars. When incorporated into digital care delivery, conversational agents can enable continuous monitoring and alignment of interventions with the dynamic circumstances of people self-managing CMD through providing contextually relevant assistance [15,17]. Conversational agents can assess user comprehension, offer real-time emotional and motivational support, translate specialized language into familiar terms, integrate relevant dialogues with on-screen instructions, and provide voice-based interaction options [14,16]. For example, a person with hypertension receiving a reading of 150/95 mm Hg may not understand whether this requires immediate action, medication, or lifestyle changes. A multimodal conversational agent can assess their comprehension through dialogue (“What does this reading mean to you?”), provide personalized explanations based on their medication regimen, and offer contextual guidance while displaying visual aids showing normal versus elevated ranges. This is just one example of how a specific design feature, multimodality, can improve a person’s engagement with an intervention supporting them to self-manage their condition. Capabilities of agents have been advancing, as these technologies have undergone rapid progress since their emergence in the 1960s [18,19]. Despite this progress, the ongoing challenge of sustaining user engagement, a common issue in digital health, continues to limit the use of conversational agents in clinical contexts, since sustained interaction is key to their effectiveness [20-22].

The effectiveness of conversational agent–enabled interventions in supporting self-management is dependent on user engagement, which describes the depth and frequency of a user’s interactions with an intervention [23-25]. While engagement is a well-recognized concept, its application in digital health often varies based on different perspectives and methodologies for measurement [26-28]. For example, some researchers equate engagement directly with user activity, while others consider how well an intervention aligns with a user’s cognitive beliefs [27]. To achieve a change in health-related behaviors, digital health interventions require sufficient participant engagement so users not only understand and apply intervention content but also develop the skills, motivation, and confidence necessary to make sustainable lifestyle changes [24,29,30]. However, the low rates of engagement with most digital health interventions for self-management of chronic diseases are well known [21,31,32].

The design features of digital health interventions are important determinants of user engagement and are typically informed by theory and empirical evidence, such as the self-reference encoding model [25,33,34]. For example, researchers observed that when self-relevant cues were used explicitly to refer to users’ smoking habits, rather than using them implicitly to tailor content, improved smoking cessation outcomes occurred [23]. This work demonstrates both the benefits and the limitations of applying theory to design. In particular, reflecting on how the theory was applied in practice provided a more nuanced understanding of the role of self-relevant cues as a design feature.

The design of conversational agent–enabled interventions can benefit from empirical evidence on design features and their impacts on engagement in people self-managing CMDs [23,25]. Empirical evidence of this kind is sparse, primarily found in the evaluation of prototypes within qualitative and quasi-experimental studies [35-38]. This review fills this gap through a thematic synthesis that explicitly pairs individual design features with the engagement outcomes they influence. Synthesizing the empirical evidence could enable the recognition of broad trends, reduce bias in design reflections, and contribute to more nuanced applications of theories and models. This synthesis can facilitate the development of more engaging and effective conversational agent–enabled interventions for supporting people with CMDs to self-manage their condition.

Methods

Protocol and Registration

The methodology used in this study has been detailed in the protocol paper [39]. This review was conducted in accordance with the Cochrane Handbook for Systematic Reviews of

XSL•FO
RenderX

Interventions and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [40,41] (Multimedia Appendix 1). The review was initiated in May 2023 and registered in PROSPERO (CRD42023431579).

Information Sources

We systematically searched Ovid (MEDLINE), Web of Science, and Scopus databases for relevant papers from inception until April 2024. Ovid (MEDLINE) was selected for its medical and health science focus to capture studies that included CMDs. Web of Science and Scopus were selected for their broader focus, which is useful for capturing research on conversational agents and engagement topics. We excluded certain databases after preliminary assessments. For example, Embase was excluded due to its significant content overlap with MEDLINE, and the Cochrane Library was excluded because it lacked a significant number of relevant exemplar papers found in the other databases. As this is a rapidly developing field, final searches were conducted prior to the submission of this manuscript. Reference list searches were conducted on all papers included in the full-text review.

Although IEEE Xplore and the ACM Digital Library do contain work on conversational agents, a scoping search showed that almost all health-related records from those sources were already captured in Ovid (MEDLINE), Web of Science, or Scopus. The unique papers we found in the computing databases focused on technical architecture and system development, with minimal reporting on end-user engagement, the central outcome of this review. Because our objective was to synthesize evidence on how broad design features influence participant engagement (regardless of the underlying algorithm), and not to catalogue implementation details, the search strategy was confined to the 3 multidisciplinary databases, Ovid (MEDLINE), Web of Science, and Scopus. This focused approach ensured capture of health-focused literature while maintaining feasibility, as our pilot searches confirmed that the selected databases provided comprehensive coverage of conversational agent research with engagement outcomes in health care populations.

Search Strategy

An extensive set of search terms was used, which related to the 3 central topics of the review: CMD, conversational agents, and engagement. Boolean operators were used to combine search terms, including: (“Cardiovascular Diseases” [MeSH Terms] OR metabolic) AND (“conversational agent*” OR chatbot*) AND (accept* OR perceived). This approach was intended to yield a comprehensive collection of literature that explores the intersection of these central topics. The complete search strategy can be found in Multimedia Appendix 2.

Eligibility Criteria

Population

Studies included adult participants (≥18 years of age) with a CMD diagnosis. Populations reported with comorbidities, such as mental health disorders or multiple CMDs, were also included.

Intervention

Studies reported on conversational agent–enabled interventions, which supported people with CMD to self-manage their condition. For example, by offering emotional or educational support or assisting individuals in monitoring their symptoms.

Outcomes

Studies reported on engagement outcomes, such as ratings, interviews, analytics, and focus groups.

Exclusion Criteria

Excluded from this study were technical feasibility studies without user engagement data, studies where CMD populations were not reported separately, and nonpeer-reviewed studies (theses, preprints, abstract only, etc).

Study Design

Studies were primary research, including qualitative studies, quasi-experimental studies, and randomized controlled trials (RCTs). Reviews, editorials, protocols, and non-English publications were excluded.

Data Management

All search results were imported into PaperPile (Paperpile LLC), and duplicates were automatically removed. The search results were then imported into Covidence (Veritas Health Innovation), where 53 more duplicates were removed. Covidence was used to store PDF files of papers considered during the full-text review and to conduct study selection, data extraction, and quality appraisal, with data then being exported to Microsoft Excel.

Study Selection

During initial screening, 2 reviewers (NK and ATS) independently examined the titles and abstracts of all studies collected from the search strategy. Screening was based on the defined inclusion and exclusion criteria. Conflict was resolved by meetings between the 2 reviewers (NK and ATS). If a conflict was not able to be resolved, it was to be arbitrated by a third reviewer to achieve consensus, though this did not occur. Remaining studies were independently assessed by 2 reviewers (NK and ATS) in full-text screening, using the defined inclusion and exclusion criteria. Any conflict between 2 reviewers (NK and ATS) was settled during meetings, with arbitration by a third reviewer if required though this did not occur.

Data Extraction

Data extraction from all included studies, including any supplementary material, was conducted and documented by 2 reviewers (NK and ATS) independently. Where further information was required, study authors were contacted for clarification. The data extraction form was designed to collect comprehensive information on how people with CMD engage with different design features in conversational agent–enabled interventions. Details included bibliographic information, study design, participants and population, intervention features, and how design features influenced engagement outcomes.

XSL•FO
RenderX

Quality Appraisal

Due to the mixed study designs of the included literature, the following Joanna Briggs Institute critical appraisal tools were used: "Checklist for Qualitative Research," "Checklist for Quasi-Experimental Studies," and "Assessment of Risk of Bias for Randomized Controlled Trials" [42-44].

Data Synthesis

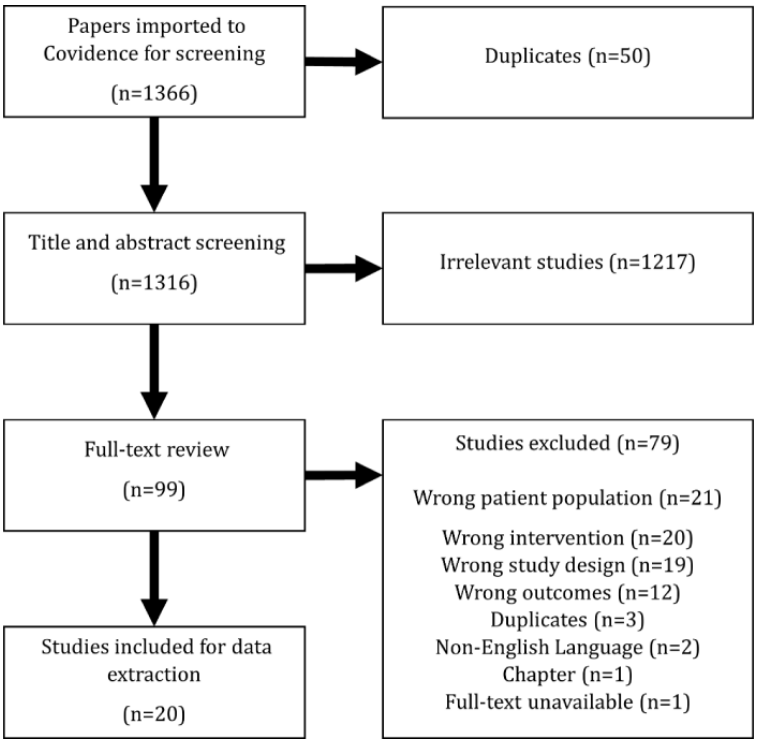
An adapted version of the thematic synthesis analysis method developed by Thomas and Harden [45] was used. This method focused on data extracted from studies that detailed design features and engagement outcomes. During the initial phase, data were categorized under multiple domains. Following the extraction, domains were iteratively consolidated until further merging would compromise the descriptive accuracy of each dataset. Then, within each domain, data were categorized under multiple subdomains. Analysis of these domains and subdomains contributed to informing design choices for conversational agent–enabled interventions and identifying research gaps in how engagement is reported in the literature.

Results

Overview

Of 1366 studies imported for screening, 20 studies were included in the review (Figure 1). All included studies were published between 2015 and 2023 (Table 1). In total, 10 studies focused on cardiovascular disease (heart failure, atrial fibrillation, acute coronary syndrome, hypertension, and stroke), 9 studies focused on diabetes (type 2 diabetes and gestational diabetes mellitus), and 1 study focused on chronic kidney disease (Table S1 in Multimedia Appendix 2) [35-38,46-62]. The sample sizes ranged from 8 to 187, and the age of the participants ranged from 21 to 87 years (Table S1 in Multimedia Appendix 2). A total of 13 of the conversational agents were embodied (ranging from simple, cartoon-like avatars in a digital interface to sophisticated, humanoid robots capable of mimicking human gestures and expressions), and 7 were chatbots (text-only interface; Table 1). Overall, studies using embodied agents revealed broader effects of design features on engagement outcomes than studies using chatbots. For further information about study designs, populations, and interventions, refer to Multimedia Appendix 2.

Figure 1. Flowchart illustrating the study selection process.



XSL•FO
RenderX

Table 1. Characteristics of included studies.

Author (year)	Study design	Conversational agent role	Conversational agent type, platform, and dialogue system	Measures of engagement (total sample size)
Apergi et al (2021) [37]	Quasi-experimental	To ask the patients the same series of questions related to their heart failure treatment and symptoms and provide feedback.	Embodied, app-based+Alexa, tree-based	Surveys, questionnaires, analytics (n=30)
Balsa et al (2019) [46]	Qualitative	To assist older people with type 2 diabetes mellitus in medication adherence and lifestyle changes.	Embodied, app-based, tree-based	Self-administered questionnaire (n=28)
Balsa et al (2020) [47]	Qualitative	To support older people with type 2 diabetes mellitus in medication adherence and lifestyle changes.	Embodied, app-based, tree-based	Questionnaire, open questions, diaries, digital notes, telephone follow-ups (n=20)
Baptista et al (2020) [36]	Quasi-experimental	To deliver self-management education and support to adults with type 2 diabetes mellitus.	Embodied, app-based, tree-based	Web-based surveys, interview (n=93)
Cheng et al (2018) [35]	Qualitative	To allow for a less cumbersome way for patients with geriatric type 2 diabetes mellitus to effectively adhere to guidelines.	Embodied, web-based+Google Home, AI[a]-based	Subjective tests for qualitative assessment of effectiveness and satisfaction, gathering feedback from experts and potential users (n=10)
Echeazarra et al (2021) [48]	Randomized controlled trial	To help patients with hypertension to self-monitor their blood pressure.	Chatbot, messaging service, tree-based	Satisfaction survey (n=112)
Epalte et al (2023) [38]	Qualitative	To counsel, educate, and train people who had a stroke and their family on stroke, rehabilitation, care, and other related issues.	Chatbot, app-based, AI-based	Semistructured interviews (n=12)
Gingele et al (2023) [49]	Qualitative	To evaluate patient's health status, provide patient education, and improve their ability to communicate with heart failure nurses.	Embodied, app-based, AI-based	System usability scale, self-developed patient satisfaction scale, general feedback form (n=37)
Gong et al (2020) [50]	Randomized controlled trial	To provide more accessible and engaging self-management support, monitoring, and coaching to adults with type 2 diabetes mellitus in Australia.	Embodied, app-based, tree-based	Data analytics, process evaluation (n=187)
Guhl et al (2020) [51]	Quasi-experimental	To augment patient-centered health care by providing health education, monitoring, and problem-solving for users.	Embodied, app-based, tree-based	Measures, self-reports (n=120)
Kimani et al (2016) [52]	Quasi-experimental	To provide education on atrial fibrillation and promote adherence to daily heart rhythm monitor readings.	Embodied, app-based, tree-based	Data analytics, self-report scale, semistructured interview (n=16)
Magnani et al (2017) [53]	Quasi-experimental	To promote education, motivation, and monitor patient's symptoms and adherence to behaviors.	Embodied, app-based, tree-based	Qualitative interviews, questionnaire (n=31)
Nassar et al (2025) [54]	Quasi-experimental	To deliver education and support to adults with type 2 diabetes mellitus.	Chatbot, web-based, tree-based	App data analytics, surveys (n=150)
Pienkowska et al (2023) [55]	Qualitative	To educate adults living with type 2 diabetes about their condition.	Chatbot, app-based, tree-based	Web-based survey, structured interview (n=8)
Roca et al (2021) [56]	Quasi-experimental	To improve medication adherence in patients with comorbid type 2 diabetes mellitus and depressive disorder.	Chatbot, messaging service, tree-based	Interviews, questionnaires (n=19)
Sagstad et al (2022) [57]	Qualitative	An informational chatbot addition to established care for women with gestational diabetes mellitus.	Chatbot, web-based, AI-based	Chatbot dialogues, data analytics (n=N/A[b])
Ter Stal et al (2021) [58]	Qualitative	To support users in self-management of chronic diseases in a long-term, daily life setting.	Embodied, app-based, tree-based	Ratings, data analytics, semistructured interviews (n=11)

XSL•FO
RenderX

Author (year)	Study design	Conversational agent role	Conversational agent type, platform, and dialogue system	Measures of engagement (total sample size)
Tongpeth et al (2018) [59]	Qualitative	To improve patients' knowledge of, and response to, acute coronary syndrome symptoms.	Embodied, app-based, tree-based	Focus groups, satisfaction questionnaire (n=22)
Tsai et al (2022) [60]	Qualitative	To support patients with chronic kidney disease and manage their condition.	Chatbot, app-based, AI-based	Structured questionnaire, one-on-one interviews (n=26)
Zhang et al (2015) [61]	Qualitative	To counsel patients on their diagnoses and medications specified by a clinician.	Embodied, app-based, unreported	Scales (n=10)

[a]AI: artificial intelligence.

[b]N/A: not available.

Study Quality

Included studies met most criteria outlined in the relevant Joana Briggs Institute critical appraisal tools with 2 exceptions, Echeazarra et al [48] and Tsai et al [60] (Multimedia Appendix 2). The most common unmet criterion for RCTs was a lack of ensured blindness, for quasi-experimental studies, it was a lack of pre- and postmeasurements, and for qualitative studies, it was unaddressed researcher influence (Multimedia Appendix 2).

Several other criteria relating to ethical considerations and data privacy were also assessed for each study: 14 of 20 studies reported on whether institutional review board approval was received, 2 of 20 studies reported on whether interventions complied with local data protection regulations, 7 of 18 studies reported on user consent procedures (participant consent was not required for 2 of the studies), 0 of 20 studies reported on user data control options, 0 of 6 studies reported whether artificial intelligence (AI) agents used explainable AI, and 0 of 6 studies reported whether efforts were made to address potential biases in AI models. These gaps highlight a critical concern: while conversational agents in health care handle sensitive personal data and influence health decisions, most studies (18/20) lacked comprehensive ethical frameworks. This represents a significant research gap, requiring urgent attention, as AI-driven health interventions become more prevalent.

Domains of Design Features

In total, 5 domains and 13 subdomains pertaining to design features that impact user engagement with conversational agent–enabled interventions emerged (Table 2). The 5 domains (and corresponding subdomains) were communication style (anthropomorphism, multiple choice, pacing, and redundancy), functionality (miscellaneous, embodiment, and personalization), accessibility (platforms and tutorials), visual appearance (anthropomorphism and gender), and personality (anthropomorphism and mentorship). Data that reported on design features or user engagement separately were used to identify proposed features and research gaps. Some studies did not directly address the impact of design features on user engagement but were useful in identifying research gaps [52,53,61]. Key findings are reported here; for comprehensive descriptions of how design features impacted engagement, refer to Multimedia Appendix 3 [35-38,46-61].

XSL•FO
RenderX

Table 2. Domains and subdomains of conversational agents and the studies that reported their impacts on engagement.

Domains and subdomains	Reference
Communication style	
Anthropomorphism	[35,36,38,47,58]
Multiple choice	[35,54-56,58,60]
Pacing	[38,47-49,54,56,58,59]
Redundancy	[47,51,54,56,57]
Functionality	
Miscellaneous	[50,54,56,59]
Embodiment	[36,37]
Personalization	[51,58]
Accessibility	
Platforms	[35,38,48,60]
Tutorials	[35,38,46,49,56]
Visual appearance	
Anthropomorphism	[36,46,55,58,59]
Gender	[58]
Personality	
Anthropomorphism	[36]
Mentorship	[36,51,55,58,59]

Communication Style

Communication style describes how a conversational agent interacts with the user. A natural flow of conversation was credited by most participants, with a favorable view of the intervention by Cheng et al [35] as the reason for its high usability. Additionally, voice quality, including factors like tone, cadence, and pronunciation, impacted user satisfaction and the perceived anthropomorphism of the agent [36,47]. Users appreciated being able to ask agents about information already provided by the health care service or the internet, “You can bring up information on [atrial fibrillation] on the internet and read most of the same stuff on your own, but with Tanya you have a structured presentation that you’re led through, so you end up getting the information you should be getting” [51,57]. Many users found multiple choice dialogues limiting and expressed frustration, leading to early disengagement [35,56,58]. Inundating users with “too much reading” or “repetitive” content caused burnout or disinterest [38,47,58]. Preferences regarding communication frequency varied among users, with a tendency for users preferring weekly communications [49]. Studies using both chatbots and embodied agents contributed similar amounts of evidence on engagement outcomes for the accessibility domain. Most engagement outcomes reported in studies using chatbots related to design features within the communication style domain.

Functionality

Functionality describes the practical capabilities of the agent. A range of functions improved user engagement, including gamification elements, agent-enabled quizzes, and weather forecasting [36,56,59]. There were conflicting reports pertaining to voice assistant functionality conversational agents, with 1 study showing no effect, and another had users reporting voice added depth to communications “instead of reading it, you’re hearing it” [36,37]. Agents that did not account for user context and time constraints led to user dissatisfaction, with irrelevant content reducing user engagement, “When I have to go to work, I do not have time to watch a 15-minute video” [51,58]. Most engagement outcomes for design features within the functionality domain were provided by studies using embodied agents.

Accessibility

Accessibility describes the measures taken to ensure that different types of users can interact with the agent. Technical challenges and a lack of familiarity with technology were issues among some users [35,46]. Low digital literacy resulted in various barriers to engaging with interventions, like users accidentally disabling the sound of reminders and then being unable to unmute them [38,56]. Even with thorough out-patient tutorials, certain demographics, such as older patients with limited digital experience, still faced challenges using the technology independently at home [49]. Older patients tended to prefer voice-activated platforms like Google Home over smartphones for conversational agents, but they faced challenges like physical inconvenience, language barriers, and technology support needs [35]. A platform that minimized installation requirements, by hosting their agent on an app users already had downloaded, was appreciated by users [60]. Studies using chatbots and embodied agents contributed similar amounts of evidence on engagement outcomes for the accessibility domain.

XSL•FO
RenderX

Visual Appearance

Visual appearance describes the physical design and visual representation of the agent. While some users favored anthropomorphic agents due to their perceived credibility, others leaned toward cartoon figures for a more "fun" and user-friendly experience [36]. A photo-realistic design, especially one that fits the context, such as a nurse in a health care setting, was reported to make interactions feel more genuine and personal [58]. Ensuring human-like facial expressions and gestures was important for avoiding negative impacts to the user experience [36,46,59]. Sylvia's identity as a woman provoked mixed reactions, with one user calling her a "stupid woman," while another liked her identity as they "hated" listening to men [58]. Only studies using embodied agents contributed evidence on engagement outcomes for the visual appearance domain.

Personality

Personality describes the character and demeanor that the agent embodies. Incorporating human-like mannerisms increased users' engagement with Laura, but her backstory led to frustration for some users, who felt that it was excessive and a waste of time, as Laura was not real [36]. Several users who perceived Laura to be authoritative developed feelings of guilt, leading some to temporarily cease communications and others to consider reporting only their best readings [36]. Whereas users who viewed Laura as friendly and supportive tended to be more engaged during communications [36]. Users explained that because Tanya and Nurse Cora did not say "very drastic things," they felt encouraged to take manageable actions in their daily lives to "do something simple daily" [51,59]. One user expressed a preference for an authoritative agent over a friendly agent, "She could be your girl next door...If I have medical complaints, I prefer an authority to explain what to do or not to do" [58]. Only studies using embodied agents contributed evidence on engagement outcomes for the visual appearance domain.

Proposed Features

Proposed features describe features suggested by users that they felt would enhance their impression of or engagement with the intervention and represent ideas for further exploration. While in-app navigation was intuitive for most users of Nurse Cora, some users indicated that they would have liked instructions on how to navigate the app containing Nurse Cora at the beginning of the trial [59]. Multiple-choice only functionality was frequently a point of frustration for users of conversational agent–enabled interventions, making one participant "consider [them]self a passive participant" in the conversation [35,38,47,54-56,58,60]. Molly users suggested that she omit punctuation when speaking and incorporate voice recognition [49]. Users of Nurse Cora wanted her to emphasize keywords using larger font sizes and to provide the option to repeat dialogues [59]. Vitória users suggested developing a "strategy to remind users to use the app" and offering specialized content such as "recipes for people with diabetes" [47]. Sylvia users desired more follow-up and checks on prior interactions, "Did you read that? Did you do this?" [58]. This sentiment was also expressed by Nurse Cora users who desired more feedback during the quizzes [59].

Research Gaps

Research gaps were found in the way engagement of conversational agent–enabled interventions tended to be reported. Conversational agent–enabled interventions were frequently rated in terms of their engagement, but the rationale behind these ratings was not explored except for the study by Cheng et al [35,38,46,48,50-54,56,58,59,61]. For this study, a natural flow of conversation was credited by most participants, with a favorable view of the app as the reason for its high usability [35]. Additionally, differing levels of engagement, trust, or perceived care based on ethnicity, medication regimes, age, education, and computer literacy were identified but not explained [37,61].

Discussion

Principal Findings

This systematic review synthesized evidence pertaining to conversational agent–enabled intervention design features and their impacts on the engagement of people self-managing CMD. Integrating redundancy and anthropomorphism to increase user autonomy and investment, respectively, emerged as 2 strategies for improving engagement relevant to all 5 emergent domains. However, care must be taken to avoid inundating users with excessive redundancy or anthropomorphism, as these were common frustrations among users across studies. The findings from this review can be applied to optimizing the development and evaluation of conversational agent–enabled interventions, facilitating more engaging and effective strategies for supporting people self-managing CMDs. Additionally, this review establishes a foundation for upcoming research by identifying research gaps and synthesizing an array of innovative design features suggested by users (Textbox 1).

Textbox 1. Design checklist.

- Provide multiple input and output modalities so users can use those that match their preferences or needs. Within each modality, allow users to customize design features, such as increasing text sizes or toggling agent accents.
- Offer a "Skip" button so users can bypass dialogues they perceive as frustrating and time-consuming, such as character backstories.
- Ensure sensitivity to how much time users have for a session and adapt content accordingly.
- Provide expandable dialogue options that users can ignore to determine the pacing of conversations.
- Avatar choices should reflect agent functions, users expect serious activities to be delivered by more realistic avatars, and fun activities to be delivered by more cartoonish avatars.
- Provide multiple options for users to learn how to use agent features, such as in-person, in-app, or helplines.

XSL•FO
RenderX

Redundant design features are those that offer additional input options and delivery channels of content. Redundancy is relevant to all 5 emergent domains, as it can be incorporated into modes of communication, offering overlapping functions and inputs and supporting adaptive interaction styles and various other aspects of intervention design. By offering redundant design features, users are empowered to interact with interventions on their own terms, thereby improving sustained engagement. For example, Roca et al [56] reported that one patient was unable to respond to medication reminders, confirming they had taken their medicine because the agent did not understand their messages. In total, the authors recorded 34 such misunderstood interactions between this patient and the agent. Despite this issue, the patient managed to manually log their medication adherence, illustrating the critical role of redundancy within interaction options [56]. A lack of redundancy in accessibility led to problems for Gingele et al [49], who found that despite providing comprehensive outpatient tutorials, users encountered significant challenges when attempting to use conversational agent–enabled interventions independently at home. These challenges could be addressed by offering additional tutorial options like helplines, outpatient sessions, and in-app guides [49]. Despite the benefits of redundancy in promoting user autonomy, empirical evidence identified within this review also underscored the importance of ensuring that users are not inundated with irrelevant content [38,47,58]. For example, Ter Stal et al [58] highlighted instances where participants felt overwhelmed by excessive and inappropriate communication frequency, leading to frustration and disengagement, “When I have to go to work, I do not have time to watch a 15-minute video.” Furthermore, Epalte et al [38] reported a participant complaining of “too much reading,” and Balsa et al [47] reported users identifying information overload and excessive queries as areas for improvement. Similarly, Pop-Eleches et al [63] found that weekly SMS text message reminders significantly improved antiretroviral adherence, whereas daily reminders did not.

The assertion that offering redundant design features improves user engagement by enabling flexibility in interaction aligns with findings grounded in self-determination theory from a tutorial paper by Yardley et al [26]. Self-determination theory posits that engagement can be improved by supporting user autonomy, competence, and relatedness [64]. By providing redundancy in options and content, users are empowered to choose their interaction methods, access alternative options, and find content that feels personally relevant, thereby supporting user autonomy, competence, and relatedness [26]. An alternative explanation for why redundant design features improved user engagement is that they create a more consistent user environment where repeated exposures to similar content can make new information more familiar and accessible [65]. This explanation links into the concept of omnichannel engagement, describing uniformity between content channels, which is elaborated in a viewpoint paper by Blasiak et al [65]. There were several examples where users appreciated interventions providing content, which they could access elsewhere [51,56,57]. For example, Sagstad et al [57] and Guhl et al [51] reported that users appreciated accessing information through the conversational agents within the interventions, which they knew they could access on the web or from their doctor [51,57]. Furthermore, all participants within the study by Roca et al [56] gave an affirmative response to the statement, “From a certain moment, the virtual assistant began to give the weather forecast. Has this helped you to use the virtual assistant more frequently?”

Anthropomorphic design features are those that contribute to users attributing human traits, emotions, or intentions to the intervention. Anthropomorphism is relevant to all 5 emergent domains, as it can be incorporated into the conversational style, personality, visual design, and various other aspects of intervention design. By offering anthropomorphic design features, users can become more invested in interventions, thereby improving sustained engagement. For example, Ter Stal et al [58] reported a participant stating that the realistic depiction of the conversational agent made the interaction more personal. Similarly, Baptista et al [36] reported user preference for Laura’s human character over a cartoon character, “I’m not sure I would have given the same level of credibility to, for example, a dog or a cat or something like that.” Furthermore, a common point of frustration among users across studies was with conversational agents which were “not human enough”; these issues pertained to the speech, mannerisms, gestures, and expressions of conversational agents [36,46,59]. Despite the benefits of anthropomorphism in promoting user investment, empirical evidence identified within this review also identified instances where efforts toward this goal were excessive or came at the expense of other functionalities [36,58]. An example of an excessive pursuit of anthropomorphism was identified by Baptista et al [36], who observed that providing a backstory for their conversational agent led to frustration for some users. Here, a backstory feature in the personality domain, when projected through an anthropomorphized avatar in the visual appearance domain, resulted in this avatar increasing frustration rather than investment with the intervention itself. Furthermore, Ter Stal et al [58] reported that most participants expressed that they were not interested in the small talk from their conversational agent.

The contention that offering anthropomorphic design features improves user engagement by promoting user investment in interactions aligns with the computers are social actors framework. The computers are social actors framework posits that engagement can be improved by supporting user investment by exploiting users’ innate social expectations to make interactions more meaningful [66]. However, the computers are social actors framework fails to explain why excessive anthropomorphism leads to user frustration. Ruslana and Ning [67] identified excessive anthropomorphism as problematic because it caused confusion for users. However, they also suggested an alternative explanation for its positive effect on engagement: anthropomorphic design features may make interventions feel more approachable and less formal. This suggests that developers should carefully consider the function of anthropomorphism and be willing to forego it if simpler methods can achieve more approachable interventions. For example, Baptista et al [36] observed that certain users favored cartoon figures over anthropomorphic agents, as they were more “fun.” Similarly, Thunström et al [68] found that a mental health

XSL•FO
RenderX

chatbot was rated more usable than an anthropomorphic embodied conversational agent, though the usability scale used in this study did not have the fidelity to attribute this to a specific feature differing between the agents.

Limitations

This systematic review faced limitations primarily due to the heterogeneity of included interventions, outcomes, and study types. In terms of interventions, the field of conversational agent–enabled interventions is rapidly evolving, with new technologies and platforms emerging regularly. This resulted in a wide variety of conversational agent–enabled interventions included in the review, ranging from basic chatbots with multiple choice responses to embodied conversational agents. In terms of outcomes, engagement in digital health lacks a standardized definition, leading to different evaluations and operationalizations between studies. For example, some studies equated it directly with use, while others conceptualized it as a composite of behavioral, cognitive, and affective components. The diversity in how engagement was defined and measured across studies limited the synthesis of findings, pointing toward the need for a standardized operationalization of engagement within digital health contexts, to enable more precise and comparable measurements. In terms of study types, papers included in this review were primarily qualitative or quasi-experimental studies, restricting our ability to compare design features with quantitative user engagement metrics through meta-analysis. Together, these heterogeneities limited our ability to compare results across studies and draw definitive conclusions about the impacts of certain design features on engagement, so findings should be interpreted as directional rather than quantitative. Furthermore, studies were conducted in home-based self-management contexts, which are not equally applicable across other health care contexts, such as hospitals and primary care settings. However, the review proceeded with these limitations as more specific inclusion criteria for interventions, outcomes, and study types would have reduced the sensitivity of this systematic review in identifying insights pertaining to design features and their impacts on engagement and in identifying research gaps and opportunities for further research. To address these limitations, we recommend that future systematic reviews focus on specific agent types (eg, embodied agents only) with standardized engagement definitions and conduct individual participant data meta-analyses when sufficient high-quality RCTs become available.

Areas for further research were identified based on the proposed features, research gaps, and limitations identified within this systematic review. A research gap that was consistently identified across studies was that engagement ratings from participants were frequently collected, though the qualitative processes underlying these ratings were rarely explored, hindering a deeper understanding of engagement. For example, Echeazarra et al [48] asked users to rate ease of use and usefulness of their intervention on a Likert scale but did not ask users to explain which features contributed to these ratings. The study by Cheng et al [35] provided a protocol for addressing this research gap, and users were prompted to explain which features contributed to their ratings. Another research gap was that clear and unexplained disparities in how different demographics engage with conversational agent–enabled interventions were identified in multiple studies [37,61]. For example, differing levels of engagement, trust, or perceived care were observed based on ethnicity, medication regimes, age, education, and computer literacy [37,61]. Further research could investigate how design features could be tailored for these different groups, such as using co-design methodologies to explore the role of redundancy in improving engagement of people with poor computer literacy through emphasizing multiple interaction modalities. We recommend factorial RCTs testing specific combinations of design features (eg, 2×2 design testing high or low redundancy×high or low anthropomorphism) across diverse populations, with standardized engagement metrics including both behavioral analytics and qualitative experience measures.

Conclusions

This systematic review provides key insights and rationales for developing conversational agent–enabled interventions for people self-managing CMD. However, because the included studies varied widely in agent modalities and engagement definitions and were largely qualitative, the generalizability of these insights is limited. Designing conversational agent–enabled interventions requires a balance of redundancy and personalization to support user autonomy without overwhelming users. Additionally, a balance between anthropomorphism and functionality is required to support user investment without frustrating users. Maintaining these balances is a key to improving engagement with conversational agent–enabled interventions and supporting people with CMD to self-manage their condition.

Acknowledgments

The authors thank the National Health and Medical Research Council (NHMRC) Centre of Research Excellence in Digital Technology to Transform Chronic Disease Outcomes, which aims to transform Australia’s digital health ecosystem. It is funded by the NHMRC (2020-2024: ID 1170937). NK accepts full responsibility for the work and conduct of this study.

Authors' Contributions

All authors contributed significantly to the research project and manuscript. NK and ATS performed the data collection. NK conducted the data analysis. All authors were involved in writing the manuscript and had final approval of the submitted and published versions.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 72 KB - mhealth_v13i1e67913_app1.pdf]

Multimedia Appendix 2
The complete search strategy.
[DOCX File , 47 KB - mhealth_v13i1e67913_app2.docx]

Multimedia Appendix 3
Comprehensive descriptions of how design features impacted engagement.
[DOCX File , 37 KB - mhealth_v13i1e67913_app3.docx]

References

1. Australian burden of disease study 2022. Australian Institute of Health and Welfare. 2022. URL: https://www.aihw.gov.au/reports/burden-of-disease/australian-burden-of-disease-study-2022/contents/interactive-data-on-disease-burden/burden-of-disease-in-australia [accessed 2024-05-06]
2. Yang Q, Cogswell ME, Flanders WD, Hong Y, Zhang Z, Loustalot F, et al. Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults. JAMA 2012;307(12):1273-1283 [FREE Full text] [doi: 10.1001/jama.2012.339] [Medline: 22427615]
3. Russell S, Ogunbayo OJ, Newham JJ, Heslop-Marshall K, Netts P, Hanratty B, et al. Qualitative systematic review of barriers and facilitators to self-management of chronic obstructive pulmonary disease: views of patients and healthcare professionals. NPJ Prim Care Respir Med 2018;28(1):2 [FREE Full text] [doi: 10.1038/s41533-017-0069-z] [Medline: 29343739]
4. Sattar N, Gill JMR, Alazawi W. Improving prevention strategies for cardiometabolic disease. Nat Med 2020;26(3):320-325 [FREE Full text] [doi: 10.1038/s41591-020-0786-7] [Medline: 32152584]
5. NCD Countdown 2030 Collaborators. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards sustainable development goal target 3.4. Lancet 2018;392(10152):1072-1088 [FREE Full text] [doi: 10.1016/S0140-6736(18)31992-5] [Medline: 30264707]
6. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. J Am Coll Cardiol 2017;70(1):1-25 [FREE Full text] [doi: 10.1016/j.jacc.2017.04.052] [Medline: 28527533]
7. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016;387(10027):1513-1530 [FREE Full text] [doi: 10.1016/S0140-6736(16)00618-8] [Medline: 27061677]
8. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002;288(14):1775-1779. [doi: 10.1001/jama.288.14.1775] [Medline: 12365965]
9. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
10. Fisher EB, Brownson CA, O'Toole ML, Shetty G, Anwuri VV, Glasgow RE. Ecological approaches to self-management: the case of diabetes. Am J Public Health 2005;95(9):1523-1535. [doi: 10.2105/AJPH.2005.066084] [Medline: 16051929]
11. Settineri S, Frisone F, Merlo EM, Geraci D, Martino G. Compliance, adherence, concordance, empowerment, and self-management: five words to manifest a relational maladjustment in diabetes. J Multidiscip Healthc 2019;12:299-314 [FREE Full text] [doi: 10.2147/JMDH.S193752] [Medline: 31118655]
12. Nguyen TNM, Whitehead L, Saunders R, Dermody G. Systematic review of perception of barriers and facilitators to chronic disease self-management among older adults: implications for evidence-based practice. Worldviews Evid Based Nurs 2022;19(3):191-200. [doi: 10.1111/wvn.12563] [Medline: 35032152]
13. Lawn S, Battersby M, Lindner H, Mathews R, Morris S, Wells L, et al. What skills do primary health care professionals need to provide effective self-management support? Seeking consumer perspectives. Aust J Prim Health 2009;15(1):37. [doi: 10.1071/py08053]
14. Dingler T, Kwasnicka D, Wei J, Gong E, Oldenburg B. The use and promise of conversational agents in digital health. Yearb Med Inform 2021;30(1):191-199 [FREE Full text] [doi: 10.1055/s-0041-1726510] [Medline: 34479391]
15. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ 2023;103:102274. [doi: 10.35542/osf.io/5er8f]

XSL•FO
RenderX

16. Chew HSJ. The use of artificial intelligence-based conversational agents (chatbots) for weight loss: scoping review and practical recommendations. JMIR Med Inform 2022;10(4):e32578 [FREE Full text] [doi: 10.2196/32578] [Medline: 35416791]
17. Noh E, Won J, Jo S, Hahm DH, Lee H. Conversational agents for body weight management: systematic review. J Med Internet Res 2023;25:e42238 [FREE Full text] [doi: 10.2196/42238] [Medline: 37234029]
18. Schöbel S, Schmitt A, Benner D, Saqr M, Janson A, Leimeister JM. Charting the evolution and future of conversational agents: a research agenda along five waves and new frontiers. Inf Syst Front 2023;26(2):729-754. [doi: 10.1007/s10796-023-10375-9]
19. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
20. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
21. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
22. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
23. Dijkstra A. Working mechanisms of computer-tailored health education: evidence from smoking cessation. Health Educ Res 2005;20(5):527-539. [doi: 10.1093/her/cyh014] [Medline: 15701665]
24. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
25. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
26. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
27. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
28. Bijkerk LE, Spigt M, Oenema A, Geschwind N. Engagement with mental health and health behavior change interventions: an integrative review of key concepts. J Context Behav Sci 2024;32:100748. [doi: 10.1016/j.jcbs.2024.100748]
29. Couper MP, Alexander GL, Zhang N, Little RJA, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
30. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
31. Hong YA, Jiang S, Liu PL. Use of patient portals of electronic health records remains low from 2014 to 2018: results from a national survey and policy implications. Am J Health Promot 2020;34(6):677-680. [doi: 10.1177/0890117119900591] [Medline: 32030989]
32. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
33. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. Internet Interv 2022;27:100495 [FREE Full text] [doi: 10.1016/j.invent.2022.100495] [Medline: 35059305]
34. Ting Y. The pitfalls of mobile devices in learning: a different view and implications for pedagogical design. J Educ Comput Res 2012;46(2):119-134. [doi: 10.2190/ec.46.2.a]
35. Cheng A, Raghavaraju V, Kanugo J, Handrianto Y, Shang Y. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. 2018 Presented at: 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC); January 12-15, 2018; Las Vegas, NV, United States p. 1-5. [doi: 10.1109/ccnc.2018.8319283]
36. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. Acceptability of an embodied conversational agent for type 2 diabetes self-management education and support via a smartphone app: mixed methods study. JMIR Mhealth Uhealth 2020;8(7):e17038 [FREE Full text] [doi: 10.2196/17038] [Medline: 32706734]

XSL•FO
RenderX

37. Apergi LA, Bjarnadottir MV, Baras JS, Golden BL, Anderson KM, Chou J, et al. Voice interface technology adoption by patients with heart failure: pilot comparison study. JMIR Mhealth Uhealth 2021;9(4):e24646 [FREE Full text] [doi: 10.2196/24646] [Medline: 33792556]
38. Epalte K, Tomsone S, Vētra A, Bērziņa G. Patient experience using digital therapy "Vigo" for stroke patient recovery: a qualitative descriptive study. Disabil Rehabil Assist Technol 2023;18(2):175-184. [doi: 10.1080/17483107.2020.1839794] [Medline: 33155507]
39. Kashyap N, Sebastian A, Lynch C, Jansons P, Maddison R, Dingler T, et al. Engagement with conversational agent-enabled interventions in cardiometabolic disease management: protocol for a systematic review. JMIR Res Protoc 2024;13:e52973 [FREE Full text] [doi: 10.2196/52973] [Medline: 39110504]
40. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.3. Cochrane Training. 2022. URL: https://www.cochrane.org/authors/handbooks-and-manuals/handbook [accessed 2023-07-01]
41. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
42. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc 2015;13(3):179-187. [doi: 10.1097/XEB.0000000000000062] [Medline: 26262565]
43. Tufanaru C, Munn Z, Aromataris E. Systematic reviews of effectiveness. In: JBI Manual for Evidence Synthesis. Adelaide: Joanna Briggs Institute; 2017.
44. Barker TH, Stone JC, Sears K, Klugar M, Tufanaru C, Leonardi-Bee J, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evid Synth 2023;21(3):494-506. [doi: 10.11124/JBIES-22-00430] [Medline: 36727247]
45. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
46. Balsa J, Neves P, Félix I, Pereira GM, Alves P, Carmo M, et al. Intelligent virtual assistant for promoting behaviour change in older people with T2D. 2019 Presented at: 19th EPIA Conference on Artificial Intelligence, EPIA 2019; September 3-6, 2019; Vila Real, Portugal p. 372-383. [doi: 10.1007/978-3-030-30241-2_32]
47. Balsa J, Félix I, Cláudio AP, Carmo MB, Silva ICE, Guerreiro A, et al. Usability of an intelligent virtual assistant for promoting behavior change and self-care in older people with type 2 diabetes. J Med Syst 2020;44(7):130. [doi: 10.1007/s10916-020-01583-w] [Medline: 32533367]
48. Echeazarra L, Pereira J, Saracho R. TensioBot: a chatbot assistant for self-managed in-house blood pressure checking. J Med Syst 2021;45(4):54. [doi: 10.1007/s10916-021-01730-x] [Medline: 33723721]
49. Gingele AJ, Amin H, Vaassen A, Schnur I, Pearl C, Brunner-La Rocca H, et al. Integrating avatar technology into a telemedicine application in heart failure patients: a pilot study. Wien Klin Wochenschr 2023;135(23-24):680-684 [FREE Full text] [doi: 10.1007/s00508-022-02150-8] [Medline: 36732377]
50. Gong E, Baptista S, Russell A, Scuffham P, Riddell M, Speight J, et al. My diabetes coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: randomized effectiveness-implementation trial. J Med Internet Res 2020;22(11):e20322 [FREE Full text] [doi: 10.2196/20322] [Medline: 33151154]
51. Guhl E, Althouse AD, Pusateri AM, Kimani E, Paasche-Orlow MK, Bickmore TW, et al. The atrial fibrillation health literacy information technology trial: pilot trial of a mobile health app for atrial fibrillation. JMIR Cardio 2020;4(1):e17162 [FREE Full text] [doi: 10.2196/17162] [Medline: 32886070]
52. Kimani E, Bickmore T, Trinh H, Ring L, Paasche-Orlow M, Magnani J. A smartphone-based virtual agent for atrial fibrillation education and counseling. 2016 Presented at: Intelligent Virtual Agents, IVA 2016; September 20-23, 2016; Los Angeles, CA, United States p. 120-127. [doi: 10.1007/978-3-319-47665-0_11]
53. Magnani JW, Schlusser CL, Kimani E, Rollman BL, Paasche-Orlow MK, Bickmore TW. The atrial fibrillation health literacy information technology system: pilot assessment. JMIR Cardio 2017;1(2):e7 [FREE Full text] [doi: 10.2196/cardio.8543] [Medline: 29473644]
54. Nassar CM, Dunlea R, Montero A, Tweedt A, Magee MF. Feasibility and preliminary behavioral and clinical efficacy of a diabetes education chatbot pilot among adults with type 2 diabetes. J Diabetes Sci Technol 2025;19(1):54-62. [doi: 10.1177/19322968231178020] [Medline: 37278191]
55. Pienkowska A, Ang C, Mammadova M, Mahadzir M, Car J. A diabetes education app for people living with type 2 diabetes: co-design study. JMIR Form Res 2023;7:e45490 [FREE Full text] [doi: 10.2196/45490] [Medline: 37721799]
56. Roca S, Lozano ML, García J, Alesanco Á. Validation of a virtual assistant for improving medication adherence in patients with comorbid type 2 diabetes mellitus and depressive disorder. Int J Environ Res Public Health 2021;18(22):12056 [FREE Full text] [doi: 10.3390/ijerph182212056] [Medline: 34831811]
57. Sagstad M, Morken N, Lund A, Dingsør LJ, Nilsen A, Sorbye L. Quantitative user data from a chatbot developed for women with gestational diabetes mellitus: observational study. JMIR Form Res 2022;6(4):e28091 [FREE Full text] [doi: 10.2196/28091] [Medline: 35436213]

XSL•FO
RenderX

58. Ter Stal S, Sloots J, Ramlal A, Op den Akker H, Lenferink A, Tabak M. An embodied conversational agent in an eHealth self-management intervention for chronic obstructive pulmonary disease and chronic heart failure: exploratory study in a real-life setting. JMIR Hum Factors 2021;8(4):e24110 [FREE Full text] [doi: 10.2196/24110] [Medline: 34734824]
59. Tongpeth J, Du HY, Clark RA. Development and feasibility testing of an avatar-based education application for patients with acute coronary syndrome. J Clin Nurs 2018;27(19-20):3561-3571. [doi: 10.1111/jocn.14528] [Medline: 29777554]
60. Tsai WC, Li WY, Tsai JY, Yang JS. Improving self-diet management of chronic kidney disease patients through chatbots. 2022 Presented at: Human Aspects of IT for the Aged Population: Technology in Everyday Living, Part II; June 26-July 1, 2022; Virtual Event p. 268-276. [doi: 10.1007/978-3-031-05654-3_18]
61. Zhang Z, Trinh H, Chen Q, Bickmore T. Adapting a geriatrics health counseling virtual agent for the Chinese culture. 2015 Presented at: Intelligent Virtual Agents, IVA 2015; August 26-28, 2015; The Netherlands p. 275-278. [doi: 10.1007/978-3-319-21996-7_28]
62. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. User experiences with a type 2 diabetes coaching app: qualitative study. JMIR Diabetes 2020 Jul 17;5(3):e16692 [FREE Full text] [doi: 10.2196/16692] [Medline: 32706649]
63. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
64. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
65. Blasiak A, Sapanel Y, Leitman D, Ng WY, De Nicola R, Lee VV, et al. Omnichannel communication to boost patient engagement and behavioral change with digital health interventions. J Med Internet Res 2022;24(11):e41463 [FREE Full text] [doi: 10.2196/41463] [Medline: 36383427]
66. Nass C, Steuer J, Tauber ER. Computers are social actors. 1994 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 24-28, 1994; Boston, MA, United States p. 72-78. [doi: 10.1145/191666.191703]
67. Ruslana K, Ning MA. Research on anthropomorphic visual image of digital avatar of intelligent chatbot. Kyiv National University of Technologies and Design. 2023. URL: https://er.knutd.edu.ua/bitstream/123456789/24651/1/APSD_2023_V1_P275-278.pdf [accessed 2024-05-06]
68. Thunström AO, Carlsen HK, Ali L, Larson T, Hellström A, Steingrimsson S. Usability comparison among healthy participants of an anthropomorphic digital human and a text-based chatbot as a responder to questions on mental health: randomized controlled trial. JMIR Hum Factors 2024;11:e54581 [FREE Full text] [doi: 10.2196/54581] [Medline: 38683664]

Abbreviations

AI: artificial intelligence
CMD: cardiometabolic disease
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by Z Ni; submitted 23.10.24; peer-reviewed by TF Ojo, CL Sánchez Bocanegra; comments to author 28.04.25; revised version received 29.05.25; accepted 11.07.25; published 18.09.25.

Please cite as:
Kashyap N, Sebastian AT, Lynch C, Jansons P, Maddison R, Dingler T, Oldenburg B
Engagement With Conversational Agent–Enabled Interventions in Cardiometabolic Disease Self-Management: Systematic Review
JMIR Mhealth Uhealth 2025;13:e67913
URL: https://mhealth.jmir.org/2025/1/e67913
doi:10.2196/67913
PMID:

©Nick Kashyap, Ann Tresa Sebastian, Chris Lynch, Paul Jansons, Ralph Maddison, Tilman Dingler, Brian Oldenburg. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effects of Remote Patient Monitoring on Health Care Utilization in Patients With Noncommunicable Diseases: Systematic Review and Meta-Analysis

Geir Smedslund[1], PhD; Nina Østerås[1], PhD; Christine Hillestad Hestevik[1,2], PhD

[1]Centre for Treatment of Rheumatic and Musculoskeletal Diseases (REMEDY), Diakonhjemmet Hospital, Oslo, Norway

[2]Health and Social Care Interventions, Division of Health Services, Norwegian Institute of Public Health, Oslo, Norway

Corresponding Author:
Geir Smedslund, PhD
Centre for Treatment of Rheumatic and Musculoskeletal Diseases (REMEDY)
Diakonhjemmet Hospital
Postboks 23 Vinderen
Oslo, 0319
Norway
Phone: 47 2245 4862
Email: geir.smedslund@dmp.no

Abstract

Background: Management of noncommunicable diseases (NCDs) is an increasing challenge for health care systems. Although remote patient monitoring presents a promising solution by utilizing technology to monitor patients outside clinical settings, there is a lack of knowledge about the effect on resource utilization.

Objective: This systematic review aimed to review the effects of remote patient monitoring on health care resource utilization by patients with NCDs.

Methods: Eligible randomized controlled trials (RCTs) involved digital transmission of health data from patients to health care personnel. Outcomes included hospitalizations, length of stay, outpatient visits, and emergency visits. A systematic literature search was performed in Medline, Embase, and Cochrane Central Register of Controlled Trials in June 2024. Titles, abstracts, and full texts were screened individually by 2 authors. Risk of bias was assessed, and data were extracted, analyzed, and pooled in meta-analysis when possible. Confidence in the estimates was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

Results: We included 40 RCTs published between 2017 and 2024. The largest group of NCDs was cardiovascular disease (16 studies). Remote patient monitoring may slightly decrease the proportion of hospitalizations compared with usual care (risk ratio [RR] 0.86, 95% CI 0.77 to 0.95; low certainty). Compared with usual care, remote patient monitoring had fewer or an equal number of hospitalizations (mean difference –0.13, 95% CI –0.29 to 0.03; low certainty). Hospital length of stay may be slightly reduced with remote patient monitoring compared with usual care (mean difference –0.84, 95% CI –1.61 to –0.06 days; low certainty). The proportion of outpatient visits showed probably little to no difference between remote patient monitoring and usual care (RR 0.94, 95% CI 0.87 to 1.02; moderate certainty). Compared with usual care, remote patient monitoring had slightly more outpatient visits, but the CI was wide (mean difference 0.41, 95% CI –0.22 to 1.03; low certainty). The results indicate a small or no difference between remote patient monitoring and usual care regarding proportion of emergency visits (RR 0.91, 95% CI 0.79 to 1.05; low certainty). We are uncertain whether remote patient monitoring increases or decreases the number of emergency visits, as the evidence was of very low certainty.

Conclusions: This systematic review showed that remote patient monitoring possibly led to lower proportions of patients being hospitalized, fewer hospitalizations, and shorter hospital length of stay compared with usual care. Patients undergoing remote monitoring had possibly more outpatient visits compared with usual care. The proportions of patients with outpatient visits or emergency visits were probably similar. Finally, we had very low certainty in the number of emergency visits. The results should be considered with caution as the certainty of evidence was moderate to very low. We did not find results regarding institutional stay.

(JMIR Mhealth Uhealth 2025;13:e68464) doi:10.2196/68464

XSL•FO
RenderX

KEYWORDS

remote patient monitoring; health care resource utilization; noncommunicable diseases; systematic review; hospitalizations; outpatient visits; remote consultation; chronic disease; patient monitoring; length of stay; emergency service; hospital

Introduction

Noncommunicable diseases (NCDs) including diabetes, cardiovascular diseases, chronic respiratory diseases, and mental health conditions lead to widespread human suffering globally and impose a high burden on health care systems, both financially and structurally [1,2]. More than 70% of all deaths globally are caused by NCDs, and cardiovascular disease accounts for most of the deaths [3]. The already substantial societal costs of NCDs are expected to rise even further in the years to come [4]. NCDs contribute to reduced productivity, shortened working lives, and premature mortality, with estimates suggesting a cumulative global loss of between US $30 trillion and US $47 trillion in economic output between 2011 and 2030 [5]. Furthermore, NCDs lead to high treatment costs, imposing a direct economic burden on health systems [1,6].

The follow-up of these individuals is resource-intensive and requires a high degree of interdisciplinary and holistic care over time. This leads to increased demands on both capacity and competence in health care services, as these patients often need both continuous monitoring and acute interventions [1,2]. As a result, there is significant interest in exploring and testing new care models aimed at extending the services to patients' homes through innovative technologies, which may contribute to preventing avoidable use of health care resources [7-9]. Welfare technology is highlighted as an important element in the development of the health and care sector [9].

Among these technologies, remote patient monitoring presents significant opportunities to improve the timeliness of care, enhance health outcomes, and potentially reduce unnecessary use of health care resources [9,10]. Remote patient monitoring utilizes technological tools to track patients' health status outside of typical clinical settings, including patients' homes or other remote locations. This method involves sending health data directly to health care providers via automated electronic systems or web- and phone-based interfaces, facilitating timely interventions if health status deteriorates [11,12].

This process ensures that health care providers can respond rapidly and potentially prevent hospitalizations. Unlike traditional calendar-based hospital follow-up appointments, remote monitoring enables customized treatment plans tailored to each patient's specific needs and disease activity [11,12]. Routine monitoring of patient's health data and receiving alerts when measurements surpass a predetermined threshold can enable health care providers to intervene promptly. This approach offers potential advantages for both individuals and society by lessening the burden on patients and decreasing the need for unnecessary medical visits [10,13].

Moreover, the use of welfare technology hopefully can free up resources within health care services, allowing them to be allocated more effectively to patients with greater needs. However, findings from studies examining patients' utilization of health care services with remote patient monitoring compared with usual care have shown mixed results [14]. Therefore, we lack definitive knowledge about the effect of remote monitoring on resource utilization. [14] The purpose of this systematic review was to examine the effects of remote patient monitoring on health care resource utilization for individuals with NCDs, compared with usual care. The protocol for this review was preregistered in PROSPERO (CRD42023431366) on June 19, 2023. This review is an update of a previous review [15] published in Norwegian with the key messages and executive summary provided in English.

Methods

We conducted a systematic review according to the guidelines described in the Cochrane handbook [16] and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [17]. The completed PRISMA checklist is provided in Multimedia Appendix 1.

Selection Criteria

Eligible studies were randomized controlled trials investigating the effect of remote monitoring of patients with NCDs (including chronic pulmonary diseases, diabetes, cardiovascular disease, cancer, mental health disorders, and comorbidities) on health care utilization, compared with usual care, published in English or a Scandinavian language.

The interventions had to involve digital transmission of health data from the patient to health care personnel, who would then contact the patient and provide medical support and guidance if the measurements indicated signs of deterioration or fell outside normal values. We also included studies that incorporated additional components alongside remote patient monitoring, such as rehabilitation guidance, exercise prescriptions, and other forms of support.

Studies involving digital follow-up without any element of self-reporting, such as implants that transmit data automatically, were excluded.

Given the rapid advancements in digital health technologies, including remote monitoring tools, software platforms, and communication infrastructure, we prioritized newer studies to ensure the findings reflected current clinical practice and technological capabilities. Older studies may be based on outdated technologies or care models that no longer represent the current standard of care, potentially limiting the applicability and relevance of their results. Hence, only studies published from 2017 and onwards with interventions conducted in 2013 or later (within the last 10 years at the time of our first systematic search) were included.

The studies had to report one or more of the following outcomes: hospitalizations (proportions or numbers), length of hospital stay, outpatient consultations (proportions or numbers), use of emergency services (proportions or numbers), visits to general

XSL•FO
RenderX

practitioners, use of home health care services, and use of institutional stay.

Search Strategy

A librarian developed a search strategy corresponding to the inclusion criteria and searched the following databases: Medline (Ovid), Embase (Ovid) and Cochrane Central Register of Controlled Trials (Wiley). Searches were conducted in 3 rounds: April 2022, May 2023, and June 2024 (Multimedia Appendix 2).

Inclusion Process

The authors independently screened the references identified in the previous review and the updated searches. We utilized the "priority screening" function in EPPI-Reviewer [18] to enhance the efficiency of screening titles and abstracts. This machine learning feature identifies the characteristics of included and excluded studies and predicts the likelihood of relevance for each reference. Priority screening prioritizes relevant records at the beginning of the screening process and relegates irrelevant records toward the end. Two authors (GS and CHH) independently screened the titles and abstracts to determine their eligibility according to our inclusion criteria. Disagreements were solved through discussion. We decided to stop screening after assessing 400 titles and abstracts in a row without including any additional references, as the inclusion rate had plateaued, suggesting that the algorithm had already identified all relevant studies

Subsequently, we obtained the full texts of references deemed relevant. In pairs, the authors (GS, CHH, NØ) then independently reviewed these full texts and made a final assessment to determine which studies to include. Disagreements were solved through consensus.

Risk of Bias

Two authors (CHH and GS) independently assessed risk of bias for each outcome with the Cochrane risk-of-bias tool (RoB 2) for randomized trials [19]. Disagreements were solved through discussion. RoB 2 assesses risk of bias within 5 domains: randomization process, deviations from the intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Overall risk of bias for each outcome was classified as low risk, some concerns, or high risk based on the domain with the highest risk of bias.

Data Extraction

One author (GS) extracted data from the included studies, and another (CHH or NØ) checked the data against the publication. We extracted the following data: study and participant characteristics, information about the interventions, observation period, and outcomes related to resource use. These outcomes included hospitalizations (including readmissions), length of hospital stay, and number of outpatient visits (including clinic visits, outpatient visits, and medical consultations with private physicians and general practitioners), and emergency services. In case of missing information, we contacted the study authors.

Statistical Analysis

We conducted meta-analyses combining studies reporting on the same outcome in RevMan Web [20]. Some of the effect sizes were transformed, as the original data were not reported in a format ready for effect size calculation. For dichotomous outcomes (proportion hospitalized, proportion of outpatient visits, proportion with emergency visits) we performed random-effects analyses with the risk ratio (RR) as the measure of relative effect. For continuous outcomes (number of hospitalizations, hospital length of stay, number of outpatient visits, number of emergency visits) we performed random-effects analyses with the mean difference as the measure of relative effect. A decision taken after the protocol was published was to add results of absolute effects on all outcomes to enhance interpretability and clinical relevance. We believe this can reduce the risk of misinterpreting large relative effects with low absolute benefit and facilitate meaningful comparisons and meta-analyses across studies with varying baseline risks

We did not plan any subgroup analyses or meta-regression in the protocol, but we did the following post hoc analyses: If there were 2 or more studies with the same NCD for an outcome, we subgrouped the meta-analysis per disease group. If there were 10 or more studies in a meta-analysis, we performed mixed effects meta-regression using the R-package [metafor] [21]. Explanatory variables were mean age and publication year. In addition, if there were 10 or more studies in a meta-analysis, we produced funnel plots and performed the Egger test in R [metafor].

Grading of Recommendations Assessment, Development, and Evaluation

We used the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach [22] to assess our certainty of the effect estimates per outcome within the following 5 domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. The GRADE approach specifies 4 levels of certainty: high, moderate, low, or very low. Table 1 illustrates how the certainty of the evidence and the importance of the outcome influence the description of the direction and size of an intervention's effect.

XSL•FO
RenderX

Table 1. How certainty of the evidence and importance of the outcome influence the description of an intervention's effect, as described in [23].

Level of certainty	Important benefit or harm	Less important benefit or harm	No important benefit or harm or null effect
High certainty	Increases or decreases	Increases or decreases slightly	Little to no difference
Moderate certainty	Probably increases or decreases	Probably increases or decreases slightly	Probably little to no difference
Low certainty	May increase or decrease	May increase or decrease slightly	May make little to no difference
Very low certainty	Uncertain whether intervention increases or decreases outcome	Uncertain whether intervention increases or decreases outcome	Uncertain whether intervention increases or decreases outcome

Results

Included Studies

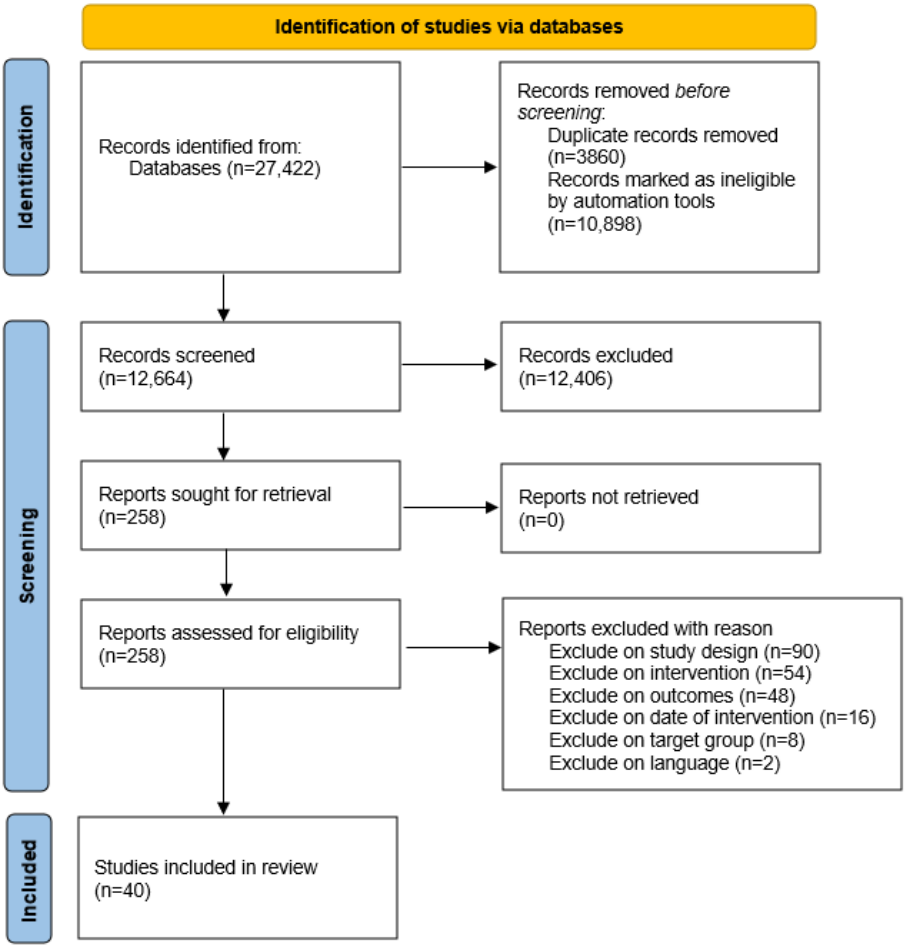
We identified 23,562 unique references in the literature search, of which 258 studies were evaluated in full text. In total, 218 of these studies were excluded for the following reasons: wrong study design (n=90), wrong intervention (n=54), wrong outcomes (n=48), observation period before 2013 (n=16), wrong target group (n=8), and wrong language (n=2; Figure 1).

We included 40 studies [24-62] published between 2017 and 2024 (Tables 2 and 3).

The studies were distributed across various countries, mainly representing high-income countries, with the highest number conducted in the United States and Denmark (Table 2).

The most frequently reported outcomes were proportion of hospitalized patients (including readmissions), proportion with emergency visits, and length of hospital stay.

Figure 1. Flowchart of study selection process.



XSL•FO
RenderX

Table 2. Key characteristics of included studies (N=40).

Characteristic	Details
Total participants, n (range in individual studies)	11,617 (31-1225)
Female participants (N=11,617), n (%)	4888 (42)
Studies with mean age ≥50 years, n (%)	32 (80)
Chronic conditions , n (%) of studies	
Cardiovascular disease	16 (40)
Chronic obstructive pulmonary disease	8 (20)
Cancer	5 (13)
Multiple chronic conditions	4 (10)
Mental illness	3 (8)
Cystic fibrosis	2 (5)
Diabetes mellitus	2 (5)
Countries represented, n (%) of studies	
United States	5 (13)
Denmark	5 (13)
Canada	4 (10)
Spain	3 (8)
Australia	3 (8)
United Kingdom	2 (5)
Poland	2 (5)
Japan	2 (5)
Netherlands	1 (3)
Norway	1 (3)
Germany	1 (3)
Colombia	1 (3)
China	1 (3)
New Zealand	1 (3)
France	1 (3)
Russia	1 (3)
Singapore	1 (3)
Korea	1 (3)
Brazil	1 (3)
Argentina	1 (3)
Hong Kong	1 (3)
Multinational	1 (3)

XSL•FO
RenderX

Table 3. Study characteristics (N=40).

First author (year)/country	Name of trial or program (app)	Sample size, n	Diagnoses	Age (years), mean (SD)	Women, %	Intervention	Relevant outcomes	Follow-up
Abelsen (2022)/Norway [63,64]	NR[a]	538	Chronic diseases (COPD[b], T2DM[c], HF[d], cancer, mental illness)	69.7	49.0	Tablet for patients to answer simple questions about their health or to monitor measurements related to health status (such as BP[e], blood glucose, oxygen saturation, or weight); automatic transfer of measurements from the home telemonitoring devices to the patient's tablet and follow-up center	Number of hospitalizations, length of hospital stay, use of acute services, outpatient visits and home care services	6 and 12 months
Absolom (2021)/England [24]	App: eRAPID	508	Colorectal, breast, or gynecological cancers	56.0 (11.8)	79.9	An online eHealth system for patients to self‐report symptoms during cancer treatment; provision of automated severity-dependent patient advice guiding self-management or medical contact	Proportion hospitalized and proportion with emergency visits	6, 12, and 18 weeks
Achury-Saldana (2024)/Colombia [25]	App: ControlVit	175	HF	66 (11.9)	28.6	Daily (weight, arterial pressure, HR[f], and symptoms) questionnaire; alerts when an increase in weight or change in HR or arterial pressure was detected	Proportion hospitalized	6 months
Ackermann (2021)/Australia [26]	Trial: MEL-SELF pilot study; app: ASICA[g] SkinChecker app	100	Localized melanoma	58.7 (12.0)	46	Usual care plus patient-led surveillance with instructional videos on how to perform SSE[h], reminders to undertake SSE, a mobile dermatoscope attached to their smartphone, and an app that facilitated store-and-forward teledermatology and fast-tracked unscheduled clinic visits	Number of outpatient visits	6 months
Agarwal (2019)/Canada [27]	App: BlueStar	240	T2DM	51.8 (10.7)	48.0	Information related to T2DM management entered into the app by patients who received customized, evidence-based messages to impact motivation, behavior, and education; the transfer of health data to the user's clinician also facilitated by the app	Proportion hospitalized and proportion with outpatient and emergency visits	3 and 6 months
Alshahrani (2024)/UK [28]	Trial: Tele-ACS study	337	Acute coronary syndrome	58.1	13.9	12-lead ECG[i] belt with automatic monitoring of BP and pulse oximeter; remote consultation with clinician	Length of hospital stay and proportion with emergency visits	6 months
Andersen (2023)/Denmark [29]	NR	222	COPD	70 (64-76)[j]	61	Home measurement of oxygen saturation, HR, peak expiratory flow, body weight, and a questionnaire on dyspnea, cough, sputum volume, and color; results sent to a server with green, yellow, and red alerts	Proportion hospitalized, and length of hospital stay	6 and 24 months

XSL•FO
RenderX

First author (year)/country	Name of trial or program (app)	Sample size, n	Diagnoses	Age (years), mean (SD)	Women, %	Intervention	Relevant outcomes	Follow-up
Angermann (2023)/Germany [30]	Trial: randomized INH[k] trial	715	Acute systolic HF	70 (60-78)[j]	29	Structured telephone monitoring based on patient-recorded vital signs and a 19-item standardized questionnaire addressing factors such as adverse events since previous contact, medication adherence, and satisfaction with remote patient monitoring	Proportion hospitalized	18 and 60 months
Benzo (2022)/US [31]	NR	375	COPD	69.0 (9.5)	56.5	Weekly HC[l] calls and a remote monitoring system that had a computer tablet, an activity monitor, and an oximeter; tablet-gathered data transmitted to an online patient portal to monitor compliance with the rehabilitation routine and daily physical activity; aim of weekly HC call to be a behavior change intervention supported by the motivational interviewing and self-efficacy theories	Proportion hospitalized and with emergency visits	3 months
Broadbent (2018)/New Zealand [32]	App: iRobi robot	60	COPD	69.8 (10.1)	61.5	Robot provided to patients at home programmed to deliver COPD management consisting of several components guided by a clinical pathway; overall program designed to monitor health and prompt medical contact if health was deteriorating	Proportion hospitalized and hospital length of stay	4 months
Dorsch (2021)/US [33]	App: Manage-HF4Life	83	Left ventricular ejection fraction	61.1 (9.0)	35.0	Self-monitoring with a mobile app, along with a Fitbit physical activity monitor and scale	Proportion hospitalized	12 weeks
Faurholt-Jepsen (2021)/Denmark [34]	Trial: RADMIS; app: Monsenso system smartphone app	98	Bipolar disorder	42.7 (13.5)	25.6	Daily self-monitoring of mood, level of activity, duration of sleep, and medication intake; collected smartphone data on a webpage regularly examined by study nurses who, based on the patients' needs and collected data, contacted the patients and gave advice	Proportion hospitalized and duration of psychiatric readmissions	3 and 6 months
Franz (2022)/US [35]	Trial: eICE[m] study	267	Cystic fibrosis	28.8% <18 years	50.9	Symptoms measured by patients using the CFRSD[n] twice per week; study sites automatically alerted if a participant's FEV1[o] fell below baseline >10% or if CFRSD worsened in ≥2 of 8 respiratory symptoms	Proportion hospitalized, length of hospital stay, and number of outpatient and emergency visits per participant per year of follow-up	12 months

XSL•FO
RenderX

First author (year)/country	Name of trial or program (app)	Sample size, n	Diagnoses	Age (years), mean (SD)	Women, %	Intervention	Relevant outcomes	Follow-up
Galinier (2020)/France [36]	Trial: OSICAT	937	Acute HF	69.9 (12.4)	27.8	Telemonitoring program with 2 elements: electronic devices and a personalized educational component; telemonitoring group: identical set of electronic scales to measure body weight and a device to answer 8 symptom questions provided to each participant; measurement of body weight and recording of HF symptoms communicated daily to a secure server and analyzed automatically by an expert system that generated alerts, as necessary, with the aim of predicting episodes of cardiac decompensation	Proportion hospitalized and hospital length of stay	18 months
Garanin (2022)/Russia [37]	App: INME-01 tonometer	392	HF and acute coronary syndrome	66.3 (47.9)	49.4	Monitoring carried out with the certified INME-01 tonometer with an integrated GSM[p] module that allows BP and pulse rate measurements and ability to transmit results via a cellular communication channel to a research center; based on the results of the received data, doctor able to contact the patient and adjust the previously prescribed treatment	Proportion hospitalized and length of hospital stay	3 months
Gomis-Pastor (2023)/Spain [38]	App: mHeart	134	Heart transplant	45 (16)	31	Multifaceted theory-based interventions provided during the study period to optimize therapy management using a mobile app	Proportion hospitalized and proportion with outpatient and emergency visits	6 and 12 months
Greer (2020)/US [39]	NR	181 (randomized)	Cancer	53.3 (12.9)	53.6	Smartphone mobile app with medication plan, reminders, symptom-reporting module, patient education, and Fitbit integration for tracking physical activity; used for 12 weeks	Number of hospitalizations and emergency visits	12 weeks
Hernandez-Quiles (2024)/Spain [40]	TELECARE	510	Advanced heart and lung disease	76 (14)	54.5	Monitoring of biological and medical questionnaires available in real time for the health care team; alarm routines	Proportion hospitalized, length of hospital stay, and proportion and number with emergency visits	45, 90, and 180 days
Indraratna (2022)/Australia [41]	App: TCC[q] smartphone app	164	Acute coronary syndrome or HF	61.5 (12.5)	21.0	Smartphone app with Bluetooth-enabled devices for daily monitoring of weight, BP, and physical activity, along with educational push notifications	Proportion hospitalized	30 days
Jiang (2022)/Singapore [43]	App: NSSMP[r]	114	T2DM	52.8 (11.0)	30.0	6-month NSSMP consisting of an education session and a smartphone app for self-monitoring and recording of blood glucose, diet, and exercise	Number of hospitalizations and number of outpatient and emergency visits	6 months

XSL•FO
RenderX

First author (year)/country	Name of trial or program (app)	Sample size, n	Diagnoses	Age (years), mean (SD)	Women, %	Intervention	Relevant outcomes	Follow-up
Jiang (2024)/China [42]	NR	148	COPD with hypercapnic chronic respiratory failure	72.7 (6.79)	14.2	Internet of Things–based management based on telemonitoring of clinical and ventilator parameters for 12 months	Number with readmission	12 months
Krzesinski (2021)/Poland [44]	Trial: AMULET	605	HF and left ventricular ejection fraction	67 (14)[s]	21	Nurse-led noninvasive assessments supporting remote therapeutic decisions for patients after an episode of acute HF, including vital sign monitoring, assessments of HF symptoms, and physician recommendations	Proportion hospitalized	12 months
Laursen (2021)/Denmark [45]	Trial: ENTER; app: mDiary	78	Borderline personality disorder	26.8 (18.1)	87.0	Mobile diary app: psychoeducational material (dialectical behavior therapy) and visualizations of participant's data; recording of symptom information (eg, emotional dysregulation and thoughts of suicide and self-harm) by participants so health care professionals could follow them in real time and use the registrations in the treatment	Proportion hospitalized and number of outpatient visits	1 month
Lee (2023)/Korea [46]	App: PRO-CTCAE	222	Breast, lung, head and neck, esophageal, or gynecologic cancer	55.9 (9.7)	60.5	Completed symptom questionnaire every 7 days for 8 weeks; data accessible by clinicians using a web dashboard	Proportion hospitalized and proportion with outpatient and emergency visits (at least once)	2 months
Mizukawa (2019)/Japan [47]	NR	39	HF	72.4 (12.7)	48.7	3 arms: telemonitoring (collaborative management), self-management education, and usual care; all involved self-monitoring weight, BP, and pulse; 12-month disease management program also part of intervention groups; telemonitoring group: noninvasive physiologic telemonitoring devices to measure BP, pulse rate, and body weight for 12 months and data transmitted to the nurse's computer; self-management group: described as "self-management"	Proportion hospitalized and length of hospital stay	24 months
Piotrowicz (2020)/Poland [48]	Trial: TELEREH-HF trial/CardioMessenger transmitter of the CareLink network, or erlin@home wireless transmitter	850	HF	62.4 (10.5)	11.4	Hybrid comprehensive telerehabilitation: supervised exercise training monitored with tele-ECG; monitoring of BP and body weight	Proportion hospitalized	9 weeks

XSL•FO
RenderX

First author (year)/country	Name of trial or program (app)	Sample size, n	Diagnoses	Age (years), mean (SD)	Women, %	Intervention	Relevant outcomes	Follow-up
Rohde (2024)/Brazil [49]	Trial: MESSAGE-HF study; app: Clever Care	699	HF (with reduced left ventricular ejection fraction [<40%])	61.2 (14.5)	34.2	Telemonitoring group: 4 daily short text messages to optimize self-care, active engagement, and early intervention; red flags based on feedback messages triggering automatic diuretic adjustment and/or telephone call from the health care team	Proportion hospitalized and with emergency visits	180 days
Shimoyama (2023)/Japan [50]	Trial: Telenursing system (COMPAS)	31	Chronic respiratory failure	73.0 (10.2)	40	Once a day, questions answered by each participant from a list of options; data on physical condition, including vital signs, respiratory symptoms, food intake, excretion, medication use, physical symptoms other than respiratory symptoms, and questions to the medical staff sent to the server and collected; transmitted data monitored remotely, and nursing support determined based on the results of each question item set in advanced consultation with the physician in charge	Number of hospitalizations and length of hospital stay	3 months
Soriano (2018)/Spain [51]	Trial: PROMETE II	229	COPD	71.4 (8.4)	19.6	A pulse oximeter, BP gauge, spirometer, and respiratory rate and oxygen therapy compliance monitor connected to a modem that uploaded readings to secure servers—performed daily for 12 months; alerts automatically triggered when measurements were exceeding thresholds or missing	Proportion hospitalized and length of hospital stay	12 months
Stamenova (2020)/Canada [52]	App: Cloud DX Connected Health Kit	81	COPD	72.4 (9.3)	46.0	Cloud DX Connected Health Kit: custom tablet computer, Pulse wave wrist cuff monitor (measured BP), oximeter, weighing scale and thermometer; data from all devices transmitted to a database; 3-armed study, including remote patient monitoring group and self-monitoring group: oximetry and BP daily for 6 months; when readings fell outside predetermined thresholds, clinical project specialist notified who took appropriate action	Number of hospitalizations, hospital length of stay, proportion with outpatient visits, and number of emergency visits	6 months

XSL•FO
RenderX

First author (year)/country	Name of trial or program (app)	Sample size, n	Diagnoses	Age (years), mean (SD)	Women, %	Intervention	Relevant outcomes	Follow-up
Temple-Oberle (2023)/Canada [53]	App: RecoverWell	72	Cancer (breast, benign ovarian or uterine mass, endometrial, or cervical)	54.90 (11.18)	100	Data input: drain outputs, surgical site photographs, and QoR15[t] and EORTC[u] scales; computer used by trained and designated portal monitor and patient's surgeon to access the data and monitor patients; any red flags brought to the surgeon's attention and prompted action either via a telephone call or an in-person appointment depending on the issue	Proportions with outpatient and emergency visits	6 weeks
Tønning (2020/2021)/Denmark [54]	Trial: RADMIS; app: Monsenso system	120	Unipolar depressive disorder	43.9 (14.2)	52.5	Multimodal monitoring and treatment based on a smartphone-based system (Monsenso system); patient monitoring of mood, sleep, and activity daily, and study nurse checking the data 3 times weekly and reacting according to the data presented	Proportion hospitalized and length of hospital stay	6 monhts
Wagenaar (2019)/Netherlands [55]	App/program: EACP[v] including "Heartfailurematters.org" website	450	HF	66.8 (11.0)	25.8	3-arm trial: (1) EACP, (2), website heartfailurematters.org, (3) usual care; EACP group: use of the e-Vita platform with telemonitoring of weight, BP, and HR and nurse alerts if values were abnormal; website group encouraged to use the website, information leaflet, and a reminder to use the website	Proportion hospitalized	12 months
Walker (2018)/ Spain, United Kingdom, Slovenia, Estonia, and Sweden [56]	Trial: CHROMED; app: RESMON PRO DIARY using FOT, a touchscreen computer, and a mobile modem	312	COPD	71[w]	34.0	Monitoring platform used for 9 months; measurements of within-breath respiratory mechanical impedance, BP, oxygen saturation, HR, and body temperature; contact with the study nurse triggered by abnormal values	Proportion hospitalized and length of hospital stay	12 months
Wang (2023/Hong Kong [57]	App: telemedicine app and system (HealthCap)	49	Hypertension	59.9 (8.4)	53	BP readings reported by patients twice daily for a week before their consultation, with data sent to a secure computer; review of results by physicians prior to appointments; and optimal 7-day average BP and no issues, appointment deferred for 3 months; suboptimal readings or presence of symptoms, consultation attendance by patients mandatory, and doctors able to adjust medication according to guidelines	Proportion hospitalized and proportion or number with outpatient visits	6 months
Ware (2022)/Canada [58]	App: Medly smartphone app	96	T2DM, hypertension, or HF	59.0 (12.6)	44.0	Telemonitoring group: patient measurements of symptoms; alerts to the clinical team via email and a secure web portal when recorded values (weight, BP, blood glucose) exceeded target thresholds	Proportion hospitalized, number of hospitalizations, and proportions and numbers of emergency and outpatient visits	6 months

XSL•FO
RenderX

First author (year)/country	Name of trial or program (app)	Sample size, n	Diagnoses	Age (years), mean (SD)	Women, %	Intervention	Relevant outcomes	Follow-up
Widmer (2017)/US [59]	NR	71	Acute coronary syndrome	63.0 (10.8)	18.6	Digital health intervention involving reporting of dietary and exercise habits throughout cardiac rehabilitation and educational information for patients' healthy lifestyles; patient-entered values in the app for BP, lipids, glucose, or weight that were out of normal range, patients asked to consult their physician	Proportion hospitalized and proportion with emergency visits	3 and 6 months
Witt Udsen (2017)/Denmark [60]	TeleCare North	1225	COPD	69.5 (9.4)	51.7	A tablet at home; information to participants about COPD management and transmission by participants of data on BP, pulse, blood oxygen saturation, and body weight to health personnel	Number of hospitalizations, length of hospital stay, numbers of outpatient and emergency visits, and use of home care services	12 months
Wood (2020)/Australia [61]	NR	60	Cystic fibrosis	31.0 (9.0)	48.7	Smartphone app with 14 questions about respiratory symptoms and well-being to be answered weekly or sooner when worsening of symptoms; if symptoms worsened, call from a nurse	Proportion hospitalized, length of hospital stay, and number of outpatient visits	12 months

XSL•FO
RenderX

First author (year)/country	Name of trial or program (app)	Sample size, n	Diagnoses	Age (years), mean (SD)	Women, %	Intervention	Relevant outcomes	Follow-up
Yanicelli (2021)/Argentina [62]	NR	30	HF	52.0 (NR)	20.0	Home telemonitoring system consisting of an app collecting weight, BP, and HR measurements and symptoms (eg, swelling in ankles, legs, shortness of breath); app-based alert to a physician if measurements are outside normal ranges; also included educational functionality	Proportion hospitalized	3 months

[a]NR: not reported.
[b]COPD: chronic obstructive pulmonary disease.
[c]T2DM: type 2 diabetes mellitus.
[d]HF: heart failure.
[e]BP: blood pressure.
[f]ASICA: Achieving Self-directed Integrated Cancer Aftercare.
[g]HR: heart rate.
[h]SSE: skin self - examination.
[i]ECG: electrocardiogram.
[j]Median (range).
[k]INH: Interdisciplinary Network Heart Failure.
[l]HC: health coach.
[m]eICE: early Intervention in Cystic Fibrosis Exacerbation.
[n]CFRSD: Cystic Fibrosis Respiratory Symptom Diary.
[o]FEV1: forced expiratory volume in one second.
[p]GSM: Global System for Mobile Communications.
[q]TCC: TeleClinical Care.
[r]NSSMP: nurse-led smartphone-based self-management program.
[s]Median (IQR).
[t]QoR15: quality of recovery questionnaire.
[u]EORTC: European Organisation for Research and Treatment of Cancer.
[v]EACP: eHealth adjusted care pathway.
[w]Median.

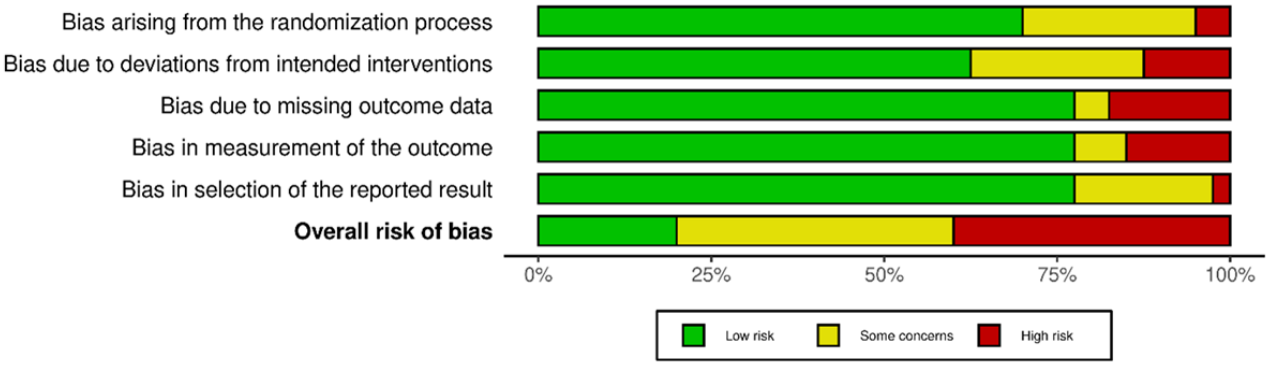
For one study by Abelsen et al [64], we were unable to calculate effect estimates. As a result, the study was not included in the meta-analyses, and we did not assess our confidence in the results. The findings from this study are described narratively. Table 2 shows the main characteristics of the included studies. Some of the trials had specific names, and some utilized a specific app, which are listed in the second column. The interventions were mostly similar, but we provided a detailed description of each. Although the trials had many outcomes, we included only those relevant for this review. Follow-up times ranged from 1 month to 60 months, with 37 studies having follow-up times between 3 months and 12 months (Table 3).

Risk of Bias

Of 40 studies, 8 were assessed as having a low risk of bias, 16 had some concerns, and the remaining 16 were assessed as having a high risk of bias. Figure 2 shows a summary of the risk of bias assessments.

XSL•FO
RenderX

Figure 2. Summary risk of bias assessment.



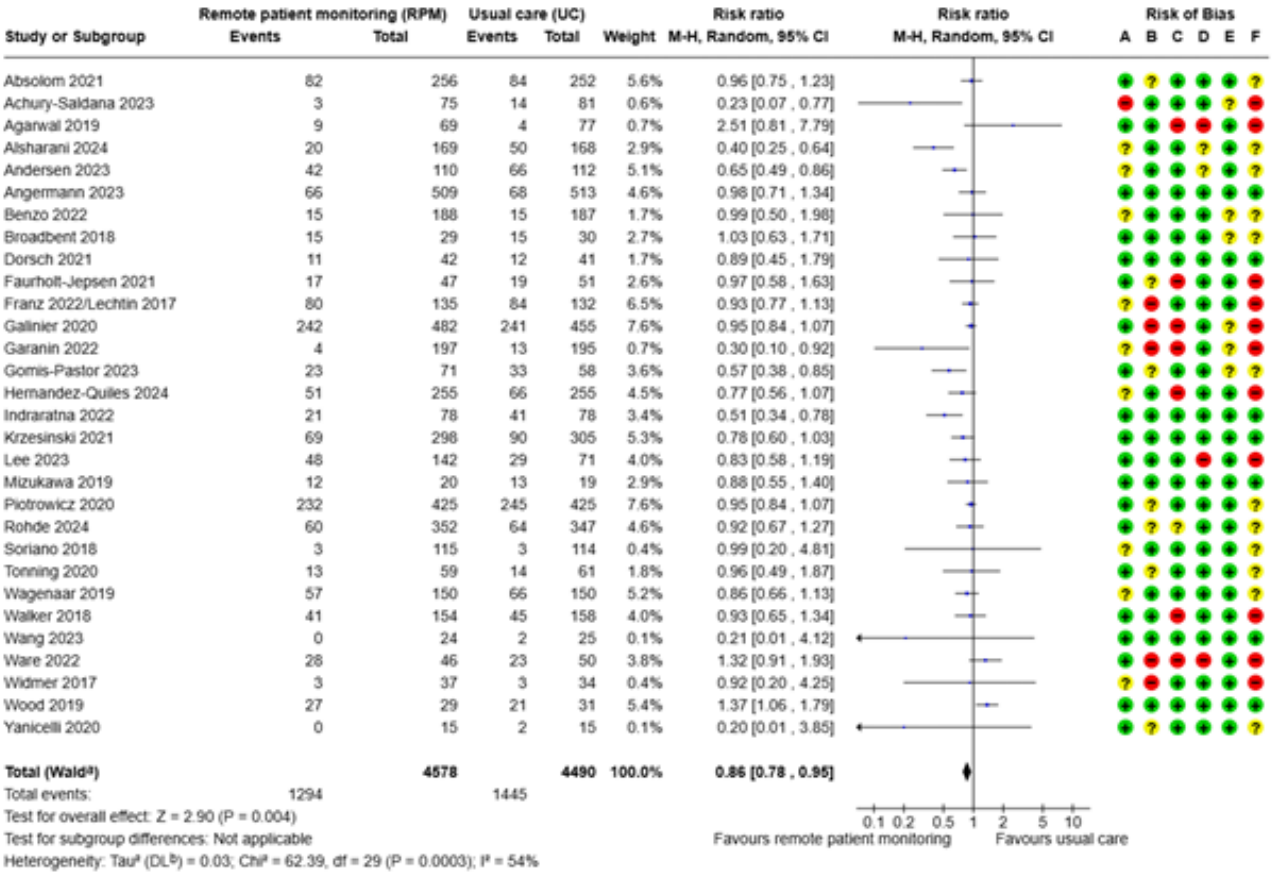
Effects of Interventions

Proportion of Hospitalized Patients

Of the included studies, 30 reported the proportion of hospitalized patients. The RR was 0.86 (95% CI 0.77 to 0.95; low certainty; Figure 3, Multimedia Appendix 2). There was significant heterogeneity (I^2: 55%). Mixed effects meta-regression analysis indicated that newer studies were more in favor of remote patient monitoring (coefficient for publication year: –0.06, *P*=.008; meaning that, for every increase of 1 publication year, the RR decreased by 0.06.) Mean age did not have a significant effect on proportion hospitalized. When both explanatory variables were entered together (full model), publication year was still significant (coefficient: –0.07, *P*=.006). In the full model, I^2 was 46%. Therefore, there was still some unexplained heterogeneity. The Egger test was not significant (*P*=.07), not indicating publication bias. In the subgroup of cardiovascular disease (16 studies), the RR was 0.76 (95% CI 0.65 to 0.88). We did not use GRADE for this subgroup result. The only other diagnostic subgroup with more than 2 studies was chronic obstructive pulmonary disease (COPD; 5 studies). The RR was 0.81 (95% CI 0.66 to 1.00). For the other subgroups, there were no important differences between remote patient monitoring and usual care.

XSL•FO
RenderX

Figure 3. Meta-analysis of the proportion of hospitalized patients in groups randomized to remote patient monitoring or usual care, with the risk of bias assessments for the outcome shown on the right for each study.



Footnotes

[a]CI calculated by Wald-type method.

[b]Tau² calculated by DerSimonian and Laird method.

Risk of bias legend

(A) Bias arising from the randomization process

(B) Bias due to deviations from intended interventions

(C) Bias due to missing outcome data

(D) Bias in measurement of the outcome

(E) Bias in selection of the reported result

(F) Overall bias

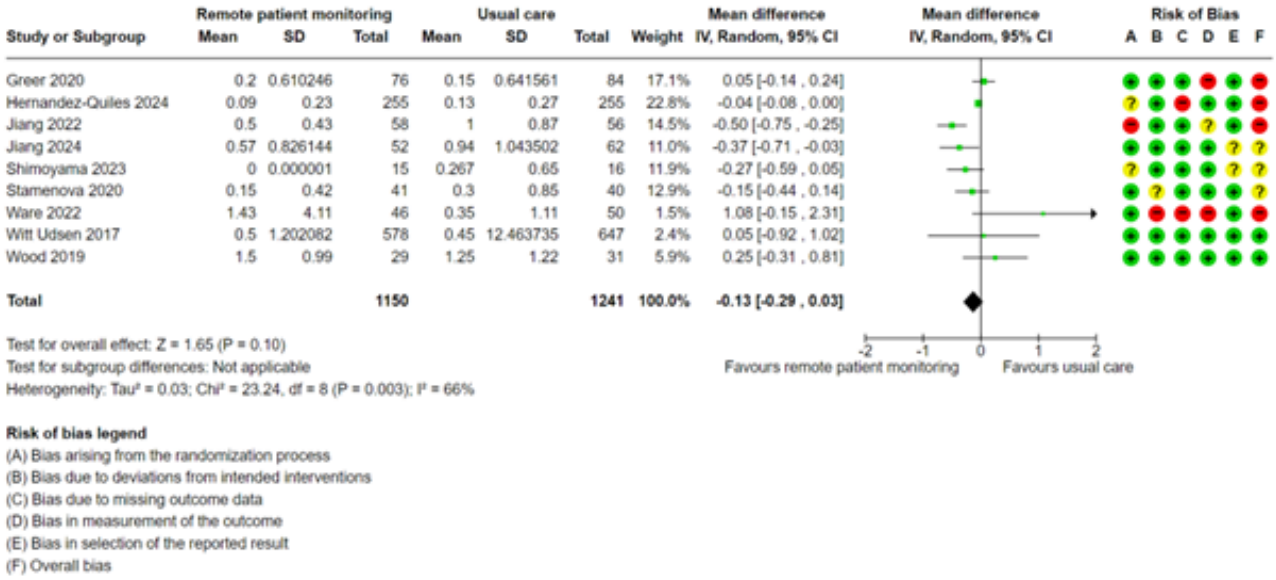
Laursen et al [45] reported the proportion with hospitalizations, but it was only reported as <5. Therefore, this could not be included in the meta-analysis. Jiang et al [43] stated that the 12-month risk of readmission was 34.3% in intervention group compared with 56.0% in the control group, but these numbers could not be included in the meta-analysis.

Number of Hospitalizations

Patients receiving usual care had a mean number of 0.42 hospitalizations. A meta-analysis with 9 studies showed that the group with remote patient monitoring had 0.13 fewer hospitalizations (95% CI –0.29 to 0.03; low certainty; Figure 4, Multimedia Appendix 2). We did not perform mixed effects meta-regression, subgroup, or test for publication bias for this outcome.

XSL•FO
RenderX

Figure 4. Meta-analysis of the number of hospitalizations in groups randomized to remote patient monitoring or usual care, with the risk of bias assessments for the outcome shown on the right for each study.

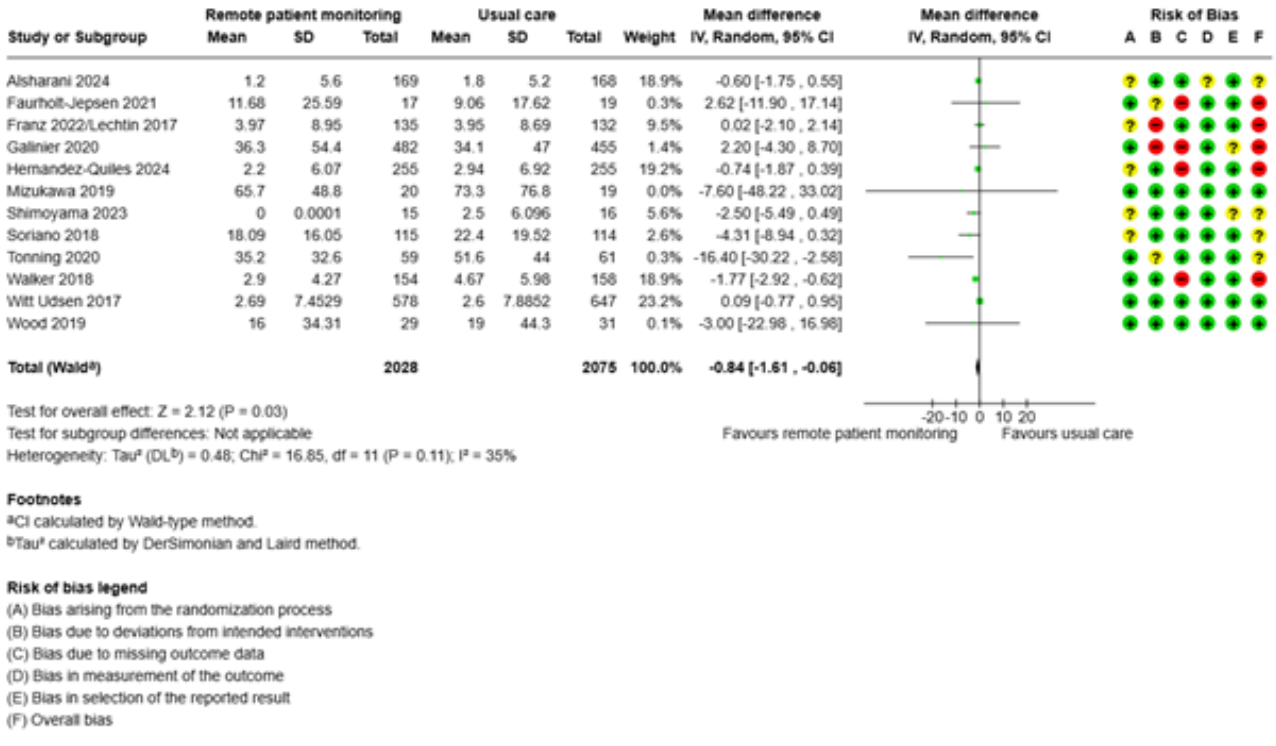


Abelsen et al [64] measured the number of unplanned hospital admissions after 6 months and 12 months and reported no difference between the groups after 6 months. However, they found an increase in the number of admissions in the remote patient monitoring group compared with the usual care group after 12 months.

Length of Hospital Stay

The mean hospital length of stay in the usual care group was 6.82 days. A meta-analysis with 12 studies showed that the length of hospital stay was slightly shorter for the remote patient monitoring group, at 0.84 days shorter (95% CI –1.61 to –0.06; low certainty; Figure 5, Multimedia Appendix 2). None of the explanatory variables had a significant effect on length of hospital stay. The Egger test was not significant (*P*=.08).

Figure 5. Meta-analysis of the length of hospital stay (in days) in groups randomized to remote patient monitoring or usual care, with the risk of bias assessments for the outcome shown on the right for each study.



Abelsen et al [63,64] measured the number of hospital bed days for unplanned admissions after 6 months and 12 months and reported no difference between the groups. Andersen et al [29] reported the median length of stay in the hospital for 0 months through 6 months as 5 days in the telegroup and 7 days in the control group. For a period of 0 months through 24 months,

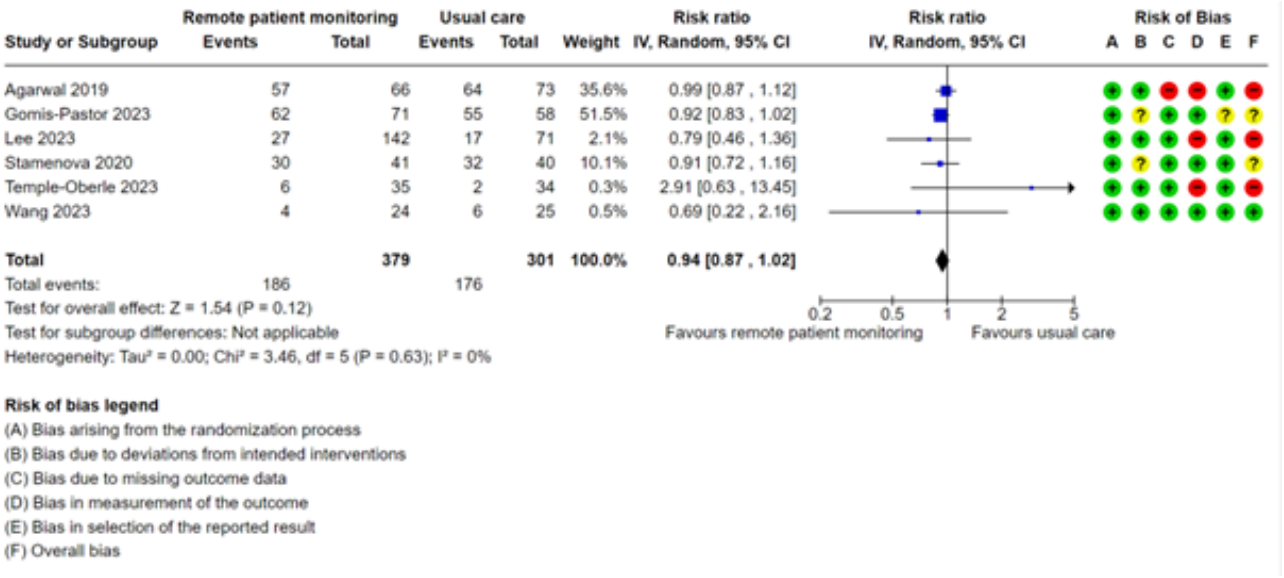
XSL•FO
RenderX

they reported 4 days in both groups. It was not possible to compute an effect size. Broadbent et al [32] reported the total number of days in hospital for respiratory problems in the intervention and control groups. It was not possible to compute an effect size. Garanin et al [37] reported that the intervention group was hospitalized for a mean number of 7.5 days and the usual care group was hospitalized for a mean of 10.2 days. It was not possible to compute an effect size.

Proportion With Outpatient Visits

A meta-analysis with 6 studies showed almost no difference in the proportion of patients with outpatient visits between the 2 groups (RR 0.94, 95% CI 0.87 to 1.02; moderate certainty; Figure 6, Multimedia Appendix 2). We did not perform mixed effects meta-regression, subgroup based on condition, or test for publication bias for this outcome.

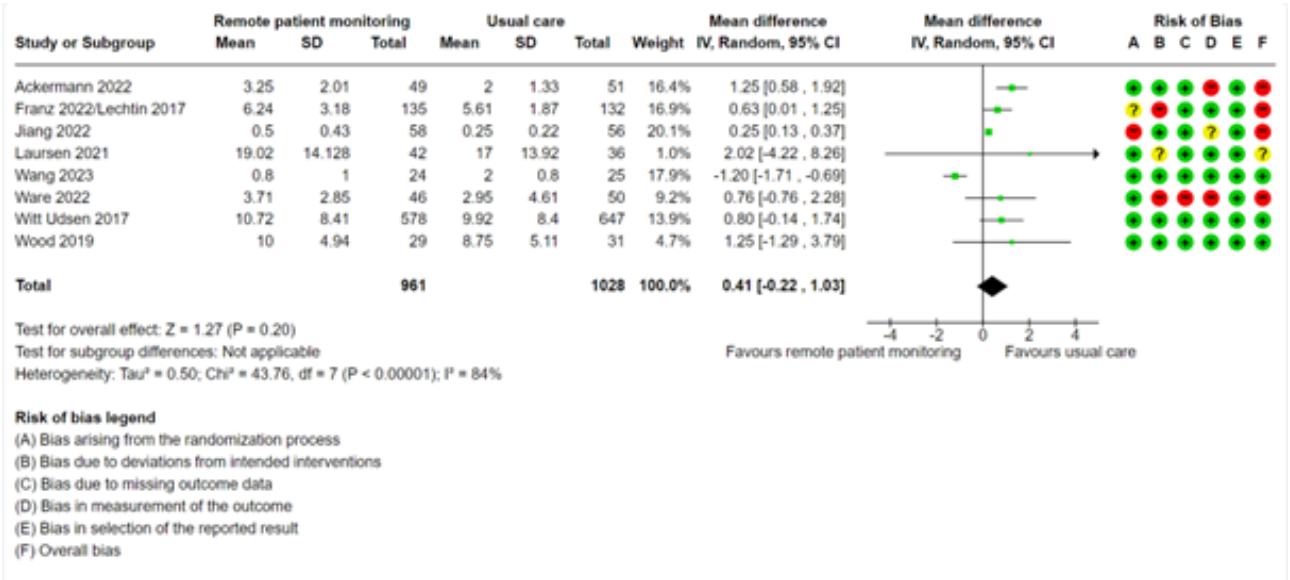
Figure 6. Meta-analysis of the proportion of outpatient visits in groups randomized to remote patient monitoring or usual care, with the risk of bias assessments for the outcome shown on the right for each study.



Mean Number of Outpatient Visits

The studies came from many different countries with diverse health care systems. We decided not to distinguish between private or public clinic visits, medical consultations, private physicians, or general practitioners. Instead, we categorized all these consultations as "outpatient visits." The usual care group had a mean number of 8.13 outpatient visits. A meta-analysis with 8 studies showed that the remote patient monitoring group had slightly more outpatient visits (0.41 more), thus favoring the usual care group, but the 95% CI crossed zero (95% CI –0.22 to 1.03; low certainty; Figure 7, Multimedia Appendix 2). We did not perform mixed effects meta-regression, subgroup based on condition, or test for publication bias for this outcome.

Figure 7. Meta-analysis of the mean number of outpatient visits in groups randomized to remote patient monitoring or usual care, with the risk of bias assessments for the outcome shown on the right for each study.



Abelsen et al [63,64] measured the number of outpatient visits and number of contacts with general practitioners after 6 months and 12 months. They found no difference in the number of outpatient visits and a tendency toward a greater increase in the

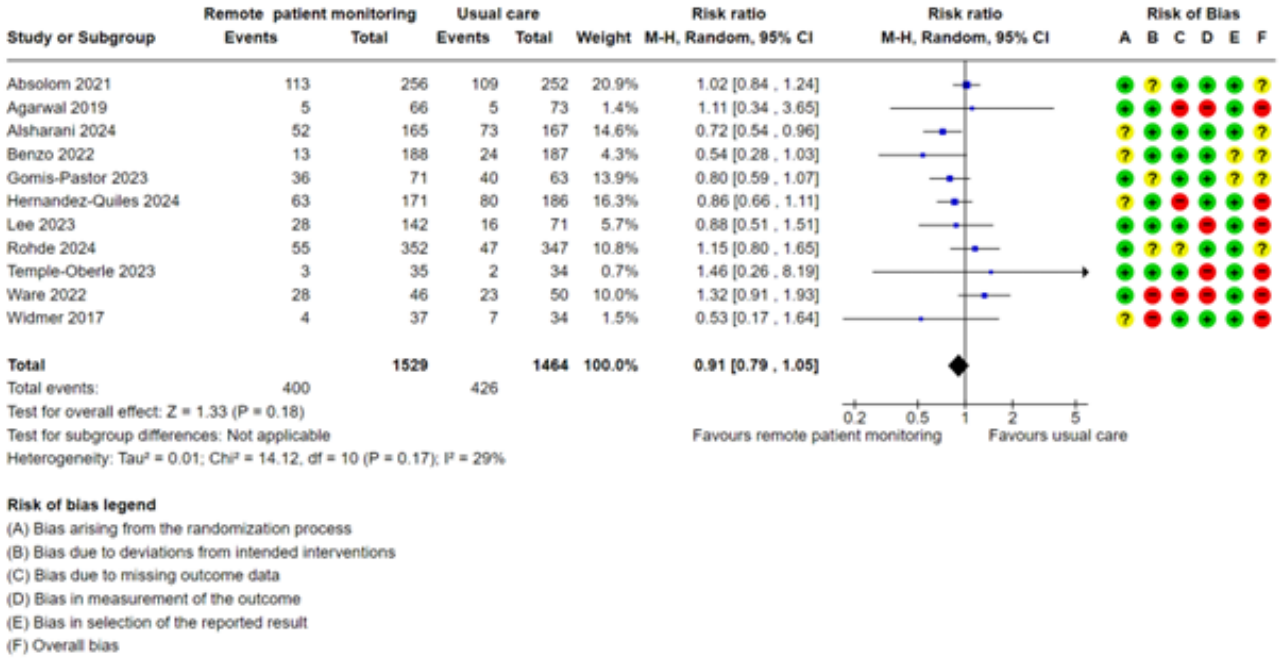
XSL•FO
RenderX

number of general practitioner consultations in the remote patient monitoring group compared with the usual care group after 6 months and 12 months.

Proportion With Emergency Visits

A meta-analysis with 11 studies showed little or no difference between the groups on proportion of emergency visits (RR 0.91, 95% CI 0.79 to 1.05; low certainty; Figure 8, Multimedia Appendix 2). There was no indication of publication bias (Egger test, *P*=.75). We did not perform mixed effects meta-regression or subgroup analysis based on condition because the heterogeneity was low.

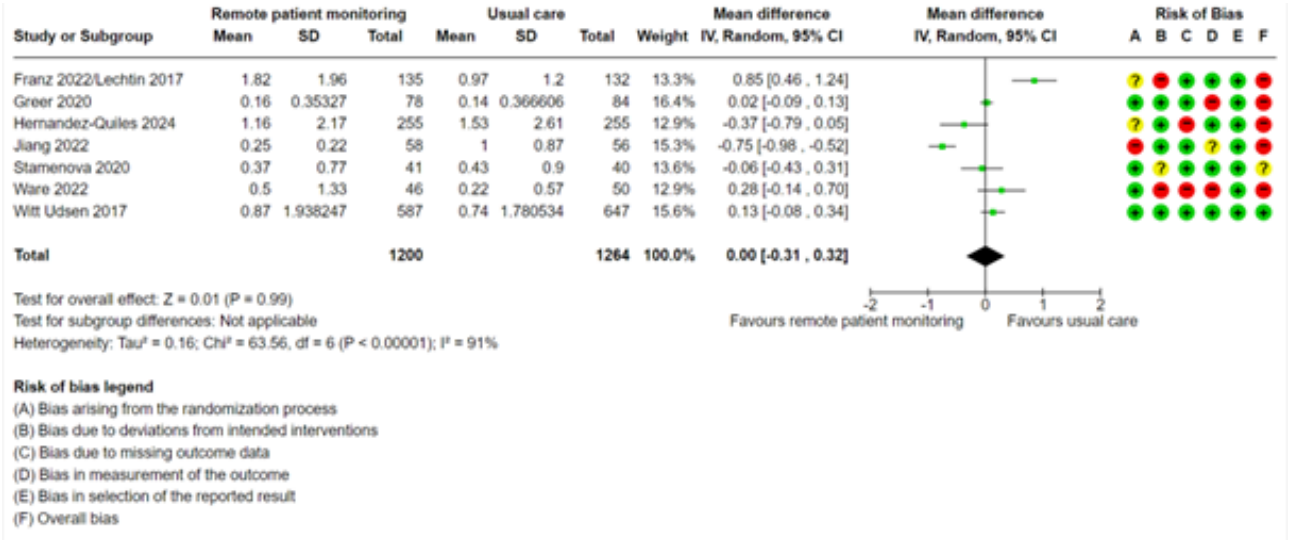
Figure 8. Meta-analysis on the proportion of emergency visits in groups randomized to remote patient monitoring or usual care, with the risk of bias assessments for the outcome shown on the right for each study.



Number of Emergency Visits

The usual care group had a mean number of 0.86 emergency visits. A meta-analysis with 7 studies showed very low certainty evidence for the number of emergency visits (mean difference 0.00, 95% CI –0.31 to 0.32; very low certainty; Figure 9, Multimedia Appendix 2). We did not perform mixed effects meta-regression, perform subgroup analysis based on condition, or test for publication bias.

Figure 9. Meta-analysis on the number of emergency visits in groups randomized to remote patient monitoring or usual care, with the risk of bias assessments for the outcome shown on the right for each study.



Abelsen et al [63,64] examined both the number of stays in municipal acute care and the number of contacts with emergency services after 6 months and 12 months. The study found no differences between the groups for either measure.

XSL•FO
RenderX

Home Care Services

Witt Udsen 2017 [60] measured the average hours spent on health care services at home over a 12- month period, finding that more hours were spent on home health care services in the remote patient monitoring group compared with the usual care group (mean difference 523.2, 95% CI –152.6 to 1199.0 hours).

Similarly, Abelsen et al [63,64] measured the average number of home health care visits after 6 months and 12 months and reported no differences between groups.

Institutional Stay

None of the included studies reported on institutional stay, only hospitalizations.

Absolute Effects

So far, we presented relative effects of remote patient monitoring compared with usual care, reporting RRs and mean differences. However, to fully grasp the impact and understand how many individuals are actually affected, it is important to also consider the absolute numbers. Table 4 shows the proportions of hospitalizations, outpatient visits, and emergency visits for the 2 groups, and Table 5 shows the numbers of hospitalizations, outpatient visits, and emergency visits and length of hospital stay in days.

Table 4. Proportions with hospitalizations, outpatient visits, and emergency visits in remote patient monitoring versus usual care.

Outcomes	Usual care	Remote patient monitoring
Proportion hospitalized	0.32	0.28
Proportion with outpatient visits	0.59	0.55
Proportion with emergency visits	0.29	0.26

Table 5. Absolute numbers of hospitalizations, days in hospital, outpatient visits, and emergency visits in remote patient monitoring versus usual care.

Outcomes	Usual care	Remote patient monitoring
Number of hospitalizations	0.42	0.29
Length of hospital stay	6.82	6.01
Number of outpatient visits	8.13	8.54
Number of emergency visits	0.86	0.86

Discussion

Principal Findings

Remote patient monitoring may slightly reduce the proportion of hospital stays, but it may make little or no difference to the number or length of hospital stays. It likely has little or no effect on the proportion of patients with outpatient visits and may result in little or no difference in the mean number of outpatient visits or the proportion with emergency visits. The evidence is uncertain regarding the effect of remote patient monitoring on the number of emergency visits.

Other systematic reviews have similar findings. Iqbal and colleagues [65], who examined clinical outcomes of digital alerting systems in remote monitoring, also found reductions in hospitalizations and length of stay for certain cohorts. However, they found no important benefit on emergency department or outpatient visits. Similarly, a review by Auener et al [66] on telemonitoring for chronic heart failure patients showed no clear differences in health care utilization between the remote monitoring group and usual care, except for an increase in nonemergency outpatient department visits for the remote monitoring group. These findings suggest that closer monitoring through remote patient monitoring may lead to more frequent health care interactions in an outpatient setting. Improved condition management through continuous monitoring and timely intervention can potentially lower the incidence of critical health episodes, reducing the demand for more costly interventions like hospitalizations. Patients who are remotely monitored may also feel a greater sense of security or may be more inclined to seek medical attention for minor symptoms or concerns that they might otherwise have ignored. This increased health care–seeking behavior could lead to higher health care utilization rates of, for example, outpatient clinical visits and potentially counteract savings from remote monitoring.

Hospitalizations are highly resource-intensive and costly [67]. Although our findings on the absolute reduction in the proportion of hospitalized patients may seem small, with a reduction from 32% to 28%, this modest reduction is expected to result in significant cost savings. Fewer hospitalizations and shorter hospital stays can lead to decreased health care costs per patient and allow for more efficient use of hospital resources, such as freeing up beds and reducing the burden on health care staff. This, in turn, enables hospitals to accommodate more patients and potentially reduces wait times for critical care.

Previous reviews have shown mixed results regarding the cost-effectiveness of remote patient monitoring [68], and remote patient monitoring systems sometimes lead to increased utilization of health care services rather than the anticipated reduction. The effectiveness of remote monitoring in reducing health care resource utilization may vary depending on the specific patient population, disease condition, and health care setting [69]. Thomas et al [14] identified several factors contributing to increased health care utilization. These included recruiting high-risk populations, failure to integrate remote

XSL•FO
RenderX

patient monitoring into workflows, and system measurement errors. Specific issues such as slow responses to alerts and low adherence to the remote follow-up by patients or clinicians were significant contributors to increased acute care use in the remote patient monitoring group. To reduce hospitalizations, it is crucial that remote patient monitoring systems accurately predict disease exacerbations by detecting changes in symptoms related to disease worsening. This can be particularly challenging in populations with unpredictable disease progression [14]. Patients who are more prone to frequent hospitalizations may experience a greater reduction in admissions with timely interventions made possible by remote monitoring. However, in practice, clinicians may hesitate to rely on remote assessments for their most vulnerable and critically ill patients [14]. The subgroup analysis in our review indicated that remote patient monitoring may reduce hospitalizations for patients with heart disease and COPD, but among the other NCDs, there were little or no differences. A systematic review of evaluations of noninvasive remote patient monitoring for chronic diseases [69] found that remote patient monitoring was highly cost-effective for hypertension. However, for COPD and heart failure, the cost-effectiveness varied depending on the severity of the disease. These findings suggest that certain patient populations may benefit more from remote monitoring than others and that variations in outcomes are influenced by factors such as diagnosis severity, health care setting, and the specific design of the remote patient monitoring intervention. Consequently, the impact on health care resource utilization may differ across contexts.

Regarding organizational aspects, remote patient monitoring can potentially streamline care, but it can also introduce complexities that may offset resource savings. For instance, remote patient monitoring generates large volumes of health data, which must be processed, cleaned, analyzed, and managed before it can be reliably used for clinical decision support. This not only requires robust data handling procedures but also the allocation of resources and adaptation of existing health care infrastructure to effectively integrate these systems into routine care [70,71]. Moreover, following up missing data due to lack of patient adherence or problems with devices presents additional, resource-intensive challenges that must be addressed to ensure safe and equitable care [72]. Alignment and adaptation of infrastructure, along with training and upskilling of staff, seem crucial to meet the demands of remote monitoring, ensuring continuity of patient care without creating gaps. Substantial investments in technology, training, and support are required to implement remote patient monitoring, and these costs may not be immediately offset by savings [69]. There is also a rapid advancement in digital solutions, and our results show that newer studies suggest better results in terms of resource utilization than older studies.

Strengths and Limitations

This systematic review's selection of eligible studies relied on a clearly defined and systematic search strategy. Although it is possible that we overlooked studies due to insufficient descriptions of interventions or populations in the title or abstract, we believe it is likely that we identified the most relevant studies. We only included studies in which the intervention was conducted in 2013 or later. This may have excluded valuable research from earlier periods. However, given the significant advancements in technology in recent years, we prioritized newer studies to provide more realistic and up-to-date results. We did not examine health care personnels' resource use related to remote follow-up, though previous studies suggest that handling incoming data from patient-reported outcome measures can be time-consuming for staff and may place a significant burden on health care providers in terms of time and resources [73].

Further, we assessed the risk of bias and certainty of the evidence transparently and systematically. These assessments allow readers to make their own judgments of the strength of the evidence. However, in many of the studies, the settings, interventions, and populations were poorly described, which may introduce limitations when synthesizing results from studies with different populations and intervention types. We attempted to account for this in our GRADE assessments, but due to poor reporting and a lack of detailed information about the settings, there may still be challenges related to heterogeneity. This could affect the overall interpretation of the findings and reduce confidence in the conclusions drawn. Moreover, regarding the outcome of outpatient visits, it is somewhat unclear what specific terms like "outpatient clinic visits," "outpatient visits," "medical consultations," and visits to "private physicians" or "general practitioners" precisely mean in the studies we reviewed. This inconsistency in terminology represents a limitation and should be considered carefully when interpreting this outcome, as it may affect the comparability of health care utilization results across the studies.

The evidence base is heavily skewed toward studies conducted in high-income countries, which limits the applicability of the findings to low- and middle-income settings. Additionally, considerable heterogeneity in patient populations and interventions across studies complicates the ability to draw broader, generalizable conclusions. The lack of economic evaluations in this review also limits our ability to assess cost-effectiveness, which is a critical factor for informing health care policy and implementation decisions.

Implications for Future Research

Future research should explore ways to streamline the processing of data from patient-reported outcomes measures through automated tools and improved system integration to optimize efficiency. Understanding the relationship between remote monitoring and resource utilization—along with identifying barriers and facilitators—is crucial to develop services that can maintain or improve current standards of care while being more resource-efficient. There is also a need for systematic reviews that examine more homogeneous patient groups and forms of digital home follow-up to better understand which groups may benefit from specific treatments. Additionally, a more comprehensive health-economic analysis is needed, considering resource use more holistically, including costs associated with technology acquisition, maintenance, follow-up on incoming data, service usage, and impacts on work absence, among other factors.

XSL•FO
RenderX

Conclusions

This systematic review showed that remote patient monitoring possibly led to lower proportions of patients being hospitalized, fewer hospitalizations, and shorter hospital length of stay compared with usual care. The proportion of patients with outpatient visits was probably similar for remote patient monitoring and usual care. Remote patient monitoring had possibly more outpatient visits compared with usual care. The proportion with emergency visits was possibly similar between the groups. We are uncertain about the effect on number of emergency visits. The results should be considered with caution as the certainty of evidence was moderate to very low. We did not find results regarding institutional stay.

Acknowledgments

We would like to thank Gyri Hval for conducting the literature searches.

The study was funded by the Research Council of Norway (328657) and the Norwegian Institute of Public Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 142 KB - mhealth_v13i1e68464_app1.pdf]

Multimedia Appendix 2
Search strategy and Grading of Recommendations Assessment, Development, and Evaluation (GRADE) assessments.
[PDF File (Adobe PDF File), 302 KB - mhealth_v13i1e68464_app2.pdf]

References

1. Vandenberghe D, Albrecht J. The financial burden of non-communicable diseases in the European Union: a systematic review. Eur J Public Health 2020 Aug 01;30(4):833-839. [doi: 10.1093/eurpub/ckz073] [Medline: 31220862]
2. Atun R, Jaffar S, Nishtar S, Knaul FM, Barreto ML, Nyirenda M, et al. Improving responsiveness of health systems to non-communicable diseases. Lancet 2013 Feb 23;381(9867):690-697. [doi: 10.1016/S0140-6736(13)60063-X] [Medline: 23410609]
3. Noncommunicable diseases. World Health Organization. 2024 Dec 23. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2025-09-19]
4. Healthier Together: EU Non-Communicable Diseases Initiative,. European Union. 2022. URL: https://health.ec.europa.eu/document/download/d843d53e-c1c1-4664-b31e-febf618d011a_en?filename=eu-ncd-initiative_publication_en_0.pdf [accessed 2025-09-20]
5. Bloom DE, Cafiero ET, Jané-Llopis E, Abrahams-Gessel S, Bloom LR, Fathima S, et al. The Global Economic Burden of Noncommunicable Diseases. Geneva. World Economic Forum. 2011 Sep. URL: https://www3.weforum.org/docs/WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf [accessed 2025-09-20]
6. Health Promotion and Disease Prevention Knowledge Gateway. European Commission. URL: https://knowledge4policy.ec.europa.eu/health-promotion-knowledge-gateway/cost-non-communicable-diseases-eu_en [accessed 2025-04-24]
7. Albahri OS, Albahri AS, Mohammed KI, Zaidan AA, Zaidan BB, Hashim M, et al. Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: taxonomy, open challenges, motivation and recommendations. J Med Syst 2018 Mar 22;42(5):80. [doi: 10.1007/s10916-018-0943-4] [Medline: 29564649]
8. Peyroteo M, Ferreira IA, Elvas LB, Ferreira JC, Lapão LV. Remote monitoring systems for patients with chronic diseases in primary health care: systematic review. JMIR Mhealth Uhealth 2021 Dec 21;9(12):e28285 [FREE Full text] [doi: 10.2196/28285] [Medline: 34932000]
9. The ongoing journey to commitmenttransformation: digital health in the WHO European Region. World Health Organization. 2023. URL: https://cdn.who.int/media/docs/librariesprovider2/data-and-evidence/english-ddh-260823_7amcet.pdf [accessed 2025-09-20]
10. Recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://iris.who.int/bitstream/handle/10665/311941/9789241550505-eng.pdf [accessed 2025-09-20]
11. Solomon DH, Rudin RS. Digital health technologies: opportunities and challenges in rheumatology. Nat Rev Rheumatol 2020 Sep;16(9):525-535. [doi: 10.1038/s41584-020-0461-x] [Medline: 32709998]
12. Le KH, La TXP, Tykkyläinen M. Service quality and accessibility of healthcare facilities: digital healthcare potential in Ho Chi Minh City. BMC Health Serv Res 2022 Nov 19;22(1):1374 [FREE Full text] [doi: 10.1186/s12913-022-08758-w] [Medline: 36403031]

XSL•FO
RenderX

13. Hewlett S, Mitchell K, Haynes J, Paine T, Korendowych E, Kirwan JR. Patient-initiated hospital follow-up for rheumatoid arthritis. Rheumatology (Oxford) 2000 Sep;39(9):990-997. [doi: 10.1093/rheumatology/39.9.990] [Medline: 10986304]
14. Thomas EE, Taylor ML, Banbury A, Snoswell CL, Haydon HM, Gallegos Rejas VM, et al. Factors influencing the effectiveness of remote patient monitoring interventions: a realist review. BMJ Open 2021 Aug 25;11(8):e051844 [FREE Full text] [doi: 10.1136/bmjopen-2021-051844] [Medline: 34433611]
15. Hestevik CN, Bjerk M, Klem H, Hval G, Kleven L. Digital hjemmeoppfølging og ressursbruk i primær- og spesialisthelsetjenestenn systematisk oversikt. Folkehelseinstituttet. 2022 Aug 29. URL: https://www.fhi.no/publ/2022/digital-hjemmeoppfolging-og-ressursbruk-i-primar--og-spesialisthelsetjenest/ [accessed 2025-09-20]
16. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane. URL: https://training.cochrane.org/handbook [accessed 2025-09-20]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
18. Thomas J, Brunton J, Graziosi S. EPPI-Reviewer 4: software for research synthesis. EPPI-Centre Software. London, England: Social Science Research Unit, UCL Institute of Education; 2010. URL: https://eppi.ioe.ac.uk/cms/Default.aspx [accessed 2025-09-20]
19. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
20. RevMan Web (Review Manager). The Cochrane Collaboration. URL: https://revman.cochrane.org [accessed 2025-09-20]
21. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 2010;36(3):1-48 [FREE Full text] [doi: 10.18637/jss.v036.i03]
22. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
23. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, GRADE Working Group. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. J Clin Epidemiol 2020 Mar;119:126-135 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.10.014] [Medline: 31711912]
24. Absolom K, Warrington L, Hudson E, Hewison J, Morris C, Holch P, et al. Phase III randomized controlled trial of eRAPID: eHealth intervention during chemotherapy. J Clin Oncol 2021 Mar 01;39(7):734-747 [FREE Full text] [doi: 10.1200/JCO.20.02015] [Medline: 33417506]
25. Achury-Saldaña DM, Gonzalez RA, Garcia A, Mariño A, Bohorquez WR. Efficacy of a telemonitoring system as a complementary strategy in the treatment of patients with heart failure: randomized clinical trial. Comput Inform Nurs 2024 Jul 01;42(7):522-529. [doi: 10.1097/CIN.0000000000001115] [Medline: 38657019]
26. Ackermann DM, Dieng M, Medcalf E, Jenkins MC, van Kemenade CH, Janda M, et al. Assessing the potential for patient-led surveillance after treatment of localized melanoma (MEL-SELF): a pilot randomized clinical trial. JAMA Dermatol 2022 Jan 01;158(1):33-42 [FREE Full text] [doi: 10.1001/jamadermatol.2021.4704] [Medline: 34817543]
27. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321 [FREE Full text] [doi: 10.2196/10321] [Medline: 30632972]
28. Alshahrani NS, Hartley A, Howard J, Hajhosseiny R, Khawaja S, Seligman H, et al. Randomized trial of remote assessment of patients after an acute coronary syndrome. J Am Coll Cardiol 2024 Jun 11;83(23):2250-2259 [FREE Full text] [doi: 10.1016/j.jacc.2024.03.398] [Medline: 38588928]
29. Andersen FD, Trolle C, Pedersen AR, Køpfli ML, Børgesen S, Jensen MS, et al. Effect of telemonitoring on readmissions for acute exacerbation of chronic obstructive pulmonary disease: a randomized clinical trial. J Telemed Telecare 2023 Jan 23;30(9):1417-1424. [doi: 10.1177/1357633x221150279]
30. Angermann CE, Sehner S, Faller H, Güder G, Morbach C, Frantz S, INH Study Group of the Competence Network Heart Failure. Longer-term effects of remote patient management following hospital discharge after acute systolic heart failure: the randomized E-INH trial. JACC Heart Fail 2023 Feb;11(2):191-206 [FREE Full text] [doi: 10.1016/j.jchf.2022.10.016] [Medline: 36718715]
31. Benzo R, Hoult J, McEvoy C, Clark M, Benzo M, Johnson M, et al. Promoting chronic obstructive pulmonary disease wellness through remote monitoring and health coaching: a clinical trial. Annals ATS 2022 Nov;19(11):1808-1817. [doi: 10.1513/annalsats.202203-214oc]
32. Broadbent E, Garrett J, Jepsen N, Li Ogilvie V, Ahn HS, Robinson H, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
33. Dorsch MP, Farris KB, Rowell BE, Hummel SL, Koelling TM. The effects of the ManageHF4Life mobile app on patients with chronic heart failure: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 07;9(12):e26185 [FREE Full text] [doi: 10.2196/26185] [Medline: 34878990]

XSL•FO
RenderX

34. Faurholt-Jepsen M, Lindbjerg Tønning M, Fros M, Martiny K, Tuxen N, Rosenberg N, et al. Reducing the rate of psychiatric re-admissions in bipolar disorder using smartphones-the RADMIS trial. Acta Psychiatr Scand 2021 May 26;143(5):453-465. [doi: 10.1111/acps.13274] [Medline: 33354769]
35. Franz N, Rapp H, Hansen RN, Gold LS, Goss CH, Lechtzin N, et al. Health care costs related to home spirometry in the eICE randomized trial. J Cyst Fibros 2022 Jan;21(1):61-69 [FREE Full text] [doi: 10.1016/j.jcf.2021.02.014] [Medline: 33715993]
36. Galinier M, Roubille F, Berdague P, Brierre G, Cantie P, Dary P, OSICAT Investigators. Telemonitoring versus standard care in heart failure: a randomised multicentre trial. Eur J Heart Fail 2020 Jun;22(6):985-994 [FREE Full text] [doi: 10.1002/ejhf.1906] [Medline: 32438483]
37. Garanin AA, Mullova IS, Shkaeva OV, Duplyakova PD, Duplyakov DV. Remote monitoring of outpatients discharged from the emergency cardiac care department. Russian Journal of Cardiology 2022 Sep 13;27(3S):5072. [doi: 10.15829/1560-4071-2022-5072]
38. Gomis-Pastor M, Mirabet Perez S, De Dios Lopez A, Brossa Loidi V, Lopez Lopez L, Pelegrin Cruz R, et al. Does an eHealth intervention reduce complications and healthcare resources? A mHeart single-center randomized-controlled trial. J Cardiovasc Dev Dis 2023 Feb 10;10(2):77 [FREE Full text] [doi: 10.3390/jcdd10020077] [Medline: 36826572]
39. Greer JA, Jacobs JM, Pensak N, Nisotel LE, Fishbein JN, MacDonald JJ, et al. Randomized trial of a smartphone mobile app to improve symptoms and adherence to oral therapy for cancer. J Natl Compr Canc Netw 2020 Feb;18(2):133-141. [doi: 10.6004/jnccn.2019.7354] [Medline: 32023526]
40. Hernandez-Quiles C, Bernabeu-Wittel M, Barón-Franco B, Palacios AA, Garcia-Serrano MR, Lopez-Jimeno W, et al. A randomized clinical trial of home telemonitoring in patients with advanced heart and lung diseases. J Telemed Telecare 2021 Dec 01;30(2):356-364. [doi: 10.1177/1357633x211059707]
41. Indraratna P, Biswas U, McVeigh J, Mamo A, Magdy J, Vickers D, et al. A smartphone-based model of care to support patients with cardiac disease transitioning from hospital to the community (TeleClinical Care): pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 28;10(2):e32554 [FREE Full text] [doi: 10.2196/32554] [Medline: 35225819]
42. Jiang W, Jin X, Du C, Gu W, Gao X, Zhou C, et al. Internet of things-based management versus standard management of home noninvasive ventilation in COPD patients with hypercapnic chronic respiratory failure: a multicentre randomized controlled non-inferiority trial. EClinicalMedicine 2024 Apr;70:102518 [FREE Full text] [doi: 10.1016/j.eclinm.2024.102518] [Medline: 38495520]
43. Jiang Y, Ramachandran HJ, Teo JYC, Leong FL, Lim ST, Nguyen HD, et al. Effectiveness of a nurse-led smartphone-based self-management programme for people with poorly controlled type 2 diabetes: a randomized controlled trial. J Adv Nurs 2022 Apr 15;78(4):1154-1165. [doi: 10.1111/jan.15178] [Medline: 35170786]
44. Krzesiński P, Jankowska EA, Siebert J, Galas A, Piotrowicz K, Stańczyk A, et al. Effects of an outpatient intervention comprising nurse-led non-invasive assessments, telemedicine support and remote cardiologists' decisions in patients with heart failure (AMULET study): a randomised controlled trial. Eur J Heart Fail 2022 Mar;24(3):565-577 [FREE Full text] [doi: 10.1002/ejhf.2358] [Medline: 34617373]
45. Laursen SL, Helweg-Jørgensen S, Langergaard A, Søndergaard J, Sørensen SS, Mathiasen K, et al. Mobile diary app versus paper-based diary cards for patients with borderline personality disorder: economic evaluation. J Med Internet Res 2021 Nov 11;23(11):e28874 [FREE Full text] [doi: 10.2196/28874] [Medline: 34762057]
46. Lee M, Kang D, Kang E, Kim S, Kim Y, Ahn JS, et al. Efficacy of the PRO-CTCAE mobile application for improving patient participation in symptom management during cancer treatment: a randomized controlled trial. Support Care Cancer 2023 May 06;31(6):321 [FREE Full text] [doi: 10.1007/s00520-023-07779-3] [Medline: 37148373]
47. Mizukawa M, Moriyama M, Yamamoto H, Rahman MM, Naka M, Kitagawa T, et al. Nurse-led collaborative management using telemonitoring improves quality of life and prevention of rehospitalization in patients with heart failure. Int Heart J 2019 Nov 30;60(6):1293-1302 [FREE Full text] [doi: 10.1536/ihj.19-313] [Medline: 31735786]
48. Piotrowicz E, Pencina MJ, Opolski G, Zareba W, Banach M, Kowalik I, et al. Effects of a 9-week hybrid comprehensive telerehabilitation program on long-term outcomes in patients with heart failure: the Telerehabilitation in Heart Failure Patients (TELEREH-HF) randomized clinical trial. JAMA Cardiol 2020 Mar 01;5(3):300-308 [FREE Full text] [doi: 10.1001/jamacardio.2019.5006] [Medline: 31734701]
49. Rohde LE, Rover MM, Hoffmann Filho CR, Rabelo-Silva ER, Silvestre OM, Martins SM, MESSAGE-HF Investigators. Multifaceted strategy based on automated text messaging after a recent heart failure admission: the MESSAGE-HF randomized clinical trial. JAMA Cardiol 2024 Feb 01;9(2):105-113. [doi: 10.1001/jamacardio.2023.4501] [Medline: 38055237]
50. Shimoyama M, Yoshida S, Takahashi C, Inoue M, Sato N, Sato F. Effectiveness of a telenursing intervention program in reducing exacerbations in patients with chronic respiratory failure receiving noninvasive positive pressure ventilation: a randomized controlled trial. PLoS One 2023 Oct 26;18(10):e0269753 [FREE Full text] [doi: 10.1371/journal.pone.0269753] [Medline: 37883428]
51. Soriano JB, García-Río F, Vázquez-Espinosa E, Conforto JI, Hernando-Sanz A, López-Yepes L, et al. A multicentre, randomized controlled trial of telehealth for the management of COPD. Respir Med 2018 Nov;144:74-81 [FREE Full text] [doi: 10.1016/j.rmed.2018.10.008] [Medline: 30366588]

XSL•FO
RenderX

52. Stamenova V, Liang K, Yang R, Engel K, van Lieshout F, Lalingo E, et al. Technology-enabled self-management of chronic obstructive pulmonary disease with or without asynchronous remote monitoring: randomized controlled trial. J Med Internet Res 2020 Jul 30;22(7):e18598 [FREE Full text] [doi: 10.2196/18598] [Medline: 32729843]
53. Temple-Oberle C, Yakaback S, Webb C, Assadzadeh GE, Nelson G. Effect of smartphone app postoperative home monitoring after oncologic surgery on quality of recovery: a randomized clinical trial. JAMA Surg 2023 Jul 01;158(7):693-699 [FREE Full text] [doi: 10.1001/jamasurg.2023.0616] [Medline: 37043216]
54. Tønning ML, Faurholt-Jepsen M, Frost M, Martiny K, Tuxen N, Rosenberg N, et al. The effect of smartphone-based monitoring and treatment on the rate and duration of psychiatric readmission in patients with unipolar depressive disorder: the RADMIS randomized controlled trial. J Affect Disord 2021 Mar 01;282:354-363. [doi: 10.1016/j.jad.2020.12.141] [Medline: 33421863]
55. Wagenaar KP, Broekhuizen BDL, Jaarsma T, Kok I, Mosterd A, Willems FF, et al. Effectiveness of the European Society of Cardiology/Heart Failure Association website 'heartfailurematters.org' and an e-health adjusted care pathway in patients with stable heart failure: results of the 'e-Vita HF' randomized controlled trial. Eur J Heart Fail 2019 Feb;21(2):238-246 [FREE Full text] [doi: 10.1002/ejhf.1354] [Medline: 30485612]
56. Walker PP, Pompilio PP, Zanaboni P, Bergmo TS, Prikk K, Malinovschi A, et al. Telemonitoring in chronic obstructive pulmonary disease (CHROMED). A randomized clinical trial. Am J Respir Crit Care Med 2018 Sep 01;198(5):620-628. [doi: 10.1164/rccm.201712-2404OC] [Medline: 29557669]
57. Wang S, Leung M, Leung S, Han J, Leung W, Hui E, et al. Safety, feasibility, and acceptability of telemedicine for hypertension in primary care: a proof-of-concept and pilot randomized controlled trial (SATE-HT). J Med Syst 2023 Mar 11;47(1):34 [FREE Full text] [doi: 10.1007/s10916-023-01933-4] [Medline: 36905441]
58. Ware P, Shah A, Ross HJ, Logan AG, Segal P, Cafazzo JA, et al. Challenges of telemonitoring programs for complex chronic conditions: randomized controlled trial with an embedded qualitative study. J Med Internet Res 2022 Jan 26;24(1):e31754 [FREE Full text] [doi: 10.2196/31754] [Medline: 35080502]
59. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: a randomized controlled trial. Am Heart J 2017 Jun;188:65-72. [doi: 10.1016/j.ahj.2017.02.016] [Medline: 28577682]
60. Witt Udsen F, Lilholt PH, Hejlesen O, Ehlers L. Cost-effectiveness of telehealthcare to patients with chronic obstructive pulmonary disease: results from the Danish 'TeleCare North' cluster-randomised trial. BMJ Open 2017 May 17;7(5):e014616 [FREE Full text] [doi: 10.1136/bmjopen-2016-014616] [Medline: 28515193]
61. Wood J, Jenkins S, Putrino D, Mulrennan S, Morey S, Cecins N, et al. A smartphone application for reporting symptoms in adults with cystic fibrosis improves the detection of exacerbations: results of a randomised controlled trial. J Cyst Fibros 2020 Mar;19(2):271-276 [FREE Full text] [doi: 10.1016/j.jcf.2019.09.002] [Medline: 31522924]
62. Yanicelli LM, Goy CB, González VDC, Palacios GN, Martínez EC, Herrera MC. Non-invasive home telemonitoring system for heart failure patients: a randomized clinical trial. J Telemed Telecare 2020 Jan 23;27(9):553-561. [doi: 10.1177/1357633x19899261]
63. Abelsen B, Godager G, Harsheim I, Iversen T, Kristiansen I, Løyland H. Evaluering av utprøving av digital hjemmeoppfølging: Delrapport II. Oslo Economics. 2021 Apr 30. URL: https://osloeconomics.no/wp-content/uploads/2021/05/Evaluering-av-utproving-av-digital-hjemmeoppfolging-delrapport-II-1.pdf [accessed 2025-09-20]
64. Abelsen B, Godager G, Harsheim I, Iversen T, Kristiansen I, Løyland H. Evaluering av utprøving av digital hjemmeoppfølging: Sluttrapport. Oslo Economics. 2022 Jan 07. URL: https://tinyurl.com/f3kjeznf [accessed 2025-09-20]
65. Iqbal FM, Lam K, Joshi M, Khan S, Ashrafian H, Darzi A. Clinical outcomes of digital sensor alerting systems in remote monitoring: a systematic review and meta-analysis. NPJ Digit Med 2021 Jan 08;4(1):7 [FREE Full text] [doi: 10.1038/s41746-020-00378-0] [Medline: 33420338]
66. Auener SL, Remers TEP, van Dulmen SA, Westert GP, Kool RB, Jeurissen PPT. The effect of noninvasive telemonitoring for chronic heart failure on health care utilization: systematic review. J Med Internet Res 2021 Sep 29;23(9):e26744 [FREE Full text] [doi: 10.2196/26744] [Medline: 34586072]
67. WHO-CHOICE estimates of cost for inpatient and outpatient health service delivery. World Health Organization. 2021. URL: https://cdn.who.int/media/docs/default-source/health-economics/who-choice-estimates-of-cost-for-inpatient-and-outpatient-health-service-delivery.pdf?sfvrsn=b814d37e_3&download=true [accessed 2025-09-19]
68. Eze ND, Mateus C, Cravo Oliveira Hashiguchi T. Telemedicine in the OECD: an umbrella review of clinical and cost-effectiveness, patient experience and implementation. PLoS One 2020;15(8):e0237585 [FREE Full text] [doi: 10.1371/journal.pone.0237585] [Medline: 32790752]
69. De Guzman KR, Snoswell CL, Taylor ML, Gray LC, Caffery LJ. Economic evaluations of remote patient monitoring for chronic disease: a systematic review. Value Health 2022 Jun;25(6):897-913 [FREE Full text] [doi: 10.1016/j.jval.2021.12.001] [Medline: 35667780]
70. Abdolkhani R, Gray K, Borda A, DeSouza R. Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open 2019 Dec;2(4):471-478 [FREE Full text] [doi: 10.1093/jamiaopen/ooz036] [Medline: 32025644]

XSL•FO
RenderX

71. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]
72. Vandenberk B, Raj SR. Remote patient monitoring: what have we learned and where are we going? Curr Cardiovasc Risk Rep 2023 Apr 22;17(6):103-115 [FREE Full text] [doi: 10.1007/s12170-023-00720-7] [Medline: 37305214]
73. Gu Y, Guan Y, Meng Z. Health providers' perceptions and experiences of using mHealth for chronic noncommunicable diseases: qualitative systematic review and meta-synthesis. J Med Internet Res 2023 Sep 12;25:e45437 [FREE Full text] [doi: 10.2196/45437] [Medline: 37698902]

Abbreviations

COPD: chronic obstructive pulmonary disease
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
GSM: Global System for Mobile Communications module
NCD: noncommunicable disease
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RoB 2: Cochrane risk-of-bias Tool
RR: risk ratio

Edited by L Buis; submitted 06.11.24; peer-reviewed by Y Li, ADS Menezes; comments to author 20.03.25; revised version received 22.05.25; accepted 16.09.25; published 01.10.25.

Please cite as:
Smedslund G, Østerås N, Hestevik CH
Effects of Remote Patient Monitoring on Health Care Utilization in Patients With Noncommunicable Diseases: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2025;13:e68464
URL: https://mhealth.jmir.org/2025/1/e68464
doi:10.2196/68464
PMID:

©Geir Smedslund, Nina Østerås, Christine Hillestad Hestevik. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 01.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

AI-Driven Real-Time Monitoring of Cardiovascular Conditions With Wearable Devices: Scoping Review

Ali Abedi[1,2], PhD; Anshul Verma[1,2], MSc; Dherya Jain[1,2], BSc; Jathushan Kaetheeswaran[1,2,3], BASc; Cynthia Chui[4], MI; Milad Lankarany[1,2,3], PhD; Shehroz S Khan[5], PhD

[1]KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

[2]Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada

[3]Krembil Brain Institute, University Health Network, Toronto, ON, Canada

[4]Library & Information Services, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

[5]College of Engineering and Technology, American University of the Middle East, Egaila, Kuwait

Corresponding Author:
Ali Abedi, PhD
KITE Research Institute
Toronto Rehabilitation Institute
University Health Network
550 University Ave
Toronto, ON, M5G 2A2
Canada
Phone: 1 4165973422 ext 7971
Email: ali.abedi@uhn.ca

Abstract

Background: Cardiovascular diseases remain the leading cause of mortality worldwide, accounting for 18 million deaths annually. Detection and prediction of cardiovascular conditions are essential for timely intervention and improved patient outcomes. Wearable devices offer a promising, noninvasive solution for continuous monitoring of cardiovascular signals, vital signs, and physical activity. However, the large data volumes generated by these devices and the rapid fluctuations in cardiovascular signals necessitate advanced artificial intelligence (AI) techniques for real-time analysis and effective clinical decision-making.

Objective: The objective of this scoping review was to identify the main challenges of AI-driven platforms for real-time cardiovascular condition monitoring with wearable devices and explore potential solutions. In addition, this review aimed to examine how AI algorithms are developed for robust monitoring and how deployment pipelines are optimized to enable real-time cardiovascular condition monitoring.

Methods: A comprehensive search was conducted in the following electronic databases: MEDLINE(R) ALL (Ovid), Embase (Ovid), Cochrane Central Register of Controlled Trials (Ovid), Web of Science Core Collection (Clarivate), IEEE Xplore, and ACM Digital Library, yielding 2385 unique records. Inclusion criteria focused on studies that used wearable devices for participant data collection and applied AI algorithms for real-time analysis to detect or predict cardiovascular events and diseases. After title and abstract screening, 153 papers remained, and following a full-text review, 19 studies met the inclusion criteria.

Results: The findings indicate that despite the promise of AI and wearable devices, research on real-time cardiovascular monitoring remains limited and lacks comprehensive validation. Most studies relied on publicly available wearable datasets rather than real-world validation with recruited participants in community settings. Studies that deployed AI algorithms in real time frequently failed to report operational characteristics and challenges. Electrocardiography-based wearable sensors were the most frequently used devices, primarily in hospital settings. A variety of AI techniques, ranging from traditional machine learning to lightweight deep learning algorithms, were deployed either on wearable devices or via cloud-based processing.

Conclusions: Robust, interdisciplinary research is needed to harness the full potential of AI-driven, real-time cardiovascular health management using wearable devices. This includes the development and validation of scalable solutions for continuous community-based deployment. Furthermore, real-world challenges such as participant compliance, hardware and connectivity constraints, and AI model optimization for real-time continuous monitoring must be carefully addressed.

(JMIR Mhealth Uhealth 2025;13:e73846) doi:10.2196/73846

XSL•FO
RenderX

KEYWORDS

cardiovascular diseases; artificial intelligence; real-time monitoring; continuous monitoring; wearable devices; machine learning; deep learning; smartwatches; electrocardiography (ECG); atrial fibrillation; mobile phone

Introduction

Background

Cardiovascular diseases (CVDs) continue to be the leading cause of mortality worldwide, causing more than 18 million deaths each year [1]. The number of deaths attributed to CVDs has increased significantly over the past decades, rising by 53.7% from 12.1 million in 1990 to 18.6 million in 2019 [2]. This trend is expected to continue, with an estimated 35.6 million cardiovascular deaths by 2050, an increase of 90% compared with 2025 [2]. This escalating trend underscores the urgent need for comprehensive prevention and management strategies to address the global impact of CVDs.

Detection and prediction of CVDs and cardiovascular events are critical for reducing the morbidity and mortality associated with these conditions [3-5]. CVDs, such as coronary artery disease, hypertensive heart disease, and cardiomyopathies, often develop silently over years, with patients exhibiting minimal or no symptoms until a significant event occurs [6]. Identifying risks at an early stage allows for timely intervention, lifestyle modifications, and treatment, which can dramatically reduce the likelihood of severe outcomes [7]. Real-time detection and prediction of cardiovascular events, such as heart attacks, cardiac arrest, and strokes, is particularly essential because it provides actionable insights during critical windows when interventions can have the greatest impact. Delays in recognizing and responding to subtle physiological changes can mean the difference between prevention and a life-threatening event [8,9].

Wearable devices equipped with sensors for continuous monitoring of vital signs such as heart rate, blood pressure, and oxygen saturation offer the potential to identify abnormal patterns indicative of cardiovascular stress or disease [10]. These devices enable remote monitoring for people in the community, allowing health care providers to track patients' health in real time and intervene promptly, even outside of clinical settings. Coupled with artificial intelligence (AI), these devices can analyze large amounts of data in real time, detecting and predicting potential cardiovascular events before they occur [9]. Advanced machine learning and deep learning techniques play a critical role in processing and interpreting the large and complex data captured by wearable devices, identifying subtle patterns and trends that may go unnoticed by human observation. Their ability to learn from historical data and continuously improve predictions ensures timely and accurate alerts, enabling health care providers to intervene before critical events unfold, thereby reducing the risk of adverse outcomes [9].

An AI-driven, wearable sensor-based platform for real-time cardiovascular monitoring can be equipped with single or multimodal sensors to collect various data, including electrocardiography (ECG), photoplethysmography (PPG), heart rate, blood pressure, and mobility indicators. These data are analyzed by AI algorithms, trained on historical datasets, and optimized for real-time inference. Depending on the system's design, these algorithms can run either directly on wearable devices or on cloud servers (requiring data transfer). Analysis results are then communicated to clinicians, participants, or caregivers, enabling timely interventions to prevent or reduce adverse cardiovascular outcomes.

By prioritizing early detection and prediction, health care systems can shift their focus from reactive to proactive care, reducing hospitalizations, improving health outcomes and patients' quality of life, and lowering health care costs [11]. Ultimately, early intervention is the foundation of effective CVD management and a pathway to saving millions of lives annually [3-5].

Related Reviews

Table 1 outlines recent reviews on the applications of wearable devices and AI algorithms for cardiovascular monitoring [10,12-22], as well as relevant reviews in the broader domain of health care monitoring [23-25]. Each review is marked to indicate which of the 4 key concepts (CVD, real-time monitoring, AI, and wearable devices) it covers. While previous reviews on cardiovascular monitoring addressed 1 or 2 concepts, wearable devices [10,12-17,20-22], AI [18,20-22,25], or real-time monitoring [10,23-25], none comprehensively covered all 3. In addition, reviews on real-time monitoring in general health care [23-25] lack a specific focus on the detection and prediction of cardiovascular conditions. Most of these reviews included studies that used precollected public datasets for developing AI algorithms, which do not deal with the real-world challenges of deploying AI for real-time monitoring.

XSL•FO
RenderX

Table 1. A comparison between this scoping review and previous related reviews, sorted by the inclusion of focus on cardiovascular diseases, real time, artificial intelligence, and wearable concepts[a].

Review	Year	CVD[b]	Real time	AI[c]	Wearable
Albahri et al [23]	2018	✗	√	✗	✗
Uddin and Koo [24]	2024	✗	√	✗	✗
Paganelli et al [25]	2022	✗	√	√	✗
Bogar et al [12]	2024	√	✗	✗	√
Ho et al [13]	2024	√	✗	✗	√
Duncker et al [14]	2021	√	✗	✗	√
Bayoumy et al [15]	2021	√	✗	✗	√
Nazarian et al [16]	2021	√	✗	✗	√
Scholte et al [17]	2024	√	✗	✗	√
Ahsan and Siddique [18]	2022	√	✗	√	✗
Safdar et al [19]	2018	√	✗	√	✗
Lee et al [20]	2022	√	✗	√	√
Huang et al [21]	2022	√	✗	√	√
Moshawrab et al [22]	2023	√	✗	√	√
Lin et al [10]	2021	√	√	✗	√
This review	2025	√	√	√	√

[a]This review is the only one that integrates all 4 concepts comprehensively.

[b]CVD: cardiovascular disease.

[c]AI: artificial intelligence.

Research Questions

AI [18,20-22,25] plays a crucial role in processing data from wearable devices [10,12-17,20-22] facilitating real-time [10,23-25] detection, prediction, and diagnosis of cardiovascular conditions. This empowers clinicians to intervene swiftly, helping to prevent or mitigate adverse outcomes, reduce hospitalizations and mortality, and improve patient quality of life. In this paper, a scoping review was conducted to systematically map the current body of research on AI-driven platforms using wearable sensors for real-time cardiovascular monitoring and to identify knowledge gaps and associated challenges. The scoping review was guided by the following research questions: (1) What are the primary challenges of AI-driven platforms for real-time cardiovascular condition monitoring with wearable devices, and how are these challenges addressed? (2) How are AI algorithms developed to support robust cardiovascular condition monitoring? (3) How are AI algorithms and deployment pipelines optimized to enable real-time cardiovascular condition monitoring?

While the first research question focuses on platform-level challenges and solutions for real-time cardiovascular condition monitoring, the second and third questions specifically examine AI algorithms, one of the core components of these platforms, in terms of ensuring robustness and enabling real-time deployment.

Methods

Study Design

This study used a scoping review methodology to address the broad scope of the research questions, the diversity of the studies and populations, and the absence of previously conducted comprehensive reviews [26,27]. The scoping review followed the framework outlined by Arksey and O'Malley [26] and was reported in compliance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [27].

Eligibility Criteria

Inclusion Criteria

Peer-reviewed journal and conference papers written in English, including quantitative, qualitative, and mixed method studies, were considered for inclusion. To be eligible, studies had to describe either the development of a new platform (research-based or commercial) or the use of an existing platform for AI-driven real-time monitoring of cardiovascular conditions through wearable devices. The platform was required to meet all four criteria: (1) It must collect continuous data from adult participants using wearable sensing devices. (2) It must be AI-driven, incorporating machine learning and deep learning algorithms to analyze wearable data. (3) The AI algorithms must perform data analysis and inference-making in real time or near real time. (4) The inferences made by AI algorithms must

XSL•FO
RenderX

involve the detection or prediction of cardiovascular conditions, diseases, or events. Studies could also look at other comorbidities, such as diabetes, in addition to CVD.

Real-time analysis involves continuous processing with minimal delay, while near–real time refers to intermittent or slightly delayed processing due to computational or design limitations. Although "detection" and "prediction" are often used interchangeably in the literature, this paper distinguishes between them: detection infers current conditions from current wearable data, whereas prediction uses current data to forecast future conditions.

Using the SPICE framework [28], the setting; population; intervention; comparison; and evaluation criteria were identified as participants' homes, nursing homes, or hospitals; adult participants; monitoring (predicting or detecting) any cardiovascular condition; real-time AI algorithms; and the accuracy and latency of real-time AI algorithms, respectively.

Exclusion Criteria

Non–peer-reviewed and non-English publications or resources were excluded. Studies were excluded if they (1) did not use a platform in which wearable sensor devices were used for data collection from adult participants, (2) did not use AI algorithms for wearable sensor data analysis, (3) performed offline data analysis, or (4) evaluated their platform only on wearable sensor datasets already collected, such as publicly available datasets.

If 1 or more of the exclusion criteria were met, studies were excluded.

Information Sources and Search Strategy

To identify relevant studies, a comprehensive literature search was designed in collaboration with an information specialist (CC) and further refined through team discussions. AA and AV initially provided CC with a list of keywords and 20 target papers that needed to be retrieved from the search. Following this, CC, AA, and AV collaboratively refined the keyword list and developed a search strategy using a combination of subject headings and text words for MEDLINE, which was subsequently translated to other databases. The search strategy consisted of 4 concepts: wearable devices, CVDs, AI algorithms, and real-time data analysis. The search was conducted in the following electronic bibliographic databases: MEDLINE(R) ALL (Ovid), Embase (Ovid), Cochrane Central Register of Controlled Trials (Ovid), Web of Science Core Collection (Clarivate), IEEE Xplore, and ACM Digital Library, covering records from inception to July 5, 2024. Search filters were used to remove studies on children and animals when possible. The results were limited to the English language. Table 2 provides a subset of unique search keywords used for database queries. The complete search strategy and associated keywords for databases are detailed in Multimedia Appendix 2. The search results were exported as RIS files, consolidated, imported into the Covidence web application for literature review, and deduplicated. The reference lists of the included studies were also examined to identify any additional relevant studies.

Table 2. A subset of the unique search keywords used to search the databases.

Concept	Sample keywords
Cardiovascular diseases	Cardiovascular Diseases, Heart Diseases, Auriculo-Ventric, Irregular Heart Rate, Small Vessel Disease, Cardiomyopath, Thrombophlebit, High Blood Pressure, Stroke, Heart Attack
Real time	Real-Time, Live, Online, Dynamic, Synchronous, Streaming, Simultaneous, Cloud, Fog, Edge, Continuous
Artificial intelligence	Artificial Intelligence, Pattern Recognition, Decision Trees, Supervised Learning, Natural Language Processing, Large Language Model, ChatGPT, Transformer Architect, Neural Network, Intelligent System, Deep Learning
Wearable devices	Wearable Devices, Remote Sensing Technology, Internet of Things, Wearable, Smart sensor, Garment, Smart eyewear, Smart jewellery, Electronic Textile, Electronic Skin, FitBit

Selection of Sources of Evidence

A group of 4 independent reviewers (AV, DJ, JK, and ZH) conducted the title and abstract screening using the Covidence web application. Each study was independently reviewed by at least 2 of these reviewers. Relevant studies identified during this process underwent full-text review and data charting, performed by at least 2 reviewers from the same group (AV, DJ, JK, and ZH). Any conflicts arising during the title and abstract screening or full-text review were resolved by AA.

Data Charting Process and Data Items

To address the research questions for this scoping review, a data-charting form was designed to extract relevant information from the reviewed studies. The form consisted of five sections: (1) study characteristics, participants, and settings, as well as cardiovascular condition under monitoring; (2) study aims, methodologies, and key findings; (3) characteristics of wearable devices; (4) characteristics of real-time AI algorithms; and (5) robustness analysis of AI algorithms.

Synthesis of Results

To address the research questions, a descriptive analysis was performed, followed by a narrative summary of relevant study characteristics presented using tables. Studies were analyzed based on the data characteristics outlined in the previous subsection. Due to the heterogeneity in populations, CVDs, outcome measurement tools, and measurement times, a meta-analysis was not conducted [29].

Results

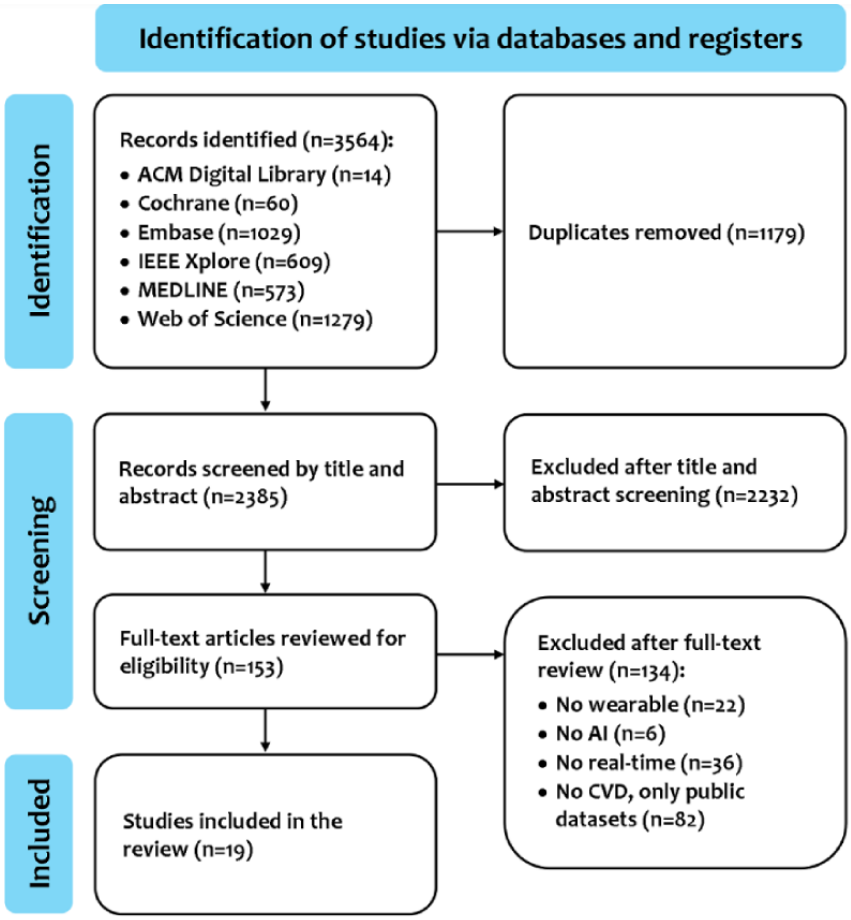
Selection of Sources of Evidence

Figure 1 illustrates the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram, which describes the study selection process. Upon removing duplicates,

XSL•FO
RenderX

a total of 2385 studies were identified through comprehensive literature searches of electronic databases. Following title and abstract screening, 2232 studies were excluded, and 153 full-text studies were retrieved for full-text review. Among these 153 studies, 134 were excluded due to the absence of 1 or more of the 4 aforementioned inclusion criteria. This resulted in the inclusion of 19 studies.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the scoping review. Of the 2385 unique titles and abstracts initially screened, 153 full-text studies were further evaluated, resulting in 19 studies being included in the scoping review. ACM: Association for Computing Machinery; AI: artificial intelligence; CVD: cardiovascular disease; IEEE: Institute of Electrical and Electronics Engineers.



Characteristics of Sources of Evidence

Multimedia Appendices 3-7 show the 5 aforementioned categories of data items for the studies included in this scoping review. A "—" in the tables indicates that the respective item was not mentioned or discussed in the paper.

Characteristics of Studies

As outlined in Multimedia Appendix 3, the included studies (n=19) were published between 2010 and 2024, with the majority, 17 (89.5%), published between 2019 and 2024 [30-46], highlighting the increasing shift toward real-time AI solutions for monitoring cardiovascular conditions. Among the included studies, 5 (26.3%) were conducted in the United States [32,40,44,45,47], 4 (21.1%) in Taiwan [33,35,36,48], 3 (15.8%) in China [31,34,46], 2 (10.5%) in India [37,43], and 1 (5.3%) each in Japan [30], Turkey [38], Pakistan [39], Bangladesh [41], and Sri Lanka [42]. Of the included studies, 14 (73.7%) were journal papers, while 5 (26.3%) were peer-reviewed conference publications.

Characteristics of Participants

One of the inclusion criteria for this scoping review was that the selected studies must have involved the collection of data from adult participants specifically for monitoring cardiovascular conditions. Consequently, a significant number of studies were excluded during the title and abstract screening, as well as the full-text review (n=82). The majority of these excluded studies focused on the development and evaluation of AI algorithms or AI-driven platforms for cardiovascular condition monitoring using publicly available datasets, such as the Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) Arrhythmia Database [49]. Such studies, while valuable for algorithmic advancements, do not provide insights into the feasibility and practical challenges associated with deploying AI models for real-time cardiovascular condition monitoring in real-world settings.

The included studies generally lacked comprehensive details regarding the demographic information of participants whose data were collected and analyzed for cardiovascular condition monitoring. Notably, as can be seen in Multimedia Appendix 3, 7 out of the 19 included studies did not provide any participant

XSL•FO
RenderX

demographic information [31,33,34,36,39,43,47]. A highly variable number of participants was reported in the 12 studies that disclosed participant counts, ranging from as few as 2 participants [38] to as many as 350 participants [46]. For studies that reported sex distribution, the balance varied. Some studies included more females [30,32,46], while others reported more males [40,44,45]. Age ranges also varied significantly across studies, with the youngest participants reporting having a mean age of 33.2 years [30] and the oldest having a mean age of 72.1 years [45].

Depending on the study design, some studies included only those participants with CVD [31,33-37,40,41,43-47], others involved a mixed population of both healthy individuals and those with CVD [32,38,42,48], and 1 study included only healthy participants [30]. With the exception of Colombage et al [42], who monitored participants with both heart failure and diabetes, all other studies focused exclusively on monitoring 1 or more cardiovascular conditions without accounting for comorbidities beyond CVDs.

Characteristics of Study Settings

Unexpectedly, with the exception of a few studies [32,43-45], which reported study durations of 1, 2, 4, and 16 weeks, the other studies involved data collection lasting only for 1 day [33,46] or a single session [35,40,48]. In addition, while 3 studies were conducted in home settings [30,32,44], all other studies took place in hospital or laboratory environments.

As outlined in Multimedia Appendix 3, in 84.2% (16/19) of the studies, the cardiovascular conditions under monitoring predominantly involved the detection of various types of cardiac arrhythmias, with a primary focus on atrial fibrillation (AF) [30-36,38,41,43-48]. Among these, 2 studies additionally measured the AF burden [32,44], and 1 study aimed to predict AF by detecting early warnings or the onset of AF [46]. One study concentrated on the measurement of left ventricular ejection fraction [40]. In addition, 2 studies focused on detecting heart disease [37] and heart failure [42], while 1 study investigated the prediction of cardiac arrest [39].

Characteristics of Wearable Devices

Depending on the data modalities required for AI-driven cardiovascular condition monitoring, primarily ECG data, a variety of wearable devices were used in the included studies. As outlined in Multimedia Appendix 5, the studies either developed custom hardware sensors or used commercially available wearable devices, such as Polar [30], Samsung [32], Xiaomi [42], Verily [44], and Apple [45] smartwatches. Typically, studies that collected data over extended periods (at least 1 week) relied on commercial smartwatches suitable for everyday use, whereas those conducted in controlled laboratory settings used custom-built hardware wearable sensors [33-40,48].

Seven out of 19 (36.8%) studies used only ECG signals [31,33,34,36,43,46,48], while another 7 (36.8%) studies incorporated ECG signals along with additional physiological or motion signals, such as PPG, acceleration, blood pressure, galvanic skin response, skin temperature, and oxygen saturation [32,39-41,44,45,47]. Other studies exclusively used non-ECG signals, such as PPG [35], heart rate, skin temperature, and acceleration [37], or electrical signals from skin-muscle interfaces [38]. The data collection frequency ranged from every hour [42] to 1200 Hz [33]. Either raw signals or features extracted from signals were input to AI models for analysis.

Of the 19 included studies, 8 (42.1%) deployed AI models directly on wearable devices for data analysis and inference [31,32,35,37,40,44,45,48]. Other studies transferred wearable sensor data to a cloud for AI-based analysis [33,34,39,41-43], while a few used a hybrid approach, deploying AI models on both wearable devices and the cloud [35,44].

Characteristics of AI Algorithms

As outlined in Multimedia Appendix 6, except for 2 studies that performed prediction tasks, one predicting the early warning or onset of AF [46] and another predicting the risk of cardiac arrest [39], all other studies focused on detection. Except for 3 studies using supervised AI models for regression tasks to measure left ventricular ejection fraction [40] and AF burden [32,44], other studies used supervised AI models for classification. These classification tasks varied, including binary classifications (eg, normal vs abnormal cardiovascular conditions) [31,32,35,37,38,43,45] and categorical classifications distinguishing between multiple cardiovascular conditions [30,33,34,36,41,42,44,46-48].

Most of the included studies used preexisting machine learning and deep learning models rather than developing novel AI architectures. The machine learning models applied included rule-based classifiers, logistic regression, decision tree, k-nearest neighbor, random forest, gradient-boosting machines, support vector machine (SVM), feed-forward neural networks, and hidden Markov model [30,32,36,37,39,42,47,48]. The deep learning models primarily comprised low-complexity 1D or 2D Convolutional Neural Networks (CNNs) [31,33-35,38,40,43,45,46], with a few studies incorporating architectures such as residual layers [44] and attention mechanisms [41]. In machine learning models, feature extraction is a separate step where meaningful characteristics are manually engineered from raw wearable data before being input into the model for analysis. In contrast, deep learning models, such as CNNs, take raw data directly as input. These models automatically learn and extract relevant features through their convolutional layers during training, eliminating the need for manual feature engineering. However, deep learning models are typically more complex, with higher computational demands and longer inference times, which is a key consideration for real-time cardiovascular monitoring.

The annotation of wearable sensor data for supervised AI model development was predominantly performed by cardiologists in most studies. However, some studies used public datasets, such as the MIT-BIH Arrhythmia Database [49], for training AI models, which were subsequently tested on data collected from their study participants [30,31,33,37,41,45,47]. Depending on the specific task, a variety of evaluation metrics were used to assess the performance of AI algorithms, with accuracy being the most commonly used.

XSL•FO
RenderX

Synthesis of Results

Overview

Research question 1: What are the primary challenges of AI-driven platforms for real-time cardiovascular condition monitoring with wearable devices, and how are these challenges addressed?

Several challenges associated with AI-driven platforms for real-time cardiovascular condition monitoring using wearable devices were identified. This section discusses the identified challenges along with strategies proposed in the reviewed studies that addressed or explored potential solutions to mitigate their impact and improve the performance of cardiovascular condition monitoring.

Latency in Decision-Making

A significant limitation of deploying AI-driven platforms for real-time cardiovascular monitoring is the delay in making inferences about cardiovascular conditions. This delay arises from computational latency, or *inference time*, which refers to the time required for machine learning and deep learning models to extract features and analyze data using pretrained algorithms. In addition, some platforms implement near real-time inference, where data analysis occurs intermittently rather than continuously. Near real-time analysis is often used to balance clinical relevance with computational efficiency, reducing battery consumption and minimizing web-based data usage when data transfer is required for cloud-based processing. This is especially critical for wearable devices, as battery depletion leads to data gaps and interrupted monitoring. For example, the platform developed by Zhu et al [32] processes data in 5-minute intervals to address these trade-offs. However, they noted that such intervals are insufficient to capture brief arrhythmic episodes that occur within shorter time frames, presenting a significant limitation.

As outlined in Multimedia Appendix 6, AI inference times varied widely in the included studies, ranging from milliseconds to several seconds, depending on the platform and implementation. The fastest reported inference time was 100 milliseconds [46], followed by times under a second [30,31,47], 2 seconds [39], less than 5 seconds [33], and under 6 seconds [48]. Inference frequencies, which often correspond to the size of data window lengths, also differed significantly in the included studies. These include processing every second [30,47], every 5 seconds [45], and every 15 seconds [46]. These variations highlight the trade-offs AI platform developers must consider, balancing real-time responsiveness, computational demands, and energy efficiency on wearable and cloud-based platforms.

Motion Artifacts

Motion artifacts are a significant challenge in wearable-based cardiovascular monitoring, particularly for ECG and PPG-based platforms, as they introduce noise that disrupts the detection of true physiological signals. Commonly caused by user movements or improper sensor-skin contact during everyday activities or dynamic conditions such as exercise, these artifacts degrade the signal-to-noise ratio and often overlap with the frequency range of cardiac signals [32]. This overlap complicates the differentiation between true heart rhythms and noise, compromising accurate rhythm detection and increasing the likelihood of missed detections. In addition, preprocessing to filter motion artifacts and noise introduces further latencies, particularly during periods of high physical activity when signal corruption is prevalent.

Hu et al [47] addressed motion artifacts by integrating accelerometer data to classify user activity and differentiate artifacts from true ECG signals. Ye et al [31] optimized motion artifact management by reconstructing multicycle ECG segments to retain correlation features. In the platform developed by Nguyen et al [35], motion artifacts were addressed through a hybrid deep neural network, where a 1D CNN filtered out low-quality PPG signals caused by motion or ambient light interference, ensuring that only reliable data were forwarded to a 2D CNN for the primary task of AF detection.

Participant Compliance

Adherence and participant compliance remain critical challenges in wearable-based cardiovascular monitoring. Participants often face difficulties consistently recharging the device on time, wearing it, and maintaining proper sensor-skin contact, which is essential for obtaining reliable PPG and ECG data. While features such as battery recharge reminders and sensor contact monitors can help notify users of low battery or improper device wear, compliance rates have shown significant variability [32]. Notifications must also be limited to prevent overwhelming participants, which can sometimes delay necessary corrections. In addition, adherence variability has been linked to demographic differences and individual behaviors, underscoring the importance of developing personalized notification systems to improve compliance [32].

Poh et al [44] highlighted participant compliance, reporting a median device wear time of 18.3 hours daily. Compliance was supported by automated notifications prompting users to perform ECG acquisitions when AF was detected through continuous PPG. These notifications encouraged consistent use and improved data quality for continuous monitoring in real-world ambulatory environments.

Internet and Bluetooth Connection

Maintaining reliable internet and Bluetooth connections is a significant challenge for wearable-based monitoring systems, especially for platforms requiring data transfer to the cloud for AI-based analysis or to a base device, such as a smartphone or computer, acting as a hub [30]. Interruptions or instability in these connections can lead to delays, data loss, or incomplete uploads, compromising the system's effectiveness and accuracy.

Battery Life

Battery life is a critical challenge in wearable-based monitoring, as the need for frequent charging directly impacts usability and participant adherence [32]. While notifications can remind users to recharge devices, particularly to ensure the collection of overnight data, maintaining consistent compliance remains difficult. Participants often forget or neglect to charge or wear

XSL•FO
RenderX

the device regularly, resulting in data gaps that undermine the system's ability to provide effective continuous monitoring.

To enhance battery life, the wearable sensor node developed by Hu et al [47] integrated energy-efficient hardware components and low-power Bluetooth communication modules, enabling 24-hour monitoring with a 500-mAh battery. Reliable data transmission was ensured through Bluetooth Low Energy to a smartphone gateway, where AI-based analyses were conducted. Ye et al [31] addressed battery limitations in their wearable cardiac arrhythmia–monitoring processor using an ultra–low-power design.

Research question 2: How are AI algorithms developed to support robust cardiovascular condition monitoring?

Robustness, in the context of AI for cardiovascular monitoring, refers to an algorithm's ability to sustain performance under variable, imperfect, or unforeseen conditions. Balendran et al [50] conducted a scoping review to explore robustness concepts in the application of machine learning models in health care. They identified 8 key robustness concepts, which are detailed in the following subsections and outlined in Multimedia Appendix 7, highlighting whether they were addressed in the reviewed studies. While all 8 concepts are considered in this section, the eighth concept is not included in Multimedia Appendix 7, as none of the included studies addressed it.

Input Perturbations and Alterations

Perturbations such as noise in data collected from wearable devices (eg, ECG and PPG) are often caused by factors such as motion artifacts from body movements. These alterations were addressed in the methodologies of the reviewed studies, including applying signal-processing techniques such as filtering, using acceleration data to compensate for motion, or analyzing ECG and PPG signals only when motion artifacts were absent. However, the extent to which the reviewed studies addressed these artifacts was limited to the methodology level, without comprehensively validating their platforms against the full range of alterations that can occur continuously in everyday life.

Missing Data

Missing input data, such as robustness to random and nonrandom data missingness, is a critical challenge in AI-driven platforms relying on wearable devices. Missing data frequently occurs due to various reasons, including participants not wearing the devices, devices being out of charge, or lack of internet or Bluetooth connectivity [51,52]. However, the reviewed studies did not address this robustness concept on their proposed platforms. To minimize missing data, participant adherence must be improved by encouraging consistent wearing, regular charging, and maintaining internet or Bluetooth connectivity for wearable devices. In addition, when missing data occur, imputation strategies [53] should be used to mitigate their negative impact on the performance of AI algorithms [54].

Label Noise

Label noise refers to the uncertainty or inaccuracies in the labels used to train AI models, such as inconsistencies arising from differing expert opinions [55]. Some of the reviewed studies used publicly available datasets to train their AI models and then applied these trained models for cardiovascular condition monitoring on their recruited participants. However, these studies failed to examine label noise caused by discrepancies between the public datasets and the participant-collected data. For example, Pramukantoro and Gofuku [30] trained their models on the MIT-BIH dataset [49] collected in a hospital setting and deployed them for monitoring participants in home environments. In addition, while some studies reported annotating their datasets with cardiologists, they did not provide detailed information about the annotation protocols [56], such as interrater agreement among annotators, or analyze how annotation variability impacted dataset quality and AI model performance.

Imbalanced Data

Imbalanced data refers to scenarios where samples from 1 or more classes are significantly underrepresented compared with others [57]. This issue is critical when the minority class represents important events, such as the occurrence of cardiac arrest versus nonoccurrence, or in regression problems, where underrepresented values, such as very low heart rates, are of interest. If left unaddressed, AI models trained on such data are prone to overfitting to the majority class, resulting in the failure to detect or predict crucial cardiac events [58]. While some studies acknowledged and addressed this issue using techniques such as the Synthetic Minority Oversampling Technique or weighted loss functions favoring minority classes [30-32,42,47], the majority of the included studies failed to discuss or mitigate this challenge, limiting the robustness of their models. Recent advancements such as contrastive learning [59] and in-context learning [60] offer promising avenues for improving model performance under imbalanced conditions by leveraging richer representations and context-aware adaptation.

In the context of imbalanced datasets, reliance on accuracy alone can be misleading, as it may reflect performance on the majority class while obscuring poor classification of the minority class. Metrics more appropriate for imbalanced scenarios include the area under the precision-recall curve, area under the receiver operating characteristic curve, F_1-score, precision, recall, specificity, true negative rate, and false-negative rate. These metrics provide a more comprehensive assessment of model performance, particularly in high-stakes applications such as cardiovascular monitoring, where failure to identify rare but critical conditions can lead to serious consequences.

Feature Selection

Using feature selection techniques on features extracted from wearable data allows models to focus on the most significant and relevant features for inference in cardiovascular conditions [61]. Selecting the most important features and removing redundant ones enhance model generalizability and reduce overfitting. In addition, reducing the number of features helps avoid the curse of dimensionality, making AI pipelines lighter and more efficient for real-time inference. Among the reviewed studies, none explicitly implemented feature selection techniques; instead, they used either raw sensor data or extracted features directly as input to AI models. One study used attention mechanisms in CNNs [41], enabling the network to focus on

XSL•FO
RenderX

specific portions of input data or features that were more effective at distinguishing between classes.

Model Specification and Learning

It pertains to robustness against variability arising from the selection, parameterization, and training of AI models [62]. A few of the included studies addressed this robustness concept by using various techniques. For instance, some studies used multiple machine learning algorithms or different versions of the same algorithm to enhance platform robustness against model type and parameter variability. Others implemented cross-validation techniques during the development and validation phases of their AI algorithms to ensure stability and reliability in their performance.

External Data and Domain Shift

This robustness concept involves evaluating a model's performance across diverse populations, tasks, and care settings, such as different environments, age groups, sexes, and ethnicities [63,64]. It is linked to bias, as models trained on homogeneous datasets may not generalize well to underrepresented groups [64]. Ensuring robustness helps mitigate bias and improve fairness and reliability in real-world applications. The distribution and diversity of demographics varied across the reviewed studies. Among those that reported demographic information, the data are often imbalanced, introducing bias into AI models that skew inferences toward majority age groups, sexes, ethnicities, or health conditions. It is crucial to include a representative proportion of demographics during the training process, ensuring alignment with the populations where the AI models will be deployed in real-world settings.

Adversarial Attacks

Adversarial attacks involve deliberate alterations to input data to manipulate the predictions of a model [65]. In wearable cardiovascular monitoring, such attacks could subtly distort ECG or PPG signals, leading to misclassification of arrhythmias or missed detection of conditions such as AF. These small changes, invisible to humans, can severely impact model performance. However, none of the reviewed studies addressed or evaluated the potential impact of adversarial attacks on their AI models.

Research question 3: How are AI algorithms and deployment pipelines optimized to enable real-time cardiovascular condition monitoring?

Methods for optimizing AI algorithms and deployment pipelines to enable real-time cardiovascular condition monitoring are examined in this section. Strategies proposed in the reviewed studies, particularly those that explored optimization techniques, are discussed in relation to their effectiveness in improving platform performance, computational efficiency, and real-world applicability.

AI Algorithm Optimization

Zhu et al [32] optimized their AI algorithm for real-time inference through a hybrid decision model that combined machine learning classifiers for clean data with statistical heuristics for noisy data. Advanced signal-preprocessing techniques, including noise filtering and motion artifact detection, enhanced robustness during real-world use. The model processed 5-minute data intervals to balance computational efficiency with accuracy, achieving high AF detection accuracy. Pramukantoro and Gofuku [30] demonstrated the use of lightweight machine learning models, enabling inference times under 1 second for real-time cardiovascular monitoring. Ye et al [31] implemented an event-driven neural network architecture, where data analysis is triggered only when predefined signal thresholds are reached. This selective activation minimizes unnecessary computation, enabling efficient real-time processing. Mary et al [43] used modular linear discriminant analysis for dimensionality reduction in their ECG classification platform. Modular linear discriminant analysis extracted the most relevant shape, texture, and statistical features from ECG signals, reducing dimensionality while retaining critical information. This process enhanced their deep neural network's efficiency and accuracy in real-time analysis.

Backend or Server-Side Processing

Hu et al [47] optimized AI deployment by running arrhythmia classification algorithms on a smartphone gateway rather than a wearable device. The smartphone processed ECG data transmitted via Bluetooth, enabling real-time inference through a layered hidden Markov model, reducing computational demands on the wearable and enhancing overall system efficiency and scalability. Lin et al [33] used a wearable ECG device with Bluetooth Low Energy to transmit real-time data to a smartphone app, which provided immediate classification of ECG signals as normal or abnormal. For detailed analysis, the data were transmitted to a cloud server via Wi-Fi or mobile internet, enabling remote, high-precision classification and storage in a cloud database. This 2-tiered system ensured seamless integration of on-device and cloud-based inference. Nguyen et al [35] used a 1D CNN deployed on the wearable device for quality assessment, which evaluated PPG signals in real time to determine whether they met the required standards for further analysis, while a 2D CNN in the cloud performed the main task of AF detection.

Hardware Efficiency Improvements

Zhu et al [32] designed their system to focus on efficient data handling and real-time inference while minimizing computational and energy demands, enabling seamless monitoring in diverse settings. Ye et al [31] reported that their system achieved a rapid 0.2-second response time and significantly reduced energy consumption to 842 nW in classification mode.

Discussion

Principal Findings

In this scoping review, 19 studies were identified that reported the development or utilization of AI-driven platforms incorporating wearable devices for real-time cardiovascular condition monitoring. These studies leveraged various wearable sensors to acquire multimodal physiological data, which were subsequently processed by AI algorithms to detect or predict cardiovascular events and conditions. Previous reviews have separately demonstrated the feasibility and success of AI-based

XSL•FO
RenderX

[18,20-22,25] wearable [10,12-17,20-22] platforms for cardiovascular monitoring, as well as real-time monitoring [23-25] in general health care applications. Building upon these findings, the current review comprehensively integrates AI, wearable technology, real-time capabilities, and cardiovascular conditions to investigate the state of research in this area. The findings align with earlier literature in supporting the promise of such integrated platforms, while also revealing critical areas that remain underexplored, particularly regarding generalizability, robustness, and real-world deployment.

The included studies were conducted across various countries and study settings, ranging from controlled laboratory environments to real-world clinical and home-based contexts. Participant information varied in completeness, with several studies lacking consistent reporting on age and sex distributions. Most studies focused on detecting AF, suggesting a limited exploration of other cardiovascular conditions. A range of wearable devices was used, including commercial systems and custom-built prototypes, capturing physiological data from modalities such as ECG, PPG, and motion sensors. The AI models used spanned from traditional machine learning techniques, such as SVMs and decision trees, to more complex architectures, such as CNNs equipped with attention mechanisms and hybrid models. While most studies reported favorable performance metrics, such as high accuracy, sensitivity, or specificity, these metrics were not always uniformly provided. Moreover, considerations of robustness were often underrepresented, with only a few studies addressing input noise, missing data, class imbalance, or domain shift through approaches such as data augmentation, weighted loss functions, or external validation. These observations underscore the variability in methodological quality across studies and highlight areas where important limitations and challenges remain to be addressed.

Limitations and Future Recommendations

Platform Validation in Real-World Settings

A wearable sensor platform designed to monitor cardiovascular conditions, particularly those critical enough to risk rehospitalization or death if unmanaged, requires thorough performance evaluation. Furthermore, its validation in real-world settings is essential before being implemented for clinical use with participants. This is especially important when the platform is intended for continuous use in community settings, where it must reliably operate over extended periods. The need for validation becomes even more critical when the platform aims to provide real-time inferences about cardiovascular conditions, as this demands high accuracy, low latency, participant compliance, and reliability under dynamic, real-world conditions. There is a significant gap in the current research literature regarding the validation of proposed platforms in real-world settings, highlighting the need for further studies to assess and improve their performance, reliability, and scalability under practical conditions.

Lack of Evaluation on Real-World Participants

Many studies were excluded during screening because they relied solely on publicly available datasets, such as subsets of the MIT-BIH dataset [49], without recruiting participants for primary data collection. As a result, these evaluations do not assess the platform's usability, comfort, or effectiveness on actual users. For instance, a deep learning model may demonstrate high accuracy for AF detection using benchmark datasets but may not account for variability in signal quality, user behavior, or device adherence observed in real-world participants. Similarly, while custom ECG hardware may perform well in technical terms, it might be unsuitable for continuous wear due to discomfort or practical limitations when tested on real individuals.

Lack of Evaluation in Real-World Environments

Most included studies conducted testing in controlled hospital or laboratory settings, with only a few evaluating their platforms in unsupervised community environments [30,32,44]. However, understanding the effectiveness and usability of AI-driven wearable platforms requires deployment in real-world settings, where participants live independently and are not overseen by researchers or clinicians. Real-world environments introduce critical factors such as motion artifacts, inconsistent device usage, battery constraints, and variable connectivity, all of which affect platform reliability. Importantly, cardiovascular events often occur outside clinical settings [66], reinforcing the need for evaluations that reflect these real-use conditions.

Lack of Platform Usability, Acceptability, and Adherence Validation and Co-Designing

The reviewed studies did not sufficiently evaluate their platforms in terms of usability, acceptability, and adherence from the perspectives of key stakeholders, including patients, caregivers, and clinicians [67,68]. Important factors such as the physical characteristics of the wearable devices (eg, shape, size, and weight), patient adherence to wearing the devices consistently, the frequency and type of alerts (such as cardiac arrest warnings), and the usability of companion smartphone or web apps require co-design with end users to ensure successful deployment and acceptance in real-world settings [69,70]. A major consideration in improving real-world usability is the frequency and reliability of platform-generated alerts, which are closely tied to the robustness of the underlying AI algorithms. High false-positive rates can lead to alert fatigue [71], causing users to lose trust and potentially ignore even critical alarms.

To mitigate these risks, future work should incorporate participatory design approaches and iterative feedback from patients, caregivers, and clinicians [69]. Continuous involvement of stakeholders in the development and postdeployment refinement of platforms can improve usability, foster trust, and increase adherence, thereby supporting successful real-world adoption.

Limitations in the Development and Validation of Real-Time and Robust AI Algorithms

Trade-Offs Between Shallow and Deep Learning Models for Real-Time Cardiovascular Monitoring

Most of the reviewed studies used traditional machine learning models, such as decision trees, SVMs, and lightweight CNNs.

XSL•FO
RenderX

These models were favored due to their lower computational demands and ability to support real-time inference, which is essential for timely cardiovascular condition monitoring. In many cases, shallow models demonstrated comparable performance metrics to deep learning architectures, raising the question of whether increased model complexity is justified. From a deployment perspective, especially on edge devices, simpler models offer advantages in terms of lower latency, reduced power consumption, and ease of integration. Nevertheless, recent advancements in AI research have introduced more sophisticated architectures, such as Transformers [72], which have demonstrated improved performance in analyzing sequential data, such as ECG and PPG signals. These models are capable of capturing long-range temporal dependencies, which can be valuable in detecting subtle patterns in cardiovascular signals. However, their high computational cost often limits their use in real-time or on-device scenarios unless optimization techniques are applied. Approaches such as model pruning, quantization, and knowledge distillation can be used to reduce model size and inference time [73], making these architectures more suitable for deployment in resource-constrained environments.

Edge Versus Cloud Deployment Models for AI-Enabled Cardiovascular Platforms

While edge deployment supports on-device processing with minimal delay, several studies used offloading techniques, sending data to smartphone gateways or cloud servers for analysis [74]. Despite the potential of cloud infrastructure for scalable, low-latency AI deployment, modern cloud-based tools specifically designed for AI, such as Google Cloud Vertex AI, AWS SageMaker, and Azure AI [75], were not used in the reviewed studies. These platforms provide capabilities for efficient training, deployment, and maintenance of AI models, including support for real-time inference and compliance with health care data standards. Their use could significantly enhance the flexibility, performance, and maintainability of AI-enabled cardiovascular monitoring systems. Future studies should consider the trade-offs between edge and cloud deployment models when designing AI-driven platforms. While edge computing ensures privacy and immediate feedback, cloud deployment enables more complex models and centralizes updates and maintenance. Both deployment strategies involve specific considerations around latency, energy consumption, data privacy, and reliability, which must be evaluated based on the target application and environment.

Robustness and Ethical Considerations in AI Algorithms for Cardiovascular Monitoring

The robustness of the included studies was examined under research question 3, based on the 8 robustness dimensions defined by Balendran et al [50]. Input perturbations, such as motion artifacts or physiological noise in ECG and PPG signals, were minimally addressed. Some studies applied basic signal filtering, but few used advanced noise mitigation strategies. Missing data, often resulting from sensor detachment or transmission issues, were rarely discussed. Models that ignore such gaps risk biased outputs. Incorporating imputation methods or continuity-aware algorithms could improve resilience. Data imbalance was common, with normal cardiovascular states vastly outnumbering abnormal ones. Many studies reported only accuracy, which is insufficient for imbalanced datasets. Metrics such as precision, recall, specificity, F_1-score, and area under the precision-recall curve provide a more comprehensive view. External validation was often absent, raising concerns about generalizability and algorithmic bias when models are applied across diverse populations or devices.

Adversarial robustness, although critical, was not explicitly explored. Wearable data can be susceptible to manipulation or distortion. To mitigate such vulnerabilities, techniques such as adversarial training or input preprocessing can be used [76]. Cloud-deployed models were typically supported by encrypted data transfer and Health Insurance Portability and Accountability Act–compliant storage. However, more advanced privacy-preserving strategies, such as federated learning and split learning with differential privacy, offer better protection by avoiding transmission of raw data [77]. These approaches enable collaborative model training across devices, sharing only model parameters or components.

Interpretability was another underaddressed dimension. Machine learning and deep learning models, including CNNs and Transformers, are often treated as black boxes. To promote clinical adoption, interpretability techniques should be applied. Tools such as attention mechanisms [72], Grad-CAM [78], and SHAP [79] can help identify critical signal segments or features contributing to a prediction. This allows clinicians not only to detect cardiovascular conditions but also to understand potential underlying causes, supporting more informed decision-making.

Together, these observations highlight the need for more robust, transparent, and ethically aware AI systems. Addressing these robustness dimensions is essential for advancing AI-driven cardiovascular monitoring platforms that are safe, reliable, and applicable in real-world health care settings.

Other Methodological and Reporting Limitations

Among all cardiovascular conditions requiring real-time monitoring, the included studies addressed only a limited subset, with most focusing on AF detection. Critical conditions, such as stroke, remain unexamined. Many studies also lacked essential methodological details, including complete demographic information [31,33-39,41,43,47], study duration [30,31,34,36-39,41,42,47], and data collection frequency [36,38,39]. These omissions limit the interpretability and reproducibility of findings.

There is also inconsistency in how AI algorithms are applied to wearable data for cardiovascular monitoring. Although continuous data collection and inference are desirable, practical constraints such as battery life, network reliability, and participant adherence often necessitate compromises. To determine appropriate monitoring frequency and system design, collaboration between platform engineers and cardiovascular clinicians is essential, as requirements vary across conditions, such as AF detection, cardiac arrest detection, and early warning applications.

XSL•FO
RenderX

Methodological Limitations and Scope of the Review

This review used a systematic and reproducible approach, guided by a well-established scoping review framework. The search strategy was developed in collaboration with a Library Sciences Expert to ensure comprehensive coverage of relevant literature from the inception of the technology to the present. However, some methodological limitations should be acknowledged. The exclusion of non–English language papers may have led to the omission of relevant studies published in other languages. Gray literature and preprints were also excluded, potentially contributing to selection bias. Although multiple databases were searched, some degree of publication bias or incomplete database coverage cannot be ruled out. In addition, the generalizability of findings may be limited due to the predominant focus on AF in the included studies and the underrepresentation of diverse cardiovascular conditions and populations.

Conclusions

AI-powered platforms integrated with wearable devices show strong potential for transforming real-time cardiovascular condition monitoring in community settings. By leveraging AI algorithms to analyze data from wearable sensors, these systems enable early detection of conditions such as AF and cardiac arrest, allowing timely intervention by clinicians or caregivers. However, most studies reviewed were limited to short-term evaluations in hospital environments, with minimal validation in unsupervised, real-world contexts. Key challenges include inconsistencies in data collection and inference frequency, lack of model robustness evaluation, and limited coverage of ethical considerations such as privacy, interpretability, and algorithmic bias. Few studies addressed robustness dimensions such as resilience to input noise, dataset imbalance, or domain shift, which are critical for safe deployment. Usability and adherence validation across diverse populations remain underexplored, as do issues related to participant compliance, battery constraints, and user-centered design. To advance this field, future research should (1) validate platforms in diverse, community-based settings with long-term use; (2) optimize and test robust AI models under real-world constraints; (3) address usability, acceptability, and co-design with stakeholders; (4) evaluate privacy-preserving techniques and interpretability tools to foster trust; and (5) develop standardized protocols for data collection and model evaluation.

Interdisciplinary collaboration will be essential to enhance the practicality, equity, and reliability of AI-enabled cardiovascular monitoring systems for real-world use.

Acknowledgments

The authors extend their sincere thanks and appreciation to the Natural Sciences and Engineering Research Council of Canada for providing funding for this research. Special gratitude is also due to Zara Hasan and Sahail Barakzi for their invaluable support during the title and abstract screening and full-text phases of this review.

Authors' Contributions

AA and SSK developed the research questions. AA designed the review framework and protocol, building on an initial version conceived by SSK and ML. CC conducted the systematic database search. AA led the review process, including title and abstract screening, full-text review, and data extraction, and also refined the data extraction tables and drafted all manuscript sections. AV contributed to developing the review protocol. AV, DJ, and JK participated in title and abstract screening, full-text review, and data extraction. In addition, they helped prepare the data extraction tables included in the manuscript. All authors reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist and explanation.
[DOCX File , 106 KB - mhealth_v13i1e73846_app1.docx]

Multimedia Appendix 2
Search strategy and associated keywords.
[DOCX File , 33 KB - mhealth_v13i1e73846_app2.docx]

Multimedia Appendix 3
Baseline characteristics of the reviewed studies, including study characteristics, participants, and settings, as well as the cardiovascular condition under monitoring.
[DOCX File , 33 KB - mhealth_v13i1e73846_app3.docx]

Multimedia Appendix 4

XSL•FO
RenderX

Aims, methodologies, and key findings of the included studies.
[DOCX File , 33 KB - mhealth_v13i1e73846_app4.docx]

Multimedia Appendix 5
Characteristics of wearable devices in the included studies.
[DOCX File , 33 KB - mhealth_v13i1e73846_app5.docx]

Multimedia Appendix 6
Characteristics of artificial intelligence algorithms in the included studies.
[DOCX File , 48 KB - mhealth_v13i1e73846_app6.docx]

Multimedia Appendix 7
Robustness analysis of the artificial intelligence algorithms in the included studies.
[DOCX File , 48 KB - mhealth_v13i1e73846_app7.docx]

References

1. Li Y, Cao G, Jing W, Liu J, Liu M. Global trends and regional differences in incidence and mortality of cardiovascular disease, 1990-2019: findings from 2019 global burden of disease study. Eur J Prev Cardiol 2023;30(3):276-286. [doi: 10.1093/eurjpc/zwac285] [Medline: 36458973]
2. Wang MS, Deng JW, Geng WY, Zheng R, Xu HL, Dong Y, et al. Temporal trend and attributable risk factors of cardiovascular disease burden for adults 55 years and older in 204 countries/territories from 1990 to 2021: an analysis for the Global Burden of Disease Study 2021. Eur J Prev Cardiol 2025;32(7):539-552. [doi: 10.1093/eurjpc/zwae384] [Medline: 39591503]
3. Buxton AE, Lee KL, Fisher JD, Josephson ME, Prystowsky EN, Hafley G. A randomized study of the prevention of sudden death in patients with coronary artery disease. Multicenter Unsustained Tachycardia Trial Investigators. N Engl J Med 1999;341(25):1882-1890. [doi: 10.1056/NEJM199912163412503] [Medline: 10601507]
4. Chugh SS. Early identification of risk factors for sudden cardiac death. Nat Rev Cardiol 2010;7(6):318-326 [FREE Full text] [doi: 10.1038/nrcardio.2010.52] [Medline: 20421887]
5. Eckersley L, Sadler L, Parry E, Finucane K, Gentles TL. Timing of diagnosis affects mortality in critical congenital heart disease. Arch Dis Child 2016;101(6):516-520. [doi: 10.1136/archdischild-2014-307691] [Medline: 26130379]
6. Gheorghiade M, Bonow RO. Chronic heart failure in the United States: a manifestation of coronary artery disease. Circulation 1998;97(3):282-289. [doi: 10.1161/01.cir.97.3.282] [Medline: 9462531]
7. Ornish D, Scherwitz LW, Billings JH, Brown SE, Gould KL, Merritt TA, et al. Intensive lifestyle changes for reversal of coronary heart disease. JAMA 1998;280(23):2001-2007. [doi: 10.1001/jama.280.23.2001] [Medline: 9863851]
8. Lown B, Wolf M. Approaches to sudden death from coronary heart disease. Circulation 1971;44(1):130-142. [doi: 10.1161/01.cir.44.1.130] [Medline: 4104697]
9. Armoundas AA, Narayan SM, Arnett DK, Spector-Bagdady K, Bennett DA, Celi LA, American Heart Association Institute for Precision Cardiovascular Medicine; Council on CardiovascularStroke Nursing; Council on Lifelong Congenital Heart DiseaseHeart Health in the Young; Council on Cardiovascular RadiologyIntervention; Council on Hypertension; Council on the Kidney in Cardiovascular Disease;Stroke Council. Use of artificial intelligence in improving outcomes in heart disease: a scientific statement from the American Heart Association. Circulation 2024;149(14):e1028-e1050 [FREE Full text] [doi: 10.1161/CIR.0000000000001201] [Medline: 38415358]
10. Lin J, Fu R, Zhong X, Yu P, Tan G, Li W, et al. Wearable sensors and devices for real-time cardiovascular disease monitoring. Cell Reports Physical Science 2021;2(8):100541. [doi: 10.1016/j.xcrp.2021.100541]
11. Waldman SA, Terzic A. Health care evolves from reactive to proactive. Clin Pharmacol Ther 2019;105(1):10-13 [FREE Full text] [doi: 10.1002/cpt.1295] [Medline: 30597564]
12. Bogár B, Pető D, Sipos D, Füredi G, Keszthelyi A, Betlehem J, et al. Detection of arrhythmias using smartwatches—a systematic literature review. Healthcare (Basel) 2024;12(9):892 [FREE Full text] [doi: 10.3390/healthcare12090892] [Medline: 38727449]
13. Ho JSY, Ho ESY, Yeo LLI, Kong WKF, Li TYW, Tan BYQ, et al. Use of wearable technology in cardiac monitoring after cryptogenic stroke or embolic stroke of undetermined source: a systematic review. Singapore Med J 2024;65(7):370-379 [FREE Full text] [doi: 10.4103/singaporemedj.SMJ-2022-143] [Medline: 38449074]
14. Duncker D, Ding WY, Etheridge S, Noseworthy PA, Veltmann C, Yao X, et al. Smart wearables for cardiac monitoring-real-world use beyond atrial fibrillation. Sensors (Basel) 2021;21(7):2539 [FREE Full text] [doi: 10.3390/s21072539] [Medline: 33916371]
15. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021;18(8):581-599 [FREE Full text] [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]

XSL•FO
RenderX

16. Nazarian S, Lam K, Darzi A, Ashrafian H. Diagnostic accuracy of smartwatches for the detection of cardiac arrhythmia: systematic review and meta-analysis. J Med Internet Res 2021;23(8):e28974 [FREE Full text] [doi: 10.2196/28974] [Medline: 34448706]
17. Scholte NTB, van Ravensberg AE, Shakoor A, Boersma E, Ronner E, de Boer RA, et al. A scoping review on advancements in noninvasive wearable technology for heart failure management. NPJ Digit Med 2024;7(1):279 [FREE Full text] [doi: 10.1038/s41746-024-01268-5] [Medline: 39396094]
18. Ahsan MM, Siddique Z. Machine learning-based heart disease diagnosis: a systematic literature review. Artif Intell Med 2022;128:102289. [doi: 10.1016/j.artmed.2022.102289] [Medline: 35534143]
19. Safdar S, Zafar S, Zafar N, Khan NF. Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. Artif Intell Rev 2017;50(4):597-623. [doi: 10.1007/s10462-017-9552-8]
20. Lee S, Chu Y, Ryu J, Park YJ, Yang S, Koh SB. Artificial intelligence for detection of cardiovascular-related diseases from wearable devices: a systematic review and meta-analysis. Yonsei Med J 2022;63(Suppl):S93-S107 [FREE Full text] [doi: 10.3349/ymj.2022.63.S93] [Medline: 35040610]
21. Huang J, Wang J, Ramsey E, Leavey G, Chico TJA, Condell J. Applying artificial intelligence to wearable sensor data to diagnose and predict cardiovascular disease: a review. Sensors (Basel) 2022;22(20):8002 [FREE Full text] [doi: 10.3390/s22208002] [Medline: 36298352]
22. Moshawrab M, Adda M, Bouzouane A, Ibrahim H, Raad A. Smart wearables for the detection of cardiovascular diseases: a systematic literature review. Sensors (Basel) 2023;23(2):828 [FREE Full text] [doi: 10.3390/s23020828] [Medline: 36679626]
23. Albahri OS, Zaidan AA, Zaidan BB, Hashim M, Albahri AS, Alsalem MA. Real-time remote health-monitoring systems in a medical centre: a review of the provision of healthcare services-based body sensor information, open challenges and methodological aspects. J Med Syst 2018;42(9):164. [doi: 10.1007/s10916-018-1006-6] [Medline: 30043085]
24. Uddin R, Koo I. Real-time remote patient monitoring: a review of biosensors integrated with multi-hop IoT systems via cloud connectivity. Applied Sciences 2024;14(5):1876. [doi: 10.3390/app14051876]
25. Paganelli AI, Mondéjar AG, da Silva AC, Silva-Calpa G, Teixeira MF, Carvalho F, et al. Real-time data analysis in health monitoring systems: a comprehensive systematic literature review. J Biomed Inform 2022;127:104009 [FREE Full text] [doi: 10.1016/j.jbi.2022.104009] [Medline: 35196579]
26. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
27. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Booth A. Clear and present questions: formulating questions for evidence based practice. Library Hi Tech 2006;24(3):355-368. [doi: 10.1108/07378830610692127]
29. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]
30. Pramukantoro ES, Gofuku A. A heartbeat classifier for continuous prediction using a wearable device. Sensors (Basel) 2022;22(14):5080 [FREE Full text] [doi: 10.3390/s22145080] [Medline: 35890769]
31. Ye Z, Lu X, Wang S, Li B. An 842 nW wearable inter-patient cardiac arrhythmia monitoring processor with a feature engine-based artificial neural network. : IEEE; 2023 Presented at: IEEE 15th International Conference on ASIC (ASICON); October 24-27, 2023; Nanjing, China p. 1-4. [doi: 10.1109/asicon58565.2023.10396056]
32. Zhu L, Nathan V, Kuang J, Kim J, Avram R, Olgin J, et al. Atrial fibrillation detection and atrial fibrillation burden estimation via wearables. IEEE J Biomed Health Inform 2022;26(5):2063-2074. [doi: 10.1109/JBHI.2021.3131984] [Medline: 34855603]
33. Lin YJ, Chuang CW, Yen CY, Huang SH, Huang PW, Chen JY. Artificial intelligence of things wearable system for cardiac disease detection. : IEEE; 2019 Presented at: IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS); March 18-20, 2019; Hsinchu, Taiwan. [doi: 10.1109/aicas.2019.8771630]
34. Fu Z, Hong S, Zhang R, Du S. Artificial-intelligence-enhanced mobile system for cardiovascular health management. Sensors (Basel) 2021;21(3):773 [FREE Full text] [doi: 10.3390/s21030773] [Medline: 33498892]
35. Nguyen DH, Chao PC, Chung C, Horng R, Choubey B. Detecting atrial fibrillation in real time based on PPG via two CNNs for quality assessment and detection. IEEE Sensors J 2022;22(24):24102-24111. [doi: 10.1109/jsen.2022.3217037]
36. Lin YJ, Chuang CW, Yen CY, Huang SH, Lee SY, Chen JY. An AIoT wearable ECG patch with decision tree for arrhythmia analysis. : IEEE; 2019 Presented at: IEEE Biomedical Circuits and Systems Conference (BioCAS); October 17-19, 2019; Nara, Japan p. 1-1. [doi: 10.1109/biocas.2019.8919141]
37. Jenifer A, Jeba G, Paulraj L, Kumar K N, Yuvaraj T, Alen G. Edge-based heart disease prediction device using internet of things. : IEEE; 2022 Presented at: International Conference on Applied Artificial Intelligence and Computing (ICAAIC); May 9-11, 2022; Salem, India p. 1500-1504. [doi: 10.1109/icaaic53929.2022.9793104]

XSL•FO
RenderX

38. Ergen O. Robust atrial fibrillation monitoring utilizing graphene aerogel-based nano-tattoo. Mater Lett 2021;291:129525. [doi: 10.1016/j.matlet.2021.129525]
39. Hannan A, Cheema SM, Pires IM. Machine learning-based smart wearable system for cardiac arrest monitoring using hybrid computing. Biomed Signal Process Control 2024;87:105519. [doi: 10.1016/j.bspc.2023.105519]
40. Howard-Quijano K, Saraf K, Borgstrom P, Baek C, Wasko M, Zhang X, et al. Evaluation of wearable acoustic sensors and machine learning algorithms for automated measurement of left ventricular ejection fraction. Am J Cardiol 2023;200:87-94. [doi: 10.1016/j.amjcard.2023.04.047] [Medline: 37307784]
41. Islam MR, Kabir MM, Mridha MF, Alfarhood S, Safran M, Che D. Deep learning-based IoT system for remote monitoring and early detection of health issues in real-time. Sensors (Basel) 2023;23(11):5204 [FREE Full text] [doi: 10.3390/s23115204] [Medline: 37299933]
42. Colombage L, Amarasiri T, Sanjeewani T, Senevirathne C, Panchendrarajan R. SmartCare: detecting heart failure and diabetes using smartwatch. 2022 Presented at: IEEE International Conference on Smart Internet of Things (SmartIoT); August 19-21, 2022; Suzhou, China p. 20-27. [doi: 10.1109/smartiot55134.2022.00013]
43. Mary GAA, Sathyasri B, Murali K, Prabhu LAJ, Bharatha Devi N. Electrocardiogram signal classification in an IoT environment using an adaptive deep neural networks. Neural Comput Appl 2023;35(21):15333-15342. [doi: 10.1007/s00521-023-08534-9]
44. Poh M, Battisti AJ, Cheng L, Lin J, Patwardhan A, Venkataraman GS, et al. Validation of a deep learning algorithm for continuous, real-time detection of atrial fibrillation using a wrist-worn device in an ambulatory environment. J Am Heart Assoc 2023;12(19):e030543 [FREE Full text] [doi: 10.1161/JAHA.123.030543] [Medline: 37750558]
45. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. Circ Arrhythm Electrophysiol 2019;12(6):e006834 [FREE Full text] [doi: 10.1161/CIRCEP.118.006834] [Medline: 31113234]
46. Gavidia M, Zhu H, Montanari AN, Fuentes J, Cheng C, Dubner S, et al. Early warning of atrial fibrillation using deep learning. Patterns (N Y) 2024;5(6):100970 [FREE Full text] [doi: 10.1016/j.patter.2024.100970] [Medline: 39005489]
47. Hu S, Wei H, Chen Y, Tan J. A real-time cardiac arrhythmia classification system with wearable sensor networks. Sensors (Basel) 2012;12(9):12844-12869 [FREE Full text] [doi: 10.3390/s120912844] [Medline: 23112746]
48. Lin CT, Chang KC, Lin CL, Chiang CC, Lu SW, Chang SS, et al. An intelligent telecardiology system using a wearable and wireless ECG to detect atrial fibrillation. IEEE Trans Inf Technol Biomed 2010;14(3):726-733. [doi: 10.1109/TITB.2010.2047401] [Medline: 20371411]
49. Moody G, Mark R. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag 2001;20(3):45-50. [doi: 10.1109/51.932724] [Medline: 11446209]
50. Balendran A, Beji C, Bouvier F, Khalifa O, Evgeniou T, Ravaud P, et al. A scoping review of robustness concepts for machine learning in healthcare. NPJ Digit Med 2025;8(1):38 [FREE Full text] [doi: 10.1038/s41746-024-01420-1] [Medline: 39824951]
51. Abedi A, Dayyani F, Chu C, Khan SS. MAISON—multimodal AI-based sensor platform for older individuals. : IEEE; 2022 Presented at: IEEE International Conference on Data Mining Workshops (ICDMW); November 28, 2022; Orlando, FL p. 238-242 URL: https://doi.org/10.1109/ICDMW58026.2022.00040 [doi: 10.1109/icdmw58026.2022.00040]
52. Baek HJ, Shin J. Effect of missing inter-beat interval data on heart rate variability analysis using wrist-worn wearables. J Med Syst 2017;41(10):147. [doi: 10.1007/s10916-017-0796-2] [Medline: 28812280]
53. Tlija A, v KW, Istrate D. Missing-data imputation using wearable sensors in heart rate variability. Bull Pol Acad Sci Tech Sci 2020;68(2):255-261. [doi: 10.24425/bpasts.2020.133118]
54. Hussein D, Belkhouja T, Bhat G, Doppa JR. Energy-efficient missing data imputation in wearable health applications: a classifier-aware statistical approach. 2024 Presented at: IJCAI '24: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence; August 3-9, 2024; Jeju, Korea p. 7296-7304. [doi: 10.24963/ijcai.2024/807]
55. Algan G, Ulusoy I. Label noise types and their effects on deep learning. ArXiv ;1:1-6 Preprint posted online on March 23 2020. [doi: 10.48550/arXiv.2003.10471]
56. Aslan S, Emine Mete S, Okur E, Oktay E, Alyuz N, Ergin Genc U, et al. Human Expert Labeling Process (HELP): towards a reliable higher-order user state labeling process and tool to assess student engagement. J Educ Technol 2017;57(1):53-59.
57. Yang Y, Zha K, Chen Y. Delving into deep imbalanced regression. PMLR 2021;139:11842-11851.
58. Bektas J, Ibrikci T, Ozcan IT. Classification of real imbalanced cardiovascular data using feature selection and sampling methods: a case study with neural networks and logistic regression. Int J Artif Intell Tools 2017;26(06):1750019. [doi: 10.1142/s0218213017500191]
59. Keramati M, Meng L, Evans RD. ConR: contrastive regularizer for deep imbalanced regression. ArXiv ;1:1-24 Preprint posted online on September 13, 2023.
60. Nejjar I, Ahmed F, Fink O. Im-context: in-context learning for imbalanced regression tasks. ArXiv ;1:1-19 Preprint posted online on May 28, 2024.
61. Ghosh P, Azam S, Jonkman M, Karim A, Shamrat FMJM, Ignatious E, et al. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access 2021;9:19304-19326. [doi: 10.1109/access.2021.3053759]

XSL•FO
RenderX

62. Aliferis C, Simon G. Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. In: Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls. Cham: Springer; 2024:477-524.
63. Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine 2021;67:103358 [FREE Full text] [doi: 10.1016/j.ebiom.2021.103358] [Medline: 33962897]
64. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. Patterns (N Y) 2021;2(10):100347 [FREE Full text] [doi: 10.1016/j.patter.2021.100347] [Medline: 34693373]
65. Newaz AI, Haque NI, Sikder AK, Rahman MA, Uluagac AS. Adversarial attacks to machine learning-based smart healthcare systems. 2020 Presented at: GLOBECOM 2020-2020 IEEE Global Communications Conference; December 07-11, 2020; Taipei, Taiwan p. 1-6. [doi: 10.1109/globecom42002.2020.9322472]
66. Brooks SC, Clegg GR, Bray J, Deakin CD, Perkins GD, Ringh M, International Liaison Committee on Resuscitation. Optimizing outcomes after out-of-hospital cardiac arrest with innovative approaches to public-access defibrillation: a scientific statement from the international liaison committee on resuscitation. Circulation 2022;145(13):e776-e801 [FREE Full text] [doi: 10.1161/CIR.0000000000001013] [Medline: 35164535]
67. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
68. Claes J, Cornelissen V, McDermott C, Moyna N, Pattyn N, Cornelis N, et al. Feasibility, acceptability, and clinical effectiveness of a technology-enabled cardiac rehabilitation platform (physical activity toward health-I): randomized controlled trial. J Med Internet Res 2020;22(2):e14221 [FREE Full text] [doi: 10.2196/14221] [Medline: 32014842]
69. Kilfoy A, Hsu TC, Stockton-Powdrell C, Whelan P, Chu CH, Jibb L. An umbrella review on how digital health intervention co-design is conducted and described. NPJ Digit Med 2024;7(1):374 [FREE Full text] [doi: 10.1038/s41746-024-01385-1] [Medline: 39715947]
70. Perego P, Scagnoli M. Co-design the acceptability of wearables in the healthcare field. In: EAI International Conference on IoT Technologies for HealthCare. Cham: Springer; 2021:21-32.
71. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R, with the HITEC Investigators. Correction to: effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak 2019;19(1):227 [FREE Full text] [doi: 10.1186/s12911-019-0971-0] [Medline: 31739801]
72. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA p. 6000-6010.
73. Chitty-Venkata KT, Mittal S, Emani M, Vishwanath V, Somani AK. A survey of techniques for optimizing transformer inference. J Syst Arch 2023;144:102990. [doi: 10.1016/j.sysarc.2023.102990]
74. Rodrigues TK, Suto K, Nishiyama H, Liu J, Kato N. Machine learning meets computation and communication control in evolving edge and cloud: challenges and future perspective. IEEE Commun Surv Tutorials 2020;22(1):38-67. [doi: 10.1109/comst.2019.2943405]
75. Thallam NST. Comparative analysis of public cloud providers for big data analytics: AWS, azure, and google cloud. Int J AI Bigdata Comput Manag Stud 2023;4:18-29. [doi: 10.63282/3050-9416.ijaibdcms-v4i3p103]
76. Bai T, Luo J, Zhao J, Wen B, Wang Q. Recent advances in adversarial training for adversarial robustness. 2021 Presented at: Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21); August 21, 2021; Montreal, Canada p. 4312-4321. [doi: 10.24963/ijcai.2021/591]
77. Abedi A, Khan SS. FedSL: federated split learning on distributed sequential data in recurrent neural networks. Multimedia Tools Appl 2023;83(10):28891-28911. [doi: 10.1007/s11042-023-15184-5]
78. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 Presented at: IEEE International Conference on Computer Vision (ICCV); October 22-29, 2017; Venice, Italy. [doi: 10.1109/iccv.2017.74]
79. Lundberg SM, Lee S. A unified approach to interpreting model predictions. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA p. 4768-4777. [doi: https://dl.acm.org/doi/10.5555/3295222.3295230]

Abbreviations

AF: atrial fibrillation
AI: artificial intelligence
CNN: Convolutional Neural Network
CVD: cardiovascular disease
ECG: electrocardiography
MIT-BIH: Massachusetts Institute of Technology-Beth Israel Hospital
PPG: photoplethysmography

XSL•FO
RenderX

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SVM: support vector machine

Edited by M Dorsch; submitted 18.03.25; peer-reviewed by NH Hoang, AM Ibrahim; comments to author 25.07.25; revised version received 13.08.25; accepted 12.09.25; published 11.11.25.

Please cite as:
Abedi A, Verma A, Jain D, Kaetheeswaran J, Chui C, Lankarany M, Khan SS
AI-Driven Real-Time Monitoring of Cardiovascular Conditions With Wearable Devices: Scoping Review
JMIR Mhealth Uhealth 2025;13:e73846
URL: https://mhealth.jmir.org/2025/1/e73846
doi:10.2196/73846
PMID:

©Ali Abedi, Anshul Verma, Dherya Jain, Jathushan Kaetheeswaran, Cynthia Chui, Milad Lankarany, Shehroz S Khan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of a Smartphone-Based Positive Reflection Diary on Work Engagement Among Japanese Workers: Randomized Controlled Trial

Masahito Tokita[1], MSc; Shuichiro Kobayashi[2], BA; Daisuke Miyanaka[3,4], MPH; Nobuyuki Takate[5], MA; Hiroyuki Nakano[5], BA; Kazuki Takeuchi[6], MA; Akihito Shimazu[4], PhD

[1]Keio University, Keio Research Institute at SFC, Fujisawa, Japan

[2]Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan

[3]Better Options, Inc, Tokyo, Japan

[4]Faculty of Policy Management, Keio University, 5322 Endo, Fujisawa, Japan

[5]Japan Productivity Center, Tokyo, Japan

[6]Labor Trust LCC, Tokyo, Japan

Corresponding Author:
Akihito Shimazu, PhD
Faculty of Policy Management, Keio University, 5322 Endo, Fujisawa, Japan

Abstract

Background: Work engagement is an important determinant of workers' well-being. According to the job demands–resources model, personal resources are one of the key antecedents of work engagement. Enhancing personal resources leads to improved work engagement. Furthermore, reflecting positively on one's achievements at work may enhance personal resources. Hence, there is a need for a simple, self-guided tool such as a smartphone app that can be used by employees to record and reflect on their work accomplishments.

Objective: This study aimed to evaluate the effect of a smartphone-based diary (ie, Work Engagement Diary) that promotes positive reflection in daily working life on work engagement among Japanese workers in a randomized controlled trial.

Methods: Six hundred Japanese workers who met the inclusion criteria were randomly allocated to either the intervention group or the wait-list control group (300 participants each) via a web survey company. Participants in the intervention group kept the Work Engagement Diary for 2 weeks, whereby they were required to set a weekly goal at the beginning of the week and fill in their work achievements at the end of the day. Those in the waitlist control group did not receive any interventions until they completed a follow-up survey. Work engagement was assessed at preintervention (T1), postintervention (T2), and 3 weeks after intervention (T3) in both groups.

Results: A mixed model for repeated measures conditional growth model analysis using intention-to-treat revealed a significant improvement in work engagement for the intervention group compared to the wait-list control group throughout the study period (P=.04). Effect sizes were small in work engagement for T1 versus T2 (Cohen d=0.11 [95% CI −0.06 to 0.28]) and for T1 versus T3 (Cohen d=0.12 [95% CI −0.06 to 0.28]).

Conclusions: This randomized controlled trial demonstrated that our newly developed smartphone-based positive reflection diary at work effectively improved work engagement among Japanese workers. Future research needs to clarify longer-term intervention effects and detailed mechanisms of the intervention effects.

Trial Registration: University Hospital Medical Information Network Clinical Trials Registry UMIN000038430; https://center6.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000043763

(JMIR Mhealth Uhealth 2025;13:e55664) doi:10.2196/55664

KEYWORDS

work engagement; positive reflection; self-monitoring; smartphone-based diary; randomized controlled trial

XSL•FO
RenderX

Introduction

Work Engagement

Work engagement is beneficial for employees and organizations alike [1,2] and is associated with work- and health-related positive outcomes. Work engagement refers to employees' strength and optimal functioning and positive work-related states characterized by vigor, dedication, and absorption in their work [3]. Vigor is defined as high levels of energy and mental resilience in one's work. Dedication is a strong involvement in one's work combined with a sense of significance, enthusiasm, and challenge. Absorption is defined as having higher concentration and being happily engrossed in one's work. It is viewed as a psychological state in which time seems to pass quickly and it is difficult to leave work [1]. A recent meta-analysis [4] reported that work engagement was positively related to job performance, job satisfaction, and organizational commitment and negatively related to turnover intention. Furthermore, work engagement as an antecedent of health- and work-related outcomes is positively associated with variables such as job performance and job satisfaction and negatively associated with variables such as work withdrawal behavior and physical and mental ill health [5,6].

Job Demands–Resources Model

According to the job demands–resources (JD-R) model [1,7], job resources foster work engagement in a positive motivational process wherein an increase in job resources leads to an increase in work engagement and positive health and organizational outcomes. On the contrary, job demands cause exhaustion in an energy-sapping process [8,9] wherein an increase in job demands (such as stressors) leads to an increase in burnout and negative health and organizational outcomes. Personal resources (such as positive emotions, vitality, and mental acuity) as antecedents have been included in the extension of the JD-R model and may be directly or indirectly related to work engagement [10-12]. Positive work reflection is considered a personal resource in the JD-R model and plays a key role in enhancing work engagement [4]. Research suggests that when employees reflect on their work achievements and growth, their personal resources increase, leading to higher work engagement [13]. Furthermore, positive reflection helps sustain work engagement—even under high job demand—highlighting its importance in promoting work engagement [14]. Therefore, interventions that focus on increasing personal resources, such as focusing on what workers have achieved rather than what they have not achieved, may improve work engagement.

Positive Reflection

Increasing awareness of one's own achievements at work leads to positive health outcomes among workers [12]. Reflecting positively on one's work, which focuses on achievement at work, leads to a reevaluation of potentially stressful work situations [15]. Positive work reflection refers to thinking about the positive aspects of one's job and includes thoughts about pleasurable events such as successful task performance and supportive work relationships [16]. A recent meta-analysis [17] reported that off-job positive work-related thoughts had a positive relationship with work engagement (ρ=0.493, 95% CI 0.407-0.579). Thus, positive reflection about one's achievement at work during off-job time can lead to improved work engagement. Furthermore, self-monitoring may be helpful for reflecting positively on one's work. Keeping a diary of positive work achievements, including increasing access to positive emotions, and positive work-related thoughts can be a strategy for self-monitoring. According to social cognitive theory, self-monitoring influences people's motivation and behavior by increasing their interest in their actions and their occurrence and effects [18,19].

Smartphone-Based Apps

There are two types of smartphone apps: those that run on a web browser (web apps) and those that run as native apps. In general, native apps are considered to have an advantage over web apps in terms of CPU and memory usage [20]. However, few health-related native apps are supported by evidence [21]. Paganin and Simbula [22] reviewed 31 studies that examined the impact of workplace smartphone-based apps on improving employees' physical health, targeted symptoms (eg, depressive symptoms), and well-being. Of the eight interventions specifically designed to improve worker well-being, only four—namely, Headspace, Hello Mindcare, Viary, and Smartphone Resiliency Training—were strictly native smartphone apps. The other four were intervention programs accessible via web browsers compatible with smartphones. These findings suggest that native smartphone apps explicitly designed to improve worker well-being are limited.

In a meta-analysis of intervention studies designed to improve work engagement among employees published between 2009 and 2017 [23], 15 online (web-based or web apps) intervention studies were included, yielding an effect size of 0.14 (P<.05). Furthermore, while app stores offer numerous health-related apps claiming to improve well-being, only a small proportion of them have undergone rigorous scientific evaluation. According to Lau et al [24], of the 1009 psychosocial wellness and stress management smartphone apps available for self-help, only 21 (2.08%) were supported by peer-reviewed evidence regarding feasibility or effectiveness. Thus, the intervention proposed in the present study represents a significant advancement as it is both a native smartphone app and has undergone scientific validation, which distinguishes it as a robust tool for improving worker well-being.

According to a meta-analysis of work engagement interventions [12], most personal resource-building interventions were delivered in group or face-to-face sessions (eg, [25-27]), which had a significant effect compared to online sessions [22] and the advantage of direct instruction and interaction with others. However, they also have the disadvantage of requiring coordination of time and place and trained staff. One way to overcome these disadvantages is to use smartphone apps. Smartphone apps offer several advantages, such as being easy to start and complete in a short time, being intuitive with minimal explanation, and being ubiquitous without worrying about time and place, offering just‐in‐time interventions [28]. Bartels et al [29] found that mobile technologies are suitable for flexible self-monitoring in everyday life. However, while there are numerous smartphone apps designed to enhance

XSL•FO
RenderX

occupational health, safety, and well-being in workplaces, only a limited number of them have been scientifically evaluated [30]. Therefore, it is worthwhile to conduct an evaluation of the effects of a newly developed smartphone-based app.

Aim

The present study aimed to evaluate the effect of a newly developed smartphone-based positive reflection diary, the Work Engagement Diary (WEDiary), to improve work engagement as a primary outcome among Japanese workers using a randomized controlled design.

Methods

Recruitment

Participants were recruited from those registered with a web survey company (Macromill Inc., Tokyo, Japan). Target participants were workers who met the following four inclusion criteria: (1) aged from 20 to 59 years, (2) had two consecutive days off in a week, (3) were not going to quit or change jobs within the next 3 months, and (4) used iOS 10.0 or later or Android 5.1 or later smartphones. First, candidates who were registered with the web survey company received an email invitation from the survey company that included study information and an informed consent form. Respondents read the research explanation, and those who agreed to participate in the study clicked on the "Agree" button on the webpage to indicate their consent. Subsequently, candidates who met the inclusion criteria and agreed to participate in the study were asked to complete a web-based baseline survey. This survey was continued until the number of participants reached 600. The baseline survey began on October 31, 2019, and ended on November 5, 2019. After completing the baseline survey, participants received an email informing them of their allocation group: the intervention group or the wait-list control group. At the end of the study period, participants received redeemable points from the web survey company that could be used for shopping as a reward for participating in the study.

Sample Size

A sample size of 394 participants per group was calculated, entering α error probability of .05, power (1–β) of 0.80, and small effect size of 0.20 [12,31] using G*Power 3.1 [32]. However, funding constraints led to a sample size of 300 per group.

Trial Design and Randomization

We conducted a two-group parallel-group, nonblinded, randomized controlled trial (RCT). The study protocol was registered at the University Hospital Medical Information Network Clinical Trials Registry (UMIN000038430). After the baseline survey, participants were randomly allocated to either the intervention group or the wait-list control group at a 1:1 ratio at the individual level by the survey company. The survey company was asked to generate the allocation table and register and assign participants. The researchers were not involved in any of the following activities undertaken by the survey company: random allocation sequence generation and implementation and allocation concealment. The log data of the WEDiary were managed by NT, which were subsequently passed on to MT without being processed. In addition, NT did not have access to the questionnaire data at any of the three time points. Therefore, they were unable to link the log data to the questionnaire data at any of these time points. The study protocol was written according to the Consolidated Standards of Reporting Trials checklist [33].

Development of Smartphone-Based App

Overview

We developed a smartphone-based positive reflection diary app (the WEDiary) for both Android and iOS. Figure 1 shows screenshots of the WEDiary. The intervention period was 2 weeks from November 11 to 24, 2019. We determined the intervention and follow-up periods based on a meta-analysis conducted by Vîrgă et al [23] who found a larger effect size for a 2-week intervention (d=0.55, P<.05) compared to those of other durations. The meta-analysis also revealed that follow-ups within 1 month postintervention showed a larger effect size (d=0.34, P<.05) than those conducted at longer intervals. The WEDiary is designed to incorporate a post-work prompt, "Please write about what you were able to do at work today," aimed at fostering positive reflection on daily work activities. Through consistent engagement with this reflective practice, participants are expected to cultivate and accumulate positive emotions over time. According to the Broaden-and-Build theory, positive emotions expand individuals' attention and cognitive flexibility, thereby facilitating the generation of diverse ideas and the development of critical personal resources [34]. However, the transient experience of isolated positive emotions alone is insufficient for resource building; instead, sustained accumulation of such emotions is essential [35,36].

XSL•FO
RenderX

Figure 1. Screenshots of the Work Engagement Diary (WEDiary).

A. Setting a goal, daily achievement, and answering four questions on daily work engagement.

B. Clicking the "Graph" button (dotted square) opens a graph of the daily values.

C. Clicking the "List" button (dotted square) opens the "Daily achievement" screen once a day to upload the data.

On the morning of the first day, participants in the intervention group were required to set a goal for that week by filling in the relevant box in the WEDiary (Figure 1A). At the end of the working day, they filled in "Daily achievement" (serving as a daily reflective diary) and rated the following four questions on daily work engagement, using a 7-point scale ranging from 1 (not at all) to 7 (very well): (1) "I felt energetic at work today (for vigor)," (2) "I felt like going to work (for vigor)," (3) "I was inspired by my work today (for dedication)," and (4) "I am proud of the things I did at work today (for dedication)" (Figure 1). These four questions were adopted to measure daily work engagement based on previous diary studies [10,36,37]. Vîrgă et al.'s [23] meta-analysis of the effects of controlled interventions on work engagement found that the effect size for absorption was small (d=0.21, P<.01) as opposed to those for vigor (d=0.36, P<.01) and dedication (d=0.35, P<.01). Thus, we adopted two items each for "vigor" and "dedication" but none for "absorption." A line graph of the daily values was available when users clicked the button with the graph icon (Figure 1B). The line graph helped participants enhance positive reflections from their visual perspective. On the last day of the 5 working days, they filled in their weekly reflections and goals for the next week in the relevant box. Finally, they pressed the "List" button on the home screen to open the "Daily achievement" screen once a day (Figure 1C) to check their use of the WEDiary and to upload the data to the server. Empirical evidence supporting this notion includes a study employing the experience sampling method, which demonstrated that positive emotional experiences are associated with enhanced hope and increased work engagement [36]. The WEDiary allows users to record, accumulate, visualize, and reinforce these positive emotional experiences. Complementary findings from a meta-analysis [17] highlight the positive association between post-work positive thoughts and work engagement (ρ=0.493, 95% CI 0.407-0.579) as well as between positive reflection and engagement (ρ=0.482, 95% CI 0.390-0.573). Additionally, positive rumination about work was found to significantly contribute to the accumulation of positive emotions (ρ=0.610, 95% CI 0.51-0.709). The WEDiary's ability to graphically track work engagement is aimed to further reinforce the accumulation of positive emotional experiences. However, the act of responding to the prompt "What did you do well at work today?" could potentially introduce stress for some users. To tackle this, the WEDiary emphasizes cultivating a habitual focus on positive recollection rather than limiting reflection to specific daily achievements. The tool encourages users to acknowledge progress even if predetermined goals are not fully attained. Additionally, guidance provided within the WEDiary directs users to concentrate on the positive aspects of their experiences, with the option to discontinue its use at any time. Its comprehensive design aims to support the sustainable enhancement of positive emotional states and work engagement, contributing to long-term personal resource development.

Intervention Group

A few days before starting the program, participants in the intervention group received an email including the URL/QR code of the WEDiary download website/app store and were required to install the app on their smartphones. Furthermore, on the morning of the first day of the intervention period, they received an email to start the program with user IDs and authentication codes for activating the WEDiary. User

XSL•FO
RenderX

registration consisted of three steps. First, participants entered their user IDs, gender, year of birth, and authentication code. Subsequently, they were asked to read and agree with the privacy policy. Finally, they were asked to enter their starting and ending business days for a week. Participants received emails requesting them to download, install, and activate the WEDiary if activation of the WEDiary was not confirmed. Participants in the intervention group received email reminders to use the app every day during the intervention period.

Wait-List Control Group

Participants in the wait-list control group were requested to be as usual until the completion of the survey 3 weeks after the intervention. They could use the WEDiary after the 3-week follow-up survey period had concluded, that is, on December 23, 2019, or later.

Measures

Data Collection

The intervention period was from November 11 to 24, 2019. All participants in both groups were asked to complete web-based self-report surveys thrice to measure the effects of the 2-week intervention. The preintervention survey period was from October 31 to November 5, 2019 (T1); the postintervention survey period after the intervention was from November 25 to December 5, 2019 (T2); and the 3-week follow-up survey period was from December 16 to 22, 2019 (T3). Participants were not allowed to proceed to the next question unless they answered the previous question.

Primary Outcome

As the primary outcome, work engagement was measured using the Japanese version of the Utrecht Work Engagement Scale (UWES-J), which was reported to be reliable and valid [38]. The UWES-J consists of three subscales of three items each (eg, vigor: “At my work, I feel bursting with energy;” dedication: “My job inspires me;” and absorption: “I get carried away when I am working”). Each item was assessed on a 7-point Likert scale ranging from 0 (never) to 6 (always). The scale score was calculated by adding all the scores and dividing them by the number of items, ranging from 0 to 6. A higher UWES-J score indicated a higher degree of work engagement.

Demographic Characteristics

Gender, age, marital status, education, job, employment status, working hours per week during the past month, and years of employment tenure were collected at T1. Only education was collected at T2 because we missed out on collecting it at T1.

Statistical Analysis

Descriptive statistics for participants’ characteristics at T1 were calculated to identify the differences between the intervention and wait-list control groups. *T*-tests were conducted for continuous variables such as age, working hours per week, and years of employment tenure. χ^2 tests were conducted for categorical variables such as gender, marital status, education, and job. The mean scores and SDs of work engagement at the three time points were calculated.

We conducted intention-to-treat (ITT) and per-protocol analyses (not adjusted) to evaluate the effect of the WEDiary using a mixed model for repeated measures conditional growth model analyses with unstructured covariance and full information maximum likelihood estimation for the data with missing values, respectively. A group (intervention and wait-list control)×time (T1, T2, and T3) interaction was analyzed as an indicator of the intervention effect. Since some participants did not consistently use the WEDiary at the three time points, only those who used it every working day during the intervention period (2 weeks), excluding weekends, and responded to all three surveys were analyzed for per-protocol analysis. In addition, a mixed model for repeated measures conditional growth model analysis was conducted for subgroup analyses, dividing each of the following into two groups: work engagement, age, and working hours (divided into two groups based on their median values); gender (male/female); educational background (university graduate or higher/high school graduate or lower); white-collar/non-white-collar; and employment status (regular employees/nonregular employees).

The effect sizes (95% CIs) were calculated using Cohen *d* only among the participants for per-protocol analysis, although the effect sizes may be biased due to dropouts. Values of 0.2, 0.5, and 0.8 are generally interpreted as small, medium, and large effects, respectively [39].

To evaluate the process evaluation, the number and percentage of those who installed the WEDiary and active users of the WEDiary were calculated. Active users were defined as those who clicked the “List” button to upload the diary data, such as “Daily achievement,” and rated the four questions on daily work engagement. However, those who clicked the “List” button without uploading the diary data or did not rate the four questions were excluded.

To evaluate the trajectories of the four questions in the WEDiary for 10 days (excluding weekends), a mixed model analysis was conducted using the total score of the four questions as the dependent variable and time as the independent variable.

Two-tailed test with a *P* value set to .05 was used in all statistical analyses. All statistical analyses were performed using the Statistical Package for Social Sciences for Mac version 27 (IBM Corp., Armonk, NY, USA).

Ethical Considerations

This study was reviewed and approved by the Research Ethics Review Board of the Faculty of Policy Management and the Faculty of Environment and Information Studies at Keio University (approval no.: 265), in accordance with the Declaration of Helsinki and relevant ethical guidelines. All participants were provided with detailed information about the study on a designated web page, including its purpose, procedures, potential benefits, and their rights. Informed consent was obtained electronically from those who agreed to participate. Participation was entirely voluntary, and participants were free to withdraw from the study at any time without penalty. To ensure privacy and confidentiality, all questionnaire data were collected by a survey company, anonymized (decoded), and subsequently provided to the researchers. Personally identifiable

XSL•FO
RenderX

information was retained solely by the survey company and was not accessible to the research team. Participants had the opportunity to benefit directly, as the intervention had the potential to improve work engagement. In addition, participants received a reward of points equivalent to 1000 yen (approximately US $9 at the time of the study), which could be used for shopping upon completion of the study.

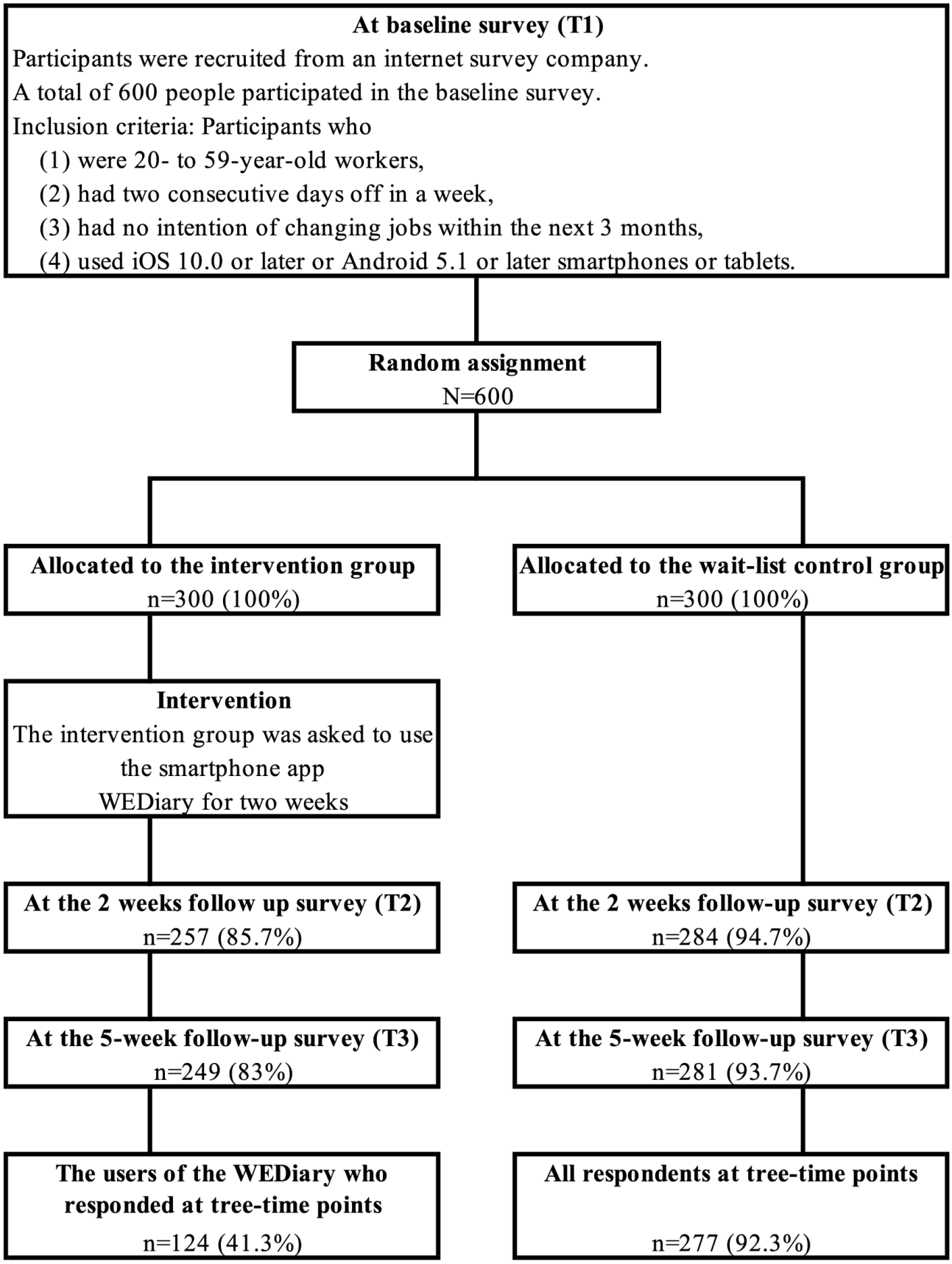
Results

Participants, Randomization, and Attrition

Figure 2 presents the flowchart and randomization of the participants. After the baseline survey, 600 participants were randomly allocated to either the intervention group or the wait-list control group (300 participants per group). After the 2-week intervention period, participants answered the questionnaires again at T2 and at T3. The number of respondents in the intervention group was 257 (85.7%) at T2 and 249 (83%) at T3, while that in the wait-list control group was 284 (94.7%) at T2 and 281 (93.7%) at T3.

XSL•FO
RenderX

Figure 2. Participant flowchart.



Demographic Characteristics

Table 1 presents participants' demographic characteristics at T1. There were no significant differences in characteristics between the intervention and wait-list control groups, except for education. Most participants were male (the intervention group: 68.3%; the control group: 67%) and married (intervention group: 62%; control group: 65.7%). Education was significantly

XSL•FO
RenderX

different between the intervention and wait-list control groups (P<.001). More participants in the intervention group, as opposed to the wait-list control group, had completed high school or less (19.7% for the intervention group and 15.7% for the control group) and graduate school (8.3% for the intervention group and 7.7% for the control group).

XSL•FO
RenderX

Table . Baseline characteristics of participants in the intervention and wait-list control groups.

	Intervention group (n=300)	Wait-list control group (n=300)	*P* value[a]
Gender, n (%)			.73
Male	205 (68.3)	201 (67)	
Female	95 (31.7)	99 (33)	
Age, mean (SD)	40.76 (9.5)	41.69 (10.3)	.25
Marital status, n (%)			.35
Married	186 (62)	197 (65.7)	
Unmarried	114 (38)	103 (34.3)	
Education[b], n (%)			<.001
Under high school	59 (19.7)	47 (15.7)	
College	41 (13.7)	48 (16)	
University	130 (43.3)	166 (55.3)	
Graduate school	25 (8.3)	23 (7.7)	
Other/don’t want to answer	2 (0.7)	0 (0)	
N/A[c]	43 (14.3)	16 (5.3)	
Job, n (%)			.56
Professional/engineer	68 (22.7)	73 (24.3)	
Manager	34 (11.3)	47 (15.7)	
Clerical	91 (30.3)	86 (28.7)	
Sales	37 (12.3)	30 (10)	
Service	11 (3.7)	11 (3.7)	
Production/skills	28 (9.3)	31 (10.3)	
Security officer	2 (0.7)	3 (1)	
Forestry/fishery	1 (0.3)	0 (0)	
Transportation/communications	7 (2.3)	2 (0.7)	
Others	21 (7.0)	17 (5.7)	
Employment status, n (%)			.37
Executive officer	6 (2)	3 (1)	
Self-employed	4 (1.3)	2 (0.7)	
Self-employed (help)	1 (0.3)	0 (0)	
Regular employee	261 (87)	259 (86.3)	
Part-time employee	5 (1.7)	2 (0.7)	
Temporary employee	9 (3)	14 (4.7)	
Contract employee	11 (3.7)	19 (6.3)	
Others	3 (1)	1 (0.3)	
Working hours per week, mean (SD)	41.5 (15.5)	41.6 (17.1)	.92
Employment years, mean (SD)	11.5 (9.5)	12.5 (10.7)	.23

[a]*t* test or χ^2 test.

[b]Measured at T2.

[c]N/A: not available.

XSL•FO
RenderX

Outcome Mean Scores and SDs

Table 2 shows the means and SDs of the outcome variables at T1, T2, and T3 in the intervention and wait-list control groups. The means (SDs) for the intervention group were 2.6 (1.1) at T1, 2.7 (1.1) at T2, and 2.7 (1.2) at T3. The means (SDs) for the wait-list control group were 2.6 (1.1) at T1, 2.6 (1.1) at T2, and 2.5 (1.2) at T3. The reliability estimated by Cronbach α was .93 at T1, .95 at T2, and .95 at T3 for this study.

Table . Means (SDs) of outcome variables at T1, T2, and T3 in the intervention and wait-list control groups for the whole sample.

	Intervention group		Wait-list control group	
	Participants, n	Mean (SD)	Participants, n	Mean (SD)
T1	300	2.6 (1.1)	300	2.6 (1.1)
T2	257	2.7 (1.1)	284	2.6 (1.1)
T3	249	2.7 (1.2)	281	2.5 (1.2)

Effects of the WEDiary

Intention-to-Treat Analyses

Table 3 shows the estimated effect of the WEDiary on work engagement based on the mixed model analyses for the ITT analyses. There was a significant effect on work engagement between the two groups (t=2.03, P=.04). For the subscales, significant effects on vigor (t=2.30, P=.02) and dedication (t=2.78, P=.01) were found, although no significant difference on absorption (t=0.35, P=.73) was found between the two groups.

Table . Result of mixed model for repeated measures conditional growth model analyses for the whole sample throughout the study period.

	Effect	SE	*df*	*t*	*P* value	95% CI (lower to upper)
Work engagement	0.07	0.03	545	2.03	.04	0.00 to 0.13
Vigor	0.08	0.04	549	2.30	.02	0.01 to 0.16
Dedication	0.11	0.04	546	2.78	.01	0.03 to 0.18
Absorption	0.01	0.04	546	0.35	.73	–0.06 to 0.09

Per-Protocol Analyses

In per-protocol analyses, the number of participants was 124 for the intervention group and 277 for the wait-list control group. A significant effect on work engagement (t=2.61, P=.01) was found between the two groups. For subscales, significant effects on vigor (t=2.32, P=.02) and dedication (t=3.26, P=.001), but not on absorption (t=1.22, P=.22), were found between the two groups (Table 4).

Table . Effect of the Work Engagement Diary (WEDiary) on work engagement between the intervention group using the WEDiary for every working day (n=124) and the wait-list control group (n=277).

	Effect	SE	*df*	*t*	*P* value	95% CI (lower to upper)
Work engagement	0.11	0.04	399	2.61	.01	0.03 to 0.19
Vigor	0.11	0.05	399	2.32	.02	0.02 to 0.20
Dedication	0.15	0.05	399	3.26	.001	0.06 to 0.24
Absorption	0.06	0.05	399	1.22	.22	–0.04 to 0.16

Subgroup Analyses

A mixed model analysis was conducted on work engagement as the dependent variable, dividing each of the following into two groups: work engagement, gender, age, working hours, educational background, white-collar/non-white-collar, and employment status. The results showed that work engagement significantly improved among females, university graduates or higher, white-collar workers, and regular employees (t=2.85, P=.005; t=2.37, P=.02; t=2.37, P=.02; t=2.37, P=.02, respectively).

Effect Size

The effect size was small at each time point (T1 vs T2 and T1 vs T3; Table 5). The number of questionnaire respondents at both T1 and T2 was 257 for the intervention group and 284 for the wait-list control group, respectively, whereas that at both T1 and T3 was 249 for the intervention group and 281 for the wait-list control group, respectively.

XSL•FO
RenderX

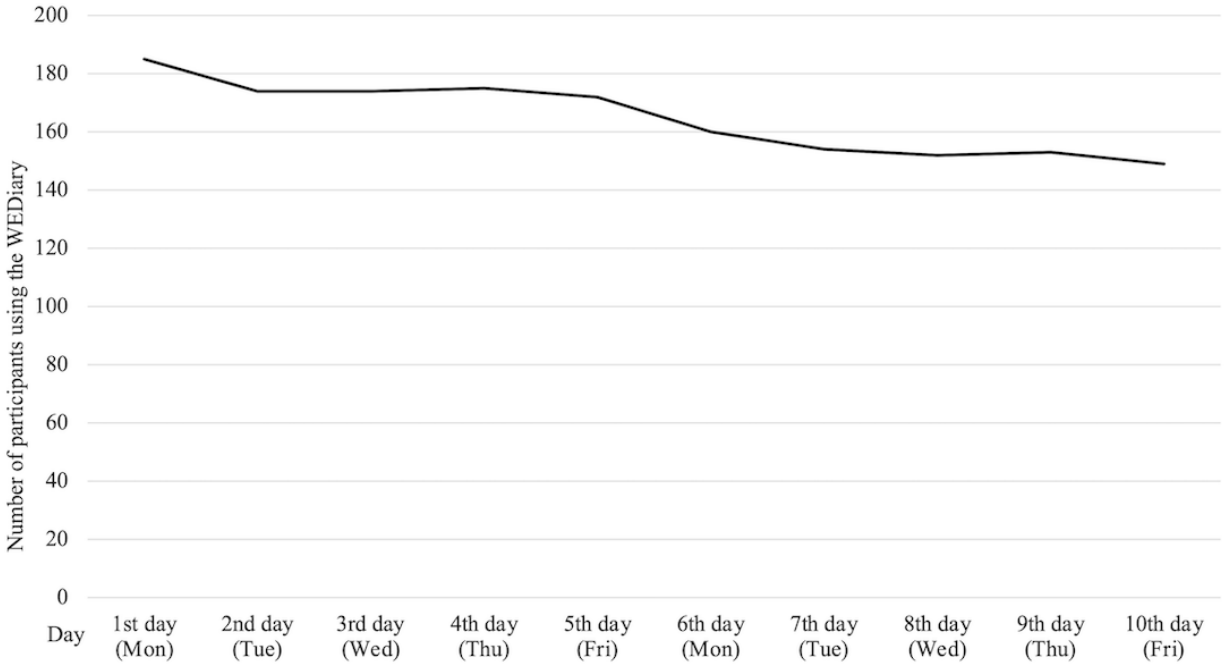
Table . The effect sizes of the Work Engagement Diary (WEDiary).

	T1 versus T2	T1 versus T3
	Cohen *d* (95% CI [lower to upper])	Cohen *d* (95% CI [lower to upper])
Work engagement	0.11 (−0.06 to 0.28)	0.12 (−0.06 to 0.28)
Vigor	0.06 (−0.11 to 0.23)	0.14 (−0.03 to 0.31)
Dedication	0.16 (−0.02 to 0.32)	0.17 (−0.002 to 0.34)
Absorption	0.06 (−0.11 to 0.23)	0.01 (−0.16 to 0.18)

Process Evaluation

Of the 241 (80.3%) participants in the intervention group (n=300) who installed the WEDiary on their smartphones, 191 (63.7%) used it at least once. The number of daily users gradually decreased during the intervention period. The first day of the intervention period had 185 users (61.7%), which was the highest number. In contrast, the number of users on the last day of the intervention period was 149 (49.7%), which was the lowest (Figure 3).

Figure 3. The number of daily WEDiary users.



Trajectories of the Four WEDiary Questions

The total scores of the four questions included in the WEDiary were analyzed using a mixed model (linear regression model), with the total score as the dependent variable and the time point as the independent variable. The intercept (mean value) at the 10 time points was 3.58, and the slope was 0.035 (P<.01), showing a significant increase over time. Each of the four items also had a significant positive slope; however, the slope parameters were small.

Discussion

Effect on Work Engagement

The present study showed that our newly developed smartphone-based positive diary app for enhancing work engagement (WEDiary) among Japanese workers had a significantly favorable effect on work engagement in the intervention group compared to the wait-list control group during the study period. Thus, the WEDiary was effective in improving work engagement among participants. Subgroup analysis showed that work engagement significantly improved among females, university graduates or higher, white-collar workers, and regular employees. This suggests that providing the WEDiary to specific target groups is effective. Females tend to ruminate more than males in response to stress [40]. WEDiary may have guided female participants toward positive rumination (ie, positive reflection). This result can also be explained by the Broaden-and-Build theory, which posits that positive emotions expand individuals' attention and cognitive flexibility, thereby facilitating the generation of diverse ideas and the development of critical personal resources [34]. Participants with a university degree or higher are likely to be self-regulated learners and tend to be highly capable of utilizing self-efficacy [41]. As a result of using WEDiary and enhancing self-efficacy as a personal resource, work engagement may have increased. The JD-R model is a valid theoretical framework regardless of job type; however, the importance of specific demands and resources

XSL•FO
RenderX

may vary depending on the job type [42]. In this study, in terms of enhancing personal resources, white-collar workers may have benefited more from using WEDiary compared to non-white-collar workers. Future research should examine job types rather than the two categories of white-collar and non-white-collar workers to potentially refine the WEDiary. Furthermore, in the JD-R model, nonregular employees may represent a vulnerable group with fewer job resources [43], while regular employees, who typically possess relatively higher job resources, may experience greater enhancement of personal resources through using WEDiary. Thus, the WEDiary, which focuses on one's strengths and positive experiences at work, is likely to be particularly effective for females with a high tendency to ruminate, highly educated individuals with self-control, white-collar workers, and regular employees possessing high job resources.

Effect on Vigor and Dedication

The WEDiary might have facilitated participants to gain a greater sense of achievement, such as through "Daily achievement" (serving as a daily reflective diary) or the visualized line graph, leading to enhanced work engagement. Participants set a weekly goal at work, filled in one daily achievement at work, and answered four questions about work engagement. A text box was provided to indicate daily achievement in one short sentence. Answering the four questions about work engagement and the text input of what was accomplished provided an opportunity to reflect on the day's work. Through this process, participants may gain a greater sense of achievement at work, leading to improved work engagement [17]. In other words, this process may enhance the levels of energy and mental resilience in one's work (vigor) and strong involvement in one's work combined with a sense of significance, enthusiasm, and challenge (dedication).

On the other hand, absorption (ie, having higher concentration, being happily engrossed in one's work, and being in a state of mind in which time seems to pass quickly and in which it is difficult to leave work) may have had delayed effects. The fact that the WEDiary was used at the end of the working day may be why the effect on absorption was not immediately evident. Future research with a longer study period than that of the current study may have more success in observing the effect on absorption. Continuous use of the WEDiary may improve work engagement and, in the long run, absorption. Another possibility is that the WEDiary may have a low effect on absorption. Prior meta-analyses of studies on work engagement improvement also found absorption to be lower than vigor and dedication [23,44]. It may therefore be necessary to consider an improved version of the WEDiary that takes absorption improvement into account.

Comparison With Prior Work

In comparison to prior interventions focusing on building personal resources, the strength of our study lies in targeting diverse workers (hotel employees [45], caregivers [45]), with work engagement (exhaustion [45,46]) as its outcome. Hence, the generalizability of the WEDiary may be higher compared to prior interventions focusing on building personal resources. Vîrgă et al's [23] meta-analysis on the effectiveness of interventions intended to improve work engagement found a modest but significant overall improvement in work engagement (d=0.21, P<.01), with notable gains in vigor (d=0.36, P<.05) and dedication (d=0.35, P<.05), though absorption showed a smaller effect (d=0.16, P<.05). In addition, it was found that intervention effects peaked within the first month (d=0.34, P<.05) but declined over time, becoming negligible by 6 - 12 months. Interventions of up to 2 weeks (d=0.55, P<.05) or those between 12 and 16 weeks (d=0.46, P<.05) were deemed most effective, while very short or extended interventions were less impactful. The WEDiary, a smartphone app based on positive psychology, aims to improve work engagement through a 2-week intervention and a 3-week follow-up, suggesting potential effectiveness based on these findings.

Time of Appearance of the Effect

Effect sizes were relatively larger for T1 versus T3 than for T1 versus T2 (except for absorption). This could be a delayed effect, which may be a sleeper effect [47]. It occurs when an intervention does not have an immediate effect and requires incubation time to produce its effect [47]. It may take time for work engagement to improve through the use of the WEDiary. Regarding previous studies on improving work engagement by focusing on personal resources, those with an intervention period of 30 days [48] and 8 weeks [49] significantly improved work engagement at postintervention, while those with an intervention period of 2 weeks [25] from baseline to follow-up survey and half a day [50] from baseline to 1-month follow-up showed no statistically significant effects. On the contrary, a meta-analysis of intervention studies aimed at improving work engagement [23] reported that the effect size for a 2-week intervention period was relatively high at d=0.55 (P<.05) compared to other intervention periods and that the effect size for a follow-up period of less than 1 month was relatively higher (d=0.034, P<.05) than those for other periods (1 - 3 months: d=0.28; 3 - 6 months: d=0.07; 6 - 12 months: d=0.03; all P<.05). The strength of the present study is that the WEDiary with a 2-week intervention period showed a sustained significant effect throughout the study period of 4 weeks. The intervention and follow-up periods of the WEDiary may have been suitable for evaluating the effect. However, while the intervention aimed at encouraging work engagement improved two of the three subdimensions of work engagement (ie, vigor and dedication), improvements in the third subdimension (ie, absorption) remained limited. This result is in line with recent research that has reported a similar trend. There are several possible reasons for the limited improvement observed in absorption, including work environment characteristics [51] and individual intrinsic motivation [52]. Therefore, it is suggested that the approach to improving immersion should be based not only on psychological intervention but also on a thorough review of work design and the interests and concerns of individual employees.

Small Effect Size

Unfortunately, the effect size was small in this study. However, despite this, effect sizes were relatively larger for T1 versus T3 (Cohen d=0.12) than for T1 versus T2 (Cohen d=0.11). This may be because it takes time to accumulate positive emotions. A prior meta-analysis of interventions aimed at improving work

XSL•FO
RenderX

engagement [23] reported effect sizes of *d*=0.14 (95% CI 0.05-0.22) and *d*=0.29 (95% CI 0.16-0.41) for online and face-to-face interventions, respectively. Considering that the WEDiary is a self-learning program with no feedback function and that the intervention period was 2 weeks long, an effect size of 0.11 immediately after the intervention and 0.12 at follow-up is in line with previous studies and suggests the possibility of participants being in the middle of accumulating positive emotions. This could be because the WEDiary usage rate gradually decreased during the intervention period (2 weeks); users who used the app infrequently or not at all may have weakened the effect of the intervention. Allocated interventions such as the RCT have a smaller effect than volunteered interventions [11]. This speculation could be supported by the fact that the intervention effect improved in the per-protocol analyses compared to ITT analyses. In the future, we need to investigate how to enhance the installation and use of an app, such as using push notifications or messages. However, the significance level was set at α=.05 in this study, and there is a possibility that the statistical power (1–β) may be low, necessitating caution when interpreting the results. A low power increases the risk of Type II errors, potentially leading to false-negative conclusions. Additionally, it is essential to assess whether the effect size used in the sample size calculation (Cohen *d*=0.20) was appropriate, particularly when comparing conditions such as face-to-face versus online formats or the presence versus absence of support. An underestimated or overestimated effect size could impact the validity of the findings and the robustness of the statistical conclusions. Increasing sample sizes has recently become a critical concern in online psychological interventions aimed at behavior change. Previous studies have identified several effective strategies for expanding participant pools. Utilizing digital platforms, such as chatbots, has significantly enhanced sample sizes by improving accessibility and user engagement [53]. Additionally, large-scale online recruitment methods, including the use of social media and targeted advertising, have proven effective in reaching diverse populations and increasing the number of participants [54].

User Attrition

Process evaluation clearly showed user attrition. Positive psychology interventions such as the one used in our study, in particular, often have low participation rates and high attrition [11]. In a positive psychology study involving workers similar to those in the present study, attrition may have been due to participants perceiving the required time commitment as burdensome and to the absence of face-to-face interactions, as WEDiary was self-learning [55]. The following solutions to user attrition have been examined: self‐selection [56], participant autonomy [57], and belief [58]. To apply our results to different work settings, the WEDiary with the addition of the above strategies may have the potential to increase the promotion of use. This study did not include a qualitative survey on user experience such as satisfaction. Further detailed qualitative surveys on user experience are needed in the future. This may contribute to future app updates.

Questionnaire Response Rate

The questionnaire response rate was higher in the wait-list control group than in the intervention group (85.7% vs 94.7% at T2, 83% vs 93.7% at T3). This may be because the expectation of using the WEDiary in the wait-list control group may have encouraged the participants to respond to the questionnaires. In this study, data from the wait-list control group were not collected. However, including a plan to collect and analyze data from the wait-list control group after using the app in such intervention studies may help clarify the mechanisms of improvement in outcomes such as work engagement in more detail.

Limitations

The present study has several limitations. First, the participants were all registered members of the online research company, which requires caution regarding the generalizability of our findings. Second, internet users' socioeconomic and educational status is usually high on average, compared to that of the general population [59,60]. Indeed, participants in the present study had a higher educational status than those completing nationwide paper-and-pencil surveys in Japan [61]. Third, the small sample size compared to the calculated sample size might have lowered statistical power compared to expectations. The sample size of the study was 300 participants per group compared to the estimated number of 394 participants per group needed to detect an effect size of 0.20 for work engagement. Thus, this study had a lower statistical power. Fourth, the low adherence rate to the study protocol may have weakened the effect of this study. Of the 300 participants, 241 (80.3%) installed the app but only 191 (63.7%) used the WEDiary. Previous studies on smartphone-based interventions did not indicate the installation rate because it was required that the smartphone app be installed in advance [62-64]. To evaluate the effects of the intervention, it may be necessary to require participants to download and install the WEDiary prior to the presurvey. Alternatively, we may need to develop strategies to encourage participants to download and install the WEDiary actively. Further research is required to increase the installation rate. Fifth, participants with social desirability or the dropout of nonresponders might cause a biased estimate. Sixth, this study employed a nonblinded RCT design, which may introduce potential bias in both the intervention and wait-list control groups. Seventh, it is possible that work engagement measured using the UWES-J at the three time points correlates with the four questions included in the WEDiary (modified UWES-J). In other words, responding to the UWES-J at T2 and T3 may lead to improving positive emotions in the participants. However, UWES-J items use the present tense to measure the present state, with no items reflecting on the state of only the day. The UWES-J may be unlikely to improve work engagement. Furthermore, of the four WEDiary questions, three include the word "today;" hence, each of these three questions asks for an evaluation of today. Eighth, this study analyzed the positive motivational process of the two processes of the JD-R model. The analysis did not control for factors such as the job demands and job resources of the JD-R model. No control for factors could limit the accurate determination of the effectiveness of the intervention and the potential buffering or mediating effects of these

XSL•FO
RenderX

variables; hence, further research is needed to clarify the mechanisms.

Finally, there are limitations in the generalizability of the results because the present study was conducted for Japanese workers. It is unclear whether similar results can be obtained for non-Japanese workers. Further research is needed for a variety of workers.

Conclusions

This RCT demonstrated that our newly developed smartphone-based positive reflection diary (WEDiary) was effective throughout the three time points in improving work engagement among Japanese workers. Effect sizes were relatively larger for T1 versus T3 than for T1 versus T2. Future research needs to clarify longer-term intervention effects and detailed mechanisms of the intervention effects.

Acknowledgments

This research was funded by a grant from the Japan Productivity Center. The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

MT, SK, KT, and AS developed the study design and collected data. MT, SK, and AS drafted the manuscript. MT, SK, and DM performed the statistical analyses. NT and HN managed the WEDiary. All authors reviewed and approved the final manuscript.

Conflicts of Interest

The authors declare that this research was funded by a grant from the Japan Productivity Center, to which the co-authors NT and HN belong, which could be construed as a potential conflict of interest. Co-author DM manages Better Options, Inc., which did not fund this research. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Checklist 1
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File, 488 KB - mhealth_v13i1e55664_app1.pdf]

References

1. Schaufeli WB, Bakker AB. Job demands, job resources, and their relationship with burnout and engagement: a multi‐sample study. J Organ Behavior 2004 May;25(3):293-315 [FREE Full text] [doi: 10.1002/job.248]
2. Hakanen JJ, Ropponen A, Schaufeli WB, De Witte H. Who is engaged at work? A large-scale study in 30 European countries. J Occup Environ Med 2019 May;61(5):373-381. [doi: 10.1097/JOM.0000000000001528] [Medline: 30557226]
3. Schaufeli WB, Salanova M, González-romá V, Bakker AB. The measurement of engagement and burnout: a two sample confirmatory factor analytic approach. J Happiness Stud 2002 Mar;3(1):71-92. [doi: 10.1023/A:1015630930326]
4. Borst RT, Kruyen PM, Lako CJ, de Vries MS. The attitudinal, behavioral, and performance outcomes of work engagement: a comparative meta-analysis across the public, semipublic, and private sector. Rev Public Pers Adm 2020 Dec;40(4):613-640. [doi: 10.1177/0734371X19840399]
5. Garg N, Singh P. Work engagement as a mediator between subjective well-being and work-and-health outcomes. MRR 2019 Dec 16;43(6):735-752. [doi: 10.1108/MRR-03-2019-0143]
6. Purba SD, Chaterine C, Hardjono S, Clarissa B. Psychological meaningfulness and work engagement effect on doctor's job satisfaction. JDM 2019 Sep 28;10(2):229-239 [FREE Full text] [doi: 10.15294/jdm.v10i2.20835]
7. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol 2001 Jun;86(3):499-512. [Medline: 11419809]
8. Schaufeli WB. Applying the job demands-resources model. Organ Dyn 2017 Apr;46(2):120-132. [doi: 10.1016/j.orgdyn.2017.04.008]
9. Kwon K, Kim T. An integrative literature review of employee engagement and innovative behavior: revisiting the JD-R model. Hum Resour Manag Rev 2020 Jun;30(2):100704. [doi: 10.1016/j.hrmr.2019.100704]
10. Xanthopoulou D, Bakker AB, Demerouti E, Schaufeli WB. Work engagement and financial returns: a diary study on the role of job and personal resources. J Occupat & Organ Psyc 2009 Mar;82(1):183-200. [doi: 10.1348/096317908X285633]
11. Gilbert E, Foulk T, Bono J. Building personal resources through interventions: an integrative review. J Organ Behavior 2018 Feb;39(2):214-228 [FREE Full text] [doi: 10.1002/job.2198]
12. Knight C, Patterson M, Dawson J. Building work engagement: a systematic review and meta-analysis investigating the effectiveness of work engagement interventions. J Organ Behav 2017 Jul;38(6):792-812. [doi: 10.1002/job.2167] [Medline: 28781428]

XSL•FO
RenderX

13. Sonnentag S, Tian AW, Cao J, Grushina SV. Positive work reflection during the evening and next‐day work engagement: testing mediating mechanisms and cyclical processes. J Occupat & Organ Psyc 2021 Dec;94(4):836-865 [FREE Full text] [doi: 10.1111/joop.12362]
14. Salmela-Aro K, Upadyaya K. Role of demands-resources in work engagement and burnout in different career stages. J Vocat Behav 2018 Oct;108:190-200. [doi: 10.1016/j.jvb.2018.08.002]
15. Bono JE, Glomb TM, Shen W, Kim E, Koch AJ. Building positive resources: effects of positive events and positive reflection on work stress and health. AMJ 2013 Dec;56(6):1601-1627. [doi: 10.5465/amj.2011.0272]
16. Meier LL, Cho E, Dumani S. The effect of positive work reflection during leisure time on affective well‐being: results from three diary studies. J Organ Behavior 2016 Feb;37(2):255-278 [FREE Full text] [doi: 10.1002/job.2039]
17. Jimenez WP, Hu X, Xu XV. Thinking about thinking about work: a meta-analysis of off-job positive and negative work-related thoughts. J Bus Psychol 2022 Apr;37(2):237-262. [doi: 10.1007/s10869-021-09742-7]
18. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191]
19. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991 Dec;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
20. Horn R, Lahnaoui A, Reinoso E, et al. Native vs web apps: comparing the energy consumption and performance of android apps and their web counterparts. Presented at: 2023 IEEE/ACM 10th International Conference on Mobile Software Engineering and Systems (MOBILESoft); May 14-15, 2023; Melbourne, Australia p. 44-54. [doi: 10.1109/MOBILSoft59058.2023.00013]
21. Koh J, Tng GYQ, Hartanto A. Potential and pitfalls of mobile mental health apps in traditional treatment: an umbrella review. J Pers Med 2022 Aug 25;12(9):1376. [doi: 10.3390/jpm12091376] [Medline: 36143161]
22. Paganin G, Simbula S. Smartphone-based interventions for employees' well-being promotion: a systematic review. Electron J Appl Stat Anal 2020;13(3):682-712. [doi: 10.1285/i20705948v13n3p682]
23. Vîrgă D, Maricuţoiu LP, Iancu A. The efficacy of work engagement interventions: a meta-analysis of controlled trials. Curr Psychol 2021 Dec;40(12):5863-5880. [doi: 10.1007/s12144-019-00438-z]
24. Lau N, O'Daffer A, Colt S, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798. [doi: 10.2196/17798] [Medline: 32357125]
25. Chen S, Westman M, Eden D. Impact of enhanced resources on anticipatory stress and adjustment to new information technology: a field-experimental test of conservation of resources theory. J Occup Health Psychol 2009 Jul;14(3):219-230. [doi: 10.1037/a0015282] [Medline: 19586218]
26. Sodani M, Yadigari E, Shfia-Abadi A, Mohammadi K. Effectiveness of group-based creativity acquisition on job self-efficacy in a welfare organization in iran. 2011. URL: https://tinyurl.com/mb2akkbh [accessed 2023-04-28]
27. Vuori J, Toppinen-Tanner S, Mutanen P. Effects of resource-building group intervention on career management and mental health in work organizations: randomized controlled field trial. J Appl Psychol 2012 Mar;97(2):273-286. [doi: 10.1037/a0025584] [Medline: 21942405]
28. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018 Oct;17(3):276-277. [doi: 10.1002/wps.20550] [Medline: 30192103]
29. Bartels SL, van Knippenberg RJM, Dassen FCM, et al. A narrative synthesis systematic review of digital self-monitoring interventions for middle-aged and older adults. Internet Interv 2019 Dec;18:100283. [doi: 10.1016/j.invent.2019.100283] [Medline: 31890630]
30. Karlsen IL, Svendsen PA, Abildgaard JS. A review of smartphone applications designed to improve occupational health, safety, and well-being at workplaces. BMC Public Health 2022 Aug 10;22(1):1520. [doi: 10.1186/s12889-022-13821-6] [Medline: 35948904]
31. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271. [doi: 10.2196/jmir.7583] [Medline: 28747293]
32. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
33. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
34. Fredrickson BL. What good are positive emotions? Rev Gen Psychol 1998 Sep;2(3):300-319. [doi: 10.1037/1089-2680.2.3.300] [Medline: 21850154]
35. Fredrickson BL, Joiner T. Positive emotions trigger upward spirals toward emotional well-being. Psychol Sci 2002 Mar;13(2):172-175. [doi: 10.1111/1467-9280.00431] [Medline: 11934003]
36. Ouweneel E, Le Blanc PM, Schaufeli WB, van Wijhe CI. Good morning, good day: a diary study on positive emotions, hope, and work engagement. Human Relations 2012 Sep;65(9):1129-1154. [doi: 10.1177/0018726711429382]

XSL•FO
RenderX

37. Rodríguez-Sánchez AM, Schaufeli W, Salanova M, Cifre E, Sonnenschein M. Enjoyment and absorption: an electronic diary study on daily flow patterns. Work & Stress 2011 Jan;25(1):75-92. [doi: 10.1080/02678373.2011.565619]
38. Shimazu A, Schaufeli WB, Kosugi S, et al. Work engagement in Japan: validation of the Japanese version of the Utrecht Work Engagement Scale. Applied Psychology 2008 Jul;57(3):510-523 [FREE Full text] [doi: 10.1111/j.1464-0597.2008.00333.x]
39. Cohen J. A power primer. In: Kazdin AE, editor. Methodological Issues and Strategies in Clinical Research, 4th edition: American Psychological Association; 2016:279-284 URL: http://content.apa.org/books/14805-018 [accessed 2023-04-28] [doi: 10.1037/14805-018]
40. Nolen-Hoeksema S. Gender differences in depression. Curr Dir Psychol Sci 2001 Oct;10(5):173-176. [doi: 10.1111/1467-8721.00142]
41. Zimmerman BJ. Becoming a self-regulated learner: an overview. Theory Pract 2002 May;41(2):64-70. [doi: 10.1207/s15430421tip4102_2]
42. Van den Broeck A, Elst TV, Baillien E, et al. Job demands, job resources, burnout, work engagement, and their relationships: an analysis across sectors. J Occup Environ Med 2017 Apr;59(4):369-376. [doi: 10.1097/JOM.0000000000000964] [Medline: 28157768]
43. Hakanen JJ, Ropponen A, De Witte H, Schaufeli WB. Testing demands and resources as determinants of vitality among different employment contract groups. a study in 30 European countries. Int J Environ Res Public Health 2019 Dec 6;16(24):4951. [doi: 10.3390/ijerph16244951] [Medline: 31817619]
44. Mazzetti G, Robledo E, Vignoli M, Topa G, Guglielmi D, Schaufeli WB. Work engagement: a meta-analysis using the job demands-resources model. Psychol Rep 2023 Jun;126(3):1069-1107. [doi: 10.1177/00332941211051988] [Medline: 34886729]
45. Xu S, Martinez LR, Van Hoof H, Estrella Duran M, Maldonado Perez G, Gavilanes J. Emotional exhaustion among hotel employees: the interactive effects of affective dispositions and positive work reflection. Cornell Hosp Q 2018 Aug;59(3):285-295. [doi: 10.1177/1938965517748774]
46. Clauss E, Hoppe A, O'Shea D, González Morales MG, Steidle A, Michel A. Promoting personal resources and reducing exhaustion through positive work reflection among caregivers. J Occup Health Psychol 2018 Jan;23(1):127-140. [doi: 10.1037/ocp0000063] [Medline: 27936830]
47. Zapf D, Frese M. Methodological issues in the study of work stress: objective versus subjective measurement of work stress and the question of longitudinal studies. In: Cooper CL, Payne R, editors. Causes, Coping, and Consequences of Stress at Work: John Wiley & Sons; 1988:375-411.
48. Bishop M. Work engagement of older registered nurses: the impact of a caring-based intervention. J Nurs Manag 2013 Oct;21(7):941-949. [doi: 10.1111/jonm.12182] [Medline: 24131082]
49. Ouweneel E, Le Blanc PM, Schaufeli WB. Do‐it‐yourself: an online positive psychology intervention to promote positive emotions, self‐efficacy, and engagement at work. Career Dev Int 2013 May 17;18(2):173-195. [doi: 10.1108/CDI-10-2012-0102]
50. Meyers MC, van Woerkom M. Effects of a strengths intervention on general and work-related well-being: the mediating role of positive affect. J Happiness Stud 2017 Jun;18(3):671-689. [doi: 10.1007/s10902-016-9745-x]
51. Bakker AB, Demerouti E. The job demands‐resources model: state of the art. J Manag Psychol 2007 Apr 3;22(3):309-328. [doi: 10.1108/02683940710733115]
52. Extremera N, Mérida-López S, Sánchez-Álvarez N, Quintana-Orts C. How does emotional intelligence make one feel better at work? The mediational role of work engagement. Int J Environ Res Public Health 2018 Sep 2;15(9):1909. [doi: 10.3390/ijerph15091909] [Medline: 30200548]
53. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018. [doi: 10.2196/15018] [Medline: 31674920]
54. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
55. Kaplan S, Bradley-Geist JC, Ahmad A, Anderson A, Hargrove AK, Lindsey A. A test of two positive psychology interventions to increase employee well-being. J Bus Psychol 2014 Sep;29(3):367-380. [doi: 10.1007/s10869-013-9319-4]
56. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
57. Trougakos JP, Hideg I, Cheng BH, Beal DJ. Lunch breaks unpacked: the role of autonomy as a moderator of recovery during lunch. AMJ 2014 Apr;57(2):405-421. [doi: 10.5465/amj.2011.1072]
58. Ng W. Boosting well-being with positive psychology interventions: moderating role of personality and other factors. J Contemp Psychother 2015 Jun;45(2):79-87. [doi: 10.1007/s10879-014-9291-y]
59. Gosling SD, Mason W. Internet research in psychology. Annu Rev Psychol 2015 Jan 3;66(1):877-902. [doi: 10.1146/annurev-psych-010814-015321] [Medline: 25251483]

XSL•FO
RenderX

60. Smith MA, Leigh B. Virtual subjects: using the Internet as an alternative source of subjects and research environment. Behav Res Methods Instrum Comput 1997 Dec;29(4):496-505. [doi: 10.3758/BF03210601]
61. Oshio T, Kobayashi M. Income inequality, perceived happiness, and self-rated health: evidence from nationwide surveys in Japan. Soc Sci Med 2010 May;70(9):1358-1366. [doi: 10.1016/j.socscimed.2010.01.010] [Medline: 20167407]
62. Kristjánsdóttir OB, Fors EA, Eide E, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013 Jan 7;15(1):e5. [doi: 10.2196/jmir.2249] [Medline: 23291270]
63. Toro-Ramos T, Lee DH, Kim Y, et al. Effectiveness of a smartphone application for the management of metabolic syndrome components focusing on weight loss: a preliminary study. Metab Syndr Relat Disord 2017 Nov;15(9):465-473. [doi: 10.1089/met.2017.0062] [Medline: 29035677]
64. Lukas CA, Eskofier B, Berking M. A gamified smartphone-based intervention for depression: randomized controlled pilot trial. JMIR Ment Health 2021 Jul 20;8(7):e16643. [doi: 10.2196/16643] [Medline: 34283037]

Abbreviations

ITT: intention-to-treat
JD-R: job demands–resources
RCT: randomized controlled trial
UWES-J: Utrecht Work Engagement Scale-Japanese version
WEDiary: Work Engagement Diary

Edited by L Buis; submitted 29.07.24; peer-reviewed by CY Chen, H Li; revised version received 07.08.25; accepted 08.08.25; published 25.09.25.

Please cite as:
Tokita M, Kobayashi S, Miyanaka D, Takate N, Nakano H, Takeuchi K, Shimazu A
Effects of a Smartphone-Based Positive Reflection Diary on Work Engagement Among Japanese Workers: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e55664
URL: https://mhealth.jmir.org/2025/1/e55664
doi:10.2196/55664

©Masahito Tokita, Shuichiro Kobayashi, Daisuke Miyanaka, Nobuyuki Takate, Hiroyuki Nakano, Kazuki Takeuchi, Akihito Shimazu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Automatic Image Recognition Meal Reporting Among Young Adults: Randomized Controlled Trial

Prasan Kumar Sahoo[1,2], PhD; Sherry Yueh-Hsia Chiu[3,4], PhD; Yu-Sheng Lin[5], MD; Chien-Hung Chen[6], MSc; Denisa Irianti[7], MSc; Hsin-Yun Chen[8], MSc; Mekhla Sarkar[1], MTech; Ying-Chieh Liu[5,7], PhD

[1]Department of Computer Science and Information Engineering, College of Engineering, Chang Gung University, Taoyuan, Taiwan

[2]Department of Neurology, Chang Gung Memorial Hospital, Linkou Medical Center, Taoyuan, Taiwan

[3]Department of Health Care Management, College of Management, Chang Gung University, Taoyuan, Taiwan

[4]Division of Hepato-Gastroenterology, Department of Internal Medicine, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan

[5]Healthcare Center, Department of Internal Medicine, Taoyuan Chang Gung Memorial Hospital, Taoyuan, Taiwan

[6]Digital Transformation Research Institute, Institute for Information Industry, Taipei, Taiwan

[7]Department of Industrial Design, College of Management, Chang Gung University, 259 Wenhua 1st Road, Guishan District, Taoyuan, Taiwan

[8]Department of Nutrition Therapy, Chang Gung Memorial Hospital, Taoyuan, Taiwan

Corresponding Author:
Ying-Chieh Liu, PhD
Healthcare Center, Department of Internal Medicine, Taoyuan Chang Gung Memorial Hospital, Taoyuan, Taiwan

Abstract

Background: Advances in artificial intelligence technology have raised new possibilities for the effective evaluation of daily dietary intake, but more empirical study is needed for the use of such technologies under realistic meal scenarios. This study developed an automated food recognition technology, which was then integrated into its previous design to improve usability for meal reporting. The newly developed app allowed for the automatic detection and recognition of multiple dishes within a single real-time food image as input. App performance was tested using young adults in authentic dining conditions.

Objective: A 2-group comparative study was conducted to assess app performance using metrics including accuracy, efficiency, and user perception. The experimental group, named the automatic image-based reporting (AIR) group, was compared against a control group using the previous version, named the voice input reporting (VIR) group. Each application is primarily designed to facilitate a distinct method of food intake reporting. AIR users capture and upload images of their selected dishes, supplemented with voice commands where appropriate. VIR users supplement the uploaded image with verbal inputs for food names and attributes.

Methods: The 2 mobile apps were subjected to a head-to-head parallel randomized evaluation. A cohort of 42 young adults aged 20 - 25 years (9 male and 33 female participants) was recruited from a university in Taiwan and randomly assigned to 2 groups, that is, AIR (n=22) and VIR (n=20). Both groups were assessed using the same menu of 17 dishes. Each meal was designed to represent a typical lunch or dinner setting, with 1 staple, 1 main course, and 3 side dishes. All participants used the app on the same type of smartphone, with the interfaces of both using uniform user interactions, icons, and layouts. Analysis of the gathered data focused on assessing reporting accuracy, time efficiency, and user perception.

Results: For the AIR group, 86% (189/220) of dishes were correctly identified, whereas 68% (136/200) of dishes were accurately reported. The AIR group exhibited a significantly higher degree of identification accuracy compared to the VIR group (P<.001). The AIR group also required significantly less time to complete food reporting (P<.001). System usability scale scores showed both apps were perceived as having high usability and learnability (P=.20).

Conclusions: The AIR group outperformed the VIR group concerning accuracy and time efficiency for overall dish reporting within the meal testing scenario. While further technological enhancement may be required, artificial intelligence vision technology integration into existing mobile apps holds promise. Our results provide evidence-based contributions to the integration of automatic image recognition technology into existing apps in terms of user interaction efficacy and overall ease of use. Further empirical work is required, including full-scale randomized controlled trials and assessments of user perception under various conditions.

Trial Registration: International Standard Randomized Trial Registry ISRCTN27511195; https://www.isrctn.com/ISRCTN27511195

(JMIR Mhealth Uhealth 2025;13:e60070) doi:10.2196/60070

XSL•FO
RenderX

KEYWORDS

automatic food image recognition; speech recognition; artificial intelligence; usability evaluation; mHealth; image recognition; randomized controlled trial; recognition; nutrition; accuracy; efficacy; user perception; Taiwan; user interaction; vision technology

Introduction

Background

In 2022, the World Health Organization [1,2] classified 2.5 billion adults (aged 18 years and older) as overweight and thus susceptible to chronic diseases associated with obesity. Failure to maintain appropriate nutrition among young people leads to a range of health issues later in life [3,4]. Young people also increasingly integrate smartphones into their daily lives, raising a growing interest in using this technology platform to deliver health-improving behavioral interventions, including healthy eating among young adults [5,6]. Such mobile health (mHealth) interventions are increasingly used to encourage healthy eating behaviors [7-10] and are increasingly popular among young users [11,12]. mHealth technologies already play a significant role in reshaping health care access [13,14] and allow for broad access to scalable and cost-effective solutions [8].

Artificial intelligence (AI)-based services are an emerging trend [15,16], with an increasingly substantial impact on various health care domains, providing enhanced accuracy, improved outcomes, and cost-effectiveness [17-19], and health care professionals have been found to hold favorable views toward AI [20,21]. Machine learning, an increasingly mature AI application, has the potential to revolutionize the mHealth domain [11,14].

Challenges in Dietary Intake Input

An ideal mHealth app should prioritize ease of use, reliability, and long-term engagement [13,22,23]. However, manually entering dietary intake information poses significant usability challenges for mHealth app users [24], potentially resulting in inaccurate or incomplete reporting, and thus undermining the efficacy of managing healthy eating habits [16]. This raises an urgent need to minimize the operational loading of users [25,26], as the ease and effectiveness of food data entry methods have a direct and significant impact on the usability of dietary tracking applications [23]. Recent advancements in computer vision and deep learning show potential for replacing traditional input methods [27,28]. This study integrates automatic image recognition technology [29-31] into an mHealth app, which originally used voice-based inputs, seeking to improve accuracy and time efficiency in meal reporting.

Objectives

This study assesses a parallel 2-group randomized trial designed to evaluate the relative effectiveness of a new automatic image-based reporting (AIR) app compared to the existing voice input reporting (VIR) app [32], in terms of reporting accuracy, time efficiency, and user perception. The key technological components use newly developed AI features to integrate a set of food images representing both single and mixed dishes [31].

Methods

General Overview of the Approach

The original app was designed to aid individuals in improving their dietary habits [33]. The previous iteration included voice inputs to enhance meal reporting, using AI services, specifically Google AI, to transcribe speech into text [32]. Users could use this feature to vocally report food ingredients, portion sizes, cooking methods, and other attributes for individual dishes. While the existing design was shown to be positive in terms of accuracy and was generally well received by users, there were concerns regarding the accuracy and time-consuming nature of completing meal reporting for an entire meal. Furthermore, in authentic dietary intake scenarios, voice reporting during meal consumption was not always convenient. Consequently, we developed the latest version to enhance the existing design.

The iterative development process was rooted in a user-centered design model [34] and included research, ideation, and implementation stages. We reviewed the relevant literature and commercially available apps, along with extensive brainstorming sessions among team members to generate diverse design ideas. AI techniques have been increasingly applied to food identification and nutrition-related applications; thus, one idea raised in the ideation stage was to allow users to take a single photo of an entire meal for analysis using AI-based recognition, rather than to process individual dishes in sequence. The initial automated meal recognition system was developed and validated using our developed AI recognition model, extracting features from convolutional neural networks [31], and tested under a laboratory setting. The results offered relatively high mean average precision for a range of dish types. The newly developed feature allowed users to upload a meal image for automatic recognition using the AI engine located on a remote server. This functionality was then integrated into the previous app version to improve convenience, accuracy, and time efficiency for meal reporting. The previous version, the voice-only reporting app, allowed users to simultaneously use verbal reporting of food names and attributes [32].

App Implementation

The 2 apps were implemented in a 6.8-inch smartphone using the Android (Google, Inc) operating system. The AI server used in our previous research was improved to simultaneously recognize a set of multiple dishes and achieve a relatively high accuracy of food-image recognition under a laboratory setting [31]. Both apps used the Google Speech Cloud service for continuous speech recognition. The developed interfaces included user-friendly design elements, such as large-sized buttons and text, a simple layout, and high-contrast colors. Based on recommended design guidelines [32], the 2 apps shared a common interface design including the placement of buttons, text, and icons. Clear and intuitive visual cues were used to facilitate user interaction.

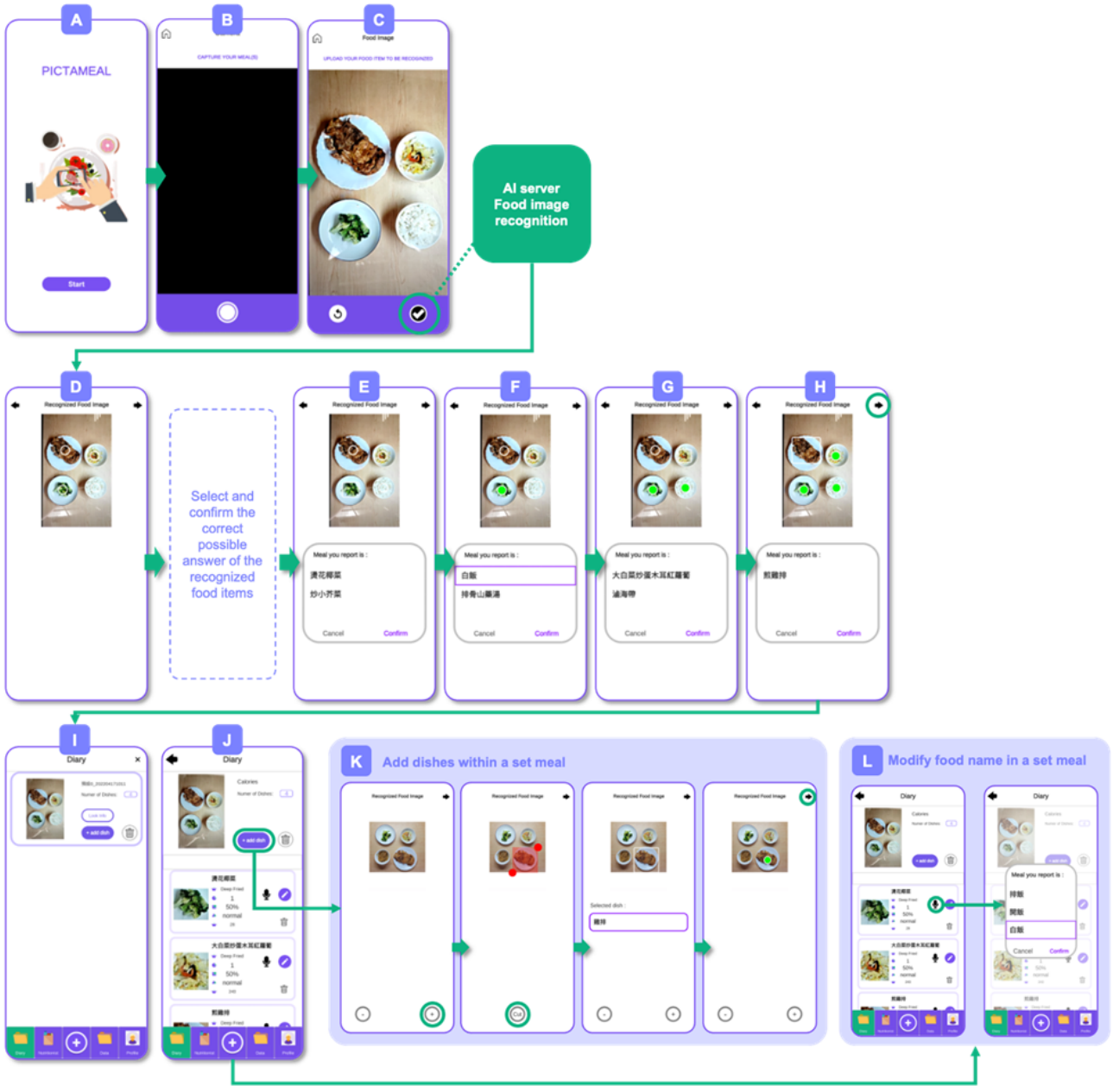
XSL•FO
RenderX

App Operation

Overview

Figure 1 summarizes the 4 major stages in user interaction for the AIR design. In the first stage, users take and upload a photo of a meal (Figure 1A-C), followed by modification of the food ingredients and cooking method for each of the dishes and the provision of additional information, for example, portion size (Figure 1I,J). Users reviewed, revised, and confirmed the image, calorie content, and macronutrient information for each dish (Figure 1J,K) and were given the opportunity to add missing dishes (Figure 1J,L).

Figure 1. Workflow of the automatic image-based reporting (AIR) app: the user captures a food image, and the app automatically detects and works to recognize the dish (illustrated here with steamed rice as an example).



About AIR

To activate the app, users click the "start" button (Figure 1A) and then activate the smartphone's built-in camera (Figure 1B). To capture the meal photo, the user clicks the icon in the lower-right corner (Figure 1C) and can retake the photo using the icon on the lower-left. Once confirmed by the user using the upload button (lower-right in Figure 1D), the meal photo is uploaded to the remote AI server for analysis. Analysis results are then sent to the app, with white dots appearing above each dish that was successfully recognized. In Figure 1, four dishes were successfully recognized (Figure 1D). Figures 1E-1H illustrate the steps by which users select the correct dish name.

For each recognized dish, up to 3 possible names with the highest confidence scores are provided (see Figure 1E). The user then selects the correct name. Upon confirmation, the color of the small circle switches from white to green, indicating that the dish entry has been completed (Figure 1E-H).

If the dish item is not recognized, or if the selection provided does not include an appropriate answer, the user resorts to an alternative input method, that is, voice input. To use the voice

XSL•FO
RenderX

input method, the user drags a boundary square to surround a specific dish, then clicks the "voice" button and recites the dish name aloud (Figure 1I-K). For example, Figure 1J shows an unrecognized dish with 1 to 3 possible answers, prompting the user to respond by selecting an option using the "check" button or to return to the previous step using the "undo" button. Once all the dishes are reported, the user proceeds to the Food Diary Page to input related information such as portion size and cooking method (Figure 1J). Once dish reporting is complete, detailed information is provided for the entire meal (Figure 1L top) and for each individual item (Figure 1L bottom), including calorie count, food item name, and food attributes.

About VIR

The VIR process initially follows the procedure depicted in Figure 1A-C. However, rather than uploading the image to the server for automatic recognition (as depicted in Figure 1E-H), users manually manipulate the cursor (depicted as a red square in Figure 1K) to scroll down and adjust the size of the area displaying the dish image. The scrolled image is then stored for later visualization on the subsequent page (Figure 1H). Once the desired dish is identified, the user clicks the microphone icon and records the dish name. Then the user uses the voice input function to report the dish ingredients and cooking method, prompting the system to provide up to 3 options for selection and confirmation by clicking the check button, thereby concluding the reporting process. If none of the supplied answers are correct, the user clicks the microphone icon and repeats the process until the correct answer is determined. Once each dish reporting is finalized, a green circle appears above the dish. Once all dish reporting is complete, the user proceeds to the Food Diary Page to input related food information using the voice reporting process previously discussed.

Ethical Considerations

The study protocol was reviewed by the ethics committee of Chang Gung Memorial Hospital and received institutional review board approval (202101985B0C501). A total of 42 young adults were recruited and provided informed consent. Participants received a small incentive (ie, New Taiwanese $150, roughly US $5) upon completing the study sessions. All data were handled in deidentified form to ensure participant privacy.

Study Design and Participant Recruitment

We conducted a parallel 2-group randomized trial to compare the performance of the AIR app against the VIR app in terms of accuracy, time efficiency, and user perception. Study participant recruitment was conducted through notices placed on bulletin boards in Chang Gung University in Taoyuan City, Taiwan. Registration, schedule arrangement, and collection of background information were conducted through a web-based form. Biographic data were used to allocate participants into the AIR and VIR groups. Self-reported baseline information included gender, age, BMI, experience in nutrition education, use of nutrition-related apps, cooking experience, and experience using mobile phones or tablets. Eligible participants were (1) aged from 20 to 25 years and (2) capable of operating the app on their mobile phones. Participants currently under any form of dietary control, currently engaged in deliberate weight loss, or following a vegetarian diet were excluded. The assessment was conducted in a cafeteria at Chang Gung University.

Dishes for the experiment were selected under the supervision of a senior nutritionist (HYC) and consisted of typical local foods familiar to study participants. Foods were presented in terms of set meals involving 17 food items that were used to represent lunch and dinner. Each set meal contained 5 food items (ie, a staple food; a main course; a dish with 1 ingredient, such as stir-fried broccoli; a dish with 2 ingredients, such as stir-fried egg with tomato; and a dish with 3 ingredients).

Sample Size Estimation

Drawing from our prior experience evaluating the young population [35], mobile app usage within this demographic exhibited nearly perfect accuracy. Consequently, in addition to assessing accuracy, we aimed to explore potential differences in time efficiency when young individuals used the newly designed app. Our previous study of customized dietary recording among young participants [35] found that the mean duration difference for the assessment of a fried pork chop was 0.7464 (2.60 vs 3.34 seconds) with SD=0.93. The sample size of that study provided a statistical power of 80% and a 2-tailed alpha level of 5%, indicating the minimum sample size required was 26 participants each for the AIR and VIR groups.

Randomization

SAS software (SAS Institute) was used to generate randomized lists of equal size with a 1:1 ratio for the 2 studies, with 22 and 20 participants, respectively, assigned to the AIR and VIR groups.

Evaluation Outcomes

Three outcome measures were evaluated to assess the respective performance of the 2 mobile apps. Data were collected automatically from each participant's interaction, including the tapping of function buttons, and each interaction was logged with a timestamp. In addition, all suggested results were recorded, along with the user's subsequent actions.

Accuracy

Accuracy was defined as the degree to which the response provided by the participant during app operation matched the predefined answer. Conversely, an incorrect response occurred if the app output did not match the predefined answer. The accuracy rate for a certain dish, for example, fried rice, was calculated as the overall number of correct responses divided by the total number. Furthermore, following our previous study [32], four error types were defined as follows: (1) "missing cooking method," where the app provided an incorrect answer due to the absence of a cooking method (eg, "stir-fry"); (2) "incorrect cooking method," where the app provided an incorrect answer due to the use of the wrong cooking method; (3) "irrelevant food name," where the incorrect answer was due to a food name not matching the desired one; and (4) "missing food ingredients," defined as when essential food ingredients (eg, rice) were not included in the reported answer.

XSL•FO
RenderX

Task Duration

The operating duration began when the participant started to report a food dish until the reporting task was completed. For the AIR group, having uploaded the image, the task duration was calculated from the time the participant clicked on the dish displayed on the screen to when the participant clicked the "complete" button. For the VIR group, the task duration was calculated from the time the participant clicked the "voice" button to begin recording to when the participant clicked the "complete" button.

Perception

Brooke's system usability scale (SUS) [36] measures participant perception using a questionnaire comprising 10 items on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Consistent with the approach outlined by Bangor et al [37], the mean SUS score was used alongside an adjective rating scale in which mean scores of 35.7, 50.9, 71.4, and 85.5, respectively, corresponded to the adjective scales "poor," "ok," "good," and "excellent." There is no fifth adjective category, and scores above 85.5 are generally considered "excellent."

Assessment Procedures

The experiment was conducted by 2 research assistants (DI and MS) who obtained informed consent from all participants. All participants used an identical 6.8-inch Android smartphone, and all trials took place in the same university cafeteria, with individual participant sessions scheduled by appointment. Before the experiment, each participant underwent training by watching an instructional video illustrating how to use the food reporting app. Each participant was allowed to interact with the app for several minutes to ensure familiarity before the trial and was informed that task completion time was included as a performance metric.

Each individual trial included 2 consecutive sessions, each requiring the reporting of a set meal. Each set meal included a staple food, a main course, a dish with 1 ingredient, a dish with 2 ingredients, and a dish with 3 ingredients. Participants were allowed to rest for up to 3 minutes between each test. The total test duration for each participant was approximately 1 hour, and all participants successfully completed the assessment.

Statistical Analysis

Chi-square and *t* tests were applied to examine the baseline characteristics of participants for categorical and continuous variables, respectively. The accuracy between different groups was reported as the proportion of error, calculated as the number of errors or total answer items. The time duration for operating assessment was also used to evaluate efficiency. As the time duration is a continuous variable, a *t* test was used to assess and compare the difference between the 2 groups. This comparison was also applied to dishes with different ingredients. SAS (version 9.1.4) was used to conduct all statistical analyses. In all 2-tailed statistical tests, $P<.05$ was considered statistically significant.

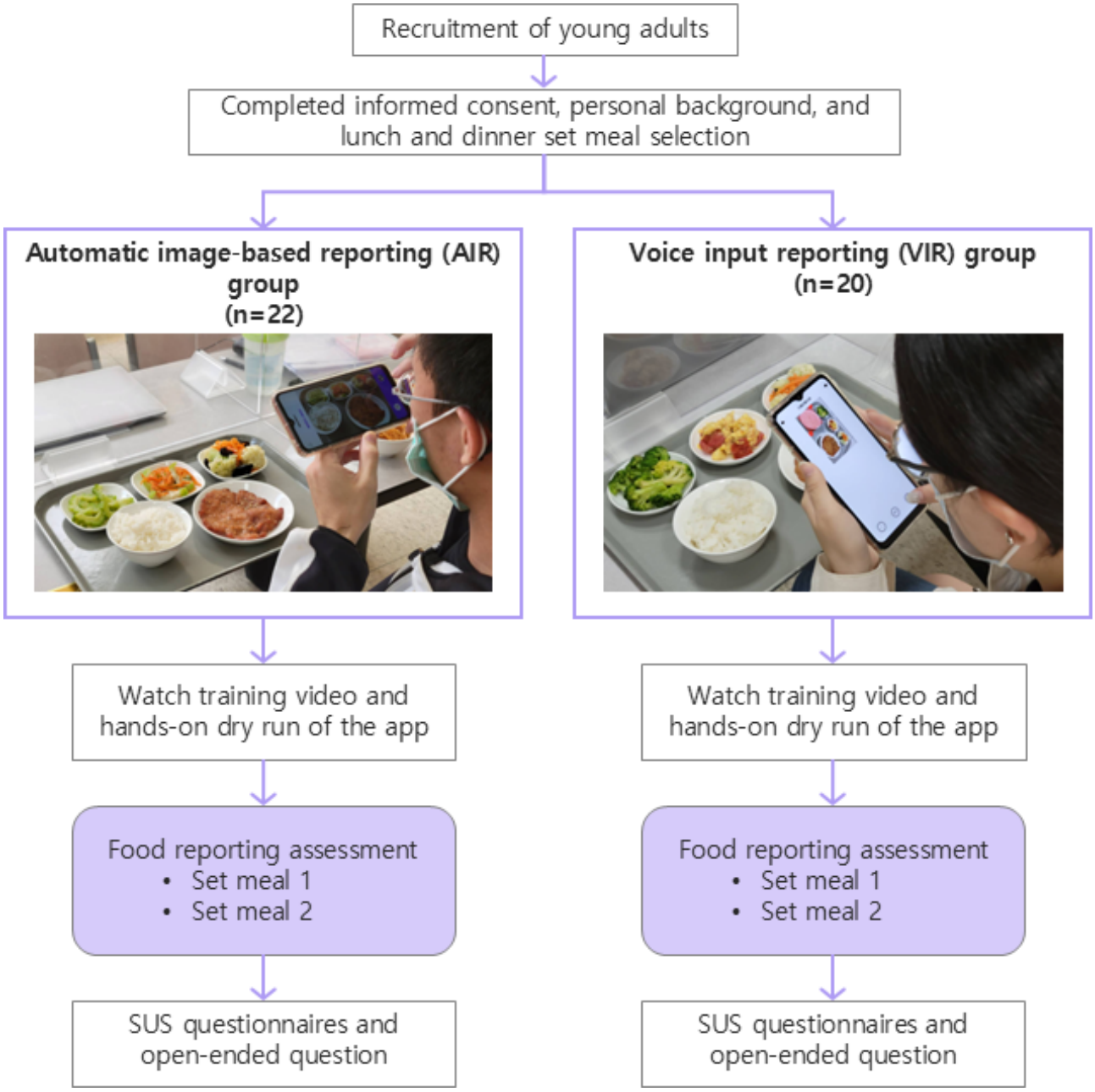
Results

Participant Characteristics

Following the assessment procedures shown in Figure 2, all 42 participants completed the experiment, including 9 male and 33 female participants, with a mean age of 21.21 (SD 1.39) years. As shown in Table 1, 22 and 20 respondents were randomly assigned to the AIR and VIR groups.

XSL•FO
RenderX

Figure 2. App evaluation flow using a randomized design. SUS: system usability scale.



XSL•FO
RenderX

Table . Participant characteristics of the AIR[a] and VIR[b] groups.

Variables	Total (N=42)	AIR (n=22)	VIR (n=20)	*P* value
Gender, n (%)				.60
Female	33 (79)	18 (82)	15 (76)	
Male	9 (21)	4 (18)	5 (24)	
Age (years)[c], mean (SD)	21.21 (1.39)	21.32 (1.29)	21.10 (1.52)	.62
BMI (kg/m^2)[c], mean (SD)	21.10 (3.17)	20.50 (2.77)	21.80 (3.51)	.21
Education level, n (%)				.60
Bachelor's degree	34 (81)	17 (77)	17 (85)	
Master's degree	8 (19)	5 (23)	3 (15)	
Q1. Experience with nutrition-related courses, n (%)				.59
Yes	27 (64)	15 (68)	12 (60)	
No	15 (36)	7 (32)	8 (40)	
Q2. Experience with health education, n (%)				.98
Yes	23 (55)	12 (55)	11 (55)	
No	19 (45)	10 (45)	9 (45)	
Q3. Experience in cooking, n (%)				.51
Yes	39 (93)	21 (95)	18 (90)	
No	3 (7)	1 (5)	2 (10)	
Q4. Experience using nutrition-related apps, n (%)				.75
Yes	22 (52)	11 (50)	11 (55)	
No	20 (48)	11 (50)	9 (45)	

[a]AIR: automatic image-based reporting.

[b]VIR: voice input reporting.

[c]Age and BMI data were analyzed with ANOVA.

Overall Accuracy

The AIR and VIR groups achieved respective overall accuracy levels of 86% (189/220) and 68% (136/200) for all 17 food dishes (Table 2). Within the food categories, the AIR group was significantly more accurate than the VIR group for the staple food (P<.05), main course (P<.05), and side dish with 3 ingredients (P<.05). No significant differences were found for the other food categories (ie, side dishes with 1 or 2 ingredients).

XSL•FO
RenderX

Table . Overall accuracy comparison in the AIR[a] and VIR[b] groups.

Type and dishes	Method		*P* value
	AIR (n=22)	VIR (n=20)	
Overall			<.001
Correct	189 (86)	136 (68)	
Incorrect	31 (14)	64 (32)	
Staple food, n (%)			
Overall			.01
Correct	43 (98)	32 (80)	
Incorrect	1 (2)	8 (20)	
Stir-fried noodle			.007
Correct	20 (95)	11 (58)	
Incorrect	1 (5)	8 (42)	
Steamed rice			N/A[c]
Correct	19 (100)	17 (100)	
Incorrect	0 (0)	0 (0)	
Fried rice			N/A
Correct	4 (100)	4 (100)	
Incorrect	0 (0)	0 (0)	
Main course, n (%)			
Overall			.04
Correct	40 (91)	29 (73)	
Incorrect	4 (9)	11 (28)	
Pan-fried chicken breast			.06
Correct	24 (92)	13 (68)	
Incorrect	2 (8)	6 (32)	
Braised pork chop			.42
Correct	16 (89)	16 (76)	
Incorrect	2 (11)	5 (24)	
Side dish with 1 ingredient, n (%)			
Overall			.06
Correct	31 (70)	20 (50)	
Incorrect	13 (30)	20 (50)	
Stir-fried eggplant			.10
Correct	14 (82)	5 (50)	
Incorrect	3 (18)	5 (50)	
Stir-fried kelp			.63
Correct	11 (85)	8 (73)	
Incorrect	2 (15)	3 (27)	
Stir-fried cauliflower			>.99
Correct	4 (50)	6 (46)	
Incorrect	4 (50)	7 (54)	
Stir-fried broccoli			>.99

XSL•FO
RenderX

Type and dishes	Method		*P* value
	AIR (n=22)	VIR (n=20)	
Correct	0 (0)	1 (25)	
Incorrect	4 (100)	3 (75)	
Stir-fried bitter melon			.33
Correct	2 (100)	0 (0)	
Incorrect	0 (0)	2 (100)	
Side dish with 2 ingredients, n (%)			
Overall			.30
Correct	37 (84)	30 (75)	
Incorrect	7 (16)	10 (25)	
Stir-fried egg with tomato			>.99
Correct	15 (94)	14 (93)	
Incorrect	1 (6)	1 (7)	
Stir-fried loofah with carrot			>.99
Correct	9 (69)	8 (67)	
Incorrect	4 (31)	4 (33)	
Stir-fried pork with bell pepper			.03
Correct	10 (100)	5 (50)	
Incorrect	0 (0)	5 (50)	
Stir-fried bitter melon with carrot			.46
Correct	3 (60)	3 (100)	
Incorrect	2 (40)	0 (0)	
Side dish with 3 ingredients, n (%)			
Overall			.02
Correct	38 (86)	25 (63)	
Incorrect	6 (14)	15 (38)	
Stir-fried Chinese cabbage with carrot and black fungus			.11
Correct	18 (90)	11 (65)	
Incorrect	2 (10)	6 (35)	
Stir-fried cauliflower with carrot and black fungus			.46
Correct	16 (80)	11 (65)	
Incorrect	4 (20)	6 (35)	
Stir-fried bean sprout with carrot and black fungus			.20
Correct	4 (100)	3 (50)	
Incorrect	0 (0)	3 (50)	

[a]AIR: automatic image-based reporting.

[b]VIR: voice input reporting.

[c]N/A: not applicable.

The AIR group achieved accuracy rates exceeding 95% for individual dishes, including stir-fried noodle, steamed rice, fried rice, stir-fried bitter melon, stir-fried bitter melon with carrot, and stir-fried bean sprout with carrot and black fungus. The VIR group achieved accuracy rates exceeding 95% for stir-fried noodle, steamed rice, fried rice, and stir-fried bitter melon with carrot. The lowest accuracy for both groups was for stir-fried broccoli (less than 30% accuracy).

XSL•FO
RenderX

Time Efficiency

Compared to the VIR group, the AIR group required significantly less time for task completion (P<.001), requiring around 2-25 seconds per task, with “steamed rice” being the fastest (mean 2.05, SD 1.43 seconds) and “stir-fried cauliflower with carrot and black fungus” the slowest (mean 24.90, SD 20.60 seconds; Table 3). In the VIR group, the operation time ranged from around 12 to 22 seconds per task, with “stir-fried kelp” being the fastest (mean 12.76, SD 4.12 seconds) and “stir-fried cauliflower with carrot and black fungus” the slowest (mean 21.65, SD 5.25 seconds).

Table . Dish reporting time in the AIR[a] and VIR[b] groups.

Dish	Reporting time (seconds)				*P* value
	AIR (n=22)		VIR (n=20)		
	n	Mean (SD)	n	Mean (SD)	
Overall	—[c]	12.43 (12.42)	—	16.25 (5.22)	<.001
Staple food	44	5.61 (6.21)	40	14.53 (5.29)	<.001
Stir-fried noodle	21	8.76 (7.42)	19	15.40 (5.75)	.003
Steamed rice	19	2.05 (1.43)	17	13.85 (5.37)	<.001
Fried rice	4	6.00 (4.90)	4	13.50 (1.76)	.03
Main course	44	9.32 (6.87)	40	15.53 (4.19)	<.001
Pan-fried chicken breast	26	6.27 (5.33)	19	16.37 (5.03)	<.001
Braised pork chop	18	13.72 (6.56)	21	14.76 (3.19)	.52
Side dish with 1 ingredient	44	12.16 (9.55)	40	15.93 (5.38)	.03
Stir-fried eggplant	17	15.06 (11.51)	10	20.40 (6.73)	.20
Stir-fried kelp	13	9.08 (6.03)	11	12.76 (4.12)	.10
Stir-fried cauliflower	8	8.00 (6.18)	13	14.54 (3.49)	<.001
Stir-fried broccoli	4	20.33 (14.64)	4	16.65 (4.25)	.57
Stir-fried bitter melon	2	14.00 (1.41)	2	18.70 (0.42)	.05
Side dish with 2 ingredients	44	15.27 (13.81)	40	15.48 (3.58)	.93
Stir-fried egg with tomato	16	12.44 (4.49)	15	15.36 (3.21)	.05
Stir-fried loofah with carrot	13	20.62 (22.68)	12	16.92 (3.63)	.58
Stir-fried pork with bell pepper	10	11.00 (4.52)	10	14.58 (3.91)	.07
Stir-fried bitter melon with carrot	5	19.00 (14.09)	3	13.33 (3.56)	.53
Side dish with 3 ingredients	44	19.80 (17.20)	40	19.76 (5.88)	.99
Stir-fried Chinese cabbage with carrot and black fungus	20	16.85 (13.35)	17	18.18 (5.52)	.71
Stir-fried cauliflower with carrot and black fungus	20	24.90 (20.60)	17	21.65 (5.25)	.53
Stir-fried bean sprout with carrot and black fungus	4	9.00 (6.22)	6	18.87 (7.91)	.07

[a]AIR: automatic image-based reporting.

[b]VIR: voice input reporting.

[c]Not applicable.

XSL•FO
RenderX

In the 3 food categories, that is, the staple food, main course, and dishes with 1 ingredient, the AIR group required significantly less time than the VIR group. Reporting performance for the 2 food categories did not differ significantly for dishes with 2 or 3 ingredients.

Within the 6 dishes in the categories of staple food and main course, except for "braised pork chop," the other 5 dishes in the AIR required significantly less time than the VIR group. Within the five 1-ingredient dishes, the AIR group required significantly less time to report "stir-fried cauliflower" ($P<.001$) and was faster for "stir-fried bitter melon" ($P<.05$). For dishes with 2 ingredients, the AIR group required significantly less time in "stir-fried egg with tomato" ($P<.05$). For 3 dishes with 3 ingredients, the 2 groups did not reveal significant differences.

SUS and Subjective Perception

Table 4 summarizes the SUS score and its 2 divisions in terms of usability and learnability. Overall scores showed no significant differences between the AIR and VIR groups, but both groups had the overall score at 84.72 and 83, respectively, indicating that the participants in both groups considered the app to be generally easy-to-use and easy-to-learn.

Table . System usability scale of the AIR[a] and VIR[b] groups.

Score[c,d,e]	Assessment score, mean (SD)		*P* value
	AIR (n=22)	VIR (n=20)	
Overall score	84.72 (12.66)	83.00 (11.95)	.66
Usability score	84.43 (11.34)	83.50 (9.30)	.77
Learnability score	85.00 (15.66)	82.50 (16.18)	.61

[a]AIR: automatic image-based reporting.

[b]VIR: voice input reporting.

[c] The questionnaire was presented in Chinese.

[d]Mean scores for system usability with adjective ratings are as follows: 35.7 (poor), 50.9 (ok), 71.4 (good), and 85.5 (excellent).

[e]The questionnaire's Cronbach α for AIR (α=0.91) and VIR (α=0.89) exceeded 0.70, indicating good internal consistency.

Discussion

Principal Findings

While automatic image recognition is increasingly integrated into mobile meal reporting [28,38,39], few comparative studies have examined how this technology creates additional value for existing apps. Unlike previous studies, such as MyDietCam [26] and other existing apps [40], which primarily assessed usability through user-reported opinions or heuristic evaluation methods without using comparative experimental designs, our study used a randomized controlled trial to quantitatively evaluate differences in accuracy and time efficiency between automatic image recognition combined with voice input and voice-only reporting methods. Additional image capture techniques include wearable sensors. For example, Hussain et al [38] developed a system using a wearable camera and convolutional neural networks for food recognition.

In this study, we included VIR as the control condition because it represented the previous version of our app and served as the existing standard for comparison. At the time the study was designed, it was not known how VIR would perform relative to the newly developed AIR version. Our goal was to quantify the added value of the image recognition feature by comparing AIR against a practical and relevant baseline, namely, VIR. The results of this study indicate that combining automatic image and voice recognition is not only feasible but also provides improvements over voice-only versions in terms of accuracy and time efficiency. Our evidence-based findings provide insight for researchers or practitioners developing the next generation of dietary intake reporting applications, with implications as follows.

Accuracy in AIR Versus VIR

Correctly identifying and reporting the dish names and cooking methods is crucial for accurate daily intake management. Table 2 shows that user performance with AIR is significantly more accurate than with VIR due to the integration of automatic image recognition. However, in the AIR group, the dishes that had been correctly recognized remain limited. Furthermore, the recognition accuracy rate of each dish for the 5 food categories fluctuated significantly within the AIR group. Image recognition errors in the trial may have resulted from the AI server failing to adequately recognize the dish or providing incorrect possible answers. In addition, errors could occur when participants were required to use voice input to complete the reporting task when the image recognition did not work properly for some dishes. Failed image recognition could also occur due to poor image quality in the uploaded file, inappropriate lighting, the technical algorithm, food recognition technologies [31,41], limited food datasets [25,26], and the contexts of use [42].

Time Efficiency in AIR Versus VIR

Task completion for the AIR group was significantly faster than for the VIR group, suggesting that the integration of automatic image recognition can effectively reduce reporting time, though such performance improvements were inconsistent across food categories, and further enhancements are needed.

Participant Perception

Overall, SUS results indicated participants found both apps to be easy to use. Regardless of the differences in accuracy rate and time difference achieved by the 2 groups, the app in each of the groups demonstrated high overall SUS scores.

XSL•FO
RenderX

Limitations and Future Research

The study design assumed that users would take a photo, upload it, and report the meal in order to complete the meal reporting task. However, there are potentially other reporting scenarios to be considered in future research. For example, users may forget to take a photo before eating, or they may consume meals that are the same as or similar to previous ones. Another scenario could involve users completing the report at a later time. Adding features to the app—such as allowing users to select a previously recorded meal—could enhance flexibility and support these situations. These cases could affect the performance of automatic image recognition and warrant further investigation.

Regarding the sample size, this study was conducted during a period of severe COVID-19 restrictions, and the university was under semilockdown, making it challenging to recruit additional participants. Based on the same parameters used for the original sample size calculation, the final sample of 42 young participants yielded an estimated statistical power of 72%, which is slightly below the conventional threshold of 80%. Even in this context, the primary outcomes—accuracy and time—still showed statistically significant differences. However, these results should be interpreted with caution, as the study may have been underpowered for these measures. To generalize the findings, further confirmation with a larger and more diverse participant sample is recommended. Future research should also include a broader range of authentic dishes and longer reporting periods.

Conclusions

Integrating AI image recognition in a voice-based meal reporting application was found to significantly improve reporting accuracy and time efficiency among young adult users. Further design improvements are required, as is testing in a broader range of authentic dining environments and a more varied array of food items.

Acknowledgments

The authors are thankful for all the research assistants at the Food Interaction Design Lab and the ABC Lab of Chang Gung University. This research was funded by the Research Fund of Chang Gung Memorial Hospital and Chang Gung University (BMRPD 067) and the National Science and Technology Council, Taiwan (NSTC 112-2221-E-182-048-MY2 and NSTC 113-2221-E-182-028).

Conflicts of Interest

None declared.

Checklist 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File, 1105 KB - mhealth_v13i1e60070_app1.pdf]

References

1. Boutari C, Mantzoros CS. A 2022 update on the epidemiology of obesity and a call to action: as its twin COVID-19 pandemic appears to be receding, the obesity and dysmetabolism pandemic continues to rage on. Metab Clin Exp 2022 Aug;133:155217. [doi: 10.1016/j.metabol.2022.155217] [Medline: 35584732]
2. Haththotuwa RN, Wijeyaratne CN, Senarath U. Worldwide epidemic of obesity. In: Obesity and Obstetrics: Elsevier; 2020:3-8. [doi: 10.1016/B978-0-12-817921-5.00001-1]
3. Chaudhary A, Sudzina F, Mikkelsen BE. Promoting healthy eating among young people—a review of the evidence of the impact of school-based interventions. Nutrients 2020 Sep 22;12(9):2894. [doi: 10.3390/nu12092894] [Medline: 32971883]
4. Winpenny EM, van Sluijs EMF, White M, Klepp KI, Wold B, Lien N. Changes in diet through adolescence and early adulthood: longitudinal trajectories and association with key life transitions. Int J Behav Nutr Phys Act 2018 Sep 10;15(1):86. [doi: 10.1186/s12966-018-0719-8] [Medline: 30200990]
5. Montero K, Kelly P. Young People and the Aesthetics of Health Promotion: Beyond Reason, Rationality and Risk: Routledge; 2016.
6. Wickman ME, Anderson NLR, Greenberg CS. The adolescent perception of invincibility and its influence on teen acceptance of health promotion strategies. J Pediatr Nurs 2008 Dec;23(6):460-468. [doi: 10.1016/j.pedn.2008.02.003] [Medline: 19026914]
7. Mattei J, Alfonso C. Strategies for healthy eating promotion and behavioral change perceived as effective by nutrition professionals: a mixed-methods study. Front Nutr 2020;7:114. [doi: 10.3389/fnut.2020.00114] [Medline: 32923451]
8. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317. [doi: 10.2196/jmir.5927] [Medline: 27986647]
9. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012 May 14;14(3):e72. [doi: 10.2196/jmir.1977] [Medline: 22584372]

XSL•FO
RenderX

10. Chen Y, Perez-Cueto FJA, Giboreau A, Mavridis I, Hartwell H. The promotion of eating behaviour change through digital interventions. Int J Environ Res Public Health 2020;17(20):7488. [doi: 10.3390/ijerph17207488]
11. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98. [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
12. Nikolaou CK, Tay Z, Leu J, et al. Young people's attitudes and motivations toward social media and mobile apps for weight control: mixed methods study. JMIR mHealth uHealth 2019 Oct 10;7(10):e11205. [doi: 10.2196/11205] [Medline: 31603431]
13. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86. [doi: 10.2196/jmir.2583] [Medline: 23598614]
14. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence–empowered mHealth: scoping review. JMIR mHealth uHealth 2022 Jun 9;10(6):e35053. [doi: 10.2196/35053] [Medline: 35679107]
15. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare: Elsevier; 2020:25-60. [doi: 10.1016/B978-0-12-818438-7.00002-2]
16. Torbjørnsen A, Ribu L, Rønnevig M, Grøttland A, Helseth S. Users' acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. BMC Health Serv Res 2019 Sep 6;19(1):641. [doi: 10.1186/s12913-019-4486-2] [Medline: 31492176]
17. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]
18. Krishnan G, Singh S, Pathania M, et al. Artificial intelligence in clinical medicine: catalyzing a sustainable global healthcare paradigm. Front Artif Intell 2023;6:1227091. [doi: 10.3389/frai.2023.1227091] [Medline: 37705603]
19. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023 Sep 22;23(1):689. [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
20. Thomas LB, Mastorides SM, Viswanadhan NA, Jakey CE, Borkowski AA. Artificial intelligence: review of current and future applications in medicine. Fed Pract 2021 Nov;38(11):527-538. [doi: 10.12788/fp.0174] [Medline: 35136337]
21. Oh S, Kim JH, Choi SW, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: an online mobile survey. J Med Internet Res 2019 Mar 25;21(3):e12422. [doi: 10.2196/12422] [Medline: 30907742]
22. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158. [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
23. Bergevi J, Andermo S, Woldamanuel Y, Johansson UB, Hagströmer M, Rossen J. User perceptions of eHealth and mHealth services promoting physical activity and healthy diets: systematic review. JMIR Hum Factors 2022 Jun 28;9(2):e34278. [doi: 10.2196/34278] [Medline: 35763339]
24. Almiron-Roig E, Navas-Carretero S, Emery P, Martínez JA. Research into food portion size: methodological aspects and applications. Food Funct 2018 Feb 21;9(2):715-739. [doi: 10.1039/c7fo01430a] [Medline: 29219156]
25. Howes E, Boushey CJ, Kerr DA, Tomayko EJ, Cluskey M. Image-based dietary assessment ability of dietetics students and interns. Nutrients 2017 Feb 7;9(2):114. [doi: 10.3390/nu9020114] [Medline: 28178196]
26. Kong NA, Moy FM, Ong SH, Tahir GA, Loo CK. MyDietCam: development and usability study of a food recognition integrated dietary monitoring smartphone application. Digit Health 2023;9. [doi: 10.1177/20552076221149320] [Medline: 36644664]
27. Kaushal S, Tammineni DK, Rana P, Sharma M, Sridhar K, Chen HH. Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: new insights and advances. Trends Food Sci Technol 2024 Apr;146:104408. [doi: 10.1016/j.tifs.2024.104408]
28. Amugongo LM, Kriebitz A, Boch A, Lütge C. Mobile computer vision-based applications for food recognition and volume and calorific estimation: a systematic review. Healthcare (Basel) 2023;11(1):59. [doi: 10.3390/healthcare11010059]
29. Van Asbroeck S, Matthys C. Use of different food image recognition platforms in dietary assessment: comparison study. JMIR Form Res 2020 Dec 7;4(12):e15602. [doi: 10.2196/15602] [Medline: 33284118]
30. Rantala E, Balatsas-Lekkas A, Sozer N, Pennanen K. Overview of objective measurement technologies for nutrition research, food-related consumer and marketing research. Trends Food Sci Technol 2022 Jul;125:100-113. [doi: 10.1016/j.tifs.2022.05.006]
31. Liu YC, Onthoni DD, Mohapatra S, Irianti D, Sahoo PK. Deep-learning-assisted multi-dish food recognition application for dietary intake reporting. Electronics (Basel) 2022;11(10):1626. [doi: 10.3390/electronics11101626]
32. Liu YC, Chen CH, Lin YS, et al. Design and usability evaluation of mobile voice-added food reporting for elderly people: randomized controlled trial. JMIR mHealth uHealth 2020 Sep 28;8(9):e20317. [doi: 10.2196/20317] [Medline: 32985999]
33. Liu YC, Wu ST, Lin SJ, Chen CH, Lin YS, Chen HY. Usability of food size aids in mobile dietary reporting apps for young adults: randomized controlled trial. JMIR mHealth uHealth 2020 Apr 29;8(4):e14543. [doi: 10.2196/14543] [Medline: 32347805]
34. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]

XSL•FO
RenderX

35. Liu YC, Chen CH, Tsou YC, et al. Evaluating mobile health apps for customized dietary recording for young adults and seniors: randomized controlled trial. JMIR mHealth uHealth 2019 Feb 15;7(2):e10931. [doi: 10.2196/10931] [Medline: 30767906]
36. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:189-194. [doi: 10.1201/9781498710411]
37. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J User Experience 2009 May 1;4(3):114-123 [FREE Full text]
38. Hussain G, Maheshwari MK, Memon ML, Jabbar MS, Javed K. A CNN based automated activity and food recognition using wearable sensor for preventive healthcare. Electronics (Basel) 2019;8(12):1425. [doi: 10.3390/electronics8121425]
39. Fakhrou A, Kunhoth J, Al Maadeed S. Smartphone-based food recognition system using multiple deep CNN models. Multimed Tools Appl 2021 Sep;80(21-23):33011-33032. [doi: 10.1007/s11042-021-11329-6]
40. Li X, Yin A, Choi HY, Chan V, Allman-Farinelli M, Chen J. Evaluating the quality and comparative validity of manual food logging and artificial intelligence-enabled food image recognition in apps for nutrition care. Nutrients 2024 Aug 5;16(15):2573. [doi: 10.3390/nu16152573] [Medline: 39125452]
41. Boushey CJ, Spoden M, Zhu FM, Delp EJ, Kerr DA. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. Proc Nutr Soc 2017 Aug;76(3):283-294. [doi: 10.1017/S0029665116002913] [Medline: 27938425]
42. Papathanail I, Rahman LA, Brigato L, et al. The nutritional content of meal images in free-living conditions-automatic assessment with goFOOD. Nutrients 2023 Sep 2;15(17):3835. [doi: 10.3390/nu15173835] [Medline: 37686866]

Abbreviations

AI: artificial intelligence
AIR: automatic image-based reporting
mHealth: mobile health
SUS: system usability scale
VIR: voice input reporting

Edited by L Buis; submitted 21.05.24; peer-reviewed by AK Vadathya, KH Tan, RT Khomeiran; revised version received 21.04.25; accepted 06.06.25; published 14.08.25.

Please cite as:
Sahoo PK, Chiu SYH, Lin YS, Chen CH, Irianti D, Chen HY, Sarkar M, Liu YC
Automatic Image Recognition Meal Reporting Among Young Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e60070
URL: https://mhealth.jmir.org/2025/1/e60070
doi:10.2196/60070

© Prasan Kumar Sahoo, Sherry Yueh-Hsia Chiu, Yu-Sheng Lin, Chien-Hung Chen, Denisa Irianti, Hsin-Yun Chen, Mekhla Sarkar, Ying-Chieh Liu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Utility of Digital Voice Assistants for Primary Care Patients, Including Those With Physical and Visual Disabilities: Cross-Sectional Study

Maya Rajan[1], BA; Allison Furgal[2], PhD; Reema Kadri[1], MLIS; Omar Arman[1], MD; Kate Panzer[1,3], MD; Donna Wicker[4,5], OD; Michael M McKee[1], MD, MPH; Melissa Plegue[6], MA; Alexandria Degner[1], MA; Lorraine R Buis[1], PhD

[1]Department of Family Medicine, University of Michigan, 1018 Fuller Street, Ann Arbor, MI, United States

[2]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States

[3]Department of Orthopedic Surgery, University of Virginia, Charlottesville, VA, United States

[4]Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, MI, United States

[5]Detroit Mercy Eye Institute, University of Detroit Mercy, Novi, MI, United States

[6]Department of Pediatrics, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Lorraine R Buis, PhD
Department of Family Medicine, University of Michigan, 1018 Fuller Street, Ann Arbor, MI, United States

Abstract

Background: Today, most smartphones provide a digital voice assistant (DVA) for their user, and it is estimated that about 91% of adults report owning and operating a smartphone. A DVA is an automated system preinstalled on technological devices, such as smartphones, computers, tablets, and speakers, which serves to aid users in performing tasks like answering questions, managing smart devices at home, playing music, managing schedules, sending messages, and more. Research with DVA is emerging, and its applicability to health and health care needs to be elucidated.

Objective: The objective of this study was to describe the use of DVAs among primary care patients, as well as purposely sampled clinics including patients with visual and physical disabilities.

Methods: A convenience sample of adult participants was recruited to complete a needs assessment survey to ascertain the interest and possible utility of DVAs to promote and enhance health from among three populations at an academic medical center: (1) general primary care patients recruited from a primary care clinic, (2) patients with visual disabilities recruited from a low vision clinic, and (3) patients with physical disabilities recruited from a physical medicine and rehabilitation clinic. The survey used in this study was a 46-item investigator-developed instrument administered to participants assessing knowledge, use, and perceptions of DVAs, participant interest to participate in related future studies, and demographics.

Results: The results of the survey showed that the majority of participants have used a DVA before (69.7%, 152/218) and were or might be willing to use them in the future (84.0%, 178/212). Participants reported moderate to high concern about the privacy (47.8%, 97/203), security (54.5%, 110/202), and confidentiality (51.7%, 105/203) of DVAs. A greater proportion of those with visual disabilities reported having never used DVA than those without visual disabilities (39.0% vs 24.6%, *P*=.03). There was no significant difference in reliance on DVAs for participants with and those without physical disabilities (45.0% vs 34.9%, *P*=.31), indicating that they do not require it for everyday needs.

Conclusions: DVA use remains low among the surveyed participants with physical and visual disabilities. For those with visual disabilities, DVA use was seen to be advantageous in everyday life for tasks such as answering questions and seeking information, but not for those with physical disabilities. However, further research should be conducted that focuses on the use of DVAs by accessing data that represent an individual's DVA use without being biased by knowledge of a research study. In addition, research is needed on DVA use that includes diverse samples of participants with physical and visual disabilities, which address the barriers to using DVAs for these adult populations.

(JMIR Mhealth Uhealth 2025;13:e66185) doi:10.2196/66185

KEYWORDS

smartphone; mobile applications; disabled persons; visually impaired persons; primary health care; digital voice assistant

XSL•FO
RenderX

Introduction

Currently, about 91% of American adults report owning and operating a smartphone [1] and the overwhelming majority of those smartphones provide users with a digital voice assistant (DVA). DVAs are software that comes preinstalled on many different mobile devices, such as smartphones, computers, tablets, and laptops, or that operates as a stand-alone device such as a smart speaker. DVAs aid users in performing tasks like answering questions, managing smart devices in the home, playing music, managing schedules, sending messages, and more. To recognize and execute these tasks, DVAs communicate with the user in their natural language by using voice recognition while using either native or third-party services to obtain the requested information or perform the task at hand. The first DVA, Siri, was invented in 2011 by Apple and additional companies brought their own DVAs to market in the following years. DVAs not only exist on smartphones, but also within other smart devices such as speakers, TVs, and other various automated home control systems [2]. Studies investigating the use and quality of these smart home devices have shown that they increase convenience, reduce loneliness, and improve quality of life in older adults and adults with disabilities [3]. However, barriers exist in installing, troubleshooting, and updating these technologies for long term use [4]. The most prevalent examples of DVAs in the United States across smartphones and home devices include Siri by Apple, Alexa by Amazon, and Google Assistant [5]. DVAs have experienced explosive growth within the past few years, with recent estimates indicating 62% of American adults having used one on a device [6].

In the past, most research involving DVAs has been narrowly focused and limited to economic analyses focused on potential market value [5], as well as consumer perceptions of privacy and security in regards to DVAs [7]. DVA applications within health care are becoming more widely commercially available and have been primarily focused on demonstrative projects for memory support in patients with dementia [8], older adults with diabetes [9], and communicative apps and technologies for individuals with intellectual and developmental disability or autism spectrum disorder [10]. DVAs are ripe for study within health care contexts, and they pose a particularly exciting innovation in the realm of assistive technologies for older populations, as well as those with various disabilities. Indeed, the potential for DVAs to aid older adults is becoming more clear and emerging research, especially during the COVID-19 pandemic, has investigated the ability of DVAs to reduce loneliness and social isolation of older adults [11].

Aside from assisting with disabilities, older populations engage with DVAs in more relational ways, such as through asking the DVA “personal questions” to get to know them, asking for advice, and engaging with the DVA to alleviate stress [12]. Indeed, as DVAs are typically controlled with voice activation and voice prompts, they may be more easily used by populations who have physical or visual disabilities that result in difficulties manipulating or navigating their mobile devices. DVAs have the potential to deliver interventions specifically for patient populations with a broad range of physical and visual disabilities, both of which can make using mobile devices challenging. For individuals whose physical disabilities affect their endurance, mobility, or dexterity, DVAs could enhance their ability to complete daily tasks and acquire information by mitigating detrimental exertion. Likewise, those with visual disabilities may benefit from DVAs ability to read information out loud instead of being required to visually read online information.

DVAs have great potential for use in health care and could potentially facilitate clearer communication between patients and providers through language translation capabilities, hands-free calling and messaging, and reminders for medications and rehabilitation exercises [13]. For those with visual disabilities, as well as physical disabilities that prevent one from holding a device, DVAs could be transformative in bridging their connection to online health information, family members, and health care providers [14]. The technology can provide greater independence and manage safety risks for those who fall under these specific populations [15]. Given the widespread penetration of DVAs, coupled with their potential for use among populations who may struggle with traditional mobile devices, research to explore potential applications for use within these contexts is needed. In addition, recent and ongoing research exists to evaluate the accuracy and quality of health information that DVAs provide to users [16]. As a first step at understanding the potential for greater use within health care settings, it is integral to investigate the behaviors and opinions on DVAs and barriers to access and use of this technology.

To better understand the potential for use in various health care settings and populations, we sought to describe general DVA use among a broad range of adult patients including primary care patients, as well as those seeking care from specialty clinics where there are high concentrations of patients with either visual or physical disabilities. The results of this study carry importance to inform the future innovations in DVA technology. Depending on the use type, or lack thereof, these tools can be better designed to address the needs of these populations and aid them in their daily functioning.

Methods

Participants and Setting

A cross-sectional study using a convenience sample of adult participants was used to survey the use and perceptions of DVAs from patients at Michigan Medicine, the academic medical center affiliated with the University of Michigan. Patients from 3 purposely sampled clinics in a single city were recruited. The clinics included a family medicine (Briarwood Family Medicine Center), low vision (Kellogg Eye Center Low Vision and Rehabilitation Clinic), and physical medicine & rehabilitation (PM&R) clinic at Michigan Medicine. The family medicine clinic was selected to reflect a typical primary care setting with a range of patients. The low vision clinic was selected for its access to patients with visual disabilities. Finally, the PM&R clinic was selected to enrich the sample for individuals who had physical-related disabilities. Potential participants younger than 18 years were excluded from participation.

XSL•FO
RenderX

Participant Recruitment

Potential participants at the Kellogg and PM&R clinics were informed of the study by clinic staff members during check in. If interested, participants were given a paper survey to complete. Completed surveys were returned to clinic staff members. In addition, participants were recruited at the monthly low-vision support group offered at the Kellogg Eye Center. At the family medicine clinic, study staff members were posted in the lobby and informed potential participants of the study. Those interested in participating were handed a survey, which was returned to study staff at completion. Recruitment occurred over 19 months from September 2019 to March 2021, with recruitment delays experienced due to the COVID-19 pandemic.

Procedures

At each of the 3 clinic sites, an information sheet explaining the purpose of the survey, as well as a self-administered paper copy of the survey, were given to potential participants by clinic or study staff members. Participants then had the option to fill out the survey at any point during their clinic visit and return the paper survey to the clinic or study staff members upon completion. The assessment survey was self-administered by the respondent; however, in cases where assistance was needed for survey completion, participants could select a friend or caregiver accompanying them for assistance. In the case of participants recruited from the low vision clinic, clinic staff members were available to help respondents complete the survey.

Survey Instrument

The study survey used was a 46-item investigator-developed paper-based instrument that surveyed participants about their knowledge, use, and perceptions of DVAs, their interest to participate in related future studies, and demographic questions (Multimedia Appendix 1). Questions regarding knowledge and use of DVAs were presented in 6 closed, multiple-choice format questions with some including an “other” option for further elaboration. There was one additional open-ended question that asked participants to give examples of potential use cases of DVAs that aligned with their interests. Participants were also asked to rate their concerns about 5 specific issues related to the use of DVAs, and one opportunity to voice their own concern. These ratings used a 5-point Likert scale (0=not concerned at all; 4=highly concerned) and were used to measure concerns relating to privacy, security, confidentiality, accuracy, and reliability of the DVAs. The questions about privacy, security, and confidentiality included definitions of the respective terms to clearly define the specific concern. “Privacy” referred to freedom to opt out of monitoring and to determine how personal data is used. “Security” referred to protection of personal data from loss or theft. “Confidentiality” was defined as the obligation of trusted persons to keep data safe and held in confidence. To assess participants’ interest in participating in future studies relating to DVAs, 3 closed multiple choice questions were asked.

The next set of questions presented assessed whether participants had a visual disability. If participants endorsed having a visual disability, 9 additional questions were presented asking about the nature of their disability, access to care, and access to resources. These questions were presented in mixed format. Participants who did not have visual disabilities were asked to skip to the next section, which was focused on physical disability.

The physical disability portion of the survey asked questions generically about disability, which allowed participants to self-identify and describe the nature of their disability. Those endorsing no physical disability were asked to skip to the final set of questions on demographics; whereas those who did endorse a physical disability were asked mixed format questions about the nature of their disability, access to care, and access to resources. As participants were allowed to self-identify with having a physical disability, this study included a wide range of physical disabilities and documented disability within the electronic health record was not required by any participants; rather, participants were asked to endorse whether they had any kind of physical disability and if so, questions probed into the nature of their self-reported disability.

Given that participants with visual disabilities were expected to have more difficulty taking the survey and would require assistance, a shortened 34-item survey presented in large print was administered to participants recruited from the Kellogg low vision clinic (Multimedia Appendix 2). The 34-item survey contained the same questions as the aforementioned 46-item instrument, except it excluded the question set focused on physical disabilities. Both versions of the survey were designed to be completely anonymous, with no protected health information or HIPAA (Health Insurance Portability and Accountability Act)-protected data collected. See Multimedia Appendices 1 and 2 for the survey instrument.

Participants gave their completed surveys to study or clinic staff members, who then placed the surveys into a box clearly labeled with the study title. After the collection of the completed surveys, the data were entered into a secure web application REDCap (Research Electronic Data Capture; Vanderbilt University) by approved study personnel. The entered data were verified through a 10% random sampling for entry errors. Incomplete survey responses were also retained and collected during this study.

Statistical Analysis

The survey data were analyzed using descriptive statistics including frequencies, means, and SD. To determine differences in responses between subgroups, bivariate associations between subgroups and demographics, DVA use, and DVA concerns were performed using chi-square tests or Fisher exact tests as appropriate depending on the numbers in each group. Differences were examined by previous DVA use, clinic site, visual disability status, physical disability status, and willingness to participate in future research studies. All analyses were performed using SAS Software (version 9.4; SAS Institute Inc).

Ethical Considerations

Our study was registered under a self-exemption process at the University of Michigan, classified under EXEMPTION 2 (i) and 2 (ii) as described in 45 CFR 46.104(d)(IRBMED HUM00156811). Participants were handed an information sheet

XSL•FO
RenderX

outlining the risks and benefits of the study before completing the survey. All study materials were stored in locked file cabinets in the locked office of the principal investigator. All data were collected anonymously and reported in the aggregate, and participant responses cannot be linked to individual participants. No financial or nonfinancial incentives were offered for study participation.

Results

Demographics

A total of 218 eligible participants completed the survey. The majority of the sample was White (71.6%, 156/218) and female (61.7%, 124/201) with an average age of 52.6 (SD 18.9) years. The education levels of the participants varied, with the largest portion having a graduate degree (35.0%, 71/203), followed by a split between those who have completed a bachelor's degree (20.2%, 41/203) and those who have completed some college education (19.7%, 40/203). Participants' employment status was primarily full-time (40.8%, 82/201) followed by retired (29.9%, 60/201). Roughly half of the participant population earned an annual income of US $75,000 or higher (46.3%, 88/190) and was insured through their employer (54.1%, 118/218). Almost half of the participant population (49.5%, 101/204) indicated interest in participating in an additional research study that provides DVAs for use (Table 1).

XSL•FO
RenderX

Table . Demographics of all study participants.

Variable	Values
Age (years; n=217), mean (SD)	52.6 (18.9)
Site (n=218), n (%)	
Physical Medicine and Rehabilitation	84 (38.5)
Briarwood Family Medicine	80 (36.7)
Kellogg Eye Center	54 (24.8)
Gender (n=201), n (%)	
Male	76 (37.8)
Female	124 (61.7)
Other	1 (0.5)
Race[a] (n=218), n (%)	
White	156 (71.6)
Black or African American	34 (15.6)
Asian	10 (4.6)
American Indian or Alaska Native	3 (1.4)
Native Hawaiian or other Pacific Islander	0 (0)
Other	8 (3.7)
Not reported	13 (6.0)
Hispanic, Latino, or Spanish origin (n=199), n (%)	
Yes	14 (7.0)
No	185 (93.0)
Living situation[a] (n=218), n (%)	
Live alone	43 (19.7)
With spouse or other companion	121 (55.5)
With adult children	24 (11.0)
With young children	28 (12.8)
With siblings, parents, or other guardian	21 (9.6)
Other	11 (5.0)
Not reported	12 (5.5)
Income (n=190), n (%)	
Under US $25,000	31 (16.3)
US $25,000 - US $49,000	23 (12.1)
US $50,000 - US $74,999	33 (17.4)
US $75,000 or higher	88 (46.3)
Do not know	15 (7.9)
Insurance[a] (n=218), n (%)	
Insured through an employer	118 (54.1)
Insurance purchased directly from company	18 (8.3)
Medicare	71 (32.6)
Medicaid	25 (11.5)
No insurance	2 (0.9)
Other	11 (5.0)

XSL•FO
RenderX

Variable	Values
Not reported	16 (7.3)
Education (n=203), n (%)	
High school diploma or less	33 (16.3)
Some college	40 (19.7)
Associate's degree, trade school or apprenticeship	18 (8.9)
Bachelor's degree	41 (20.2)
Graduate degree	71 (35.0)
Employment status (n=201), n (%)	
Employed part-time	17 (8.5)
Employed full-time	82 (40.8)
Retired	60 (29.9)
Currently on disability	26 (12.9)
Laid off or unemployed	16 (8.0)
General health (n=206), n (%)	
Poor	11 (5.3)
Fair	40 (19.4)
Good	98 (47.6)
Very good	50 (24.3)
Excellent	7 (3.4)
Previously Used DVA[b] (n=218), n (%)	
Yes	152 (69.7)
No	66 (30.3)
Willing to use DVA in the future (n=212), n (%)	
Yes	131 (61.8)
Maybe	47 (22.2)
No	34 (16.0)
Has any vision-related disabilities (n=211), n (%)	
Yes	77 (36.5)
No	134 (63.5)
Has any physical disabilities[c] (n=156), n (%)	
Yes	47 (30.1)
No	109 (69.9)

[a]Participants could select multiple categories in these questions resulting in percentages that may add to more than 100%.

[b]DVA: digital voice assistant.

[c]Questions about physical difficulties were not asked at the Kellogg site. Number missing reflects this.

General Use of DVAs

The results of the survey showed that DVAs are being used for various tasks and there is a willingness to use them in the future. The majority of all participants had previously used a DVA (69.7%, 152/218) and were or might be willing to use DVAs in the future (84.0%, 178/212) including 57.1% (36/63) of participants who had never used a DVA before (Table 1). Among the 152 participants who had used a DVA, Siri was the most highly used DVA (67.8%, 103/152). Other commonly used DVAs were Alexa and Google Assistant which were both used similarly by about nearly half (70/152 and 71/152, respectively) of participants. Of those who did report DVA use, the majority accessed the tool from their personal cell phone (88.8%, 135/152), and the main task was to answer questions (75.0% 114/152), with entertainment (57.2%, 87/152), accessing information online (55.9%, 85/152), and setting timers (51.3%, 78/152) also being frequent tasks done with a DVA. While many of the participants have used DVAs, the frequency of this use was split between those who use it rarely or sometimes

XSL•FO
RenderX

(57.4%, 85/148) and those who use it often or always (42.6%, 63/148). Although the participants use DVAs, the majority did not rely on the DVA for everyday needs (57.0%, 85/149; Table 2).

Table . Digital voice assistant use characteristics among those who had previously used a digital voice assistant.

Variable	Participants, n (%)
DVA[a] use (n=152)	
Alexa or Amazon Echo	70 (46.1)
Cortana	17 (11.2)
Siri	103 (67.8)
Google assistant	71 (46.7)
Google now	15 (9.9)
Other	6 (3.9)
None selected	0 (0)
DVA device (n=152)	
Personal cell phone	135 (88.8)
Personal Google home or Google Home Mini	23 (15.1)
Personal Echo or Echo Dot	43 (28.3)
Personal iPad or iPod	40 (26.3)
Personal tablet, laptop, or desktop computer	47 (30.9)
Device owned by another person	16 (10.5)
Other	9 (5.9)
None selected	2 (1.3)
DVA[a] tasks (n=152)	
Answering questions	114 (75.0)
Managing schedules or calendars	61 (40.1)
Setting timers	78 (51.3)
Accessing information online	85 (55.9)
Entertainment	87 (57.2)
Other	25 (16.4)
None selected	1 (0.7)
DVA use frequency (n=148)	
Rarely	47 (31.8)
Sometimes	38 (25.7)
Often	40 (27.0)
Always	23 (15.5)
Rely on DVA for everyday needs, (n=149)	
Yes	41 (27.5)
Sometimes	23 (15.4)
No	85 (57.0)

[a]DVA: digital voice assistant; participants could select multiple categories in these questions resulting in percentages that may add to more than 100%.

DVA Use and Visual Disabilities

Among respondents, 36.5% (77/211) reported having visual disabilities. Of the participants who reported having any visual disability, 81.6% (62/76) found it mildly to extremely difficult to perform activities because of their vision impairment and 41.9% (31/74) reported having vision-related needs not being met. Only 38.2% (26/68) of participants with visual disabilities indicated being very or highly satisfied with current services and devices, revealing a crucial need for improvement.

XSL•FO
RenderX

A greater proportion of participants with visual disabilities reported having never used DVA (39.0%, 30/77) than those without visual disabilities (24.6%, 33/134; *P*=.03). Most participants with visual disabilities who had used a DVA did so on their personal cell phone (53.2%, 41/77) but this was significantly less than the proportion of those without visual disabilities who use DVAs on cell phones (68.7%, 92/134) (*P*=.03). As in the entire sample, the most common DVA task was answering questions regardless of visual disability status, but this task was done by fewer people with visual disabilities (40.3%, 31/77) than those without visual disabilities (60.4%, 81/134; *P*=.005). Those with visual disabilities indicated relying on DVA sometimes or always (63.8%, 30/47), more often than those without visual disabilities (32.7%, 33/101; *P*=.001). In addition, participants with visual disabilities were equally as willing to potentially use DVAs in the future (85.3%, 64/75) as those without visual disabilities (85.0%, 113/133; *P*=.78). Within the group of participants who had never used a DVA, those with and those without visual disabilities had similar willingness to potentially use DVAs in the future (62.1%, 18/29 vs 56.3%, 18/32, respectively; *P*=.64). Those with visual disabilities were also more willing to grant access to past data for DVA studies than those without visual disabilities (22.5%, 16/71 vs 13.1%, 17/130; *P*=.01; Table 3).

XSL•FO
RenderX

Table . Digital voice assistant use characteristics by disability status. *P* values are derived from chi-square tests unless otherwise noted.

	Visual disabilities			Physical disabilities		
	No (N=134), n (%)	Yes (N=77), n (%)	*P* value	No (N=109), n (%)	Yes (N=47), n (%)	*P* value
DVA[a] use						
Alexa or Amazon Echo	49 (36.6)	21 (27)	.17	46 (42.2)	10 (21)	.01
Cortana	12 (9.0)	5 (6)	.53	9 (8.3)	5 (11)	.76[b]
Siri	72 (53.7)	28 (36)	.02	63 (57.8)	20 (43)	.08
Google assistant	48 (35.8)	21 (27)	.20	39 (35.8)	17 (36)	.96
Google now	10 (7.5)	5 (6)	.79	12 (11.0)	1 (2)	.11[b]
Never used DVA	33 (24.6)	30 (39)	.03	23 (21.1)	16 (34)	.09
Other	1 (0.7)	5 (6)	.03[b]	1 (0.9)	0 (0)	.99[b]
None selected	0 (0)	1 (1)	.36[b]	0 (0)	0 (0)	N/A[c]
DVA device						
Personal cell phone	92 (68.7)	41 (53)	.03	80 (73.4)	27 (57)	.049
Personal Google home or Google home mini	15 (11.2)	8 (10)	.86	14 (12.8)	4 (9)	.44
Personal Echo or Echo Dot	31 (23.1)	12 (16)	.19	26 (23.9)	6 (13)	.12
Personal iPad or iPod	23 (17.2)	15 (19)	.67	23 (21.1)	6 (13)	.22
Personal tablet, laptop, or desktop computer	26 (19.4)	21 (27)	.19	21 (19.3)	7 (15)	.51
Device owned by another person	11 (8.2)	5 (6)	.65	9 (8.3)	3 (6)	.99[b]
Other	4 (3.0)	5 (6)	.29[b]	5 (4.6)	0 (0)	.32[b]
None selected	33 (24.6)	30 (39)	.03	23 (21.1)	16 (34)	.09
DVA tasks						
Answering questions	81 (60.4)	31 (40)	.005	68 (62.4)	28 (60)	.74
Managing schedules or calendars	41 (30.6)	19 (25)	.36	38 (34.9)	11 (23)	.16
Setting timers	54 (40.3)	23 (30)	.13	50 (45.9)	14 (30)	.06
Accessing information online	54 (40.3)	30 (39)	.85	47 (43.1)	20 (43)	.95
Entertainment	61 (45.5)	25 (32)	.06	54 (49.5)	19 (40)	.30
Other	13 (9.7)	12 (16)	.20	12 (11.0)	2 (4)	.23[b]
None selected	0 (0)	1 (1)	.37[b]	0 (0)	0 (0)	N/A[c]
DVA use frequency			.02			.30
Rarely or sometimes	64 (64.0)	20 (43)		53 (62.4)	16 (52)	
Often or always	36 (36.0)	26 (57)		32 (37.6)	15 (48)	
Not reported[d]	34 (25.4)	31 (40)		24 (22.0)	16 (34)	
Rely on DVA for everyday needs			.001			.08

XSL•FO
RenderX

	Visual disabilities			Physical disabilities		
	No (N=134), n (%)	Yes (N=77), n (%)	*P* value	No (N=109), n (%)	Yes (N=47), n (%)	*P* value
Yes	22 (21.8)	18 (38)		17 (19.8)	12 (39)	
Sometimes	11 (10.9)	12 (26)		13 (15.1)	2 (6)	
No	68 (67.3)	17 (36)		56 (65.1)	17 (55)	
Not reported[d]	33 (24.6)	30 (39)		23 (21.1)	16 (34)	
Willing to use DVA in future			.78			.82
Yes	81 (60.9)	49 (65)		70 (64.2)	27 (60)	
Maybe	32 (24.1)	15 (20)		24 (22.0)	12 (27)	
No	20 (15.0)	11 (15)		15 (13.8)	6 (13)	
Not reported[d]	1 (0.7)	2 (3)		0 (0)	2 (4)	
Willing to share existing DVA data			.01			.36
Yes	17 (13.1)	16 (23)		15 (14.2)	7 (15)	
Maybe	25 (19.2)	9 (13)		21 (19.8)	7 (15)	
No	55 (42.3)	17 (24)		47 (44.3)	16 (35)	
Do not have a DVA	33 (25.4)	29 (41)		23 (21.7)	16 (35)	
Not reported[d]	4 (3.0)	6 (8)		3 (2.8)	1 (2)	

[a]DVA: digital voice assistant participants; could select multiple categories in these questions resulting in percentages that may add to more than 100%. *P* values for these variables are testing the difference in proportion selecting "Yes" in an individual category by disability status.

[b]*P* value from the Fisher exact test.

[c]N/A: not applicable.

[d]These values were excluded from the chi-square tests, as well as the % calculations.

DVA Use and Physical Disabilities

Among respondents, 47 (30.1%, 47/156) reported having a physical disability. Those with physical disabilities were more likely to report also having visual disabilities (29.8%, 14/47) than those who did not report having any physical disabilities (9.2%, 10/109, respectively; *P*=.001). Among the participants who possessed physical disabilities, 63.0% (29/46) found it moderately, very, or extremely difficult to perform everyday activities. The most common activities affected by the physical disability were walking or climbing stairs (87.2%, 41/47) and holding objects in one's hands (31.9%, 15/47). Most participants (80.4%, 37/46) reported that their physical needs are being met. Still, only 55.6% (10/18) of participants with physical disabilities report being very or highly satisfied with the services or devices meeting their needs.

Participants with physical disabilities reported having never used DVA (34.0%, 16/47) more often than those without a physical disability (21.1%, 23/109), although the difference is not statistically significant (*P*=.09). The proportion of participants that used DVAs for certain tasks tended to be lower for those with physical disabilities than those without physical disabilities but again these differences were not statistically significant. There was a statistically significant difference (*P*=.01) in the use of Alexa or Amazon Echo for those with physical disabilities vs those without physical disabilities (21.3%, 10/47 vs 42.2%, 46/109, respectively). Similarly to those with visual disabilities, the most common task participants with physical disabilities used DVAs for was answering questions (59.6%, 28/47). There was no significant difference in the participants' reliance on DVA for everyday needs by physical disability status. Most participants with physical disabilities were willing, or possibly willing, to use DVA in the future (86.7%, 39/45). There is a significant difference of reliance on DVA with those without either disability being less reliant and those with visual or both visual and physical disabilities being more reliant (*P*=.01; Table 3).

Concerns With DVAs

Overall, a large proportion of participants had at least a slight concern about privacy (86.2%, 175/203), security (89.1%, 180/202), confidentiality (86.7% 176/203), accuracy of information (81.7%, 165/202), and reliability of information (85.1%, 171/201). More specifically, roughly half of the participants had moderate to high concerns about the privacy (47.8%, 97/203), security (54.5%, 110/202), and confidentiality (51.7%, 105/203) of DVAs. The degree of concern about privacy, security, and confidentiality had no significant differences between the participants with and those without visual disabilities. We found a significant difference between concerns about accuracy (*P*=.04) and reliability (*P*=.048) of information from DVAs based on visual disability status with participants with visual disabilities tending to be moderately or highly concerned more often. The degree of concern about privacy, security, confidentiality, accuracy, and reliability had

XSL•FO
RenderX

no significant differences between the participants with and those without physical disabilities. Despite these concerns, we generally did not see a significant decrease in DVA use, frequency, or reliance for those with higher levels of concern. The one exception was in those with high levels of concern about security who tended to use DVAs less frequently than those with slight or no concern about security (P=.04; Table 4).

XSL•FO
RenderX

Table . Perceptions of digital voice assistants overall and by disability statuses. *P* values are derived from chi-square tests.

Concerns	Overall (N=218), n (%)	Visual disabilities			Physical disabilities[a]		
		No (N=134), n (%)	Yes (N=77), n (%)	*P* value	No (N=109), n (%)	Yes (N=47), n (%)	*P* value
Privacy				.24			.71
Not concerned at all	28 (13.8)	14 (10.8)	14 (19)		9 (8.5)	4 (9)	
Slightly or somewhat concerned	78 (38.4)	51 (39.2)	27 (37)		42 (39.6)	15 (33)	
Moderately or highly concerned	97 (47.8)	65 (50.0)	32 (44)		55 (51.9)	27 (59)	
Not reported[b]	15 (6.9)	4 (3.0)	4 (5)		3 (2.8)	1 (2)	
Security				.22			.72
Not concerned at all	22 (10.9)	11 (8.5)	11 (15)		6 (5.7)	4 (9)	
Slightly or somewhat concerned	70 (34.7)	49 (38.0)	21 (29)		37 (35.2)	14 (30)	
Moderately or highly concerned	110 (54.5)	69 (53.5)	41 (56)		62 (59.0)	28 (61)	
Not reported[b]	16 (7.3)	5 (3.7)	4 (5)		4 (3.7)	1 (2)	
Confidentiality				.29			.54
Not concerned at all	27 (13.3)	14 (10.8)	13 (18)		8 (7.5)	6 (13)	
Slightly or somewhat concerned	71 (35.0)	49 (37.7)	22 (30)		37 (34.9)	14 (30)	
Moderately or highly concerned	105 (51.7)	67 (51.5)	38 (52)		61 (57.5)	26 (57)	
Not reported[b]	15 (6.9)	4 (3.0)	4 (5)		3 (2.8)	1 (2)	
Accuracy of information				.04			.21
Not concerned at all	37 (18.3)	21 (16.2)	16 (22)		14 (13.2)	9 (20)	
Slightly or somewhat concerned	83 (41.1)	62 (47.7)	21 (29)		53 (50.0)	16 (35)	
Moderately or highly concerned	82 (40.6)	47 (36.2)	35 (49)		39 (36.8)	21 (46)	
Not reported[b]	16 (7.3)	4 (3.0)	5 (6)		3 (2.8)	1 (2)	
Reliability of information				.048			.13
Not concerned at all	30 (14.9)	16 (12.4)	14 (19)		9 (8.5)	9 (20)	
Slightly or somewhat concerned	81 (40.3)	60 (46.5)	21 (29)		49 (46.2)	17 (38)	

XSL•FO
RenderX

Concerns	Overall (N=218), n (%)	Visual disabilities			Physical disabilities[a]		
		No (N=134), n (%)	Yes (N=77), n (%)	*P* value	No (N=109), n (%)	Yes (N=47), n (%)	*P* value
Moderately or highly concerned	90 (44.8)	53 (41.1)	37 (51)		48 (45.3)	19 (42)	
Not reported[b]	17 (7.8)	5 (3.7)	5 (6)		3 (2.8)	2 (4)	

[a]Participants recruited from Kellogg Eye Center were not asked questions pertaining to physical disabilities, so the total number of people reporting physical disability status is a subset of the total sample.

[b]These values were excluded from the chi-square tests, as well as the % calculations.

Discussion

Principal Findings

In this study, we found the majority of participants had previous experience with DVAs and that there was interest in using them in the future. Looking specifically at those with self-reported disabilities, many adults with physical disabilities and visual disabilities do not regularly use DVAs, however, those with visual disabilities have a greater reliance on the technology. Overall, participants were moderately concerned with the issues of privacy, security, confidentiality, accuracy, and reliability of DVAs. Roughly half of those with visual disabilities had moderate or high concern for privacy, security, and confidentiality, but these numbers do not drastically differ from those without visual disabilities, or with or without physical disabilities, nor does it seem to affect their use and reliance on the technology. Previous literature has discussed smartphone-based low vision aids where DVAs are briefly examined among other smartphone tools for low vision usability [17,18]. The use of DVAs for people with disabilities, especially for the purposes of living independently, is a rapidly expanding field of study [15]. Many recent studies have focused on the accessibility, perceptions, and feasibility of the applications of DVAs, but few have evaluated the actual use cases and frequency of DVA use among individuals with visual or physical disabilities. This is an important step toward understanding and developing technology that better meets the needs of these individuals to improve health outcomes and quality of life.

Regularity of DVA Use

Overall, it was observed that more than half of the participants used DVAs only “sometimes” or “rarely.” Despite the lack of current use when the survey instrument was completed, more than half of the participants were willing to use DVAs in the future. This was perhaps a function of age, seeing that the average age of participants was older at 52.6 years old. From the survey, it was observed that the dominant use of DVAs was found in smartphones. Smartphone use among older age groups differs from those younger as a 2024 study from the Pew Research Center reported that 91% of adults aged between 50 - 64 years and 79% of adults aged 65 years and older own smartphones in the United States, compared with 99% of adults aged 18 - 29 years [1]. Using smart speakers as DVAs in populations of older adults in low socioeconomic positions has also proved to be promising, though more research should be conducted to promote the answering of health-related questions [19]. The additional smartphone functions and tools available for consumers are not necessarily used by older populations. While lower DVA use is more common among older adults, this could be related to a number of factors including lesser comfort with the technology, lower digital literacy, usability factors, technology learning curves, and other personal hesitations.

Perceptions of DVAs

The perceptions of DVAs were quite similar across those with and those without physical and visual disabilities. Overall, about half of participants expressed a moderate to high concern with the security and reliability of DVAs. It was interesting to note that those with visual disabilities were less concerned with privacy and confidentiality issues than those without visual disabilities. Concerns regarding online data privacy, security, and reliability are not novel, as literature spanning decades has documented these concerns. With specific regards to data privacy, DVAs represent an interesting use case as large data giants such as Facebook (Meta) and Amazon are increasing their use of personal user data for advertising, as well as individual targeting and tailoring of products and services. DVA companies have a reputation for lacking transparency with users about privacy protection which can lead to a lack of trust among users [20]. Many DVAs rely on the “activation keyword approach,” where devices listen for a specific keyword before they start recording, which means that these technologies are always listening, unless muted. It also means that DVAs are susceptible to false activation, which means that recording could be unintentionally turned on at a time when users are not aware. Overall, there needs to be improved transparency and additional user controls for personal recording for DVAs and their devices. However, despite the concern over data privacy, 23% of participants reporting visual disabilities and 15% of participants reporting physical disabilities in our study were willing to grant access to their past existing data or share future data for future DVA studies.

The landscape of DVAs along with the integration of generative artificial intelligence (AI) is everchanging, which could lead to evolving data privacy and data quality concerns regarding DVAs. Since the release of ChatGPT by OpenAI in 2022, the interest and number of publications regarding “trust” in conversational agents has risen rapidly. Before 2024, computer science held the vast majority of publications regarding trust in conversational agents; however, the field of medicine has shown

XSL•FO
RenderX

greater growth in interdisciplinary research with conversational agents over the last 8 years [21]. Applications of AI in health systems will require close monitoring and transparency for patients to feel comfortable granting access [22]. Recent research has shown that many individuals feel relatively comfortable sharing their data with national health organizations and universities, but much less willing to share with private commercial organizations such as insurance and technological companies [23]. This further emphasizes the importance of exploring utility and applications of DVAs and AI in health care and research contexts.

Reliance on DVAs by Those With Visual Disabilities

The survey showed that those with visual disabilities had a greater reliance on DVAs than those without visual disabilities. This is likely due to those with visual disabilities relying on voice commands to read and assist them with tasks they otherwise would not be able to do. In terms of specific activities, participants with visual disabilities reported using DVAs more frequently for answering questions than those without visual disabilities.

DVA Use by Those With Physical Disabilities

The survey showed that the DVA use cases of those with physical disabilities did not differ significantly from those without physical disabilities. Due to the nature of the relatively small convenience sample, there was not a large degree of heterogeneity in terms of types of physical disabilities. Those with physical disabilities were shown to rely more on DVAs than those without. The finding of increased reliance is supported by previous interview-based research that noted DVAs promote autonomy by increasing the ability to perform daily activities and tasks that were normally obstructed by mobility or other barriers [24].

Strengths and Limitations

To our knowledge, this is one of the first studies to assess the utility and perceptions of DVAs for adults with specific focus on those with visual and physical disabilities. Despite the novelty of the study, there are multiple limitations to consider that can guide future exploration. It is crucial to note that due to the COVID-19 pandemic, data collection was disrupted with delays in analysis. Despite the lag experienced, the study begins to fill a literature gap on the use of DVAs, particularly from the perspective of those living with visual and physical disabilities.

This study is not without limitations. The study findings relied on participant self-reported information, which may introduce response or recall bias, such as participants not remembering, or choosing not to share information related to disabilities or DVA use. Moreover, this study used a convenience sample, which is a limitation to our approach. In addition, participants were recruited from one health system, so results may not be generalizable to other populations. To reduce participant burden among those with visual disabilities, the survey was shortened. This may have overlooked potential dual disability concerns (physical disabilities among those with visual disabilities). In addition, the study survey did not capture the specifics related to the reported disability, such as etiology, onset, severity and overall quality of life. In addition, the survey instrument used was created specifically for this study and not validated. Despite these limitations, our study was intended to be exploratory in nature and serve as a foundation for future work in this field.

Future Perspectives

Future work should examine DVA use through device use data recordings and compare this across different disabilities and functional levels. Additional areas of investigation should incorporate qualitative approaches to more fully elucidate the ways that DVAs have been used, or could be used, to support the health of older adults and those individuals living with disabilities. Not only does this have great implications for the more than 25% of individuals in the United States living with a disability [25], but it also has great implications for the number of aging individuals who may benefit from DVA use (eg, social isolation and improved independence).

Conclusions

Self-reported use of DVA varied across the surveyed participants. Tasks such as answering questions and seeking more information were the most used functions of DVAs by individuals with physical and visual disabilities. Participant perceptions of DVAs, regardless of disability status, suggest that concerns with data privacy, security, confidentiality, reliability, and accuracy exist.

Acknowledgments

The authors of this study wish to thank Sherry Day and Ebony Parker-Featherstone for their assistance with recruitment for this study, as well as Melanie Herschel, Kelley Wyse, and Katie Grode for administrative support. Additional thanks to the University of Michigan Briarwood Family Medicine Clinic, the Kellogg Low Vision Clinic, and the Department of Physical Medicine & Rehabilitation for supporting recruitment for this study. This research was unfunded. Generative artificial intelligence was not used in the production of this manuscript.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AF handled formal analysis. OA, KP, SD, and EPF conducted investigation. LRB, RK, MM, DW, and MP conducted methodology. LRB and RK performed project administration. LRB and RK handled supervision. AF conducted validation. AF handled

XSL•FO
RenderX

visualization. MR, LRB, AF, and RK performed writing–original draft. MR, LRB, AF, RK, and AD managed writing–reviewing and editing.
LRB, MM, DW, RK, and MP conceptualized the study.

Conflicts of Interest

LRB is the Editor-in-Chief of *JMIR mHealth and uHealth*.

Multimedia Appendix 1
Digital personal assistant pilot study.
[DOCX File, 171 KB - mhealth_v13i1e66185_app1.docx]

Multimedia Appendix 2
Digital personal assistant pilot study - large print.
[DOCX File, 148 KB - mhealth_v13i1e66185_app2.docx]

References

1. Pew Research Center. Mobile Fact Sheet. URL: https://www.pewresearch.org/internet/fact-sheet/mobile [accessed 2025-04-22]
2. Jamwal R, Jarman HK, Roseingrave E, Douglas J, Winkler D. Smart home and communication technology for people with disability: a scoping review. Disabil Rehabil Assist Technol 2022 Aug;17(6):624-644. [doi: 10.1080/17483107.2020.1818138] [Medline: 32924660]
3. Chang F, Sheng L, Gu Z. Investigating the integration and the long-term use of smart speakers in older adults' daily practices: Qualitative study. JMIR Mhealth Uhealth 2024 Feb 12;12:e47472. [doi: 10.2196/47472] [Medline: 38345844]
4. Ding D, Morris L, Messina K, Fairman A. Providing mainstream smart home technology as assistive technology for persons with disabilities: a qualitative study with professionals. Disabil Rehabil Assist Technol 2023 Oct;18(7):1192-1199. [doi: 10.1080/17483107.2021.1998673] [Medline: 34752169]
5. Budzinski O, Noskova V, Zhang X. The brave new world of digital personal assistants: benefits and challenges from an economic perspective. NETNOMICS 2019 Dec;20(2-3):177-194. [doi: 10.1007/s11066-019-09133-4]
6. Smart Audio Report 2022 from NPR and Edison Research. National Public Radio and Edison Research. URL: https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/ [accessed 2025-04-22]
7. Liao Y, Vitak J, Kumar P, Zimmer M, Kritikos K. Understanding the role of privacy and trust in intelligent personal assistant adoption. In: Taylor, N, Christian-Lamb, C, Martin, M, Nardi, B (Eds) Information in Contemporary Society iConference 2019 Lecture Notes in Computer Science 2019, Vol. 11420:102-113. [doi: 10.1007/978-3-030-15742-5_9]
8. Van der Roest HG, Wenborn J, Pastink C, Dröes RM, Orrell M. Assistive technology for memory support in dementia. Cochrane Database Syst Rev 2017 Jun 11;6(6):CD009627. [doi: 10.1002/14651858.CD009627.pub2] [Medline: 28602027]
9. Balasubramanian GV, Beaney P, Chambers R. Digital personal assistants are smart ways for assistive technology to aid the health and wellbeing of patients and carers. BMC Geriatr 2021 Nov 15;21(1):643. [doi: 10.1186/s12877-021-02436-y] [Medline: 34781881]
10. Conte C, López Pell AF. Digital applications and support devices for persons with intellectual disability and autism spectrum disorder. Vertex 2018 Jan;29(137):65-71. [Medline: 30605197]
11. Marziali RA, Franceschetti C, Dinculescu A, et al. Reducing loneliness and social isolation of older adults through voice assistants: Literature review and bibliometric analysis. J Med Internet Res 2024 Mar 18;26:e50534. [doi: 10.2196/50534] [Medline: 38498039]
12. Oewel B, Ammari T, Brewer RN. Voice assistant use in long-term care. 2023 Jul 19 Presented at: Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23); Jul 19, 2023; New York, NY URL: https://dl.acm.org/doi/proceedings/10.1145/3571884 [doi: 10.1145/3571884.3597135]
13. Shade M, Rector K, Kupzyk K. Voice assistant reminders and the latency of scheduled medication use in older adults with pain: descriptive feasibility study. JMIR Form Res 2021 Sep 28;5(9):e26361. [doi: 10.2196/26361] [Medline: 34581677]
14. Sezgin E, Militello LK, Huang Y, Lin S. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. Transl Behav Med 2020 Aug 7;10(3):606-628. [doi: 10.1093/tbm/ibz141] [Medline: 32766865]
15. Esquivel P, Gill K, Goldberg M, Sundaram SA, Morris L, Ding D. Voice assistant utilization among the disability community for independent living: a rapid review of recent evidence. Human Behavior and Emerging Technologies 2024 Apr 12;2024(1):1-39. [doi: 10.1155/2024/6494944]
16. Alagha EC, Helbing RR. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform 2019 Nov;26(1):e100075. [doi: 10.1136/bmjhci-2019-100075] [Medline: 31767629]

XSL•FO
RenderX

17. Robinson JL, Braimah Avery V, Chun R, Pusateri G, Jay WM. Usage of accessibility options for the iPhone and iPad in a visually impaired population. Semin Ophthalmol 2017;32(2):163-171. [doi: 10.3109/08820538.2015.1045151] [Medline: 26154395]
18. Yeo JH, Bae SH, Lee SH, Kim KW, Moon NJ. Clinical performance of a smartphone-based low vision aid. Sci Rep 2022 Jun 24;12(1):10752. [doi: 10.1038/s41598-022-14489-z] [Medline: 35750770]
19. McCloud R, Perez C, Bekalu MA, Viswanath K. Using smart speaker technology for health and well-being in an older adult population: Pre-post feasibility study. JMIR Aging 2022 May 9;5(2):e33498. [doi: 10.2196/33498] [Medline: 35532979]
20. Lau J, Zimmerman B, Schaub F. Alexa, are you listening?: privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Presented at: Proceedings of the ACM on Human-Computer Interaction; Nov 1, 2018. [doi: 10.1145/3274371]
21. Aksoy M, Bush A. A bibliometric analysis of trust in conversational agents over the past fifteen years. arXiv. Preprint posted online on 2024. [doi: 10.48550/arXiv.2408.16837]
22. Liyanage H, Liaw ST, Jonnagaddala J, et al. Artificial intelligence in primary health care: Perceptions, issues, and challenges. Yearb Med Inform 2019 Aug;28(1):41-46. [doi: 10.1055/s-0039-1677901] [Medline: 31022751]
23. Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A. Patient perceptions on data sharing and applying artificial intelligence to health care data: Cross-sectional survey. J Med Internet Res 2021 Aug 26;23(8):e26162. [doi: 10.2196/26162] [Medline: 34236994]
24. Vieira AD, Leite H, Volochtchuk AVL. The impact of voice assistant home devices on people with disabilities: A longitudinal study. Technol Forecast Soc Change 2022 Nov;184:121961. [doi: 10.1016/j.techfore.2022.121961]
25. Prevalence of disabilities and health care access by disability status and type among adults. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/disability-and-health/articles-documents/disabilities-health-care-access.html [accessed 2025-04-23]

Abbreviations

AI: artificial intelligence
DVA: digital voice assistant
HIPAA: Health Insurance Portability and Accountability Act
PM&R: physical medicine and rehabilitation
REDCap: Research Electronic Data Capture

Edited by TDA Cardoso; submitted 05.09.24; peer-reviewed by D Reifegerste, MD Rosa; revised version received 07.02.25; accepted 20.02.25; published 14.08.25.

Please cite as:
Rajan M, Furgal A, Kadri R, Arman O, Panzer K, Wicker D, McKee MM, Plegue M, Degner A, Buis LR
Exploring the Utility of Digital Voice Assistants for Primary Care Patients, Including Those With Physical and Visual Disabilities: Cross-Sectional Study
JMIR Mhealth Uhealth 2025;13:e66185
URL: https://mhealth.jmir.org/2025/1/e66185
doi:10.2196/66185

© Maya Rajan, Allison Furgal, Reema Kadri, Omar Arman, Kate Panzer, Donna Wicker, Michael McKee, Melissa Plegue, Alexandria Degner, Lorraine R Buis. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Objectively and Subjectively Measured Physical Activity and Their Associations With Cardiometabolic Risk in the UK Biobank: Retrospective Cohort Study

Charlyne Bürki[1,2], MS; Caiwei Tian[3], BS; Kenneth Westerman[4,5], PhD; Chirag Patel[1], PhD

[1]Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck St, Boston, MA, United States

[2]Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland

[3]Faculty of Arts and Sciences, Harvard University, Cambridge, MA, United States

[4]Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA, United States

[5]Programs in Metabolism and Medical Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, United States

Corresponding Author:
Chirag Patel, PhD
Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck St, Boston, MA, United States

Abstract

Background: The association between physical activity (PA) behavior and cardiometabolic risk factors has depended largely on questionnaire-based reporting. More studies are turning to mobile health (mHealth) device solutions to measure PA. While there are differences between self-reported activity levels and objectively measured accelerometer-based activity, how these differences manifest in disease risk is unknown.

Objective: Here, we sought to evaluate these differences between self-reported and mHealth-based PA and to model the impact on their association with cardiometabolic factors. Our study provides a framework to assess the quality of relationships measured by mHealth technologies, which is generalizable to other sensors or activity-measuring devices.

Methods: We assessed PA using both wrist-worn accelerometer data and self-reported questionnaires in 16,000 participants of the UK Biobank (UKB) between 2013 and 2015, focusing on walking, sleeping, sedentary, and moderate-to-vigorous physical activity (MVPA). We compared the concordance between self-reported and objective measures of PA. We also compared the association between objectively measured or self-reported PA and future clinical biomarker levels (eg, BMI, pulse rate, glucose control, and cholesterol).

Results: Participants underestimated their weekly sedentary duration on average of 2.86 hours, and the coefficient of correlation (r) between subjective and objective activity was 0.12 for sedentary time, 0.16 for MVPA, 0.18 for walking, and 0.13 for sleeping. We found an inverse association between objectively measured MVPA and cardiometabolic biomarkers such as BMI and pulse rate, but found no association between subjectively reported activity and cardiometabolic biomarkers. We estimated that there is a 6% larger association between subjectively measured MVPA and BMI in healthy adults (vs the objective counterpart). We also estimated a 2%-3% difference on a healthy adult heartbeat (healthy range: 60-100 bpm) if relying on subjectively reported observations instead of measured PA.

Conclusions: These findings suggest that the association based on self-reported activity is likely overestimated and biased compared with objectively measured PA. Therefore, care should be taken when assessing the effects of self-reported PA on key cardiometabolic factors, such as BMI and pulse rate. We emphasize that while the associations are biased when comparing PA modalities, we cannot conclude which method more closely reflects the daily activity load.

(JMIR Mhealth Uhealth 2025;13:e54820) doi:10.2196/54820

KEYWORDS

accelerometer; exercise; risk factor; sedentary; biobank; physical activity; cardiometabolic risk; cardiometabolic; observational study; clinical biomarker; biomarkers; body mass index; pulse rate; glucose; cholesterol

Introduction

It is well established that physical activity (PA) is paramount in the maintenance of health and reduction of risk of chronic disease [1]. PA, whether self-reported or measured by mobile health (mHealth) devices, is consistently associated with higher risks of cardiovascular diseases, type 2 diabetes, low life satisfaction, and general worse health [2-5]. A better understanding of the relationship between specific PA behaviors and health is critical for disease prevention, but efforts are hindered by the complexity of PA assessment.

XSL•FO
RenderX

PA is characterized by 4 dimensions (type, frequency, duration, and intensity), is carried out in at least 4 different contexts (home activities, leisure, transport, and occupational), and can be reported in at least 3 different manners (using energy expenditure in kilocalories, categorizing time spent in different types of activities over a period of time, or using the metabolic equivalent [MET] of a task) [6]. The heterogeneity in reported PA contributes to the conflicting results in this field and adds to the difficulty of discerning a clear link between PA and diseases or biomarkers predictors of diseases. For instance, in the US Centers for Disease Control and Prevention National Health and Nutrition Examination Survey (NHANES), Troiano et al [7] found that less than 5% of US adults met PA directives when relying on accelerometer-measured PA, but more than 51% did when basing on self-reported questionnaire data. Using the same survey, Atienza et al [8] found that physiological and anthropometric biomarkers associated more strongly with accelerometer-derived moderate-to-vigorous physical activity (MVPA) measurements than self-reported PA. Other meta-analyses also reported systematic underestimations of measured sedentary time (as much as 1.74 hours per day) when using self-reported measures [9,10]. As a result, there is conflicting evidence for the validity of relying on self-reported PA.

Sensors and wearable devices (mHealth devices), including accelerometric devices, have progressively integrated into cohort studies, allowing an increasingly diversified field of participants and an easier framework to collect measured data. While these wearable devices can measure PA, they cannot determine disease risk outcomes directly from measured PA. Instead, physiologically relevant metrics modulating diseases must be measured.

Then, a relationship between measured (mHealth) and self-reported PA levels and disease can be quantitatively compared. If these are assessed in the same cohort study, then it is possible to determine whether mHealth devices allow obtaining additional signals with lower error in comparison to traditional self-reporting PA. Furthermore, it is also possible to demonstrate the added value of mHealth-like devices for research through such quantitative comparisons.

In this study, we evaluated the concordance between objectively-measured (via mHealth devices) and subjectively-reported PA in their association with biomarkers (eg, BMI, pulse rate) of cardiometabolic disease in the UK Biobank (UKB) cohort for a quantitative comparison of these measurement techniques in a research setting. We examined the differences between objective and subjective activity population means at an individualized level. Finally, we compared the different types of activities in the cohort, highlighting the discrepancies found when predicting cardiometabolic biomarker levels from activity levels. We anticipate that the approaches developed here to systematically study associations between PA data modalities and clinical outcomes will be generalizable to new approaches and sensors that measure activity. Thematically, these findings demonstrate addressing key themes in mHealth, including mHealth for Data Collection and Research, Evaluation and Research Methodology for mHealth, and Fitness Trackers.

Methods

Physical Activity Measurements

Measured Activity Derived From an Accelerometer

It is hypothesized that weekend and weekday patterns of MVPA are different [5]. Therefore, we sought to analyze the population on both weekends and weekdays separately. Hourly proportion of time spent in weekly and weekend activity was extracted and preprocessed by Le Goallec et al [11] from the publicly available software [12,13] and converted to weekly minutes of types of activity. The weekly total activity of the 4 types of activity—walking, sleeping, sedentary, and MVPA—was obtained by summing all the hourly proportion of weekday and weekend separately, weighting by 5 and 2, respectively, and converted to weekly minutes by multiplying by 60. Following the filtering out of missing or unusable accelerometer-derived data, the number of participants was 16,467. For sensitivity analysis, we also subdivided the study population to only retain individuals who had a nonwear time of zero days, meaning that all the included participants wore the accelerometer device for an uninterrupted period of 7 days. This allowed for a quality assessment of the objective PA in 5587 individuals.

Self-Reported Activity Derived From the Physical Activity Questionnaire

Weekly minutes of activity estimated by questionnaires were obtained by multiplying the weekly frequency by the daily amount (see Table S1 in Multimedia Appendix 1 for all the questions) only if participants had answered both questions. The questionnaire answers were part of a short-form International Physical Activity Questionnaire (IPAQ) filled out at every visit. If a participant did not recall the duration of an activity but had at least practiced it for 10 minutes, the participant's average was substituted. Any participant with an incomplete answer was dropped. To compare to the measured MVPA, weekly minutes of moderate activity were added to vigorous physical activity (VPA) reported by participants. Weekly minutes of sleep were calculated by converting the reported daily hours of sleep to daily minutes and multiplying by 7. Self-reported time spent watching television, on the computer, and driving was added and capped at 7560 weekly minutes (corresponding to 18 daily hours) as other studies have reported doing [3] to obtain weekly sedentary minutes. If participants reported spending less than 1 hour in either category, the total for that category was encoded as 0 minutes.

Anthropometric and Biomarker Data

Response Variable Encoding

We chose "response" variables, or dependent clinical variables that include anthropometrics and plasma biomarkers, such as BMI and pulse rate [3,14,15]. We also chose "negative controls," such as sunscreen use. Blood pressure readings were averaged for the visit. Answers containing "prefer not to answer" were excluded, and those that answered "do not know" were replaced with the average of the population to maximize the study population size.

XSL•FO
RenderX

Covariate Selection to Adjust for Lifestyle and Comorbidities

Dietary factors were included as covariates in regression models incorporated as 6 separate regression terms in the final model, each accounting for weekly intakes of vegetables, fruits, poultry, fish, meat, and bread-type. We included a variable for red meat (eg, pork, beef, lamb/mutton) and another for poultry. Bread type was encoded as a binary indicator of nonwhite or white bread.

We used a modified version of the Charlson Comorbidity Index (CCI), which measures the total comorbidities for a participant, to account for frailty and disease. We also included in the modified CCI any self-reported diabetes diagnosed by a practitioner or a medical record for type I or type II diabetes, any self-reported cancer diagnosed, any *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) record for Chronic Obstructive Pulmonary Disease, and any record of major cardiovascular diseases.

Statistical Analysis

Study Population Statistics

We reported the mean and SD for all measures where possible. For proportions, we provided a percentage and the nonsimplified fraction.

Quantifying the Measured/Self-Reported Relationship

The median and IQR of each activity, for both self-reported and measured, were calculated. Pearson correlation *r* was reported between measured/self-reported pairs of each activity type.

Estimating Associations Between Biomarker Levels and PA in 2 Separate Models for Measured and Self-Reported Activity

Models predicting individual biomarker levels from self-reported PA only (MVPA and sedentary time) were fit and adjusted for self-reported dietary intake, modified CCI, and smoking status. Separately, we modeled individual biomarker levels from measured PA, for both MVPA and sedentary time, adjusting for dietary intake, adapted CCI, and smoking status. The models using only measured PA values weighted each observation by a decay function that mapped individuals' time interval between their measurement and the subsequent date at which the biomarker was recorded. This mitigated crosstalk from other factors prone to vary as time passes and affect biomarker levels, such as medication use, by upweighting individuals with biomarker measurements closer to the date of activity measurement. The decay function, fΔdays=12πe-Δdays22(2*365)2+0.5, is a Gaussian with a SD of 2 years, translated vertically by 0.5 units and normalized.

We extracted the β coefficients and their 95% CI of respectively measured and self-reported PA to compare effect sizes for every permutation of biomarker/activity combination, stratifying the results by sex and age. To calculate the average effect size difference between subjectively and objectively reported categories of PA, we simply multiplied the effect sizes by, respectively, the subjective and objective population means of MVPA and sedentary time.

Quantifying the Independent Contribution of Objective PA Versus Self-Reported PA

To compare the effect size and the independent contribution of objective versus self-reported PA, we estimated the 95% CI of measured and self-reported estimates. Specifically, we fit a model predicting the biomarker level ("response" variable) as a function of measured PA or self-reported PA, and other adjusting variables. In summary, if the CI of the PA estimate after adjusting for lifestyle and the CCI variables did not contain 0 (was *P* value significant), we concluded that measured PA added additional signal independent of self-reported PA. The impact of PA on selected biomarkers was ranked based on the estimates' magnitudes.

All statistical analyses were carried out on software R version 4.2.1 (R Core Team) with RStudio 2022.02.1+461 "Prairie Trillium" (Posit, PBC) release for Chrome (Google LLC)/111.0.0.0.

Ethical Considerations

- We applied for internal review board (IRB) approval to the Harvard University IRB. The Harvard University IRB approved of this study (IRB: IRB16-245). The project number for this manuscript is 22881. The UKB investigative team had informed consent, and individuals who have opted out of the study have been removed from the analysis. The consent form is provided on the web [16]. The UKB data are deidentified data. UKB recipients were not directly compensated for participation; however, participants were reimbursed for time and travel. More information can be found here [17].

The UKB is a prospective cohort study of 502,507 recruited between 2007 and 2010 [18]. Participants were asked to participate in follow-up studies (at random). Between 2013 and 2015, a total of 103,661 participants consented to wear a wrist accelerometer for 7 consecutive days [19]. The study participants all took part in subsequent visits between 2013 and 2015, following their participation in the accelerometer measurement.

Results

Study Population Description

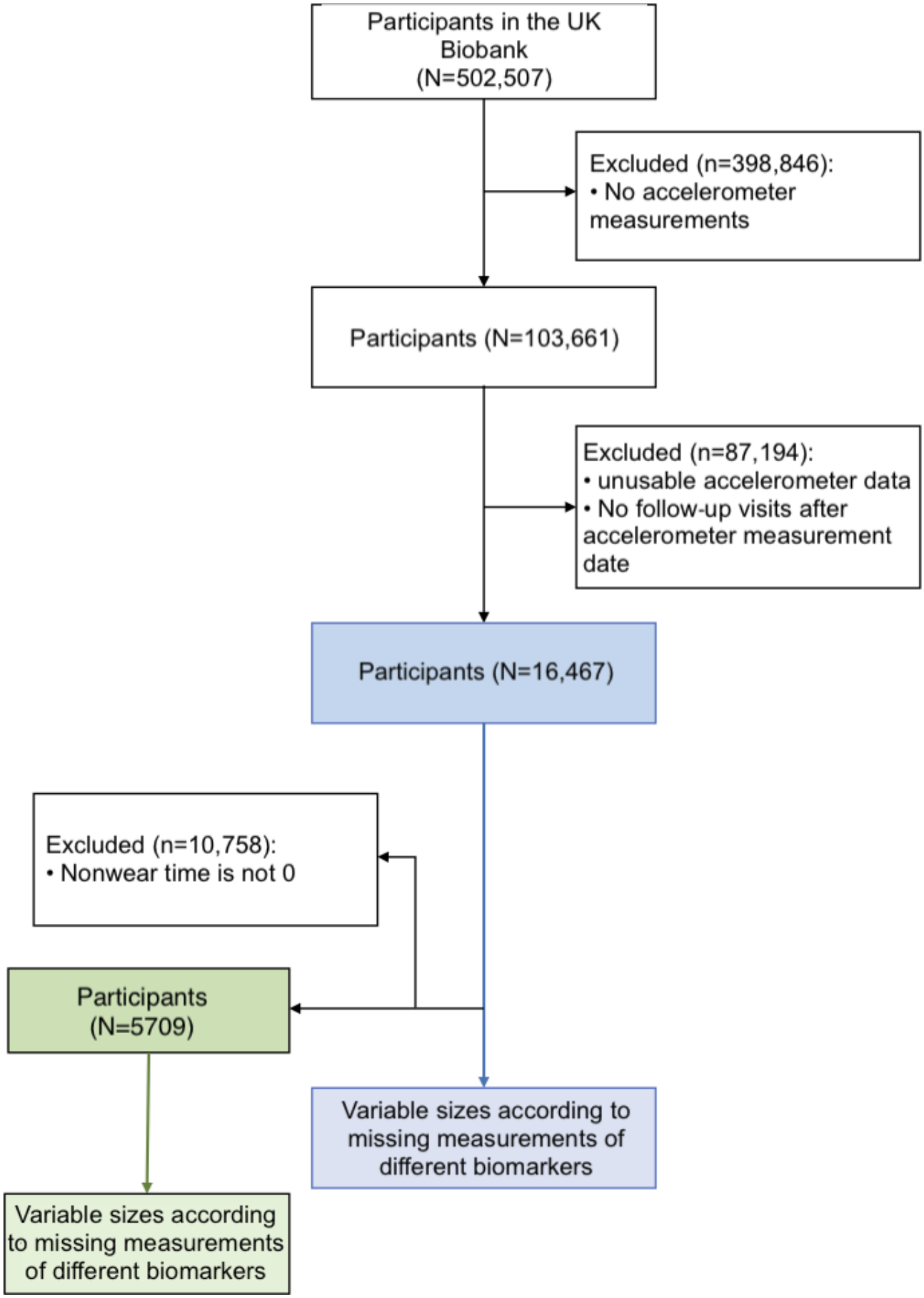
Of the 502,507 participants included in the UKB, 103,661 had accelerometer records available. Of those, 16,467 had valid accelerometric data and had anthropometric and biomarker data available following the accelerometer measurement visit (see Figure 1 for the flowchart of the study). The study cohort was comparable to the whole UKB population, in that it was primarily white (97% vs 95%), had similar measured anthropometrics such as BMI (26.29 [4.39] kgm2 vs 27.43 [4.8] kgm2) and similar proportions of males and females (53% females vs 54%). However, the study population was more educated (56% in the highest education category vs 33%), had higher income (31.5% in the top 2 income categories vs 25.4%), and healthier lifestyle habits in that the majority had never smoked (63% vs 55%). The studied cohort was also on average 4 years older than the whole UKB population at baseline, even after accounting for the time elapsed between initial recruitment

XSL•FO
RenderX

and the time of measured PA. Refer to Table 1 for all evaluated characteristics.

XSL•FO
RenderX

Figure 1. Flowchart of the study: in blue the main study population; in green the subpopulation that wore the accelerometer for an uninterrupted period of 7 days.



XSL•FO
RenderX

Table . Descriptive data comparing the study population with the UK Biobank cohort.

Characteristics	All, (N=502,406)[a]	Study, (N=16,467)[b]
Sex, n (%)		
Female	273,324/502,406 (54)	8726/16,467 (53)
Male	229,082/502,406 (46)	7741/16,467 (47)
Ethnic background		
Bangladeshi	236/499,630 (<0.1)	6/16,425 (<0.1)
Black	8058/499,630 (1.6)	90/16,425 (0.5)
Chinese	1573/499,630 (0.3)	50/16,425 (0.3)
Indian	5951/499,630 (1.2)	97/16,425 (0.6)
Mixed	2954/499,630 (0.6)	71/16,425 (0.4)
Other Asian	1857/499,630 (0.4)	20/16,425 (0.1)
Other ethnic group	4557/499,630 (0.9)	71/16,425 (0.4)
Pakistani	1835/499,630 (0.4)	15/16,425 (<0.1)
White	472,609/499,630 (95)	16,005/16,425 (97)
Prefer not to answer, n	2776	42
Age (years), mean (SD)	56.57 (8.10)	64.32 (7.76)
Above 65 years old	74,746/502,406 (15)	6287/16,467 (38)
BMI	27.43 (4.80)	26.29 (4.39)
Missing	3104	517
Smoking		
Never	273,473/499,458 (55)	10,358/16,467 (63)
Previous	173,023/499,458 (35)	5630/16,467 (34)
Current	52,962/499,458 (11)	479/16,467 (2.9)
Prefer not to answer, n	2948	0
Highest education level (years)		
7	84,140/485,067 (17)	634/14,319 (4.4)
10	81,390/485,067 (17)	1558/14,319 (11)
13	26,247/485,067 (5.4)	712/14,319 (5.0)
15	56,956/485,067 (12)	1361/14,319 (9.5)
19	75,217/485,067 (16)	2075/14,319 (14)
20	161,117/485,067 (33)	7979/14,319 (56)
Prefer not to answer, n	17,339	2148
Resting pulse rate	69.36 (11.26)	68.35 (11.54)
Missing, n	30,109	2870
Body fat, n (%)	31.45 (8.55)	30.78 (8.13)
Missing, n	10,405	847
Waist circumference	90.31 (13.49)	87.74 (12.72)
Missing, n	2160	481
Alcohol consumption		
Never	40,627/502,406 (8.1)	1062/16,467 (6.4)
Occasionally	113,836/502,406 (23)	3624/16,467 (22)
Frequently	346,443/502,406 (69)	11,781/16,467 (72)
Prefer not to answer	1,500/502,406 (0.3)	0/16,467 (0)

XSL•FO
RenderX

Characteristics	All, (N=502,406)[a]	Study, (N=16,467)[b]
Household income (in Great British Pounds)[d]		
<18,000	97,180/425,265 (23)	1690/14,849 (11)
18,000-30,999	108,157/425,265 (25)	4053/14,849 (27)
31,000-51,999	110,751/425,265 (26)	4471/14,849 (30)
52,000-100,000	86,250/425,265 (20)	3515/14,849 (24)
>100,000	22,927/425,265 (5.4)	1120/14,849 (7.5)
Prefer not to answer	77,141	1618
Blood pressure (systolic/diastolic)	137.82(18.68)/82.22(10.16)	138.79(18.77)/78.54(10.05)
Missing, n	30,109	2870
Self-reported weekly PA[c] (mins)		
Moderate to vigorous	362.56 (550.29)	377.58 (463.31)
Sedentary	1,900.43 (1,065.78)	2051.01 (1000.46)
Walking	358.38 (502.41)	358.86 (416.73)
Sleeping	3,006.04 (456.83)	3005.02 (426.04)
Missing, n	46,972	0
Self-reported nutrition		
Daily vegetables (heaped table spoon)	4.88 (3.37)	5.08 (3.27)
Weekly fruits (serving)	3.05 (2.63)	3.20 (2.54)
Weekly fish (serving)	2.27 (1.60)	2.36 (1.57)
Weekly meat (serving)	3.60 (2.24)	3.24 (2.10)
Weekly poultry (serving)	1.92 (1.26)	1.93 (1.29)
Missing, n	893	0

[a]Measures taken at recruitment.

[b]Measures taken after the visit wherein physical activity was measured.

[c]PA: physical activity.

[d]Great British Pound to US Dollar at the time of the study: US $1.21=GBP 1.

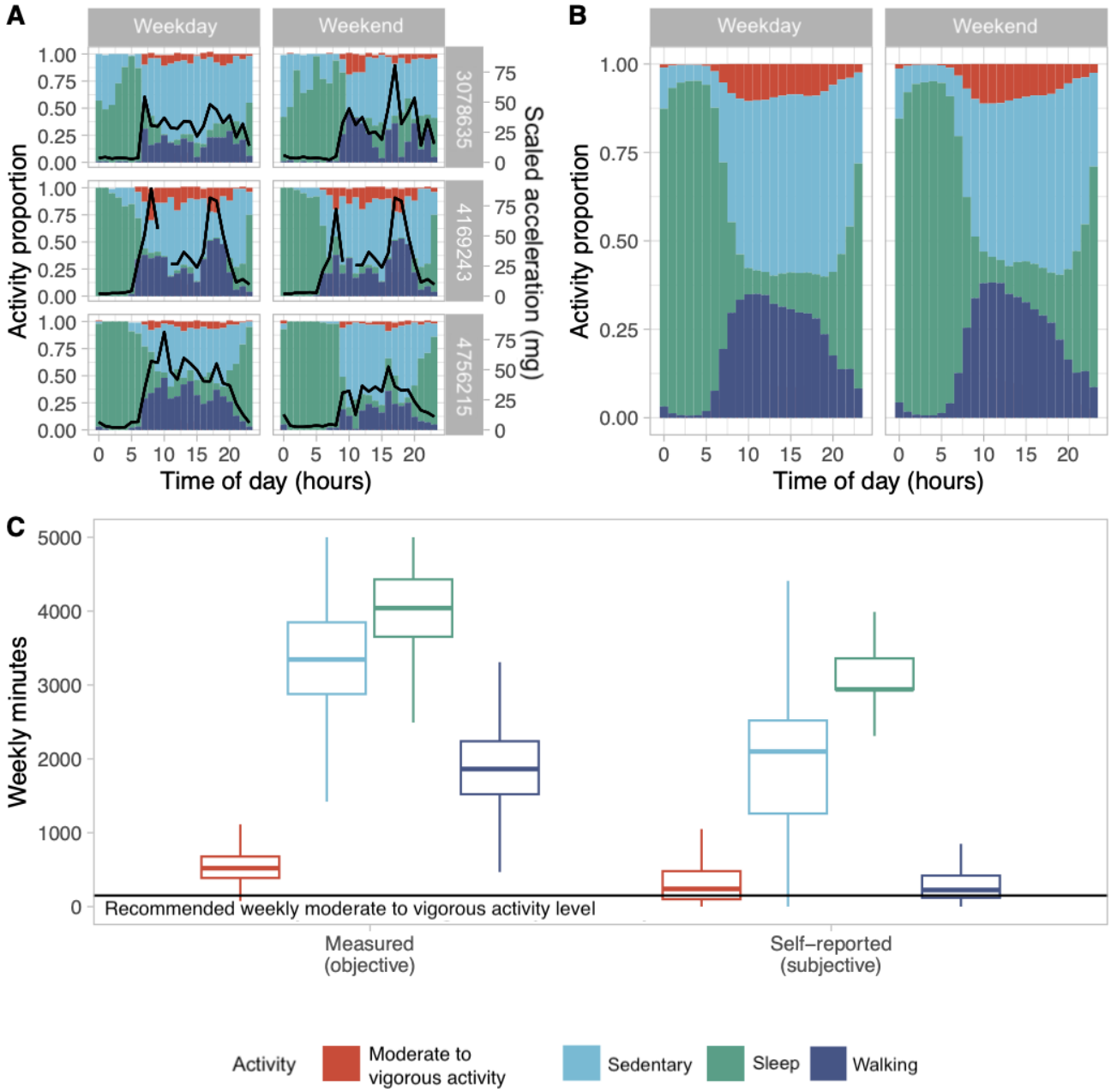
Relationship Between Measured and Self-Reported PA

Measured or objective PA was assessed at both the individual level (Figure 2A). We observed concordance between 3 random participants' hourly classified activities and the averaged triaxial acceleration, and at the population level in Figure 2B. At the population level, and even at the individuals' level, we found that a little proportion of time was spent in MVPA. Throughout the day, individuals spent at most 12.5% of an hour engaged in MVPA, whereas 50% of their time was spent in sedentary status outside of their sleep. The weekly median number of minutes spent in sedentary time was 3200 minutes (7.62 daily hours), 2000 minutes (4.76 daily hours) in walking, and 500 minutes (1.19 daily hours) in MVPA (Figure 2C). Participants averaged 24-hour period habits during the weekday did not differ from habits during the weekend.

The self-reported levels of sedentary time were not concordant with the objectively measured sedentary time. Participants reported 2000 weekly minutes (4.76 daily hours) of sedentary time, or about 1200 weekly minutes (2.85 daily hours) less than the measured level. The population mean of self-reported walking was similar to the self-reported MVPA population mean, each averaging 250 weekly minutes (0.6 daily hours). However, these self-reported values deviated from their measured counterparts by 1750 minutes for walking (4.17 daily hours) and 250 minutes for MVPA (0.60 daily hours; Figure 2C).

XSL•FO
RenderX

Figure 2. (A) Three individuals' average measured hourly activity over the course of a day overlaid with their scaled averaged recorded triaxial acceleration (black line) separated by weekday/weekend (black line discontinuities correspond to values in excess of the figure window), (B) the averaged (mean) classified measured hourly activity of the study cohort over a 24-hour period separated by weekday/weekend, and (C) the cohort's weekly measured and self-reported physical activity, stratified by activity type.

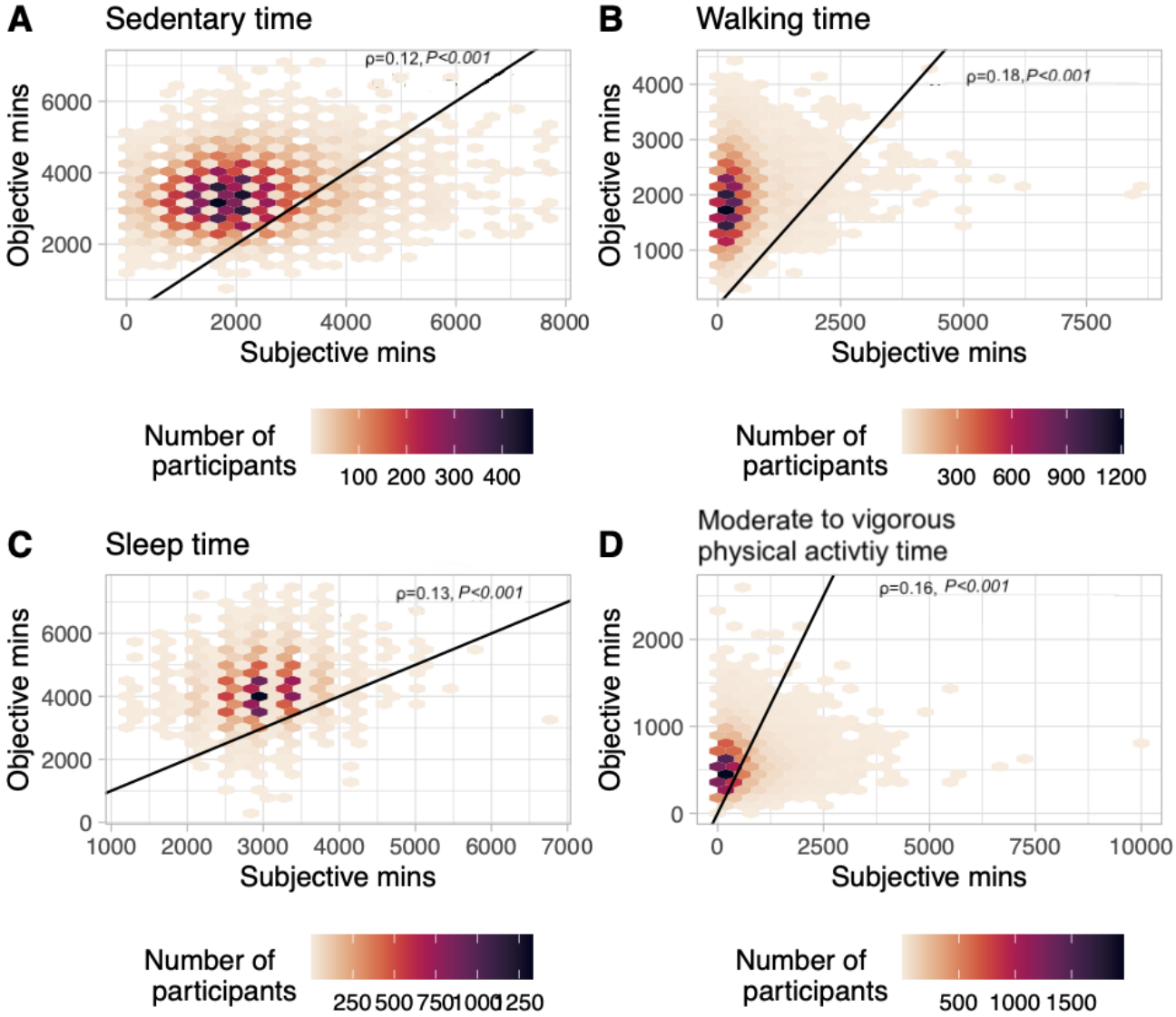


Quantifying the Measured/Self-Reported PA Relationship

Objective PA was modestly associated with subjective PA in Figure 3A-D (r=0.12; P<.001 for sedentary; r=0.18; P<.001 for walking; r=0.16; P<.001 for MVPA; and r=0.13; P<.001 for sleeping). Interestingly, the correlation was higher for walking and MVPA activities than for sedentary time and sleeping. We found the same correlations for the subpopulation that had noninterrupted accelerometer wear time during the study period (figures not included).

XSL•FO
RenderX

Figure 3. Scatter plots of the study population comparing the reported subjective minutes to the objectively measured minutes in each type of physical activity. The correlation coefficient (*rho*) is reported in each grid. The black line is the identity line (y=x), corresponding to a perfect one-to-one match. Hexagons are color-coded by their density: (A) weekly minutes of sedentary time, (B) weekly minutes of walking, (C) weekly minutes of sleeping, and (D) weekly minutes of moderate to vigorous activity.

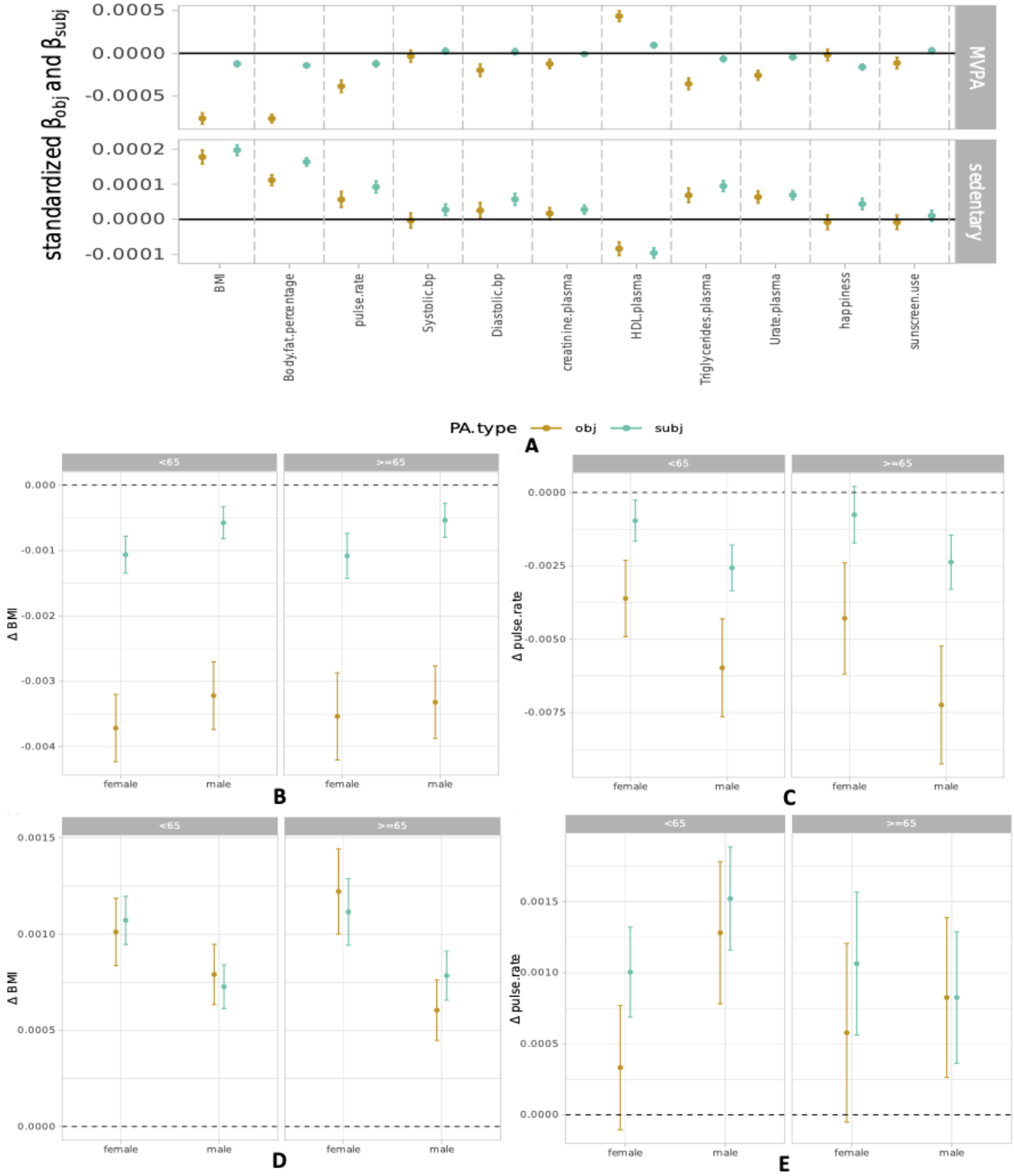


Effect of PA on Cardiometabolic Biomarkers

Figure 4A depicts the standardized effect size of measured and self-reported MVPA and sedentary time on selected cardiometabolic-related biomarker or "response" levels (see Figure S1 in Multimedia Appendix 1). The subjective and objective PA are modeled jointly, such that each measurement's effect estimate assumes the other is held constant; in other words, the coefficients measure the effect of one PA measure independent of the other. The objective PA association sizes were greatest for anthropometric markers (BMI, waist circumference, body fat mass), certain physiological markers such as pulse rate, and certain plasma biomarkers such as high-density lipoprotein and triglycerides. As a negative control, objective sedentary time did not predict other biomarkers presumably unrelated to physical fitness, such as sunscreen usage (β=−8.42e-6; 95% CI −2.89e to 5,1.20e0; *P*=.42); however, objective MVPA was nominally significant (β=−1.13e−4; 95% CI −1.79e−4 to 4.72e−5; *P*=.009). Also, BMI and pulse rate levels had an inverse association with MVPA (eg, lower BMI and pulse rate for higher levels of MVPA) and a positive association with sedentary time. Further, subjective MVPA estimates were null across all biomarkers, but subjective sedentary time estimates were similar to the estimates predicted by objective sedentary time.

XSL•FO
RenderX

Figure 4. (A) Standardized effect sizes of objective physical activity adjusted for dietary intake, adapted CCI, and subjective physical activity on selected biomarkers. The subjective physical activity coefficients are plotted as a comparison; (B-E) effect sizes of PA on selected biomarkers issued from two separate models of objective and subjective physical activity, stratified by age group: (B) moderate-to-vigorous physical activity BMI, (C) moderate-to-vigorous physical activity pulse rate, (D) sedentary BMI, (E) and sedentary pulse rate.



The association sizes of objective MVPA were directionally consistent and stronger in magnitude than those of subjective MVPA in all groups (above/below 65, males/females; Figures 4B and C, Figure S2 in Multimedia Appendix 1). There was also no substantial difference in effect size between the age groups.

In contrast to MVPA, objective and subjective effect sizes for sedentary time coincided almost perfectly for all anthropometric and physiological measurements (see Figure S3 in Multimedia Appendix 1 for all tested biomarkers). Sedentary time was associated with an increase in BMI and pulse rate (Figures 4D and E). We also replicated these tests in the subpopulation that

XSL•FO
RenderX

wore the accelerometer without interruption and found the same results (figures not included).

Finally, we calculated the differences of biomarker effect sizes between each subjective/objective PA pair stratified by sex and age (above/below 65 years old) in original and nonstandardized units (eg, beats per minute for pulse rate and kgm2 for BMI). For BMI estimations from weekly MVPA, we found that subjective MVPA was associated with an increase of 1.5 kgm2 more than objective MVPA (1.59kgm2 [1.41, 1.78] for females <65, 1.47kgm2 [1.27, 1.67] for males <65, 1.5 kgm2[1.25, 1.75] for females >=65, and 1.53kgm2 [1.31,1.74] for males >=65). We also found that subjective MVPA was associated with an increase of approximately 2 bpm more than objective MVPA (1.56 bpm [1.09, 2.04] for females <65, 2.34 bpm [1.71, 2.98] for males <65, 1.95 bpm [1.25, 2.66] for females >=65, and 3.03 bpm [2.25, 3.8] for males >=65). For BMI predicted by sedentary time, we found that subjective estimates were lower than objective measurements by, on average, 1kgm2 for all subsets. See Figure S4 in Multimedia Appendix 1 for all tested biomarkers.

Discussion

Principal Findings

Here, we quantified the relationship between objective and subjective PA in two ways: (1) in a direct, one-to-one comparison and (2) in their association with cardiometabolic biomarkers. The former yielded similar results to previously reported studies assessing the differences between measured PA assessed with wearable devices and questionnaires [7,10,20,21].

First, the reported low Pearson correlations between objective and subjective PA measures were unsurprising given other reported levels below r=0.30 [7,21,22]. A European cohort reported a Pearson correlation of r=0.28 [21] while a Norwegian study reported correlations of r=0.46 for sedentary time and r=0.20 for MVPA [20]. Our results were lower. Further, participants underestimated their sedentary durations by almost 1200 weekly minutes (2.86 hours per day), a larger difference than what Prince et al [10] reported in their systematic meta-study review, where an underestimation of 731 weekly minutes (1.74 hours per day) was reported.

We found great mismatches when matching each individual's subjective/objective measurements. In particular, we found that the mismatch was greatest in those who self-reported higher values of MVPA. This may have indicated that either participants with high activity levels over-reported their MVPA workout activity, either because (1) the accelerometers were less adapted to validate highly-active participants' levels, or (2) that other factors like education level, age, or physical ability [20], affected participants' self-reporting values. A reasonable explanation is probably a combination of these hypotheses. For example, at the population level, measured activity medians of all types were higher than the questionnaire-based ones. In fine, discrepancies in perceived PA durations are sure to vary between individuals given their different mobility abilities, which were not assessed in the cohort.

The way that we assess PA has a large influence on the size and robustness of the association with clinical biomarkers of disease. Our findings are in agreement with other studies that describe the inverse association between increased MVPA and BMI and heart rate (respectively, the regular association between sedentary time and BMI and heart rate) [3,14]. However, we observed that objective measurements were more strongly associated with cardiometabolic biomarkers than has been previously appreciated. In the case of increased MVPA, the negative association of measured PA was of a larger magnitude than that of subjective PA by 4e-3 standard units on anthropometric markers and on the order of 1e-3 standard units in the case of plasma biomarkers and pulse rate. In the case of increased sedentary PA, the magnitudes of objective and subjective associations were largely invariant to age and sex/gender.

Moreover, we found discrepancies between levels of BMI and pulse rate estimated with subjective and objective MVPA. Subjective weekly MVPA was estimated on average, in the different population subsets, 1.5 kgm2 and 2 bpm more than objective MVPA. Contextualized in healthy adults, these numbers reflect a 6% overestimation of MVPA association on BMI, if relying on subjectively-reported observations rather than measured MVPA. The MVPA impact on pulse rate differences was less, varying between 2% and 3% difference on an adult heart beating 60-100 bpm. Though small, these variations may present key influences on clinical biomarkers.

Comparison With Previous Work

Due in part to the ease of implementation of IPAQ (long or short form) or equivalent questionnaires such as 7-day Physical Activity Recall [23], and in part by the relatively new widespread use of wearable devices, most cohort studies still rely on only subjective PA measurements to assess physical fitness. For instance, a recent study using the UKB found that participants adhering to frequent MVPA patterns based on subjective PA were associated with reduced risk of dementia [24]. Given the discrepancies between subjective and objective PA quantified in this study, however, it is questionable whether the reported effect is discernible in objectively measured PA in the UKB cohort.

PA investigation on cardiometabolic disease may benefit from increased power and precision from the use of accelerometer data. These findings support the hypothesis that the difference observed between subjective and objective MVPA measurements was more likely due to misjudging MVPA duration according to fitness levels and not age. Indeed, we did not observe any objective MVPA activity difference between the 2 age groups but found diluted pulse rate associations with increased MVPA activity in both groups (similarly, increased pulse rate with increased sedentary time). Given that the objectively reported effect sizes had greater magnitudes (and perhaps less error) than the subjective ones, greater power may be found with accelerometer-based measurements.

The measured PA data values were extracted and categorized from the freely available software [12,13,19]. Although measured PA was shown to strongly correlate with daily observed and reported behavior, the PA categorizations are only

XSL•FO
RenderX

approximations of MET, which is defined as using 3.5 mL of oxygen per minute per kilogram of body weight. By definition, a participant is classified as doing MVPA if the activity duration intensity is greater than or equal to 3 METs. While gold-standard methods such as doubly labeled water [25] exist and result in precise, individual-based calibrations, the method used to process accelerometric data did not benefit from an individual-centric calibration. This, therefore, is one source of the discrepancies between PA classifications.

We hypothesize that individuals with different baseline cardiometabolic health may perceive their PA differently, and this will bias associations between reported activity and cardiometabolic biomarkers. We describe a plausible scenario: for instance, if 2 individuals of different baseline metabolic capacities are carrying out the same activity intensity, such as running a certain distance at a certain pace for a certain time, an individual-centric method would result in different MVPA duration classifications for each individual. Equivalently, an individual with a healthier baseline metabolic capacity would have less MVPA duration than an individual with a lower metabolic capacity. Hence, in a scenario where both individuals run the same distance at the same pace, the accelerometer reading will be identical. However, the self-reported perception of PA may differ; the individual with better cardiometabolic fitness may have perceived and reported doing less MVPA than the other individual. This could translate to the reported discrepancy observed in the study participants and therefore influence the magnitude of the coefficients of the biomarkers studied.

Unlike MVPA, we observed concordant associations between sedentary activity (measured by both accelerometer and self-reported) on biomarker levels. We hypothesize that measurement "error" in the aggregate may be lower for activities related to inactivity, which sum over behaviors such as watching television and using a computer versus reporting one MVPA-related activity carried out weekly [9,20].

With the new generation of mHealth devices that have been developed since the study's conception, which incorporate personalized thresholds to categorize activity intensities [5]. There is evidence that activity intensity is assessed more objectively and therefore reflects more precise metabolic exertion. Leveraging these threshold personalization allows us to distinguish new behavioral patterns. For instance, McConnell et al [5] identified a cluster of adults who were more active during the weekend (called weekend warriors) in comparison to weekdays. Implementing studies with such devices is, however, expensive, but given the discrepancies highlighted in this study between self-reported and measured PA and predicting cardiometabolic biomarker levels, it may be crucial to re-evaluate PA assessments, especially when these are translated to prevent disease outcomes, such as in guidelines [26]. While we acknowledge the development of new technologies for measuring PA, the core contribution of our paper lies in the method we present to quantify the quality of relationships produced by mHealth technologies. This approach is not limited by specific devices or data sources and, importantly, should generalize to newer technologies as they emerge.

It may be possible, in the future, to create a calibration between self-reported and objective PA to reanalyze results from studies that have one measure and not the other. Many studies rely on self-reported measures. In this study, we are primarily focused on how the discrepancies between self-reported and objective measures impact associations with clinical outcomes, rather than directly translating between the two measures. As we show in Figure 3, the correlation between self-reported and objectively measured PA is very modest, below 0.1, highlighting the challenges inherent in creating such a mapping. This low correlation, in itself, demonstrates that providing a simple or reliable translation will require a formal study and potentially more data.

Use of self-reported versus accelerometer-based approaches for measuring PA has implications in biobank-scale genetic studies, such as "genome-wide association studies" (GWASs) [27]. GWASs are a discovery-based approach to broadly uncover inherited genetic loci, or places along the genome that predispose individuals to perform PA. However, we hypothesize that the scale of genetic discovery (the number of loci found) and their association sizes are influenced by the type of PA being interrogated. A number of papers have begun to interrogate the genetic contributions, via GWAS, in PA using both self-reported [28] and/or objective/accelerometer-based modalities.

For example, Klimentidis et al [29] conducted a GWAS of habitual PA of both self-reported PA and accelerometry data in the same participants analyzed in our study, individuals from the UKB. Briefly, they associated genetic loci with self-reported MVPA and VPA. They also associated genetic loci with average acceleration and fraction of time with higher accelerations as proxies of vigorous activity. In brief, while the sample sizes from each of the GWASs were different, they found different numbers and genetic loci from GWASs of self-reported and accelerometer-based measurement of PA. Specifically, the self-reported and objective indicators had an overlap in some loci found (particularly near gene *Cell Adhesion Molecule 2*), but the concordance was not consistent across all assessments of self-reported and accelerometer-based modalities. At a relaxed genome-wide *P* value threshold, the authors observed 4 genetic loci in common between self-reported MVPA and accelerometer and 3 loci in common between vigorous activity and accelerometer GWASs. Studying the difference in the genetic "architecture," or the number and biological implications of genetic loci, is an aim for a future study.

Limitations

Some caution should be used when interpreting these results. First, objective and subjective measurements were not collected at the same time. Participants of the UKB were invited to fill out IPAQ forms during their medical appointments independently of their participation in the accelerometer study. Furthermore, participants in the accelerometer study were not required to schedule a medical appointment during the week of their participation. Thus, self-reported PA was not collected simultaneously with the measured PA. While most participants had self-reported PA measures within weeks of their measured PA recording, some had at most 2 years between the measures.

XSL•FO
RenderX

Though we applied a weight correction accounting for time elapsed between the measures (self-reported PA and measured PA) and considered additional lifestyle factors, there were other factors that could not be captured, which may have influenced the observed differences between objective and subjective MVPA.

Conclusions

In sum, by associating both objective and subjective PA with health indicators and clinical risk variables, we quantified the epidemiological implications of the relationship between subjective and objective PA versus just studying the concordance between objective and subjective PA. We found that BMI and pulse rate followed a positive relation with MVPA and a negative association in the case of increased MVPA. Further, we showed that objective MVPA measurements, even in the presence of subjective MVPA, captured more robust signals than subjective MVPA. Importantly, the framework we presented here allows us to quantify the quality of relationships produced by mHealth versus self-reported approaches. We showed that our comparison should be carried out to evaluate the degree of bias of self-reported versus mHealth devices in large cohort studies. This has implications beyond the accelerometer devices used in this study and is generalizable to other mHealth devices. Thematically, these findings demonstrate addressing key themes in mHealth, including mHealth for Data Collection and Research, Evaluation and Research Methodology for mHealth, and Fitness Trackers.

Acknowledgments

This research is supported by National Institutes of Health's National Institute of Environmental Health Sciences (R01ES032470) and the funders had no role in the research.

Multimedia Appendix 1
Additional table and figures.
[DOCX File, 1175 KB - mhealth_v13i1e54820_app1.docx]

References

1. Sui X, LaMonte MJ, Laditka JN, et al. Cardiorespiratory fitness and adiposity as mortality predictors in older adults. JAMA 2007 Dec 5;298(21):2507-2516. [doi: 10.1001/jama.298.21.2507] [Medline: 18056904]
2. Cassidy S, Chau JY, Catt M, et al. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233,110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. BMJ Open 2016 Mar 15;6(3):e010038. [doi: 10.1136/bmjopen-2015-010038] [Medline: 27008686]
3. Emaus A, Degerstrøm J, Wilsgaard T, et al. Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromso study. Scand J Public Health 2010 Nov;38(5 Suppl):105-118. [doi: 10.1177/1403494810378919] [Medline: 21062845]
4. Master H, Annis J, Huang S, et al. Association of step counts over time with the risk of chronic disease in the all of us research program. Nat Med 2022 Nov;28(11):2301-2308. [doi: 10.1038/s41591-022-02012-w] [Medline: 36216933]
5. McConnell MV, Shcherbina A, Pavlovic A, et al. Feasibility of obtaining measures of lifestyle from a smartphone app: the MyHeart counts cardiovascular health study. JAMA Cardiol 2017 Jan 1;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
6. Strath SJ, Kaminsky LA, Ainsworth BE, et al. Guide to the assessment of physical activity: clinical and research applications: a scientific statement from the American Heart Association. Circulation 2013 Nov 12;128(20):2259-2279. [doi: 10.1161/01.cir.0000435708.67487.da] [Medline: 24126387]
7. Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
8. Atienza AA, Moser RP, Perna F, et al. Self-reported and objectively measured activity related to biomarkers using NHANES. Med Sci Sports Exerc 2011 May;43(5):815-821. [doi: 10.1249/MSS.0b013e3181fdfc32] [Medline: 20962693]
9. Prince SA, Cardilli L, Reed JL, et al. A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2020 Mar 4;17(1):31. [doi: 10.1186/s12966-020-00938-3] [Medline: 32131845]
10. Prince SA, Adamo KB, Hamel ME, et al. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 6;5(1):56. [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
11. Le Goallec A, Collin S, Jabri M, et al. Machine learning approaches to predict age from accelerometer records of physical activity at biobank scale. PLOS Digit Health 2023 Jan;2(1):e0000176. [doi: 10.1371/journal.pdig.0000176] [Medline: 36812610]
12. Doherty A, Smith-Byrne K, Ferreira T, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. Nat Commun 2018 Dec 10;9(1):5257. [doi: 10.1038/s41467-018-07743-4] [Medline: 30531941]

XSL•FO
RenderX

13. Willetts M, Hollowell S, Aslett L, et al. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. Sci Rep 2018 May 21;8(1):7961. [doi: 10.1038/s41598-018-26174-1] [Medline: 29784928]
14. Harris TJ, Owen CG, Victor CR, et al. What factors are associated with physical activity in older people, assessed objectively by accelerometry? Br J Sports Med 2009 Jun;43(6):442-450. [doi: 10.1136/bjsm.2008.048033] [Medline: 18487253]
15. Lippi G, Schena F, Montagnana M, et al. Significant variation of traditional markers of liver injury after a half-marathon run. Eur J Intern Med 2011 Oct;22(5):e36-e38. [doi: 10.1016/j.ejim.2011.02.007] [Medline: 21925040]
16. Consent Form: UK Biobank. URL: https://www.ukbiobank.ac.uk/wp-content/uploads/2025/01/Participant-original-consent-form.pdf [accessed 2025-08-11]
17. UK Biobank Information Leaflet. URL: https://www.ukbiobank.ac.uk/wp-content/uploads/2025/01/Participant-original-information-leaflet.pdf [accessed 2025-08-11]
18. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015 Mar;12(3):e1001779. [doi: 10.1371/journal.pmed.1001779] [Medline: 25826379]
19. Doherty A, Jackson D, Hammerla N, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. PLOS ONE 2017;12(2):e0169649. [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
20. Dyrstad SM, Hansen BH, Holme IM, et al. Comparison of self-reported versus accelerometer-measured physical activity. Med Sci Sports Exerc 2014 Jan;46(1):99-106. [doi: 10.1249/MSS.0b013e3182a0595f] [Medline: 23793232]
21. Steene-Johannessen J, Anderssen SA, van der Ploeg HP, et al. Are self-report measures able to define individuals as physically active or inactive? Med Sci Sports Exerc 2016 Feb;48(2):235-244. [doi: 10.1249/MSS.0000000000000760] [Medline: 26322556]
22. Spartano NL, Demissie S, Himali JJ, et al. Accelerometer-determined physical activity and cognitive function in middle-aged and older adults from two generations of the Framingham heart study. Alzheimers Dement (N Y) 2019;5(1):618-626. [doi: 10.1016/j.trci.2019.08.007] [Medline: 31660424]
23. Sylvia LG, Bernstein EE, Hubbard JL, et al. Practical guide to measuring physical activity. J Acad Nutr Diet 2014 Feb;114(2):199-208. [doi: 10.1016/j.jand.2013.09.018] [Medline: 24290836]
24. Zhu J, Ge F, Zeng Y, et al. Physical and mental activity, disease susceptibility, and risk of dementia: a prospective cohort study based on UK biobank. Neurology (ECronicon) 2022 Aug 23;99(8):e799-e813. [doi: 10.1212/WNL.0000000000200701] [Medline: 35896434]
25. Westerterp KR. Assessment of physical activity: a critical appraisal. Eur J Appl Physiol 2009 Apr;105(6):823-828. [doi: 10.1007/s00421-009-1000-2] [Medline: 19205725]
26. Romeres D, Schiavon M, Basu A, et al. Exercise effect on insulin-dependent and insulin-independent glucose utilization in healthy individuals and individuals with type 1 diabetes: a modeling study. Am J Physiol Endocrinol Metab 2021 Jul 1;321(1):E122-E129. [doi: 10.1152/ajpendo.00084.2021] [Medline: 33998292]
27. Uffelmann E, Huang QQ, Munung NS, et al. Genome-wide association studies. Nat Rev Methods Primers 2021;1(1):1-21. [doi: 10.1038/s43586-021-00056-9]
28. Wang Z, Emmerich A, Pillon NJ, et al. Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention. Nat Genet 2022 Sep;54(9):1332-1344. [doi: 10.1038/s41588-022-01165-1] [Medline: 36071172]
29. Klimentidis YC, Raichlen DA, Bea J, et al. Genome-wide association study of habitual physical activity in over 377,000 UK biobank participants identifies multiple variants including CADM2 and APOE. Int J Obes (Lond) 2018 Jun;42(6):1161-1176. [doi: 10.1038/s41366-018-0120-3] [Medline: 29899525]

Abbreviations

bpm: beats per minute
CCI: Charlson Comorbidity Index
GWAS: genome-wide association study
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
IPAQ: International Physical Activity Questionnaire
IRB: internal review board
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
NHANES: Disease Control and Prevention National
PA: physical activity
UKB: UK Biobank
VPA: vigorous physical activity

XSL•FO
RenderX

Edited by L Buis; submitted 22.11.23; peer-reviewed by A Justice, H Namba; revised version received 21.12.24; accepted 13.03.25; published 27.08.25.

Please cite as:
Bürki C, Tian C, Westerman K, Patel C
Objectively and Subjectively Measured Physical Activity and Their Associations With Cardiometabolic Risk in the UK Biobank: Retrospective Cohort Study
JMIR Mhealth Uhealth 2025;13:e54820
URL: https://mhealth.jmir.org/2025/1/e54820
doi:10.2196/54820

© Charlyne Bürki, Caiwei Tian, Kenneth Westerman, Chirag Patel. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Wear Time for the Analysis of Consumer-Grade Wearables' Data: Case Study Using Fitbit Data

Loubna Baroudi[1], MSc; Ronald Fredrick Zernicke[2,3], MSc, ND, PhD; Muneesh Tewari[4,5,6,7], PhD; Noelle E Carlozzi[8], PhD; Sung Won Choi[9], MSc, MD; Stephen M Cain[10], PhD

[1]Department of Mechanical Engineering, University of Michigan–Ann Arbor, 2505 Hayward St, Ann Arbor, MI, United States
[10]Department of Chemical and Biomedical Engineering, West Virginia University, Morgantown, WV, United States
[2]Department of Orthopedic Surgery, University of Michigan–Ann Arbor, Ann Arbor, MI, United States
[3]Exercise & Sport Science Initiative, University of Michigan–Ann Arbor, Ann Arbor, MI, United States
[4]Center for Computational Medicine and Bioinformatics, University of Michigan–Ann Arbor, Ann Arbor, MI, United States
[5]Department of Internal Medicine, University of Michigan–Ann Arbor, Ann Arbor, MI, United States
[6]Department of Biomedical Engineering, University of Michigan–Ann Arbor, Ann Arbor, MI, United States
[7]Veterans Administration Ann Arbor Healthcare System, University of Michigan–Ann Arbor, Ann Arbor, MI, United States
[8]Department of Physical Medicine and Rehabilitation, University of Michigan–Ann Arbor, Ann Arbor, MI, United States
[9]Department of Pediatrics, University of Michigan–Ann Arbor, Ann Arbor, MI, United States

Corresponding Author:
Loubna Baroudi, MSc
Department of Mechanical Engineering, University of Michigan–Ann Arbor, 2505 Hayward St, Ann Arbor, MI, United States

Abstract

Background: Consumer-grade wearables allow researchers to capture a representative picture of human behavior in the real world over extended periods. However, maintaining users' engagement remains a challenge and can lead to a decrease in compliance (eg, wear time in the context of wearable sensors) over time (eg, "wearables' abandonment").

Objective: In this work, we analyzed datasets from diverse populations (eg, caregivers for various health issues, college students, and pediatric oncology patients) to quantify the impact that wear time requirements can have on study results. We found evidence that emphasizes the need to account for participants' wear time in the analysis of consumer-grade wearables data. In Aim 1, we demonstrate the sensitivity of parameter estimates to different data processing methods with respect to wear time. In Aim 2, we demonstrate that not all research questions necessitate the same wear time requirements; some parameter estimates are not sensitive to wear time.

Methods: We analyzed 3 Fitbit datasets comprising 6 different clinical and healthy population samples. For Aim 1, we analyzed the sensitivity of average daily step count and average daily heart rate at the population sample and individual levels to different methods of defining "valid" days using wear time. For Aim 2, we evaluated whether some research questions can be answered with data from lower compliance population samples. We explored (1) the estimation of the average daily step count and (2) the estimation of the average heart rate while walking.

Results: For Aim 1, we found that the changes in the population sample average daily step count could reach 2000 steps for different methods of analysis and were dependent on the wear time compliance of the sample. As expected, population samples with a low daily wear time (less than 15 hours of wear time per day) showed the most sensitivity to changes in methods of analysis. On the individual level, we observed that around 15% of individuals had a difference in step count higher than 1000 steps for 4 of the 6 population samples analyzed when using different data processing methods. Those individual differences were higher than 3000 steps for close to 5% of individuals across all population samples. Average daily heart rate appeared to be robust to changes in wear time. For Aim 2, we found that, for 5 population samples out of 6, around 11% of individuals had enough data for the estimation of average heart rate while walking but not for the estimation of their average daily step count.

Conclusions: We leveraged datasets from diverse populations to demonstrate the direct relationship between parameter estimates from consumer-grade wearable devices and participants' wear time. Our findings highlighted the importance of a thorough analysis of wear time when processing data from consumer-grade wearables to ensure the relevance and reliability of the associated findings.

(JMIR Mhealth Uhealth 2025;13:e46149) doi:10.2196/46149

XSL•FO
RenderX

KEYWORDS

wear time; wearables; smartwatch; mobile health; physical activity; engagement; walking; dataset; wearable devices; reliability; behavior; caregiver; students; Fitbit; users

Introduction

Physical activity can be used to prevent and treat multifarious health issues. In the clinical setting and in research studies, the assessment of physical activity level is often carried out using self-report questionnaires. However, because self-reports are low in resolution, subjective, and often limited in scope, researchers have been looking for solutions to objectively quantify physical activity [1,2]. In the early 21st century, smartwatches and fitness trackers entered the consumer market [3]. These consumer-grade devices integrate a combination of sensors to measure key health metrics such as step count or heart rate. With the realization that fitness trackers can motivate users to exercise by providing them with real-time feedback, researchers and clinicians are now increasingly using these devices to monitor individuals [4-7]. In particular, Fitbit (Google Inc) smartwatches have imposed themselves in the clinical world due to their low price, high battery life, user-friendliness, and compatibility with most smartphones on the market. However, useful data can only be collected if participants are compliant and wear the watch.

With longitudinal studies, it can be challenging to maintain users' engagement and compliance. It is common to observe what is called "wearables' abandonment"; as the excitement from having a new gadget wears off, people start to wear the sensors less and less [8]. In the accelerometer research—where accelerometers are commonly used to monitor physical activity—researchers have established different methods to account for wear time in the data processing pipeline [9-12]. In both the Fitbit and the accelerometer literature, when wear time is considered, researchers use different thresholds on daily wear time or on step count to preprocess their data and extract valid days [13-16]. However, as wear time and activity levels vary for each study population, those thresholds might not always be generalizable or may be overly strict for some research questions. For instance, a very active population might need a higher step count threshold compared to a sedentary population. This suggests that researchers should evaluate whether their selected wear time requirements are appropriate for their study population and research question to ensure the validity of their data processing [17]. Researchers should also present a justification for their wear time requirements. Furthermore, as different research questions might require a different amount of data, the required amount of data will need to be adapted.

In this work, we aimed to further the understanding of the effects of compliance on parameter estimates. Our objective was to (Aim 1) quantify the sensitivity of calculations to different wear time requirements between and within different population samples and (Aim 2) demonstrate that not all research questions may require identical wear time requirements. We leveraged three Fitbit datasets from six different populations to illustrate this work: the estimation of the average (1) daily step count, (2) daily heart rate, and (3) average heart rate while walking. As the use of consumer-grade wearables is rapidly increasing, this work provides a thorough demonstration of the potential impact of wear time on essential parameter estimates in a diversity of populations (clinical and nonclinical) with varying degrees of compliance. This work provides quantitative evidence of the need to use wear time requirements in data analyses, specifically for study population samples with low compliance.

Methods

Description of Datasets

For this study, 3 Fitbit datasets from 6 different populations were used (Table 1).

Table . Dataset details.

Dataset	Caregivers			Students	Pediatric oncology	
	Hematopoietic cell transplantation	Huntington disease	Spinal cord injury		Caregivers	Patients
Sample size, n	30	21	19	2107	49	44
Study length, days	90	90	90	90	120	120
Fitbit type	Fitbit Inspire 2	Fitbit Inspire 2	Fitbit Inspire 2	Fitbit Charge	Fitbit Charge	Fitbit Charge
Fitbit wear instructions	Wear as much as possible	Wear as much as possible	Wear as much as possible	40 h/wk	Wear as much as possible	Wear as much as possible
Fitbit wear incentives	Fitbit given + US $1/day if some Fitbit OR survey data	Fitbit given + US $1/day if some Fitbit OR survey data	Fitbit given + US $1/day if some Fitbit OR survey data	Fitbit given	Fitbit given	Fitbit given

Three Different Caregiver Groups

The details of this study protocol can be found in Carlozzi et al [18]. Briefly, caregivers for persons with Huntington disease (HD), spinal cord injury (SCI), and hematopoietic cell transplantation (HCT) from different clinics at the University of Michigan were recruited between November 2020 and June 2021. This study's objective was to evaluate a just-in-time adaptive intervention to promote caregivers' self-care. This intervention leveraged the combination of Fitbit and survey data collected using an app (TBI-CareQOL). It is important to note

XSL•FO
RenderX

that participants were reminded to sync their Fitbit data every Monday and Friday if they had not already done so.

College Students

The details of this study protocol can be found in Cislo et al [19]. Briefly, graduate and undergraduate students at the University of Michigan were recruited between September 2020 and December 2020 to study students' mental health and physical activity during the COVID-19 pandemic. Different measures were collected using surveys, a smartphone app (Roadmap 2.0), and a Fitbit smartwatch.

Pediatric Oncology Caregivers and Patients

This dataset was collected in a study evaluating the use of a mobile health app (Roadmap 2.0) intervention for cancer caregivers and their patients [20]. Participants were recruited between September 2020 and September 2021 from the Adult and Pediatric Hematology and Oncology Units of Mott Children's Hospital in Ann Arbor, MI.

Data Analysis and Statistics

Compliance Levels for Each Population

We first compared the level of compliance and wear time for each population. To calculate wear time, we used the heart rate data from the Fitbit watch [13]. A heart rate value was registered by the watch every minute if the watch was worn. Thus, wear time was calculated as follows:

$$\text{Wear time} = \frac{\#\text{ Minutes of registered heart rate}}{\text{Total minutes}} \quad (1)$$

Total minutes depended on the time frame of interest. For instance, if we wanted to calculate wear time over 24 hours, the total number of minutes was 1440 minutes. Because our groups did not have a similar variance, and the sample sizes were largely different, the difference in wear time between groups was evaluated using a Kruskal-Wallis test [21]. If the omnibus test was significant, the Dunn test was used to evaluate pairwise differences [22]. Significance was evaluated at the .05 level.

Aim 1: Sensitivity of Parameter Estimates to Different Wear Time Requirements

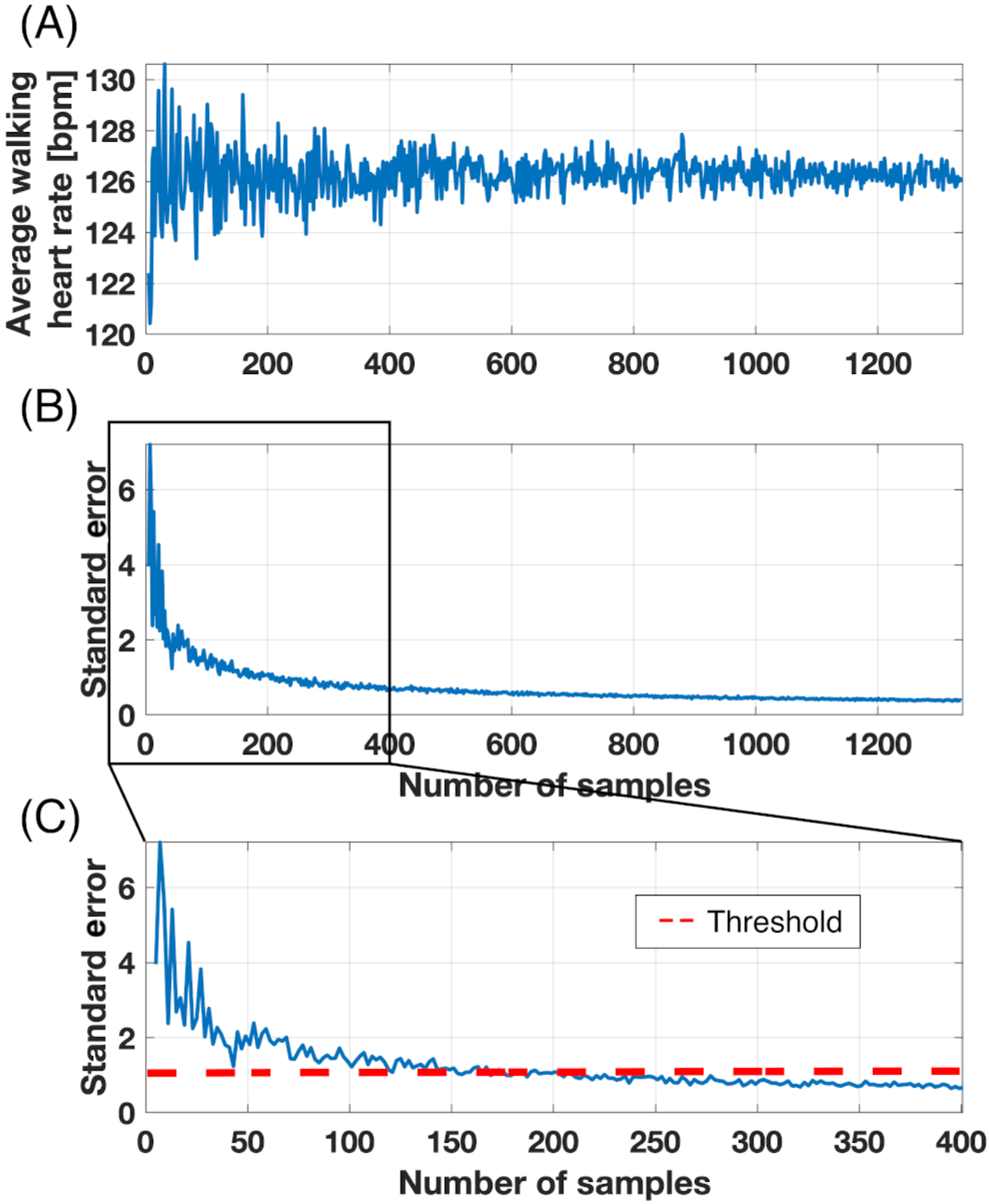
For Aim 1, three different definitions of compliance were used to evaluate the impact on the evaluation of the average daily step count and average daily heart rate for the population and the individual. Both of these measures have been used to determine fitness levels in multifarious populations [20-22]. Compliance was expressed in the form of valid days, where a valid day corresponded to a day that met a certain criterion and was kept for analysis. The 3 different definitions of a valid day were as follows: (*None*) all days were considered valid, (*StepCount1000*) a day was valid if the step count registered for that day was greater than 1000, and (*WearTime80*) a day was valid if the wear time that day (24 hours) was greater than 80% (19.2 hours). For that last definition, wear time was calculated using equation (1). The definitions of a valid day *StepCount1000* and *WearTime80* were based on studies found in the literature [13,14]. These 2 definitions also offer different perspectives on wear time by leveraging 2 distinct data types (step count vs wear time, defined here using heart rate). We also added the definition *None* to demonstrate the impact of not including wear time in analyses, which is the case in a large number of studies that use consumer-grade wearables [17]. The average daily step count and heart rate were calculated for the population and each individual using these 3 different definitions. The code to extract participants' wear time using Fitbit heart rate data and evaluate effects on step count is available on GitHub [23]. This repository includes the data preprocessing scripts and analysis code using an open-source Fitbit dataset.

Aim 2: Effect of Compliance on the Research Objective

For Aim 2, we examined the effect of compliance on the ability to address different research questions. We wanted to investigate whether some research questions can be answered, even when subject compliance is poor. Two research objectives were compared for this aim: (1) the evaluation of average daily step count and (2) the evaluation of average heart rate while walking. Heart rate while walking can be an indicator of cardiac health and exercise intensity [24-26]. To obtain heart rate while walking, we used Fitbit's number of steps taken in a minute and isolated instances where that value was above 80 steps taken in a minute. That indicated a high likelihood for an individual to be on a walk [27]. Then, we isolated the values of heart rate corresponding to those instances that we named walking heart rate. To determine the minimal number of samples of walking heart rate (eg, the number of data points of heart rate corresponding to our definition of walking heart rate) needed to converge to a confident average, the standard error was calculated for an increasing random number of samples. A threshold of 1 bpm (beats per minute) was set for convergence (Figure 1). Once the minimum number of samples necessary for each individual was found, the number of days needed to get this amount of samples was extracted.

XSL•FO
RenderX

Figure 1. Evolution of average steps taken in a minute (A) and standard error (in steps) (B and C) with an increase in the number of samples for a representative subject. (C) is the expansion of (B) to visualize the number of samples necessary for this individual to obtain a standard error below 2. bpm: beats per minute.



Ethical Considerations

All datasets were generated by teams at Michigan Medicine at the University of Michigan, and the different data collection protocols were each approved by the University of Michigan institutional review board, under the review numbers HUM00186436, HUM00176584, HUM00184455, and HUM00185391. All participants provided informed consent prior to their participation in the study. The details of the compensation for each study are provided in Table 1.

XSL•FO
RenderX

Results

Wear Time

The average wear time of the device varied by population (Figure 2). The most compliant sample was the HCT caregivers with an average daily wear time of 20.8 hours. The least compliant sample was the pediatric oncology patients with an average daily wear time of 9.8 hours. Differences were also observed in the spread of the data. For the pediatric oncology patients and caregivers and the student sample, the distribution went from 0% to above 90% of wear time. A 0% wear time indicated that participants received the Fitbit but never wore it.

There were significant differences in the average wear time between populations (Table 2). Notably, the pediatric oncology patients had significantly lower wear time than all other studied samples except for the pediatric oncology caregivers. The HCT caregivers also had significantly higher wear time than the other groups, except for the SCI and HD caregivers.

A decrease was observed in monthly wear time for all recruited groups. The average decrease was a little more than 10%, with the highest decrease recorded for the pediatric oncology patients (−19%) and the lowest decrease recorded for the HCT caregivers (−8%).

Figure 2. Distribution of average daily wear time in percentages (left) and hours (right) for each recruited sample. The pediatric oncology patients and caregivers population, as well as the student population, had participants that were given a Fitbit and never wore it. As such, there are instances with 0 hours of daily wear time in the distributions. HCT: hematopoietic cell transplantation; HD: Huntington disease; SCI: spinal cord injury.

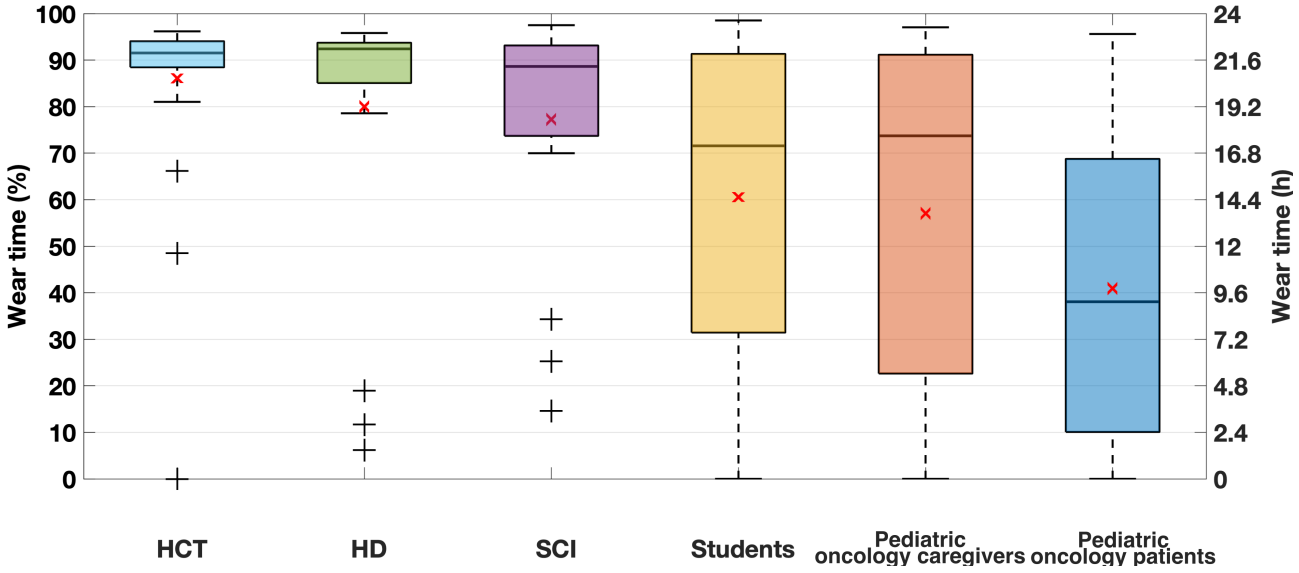


Table . Model coefficients and significance from post hoc test evaluating the difference in wear time between samples.

	HCT[a]	HD[b]	SCI[c]	Students	Pediatric oncology caregivers
HD	0.52	—[d]	—	—	—
SCI	1.04	0.5	—	—	—
Students	4.23[e]	2.88	2.04	—	—
Pediatric oncology caregivers	3.68[e]	2.7	2.02	0.52	—
Pediatric oncology patients	5.83[e]	4.65[e]	3.91[e]	3.96[e]	2.54

[a]HCT: hematopoietic cell transplantation.

[b]HD: Huntington disease.

[c]SCI: spinal cord injury.

[d]Not applicable.

[e]$P<.001$.

Aim 1

Population Level

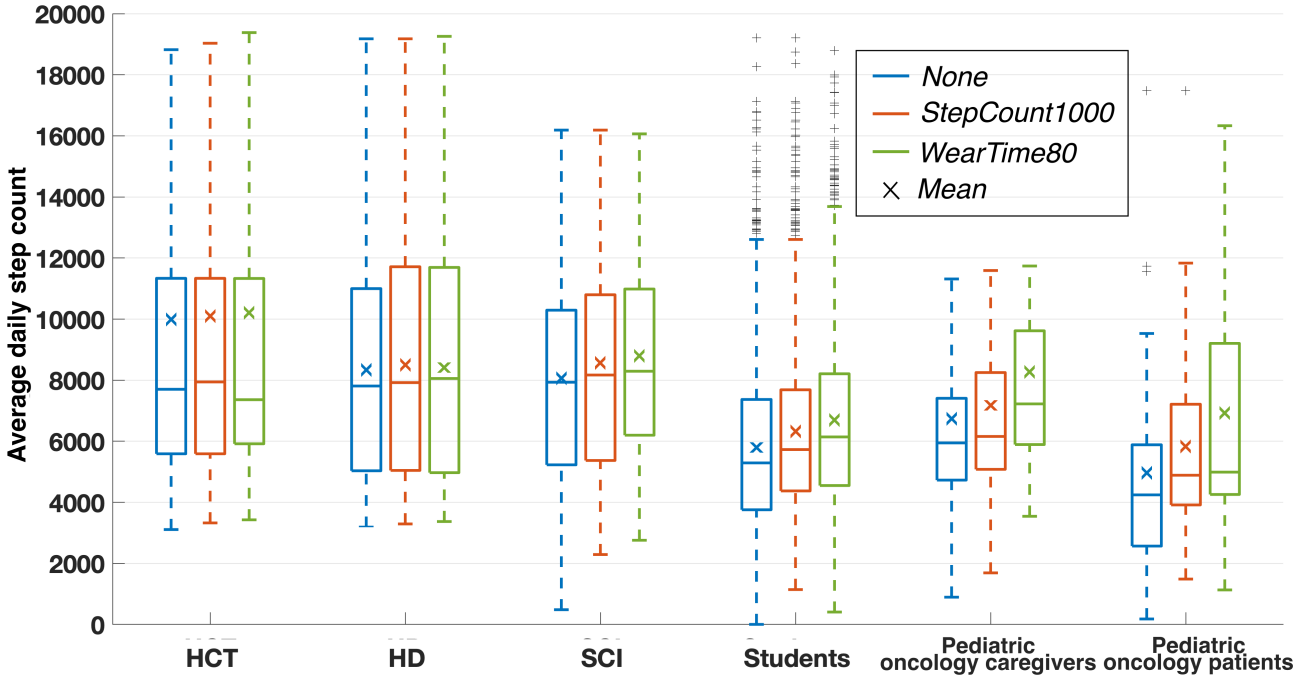
First, we observed the changes in the sample average daily step count for the 3 chosen definitions of a valid day (Figure 3). The average number of valid days decreased between 15% and 33% from the definition *None* to *WearTime80*. The decrease from *None* to *StepCount1000* was around 12% for the SCI caregivers and the pediatric oncology patients and close to 0% for the other samples. There was also a decrease from *StepCount1000* to *WearTime80* from 12% to 21%. The change in average daily

XSL•FO
RenderX

step count was dependent on the sample. HCT and HD caregivers both showed a difference smaller than 15 steps between *None* and *WearTime80*. On the other hand, the pediatric oncology patients showed a difference of almost 2000 steps between *None* and *WearTime80*. The sample sizes decreased from *None* to *WearTime80* for the pediatric oncology patients, caregivers, and students (n=6, −14%; n=10, −20%; and n=7, −8%; respectively).

We also analyzed changes in the sample average daily heart rate for the 3 definitions. We found only small differences in average daily heart rate for all samples. The student sample showed the largest difference of 3 bpm between *None* and *WearTime80*.

Figure 3. Average daily step count across samples for the different definitions of a valid day. The effect of these different definitions was dependent on the sample, going from less than 80 steps for the HD sample to almost 2000 steps for the pediatric oncology patients. HCT: hematopoietic cell transplantation; HD: Huntington disease; SCI: spinal cord injury.



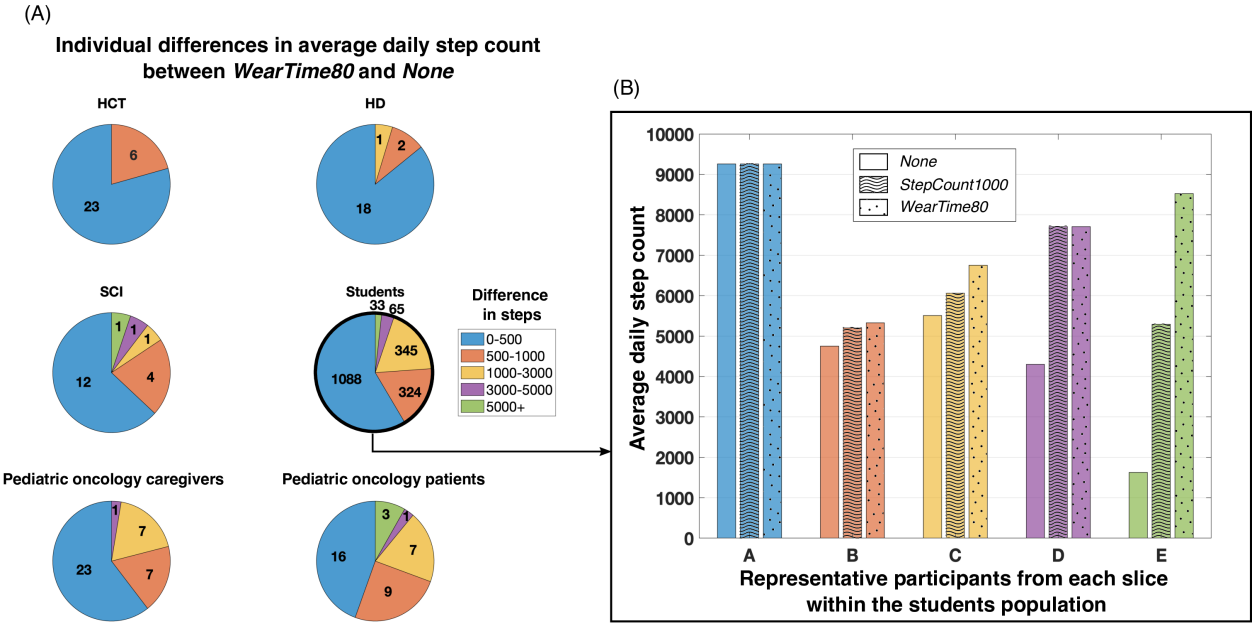
Individual Level

We assessed the changes in each individual's average daily step count (Figure 4). We binned individuals based on the average daily step count difference between *None* and *WearTime80*, since those definitions were the least and most stringent, respectively (Figure 4A). For all samples, the majority of individuals had a difference in average daily step count between 0 and 500 steps. The HCT caregivers had the least differences in individual steps, followed by the HD caregivers. Approximately, a quarter of the SCI caregivers, pediatric oncology patients and caregivers, and students had a step count difference higher than 1000. By taking a closer look at the individuals in each bin for a group (Figure 4B), we saw that participants like participant A did not display any difference in average daily step count between each definition. However, participants like participant E showed a difference higher than 3000 steps between each definition, with that difference reaching almost 7000 steps between *None* and *WearTime80*.

Little to no variation was observed within participants for the average daily heart rate between *None* and *WearTime80*. Only 3% (n=1) of the pediatric oncology patient participants had a difference between 5 and 10 bpm, and less than 1% (n=21) of students had a difference between 10 and 15 bpm.

XSL•FO
RenderX

Figure 4. (A) Individual differences in average daily step count calculated using *None* versus *WearTime80*. For example, 21% (n=6) of the HCT sample had a difference in average daily step count between 500 and 999 steps when calculations were made using *WearTime80* versus *None*. The grouping of the step differences was inspired by the physical activity level categories presented by Tudor-Locke and Bassett [28]. (B) Five representative participants' average daily step count from the different slices of difference in step count, using the 3 definitions of a valid day. For example, participant E was representative of students who had a difference in average daily step count larger than 5000 when calculated with *WearTime80* compared to *None*. HCT: hematopoietic cell transplantation; HD: Huntington disease; SCI: spinal cord injury.

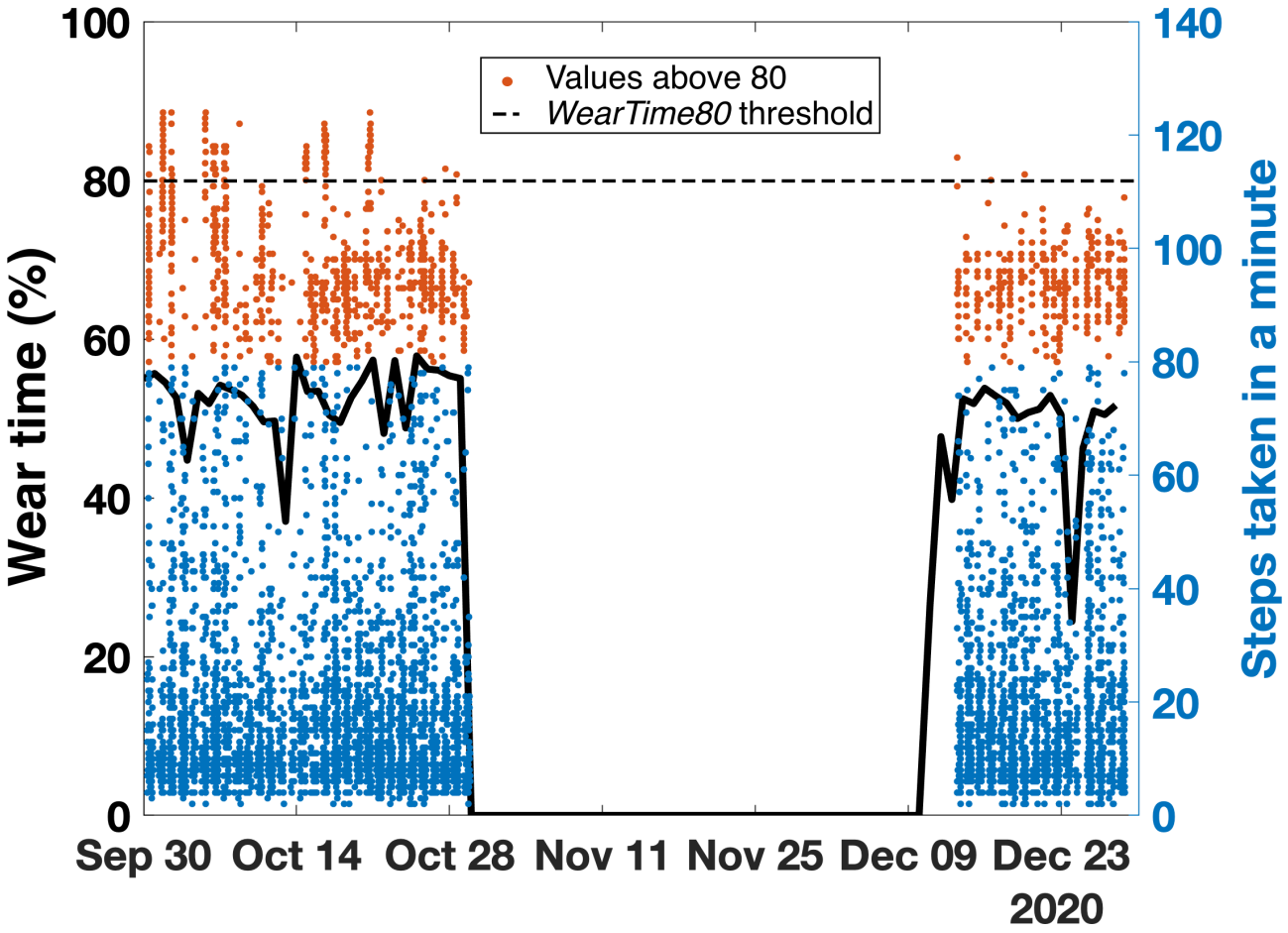


Aim 2

On average, across all population samples, 304 samples (minutes of walking) were needed to obtain an estimate of the walking heart rate with a standard error of 1 bpm. For most individuals, the minimal number of samples was reached in under 26 days of data collection. For the pediatric oncology caregivers, 1 participant did not have enough samples to obtain a confident estimate of the walking heart rate, but an estimate of that caregiver's average daily step count using *WearTime80* was obtained. On the other hand, 5 (17%) HCT, 3 (14%) HD, 2 (11%) SCI, 186 (10%) students, and 3 (7%) of the pediatric oncology patients did not have any valid days to calculate their average daily step count but presented enough samples to confidently obtain their walking heart rate. Figure 5 illustrates an individual's data from the student sample within that 10%. That individual wore the Fitbit at the beginning and the end of the study for around 50% of the day. That student had 1314 samples of steps taken in a minute—enough to estimate walking heart rate—while presenting no days with a wear time higher than 80%.

XSL•FO
RenderX

Figure 5. Wear time (black line, left axis) and steps taken in a minute (blue and orange dots, right axis) over time for an individual from the student sample. Average daily step count using *WearTime80* with this participant was not possible as no days presented a wear time larger than 80%. However, there was a sufficient number of samples to confidently estimate walking heart rate based on the number of samples of steps taken in a minute that was above 80.



Discussion

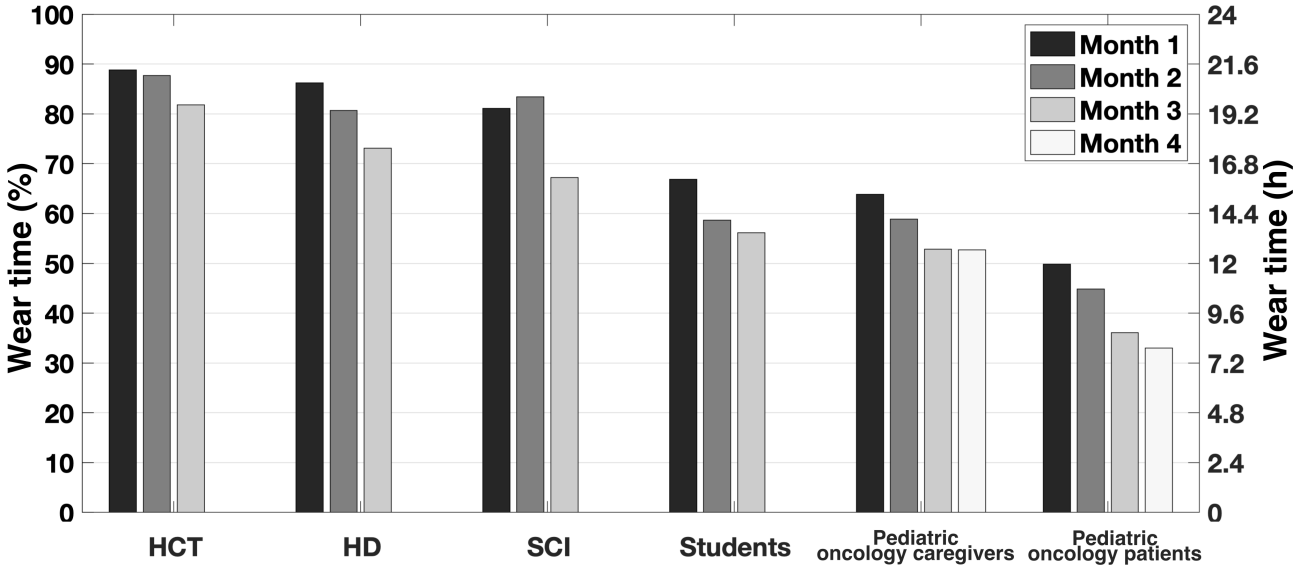
Principal Results

The purpose of this study was to quantify how wear time could impact parameter estimates in studies using consumer-grade wearables by using real-world samples with varying participant compliance. This work is distinguished from the literature by demonstrating the conditional nature of wear time's influence in consumer-grade wearables' data, which varies significantly across research samples and research questions. By leveraging a broad array of datasets, including vulnerable patient populations, we showed evidence of the nuanced effect of wear time, highlighting its critical role in research using wearables, especially in samples with inconsistent wear patterns such as the pediatric oncology caregiver and patient participants used in this study. It is important to note that the pediatric oncology patients' ages ranged from 5 to 20 years, with a mean of 11.9 years. The young age combined with illness for these subjects likely explained this population sample's low level of wear time with large variability.

Unsurprisingly, implementing different wear time requirements to define a valid day can significantly impact parameter estimates, especially in samples with low compliance. Wear time compliance for the HD and HCT caregivers was high (Figure 6); therefore, the estimates of the samples' average daily step counts were minimally affected by different definitions of a valid day (Figure 3). In comparison, the pediatric oncology patient sample had low wear time compliance (Figure 6); therefore, the estimate of the sample average daily step count was significantly impacted by different definitions of a valid day (Figure 3). Importantly, our analysis quantifies these differences so that other researchers can better understand the potential implications of different valid day definitions. The definition of a valid day had almost no impact on estimates of the average daily heart rate across samples. This also is not surprising as heart rate is a more stable measure that does not necessarily correlate with wear time; any biases from missing data are averaged out.

XSL•FO
RenderX

Figure 6. Monthly average daily wear time in percentages (left) and hours (right). We observed a monthly decrease in wear time for all groups symptomatic of the phenomenon of wearables' abandonment. The pediatric oncology dataset was collected over 120 days versus 90 days for the other datasets. HCT: hematopoietic cell transplantation; HD: Huntington disease; SCI: spinal cord injury.



A significant novelty in this work is demonstrating the flexibility of data usage in real-world settings (ie, outside of the hospital). In fact, we showed that even if an individual's data were not sufficient to answer a specific research question, those data could potentially still be leveraged toward a different analysis (eg, for the estimate of walking heart rate as shown in this study). In particular, the variations in the research question time frame (eg, daily, nighttime, and weekly) can lead to different levels of required compliance. For example, to accurately estimate the average number of steps per day, we need a sufficient number of days with high wear time. However, for the example we used in Aim 2—the estimation of walking heart rate—we need a sufficient number of steps taken in a bout of walking by an individual. As such, we show in our example how an individual can have a large amount of walking data throughout an entire study while not having enough individual days with high wear time (Figure 5). These findings suggest that even samples with low compliance may still be useful for answering research questions. In this case, even a day with low wear time (ie, not a valid day) may contain several valid walks useful for estimating average walking heart rate.

Secondary Findings

We also found that depending on whether one is interested in studying a population as a whole or the individuals within a population, the impact of wear time differed. For instance, in the student population, we observed only a small change in the estimated average daily step count in the population with the different methods of analysis (Figure 3). Depending on the objective of a study, that change could be negligible. However, if research is conducted to design a physical activity intervention or to evaluate the efficacy of an intervention, for example, the processing of an individual's data without the inclusion of wear time could lead to critical errors (Figure 4). In particular, if one considers the standard classification of participants into levels of physical activity (0 - 5000 daily steps: sedentary; 5000 - 7500: low active; 7500 - 10,000: somewhat active; 10,000+: highly active) [28], those errors can induce a misclassification. Ultimately, the intervention design or the evaluation of the intervention could be erroneous.

We quantified the phenomenon of wearables' abandonment within our different populations (Figure 6). That phenomenon can potentially create a bias in results over time. For example, a monthly decrease in the population sample level of physical activity can be explained by an actual decline in people's movements or by the fact that participants were not wearing their sensors as much. An appropriate quantification of wear time in the analysis can potentially remove that bias.

Comparison With Prior Work

While the literature on accelerometer usage presents various methodologies for accounting for wear time, it is essential to acknowledge that raw acceleration data or activity counts, typical outputs from accelerometers, are not available from consumer-grade wearables. Therefore, the strategies for considering wear time in the context of these devices necessitate their own approach. Consumer-grade wearables often offer summarized or processed data, which requires tailored methods for wear time analysis to ensure accurate and meaningful interpretation in research contexts. In general, research data on the analysis of wear time and compliance with consumer-grade wearables are sparse in the literature. Often, studies using consumer-grade wearables, such as smartwatches, rely on threshold-based methodologies for processing wear time [13-16]. However, these thresholds are often implemented without sufficient justification based on the specific dataset being analyzed and the research objectives.

Claudel et al [17] developed a novel method to calculate wear time and determine a valid assessment day. Their proposed method calculated wear time using heart rate detection, step count, and the range of heart rate values. A valid day was defined as a day with more than 10 hours of wear time, with the removal of sleeping hours (from 11 PM to 5 AM). Claudel and colleagues present a structured and detailed methodology that can potentially be replicated for other studies. It highlights

XSL•FO
RenderX

the importance of nighttime in wear time calculations and how researchers can conduct a sensitivity analysis to determine the impact of different nighttime limits on the outcome variables. Our study also demonstrates the potential impact of different calculations of wear time on parameter estimates. Both studies illustrate the importance of considering how the definition of a valid day may impact sample parameter estimates.

Studies have explored the factors that are related to compliance levels [13]. The NetHealth study [13] collected data from a student population using Fitbit devices and showed that some personality traits (eg, extraversion and openness) and Fitbit metrics (eg, minutes asleep and sedentary minutes) were related to compliance levels. Understanding the underlying reasons that may lead an individual to be more or less compliant can be useful in improving the overall wear time levels in these types of studies.

Another, perhaps more, direct way to increase compliance is through the use of incentives. There is a range of incentive models one can use to maintain compliance in longitudinal studies [29]. Although improving compliance allows for parameter estimates to be less sensitive to wear time, it is difficult to imagine a "one size fits all" incentive model that would work on every single participant of a study. Thus, the analysis of wear time is likely to remain a crucial step in the analysis of consumer-grade wearables data to ensure accurate parameter estimates.

Limitations

In the accelerometer literature, raw outputs from the sensors are most often used to determine wear time [10,11], which removes a layer of potential error. The outputs provided by Fitbit are processed (as opposed to raw) with unknown algorithms. Here, we assumed those outputs, and in particular minute-by-minute heart rate, were reliable and could be used to accurately estimate wear time. Claudel et al [17] formulated a more complex wear time calculation algorithm fusing both heart rate detection and heart rate values revealing a potential overestimation of wear time when using heart rate detection only. Determining the most accurate estimate of wear time was beyond the scope of this study. However, further research should be conducted to validate methods for the calculation of wear time using the processed outputs of consumer-grade wearables as it may influence parameter estimates.

We assumed in Aim 2 that wear time was random. In other words, we assumed that participants did not choose a particular moment in their day to wear the device, which would bias the estimates of walking heart rate. However, this does not influence the conclusion of Aim 2, as we primarily focused on the amount of data needed to estimate a parameter with confidence, and not whether we could accurately estimate true walking heart rate.

It is important to note that the different datasets we used came from 2 different Fitbit devices (Fitbit Charge and Fitbit Inspire 2). It is possible that the versions of software and hardware are different between these 2 devices. Since the software information is proprietary information, we do not know whether the model differences impacted the used outputs. However, we expect and assume that the measurements between the 2 models are comparable and did not have a significant impact on our analysis.

Additionally, although we analyzed samples from different populations, our results may not translate to similar populations. First, each population sample was given a different set of instructions, which influenced their wear time. Second, Fitbit data syncing can be difficult and depends on the software used as well as the Fitbit model itself. Thus, the amount of data retrieved for the same population sample can be variable from one study to another. Finally, the different datasets used in this study were all collected during the COVID-19 pandemic, which may have affected the health-related behaviors of participants. We recommend that researchers conduct an analysis of wear time systematically when using sensing technologies to derive conclusions.

Conclusions

One of the major benefits of leveraging fitness trackers is the potential to gain a unique and objective insight into an individual's natural behavior. However, if the analysis of wear time is not appropriately integrated into the data processing pipeline, that purported insight can become biased and imprecise, driving the results away from reality. Herein, we presented evidence of the critical impact of wear time on parameter estimates and how that impact may vary depending on the population sample, research question, study design (ie, incentives), and participant compliance. Future research could significantly benefit from our findings and methodological examples to incorporate the following recommendations: first, it is essential to conduct a thorough analysis of the wear time of the population in any given study that uses wearables, with particular attention to the data available when using consumer-grade wearables; second, it is equally important to systematically document and provide comprehensive details on that analysis. Our work provides quantitative data on the impact that different valid day definitions can have on sample parameter estimates in samples with varying wear time compliance.

Acknowledgments

The Roadmap projects that leveraged the Fitbit wearable sensors across the pediatric oncology and college student populations were supported by the National Institute of Health and National Heart, Lung, and Blood Institute (grants 1R01HL146354 and K24HL156HL156896), the Edith S Briskin and Shirley K Schlafer Foundation, the University of Michigan Taubman Institute (Taubman Institute Innovation Project), and by a Scientific Research Initiative grant from the Biosciences Initiative of the University of Michigan Biosciences Initiative. The CareQOL JITAI project that used Fitbit wearable sensors across hematopoietic cell transplantation, spinal cord injury, and traumatic brain injury populations was supported by the following grants: R01NR013658 from the National Institutes of Health (NIH)/National Institute of Nursing Research; R01HL146354 and K24HL156896 from the

XSL•FO
RenderX

National Heart, Lung and Blood Institute; and UL1TR002240 from the National Center for Advancing Translational Sciences. We are grateful to the investigators, coordinators, and research associates and assistants who worked on this study, the study participants, and organizations who supported recruitment efforts.

Conflicts of Interest

None declared.

References

1. Harris TJ, Owen CG, Victor CR, Adams R, Ekelund U, Cook DG. A comparison of questionnaire, accelerometer, and pedometer: measures in older people. Med Sci Sports Exerc 2009 Jul;41(7):1392-1402. [doi: 10.1249/MSS.0b013e31819b3533] [Medline: 19516162]
2. Beagle AJ, Tison GH, Aschbacher K, Olgin JE, Marcus GM, Pletcher MJ. Comparison of the physical activity measured by a consumer wearable activity tracker and that measured by self-report: cross-sectional analysis of the health eHeart study. JMIR Mhealth Uhealth 2020 Dec 29;8(12):e22090. [doi: 10.2196/22090] [Medline: 33372896]
3. Wright R, Keith L. Wearable technology: if the tech fits, wear it. J Electron Resour Med Libr 2014 Oct 2;11(4):204-216. [doi: 10.1080/15424065.2014.969051]
4. Lymberis A, Dittmar A. Advanced wearable health systems and applications. IEEE Eng Med Biol Mag 2007;26(3):29-33. [doi: 10.1109/memb.2007.364926] [Medline: 17549917]
5. Tröster G. The agenda of wearable healthcare. Yearb Med Inform 2005;14(1):125-138. [doi: 10.1055/s-0038-1638446] [Medline: 27706307]
6. Bravata DM, Smith-Spangler C, Sundaram V, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
7. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
8. Lazar A, Koehler C, Tanenbaum TJ, Nguyen DH. Why we use and abandon smart devices. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15): Association for Computing Machinery:635-646. [doi: 10.1145/2750858.2804288]
9. Catellier DJ, Hannan PJ, Murray DM, et al. Imputation of missing data when measuring physical activity by accelerometry. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S555-S562. [doi: 10.1249/01.mss.0000185651.59486.4e] [Medline: 16294118]
10. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. How many hours are enough? Accelerometer wear time may provide bias in daily activity estimates. J Phys Act Health 2013 Jul;10(5):742-749. [doi: 10.1123/jpah.10.5.742] [Medline: 23036822]
11. Rich C, Geraci M, Griffiths L, Sera F, Dezateux C, Cortina-Borja M. Quality control methods in accelerometer data processing: defining minimum wear time. PLoS One 2013;8(6):e67206. [doi: 10.1371/journal.pone.0067206] [Medline: 23826236]
12. McCarthy M, Bury DP, Byrom B, Geoghegan C, Wong S. Determining minimum wear time for mobile sensor technology. Ther Innov Regul Sci 2021 Jan;55(1):33-37. [doi: 10.1007/s43441-020-00187-3] [Medline: 32588369]
13. Faust L, Purta R, Hachen D, et al. Exploring compliance: observations from a large scale Fitbit study. In: Proceedings of the 2nd International Workshop on Social Sensing (SocialSens'17): Association for Computing Machinery:55-60. [doi: 10.1145/3055601.3055608]
14. Hardcastle SJ, Jiménez-Castuera R, Maxwell-Smith C, Bulsara MK, Hince D. Fitbit wear-time and patterns of activity in cancer survivors throughout a physical activity intervention and follow-up: exploratory analysis from a randomised controlled trial. PLoS One 2020;15(10):e0240967. [doi: 10.1371/journal.pone.0240967] [Medline: 33075100]
15. Giurgiu M, Timm I, Becker M, et al. Quality evaluation of free-living validation studies for the assessment of 24-hour physical behavior in adults via wearables: systematic review. JMIR Mhealth Uhealth 2022 Jun 9;10(6):e36377. [doi: 10.2196/36377] [Medline: 35679106]
16. Giurgiu M, Kolb S, Nigg C, et al. Assessment of 24-hour physical behaviour in children and adolescents via wearables: a systematic review of free-living validation studies. BMJ Open Sport Exerc Med 2022;8(2):e001267. [doi: 10.1136/bmjsem-2021-001267] [Medline: 35646389]
17. Claudel SE, Tamura K, Troendle J, et al. Comparing methods to identify wear-time intervals for physical activity with the Fitbit Charge 2. J Aging Phys Act 2021 Jun 1;29(3):529-535. [doi: 10.1123/japa.2020-0059] [Medline: 33326935]
18. Carlozzi NE, Choi SW, Wu Z, et al. An app-based just-in-time adaptive self-management intervention for care partners (CareQOL): protocol for a pilot trial. JMIR Res Protoc 2021 Dec 9;10(12):e32842. [doi: 10.2196/32842] [Medline: 34889775]
19. Cislo C, Clingan C, Gilley K, et al. Monitoring beliefs and physiological measures in students at risk for COVID-19 using wearable sensors and smartphone technology: protocol for a mobile health study. JMIR Res Protoc 2021 Jun 4;10(6):1-9. [doi: 10.2196/29561] [Medline: 34115607]
20. Koblick SB, Yu M, DeMoss M, et al. A pilot intervention of using a mobile health app (ONC Roadmap) to enhance health-related quality of life in family caregivers of pediatric patients with cancer. Mhealth 2023;9:5. [doi: 10.21037/mhealth-22-24] [Medline: 36760786]

XSL•FO
RenderX

21. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952 Dec;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
22. Dunn OJ. Multiple comparisons among means. J Am Stat Assoc 1961 Mar;56(293):52-64. [doi: 10.1080/01621459.1961.10482090]
23. Fitbit data processing. GitHub. URL: https://github.com/Loubna-brd/Fitbit_data_processing [accessed 2025-01-29]
24. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul;47(7):1303-1315. [doi: 10.1007/s40279-016-0663-1]
25. Kraus WE, Janz KF, Powell KE, et al. Daily step counts for measuring physical activity exposure and its relation to health. Med Sci Sports Exerc 2019 Jun;51(6):1206-1212. [doi: 10.1249/MSS.0000000000001932] [Medline: 31095077]
26. Wang C, Lizardo O, Hachen DS. Using Fitbit data to monitor the heart rate evolution patterns of college students. J Am Coll Health 2022 Apr 3;70(3):875-882. [doi: 10.1080/07448481.2020.1775610]
27. Tudor-Locke C, Rowe DA. Using cadence to study free-living ambulatory behaviour. Sports Med 2012 May;42(5):381-398. [doi: 10.2165/11599170-000000000-00000]
28. Tudor-Locke C, Bassett DR. How many steps/day are enough? Sports Med 2004;34(1):1-8. [doi: 10.2165/00007256-200434010-00001]
29. Jaimes LG, Vergara-Laurens IJ, Raij A. A survey of incentive techniques for mobile crowd sensing. IEEE Internet Things J 2015;2(5):370-380. [doi: 10.1109/JIOT.2015.2409151]

Abbreviations

bpm: beats per minute
HCT: hematopoietic cell transplantation
HD: Huntington disease
SCI: spinal cord injury

Edited by L Buis; submitted 31.01.23; peer-reviewed by A Martinko, MJ Toledo, YC Wang; revised version received 02.09.24; accepted 20.12.24; published 21.03.25.

Please cite as:
Baroudi L, Zernicke RF, Tewari M, Carlozzi NE, Choi SW, Cain SM
Using Wear Time for the Analysis of Consumer-Grade Wearables' Data: Case Study Using Fitbit Data
JMIR Mhealth Uhealth 2025;13:e46149
URL: https://mhealth.jmir.org/2025/1/e46149
doi:10.2196/46149

© Loubna Baroudi, Ronald Fredrick Zernicke, Muneesh Tewari, Noelle E Carlozzi, Sung Won Choi, Stephen M Cain. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

An Actor-Partner Interdependence Mediation Model for Assessing the Association Between Health Literacy and mHealth Use Intention in Dyads of Patients With Chronic Heart Failure and Their Caregivers: Cross-Sectional Study

Xiaorong Jin[1,2], MSN; Yimei Zhang[2], BSN; Min Zhou[2], BSN; Qian Mei[3], BSN; Yangjuan Bai[4], MSN; Qiulan Hu[5], MSN; Wei Wei[6], MSN; Xiong Zhang[2], MSN; Fang Ma[2], PhD

[1]Department of General Surgery I, The People's Hospital of ChuXiong Yi Autonomous Prefecture, ChuXiong, Yunnan Province, China

[2]Department of Nursing, The First Affiliated Hospital of Kunming Medical University, 295 Xichang Road, Kunming City, Yunnan Province, China

[3]Coronary Heart Disease Center, Fuwai Yunnan Cardiovascular Hospital, Kunming City, Yunnan Province, China

[4]Cardiology Department, The First Affiliated Hospital of Kunming Medical University, Kunming City, Yunnan Province, China

[5]Department of Geriatric Intensive Medicine, The First Affiliated Hospital of Kunming Medical University, Kunming City, Yunnan Province, China

[6]Digestive Surgery Department, The First Affiliated Hospital of Kunming Medical University, Kunming City, Yunnan Province, China

Corresponding Author:
Fang Ma, PhD
Department of Nursing, The First Affiliated Hospital of Kunming Medical University, 295 Xichang Road, Kunming City, Yunnan Province, China

Abstract

Background: Chronic heart failure (CHF) has become a serious threat to the health of the global population. Self-management is the key to treating CHF, and the emergence of mobile health (mHealth) has provided new ideas for the self-management of CHF. Despite the many potential benefits of mHealth, public utilization of mHealth apps is low, and poor health literacy (HL) is a key barrier to mHealth use. However, the mechanism of the influence is unclear.

Objective: The aim of this study is to explore the dyadic associations between HL and mHealth usage intentions in dyads of patients with CHF and their caregivers, and the mediating role of mHealth perceived usefulness and perceived ease of use in these associations.

Methods: This study had a cross-sectional research design, with a sample of 312 dyads of patients with CHF who had been hospitalized in the cardiology departments of 2 tertiary care hospitals in China from March to October 2023 and their caregivers. A general information questionnaire, the Chinese version of the Heart Failure-Specific Health Literacy Scale, and the mHealth Intention to Use Scale were used to conduct the survey; the data were analyzed using the actor-partner interdependence mediation model.

Results: The results of the actor-partner interdependent mediation analysis of HL, perceived usefulness of mHealth, and mHealth use intention among patients with CHF and their caregivers showed that all of the model's actor effects were valid (β=.26 - 0.45; *P*<.001), the partner effects were partially valid (β=.08 - 0.20; *P*<.05), and the mediation effects were valid (β=.002 - 0.242, 95% CI 0.003 - 0.321; *P*<.05). Actor-partner interdependent mediation analyses of HL, perceived ease of use of mHealth, and mHealth use intention among patients with CHF and caregivers showed that the model's actor effect partially held (β=.17 - 0.71; *P*<.01), the partner effect partially held (β=.15; *P*<.01), and the mediation effect partially held (β=.355 - 0.584, 95% CI 0.234 - 0.764; *P*<.001).

Conclusions: Our study proposes that the HL of patients with CHF and their caregivers positively contributes to their own intention to use mHealth, suggesting that the use of mHealth by patients with CHF can be promoted by improving the HL of patients and caregivers. Our findings also suggest that the perceived usefulness of patients with CHF and caregivers affects patients' mHealth use intention, and therefore patients with CHF and their caregivers should be involved throughout the mHealth development process to improve the usability of mHealth for both patients and caregivers. This study emphasizes the key role of patients' perception that mHealth is easy to use in facilitating their use of mHealth. Therefore, it is recommended that the development of mHealth should focus on simplifying operational procedures and providing relevant operational training according to the needs of the patients when necessary.

(JMIR Mhealth Uhealth 2025;13:e63805) doi:10.2196/63805

XSL•FO
RenderX

KEYWORDS

chronic heart failure; caregivers; health literacy; mHealth; actor-partner interdependence mediation model; mobile health

Introduction

Heart failure is a condition of signs and symptoms triggered by structural or functional abnormalities of the heart, as evidenced by elevated natriuretic peptide levels and cardiogenic, pulmonary, or systemic congestion [1]. Chronic heart failure (CHF) manifests as a persistent state of heart failure that may be stable, worsening, or decompensated [2]. As a severe or advanced manifestation of a wide range of cardiovascular diseases, CHF is known for its high morbidity and mortality and has become an important public health problem threatening the health and lives of the global population [3]. According to the Global Heart Failure Survey, the average incidence of CHF reaches 460/100,000 persons/year, while the 5- and 10-year survival rates of patients after diagnosis are only 57% and 35%, respectively [4,5]. More worryingly, approximately 60% of patients with CHF die within 5 years of diagnosis [6,7].

Studies have shown that self-management is important for patients with CHF to improve cardiac function, improve disease prognosis, and reduce mortality [8]. With the increasing popularity of smartphones and wearable devices, mobile health (mHealth) has gradually become a new means to support self-management in patients with CHF [9]. mHealth refers to the integrated application of mobile phones and other wireless technologies in medical practice, and its functions cover a wide range of aspects such as booking appointments, health information inquiry, and vital signs monitoring [9]. Although the role of mHealth in facilitating self-management in patients with CHF has been widely recognized, the actual usage of mHealth by patients is generally low [10,11].

Health literacy (HL) refers to an individual's ability to access, understand, process, and utilize health information to promote one's health [12]. As information largely moves to a digital medium, HL takes on an added dimension in supporting the skills needed to understand online health information, which makes digital health a significant component of HL [13-15]. It has been evidenced that digital HL is an important indicator of mHealth use intention, indicating that HL is an important factor influencing mHealth use intention [16]. On the other hand, mHealth use intention would reinforce individual motivation and confidence to broadly develop their literacy about digital health and HL as well [17]. Additionally, the study indicates that individuals who intend to seek health-related information or advice on the internet exhibit higher digital HL levels, suggesting that mHealth use intention could contribute to an increase in HL [18]. Existing studies have shown that HL is positively associated with mHealth use intention and behavior. Specifically, the higher the level of HL of patients, the more likely they are to adopt and use mHealth technologies, and vice versa [19,20]. Although the association between HL and willingness to use mHealth has been reported several times, the exact mechanism remains to be further explored. According to the Technology Acceptance Model (TAM) theory, users' actual adoption or rejection of a technology such as mHealth is primarily affected by 2 factors, namely perceived usefulness and perceived ease of use [21]. The same researchers determined that perceived ease of use and perceived usefulness of telemedicine services have a substantial effect on the telemedicine behavioral intention of older patients [21]. Additionally, it has been evidenced that HL can influence the perceived usefulness and perceived ease of use of mHealth [22,23]. Hence, perceived usefulness and perceived ease of use of mHealth might act as mediators of HL and mHealth use intention.

In addition, previous studies have pointed out that caregivers play an integral role in the self-management of patients with CHF [24,25]. Cheng et al suggested that caregivers bear 70% of the responsibility for chronic disease management [26]. According to Interdependence Theory, in a dyadic relationship, each person has the ability to influence the outcomes of the other, suggesting partners affect each other's motives, preferences, behaviors, and health outcomes [27,28]. Patients with CHF and their caregivers constitute a dyad, and the attributes of one dyad member can influence the motive and behavior of the other [26], hence, patients' and caregivers' HL might influence the partners' mHealth use intentions. The social influence variable in the Unified Theory of Acceptance and Use of Technology model also suggests that a patient's willingness to adopt a technology may be influenced by the perceptions of important people around them, such as caregivers [29,30]. However, current research has mostly explored mHealth use intentions from the single perspective of either the patient or the caregiver, ignoring the interactions between the two.

Therefore, this study explored in depth the underlying mechanism linking HL and mHealth use intention of patients with CHF and their caregivers from a dyad perspective, which can help provide targeted intervention for future mHealth use intention studies, thus promoting the use of mHealth among patients with CHF.

Methods

Design and Sample

This study adopted a cross-sectional research design with a convenience sampling method. Patients with CHF admitted to the cardiology departments of 2 tertiary care hospitals in Yunnan Province, China, from March to October 2023 and their caregivers were selected for the study. The inclusion criteria for patients with CHF were as follows: (1) age ≥18 years old; (2) compliance with the 2022 AHA/ACC/HFSA Guidelines for the Management of Heart Failure [2] diagnostic criteria for CHF; (3) meeting New York Heart Association (NYHA) cardiac function class I-III; (4) at least one caregiver; (5) being able to complete the questionnaire independently or under the guidance of the researchers; and (6) voluntarily participating in this study. The exclusion criteria for patients with CHF were as follows: (1) acute myocardial infarction or acute pulmonary embolism in the past month; (2) with other critical diseases, such as malignant tumor, renal failure, respiratory failure; (3) not using smartphones; (4) cognitive disorders or psychiatric disorders;

XSL•FO
RenderX

and (5) being involved in other research projects. Caregiver inclusion criteria were as follows: (1) informal caregivers who are the primary caregiver and provide care and support to the patient without compensation [26,31]; (2) age ≥18 years old; (3) being able to complete the questionnaire independently or under the guidance of the researchers; and (4) voluntarily participating in the study. Caregiver exclusion criteria were as follows: (1) cognitive impairment or mental illness; (2) serious physical disease, such as cancer or vital organ failure; (3) not using smartphones; and (4) participating in other studies. According to Ledermann et al [32], in the actor-partner interdependence model (APIM), a sample size of approximately 93 to 241 dichotomies is recommended, while for the actor-partner interdependence mediation model (APIMeM), a sample size of approximately 120 dichotomies is required for good mediation. Of 367 eligible patient-caregiver dyads, 31 were excluded from enrollment because one member of the dyad refused to participate, and 24 were excluded from the analysis because of missing data. Thus, 312 dyads were included in the final data analysis. This sample size met the requirement for structural equation modeling and APIMeM.

Data Collection

Before the survey, all of the researchers received unified training to familiarize themselves with the content of the questionnaire scale and to master the way of filling it out and the precautions to be taken. Data collection was carried out by patients with CHF and their caregivers alone, with the researcher providing guidance but not interfering with the study subjects. After the survey was completed, researchers checked the data on the spot; if they found any missing or wrongly filled items, they verified the data with the research subjects immediately and assisted them in completing the data.

Measures

Demographic and Clinical Characteristics

A questionnaire, which was designed by the researchers based on a large amount of literature, was used in the survey, and it included patients' information such as gender, age, literacy level, marital status, smoking history, drinking history, duration of heart failure, number of hospitalizations, NYHA cardiac function classification, previous history of cardiovascular disease, left ventricular ejection fraction, and BMI, as well as the caregiver's information such as gender, age, relationship with the patient, education level, occupation, marital status, monthly income, total caregiving time, caregiving time per day, and illness.

Health Literacy

The Heart Failure–Specific Health Literacy Scale (HF-SHLS) was used to measure HL; it was initially developed by Matsuoka et al [33] and Chineseized by Yue Meng et al [34] in 2016. The HF-SHLS consists of 3 dimensions with 12 entries, namely functional HL, interactive HL, and critical HL; each entry ranks from "not applicable" (score of 1) to "very applicable" (score of 4), and the total score ranges between 12 - 48, with a higher score representing a higher level of HL. The Cronbach values of the overall scale and 3 dimensions were 0.87, 0.84, 0.72, and 0.69, respectively [33,34], with acceptable internal consistency, and the scale was used with permission from the original authors.

mHealth Use Intention

In this study, "mHealth" refers to the integrated application of smartphones in medical practice, including activities such as booking appointments and inquiring about health information. The mHealth Service Use Intention Questionnaire, which was prepared by Shi Lin [35] was used to measure the perceived usefulness of mHealth, perceived ease of use of mHealth, and mHealth use intentions, which consists of 7 dimensions and 26 entries. Cronbach values for each dimension ranged from 0.822 - 0.968. In this study, the 3 dimensions measuring the perceived usefulness of mHealth, perceived ease of use of mHealth, and mHealth use intentions were used, with Cronbach values of each dimension being 0.904, 0.846, and 0.968, respectively. The indicators in the questionnaire are scored on a 5-point Likert scale, with scores of 1 - 5 representing strongly disagree, disagree, undecided, agree, and strongly agree; a higher score in the related dimension means a higher level of perceived usefulness of mHealth, perceived ease of use of mHealth, and mHealth use intentions. The scale was used with permission from the original authors. For the detailed items and response options of this questionnaire, please refer to Multimedia Appendix 1.

Statistical Analyses

SPSS (version 26.0; IBM Corp) and AMOS (version 26.0; IBM Corp) were used for data analysis. A value of $P<.05$ based on a 2-tailed test was considered statistically significant. Data were tested for normality (results are detailed in Multimedia Appendix 2), which was based on the Kolmogorov-Smirnov test when the sample size was greater than 50, and when $P<.05$, it was considered as not conforming to normal distribution [36]. The results showed that the data in this study did not conform to normal distribution. Thus, we used median and interquartile spacing to describe the skewed distribution. The Wilcoxon signed-rank sum test was used to detect differences in scores for HL, perceived usefulness of mHealth, perceived ease of use of mHealth, and mHealth use intention among patients with CHF and their caregivers. Spearman correlation analysis was used to test the relationship between HL, mHealth perceived usefulness, and perceived ease of mHealth in the patient-caregiver dyad in the context of CHF. APIMeM was used to construct the APIMeM of HL-perceived usefulness of mHealth-intention to use mHealth in patients with CHF and their caregivers and the APIMeM of HL-perceived ease of use of mHealth-mHealth use intention in patients with CHF and their caregivers. The nonnormal distributions were calculated using the bootstrap method. The fit of each model was assessed using X^2/df (<3), root mean square error of approximation (RMSEA; ≤0.08), comparative fit index (CFI; ≥0.90), and Tucker-Lewis index (TLI; ≥0.90). The mediating effects in APIMeM were tested using bootstrap analyses, with a sample size of 5000 and a confidence interval of 95%. Furthermore, we conducted multigroup APIMeM to analyze whether age and relationship with the patient moderated the dyadic effects. We used a *t* test, Mann-Whitney test, or Kruskal-Wallis test to determine whether there was a significant difference in the path

XSL•FO
RenderX

coefficients between groups based on the normality of the path coefficients of each group; significance tests were set at a 2-sided, $P \leq .05$.

Ethical Considerations

The study was approved by the Ethics Committees of the First Affiliated Hospital of Kunming Medical University (2022-L-304) and Fuwai Cardiovascular Disease Hospital of Yunnan Province (2023-033-01) and was conducted in accordance with the principles of the Declaration of Helsinki. Informed consent was obtained from all participants before the investigation. Participation in the study was entirely voluntary, and no compensation was provided for their involvement. Participant data were deidentified and stored on a local secure server. Further, the participants had the right to withdraw from the study at any time, without giving any reasons. In addition, no identification of individual participants or users in any images occurred in the manuscript or supplementary material.

Results

Characteristics of the Patient-Caregiver Dyads

A total of 312 dyads of patients with CHF and their caregivers were included in the final analysis. The age of patients with CHF ranged from 18 to 89 years old; most of them were male (65.7%), their education level was primary school or below (32%), their NYHA classification was mostly grade II (46.8%), 130 of them had a left ventricular ejection fraction of more than 50% (41.7%), and almost half of the patients' BMIs were in the range of 18.5 - 23.9 (49%). The age of caregivers ranged from 18 to 84 years old; most of them were female (62.2%), their education level was mainly junior high school (25.6%) and bachelor's degree (27.6%), and most of them were the children of patients (46.6%); for 41.35% of the caregivers, the average per capita monthly income was 2000 - 4999 yuan (US $274-$686), and for most caregivers, the total time of caregiving was less than 1 year (76%). Detailed information is shown in Table 1.

XSL•FO
RenderX

Table . General demographic and disease-related information for patients with chronic heart failure and their caregivers.

Variable and clusters		Patients with chronic heart failure (N=312）		Caregivers (N=312）	
		Frequency	Percentage	Frequency	Percentage
Sex					
	Male	205	65.7	118	37.8
	Female	107	34.3	194	62.2
Age group (years)					
	18 - 44	41	13.1	152	48.7
	45 - 59	114	36.5	111	35.6
	60 - 74	107	34.3	39	12.5
	75 - 89	50	16	10	3.2
Education					
	≤Primary school	100	32	32	10.3
	Junior high school	86	27.6	80	25.6
	Senior high school	48	15.4	51	16.3
	Associate degree	33	10.6	55	17.6
	Bachelor’s degree	41	13.1	86	27.6
	≥Master’s degree	4	1.3	8	2.6
New York Heart Association classification					
	I	43	13.8	—[a]	—
	II	146	46.8	—	—
	III	123	39.4	—	—
Left ventricular ejection fraction					
	<30	55	17.6	—	—
	30 - 50	127	40.7	—	—
	>50	130	41.7	—	—
BMI					
	<18.5	25	8	—	—
	18.5 - 23.9	153	49	—	—
	24.0 - 27.9	93	29.8	—	—
	≥28.0	41	13.1	—	—
Relationship of caregiver to patient					
	Parents	—	—	8	2.6
	Spouse	—	—	131	42.0
	Son/daughter	—	—	144	46.2
	Friend	—	—	6	1.9
	Other kinships	—	—	23	7.4
Monthly per capita income (yuan)[b]					
	2000	—	—	64	20.5
	2000 - 4999	—	—	129	41.3
	≥5000	—	—	119	38.1
Duration of caregiving (years)					

XSL•FO
RenderX

Variable and clusters		Patients with chronic heart failure (N=312）		Caregivers (N=312）	
		Frequency	Percentage	Frequency	Percentage
	<1	—	—	237	76.0
	1 - 2.9	—	—	23	7.4
	3 - 5.9	—	—	10	3.2
	≥6	—	—	42	13.5

[a]Not applicable.

[b]A currency exchange rate of 1 yuan=US $0.14 is applicable.

The Scores and Differences in the Study Variables in the Dyadic Group

The median HL score of patients with CHF was 34.00 (IQR27.00-40.00), which was lower than that of their caregivers (median 37.00, IQR 31.00-42.00), with a statistically significant difference between them (*P*<.001). The median (IQR) scores of perceived usefulness, perceived ease of use, and mHealth use intention in patients with CHF were 20.00 (17.00-23.00), 12.00 (6.00-15.00), and 14.00 (11.00-15.00), respectively; the median (IQR) perceived usefulness, perceived ease of use, and mHealth use intention in caregivers were 21.00 (19.00-25.00), 15.00 (12.00-15.00), and 15.00 (14.00-15.00), respectively. The patients' scores were lower than those of carers, with statistically significant differences between them (*P*<.001), as shown in Table 2.

Table . Scores and differences in variables among patients with chronic heart failure and their caregivers.

Dimension	Patients with chronic heart failure (N=312), median (IQR)	Caregivers (N=312), median (IQR)	*Z*-score	*P* value
Health literacy	34.00 (27.00-40.00)	37.00 (31.00-42.00）	4.90	<.001
Perceived usefulness	20.00 (17.00-23.00)	21.00 (19.00-25.00）	4.38	<.001
Perceived ease of use	12.00 (6.00-15.00)	15.00 (12.00-15.00）	7.88	<.001
Mobile health use intentions	14.00 (11.00-15.00)	15.00 (14.00-15.00）	6.86	<.001

Correlation Coefficients of Study Variables

Among patients with CHF, mHealth use intention was positively correlated with their own HL, mHealth perceived usefulness, and perceived ease of use (*r*=0.512 - 0.682, *P*<.05), and positively correlated with the caregiver's mHealth perceived usefulness and mHealth use intention (*r*=0.222 - 0.310, *P*<.05); caregiver's mHealth use intention was positively correlated with their own HL, mHealth perceived usefulness, and perceived ease of use (*r*=0.393 - 0.572, *P*<.01), and positively correlated with patients' HL, mHealth perceived usefulness, perceived ease of use, and mHealth use intention (*r*=0.170 - 0.310, *P*<.01). The results of the correlation analyses are detailed in Table 3.

XSL•FO
RenderX

Table . Correlations between health literacy, perceived usefulness, perceived ease of use, and mobile health use intentions in patients with chronic heart failure and their caregivers.

Variable	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]
1	1							
2	0.432[i]	1						
3	0.717[i]	0.465[i]	1					
4	0.577[i]	0.512[i]	0.682[i]	1				
5	0.211[i]	0.089	0.142[j]	0.094	1			
6	0.144[j]	0.361[i]	0.132[j]	0.222[i]	0.327[i]	1		
7	0.124[j]	0.164[i]	0.130[j]	0.109	0.466[i]	0.438[i]	1	
8	0.170[i]	0.282[i]	0.204[i]	0.310[i]	0.393[i]	0.527[i]	0.572[i]	1

[a]Health literacy of patients with chronic heart failure.

[b]Patients' perceived usefulness of mobile health.

[c]Patients' perceived ease of use of mobile health.

[d]Patients' mobile health use intention.

[e]Caregivers' health literacy.

[f]Caregivers' perceived usefulness of mobile health.

[g]Caregivers' perceived ease of use of mobile health.

[h]Caregivers' mobile health use intention.

[i]*P*<.01.

[j]*P*<.05.

The APIMeM of HL-Perceived Usefulness of mHealth-mHealth Use Intention in Patients With CHF and Caregivers

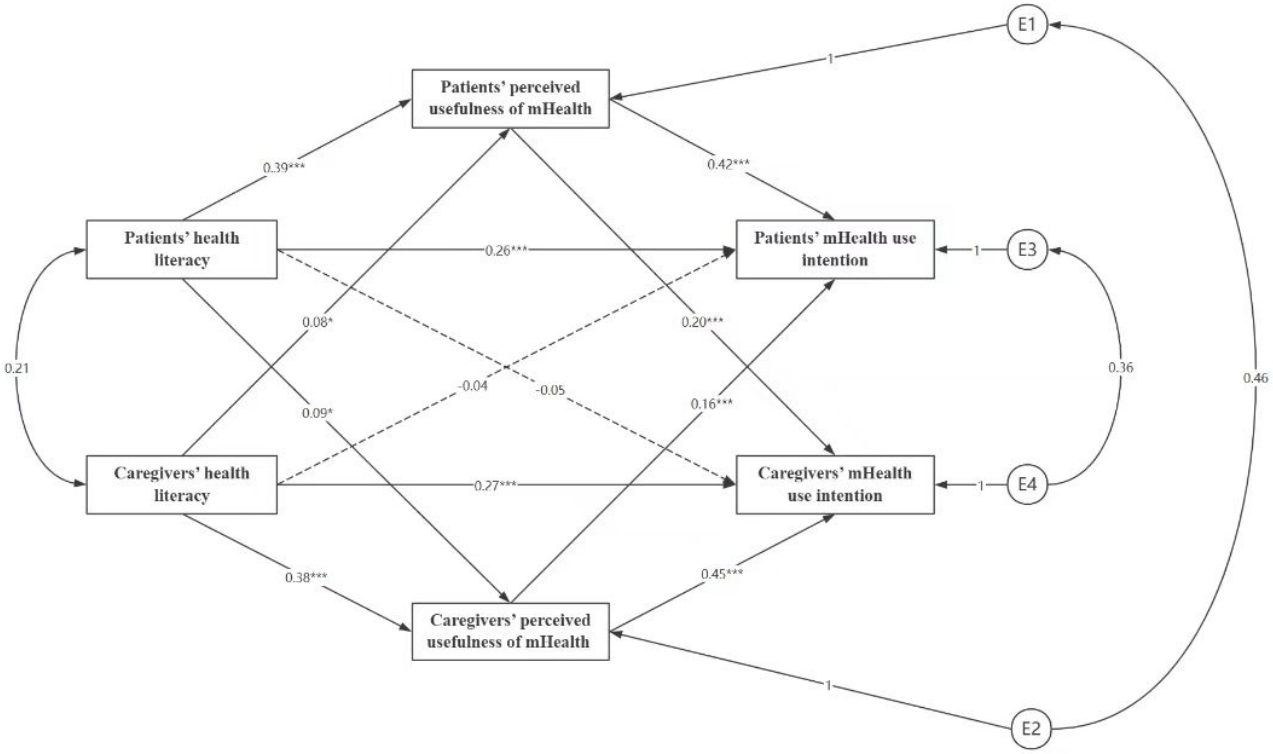
In this study, the model fit index values were X^2 (*df*)=1.175 (<3), RMSEA=0.024 (≤0.08), CFI=0.992 (≥0.90), TLI=0.995 (≥0.90), which indicated that the model was well fitted.

Direct Effect Analysis

The results showed that (1) patients' and caregivers' HL was a positive predictor of their own perceived usefulness of mHealth (β=.38 - 0.39; *P*<.001); (2) patients' and caregivers' perceived usefulness of mHealth was a positive predictor of their own mHealth use intention (β=.42 - 0.45; *P*<.001); (3) patients' and caregivers' HL had a positive predictive effect on their own mHealth use intention (β=.26 - 0.27; *P*<.001), that is, the subject effects of the model all held; (4) HL of patients and caregivers had a positive predictive effect on each other's perceived usefulness of mHealth (β=.08 - 0.09; *P*<.05); (5) perceived usefulness of mHealth of patients and caregivers had a positive predictive effect on each other's mHealth use intention (β=.16 - 0.20; *P*<.001); (6) patients' and caregivers' HL had a nonsignificant effect on each other's mHealth use intention, that is, the object effect was partially established (see Figure 1 for details).

XSL•FO
RenderX

Figure 1. APIMeM results of health literacy and perceived usefulness of mHealth on mHealth use intention in patient-caregiver dyads. Dashed lines indicate insignificant path coefficients and solid lines indicate significant path coefficients (*P<.05, **P<.01, ***P<.001). APIMeM: actor-partner interdependence mediation model; mHealth: mobile health.



Mediation Analysis

Bootstrap tests for mediating effects showed the patients' and caregivers' HL directly positively predicted their own mHealth use intention (β=.381, 95% CI 0.279 - 0.495; P<.001) and also positively predicted both their own (β=.242, 95% CI 0.171 - 0.321; P<.001) and their counterpart's mHealth perceived usefulness (β=.022, 95% CI 0.003 - 0.052; P=.02), which meant that the perceived usefulness of mHealth for patients with CHF and caregivers played a partial mediating role between their own and their counterpart's HL and mHealth use intention. In addition, patients' and caregivers' HL did not directly predict their counterpart's mHealth use intention (β=−.068, 95% CI −0.168 - 0.038; P=.21), but it could be positively predicting own perceived usefulness (β=.102, 95% CI 0.049 - 0.167; P<.001) and their counterpart's perceived usefulness (β=.053, 95% CI 0.006 - 0.107; P=.02), further positively predicting their own intention to use mHealth and their counterpart's intention to use mHealth, which meant that mHealth perceived usefulness of patients with CHF and caregivers played a fully mediating role in the relationship between their own HL and their counterpart's mHealth use intention (see Table 4 for details).

XSL•FO
RenderX

Table . The direct, indirect, and total indirect effects for patients with chronic heart failure and caregivers in the actor-partner interdependence mediation model (N=312 dyads).

	β [a]	SE	95% CI	*P* value
Actor effect				
Patients				
Total indirect effect	.264	0.046	0.182 to 0.356	<.001
1[b]HL[c]→1PU[d]→1UI[e]	.242	0.039	0.171 to 0.321	<.001
1HL→2[f] PU→1 UI	.022	0.012	0.003 to 0.052	.02
Direct effect	.381	0.054	0.279 to 0.495	<.001
Caregivers				
Total indirect effect	.264	0.046	0.182 to 0.356	<.001
2HL→1PU→2 UI	.022	0.012	0.003 to 0.052	.02
2HL→2PU→2 UI	.242	0.039	0.171 to 0.321	<.001
Direct effect	.381	0.054	0.279 to 0.495	<.001
Partner effect				
Patients				
Total indirect effect	.154	0.044	0.072 to 0.245	<.001
2HL→1PU→1 UI	.053	0.026	0.006 to 0.107	.02
2HL→2PU→1 UI	.102	0.030	0.049 to 0.167	<.001
Direct effect	−.068	0.052	−0.168 to 0.038	.21
Caregivers				
Total indirect effect	.154	0.044	0.072 to 0.245	<.001
1HL→1PU→2 UI	.102	0.030	0.049 to 0.167	<.001
1HL→2PU→2 UI	.053	0.026	0.006 to 0.107	.02
Direct effect	−.068	0.052	−0.168 to 0.038	.21

[a]β: standardized estimate.

[b]1: patients.

[c]HL: health literacy.

[d]PU: perceived usefulness.

[e]UI: use intention.

[f]2: caregivers.

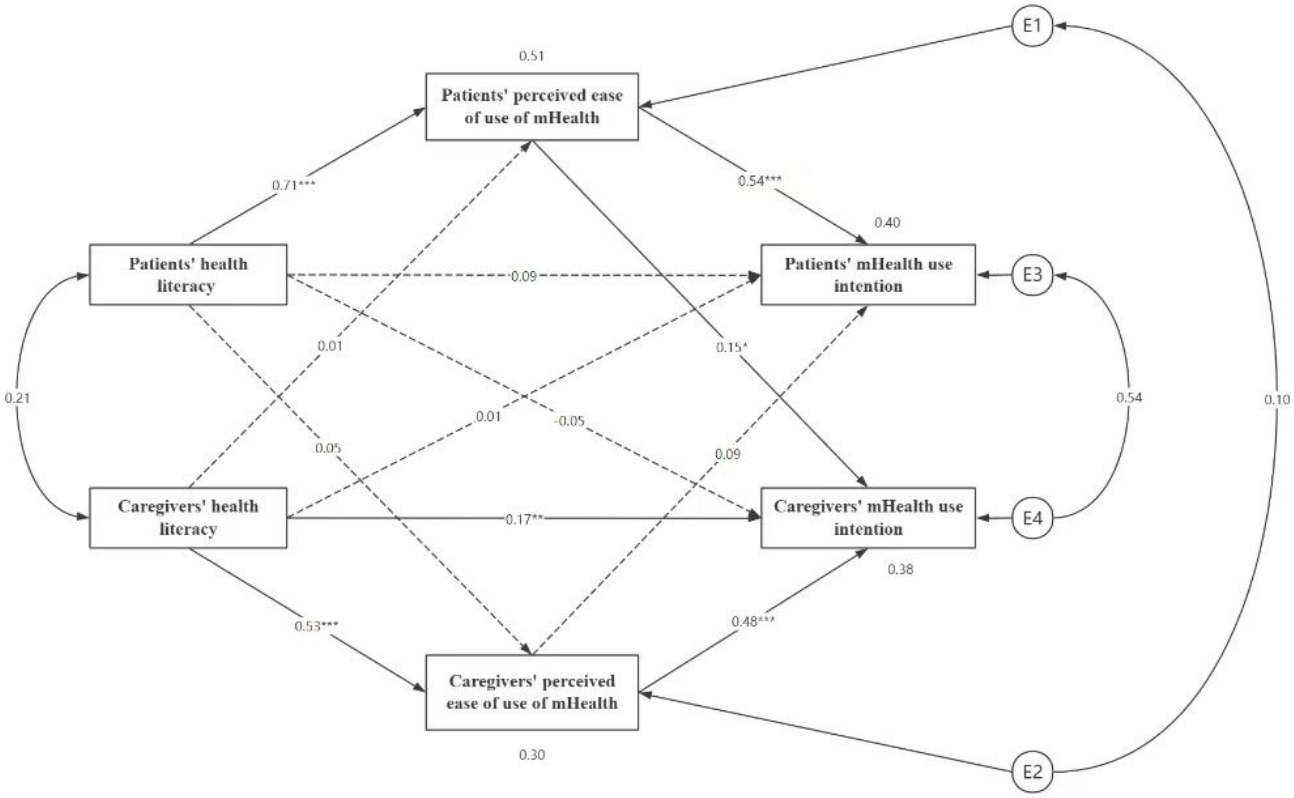
The APIMeM of HL-Perceived Ease of Use of mHealth-mHealth Use Intention in Patients With CHF and Caregivers

The model was a saturated model with 0 degrees of freedom [37], so its fit indices would no longer be estimated and only its path coefficients would be of interest [38].

The results showed that (1) patients' and caregivers' HL was a positive predictor of their own perceived ease of use (β=.53 - 0.71; *P*<.001); (2) patients' and caregivers' perceived ease of use was a positive predictor of their own mHealth use intention (β=.48 - 0.54; *P*<.001); (3) caregivers' HL had a positive predictive effect on their own mHealth use intention (β=.17; *P*<.01), that is, the main effect of the model was partially established; and (4) perceived ease of use in patients had a positive predictive effect on the caregiver's mHealth use intention (β=.15; *P*<.01); the rest of the object effects were not significant (see Figure 2 for details).

XSL•FO
RenderX

Figure 2. APIMeM results of health literacy and perceived ease of use on mHealth use intention in patient-caregiver dyads. Dashed lines indicate insignificant path coefficients and solid lines indicate significant path coefficients (*P<.05, **P<.01, ***P<.001). APIMeM: actor-partner interdependence mediation model; mHealth: mobile health.



Mediation Analysis

Bootstrap tests of mediating effects were conducted, and the results showed that patients' HL positively predicted their own perceived ease of use, which then positively predicted their own mHealth use intention (β=.584, 95% CI 0.443 - 0.764; P<.001), with perceived ease of use playing a fully mediating role. Caregivers' HL directly positively predicted their own mHealth use intention (β=.235, 95% CI 0.079 - 0.402; P=.003), as well as positively predicting their own perceived ease of use, which predicted their mHealth use intention (β=.355, 95% CI 0.234 - 0.504; P<.001), and perceived ease of use played a partial mediating role. In addition, patients' HL positively predicted their own perceived ease of use, which positively predicted their counterpart's mHealth use intention (β=.128, 95% CI 0.012 - 0.281; P=.02; see Table 5 for details).

XSL•FO
RenderX

Table . The direct, indirect, and total indirect effects for patients with chronic heart failure and caregivers in the actor-partner interdependence mediation model (N=312 dyads).

	β [a]	SE	95% CI	*P* value
Actor effect				
Patients				
Total indirect effect	.591	0.082	0.449 to 0.775	<.001
1[b]HL[c]→1PEOU[d]→1UI[e]	.584	0.081	0.443 to 0.764	<.001
1HL→2[f]PEOU→1UI	.007	0.010	−0.004 to 0.043	.19
Direct effect	.138	0.087	−0.036 to 0.307	.11
Caregivers				
Total indirect effect	.357	0.071	0.232 to 0.509	.23
2HL→1PEOU→2UI	.002	0.008	−0.013 to 0.020	.63
2HL→2PEOU→2UI	.355	0.069	0.234 to 0.504	<.001
Direct effect	.235	0.082	0.079 to 0.402	.003
Partner effect				
Patients				
Total indirect effect	.089	0.073	−0.043 to 0.243	.19
2HL→1PEOU→1UI	.008	0.035	−0.060 to 0.080	.79
2HL→2PEOU→1UI	.082	0.064	−0.036 to 0.215	.17
Direct effect	.024	0.093	−0.160 to 0.205	.78
Caregivers				
Total indirect effect	.159	0.074	0.033 to 0.323	.01
1HL→1PEOU→2UI	.128	0.066	0.012 to 0.281	.02
1HL→2PEOU→2UI	.031	0.030	−0.023 to 0.094	.26
Direct effect	−.063	0.087	−0.244 to 0.095	.48

[a]β: standardized estimate.

[b]1: patients.

[c]HL: health literacy.

[d]PEOU: perceived ease of use.

[e]UI: use intention.

[f]2: caregivers.

Moderated Effects of Age and Patient-Caregiver Relationship on Dyadic Effects

We regrouped the data into older and younger groups based on patient and caregiver age, and into spouse, child, and other groups based on patient-caregiver relationship, and used multigroup APIMeM to explore further whether the dichotomous effects of HL, perceived usefulness of mHealth, and perceived ease of use on mHealth use intention were moderated by patient and caregiver age and relationship. The results showed that the path coefficients of the APIMeM were not significantly different across groups (P>.05), meaning that no moderate influence of patient and caregiver age and relationship on the dichotomous effects of HL and mHealth perceived usefulness and perceived ease of use on mHealth use intention was observed. The details are shown in Multimedia Appendix 3.

Discussion

Principal Findings

The study took place during the heart of the COVID-19 pandemic, which forced much of the world to rely on eHealth and mHealth during this time. Some studies have reported an increase in internet dependence as a result of the COVID-19 pandemic, which has led to an increase in digital connectivity among older adults, thereby increasing the feasibility of their use of mHealth services [39]. Our study showed that the HL, perceived usefulness, perceived ease of use, and mHealth use intention of patients with CHF were significantly lower than caregivers' scores, which might be due to the following reasons. First, most of the caregivers in the study were the children of the patients and were therefore younger. It has been evidenced that bodily functions and cognitive capacities deteriorate with age, resulting in decreased vision flexibility and memory loss,

XSL•FO
RenderX

among other effects [40,41]. In addition, the younger generation, called "digital natives," had earlier and more frequent access to the internet, whereas the ability of older adults to adapt to modern digital life wanes over time, and this "digital divide" of older adults means that they face greater difficulties and obstacles in the process of using mHealth [42]. Furthermore, previous studies have pointed out that HL declines with age [43]. Therefore, compared with younger caregivers, patients with CHF have more difficulties grasping and using mHealth, and they may show lower perceived usefulness, perceived ease of use, mHealth use intention, and HL as well. Second, research has shown that [44] physical limitations associated with illness might impede an individual's ability to use technology, making it more arduous and challenging to use mHealth. The more severe the condition, the more likely the individual is to perceive more impediments to the use of mHealth [45]; therefore, compared with caregivers, people with CHF might exhibit lower levels of perceived usefulness, perceived ease of use, and mHealth use intention.

Our study showed that the HL of patients with CHF and their caregivers positively contributed to their own mHealth use intention. HL has emerged as a powerful predictor of self-care behaviors including mHealth use in the context of chronic illness, which can empower individuals to navigate the adoption of mHealth use [46]. HL entails people's motivation to understand and apply health information in order to make decisions in everyday life to maintain or improve quality of life [47]; therefore, people with higher HL might be more motivated to use mHealth to manage disease. Evidence also suggests that individuals with low HL face greater challenges in computer use and reading, manipulating, and assessing health information [48], whereas individuals with higher levels of HL are more likely to comprehend and use mHealth [49], leading to a high level of mHealth use intention. Our study also found that the HL of patients with CHF and their caregivers positively influenced their counterparts' mHealth use intention by the fully mediating role of perceived usefulness. Therefore, to improve patients' mHealth use, not only is the improvement of their HL necessary but their caregivers' enhanced HL is also highlighted.

Perceived usefulness was identified as a partial mediator of HL and mHealth use intention in patients with chronic heart disease and their caregivers in our study. This aligns with the broader literature emphasizing the importance of perceived usefulness in shaping mHealth adoption intention. As an important attribute of HL, reading skills entail a multitude of complex cognitive processes that require the ability to obtain meaning from the text being read. Additionally, HL lies in being able to use the information to make appropriate health care decisions [50]. All of the above suggests that higher HL might facilitate the understanding of mHealth, including the design, purpose, precautions, and adoption, leading to a higher level of perceived usefulness. It has also been evidenced by Wu et al [22,23] that HL positively affects the perceived usefulness of mHealth. Furthermore, usability issues, such as lack of an easy overview and nonintuitive app design, text, buttons, and icon elements, have been reported to be negatively correlated with mHealth utilization [51,52]. In addition, TAM theory also supports the positive relationship between perceived usefulness and mHealth use intention [53]. It was also suggested by the study of Dunn Lopez et al that mHealth apps must be readable, provide useful functions, and be based on evidence to improve patients' intention to use mHealth [54]. We also identified that the perceived usefulness of patients acted as the full mediator of patients' HL and caregivers' mHealth use intention, and the perceived usefulness of caregivers acted as the full mediator of caregivers' HL and patients' mHealth use intention. The findings suggest that the perceived usefulness of the patients and caregivers can both influence the mHealth use intention of patients, highlighting that patients and caregivers should be involved during the redevelopment process of mHealth to improve its usability from the perspectives of patients and caregivers. The app content must be readable, provide useful functions, and be based on evidence.

In our study, perceived ease of use was identified as a partial mediator of HL and mHealth use intention in patients with chronic heart disease, and it was a full mediator of HL and mHealth use intention in caregivers. HL includes individuals' capabilities to access and use available resources within the health care system, which allows individuals to increase their abilities to use preventive health services including mHealth to manage their health [55]. Thus, people with higher HL will show increased perceived ease of use of mHealth. In addition, perceived ease of use influencing individuals' mHealth use intention is also attested by other researchers and is supported by TAM theory. Patients' and caregivers' concerns about mHealth apps being difficult and time-consuming to use, along with being not easy to access through the use of complex keywords, led to their unwillingness to use mHealth [56]. The actor-partner analysis showed that perceived ease of use of the patients was a full mediator of patients' HL and their caregivers' mHealth use intention; however, perceived ease of use of the caregivers was not a mediator of caregivers' or patients' HL and patients' mHealth use intention. Therefore, caregivers' perceived ease of use of mHealth will not induce patients' willingness to adopt mHealth, and patients will choose to use it based on their own experience, highlighting the key role of perceived ease of use of the patients in facilitating their mHealth use. Nevertheless, some patients with CHF have expressed difficulty in downloading electronic health–related equipment and procedures and in operating health-monitoring equipment due to their complexity, which has increased their anxiety about its use [57]. Other research has also pointed out that the target population of mHealth expressed the desire for a simple and intuitive-to-use mobile app with nonmedical language and easy-to-understand material [58]. Therefore, it is recommended that the development of mHealth should focus on simplifying operational procedures and providing relevant operational training when necessary based on the needs of patients.

Limitations

This study also has limitations. First, it was not possible to infer a causal relationship between the variables as this was a cross-sectional study. Second, this study only focused on the mediating effects of 2 variables, perceived usefulness and perceived ease of use of mHealth, and the mediation of the relationship between HL and mHealth use intention, ignoring other possible variables. To better explain how HL affects

XSL•FO
RenderX

mHealth use intentions in the patient-caregiver dyad, future research should focus on other possible mediating or moderating variables. In addition, the increase in internet and digital technology use spurred on by the pandemic might have impacted the findings; therefore, our study findings should be considered with caution. Third, participants were from only 2 tertiary hospitals in Kunming, Yunnan Province, China, and most of the patient-caregiver relationships were between parents and children, who had a certain age gap. Therefore, the population characteristics and cultural differences between countries and regions may limit the generalizability of our findings.

Conclusions

Our study proposes that patients with CHF will exhibit lower HL, perceived usefulness, perceived ease of use, and intention to use mHealth compared to their caregivers. Furthermore, the HL of patients with CHF and their caregivers positively contributes to their own mHealth use intention, with the full mediating effect of perceived usefulness. Therefore, mHealth utilization among patients with CHF could be promoted by improving their own HL and that of their caregivers as well. The findings suggest that both patients' and caregivers' perceived usefulness affects patients' mHealth use intention, and therefore patients and caregivers should be involved in the entire mHealth development process in order to improve the usability of mHealth for patients and caregivers. The study highlights the key role of patients' perception of mHealth as easy to use in facilitating their use of mHealth. Therefore, it is recommended that the development of mHealth should focus on simplifying operational procedures and providing relevant training about operations according to the needs of the patients when necessary.

Acknowledgments

The authors would like to express their heartfelt gratitude to all persons who participated in this study as well as the members of the research team. This study was supported by the Yunnan Provincial Department of Education Science Research Fund Project, China (number 2023Y0749).

Data Availability

The full interview transcripts and audio files are not available to the public due to data protection reasons. However, transcripts and findings are available on request from the corresponding author.

Authors' Contributions

XJ (first author) participated in study design, data collection and analysis, and manuscript writing. YZ and MZ participated in data collection and data analysis. QM, YB, QH, and WW participated in reviewing the research design manuscript. XZ participated in data analysis. FM (corresponding author) participated in study design, data analysis, manuscript editing, and review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile Health Service Use Intention Questionnaire.
[PDF File, 250 KB - mhealth_v13i1e63805_app1.pdf]

Multimedia Appendix 2
The results of the normality test.
[PDF File, 77 KB - mhealth_v13i1e63805_app2.pdf]

Multimedia Appendix 3
The results of the multigroup actor-partner interdependence mediation model.
[PDF File, 981 KB - mhealth_v13i1e63805_app3.pdf]

References

1. Bozkurt B, Coats AJS, Tsutsui H, et al. Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. European J of Heart Fail 2021 Mar;23(3):352-380 [FREE Full text] [doi: 10.1002/ejhf.2115]
2. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: executive summary: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2022 May 3;145(18):e876-e894. [doi: 10.1161/CIR.0000000000001062] [Medline: 35363500]

XSL•FO
RenderX

3. The Writing Committee of the Report on Cardiovascular Health and Diseases in China. Report on cardiovascular health and diseases in China 2022: an updated summary. Biomed Environ Sci 2023 Aug 20;36(8):669-701. [doi: 10.3967/bes2023.106] [Medline: 37711081]
4. Emmons-Bell S, Johnson C, Roth G. Prevalence, incidence and survival of heart failure: a systematic review. Heart 2022 Aug 11;108(17):1351-1360. [doi: 10.1136/heartjnl-2021-320131] [Medline: 35042750]
5. Kitsiou S, Vatani H, Paré G, et al. Effectiveness of mobile health technology interventions for patients with heart failure: systematic review and meta-analysis. Can J Cardiol 2021 Aug;37(8):1248-1259. [doi: 10.1016/j.cjca.2021.02.015] [Medline: 33667616]
6. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail 2013 May;6(3):606-619. [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]
7. Writing Group Members, Mozaffarian D, Benjamin EJ, et al. Heart disease and stroke statistics-2016 update: a report from the American Heart Association. Circulation 2016 Jan 26;133(4):e38-360. [doi: 10.1161/CIR.0000000000000350] [Medline: 26673558]
8. Riegel B, Driscoll A, Suwanno J, et al. Heart failure self-care in developed and developing countries. J Card Fail 2009 Aug;15(6):508-516. [doi: 10.1016/j.cardfail.2009.01.009] [Medline: 19643362]
9. Leigh JW, Gerber BS, Gans CP, Kansal MM, Kitsiou S. Smartphone ownership and interest in mobile health technologies for self-care among patients with chronic heart failure: cross-sectional survey study. JMIR Cardio 2022 Jan 14;6(1):e31982. [doi: 10.2196/31982] [Medline: 35029533]
10. Song T, Deng N, Cui T, et al. Measuring success of patients' continuous use of mobile health services for self-management of chronic conditions: model development and validation. J Med Internet Res 2021 Jul 13;23(7):e26670. [doi: 10.2196/26670] [Medline: 34255685]
11. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR mHealth uHealth 2015 Nov 4;3(4):e101. [doi: 10.2196/mhealth.4924] [Medline: 26537656]
12. Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Institute of Medicine (US) committee on health literacy. In: Health Literacy: A Prescription to End Confusion: National Academies Press; 2004. [doi: 10.17226/10883] [Medline: 25009856]
13. Manafò E, Wong S. Promoting eHealth literacy in older adults: key informant perspectives. Can J Diet Pract Res 2013;74(1):37-41. [doi: 10.3148/74.1.2013.37] [Medline: 23449214]
14. Jung SO, Son YH, Choi E. E-health literacy in older adults: an evolutionary concept analysis. BMC Med Inform Decis Mak 2022 Jan 31;22(1):28. [doi: 10.1186/s12911-022-01761-5] [Medline: 35101005]
15. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 1;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
16. Andersen MH, Hermansen Å, Dahl KG, et al. Profiles of health literacy and digital health literacy in clusters of hospitalised patients: a single-centre, cross-sectional study. BMJ Open 2024 May 20;14(5):e077440. [doi: 10.1136/bmjopen-2023-077440] [Medline: 38772592]
17. Kasaye MD, Kebede N, Kalayou MH, Kebede SD, Molla A. Digital health literacy and associated factors among health professionals during the outbreak of corona virus pandemic in Ethiopia: a systematic review and meta-analysis. Digit Health 2024;10:20552076241271799. [doi: 10.1177/20552076241271799] [Medline: 39148812]
18. Zhao BY, Huang L, Cheng X, et al. Digital health literacy and associated factors among internet users from China: a cross-sectional study. BMC Public Health 2024 Mar 27;24(1):908. [doi: 10.1186/s12889-024-18324-0] [Medline: 38539176]
19. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014 May 9;16(5):e125. [doi: 10.2196/jmir.3283] [Medline: 24824062]
20. Schillinger D, Grumbach K, Piette J, et al. Association of health literacy with diabetes outcomes. JAMA 2002;288(4):475-482. [doi: 10.1001/jama.288.4.475] [Medline: 12132978]
21. Zhou M, Zhao L, Kong N, Campy KS, Qu S, Wang S. Factors influencing behavior intentions to telehealth by Chinese elderly: an extended TAM model. Int J Med Inform 2019 Jun;126:118-127. [doi: 10.1016/j.ijmedinf.2019.04.001] [Medline: 31029253]
22. Wu Y, Zhang M, Zhu L. The research on the evaluation of consumer's electronic health literacy and application in mobile health. Consum Econ 2017;33(1):90-96 [FREE Full text]
23. Zhang M, Liu X, Zhang Y. Users' medical information help-seeking behavior in online health community–external factors, individual motivation and formation path. Modern Inform 2018;11:18-24. [doi: 10.3969/j.issn.1008-0821.2018.11.004]
24. McHorney CA, Mansukhani SG, Anatchkova M, et al. The impact of heart failure on patients and caregivers: a qualitative study. PLoS One 2021;16(3):e0248240. [doi: 10.1371/journal.pone.0248240] [Medline: 33705486]
25. Wingham J, Frost J, Britten N, et al. Needs of caregivers in heart failure management: a qualitative study. Chronic Illn 2015 Dec;11(4):304-319. [doi: 10.1177/1742395315574765] [Medline: 25795144]
26. Cheng M, Zhu C, Meng J, Pu C, Chen G, Liu H. Association between stigma and quality of life among chronic refractory wounds patients and informal caregivers: an actor-partner interdependence model analysis. Qual Life Res 2024 Nov;33(11):3131-3140. [doi: 10.1007/s11136-024-03759-0] [Medline: 39162970]

XSL•FO
RenderX

27. Manne S, Etz RS, Hudson SV, et al. A qualitative analysis of couples' communication regarding colorectal cancer screening using the Interdependence Model. Patient Educ Couns 2012 Apr;87(1):18-22. [doi: 10.1016/j.pec.2011.07.012] [Medline: 21835574]
28. Ellis KR, Janevic MR, Kershaw T, Caldwell CH, Janz NK, Northouse L. Engagement in health-promoting behaviors and patient-caregiver interdependence in dyads facing advanced cancer: an exploratory study. J Behav Med 2017 Jun;40(3):506-519. [doi: 10.1007/s10865-016-9819-6] [Medline: 28078502]
29. Dwivedi YK, Shareef MA, Simintiras AC, Lal B, Weerakkody V. A generalised adoption model for services: a cross-country comparison of mobile health (m-health). Gov Inf Q 2016 Jan;33(1):174-187. [doi: 10.1016/j.giq.2015.06.003]
30. Sun Y, Wang N, Guo X, Peng J. Understanding the acceptance of mobile health services: a comparison and integration of alternative models. Journal of Electronic Commerce Research 2013 Jan 1;14:183-200 [FREE Full text]
31. Pasek M, Goździalska A, Jochymek M, Caruso R. Social support in a cancer patient-informal caregiver dyad: a scoping review. Cancers (Basel) 2023 Mar 14;15(6):1754. [doi: 10.3390/cancers15061754] [Medline: 36980639]
32. Ledermann T, Rudaz M, Wu Q, Cui M. Determine power and sample size for the simple and mediation actor–partner interdependence model. Fam Relat 2022 Oct;71(4):1452-1469 [FREE Full text] [doi: 10.1111/fare.12644]
33. Matsuoka S, Kato N, Kayane T, et al. Development and validation of a heart failure-specific health literacy scale. J Cardiovasc Nurs 2016;31(2):131-139. [doi: 10.1097/JCN.0000000000000226] [Medline: 26049813]
34. Yue M, Zhang L, Lu Y, Jin C. Translation and psychometric evaluation of the Chinese version of the Heart Failure-Specific Health Literacy Scale. International Journal of Nursing Sciences 2016 Dec;3(4):342-346. [doi: 10.1016/j.ijnss.2016.11.004]
35. Lin S. Research on Elderly Patients' Willingness to Use mHealth Services: Henan University; 2022. [doi: 10.27114/d.cnki.ghnau.2022.000477]
36. Kolmogorov–Smirnov test. In: The Concise Encyclopedia of Statistics: Springer; 2008:283-287. [doi: 10.1007/978-0-387-32833-1_214]
37. Steeger CM, Gondoli DM. Mother-adolescent conflict as a mediator between adolescent problem behaviors and maternal psychological control. Dev Psychol 2013 Apr;49(4):804-814. [doi: 10.1037/a0028599] [Medline: 22612432]
38. Li Z, Xiang-Juan XUE, Jing-Xin D. Discrimination perception, academic achievement in rural left behind children: longitudinal mediation role of depression. J Psychol Sci 2019 May 20(3):584. [doi: 10.16719/j.cnki.1671-6981.20190311]
39. Jaana M, Paré G. Comparison of mobile health technology use for self-tracking between older adults and the general adult population in Canada: cross-sectional survey. JMIR mHealth uHealth 2020 Nov 27;8(11):e24718. [doi: 10.2196/24718] [Medline: 33104517]
40. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
41. Chen K, Chan AHS. A review of technology acceptance by older adults. Gerontechnology 2011 Jan 1;10(1):1-12 [FREE Full text] [doi: 10.4017/gt.2011.10.01.006.00]
42. Liu L, Wu F, Tong H, Hao C, Xie T. The digital divide and active aging in China. Int J Environ Res Public Health 2021 Dec 1;18(23):12675. [doi: 10.3390/ijerph182312675] [Medline: 34886400]
43. Levin JB, Peterson PN, Dolansky MA, Boxer RS. Health literacy and heart failure management in patient-caregiver dyads. J Card Fail 2014 Oct;20(10):755-761. [doi: 10.1016/j.cardfail.2014.07.009] [Medline: 25072623]
44. Foster M, Xiong W, Quintiliani L, Hartmann CW, Gaehde S. Preferences of older adult veterans with heart failure for engaging with mobile health technology to support self-care: qualitative interview study among patients with heart failure and content analysis. JMIR Form Res 2022 Dec 20;6(12):e41317. [doi: 10.2196/41317] [Medline: 36538348]
45. Schomakers EM, Lidynia C, Vervier LS, Calero Valdez A, Ziefle M. Applying an extended UTAUT2 model to explain user acceptance of lifestyle and therapy mobile health apps: survey study. JMIR mHealth uHealth 2022 Jan 18;10(1):e27095. [doi: 10.2196/27095] [Medline: 35040801]
46. Magi CE, Bambi S, Rasero L, et al. Health literacy and self-care in patients with chronic illness: a systematic review and meta-analysis protocol. Healthcare (Basel) 2024 Mar 31;12(7):762. [doi: 10.3390/healthcare12070762] [Medline: 38610184]
47. Jaensson M, Josefin W, Dahlberg K. Health literacy friendly organizations - a scoping review about promoting health literacy in a surgical setting. Patient Educ Couns 2024 Aug;125:108291. [doi: 10.1016/j.pec.2024.108291] [Medline: 38626578]
48. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790. [doi: 10.2196/34790] [Medline: 35171099]
49. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019;5:2055207619882181. [doi: 10.1177/2055207619882181] [Medline: 31656632]
50. Sierra M, Cianelli R. Health literacy in relation to health outcomes: a concept analysis. Nurs Sci Q 2019 Oct;32(4):299-305. [doi: 10.1177/0894318419864328] [Medline: 31514613]
51. Wong AKC, Bayuo J, Wong FKY, et al. Experiences of receiving an mHealth application with proactive nursing support among community-dwelling older adults: a mixed-methods study. BMC Nurs 2024 Apr 7;23(1):232. [doi: 10.1186/s12912-024-01909-w] [Medline: 38584270]

XSL•FO
RenderX

52. Giebel GD, Abels C, Plescher F, et al. Problems and barriers related to the use of mHealth apps from the perspective of patients: focus group and interview study. J Med Internet Res 2024 Apr 23;26:e49982. [doi: 10.2196/49982] [Medline: 38652508]
53. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
54. Dunn Lopez K, Chae S, Michele G, et al. Improved readability and functions needed for mHealth apps targeting patients with heart failure: an app store review. Res Nurs Health 2021 Feb;44(1):71-80. [doi: 10.1002/nur.22078] [Medline: 33107056]
55. Parnell TA, Stichler JF, Barton AJ, Loan LA, Boyle DK, Allen PE. A concept analysis of health literacy. Nurs Forum 2019 Jul;54(3):315-327. [doi: 10.1111/nuf.12331] [Medline: 30793314]
56. Bendixen RM, Fairman AD, Karavolis M, Sullivan C, Parmanto B. A user-centered approach: understanding client and caregiver needs and preferences in the development of mHealth apps for self-management. JMIR mHealth uHealth 2017 Sep 26;5(9):e141. [doi: 10.2196/mhealth.7136] [Medline: 28951378]
57. Zhang N, Li Q, Chen S, et al. Effectiveness of nurse-led electronic health interventions on illness management in patients with chronic heart failure: a systematic review and meta-analysis. Int J Nurs Stud 2024 Feb;150:104630. [doi: 10.1016/j.ijnurstu.2023.104630] [Medline: 38029453]
58. Hafdi M, Eggink E, Hoevenaar-Blom MP, et al. Design and development of a mobile health (mHealth) platform for dementia prevention in the prevention of dementia by mobile phone applications (PRODEMOS) project. Front Neurol 2021;12:733878. [doi: 10.3389/fneur.2021.733878] [Medline: 34975710]

Abbreviations

APIM: actor-partner interdependence model
APIMeM: actor-partner interdependence mediation model
CFI: comparative fit index
CHF: chronic heart failure
HF-SHLS: Heart Failure–Specific Health Literacy Scale
HL: health literacy
mHealth: mobile health
RMSEA: root mean square error of approximation
TAM: Technology Acceptance Model
TLI: Tucker-Lewis index

Edited by M Dorsch; submitted 30.06.24; peer-reviewed by A Bugajski, H Wang; revised version received 25.12.24; accepted 24.01.25; published 06.03.25.

Please cite as:
Jin X, Zhang Y, Zhou M, Mei Q, Bai Y, Hu Q, Wei W, Zhang X, Ma F
An Actor-Partner Interdependence Mediation Model for Assessing the Association Between Health Literacy and mHealth Use Intention in Dyads of Patients With Chronic Heart Failure and Their Caregivers: Cross-Sectional Study
JMIR Mhealth Uhealth 2025;13:e63805
URL: https://mhealth.jmir.org/2025/1/e63805
doi:10.2196/63805

© Xiaorong Jin, Yimei Zhang, Min Zhou, Qian Mei, Yangjuan Bai, Qiulan Hu, Wei Wei, Xiong Zhang, Fang Ma. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Participant Adherence in Repeated-Dose Clinical Studies Using Video-Based Observation: Retrospective Data Analysis

Jihong Song[1], BSN; Sungpil Han[1,2], MD, PhD; Suein Choi[1,2], MD, PhD; Jonghyuk Lim[3], BE; Byeong Yeob Oh[3], BBA; Dongoh Shin[3], MEng; Seunghoon Han[1,2], MD, PhD

[1]Department of Pharmacology, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[2]Department of Clinical Pharmacology and Therapeutics, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[3]CareSquare, Inc., Seoul, Republic of Korea

Corresponding Author:
Seunghoon Han, MD, PhD
Department of Clinical Pharmacology and Therapeutics
Seoul St. Mary's Hospital, College of Medicine
The Catholic University of Korea
222 Banpo-daero, Seocho-gu
Seoul, 06591
Republic of Korea
Phone: 82 10 9484 2415
Fax: 82 2 3147 8357
Email: waystolove@catholic.ac.kr

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2025/1/e79507

Abstract

Background: Maintaining accurate medication records in clinical trials is essential to ensure data validity. Traditional methods such as direct observation, self-reporting, and pill counts have shown limitations that make them inaccurate or impractical. Video-based monitoring systems, available as commercial or proprietary mobile applications for smartphones and tablets, offer a promising solution to these traditional limitations. In Korea, a system applicable to the clinical trial context has been developed and used.

Objective: This study aimed to evaluate the usefulness of an asynchronous video-based self-administration of the investigational medicinal product (SAI) monitoring system (VSMS) in ensuring accurate dosing and validating participant adherence to planned dosing times in repeated-dose clinical trials.

Methods: A retrospective analysis was conducted using data from 17,619 SAI events in repeated-dose clinical trials using the VSMS between February 2020 and March 2023. The SAI events were classified into four categories: (1) Verified on-time dosing, (2) Verified deviated dosing, (3) Unverified dosing, and (4) Missed dosing. Analysis methods included calculating the success rate for verified SAI events and analyzing trends in difference between planned and actual dosing times (PADEV) over the dosing period and by push notification type. The mean PADEV for each subsequent dosing period was compared with the initial period using either a paired *t* test or a Wilcoxon signed-rank test to assess any differences.

Results: A comprehensive analysis of 17,619 scheduled SAI events across 14 cohorts demonstrated a high success rate of 97% (17,151/17,619), with only 3% (468/17,619) unsuccessful due to issues like unclear video recordings or technical difficulties. Of the successful events, 99% (16,975/17,151) were verified as on-time dosing, confirming that the dosing occurred within the designated SAI time window with appropriate recorded behavior. In addition, over 90% (367/407) of participants consistently reported dosing videos on all analyzed SAI days, with most days showing over 90% objective dosing data, underscoring the system's effectiveness in supporting accurate SAI. There were cohort differences in the tendency to dose earlier or later, but no associated cohort characteristics were identified. The initial SAI behaviors were generally sustained during the whole period of participation, with only 16% (13/79) of study days showing significant shifts in actual dosing times. Earlier deviations in SAI times were observed when only dosing notifications were used, compared with using reminders together or no notifications.

XSL•FO
RenderX

Conclusions: VSMS has proven to be an effective tool for obtaining dosing information with accuracy comparable to direct observation, even in remote settings. The use of various alarm features and appropriate intervention by the investigator or observer was identified as a way to minimize adherence deterioration. It is expected that the usage and usefulness of VSMS will be continuously improved through the accumulation of experience in various medical fields.

(JMIR Mhealth Uhealth 2025;13:e65668) doi:10.2196/65668

KEYWORDS

adherence; mobile health; self-administration; repeated-dose clinical trials; video-based monitoring; mobile phone

Introduction

Repeated dose studies in healthy volunteers are essential for obtaining basic information about the pharmacokinetics of a drug. These studies are performed during the development of a specific investigational medicinal product (IMP) for which the dose of the active ingredient, formulation and route of administration has been determined. They include repeated dose escalation studies, drug-drug interaction studies, and some bioequivalence studies [1-3]. The changes in drug concentration that occur when an active ingredient is administered using a specific IMP are usually described in terms of the time elapsed after administration, from which various pharmacokinetics relationships are defined. Therefore, whether the IMP was actually administered and the time of administration are considered pivotal information to be obtained in such studies [4].

Historically, a variety of techniques have been used to ensure precise dosing information in clinical trials. These include direct observation by investigators, self-reporting through diaries or questionnaires, and pill counts [4,5]. Although direct observation is the most reliable method, it may be impractical and burdensome due to the necessity of frequent hospital visits, which increases costs and inconveniences participants [4,6-10]. Self-reporting methods are susceptible to biases such as recall and social desirability biases, which can lead to inaccuracies [5,6,11-15]. While pill counts are a straightforward approach, they cannot confirm actual administration, making them less reliable for studies requiring precise dosing records [12]. The inherent limitations of existing methodologies have consistently underscored the necessity for novel approaches capable of addressing the demands of repeated-dose studies.

The advent of mobile technologies has brought about innovative methodologies for the monitoring of the IMP administration in clinical trials [4,16]. One such advancement is the use of video-based observation, which permits the remote assessment of self-administration of the investigational medicinal product (SAI) [4]. Video-based observation uses either synchronous or asynchronous methodologies. In the former, the investigator observes the SAI in real-time [6,16-21], whereas in the latter, the participant uploads a video of the SAI for subsequent review [4,6,16,17,22-25]. This technology offers a degree of verification that is comparable to that of direct observation while reducing the necessity for frequent hospital visits, thus enhancing the convenience of participants and reducing the costs associated with the trial [4].

This study was to assess the usefulness of an asynchronous video-based SAI monitoring system (VSMS) through a simple and straightforward outcome measure: “Was the investigator able to reliably verify SAI using the videos uploaded by the participant?” This outcome measure is ultimately a direct indicator of whether the system performed as targeted, and was considered to be a composite of technical influences such as system reliability and sociodemographic influences such as digital literacy. In addition, if the results showed that the VSMS was sufficient to verify SAIs, it was planned to explore participants’ adherence to planned SAI times when using these systems. These evaluations will enhance our understanding of the validity and reliability of asynchronous VSMS in a range of repeated-dose trial settings.

Methods

The System Overview

The VSMS used in this study comprises a web-based system for investigators and a mobile app for participants. The web-based system allows investigators to (1) search for a QR code to match a participant to a specific participant number in the study; (2) view the SAI schedule, which should be assessed by calendar date; (3) view SAI-related videos uploaded by the participant; (4) determine if a valid administration is made and enter corresponding data; and (5) send the participant feedback on the SAI.

The system is designed to allow the investigator to make the final confirmation about the SAI results, with outcomes recorded in one of four categories (1)-(4):

(1) **Verified on-time dosing**

- The actual dosing time Provided by the system is within the allowed SAI time window and the dosing behavior recorded in the video is appropriate.
- The actual dosing time Provided by the system is outside the limits of the allowed SAI time window, but based on the video, it is verified that the appropriate dosing behavior occurred within the window (eg, video recording started before the time window range, but the actual dosing occurs within the window). In this case, the investigator manually corrected the actual dosing time.

(2) **Verified deviated dosing**

The actual dosing time provided by the system is outside the limits of the allowed SAI time window, and the appropriate SAI behavior is verified based on the video, but the timing is outside the window.

XSL•FO
RenderX

(3) **Unverified dosing**

SAI cannot be verified via video due to technical issues, participant nonadherence to recording procedures (eg, too dark recording environment, skipping oral cavity disclosure), and so on, but there is evidence to believe the participant performed the SAI (eg, the participant's statement from a phone call with the reason of inappropriate video recording)

(4) **Missed dosing**

SAI cannot be verified via video and no evidence that the participants performed SAI can be obtained.

The mobile app is designed with a simple and user-friendly interface. The system may provide push notifications to participants before or after the scheduled dosing time, as specified in the protocol. Upon opening the application, the participant is immediately instructed to start the video recording and proceed with the SAI. The server connection time, recording start time, and recording end time were always captured based on the participants' mobile device time to ensure that the data remained accurate even in situations such as slow internet speeds or transmission failures. In these cases, the recorded video was sent to the server as soon as mobile network conditions improved. Of the stored time data, the recording start time was displayed on the investigator's web system as the default value for the actual dosing time. The recorded video was temporarily cached on the participants' mobile device and deleted once the transfer to the server was confirmed. The behavior of the system according to the SAI schedule and procedures, storage and management of SAI data, and processing of other communication-related information was handled through servers operated by the service provider, which complied with all applicable quality and regulatory standards (encrypting network communications, ensuring physical security, and restricting video access only to authorized users).

Clinical Trials and Cohorts

The data presented in this study were derived from repeated-dose clinical trials in healthy volunteers who used DoseEase. These trials were conducted between February 2020 and March 2023 and the IMPs were all self-administrable formulations (oral or ophthalmic). Trials that included groups with 2 or more different SAI schedules were also included in the analysis, and each group was treated as a separate Cohort. The eligibility criteria for participants across all trials were consistent and included the following: age between 19 years and 55 years, a minimum weight of 50 kg for men and 45 kg for women, with a BMI between 18.5 and 24.9, and confirmation of healthy condition in a comprehensive medical examination, which included an assessment of medical history, physical examination, and laboratory tests. Participants were excluded if they exhibited significant active disease at the screening stage or a history of previous disease that could affect the administration and absorption of the IMP.

Before enrollment began for the clinical trial, the service provider received a protocol from the research staff members to set up the system that specified the following;

- The number of participants and the principle of participant numbering.
- Cohort categorization according to the dosing schedule.
- SAI schedules by cohort and overall dosing schedule.
- Allowable time deviations per SAI event.
- Plan to deliver push notifications per SAI event.

The finalized system was piloted and validated by the investigators before the start of the trial.

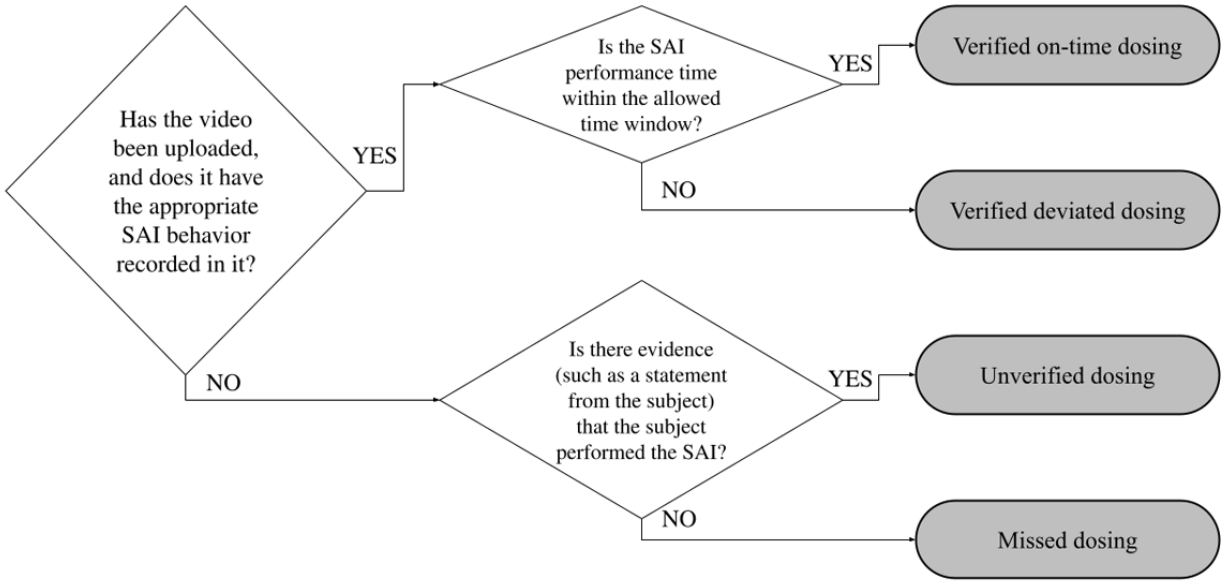
At the first participant visit, the investigator accessed the VSMS web, retrieved and printed a QR code for each participant number, which was then scanned by the participant with that number using his or her mobile device (this was a key procedure to ensure that each participant was correctly identified without obtaining additional sensitive personal information). This registered the participants device in the system at the same time as the VSMS application was installed on the participants mobile device. This procedure was followed by training, which included 1-2 mock SAIs to familiarize the participant with the system. The mock SAI underscored the significance of securely positioning the mobile device and executing all procedures associated with the SAI (eg, the IMP preparation with a specified volume of water, the oral administration, and the disclosure of the oral cavity after dosing for oral administration) within the boundaries of the recording screen. Participants were also instructed to protect their privacy by wearing appropriate clothing and recording the video in a place where their home environment was not visible. At the visit immediately before the scheduled SAI, the participant received the IMP for the number of scheduled SAIs before the next visit, after which the IMP was managed by the participant at a location other than the hospital.

On days when SAI was scheduled, participants were either alerted via push notification on the mobile application or not, as specified in the protocol. The timing (before the planned dosing time [dosing notification], after the planned dosing time [dosing reminder], or both) and frequency of these alerts varied by cohort. As trained in advance, at their own available time (preferably within the allowable SAI time window specified in the protocol), participants performed the SAI while recording a video. If mobile network conditions were adequate, the video was sent to the server as soon as it was recorded, at which point it was made available to the investigator. If the video was not uploaded after the planned dosing time, investigators were able to contact participants by phone to remind them to perform SAI or to check for any technical issues.

For each SAI event, investigators evaluated the uploaded videos to assess whether the SAI was appropriate for each participant. A decision tree was provided to help the investigator confirm SAI videos and related records into the appropriate category (Figure 1). For cases other than verified on-time dosing, investigators were able to send proper feedback to the participant to improve adherence.

XSL•FO
RenderX

Figure 1. Investigator's outcome decision-tree based on self-administered videos. SAI: self-administration of the investigational medicinal product.



Dataset for Statistical Analysis

Each SAI record included the planned Study Day (with the first dose day equal to 1), the Planned Dosing Time, the time the video recording started, and the investigator-confirmed actual dosing time. The difference between planned dosing time and actual dosing time (PADEV) was defined as the dosing time deviation. Study day values of 15 or more were recorded as "≥15" since the typical dosing duration was under 2 weeks. The investigator's final decision on SAI (categorized as verified on-time dosing, verified deviated dosing, unverified dosing, or missed dosing) and the type of push notification delivery (categorized as none, dosing notification, dosing reminder, or both) were also included. All available data from participants who dropped out before completing all study procedures were included in the analysis. Dataset handling and statistical analyses were performed using R software (version 4.2.2; R Foundation for Statistical Computing). The significance level was set at .05, and all tests were 2-tailed.

Fundamental Evaluation of Usefulness-Success Rate of the SAI Events

The primary goal of a VSMS is to ensure that the participant's video-recorded SAI behavior is successfully transmitted to the investigator, who can then determine the SAI outcome. This goal cannot be achieved if the participant or investigator is unskilled in using the system, or if the data is incomplete due to technical problems with the system. Therefore, a VSMS may be considered "useful" when it adequately fulfills its intention of use. In this study, cases where the investigator's final decision for the SAI is either verified on-time dosing or verified deviated dosing are collectively referred to as validated dosing. The percentage of validated dosing among all SAI records (referred to as the success rate, %) is evaluated as a measure of usefulness. The success rate was determined for each study day in which SAIs were performed in each cohort (considered separate cohorts if the same study included arms with 2 or more different dosing schedules). The success rate for each study day in the overall cohort and across all study days in each cohort were also calculated.

Trends in Dosing Time Deviations by Cohort

VSMS allows investigators to view SAI records in real-time, unlike other existing methods. This means that investigator-participant interaction is possible in real time throughout the study period, with features such as push notifications (dosing notification or dosing reminder) and appropriate participant management to provide input on adherence. To assess the usefulness of this aspect, the trend of PADEV per study day was analyzed for each cohort. Only validated dosing data records from cohorts with the success rate of 90% or higher were used in this analysis (even verified on-time dosing records may have deviations between planned dosing time and actual dosing time). In the first analysis, a paired *t* test or a Wilcoxon signed-rank test (depending on the sample size) was performed on the difference between the mean of PADEV in the reference study days (the first 3 study days in which SAIs were performed) and the mean of PADEV in each subsequent study day to assess whether the dosing trend identified early in the dosing period was maintained thereafter. In the second analysis, we performed a descriptive statistical analysis of the trend in dosing deviations over the entire SAI period according to the type of push notification (type 1 [dosing notification only], type 2 [both dosing notification and dosing reminder], and type 3 [no push notifications]) provided, followed by *t* tests of means between groups.

Ethical Considerations

This study used a retrospective SAI dataset gathered through the use of DoseEase (CareSquare, Inc), the inaugural

XSL•FO
RenderX

asynchronous VSMS developed in Korea. Each individual clinical trial that provided data for this study was approved by the hospital's institutional review board (IRB) before conduct and documented informed consent was obtained from all participants; however, this study is a secondary analysis of anonymized data already available. The dataset for this study included only SAI results interpreted by the investigators in each trial, and no personally identifiable information, including videos, was accessed or analyzed. We used the trial protocol number as the study identifier and the participant number assigned per protocol as the participant identifier. Due to the nature of this study, no additional consent was required for the individual participants who provided data in each trial, and the overall study design, including the consent waiver, was reviewed and approved by The Catholic University of Korea Songeui Campus IRB (MC23EASI0052, approval date: July 11, 2023).

Results

Cohort Characteristics

There were 10 trials included in the analysis based on the selection criteria, resulting in 14 unique cohort datasets. The Cohort characteristics could be summarized based on IMP characteristics, enrollment and completion rates, SAI schedule, push notification settings, and allowed time window for SAI. The route of administration was oral in 12 cohorts (86%, 12/14) and ophthalmic in 2 cohorts (14%, 2/14). The IMPs were formulated in various forms, including tablets, capsules, and powders. High completion rates were observed in 13 cohorts (93%, 13/14), with more than 80% of participants completing the study, and some cohorts achieving 100% completion (Table 1).

Table 1. Investigational medicinal product types and participant disposition of each Cohort.

Cohort	Dosing route	Formulation	Number of participants		
			Enrolled, n	Dropout, n	Completed n/N (%)
1	Oral	Tablet	90	8	82/90 (91)
2	Oral	Tablet	15	3	12/15 (80)
3	Oral	Tablet	23	4	19/23 (83)
4	Oral	Capsule or tablet	36	5	31/36 (86)
5	Oral	Tablet	40	9	31/40 (78)
6	Oral	powder	44	1	43/44 (98)
7	Oral	Capsule	8	1	7/8 (88)
8	Oral	Powder	45	0	45/45 (100)
9	Oral	Tablet	15	0	15/15 (100)
10	Oral	Tablet	15	2	13/15 (87)
11	Oral	Tablet	15	1	14/15 (93)
12	Ophthalmic	Eye drops	16	2	14/16 (88)
13	Ophthalmic	Eye drops	16	2	14/16 (88)
14	Oral	Tablet	29	2	27/29 (93)

The SAI schedules varied widely, ranging from 5 to 94 days, with daily frequencies varying from 1 to 4 times. In some cohorts, interinstitutional administration and SAIs were performed at different times on the same day. The planned dosing time was centered around a reference point (eg, 9 AM) for each cohort, with only minor variability in minutes across participants (eg, 9:02, 9:04, 9:06, etc). Push notifications were used in 11 cohorts (79%, 11/14), with the majority (73%, 8/11) using both dosing notifications and dosing reminders (Table 2).

XSL•FO
RenderX

Table 2. Self-administration of the investigational medicinal product-specific characteristics by cohort.

Cohort	SAI[a] schedule				Dosing notification (min before the planned dosing time)	Dosing reminder (min after the planned dosing time)	Allowed time window for SAI (min)
	SAI day	Number of total SAI days	Number of SAIs per day	Planned dosing time[b]			
1	2-4, 6-8, 10-12, 14-16, 18, 19	14	1	9 AM	–60, –30, 0	15, 30, and 45	–60 to 60
2	1, 2, 6, 7, 9-11	7	1	9 AM	N/A[c]	N/A	–60 to 60
3	1, 2, 4-6, 10, 11	7	1	9 AM	N/A	N/A	–60 to 60
4	1, 2, 12-15, 17-20, 24, 25	12	1	9 AM	N/A	N/A	–60 to 60
5	1-4, 13-17, 21-23	12	1	9 AM	–60, –30, 0	15, 30, and 45	–60 to 60
6	1-32	32	2	Variable[d]	–60, –15, 0	60 and 110	–120 to 120
7	3, 5, 7, 9, 10, 12-14, 16-21, 23-30, 32-45, 47-60, 62-75, 77-90	78	1	9 AM	–30, –5, 0	10, 20, 165, and 665	–180 to 180
8	1-28	28	2	Variable[d]	–60, –10, 0	60, 180, and 350	–120 to 360
9	3-7	5	2[e], 3	8 AM, 2 PM, and 8 PM	–30, 0	N/A	–120 to 120
10	4-8, 11	6	2[e], 3	8 AM, 2 PM, and 8 PM	–30, 0	N/A	–120 to 120
11	5-9, 12	6	2[e], 3	8 AM, 2 PM, and 8 PM	–30, 0	N/A	–120 to 120
12	2-4, 8-96	92	1[e], 2[e], 4	1 AM, 9 AM, 9:10 AM, 5 PM	–10, 0	10	N/A
13	1-89, 92-96	94	2[e], 3	1 AM, 9 AM, 5 PM	–10, 0	10	N/A
14	2, 3, 5, 6, 8, 9, 11	7	1	9 AM	–30, 0	30	–60 to 60

[a]SAI: self-administration of the investigational medicinal product.

[b]The presented data refers to the reference point for the planned dosing time (eg, 9:00 AM) in each cohort, with minor variability across participants (eg, 9:02, 9:04, 9:06, etc).

[c]N/A: not applicable.

[d]The planned dosing times vary across participants, and no common reference point can be provided.

[e]SAI event and dosing by investigator performed on the same day.

Fundamental Evaluation of the Usefulness-Success Rate of the SAI Events

A comprehensive analysis was conducted on 17,619 scheduled SAI events across the 14 cohorts to evaluate the system's ability to fulfill its intended use. The analysis revealed a high success rate of 97% (17,151/17,619), while 3% (468/17,619) were unsuccessful due to issues such as unclear video recordings or technical difficulties. Among the successful events, 99% (16,975/17,151) were validated by investigators as verified on-time dosing, indicating that the actual dosing time was within the allowed SAI time window and the dosing behavior recorded in the video was appropriate.

A total of 127 SAI days were analyzed, with over 90% (367/407) of participants consistently reporting dosing videos on all SAI days. The majority of SAI days (95%, 121/127) had more than 90% objective dosing data, with some days requiring additional verification via phone to ensure accurate data capture. This high rate of validated dosing (verified on-time dosing and verified deviated dosing) indicates that the system effectively supported the accurate SAI (Table 3 and Multimedia Appendix 1).

XSL•FO
RenderX

Table 3. Self-administration of the investigational medicinal product success rate by Cohort and by study day. Success rate (%)=(validated dosing events/planned SAIs[a] in each study day) × 100.

Cohort	Study day															Cohort overall
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	≥15	
1	N/A[b]	99	99	99	N/A	100	99	97	N/A	99	100	98	N/A	99	99	99
2	93	100	N/A	N/A	N/A	92	100	N/A	100	100	100	N/A	N/A	N/A	N/A	98
3	100	100	N/A	96	100	100	N/A	N/A	N/A	100	95	N/A	N/A	N/A	N/A	98
4	100	100	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	100	100	100	100	100
5	95	98	100	100	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	97	100	100	99
6	83[c]	91	95	93	97	91	89[c]	90	94	92	93	95	91	93	95	93
7	N/A	N/A	100	N/A	100	N/A	100	N/A	100	100	N/A	100	100	100	100	100
8	0[c]	92	92	94	97	96	94	97	98	97	94	97	99	97	97	93
9	N/A	N/A	87[c]	91	98	96	93	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	93
10	N/A	N/A	N/A	100	98	100	100	100	N/A	N/A	100	N/A	N/A	N/A	N/A	100
11	N/A	N/A	N/A	N/A	100	96	98	89[c]	96	N/A	N/A	79[c]	N/A	N/A	N/A	93
12	N/A	100	100	100	N/A	N/A	N/A	100	100	100	100	96	100	98	100	99
13	100	100	94	100	100	100	100	100	100	98	100	100	100	98	99	99
14	N/A	100	100	N/A	100	100	N/A	100	100	N/A	100	N/A	N/A	N/A	N/A	100
By study day	66	96	96	96	97	97	96	95	98	97	98	96	97	97	98	N/A

[a]SAI: self-administration of the investigational medicinal product.

[b]N/A (not applicable) indicates SAI was not performed on the study day.

[c]Success rate is less than 90%.

Trends in Dosing Time Deviations by Cohort

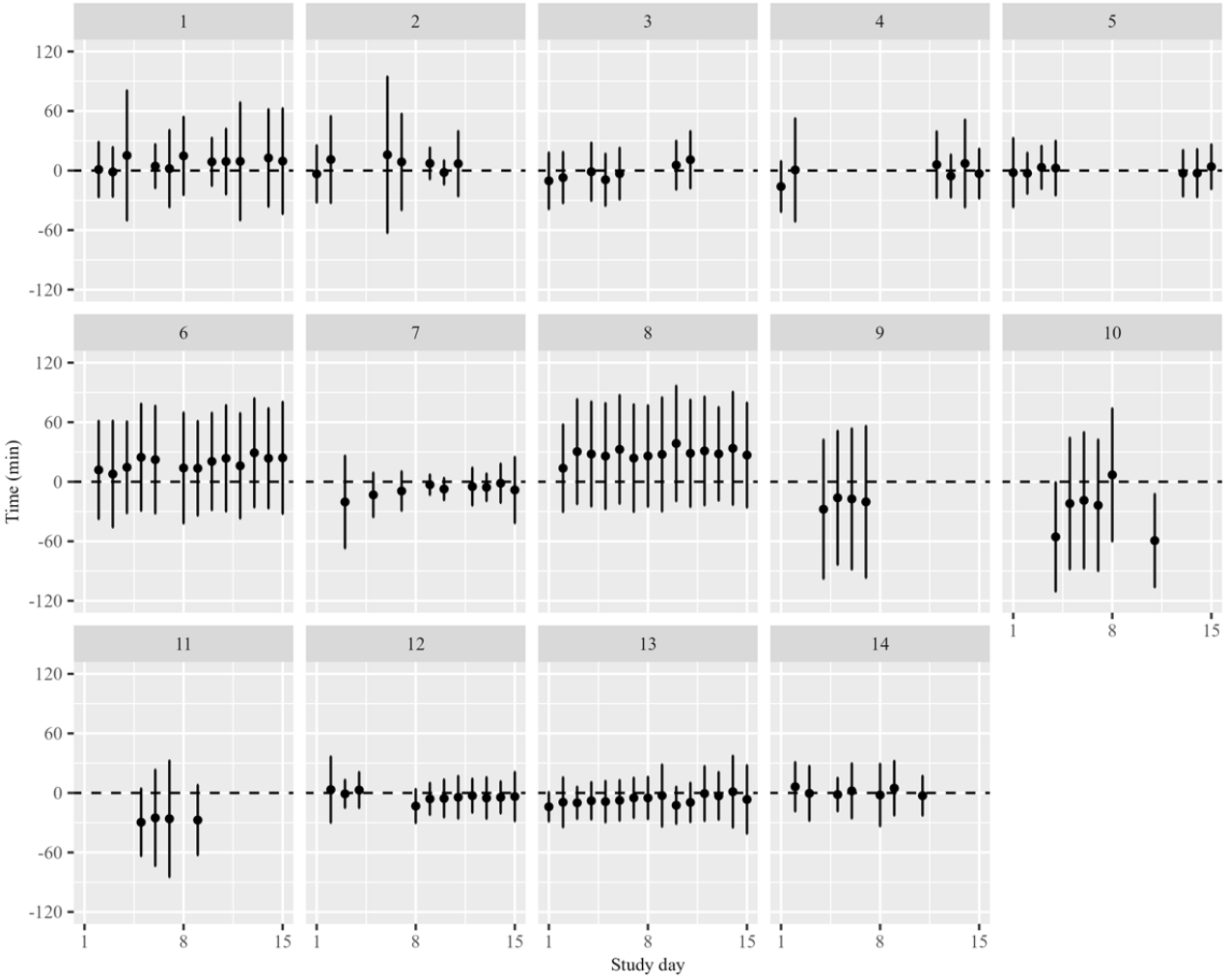
The pattern of PADEV by cohort appeared to be random and did not show any consistent trend depending on the characteristics of the IMP or the scheme of the SAI. For almost all SAI events, the SD of PADEV was greater than the mean, showing that dosing deviations per participant were highly variable around the planned dosing time. Cohort 10 had the earliest mean PADEV during the initial period at –29.6 (SD 65.8) minutes, and this trend remained consistent in subsequent periods, with a mean deviation of –27.2 (SD 65.9) minutes. Similarly, cohort 8 had the latest mean PADEV of 24.0 (SD 50.5) minutes during the initial period and maintained a similar pattern thereafter with a deviation of 27.6 (SD 53.1) minutes.

A trend toward delayed dosing was observed in cohorts 1, 2, 6, 8, 12, and 14, especially during the initial period. For example, cohort 1 had a mean deviation of 5.0 (SD 44) minutes in the initial period, with subsequent deviations ranging from 1.9 (SD 38.9) minutes to 14.8 (SD 39.5) minutes. Cohort 6 had an initial mean deviation of 11.3 (SD 49.9) minutes, which increased to 21.5 (SD 54.5) minutes in the second half. Conversely, cohorts 3, 4, 5, 7, 9, 10, 11, and 13 tended to dose earlier than the planned time, and this trend was maintained over time. For example, cohort 3 had an initial mean deviation of –6.3 (SD 27.8) minutes, which was followed by a deviation of –2.7 (SD 27.5) minutes. Cohort 7 had an initial deviation of –14.3 (SD 30.9) minutes and remained at the same deviation of –8.3 (SD 32.4) minutes afterward (Figure 2 and Multimedia Appendix 2). While some SAIs were conducted at times that may be outside of the participants' daily routine (after 10 PM and before 7 AM), the overall SAI pattern was not significantly different from the typical time of day.

Tests of comparison between reference (the first 3 study days in which SAIs were performed) and subsequent study days showed no significant differences in dosing time deviations for most cohorts. However, significant differences (P<.05) were observed on 13 of 79 (16%) days, indicating occasional deviations from the initial trend (Table 4). For example, cohort 12 had a significant difference on day 8 with a mean deviation of –13.3 (SD 17.1) minutes compared with 1.8 (SD 23.0) minutes in the initial period. The detailed results of the analysis are presented in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 2. Distribution of the dosing time deviation by study day in each cohort. The distribution illustrates the trends in self-administration times among participants over the course of the study, demonstrating the extent to which dosing occurs earlier or later in comparison to on-time dosing (0 minutes).



XSL•FO
RenderX

Table 4. Proportion of study days with the significant difference in mean dosing time from reference day.

Cohort	Proportion, % (n/N)[a]
1	0 (0/8)
2	0 (0/4)
3	25 (1/4)
4	0 (0/3)
5	0 (0/4)
6	40 (4/10)
7	0 (0/6)
8	9 (1/11)
9	0 (0/1)
10	67 (2/3)
11	0 (0/1)
12	13 (1/8)
13	25 (3/12)
14	0 (0/4)
Total	16 (13/79)

[a]Proportion (number of study days with a significant mean PADEV [difference between planned and actual dosing time] difference from the reference period/total study days with SAI excluding the reference period).

For type 1 cohorts, which included Cohorts 9, 10, and 11 with a total of 525 SAI data points, the average PADEV was –24.9 (SD 63) minutes. This indicates a tendency to administer doses earlier than planned when only notifications were used. type 2 cohorts included cohorts 1, 5, 7, 12, 13, and 14; cohorts 6 and 8 were excluded due to potential bias in PADEV due to formulation differences (powder). The mean and SD of the PADEV was –3.5 (SD 31.3) minutes, and it should be noted that this type contained significantly more SAI events than the other types. This suggests that the SD of other types is likely overestimated compared to this type. In a mean difference test performed to account for this possibility (without the equal variance assumption), this type was significantly different from type 1, but not from type 3. The cohorts with no push notifications (type 3), including cohorts 2, 3, and 4 with 639 SAI data points, showed an average PADEV of –1.3 (SD 32.2) minutes. Despite the lack of notifications, these cohorts also maintained relatively small mean deviation, similar to type 2 (Table 5 and Figure 3).

XSL•FO
RenderX

Table 5. Distribution of dosing time deviation by push notification type and study day.

Study day	Deviation (minutes), mean (SD)		
	Type 1[a] (n[d]=525)	Type 2[b] (n[d]=11,297)	Type 3[c] (n[d]=639)
1	N/A[e]	–7.5 (28)	–11.9 (27.2)
2	N/A	–1.1 (26.4)	0.4 (43.7)
3	N/A	–2.8 (24.1)	N/A
4	39.5 (65.2)	5.8 (48.3)	–1.2 (29.3)
5	–21.8 (60)	–6.8 (19.9)	–9.3 (26.3)
6	–20.5 (63.1)	0.6 (23.2)	3.6 (50.9)
7	–23.4 (66.9)	–1.0 (33.0)	8.7 (48.6)
8	6.9 (67)	3.9 (34.3)	N/A
9	–27.4 (35.4)	–2.6 (24.5)	7.2 (16)
10	N/A	–1.0 (22.9)	2.4 (20.6)
11	–59.4 (47)	0.0 (27.3)	9.2 (30.1)
12	N/A	2.9 (42.9)	5.9 (33.6)
13	N/A	–3.9 (22.1)	–5.5 (21.6)
14	N/A	3.3 (36.8)	7.1 (44.2)
15	N/A	–4.5 (31.2)	–3.1 (25.1)
Total period	–24.9 (63)	–3.5 (31.3)[f]	–1.3 (32.2)[g]

[a]Type 1 includes cohorts that received only dosing notification.

[b]Type 2 includes cohorts that received both dosing notification and dosing reminder.

[c]Type 3 includes cohorts that did not receive any push notifications.

[d]Number of self-administration events.

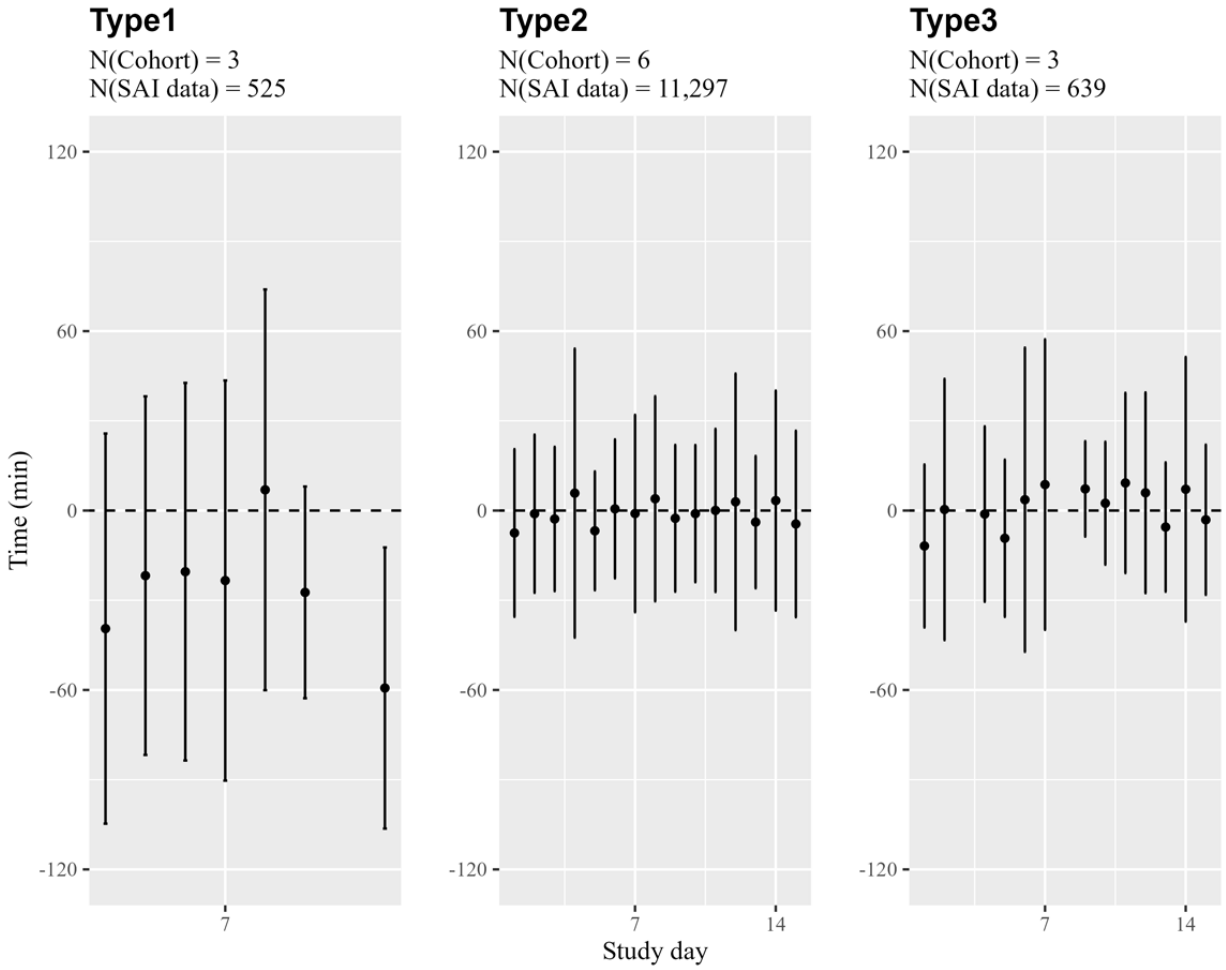
[e]N/A: not applicable.

[f]$P<.001$ on comparing with type 1.

[g]$P<.001$ on comparing with type 1 and $P=.09$ on comparing with type 2.

XSL•FO
RenderX

Figure 3. Trends in the dosing time deviation by push notification type. SAI: self-administration of the investigational medicinal product. Push notification types are as follows: type 1 (dosing notification only), type 2 (both dosing notification and dosing reminder), and type 3 (no push notifications). The distribution illustrates the trends in self-administration times for each type based on the push notifications sent through the VSMS. VSMS: video-based self-administration of the investigational medicinal product monitoring system.



Discussion

Principal Findings

The cohort characteristics of this study highlight key considerations for the implementation and scalability of VSMS. In all cohorts, high adherence was mandatory to achieve the study objectives, and participants provided informed consent before participation. This is likely the primary reason for the relatively high adherence observed across cohorts compared with that observed in later clinical trials or real-world settings [5,26]. Nevertheless, the clear benefits of using a VSMS were consistent with previous studies of similar systems [16-18,20-25,27-30]. By providing visual evidence of SAI, the VSMS enabled SAI instead of direct observation, and the video-based, real-time management allowed investigators to monitor adherence during the dosing period and intervene in a timely manner to minimize noncompliance [17,23-25,31]. In this study, additional measures to improve the reliability of the data were proposed by clarifying the outcome categorization system of verified on-time dosing, verified deviated dosing, unverified dosing, and missed dosing. While VSMS increases the participant's performance burden for the SAI itself, it reduces the inconvenience of site visits for dosing and minimizes the potential for participant dropout due to nonadherence [6,27,32], expanding the feasibility of a variety of clinical studies that require full adherence and its evidence. Conversely, the use of VSMS in studies that only investigate the percentage of planned doses actually taken and do not require thorough evidence of adherence will need to be determined on a cost-benefit basis.

Adherence in cohorts with SAI using VSMS was similar to that observed in repeated-dose early-phase clinical trials based on direct observation (where almost 100% adherence is required for participants). Furthermore, the concept of adherence extends not only to the performance of SAIs, but also to the use of the provided system to record SAI behavior and transmit it to the researcher. Therefore, high adherence in studies using VSMS indicates that the essential requirements of the system have been met, which (1) technical reliability: the system operated without significant technical errors, ensuring that data capture and processing was accurate and reliable; (2) participant acceptability: participants were able to use the mobile-based application without difficulty, which is critical to maintaining high levels of compliance in clinical trials; and (3)

XSL•FO
RenderX

comprehensiveness for investigators: the system provided sufficient information in an appropriate manner to enable investigators to make informed decisions regarding the validation of SAI. Along with the intuitive user interface, user acceptability can be attributed in part to the initial training during the first visit, including a mock SAI session.

The study observed that SAI behaviors established early in the trial were generally maintained throughout the study period. Only 16% (13/79) of events showed statistically significant differences in dosing times compared with the initial SAI events. When using dosing notifications alone, a deviation toward faster SAI times was observed compared with using dosing reminders together or no push notifications at all. The use of dosing notification alone was associated with a faster deviation to SAI time compared with the use of dosing reminder or no push notification at all. Even accounting for the difference in the number of events in each group, the variability in deviation was greatest for dosing notification alone. The fact that the distribution of deviations was similar between no push notification group and both dosing notification and reminder group suggests that participants were highly motivated to adhere to their planned dosing times. In this situation, using only the dosing notification would have resulted in participants choosing to perform SAI before they forgot to do it, and there would have been significant interparticipant variation in this choice. Although push notifications were preset and remained constant throughout the study period in the cohorts included in these studies, given the nature of SAI behavior and the impact of push notifications, it is possible that better compliance could be achieved by adjusting push notifications based on the initial SAI pattern of the cohort. Particularly, it is considered necessary to modify the strategy of push notifications by identifying the tendencies of the participants at the beginning of the SAI period when only dosing notification is planned.

The VSMS used in this study is compatible with any type of mobile device with video recording and wireless transmission capabilities. As an alternative to installing the VSMS application on a participant's smartphone, a separate mobile device capable of performing VSMS could be provided for the entire study period to conduct the SAI. Some participants in the cohort included in this study participated in this manner. This fact means that the participant or patient does not necessarily need to own a smartphone and agree to install the VSMS application in order to use VSMS. In other words, the usefulness of VSMS depends on the participant's willingness to perform the SAI and the appropriate use of the VSMS application, rather than the ownership of a mobile device.

Limitations

The findings were derived from a population with limited demographic characteristics. The usefulness of VSMS is likely to be contingent upon the participant's or patient's capability to use mobile apps and the mobile environment in which they reside. It is expected that prospective evaluations in larger and more diverse populations, including different participant characteristics in terms of age, dosing regimens, or indications, will provide even more reliable evidence.

Conclusions

VSMS proved to be an effective tool for obtaining dosing information with accuracy comparable to direct observation, even in remote situations. The potential to minimize adherence deterioration through various alarm features and appropriate investigator or observer intervention was also identified. In many research or clinical situations where participant adherence is essential, but direct observation is difficult, VSMS may be a key element in obtaining dosing information. Proper usage of participants' smartphones or other mobile devices and previous training of participants and observers may improve the accuracy of SAI recording via VSMS. It is expected that the usage and usefulness of VSMS will be continuously improved through the accumulation of experience in various medical fields.

Acknowledgments

This research was supported by the "Korea National Institute of Health" (KNIH) research project (2023-ER0809-01).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The proportion of SAI (self-administration of the investigational medicinal product) outcome by category on each cohort.
[XLSX File (Microsoft Excel File), 80 KB - mhealth_v13i1e65668_app1.xlsx]

Multimedia Appendix 2
The results of the comparison between the mean PADEV (difference between planned and actual dosing time) on the reference study day and the mean PADEV on each subsequent study day, along with the distribution of daily PADEV.
[XLSX File (Microsoft Excel File), 107 KB - mhealth_v13i1e65668_app2.xlsx]

References

XSL•FO
RenderX

1. Kubitza D, Becka M, Wensing G, Voith B, Zuehlsdorf M. Safety, pharmacodynamics, and pharmacokinetics of BAY 59-7939-an oral, direct Factor Xa inhibitor-after multiple dosing in healthy male subjects. Eur J Clin Pharmacol 2005;61(12):873-880. [doi: 10.1007/s00228-005-0043-5] [Medline: 16328318]
2. Lee HW, Kang WY, Jung W, Gwon M, Cho K, Yang DH, et al. Evaluation of the pharmacokinetic drug-drug interaction between micronized fenofibrate and pitavastatin in healthy volunteers. Pharmaceutics 2020;12(9):869 [FREE Full text] [doi: 10.3390/pharmaceutics12090869] [Medline: 32932576]
3. Wymer J, Apple S, Harrison A, Hill BA. Pharmacokinetics, bioavailability, and swallowing safety with riluzole oral film. Clin Pharmacol Drug Dev 2023;12(1):57-64 [FREE Full text] [doi: 10.1002/cpdd.1168] [Medline: 36168148]
4. Park LG, Howie-Esquivel J, Dracup K. Electronic measurement of medication adherence. West J Nurs Res 2015;37(1):28-49. [doi: 10.1177/0193945914524492] [Medline: 24577868]
5. Matsui D. Strategies to Measure and Improve Patient Adherence in Clinical Trials. Pharm Med 2012;23(5-6):289-297 [FREE Full text] [doi: 10.1007/BF03256784]
6. Mason M, Cho Y, Rayo J, Gong Y, Harris M, Jiang Y. Technologies for medication adherence monitoring and technology assessment criteria: narrative review. JMIR Mhealth Uhealth 2022;10(3):e35157 [FREE Full text] [doi: 10.2196/35157] [Medline: 35266873]
7. Poongothai S, Anjana R, Aarthy R, Unnikrishnan R, Narayan KMV, Ali M, et al. Strategies for participant retention in long term clinical trials: a participant -centric approaches. Perspect Clin Res 2023;14(1):3-9 [FREE Full text] [doi: 10.4103/picr.picr_161_21] [Medline: 36909219]
8. Vallejos X, Wu C. Digital medicine: innovative drug-device combination as new measure of medication adherence. J Pharm Technol 2017;33(4):137-139 [FREE Full text] [doi: 10.1177/8755122517704212] [Medline: 34860983]
9. Cross A, Gupta N, Liu B, Nair V, Kumar A, Kuttan R. 99DOTS: a low-cost approach to monitoring and improving medication adherence. 2019 Presented at: ICTD '19: Proceedings of the Tenth International Conference on Information and Communication Technologies and Development; 2019 January 4; Ahmedabad, India URL: https://dl.acm.org/doi/10.1145/3287098.3287102
10. Liu X, Blaschke T, Thomas B, De Geest S, Jiang S, Gao Y, et al. Usability of a medication event reminder monitor system (MERM) by providers and patients to improve adherence in the management of tuberculosis. Int J Environ Res Public Health 2017;14(10):1115 [FREE Full text] [doi: 10.3390/ijerph14101115] [Medline: 28946683]
11. Stirratt MJ, Dunbar-Jacob J, Crane HM, Simoni JM, Czajkowski S, Hilliard ME, et al. Self-report measures of medication adherence behavior: recommendations on optimal use. Transl Behav Med 2015;5(4):470-482 [FREE Full text] [doi: 10.1007/s13142-015-0315-2] [Medline: 26622919]
12. McCann DJ, Petry NM, Bresell A, Isacsson E, Wilson E, Alexander RC. Medication nonadherence, "Professional Subjects," and apparent placebo responders: overlapping challenges for medications development. J Clin Psychopharmacol 2015;35(5):566-573 [FREE Full text] [doi: 10.1097/JCP.0000000000000372] [Medline: 26244381]
13. Garfield S, Clifford S, Eliasson L, Barber N, Willson A. Suitability of measures of self-reported medication adherence for routine clinical use: a systematic review. BMC Med Res Methodol 2011;11:149 [FREE Full text] [doi: 10.1186/1471-2288-11-149] [Medline: 22050830]
14. Nguyen T, La Caze A, Cottrell N. What are validated self-report adherence scales really measuring?: a systematic review. Br J Clin Pharmacol 2014;77(3):427-445 [FREE Full text] [doi: 10.1111/bcp.12194] [Medline: 23803249]
15. Garber MC, Nau DP, Erickson SR, Aikens JE, Lawrence JB. The concordance of self-report with other measures of medication adherence: a summary of the literature. Med Care 2004;42(7):649-652. [doi: 10.1097/01.mlr.0000129496.05898.02] [Medline: 15213489]
16. DiStefano MJ, Schmidt H. mHealth for tuberculosis treatment adherence: a framework to guide ethical planning, implementation, and evaluation. Glob Health Sci Pract 2016;4(2):211-221 [FREE Full text] [doi: 10.9745/GHSP-D-16-00018] [Medline: 27353615]
17. Holzman SB, Atre S, Sahasrabudhe T, Ambike S, Jagtap D, Sayyad Y, et al. Use of smartphone-based Video Directly Observed Therapy (vDOT) in tuberculosis care: single-arm, prospective feasibility study. JMIR Form Res 2019;3(3):e13411 [FREE Full text] [doi: 10.2196/13411] [Medline: 31456581]
18. Chuck C, Robinson E, Macaraig M, Alexander M, Burzynski J. Enhancing management of tuberculosis treatment with video directly observed therapy in New York City. Int J Tuberc Lung Dis 2016;20(5):588-593. [doi: 10.5588/ijtld.15.0738] [Medline: 27084810]
19. DeMaio J, Schwartz L, Cooley P, Tice A. The application of telemedicine technology to a directly observed therapy program for tuberculosis: a pilot project. Clin Infect Dis 2001;33(12):2082-2084. [doi: 10.1086/324506] [Medline: 11698993]
20. Mirsaeidi M, Farshidpour M, Banks-Tripp D, Hashmi S, Kujoth C, Schraufnagel D. Video directly observed therapy for treatment of tuberculosis is patient-oriented and cost-effective. Eur Respir J 2015;46(3):871-874 [FREE Full text] [doi: 10.1183/09031936.00011015] [Medline: 25792632]
21. Wade VA, Karnon J, Eliott JA, Hiller JE. Home videophones improve direct observation in tuberculosis treatment: a mixed methods evaluation. PLoS One 2012;7(11):e50155 [FREE Full text] [doi: 10.1371/journal.pone.0050155] [Medline: 23226243]

XSL•FO
RenderX

22. Holzman SB, Zenilman A, Shah M. Advancing patient-centered care in tuberculosis management: a mixed-methods appraisal of video directly observed therapy. Open Forum Infect Dis 2018;5(4):ofy046 [FREE Full text] [doi: 10.1093/ofid/ofy046] [Medline: 29732378]
23. Garfein RS, Collins K, Muñoz F, Moser K, Cerecer-Callu P, Raab F, et al. Feasibility of tuberculosis treatment monitoring by video directly observed therapy: a binational pilot study. Int J Tuberc Lung Dis 2015;19(9):1057-1064 [FREE Full text] [doi: 10.5588/ijtld.14.0923] [Medline: 26260824]
24. Shields MD, ALQahtani F, Rivey MP, McElnay JC. Mobile direct observation of therapy (MDOT) - A rapid systematic review and pilot study in children with asthma. PLoS One 2018;13(2):e0190031 [FREE Full text] [doi: 10.1371/journal.pone.0190031] [Medline: 29401500]
25. Sekandi JN, Buregyeya E, Zalwango S, Dobbin KK, Atuyambe L, Nakkonde D, et al. Video directly observed therapy for supporting and monitoring adherence to tuberculosis treatment in Uganda: a pilot cohort study. ERJ Open Res 2020;6(1):00175-02019 [FREE Full text] [doi: 10.1183/23120541.00175-2019] [Medline: 32280670]
26. Le Flohic E, Vrijens B, Hiligsmann M. The impacts of undetected nonadherence in phase II, III and post-marketing clinical trials: an overview. Br J Clin Pharmacol 2024;90(8):1984-2003. [doi: 10.1111/bcp.16089] [Medline: 38752447]
27. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
28. Sekandi JN, Onuoha NA, Buregyeya E, Zalwango S, Kaggwa PE, Nakkonde D, et al. Using a mobile health intervention (DOT Selfie) With transfer of social bundle incentives to increase treatment adherence in tuberculosis patients in Uganda: protocol for a randomized controlled trial. JMIR Res Protoc 2021;10(1):e18029 [FREE Full text] [doi: 10.2196/18029] [Medline: 32990629]
29. Krueger K, Ruby D, Cooley P, Montoya B, Exarchos A, Djojonegoro BM, et al. Videophone utilization as an alternative to directly observed therapy for tuberculosis. Int J Tuberc Lung Dis 2010;14(6):779-781. [Medline: 20487619]
30. Creary SE, Gladwin MT, Byrne M, Hildesheim M, Krishnamurti L. A pilot study of electronic directly observed therapy to improve hydroxyurea adherence in pediatric patients with sickle-cell disease. Pediatr Blood Cancer 2014;61(6):1068-1073. [doi: 10.1002/pbc.24931] [Medline: 24436121]
31. Hoffman JA, Cunningham JR, Suleh AJ, Sundsmo A, Dekker D, Vago F, et al. Mobile direct observation treatment for tuberculosis patients: a technical feasibility pilot using mobile phones in Nairobi, Kenya. Am J Prev Med 2010;39(1):78-80. [doi: 10.1016/j.amepre.2010.02.018] [Medline: 20537846]
32. Zweben A, Fucito LM, O'Malley SS. Effective Strategies for Maintaining Research Participation in Clinical Trials. Drug Inf J 2009 Dec 30;43(4):459-467 [FREE Full text] [doi: 10.1177/009286150904300411] [Medline: 24311825]

Abbreviations

IMP: investigational medicinal product
IRB: institutional review board
PADEV: difference between planned and actual dosing time
SAI: self-administration of the investigational medicinal product
VSMS: video-based self-administration of the investigational medicinal product monitoring system

Edited by L Buis; submitted 22.08.24; peer-reviewed by PJ Rudzki, RL Randell; comments to author 04.10.24; revised version received 03.11.24; accepted 22.02.25; published 07.04.25.

Please cite as:
Song J, Han S, Choi S, Lim J, Oh BY, Shin D, Han S
Participant Adherence in Repeated-Dose Clinical Studies Using Video-Based Observation: Retrospective Data Analysis
JMIR Mhealth Uhealth 2025;13:e65668
URL: https://mhealth.jmir.org/2025/1/e65668
doi:10.2196/65668
PMID:

©Jihong Song, Sungpil Han, Suein Choi, Jonghyuk Lim, Byeong Yeob Oh, Dongoh Shin, Seunghoon Han. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 07.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Enhancing Cardiopulmonary Resuscitation Quality Using a Smartwatch: Neural Network Approach for Algorithm Development and Validation

Gaurav Rao[1], PhD; David W Savage[2], MD; Gabrielle Erickson[2], MD; Nathan Kyryluk[2], MD; Pawan Lingras[1], PhD; Vijay Mago[3], PhD

[1]Department of Mathematics and Computing, Faculty of Science, Saint Mary's University, Halifax, NS, Canada

[2]Emergency Medicine, Faculty of Family and Emergency Medicine, NOSM University, Thunder Bay, ON, Canada

[3]School of Health Policy and Management, Faculty of Health, York University, Toronto, ON, Canada

Corresponding Author:
Gaurav Rao, PhD
Department of Mathematics and Computing
Faculty of Science
Saint Mary's University
923 Robie St
Halifax, NS, B3H3C3
Canada
Phone: 1 9024205400
Email: gaurav.rao@smu.ca

Abstract

Background: Sudden cardiac arrest is a major cause of mortality, necessitating immediate and high-quality cardiopulmonary resuscitation (CPR) for improved survival rates. High-quality CPR is defined by chest compressions at a rate of 100-120 per minute and a depth of 50-60 mm. Monitoring and maintaining these parameters in real time during emergencies remain a challenge.

Objective: This study introduces a neural network model designed to predict and assess CPR quality using accelerometer data from a smartwatch.

Methods: The study involved 83 participants performing CPR on mannequins, with accelerometer data collected via smartwatches worn by the participants. These data were aligned with gold-standard data from the mannequins. The accelerometer-derived compression data were segmented into 5-second intervals for training the neural network models. A total of 1226 neural network models were developed, incorporating variations in hyperparameters and dataset configurations to optimize performance.

Results: The optimal model demonstrated the capability to accurately predict the number of compressions and the average compression depth within a 5-second interval. The model achieved an accuracy of ±3.8 mm for compression depth and an average deviation of 0.8 compressions. The results indicated that the neural network model could accurately assess CPR quality metrics, surpassing other models discussed in the literature. The large and diverse dataset used in this study contributed to the robustness and reliability of the model.

Conclusions: This study validates the efficacy of a neural network model in accurately predicting CPR metrics using smartwatch accelerometer data. The model outperforms previous methods and shows promise for real-time feedback during CPR. Future work involves deploying the model directly on smartwatches for real-time application, potentially improving sudden cardiac arrest survival rates through immediate and accurate feedback on CPR quality.

(JMIR Mhealth Uhealth 2025;13:e57469) doi:10.2196/57469

KEYWORDS

CPR performance; CPR feedback; neural network; smart health; smartwatch; sudden cardiac arrest; mobile phone; cardiac arrest; cardiopulmonary resuscitation; compressions; monitoring; emergency; network model; wearables; chest compression; efficacy; data collection

XSL•FO
RenderX

Introduction

Sudden cardiac arrest (SCA) is identified as a principal cause of mortality in North America, particularly among young athletes [1-4]. It can affect individuals irrespective of their lifestyle or health status, leading to either irregular or nonexistent heart rhythms [5,6]. As a consequence, the flow of blood to major organs is halted, depriving them of essential oxygen, resulting in tissue damage and potentially culminating in organ failure. The probability of survival decreases by about 10% with each passing minute without intervention; therefore, immediate and effective treatment is crucial to enhance survival rates by minimizing damage to tissues and organs [7,8].

Efforts to deliver prompt care to SCA victims continue, with emergency medical services (EMS) prioritizing SCAs to expedite emergency responses [9,10]. Public data indicate that the target for EMS response time to an SCA event is within 8-10 minutes [5,11-14]. The initial treatment involves performing cardiopulmonary resuscitation (CPR) and using an automated external defibrillator (AED) to assist in heart pumping and maintain blood flow to the brain and other vital organs. In certain instances, an AED can also restore the heart's electrical activity [6,15].

Organizations such as the American Heart Association (AHA) and the Red Cross play a pivotal role in providing CPR and AED training to the public, empowering bystanders to administer early care to SCA victims until EMS arrives. Studies indicate that bystander-administered CPR significantly improves survival rates compared with scenarios where CPR is not administered [16-18]. The quality of CPR is critical, as high-quality CPR is associated with increased chances of survival. The AHA defines high-quality CPR as having a compression depth of 5-6 cm and a rate of 100-120 compressions per minute, standards that are also supported by the European Resuscitation Council [19,20].

Training the general public in CPR administration has emerged as a key objective for numerous organizations aiming to improve survival rates for SCA patients [21,22]. If a significant portion of the populace receives training in CPR, SCA victims stand a higher chance of receiving necessary immediate care until EMS arrive, thereby enhancing their survival prospects. Public training programs instruct participants on identifying SCA victims and initiating high-quality CPR. In these sessions, participants engage in CPR practice on mannequins while trainers offer real-time feedback on their performance, ensuring proficiency [23-25]. With advancements in technology, CPR performance can now be quantified, allowing feedback to be grounded in these measurements. Technologies used in training encompass sensor-equipped mannequins and CPR feedback devices.

Mannequins equipped with sensors accurately assess compression depth and frequency, providing trainers with metrics to deliver precise feedback to trainees. Beyond training devices, the industry also offers devices that can be placed on the chest of a mannequin or patient. These devices are equipped with accelerometers and pressure sensors to accurately assess CPR performance and transmit the information to another device for providing real-time feedback. The Laerdal CPRMeter2 is an example of such a device [26].

For SCA incidents, emergency call operators use technologies like tele CPR and video CPR to assist individuals in achieving high-quality CPR [27-29]. Upon identifying an SCA situation, EMS is dispatched, and the caller is guided over the phone on performing CPR. Nevertheless, the operator cannot gauge the quality of compressions administered, providing only verbal instructions to help the caller maintain an appropriate pace, which leaves compression depth uncertain. Video CPR, as an advanced approach, transitions communication from audio to video, allowing the operator to observe the CPR performance and offer real-time feedback. This technology requires advancements in emergency response systems to support video calling capabilities and necessitates video capability on the caller's part. A limitation is its dependency on 2 bystanders—one to execute CPR and another to capture it.

Researchers have delved into next-generation CPR feedback technologies, including virtual reality and augmented reality–based devices [30-33]. These devices employ integrated cameras to capture and analyze compression depth and rate, displaying real-time statistics on the device screen to aid users in enhancing their CPR performance [34,35]. However, the widespread adoption of such advanced technologies encounters obstacles: they are not always accessible to those who do use them.

Therefore, the challenge lies in identifying a device suitable for real-world emergencies that can precisely measure CPR performance and provide appropriate feedback without the necessity for specialized hardware. The literature identifies smartphones and smartwatches as 2 potential solutions [36-38]. Both types of devices are equipped with sensors capable of evaluating CPR performance. Song et al [39] introduced a mobile app that leverages accelerometer data from smartphones to assess compression quality and provide feedback via screen displays and audio cues. Nonetheless, this approach faces limitations; it necessitates attaching the smartphone to the user's arm, lacks details on data cleaning and noise removal, and does not consider variations in device orientation [40]. Similar challenges are evident in other algorithms designed for mobile CPR apps [41-43], with many focusing solely on training scenarios rather than real emergencies.

Gruenerbl et al [44] proposed a smartwatch app capable of measuring CPR parameters and offering visual feedback. This app analyzes accelerometer data to evaluate compression quality, identifying each positive peak on the y-axis as a compression and calculating the differences in y-axis peaks to determine compression depth. However, the study does not provide detailed algorithmic and data-cleaning methodologies for replication and comparison.

Lu et al [45] also proposed a smartwatch app alongside an algorithm for evaluating compression metrics. They tested using a Resusci Anne QCPR training manikin (Laerdal) and an android ASUS ZenWatch 2 (model WI501Q; ASUSTeK Computer Inc). The developed polynomial model predicts compression depth and rate from smartwatch accelerometer data. Although data were collected, its limited variability—compression counts

XSL•FO
RenderX

between 80-140 and depths of 4-7 cm—fails to cover the wider range expected in real-world scenarios, nor does it elaborate on data cleaning or handling irregularities.

Using smartwatch accelerometer data presents several challenges, such as sensor white noise, gravitational effects, hand movements, and shifts in the watch's position due to a loose fit. Filtering such noise from sensor data remains a significant challenge [46]. Employing noisy datasets can result in substantial variations in output results over time. To tackle noise filtration, we propose a neural network model trained on noisy data to predict compression performance, as detailed in the Methods section.

Methods

Overview

This section presents the technology and equipment used during the data collection, cleaning, and processing phases. A total of 83 participants from Thunder Bay Regional Health Sciences Centre (TBRHSC) took part in the data collection by performing CPR in a controlled simulation setting, using an Apple Watch, Apple iPhone, and a Laerdal mannequin for data capture [26,47].

Equipment and Hardware

Participants collected accelerometer data with an Apple Watch Series 7 while performing CPR, which was then stored on an Apple iPhone 14. These devices, chosen for their significant presence in the wearable segment, were equipped with specially developed apps for data collection and analysis [48,49]. This study differs from previous ones by focusing on the potential of the Apple ecosystem for health care apps, rather than on Android devices. The apps were designed for compatibility with Apple iPhone 10 and newer, as well as Apple Watch Series 3 and later models.

CPR performance practice occurred on the Laerdal SimMan 3G mannequin, which is outfitted with sensors that record and provide feedback on key compression quality metrics such as depth, rate, and recoil. This aids in improving CPR quality [50]. Laerdal mannequins were chosen, as in other studies, for their precise data capture capabilities [39,40,44].

Data analysis was performed using a laptop with an i7 Intel processor, 32 GB RAM, and a 1TB hard drive, using Microsoft Excel and Python for data cleaning, processing, model training, and evaluation. Final analyses were done on a cloud server (Amazon Web Services) equipped with an NVIDIA A10G GPU (24 GB GPU Memory), 32 GB RAM, and 8 vCPUs [51].

Data Collection

A convenience sampling method was used in this study. A notice was sent around the hospital and health care providers including doctors, nurses, respiratory therapists, and paramedics were recruited to participate. Data for this study were collected from 83 individuals over a 3-month period. Staff members from TBRHSC were invited to participate in the research, with their CPR experience ranging from none to more than 30 instances. Initially, participants received a brief orientation on CPR techniques and the benchmarks for high-quality CPR performance.

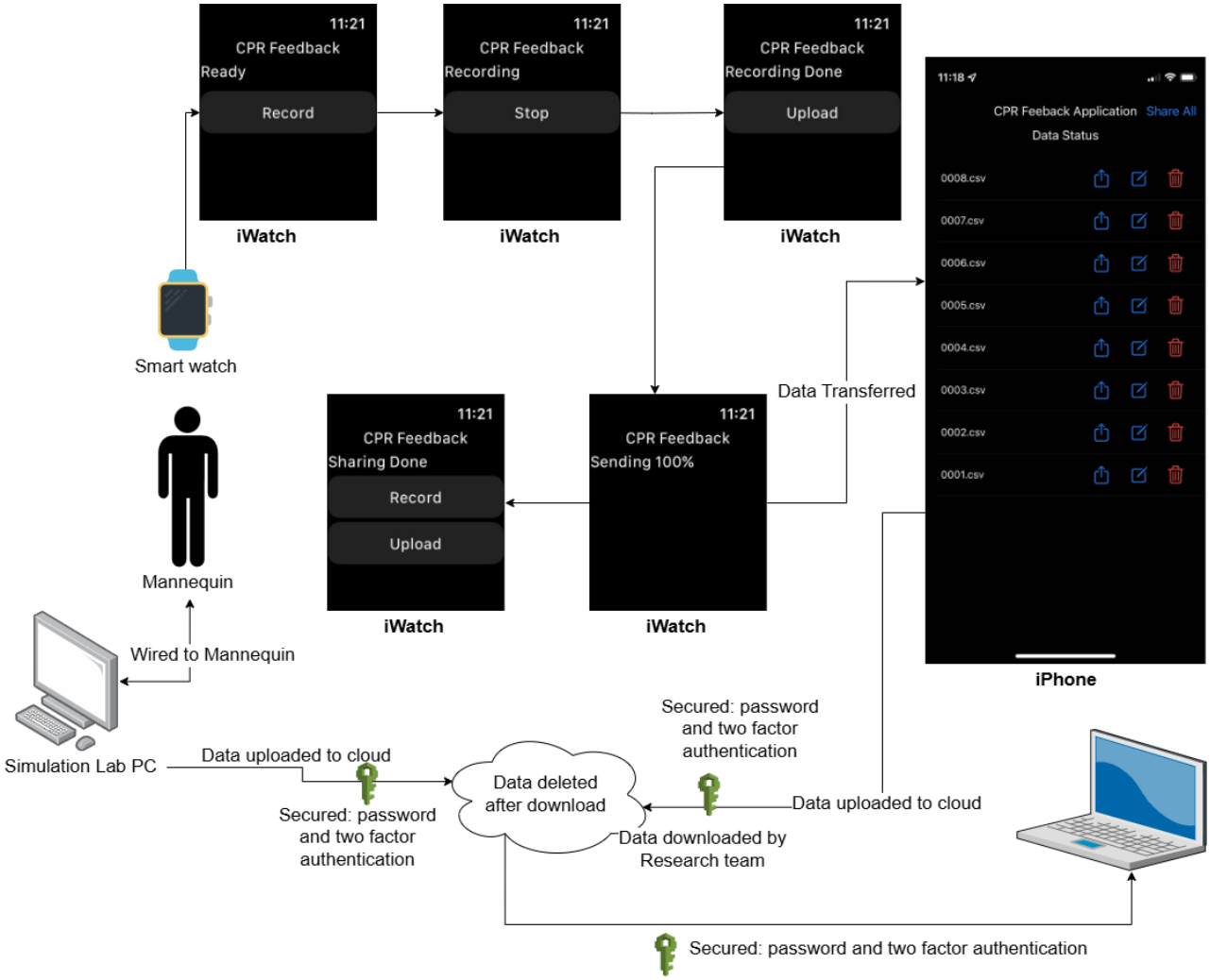
The simulation laboratory at TBRHSC, equipped with a Laerdal SimMan 3G mannequin capable of measuring compression depth, rate, and recoil, served as the setting for data collection. Upon arriving at the laboratory, participants received an Apple Watch containing the necessary data collection application. The participants were given training on how to perform high-quality CPR. After training, the participants were asked to perform CPR for 2 minutes, during which the data would be recorded. During these 2 minutes, participants were not provided with any feedback on their performance. A limited number of participants (n=41) performed another round of 2 minutes of CPR after taking a short 10-minute break.

When participants were ready to perform CPR after training, they were positioned correctly beside the mannequin, placing their hands on its chest, and initiated CPR compressions upon commencing data collection via the Apple Watch and mannequin. This procedure was carried out for 2 minutes, after which participants were advised to keep their hands in place on the mannequin to mitigate the potential introduction of noise in the data caused by removing their hands. The protocol was approved by TBRHSC and is attached as a Multimedia Appendix 1.

Subsequently, data from the Apple Watch and the mannequin were exported with a unique anonymous identifier to link the datasets without disclosing the identities of the participants. The process of data collection and the following data transmission is delineated in Figure 1. The data were securely transferred daily to the lead author's workstation. The app on the Apple Watch recorded accelerometer data, including timestamps at a 50 Hz frequency. The mannequin's data yielded detailed insights into each compression, providing metrics on depth, recoil completeness, start and end times of compression, and the peak time of compression. It is important to note that the mannequin's time stamps were relative to the onset of the data collection period rather than an absolute start time.

XSL•FO
RenderX

Figure 1. The flow diagram shows the screenshots of iPhone and Apple Watch apps, and the data flow used to collect the data during the study. CPR: cardiopulmonary resuscitation.



In aggregate, the study compiled 256 minutes of CPR data, which included 27,844 compressions with an average depth of 44 (SD 4) mm. Of the 83 contributing participants, 27 were male, and 56 were female; 48 of these individuals had previously performed CPR on more than 10 occasions.

The sample size was determined by the number of eligible participants available during the data collection period, using a convenience sampling approach. While no formal power analysis was conducted, the sample size of 83 participants was considered sufficient to meet the study's objectives. Notably, previous studies in similar domains have used smaller sample sizes—Gruenerbl et al [44] included 50 participants, and Lu et al [45] conducted their study with just 8 participants. In addition, some participants in our study performed CPR twice, resulting in 124 distinct datasets, further enhancing the dataset's richness and utility for training and evaluation.

Data Cleaning and Preparation for Model Training

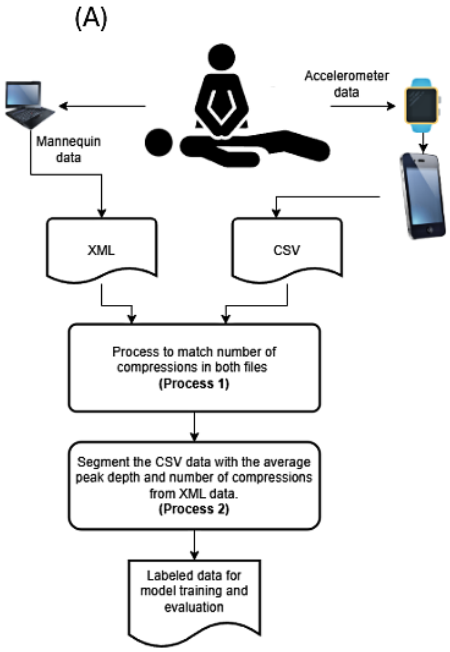
Overview

This study used data from 2 sources—an Apple Watch app and exported data from a mannequin. Each source's data were recorded in separate files, but a unique identifier linked them for consistency throughout the data cleaning and preprocessing stages. Figure 2 illustrates the process of preparing the raw data for model training and the collected sample files.

The mannequin's data, exported as a comprehensive list of compressions during each session, served as the gold standard due to the sensors' precise compression depth, rate, and recoil measurements. Meanwhile, the Apple Watch provided time-stamped accelerometer data across the x-, y-, and z-axes, formatted as comma-separated values (CSVs). Both datasets included detailed compression information, enabling analysis of 2-minute CPR sessions.

XSL•FO
RenderX

Figure 2. (A) The flow diagram illustrates the process for preparing the data for model training and sample data from XML and CSV files. (B) XML data file sample, showing data for 1 compression collected from the Laerdal mannequin. (C) CSV file sample, showing data collected from the Apple smartwatch. CSV: comma-separated value.



(B)

```
<evt type="CprCompEvent" msecs="0">
    <param type="endTime" value="330"/>
    <param type="compDepth" value="43" units="mm"/>
    <param type="compPeakTime" value="160" units="ms"/>
    <param type="compInstantaneousPeriod" value="0" units="ms"/>
    <param type="compAbsolutePeakDepth" value="43" units="mm"/>
    <param type="compLeaningDepth" value="0" units="mm"/>
    <param type="compReleaseDepth" value="7" units="mm"/>
    <param type="compHandposXiphoid" value="0"/>
    <param type="compHandposError" value="0"/>
    <param type="compIncompleteRelease" value="1"/>
    <param type="compMeanRate" value="110"/>
</evt>
```

(C)

id	datetime	x	y	z
1	2023-06-02 11:37:41 +0000	0.496582031	0.469970703	-0.742675781
2	2023-06-02 11:37:41 +0000	0.523681641	0.472412109	-0.733398438
3	2023-06-02 11:37:41 +0000	0.550292969	0.47265625	-0.738525391
4	2023-06-02 11:37:41 +0000	0.530029297	0.451660156	-0.724853516
5	2023-06-02 11:37:41 +0000	0.512695313	0.448730469	-0.670410156
6	2023-06-02 11:37:41 +0000	0.494384766	0.452636719	-0.706054688
7	2023-06-02 11:37:41 +0000	0.503173828	0.469726563	-0.762451172
8	2023-06-02 11:37:41 +0000	0.505126953	0.445068359	-0.701904297
9	2023-06-02 11:37:41 +0000	0.518554688	0.481201172	-0.762451172
10	2023-06-02 11:37:41 +0000	0.506835938	0.451416016	-0.734619141

For each session involving a pair of mannequin and smartwatch datasets, a 2-minute duration was recorded. The data were segmented into smaller portions to facilitate prompt feedback, potentially enabling real-time feedback. This segmentation process involved two critical steps: (1) aligning the datasets from the mannequin and the smartwatch to ensure their data corresponded accurately for analysis, detailed in subsection “Data Matching (Process 1)”; and (2) executing the chunking algorithm, detailed in the subsection “Data Chunking (Process 2).”

While real time user feedback was beyond this project’s scope, the data were examined to assess future applicability for this purpose. A 5-second analysis window was selected for potential feedback, where high-quality CPR—defined by approximately 10 compressions within this period—was used as the feedback standard. After receiving feedback, users would need time to comprehend and adjust their technique before the next feedback prompt.

Data Matching (Process 1)

The 2 data sources have different structures: the mannequin (XML file) records a sequence of compression events and their properties, while the smartwatch logs (CSV file) contain time-series accelerometer data. Because of this difference, synchronization cannot be based on time stamps. Instead, the datasets are aligned based on compression events, as the focus is on analyzing compression depth and frequency. The first step is to identify when compressions occurred using the accelerometer data, as explained further in this study. Once these events are detected, they can be matched with the corresponding compression depth and other properties from the XML file.

A chest compression pattern typically indicates a peak in acceleration when the CPR performer compresses the chest, as shown in Figure 3. Due to sensor noise and the performer’s hand movement vibrations, various peaks can appear in the data; however, only a few of these should be considered valid compressions. The XML file from the mannequin specifies the number of compressions performed, which serves as the gold standard for creating an algorithm to identify compression peaks in the CSV file.

After analyzing the accelerometer data from the compressions, 3 variables were identified to help determine if a peak is a valid compression peak. First, the acceleration difference: the change in acceleration at the compression peak should be significant.

XSL•FO
RenderX

Second, the window size is the time frame within which a peak can occur; for example, 2 compressions cannot happen within 5 microseconds. Third, the minimum acceleration peak helps filter out recoil peaks or peaks caused by vibrations or sensor noise. These values may vary from one compression to another, even if performed by the same person, due to factors such as user fatigue, incomplete recoil, changes in compression rate, or varying compression force.

To handle variability, algorithm 1 (Figure 4) was designed to match the "peaksFoundInCSV" with the peaks recorded in the XML file. The authors used SciPy's *find_peaks* function to detect peaks in the CSV file based on 3 key variables—acceleration difference, window size, and minimum acceleration peak. Since these values varied across files, a script iterated through a range of values—acceleration difference (0.01-2), window size (1-40), and minimum acceleration peak (–1 to 2)—with step sizes of 0.2, 3.9, and 0.3, respectively, until the detected peaks matched those in the XML file. To verify accuracy, random manual checks were performed by generating graphs, similar to Figure 3, to ensure the peaks were correctly identified and aligned.

Figure 3. Process 1: The diagram illustrates a representative output showing the mapping of compressions in the XML data (mannequin data) with the compressions detected in the accelerometer data (smartwatch data).

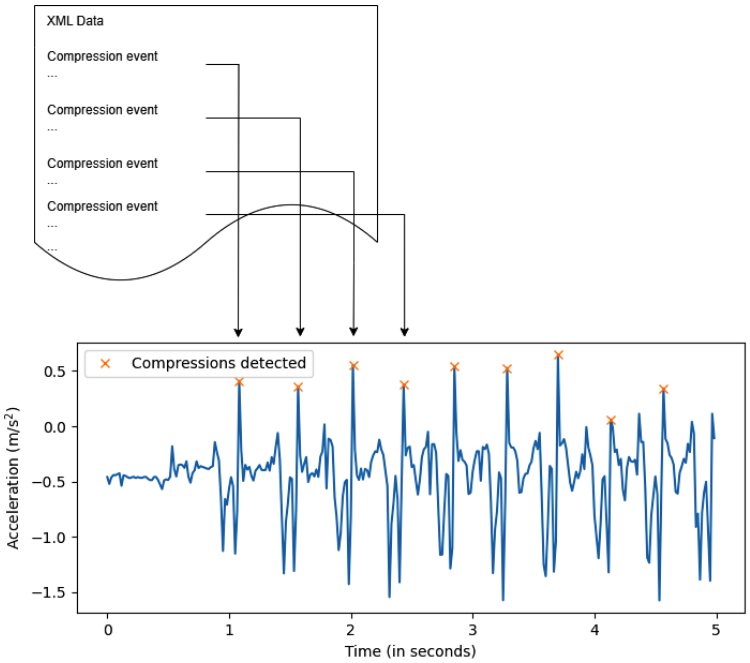


Figure 4. Pseudocode to generate matches between compressions in mannequin and accelerometer data files (algorithm 1).

Input: AccelerometerFile (C); MannequinFile (X); AccelerationDifferenceRange (AD), MinimumAccelerationPeakRange (MAD), WindowSizeRange (WS)

Output: Dictionary (CompressionMatching) mapping compression peaks in accelerometer data to corresponding Records in mannequin data

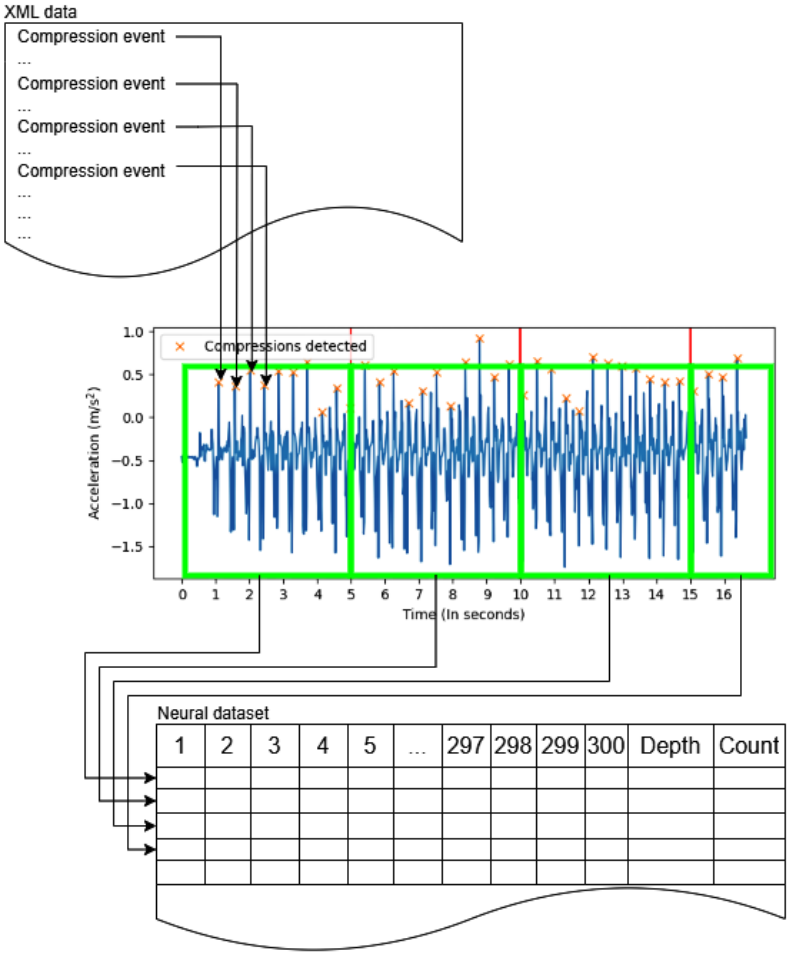
```
PeaksInMannequin = findCompressionsCountInMannequin(X)        ▷ Collecting Mannequin data
for ad ∈ AD do                                                ▷ Looping through range of AD values
    for mad ∈ MAD do                                          ▷ Looping through range of MAD values
        for ws ∈ WS do                                        ▷ Looping through range of WS values
            peaksFoundInCSV = findPeaks(C, ad, mad, ws)       ▷ Finding peaks in CSV data using findpeak function
            if peaksFoundInCSV = PeaksInMannequin then        ▷ If compression match, then return data
                CompressionMatching = getMatchingData(X, C, peaksFoundInCSV )
                return CompressionMatching
            else
                updateRangeVariables(AD, MAD, WS)             ▷ update the variables to find peaks again
            end if
        end for
    end for
end for
```


XSL•FO
RenderX

Data Chunking (Process 2)

A 5-second interval was chosen as the optimal chunking duration, with the rationale for this selection explained in the previous section. Subsequently, the aligned datasets underwent segmentation into chunks for detailed analysis. Figure 5 illustrates the segmentation process applied to a representative dataset. Following the alignment of accelerometer data with the mannequin's compression peaks, the accelerometer readings were divided into 300-point segments (indicated by a green box). Each segmented block (denoted by a green box) was then documented as a single record within the dataset used for training the proposed model, formatted in CSV. These records encompassed acceleration data for a 5-second interval (equivalent to 300 data points at a 60 Hz sampling rate), alongside the average compression depth and the total number of compressions recorded within that timeframe. For instances where the dataset was less than 5 seconds in length, the acceleration data was padded with zeros at the end to standardize it to a 5-second duration. The zero padding was performed on 124 records. The datasets prepared through this method were subsequently used to train the neural network model.

Figure 5. Process 2: Illustration of combining accelerometer and mannequin data to create the dataset for training the neural network model.



Model to Predict CPR Performance

The aim of this study was to evaluate CPR performance by analyzing accelerometer data collected from smartwatches, which inherently included various forms of sensor noise, such as vibrations due to hand movements and accelerations not associated with chest compressions. The direct elimination of this noise was deemed impractical; therefore, the decision was made to use a supervised learning approach, specifically a neural network model, to tackle this intricate challenge.

During model testing, the training data was randomly selected from 5-second chunks (2701 records), which could result in partial data from a participant appearing in both the training and test sets. However, since these chunks were not directly comparable, as determined during the review, the data were not split by participants. Although neural networks typically assume that data are independent and identically distributed, this study's approach does not violate this assumption. Figure 6 illustrates the CPR performance of 3 participants over 100 seconds, summarized in 5-second intervals. Each chunk is treated as an independent data point, with no relationship to the previous or next chunk. The objective of this research is to randomly select a 5-second chunk and predict compression depth and count based on the accelerometer data, independent of past or future data. Since the model is trained to make predictions on isolated chunks rather than sequential patterns, the independent and identically distributed assumption holds, and data splitting by participants is not necessary. Figure 6 shows that compression depth declined over time and compression counts varied significantly. However, since the model does not rely on temporal dependencies, each chunk remains independent,

XSL•FO
RenderX

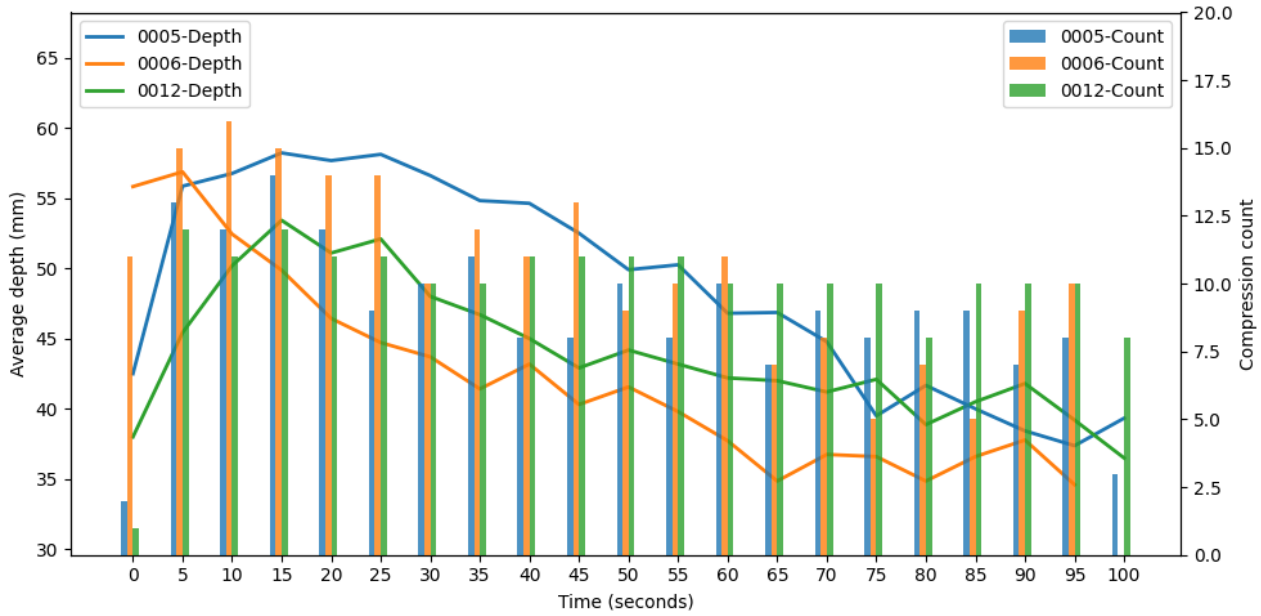
justifying the decision not to split data by participants for training and testing.

The dataset was randomly split into 80% (2161/2701) training and 20% (540/2701) validation test sets. Each record consisted of 300 accelerometer data points, corresponding to a duration of 5 seconds at a 60 Hz sampling rate. The objective of the model is to predict both the number of compressions and their average depth during this interval.

The development of the model used the TensorFlow library within the Python programming environment [52]. To optimize the model's performance, a combination of grid search and Hyperband techniques was used. In each experiment, grid search was used to explore a broad range of hyperparameters, identifying configurations that performed well or poorly. In addition, the Hyperband technique was applied to refine the search by excluding poorly performing hyperparameter sets, thereby narrowing the search space. This process was repeated iteratively, with grid search expanding the search range in each new experiment, followed by Hyperband reducing it to eliminate ineffective configurations. The key hyperparameters adjusted included the number of hidden layers, layer sizes, number of epochs, batch sizes, and dropout rates. The objective was to determine the most effective combination of hyperparameters that would allow the model to accurately interpret accelerometer data, despite sensor noise, and reliably assess CPR quality based on smartwatch data.

To evaluate the normality of the data and assess the correlation between actual and predicted values, 2 statistical tests were conducted. First, the Shapiro-Wilk test was performed on the expected depth and expected compression count to determine if the data followed a normal distribution. The results showed P values of 2.69×10^{-3} and 2.80×10^{-4}, respectively, indicating that both datasets deviated significantly from normality. Given this, Spearman rank correlation was chosen to measure the relationship between expected versus predicted depth and compression counts since it does not assume normality.

Figure 6. Variation in compression depth and compression rate among three participants over 100 seconds of CPR performance.



Ethical Considerations

The research ethics evaluation for this study was reviewed and approved by the TBRHSC Research Ethics Board (2022519), with all procedures conducted in accordance with the ethical standards of the responsible institutional committee and the principles outlined in the World Medical Association Declaration of Helsinki. Written informed consent was obtained from all participants before their inclusion in the study. To maintain privacy and confidentiality, all data collected were deidentified at the point of collection, with each participant assigned a unique identifier; a separate, secure document mapping participant identities to these identifiers was maintained independently and was not linked to the research data. Participants were compensated with a US $5 gift card for their participation.

Results

Table 1 contains the demographic details of participants involved in the data collection process. A total of 83 individuals contributed, resulting in 124 mannequin-smartwatch datasets, with 41 participants completing 2 rounds of CPR. The sample included 56 females and 27 males, spanning various age groups, with the largest group being younger than 30 years old. Participants also varied in their previous CPR experience, ranging from no previous attempts to over 30, offering a diverse dataset in terms of skill level and background.

XSL•FO
RenderX

Table 1. Demographic details of participants in the data collection process.

Description	Number of participants, n
Number of mannequin-smartwatch datasets	124
Number of participants	83
Participants performing 2 rounds of CPR[a]	41
Sex	
Female	56
Male	27
Age groups (y)	
<30	42
30-39	27
40-49	8
>50	6
Previous CPR attempts	
0	12
1-9	23
10-19	10
20-29	14
30+	24

[a]CPR: cardiopulmonary resuscitation.

A series of experiments were systematically conducted, exploring various hyperparameters and datasets, until no notable improvements in outcomes were observed. This investigative process entailed 12 different experiments to find the optimal set of hyperparameters. Within each trial, data on the variations in hyperparameters were documented, noting the discrepancies between anticipated and actual compression depths and counts. Given that the variance between expected and observed results could manifest in negative figures, potentially skewing average calculations, absolute values were used for a more accurate determination of the average and median. These analytical steps informed adjustments in hyperparameters for ensuing experiments aimed at enhancing performance.

Throughout this extensive testing phase, a total of 1226 iterations were conducted, each exploring different hyperparameter settings. The specific settings and the best outcomes from each experiment are presented in Table 2, while detailed information on each experiment is provided in Multimedia Appendix 2. Summarizing the testing, the most effective model across all iterations was characterized by the following hyperparameters: hidden layers configured as (2000, 1000, 250, 100, 50), a batch size of 2000 epochs, and a dropout rate of 10% after every 2 layers. Predictions from this model demonstrated an average absolute loss in compression depth of 3.8 (SD 0.16) mm and in compression count of 0.8 (SD 0.05) counts, with a median absolute loss for compression depth at 2.59 (IQR 0.24) mm and for compression count at 0.03 counts. Throughout the course of the experiments, 27 models were identified as capable of predicting an average absolute compression depth below 4.0 mm.

The Spearman correlation coefficients were 0.80 and 0.78, respectively, with *P* values of 1e-3, indicating a strong and statistically significant positive correlation between actual and predicted values for both compression depth and compression count. These findings suggest that the model effectively predicts compression depth and count, even though the underlying data distribution is not normal.

XSL•FO
RenderX

Table 2. Summarizing hyperparameters for each experiment and the best results achieved.

Experiment	Epochs	Batch size	Layer sizes	Number of layers	Connection dropout	Best result (compression depth, count)
1	10	3, 5, 7, 9, 11	10, 35, 60, 85, 100, 400, 700, 1000, 1200	3	10% after every 2 layers	±4.8 mm, 1.2 counts
2	100, 500	6, 12, 18	10, 35, 60, 85, 100, 400, 700, 1000, 1200	3, 4	10% after every 2 layers	±4.0 mm, 0.9 counts
3	500, 750	18, 24	5, 40, 100, 500, 900, 1300	4, 6	10% after every 2 layers	±4.8 mm, 1.2 counts
4	500, 750	18, 24	5, 40, 100, 500, 900, 1300	4, 6	10% after every 2 layers	±3.9 mm, 0.9 counts
5	500, 750	18, 24	5, 10, 60, 100, 400, 700, 1000	4	10% after every 2 layers	±3.9 mm, 0.8 counts
6	500, 750	6, 24	5, 10, 50, 100, 500, 1000, 4000	4, 5	10% after every 2 layers	±6.2 mm, 1.2 counts
7	1000	3	5, 10, 50, 100, 500, 1000, 4000	5	10% after every 2 layers	±4.0 mm, 0.9 counts
8	1000	12	5, 10, 50, 100, 500, 1000, 4000	5	No dropouts	±3.9 mm, 0.9 counts
9	1000	3, 12	5, 10, 50, 100, 500, 1000, 4000	3	No dropouts	±4.1 mm, 0.9 counts
10	100, 500, 1000	3, 9	5, 10, 25, 50, 75, 100, 250, 500, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, 3500, 3750, 4000	3	10% after every 2 layers	±4.1 mm, 0.9 counts
11	100, 1000	3, 128	5, 10, 25, 50, 75, 100, 250, 500, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, 3500, 3750, 4000	5	10% after every 2 layers	±3.9 mm, 0.8 counts
12	1000, 2000	256, 1024	5, 10, 25, 50, 75, 100, 250, 500, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, 3500, 3750, 4000	5	10% after every 2 layers	±3.8 mm, 0.8 counts

Discussion

Overview

This study introduces a neural network model capable of using accelerometer data from a smartwatch to guide users in delivering high-quality CPR. To facilitate this, apps for both Apple Watch and iPhone were developed to gather accelerometer data during CPR. This research compiled data from 83 participants using Apple Watches to perform CPR on a mannequin, resulting in a total of 27,844 chest compressions. A total of 1226 experiments were conducted to evaluate different hyperparameter settings. The most effective model used hidden layers (2000,1000,250,100,50), a batch size of 2000, and a 10% dropout after every 2 layers. This model achieved an average absolute loss of 3.8 mm in compression depth and 0.8 in compression count, with median losses of 2.59 mm and 0.03 counts, respectively. In total, 27 models achieved an average absolute compression depth loss of less than 4.0 mm.

Principal Results

Considering mannequin data as the benchmark, the neural network was trained with the accelerometer data to emulate the mannequin's feedback. After undergoing 1226 iterations of model refinement, the optimal model emerged, capable of estimating compression depth with an average discrepancy of 3.8 mm and compression count with an average discrepancy of 0.8 counts across 5-second intervals. This investigation underscores the potential of a neural network model to surpass existing models in predicting CPR quality, underpinned by a more extensive and realistic dataset than those previously reported in the literature [40,44,45,53-55].

Comparison With Previous Work

Numerous studies have highlighted significant enhancements in CPR quality when performers receive real-time feedback through intelligent technologies [40,53,56]. Gruenerbl et al [44] proposed a smartwatch app capable of measuring CPR parameters and offering visual feedback. This app analyzes accelerometer data to evaluate compression quality, providing user feedback to achieve high-quality compressions. Lu et al [45] also proposed a smartwatch app alongside an algorithm for evaluating compression metrics. However, these proposed solutions are developed on limited datasets and do not provide sufficient details to reproduce or enhance the system. Dedicated devices, such as the Laerdal CPRMeter2, which can be placed on the chest of a mannequin or patient, are available for purchase. These devices are equipped with accelerometers and pressure sensors to accurately assess CPR performance and provide real-time feedback [26]. However, these devices are expensive and must be carried by the user at all times, as emergencies can occur at any moment.

Limitations

This study demonstrates that the accelerometer data from the smartwatch can be used to provide real-time feedback to the CPR performer and can help increase the survival chances of the patient. However, there are a few limitations worth noting. First, the developed model focuses on measuring compression

XSL•FO
RenderX

depth and count but does not account for compression recoil. Compression recoil plays an important role in performing high-quality CPR, as it allows the patient's chest to fully expand. This full expansion enables more air to enter the patient's body. Compression recoil can also be measured using acceleration data from the smartwatch and compared with sensor data from the mannequin, similar to how compression events are measured. In future research, the authors plan to include this factor as well.

Second, the Apple Watch Series 7 was the exclusive device used to collect data during this study. However, sensors in smartwatches vary between brands, models, and versions, each having its own set of white noise that may impact the results. Therefore, it is necessary to investigate how other devices with different sensors perform with the presented model and whether the model needs to be tuned for each device type. In future studies, the authors plan to use an Android device to cover a broader spectrum of sensors.

Third, during data collection, it was observed that the sensor data frequency fluctuated between 48-60 Hz, although the collection rate was set to 60 Hz. This fluctuation could be due to several reasons: slower processing on the device, the device's inability to write data at that frequency to the file, or the device's processing speed being unable to read the data from the sensors at this speed. Therefore, it is important to investigate how the model's performance would be affected if the data frequency were lowered.

Fourth, while working with the Apple Watch, the authors found a limitation that restricts its use during emergency events. The Apple Watch only allows an app to run in the background for 30 seconds before it is suspended, after which the app can only execute its tasks intermittently. This restriction hinders the use of the smartwatch for providing real-time feedback during emergencies.

Finally, in data segmentation, the acceleration data were chunked based on a fixed number of data points (300, representing 5 seconds of data at 60 Hz). This chunking may occur during a compression event, potentially splitting the compression event into 2 parts, which the algorithm may not recognize, resulting in missed compressions. In future research, the authors plan to address this problem by introducing an advanced algorithm that can recognize the completion of a compression event and then segment the data accordingly to avoid splitting compression data.

Conclusions

This study presents a neural network model designed to gauge the rate and depth of CPR compressions accurately. The model has undergone training on a large dataset, ensuring a robust foundation for its predictive capabilities. Its accuracy is remarkably close to the measurements obtained from a CPR mannequin, serving as a benchmark for comparison. Notably, this model demonstrates superior performance when compared with other models discussed within the existing literature, establishing it as a significant advancement in the field of medical emergency training and response analysis. Implementation of this model in real-world scenarios could significantly improve SCA survival rates.

Acknowledgments

We would like to thank Leanne Baird and the Ibn Sina Simulation Lab at the Thunder Bay Regional Health Sciences Centre for providing technical support and access to the simulation mannequin. The authors used the ChatGPT-4 language model (developed by OpenAI) to correct grammar and typographical errors in the manuscript. All edits were reviewed and approved by the authors to ensure accuracy and integrity.

This research was funded by the MITACS grant and received by Gaurav Rao and Avive Solutions Inc. Other research costs are covered by the NSERC Discovery Grant (RGPIN-2017-05377), held by VM.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request. Access to the data may be granted for research purposes, subject to ethical considerations and approval.

Authors' Contributions

Conceptualization and resources were contributed by GR, DWS, VM, and PL. Methodology was contributed by GR, DWS, and VM. Software, visualization, and investigation were handled by GR. Validation and formal analysis were contributed by GR and VM. Data curation was contributed by GR, GE, and NK. Writing—original draft was contributed by GR. Writing—review and editing was contributed by GR, DWS, VM, and PL. Supervision and project administration were handled by VM and PL. Funding acquisition was handled by GR, GWS, VM, and PL.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Informed consent form.
[PDF File (Adobe PDF File), 543 KB - mhealth_v13i1e57469_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Experiment details.
[DOCX File , 16 KB - mhealth_v13i1e57469_app2.docx]

References

1. Hosey RG, Armsey TD. Sudden cardiac death. Clin Sports Med 2003;22(1):51-66. [doi: 10.1016/s0278-5919(02)00071-6] [Medline: 12613086]
2. Medlineplus.gov . Sudden Cardiac Arrest ? Sudden Cardiac Death ? MedlinePlus. 2022. URL: https://medlineplus.gov/suddencardiacarrest.html, [accessed 2022-09-16]
3. Heart.org F. 2025 Heart Disease and Stroke Statistics Update Fact Sheet. 2025. URL: https://www.heart.org/en/-/media/PHD-Files-2/Science-News/2/2025-Heart-and-Stroke-Stat-Update/2025-Statistics-At-A-Glance.pdf [accessed 2025-05-01]
4. Research CIoH. Research program intended to help Canadians survive sudden cardiacarrest expands to all 10 Canadian provinces - CIHR. 2019. URL: https://cihr-irsc.gc.ca/ [accessed 2022-09-16]
5. Heart.org C. What is CPR. 2022. URL: https://cpr.heart.org/en/resources/ [accessed 2022-09-16]
6. Heart.org. (n.d.). Emergency Treatment of Cardiac Arrest | American Heart Association. Heart.Org. 2021. URL: https://www.heart.org/en/health-topics/cardiac-arrest/emergency-treatment-of-cardiac-arrest [accessed 2021-08-24]
7. Rescue. RescueOne. (2013). Every Second Counts. 2013. URL: https://rescue-one.com/wp-content/uploads/aha-ruralcommunityaccess.pdf [accessed 2025-05-01]
8. Valenzuela TD, Roe DJ, Cretin S, Spaite DW, Larsen MP. Estimating effectiveness of cardiac arrest interventions: a logistic regression survival model. Circulation 1997;96(10):3308-3313. [doi: 10.1161/01.cir.96.10.3308] [Medline: 9396421]
9. Weisfeldt ML, Sitlani CM, Ornato JP, Rea T, Aufderheide TP, Davis D, et al. Survival after application of automatic external defibrillators before arrival of the emergency medical system: evaluation in the resuscitation outcomes consortium population of 21 million. J Am Coll Cardiol 2010;55(16):1713-1720 [FREE Full text] [doi: 10.1016/j.jacc.2009.11.077] [Medline: 20394876]
10. Brooks SC, Simmons G, Worthington H, Bobrow BJ, Morrison LJ. The PulsePoint Respond mobile device application to crowdsource basic life support for patients with out-of-hospital cardiac arrest: challenges for optimal implementation. Resuscitation 2016;98:20-26. [doi: 10.1016/j.resuscitation.2015.09.392] [Medline: 26475397]
11. Van de Voorde P, Gautama S, Momont A, Ionescu CM, De Paepe P, Fraeyman N. The drone ambulance [A-UAS]: golden bullet or just a blank? Resuscitation 2017;116:46-48. [doi: 10.1016/j.resuscitation.2017.04.037] [Medline: 28487253]
12. Chien CY, Tsai SL, Tsai LH, Chen CB, Seak CJ, Weng YM, et al. Impact of transport time and cardiac arrest centers on the neurological outcome after out‐of‐hospital cardiac arrest: a retrospective cohort study. JAHA 2020;9(11). [doi: 10.1161/jaha.119.015544]
13. Ko SY, Ro YS, Shin SD, Song KJ, Hong KJ, Kong SY. Effect of a first responder on survival outcomes after out-of-hospital cardiac arrest occurs during a period of exercise in a public place. PLoS One 2018;13(2):e0193361 [FREE Full text] [doi: 10.1371/journal.pone.0193361] [Medline: 29489877]
14. Andelius L, Hansen C, Lippert F, Karlsson L, Torp-Pedersen C, Gislason GH, et al. 40 Long ambulance response time is associated with higher incidence of cardiopulmonary resuscitation and defibrillation by dispatched citizen first-responders. BMJ Open 2019;9(Suppl 2):A15. [doi: 10.1136/bmjopen-2019-ems.40]
15. Rea T, Blackwood J, Damon S, Phelps R, Eisenberg M. A link between emergency dispatch and public access AEDs: potential implications for early defibrillation. Resuscitation 2011;82(8):995-998. [doi: 10.1016/j.resuscitation.2011.04.011] [Medline: 21570169]
16. Sanko S, Kashani S, Lane C, Eckstein M. Implementation of the los angeles tiered dispatch system is associated with an increase in telecommunicator-assisted CPR. Resuscitation 2020;155:74-81. [doi: 10.1016/j.resuscitation.2020.06.039] [Medline: 32721415]
17. Andelius L, Hansen CM, Lippert F, Karlsson L, Torp-Pedersen C, Gislason G, et al. Will shorter distance between dispatched lay-rescuer and out-of-hospital cardiac arrest increase cardiopulmonary resuscitation and early defibrillation rates? Resuscitation 2018;130:e56-e57. [doi: 10.1016/j.resuscitation.2018.07.104]
18. Pijls RW, Nelemans PJ, Rahel BM, Gorgels AP. A text message alert system for trained volunteers improves out-of-hospital cardiac arrest survival. Resuscitation 2016;105:182-187 [FREE Full text] [doi: 10.1016/j.resuscitation.2016.06.006] [Medline: 27327230]
19. Heart.org G. New resuscitation guidelines update CPR chest pushes. 2015. URL: https://www.heart.org/en/news/2018/07/23/new-resuscitation-guidelines-update-cpr-chest-pushes, [accessed 2025-05-01]
20. Perkins GD, Gräsner JT, Semeraro F, Olasveengen T, Soar J, Lott C, Autoren der Leitlinien des European Resuscitation Council. [Executive summary]. Notf Rett Med 2021;24(4):274-345 [FREE Full text] [doi: 10.1007/s10049-021-00883-z] [Medline: 34093077]
21. Zègre-Hemsey JK, Grewe ME, Johnson AM, Arnold E, Cunningham CJ, Bogle BM, et al. Delivery of automated external defibrillators via drones in simulated cardiac arrest: users' experiences and the human-drone interaction. Resuscitation 2020;157:83-88 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.10.006] [Medline: 33080371]

XSL•FO
RenderX

22. Zègre-Hemsey JK, Bogle B, Cunningham CJ, Snyder K, Rosamond W. Delivery of automated external defibrillators (AED) by drones: implications for emergency cardiac care. Curr Cardiovasc Risk Rep 2018;12(11) [FREE Full text] [doi: 10.1007/s12170-018-0589-2] [Medline: 30443281]
23. Tanaka S, Tsukigase K, Hara T, Sagisaka R, Myklebust H, Birkenes TS, et al. Effect of real-time visual feedback device 'Quality cardiopulmonary resuscitation (QCPR) classroom' with a metronome sound on layperson CPR training in Japan: a cluster randomized control trial. BMJ Open 2019;9(6):e026140 [FREE Full text] [doi: 10.1136/bmjopen-2018-026140] [Medline: 31189674]
24. Baldi E, Cornara S, Contri E, Epis F, Fina D, Zelaschi B, et al. Real-time visual feedback during training improves laypersons' CPR quality: a randomized controlled manikin study. CJEM 2017;19(6):480-487. [doi: 10.1017/cem.2016.410] [Medline: 28115027]
25. Wang SA, Su CP, Fan HY, Hou WH, Chen YC. Effects of real-time feedback on cardiopulmonary resuscitation quality on outcomes in adult patients with cardiac arrest: a systematic review and meta-analysis. Resuscitation 2020;155:82-90. [doi: 10.1016/j.resuscitation.2020.07.024] [Medline: 32755666]
26. Laerdal.com . Laerdal Medical. 2023. URL: https://laerdal.com/ [accessed 2025-04-05]
27. Rao G, Savage DW, Mago V, Lingras P. A Survey on Technologies Used During out of Hospital Cardiac Arrest. 2023 Presented at: 16th International Conference on Health Informatics; 2023 February 16 - 18; Lisbon, Portugal. [doi: 10.5220/0011749800003414]
28. Hardeland C, Skåre C, Kramer-Johansen J, Birkenes TS, Myklebust H, Hansen AE, et al. Targeted simulation and education to improve cardiac arrest recognition and telephone assisted CPR in an emergency medical communication centre. Resuscitation 2017;114:21-26. [doi: 10.1016/j.resuscitation.2017.02.013] [Medline: 28236428]
29. Sanko S, Feng S, Lane C, Eckstein M. Comparison of emergency medical dispatch systems for performance of telecommunicator-assisted cardiopulmonary Resuscitation among 9-1-1 callers with limited English proficiency. JAMA Netw Open 2021;4(6):e216827 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.6827] [Medline: 34076700]
30. Hambly H, Rajabiun R. Rural broadband: gaps, maps and challenges. Telematics and Informatics 2021;60:101565. [doi: 10.1016/j.tele.2021.101565]
31. Lee HS, You K, Jeon JP, Kim C, Kim S. The effect of video-instructed versus audio-instructed dispatcher-assisted cardiopulmonary resuscitation on patient outcomes following out of hospital cardiac arrest in Seoul. Sci Rep 2021;11(1):15555 [FREE Full text] [doi: 10.1038/s41598-021-95077-5] [Medline: 34330969]
32. Riou M, Ball S, O'Halloran KL, Whiteside A, Williams TA, Finn J. Hijacking the dispatch protocol: when callers pre-empt their reason-for-the-call in emergency calls about cardiac arrest. Discourse Studies 2018;20(5):666-687. [doi: 10.1177/1461445618754435]
33. Case R, Cartledge S, Siedenburg J, Smith K, Straney L, Barger B, et al. Identifying barriers to the provision of bystander cardiopulmonary resuscitation (CPR) in high-risk regions: a qualitative review of emergency calls. Resuscitation 2018;129:43-47. [doi: 10.1016/j.resuscitation.2018.06.001] [Medline: 29864455]
34. Meinich-Bache O, Engan K, Birkenes TS, Myklebust H. Real-time chest compression quality measurements by smartphone camera. J Healthc Eng 2018;2018:6241856 [FREE Full text] [doi: 10.1155/2018/6241856] [Medline: 30581549]
35. Ali S, Athar M, Ahmed S. A randomised controlled comparison of video versus instructor-based compression only life support training. Indian J Anaesth 2019;63(3):188. [doi: 10.4103/ija.ija_737_18]
36. Jeon SA, Chang H, Yoon SY, Hwang N, Kim K, Yoon H, et al. Effectiveness of smartwatch guidance for high-quality infant cardiopulmonary resuscitation: A simulation study. Medicina (Kaunas) 2021;57(3):193 [FREE Full text] [doi: 10.3390/medicina57030193] [Medline: 33668789]
37. Sevil H, Bastan V, Gültürk E, El Majzoub I, Göksu E. Effect of smartphone applications on cardiopulmonary resuscitation quality metrics in a mannequin study: a randomized trial. Turk J Emerg Med 2021;21(2):56-61 [FREE Full text] [doi: 10.4103/2452-2473.313333] [Medline: 33969240]
38. An M, Kim Y, Cho WK. Effect of smart devices on the quality of CPR training: a systematic review. Resuscitation 2019;144:145-156 [FREE Full text] [doi: 10.1016/j.resuscitation.2019.07.011] [Medline: 31325556]
39. Song Y, Chee Y, Oh J, Ahn C, Lim TH. Smartwatches as chest compression feedback devices: a feasibility study. Resuscitation 2016;103:20-23. [doi: 10.1016/j.resuscitation.2016.03.014] [Medline: 27004719]
40. Semeraro F, Taggi F, Tammaro G, Imbriaco G, Marchetti L, Cerchiari EL. iCPR: a new application of high-quality cardiopulmonary resuscitation training. Resuscitation 2011;82(4):436-441. [doi: 10.1016/j.resuscitation.2010.11.023] [Medline: 21227560]
41. Chan T, Wan K, Chan J, Lam H, Wong Y, Kan P. New era of CPR: application of I-Technology in resuscitation. Hong Kong Journal of Emergency Medicine 2017;19(5):305-311. [doi: 10.1177/102490791201900502]
42. Park CS, Kang IG, Heo SJ, Chae Y, Kim H, Park S, et al. A Randomised, cross over study using a mannequin model to evaluate the effects on CPR quality of real-time audio-visual feedback provided by a smartphone application. Hong Kong Journal of Emergency Medicine 2017;21(3):153-160. [doi: 10.1177/102490791402100304]
43. King CE, Sarrafzadeh M. A survey of smartwatches in remote health monitoring. J Healthc Inform Res 2018;2(1-2):1-24 [FREE Full text] [doi: 10.1007/s41666-017-0012-7] [Medline: 30035250]

XSL•FO
RenderX

44. Gruenerbl A, Pirkl G, Monger E, Gobbi M, Lukowicz P. Smart-Watch life savermartWatch interactive-feedback system for improving bystander CPR. 2015 Presented at: ISWC '15: Proceedings of the 2015 ACM International Symposium on Wearable Computers; 2015 September 7 - 11; Osaka Japan p. 19. [doi: 10.1145/2802083.2802086]
45. Lu TC, Chen Y, Ho TW, Chang Y, Lee Y, Wang Y, et al. A novel depth estimation algorithm of chest compression for feedback of high-quality cardiopulmonary resuscitation based on a smartwatch. J Biomed Inform 2018;87:60-65 [FREE Full text] [doi: 10.1016/j.jbi.2018.09.014] [Medline: 30268843]
46. Asada HH, Jiang HH, Gibbs P. Active noise cancellation using MEMS accelerometersfor motion-tolerant wearable bio-sensors. 2004 Presented at: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2004 September 01-05; San Francisco, CA, USA. [doi: 10.1109/iembs.2004.1403631]
47. Apple . App Store - Apple (CA). 2023. URL: https://www.apple.com/ [accessed 2025-04-05]
48. Statista.com . Statista Mobile OS share. 2023. URL: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/ [accessed 2025-04-05]
49. Statista.com . Statista Watch OS share. 2023. URL: https://www.statista.com/statistics/910862/worldwide-smartwatch-shipment-market-share/ [accessed 2025-04-05]
50. Laerdal.com . Laerdal Medical AnneQCPR. 2023. URL: https://laerdal.com/ca/products/simulation-training/resuscitation-training/little-anne-qcpr/ [accessed 2025-04-05]
51. Amazon Web Services. G5 instance types. 2023. URL: https://aws.amazon.com/ec2/instance-types/g5/ [accessed 2025-04-05]
52. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale machine learning on heterogeneous systems. arXiv:1603.04467 2015 [FREE Full text]
53. Song Y, Oh J, Chee Y. A new chest compression depth feedback algorithm for high-quality CPR based on smartphone. Telemed J E Health 2015;21(1):36-41 [FREE Full text] [doi: 10.1089/tmj.2014.0051] [Medline: 25402865]
54. Ahn C, Lee J, Oh J, Song Y, Chee Y, Lim TH, et al. Effectiveness of feedback with a smartwatch for high-quality chest compressions during adult cardiac arrest: a randomized controlled simulation study. PLoS One 2017;12(4):e0169046 [FREE Full text] [doi: 10.1371/journal.pone.0169046] [Medline: 28369055]
55. Lu TC, Chang YT, Ho TW, Chen Y, Lee YT, Wang YS, et al. Using a smartwatch with real-time feedback improves the delivery of high-quality cardiopulmonary resuscitation by healthcare professionals. Resuscitation 2019;140:16-22. [doi: 10.1016/j.resuscitation.2019.04.050] [Medline: 31078650]
56. Truszewski Z, Szarpak L, Kurowski A, Evrin T, Zasko P, Bogdanski L, et al. Randomized trial of the chest compressions effectiveness comparing 3 feedback CPR devices and standard basic life support by nurses. Am J Emerg Med 2016;34(3):381-385. [doi: 10.1016/j.ajem.2015.11.003] [Medline: 26612703]

Abbreviations

AED: automated external defibrillator
AHA: American Heart Association
CPR: cardiopulmonary resuscitation
CSV: comma-separated value
EMS: emergency medical services
SCA: sudden cardiac arrest
TBRHSC: Thunder Bay Regional Health Sciences Centre

Edited by T de Azevedo Cardoso, L Buis; submitted 17.02.24; peer-reviewed by T Tangpaisarn, Q Dong, W Kern; comments to author 15.05.24; revised version received 03.07.24; accepted 31.03.25; published 05.05.25.

Please cite as:
Rao G, Savage DW, Erickson G, Kyryluk N, Lingras P, Mago V
Enhancing Cardiopulmonary Resuscitation Quality Using a Smartwatch: Neural Network Approach for Algorithm Development and Validation
JMIR Mhealth Uhealth 2025;13:e57469
URL: https://mhealth.jmir.org/2025/1/e57469
doi:10.2196/57469
PMID:

©Gaurav Rao, David W Savage, Gabrielle Erickson, Nathan Kyryluk, Pawan Lingras, Vijay Mago. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited.

XSL•FO
RenderX

The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Relationship Between Activity Tracker Metrics and the Physical Activity Index and Their Association With Cardiometabolic Phenotypes, Subclinical Atherosclerosis, and Cardiac Remodeling: Cross-Sectional Study

Weiting Huang[1], MBBS; Mark Kei Fong Wong[2], PhD; Enver De Wei Loh[2], MBBS; Tracy Koh[1], MD; Alex Weixian Tan[1], MBBS; Xiayan Shen[1], MBBS; Onur Varli[1], MD; Siew Ching Kong[1], MSc; Calvin Woon Loong Chin[1], PhD; Swee Yaw Tan[1], MBBS; Jonathan Jiunn Liang Yap[1], MBBS; Eddie Yin Kwee Ng[2], PhD; Khung Keong Yeo[1], MBBS

[1]National Heart Centre Singapore, Singapore, Singapore

[2]Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Weiting Huang, MBBS
National Heart Centre Singapore
5 Hospital Drive
Singapore, 169609
Singapore
Phone: 65 96370396
Email: huang.weiting@singhealth.com.sg

Abstract

Background: Consumer wearable technology quantifies physical activity; however, the association between these metrics and cardiometabolic health requires further elucidation.

Objective: This study identified latent factors derived from Fitbit heart rate metrics and their relationship with cross-sectional cardiovascular phenotypes.

Methods: This cross-sectional analysis included 457 participants from the SingHEART study, a multiethnic, population-based study of Asian individuals aged 21 to 69 years recruited in Singapore. Participants wore the Fitbit Charge HR for 7 days, and data on physical activity metrics, self-reported physical activity index (PAI), blood tests, coronary artery calcium scores, and cardiac magnetic resonance imaging were collected. Exploratory factor analysis identified latent factors from Fitbit metrics, and multivariate regression analysis assessed associations with blood and cardiovascular imaging phenotypes.

Results: Higher levels of self-reported PAI were significantly associated with a higher number of calories burned (P=.008), number of steps and floors climbed, distance, number of activity calories, and number of very active minutes (P<.001). However, there was no association between PAI and other Fitbit metrics. Using exploratory factor analysis, we identified three latent factors measured by Fitbit metrics: (1) elevated metabolic equivalents of task (METs; calories burned per day, minutes per day spent fairly active in 3-6 METs and very active in ≥6 METs, and activity calories), (2) total activity (steps per day, distance in kilometers per day, and number of floors per day), and (3) others, all with a Cronbach α of >0.7. Higher total activity was associated with increased high-density lipoprotein levels (β=0.06; P<.001), decreased triglyceride levels (β=−0.10; P=.006), and lower BMI (β=−0.63; P<.001) after adjustment for age, gender, systolic blood pressure, total cholesterol, and family history of heart disease. The interaction between total activity and elevated METs was associated with lower fasting glucose (β=−0.07; P=.004). Elevated METs were associated with higher log(coronary artery calcium+1) and higher BMI (P<.001). Total activity was significantly associated with higher indexed biventricular systolic (P=.01 for left and P=.006 for right) and diastolic volumes (P<.001) and higher indexed left ventricular mass (P=.005).

Conclusions: We identified 3 groups of wearable metrics with distinct characteristics. While total activity had a significant relationship with self-reported PAI, most metrics of elevated METs did not. Total activity had a consistent and favorable association with lipid and glucose profiles and a dose-dependent association with cardiac remodeling. Elevated METs alone did not appear to have a significant association with favorable cardiovascular profiles. This study suggests that the total activity metrics are robust and dependable when interpreting an individual's activity levels, with construct validity according to self-reported PAI and a positive association with lipid and glucose profiles, and demonstrate dose-dependent associations with cardiac remodeling

XSL•FO
RenderX

after adjustment for demographics and risk factors. Findings related to elevated METs may be due to the Hawthorne effect and require further studies.

(JMIR Mhealth Uhealth 2025;13:e71213) doi:10.2196/71213

KEYWORDS

digital health; wearables; cardiovascular health; psychometric properties; metabolic health

Introduction

Background

Consumer health wearables are being rapidly adopted by the general population. Surveys have shown that up to 60% of the population own and use wearable devices [1-3]. Physical activity information is the most common measurement provided by current wearable devices. In addition, wearable-reported physical activity data have been studied and validated in various settings to ensure accuracy [4,5]. The abundance of accurate physical activity data presents an opportunity to transform them into actionable information.

Physical activity is a modifiable risk factor toward reducing cardiovascular disease, with up to 30% risk reduction in events among middle-aged adults [6]. Guideline recommendations on physical activity are largely derived from self-reported physical activity in epidemiological studies and its links to cardiovascular risk factors and events [6-13]. Self-reported physical activity has only been moderately correlated at best with actual physical activity [14], suggesting the imprecision of self-reported physical activity. Studies comparing self-reported and wearable-quantified physical activity have mainly looked at their correlation [15,16], but literature on their differences in association with cross-sectional health states is lacking. Less precise self-reported physical activity may mask significant associations with cardiometabolic profiles, which change subtly with physical activity. The inclusion of health wearables in recent large population studies [14,17,18] presents an opportunity to quantify physical activity more accurately. Moreover, there is a strong body of evidence suggesting that interventions using wearable activity trackers result in improvements in the physical activity of users; however, the effectiveness of these interventions on physiological outcomes remains uncertain [19].

Objectives

Despite widespread interest in the utility of wearable devices in health care, one of the challenges is synthesizing the vast amounts of data collected and integrating them into clinical practice [20]. While most studies have shown that step count consistently shows a positive relationship with cardiovascular health [8,21,22], other metrics such as exercise intensity and sedentary time have been less studied [22]. This study aimed to identify latent factors derived from Fitbit Charge HR (Google) physical activity metrics and correlate them with cross-sectional health outcomes. Considering the differences in activity behavior between weekends and weekdays [7], participants were given the fitness tracker for 1 week to ensure representative physical activity behavior, similar to that in previously conducted wearable studies [18,23].

Methods

Study Design

SingHEART (ClinicalTrials.gov NCT02791152) [17] was the first multiethnic prospective population-based study of healthy Asian individuals who completed detailed lifestyle and medical history questionnaires and had their activity levels recorded via wearables and their cross-sectional cardiovascular status detailed via blood tests and imaging (described in the following sections). A total of 457 individuals from SingHEART [17] were included in this cross-sectional study. Healthy male and female participants aged 21 to 69 years without any previous cardiovascular disease (eg, ischemic heart disease, stroke, or peripheral vascular disease) or diabetes mellitus were recruited from the Singapore general public via advertisements (eg, posters and the local newspaper).

Baseline Blood Tests and Measurements

A comprehensive set of assessments was conducted at baseline. Basic blood tests, including fasting lipids and glucose, and clinical parameters of height, weight, and hip and waist circumference were measured.

Physical Activity

Participants were given the Fitbit Charge HR wearable device, which tracked heart rate, step count, and intensity of physical activity. The device was worn over a course of 7 days [16]. Step counts were retrieved as daily totals, and the mean number of steps in each day was derived. There were 9 averaged metrics on physical activity presented by Fitbit Charge HR on a per-day basis: calories burned, number of floors, number of steps, distance, minutes spent sedentary, minutes spent lightly active, minutes spent active, minutes spent very active, and activity calories. The various activity intensity levels were previously defined by Fitbit: sedentary, lightly active (1.5-3 metabolic equivalents of task [METs]), fairly active (3-6 METs in at least 10-minute bouts), and very active (≥6 METs or ≥145 steps per minute in at least 10-minute bouts) [24]. Participants were asked to carry on with their regular jobs and physical activity and not to artificially inflate their activity levels so that the wearable was able to capture an accurate representation of their usual lifestyle. Self-reported physical activity was also quantified using the physical activity index (PAI) [25] derived from the General Practice Physical Activity Questionnaire [26].

Blood Pressure Monitoring

Ambulatory blood pressure (BP) was measured using a cuff monitor (Spacelabs Healthcare models 90227 and 90217A). BP was then analyzed as a continuous variable.

XSL•FO
RenderX

Coronary Artery Calcium Scoring

All participants underwent noncontrast cardiac computed tomography scans using a 320-slice computed tomography scanner. Coronary artery calcium (CAC) was quantified using the Agatston score [12] via the Vitrea Workstation. CAC was logarithmically transformed into a continuous variable (log[CAC+1]). As this cohort of individuals was free of known cardiovascular disease, the CAC helped detect subclinical, early atherosclerosis.

Magnetic Resonance Imaging Cardiac Measurements

Cardiovascular magnetic resonance was conducted for all participants using either the Ingenia 3.0T (Philips) or MAGNETOM Aera 1.5T (Siemens Healthineers) scanners. Parameters of cardiac mass, volumes, and function were measured using the cvi42 software (Circle Cardiovascular Imaging) and standardized protocols [27]. Left ventricular mass without papillary muscles was indexed to body surface area according to the DuBois formula, and left ventricular mass index was used for analysis [28].

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki [29]. Written informed consent was obtained from all participants, and the study was approved by the SingHealth Centralised Institutional Review Board (reference 2015/2601). The data used for analysis were anonymized to ensure participants' privacy and confidentiality. Participants were given SGD $50 (US $39) as compensation for their transport and inconvenience.

Statistical Analysis: Identifying Latent Factor Measures Using Fitbit Metrics

In this paper, *items* refer to the Fitbit-derived metrics. Using factor analysis, we determined the optimal number of factors using the Kaiser rule (ie, eigenvalue>1). To identify items contributing to factors, we used the criterion of factor loading of >0.4 as a cutoff [30].

Internal consistency reliability (the extent to which the items grouped together are appropriate to measure the same property) was assessed for each factor using the Cronbach α [31]. Construct validity of factors (the ability to measure the intended property in reference to acknowledged standards) was assessed against the self-reported PAI. A value of ≥0.6, 0.4 to 0.5, and 0.2 to 0.3 was considered a strong, moderate, and weak correlation, respectively [32].

Structural equation modeling [33] was used to analyze the relationship between measured items and latent constructs.

After factor analysis using oblique rotation, the identified factors were predicted using regression scoring [34]. Each individual factor was obtained as a weighted sum of standardized versions of the items with a mean of 0.

Intergroup differences in risk factors were evaluated using chi-square tests for categorical variables and ANOVA for continuous variables with Bonferroni correction. Statistical significance was considered as $P<.05$.

Multivariate regression analysis adjusting for age, gender, BMI, 24-hour average systolic BP, total cholesterol, and family history of coronary artery disease testing the identified latent factors was conducted for the lipid and glucose profiles, log(CAC+1), and cardiac magnetic resonance imaging (MRI) variables.

All analyses were conducted using Stata (version 17.0; StataCorp) [35].

Results

Overview

The baseline characteristics of the population as a whole and divided by self-reported PAI categories are shown in Table 1. The average age of the participants was 50.05 (SD 9.42) years, and 46.6% (213/457) were male. Most of the participants (234/457, 51.2%) had sedentary jobs, and 4.4% (20/457) had highly active jobs. There was a significantly higher proportion of male individuals in the moderately active and active categories. There was also a significant trend of increasing calories burned, number of steps, distance, number of floors, activity calories, and very active minutes recorded via Fitbit with increasing PAI levels but not for the activity minutes in other MET categories. Significant trends of increasing cardiac remodeling via indexed left and right ventricular volumes and indexed left ventricular mass measured through cardiac MRI with increasing PAI levels were observed. There was no significant association between increasing PAI levels and BP, lipid and glucose profiles, BMI, or log(CAC+1).

XSL•FO
RenderX

Table 1. Demographic and clinical characteristics of participants in general and across physical activity index categories (N=457).

	All	Sedentary (n=105)	Moderately inactive (n=106)	Moderately active (n=108)	Active (n=106)	*P* value
Age (y), mean (SD)	50.05 (9.42)	49.62 (9.86)	51.78 (9.51)	49.31 (9.31)	49.81 (8.57)	.63
Male sex, n (%)	213 (46.6)	38 (36.2)	38 (35.8)	49 (45.4)	63 (59.4)	<.001
Job type, n (%)						<.001
Sedentary	234 (51.2)	70 (66.7)	38 (35.8)	68 (63)	54 (50.9)	
Active	135 (29.5)	0 (0)	60 (56.6)	17 (15.7)	35 (33)	
Highly active	20 (4.4)	0 (0)	0 (0)	6 (5.6)	5 (4.7)	
Calories burned, mean (SD)	938.74 (353.20)	2065.53 (510.64)	2077.66 (438.31)	2124.99 (451.90)	2264.87 (478.37)	.008
Number of steps, mean (SD)	938.74 (353.20)	8398.25 (3220.46)	9506.717 (3238.37)	9452.63 (2692.92)	10,525.69 (3591.07)	<.001
Distance (km), mean (SD)	6.55 (2.35)	5.70 (2.28)	6.50 (2.34)	6.51 (1.95)	7.44 (2.58)	<.001
Number of floors, mean (SD)	9.03 (7.91)	6.59 (4.59)	9.62 (8.99)	8.65 (6.61)	11.26 (9.74)	<.001
Sedentary minutes, mean (SD)	873.10 (119.08)	889.45 (110.78)	870.34 (120.07)	859.88 (122.64)	873.14 (122.10)	.34
Lightly active minutes, mean (SD)	222.40 (64.46)	216.65 (62.20)	231.42 (63.59)	225.05 (65.15)	216.39 (66.49)	.26
Fairly active minutes, mean (SD)	18.51 (17.43)	15.05 (16.97)	17.76 (15.65)	19.95 (16.80)	21.21 (19.66)	.05
Very active minutes, mean (SD)	19.11 (19.26)	14.49 (18.04)	17.57 (18.78)	18.35 (14.99)	26.01 (22.83)	<.001
Activity calories, mean (SD)	931.46 (353.70)	854.76 (378.15)	919.14 (320.72)	934.51 (306.16)	1016.63 (389.61)	.01
24-h SBP[a] (mm Hg), mean (SD)	116.37 (13.80)	116.07 (14.53)	116.01 (13.09)	116.98 (15.19)	116.41 (12.36)	.95
24-h DBP[b] (mm Hg), mean (SD)	73.88 (9.12)	73.88 (9.58)	73.89 (8.88)	73.79 (9.63)	73.96 (8.46)	>.99
24-h HR[c] (bpm[d]), mean (SD)	71.43 (8.71)	73.96 (7.63)	71.45 (9.60)	72.00 (8.34)	68.33 (8.32)	<.001
Total cholesterol (mmol/L), mean (SD)	5.44 (0.93)	5.40 (0.88)	5.38 (0.85)	5.42 (0.83)	5.57 (1.06)	.42
LDL[e] (mmol/L), mean (SD)	3.40 (0.82)	3.40 (0.77)	3.27 (0.86)	3.357 (0.71)	3.57 (0.92)	.06
HDL[f] (mmol/L), mean (SD)	1.493 (0.31)	1.44 (0.29)	1.47 (0.32)	1.55 (0.35)	1.49 (0.30)	.11
Triglycerides (mmol/L), mean (SD)	1.19 (0.70)	1.2464762 (0.75)	1.2806604 (0.75)	1.12 (0.60)	1.13 (0.70)	.26
Glucose (mmol/L), mean (SD)	5.28 (0.75)	5.34 (1.10)	5.34 (0.55)	5.25 (0.48)	5.21 (0.71)	.51
BMI (kg/m^2), mean (SD)	23.49 (3.49)	24.05 (4.40)	23.42 (3.2)	23.24 (3.54)	23.26 (2.56)	.30
Log(CAC[g]+1), mean (SD)	1.16 (2.01)	0.97 (1.89)	1.08 (2.03)	1.13 (1.91)	1.45 (2.18)	.35
Indexed RVEDV[h] (mL/m^2), mean (SD)	2087.94 (362.35)	1940.40 (273.89)	2046.38 (375.31)	2115.68 (339.49)	2249.46 (384.74)	<.001
Indexed RVESV[i] (mL/m^2), mean (SD)	891.14 (256.14)	777.37 (188.92)	869.63 (279.37)	918.49 (231.37)	999.50 (267.26)	<.001

XSL•FO
RenderX

	All	Sedentary (n=105)	Moderately inactive (n=106)	Moderately active (n=108)	Active (n=106)	*P* value
Indexed LVEDV[j] (mL/m^2), mean (SD)	1994.67 (288.64)	1889.03 (238.83)	1953.58 (284.17)	2040.58 (277.55)	2092.64 (309.52)	<.001
Indexed LVESV[k] (mL/m^2), mean (SD)	767.64 (193.41)	698.47 (168.88)	744.28 (192.64)	793.29 (182.67)	832.72 (203.37)	<.001
LV[l] mass (g/m^2), mean (SD)	1227.55 (252.72)	1164.41 (221.55)	1199.44 (246.78)	1262.25 (264.28)	1281.77 (260.99)	.002

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]HR: heart rate.

[d]bpm: beats per minute.

[e]LDL: low-density lipoprotein.

[f]HDL: high-density lipoprotein.

[g]CAC: coronary artery calcium.

[h]RVEDV: right ventricular end diastolic volume.

[i]RVESV: right ventricular end systolic volume.

[j]LVEDV: left ventricular end diastolic volume.

[k]LVESV: left ventricular end systolic volume.

[l]LV: left ventricle.

Exploratory Factor Analysis

Using exploratory factor analysis with a cutoff of eigenvalue>1, we identified 3 latent factors measured via the Fitbit metrics on physical activity. The 3 factors identified were elevated METs, total activity, and others. While elevated METs represent the time engaged in moderate- and higher-level physical activity and catabolism measured in calories, total activity describes the average activity, including activity during nonexercise time, throughout the day.

Under factor 1, elevated METs, items with loadings of >0.4 included calories burned per day; minutes per day spent fairly active, defined as 3 to 6 METs; minutes per day spent very active, defined as ≥6 METs; and activity calories. For factor 2, total activity, items with loadings of >0.4 included number of steps per day, distance in kilometers per day, and number of floors per day. Factor 3 (others) items with loadings of >0.4 included sedentary time minutes and minutes spent lightly active, defined as 1.5 to 3 METs, where the loading value for sedentary time minutes was negative (Table 2).

Structural equation modeling was conducted on items and constructs, and the results are shown in Figure 1. There was a strong correlation (>0.6) between elevated METs and total activity.

XSL•FO
RenderX

Table 2. Item-to-scale (latent factor) correlation and Cronbach α of items.

	Factor loading	Item-to-scale correlation	Cronbach α
Factor 1: elevated MET[a] density[b]			
Calories burned per day	0.8394	0.7789	0.7937
Minutes spent fairly active per day	0.7769	0.7911	0.7849
Minutes spent very active per day	0.7511	0.7856	0.7888
Activity calories per day	0.7452	0.8683	0.7234
Factor 2: total activity[c]			
Number of steps per day	0.7061	0.8966	0.5589
Distance (km) per day	0.6916	0.9105	0.5217
Number of floors per day	0.8602	0.6956	0.9423
Factor 3: other nonsedentary time[d]			
Sedentary time per day	−0.8592	—[e]	—
Minutes spent lightly active per day	0.8896	—	—

[a]MET: metabolic equivalent of task.

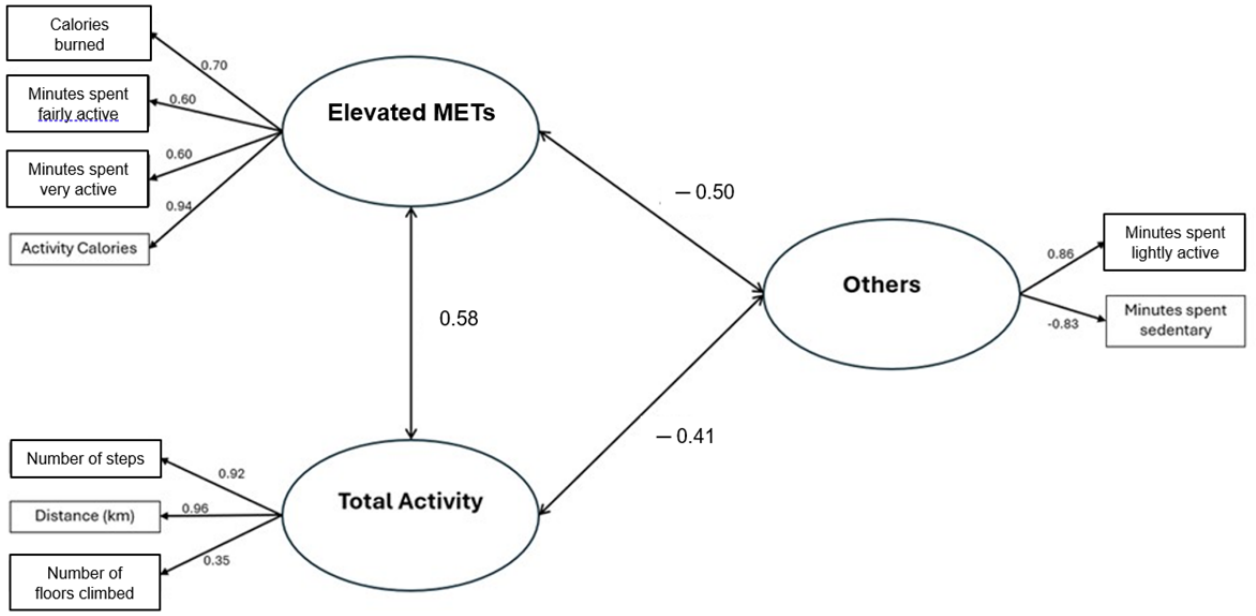
[b]Overall Cronbach α=0.8202.

[c]Overall Cronbach α=0.7815.

[d]Overall Cronbach α=0.7452.

[e]Not applicable.

Figure 1. Confirmatory factor analysis and sequential equation modeling of the 9 Fitbit metrics. The values show the strength of association between the metrics and the factor and between the different factors. MET: metabolic equivalent of task.



Reliability

For assessing internal consistency reliability, the Cronbach α exceeded 0.70 for all 3 constructs (elevated METs=0.820; total activity=0.781; others=0.745). Correlations between the factors and hypothesized items are shown in Table 2.

Construct Validity

The correlation between self-reported PAI [36] and elevated METs and total activity was significant but weak (0.23 and 0.26, respectively; *P*<.001). The *others* factor was not correlated with PAI (*P*=.64).

Regression Analysis

As there was a strong correlation between elevated METs and total activity, we included and tested the interaction term of elevated METs and total activity in the multivariate regression analysis.

XSL•FO
RenderX

Lipid and Glucose Profiles and BMI

In contrast to the PAI, which did not show any significant trend regarding lipid and glucose profiles and BMI, total activity was associated with significantly higher high-density lipoproteins (HDLs; β=0.06; *P*<.001) and lower triglycerides (β=−0.10; *P*=.006); total activity trended toward being associated with lower fasting glucose, and the interaction term of elevated METs and total activity was significantly associated with lower fasting glucose levels (β=−0.07; *P*=.004). The total activity and other factors were significantly associated with lower BMI (β=−0.63 and −0.82, respectively; *P*<.001), whereas the elevated METs factor was associated with higher BMI (β=1.47; *P*<.001). There was no significant relationship between the various factors and low-density lipoprotein (LDL; Table 3). The coefficient plots of the variables against the lipid and glucose profiles and BMI are shown in Figure 2.

Table 3. Linear regression of lipid and glucose profiles and BMI against demographics, clinical variables, and activity metrics.

	HDL[a] (mmol/L)[b]			Triglycerides (mmol/L)[c]			LDL[d] (mmol/L)[e]			Glucose (mmol/L)[f]			BMI (kg/m^2)[g]		
	β (SE)	*t* test (*df*)	*P* value	β (SE)	*t* test (*df*)	*P* value	β (SE)	*t* test (*df*)	*P* value	β (SE)	*t* test (*df*)	*P* value	β (SE)	*t* test (*df*)	*P* value
Male sex	−0.20 (0.03)	−5.97 (10)	<.001	0.21 (0.08)	2.64 (10)	.009	0.09 (0.04)	2.10 (10)	.04	0.03 (0.08)	0.43 (10)	.67	−0.39 (0.36)	−1.08 (10)	.28
Age (y)	0.00 (0.00)	0.37 (10)	.71	0.00 (0.00)	1.10 (10)	.27	0.00 (0.00)	−0.46 (10)	.65	0.01 (0.00)	3.17 (10)	.002	−0.03 (0.02)	−1.91 (10)	.06
BMI (kg/m^2)	−0.03 (0.00)	−6.05 (10)	<.001	0.03 (0.01)	2.83 (10)	.005	0.01 (0.01)	2.46 (10)	.01	0.03 (0.01)	2.60 (10)	.01	—[h]	—	—
24-h SBP[i]	0.00 (0.00)	−0.62 (10)	.53	0.00 (0.00)	1.21 (10)	.23	0.00 (0.00)	−0.58 (10)	.56	0.00 (0.00)	1.27 (10)	.20	0.06 (0.01)	5.41 (10)	<.001
Total cholesterol (mmol/L)	0.10 (0.01)	7.37 (10)	<.001	0.15 (0.03)	4.62 (10)	<.001	0.81 (0.02)	42.59 (10)	<.001	0.01 (0.03)	0.21 (10)	.83	0.19 (0.15)	1.23 (10)	.22
Family history of CAD[j]	0.00 (0.03)	0.05 (10)	.96	0.07 (0.08)	0.97 (10)	.34	−0.01 (0.04)	−0.23 (10)	.82	0.06 (0.08)	0.75 (10)	.45	−0.29 (0.35)	−0.82 (10)	.41
Elevated METs[k]	0.00 (0.02)	0.20 (10)	.84	0.08 (0.05)	1.75 (10)	.08	−0.03 (0.03)	−1.15 (10)	.25	0.21 (0.05)	4.46 (10)	<.001	1.48 (0.20)	7.55 (10)	<.001
Total activity	0.06 (0.02)	3.85 (10)	<.001	−0.10 (0.04)	−2.74 (10)	.006	−0.02 (0.02)	−0.76 (10)	.45	−0.07 (0.04)	−1.86 (10)	.06	−0.63 (0.17)	−3.81 (10)	<.001
Others	−0.01 (0.02)	−0.36 (10)	.72	−0.01 (0.04)	−0.29 (10)	.78	0.01 (0.02)	0.43 (10)	.67	−0.03 (0.04)	−0.81 (10)	.42	−0.82 (0.17)	−4.93 (10)	<.001
Interaction term of elevated METs and total activity	−0.01 (0.01)	−0.98 (10)	.33	−0.02 (0.02)	−0.91 (10)	.36	0.01 (0.01)	1.01 (10)	.31	−0.07 (0.02)	−2.92 (10)	.004	−0.19 (0.10)	−1.91 (10)	.06
_cons (constant of the model)	1.68 (0.16)	10.60 (10)	<.001	−0.98 (0.38)	−2.58 (10)	.01	−1.22 (0.21)	−5.71 (10)	<.001	3.63 (0.39)	9.25 (10)	<.001	17.22 (1.53)	11.22 (10)	<.001

[a]HDL: high-density lipoprotein.

[b]Adjusted R^2=0.294.

[c]Adjusted R^2=0.164.

[d]LDL: low-density lipoprotein.

[e]Adjusted R^2=0.807.

[f]Adjusted R^2=0.159.

[g]Adjusted R^2=0.263.

[h]Not applicable.

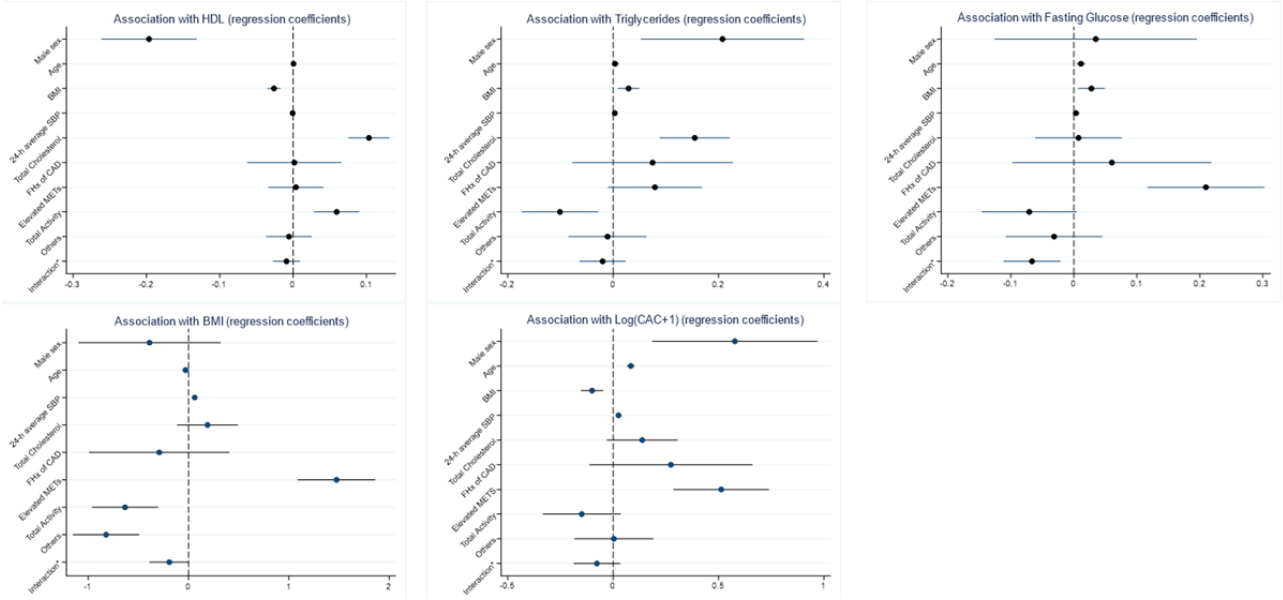
[i]SBP: systolic blood pressure.

[j]CAD: coronary artery disease.

[k]MET: metabolic equivalent of task.

XSL•FO
RenderX

Figure 2. Regression coefficient plot of activity metric latent factors against lipid and glucose profiles, BMI and log(coronary artery calcium+1) adjusting for demographics and risk factors. CAC: coronary artery calcium; CAD: coronary artery disease; FHx: family history; HDL: high-density lipoprotein; MET: metabolic equivalent of task; SBP: systolic blood pressure.



Cardiac MRI and Coronary Calcium Score

The elevated METs factor was significantly associated with higher log(CAC+1) (β=−0.51; *P*<.001; Table S1 in Multimedia Appendix 1). Total activity was significantly associated with higher indexed biventricular diastolic and systolic volumes and higher indexed left ventricular mass (Table 4). The coefficient plots of the variables against the log(CAC+1) and cardiac MRI variables are shown in Figures 2 and 3, respectively.

XSL•FO
RenderX

Table 4. Linear regression of cardiac magnetic resonance imaging dimensions against demographics, clinical variables, and activity metrics.

	Indexed RVEDV[a,b]			Indexed RVESV[c,d]			Indexed LVEDV[e,f]			Indexed LVESV[g,h]			Indexed LVM[i,j]		
	β (SE)	*t* test (*df*)	*P* value	β (SE)	*t* test (*df*)	*P* value	β (SE)	*t* test (*df*)	*P* value	β (SE)	*t* test (*df*)	*P* value	β (SE)	*t* test (*df*)	*P* value
Male sex	314.80 (46.29)	6.80 (10)	<.001	251.07 (32.26)	7.78 (10)	<.001	208.45 (32.06)	6.50 (10)	<.001	156.81 (20.88)	7.51 (10)	<.001	218.37 (25.34)	8.62 (10)	<.001
Age (y)	−12.73 (2.02)	−6.29 (10)	<.001	−8.53 (1.41)	−6.05 (10)	<.001	−9.78 (1.38)	−7.09 (10)	<.001	−5.94 (0.90)	−6.61 (10)	<.001	−4.12 (1.10)	−3.74 (10)	<.001
BMI (kg/m^2)	5.32 (6.16)	0.87 (10)	.39	0.02 (4.29)	0.00 (10)	>.99	−0.42 (4.20)	−0.10 (10)	.92	−4.50 (2.73)	−1.65 (10)	.10	2.28 (3.31)	0.69 (10)	.49
24-h SBP[k]	−2.79 (1.53)	−1.82 (10)	.07	−1.98 (1.07)	−1.86 (10)	.06	−1.07 (1.03)	−1.04 (10)	.30	−1.00 (0.67)	−1.50 (10)	.14	6.25 (0.82)	7.67 (10)	<.001
Total cholesterol (mmol/L)	−13.91 (20.70)	−0.67 (10)	.50	−4.81 (14.42)	−0.33 (10)	.74	−32.76 (13.65)	−2.40 (10)	.02	−17.70 (8.89)	−1.99 (10)	.047	3.77 (10.80)	0.35 (10)	.73
Family history of CAD[l]	51.70 (45.16)	1.14 (10)	.25	22.65 (31.47)	0.72 (10)	.47	51.31 (31.65)	1.62 (10)	.11	39.50 (20.61)	1.92 (10)	.06	−1.38 (25.09)	−0.06 (10)	.96
Elevated METs[m]	−46.04 (27.09)	−1.70 (10)	.09	−11.82 (18.88)	−0.63 (10)	.53	−32.96 (18.46)	−1.79 (10)	.08	0.52 (12.02)	0.04 (10)	.97	0.53 (14.58)	0.04 (10)	.97
Total activity	86.23 (22.27)	3.87 (10)	<.001	43.23 (15.52)	2.79 (10)	.006	61.14 (14.99)	4.08 (10)	<.001	23.97 (9.76)	2.46 (10)	.01	33.25 (11.82)	2.81 (10)	.005
Others	6.79 (21.93)	0.31 (10)	.76	3.61 (15.28)	0.24 (10)	.81	3.48 (15.23)	0.23 (10)	.82	2.16 (9.92)	0.22 (10)	.83	−4.71 (12.18)	−0.39 (10)	.70
Interaction term of elevated METs and total activity	−11.83 (15.78)	−0.75 (10)	.45	−13.12 (11.00)	−1.19 (10)	.23	−3.31 (8.94)	−0.37 (10)	.71	−5.47 (5.82)	−0.94 (10)	.35	−5.84 (7.05)	−0.83 (10)	.41
_cons (constant of the model)	2844.39 (226.35)	12.57 (10)	<.001	1455.21 (157.73)	9.23 (10)	<.001	2702.53 (154.30)	17.51 (10)	<.001	1312.02 (100.47)	13.06 (10)	<.001	548.18 (122.74)	4.47 (10)	<.001

[a]RVEDV: right ventricular end diastolic volume.

[b]Adjusted R^2=0.268.

[c]RVESV: right ventricular end systolic volume.

[d]Adjusted R^2=0.286.

[e]LVEDV: left ventricular end diastolic volume.

[f]Adjusted R^2=0.225.

[g]LVESV: left ventricular end systolic volume.

[h]Adjusted R^2=0.250.

[i]LVM: left ventricular mass.

[j]Adjusted R^2=0.385.

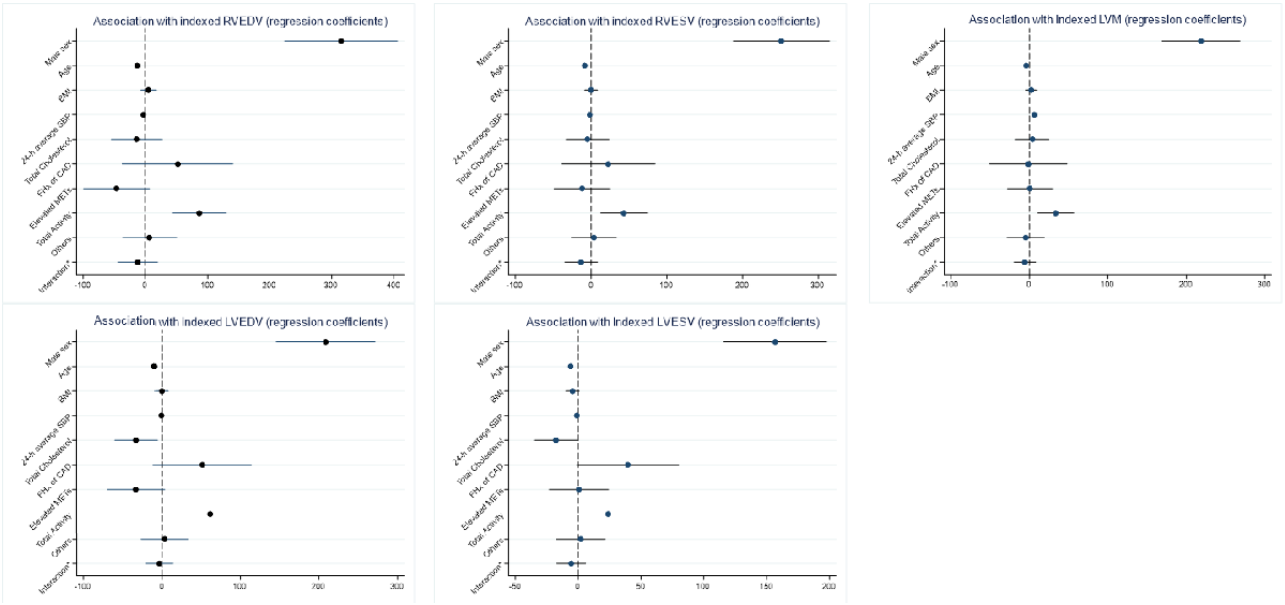
[k]SBP: systolic blood pressure.

[l]CAD: coronary artery disease.

[m]MET: metabolic equivalent of task.

XSL•FO
RenderX

Figure 3. Regression coefficient plot of activity metric latent factors against cardiac magnetic resonance imaging dimensions adjusting for demographics and risk factors. CAD: coronary artery disease; FHx: family history; LVEDV: left ventricular end diastolic volume; LVESV: left ventricular end systolic volume; LVM: left ventricular mass; MET: metabolic equivalent of task; RVEDV: right ventricular end diastolic volume; RVESV: right ventricular end systolic volume; SBP: systolic blood pressure.



Quantification of Change in Lipid Profile and BMI per 1000 Increase in Steps

As the most used metric in total activity was step count, we sought to quantify the change in lipid profile and BMI with every 1000-step change. After adjusting for demographics and risk factors, every 1000-step increase per day increased HDL by 0.01 mmol/L, decreased triglycerides by 0.02 mmol/L, and reduced BMI by 0.11 kg/m^2 (Table S2 in Multimedia Appendix 1).

Discussion

Principal Findings

The purpose of this study was to identify the underlying factors measured through an activity tracker, the Fitbit Charge HR, and extrapolate these factors to cross-sectional cardiovascular phenotypes. There has been no previous study to our knowledge conducted for this. Understanding the underlying factors measured by an activity tracker is important to set specific activity goals and prescribe precise physical activity domains using activity trackers. Previous studies have looked at the effectiveness of activity trackers in increasing physical activity [37] and largely focused on step counts in reduction of all-cause mortality [8,38].

This study revealed three factors measured by the Fitbit Charge HR: (1) elevated METs (calories burned per day, minutes per day spent fairly active, 3-6 METs; minutes per day spent very active, ≥6 METs; and activity calories), (2) total activity (number of steps per day, distance in kilometers per day, and number of floors per day), and (3) others (sedentary time minutes and minutes spent lightly active, 1.5-3 METs).

Total activity, represented by steps, distance, and floors, appeared to be consistently associated with a favorable cardiovascular risk profile in our study. It appears that the absolute volume of activity, quantified via total activity metrics, was more consistent in terms of positive cardiovascular benefit [39-41] compared to elevated METs. There was a dose-dependent increase in HDL and a reduction in triglycerides and BMI. HDL is more sensitive to exercise compared to LDL and triglycerides [42]. The lack of decreased LDL with total activity may be due to the study being observational rather than interventional in nature, where an exercise program was shown to reduce LDL, especially when the mean LDL of the participants was normal at 3.41 (SD 0.83) mmol/dL. Given that the adjusted R^2 of the model was high at >0.80, it might also be that other factors played a much larger role in influencing LDL levels compared to physical activity. Compared to self-reported PAI, which had no association with lipid and glucose profiles and BMI, these results suggest that wearable-quantified activity metrics are more accurate and precise in recording activity level, as evidenced by significant associations with cross-sectional cardiometabolic profiles.

There was also significant cardiovascular remodeling based on total activity and self-reported PAI, with significantly larger indexed biventricular systolic and diastolic volumes and indexed left ventricular mass after adjustment for baseline demographics and risk factors. This finding was also consistent with those of previous exercise studies [43,44]. In our study, we can only conclude that total activity was associated with physiological adaptation of the heart, but the functional consequence is unknown.

Regarding the association between elevated METs and higher log(CAC+1), one possible hypothesis is the reverse J curve of exercise [45,46], where excessive exercise gradually precipitates cardiovascular damage. A UK Biobank study demonstrated that the cardiovascular protective effect was lost in the 75th percentile of moderate to vigorous physical activity and above [47], and a systematic review showed that endurance training increased the odds of having high CAC but without increased

XSL•FO
RenderX

prevalence of high-risk plaque [48]. Another possible explanation is a change in participant behavior due to the Hawthorne effect [49]. The METs calculated via Fitbit were based on heart rate and step count over a prespecified period [50]; it is possible that less healthy participants attempted to put in extra exercise time over the 7-day observation period, and with poorer heart rate recovery [51], this would increase their scores in elevated METs. Total activity detailed the average step count and distance covered daily and was harder to inflate compared to elevated METs. Hence, this finding warrants further prospective observational studies or longer periods of monitoring before drawing conclusions.

There are also very few studies that have correlated step count and activity levels to reduced subclinical atherosclerosis [52]. However, there is consistent literature suggesting that increasing step count reduces mortality and cardiovascular events [8,38,53]. Previous population studies have suggested that exercise is a risk modifier, especially in patients with high CAC [54]. As shown in our study, this could be mediated through increased HDL and reduced triglycerides, BMI, and fasting glucose [55,56]. While the pathophysiology might be hard to prove in human studies, animal studies have shown that exercise training reduces glucose and triglyceride levels in LDL-knockout mice [57] and reduces hyperlipidemia-induced cardiac damage through reducing oxidative stress and inflammation in apolipoprotein E–deficient mice [58], hence reducing mortality in atherosclerosis-predisposed animals. HDL is an antioxidant that provides potent protection to LDL from oxidative damage by free radicals, hence inhibiting the generation of proinflammatory oxidized lipids [59], and aerobic exercise has been found to increase both HDL quantity and quality [60]. This might be a reason for the lack of significant association between total activity and presence of atherosclerosis quantified through log(CAC+1) as atherosclerosis might have been caused by other genetic and epigenetic factors and physical activity is but a risk modifier.

Our study showed that there are distinct characteristics conveyed by the different Fitbit activity metrics. Interpreting them individually may not be as helpful; there is a need to recognize the underlying factor described by each metric to make more precise recommendations for exercise and physical health. This study shows that increased activity regardless of MET intensity, represented by total activity, has significant beneficial associations with cardiometabolic profiles. After accounting for gender, BP, and other risk factors, every 1000-step increase per day increased HDL by 0.01 mmol/L, reduced triglycerides by 0.02 mmol/L, and reduced BMI by 0.11 kg/m^2. From a clinical perspective, it may be reasonable to look at the metrics represented sequentially and hierarchically, first ensuring that the individual has achieved reasonable total activity before focusing on the intensity of the activity quantified through the metrics in the elevated METs factor. For example, for patients with no CAC, physicians can recommend 10,000 steps per day [61] and no restriction on activity; conversely, for patients with high CAC who are looking for exercise recommendations, focusing mainly on 10,000 steps per day and moderating very vigorous activity based on effort limitation, especially if the patient had no exercise regime before, may be reasonable.

This was an observational study and not an interventional study, hence limiting it to associations. The 7-day observation period, due to limitations in device availability, may also not be a true representation of activity levels in all patients, although the study team counseled participants to continue their usual activity levels. In addition, this study comprised mainly an Asian population; hence, the generalizability of these findings to other populations may be limited. It will be helpful to use the same framework and explore these associations in other populations. The upcoming 10-year follow-up for this prospective cohort will also shed light on the impact of wearable-quantified physical activity on cardiovascular events stratified by baseline CAC scores.

Conclusions

There are different characteristics conveyed by the different Fitbit physical activity metrics. This study yielded 3 main latent factors and described their associations with cardiovascular health. Total activity, rather than elevated METs, was associated with beneficial effects on cardiometabolic profiles. Total activity appears to be more sensitive and precise than self-reported PAI, with significant associations with lipid and glucose profiles and BMI. One can consider interpreting activity metrics on wearables sequentially and hierarchically, focusing first on metrics in the total activity factor before considering metrics in the elevated METs factor when providing physical activity advice.

Acknowledgments

This work was supported by the Open Fund – Young Individual Research Grant from the National Medical Research Council, Ministry of Health, Republic of Singapore (grant MOH-OFYIRG21jun-0004), and the National Research Foundation, Singapore, under its AI Singapore program (grant AISG-GC-2019-001-2B). The authors would also like to thank the Lee Foundation for their grant support for the SingHEART study conducted at the National Heart Centre Singapore. The funding bodies were not involved in design, conduct, data interpretation, or writing of the manuscript.

Data Availability

The datasets generated or analyzed during this study are not publicly available as recruitment is still ongoing and the participants have not completed their follow-up but are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

KKY, CWLC, and SYT contributed to study conception and design. SCK, XS, AWT, and OV contributed to data collection. WH, TK, EDWL, and MKFW analyzed and interpreted the results. WH, EYKN, and JJLY prepared the draft manuscript. All authors reviewed the results and approved the final version of the manuscript.

Conflicts of Interest

KKY has received research funding from Amgen, AstraZeneca, Abbott Vascular, Bayer, Boston Scientific, Shockwave Medical, and Novartis (via the National Heart Centre Singapore); consulting fees from Abbott Vascular, Medtronic, Novartis, and Peijia Medical; and speaker fees from Shockwave Medical, Abbott Vascular, Boston Scientific, Medtronic, Alvimedica, BIOTRONIK, OrbusNeich, Shockwave Medical, Amgen, Novartis, AstraZeneca, MicroPort, Terumo, and Omnicare. KKY is also cofounder of and owns equity in Trisail Medical, of which Orbus Neich is an investor. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 28 KB - mhealth_v13i1e71213_app1.docx]

References

1. Wearable devices ownership in the United States in 2021, by age. Statista. URL: https://www.statista.com/statistics/1310364/wearable-devices-ownership-age-us/ [accessed 2024-04-16]
2. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443 [FREE Full text] [doi: 10.2196/22443] [Medline: 33064083]
3. Shandhi MM, Singh K, Janson N, Ashar P, Singh G, Lu B, et al. Assessment of ownership of smart devices and the acceptability of digital health data sharing. NPJ Digit Med 2024 Feb 22;7(1):44 [FREE Full text] [doi: 10.1038/s41746-024-01030-x] [Medline: 38388660]
4. Adam Noah J, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. J Med Eng Technol 2013 Oct;37(7):456-462. [doi: 10.3109/03091902.2013.831135] [Medline: 24007317]
5. Sasaki JE, Hickey A, Mavilia M, Tedesco J, John D, Kozey Keadle S, et al. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. J Phys Act Health 2015 Feb;12(2):149-154. [doi: 10.1123/jpah.2012-0495] [Medline: 24770438]
6. Shiroma EJ, Lee IM. Physical activity and cardiovascular health: lessons learned from epidemiological studies across age, gender, and race/ethnicity. Circulation 2010 Aug 17;122(7):743-752. [doi: 10.1161/circulationaha.109.914721]
7. To QG, Stanton R, Schoeppe S, Doering T, Vandelanotte C. Differences in physical activity between weekdays and weekend days among U.S. children and adults: cross-sectional analysis of NHANES 2011-2014 data. Prev Med Rep 2022 Aug;28:101892 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101892] [Medline: 35813401]
8. Stens NA, Bakker EA, Mañas A, Buffart LM, Ortega FB, Lee DC, et al. Relationship of daily step counts to all-cause mortality and cardiovascular events. J Am Coll Cardiol 2023 Oct 10;82(15):1483-1494 [FREE Full text] [doi: 10.1016/j.jacc.2023.07.029] [Medline: 37676198]
9. Sattelmair J, Pertman J, Ding EL, Kohl HWIII, Haskell W, Lee IM. Dose response between physical activity and risk of coronary heart disease. Circulation 2011 Aug 16;124(7):789-795. [doi: 10.1161/circulationaha.110.010710]
10. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. BMJ 2016 Aug 09;354:i3857 [FREE Full text] [doi: 10.1136/bmj.i3857] [Medline: 27510511]
11. Hamer M, Chida Y. Walking and primary prevention: a meta-analysis of prospective cohort studies. Br J Sports Med 2008 Apr 29;42(4):238-243. [doi: 10.1136/bjsm.2007.039974] [Medline: 18048441]
12. Ekelund U, Steene-Johannessen J, Brown WJ, Fagerland MW, Owen N, Powell KE, et al. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. Lancet 2016 Sep 24;388(10051):1302-1310. [doi: 10.1016/S0140-6736(16)30370-1] [Medline: 27475271]
13. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. Circulation 2019 Sep 10;140(11):e596-e646 [FREE Full text] [doi: 10.1161/CIR.0000000000000678] [Medline: 30879355]
14. Quinlan C, Rattray B, Pryor D, Northey JM, Anstey KJ, Butterworth P, et al. The accuracy of self-reported physical activity questionnaires varies with sex and body mass index. PLoS One 2021 Aug 11;16(8):e0256008 [FREE Full text] [doi: 10.1371/journal.pone.0256008] [Medline: 34379676]

XSL•FO
RenderX

15. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus accelerometer-measured physical activity. Med Sci Sports Exerc 2014 Jan;46(1):99-106. [doi: 10.1249/MSS.0b013e3182a0595f] [Medline: 23793232]
16. Beagle AJ, Tison GH, Aschbacher K, Olgin JE, Marcus GM, Pletcher MJ. Comparison of the physical activity measured by a consumer wearable activity tracker and that measured by self-report: cross-sectional analysis of the health eHeart study. JMIR Mhealth Uhealth 2020 Dec 29;8(12):e22090 [FREE Full text] [doi: 10.2196/22090] [Medline: 33372896]
17. Yap J, Lim WK, Sahlén A, Chin CW, Chew KM, Davila S, et al. Harnessing technology and molecular analysis to understand the development of cardiovascular diseases in Asia: a prospective cohort study (SingHEART). BMC Cardiovasc Disord 2019 Nov 21;19(1):259 [FREE Full text] [doi: 10.1186/s12872-019-1248-3] [Medline: 31752689]
18. Cochrane SK, Chen SH, Fitzgerald JD, Dodson JA, Fielding RA, King AC, et al. Association of accelerometry-measured physical activity and cardiovascular events in mobility-limited older adults: the LIFE (Lifestyle Interventions and Independence for Elders) study. J Am Heart Assoc 2017 Dec 02;6(12):e007215 [FREE Full text] [doi: 10.1161/JAHA.117.007215] [Medline: 29197830]
19. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/s2589-7500(22)00111-x]
20. Hughes A, Shandhi MM, Master H, Dunn J, Brittain E. Wearable devices in cardiovascular medicine. Circ Res 2023 Mar 03;132(5):652-670. [doi: 10.1161/circresaha.122.322389]
21. Lee IM, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 01;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
22. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
23. Jefferis BJ, Parsons TJ, Sartini C, Ash S, Lennon LT, Papacosta O, et al. Does total volume of physical activity matter more than pattern for onset of CVD? A prospective cohort study of older British men. Int J Cardiol 2019 Mar 01;278:267-272 [FREE Full text] [doi: 10.1016/j.ijcard.2018.12.024] [Medline: 30578094]
24. Semanik P, Lee J, Pellegrini CA, Song J, Dunlop DD, Chang RW. Comparison of physical activity measures derived from the Fitbit flex and the ActiGraph GT3X+ in an employee population with chronic knee symptoms. ACR Open Rheumatol 2020 Jan 02;2(1):48-52 [FREE Full text] [doi: 10.1002/acr2.11099] [Medline: 31943970]
25. Paffenbarger RSJ, Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. Am J Epidemiol 1978 Sep;108(3):161-175. [doi: 10.1093/oxfordjournals.aje.a112608] [Medline: 707484]
26. General practice physical activity questionnaire (GPPAQ). Department of Health and Social Care United Kingdom Government. 2013 Apr 23. URL: https://www.gov.uk/government/publications/general-practice-physical-activity-questionnaire-gppaq [accessed 2025-04-24]
27. Le TT, Tan RS, De Deyn M, Goh EP, Han Y, Leong BR, et al. Cardiovascular magnetic resonance reference ranges for the heart and aorta in Chinese at 3T. J Cardiovasc Magn Reson 2016 Apr 12;18(1):21 [FREE Full text] [doi: 10.1186/s12968-016-0236-3] [Medline: 27071974]
28. Ruggieri G, Rocca AR. Analysis of past and present methods of measuring and estimating body surface area and the resulting evaluation of its doubtful suitability to universal application. Blood Purif 2010 Nov 25;30(4):296-305. [doi: 10.1159/000321072] [Medline: 21109735]
29. WMA Declaration of Helsinki – ethical principles for medical research involving human participants. World Medical Association. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki/ [accessed 2025-05-27]
30. Rosenblad A. Applied Multivariate Statistics for the Social Sciences, Fifth Edition by James P. Stevens. Int Stat Rev 2009 Dec;77(3):476. [doi: 10.1111/j.1751-5823.2009.00095_13.x]
31. Luo N, Seng BK, Xie F, Li SC, Thumboo J. Psychometric evaluation of the Schizophrenia Quality of Life Scale (SQLS) in English- and Chinese-speaking Asians in Singapore. Qual Life Res 2008 Feb 21;17(1):115-122. [doi: 10.1007/s11136-007-9278-1] [Medline: 18030594]
32. Cohen RD. Validation of health-related quality of life instruments. Hepatology 1999 Jun;29(6 Suppl):7S-8S. [Medline: 10386075]
33. Muthén B. Latent variable structural equation modeling with categorical data. J Econom 1983 May;22(1-2):43-65. [doi: 10.1016/0304-4076(83)90093-3]
34. StataCorp. Stata 19 User's Guide. College Station, TX: StataCorp LLC; 2025.
35. StataCorp. Stata: Release 17. College Station, TX: StataCorp LLC; 2021.
36. Heron N, Tully MA, McKinley MC, Cupples ME. Physical activity assessment in practice: a mixed methods study of GPPAQ use in primary care. BMC Fam Pract 2014 Jan 15;15:11 [FREE Full text] [doi: 10.1186/1471-2296-15-11] [Medline: 24422666]
37. Tang MS, Moore K, McGavigan A, Clark RA, Ganesan AN. Effectiveness of wearable trackers on physical activity in healthy adults: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e15576 [FREE Full text] [doi: 10.2196/15576] [Medline: 32706685]

XSL•FO
RenderX

38. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022 Mar;7(3):e219-e228 [FREE Full text] [doi: 10.1016/S2468-2667(21)00302-9] [Medline: 35247352]
39. O'Donovan G, Owen A, Bird SR, Kearney EM, Nevill AM, Jones DW, et al. Changes in cardiorespiratory fitness and coronary heart disease risk factors following 24 wk of moderate- or high-intensity exercise of equal energy cost. J Appl Physiol (1985) 2005 May;98(5):1619-1625 [FREE Full text] [doi: 10.1152/japplphysiol.01310.2004] [Medline: 15640382]
40. Kraus WE, Houmard JA, Duscha BD, Knetzger KJ, Wharton MB, McCartney JS, et al. Effects of the amount and intensity of exercise on plasma lipoproteins. N Engl J Med 2002 Nov 07;347(19):1483-1492. [doi: 10.1056/NEJMoa020194] [Medline: 12421890]
41. Nybo L, Sundstrup E, Jakobsen MD, Mohr M, Hornstrup T, Simonsen L, et al. High-intensity training versus traditional exercise interventions for promoting health. Med Sci Sports Exerc 2010 Oct;42(10):1951-1958. [doi: 10.1249/MSS.0b013e3181d99203] [Medline: 20195181]
42. Wang Y, Xu D. Effects of aerobic exercise on lipids and lipoproteins. Lipids Health Dis 2017 Jul 05;16(1):132 [FREE Full text] [doi: 10.1186/s12944-017-0515-5] [Medline: 28679436]
43. Pelliccia A, Caselli S, Sharma S, Basso C, Bax JJ, Corrado D, et al. European Association of Preventive Cardiology (EAPC) and European Association of Cardiovascular Imaging (EACVI) joint position statement: recommendations for the indication and interpretation of cardiovascular imaging in the evaluation of the athlete's heart. Eur Heart J 2018 Jun 01;39(21):1949-1969 [FREE Full text] [doi: 10.1093/eurheartj/ehx532] [Medline: 29029207]
44. D'Andrea A, Morello A, Iacono AM, Scarafile R, Cocchia R, Riegler L, et al. Right ventricular changes in highly trained athletes: between physiology and pathophysiology. J Cardiovasc Echogr 2015;25(4):97-102 [FREE Full text] [doi: 10.4103/2211-4122.172486] [Medline: 28465945]
45. O'Keefe EL, Torres-Acosta N, O'Keefe JH, Lavie CJ. Training for longevity: the reverse J-curve for exercise. Mo Med 2020;117(4):355-361 [FREE Full text] [Medline: 32848273]
46. Merghani A, Maestrini V, Rosmini S, Cox AT, Dhutia H, Bastiaenan R, et al. Prevalence of subclinical coronary artery disease in masters endurance athletes with a low atherosclerotic risk profile. Circulation 2017 Jul 11;136(2):126-137. [doi: 10.1161/circulationaha.116.026964]
47. Khurshid S, Al-Alusi MA, Churchill TW, Guseh JS, Ellinor PT. Accelerometer-derived "weekend warrior" physical activity and incident cardiovascular disease. JAMA 2023 Jul 18;330(3):247-252 [FREE Full text] [doi: 10.1001/jama.2023.10875] [Medline: 37462704]
48. Guarnieri G, Pozzi FE, Conte E, Righetto M, Bartorelli A, Andreini D. Extreme endurance training and coronary artery disease: a systematic review and a meta-analysis. Int J Cardiol 2025 Jun 15;429:133172 [FREE Full text] [doi: 10.1016/j.ijcard.2025.133172] [Medline: 40097033]
49. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]
50. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
51. Liu Z, Zheng P, Fang Y, Huang J, Huang J, Chen L, et al. Joint association of sedentary time and physical activity with abnormal heart rate recovery in young and middle-aged adults. BMC Public Health 2024 Jul 04;24(1):1787 [FREE Full text] [doi: 10.1186/s12889-024-19298-9] [Medline: 38965484]
52. Storti KL, Pettee Gabriel KK, Underwood DA, Kuller LH, Kriska AM. Physical activity and coronary artery calcification in two cohorts of women representing early and late postmenopause. Menopause 2010;17(6):1146-1151 [FREE Full text] [doi: 10.1097/gme.0b013e3181e3a356] [Medline: 20651620]
53. Saint-Maurice PF, Troiano RP, Bassett DRJ, Graubard BI, Carlson SA, Shiroma EJ, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020 Mar 24;323(12):1151-1160 [FREE Full text] [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
54. Arnson Y, Rozanski A, Gransar H, Hayes SW, Friedman JD, Thomson LE, et al. Impact of exercise on the relationship between CAC scores and all-cause mortality. JACC Cardiovasc Imaging 2017 Dec;10(12):1461-1468 [FREE Full text] [doi: 10.1016/j.jcmg.2016.12.030] [Medline: 28528154]
55. Orozco-Beltran D, Gil-Guillen VF, Redon J, Martin-Moreno JM, Pallares-Carratala V, Navarro-Perez J, et al. Lipid profile, cardiovascular disease and mortality in a Mediterranean high-risk population: the ESCARVAL-RISK study. PLoS One 2017 Oct 18;12(10):e0186196 [FREE Full text] [doi: 10.1371/journal.pone.0186196] [Medline: 29045483]
56. Wang R, Yu X, Wang Z, Liu Y, Chen H, Liu S, et al. Blood lipid levels and all-cause mortality in older adults: the Chinese Longitudinal Healthy Longevity Survey 2008-2018. Epidemiol Health 2022 Jul 05;44:e2022054 [FREE Full text] [doi: 10.4178/epih.e2022054] [Medline: 35798025]
57. da Silva Ferreira G, Bochi AP, Pinto PR, Del Bianco V, Rodrigues LG, Morais MR, et al. Aerobic exercise training prevents insulin resistance and hepatic lipid accumulation in LDL receptor knockout mice chronically fed a low-sodium diet. Nutrients 2021 Jun 24;13(7):2174 [FREE Full text] [doi: 10.3390/nu13072174] [Medline: 34202724]

XSL•FO
RenderX

58. Pei Z, Ji J, Gao Y, Wang H, Wu Y, Yang J, et al. Exercise reduces hyperlipidemia-induced cardiac damage in apolipoprotein E-deficient mice via its effects against inflammation and oxidative stress. Sci Rep 2023 Jun 05;13(1):9134 [FREE Full text] [doi: 10.1038/s41598-023-36145-w] [Medline: 37277452]
59. Brites F, Martin M, Guillas I, Kontush A. Antioxidative activity of high-density lipoprotein (HDL): mechanistic insights into potential clinical benefit. BBA Clin 2017 Dec;8:66-77 [FREE Full text] [doi: 10.1016/j.bbacli.2017.07.002] [Medline: 28936395]
60. Franczyk B, Gluba-Brzózka A, Ciałkowska-Rysz A, Ławiński J, Rysz J. The impact of aerobic exercise on HDL quantity and quality: a narrative review. Int J Mol Sci 2023 Feb 28;24(5):4653 [FREE Full text] [doi: 10.3390/ijms24054653] [Medline: 36902082]
61. Ahmadi MN, Rezende LF, Ferrari G, Del Pozo Cruz B, Lee IM, Stamatakis E. Do the associations of daily steps with mortality and incident cardiovascular disease differ by sedentary time levels? A device-based cohort study. Br J Sports Med 2024 Mar 08;58(5):261-268 [FREE Full text] [doi: 10.1136/bjsports-2023-107221] [Medline: 38442950]

Abbreviations

BP: blood pressure
CAC: coronary artery calcium
HDL: high-density lipoprotein
LDL: low-density lipoprotein
MET: metabolic equivalent of task
MRI: magnetic resonance imaging
PAI: physical activity index

Edited by L Buis; submitted 12.01.25; peer-reviewed by A Woodward, M Fraser; comments to author 23.04.25; revised version received 15.05.25; accepted 02.09.25; published 24.09.25.

Please cite as:
Huang W, Wong MKF, Loh EDW, Koh T, Tan AW, Shen X, Varli O, Kong SC, Chin CWL, Tan SY, Yap JJL, Ng EYK, Yeo KK
Relationship Between Activity Tracker Metrics and the Physical Activity Index and Their Association With Cardiometabolic Phenotypes, Subclinical Atherosclerosis, and Cardiac Remodeling: Cross-Sectional Study
JMIR Mhealth Uhealth 2025;13:e71213
URL: https://mhealth.jmir.org/2025/1/e71213
doi:10.2196/71213
PMID:

©Weiting Huang, Mark Kei Fong Wong, Enver De Wei Loh, Tracy Koh, Alex Weixian Tan, Xiayan Shen, Onur Varli, Siew Ching Kong, Calvin Woon Loong Chin, Swee Yaw Tan, Jonathan Jiunn Liang Yap, Eddie Yin Kwee Ng, Khung Keong Yeo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Within- and Between-Individual Compliance in Mobile Health: Joint Modeling Approach to Nonrandom Missingness in an Intensive Longitudinal Observational Study

Young Won Cho[1], MA; Sy-Miin Chow[1], PhD; Jixin Li[2], MA; Wei-Lin Wang[3], PhD; Shirlene Wang[4], PhD; Linying Ji[5], PhD; Vernon M Chinchilli[6], PhD; Stephen S Intille[2], PhD; Genevieve Fridlund Dunton[7], PhD

[1]Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, United States

[2]Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States

[3]Department of Preventive Medicine, University of Southern California, Los Angeles, CA, United States

[4]The Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[5]Department of Psychology, Montana State University, Bozeman, MT, United States

[6]Department of Public Health Sciences, The Pennsylvania State University, Hershey, PA, United States

[7]Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Young Won Cho, MA
Department of Human Development and Family Studies
The Pennsylvania State University
119 Health and Human Development Building
221 E. College Avenue
University Park, PA, 16802
United States
Phone: 1 8148638000
Email: yfc5309@psu.edu

Abstract

Background: Missing data are inevitable in mobile health (mHealth) and ubiquitous health (uHealth) research and are often driven by distinct within- and between-person factors that influence compliance. Understanding these distinct mechanisms underlying nonresponse can inform strategies to improve compliance and strengthen the validity of inferences about health behaviors. However, current missing data handling techniques rarely disentangle these different sources of nonresponse, especially when data are missing not at random.

Objective: We demonstrate the usability of joint modeling in the mHealth context, showing how simultaneously accounting for the dynamics of health behavior and both within- and between-person missingness mechanisms can affect the validity of health behavior inferences. We also illustrate how joint modeling can inform distinct sources of (possibly nonignorable) missingness in studies using ecological momentary assessment and wearable devices. We provide a practical workflow for applying joint models to empirical data.

Methods: We applied joint modeling on empirical data comprising 1 year of daily smartphone-based ecological momentary assessment data (affect and energetic feeling) and smartwatch-tracked physical activity (PA). The approach combined (1) a multilevel vector autoregressive model for examining the reciprocal influences between daily affect and PA, and (2) a multilevel probit model for missingness. Unlike conventional 2-stage imputation methods—which first impute missing data before fitting the main model—joint modeling handles missingness during model fitting without explicit imputation. Sensitivity analyses compared results from the proposed method to other missing data approaches that do not explicitly model missingness. A simulation study designed to mirror the temporally clustered (eg, consecutive days of missing data) and person-specific missingness patterns of the empirical data validated the feasibility of the proposed approach.

Results: Sensitivity analysis indicated relative robustness of the autoregressive effects across missing data handling approaches, whereas cross-regressive effects could be detected only under the joint modeling but not with methods that did not simultaneously model missingness mechanisms. Specifically, under joint modeling approaches, participants had higher levels of PA on days following a previous day with higher self-report energy levels (95% credible interval [CrI] 0.012-0.049). Furthermore, the missing data model revealed both missing not at random and missing at random mechanisms. For example, lower PA predicted higher

XSL•FO
RenderX

missingness in PA at the within-person level (95% CrI –1.528 to –1.441). Being employed was associated with higher missingness in device-tracked PA at the between-person level (95% CrI 0.148-0.574). Finally, simulation showed that joint modeling could improve the accuracy of estimates and identify nonignorable missingness.

Conclusions: We recommend joint modeling with multilevel decomposition for addressing nonignorable missingness in mHealth/uHealth studies collecting intensive longitudinal data. We also suggest using a missing data model to explore the missingness mechanism and inform data collection strategies.

(JMIR Mhealth Uhealth 2025;13:e65350) doi:10.2196/65350

KEYWORDS

missing data; joint model; intensive longitudinal data; longitudinal data; mHealth; mobile health; smartwatch; digital health; digital intervention

Introduction

Background

As technology advances, mobile health (mHealth) and ubiquitous health (uHealth) data collection via mobile and wearable devices has become increasingly common. This typically involves continuous tracking of physiological data, such as step counts and heart rate, through smartwatches [1], alongside self-reported ecological momentary assessment (EMA) surveys prompted on smartphones or web interfaces multiple times a day to measure self-reported behavioral and psychological variables [2,3]. Such intensive longitudinal data (ILD) [4,5] offer numerous advantages. They facilitate remote monitoring for timely interventions, reduce recall bias by capturing real-time data, and allow for examination of behavioral changes on more granular timescales. These features have the potential to enhance the effectiveness of health monitoring and research. Missing data are inevitable in mHealth and uHealth studies as triggered by myriad reasons, including variations in participants' compliance, engagement, contextual factors, and other measurements and technical issues [6,7]. Whereas various modeling methods have been proposed to deal with missing data in ILD, few mHealth or uHealth studies have systematically leveraged and integrated passive device and subjective EMA responses to inform missingness mechanisms that drive across-participant and across-time variations in compliance rates and possible consequences on health behavior inference [7-10].

Why Joint Modeling for mHealth and uHealth Studies?

Overview

Joint modeling is a technique that handles nonresponse bias by simultaneously fitting 2 interconnected models: 1 for the main analysis (substantive model) and 1 for understanding how missing data might occur (missing data model). The substantive model could be any statistical model for testing the primary research question, such as linear regression, survival analysis, and path analysis. The missing data model is a model that examines whether missingness is linked to certain factors by modeling binary missing indicators of the variables of interest (see Figure 1 for an overview).

Joint modeling is commonly used to address nonignorable, also known as nonrandom missingness, which occurs when the missingness depends on unobserved outcomes. For example, if some participants wear their smartwatches only when they are physically active, the device-tracked physical activity (PA) data are said to be missing not at random (MNAR), as the likelihood of missingness depends on the unobserved PA levels, which ought to be measured but were missing on the less active days (alternatively, more germane scenarios include missingness completely at random [MCAR], in which the missingness does not depend on any variable, and missingness at random [MAR], in which the missingness depends only on variables that are observed in the measured data). Overlooking such nonignorable missingness by examining the substantive processes only with the available data often results in incorrect inferential conclusions [11].

Past mHealth and uHealth studies that sought to identify predictors of participant compliance [8,12] have typically focused exclusively on the "missing data models." As such, the researchers miss opportunities to allow the substantive and missing data models to iteratively inform each other during model fitting, especially in MNAR scenarios.

Such a joint inferential approach has several advantages. First, by explicitly accounting for when and why missingness occurs, the missing data model informs substantive models with more plausible values for unobserved data, leading to enhanced estimates of key substantive processes (such as PA and energy levels), even when data are incomplete. Second, the missing data model also benefits from the substantive model, especially when the likelihood of missingness depends on variations in the substantive processes (ie, MNAR), as noted in several EMA studies involving subjective self-reports [13,14].

Third, in mHealth or uHealth contexts, this mutual reinforcement between models can be further enhanced by leveraging passively collected data. Passive data such as device usage indicators (eg, app engagement records) can be collected unobtrusively and may provide valuable auxiliary information for modeling missingness mechanisms.

Fourth, the enhanced ability to detect nonignorable missingness can inform better data collection strategies. For instance, if the model reveals that PA data are more likely to be missing during low-activity days (ie, less inclined to wear devices on inactive days; MNAR), researchers might implement targeted strategies to improve data quality, such as sending adaptive reminders encouraging device use particularly during low-activity periods. While direct follow-up with nonrespondents is ideal [15], it is often costly or infeasible, making joint modeling a valuable

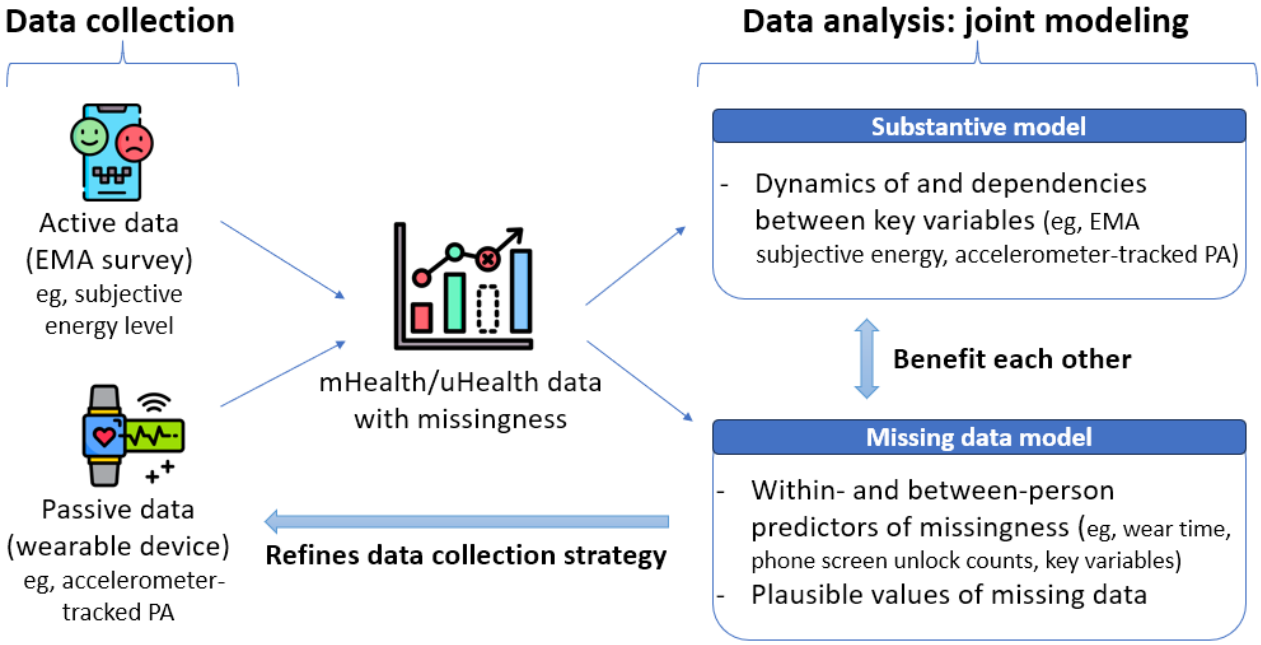
XSL•FO
RenderX

alternative for identifying potential MNAR without extra data collection.

Finally, even highly engaged participants exhibit occasional missingness in longitudinal studies, but no studies to date have considered the implications of conflating within- and between-participant variations in compliance. Examining how compliance fluctuations relate to PA and energy levels can inform the design of more sustainable health behavior interventions.

This study capitalizes on a 1-year intensive longitudinal study to demonstrate how device-tracked PA and subjective EMA reports (eg, energy levels) can be integrated within a joint modeling framework that captures both within- and between-person variations. Passive indicators such as cell phone unlock counts are incorporated as auxiliary variables to aid modeling missingness. This approach highlights the benefits of leveraging multiple data sources and the risks of ignoring key sources of variation.

Figure 1. Conceptual framework of joint modeling. EMA: ecological momentary assessment; mHealth: mobile health; PA: physical activity; uHealth: ubiquitous health.



Existing Joint Modeling Approaches

Three common approaches for handling missing data within the joint modeling framework are the selection model, shared parameter model (SPM), and pattern mixture model (PMM). We focus on the selection model and SPM, as PMM is less applicable to ILD context. PMM stratifies data by missing patterns and forms distinct models for each [16], which becomes impractical for ILD, given its diverse missing patterns.

In the selection model, we assume that the missingness depends on the unobserved values of the dependent variables (DVs). Thus, the selection model directly models the probability of missingness as a function of the (possibly unobserved) DV and uses a joint likelihood to link the substantive and missing data models [17]. For example, participants may be inclined to wear their smartwatches tracking PA only when they are physically active, leading to missing PA data when they are sedentary. To address this MNAR issue, selection models directly incorporate PA levels as a predictor in the PA missingness model, thereby correcting the bias caused by nonignorable missingness. Additionally, by integrating auxiliary variables, additional variables not part of the main analysis but providing valuable information about the missing data mechanism, as covariates [18] (eg, time of day as a proxy for habitual exercise routines), the model can also capture and adjust for MAR mechanisms. This approach enables the missing data model to serve as a tool not only for correcting bias but also for investigating potential contributors to the missingness, such as how PA levels influence the likelihood of PA data being missing.

In SPM, the missingness and substantive models are linked through shared latent parameters (eg, random effects) [19], and it assumes that the substantive and missingness processes are independent, given the shared or correlated random effects. For instance, in a longitudinal treatment study, patients may be more likely to drop out as their health deteriorates (MNAR). Here, the patient's latent health trajectory—represented by a random slope (ie, person-specific effects of time on health)—is associated with data missingness. By incorporating an individual's random slope as a predictor in the missing data model, 2 processes are linked [20]. When the missing data model's regression coefficient for the random slope is nonzero, it indicates dependence between the substantive and missingness processes.

While SPM and the selection model are often described as distinct approaches, SPM is sometimes viewed as a variant of the selection model [21,22]. Specifically, when applying a selection model within a multilevel framework, the structure closely resembles that of the SPM [22]. This similarity arises when the DV is decomposed into within- and between-person components, and the between-person component (ie, the random

XSL•FO
RenderX

intercept) is used as a predictor in the missing data model. In this case, it mirrors SPM's use of a latent random effect to link missingness and substantive processes.

These approaches form the conceptual basis of our proposed joint modeling, which specifically targets within- and between-person heterogeneity, as well as temporally clustered missingness—the tendency for missing days to be followed by more missing days. We describe each of these characteristics in turn.

Missing Data Issues in ILD

Heterogeneity in Missing Data Patterns Within and Across Individuals

Unlike conventional panel longitudinal data, ILD is characterized by a much higher frequency of assessments, leading to greater variability in missing data patterns. First, a high frequency of data collection can amplify within-person variations in compliance because the number of possible missing data patterns increases with an increase in the number of assessment occasions. Since data are collected daily or multiple days a time, more factors can influence daily compliance rates in mHealth and uHealth studies [8,9], such as daily variations in affect, technical issues, forgetting to wear a watch [23], or any disruptions in daily lives (eg, illness, vacations). Nahum-Shani et al [7] discussed how engagement in digital interventions using mobile devices can be influenced by contextual factors such as emotional, physical, and cognitive energy. As more concrete examples, nonresponse to EMA push notifications could be influenced by the emotional energy of the survey moment, as well as PA levels or sleep duration on the day, which is proxy of physical energy throughout the day.

Second, participants may miss assessment points due to baseline individual differences in compliance, such as levels of conscientiousness or familiarity with mobile technology. This between-person variation can become more pronounced with more repeated measures. Martinez et al [12] reported that compliance rates for wearable trackers and mobile EMA were related to person-specific characteristics. Older participants tended to have more difficulties using smartphones, leading to lower compliance rates [24]. A higher BMI was related to lower compliance rates in a weight intervention study because of physical discomfort with wearing the device and the collection of weight data. In summary, participants' compliance to mHealth and uHealth data collection changes not only because of their person-level traits but also due to changes in daily contextual factors. As sample size and measurement occasions increase, so does the variability in compliance rates both within and across individuals.

Our proposed joint modeling approach allows for between- and within-person heterogeneity in missingness by combining a multilevel missing data model with a multilevel vector autoregressive (VAR) model for the substantive processes of interest. Applications of such multilevel decomposition of missingness are still nascent in the literature and are primarily restricted to cross-sectional settings [25], or iterations through missing data imputation and substantive process modeling in 2 stages [23]. Although a few studies in the ILD literature have adopted joint modeling approaches (with substantive process models that were similar to that adopted here) [26], these studies did not apply multilevel decomposition to the missing data model itself.

Temporally Clustered Missingness

In our empirical data, we observed significant patterns of temporally clustered missingness (see Figure 2 for examples). Specifically, 25.2% (55/218) of the participants had at least 1 period of 7 consecutive days of missingness for EMA-measured energetic feelings (ENs), and 41.3% (90/218) of participants had at least 1 period of 7 consecutive days of missingness for smartwatch-tracked accelerometer-measured PA (see section "Motivating Empirical Data" for the valid data criterion). These patterns highlight a notable feature of missing data in ILD: burst or temporally clustered missingness, where 1 day of missingness tends to be followed by more missing days. This clustering can occur for various reasons. For instance, during busy periods or while traveling, participants may ignore survey prompts for several consecutive days. If any technical issues occur, the data can be continuously missing until the problem is resolved. Such burst missingness suggests that each instance of missing data may not be independent of the previous instances.

Incorporating this pattern into missingness models could improve the accuracy of analyses. However, existing methods often overlook these patterns, likely because it is not commonly observed in cross-sectional or conventional panel longitudinal data. To our knowledge, only 1 study has considered the temporal dependency of missingness when modeling missing processes but without substantive models [27]. In contrast, an ILD study with a joint modeling approach [26] did not incorporate such a temporal component in its missing data model. To address this gap, we integrate autoregressive (AR) terms into our proposed missing data model within the framework of the joint model. AR terms account for the autocorrelation in missing data by relating the current missing data status to the previous time points, thereby capturing the tendency for missingness to occur in clusters.

Our approach is a multilevel joint model based on the selection model framework, assuming the missingness in DV depends on unobserved variables or parameters (MNAR), and other observed covariates (thus allowing for MAR as a special case). Under this approach, substantive and missingness models are connected via joint likelihood and correlated residuals. While rooted in the selection model framework, our multilevel joint model shares some similarities with SPM: DVs are decomposed into within- and between-person components in substantive models, and the latent between-person components (ie, random intercepts) serve as predictors in missing data models. Our missing data model can be further extended from an SPM perspective by incorporating additional random effects (eg, random AR effects in VAR models). However, such an extension was not implemented, due to the lack of theoretical background to expect that person-specific AR coefficients would be systemically related to missingness. Table 1 compares the selection model, SPM, and our multilevel joint model, emphasizing their suitability for different scenarios and highlighting the unique contributions of our approach.

XSL•FO
RenderX

Figure 2. Data missingness in bursts from 2 participants. EMA: ecological momentary assessment; EN: energetic feeling; PA: physical activity.

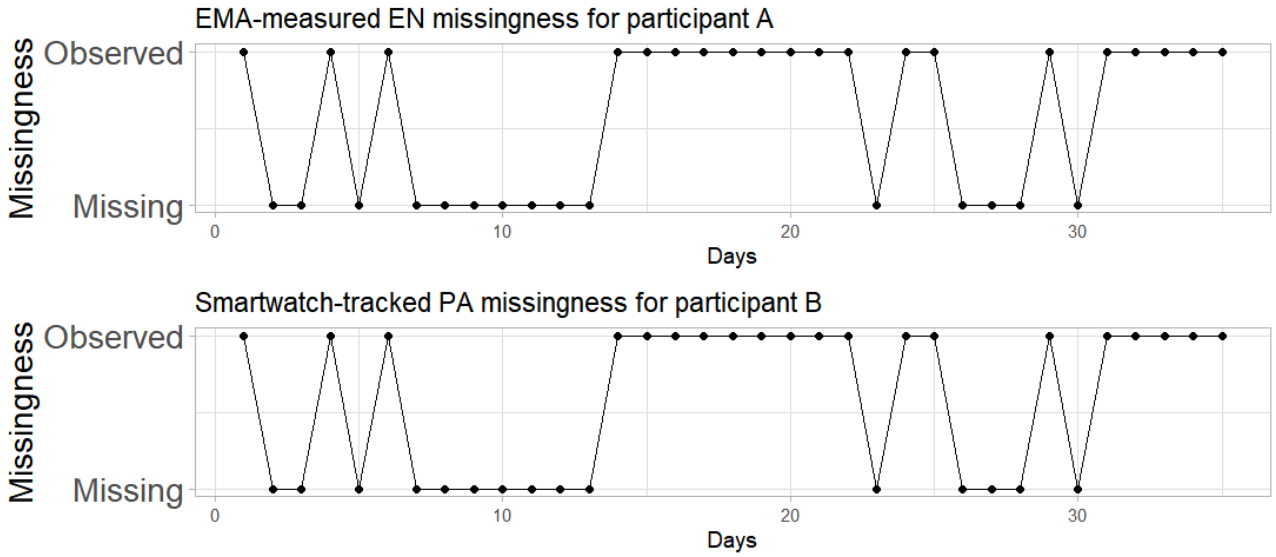


Table 1. Comparison of joint modeling missing data handling approaches.

	Key properties	When to use	Common use cases
Selection model	• Links outcome and missingness models via joint likelihood, often with correlated residuals.	When missingness depends on unobserved DVs[a].	• Widely applied in cross-sectional and panel data.
Shared parameter model	• Models share latent parameters (eg, random effects). • Assumes conditional independence given shared parameters.	When missingness relates to latent change components (eg, random slope).	• Typically to model dropout or attrition in longitudinal panel studies. • Terminal dropout or attrition does not require consideration of overtime variations in missingness.
Multilevel joint model (this study)	• Link 2 models via joint likelihood and correlated residuals. • Does not assume conditional independence given shared parameters. • Decomposes missingness into within- and between-person levels, disentangling momentary and habitual missing tendencies.	When missingness depends on either unobserved variable values or latent changing dynamics (eg, random AR[b] effects), with temporally clustered missing patterns.	• Tailored for ILD[c] with intermittent missingness. • Useful for ILD-specific challenges: mixed sources of missingness (person- and context-level), temporal dependency of missingness.

[a]DVs: dependent variables.

[b]AR: autoregressive.

[c]ILD: intensive longitudinal data.

This Study

Due to the distinct patterns of missingness in ILD, specialized approaches are required for handling missing data. In this study, we apply multilevel joint modeling to empirical mHealth data consisting of up to a year of daily self-reports on feeling energetic and PA, tracked by smartwatches. Especially, through this empirical analysis, we provide practical guidance on building missing data models and discuss the use of joint modeling of the dynamics of the substantive health processes and missingness mechanisms.

Additionally, by comparing the empirical results obtained across different missing data handling methods—such as listwise deletion, last observation carried forward (LOCF), mean imputation, Bayesian full information maximum likelihood (FIML), and joint modeling with alternative missingness models—we highlight the importance of conducting sensitivity analysis to evaluate the robustness of substantive conclusions.

Given that true substantive and missing data models are never known in empirical analysis, we further supplement the empirical analysis with a simulation study. This simulation aims to verify the efficacy of the joint modeling approach under known data-generating mechanisms and missing data patterns designed to mirror the intricacies of empirical ILD, such as individual variations in compliance rates, clustered patterns of missingness, and the presence of ordinal DV (eg, Likert scale).

Methods

Motivating Empirical Data

Data from the Temporal Influences on Movement and Exercise study [28], collected over 1 year, were analyzed. The study

XSL•FO
RenderX

recruited young adults (aged 18-30 years; N=246) across the United States. Participants completed smartphone EMA surveys on EN at the end of the day and wore smartwatches to track PA throughout their participation. Each participant's study start date was defined as the first day with both EMA and wearable data. If EMA responses continued after wearable data had permanently stopped (eg, due to technical issues), those trailing EMA entries were excluded to preserve meaningful variation in both affect and PA.

Data were retained for participants who had at least 14 days of nonmissing data for both EN and PA and at least 30 days of participation. Participants with more than 80% missing data in either EN or PA were considered noninformative for the intended analysis and excluded to prevent excessive imputation bias. This yielded final data from 218 out of 246 individuals, with data ranging from 39 to 401 (median 323.5, IQR 153.5-347) days and a total of 55,716 data points.

Measurement

Participants rated their daily EN on a 5-point Likert scale via smartphone app at the end of the day. Smartwatch-tracked PAs were first summarized using the motion independent movement summary units [29], based on raw acceleration data from Fossil Sport 4/5 (initially in 1-second epoch), and were aggregated daily. To facilitate efficient model fitting, daily PA was scaled by its standard deviation to align with other variables, as large discrepancies in variable scales can lead to numerical instability and slower convergence in complex models [30].

Time-varying covariates included daily device wear time, self-reported daily sleep duration (in hours), a season indicator, and phone screen unlock counts per day. Device wear time was determined from the SWaN (Sleep Wake and Nonwear) algorithm [28], which categorized available data into Wear, Sleep, or Nonwear. Demographic and between-person variables were also considered: sex (female=1 and male=0), age, employment status (employed for wages=1 and no=0), self-reported depression diagnosis (yes=1 and no=0), and alcohol consumption habit (number of days of drinks per week).

Missing Data

Individual missing data percentages for EN (the proportion of missing days relative to each participant's total participation) ranged from 0% to 71.2%, with an average of 16.7% per person. For the entire dataset, the missing rate was 14.9% (8319/55,716), with continuous missing days ranging from 1 to 38 (mean 1.6, SD 1.98 days). For daily PA, missing data were defined as less than 10 hours of smartwatch wear [31]. If no data were available, it was treated as 0 wear time, and PA was marked as missing. Individual missing data percentages for PA ranged from 0% to 68.5% (average=19.5% per person). For the entire dataset, the missing rate was 18.4% (10,265/55,716), with continuous missing days ranging from 1 to 100 (mean 2.04, SD 3.56 days). The median of the continuous missing days was 6, and the 90th percentile was 30, indicating that most of the gaps were short, with only 10% exceeding a month. Notably large gaps (eg, 100), upon reviewing research notes, were confirmed to be due to technical issues, with participants resuming after the issues were resolved.

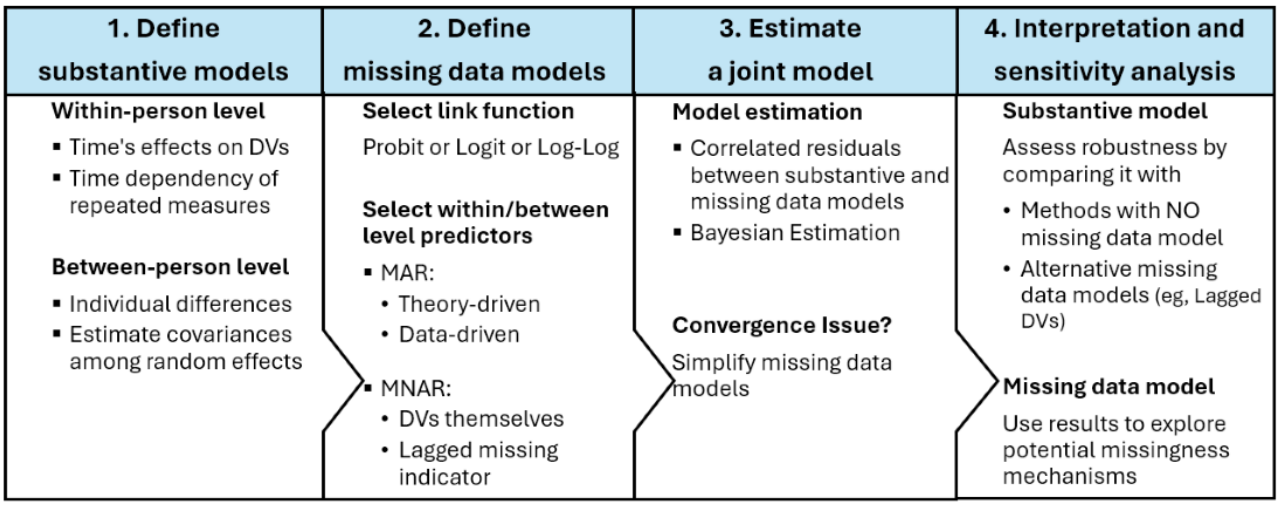
Ethical Considerations

This study analyzed fully anonymized secondary data obtained from a previous study [28] conducted at the University of Southern California (HS-18-00605). The original study was approved by the University of Southern California Institutional Review Board and conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All participants provided informed consent to have their deidentified data published in journals. As this study involved only anonymized secondary data and did not include direct participant contact, additional ethical review was not required.

Statistical Analysis

Figure 3 provides an overview of our analytic framework. It outlines the key steps for specifying the substantive and missing data models, estimating them jointly, conducting sensitivity analysis, and interpreting the results. The following subsections elaborate on each step of this workflow. Code for replicating empirical analysis is available in Multimedia Appendix 1.

Figure 3. Workflow of multilevel joint modeling for handling missing not at random intermittent missingness. DVs: dependent variables; MAR: missing at random; MNAR: missing not at random.



XSL•FO
RenderX

Substantive Model: Multilevel VAR (1) Model

Joint modeling involves 2 components: substantive and missing data models. The substantive model, driven by research questions, is a statistical model that tests research hypotheses. In our case, the research question focused on the reciprocal relationship between daily PA and EN over time. Specifically, we aimed to understand how the PA and EN from the previous day influenced those on the following day. This exploration of lead-lag relations can provide insights into how changes in one process may influence the other, potentially informing interventions to improve physical and mental well-being. For this, we used a VAR model as our substantive model. The VAR (1) model for EN ($EN_{i,t}$) and PA ($PA_{i,t}$) is described by the following equations. The subscripts i and t denote the values for ith individual at time t.

In these equations, $\mu_{1,i}$ and $\mu_{2,i}$ are individual baseline levels of EN and PA, when all predictors are at their average values. Individual variations in baseline were allowed and predicted by person-specific variables (sex, age, employment status, depression, and alcohol consumption habits), with EN baseline additionally predicted by person-average sleep duration, and PA baseline by person-average device wear time.

The coefficients $a_{1,i}$ and $a_{2,i}$ represent the AR effects for person i, namely, the extent of yesterday's energy (PA) on today's energy (PA), respectively. The coefficients, $b_{1,i}$ and $b_{2,i}$, termed cross-regressive (CR) coefficients, represent yesterday's PA → today's energy and yesterday's energy → today's PA effects for person i. Thus, temporal dependencies within individuals were explicitly accounted for through individual differences in AR and CR effects, which include the sample's average AR and CR coefficients (termed "fixed effects"), and stochastic parts that capture each person's deviations from the sample averages (termed "random effects"). All the random effects were allowed to covary with each other.

Daily sleep duration, a seasonal indicator, and device wear time were included as time-varying predictors. The terms $e_{1,i,t}$ and $e_{2,i,t}$ represent dynamic noises, or residuals in the 2 processes that were not accounted for by other variables, but whose effects carried on to future time points through the AR and CR terms. These dynamic noises were allowed to covary with each other.

Given the importance of modeling time effects in ILD analyses, we examined potential temporal trends in PA. We found no significant linear trend in PA over the year. However, to account for seasonal variations in PA, we included a season variable in the model [32]. Exploratory analysis showed that PA levels were higher in summer and lower in winter, with no significant difference between spring and fall. Thus, winter was coded as –1, spring or fall as 0, and summer as 1. It is worth noting that the specification of the substantive model and the omission or misestimation of covariances between random effects can impact parameter estimates, particularly in the presence of substantial missing data [33].

Missing Data Model: Multilevel Probit Model

Choice of Models for Binary Outcomes

We constructed a missing data model for each of 2 DVs (EN and PA). Each was constructed as a multilevel univariate probit regression model [27], with the binary missing indicators of whether a data point is missing (0=observed and 1=missing) on a particular day t for person i and construct d serving as the ordinal DV. Underlying the ordered binary responses for each construct were a normally distributed, time- and person-varying latent variable that represented each individual's underlying and continuous tendency missingness on a given day, with missingness occurring if this tendency exceeds an estimated threshold parameter.

While logit and log-log links are viable alternatives for modeling binary outcomes, we selected the probit model for both theoretical and practical reasons. First, although logit and probit models often yield similar results [34], theoretical considerations favor the probit model. Theoretically, the probit model aligns well with ILD missing data contexts, where a latent tendency for missingness at person level (habitual missing) and day level (short-term tendency) is plausible. Second, prior research suggests that in complex settings—such as multivariate binary response models with random effects—probit often provides better fit than logit [35]. Third, the log-log link is tailored to rare - event or survival - type dropout contexts, whereas our focus is on intermittent nonresponse. Finally, software constraints—such as Mplus's support for multilevel Bayesian probit but not logit—made probit the most feasible option in practice.

Across-Time and Person Heterogeneity in Missingness

Since missing data indicators were themselves longitudinal and nested within each person, the proposed multilevel probit regression model extended standard probit regression by decomposing the missingness mechanisms into day- and person-level missingness. The within-person part modeled day-to-day variations in missing tendency related to time-varying covariates, such as energy and PA levels themselves, and phone usage. The between-person part captures individual differences in overall missingness tendencies, linked to person-specific predictors (eg, person mean of phone usage) and demographic factors. Table 2 summarizes representative hypotheses about within- and between-person missingness mechanisms and shows how each is reflected in the multilevel missing data model.

XSL•FO
RenderX

Table 2. Plausible missingness-related hypotheses.

Levels and hypotheses	Reflecting missing data models	Mechanisms
Within-person		
On days when they feel less energetic than usual, they tend to miss more EMA[a] prompts.	EN missing$_{i,t}$ = β_1 × EN[b]$_{i,t}$	• MNAR[c]
On days when they engage in more physical activity than usual, they tend to miss more EMA prompts.	EN missing$_{i,t}$ = β_2 × PA[d]$_{i,t}$	• MAR[e], if PA missingness is not related to energy level. • MNAR, if PA missingness is related to energy level itself (eg, on low-energy days, people are less likely to wear smartwatches).
On days when they use smartphones more than usual, they tend to miss fewer EMA prompts.	EN missing[f]$_{i,t}$ = β_3 × Phone[g]$_{i,t}$	• MAR
Between-person		
People who have lower energy than others tends to miss more EMA prompts.	EN missing[h]$_i$ = β_5 × EN$_i$	• MNAR
People who are generally more physically active than others tend to miss more EMA prompts.	EN missing$_i$ = β_4 × PA$_i$	• MAR, if overall PA missingness is not related to overall energy level. • MNAR, if overall PA missingness is related to overall EN level (eg, those with generally lower energy are less likely to wear smartwatches).
People who generally use phones more than others tend to miss fewer EMA prompts.	EN missing$_i$ = β_6 × Phone$_i$	• MAR

[a]EMA: ecological momentary assessment.

[b]EN$_{i,t}$: energetic feelings. Subscripts *i* and *t* denote the values for i_{th} individual at time *t*.

[c]MNAR: missing not at random.

[d]PA$_{i,t}$: physical activity. Subscripts *i* and *t* denote the values for i_{th} individual at time *t*.

[e]MAR: missing at random.

[f]Missing$_{i,t}$: missing indicator at time *t* for person *i* (1=unobserved and 0=observed).

[g]Phone$_{i,t}$: the number of times the phone screen was unlocked. Subscripts *i* and *t* denote the values for i_{th} individual at time *t*.

[h]Missing$_i$: between-person level missingness for person *i*.

Selection of Predictors for Missing Data Models

Building missing data models requires identifying appropriate predictors of missingness, a process closely tied to hypothesis generation, as illustrated in Table 2. Theory-driven approaches use prior research to inform potential predictors of missingness or compliance. For example, age was related to lower compliance rates [24]. Daily emotional, physical, and cognitive energy may influence engagement in mHealth studies [7]. In our context, daily EN, PA, and sleep duration, which are proxies of daily energy levels, can be considered potential predictors of missingness. Also, smartphone usage had a significant impact on EMA response rates [8].

When theoretical guidance is limited, statistical exploration helps identify potential predictors of missingness, particularly under MAR assumptions. To inform the multilevel probit model, we conducted correlational analyses between missingness indicators and person-mean centered (within-person) and person-mean (between-person) variables. Predictors significantly associated with missingness at each level were selected for inclusion. The full list of predictors and their correlation results is reported in the “Results” section.

Burst Missingness

To capture burst patterns in missingness observed in the data, we included lagged (yesterday’s) latent missingness tendencies of yesterday’s missingness for EN and PA on today’s missingness for EN and PA in the missing data model. By incorporating temporal dependencies (ie, AR effects) of yesterday’s missingness for DVs on today’s missingness for DVs, the model allowed a participant’s missingness tendency on 1 day to influence missingness on the following day.

Estimation of the Joint Model

After constructing the substantive and missing data models, we jointly estimated them by linking them through joint likelihood and correlated residuals between the substantive and missing data models. For estimation, we used Bayesian methods due to its pragmatic advantages in handling complex models with multilevel structures [36-38]. Following practices for Bayesian reporting [39], we presented the median of each posterior distribution as the point estimate, alongside standard deviations and 95% credible intervals (CrIs). Parameters were considered significant if their CrI did not include zero.

Model estimation was conducted on a PC (Intel i7 CPU, 32GB RAM) and took 2 hours to converge. If there are too many

XSL•FO
RenderX

predictors in the missing data model, the run time may increase and convergence issues may arise. In such cases, simplifying the model by reducing the number of predictors or restricting the random effects may help. In our experience, fitting the missing data model alone without the substantive model often led to convergence issues. This highlighted 1 practical advantage of using the joint modeling approach to allow inclusion of values of substantive processes in the missing data and vice versa to facilitate the estimation process. Details of the default Bayesian estimation settings in Mplus and the convergence criteria are outlined in Multimedia Appendix 1.

Sensitivity Analysis

To assess sensitivity to different missing data methods, we compared our joint modeling result with other commonly used techniques in behavioral science, including listwise deletion, LOCF, mean imputation, and Bayesian FIML. Listwise deletion (complete case analysis) removes any time points with missing values, resulting in the greatest extent of data loss among the methods considered. LOCF fills in missing values by carrying forward the last observed value, assuming no change since the last entry. Mean imputation replaces missing values with the variable's overall mean, disregarding individual and temporal variation. FIML is a more sophisticated technique that leverages all available data, including both fully and partially observed patterns (where some variables are measured at a specific time point while others are missing). It is considered more robust than traditional methods, such as mean imputation, and effective under the MAR assumption [17]. Bayesian FIML is an approach that uses Bayesian inference to handle missing data similarly to FIML, leveraging the computational advantages of Bayesian estimation [40,41].

Additionally, given recent advancements in deep learning for time series imputation—particularly methods that explicitly model temporal dependencies—we conducted a sensitivity analysis using 2 methods that have shown promising results: SAITS (Self-Attention-based Imputation for Time Series [42]) and BRITS (Bidirectional Recurrent Imputation for Time Series [43]). SAITS is a transformer-based model that leverages self-attention mechanisms, while BRITS is based on bidirectional recurrent neural networks. We applied SAITS-base [42] and BRITS to impute missing values in our data before estimating the substantive model. Imputation was implemented using PyPOTS [44].

Notably, the methods discussed do not incorporate explicit missing data models that we can interpret. To further assess the sensitivity of our joint modeling approach, we examined the robustness of key conclusions under an alternative missing data model specification by extending the proposed missing data model to include additional lagged variables of the DVs as predictors.

Among the approaches mentioned, listwise deletion, LOCF, mean imputation, SAITS, and BRITS are 2-stage processes, where missing data are imputed first, followed by fitting the substantive model using the complete dataset. In contrast, Bayesian FIML and joint modeling are a 1-stage process that handles missing data while fitting the substantive model, without explicit data imputation process.

Results

In this section, we first present the results from our proposed joint modeling approach, which addresses the possibility of data being MNAR, through both the substantive model and the missing data model. We then report the results of sensitivity analyses, comparing our approach with alternative missing data handling techniques and evaluating the robustness of our findings under different specifications of the missing data model within the joint modeling framework.

Correlation Analysis for Specifying the Missing Data Model

To inform the multilevel probit model, we conducted correlational analyses between missingness indicators and both person-mean centered (within-person) and person-mean (between-person) variables. At the within-person level, daily phone usage (r=–0.065; P<.001) and sleep durations (r=0.021; P<.001) were correlated with EN missingness. Similarly, daily phone usage (r=–0.022; P<.001) and sleep durations (r=0.031; P<.001) were correlated with PA missingness. At the between-person level, average phone usage was negatively correlated with both EN (r=–0.255; P<.001) and PA (r=–0.14; P=.039). Average sleep duration was related to EN missingness (r=–0.136; P=.045). It should be noted that these correlational results are based on observed data only and may differ from results obtained under joint modeling; their purpose was to inform predictor selection rather than to test missingness mechanisms.

Based on the theoretical considerations and these correlational findings the final missing data models included the following predictors: at the within-person level, daily EN, PA, sleep, and phone usage, and at the between-person level, the person means of EN, PA, sleep, and phone usage, along with participants' sex, age, employment status, self-reported depression, and alcohol consumption habit.

Substantive Model Results

To examine the daily reciprocal relationship between PA and EN, the multilevel VAR (1) model was fitted jointly with the multilevel probit regression model to represent the missingness of 2 DVs. Results showed that higher than usual levels of PA or EN on 1 day were significantly associated with higher levels of the same variable on the following day, as indicated by positive AR coefficients (EN 0.226, 95% CrI 0.196-0.254; PA 0.081, 95% CrI 0.066-0.097). This suggests that both PA and EN exhibited a certain level of continuity over time. Regarding reciprocal influence—the focus of this analysis—CR coefficients indicated that higher energy levels on 1 day predicted higher PA levels the next day (b=0.03; 95% CrI 0.012-0.049), and vice versa: higher PA levels also predicted higher energy levels on the following day (b=0.03; 95% CrI 0.006-0.054).

Covariate effects were as follows. PA levels tended to increase during warmer seasons (b=0.031; 95% CrI 0.025-0.038). Individuals with self-reported depression exhibited generally lower EN levels (b=–0.411; 95% CrI–0.735 to –0.078). Individuals who were employed (b=0.380; 95% CrI 0.178-0.582) and those who reported higher frequencies of alcohol

XSL•FO
RenderX

consumption (*b*=0.069; 95% CrI 0.009-0.128) showed higher overall PA levels. Additionally, sex differences were observed in PA baseline, with females exhibiting higher baseline PA than males (*b*=0.219; 95% CrI 0.017-0.420).

Missing Data Model Results

In this section, we present results from missing data models to explore plausible hypotheses about missingness mechanisms. Full probit model results are shown in Table S1 in Multimedia Appendix 2. It is worth noting that the estimated missingness mechanisms here are not definitive; rather, they provide preliminary insights that warrant further discussion with experts in the field.

Missing Not at Random

The results indicated that the PA level was significantly associated with missingness in both within- and between-person levels of EN and PA. Lower than usual PA levels or overall lower PA levels compared with others were linked to more missingness in both EN and PA. Since PA was passively recorded via smartwatch, the observed negative association between same day PA and missingness likely reflects that more active days coincide with consistent device wear.

Unlike PA, EMA self-reported EN levels were not associated with EN missingness at either the within- or between-person level, indicating that individuals with lower energy—whether generally or on a given day—were not more likely to miss reporting. Interestingly, at the within-person level, missingness in smartwatch-tracked PA was higher on days they reported higher than usual EN. While unexpected, this result might reflect nuanced interactions between study and other unmeasured factors that were not accounted for by the model. Follow-up sensitivity analysis (see the "Sensitivity Analysis" section) showed that this result, while unexpected, was robust to changes in variations in the missing data models.

Missing at Random

As expected, the amount of phone usage was negatively associated with missingness in EMA-measured EN at both the within-person and between-person levels. More frequent phone usage than usual or compared with others was associated with less missingness in EN. However, phone usage was not related to the missingness of smartwatch-tracked PA. Self-reported sleep duration was not significantly related to missingness of either PA or EN.

We investigated the relationship between demographic variables (sex, age, employment status, self-reported depression, and alcohol consumption) and missingness in the data. First, we found that none of the demographic variables were significantly associated with missingness in EMA-measured EN. Notably, self-reported depression and age did not affect missingness. This may be because the participants were a nonclinical sample and they were all from generations familiar with smart devices, which could have minimized the impact of these variables on missingness. Conversely, people who were employed for wages showed higher missingness in smartwatch-tracked PA, suggesting that occupational demands may interfere with consistent device wear. We address the importance of considering job-related factors in influencing uHealth missing data in the "Discussion" section.

Burst Missingness

To account for burst missingness, we included the previous time point's missing indicators as predictors. A positive AR coefficient indicates temporally clustered missingness. In the EN missing data model, the AR coefficient was positive and significant (*b*=0.332; 95% CrI 0.312-0.352), suggesting that if a participant missed responding to EMA once, they were more likely to miss responding the next day as well. However, the AR coefficient in the PA missingness model was not positive (*b*=–0.011; 95% CrI –0.019 to –0.003), which will be further discussed in the "Sensitivity Analysis" section. The different AR patterns between EMA-measured EN and smartwatch-tracked PA imply that missing data patterns may differ between active self-report EMA measures and passively recorded smartwatch-tracked measures.

Sensitivity Analysis

We compared our findings from the proposed joint model (joint model 1) with other missing data handling techniques that do not explicitly incorporate missing data models: (1) listwise deletion, (2) LOCF, (3) mean imputation, (4) Bayesian FIML, (5) SAITS base, and (6) BRITS. Furthermore, to examine the sensitivity of our conclusions to the specification of the missing data model itself, we tested an alternative joint model (7), joint model 2, that extended joint model 1 by including an additional predictor—a lagged DV—into the missingness models.

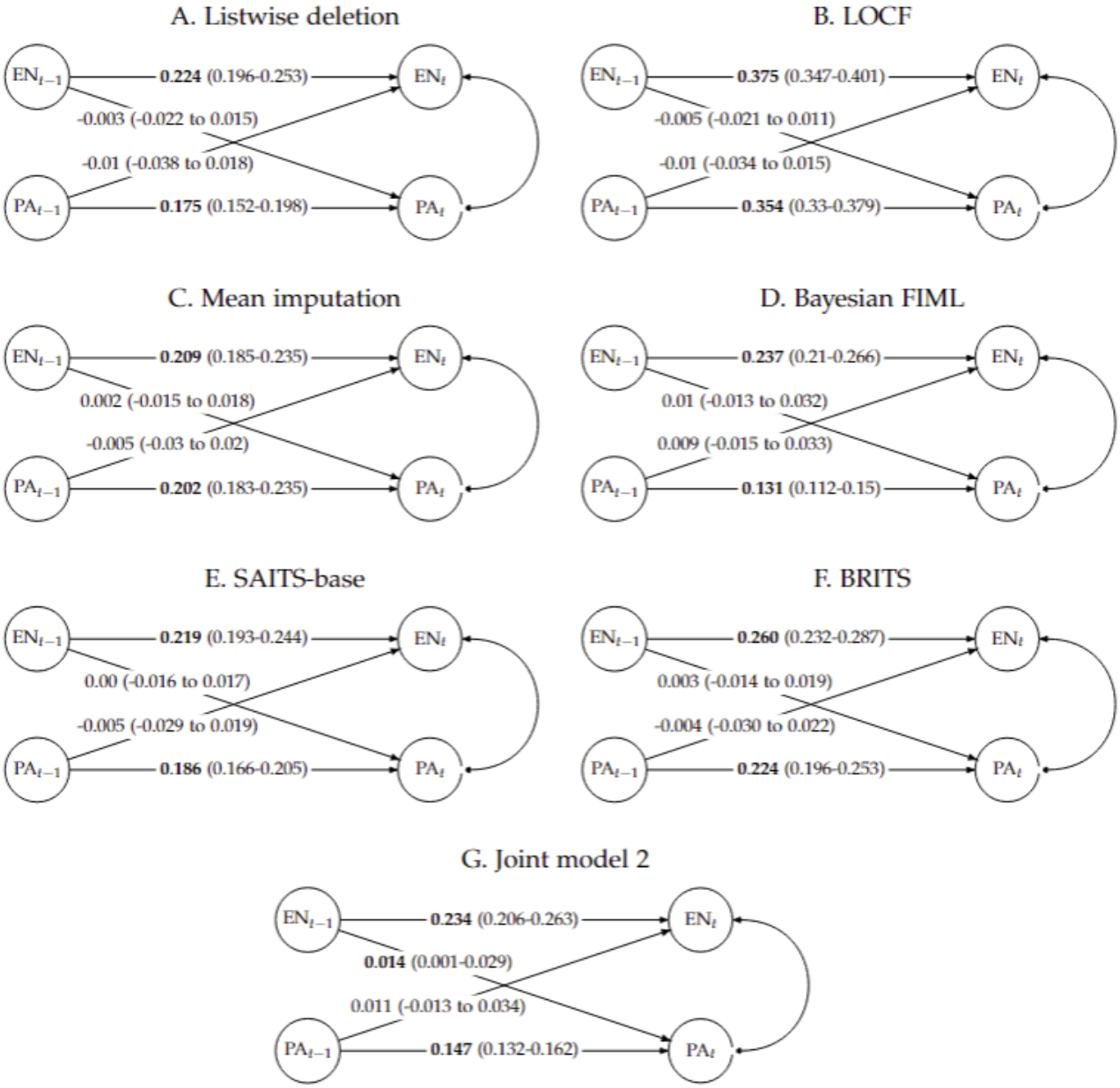
Substantive Model

AR Effect

Comparison of AR and CR effects—the key substantive model results—is summarized in Figure 4. For both PA and EN, significant positive AR coefficients were observed across all methods. The magnitudes were generally similar for EN across methods, except for LOCF, which overestimated AR parameters due to its characteristic of repeating the same value. Conversely, the AR estimates for PA were notably influenced by missing data handling methods. This appears to stem from the high incidence of MNAR missingness in PA, as discussed in the previous "Missing Data Model Result" section. As expected, AR coefficients were more similar between the Bayesian FIML and joint models than with other 2-step approaches.

XSL•FO
RenderX

Figure 4. Sensitivity analysis of autoregressive (AR) and cross-regressive (CR) parameters across missing data handling methods. Each node represents EN or PA at time t or t−1. Arrows show AR and CR paths with point estimates and 95% credible intervals (within parentheses). Labels in boldface indicate significance (95% credible interval excludes 0). (A) Listwise deletion, (B) LOCF, (C) mean imputation, (D) Bayesian FIML, (E) SAITS-base, (F) BRITS, and (G) joint model 2. BRITS: Bidirectional Recurrent Imputation for Time Series; EN: energetic feeling; FIML: full information maximum likelihood; LOCF: last observation carried forward; PA: physical activity; SAITS: Self-Attention-based Imputation for Time Series.



CR Effect

In contrast to AR effects, CR coefficients—representing the dynamic interplay between energy levels and PA—were much closer to 0, thus yielding inferential conclusions that were sensitive to changes in missing data handling methods (Figure 4). Across approaches without an explicit missing data model, in models (A) to (F), CR estimates were not statistically significant. However, when joint modeling was applied (joint models 1 and 2), a significant CR effect emerged. Similar to joint model 1, joint model 2 also showed that higher energy levels on one day predicted higher PA levels on the next, demonstrating the robustness of this finding. However, the reverse CR effect (from PA to energy) was not significant in joint model 2, unlike in joint model 1.

These findings illustrate how joint modeling can uncover dynamic associations that might be obscured when missing data are not explicitly modeled. They also show how the estimation of CR effects can be influenced by the specific construction of the missing data model, highlighting the importance of sensitivity analysis across missing data model specifications.

Person-Level Covariate Effect

We investigated whether the baseline levels of DVs are associated with person-specific characteristics such as sex, age, employment status, depression, and alcohol consumption habit. Overall, the estimates were consistent across different methods. One exception was the effect of sex on PA baseline: although estimated sex differences from other missing data handling methods were in the same direction (ie, positive), it was

XSL•FO
RenderX

significant only in 2 joint models and SAITS, where females exhibited a higher baseline PA (joint model 1: *b*=0.219, 95% CrI 0.017-0.42; joint model 2: *b*=0.226, 95% CrI 0.025-0.43; and SAITS: *b*=0.0.190, 95% CrI 0.017-0.353). This aligns with our simulation results, which are elaborated later, and indicates that while the estimates of person-level covariate effects on baseline levels are generally less sensitive than CR effects, some parameters can still be influenced by missing data handling methods.

Missing Data Model

The probit model results for joint model 2 are shown in Table S2 in Multimedia Appendix 2. Compared with joint model 1, here we address some robust findings and note key differences from joint model 2. Under joint model 2, the previous day's EMA-measured EN level was not significantly associated with EN missingness, and the overall estimation on EN missingness mechanisms remained unchanged. In contrast, the previous day's smartwatch-tracked PA level was positively associated with PA missingness (95% CrI 0.214-0.249), resulting in some shifts in parameter estimates— already noted in the "Substantive Model" subsection of "Sensitivity Analysis."

Missing Not at Random

Consistent with joint model 1, joint model 2 showed that lower smartwatch-tracked PA levels were associated with higher missingness in both EMA self-reported EN and device-tracked PA measures at the within- and between-person levels, supporting the robustness of this pattern. Additionally, both models revealed an unexpected within-person effect: on days when individuals reported higher than usual EN, there was more missingness in PA measures. While unexpected, this result remained robust across different model specifications. To investigate further, we fitted another joint model with a weekend indicator to predict outcome missingness. This revealed higher PA missingness on weekends (*b*=0.093, 95% CrI 0.068-0.119), but the other main results remained unchanged.

Missing at Random

A higher frequency of alcohol consumption was associated with more missingness in smartwatch-tracked PA in joint model 1, but not in joint model 2. While individuals who consume alcohol more frequently may experience forgetfulness or have lifestyle choices prioritizing social activities over research participation [45], this association appeared sensitive to the missing data model specification. Another less robust finding was a sex difference in smartwatch-tracked PA missingness: females showed higher missingness in joint model 2 but not in joint model 1. This suggests that certain between-person effects—alcohol consumption and sex—may be more model-dependent and should be interpreted with extra caution.

Burst Missingness

In EN missingness model, similar positive AR coefficients (0.332 in joint model 1 and 0.331 in joint model 2) suggest a robust pattern of burst EN missingness: participants who missed an EMA-measured EN response were more likely to miss the following day as well. However, the AR coefficient in the PA missingness in joint model 2 (*b*=0.031, 95% CrI 0.023-0.039) differed from joint model 1. This difference suggests that burst missingness in PA is partially driven by the PA levels themselves (ie, MNAR), as well as by other unexplained autocorrelated processes. In joint model 2, we included lagged PA as a predictor, allowing the model to attribute part of the burst missingness to the autocorrelated nature of PA. As a result, the AR term—which previously captured both structured (ie, MNAR-related) and unstructured burst missingness—changed, now reflecting primarily the unstructured component of burst missingness that is unrelated to PA levels.

Simulation Study Results

Simulated Substantive Model Results

To evaluate the consequences of omitting components or the entire missing data model, we conducted a simulation study designed to closely mirror the structure and characteristics of our empirical data (see Multimedia Appendix 3 for full simulation design). By anchoring the data generation on true parameter values, simulation lets us quantify bias when omitting specific components of the missing data model or entire missing data models.

Briefly, we generated complete longitudinal datasets for 200 participants over 300 days using a 2-outcome multilevel VAR (1) model, with known AR, CR, and covariate effects. Parameter values were set to approximate those estimated in the empirical analysis. We then introduced 30% missingness on both outcomes and a time-varying covariate under missing completely at random (MCAR), MAR, and MNAR mechanisms—incorporating within- and between-person predictors, individually different overall compliance rates, and burst missingness. Finally, we compared 5 analysis strategies (complete data, Bayes-FIML, and 3 joint model variants) across 100 replications. Results are summarized graphically in Figure S1 in Multimedia Appendix 3.

When There Were No Missing Data

Fitting the substantive model to fully observed data ("full data" condition in Figure S1 in Multimedia Appendix 3) provides a benchmark for other models. Under this ideal condition, estimations were unbiased: all closely approximate the true values. Only the covariate effects on AR were slightly underestimated, within an acceptable range.

Complete Omission of Missing Data Model

What happens if missing data models are not included at all? When missing data were handled solely through Bayesian FIML, estimates were severely biased across all parameters compared with models with explicit missing data models (see Bayes FIML condition in Figure S1 in Multimedia Appendix 3). These findings underscore the importance of incorporating dedicated missing data models, especially with MNAR.

Joint Modeling: When the Missing Data Model Was Included

When results were obtained using our proposed joint model (joint model 1; Joint DV condition in Figure S1 in Multimedia Appendix 3), which included missingness models for 2 DVs, estimation results from the joint model were almost comparable to fitting models to fully observed data despite the use of MNAR

XSL•FO
RenderX

data. Notably, despite being small (0.05), the CR coefficients were accurately estimated.

Adding a Missing Data Model for Time-Varying Covariates

In many cases, time-varying covariates also have missingness. Should we incorporate missingness models for them, too? The short answer is: when time-varying covariates have MNAR missingness, proper missing data models can improve estimation. However, compared with joint models that accounted for missingness only in DVs, adding missing data models for covariates did not always lead to substantial changes in statistical inference. Therefore, a pragmatic approach may be to include missing data models for the DVs under thorough explorations and as guided by theories but incorporate simpler missing data handling methods on the time-varying covariates (eg, partial FIML proposed by Ji et al [26]).

When Clustered Missingness Was Not Accounted for

What happens when missingness is temporally clustered but AR terms to capture the burst missingness are not included in the missing data model? Our simulation shows that while omitting AR terms is not as detrimental as excluding the missing data model entirely, it can still lead to subtle but meaningful bias in certain parts of the substantive model. For detailed results and more discussion, see Multimedia Appendix 3.

Simulated Missing Data Model Results

The missing data model estimates under the "Joint DV" condition are shown in Table S1 in Multimedia Appendix 3. Overall, the results suggested that all parameters from the multilevel probit model were well recovered under the joint modeling framework. At the within-person level, all coefficients were estimated accurately, regardless of whether the missing mechanism was MAR or MNAR, with the absolute relative bias ranging from 0.0 to 0.08—well below the commonly accepted threshold of 0.1. This demonstrated the probit model's adeptness in identifying both MAR and MNAR. Also, the AR coefficients were accurately estimated, indicating that the model can capture clustered missing patterns if they existed. At the between-person level, however, the coefficients were generally overestimated, with relative biases falling within a range typically considered unacceptable (0.1). Despite these biases, the coefficients still correctly captured the directions and approximate magnitudes of the true values. Thus, while person-level coefficients may lack some accuracy, they remain practically meaningful in terms of direction and magnitude.

Therefore, when used in conjunction with an appropriate substantive model, the proposed missing data model can offer useful insights into missing data mechanisms, even for MNAR scenarios. While testing for MNAR is widely recognized as challenging, using joint models may help in gaining a better understanding of these mechanisms. However, when applied to empirical data, caution is always necessary, as the true model for both the substantive and missing processes is unknown.

Discussion

Principal Results

We illustrated the process of fitting a multilevel VAR model to ILD, addressing potential MNAR missingness using a multilevel joint modeling framework. Our approach extends prior selection models to specifically handle intermittent missingness in ILD by incorporating within- and between-person decomposition and by capturing the temporal dependencies in the missing data model. We further conducted a sensitivity analysis comparing this approach with commonly used missing data methods, a deep leaning–based imputation technique, and alternative specifications of the missing data model—thus evaluating the robustness of findings on both substantive behavioral processes and underlying missing mechanisms.

We proposed examining the plausible missingness mechanisms through the missing data model. These insights are valuable for future mHealth and uHealth designs, as identifying patterns in missingness under certain conditions can help reduce such occurrences or address technological challenges that may disproportionately affect specific subgroups. Specifically, our analysis consistently revealed, across different missing data model specifications, that missingness in smartwatch-tracked PA and EMA-measured EN is not random but rather MNAR. That is, lower than usual PA levels, or overall lower PA levels, were associated with increased missingness in both PA and EN. This suggests that people are more likely to have missing data when they are less active, possibly because they are less mindful of wearing tracking devices during low activity. This also aligns with prior discussions in the study by Nahum-Shani et al [7], which suggest that insufficient physical energy levels may contribute to higher noncompliance rates in responding to EMA prompts or wearing wearable devices. Interestingly, yesterday's high PA was linked to more missingness in PA today. This could be due to factors such as smartwatch battery issues—where a day of high PA leads to excessive battery consumption, leaving the smartwatch uncharged the next day and resulting in missing data. Alternatively, it could be a pattern where 1 day of high PA leads to a PA drop the following day, which in turn increases the likelihood of not wearing the smartwatch. While the PA process exhibited an AR pattern in the overall sample, excessive PA on one day can reduce the next day's PA [46]. Confirming these potential reasons would require follow-up interviews or further investigation.

Additionally, employed individuals showed higher PA missing rates, suggesting that employment status may influence mHealth and uHealth data compliance. Job-related factors, such as demanding schedules, limited time to resolve technical issues, and stress, could affect missingness. For example, based on our research note, issues such as device malfunctions or charger problems often require contact with research staff for resolution. However, individuals with regular jobs may have fewer flexible schedules, making it difficult to address these issues. Additionally, Martinez et al [12] found that individuals in supervisory roles, who experience higher busyness, exhibited lower compliance. Researchers may need to consider these factors when designing uHealth studies.

XSL•FO
RenderX

Beyond identifying missingness mechanisms, we also examined how accounting for these mechanisms through joint modeling influenced the estimation of substantive behavioral inferences. Our substantive model included 2 key components: AR effects, capturing temporal stability, and CR effects, capturing the reciprocal influences between PA and EN. While AR effects remained consistent across different missing data handling methods, CR effects were significant only when using the joint modeling approach. The CR effects, although relatively small, reflected the reciprocal dynamics between variables and were of central interest in studies of health behaviors. Thus, this finding highlights the value of joint modeling and the importance of sensitivity analysis in assessing the robustness of parameter estimates in the presence of MNAR. Importantly, follow-up sensitivity analysis using alternative missing data model specification suggested that only the CR from EN to PA was robust to variations in the missing data model. The opposite CR effect (from PA to EN) was not. This underscored the need for sensitivity analyses with alternative missing data model specifications, even when using advanced missing data techniques such as joint modeling.

Nonetheless, our simulation study suggested that the joint modeling approach provided more reliable estimates than other methods under MNAR conditions. The proposed joint model yielded results comparable with those from fully observed data despite MNAR data. In contrast, simply fitting the substantive model alone, without explicitly modeling missingness, led to overall biased estimates. CR effects were particularly vulnerable due to their typically smaller magnitude than AR effects, resulting in low power (around 30%) when MNAR was not dealt with in the model. This aligned with our empirical results, where the CR effects were significant only in the joint models.

While joint modeling offers a principled approach grounded in theory, there has been a growing interest in mHealth and uHealth studies to use machine learning (eg, deep learning–based time series imputations) to handle missing data. Most of these approaches, including the approaches considered in our sensitivity analysis, typically handle the missingness and substantive model fitting as a 2-step process (imputation followed by model fitting), thus missing the opportunity to leverage the interdependence between the substantive and missing data models. Also, deep learning models often lack interpretability of conventional statistical models, making it difficult to explore missing mechanisms. They also require large sample sizes and extensive hyperparameter tuning, which can be a barrier in some studies. This may explain why SAITS and BRITS, in our sensitivity analysis, performed similarly to methods without explicit missing data models. While these models hold potential, their current performance may be limited by their assumptions about MCAR [42], the need for further hyperparameter tuning and larger samples, which may require additional refitment for behavioral data contexts. Nevertheless, the proposed joint framework can, in principle, be extended to deep learning architectures—allowing for simultaneous modeling for the substantive and missingness processes and thus “sharing” of information from these different data streams to strengthen the overall prediction as well as classification results.

Limitations

In this study, we leveraged model-implied values for the substantive processes afforded by the multilevel VAR model to shed light on the missing data processes. In practice, there is no guarantee that the multilevel VAR model reflects the true change mechanisms of the substantive processes. Thus, results from modeling, including those from the missing data model, should always be interpreted with some caution. Nevertheless, our view is that the multilevel VAR model, as one of the most commonly adopted models for describing ILD processes [47], serves as a useful proxy for describing the unfolding of PA and affective processes in the study participants.

Our proposed missing data model was also limited in that we did not consider participant attrition. The missing data model used in our example mainly addressed intermittent missingness. To address dropout as well, a multinomial logit model [11] may be used as an extension to our proposed model to categorize participants’ missingness tendency into 3 nominal categories: observed, intermittent missing, and dropout, thereby providing a more comprehensive approach to handling missing data in longitudinal studies.

Additionally, our analysis was based on daily aggregated data and did not delve into specific reasons behind the missing accelerometer data, such as data transfer issues, battery drain, or the watch being powered off. Instead, we focused on the broader issue of missingness, emphasizing psychological, physical, and individual factors. Exploring reasons related to technical issues in depth would require analysis at a finer granularity, such as minute-by-minute or second-by-second data. However, this level of detail was beyond the scope of our study.

As a future direction aligned with a finer timescale analysis, it would also be beneficial to differentiate between partial missing data (eg, 2-3 hours of device wear time less than 10 hours) and complete missing data (0 hours of wear time) for PA, as there may be qualitative differences between these scenarios. Similarly, while this paper focuses on day level, a multilevel joint modeling approach could be more advantageous in studies where EMA items are measured multiple times a day. For example, the relationship between PA and EMA missingness might appear weak or inconsistent at the day level, but at a finer scale, individuals who engage in high levels of PA during EMA prompts may be less likely to respond to those prompts [9]. Finally, we note that the results of the missing data model may depend on the choice of PA measure. Notably, there is no universally accepted convention for measuring PA [48]. Researchers who measured PA using accelerometers have constructed PA metrics in several ways, including (1) the total volume of activity counts or acceleration signals accumulated throughout a specified time interval (eg, day) [49]—similar to total daily step count; (2) the mean activity counts or signals per hour or minute, which offer an average measure of activity intensity [50]; and (3) the duration of time spent in PA categorized by intensity thresholds, such as moderate to vigorous PA [51]. Depending on the PA measure used, the definition of missing or valid data can vary, leading to potentially divergent model estimation results [52].

XSL•FO
RenderX

Additionally, PA data are often positively skewed, with a concentration of data at the lower end of the scale [53]. This deviation from normality may pose a potential limitation, as joint modeling approaches typically rely on the normality assumption [15]. While addressing this limitation is beyond the scope of this study, future research should consider modeling strategies that accommodate nonnormal data distribution, such as transformation or using a gamma link function [53].

Conclusions

This study enriches our understanding of missing data mechanisms and strategies for managing them in ILD. Key contributions include the following: First, incorporating missing data mechanisms into the model not only improves the properties of the estimation results but also provides a nuanced understanding of missing data mechanisms. Understanding the reasons for missing data holds significant value for several reasons. If the reasons are within the researcher's control, it can provide insights into how to collect data in future studies. Detecting contextual factors related to missingness can be used to pinpoint ways to improve participants' engagement. Additionally, understanding the mechanisms behind missing data can sometimes provide deeper insights into the variables themselves, thus hastening researchers' and practitioners' understanding of the phenomenon under study.

Second, we highlight the importance of sensitivity analysis. Evaluating the sensitivity and robustness of specific model aspects in relation to missing data provides crucial insights for informed medical decisions based on model outcomes. If the results exhibit sensitivity to methodological changes, it necessitates a thoughtful interpretation of the conclusions regarding their effects and suggests a potential need for further research.

Third, we demonstrated the effectiveness of our approach using simulated data that closely resemble actual datasets. We showed that our multilevel joint modeling is effective even with ordinal variables, such as Likert scale variables, and could reveal even strong MNAR mechanisms, as evidenced by the large coefficients observed in the empirical results. Furthermore, AR terms in the missing data model were able to capture frequently observed temporal dependency in missingness in ILD. However, it is important to note that the outcome of joint modeling may differ based on the chosen missing data model specification and still can give biased results if the specified model is not correct. Therefore, we strongly recommend performing sensitivity analysis with multiple missing data models to ensure the robustness of findings and avoid overreliance on a single model.

In summary, our findings highlight the importance of appropriately handling nonignorable missing data, particularly through multilevel joint modeling, followed by sensitivity analysis. This workflow will improve the validity of substantive conclusions and enhance our understanding of the underlying missing data mechanisms, which can inform better study designs and intervention strategies. While our analysis focused on a limited set of plausible variables, researchers would benefit from exploring a broader array of the psychological and behavioral factors that may influence compliance in mHealth studies. We hope this study serves as a framework for future investigations.

Acknowledgments

This research was funded by the National Institutes of Health Intensive Longitudinal Health Behavior Cooperative Agreement Program under U24AA027684, the National Center for Advancing Translational Sciences under UL1TR002014-06, the National Institute of Diabetes and Digestive and Kidney Diseases under U01DK135126, and the National Institute of General Medical Sciences of the National Institutes of Health under P20GM103474. Figure 1 was designed using image resources made by Freepik, srip, and Uniconlabs from www.flaticon.com.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Code for empirical analysis.
[DOCX File , 19 KB - mhealth_v13i1e65350_app1.docx]

Multimedia Appendix 2
Missing data model results.
[DOCX File , 25 KB - mhealth_v13i1e65350_app2.docx]

Multimedia Appendix 3
Simulation study materials.
[DOCX File , 187 KB - mhealth_v13i1e65350_app3.docx]

References

XSL•FO
RenderX

1. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
2. Dunton GF, Dzubur E, Intille S. performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity. J Med Internet Res 2016 Jun 01;18(6):e106 [FREE Full text] [doi: 10.2196/jmir.5398] [Medline: 27251313]
3. Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L. delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR Mhealth Uhealth 2021 Nov 19;9(11):e22890 [FREE Full text] [doi: 10.2196/22890] [Medline: 34806995]
4. Dunton GF, Rothman AJ, Leventhal AM, Intille SS. How intensive longitudinal data can stimulate advances in health behavior maintenance theories and interventions. Transl Behav Med 2021 Feb 11;11(1):281-286 [FREE Full text] [doi: 10.1093/tbm/ibz165] [Medline: 31731290]
5. Gates K, Chow S, Molenaar P. Intensive Longitudinal Analysis of Human Processes. First edition. Boca Raton, FL: Chapman & Hall/CRC Press; 2023.
6. Wang S, Yang C, Brown D, Cheng A, Kwan MYW. compliance with ecological momentary assessment in movement behavior research among adolescents and emerging adults: systematic review. JMIR Mhealth Uhealth 2025 Feb 11;13(1):e52887. [doi: 10.2196/52887]
7. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Mar 17;77(7):836-852. [doi: 10.1037/amp0000983] [Medline: 35298199]
8. van Berkel N, Goncalves J, Hosio S, Sarsenbayeva Z, Velloso E, Kostakos V. Overcoming compliance bias in self-report studies: a cross-study analysis. Int J Hum-Comput Stud 2020 Feb;134:1-12. [doi: 10.1016/j.ijhcs.2019.10.003]
9. Ponnada A, Li J, Wang SD, Wang W, Dunton GF, Intille SS. biases in microinteraction ecological momentary assessment (µEMA) non-response. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022 Mar;6(1):26. [doi: 10.1145/3517259] [Medline: 39866712]
10. Stone AA, Schneider S, Smyth JM. Evaluation of pressing issues in ecological momentary assessment. Annu Rev Clin Psychol 2023 May 09;19:107-131 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080921-083128] [Medline: 36475718]
11. Yuan C, Hedeker D, Mermelstein R, Xie H. A tractable method to account for high-dimensional nonignorable missing data in intensive longitudinal data. Stat Med 2020 Sep 10;39(20):2589-2605 [FREE Full text] [doi: 10.1002/sim.8560] [Medline: 32367549]
12. Martinez GJ, Mattingly SM, Robles-Granda P, Saha K, Sirigiri A, Young J, et al. Predicting participant compliance with fitness tracker wearing and ecological momentary assessment protocols in information workers: observational study. JMIR Mhealth Uhealth 2021 Nov 12;9(11):e22218 [FREE Full text] [doi: 10.2196/22218] [Medline: 34766911]
13. Li Y, Oravecz Z, Ji L, Chow S. multiple imputation with factor scores: a practical approach for handling simultaneous missingness across items in longitudinal designs. Multivariate Behav Res 2024 Jul 12;60(1):61-89. [doi: 10.1080/00273171.2024.2371816]
14. Tang N, Chow S, Ibrahim JG, Zhu H. Bayesian sensitivity analysis of a nonlinear dynamic factor analysis model with nonparametric prior and possible nonignorable missingness. Psychometrika 2017 Dec;82(4):875-903. [doi: 10.1007/s11336-017-9587-4] [Medline: 29030749]
15. Little R, Rubin D. Statistical Analysis With Missing Data. 2nd Edition. Hoboken, NJ: Wiley; 2002.
16. Little RJA. Pattern-mixture models for multivariate incomplete data with covariates. J Am Stat Assoc 1993 Mar;88(421):125-134 [FREE Full text]
17. Enders C. Applied Missing Data Analysis. Second edition. New York, NY: Guilford Press; Aug 2022.
18. Savalei V, Bentler PM. A two-stage approach to missing data: theory and application to auxiliary variables. Struct Equation Model Multidisciplinary J 2009 Jul 14;16(3):477-497. [doi: 10.1080/10705510903008238]
19. Albert PS. Shared random parameter models: a legacy of the biostatistics program at the National Heart, Lung, and Blood Institute. Stat Med 2019 Feb 20;38(4):501-511. [doi: 10.1002/sim.8011]
20. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics 1988 Mar;44(1):175-188. [doi: 10.2307/2531905] [Medline: 2486189]
21. Enders CK. Missing data: an update on the state of the art. Psychol Methods 2023 Mar 16:322-339. [doi: 10.1037/met0000563] [Medline: 36931827]
22. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. J Am Stat Assoc 1995 Sep;90(431):1112-1121. [doi: 10.1080/01621459.1995.10476615]
23. Yue Xu S, Nelson S, Kerr J, Godbole S, Patterson R, Merchant G, et al. Statistical approaches to account for missing values in accelerometer data: applications to modeling physical activity. Stat Methods Med Res 2018 Apr;27(4):1168-1186. [doi: 10.1177/0962280216657119] [Medline: 27405327]
24. Orstad SL, Gerchow L, Patel NR, Reddy M, Hernandez C, Wilson DK, et al. Defining valid activity monitor data: a multimethod analysis of weight-loss intervention participants' barriers to wear and first 100 days of physical activity. Informatics (MDPI) 2021 Jun;8(2):39 [FREE Full text] [doi: 10.3390/informatics8020039] [Medline: 36530339]

XSL•FO
RenderX

25. Yucel RM. Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. Phil Trans R Soc A 2008 Jul 13;366(1874):2389-2403. [doi: 10.1098/rsta.2008.0038]
26. Ji L, Chen M, Oravecz Z, Cummings EM, Lu Z, Chow S. A Bayesian vector autoregressive model with nonignorable missingness in dependent variables and covariates: development, evaluation, and application to family processes. Struct Equ Modeling 2020 May;27(3):442-467. [doi: 10.1080/10705511.2019.1623681] [Medline: 32601517]
27. McNeish D, Somers JA, Savord A. Dynamic structural equation models with binary and ordinal outcomes in Mplus. Behav Res Methods 2024 Mar;56(3):1506-1532 [FREE Full text] [doi: 10.3758/s13428-023-02107-3] [Medline: 37118647]
28. Wang S, Intille S, Ponnada A, Do B, Rothman A, Dunton G. Investigating microtemporal processes underlying health behavior adoption and maintenance: protocol for an intensive longitudinal observational study. JMIR Res Protoc 2022 Jul 14;11(7):e36666 [FREE Full text] [doi: 10.2196/36666] [Medline: 35834296]
29. John D, Tang Q, Albinali F, Intille S. An open-source monitor-independent movement summary for accelerometer data processing. J Meas Phys Behav 2019 Dec;2(4):268-281 [FREE Full text] [doi: 10.1123/jmpb.2018-0068] [Medline: 34308270]
30. Gelman A. Parameterization and Bayesian modeling. J Am Stat Assoc 2004 Jun;99(466):537-545. [doi: 10.1198/016214504000000458]
31. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
32. Garriga A, Sempere-Rubio N, Molina-Prados MJ, Faubel R. Impact of seasonality on physical activity: a systematic review. Int J Environ Res Public Health 2021 Dec 21;19(1):2 [FREE Full text] [doi: 10.3390/ijerph19010002] [Medline: 35010262]
33. Thomadakis C, Meligkotsidou L, Pantazis N, Touloumi G. Misspecifying the covariance structure in a linear mixed model under MAR drop‐out. Stat Med 2020 May 25;39(23):3027-3041. [doi: 10.1002/sim.8589]
34. Chen G, Tsurumi H. Probit and logit model selection. Commun Stat Theory Methods 2011 Jan;40(1):159-175. [doi: 10.1080/03610920903377799]
35. Hahn E, Soyer R. Probit versus logit models: the role of the link function in the multivariate realm. Technical Report TR-2008-17. The Institute for Integrating Statistics in Decision Sciences, The George Washington University. 2008. URL: https://business.gwu.edu/sites/g/files/zaxdzs5326/files/downloads/TR-2008-17.pdf [accessed 2025-08-12]
36. Cho YW, Chow SM, Marini CM, Martire LM. Multilevel latent differential structural equation model with short time series and time-varying covariates: a comparison of frequentist and Bayesian estimators. Multivariate Behav Res 2024;59(5):934-956 [FREE Full text] [doi: 10.1080/00273171.2024.2347959] [Medline: 38821115]
37. Oravecz Z, Tuerlinckx F, Vandekerckhove J. A hierarchical latent stochastic differential equation model for affective dynamics. Psychol Methods 2011 Dec;16(4):468-490. [doi: 10.1037/a0024375] [Medline: 21823796]
38. Chen M, Chow S, Oravecz Z, Ferrer E. Fitting Bayesian stochastic differential equation models with mixed effects through a filtering approach. Multivariate Behav Res 2023 Feb 27;58(5):1014-1038. [doi: 10.1080/00273171.2023.2171354]
39. Kruschke JK. Bayesian analysis reporting guidelines. Nat Hum Behav 2021 Oct;5(10):1282-1291 [FREE Full text] [doi: 10.1038/s41562-021-01177-7] [Medline: 34400814]
40. Hamaker EL, Asparouhov T, Brose A, Schmiedek F, Muthén B. At the frontiers of modeling intensive longitudinal data: dynamic structural equation models for the affective measurements from the COGITO study. Multivariate Behav Res 2018;53(6):820-841 [FREE Full text] [doi: 10.1080/00273171.2018.1446819] [Medline: 29624092]
41. Asparouhov T, Hamaker EL, Muthén B. Dynamic structural equation models. Struct Equ Modeling 2017 Dec 27;25(3):359-388. [doi: 10.1080/10705511.2017.1406803]
42. Du W, Côté D, Liu Y. SAITS: self-attention-based imputation for time series. Expert Syst Appl 2023 Jun 01;219:119619. [doi: 10.1016/j.eswa.2023.119619]
43. Cao W, Wang D, Li J, Zhou H, Li Y, Li L. BRITS: bidirectional recurrent imputation for time series. Red Hook, NY, USA: Curran Associates Inc; 2018 Presented at: Proceedings of the 32nd International Conference on Neural Information Processing Systems Red Hook, NY; December 3-8, 2018; Montréal, Canada p. 6776-6786. [doi: 10.48550/arXiv.1805.10572]
44. Du W. PyPOTS: A Python Toolbox for Data Mining on Partially-Observed Time Series. In: arXiv. Preprint posted online on May 30, 2023. Ithaca, NY, USA: Cornell University; May 30, 2023.
45. Wetherill RR, Fromme K. Acute alcohol effects on narrative recall and contextual memory: an examination of fragmentary blackouts. Addict Behav 2011 Aug;36(8):886-889 [FREE Full text] [doi: 10.1016/j.addbeh.2011.03.012] [Medline: 21497445]
46. Funabashi D, Dobashi S, Sameshima K, Sagayama H, Nishijima T, Matsui T. Acute vigorous exercise decreases subsequent nonexercise physical activity and body temperature linked to weight gain. Med Sci Sports Exerc 2024 Oct 01;56(10):1964-1975. [doi: 10.1249/MSS.0000000000003487] [Medline: 38758538]
47. Lafit G, Meers K, Ceulemans E. A systematic study into the factors that affect the predictive accuracy of multilevel VAR(1) models. Psychometrika 2022 Jun;87(2):432-476. [doi: 10.1007/s11336-021-09803-z] [Medline: 34724142]
48. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle NC, Mora-Gonzalez J, Löf M, et al. Data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017 Sep;47(9):1821-1845. [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]

XSL•FO
RenderX

49. Choudhury R, Park J, Banarjee C, Coca MG, Fukuda DH, Xie R, et al. Associations between monitor-independent movement summary (MIMS) and fall risk appraisal combining fear of falling and physiological fall risk in community-dwelling older adults. Front Aging 2024;5:1284694 [FREE Full text] [doi: 10.3389/fragi.2024.1284694] [Medline: 38660534]
50. Do B, Hedeker D, Wang W, Mason TB, Belcher BR, Miller KA, et al. Investigating the day-level associations between affective variability and physical activity using ecological momentary assessment. Psychol Sport Exerc 2024 Jan;70:102542. [doi: 10.1016/j.psychsport.2023.102542] [Medline: 37805039]
51. Colley RC, Tremblay MS. Moderate and vigorous physical activity intensity cut-points for the Actical accelerometer. J Sports Sci 2011 May;29(8):783-789. [doi: 10.1080/02640414.2011.557744] [Medline: 21424979]
52. Baroudi L, Zernicke RF, Tewari M, Carlozzi NE, Choi SW, Cain SM. Using wear time for the analysis of consumer-grade wearables' data: case study using Fitbit data. JMIR Mhealth Uhealth 2025 Mar 21;13(1):e46149. [doi: 10.2196/46149]
53. Baldwin SA, Fellingham GW, Baldwin AS. Statistical models for multilevel skewed physical activity data in health research and behavioral medicine. Health Psychol 2016 Jun;35(6):552-562. [doi: 10.1037/hea0000292] [Medline: 26881287]

Abbreviations

AR: autoregressive
BRITS: Bidirectional Recurrent Imputation for Time Series
CR: cross-regressive
CrI: credible interval
DV: dependent variable
EMA: ecological momentary assessment
EN: energetic feeling
FIML: full information maximum likelihood
ILD: intensive longitudinal data
LOCF: last observation carried forward
MAR: missing at random
MCAR: missing completely at random
mHealth: mobile health
MNAR: missing not at random
PA: physical activity
PMM: pattern mixture model
SAITS: Self-Attention-based Imputation for Time Series
SPM: shared parameter model
SWaN: Sleep Wake and Nonwear algorithm
uHealth: ubiquitous health
VAR: vector autoregressive

Edited by Z Liang; submitted 15.10.24; peer-reviewed by C Thomadakis, J Sena; comments to author 06.03.25; revised version received 01.05.25; accepted 16.07.25; published 30.10.25.

Please cite as:
Cho YW, Chow SM, Li J, Wang WL, Wang S, Ji L, Chinchilli VM, Intille SS, Dunton GF
Within- and Between-Individual Compliance in Mobile Health: Joint Modeling Approach to Nonrandom Missingness in an Intensive Longitudinal Observational Study
JMIR Mhealth Uhealth 2025;13:e65350
URL: https://mhealth.jmir.org/2025/1/e65350
doi:10.2196/65350
PMID:

©Young Won Cho, Sy-Miin Chow, Jixin Li, Wei-Lin Wang, Shirlene Wang, Linying Ji, Vernon M Chinchilli, Stephen S Intille, Genevieve Fridlund Dunton. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating the Clinical Effectiveness and Patient Experience of a Large Language Model–Based Digital Tool for Home-Based Blood Pressure Management: Mixed Methods Study

Alan Jelic[1], MD; Igor Sesto[1], MD; Luka Rotkvic[1], MD; Luka Pavlovic[2], MSc; Nikola Erceg[3], PhD; Nina Sesto[2], PhD; Zeljko Kraljevic[4], PhD; Joshua Au Yeung[5], MD; Amos Folarin[4], PhD; Richard Dobson[4], PhD; Petroula Laiou[4], PhD

[1]Magdalena Clinic for Cardiovascular Diseases, Zagreb, Croatia

[2]Megi Health Ltd, London, United Kingdom

[3]Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia

[4]King's College London, 16 De Crespigny Park, London, United Kingdom

[5]King's College Hospital NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Petroula Laiou, PhD
King's College London, 16 De Crespigny Park, London, United Kingdom

Abstract

Background: Hypertension, one of the most common cardiovascular conditions worldwide, necessitates comprehensive management due to its association with multiple health risks. Effective control often involves lifestyle changes and continuous monitoring, yet many individuals struggle to adhere to traditional management approaches. Digital health tools are emerging as promising alternatives, offering remote monitoring and real-time support. This study focuses on evaluating a digital tool specifically designed for hypertension management, analyzing its effectiveness, and gathering user perspectives on its functionality and impact.

Objective: The primary objective of this study is to assess the effectiveness of a digital health tool in managing hypertension. Additionally, the study aims to understand user experiences and satisfaction levels to gauge the tool's acceptance and potential for long-term use. By analyzing data from a large cohort, we seek to determine whether the tool can contribute to meaningful reductions in blood pressure and support sustained engagement over time.

Methods: The study includes a cohort of 5136 participants who used the digital hypertension management tool. This tool provides continuous blood pressure monitoring, real-time feedback, and personalized health recommendations, which are crucial for tailored intervention. Participants recorded their blood pressure values over time, and we tracked retention rates to measure adherence. An online survey was administered to gather user feedback, focusing on ease of use, satisfaction levels, and perceived health benefits.

Results: Our analysis indicates a significant reduction in blood pressure values among users, with a positive correlation observed between the duration of use and the extent of blood pressure reduction. We performed a 1-sided Wilcoxon Rank Sum test to compare systolic blood pressure values in the first and last biweekly use intervals, and descriptive statistics were used to assess survey responses. High retention rates were observed, with 2583 (50.3%) participants using the tool after 1 year. Survey responses revealed high satisfaction, with users highlighting the tool's ease of use and noting reduced anxiety related to blood pressure management. These results suggest that users found the digital tool both effective and convenient.

Conclusions: This study demonstrates the potential benefits of digital health tools in managing hypertension, emphasizing their ability to engage users over long periods and support blood pressure reduction. The high satisfaction rates and positive user feedback underscore the importance of user-centered design in creating effective health interventions. Overall, the findings suggest that digital tools, when designed with a focus on user experience, could be a valuable component in hypertension management strategies, complementing traditional health care approaches.

(JMIR Mhealth Uhealth 2025;13:e68361) doi:10.2196/68361

KEYWORDS

hypertension; large language models; LLMs; digital tools; artificial intelligence; AI

XSL•FO
RenderX

Introduction

Background

According to the American Heart Association, high blood pressure, called hypertension, is defined as systolic blood pressure (SBP) ≥130 mm Hg or diastolic blood pressure ≥80 mm Hg that remains elevated across time [1]. Hypertension affects over 1.2 billion people worldwide [2] and, in the last decades, the number of people with hypertension has doubled [3]. Hypertension is a major risk factor for cardiovascular diseases such as heart attack and heart failure, stroke, and chronic kidney disease [4]. Hence, effective management of hypertension is a key factor in reducing morbidity and mortality associated with cardiovascular complications.

Traditional approaches to hypertension management usually require regular visits to health care facilities for blood pressure monitoring and consultations. Such approaches can be burdensome for patients and place a strain on health care systems. Moreover, patients usually have difficulty with medication adherence and struggle to make lifestyle changes, with half of the patients stopping their treatment within a year [5], making this a complex and multifactorial challenge. These factors result in poor blood pressure control, leading to increased health risks, complications, mortality, and hospital admissions. To address these health issues, there is a growing interest in leveraging artificial intelligence (AI) and digital health technologies (eg, mobile apps and smartwatches) to facilitate home-based hypertension management. Such technologies offer long-term blood pressure monitoring, send reminders for the medication intake, provide educational content to support lifestyle changes, and provide personalized health advice.

Recent trials have demonstrated that digital applications can significantly lower baseline 24-hour ambulatory, home, and office SBP compared with the control group [6,7]. Since 2022, there has been an explosion of interest in large language models (LLMs) for their conversational ability and their potential as clinical AI chatbots. Current LLMs built on the transformer architecture have demonstrated comprehensive medical knowledge [8], human-level empathy [9], and comparable bedside manner [10] and medical safety to clinicians. Implementation of an LLM-enhanced digital solution may improve user interaction and efficacy.

While the clinical benefits of long-term blood pressure monitoring with digital technology are well documented [11], it is of paramount importance to also understand the effectiveness for blood pressure reduction as well as perspectives of users for such digital tools. Investigating users' perspectives for digital home–based blood pressure monitoring can contribute to the optimization and improvement of digital tools for hypertension as well as to ensure that such digital tools fulfill the needs and preferences of the users [12].

Objectives

In this work, we analyze the effectiveness for blood pressure reduction, user retention, and satisfaction with an innovative digital health tool that operates as a chatbot called Megi. Megi actively communicates with patients on a daily basis, reminds them to monitor their blood pressure, submits other relevant biometric data, takes their medications, and enhances their lifestyle. Specifically, we analyze the retention and engagement of 5136 Megi users and investigate whether they can achieve blood pressure reduction. In addition, we present the results of an online survey that quantifies users' perspectives about Megi. The insights obtained from this study can aid in the design, development, and effectiveness of digital tools for hypertension and enhance users' satisfaction.

Methods

Study Design

We conducted an observational cohort study with 5136 Megi users who used the platform between June 2022 and the end of July 2024 and fully completed the onboarding process. A user was considered fully onboarded once they accepted the consent; entered their first name, last name, and date of birth; and set the blood pressure measurement reminders. These users were not recruited through traditional methods; rather, they joined the platform organically through clinical referrals, health campaigns, or personal interest in self-managing hypertension. First, we analyzed user engagement and retention on the platform across the whole user base to assess the platform's overall success in retaining patients engaged and adherent with at-home blood pressure monitoring. We define retention rate as the percentage of users who continued to use Megi beyond specific time intervals after their initial use of the platform. In this study, we considered the intervals of 3, 6, 9, 12, and 24 months. Accordingly, users were grouped into cohorts based on the duration of their use of Megi. These durations correspond to the specified retention periods. We analyzed the retention based on two types of events: (1) blood pressure reporting and (2) any interaction with Megi.

Subsequently, we randomly selected a cohort of active users that had been using Megi for at least 3 months to evaluate their perspectives about Megi. The randomly selected users had to fill out an extensive online survey that consisted of 20 questions, which were designed to (1) evaluate the usefulness of Megi in improving self-management of hypertension, (2) assess their satisfaction, (3) describe their willingness to use Megi, and (4) assess their behavioral response to Megi.

Ethical Considerations

The survey was conducted in June 2023, and all participants gave written informed consent that they agreed to participate in the survey. Data were stored on a secure cloud platform that adheres to protected health information management standards. The study was approved by the Magdalena University Hospital Research Ethics Committee (195/Inf-701/22).

Digital Tool for the Management of Hypertension

All study participants used Megi, an AI-based digital health tool for hypertension management. Megi is a chatbot equipped with a specialized transformer-based AI model and is accessed via messaging apps such as WhatsApp and Viber. The chat is built on LLMs from the GPT family (GPT-3.5, GPT-4, and GPT-4o [OpenAI]). To date, various versions of OpenAI's models have been used. These models are accessed via the

XSL•FO
RenderX

Microsoft Azure OpenAI Service. Via this chatbot and through images, Megi gathers on a daily basis multimodal behavioral and biometric data from users that range from blood pressure and heart rate measurements to mood and stress levels. In addition, it provides personalized advice for lifestyle changes. On a daily basis, users receive reminders to measure their blood pressure and take their medications. All data are stored in secured cloud platforms, and clinicians have 24-7 access to all data via an online platform. Visual examples of the Megi platform and further information about Megi can be found on the Megi website [13].

Survey Questions

The online survey consisted of 19 questions. There were 3 questions that collected information about the users' age, sex, and employment status. In addition, there was 1 question that evaluated the impact of hypertension on the users' daily life. The survey had 3 questions that assessed the usefulness of the technology in improving self-management of hypertension (ie, "I understand why it is important to measure your blood pressure daily"; "I would often forget to measure my blood pressure if Megi didn't remind me"; and "With Megi, I will develop the habit of regular blood pressure measurement and monitoring"). Moreover, there were 3 questions that evaluated the users' satisfaction (ie, "I can see the benefits of using Megi"; "I have confidence in Megi"; and "I would continue to use Megi even if my doctor did not access my data"). The willingness to use the digital tool was evaluated by 4 questions (ie, "I would definitely recommend Megi to my friends"; "Anyone who cares about their health should use Megi"; "For the sake of myself and my loved ones, it is important for me to take care of my health"; and "My loved ones and friends would encourage me to use Megi"), while the behavioral response of the users was also assessed by 4 questions (ie, "Since using Megi, I feel safer and more relaxed"; "I feel happy when I receive messages from Megi"; "Megi makes me feel less lonely"; and "I wouldn't want to lose access to Megi"). Finally, the survey had 1 question that asked users to provide the unmet needs of the digital tool in free text. The answers were provided either in a 5-point Likert scale (1: I strongly disagree, 2: disagree, 3: neither agree nor disagree, 4: agree, and 5: strongly agree) or in free text.

Statistical Analysis

We analyzed the blood pressure values of all users in the first 2 weeks after their enrollment in the Megi platform and in the last 2 weeks before they stopped reporting their blood pressure values. For every user, we computed the average blood pressure values in the first and last biweekly segment. To quantify the statistical differences between the first and last biweekly segments, we applied a 1-sided nonparametric Wilcoxon Rank Sum test. Results were considered statistically significant if their *P* value was less than .05. In addition, we performed a descriptive analysis to assess the results of the online survey. Informative quotes were provided for the answers that were given in free text. All analysis was performed in MATLAB R2022a.

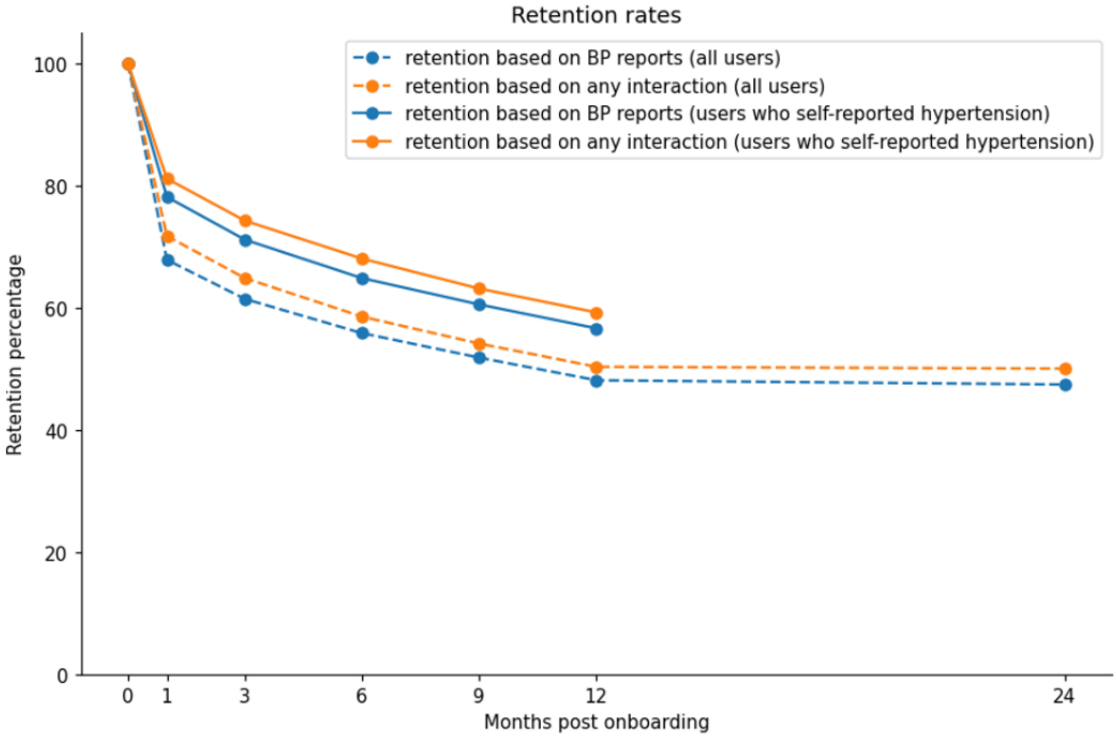
Results

User Retention

We first analyzed the retention of 5136 Megi users (n=2399, 46.71% female; n=2733, 53.21% male; n=4, 0.001% other; median age 59, range 17 - 95 y). Retention was quantified based on (1) blood pressure reporting and (2) any interaction with Megi. Figure 1 demonstrates that the retention rates after one year of using Megi were 50.3% for users based on all their interactions with Megi (eg, updating weight data, reacting to therapy reminders, and submitting feedback) and 48.1% for users based on their blood pressure measurement reports. The retention rates were even higher, that is, 60.5 and 63.1% respectively, for users who self-reported being hypertensive. Interestingly, user retention remained stable even after 2 years of using Megi.

XSL•FO
RenderX

Figure 1. User retention (N=5136) across months after onboarding based on blood pressure (BP) reporting (orange) and any interaction with Megi (blue). Dotted lines correspond to all users, while solid lines correspond to users who self-reported that they had hypertension upon their enrollment (this feature was implemented in the second year).



Blood Pressure Analysis

In total, 4736 (92.21%) of 5136 users reported their blood pressure values. We first sought to uncover whether there was a statistically significant difference in the average SBP measurements between the first and last biweekly segment. Figure 2 illustrates the distributions of the average SBP values of all users for the first biweekly segment after their enrollment and the last biweekly segment for which users reported their blood pressure values. We applied a 1-sided Wilcoxon rank sum test and found that the SBP values in the last biweekly segment were statistically significantly reduced compared to the first biweekly interval (P<.001).

Afterward, we wanted to explore whether such SBP differences were stronger for users whose average SBP values were at least 140 mmHg in the first biweekly segment (Figure 3). We saw that such differences between the first and last biweekly interval were much stronger (P=1.7×10^{-9}). We also examined the correlation between the SBP drop (ie, the difference of the average SBP during the first biweekly interval and the last biweekly interval) and the period for which users report their blood pressure values. The Spearman correlation was 0.53, and it was statistically significant (P=4.5^{-32}), which means that larger SBP reduction is positively correlated with longer use of Megi. Figure 3 illustrates the SBP drop across the months for which users reported their blood pressure values. The maximum duration in blood pressure reporting was 24 months, and we observed that there was an increasing trend of the SBP drop across time.

We performed a similar analysis for the users whose average SBP was at least 150 mm Hg in the first biweekly segment (Figure 4). For users with such high blood pressure, the differences between the first and last biweekly interval were also strong (P value=1.7×10^{-9}) and the correlation between the SBP drop and the duration for which users report their blood pressure measurements was even stronger (ρ=0.75). Figure 4 also illustrates a clear increase in the SBP drop across time.

XSL•FO
RenderX

Figure 2. Distributions of the average systolic blood pressure (SBP) measurements during the first biweekly segment after study enrollment (blue) and last biweekly segment (orange) before users paused reporting blood pressure values.

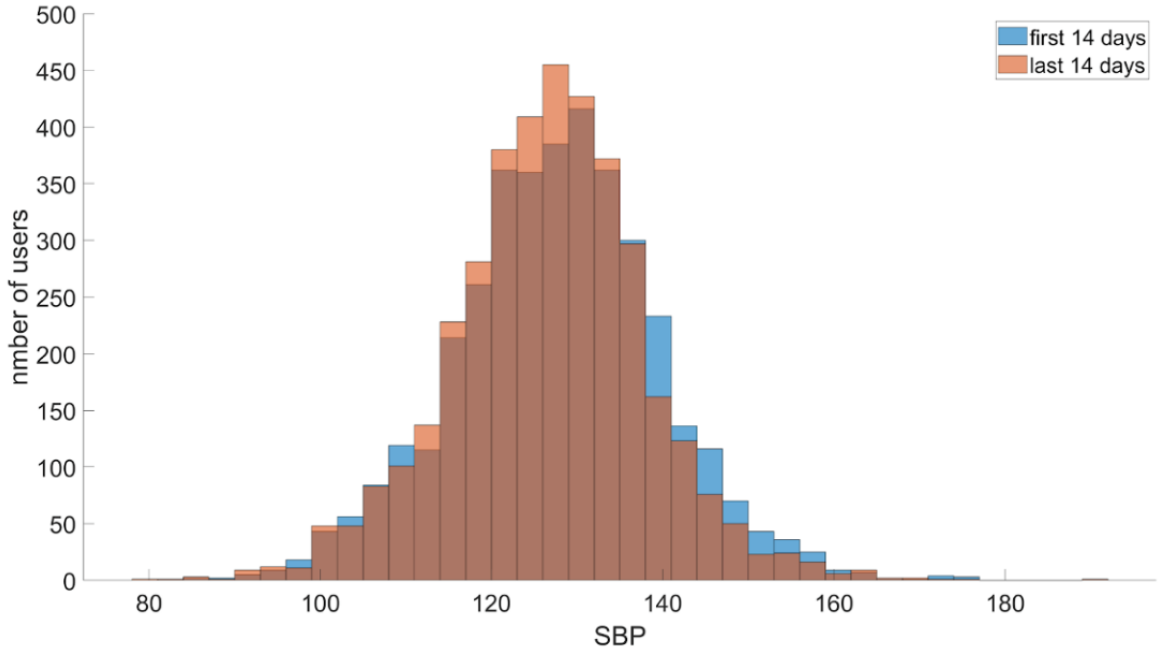


Figure 3. Left: distributions of systolic blood pressure (SBP) measurements for users whose average SBP value was at least 140 mm Hg during the first weekly biweekly interval after enrollment (blue). The orange violin plot depicts the average SBP values during the last biweekly interval for which users reported their blood pressure values. Center: SBP drop across the time (in mo) for which users report their BP measurements. Each dot corresponds to a different user. Right: distributions of the SBP drop across different time intervals (in mo) for which users report the blood pressure measurements.

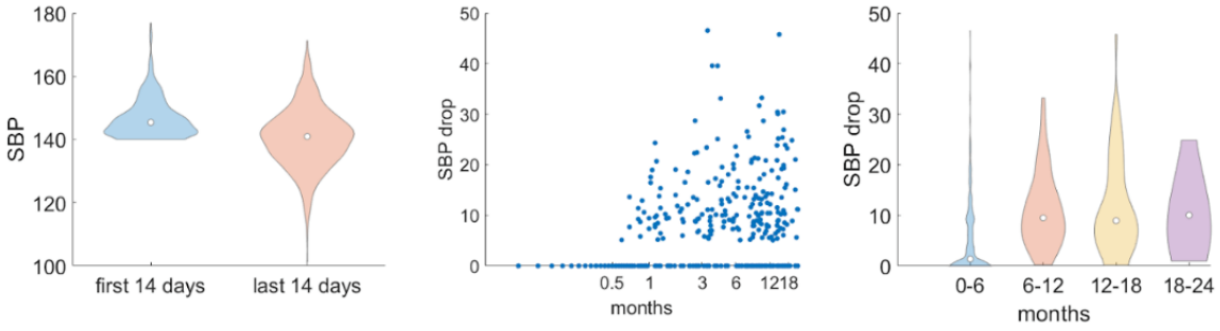
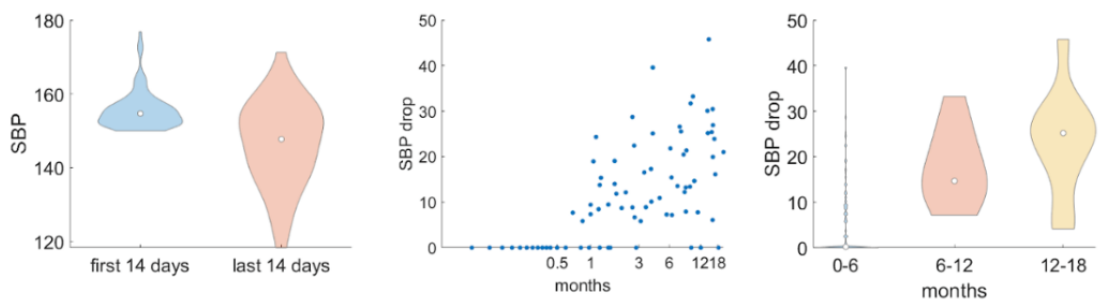


Figure 4. Users whose average systolic blood pressure (SBP) was at least 150 mm Hg during the first biweekly interval after their enrollment.



XSL•FO
RenderX

Online Survey Analysis: Dataset Characteristics

A total of 129 people clicked on a link to access the survey (June 2023). The total number of analyzed participants is 125 (96.9%; n=3, 2.33% participants were excluded from the analysis due to the presence of missing values). Summary demographics regarding participants' age, sex, and employment status are provided in Table 1 and Figure 1. Out of the 125 participants, 55 (44%) were female, and the age of all survey participants ranged from 23 to 82 (mean 60, SD 11.6) years.

At first, study participants were asked whether high blood pressure has a significant impact on their life. The vast majority of the participants strongly agreed (n=58/125, 47%) or agreed (36/125, 29%) while just 23 of 125 (19%) strongly disagreed or disagreed.

Table . Dataset characteristics (N=125).

Characteristics	Users, n (%)
Age group (y)	
20-29	2 (1.6)
30-39	3 (2.4)
40-49	18 (14.4)
50-59	27 (21.6)
60-69	52 (41.6)
70-79	20 (16)
80-89	1 (0.8)
Sex	
Male	70 (44)
Female	55 (56)
Employment status	
Employed	54 (43.2)
Unemployed	5 (4)
Retired	64 (51.2)
Student	1 (0.8)

Participants were asked a series of questions to evaluate the usefulness of technology in managing their hypertension. The responses indicated a high level of perceived usefulness. Overall, 80% (100/125) of the participants agreed or strongly agreed that Megi helped them understand their hypertension better. Regarding the importance of blood pressure monitoring, 98% (122/125) found the technology useful for regularly monitoring their blood pressure. When the participants were asked whether they would forget to measure their blood pressure if Megi did not remind them, 80% (100/125) responded that they would. In addition, study participants felt that using Megi made them develop the habit of measuring their blood pressure regularly (118/125, 95% responded positively).

When participants answered questions that assessed their overall perception, confidence in the digital tool, and willingness to continue using it independently of their health care provider, again their overall satisfaction was very high. Specifically, regarding the statement "I can see the benefits of using Megi," 98% (122/125) of the study participants strongly agreed or agreed. In addition, when participants were asked whether they had confidence in Megi, 92% (115/125) responded positively, while when they were asked whether they were willing to use Megi independently, more than 80% (100/125) agreed or strongly agreed.

Study participants were also assessed regarding their willingness to use or recommend Megi. Specifically, they asked whether they (1) would recommend Megi to their friends (112/125, 90% of the users responded positively) and (2) would agree with the following statements: "Anyone who cares about their health should use Megi" (100/125, 80% of the users strongly agreed or agreed); "For the sake of myself and my loved ones, it is important for me to take care of my health" (112/125, 90% of the users responded positively); and "My loved ones and friends would encourage me to use Megi" (more than 55% of the users responded positively).

At the end of the survey, participants had to answer questions regarding their behavioral response with the digital tool. They had to answer the following questions: "Since using Megi, I feel safer and more relaxed"; "I feel happy when I receive messages from Megi"; "I wouldn't want to lose access to Megi"; and "Megi makes me feel less lonely." Across all questions but the last, more than 80% (100/125) of the participants felt that Megi reduced their anxiety and made them feel safe.

Unmet Needs

Study participants were also asked whether there is something that they would change about Megi. Answers were given in free text. Some users were asking for more personalized treatment (eg, "I think maybe she should adjust according to the specifics

XSL•FO
RenderX

of individual patients. I can illustrate this with my own example. I have been hypertensive for about thirty years, I regularly measure my blood pressure, so I don't think it's necessary to report my condition every two weeks"; "I would like to determine more precisely what type of pressure I have...when it is really low and when it is high, and based on the measurements…") or the possibility of clinical consultations with cardiologists (eg, "Maybe the availability of a cardiologist to guide me"; "If there were an opportunity to occasionally discuss the results with a doctor, I think I would take regular measurements more seriously"). On the basis of the qualitative feedback, several common themes emerged that can inform future improvements to Megi. These user-driven recommendations are summarized here—(1) personalization of reminders: allow users to adjust the frequency or timing of blood pressure measurement reminders based on their experience and needs; (2) flexible reporting options: give experienced users the option to reduce or customize the frequency of data entry, especially if their condition is stable; and (3) clinical support integration: offer optional access to health care professionals (eg, cardiologists or general practitioners) for periodic review or consultations.

Discussion

Principal Findings

The management of hypertension at home using digital apps presents a promising avenue for improving patient outcomes and alleviating the burden on health care systems. This study explored users' perspectives on Megi, a digital tool designed to facilitate long-term home-based blood pressure monitoring, and demonstrated the effectiveness of Megi on blood pressure reduction.

Analyzing the activity of 5136 users, our results show very high engagement and adherence to blood pressure readings that remain stable even after one year of using Megi (Figure 1). Specifically, more than 50% of the users are active on the Megi digital health platform for 12 months, which demonstrates higher user retention when compared to previously reported industry standards. For example, consistent with other large-scale digital health studies, the notable Stanford-led MyHeart Counts study experienced substantial dropout rates; mean engagement with the app was only 4.1 days [14]. Statista research into the worldwide retention rate of mobile app installs reveals that, on the day of installation, the average retention rate across the 31 categories was 25.3% [15]. Worryingly, by day 30, this percentage drops to just 5.7%, and when it comes to the health and wellness industry specifically, it is even lower at 2.8% [15].

In addition, our analysis showed the effectiveness of Megi as a digital tool for hypertension management (Figures 2-4). We demonstrated that the average SBP of the last biweekly interval for which users reported their blood pressure values was statistically significantly smaller compared to the first biweekly segment after users' enrollment (Figure 2). This difference was even stronger for users who had high (>140 mm Hg; Figure 3A) and very high SBP measurements (>150 mm Hg; Figure 4A) during the first biweekly segment. Furthermore, we showed that there is a positive correlation between the SBP drop and the duration for which users report their blood pressure values. This means that larger SBP reductions are associated with longer use of Megi (Figures 3, 4B, and 4C).

While our results demonstrate a significant association between longer Megi use and reductions in SBP, it is important to note that the observational nature of the study limits causal inference. Without a control group, we cannot fully exclude the impact of confounding variables such as concurrent lifestyle changes, improved medication adherence independent of Megi, or parallel support from health care providers. These factors may have contributed, in part, to the observed improvements. Future randomized controlled trials are needed to validate these findings and better isolate the direct effect of the digital tool on clinical outcomes.

Through an online survey, we evaluated users' perspectives on the technology's usefulness in improving self-management of hypertension, their satisfaction with the tool, their willingness to use it, and their behavioral responses to its implementation. The findings indicate a high level (ie, more than 90%) of user satisfaction with Megi. Participants strongly acknowledged the technology's usefulness in enhancing their ability to manage hypertension independently (Figures 3 and 4), as well as expressed their willingness to suggest Megi to other users. The tool's integration into daily routines led to a noticeable reduction in anxiety and an increase in feelings of relaxation and happiness, suggesting a positive impact on mental well-being. The behavioral response data—particularly users reporting feeling safer, less anxious, and more connected—aligns with previous findings on the importance of emotional engagement in digital health interventions. Previous studies have shown that user adherence improves when digital tools provide empathic, human-like interaction, particularly in long-term chronic condition management [9,12]. Megi's use of LLM-powered messaging may have contributed to these outcomes by delivering timely, personalized, and conversational support that mimics human dialog. This is consistent with recent research demonstrating that LLMs can convey human-like empathy and reassurance in health contexts [8,10]. Furthermore, the willingness of users to recommend Megi to friends reflects strong approval and confidence in its efficacy. Overall, the positive feedback from users highlights the potential of Megi as a valuable digital tool for hypertension management.

Limitations

Despite the valuable insights, this study has limitations that must be acknowledged. First, its observational design limits the ability to draw causal conclusions. Second, the online survey relied on convenience sampling and included only users with at least 3 months of continuous use, excluding those with intermittent engagement. This may introduce selection bias and limit generalizability. In addition, the single ethnicity of the analyzed cohort limits the generalizability of the findings. Cultural attitudes toward technology, patient-clinician relationships, and the structure of local health care systems may influence user engagement, retention, and perceived value of digital tools. Moreover, the evaluation focused exclusively on one digital tool, that is, Megi, which may not provide a comprehensive understanding of the broader landscape of digital

XSL•FO
RenderX

tools for hypertension management. These constraints may affect the generalizability of the findings and highlight the need for further research with more diverse and larger populations to validate and expand upon these results.

Recommendations

In addition, we recommend that future research include more diverse populations to enhance the generalizability of results and evaluate clinically validated outcomes such as cardiovascular events and treatment modifications. Incorporating structured metrics for therapeutic compliance and integrating the platform with clinical workflows would further strengthen its clinical utility. In addition, prospective studies—particularly randomized controlled trials—are needed to establish causal relationships and assess the long-term impact of digital tools such as Megi on hypertension management and overall cardiovascular health.

Conclusions

Our study demonstrated the effectiveness of a digital tool for hypertension management and highlighted the importance of understanding user views in the design and implementation of digital health interventions for hypertension management. In conclusion, the insights obtained from this study highlight the potential of digital apps to enhance hypertension management by providing continuous monitoring, real-time feedback, and personalized health recommendations.

Conflicts of Interest

NS declares financial support from Megi Health Ltd. PL declares a financial interest in Megi Health Ltd.

References

1. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018 Jun;71(6):1269-1324. [doi: 10.1161/HYP.0000000000000066] [Medline: 29133354]
2. Hypertension. World Health Organization (WHO). URL: https://www.who.int/news-room/fact-sheets/detail/hypertension [accessed 2024-03-21]
3. Zhou B, Carrillo-Larco RM, Danaei G, et al. Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep;398(10304):957-980. [doi: 10.1016/S0140-6736(21)01330-1]
4. Chobanian AV, Bakris GL, Black HR, et al. The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 2003 May 21;289(19):2560-2572. [doi: 10.1001/jama.289.19.2560] [Medline: 12748199]
5. GuidanceHealth matters: combating high blood pressure. Government of UK. URL: https://www.gov.uk/government/publications/health-matters-combating-high-blood-pressure/health-matters-combating-high-blood-pressure [accessed 2024-07-02]
6. Nomura A, Tanigawa T, Kario K, Igarashi A. Cost-effectiveness of digital therapeutics for essential hypertension. Hypertens Res 2022 Oct;45(10):1538-1548. [doi: 10.1038/s41440-022-00952-x] [Medline: 35726085]
7. Kishi T, Usui H, Nagayoshi S. A prospective pilot trial of the health data monitoring system without educational content for patients with hypertension. Hypertens Res 2024 Feb;47(2):529-532. [doi: 10.1038/s41440-023-01510-9] [Medline: 38017186]
8. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature New Biol 2023 Aug;620(7972):172-180. [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
9. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 1;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
10. Mukherjee S, et al. Polaris: a safety-focused LLM constellation architecture for healthcare. arXiv. Preprint posted online on Mar 20, 2024. [doi: 10.48550/arXiv.2403.13313]
11. Bashshur RL, Howell JD, Krupinski EA, Harms KM, Bashshur N, Doarn CR. The empirical foundations of telemedicine interventions in primary care. Telemed J E Health 2016 May;22(5):342-375. [doi: 10.1089/tmj.2016.0045] [Medline: 27128779]
12. Orji R, Mandryk RL, Vassileva J, Gerling K. Tailoring persuasive health interventions to young adults: the role of user characteristics, goal relevance, and motivational differences. Pers Ubiquitous Comput 2018;22(4):647-664. [doi: 10.1007/s00779-017-1020-4]
13. Megi - Heart Health for Women. Megi. URL: https://www.megi.ai/ [accessed 2025-10-10]
14. Hershman SG, Bot BM, Shcherbina A, et al. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the MyHeart Counts Study. Sci Data 2019 Apr 11;6(1):24. [doi: 10.1038/s41597-019-0016-7] [Medline: 30975992]
15. Retention rate on day 30 of mobile android app installs worldwide in 3rd quarter 2024, by category. Statista. URL: https://www.statista.com/statistics/259329/ios-and-android-app-user-retention-rate/ [accessed 2024-07-26]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
LLM: large language model
SBP: systolic blood pressure

Edited by Z Ni; submitted 04.11.24; peer-reviewed by GK Gupta, KL Du, R Verma; revised version received 07.08.25; accepted 06.09.25; published 03.11.25.

Please cite as:
Jelic A, Sesto I, Rotkvic L, Pavlovic L, Erceg N, Sesto N, Kraljevic Z, Au Yeung J, Folarin A, Dobson R, Laiou P
Evaluating the Clinical Effectiveness and Patient Experience of a Large Language Model–Based Digital Tool for Home-Based Blood Pressure Management: Mixed Methods Study
JMIR Mhealth Uhealth 2025;13:e68361
URL: https://mhealth.jmir.org/2025/1/e68361
doi:10.2196/68361

© Alan Jelic, Igor Sesto, Luka Rotkvic, Luka Pavlovic, Nikola Erceg, Nina Sesto, Zeljko Kraljevic, Joshua Au Yeung, Amos Folarin, Richard Dobson, Petroula Laiou. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adolescent Engagement With a Multicomponent mHealth Tool: Identifying Usage Patterns, Determinants, and Health Behavior Change in an Intervention Trial

Carmen Peuters[1,2,3], PhD; Ann DeSmet[4,5], PhD; Laura Maenhout[2], PhD; Greet Cardon[2], PhD; Dries Debeer[6], PhD; Geert Crombez[1], PhD

[1]Department of Experimental-Clinical and Health Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Henri Dunantlaan 2, Ghent, Belgium

[2]Department of Movement and Sports Sciences, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

[3]ISGlobal, Barcelona Institute for Global Health - Campus MAR, Barcelona Biomedical Research Park (PRBB), Carrer del Doctor Aiguader, 88, Barcelona, Spain

[4]Faculty of Psychology, Educational Sciences and Speech Therapy, Université Libre de Bruxelles, Brussels, Belgium

[5]Department of Communication Studies, University of Antwerp, Antwerp, Belgium

[6]Department of Experimental Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

Corresponding Author:
Carmen Peuters, PhD
Department of Experimental-Clinical and Health Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Henri Dunantlaan 2, Ghent, Belgium

Abstract

Background: Research about the engagement of adolescents with mobile health (mHealth) interventions is scarce, while it is generally assumed that the engagement affects the intervention efficacy.

Objective: Using an mHealth intervention that targets the general population of adolescents to promote healthy behaviors (physical activity, low sedentary time, adequate sleep, and taking breakfast) and mental health, we aimed to investigate (1) how adolescents engage with the intervention, (2) which engagement styles can be identified and how these differ according to personal characteristics, and (3) which style of engagement predicts behavior change. The intervention used, #LIFEGOALS, includes self-regulation techniques, a support chatbot, narrative videos, and gamification, brought together in an app coupled to an activity tracker.

Methods: Logged usage data and self-reports of experience with #LIFEGOALS were collected from 159 adolescents (mean age 13.54, SD 0.95 years) over a 12-week intervention period and used to describe behavioral and experiential engagement with the intervention components over time. Baseline data on sociodemographic variables, mental health, and behavioral determinants were explored as determinants of engagement and were used to characterize engagement styles that were identified through exploratory cluster analysis on the frequency of usage of the components. Linear mixed-effects regression was used to analyze the effect of engagement style on health behavior change.

Results: Average time in the app was 26 minutes (SD 26) over the 12-week period, with usage decreasing substantially after the first week. The use of self-regulation techniques and gamification was strongly interrelated ($0.65 < r < 0.70$), whereas use of Fitbit showed weaker correlations with other component usage ($0.15 < r < 0.31$). Exploratory analyses suggest that engagement was influenced by immigration background and by adolescents' attitudes, self-efficacy, and intentions toward healthy living. Younger participants tended to use the Fitbit more frequently. Cluster analysis identified 4 engagement styles: narrative usage (n=19), app usage (n=36), Fitbit usage (n=32), and no usage (n=72), which were associated with differences in age, peer support, and mental health. Engagement style did not affect change in health behavior outcomes from preintervention to postintervention.

Conclusions: Different engagement styles were identified based on the frequency and type of components used. Findings support the relevance of tailoring mHealth to individual, interpersonal, and contextual characteristics. The overall low engagement with the intervention may have limited the detection of differences in health effects between engagement styles.

Trial Registration: Clinicaltrials.gov NCT04719858; https://clinicaltrials.gov/study/NCT04719858

(JMIR Mhealth Uhealth 2025;13:e59041) doi:10.2196/59041

XSL•FO
RenderX

KEYWORDS

behavioral engagement; experiential engagement; multicomponent interventions; unstructured apps; cluster analysis; mobile health; behavior change; system log data; law of attrition; tailoring digital health interventions; teens; youth; adolescents; engagement; support chatbot; gamification; activity tracker; logistic regression; sleep quality; mobile phone

Introduction

Mobile health (mHealth) is defined as the use of mobile devices such as smartphones and activity trackers in health care. Compared with in-person care, mHealth may offer advantages in promoting health by reducing barriers such as cost, stigma, confidentiality concerns, and scheduling or distance constraints [1,2]. mHealth may also facilitate health empowerment by providing personalized, real-time information and support [3] and by enabling independent health management that can bypass parental dependence [2,4]. Health apps may be particularly suitable for adolescents [5]. Following statistics from Western countries, between 93% and 98% of teens aged 12-17 years own a smartphone [6-8]. In Flanders, Belgium, 96% of adolescents use their smartphone more than an hour, and 45% even more than 4 hours on days without school [9]. The integration of smartphones in adolescents' daily life may facilitate the adoption of a health app. Nevertheless, it remains a challenge to get adolescents to initiate and sustain engagement with digital health interventions. In research trials, a high percentage of adolescents do not use the app or drop out early [10-13]. These data are estimated to be even higher in real-world usage [14]. Low intervention usage relates to fewer positive health effects [15-17], but little is known about the quality of engagement and how this relates to health outcomes.

Intervention log analytics is an approach to study and interpret engagement with digital interventions. For unstructured apps (ie, consisting of variable components that users can access and use at will), useful indicators are the frequency of interactions logged, the number of app features accessed, the number of activities completed, and the number of minutes spent in a web-based program [18,19]. Logged usage data have been used to determine user styles, either via cluster analysis or by transforming a continuous engagement outcome into discrete categories [16,18]. Users are mainly classified according to the extent of usage of the intervention [15-17,20], but user patterns can also be explored in terms of engagement with different intervention components or features. This might be relevant for multicomponent and (partly) unstructured interventions where it can reveal naturalistic user patterns, such as preferences for (the combination of) certain intervention features and their relationship with intervention effects. A few studies have identified clusters that reflect a difference in the use of different components [17,20]. Besides a cluster with overall low usage, the remaining clusters distinguished between using mainly 1 component of the intervention and using that component in combination with other features of the intervention. Effects of these types of clusters on health outcomes are mixed and likely depend on the intervention content and components. For example, the analysis of engagement with a digital intervention for young adults recovering from first-episode psychosis found health improvements only for the group with maintained use of the therapy and the social component but not for the group with maintained use of only the social component [17]. Another web-based intervention for adults with mild to moderate depression, anxiety, and stress did not find an effect of the clusters on health outcomes [18].

With the aim of maximizing intervention effects, an identification of different engagement styles and their health effects should be complemented with an understanding of the factors that influence patterns of engagement. Intervention characteristics (eg, novelty, autonomy, or complexity) or intervention components (eg, reminders or social support features) can have a different impact on individuals' engagement depending on user characteristics (eg, motivation for the health behavior, mood, or education level) [21]. More research is needed on the determinants of different aspects of engagement to inform future design and tailoring of effective mHealth interventions.

This study uses the multicomponent mHealth intervention #LIFEGOALS as a case study to (1) describe engagement with the intervention, (2) identify and characterize clusters of engagement, and (3) analyze their influence on health behavior change. #LIFEGOALS is an app coupled with an activity tracker to guide adolescents toward healthier lifestyles for promoting mental health. The intervention showed beneficial effects on physical activity, sedentary behavior, sleep quality, and emotions (ie, in a close-to-normal situation, not in case of COVID-19 pandemic restrictions) [22].

Methods

Study Design

This study used data from adolescents who participated in a quasi-randomized controlled effect study of the #LIFEGOALS intervention [22]. Participants were recruited from the general population of adolescents enrolled in nonspecial needs secondary education in Flanders, Belgium. The recruitment process was designed to ensure adequate representation of students from the vocational education track, which focuses on practical and job-oriented training. In Flanders, adolescents in the vocational track typically come from lower socioeconomic backgrounds compared with their peers in the general or technical tracks [13] and are at greater risk for engaging in unhealthy behaviors [23]. From the 184 participants allocated to the intervention group, 25 could not install the app due to technical reasons and were not included in this study, leaving a final sample of 159 adolescents (mean age 13.54, SD 0.95 years; 82/159, 51.6% girls). Data were collected in 3 waves. All waves included 2 intervention schools and started baseline assessment in October 2020, November 2020, and January 2021, respectively. At baseline, and after 2, 6, and 12 weeks of intervention, participants completed a web-based survey. Following the baseline assessment, participants were assisted to install the #LIFEGOALS app on their smartphone and were given a Fitbit Charge 2 or 3 on-lend to link with the app. The intervention

XSL•FO
RenderX

was introduced as a health app to improve lifestyle, and participants were invited to use the intervention in the next 12 weeks. As it concerns an unstructured app, users could access all components at all times and could engage with the intervention as they wished. The working of the app was explained in the classroom by means of a slide show. In the last wave of data collection (n=62), the first episode of the narrative was shown in the classroom because low viewership numbers for the narrative were observed in waves 1 and 2.

Intervention

#LIFEGOALS is an mHealth promotion multicomponent intervention for Flemish adolescent youth, which aims to promote mental health by improving healthy lifestyle behaviors. The app focuses on increasing physical activity, reducing sedentary behavior, improving adequate sleep, and daily breakfast consumption. The intervention is theory-based (Health Action Process Approach [24], Persuasive Systems Design [25], and extended Elaboration Likelihood Model [26]) and was developed using a participatory design approach. It includes (1) self-regulation techniques guiding users through goal setting, action and coping planning, and self-monitoring (using an activity tracker and graphs) to help bridge the intention-behavior gap; (2) a health narrative in the format of an entertaining storyline offered in weekly short video episodes to facilitate health behavior change through identification with the characters; (3) an automated, rule-based (non–artificial intelligence) chatbot for individual support by providing information and sending supporting messages; and (4) other content and features such as health information and gamification. The use of the #LIFEGOALS intervention was stimulated by placing a roll-up banner in the participating classes to remind participants of using the app. The intervention is described in more detail elsewhere [27], and the home screen is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Home screen of the #LIFEGOALS app.

XSL•FO
RenderX

Ethical Considerations

The study was approved by the ethics committee of the Faculty of Psychology and Educational Sciences of Ghent University (registration no.: 2020/70). Participants, their parent or guardian, and the school director received an information form tailored to their role detailing the study process, data protection measures, and the right of the participants to withdraw. Written consent was obtained before participation. During data collection, participants' identities were safeguarded through pseudonymization, using an encrypted key file to link participant codes with identities. After data collection and merging of user log data, questionnaire follow-ups, and baseline data, the dataset was fully anonymized and the key file was permanently destroyed. As an incentive, participants received a power bank after the baseline assessment and a cinema ticket upon completing the follow-up measurement.

Measures

Behavioral Engagement

By means of a unique identifier that had to be entered when installing the #LIFEGOALS app, user app activity was continuously logged throughout the 12-week intervention period. Log files included a date stamp and time stamp, the sort of activity (eg, filling out a certain section of the app, setting a certain choice or content, or receiving a push message by the app), and the content of the activity, if relevant (eg, text filled in by the participant, the option that was chosen, or the content of the push message). The usage of different components was operationalized by the frequency of opening each component section in the app (Table 1). Usage of the narrative was additionally operationalized by the self-reported number of watched episodes after week 1 (answer options: 0, 1, or all 2), after week 6 (0, 1 or 2, or 3 or more), and after week 12 (0, 1 or 2, 3-11, or all 12). Likewise, for the Fitbit, a unique login provided during installation gave the researchers access to the Fitbit data after participants synchronized their Fitbit in the Fitbit app. A total number of daily steps that exceeded zero was coded as a day with Fitbit usage.

Table . The different app components.

App component	Description
Action planning	Self-regulation technique. Via the button "Set mission" users are guided to translate a goal (eg, increase physical activity) into an action plan (eg, going to school by bike on Friday at 8 AM).
Coping planning	Self-regulation technique. When creating their action plan, users are asked to think of possible obstacles that may occur (eg, forgetting) and to think ahead of possible solutions for that obstacle (eg, hang post-its). Coping planning is, furthermore, accessible via the "Tools" button.
Agenda	The button "My agenda" offers a week overview of all planned missions, as well as a list "My missions" where—for every mission in the past—users can tick off whether they managed to fulfill the mission.
Chatbot	"Botty" is a chatbot to which users can ask questions (about a healthy lifestyle or the functioning of the app) and which sends encouraging messages and an announcement when a new episode of the narrative is available.
Narrative	Twelve short video episodes (3-6 minutes each) screen the everyday life of 4 adolescent classmates with a focus on entertainment. Implicitly, the storyline shows the importance of a healthy lifestyle for mental well-being. Each week, a new episode comes available in the app.
Gamification	Users can earn "coins" by planning and completing missions. Earned coins can be used to purchase attributes to personalize an avatar. Gamification usage was operationalized by summing the frequency of opening the app sections "Coins" and "Avatar."
Progress graphs	The "Graphs" button presents graphical representations of user progress to self-monitor their behavior (number of steps, minutes spent in physical activities, sleep duration, and frequency of taking breakfast).
Information	Under the "Tips & Tricks" button are included: information about the health benefits of the targeted behaviors, inspiration on how to perform the health behaviors, and referral to other organizations and websites for more information or guidance.

Experiential Engagement

After week 1, week 6, and postintervention, participants reported whether they had watched any episode of the narrative and whether they had used the other parts of the app. Participants who reported nonusage were asked in an open question for the reason. For participants who indicated at least minimal usage, their user experience was measured for, respectively, the narrative component and the other parts of the app. Items from the Digital Behavior Change Intervention Engagement Scale

XSL•FO
RenderX

[28] were translated and specified to the #LIFEGOALS narrative or the other parts of the app. After factor analysis on data of the pilot study (n=35), items were limited to 3 (interest, attention, and affect) for *experiential engagement with the narrative*, and 4 (interest, attention, affect, and usefulness) for *experiential engagement with the other parts of the app*. Answers on a 5-point Likert scale (totally disagree—totally agree) were averaged with higher scores indicating higher experiential engagement.

Behavioral Determinants

Attitude (ie, the tendency to evaluate behavior with some degree of favor or disfavor), *self-efficacy* (ie, the belief in one's capability to perform the behavior), and *intention* (ie, the explicit decision to act in a certain way) were assessed separately for physical activity, sedentary behavior, sleep, and frequency of breakfast consumption. The items were based on scales used in other studies ([29] for physical activity and sedentary behavior; [30] for sleep) from which the highest loading items from factor analysis were selected (retaining 3 items for attitude, 3 for self-efficacy, and 1 for intention). The scores from 1 (not agree) to 5 (totally agree) for the different behaviors were averaged to reflect the determinant toward a healthy lifestyle in general.

Health Behaviors and Mental Health

Physical activity, *sedentary behavior,* and *sleep time* were measured at baseline and postintervention using Axivity AX3 accelerometers worn on the nondominant wrist for 7 consecutive days. Raw accelerometry data were processed in R to obtain the average Euclidean Norm Minus One per minute (ie, volume of physical activity) expressed in milligravity-based acceleration units (mg), the average of sedentary minutes per day (ie, sedentary time), and weekday-to-weekend sleep time difference (ie, sleep regularity) (see the study by Peuters et al [22]). Phillips cut points [31] were used to classify each activity minute into light or moderate to vigorous intensity physical activity. *Sleep quality* (ie, a composite score of items on sleep latency, sleep interruption, daytime sleepiness, and experience of sleep quality, based on the studies by Inchley et al [32] and Essner et al [33]) and *frequency of breakfast consumption* (ie, the average days per week taking breakfast [32]) were assessed in the web-based survey at baseline and postintervention.

Health-related quality of life, *psychological well-being*, *emotions*, *self-perception,* and *peer support* were assessed using the KIDSCREEN-10 [34] and subscales from the KIDSCREEN-27 [35] and KIDSCREEN-52 [36]. Higher *t*-scores (mean 50, SD 10) indicate better mental health. *Resilience* was measured with an adapted Dutch version of the Brief Resilience Scale [37], with higher averaged scores indicative of greater resilience. *Depressed feelings* were measured using the Custom Short Form of the Dutch version of the PROMIS-PedDepSx (Patient-Reported Outcomes Measurement Information System Pediatric Bank Depressive Symptoms v2.0) [38], with higher *t*-scores (mean 50, SD 10) indicating more depressed feelings.

Pandemic-Related Education Measures

The data collection took place during the second wave of the COVID-19 pandemic in Belgium. In this period in Belgium, people were allowed to leave the house by foot or bike, but other governmental measures were implemented to limit the spread of infections. One of these measures that might have affected adolescent usage of the intervention was a minimum of 50% per week distance learning for adolescents in the third year of secondary education and above (not affecting first and second years). A binary variable was created representing the presence of education restrictions (normal in-school education vs partly remote education).

Statistical Analyses

Describing Engagement

Linear mixed models with school class as random intercept (to deal with the nested nature of the data) were used to analyze differences in engagement according to individual characteristics. Frequency variables were log-transformed to account for their skewed distribution. The reported statistical tests use Satterthwaite's method to approximate degrees of freedom, as implemented in the *lmerTest* R package [39]. Given the exploratory nature of these analyses including multiple comparisons, *P* values are reported descriptively and not interpreted for significance.

Identifying Engagement Styles

The *cluster* R package [40] and R (version 4.0.5) were used to identify clusters in the engagement data. The selection of which engagement indicators to include in the final cluster analysis was based on the descriptive statistics of engagement and on the theoretical and evidence-based rationale for inclusion of the different components in the intervention (Multimedia Appendix 1). Because data of mixed types (continuous for frequency and ordinal for narrative episodes) were used, Gower distance [41] was used as the distance measure. The frequency variables were right-skewed and, therefore, log-transformed. Partitioning Around Medoids was used as an iterative clustering procedure. Average Silhouette Width was used to decide on the optimal number of clusters. Only those solutions with 3 or more clusters (to explore patterns of engagement outside of a dichotomous high vs low use range) and only those solutions where all clusters consisted of ≥15 participants (to be able to perform follow-up analyses) were considered.

Determinants of Engagement Style

Differences between clusters on baseline characteristics were investigated with multinomial logistic regression using the *nnet* R package [42]. Given the nested nature of the data (ie, students nested within classroom), we ran multilevel versions of the models with school class as random intercept. However, these models did not explain additional variance and resulted in the same conclusions, for which only the results from the models without multilevel structure are presented. Due to the higher number of missing values in the accelerometry data, a separate model was fitted for the health behaviors versus the sociodemographic variables (age, gender, and family affluence), behavioral determinants (attitude, self-efficacy, and intention), and mental health indicators. Education restrictions (ie, the pandemic-related measure of half-time remote education) were included as a covariate because they might have influenced engagement with the intervention by reducing live peer

XSL•FO
RenderX

interactions. The results of the full models are added in Multimedia Appendix 2. Variables that were significantly associated with the clusters (ie, age, intention, self-perception, peer support, depressed feelings, and education restrictions) were standardized and included in the final analysis.

Predicting Health Outcomes Based on Engagement Style

For each behavioral outcome, the *lmerTest* R package [39] was used to fit a linear mixed model with random intercepts for participant and school class to test the interaction between time (preintervention vs postintervention) and cluster. For the analysis of physical activity, 1 outlier was detected by inspecting residual plots and was left out of the analysis.

Results

Participant Characteristics

Sample characteristics are shown in Table 2. The participants were evenly divided over the first, second, and third year of secondary education. The proportion of participants following the vocational education track was comparable with the 24.5% of the overall Flemish population in 2020‐2021 [43]. Distribution of family affluence into low, middle, and high affluence was comparable with the general Flemish population [32]. See the study by Peuters et al [22] for sample characteristics of mental health and lifestyle behaviors.

Table . Sample descriptives (N=159).

Characteristics	Values, n (%)
Gender	
Girl	82 (51.6)
Boy	75 (47.2)
Other	2 (1.3)
Education type	
General or technical	120 (75.5)
Vocational	39 (24.5)
Grade	
First year	51 (32.1)
Second year	56 (35.2)
Third year	52 (32.7)
Family affluence	
Low	47 (29.6)
Middle	77 (48.4)
High	35 (22.0)
Language at home	
Dutch	129 (81.1)
Other	30 (18.97)

Describing Engagement

Usage Over Time

The components that were used the most were the action and coping planning and the Fitbit, followed by the gamification and the chatbot. Users visited the narrative section in the app only a few times, but it should be taken into account that weekly only 1 new episode of the narrative appeared in the app. Behavioral engagement with the app and Fitbit drastically decreased after the first week (see Multimedia Appendix 3 for boxplots of the usage frequency of the different intervention components). Of the participants who installed the app, 76.1% (121/159) never opened the app in the last half of the intervention period. Continued usage with the Fitbit was also low, with 42.8% (68/159) of the participants who never synchronized the Fitbit with the Fitbit app in the second half of the intervention. Experiential engagement was more stable over time (Multimedia Appendix 3). Average experiential engagement at the end of the first week was positive with the app (mean 3.55, SD 0.71) and with the narrative component (mean 3.68, SD 0.68) and dropped slightly by the end of weeks 6 and 12. Of the participants who completed the postsurvey, 25.8% (38/147) indicated that they had not really used the app. The most common reasons for not or no longer using the app after week 1, week 6, or week 12 are shown in Table 3, together with the number of participants reporting this reason.

XSL•FO
RenderX

Table . Participant reasons for low usage of the #LIFEGOALS app after week 1, week 6, and week 12.

Reason for nonusage	Week 1[a], n (%)	Week 6[b], n (%)	Week 12[c], n (%)
Low ease of use (app was unclear or complicated to use)	4 (20.0)	3 (15.8)	7 (18.4)
Lack of time (having no time or too little time to use the app)	3 (15.0)	7 (36.8)	4 (10.5)
Forgotten to use the app	5 (25.0)	2 (10.5)	6 (15.8)
Technical problems (crashes, slow, or not being able to install)	4 (20.0)	1 (5.3)	6 (15.8)
No interest or not useful (no motivation, no need to use the app)	1 (5.0)	3 (15.8)	6 (15.8)
App unattractive (was no fun or boring)	0 (0.0)	0 (0.0)	3 (7.9)
Fitbit app was easier to use than the #LIFEGOALS app	1 (5.0)	0 (0.0)	1 (2.6)
Other (eg, did not know why)	2 (10.0)	3 (15.8)	7 (18.4)

[a]n=20; missing=14.5% (23/159).

[b]n=19; missing=23.9% (38/159).

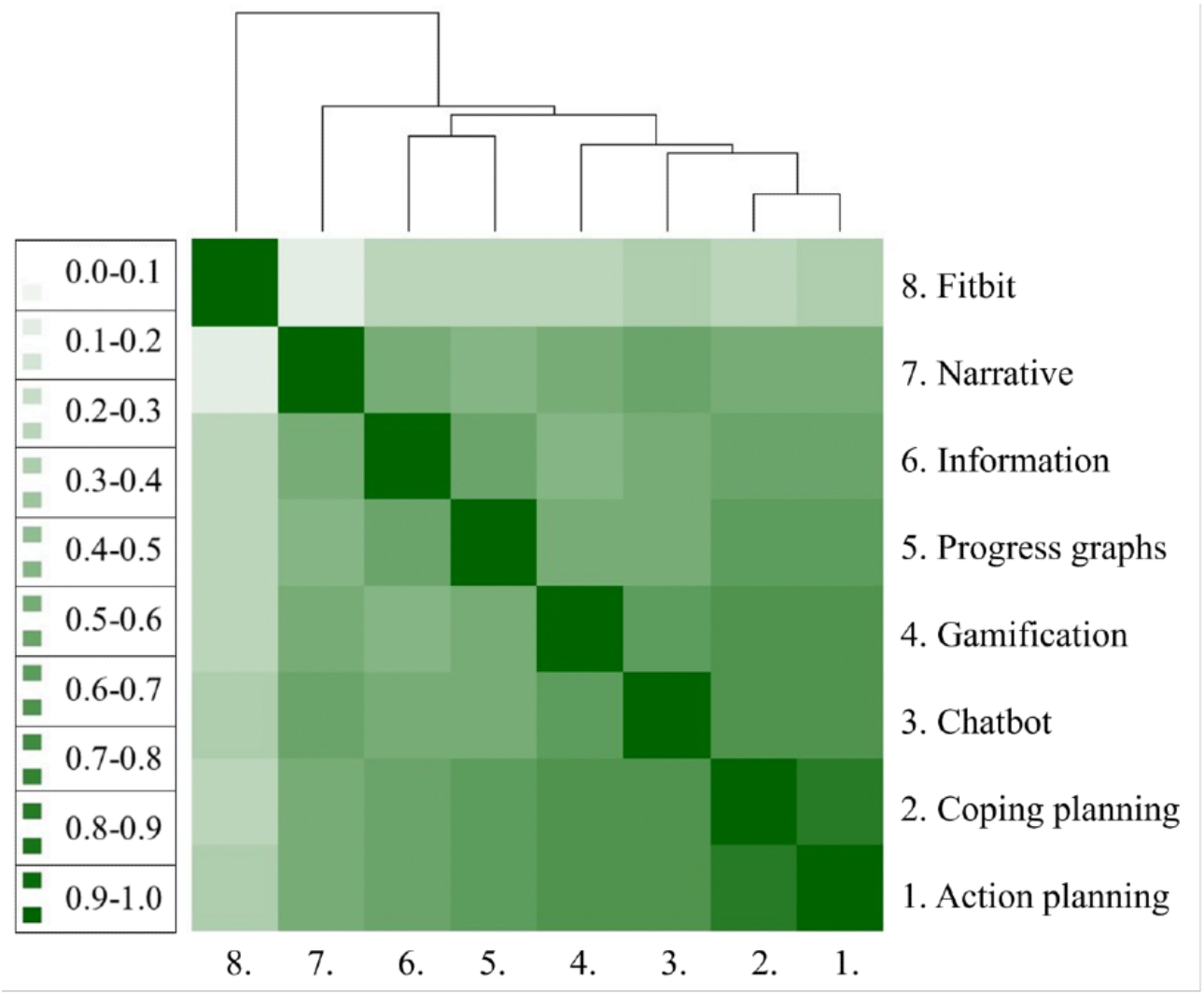
[c]n=38; missing=7.5% (12/159).

Associations Between Intervention Components

Interdependence between the intervention components was calculated via Spearman rank order correlations between the frequencies of usage (Multimedia Appendix 3), and is shown in an association heat map (Figure 2). The frequency with which the action and coping planning were used was highly associated with the usage of the gamification and the chatbot. The interrelationship between usage of the progress graphs and information was higher than their association with the rest of the intervention. The number of days the Fitbit was used was independent from the overall app use frequency.

XSL•FO
RenderX

Figure 2. App component associations. Heat map (rank-order correlations) and dendrogram of the frequency of usage of the different app components. Darker colors represent stronger positive associations.

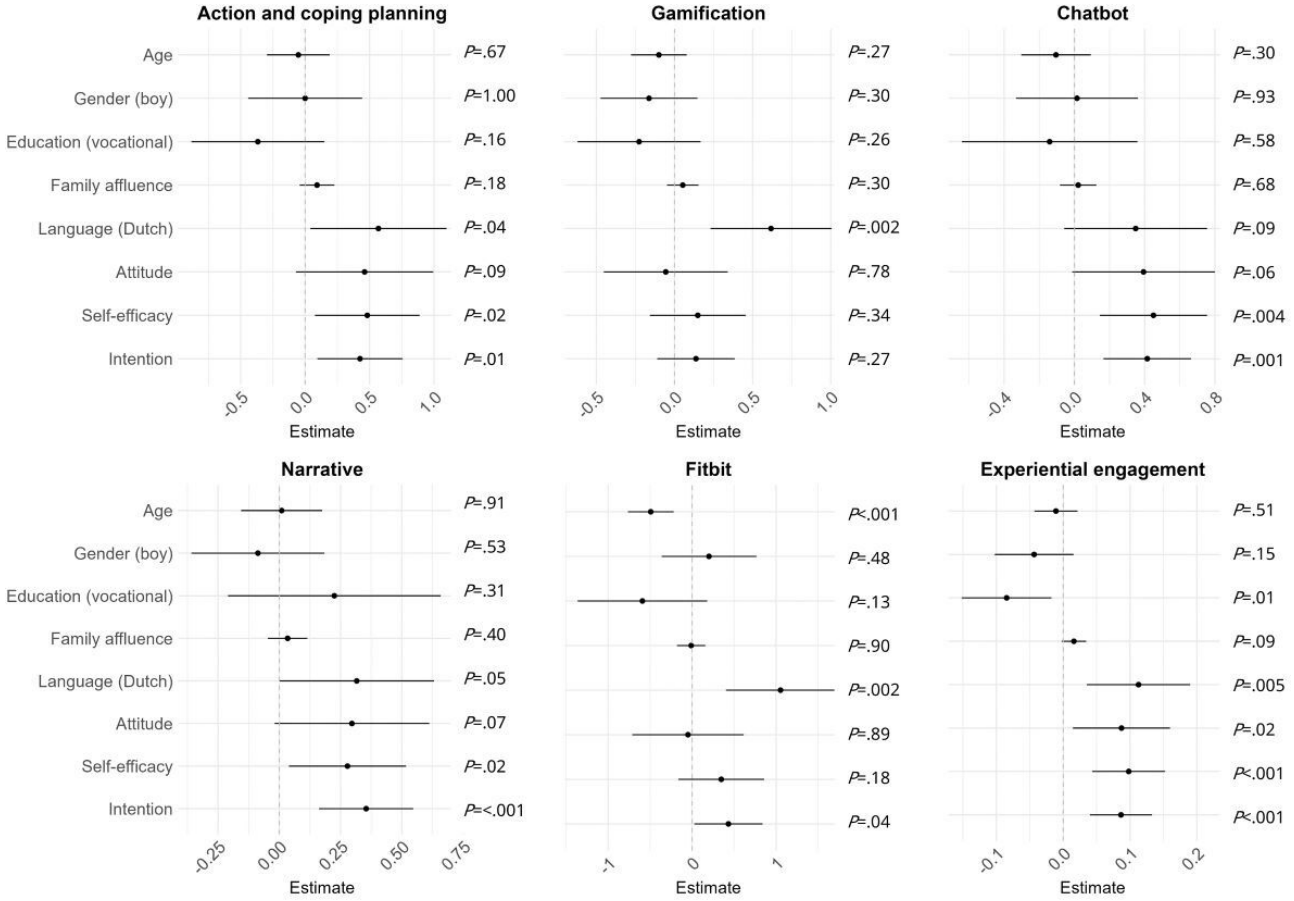


Exploring Determinants of Engagement

For each participant characteristic and baseline behavioral determinant, we ran a separate univariate linear mixed model including school class as random intercept to predict the engagement indicators. Results are shown in Figure 3, showing the estimates with 95% CIs and corresponding *P* values. Descriptive statistics for the engagement indicators across levels of the different determinants are shown in Multimedia Appendix 4. Fitbit usage varied as a function of age, with older adolescents using the Fitbit for fewer days. Age did not influence the other engagement indicators. Participants who did not speak Dutch as their primary language at home exhibited lower engagement with all components of the intervention. Gender, education type, and family affluence did not predict engagement. A positive attitude toward a healthier lifestyle, self-efficacy to change health behaviors, and the intention to change health behaviors were associated with higher usage of the action and coping planning, the chatbot, and the narrative components. However, for the gamification and Fitbit components, attitude and self-efficacy did not predict usage and intention to a lesser extent. Regarding experiential engagement with the app, participants in the general or technical education track, as well as those who spoke Dutch as their primary language at home, provided more positive experiential evaluations than participants in the vocational track or those speaking another maternal language. Additionally, higher attitude, self-efficacy, and intention were linked to more positive reporting of experiential engagement.

XSL•FO
RenderX

Figure 3. Exploratory analysis of participant characteristics predicting engagement indicators using univariate linear mixed models (1 predictor per model) with school class as a random intercept (N=159). Frequency outcomes were log-transformed. Estimates shown with 95% CIs and *P* values.



Identifying Engagement Styles

Cluster solutions with 3, 4, or 5 clusters had comparable average silhouette width (Figure 4). Exploring these cluster solutions, it was observed that different clusters were formed based on the extent to which the Fitbit was used, the narrative was watched, and the rest of the app was used. There was no clear separation between clusters according to the extent of usage of the chatbot. Because a 5-cluster solution had the greatest silhouette width but had 1 cluster that was considered too small for follow-up analyses (n=9), the 4-cluster solution was chosen.

Median usage of the intervention components per cluster is shown in Figure 5. The 4 clusters were labeled according to what distinguishes them from each other. Participants classified in the “narrative usage” cluster (n=19) were characterized by having watched multiple episodes of the narrative, those in the “Fitbit usage” cluster (n=32) by a relatively high and sustained usage of the Fitbit, those in the “app usage” cluster (n=36) had used the intervention but with less explicit usage of the Fitbit or narrative, and participants classified in the “no usage” cluster (n=72) had overall low to no usage of the intervention.

XSL•FO
RenderX

Figure 4. Average silhouette width for different numbers of clusters.

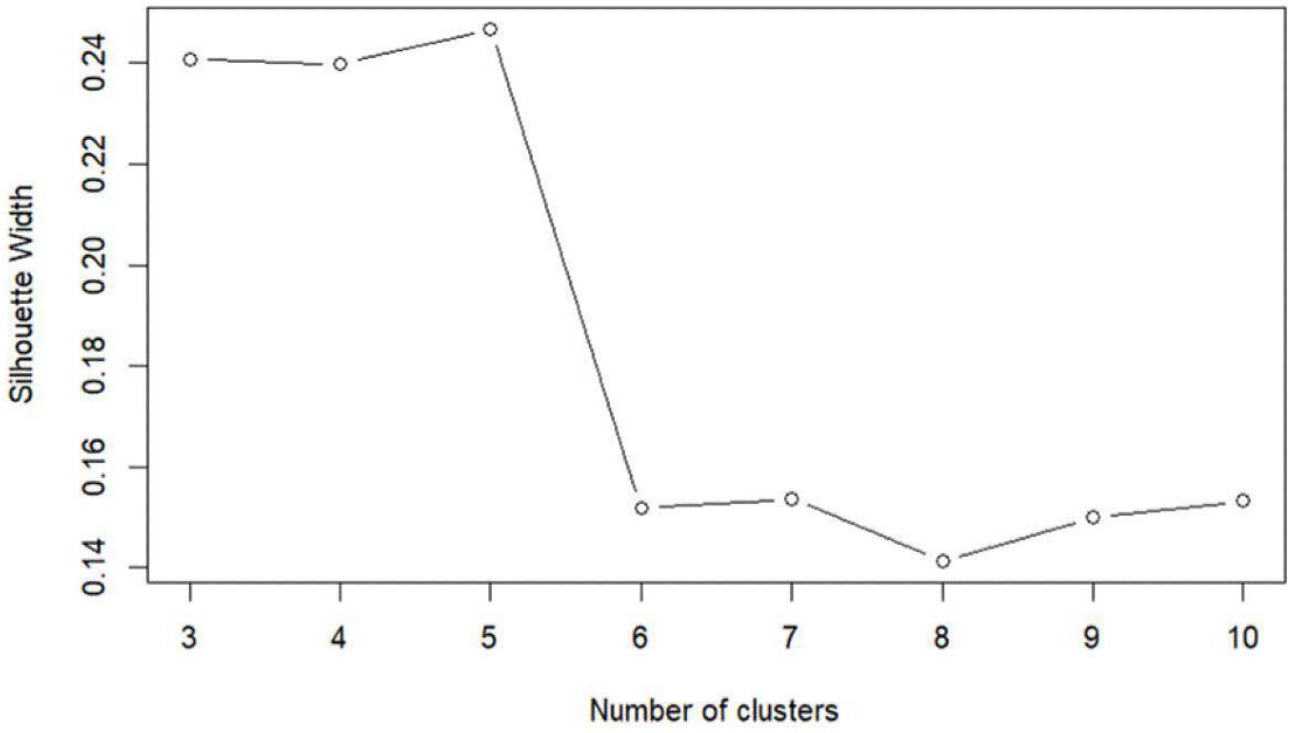
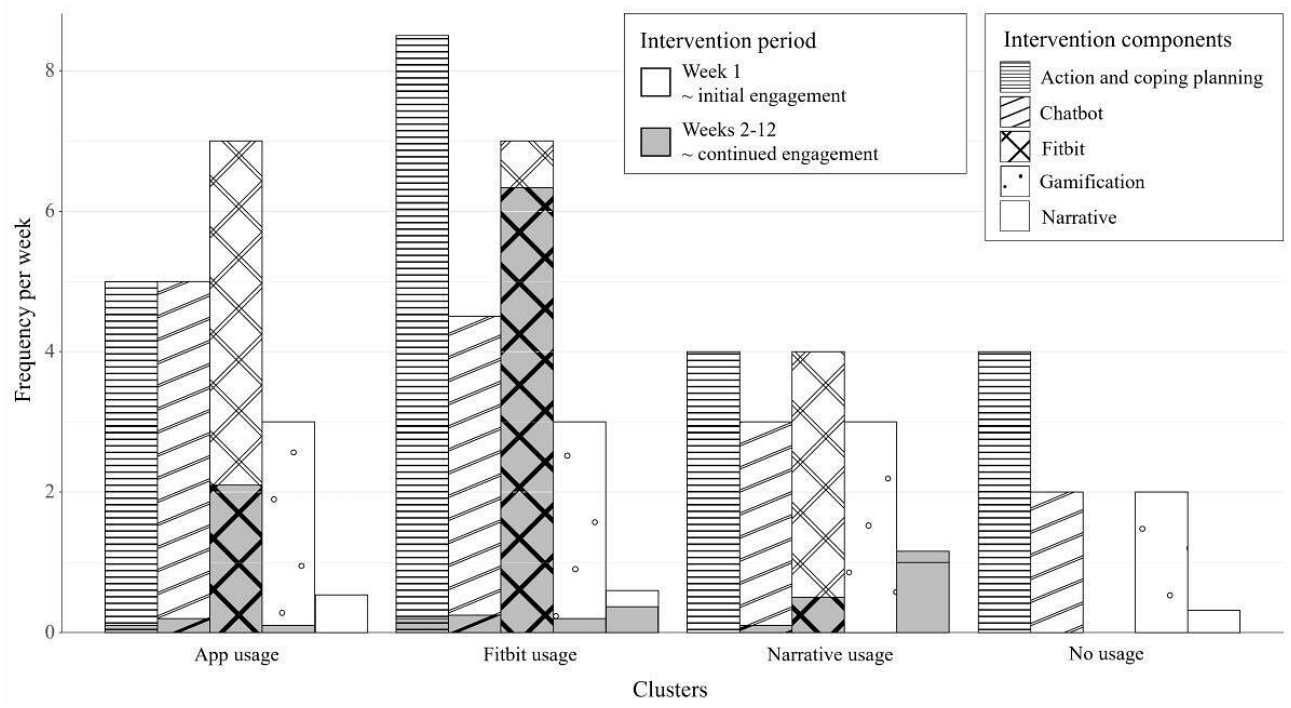


Figure 5. Average usage of the main intervention components per engagement style. Usage is presented separately for week 1 (white bars) and the weekly average of weeks 2-12 (gray bars). Presented are the median of the frequency per week using the component or the days per week wearing the Fitbit. For the narrative, the mean was calculated on the self-reported number of watched episodes after week 1 (0=none, 1=1, or 2=all 2) and after week 12 (0=less than 3, 1=3-11, and 2=all 12).



Determinants of Engagement Style

Statistics of the final multinomial logistic regression model are shown in Multimedia Appendix 2 and the statistically significant associations are reported here. Younger adolescents were more likely to be in the Fitbit usage group than in the no usage group (odds ratio [OR] 0.30, 95% CI 0.14 - 0.65; *P*=.002). Participants who reported lower peer support were more likely to be in the app usage group than in the no usage group (OR 0.43, 95% CI 0.25 - 0.74; *P*=.002) and more likely to be in the Fitbit usage group than in the app usage group (OR 1.96, 95% CI 1.08 - 3.55; *P*=.03). Participants who reported more depressed feelings were more likely to be in the no usage group than in the app usage group (OR 0.33, 95% CI 0.17 - 0.66; *P*=.002). No other baseline characteristics were significant predictors of engagement cluster.

XSL•FO
RenderX

Predictive Value of Engagement Style

The changes in the health behaviors from preintervention to postintervention did not differ significantly between clusters (Table 4).

Table . Effects of engagement cluster on preintervention to postintervention change in health behaviors[a].

	N	Chi-square (*df*)	*P* value
Physical activity, ENMO[b] (mg)	147	1.7 (3)	.63
Sedentary time (minutes)	148	0.9 (3)	.82
Sleep regularity (minutes)	132	1.5 (3)	.68
Sleep quality (scale 1 - 5)	159	2.0 (3)	.57
Breakfast (days per week)	159	4.6 (3)	.21

[a]Linear mixed models with subject and school class as random intercepts.

[b]ENMO: Euclidean Norm Minus One.

Discussion

Differential Engagement Between Components Over Time

This study investigated engagement with the multicomponent mHealth intervention #LIFEGOALS. #LIFEGOALS consisted of an app coupled with an activity tracker and included self-regulation guidance, gamification elements, health information, a support chatbot, and narrative videos. We aimed to gain insight into how the intervention is used, how components relate to each other, which users apply which engagement style, and how engagement style relates to health outcomes. The combination of system usage and questionnaire data allowed for the investigation of both behavioral and experiential engagement.

On average, users rated their initial experience when using the app as positive. The components that were used the most were the action and coping planning and the Fitbit, followed by the gamification and the chatbot. In the first week, many users used nearly all the intervention components (ie, self-regulation techniques, gamification, chatbot, and Fitbit). However, overall usage dropped steeply after the first week. The decrease in usage was least observed in the number of days wearing the Fitbit, although the use of the Fitbit decreased over time. This is in line with previous research showing that providing a Fitbit is not enough to achieve sustained usage [44]. It may well be that initial engagement is exploratory in nature, given that for all components, median usage was the highest in the first week, after which the extent of usage decreased. The rating of experiential engagement (ie, interest, affect, attention, and usefulness) slightly decreased only after the first week, potentially indicating that mHealth acceptability does not predict sustained mHealth usage. However, common self-reported reasons for low usage of the #LIFEGOALS intervention were low ease of use, technical problems, next to lack of time, forgetting, or no interest in the app. These findings confirm previous research highlighting the importance of technical quality and ease of use for mHealth engagement [45,46]. This might particularly play a role for adolescents in the vocational education track. They, along with participants whose maternal language was not Dutch, rated their experience while using the app less positively than participants in the general or technical education track. Another paper on this trial, moreover, found adolescents in the vocational education track to be more likely to drop out from using the intervention [13]. Overall, health is not a main concern for many adolescents, and vocational students may be less motivated to change health behaviors [13]. During data collection, it was indeed observed that vocational students showed lower motivation to participate in the study and to use the app than participants in the general track. A lower education level and Dutch language proficiency may have also contributed to greater difficulties in using the app. One proposed strategy for tailoring interventions to individuals with lower education attainment is to provide a support person with whom they can have direct contact [13].

Furthermore, exploration of engagement with #LIFEGOALS provided insight into the popularity and interconnectedness of specific intervention components. Engagement was the highest with the intervention strategies intended to enhance *self-regulation*. That is, action and coping planning were visited the most in the app, and sustained usage was the highest with the self-monitoring device (ie, the Fitbit). These components are theoretically considered important for behavior change, as has been shown in other studies [47,48]. Furthermore, usage of the action and coping planning was moderately to highly associated with usage of the *gamification*, suggesting that the gamification succeeded in its role to motivate users to set and complete action plans. Especially for an adolescent population, gamification has been put forward as an essential element within mHealth [46,49]. Results suggest that a points system with avatar appears to be an effective combination of gamification techniques. Notably, baseline attitude toward a healthy lifestyle and intentions to change health behaviors were not associated with subsequent use of the gamification features or the Fitbit. However, they did influence engagement with other components of the intervention. This pattern may suggest that reward-based elements, such as gamification and Fitbit features, have the potential to engage adolescents with lower behavioral motivation. In contrast, components such as the chatbot and narrative, which could also be considered motivational, may have less potential in this regard. There was a particularly low usage of the narrative video episodes, as well as with the tips with information for healthy lifestyles. In #LIFEGOALS, the gamification features were not designed to stimulate engagement with the health narrative or information in the app, which could

XSL•FO
RenderX

partly explain the low usage of these elements. Postintervention interviews with users suggested that tailored advertisement could improve engagement with the narrative videos [22]. Future research is needed to formulate an answer to whether increased embeddedness in the app, linkage with gamification, or other techniques could push engagement with health narrative video episodes as part of mHealth.

Engagement Styles

Cluster analyses using multiple indicators of engagement classified users according to the extent of overall usage and to the extent of usage of different intervention components. Besides a no usage cluster, the remaining clusters were characterized by relatively high usage of the health narrative, high usage of the Fitbit, or no explicit usage of the narrative or Fitbit. The identification of different clusters depending on the extent of usage of a certain intervention component was also found in other studies investigating clusters of engagement with a multicomponent intervention [17,20]. Experiential engagement did not influence the clustering in this study. It might be that experiential engagement is reflected in the amount of usage (considering, eg, that the experience of flow implies a loss of track of time, which will translate in longer in-app duration), thus explaining little additional variance on which to base cluster grouping. Indeed, previous research has indicated significant correlations between self-reported experiential engagement and logged behavioral engagement [28]. In what follows, we discuss the 4 engagement styles that were distinguished in this study, together with their determinants.

A small group of users (19/159, 12%) was classified with a *narrative usage* style, characterized by the relatively high number of health narrative episodes watched. They all reported having watched 3 or more (out of 12) episodes, did not consistently wear the Fitbit, and had average usage of the self-regulation techniques and chatbot. Based on theories on narrative persuasion, we expected that adolescents with low motivation and ability to cognitively elaborate on how to live healthier would more easily be persuaded by a health message embedded in a narrative [26]. We found no evidence that low motivation or low ability increased the likelihood of watching the narrative: low intention to live healthier (as a proxy for motivation) or vocational education type (indicative of lower ability for cognitive processing) did not increase the odds of being in the narrative usage cluster. In contrast, exploratory linear regressions showed greater narrative usage among participants with higher intention. It could be that motivation and ability do not predict who will watch the narrative but rather who would be more easily influenced by watching the narrative. Future studies with a larger sample size are needed to investigate this question.

The group of users classified with a *Fitbit usage* engagement style (32/159, 20% of the participants) wore the Fitbit nearly every day of the 12 weeks. Their usage of the app ranged from a little exploration of some components to high initial and sustained engagement with the entire app. The probability of engaging with a Fitbit usage style compared with a no usage style decreased with age (OR=0.30). Evidence indicates that although adolescents perceive the Fitbit as easy to use, they also consider it as effort-demanding [44,50]. It could be that young adolescents are more easily entertained by the Fitbit features, hence outweighing the effort required to use the Fitbit. Participants who reported higher perceived peer support were 96% more likely (OR=1.96) to continuously use the Fitbit (Fitbit usage style) than the app components (app usage style). It could be that supportive peers encouraged usage of the Fitbit by providing the opportunity to compare progress and achievements or to use the device together. Indeed, social support has been found to facilitate positive change in physical activity [51,52].

A quarter of the participants (36/159, 23%) showed usage of most or all the intervention components and were categorized in the *app usage* engagement style. The probability of engaging with an app usage style compared with a no usage style differed according to peer support and depressed feelings. Adolescents who perceived higher peer support were less likely to engage with an app usage style than with a no usage style (OR=0.42) or a Fitbit usage style (OR=0.51). It could be that the experience of sufficient support in one's social network makes a health app seem redundant. Also, participants reporting more depressed feelings were less likely to engage with an app usage style than with a no usage style (OR=0.34). Individuals with a low mood may lack the energy required to engage with the app and have been found to engage less with mHealth also in other studies [45].

The largest cluster was the *no usage* cluster, consisting of 45% (72/159) of the participants who had not or rarely used any component of the intervention after the first week. Previous studies analyzing clusters of engagement have also identified 1 group with very low to no usage [15-17,20]. Not using the intervention should not cause concern if this reflects that the individual already meets the objectives of the intervention [53]. Individuals who already perform the behaviors that are targeted by the mHealth intervention, for instance, may perceive lower usefulness and use the intervention to a lesser extent than individuals who are aware that they could make important health behavior improvements. Baseline descriptive statistics indeed indicated that a part of the sample already maintained a healthy lifestyle. However, healthier behaviors at baseline did not significantly increase the odds of a no usage engagement style, but this might be due to low statistical power given the many missing accelerometry data and relatively small sample sizes of the clusters. It is an interesting avenue to further investigate baseline levels of health behaviors as predictors of engagement style, as this could also provide an answer to the question whether the intervention reaches the target group with the highest need.

Effective Engagement

To gain insight into the characteristics of engagement that are most effective for achieving intervention effects, we analyzed whether health behavior change differed according to engagement style. Our results did not show an effect of engagement style on change in physical activity, sedentary behavior, sleep duration, sleep quality, or frequency of breakfast consumption. Among others, the overall low engagement with #LIFEGOALS may have complicated the detection of differences in health effects between engagement styles, given

XSL•FO
RenderX

that also the clusters other than the no usage cluster contained many participants with little overall usage and low continued use. Low engagement, however, does not necessarily imply a failure of the intervention to reach health effects. Usage of the intervention components (ie, behavioral engagement) and the experience while using the intervention (ie, experiential engagement) should lead to engagement with the behavior change processes that are related to the intervention goals [54]. A decrease in engagement with the intervention may then be a positive phenomenon if this reflects continued engagement with the behavioral processes. This could provide an explanation of why there exists inconsistency on whether more sustained engagement is related to greater intervention effects [15,55]. To understand "effective engagement," more studies on different mHealth contexts are needed to clarify the type and amount of engagement that are minimum required for reaching health effects.

Limitations

System usage data are a complex entity, introducing challenges to adequately register and extract engagement indicators. In our study, all registered activity in the app or with the activity tracker was assigned to the participant who logged in on these devices during installation, for which we cannot rule out that these logged data may have included activity of other individuals using those devices. Furthermore, for the duration spent in the app, we detected an error for defining the end of a session (ie, in specific circumstances, no end duration was logged). Therefore, frequency data were used instead. There are multiple ways to operationalize behavioral engagement, and we could have also opted for the intensity of use (eg, the rate of visited pages or the number of action plans created) [56]. Because different indicators of behavioral engagement might reveal different effects [18], the multiple ways to operationalize the usage of mHealth components make it difficult to compare log usage data between studies. Next, experiential engagement was measured by self-report ratings of interest, affect, attention, and usefulness. It might be that this measure did not adequately reflect individual variance in the concepts it aims to measure, which could be why experiential engagement did not influence clustering into engagement styles in this study. Future research could consider alternative measures of experiential engagement, such as ecological momentary assessment or questionnaires with more tangible items (eg, ease of use and satisfaction as in the USE Questionnaire [57]) rather than relying on more abstract items such as "I was focused when I used the app," which may be difficult for young adolescents to self-report retrospectively. Additional research is needed to investigate how best to measure the user experience when interacting with the technology and its additional value for understanding mHealth engagement. Furthermore, the data collection took place during the COVID-19 pandemic, but for the behavior change predictions, we did not control for COVID-19–related measures due to the small sizes of the subsamples of engagement styles. Generalization of the findings to a nonpandemic situation should be done with caution, given that the effect study of #LIFEGOALS showed moderation by pandemic-related restrictions for education and sports [22]. Moreover, the elevated missing data caused by using accelerometers to assess activity-related behaviors limited the statistical power to study health behaviors as determinants or outcomes of engagement styles. Overall, the low usage of the intervention limited the possibilities to identify engagement styles and detect differences between styles. A high percentage of nonusage is also observed in other mHealth studies [5,12,58], but several participants mentioned low ease of use, and the intervention might indeed have lacked in-app guidance and tutorials on what to do in the app. The overall look and feel of apps have been found to be an important factor for adolescents to adopt health apps [59], and the design and technical functioning of #LIFEGOALS were likely of lower quality than those of commercial apps. The high technical requirements of end users challenge the possibilities to investigate mHealth engagement in noncommercial research contexts.

Future Directions and Conclusions

This study was the first, to our knowledge, to focus specifically on user styles of engagement with mHealth among adolescents. Four engagement styles were identified: narrative usage, Fitbit usage, app usage, and no usage. The styles differ in extent of usage and in preferences for certain intervention components. For instance, adolescents with lower perceived social support showed higher engagement with the app, while depressed feelings, consistent with prior research, were inversely associated with mHealth usage. Interestingly, Fitbit usage was independent of app usage, with younger adolescents more likely to use the Fitbit than the older ones. Moreover, our results reaffirm the important role of social support in sustaining engagement with wearable activity trackers [51]. Interventions designed to promote physical activity are likely to benefit from incorporating interactive components to foster continued physical activity.

Furthermore, our findings support previous research emphasizing the importance of technical quality and intuitive design in fostering mHealth engagement [45,46]. These aspects may be particularly crucial for adolescents with lower education levels or from immigration backgrounds, who are already at higher risk for unhealthy behaviors. To avoid leaving behind those who might benefit the most, digital health interventions must ensure optimal navigability and ease of use. For these populations, offering in-person support may further improve accessibility and uptake [13].

Moreover, our results provide support for the inclusion of reward-based elements and self-regulation. The findings from our exploratory analyses suggest that adolescents' attitudes, self-efficacy, and intentions toward healthy living may influence engagement with mHealth behavior change interventions. In primary prevention, where intrinsic motivation to change behavior may be low, reward-based elements such as gamification and Fitbit features appear especially promising. In our study, engagement with these components did not depend on behavioral motivation, while our findings suggest that reward-based components can engage users to set and complete action plans. In contrast, components such as the chatbot and narrative, although potentially motivational, may be less effective in engaging less motivated users. Self-regulation techniques, previously shown to be effective in promoting

XSL•FO
RenderX

behavior change [47,48], were used most frequently in our study, underscoring their value in mobile behavior change interventions. The identified variations in usage of the mHealth intervention across individual characteristics underscore the importance of tailoring mHealth to the needs and motivations of diverse populations in order to ensure accessibility and sustained engagement.

Our results encourage the continuation of upcoming research aiming to tailor mHealth interventions to user profiles. Addressing the challenge of generally low engagement of mHealth remains crucial, as higher usage could provide more robust findings about the type and duration of engagement necessary to exert meaningful health effects. It can then be investigated for whom and under which circumstances a nudging toward a promising engagement style may be both feasible and beneficial.

Acknowledgments

The authors would like to thank Armand De Clercq for the technical support in gathering the log usage data, and all participants for their participation in the study. This research was conducted with the support of the FWO (grant G043319N) to Geert Crombez and Ilse De Bourdeaudhuij and a grant for the development and evaluation of the method MOV-E-STAR of the Flemish Government to Ghent University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Indicators of engagement included in the cluster analyses.
[DOCX File, 16 KB - mhealth_v13i1e59041_app1.docx]

Multimedia Appendix 2
Multinomial logistic regressions of determinants of engagement style.
[DOCX File, 38 KB - mhealth_v13i1e59041_app2.docx]

Multimedia Appendix 3
Descriptive statistics and correlation matrix of engagement with the different intervention components.
[DOCX File, 6275 KB - mhealth_v13i1e59041_app3.docx]

Multimedia Appendix 4
Exploring determinants of engagement.
[DOCX File, 39 KB - mhealth_v13i1e59041_app4.docx]

Checklist 1
CONSORT e-HEALTH (V1.6.1) checklist.
[PDF File, 239 KB - mhealth_v13i1e59041_app5.pdf]

References

1. Khosravi M, Azar G. A systematic review of reviews on the advantages of mHealth utilization in mental health services: a viable option for large populations in low-resource settings. Glob Ment Health (Camb) 2024;11:e43. [doi: 10.1017/gmh.2024.39] [Medline: 38690573]
2. Messner EM, Probst T, O'Rourke T, Stoyanov S, Baumeister H. MHealth applications: potentials, limitations, current quality and future directions. In: Baumeister H, Montag C, editors. Digital Phenotyping and Mobile Sensing: Springer International Publishing; 2019:235-248. [doi: 10.1007/978-3-030-31620-4_15]
3. Zapata-Lamana R, Lalanza JF, Losilla JM, Parrado E, Capdevila L. mHealth technology for ecological momentary assessment in physical activity research: a systematic review. PeerJ 2020;8:e8848. [doi: 10.7717/peerj.8848]
4. Kaufman MR, Levine D, Casella A, DuBois DL. E-mentoring to address youth health: a systematic review. Adolesc Res Rev 2022;7(1):63-78. [doi: 10.1007/s40894-021-00172-3] [Medline: 34568546]
5. Dute DJ, Bemelmans WJE, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: a review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 5;4(2):e39. [doi: 10.2196/mhealth.3559] [Medline: 27150850]
6. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307. [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]

XSL•FO
RenderX

7. Children and parents: media use and attitudes 2023. Ofcom. 2023. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0027/255852/childrens-media-use-and-attitudes-report-2023.pdf [accessed 2025-4-20]
8. Pew Research Center. Teens, social media & technology 2018. 2018 May 31 URL: https://apo.org.au/node/175701 [accessed 2025-04-20]
9. Vanwynsberghe H, Joris G, Waeterloos C, et al. Onderzoeksrapport apestaartjaren: de digitale leefwereld van kinderen en jongeren [Research report apestaartjaren. The digital world of children and youth]. : Mediaraven: Gent; 2022 URL: https://www.mediawijs.be/nl/onderzoek/apenstaartjaren [accessed 2025-04-20]
10. Leech T, Dorstyn D, Taylor A, Li W. Mental health apps for adolescents and young adults: a systematic review of randomised controlled trials. Child Youth Serv Rev 2021 Aug;127:106073. [doi: 10.1016/j.childyouth.2021.106073]
11. Shin Y, Kim SK, Lee M. Mobile phone interventions to improve adolescents' physical health: a systematic review and meta-analysis. Public Health Nurs 2019 Nov;36(6):787-799. [doi: 10.1111/phn.12655] [Medline: 31456259]
12. Stassen G, Grieben C, Froböse I, Schaller A. Engagement with a web-based health promotion intervention among vocational school students: a secondary user and usage analysis. Int J Environ Res Public Health 2020 Mar 25;17(7):2180. [doi: 10.3390/ijerph17072180] [Medline: 32218251]
13. Maenhout L, Peuters C, Cardon G, Crombez G, DeSmet A, Compernolle S. Nonusage attrition of adolescents in an mHealth promotion intervention and the role of socioeconomic status: secondary analysis of a 2-arm cluster-controlled trial. JMIR Mhealth Uhealth 2022 May 10;10(5):e36404. [doi: 10.2196/36404] [Medline: 35536640]
14. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
15. Wu D, An J, Yu P, et al. Patterns for patient engagement with the hypertension management and effects of electronic health care provider follow-up on these patterns: cluster analysis. J Med Internet Res 2021 Sep 28;23(9):e25630. [doi: 10.2196/25630] [Medline: 34581680]
16. Perski O, Watson NL, Mull KE, Bricker JB. Identifying content-based engagement patterns in a smoking cessation website and associations with user characteristics and cessation outcomes: a sequence and cluster analysis. Nicotine Tob Res 2021 Jun 8;23(7):1103-1112. [doi: 10.1093/ntr/ntab008] [Medline: 33433609]
17. O'Sullivan S, Schmaal L, D'Alfonso S, et al. Characterizing use of a multicomponent digital intervention to predict treatment outcomes in first-episode psychosis: cluster analysis. JMIR Ment Health 2022 Apr 7;9(4):e29211. [doi: 10.2196/29211] [Medline: 35389351]
18. Donkin L, Hickie IB, Christensen H, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231. [doi: 10.2196/jmir.2771] [Medline: 24135213]
19. Pham Q, Graham G, Carrion C, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019;7(1):e11941. [doi: 10.2196/11941]
20. Sanatkar S, Baldwin PA, Huckvale K, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019;21(11):e14728. [doi: 10.2196/14728]
21. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
22. Peuters C, Maenhout L, Cardon G, et al. A mobile healthy lifestyle intervention to promote mental health in adolescence: a mixed-methods evaluation. BMC Public Health 2024 Jan 2;24(1):44. [doi: 10.1186/s12889-023-17260-9] [Medline: 38166797]
23. Gautam N, Dessie G, Rahman MM, Khanam R. Socioeconomic status and health behavior in children and adolescents: a systematic literature review. Front Public Health 2023;11:1228632. [doi: 10.3389/fpubh.2023.1228632]
24. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
25. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24. [doi: 10.17705/1CAIS.02428]
26. Slater MD, Likelihood E. Entertainment-education and elaboration likelihood: understanding the processing of narrative persuasion. Commun Theory 2002;12(2):173-191. [doi: 10.1111/j.1468-2885.2002.tb00265.x]
27. Peuters C, Maenhout L, Crombez G, Cardon G, DeSmet A. Development of a mobile healthy lifestyle intervention for promoting mental health in adolescence: #LIFEGOALS. OSF 2022. [doi: 10.17605/OSF.IO/PWUEH]
28. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019;5:2055207619880676. [doi: 10.1177/2055207619880676] [Medline: 31620306]
29. D'Haese S. Insights Into Individual and Environmental Correlates of Physical Activity Among Youth [Doctoral dissertation]. : Ghent University; 2016 URL: http://hdl.handle.net/1854/LU-7160265 [accessed 2025-4-20]

XSL•FO
RenderX

30. Vandendriessche A, Verloigne M, Boets L, et al. Factoren die samenhangen met slaapgedrag bij adolescenten: een focusgroeponderzoek in Vlaanderen [Factors associated with sleep behavior in adolescents: a focus group study in Flanders]. JGZ Tijdschr Voor Jeugdgezondheidsz 2020;52(5-6):125-131. [doi: 10.1007/s12452-020-00221-8]
31. Phillips LRS, Parfitt G, Rowlands AV. Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. J Sci Med Sport 2013 Mar;16(2):124-128. [doi: 10.1016/j.jsams.2012.05.013] [Medline: 22770768]
32. Inchley JC, Stevens G, Samdal O, Currie DB. Enhancing understanding of adolescent health and well-being: the health behaviour in school-aged children study. J Adolesc Health 2020 Jun;66(6S):S3-S5. [doi: 10.1016/j.jadohealth.2020.03.014] [Medline: 32446607]
33. Essner B, Noel M, Myrvik M, Palermo T. Examination of the factor structure of the Adolescent Sleep-Wake Scale (ASWS). Behav Sleep Med 2015;13(4):296-307. [doi: 10.1080/15402002.2014.896253] [Medline: 24742264]
34. Ravens-Sieberer U, Erhart M, Rajmil L, et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. Qual Life Res 2010 Dec;19(10):1487-1500. [doi: 10.1007/s11136-010-9706-5] [Medline: 20668950]
35. Ravens-Sieberer U, Auquier P, Erhart M, et al. The KIDSCREEN-27 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Qual Life Res 2007 Oct;16(8):1347-1356. [doi: 10.1007/s11136-007-9240-2] [Medline: 17668292]
36. Ravens-Sieberer U, Gosch A, Rajmil L, et al. The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Value Health 2008;11(4):645-658. [doi: 10.1111/j.1524-4733.2007.00291.x] [Medline: 18179669]
37. Soer R, Six Dijkstra MWMC, Bieleman HJ, et al. Measurement properties and implications of the Brief Resilience Scale in healthy workers. J Occup Health 2019 May;61(3):242-250. [doi: 10.1002/1348-9585.12041] [Medline: 30903648]
38. Irwin DE, Stucky B, Langer MM, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Qual Life Res 2010 May;19(4):595-607. [doi: 10.1007/s11136-010-9619-3] [Medline: 20213516]
39. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: tests in linear mixed effects models. J Stat Soft 2017;82(13). [doi: 10.18637/jss.v082.i13]
40. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: cluster analysis basics and extensions. R package version 2.1.2. 2021. URL: https://cran.r-project.org/web/packages/cluster/index.html [accessed 2022-02-07]
41. Gower JC. A general coefficient of similarity and some of its properties. Biometrics 1971 Dec;27(4):857. [doi: 10.2307/2528823]
42. Elff M. mclogit: multinomial logit models, with or without random effects or overdispersion. 2022. URL: https://CRAN.R-project.org/package=mclogit [accessed 2022-4-20]
43. Statistisch jaarboek van het vlaams onderwijs: schooljaar 2021-2022. Vlaams Ministerie Van Onderwijs En Vorming. 2023. URL: www.vlaanderen.be/publicaties/statistisch-jaarboek-van-het-vlaams-onderwijs [accessed 2025-4-20]
44. Koorts H, Salmon J, Timperio A, et al. Translatability of a wearable technology intervention to increase adolescent physical activity: mixed methods implementation evaluation. J Med Internet Res 2020 Aug 7;22(8):e13573. [doi: 10.2196/13573] [Medline: 32763872]
45. Psihogios AM, King-Dowling S, O'Hagan B, et al. Contextual predictors of engagement in a tailored mHealth intervention for adolescent and young adult cancer survivors. Ann Behav Med 2021 Nov 18;55(12):1220-1230. [doi: 10.1093/abm/kaab008] [Medline: 33674863]
46. Jeminiwa RN, Hohmann NS, Fox BI. Developing a theoretical framework for evaluating the quality of mHealth apps for adolescent users: a systematic review. J Pediatr Pharmacol Ther 2019;24(4):254-269. [doi: 10.5863/1551-6776-24.4.254] [Medline: 31337988]
47. Schroé H, Van Dyck D, De Paepe A, et al. Which behaviour change techniques are effective to promote physical activity and reduce sedentary behaviour in adults: a factorial randomized trial of an e- and m-health intervention. Int J Behav Nutr Phys Act 2020 Oct 7;17(1):127. [doi: 10.1186/s12966-020-01001-x] [Medline: 33028335]
48. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949. [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
49. Bosworth KT, Flowers L, Proffitt R, et al. Mixed-methods study of development and design needs for CommitFit, an adolescent mHealth app. Mhealth 2023;9:22. [doi: 10.21037/mhealth-22-35] [Medline: 37492126]
50. Drehlich M, Naraine M, Rowe K, et al. Using the technology acceptance model to explore adolescents' perspectives on combining technologies for physical activity promotion within an intervention: usability study. J Med Internet Res 2020 Mar 6;22(3):e15552. [doi: 10.2196/15552] [Medline: 32141834]
51. Beets MW, Vogel R, Forlaw L, Pitetti KH, Cardinal BJ. Social support and youth physical activity: the role of provider and type. Am J Health Behav 2006;30(3):278-289. [doi: 10.5555/ajhb.2006.30.3.278] [Medline: 16712442]
52. Kelly SA, Melnyk BM, Jacobson DL, O'Haver JA. Correlates among healthy lifestyle cognitive beliefs, healthy lifestyle choices, social support, and healthy behaviors in adolescents: implications for behavioral change strategies and future research. J Pediatr Health Care 2011;25(4):216-223. [doi: 10.1016/j.pedhc.2010.03.002] [Medline: 21700136]
53. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20. [doi: 10.2196/jmir.8.3.e20]

XSL•FO
RenderX

54. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015]
55. Bricker JB, Sridharan V, Zhu Y, et al. Trajectories of 12-month usage patterns for two smoking cessation websites: exploring how users engage over time. J Med Internet Res 2018;20(4):e10143. [doi: 10.2196/10143]
56. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018;20(11):e292. [doi: 10.2196/jmir.9397]
57. Lund AM. Measuring usability with the USE questionnaire. Usability Interface. 2001. URL: https://tinyurl.com/4wesxuab [accessed 2019-09-09]
58. Chew CSE, Davis C, Lim JKE, et al. Use of a mobile lifestyle intervention app as an early intervention for adolescents with obesity: single-cohort study. J Med Internet Res 2021 Sep 28;23(9):e20520. [doi: 10.2196/20520] [Medline: 34581672]
59. Chan A, Kow R, Cheng JK. Adolescents' perceptions on smartphone applications (apps) for health management. J Mobile Technol Med 2017;6(2):47-55. [doi: 10.7309/jmtm.6.2.6]

Abbreviations

mHealth: mobile health
OR: odds ratio
PROMIS-PedDepSx : Patient-Reported Outcomes Measurement Information System Pediatric Bank Depressive Symptoms v2.0

Edited by L Buis; submitted 31.03.24; peer-reviewed by A Ortega, C Sayegh, C Bowen-Forbes, C Baur; revised version received 22.04.25; accepted 02.06.25; published 18.08.25.

Please cite as:
Peuters C, DeSmet A, Maenhout L, Cardon G, Debeer D, Crombez G
Adolescent Engagement With a Multicomponent mHealth Tool: Identifying Usage Patterns, Determinants, and Health Behavior Change in an Intervention Trial
JMIR Mhealth Uhealth 2025;13:e59041
URL: https://mhealth.jmir.org/2025/1/e59041
doi:10.2196/59041

© Carmen Peuters, Ann DeSmet, Laura Maenhout, Greet Cardon, Dries Debeer, Geert Crombez. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Fit in a Mobile Health Intervention for Alcohol Use Disorder: Qualitative Study

Nora Jacobson[1,2], PhD; Linda S Park[3], PhD; Alice Pulvermacher[3], MS; Samantha Voelker[3], BS; Mallory Herzog[3]; Andrew Quanbeck[3], PhD

[1]Institute for Clinical and Translational Research, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States

[2]School of Nursing, University of Wisconsin-Madison, Madison, WI, United States

[3]Department of Family Medicine and Community Health, School of Medicine and Public Health, University of Wisconsin-Madison, 610 N Whitney Way, STE 200, Madison, WI, United States

Corresponding Author:
Andrew Quanbeck, PhD
Department of Family Medicine and Community Health, School of Medicine and Public Health, University of Wisconsin-Madison, 610 N Whitney Way, STE 200, Madison, WI, United States

Abstract

Background: Implementation frameworks such as the Exploration, Preparation, Implementation, Sustainment model emphasize the importance of the *fit* between an intervention and its context, which includes the needs of its target population, as well as the culture, resources, and capabilities of the implementing organization. Although lack of fit is a major barrier to implementation, fit has not often been a focus of implementation research. This paper uses fit as a lens to examine the implementation of Tula, a mobile health app aimed at reducing risky drinking days among individuals meeting the criteria for mild to moderate alcohol use disorder, in a 3-arm (app alone, app plus peer mentoring, and app plus health coaching) randomized controlled trial.

Objective: We sought to better understand the trial results and to provide actionable guidance for future implementation of digital health interventions in health care organizations.

Methods: Semistructured interviews with 18 trial participants and 7 Tula implementers were conducted. Trial participants were pulled equally from each arm of the trial and represented participants who demonstrated both high and low engagement with the app. Implementers consisted of a project manager, 4 peer mentors, and 2 health coaches. Interviews with participants focused on their motivations, opinions, and experiences of the intervention and their perception of their drinking behavior following the intervention, including how their use of the app worked to change that behavior. Interviews with implementers were centered on their roles, theories of change, perceptions of intervention, and areas for improvement. All interviews were analyzed using rapid qualitative analysis with deductive and inductive components.

Results: We identified areas of both fit and misfit. For example, there was a good fit between implementers' theories of change and participants' description of how change occurred. Fit was improved by the versatility of the app, which allowed participants to customize their experiences. Conversely, misfit was noted in the app's inability to cultivate connection for many participants and a disjunction between the role of peer mentors in the intervention and their broader professional ethos.

Conclusions: Focusing on fit provides a useful guide to enhance future iterations of the Tula app that lead to better sustainment of the intervention.

Trial Registration: ClinicalTrials.gov NCT04011644; https://clinicaltrials.gov/study/NCT04011644

International Registered Report Identifier (IRRID): RR2-10.2196/31109

(JMIR Mhealth Uhealth 2025;13:e65218) doi:10.2196/65218

KEYWORDS

alcohol use disorder; EPIS framework; fit; mHealth; mHealth app; effectiveness; implementation trial; qualitative study; motivations; opinions; drinking behavior

Introduction

Excessive alcohol use is linked to increased injury risk and poor health outcomes, causing approximately 140,000 US deaths and 3.6 million years of potential life lost annually [1,2]. Alcohol use disorder (AUD) spans a spectrum of risk, necessitating interventions tailored to an individual's likelihood of developing severe AUD [3]. Despite its high prevalence and significant impact on health, alcohol use is often undertreated in general medical settings. Many patients with AUD go unrecognized or do not receive appropriate intervention due to time constraints,

XSL•FO
RenderX

limited training, and stigma associated with addressing substance use in primary care [4].

Health coaching is increasingly used in health care organizations as a strategy to promote behavior change, support chronic disease management, and enhance patient engagement. Its integration into clinical care is growing, particularly within primary care and population health initiatives [5]. Peer support models are increasingly integrated into health care systems, particularly within mental health services, to enhance recovery-oriented care and patient engagement. This integration reflects a global shift toward incorporating lived experiences into treatment frameworks to improve outcomes and support system transformation [6].

Improving access to effective interventions is crucial due to AUD's widespread and destructive nature [1,7]. Digital medicine apps show promise in treating and providing ongoing care for patients with AUD, offering a widely accessible means for self-management and clinical monitoring [8]. However, research on implementing these technologies for early intervention in AUD is limited [3]. A key research question remains: how can human support enhance the effectiveness of digital medicine apps [9-11]?

This paper reports results from the qualitative component of a hybrid effectiveness–implementation trial that systematically varied the degree of human touch offered to support use of an evidence-based mobile health (mHealth) intervention for AUD (NCT04011644) [12]. Implementation of the intervention was guided by the EPIS (Exploration, Preparation, Implementation, Sustainment) framework, which delineates 4 phases in the implementation of evidence-based interventions: *exploration* to consider target population needs, *preparation* to assess barriers and facilitators to implementation, *implementation* to assess progress and determine adjustments, and *sustainment* to assess how to maintain the context structures and supports of the intervention [12]. For successful implementation, there needs to be a high degree of "fit" between the intervention and the contexts of implementation [13,14]. Fit, which has also been described as "appropriateness" and "compatibility" [14-16], has been conceptualized as encompassing multiple domains, including alignment between the intervention and the needs and preferences of the target population and alignment between the intervention and the culture, resources, and capabilities of the implementing organization [14]. The lack of fit is one of the most often cited barriers to implementing an intervention, but the construct has not often been the focus of implementation research [14].

In this paper, we examined the experiences of the users and implementers of an mHealth intervention called Tula (Sanskrit for "balance") for AUD that was tested in a recent 12-month randomized controlled trial [17,18]. We deploy the notion of fit as a lens to explore the alignment between the Tula intervention, the various digital support models, the target population, and the implementing organizations to better understand the trial results and to provide actionable guidance for future implementation of digital health interventions in health care organizations.

Methods

Intervention Design

The Tula intervention used a harm reduction approach to support individuals with mild to moderate AUD to decrease the number of heavy drinking days (defined as days with 5 or more standard drink units for males and 4 or more for females). Trial volunteers were randomized to 1 of 3 arms that included different levels of human touch: app only, app plus peer support, or app plus health coaching (see Table 1 for details of these arms). Participants in each arm received a monetary incentive to complete periodic study surveys that assessed the number of heavy drinking days and several quality of life measures. A full description of the protocol can be found in a study by Park et al [17].

Full trial results are reported in a study by Quanbeck et al [18]. Across all 3 groups, a statistically significant time effect was observed, with the percentage of self-reported heavy drinking days dropping from 38.4% (95% CI 35.8%-41%) at baseline to 22.5% (95% CI 19.5%-25.5%) at 12 months. There were no significant differences between groups in alcohol use reduction. Participants in the health coaching arm showed statistically significant improvements in mental health–related quality of life compared to the other 2 groups. However, participants in the health coaching arm dropped out of the study at a higher rate than participants in the other 2 arms, making that result difficult to verify due to risk of bias due to survivorship effects (ie, patients who had worse mental health may have been more likely to withdraw from the study).

Table . Study arms.

	App only	App+peer support	App+health coaching
Use of Tula	Unguided use of Tula	Tula use supported by a community-based peer mentor	Tula use supported by the health coach
Intervention	Study team conducts safety monitoring and technical support	Interpersonal communication and wellness monitoring via the app	Up to three 1:1 health coaching sessions via phone call
Discussion forum	No discussion forum	A discussion forum moderated by a peer mentor	A discussion forum moderated by a health coach
Communication	No communication feature with private messaging	The private messaging feature in communication routing to a peer mentor	The private messaging feature in communication routing to a health coach
Dashboard	No dashboard access	No dashboard access	Health monitoring by health coaches supported by a dashboard

XSL•FO
RenderX

Participant and Implementor Description

Eligibility criteria are listed in Textbox 1. All participants created anonymous usernames to use the app.

We partnered with a community-based organization to have certified peer specialists provide social support for participants randomly assigned to the peer support group. We provided project-specific training regarding the app and facilitating discussions. Peer specialists were not assigned to any specific participant but were available to any participant who reached out to them via a private messaging feature of the app. Peer specialists were able to draw upon their lived experiences with recovery from substance use disorders, including mental health issues, to moderate a discussion forum while encouraging participants to use the app.

In total, 2 certified health coaches—part of the health care system—worked one-on-one with participants randomly assigned to their group. The health coaches' goal was to help participants achieve their self-identified goals toward a healthier lifestyle as they reduced their alcohol consumption. Participants were encouraged to use up to three 1:1 personalized health coaching sessions via phone during the first 90 days of their trial session.

Textbox 1. Eligibility criteria.

Inclusion criteria

- 21 years old or older
- Wants to reduce drinking
- Owns a smartphone and is willing to download and use the Tula app
- Lives within the health care system service area
- Meets at least one of the following criteria:
  - · AUDIT (The Alcohol Use Disorders Identification Test) [19] screening score 8+, OR
  - · Responds "yes" to at least 2 questions on the Alcohol Use Disorder: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [AUD: DSM-5] [20], OR
  - · Reports moderate- to high-risk drinking patterns
  - · 4+ drinks on any single day and >7 drinks per wk (women and men aged >65 y)
  - · 5+ drinks on any single day and >14 drinks per wk (men)

Exclusion criteria

- Reports symptoms consistent with severe alcohol use disorder during screening (6 or more of 11 Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [DSM-5] criteria)
- Active psychotic disorder diagnosis
- Acute medical problem requiring immediate hospitalization
- Terminal illness

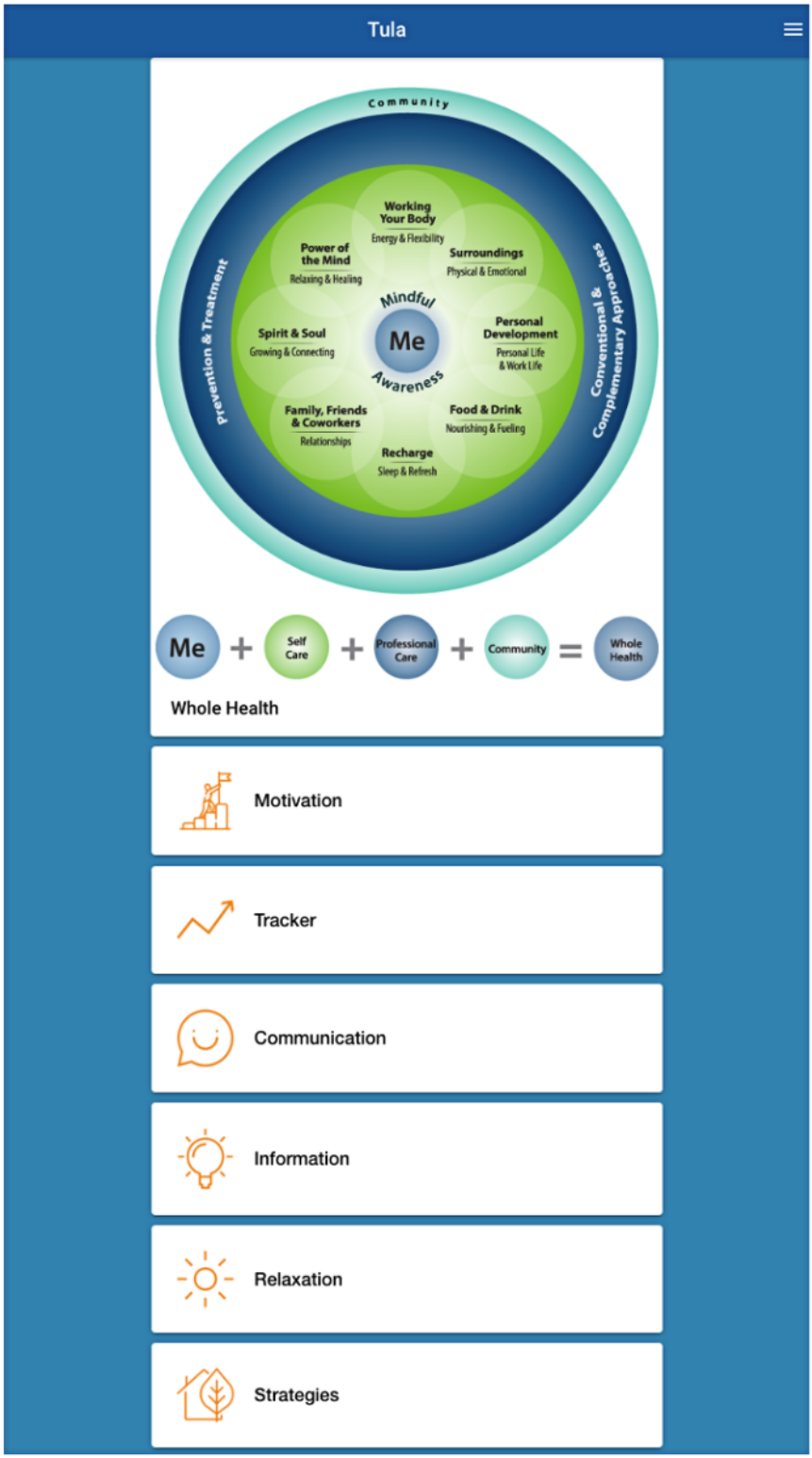
App Design

The app was designed according to the tenets of self-determination theory, which posits that adaptive functioning is optimized when people are able to meet their needs for autonomy, competence, and connection [17,21]. The app content was informed by the Whole Health Model, an approach to health promotion that views health as having physical, psychological, spiritual, and social dimensions and seeks to enable people to improve their health by empowering them to identify their own needs and priorities and set their own goals [22]. Figure 1 shows a visual depiction of the app.

The original trial protocol for this Tula intervention project was considered minimal risk for participants and was exempt from requiring a formal data monitoring committee review. All data collected via the app were stored in secure, password-protected servers. Participants were informed that no patient health information would be collected from their medical records, nor would any study data be entered into their medical records. All data were deidentified before any analysis.

XSL•FO
RenderX

Figure 1. View of the app.



Qualitative Study Sampling

We sought to elicit perspectives from implementers of the Tula intervention and research participants. The implementer sample (n=7) included 1 research team member who had project management responsibilities, all 4 individuals who worked as peer mentors, and the 2 individuals who provided health coaching. We used the level of engagement with the Tula intervention as a criterion to selectively sample research participants. Engagement was determined by total number of app page views over 12 months for participants in all 3 arms, discussion group views and posts for the participants in the peer support arm, and number of health coaching sessions completed for those in the health coaching arm. Participants were divided into high and low engagement groups (with the most engaged listed first in their group and the lowest engaged listed first in theirs). Researchers then used these rankings to invite interview participation, with the most and least engaged invited first. All

XSL•FO
RenderX

invitations were sent via email addresses provided by the research participants. Research staff attempted contact with each participant 3 times before moving on to the next participant on the list. Participants who did not engage at all—for example, those in the health coaching arm who attended zero coaching sessions—were not eligible because they would have no direct experience with the intervention.

Qualitative Interviews

Implementers were interviewed by the first author, a public health scientist and qualitative methodologist. These interviews focused on roles played in Tula, implementers' theories of change, perceptions of how the intervention did or did not work, and suggestions for improvement of the intervention. Trial participant interviews were conducted by the first author and 2 other research team members, AP, a research program manager, and SV, an undergraduate research assistant, who participated in interview training conducted by the first author. The 3 interviewers, all women, were not involved in developing the Tula intervention and had not had previous contact with the participants. Participant interviews covered participants' reasons for joining the study and their opinions of and experiences with the different components of the intervention, as well as their perceptions of whether and how Tula had affected their drinking behavior. Interviews were conducted using semistructured interview guides developed by the research team and minimally revised following early interviews. All interviews were held virtually and included only the interviewer and the research participant. Interviewers did not have knowledge of the individual participants' outcomes—that is, whether or not they had reduced their drinking.

Qualitative Analysis

All interviews were audio recorded, transcribed verbatim, and analyzed using rapid qualitative analysis, an approach that includes both deductive and inductive elements and is particularly appropriate for implementation research [23]. An initial coding matrix used domains derived from the EPIS framework: innovation, inner context, and outer context. After sorting the data into these a priori categories, a round of coding developed new categories that were specific to the data set. For example, the inner context included participants' motivations for volunteering for the study, whereas the outer context included the impact of the COVID-19 pandemic (which coincided with the trial). The notion of fit as described by Blanchard and Livet [14] was then operationalized by searching for areas of compatibility and incompatibility between the Tula intervention and the inner and outer contexts of its implementation. Other analytic activities included memo writing and discussions among the research team about the developing findings. The first author led the analysis, including an iterative process of review and feedback from the other authors.

Ethical Considerations

The full trial was approved by the Minimal Risk Institutional Review Board (IRB) at the University of Wisconsin-Madison (2019‐0337 CP012 for IRB Study 2019‐0337, NCT06305624). In addition, the trial received a Certificate of Confidentiality from the National Institutes of Health. The qualitative study was approved by the same IRB as an addendum to the trial protocol (2019‐0337 CP012 for IRB Study 2019‐0337). The IRB allowed a waiver of written consent for the trial, but all participants were provided with a study information and consent document and were given the opportunity to ask any questions about the study at the point of recruitment. Similarly, interview participants were provided with written information about the purpose of the interviews and their rights as research participants. Verbal consent was affirmed at the beginning of each interview. Interviews were anonymized during transcription, which was completed by a University of Wisconsin–based service certified in Health Insurance Portability and Accountability Act (HIPAA) regulations and human subjects protections. All research participants interviewed received US $50 gift cards in appreciation for their time. Implementer interview participants were not compensated apart from the financial support they received for their roles in the intervention.

Results

Overview

The findings reported here were based on semistructured interviews with 7 Tula implementers and 18 research participants. All implementers approached agreed to take part in interviews. The participant response rate varied considerably, from a low of 15% in the low engagement app only arm to a high of 75% in the high engagement app only arm. The average response rate across arms and engagement levels was 41.25%. Nonresponders included participants whose email addresses proved to be nonfunctioning, as well as those who never replied to 3 invitations. Implementer interviews ranged from 30 to 70 minutes in length. Participant interviews ranged from 19 to 55 minutes (mean 27.5 min). The demographics and usage data for the 18 research participants were largely reflective of the entire trial sample (n=558). Interview participants were primarily White (17/18, 94.4%), female (13/18, 72.2%), and highly educated, holding a bachelor's or master's degree (12/18, 66.7%). The average age was 47 (SD 12.8) years (see Table 2, which also shows use data for interview participants and the overall trial sample).

XSL•FO
RenderX

Table . Characteristics of the study population[a].

Characteristics	Low use (n=9)	High use (n=9)	Total interviewed (n=18)	Total (n=558)
Age, mean (SD)	43.8 (13.9)	50.2 (11.6)	47 (12.8)	42.8 (12.9)
Female, n (%)	6 (66.7)	7 (77.8)	13 (72.2)	366 (65.6)
Race, n (%)				
White	9 (100)	8 (88.9)	17 (94.4)	510 (91.4)
Black	0 (0)	1 (11.1)	1 (5.6)	27 (4.8)
AI/NA[b]	0 (0)	0 (0)	0 (0)	5 (0.9)
Asian	0 (0)	0 (0)	0 (0)	12 (2.2)
NH/PI[c]	0 (0)	0 (0)	0 (0)	2 (0.4)
Other	0 (0)	0 (0)	0 (0)	6 (1.1)
Hispanic or Latino, n (%)	0 (0)	0 (0)	0 (0)	14 (2.5)
Education, n (%)				
<High school	0 (0)	0 (0)	0 (0)	3 (0.5)
HS[d] or GED[e]	0 (0)	1 (5.6)	1 (5.6)	67 (12)
Vocation or associate	2 (22.2)	3 (33.3)	5 (27.8)	85 (15.2)
Bachelors	5 (55.6)	5 (55.6)	10 (55.6)	235 (42.1)
Masters	2 (22.2)	0 (0)	2 (11.1)	121 (21.7)
Doctorate	0 (0)	0 (0)	0 (0)	47 (8.4)
Marital status, n (%)				
Married	6 (66.7)	4 (44.4)	10 (55.6)	301 (53.9)
Widowed	1 (11.1)	1 (11.1)	1 (5.6)	10 (1.8)
Divorced	0 (0)	1 (11.1)	1 (5.6)	55 (9.9)
Separated	0 (0)	0 (0)	0 (0)	6 (1.1)
Never married	0 (0)	2 (22.2)	3 (16.7)	109 (19.5)
Living with a partner	2 (22.2)	1 (11.1)	3 (16.7)	73 (13.1)
Refused	0 (0)	0 (0)	0 (0)	1 (0.2)
Don’t know	0 (0)	0 (0)	0 (0)	3 (0.5)
Severity: mild, n (%)	4 (44.4)	3 (33.3)	7 (38.9)	356 (63.8)
Tula usage, mean (SD)				
Page views	385.8 (119.9)	2587 (1104)	1487 (1365)	806.1 (510.5)
DG[f] Read	21.5 (26.1)	320.2 (464.2)	170.8 (349.5)	76.7 (118.1)
DG Post	8 (0)	64 (42.5)	50 (44.6)	8.3 (12.1)
Health coach visits	1 (0)	3 (0)	2 (1.1)	1.7 (1.3)

[a]Mild severity was determined by a screening survey based on National Institute on Alcohol Abuse and Alcoholism thresholds and the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition, when 2 to 3 symptoms/criteria are present for AUD. A total of 4 to 5 symptoms constitute moderate AUD. Six or more symptoms constitute severe AUD—patients scoring 6 or higher were not eligible for the trial.

[b]AI/NA: American Indian/Native American.

[c]NH/PI: Native Hawaiian/Pacific Islander.

[d]HS: high school.

[e]GED: General Educational Development.

[f]DG: discussion group.

XSL•FO
RenderX

Motivations

Research participants identified 2 alcohol-specific motivations for wanting to join the study: being curious about their drinking habits and wanting to cut down on their drinking. In discussing these motivations, they alluded to feelings of being at risk for problem drinking either because of their individual circumstances—for example, a family history of alcoholism or an increase in drinking during the early days of the COVID pandemic—or because they perceived their social context as being oriented toward heavy drinking.

> *Why? I guess I was just kind of, I was curious about my own habits. Alcoholism runs in my family...and I just was kind of, it's always kind of lingered in the back of my mind, like how responsible I am with alcohol. So I was just kind of curious to see how often I really did drink...and cravings and, you know, did I need it?* [Participant 04 in the app-only arm]

> *[I already have anxiety issues, but during COVID my anxiety really went up a lot, and I found that I was drinking quite a bit, like every day actually. Much more than I had before. So when the study came about...I just thought it would be really good to try to get back on track and to decrease the, how much I was drinking, because it really affects my anxiety.]* [Participant 06 in the health coaching arm]

Participants also described being motivated to participate in the Tula intervention by a wish to improve their overall health and by an interest in research.

> *[Well, for myself, it's, you know, like some people have every year where, you know, they start the new year and they're like, okay, what do I want to try to do to be better? I'm always trying to identify things of, areas work where I can, you know, get better. When I had seen that, I knew that I was probably drinking a little bit more than what I should have, and it just looked at, you know, right time, right place, right opportunity.]* [Participant 04 in the health coaching arm]

> *[Well I just wanted to cut down on my drinking, and the reason I wanted to do that was because my cholesterol levels were kind of high. Well, I attributed that to late night snacking, and you know, just overeating because that's what alcohol kind of triggers in me. So, yeah. So that's why I wanted to participate, and just to be healthier, you know?]* [Participant 06 in the peer support arm]

> *[So I found it very interesting. I have a health background, and I've always been, you know, into health and wellness and fitness. So I just thought that it sounded like a very interesting study to be a part of. I'm always kind of interested in research studies in general, but this one in particular I thought was especially interesting just because it had, you know, it related more towards overall, you know, health and wellbeing.]* [Participant 01 in the health coaching arm]

Implementer Theories of Change

Each of the 3 components of the Tula intervention—the app, the peer support, and the health coaching—had its own theory of change.

As noted, the app's design was based on self-determination theory and the Whole Health model. The app's embedded theory of change held that people's ability to change their drinking behavior would be promoted by features that were customizable and educational and that supported relationships with others.

Peer mentors told us that behavior changes, such as drinking reduction, depend on an individual's intrinsic motivation. Motivation is prompted and reinforced by increased self-awareness of the pros and cons of the behavior and, ultimately, a decision that the cons of continuing the behavior outweigh the pros. The role of the peer mentor is to use their own experiences of substance use, recovery, and relapse to establish rapport and then to "support and love [people] where they are at" by offering nonjudgmental, highly personalized, and often intensive assistance.

The health coaches described grounding their practice in the Whole Health model and the Stages of Change model of health behavior change [24]. Similar to peers, they saw health behavior change as originating in an individual's intrinsic motivation and the coaches' role as "support[ing] people in making health change that they were ready and wanting to make." Health coaches "[get] to the heart of what really matters to people" in order to provide tools that facilitate an individual's "mindful awareness" of their health habits and the ways in which these habits are or are not consonant with what the individual values. Health coaching consists of fairly structured conversations in which coaches "help [people] go from knowing to doing," or move from contemplation to action in the stages of change, by working with the individual to identify what they value, figure out how their current habits do or do not serve these values, set goals for change, review their progress toward meeting these goals, and fine-tune their goals as needed.

Change Mechanisms

Participants described 4 main mechanisms through which the Tula intervention worked to support change: accountability, encouragement, information, and connection. These mechanisms were described by participants in each of the 3 arms, though, as we explore, some components of the Tula intervention were a better fit than others in supporting these mechanisms.

Accountability came in 2 forms. In the first form, the Tula intervention raised participants' awareness of their drinking behavior by providing visibility onto the number of drinks taken each week.

> *[[The app asks you] "How much did you drink last week?" And that, for me, was enough to just kind of keep me in check. You know, like, "Ooh, I had a lot last week"...if you have three or four drinks, then you want to have a fifth drink. You're like, "Ooh, but I gotta enter that in that app".]* [Participant 02 in the health coaching arm]

XSL•FO
RenderX

> *[Well, as you know, the application itself makes you more mindful of how much alcohol you're consuming because on a weekly basis, it asks you to tally up how much you've had, and when you actually physically do that, when at the time might seem like I just had a few beers, you know, just had a few drinks, you can see where it really starts to add up. [Laughs] And so I used it, I guess on a weekly basis and then because I was doing that on a weekly basis and seeing how much I was drinking a night... it made me more mindful day-to-day as to how much I was consuming. And so it was, it became more like, "Oh, you know, you already had a couple, you know, maybe that third one is not a good idea."]* [Participant 04 in the peer support arm]

In its second form, accountability consisted of cycles of pattern analysis and goal setting and either positive or negative reinforcement for behavior change.

> *[I also liked that it had the like, mood and sleep like trackers along with the alcohol consumption trackers. Just because it was interesting to kind of see the correlation between all three of those factors and not just looking only at alcohol consumption. It's kind of interesting to see. I noticed that like alcohol consumption tend to be a little higher, like in weeks where I rated mood or lower sleep.]* [Participant 02 in the app-only arm]

> *[[The combination of the tracker and the surveys] made me reflect on my goals... If the question was, how connected do you feel to people outside of work, or, you know, to your family. If I noticed that I connected low there, and then had like three or four drinks above my goal for that week...then that was like a pretty clear, oh, you know, things are out of balance here.]* [Participant 02 in the peer support arm]

Both high and low engagement participants described accountability as having had the most impact on their drinking behavior.

Encouragement worked by promoting positive thinking, changes in perspective, mood elevation, and hope, which in turn helped to provide motivation to change behavior.

> *[Oh, just you know, [the thought of the day] was always something positive and it just launches your day in a great manner and, you know, inspiring quotes or, you know, you've got this, you can do it. You know, tough times don't last, tough people do kind of things and just very inspiring, you know, and it wasn't preachy or anything like that. It was always just kind of nice little encouraging snippets.]* [Participant 04 in the peer support arm]

> *[I dug into the gratitude list. So it was more of a reminder. If I was having a bad day that I could really point out, remember, positives of my day to go back and reflect upon.]* [Participant 05 in the health coaching arm]

> *[[The peer mentor] kind of opened up and told us about her experiences as well. I thought she was very down to earth, and very encouraging.]* [Participant 06 in the peer support arm]

> *[There was challenges and stuff like that and it wasn't always easy, that there was always being progress made. So that, it was breaking that stuff down when it's like, oh, well, you know, and [the health coach]'d be like, no, well, you know, look what you've already done and, you know, and then come with some other tools or things where I was having difficulties that I wasn't able to utilize, showing me the way to, you know. Just different ideas and things like that. So it was like, I just felt like I was going in the right direction, and that would be probably the biggest, it was very motivating, I guess.]* [Participant 04 in the health coaching arm]

Information supported change by offering new strategies for adaptive behaviors, as well as substitutions for maladaptive behaviors.

> *[[The app contained] just different ideas of things to do to occupy yourself other ways and meditation ideas, like this all kinds of different ideas. Things to keep you busy.]* [Participant 05 in the peer support arm]

Connection allowed users to exchange practical knowledge and to gain insight into others' experiences.

> *[It was insightful to get some tips from other people and it kind of led me to something new I found out, too, which has really helped me. I, well I started making mocktails, and I'd have them at home. I rarely drink at home, but every once in a while, I get really stressed out, beat up, and just, ugh, you know? I want a cocktail, but I found that mocktails were really good for me because they kind of gave me that ritual. They really helped. So, somebody suggested trying them and, you know, try some different recipes, and yeah. So I got some good information from other members that way.]* [Participant 06 in the peer support arm]

> *[So it was useful in the fact that...challenges that I would run into, I could see others that were challenged in the same manner, and some, on some occasions you could get feedback of what they were able to use or what they did that that made that difficulty something that was easier to deal with. Then there was also ones that would have comments where you like, okay, I don't want to do that. So on the positive there was there was good things, and on the negative there was good things. It was just nice to have that different perspective.]* [Participant 04 in the health coaching arm]

Fit and Misfit

We found areas of fit and misfit across the 3 components of the Tula intervention and the contexts of implementation.

Tula was a good fit with participants' motivations for joining the trial. All interview participants described the intervention

XSL•FO
RenderX

as having been successful in giving them some insight into their drinking. Many, though not all, reported that they had reduced their drinking or otherwise changed their drinking behavior. Several noticed improvements in their overall health—for example, weight loss or improved sleep.

There was compatibility between implementers' theories of change and participants' reports of how Tula worked to effect changes in their drinking behavior. The accountability, encouragement, information, and connection mechanisms were consistent with the theories of change embedded in the 3 components of Tula. That is, when it worked, Tula appeared to have worked in ways it was theorized it would.

There was also a good fit between participants' motivations, the change processes they described, and the app's features. The feature most often associated with both forms of accountability was the weekly drink tracker, which was universally popular with low and high engagement participants. All interview participants described using the tracker to become more aware of their drinking behavior, and several attributed their ability to reduce or change their drinking behavior to the tracker. Participants who experienced the second form of accountability also cited the study surveys. The app features described as being the most encouraging were the thought of the day, the modules devoted to motivation and gratitude, and, for those in the human touch arms, interactions with the peer mentors and health coaches. The Tula components that seemed most closely linked to information were the app modules focused on relaxation and drinking reduction strategies and the discussion groups, particularly the group associated with the peer support component. The Tula components mentioned as important to fostering connection were the peer support and the health coaching.

All the peers and health coaches indicated that the harm reduction philosophy underlying Tula was compatible with their own beliefs and approaches. The health coaches described their Tula work as very similar in approach to their non-Tula coaching and as personally fulfilling.

A significant misfit was the dissonance between the peer mentors' conceptualization of the peer role and the role as conceived in the Tula intervention. Specifically, the peer mentors contrasted the highly personal nature of their services as offered in the non-Tula context with the distanced and anonymous quality of the Tula work, which one peer mentor described as "dry." All peers noted this difference and suggested that it had hindered their effectiveness.

> *[I didn't have enough in-depth dialogue with each participant where I could understand where they're coming from...in a more in-depth way. And...I couldn't really share my experiences as deeply as I would if it was face-to-face.]* [Peer mentor 01]

> *[There was no face-to-face contact...there was not much of a connection made there. Whereas, you know, with [traditional] peer support, we meet in person, face-to-face. You know, we can, you know, we can see the, the body language...that's how you build a rapport, connection and trust. And so through an app there just wasn't anything like that... It's just not enough. It was, it was just really almost non-human interaction in a way.]* [Peer mentor 03]

One peer mentor experienced this misfit so acutely that they provided their contact information to trial participants and held one-on-one telephone conversations.

The discussion feature, included in the app for those trial participants randomized to the peer support or the health coaching arms, was successful for some users, especially to the extent that it was a source of information and encouragement. However, it appeared to be largely a misfit when it came to providing deeper feelings of connection. Both the health coaches and the peer mentors said they felt poorly equipped to facilitate meaningful discussions. Even the highest engagement participants in the peer support component, those who spent the most time on the discussion pages, were lukewarm about the discussions. Several participants said they had other outlets, such as AA, that met their needs for connection. Others noted that they felt they had little in common with participants who posted in the discussions.

> *[I only really read. I never really posted, to be honest... It's easier just to kind of read other peoples'.]* [Participant 03 in the peer support arm]

> *[To be honest, I did not play on the social side of it at all...it was people who weren't going to quit, and they didn't really want to. They were looking for someone to talk them into it. That's kind of the way I looked at those comments...in my opinion, if you want to do something, you just got to do it. It's not like somebody should be talking to you or anything like that. It's got to come from inside.]* [Participant 01 in the peer support arm]

Implementers identified a misfit attributable to the research context. They perceived that some trial participants were motivated not by an authentic wish to change their behavior but by the monetary compensation offered by the study. Thus, they lacked "skin in the game," and, while these participants completed the surveys that were the basis for receiving compensation, they did not really engage with any of the Tula components.

The app's technology and the volume of its content showed fit and misfit. While most implementers and participants praised how well the app functioned—"everything you click on, it works," others described it as clunky and unintuitive or reported technical glitches that interfered with use. In total, 2 participants expressed a lack of enthusiasm for any app: one because their life during COVID was taking place almost entirely online and they wanted a break from screens and another because they were trying to reduce their phone usage.

Many participants liked how much information the app contained; others felt overwhelmed or wanted a more curated experience.

> *[And so for me it was frustrating because there was so much in abundance. And I felt like you had to do it all. It wasn't like I could do a subset and be successful. So then I found myself not being as*

XSL•FO
RenderX

> *successful. That's a terrible thing to say, but I did...it became overwhelming and I felt more guilty and worse off.]* [Participant 03 in the health coaching arm]

Other domains that mixed fit and misfit were preferences related to intervention intensity and privacy.

While most participants seemed to appreciate that the Tula experience was fairly relaxed, others felt that they would have gotten greater benefit from a more intensive intervention.

> *[I probably would have benefited from it all. You know, having the group, having the counselor and the app, all three of those, would be a pretty high success rate I think, for myself personally.]* [Participant 03 in the app-only arm]

> *[I felt [the health coaching sessions] weren't memorable. I felt they were short, weren't in-depth. It was like, "what do you want your goals to be? Here are your problems. OK, that's a good goal." I just fell off the map. It really wasn't, I thought it would be more invasive at that level with the coach, and it was 15 minutes maybe each... I felt the coach didn't coach... You want someone who is skilled but is also going to provide you with expertise and solutions and ideas. I didn't feel that was done.]* [Participant 05 in the health coaching arm]

Many participants stated that they did not care to reveal themselves or make themselves "vulnerable" to strangers online, and thus, they appreciated having the ability to remain anonymous.

> *[Wanting to kind of self-improve on your own terms was really helpful for me in just having it be private, but knowing that, you know, there were strategies in there I could look at and I didn't have to, like reach out to somebody. For that, personally, for me, it was helpful.]* [Participant 01 in the app-only arm]

> *[And that way when you use the app, it's private, you know? No matter what you enter into it, it's just you and your entering info, you know? No one else around and it can, as bad as it may seem, who cares? Because no one sees it but you, you know what I mean? It goes to data for whatever you guys are doing, but at the time when you're entering stuff it don't matter, because it's just you.]* [Participant 05 in the app-only arm]

The fact that health coaching required some self-revelation during one-on-one conversations with coaches was uncomfortable for some participants in that arm.

> *[The other most difficult thing is talking to the coach on the phone whom I didn't have a relationship with, didn't know. And for me, it was already an embarrassing topic…I felt very probed and very kind of like a scientific experiment…I know it was anxiety producing for me.]* [Participant 03 in the health coaching arm]

Notably, the research team member who informed participants of which arm they had been assigned to reported that many people randomized to the coaching arm expressed disappointment with their assignments. As noted earlier, the health coaching arm had a higher rate of attrition than the other 2 arms.

Discussion

Principal Findings

While most of the literature on alcohol reduction mHealth interventions has been inconclusive, there is some evidence of promise across a range of studies [25-28]. Our examination of fit suggests some explanations for that promise.

We found a good fit between the Tula intervention, inclusive of the app, peer support, and health coaching, and many aspects of the contexts of implementation. There was a compatibility between the implementers' theories of change and participants' descriptions of how change occurred and between these change processes and several of the app's features. The best fit was between the change process participants called accountability and the app's tracking feature. This is consistent with findings from other studies, which have demonstrated positive behavior change from logging alcohol consumption as participants see real-time progress via the tracking function [26]. Giroux et al [29] conducted a study of a location-based monitoring and intervention system for AUDs, and Osth et al [30] compared 2 different smartphone apps to complement alcohol treatment. Both reported that participants found the accountability function and tracking feature increased their awareness of drinking consumption and increased their motivation for behavior change.

We also uncovered several areas of misfit. There was discordance between the peer mentor role in the Tula intervention and key features of peer mentorship in the non-Tula environment. The app's discussion feature was only minimally effective at fostering connection.

In addition, we observed several domains of mixed fit and misfit, specifically pertaining to the app technology's ease of use and its content and to participants' preferences related to intervention intensity and privacy. Patient concerns over data privacy and security are important when it comes to issues of choice in digital health. Patients expect clear guidance on how their data will be used and value the ability to control data sharing [31]. That the Tula intervention offered an array of options—"something for everyone"—was cited by both participants and implementers as an important strength. Having choices meant that participants could customize their experiences, including the choice to ignore parts of the intervention that they did not find helpful or interesting. A narrative review of 71 articles on barriers to and facilitators of digital health adoption found that patient empowerment, self-management, and personalization were the key drivers in the adoption of digital health tools [32].

Another key finding was the ways in which the research context affected both participants' and implementers' experiences. For some participants, random assignment constrained choice, pushing them into interventions they had no interest in exploring. From the implementers' perspective, the incentives offered to trial participants appealed to motivations that had nothing to do with drinking or health. In both cases, the misfit

XSL•FO
RenderX

between the intervention and the research context appeared to have threatened Tula's effectiveness. Conversely, the survey that participants were asked to fill out as part of the research seemed to have functioned as an effective part of the intervention for some participants by reinforcing accountability.

This qualitative examination helps provide an explanation for the main results of the Tula trial. The dominant mechanism of change described by the interview participants was accountability, which was closely tied to the app's tracker, a feature that trial participants in all three arms were able to access. The fact that the peer support component was not more effective than the app alone may be because what happened in the Tula intervention bore little resemblance to the highly personalized practices of outreach and response that are part of the ethos of peer support as it has developed as a profession [33,34]. The higher rate of attrition in the health coaching group compared to the other study arms may be at least in part explained by a misfit between the privacy preferences of a subset of participants and the demands of the health coaching intervention. However, it should be kept in mind that related research on other digital health apps for alcohol reduction has also shown very high attrition rates; Attwood et al's examination of the Drinkaware app for alcohol use reduction observed high attrition, with only 42.6% of users using the app after the first week and only 5% retained after 12 weeks [35]. The improvement in mental health–related quality of life among participants who stuck with health coaching seems to reinforce the idea that allowing people to customize their own intervention experiences may garner the best outcomes.

Limitations

We refer readers to a report of the primary quantitative analysis for a discussion of limitations of the parent trial [18]. The design of this qualitative study did not allow us to determine the exact nature of the relationship between fit and engagement for individuals in the primary quantitative analysis, but the data do suggest that the two are related. Participants for whom the Tula components to which they were assigned were a good fit were enthusiastic about them and appear to have been more engaged. Lower engagement users seem to have experienced more misfit.

Although the demographics of the interview participants were largely consistent with the overall Tula sample, individuals who agreed to participate in interviews might have differed from the larger sample in ways we did not measure. Our decision not to recruit participants who did not use the Tula components to which they were assigned means that the sample does not represent the group of participants with the lowest levels of engagement. However, the consistency of responses suggests that we did achieve informational redundancy and thus are able to represent the experience of participants who met our inclusion criteria. Because the overall sample for the trial was predominantly female, White, and highly educated, the findings reported here may not be generalizable to a more diverse population. Improvements in mental health–related quality of life may have been related to the personalized social support shown in the coaching group, although we do not have specific quantitative data to substantiate this claim. Finally, it could be that participants were interested in reducing the number of heavy drinking days, hence their participation in the study, and that the reduction seen was a result of their own personal motivations—not the app, peer support, or health coaching.

Future Directions

Further analysis of the specific intervention components that participants engaged with may help improve the intervention design in future research. The study found that different users have different preferences for intervention intensity. Dynamic adjustment of intervention intensity could feasibly be achieved in future intervention designs to meet individual needs by using sequential, multiple assignment randomized trial designs or microrandomized trial designs [36,37]. Future research should also explore the impact of comorbid factors common in patients with AUD (eg, depression, anxiety) on the fitness and effectiveness of mHealth interventions.

Conclusions

The results reported in this paper suggest that the concept of fit can go some way toward explaining the outcomes of the Tula trial and may provide insights into participant experiences of mHealth interventions in general. Using fit as a lens through which to examine these experiences provides insights that can be applied to iterative intervention improvement, which in turn will increase effectiveness and promote sustainment.

Acknowledgments

The authors would like to thank our community partners: Safe Community of Madison and Dane County and the UW Health Wellness Center. In particular, the authors want to recognize Tanya Kraege, Charlestine "Charlie" Daniel, Cheryl Wittke, and the peer support specialists who were integral to moderating and facilitating discussion forums with participants in this study. The authors also want to recognize Diane Scherschel, the Integrative Health and Wellness Manager at the UW Health Wellness Center, as well as Katie Eby and Ellen Houston, who provided health coaching to our participants on their journey toward a healthier lifestyle. This research was funded by the National Institute on Alcohol Abuse and Alcoholism (grant number R01-AA024150). The funder had no role in any aspect of the development, conduct, analysis, or reporting of the study.

Data Availability

The dataset containing deidentified data used to generate the results in this paper is available on request sent to corresponding author Dr. Andrew Quanbeck.

XSL•FO
RenderX

Authors' Contributions

NJ conceived the paper, led data collection and analysis, and drafted the paper. LSP contributed to the analysis and drafted the paper. AP and SV participated in data collection and analysis and contributed to the paper. MH drafted the paper and assisted with editing for submission. AQ led the Tula RCT and provided critical revisions to the manuscript. All authors read, reviewed, and approved of the final submission.

Conflicts of Interest

AQ has a shareholder interest in CHESS Health, a company that disseminates software technology to the addiction treatment field (CHESS Health did not develop the app tested in the manuscript) and consults on implementation of evidence-based health care interventions for the non-profit NIATx Foundation. These relationships are managed by the University of Wisconsin–Madison's Conflict of Interest Committee. All other authors have no conflicts of interest to disclose.

References

1. Data on excessive drinking. Centers for Disease Control and Prevention. 2022 Nov 28. URL: https://tinyurl.com/ybcek243 [accessed 2023-05-23]
2. Behavioral risk factor surveillance system survey data. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/brfss/index.html [accessed 2024-06-06]
3. Miller AP, Kuo SIC, Johnson EC, et al. Diagnostic criteria for identifying individuals at high risk of progression from mild or moderate to severe alcohol use disorder. JAMA Netw Open 2023 Oct 2;6(10):e2337192. [doi: 10.1001/jamanetworkopen.2023.37192] [Medline: 37815828]
4. Jonas DE, Garbutt JC, Brown JM, et al. Screening, behavioral counseling, and referral in primary care to reduce alcohol misuse. Agency for Healthcare Research and Quality (US). 2012. URL: http://www.ncbi.nlm.nih.gov/books/NBK99199 [accessed 2025-05-23]
5. Boehmer KR, Álvarez-Villalobos NA, Barakat S, et al. The impact of health and wellness coaching on patient-important outcomes in chronic illness care: a systematic review and meta-analysis. Patient Educ Couns 2023 Dec;117:107975. [doi: 10.1016/j.pec.2023.107975] [Medline: 37738790]
6. Mulvale G, Wilson F, Jones S, et al. Integrating mental health peer support in clinical settings: lessons from Canada and Norway. Healthc Manage Forum 2019 Mar;32(2):68-72. [doi: 10.1177/0840470418812495] [Medline: 30744437]
7. Bertha M, Shedden K, Mellinger J. Trends in the inpatient burden of alcohol-related liver disease among women hospitalized in the United States. Liver Int 2022 Jul;42(7):1557-1561. [doi: 10.1111/liv.15277] [Medline: 35451173]
8. Quanbeck A, Chih MY, Isham A, Gustafson D. Mobile delivery of treatment for alcohol use disorders: a review of the literature. Alcohol Res 2014;36(1):111-122. [doi: 10.35946/arcr.v36.1.11] [Medline: 26259005]
9. Domhardt M, Geßlein H, von Rezori RE, Baumeister H. Internet- and mobile-based interventions for anxiety disorders: a meta-analytic review of intervention components. Depress Anxiety 2019 Mar;36(3):213-224. [doi: 10.1002/da.22860] [Medline: 30450811]
10. Lepore SJ, Collins BN, Killam HW, Barry B. Supportive accountability and mobile app use in a tobacco control intervention targeting low-income minority mothers who smoke: observational study. JMIR Mhealth Uhealth 2021 Jul 2;9(7):e28175. [doi: 10.2196/28175] [Medline: 34255698]
11. Versluis A, Verkuil B, Spinhoven P, van der Ploeg MM, Brosschot JF. Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jun 27;18(6):e152. [doi: 10.2196/jmir.5642] [Medline: 27349305]
12. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health 2011 Jan;38(1):4-23. [doi: 10.1007/s10488-010-0327-7] [Medline: 21197565]
13. Moullin JC, Dickson KS, Stadnick NA, Rabin B, Aarons GA. Systematic review of the exploration, preparation, implementation, sustainment (EPIS) framework. Implement Sci 2019 Jan 5;14(1):1. [doi: 10.1186/s13012-018-0842-6] [Medline: 30611302]
14. Blanchard CM, Livet M. Ensuring intervention success: assessing fit as an overlooked step of the implementation process. Pharm Pract (Granada) 2020;18(4):2235. [doi: 10.18549/PharmPract.2020.4.2235] [Medline: 33343775]
15. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
16. Rogers EM. Diffusion of Innovations, 4th edition: New York, NY: Free Press; 1995.
17. Park LS, Chih MY, Stephenson C, et al. Testing an mHealth system for individuals with mild to moderate alcohol use disorders: protocol for a type 1 hybrid effectiveness-implementation trial. JMIR Res Protoc 2022 Feb 18;11(2):e31109. [doi: 10.2196/31109] [Medline: 35179502]
18. Quanbeck A, Chih MY, Park L, et al. Testing support models for implementing an evidence-based digital intervention for alcohol use disorder: results of a pragmatic hybrid implementation-effectiveness trial. Res Sq 2024 Mar 28:rs.3.rs-4004555. [doi: 10.21203/rs.3.rs-4004555/v1] [Medline: 38585768]

XSL•FO
RenderX

19. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT: the alcohol use disorders identification test: guidelines for use in primary health care. World Health Organization. 2001. URL: https://iris.who.int/handle/10665/67205 [accessed 2025-10-03]
20. Alcohol use disorder: a comparison between DSM-IVDSM-5. National Institute on Alcohol Abuse and Alcoholism. 2021. URL: https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/alcohol-use-disorder-comparison-between-dsm [accessed 2023-12-19]
21. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68] [Medline: 11392867]
22. Kligler B. Whole health in the veterans health administration. Glob Adv Health Med 2022;11:2164957X221077214. [doi: 10.1177/2164957X221077214] [Medline: 35273831]
23. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516. [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
24. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
25. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384. [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
26. Mubin O, Cai B, Al Mahmud A, Kharub I, Lwin M, Khan A. A preliminary evaluation of mobile phone apps to curb alcohol consumption. Int J Environ Res Public Health 2021 Dec 23;19(1):135. [doi: 10.3390/ijerph19010135] [Medline: 35010392]
27. Colbert S, Thornton L, Richmond R. Smartphone apps for managing alcohol consumption: a literature review. Addict Sci Clin Pract 2020 May 7;15(1):17. [doi: 10.1186/s13722-020-00190-x] [Medline: 32381062]
28. Bertholet N, Schmutz E, Studer J, et al. Effect of a smartphone intervention as a secondary prevention for use among university students with unhealthy alcohol use: randomised controlled trial. BMJ 2023 Aug 16;382:e073713. [doi: 10.1136/bmj-2022-073713] [Medline: 37586742]
29. Giroux D, Bacon S, King DK, Dulin P, Gonzalez V. Examining perceptions of a smartphone-based intervention system for alcohol use disorders. Telemed J E Health 2014 Oct;20(10):923-929. [doi: 10.1089/tmj.2013.0222] [Medline: 25243480]
30. Östh J, Danielsson AK, Lundin A, Wennberg P, Andréasson S, Jirwe M. Keeping track of my drinking - patient perceptions of using smartphone applications as a treatment complement for alcohol dependence. Subst Use Misuse 2024;59(2):291-299. [doi: 10.1080/10826084.2023.2269578] [Medline: 37876238]
31. Grande D, Mitra N, Shah A, Wan F, Asch DA. Public preferences about secondary uses of electronic health information. JAMA Intern Med 2013 Oct 28;173(19):1798-1806. [doi: 10.1001/jamainternmed.2013.9166] [Medline: 23958803]
32. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients' perspectives on digital health tools. PEC Innov 2023 Dec;2:100171. [doi: 10.1016/j.pecinn.2023.100171] [Medline: 37384154]
33. Fortuna KL, Solomon P, Rivera J. An update of peer support/peer provided services underlying processes, benefits, and critical ingredients. Psychiatr Q 2022 Jun;93(2):571-586. [doi: 10.1007/s11126-022-09971-w] [Medline: 35179660]
34. Chapman SA, Blash LK, Mayer K, Spetz J. Emerging roles for peer providers in mental health and substance use disorders. Am J Prev Med 2018 Jun;54(6 Suppl 3):S267-S274. [doi: 10.1016/j.amepre.2018.02.019] [Medline: 29779551]
35. Attwood S, Parke H, Larsen J, Morton KL. Using a mobile health application to reduce alcohol consumption: a mixed-methods evaluation of the drinkaware track & calculate units application. BMC Public Health 2017 May 17;17(1):394. [doi: 10.1186/s12889-017-4358-9] [Medline: 28511698]
36. Kidwell KM, Almirall D. Sequential, multiple assignment, randomized trial designs. JAMA 2023 Jan 24;329(4):336-337. [doi: 10.1001/jama.2022.24324] [Medline: 36692577]
37. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing digital integrated care via micro-randomized trials. Clin Pharmacol Ther 2018 Jul;104(1):53-58. [doi: 10.1002/cpt.1079] [Medline: 29604043]

Abbreviations

AUD: alcohol use disorder
IRB: institutional review board
mHealth: mobile health

XSL•FO
RenderX

Edited by L Buis; submitted 21.08.24; peer-reviewed by B Riggin, D Singh, WQ Chen; revised version received 23.05.25; accepted 25.08.25; published 14.10.25.

Please cite as:
Jacobson N, Park LS, Pulvermacher A, Voelker S, Herzog M, Quanbeck A
Exploring Fit in a Mobile Health Intervention for Alcohol Use Disorder: Qualitative Study
JMIR Mhealth Uhealth 2025;13:e65218
URL: https://mhealth.jmir.org/2025/1/e65218
doi:10.2196/65218

© Nora Jacobson, Linda S Park, Alice Pulvermacher, Samantha Voelker, Mallory Herzog, Andrew Quanbeck. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Barriers and Facilitators to Implementing a Smartphone App for Physicians to Improve the Management of Acute Myocardial Infarctions: Multicenter, Mixed Methods, Observational Study

Katelyn J Cullen[1,2*], MPH; Hassan Mir[3,4*], MPH, MHI, MD; Madhu K Natarajan[1,2,5], MSc, MD; Marija Corovic[1,2], MPH; Karen Mosleh[1,2], MSc; Jacob Crawshaw[6], PhD; Mathew Mercuri[1,7,8,9], PhD; Hassan Masoom[10], MASc, MD; JD Schwalm[1,2,5], MSc, MD

[1]Department of Medicine, McMaster University, Hamilton, ON, Canada

[10]Division of Cardiology, William Osler Health System, Brampton, ON, Canada

[2]Hamilton Health Sciences Knowledge Centre, Hamilton, ON, Canada

[3]Division of Cardiology, Ottawa Heart Institute Research Corporation, University of Ottawa Heart Institute, 40 Ruskin St, Ottawa, ON, Canada

[4]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[5]Population Health Research Institute, Hamilton, ON, Canada

[6]Centre for Implementation Research, Ottawa Hospital Research Institute, The Ottawa Hospital, Ottawa, ON, Canada

[7]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[8]Centre for Philosophy of Epidemiology, Medicine and Public Health, Department of Philosophy, University of Johannesburg, Auckland Park, South Africa

[9]Department of Philosophy, McMaster University, Hamilton, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Hassan Mir, MPH, MHI, MD
Division of Cardiology, Ottawa Heart Institute Research Corporation, University of Ottawa Heart Institute, 40 Ruskin St, Ottawa, ON, Canada

Abstract

Background: Timely and appropriate care is critical for patients with ST-elevation myocardial infarction (STEMI). Effective communication and prompt sharing of test results, particularly electrocardiograms (ECGs), between the referring emergency medicine (EM) physician or emergency medical service (EMS) paramedic and the interventional cardiologist (IC) are essential. This exchange relies on fax or SMS text messages. The SmartAMI-ACS (Strategic Management of Acute Reperfusion and Therapies in Acute Myocardial Infarction) App was developed to streamline communication. It is user friendly and privacy compliant, and enables rapid, secure ECG sharing to support faster, informed clinical decision-making.

Objective: This paper details the results of targeted preimplementation surveys to establish barriers and enablers of using a smartphone app to transmit ECG images among ICs, EM physicians, and EMS paramedics to help tailor implementation interventions.

Methods: To assess the proposed acceptability and uptake of the app, preimplementation surveys were disseminated to ICs, EM physicians, and EMS paramedics in one region of Ontario, Canada. Questions were generated based on selected components of the Consolidated Framework for Implementation Research, results from a pilot study carried out at a regional hospital where the SmartAMI-ACS app was previously implemented, and predicted barriers based on expert guidance. The preimplementation surveys consisted of 7-point Likert scale questions (1=strongly disagree and 7=strongly agree) and open-ended questions. Open-ended data were extracted verbatim and analyzed using an inductive qualitative approach, with transcripts coded into descriptive qualitative codes and then collapsed into themes.

Results: Survey uptake was acceptable, with 9 of the invited 10 ICs, 51 of the invited 223 EM physicians, and 93 of the invited 1138 EMS paramedics responding. All groups recognized that current practices for sharing ECGs allowed room for improvement, accepting that fax can be inconvenient and SMS text messages may not be secure. When asked whether there was a need for a smartphone app to transmit ECGs, ICs (mean 6.67, SD 0.5), EM physicians (mean 5.57, SD 1.3), and EMS paramedics (mean 5.79, SD 1.45) consistently agreed. Commonly reported barriers were concerns over technological challenges, privacy issues,

XSL•FO
RenderX

and cell phone reception strength. Through the identification of the barriers in each stakeholder group, implementation strategies were developed that facilitated the scale-up of this system-change intervention.

Conclusions: Results from the 3 web-based preimplementation surveys to identify key barriers and enablers to the implementation of the app helped inform the selection of tailored implementation strategies to support the rollout of the app across the health region. The surveys identified key barriers around technology, privacy, and access to required Wi-Fi that needed to be addressed during app implementation to facilitate uptake and use. Results from the surveys, and ongoing evaluation of effectiveness, are informing the expansion of the app intervention to local ambulance services and other health regions.

Trial Registration: ClinicalTrials.gov NCT05290389; https://clinicaltrials.gov/study/NCT05290389

International Registered Report Identifier (IRRID): RR2-10.2196/55506

(JMIR Mhealth Uhealth 2025;13:e60173) doi:10.2196/60173

KEYWORDS

ST-elevation myocardial infarction; digital health; mHealth; barriers and enablers; myocardial infarctions; myocardial; decision-making; decision; decision support; Ontario; Canada; survey; paramedics; privacy; app; application; cardiology; interprofessional communication; intervention; emergency medicine; ECG

Introduction

Each year, approximately 7000 people in Ontario experience ST-elevation myocardial infarctions (STEMIs) requiring timely reperfusion of the affected area of the heart to restore blood flow through the coronary arteries [1]. Reperfusion can occur by a procedure called percutaneous coronary intervention (PCI) or with thrombolytic medications. When possible, PCI is preferable, and guidelines recommend that a patient presenting to a center that can perform PCI (PCI-capable center) should have this procedure performed within 90 minutes of the patient presenting to the hospital. Those who present to a non–PCI-capable center should be transferred to a center that can perform this procedure within 120 minutes [2]. However, among the 270 health care institutions in Ontario, only 20 centers have PCI capabilities on site [3]. Thus, achieving timely care can be challenging, particularly for patients who live in rural areas where access to a PCI-capable center is limited. Where on-site PCI is not available, patients must be transferred to a partner PCI-capable hospital. Patients receiving care above evidence-based best practice time thresholds show an increased risk of mortality, repeat myocardial infarction, congestive heart failure, and hospitalization [4-7].

STEMI is diagnosed based on an electrocardiogram (ECG), and care for patients with STEMI requires rapid coordination of several services, including emergency medical services (EMS), emergency medicine (EM) physicians and staff, interventional cardiologists (ICs), and the PCI laboratory staff. This can range across several centers and jurisdictions that rarely share electronic medical records. Previous research shows that efficient coordination of these services through timely communication leads to improved outcomes for patients with STEMI [2]. The current process for STEMI activation for many centers in Canada relies on a variety of technologies that are poorly integrated or inefficient. For example, STEMI activation at the Hamilton General Hospital, a high-volume quaternary cardiac care hospital in Ontario, Canada, used fax for ECG transmission, which is secure but slow, inconvenient, and sometimes unavailable. To improve communication and efficiency in this pathway, a quick, easy-to-use, privacy-compliant smartphone app (SmartAMI-ACS [Strategic Management of Acute Reperfusion and Therapies in Acute Myocardial Infarction]) was developed for rapid ECG sharing between health care providers, enabling timely decision-making.

The SmartAMI-ACS app has been pilot tested with ICs and select EM physicians since May 2020 for patients presenting to Niagara Health and requiring transfer to Hamilton General Hospital for primary PCI. Although preliminary data show promise [8], additional research is required to promote the technology more widely as an evidence-based innovation for care. Similar smartphone apps with capabilities of sending ECG images to ICs for decision-making have demonstrated outcome improvement in patients with STEMI and a reduction in decision time [9-13]. However, there is limited evidence on users' perspectives regarding the barriers and enablers to implementing an app that allows for the sharing of ECG images in the care of patients with STEMI, which may influence its acceptability and long-term use. To promote the app as the standard means to communicate ECGs in the setting of patients with STEMI beyond the pilot sites in the Niagara region, we sought to gather an understanding of the perceived usefulness and acceptability of the technology in routine practice from prospective app users across a wider region (ie, Southeastern Ontario) [14]. This paper details the findings from the preimplementation surveys, which will be used to assist with the potential scale-up and implementation of the app across other provincial STEMI systems in Ontario and beyond.

Methods

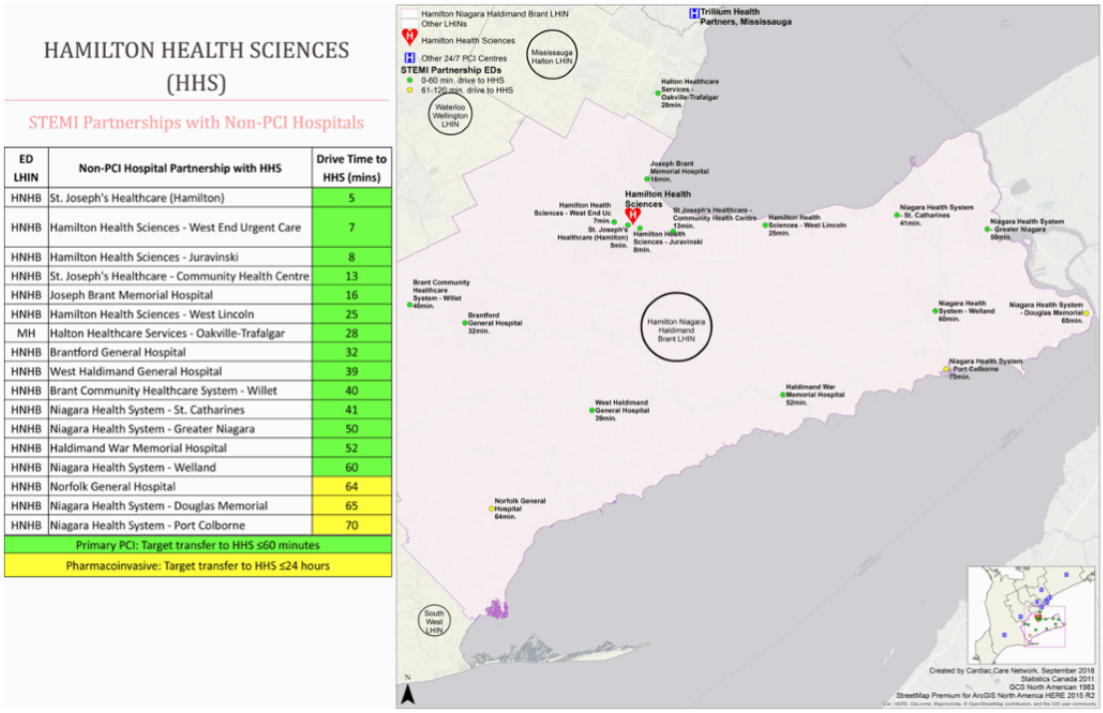
Study Design

The development and administration of the survey were led by a collaborative team consisting of medical professionals, public health researchers, and experts in implementation science. A multicenter cross-sectional study was conducted to assess the need, usefulness, and acceptability of the SmartAMI-ACS app in practice through the development and distribution of web-based surveys to key stakeholders, including ICs, EM physicians, and EMS paramedics who care for patients presenting with STEMI within the Hamilton General Hospital cardiac catheterization laboratory catchment area. The catchment area includes the Hamilton, Niagara, Haldimand, and Brant

XSL•FO
RenderX

regions, which cover an area of 6473 km^2 and a population of 1.8 million people [15,16]. Figure 1 illustrates the partner hospitals from which EM physicians were sent survey invitations, as well as a geographical map depicting the locations where sampled EMS paramedics respond to emergency calls. ICs were sampled from the primary PCI hospital indicated on the map as Hamilton Health Sciences.

Figure 1. Primary PCI hospital STEMI partnership map, adapted from the 2017 - 2018 Ontario Local Health Integration Network Annual Report [17]. PCI: percutaneous coronary intervention; STEMI: ST-elevation myocardial infarction.



ED LHIN	Non-PCI Hospital Partnership with HHS	Drive Time to HHS (mins)
HNHB	St. Joseph's Healthcare (Hamilton)	5
HNHB	Hamilton Health Sciences - West End Urgent Care	7
HNHB	Hamilton Health Sciences - Juravinski	8
HNHB	St. Joseph's Healthcare - Community Health Centre	13
HNHB	Joseph Brant Memorial Hospital	16
HNHB	Hamilton Health Sciences - West Lincoln	25
MH	Halton Healthcare Services - Oakville-Trafalgar	28
HNHB	Brantford General Hospital	32
HNHB	West Haldimand General Hospital	39
HNHB	Brant Community Healthcare System - Willet	40
HNHB	Niagara Health System - St. Catharines	41
HNHB	Niagara Health System - Greater Niagara	50
HNHB	Haldimand War Memorial Hospital	52
HNHB	Niagara Health System - Welland	60
HNHB	Norfolk General Hospital	64
HNHB	Niagara Health System - Douglas Memorial	65
HNHB	Niagara Health System - Port Colborne	70

Survey Development

In 2021, three surveys were developed and tailored for each stakeholder group (ICs, EM physicians, and EMS paramedics). These surveys were focused on gathering a better understanding of the barriers and facilitators that may impact the successful implementation of the SmartAMI-ACS app intervention on a larger scale (see Multimedia Appendices 1-3 for the surveys). Development of survey questions, analysis, and interpretation were guided by results from the Niagara Health pilot study, predicted barriers based on expert guidance from study team members (including implementation science researchers, ICs, EMS paramedics, and EM physicians) and the Consolidated Framework for Implementation Research (CFIR) [18]. The CFIR is a determinant framework, with a list of 39 constructs organized across 5 major domains, and provides a practical structure to identify multilevel factors that can impact implementation [19]. CFIR is a widely used framework in implementation science, with great breadth in its application across study design, objectives, settings, and outcome measures [20]. Through an iterative process involving input from research team members, questions were drafted and selected for inclusion in the surveys around the potential need for an app to transmit ECG images, the fit of an app into existing workflow processes, the potential for an app to improve patient care and reduce treatment times, and the overall comfort level of using an app in the treatment of suspected STEMI cases. The preimplementation surveys consisted of both 7-point Likert scale questions and open-ended questions. Likert scales were used to identify participant opinions on app implementation and usefulness. We applied a 7-point scale (1=strongly disagree and 7=strongly agree), as this has been shown to reach the upper limits of the scale's reliability, in comparison to a 5-point scale [21]. The surveys developed for each group are fully presented in Multimedia Appendices 1-3. Additional questions on the usage of the STEMI activation system (known locally as the Heart Investigation Unit Hotline) were included; however, results from those questions will not be reported in this paper.

Sample Population

Our study targets the health care providers who use the STEMI activation system for a quaternary care hospital to activate and respond to a STEMI. These include ICs, EM physicians, and EMS paramedics. All IC and EM physicians working at one of the referring centers (Figure 1).

Survey Distribution

The preimplementation surveys were hosted in REDCap (Research Electronic Data Capture; Vanderbilt University), and questionnaires were designed for anonymous completion. Senior

XSL•FO
RenderX

leaders of each survey group used a list of physician and EMS paramedic contact details provided by administrative distribution lists to develop the survey sampling frame. Senior leaders were asked to distribute a survey invitation by email to their staff, which included a summary of the evaluation project, a link to the survey, and details about consent; they were also encouraged to send out a reminder email to staff approximately 1 month after the initial invitation was sent.

Data Analysis

Quantitative data were analyzed using descriptive statistics of frequencies through Excel Microsoft 365, with continuous variables presented as means and SDs. Open-ended data were extracted verbatim and analyzed using an inductive qualitative approach, with transcripts coded into descriptive qualitative codes and then collapsed into themes using qualitative data analysis software (NVivo 12, QSR International). Initial coding was completed independently by 2 researchers (KJC and KM), and the development of themes was carried out collaboratively by 2 researchers (KJC and KM), with themes then validated by a third researcher (JC) to ensure rigor and enhance the trustworthiness of data analysis.

Ethical Considerations

Consent was obtained from participants through the use of a question at the beginning of each survey. As survey responses were anonymous, participants could not withdraw their data after submission; however, participants were notified that they could withdraw consent by exiting the survey without submitting any responses. There was no compensation provided for participation. This study was approved by the Hamilton Integrated Research Ethics Board (REB #7691). That approval did not include the ability to publish verbatim survey responses.

Results

Quantitative

Between the period of June and August 2021, senior leaders of each survey group sent out survey requests to physicians and EMS paramedic staff, with response rates varying among each of the stakeholder groups sampled. Of the 10 ICs at the Hamilton General Hospital, 9 (90%) completed the survey. Of the 223 invited EM physicians, 51 (22.9%) completed the survey, and 93 (8.2%) of the 1138 invited EMS paramedics completed the survey.

Of the 51 EM physicians, 49 (96%) reported that they currently submit ECGs to ICs. Nearly all (50/51, 98%) did so through SMS text messages, while 39% (20/51) also reported sending ECGs via fax. All 9 ICs indicated that they received ECGs via SMS text message, while 33% (3/9) reported receiving ECGs also via fax. EMS paramedics reported that they rarely submitted prehospital ECGs to ICs (7/93, 8%), and physical ECGs were typically exchanged in person at patient handoff.

Among all groups, there was recognition that current practices for sharing ECGs allowed room for improvement, accepting that fax can be inconvenient and SMS text messages are not secure. When asked whether there was a need for a smartphone app to transmit ECGs, ICs (mean 6.67, SD 0.50), EM physicians (mean 5.57, SD 1.30), and EMS paramedics (mean 5.79, SD 1.45) consistently agreed one was needed. Furthermore, among ICs, EM physicians, and EMS paramedics, use of the app was believed to fit well into existing workflow processes, with mean Likert responses of 6.67 (SD 0.50), 5.65 (SD 1.31), and 5.56 (SD 1.77), respectively. Staff also noted feeling comfortable with using an app for ECG transmission and review, with mean Likert responses of 6.44 (SD 0.88), 5.51 (SD 1.92), and 5.22 (SD 2.05) reported. Table 1 highlights viewpoints on the implementation of a smartphone app to transmit ECGs from each group.

XSL•FO
RenderX

Table . Mean and SD scores from 7-point Likert scale responses on the implementation of a smartphone app to transmit ECGs[a] in cases of suspected STEMI[b].

	Interventional cardiologists, mean (SD)	EM[c] physicians, mean (SD)	EMS[d] paramedics, mean (SD)
There is a need for a smartphone app for the transmission and review of ECGs	6.67 (0.5)	5.57 (1.3)	5.79 (1.45)
The implementation of a smartphone app fits well within existing work processes in my setting	6.67 (0.5)	5.65 (1.31)	5.56 (1.77)
Using a smartphone app for synchronous review of ECGs will help to reduce the time to STEMI treatment	6 (1.32)	5 (1.55)	5.17 (1.65)
Using a smartphone app for synchronous review of ECGs will help to reduce the number of false STEMI activations	6.33 (1.12)	4.59 (1.66)	5.69 (1.22)
How comfortable are you with the idea of using a smartphone app for the transmission and review of ECGs?	6.44 (0.88)	5.51 (1.92)	5.22 (2.05)

[a]ECG: electrocardiogram.

[b]STEMI: ST-elevation myocardial infarction.

[c]EM: emergency medicine.

[d]EMS: emergency medical service.

Qualitative

Through inductive qualitative analysis, 24 codes were identified for ICs, 46 codes were identified for EM physicians, and 57 codes were identified for EMS paramedics (Table 2). The codes were grouped into four key themes, which housed both barriers and enablers to app implementation: (1) technological issues and constraints; (2) perceived capabilities, alternatives, and comparisons of the app; (3) attitudes and beliefs toward the app; and (4) impact on current workflow and requirements for implementation (Table 3). Table 4 demonstrates how the identified barriers were mapped to implementation strategies.

Commonly reported barriers to implementation were concerns over technological failures within the app. Given that STEMI activation is a medical emergency, there were concerns that app updates could be required during urgent scenarios, making the app unusable and thus affecting clinical care. EMS paramedic providers and physicians practicing at hospitals with unreliable internet signals also expressed concerns about their ability to reliably call and transmit ECGs using their smartphones. Another important barrier was the uncertainty about the proven risks and benefits of the app, as this is a new technology without extensive real-world data. Privacy and security of the tool to transmit personal health information was raised. There was concern that the app could take more time than the current clinical workflows, thus worsening the process of STEMI activation. Respondents also expressed a desire to follow up with referring or accepting physicians to discuss the case in the future, which is a feature not offered via the app.

XSL•FO
RenderX

Table . Codebook created for the analysis of themes on the barriers and enablers to app implementation.

	Interventional cardiologists	EM[a] physicians	EMS[b] paramedics
Technological issues and constraints	• App needs to work well • App not working • App requires redownloading • Biometric authentication • Log-in and password issues	• App data security • App not working • App registration process • Backup option needed in case app fails • Cell signal issues • Concerns over alignment of app with privacy standards • Patient data privacy • Reliability of app technology • Speed of ECG[c] transmission • Technology failure • User log-in and password issues • Wi-Fi signal issues	• App data security • App not working • Backup option needed in case app fails • Cell signal issues • Concerns over alignment of app with privacy standards • No time to update app • Patient data privacy • Reliability of app technology • Technology failure
Perceived capabilities, alternatives, and comparisons of the app	• Alternative to app: critical hotline • Comparison to current practices • Desire to send or receive additional clinical information • Follow-up capacity to referring physician • Impact on patient care • Storage for physician dialogue	• Access to app by temporary physicians • Comparison with current practice • Designated phone in emergency department • ECG image quality • Follow-up capability to referring physician • Impact on patient care • In favor of app • Relative advantage • Storage for physician dialogue • Text/fax along with dialogue considered safest • Usability across multiple platforms	• Impact on patient care • Alternative to app – ECG monitor direct submission • Alternative to app – iPad • Alternative to app – cellular enabled computers • Alternative to app – video telemedicine • Alternative to app – email • Benefit of IC[d] consultation • Comparison with current practice • Confirmation needed that ECG sent successfully • ECG image quality • Limitations of point-in-time ECGs
Attitudes and beliefs toward the intervention	• Ease of use • Evaluate adoption in clinical practice • Evaluate unexpected consequences • Interventionalist requested texting • Positive experience with app	• Alternative to app: Epic Systems Corporation's electronic medical records • Alternative to app: hypercare • App implementation creates opportunity for error • Data security on personal device • Ease of use • Frequency of use • Interventionalist requested texting • Positive experience with app • Resistance to new system process • Responsibility – unit staff/coordinators • Uncertainty about harms and benefits • Use of personal cellular data • Use of personal device	• App not needed • Appropriately shifts decision-making • Comfort with using technology • Concerns over liability if app does not work • Ease of use • Impact on paramedic ECG interpretation skills • Impact on patient care • In favor of app • Negative response from ICs • Not in favor of app • Paramedics can diagnose well without the app • Reduces paramedic autonomy • Testing of app required before implementation • Uncertainty about harms and benefits • Use of personal cellular data • Use of personal device • Use of service provided smartphone

XSL•FO
RenderX

	Interventional cardiologists	EM[a] physicians	EMS[b] paramedics
Impact on current workflow and requirements for implementation	• Managing expectations of emergency medicine physicians • More frequent review by ICs • Need for contact information of referring physician • Opportunity for physician dialogue • Scope of appropriate patient cases • Time to complete app process • Time to reset password • Timely access to ECG by IC	• Administrative burden • Asynchronous training • Need for physician dialogue • No time to update app • Quick training for new physicians • Response time by IC • Scope of appropriate patient cases • Time to complete app process • Training recipients – unit coordinators • Video training	• Increased scene time • More frequent review by ICs • Need for physician dialogue • Need for training • Outcome – transport time • Outcome – false activation • Outcome – increase in-field time • Outcome – cost • Outcome – timely treatment • Practicalities of using phone while in the ambulance • Practicalities of using phone while in the field • Practicalities of using phone while wearing PPE[e] • Public image of using phone on scene • Response time by IC • Time to complete app process • Training – case scenario • Training – ECG interpretation • Training – in-person • Training – phone and text support • Training – video training

[a]EM: emergency medicine.

[b]EMS: emergency medical service.

[c]ECG: electrocardiogram.

[d]IC: interventional cardiologist.

[e]PPE: personal protective equipment.

XSL•FO
RenderX

Table . Barriers and enablers related to the implementation of a smartphone app to transmit ECGs[a] in cases of suspected STEMI[b] across all sampled groups.

Themes and barriers	Enabler	CFIR[c] construct	CFIR construct definition [18]	ICs[d]	EM[e] physicians	EMS[f] paramedics
Technological issues and constraints						
Technology failure	—[g]	Complexity	Perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy, and number of steps required to implement.	✓	✓	✓
Cell signal issues	—	Complexity	Perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy, and number of steps required to implement.	—	✓	✓
User log-in, password recall, and registration issues	—	Complexity	Perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy, and number of steps required to implement.	✓	✓	✓
—	Speed of ECG submission	Relative advantage	Stakeholders' perception of the advantage of implementing the intervention versus an alternative solution.	✓	✓	✓
ECG image quality	—	Knowledge and beliefs about the intervention	Individuals' attitudes toward and value placed on the intervention as well as familiarity with facts, truths, and principles related to the intervention.	—	✓	✓
—	ECG image quality	Knowledge and beliefs about the intervention	Individuals' attitudes toward and value placed on the intervention as well as familiarity with facts, truths, and principles related to the intervention.	✓	—	—

XSL•FO
RenderX

Themes and barriers		Enabler	CFIR[c] construct	CFIR construct definition [18]	ICs[d]	EM[e] physicians	EMS[f] paramedics
	Privacy and security	—	Knowledge and beliefs about the intervention	Individuals' attitudes toward and value placed on the intervention as well as familiarity with facts, truths, and principles related to the intervention.	—	✓	✓
	Use of personal data allowance	—	Knowledge and beliefs about the intervention	Individuals' attitudes toward and value placed on the intervention as well as familiarity with facts, truths, and principles related to the intervention.	—	✓	—
Perceived capabilities, alternatives, and comparisons of the app							
	Time to complete app process	—	Complexity	Perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy, and number of steps required to implement.	✓	✓	✓
	Desire to send/receive additional clinical information	—	Compatibility	The degree of tangible fit between meaning and values attached to the intervention by involved individuals, how those align with individuals' own norms, values, and perceived risks and needs, and how the intervention fits with existing workflows and systems.	—	—	✓

XSL•FO
RenderX

Themes and barriers	Enabler	CFIR[c] construct	CFIR construct definition [18]	ICs[d]	EM[e] physicians	EMS[f] paramedics
—	Opportunity for physician dialogue to discuss borderline cases	Compatibility	The degree of tangible fit between meaning and values attached to the intervention by involved individuals, how those align with individuals' own norms, values, and perceived risks and needs, and how the intervention fits with existing workflows and systems.	—	✓	✓
Opportunity for follow-up between referring physician and IC	—	Compatibility	The degree of tangible fit between meaning and values attached to the intervention by involved individuals, how those align with individuals' own norms, values, and perceived risks and needs, and how the intervention fits with existing workflows and systems.	✓	✓	—
Alternative practices (fax) believed to align more with privacy standards	—	Relative advantage	Stakeholders' perception of the advantage of implementing the intervention versus an alternative solution.	—	✓	✓
—	Secure solution compared to texting images	Relative advantage	Stakeholders' perception of the advantage of implementing the intervention versus an alternative solution.	✓	✓	—
Attitudes and beliefs toward the intervention						
—	Ease of use	Complexity	Perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy, and number of steps required to implement.	✓	—	—

XSL•FO
RenderX

Themes and barriers	Enabler	CFIR[c] construct	CFIR construct definition [18]	ICs[d]	EM[e] physicians	EMS[f] paramedics
—	In favor of app	Knowledge and beliefs about the intervention	Individuals' attitudes toward and value placed on the intervention as well as familiarity with facts, truths, and principles related to the intervention.	✓	✓	✓
Resistance to new system procedure	—	Individual stage of change	Characterization of the phase an individual is in, as he or she progresses toward skilled, enthusiastic, and sustained use of the intervention.	—	✓	✓
—	Positive experience with using app	Individual stage of change	Characterization of the phase an individual is in, as he or she progresses toward skilled, enthusiastic, and sustained use of the intervention.	✓	—	—
Uncertainty about harms and benefits	—	Evidence, strength, and quality	Stakeholders' perceptions of the quality and validity of evidence supporting the belief that the intervention will have desired outcomes.	—	✓	✓
Need for evaluation of app in clinical practice	—	Evidence, strength, and quality	Stakeholders' perceptions of the quality and validity of evidence supporting the belief that the intervention will have desired outcomes.	—	✓	✓
—	Potentially improve patient outcomes	Relative advantage	Stakeholders' perception of the advantage of implementing the intervention versus an alternative solution.	✓	✓	—
Impact on current workflow and requirements for implementation						

XSL•FO
RenderX

Themes and barriers	Enabler	CFIR[c] construct	CFIR construct definition [18]	ICs[d]	EM[e] physicians	EMS[f] paramedics
Lead to more frequent review by ICs	—	Complexity	Perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy, and number of steps required to implement.	✓	—	—
Administrative burden	—	Complexity	Perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy, and number of steps required to implement.	—	✓	✓
Time needed to update app or reset password if needed	—	Complexity	Perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy, and number of steps required to implement.	✓	✓	✓
—	Response time by IC	Relative advantage	Stakeholders' perception of the advantage of implementing the intervention versus an alternative solution.	—	✓	✓
—	Timely access to ECG by IC	Relative advantage	Stakeholders' perception of the advantage of implementing the intervention versus an alternative solution.	✓	✓	✓
—	Diverse training options available	Engaging	Attracting and involving appropriate individuals in the implementation and use of the intervention through a combined strategy of social marketing, education, role modeling, training, and other similar activities.	✓	✓	✓

XSL•FO
RenderX

Themes and barriers	Enabler	CFIR[c] construct	CFIR construct definition [18]	ICs[d]	EM[e] physicians	EMS[f] paramedics
Increase on-scene time in the field	—	Compatibility	The degree of tangible fit between meaning and values attached to the intervention by involved individuals, how those align with individuals' own norms, values, and perceived risks and needs, and how the intervention fits with existing workflows and systems.	—	—	✓
Degradation of ECG interpretation skills	—	Other personal Attributes	A broad construct to include other personal traits such as tolerance of ambiguity, intellectual ability, motivation, values, competence, capacity, and learning style.	—	—	✓

[a]ECG: electrocardiogram.

[b]STEMI: ST-elevation myocardial infarction.

[c]CFIR: Consolidated Framework for Implementation Research.

[d]IC: interventional cardiologist.

[e]EM: emergency medicine.

[f]EMS: emergency medical service.

[g]Not applicable.

Discussion

Principal Results

Responses from the preimplementation surveys addressed the objectives of this study, which sought to develop an understanding of the barriers and enablers to implementing an app that facilitates emergency cardiac communication. The primary use case is to securely transmit ECG images and enable telephone calls directly via the app in the case of suspected STEMI in a large health region in Ontario, Canada. Although previous studies have examined the diagnostic use of ECG image transmission via an app for STEMI care [11,22,23], this paper is the first to report on the barriers and enablers to implementing such an app from target users, allowing for tailored implementation strategies to be used in future to increase uptake. Overall, survey findings demonstrated agreement on the perceived need for an app and consensus that the app would fit well within existing workflow processes, reduce time to STEMI treatment, and reduce inappropriate activation of the on-call IC team. Commonly reported barriers were concerns over technological challenges, privacy issues, and cell phone reception strength.

While all 3 groups shared concerns over potential technological challenges, privacy issues, and cell phone reception strength, there were differences in the barriers and enablers reported by each group (Table 3). As the app was previously implemented in the IC group during the pilot phase in 2020, several of the barriers expressed by the EM physicians and EMS paramedics were not reported by the ICs. For example, the ICs were aware of app functionality and ECG image quality and so did not report these as a barrier to use. However, the majority of EM physicians and all EMS paramedics had no prior experience with using the app at the time of survey collection, and so responses illustrate apprehension of the unknown. Across each stakeholder group, common enablers were the expected speed of ECG image sharing and the associated timely access to that knowledge distribution. All groups reported comments that were both in favor and against app implementation. Given the rapid growth of smartphone technology in recent decades and its increasing adoption among consumers, it is unsurprising that survey responses reflected a high level of comfort with using an app to transmit ECG images. Furthermore, mobile health technology has the potential to enhance system efficiencies and streamline workflow processes, which may account for our

XSL•FO
RenderX

findings indicating both a recognized need for an app to transmit ECG images and its perceived compatibility with existing practices. From the barriers and enablers that were identified in the survey responses, select CFIR constructs were found to be represented more frequently than others (Table 3). Under the technological issues and constraints theme, CFIR constructs such as complexity and knowledge and beliefs about the intervention were commonly referenced. This indicated that the individual's personal perception on how difficult the app would be to use and their personal attitudes toward using the app for cases of suspected STEMI were going to be contributing factors to whether they took the necessary steps to download and use the app. In the perceived capabilities and comparisons theme, CFIR constructs of compatibility and relative advantage were frequently cited. Participants' willingness to use the app depended on whether they saw it as valuable, fulfilling a clinical need, and offering advantages over existing ECG transmission methods. In the attitudes and beliefs toward the intervention theme, there were no dominant CFIR constructs; however, responses indicated that participant perception on the evidence supporting the need for an app to transmit ECG images influenced whether they themselves would download and use it. Finally, in the impact on current workflow and requirements for implementation theme, complexity and relative advantage were again most commonly referenced, demonstrating that the perceived difficulty of incorporating the app into their current workflow processes and their perception on the benefits to using the app, would impact uptake. Our findings of the barriers identified in the survey responses, combined with our analysis of the most commonly cited CFIR constructs, facilitated the development of implementation strategies for app use and scale-up.

Results from the web-based surveys allowed for the application of stakeholder-specific implementation science interventions that supported the uptake of the app on a wider scale across the health region. For example, we obtained an understanding of the preferred mechanism of training on app use and applied appropriate education strategies, such as offering Zoom drop-in question and answer training sessions at accessible times during the day, to accommodate shift work hours. Furthermore, we cultivated relationships with opinion leaders (EM chiefs) and early adopters and sought advice from experts in implementation science research to maximize survey responses and increase uptake and use of the app. To address password recall concerns among ICs, biometric authentication was introduced to streamline log-in. Finally, key barriers related to concerns over technological challenges and privacy issues were addressed with the development and distribution of educational materials, such as user guides that highlighted multiple successful security tests from internal privacy, security, and information technology staff along with external third-party vendors. Table 4 illustrates the barriers by theme and the implementation strategies that were used to facilitate the uptake and use of the app.

This study focuses on a specific region in Ontario, Canada; however, the results can be generalized to other regions or countries where an app is being considered for inclusion in patient care by health care providers, as noted by the similarity of findings to published literature. A systematic review of barriers and facilitators to using digital health technologies representing North America, Europe, Asia, Africa, and Latin America highlighted infrastructure and technical barriers as well as privacy and security as commonly reported concerns by health care providers when using app-based technology [24]. As reported in our findings, concerns over insufficient network and connectivity speed were found, with a higher prevalence of these barriers being reported by those working in rural or remote regions. Furthermore, a published study on app use among nurses reported similar findings, highlighting perceived usefulness and ease of use as key factors influencing adoption [25]. The alignment of this study's findings with previously published research on barriers and enablers to app use by health care providers underscores the importance of engaging users before implementation, which can assist in developing targeted strategies to enhance the adoption and integration of new technology.

These findings are relevant to clinical practice improvements as they highlight the potential for mobile health technology to enhance communication in STEMI care and improve coordination among health care providers. By addressing identified barriers and leveraging enablers, the successful integration of an app for ECG transmission could improve the efficiency of STEMI activation, reduce unnecessary patient transfers, and ultimately enhance patient outcomes by minimizing delays in critical care. While emerging technology can allow for the integration of a machine learning model to enhance the ECG diagnosis of acute coronary occlusion, research testing the precision and sensitivity of such models combined the judgment of trained health professionals for ECG review [26]. The findings presented in this paper can support the integration of new technologies, including artificial intelligence–enabled apps, into acute STEMI care, as without understanding how to effectively implement and maintain an app in the acute management of STEMI, similar barriers are likely to hinder the successful adoption of artificial intelligence technologies in this setting.

Following this analysis of barriers and enablers to implementation, further research is needed to evaluate the impact of mobile health technology on the care of suspected STEMI cases. Specifically, it is important to assess whether addressing the identified barriers and leveraging enablers leads to successful implementation strategies that enhance app adoption and sustained use. Further analysis from this implementation will explore data on app uptake, usage, and effectiveness, with findings to be published in subsequent papers.

XSL•FO
RenderX

Table . Mapping barriers identified from survey responses to implementation strategies.

Theme	Barrier	Implementation strategy
Technological issues and constraints	• Potential to forget log-in credentials which could delay care • Privacy and security • Time needed to update app • Technology failure	• Biometric authentication (face/fingerprint ID) was enabled for ICs[a] to assist with faster log-in and to address concerns over password recollection • Targeted education messages were developed, which noted third-party security testing to relay privacy concerns • App updates automatically • In the app invitation email, users are encouraged to notify the study team should they have any issues with the app functionality or logging-in to the app
Perceived capabilities, alternatives, and comparisons of the app	• Follow-up capacity to referring physician • Usability across multiple platforms	• Referring physician required to call the STEMI[b] activation number through the app and speak to on-call IC, allowing for the opportunity to discuss patient care • User guide and app invitation email provide links to both the Apple App Store and the Google Play Store
Attitudes and beliefs toward the app	• Ease of use • Resistance to new system process • App believed to be unnecessary	• User guide contains images depicting app simplicity • Cultivated relationships with opinion leaders (EM[c] chiefs) and early adopters from the pilot study • With the support of EM chiefs, a newsletter was sent out to EM physicians that included important reminders about app use, a link to the education session recording, and a general update on app downloads across the region to demonstrate uptake by peers
Impact on current workflow and requirements for implementation	• Concerns about how app would fit into existing workflows • Asynchronous training	• Developed a user guide with clearly structured guidance and step-by-step instructions on how and when the app should be used • Hosted a collaborative web-based education session for EM physicians discussing the regional STEMI program and the app. A recording of the session was made available for asynchronous learning on a dedicated web page that also housed additional educational materials. In addition, scheduled web-based drop-in Q&A sessions at multiple time periods throughout the day during time of app launch, to accommodate shift work hours

[a]IC: interventional cardiologist.

[b]STEMI: ST-elevation myocardial infarction.

[c]EM: emergency medicine.

Limitations

As the surveys were distributed prior to the majority of EM physicians and EMS paramedics gaining access to the app, there was limited understanding of app features, capabilities, and usability among these 2 groups. This made it difficult for participants to answer certain questions and was reflected in survey responses that noted inadequate knowledge affected their ability to reliably answer some questions. Nonetheless, this provided valuable, unbiased responses to questions aimed at understanding the need for such a tool. Additionally, as links to the survey were shared directly by senior leaders of each stakeholder group, it was challenging for the research team to manage follow-up emails that may have improved survey

XSL•FO
RenderX

completion rates, as well as calculate accurate response rates. However, this was the preferred means of communication by the senior leaders to maintain anonymity, confidentiality, and avoiding coercion to respond to questionnaires. It is also possible that response rates were improved as participants were being directly asked to complete the survey by the chief or chair of their department. This study leveraged a convenience sampling strategy, which can introduce selection and sampling bias. As such, caution is necessary in the generalization of the results. Reassuringly, there was significant representation from each hospital and paramedic region with similar themes and responses for both the qualitative and quantitative questions.

Conclusions

The implementation of system-level interventions in health care can be affected by unexpected barriers, leading to poor uptake and acceptance. In order to circumvent the challenges of implementing change to an organization's culture, understanding the variables that influence or have an impact on the population's willingness to accept and adopt the change is critical. This study exemplifies the importance of identifying barriers faced by the target population and using that knowledge to create successful implementation strategies to optimize research interventions. Results from the preimplementation surveys disseminated to ICs, EM physicians, and EMS paramedics helped inform the strategies for the SmartAMI-ACS app implementation program in Hamilton and the surrounding region. They will be vital in informing active scaling of the app into additional areas throughout the province and local ambulance services.

Acknowledgments

We wish to thank the emergency medicine physicians and emergency medical service paramedic chiefs who assisted in distributing the survey invitation email. Their contribution made it possible to gather important insight into the barriers and facilitators to the implementation of the smartphone app technology. This study was funded through the Hamilton Academic Health Sciences Organization Innovation Fund received in 2021 - 2024 (HAH-21 - 015).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. All relevant data are included in this article.

Authors' Contributions

HM, MKN, JDS, and MM contributed to the design of the study. KJC, KM, and MC managed the ethics application, implementation, and evaluation of the app. KJC, KM, and JC contributed to the analysis of survey findings. All authors contributed to writing the manuscript and approved the final version.

Conflicts of Interest

HM is an inventor and sole owner of the background intellectual property of the smartphone app used in this study. There is no specific financial conflict of interest to declare as the tool is being used for research. However, it may be commercialized in the future. None of the other authors have a relevant conflict of interest to declare.

Multimedia Appendix 1
STEMI activation questionnaire: interventional cardiologists. STEMI: ST-elevation myocardial infarction.
[DOCX File, 23 KB - mhealth_v13i1e60173_app1.docx]

Multimedia Appendix 2
STEMI activation questionnaire: emergency medicine physicians. STEMI: ST-elevation myocardial infarction.
[DOCX File, 25 KB - mhealth_v13i1e60173_app2.docx]

Multimedia Appendix 3
STEMI activation questionnaire: paramedics. STEMI: ST-elevation myocardial infarction.
[DOCX File, 24 KB - mhealth_v13i1e60173_app3.docx]

References

1. Training bulletin STEMI hospital bypass 118. Ministry of Long-Term Care. 2016 Jan. URL: https://www.health.gov.on.ca/en/pro/programs/emergency_health/docs/ehs_training_blltn118_en.pdf [accessed 2025-06-02]
2. Wong GC, Welsford M, Ainsworth C, et al. 2019 Canadian Cardiovascular Society/Canadian Association of Interventional Cardiology guidelines on the acute management of ST-elevation myocardial infarction: focused update on regionalization and reperfusion. Can J Cardiol 2019 Feb;35(2):107-132. [doi: 10.1016/j.cjca.2018.11.031] [Medline: 30760415]

XSL•FO
RenderX

3. Cardiac, stroke & vascular centres. CorHealth Ontario. URL: https://www.corhealthontario.ca/cardiac-stroke-&-vascular-centres [accessed 2025-06-02]
4. Lambert L, Brown K, Segal E, Brophy J, Rodes-Cabau J, Bogaty P. Association between timeliness of reperfusion therapy and clinical outcomes in ST-elevation myocardial infarction. JAMA 2010 Jun 2;303(21):2148-2155. [doi: 10.1001/jama.2010.712] [Medline: 20516415]
5. McNamara RL, Wang Y, Herrin J, et al. Effect of door-to-balloon time on mortality in patients with ST-segment elevation myocardial infarction. J Am Coll Cardiol 2006 Jun 6;47(11):2180-2186. [doi: 10.1016/j.jacc.2005.12.072] [Medline: 16750682]
6. Brodie BR, Gersh BJ, Stuckey T, et al. When is door-to-balloon time critical? Analysis from the HORIZONS-AMI (Harmonizing Outcomes with Revascularization and Stents in Acute Myocardial Infarction) and CADILLAC (Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications) trials. J Am Coll Cardiol 2010 Jul 27;56(5):407-413. [doi: 10.1016/j.jacc.2010.04.020] [Medline: 20650362]
7. Sharing best practices: timely access to reperfusion therapy for STEMI in Ontario. CorHealth Ontario. 2018 Jan. URL: https://www.corhealthontario.ca/Sharing-Best-Practices.pdf [accessed 2025-06-02]
8. Mir H, Syed T, Schwalm J, et al. Smartphone application for STEMI activation: a pilot study. Can J Cardiol 2021 Oct;37(10):S8. [doi: 10.1016/j.cjca.2021.07.032]
9. Park JJ, Yoon CH, Suh JW, et al. Reduction of ischemic time for transferred STEMI patients using a smartphone social network system. J Am Coll Cardiol 2016 Sep 27;68(13):1490-1492. [doi: 10.1016/j.jacc.2016.07.733] [Medline: 27659471]
10. Astarcioglu MA, Sen T, Kilit C, et al. Time-to-reperfusion in STEMI undergoing interhospital transfer using smartphone and WhatsApp messenger. Am J Emerg Med 2015 Oct;33(10):1382-1384. [doi: 10.1016/j.ajem.2015.07.029] [Medline: 26299691]
11. Bradley PR, Durej MT, Gessner AE, et al. Smartphone application for transmission of ECG images in pre-hospital STEMI treatment. Presented at: 2012 IEEE Systems and Information Engineering Design Symposium (SIEDS); Apr 27, 2012; Charlottesville, VA, USA p. 118-123. [doi: 10.1109/SIEDS.2012.6215144]
12. Chao CC, Chen YC, Shih CM, et al. Smartphone transmission of electrocardiography images to reduce time of cardiac catheterization laboratory activation. J Chin Med Assoc 2018 Jun;81(6):505-510. [doi: 10.1016/j.jcma.2017.11.009] [Medline: 29287703]
13. Studencan M, Alusik D, Plachy L, et al. Significant benefits of new communication technology for time delay management in STEMI patients. PLOS ONE 2018;13(11):e0205832. [doi: 10.1371/journal.pone.0205832] [Medline: 30388116]
14. Mir H, Cullen KJ, Mosleh K, et al. Smartphone app for prehospital ECG transmission in ST-elevation myocardial infarction activation: protocol for a mixed methods study. JMIR Res Protoc 2024;13:e55506. [doi: 10.2196/55506]
15. Census Profile, 2021 Census of Population: profile table. Statistics Canada. URL: https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&GENDERlist=1,2,3&STATISTIClist=1,4&HEADERlist=0&DGUIDlist=2021A00053520005&SearchText=toronto [accessed 2025-06-02]
16. Census Profile, 2016 Census: Hamilton Niagara Haldimand Brant, [Health region, December 2017], Ontario and Ontario [Province]. Statistics Canada. URL: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=HR&Code1=3048&Geo2=PR&Code2=35&SearchText=Hamilton%20Niagara%20Haldimand%20Brant&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=HR&GeoCode=3048&TABID=1&type=0 [accessed 2025-06-02]
17. Hamilton Niagara Haldimand Brant LHIN: 2017-2018 annual report. Ontario Local Health Integration Network. 2018. URL: https://healthcareathome.ca/wp-content/uploads/2022/11/04-2017-18-HNHBLHIN-AR-EN.pdf [accessed 2025-6-2]
18. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4(1):50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
19. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
20. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the Consolidated Framework for Implementation Research. Implement Sci 2016 May 17;11(1):72. [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]
21. Abdul Malik MA, Mustapha MF, Mohamad Sobri N, et al. Optimal reliability and validity of measurement model in confirmatory factor analysis: different Likert point scale experiment. J Contemp Issues Thought 2021 Jul;11:105-112. [doi: 10.37134/jcit.vol11.9.2021]
22. Barbagelata A, Bethea CF, Severance HW, et al. Smartphone ECG for evaluation of ST-segment elevation myocardial infarction (STEMI): design of the ST LEUIS international multicenter study. J Electrocardiol 2018;51(2):260-264. [doi: 10.1016/j.jelectrocard.2017.10.011] [Medline: 29174099]
23. Mahajan S, Garg S, Sharma R, et al. Smartphone based 12 lead ECG as decision support in STEMI: a prospective, cross sectional study, non-randomised, single blinded and single-center study. Int J Community Med Public Health 2023 Mar;10(4):1410-1414. [doi: 10.18203/2394-6040.ijcmph20230916]

XSL•FO
RenderX

24. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
25. Nezamdoust S, Abdekhoda M, Rahmani A. Determinant factors in adopting mobile health application in healthcare by nurses. BMC Med Inform Decis Mak 2022 Feb 22;22(1):47. [doi: 10.1186/s12911-022-01784-y] [Medline: 35193552]
26. Al-Zaiti SS, Martin-Gill C, Zègre-Hemsey JK, et al. Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. Nat Med 2023 Jul;29(7):1804-1813. [doi: 10.1038/s41591-023-02396-3] [Medline: 37386246]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
ECG: electrocardiogram
EM: emergency medicine
EMS: emergency medical service
IC: interventional cardiologist
PCI: percutaneous coronary intervention
REDCap: Research Electronic Data Capture
SmartAMI: Strategic Management of Acute Reperfusion and Therapies in Acute Myocardial Infarction
STEMI: ST-elevation myocardial infarction

Edited by L Buis; submitted 06.05.24; peer-reviewed by A Alshatti, A Jain, D Yoo, H Al-Zayer; revised version received 25.04.25; accepted 29.04.25; published 08.07.25.

Please cite as:
Cullen KJ, Mir H, Natarajan MK, Corovic M, Mosleh K, Crawshaw J, Mercuri M, Masoom H, Schwalm JD
Exploring the Barriers and Facilitators to Implementing a Smartphone App for Physicians to Improve the Management of Acute Myocardial Infarctions: Multicenter, Mixed Methods, Observational Study
JMIR Mhealth Uhealth 2025;13:e60173
URL: https://mhealth.jmir.org/2025/1/e60173
doi:10.2196/60173

© Katelyn J Cullen, Hassan Mir, Madhu K Natarajan, Marija Corovic, Karen Mosleh, Jacob Crawshaw, Mathew Mercuri, Hassan Masoom, JD Schwalm. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy, Safety, and Cost-Effectiveness of "Internet + Pharmacy Care" Via the Alfalfa App in Warfarin Therapy Management After Cardiac Valve Replacement: Randomized Controlled Trial

Yiyi Qian[1], MS; Weizhao Chen[2], MS; Bin Zhou[2], MS; Jiangya Li[3], MS; Yuanyuan Guo[4], PhD; Zhiying Weng[3], PhD; Jinhua Zhang[5], PhD

[1]Clinical Research and Trial Center, The Second Affiliated Hospital of Kunming Medical University, 1st floor, Building 1, Inpatient Building, #374 Dianmian Road, Kunming, China

[2]The Second Clinical College of Kunming Medical University, Kunming, China

[3]School of Pharmacy, Kunming Medical University, Yunnan, Kunming, China

[4]Department of Vascular Surgery, Fuwai Yunnan Hospital, Chinese Academy of Medical Sciences, Yunnan, Kunming, China

[5]Department of Pharmacy, Fujian Maternity and Child Health Hospital College of Clinical Medicine for Obstetrics and Gynecology and Pediatrics, Fuzhou, China

Corresponding Author:
Yiyi Qian, MS
Clinical Research and Trial Center, The Second Affiliated Hospital of Kunming Medical University, 1st floor, Building 1, Inpatient Building, #374 Dianmian Road, Kunming, China

Abstract

Background: Anticoagulation management is important in preventing complications in patients undergoing cardiac valve replacement. The development of mobile apps offers new opportunities for the management of long-term anticoagulants. However, there is a lack of randomized controlled trials evaluating the effectiveness, safety, cost-effectiveness, and user demand for internet-based anticoagulation management.

Objective: This study aimed to evaluate the efficacy, safety, and cost-effectiveness of a 3-month warfarin dose adjustment mobile app Alfalfa compared to offline management in patients postcardiac valve replacement. We also explored the app's feasibility on user satisfaction and demand.

Methods: This study was a randomized controlled trial with assessments conducted at baseline and at a 3-month follow-up. Participants were eligible if they had been on warfarin therapy for at least 3 months, received warfarin management either through the Alfalfa app or through pharmacist-led anticoagulation outpatient clinic visits, consented to regular follow-ups, and had not experienced serious bleeding or thrombotic events in the 3 months before warfarin treatment. A *P* value of ≤.05 was considered statistically significant.

Results: A total of 405 participants were included in the analysis. The time in therapeutic range was significantly higher in the Alfalfa app group than in the offline group (66.46% vs 46.65%, *P*<.001). Participants in the Alfalfa app group had a higher monitoring frequency (8.14 vs 4.47, *P*<.001) and a greater percentage of international normalized ratio values within the target range (896/1660, 53.98% vs 346/899, 38.49%; *P*<.001) than those in the offline group. In addition, the Alfalfa app group exhibited lower rates of subtherapeutic (235/1660, 14.16% vs 152/899, 16.91%; *P*<.05) and extreme subtherapeutic international normalized ratio values (273/1660, 16.45% vs 186/899, 20.69%; *P*<.05) than the offline group. However, the incidence of minor bleeding was higher in the Alfalfa app group (12/204, 5.9% vs 3/201, 1.5%; *P*=.02). In terms of cost-effectiveness, the Alfalfa app group had a significantly lower average cost per test (42.37 vs 78.3, *P*<.001), average time per test (47.42 vs 90.74, *P*<.001), and cost-effectiveness ratio (385.9 vs 662.9) than the offline group. A total of 86 participants completed the satisfaction questionnaire, and the vast majority of participants expressed high levels of satisfaction with the Alfalfa App, while also providing further suggestions for improvement.

Conclusions: The integration of "Internet+Pharmacy Care" using the Alfalfa App can improve the effectiveness of warfarin anticoagulation management in patients following heart valve surgery. The Alfalfa app provides a more efficient, secure, and cost-effective solution to warfarin management than traditional offline methods.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900021920; https://www.chictr.org.cn/showproj.html?proj=36832

(JMIR Mhealth Uhealth 2025;13:e53586) doi:10.2196/53586

XSL•FO
RenderX

KEYWORDS

smartphone; mobile phone; alfalfa app; warfarin; anticoagulation; internet + pharmacy care

Introduction

Valvular heart disease is one of the most prevalent cardiac conditions in China. According to 2019 extracorporeal circulation statistics, the number of heart valve surgeries continues to rise, with approximately 73,000 procedures performed annually, accounting for nearly 30% of all cardiac surgeries [1,2]. The primary treatment modalities for valvular heart disease include artificial valve replacement and valve repair surgery, both of which require long-term oral anticoagulant therapy postoperatively. Warfarin, a traditional oral anticoagulant, which widely used in clinical practice, remains the primary medication for reducing the risk of thrombosis and stroke after heart valve replacement surgery [3-5]. Patients undergoing biological valve replacement typically require 3 to 6 months of oral warfarin therapy, while those with mechanical valves need lifelong anticoagulation. Managing warfarin therapy poses challenges due to its narrow therapeutic range, interactions with food and drugs, and significant interindividual variability in dosage requirements. These challenges have been observed both domestically and internationally [6-8].

In recent years, anticoagulation clinics and home-based follow-up management models have been established in various countries. In developed regions of China, pharmaceutical services led by pharmacists in hospitals have moderately enhanced the effectiveness of anticoagulation treatment. However, both globally and in China, issues persist regarding the low efficiency and high use of human and material resources in postoperative anticoagulation management for patients with valve disease [6,9]. Many patients receiving oral anticoagulation therapy reside in rural areas where access to quality medical resources is limited. Even urban patients often face constraints such as doctor-to-patient ratios and limited consultation time. Outpatient physicians always struggle to provide comprehensive medication guidance and follow-up for every oral anticoagulation therapy patient, which can negatively impact the long-term treatment outcomes for these individuals [10,11].

With the development of information technology, countries worldwide are actively exploring personalized, intelligent and efficient eHealth tools based on their own anticoagulation management models. Some guidelines emphasize the potential of eHealth tools in improving anticoagulation management and medical outcomes [12]. The use of these eHealth tools to assist physicians with anticoagulation decision-making is recommended [13]. In April 2018, the General Office of the State Council of China issued the “Opinions on Promoting the Development of “Internet+Healthcare,” which clearly supports medical institutions in providing telemedicine services and promoting effective communication between doctors and patients.

In 2020, our research team developed an app called Alfalfa for warfarin dose adjustment [14], which has demonstrated significant advantages compared with other eHealth tools in China [15]. Through the app, patients can report coagulation results and health status to their doctors or pharmacists through mobile devices. Medication is then adjusted based on the recommendations, and international normalized ratio (INR) values are monitored accordingly [14]. Our previous retrospective observational cohort study explored the safety and efficacy of warfarin anticoagulation management through the Alfalfa App. The study confirmed that patients who use the app to manage experienced improved anticoagulation quality, reduced major bleeding events, and lower emergency and hospitalization rates compared with those receiving traditional offline management. This model has been demonstrated to be particularly suitable for patients in rural and remote areas [15].

Based on the previous study and the “Internet+pharmaceutical service” model we established, we further designed this prospective randomized controlled study in the Yunnan region of western China. The aim is to further investigate the safety, effectiveness, and cost-efficiency of the Alfalfa app in managing warfarin anticoagulation in patients with valvular disease and to better understand the app’s usage requirements.

Methods

Study Design

This study was a single-center randomized control trial (trial registration: ChiCTR1900021920) with assessments conducted at baseline and post intervention (3 months). Participants randomized to the intervention group received anticoagulation management through the Alfalfa App (“Internet+pharmacy care”), while the control group received offline management at a pharmacist-led anticoagulation clinic. Both groups were followed up for 3 months.

Sample Size

The primary outcome indicator, time in therapeutic range (TTR), was used to calculate the required sample size. Our previous research suggested that TTR for the web-based and offline groups was 73% and 60%, respectively [7]. With a power of 80% and a significance level of .05, 1:1 ration of test and control groups, a minimum of 201 patients per group is required, calculated using PASS 27 software. Assuming 10% dropout rate, each group needed at least 223 patients, for a total of 446 patients.

Recruitment and Selection

Participants were recruited from Fuwai Yunnan Cardiovascular Hospital. Patients diagnosed with valvular heart disease, who underwent valve surgery at the hospital received warfarin education from clinical pharmacists before discharge. After training, a validated questionnaire [16] was used to assess the patient’s understanding of warfarin management. If patients missed a question, they would be retrained and answered until all the questions were answered correctly. These patients were eligible for the study if they had been on warfarin treatment for at least 3 months, received warfarin management through either the Alfalfa app or pharmacist-led outpatient clinic visits, agreed

XSL•FO
RenderX

to regular follow-up, had no serious bleeding or thrombotic events in the 3 months before warfarin treatment. The exclusion criteria included pregnancy, switching to another anticoagulant during the study period, and having 3 or fewer INR records during the follow-up period. After obtaining informed consent and completing the baseline questionnaire, participants were randomized into either the web-based group or offline group.

Randomization and Blinding

Participants were randomized using a Microsoft Excel–generated randomization list in a 1:1 ratio. After participants completed baseline assessments, a research team member (unblinded) assigned participants to a group based on the randomization list. Due to the nature of the intervention, the study was open labeled and blinding was not feasible.

Web-Based Group (Intervention Group）

After randomization, participants in the web-based group were instructed through WeChat (Tencent) to download and install the Alfalfa app on their smartphones. Detailed instructions on app usage were provided. The Alfalfa app (Alfalfa Health Management) was used for remote warfarin management in the web-based group; a detailed description of the app has been published earlier [14,15]. Patients selected a pharmacist within the app who would manage their warfarin dosage. The chosen pharmacist reviewed INR results uploaded by the patient and provided recommendations on dosage adjustments and the timing of subsequent INR tests. The coagulation management team consist of 4 cardiologists and clinical pharmacists specializing in anticoagulation from Yunnan Cardiovascular Hospital. A senior clinical pharmacist responded to patient queries within 30 minutes of receiving uploads.

Offline Group (Control Group)

Participants in the offline group were instructed to visit the pharmacist's office for periodic INR testing. Warfarin dosage adjustments were supervised by the same senior clinical pharmacist who managed the web-based group.

INR Testing and Warfarin Dose Adjustment

All participants adjusted their warfarin dosage as needed and underwent periodic INR testing throughout the study. Participants in the web-based group could test INR at nearby hospitals and receive adjustment suggestions from clinical pharmacists through the app. Patients in the offline group were required to visit the hospital's anticoagulation clinic. Warfarin dose adjustment protocols were standardized by the coagulation management team. Monitoring was conducted weekly during the first 2 weeks of therapy and then every 2-4 weeks. The target INR range was 1.7 - 2.5 for patients with valve replacement.

Measures and Incentives

All participants received free pharmaceutical care and advice during the study period. Offline group participants were exempted from consultation fees in the anticoagulation pharmacy clinic during the study period. After the study, offline group participants were recommended to use the Alfalfa app. Registration and use of the Alfalfa app remained free for participants during and after the study. And the app also provided a free "anticoagulation circle" for patients to access educational resources of anticoagulation treatment and communicate with one another.

Primary Outcome

The primary outcome was the percentage of TTR, calculated using the standard linear interpolation method [17]. TTR is a widely recognized measure of anticoagulant control quality, with TTR ≥60% indicating high-quality anticoagulation and TTR <60% indicating low-quality anticoagulation [18,19].

Secondary Outcomes

Monitoring Frequency and INR Value Distribution

Monitoring frequency and INR distribution were assessed to evaluate compliance rates. According to the patient's target range, INR values were categorized into 5 levels referring to previous studies [20,21]: extreme subtherapeutic, subtherapeutic, therapeutic, supratherapeutic, and extreme supratherapeutic. Definitions of these categories are provided in the Multimedia Appendix 1. Patients with multiple anticoagulation indications were assigned the highest INR target based on their conditions.

Clinical Events

The incidences of clinical events were used to evaluate safety differences between the 2 groups, including bleeding events, thrombotic events, and other related incidents. Bleeding events were classified as either minor or major bleeding events according to the International Society on Thrombosis and Haemostasis (ISTH) criteria. Minor bleeding events included epistaxis, gingival or oral bleeding, skin ecchymosis, fundus bleeding, increased or prolonged menstruation, hematuria and other bleeding symptoms that could be quickly stopped. Major bleeding events were defined as any bleeding requiring hospitalization or blood transfusion, including gastrointestinal bleeding and intracranial hemorrhage [22]. Thrombotic events included transient ischemic attack, ischemic strokes, venous thromboembolism and valve thrombosis. Other clinical events included warfarin-related emergency department visits and hospitalizations.

Costs

We conducted a cost analysis for both groups to demonstrate the economic differences observed during the follow-up period. The cost-effectiveness analysis was used to identify the most optimal solution. Economic indicators included the average cost per test, total cost per capita, average time per test, overall total cost, and the cost-effectiveness ratio. The average cost per test was defined as the mean expense per test per patient, including transportation costs. The total cost per capita referred to the average expense for all tests over a 3-month period per patient, also including transportation costs. The average time per test indicated the mean time required for 1 test per patient. The cost-effectiveness ratio was calculated as the ratio of total costs to efficiency, with costs including direct expenses such as INR testing fees and transportation expenses. A patient with TTR ≥60% was considered to have achieved an effective outcome. A lower cost-effectiveness ratio suggests that the intervention is more cost-efficient, indicating a more favorable and rational solution. All currency conversions in this paper are based on the exchange rate as of December 31, 2021.

XSL•FO
RenderX

Feasibility

Feasibility measures included acceptability and demand. Acceptability was measured through a satisfaction survey administered post intervention. A questionnaire designed by our previous study [23,24] was distributed to participants through WeChat in the form of a link to assess patient satisfaction and further demand for the Alfalfa app. Since the Alfalfa App did not support real-time one-on-one chats, we distributed survey links twice daily for 1 week in a WeChat group that included all participants. Completion of the surveys was voluntary, and each participant was allowed to submit only once (Multimedia Appendix 2). The surveys were distributed to all participants who had been using the Alfalfa App for at least 3 months (not limited to the web-based group). The percentage of responses for each question was calculated once the questionnaires were returned.

Data Collection

Basic information, including demographic characteristics of all participants, was collected from electronic medical records. The data automatically recorded in the Alfalfa app included duration of anticoagulation, target INR range, INR values and warfarin doses at each report, bleeding events, thromboembolic events, emergency department visits, hospitalizations, disease state, concurrent medications, and dietary habits. Information provided by clinical pharmacists, such as adjusted warfarin dosage, timing for subsequent INR tests, and recommendations on diet and exercise, was automatically collected in the Alfalfa app. Data from the offline warfarin management service, as well as costs for all participants, were collected from electronic medical records and telephone follow-ups.

Statistical Analyses

Data were analyzed using Excel 2010 and SPSS (version 17.0, IBM Corp) statistical software. Different statistical methods were applied depending on the variable type results. An independent 2-sample, 2-tailed *t* test was used for statistical analysis of continuous variables, such as age, TTR, and costs. A chi-square test or Fisher exact test was used for categorical variables, such as sex, conditions, diseases, target INR, efficacy outcomes, adverse events, and INR results. Continuous variables are presented as mean (SD), while categorical variables are reported as frequency (percentage). Statistical significance was set at $P<.05$.

Ethical Considerations

This study complied with the Declaration of Helsinki and was approved by the Ethics Committee of Fuwai Yunnan Cardiovascular Hospital, Kunming, Yunnan, China (Ethics No. 2020-010-01). It was also registered with the Chinese Clinical Trial Registry (ChiCTR1900021920). All participants provided written informed consent before the study and signed an electronic informed consent form again upon registering for the Alfalfa App. Participants were allowed to withdraw from the study at any time during the research process. All medical records and test reports of the participants were stored in the hospital where they received care, and personal privacy information was strictly protected. No additional compensation was provided to the patients for participating in this study; however, some free services were offered during the research period (see the *Measures and Incentives* subsection). The datasets generated or analyzed during the study are available from the corresponding author upon reasonable request.

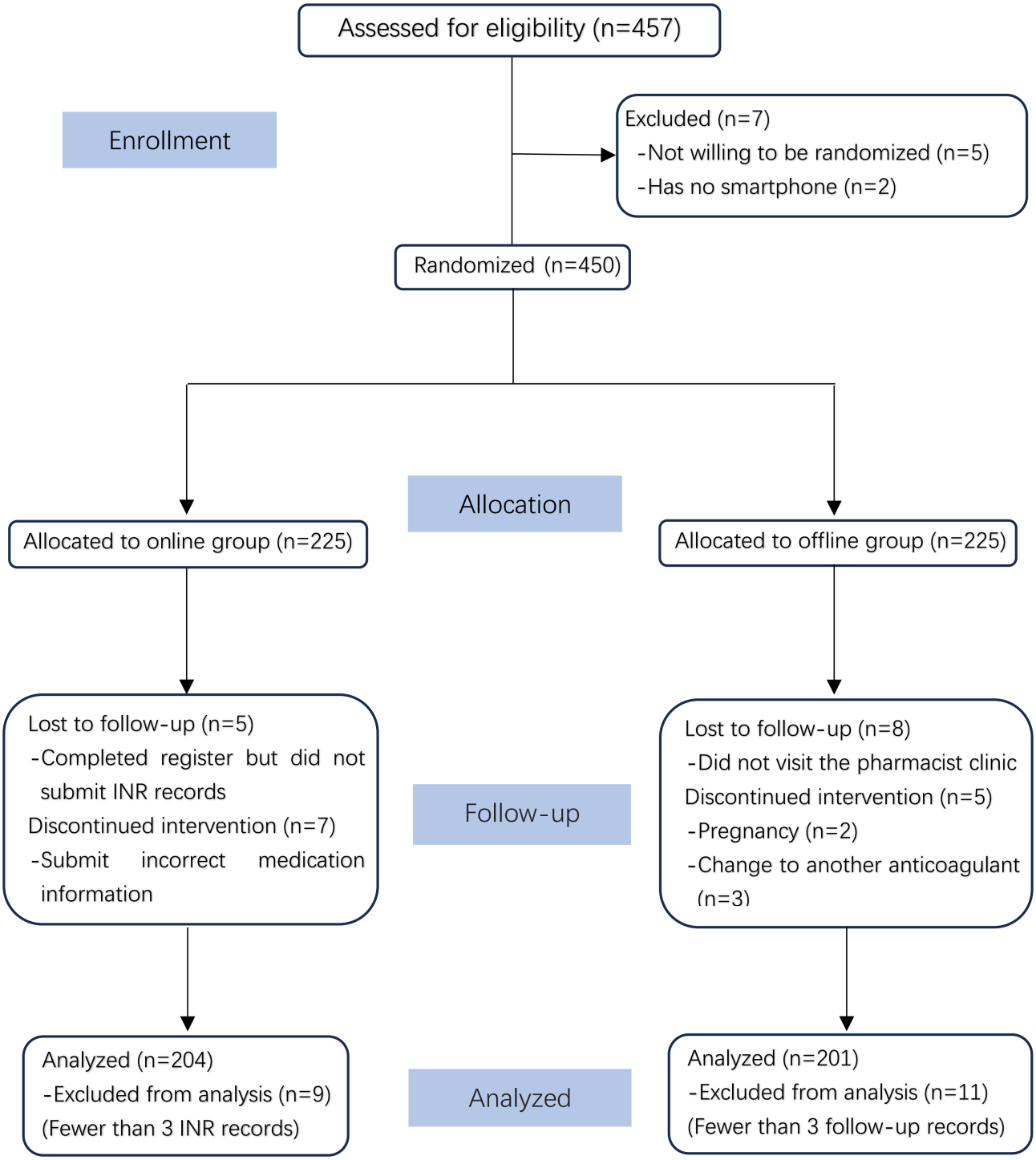
Results

Participant Enrollment

Between January 2020 and December 2021, a total of 457 patients from Fuwai Yunnan Cardiovascular Hospital were recruited for eligibility assessment, and 450 were enrolled in the study. Of these, 225 participants were randomized to the web-based group and 225 to the offline group. Participants were excluded if they completed registration but did not submit INR records (n=5), did not visit the pharmacist clinic (n=8), provided incorrect medication information (n=7), were pregnant (n=2), switched to another anticoagulant (n=3), or had fewer than 3 follow-up visits or INR records (n=20). This resulted in a total of 405 participants included in the final analysis (Figure 1). There were no missing data among these participants, and baseline characteristics were comparable across study groups (Table 1).

XSL•FO
RenderX

Figure 1. Enrollment flow chart. INR: international normalized ratio.



XSL•FO
RenderX

Table . Patients' characteristics.

Characteristics	Web-based (n=204)	Offline (n=201)	*P* value
Age (years), mean (SD)	49.66 (12.37)	51.97 (13.58)	.07
Sex (male), n (%)	115 (56.4)	101 (50.2)	.22
Height (cm), mean (SD)	163.2 (7.91)	161.78 (9.78)	.11
Weight (kg), mean (SD)	62.95 (15.53)	61.12 (13.38)	.21
Baseline INR, mean[a] (SD)	1.09 (0.13)	1.11 (0.14)	.14
Diseases, n (%)			
Hypertension	33 (16.18)	31 (15.42)	.29
Diabetes	8 (3.92)	7 (3.48)	.12
Hyperuricemia or gout	2 (1.0)	1 (0.5)	.99
Co-antiplatelet therapy	14 (6.86)	21 (10.45)	.20
Smoking, n (%)	66 (32.4)	59 (29.4)	.51
Alcohol intake, n (%)	20 (9.8)	26 (12.9)	.32
Regional Distribution, n (%)			.50
Kunming	70 (34.3)	64 (31.8)	
Prefecture-level cities	128 (62.8)	134 (66.7)	
Outside Yunnan province	6 (2.9)	3 (1.5)	

[a]INR: international normalized ratio.

Primary Outcomes (TTR)

Table 2 shows that the mean TTR was significantly higher in the web-based group compared with the offline group (66.46% vs 46.65%, *P*<.001). In addition, the percentage of INR values within the therapeutic range was higher in the web-based group compared with the offline group (124/204, 60.78% vs 67/201, 33.33%; *P*<.001; Table 2).

Table . Percentage of time in therapeutic range per study group.

Variable	Web-based (n=204)	Offline (n=201)	*P* value
TTR[a] (%), mean (SD)	66.46 (25.22)	46.65 (39.13)	<.001
Control ratio, n (%)			<.001
TTR≥60%	124 (60.78)	67 (33.33)	
TTR<60%	80 (39.22)	134 (66.67)	

[a]TTR: time in therapeutic range.

Secondary Outcomes

Monitoring Frequency and INR Value Distribution

The Alfalfa app group had a higher monitoring frequency compared to the offline group (8.14 vs 4.47, *P*<.001) and a greater percentage of INR values within the therapeutic range (896/1660, 53.98% vs 346/899, 38.49%; *P*<.001). Participants managed through the Alfalfa app also had lower rates of subtherapeutic INR values (235/1660, 14.16% vs 152/899, 16.91%; *P*<.05), and extreme subtherapeutic INR values (273/1660, 16.45% vs 186/899, 20.69%; *P*<.05) compared with those managed by offline services. The incidence of supratherapeutic and extreme supratherapeutic INR values was similar between the 2 groups (Table 3).

XSL•FO
RenderX

Table . Monitoring frequency and distribution of international normalized ratio values.

Variable	Web-based (n=204/1660)	Offline (n=201/899)	*P* value
Monitoring frequency (times), mean (SD)	8.14 (4.75)	4.47 (3.84)	<.001
Extreme subtherapeutic, n (%)	273 (16.45)	186 (20.69)	.003
Subtherapeutic, n (%)	235 (14.16)	152 (16.91)	.02
Therapeutic, n (%)	896 (53.98）	346 (38.49)	<.001
Supratherapeutic, n (%)	251 (15.12)	204 (22.69)	.11
Extreme supratherapeutic, n (%)	5 (0.3)	10 (1.11)	.25

Clinical Events

The incidence of minor bleeding events vs was higher in the Alfalfa app group compared with the offline management group (12/204, 5.9% vs 3/201, 1.5%; *P*=.02). However, there were no significant differences between the groups in major bleeding events (3/204, 1.5% vs 1/201, 0.5% *P*=.63), thromboembolic events (1/204, 0.5% vs 0/201, 0%; *P*=.99), warfarin-related emergency department visits (1/204, 0.5% vs 2/201, 1.0%; *P*=.99) or warfarin-related hospital admissions (3/204, 1.5% vs 2/201, 1.0%; *P*=.99; Table 4).

Table . The incidences of clinical events.

Variable	Web-based (n=204)	Offline (n=201)	*P* value
Minor bleeding, n (%)	12 (5.9)	3 (1.5)	.02
Major bleeding, n (%)	3 (1.5)	1 (0.5)	.63
Thromboembolic, n (%)	1 (0.5)	0 (0)	.99
Warfarin-related emergency department visits, n (%)	1 (0.5)	2 (1)	.99
Warfarin-related hospital admissions, n (%)	3 (1.5)	2 (1)	.99

Costs

The *t* test revealed that the average cost per test (42.37 vs 78.3, *P*<.001) and the average time per test (47.42 vs 90.74, *P*<.001) were significantly lower for the web-based group compared with the offline group. While the total cost per capita was similar between the 2 groups (234.55 vs 220.98, *P*=.58), the cost-effectiveness ratio was lower for the web-based group (385.9 vs 662.9; Table 5).

Table . Economic differences.

Variable	Web-based (n=204)	Offline (n=201)	*P* value
Average cost per test, mean (SD)	42.37 (31.6)	78.3 (77.91)	<.001
Total cost per capita, mean (SD)	234.55 (270.25)	220.98 (224.74)	.58
Average time per test, mean (SD)	47.42 (41.01)	90.74 (74.57)	<.001
Total cost	¥47848.7 (US $7502.68)	¥44417.6 (US $6964.68)	—[a]
Effective, n (%)	124 (60.8)	67 (33.3)	—
Cost-effect ratio (C/E; as in Table 1)[b]	385.9	662.9	—

[a]Not applicable.

[b]C/E: total cost/effective value.

Acceptability and Demand

After 1 week of distributing and reminding participants to complete the survey links, a total of 86 voluntarily completed questionnaires were collected. Of these, 59.3% (51/86) had used the Alfalfa app for over a year, and 77.9% (67/86) preferred to stay in contact with their doctor through the app. The vast majority (83/86, 96.5%) found the app convenient, noting the ability to receive timely responses from doctors without needing to visit the hospital and to view all their report records within the app. Approximately 80% (69/86) of the patients expressed positive experiences with the doctor's responses (82/86, 95.3%), the completeness of records (75/86, 87.2%), and the free services (69/86, 80.2%).

XSL•FO
RenderX

All participants were satisfied with warfarin dose adjustment (100%, 86/86), timely responses from doctors (86/86, 100%), the report submission process (78/86, 90.7%), and the history records functions in the app (85/86, 98.8%). Nearly half of the respondents (42/86, 48.8%) indicated they would be willing to pay a fee ranging from ¥3 (US $0.4704) to ¥10 (US $1.568) for the app's services. A large majority (77/86, 89.5%) expressed a preference for services provided by experts with senior professional titles. A few participants (10/86, 11.6%) suggested improvements, such as simplifying the app's operation, adding a feature for video interactions with doctors (3/86, 3.49%), and establishing a more comprehensive medical history record function (1/86, 1.16%; Multimedia Appendix 2).

Discussion

Principal Findings

The main goal of this study was to evaluate the efficacy, safety, and economic impact of using the Alfalfa app, a smartphone app designed for dose consultation and adjustment in patients undergoing long-term anticoagulation therapy with warfarin. TTR was selected as the primary outcome, with a TTR of 60% or higher generally considered indicative of high-quality anticoagulant therapy [25]. In our study, TTR was significantly higher in the Alfalfa app group compared with the offline group (66.46% vs 46.65%, $P<.001$). While a study by Lee et al [26] reported contrary findings, our results align with the majority of telemedicine-based anticoagulation studies, which show improved TTR outcomes compared with traditional management methods [27-32]. Previous cohort studies in Eastern China also supported the efficacy of the Alfalfa App, reporting significantly higher TTRs in their intervention groups (61% and 79.35%, respectively) [15,33]. Our findings further confirmed the effectiveness of Alfalfa App in improving the quality of anticoagulation management.

Our research also demonstrated that participants in the web-based group had a significantly higher frequency of INR monitoring compared with those in the offline group (8.14 vs 4.47, $P<.001$). This indicates that the Alfalfa App facilitated better interaction between patients and health care providers, enabling participants to easily upload their INR records and receive timely feedback from clinical pharmacists. In addition, the app's supervision and tracking features likely contributed to the regular monitoring observed. These findings are consistent with 2 other retrospective cohort studies, which also reported higher rates of therapeutic INR values in the web-based group [15,33]. However, compared with the other 2 retrospective studies, there was no consistent pattern in the distribution of extreme subtherapeutic, subtherapeutic, supratherapeutic, and extreme supratherapeutic INR values between the groups. The result indicates that the web-based management model helps improve the percentage of INR values within the therapeutic range (indirectly equivalent to TTR), but when the INR values fall outside the therapeutic range, they do not necessarily tend to be too high or too low. This suggests that there may be no direct association between INR value distribution and anticoagulation management methods, indicating a need for further exploration with larger sample sizes.

Our research also revealed the higher incidence of minor bleeding events in the Alfalfa app group, consistent with the results reported in a similar cohort study [15]. According to a meta-analysis of 12 randomized controlled trials, telemedicine combined with portable coagulometers significantly improved TTR and reduced the incidence of thrombotic events in cardiovascular patients [34]. However, several studies reported no statistically significant difference ($P>.05$) in bleeding or thrombotic events between the eHealth tools group and the conventional management groups [28,30,35]. eHealth tools can be broadly categorized into 4 types, namely computer-based support systems, electronic health records, telemedicine platforms, and mobile apps. Among these, mobile apps offer the greatest convenience, allowing patients to communicate with doctors or pharmacists at any time [36]. This may explain the higher detection rate of minor bleeding events in the Alfalfa app group, as patients were more likely to report symptoms to clinical pharmacists through the app in time, especially during the period of COVID-19 pandemic, when patients were less willing to go out for medical treatment. This made minor bleeding in the web-based group of the study easier to detect. The pandemic also limited our ability to fully assess clinical event outcomes due to the reduced sample size and limited participant enrollment.

Our findings also suggest that the Alfalfa App may reduce the cost per visit for anticoagulation management and save participants time in return visits. The convenience of using the app led to higher adherence to follow-up visits in the web-based group, resulting in a greater number of return visits per capita during the follow-up period. Although we did not calculate the incremental cost-effectiveness ratio due to a lack of authoritative data, the cost-effectiveness value suggests a positive economic outlook for the Alfalfa App. It should be noted that patients in both the web-based and offline groups were exempt from registration fees. If web-based services were to be charged, combined with the increased number of follow-up visits, this could potentially offset the cost savings from not visiting the hospital, and the overall expenses might even be higher [37], and more research is needed to explore the full economic impact of the app.

Two studies revealed that most patients were satisfied with the use of eHealth tools, which also improved their quality of life [38,39]. Consistent with these findings, our survey results of the experience and satisfaction with pharmaceutical care and Alfalfa App also showed that the vast majority of patients were highly satisfied with the dose adjustment, timely responses, and free service provided. The timeliness and convenience of the Alfalfa App were particularly appreciated. However, a small number of participants complained the app difficult to use. During the follow-up, 5 participants discontinued to submit their results, resulting in withdrawals. This suggests that both the app and anticoagulant management services have areas for improvement. With the advancements of machine learning, the integration of deep learning, reinforcement learning, and ensemble learning in warfarin management after heart valve surgery is becoming more prevalent [40]. Internet+pharmacy care through app has the potential to enhance efficiency, reduce time and economic costs, and improve the quality of medical

XSL•FO
RenderX

care, especially by reducing the disparities among primary care providers.

Limitations

Although our study demonstrates promising results regarding the efficacy of internet+pharmacy care through Alfalfa Appin improving anticoagulant quality among patients following heart valve surgery, several limitations must be acknowledged. First, because the study's enrollment period overlapped with the COVID-19 lockdown, patients reduced their hospital visits. The significant decline in hospital patient flow from 2020 to 2022 confirms this trend. As a result, the study's sample size was limited, preventing us from conducting subgroup analyses. Second, this was a single-center study conducted in southwestern China, which may limit the generalizability of the findings to other populations. Further prospective, randomized, multicenter studies are needed to confirm these results.

We estimated a sample size of 446, assuming a low dropout rate of 10%. By the end of the study, 405 participants were included in the analysis, and the actual dropout rate was within expectations. With strong effect sizes, we maintained adequate statistical power. However, due to limitations with the Alfalfa App, we were unable to send private information to participants, which prevented us from distributing the satisfaction questionnaire individually. As a result, only 86 questionnaires were collected. In the future, we aim to address this limitation to enhance the completeness of research data.

Conclusions

In conclusion, the "Internet+Pharmacy Care" model through the Alfalfa App can improve the quality of warfarin anticoagulation in patients after heart valve surgery. Compared with traditional outpatient service, web-based management does not increase the risk of major bleeding, embolism, or hospitalization events, and is both economical and cost-effective. Looking forward, we anticipate that machine learning and artificial intelligence will bring new opportunities for innovation in anticoagulation management. Future studies with larger sample sizes, multicenter designs, and long-term follow-up are needed to further validate the role of these technologies.

Acknowledgments

This study was funded by the First-Class Discipline Team of Kunming Medical University (Research and Practice Team of Pharmaceutical Policy), 2024 (project number 2024XKTDPY21), Yunnan Provincial Clinical Medical Research Center (project number FZX2019-06-01,2022YFKY010), Research and Application of Epidemiology, Pathogenesis, Diagnosis and Treatment of Cardiovascular Diseases in High Altitude of Yunnan Province Project (number 202103AC100004) and the National Clinical Research Center for Cardiovascular Diseases-Yunnan Provincial Sub-Center.

Data Availability

The datasets used and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YQ designed the study and wrote the manuscript. YG assisted with the recruitment of subjects. JZ provided the permission to use the Alfalfa app. ZW gave suggestions for the cost-benefit analysis in the study. JL performed the statistical analysis. WC and BZ participated in the study and contributed to data analysis, drafting, data interpretation, and manuscript revision. YQ is the first author and corresponding author. YG, ZW, and JZ are joint corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The classification criteria for international normalized ratio.
[DOCX File, 13 KB - mhealth_v13i1e53586_app1.docx]

Multimedia Appendix 2
Alfalfa app user satisfaction survey and survey results of Alfalfa app usage.
[DOCX File, 25 KB - mhealth_v13i1e53586_app2.docx]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File, 14224 KB - mhealth_v13i1e53586_app3.pdf]

References

1. Santangelo G, Bursi F, Faggiano A, et al. The global burden of valvular heart disease: From clinical epidemiology to management. J Clin Med 2023 Mar 10;12(6):2178. [doi: 10.3390/jcm12062178] [Medline: 36983180]

XSL•FO
RenderX

2. Chinese Society of Extracorporeal Circulation. White book of Chinese cardiovascular surgery and extracorporeal circulation in 2019. Chinese Journal of Extracorporeal Circulation 2020 Aug 28;20(4):196-199. [doi: 10.13498/j.cnki.chin.j.ecc.2022.04.02]
3. Zhang J, Wu T, Chen W, Fu J, Xia X, Chen L. Effect of gene-based warfarin dosing on anticoagulation control and clinical events in a real-world setting. Front Pharmacol 2020 Jan 17;10. [doi: 10.3389/fphar.2019.01527]
4. Tsoumas I, Oz A, Lampropoulos K. Optimal anticoagulation on TAVI patients based on thrombotic and bleeding risk and the challenge beyond: A systematic review and meta-analysis. Curr Probl Cardiol 2023 Jun;48(6):101632. [doi: 10.1016/j.cpcardiol.2023.101632] [Medline: 36724817]
5. Oz A, Tsoumas I, Lampropoulos K, Xanthos T, Karpettas N, Papadopoulos D. Cardiac rehabilitation after TAVI -A systematic review and meta-analysis. Curr Probl Cardiol 2023 Mar;48(3):101531. [doi: 10.1016/j.cpcardiol.2022.101531] [Medline: 36493915]
6. Cromheecke ME, Levi M, Colly LP, et al. Oral anticoagulation self-management and management by a specialist anticoagulation clinic: a randomised cross-over comparison. Lancet 2000 Jul 8;356(9224):97-102. [doi: 10.1016/S0140-6736(00)02470-3] [Medline: 10963245]
7. Zhang J, Liu M, Chen Q, Wu J, Cao H. Outcomes of an online pharmacist-managed anticoagulation clinic for individuals on warfarin therapy living in rural communities. Thromb Res 2017 Sep;157:136-138. [doi: 10.1016/j.thromres.2017.07.016] [Medline: 28746903]
8. Zhu Z, Li Y, Meng X, et al. New warfarin anticoagulation management model after heart valve surgery: rationale and design of a prospective, multicentre, randomised trial to compare an internet-based warfarin anticoagulation management model with the traditional warfarin management model. BMJ Open 2019 Dec 5;9(12):e032949. [doi: 10.1136/bmjopen-2019-032949] [Medline: 31811010]
9. Xia X, Wu J, Zhang J. The effect of online versus hospital warfarin management on patient outcomes: a systematic review and meta-analysis. Int J Clin Pharm 2018 Dec;40(6):1420-1429. [doi: 10.1007/s11096-018-0731-8] [Medline: 30367374]
10. McCahon D, Baker JM, Murray ET, Fitzmaurice DA. Assessing the utility of an online registry for patients monitoring their own warfarin therapy. J Clin Pathol 2016 Apr;69(4):331-336. [doi: 10.1136/jclinpath-2015-203168] [Medline: 26519487]
11. Cao H, Wu J, Zhang J. Outcomes of warfarin therapy managed by pharmacists via hospital anticoagulation clinic versus online anticoagulation clinic. Int J Clin Pharm 2018 Oct;40(5):1072-1077. [doi: 10.1007/s11096-018-0674-0] [Medline: 29956134]
12. Jain R, Stone JA, Agarwal G, et al. Canadian Cardiovascular Harmonized National Guideline Endeavour (C-CHANGE) guideline for the prevention and management of cardiovascular disease in primary care: 2022 update. CMAJ 2022 Nov 7;194(43):E1460-E1480. [doi: 10.1503/cmaj.220138] [Medline: 36343954]
13. Holbrook A, Schulman S, Witt DM, et al. Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012 Feb;141(2 Suppl):e152S-e184S. [doi: 10.1378/chest.11-2295] [Medline: 22315259]
14. Xia X, Fu J, Wu T, Chen W, Jinhua Z. Comparison of the outcomes of warfarin therapy and economics by online and offline anticoagulation management models: protocol for a randomised controlled trial. BMJ Open 2019 Sep 5;9(9):e031730. [doi: 10.1136/bmjopen-2019-031730] [Medline: 31492796]
15. Cao H, Jiang S, Lv M, Wu T, Chen W, Zhang J. Effectiveness of the Alfalfa app in warfarin therapy management for patients undergoing venous thrombosis prevention and treatment: Cohort study. JMIR Mhealth Uhealth 2021 Mar 2;9(3):e23332. [doi: 10.2196/23332] [Medline: 33650976]
16. Briggs AL, Jackson TR, Bruce S, Shapiro NL. The development and performance validation of a tool to assess patient anticoagulation knowledge. Res Social Adm Pharm 2005 Mar;1(1):40-59. [doi: 10.1016/j.sapharm.2004.12.002] [Medline: 17138465]
17. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briët E. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost 1993 Mar 1;69(3):236-239. [doi: 10.1055/s-0038-1651587] [Medline: 8470047]
18. Macedo AF, Bell J, McCarron C, et al. Determinants of oral anticoagulation control in new warfarin patients: analysis using data from Clinical Practice Research Datalink. Thromb Res 2015 Aug;136(2):250-260. [doi: 10.1016/j.thromres.2015.06.007] [Medline: 26073321]
19. Chen L, Zhou YZ, Zhou XM, et al. Evaluation of the “safe multidisciplinary app-assisted remote patient-self-testing (SMART) model” for warfarin home management during the COVID-19 pandemic: study protocol of a multi-center randomized controlled trial. BMC Health Serv Res 2021 Aug 26;21(1):875. [doi: 10.1186/s12913-021-06882-7] [Medline: 34445995]
20. Hadlock GC, Burnett AE, Nutescu EA. Warfarin. In: GD B, MB S, editors. Anticoagulation Therapy: Springer; 2018. [doi: 10.1007/978-3-319-73709-6_2]
21. Prochaska JH, Göbel S, Keller K, et al. Quality of oral anticoagulation with phenprocoumon in regular medical care and its potential for improvement in a telemedicine-based coagulation service--results from the prospective, multi-center, observational cohort study thrombEVAL. BMC Med 2015 Jan 23;13(1):14. [doi: 10.1186/s12916-015-0268-9] [Medline: 25616558]

XSL•FO
RenderX

22. Schulman S, Kearon C, Subcommittee on Control of Anticoagulation of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost 2005 Apr;3(4):692-694. [doi: 10.1111/j.1538-7836.2005.01204.x] [Medline: 15842354]
23. Qian Y, Li C, Guo Y. Quality evaluation of anticoagulation therapy management by clinical pharmacists in patients taking warfarin. chinese journal of clinical pharmacy 2018;27(3):172-175.
24. He X, D, You J, Qian Y. Investigation and analysis of patients' satisfaction for clinical pharmacists in anticoagulation clinics. Evaluation and Analysis of Drug Use in Hospitals of China 2019;19(9). [doi: 10.14009/j.issn.1672-2124.2019.09.033]
25. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med 2007 Jun 19;146(12):857-867. [doi: 10.7326/0003-4819-146-12-200706190-00007] [Medline: 17577005]
26. Lee JC, Horner KE, Krummel ML, McDanel DL. Clinical and financial outcomes evaluation of multimodal pharmacist warfarin management of a statewide urban and rural population. J Pharm Pract 2018 Apr;31(2):150-156. [doi: 10.1177/0897190017703506] [Medline: 28441899]
27. Manotti C, Moia M, Palareti G, Pengo V, Ria L, Dettori AG. Effect of computer-aided management on the quality of treatment in anticoagulated patients: a prospective, randomized, multicenter trial of APROAT (Automated PRogram for Oral Anticoagulant Treatment). Haematologica 2001 Oct;86(10):1060-1070. [Medline: 11602412]
28. Thompson JL, Burkhart HM, Daly RC, et al. Anticoagulation early after mechanical valve replacement: improved management with patient self-testing. J Thorac Cardiovasc Surg 2013 Sep;146(3):599-604. [doi: 10.1016/j.jtcvs.2012.03.088] [Medline: 22921821]
29. Zhu Z, Li C, Shen J, et al. New internet-based warfarin anticoagulation management approach after mechanical heart valve replacement: Prospective, multicenter, randomized controlled trial. J Med Internet Res 2021 Aug 13;23(8):e29529. [doi: 10.2196/29529] [Medline: 34397393]
30. Christensen H, Lauterlein JJ, Sørensen PD, Petersen ERB, Madsen JS, Brandslund I. Home management of oral anticoagulation via telemedicine versus conventional hospital-based treatment. Telemed J E Health 2011 Apr;17(3):169-176. [doi: 10.1089/tmj.2010.0128] [Medline: 21254841]
31. Fitzmaurice DA, Hobbs FD, Murray ET, Bradley CP, Holder R. Evaluation of computerized decision support for oral anticoagulation management based in primary care. Br J Gen Pract 1996 Sep;46(410):533-535. [Medline: 8917873]
32. Fitzmaurice DA, Hobbs FDR, Murray ET, Holder RL, Allan TF, Rose PE. Oral anticoagulation management in primary care with the use of computerized decision support and near-patient testing: a randomized, controlled trial. Arch Intern Med 2000;160(15):2343-2348. [doi: 10.1001/archinte.160.15.2343] [Medline: 10927732]
33. Jiang S, Lv M, Zeng Z, et al. Efficacy and safety of app-based remote warfarin management during COVID-19-related lockdown: a retrospective cohort study. J Thromb Thrombolysis 2022 Jul;54(1):20-28. [doi: 10.1007/s11239-021-02630-0] [Medline: 35092566]
34. Huang Y, Xie Y, Huang L, Han Z. The value of anticoagulation management combining telemedicine and self-testing in cardiovascular diseases: A meta-analysis of randomized controlled trials. Ther Clin Risk Manag 2023;19:279-290. [doi: 10.2147/TCRM.S395578] [Medline: 36941980]
35. Poller L, Keown M, Ibrahim S, et al. An international multicenter randomized study of computer-assisted oral anticoagulant dosage vs. medical staff dosage. J Thromb Haemost 2008 Jun;6(6):935-943. [doi: 10.1111/j.1538-7836.2008.02959.x] [Medline: 18489430]
36. Wu Y, Wang X, Zhou M, Huang Z, Liu L, Cong L. Application of eHealth tools in anticoagulation management after cardiac valve replacement: Scoping review coupled with bibliometric analysis. JMIR Mhealth Uhealth 2024 Jan 5;12:e48716. [doi: 10.2196/48716] [Medline: 38180783]
37. Stephan LS, Dytz Almeida E, Guimaraes RB, et al. Processes and recommendations for creating mHealth apps for low-income populations. JMIR Mhealth Uhealth 2017 Apr 3;5(4):e41. [doi: 10.2196/mhealth.6510] [Medline: 28373155]
38. Soliman Hamad MA, van Eekelen E, van Agt T, van Straten AHM. Self-management program improves anticoagulation control and quality of life: a prospective randomized study. Eur J Cardiothorac Surg 2009 Feb;35(2):265-269. [doi: 10.1016/j.ejcts.2008.10.020] [Medline: 19041254]
39. Testa S, Alatri A, Paoletti O, et al. Reorganisation of an anticoagulation clinic using a telemedicine system: description of the model and preliminary results. Intern Emerg Med 2006;1(1):24-29. [doi: 10.1007/BF02934716] [Medline: 16941809]
40. Zeng J, Shao J, Lin S, et al. Optimizing the dynamic treatment regime of in-hospital warfarin anticoagulation in patients after surgical valve replacement using reinforcement learning. J Am Med Inform Assoc 2022 Sep 12;29(10):1722-1732. [doi: 10.1093/jamia/ocac088] [Medline: 35864720]

Abbreviations

INR: international normalized ratio
ISTH: International Society on Thrombosis and Haemostasis
TTR: time in therapy range

XSL•FO
RenderX

Edited by L Buis, TDA Cardoso; submitted 11.10.23; peer-reviewed by F Rahmillah, S Arsić; revised version received 24.11.24; accepted 25.03.25; published 20.05.25.

Please cite as:
Qian Y, Chen W, Zhou B, Li J, Guo Y, Weng Z, Zhang J
Efficacy, Safety, and Cost-Effectiveness of "Internet + Pharmacy Care" Via the Alfalfa App in Warfarin Therapy Management After Cardiac Valve Replacement: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e53586
URL: https://mhealth.jmir.org/2025/1/e53586
doi:10.2196/53586

© Yiyi Qian, Weizhao Chen, Bin Zhou, Jiangya Li, Yuanyuan Guo, Zhiying Weng, Jinhua Zhang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Preferences for Mobile Apps That Aim to Modify Alcohol Use: Thematic Content Analysis of User Reviews

Megan Kirouac, PhD; Christina Gillezeau, MPH, MS

Center on Alcohol, Substance use, And Addictions, The University of New Mexico, Albuquerque, NM, United States

Corresponding Author:
Megan Kirouac, PhD
Center on Alcohol, Substance use, And Addictions, The University of New Mexico, Albuquerque, NM, United States

Abstract

Background: Nearly one-third of adults in the United States will meet criteria for alcohol use disorder in their lifetime, yet fewer than 10% of individuals who meet for alcohol use disorder criteria will receive treatment for it. Mobile health (mHealth) applications (apps) have been suggested as a potential mechanism for closing this treatment gap, yet there is a wide variety of quality and integrity within these apps, leading to potential harms to users.

Objectives: The aim of this paper is to systematically record and qualitatively examine user reviews or mHealth apps to identify features in the existing apps that may impact usefulness and adoption of them.

Methods: The researchers used Apple App and Google Play stores to identify mHealth apps that were focused on modifying alcohol use and treating common comorbidities. Apps that were free without in-app purchases and provided multiple features for users were included. User reviews from the apps were downloaded and coded using content analysis.

Results: A total of 425 unique apps were found in our search. Of these, the majority of apps (n=301) were excluded from the present analyses for not focusing on reducing alcohol-related concerns (eg, many apps were for purchasing alcohol). Eight apps were identified and had user reviews downloaded. The apps examined in this study were VetChange, SMART, DrinkCoach, SayingWhen, AlcoStat, Celebrate Recovery, TryDry, and Construction Industry Helpline. A total of 370 reviews were downloaded and 1353 phrases were coded from those reviews into a total of 11 codes. The 5 most common themes identified were praise (498 counts coded; 36.831%), tools (150 counts coded; 11.062%), suggestions for improvement (118 counts coded; 8.756%), criticism (105 counts coded; 7.768%), and tracking (104 counts coded; 7.724%).

Conclusions: The current findings suggest that alcohol mobile app users broadly found the apps helpful in reducing their drinking or meeting their drinking goals. Users were able to identify features that they liked or found helpful in the apps, as well as provide concrete feedback about features that they would like included or improved. Specifically, flexible and expansive tracking features and comprehensive whole health tools were cited as valuable and desired. App developers and those looking to expand access to and uptake of alcohol reduction apps may find these user reviews helpful in guiding their app development.

(JMIR Mhealth Uhealth 2025;13:e63148) doi:10.2196/63148

KEYWORDS

alcohol mobile app; mHealth; alcohol use disorder; user-centered design; alcohol; user-centered; user; reviews; usefulness; mobile health app; content analysis; drinking; health tool

Introduction

It has recently been estimated that a third of US adults engage in risky alcohol use and just under a third of US adults (29%) will meet criteria for alcohol use disorder (AUD) in their lifetime [1]. Moreover, only 7.9% (2.3 million people aged 12 years and older) of individuals who met criteria for AUD in the United States received any kind of formal treatment [2]. Common reasons for not engaging in treatment include the fear of stigma or privacy concerns and lack of access to in-person resources [3]. Further complicating treatment utilization is the fact that individuals with AUD have a higher prevalence of other co-occurring mental health conditions than individuals with nonsubstance use disorder diagnoses [4]. Rates of co-occurring anxiety, depression, posttraumatic stress disorder (PTSD), or attention-deficit/hyperactivity disorder (ADHD) have been found to be particularly high among individuals diagnosed with AUD [5]. For these individuals with co-occurring diagnoses, formal treatment is often restricted to providers qualified to treat those co-occurring diagnoses simultaneously or to health care systems that are capable of providing coordinated care [4]. Unsurprisingly, having co-occurring AUD and other mental health diagnoses is often associated with poorer treatment outcomes [6].

Mobile apps and other mobile health (mHealth) interventions have been proposed as one potential solution to help overcome many of the current barriers for receiving help with one's

XSL•FO
RenderX

relationship to alcohol [7]. First, mHealth interventions are highly accessible, given that at least 85% of all US adults report owning a smartphone, including members of historically underserved communities [8]. Second, mHealth interventions are also ready 24/7 and available for brief, impromptu interactions compared with scheduled appointments during business hours associated with traditional in-person treatment. Accordingly, mHealth interventions hold promise to overcome access barriers traditional treatments face: geographic and provider restrictions [7].

While some critics argue that mHealth interventions are limited in that clients will not feel connected to a mobile app the same way they would to an in-person provider, some research has shown that clients can develop a therapeutic alliance with their mHealth tools [9,10]. Furthermore, mHealth interventions have been touted as particularly well suited to reduce stigma-related barriers compared with in-person treatments where individuals have to attend treatment visits or clinics where they must interact with others who may judge or stigmatize them [7]. Finally, mHealth interventions allow clients to have levels of autonomy, flexibility, personalization, and depth that is often unparalleled in traditional individual or group treatments [7]. For example, mHealth interventions are often designed with tools and resources sections of the intervention that clients may navigate at their own pace and based on their own individual interests and needs. Instead of having to navigate complex health care systems or potentially inaccurate or unreliable web-based resources, mHealth interventions can be programmed to contain reliable and accurate information that may be relevant for some clients (eg, psychoeducation about PTSD symptoms and PTSD symptom monitoring contained on the National Center for PTSD's VetChange mobile app for alcohol use).

Several mobile apps have been developed to target alcohol use concerns, yet the quality and integrity of those apps are highly varied. For instance, many alcohol-focused apps are created commercially with limited oversight from government and scientific communities. A recent review highlighted the myriad potential harms of unregulated health-focused apps, including apps that are falsely advertised as "safety apps" but instead "promote binge drinking" [11]. Even apps that are not designed for-profit or that do not contain overtly harmful content are potentially harmful due to the lack of oversight and regulation since they may not be evidence-based and may, therefore, fail to work as desired. Not only do these apps potentially harm users by taking their time to no positive outcome but may also cause secondary harms by making app users feel that there is no help possible for them. As pointed out by Businelle and colleagues [7], "nearly all of the mHealth interventions that are available for download today have not been evaluated for efficacy or effectiveness. Thus, individuals are downloading interventions that may not be effective for addressing their problems, which could delay or discourage the use of future well-studied and effective interventions" [7]. Moreover, recent studies suggest that the utilization of existing mobile apps remains low; yet when apps are used, there is a positive association between user engagement and intervention effectiveness [7]. Accordingly, there is a critical need to evaluate which mobile apps may be evidence-based and which apps are well used in order to inform the public about which apps may be most helpful. One such strategy to accomplish this goal is to examine user reviews of existing apps since user ratings in digital marketplaces have been found to influence app quality and, therefore, impact the usage of that app [7]. Information from these publicly available app user reviews may be used to inform app refinement without associated costs of User Experience (UX) or other User-Centered Design research projects. Therefore, the aim of this paper was to systematically review freely available alcohol mobile apps, identify potentially evidence-based options, and conduct qualitative analyses of user reviews to identify features of existing apps that may impact use and helpfulness of mHealth interventions for alcohol use.

Methods

Ethical Considerations

All data used in this study were publicly available as posted by users themselves on Apple App and Google Play app stores. Accordingly, it is within the institutional review board's policies for the authors' home institution to consider this research exempt from institutional review board review. Moreover, all data presented in this study were analyzed without usernames and findings herein are presented in aggregate. Where specific quotes are used as examples of thematic codes, no identifying information was included.

Data Selection Procedures

Both authors for this study have prior qualitative data analytic experience and met at the beginning of the study to discuss procedures for this study aims in February 2024. The first step in achieving the proposed study aims was to identify apps that were potentially evidence-based. The authors searched in the Google Play and Apple App Stores for mHealth apps focused on alcohol. Given the high degree of comorbidity of AUD with other mental health diagnoses, and the stigma associated with AUD specifically, app search terms were extended beyond alcohol alone. Five search terms were used in Google Play and the Apple App Stores: "alcohol," "alcohol and depression," "alcohol and anxiety," "alcohol and PTSD," and "alcohol and ADHD." Depression, anxiety, PTSD, and ADHD were selected since those are the mental health diagnoses with the highest prevalence of comorbidity with alcohol [5]. The authors independently reviewed their results to exclude duplicates and restricted the results to apps that were available for free, without in-app purchases, to prevent potentially harmful, for-profit corporate apps from being included in final analyses. Next, authors proposed to exclude apps that were solely for single purposes, such as blood alcohol concentration calculating apps, and any disagreements were resolved through mutual discussion and reaching consensus. Finally, since the current authors are fluent only in English, all non-English apps were excluded from this study. Of the remaining apps, reviews were extracted using the Export Comments website. Since several apps contained only few reviews and other apps contained hundreds of reviews, the authors restricted the number of exported reviews to only the most recent 50 reviews in Google Play and 50 reviews in the Apple App Store. Although this restricted the available data, it allowed results to be more balanced toward the apps generally,

XSL•FO
RenderX

rather than app-specific for the most widely reviewed apps. Moreover, reviews limited to the most recent reviews ensured that outdated reviews were not considered in this study, which is particularly important, given that the widely reviewed apps may have been updated such that previous reviews are no longer relevant. App store searches and user review extractions were completed in March 2024.

Qualitative Analyses

Exported user reviews were downloaded into a spreadsheet program in which coding was conducted. Conventional content analysis was used to examine thematic content to user reviews [12,13], which is a method of qualitative data analysis used to identify themes in text data through systematic classification. In this approach, the researchers examine the data with no preconceived, theory-based hypotheses about what kinds of codes or themes will be found and instead use the data itself to drive the codes identified.

Initial themes were identified by both authors independently generating codes for a third of the reviews from each app identified for data analysis. These codes were then compared by the authors, discussed, and a reviewed codebook was agreed upon. Based on qualitative methodology recommendations and this study aims, the following codebook rules were agreed upon by the authors: (1) assign only 1 code per utterance, with "utterances" distinguished by parts of speech (eg, "and") or punctuation; (2) emojis are not to be coded, given the propensity for emojis to be transcribed as nonsubstantive text in the data extraction (ie, smiley faces were distorted into uninterpretable characters); (3) specific features identified within coded themes should be noted (eg, when concrete tools or specific tracking features are mentioned); and (4) any time a user review is explicitly not referring to the mobile app itself should be coded as "null" (eg, reference to in-person meetings being helpful vs the app itself). These codebook rules were established to ensure clarity, prevent overlap, and prevent artificially inflating interrater reliability [14,15].

The authors then completed coding using the revised codebook for a different third of the reviews from each app. The selection was made to code a third of reviews at each of these 2 stages based on the authors' prior experience with qualitative data analysis and expectation for it to take approximately 3 iterations to finalize a codebook. However, the second iteration of the codebook was agreed to be sufficient after the second third of data were independently coded and then discussed by the authors. This final codebook was used to code all of the app user review data by each author until 80% interrater consistency was achieved, which is considered the standard consistency for these types of analyses [16]. Specifically, both authors used an iterative process of coding to create the final codebook and increase rater agreement on the application of each code to each utterance in user reviews. Once the codebook was finalized, raters compared their codes and discrepant codes were flagged and rereviewed by both raters who were then allowed to reconsider their coding upon closer review. Any final discrepancies were discussed by both raters until agreement was achieved and overall interrater consistently met the predetermined 80% guidance from the empirical literature.

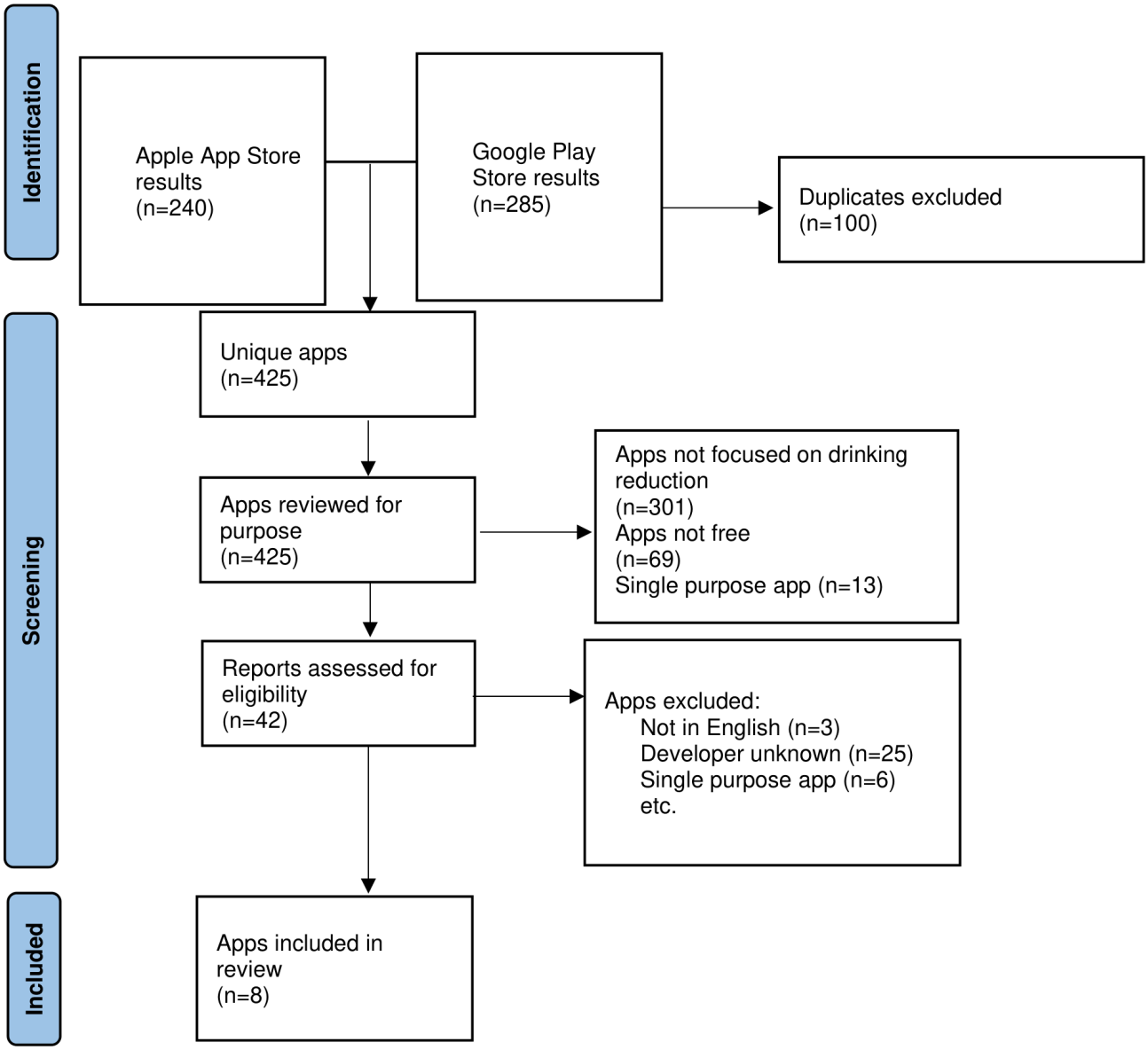
Results

Apps Analyzed

Data extraction was completed in early 2024; see Figure 1 for a flow diagram of app store search results. More than 200 apps were initially returned in search results within Google Play (n=285) and Apple App Stores for the specified search terms (n=240; n=425 unique results between both stores). The majority of these apps were excluded because they were for-profit (ie, had in-app purchases), were not mHealth apps (eg, alcohol-purchasing apps), or were explicitly harmful (ie, alcohol-drinking game apps). The remaining apps were excluded because they were of single purpose (eg, blood alcohol concentration calculators), or were not in English. The remaining results were researched by the authors to determine whether they were made by reputable and legitimate entities (eg, governmental agencies such as the National Center for PTSD) as a proxy for potential evidence basis of the app content. Findings were discussed until a consensus was established for which apps to include. A total of 8 apps were retained for data analysis: VetChange, SMART, DrinkCoach, SayingWhen, AlcoStat, Celebrate Recovery, TryDry, and Construction Industry Helpline. See Table 1 for a list of each of these apps and relevant descriptors. Each app had varying levels of user reviews provided, with Celebrate Recovery and TryDry being the only 2 apps that had more than 100 user reviews on each system platform. Accordingly, these 2 apps' user review data extraction was limited to only the most recent 50 reviews for Google and Apple (total of 100 reviews for each app). The other apps contained the following number of reviews: VetChange (n=14), SMART (n=35), DrinkCoach (n=51), Saying When (n=56), AlcoStat (n=7), and Construction Industry Helpline (n=7). A total of 370 reviews were included in this study (N=370).

XSL•FO
RenderX

Figure 1. Flow diagram of app store search results and inclusion in the present analyses.



XSL•FO
RenderX

Table . Name and description of app store search results included in the present analyses.

App name	Number of user reviews	Platform	Developer	Description
AlcoStat	7	iOS	Andrey Shestakov	Categorized as a "lifestyle" app on Apple App Store. Features include alcohol calendar to track and personal statistics of alcohol use tracked in the app such as number of ounces and cost of alcohol consumed.
Celebrate Recovery	>100	Android; iOS	Bobbi McWilliams	Categorized as a "lifestyle" app on Google Play and Apple App Stores. Features include Celebrate Recovery meeting locator, Celebrate Recovery tools, and other resources.
Construction Industry Helpline	7	Android; iOS	Construction Industry Solutions Limited	Categorized as a "health & fitness" app on Google Play Store and as a "lifestyle" app on Apple App Store. Features include coping tools, mood-tracking features, resource locator, and psychoeducation.
DrinkCoach	51	Android; iOS	Humankind	Categorized as a "health & fitness" app on Google Play and Apple App Stores. Features include alcohol tracking, calorie and cost tracking, goal-setting features, and coping skill videos.
SayingWhen	56	Android; iOS	Centre for Addiction and Mental Health (CAMH), Education	Categorized as a "health & fitness" app on Google Play and a "medical" app on Apple App Store. Features include alcohol use tracking, urge tracking, coping skills, and goal setting.
SMART Recovery	35	Android; iOS	SMART Recovery USA	Categorized as a "lifestyle" app on Google Play and as a "health & fitness" app on the Apple App Store. Features include SMART Recovery meeting locator, coping tools, and psychoeducation.
TryDry	>100	Android; iOS	Alcohol Change UK	Categorized as a "health & fitness" app on Google Play and Apple App Stores. Features include alcohol use tracking, coping skills, goal setting, and "badges" for reinforcement of achievements logged in the app.
VetChange	14	Android; iOS	US Department of Veterans Affairs	Categorized as a "health & fitness" app on Google Play and Apple App Stores. Features include alcohol tracking, coping skills, mental health tools, and goal setting.

XSL•FO
RenderX

Themes From User Reviews

Reviews varied in length and each new idea or utterance was given a unique code for a total of 1353 coded phrases. Initially, 14 themes were identified but codes that contained <1% of the total data were collapsed into other, broader codes. Specifically, the theme of technological distrust was collapsed into "criticism" and the themes of "credibility" and "nonjudgmental" were collapsed into "praise." The final codebook contained 11 codes, listed in Table 2.

Table . Rank-ordered list of thematic codes and their frequencies.

Rank	Theme	Examples	Frequency	%
1	Praise	"Absolutely fantastic," "Definitely a useful little app," "This app helped me just raise my awareness of how much I was drinking"	498	36.831
2	Tools	"Useful links to pdfs, "lots of good tools"	150	11.062
3	Suggestion	"Being able to create my own drink category, for example Hard Seltzer," "it doesn't allow me to manually input drink log data for any date prior to the current month. Please [sic] fix this!"	118	8.756
4	Criticism	"It's not very flashy or feature rich," "Terrible interface."	105	7.768
5	Tracking	"Keeps you accountable by having you log how many drinks you have to keep up with daily," "it helps keep track of dry days, mood, energy levels and alcohol cravings"	104	7.724
6	Personal Story	"HI!! I'M GRATEFUL BELIEVER IN JESUS WHO use to struggle with alcohol & Denial," "I personally didn't get any benefit from trying to put my recovery in the hands of a 'higher power', but I don't mind that some people do."	86	6.366
7	Technical Issue	"It also erased data from two days ago," "but it it [sic] quit working after the ios update that just came out"	79	5.808
8	Usability	"Easy to use, beautifully presented and functional app," "really easy to use"	77	5.697
9	Null	"U r funny," "I Cannot Cannot Get Pads Your Bernhardt"	66	4.885
10	Advice	"Get Help!," "If you want to recover you've got to come out of isolation."	38	2.812
11	Gamification	"There are badges to claim," "I love getting entering [sic] 'stayed dry' in the calendar and seeing the confetti. The reward really works!"	31	2.291

XSL•FO
RenderX

Top 5 Themes

The top 5 most commonly occurring themes were praise, tools, suggestions for improvement, criticism, and tracking (see Table 2 for a rank-ordered list of all the themes coded and their frequencies). Within the theme of "praise," reviews frequently contained general praise for the app such as "absolutely fantastic" but reviews within this category also frequently mentioned specific, positive outcomes they attributed to the app. For instance, many reviews contained concrete drinking outcome data as well as other concrete outcomes they experienced: "52 days not drinking today and soon 4 month not smoking too!!!!!," "No doubt sleeping better," and "Has massively helped to cut down!"

Closely related to "praise" was the second most frequently coded theme of "tools," which was used to identify reviews that spoke specifically to functional tools programmed into the app that were not just tracking tools (see "tracking" code descriptions in the final paragraph of this section). To better inform future app development and refinement, any time "tools" were coded in the present analyses, the authors noted which specific tools were referenced as helpful by the app users. Future app programming efforts are recommended to consider these tools in their apps. Example tools that were commonly mentioned as being enjoyed and helpful were time- and location-specific reminders and specific coping tools such as guided mindfulness exercises and ways of coping with cravings, goal-setting functions, community resources, and psychoeducational resources. Examples of reviews that contained "tools" content are "useful links to pdfs and videos which are supplementary," "There's access to podcasts and videos, local meetings, motivational memes," and "gives you info on where you are compared to city wide average drinkers."

Similar to the "tools" code, the code for "suggestions for improvement" was used to indicate when users desired tools or features of the app that did not exist in the current version of the app being used. The authors also noted which specific suggestions were requested to best inform future app development efforts. The most common suggestion for improvement in apps was related to personalization and flexibility in the tracking components of the app. For instance, several instances of "suggestion for improvement" were coded for TryDry in relation to US users requesting US-specific quantities and standard drinks (ounces, 14-g standard drinks) since the app was developed in the United Kingdom (milliliters, 8-g standard drink units). One user of the TryDry app said in their review: "The one thing that could be better is using oz for the US but I understand it is for UK." Other app reviews for TryDry and the other apps analyzed in this study called for personalization and flexibility around entering their drink information and having the app calculate standard drinks for them. Examples of this suggestion include "Have a calculator to convert units to abv% ," and "Really don't understand why it can't show half units on the calendar." Other suggestions for improvement included increased tracking features such as the ability to add notes to document particular circumstances that may explain their tracked event (eg, specific events that related to increased drinking and different moods), the abilities to customize the display settings on the tracking calendar (eg, start the week on a Monday vs a Sunday), and the option for daily reminders or notifications to ensure that users accurately tracked their drinking (eg, several apps automatically assumed that a day without alcohol consumption entered was an alcohol-free day, and several users mentioned that this led to inaccurate tracking during times when they forgot to enter drinks consumed, so a daily notification to confirm lack of alcohol consumption would be helpful instead).

It is intuitive that one may take the time to write reviews only for apps about which they have strong opinions; therefore, it is unsurprising that "criticism" was one of the top 5 most frequently coded themes. In contrast to "suggestions for improvement," reviews that contained utterances coded as "criticism" did not provide specific, constructive feedback. As such, the authors did not note specific subthemes within this category and the majority of reviews coded with "criticism" did not yield particularly helpful feedback for future app development or improvement. Examples of reviews that contained "criticism" included "Having 1 drink in the past 30 days means I am drinking more than "28%" what bunk statistics is this based on. It acts like I'm abusing substances because I had one drink," "I really wish I could rate this better, but it just doesn't work at all," and "Very frustrating."

Nearly equal to the frequency with which "criticism" was coded was the code for "tracking" in the present analyses. While potentially considered a "tool," the mention of "tracking" specifically was so common that the authors agreed that it warranted its own code. Examples of reviews containing "tracking" codes were "keeps you accountable by having you log how many drinks you have to keep up with daily; and weekly goals," "Good tool to track drinks per day," and "has helped me a lot to track my drinking habits." However, tracking capabilities across the reviewed apps were not limited solely to tracking drinking behavior, and several reviews specifically mentioned the helpfulness of tracking more than drinking behavior alone. The most commonly mentioned tracking features that were mentioned as helpful to track included cost, calories, cravings, and mood. That "tracking" was among the top 5 most frequently mentioned codes beyond "tools" more broadly highlights just how valuable and well-liked tracking features are in alcohol apps.

Additional Themes

The remaining 6 codes were observed in moderate and low frequency, compared with the codes of the top 5. "Personal story," "technical issues," "useability," and "null" codes were found in moderate frequency, comprising around 5% of the total codes each. The "personal story" and "null" codes were used when the review contained information that was not clearly related to the app. For instance, many user reviews contained personal anecdotes about their lives that may have led them to seek help through the mobile app but did not indicate any information for others about the app per se (eg, "Other 12 step programs weren't doing it for me."). The "null" code was used for reviews that were non-English or contained nonsensical text (eg, "I Cannot Cannot Get Pads Your Bernhardt") or where the text clearly was reviewing something outside of the app itself such as reviews of in-person meetings associated with but not

XSL•FO
RenderX

part of the app directly (ie, SMART Recovery or Celebrate Recovery in-person meetings). Complaints about particular "technical issues" or glitches experienced in the apps were also found at moderate frequency, likely highlighting a need for apps to have tech support options available. Conversely, "useability" was coded when reviews mentioned positive qualities about the user experience from an app interface perspective.

At lower but still prominent frequencies were the codes for "advice" and "gamification." While less useful feedback for mobile app development and refinement, many reviews contained advice to others that may or may not have been explicitly about the app itself. The "advice" code was found for most but not all of the apps reviewed and so may speak to population-level differences (eg, veterans advocating for other veterans and in-person meeting-focused apps advocating for community). For TryDry and DrinkCoach, several reviews mentioned the gamification that has been programmed to reinforce abstinence days such as confetti appearing on the screen to celebrate abstinent day entry. Other app reviewers also highlighted that color coding of drinking data tracking that aligns with their drinking goals was helpful. The fact that only 2 of the 8 total apps evaluated in this study contained mention of gamification and yet still was coded 31 times speaks to the high desirability of gamified features for mobile apps and suggests that other apps should program gamified functionality.

Discussion

Principal Results

Central to the primary aims of the study, qualitative analyses yielded significant insights into features of alcohol-focused mobile apps that may inform app development and enhancement to make apps more accessible and helpful. Broadly, the central research question of this project was to identify overarching themes across alcohol app user reviews. We found common topics of reviews shown in Table 2. However, while conducting formal qualitative analyses of these overarching themes, both individual coders also took note of any mention of specific features that were praised, suggested, or criticized that may provide deeper insights for future app developers. Given these were subthemes and counts were inherently very small compared with the overall themes identified in Table 2, formal qualitative analyses were not completed to obtain counts of each specific feature highlighted by reviewers. Nevertheless, these subthemes may be particularly helpful for guiding future app development and enhancement, given their specificity. Features that were commented in user reviews as helpful or desired include tools to help access resources (eg, in-person meeting locators and psychoeducational materials) or coping skills, customizability to add personalized notes, and tracking functionality. While several user reviews commented on the helpfulness and importance of tracking alcohol use itself, being able to calculate standard drinks flexibly and track nondrinking data such as mood, craving, costs, and calories was prominent across mobile app reviews. The presence of reminder capabilities was also frequently commented on as a helpful tool to enhance the benefits of tracking features of apps. Moreover, including gamification (eg, color changes of tracking data when goals are met and rewarding displays such as confetti for achieving goals) explicitly into app programming was well received by app users (gamification was coded 31 times despite the majority of the reviewed apps lacking gamification features). Of note, only 2 of the 8 apps included in this study contained user reviews of "gamification" codes, yet "gamification" alone consisted of 2.29% of all coded responses. If gamification was programmed into more of the apps, one would expect that the proportion of reviews mentioning "gamification" would have increased significantly.

Additional Insights on mHealth Mobile Apps

A number of mHealth mobile apps are available for individuals interested in changing their relationship with alcohol. On one hand, the ubiquity of mobile apps provides individuals with many options to access resources that may not otherwise be available to them. On the other hand, without research and regulations, serious negative consequences can occur if people find the apps unhelpful or harmful [11]. In this study, we reviewed freely available mobile apps that may help individuals reduce their alcohol use and alcohol-related harms. We found hundreds of results, yet the vast majority of these mobile apps were available for-profit, for entertainment and potential harm (eg, drinking game apps), or for single purposes that restrict the potential usefulness of an app (eg, blood alcohol concentration calculators). This information alone highlights one current problem for the state of app development.

In addition to the current potential harms for app stores providing such limited mHealth interventions as opposed to predatory products, another important finding from this study is the inherent limitations in app store search features. For instance, Apple App yielded drastically different search results compared with Google Play. Moreover, apps known to exist failed to be returned in search results (eg, DrinkLess and DrinksRation from the United Kingdom, and Less, which is currently available on iPhone in the United States). Given the present methodology for searching not only for "alcohol" but also for other highly related terms (ie, "alcohol and anxiety"), this finding is concerning. One of the primary advantages of mobile apps is their accessibility; however, if the majority of available apps are for-profit or not viewable with very basic search terms, many individuals who may benefit from mHealth interventions may fail to access them.

One anticipated theme for this study was that privacy concerns may play a large role in user reviews, given concerns regarding privacy and mHealth in the academic literature [17]. However, this concern did not feature prominently in user reviews, with only 1 review mentioning concerns regarding location tracking in the app. This suggests that these users did not find privacy concerns to be a barrier to mHealth app adoption. However, it also highlights that mHealth app users may not be attending to their data privacy, and thus it is important for mHealth apps to provide transparent data usage policies and follow best practices regarding maintaining mHealth data [18]. Other primary takeaways from the user reviews include areas of criticism that highlight the need for improvements in existing apps (criticism was coded 105 times). Specifically, user interface, the ability for the app to perform multiple tasks (ie, not simply a meeting

XSL•FO
RenderX

locator and readings interface), availability of technical support to report glitches with the app, and customizability were highly sought (eg, ability to change volume measurements from milliliters to ounces). Future app developers and organizations with active app portfolios are recommended to use these findings to develop more well-received apps.

Limitations

While many of the present findings have significant and novel implications, the study is not without limitations. One limitation is that not all known alcohol-related apps were included in the present analyses. However, this is useful information since it highlights the inadequacies of current search engines in mobile app stores and shows a current gap that needs to be addressed by app developers and app stores in order for their products to be useful to real-world users. A second limitation is that not all existing app reviews were examined in this study. Specifically, TryDry and Celebrate Recovery had several hundred of reviews across Google and Apple but the present analyses were restricted to only the most recent 50 reviews from each store (100 reviews total for each app). Accordingly, additional nuances may have been missed that would have resulted in different codes in our qualitative analyses. However, had all reviews been included, the current findings would not generalize to other alcohol-related apps since the codes would be overly specific to those 2 apps alone. The fact that "gamification" occurred frequently enough to warrant a code and, yet, it was only a feature in 2 of the 8 apps highlights the appeal of gamified features within apps to users. Furthermore, many "technical issues" that were found may have been specific to unique versions of an app, so analyzing all existing reviews may have resulted in skewed findings about glitches or other app features that no longer exist in the current apps. Another limitation of this study was the restriction to English language apps. In our search for freely available mobile apps, we did find non-English apps such as "Oz Ensemble" developed by Fabrique numérique des ministères sociaux. The inclusion of non-English apps may have resulted in different findings, and future studies should be done to examine and compare non-English apps with the present results. Another point of potential limitation in our findings was the reliance on limited tools to examine the mobile app stores more systematically such as machine learning or other techniques that require different expertise or software support. Given the present research was completed without independent funding, our methodology was restricted to approaches that were freely available and fell within the scope of our existing expertise; it is possible that purchasable or other tools that conduct systematic reviews of app stores may have yielded different findings. A final limitation of this study is the inherent limitations of using publicly available user review data as opposed to other methodology for app development such as UX studies [19]. Yet, the advantages of UX studies may also be considered disadvantages that the present methodology addresses: UX studies are costly and formal. The current data were obtained freely and without the same potential desirability that may result from formal research, which may have resulted in different findings than if this study had been conducted through formal research methods such as UX studies [19].

Comparison to Prior Work

The present findings echo previous research regarding mHealth interventions for substance use. Specifically, that "nearly all of the mHealth interventions that are available for download today have not been evaluated for efficacy or effectiveness" [7]. In addition, the wide difference in the number of available reviews across apps evaluated in this study highlights findings from previous research that user's ratings influence perceptions about the apps' overall quality [7]. TryDry and Celebrate Recovery had significantly more user ratings and reviews than the other apps, and those apps have more downloads. Moreover, our findings are consistent with previous research examining app user reviews of nonsubstance use mental health apps, which highlight user preferences for usability, minimization of technical issues, and personalization [20-22]. As such, app developers should listen to what user reviews say in order to learn what makes one app successful compared with another from user experience and access perspectives. Also consistent with previous studies of formal AUD treatment and behavior change literature more broadly, skills-based tools and tracking capabilities of alcohol-related mobile apps arose as 2 of the most widely referenced content in the present qualitative analyses.

Conclusions

Mobile apps hold great promise to help the millions of individuals worldwide who experience serious harms associated with alcohol use. They may reach a wider audience than traditional, in-person therapy due to the ubiquity of smartphones and the free nature of these apps. Moreover, as compared with in-person treatment, receiving help in changing one's relationship with alcohol through a mobile app provides the app user with inherently increased privacy and autonomy. However, many apps that have been developed are not accessed or are not science-based: this study highlights ways to translate evidence-based techniques to mobile apps by listening to what the app users themselves have to say. Although many of the reviews for these apps are similar to reviews of any product by being either generic praise or generic, nonconstructive criticism, many of the present findings highlight concrete features that users want in an app to help change their relationship to alcohol. Flexible and versatile tools to access resources such as coping skills, psychoeducation, and community support; easy to use and rewarding or fun tracking capabilities; and ways for users to individualize their experience stand out among the more than 1300 coded utterances contained in the reviews that were examined. Similar to traditional therapy, skills-based and behaviorally based interventions that allow users autonomy were widely recognized as being helpful, appreciated, or, where lacking in an app, desired.

XSL•FO
RenderX

Acknowledgments

Preparation of this manuscript was supported in part by grants from the National Institute on Alcohol Abuse and Alcoholism (T32 AA018108-14 and R01 AA031159-01) and the National Institute on Drug Abuse (RM1 DA055301-02). The authors would like to thank Dr Claire Garnett for her consultation and expertise in alcohol mHealth intervention development research and mobile apps from the United Kingdom.

Conflicts of Interest

None declared.

Checklist 1
PRISMA checklist.
[PDF File, 60 KB - mhealth_v13i1e63148_app1.pdf]

References

1. Patel AK, Balasanova AA. Unhealthy alcohol use. JAMA 2021 Jul 13;326(2):196. [doi: 10.1001/jama.2020.2015] [Medline: 34255006]
2. National Institute on Alcohol Abuse and Alcoholism. Alcohol treatment in the United States. 2024. URL: https://www.niaaa.nih.gov/alcohols-effects-health/alcohol-topics/alcohol-facts-and-statistics/alcohol-treatment-united-states#:~:text=Among%20an%20estimated%2029.5%20million,alcohol%20in%20the%20past%20year [accessed 2024-04-08]
3. Venegas A, Donato S, Meredith LR, Ray LA. Understanding low treatment seeking rates for alcohol use disorder: a narrative review of the literature and opportunities for improvement. Am J Drug Alcohol Abuse 2021 Nov 2;47(6):664-679. [doi: 10.1080/00952990.2021.1969658] [Medline: 34464542]
4. Yule AM, Kelly JF. Integrating treatment for co-occurring mental health conditions. Alcohol Res 2019 Oct 24;40(1):arcr.v40.1.07. [doi: 10.35946/arcr.v40.1.07] [Medline: 31649837]
5. Castillo-Carniglia A, Keyes KM, Hasin DS, Cerdá M. Psychiatric comorbidities in alcohol use disorder. Lancet Psychiatry 2019 Dec;6(12):1068-1080. [doi: 10.1016/S2215-0366(19)30222-6] [Medline: 31630984]
6. Hunt GE, Large MM, Cleary M, Lai HMX, Saunders JB. Prevalence of comorbid substance use in schizophrenia spectrum disorders in community and clinical settings, 1990-2017: systematic review and meta-analysis. Drug Alcohol Depend 2018 Oct 1;191:234-258. [doi: 10.1016/j.drugalcdep.2018.07.011] [Medline: 30153606]
7. Businelle MS, Perski O, Hébert ET, Kendzor DE. Mobile health interventions for substance use disorders. Annu Rev Clin Psychol 2024 Jul;20(1):49-76. [doi: 10.1146/annurev-clinpsy-080822-042337] [Medline: 38346293]
8. Pew Research Center. Mobile fact sheet. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-01-02]
9. Hébert ET, Ra CK, Alexander AC, et al. A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial. J Med Internet Res 2020 Mar 9;22(3):e16907. [doi: 10.2196/16907] [Medline: 32149716]
10. Tami-Maury I, Klaff R, Hussin A, et al. A text-based smoking cessation intervention for sexual and gender minority groups: protocol for a feasibility trial. JMIR Res Protoc 2022 Dec 9;11(12):e42553. [doi: 10.2196/42553] [Medline: 36485022]
11. Ghassemlou S, Marini C, Chemi C, Ranjit YS, Tofighi B. Harmful smartphone applications promoting alcohol and illicit substance use: a review and content analysis in the United States. Transl Behav Med 2020 Oct 12;10(5):1233-1242. [doi: 10.1093/tbm/ibz135] [Medline: 33044528]
12. Shelley M, Krippendorff K. Content analysis: an introduction to its methodology. J Am Stat Assoc 1984 Mar;79(385):240. [doi: 10.2307/2288384]
13. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
14. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
15. Stemler S. An overview of content analysis. Pract Assess Res Eval 2000;7:1. [doi: 10.7275/z6fm-2e34]
16. Shek DTL, Tang VMY, Han XY. Evaluation of evaluation studies using qualitative research methods in the social work literature (1990-2003): evidence that constitutes a wake-up call. Res Soc Work Pract 2005 May;15(3):180-194. [doi: 10.1177/1049731504271603]
17. Sampat BH, Prabhakar B. Privacy risks and security threats in mHealth apps. J Int Technol Inf Manag 2017 Dec;26(4):126-153. [doi: 10.58729/1941-6679.1353]
18. Mitchell S, Ridley S, Tharenos C, Varshney U, Vetter R, Yaylacicegi U. Investigating privacy and security challenges of mHealth applications. 2013 URL: https://aisel.aisnet.org/amcis2013/HealthInformation/RoundTablePresentations/6 [accessed 2025-03-07]
19. Frith KH. User experience design: the critical first step for app development. Nurs Educ Perspect 2019;40(1):65-66. [doi: 10.1097/01.NEP.0000000000000451] [Medline: 30575692]

XSL•FO
RenderX

20. Oyebode O, Alqahtani F, Orji R. Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews. IEEE Access 2020;8:111141-111158. [doi: 10.1109/ACCESS.2020.3002176]
21. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066. [doi: 10.1177/1460458219896492] [Medline: 31920160]
22. Funnell EL, Spadaro B, Martin-Key N, Metcalfe T, Bahn S. mHealth solutions for mental health screening and diagnosis: a review of app user perspectives using sentiment and thematic analysis. Front Psychiatry 2022;13:857304. [doi: 10.3389/fpsyt.2022.857304] [Medline: 35573342]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
AUD: alcohol use disorder
mHealth: mobile health
PTSD: posttraumatic stress disorder
UX: User Experience

Edited by L Buis; submitted 11.06.24; peer-reviewed by A Sriram, FW Chi, R Pryss; revised version received 21.11.24; accepted 11.02.25; published 19.03.25.

Please cite as:
Kirouac M, Gillezeau C
Preferences for Mobile Apps That Aim to Modify Alcohol Use: Thematic Content Analysis of User Reviews
JMIR Mhealth Uhealth 2025;13:e63148
URL: https://mhealth.jmir.org/2025/1/e63148
doi:10.2196/63148

© Megan Kirouac, Christina Gillezeau. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Insights Into How mHealth Applications Could Be Introduced Into Standard Hypertension Care in Germany: Qualitative Study With German Cardiologists and General Practitioners

Susann May[1,2], MPH; Frances Seifert[1,2], MSc; Dunja Bruch[2], MSc; Martin Heinze[3], Prof Dr Med; Sebastian Spethmann[4,5*], Prof Dr Med; Felix Muehlensiepen[1,2,6*], Dr rer med, PhD

[1]Center for Health Services Research, Brandenburg Medical School Theodor Fontane, Rüdersdorf, Germany

[2]Department of Cardiovascular Surgery, Heart Center Brandenburg, Faculty of Health Sciences, Brandenburg Medical School Theodor Fontane, Bernau bei Berlin, Germany

[3]Department of Psychiatry and Psychotherapy, Immanuel Klinik Rüdersdorf, University Hospital of Brandenburg Medical School Theodor Fontane, Rüdersdorf, Germany

[4]Deutsches Herzzentrum der Charité – Department of Cardiology, Angiology and Intensive Care Medicine, Berlin, Germany

[5]Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[6]AGEIS, Université Grenoble-Alpes, Grenoble, France

[*]these authors contributed equally

Corresponding Author:
Susann May, MPH
Center for Health Services Research
Brandenburg Medical School Theodor Fontane
Seebad 82/83
Rüdersdorf, 15562
Germany
Phone: 49 3391 39 145 91
Email: susann.may@mhb-fontane.de

Abstract

Background: Mobile health (mHealth) apps provide innovative solutions for improving treatment adherence, facilitating lifestyle modifications, and optimizing blood pressure control in patients with hypertension. Despite their potential benefits, the adoption and recommendation of mHealth apps by physicians in Germany remain limited. This reluctance may be due to a lack of understanding of the factors influencing physicians' willingness to incorporate these digital tools into routine clinical practice. Understanding these factors is crucial for fostering greater integration of mHealth apps in hypertension care.

Objective: The aim of this study was to explore the relationship between physicians' information needs and acceptance factors, and how these elements can support the effective integration of mHealth apps into daily medical routines.

Methods: We conducted a qualitative study involving 24 semistructured telephone interviews with physicians, including 14 cardiologists and 10 general practitioners, who are involved in the treatment of hypertensive patients. Participants were selected through purposive sampling to ensure a diverse range of perspectives. Thematic analysis was conducted using MAXQDA software (Verbi GmbH) to identify key themes and subthemes related to the acceptance and use of mHealth apps.

Results: The analysis revealed significant variability in physicians' information needs regarding mHealth apps, particularly concerning their functionalities, clinical benefits, and potential impact on patient outcomes. These informational gaps play a critical role in determining whether physicians are willing to recommend mHealth apps to their patients. Key determinants influencing acceptance were identified, including the availability of robust knowledge about the apps, high-quality and reliable data, generational shifts within the medical profession, solid evidence supporting the effectiveness of the mHealth apps, and clearly defined areas of application and responsibilities within the physician-patient relationship. The study found that acceptance of mHealth apps could be significantly increased through targeted educational initiatives, enhanced data quality, and better integration of these tools into existing clinical workflows. Furthermore, younger physicians, more familiar with digital technologies, demonstrated greater openness to using mHealth apps, suggesting that generational changes may drive future increases in adoption.

Conclusions: The successful integration of mHealth apps into hypertension management requires a multifaceted approach that addresses both the informational and practical concerns of physicians. By disseminating comprehensive knowledge about the

XSL•FO
RenderX

variety, functionality, and proven efficacy of hypertension-related mHealth apps, health care providers can be better equipped to use these tools effectively. This approach necessitates the implementation of various knowledge transfer strategies, such as targeted training programs, peer learning opportunities, and active engagement with digital health technologies. As physicians become more informed and confident in the use of mHealth apps, their acceptance and recommendation of these tools are likely to increase, leading to more widespread adoption. Overcoming current barriers related to information deficits and data quality is essential for ensuring that mHealth apps are optimally used in routine hypertension care, ultimately improving patient outcomes and enhancing the overall quality of care.

Trial Registration: German Clinical Trials Register DRKS00029761; https://drks.de/search/de/trial/DRKS00029761

International Registered Report Identifier (IRRID): RR2-10.3389/fcvm.2022.1089968

(JMIR Mhealth Uhealth 2025;13:e56666) doi:10.2196/56666

KEYWORDS

hypertension; mHealth apps; physicians; qualitative study; digitalization; app; application; Germany; blood pressure; cardiologists; thematic analysis; general practitioner; mobile phone

Introduction

Advancing digitalization of health care has a transformative impact on the prevention and treatment of diseases, particularly in the realm of hypertension. In this context, mobile health (mHealth) apps are increasingly gaining significance, offering innovative means for monitoring, maintaining, and preventing high blood pressure [1-3].

Hypertension poses a substantial health challenge, affecting millions of people worldwide [4]. In Germany, it is one of the most prevalent health conditions, affecting a considerable portion of the population. In Germany, 19 million people with statutory health insurance (26.3%) were diagnosed with high blood pressure in 2018 [5]. Prevention and successful management of hypertension are imperative in averting potentially severe complications, including ischemic heart disease, stroke, and renal disorders. In addition, these measures play a pivotal role in enhancing the overall quality of life of patients with hypertension [6-8].

Preventing high blood pressure using a smartphone app that can be accessed at any time appears to be a promising tool [9]. Lifestyle interventions for preventive behavior, such as improving medication adherence, restricting food intake, or promoting activity, can generally be encouraged with the help of mHealth apps [10-17]. A systematic review concluded that the use of mHealth apps in hypertension care by general practitioners could be very useful, although various barriers, such as generational differences in willingness to use digital devices and the lack of knowledge of available mHealth apps, need to be overcome [18]. In the future, an increasing number of digital technologies will be used in health care [19]. To address this development, it is necessary to understand the current health care situation, especially the insights of physicians and their perspectives on benefits and challenges, as well as their information needs, to best integrate digital technologies into care.

To date, there has been limited research on physicians' perceptions of mHealth apps and devices and their willingness to prescribe them for hypertension care. In general, research has shown that physician adoption of mHealth apps depends on several factors, including perceived usefulness, ease of use, and trustworthiness of the technology. For example, Gagnon et al [20] found that the acceptance of mHealth apps by health care providers was highly dependent on their perception of clinical usefulness and ease of use. An Australian study highlighted concerns, including the digital divide between generations, lack of knowledge about prescription mHealth apps and devices, additional working hours, and data security concerns [21]. A French study [22] investigated general practitioners' attitudes toward prescribing mHealth apps and identified three main attitudes among general practitioners: (1) digital engagement, emphasizing the benefits of mHealth as a complementary tool for medical practice; (2) patient protection, focusing on concerns about data security, misinformation, and patient safety; and (3) doctor protection, highlighting concerns about increased workload and the impact on the doctor-patient relationship. A recent study on attitudes toward the use of digital health applications (in German: digitale Gesundheitsanwendungen, DiGA; digital therapeutics that can be prescribed by physicians in Germany) in Germany revealed information deficits. Both doctors and patients are insufficiently informed about mHealth apps and their health benefits, leading to a lack of trust [23]. A descriptive study of German general practitioners found that 60% of general practitioners recognized mHealth apps as beneficial for the health management of their patients, but only 36% had an overall positive opinion, and only 18% regularly recommended such mHealth apps [24]. However, despite these findings regarding diverse information needs, there is little evidence specifically examining the acceptance of mHealth apps by physicians in the context of hypertension management in Germany. Although physicians generally have a positive perception of such apps, this has not yet translated into widespread adoption in practice. This discrepancy suggests the presence of underlying, unaddressed barriers—whether related to information gaps, workflow integration issues, or confidence in the technology—which prevents clinicians from actively recommending or using these tools. It is these hidden factors that our research aims to uncover and address in order to gain a deeper understanding of the dynamics influencing the adoption of mHealth apps in hypertension care. Most previous studies have either focused on other chronic diseases or on the patient's perspective, leaving a gap in understanding physician-specific barriers related to hypertension management.

XSL•FO
RenderX

In addition, the German health care system has specific structural and regulatory characteristics that can affect the integration of mHealth apps. Germany is characterized by strict data protection regulations, as laid down in the European Union's General Data Protection Regulation (in German: Datenschutz-Grundverordnung, DSGVO). Especially in the sensitive context of mHealth apps, this leads to additional uncertainties and reservations among physicians who want to integrate these technologies into their clinical practice. Furthermore, fragmented digital infrastructures, complex reimbursement mechanisms for digital health applications, and varying degrees of digitalization in the different health care sectors further complicate acceptance. It is crucial to address these barriers by exploring the specific information needs of physicians regarding mHealth apps to ensure their effective implementation. Evidence shows that eHealth implementations often fail because physicians are not adequately involved in the development and integration processes [25]. These structural and informational gaps contribute to the disconnect between the theoretical openness toward mHealth apps and their practical adoption in routine care.

A comprehensive understanding of physicians' concerns, expectations, and knowledge gaps could guide the development of training programs that enhance their skills in handling digital preventive tools. Furthermore, investigating the acceptance of these technologies provides insights into the potential challenges that medical staff may face, enabling the design of user-friendly mHealth apps that can seamlessly integrate into clinical practice.

This study focused on investigating the information needs and acceptability of mHealth apps for the prevention of hypertension by health care providers. The aim of this study was to develop an understanding of how information needs and acceptance factors are related to the best integration of mHealth apps into hypertension care. This study addressed the following research questions: (1) What is the self-perceived acceptability of mHealth apps in hypertension care by physicians, and how could it be increased? (2) What information needs do physicians have, and how can these be met? (3) What is needed to integrate mHealth apps into the health care system in the long term, and what practical implications can be derived from this?

Methods

Study Design

This paper presents an investigation of physicians' perspectives on the use of digital technologies in preventing hypertension. Using qualitative data collection methods, comprehensive interviews were conducted with general practitioners and cardiologists.

Ethical Considerations

All the ethical issues were addressed. All experimental protocols were approved by the institutional or licensing committee. All methods were performed in accordance with relevant guidelines and regulations. Written consent was obtained from all participants after they were given the opportunity to ask clarifying questions. This study was approved by the Ethics Committee of the Brandenburg Medical School Theodor Fontane (E-02-20220620). Participants were informed verbally and in writing about the purpose, the procedure, the significance of the study, and the benefits and risks that may be associated with it and had the opportunity to ask questions. They were also informed that they had the right to withdraw their consent to participate in the study at any time, either verbally or in writing, without giving reasons. They were also informed that personal data would be recorded and stored, whereby the data would be pseudonymized, but that no data would be published that could be used to identify the person. For security reasons, the data received from the participants was always stored in a password-protected folder on a secure desktop computer. Written consent was obtained after participants were given the opportunity to ask questions. Patients were not involved in designing this study. Participants were offered €75 (approximately US $77.89) as an incentive for their participation in the study.

Participant Recruitment

The participants were selected using purposive sampling [26]. We included both general practitioners and cardiologists. Further inclusion criteria were current involvement in the treatment of patients with hypertension and interest in participating in an interview. The participants were recruited by snowball sampling. The following sources were used for the recruitment:

- Social media accounts of the university (Brandenburg Medical School Theodor Fontane): Facebook, Instagram, and Twitter.
- Association of Statutory Health Insurance Physicians Brandenburg: newsletter.
- General Practitioners Association Brandenburg: information study participation at events.
- Personal contact among colleagues.

Physicians were offered €75 (US $77.89) as an incentive for their participation in the study, which was financed from the earmarked research funds. This measure was taken to reward the time spent by doctors and to ensure sufficient willingness to participate. This study was conducted independently and was not directly related to the Innovation Committee of the Federal Joint Committee (G-BA: Committee in Germany that determines which medical services are provided to people with statutory health insurance). There were no external influences on the design, conduct, or evaluation of the study, and there were no conflicts of interest that could compromise the independence and reliability of the results.

Data Collection

A preliminary semistructured interview guide was collaboratively developed by a multiprofessional team (SM, FM, DB, and SS). The interview guide comprised open-ended questions exploring participants' familiarity with digital applications, particularly mHealth apps, their perceived benefits and barriers, their information needs for prevention, and aspects that influence their acceptance of mHealth apps. Sample questions included inquiries about the use of mHealth apps in hypertension treatment, perceived benefits, encountered barriers, and level of knowledge about mHealth apps (please see the interview guide in Multimedia Appendix 1). In addition,

XSL•FO
RenderX

sociodemographic data, such as profession, gender, age, number of inhabitants, duration of professional activity, and setting, were collected. The present findings report on information needs and aspects of acceptance.

To ensure the clarity and relevance of the questions, a pilot test of the interview guide was conducted with 5 eligible participants recruited from clinics and outpatient physicians in the study's catchment area. All interviews were conducted through telephone between October 2022 and March 2023 and were recorded and transcribed verbatim. Data collection continued until no new findings emerged and content saturation was achieved. Saturation was defined as code saturation, indicating no additional issues identified, and meaning saturation, indicating no further dimensions, nuances, or insights [27]. Field notes were recorded for each interview. They were used to understand the interview situation better.

Data Analysis

Qualitative analysis of the interviews was performed iteratively by SM and FM based on thematic analysis [19] using MAXQDA Analytics Pro 2022 (release 22.1.0; Verbi GmbH). After transcription of the audio material, the analysis began with familiarization with the interviews. A sample of the transcripts was then independently analyzed using SM and FM to iterate and finalize a comprehensive codebook. The main categories and subcategories are formed from these codes. The thematic analysis included reading each transcript, identifying patterns, assigning codes, and formulating themes and subthemes from the data. After this process, the data were checked, compared, and discussed to eliminate discrepancies [28]. If no new data, themes, or relationships could be identified, data saturation was achieved [29]. To present these findings, significant excerpts from the transcriptions were chosen as representative quotes. These quotes have been translated into English and incorporated into the manuscript. The manuscript was compiled in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research; see Multimedia Appendix 2) [30].

Results

Overview

A total of 24 interviews were conducted and analyzed. The mean duration of the interviews was 42 (SD 5.75; minimum 33, maximum 56) minutes. The mean age of the participants was 50 (range 35-74) years. Most participants were male (13/24, 54%). In total, 14 cardiologists and 10 general practitioners participated in the study. A total of 7 individuals worked in the inpatient setting, and 18 people worked in the outpatient setting. Furthermore, 1 person worked in both settings. In addition, 4 of the physicians already had experience with digital applications in daily practice or were working with them. Table 1 shows the detailed characteristics of the study participants.

XSL•FO
RenderX

Table 1. Detailed characteristics of study participants.

ID	Profession	Duration of professional activity (years)	Age (years)	Sex	Number of inhabitants (workplace), n	Setting
K1	Cardiologist	14	45	Male	3.6 million	Inpatient
K2	Cardiologist	49	74	Male	1.4 million	Outpatient
K3	Cardiologist	7	48	Male	60,000	Outpatient and inpatient
K4	Cardiologist	20	57	Male	10,000	Outpatient
K5	Cardiologist	8	44	Male	310,000	Outpatient
K6	Cardiologist	29	62	Male	3.4 million	Inpatient
K7	Cardiologist	38	55	Female	90,000	Outpatient
K8	Cardiologist	4	41	Female	3.4 million	Inpatient
K9	Cardiologist	18	55	Female	20,000	Outpatient
K10	Cardiologist	7	41	Male	30,000	Outpatient
K11	Cardiologist	7	40	Female	13,000	Outpatient
K12	Cardiologist	11	52	Male	300,000	Outpatient
K13	Cardiologist	20	51	Female	130,000	Inpatient
K14	Cardiologist	20	51	Male	3.6 million	Inpatient
H1	General practitioner	21	45	Male	3.6 million	Outpatient
H2	General practitioner	29	57	Female	15,000	Outpatient
H3	General practitioner	5	40	Male	70,000	Outpatient
H4	General practitioner	5	40	Female	70,000	Outpatient
H5	General practitioner	30	60	Female	3.6 million	Outpatient
H6	General practitioner	4	35	Female	3.6 million	Outpatient
H7	General practitioner	5	56	Male	8000	Outpatient
H8	General practitioner	12	36	Male	7000	Inpatient
H9	General practitioner	30	63	Female	900	Outpatient
H10	General practitioner	21	54	Female	3.6 million	Outpatient

Self-Perceived Acceptability and Acceptance-Increasing Factors

Neither a particularly positive nor negative attitude toward mHealth apps was identified. Rather, physicians' response behavior is characterized by heterogeneity. mHealth apps are generally accepted by physicians, as long as they are not part of their daily medical routines. This means that they are certainly open-minded but not in their own care settings, especially due to a lack of experience in using mHealth apps.

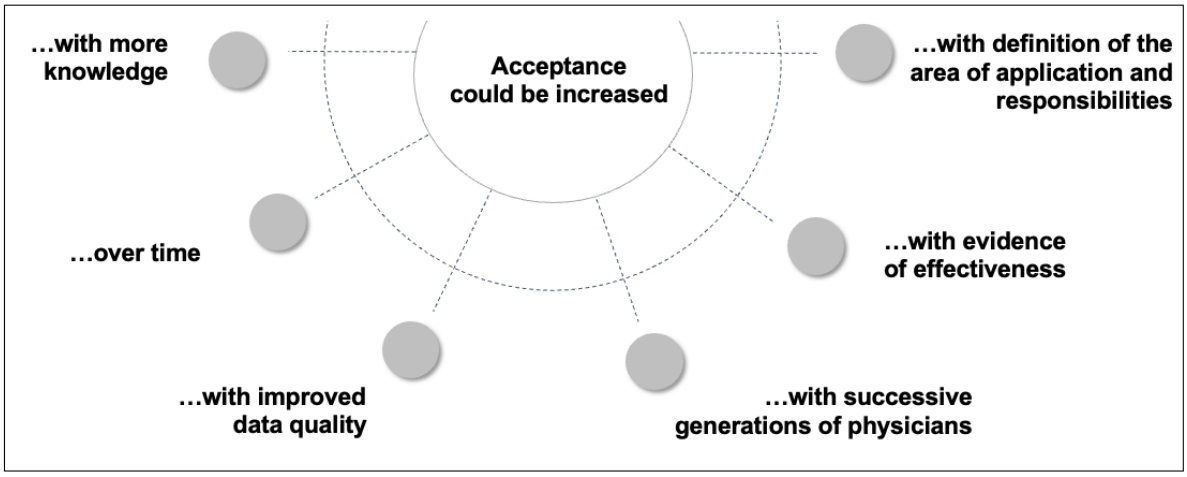
> *Well, I have to answer that in such a way that it is neutral, so to speak, because I simply do not have enough experience so far.* [K3]

> *I do not think I need it in my daily routine. I know a bit about high blood pressure and about general measures and therapies, and I try to convey that to the patients in conversation and then use it. So, I do not think I have discovered the need for it at the moment, to what extent it could take work off my hands or to what extent the patients could be better blood pressure controlled, or something else.* [K3]

Nevertheless, physicians emphasized that acceptance could increase. Figure 1 shows the factors associated with the increase in acceptance, and are further discussed in the following subsections.

XSL•FO
RenderX

Figure 1. Identified subthemes of the acceptance-increasing factors (coding tree).



Acceptance Could Be Increased With More Knowledge

Overall, the data revealed that physicians' acceptance of hypertension apps could increase with a deeper understanding of their benefits, targeted training, and seamless integration into existing clinical workflows.

> *And these reservations are simply ignorance and also ignorance of the actual tools that are available.* [K2]

Acceptance Could Be Increased Over Time

The interviewed physicians emphasized that the acceptance of digital technologies and mHealth apps in medical practice is often subject to a natural maturation process in which initial concerns and uncertainties diminish over time. This also applies to hypertension apps, which serve as technologies for the monitoring and self-management of blood pressure. For example, physicians indicated that increasing acceptance is significantly influenced by time as experience accumulates, research findings grow, and technology proves its reliability. In the early stages of mHealth app adoption, some physicians may be hesitant due to uncertainties about data accuracy, privacy concerns, and integration into existing workflows. However, over time, positive experiences and feedback from colleagues who have already implemented such technologies will help alleviate concerns and build confidence in the effectiveness of mHealth apps.

> *It will be unavoidable (the use of apps in daily routines). Thus, the industry is flooding us with apps. Studies are certainly required to be conducted. That's all true. We will certainly be doing a lot of things digitally in ten or 20 years.* [K14]

Acceptance Could Be Increased With Improved Data Quality

A key factor influencing the acceptance of hypertension apps among physicians is the data quality offered by these apps. Physicians reported concerns regarding the reliability of the collected data, leading to reluctance to rely on such technologies. In particular, it is not always guaranteed that patients enter their blood pressure data correctly. Overall, the ongoing improvement in data quality will help reduce existing reservations about hypertension apps and further increase their acceptance among physicians.

> *I believe that we can foresee that this will be accepted because the data [on use and effectiveness] will presumably also have better quality in the future.* [K2]

Acceptance Could Be Increased With Successive Generations of Physicians

The ongoing generational shift in the medical profession plays a decisive role in increasing acceptance of hypertension apps. Physicians emphasize that younger physicians who have grown up in an era of digital innovation are often more open to the use of modern technology in medical practice. Unlike the older generations, who may be more familiar with traditional methods of patient monitoring, young physicians already have an affinity for digital health solutions. Digital natives recognize the potential of mHealth apps as efficient technologies for improving patient care. Their willingness to accept and integrate innovative technologies is fueled by the realization that these apps can not only promote patient autonomy but also increase efficiency in clinical practice. The younger generation's openness to new technologies is driving the integration of mHealth apps into the modern medical landscape. As the digital-savvy generation enters medical practice, the use of mHealth apps is likely to be seen as a natural part of treatment. Interestingly, participants in all age groups expressed this impression.

> *Of course, this [use of mHealth apps] comes with younger physicians who have grown digitally in a completely different way. For them, it will be a matter of course that they send their data, laboratory results, and everything else digitally.* [K2]

Acceptance Could Be Increased With Evidence of Effectiveness

The physicians reported that they need solid evidence that the use of hypertension apps is not only technologically advanced but also achieves positive results in clinical practice. Clear evidence of improvements in blood pressure control, reduction

XSL•FO
RenderX

of complications, and overall better health care is essential from the point of view of strengthening physician conviction.

> *Certainly, scientific proof that it makes sense. I am always in favor of prescribing something like this if it is proven to be beneficial for the patient. However, I do not see that yet.* [K9]

Acceptance Could Be Increased With a Definition of the Area of Application and Responsibilities

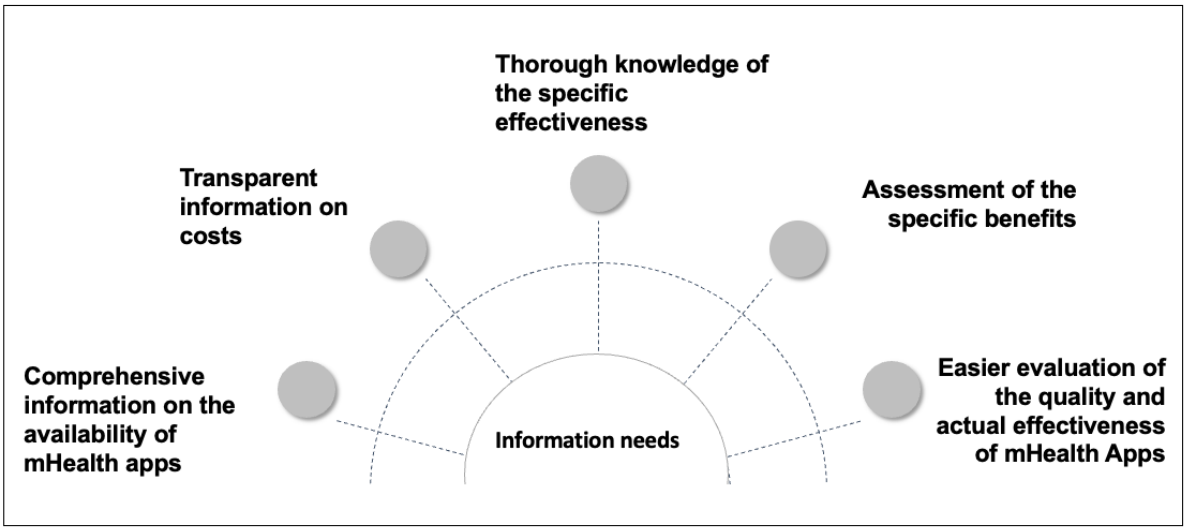
In the interviews, the physicians addressed the fact that the purpose and functions of the use of mHealth apps are often not defined, either for the patients themselves or within physician-patient communication. Physicians point out that it should be clear to both patients and physicians how an mHealth app is integrated into care. Whether the mHealth app is used as a lifestyle product or a monitoring tool. Whether physicians are actively involved in the evaluation of the data, whether they should take on a controlling function, and whether the use of the mHealth app is only for observation or should even have consequences, for example, through advice from the physician.

> *I always think the question is interesting: what do I want to achieve with the app? Do I want truest for therapy management? These are the patients we were just talking about, who are not health-conscious per se. Do we want people to feel comfortable with their illness and feel that they are doing an incredible amount for themselves? That brings us to advice, offers, and lifestyle apps - that is, what I call it, right? Somewhere along the line, if it is from a medical context, it should take both facets into account but focus more on the medical context.* [K14]

Information Needs

A variety of information needs became apparent in the interviews. Figure 2 shows the subthemes of information needs.

Figure 2. Identified subthemes of the information needs (coding tree). mHealth: mobile health.



Comprehensive information on the Availability of mHealth Apps

Physicians are generally interested in using mHealth apps in their daily routine. However, there is a lack of information on the range of mHealth apps that can be integrated into hypertension care. There is a knowledge gap regarding the functions, data security, accuracy, and clinical benefits of hypertension apps.

> *In my opinion, the market has been overly flooded. And then you do not even know which apps are actually certified and which are available. Thus, we cannot know all apps that are buzzing around the market.* [K7]

Transparent Information on Costs

A challenging aspect in the context of integrating mHealth apps into medical practice is that many physicians are often not fully informed about the costs associated with these mHealth apps. Nontransparent pricing and diverse pricing models of various providers lead to uncertainty. This lack of clarity regarding financial aspects can affect physicians' decision-making and even make them reluctant to integrate such mHealth apps into their treatment plans.

> *I do not know enough about what is available or what you can choose from and what that means in terms of costs.* [K3]

Thorough Knowledge of the Specific Effectiveness

The lack of knowledge regarding the specific effectiveness of hypertension apps poses a further challenge for physicians. There is often a lack of clear information about how these apps can have a positive effect on the treatment of patients with high blood pressure in practice. In contrast, physicians emphasize that they know and can assess the use and effects of medication. Physicians need detailed insights into the algorithms, data analysis methods, and interpretation of the information generated by the mHealth apps to understand and assess their effectiveness.

> *Yes, and I mean, when I prescribe ramipril, it's because I know this medication, I know its effects, side effects, I have used it for a long time, and I know*

XSL•FO
RenderX

it. However, these apps are expensive. They cost 500 euros, so I want to know in advance what I am prescribing and how they work. [H10]

Assessment of the Specific Benefits

The limitations in assessing the specific benefits of hypertension apps pose a challenge to physicians. There is often a lack of clear and convincing information about how these mHealth apps actually benefit patients. In addition, physicians are unable to assess the extent to which apps benefit their daily practice.

So, I think at the moment I just have not identified the need for myself to what extent this could take work off my hands, how it would benefit the patients, and to what extent the patients could be better blood pressure adjusted or something. [K3]

Easier Evaluation of the Quality and Actual Effectiveness of mHealth Apps

Physicians' difficulty in assessing the quality of mHealth apps is a key barrier. There is often a lack of clear benchmarks and objective criteria for evaluating the actual benefits of these mHealth apps for individual patient care. The lack of standardized evaluation methods makes it difficult for physicians to determine the degree of effectiveness of a particular hypertension app and assess its clinical relevance.

And I would always look at the study situation first because many of them have only been proven with very poor studies and are, therefore, only provisionally approved. I do not think I would prescribe them at the moment because I think they cost money and have not been sufficiently tested. I do not even know how good they are. [H6]

Evaluating the effectiveness of hypertension apps is a complex challenge for physicians. The lack of standardized criteria for collecting outcome data and comprehensive clinical trials makes it difficult to draw objective conclusions about the effectiveness of these mHealth apps. Physicians often struggle to assess how well hypertension apps actually work in practice and what concrete contribution they can make to improving blood pressure control and the health of their patients. To make matters worse, many of the available hypertension apps have different functions and approaches, which further complicate a standardized assessment.

As I said, I'm too uneducated when it comes to this. I am unfamiliar with any studies and do not know how they work or how effective they are. [K3]

Discussion

Principal Findings

This study provides new insights into the factors influencing physicians' adoption of mHealth apps in the context of hypertension management in Germany. Our findings reveal a complex interaction of factors—from information gaps and workflow integration challenges to uncertainties around data security and reimbursement structures—which influence physicians' willingness to integrate these technologies into their daily routines.

This study provides a nuanced view of the acceptance of mHealth apps among cardiologists and general practitioners, particularly in the context of hypertension management. While there is a general openness to digital technologies among physicians, hesitation persists, primarily due to a lack of familiarity and experience with these technologies. A significant barrier to the adoption of mHealth apps is a limited understanding of their benefits and functionalities. Physicians have indicated that targeted training could greatly enhance acceptance. Over time, acceptance is expected to increase as experience grows and positive outcomes are observed. However, concerns regarding the accuracy of patient-entered data remain a major issue, with improved data reliability being essential for building trust in these technologies. Younger physicians, who are more familiar with digital technologies, tend to be more receptive to mHealth apps, suggesting that acceptance is likely to grow as the medical field evolves. Nevertheless, substantial information gaps, including a lack of awareness of the range and benefits of available apps, hinder adoption. In addition, unclear cost structures and the absence of standardized methods for assessing quality and effectiveness further contribute to physicians' reluctance to recommend and use these technologies.

Comparison With Previous Work and Theoretical Embedding

Self-Perceived Acceptability of Physicians Toward mHealth Apps in Hypertension Care

A major obstacle to the introduction of mHealth apps into standard hypertension care is a lack of understanding of their benefits and functions. The aspect "Acceptance could be increased with more knowledge" can be linked to the theoretical concept of the "Technology Acceptance Model" (TAM). TAM, developed by Davis [31], postulates that the acceptance of technology depends largely on 2 factors: its perceived usefulness and perceived ease of use. In terms of mHealth apps, physicians with a sound understanding of the specific benefits and ease of integration of these technologies into their clinical practice are more likely to accept and use them. When physicians are better informed about the positive clinical outcomes that can be achieved through mHealth apps, as well as aspects related to practical application, such as data integrity and privacy, their perceived usefulness and ease of use increase. This increased knowledge reduces uncertainty and skepticism, which ultimately leads to greater acceptance and integration of technology into everyday medical practice. A systematic review examining factors influencing acceptance concluded that the lack of routine use of mHealth apps is primarily due to insufficient knowledge and information regarding their integration into daily routines. [32]. The need for information is a key component of acceptance, particularly in telemedicine [33], as well as in mHealth apps. The way in which information is provided and how well users' information needs are met have a significant influence on their willingness to accept and use a particular technology. A recent study [34] pointed out that the level of awareness of mHealth apps for hypertension should be increased through targeted information campaigns and physician training. In addition, physicians need to test mHealth apps, and experience and acceptance would increase.

XSL•FO
RenderX

Acceptance is expected to increase over time as more experience is gained and positive outcomes are observed. This aspect can be linked to the diffusion of innovation theory proposed by Rogers [35]. According to Rogers' theory, the adoption of new technologies follows a specific pattern over time and is categorized into 5 stages: innovators, early adopters, the early majority, the late majority, and laggards. Initially, only a small group of innovators and early adopters might embrace mHealth apps; however, as more physicians witness positive outcomes and share their experiences, the early and late majority will gradually adopt these technologies. Over time, as research continues to provide robust evidence of the effectiveness and reliability of mHealth apps and as the technology becomes more integrated into standard medical practices, the remaining skeptical physicians (the laggards) will also start to accept and use these technologies. The stronger the habit of using mHealth apps, the stronger the connection between user satisfaction and willingness to continue using the apps [36]. People who are satisfied with and develop habits related to new technologies are more likely to use these apps in the long term [36].

Another important finding of our study is that there are concerns about the accuracy of the data entered by patients, and improved data reliability is crucial for building trust in these technologies. This is in line with the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) [37]. One of the key determinants within this framework is the concept of trust, which is strongly influenced by the quality of the data provided by technology. High-quality data, characterized by accuracy, reliability, timeliness, and relevance, can significantly increase physician trust in mHealth apps. When data quality is assured, physicians are more likely to perceive technology as useful and reliable, reducing their uncertainty and increasing their willingness to adopt and integrate these technologies into their clinical practice. This improved confidence and the perceived benefits of better data quality can lead to higher overall adoption of mHealth apps among healthcare professionals. A recent study [38] showed that doctors were willing to use mHealth apps if they felt comfortable and safe. This highlights the importance of user experience in the adoption of mHealth apps by health care professionals. The intention to use mHealth apps in everyday health care depends on the perceived reliability of these technologies. If physicians perceive mHealth apps as reliable and trustworthy, they are more likely to regularly integrate them into their workflows [36].

The identified aspect, "Acceptance could be increased with successive generations of physicians," can be linked to the concept of the "Digital Divide," which highlights the disparities in access to and use of digital technologies across different demographic groups [39-41]. Younger physicians with digital technologies are less likely to experience the barriers associated with the digital divide. They generally have better access to digital technologies, higher levels of digital literacy, and greater confidence in using technology in medical practice. As these younger, more technologically adept physicians enter the workforce, acceptance and integration of mHealth apps will naturally increase. Several studies have shown that younger doctors are more likely to use mHealth apps, and in an Australian study, both patients and physicians cited age as a barrier to mHealth app prescription but also provided examples of exceptions to this age-based divide in digital readiness [21]. A German study also identified a trend in physician age. Younger physicians are often more open to digitalization and welcome innovations such as digital health applications [42]. Another study from Germany specifies the age aspect and shows that the highest acceptance rates of digital health applications are expected in the 30-49 years age group [43].

Physicians' Information Needs and How to Address Them

Based on the assumption that acceptance could be increased with more knowledge, it is particularly important to determine what information needs exist in the context of mHealth apps in hypertension care. The interviewed physicians expressed a high demand for information regarding mHealth apps for hypertension care and various information deficits. At the same time, however, there are reservations due to uncertainties regarding effectiveness, data security, and integration into the clinical workflow. A thorough and, at the same time, a low-threshold explanation of the benefits and security aspects of such applications is therefore crucial to increase acceptance of mHealth.

The lack of information also explains why mHealth apps are not yet integrated into the health care infrastructure. Few specialists currently use or recommend mHealth apps in their daily routines [44]. This leads to a discrepancy between the available digital technologies, such as mHealth apps, and their actual use in clinical practice. The gap between the daily work with patients and the availability of mHealth apps is becoming increasingly apparent, and in this context, the "digital divide," which refers to differences in access, use, and understanding of digital technologies in health care, is also prevalent among health care practitioners [40]. On the one hand, there is the daily interaction with patients, which is characterized by personal contact, traditional diagnostic methods, and conventional treatment approaches. On the other hand, a growing number of mHealth apps offer digital solutions for monitoring health parameters, managing diseases, and accessing health information. Our results show that physicians do not know what mHealth apps can do, what it means to use them, or how much effort is involved.

In addition, a lack of information regarding costs also contributes to this. Reviews demonstrate that physicians are less willing to use eHealth technologies if the reimbursement is not clearly defined or communicated transparently [45,46]. This uncertainty surrounding financial compensation not only hinders the adoption of mHealth apps but also undermines their potential integration into routine clinical practice. Clear guidelines and transparent communication regarding reimbursement policies are essential to enhance healthcare providers' confidence in adopting these technologies.

An approach to closing the information gap resulting from the lack of assessment of the quality and effectiveness of mHealth apps is to use assessment tools. Various studies have already examined this and concluded that quality assessment tools are a viable approach to distinguish high-quality mHealth apps from lower-quality mHealth apps [47,48]. By implementing

XSL•FO
RenderX

standardized evaluation criteria, these tools can provide health care professionals with the necessary insights to make informed decisions about which apps should be integrated into their practice, ultimately enhancing patient care. Establishing clear and transparent evaluation metrics covering aspects such as design, information/content, usability, functionality, ethical issues, security and privacy, and user-perceived value would not only build trust among health care providers but also facilitate broader acceptance and integration into routine care [47].

Integrating mHealth Apps Into the Health Care System: Practical Implications

Based on our findings, we derived 4 fields of action (Multimedia Appendix 3) to achieve a more successful implementation of mHealth apps in physicians' daily routines.

Definition of Areas of Application of Responsibilities

Before mHealth apps are used for hypertension treatment, their functions should be specified in advance. The physician and patient should discuss the area of application, therapy goal, and type of collaboration. If the mHealth app is used as a lifestyle product, the patient tracks their parameters alone, and the physician is not directly involved in medical treatment. This means that the physician has no control, and there are no responsibilities for the physician; that is, they do not have to actively intervene. If the mHealth app is used as medical advice, the patient and physician monitor the parameters together, and the effectiveness and benefits should be proven in advance. This implies that physicians are directly involved in medical treatment. The physician and the patient discuss the parameters together, and the physician assumes a certain degree of control and can actively intervene. This approach could lead to mHealth apps being better accepted by physicians as medical advice and used by them more frequently. Multimedia Appendix 4 illustrates the proposed approach.

Specific Recommendations for Integration Into Daily Routines With Guidelines and Options for Measuring Outcomes

A consistent methodology for collecting outcome data and establishing evidence-based criteria is essential to enable clinicians to better understand the performance of these applications. Creating clear guidelines and evaluation tools will help reduce uncertainty about their effectiveness and provide physicians with a solid foundation for making decisions about integrating hypertension apps into their practice. In addition, interfaces and clear guidelines for the integration of such technologies must be created to promote their use by physicians. Intensive efforts should be made to obtain more information about the range, effects, and modes of action of mHealth apps [34].

Information Campaigns

In health care, knowledge translation plays a key role in the integration of new medical technologies into everyday care [49]. The acceptance of new technologies depends heavily on how well physicians understand the underlying scientific evidence, how effectively these are implemented in clinical practice, and how they bridge the gap between research and practice. Knowledge translation is essential as it facilitates the dissemination of evidence-based practices, ensuring that the latest research findings and medical innovations reach health care practitioners, policymakers, and, most importantly, the general public. Information campaigns serve as powerful vehicles for knowledge translation, acting as conduits for conveying complex medical information in an accessible and understandable manner. As the acceptance of mHealth apps is also generation-dependent, strategies that appeal to both younger and more experienced physicians should be developed. This can be achieved through various approaches, such as peer learning or mentoring. Furthermore, there should be special training programs for physicians that provide detailed information about the available mHealth apps, their functions, and scientific evidence for their effectiveness [22].

Active Use of the mHealth Apps in Collaboration With the Developers

For physicians, actively engaging with and experimenting with these apps is both practical and crucial. By exploring digital technologies, such as mHealth apps, physicians can gain valuable insights into their functions, usability, and potential benefits [50]. By actively using health apps, physicians can develop a nuanced understanding of how these technologies operate in real-world medical settings. This hands-on experience not only equips physicians with the knowledge to recommend suitable apps to patients but also fosters a deeper appreciation of the role these technologies can play in modern health care. Furthermore, physicians acting as early adopters can contribute significantly to the overall acceptance and integration of health apps within the medical community.

Strengths

To the best of our knowledge, this is the first qualitative thematic analysis to provide an in-depth understanding of the acceptance and information needs related to mHealth apps for hypertension prevention. While our study specifically examined physicians' acceptance and information needs regarding mHealth apps in hypertension management, the findings suggest that many of the factors identified, such as the need for reliable data, evidence of effectiveness, and clearly defined areas of application, may also be applicable to the management of other chronic diseases such as diabetes. This potential transferability indicates that barriers and facilitators to adopting mHealth apps in medical practice share a common foundation that extends beyond individual conditions.

Limitations

The findings could potentially be influenced by selection bias, as individuals with a greater interest in the topic might have been more inclined to participate in the study. The sampling strategy employed may not have reached a representative cross-section, leading to a lack of recording of the essential aspects. Consequently, the generalizability of the results is hindered. It should also be mentioned that the interviews were conducted over the telephone. The lack of nonverbal cues, such as body language, facial expressions, and eye contact, can make it difficult to capture the full range of emotions and reactions.

XSL•FO
RenderX

Another limitation of this study is that only 4 participating physicians had experience with mHealth apps in their daily routines. This could limit the validity of the results, as the perspectives of physicians without experience using such apps could be less informative. In addition, it is important to note that the patient interviews were not conducted during this study. Including patient perspectives could have provided a more rounded understanding of the acceptance of and information needs related to mHealth apps in hypertension care. Their absence from our research presents an area for future exploration to enrich and validate our findings. In other parts of the study, in which this study was also conducted, the patients are specifically interviewed [32].

Recognizing the preliminary nature of this broad exploration into "hypertension mHealth apps," the authors acknowledge that this serves as an initial foray. Given the extensive and diverse nature of the topic, future studies should be tailored to specific health care contexts to expand on these initial findings. These limitations present a clear avenue for future research. Plans are underway to broaden the study scope by including a larger sample in Germany [51] and to validate the qualitative outcomes through a comprehensive questionnaire survey. Subsequent investigations should focus on the identification of patients suitable or unsuitable for mHealth apps in hypertension care. Attention should also be directed toward effective strategies for educating health care professionals about mHealth and digital health in a broader sense. In addition, exploring reimbursement strategies that reinforce the optimal use of mHealth for managing chronic conditions is imperative.

Finally, the potential barriers identified here regarding the prescription of apps for hypertension by physicians pertaining to the German health care system. Further studies are required to determine whether these aspects are applicable to physicians in other countries.

Conclusion

The use of mHealth apps in hypertension management is currently characterized by information deficits and reduced acceptance among physicians. This study underlines the need for intensive knowledge translation and transfer. The active involvement of physicians in the development and implementation of mHealth apps in hypertension care is particularly important to move from individual self-management by patients to formalized mHealth app-supported health care to prevent hypertension.

Acknowledgments

The authors would like to thank all interview partners and participants for their valuable contributions and time. We would like to express special thanks to the Association of Statutory Health Insurance Physicians Brandenburg (KVBB) and the General Practitioners' Association Brandenburg for their support during the recruitment process. This study was supported by the Innovation Committee, the Federal Joint Committee (Innovation Fund; grant 01VSF18014).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. For further questions regarding the reuse of data, please contact the corresponding author.

Authors' Contributions

SM contributed to the design of the study, data acquisition and analysis, interpretation of data, and drafting of the manuscript, and had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. FS contributed to data analysis and substantively revised the manuscript. DB contributed to the design of the study, obtaining funding, data acquisition, and analysis, and substantively revising the manuscript. MH contributed to the design of the study, obtained funding, and substantively revised the manuscript. SS contributed to the study design, obtaining funding, data acquisition and analysis, and substantively revised the manuscript. FM contributed to the design of the study, data analysis, data interpretation, and substantially revised the manuscript. All authors approved the submitted manuscript. In addition, the authors have agreed to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even those in which the author was not personally involved, were appropriately investigated and resolved, and the resolution documented in the literature.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 44 KB - mhealth_v13i1e56666_app1.pdf]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 50 KB - mhealth_v13i1e56666_app2.pdf]

Multimedia Appendix 3
The 4 fields of actions to achieve a more successful implementation of mobile health apps in physicians' daily routines.
[PNG File , 70 KB - mhealth_v13i1e56666_app3.png]

Multimedia Appendix 4
Specification of the use of mobile health apps.
[PNG File , 89 KB - mhealth_v13i1e56666_app4.png]

References

1. Thangada ND, Garg N, Pandey A, Kumar N. The emerging role of mobile-health applications in the management of hypertension. Curr Cardiol Rep 2018;20(9):78. [doi: 10.1007/s11886-018-1022-7] [Medline: 30046971]
2. McLean G, Band R, Saunderson K, Hanlon P, Murray E, Little P, DIPSS co-investigators. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. J Hypertens 2016;34(4):600-612 [FREE Full text] [doi: 10.1097/HJH.0000000000000859] [Medline: 26845284]
3. Kumar N, Khunger M, Gupta A, Garg N. A content analysis of smartphone-based applications for hypertension management. J Am Soc Hypertens 2015;9(2):130-136. [doi: 10.1016/j.jash.2014.12.001] [Medline: 25660364]
4. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990-2015. JAMA 2017;317(2):165-182 [FREE Full text] [doi: 10.1001/jama.2016.19043] [Medline: 28097354]
5. Holstiege J, Akmatov MK, Steffen A, Bätzing J. Diagnoseprävalenz der Hypertonie in der vertragsärztlichen Versorgung ? aktuelle deutschlandweite Kennzahlen. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi) [Article in German]. Versorgungsatlas-Bericht Nr. 20/01. Berlin 2020. [doi: 10.20364/VA-20.01]
6. Neuhauser H, Kuhnert R, Born S. 12-Month prevalence of hypertension in Germany. J Health Monit 2017;2(1):51-57 [FREE Full text] [doi: 10.17886/RKI-GBE-2017-016] [Medline: 37151301]
7. Zhou B, Perel P, Mensah GA, Ezzati M. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. Nat Rev Cardiol 2021;18(11):785-802 [FREE Full text] [doi: 10.1038/s41569-021-00559-8] [Medline: 34050340]
8. Olsen MH, Angell SY, Asma S, Boutouyrie P, Burger D, Chirinos JA, et al. A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on hypertension. Lancet 2016;388(10060):2665-2712. [doi: 10.1016/S0140-6736(16)31134-5] [Medline: 27671667]
9. Ek A, Alexandrou C, Söderström E, Bergman P, Delisle Nyström C, Direito A, et al. Effectiveness of a 3-month mobile phone-based behavior change program on active transportation and physical activity in adults: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(6):e18531 [FREE Full text] [doi: 10.2196/18531] [Medline: 32510462]
10. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
11. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
12. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? an evidence review. J Med Internet Res 2016;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
13. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
14. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
15. Omboni S, Caserini M, Coronetti C. Telemedicine and M-Health in hypertension management: technologies, applications and clinical evidence. High Blood Press Cardiovasc Prev 2016;23(3):187-196. [doi: 10.1007/s40292-016-0143-6] [Medline: 27072129]
16. Parati G, Dolan E, McManus RJ, Omboni S. Home blood pressure telemonitoring in the 21st century. J Clin Hypertens (Greenwich) 2018;20(7):1128-1132 [FREE Full text] [doi: 10.1111/jch.13305] [Medline: 30003701]
17. Mancia G, Kreutz R, Brunström M, Burnier M, Grassi G, Januszewicz A, et al. 2023 ESH guidelines for the management of arterial hypertension the task force for the management of arterial hypertension of the European society of hypertension: endorsed by the International Society of Hypertension (ISH) and the European Renal Association (ERA). J Hypertens 2023;41(12):1874-2071. [doi: 10.1097/HJH.0000000000003480] [Medline: 37345492]

XSL•FO
RenderX

18. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020;8(10):e21759 [FREE Full text] [doi: 10.2196/21759] [Medline: 33074161]
19. Mandrola J. Digital health: should we be concerned? Methodist Debakey Cardiovasc J 2020;16(4):309-313 [FREE Full text] [doi: 10.14797/mdcj-16-4-309] [Medline: 33500760]
20. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
21. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mhealth apps in Australian general practice: qualitative study. JMIR Mhealth Uhealth 2020;8(7):e17447 [FREE Full text] [doi: 10.2196/17447] [Medline: 32729839]
22. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021;9(3):e21795 [FREE Full text] [doi: 10.2196/21795] [Medline: 33661123]
23. Schroeder T, Haug M, Georgiou A, Seaman K, Gewald H. Evidence of how physicians and their patients adopt mHealth apps in Germany: exploratory qualitative study. JMIR Mhealth Uhealth 2024;12:e48345 [FREE Full text] [doi: 10.2196/48345] [Medline: 38231550]
24. Wangler J, Jansky M. The use of health apps in primary care-results from a survey amongst general practitioners in Germany. Wien Med Wochenschr 2021;171(7-8):148-156 [FREE Full text] [doi: 10.1007/s10354-021-00814-0] [Medline: 33570692]
25. Hannan TJ, Celia C. Are doctors the structural weakness in the e-health building? Intern Med J 2013;43(10):1155-1164. [doi: 10.1111/imj.12270] [Medline: 24134174]
26. Etikan I. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Stat 2016;5(1):1-4 [FREE Full text] [doi: 10.11648/j.ajtas.20160501.11]
27. Hennink MM, Kaiser B, Marconi V. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017;27(4):591-608 [FREE Full text] [doi: 10.1177/1049732316665344] [Medline: 27670770]
28. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006 Jun 01;27(2):237-246. [doi: 10.1177/1098214005283748]
29. Fusch P, Ness L. Are we there yet? data saturation in qualitative research. Qual Rep 2015:1408-1416 [FREE Full text]
30. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
31. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]
32. Alsahli S, Hor S, Lam M. Factors influencing the acceptance and adoption of mobile health apps by physicians during the covid-19 pandemic: systematic review. JMIR Mhealth Uhealth 2023;11:e50419 [FREE Full text] [doi: 10.2196/50419] [Medline: 37938873]
33. Muehlensiepen F, Petit P, Knitza J, Welcker M, Vuillerme N. Factors associated with telemedicine use among German general practitioners and rheumatologists: secondary analysis of data from a nationwide survey. J Med Internet Res 2022;24(11):e40304 [FREE Full text] [doi: 10.2196/40304] [Medline: 36449333]
34. Breil B, Salewski C, Apolinário-Hagen J. Comparing the acceptance of mobile hypertension apps for disease management among patients versus clinical use among physicians: cross-sectional survey. JMIR Cardio 2022;6(1):e31617 [FREE Full text] [doi: 10.2196/31617] [Medline: 34989683]
35. Rogers EM. Diffusion of Innovations (5th ed.). Washington, DC: Free Press; 2003.
36. Wu P, Zhang R, Zhu X, Liu M. Factors influencing continued usage behavior on mobile health applications. Healthcare (Basel) 2022;10(2):208 [FREE Full text] [doi: 10.3390/healthcare10020208] [Medline: 35206823]
37. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
38. Wu P, Zhang R, Luan J, Zhu M. Factors affecting physicians using mobile health applications: an empirical study. BMC Health Serv Res 2022;22(1):24 [FREE Full text] [doi: 10.1186/s12913-021-07339-7] [Medline: 34983501]
39. Riggins FJ, Dewan S. The digital divide: current and future research directions. J Assoc Inf Syst 2005;6(12):298-337 [FREE Full text]
40. Cornejo Müller A, Wachtler B, Lampert T. [Digital divide-social inequalities in the utilisation of digital healthcare]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020;63(2):185-191 [FREE Full text] [doi: 10.1007/s00103-019-03081-y] [Medline: 31915863]
41. Cullen R. Addressing the digital divide. Online Inf Rev 2001;25(5):311-320. [doi: 10.1108/14684520110410517]
42. Schroeder T, Seaman K, Nguyen A, Gewald H, Georgiou A. Enablers and inhibitors to the adoption of mHealth apps by patients - a qualitative analysis of German doctors' perspectives. Patient Educ Couns 2023;114:107865 [FREE Full text] [doi: 10.1016/j.pec.2023.107865] [Medline: 37356116]
43. Uncovska M, Freitag B, Meister S, Fehring L. Patient acceptance of prescribed and fully reimbursed mHealth apps in Germany: an UTAUT2-based online survey study. J Med Syst 2023;47(1):14 [FREE Full text] [doi: 10.1007/s10916-023-01910-x] [Medline: 36705853]

XSL•FO
RenderX

44. Obermann K, Brendt I, Müller P. Ein Jahr digitale Gesundheitsanwendungen in der Praxis: Erkenntnisse und Erfahrungen [in German]. URL: https://www.stiftung-gesundheit.de/pdf/studien/aerzte-im-zukunftsmarkt-gesundheit_2021_2.pdf [accessed 2025-03-13]
45. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
46. Giebel GD, Speckemeier C, Abels C, Plescher F, Börchers K, Wasem J, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023;25:e43808 [FREE Full text] [doi: 10.2196/43808] [Medline: 37171838]
47. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
48. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
49. Lang ES, Wyer PC, Haynes RB. Knowledge translation: closing the evidence-to-practice gap. Ann Emerg Med 2007;49(3):355-363. [doi: 10.1016/j.annemergmed.2006.08.022] [Medline: 17084943]
50. Affinito L, Fontanella A, Montano N, Brucato A. How physicians can empower patients with digital tools. J Public Health (Berl.) 2020;30(4):897-909 [FREE Full text] [doi: 10.1007/s10389-020-01370-4]
51. Bruch D, Muehlensiepen F, May S, Wengemuth E, Johannsen O, Reber KC, et al. Digital preventive measures for arterial hypertension (DiPaH) - a mixed-methods study protocol for health services research. Front Cardiovasc Med 2022;9:1089968 [FREE Full text] [doi: 10.3389/fcvm.2022.1089968] [Medline: 36704456]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
mHealth: mobile health
TAM: Technology Acceptance Model
UTAUT2: Unified Theory of Acceptance and Use of Technology 2

Edited by L Buis; submitted 23.01.24; peer-reviewed by P Zhang, N Trehan, M Lopez-Sublet, K Maeshima, M Haug; comments to author 12.06.24; revised version received 30.08.24; accepted 18.02.25; published 28.03.25.

Please cite as:
May S, Seifert F, Bruch D, Heinze M, Spethmann S, Muehlensiepen F
Insights Into How mHealth Applications Could Be Introduced Into Standard Hypertension Care in Germany: Qualitative Study With German Cardiologists and General Practitioners
JMIR Mhealth Uhealth 2025;13:e56666
URL: https://mhealth.jmir.org/2025/1/e56666
doi:10.2196/56666
PMID:

©Susann May, Frances Seifert, Dunja Bruch, Martin Heinze, Sebastian Spethmann, Felix Muehlensiepen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Users' Perceptions and Trust in AI in Direct-to-Consumer mHealth: Qualitative Interview Study

Katie Ryan[1], MA; Justin Hogg[1], BA; Max Kasun[1], BA; Jane Paik Kim[1], PhD

Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Katie Ryan, MA
Department of Psychiatry and Behavioral Sciences
Stanford University School of Medicine
Stanford University
401 Quarry Rd
Palo Alto, CA, 94304
United States
Phone: 1 (650) 723 6180
Email: kryan2@stanford.edu

Abstract

Background: The increasing use of direct-to-consumer artificial intelligence (AI)–enabled mobile health (AI-mHealth) apps presents an opportunity for more effective health management and monitoring and expanded mobile health (mHealth) capabilities. However, AI's early developmental stage has prompted concerns related to trust, privacy, informed consent, and bias, among others. While some of these concerns have been explored in early stakeholder research related to AI-mHealth, the broader landscape of considerations that hold ethical significance to users remains underexplored.

Objective: Our aim was to document and explore the perspectives of individuals who reported previous experience using mHealth apps and their attitudes and ethically salient considerations regarding direct-to-consumer AI-mHealth apps.

Methods: As part of a larger study, we conducted semistructured interviews via Zoom with self-reported users of mHealth apps (N=21). Interviews consisted of a series of open-ended questions concerning participants' experiences, attitudes, and values relating to AI-mHealth apps and were conducted until topic saturation was reached. We collaboratively reviewed the interview transcripts and developed a codebook consisting of 37 codes describing recurring or otherwise noteworthy sentiments that inductively arose from the data. A single coder coded all transcripts, and the entire team contributed to conventional qualitative analysis.

Results: Our qualitative analysis yielded 3 major categories and 9 subcategories encompassing participants' perspectives. Participants described attitudes toward the impact of AI-mHealth on users' health and personal data (ie, influences on health awareness and management, value for mental vs physical health use cases, and the inevitability of data sharing), influences on their trust in AI-mHealth (ie, endorsements and guidance from health professionals or health or regulatory organizations, attitudes toward technology companies, and reasonable but not necessarily explainable output), and their preferences relating to the amount and type of information that is shared by AI-mHealth apps (ie, the types of data that are collected, future uses of user data, and the accessibility of information).

Conclusions: This paper provides additional context relating to a number of concerns previously posited or identified in the AI-mHealth literature, including trust, explainability, and information sharing, and revealed additional considerations that have not been previously documented, that is, users' differentiation between the value of AI-mHealth for physical and mental health use cases and their willingness to extend empathy to nonexplainable AI. To the best of our knowledge, this study is the first to apply an open-ended, qualitative descriptive approach to explore the perspectives of end users of direct-to-consumer AI-mHealth apps.

(JMIR Mhealth Uhealth 2025;13:e64715) doi:10.2196/64715

KEYWORDS

artificial intelligence; mobile health; mHealth; trust; ethics; end users; qualitative study; semistructured interviews

XSL•FO
RenderX

Introduction

Background

Direct-to-consumer mobile health (mHealth) apps present an opportunity for widely accessible health management and monitoring, with the integration of artificial intelligence (AI) offering immense promise to enhance the effectiveness and expand the capabilities of these apps [1]. Indeed, many direct-to-consumer AI-enabled mHealth (AI-mHealth) apps are already in widespread use by individuals seeking to address particular health concerns, obtain personalized insights into their health, promote health-seeking behaviors, and help set and achieve well-being goals [2].

While AI-mHealth apps offer potential for improving proactive health management and monitoring, the rapid pace of AI innovation has outpaced research efforts aiming to facilitate the ethical development and adoption of AI-based apps and maximize their benefits. Recent normative work has articulated concerns related to privacy, informed consent, and bias, among others [3-5]. For example, integral to the training of AI models and the appeal of AI use cases (eg, personalized health recommendations and digital phenotyping based on ecological momentary assessments) is the collection of behavioral data, many times more than what are typically analyzed by mHealth apps that do not use AI, and these data may encompass a wider range of tracking modalities [6,7]. Mobile devices' capabilities to passively collect biometric data (eg, motion, heart rate, and sleep) and contextual data (eg, location and browser activity) have raised novel considerations around data privacy, including users' ownership of these data, consent for data collection, AI capabilities of inferring private information, and potential for surveillance [8-10]. Another challenge for AI-mHealth is unwanted algorithmic bias, which, if inadequately addressed, could perpetuate existing health inequities and even widen health disparities, conflicting with the goals of personalized medicine [11-13]. A further set of considerations relates to how AI use cases may shape how users think and feel about their health status and health behaviors, for example, whether self-monitoring features, depending on how they are framed and implemented, could support or threaten experiences of autonomy and agency [6,14].

These considerations relating to trust, privacy, informed consent, bias, and autonomy have been repeatedly highlighted in the AI-mHealth literature; however, the perspectives on these issues of the stakeholders in direct-to-consumer AI-mHealth who arguably stand to assume most of its benefits and risks (ie, end users) are yet to be adequately explored in the empirical literature [15]. Empirical work assessing trust in *health care provider*-to-consumer conversational agents has shown that trust has interpersonal (eg, trust in companies and health care providers), social (eg, testimonies from loved ones and health care providers), and technological dimensions (eg, trust in the capabilities of the tool and design cues) [16]. A scoping review of studies assessing user views of AI-mHealth apps identified a number of likely barriers to adoption related to user-centered explainability, trust, empathy, usability, privacy, AI accountability, and diversity of model training and test populations [17]. Importantly, the study revealed that direct-to-consumer AI-mHealth user research has been largely limited to quantitative studies that evaluated specific dimensions of user experience collected via structured questionnaires and performance metrics generated by the apps, which tend to rely heavily on researchers' a priori assumptions about what the major user considerations are.

The small body of published qualitative work has used interviews of actual or anticipated users of direct-to-consumer AI-mHealth apps to shed light on their attitudes and perspectives, yielding several ethically salient insights. An interview study of university students by Nadarzynski et al [18] explored their attitudes toward hypothetical conversational AI apps. The study identified several sources of hesitancy related to AI interactions (a lack of empathy and understanding, especially in mental health use cases) and accuracy and quality concerns as well as perceived advantages around anonymity, convenience, accessibility, and guidance toward relevant health services. In semistructured focus groups with young people about their attitudes toward AI-mHealth apps, Götzl et al [19] found nuanced attitudes toward data sharing and safety; participants reported willingness to share their data provided that they would truly support their health interests and expected transparency about how their data were being used, discretion in requesting more sensitive data, and user controls for personalization features. In a report by Tsai et al [20], focusing on users' needs related to explainability in AI-enabled symptom-checking tools, interviewees reported that they often felt confused by the order and content of the differential diagnosis questions asked by these apps and hesitated to trust the accuracy of their results due to low transparency about their reasoning procedures and underlying data.

Objectives

As AI advances, direct-to-consumer AI-mHealth apps may create more opportunities to improve population-wide health awareness, disease prevention, and well-being. To realize these aspirations, it is essential to engage with the end users whom the tools are intended to help and develop a comprehensive picture of the views of the end users of direct-to-consumer AI-mHealth apps, including their own reservations or anxieties that may remain overlooked and pose barriers to establishing trust and fostering adoption. Although previous qualitative study designs have focused on exploring users' views on specific use cases and topics of ethical importance, no single study has examined their ethical perspectives on AI-mHealth broadly. Therefore, in this paper, we set out to describe, using semistructured interviews and a qualitative descriptive approach, the topology of ethical considerations of individuals who have experience interacting with mHealth apps with respect to AI-mHealth.

Methods

Study Design

The findings presented in this paper were collected and analyzed as part of a larger study that aimed to identify and describe ethically laden sentiments and anticipated issues of AI well-being apps' use in work environments from the

XSL•FO
RenderX

perspectives of current and potential future end users. This study addressed a supplemental aim to an ongoing study about the ethics of AI use in medicine (NCATS R01-TR-003505) [21]. To accomplish this aim, we conducted semistructured interviews with existing users of AI-mHealth apps (N=21) to identify and describe their perspectives relating to ethical dimensions associated with the use of such apps, specifically in the workplace.

A qualitative descriptive approach was applied to the study design, as this approach emphasizes the description and analysis of stakeholders' experiences and perceptions using language and ideas that emerge directly from the stakeholders themselves [22,23]. While other methods of qualitative research and analysis prioritize the development or advancement of theory, qualitative description allows for the emergence of new knowledge that may not be readily available from theoretical deduction and is grounded in the experiences of participants.

Recruitment and Participants

Adults aged ≥18 years who reported current or previous use of an mHealth app were eligible to participate in this study. Participants were recruited for this study via an electronic advertisement posted to the Stanford Psychiatry Department's *Currently Recruiting Studies* web page. The advertisement described the study and contained a link to a web-based screening survey and was posted from August 1, 2022, through August 29, 2022.

A total of 32 individuals completed the screening survey; 31 (97%) individuals met the eligibility criteria for participation. All eligible individuals were contacted via email to schedule a Zoom (Zoom Communications, Inc) interview; 21 (68%) of these 31 eligible individuals scheduled and completed an interview. Interviews continued until the minimum target sample size of 20 was exceeded. At that point, interviewers jointly determined that content saturation had been reached, as no new substantive themes were emerging from additional data collection [24].

Procedures

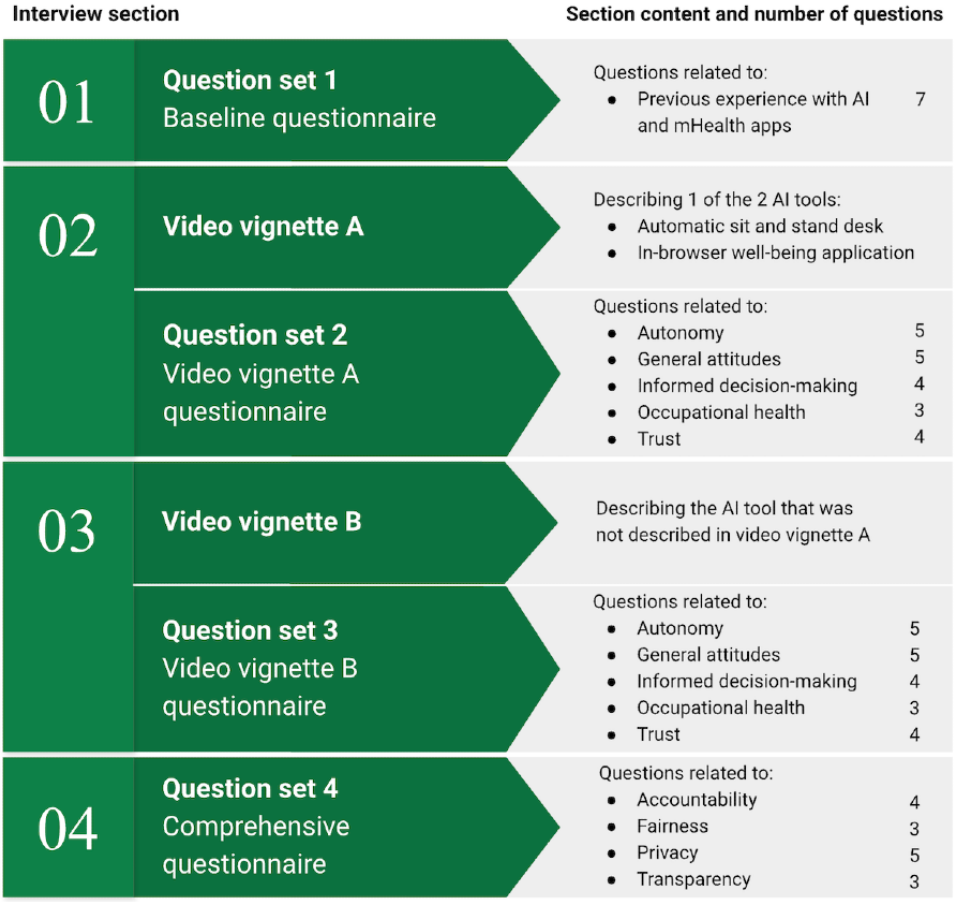
Semistructured interviews were used to facilitate discussion that encouraged participants to reflect on their experiences and opinions and allow them the flexibility to focus on topics that they found the most important. This approach allowed interviewers to ask unscripted follow-up questions where relevant or skip questions that did not apply in the context of the conversation. By conducting interviews via Zoom, participants outside of the local area were able to participate in the study.

The semistructured interviews contained 4 sections: a baseline question set, 2 video vignettes and associated question sets, and a comprehensive question set (Figure 1). In order to stimulate discussion of ethical considerations among a population that may not have encountered AI use cases in mHealth previously, 2 video vignettes demonstrating the use of example AI-enabled workplace health interventions were used. The presentation of these video vignettes was followed by ethically salient questions related to the presented technology (eg, questions relating to understanding, trust, and autonomous decision-making).

Interviews were conducted between August 2022 and September 2022 by 1 of our team's 3 trained interviewers. The interviews lasted 75 minutes and 35 seconds on average (SD 24 minutes, 28 seconds; range 35-131 minutes) and were audio recorded for the purposes of transcribing.

XSL•FO
RenderX

Figure 1. Outline of the semistructured interviews. AI: artificial intelligence; mHealth: mobile health.



Data Coding and Analysis

Auto-transcriptions created by Zoom were reviewed, edited, and deidentified by a member of the research team. Data analysis was guided by the principles of qualitative content analysis [25]. This inductive approach, in which transcribed data are broken down into descriptive units that are named and sorted based on their content, allows for the emergence of codes and themes directly from the dataset [26]. As we were primarily interested in identifying and describing ethical considerations of AI-mHealth users that have been underexplored in the existing literature, this inductive approach was selected over a deductive approach, which would have relied more heavily on a priori assumptions about user priorities and perspectives.

An initial round of open coding was performed on each transcript by 2 authors. The authors reviewed their assigned transcripts independently, identified interview content that they felt was substantive, and suggested descriptive codes for this content. The authors then met as a group to discuss the descriptive codes that emerged during open coding and compare the content associated with each code. They collaboratively identified topics that recurred throughout the transcripts and established preliminary code names and definitions for these topics. The transcripts were then rereviewed by the same 2 authors using the preliminary codes and definitions as a guide. At the completion of this phase of intermediate coding, the authors met as a group to compare the coded units, further refine the code names and definitions, and draft the final version of the codebook.

The final version of the codebook contained 37 codes derived directly from the content of the interviews describing baseline characteristics and tool-specific, personal, and contextual factors influencing participant attitudes toward AI-mHealth. The transcripts and codebook were then uploaded to NVivo (version 1.0; Lumivero) for final coding, which was completed by a single research team member and then reviewed for consistency by a different team member.

Data analysis began upon the completion of coding. A conventional qualitative analysis approach was used to guide the analysis [25,26]. At the completion of final coding, all authors met as a group to analyze the content and meaning of the coded units and develop categories and themes that described the associations between units. This process resulted in the identification of 3 “buckets” detailing different aspects of participants’ experiences and attitudes toward well-being AI (Table 1). For the purpose of providing an accurate and in-depth analysis of the content of these interviews, this paper will only address the codes, categories, and themes associated with bucket 1. Those associated with buckets 2 and 3 will be analyzed in future publications.

XSL•FO
RenderX

Table 1. Inductive coding groups and codes describing users' perspectives toward well-being artificial intelligence (AI; N=21).

The codes used in the development of the bucket	Interviews cited, n (%)
Bucket 1: user attitudes toward AI-enabled mHealth[a]	
Data collection	21 (100)
Data use and sharing	21 (100)
Availability of information	20 (95)
Experience with mHealth apps	20 (95)
Explainability of the model	20 (95)
Regulation	20 (95)
Research participation	20 (95)
Potential benefits	19 (90)
Accountability	18 (86)
Attitudes toward AI	17 (81)
Potential risks	17 (81)
Trust in institutions or entities	16 (76)
Understanding of and familiarity with AI	13 (62)
Potential of AI	12 (57)
Value	11 (52)
Perceived ubiquity of technology	9 (43)
Financial or economic factors	8 (38)
Influence on the mind or behavior	3 (14)
Bucket 2: ethical concerns relating to well-being AI in workplace settings	
Workplace-related factors	21 (100)
Trust in the tool	20 (95)
Bias	18 (86)
Data security and privacy	18 (86)
Autonomy and agency	16 (76)
Social identity and stigma	15 (71)
Trust in intentions	10 (48)
Personal health	9 (43)
Personal qualities or traits	9 (43)
Life stage	7 (33)
Social or historical context	6 (29)
Accessibility	3 (14)
Bucket 3: user attitudes toward features of well-being AI tools in workplace settings	
Degree of user's control	21 (100)
Intrusiveness	19 (90)
Personal context	18 (86)
User experience	18 (86)
Optimization	17 (81)
Type of intervention	17 (81)
Physical context	3 (14)

[a]mHealth: mobile health.

XSL•FO
RenderX

Ethical Considerations

This study obtained human participant research approval from the institutional review board of Stanford University on June 21, 2022 (58118). A copy of the institutional review board–approved informed consent form was emailed to all potential participants before the interview. Interviewers then reviewed and explained the content of the consent form to all participants at the start of the interview using the screen share function on Zoom. All participants signed an electronic consent form before the start of the study-related procedures. Participants were able to opt out at any point during the consent process or during the research interview. Identifiers were removed from the dataset during the transcription process, before analysis. Participants were compensated in the form of a US $50 electronic gift card.

Results

Overview

The demographic information of participants can be found in Table 2. The final cohort of participants included young adults, aged between 20 and 36 years, and was primarily composed of individuals whose reported sex was female (15/21, 71%) and whose reported race was Asian (8/21, 38%), Black (5/21, 24%), or multiracial (4/21, 19%). In their open-ended responses, participants reported having experience using AI-mHealth apps, such as Apple Health, Headspace, Calm, and Flo.

All participants in this study commented on considerations related to bucket 1: user attitudes toward and understanding of AI-mHealth (Table 2). These considerations primarily arose from responses that emerged in parts 1 and 4 of the interview question set (ie, the baseline questions at the start of the interview and the comprehensive questions at the end of the interview; Figure 1). In discussions relating to attitudes and understanding of AI-mHealth, 3 major categories, each with 3 related subcategories, emerged (Figure 2). The content of these categories and subcategories is described in detail in the subsequent sections.

XSL•FO
RenderX

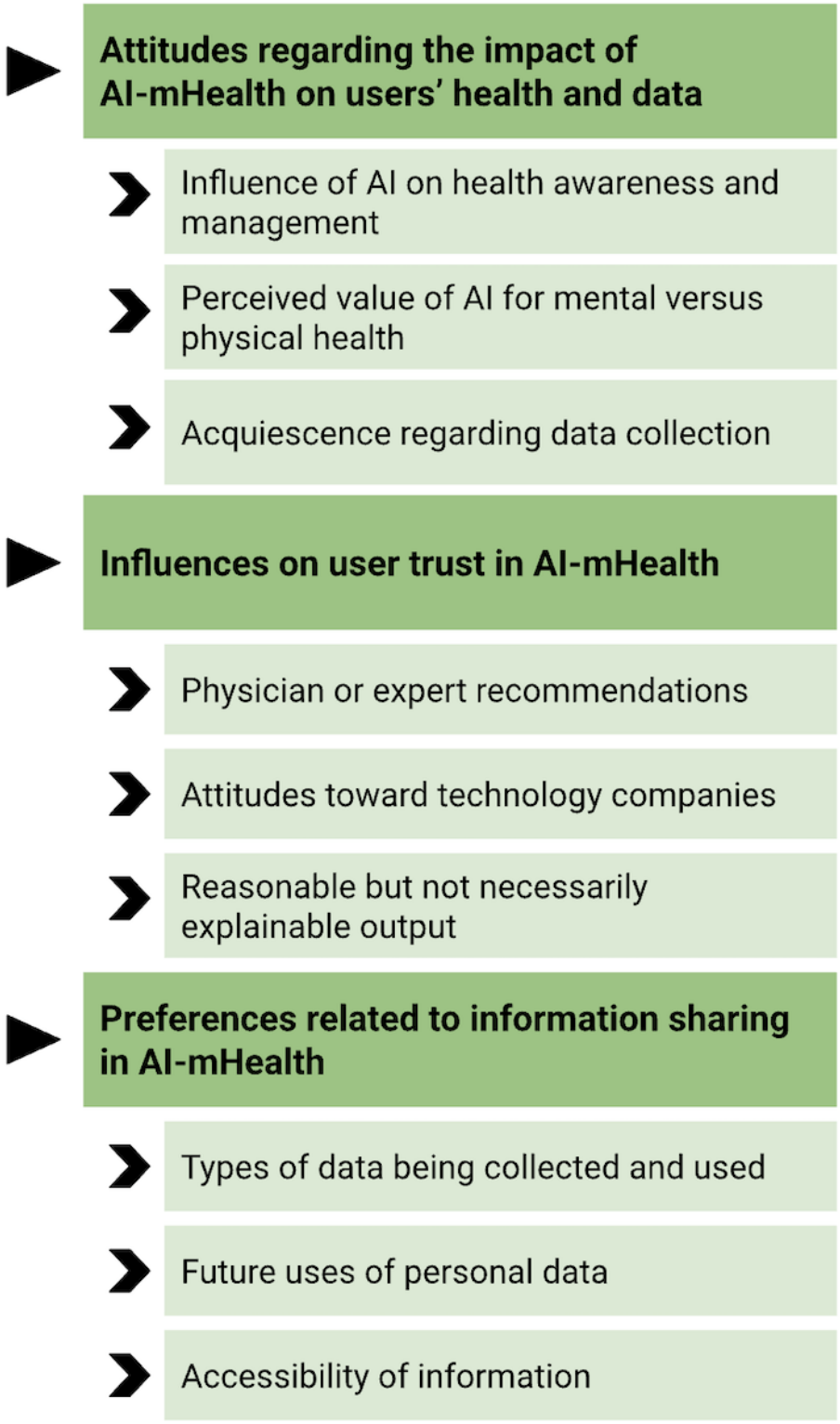
Table 2. Participant demographics (N=21).

Demographics	Values
Age (y)	
Value, mean (SD)	26.8 (4.98)
Value, median (range)	26.0 (20.0-36.0)
Missing, n (%)	2 (10)
Sex, n (%)	
Female	15 (71)
Male	6 (29)
Race, n (%)	
Asian	8 (38)
Black or African American	5 (24)
White	3 (14)
Multiracial	4 (19)
Other	1 (5)
Ethnicity, n (%)	
Hispanic or Latino	1 (5)
Not Hispanic or Latino	20 (95)
Education[a], n (%)	
High school degree or equivalent	1 (5)
Some college; no degree	5 (24)
Bachelor's degree	6 (29)
Master's degree	4 (19)
Doctorate	1 (5)
Associate, technical, or vocational degree	3 (14)
Missing	1 (5)
Employment	
Employed full time (≥35 h per wk)	10 (48)
Employed part time (up to 35 h per wk)	4 (19)
Unemployed and currently looking for work	1 (5)
Student	8 (38)
Stay-at-home parent	1 (5)

[a]Participants were allowed to select >1 answer for the education question; thus, the total percentage may be >100%.

XSL•FO
RenderX

Figure 2. Categories and subcategories describing users' attitudes toward artificial intelligence (AI)–enabled mobile health (AI-mHealth) apps.



Attitudes Regarding the Impact of AI-mHealth on Users' Health and Data

Influence of AI on Health Awareness and Management

Participants expressed positive sentiment toward the potential for AI-mHealth apps to support health awareness, health maintenance, and preventative care. Several users acknowledged the value of AI-mHealth in encouraging users to take steps to manage their own health, with one participant noting that "Sometimes we can't acknowledge the moment ourselves" (P101) and another noting that a major benefit of the incorporation of AI into mHealth is its ability to "[help] you know when you should be thinking about your mental health and wellbeing" (P130).

While improved health awareness was perceived as a benefit of AI-mHealth apps, participants also described ways in which they felt such technology could undesirably influence their health behavior. While some participants referenced having

XSL•FO
RenderX

"blind trust" (P096 and P117) in AI-mHealth to manage aspects of their health, others emphasized that AI-mHealth apps should act as a supplement to an individual's health management, with one user noting, "Any AI use should just be a healthy sidestep to us taking control of our mental health versus taking over and trying to do it for you" (P130). Another described a desire to feel in control over the management of their own health and explained how AI-mHealth could interfere with feelings of autonomy:

> *Part of me doesn't like to have too many orders. I like having the ability to choose a bit...When it comes down to health, my body and my movements, that's one of the main things that we can control, that we should control. And someone telling me to "Do this right now," it's just eh. The way I view health is as an investment in myself, but I also want to do that, I want to invest in my health. I want it to be voluntary in that sense. And just having [technology do that instead] makes me feel like a robot, just not completely autonomous.* [P101]

Perceived Value of AI for Mental Versus Physical Health

When discussing their attitudes toward AI-mHealth apps, users differentiated between the domains of mental health and physical health. Multiple users noted that physical health could be more easily quantified than mental health, thus making mental health more complex and difficult to measure. Users commented that "Physical health is more quantitative and mental health is more qualitative...there is so much more to unpack from it than just a number for physical health" (P101); "Mental health is, more often than not, harder to measure than physical health" (P104); and "Technology can compensate for physical health, as opposed to, mental health can become so complicated" (P130). Several users noted that, because of mental health's resistance to measurement and quantification, they would be more skeptical of AI-mHealth recommendations for mental health and more accepting of recommendations for physical health:

> *[AI] could work for certain things but...It maybe needs more work on other aspects, especially when it comes to emotions and people's decisions. In math, it totally works. But if it can tell you when you're going to have this episode of sadness or happiness, I'm not very sure about that.* [P118]

> *Physical health has been a lot more black and white. When I get a suggestion for my physical health, I don't question it as much. I am more likely to accept it...I am less hesitant about software that is telling me to do a physical activity.* [P101]

Users further emphasized that they would desire the ability for greater personalization, customization, or control of AI-mHealth apps that target mental health, with the aforementioned user noting the following:

> *With the mental health app, I want it to be more customizable. I feel like it requires more factors to consider before suggesting or working.* [P101]

Acquiescence Regarding Data Collection

Users expressed a reluctant acceptance toward the amount and type of personal data collection that occurred while using AI-mHealth apps. Although many described discomforts regarding the volume of data collection, they acknowledged that it felt "unavoidable" due to how deeply ingrained this process is in modern technology use (Textbox 1).

Textbox 1. User quotes regarding data collection.

- "[Sharing data] feels unavoidable to me unfortunately. I don't feel great about this but it feels like there's some of my data that I will just have to give up." [P097]
- "I feel like our generation is so used to using applications and we don't really question what happens with [our data]." [P101]
- "Ideally I don't want that to happen, but I think in reality they already have access to all of our data." [P105]
- "I am very into providing my data for research purposes, because I know that [data collection] is still happening anyway." [P106]
- "I already know companies use my data to sell, but that's just also inevitable." [P114]
- "I honestly just think it's so much a part of everyday life. I would not say I am 100% comfortable with all the data collection that is happening." [P124]
- "I don't particularly want data that isn't already out there to be going out there...I deeply hate that expectation of handing over [my] data about what I am doing." [P131]

Several users admitted a conflict between their theoretical preferences relating to data collection and sharing and their actual use of technology, with one user summarizing the following:

> *I feel like in theory I care a lot about how my data is being used, what type of data, who it's being sold to. In practice, I feel like I don't really know what kind of data [the apps I use] collect.* [P097]

A different participant acknowledged that, while data sharing likely would not incur any real risks, their negative feelings toward these practices were based "on principle" and their desire to feel as if they had control over information about themselves:

> *When I think about data privacy, it comes more down to the principle of wanting my information to stay private and not feel like I am being tracked all the time. Because in reality when I think about it, nothing terrible seems like it is going to happen if my information is going to be used other than them telling me what to buy. It is not a huge issue, but it is almost the principle of the thing, and wanting to feel like you*

XSL•FO
RenderX

> *have ownership over your day-to-day and it not being up to outside controlling factors.* [P130]

Influences on User Trust in AI-mHealth

Physician or Expert Recommendations

Multiple users agreed that endorsements from physicians or guidance from health or regulatory organizations would increase their trust in an AI-mHealth app and their willingness to use it, with one stating that "having some kind of affiliation with an actual medical professional does actually instill a little more trust and make you more willing to share your information" (P130). Several users specifically mentioned that they would be more comfortable sharing personal health data with apps that were recommended to them by a trusted physician. One user described the following:

> *Depending on the original intent of the app, if it greatly benefits my health, if my therapist told me to use this specific app and then I saw that it had great improvements in my health because it tracked my specific triggers or mental health and it took record of everything, I don't think I'd question the data that is being collected.* [P114]

One user specifically referenced the Food and Drug Administration approval of AI-mHealth apps, commenting on how the regulatory process of Food and Drug Administration approval gave them confidence and had the potential to give other users more trust about using such apps:

> *One of the applications I used is a FDA approved method for tracking something. That gives me a lot more trust. For AI situations where it is making recommendations based on your health, it should be like any other thing. Drugs that increase your mental wellbeing, those are FDA approved. It should go through the same scrutinizing process and the same regulatory process as well. I think the public would be more willing to use it too.* [P096]

Attitudes Toward Technology Companies

Users expressed mixed levels of trust in companies that develop AI-mHealth apps. Several users saw the reputation of large companies as a reason to have trust in being a consumer there because these companies were "reputable and reliable" (P114). Two participants specifically referenced feeling more comfortable sharing their data with larger, established companies, noting that "I would feel safe...if it's a reliable company like Google" (P127) and "A big company that people would know, I would feel comfortable sharing the information" (P130). Although most users did not provide additional details about why they trusted larger companies, one participant acknowledged the advantages that large companies have that may keep user data safer:

> *The paradoxical part is I'd rather almost want it to be Amazon other than somebody I don't know, because I know that Amazon is large enough and asks the employees to do a better job at keeping my data safe from outside people.* [P131]

Others described mistrust in technology companies based on their perceptions about how these companies collect, use, distribute, and profit from users' data, with one user stating the following:

> *They use all this data to benefit their company and they have all this money, but none of the users receive it...I feel like they are just using [our data] for personal selfish gain without it really benefiting anybody except their own company.* [P114]

Another user cited the history of such companies benefiting from the collection of personal user data as evidence against trusting them to genuinely work toward creating benefit for their consumers:

> *All the mental health applications, we can see that they want to earn money. I think the consumers that are looking for mental health services in any kind of different formats, including applications, are essentially looking for a little more authenticity in their services because it is mental health care. If the consumers can see they profit over us, once they see that, it's like, "I don't really trust this service."* [P106]

Reasonable but Not Necessarily Explainable Output

Multiple users indicated that understanding the rationale behind an AI-mHealth app's recommendation was not crucial to their trust in the app. Users acknowledged the complexity and black box nature of AI reasoning and did not think that having increased access to weights or other data used by the AI would be beneficial to their own trust or understanding. Instead, there was agreement that the perceived reasonableness of an AI recommendation had a greater influence on their trust and willingness to implement a recommendation (Textbox 2).

Textbox 2. User quotes regarding the explainability of artificial intelligence output.

- "I can appreciate the fact that a complex algorithm is a black box, but as long as I know that their training sample was diverse and big enough and their results are interpretable and make sense I don't mind as much. I think it's fine that they don't know what is happening." [P97]
- "I've learned that the neural network is a black box. You can only know what you add in and what you get out of it. It's still a mystery to everybody. I have accepted that with the AI tools that I have used in the past. If the conclusion isn't outrageous and I think it's useful, I'll do it." [P104]
- "I believe the rationale for the AI is based on a bunch of different factors and I don't think I'm interested in learning which factors weigh a little bit more than the other." [P105]
- "It's the zeros and ones in the computer. How can I interpret that and equate that with the human thought process? It's not possible." [P124]
- "I honestly don't care as long as the actions and recommendations don't seem ridiculous, I would not care about the exact reasoning." [P126]

XSL•FO
RenderX

In addition, several users conveyed tolerance for the "black box" limitations of AI recommendations, noting that humans often cannot explain exactly what drives their intuitions and judgments, and extending this idea to AI:

> *There are some things that we think about and we can't really explain our reasons for it...We don't really know how to quantify certain things, so I can't completely blame the AI for having a black box concerning my data either.* [P114]

Multiple other users similarly compared the limits of explainable AI to the covert processes of the human brain:

> *It is sometimes really hard for even the software engineers to design the specific algorithm to know exactly what is going on in that thought process. It is really hard for even us to walk someone through, "I thought about this and then this thought led to that, because of the specific factors..." It is really hard even for ourselves to talk that through.* [P106]

> *[Humans] can't always pinpoint exactly where something comes from, where an idea comes from, where feelings come from. If I'm being empathetic to the AI, I would be like, "Yeah I get that. You don't know where this suggestion is coming from but you feel like it's correct." I would probably trust that.* [P115]

Preferences Related to Information Sharing in AI-mHealth

Type of Data Being Collected and Used

Several participants clarified that, while they felt comfortable with not knowing *how* an AI app makes a decision, it was important for them to understand *what personal* data were being collected and used in such decisions, as described by these participants in the following statements:

> *The initial information they are collecting I would want to know. How those decisions are made, I definitely can understand that that would be too complicated to be explained. That I don't really mind.* [P130]

> *It wouldn't be too much of an issue if they can't explain how it's making the decision, but at least knowing what is being considered in the process would give me some peace of mind.* [P101]

One of these users wanted to know, "What information do they need from me to decide?" (P130), while other users summarized, "I would love to know what kind of data is being taken in" (P127) and "I would really appreciate knowing what behaviors and responses they are using to create that output...I think people 100% deserve to know every piece of data they are using and knowing what that is for" (P097). Several others described how being provided with information regarding the type of personal data that are being collected and used in an algorithm's decisions would increase their willingness to consistently use an app (P096, P117, and P120).

Future Uses of Personal Data

Many users expressed a desire to have access to information regarding the potential future uses of data collected via their use of AI-mHealth apps. Specifically, they wanted to know if personal data collected by an app would be sold to or shared with other corporations or entities (Textbox 3).

Textbox 3. User quotes regarding future uses of personal data.

- "I would like to know if the applications that I'm sharing data with have plans to sell that data." [P097]
- "[I want to know] what it's being used for. And I think it should only be used for the optimization and my de-identified data should be used for the training of the model to make it more accurate, better for future uses." [P104]
- "I really would like to know what data are being collected and how exactly they are being used. Because I don't want my health data to be floating around." [P106]
- "I would love to know who is the end user of my data. Has my data been sold to another person?" [P111]
- "I really want to know specifically how they use it. Are they selling my data, what agencies or companies have access to that? What is the background of these companies? Are they just going to spam email me, put it on the dark web? I really want to know the security of that." [P114]
- "How are you using my data, is it going anywhere where it's identifiable?" [P120]
- "We are hearing stories that they are selling data. I would love to know if the data that is collected, is it being sold to hospitals or is it being sold to the black market? If it is sold to a hospital for further research then I will be okay with it, but if the app is using it for other malicious activities, that would be very wrong." [P127]

Multiple users discussed how the collection of highly sensitive health data increased their level of concern about potential sharing or selling of the data to third parties and their desire to know whether this was occurring. Three users specifically mentioned menstruation-tracking apps as examples of health apps where concerns of data being shared or sold involved additional or elevated risks (P097, P114, and P115). One participant argued that exposing sensitive health data to third parties could lead to serious social, legal, and medical harms:

> *[I would want to know] who has access to that data. I read some discourse about the banning of abortion and then they were telling people to delete period tracking applications because if they're able to detect when you're on your period or trying to get an abortion, you could legally be arrested for that...If [the company is] able to sell that data or give the government access to that, it could potentially put certain people in danger.* [P114]

XSL•FO
RenderX

Accessibility of Information

Users expressed a range of suggestions about how companies might better communicate to users about the collection and use of their personal data. Multiple users described the lack of engagement that users have with traditional terms and conditions pages, noting that "Many people just don't read them, and I am one of those people usually" (P115); "For me, and probably a lot of people, [when] reading those small text notices, I always push 'agree' and move on" (P096); and "A lot of people blow through that stuff and are like, 'Yes, yes, accept'" (P120). The length and complexity of these agreements were frustrating to users, with one commenting that "I just don't like how a lot of us aren't really informed how our data is being used, and then they make it so complicated to the point where we don't really look through it" (P114).

Several users described wanting to receive clear and digestible information regarding the risks and benefits of AI-mHealth apps as a more straightforward way to understand considerations related to their data. Two users described that, "It would be good to know the risks and benefits" (P096) and "[I would like] an explanation of pros and cons, rather than a sales pitch-type view of it" (P124). Other participants commented that companies should provide a "condensed version" of the terms and service agreement that "just gives the bullet points" (P101), with one participant noting how a shift toward providing more accessible information could demonstrate a company's concern for and efforts to safeguard users:

> *Privacy policies are often not very accessible. I want to see, not just do you provide the information, but do you provide it in a way that is clear with its language which to me also expresses that you care about what you are doing with the data? [To see] if you take the time and energy to make something that people understand because you care about conveying that information.* [P131]

One participant further expanded on their request for more accessible information regarding the risks, benefits, and data policies of an AI-mHealth app and posed an analogy to the informed consent process they completed for this interview study:

> *I want the app to highlight the very important data that they will be collecting and to put that into layman's terms so that I can just understand it very easily...I am thinking of something like the informed consent procedure that you just did to me. The document itself looked pretty intimidating...[but] you highlighted the components I really needed to know as a study participant...If the app could do something like that...if it can be that transparent and simple, that will definitely boost my trust and credibility as the app's consumer.* [P106]

Some participants suggested technology- and media-based alternatives to the current text-based terms of service agreements that they felt would provide information relating to risks, benefits, and how personal data are collected and used in a more digestible and accessible format. Two suggested videos as an alternative to text (P096 and P115), with one noting that this format could be useful for "forcing" users to engage with the information:

> *If it were presented in a really digestible way where you were basically forced to read it or hear it if you use the app, like a video where you couldn't move on unless you play the whole thing.* [P096]

Multiple other participants commented that apps could have associated websites where users could look up information if questions arose through their use of the app, with one noting the following:

> *It would be great to have a resource that is just live all the time. A website where you can access that information and then contact them if you have questions that aren't answered on their website.* [P120]

Discussion

Principal Findings

In this semistructured interview study, users expressed attitudes and concerns about a number of ethical considerations related to the use of AI-mHealth apps, including considerations related to autonomy, privacy, trust, transparency, and information sharing. The major findings are discussed in detail in the subsequent sections.

The Value of AI-mHealth: Perceived Limitations of Apps for Mental Health

Interviewees expressed comfort with the idea of using AI-mHealth apps to supplement their health management, citing their value in organizing and tracking health data and providing users with prompts or knowledge that could benefit them. Several users were adamant that AI-mHealth should supplement rather than replace health care services and preferred technologies that encouraged informed and autonomous health decision-making, as opposed to more assertive recommendations, which they felt could be perceived as orders that could ultimately detract from health awareness. This finding aligns with previous research by Almourad et al [27], which found that users' feelings of increased self-awareness and a sense of control over the use were acceptance factors in their use of mHealth technologies.

Users in this study notably distinguished between the potential value of AI-mHealth apps designed for physical health and those intended for mental health. Strikingly, some were less confident that AI could be as effective for mental health use cases compared to physical health use cases. They perceived mental experience as highly personal, intimate, and resistant to observation and measurement (admitting also the limits of human introspection) and therefore did not feel that AI-mHealth apps could "know" them well enough to make accurate inferences that lead to helpful predictions or guidance. These attitudes deviate from previous findings among avid users of AI mental health chatbots, who report high engagement and trust in such apps [28], and suggest that less experienced or laypersons, such as those in our study, may be more hesitant to accept predictions and guidance that rely on AI inferences

XSL•FO
RenderX

regarding more subjective and phenomenal mental processes (eg, underlying emotions or affective states) compared with apps that target physical health. The design and implementation of AI use cases for mental health will benefit from additional empirical inquiry to help understand how users with less experience think AI can benefit this domain and develop a clearer picture of ethically salient factors that may influence adoption.

The Collection and Sharing of User Data: Acquiescence and Mistrust

There was a general admission among interviewees that, as users of modern technology, they often must concede control over their data and their future uses. They acknowledged discomfort with the fact that personal data are systematically collected through their devices, partly to help advance corporate and business interests, but expressed acquiescence about their ability to prevent or change this reality. These attitudes have been documented frequently in the technology literature, having been described as privacy cynicism [29], digital resignation [30], privacy apathy [31], and privacy fatigue [32]. Notably, participants in our study were young adults whose interactions with technology have likely always involved concessions relating to privacy, possibly resulting in a reluctant acceptance that the use of technology by default incurs the loss of certain privacies. Future studies assessing feelings of acquiescence across technology users of different age groups would be of interest.

In addition, participants in this study described how the pervasiveness of data collection and sharing influenced their trust not just with specific technologies but also with the corporations or entities associated with them. Some expressed greater trust in large companies that they believed were better suited to protect their data, while others asserted no trust in these same companies due to their historical collection of user data. This, too, is consistent with findings from the technology literature, which found mistrust in corporations that users associate with data sharing. As was described by Lutz et al [33] in their survey study, the potential benefits of an individual technology or app may be trumped by "mistrust and powerlessness in relation to those platforms ostensibly providing the infrastructure." Although a recent user-review analysis found an overall high degree of public trust in individual AI-mHealth apps for mental health, our results, along with findings from the technology literature, indicate that there are users whose attitudes toward AI-mHealth are influenced by their existing trust, or lack thereof, in large technology companies [34]. More research is needed to assess how the systematic collection and distribution of personal data have contributed to users' trust or mistrust in technology companies and the impact that this may have on their trust in and adherence to AI-mHealth apps.

Users' Desire for Information: Understanding What, Not How

In a previous interview study with users of AI-mHealth apps representing an array of mHealth use cases, participants reported desiring more straightforward and precise information about the apps, including how they used AI, how the AI-enabled features worked, and how well the AI might solve the health needs the apps purported to address [35]. Participants in this study consistently expressed similar high-level concerns; however, they were less concerned with *how* an AI-mHealth app generated its outputs than they were about *what* types of data it collected and used to generate its feedback or guidance. A related, frequently expressed worry was whether their personal data could be used by additional parties or for other purposes in the future. These findings suggest that AI-mHealth users may put less stake in understanding the underlying technology and reasoning procedures driving an app and have more reservations about the collection and use of their personal data.

Experimental studies with users of AI tools in health care and other sectors engaging in AI innovation have shown that enhanced explainability is not always favorable to user understanding, decision-making, or trust [36-38]. Our results contribute much-needed stakeholder data to this literature and reveal that explainability tools or technical explanations of algorithms and models may not be crucial to secure users' trust and facilitate adoption, provided that they have access to resources that assuage personal data–related concerns.

Beyond noting that understanding an AI's reasoning procedure was not essential to their sense of trust, several participants in this study strikingly related black box algorithms to human intelligence and decision-making processes. Specifically, they described how both algorithmic and human reasoning involve processes that are obscured from the observer. A previous interview study by Gkinko and Elbanna [39] documented the tendency of people to extend empathy to AI chatbots into their workplace environments; however, the association of black box AI decision-making and human intelligence appears to be a novel finding that has not been previously documented among users of AI-mHealth apps. This association provides a reminder that humans are accustomed to making decisions based on mental processes that are not necessarily able to be articulated or explained. Taking action based on incomplete or obscured information is not a novel experience for humans, and as our participants acknowledged, this willingness could possibly extend to their interactions with AI.

Informing Users: Concerns and Recommendations

Users in this study repeatedly acknowledged that they do not read terms and conditions agreements, which they felt were overly lengthy and not designed to inform users. Participants offered a range of alternatives for how they would prefer to be informed about the use and future distribution of their personal data. User-friendly summaries of risks and benefits of an app were suggested, as were videos or accessible websites. As highlighted by several participants, the inclusion of easily digestible lists or videos regarding risks and benefits of use is a fairly minor addition to an app that could be incorporated before or after existing terms and conditions pages. Comments from our participants indicate support for concepts, such as the "Nutrition Label for Privacy" proposed by Kelley et al [40]. As AI-enabled technology that collects increasing amounts of user data becomes more accessible to the public, offering users clear and comprehensible information regarding the use and sharing of their data may influence their trust in a tool and willingness to incorporate its recommendations.

XSL•FO
RenderX

When discussing how information sharing in digital health could be improved, one participant referenced the informed consent method that is typically used in medicine and research. Informed consent forms share similar challenges to terms and conditions documents, in that they often include content that is not necessarily relevant to participant understanding or decision-making, but that instead is required for the purpose of protecting the interests of the funding or managing institutions. Studies testing shorter consent forms have found that participants were more likely to fully read shorter forms and that those exposed to shorter forms comprehended more of the information, indicating that the length of the terms and conditions pages could be a primary factor in users' tendency to bypass them without engagement [41]. Furthermore, the accompanying informed consent process in which the information from informed consent forms is distilled into accessible language that highlights the risks and benefits is often positively rated by participants, possibly indicating that terms and conditions pages that are followed by summaries highlighting the information relating to specific risks and benefits may be well-received by users of AI-mHealth apps [42,43]. Informed consent processes are established within medical practice and research and could provide a useful road map for AI-mHealth apps, which may require both templated legal disclosures and access to product information that encourages participant engagement and understanding.

Limitations

This study has several limitations relating to its sample. Most notably, all participants were aged between 20 and 36 years at the time of the interview. It is documented that attitudes toward technology are influenced by age and life stage, so the findings from this study should therefore not be generalized to older populations [44,45]. Furthermore, our sample primarily consisted of individuals who identified as female and belonged to a racial or ethnic minority group. As we did not specifically intend to prioritize the interviewing of these populations, the attitudes documented in the interviews and analysis should not be assumed to be representative of all users of mHealth apps. In addition, our findings are limited by the fact that interviews were completed in 2022, before the first release of ChatGPT, which represented a major shift in the public's perceptions toward AI and AI-enabled technologies.

Conclusions

In this qualitative study using semistructured interviews, users expressed their attitudes and concerns about a number of ethical considerations in the use of AI-mHealth apps, including considerations related to autonomy, privacy, trust, transparency, and information sharing. Notably, users distinguished between the potential value of mental and physical AI-mHealth apps, cited existing feelings of loss of control and ownership of their data in everyday use of technology that appeared to influence their trust in AI-mHealth broadly, and expressed a desire for more proactive and accessible information sharing about aspects of AI-mHealth apps. These findings present recommendations for consideration in the development and distribution of AI-mHealth apps that may enable greater ethical alignment between producers and consumers of AI-mHealth apps. Future evidence-based research examining the intersections between mHealth and AI with the attitudes of relevant stakeholders is still needed as is research that analyzes how the actual use of AI-mHealth apps aligns with the ethical perspectives identified in this study.

Acknowledgments

This study was supported by the National Center for Advancing Translational Sciences as a 1-year administrative supplement to a larger project (R01-TR-003505).

Data Availability

The datasets generated or analyzed during this study are not publicly available in order to protect the identities of participants but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Khan ZF, Alotaibi SR. Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective. J Healthc Eng 2020;2020:8894694. [doi: 10.1155/2020/8894694] [Medline: 32952992]
2. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence-empowered mHealth: scoping review. JMIR Mhealth Uhealth 2022 Jun 09;10(6):e35053 [FREE Full text] [doi: 10.2196/35053] [Medline: 35679107]
3. Yeung AW, Torkamani A, Butte AJ, Glicksberg BS, Schuller B, Rodriguez B, et al. The promise of digital healthcare technologies. Front Public Health 2023 Sep 26;11:1196596 [FREE Full text] [doi: 10.3389/fpubh.2023.1196596] [Medline: 37822534]
4. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MD, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 04;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]

XSL•FO
RenderX

5. Gerke S, Rezaeikhonakdar D. Privacy aspects of direct-to-consumer artificial intelligence/machine learning health apps. Intell Based Med 2022;6:100061. [doi: 10.1016/j.ibmed.2022.100061]
6. Rubeis G. iHealth: the ethics of artificial intelligence and big data in mental healthcare. Internet Interv 2022 Apr;28:100518 [FREE Full text] [doi: 10.1016/j.invent.2022.100518] [Medline: 35257003]
7. Ramkumar PN, Haeberle HS, Bloomfield MR, Schaffer JL, Kamath AF, Patterson BM, et al. Artificial intelligence and arthroplasty at a single institution: real-world applications of machine learning to big data, value-based care, mobile health, and remote patient monitoring. J Arthroplasty 2019 Oct;34(10):2204-2209. [doi: 10.1016/j.arth.2019.06.018] [Medline: 31280916]
8. Thapa C, Camtepe S. Precision health data: requirements, challenges and existing techniques for data security and privacy. Comput Biol Med 2021 Feb;129:104130. [doi: 10.1016/j.compbiomed.2020.104130] [Medline: 33271399]
9. Davies B. 'Personal health surveillance': the use of mHealth in healthcare responsibilisation. Public Health Ethics 2021 Nov;14(3):268-280 [FREE Full text] [doi: 10.1093/phe/phab013] [Medline: 34899983]
10. Kotz D, Gunter CA, Kumar S, Weiner JP. Privacy and security in mobile health: a research agenda. Computer (Long Beach Calif) 2016 Jun;49(6):22-30 [FREE Full text] [doi: 10.1109/MC.2016.185] [Medline: 28344359]
11. Brault N, Saxena M. For a critical appraisal of artificial intelligence in healthcare: the problem of bias in mHealth. J Eval Clin Pract 2021 Jun 23;27(3):513-519. [doi: 10.1111/jep.13528] [Medline: 33369050]
12. Gloria K, Rastogi N, DeGroff S. Bias impact analysis of AI in consumer mobile health technologies: legal, technical, and policy. arXiv Preprint posted online on August 29, 2022 [FREE Full text] [doi: 10.48550/arXiv.2209.05440]
13. Hunt X, Tomlinson M, Sikander S, Skeen S, Marlow M, du Toit S, et al. Artificial intelligence, big data, and mHealth: the frontiers of the prevention of violence against children. Front Artif Intell 2020 Oct 22;3:543305 [FREE Full text] [doi: 10.3389/frai.2020.543305] [Medline: 33733202]
14. Kühler M. Exploring the phenomenon and ethical issues of AI paternalism in health apps. Bioethics 2022 Feb 24;36(2):194-200. [doi: 10.1111/bioe.12886] [Medline: 34031908]
15. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Med Ethics 2021 Feb 15;22(1):14 [FREE Full text] [doi: 10.1186/s12910-021-00577-8] [Medline: 33588803]
16. Wutz M, Hermes M, Winter V, Köberlein-Neu J. Factors influencing the acceptability, acceptance, and adoption of conversational agents in health care: integrative review. J Med Internet Res 2023 Sep 26;25:e46548 [FREE Full text] [doi: 10.2196/46548] [Medline: 37751279]
17. He X, Zheng X, Ding H. Existing barriers faced by and future design recommendations for direct-to-consumer health care artificial intelligence apps: scoping review. J Med Internet Res 2023 Dec 18;25:e50342 [FREE Full text] [doi: 10.2196/50342] [Medline: 38109173]
18. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
19. Götzl C, Hiller S, Rauschenberg C, Schick A, Fechtelpeter J, Fischer Abaigar U, et al. Artificial intelligence-informed mobile mental health apps for young people: a mixed-methods approach on users' and stakeholders' perspectives. Child Adolesc Psychiatry Ment Health 2022 Nov 17;16(1):86 [FREE Full text] [doi: 10.1186/s13034-022-00522-6] [Medline: 36397097]
20. Tsai CH, You Y, Gui X, Kou Y, Carroll JM. Exploring and promoting diagnostic transparency and explainability in online symptom checkers. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445101]
21. Kim JP, Yang HJ, Kim B, Ryan K, Roberts LW. Understanding physician's perspectives on AI in health care: protocol for a sequential multiple assignment randomized vignette study. JMIR Res Protoc 2024 Apr 04;13:e54787 [FREE Full text] [doi: 10.2196/54787] [Medline: 38573756]
22. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
23. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
24. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
25. Downe-Wamboldt B. Content analysis: method, applications, and issues. Health Care Women Int 1992;13(3):313-321. [doi: 10.1080/07399339209516006] [Medline: 1399871]
26. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
27. Almourad MB, Alrobai A, Skinner T, Hussain M, Ali R. Digital wellbeing tools through users lens. Technol Soc 2021 Nov;67:101778. [doi: 10.1016/j.techsoc.2021.101778]

XSL•FO
RenderX

28. Song I, Pendse SR, Kumar N, De Choudhury M. The typing cure: experiences with large language model chatbots for mental health support. arXiv Preprint posted online on January 25, 2024 [FREE Full text]
29. Hoffmann CP, Lutz C, Ranzini G. Privacy cynicism: a new approach to the privacy paradox. Cyberpsychology (Brno) 2016;10(4) [FREE Full text] [doi: 10.2139/ssrn.3319830]
30. Draper NA, Turow J. The corporate cultivation of digital resignation. New Media Soc 2019 Mar 08;21(8):1824-1839. [doi: 10.1177/1461444819833331]
31. Hargittai E, Marwick A. "What can I really do?" Explaining the privacy paradox with online apathy. Int J Commun 2016;10 [FREE Full text]
32. Choi H, Park J, Jung Y. The role of privacy fatigue in online privacy behavior. Comput Hum Behav 2018 Apr;81:42-51. [doi: 10.1016/j.chb.2017.12.001]
33. Lutz C, Hoffmann CP, Ranzini G. Data capitalism and the user: an exploration of privacy cynicism in Germany. New Media Soc 2020 Jul 22;22(7):1168-1187. [doi: 10.1177/1461444820912544]
34. Shan Y, Ji M, Xie W, Lam KY, Chow CY. Public trust in artificial intelligence applications in mental health care: topic modeling analysis. JMIR Hum Factors 2022 Dec 02;9(4):e38799 [FREE Full text] [doi: 10.2196/38799] [Medline: 36459412]
35. Su Z, Figueiredo MC, Jo J, Zheng K, Chen Y. Analyzing description, user understanding and expectations of AI in mobile health applications. AMIA Annu Symp Proc 2020;2020:1170-1179 [FREE Full text] [Medline: 33936493]
36. Jiang J, Kahai S, Yang M. Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. Int J Hum Comput Stud 2022 Sep;165:102839. [doi: 10.1016/j.ijhcs.2022.102839]
37. Smith-Renner A, Fan R, Birchfield M, Wu T, Boyd-Graber J, Weld DS, et al. No explainability without accountability: an empirical study of explanations and feedback in interactive ML. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376624]
38. Branley-Bell D, Whitworth R, Coventry L. User trust and understanding of explainable AI: exploring algorithm visualisations and user biases. In: Proceedings of the Human-Computer Interaction. Human Values and Quality of Life: Thematic Area. 2020 Presented at: HCI 2020; July 19, 2020; Copenhagen, Denmark. [doi: 10.1007/978-3-030-49065-2_27]
39. Gkinko L, Elbanna A. Hope, tolerance and empathy: employees' emotions when using an AI-enabled chatbot in a digitalised workplace. Inf Technol People 2022 May 03;35(6):1714-1743. [doi: 10.1108/itp-04-2021-0328]
40. Kelley PG, Bresee J, Cranor LF, Reeder RW. A "nutrition label" for privacy. In: Proceedings of the Symposium on Usable Privacy and Security. 2009 Presented at: SOUPS 2009; July 15-17, 2009; Mountain View, CA URL: https://cups.cs.cmu.edu/soups/2009/proceedings/a4-kelley.pdf [doi: 10.1145/1572532.1572538]
41. Perrault EK, Nazione SA. Informed consent-uninformed participants: shortcomings of online social science consent forms and recommendations for improvement. J Empir Res Hum Res Ethics 2016 Jul 31;11(3):274-280. [doi: 10.1177/1556264616654610] [Medline: 27329533]
42. O' Sullivan L, Feeney L, Crowley RK, Sukumar P, McAuliffe E, Doran P. An evaluation of the process of informed consent: views from research participants and staff. Trials 2021 Aug 18;22(1):544 [FREE Full text] [doi: 10.1186/s13063-021-05493-1] [Medline: 34407858]
43. Hayman RM, Taylor BJ, Peart NS, Galland BC, Sayers RM. Participation in research: informed consent, motivation and influence. J Paediatr Child Health 2001 Feb 21;37(1):51-54. [doi: 10.1046/j.1440-1754.2001.00612.x] [Medline: 11168870]
44. Schuster AM, Cotten SR. Differences between employed and retired older adults in information and communication technology use and attitudes. Work Aging Retire 2024 Jan;10(1):38-45. [doi: 10.1093/workar/waac025]
45. Elias SM, Smith WL, Barney CE. Age as a moderator of attitude towards technology in the workplace: work motivation and overall job satisfaction. Behav Inf Technol 2012 May;31(5):453-467. [doi: 10.1080/0144929x.2010.513419]

Abbreviations

AI: artificial intelligence
AI-mHealth: artificial intelligence–enabled mobile health
mHealth: mobile health

Edited by L Buis; submitted 24.07.24; peer-reviewed by B Thies, D Singh; comments to author 09.01.25; revised version received 14.02.25; accepted 25.04.25; published 20.05.25.

Please cite as:
Ryan K, Hogg J, Kasun M, Kim JP
Users' Perceptions and Trust in AI in Direct-to-Consumer mHealth: Qualitative Interview Study
JMIR Mhealth Uhealth 2025;13:e64715
URL: https://mhealth.jmir.org/2025/1/e64715
doi:10.2196/64715
PMID:

XSL•FO
RenderX

©Katie Ryan, Justin Hogg, Max Kasun, Jane Paik Kim. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Population-Level Dissemination of a Smoking Cessation Smartphone App: Quasi-Experimental Comparison of Values-Based Messages in Social Media Advertisements

Jonathan B Bricker[1,2], PhD; Margarita Santiago-Torres[1], PhD; Kristin E Mull[1], MS; Brianna M Sullivan[1], MS; Ravi Mehrotra[1,3], MD

[1]Cancer Prevention Program, Division of Public Health Sciences, Fred Hutchinson Cancer Center, 1100 Fairview Avenue North, Seattle, WA, United States

[2]Department of Psychology, University of Washington, Seattle, Washington, United States

[3]Centre for Health, Innovation and Policy Foundation, Noida, Uttar Pradesh, India

Corresponding Author:
Jonathan B Bricker, PhD
Cancer Prevention Program, Division of Public Health Sciences, Fred Hutchinson Cancer Center, 1100 Fairview Avenue North, Seattle, WA, United States

Abstract

Background: Cigarette smoking is prevalent in many countries worldwide, especially in low- and middle-income countries (LMICs), presenting an urgent public health challenge. Disseminating freely available smoking cessation treatments that effectively decrease cigarette smoking globally is urgently needed.

Objective: Identify the highest impact and most cost-effective values-based social media advertisements to disseminate our smoking cessation smartphone app, “iCanQuit”, among adults living in 7 major cities of India. Values represented in the advertisements included family, relationships, self-care, health, and self-control. Using a quasi-experimental design, we aimed to determine (1) which values-based advertisements had the highest smoking cessation app dissemination reach, as measured by click-through rate (CTR), app installs, and app usage metrics; and (2) which values-based message advertisements were more cost-effective as measured by cost-per-impression, cost-per-click, and cost-per-install. The study population included a selected media market of individuals living in 7 metro cities of India – Delhi, Mumbai, Kolkata, Chennai, Bengaluru, Hyderabad, and Pune – who were exposed to one of 6 social media advertisements from January 16 to May 5, 2024.

Methods: The advertisement campaign design for each of the identified values, based on previous smoking cessation trial data, followed a collaborative iterative process. Advertisements ran sequentially for 16 weeks. Advertisement exposure and app usage data were objectively collected via Google’s Display & Video 360 advertisements campaign management and Firebase app development platforms. Advertisement exposure impact on app engagement was measured via several metrics, including click-through rate (CTR, ie, the likelihood of user clicks on an advertisement after seeing it), the number of app installs (ie, a user opening the app for the first time after downloading it), and the number of app sessions (ie, app usage). Cost efficiency was measured via cost per click and cost per install for each ad.

Results: Overall, the CTR was 5%. The app was installed 5111 times. The average cost per click and cost per app install across all advertisements were US $ 0.006 and US $ 6.43, respectively. The advertisements with the lowest cost per install (range: US $4.83-US $5.16) and highest CTR (between 6% and 9%) focused on the values of family, health, and self-control. Advertisements focused on the values of relationships and self-care had modestly higher levels of engagement.

Conclusions: Advertisements focusing on the values of family, health, and self-control had the highest potential reach at the lowest cost. Overall, these findings provide insights into the reach and cost-effectiveness of values-based messages in social media advertisements, guiding future outreach efforts for population-level dissemination of smoking cessation apps.

(JMIR Mhealth Uhealth 2025;13:e71619) doi:10.2196/71619

KEYWORDS

advertisement; dissemination; iCanQuit; India; smartphone applications; smoking cessation; social media; values-based messages

XSL•FO
RenderX

Introduction

Globally, cigarette smoking is a significant public health challenge across many demographic groups [1]. Worldwide, one in 4 adults currently smokes cigarettes and men have almost 5 times higher smoking rates than women [2,3]. India, China, and Indonesia, which are low- and middle-income countries (LMICs), account for 51.4% of the world's men who smoke. More than half of all smoking-attributed global deaths occurred in China, India, Russia, and the United States combined, with men alone accounting for 75% of these deaths [4].

Quitting smoking is the single most effective method for preventing cigarette smoking-related morbidity and mortality—especially in LMICs [5,6]. Unfortunately, barriers to access to health care, lack of health professionals trained in evidence-based smoking cessation treatments, lack of access to and subsequent usage of such treatments, and transportation-related barriers contribute to smoking cessation programs being out of reach for those who are most in need. Implementing free, remotely accessible digital interventions may improve access to evidence-based smoking cessation programs at the population level.

Digital interventions, such as smartphone apps, offer a newer approach to in-person interventions, with the potential for broader reach and greater acceptance [7,8]. They are freely available on app stores, accessible anytime, and do not always require internet access [9-12]. An efficacious smoking cessation application could be a feasible method for community, clinical, and public health programs to offer scalable smoking cessation services—thereby addressing health inequities in tobacco control.

Our group developed iCanQuit, an Acceptance and Commitment Therapy (ACT)-based smartphone smoking cessation app [13]. Unlike standard interventions for smoking cessation [14], ACT-based smoking cessation interventions teach people to observe, acknowledge, and accept their cravings to smoke rather than avoid them, and focus on life values, instead of expectations, as motivation to quit [13,15]. iCanQuit targets 2 core theoretical processes of ACT, acceptance and values. The acceptance component teaches skills to accept sensations, emotions, and thoughts that cue smoking via distancing from thoughts about smoking, mindfulness, and flexible perspective taking. The ACT-based emphasis on values teaches skills for determining the importance of valued life domains (eg, family, health, and self-care) that motivate quitting smoking.

We tested the efficacy of iCanQuit against a US Clinical Practice Guidelines-based app (QuitGuide) for cigarette smoking cessation in a full-scale randomized clinical trial (RCT) among 2415 adults residing in the United States [16]. iCanQuit was 1.5 times more efficacious than QuitGuide for smoking cessation at 12 months (P<.001) [16]. In addition, the intervention's impact on smoking cessation was mediated by an increase in acceptance of cues to smoke and higher engagement with the app [17]. Although iCanQuit is the only known smoking cessation app proven efficacious for smoking cessation in a full-scale randomized trial, we are aware of no published studies on methods to disseminate any smoking cessation app.

Historically, exposure to antismoking mass media (ie, television) advertisement campaigns has been deployed to disseminate smoking cessation programs at the population level [18-21]. The most used approach is "negative advertising" or the use of advertisements with graphic content that provoke a strong negative emotional response, such as fear, disgust, and sadness about the serious harms of smoking [22]. Tobacco control advertisement campaigns are also more effective when they provide actionable resources for cessation (ie, telephone number for a support line) [23].

An alternative approach to negatively framed advertisements is to design an advertisement campaign with personal values known to motivate behavior change [24]. According to the ACT approach underlying the iCanQuit app, values are "chosen qualities of an action with purpose" [25,26]. Personal values motivate people's choices, depending on the importance of the value-behavior domain [24]. Personal values can influence behavior change and an individual's response to health-related information [27,28]. However, little is known about the use of values-based messages in tobacco control population-level advertisement campaigns. A recent study aimed at developing antismoking communication campaigns examined the associations between personal values and tobacco control policies [29]. The most frequently selected values by participants included (1) "assuring my family is safe and secure," which reflects the value of family and (2) "keeping myself in good health," which reflects the value of health. These results suggest that values-based messaging is a promising approach in public health communication efforts to increase the adoption of smoking cessation programs at the population level.

Building on this research, we tested the use of values-based messages to disseminate the iCanQuit smoking cessation app at the population level. For this, we selected adults in India, the largest and fastest growing population globally (surpassing China in 2023) [30], with disproportionately high rates of cigarette smoking, especially among men (17.4%, equivalent to >100 million men in India), who have over 6 times higher rates of cigarette smoking than women (2.8%) [4]. Therefore, there is an urgent need to access evidence-based smoking cessation programs.

Given that cigarette smoking is a significant public health challenge in India [31], and to address the need for freely available smoking cessation resources, we tested the dissemination potential of the iCanQuit app to individuals in India. We first conducted user-centered qualitative research to inform the tailoring of iCanQuit to users in India. We then conducted user-centered qualitative research to inform the selection of values-based messages in social media advertisement campaigns. Subsequently, we conducted a pilot social media advertisement campaign to determine the most efficient and cost-effective strategies for reaching a broad audience of adults who smoke. The 16-week advertisement campaign aimed to compare the reach of 5 personal values-based message advertisements (ie, family, relationships, self-care, health, and self-control) known to motivate behavior change in promoting installation and usage of the iCanQuit app [24,32].

XSL•FO
RenderX

The study's main objective was to identify the highest impact values-based social media advertisements to disseminate our smoking cessation smartphone app, "iCanQuit," among adults living in 7 major cities of India. Using a quasi-experimental design, the goals were to (1) determine which values-based advertisements had the highest smoking cessation app dissemination reach, as measured by click-through rate (CTR), app installs, and app usage metrics; and (2) determine which values-based message advertisements were more cost-effective as measured by cost-per-impression, cost-per-click, and cost-per-install.

Methods

Overview

The main objective of this quasi-experimental comparative study was to identify the highest reaching and most cost-effective values-based social media advertisement to disseminate a smoking cessation app already proven to be efficacious to help adults quit cigarette smoking [16]. The visual advertisement messaging was focused on various personal values hypothesized to motivate adults to quit smoking. In the context of this campaign, we tested various personal values messaging that may motivate viewers exposed to the advertisements to install the app, thereby promoting app engagement.

Adults in a selected media market in India were exposed to one of 6 advertisements that ran sequentially from January 16, 2024, to May 5, 2024 (Table 1). The user sees the ad, finds the message interesting, clicks on it, and is directed to the advertisement's website (ie, the app store iCanQuit page). Over the course of 16 weeks, the digital advertisement campaign sequentially deployed each advertisement. The generic control advertisement ran for about 6 weeks, while each values-based advertisement ran for about 2 weeks.

Table . Advertisement campaign active dates.

Type of value-based advertisement	Dates advertisement was active	Number of days
Generic control ad	01/16/2024-02/26/2024	42
Family	02/22/2024-03/12/2024	20
Relationships	02/22/2024-02/26/2024 03/16/2024-03/23/2024	13
Self-care	03/25/2024-04/07/2024	14
Health	04/08/2024-04/22/2024	15
Self-control	04/23/2024-05/05/2024	13

Priority Population

The study population included individuals living in 7 metro cities of India – Delhi, Mumbai, Kolkata, Chennai, Bengaluru, Hyderabad, and Pune—as the test areas due to their high population density as well as cultural and linguistic diversity [33].

Values-Based Visual Advertisements

Personal values were identified based on empirical research data from previous digital smoking cessation intervention trials conducted by our group [16,34]. We selected the most frequently reported personal values that participants cited as motivations for quitting cigarette smoking. These values include: (1) family, (2) health, (3) money, (4) spirituality, (5) relationships, (6) pets, (7) self-growth, and (8) friends.

The advertisement campaign design for each of the identified values, based on previous smoking cessation trial data, followed a collaborative 2-step process. Step 1 involved an expert-guided development of values-based messages to motivate adults who smoke to quit smoking. Our research and clinical team of investigators and colleagues with expertise in ACT worked together to generate values-based messages based on their expertise in delivering behavior-based interventions [13,26]. Step 2 involved developing visual advertisements for each of the 5 selected messages through an iterative design process. We collaborated with a vendor specializing in visual advertisement creation (GroupM) [35], ensuring each advertisement was aligned with a specific message.

Visual advertisements were created to align with the 5 personal values of family, relationships, self-care, health, and self-control. A sixth generic control advertisement that did not correspond to a specific value served as a reference (Figure 1). Advertisements ran in the English language. Each visual advertisement ran in 4 mobile-friendly image sizes: 250 × 250, 300 × 250, 320 × 480, and 320 × 50 pixels.

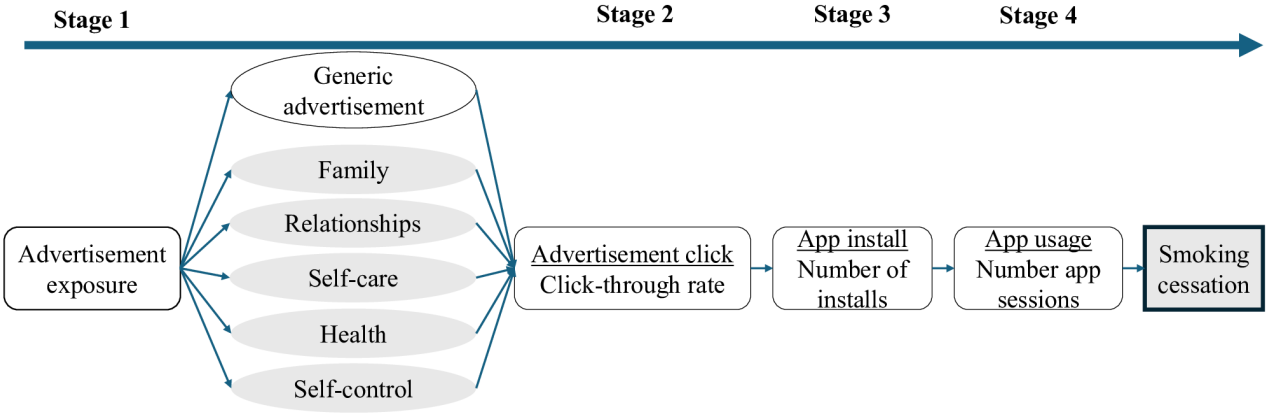
XSL•FO
RenderX

Figure 1. Visual advertisements used in the campaign. Top row, from left to right: generic, family, relationships. Bottom row, from left to right: self-care, health, self-control.

Conceptual Model

We hypothesized that exposure to values-based messages via social media advertisements results in app engagement via a 4-stage process (Figure 2). In stage 1, adults in a selected media market in India are exposed to one of 6 advertisements—a generic control advertisement and 5 values-based advertisements—that run sequentially. In stage 2, the values-based messaging in the ads, which is theorized to resonate with adults who smoke, leads to advertisement clicks, which guide users to download the app. In stage 3, users download and install the app on their smartphones. In stage 4, users engage with the app, and it is theorized that engagement with the smoking cessation app leads to smoking cessation given that engagement is a strong predictor of smoking cessation [17,36]. Note that smoking cessation is not an outcome of this study and thus was not measured.

Figure 2. Conceptual model: values-based messages in social media advertisement.



Data Collection Overview

Data were objectively collected via Google's Display & Video 360 (DV360) advertisements campaign management and Firebase app development platforms. Advertisement exposure and advertisement clicks were measured through DV360, while app installs and app usage metrics (user usage of the iCanQuit app once installed) were tracked via Firebase. Once an app was successfully installed, the user was linked to the advertisement they clicked as captured via the DV360 platform.

XSL•FO
RenderX

Stage 1. Advertisement Exposure

We used Google's Display & Video 360 (DV360) platform for content-oriented advertising. DV360 is an advertising platform within the Google Marketing Platform, which is designed for managing display and video advertisement campaigns. DV360 seamlessly integrates advertisements into the user's web browsing experience through tailoring and real-time bidding processes, ensuring relevant advertisements are shown to the right users at the right time. The details of each step in this process are outlined in Table 2.

Table . Steps in visual advertisement exposure through DV360.

Step	Description
Browsing online	A user browses the web by visiting a blog or web page.
Real-time bidding and advertisement request sent	As the page loads, it sends an advertisement request to an advertisement exchange (ie, DV360 platform).
User profile matching and DV360 bids	DV360 identifies the user as "interested" or "high intent" based on the set of content-oriented parameters and places a bid. Advertisements are also triggered on web pages that include relevant keywords (ie, content-oriented keywords).
Auction and advertisement selection	DV360 wins the auction, and an advertisement is selected.
Advertisement display	The advertisement appears in a banner slot on the blog or web page that the user is browsing.
Post advertisement user interaction	The user sees the ad, finds the message interesting, clicks on it, and is directed to the ad's website (ie, the app store iCanQuit page).

Advertisements were tailored to affinity and in-market audience segments, as well as to users of 48 competitor smoking cessation apps, for their likely interest in smoking cessation to maximize the impact and cost-efficiency of the campaign. Affinity groups used for advertisement tailoring included the following: food and dining, frequently dines out, frequently eats breakfast out, frequently eats lunch out, frequently eats dinner out, coffee shop regulars, fashionistas, beauty mavens, shutterbugs, technophiles, frequently attends live events, pop music fans, theater aficionados, and media and entertainment. In-market segments selected included makeup and cosmetics, soccer tickets, party supplies and planning, event planning services, business financial services, and vacation packages. Advertisements were compared descriptively on several metrics. For instance, they were compared on impressions, which is the number of times a visual advertisement is displayed to a user on a webpage or blog.

The full advertisement campaign ran from January 16, 2024, to May 5, 2024 (Table 1). Over the course of 16 weeks, the digital advertisement campaign sequentially deployed each ad. The generic control advertisement ran for about 6 weeks, while each values-based advertisement ran for about 2 weeks.

Stage 2. Advertisement Clicks

Clicks are the number of times users interact with an advertisement by clicking on it. CTR is the ratio of clicks to impressions, expressed as a percentage. It measures the likelihood of users clicking on an advertisement after being exposed to it, indicating interest in the message.

Stage 3. App Installs

App installs are the number of times an app is successfully downloaded and installed on a device. The conversion rate is the ratio of installations to clicks. The cost per install is the average cost for each app installation, calculated by dividing the total advertisement campaign cost by the number of app installs. To monitor app installs and assess the impact of the advertisement campaign on app usage metrics as a measure of app engagement, the mobile measurement platform, Firebase, was integrated with the advertisement serving platform Campaign Manager DV360. Firebase is a Backend-as-a-Service (BaaS) platform developed by Google that provides a suite of tools and services to help developers build, deploy, and manage applications, particularly mobile and web apps, with features like real-time database, authentication, and cloud storage. This integration facilitated the tracking of app installations and app usage metrics described below.

Stage 4. App Usage Metrics

Once the iCanQuit app was successfully installed, all in-app activity was tracked by Firebase. Several usage metrics were summarized for the first 30 days for each app user, including the (1) number of app sessions, (2) time per session, (3) unique days of use, and (4) length of use from the first to last session. Smoking status and cessation were not measured.

Statistical Analysis

Users were linked to the advertisement they viewed through advertisement tracking within Firebase, the Google platform used to track app installs and app usage. When an app was successfully installed, the user was assigned an advertisement tracking ID corresponding to the Google advertisement clicked to initiate the app download. For users who had advertisement tracking disabled, we assigned the advertisement campaign based on the installation date, selecting the campaign closest to that date.

Given the quasi-experimental design, there were no inclusion or exclusion criteria nor missing data. However, there were 34 users who had advertisement tracking features disabled and installed the app between February 22 - 26, 2024, when 3 advertisements were running simultaneously. As it was

XSL•FO
RenderX

impossible to assign these users to a specific advertisement, they were excluded from the analysis.

To compare the values-based advertisements on advertisement metrics, we used chi-square tests for CTR and cost per install and ANOVA for the number of installs per day. Advertisements were compared on app usage counts using negative binomial models for the number of app sessions, unique days of use, and length of use, given their heavily right-skewed distributions. Finally, advertisements were compared on time per session using ANOVA.

Ethical Considerations

All study procedures, protocols, and procedures for the advertisement campaign were approved by the Fred Hutchinson Cancer Center's Institutional Review Board and determined to be a nonhuman subject research since we did not interact directly with individuals, nor collect any personally identifiable information.

Results

Advertisement Clicks

The advertisement campaigns resulted in a total of 119,258,689 impressions and 5,635,114 clicks (Table 3). CTR significantly varied across advertisements (P<.001), averaging 5% across ads. The relationships and health values-based advertisements had higher CTRs, at 9% each.

Table . Advertisement metrics.

Variables	Type of ad							*P* value
	Overall	Generic	Family	Relationships	Self-care	Health	Self-control	
Number of impressions	119,258,689	25,939,433	21,783,538	14,292,366	26,077,665	14,093,203	17,072,484	NA[a]
Number of clicks	5,635,114	911,843	1,261,810	494,136	153,747	1,285,553	1,528,025	NA
Click-through rate[b], %	5%	4%	6%	3%	1%	9%	9%	<.001
Number of app installs	5111	1149	1192	509	346	946	969	NA
Average number of installs per day, mean (SD)	46.0 (28.3)	27.4 (13.8)	59.6 (27.6)	39.2 (19.2)	24.7 (19.9)	63.1 (30.5)	74.5 (17.2)	<.001
Conversion rate[c], %	0.09%	0.13%	0.09%	0.10%	0.23%	0.07%	0.06%	<.001

[a]NA: not applicable

[b]Click-through rate = Number of clicks/Number of impressions.

[c]Conversion rate = Number of app installs/Number of clicks.

App Installs

The advertisement campaigns resulted in a total of 5111 app installations (app installs; Table 3). App installs per day varied significantly across advertisements (P<.001). The self-control and health value-based advertisements averaged the greatest number of installs per day, with means of 74.5 (SD=17.2) and 63.1 (30.5), respectively. The conversion rate, or the ratio of installations to clicks, also varied significantly between advertisements (P<.001). The self-care value-based advertisement had the highest conversion rate of 0.23%, while the self-control value-based advertisement had the lowest conversion rate of 0.06%.

App Usage Metrics

The number of app sessions, minutes spent per session, unique days of use, and days from the first to last session were summarized for the first 30 days for each of the 5111 app users in India. Summary usage statistics are given in Table 4 for all app users combined, and separately by advertisement viewed. Most app usage metrics were similar between values-based advertisement campaigns, except number of app sessions varied significantly across campaigns (P=.04). The relationships value-based advertisement averaged a higher number of app sessions per user (mean=2.3, median=1). The mean length of app use in the first 30 days was 9.3 (SD=15.8) days

XSL•FO
RenderX

Table . App usage metrics.

Usage metric, mean (SD)	Overall (n=5111)	Type of advertisement						*P* value
		Generic n=1149	Family n=1192	Relationships n=509	Self-care n=346	Health n=946	Self-control n=969	
Number of app sessions	2.1 (2.7)[a]	2.2 (3.1)[a]	2.0 (2.2)[a]	2.3 (3.3)[a]	2.2 (3.9)[a]	2.1 (2.3)[a]	2.1 (2.4)[a]	.04
Time per session, minutes	0.9 (2.0)[b]	0.9 (1.9)[b]	0.9 (1.9)[b]	1.0 (1.7)[b]	0.8 (1.8)[c]	0.8 (1.8)[b]	0.9 (2.6)[b]	.85
Unique days of use, days	1.8 (1.9)[a]	1.9 (2.1)[a]	1.8 (1.6)[a]	2.0 (2.1)[a]	1.9 (2.5)[a]	1.8 (1.7)[a]	1.8 (1.7)[a]	.17
Length of use from first to last session, days	9.3 (15.8)[a]	9.3 (16.0)[a]	9.4 (15.9)[a]	10.2 (16.6)[a]	10.3 (16.7)[a]	9.0 (15.4)[a]	8.5 (15.2)[a]	.14

[a]Median=1.

[b]Median=0.3.

[c]Median=0.2

[d]During the first 30 days after downloading the app.

Advertisement Budget and Cost

The campaign budget was evenly divided across the 5 values-based advertisements to ensure equal exposure and allow for unbiased results. The total budget for the iCanQuit India initiative was Indian Rupees 26 lakhs, equivalent to 2,600,000 rupees or US$ 34,211 (exchange rate on February 27, 2024). The total campaign media advertisement spend was US $32,847.18. The 3 top-performing advertisements in terms of cost per install and number of installations were family, health, and self-control value-based ads, while the worst-performing advertisements were relationships and self-care value-based advertisements. The full cost breakdown by advertisement is given in Table 5. The average cost per install across all advertisements was US $6.43; the average cost per install across the family, health, and self-control value-based advertisements was US $4.97, with the family value-based advertisement having the lowest cost per install at US $4.83.

Table . Advertisement budget and cost.

Variables	Type of advertisement						
	Overall	Generic	Family	Relationships	Self-care	Health	Self-control
Advertisement spending, USD	$32,847.18	$6016.61	$5756.81	$4976.20	$6400.43	$4885.89	$4811.24
Cost-per 1000 impressions, USD	$0.275	$0.232	$0.264	$0.348	$0.245	$0.347	$0.282
Cost-per-click, USD	$0.006	$0.007	$0.005	$0.010	$0.042	$0.004	$0.003
Cost-per-install, USD	$6.43	$5.24	$4.83	$9.78	$18.50	$5.16	$4.97

Discussion

Principal Findings

The success of the iCanQuit values-based advertisement campaign was measured by its ability to effectively communicate personal values hypothesized to motivate change in behavior (ie, smoking cessation) and its yield of app installs by the selected audience (users in India). The advertisement campaign generated approximately 119 million impressions and 5.6 million clicks with an average click-through rate of 5% total. This click-through rate was greater than industry standards, which range between 0.5% and 1% [37]. The campaign resulted in approximately 5111 app installs, with reach (click-through rate and conversion rate) and cost efficiency of the campaign varying across advertisements.

The self-control value-based advertisement achieved the highest reach and cost-effectiveness, followed by the health value-based advertisement, suggesting that self-control and health are key motivators for smoking cessation among people in India. In contrast, the self-care and relationships value-based advertisements showed relatively low reach and cost-effectiveness. The self-care value-based advertisement had the highest conversion rate of all advertisements, but its absolute volume of conversions remained the lowest. We are unsure why the self-care advertisement did not perform well, though we believe the advertisement's high conversion rate is due to its

XSL•FO
RenderX

absolute number of impressions. Specifically, the self-care advertisement had nearly twice as many impressions (26,077,665) as the health advertisement (14,093,203), even though both advertisements ran for a similar number of days (14 and 15 days, respectively). There are a few plausible explanations for this. First, it is possible that the algorithm favored the self-care advertisement more than the health advertisement, giving it more exposure. Second, the self-care advertisement may have won more advertisement auctions, which is an iterative process in which both tailoring and real-time bidding processes ensure relevant advertisements are shown to the right users at the right time. For instance, if the self-care advertisement initially generated a higher click-through rate, the algorithm would predict strong performance and therefore show it more often. In other words, the advertisement received more screen time because it was perceived as more engaging and better suited for the audience.

While the relationship values-based advertisement resulted in the greatest engagement with the app, the advertisement had relatively low reach and cost-effectiveness. Finally, while the generic control advertisement had a relatively low cost per install (US $5.24), this advertisement yielded a total of 1149 installs in 6 weeks compared with the family value-based advertisement which yielded 1192 in half the time.

Importance of Findings

This is the first study, to our knowledge, to draw a direct comparison of the reach and cost-effectiveness of values-based advertisement campaigns for disseminating a smoking cessation app at the population level. We found that advertisements focusing on the values of self-control and health yielded the highest reach at the lowest cost, suggesting potential for broad dissemination and scalability. The high performance of self-control-based messages indicates that the ability to regulate urges to smoke as a strategy to quitting smoking may resonate with adults in India. The health-based messaging confirms previous research suggesting health is an important motivator to quitting smoking [38].

Comparison With Previous Work

To put the findings of this study in context, the only known previous study of a mobile-based smoking cessation program is a media campaign for SmokeFreeTXT (SFT), a text messaging program for adolescents in the United States [39]. This was a nationwide campaign that included traditional (television and radio), online, and social media outreach with action-based advertisement messaging. The SFT social media campaign resulted in a click-through rate of 0.14%, as compared with 5% in our study. In terms of cost, the cost-per-click in the SFT social media campaign was US $0.63 compared with a cost-per-click of US $0.0006 in our study. These results suggest that our social media campaign was more effective in reach and cost among adults in India. The SFT measure of engagement was the number of new website sign-ups, which was 2384 new sign-ups in 8 weeks. In comparison, our study achieved 5111 app installs in a period of 16 weeks, demonstrating the feasibility and dissemination potential of our values-based social media campaign to LMICs in a relatively shorter period of time.

Our work extends previous research by demonstrating the feasibility of a social media advertisement campaign as a cost-effective strategy to disseminate a smoking cessation app in an LMIC such as India. While most tobacco control advertisement campaigns, whether digital or traditional, have been conducted and evaluated in high-income countries (ie, the United States of America, Canada, Australia, and the United Kingdom) [40], our study focused on a priority population in an LMIC. Second, antitobacco campaigns have historically focused on the negative health consequences of smoking, whereas our study provides first-time data on the potential reach and cost-effectiveness of a mobile-based visual advertisement campaign for a smoking cessation app that focused on personal values known to motivate behavior change. Finally, beyond providing data on advertisement exposure, our study also extends this literature by further evaluating app usage metrics provided, an overall measure of engagement, known to contribute to smoking cessation [17,36].

The results also offer specific insights for future media advertisement campaigns. First, close monitoring of advertisement campaign performance is essential. While some advertisements may initially show higher engagement and be favored by platform algorithms, the click-through rates need to be put in perspective with other important metrics. For instance, these early signals, such as the click-through rate, must be interpreted alongside other outcomes such as the conversion rate and cost-effectiveness. Second, greater exposure does not necessarily translate to better performance. For example, although the generic advertisement ran longer (6 weeks vs 2 weeks on average) and had a lower cost per impression, it yielded modest click-through rate and conversion rate compared to shorter values-based advertisement campaigns. Finally, the resonance of advertisement content with the target audience matters. Tailoring advertisement messages to the cultural and motivational context of the audience is critical, and user-centered research in the intended market is a key step for campaign success.

Limitations

This study has several key limitations. First, the nonrandomized design introduces the potential for selection bias, as users who engaged with a particular type of advertisement may differ systematically from those exposed to other advertisements in this study. In addition, user characteristics were not collected, meaning unmeasured confounders, such as socioeconomic status or smoking behaviors, could have influenced both advertisement engagement and app usage. Second, individuals may have seen multiple advertisements or viewed the same advertisement multiple times, and thus, previous exposure to earlier advertisements and their potential influence on engagement could not be measured. Third, in rare cases when multiple advertisement campaigns ran simultaneously, it was not possible to objectively determine which advertisement(s) an individual was exposed to if they had advertisement tracking turned off (n=34 users excluded, 0.7% of the sample). Fourth, findings are limited to individuals in 7 selected urban areas in India comprising a total population of 107.3 million people. Replication is needed in other regions, including rural areas in India and other countries. Finally, this study did not evaluate

XSL•FO
RenderX

the impact of advertisements on key smoking cessation outcomes, such as intentions to quit, quit attempts, or sustained cessation. Future research should assess whether these types of advertisement campaigns ultimately lead to smoking cessation.

Conclusions

The results of this study indicate that exposure to advertisements focusing on the values of family, health, and self-control had the highest potential reach at the lowest cost. Overall, these findings provide insights into the reach and cost-effectiveness of values-based messages in social media advertisement, guiding future outreach efforts for population-level dissemination of smoking cessation apps.

Acknowledgments

We appreciate the tireless contributions of the entire study staff and the development services of Moby, Inc. We are very appreciative of the study participants. Funding statement The study was funded by a generous donation from the Ladir Foundation to the Fred Hutchinson Cancer Center.

Disclaimer

There was no use of generative artificial intelligence (AI) technology in the generation of text, figures, or other informational content of this manuscript.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

JBB was responsible for conceptualization, funding acquisition, investigation, methodology, supervision, and writing – review and editing. MST contributed to writing – original draft. BMS handled project administration and writing – review and editing. KEM was involved in formal analysis and writing – review and editing. RM contributed to writing – review and editing.

Conflicts of Interest

None declared.

References

1. Verghese C, Redko C, Fink B. Screening for lung cancer has limited effectiveness globally and distracts from much needed efforts to reduce the critical worldwide prevalence of smoking and related morbidity and mortality. J Glob Oncol 2018 Sep;4:1-7. [doi: 10.1200/JGO.2017.1700016] [Medline: 30241213]
2. Reitsma MB, Kendrick PJ, Ababneh E. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet 2021 Jun 19;397(10292):2337-2360. [doi: 10.1016/S0140-6736(21)01169-7] [Medline: 34051883]
3. Bilano V, Gilmour S, Moffiet T, et al. Global trends and projections for tobacco use, 1990-2025: an analysis of smoking indicators from the WHO Comprehensive Information Systems for Tobacco Control. Lancet 2015 Mar 14;385(9972):966-976. [doi: 10.1016/S0140-6736(15)60264-1] [Medline: 25784347]
4. Reitsma MB, Fullman N, Ng M. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. Lancet 2017 May 13;389(10082):1885-1906. [doi: 10.1016/S0140-6736(17)30819-X] [Medline: 28390697]
5. U.S. Department of Health and Human Services. Smoking Cessation A Report of the Surgeon General: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 2020.
6. US department of health and human services. Tobacco use among U.S racial/ethnic minority groups—African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, and Hispanics: a report of the surgeon general Atlanta, Georgia: U.S department of health and human services, centers for disease control and prevention, national center for chronic disease prevention and health promotion, office on smoking and health. 1998. URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/00055081.htm [accessed 2025-07-08]
7. Wasil AR, Gillespie S, Shingleton R, Wilks CR, Weisz JR. Examining the reach of smartphone apps for depression and anxiety. Am J Psychiatry 2020 May 1;177(5):464-465. [doi: 10.1176/appi.ajp.2019.19090905] [Medline: 32354266]
8. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]

XSL•FO
RenderX

9. Chen YF, Madan J, Welton N, et al. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. Health Technol Assess 2012;16(38):1-205. [doi: 10.3310/hta16380] [Medline: 23046909]
10. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. Telemed J E Health 2014 Mar;20(3):206-214. [doi: 10.1089/tmj.2013.0169] [Medline: 24350804]
11. Harrison R, Flood D, Duce D. Usability of mobile applications: literature review and rationale for a new usability model. J Interact Sci 2013;1(1):1. [doi: 10.1186/2194-0827-1-1]
12. Broderick J, Devine T, Lemerise AJ, et al. Designing health literate mobile apps. NAM Perspectives 2014;4(1) [FREE Full text] [doi: 10.31478/201401a]
13. Hayes SC, Levin ME, Plumb-Vilardaga J, Villatte JL, Pistorello J. Acceptance and commitment therapy and contextual behavioral science: examining the progress of a distinctive model of behavioral and cognitive therapy. Behav Ther 2013 Jun;44(2):180-198. [doi: 10.1016/j.beth.2009.08.002] [Medline: 23611068]
14. Fiore M, Jaén CR, Baker TB. A clinical practice guideline for treating tobacco use and dependence: 2008 update. Am J Prev Med 2008 Aug;35(2):158-176. [doi: 10.1016/j.amepre.2008.04.009]
15. Webb Hooper M, Dietz NA, Wilson JC. Smoking urges during treatment and long-term cessation among low-income African Americans. Ethn Dis 2017;27(4):395-402. [doi: 10.18865/ed.27.4.395] [Medline: 29225440]
16. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020 Nov 1;180(11):1472-1480. [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
17. Bricker JB, Levin M, Lappalainen R, et al. Mechanisms of smartphone apps for cigarette smoking cessation: results of a serial mediation model from the iCanQuit randomized trial. JMIR Mhealth Uhealth 2021 Nov 9;9(11):e32847. [doi: 10.2196/32847] [Medline: 34751662]
18. Durkin SJ, Biener L, Wakefield MA. Effects of different types of antismoking ads on reducing disparities in smoking cessation among socioeconomic subgroups. Am J Public Health 2009 Dec;99(12):2217-2223. [doi: 10.2105/AJPH.2009.161638] [Medline: 19833980]
19. Emery S, Kim Y, Choi YK, Szczypka G, Wakefield M, Chaloupka FJ. The effects of smoking-related television advertising on smoking and intentions to quit among adults in the United States: 1999-2007. Am J Public Health 2012 Apr;102(4):751-757. [doi: 10.2105/AJPH.2011.300443] [Medline: 22397350]
20. Hyland A, Wakefield M, Higbee C, Szczypka G, Cummings KM. Anti-tobacco television advertising and indicators of smoking cessation in adults: a cohort study. Health Educ Res 2006 Apr;21(2):296-302. [doi: 10.1093/her/cyh068] [Medline: 16286480]
21. Wakefield MA, Spittal MJ, Yong HH, Durkin SJ, Borland R. Effects of mass media campaign exposure intensity and durability on quit attempts in a population-based cohort study. Health Educ Res 2011 Dec;26(6):988-997. [doi: 10.1093/her/cyr054] [Medline: 21730252]
22. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 9;376(9748):1261-1271. [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]
23. Bala MM, Strzeszynski L, Topor-Madry R. Mass media interventions for smoking cessation in adults. Cochrane Database Syst Rev 2017 Nov 21;11(11):CD004704. [doi: 10.1002/14651858.CD004704.pub4] [Medline: 29159862]
24. Bardi A, Schwartz SH. Values and behavior: strength and structure of relations. Pers Soc Psychol Bull 2003 Oct;29(10):1207-1220. [doi: 10.1177/0146167203254602] [Medline: 15189583]
25. Wilson KG, Hayes SC, Gifford EV. Cognition in behavior therapy: agreements and differences. J Behav Ther Exp Psychiatry 1997 Mar;28(1):53-63. [doi: 10.1016/s0005-7916(96)00043-2] [Medline: 9110239]
26. Zhang CQ, Leeming E, Smith P, et al. Acceptance and commitment therapy for health behavior change: a contextually-driven approach. Front Psychol 2017;8:2350. [doi: 10.3389/fpsyg.2017.02350] [Medline: 29375451]
27. McClure JB, Anderson ML. Evaluation of a population-level strategy to promote tobacco treatment use among insured smokers: a pragmatic, randomized trial. BMC Public Health 2018 Feb 8;18(1):228. [doi: 10.1186/s12889-018-5119-0] [Medline: 29422026]
28. Chittamuru D, Daniels R, Sarkar U, Schillinger D. Evaluating values-based message frames for type 2 diabetes prevention among Facebook audiences: Divergent values or common ground? Patient Educ Couns 2020 Sep 2. [doi: 10.1016/j.pec.2020.08.042] [Medline: 32938563]
29. Heley K, D'Angelo H, Oh A, et al. Understanding associations of personal values with support for tobacco and alcohol control policies. Am J Prev Med 2023 Sep;65(3):448-457. [doi: 10.1016/j.amepre.2023.03.010] [Medline: 36933752]
30. Key facts as india surpasses china as the world's most populous country. Per Research Center. 2023. URL: https://www.pewresearch.org/short-reads/2023/02/09/key-facts-as-india-surpasses-china-as-the-worlds-most-populous-country/#:~:text=Other%20sources%20of%20information%20for,the%20UN%20population%20data%20begins [accessed 2024-10-01]
31. Singh SK, Kumar S, Kashyap GC. Tobacco use and cessation among a nationally representative sample of men in India, 2019-2021. J Smok Cessat 2023;2023:4292647. [doi: 10.1155/2023/4292647] [Medline: 37006795]

XSL•FO
RenderX

32. Schwartz SH, Cieciuch J, Vecchione M, et al. Value tradeoffs propel and inhibit behavior: validating the 19 refined values in four countries. Euro J Social Psych 2017 Apr;47(3):241-258 [FREE Full text] [doi: 10.1002/ejsp.2228]
33. Tata institute of social sciences (TISS) mamohafw, india. Go Global Adult Tobacco Survey GATS 2 India 2016-17. 2018. URL: https://ntcp.mohfw.gov.in/assets/document/surveys-reports-publications/Global-Adult-Tobacco-Survey-Second-Round-India-2016-2017.pdf [accessed 2025-07-18]
34. Bricker JB, Mull KE, McClure JB, Watson NL, Heffner JL. Improving quit rates of web-delivered interventions for smoking cessation: full-scale randomized trial of WebQuit.org versus Smokefree.gov. Addiction 2018 May;113(5):914-923. [doi: 10.1111/add.14127] [Medline: 29235186]
35. WPP Media. URL: https://www.groupm.com [accessed 2025-06-17]
36. Bricker JB, Mull KE, Santiago-Torres M, et al. Smoking cessation smartphone app use over time: predicting 12-month cessation otcomes in a 2-arm randomized trial. J Med Internet Res 2022 Aug 18;24(8):e39208. [doi: 10.2196/39208] [Medline: 35831180]
37. Rich media gallery. April 2013-June 2013 CTR standard media benchmarks. 2015. URL: http://www.richmediagallery.com/tools/benchmarks [accessed 2025-07-18]
38. McCaul KD, Hockemeyer JR, Johnson RJ, et al. Motivation to quit using cigarettes: a review. Addict Behav 2006 Jan;31(1):42-56. [doi: 10.1016/j.addbeh.2005.04.004]
39. Sanders A, Robinson C, Taylor SC, et al. Using a media campaign to increase engagement with a mobile-based youth smoking cessation program. Am J Health Promot 2018 Jun;32(5):1273-1279. [doi: 10.1177/0890117117728608] [Medline: 28925292]
40. Chan L, O'Hara B, Phongsavan P, Bauman A, Freeman B. Review of evaluation metrics used in digital and traditional tobacco control campaigns. J Med Internet Res 2020 Aug 11;22(8):e17432. [doi: 10.2196/17432] [Medline: 32348272]

Abbreviations

ACT: acceptance and commitment therapy
CTR: click-through rate
LMICs: low- and middle-income countries
RCT: randomized controlled trial
SFT: SmokeFreeTXT

Edited by L Buis; submitted 22.01.25; peer-reviewed by A Eisingerich, K Kaphingst; revised version received 23.05.25; accepted 04.06.25; published 28.07.25.

Please cite as:
Bricker JB, Santiago-Torres M, Mull KE, Sullivan BM, Mehrotra R
Population-Level Dissemination of a Smoking Cessation Smartphone App: Quasi-Experimental Comparison of Values-Based Messages in Social Media Advertisements
JMIR Mhealth Uhealth 2025;13:e71619
URL: https://mhealth.jmir.org/2025/1/e71619
doi:10.2196/71619

© Jonathan B Bricker, Margarita Santiago-Torres, Kristin E Mull, Brianna M Sullivan, Ravi Mehrotra. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Leveraging Smartphone Mobility Data to Understand HIV Risk Among Rural South African Young Adults: Feasibility Study

Thulile Mathenjwa[1,2], MA; Elphas Luchemo Okango[1], PhD; Khai Hoan Tram[3], MD; Maxime Inghels[4,5], PhD; Diego Cuadros[6], PhD; Hae-Young Kim[7], PhD; Fiona Walsh[8], MPH; Till Barnighausen[1,8], PhD; Adrian Dobra[9*], PhD; Frank Tanser[1,10,11*], PhD

[1]Africa Health Research Institute, Krith Building, Third Floor, 719 Umbilo Rd, Durban, South Africa

[10]Centre for Epidemic Response and Innovation, School of Data Science and Computational Thinking, Stellenbosch University, Stellenbosch, South Africa

[11]South African Centre for Epidemiological Modelling and Analysis (SACEMA), School of Data Science and Computational Thinking, Stellenbosch University, Stellenbosch, South Africa

[2]Faculty of Health Sciences, School of Public Health, Division of Epidemiology & Biostatistics, University of Cape Town, Cape Town, South Africa

[3]Department of Medicine, University of Washington, Washington, WA, United States

[4]Lincoln International Institute for Rural Health, University of Lincoln, Lincoln, United Kingdom

[5]Centre Régional de Recherche et de Formation à la Prise en Charge Clinique de Fann, Hôpital Fann, Dakar, Senegal

[6]Digital Epidemiology Laboratory, University of Cincinnati, Cincinnati, OH, United States

[7]Department of Population Health, New York University School of Medicine, New York, NY, United States

[8]Medical Faculty and University Hospital, Heildelberg University, Heidelberg Institute of Global Health, Heidelberg, Germany

[9]Department of Statistics, University of Washington, Washington, WA, United States

[*]these authors contributed equally

Corresponding Author:
Thulile Mathenjwa, MA
Africa Health Research Institute, Krith Building, Third Floor, 719 Umbilo Rd, Durban, South Africa

Abstract

Background: Smartphones provide a precise method to study human mobility at an unprecedented scale, allowing researchers to explore the links between mobility, HIV risk, and treatment outcomes. However, leveraging smartphone technology to study HIV risk in rural settings presents unique challenges and opportunities.

Objective: This study assessed the feasibility of using smartphone GPS technology to collect mobility data from young adults in rural KwaZulu Natal, South Africa. We also present key lessons learned during the study.

Methods: The study was conducted in 2 phases (June 2021-May 2023) with males and females aged 20 - 30 years old. In phase I, participants received smartphones with a customized study app (Avicenna research software). In phase II, they used their personal smartphones and installed the study app. The app used Android location services to record the smartphone location every 30 minutes and send it to a secure study server hourly. Participants were followed up for 6 months (26 wk). If location data were missing for 48 - 72 hours, participants were contacted for troubleshooting. Engagement strategies, including reverse billing and gamification (Wheel of Fortune), were implemented to address internet connection barriers and aid data collection.

Results: A total of 207 participants were enrolled (phase I: 163; phase II: 44) with 204 providing mobility data. Participants recorded 27.6 million location points with a median number of 74,865 (IQR 28,471 - 186,578) per participant. The mean weekly location points recorded was 95.3 out of 336 possible half-hour intervals (28.4%). Phase II saw more stable data collection in the latter half of the study, due to increased user engagement with the app. Challenges included phone-related issues (screen malfunctions, lost and broken phone), app terminations, and limited internet connectivity. Reverse billing and gamification strategies improved location data collection through increased user engagement.

Conclusions: This study demonstrates that the use of smartphone-based GPS technology is feasible among young adults in a rural South African setting. Although only 28.4% (95.3/336) of expected weekly location data were collected, the study offers insights into engagement strategies that can be used to enhance location data collection in similar contexts. Continuous troubleshooting identified challenges and informed solutions to data collection gaps. Reverse billing system and gamification resulted in significant increases in location data received. These findings underscore the potential of integrating mobile health tools into health systems to better support high-risk mobile populations.

(JMIR Mhealth Uhealth 2025;13:e67519) doi:10.2196/67519

XSL•FO
RenderX

KEYWORDS

geographic mobility; migration; HIV; global positioning system; smartphone; young people; GPS

Introduction

Mobility and HIV Prevention and Care Outcomes

The scale-up of antiretroviral therapy (ART) has significantly increased adult life expectancy for people living with HIV and reduced HIV transmission [1-3]. Despite this, in 2022, there were 1.3 million new HIV infections and 630,000 HIV-related deaths globally, with the majority occurring in sub-Saharan Africa [4]. Migration, including short-distance and frequent geographic relocations, is sustaining the HIV epidemic [5,6]. Previous studies indicate that highly mobile individuals face a greater risk of acquiring HIV [7-9] and once living with HIV, are less likely to adhere to treatment or achieve viral suppression compared to less mobile individuals [10,11]. Consequently, highly mobile people are at an increased risk of developing AIDS [12] and dying from HIV-related causes [13]. In addition, mobility is linked to high-risk behaviors (eg, multiple and concurrent partners and substance use), which can increase the risk of HIV transmission if individuals are not virally suppressed [14-17]. For example, the treatment and prevention trial in rural KwaZulu-Natal showed that the new HIV infections among transient individuals within the trial communities slowed down efforts to increase ART coverage and population viral load suppression [18]. Therefore, understanding mobility patterns, including destinations, duration, and frequency, is important to inform the development of effective interventions to improve HIV prevention and treatment outcomes.

Potential of Location-Aware Technologies

Location-aware technologies, such as smartphones, offer simple and accurate tools for capturing both spatial and temporal dimensions of human mobility. A growing body of research has used portable GPS devices, smartwatches, and phones to examine relationships between mobility and health-related outcomes [19-21]. These studies reported sufficient acceptability among participants for using such technologies. However, there were challenges with standalone GPS devices, including satellite signal loss [22], the burden of carrying devices [21], complaints about discomfort or appearance [19], privacy and confidentiality concerns [20], and the need for active user effort to facilitate data collection [23]. Familiar technologies such as smartphones and wearables can mitigate some of these challenges. For instance, potential participants in rural South Africa perceived smartphone GPS valuable and appealing for studying human mobility and HIV risk due to their familiarity with smartphones [24]. Another United States study tested the use of smartphones and smartwatches in chronically ill patients with noncommunicable diseases [23]. Participants preferred smartphones because of ease of use. Also, there was more complete location data from the smartphones compared to smartwatches [23]. However, this study only assessed short-term use (28 days) and focused on noncommunicable diseases, which are typically less stigmatized than infectious diseases like HIV.

We conducted a pilot study to assess the feasibility of using smartphone GPS technology to study linkages between human mobility and poor HIV outcomes among young adults in rural KwaZulu-Natal, South Africa. Young adults in this setting are highly mobile [7] and experience complex and intersecting vulnerabilities (poor risk perception, economic hardships, and lack of social support), increasing their risk of poor HIV care outcomes [14,25]. This paper evaluates the long-term use of smartphone GPS for collecting mobility data among these young adults. We also present key lessons learned from this study.

Methods

Study Setting

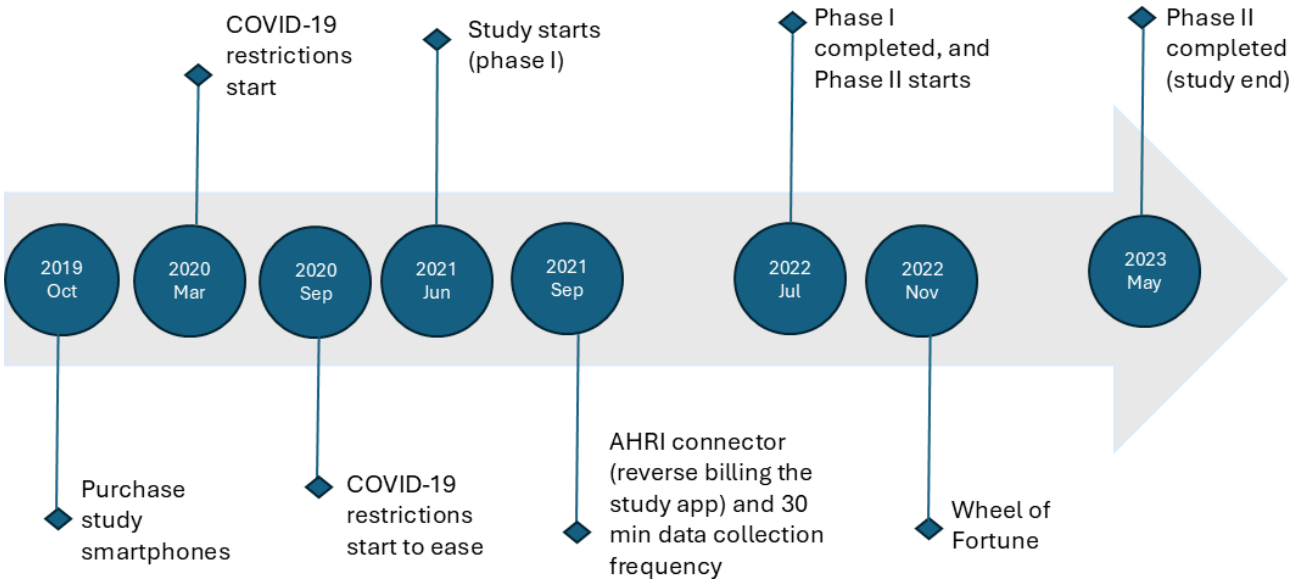
This study was conducted in uMkhanyakude district, northern KwaZulu-Natal, South Africa, using the Africa Health Research Institute (AHRI) population-based HIV surveillance program. The study area is predominantly rural, with limited access to water and electricity [26]. Internet connectivity is low, mainly accessed via mobile phones on a prepaid basis. Over 85% of youth aged 20 - 24 are unemployed [27], and 20% of young adults migrate annually for economic and educational opportunities [7]. The area has high HIV burdens with an incidence rate of 8% among women aged 20 - 24 and 4% among men aged 25 - 29 years [28].

Study Design and Dates

The study was conducted in 2 phases from June 2021 to May 2023 (see Figure 1). In phase I, participants received an Android smartphone (Proline Falcon XL, Proline, South Africa) with a customized Avicenna app. These smartphones, purchased in October 2019, were chosen for their low cost and compatibility with a proposed app developed for the study. However, this app was found unusable in March 2020, leading to delays in the study to search for an alternative. The COVID-19 pandemic led to further delays, and the study began in June 2021. In phase II, participants used their personal smartphones to address operational challenges faced in phase I.

XSL•FO
RenderX

Figure 1. Sesikhona study timelines. AHRI: Africa Health Research Institute.



Recruitment, Enrollment, and Follow-Up

Individuals were eligible if they: (1) were aged 20 to 30 years old; (2) participated in the 2019 annual AHRI HIV surveillance round; (3) resided in the southern AHRI HIV surveillance area; (4) were willing to participate in the study; and (5) owned a compatible smartphone (minimum random access memory [RAM]: 1GB, sufficient free space for study app installation - phase II only). Eligibility was not limited by HIV status; the study aimed to include both those living with HIV and those who were not. The age criterion excluded individuals aged 18-19 years because they were more likely to be in school, raising parental concerns about smartphone use.

Using the 2019 AHRI HIV surveillance program population as a sampling frame, we generated a random sample of 1467 eligible individuals. Fieldworkers contacted them by telephone to introduce the study, assess eligibility, and interest. Data collection was conducted using Android tablets. Interested individuals were scheduled for a home visit. During the home visit, fieldworkers provided more comprehensive study information, including an informational video addressing privacy and confidentiality concerns identified in preliminary research [24]. The video used animations to explain GPS technology, study procedures, and data protection measures. Participants had the opportunity to ask questions before giving informed consent. After obtaining informed consent, participants were guided through the app installation (phase II) and registration. They were also provided with instructions on the app's functionality and usage. Participants were followed for 6 months (26 weeks).

Participants received mobile data bundles (phase I: 1GB, phase II: 2GB) at enrollment and monthly thereafter. A portion of the data was intended for app connectivity with the study server, but this was not implemented due to changes with the app service providers. In phase II, monthly mobile was provided only if participants provided location data for at least 10 days. Figure 2 illustrates the study enrollment, participation, and follow-up process.

XSL•FO
RenderX

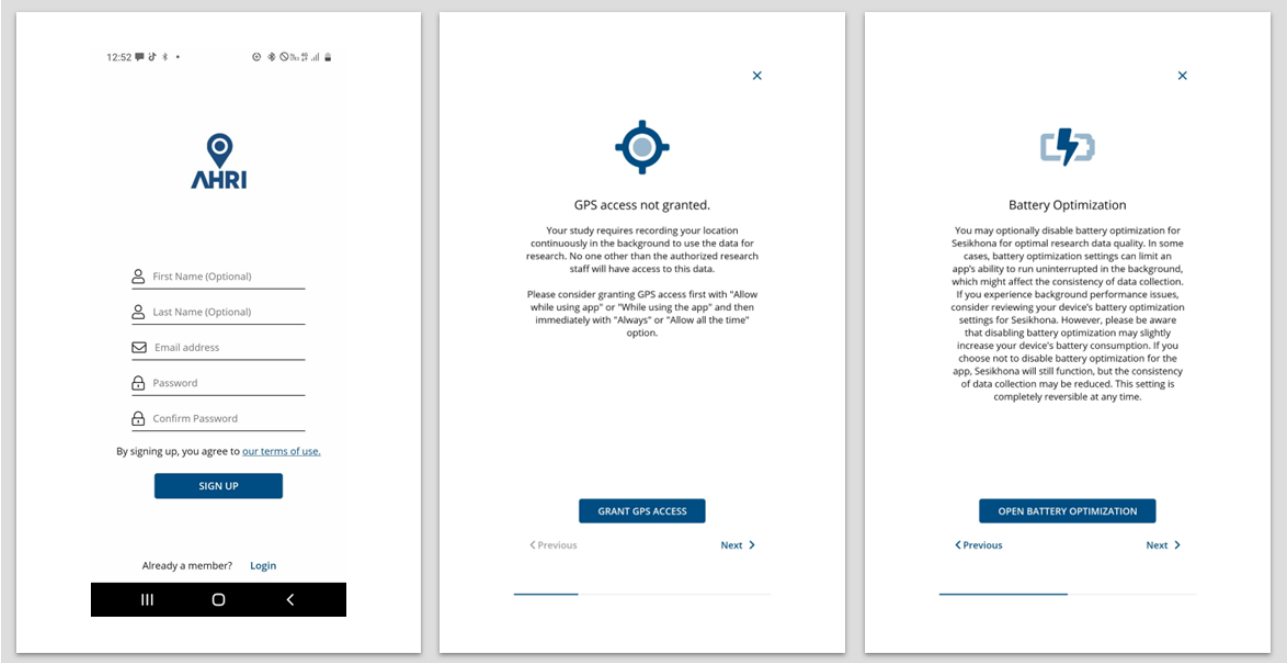
Figure 2. Description of the study enrollment and follow-up process.

Study App and Data Collection

We used the Avicenna software platform [29], customized into a smartphone app called "Sesikhona." The Sesikhona app, available on Google Play Store, used Android location services to capture GPS coordinates of the device using a combination of Wi-Fi routers, cell towers, and GPS satellites. Initially, the app collected location data every 5 minutes but was later adjusted to 30 minutes (refer to Figure 1), to capture at least 48 location points per smartphone daily. For accuracy, the app recorded GPS data over 60 seconds until 3 precise location points were obtained.

For the app to collect location data, participants needed to: (1) grant the app GPS access; (2) enable location services; and (3) exclude the app from battery optimization (see Figure 3). Participants could revoke these permissions at any time to prevent location data collection. The app uploaded collected location data hourly to a secure study server hosted by Avicenna. If the smartphone lacked internet connection, data was supposed to be stored locally and uploaded during the next scheduled cycle. The data were then transferred from Avicenna to a MySQL database at AHRI.

Figure 3. Screenshots of the Sesikhona app showing the registration page, GPS, and battery optimization permissions.



XSL•FO
RenderX

App Registration, Troubleshooting, and App Updates

To ensure confidentiality, the app did not collect personal identifying information. Instead, participants were registered with an alias, comprising an AHRI internal number, a study-specific email address, and password. The fieldworker aided the initial app setup by tethering the smartphone to their tablet for internet access. To ensure data integrity, participants were contacted by phone or visited at home if location data was missing for 48 - 72 hours. Troubleshooting involved checking app settings such as location permissions, battery optimization, and internet connectivity using REDCap (version 14.044). Fieldworkers used tablets as temporary internet hotspots, enabling the app to resume data transmission.

App updates were managed in 2 processes. Study modifications, such as changes to location data collection frequency, were automatically pushed from the Avicenna platform to participants' devices in the background. These updates required an internet connection. If offline, participants had to manually initiate the update by selecting the "reload studies from the server" option in the app. Updates to the Sesikhona app, including performance enhancements or bug fixes, were provided through the Google Play Store. Participants had to manually approve and install these updates. Initially, fieldworkers conducted home visits for assistance but later provided remote support via phone calls and data bundles to ensure internet access. App updates were discussed in weekly debriefing meetings with the project operations team, reviewing recruitment, enrollment and follow-up progress, technical challenges, and study procedures. Notes from these meetings informed revisions and improvements.

Data Analysis

We report on participants' enrollments, key demographics, HIV status at enrollment, age group, education years, and socioeconomic status. Feasibility was assessed using collected location data, including the number of location data points and weekly averages per participant. Also, we reviewed project operation meeting notes to construct a narrative of our study implementation experiences. This review highlights challenges in data collection, like phone replacement and device settings, and strategies to improve data collection.

Ethical Considerations

The study was designed with special attention to ethical considerations related to location data collection. We conducted formative research and applied the Emanuel ethical framework to identify the ethical issues and used the findings to inform the development of the informed consent materials and overall study design [24]. The study was approved by the Biomedical Research Ethics Committee of the University of KwaZulu Natal (BREC Ref 460/15). All participants provided written informed consent before participation, and study participation was entirely voluntary. To ensure privacy, all data were deidentified, with no personal identifying information in the final dataset. Participants received smartphones in phase I to keep after the study, along with data bundles in both phases as incentives for participation.

Results

We first describe participant characteristics, then summarize the collected location data. Finally, we discuss the operational challenges faced and opportunities identified during the study implementation.

Enrollment and Participant characteristics

A total of 516 individuals were contacted (including 88 in phase II), of whom 207 consented to participate (phase I: 163 and phase II: 44). Of the 207 participants, 3 did not provide location data (Table 1): one due to incomplete registration and 2 due to device incompatibility. Among those who provided location data, 131 out of 204 (64.2%) were female. Most participants, 95.2% (160 out of 168 that reported education) completed 10 to 17 years of education, and 39 out of 204 (19%) were living with HIV. Regarding age distribution, 80 out of 204 (39.2%). However, phase II participants were slightly older with a higher socio-economic status (Table 1).

XSL•FO
RenderX

Table . Demographic characteristics of study participants[a].

Characteristic	Phase I (n=162), n (%)	Phase II (n=42), n (%)	Total (n=204), n (%)
Sex			
Female	99 (61.1)	32 (76.2)	131 (64.2)
Male	63 (38.9)	10 (23.8)	73 (35.8)
HIV status			
Negative	129 (80.1)	35 (83.3)	164 (80.8)
Positive	32 (19.9)	7 (16.7)	39 (19.2)
Missing	1	0	1
Age			
20 - 24	68 (42.0)	12 (28.6)	80 (39.2)
25 - 30	94 (58.0)	30 (71.4)	124 (60.8)
Years of education			
[0,10]	7 (5.4)	1 (2.6)	8 (4.8)
[10,17]	123 (94.6)	37 (97.4)	160 (95.2)
Missing	32	4	36
Socioeconomic Index[b]			
[1–2]	7 (4.5)	3 (7.3)	10 (5.1)
[2–3]	103 (66.9)	21 (51.2)	124 (63.6)
[3–4]	44 (28.6)	17 (41.5)	61 (31.3)
Missing	8	1	9

[a]Excluding the 3 that did not contribute location data.

[b]Socioeconomic index with [1-2] representing very low socio-economic status, [2-3] low socioeconomic status and [3-4] lower-middle socioeconomic status.

Location Data

A total of 204 participants recorded 27,589,855 location points, with a median number of 74,865 (IQR 28,471 - 186,578) points per participant. Participants provided a median of 2906 (IQR 856 - 6688) location points weekly, though consistency varied. Ideally, location data were collected every 30 minutes, equating to a minimum of 336 location points weekly (48 daily). However, on average, participants collected 95.3 out of the expected 336 points (28.4%) weekly. The range and volume of location data varied by participant (Figure 4).

In phase I, weekly location data peaked at 45.4% (SD 29.7%; 152.5/336) in Week 2 and declined by more than half to 21.5% (SD 22.2%; 72.3/336) by Week 26. In phase II, data collection fluctuated with participant engagement, increasing in the latter half of the study and peaking at 43.3% (SD 36.5%; 145.5/336) in Week 26. However, the overall difference between the 2 phases was 28.3% (SD 15.6%; 95.0/336) versus 29.0% (SD 26.6%; 97.4/336) was not statistically significant (P=.82) using a 2-sample t test. (Figure 5).

XSL•FO
RenderX

Figure 4. Range and density of weekly location data.

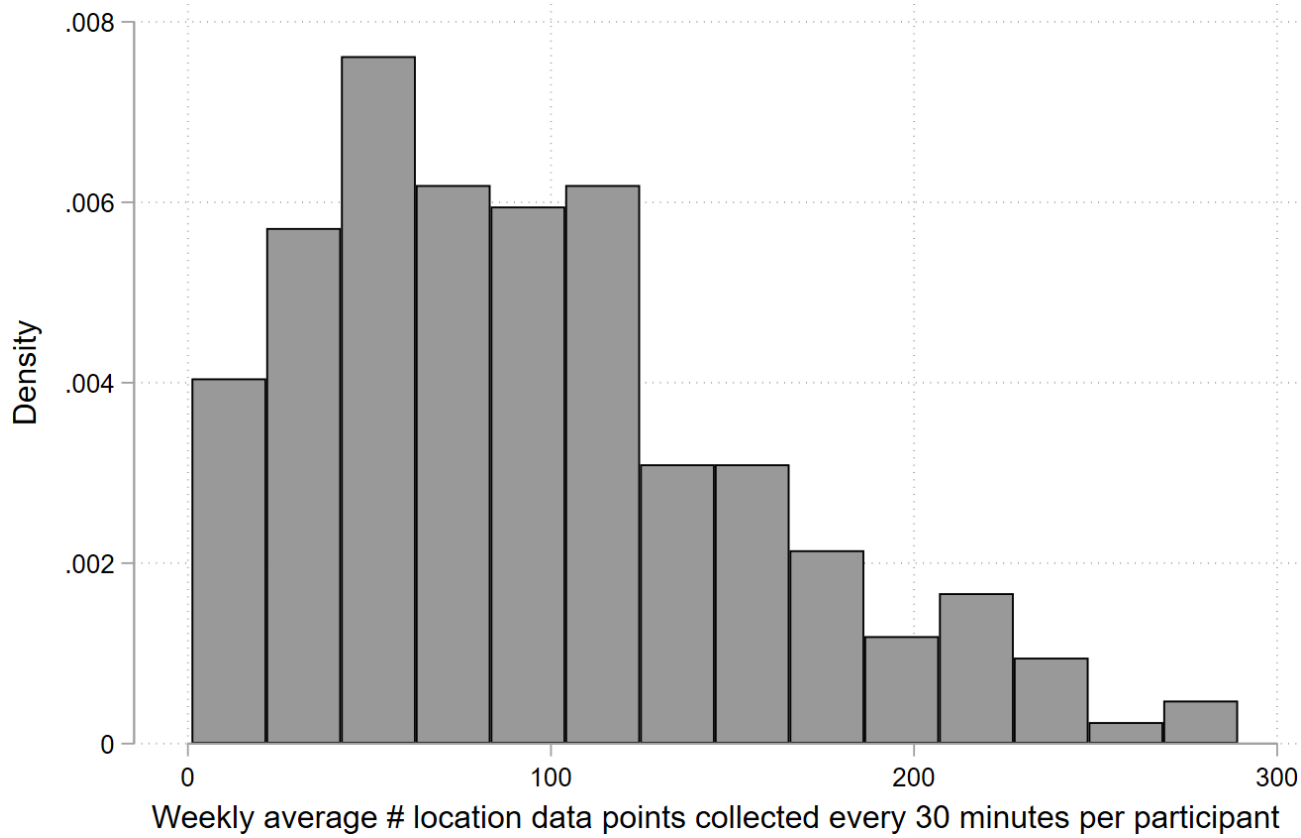
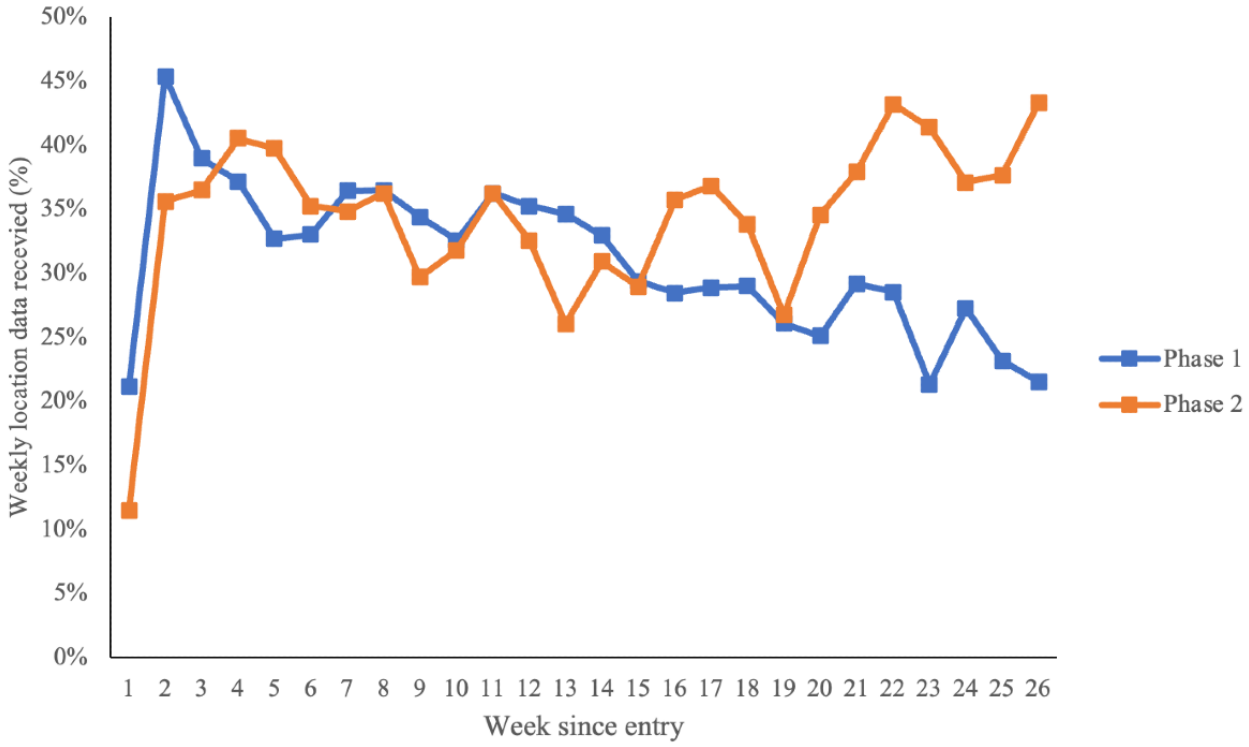


Figure 5. Mean percentage of location data points received per week by phase.



Operational Challenges and Opportunities

Several operational challenges emerged during phase I of the study. A major issue was the app's reliance on internet connectivity to upload location. In the first month, we found that many participants quickly exhausted their prepaid data bundles, leading to gaps in location data. To address this, we partnered with Datafree technologies [30] to implement a reverse-billed app (AHRI connector) which allowed the Sesikhona app to upload collected location data without an

XSL•FO
RenderX

active data bundle on participants' devices. However, deploying the AHRI connector app (available as an APK file) required home visits to install via Bluetooth. While this was a logistical challenge, it presented an opportunity to initiate the study update to the increased 30-minute data collection frequency. Reverse billing initial data bundle balance requirements were bypassed by tethering participant devices to the fieldworker tablet. Another challenge stemmed from hardware limitations. The study smartphones, purchased in 2019, faced significant technological issues due to their prolonged storage and aging hardware. Common malfunctions included frozen touchscreens, loss of network signal, and hardware failures, resulting in 77 phone replacements, with some 20 participants requiring multiple replacements (refer to Multimedia Appendix 1). Some participants lost or broke their phones. Notably, 16 participants opted to install the app on their personal phones, demonstrating early acceptability of personal phone use for the study.

In addition, the app was prone to frequent app terminations resulting in missing location data. Despite participants maintaining correct app settings, including 90% (324/360) with location services enabled (on), 95.5% (345/361) with battery optimization disabled, and 76.8% (274/357) with AHRI connector on (refer to Multimedia Appendix 2), the app was not collecting location data. Subsequent analysis by the Avicenna team determined that these disruptions were due to the smartphone operating system background task management, which automatically terminated apps running in the background, including Sesikhona, to optimize system performance for active apps.

Adaptations and Phase II

In response to the phone-related challenges and participants' preference for switching to personal phones, phase II adopted a revised approach. Unlike phase I, where the study provided smartphones, phase II required participants to install the Sesikhona app on their own devices. Of the 44 participants who consented, 2 had smartphones with 512 MB RAM, insufficient for the app. This issue was not initially identified, but discussions with Avicenna established that at least 1 GB RAM was needed for optimal app performance. Similarly to phase I, participants reported losing (n=2) or breaking (n=2) their phones, and 8 participants switched to different phones. In addition, a few participants changed their phone numbers. Nine participants deleted the app across both phases (5 in phase I and 4 in phase II), 6 reinstalled it, and 3 were no longer interested. Reasons for deletions included unintentional removal by children, app performance issues, or insufficient storage. Some intended to reinstall but were unable to due to forgotten login details. Fieldworkers provided reinstallation assistance and technical support to minimize data collection disruptions.

Location data gaps persisted in phase II despite correct app settings. To enhance user engagement and reduce app terminations, we introduced a gamification (Wheel of Fortune) in November 2022. This daily lottery game (available from 9 AM to 4 PM) with additional plays on Saturdays offered participants an opportunity to win prizes like airtime, data bundles, and shopping vouchers (see Figure 6).

Engaging with the game was positively correlated with the frequency of location data collection. The probability of participants providing location data was 77.6% days when they played the game, versus 27.5% if they did not play (refer to Multimedia Appendix 3). The odds ratio was 9.12 (95% CI 5.44 - 15.29) with a $P<.001$. In addition, playing the game was associated with +15.7% (7.5/48; $P=.007$) increase in the amount of location data received per day (refer to Multimedia Appendix 4).

Table 2 summarizes the key findings from phase I and II. While phase II had few participants, having AHRI connector at the outset and introduction of the gamification enhanced user engagement and data collection. Overall, there were fewer technical challenges reported in phase II compared to phase I.

XSL•FO
RenderX

Figure 6. Wheel of Fortune aimed to increase user engagement with the app. Screen 1: how it appears on the Sesikhona app homepage; screen 2: spinning wheel with an example of a win; and screen 3: the prizes.

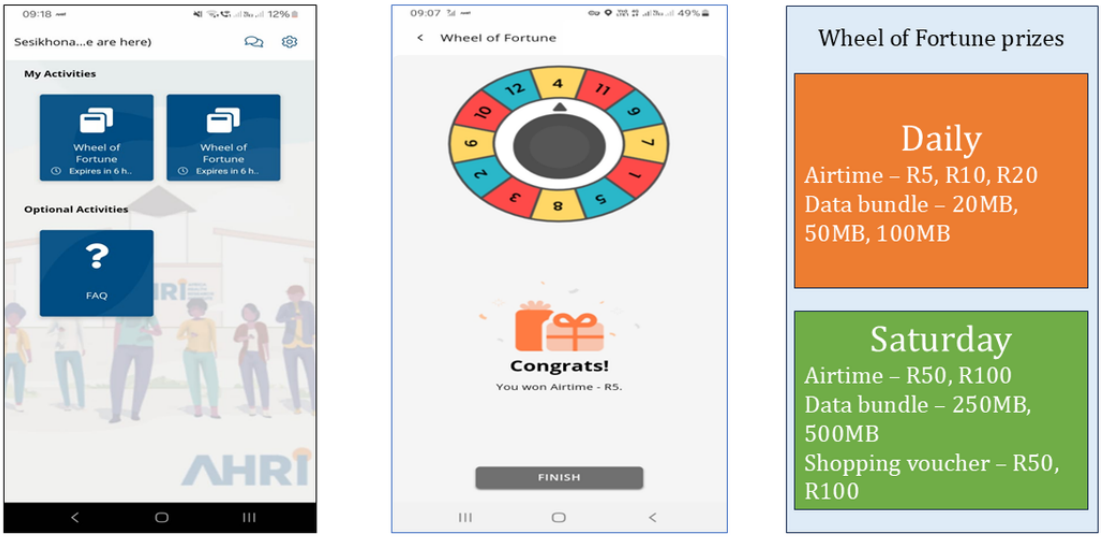


Table . Summary of phase I and phase II results.

Parameter	Phase I (June 2021 to July 2022)	Phase II (July 2022 to May 2023)
Participants enrolled, n	163	44
Participants contributing location data, n	162	42
Average number of location points/participant, median (IQR)	67,398 (29,771 - 168961)	47,510 (12,252 - 172,613)
Range of location points/participant	389 - 670,384	53 - 950,597
Percentage of location points per week (out of 336 half-hour intervals), mean (SD)	28.3 (15.6)	29.0 (26.6)
Introduction of AHRI[a] connector (reverse billing app)	Implemented midway	Implemented at the start of the phase
Phone replacements	77 phones replaced due to technical issues	2 participants' phones were incompatible, with 4 reporting broken/lost phones
Gamification (Wheel of Fortune) impact	N/A[b]	Increased probability of receiving location data by 77.6% if the participant played the game, versus 27.5% if they did not play and associated with +15.7% more data received in a per day compared to not playing (*P*=.007)
Challenges faced	Aging hardware, data bundle exhaustion, app termination	Some phone compatibility issues, but fewer technical problems overall
Engagement strategy	Relied on troubleshooting and participant support	Gamification introduced to enhance engagement and data collection

[a]AHRI: Africa Health Research Institute.

[b]not applicable.

Discussion

Principal Findings

In this study, we found that smartphone-based GPS technology is a feasible approach for studying human mobility in rural settings using commercial software, provided that key implementation challenges are addressed. Individuals living with HIV and without accepted both study-provided and personal smartphone use for location data collection. Although most participants complied with study protocols, only 28.4% (95.3/336) of the expected weekly location data were collected,

XSL•FO
RenderX

largely due to operational challenges such as hardware malfunctions, lack of internet connection, and frequent app terminations, particularly during phase I. Frequent troubleshooting provided insights into context-specific adaptations needed for rural settings, including implementation of the reverse billing to facilitate data transmission with an active data bundle.

The introduction of gamification in phase II significantly increased the likelihood of receiving location data from participants who engaged with the game. This highlights the potential of gamified strategies to enhance participant engagement and improve data completeness in long-term studies. Notably, phase II participants were slightly older and from higher income brackets, raising important considerations for equity in future studies. Requiring smartphone ownership could exclude younger or lower-income populations, who may be at higher risk for HIV and stand to benefit most from mobility-informed interventions.

Comparison to Prior Work

Our results align with previous studies demonstrating feasibility and acceptance of smartphone GPS apps to understand HIV risk [31-33], substance use and violence among people living with HIV [34]. The lower median data collection of 28.4% highlights challenges in rural settings. Two South African smartphone-based studies reported significant technological and operational issues with using personal phones, where inadequate RAM, storage, and device type affected implementation [35,36]. Furthermore, Clouse et al noted how lack of internet contributed to missing location data [36].

Although Hardy et al [23] found that smartphones provided more complete location data with less user effort, our study showed that the app required significant user effort to collect data. The high number of days with missing location data, especially later in the study, is comparable to findings from studies using standalone GPS devices [20,21]. However, these studies had smaller sample sizes, shorter follow-up periods, and different settings. Finally, other research shows that game elements enhance user engagement in mobile apps [37].

Implications for Future HIV Research and Interventions

The mobility data from this study will provide a granular understanding of how mobility intersects with HIV risk. Identifying geographic hotspots and mobility patterns linked to high-risk behaviors, such as multiple concurrent partnerships or inconsistent condom use, can enable more targeted and efficient allocation of prevention resources. For example, interventions can be tailored to specific locations and times where risk is highest to maximize impact.

Mobility data can also reveal disruptions in HIV care, such as missed clinic appointments due to travel, informing the design of tailored interventions like mobile clinics or targeted outreach programs at critical locations. Integrating mobile health tools, like the Sesikhona app, into broader health systems offers opportunities to prospectively monitor human mobility and deliver real-time support to individuals at risk of disengagement. For example, automated reminders or check-ins could be triggered on an individual's location to encourage care engagement or treatment adherence protocols. However, for these strategies to be effective, technological and access barriers must be addressed, and the interventions should be co-designed with communities to align with privacy expectations and ethical standards. Overcoming these challenges will be critical to leveraging mobility-informed approaches to achieving the UNAIDS 95-95-95 targets, particularly among highly mobile and hard-to-reach populations.

Study Strengths and Limitations

This study provides important evidence on the feasibility of using commercially available smartphone GPS apps to collect mobility data in a rural, resource-limited setting. While the high adoption rate and engagement demonstrate potential, feasibility was constrained by technological and operational challenges related to device and internet connectivity. These findings suggest that use of smartphone-based GPS to study mobility is feasible if key barriers are addressed. However, generalizability may be limited. The study was conducted within a cohort already engaged through the AHRI HIV population-based surveillance program, where established trust likely enhanced participant acceptance. In addition, reliance on smartphone technology may have introduced access and literacy barriers, which could affect scalability in other settings.

Conclusion

This study provides critical insights into the feasibility of using smartphone-based GPS technology to collect mobility data among young adults in a rural, HIV-hyperendemic setting in South Africa. By comparing the outcomes of 2 distinct phases, the study highlights the success and challenges of using both study-provided and personal smartphones. Although the collected data was lower than expected, high adoption rates, sustained engagement, and the success of solutions such as reverse billing and gamification demonstrate the potential of smartphone-based approaches to study mobility, provided technological barriers are addressed.

The implications of this study extend beyond its immediate context, offering valuable lessons for the broader field of HIV research and intervention. Our findings underscore the potential of integrating mobile health tools into health systems to support high-risk mobile populations. Mobility-informed approaches could enhance real-time monitoring, targeted service delivery, and ultimately improve HIV prevention and care outcomes. Further research is needed to refine engagement strategies, address technological and access barriers, and optimize data integration methods to fully leverage mobility data for achieving global HIV epidemic control.

XSL•FO
RenderX

Acknowledgments

The authors would like to acknowledge the study participants for their time. We wish to appreciate Ms Zinhle Mkhwanazi, Mr Velaphi Gumede, Mrs Nqobile Nkosi, Ms Onie Mgu, Mrs Neli Mtshali, and Mr Mareca Sithole for participant recruitment, data collection, and data management. We would like to acknowledge Mr Brendan Gilbert for the suggestion of a lottery game. This study was supported by grants from the National Institutes of Health (R01HD084233), Academy of Medical Sciences Newton Advanced Fellowship, and German research foundation (BA2067/14-1). The Africa Health Research Institute's (AHRI) Demographic Surveillance Information System and Population Intervention Program is funded by the Wellcome Trust (227167/A/23/Z) and the South Africa Population Research Infrastructure Network (funded by the South African Department of Science and Technology and hosted by the South African Medical Research Council). FT and TM are also supported by the National Institute of Mental Health (R01MH131480). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding bodies.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to them containing sensitive information that could compromise the privacy of research participants but are available from the principal investigator (FT) on reasonable request.

Authors' Contributions

TM: conceptualization, methodology, formal analysis, project administration, writing – original draft, writing – review & editing. EO: formal analysis, methodology, data curation, software, validation, visualization, writing – original draft, writing – review & editing. KHT: formal analysis, methodology, data curation, software, validation, visualization, writing – original draft, writing – review & editing. MI: writing – original draft, writing – review & editing. DC: writing – original draft, writing – review & editing. HYK: writing – original draft, writing – review & editing. FW: writing – original draft, writing – review & editing. TB: conceptualization, funding acquisition, investigation, writing – original draft, writing – review & editing. AD: conceptualization, investigation, methodology, writing – original draft, writing – review & editing. FT: conceptualization, funding acquisition, investigation, methodology, supervision, writing – original draft, writing – review & editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Phase I phone replacement reasons.
[PNG File, 17 KB - mhealth_v13i1e67519_app1.png]

Multimedia Appendix 2
Phone settings during troubleshooting.
[PNG File, 85 KB - mhealth_v13i1e67519_app2.png]

Multimedia Appendix 3
Probability of receiving location data on a given day.
[PNG File, 54 KB - mhealth_v13i1e67519_app3.png]

Multimedia Appendix 4
Expected percentage of data received (out of 48 daily half-hour blocks).
[PNG File, 59 KB - mhealth_v13i1e67519_app4.png]

References

1. Bor J, Rosen S, Chimbindi N, et al. Mass HIV treatment and sex disparities in life expectancy: demographic surveillance in rural South Africa. PLoS Med 2015 Nov;12(11):e1001905. [doi: 10.1371/journal.pmed.1001905] [Medline: 26599699]
2. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505. [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
3. Tanser F, Bärnighausen T, Grapsa E, Zaidi J, Newell ML. High coverage of ART associated with decline in risk of HIV acquisition in rural KwaZulu-Natal, South Africa. Science 2013 Feb 22;339(6122):966-971. [doi: 10.1126/science.1228160] [Medline: 23430656]
4. UNAIDS 2022 Epidemiological Estimates: UNAIDS; 2022.
5. Camlin CS, Cassels S, Seeley J. Bringing population mobility into focus to achieve HIV prevention goals. J Int AIDS Soc 2018 Jul;21(Suppl 4):e25136. [doi: 10.1002/jia2.25136] [Medline: 30027588]

XSL•FO
RenderX

6. Cuadros DF, de Oliveira T, Gräf T, et al. The role of high-risk geographies in the perpetuation of the HIV epidemic in rural South Africa: A spatial molecular epidemiology study. PLOS Glob Public Health 2022;2(2):e0000105. [doi: 10.1371/journal.pgph.0000105] [Medline: 36962341]
7. Dobra A, Bärnighausen T, Vandormael A, Tanser F. Space-time migration patterns and risk of HIV acquisition in rural South Africa. AIDS 2017 Jan 2;31(1):137-145. [doi: 10.1097/QAD.0000000000001292] [Medline: 27755099]
8. Dzomba A, Tomita A, Vandormael A, Govender K, Tanser F. Effect of ART scale-up and female migration intensity on risk of HIV acquisition: results from a population-based cohort in KwaZulu-Natal, South Africa. BMC Public Health 2019 Feb 14;19(1):196. [doi: 10.1186/s12889-019-6494-x] [Medline: 30764786]
9. Tomita A, Vandormael AM, Bärnighausen T, de Oliveira T, Tanser F. Social disequilibrium and the risk of HIV acquisition: a multilevel study in rural KwaZulu-Natal Province, South Africa. J Acquir Immune Defic Syndr 2017 Jun 1;75(2):164-174. [doi: 10.1097/QAI.0000000000001349] [Medline: 28291049]
10. Thorp M, Ayieko J, Hoffman RM, Balakasi K, Camlin CS, Dovel K. Mobility and HIV care engagement: a research agenda. J Int AIDS Soc 2023 Mar;26(3):e26058. [doi: 10.1002/jia2.26058] [Medline: 36943731]
11. Tomita A, Vandormael A, Bärnighausen T, et al. Sociobehavioral and community predictors of unsuppressed HIV viral load: multilevel results from a hyperendemic rural South African population. AIDS 2019 Mar 1;33(3):559-569. [doi: 10.1097/QAD.0000000000002100] [Medline: 30702520]
12. Ross J, Cunningham CO, Hanna DB. HIV outcomes among migrants from low-income and middle-income countries living in high-income countries: a review of recent evidence. Curr Opin Infect Dis 2018 Feb;31(1):25-32. [doi: 10.1097/QCO.0000000000000415] [Medline: 29095720]
13. Tanser F, Bärnighausen T, Vandormael A, Dobra A. HIV treatment cascade in migrants and mobile populations. Curr Opin HIV AIDS 2015 Nov;10(6):430-438. [doi: 10.1097/COH.0000000000000192] [Medline: 26352396]
14. Ngwenya N, Bernays S, Nkosi B, et al. Making sense of uncertainty: The precarious lives of young migrants from rural Kwazulu-Natal, South Africa. Glob Public Health 2023 Jan;18(1):2229895. [doi: 10.1080/17441692.2023.2229895] [Medline: 37431745]
15. Kwena Z, Nakamanya S, Nanyonjo G, et al. Understanding mobility and sexual risk behaviour among women in fishing communities of Lake Victoria in East Africa: a qualitative study. BMC Public Health 2020 Jun 15;20(1):944. [doi: 10.1186/s12889-020-09085-7] [Medline: 32539818]
16. Okano JT, Busang L, Seipone K, Valdano E, Blower S. The potential impact of country-level migration networks on HIV epidemics in sub-Saharan Africa: the case of Botswana. Lancet HIV 2021 Dec;8(12):e787-e792. [doi: 10.1016/S2352-3018(21)00267-8] [Medline: 34774183]
17. Dzomba A, Tomita A, Govender K, Tanser F. Effects of migration on risky sexual behavior and HIV acquisition in South Africa: a systematic review and meta-analysis, 2000-2017. AIDS Behav 2019 Jun;23(6):1396-1430. [doi: 10.1007/s10461-018-2367-z] [Medline: 30547333]
18. Larmarange J, Diallo MH, McGrath N, et al. The impact of population dynamics on the population HIV care cascade: results from the ANRS 12249 Treatment as Prevention trial in rural KwaZulu-Natal (South Africa). J Int AIDS Soc 2018 Jul;21(Suppl 4):e25128. [doi: 10.1002/jia2.25128] [Medline: 30027600]
19. Duncan DT, Kapadia F, Regan SD, et al. Feasibility and acceptability of global positioning system (GPS) methods to study the spatial contexts of substance use and sexual risk behaviors among young men who have sex with men in New York City: a P18 cohort sub-study. PLoS One 2016;11(2):e0147520. [doi: 10.1371/journal.pone.0147520] [Medline: 26918766]
20. Duncan DT, Chaix B, Regan SD, et al. Collecting mobility data with GPS methods to understand the hiv environmental riskscape among young Black men who have sex with men: a multi-city feasibility study in the Deep South. AIDS Behav 2018 Sep;22(9):3057-3070. [doi: 10.1007/s10461-018-2163-9] [Medline: 29797163]
21. Paz-Soldan VA, Stoddard ST, Vazquez-Prokopec G, et al. Assessing and maximizing the acceptability of global positioning system device use for studying the role of human movement in dengue virus transmission in Iquitos, Peru. Am J Trop Med Hyg 2010 Apr;82(4):723-730. [doi: 10.4269/ajtmh.2010.09-0496] [Medline: 20348526]
22. Williams J, Townsend N, Duncan G, Drewnowski A. Participant experience using GPS devices in a food environment and nutrition study. J Hunger Environ Nutr 2016 Jul 2;11(3):414-427. [doi: 10.1080/19320248.2015.1112754]
23. Hardy J, Veinot TC, Yan X, et al. User acceptance of location-tracking technologies in health research: Implications for study design and data quality. J Biomed Inform 2018 Mar;79:7-19. [doi: 10.1016/j.jbi.2018.01.003] [Medline: 29355784]
24. Mathenjwa T, Nkosi B, Kim HY, Bain LE, Tanser F, Wassenaar D. Ethical considerations in using a smartphone-based GPS app to understand linkages between mobility patterns and health outcomes: The example of HIV risk among mobile youth in rural South Africa. Dev World Bioeth 2023 Dec;23(4):321-330. [doi: 10.1111/dewb.12376] [Medline: 36201706]
25. Bernays S, Lanyon C, Dlamini V, Ngwenya N, Seeley J. Being young and on the move in South Africa: how 'waithood' exacerbates HIV risks and disrupts the success of current HIV prevention interventions. Vulnerable Child Youth Stud 2020 Oct 1;15(4):368-378. [doi: 10.1080/17450128.2020.1739359]
26. Herbst K, Law M, Geldsetzer P, Tanser F, Harling G, Bärnighausen T. Innovations in health and demographic surveillance systems to establish the causal impacts of HIV policies. Curr Opin HIV AIDS 2015 Nov;10(6):483-494. [doi: 10.1097/COH.0000000000000203] [Medline: 26371462]

XSL•FO
RenderX

27. Chimbindi N, Mthiyane N, Birdthistle I, et al. Persistently high incidence of HIV and poor service uptake in adolescent girls and young women in rural KwaZulu-Natal, South Africa prior to DREAMS. PLoS One 2018;13(10):e0203193. [doi: 10.1371/journal.pone.0203193] [Medline: 30325932]
28. Akullian A, Vandormael A, Miller JC, et al. Large age shifts in HIV-1 incidence patterns in KwaZulu-Natal, South Africa. Proc Natl Acad Sci USA 2021 Jul 13;118(28). [doi: 10.1073/pnas.2013164118]
29. Avicenna. 2024. URL: https://avicennaresearch.com/ [accessed 2025-08-13]
30. : Datafree Technologies; 2024. URL: https://datafree.tech/ [accessed 2025-08-13]
31. Goedel WC, Reisner SL, Janssen AC, et al. Acceptability and feasibility of using a novel geospatial method to measure neighborhood contexts and mobility among transgender women in New York City. Transgend Health 2017;2(1):96-106. [doi: 10.1089/trgh.2017.0003] [Medline: 29082330]
32. Duncan DT, Park SH, Goedel WC, Sheehan DM, Regan SD, Chaix B. Acceptability of smartphone applications for global positioning system (GPS) and ecological momentary assessment (EMA) research among sexual minority men. PLoS One 2019;14(1):e0210240. [doi: 10.1371/journal.pone.0210240] [Medline: 30689651]
33. Nanyonjo G, Kwena Z, Nakamanya S, et al. Finding women in fishing communities around Lake Victoria: “Feasibility and acceptability of using phones and tracking devices”. PLoS One 2024;19(1):e0290634. [doi: 10.1371/journal.pone.0290634] [Medline: 38206982]
34. Theall KP, Felker-Kantor E, Wallace M, Zhang X, Morrison CN, Wiebe DJ. Considering high alcohol and violence neighborhood context using daily diaries and GPS: A pilot study among people living with HIV. Drug Alcohol Depend 2018 Jun 1;187:236-241. [doi: 10.1016/j.drugalcdep.2018.03.005] [Medline: 29684891]
35. Venter W, Coleman J, Chan VL, et al. Improving linkage to HIV care through mobile phone apps: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 17;6(7):e155. [doi: 10.2196/mhealth.8376] [Medline: 30021706]
36. Clouse K, Noholoza S, Madwayi S, et al. The implementation of a GPS-based location-tracking smartphone app in South Africa to improve engagement in HIV care: randomized controlled trial. JMIR Mhealth Uhealth 2023 May 19;11(1):e44945. [doi: 10.2196/44945] [Medline: 37204838]
37. Darejeh A, Salim SS. Gamification Solutions to Enhance Software User Engagement—A Systematic Review. Int J Hum Comput Interact 2016 Aug 2;32(8):613-642. [doi: 10.1080/10447318.2016.1183330]

Abbreviations

AHRI: Africa Health Research Institute
ART: antiretroviral therapy
RAM: random access memory

Edited by L Buis; submitted 01.11.24; peer-reviewed by HJ Gilmore, K Clouse; revised version received 06.05.25; accepted 06.06.25; published 25.08.25.

Please cite as:
Mathenjwa T, Okango EL, Tram KH, Inghels M, Cuadros D, Kim HY, Walsh F, Barnighausen T, Dobra A, Tanser F
Leveraging Smartphone Mobility Data to Understand HIV Risk Among Rural South African Young Adults: Feasibility Study
JMIR Mhealth Uhealth 2025;13:e67519
URL: https://mhealth.jmir.org/2025/1/e67519
doi:10.2196/67519

© Thulile Mathenjwa, Elphas Luchemo Okango, Khai Hoan Tram, Maxime Inghels, Diego Cuadros, Hae-Young Kim, Fiona Walsh, Till Barnighausen, Adrian Dobra, Frank Tanser. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adaptation of an mHealth Solution for the Nutritional Management of Diabetes in a Low- and Middle-Income Country: Pre-Post Mixed Methods Pilot Study

Cecilia Anza-Ramirez[1*], MSc; Karen Bonilla-Aguilar[1*], MSc; David Beran[2], PhD; Jean-Luc Mando[3], MSc; Lorena Saavedra-Garcia[1], MSc; Monica Julissa Angulo-Barranca[4], MD; Marco Alexi Taboada Garcia[5,6], MSc; Leonardo Albitres-Flores[1], MD; Alejandro Loayza[7], BSc; Jessica Hanae Zafra-Tanaka[1,2], PhD; Olivia Heller[8], MSc; María Lazo-Porras[1], PhD; Montserrat Castellsague Pique Perolini[8], DSN

[1]CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Av. Armendáriz 445, Miraflores, Lima, Peru

[2]Division of Tropical and Humanitarian Medicine, University of Geneva and Geneva University Hospitals, Geneva, Switzerland

[3]KmKoncept Sàrl, Veyrier, Switzerland

[4]Pediatric Diabetes Reference Center, Nuna Integral Pediatrics Clinic, Lima, Peru

[5]Nutrition, Faculty of Science and Engineering, Universidad Peruana Cayetano Heredia, Lima, Peru

[6]EsSalud, Lima, Peru

[7]Nutrition and Dietetics, Universidad Peruana de Ciencias Aplicadas, Lima, Peru

[8]Division of Tropical and Humanitarian Medicine, Geneva University Hospitals, Geneva, Switzerland

[*]these authors contributed equally

Corresponding Author:
María Lazo-Porras, PhD
CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Av. Armendáriz 445, Miraflores, Lima, Peru

Abstract

Background: Carbohydrate counting (CC) is vital for individuals living with type 1 diabetes mellitus (T1DM); yet, formal training is often lacking in many contexts. To bridge this gap, the parents of a person living with diabetes and a team at the Geneva University Hospital (HUG) developed WebDia, a free-access app that helps patients with T1DM assess meal carbohydrates and make informed decisions regarding insulin dosage. In the context of Peru, where dietary patterns and meal compositions may differ, customizing WebDia to suit the components of the Peruvian diet becomes particularly relevant.

Objective: This study aimed to customize WebDia according to the composition of the Peruvian diet to facilitate CC, and to provide training to health care workers (HCWs), children and adolescents living with T1DM, and their caregivers in the proficient use of WebDia-Mundi (new version of the Swiss app WebDia adapted to other geographic contexts).

Methods: A dietitian compiled a database of Peruvian foods and their carbohydrate content. This was reviewed by a Swiss nurse specialized in diabetes, a Peruvian pediatric endocrinologist, and 2 researchers. Validation was conducted with a small group of children and adolescents living with T1DM and their caregivers. Subsequently, a 3-day workshop was held in 3 Peruvian regions for HCW and children and adolescents living with T1DM. The first 2 days were a training course for HCW to gain knowledge in T1DM and learn CC skills. This was followed by a 1-day workshop involving HCW, children and adolescents living with T1DM, and their caregivers. At the end of the workshop and 3 months later, an evaluation was performed to assess the app's usability, glycated hemoglobin, quality of life, and knowledge perception for children and adolescents living with T1DM and their caregivers. Furthermore, changes in knowledge among HCWs and overall workshop satisfaction were measured.

Results: WebDia-Mundi was customized for the Peruvian context in 2022-2023. The training was attended by 25 HCWs, 25 children and adolescents living with T1DM, and 31 caregivers. Following the training, HCWs exhibited a significant 3.5-point increase in their knowledge of T1DM, while achieving positive results regarding the usability of WebDia-Mundi. Children and adolescents living with T1DM and their caregivers also reported a favorable perception of the ease of use and functionality of WebDia-Mundi, which enhanced their CC skills.

Conclusions: This study underscores the importance of collaboration among multidisciplinary teams and the involvement of individuals with T1DM. Adapting mobile health solutions to new contexts and sharing experiences can help standardize this process.

(JMIR Mhealth Uhealth 2025;13:e58029) doi:10.2196/58029

XSL•FO
RenderX

KEYWORDS

diabetes mellitus type 1; user-centered design; counting carbohydrates; eHealth; mobile applications; low-income countries; middle-income countries; diabetes; insulin; quality of life

Introduction

Diabetes represents a global health challenge, directly contributing to 1.5 million deaths annually, with a significant proportion of these deaths occurring in low- and middle-income countries (LMICs) [1]. While type 2 diabetes mellitus constitutes most diabetes cases, there are still approximately 9 million individuals grappling with type 1 diabetes mellitus (T1DM) [2]. Nonetheless, the persistent barrier of the cost of insulin and the supplies needed to maintain treatment for T1DM has a direct impact on individuals' daily lives [3].

A notable barrier within the health care system revolves around the inadequate nutrition education, particularly the emphasis on carbohydrate counting (CC) and the management of T1DM through a therapeutic patient education approach [4]. CC, in conjunction with insulin dosage adjustments, plays a pivotal role in the effective management of T1DM, improving metabolic control and enhancing the quality of life [5]. CC involves a dietary management strategy that quantifies the carbohydrate content of food, enabling individuals to administer the appropriate insulin dose to maintain optimal blood glucose levels [6]. The use of technology has proven effective in enhancing CC education and the overall learning experience [7,8]. To respond to the urgent demand to reach a balance in the treatment of T1DM, WebDia is an app designed to speed up CC by providing a complete list of foods, accompanied by their corresponding carbohydrate content [9]. The implementation of WebDia in Switzerland has yielded promising outcomes among individuals with T1DM, resulting in a reduction of glycated hemoglobin levels by 0.33% without a concurrent increase in hypoglycemia prevalence [10]. Moreover, users of WebDia have reported increased independence and self-confidence in managing their condition, leading to enhanced self-efficacy and improved self-management of T1DM [9].

Deficiencies in the health care system, such as the exclusive availability of specialized care at the tertiary level and insufficient training of health care workers (HCWs) in managing and diagnosing T1DM, impact the quality of life for individuals with this condition in LMICs, including Peru [3]. Peru's health care system faces a number of challenges, including poor accessibility to information at different health care levels, insufficient financial resources, and a shortage of HCWs with adequate training in diabetes management [11]. Furthermore, the health care system's response to individuals living with T1DM is notably deficient in terms of understanding the disease, and the information provided does not align with the unique needs and characteristics of this particular population [12]. Additionally, dietary patterns vary widely across different cultural and geographical contexts, influencing how people with T1DM approach CC. In Peru, the complexity of gastronomy, characterized by diverse ingredient combinations and traditional dishes, adds an additional challenge.

In this context, Peru finds itself in a favorable position to benefit from a tailored adaptation of WebDia, accompanied by the essential training of HCWs. The study had a dual purpose: first, to adapt WebDia to the Peruvian context as previously detailed, and second, to empower HCWs to enhance the quality of care they provide to individuals living with T1DM. Additionally, the study assessed the impact of WebDia-Mundi (new version of the Swiss app WebDia adapted to other geographic contexts) on the quality of life and glycated hemoglobin levels of children and adolescents living with T1DM.

Methods

Study Design

The study design was pre-post and usability study using mixed methods. The study used both qualitative and quantitative approaches across 3 regions of Peru with distinct demographic characteristics that contribute to their cultural and socioeconomic diversity (Arequipa, Lima, and Trujillo).

This study was carried out in 3 phases. The first phase involved adapting the WebDia app to the Peruvian context, spanning from March to December 2022. The second phase centered on training HCWs, children and adolescents living with T1DM, and their caregivers. A series of workshops was conducted to enhance diabetes care management, with a particular emphasis on refining skills related to CC. The third phase consisted of the evaluation of usability, knowledge acquisition, glycated hemoglobin levels, quality of life, perception of knowledge, and workshop satisfaction. Both phases were developed in March 2023. Furthermore, follow-up sessions were conducted 3 months later, from June to July 2023, to assess the sustainability of the acquired knowledge and skills.

Participants

The study had different participants for each phase. The first phase (adaptation) involved a collaboration with a specialized dietitian and comprised individuals older than 18 years living with T1DM and caregivers of children with T1DM. In the second phase, HCWs responsible for managing individuals requiring insulin for diabetes were actively involved, alongside children and adolescents living with T1DM, 6 to 18 years old, and their corresponding caregiver. Subsequently, during the third phase (evaluation), all participants from the second phase were considered. Recruitment of participants across all 3 phases was conducted through a strategic social media outreach, using a snowballing technique to expand the participant pool.

Study Procedures

First Phase: Adaptation

WebDia, an innovative mobile health (mHealth) app developed by JLM in response to his daughter's diagnosis of T1DM, assists individuals with diabetes who use insulin to count carbohydrates and adjust insulin treatment [10]. Originally in French and tailored to the Swiss diet, the app required adaptation to suit the

XSL•FO
RenderX

Peruvian context. One of the primary modifications was the translation of WebDia into Spanish, ensuring accessibility for most of the Peruvian population. Additionally, user-friendly features were adjusted, including redesigning the logo. The most critical adaptation involved integrating a food database that reflects the diversity of Peruvian cuisine, incorporating locally available foods, traditional dishes, and commonly consumed ingredients with culturally appropriate portion sizes.

For this purpose, the most prevalent foods consumed in Peru were identified by using food data from a previous CRONICAS project [13] and consulting a recipe book authored by a dietitian specializing in macronutrient counting. The selection of typical dishes was based on popular local foods from Lima, Arequipa, and Trujillo. However, we did not have access to a systematically compiled dataset of the most commonly consumed dishes by region. Instead, our selection was based on available data sources and expert consultation. Information on carbohydrate composition was then incorporated using different sources. For natural foods and traditional dishes, we used the Peruvian table of food composition [14] and values extracted from WebDia. For processed and ultraprocessed food, carbohydrates information was taken from the database developed by CRONICAS Center of Excellence in Chronic Disease at Universidad Peruana Cayetano Heredia, in collaboration with the University of North Carolina at Chapel Hill [13]. Additionally, although accessibility was considered in the app's adaptation process, WebDia-Mundi is not currently designed to support multiple languages beyond Spanish.

This adaptation was then validated with a small group of participants, including people living with T1DM from different ages and their caregivers. The validation process consisted of collecting first impressions about the app, such as missing foods on the list, technical issues, and overall usability feedback. During this process, we also addressed specific needs to enhance accessibility, ensuring that the app could be used even by individuals who do not speak Spanish or who are illiterate. In the Peruvian version, we also incorporated multiple food photos to help users estimate food portions and make the weight of foods more visible. Based on this input, an improved version, now called WebDia-Mundi [15], was developed. Finally, a pilot usability study was conducted with the same group of participants, generating data that facilitated enhancements to various app functions, considering diverse perspectives (Figure S1 in Multimedia Appendix 1).

Second Phase: Training

Workshops were conducted across 3 regions of Peru (Arequipa, Lima, and Trujillo) targeting HCWs, children, and adolescents living with T1DM, and their caregivers (Table S1 in Multimedia Appendix 2). These workshops primarily focused on comprehensive training in CC, instructions on using the app to perform CC as part of self-monitoring, and how to adjust insulin doses in a personalized way.

For this purpose, the HCWs were involved in a 2-day face-to-face theoretical and practical training during which they lived like a person living with T1DM (they practiced doing CC and applying to themselves several insulin injections simulated with sodium chloride). HCWs were instructed to previously watch 3 short videos produced at Geneva University Hospitals (HUG) to lay the foundations for diabetes management using CC [16-18]. These instructional days were followed by a one-day practical workshop, bringing together HCWs, children and adolescents living with T1DM, and their caregivers. During this session, HCWs who were already acquainted with the WebDia-Mundi app collaborated with children and adolescents living with T1DM and their caregivers to provide guidance on CC and how to adapt insulin treatment. To ensure effective participation, participants were encouraged to download the app from either the Apple Store [19] or the Play Store [20] to ensure effective use during the session.

Third Phase: Evaluation

During the 2-day face-to-face session, a knowledge test was applied before and after the training. In addition, at the beginning of the third-day practical workshop, sociodemographic data, perceptions of knowledge, and quality of life (self-reported by the children and adolescents living with T1DM and the evaluation of them provided by their caregiver) were collected. If consent was provided by children and adolescents living with T1DM and their caregivers, a blood sample was collected for glycated hemoglobin analysis. At the end of the workshop, the usability of WebDia-Mundi and workshop satisfaction were assessed. Three months later, participants were recontacted for follow-up evaluations to assess the usability of WebDia-Mundi, perceptions of knowledge, quality of life, and glycated hemoglobin (specifically for children and adolescents living with T1DM).

Outcomes

The primary outcome was the usability of WebDia-Mundi to explore the main issues in the app from the end users' perspective. The usability assessment was carried out using the System Usability Scale (SUS) for children and adolescents living with T1DM and their caregivers [21]. This questionnaire is considered a reliable tool to measure the usability of products and services, including hardware, software, mobile devices, and apps. In addition, the questionnaires were supplemented with questions about perceived operability of the app, perceived ease of the app, satisfaction with the app, and open-ended questions about the participants' experience of use. All questions were adapted and translated into Spanish using simple terminology to enhance comprehension.

Another primary outcome was enhancing knowledge of the HCWs. For this purpose, a pre and posttest was conducted at the beginning and at the end of the 2-day workshop sessions, to assess their knowledge. The test was developed by a diabetes educator. The questions were related to the topics taught during the training, such as diabetes, diet, CC, insulin adjustment, etc.

The following were predefined as secondary outcomes: glycated hemoglobin assessed by blood sampling using a validated clinical laboratory (SynLab ISO 9001:2015) in the 3 regions. Quality of life was measured using the Spanish validated PedsQL Diabetes Module (version 3.0; Mapi Research Trust) [22] using child self-report and parent proxy-report scales tailored to the respective age groups of the children. Perceptions of knowledge through specific questions. Workshop satisfaction

XSL•FO
RenderX

was assessed by a questionnaire developed by the research team. To ensure accuracy and appropriateness, we carefully reviewed the Spanish version. Knowledge perception and workshop satisfaction were assessed using specific questions developed by the research team (Table S5 in Multimedia Appendix 2).

Data Evaluation

The SUS

The SUS was administered to all participants, using a Likert scale ranging from 1 to 5, where 1 indicates strongly disagree, and 5 indicates strongly agree. For the analysis, we used the standard approach to generate a final score ranging from 0 to 100. To do so, we converted the scores with a different setting for odd and even items (respectively, positive and negative tone items). In addition, if a respondent leaves an item blank, it is assigned a raw score of 3 (the center of the 5-point scale). The adjusted scores were then summed and multiplied by 2.5 to obtain the standard SUS score [21].

Knowledge Test (Pre-Post)

A pre-post knowledge test was administered to HCWs, and the final score, ranging from 0 to 20, was calculated by adding the individual scores for each question. For more details on the content and level of the test, please refer to the questions in Table S5 in Multimedia Appendix 2.

Glycated Hemoglobin

The assessment of the glycated hemoglobin test was conducted in children and adolescents living with T1DM. The results were obtained from blood samples analyzed in a laboratory that had been validated across all 3 regions.

Quality of Life

The PedsQL 3.0 Diabetes was used to assess the quality of life in children and adolescents living with T1DM, considering perspectives from both children and adolescents living with T1DM and their caregivers. Like the SUS, a Likert scale (0 - 4) was used across five domains: (1) diabetes symptoms, (2) treatment I, (3) treatment II, (4) worry, and (5) communication. Treatment I is used to indicate the presence or absence of a problem, while Treatment II is used to measure the degree of difficulty associated with the problems identified. The items were then reverse transformed to a scale of 0 - 100 (0=100, 1=75, 2=50, 3=25, 4=0), with higher scores indicating better quality of life. If a respondent left an item blank, it was assigned a score of 2 (the center of the 5-point scale) [22].

Perception of Knowledge

Knowledge perception questions were assessed for all participants, using a knowledge rating scale from 1 to 5, with 5 representing the highest score.

Workshop Satisfaction

Satisfaction with the workshops was evaluated for all participants by extracting the most recurrent phrases from the open-ended questions.

Statistical Methods

Data analysis was conducted in STATA for Windows (version 17; StataCorp). The description of the study population was carried out using absolute and relative frequencies for categorical variables, while numerical variables were summarized using the median and the 25th and 75th percentiles. To assess changes in quality of life, perceived knowledge, and usability as reported by children and adolescents living with T1DM and their caregivers, the Wilcoxon matched pairs test and McNemar's chi-square test were used. A P value <.05 was considered significant. In addition, we used line graphs to visualize the progression of acquired knowledge before and after the training workshop, as well as to illustrate changes in glycated hemoglobin levels at the 3-month mark.

Ethical Considerations

The study was approved by the Institutional Review Board of Universidad Peruana Cayetano Heredia (approval number 513-35-22, registration code 208867). All participants provided their consent, with the option to withdraw at any time. Given that the study population was from urban areas and no specific cultural adaptations were required, the main consideration was ensuring that the language used in the consent form was clear and easily understood by participants. For children younger than 18 years, assent was obtained after a detailed explanation of the study's objectives, benefits, and risks. During the workshop phase, consent from only one caregiver was required. In cases where the child was below 6 years old, both the child and the caregiver had the option to participate in the activities; however, the data collected from this group were excluded from the analysis. All collected data were deidentified, and only the first authors had access to safeguard participant information. While no direct financial compensation was provided, participants received support for travel, accommodation, and meals as needed, particularly for those coming from regions outside the study site.

Results

Demographic Characteristics

Participants for all study phases were selected using a convenience sampling approach, based on their availability and willingness to participate. A total of 25 HCWs and children and adolescents living with T1DM, along with their caregivers, were initially enrolled. Subsequently, 24 HCWs and 19 children and adolescents living with T1DM and their caregiver were included in the study. For the follow-up phase, 12 children and adolescents living with T1DM and their caregivers were retained in the study (Figures S2 to S4 in Multimedia Appendices 3-5). Of the total HCWs, 18 (75%) were female. They reported caring for a diverse range of diabetes patients, with a median of 50 (IQR 5-150) patients with T2DM and a median of 5 (IQR 2-12) patients with T1DM. The perceived knowledge of T1DM was 12 (50%) fair and 9 (38%) good, perceived knowledge of CC was rated as 6 (25%) poor and 11 (46%) fair, and perceived knowledge of insulin therapy was reported as 11 (46%) poor and 7 (29%) good (Table 1).

XSL•FO
RenderX

Regarding children and adolescents living with T1DM, the median age was 12 (IQR 10 - 16) years, and their caregiver had a median age of 47 (IQR 40 - 51) years, with 15 (79%) caregivers being female. The primary caregiver of children and adolescents living with T1DM included other family members in addition to parents, but most of the attendees were mothers. Around 13 (72%) caregivers had a university or technical education. In terms of perceived knowledge, 13 (68%) caregivers had a fair perception of T1DM, and 4 (21%) caregivers considered it good. For CC, 7 (37%) caregivers rated their knowledge as poor, and 6 (32%) caregivers rated it as fair, while for insulin therapy, 8 (42%) caregivers considered it fair, and 8 (42%) caregivers considered it good (Table 2).

XSL•FO
RenderX

Table . Sociodemographic characteristics of participants.

Health care workers (n=24)	Values
Age, median (P25-P75)	48 (41-54)
Sex[a], n (%)	
Female	18 (75)
Male	5 (21)
Profession[a], n (%)	
Pediatric endocrinologist	3 (13)
Adult endocrinologist	2 (8)
Medicine	6 (25)
Nursing	7 (29)
Nutrition	3 (13)
Pharmacist	1 (4)
Diabetes educator[a], n (%)	
Yes	10 (42)
No	13 (54)
T1DM knowledge[a], n (%)	
Very good	1 (4)
Good	14 (58)
Fair	8 (33)
Poor	0 (0)
Very poor	0 (0)
T1DM[b] knowledge[a], n (%)	
Very good	0 (0)
Good	9 (38)
Fair	12 (50)
Poor	2 (8)
Very poor	0 (0)
Carbohydrate counting knowledge[a], n (%)	
Very good	0 (0)
Good	5 (21)
Fair	11 (46)
Poor	6 (25)
Very poor	1 (4)
Insulin therapy knowledge[a], n (%)	
Very good	3 (13)
Good	7 (29)
Fair	11 (46)
Poor	1 (4)
Very poor	1 (4)

[a]Results may not add due to missing values.

[b]T1DM: type 1 diabetes mellitus.

XSL•FO
RenderX

Table . Sociodemographic characteristics of children and adolescents living with type 1 diabetes mellitus and their caregiver/parent.

Characteristics	Values
Children living with T1DM[a] (n=19)	
Age, median (P25-P75)	12 (10-16)
Sex, n (%)	
Female	10 (53)
Male	9 (47)
Time since T1DM[a] diagnosis (months), median (P25-P75)	36 (15-56)
Type of insurance[b], n (%)	
Integrated Health Insurance	5 (28)
Social Security (EsSalud)	9 (50)
Private	4 (22)
Chronic complications[b], n (%)	
Yes	2 (11)
No	16 (89)
Type of insulin treatment[b], n (%)	
Human insulin	2 (11)
Analog insulin	16 (89)
Glucometer use, n (%)	
>5 times per day	7 (37)
2 - 5 times per day	7 (37)
Once a day	3 (16)
<1time per day	1 (5)
Do not use	1 (5)
Use of continuous glucose monitor[b], n (%)	
Yes	8 (44)
No	10 (56)
Medical visits in the past 12 months, median (P25-P75)	4 (1-6)
Medical visits in the past 3 months, median (P25-P75)	1 (0 - 2)
Hospitalizations in the last 12 months, median (P25-P75)	0 (0 - 1)
Hospitalizations in the last 3 months, median (P25-P75)	0 (0 - 0)
Severe hypoglycemia in the last 12 months[b], median (P25-P75)	1 (0 - 4)
Severe hypoglycemia in the last 3 months, median (P25-P75)	0 (0 - 2)
Glycated hemoglobin (%), median (P25-P75)	7.1 (5.7 - 8.0)
Carbohydrate counting knowledge[b], n (%)	
Very good	0 (0)
Good	1 (6)
Fair	11 (61)
Poor	3 (17)
Very poor	3 (17)
Parents/caregivers (n=19)	
Age, median (P25-P75)	47 (40-51)

XSL•FO
RenderX

Characteristics	Values
Sex[b], n (%)	
Female	15 (79)
Male	3 (16)
Highest level of education[b], n (%)	
Primary incomplete	1 (5)
Secondary complete	4 (21)
Technical	5 (26)
University	8 (46)
Relationship with the person with T1DM[ab], n (%)	
Mother	16 (84)
Father	1 (5)
Aunt/Uncle	1 (5)
T1DM[a] knowledge[b], n (%)	
Very good	0 (0)
Good	4 (21)
Fair	13 (68)
Poor	1 (5)
Very poor	0 (0)
Carbohydrate counting knowledge[b]	
Very good	0 (0)
Good	2 (10)
Fair	6 (32)
Poor	7 (37)
Very poor	3 (16)
Insulin therapy knowledge[b]	
Very good	1 (5)
Good	8 (42)
Fair	8 (42)
Poor	1 (5)
Very poor	0 (0)

[a]T1DM: type 1 diabetes mellitus.

[b]Results may not add due to missing values.

WebDia-Mundi Usability

HCWs achieved a median overall usability score of 76%, reflecting positive results for perceived operability of the app and partially positive findings for perceived app, ease of the app, and satisfaction with the app (Table S2 in Multimedia Appendix 2).

HCWs also highlighted the practicality of CC in the diet as follows:

> *Very useful in our country, where carbohydrates are prevalent, and misconceptions about their restriction in diabetes cases are common.*
>
> *It's user-friendly, allows saving calculation data, and the dish photos are highly beneficial.*
>
> *It fosters creativity, letting me combine available ingredients to meet carbohydrate limits.*

On the other hand, the app recorded an overall usability score of approximately 81% for children and adolescents living with T1DM, which significantly dropped to a median of 57% at 3 months. Likewise, caregivers had an initial median score of 74% during the workshop, which decreased significantly to 54% at 3 months. Notably, positive results were observed for perceived operability of the app, perceived ease of the app, and

XSL•FO
RenderX

satisfaction with the app in both groups, with no significant changes noted at the 3-month mark (Table 3).

Children and adolescents living with T1DM and their caregivers consistently emphasized the app's advantages. Statements from children and adolescents living with T1DM included:

> *It helps to calculate carbohydrates more accurately.*
>
> *It improves control of quantities at mealtimes.*
>
> *It promotes independence, gradually building knowledge in carb counting.*

Moreover, they expressed that they found the app highly useful and user-friendly. In parallel, caregivers mentioned their perspectives with comments such as:

> *It makes it easier for me to understand carbohydrate amounts for my children.*
>
> *It has the potential to prevent hypoglycemia and hyperglycemia.*
>
> *I can refine carbohydrate counting for enhanced control.*
>
> *It increases my confidence in accurately calculating carbohydrates.*

XSL•FO
RenderX

Table . Usability of WebDia reported by the children and by their parent/caregiver at baseline and 3 months later.

Characteristics	Baseline	3 months follow-up	*P* value
Children living with T1DM[a]			
Overall score, median (P25-P75)[b]	80.8 (51.9 - 85.6)	56.7 (3.8 - 71.2)	.02
Perceived operability of the app			
I could use it by myself[c]			.10
Yes	11 (58)	6 (50)	
No	3 (16)	5 (42)	
I believe that offers benefits[c]			.32
Yes	14 (74)	10 (83)	
No	0 (0)	1 (8)	
Perceived ease of the app			
Difficulties using the app[c]			.32
Yes	0 (0)	5 (42)	
No	14 (74)	5 (42)	
Difficulties with carbohydrate counting[c]			.32
Yes	4 (21)	5 (42)	
No	11 (58)	6 (50)	
Satisfaction with the app			
Instructions were adequate[c]			.32
Yes	15 (79)	10 (83)	
No	0 (0)	1 (8)	
Recommended app[c]			.32
Yes	15 (79)	10 (83)	
I don’t know	0 (0)	1 (8)	
Parents/caregivers			
Overall score, median (P25-P75)[b]	73.6 (51.9 - 85.6)	54.3 (18.3 - 68.8)	.03
Perceived operability of the app			
I could use it by myself[c]			.32
Yes	14 (74)	7 (58)	
No	4 (21)	5 (42)	
I believe that offers benefits[c]			>.99
Yes	18 (95)	11 (92)	
No	0 (0)	1 (8)	
Perceived ease of the app			
Difficulties using the app[c]			.03
Yes	1 (5)	7 (58)	
No	17 (90)	5 (42)	
Difficulty of carbohydrate counting[c]			.71
Yes	8 (42)	5 (42)	
No	10 (53)	6 (50)	
Satisfaction with the app			

XSL•FO
RenderX

Characteristics	Baseline	3 months follow-up	*P* value
Instructions were adequate[c]			>.99
Yes	18 (95)	11 (92)	
No	0 (0)	1 (8)	
Recommended app[c]			>.99
Yes	18 (95)	11 (92)	
I don’t know	0 (0)	1 (8)	

[a]T1DM: type 1 diabetes mellitus.

[b]Wilcoxon matched pairs test was used.

[c]McNemar chi-square test was used. Results may not add due to missing values.

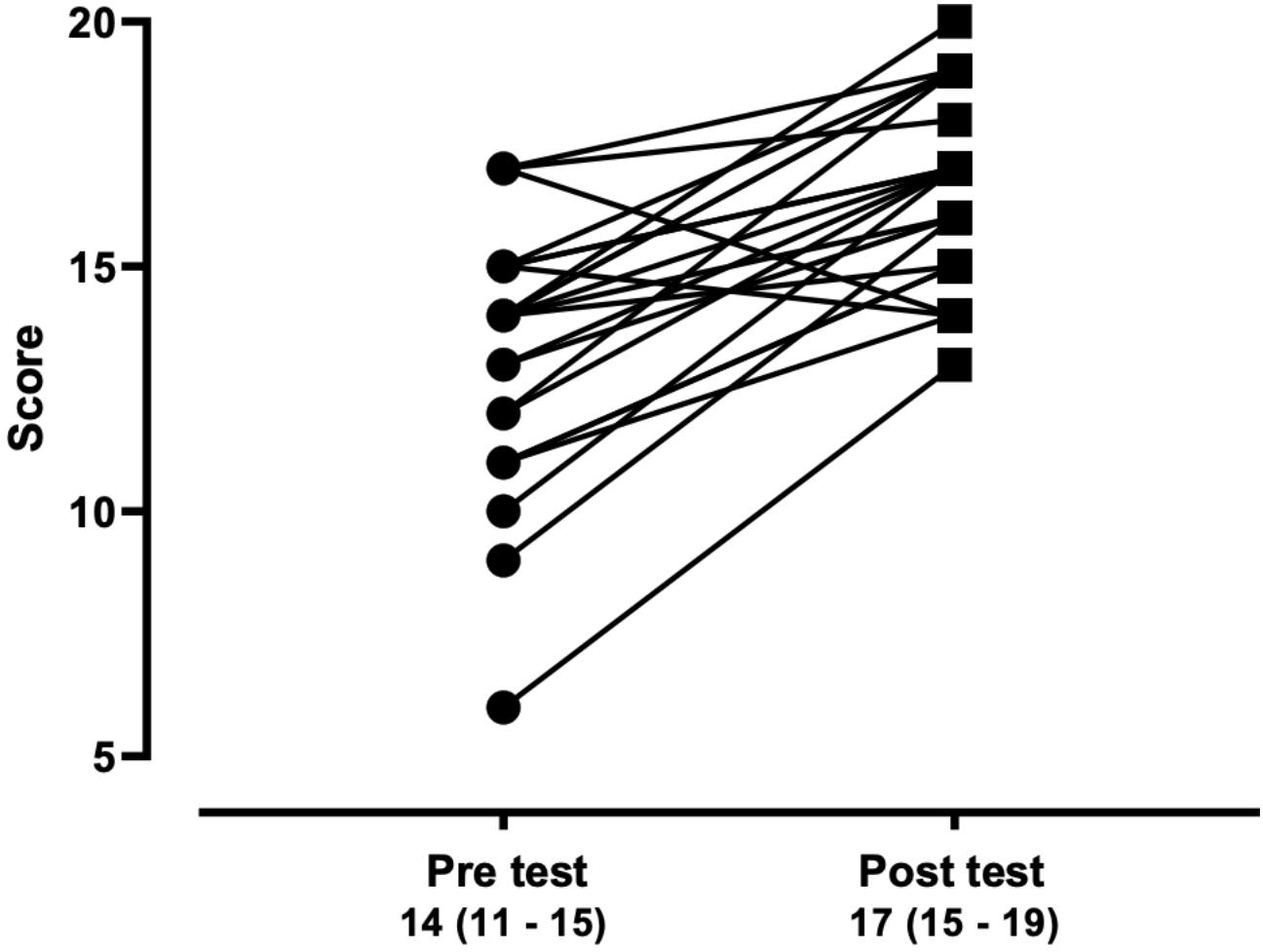
Common drawbacks identified during the workshop included the absence of specific dishes and foods, the necessity for adaptations tailored to different cities, the significance of differentiating between rapidly absorbed and complex carbohydrates, the demand for more detailed information regarding individual meal ingredients, and the desire to compare multiple meal options simultaneously based on their carbohydrate content. Furthermore, technology access was acknowledged as a potential limitation, particularly for individuals residing in areas with limited internet or smartphone connectivity. However, both children and adolescents living with T1DM and their caregivers expressed a strong motivation to frequently use WebDia-Mundi.

Knowledge Obtained in the Workshops

Overview

The HCW pretest scores ranged from 6 to 17, with a median of 14.0 (IQR 11.5 - 15.0), while the posttest scores ranged from 13 to 20, with a median of 17.0 (IQR 15.0 - 19.0). There was a significant increase between pre- and posttest scores (*P*=.001, Figure 1).

Figure 1. Pre- and posttest scores from health care workers (n=23) attending the workshops.

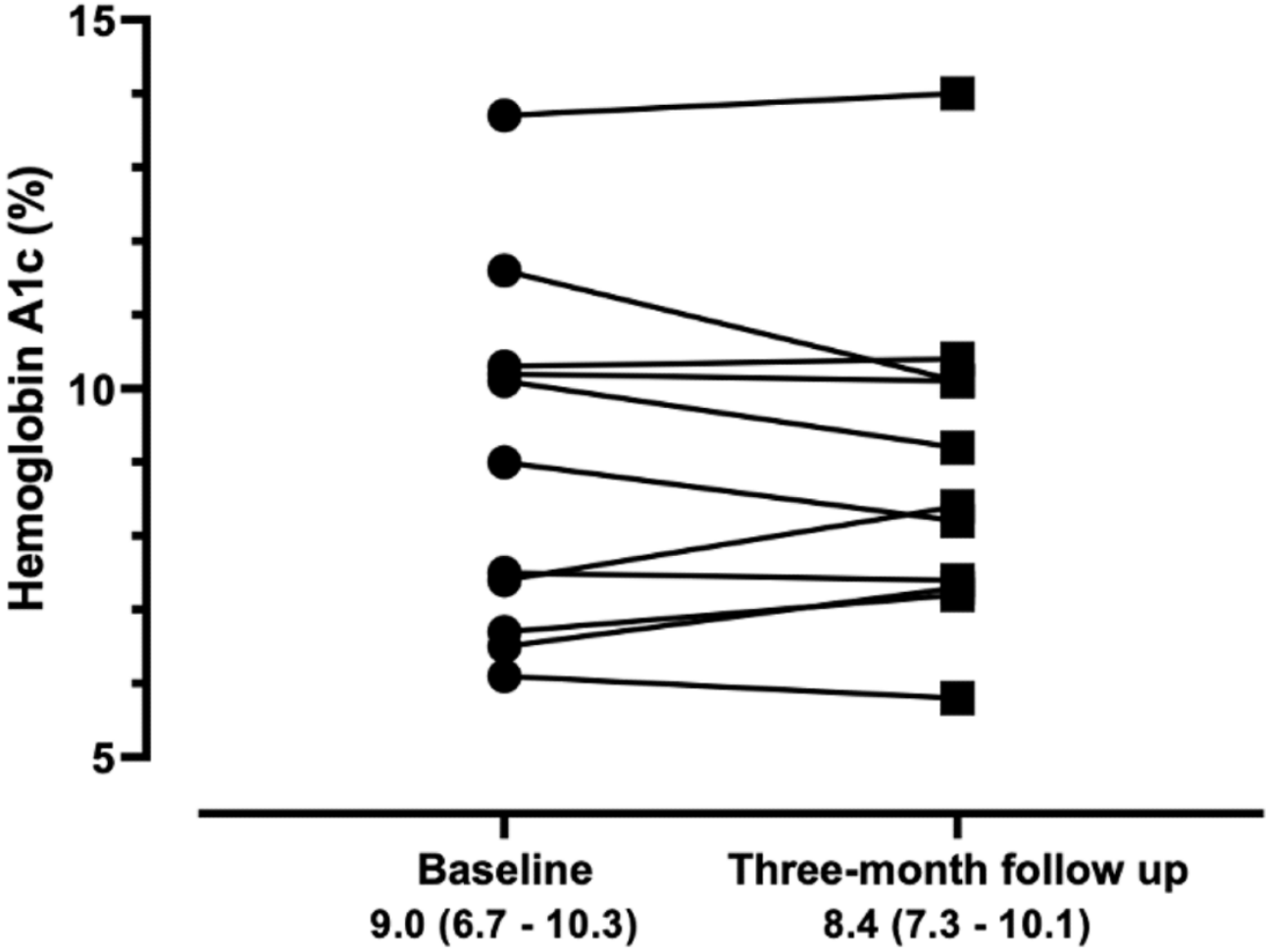


XSL•FO
RenderX

Glycated Hemoglobin

Glycated hemoglobin levels were measured in children and adolescents living with T1DM at the workshop and 3 months later. On the workshop day, the glycated hemoglobin level was 9.0% (IQR 6.7% - 10.3%), ranging from 5.4% to 13.7%. At the 3-month follow-up, the glycated hemoglobin level was 8.4% (IQR 7.3% - 10.1%), ranging from 5.8% to 14.0%, without changes in comparison to the baseline (*P*=.789, Figure 2).

Figure 2. Comparison of hemoglobin A_{1c} levels in participating children during the workshop and at the 3-month follow-up.



Quality of Life

Across all domains, scores were close to 60, with a slight increase in worry and a marginal decrease in diabetes symptoms, communication, and treatment-related aspects. The overall score showed a slight increase for both, but the changes observed from baseline to follow-up were not statistically significant (Table 4). On the other hand, no significant differences were found between the perception provided by children and adolescents living with T1DM and their caregivers (Table 3 in Multimedia Appendix 2).

XSL•FO
RenderX

Table . Quality of life reported by the children and by their parent/caregiver at baseline and 3 months later. Wilcoxon matched pairs test was used.

Characteristics	Baseline	3 months follow-up	*P* value
Children living with T1DM[a] (P25-P75)			
Diabetes symptoms	59.1 (56.8 - 77.3)	54.6 (50.0 - 72.7)	.35
Treatment I	56.3 (50.0 - 81.3)	50.0 (50.0 - 75.0)	.44
Treatment II	67.9 (50.0 - 82.1)	53.6 (50.0 - 78.6)	.27
Worry	41.7 (25.0 - 50.0)	50.0 (50.0 - 50.0)	.20
Communication	66.7 (33.3 - 91.7)	50.0 (50.0 - 75.0)	.95
Overall	60.5 (43.9 - 69.7)	61.0 (52.3 - 70.9)	.16
Parents/caregivers, median (P25-P75)			
Diabetes symptoms	65.9 (50.0 - 75.0)	50.0 (50.0 - 59.1)	.08
Treatment I	50.0 (37.5 - 68.8)	50.0 (50.0 - 75.0)	.56
Treatment II	60.7 (46.4 - 78.6)	50.0 (50.0 - 75.0)	.52
Worry	41.7 (33.3 - 50.0)	50.0 (50.0 - 58.3)	.11
Communication	54.2 (16.7 - 83.3)	50.0 (50.0 - 75.0)	.40
Overall	56.9 (42.6 - 68.9)	62.4 (56.4 - 75.8)	.11

[a]T1DM: type 1 diabetes mellitus.

Perception of knowledge

The perception of CC and T1DM knowledge, when compared with 3 months after the workshop, showed a significant decrease in median reported knowledge among both children and adolescents living with T1DM and their caregivers (Table 5).

Table . Difference in knowledge at 3-month follow-up in children living with type 1 diabetes mellitus and their caregivers. Wilcoxon matched pairs test was used.

Characteristics	Baseline	3 months follow-up	*P* value
Children living with T1DM[a], median (P25-P75)			
Carbohydrate counting knowledge	3 (3-4)	3 (2-3)	.03
Parents/caregivers, median (P25-P75)			
T1DM[a] knowledge	3 (3-3)	2 (2-3)	.02
Carbohydrate counting knowledge	3 (3-4)	3 (2-4)	.04
Insulin therapy knowledge	2 (2-3)	2 (2-2)	.75

[a]T1DM: type 1 diabetes mellitus.

Workshop Satisfaction

HCWs found that the workshops not only taught them about CC and the use of WebDia-Mundi but also allowed them to see T1DM from the perspective of children and adolescents living with T1DM and their caregivers. They expressed an understanding that there is no need to impose food restrictions when managing T1DM, and those with T1DM can lead normal lives if they practice CC. Some HCWs mentioned the following:

> *No need for prohibitions or excessive restrictions.*
>
> *Type 1 diabetics can lead a normal life.*

Regarding changes they plan to make in their practice after the workshop, HCWs expressed a commitment to active listening, considering patients' contexts, and practicing empathy. They mentioned intentions such as:

> *Listen and provide more support.*
>
> *Empathize and put myself in their shoes.*
>
> *Help caregivers understand it's not their fault, is a manageable condition.*
>
> *Show greater interest in their well-being and challenges.*
>
> *Empower them to manage the condition and live a normal life.*
>
> *Always prioritize the patient's feelings.*

HCWs found the most positive aspect of the workshop to be the valuable interactions with children and adolescents living with T1DM and their caregivers. On the other hand, the most negative aspect, as reported by HCWs, was the discomfort of self-administering glucose measurements and injections during the practical sessions. Their feedback included:

XSL•FO
RenderX

> *The hands-on experience was enriching, allowing us to experience it ourselves, administering NaCl and measuring glucose. It was a unique and fascinating practical course in my professional life.*
>
> *Living as a diabetic [person living with diabetes], understanding the effort required to measure glucose, calculate meals, and make injections, highlighted the time and dedication patients invest in normalising their glucose.*
>
> *Practising self-monitoring of glucose and experiencing it first-hand truly deepened our understanding of the patient's perspective.*

Finally, the comments and reflections of the HCWs underscored the workshop's significant influence on their perspectives and professional approaches. They emphasized the need to expand participation, sharing this knowledge with professionals from various health care organizations, expressing interest in specialized workshops, and proposing the creation of camps and activities for children and adolescents living with T1DM.

The caregiver also showed significant interest in the workshop. Most of them found the workshop duration suitable, and all reported valuable knowledge gained, resulting in a positive learning outcome. Their feedback emphasized the importance of CC and the usefulness of CC apps. Some comments mentioned were:

> *I gained a better understanding of carbohydrates.*
>
> *The carbohydrate counting app proved to be incredibly useful.*
>
> *Maintaining a positive outlook is crucial; it helps build confidence that things happen for a reason.*

In summary, the lifestyle changes considered by the children and adolescents living with T1DM and their caregivers were predominantly centered around dietary choices and the use of WebDia-Mundi (Table S4 in Multimedia Appendix 2).

Discussion

Principal Results

Our findings indicate a favorable response to the WebDia-Mundi app (an adaptation of the WebDia app tailored to the Peruvian context) among HCWs, children and adolescents living with T1DM, and their caregivers. The incorporation of popular local foods and an easy-to-use interface were key factors in ensuring the app's effective adaptation in Peru. Additionally, the app played an important role in enhancing users' understanding of CC, which is essential for better nutritional management and quality of life. The training provided to HCWs improved their knowledge regarding the care they deliver to individuals with T1DM. It is important to note that most HCWs are more familiar and comfortable providing dietary advice for individuals with type 2 diabetes and often feel less confident when counseling patients with T1DM. Moreover, children and adolescents living with T1DM, and their caregivers, exhibited a generally positive perception of their quality of life. While the use of WebDia-Mundi resulted in slight improvements in specific domains, these enhancements did not reach statistical significance. Furthermore, there were no notable changes observed in the levels of glycosylated hemoglobin among children and adolescents living with T1DM. In the management of T1DM in LMICs, having tools that alleviate the burden of self-managing glucose levels is crucial. One of the key challenges faced by people living with T1DM is understanding how to manage their meals effectively. Traditionally, many individuals with T1DM followed highly restrictive dietary patterns due to limited knowledge about CC. Providing an accessible and practical app empowers both patients and caregivers, while improving HCWs' understanding of CC contributes to better overall care.

Comparison With Prior Work

The general knowledge was low among the participants of the workshop. Other studies with physicians and nurses show similar results on the inadequate training in the treatment of T1DM [23-25]. A parallel scenario emerged among children and adolescents living with T1DM and their caregivers, highlighting a perceived inadequate knowledge concerning self-management and dietary practices [26]. It is noteworthy that enhancing the knowledge of CC has been proven to positively impact the self-management of the disease [27]. Nevertheless, 3 months into the training, there was a notable decline in the perceived knowledge regarding CC and T1DM. This decline could be attributed to diverse postintervention factors that might have adversely impacted self-perceived knowledge when attempting to implement the acquired skills from the workshops, especially in the practical application of CC in daily life, or they discovered there was much more to learn than they initially thought as they underwent the training [28,29].

Regarding the usability of WebDia-Mundi, HCWs expressed a positive attitude towards operability, ease, and satisfaction, emphasizing the practicality of the CC in the diet and the promotion of an unrestricted diet. Similarly, both children and adolescents living with T1DM and their caregivers initially shared a similar attitude. However, over the course of 3 months, this positive attitude decreased significantly by 24 and 20 percentage points for children and adolescents living with T1DM and their caregivers, respectively, resulting in an attitude close to neutral. Despite this change in attitude, there were no statistically significant differences observed in perceived operability, satisfaction, and ease. The lack of significance in these measures may be attributed to the small sample size, as numerical values are more vulnerable to minor fluctuations compared with relative frequency. Additionally, the observed lack of significance may be linked to a disinterest in using the app, given that a low proportion of adolescents have been reported to adhere to the use of such apps [30]. The mHealth adherence rate averages 56%, highlighting the necessity for enhancement through customization options and reminder notifications for improved usability [31].

Furthermore, our results show no significant changes in the mean glycosylated hemoglobin level of children and adolescents living with T1DM after 3 months. These findings are consistent with previous reports suggesting that while changes have been observed with CC support apps at 3 months in adults, no similar changes were found in the pediatric age group, possibly due to

XSL•FO
RenderX

the inherent difficulty in controlling diet in this population [32,33]. It is essential to note that the patient management approach in Peru exhibited clear restrictions, suggesting that introducing an alternative management style might necessitate an adaptation period for patients to assimilate the concept of CC, potentially extending beyond 3 months. While certain apps, such as the Swiss version (ie, WebDia), have explored the option of displaying insulin amounts to enhance user-friendliness [10,34], this was not implemented in Peru due to regulatory requirements and the need for supervised implementation by trained HCWs to mitigate potential clinical consequences [35]. Furthermore, there was limited awareness among patients' treating physicians about WebDia-Mundi, and the research team did not provide close follow-up.

Most existing mHealth apps for CC are free; however, their food databases are often generic, lacking cultural specificity [36,37]. The complexity of CC varies significantly depending on the dietary context of each country. One widely recommended feature in previous studies is the use of visual food representations, which was also preferred by participants in our validation process [38]. As a result, we incorporated more visual photos to enhance usability. While the original Swiss version included only 20 food images, the Peruvian version was enriched with 108 photos to better support users in estimating portion sizes and improve the app's practical utility. Furthermore, previous studies have not been conducted in LMICs, where challenges such as limited access to health care and medications can significantly impact outcomes related to quality-of-life improvements for people living with T1DM. However, despite the lack of changes in glycosylated hemoglobin, participants reported improvements in the management and control of their disease, citing increased self-confidence [39]. Some participants mentioned feeling more freedom in their dietary choices and greater confidence when deciding what to eat. Additionally, children were able to assist their caregivers and became more aware of CC, making the process less stressful for both. Consequently, our results on the perception of the quality of life of both children and adolescents living with T1DM and their caregivers, although not statistically significant, revealed a slight increase in the overall score.

A frequently suggested enhancement for WebDia-Mundi involves expanding the information provided for each preparation or food, such as displaying ingredients considered, carbohydrate typology, and comparisons with similar preparations. While this could enhance user decision-making, there's a potential trade-off, as it might increase the complexity of the app [34]. Access to technology also surfaced as a potential limitation, particularly for individuals residing in areas with restricted internet or smartphone connectivity. This is one of the most common challenges encountered in the implementation of digital health services or mHealth solutions [40].

Regarding the workshop, the results indicate a noteworthy increase in knowledge acquisition among HCWs. Beyond demonstrating improved understanding of CC and the use of WebDia-Mundi, they also expressed gaining insights into the lived experience of an individual living with T1DM. These workshops adhere to the basic education outlined in the principles of the National Assessment Framework, which is derived from the National Diabetes Services Plan in Australia [41]. This educational approach aims to advance knowledge and understanding, but to achieve significant behavioral changes in HCWs, it is essential to implement a comprehensive training program [41]. Similarly, children and adolescents living with T1DM and their caregivers provided positive feedback about the workshop, citing the acquisition of valuable knowledge, particularly in the realm of making informed food choices. The involvement of children and adolescents living with T1DM and their caregivers in these workshops facilitates a thorough assessment of the needs and expectations for enhancing educational programs tailored for individuals living with T1DM [42].

Public Health and Clinical Relevance

The adapted app, WebDia-Mundi, holds the potential to enhance access to diabetes management tools, especially in LMICs. By offering an adaptable tool for diverse cultural settings, it could contribute to reducing disparities in health care and improving equity in access to quality resources for T1DM control [40]. Additionally, the app can serve as a valuable tool for health promotion and education, as the integration of mHealth and health coaching for individuals with chronic diseases has demonstrated a positive impact on both the quality of life and self-management of the disease [43]. Facilitating CC and insulin management can empower individuals living with T1DM and their caregivers to make informed decisions about their treatment and enhance self-care, regardless of their location. However, the adaptation process posed significant challenges, highlighting the need for more agile and scalable approaches to effectively adapt mHealth tools to different countries. Cultural and linguistic differences, as well as variations in dietary habits and health systems, require a structured but flexible adaptation framework. Multistakeholder involvement was essential to ensure that the tool meets the needs of the target population. In addition, e-learning strategies could facilitate the implementation of WebDia-Mundi, improving its accessibility and ease of use for both HCWs and end users in a variety of settings. On the clinical side, WebDia-Mundi could be an effective resource for HCWs in clinical T1DM management by simplifying carbohydrate intake tracking and insulin dosing, ensuring that CC education remains accessible to a wide range of users, including those who are illiterate or speak a native language other than Spanish [44]. Furthermore, WebDia Mundi might be beneficial for caregivers by enabling them to better understand diabetes management and providing practical tools to support patients more effectively [10].

Limitations

This study has several strengths that significantly contribute to its impact and reliability. First, the successful adaptation of the WebDia app to the Peruvian context highlights the adaptability of WebDia-Mundi. Its potential to address and integrate diverse dietary and cultural needs highlights the possibility of scalability to other countries with no costs for the users. Second, the comprehensive approach, incorporating both quantitative and qualitative methodologies, offers a more complete assessment of the app's effectiveness. Additionally, the active involvement and collaboration among HCWs, children and adolescents living

XSL•FO
RenderX

with T1DM, and their caregivers throughout the study phases ensure validity and reflect a strong commitment to enhancing T1DM care within the Peruvian context. It is important to highlight that the collaboration among dietitians, endocrinologists, app developers, and the diabetes community throughout the study phases ensured that the adaptation process was user-centered and aligned with the real needs of the target population. The participatory approach ensured that WebDia-Mundi met the dietary and cultural needs of the target population. Furthermore, the inclusion of 3 distinct cities in Peru, each with its unique characteristics, although not exhaustive of Peru's geographical diversity, enhances the heterogeneity of our findings. However, it is important to acknowledge certain limitations. The absence of a control group in specific aspects of the study restricts direct comparisons with individuals who did not participate in the research. Additionally, the short follow-up period may have limited our ability to observe significant changes in HbA_{1c} and long-term behavioral adaptations. Another key limitation is the potential bias inherent in using a convenience and snowball sampling method and recruiting participants through social media. This approach may have inadvertently led to the inclusion of individuals already inclined towards undertaking the course due to their existing interest, potentially skewing our sample towards those with lower levels of knowledge. The lack of heterogeneity in education levels among caregivers may have influenced the feedback on the app's usability and accessibility, leading to different perspectives on its ease of use and practical applicability. Furthermore, this recruitment strategy could impact the external validity of our findings. However, it is important to highlight that recruiting people with T1DM presents significant challenges, even when using traditional recruitment methods. Based on our experience in patient recruitment, social media and community networks have proven to be among the most effective ways to engage participants in this context. Nevertheless, it is crucial to emphasize that this study serves as a pilot project, not intended to yield universally applicable results. Moreover, a technical barrier in LMICs is the potential existence of areas with lower mobile literacy levels or limited internet access. This might prevent widespread and equitable use of app-based tools, possibly excluding certain populations from fully benefiting from technological health solutions [45]. Finally, the initial sample size was not reached, and therefore, the power is lower, or the ability to detect small changes is lower.

Conclusion

WebDia-Mundi is a tool for enhancing the self-management of people living with T1DM, covering both children and adolescents. Although clinically significant improvements in glycosylated hemoglobin have not been achieved, the application has received favorable acceptance and demonstrated efficacy in increasing dietary education, enabling a life free from unnecessary restrictions. It is crucial to acknowledge that additional efforts are necessary to enhance implementation, which requires further pretesting. Nonetheless, this study serves as an example of how tailoring mHealth solutions to the specific context of LMICs can contribute significantly to enhancing the quality of life for individuals with T1DM. Future research is needed to assess the long-term impact of the app and explore its adaptation to other cultural settings and evaluate strategies to reach a more diverse population, including people with different languages, literacy levels, and educational backgrounds, to ensure equitable access and ease of use.

Acknowledgments

We would like to acknowledge the many individuals who were involved in this study. A special thanks to Maria Dextre Jauregui and Ursula Bonilla for their collaboration in this study. We would like to acknowledge the "Fond de Péréquation" from the Geneva University Hospitals and The Leona M and Harry B Helmsley Charitable Trust for granting the funds for this project. The database used for carbohydrate counting content in processed and ultraprocessed foods originated from a prior study conducted within the Food Policy Program, funded by Bloomberg Philanthropies (grants 46129 and 2019 - 71181). Additionally, support was provided by the SALURBAL project, funded by Wellcome Trust (grant 205177/Z/16/Z).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request. Deidentified data supporting the findings of this study can be provided for research purposes in line with JMIR's data sharing policy.

Authors' Contributions

CAR collected the study data and contributed to manuscript preparation. KBA drafted and wrote the manuscript. DB provided overall guidance on the study design and manuscript development and contributed to study conceptualization and manuscript writing. JLM developed the WebDia application. LSG contributed to the carbohydrate estimation of processed foods. MJAB and MATG contributed to training activities. LAF and AL participated in the validation process during the first phase of adaptation. JZT and OH contributed to study conceptualization and manuscript writing. MLP coordinated the project, provided overall supervision, and contributed to study conceptualization and manuscript writing. MCP supervised the entire study process, provided critical guidance and training, and contributed to study conceptualization and manuscript writing. All authors reviewed and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Process flowchart for adapting WebDia to the Peruvian context and create WebDia - Mundi.
[PNG File, 270 KB - mhealth_v13i1e58029_app1.png]

Multimedia Appendix 2
Overview of session design, WebDia Mundi usability, quality of life outcomes, workshop satisfaction, and knowledge perception.
[DOCX File, 37 KB - mhealth_v13i1e58029_app2.docx]

Multimedia Appendix 3
Flowchart of health care workers attending the workshops by study site.
[PNG File, 33 KB - mhealth_v13i1e58029_app3.png]

Multimedia Appendix 4
Flowchart of children living with type 1 diabetes mellitus attending the workshops by study site.
[PNG File, 69 KB - mhealth_v13i1e58029_app4.png]

Multimedia Appendix 5
Flowchart of caregivers/parents of individuals living with type 1 diabetes mellitus attending the workshops by study site.
[PNG File, 60 KB - mhealth_v13i1e58029_app5.png]

References

1. Global report on diabetes. World Health Organization. 2016. URL: https://www.who.int/publications-detail-redirect/9789241565257 [accessed 2023-11-04]
2. Green A, Hede SM, Patterson CC, et al. Type 1 diabetes in 2017: global estimates of incident and prevalent cases in children and adults. Diabetologia 2021 Dec;64(12):2741-2750. [doi: 10.1007/s00125-021-05571-8] [Medline: 34599655]
3. Abdraimova A, Besançon S, Portocarrero J, et al. Management of type 1 diabetes in low- and middle-income countries: comparative health system assessments in Kyrgyzstan, Mali, Peru and Tanzania. Diabet Med 2022 Aug;39(8):e14891. [doi: 10.1111/dme.14891] [Medline: 35621029]
4. Clerc A. Nutrition education to type 1 diabetes patients: few changes over the time. Front Clin Diabetes Healthc 2023;4:1243237. [doi: 10.3389/fcdhc.2023.1243237] [Medline: 37705724]
5. Tascini G, Berioli MG, Cerquiglini L, et al. Carbohydrate counting in children and adolescents with type 1 diabetes. Nutrients 2018 Jan 22;10(1):109. [doi: 10.3390/nu10010109] [Medline: 29361766]
6. Dimitriades ME, Pillay K. Carbohydrate counting in type 1 diabetes mellitus: dietitians' perceptions, training and barriers to use. S Afr J Clin Nutr 2022 Sep 30;35(3):94-99. [doi: 10.1080/16070658.2021.1979764]
7. Amorim D, Miranda F, Ferreira L, Abreu C. Data-driven carbohydrate counting accuracy monitoring: a personalized approach. Procedia Comput Sci 2022;204:900-906. [doi: 10.1016/j.procs.2022.08.109]
8. Chotwanvirat P, Hnoohom N, Rojroongwasinkul N, Kriengsinyos W. Feasibility study of an automated carbohydrate estimation system using thai food images in comparison with estimation by dietitians. Front Nutr 2021;8:732449. [doi: 10.3389/fnut.2021.732449] [Medline: 34733876]
9. WebDia. Geneva University Hospitals. URL: https://www.hug.ch/application-mobile/webdia [accessed 2023-04-21]
10. Klee P, Bussien C, Castellsague M, et al. An intervention by a patient-designed do-it-yourself mobile device app reduces HbA1c in children and adolescents with type 1 diabetes: a randomized double-crossover study. Diabetes Technol Ther 2018 Dec;20(12):797-805. [doi: 10.1089/dia.2018.0255] [Medline: 30403495]
11. Cardenas MK, Miranda JJ, Beran D. Delivery of type 2 diabetes care in low- and middle-income countries: lessons from Lima, Peru. Diabet Med 2016 Jun;33(6):752-760. [doi: 10.1111/dme.13099] [Medline: 27194174]
12. Insulin access and affordability. Health Action International. URL: https://haiweb.org/projects/acciss-study [accessed 2023-11-04]
13. Saavedra-Garcia L, Meza-Hernández M, Diez-Canseco F, Taillie LS. Reformulation of top-selling processed and ultra-processed foods and beverages in the Peruvian food supply after front-of-package warning label policy. IJERPH 2023;20(1):424. [doi: 10.3390/ijerph20010424]
14. Instituto Nacional de Salud (INS). Tablas Peruanas de Composición de Alimentos: Lima: INS; 2017. URL: https://repositorio.ins.gob.pe/bitstream/handle/20.500.14196/1034/tablas-peruanas-QR.pdf [accessed 2025-09-16]
15. WebDia-Mundi. URL: https://webdia-mundi.com/index.html [accessed 2023-11-04]

XSL•FO
RenderX

16. WebDia | la hipoglucemia qué hacer. YouTube. 2023. URL: https://www.youtube.com/watch?v=Wmx-rQ9RaUw [accessed 2023-11-04]
17. WebDia | las funciones de la insulina en la diabetes tipo 1. YouTube. 2023. URL: https://www.youtube.com/watch?v=Jz7ivlFC9yI [accessed 2023-11-04]
18. WebDia | los grupos de alimentos de la pirámide alimentaria. YouTube. 2023. URL: https://www.youtube.com/watch?v=a35cT3X3V3Q [accessed 2023-11-04]
19. Mando JL. WebDia-mundi. App Store. 2023. URL: https://apps.apple.com/us/app/webdia-mundi/id6444607213 [accessed 2023-11-25]
20. Mando JL. WebDia-mundi. Google play. 2023. URL: https://play.google.com/store/apps/details?id=com.kmkoncept.webdiamundi&hl=es [accessed 2023-11-25]
21. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
22. Garcia L, Manna TD, Passone CDG, Oliveira LD. 3.0 diabetes module in Brazil-Portuguese language. J Pediatr (Rio J) 2018;94(6):680-688. [doi: 10.1016/j.jped.2017.09.009] [Medline: 29144963]
23. Sow A, Boiro D, Sow PS, et al. Insulin therapy in childhood type 1 diabetes: knowledge and practice in Senegal. Arch Pediatr 2021 May;28(4):307-310. [doi: 10.1016/j.arcped.2021.02.006] [Medline: 33715933]
24. Stefanowicz-Bielska A, Słomion M, Rąpała M. Knowledge of school nurses on the basic principles of type 1 diabetes mellitus self-control and treatment in children. Int J Environ Res Public Health 2022 Dec 9;19(24):16576. [doi: 10.3390/ijerph192416576] [Medline: 36554455]
25. Alassaf A, Gharaibeh L, Abuna'meh L, Odeh R. Adequacy of knowledge of new medical graduates about diagnosis and management of children and adolescents with type 1 diabetes in a developing country. BMC Med Educ 2023 Apr 12;23(1):234. [doi: 10.1186/s12909-023-04234-z] [Medline: 37046268]
26. Owusu BA, Ofori-Boateng P, Forbes A, Doku DT. Knowledge of young people living with type 1 diabetes and their caregivers about its management. Nurs Open 2023 Apr;10(4):2426-2438. [doi: 10.1002/nop2.1498] [Medline: 36448367]
27. Bawazeer NM, Alshehri LH, Alharbi NM, et al. Evaluation of carbohydrate counting knowledge among individuals with type 1 diabetes mellitus in Saudi Arabia: a cross-sectional study. BMJ Nutr Prev Health 2022 Dec;5(2):344-351. [doi: 10.1136/bmjnph-2022-000553] [Medline: 36619333]
28. Leslie RD, Evans-Molina C, Freund-Brown J, et al. Adult-onset type 1 diabetes: current understanding and challenges. Diabetes Care 2021 Nov;44(11):2449-2456. [doi: 10.2337/dc21-0770] [Medline: 34670785]
29. Emberson LL. Gaining knowledge mediates changes in perception (without differences in attention): a case for perceptual learning. Behav Brain Sci 2016 Jan;39:e240. [doi: 10.1017/S0140525X15002496] [Medline: 28355843]
30. Trawley S, Browne JL, Hagger VL, et al. The use of mobile applications among adolescents with type 1 diabetes: results from diabetes MILES youth-Australia. Diabetes Technol Ther 2016 Dec;18(12):813-819. [doi: 10.1089/dia.2016.0233] [Medline: 27788032]
31. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
32. AlBabtain SA, AlAfif NO, AlDisi D, AlZahrani SH. Manual and application-based carbohydrate counting and glycemic control in type 1 diabetes subjects: a narrative review. Healthcare (Basel) 2023 Mar 24;11(7):934. [doi: 10.3390/healthcare11070934] [Medline: 37046861]
33. Trawley S, Baptista S, Browne JL, Pouwer F, Speight J. The use of mobile applications among adults with type 1 and type 2 diabetes: results from the second MILES-Australia (MILES-2) study. Diabetes Technol Ther 2017 Dec;19(12):730-738. [doi: 10.1089/dia.2017.0235] [Medline: 29028442]
34. Sze WT, Kow SG. Perspectives and needs of Malaysian patients with diabetes for a mobile health app support on self-management of diabetes: qualitative study. JMIR Diabetes 2023 Oct 23;8:e40968. [doi: 10.2196/40968] [Medline: 37870903]
35. Rhyner D, Loher H, Dehais J, et al. Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: a comparative study. J Med Internet Res 2016 May 11;18(5):e101. [doi: 10.2196/jmir.5567] [Medline: 27170498]
36. Martinez-Millana A, Jarones E, Fernandez-Llatas C, Hartvigsen G, Traver V. App features for type 1 diabetes support and patient empowerment: systematic literature review and benchmark comparison. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e12237. [doi: 10.2196/12237] [Medline: 30463839]
37. Sun C, Malcolm JC, Wong B, Shorr R, Doyle MA. Improving glycemic control in adults and children with type 1 diabetes with the use of smartphone-based mobile applications: a systematic review. Can J Diabetes 2019 Feb;43(1):51-58. [doi: 10.1016/j.jcjd.2018.03.010] [Medline: 30026048]
38. Nguyen M, Hossain N, Tangri R, et al. Systematic evaluation of Canadian diabetes smartphone applications for people with type 1, type 2 and gestational diabetes. Can J Diabetes 2021 Mar;45(2):174-178. [doi: 10.1016/j.jcjd.2020.07.005] [Medline: 33127288]

XSL•FO
RenderX

39. Castensøe-Seidenfaden P, Husted GR, Jensen AK, et al. Testing a smartphone app (young with diabetes) to improve self-management of diabetes over 12 months: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jun 26;6(6):e141. [doi: 10.2196/mhealth.9487] [Medline: 29945861]
40. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 1;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
41. Olson JL, White B, Mitchell H, et al. The design of an evaluation framework for diabetes self-management education and support programs delivered nationally. BMC Health Serv Res 2022 Jan 9;22(1):46. [doi: 10.1186/s12913-021-07374-4] [Medline: 35000599]
42. Ridge K, Thomas S, Jackson P, et al. Diabetes-oriented learning family intervention (DOLFIN): a feasibility study evaluating an intervention for carers of young persons with type 1 diabetes. Diabet Med 2014 Jan;31(1):55-60. [doi: 10.1111/dme.12333] [Medline: 24117683]
43. Obro LF, Heiselberg K, Krogh PG, et al. Combining mHealth and health-coaching for improving self-management in chronic care. a scoping review. Patient Educ Couns 2021 Apr;104(4):680-688. [doi: 10.1016/j.pec.2020.10.026] [Medline: 33143907]
44. Centenaro A, do Nascimento C, Beretta MV, da Costa Rodrigues T. Carbohydrate counting as a strategy to optimize glycemic control in type 1 diabetes mellitus. Arch Endocrinol Metab 2023 Mar 30;67(3):385-394. [doi: 10.20945/2359-3997000000596] [Medline: 36748932]
45. van Olmen J, Erwin E, García-Ulloa AC, et al. Implementation barriers for mHealth for non-communicable diseases management in low and middle income countries: a scoping review and field-based views from implementers. Wellcome Open Res 2020;5:7. [doi: 10.12688/wellcomeopenres.15581.2] [Medline: 32399497]

Abbreviations

CC: carbohydrate counting
HCW: health care worker
HUG: Geneva University Hospital
LMIC: low- and middle-income country
mHealth: mobile health
SUS: System Usability Scale
T1DM: type 1 diabetes mellitus

Edited by L Buis; submitted 03.03.24; peer-reviewed by CL Sánchez Bocanegra, R Bipat; revised version received 09.04.25; accepted 04.06.25; published 25.09.25.

Please cite as:
Anza-Ramirez C, Bonilla-Aguilar K, Beran D, Mando JL, Saavedra-Garcia L, Angulo-Barranca MJ, Taboada Garcia MA, Albitres-Flores L, Loayza A, Zafra-Tanaka JH, Heller O, Lazo-Porras M, Castellsague Pique Perolini M
Adaptation of an mHealth Solution for the Nutritional Management of Diabetes in a Low- and Middle-Income Country: Pre-Post Mixed Methods Pilot Study
JMIR Mhealth Uhealth 2025;13:e58029
URL: https://mhealth.jmir.org/2025/1/e58029
doi:10.2196/58029

© Cecilia Anza-Ramirez, Karen Bonilla-Aguilar, David Beran, Jean-Luc Mando, Lorena Saavedra-Garcia, Monica Julissa Angulo-Barranca, Marco Alexi Taboada Garcia, Leonardo Albitres-Flores, Alejandro Loayza, Jessica Hanae Zafra-Tanaka, Olivia Heller, María Lazo-Porras, Montserrat Castellsague Pique Perolini. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of a Mobile Money–Based Conditional Cash Transfer Intervention on Health Care Utilization in Southern Madagascar: Mixed-Methods Study

Mara Anna Franke[1], MSc, MD; Anne Neumann[1], MSc; Kim Nordmann[2], MD, MGH; Daniela Suleymanova[3], MSc; Onja Gabrielle Ravololohanitra[4], MSc; Julius Valentin Emmrich[1*], MD, MSc, PD; Samuel Knauss[1*], MD

[1]Charité Center for Global Health, Charité - Universitätsmedizin Berlin, Charitéplatz1, Berlin, Germany

[2]RWTH Aachen, Aachen, Germany

[3]Sciences Po, Institut d'Etudes Politiques de Paris, Paris, France

[4]Doctors for Madagascar, Antananarivo, Madagascar

[*]these authors contributed equally

Corresponding Author:
Mara Anna Franke, MSc, MD
Charité Center for Global Health, Charité - Universitätsmedizin Berlin, Charitéplatz1, Berlin, Germany

Abstract

Background: Mobile money–based cash transfer interventions are becoming increasingly utilized, especially in humanitarian settings. southern Madagascar faced a humanitarian emergency in 2021-2022, when the second wave of the COVID-19 pandemic and a severe famine affected the fragile region simultaneously.

Objective: This mixed-methods study aims to analyze the impact and factors influencing the success of a mobile money–based conditional cash transfer intervention for health care utilization at 4 primary and 11 secondary facilities in Madagascar.

Methods: We obtained quantitative data from 11 facility registers, detailing patient numbers per month, categorized into maternity care, surgical care, pediatric care, outpatient care, and inpatient care. An interrupted time series analysis, without a control group, was conducted using the end of the intervention in July 2022 as the cut off point. For qualitative data, 64 in-depth interviews were conducted with health care providers, NGO staff, policymakers, beneficiaries, and nonbeneficiaries of the intervention, and was interpreted by 4 independent researchers using reflexive thematic analysis to identify facilitators and barriers to implementation.

Results: The interrupted time series analysis showed a significant negative impact on health care utilization, indicating a reduction in health care–seeking behavior after the end of the cash transfer intervention. The effect was stronger in the slope change of patient numbers per month (defined as $P<.05$), which significantly decreased in 39 of 55 (70%) models compared to the step change at the end of the intervention, which showed a significant but lower change ($P<.05$) in 40% (22/55) of models. The changes were most pronounced in surgical and pediatric care. The key factors that influenced the success of the implementation were grouped across three levels. At the community level, outreach conducted to inform potential beneficiaries about the project by community health workers and using the radio was a decisive factor for success. At participating facilities, high intrinsic staff motivation and strong digital literacy among facility staff positively influenced the intervention. Confusion regarding previous activities by the same implementing NGO and perceptions of unfair bonus payments for health care providers included in the project negatively affected the intervention. Finally, at the NGO-level, the staff present at each facility and the speed and efficiency of administrative processes during the intervention were decisive factors that influenced the intervention.

Conclusions: The conditional cash transfer intervention was overarchingly successful in increasing health care utilization in southern Madagascar in a humanitarian setting. However, this success was conditional on key implementation factors at the community, facility, and NGO levels. In the future, similar interventions should proactively consider the key factors identified in this study to optimize the impact.

(JMIR Mhealth Uhealth 2025;13:e60811) doi:10.2196/60811

KEYWORDS

cash transfer intervention; Madagascar; Sub-Saharan Africa; health care utilization; humanitarian assistance; Africa; mobile; mixed methods study; money; quantitative; qualitative; thematic analysis; policy; service; delivery; health care system; cash; economic; financial; payment; time series

XSL•FO
RenderX

Introduction

Over the past decade, Sub-Saharan Africa (SSA) has seen a substantial rise in the number of people with access to digital technologies. Digital transformation has made a stark difference for previously disenfranchised communities, especially in the field of financial services [1].

One technology that stands out is mobile money, which allows the creation of bank accounts and enables financial transactions using unstructured supplementary service data codes without requiring an internet connection [1]. The use of mobile money has seen an unprecedented rise in SSA over the past years, with over 781 million users in 2022 [2].

Mobile money and digital payment services are increasingly used in providing humanitarian and development aid, including by large multilateral organisations such as the World Bank [3]. The COVID-19 pandemic has resulted in a boost in mobile money–based support schemes, aiming to ensure equitable access to health care during crises [4]. Mobile money–based interventions are particularly feasible in humanitarian settings as they are fast to implement, cost-efficient, and reach people who are structurally excluded or in hard-to-reach areas [1,5].

However, scientific evidence on the effectiveness of such programs in increasing access to health care is scarce. Similarly, research on implementation factors that influence their effectiveness is lacking.

Madagascar, with a population of 29 million [6], experienced two humanitarian crises in 2021 and 2022. First, the COVID-19 pandemic led to a reduction in GDP, increased unemployment, and poverty [7] Second, a severe drought and subsequent famine affected the southern regions of the island, where extreme poverty and malnutrition rates are particularly high [8,9]. Health care access in Madagascar—particularly in the south was already limited before these crises, due to a severely resource-constrained health system and high financial barriers to care [10]. In 2021 and 2022, the additional economic strain from the pandemic and famine put the financial viability of health facilities at risk, which rely on patients' out-of-pocket payments [11,12]. In response, the nongovernmental organisation (NGO), Doctors for Madagascar implemented a mobile money–based conditional cash transfer intervention to improve access to health care and financial stability of health facilities.

In this study, we aim to (1) analyze the impact of this intervention on health care utilization in a humanitarian setting, and (2) to identify factors contributing to its success.

We expect this research to provide evidence for policymakers and implementers of future health care interventions in humanitarian settings.

Methods

Study Setting

Madagascar is one of the least developed countries globally [13]. The country has a maternal mortality rate of 392 deaths per 100,000 live births, an under-5 mortality rate of 71 per 1000 live births, and an average life expectancy of 64 years [6]. Over 40% of health care expenses are paid out-of-pocket [14]. With over 80% of the population living in extreme poverty, financial barriers are a major hindrance to accessing health care [6,8].

This study draws on data from selected health facilities in 7 regions of Madagascar (see Multimedia Appendix 1), where extreme poverty rates range between 78% and 95% [8]. During the intervention, the region was hit by severe drought and famine, putting over 2 million people into acute food insecurity [9]. Simultaneously, the second wave of the COVID-19 pandemic affected the economy severely, leading to a 7% decrease in GDP and a nationwide recession [7,15].

Intervention Description

This study analyzed a mobile money–based conditional cash transfer intervention, aimed at increasing health care utilization; it was implemented from February 2021 to July 2022. The inclusion criteria for patients to benefit from the intervention were (1) presenting for care at a participating facility during the intervention timeframe, (2) belonging to one of the eligible patient groups (detailed below), and (3) registering on the digital mTOMADY platform with a mobile money account (support was available at the health facilities at point of care). The intervention covered 80% of patients' expenses for medications and medical consumables through conditional cash transfers. Patients qualified for the intervention if they sought care for (1) acutely life-threatening conditions, (2) accidents or injuries, (3) pregnancy, childbirth, or postpartum care, or (4) pediatric care. The decision to include an eligible patient in the intervention lied with the treating health care provider, and patients could decline participation at any point.

The costs covered by the intervention were limited to medical consumables and medication due to donor requirements. It did not cover consultation fees, laboratory fees (excluding laboratory consumables), hospitalization, or indirect expenses (eg, transportation costs).

Facilities could express their interest in participating with the implementing NGO, which selected facilities based on previous collaborative experience and geographical location, prioritizing those in underserved regions with high poverty rates. The NGO staff members were employed at each participating facility to support the administration of the intervention.

Once a patient completed treatment, the health facilities submitted claims for reimbursement through a digital platform for health care–related payments developed and provided by the German-Malagasy NGO, mTOMADY [16]. The claims were filed by a local employee, usually an administrative clerk at the health facility. Claims contained patient sociodemographic data (eg, age, gender, family size), medical information (diagnoses, symptoms, quantities of medication received, and consumables used), and cost data (eg, prices of each type of medication and medical consumables). All claims underwent scrutiny for accuracy by a team of registered doctors at the implementing NGO's central administrative level. Clarification was requested by this team in case of inconsistencies or missing data. Health facilities were reimbursed for approved claims using mobile money. Participating facilities received a small

XSL•FO
RenderX

bonus payment (approximately USD 0.50) for each approved claim.

Data Sources

Quantitative Data

The primary data sources for the interrupted time series (ITS) analysis were routine facility-level registers from January 2021 to December 2022. These registers were maintained by health facilities for reporting to the national health information system. The registers categorized patient numbers per month into outpatient, inpatient, surgical, maternity, and pediatric care. The register data were obtained from 11 of the 15 facilities that participated in the intervention; 4 facilities declined to share their data. One facility provided only total patient numbers per month rather than disaggregated data. The facilities compiled register data from paper-based registers into digital datasheets. We received these datasheets, which were screened for plausibility and outliers by an independent researcher. Clarification was sought from the facilities in case of inconsistencies. Cleaned data sheets from each facility were combined into one data sheet for analysis.

Qualitative Data

Qualitative data were collected through interviews with health care providers, project implementation staff, policymakers, beneficiaries and nonbeneficiaries (individuals who were eligible to participate in the intervention but who opted out) of the intervention . Interviews were conducted between September 9 and November 11, 2022. Tailored interview guides were developed for each participant group to capture their unique perspectives and experiences.

Health care providers were recruited through phone calls or direct visits. We sampled facilities across different regions, including a mix of primary, secondary, public, private, and faith-based providers. Project implementation staff were recruited through phone calls or emails. Policymakers were purposively sampled and contacted via telephone or email, with additional participants identified through snowball sampling. Beneficiaries and nonbeneficiaries were approached in person within their communities with the help of community health workers (CHWs). Communities were selected using a purposive sampling approach, to represent two communities each in proximity and farther from the 2 facilities where the intervention was implemented for the longest duration.

A Malagasy researcher fluent in local dialects and French, conducted the interviews after undergoing comprehensive training in qualitative research and ethics. The study's purpose and objectives were communicated to the participants in their preferred language, and written informed consent was obtained. Interviews were conducted in private settings chosen by participants, in either Malagasy or French, based on participant preference, and were audio-recorded with consent.

Data Analysis

We used a convergent mixed-methods design for this study, in which qualitative data were analyzed based on the findings from the quantitative analysis [17]. We analyzed qualitative data to explain the differences in the impact of the conditional cash transfer intervention on health care utilization at participating facilities.

Quantitative Data

We conducted an ITS analysis using segmented linear regression without a control group. To assess nonstationarity, we used the Durbin-Watson test to evaluate and adjust for autocorrelation [18]. We did not assess for seasonal trends in the data. Our dataset covered only 2 years, and this limited period and number of data points did not allow for a comprehensive analysis of seasonality. We used the month in which the intervention ended at each facility as the cutoff point and analyzed step and slope changes at that point. For all facilities, the preintervention period was defined as the time from when they joined the intervention until the end of May 2022 (except for one facility, which discontinued the intervention in September 2021). The postintervention period ranged from June 2022 to December 2022 (except for one facility, for which the postintervention period began in October 2021). The start date of the intervention was different for each facility, ranging from March 2021 to December 2021.

We developed separate models per facility for all patients and by patient subgroups (outpatient care, inpatient care, surgical care, maternity care, and pediatric care). As facilities were onboarded to the intervention on a rolling basis and one facility left the intervention early, we did not run a model encompassing all facilities. All statistical analyses were conducted using R Studio, (version 2023.06.1, R Foundation for Statistical Computing) [19]. A P value of $<.05$ was considered statistically significant.

Qualitative Data

For qualitative data, recordings were transcribed verbatim and translated into English by trained interpreters. To ensure accuracy, a native Malagasy speaker conducted random spot checks, comparing transcripts and translations to the original recordings. Researchers anonymized all identifying information before transcription.

Data were securely stored in a password-protected digital database. Four researchers independently coded all interviews using reflexive thematic analysis [20]. Regular meetings were held to ensure coding validity and consistency. We used NVivo (version 12, Lumivero) for all qualitative analyses [21].

Ethical Considerations

Ethical approval for all components of the study was obtained from the University of Heidelberg Ethics Committee (Heidelberg, Germany) under the registration number: S-982/2021. In addition, we obtained formal approval from the district health office, a regional subdivision of the Malagasy Ministry of Health, in each district in which data were collected. For all secondary analyses of patient data, the ethics committee waived additional informed consent as all data were anonymized before being forwarded to the research team. For primary data collection for qualitative interviews, written informed consent was obtained from each participant prior to the interview. All identifying data were removed from the interviews during transcription, and all interviews were pseudonymized before

XSL•FO
RenderX

analysis. Participants received no compensation for their participation in this study.

We referred to the STROBE (Strengthening the reporting of observational studies in epidemiology) and SRQR (Standards for reporting qualitative research) reporting guidelines in the preparation of this manuscript [22,23].

Results

Sample Description

We obtained register data from 11 out of 15 health facilities; 4 facilities refused to share their register data. Multimedia Appendix 2 contains details on these 11 facilities.

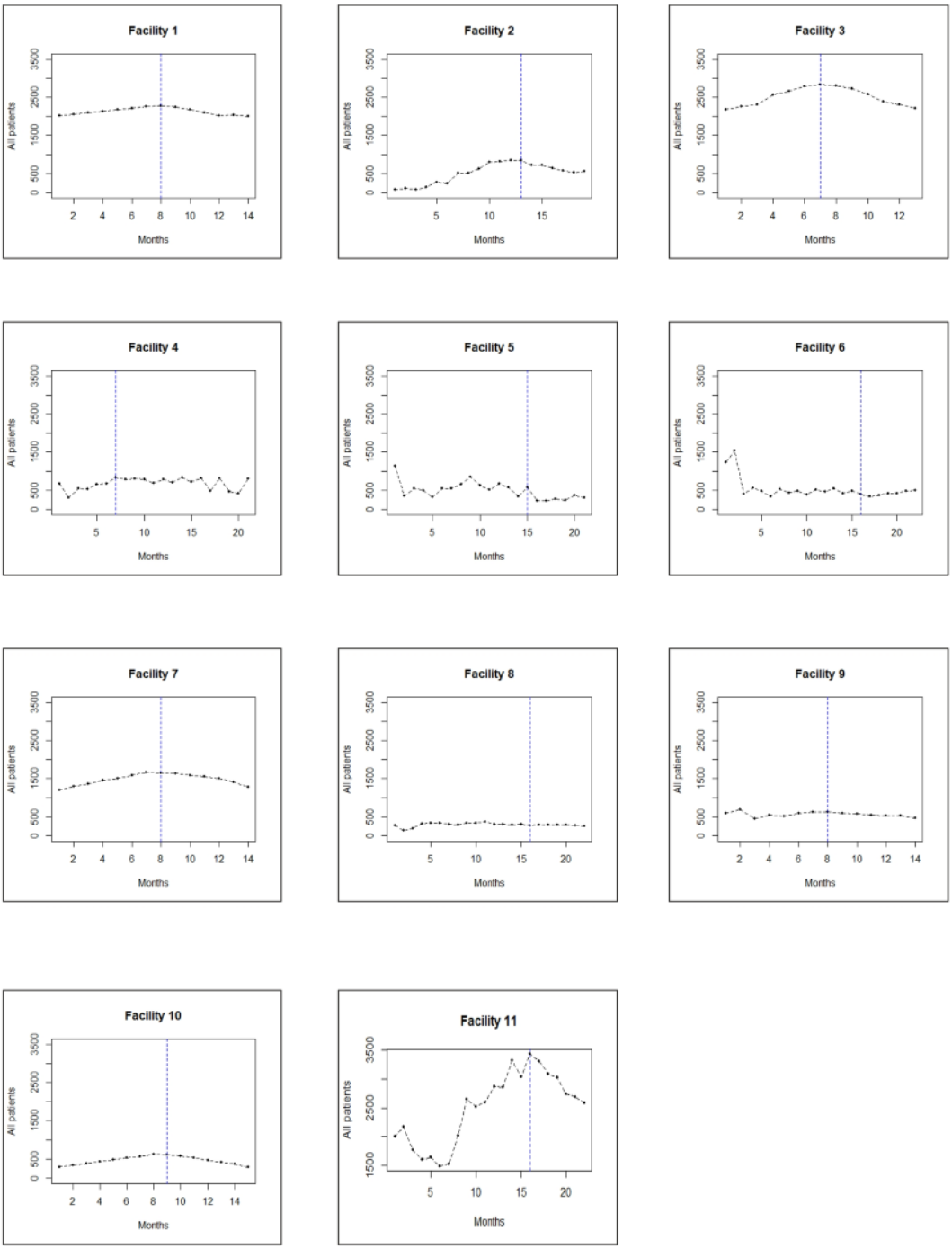
For the qualitative data, we analyzed data from 64 interviews, which lasted between 30 minutes and 1.5 hours, with an average interview duration of 47 minutes. Ten interviews were conducted with project implementation staff, 22 with health care providers, 17 with beneficiaries, 9 with nonbeneficiaries, and 6 with policymakers.

Effect of the Intervention on Health Care Utilization

Overall, health care utilization decreased significantly across most facilities after the end of the intervention. The end of the intervention had a more pronounced effect on patient numbers per month in the long term, which decreased significantly in 39 of 55 (70%) ITS models across separate facilities and patient groups. In comparison, the point effect in patient numbers only showed a significant decrease in 40% (22/55) of all models. Figure 1 illustrates total patient numbers per month at each facility.

XSL•FO
RenderX

Figure 1. Total patient numbers for all patient categories (inpatient, outpatient, surgical, maternity, and pediatric care, depending on services offered by each facility) at 11 health facilities in southern Madagascar that participated in a mobile money–based conditional cash transfer intervention for health care utilization between March 2021 and July 2022. The blue dotted lines mark the end of the intervention at each facility.



The facility with the most pronounced decrease in patient numbers after the end of the intervention was Facility 10, a private facility located in an urban area. Overall, the negative effects of the end of the intervention were most pronounced in maternity, pediatric, and surgical care; whereas, outpatient consultations were the least affected by the end of the intervention.

Table 1 shows the detailed results of the ITS models. The step change values indicate the immediate difference in patient numbers at the end of the intervention, whereas the slope change describes the change in patient numbers per month over time.

XSL•FO
RenderX

Table . Results of interrupted time series analysis using interrupted linear regression, evaluating changes in health care utilization measured by patient numbers per month across 11 health facilities in southern Madagascar after the end of a mobile money-based conditional cash transfer intervention..

Facility categories	Step change in patient numbers per month at end of intervention	*P* value[a]	Slope change in patient numbers per month after end of intervention	*P* value[a]
Facility 1				
All patients	−2.84	.02	−13.134	<.001
Outpatient consultations	−0.61	.56	−3.77	.002
Internal medicine	−2.75	.02	−13.59	.008
Surgical care	−9.10	.007	−53.94	<.001
Maternity care	−3.08	.01	−16.75	<.001
Pediatric care	−2.86	.02	−14.60	.002
Facility 2				
All patients	0.22	.83	−9.43)	<.001
Outpatient consultations	−4.83	0.004	−14.09	<.001
Internal medicine	0.30	.77	−7.51	.002
Maternity care	−3.78	<.001	−14.38	<.001
Pediatric care	1.34	.20	−3.71	<.001
Facility 3				
All patients	−4.12	0.001	−15.95	.007
Outpatient consultations	−2.36	.04	−9.17	<.001
Internal medicine	−4.69	0.003	−7.26	.003
Surgical care	−3.66	<.001	−21.22	<.001
Maternity care	−0.18	.86	−15.28	<.001
Pediatric care	−1.82	.10	−7.17	<.001
Facility 4				
All patients	−1.77	.09	−1.35	.19
Outpatient consultations	−1.77	.09	−1.35	.19
Internal medicine	−1.35	.19	−1.14	.27
Pediatric care	−2.69	.02	−1.47	.16
Facility 5				
All patients	0.69	.49	−0.27	.79
Outpatient consultations	0.71	.49	−0.24	.81
Maternity care	−0.65	.53	−1.79	.09
Facility 6				
All patients	−0.01	.99	1.24	.23
Facility 7				
All patients	−2.12	.06	−11.88	<.001
Outpatient consultations	−1.34	.21	−4.99	<.001
Internal medicine	−3.33	.003	−14.55	<.001
Surgical care	2.42	.04	−6.74	<.001
Maternity care	−0.70	.49	−8.51	<.001
Pediatric care	−2.49	.03	−20.23	<.001
Facility 8				
All patients	1.01	.33	−0.79	.43

XSL•FO
RenderX

Facility categories	Step change in patient numbers per month at end of intervention	*P* value[a]	Slope change in patient numbers per month after end of intervention	*P* value[a]
Outpatient consultations	0.82	.42	−0.77	.45
Internal medicine	0.21	.83	1.97	.06
Surgical care	−0.39	.69	−0.33	.75
Maternity care	0.99	.33	0.93	.37
Pediatric care	−3.65	.007	−3.07	.005
Facility 9				
All patients	−1.12	.29	−1.38	.19
Outpatient consultations	−0.41	.69	0.19	.86
Internal medicine	−0.79	.45	−3.08	.01
Surgical care	−1.72	.12	−1.64	.13
Maternity care	−1.71	.12	2.06	.07
Pediatric care	−0.99	.34	0.34	.74
Facility 10				
All patients	−7.13	.002	.07−49.69	<.001
Outpatient consultations	−2.03	.07	−12.65	.004
Internal medicine	−3.67	.006	−23.29	<.001
Surgical care	−0.14	.89	−8.50	<.001
Maternity care	−3.17	<.001	−28.98	<.001
Pediatric care	−5.43	.004	−21.03	<.001
Facility 11				
All patients	−1.90	.07	−2.72	.01
Outpatient consultations	−0.07	.95	−4.74	.006
Internal medicine	−2.61	.02	−1.69	.11
Surgical care	0.91	.38	−2.14	.04
Maternity care	−0.51	.62	−3.69	.004
Pediatric care	−0.88	.39	−2.62	.02

[a] *P<.05 was considered statistically significant.*

Explanatory Qualitative Findings

The qualitative interviews elucidated 7 key factors that differed between facilities where the ending of the intervention had a significant impact on health care utilization and those where it did not. These factors spanned across 3 levels: the community level, the facility level, and the level of the implementing NGO. Table 2 summarizes these factors, as determined in the qualitative analysis of in-depth interviews with key implementation stakeholders, sorted by factors and facilities, with and without significant changes in health care utilization after the end of the conditional cash transfer intervention.

XSL•FO
RenderX

Table . Key factors influencing the successful implementation of the digital cash transfer intervention..

Key factors	Facilities with significant effects[a]	Facilities without significant effects[b]
Community level		
Community Outreach	Where knowledge about the intervention was high, community members were more likely to seek care	A lack of knowledge about the intervention led to confusion and insecurity and a hesitation to utilize health care.
Facility level		
Facility staff motivation	Facilities where staff expressed high inherent motivation and altruistic mindset performed better and were more adaptive to programmatic challenges.	Facilities, where staff expressed lower intrinsic motivation, were less adaptive to programmatic challenges and performed worse.
Bonus payments paid by the NGO[c]	Bonus payments were perceived as useful to increase motivation.	Expressed that bonus payments were insufficient and demotivating, especially when unequally distributed among facility staff
Preexisting activities of the same implementing NGO	Facilities without previous interventions from the same NGO did not experience confusion about previous interventions which could thus not impact their work.	Confusion regarding the overlap and differences of the conditional cash transfer intervention with previous NGO activities led to a lack of clarity for health care providers.
Digital literacy of facility staff	High levels of digital literacy among facility staff positively impacted intervention uptake.	Low levels of digital literacy among facility staff negatively impacted intervention uptake.
NGO level		
NGO staff present at each facility	Presence of NGO staff at each facility was perceived as useful.	Challenges when NGO staff were absent, for example, on weekends or when they served multiple facilities
Speed and ease of the intervention's administrative processes	Facilities that were onboarded later in the intervention benefited from decreased challenges as frequent issues experienced during the first few months of the intervention had been resolved.	Delays and challenges in the programmatic process of the intervention (especially claims validation and reimbursement) were major barriers, especially for the first facilities onboarded.

[a]Facilities F1, F2, F3, F10, F11.

[b]Facilities F4, F5, F6, F8, F9.

[c]NGO: nongovernmental organization.

Community Level

Community Outreach

For beneficiaries, different perceptions of the outreach conducted as part of the intervention and their subsequent knowledge about the intervention influenced health care utilization. Outreach about the intervention was mostly conducted by CHWs, who held group discussions, mass outreach activities (eg, during market days), or conducted home visits and explained the intervention and its procedures to the population. These activities were usually combined with health education activities conducted by CHWs. Additionally, brief radio spots describing the intervention were broadcast in the intervention area. Additionally, at higher-performing facilities, patients expressed that hearing about the intervention through CHWs or on the radio encouraged them to seek care. Conversely, participants in proximity to lower-performing facilities expressed a lack of information about the intervention and often only heard about it once they had arrived at a health facility.

> *It was a community health worker who talked to me and explained the intervention. (He said,) that we have to go to the hospital when we are sick and not go anywhere (else) for care.* [Beneficiary 5]

> *There is no awareness (raising) here. I did not hear (about the intervention).* [Nonbeneficiary 7]

Facility Level

Facility Staff Motivation

The intrinsic motivation of individual health care providers was different between facilities. Facilities that actively sought to participate in the intervention and where health care providers had a strong interest in its success performed better. At these facilities, health care providers took considerable steps to ensure the functioning of the intervention, including covering patient copayments out of pocket or using personal phones to enroll patients.

> *For the facilities that received it involuntarily, we found several kinds of problems because like any project if there is no commitment from the people it is not possible to accomplish it. For the facilities that accepted it, these facilities welcomed it with their hearts, and we immediately saw that the project worked well and that it brought good to the whole facility.* [Project implementation staff 6]

XSL•FO
RenderX

Bonus Payments Paid by the NGO

With the intervention, facilities received a bonus payment for each claim. At some facilities, the bonus payment was perceived as useful to increase motivation among health care providers, while others expressed that the bonus payments were insufficient.

How bonus payments were distributed among staff members (ie, if everyone got a share or the money was kept by the individual responsible for filing the claims) further contributed to the frustration regarding the bonus payment and lower performance at some facilities.

> *The intervention's close collaborators benefited because they received indemnities, and they did their work well. I saw them running around doing their work and they did it perfectly.* [Health care provider 3]

> *At the time of the intervention, the work was voluminous; the indemnity for the agents is not sufficient. This is the problem. We work here; we work at night when there are many patients.* [Health care provider 15]

Preexisting Activities of the Same Implementing NGO

Overall, facilities where the intervention was more successful in increasing health care utilization were those that had not previously collaborated with the NGO. Facilities that had previously collaborated with the NGO expressed confusion about the differences between the new intervention and interventions implemented by the same NGO previously. The data further indicated an unwillingness to adapt to the new intervention's processes among some facility staff who had worked on these previous projects.

Digital Literacy of Facility Staff

The confidence and skills of health care providers and administrative clerks in using the digital tools required for the administration of the intervention equally influenced the success of the intervention. Facilities where staff had higher levels of digital literacy found the intervention easier to implement than those with lower digital literacy.

NGO Level

NGO Staff Presence at Each Facility

At high-performing facilities, active NGO staff members encouraged the uptake of the intervention by patients and aided the facilities in managing the intervention-associated workload. Because of this positive role of NGO staff, their absence was perceived as difficult. Some NGO staff served several facilities, which caused issues for facilities on the days when NGO staff were absent. Emergencies that occurred at facilities during the night or on the weekends were perceived as equally challenging because the NGO staff were absent.

Further, NGO staff placed at the facilities received a salary from the NGO, whereas the facility employees who also supported the intervention did not receive an additional salary, leading to dissatisfaction among some health care providers.

> *I encourage each patient who arrives there to discuss directly with the doctor and tell him that we have heard this and that. I did not also do a wait-and-see attitude, but I did everything, even if it is not my job, to win the hearts of these facilities, so that they recognise that I am enthusiastic about collaborating.* [Project implementation staff 7]

Speed and Ease of the Intervention's Administrative Processes

Facilities where the intervention had a significant impact on patient numbers tended to be those that joined later, likely benefiting from more established processes at the implementing NGO.

Issues related to the reimbursement of the facilities by the NGO significantly impacted the facilities, as delays in payout caused issues for the facilities related to paying their staff salaries or making prepayments for ordering medication and consumables. These delays were a significant source of frustration.

> *The payment made by (the NGO) was a bit delayed at the beginning. It was difficult for the hospital at that time. The hospital needs to run while the cash, the money was blocked there. It was settled later on, and everything was fine.* [Health care provider 11]

A key factor in causing the delay in the payout was the insufficient quality of the claims filed by the facilities, which was echoed by both health care providers and NGO staff. Health care providers expressed that additional training or guidelines on claim filing could have improved this situation.

Finally, early in the intervention, changes were required to the mTOMADY platform to adapt it for the intervention. These changes were perceived as confusing and challenging for facilities. However, the frequency of changes was reduced over the intervention's lifespan, meaning that facilities onboarded later were less likely to experience such issues.

Discussion

Principal findings

Our study aimed to describe the impact of a mobile money–based conditional cash transfer intervention on health care utilization in a humanitarian setting, as well as to identify key factors that influenced the success of the intervention. Overall, our study showed a significant decline in health care utilization following the ending of a mobile money–based conditional cash transfer intervention in southern Madagascar. Data from 11 facilities showed a marked reduction in patient visits, particularly in maternity, pediatric, and surgical services, with outpatient consultations being the least affected. The qualitative findings highlighted key factors influencing intervention outcomes across three levels. These factors included community outreach, facility staff motivation, ease of the administrative processes of the intervention, digital literacy of facility staff, and the presence of NGO staff at facilities.

Our analysis showed that health care utilization decreased at most facilities when the conditional cash transfer intervention ended. This ties in with other findings from Madagascar, which

XSL•FO
RenderX

demonstrated an increase in health care utilization when user fees were abolished [24]. This change was most pronounced for 3 patient categories: maternity patients, pediatric patients, and patients requiring surgery. Given that surgical care is particularly expensive and likely to cause catastrophic health expenditure in SSA [25-27], this is not surprising. Our findings indicate that patients might forgo necessary surgical care in the absence of financing methods for their care. The evidence on the costs of pediatric care is less comprehensive. However, several studies from SSA suggest that the costs associated with pediatric care for asthma, pediatric surgery, and pneumonia could be catastrophic for households [27-29]. Evidence on the impact of partial cost coverage or fee reduction for pediatric health services is however lacking.

Evidence on similar interventions using mobile money-based cash transfers for health care utilization in humanitarian settings is severely limited. Contrary to previous findings [30], our study revealed a significant decrease in health care utilization following the end of a cash transfer intervention. However, our study analyses a conditional cash transfer intervention, whereas previous studies examined unconditional cash transfers [30,31], in which the money may have been used for other purposes. Our findings do however align with evidence from nonhumanitarian settings, in which health care utilization increased in the presence of targeted cash transfer interventions [32,33], especially for maternity services [34]. Cash-transfer interventions in nonhumanitarian settings, especially in combination with other interventions such as health education or text-based reminders, have been shown to improve risky sexual behavior [35], and are perceived as useful in improving the adherence and management of patients with TB [36]. It must be noted that most cash transfer interventions used physical currency, limiting the comparability of our findings [30-34].

Despite an overarching impact of the conditional cash transfer intervention on health care utilization, the impact was not significant across all facilities. We identified 7 key factors that may explain these differences across facilities.

At the community level, differences in community outreach and thus the intervention-related knowledge of the population across communities was decisive. For this intervention community outreach was only conducted in certain communities. In facilities where the NGO implemented community outreach, this was done via CHWs and radio campaigns. CHWs are vital for improving health care-related knowledge, access to health care and health outcomes in the communities they serve [37,38]. As we did not interview CHWs for this sample, we cannot further elucidate the reasons behind the differences in community outreach they conducted; however, given the evidence on factors influencing CHW performance, including training, supervision, remuneration and workload [38,39], it stands to reason that similar factors were decisive in our setting.

At the facility level, facility motivation and mindset were key determining factors of the success of the intervention. Provider motivation has been described as a key factor in the implementation and adoption of new practices [38,39]. It is noteworthy that the described intervention may have impacted provider motivation and thus either positively or negatively reinforced preexisting motivation levels. For example, the intervention, leading to an increase in patient numbers at the facility, may have increased provider workload which, depending on its perceived manageability, has been identified to be a key determining factor of provider motivation [39]. Equally, the intervention required health care providers to take on additional tasks and may have negatively impacted provider motivation [39].

Bonus payments received for each claim filed, were perceived very differently across facilities. At some facilities where all staff members received benefits through bonus payments, they were perceived as motivating factors. This ties in with evidence from other contexts, which have shown that bonus payments can enhance provider motivation and potentially quality of care [40,41]. This effect is however conditional on how the distribution of bonus payments was perceived among health care providers. Bonus payments perceived as unjust or unfair decrease provider motivation [41-44 ,].

Another key factor was previous NGO interventions at the same facilities, which led to confusion at the facilities with the changes in the procedures of the new intervention. Given the sizeable percentage of public services that are delivered by NGOs in SSA [45], such negative effects of past NGO interventions on future interventions should be examined more closely.

Given the digital nature of the conditional cash transfer intervention, a lack of digital literacy at the facility level was a key factor hampering the intervention. This finding ties in with findings from other studies across SSA [46-51]. In the face of rising digitalisation of health care delivery, governments, NGOs, and multilaterals, should invest in improving the digital skills of health care providers to eliminate a key barrier to the implementation of digital health interventions in the future.

Two factors emerged as decisive at the implementing NGO level. First, NGO staff placed at each facility was perceived as an important success factor for the intervention. NGO staff alleviated health care provider workloads by supporting the intervention's administration [52]. Despite this positive effect, the fact that NGO staff were paid directly by the NGO, may have reinforced perceived inequalities in remuneration and thus furthered pre-existing frustrations [43].

Another noteworthy finding of our study is that facilities which appeared more successful were facilities, where the intervention was implemented later, and which thus benefited from more established and well-running processes. This echoes previous findings on public health interventions, which identified "implementation infrastructure," including robust administrative systems as a key success factor [40,41]. It is vital to consider these challenges for future, similar interventions, for example by incorporating piloting phases into implementation plans to establish robust administrative processes early on.

A key challenge for cash transfer interventions, especially in humanitarian settings, is sustaining the increases in health care utilization after financial support concludes. Based on the findings from our study we propose several strategies to enhance sustainability in similar interventions.

XSL•FO
RenderX

First, strengthening local health financing systems is essential for sustaining service utilization. Future interventions could partner with local governments, private, or non-profit health insurance schemes to transition conditional cash transfers into more permanent financial assistance for vulnerable populations. Developing mechanisms for partial subsidies through national health funds or community-based health insurance schemes can support continued access to essential services without relying solely on external funding.

Second, we observed that the increased health care utilization during the intervention period partly stemmed from targeted community outreach and health education efforts. Sustained CHWs in health promotion can maintain community awareness and perceived value of health care services. Future interventions should invest in CHW training and create resources to enable communities to make informed health decisions, encouraging health care utilization even without direct financial incentives.

Finally, the data analyzed in this study contained several indications that the intervention had a positive impact on the trust expressed by patients in health care providers and the health system. This was particularly observed at facilities where health care providers helped patients in signing up for the intervention, for example by allowing them to use their ID cards to obtain a SIM card, or where they supported patients who were struggling with copayments. This improved trust is crucial as it can lead to sustained improvements in health care-seeking behavior even after financial incentives end. Future implementers should prioritise trust-building measures—such as transparent communication, continuous outreach, and patient-centred care—to cultivate a health care environment where patients feel supported and motivated to seek timely medical assistance.

Limitations

We conducted an ITS analysis without a control group. Given the public health emergencies in southern Madagascar in recent years, which impacted health care utilization in the region, we opted against a historic control group. In 2020 the COVID-19 pandemic heavily impacted the Malagasy population and health system [7]. In 2018-2019, a severe measles epidemic [49,53] and in 2017, a plague outbreak affected the island [52,54]. Choosing control data from health facilities in a different part of the country was equally unreliable. Even though the health system is weak across Madagascar, the south of the island has particular challenges. Geographical access barriers are particularly pronounced [8], and facilities, human resources, and equipment are particularly scarce [10]. Equally, extreme poverty is more pronounced in the study regions than in the rest of the country [8,55]. Lastly, the severe drought and subsequent famine were specific to the study region and did not affect other parts of the country [9].

Four out of fifteen health facilities refused to share their facility registers with us for quantitative analysis. As facilities did not need to give a reason for refusing to share their data, we cannot eliminate the possibility that there is thus a selection bias in our final sample. However, we included facilities with varying levels of impact of the intervention on health care utilization, suggesting that the overarching results were not skewed. Our qualitative data draws on a convenience sample of facilities, which includes most but not all facilities for which a quantitative analysis was conducted, meaning we may not have captured all factors that influenced the success of the intervention.

Further, our qualitative data were collected in Malagasy and translated into English, which may have introduced translation errors. However, all data were translated by a trained interpreter with previous experience in qualitative research and spot-checked randomly for consistency between recording and translation by a native Malagasy speaker.

Conclusions

In conclusion, our study shows that a mobile money–based conditional cash transfer intervention led to an increase in health care utilization in a humanitarian setting, particularly in surgical and pediatric cases. This indicates that similar interventions could be useful for mitigating the effects of future humanitarian crises on health care utilization. Several factors influenced the success of the intervention. Designers and implementers of future, similar interventions should proactively mitigate these factors. Additional researchers should be invested in filling the remaining evidence gaps regarding the use of mobile money–based cash transfer interventions in humanitarian settings.

Acknowledgments

We want to thank everyone who participated in our interviews and shared their views, ideas, and thoughts openly and honestly. We would also like to thank the entire team at Doctors for Madagascar for their support of our research and for making this intervention possible. We further want to thank the local researcher who conducted the qualitative data collection for his dedication and for making this study possible. The German Cooperation for International Development (GIZ) sponsored the intervention analyzed in this study. A grant from the Theracur Foundation specifically supported the research activities. The French-translated version of the manuscript is available (See Multimedia Appendix 3).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Location of health facilities in southern Madagascar that participated in a mobile money–based conditional cash transfer intervention for health care utilization between March 2021 and July 2022. Facilities located in urban areas offering secondary care are marked

XSL•FO
RenderX

with blue dots, facilities located in rural areas offering secondary care are marked with red triangles, and facilities located in urban areas offering primary care are marked with green squares. The capital city of Antananarivo is marked in black. Several facilities (F4, F5, F10) were in the town of Fort-Dauphin in south-eastern Madagascar, leading to a clustering of symbols in the area.
[PNG File, 128 KB - mhealth_v13i1e60811_app1.png]

Multimedia Appendix 2
Level of care, setting, ownership, and intervention period of 11 out of 15 health facilities in southern Madagascar that participated in a mobile money-based conditional cash transfer intervention for health care utilization between March 2021 and July 2022. The remaining 4 facilities refused to share data for analysis and were excluded from this study.
[DOCX File, 15 KB - mhealth_v13i1e60811_app2.docx]

Multimedia Appendix 3
French translation of the manuscript.
[DOCX File, 299 KB - mhealth_v13i1e60811_app3.docx]

References

1. Ricci LA, Ahokpossi C, Belianska A. Digital currency innovations in sub-saharan africa. In: Regional Economic Outlook: Sub-Saharan Africa—Living on the Edge: International Monetary Fund (IMF); 2022. URL: https://www.imf.org/en/Publications/fintech-notes/Issues/2024/03/26/Central-Bank-Digital-Currency-and-Other-Digital-Payments-in-Sub-Saharan-Africa-A-Regional-546859 [accessed 2025-02-17] [doi: 10.5089/9798400273025.063]
2. State of the industry report on mobile money. GSMA. 2023. URL: https://www.gsma.com/mobilefordevelopment/resources/state-of-the-industry-report-on-mobile-money-2023-2 [accessed 2024-02-16]
3. Garcia MH, Moore CMT. The cash dividend: the rise of cash transfer programs in Sub-Saharan Africa. World Bank Group. 2022. URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/435291468006027351/the-cash-dividend-the-rise-of-cash-transfer-programs-in-sub-saharan-africa [accessed 2025-02-17]
4. Joint United Nations Programme on HIV/AIDS. Burkina Faso: providing cash transfers for vulnerable people living with HIV and key populations. UNAIDS and World Food Programme 2022. 2022. URL: https://www.unaids.org/sites/default/files/media_asset/2022_WCA-cash-transfers-UNAIDS-WFP-BurkinaFaso_en.pdf [accessed 2024-01-02]
5. Ahmed H, Cowan B. Mobile money and healthcare use: Evidence from East Africa. World Dev 2021 May;141:105392. [doi: 10.1016/j.worlddev.2021.105392]
6. Madagascar. The World Bank. URL: https://data.worldbank.org/country/madagascar [accessed 2024-01-02]
7. Harisoa R, David BJ. The economic impact of COVID-19 on African countries: the case of Madagascar. SHS Web Conf 2023;163:01039. [doi: 10.1051/shsconf/202316301039]
8. Institut National de la Statistique. Troisième recensement Général de la population et de l'Habitation. antananarivo, madagascar. [french]. INSAT Madagascar. 2019. URL: https://www.instat.mg/?page_id=229, Accessed January 2, 2024 [accessed 2024-01-02]
9. 2022 annual country report highlights. World Food Programme Madagascar. 2022. URL: https://docs.wfp.org/api/documents/WFP-0000148580/download [accessed 2024-01-02]
10. Ministère de la Santé Publique. Plan de Développement du secteur santé 2020-2024. World Health Organization Country Planning Cycle Database. 2019. URL: https://extranet.who.int/countryplanningcycles/planning-cycle-files/plan-de-developpement-du-secteur-sante-2020-2024 [accessed 2024-01-02]
11. Ataguba JE. COVID-19 pandemic, a war to be won: understanding its economic implications for Africa. Appl Health Econ Health Policy 2020 Jun;18(3):325-328. [doi: 10.1007/s40258-020-00580-x] [Medline: 32249362]
12. Kairu A, Orangi S, Mbuthia B, et al. The impact of COVID-19 on health financing in Kenya. PLOS Glob Public Health 2023;3(10):e0001852. [doi: 10.1371/journal.pgph.0001852] [Medline: 37889878]
13. Human development index (HDI). United Nations Human Development Reports. URL: https://hdr.undp.org/data-center/human-development-index [accessed 2024-01-02]
14. Lang E, Saint-Firmin P, Olivetti A, et al. Analyse Du Système de Financement de La Santé à Madagascar Pour Guider de Futures Réformes, Notamment La CSU: Washington [USA] : Health Policy Plus; 2018. URL: https://p4h.world/app/uploads/2023/03/2018_February_Analyse20du20systC3A8me20de20financement20de20la20SantC3A9_Madagascar.x23411.pdf [accessed 2025-01-02]
15. World Health Organization. Madagascar. URL: https://www.who.int/countries/mdg [accessed 2024-01-02]
16. Powering access to healthcare. mTOMADY. URL: https://www.mtomady.com [accessed 2024-05-13]
17. Creswell JW, Plano Clark VL, Westfall GJ. Designing and Conducting Mixed Methods Research, 3rd edition: Sage Publications; 2017.

XSL•FO
RenderX

18. Ali MM. Durbin–Watson and generalized Durbin–Watson tests for autocorrelations and randomness. J Bus Econ Stat 1987 Apr;5(2):195-203. [doi: 10.1080/07350015.1987.10509578]
19. RStudio Team. RStudio: integrated development for R.* Boston, MA. : RStudio, PBC; 2020. URL: http://www.rstudio.com [accessed 2024-02-27]
20. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
21. Dhakal K. NVivo. J Med Libr Assoc 2022 Apr 1;110(2):270-272. [doi: 10.5195/jmla.2022.1271] [Medline: 35440911]
22. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008]
23. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
24. Garchitorena A, Miller AC, Cordier LF, et al. In Madagascar, use of health care services increased when fees were removed: lessons for universal health coverage. Health Aff (Millwood) 2017 Aug 1;36(8):1443-1451. [doi: 10.1377/hlthaff.2016.1419] [Medline: 28784737]
25. Shrime MG, Dare A, Alkire BC, Meara JG. A global country-level comparison of the financial burden of surgery. Br J Surg 2016 Oct;103(11):1453-1461. [doi: 10.1002/bjs.10249] [Medline: 27428044]
26. Okoroh JS, Riviello R. Challenges in healthcare financing for surgery in sub-Saharan Africa. Pan Afr Med J 2021;38:198. [doi: 10.11604/pamj.2021.38.198.27115] [Medline: 33995804]
27. Yap A, Olatunji BT, Negash S, et al. Out-of-pocket costs and catastrophic healthcare expenditure for families of children requiring surgery in sub-Saharan Africa. Surgery 2023 Sep;174(3):567-573. [doi: 10.1016/j.surg.2023.05.010] [Medline: 37385869]
28. Zhang S, Sammon PM, King I, et al. Cost of management of severe pneumonia in young children: systematic analysis. J Glob Health 2016 Jun;6(1):010408. [doi: 10.7189/jogh.06.010408] [Medline: 27231544]
29. Ughasoro MD, Eze JN, Ayuk AC, Obumneme-Anyim I, Akubuilo U, Oguonu T. Economic burden of childhood asthma in children attending a follow-up clinic in a resource-poor setting of southeast Nigeria. Paediatr Respir Rev 2021 Mar;37:74-79. [doi: 10.1016/j.prrv.2020.01.001] [Medline: 32169438]
30. Pega F, Liu SY, Walter S, Lhachimi SK. Unconditional cash transfers for assistance in humanitarian disasters: effect on use of health services and health outcomes in low- and middle-income countries. Cochrane Database Syst Rev 2015 Sep 11;2015(9):CD011247. [doi: 10.1002/14651858.CD011247.pub2] [Medline: 26360970]
31. Tappis H, Doocy S. The effectiveness and value for money of cash-based humanitarian assistance: a systematic review. J Develop Effect 2018 Jan 2;10(1):121-144. [doi: 10.1080/19439342.2017.1363804]
32. Mesmar S, Talhouk R, Akik C, et al. The impact of digital technology on health of populations affected by humanitarian crises: Recent innovations and current gaps. J Public Health Policy 2016 Nov;37(Suppl 2):167-200. [doi: 10.1057/s41271-016-0040-1] [Medline: 27899794]
33. Briaux J, Martin-Prevel Y, Carles S, et al. Evaluation of an unconditional cash transfer program targeting children's first-1,000-days linear growth in rural Togo: a cluster-randomized controlled trial. PLOS Med 2020 Nov;17(11):e1003388. [doi: 10.1371/journal.pmed.1003388] [Medline: 33201927]
34. Hunter BM, Harrison S, Portela A, Bick D. The effects of cash transfers and vouchers on the use and quality of maternity care services: a systematic review. PLOS One 2017;12(3):e0173068. [doi: 10.1371/journal.pone.0173068] [Medline: 28328940]
35. Kuringe E, Christensen A, Materu J, et al. Effectiveness of cash transfer delivered along with combination HIV prevention interventions in reducing the risky sexual behavior of adolescent girls and young women in Tanzania: cluster randomized controlled trial. JMIR Public Health Surveill 2022 Sep 19;8(9):e30372. [doi: 10.2196/30372] [Medline: 36121686]
36. Musiimenta A, Tumuhimbise W, Atukunda E, Mugaba A, Linnemayr S, Haberer J. Digital adherence technologies and mobile money incentives for management of tuberculosis medication among people living with tuberculosis: mixed methods formative study. JMIR Form Res 2023 Apr 12;7:e45301. [doi: 10.2196/45301] [Medline: 37043263]
37. World Health Organization. What do we know about community health workers? A systematic review of existing reviews. Geneva; (human resources for health observer series no. 19). Licence: CC BY-NC-SA 3.0 IGO. 2020. URL: https://creativecommons.org/licenses/by-nc-sa/3.0/igo
38. Swider SM. Outcome effectiveness of community health workers: an integrative literature review. Public Health Nurs 2002;19(1):11-20. [doi: 10.1046/j.1525-1446.2002.19003.x] [Medline: 11841678]
39. Scott K, Beckham SW, Gross M, et al. What do we know about community-based health worker programs? A systematic review of existing reviews on community health workers. Hum Resour Health 2018 Aug 16;16(1):39. [doi: 10.1186/s12960-018-0304-x] [Medline: 30115074]
40. Milat AJ, Bauman A, Redman S. Narrative review of models and success factors for scaling up public health interventions. Implement Sci 2015 Aug 12;10(1):113. [doi: 10.1186/s13012-015-0301-6] [Medline: 26264351]

XSL•FO
RenderX

41. Bulthuis SE, Kok MC, Raven J, Dieleman MA. Factors influencing the scale-up of public health interventions in low- and middle-income countries: A qualitative systematic literature review. Health Policy Plan 2020 Mar 1;35(2):219-234. [doi: 10.1093/heapol/czz140] [Medline: 31722382]
42. Ergo A, Paina L, Morgan L, Washington, DC: United States Agency for International Development, et al. Creating stronger incentives for high-quality healthcare in low- and middle-income countries 2012, Vol. 34. URL: https://api.semanticscholar.org/CorpusID:53054518 [accessed 2024-02-28]
43. Chimhutu V, Songstad NG, Tjomsland M, Mrisho M, Moland KM. The inescapable question of fairness in pay-for-performance bonus distribution: a qualitative study of health workers' experiences in Tanzania. Global Health 2016 Nov 25;12(1):77. [doi: 10.1186/s12992-016-0213-5] [Medline: 27884185]
44. Okonofua F, Ntoimo LF, Yaya S, et al. Effect of a multifaceted intervention on the utilisation of primary health for maternal and child health care in rural Nigeria: a quasi-experimental study. BMJ Open 2022 Feb 8;12(2):e049499. [doi: 10.1136/bmjopen-2021-049499] [Medline: 35135763]
45. Nega B, Schneider G. NGOs, the state, and development in Africa. Rev Soc Econ 2014 Oct 2;72(4):485-503. [doi: 10.1080/00346764.2014.958901]
46. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018;20(5):e10235. [doi: 10.2196/10235]
47. Okano JT, Ponce J, Krönke M, Blower S. Lack of ownership of mobile phones could hinder the rollout of mHealth interventions in Africa. Elife 2022 Oct 18;11:e79615. [doi: 10.7554/eLife.79615] [Medline: 36255055]
48. Kanjo C, Hara J, Kaasbøll J. Digital empowerment for health workers, and implications on EMRs utilisation. J Health Inform Afr 2019;6(2):74-83. [doi: 10.12856/JHIA-2019-v6-i2-262]
49. Chereka AA, Demsash AW, Ngusie HS, Kassie SY. Digital health literacy to share COVID-19 related information and associated factors among healthcare providers worked at COVID-19 treatment centers in Amhara region, Ethiopia: a cross-sectional survey. Inform Med Unlocked 2022;30:100934. [doi: 10.1016/j.imu.2022.100934]
50. Tegegne MD, Tilahun B, Mamuye A, et al. Digital literacy level and associated factors among health professionals in a referral and teaching hospital: an implication for future digital health systems implementation. Front Public Health 2023;11:1130894. [doi: 10.3389/fpubh.2023.1130894] [Medline: 37113180]
51. Owoyemi A, Osuchukwu JI, Azubuike C, et al. Digital solutions for community and primary health workers: lessons from implementations in Africa. Front Digit Health 2022;4:876957. [doi: 10.3389/fdgth.2022.876957] [Medline: 35754461]
52. Muthuri R, Senkubuge F, Hongoro C. Determinants of motivation among healthcare workers in the East African community between 2009–2019: a systematic review. Healthcare (Basel) 2020;8(2):164. [doi: 10.3390/healthcare8020164]
53. Nimpa MM, Andrianirinarison JC, Sodjinou VD, et al. Measles outbreak in 2018-2019, Madagascar: epidemiology and public health implications. Pan Afr Med J 2020;35:84. [doi: 10.11604/pamj.2020.35.84.19630] [Medline: 32537087]
54. Plague – madagascar (online). World Health Organization. URL: https://www.who.int/emergencies/disease-outbreak-news/item/15-november-2017-plague-madagascar-en [accessed 2024-01-02]
55. Institut National de la Statistique et ICF. Enquête Démographique et de santé à Madagascar, 2021 [French]. The DHS Program.: INSTAT and ICF; 2022 Aug. URL: https://www.dhsprogram.com/publications/publication-FR376-DHS-Final-Reports.cfm [accessed 2024-02-02]

ABBREVIATIONS

CHW: Community health worker
ITS: Interrupted time series
NGO: Nongovernmental organisation
SRQR: Standards for reporting qualitative research
SSA: Sub Saharan Africa
STROBE: Strengthening the reporting of observational studies in epidemiology

Edited by L Buis; submitted 22.05.24; peer-reviewed by A Hidki, I Angelov; revised version received 20.11.24; accepted 16.01.25; published 03.03.25.

Please cite as:
Franke MA, Neumann A, Nordmann K, Suleymanova D, Ravololohanitra OG, Emmrich JV, Knauss S
Impact of a Mobile Money–Based Conditional Cash Transfer Intervention on Health Care Utilization in Southern Madagascar: Mixed-Methods Study
JMIR Mhealth Uhealth 2025;13:e60811
URL: https://mhealth.jmir.org/2025/1/e60811
doi:10.2196/60811

XSL•FO
RenderX

© Mara Anna Franke, Anne Neumann, Kim Nordmann, Daniela Suleymanova, Onja Gabrielle Ravololohanitra, Julius Valentin Emmrich, Samuel Knauss. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

SEARCH Study: Text Messages and Automated Phone Reminders for HPV Vaccination in Uganda: Randomized Controlled Trial

Sabrina B Kitaka[1*], MBChB, MMed Paed, PhD; Joseph Rujumba[1], SWSA, MA, PhD; Sarah K Zalwango[2,3], MBChB, MPH; Betsy Pfeffer[4], MD; Lubega Kizza[1], BSc Pharmacy; Juliane P Nattimba[1], MBChB; Ashley B Stephens[4], MD; Nicolette Nabukeera-Barungi[1], MBChB, MMed, PhD; Chelsea S Wynn[4], MPH; Juliet N Babirye[3], MD, MPH, PhD; John Mukisa[5], MBChB, MSc; Ezekiel Mupere[1], MBChB, MMED, MSc, PhD; Melissa S Stockwell[4,6*], MD, MPH

[1]Department of Pediatrics and Child Health, School of Medicine, College of Health Sciences, Makerere University, Kampala, Uganda

[2]Kampala Capital City Authority, Kampala, Uganda

[3]School of Public Health, College of Health Sciences, Makerere University, Kampala, Uganda

[4]Division of Child and Adolescent Health, Department of Pediatrics, Vagelos College of Physicians and Surgeons, Columbia University, 617 W 168th Street Suite 115, New York, NY, United States

[5]Department of Immunology and Molecular Biology, Makerere University College of Health Sciences, Kampala, Uganda

[6]Department of Population and Family Health, Mailman School of Public Health, Columbia University, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Melissa S Stockwell, MD, MPH
Division of Child and Adolescent Health, Department of Pediatrics, Vagelos College of Physicians and Surgeons, Columbia University, 617 W 168th Street Suite 115, New York, NY, United States

Abstract

Background: Cervical cancer is currently the leading female cancer in Uganda. Most women are diagnosed with late-stage disease. Human papillomavirus (HPV) vaccination is the single most important primary preventive measure. While research regarding text message vaccine reminder use is strong in the United States, their use has not yet been demonstrated in a preteen and adolescent population in subSaharan Africa or other low- and middle-income countries.

Objective: The objective of this pilot randomized controlled trial was to assess the impact of vaccine reminders with embedded interactive educational information on timeliness of HPV vaccination in Kampala, Uganda.

Methods: In this randomized controlled trial conducted in 2022, caregivers of adolescents needing a first or second HPV vaccine dose were recruited from an adolescent clinic and three community health centres in Kampala, Uganda. Families (n=154) were randomized 1:1 into intervention versus usual care, stratified by dose (ie, initiation, completion) and language (ie, English, Luganda) within each site. Intervention caregivers received a series of automated, personalized text messages or automated phone calls based on family preference. Five messages were sent before the due date, including both static and interactive educational information, with five follow-up messages for those unvaccinated. Receipt of the needed dose by 24 weeks postenrollment was assessed by χ^2, regression, and Kaplan-Meier with log-rank test. All analyses were conducted using intention-to-treat principles.

Results: Overall, 154 caregivers were enrolled (51.3% for dose 1; 48.7% for dose 2) and 64.3% (n=99) spoke Luganda. Among individuals in the intervention arm, 62% (48/78) requested SMS text message reminders and 38% (n=30) requested automated phone reminders. There was no significant difference in requested mode by HPV vaccine dose or language. Intervention adolescents were more likely to receive the needed dose by 24 weeks (51/78, 65.4% vs 27/76, 35.5%; *P*<.001; RR 1.8; 95% CI 1.3 - 2.6). There was no interaction by dose or language. There was no difference in vaccination between those requesting SMS text message versus phone reminders (32/49, 65.3% vs 19/30, 63.3%; *P*=.86). The number needed to message for one additional vaccination was 3.4 (95% CI 2.2 - 6.8). Kaplan-Meier curves demonstrated more timely vaccination in the intervention arm (*P*<.001).

Conclusions: In this novel trial, SMS text message and automated phone reminders were effective in promoting more timely HPV vaccination in this population.

Trial Registration: ClinicalTrials.gov NCT05151367; https://clinicaltrials.gov/study/NCT05151367

(JMIR Mhealth Uhealth 2025;13:e63527) doi:10.2196/63527

XSL•FO
RenderX

KEYWORDS

human papillomavirus; vaccination; adolescent medicine; global health; SMS; mHealth; HPV; vaccine; adolescent; teen; youth; teenager; Uganda; Africa; randomized controlled trial; RCT; caregiver; low income country; middle income country; subSaharan Africa; chi square; regression; mobile health; text messaging

Introduction

Worldwide, human papillomavirus (HPV) infection is the most common sexually transmitted infection [1], with the highest prevalence in young adults [2]. The majority of HPV-associated morbidity and mortality is due to cervical cancer, which is the fourth most common cancer in women worldwide, with an estimated 604,127 new cases and 341,831 deaths in 2020 [3,4]. SubSaharan Africa overall, has the highest cervical cancer incidence and mortality rates in the world [4]. Cervical cancer is the leading cancer among women in Uganda [4]. Up to 72% of women are diagnosed with late-stage disease [5], and 80% die within 5 years of diagnosis [6]. The age-standardized cervical cancer incidence rate in Uganda is more than 4 times the global average, and the age-standardized mortality rate is nearly 6 times higher [4].

HPV vaccination is the single most important primary preventive measure for HPV [7-9]. A systematic review of HPV vaccine programs across 69 countries over ten years found estimated maximal reductions of 90% for HPV 6/11/16/18 strain infection (ie, the most common cancer-causing strains), 90% for genital warts, and 45% for low-grade and 85% for high-grade cytological cervical abnormalities [10]. HPV also plays an important role in anal, vulvar, penile, and oropharyngeal cancer [4]. HPV vaccines are most effective when adolescents are vaccinated before HPV exposure. HPV vaccination is currently recommended in Uganda for adolescent girls aged 9 - 14 years in a 2-dose series and as a 3-dose series for those aged 15 - 26 years.

However, despite high efficacy and a strong safety profile, global HPV vaccination coverage rates remain low compared to other routinely recommended vaccines [11]. Uganda added HPV vaccination to the National Extended Programme on Immunization in November 2015. However, among the adolescents who receive the first dose of the HPV vaccine, many do not receive the second dose they need to complete the series for full protection. Furthermore, among those who do not receive the second dose, many do not receive it in a timely fashion—that is, within the recommended 12 months—resulting in more days at-risk. In Kampala City, only 45% - 62% of girls in the target age range have received two doses of the vaccine [12]. Similarly, many girls (56.7%) had delayed completion of HPV vaccination [13].

Reminders are an evidence-based approach to improve vaccine receipt. In Uganda, the only current reminder available is a paper appointment card, which can easily be misplaced or destroyed. A countrywide evaluation of HPV vaccination in Uganda highlighted the need for caregiver engagement to improve HPV vaccination rates. It was also noted that there was an inadequate follow-up system for second doses [14,15]. Text messaging or SMS text messaging offers low-cost, scalable opportunities to foster vaccination across a population. A Cochrane review found that text messaging, as a single-method reminder, had the largest effect size with high certainty of evidence (risk ratio [RR] 1.29, 95% CI 1.15 - 1.44; six studies; 7,772 participants) [16]. Text message reminders for HPV vaccination have shown to be effective in several studies within the United States [17-19]. Such reminders have also been used for pediatric vaccination studies in African countries such as Kenya and Nigeria [20-25], and in three other non-African low- or middle-income countries (LMIC) [26-28], but not for adolescent vaccines including HPV vaccine. Adolescent vaccination patterns, knowledge, and perceptions differ from those for pediatric vaccination, and interventions in high-income versus low- or middle-income countries also differ [29]. Therefore, text message performance cannot be extrapolated from those targeted to parents of young children or to parents in high-income settings.

In this pilot randomized controlled trial, we assessed the impact of vaccine reminders with embedded interactive educational information on timeliness of HPV vaccination in Kampala, Uganda.

Methods

Setting, Recruitment, and Informed Consent

This randomized controlled trial RCT (clinicaltrials.gov: NCT05151367; 09/12/2021), funded by the National Institutes of Health (R21CA253604), was conducted in 2022. This study took place at three health centres—Kisenyi, Kawaala, and Kiswa—in Kampala, Uganda, overseen by the Kampala Capital City Authority (KCCA) and located in the suburbs of Kampala City, as well as the Makerere/Mulago/Columbia Adolescent Health Clinic at Mulago National Referral Hospital, the teaching hospital for Makerere University College of Health Sciences. The study health facilities were selected purposively in consultation with KCCA to ensure the inclusion of different city divisions. Each of the study health facilities provides HPV vaccination services to adolescent girls and has affiliated schools, although due to the pandemic, vaccination was not occurring in schools.

Caregivers were eligible to participate if they were (1) the adult parent of an adolescent girl aged 10 - 14 years due for a first or second HPV vaccine dose during the study period, (2) resided in Kampala or the surrounding districts, (3) had their usual care at a study site, (4) spoke English or Luganda, and (5) had a cell phone with text messaging capability. Only one child per family was included. Families were recruited by study staff, including village health teams, either through the health facilities or in the community. Informed consent was obtained from a parent or legal guardian for study participation. On enrollment, a demographic survey was completed.

XSL•FO
RenderX

Randomization

We used central randomization of parent-adolescent dyads with a 1:1 allocation, stratified by dose (ie, initiation, completion), and language (English, Luganda) within each site. The randomization lists, based on permuted blocks with varying block sizes, were prepared within REDCap (Research Electronic Data Capture) electronic data capture tools hosted at Columbia University [30,31]. The statistician and those collecting outcome data remained blinded to arm assignment until analyses were complete.

Intervention

On enrollment, parents randomized to the intervention were able to select if they wanted messages sent either via SMS text messages or as an automated phone call, as well as if they wanted messages in English or Luganda. Messages included both reminders about the needed dose as well as educational information about the importance of the HPV vaccine and were based on a successful previous study [17], as well as 85 key informant interviews with caregivers (n=30), adolescents (n=30), health care workers (n=11), school (n=10), and government officials (n=4). First, team members identified the key content, timing, and framing for each message based on the interviews. The final draft messages were first designed in English, then translated into Luganda by an independent professional translator, and then reverse-translated by a second independent professional translator, while ensuring proper linguistic and cultural equivalence. The Luganda and English messages were then pretested with 24 parents with small iterative changes.

Five messages before the due date and five follow-up messages were automatically sent centrally using a custom-designed SMS text messaging or phone reminder service. Charges were reversed such that no family was charged for sending or receiving messages. Messages included both static and interactive educational information such as the need for timely receipt of HPV vaccines and differed based on which dose was due. Messages for those who had not yet received any HPV vaccine dose began right after enrollment. For those who had received one dose but had not completed the series, messages began either 4 weeks before the dose was due (ie, for those for whom the dose was not yet due) or immediately (if the dose was overdue). Families who reported that their adolescent would be away at boarding school had their message timeline targeted to the next school break. Messages were personalized with the adolescent's first name. Parents randomized to usual care received no vaccine reminders and instead received their usual care.

Measures

Outcomes

The primary outcome was the receipt of the next needed HPV vaccination (dose 1 or dose 2) by 24 weeks postenrollment. Practice staff blind to study arm assignment abstracted vaccination data from the HPV vaccine registry within each enrolling site.

Analyses

For this feasibility pilot study, with a minimal sample size of 77 in each arm, a hazards ratio of 1.67 with 80% power was observed, allowing for a 5% type I error comparing time to next dose.

To assess the primary outcome, we assessed the difference in proportions using χ^2 tests and relative risk estimated from logistic regression. We also assessed for any interaction effects of doses due or language of messages (English or Luganda). The number needed to message for one additional vaccination was then calculated. In addition, the timeliness of vaccination was assessed using Kaplan-Meier analyses with log rank test. All analyses were intention to treat. All analyses were conducted using STATA (version 14.0 MP; STATACorp). A formal statistical analysis plan was drafted and agreed upon prior to the start of analysis.

Ethical Considerations

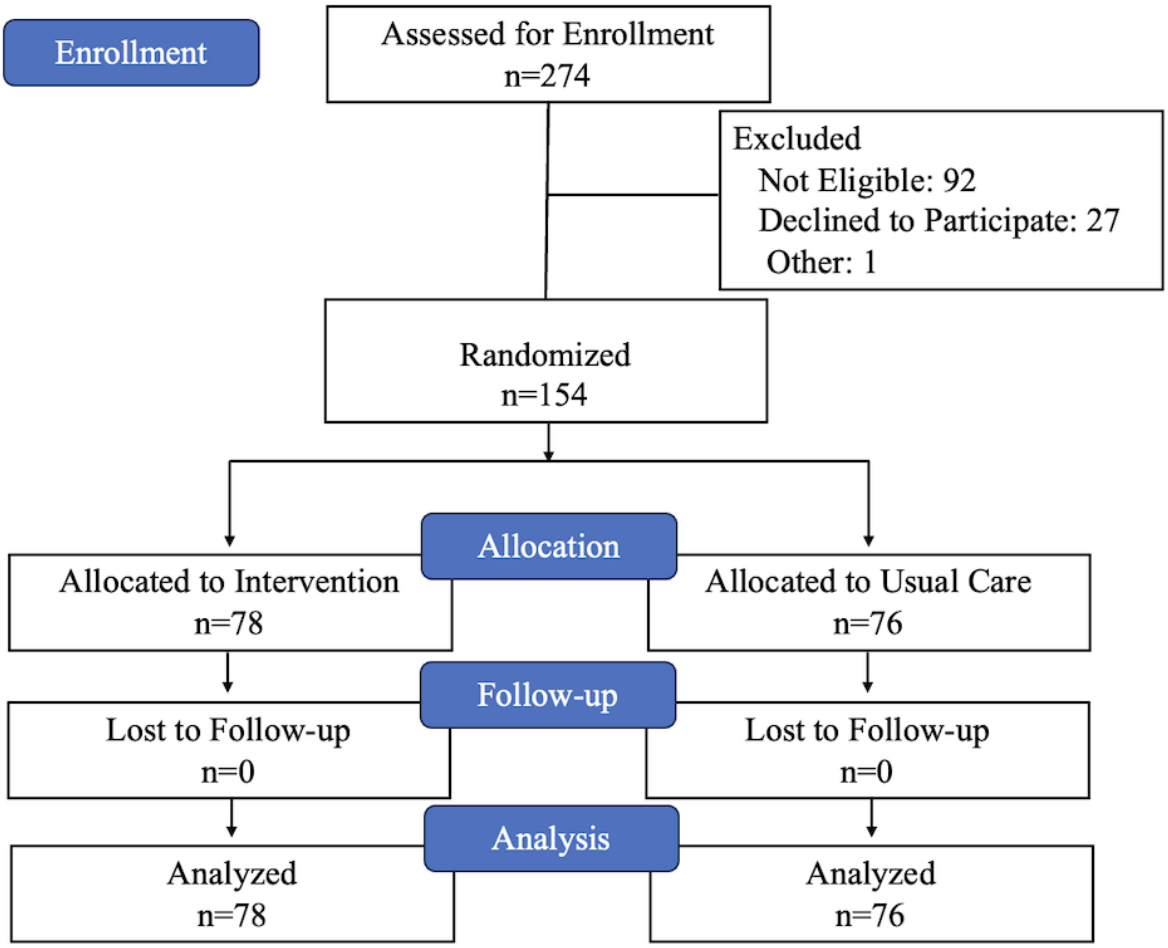
This study was approved by the Institutional Review Boards at Columbia University (AAAT1572) and Makerere University (Mak-SOMREC-2020 - 13) as well as by the Uganda National Council for Science and Technology, with written consent for study participation. Informed consent was obtained from a parent or legal guardian for study participation. All data were securely stored in a study-specific database. Those enrolled received Ugandan Shillings (UgX) 15,000 which is equivalent to approximately US $4 as compensation for their time at the initial interview.

Results

Overall, 274 families were approached and 181 were eligible; of those, 85.1% (n=154/181) were enrolled (Figure 1). Only 8 families were ineligible due to not having a cell phone. Of the 154 dyads enrolled; 51.3% (n=79) had not yet initiated the series and were enrolled to receive dose 1 reminders for initiation; 48.7% (n=75) had already received one dose and were enrolled for dose 2 reminders for completion. No family was ineligible due to not speaking English or Luganda. Nearly two-thirds of those enrolled (64.3%; n=99) spoke Luganda (Table 1), and there were relatively equal numbers enrolled by age group except for a smaller percentage of 14-year-olds. Over three-quarters (77.3%; n=119) were enrolled from the KCCA sites. There was an equal balance between arms for those in the intervention and usual care groups. Of those families randomized to the intervention, 62% (n=48/78) requested the reminders be sent as text messages, with the remaining 38% (n=30) requesting automated phone reminders. There was no significant difference in requested mode based on HPV vaccine dose or language (desired text messages for initiation reminders (22/39, 56%), versus for completion reminders (26/39, 67%; *P*=.35), desired text messages for reminders in English (21/28, 75%) versus in Luganda (27/50, 54%; *P*=.07).

XSL•FO
RenderX

Figure 1. Enrollment flow diagram.



XSL•FO
RenderX

Table . Characteristics of study participants.

Variables	Total (N=154) n (%)	Intervention (n=78), n (%)	Usual care (n=76), n (%)	*P* value
Previous HPV[a] doses received				
0 doses	79 (51.3)	39 (50.0)	40 (52.6)	.74
1 dose	75 (48.7)	39 (50.0)	36 (47.4)	
Site				
KCCA[b]	119 (77.3)	61 (78.2)	58 (76.3)	.78
Mulago	35 (22.7)	17 (21.8)	18 (23.7)	
Language				
English	55 (35.7)	28 (35.9)	27 (35.5)	.96
Luganda	99 (64.3)	50 (64.1)	49 (64.5)	
Age of adolescent				
10 years	32 (20.8)	15 (19.2)	17 (22.4)	.60
11 years	38 (24.7)	20 (25.6)	18 (23.7)	
12 years	31 (20.1)	19 (24.4)	12 (15.8)	
13 years	40 (26.0)	17 (21.8)	23 (30.3)	
14 years	13 (8.4)	7 (9.0)	6 (7.9)	
Health status of adolescent				
Fair	7 (4.5)	3 (3.8)	4 (5.3)	.41
Good	32 (20.8)	14 (17.9)	18 (23.7)	
Very good	37 (24.0)	23 (29.5)	14 (18.4)	
Excellent	78 (50.6)_	38 (48.7)	40 (52.6)	
Highest level education caregiver				
No schooling	20 (13.2)	11 (14.1)	9 (12.2)	.83
Primary or O level	102 (67.1)	53 (67.9)	49 (66.2)	
A level or higher	30 (19.7)	14 (17.9)	16 (21.6)	
Sex of caregiver				
Female	130 (84.4)	66 (84.6)	64 (84.2)	.95
Male	24 (15.6)	12 (15.4)	12 (15.8)	

[a]HPV: human papillomavirus.

[b]KCCA: Kampala Capital City Authority.

Characteristics of the Study Population

Most (130/154, 84.4%) of the caregivers were women; 13.2% (20/152) had no schooling, and 67.1% (102/152) had only completed a primary or ordinary level of education. Approximately 54.9% (84/153) had a regular (ie, not smart) cell phone. Of those with a smartphone, none preferred WhatsApp over text messaging to receive health reminders for their adolescent. Of all those surveyed, while most (136/152, 89.5%) were not worried at all about going over their text message limit, 76.6% (118/154) had run out of airtime or messages in the past 6 months. Most (125/154, 81.2%) received text messages daily. Nearly half (76/154, 49.4%) reported receiving most of their information about the HPV vaccine from a health care provider, 27.9% (n=43) from radio or television, 7.1% (n=11) from family or friends, and 2.6% (n=4) at their adolescent's school. Approximately 41.6% (64/154) knew their adolescent was due for an HPV vaccine dose, and of those 46.9% (30/64) knew when the dose was due. The HPV vaccine perceptions were generally positive, with 96.7% (147/152) reporting that the vaccine was very or somewhat safe, and 98.7% (151/153) reporting that it worked very or somewhat well. Only 44.4% (68/153) thought their adolescent would be protected with only one dose of the vaccine.

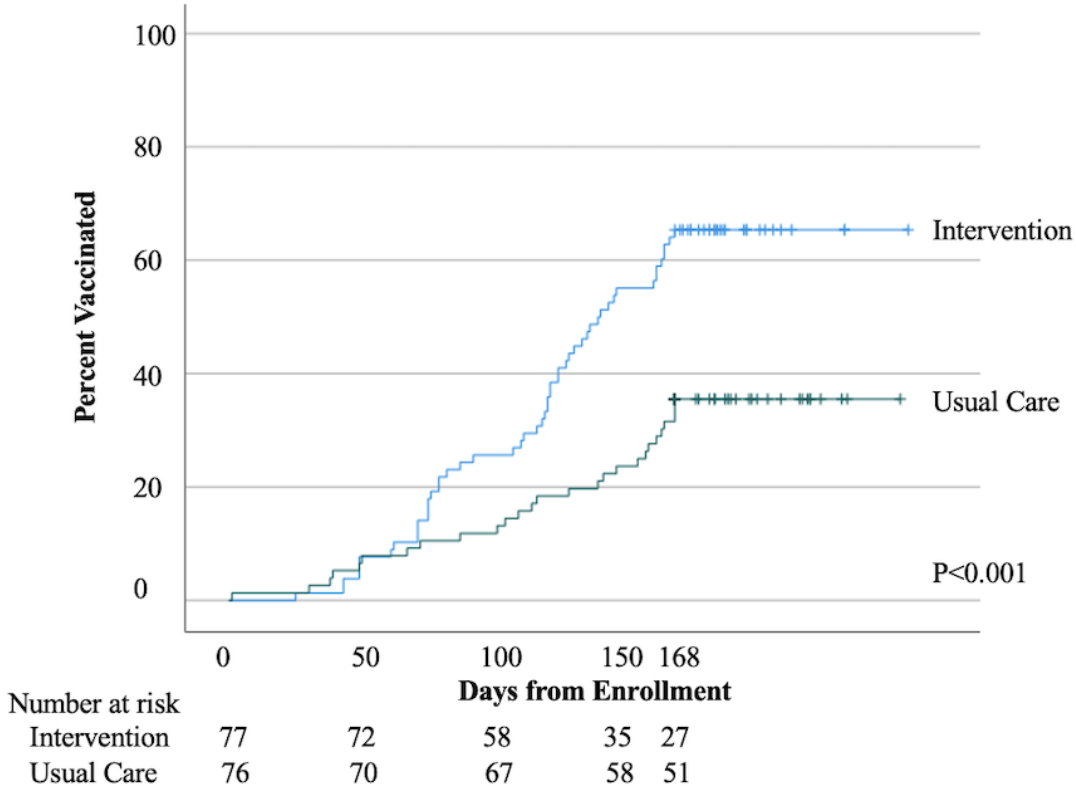
Intervention Effects

Overall, 95.5% (703/736) of text messages were delivered as expected and 90% (9/10) of replies were processed as expected.

XSL•FO
RenderX

In an intention-to-treat analysis, intervention adolescents were more likely to receive a needed dose by 24 weeks (65.4%, 51/78 vs 35.5%, 27/76; *P*<.001; RR 1.9 95% CI 1.3 - 2.6). As an exploratory analysis, there was no interaction by dose needed (*P*=.34) or language (*P*=.22). There was also no difference in vaccination between those who wanted the reminders sent as a text message or as an automated phone reminder (32/49, 65.3% vs 19/30, 63.3%; *P*=.86). The number of families needed to message for one additional girl to be vaccinated was 3.4 (95% CI 2.2 - 6.8). Kaplan-Meier curves demonstrated more timely vaccination (*P*<.001) (Figure 2). On post-hoc analysis, when split by dose, results were similar; the number of individuals who needed dose by 24 weeks for initiation was 66.7% (26/39) versus 27.5% (11/40; *P*<.001; RR 2.17 95% CI 1.34 - 3.52) and for completion was (25/39, 64.1% vs 16/36, 44.4%; *P*=.09; RR 1.54 95% CI 0.93 - 2.58).

Figure 2. Time from first message sent to receipt of next needed dose.



Kaplan-Meier estimates of time from enrollment to receipt of the needed HPV vaccine dose. The x-axis represents time in days since enrollment until receipt of the HPV vaccine dose. The y-axis represents the cumulative proportion of adolescents vaccinated. Follow-up of each child was censored at day 168 or 24 weeks, if the child had not yet received the needed dose.

Postintervention Interviews

We completed postintervention interviews with 30 parents from the intervention arm, including those in need of first versus second doses, those who requested text versus phone reminders, and those who spoke English versus Luganda. Overall, all study participants rated the reminders as highly acceptable and would recommend them to their friends. Some viewed the reminders as a demonstration of "special care" by health workers to have adolescent girls vaccinated and protected against cervical cancer. Participants noted the messages reminded them of when to take their girls for HPV vaccination and provided additional information on the safety and importance of HPV vaccine. A few participants expressed an inability to read messages on their own, receiving few messages, and having busy schedules—highlighting the continued importance of offering an option for automated phone reminders instead of solely text messages. All participants agreed that the intervention should be extended to other community members.

Discussion

This pilot trial demonstrated that SMS text message and automated phone vaccine reminders for HPV vaccination were feasible and acceptable for parents to use in a low-income country such as Uganda, with high user satisfaction. This pilot study also demonstrated effectiveness in a combined outcome of next needed dose—first or second dose—in both overall

XSL•FO
RenderX

vaccination at 24 weeks as well as timeliness of receipt of the needed dose. The overall number of families needed to message for one additional girl to be vaccinated was quite low at 3. Overall, our data suggest that integration of SMS text message vaccine reminders in LMICs is both feasible and accepted by parents and could improve HPV vaccination across a population.

Reminder strategies may be especially important for adolescents, who unlike young children in general have fewer visits with health care providers and therefore, may be less likely to be vaccinated or receive information about the vaccine. In a study conducted in Australia, SMS reminders to parents or guardians who had consented to their adolescent receiving the HPV vaccine within a school-based program led to higher HPV vaccine administration [32]. Similarly, text message reminders for HPV vaccination have been shown to be effective in several studies within the United States [17-19]. In our study, parents in the intervention arm received personalized, scheduled message reminders that included both a reminder about the needed dose as well as educational information about the importance of the HPV vaccine. This addition of information likely proved important, given that only half of the participants knew their adolescent was due for an HPV vaccine, and of those who did, only half of the participants knew when the dose was due. The countrywide evaluation of HPV vaccination in Uganda highlighted the need to increase awareness of the importance of HPV vaccination, as well as evidence-based education and outreach [14,15]. Perceived vaccine safety [5] and perceived benefits of vaccination—particularly cancer protection [33]—have all been noted before as important factors that may impact HPV vaccination in Uganda, all of which were included in our messages.

Some important lessons were learned in this pilot study that could be important for others wanting to use text message vaccine reminders. First, there was high user satisfaction, acceptability, and recommendation ratings, illustrating general interest in these sorts of reminders. Second, only 8 families were ineligible due to not having a cell phone, indicating the potential reach of such interventions. The Global System for Mobile Communications Association Intelligence data highlight that there were 30.55 million cellular mobile connections in Uganda at the start of 2023, which is equivalent to 63.8% of the total population [34]. However, other LMIC may have less cell phone penetration. Third, although a minority, some families opted to receive automated phone reminders, highlighting the potential importance of families being able to select their mode of reminder. There was also no effect on the intervention based on type of reminder selected. In a study of HPV vaccine reminders in the United States, text messages but not phone reminders were effective [18]. It is not clear if the difference here is that families had a choice in their reminder types. Another study, also in the United States but focusing on HPV vaccine series completion, had families choose the modality of the intervention and found completion rates were similar between different recall methods [35]. Of note, only half of the participants in our study had a smartphone, highlighting the continued importance of SMS versus multimedia messaging service or web-linked text messages for this population. Fourth, while most caregivers reported that they were not worried at all about going over their text message limit in a month, three-quarters of them reported that they did run out of airtime or messages in the past 6 months. In this study, we set up reverse billing. It is not clear if the parents would have opted to pay for their own messages. If such reminders were to be used across a larger population for future vaccination support, reverse billing may be needed. Of note, in previous studies in the United States, cost of messages was not highlighted as a concern [36]. Finally, we needed a long lead time to collaborate with a vendor and receive a code from the Uganda Communication Commission; therefore, others may need to keep a long lead time in mind. While not text message reminder-related, it may be important to note that only 44.4% of caregivers surveyed thought their adolescent would be very protected with only one dose of the vaccine. If Uganda moves to a one-dose HPV vaccine, as may be recommended [37], much education may be needed to reassure families that only one dose is now needed.

This study had several limitations. First, our participants were a convenience sample who agreed to take part in an HPV vaccine text messaging study and were from an urban and primarily literate population and therefore may not be representative of the larger population. Similarly, HPV vaccine perceptions regarding safety and efficacy were generally positive among participants. It is unclear if, among a population with more negative views, the intervention would be as successful. However, despite the positive views only one-third of the usual care group were vaccinated within the 24-week period. Third, not 100% of the text messages were delivered as expected. Since all analyses were intention-to-treat, this could have impacted our effect size. We were also not able to confirm delivery of automated phone calls. Fourth, we allowed participants to select text message versus automated phone reminders and therefore could not assess impact if we had not allowed selection and instead sent all participants text messages. Next, this trial took place during the COVID-19 pandemic which could have affected vaccine receipt. However, this should have affected both arms equally. Finally, this pilot study was designed for a combined outcome of next needed dose—first or second dose. Future work will involve conducting a larger, fully powered randomized control study. Despite these limitations, a key strength of this study is that it demonstrated feasibility, acceptability, and pilot effectiveness in over 150 families, recruited from different areas in Kampala, and included English and Luganda speakers. Luganda is one of the most spoken languages in Uganda.

In this novel trial, text message and automated phone vaccine reminders were effective in promoting more timely HPV vaccination. Such reminders could be considered to increase vaccination rates across a population.

XSL•FO
RenderX

Acknowledgments

We would like to thank the Kampala Capital City Authority including the Kisenyi, Kawaala, and Kiswa health centres, the Makerere/Mulago/Columbia Adolescent Health Clinic, and all of the research staff.

This work was supported by the National Institutes of Health (NIH) National Cancer Institute (NCI) grant number R21CA253604 (PI: Stockwell, Bakeera Kitaka).

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1)
[PDF File, 1030 KB - mhealth_v13i1e63527_app1.pdf]

References

1. Martinelli M, Musumeci R, Sechi I, et al. Prevalence of human papillomavirus (HPV) and other sexually transmitted infections (STIs) among Italian women referred for a colposcopy. Int J Environ Res Public Health 2019 Dec 9;16(24):5000. [doi: 10.3390/ijerph16245000] [Medline: 31818033]
2. Rohde RL, Adjei Boakye E, Christopher KM, et al. Assessing university students' sexual risk behaviors as predictors of human papillomavirus (HPV) vaccine uptake behavior. Vaccine (Auckl) 2018 Jun 14;36(25):3629-3634. [doi: 10.1016/j.vaccine.2018.05.022] [Medline: 29753605]
3. Human papillomavirus and cancer. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/human-papilloma-virus-and-cancer [accessed 2023-09-10]
4. Bruni L, Albero G, Serrano B, et al. Human papillomavirus and related diseases in the world. summary report. : ICO/IARC information centre on HPV and cancer (HPV information centre); 2023 URL: https://hpvcentre.net/statistics/reports/XWX.pdf [accessed 2023-09-10]
5. Shaping a strategy to introduce HPV vaccines in uganda-formative research results from the HPV vaccines: evidence for impact project. PATH and Child Health and Development Centre (CHDC). 2009. URL: https://screening.iarc.fr/doc/PATH_FRTS_Uganda.pdf [accessed 2023-09-10]
6. Gondos A, Brenner H, Wabinga H, Parkin DM. Cancer survival in Kampala, Uganda. Br J Cancer 2005 May 9;92(9):1808-1812. [doi: 10.1038/sj.bjc.6602540] [Medline: 15827554]
7. Markowitz LE, Dunne EF, Saraiya M, et al. Quadrivalent human papillomavirus vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2007 Mar 23;56(RR-2):1-24. [Medline: 17380109]
8. Human papillomavirus (HPV). World Health Organization. 2017. URL: https://www.who.int/europe/news-room/fact-sheets/item/human-papillomavirus-(hpv) [accessed 2023-09-20]
9. Meites E, Kempe A, Markowitz LE. Use of a 2-dose schedule for human papillomavirus vaccination - updated recommendations of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep 2016 Dec 16;65(49):1405-1408. [doi: 10.15585/mmwr.mm6549a5] [Medline: 27977643]
10. Garland SM, Kjaer SK, Muñoz N, et al. Impact and effectiveness of the quadrivalent human papillomavirus vaccine: a systematic review of 10 years of real-world experience. Clin Infect Dis 2016 Aug 15;63(4):519-527. [doi: 10.1093/cid/ciw354] [Medline: 27230391]
11. Human papillomavirus (HPV) vaccination coverage. World Health Organization. 2023. URL: https://immunizationdata.who.int/pages/coverage/hpv.html [accessed 2023-10-07]
12. Patrick L, Bakeera-Kitaka S, Rujumba J, Malande OO. Encouraging improvement in HPV vaccination coverage among adolescent girls in Kampala, Uganda. PLoS ONE 2022;17(6):e0269655. [doi: 10.1371/journal.pone.0269655] [Medline: 35679304]
13. Nakendo A, Busingye R, Kakaire O. Prevalence and factors associated with uptake of the second dose of the human papilloma virus vaccine among adolescent girls in kawempe division, kampala. In Review. Preprint posted online on 2022. [doi: 10.21203/rs.3.rs-1921193/v1]
14. HPV vaccination in Africa: lessons learned from a pilot program in Uganda. PATH, Child Health and Development Centre (CHDC), and the Uganda National Expanded Program on Immunization (UNEPI). 2011. URL: https://media.path.org/documents/RH_hpv_lessons_learned_uganda.pdf [accessed 2023-10-27]
15. Gavi Full Country Evaluations Team. Gavi full country evaluations: 2017–2018 dissemination report, uganda. : Kampala, Uganda: IDRC; 2018 URL: https://media.path.org/documents/IDRC_Uganda_Annual_Report_062718_FINAL.pdf
16. Jacobson Vann JC, Jacobson RM, Coyne-Beasley T, Asafu-Adjei JK, Szilagyi PG. Patient reminder and recall interventions to improve immunization rates. Cochrane Database Syst Rev 2018 Jan 18;1(1):CD003941. [doi: 10.1002/14651858.CD003941.pub3] [Medline: 29342498]

XSL•FO
RenderX

17. Wynn CS, Catallozzi M, Kolff CA, et al. Personalized reminders for immunization using short messaging systems to improve human papillomavirus vaccination series completion: parallel-group randomized trial. JMIR Mhealth Uhealth 2021 Dec 27;9(12):e26356. [doi: 10.2196/26356] [Medline: 34958306]
18. Rand CM, Vincelli P, Goldstein NPN, Blumkin A, Szilagyi PG. Effects of phone and text message reminders on completion of the human papillomavirus vaccine series. J Adolesc Health 2017 Jan;60(1):113-119. [doi: 10.1016/j.jadohealth.2016.09.011] [Medline: 27836533]
19. O'Leary ST, Lee M, Lockhart S, et al. Effectiveness and cost of bidirectional text messaging for adolescent vaccines and well care. Pediatrics 2015 Nov;136(5):e1220-e1227. [doi: 10.1542/peds.2015-1089] [Medline: 26438703]
20. Bangure D, Chirundu D, Gombe N, et al. Effectiveness of short message services reminder on childhood immunization programme in Kadoma, Zimbabwe - a randomized controlled trial, 2013. BMC Public Health 2015 Feb 12;15:137. [doi: 10.1186/s12889-015-1470-6] [Medline: 25885862]
21. Gibson DG, Ochieng B, Kagucia EW, et al. Mobile phone-delivered reminders and incentives to improve childhood immunisation coverage and timeliness in Kenya (M-SIMU): a cluster randomised controlled trial. Lancet Glob Health 2017 Apr;5(4):e428-e438. [doi: 10.1016/S2214-109X(17)30072-4] [Medline: 28288747]
22. Haji A, Lowther S, Ngan'ga Z, et al. Reducing routine vaccination dropout rates: evaluating two interventions in three Kenyan districts, 2014. BMC Public Health 2016 Feb 16;16:152. [doi: 10.1186/s12889-016-2823-5] [Medline: 26880141]
23. Brown VB, Oluwatosin OA, Akinyemi JO, Adeyemo AA. Effects of community health nurse-led intervention on childhood routine immunization completion in primary health care centers in Ibadan, Nigeria. J Community Health 2016 Apr;41(2):265-273. [doi: 10.1007/s10900-015-0092-3] [Medline: 26395786]
24. Eze GU, Adeleye OO. Enhancing routine immunization performance using innovative technology in an urban area of Nigeria. West Afr J Med 2015;34(1):3-10 [FREE Full text] [Medline: 26902809]
25. Obi-Jeff C, Garcia C, Adewumi F, et al. Implementing SMS reminders for routine immunization in Northern Nigeria: a qualitative evaluation using the RE-AIM framework. BMC Public Health 2022 Dec 17;22(1):2370. [doi: 10.1186/s12889-022-14822-1] [Medline: 36528596]
26. Domek GJ, Contreras-Roldan IL, O'Leary ST, et al. SMS text message reminders to improve infant vaccination coverage in Guatemala: a pilot randomized controlled trial. Vaccine (Auckl) 2016 May 5;34(21):2437-2443. [doi: 10.1016/j.vaccine.2016.03.065] [Medline: 27026145]
27. Uddin MJ, Shamsuzzaman M, Horng L, et al. Use of mobile phones for improving vaccination coverage among children living in rural hard-to-reach areas and urban streets of Bangladesh. Vaccine (Auckl) 2016 Jan 4;34(2):276-283. [doi: 10.1016/j.vaccine.2015.11.024] [Medline: 26647290]
28. Kazi AM, Ali M, Zubair K, et al. Effect of mobile phone text message reminders on routine immunization uptake in Pakistan: randomized controlled trial. JMIR Public Health Surveill 2018 Mar 7;4(1):e20. [doi: 10.2196/publichealth.7026] [Medline: 29514773]
29. Foss HS, Oldervoll A, Fretheim A, Glenton C, Lewin S. Communication around HPV vaccination for adolescents in low- and middle-income countries: a systematic scoping overview of systematic reviews. Syst Rev 2019 Aug 1;8(1):190. [doi: 10.1186/s13643-019-1100-y] [Medline: 31370886]
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
31. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
32. Tull F, Borg K, Knott C, et al. Short message service reminders to parents for increasing adolescent human papillomavirus vaccination rates in a secondary school vaccine program: a randomized control trial. J Adolesc Health 2019 Jul;65(1):116-123. [doi: 10.1016/j.jadohealth.2018.12.026] [Medline: 30879881]
33. Turiho AK, Okello ES, Muhwezi WW, Katahoire AR. Perceptions of human papillomavirus vaccination of adolescent schoolgirls in western Uganda and their implications for acceptability of HPV vaccination: a qualitative study. BMC Res Notes 2017 Aug 30;10(1):431. [doi: 10.1186/s13104-017-2749-8] [Medline: 28854964]
34. Digital 2023: Uganda. Datareportal. 2023. URL: https://datareportal.com/reports/digital-2023-uganda [accessed 2023-09-10]
35. Kempe A, O'Leary ST, Shoup JA, et al. Parental choice of recall method for HPV vaccination: a pragmatic trial. Pediatrics 2016 Mar;137(3):e20152857. [doi: 10.1542/peds.2015-2857] [Medline: 26921286]
36. Hofstetter AM, Vargas CY, Kennedy A, Kitayama K, Stockwell MS. Parental and provider preferences and concerns regarding text message reminder/recall for early childhood vaccinations. Prev Med 2013 Aug;57(2):75-80. [doi: 10.1016/j.ypmed.2013.04.007] [Medline: 23624252]
37. World Health Organization. Human papillomavirus vaccines: WHO position paper (2022 update). Weekly Epidemiological Record 2022;97(50):645-672 [FREE Full text]

Abbreviations

HPV: Human papillomavirus

XSL•FO
RenderX

KCCA: Kampala Capital City Authority
LMIC: Low- or middle- income countries
REDCap: Research Electronic Data Capture
UgX: Ugandan Shillings

Edited by L Buis; submitted 23.06.24; peer-reviewed by BO Ochieng, D M; revised version received 16.12.24; accepted 11.02.25; published 05.05.25.

Please cite as:
Kitaka SB, Rujumba J, Zalwango SK, Pfeffer B, Kizza L, Nattimba JP, Stephens AB, Nabukeera-Barungi N, Wynn CS, Babirye JN, Mukisa J, Mupere E, Stockwell MS
SEARCH Study: Text Messages and Automated Phone Reminders for HPV Vaccination in Uganda: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e63527
URL: https://mhealth.jmir.org/2025/1/e63527
doi:10.2196/63527

© Sabrina B Kitaka, Joseph Rujumba, Sarah K Zalwango, Betsy Pfeffer, Lubega Kizza, Juliane P Nattimba, Ashley B Stephens, Nicolette Nabukeera-Barungi, Chelsea S Wynn, Juliet N Babirye, John Mukisa, Ezekiel Mupere, Melissa S Stockwell. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 5.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Mobile Health Intervention (DOT Selfie) in Increasing Treatment Adherence Monitoring and Support for Patients With Tuberculosis in Uganda: Randomized Controlled Trial

Juliet Nabbuye Sekandi[1,2], MD, MS, DrPH; Esther Buregyeya[3], MBChB, MPhil, PhD; Sarah Zalwango[3,4], MBChB, MPH; Damalie Nakkonde[3], DipNurs, BPH; Patrick Kaggwa[1,2], BSc, MHI; Trang Ho Thu Quach[2], MPH, PhD; David Asiimwe[3,5], BSc; Lynn Atuyambe[3], BA, MPH, PhD; Kevin Dobbin[1], PhD

[1]Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA, United States

[2]Global Health Institute, College of Public Health, University of Georgia, Athens, GA, United States

[3]School of Public Health, College of Health Sciences, Makerere University, Kampala, Uganda

[4]Department of Public Health Services and Environment, Kampala Capital City Aunthority, Kampala, Uganda

[5]Uganda National Tuberculosis and Leprosy Control Programme, Ministry of Health, Kampala, Uganda

Corresponding Author:
Juliet Nabbuye Sekandi, MD, MS, DrPH
Department of Epidemiology and Biostatistics
College of Public Health
University of Georgia
100 Foster Road
Athens, GA, 30602
United States
Phone: 1 7065425257
Email: jsekandi@uga.edu

Abstract

Background: Directly observed therapy (DOT) is the standard method for monitoring adherence to tuberculosis (TB) treatment. However, implementing DOT poses challenges for both patients and providers due to limited financial and human resources. Increasing evidence suggests that emerging digital adherence technologies, such as video directly observed therapy (VDOT), can serve as viable alternatives.

Objective: This study aims to evaluate the effectiveness of VDOT compared with usual care directly observed therapy (UCDOT).

Methods: Between July 2020 and October 2021, we conducted a 2-arm, parallel-group, open-label randomized trial with a 1:1 assignment to receive either the VDOT intervention (n=72) or UCDOT (n=72) for treatment adherence monitoring at public health clinics in Kampala, Uganda. Each group was further stratified to ensure equal numbers of males and females. Eligible patients were aged 18-65 years, had a confirmed diagnosis of TB, and were undergoing daily treatment. The VDOT group was provided with a smartphone equipped with an app, while the UCDOT group followed the routine monitoring practices outlined by the Uganda National TB Program. We tested the hypothesis that VDOT was more effective than UCDOT for monitoring medication adherence. The primary outcome was adherence, defined as having ≥80% of the expected doses observed during the 6-month treatment period. An intention-to-treat analysis was conducted, and multivariable logistic regression was used to estimate the effect of the intervention on adherence monitoring. Adjusted relative risk ratios and their corresponding 95% CIs are presented. Secondary outcomes included treatment completion, loss to follow-up, death, and reasons for missed videos in the intervention group.

Results: The intention-to-treat analysis included 142 participants, with 2 excluded due to discontinuation of medication within the first week after enrollment. The median age of participants was 34 (IQR 26-45) years. The median fraction of expected doses observed (FEDO) was significantly higher in the VDOT group compared with the UCDOT group (100, IQR 80-100 vs 30, IQR 10-60, respectively; P<.001). When using a FEDO cutoff of ≥80% to define optimal adherence, 63 of 142 (44%) patients met the threshold, with a significant difference between the VDOT and UCDOT groups (56/71, 79% vs 7/71, 10%, P<.001). After adjusting for confounders, VDOT users were significantly more likely to achieve ≥80% of their expected doses observed compared

XSL•FO
RenderX

with UCDOT users (adjusted risk ratio 8.4, 95% CI 4.16-17.0). The most common reasons for failing to submit videos of medication intake were an uncharged phone battery, forgetting to record videos during medication intake, and losing the smartphone.

Conclusions: Enhanced VDOT was more effective than UCDOT in increasing adherence monitoring among patients with TB in Uganda. This evidence highlights the potential of digital technologies to improve treatment adherence monitoring and support in high TB burden settings with limited human resources.

Trial Registration: ClinicalTrials.gov NCT04134689; http://clinicaltrials.gov/ct2/show/NCT04134689

(JMIR Mhealth Uhealth 2025;13:e57991) doi:10.2196/57991

KEYWORDS

tuberculosis; digital adherence technologies; video-observed treatment; video directly observed treatment; directly observed therapy; adherence; mHealth; Uganda; Africa

Introduction

The End TB Strategy envisions a world free of tuberculosis (TB), with zero deaths, disease, and suffering due to TB by 2035 [1]. In 2022, an estimated 10.6 million new cases were reported, while 1.6 million people died from TB worldwide [2]. Although effective treatments for TB disease have existed for over 50 years, nonadherence to medication remains a common problem among patients and poses a significant obstacle to achieving the goals of the End TB Strategy [1,3]. Nonadherence reduces cure rates, prolongs infectiousness, and results in poor treatment outcomes, including treatment failure, drug resistance, death, and relapse [4-7]. The emergence of drug-resistant TB and multidrug-resistant TB (MDR-TB) is partly attributed to patients not adhering to their medication properly [4]. Globally, an estimated 450,000 incident cases of MDR-TB were reported in 2021, marking a 3.1% increase from 437,000 cases in 2020 [2]. According to a recent meta-analysis that included 148 studies, the global pooled prevalence of MDR-TB was 11.6% (95% CI 9.1%-14.5%), significantly higher than previous estimates [8].

The World Health Organization (WHO) and US clinical practice guidelines recommend directly observed therapy (DOT) as the standard of care for TB treatment [9,10]. DOT involves a health care worker observing a patient in person while they ingest TB medications to monitor adherence and provide support as needed [10]. When implemented correctly, in-person DOT can facilitate high levels of medication adherence and successful treatment of TB disease. Additionally, DOT can aid in the early detection of adherence issues, adverse drug reactions, and worsening TB symptoms [11]. Previous estimates indicate that approximately 50% of patients on long-term treatment for chronic diseases often fail to adhere to their prescribed medication regimens in both developed and developing countries [12]. Several barriers hinder the effectiveness of DOT, including the inconvenience of daily in-person visits to health facilities, a shortage of health care workers, high transportation costs, TB-related stigma, and long patient waiting times at health facilities [13,14].

Digital adherence technologies have recently emerged as promising tools to address barriers to patient support and treatment monitoring at the health system, patient, and structural levels [15-18]. Video directly observed therapy (VDOT) is an innovative smartphone-based system that uses an app to record videos of medication intake, enabling remote monitoring of treatment adherence. VDOT reduces the need for frequent face-to-face meetings, thereby minimizing inconvenience to patients and lowering transportation costs [19,20]. Several observational studies have evaluated the feasibility and acceptability of VDOT, including 2 conducted in Africa [20-26]. Three randomized clinical trials (RCTs) in the United States, United Kingdom, and Moldova have assessed the efficacy of VDOT, demonstrating that it is feasible, acceptable, cost-saving, and convenient for patients [16,27-29]. A pilot study in Kenya found video-observed therapy to be both technically feasible and acceptable to patients and health care professionals [24]. Similarly, studies conducted among patients and health care providers in Uganda have shown that VDOT is feasible and acceptable for monitoring and supporting patients undergoing TB treatment [25,30,31]. A recent systematic review and meta-analysis found that VDOT is effective in improving medication adherence and bacteriological resolution compared with in-person DOT care [32]. However, none of the studies included in the review were conducted in Africa, as no RCTs comparing VDOT with in-person DOT have been published from the region. The aim of our study was to compare the effectiveness of VDOT and usual care directly observed therapy (UCDOT) in observing and monitoring treatment adherence among patients with TB in Uganda.

Methods

Ethical Considerations

This study was approved by the institutional review boards of the University of Georgia (ID PROJECT00000571), Makerere University (protocol 756), and the Uganda National Council for Science and Technology (HS656ES). All participants provided written informed consent in either English or Luganda, based on their language preference. The consent process included assurances regarding the privacy and confidentiality of the data collected during the study. Participants received approximately US $10 as compensation for transportation and the time spent participating in each visit.

Trial Design

The “DOT Selfie” study was a parallel-group, open-label randomized controlled trial involving 144 adult patients with TB at selected treatment clinics in Kampala, Uganda. Participants were randomized into a control group (n=72), which received UCDOT, and an intervention group (n=72), which received VDOT for adherence monitoring. Enrollment into the

XSL•FO
RenderX

RCT occurred between July 13, 2020, and follow-up continued through October 25, 2021. The details of the trial's design and methods were described in our published protocol [33].

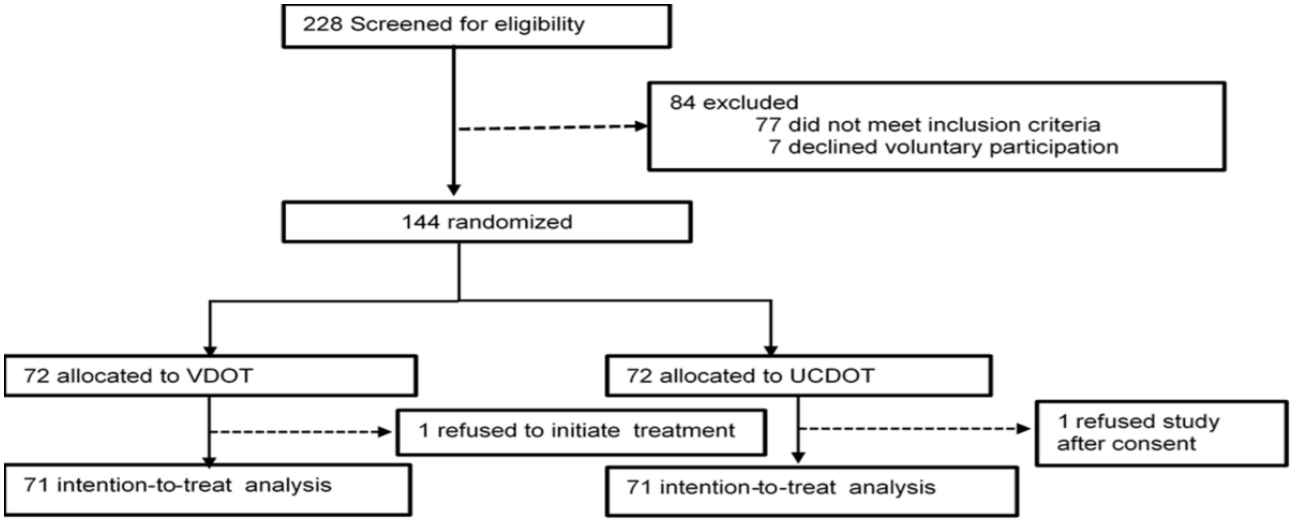
Eligibility Criteria

Eligible participants were adults aged 18-65 years with a confirmed diagnosis of drug-susceptible TB in the new or retreatment category, who had initiated treatment within the past month and planned to reside in Kampala for the entire 6-month treatment period to facilitate close study follow-up. Participants were excluded if they could not speak and read either Luganda (the local dialect) or English, were too ill to withstand the study procedures at enrollment, or had a self-reported motor, visual, hearing, or cognitive disability that would hinder proper use of a smartphone. Additionally, patients residing in areas without cellular network coverage or access to electricity for charging their smartphones were excluded.

Study Setting and Recruitment Procedures

The primary study site was the Lubaga TB clinic, with secondary recruitment sites being public clinics that provide free TB services in Kampala. All TB clinic sites are located about 10-15 km from the Kampala city center and collectively treat approximately 10,000 patients with TB annually. The study team worked closely with clinic nurses, who helped identify potentially eligible patients at the selected public clinic sites. Trained research assistants used a screening checklist to assess study eligibility, and then provided a brief description of the study. The details of the design and methods of the randomized controlled trial are described in our published protocol [33]. The baseline questionnaire was administered after consent, and a copy is provided in Multimedia Appendix 1. Participants were then randomized into the intervention or control study group. The study flow is shown in Figure 1.

Figure 1. DOT Selfie participant flow diagram. UCDOT: usual care directly observed therapy; VDOT: video directly observed therapy.



Randomization and Masking

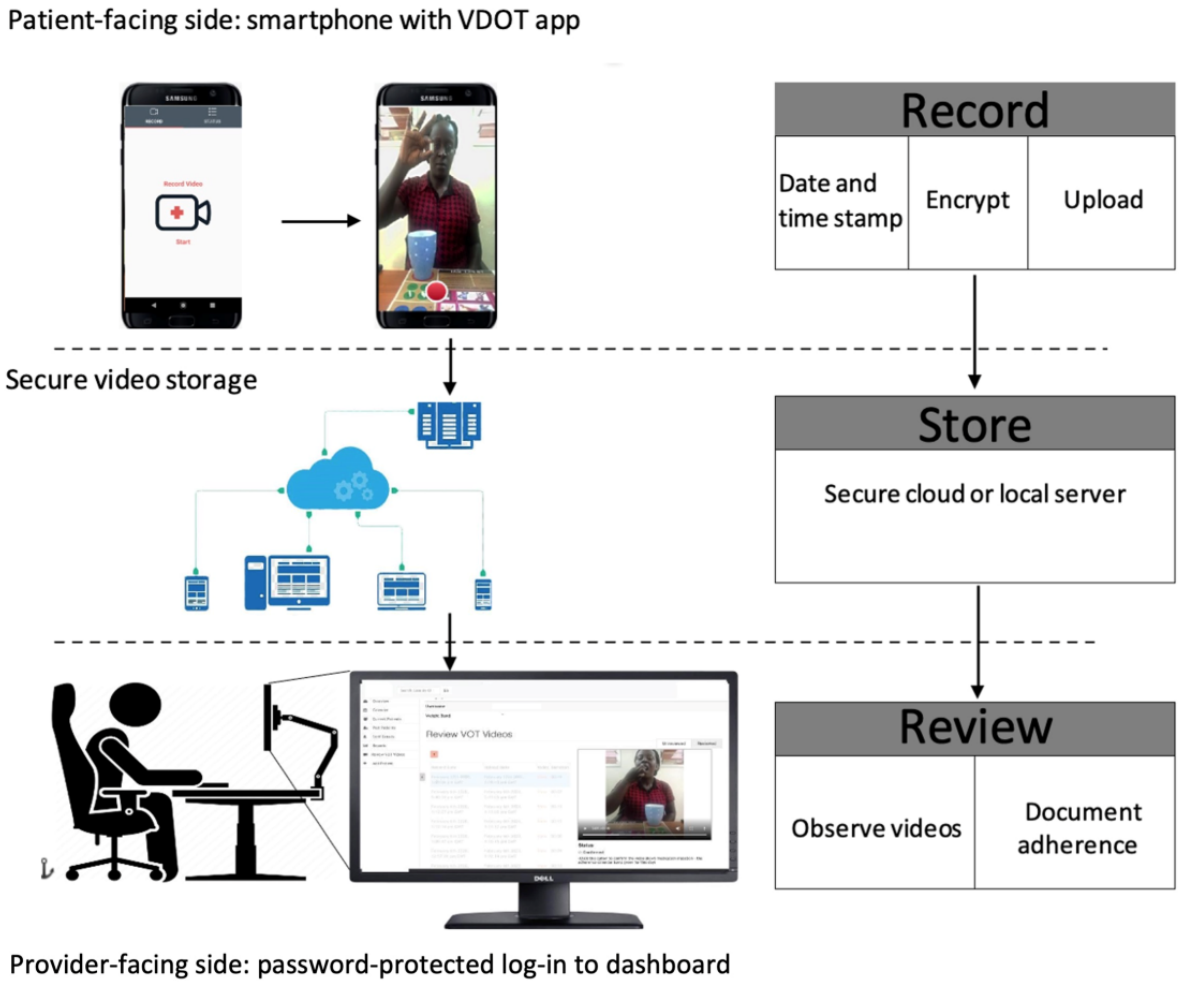
We randomized 144 participants in a 1:1 ratio to receive either VDOT or UCDOT, as shown in Figure 1. We chose equal allocation to optimize the power of the study [34]. Randomization was stratified by sex to ensure equal representation of men and women in each study group [34]. Sex stratification was justified because sex is a known factor associated with adherence to TB treatment [34,35]. Permuted block randomization with block sizes of 4 and 6 was used to ensure a balanced assignment of participants to the intervention or control groups and to minimize the predictability of study arm assignment. Each block contained a specified number of randomly ordered treatment assignments, as recommended in clinical trial methods by Friedman et al [34]. The detailed process followed during randomization is published in the protocol elsewhere [33]. Given the nature of the intervention, participants and study staff were not masked to study group assignment. Study investigators were not blinded to allocation information due to ongoing scientific reviews and the possibility of making clinical decisions about participants' treatment if it became necessary to withdraw them from the intervention. The trial statistician was blinded to study group assignment during the analyses.

Description of the Enhanced VDOT System

VDOT is a licensed Health Insurance Portability and Accountability Act (HIPAA)–compliant system originally developed by SureAdhere Mobile Technology, an initiative that started at the University of California, San Diego [36]. It consists of 3 main components, as described previously by Garfein and colleagues [19] and in our published protocol paper [33]. The components include (1) the patient-facing part, which is the smartphone app used by the patient to record medication videos; (2) the secure cloud-based server, which stores the uploaded medication videos; and (3) the provider-facing part, which is the web-based browser accessed through a secure, password-protected log-in via a computer, laptop, or tablet to review the submitted videos and daily adherence reports [33]. At the end of each video recording, an automatic, encrypted, time-stamped video was uploaded to the secure cloud server for storage and playback (Figure 2). Upon submission, the video was removed from the smartphone, ensuring that patients could not retrieve it.

XSL•FO
RenderX

Figure 2. Schematic of the asynchronous video directly observed therapy system for monitoring tuberculosis treatment (adopted from Sekandi et al [33]).



Enhancement of the VDOT System for Contextual Adaptation

The VDOT system was adapted to the local Ugandan context in 3 ways. First, we expanded the English-based SMS text message reminders to include motivational messages encouraging patients to continue taking their medications and submitting videos. Second, we translated the SMS text messages into Luganda (a local dialect) to facilitate understanding for patients who did not speak English. The following are examples of the SMS text message reminders and motivational messages we used: “It’s time to take your pills and send a video” and “Taking your pills will help you to get better.” Third, an additional SMS text message reminder was sent automatically 8 hours after the standard SMS text message reminder, which was sent to participants at 6:00 AM, if a medication video had not been received in the VDOT system. After adapting the VDOT system to the local context, it was dubbed “DOT Selfie.”

Description of the VDOT System

Description of Standard Procedures

The trial procedures are described in detail in our protocol paper, which is published elsewhere [33]. All participants had an initial face-to-face TB education session, regardless of the assigned study group. The session provided information on TB medications, including dosage, timing, the importance of taking daily pills as prescribed, and what to do in the event of a missed dose or lost pills. In addition, patients received information on cough hygiene behaviors, such as covering their cough to minimize the risk of spreading germs. Lastly, the education emphasized the importance of reporting any issues related to TB medications, including common side effects, adverse events, or new symptoms.

Intervention: Asynchronous Video Directly Observed Treatment

We used an asynchronous VDOT intervention package, which included (1) a loaner GSM (Global System for Mobile Communications) Itel P15 Android smartphone (model W5005) with a preinstalled VDOT app and a preassigned SIM card with a unique phone number; (2) a prepaid weekly internet data subscription of 350 MB at local telecom commercial rates; and (3) automated daily SMS text message reminders to take medications. Participants in the intervention group were enrolled in the VDOT system using a unique log-in personal identification number to set up an account and secure patient information in compliance with HIPAA [33]. The personal identification number also prevented access to the VDOT app by nonstudy participants. The app includes a video feature that

XSL•FO
RenderX

allows patients to record themselves while swallowing their daily doses of medication. In addition, patients received weekly incentives equivalent to US $0.30 in the form of social bundles or airtime minutes, contingent on successfully submitting videos for 7 consecutive days. The weekly incentive was specifically linked to the VDOT intervention, meaning that participants randomized to this study group were eligible to receive the incentive. However, all participants, regardless of the study arm, received a transport refund and reimbursement for their time at each visit, equivalent to about US $10. On the day of enrollment, participants received detailed training on how to use the VDOT app to record and submit videos of medication intake. They also received an instructional guide with pictorial illustrations (pill mat) showing the systematic process of recording the videos, as well as a description of the most common side effects of TB medications. The "pill mat" served as a visual aid to facilitate easy reporting of medication side effects during the video recording session (see details in Multimedia Appendix 2).

At the clinic, a trained study staff member logged into the secure VDOT system via a computer or tablet to download and review patients' daily videos, document medication adherence, and record any self-reported side effects. The study nurse also tracked missed video recordings and took follow-up actions as predefined in the protocol (see details in Multimedia Appendix 3). These actions included a follow-up call to assess missed doses, address side effects, and resolve any other issues as needed. Other secondary methods used to monitor adherence included pill counts and tracking prescription refills at the clinic.

Control: UCDOT

UCDOT was delivered according to routine practice under the Uganda National TB Program guidelines updated in 2017 [37]. The guidelines specify that the TB treatment supporter may be a health care worker, a trained workplace or community health worker, or anyone the patient chooses to observe them swallowing the tablets. The in-person interaction between the patient and the treatment supporter during treatment should occur at least five times a week, with observed medication doses documented in the treatment card. According to the guidelines, DOT services must be organized to accommodate the patients' circumstances and provided as close to their homes as possible. Some patients may opt to take their medication at the health clinic if they live nearby. The patient and treatment supporter are expected to agree on a mutually convenient meeting place for treatment observation and support. However, there is considerable variation in how DOT is delivered across clinics and among patients in Kampala city. After a confirmed diagnosis, TB medications are prescribed, and patients return for biweekly refill visits during the first 2 months (intensive treatment phase). In the continuation phase, patients return monthly for refill visits and clinical evaluation.

In usual care, adherence is assessed using patients' treatment cards, self-reports, and pharmacy prescription refills during routine clinic visits. The study research staff collaborated with TB clinic nurses or staff, community linkage facilitators, and treatment supporters to collect information on DOT. The research team combined data from all available sources and cross-validated it to the extent possible. Patients in the UCDOT study group were followed up for missed routine visits according to the National TB Program guidelines. This included a follow-up phone call if a patient missed a routine visit to refill a prescription, as well as addressing any issues, such as side effects, as needed. The COVID-19 pandemic caused some disruptions, resulting in unintended modifications to the routine delivery of UCDOT.

Follow-Up of the VDOT Group for Missed Videos and Study Visits

A predefined follow-up protocol was used to contact study participants in case of missed doses or videos (see Multimedia Appendix 3). The research staff made 2 phone call attempts within the first 24 hours of a missed video to establish the dosing history and determine the reason for the missed videos. If there was no response from the participant within 72 hours, the research team escalated the follow-up to a field visit to trace the participant at home or at their workplace. After 2 weeks of active follow-up with no success, the study team waited for the participant to return to the clinic for a routine visit. If the patient failed to show up for the scheduled clinic visit, the study team notified the national TB program staff to seek additional support. According to the standard WHO guidelines, a patient is declared lost to follow-up if they do not return to the clinic for their prescription refills for 2 consecutive months [38].

Study Measurements

Primary Outcome

The primary outcome measure was adherence, defined as the fraction of expected doses observed (FEDO) over the months of prescribed TB treatment or by the end of the study following randomization. In the VDOT group, adherence was directly measured based on medication videos submitted and reviewed by the research staff to confirm medication ingestion. In the control UCDOT group, daily medication adherence was measured through self-reports. This was also supplemented with indirect measures, such as notes on the patients' treatment card and prescription refill visits, as additional indicators of adherence. Study measurements were collected using a questionnaire at baseline and at follow-up during months 2, 4, and 6 to ensure close monitoring of all patients during treatment.

Secondary Outcomes

The secondary outcomes we selected are considered unfavorable outcomes and are also relevant indicators of performance for national TB programs. These include treatment completion, loss to follow-up, death, and treatment failure. The definitions of these outcomes were based on the WHO [38] and Uganda National TB Program guidelines [37]. Treatment completion was defined as completing all medication doses as prescribed by the TB clinic, as confirmed by the clinic records. Loss to follow-up was defined as an interruption in TB treatment for at least two consecutive months. Death was considered any death during the treatment period, regardless of the cause. Treatment failure was defined as a patient with TB whose sputum smear examination is positive at month 5 or later during TB treatment [38]. Treatment failure was difficult to ascertain due to

XSL•FO
RenderX

incomplete clinic records. However, no cases of treatment failure were documented among the participants.

Sample Size Estimation

The trial was powered at 80% to detect a 22% difference in medication adherence, the primary outcome, between the 2 groups: VDOT (85%) and UCDOT (63%), based on a 2-sided significance level of 5%. The adherence levels for the comparison groups were derived from previous DOT and VDOT studies conducted in Kampala, Uganda [25,39]. To achieve a chi-square test with a difference of 0.22 and an odds ratio of 0.30, the estimated total sample size was 124. We assumed an attrition rate of 14%, based on a randomized controlled trial comparing usual care and digital adherence interventions in Kenya, where the overall loss to follow-up rate was 11.7% (9.9% in UCDOT and 1.76% in VDOT) [40]. The calculated final sample size was 144 participants, with 72 per group. When calculating power and sample size, we treated the response as binary (0 or 1), with a cutoff value for nonadherence as the FEDO <80%, and adherence as the fraction ≥80%. This binary response variable for adherence was used in the final regression analysis to facilitate comparison with previous clinical trials [16]. Sample size tables were used to estimate precision, as described in detail elsewhere in our protocol paper [33]. All calculations were performed using the SAS statistical software package, version 9.4 (SAS Institute).

Statistical Analyses

We conducted an intention-to-treat analysis according to the study group to which patients were originally randomized. Superiority was determined by a 22% difference in the proportion of patients achieving the primary outcome (63% vs 85%). VDOT treatment observation was classified as completed if patients' medication doses were observed via video. UCDOT treatment observation was classified as completed if the patient reported that a designated treatment supporter observed the ingestion of prescribed doses. Univariate analysis was performed to describe the baseline characteristics of the study population. Multivariable logistic regression analysis was conducted to test for significant associations between the study groups (as the main exposure) and the primary binary outcome of adherence (no <80%, yes ≥80%). Age and HIV status were selected as potential confounders and were adjusted for in all models. Other covariates considered for adjustment during the primary analysis included education level, income, and smartphone ownership at baseline. We did not adjust for sex because randomization was stratified by this variable. Crude and adjusted risk ratios with 95% CIs are presented, and statistical significance was considered at *P* values less than .05.

Results

Overview

Of the 228 patients with confirmed TB who were screened, 144 met the eligibility criteria and were randomized into 1 of the 2 study groups, each with 72 participants. Two participants were excluded from the intention-to-treat analysis (n=142): 1 participant in the VDOT group refused to initiate TB treatment shortly after study enrollment, and another participant in the UCDOT group declined participation after initiating treatment. The 2 participants were included in the baseline analysis but excluded from any further analyses because they did not have the study outcomes of interest (Figure 1 and Multimedia Appendix 4 [41]). The median age of the participants was 34 (IQR 26-45) years. There was a significant difference in median age (*P*=.005) between the study groups, with participants in the UCDOT group being older; 120 of 144 (83.3%) participants owned a cell phone, while 53 of 144 (36.8%) had smartphones. Within the VDOT group, 29 of 72 (40%) owned a smartphone. Overall, 46 of 144 (31.9%) participants were HIV positive. All the measured baseline characteristics were balanced between the study groups, except for age (Table 1).

XSL•FO
RenderX

Table 1. Baseline characteristics of participants by study arm in Kampala, Uganda.

Variables	Total (N=144), n (%)	Video directly observed therapy arm (n=72), n (%)	Usual care directly observed therapy arm (n=72), n (%)	*P* value[a]
Sex, n (%)				
Male	72 (50.0)	36 (50.0)	36 (50.0)	>.99
Female	72 (50.0)	36 (50.0)	36 (50.0)	
Age (years), median (IQR)	34 (26-45)	29.5 (24-42)	38 (28-47)	*.005*
Age category (years), n (%)				
18-24	27 (18.8)	19 (26.4)	8 (11.1)	*.02*
25-34	48 (33.3)	27 (37.5)	21 (29.2)	
35-44	30 (20.8)	10 (13.9)	20 (27.8)	
45-65	39 (27.1)	16 (22.2)	23 (31.9)	
Highest level of education, n (%)				.12
No education or primary	55 (38.2)	27 (37.5)	28 (38.9)	
Secondary	64 (44.4)	28 (38.9)	36 (50.0)	
Tertiary/university	25 (17.4)	17 (23.6)	8 (11.1)	
Marital status, n (%)				.33
Currently married	55 (38.2)	23 (31.9)	32 (44.4)	
Previously married	23 (16.0)	13 (18.1)	10 (13.9)	
Never married	66 (45.8)	36 (50.0)	30 (41.7)	
Currently employed, n (%)				.61
No	54 (37.5)	29 (40.3)	25 (34.7)	
Yes	90 (62.5)	43 (59.7)	47 (65.3)	
Monthly personal income (US $[b]), median (IQR)	42.85 (14.29-85.71)	42.85 (14.29-114.30)	28.57 (11.43-85.71)	.27[c]
Monthly total household income (US $), median (IQR)	57.14 (28.57-142.86)	71.43 (28.57-157.14)	57.14 (28.57-114.23)	.56[c]
Currently owns a cell phone, n (%)				.50
No	24 (16.7)	14 (19.4)	10 (13.9)	
Yes	120 (83.3)	58 (80.6)	62 (86.1)	
Currently owns a smartphone, n (%)				.31
No	67 (46.5)	29 (40.3)	38 (52.8)	
Yes	53 (36.8)	29 (40.3)	24 (33.3)	
HIV status, n (%)				.55
Positive	46 (31.9)	24 (33.3)	22 (30.6)	
Negative	98 (68.1)	48 (66.7)	50 (69.4)	

[a]Significant values (*P*>.05) are presented in italics.

[b]US $1=3500 UGX (Uganda Shillings).

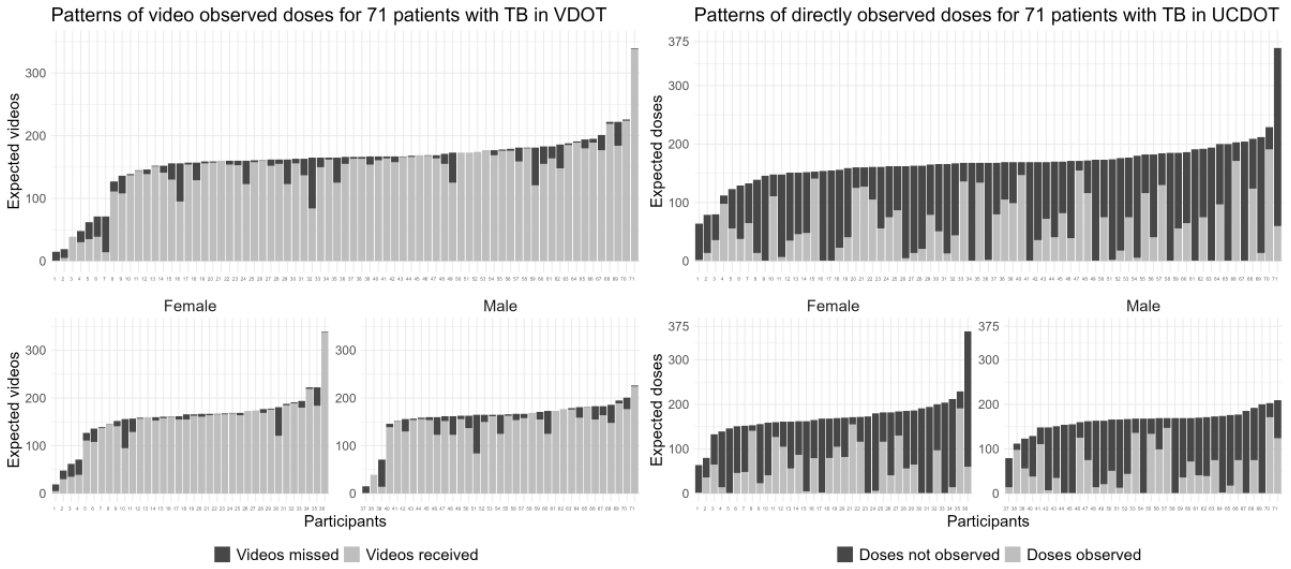
[c]Kruskal-Wallis test for differences in median.

Patterns of Treatment Observation by VDOT and UCDOT Groups and by Sex

A visual illustration of the patterns of treatment observation among study participants is shown in Figure 3. For the VDOT group, the light gray bars represent the videos received via the technology system, while the dark gray bars represent the videos that were not submitted (ie, missed). For the UCDOT group, the light gray bars show the doses that were observed (as reported by the patients) and the dark gray bars show the unobserved doses. We observed clear differences in the patterns of observation between the 2 study groups. However, when the same data are further stratified by sex, there are no obvious differences in the patterns of doses observed between men and women within each group.

XSL•FO
RenderX

Figure 3. Patterns of treatment observation among participants in video directly observed therapy (VDOT) and usual care directly observed therapy (UCDOT). TB: tuberculosis.



Median Differences in Expected and Observed Doses by Study Arm and Sex

There was a significant difference in the median number of doses observed between VDOT and UCDOT (Table 2). However, there were no significant differences in the number of doses when comparing male and female participants within the same study arm (Table 3).

Table 2. Median number of expected and observed doses by study arm.

Doses	Total, median (IQR)	Video directly observed therapy, median (IQR)	Usual care directly observed therapy, median (IQR)	*P* value
Expected videos/doses	167 (158-177)	166 (158-177)	168 (158-178)	.40
Videos/doses observed	118 (39-157)	156 (134-168)	48 (14-98)	<.001
Videos/doses not observed	37 (7-105)	7(2-23)	105 (66-151)	<.001

Table 3. Median number of expected and observed doses within study arms stratified by sex.

Arm and doses	Total, median (IQR)	Male, median (IQR)	Female, median (IQR)	*P* value
Usual care directly observed therapy arm				
Expected doses	168 (158-178)	168 (158-172)	168 (158-184)	.49
Doses observed	48 (14-98)	41 (14-86)	56 (12-99)	.61
Dose unobserved	105 (66-151)	116 (68-151)	104 (64-148)	.96
Video directly observed therapy arm				
Expected videos	166 (158-177)	165 (160-178)	166 (155-175)	.57
Video observed doses	156 (134-168)	154 (134-164)	160 (135-170)	.37
Videos missed, doses not observed	7 (2-23)	10 (3-26)	4 (2-14)	.16

Differences in Adherence Measured by the Fraction of Expected Doses Observed

We evaluated adherence using the FEDO, expressed as a percentage, as shown in Table 4. The overall median FEDO was 80% (IQR 30%-100%), but the median adherence in the VDOT arm was 100% (IQR 80%-100%), which was significantly higher (*P*<.001) than that of the UCDOT group. When using a cutoff of ≥80% for FEDO to define optimal adherence and <80% for nonadherence, only 63 of 142 (44.4%) participants, on average, achieved optimal adherence with their expected doses observed. However, in the VDOT group, 56 of 71 (79%) participants achieved an optimum FEDO, which was significantly higher (*P*<.001) than in the UCDOT group. The median FEDO did not significantly differ by sex in either study arm (*P*=.16 for VDOT and *P*=.96 for UCDOT).

XSL•FO
RenderX

Table 4. Median adherence based on FEDO[a] by study arm.

Variables	Total	Video directly observed therapy	Usual care directly observed therapy	*P* value
Adherence (%), median FEDO (IQR)	80 (30-100)	100 (80-100)	30 (10-60)	<.001
Adherence, with FEDO as a binary outcome (yes ≥80% and no <80%), n (%)				<.001
Yes, n (%)	63 (44)	56 (79)	7 (10)	
No, n (%)	79 (66)	15 (21)	64 (90)	
Total, n	142	71	71	

[a]FEDO: fraction of expected doses observed.

Factors Associated With Treatment Observation as a Measure of Adherence

We performed a multivariable logistic regression analysis to evaluate the factors associated with optimal adherence levels (≥80% doses observed), as measured by the FEDO, with the results presented in Table 5. The results from the unadjusted regression models indicated that being in the VDOT group and identifying as Muslim were independent factors significantly (*P*=.03) associated with achieving adherence of ≥80%. We found that participants in the VDOT group were 8.4 times more likely to have ≥80% of their expected doses observed compared with those in the UCDOT group, after adjusting for confounding factors such as age, religion, current cell phone ownership, and HIV status. Current cell phone ownership was significantly (*P*=.04) associated with reaching the 80% adherence level, based on expected doses observed, compared with those who did not own a cell phone, after adjusting for the study group and other covariates.

XSL•FO
RenderX

Table 5. Factors associated with treatment adherence, as measured by FEDO[a], among study participants in Kampala, Uganda[b].

Variables	Unadjusted risk ration (95% CI)	*P* value	Adjusted risk ration (95% CI)	*P* value
Arm				
Usual care directly observed therapy	1.00	N/A[c]	1.00	N/A
Video directly observed therapy	*8 (3.92 to 16.33)*	*<.001*	*8.41 (4.16 to 17)*	*<.001*
Sex				
Male	1.00	N/A	1.00	N/A
Female	1 (0.69 to 1.45)	.99	0.9 (0.61 to 1.32)	.59
Age category (years)				
45-65	1.00	N/A	1.00	N/A
35-44	0.74 (0.38 to 1.45)	.38	1.02 (0.6 to 1.74)	.95
25-34	1.18 (0.72 to 1.93)	.50	1.06 (0.73 to 1.54)	.75
18-24	1.46 (0.89 to 2.41)	.14	1.03 (0.69 to 1.52)	.89
Highest education level				
No education or primary	1.00	N/A	1.00	N/A
Secondary	0.95 (0.63 to 1.41)	.78	0.88 (0.62 to 1.25)	.46
Tertiary/University	0.90 (0.52 to 1.57)	.71	0.6 (0.36 to 1.01)	.06
Marital status				
Married	1.00	N/A	1.00	N/A
Previously married	1.17 (0.69 to 2)	.56	1 (0.63 to 1.6)	.99
Never married	1.13 (0.75 to 1.72)	.56	0.99 (0.74 to 1.31)	.93
Religion				
Catholic	1.00	N/A	1.00	N/A
Protestant	1.59 (0.93 to 2.72)	.09	1.06 (0.7 to 1.61)	.77
Muslim	*1.75 (1.04 to 2.93)*	*.03*	1.29 (0.87 to 1.91)	.20
Pentecostal/Seventh Day Adventist/Other	1.59 (0.92 to 2.76)	.10	1.01 (0.7 to 1.48)	.94
Currently own a cell phone				
No	1.00	N/A	1.00	N/A
Yes	1.22 (0.7 to 2.12)	.48	*1.62 (1.02 to 2.57)*	*.04*
HIV status				
HIV negative	1.00	N/A	1.00	N/A
HIV positive	0.87 (0.57 to 1.33)	.52	0.8 (0.56 to 1.14)	.22

[a]FEDO: Fraction of expected doses observed.

[b]Significant values (*P*>.05) are presented in italics.

[c]N/A: not applicable.

In the intention-to-treat analysis of 142 participants, the secondary TB outcomes (Table 6) evaluated showed treatment completion at 90.1% (128/142) for all participants, with significantly higher completion rates in the VDOT group compared with the UCDOT group. The rate of loss to follow-up was 6.3% (9/142) overall, but 7 of 71 (10%) participants in the UCDOT group were lost to follow-up, compared with 2 of 71 (3%) in the VDOT group. Overall, there were 5 deaths; 4 of these were HIV coinfected and 3 were male (see details in Table 7).

XSL•FO
RenderX

Table 6. Secondary tuberculosis outcomes.

Outcome	Total (n=142)	Video directly observed therapy (n=71)	Usual care directly observed therapy (n=71)	*P* value
Treatment completion, n (%)	128 (90.1)	67 (94.4)	61 (85.9)	.04
Lost to follow-up, n (%)	9 (6.3)	2 (2.8)	7 (9.9)	.02
Death, n (%)	5 (3.5)	2 (2.8)	3 (4.2)	.16
Treatment failure at 5 months, n	0	0	0	N/A[a]

[a]N/A: not applicable.

Table 7. Summary of deaths.

Participants	Study arm	Sex	Age (years)	HIV status	Number of treatment doses taken
1	VDOT[a]	Male	29	Positive	56
2	VDOT	Female	28	Positive	84
3	UCDOT[b]	Male	40	Positive	112
4	UCDOT	Female	54	Positive	84
5	UCDOT	Male	62	Negative	128

[a]VDOT: video directly observed therapy.

[b]UCDOT: Usual care directly observed therapy.

Reasons for Missed Video Submissions in the VDOT Group

Based on the standard 6-month regimen of 168 doses and the enrollment date, we expected a total of 11,928 videos. Of these, 11,041 (92.56%) were successfully submitted by participants included in the intention-to-treat analysis over the study period. Of the 887 missed videos (7.44%), the most common reason reported by participants was a lack of a charged phone battery, accounting for 183 of the 887 (20.6%) missed videos. Technical or technology issues accounted for 450 (50.7%) of the missed videos. Patient-related issues accounted for 348 (39.2%) missed videos, with the most common reason (136/887, 15.3%) being forgetting to record the videos while swallowing the medication. Additional reasons for missed videos are provided in Table 8.

Table 8. Reasons for missed videos among VDOT[a] study participants in Kampala, Uganda.

Reasons for missed videos	Patients reporting (n=212), n	Videos missed (n=887), n (%)
Technical or technology-related issues		
Phone battery is not charged	39	183 (20.6)
Phone stolen/lost	6	98 (11.0)
VDOT app errors	23	72 (8.1)
Reported lack of internet connection	22	63 (7.1)
Phone malfunction	7	18 (2.0)
Failed to use the app, needed retraining	12	16 (1.8)
Patient-related issues		
Took medications but forgot to record videos	21	136 (15.3)
Traveled to a location that has no electricity	16	74 (8.3)
Too ill to record the video but took medications	7	51 (5.7)
Too busy to record videos	13	40 (4.5)
Location is not convenient to record	7	24 (2.7)
Ran out of tuberculosis medications	16	23 (2.6)
Other not classified		
No reason specified	23	89 (10.0)

[a]VDOT: video directly observed therapy.

XSL•FO
RenderX

Discussion

Principal Findings

In this study, we found that VDOT was significantly more effective in observing medication ingestion and monitoring adherence among patients with TB. Our findings align with a recent systematic review and meta-analysis that included studies conducted in Australia, China, Moldova, the United Kingdom, and the United States. The review concluded that monitoring treatment using VDOT was significantly associated with increased adherence and better microbiological outcomes compared with UCDOT [32]. Additionally, we found that cell phone ownership was significantly associated with a higher level of observed doses, regardless of the study group. Although VDOT was effective, some participants occasionally failed to submit their expected videos of medication intake. The most common reasons for missed videos were lack of a charged phone battery, forgetting to record the medication event, and loss of the smartphone. This finding is consistent with results from our previous pilot study conducted in Kampala, Uganda [25]. To our knowledge, this study is one of the few randomized controlled trials conducted to evaluate the effectiveness of VDOT in Uganda and Africa. More studies are needed to assess the effectiveness of VDOT in rural African contexts and other low-resource settings. Additionally, it is important to identify which subgroups of patients are most likely to benefit from this technology.

There are a few published randomized controlled trials that have compared various forms of VDOT intervention packages with in-person DOT thus far [16,28,29]. Story and colleagues [16] published the first multicenter trial evaluating the efficacy of video-observed treatment compared with usual care in 2019. The study was conducted in the United Kingdom and involved a larger sample of 226 patients from 22 centers. The sample was also highly diverse, with 58% of participants having a history of drug and alcohol abuse, homelessness, and incarceration. In the UK study, 70% of patients on VDOT achieved optimal (≥80%) scheduled observations during the first 2 months, compared with 31% in the UCDOT group. The difference observed in the VDOT arm could be attributed to the difference in the timing of the measurement of the primary outcome, which was 2 months versus 6 months after randomization. The Moldova trial by Ravenscroft and colleagues [28] compared patients assigned to VDOT and clinic-based DOT. The study measured the primary outcome as the observed medication adherence every 2 weeks for the entire 6 months of treatment. The study concluded that VDOT significantly decreased nonadherence compared with standard DOT and reduced the time and money patients spent during their treatment [28]. Although there were some differences in the delivery of the VDOT intervention package and the measurement of the primary outcome, the conclusions on the effectiveness of VDOT are quite similar across studies.

The UCDOT group had a low level (21/71, 30%) of reported observed doses of medication over the 6 months of treatment. This result is very similar to the findings from the UK trial, despite differences in the study populations, delivery of DOT, and treatment durations [16]. This finding further reinforces the fact that in-person DOT is generally difficult to implement, particularly in low-resource settings [39,42]. A Cochrane review of 11 clinical trials concluded that TB cure and treatment completion rates were not improved by DOT [13]. In South Africa, a recent study found that only about 25% of patients with TB received formal DOT support, either home-based or clinic-based [43].

Cell phone ownership was significantly associated with achieving optimal observation of the expected doses of medications, regardless of the study group. Moreover, we found that 120 of 142 (84.5%) patients owned a cell phone, while 53 of 142 (37.3%) had personal smartphones at baseline. This finding suggests that having access to a cell phone could serve as an indirect facilitator of adherence support for patients on treatment. In African settings, smartphone ownership and cellular networks are rapidly proliferating, presenting opportunities to leverage modern technology to improve health care and service delivery. According to the National Institutes of Health, “there is a need to stimulate research utilizing mHealth tools aimed at the improvement of adherence to treatment, effective patient-provider communication, and self-management of chronic diseases in underserved populations” [44]. There is increased utility and effectiveness of accessible mobile technologies in enhancing the monitoring of patients with chronic diseases, which can be attributed to the affordability and reliability of smartphones in both high- and low-income settings [45,46].

The main secondary outcomes, including treatment completion and loss-to-follow-up, were statistically better in the VDOT group compared with the UCDOT group. These findings are consistent with the conclusions of a recent systematic review and meta-analysis [32]. The notably higher percentage of TB/HIV coinfected individuals among the deaths aligns with epidemiologic trends [2]. Treatment failure is an important outcome, but it is often difficult to ascertain, especially with follow-up periods of less than 1 year. To determine treatment failure, we used the records at the end of treatment at 6 months, although it should be routinely assessed at 5 months. We relied on the TB clinic records of sputum results at month 5 because our study protocol included only follow-up visits at months 2, 4, and 6. The study team did not have direct control over the standard clinical evaluations that occur in routine TB management. The clinic records were grossly deficient; therefore, we cannot conclusively state that there were no treatment failures. Larger studies are needed to compare these outcomes across the groups to ascertain the true differences.

Reasons for Failure to Submit Videos

Overall, technical and technology-related issues were the most commonly cited reasons for failure to submit videos. The lack of a charged phone battery was the single most common reason for failure to submit videos, as sometimes phones turned off during video recording. This finding is consistent with results from our pilot study in Kampala [25,47]. To minimize the phone battery issue, patients were trained and encouraged to ensure that their phones were always charged at night or before recording videos. The loss of a smartphone occurred among a

XSL•FO
RenderX

few participants, resulting in a substantial number of missed videos. The study team mitigated the disruption in data collection by replacing lost phones as soon as they were made aware. Missing adherence records were filled using self-reports from patients or clinic records on prescription refills. However, the self-report method falls short because it lacks an objective way to validate the doses taken [48].

The main patient-related reasons that mostly prevented the submission of videos included forgetting to record the medication intake, traveling to locations without access to electricity to charge the phone, or being too ill to record videos. These findings were similar to those reported in our first pilot study [25] and in other studies elsewhere [19,26]. Typically, when patients travel to rural areas in Uganda, there is a high likelihood that they will not have access to electricity, further complicating adherence monitoring. When videos were missed due to travel, the research team followed up to gather the missing information using self-reported adherence. A small number of patients reported a lack of a convenient place to record videos, which was a concern similar to our previous study [25]. This could suggest an underlying stigma associated with TB treatment. However, a study done in Turkey, which compared the levels of stigma among patients with TB using in-person DOT versus VDOT, concluded that patients who received VDOT reported experiencing less stigma [49]. The role of VDOT in mitigating stigma among patients with TB warrants further research, especially in diverse contexts.

Strengths in the Context of Prior Work

First, we evaluated the enhanced VDOT intervention, which is the only digital adherence technology offering remote visual observation of doses taken, making it comparable to the UCDOT. Other commonly used digital adherence technologies for TB monitoring, such as 99 DOTS and digital or smart pillboxes, only provide indirect signals as proxies for doses of medication intake [18]. Second, we successfully evaluated a culturally adapted VDOT in a randomized controlled trial (RCT) within the context of a low-income setting. In our study, we demonstrated that more than one-third of the participants owned a personal smartphone, and these numbers are growing globally. Third, we conducted the study in Kampala, Uganda, where the TB rates are estimated at 250 cases per 100,000, which is much higher than in the majority of settings where VDOT has been previously evaluated [16,19,29]. Fourth, the study facilitated the delivery of a patient-centered approach that was more convenient and less intrusive during the critical period of the COVID-19 pandemic. This also catered to the public health mitigation measures aimed at curtailing the transmission of COVID-19. Lastly, the RCT results generated in our study provide a basis for the systematic evaluation of the implementation process of VDOT. Although there is no gold standard for using digital adherence technologies to monitor TB treatment, our study adds to the mounting evidence that supports the digital observation of pill-taking as an alternative to in-person DOT in usual care.

Study Limitations

First, the study was conducted during the COVID-19 pandemic; therefore, some unintended disruptions in the health system could have altered the practical delivery of usual DOT care in Kampala, Uganda. This may have led to an overestimation of the effect of VDOT. However, a study comparing VDOT and UCDOT before and during COVID-19 found significantly higher adherence to TB treatment when using VDOT during both periods [50]. It is possible that adherence in the UCDOT group could have been greater if not for social distancing mandates and lockdowns. The potential effect of the pandemic period may limit the generalizability of our findings. Second, our study did not specifically measure clinical endpoints, such as sputum conversion at follow-up visits. Although we attempted to extract this information from routine clinic records, the data were incomplete. Third, our study was conducted in an urban setting, which may limit its generalizability to rural settings due to differences in key factors such as cellular network availability, internet coverage, and access to electricity for charging phone batteries. Lastly, the enhanced VDOT intervention included multiple components; however, this study was not designed to isolate the effects of specific components that may have contributed to the differences in adherence observed. Future studies could use a factorial design to assess the impact of different intervention components. There are cross-cutting limitations common to all digital health technologies, such as technology infrastructure, cost, and access to smartphones, which will impact adoption, scale-up, and sustainability, particularly in low-income settings [18].

Considerations for Future Adoption, Scale-Up, and Sustainability of VDOT

We proposed several considerations for the broader adoption, scale-up, and sustainability of VDOT in low-resource settings. Our findings support the need to carefully address the known facilitators and barriers to VDOT acceptability, such as those highlighted in the qualitative exit study of patients who used VDOT in Uganda [30]. First, VDOT requires upfront investment in the health system and local technology infrastructure to facilitate scale-up. A robust system would include the hardware—computers or smartphones—and the software—access to the internet, cellular networks, and maintenance of the VDOT system. There is also an overarching need for adequate access to electricity. Second, national TB programs could leverage public-private partnerships with entities such as telecommunication companies that manage vital resources such as internet coverage for sustainability. Third, standardized training for health providers is essential to facilitate the effective use and adoption of digital technology. Fourth, there is a need to integrate the VDOT system into the general health management information system to ensure interoperability and the timely use of adherence outcome data to improve patient care. The VDOT system could eventually be optimized for monitoring treatment adherence across multiple diseases, such as HIV/AIDS, diabetes, and hypertension. Lastly, patients would need basic training in the use of the VDOT app to ensure ease of use. In our qualitative exit interviews with users, training in the use of the smartphone app was cited as a major facilitator of ease of use and acceptability for VDOT [30]. In general, we expect that health providers and patients will grow more experienced and comfortable with technology over time.

XSL•FO
RenderX

Ethical-Legal Issues and Areas for Future Research

Confidentiality, privacy, and stigma concerns are overarching issues that must be carefully addressed to alleviate fears of intentional or unintended disclosure of a patient's disease status. The flexibility of making video recordings at the patient's desired time and place greatly enhances privacy [51]. In a VDOT study conducted in the United States and Mexico, patients agreed that VDOT promotes autonomy and a sense of control over their health [19]. The ethical concerns related to digital technologies can vary widely, particularly in relation to collecting, securely storing, and accessing videos that contain personally identifiable information, such as patients' faces. These concerns may also have legal and sociocultural contexts that need to be considered [52]. A recent qualitative study focusing on digital health technologies in sub-Saharan Africa concluded that threats to scale-up include the lack of buy-in from both patients and providers, insufficient human resources and local capacity, inadequate governmental support, overly restrictive regulations, and a lack of focus on cybersecurity and data protection [53]. Further research is needed to evaluate the impact of VDOT on patient-centered outcomes, such as satisfaction, quality of life, user engagement with the technology, and retention to prevent dropout. Other important areas of research include the effect of VDOT on health system workflows and workload, as well as its costs and cost-effectiveness compared with usual care.

Conclusions

The enhanced VDOT was more effective in increasing adherence observation to treatment than UCDOT among patients with TB in Uganda. This evidence supports the potential of digital technologies to improve monitoring and support of treatment adherence in high TB burden settings with limited human resources. Further research is needed to evaluate which subgroups are most likely to benefit from this technology.

Acknowledgments

This study is funded by the United States National Institutes of Health (grants 1R21 TW011365 and 3R21 TW011365-02S1). The funders do not have any role in the study design; in the collection, management, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication. The funders did not have any authority over the study activities. We acknowledge Drs. Christopher Whalen, Noah Kiwanuka, and Robert Kakaire for their scientific input during the implementation of the study. We acknowledge Dr. Richard Garfein at the University of California, San Diego, and Dr. Kelly Collins at SureAdhere Mobile Technology, Inc. for the technical support they provided in setting up the project. We thank the research assistants, Mitchell Geno, Daphne Kyaine, and Gloria Nassanga, who collected the data and mobilized the study participants for interviews at the Lubaga study site and other clinics in Kampala, Uganda. Lastly, we acknowledge Everwell Health Solutions for providing the VDOT mobile technology service, training the research team, and providing technical support to the project team.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline questionnaire.
[PDF File (Adobe PDF File), 536 KB - mhealth_v13i1e57991_app1.pdf]

Multimedia Appendix 2
Video directly observed therapy pill mat.
[PNG File , 74 KB - mhealth_v13i1e57991_app2.png]

Multimedia Appendix 3
Video directly observed therapy follow-up protocol.
[PDF File (Adobe PDF File), 101 KB - mhealth_v13i1e57991_app3.pdf]

Multimedia Appendix 4
CONSORT E-HEALTH checklist. CONSORT: Consolidated Standards of Reporting Trials.
[PDF File (Adobe PDF File), 1064 KB - mhealth_v13i1e57991_app4.pdf]

References

1. Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, et al. WHO's new end TB strategy. The Lancet 2015 May;385(9979):1799-1801. [doi: 10.1016/s0140-6736(15)60570-0]
2. World Health Organization (WHO). Global Tuberculosis Report. WHO. Geneva, Switzerland: WHO; 2023 Nov 30. URL: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023 [accessed 2024-01-05]

XSL•FO
RenderX

3. Dick J, Lombard C. Shared vision--a health education project designed to enhance adherence to anti-tuberculosis treatment. Int J Tuberc Lung Dis 1997 Apr;1(2):181-186. [Medline: 9441085]
4. Toczek A, Cox H, du Cros P, Cooke G, Ford N. Strategies for reducing treatment default in drug-resistant tuberculosis: systematic review and meta-analysis. Int J Tuberc Lung Dis 2013 Mar;17(3):299-307. [doi: 10.5588/ijtld.12.0537] [Medline: 23211716]
5. Zegeye A, Dessie G, Wagnew F, Gebrie A, Islam SMS, Tesfaye B, et al. Prevalence and determinants of anti-tuberculosis treatment non-adherence in Ethiopia: a systematic review and meta-analysis. PLoS One 2019;14(1):e0210422 [FREE Full text] [doi: 10.1371/journal.pone.0210422] [Medline: 30629684]
6. Gube A, Debalkie M, Seid K, Bisete K, Mengesha A, Zeynu A, et al. Assessment of Anti-TB drug nonadherence and associated factors among TB patients attending TB clinics in Arba Minch governmental health institutions, Southern Ethiopia. Tuberc Res Treat 2018;2018:3705812 [FREE Full text] [doi: 10.1155/2018/3705812] [Medline: 29670768]
7. Adane A, Alene KA, Koye DN, Zeleke BM. Non-adherence to anti-tuberculosis treatment and determinant factors among patients with tuberculosis in northwest Ethiopia. PLoS One 2013;8(11):e78791 [FREE Full text] [doi: 10.1371/journal.pone.0078791] [Medline: 24244364]
8. Salari N, Kanjoori AH, Hosseinian-Far A, Hasheminezhad R, Mansouri K, Mohammadi M. Global prevalence of drug-resistant tuberculosis: a systematic review and meta-analysis. Infect Dis Poverty 2023 May 25;12(1):57 [FREE Full text] [doi: 10.1186/s40249-023-01107-x] [Medline: 37231463]
9. World Health Organization (WHO). Guidelines for treatment of drug-susceptible tuberculosispatient care, 2017 update. WHO. Geneva, Switzerland: WHO; 2017 Apr 24. URL: https://www.who.int/publications/i/item/9789241550000 [accessed 2017-04-24]
10. Nahid P, Dorman SE, Alipanah N, Barry PM, Brozek JL, Cattamanchi A, et al. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: treatment of drug-susceptible tuberculosis. Clin Infect Dis 2016 Oct 01;63(7):e147-e195 [FREE Full text] [doi: 10.1093/cid/ciw376] [Medline: 27516382]
11. World Health Organization (WHO). Guidelines for treatment of drug-susceptible tuberculosis and patient care (2017 update). WHO. Geneva, Switzerland: WHO; 2017 Jun 30. URL: http://www.who.int/tb/publications/2017/dstb_guidance_2017/en/ [accessed 2019-09-18]
12. World Health Organization (WHO). Adherence to Long-term Therapies Evidence for Action. Geneva, Switzerland: WHO; 2003:10-15.
13. Karumbi J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev 2015 May 29;2015(5):CD003343 [FREE Full text] [doi: 10.1002/14651858.CD003343.pub4] [Medline: 26022367]
14. Yang WT, Gounder CR, Akande T, De Neve JW, McIntire KN, Chandrasekhar A, et al. Barriers and delays in tuberculosis diagnosis and treatment services: does gender matter? Tuberc Res Treat 2014;2014:461935 [FREE Full text] [doi: 10.1155/2014/461935] [Medline: 24876956]
15. Holzman SB, Atre S, Sahasrabudhe T, Ambike S, Jagtap D, Sayyad Y, et al. Use of smartphone-based video directly observed therapy (vDOT) in tuberculosis care: single-arm, prospective feasibility study. JMIR Form Res 2019 Aug 27;3(3):e13411. [doi: 10.2196/13411] [Medline: 31456581]
16. Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar 23;393(10177):1216-1224 [FREE Full text] [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30799062]
17. Cook R, Lamont T, Martin R, NIHR Dissemination Centre. Smartphones can improve adherence rates for TB treatment. BMJ 2019 Aug 29;366:l4920. [doi: 10.1136/bmj.l4920] [Medline: 31466942]
18. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
19. Garfein R, Collins K, Muñoz F, Moser K, Cerecer-Callu P, Raab F, et al. Feasibility of tuberculosis treatment monitoring by video directly observed therapy: a binational pilot study. Int J Tuberc Lung Dis 2015 Sep;19(9):1057-1064 [FREE Full text] [doi: 10.5588/ijtld.14.0923] [Medline: 26260824]
20. Krueger K, Ruby D, Cooley P, Montoya B, Exarchos A, Djojonegoro BM, et al. Videophone utilization as an alternative to directly observed therapy for tuberculosis. Int J Tuberc Lung Dis 2010 Jun;14(6):779-781. [Medline: 20487619]
21. Mirsaeidi M, Farshidpour M, Banks-Tripp D, Hashmi S, Kujoth C, Schraufnagel D. Video directly observed therapy for treatment of tuberculosis is patient-oriented and cost-effective. Eur Respir J 2015 Sep;46(3):871-874. [doi: 10.1183/09031936.00011015] [Medline: 25792632]
22. DeMaio J, Schwartz L, Cooley P, Tice A. The application of telemedicine technology to a directly observed therapy program for tuberculosis: a pilot project. Clin Infect Dis 2001 Dec 15;33(12):2082-2084. [doi: 10.1086/324506] [Medline: 11698993]
23. Gassanov M, Feldman LJ, Sebastian A, Kraguljac MJ, Rea E, Yaffe B. The use of videophone for directly observed therapy for the treatment of tuberculosis. Can J Public Health 2013 May 14;104(3):e272 [FREE Full text] [doi: 10.17269/cjph.104.3869] [Medline: 23823897]

XSL•FO
RenderX

24. Hoffman J, Cunningham JR, Suleh AJ, Sundsmo A, Dekker D, Vago F, et al. Mobile direct observation treatment for tuberculosis patients: a technical feasibility pilot using mobile phones in Nairobi, Kenya. Am J Prev Med 2010 Jul;39(1):78-80. [doi: 10.1016/j.amepre.2010.02.018] [Medline: 20537846]
25. Sekandi J, Buregyeya E, Zalwango S, Dobbin KK, Atuyambe L, Nakkonde D, et al. Video directly observed therapy for supporting and monitoring adherence to tuberculosis treatment in Uganda: a pilot cohort study. ERJ Open Res 2020 Jan;6(1):00175-2019 [FREE Full text] [doi: 10.1183/23120541.00175-2019] [Medline: 32280670]
26. Nguyen T, Pham MT, Nguyen TL, Nguyen VN, Pham DC, Nguyen BH, et al. Video directly observed therapy to support adherence with treatment for tuberculosis in Vietnam: a prospective cohort study. Int J Infect Dis 2017 Dec;65:85-89 [FREE Full text] [doi: 10.1016/j.ijid.2017.09.029] [Medline: 29030137]
27. Hayward AG, Garber E. TB Reach 5: to compare the efficacy of video observed treatment (VOT) versus directly observed treatment (DOT) in supporting adherence in patients with active tuberculosis. ISRCTN. 2014 Apr 17. URL: https://www.isrctn.com/ISRCTN26184967 [accessed 2025-01-05]
28. Ravenscroft L, Kettle S, Persian R, Ruda S, Severin L, Doltu S, et al. Video-observed therapy and medication adherence for tuberculosis patients: randomised controlled trial in Moldova. Eur Respir J 2020 Aug;56(2):2000493 [FREE Full text] [doi: 10.1183/13993003.00493-2020] [Medline: 32381495]
29. Burzynski J, Mangan JM, Lam CK, Macaraig M, Salerno MM, deCastro BR, eDOT Study Team. In-person vs electronic directly observed therapy for tuberculosis treatment adherence: a randomized noninferiority trial. JAMA Netw Open 2022 Jan 04;5(1):e2144210 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.44210] [Medline: 35050357]
30. Sekandi J, McDonald A, Nakkonde D, Zalwango S, Kasiita V, Kaggwa P, et al. Acceptability, usefulness, and ease of use of an enhanced video directly observed treatment system for supporting patients with tuberculosis in Kampala, Uganda: explanatory qualitative study. JMIR Form Res 2023 Nov 10;7:e46203 [FREE Full text] [doi: 10.2196/46203] [Medline: 37948121]
31. Sekandi J, Kasiita V, Onuoha NA, Zalwango S, Nakkonde D, Kaawa-Mafigiri D, et al. Stakeholders' perceptions of benefits of and barriers to using video-observed treatment for monitoring patients with tuberculosis in Uganda: exploratory qualitative study. JMIR Mhealth Uhealth 2021 Oct 27;9(10):e27131. [doi: 10.2196/27131] [Medline: 34704961]
32. Truong C, Tanni KA, Qian J. Video-observed therapy versus directly observed therapy in patients with tuberculosis. Am J Prev Med 2022 Mar;62(3):450-458. [doi: 10.1016/j.amepre.2021.10.013] [Medline: 34916094]
33. Sekandi J, Onuoha NA, Buregyeya E, Zalwango S, Kaggwa PE, Nakkonde D, et al. Using a mobile health intervention (DOT Selfie) with transfer of social bundle incentives to increase treatment adherence in tuberculosis patients in Uganda: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Jan 05;10(1):e18029 [FREE Full text] [doi: 10.2196/18029] [Medline: 32990629]
34. Friedman LM, Furberg CD, DeMets DL, Reboussin DM. Fundamentals of Clinical Trials. Cham, Switzerland: Springer; 2015.
35. Feng J, Huang SF, Ting WY, Chen YC, Lin YY, Huang RM, et al. Gender differences in treatment outcomes of tuberculosis patients in Taiwan: a prospective observational study. Clin Microbiol Infect 2012 Sep;18(9):E331-E337 [FREE Full text] [doi: 10.1111/j.1469-0691.2012.03931.x] [Medline: 22734962]
36. SureAdhere Mobile Technology, Inc. URL: http://www.sureadhere.com [accessed 2020-04-20]
37. Manual for management and control of tuberculosis and leprosy (3rd edition). Ministry of Health, Republic of Uganda. 2017 Mar. URL: https://www.health.go.ug/cause/manual-for-management-and-control-of-tuberculosis-and-leprosy-3rd-edition-march-2017/ [accessed 2024-03-10]
38. World Health Organization (WHO). Definitions and reporting framework for tuberculosis. WHO. Geneva, Switzerland: WHO; 2013. URL: http://www.who.int/tb/publications/definitions/en/ [accessed 2023-01-30]
39. Hassard S, Ronald A, Angella K. Patient attitudes towards community-based tuberculosis DOT and adherence to treatment in an urban setting; Kampala, Uganda. Pan Afr Med J 2017;27:1 [FREE Full text] [doi: 10.11604/pamj.2017.27.1.11119] [Medline: 28748003]
40. Yoeli E, Rathauser J, Bhanot SP, Kimenye MK, Mailu E, Masini E, et al. Digital health support in treatment for tuberculosis. N Engl J Med 2019 Sep 05;381(10):986-987. [doi: 10.1056/NEJMc1806550] [Medline: 31483974]
41. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
42. Kisambu J, Nuwaha F, Sekandi JN. Adherence to treatment and supervision for tuberculosis in a DOTS programme among pastoralists in Uganda. Int J Tuberc Lung Dis 2014 Jul;18(7):799-803. [doi: 10.5588/ijtld.13.0753] [Medline: 24902555]
43. Howell E, Kigozi NG, Heunis JC. Community-based directly observed treatment for TB patients to improve HIV services: a cross-sectional study in a South African province. BMC Health Serv Res 2018 Apr 07;18(1):255. [doi: 10.1186/s12913-018-3074-1] [Medline: 29625569]
44. mHealth tools for individuals with chronic conditions to promote effective patient-provider communication, adherence to treatment and self-management Internet. National Institutes of Health. 2016. URL: https://grants.nih.gov/grants/guide/pa-files/pa-18-386.html [accessed 2020-09-20]

XSL•FO
RenderX

45. Falzon D, Timimi H, Kurosinski P, Migliori GB, Van Gemert W, Denkinger C, et al. Digital health for the End TB Strategy: developing priority products and making them work. Eur Respir J 2016 Jul;48(1):29-45 [FREE Full text] [doi: 10.1183/13993003.00424-2016] [Medline: 27230443]
46. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
47. Mangan J, Burzynski J, deCastro BR, Salerno MM, Lam CK, Macaraig M, et al. Challenges associated with electronic and in-person directly observed therapy during a randomized trial. Int J Tuberc Lung Dis 2023 Apr 01;27(4):298-307 [FREE Full text] [doi: 10.5588/ijtld.22.0583] [Medline: 37035970]
48. Vernon A, Fielding K, Savic R, Dodd L, Nahid P. The importance of adherence in tuberculosis treatment clinical trials and its relevance in explanatory and pragmatic trials. PLoS Med 2019 Dec;16(12):e1002884 [FREE Full text] [doi: 10.1371/journal.pmed.1002884] [Medline: 31821323]
49. Kara G, Yalcin BM. Comparison of in-person vs. video directly observed therapy (VDOT) on stigma levels in tuberculosis patients. J Am Board Fam Med 2022 Aug 25:jabfm.2022.AP.210514. [doi: 10.3122/jabfm.2022.AP.210514] [Medline: 36007959]
50. Lippincott CK, Perry A, Munk E, Maltas G, Shah M. Tuberculosis treatment adherence in the era of COVID-19. BMC Infect Dis 2022 Oct 26;22(1):800 [FREE Full text] [doi: 10.1186/s12879-022-07787-4] [Medline: 36289485]
51. DiStefano M, Schmidt H. mHealth for tuberculosis treatment adherence: a framework to guide ethical planning, implementation, and evaluation. Glob Health Sci Pract 2016 Jun 20;4(2):211-221 [FREE Full text] [doi: 10.9745/GHSP-D-16-00018] [Medline: 27353615]
52. Sekandi J, Murray K, Berryman C, Davis-Olwell P, Hurst C, Kakaire R, et al. Ethical, legal, and sociocultural issues in the use of mobile technologies and call detail records data for public health in the East African region: scoping review. Interact J Med Res 2022 Jun 02;11(1):e35062 [FREE Full text] [doi: 10.2196/35062] [Medline: 35533323]
53. O'Brien N, Li E, Chaibva CN, Gomez Bravo R, Kovacevic L, Kwame Ayisi-Boateng N, et al. Strengths, weaknesses, opportunities, and threats analysis of the use of digital health technologies in primary health care in the Sub-Saharan African region: qualitative study. J Med Internet Res 2023 Sep 07;25:e45224 [FREE Full text] [doi: 10.2196/45224] [Medline: 37676721]

Abbreviations

DOT: directly observed therapy
FEDO: fraction of expected doses observed
GSM: Global System for Mobile Communications
HIPAA: Health Insurance Portability and Accountability Act
MDR-TB: multidrug-resistant tuberculosis
RCT: randomized clinical trial
TB: tuberculosis
UCDOT: usual care directly observed therapy
VDOT: video directly observed therapy
WHO: World Health Organization

Edited by L Buis; submitted 01.03.24; peer-reviewed by R Garfein, B Thies; comments to author 17.05.24; revised version received 18.09.24; accepted 18.12.24; published 16.01.25.

Please cite as:
Sekandi JN, Buregyeya E, Zalwango S, Nakkonde D, Kaggwa P, Quach THT, Asiimwe D, Atuyambe L, Dobbin K
Effectiveness of a Mobile Health Intervention (DOT Selfie) in Increasing Treatment Adherence Monitoring and Support for Patients With Tuberculosis in Uganda: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e57991
URL: https://mhealth.jmir.org/2025/1/e57991
doi:10.2196/57991
PMID:39715573

©Juliet Nabbuye Sekandi, Esther Buregyeya, Sarah Zalwango, Damalie Nakkonde, Patrick Kaggwa, Trang Ho Thu Quach, David Asiimwe, Lynn Atuyambe, Kevin Dobbin. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of the Better Operative Software Outcomes Tool for Monitoring Cataract Surgery Outcomes in Inhambane Province, Mozambique: Prospective Pilot Diagnostic Accuracy Study

Mónica Lecumberri[1,2,3], MD; Dórcia Mandlate[4], MD; Vasco Cote[5], MD; Montserrat Martín-Baranera[6], MD, PhD; Carlos L Moser[1,2], MD

[1]Department of Ophthalmology, Consorci Sanitari Integral, L'Hospitalet de Llobregat, Barcelona, Spain

[2]Eyes of the World Foundation, Barcelona, Spain

[3]Department of Surgery, Universitat Autònoma de Barcelona, Barcelona, Spain

[4]Coordinator of Research Department, Direcção Provincial de Saúde, Inhambane, Mozambique

[5]Eyes of the World Foundation, Inhambane, Mozambique

[6]Department of Clinical Epidemiology, Consorci Sanitari Integral, L´Hospitalet de Llobregat, Barcelona, Spain

Corresponding Author:
Mónica Lecumberri, MD
Department of Ophthalmology
Consorci Sanitari Integral
Av. Josep Molins, 29
L'Hospitalet de Llobregat, Barcelona, 08906
Spain
Phone: 34 934 40 75 00
Email: Molecumberri@hotmail.com

Abstract

Background: Cataract is the world's leading cause of avoidable blindness. High-quality cataract surgery is a cost-effective procedure to restore vision. In low- and middle-income countries (LMIC), there are high rates of poor outcomes after surgery and inadequate follow-up, which makes it difficult to effectively monitor surgical outcomes. To address this issue, the Better Operative Outcomes Software Tool (BOOST Cataract app) assesses the final outcome on the first postoperative day and records the reasons for poor outcomes at 6 weeks postoperatively, offering specific advice to improve outcomes.

Objective: The aim of this study is to evaluate the real-world utility of the BOOST Cataract app for monitoring cataract surgical outcomes in Inhambane Province (Mozambique) and to analyze the data collected by the app during a surgical campaign conducted therein.

Methods: This prospective study of diagnostic accuracy included patients older than 50 years who underwent manual small-incision cataract surgery. The BOOST Cataract app was used to collect data on visual acuity (VA) on the first postoperative day and to assess surgical outcomes. The sensitivity and specificity (with their 95% CIs) and area under the curve of the BOOST Cataract app (index test) were calculated to identify suboptimal outcomes (moderate and poor) compared with outcomes 6 weeks after surgery (reference standard). The causes of poor outcomes after 6 weeks were recorded and evaluated using the app.

Results: A total of 141 patients who underwent surgery during a cataract campaign in Inhambane Province (Mozambique) between April 2022 and May 2022 were included in the study. The mean age was 70 (SD 10.2) years, and 48.2% (68/141) of the participants were women. Of the 141 patients, 8 (5.7%) did not complete the study. The BOOST Cataract app had a sensitivity of 94.44% and a specificity of 59% for detecting suboptimal outcomes. The area under the curve was 0.825. The campaign outcomes were as follows: good (VA ≥6/18) in 45.9% (61/133) of cases, moderate (VA <6/18 to ≥6/60) in 33.8% (45/133), and poor (VA<6/60) in 20.3% (27/133) of cases. The main cause of a poor outcome was surgical complication, while refractive error was the main cause of a moderate outcome.

Conclusions: The BOOST Cataract app is a valuable tool for assessing suboptimal cataract surgery outcomes without waiting for the completion of postoperative treatment in settings where postoperative follow-up is incomplete. In resource-limited eye care systems, the BOOST Cataract app provides an opportunity to assess outcomes and analyze results to develop specific strategies to improve quality at a local level.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e64500) doi:10.2196/64500

KEYWORDS

cataract extraction; outcome assessment; health care; developing countries; visual acuity; software

Introduction

In 2020, cataracts remained the leading cause of avoidable blindness in people aged >50 years, followed by glaucoma and refractive error [1]. Although cataract surgery is a highly cost-effective procedure for restoring vision, only 60% of patients who undergo cataract surgery achieve visual acuity (VA) of 6/18 or better, whereas up to 18% fail to achieve a VA better than 6/60 [2,3]. Surgical outcomes are worse in low-income countries, with an estimated mean effective cataract surgical coverage of 13.9% in the African region [4].

In low- and middle-income countries (LMIC), the low rate of postoperative follow-up makes it difficult to obtain reliable surgical outcome data [5]. The first step toward improving surgical outcomes is to routinely monitor the extent and causes of poor outcomes [6]. In a prospective study conducted at Kikuyu Hospital in Kenya, prospective monitoring improved surgical outcomes [7].

The only known data on cataract surgery outcomes in Inhambane Province came from the Rapid Assessment of Avoidable Blindness (RAAB) study, which was conducted between October 2016 and November 2016. The RAAB study showed that cataract blindness and visual impairment were not controlled in the province, with only 29% of those with blindness (VA <3/60) and 20% of patients with severe visual impairment (VA <6/60) undergoing surgery. Furthermore, only 42.9% of the operated eyes had a good outcome (VA ≥6/18), and 37.5% had a poor outcome (VA <6/60) [8].

Efforts have been made to objectively evaluate surgical outcomes, particularly for residents [9]. Access to the internet in low-resource settings has prompted its use, and it has been suggested that tools such as the Better Operative Software Outcomes Tool (BOOST Cataract app) may facilitate data collection in countries with low postoperative follow-up rates and help improve surgical quality [2].

The BOOST Cataract app was launched in 2018 by a global consortium of leading eye health organizations (the Fred Hollows Foundation, Sightsavers, The International Agency for the Prevention of Blindness, ORBIS International, Aravind Eye Hospitals, and the International Council of Ophthalmology) to address the issues of surgical quality measurement and surgical data recording and to help improve cataract surgical outcomes in low-resource settings.

The BOOST Cataract app is a free and easy-to-use app that can be used for recording and analyzing surgical outcomes the day after surgery. The app does not collect personal patient data but assigns a number to each patient. There are 2 phases in the assessment process. In phase 1, postoperative uncorrected distance VA (BOOST VA) is recorded between 1 day and 3 days after surgery. The BOOST Cataract app estimates the final surgical outcome according to the World Health Organization (WHO) reference criteria, which defines a good outcome as VA ≥6/18 [10] (Table 1). The outcome estimation is based on the direct correlation between VA on postoperative day 1 and the final VA [5,11]. This overcomes the lack of postoperative follow-up, which makes outcome evaluation difficult.

Table 1. World Health Organization (WHO) cataract surgery outcome criteria according to postoperative visual acuity.

Snellen PVA[a]	Definition of distance visual impairment (ICD-11[b])	Standards for postoperative visual acuity according to WHO[c]
Good		
≥6/12	Not impaired	>80%[d]
<6/12 to ≥6/18	Mild visual impairment	>80%[d]
Suboptimal		
Moderate		
<6/18 to ≥6/60	Moderate visual impairment	<15%
Poor		
<6/60 to ≥3/60	Severe visual impairment	<5%[e]
<3/60	Blindness	<5%[e]

[a]PVA: presenting visual acuity.

[b] *ICD-11: International Classification of Diseases, 11th Revision.*

[c]WHO: World Health Organization.

[d]Assigned to only one of the “Not impaired” or “Mild visual impairment” categories, not both at once.

[e]Assigned to only one of the “Severe visual impairment” or “Blindness” categories, not both at once.

Phase 2 consists of analyzing the possible causes of poor outcomes at 6 weeks postoperatively. For this purpose, the presenting visual acuity (PVA) and best-corrected distance visual acuity (BCVA) of 20 patients with poor outcomes at least

XSL•FO
RenderX

6 weeks after surgery are recorded, and the possible etiologies (selection, complication, and refractive error) are assessed. In very low-resource settings where refractive services are unavailable, VA obtained with pinhole (Pin VA) is accepted as an approximation of BCVA. Based on these data, the tool can then identify the main cause of poor outcomes and suggests specific interventions aimed at improvements [12]. The BOOST Cataract app has been used in several countries to monitor the quality of cataract surgery [13]. A study evaluating the app's effectiveness at predicting the final surgical outcome showed that the ability of the BOOST Cataract app to correctly categorize surgical outcome was dependent upon the presence of corneal edema on postoperative day 1. However, this study was retrospective and not based on an LMIC population [14].

This study aimed to evaluate the utility of the BOOST Cataract app (software version 1.1.2) in a surgical campaign conducted in Inhambane Province and to perform an in-depth analysis of the outcomes and causes of suboptimal results.

Methods

Participants and Settings

Consecutive patients presenting with cataracts, who were aged >50 years, and who attended the Provincial Hospital of Inhambane and the District Hospital of Zavala, Mozambique, between April 2022 and May 2022 were included. Patients with cataract were recruited using convenience sampling according to the hospital's inclusion protocol for cataract surgery: patients aged >50 years with VA ≤6/60 in 1 eye and NC5 or NC6 cataract. The exclusion criteria were the absence of a pupillary light reflex and retinal pathology demonstrated by fundoscopy (NC5 cataract) or ultrasound (NC6 cataract), corneal scars affecting the visual axis, and advanced glaucoma demonstrated by optic nerve examination when it was possible to obtain.

Ethics Approval

Informed written consent was obtained from all patients who underwent surgery. All data were anonymized. As the study was carried out in the setting of standard clinical practice, the patients did not receive any financial compensation. This study adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Bioethics Committee for Health of the Republic of Mozambique (CNBS; reference: 443/CNBS/23).

Preoperative Examinations

All patients underwent a complete ophthalmological examination, including PVA, slit-lamp examination, pupillary reflexes, Goldmann applanation tonometry, and fundus examination, if lens opacity allowed. The power of the implanted intraocular lens (IOL) was calculated using a keratometer (LGM OftalTech) and mode A ultrasound (Scan sw-1000; Suoer). Patients for whom keratometry or biometry could not be performed were assigned to the IOL of the contralateral eye, if possible, or a standard 21 D IOL, if not.

VA Assessment

PVA was assessed using Tumbling E charts for distance vision, with available correction. After correctly identifying the direction of more than one-half of the optotypes in the upper line, the patient moved to the next line then to the lower lines. The lowest line at which more than one-half of the optotypes were read was recorded as the patient's VA. The patient moved one-half the distance when the direction of any line could not be identified. If the patient was unable to identify the optotypes at that distance, the assessor's fingers were placed 3 meters, 2 meters, or 1 meter from the patient. If the patient was unable to see their fingers, a light source was placed to document light perception.

Surgery and Postoperative Treatment

All surgeries were performed under retrobulbar anesthesia by the same ophthalmologist with extensive experience in manual small-incision cataract surgery (MSICS). MSICS was performed using a superior scleral tunnel. A rigid PMMA biconvex posterior chamber IOL (S3550 Aurolens; Aurolab Inc) was implanted with capsular support. Eyes with zonulolysis or extensive capsular rupture were implanted with an A5520 anterior chamber lens (Aurolab Inc). At the end of surgery, all patients received perioperative prophylaxis with intracameral 0.1 mL cefazolin (2.5 mg/mL) or 0.1 mL vancomycin (0.1 mg/mL) if they were allergic to penicillin [15].

All patients followed the same postoperative treatment protocol at the hospital, consisting of tapering doses of MFLOX-DM (moxifloxacin 0.5% + dexamethasone sodium phosphate 0.1%; Aurolab Inc) for 4 weeks. Eyes with intraocular pressure higher than 25 mm Hg received acetazolamide (250 mg Diamox).

Assessment the Day After Surgery and Phase 1 of BOOST Cataract App

In phase 1, patients were examined the day after surgery, and VA was measured in the operated eye by an independent observer (BOOST VA). The BOOST VA was entered into the app, and the surgical outcomes were calculated according to the WHO criteria (Table 1).

Slit-lamp examination was also performed to detect corneal edema. Significant corneal edema was defined as >10 Descemet folds or stromal corneal edema affecting the visual axis.

Assessment After 6 Weeks and Phase 2 of the BOOST Cataract App

All patients were instructed to return for follow-up 6 weeks after surgery and received a telephone call 2 days before the visit as a reminder to ensure complete follow-up after surgery.

All patients who completed follow-up at 6 weeks had their uncorrected VA (PVA), BCVA, and PinVA (pinhole visual acuity) assessed by an optometrist masked to the BOOST VA. A biomicroscopic examination and autorefraction analysis were performed (LGM OftalTech).

The VAs obtained at 6 weeks were classified according to the WHO criteria (Table 1): without optical correction (PVA outcome) and with pinhole (PinVA outcome).

In phase 2 of the BOOST Cataract app, the cause of poor outcomes (VA <6/60) was identified and recorded as selection, complications, or refractive error. The same was done for patients with a moderate outcome (VA <6/18 to ≥6/60).

XSL•FO
RenderX

Statistical Analyses

Statistical analyses of the data were performed using SPSS Statistics v21 (IBM Corp). Absolute and relative frequencies were obtained for qualitative variables. For the quantitative variables, depending on whether the normal distribution criteria were met, the mean and SD, median, IQR, and minimum and maximum values were considered.

Regarding the different VAs collected, an equivalent in decimal notation was assigned to the VA that could not be expressed numerically (light perception, counting fingers) [16,17]. The decimal values were converted to a LogMAR (minimum angle-of-resolution) scale for subsequent statistical analysis.

To analyze the ability of the BOOST Cataract app to predict the final surgical outcome, the results obtained on postoperative day 1 using the app (BOOST outcome) were compared with those obtained 6 weeks postoperatively (PVA outcome). For this analysis, only cases with complete follow-up were included. Both variables were grouped into 2 categories (good and moderate/poor results). Using the 6-week results as the reference standard, the sensitivity, specificity, positive predictive value, and negative predictive value of the BOOST Cataract app as a tool for predicting a negative result (moderate or poor) were estimated. All estimates are presented with the corresponding 95% CIs. A receiver operating characteristic curve was generated, and the area under the curve was calculated.

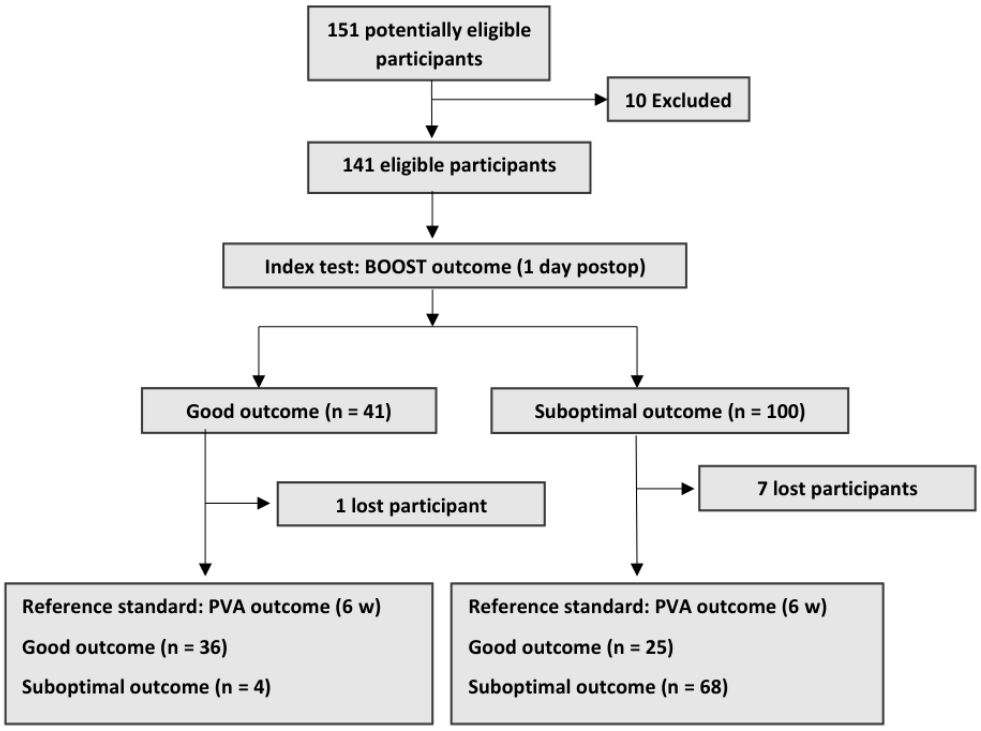
The presence of corneal edema on day 1 was tested as an independent predictor of moderate or poor outcomes at day 1 and at 6 weeks after surgery using a logistic regression model.

Results

Characteristics of the Campaign

We recruited 151 patients, and 10 were excluded because they did not meet the inclusion criteria for surgery. A study flowchart is presented in Figure 1.

Figure 1. Flow diagram of participants. BOOST: Better Operative Software Outcomes Tool; PVA: presenting visual acuity.



The study included 141 eyes of 141 patients who underwent surgery at the study site: 68 patients (68/141, 48%) were women, and 62 patients (62/141, 44%) were blind (VA <3/60 in the better eye; Table 2). This was the first eye operated on in 108 of 141 (76.6%) patients. In addition, 14 of 141 (9.7%) patients had a history of glaucoma, and 3 of 141 (2.1%) eyes had phacodonesis on biomicroscopic examination. In 123 of the 141 (87.2%) operated eyes, it was not possible to perform a fundus examination because of cataract density. Biometry could not be performed in 22 patients because of corneal irregularities or lack of fixation.

XSL•FO
RenderX

Table 2. Preoperative data.

Characteristic	Total sample (n=141)	Women (n=68)	Men (n=73)
Age (years), median (SD)	70 (10.2)	70.28 (9.7)	69.77 (10.6)
Preoperative Snellen visual acuity, n (%)			
<3/60	86 (61)	42 (618)	44 (60.3)
≥3/60 to <6/60	52 (36.9)	24 (35.3)	28 (38.4)
≥6/60	3 (2.1)	2 (2.9)	1 (1.4)
Contralateral eye status, n (%)			
Aphakia	1 (0.7)	0 (0)	1 (100)
Phakic eye	108 (77.1)	51 (47.2)	57 (52.8)
Pseudophakia	31 (22.1)	17 (54.8)	14 (45.2)
Visual status, n (%)			
Blindness[a]	62 (44.6)	27 (43.5)	35 (56.5)
Severe visual impairment[b]	13 (9.4)	8 (61.5)	5 (38.5)
Moderate visual impairment[c]	12 (8.6)	8 (66.7)	4 (33.3)
Mild visual impairment[d]	32 (23)	18 (56.3)	14 (43.8)
Not impaired[e]	20 (14.4)	7 (35)	13 (65)

[a]<3/60 in the better eye.

[b]≥3/60 to <6/60 in the better eye.

[c]<6/18 to ≥6/60 in the better eye.

[d]<6/12 to ≥6/18.in the better eye.

[e]≥6/12 in the better eye.

Outcomes of the Campaign: Phase 1 of BOOST Cataract App

Of the 14 intraoperative complications observed, 11 (7.8% of the 141 surgeries) were capsular ruptures. In addition, 10 (7.1%) anterior chamber IOL were implanted, and 1 case of aphakia was reported.

The outcomes of the BOOST Cataract app (BOOST outcomes) are presented in Table 3. Overall, 29% (41/141) of the operated eyes showed good outcomes. Of the 141 eyes, 74 (52.2%) had corneal edema on postoperative day 1. Corneal edema was associated with suboptimal (moderate and poor) BOOST outcomes (ρ<0.001).

The surgical outcome on day 1 (BOOST outcome) was worse than at 6 weeks. Of the 141 patients, 8 (5.7%) did not attend the 6-week postoperative visit. Of these, 6 underwent uneventful surgery. Of the patients who did not complete the follow-up period, 50% (4/8) had poor BOOST outcomes.

The final outcomes (PVA outcome) were good for 45.9% (61/133) of the eyes and remained poor for 20% (27/133) of the eyes. The PinVA outcomes were good for 73% (94/128) of the eyes and poor for 13% (17/128) of the eyes (Table 3).

Eyes that had a suboptimal outcome with the BOOST Cataract app were 21.44 (95% CI 6.35-72.35) times more likely to have a moderate or poor outcome at 6 weeks postoperatively compared with the eyes with a good outcome (ρ<0.001). Corneal edema was associated with the BOOST outcome (ρ<0.001) but not with the final PVA outcome (ρ=0.58).

XSL•FO
RenderX

Table 3. Surgical campaign outcomes.

Outcomes	Total sample	Women	Men
Postoperative day 1			
BOOST[a] outcome (n=141), n (%)			
Good[b]	41 (29.1)	—[c]	—
Suboptimal	100 (70.9)	—	—
Suboptimal: moderate[d]	53 (37.6)	—	—
Suboptimal: poor[e]	47 (33.3)	—	—
6 weeks postoperative			
PVA[f] outcome (n=133), n (%)			
Good	61 (45.9)	30 (46.2)	31 (45.6)
Suboptimal	72 (54.1)	35 (53.8)	37 (54.4)
Suboptimal: moderate	45 (33.8)	22 (33.8)	23 (33.8)
Suboptimal: poor	27 (20.3)	13 (20)	14 (20.6)
PinVA[g] outcome (n=128), n (%)			
Good	94 (73.4)	43 (69.4)	59 (77.3)
Suboptimal	34 (26.6)	19 (30.6)	15 (22.7)
Suboptimal: moderate	17 (13.3)	11 (17.7)	6 (13.6)
Suboptimal: poor	17 (13.3)	8 (12.9)	9 (9.1)
Postoperative spherical equivalent (D[h]), median (range)	–1.38 (–9.62 to +1.25)	—	—
Postoperative astigmatism (D), median (range)	–3 (–0.5 to –8.75)	—	—

[a]BOOST: Better Operative Outcomes Software Tool.
[b]Visual acuity (VA) ≥6/18.
[c]Not applicable.
[d]VA: <6/18 to ≥6/60.
[e]VA: <6/60.
[f]PVA: presenting visual acuity.
[g]PinVA: visual acuity with pinhole.
[h]D: diopters.

Diagnostic Accuracy of the BOOST Cataract App

The sensitivity of the BOOST Cataract app for identifying suboptimal outcomes was 94%. The specificity was 59%. The area under the curve was 0.825 (Table 4).

Table 4. Diagnostic accuracy of the Better Operative Outcomes Software Tool (BOOST Cataract app).

Performance measures	Results, % (95% CI)
Sensitivity	94.44 (86.38-98.47)
Specificity	59.02 (45.68-71.45)
Positive predictive value	73.12 (66.69-78.70)
Negative predictive value	90.00 (77.24-95.98)

Reasons for Suboptimal Outcomes: Phase 2 of the BOOST Cataract App

Surgical complications (capsular rupture and zonulolysis with vitreous loss) were the main causes of poor final outcomes, while refractive error was the main cause of a moderate outcome (Table 5).

XSL•FO
RenderX

Table 5. Causes of a suboptimal outcome according to phase 2 of the Better Operative Outcomes Software Tool (BOOST Cataract app).

Outcomes	Results, n (%)
Poor outcome[a] (n=27)	
Selection	7 (26)
Refraction	8 (30)
Complication	12 (44)
Moderate outcome[b] (n=45)	
Selection	4 (9)
Refraction	36 (80)
Complication	5 (11)

[a]Visual acuity (VA) <6/60.

[b]VA <6/18 to ≥6/60.

Discussion

Principal Findings

The BOOST Cataract app was able to detect final suboptimal outcomes (VA <6/18) on the day after surgery with a sensitivity of 94%. Due to this sensitivity, it was possible, with reasonable confidence, to identify eyes that were moderately visually impaired or worse at the end of postoperative treatment.

Resource-poor LMIC are often characterized by a lack of equipment, high staff turnover, and frequent surgical campaigns where many surgeries are performed in short periods. In this context, the routine use of the BOOST Cataract app ensures that any decrease in quality is detected early in phase 1.

The specificity of the campaign in this study was only 59%. This low specificity did not provide sufficient certainty estimating the extent of good outcomes. Therefore, the BOOST Cataract app did not determine the center's local effective cataract surgical coverage. This is because it underestimated good outcomes in phase 1 by not considering transient corneal edema that occurs in the immediate postoperative period and usually resolves without sequelae.

At 6 weeks postoperatively, 45.9% (61/133) of the operated eyes had good outcomes, and just over 20% (27/133, 20.3%) did not achieve vision better than 6/60; with pinhole, 73.4% (94/128) had a good outcome, and 13.3% (17/128) had a poor outcome. These results do not meet the quality standards recommended by the WHO (Table 1).

In the 2016 RAAB study conducted in the province, 42.9% of patients aged >50 years who underwent cataract surgery had good outcomes, and as many as 37.5% had poor outcomes; among patients assessed using PinVA, only 55.4% had good outcomes, and 34.8% had poor outcomes [8].

The 2016 RAAB study collected information on surgeries performed before 2016 when the sutureless extracapsular technique was still being introduced. MSICS is now fully established, and surgical outcomes seem to have improved. However, the RAAB study showed outcomes over time, whereas our hospital-based study only observed outcomes at 6 weeks and may not have captured long-term surgical outcomes and complications. A new RAAB study in the province is required to demonstrate the improvement in surgical outcomes.

In phase 2 of this study, surgical complications were the main cause of poor outcomes (12/27, 44%). In Inhambane Province, as in other provinces of Mozambique, cataract surgery is performed in patients with a preoperative VA ≤6/60 [18]. It is common for cataracts to present at very advanced stages with a higher rate of complications (zonulolysis and capsular rupture), despite the use of MSICS, which has been shown to be a safe technique in LMIC [19-21].

These outcomes are comparable to those reported in other areas of the country. The RAAB study performed in Sofala Province reported good outcomes in 37.8% of the patients, which improved to 45.9% when considering PinVA [18]. The main causes of poor outcomes (VA <6/60) were selection errors (55.2%) and surgical complications (34.5%). The 2018 RAAB study conducted in the province of Nampula recorded 58.2% good outcomes and 22.4% poor outcomes in terms of PVA. The main causes of outcomes worse than 6/18 vision were a lack of optical correction (34%) and surgical complications (29%). The PinVA outcomes improved to 72.4% (good) and 16.3% (poor) [22].

The BOOST Cataract app does not collect data on moderate outcomes or their causes in phase 2. This is despite these outcomes representing a postoperative VA of <6/18 to ≥6/60, equating to moderate visual impairment (Table 1).

In this study, 33.8% (45/133) of the patients had a moderate outcome, and the main cause was refractive error. The magnitude of this finding seems to indicate that, in Inhambane Province, as in other LMIC, awareness of the causes of moderate outcomes is necessary to improve them [18,23,24]. Visual impairment due to residual refractive error may be reduced if stockpiles of IOLs in a range of powers and ocular biometry are accessible [25]. In our study, the mean postoperative spherical equivalent was higher than that in other studies [26,27], even though biometry was performed in 84.4% (119/141) of the eyes and 100% of the patients had undergone IOL implantation in the range of power accorded by biometry. Cooperation of patients with poor fixation of the operated eye may lead to errors when performing biometry.

XSL•FO
RenderX

The mean postoperative astigmatism was also higher than that reported in other studies [28]. In fact, 75.2% (85/113) of the eyes had astigmatism greater than 3D against the rule. The use of a superior scleral tunnel instead of a temporal tunnel may have produced astigmatism against the rule with a worse final VA outcome [29]. Additionally, the surgical material is often in poor condition due to excessive resterilization, which may lead to more superficial sclerocorneal incisions than desired, increasing induced surgical astigmatism risk [30].

Strengths and Limitations

Although the sensitivity and specificity obtained allow reaching relevant conclusions on the real-world utility of the app at the study site, due to the small sample size, it is not possible to generalize the diagnostic accuracy of the BOOST Cataract app in other low-resource settings. In settings with a lower prevalence of suboptimal outcomes, the negative predictive value might be lower. Therefore, the app, if confirmed in larger studies, has to be restricted to settings with poor-resource health systems.

Data on surgical outcomes in Inhambane are scarce and only available from the RAAB study conducted 8 years ago. This was the first study to analyze the causes of poor and moderate outcomes in such depth.

Surgical quality criteria have recently been revised, and VA ≥6/12, representing mild visual impairment, has been adopted as the threshold for a good outcome [2]. Due to the introduction of the MSICS technique and widespread use of biometry, surgical outcomes have improved. This new threshold places higher demands on the expected outcomes and implies an improvement in the visual expectations of patients undergoing surgery in LMIC.

To analyze the diagnostic accuracy of the BOOST Cataract app, moderate and poor outcomes were grouped into the suboptimal outcome category. From a public health perspective, it is of interest to know the number of eyes that do not achieve a VA >6/60 after surgery, which is the level of severe visual impairment that has the greatest impact on patient quality of life because it affects their ability to walk independently. Grouping moderate and poor outcomes into suboptimal outcomes encourages improvements in both outcome categories without focusing only on poor outcomes.

Conclusions

RAAB studies provide reliable data for planning eye health services and evaluating programs. These are the primary sources of information for measuring surgical outcomes at the population level in LMIC [31]. However, in settings with resource-poor eye health systems, the BOOST Cataract app offers the opportunity to assess outcomes at the local level; each center needs to analyze its surgical outcomes and develop specific strategies to improve quality.

Acknowledgments

This study was conducted as part of the Doctoral Program in Surgery and Morphological Science at the Autonomous University of Barcelona and as part of the Eyes of the World Foundation's project in Mozambique, aimed at strengthening eye health services.

This work received funding from the Universitat Oberta de Catalunya (UOC) e-Health Center Project 2022.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request. All relevant data are included in this article.

Authors' Contributions

ML conceptualized the study, provided resources, curated the data, contributed to the methodology, conducted the formal analysis, prepared the original draft, and supervised the research. VC assisted with data curation and resource provision. DM contributed to conceptualization and methodology. MMB contributed to conceptualization, methodology, and formal analysis. CLM conducted the formal analysis and supervised the research. All authors participated in reviewing and editing the final manuscript. No generative artificial intelligence was used in any portion of the manuscript.

Conflicts of Interest

None declared.

References

1. GBD 2019 Blindness and Vision Impairment Collaborators, Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health 2021 Feb;9(2):e144-e160 [FREE Full text] [doi: 10.1016/S2214-109X(20)30489-7] [Medline: 33275949]
2. Burton MJ, Faal HB, Ramke J, Ravilla T, Holland P, Wang N, et al. Announcing The Lancet Global Health Commission on Global Eye Health. The Lancet Global Health 2019 Dec;7(12):e1612-e1613. [doi: 10.1016/s2214-109x(19)30450-4]
3. Taylor HR. Global blindness: the progress we are making and still need to make. Asia-Pacific Journal of Ophthalmology 2019 Nov;8(6):424-428. [doi: 10.1097/apo.0000000000000264]

XSL•FO
RenderX

4. McCormick I, Butcher R, Evans JR, Mactaggart IZ, Limburg H, Jolley E, et al. Effective cataract surgical coverage in adults aged 50 years and older: estimates from population-based surveys in 55 countries. The Lancet Global Health 2022 Dec;10(12):e1744-e1753. [doi: 10.1016/s2214-109x(22)00419-3]
5. Congdon N, Yan X, Lansingh V, Sisay A, Müller A, Chan V, et al. Assessment of cataract surgical outcomes in settings where follow-up is poor: PRECOG, a multicentre observational study. The Lancet Global Health 2013 Jul;1(1):e37-e45. [doi: 10.1016/s2214-109x(13)70003-2]
6. Limburg H, Foster A, Vaidyanathan K, Murthy GV. Monitoring visual outcome of cataract surgery in India. Bull World Health Organ 1999;77(6):455-460 [FREE Full text] [Medline: 10427929]
7. Yorston D, Gichuhi S, Wood M, Foster A. Does prospective monitoring improve cataract surgery outcomes in Africa? Br J Ophthalmol 2002 May;86(5):543-547 [FREE Full text] [doi: 10.1136/bjo.86.5.543] [Medline: 11973251]
8. Vilanculos A. Mozambique, Inhambane (2016). Rapid Assessment of Avoidable Blindness. URL: https://www.raab.world/survey/mozambique-inhambane-2016 [accessed 2025-05-02]
9. Al-Khaled T, Acaba-Berrocal L, Cole E, Ting DS, Chiang MF, Chan RP. Digital education in ophthalmology. Asia-Pacific Journal of Ophthalmology 2022 May;11(3):267-272. [doi: 10.1097/apo.0000000000000484]
10. Informal Consultation on Analysis of Blindness Prevention Outcomes, Geneva, 16-18 February 1998. World Health Organization. 1998. URL: https://iris.who.int/handle/10665/67843 [accessed 2025-05-02]
11. Briesen S, Ng EY, Roberts H. Validity of first post-operative day automated refraction following dense cataract extraction. Clin Exp Optom 2011 Mar 15;94(2):187-192 [FREE Full text] [doi: 10.1111/j.1444-0938.2010.00567.x] [Medline: 21255077]
12. Congdon N, Suburaman G, Ravilla T, Varga B, Resnikoff S, McLeod J, et al. Transforming research results into useful tools for global health: BOOST. The Lancet Global Health 2016 Feb;4(2):e96. [doi: 10.1016/s2214-109x(15)00267-3]
13. Fortané M, Bensaid P, Resnikoff S, Seini K, Landreau N, Paugam J, et al. Outcomes of cataract surgery performed by non-physician cataract surgeons in remote North Cameroon. Br J Ophthalmol 2019 Aug 05;103(8):1042-1047. [doi: 10.1136/bjophthalmol-2018-312428] [Medline: 30291137]
14. Lecumberri M, Moser CL, Loscos-Arenas J. Evaluation of the better operative outcome software tool to predict cataract surgical outcome in the early postoperative follow-up. BMC Ophthalmol 2023 Jul 13;23(1):317 [FREE Full text] [doi: 10.1186/s12886-023-03058-1] [Medline: 37442998]
15. Moser CL, Lecumberri Lopez M, Garat M, Martín-Baranera M. Prophylactic intracameral cefazolin and postoperative topical moxifloxacin after cataract surgery: endophthalmitis risk reduction and safety results in a 16-year study. Graefes Arch Clin Exp Ophthalmol 2019 Oct 7;257(10):2185-2191. [doi: 10.1007/s00417-019-04417-9] [Medline: 31392416]
16. Karanjia R, Hwang TJ, Chen AF, Pouw A, Tian JJ, Chu ER, et al. Correcting finger counting to Snellen Acuity. Neuroophthalmology 2016 Oct 22;40(5):219-221 [FREE Full text] [doi: 10.1080/01658107.2016.1209221] [Medline: 27928408]
17. Schulze-Bonsel K, Feltgen N, Burau H, Hansen L, Bach M. Visual acuities "hand motion" and "counting fingers" can be quantified with the freiburg visual acuity test. Invest Ophthalmol Vis Sci 2006 Mar 01;47(3):1236-1240. [doi: 10.1167/iovs.05-0981] [Medline: 16505064]
18. Roba A, Chagunda M, Machissa T. Blindness above and below the poverty line: Reflections form Sofala, Mozambique. J Public Health Afr 2020 Apr 29;11(1):1113 [FREE Full text] [doi: 10.4081/jphia.2020.1113] [Medline: 33209232]
19. Venkatesh R, Tan CSH, Singh GP, Veena K, Krishnan KT, Ravindran RD. Safety and efficacy of manual small incision cataract surgery for brunescent and black cataracts. Eye (Lond) 2009 May 20;23(5):1155-1157. [doi: 10.1038/eye.2008.190] [Medline: 18566610]
20. Venkatesh R, Tan CSH, Kumar TT, Ravindran RD. Safety and efficacy of manual small incision cataract surgery for phacolytic glaucoma. Br J Ophthalmol 2007 Mar 23;91(3):279-281 [FREE Full text] [doi: 10.1136/bjo.2006.105874] [Medline: 17322465]
21. Gogate P, Optom JJ, Deshpande S, Naidoo K. Meta-analysis to compare the safety and efficacy of manual small incision cataract surgery and phacoemulsification. Middle East Afr J Ophthalmol 2015;22(3):362-369 [FREE Full text] [doi: 10.4103/0974-9233.159763] [Medline: 26180478]
22. Jolley E, Cumaio M, Vilanculos A, Hassane I, Kimani K, Ogundimu K, et al. Changes in eye health and service coverage in Nampula, Mozambique between 2011 and 2018. Ophthalmic Epidemiol 2022 Feb 24;29(1):91-99 [FREE Full text] [doi: 10.1080/09286586.2021.1900279] [Medline: 33759697]
23. Ilechie AA, Boadi-Kusi BS, Ndudiri OV, Ofori EA. Evaluation of post-operative visual outcomes of cataract surgery in Ghana. International Journal of Health Research 2012;5(1):42 [FREE Full text]
24. Sengo DB, Saravila PJ, Chivinde SS, Mate LM, Faquihe MF, Moragues R, et al. Assessment of cataract surgery outcomes in Nampula (Mozambique): visual acuity, visual function and quality of life. Graefes Arch Clin Exp Ophthalmol 2023 Jun 10;261(6):1597-1608 [FREE Full text] [doi: 10.1007/s00417-022-05964-4] [Medline: 36625930]
25. McGuinness M, Moo E, Varga B, Dodson S, Lansingh VC, Resnikoff S, et al. The Better Operative Outcomes Software Tool (BOOST) prospective study: improving the quality of cataract surgery outcomes in low-resource settings. Ophthalmic Epidemiol 2025 Feb;32(1):76-86 [FREE Full text] [doi: 10.1080/09286586.2024.2336518] [Medline: 38635874]

XSL•FO
RenderX

26. Pattanayak S, Patra SHS, Nanda A, Subudhi P. Stabilization of refraction and timing of spectacle prescription following manual small-incision cataract surgery. Indian J Ophthalmol 2022 Nov;70(11):3938-3941 [FREE Full text] [doi: 10.4103/ijo.IJO_95_22] [Medline: 36308131]
27. Bernhisel A, Pettey J. Manual small incision cataract surgery. Curr Opin Ophthalmol 2020 Jan;31(1):74-79. [doi: 10.1097/ICU.0000000000000624] [Medline: 31770166]
28. Thakre U, Nikose A, Momin A. Estimation of surgically induced astigmatism after small-incision cataract surgery done by junior residents. Indian J Ophthalmol 2023 Mar;71(3):751-756 [FREE Full text] [doi: 10.4103/ijo.IJO_1616_22] [Medline: 36872671]
29. Pattanayak S, Mathur S, Nanda A, Subudhi B. Postoperative astigmatic considerations in manual small-incision cataract surgery - a review. Indian J Ophthalmol 2022 Nov;70(11):3785-3790 [FREE Full text] [doi: 10.4103/ijo.IJO_1627_22] [Medline: 36308097]
30. Pattanayak S, Nanda A, Swain A. Effect of depth of the sclerocorneal incision on postoperative corneal astigmatism in manual small-incision cataract surgery. Indian J Ophthalmol 2022 May;70(5):1612-1616 [FREE Full text] [doi: 10.4103/ijo.IJO_2649_21] [Medline: 35502037]
31. Mactaggart I, Limburg H, Bastawrous A, Burton MJ, Kuper H. Rapid assessment of avoidable blindness: looking back, looking forward. Br J Ophthalmol 2019 Nov 02;103(11):1549-1552 [FREE Full text] [doi: 10.1136/bjophthalmol-2019-314015] [Medline: 31266777]

Abbreviations

BCVA: best corrected visual acuity
BOOST: Better Operative Software Outcomes Tool
IOL: intraocular lens
LMIC: low- and middle-income countries
MSICS: manual small-incision cataract surgery
PinVA: pinhole visual acuity
PVA: presenting visual acuity
RAAB: Rapid Assessment of Avoidable Blindness
VA: visual acuity
WHO: World Health Organization

Edited by L Buis; submitted 18.07.24; peer-reviewed by J Loscos-Arenas, M Abdollahi; comments to author 09.11.24; revised version received 28.11.24; accepted 28.04.25; published 21.05.25.

Please cite as:
Lecumberri M, Mandlate D, Cote V, Martín-Baranera M, Moser CL
Effectiveness of the Better Operative Software Outcomes Tool for Monitoring Cataract Surgery Outcomes in Inhambane Province, Mozambique: Prospective Pilot Diagnostic Accuracy Study
JMIR Mhealth Uhealth 2025;13:e64500
URL: https://mhealth.jmir.org/2025/1/e64500
doi:10.2196/64500
PMID:40397505

©Mónica Lecumberri, Dórcia Mandlate, Vasco Cote, Montserrat Martín-Baranera, Carlos L Moser. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effectiveness of an Artificial Intelligence–Based Gamified Intervention for Improving Maternal Health Outcomes Among Refugees and Underserved Women in Lebanon: Community Interventional Trial

Shadi Saleh[1,2], MPH, PhD; Nour El Arnaout[1], MPH; Nadine Sabra[1], PharmD; Asmaa El Dakdouki[1], MSc; Zahraa Chamseddine[1], MSc; Randa Hamadeh[3], MPH, PhD; Abed Shanaa[4], MD; Mohamad Alameddine[5], MPH, PhD

[1]Global Health Institute, American University of Beirut, Beirut, Lebanon

[2]Department of Health Management and Policy, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

[3]Ministry of Public Health, Beirut, Lebanon

[4]United Nations Relief and Works Agency for Palestine Refugees in the Near East, Beirut, Lebanon

[5]College of Health Sciences, University of Sharjah, Sharjah, United Arab Emirates

Corresponding Author:
Mohamad Alameddine, MPH, PhD
College of Health Sciences
University of Sharjah
M23, Medical College, University City Road
Sharjah, 27272
United Arab Emirates
Phone: 971 521023423
Email: malameddine@sharjah.ac.ae

Abstract

Background: In Lebanon, disadvantaged pregnant women show poor maternal outcomes due to limited access to antenatal care (ANC) and a strained health care system, compounded by ongoing conflicts and a significant refugee population. Despite substantial efforts to improve maternal health, the provision of maternal health services in primary health care centers (PHCs) still faces significant challenges. Mobile health (mHealth) interventions, particularly those using artificial intelligence (AI) and gamification, are proving effective in addressing gaps in maternal health services by offering scalable and accessible care.

Objective: This study aimed to evaluate the effects of an AI-based gamified intervention, Gamification and Artificial Intelligence and mHealth Network for Maternal Health Improvement (GAIN MHI), on maternal health outcomes and uptake of ANC services among disadvantaged populations in Lebanon.

Methods: The study was a community interventional trial with historical controls, conducted across 19 randomly allocated PHCs in 5 Lebanese governorates. Participants included pregnant women in their first trimester visiting PHCs. The intervention used mHealth tools, including educational mobile-based messages, appointment reminders, and the GAIN MHI app, which provided AI-driven and gamified learning for health care providers (HCPs). Data collected covered demographics, medical history, and maternal and neonatal health outcomes. Key outcome measures included uptake of health care services (eg, ANC visits, supplement intake, ultrasound completion, lab tests) and maternal and neonatal outcomes (eg, term delivery, normal delivery, abortion rate, neonatal morbidity, maternal complications).

Results: This study included 3989 participants, divided between a control group (n=1993, 50%) and an intervention group (n=1996, 50%). Regression models adjusting for demographics, health, and obstetric characteristics showed significantly higher odds in the intervention group for completing 4 or more ANC visits (odds ratio [OR] 1.569, 95% CI 1.329-1.852, *P*<.05), completing lab tests (OR 1.821, 95% CI 1.514-2.191, *P*<.05), 2 or more ultrasound screenings (OR 7.984, 95% CI 6.687-9.523, *P*<.05), urine analysis (OR 4.399, 95% CI 3.631-5.330, *P*<.05), and supplement intake (OR 3.508, 95% CI 2.982-4.128, *P*<.05). Regarding outcomes, the intervention group had 29.5% increased odds of a term delivery (OR 1.295, 95% CI 1.095-1.532, *P*=.002) and 58% increased odds of avoiding neonatal morbidity (OR 1.580, 95% CI 1.185-2.108, *P*=.002). However, both groups showed decreased odds of normal delivery (intervention: OR 0.774, 95% CI 0.657-0.911; control: OR 0.823, 95% CI 0.701-0.964) and

XSL•FO
RenderX

increased odds of maternal complications (intervention: OR 0.535, 95% CI 0.449-0.637; control: OR 0.586, 95% CI 0.474-0.723; *P*<.05).

Conclusions: The GAIN MHI intervention effectively improves uptake of ANC and maternal and neonatal outcomes. Our findings highlight the potential of mHealth interventions to enhance health care delivery. To sustain these improvements, future research should focus on integrating mHealth with other interventions that address socioeconomic and contextual factors. This approach will further optimize maternal and neonatal health outcomes among disadvantaged populations.

(JMIR Mhealth Uhealth 2025;13:e65599) doi:10.2196/65599

KEYWORDS

maternal health; gamification; AI; artificial intelligence; mHealth; mobile health; neonatal outcomes; ANC; antenatal care; antenatal; maternal; mother; game; refugees; refugee; woman; Lebanon; neonatal; pregnant women; pregnancy; pregnant; effectiveness; Lebanese; mHealth tools; medical history; regression model

Introduction

Overview of Maternal Health in Lebanon

Women and girls, especially in conflict-affected and displaced communities, are highly vulnerable to poor sexual, reproductive, and maternal health (SRMH) [1-4]. Disadvantaged pregnant women in Lebanon encounter various maternal health complications and adverse pregnancy outcomes, including higher-than-usual cesarean section (C-section) deliveries due to pregnancy complications, preterm births, and hemorrhages, among others [5-7]. One key challenge is the limited access to antenatal care (ANC) services [8-10]. Recent investigations have underscored a remarkable decline in the use of ANC services among disadvantaged pregnant women in Lebanon, falling below the recommended threshold of 4 visits [1,6,11,12]. Barriers to accessing ANC include transportation issues, distance to primary health care centers (PHCs), shortage of skilled health care providers (HCPs), lack of awareness about available services, and the financial burden associated with seeking care [6,11,13-15]. Addressing these challenges and ensuring equitable access to maternal health care services are essential steps to improving pregnancy outcomes and reducing maternal mortality rates [16,17].

The Refugee Crisis in the Context of Lebanon

Addressing the identified barriers to accessing ANC services becomes particularly challenging in a fragile context, such as Lebanon, the country hosting the highest number of refugees worldwide. Around 1.5 million Syrian refugees, 489,292 Palestinian refugees, and 11,645 refugees from various other nationalities reside in Lebanon [18-20]. The continuous influx of refugees has strained the Lebanese health care system, resulting in medication shortages and high health care costs [21-25]. The country also faces a significant migration of health care professionals due to the protracted economic crisis it has witnessed over the past few years [22,26]. As a result, accessing basic health and social services has become particularly challenging for host communities, further exacerbating their vulnerability [27]. Among refugees, women of childbearing age face particular health care challenges, highlighting the importance of addressing women's health issues in Lebanon [28,29].

Primary Health Care Service Use Among Pregnant Women in Lebanon

To access essential medical services in Lebanon, disadvantaged Lebanese populations and Syrian refugees often resort to a network of PHCs, affiliated with the Ministry of Public Health (MOPH) [30,31]. The MOPH has established around 200 PHCs, managed by nongovernmental organizations (NGOs), municipalities, and governmental institutions and distributed across the country [32,33]. These PHCs provide affordable health care services for over 1 million disadvantaged Lebanese populations and Syrian refugees annually, with Lebanese patients benefiting from reduced out-of-pocket fees and Syrian refugees being supported by subsidies from the United Nations High Commissioner for Refugees (UNHCR) [31,34]. Alternatively, Palestinian refugees access health care services via a channel of 27 PHCs administered by the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) [30]. Both the MOPH and the UNRWA provide services related to reproductive health, including family planning, ANC, and postnatal care (PNC) services [35-37]. Recognizing the strains on Lebanon's health care system due to the influx of refugees and in alignment with its commitment to ensure universal health coverage, the MOPH collaborates with partners, such as the World Bank, the United Nations Populations Funds (UNFPA), and the European Union (EU), to implement projects aimed at strengthening primary health care services, including reproductive health care [23,31,32,38-41]. These initiatives provide wellness and sexual reproductive health (SRH) packages covering services such as screening, consultation, prescription medications, ANC, and PNC [39,40,42]. Similarly, to sustain its health care initiatives, the UNRWA relies on funding from member states of the United Nations, the EU, and various other states. Yet, despite the availability of these subsidized maternal health care services, compliance with standards continues to be a serious concern among disadvantaged populations in Lebanon [1]. To address the suboptimal maternal health services, mobile health (mHealth) interventions emerge as a potential solution to improve access to essential health care services [43].

Role of mHealth Interventions in Enhancing Maternal Health Outcomes

The proliferated usage of mobile devices has catalyzed the widespread adoption of mHealth initiatives, revolutionizing the landscape of health care delivery and empowering individuals

XSL•FO
RenderX

to actively engage in enhancing their well-being through accessible and innovative digital platforms [43-47]. mHealth emerges as a critical component for the worldwide effort to decrease maternal mortality and improve maternal health outcomes, mainly within disadvantaged populations [48,49]. By leveraging mobile technology, mHealth interventions offer scalable solutions to address maternal health challenges, providing vital resources, education, and support to pregnant women and new mothers, in remote or underserved areas [43]. mHealth interventions, such as Mobile Technology for Community Health (MOTECH), MOM Connect, and Mobile Alliance for Maternal Action (MAMA), highlight the effectiveness of strategies such as sending voice note messages and appointment reminders to expectant mothers and their spouses [50-52]. These interventions not only facilitate timely access to essential maternal health services but also promote greater engagement and support throughout the prenatal and postnatal journey. The integration of mHealth interventions extends beyond empowering the pregnant women and their spouses, by targeting their HCPs as well. Initiatives such as the Safe Delivery Application (SDA) show that such interventions have positively impacted HCPs' knowledge and skills, resulting in tangible improvements in maternal and neonatal outcomes.

As mHealth continues to advance, artificial intelligence (AI) innovations are being increasingly integrated into the health care sector to enhance the quality of care and address various health areas [53,54]. AI has been used in diagnostic support, as well as to enhance access to virtual health services in remote areas and address knowledge and skill gaps [55]. Building on these advancements, AI has shown significant promise in maternal health by predicting complications, such as preeclampsia, postpartum depression, gestational diabetes and anemia, stillbirths, and preterm deliveries [56]. Additionally, AI shortens the training time required for health care professionals in assessing preterm deliveries and improves access to maternal health care [57,58].

Moreover, the concept of gamification has been incorporated into mobile platforms to train HCPs and inform individuals about health and illness management [59-62]. For instance, the Maternal and Newborn Technology for Resilience in Rural Areas (MANTRA) app applies gamification to increase awareness regarding maternal and child health risk factors [63]. Users expressed satisfaction with the gamification approach, which led to improved maternal and child health knowledge and more informed decision-making [63].

These findings underscore the profound efficacy of mHealth interventions in maternal health outcomes and the accessibility of maternal health services [64,65].

mHealth Interventions for Maternal Health in Lebanon

In Lebanon, mHealth technologies have been incorporated to improve primary health care services for disadvantaged Lebanese and refugee populations [64,66]. These initiatives focus on using mHealth tools to manage noncommunicable diseases (NCDs) through interventions such as educational SMS messages and appointment reminders [67,68]. Other studies in Lebanon have investigated mHealth for weight management and the uptake of routine mammography [69,70]. In the realm of maternal health, the UNRWA electronic mother and child health (e-MCH) app represents a notable intervention. It integrates a comprehensive database on maternal and infant care and provides appointment reminders, thereby playing a critical role in ensuring the continuity of essential health care services [71].

Although certain countries in the Middle East and North Africa (MENA) region have effectively used mHealth solutions to enhance maternal health literacy and ANC visits, the focus on maternal health in Lebanon remains limited [72]. There is a scarcity of AI-driven solutions or gamified elements, highlighting a critical gap in leveraging technology to improve maternal health outcomes [73].

Study Aim

This study aimed to assess the effects of an AI-based gamified intervention, Gamification and Artificial Intelligence and mHealth Network for Maternal Health Improvement (GAIN MHI), on maternal health outcomes and uptake of ANC services among disadvantaged populations in Lebanon.

Methods

Study Design

This study was a community interventional trial with historical controls at distinct PHCs. Pre- and postintervention data were collected with the help of the assigned research assistant at each center. Data collected included demographics, medical and obstetric history, and maternal and neonatal health outcomes, including neonatal morbidities and maternal complications. The study comprised 3 distinct phases: preintervention data collection, intervention, and subsequent postintervention data collection. The preintervention phase spanned from July to December 2021. The intervention phase took place from September 2021 to December 2022. Finally, the postintervention phase extended from March to June 2023.

Center Selection

In total, 20 PHCs across 5 governorates in Lebanon—Beirut, Bekaa, North Lebanon, Mount Lebanon, and South Lebanon—were selected for this study. These centers serve disadvantaged Lebanese populations and refugees. Due to logistical constraints, 1 PHC was excluded, resulting in 19 (95%) centers. Of these, 10 (53%) were designated as control centers and 9 (47%) as intervention centers. Among the selected PHCs, 9 (47%; both control and intervention) are affiliated with the MOPH and serve Syrian refugees and disadvantaged Lebanese populations. The remaining 10 (53%) PHCs operate under the UNRWA and serve Palestinian refugees in Lebanon.

Study Population

Our target population in both pre- and postintervention phases consisted of pregnant women who presented at the selected PHCs during their first trimester and continued visiting the health centers for at least a year. All age groups were included. In the preintervention phase, pregnant women attended 1 of the PHCs for ANC during their first trimester between August 2018 and August 2020 and continued visiting the center until their delivery or termination of the pregnancy. However, in the

XSL•FO
RenderX

postintervention phase, pregnant women were already enrolled in the intervention and visited the PHCs during the period of the intervention. Controls were collected using the historical medical records of the patients at both timepoints.

The GAIN MHI Intervention

The GAIN MHI intervention (Figure 1) consisted of 2 different modalities: mobile-based messages (text and voice) and a mobile app (GAIN MHI app) aimed at enhancing maternal health outcomes (Figure 2).

Figure 1. Visual representation of the different components of the GAIN MHI intervention. ANC: antenatal care; GAIN MHI: Gamification, Artificial Intelligence and mHealth Network for Maternal Health Improvement; MOPH: Ministry of Public Health; PHC: primary health care center; UNRWA: United Nations Relief and Works Agency for Palestine Refugees in the Near East.

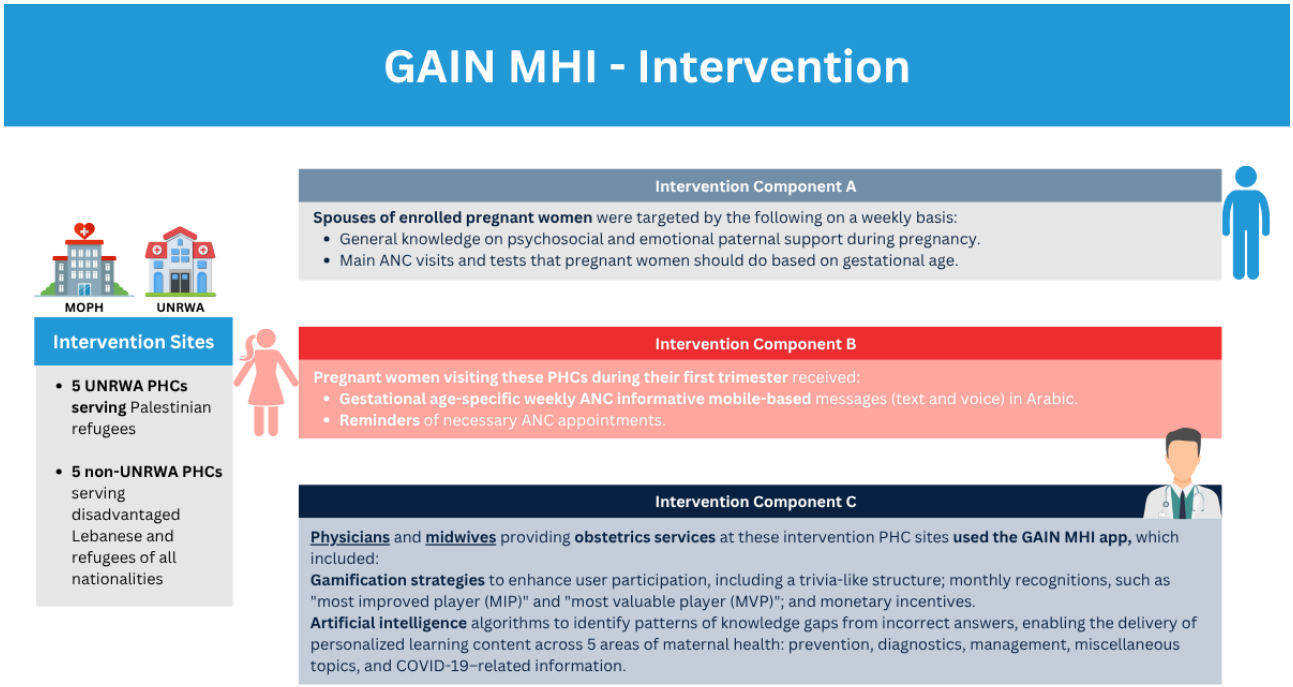
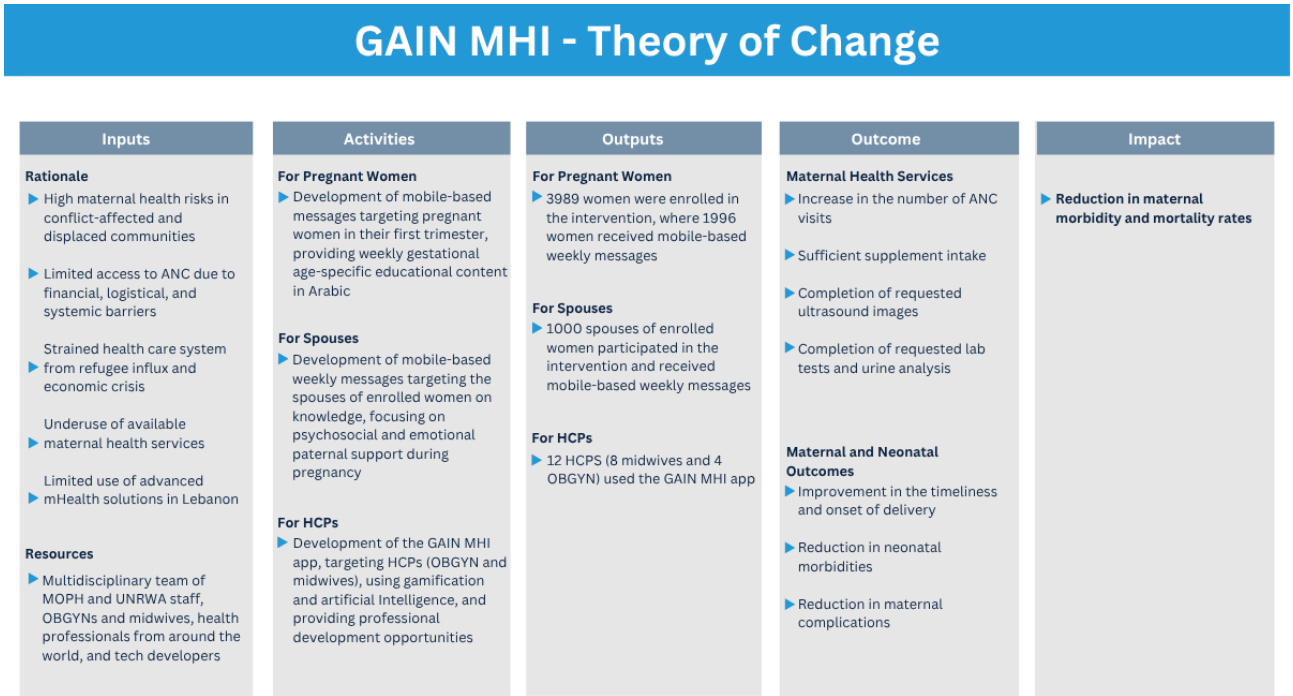


Figure 2. GAIN MHI theory of change. ANC: antenatal care; GAIN MHI: Gamification, Artificial Intelligence and mHealth Network for Maternal Health Improvement; HCP: health care provider; mHealth: mobile health; MOPH: Ministry of Public Health; OBGYN: obstetricians and gynecologists; UNRWA: United Nations Relief and Works Agency for Palestine Refugees in the Near East.



XSL•FO
RenderX

The mobile-based messages targeted pregnant women in their first trimester, including disadvantaged Lebanese populations and Syrian and Palestinian refugees, providing weekly gestational age-specific educational content in Arabic. The content covered pregnancy education, nutritional recommendations and supplementation, pregnancy danger signs, and postpartum care for 6 months after delivery, encouraging postpartum visits, breastfeeding, contraception, maternal mental health, and infant vaccinations. Reminder messages were also sent for upcoming ANC visits, vaccinations, tests, and medical examinations. Additionally, weekly messages were sent to the spouses of enrolled women on knowledge, focusing on psychosocial and emotional paternal support during pregnancy.

Although the focus was primarily on supporting pregnant women and their spouses, targeting HCPs was crucial. HCPs face various challenges, such as insufficient continuing education programs and career advancement, which hinder their work, notably in Lebanon [74,75]. To address these gaps, the GAIN MHI app, developed by the Global Health Institute at the American University of Beirut (AUB GHI), integrated gamification and AI concepts for the professional development of HCPs, particularly obstetricians and gynecologists (OBGYN) and midwives, working at the selected PHCs. This app features game principles such as a trivia-like concept, monthly recognition (most improved player [MIP] and most valuable player [MVP]), and monetary incentives to enhance learning engagement. These elements foster a sense of competition, motivation, and continuous improvement among HCPs. Additionally, the AI algorithms used in the app identify knowledge gaps based on the HCPs' incorrect answers, tailoring subsequent questions to individualize learning across 5 categories related to maternal health: prevention, diagnostic, management, miscellaneous, and COVID-19–related questions. For this purpose, a database of information on prenatal care was created. Topic modeling was used on this database to categorize the different topics of the stored content. Likewise, topic modeling was applied to the question bank, enabling the identification of topics linked to the incorrectly answered questions. Relevant information related to the identified topics was then retrieved and provided to the user. By leveraging AI, the GAINMHI app transforms traditional professional development into individualized efficient processes, minimizing the burden of standardized training programs that may not account for individual variability. In addition to this personalized learning approach, a randomized spin wheel is used for category selection, with each spin providing a question followed by explanations after the user answered. The app also includes a Knowledge Resource Center (KRC), and it communicates with users via automated WhatsApp messages, providing feedback on scores, answers, remaining attempts, monthly performance, and rewards. Available in both English and Arabic, the GAIN MHI app represents an innovative approach to advancing HCP education and maternal care.

Although the professional development of HCPs was an integral part of the intervention, it was not the primary focus of this research and is discussed in more detail in a separate paper.

Data Collection

In both pre- and postintervention phases, secondary data were collected from the medical charts of pregnant women who presented at the selected PHCs during their first trimester and continued visiting the centers for at least a year. In case a woman got pregnant more than once during the selected study period, her latest pregnancy charts were collected if they met the inclusion criteria of the project, while data on previous pregnancies were excluded. The data extracted from the medical charts of eligible pregnant women covered the following aspects: demographics, medical and surgical history, pregnancy history (supplements and ANC vaccinations taken, the first lab tests and ultrasound scans undergone, and any maternal complications encountered), neonatal complications, and breastfeeding history. Primary data on the knowledge, attitude, and practices (KAP) of women were collected before the intervention through quantitative surveys. The results of the KAP surveys will be discussed in a different paper. Primary data were also collected from pregnant women and their spouses participating in the intervention concerning their satisfaction with the mobile-based messages they were receiving. These data were collected through quantitative surveys at the end of the intervention; the results of these surveys are discussed in another paper. Satisfaction of HCPs with the GAIN MHI app was studied and captured through mixed methods surveys administered at the end of the intervention; the results of these surveys are discussed in a separate paper.

Ethical Considerations

Ethical approval was obtained from the Institutional Review Board (IRB) of the American University of Beirut (protocol number SBS-2020-0317) and the respective Ethical Committees at the MOPH and the UNRWA. Informed consent was obtained from all individuals participating in the intervention. Participants were informed of their right to opt out of the intervention at any point in time. The IRB approval covers secondary analysis without additional consent. All data were anonymized and deidentified. No compensation was received by participants.

Study Parameters

Study parameters were extracted and recorded for all pregnant women in the pre- and postintervention phases in all the centers. Table 1 shows the set of indicators that were selected to monitor the effectiveness of the intervention, along with their definitions. The selected parameters were categorized into 2 domains: uptake to health care and maternal and neonatal outcomes. Uptake of health care services included completing more than 4 ANC visits, having at least 2 ultrasound screenings taken, undergoing the available blood tests at the PHCs, undergoing the urine analysis test, and taking supplements during pregnancy. Onset of delivery on term, normal delivery, reduction in abortion, and reduction in any neonatal morbidity and any maternal complication were indicators constituting the domain of maternal and neonatal outcomes.

XSL•FO
RenderX

Table 1. Selected indicators to monitor the effectiveness of the interventions on uptake to health care and maternal and neonatal outcomes.

Domain and parameters	Definition
Uptake of health care services	
ANC[a] visits	Percentage of pregnant women who completed 4 or more ANC visits based on the World Health Organization recommendations
Supplement intake	Percentage of pregnant women taking the required vitamins that a women should take, including folic acid, iron, calcium, and magnesium
Ultrasound screenings	Percentage of pregnant women who underwent 2 or more ultrasound screenings during their pregnancy period
Blood tests	Percentage of pregnant women who underwent the available blood tests at their PHCs[b], including fasting blood sugar (FBS), blood group (+/–Rh), and the complete blood count (CBC)
Urine analysis	Percentage of pregnant women who underwent the urine analysis test
Maternal and neonatal outcomes	
Onset of delivery	Percentage of pregnant women who had a term delivery, delivering their latest child after week 37
Method of delivery	Percentage of pregnant women who had a normal delivery
Abortion	Percentage of women who had an abortion in their latest pregnancy before week 22
Neonatal morbidities	Percentage of pregnant women who faced any neonatal complications, including neonatal intensive care unit (NICU) admission, preterm delivery, Iatrogenic preterm birth injury, low birthweight, and neonatal early death
Maternal complications	Percentage of pregnant women who had any maternal complications, including gestational hypertension, gestational diabetes, perinatal depression, preterm labor, stillbirth, abortion, postpartum infection, postpartum hemorrhage, and postpartum depression

[a]ANC: antenatal care.

[b]PHC: primary health care center.

Statistical Analysis

Pregnant woman with a nationality other than Lebanese, Syrian, or Palestinian were excluded from the analysis. Sample baseline characteristics were calculated for the intervention and control arms using count and percentages or means (SDs). Chi-square (χ^2) and independent sample *t* tests were conducted to check for all possible differences in the baseline characteristics between the control and intervention arms. Given the fact that all the indicators are categorical, bivariate analysis was performed using the Pearson chi-square (χ^2) test to test the significant difference pre- and postintervention. To evaluate the impact of the intervention, separate logistic regressions were generated for each of the study indicators over the study period in each arm, while controlling for all the possible baseline differences between the control and intervention arms as age, nationality, history of chronic diseases, preterm delivery, abortion, C-section, gravida, and parity. Analysis was carried out at a .05 significance level using Stata SE software.

Results

Participants' Demographic Characteristics

Table 2 presents the baseline demographic and health characteristics of the participants in the control and intervention groups. The number of participants was almost equal between the control (n=1993, 50%) and the intervention (n=1996, 50%) group. At baseline, more than half of the sample was aged between 18 and 29 years in both intervention (801/1996, 54.3%) and control (806/1993, 53%) groups. The majority of the participants were Palestinians, constituting 71.3%(n=1423) of those in the intervention group and 63.5% (n=1267) of those in the control group. The rest of the participants were either Lebanese or Syrians. A higher number of patients in the intervention group (n=160, 8.1%) had a chronic disease. Regarding obstetric history, 6% (n=121) of participants in the intervention group had had a previous preterm delivery, 30.3%(n=605) had had a previous abortion, and 33% (n=664) had had a previous C-section delivery. In contrast, 4.3% (n=85) of participants in the control group had had a preterm delivery, 28.9% (n=575) had had an abortion, and 26.5% (n=529) had had a C-section previously. The mean gravida for the intervention and control groups was 2.98 (SD 2.02) and 3.12 (SD 1.86), respectively. There was a significant difference between the intervention and control groups in all the characteristics except for age.

XSL•FO
RenderX

Table 2. Baseline demographic and health characteristics by study group (N=3989).

Characteristics	Intervention group (n=1996)	Control group (n =1993)	*P* value
Age (years), n (%)			
<18	46 (2.3)	48 (2.4)	.50
18-29	1083 (54.3)	1057 (53.0)	—[a]
30-39	801 (40.1)	806 (40.4)	—
≥40	66 (3.3)	82 (4.1)	—
Nationality, n (%)			
Lebanese	183 (9.2)	228 (11.4)	<.001[b]
Palestinian	1423 (71.3)	1267 (63.6)	—
Syrian	390 (19.5)	498 (25.0)	—
Health and obstetric characteristics			
Have chronic diseases, n (%)	160 (8.1)	64 (3.2)	<.001[b]
Previous preterm delivery[c], n (%)	121 (6.1)	85 (4.3)	.01[b]
Previous abortion[c], n (%)	605 (30.3)	575 (28.9)	.15
Previous C-section[c,d], n (%)	664 (33.3)	529 (26.5)	<.001[b]
Gravida, mean (SD)	2.98 (2.02)	3.12 (1.86)	<.001[b]
Parity, mean (SD)	1.50 (1.34)	1.69 (1.51)	<.001[b]

[a]Not applicable.

[b]Significant at .05.

[c]Any previous maternal characteristic excludes the latest pregnancy.

[d]C-section: cesarean section.

Indicators Assessing Uptake of Health Care Services

Results of the bivariate analysis of the selected study indicators are presented in Table 3 across the 2 groups pre- and postintervention. Regarding indicators assessing uptake of health care services, during pregnancy, there was a significant rise in the percentage of women completing 4 or more ANC visits between the pre- and postintervention periods in both intervention and control groups. Specifically, the increase was higher in the intervention group, where the percentage of women completing 4 or more ANC visits increased from 71.9% (n=1436) preintervention to 79.9% (n=1251) postintervention (*P*<.001), while in the control group, the percentage increased from 68.6% (n=1367) preintervention to 74.5% (n=980) postintervention (*P*<.001). A similar trend was noticed in the percentage of women undergoing more than 2 ultrasound screenings, with a significant increase in the percentage from 44.2% (n=881) to 86.1% (n=1323) in the control group and from 46% (n=923) to 80.7% (n=1013) in the intervention group pre- and postintervention, respectively. Likewise, a significant increase was found in the percentage of women who received the intervention and completed lab tests (preintervention: n=1467, 73.5%; postintervention: n=1291, 82.5%; *P*<.001). On the contrary, in the control group, the percentage of women who completed the blood tests available at their corresponding PHCs decreased. Furthermore, there was an overall increase in the percentage of women undergoing urine analysis through their pregnancy, with a higher increase in the intervention group (preintervention: n=1276, 63.9%; postintervention: n=1379, 89.7%; *P*<.001) compared to the control group (preintervention: n=1746, 87.6%; postintervention: n=1161, 94.4%; *P*<.001). Additionally, there was an almost equivalent increase in the percentage of women taking their required supplements pre- and postintervention, respectively, in both intervention (n=906, 45.4%, to n=1162, 74.3%; *P*<.001) and control (n=768, 38.5%, to n=850, 64.6%; *P*<.001) groups.

XSL•FO
RenderX

Table 3. Bivariate analysis of the selected outcomes/indicators by study group pre- and postintervention (N=3989).

Indicators	Intervention group (n=1996)			Control group (n=1993)		
	Preintervention	Postintervention	*P* value	Preintervention	Postintervention	*P* value
Uptake of health care services						
ANC[a] visits (≥4), n (%)	1436 (71.9)	1251 (79.9)	<.001[b]	1367 (68.6)	980 (74.5)	<.001[b]
Ultrasound screenings (≥2 images), n (%)	881 (44.2)	1323 (86.1)	<.001[b]	923 (46.3)	1013 (80.7)	<.001[b]
Completed lab tests, n (%)	1467 (73.5)	1291 (82.5)	<.001[b]	1556 (78.1)	919 (69.9)	<.001[b]
Completed urine analysis, n (%)	1276 (63.9)	1379 (89.7)	<.001[b]	1746 (87.6)	1161 (94.4)	—[c]
Took the required supplements, n (%)	906 (45.4)	1162 (74.3)	<.001[b]	768 (38.5)	850 (64.6)	<.001[b]
Maternal and neonatal outcomes						
Onset of delivery: term, n (%)	1444 (72.5)	1011 (76.9)	.004[b]	1444 (85.4)	1011 (87.6)	.10
Neonatal morbidities, n (%)	159 (7.8)	85 (5.4)	.003[b]	69 (3.5)	40 (3.0)	.51
Method of delivery: normal, n (%)	962 (51.9)	652 (46.1)	.001[b]	1072 (59.4)	631 (54.0)	.004[b]
Abortion, n (%)	143 (7.16)	150 (9.6)	.009[b]	187 (9.4)	146 (11.1)	.11
Maternal complications, n (%)	309 (15.5)	407 (26.0)	<.001[b]	311 (15.6)	314 (23.9)	.02[b]

[a]ANC: antenatal care.

[b]Significant at .05.

[c]Not applicable.

Maternal and Neonatal Outcomes

Analysis of the data from the charts showed that the onset of delivery improved significantly in the intervention group but not in the control group. After the implementation of the intervention, the proportion of women who had a term delivery increased from 72.5% (n=1444) to 76.9% (n=1011) in the intervention group (*P*=.004). Similar results were observed in regard to neonatal morbidities, where the proportion of women decreased from 7.8% (n=159) to 5.4% (n=85) in the intervention group (*P*=.003). However, the proportion of women who had a normal delivery decreased in both intervention (preintervention: n=962, 51.9%; postintervention: n=652, 46.1%; *P*<.001) and control (preintervention: n=1072, 59.4%; postintervention: n=631, 54%; *P*=.004) groups, with a notable rise in the proportion of woman having C-section deliveries. Results of maternal complications were in the same direction of delivery and abortion, where a significant increase was detected in the proportion of women with maternal complications in both intervention (preintervention: n=309, 15.5%; postintervention: n=407, 26%; *P*<.001) and control (preintervention: n=311, 15.6%; postintervention: n=314, 23.9%; *P*=.02) groups.

Regression Models

Separate logistic regression models were generated for the control and intervention groups to quantify the intervention's impact over the study period (pre- and postintervention), while controlling for baseline characteristics that were different across the 2 groups (see Table 2). Each of the 2 models was adjusted for age, nationality, condition of existing chronic diseases, and obstetric history features (ie, previous preterm delivery, previous abortion, previous C-section, gravida, and parity). Tables 4 and 5 show the adjusted odds ratios (ORs) resulting from the logistic regression analysis of the selected indicators. Considering the domain of uptake of health care services, Table 4 indicates that there was a 56.9% increase in the odds of completing more than 4 ANC visits postintervention compared to preintervention (OR 1.569, 95% CI 1.329-1.852, *P*<.001), adjusting for the difference in the demographical, health, and obstetric characteristics. The OR of ANC visits was not statistically significant in the control group (*P*=.58). In contrast, concerning ultrasound images, participants in the intervention group had significantly higher odds (OR 7.984, 95% CI 6.687-9.523, *P*<.001) of having 2 or more images postintervention compared to the control group, where there was also a significant but lower improvement (OR 3.432, 95% CI 2.836-4.153, *P*<.001). A significantly high increase in the odds of pregnant women undergoing urine analysis was observed postintervention in the intervention group (OR 4.399, 95% CI 3.631-5.330, *P*<.001). Concerning lab tests, an 82.1% increase in the odds of pregnant women completing the required blood tests during their pregnancy was noted postintervention compared to preintervention (OR 1.821, 95% CI 1.514-2.191, *P*<.001).

XSL•FO
RenderX

Table 4. Adjusted OR[a] from logistic regression analysis of the selected indicators (study parameters in the uptake of health care services domain) on the study period (pre- and postintervention).

Indicator and study periods	Intervention group (n=1996)		Control group (n=1993)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
ANC[b] visits (≥4 vs <4 as reference)				
Preintervention	__[b]	—	—	—
Postintervention	1.569[c] (1.329-1.852)	<.001[d]	1.053[c] (0.876-1.266)	.58
Ultrasound images (≥2 vs <2 as reference)				
Preintervention	—	—	—	—
Postintervention	7.984[c] (6.687-9.532)	<.001[d]	3.432[c] (2.836-4.152)	<.001[d]
Supplement uptake (yes vs no as reference)				
Preintervention	—	—	—	—
Postintervention	3.508[c] (3.020-4.076)	<.001[d]	2.453[c] (2.055-2.928)	<.001[d]
Urine analysis (yes vs no as reference)				
Preintervention	—	—	—	—
Postintervention	4.399[c] (3.631-5.330)	<.001[d]	1.323[c] (0.966-1.812)	.08
Lab tests (yes vs no as reference)				
Preintervention	—	—	—	—
Postintervention	1.821[c] (1.514-2.191)	<.001[d]	0.372[c] (0.306-0.451)	<.001[d]

[a]OR: odds ratio.

[b]Reference category.

[c]Adjusting for different baseline characteristics: age, nationality, history of chronic diseases, abortion, cesarean section (C-section), gravida, and parity.

[d]Significant at .05.

XSL•FO
RenderX

Table 5. Adjusted OR[a] from logistic regression analysis of the selected indicators (study parameters in the maternal and neonatal outcomes domain) on the study period (pre- and postintervention).

Indicator and study periods	Intervention group (n=1996)		Control group (n=1993)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Delivery onset (term vs preterm as reference)				
Preintervention	__[b]	—	—	—
Postintervention	1.295[c] (1.095-1.532)	.002[d]	1.216[c] (0.944-1.565)	.13
Delivery type (normal vs C-section as reference)				
Preintervention	—	—	—	—
Postintervention	0.774[c] (0.657-0.911)	.002[d]	0.865[c] (0.716-1.046)	.14
Abortion (no vs yes as reference)				
Preintervention	—	—	—	—
Postintervention	0.702[c] (0.546-0.902)	.006[d]	0.714[c] (0.546-0.934)	.01[d]
Maternal complications (no vs yes as reference)				
Preintervention	—	—	—	—
Postintervention	0.535[c] (0.449-0.637)	<.001[d]	0.586[c] (0.474-0.723)	<.001[d]
Neonatal morbidities (no vs yes as reference)				
Preintervention	—	—	—	—
Postintervention	1.580[c] (1.185-2.108)	.002[d]	1.353[c] (0.825-2.218)	.23

[a]OR: odds ratio.

[b]Reference category.

[c]Adjusting for different baseline characteristics: age, nationality, history of chronic diseases, abortion, cesarean section (C-section), gravida, and parity.

[d]Significant at .05.

The analysis of the intervention's impact on maternal and neonatal outcomes revealed significant findings (Table 5). Specifically, the odds of experiencing a term delivery increased significantly by 29.5% (OR 1.295, 95% CI 1.095-1.532, *P*=.002) only in the intervention group. However, concerning abortion and delivery type, the study period was associated with decreased odds of normal deliveries (OR 0.774, 95% CI 0.657-0.911, *P*=.002) and decreased odds of not having an abortion (OR 0.702, 95% CI 0.657-0.911, *P*=.006). Table 5 also indicates that the odds of not having any maternal complication were 0.535 (95% CI 0.449-0.637, *P*<.001) in the intervention group and 0.586 (95% CI 0.474-0.723, *P*<.001) in the control group. In contrast, there was a significant increase of 58% in the odds of not experiencing any of the neonatal morbidities during the post- intervention period (OR=1.580, 95% CI 1.185-2.108, *P*=.002). Notably, this increase was detected as well in the control group but failed to demonstrate statistical significance.

Discussion

Principal Findings

This study assessed the impact of mHealth and AI-based gamified intervention on improving maternal health outcomes and enhancing the uptake of ANC services among disadvantaged Lebanese and refugee pregnant women. The study demonstrated significant improvements in maternal and neonatal health outcomes following the intervention. Pregnant women in the intervention group showed higher odds of completing 4 or more ANC visits (OR 1.569, 95% CI 1.329-1.852), undergoing urine analysis (OR 4.399, 95% CI 3.631-5.330), and completing required blood tests (OR 1.821, 95% CI 1.514-2.191) postintervention. Similarly, the odds of completing 2 or more ultrasound screenings were significantly higher in the intervention group (OR 7.984, 95% CI 6.687-9.523) than in the control group (OR 3.432, 95% CI 2.836-4.153). Supplement uptake also improved significantly in both groups, with a greater increase observed in the intervention group (from 45.4% to 74.3%, *P*<.05). The intervention significantly increased the odds of term delivery (OR 1.295, 95% CI 1.095-1.532) and reduced neonatal morbidities (OR 1.580, 95% CI 1.185-2.108). However, both groups showed decreased odds of a normal delivery and increased maternal complications, highlighting the need for continued efforts to mitigate these outcomes.

Uptake of Primary Health Care Services

Our study showed that there was an increase in the overall uptake of ANC visits among pregnant women in the intervention group compared to those in the control group. Such findings are in line with other studies demonstrating that implementing mHealth interventions, such as sending maternal health text messages to pregnant women, yields positive outcomes on the uptake of ANC and overall health [76-78]. Recent studies have suggested that sending reminders specifically for health care

XSL•FO
RenderX

appointments could be more effective than not sending any reminders at all [79-83]. The increased likelihood of attending at least 4 ANC visits could be explained by the additional maternal health information and psychosocial support provided by the message content generally and the ANC attendance reminder messages specifically in the GAIN MHI intervention. In a fragile context, such as Lebanon, where health care access is limited, such interventions are crucial for improving maternal and neonatal health [22]. Scalable and cost-effective mHealth solutions address systemic barriers in resource-constrained settings, ensuring better care continuity and health outcomes [43].

The higher odds of ultrasound use, completion of laboratory tests, adherence to supplements, and undergoing urinalysis among pregnant women in the intervention group emphasizes the effectiveness of mHealth interventions in promoting and actually improving ANC-seeking behavior. Our findings are consistent with other studies on mHealth and maternal health, where increased uptake of health care services was reported. A study in India demonstrated that the implementation of an mHealth intervention resulted in improved health behaviors among pregnant women, specifically by encouraging consistent adherence to supplementation. Another study in Bangladesh showed that receiving and listening to messages resulted in a significant increase in the rate of completion of laboratory tests [84]. The intervention's effectiveness might stem from its contextualized approach, which probably helped overcome logistical, linguistic, and geographical barriers, while also increasing awareness among pregnant women. It is important to highlight that substantial progress has been observed in both intervention and control groups, with notably greater improvements seen in the intervention group. This observation may be potentially due to the fact that PHCs in Lebanon benefit from external funding and ongoing projects, under the auspices of the MOPH and the UNRWA, aimed at enhancing overall health and maternal well-being [35-40,85]. The implementation of these projects at the selected PHCs may have contributed to the results obtained. In fragile contexts, such as Lebanon, mHealth interventions play a critical role in overcoming systemic barriers, improving ANC-seeking behaviors, and supporting maternal health outcomes [78]. By addressing logistical and awareness barriers, these interventions improve maternal health outcomes. Notably, progress in both intervention and control groups highlights the role of external funding and initiatives by the MOPH and the UNRWA in strengthening PHC services, further supporting maternal well-being.

Maternal Outcomes

Our study also showed a significant increase in term deliveries in the intervention group compared to the control group. One possible explanation for this could be that the increased overall uptake of health care services and supplement intake may play a potential role in decreasing the rate of preterm births. Notably, although the increase in the uptake of health care services was evident in both intervention and control groups, the likelihood was greater in the intervention group, possibly due to the encouragement provided by the GAIN MHI mHealth intervention. Additional research conducted by Gillespie [86] and Li et al [87] showed similar findings, suggesting that the rate of preterm births could potentially be decreased through the augmentation of supplement intake.

The increase in the proportion of pregnant women undergoing C-section deliveries in both intervention and control groups without a clear explanation is concerning, especially considering that the unnecessary use of C-sections may lead to negative health consequences and inefficiencies in resource allocation [88-90]. In Lebanon, understanding and addressing this trend is crucial to ensure optimal maternal health outcomes. The rise in C-section rates without medical indications is a worldwide concern and, unfortunately, is expected to persist with nearly a third (29%) of all births likely to take place by C-section by 2030 [91]. This alarming projection emphasizes the urgent need for a concerted effort to address the underlying factors driving this trend. It is essential to recognize that a singular mHealth intervention is insufficient to fully address complex maternal health challenges. This underscores the need for a comprehensive approach involving multiple stakeholders, including HCPs, policymakers, and communities. By prioritizing evidence-based practices and policies, there is an opportunity toward promoting vaginal births, when safe and appropriate, while ensuring that C-sections are reserved for cases where they are medically necessary. This should include providing continuous professional development for HCPs and implementing community education programs to raise awareness about the risks and benefits of different delivery methods. Such measures are essential for safeguarding maternal and neonatal health and reducing the burden of unnecessary surgical interventions.

In both intervention and control groups, there were no observed improvements concerning abortion and maternal complications among the disadvantaged Lebanese and refugee pregnant women. This absence of improvement underscores that maternal health outcomes are intricately influenced by various determinants that extend well beyond clinical interventions, encompassing social, economic, and environmental factors [92,93]. The disadvantaged status of the study population highlights the significant impact of socioeconomic disparities, particularly on women, which not only affect their own well-being but also shape the life prospects of their newborns. Financial constraints, in particular, may contribute to the decision to undergo abortion due to the substantial financial responsibilities associated with raising children. Although the GAIN MHI intervention has yielded promising results, it is evident that merely offering educational messages and appointment reminders is not sufficient when financial barriers and social determinants still hinder the uptake of essential maternal health care services, including addressing and treating maternal complications. Similar results were documented in a study on an mHealth intervention, the MatHealth app, indicating that financial challenges can impede the functionality of the intervention. As such, recognizing the broader determinants of maternal health outcomes, it remains important to address socioeconomic disparities and financial barriers alongside clinical interventions. By engaging in collaborative efforts, including comprehensive support, community empowerment initiatives, and adequate funding mechanisms, equitable improvements in maternal and neonatal health outcomes for

XSL•FO
RenderX

disadvantaged populations can be achieved. Persistent socioeconomic barriers, particularly among disadvantaged Lebanese and refugee women, limit the intervention's impact on maternal complications and abortion rates [22]. Financial constraints and social determinants remain significant obstacles to accessing essential care. Addressing these challenges requires a comprehensive approach that combines mHealth interventions with socioeconomic support, professional development for HCPs, and community education, ensuring equitable maternal and neonatal health outcomes.

Neonatal Outcomes

In terms of neonatal health outcomes, our study demonstrated a significant increase in the likelihood of not facing any neonatal morbidities postintervention. This aligns with the existing literature, such as the study by Bush et al [94], who found that the implementation of mHealth apps, such as the WYhealth Due Date Plus app, designed for pregnant women, resulted in a marked reduction in the incidence of low birth weight. Similarly, Nishimwe et al [95] observed an improvement in newborn outcomes with the sue of the Safe Delivery mHealth app targeting midwives. These findings underscore the multifaceted approach of the GAIN MHI intervention, which not only prioritizes support for expectant mothers but also extends its impact to their HCPs.

In light of these findings, it is essential to integrate mHealth solutions into the primary health care system, ensuring widespread uptake of maternal health information and appointment reminders. In practice, efforts should focus on contextualizing these interventions to address logistical, linguistic, and geographical barriers and include comprehensive support mechanisms to mitigate socioeconomic disparities, thereby fostering equitable maternal and neonatal health improvements. To effectively address the multifaceted nature of maternal health, it is recommended to integrate mHealth interventions within comprehensive maternal health programs. These programs should address both clinical and nonclinical factors that influence health outcomes, ensuring a holistic approach to maternal care. In fragile contexts, such as Lebanon, improving neonatal health is crucial. The GAIN MHI intervention significantly reduced neonatal morbidities, aligning with studies showing mHealth apps improve birth outcomes by enhancing maternal care and supporting HCPs [78]. Integrating mHealth solutions into primary health care, while addressing logistical and socioeconomic barriers, can foster equitable maternal and neonatal health improvements. A holistic approach that combines mHealth with broader maternal health programs is essential for sustained impact.

Equity and Gender Considerations

The GAIN MHI project prioritized equity and gender considerations across its design, implementation, and evaluation phases to ensure inclusivity and fairness in addressing maternal health challenges. All eligible women were recruited without discrimination based on race, ethnicity, nationality, religion, or socioeconomic status. The educational interventions were tailored to accommodate varying literacy levels and socioeconomic backgrounds, empowering women to make informed decisions about their health regardless of their educational or financial circumstances.

Recognizing the cultural and social dynamics of the MENA region, the project incorporated a participatory approach that engaged women and their spouses. Weekly educational messages targeted men to foster shared responsibility in maternal health, challenge traditional gender norms, and promote spousal support. This approach addressed critical barriers in underprivileged communities, where women often require the husband's consent to access health care services or purchase medications.

Additionally, the project leveraged AI and gamification through the GAIN MHI app to enhance the professional development of HCPs. By addressing knowledge gaps and tailoring training to individual needs, the app minimized inequities in access to continuing education for HCPs, particularly those serving disadvantaged populations. The integration of these innovative tools ensured that health care delivery was equitable, culturally appropriate, and responsive to the diverse needs of pregnant women and their communities.

Although the intervention aimed to promote equity, the potential for technology to exacerbate inequities was carefully mitigated. For example, the app was made available in English and Arabic to reduce language barriers, and the inclusion of automated feedback via WhatsApp ensured accessibility for users with limited digital literacy.

Limitations

Despite the positive outcomes of the GAIN MHI intervention, several limitations and potential biases must be acknowledged. The use of historical controls introduces selection bias, as differences between pre- and postintervention groups may reflect not only the intervention's impact but also other temporal changes, such as shifts in health care policies, provider practices, or patient demographics. Additionally, reliance on existing medical charts for data collection compromised data quality due to inconsistencies and missing information. Human errors in documentation and incomplete charts at PHCs, often caused by patients seeking care at private clinics or skipping appointments, posed challenges for tracking full medical histories. This fragmented documentation hindered comprehensive data collection and analysis, potentially reducing the reliability of the study's findings. The exclusive focus on objective data from medical charts also limited the study's ability to capture qualitative aspects, such as reasons for abortion or barriers to accessing health care, which are critical for understanding patients' experiences and decision-making processes. This underscores the need for complementary qualitative research to provide a holistic view of the factors influencing health care outcomes.

Cultural norms in the MENA region further complicated the intervention's implementation. Gender norms may have influenced the women's and their spouses' willingness to fully participate, and despite efforts to encourage spousal involvement through educational messages, varying levels of male engagement may have unequally affected maternal health outcomes. The study's focus on pregnant women attending

XSL•FO
RenderX

PHCs also limits the generalizability of its findings to the broader population, as women facing logistical, financial, or social barriers to accessing care were likely excluded, introducing another layer of selection bias.

Lastly, although the intervention aimed to reduce inequities by targeting underprivileged populations and incorporating AI-based tools for HCP training, reliance on technology may have inadvertently introduced disparities. Variations in access to smartphones or digital literacy levels among participants could have affected the intervention's reach and impact. These limitations highlight the need to interpret the findings with caution and emphasize the importance of further research to validate and refine the intervention for broader and more equitable.

Conclusion

In conclusion, this study highlights the effectiveness of the AI-based gamified mHealth intervention GAIN MHI in enhancing the uptake of ANC and improving maternal and neonatal outcomes. Although encouraging results were observed, the study highlights the need to couple the implementation of such interventions with other complementary interventions targeted toward addressing socioeconomic and context-specific factors that may contribute to the achievement of improved maternal and neonatal health outcomes among disadvantaged populations.

Acknowledgments

The authors would like to thank the staff and directors of the primary health care centers of the Lebanese Ministry of Public Health and the United Nations Relief and Works Agency for Palestine refugees in the Near East for their contribution to data collection and participation in the Gamification and Artificial Intelligence and mHealth Network for Maternal Health Improvement (GAIN MHI) intervention and the GAIN MHI app developer at the Global Health Institute at the American University of Beirut for designing and developing the mobile app. This project was funded by the International Development Research Centre.

Conflicts of Interest

None declared.

References

1. Benage M, Greenough PG, Vinck P, Omeira N, Pham P. An assessment of antenatal care among Syrian refugees in Lebanon. Confl Health 2015;9:8 [FREE Full text] [doi: 10.1186/s13031-015-0035-8] [Medline: 25741381]
2. Iyakaremye I, Mukagatare C. Forced migration and sexual abuse: experience of Congolese adolescent girls in Kigeme refugee camp, Rwanda. Health Psychol Rep 2016 May 11;4(3):261-271. [doi: 10.5114/hpr.2016.59590]
3. Samari G. Syrian refugee women's health in Lebanon, Turkey, and Jordan and recommendations for improved practice. World Med Health Policy 2017 Jun;9(2):255-274 [FREE Full text] [doi: 10.1002/wmh3.231] [Medline: 29051840]
4. Singh N, Aryasinghe, Smith J, Khosla R, Say L, Blanchet K. A long way to go: a systematic review to assess the utilisation of sexual and reproductive health services during humanitarian crises. BMJ Glob Health 2018;3(2):e000682 [FREE Full text] [doi: 10.1136/bmjgh-2017-000682] [Medline: 29736272]
5. Makhoul G, Falakha G, Makhoul CN, AbdelAhad A. International perspectives: impact of Syrian refugees on neonatal care in Hopital Notre Dame de la Paix, Akkar, North Lebanon. NeoReviews 2015;16(8):e456-e461. [doi: 10.1542/neo.16-8-e456]
6. Nabulsi D, Abou Saad M, Ismail H, Doumit MAA, El-Jamil F, Kobeissi L, et al. Minimum initial service package (MISP) for sexual and reproductive health for women in a displacement setting: a narrative review on the Syrian refugee crisis in Lebanon. Reprod Health 2021 Mar 08;18(1):58 [FREE Full text] [doi: 10.1186/s12978-021-01108-9] [Medline: 33685476]
7. Usta J, Masterson AR. Women and health in refugee settings: the case of displaced Syrian women in Lebanon. In: Djamba YK, Kimuna SR, editors. Gender-Based Violence: Perspectives from Africa, the Middle East, and India. Cham: Springer International Publishing; 2015:119-143.
8. Moolla A, Mdewa W, Erzse A, Hofman K, Thsehla E, Goldstein S, et al. A cost-effectiveness analysis of a South African pregnancy support grant. PLOS Glob Public Health 2024;4(2):e0002781. [doi: 10.1371/journal.pgph.0002781] [Medline: 38329926]
9. Raatikainen K, Heiskanen N, Heinonen S. Under-attending free antenatal care is associated with adverse pregnancy outcomes. BMC Public Health 2007 Sep 27;7:268-268 [FREE Full text] [doi: 10.1186/1471-2458-7-268] [Medline: 17900359]
10. Standards for maternal and neonatal care. World Health Organization. 2007 Jan 15. URL: https://www.who.int/publications/i/item/standards-for-maternal-and-neonatal-care [accessed 2025-07-10]
11. Essaid A, Shukri S, El Gharaibeh Y, Abu Taleb H, Awwad N, Nour H, et al. Gender Based Violence against Women and Girls Displaced by the Syrian Conflict in South Lebanon and North Jordan: Scope of Violence and Health Correlates. Madrid: Alianza por la Solidaridad; 2015.
12. Huster KMJ, Patterson N, Schilperoord M, Spiegel P. Cesarean sections among Syrian refugees in Lebanon from December 2012/January 2013 to June 2013: probable causes and recommendations. Yale J Biol Med 2014 Sep;87(3):269-288 [FREE Full text] [Medline: 25191143]

XSL•FO
RenderX

13. Abdin L. Challenges for pregnant Syrian refugees in Lebanon. East Mediterr Health J 2018 Dec 18;24(10):1026-1029 [FREE Full text] [doi: 10.26719/2018.24.10.1026] [Medline: 30582146]
14. Mourtada R, Schlecht J, DeJong J. A qualitative study exploring child marriage practices among Syrian conflict-affected populations in Lebanon. Confl Health 2017;11(Suppl 1):27-65 [FREE Full text] [doi: 10.1186/s13031-017-0131-z] [Medline: 29167699]
15. Health UNHCR monthly update Lebanon. UNHCR. 2013. URL: https://data.unhcr.org/en/search?text=Health+UNHCR+monthly+update+Lebanon [accessed 2025-07-10]
16. Rizkianti A, Saptarini I, Rachmalina R. Perceived barriers in accessing health care and the risk of pregnancy complications in Indonesia. IJWH 2021 Aug;13:761-772. [doi: 10.2147/ijwh.s310850]
17. Gogoi M, Unisa S, Prusty RK. Utilization of maternal health care services and reproductive health complications in Assam, India. J Public Health 2014 Apr 2;22(4):351-359. [doi: 10.1007/s10389-014-0614-y]
18. Fact sheet. Lebanon. UNHCR. 2023 Dec. URL: https://www.unhcr.org/lb/sites/lb/files/legacy-pdf/UNHCR-Lebanon-FactSheet_December-2023_Final.pdf [accessed 2025-07-10]
19. Knudsen AJ. Syria's refugees in Lebanon: brothers, burden, and bone of contention. In: Di Peri R, Meier D, editors. Lebanon Facing the Arab Uprisings: Constraints and Adaptation. London, UK: Palgrave Macmillan; 2017:135-154.
20. Where we work. UNRWA. 2023. URL: https://www.unrwa.org/where-we-work [accessed 2025-07-10]
21. Khattar G, Hallit J, El Chamieh C, Bou Sanayeh E. Cardiovascular drug shortages in Lebanon: a broken heart. Health Econ Rev 2022 Apr 11;12(1):24 [FREE Full text] [doi: 10.1186/s13561-022-00369-9] [Medline: 35403980]
22. Bou Sanayeh E, El Chamieh C. The fragile healthcare system in Lebanon: sounding the alarm about its possible collapse. Health Econ Rev 2023 Apr 04;13(1):21 [FREE Full text] [doi: 10.1186/s13561-023-00435-w] [Medline: 37014485]
23. Dumit N, Honein-AbouHaidar G. The impact of the Syrian refugee crisis on nurses and the healthcare system in Lebanon: a qualitative exploratory study. J Nurs Scholarsh 2019 May;51(3):289-298. [doi: 10.1111/jnu.12479] [Medline: 30919555]
24. Devi S. Economic crisis hits Lebanese health care. Lancet 2020 Feb;395(10224):548. [doi: 10.1016/s0140-6736(20)30407-4]
25. Nemr E, Moussallem M, Nemr R, Kosremelli Asmar M. Exodus of Lebanese doctors in times of crisis: a qualitative study. Front Health Serv 2023;3:1240052 [FREE Full text] [doi: 10.3389/frhs.2023.1240052] [Medline: 38028945]
26. Bou Sanayeh E, El Chamieh C, Saade MC, Maalouf RG, Bizri M. Post-traumatic stress symptoms experienced by healthcare workers in Lebanon four months following Beirut's ammonium nitrate explosion: a survey-based study. Arch Public Health 2022 Jun 17;80(1):156 [FREE Full text] [doi: 10.1186/s13690-022-00911-5] [Medline: 35715820]
27. Daigle M, Spencer A, Diab JL, Samneh B, Afandi A. Sex, health and rights in displacement and humanitarian response: crises upon crises in Lebanon and beyond. ODI Global. 2023 Oct 5. URL: https://odi.org/en/publications/sex-health-and-rights-in-displacement-and-humanitarian-response-crises-upon-crises-in-lebanon-and-beyond/ [accessed 2025-07-10]
28. AlArab N, Nabulsi D, El Arnaout N, Dimassi H, Harb R, Lahoud J, et al. Reproductive health of Syrian refugee women in Lebanon: a descriptive analysis of the Sijilli electronic health records database. BMC Womens Health 2023 Feb 23;23(1):81 [FREE Full text] [doi: 10.1186/s12905-023-02231-4] [Medline: 36823589]
29. El Arnaout N, Rutherford S, Zreik T, Nabulsi D, Yassin N, Saleh S. Assessment of the health needs of Syrian refugees in Lebanon and Syria's neighboring countries. Confl Health 2019;13:31 [FREE Full text] [doi: 10.1186/s13031-019-0211-3] [Medline: 31297141]
30. Health in Lebanon. UNRWA. URL: https://www.unrwa.org/activity/health-lebanon [accessed 2025-07-10]
31. Hemadeh R, Hammoud R, Kdouh O. Lebanon's essential health care benefit package: a gateway for universal health coverage. Int J Health Plann Manag 2019 Oct;34(4):e1921-e1936. [doi: 10.1002/hpm.2850] [Medline: 31271234]
32. The Ministry of Public Health. Ministry of Public Health (MOPH). URL: https://moph.gov.lb/en/Pages/9/1024/the-ministry#/en/Pages/9/1024/the-ministry [accessed 2025-07-10]
33. MOPH. Primary Health Care Department annual dashboard - 2018. Ministry of Public Health (MOPH). 2018. URL: https://www.moph.gov.lb/userfiles/files/HealthCareSystem/PHC/PHC-Dashboard%20-2018.pdf [accessed 2025-07-10]
34. Parkinson S, Behrouzan O. Negotiating health and life: Syrian refugees and the politics of access in Lebanon. Soc Sci Med 2015 Dec;146:324-331. [doi: 10.1016/j.socscimed.2015.10.008] [Medline: 26477853]
35. Abdulrahim S, El Rafei R, Beydoun Z, El Hayek GY, Nakad P, Yunis K. A test of the epidemiological paradox in a context of forced migration: low birthweight among Syrian newborns in Lebanon. Int J Epidemiol 2019 Feb 01;48(1):275-286. [doi: 10.1093/ije/dyy200] [Medline: 30357348]
36. Reproductive health services. Ministry of Public Health (MOPH). 2024. URL: https://www.moph.gov.lb/en/Pages/6/758/reproductive-health-services [accessed 2025-07-10]
37. Department of Internal Oversight Services 2022 annual report. UNRWA. 2022. URL: https://www.unrwa.org/sites/default/files/content/resources/annual_report_dios_2022_adcom_11062023.pdf [accessed 2025-07-10]
38. EU/IfS project. Conflict reduction through improving healthcare services for the vulnerable population in Lebanon. Ministry of Public Health (MOPH). URL: https://www.moph.gov.lb/en/Pages/6/2918/eu-ifs-project [accessed 2025-07-10]
39. Lebanon Health Resilience Project (P163476): second semi-annual progress report. Ministry of Public Health (MOPH). 2024 Apr 28. URL: https://www.moph.gov.lb/userfiles/files/Programs%26Projects/Lebanon%20Health%20Resilience%20Project%20-LHR/Second-Semi-Annual-Progress-Report.pdf [accessed 2025-07-10]

XSL•FO
RenderX

40. EU biggest donor to the Lebanese health sector. European Union. 2018 Apr 28. URL: https://www.eeas.europa.eu/node/41772_en [accessed 2025-07-10]
41. Our vision/mission. Ministry of Public Health (MOPH). 2024 May 15. URL: https://moph.gov.lb/en/Pages/9/1024/the-ministry#/en/Pages/9/1/our-vision-mission [accessed 2025-07-10]
42. EU and UNFPA step up support for vulnerable women in Lebanon during crisis. UNFPA. 2023 Mar 24. URL: https://tinyurl.com/25f8ywuv [accessed 2025-07-10]
43. Rahman M, Yamaji N, Nagamatsu Y, Ota E. Effects of mHealth interventions on improving antenatal care visits and skilled delivery care in low- and middle-income countries: systematic review and meta-analysis. J Med Internet Res 2022 Apr 22;24(4):e34061 [FREE Full text] [doi: 10.2196/34061] [Medline: 35451987]
44. Fedha T. Impact of mobile telephone on maternal health service care: a case of Njoro division. Open J Prev Med 2014 May;4(5):365-376. [doi: 10.4236/ojpm.2014.45044]
45. Higgs E, Goldberg AB, Labrique AB, Cook SH, Schmid C, Cole CF, et al. Understanding the role of mHealth and other media interventions for behavior change to enhance child survival and development in low- and middle-income countries: an evidence review. J Health Commun 2014;19 Suppl 1(sup1):164-189 [FREE Full text] [doi: 10.1080/10810730.2014.929763] [Medline: 25207452]
46. Miah S, Gammack J, Hasan N. Extending the framework for mobile health information systems research: a content analysis. Inf Syst 2017 Sep;69:1-24. [doi: 10.1016/j.is.2017.04.001]
47. Okolo CA, Babawarun O, Olorunsogo TO. Mobile health (mhealth) innovations for public health feedback: a global perspective. Int Med Sci Res J 2024;4(3):235-246. [doi: 10.51594/imsrj.v4i3.915]
48. De P, Pradhan MR. Effectiveness of mobile technology and utilization of maternal and neonatal healthcare in low and middle-income countries (LMICs): a systematic review. BMC Womens Health 2023 Dec 11;23(1):664 [FREE Full text] [doi: 10.1186/s12905-023-02825-y] [Medline: 38082424]
49. WHO Team. Trends in maternal mortality 2000 to 2020: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division. World Health Organization. 2023 Feb 23. URL: https://www.who.int/publications/i/item/9789240068759 [accessed 2025-07-10]
50. Barron P, Peter J, LeFevre AE, Sebidi J, Bekker M, Allen R, et al. Mobile health messaging service and helpdesk for South African mothers (MomConnect): history, successes and challenges. BMJ Glob Health 2018;3(Suppl 2):e000559 [FREE Full text] [doi: 10.1136/bmjgh-2017-000559] [Medline: 29713503]
51. Lebrun V, Dulli L, Alami SO, Sidiqi A, Sultani AS, Rastagar SH, et al. Feasibility and acceptability of an adapted mobile phone message program and changes in maternal and newborn health knowledge in four provinces of Afghanistan: single-group pre-post assessment study. JMIR Mhealth Uhealth 2020 Jul 20;8(7):e17535 [FREE Full text] [doi: 10.2196/17535] [Medline: 32706690]
52. Maliwichi P, Chigona W, Sowon K. Appropriation of mHealth interventions for maternal health care in sub-Saharan Africa: hermeneutic review. JMIR Mhealth Uhealth 2021 Oct 06;9(10):e22653 [FREE Full text] [doi: 10.2196/22653] [Medline: 34612835]
53. Shaheen MY. Applications of artificial intelligence (AI) in healthcare: a review. ScienceOpen. Preprint posted online 2021. [doi:10.14293/S2199-1006.1.SOR-.PPVRY8K.v1] . [doi: 10.14293/s2199-1006.1.sor-.ppvry8k.v1]
54. Saif-Ur-Rahman K, Islam MS, Alaboson J, Ola O, Hasan I, Islam N, et al. Artificial intelligence and digital health in improving primary health care service delivery in LMICs: a systematic review. J Evid Based Med 2023 Sep;16(3):303-320. [doi: 10.1111/jebm.12547] [Medline: 37691394]
55. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 2018;3(4):e000798 [FREE Full text] [doi: 10.1136/bmjgh-2018-000798] [Medline: 30233828]
56. Khan M, Khurshid M, Vatsa M, Singh R, Duggal M, Singh K. On AI approaches for promoting maternal and neonatal health in low resource settings: a review. Front Public Health 2022 Sep 30;10:880034 [FREE Full text] [doi: 10.3389/fpubh.2022.880034] [Medline: 36249249]
57. Rittenhouse K, Vwalika B, Keil A, Winston J, Stoner M, Price JT, et al. Improving preterm newborn identification in low-resource settings with machine learning. PLoS One 2019;14(2):e0198919 [FREE Full text] [doi: 10.1371/journal.pone.0198919] [Medline: 30811399]
58. Panda PK, Sharma R. Transforming maternal healthcare: harnessing the power of artificial intelligence for improved outcomes and access. World J Adv Res Rev 2024;23(01):662-666. [doi: 10.30574/wjarr.2024.23.1.2005]
59. Orjuela MAV, Uribe-Quevedo A, Jaimes N, Perez-Gutierrez B. External automatic defibrillator game-based learning app. 2015 Presented at: 2015 IEEE Games Entertainment Media Conference (GEM); October 14-16, 2015; Toronto, ON, Canada. [doi: 10.1109/gem.2015.7377206]
60. Guana V, Xiang T, Zhang H, Schepens E, Stroulia E. UnderControl an educational serious-game for reproductive health. 2014 Presented at: CHI PLAY '14: First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play; October 19-21, 2014; Toronto, ON, Canada. [doi: 10.1145/2658537.2662983]
61. Molnar A, Kostkova P. Learning about hygiene and antibiotic resistance through mobile games: evaluation of learning effectiveness. In: Kostkova, Learning about Hygiene and Antibiotic Resistance through Mobile Gamesvaluation of Learning

XSL•FO
RenderX

Effectiveness, in Proceedings of the 2018 International Conference on Digital Health. 2018 Presented at: DH '18: 2018 International Conference on Digital Health; April 23-26, 2018; Lyon, France. [doi: 10.1145/3194658.3194682]
62. Miller A, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Inform J 2016 Jun;22(2):184-193 [FREE Full text] [doi: 10.1177/1460458214537511] [Medline: 24986104]
63. Kayastha R, Mueller S, Yadav P, Kelman I, Boscor A, Saville N, et al. Do women in Nepal like playing a mobile game? MANTRA: a mobile gamified app for improving healthcare seeking behavior in rural Nepal. Front Public Health 2021;9:645837 [FREE Full text] [doi: 10.3389/fpubh.2021.645837] [Medline: 34805057]
64. Talhouk R, Akik C, Araujo-Soares V, Ahmad B, Mesmar S, Olivier P, et al. Integrating health technologies in health services for Syrian refugees in Lebanon: qualitative study. J Med Internet Res 2020 Jul 06;22(7):e14283 [FREE Full text] [doi: 10.2196/14283] [Medline: 32628121]
65. Yassin N, Khodor R, Baroud M. m-Health for healthcare delivery reform: prospects for Lebanese and refugee communities. In: Healthcare Systems, 1st Edition. Boca Raton, FL: CRC Press; 2018:311.
66. Braithwaite J, Mannion R, Matsuyama Y, Shekelle PG, Whittaker S, Al-Adawi S, et al. The future of health systems to 2030: a roadmap for global progress and sustainability. Int J Qual Health Care 2018 Dec 01;30(10):823-831. [doi: 10.1093/intqhc/mzy242] [Medline: 30576556]
67. Doocy S, Paik KE, Lyles E, Hei Tam H, Fahed Z, Winkler E, et al. Guidelines and mHealth to improve quality of hypertension and type 2 diabetes care for vulnerable populations in Lebanon: longitudinal cohort study. JMIR Mhealth Uhealth 2017 Oct 18;5(10):e158 [FREE Full text] [doi: 10.2196/mhealth.7745] [Medline: 29046266]
68. Saleh S, Farah A, Dimassi H, El Arnaout N, Constantin J, Osman M, et al. Using mobile health to enhance outcomes of noncommunicable diseases care in rural settings and refugee camps: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 13;6(7):e137 [FREE Full text] [doi: 10.2196/mhealth.8146] [Medline: 30006326]
69. Lakkis N, Atfeh AMA, El-Zein YR, Mahmassani DM, Hamadeh GN. The effect of two types of SMS-texts on the uptake of screening mammogram: a randomized controlled trial. Prev Med 2011 Oct;53(4-5):325-327. [doi: 10.1016/j.ypmed.2011.08.013] [Medline: 21871480]
70. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: user-centered study with employees from a Lebanese university. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e9836 [FREE Full text] [doi: 10.2196/mhealth.9836] [Medline: 30672742]
71. Seita A, Ballout G, Albeik S, Salameh Z, Zeidan W, Shah S, et al. Leveraging digital health data to transform the United Nations Systems for Palestine Refugees for the post pandemic time. Health Syst Reform 2024 Dec 17;10(2):2378505 [FREE Full text] [doi: 10.1080/23288604.2024.2378505] [Medline: 39437235]
72. Alsswey A, Al-Samarraie H, Bervell B. mHealth technology utilization in the Arab world : a systematic review of systems, usage, and challenges. Health Technol 2021;11(4):895-907. [doi: 10.1007/s12553-021-00549-3]
73. Qasrawi R, Amro M, VicunaPolo S, Abu Al-Halawa D, Agha H, Abu Seir R, et al. Machine learning techniques for predicting depression and anxiety in pregnant and postpartum women during the COVID-19 pandemic: a cross-sectional regional study. F1000Res 2022;11:390 [FREE Full text] [doi: 10.12688/f1000research.110090.1] [Medline: 36111217]
74. El-Jardali F, Alameddine M, Jamal D, Dimassi H, Dumit NY, McEwen MK, et al. A national study on nurses' retention in healthcare facilities in underserved areas in Lebanon. Hum Resour Health 2013 Sep 30;11(1):49 [FREE Full text] [doi: 10.1186/1478-4491-11-49] [Medline: 24079458]
75. Jaana M, Majdalani M, Tamim H, Rahbany R. Perceived healthcare workforce needs in Lebanon: a step towards informed human resources planning and professional development. East Mediterr Health J 2018 Dec 09;24(9):855-865 [FREE Full text] [doi: 10.26719/2018.24.9.855] [Medline: 30570118]
76. Endehabtu B, Weldeab A, Were M, Lester R, Worku A, Tilahun B. Mobile phone access and willingness among mothers to receive a text-based mHealth intervention to improve prenatal care in northwest Ethiopia: cross-sectional study. JMIR Pediatr Parent 2018 Oct 17;1(2):e9 [FREE Full text] [doi: 10.2196/pediatrics.9618] [Medline: 31518334]
77. Murthy N, Chandrasekharan S, Prakash MP, Ganju A, Peter J, Kaonga N, et al. Effects of an mHealth voice message service (mMitra) on maternal health knowledge and practices of low-income women in India: findings from a pseudo-randomized controlled trial. BMC Public Health 2020 Jun 01;20(1):820 [FREE Full text] [doi: 10.1186/s12889-020-08965-2] [Medline: 32487065]
78. Bossman E, Johansen MA, Zanaboni P. mHealth interventions to reduce maternal and child mortality in sub-Saharan Africa and southern Asia: a systematic literature review. Front Glob Womens Health 2022;3:942146 [FREE Full text] [doi: 10.3389/fgwh.2022.942146] [Medline: 36090599]
79. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
80. Gurol‐Urganci I. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013;2013(12):CD007458. [doi: 10.1002/14651858.cd007458.pub3]

XSL•FO
RenderX

81. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012 Apr;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]
82. Alam M, D'Este C, Banwell C, Lokuge K. The impact of mobile phone based messages on maternal and child healthcare behaviour: a retrospective cross-sectional survey in Bangladesh. BMC Health Serv Res 2017 Jun 24;17(1):434 [FREE Full text] [doi: 10.1186/s12913-017-2361-6] [Medline: 28645278]
83. Opon S, Tenambergen WM, Njoroge KM. The effect of patient reminders in reducing missed appointment in medical settings: a systematic review. PAMJ-OH 2020;2:2020. [doi: 10.11604/pamj-oh.2020.2.9.21839]
84. Chowdhury M, Shiblee SI, Jones HE. Does mHealth voice messaging work for improving knowledge and practice of maternal and newborn healthcare? BMC Med Inform Decis Mak 2019 Sep 05;19(1):179 [FREE Full text] [doi: 10.1186/s12911-019-0903-z] [Medline: 31488159]
85. New report: Decline of UNRWA services, consequences will be catastrophic. Euro-Med Human Rights Monitor. 2018 Mar 4. URL: https://euromedmonitor.org/en/article/2279/New-Report:-Decline-of-UNRWA-services,-consequences-will-be-catastrophic [accessed 2025-07-10]
86. Gillespie B. Nutritional status and the risk of preterm birth. In: Anumba DOC, Jayasooriya SM, editors. Evidence Based Global Health Manual for Preterm Birth Risk Assessment. Cham: Springer International Publishing; 2022:41-51.
87. Li B, Zhang X, Peng X, Zhang S, Wang X, Zhu C. Folic acid and risk of preterm birth: a meta-analysis. Front Neurosci 2019;13:1284 [FREE Full text] [doi: 10.3389/fnins.2019.01284] [Medline: 31849592]
88. Keag O, Norman JE, Stock SJ. Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: systematic review and meta-analysis. PLoS Med 2018 Jan;15(1):e1002494 [FREE Full text] [doi: 10.1371/journal.pmed.1002494] [Medline: 29360829]
89. Sandall J, Tribe RM, Avery L, Mola G, Visser GH, Homer CS, et al. Short-term and long-term effects of caesarean section on the health of women and children. Lancet 2018 Oct;392(10155):1349-1357. [doi: 10.1016/s0140-6736(18)31930-5]
90. Sobhy S, Arroyo-Manzano D, Murugesu N, Karthikeyan G, Kumar V, Kaur I, et al. Maternal and perinatal mortality and complications associated with caesarean section in low-income and middle-income countries: a systematic review and meta-analysis. Lancet 2019 May;393(10184):1973-1982. [doi: 10.1016/s0140-6736(18)32386-9]
91. Caesarean section rates continue to rise, amid growing inequalities in access. World Health Organization. 2021. URL: https://www.who.int/news/item/16-06-2021-caesarean-section-rates-continue-to-rise-amid-growing-inequalities-in-access [accessed 2025-07-10]
92. Janaki JS, Soubramanian P. Examining socioeconomic factors influencing maternal health in pregnancy. J Hum Behav Soc Environ 2024 Feb 07;35(4):469-487. [doi: 10.1080/10911359.2024.2310272]
93. Thomson K, Moffat M, Arisa O, Jesurasa A, Richmond C, Odeniyi A, et al. Socioeconomic inequalities and adverse pregnancy outcomes in the UK and Republic of Ireland: a systematic review and meta-analysis. BMJ Open 2021 Mar 15;11(3):e042753 [FREE Full text] [doi: 10.1136/bmjopen-2020-042753] [Medline: 33722867]
94. Bush J, Barlow DE, Echols J, Wilkerson J, Bellevin K. Impact of a mobile health application on user engagement and pregnancy outcomes among Wyoming Medicaid members. Telemed J E Health 2017 Nov;23(11):891-898 [FREE Full text] [doi: 10.1089/tmj.2016.0242] [Medline: 28481167]
95. Nishimwe A, Ibisomi L, Nyssen M, Conco DN. The effect of a decision-support mHealth application on maternal and neonatal outcomes in two district hospitals in Rwanda: pre - post intervention study. BMC Pregnancy Childbirth 2022 Jan 20;22(1):52 [FREE Full text] [doi: 10.1186/s12884-022-04393-9] [Medline: 35057761]

Abbreviations

AI: artificial Intelligence
ANC: antenatal care
C-section: cesarean section
EU: European Union
GAIN MHI: Gamification, Artificial Intelligence and mHealth Network for Maternal Health Improvement
HCP: health care provider
IRB: Institutional Review Board
KAP: knowledge, attitude, and practices
MENA: Middle East and North Africa
mHealth: mobile health
MOPH: Ministry of Public Health
NGO: nongovernmental organization
OBGYN: obstetricians and gynecologists
OR: odds ratio
PHC: primary health care center
PNC: postnatal care

XSL•FO
RenderX

UNRWA: United Nations Relief and Works Agency for Palestine Refugees in the Near East

Edited by L Buis; submitted 20.08.24; peer-reviewed by P Mechael, C Weerth; comments to author 20.12.24; revised version received 13.02.25; accepted 07.04.25; published 04.11.25.

Please cite as:
Saleh S, El Arnaout N, Sabra N, El Dakdouki A, Chamseddine Z, Hamadeh R, Shanaa A, Alameddine M
The Effectiveness of an Artificial Intelligence–Based Gamified Intervention for Improving Maternal Health Outcomes Among Refugees and Underserved Women in Lebanon: Community Interventional Trial
JMIR Mhealth Uhealth 2025;13:e65599
URL: https://mhealth.jmir.org/2025/1/e65599
doi:10.2196/65599
PMID:41186980

©Shadi Saleh, Nour El Arnaout, Nadine Sabra, Asmaa El Dakdouki, Zahraa Chamseddine, Randa Hamadeh, Abed Shanaa, Mohamad Alameddine. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 04.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Impact of an Adaptive mHealth Intervention on Improving Patient-Provider Health Care Communication: Secondary Analysis of the DIAMANTE Trial

Lynn Leng[1,2], MS; Marvyn R Arévalo Avalos[3], PhD; Adrian Aguilera[3,4], PhD; Courtney R Lyles[5], PhD

[1]School of Public Health, UC Berkeley-UCSF Joint Medical Program, University of California, Berkeley, 1995 University Avenue, Suite 300, Berkeley, CA, United States

[2]School of Medicine, University of California, San Francisco, San Francisco, CA, United States

[3]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[4]Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, United States

[5]Department of Public Health Sciences, Center for Healthcare Policy and Research, University of California, Davis, Davis, CA, United States

Corresponding Author:
Lynn Leng, MS
School of Public Health, UC Berkeley-UCSF Joint Medical Program, University of California, Berkeley, 1995 University Avenue, Suite 300, Berkeley, CA, United States

Abstract

Background: Depression and diabetes are highly comorbid conditions and are significant causes of global disability, particularly among individuals with low income or those from racial or ethnic minority backgrounds. While digital interventions offer promise for managing these chronic conditions (such as via lifestyle modification), there is also emerging evidence suggesting that digital support may strengthen or complement existing health care relationships, particularly by improving patient perceptions of communication and connection with their health care providers.

Objective: This study aimed to examine the impact of an adaptive mobile health (mHealth) texting-based intervention on patient ratings of communication with their health care providers among individuals with diabetes and depressive symptoms.

Methods: This study used data from the DIAMANTE (Diabetes and Mental Health Adaptive Notification Tracking and Evaluation) trial, a prospective, randomized controlled trial that varied SMS text messaging strategies to encourage physical activity to support both mental and physical health for patients with diabetes and depression. Patients were recruited from safety-net clinics in San Francisco and through social media during the COVID-19 pandemic, and were randomized into three trial arms: (1) personalized SMS text messaging about physical activity via an adaptive learning algorithm, (2) randomly selected SMS text messaging about physical activity, and (3) a control group that received no SMS text messages. As a secondary outcome, we examined pre-post changes in patient-reported health care communication, assessed via surveys with the validated Consumer Assessment of Healthcare Providers and Systems (CAHPS) communication subscale. Bivariate comparisons examined changes in CAHPS scores, including age, gender, preferred language, race or ethnicity, nativity, marital status, and education. Our primary analysis used mixed-effects modeling within an intent-to-treat analysis to determine differences in CAHPS scores by trial arm.

Results: A total of 195 patients participated in the DIAMANTE trial from 2020 to 2022. After excluding patients who had incomplete or missing baseline or exit surveys, 156 patients were included in this secondary analysis. Overall, there was a substantive but nonsignificant decrease in the average CAHPS score over the 6-month trial period (−2.6; P=.11), with similar trends across patient demographic subgroups. Upon evaluating health care communication across the three randomized controlled trial (RCT) arms, there were no significant differences in patient-provider communication.

Conclusions: Digital health interventions are rapidly increasing in clinical practice and have the ability to reach historically underserved communities by overcoming barriers such as language, geography, and time constraints. However, RCTs have primarily focused on the clinical impact of these tools, rather than how the tools may impact patients' perceptions of their relationships with providers. While our study found no significant changes in patient-provider communication by RCT arm, the temporal implications of the COVID-19 pandemic during the DIAMANTE study period remain unclear and should be further studied. This study is the first of its kind to examine the influence of an adaptive mHealth intervention on patient-reported health care communication, with insights that may be important as digitally-enabled chronic care management rapidly expands.

Trial Registration: ClinicalTrials.gov NCT03490253; https://clinicaltrials.gov/study/NCT03490253

(JMIR Mhealth Uhealth 2025;13:e64296) doi:10.2196/64296

XSL•FO
RenderX

KEYWORDS

mobile health; mhealth; chronic care management; patient engagement; patient-physician relationship; underserved; depression; diabetes; comorbidity; health literacy; self-management; application; app

Introduction

Treatment and care for depression and diabetes, two highly prevalent and often co-occurring conditions [1,2], are often siloed. Individuals with diabetes have an increased risk of developing comorbid depression compared to individuals without diabetes [3]. This comorbidity is often associated with a worse prognosis for both conditions [4], occurs at higher rates, and results in worse outcomes, particularly in marginalized communities, including individuals with low income, limited health literacy, or from racial or ethnic minority groups [5,6]. Management of chronic diseases such as diabetes requires self-management practices such as maintaining adequate physical activity and adhering to medications [6]. One important factor that affects self-management practices is positive patient-provider communication [7], which has been documented in many studies to increase trust and improve health outcomes among patients with chronic illness [8].

Mobile health (mHealth) interventions have the potential to effectively integrate with existing chronic health management plans. mHealth is defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices [9]." The advent of cost-effective mHealth apps has been effective in helping patients engage in healthy behaviors, such as increased physical activity and lowered HbA_{1c} levels [10]. For instance, a diabetes SMS text messaging program improved HbA_{1c} scores as well as provided emotional support for patients [11]. Further, the COVID-19 pandemic has expedited the use of telehealth resources and SMS text messaging through secure platforms via electronic devices, which has improved access to health care [12].

The rapidly increasing number of chronic disease mHealth apps indicates the demand for self-management tools [7]; however, there must be a consideration of digital equity in terms of individuals' internet access, skills, and support related to these platforms and tools [13]. For example, mAdherence tools (ie, mobile health adherence tools) focusing on individuals with low income as well as those who are elderly and from racial or ethnic minority backgrounds were found to be effective and with high satisfaction ratings, although these required a strong baseline adherence program to begin with, as well as active encouragement from the clinical team [14]. Similarly, mHealth tools can be tailored for patients with non-English language preferences; with the addition of culturally concordant motivational messaging, mHealth tools have the potential to reduce health care disparities [15].

Recently, there has also been a surge in mHealth apps integrating artificial intelligence to address chronic health management needs and equity concerns via more personalized communication, such as adaptive learning which tailors messaging in response to an individual's behavior and characteristics [15,16]. Importantly, personalization of messaging could also reinforce and extend the patient-provider relationship that was already established during a face-to-face interaction. Previous observational studies have found that digital communication and remote monitoring could improve the quality and quantity of existing patient-provider relationships, leading to 'digital intimacy,' characterized by familiarity made possible by electronic devices in addition to in-person encounters [17]. This was true even when participants were aware that the digital communication was computer- or digitally-generated [13,17]. While this body of literature is growing, there is a lack of randomized trial evidence about the impact of these interventions on patient-provider communication outcomes.

Therefore, we sought to examine the impact of a personalized communication intervention on participants' ratings of provider communication during the study period. Specifically, this study is a secondary analysis of an existing RCT evaluating a digital health physical activity intervention using reinforcement learning algorithms among adult patients with depression and diabetes.

Methods

Ethical Considerations

This study was reviewed and approved by the Institutional Review Board (IRB) of the University of California, San Francisco (IRB #17-22608). All procedures performed in studies involving human participants were conducted in accordance with the ethical standards of the institutional and/or national research committee and with the WMA Declaration of Helsinki. Informed consent was obtained from all individual participants included in the study. Participants were provided US $110 for completion of the study and all data was deidentified. Further details of the study are described by Aguilera et al [18].

Study Design

The DIAMANTE (Diabetes and Mental Health Adaptive Notification Tracking and Evaluation) was a prospective, IRB-approved, RCT evaluating the effectiveness of a machine-learning-guided smartphone intervention on physical activity. In brief, the intervention sent SMS text messages to participants during a 6-month period to encourage physical activity to support both mental and physical health [19]. There were three arms of the trial: (1) the control arm received weekly messages to report their mood; (2) the random messaging arm received two additional daily text messages that were randomly selected from a message bank that provided feedback on step counts as well as a motivational message related to being physically active, and (3) the adaptive messaging arm used a reinforcement learning algorithm to personalize these daily feedback and motivational messages for each participant based on their demographics as well as their recent data (eg, their steps taken on the previous day, time elapsed since the previous message type that was sent to them) [19].

XSL•FO
RenderX

Development of the SMS text messaging system was designed using a behavioral theory, and the iterative design was completed with patients in a safety-net health care setting as well as via Amazon Mechanical Turk (MTurk) crowdsourcing [20]. For example, (1) the feedback messages were created to provide positive (eg, "Great job!"), neutral (eg, "You walked 4000 steps yesterday"), and negative (eg, "You can do better!") feedback and 2) the motivational messages were categorized within theoretical categories of capability (ie, knowledge and skills to be active), opportunity (ie, advice on strategies to be active, both independently and with others), and motivation (ie, cognitive messages about beliefs and emotions that impact health behaviors). The final DIAMANTE intervention tracked participant step counts from pedometers on smartphones, via an application designed in English and Spanish for both Apple App Store and Android Google Play [19]. Further details about the DIAMANTE application design process have been published elsewhere, including details about the reinforcement learning algorithm and the main RCT findings [18].

Recruitment

English- and Spanish-speaking patients aged 18 - 75 years, diagnosed with diabetes and had documented depressive symptoms (score >5 on PHQ-8), and who had an iPhone or Android smartphone were eligible for participation. Exclusion criteria included the inability to exercise due to physical disability, active psychosis or mania, active suicidal ideation, severe cognitive impairment, inability to read and write in English and Spanish, current pregnancy, and plans to leave the country for extended periods during the 6-month trial.

From 2020 to 2022, patients were first recruited from safety-net clinics in the San Francisco Health Network, the public health care system for the city and county of San Francisco. Due to the COVID-19 pandemic, the study team then expanded recruitment to individuals on social media (ie, Craigslist, Facebook, and Google) who met the same eligibility criteria and prioritized major metropolitan areas with diabetes and larger proportions of Spanish-speaking residents [20]. All patients in the San Francisco Health Network were recruited via their existing primary care provider, and all patients recruited online self-attested that their clinician had diagnosed them with diabetes; both parameters facilitated the focus on patient-provider communication in this paper. Patients received a compensation of $40 for participation in the baseline study and an additional $70 for completion of the 6-month active intervention follow-up. Both of the time points included a detailed 50-item survey with numerous validated scales and self-reported data.

Measures

We examined the impact of the randomized intervention on pre-post changes in patient-reported health care communication via participant surveys at baseline and follow-up, based on the validated Consumer Assessment of Healthcare Providers and Systems (CAHPS) communication subscale [21]. The CAHPS score is widely used across outpatient practices to assess the quality of communication and care processes, and is a well-documented 4-item measure of how often providers explained things well, listened carefully, showed respect, and spent enough time with patients (responses ranged from 'never' to 'always,' coded from 1 to 4). Using conventional scoring, response items were linearly transformed to aid in interpretation (eg, scores of 1, 2, 3, and 4 correspond to 25, 50, 75, and 100, respectively) and averaged across questions, resulting in a total possible score from 25 to 100, with 25 being the lowest possible score, reflecting poor communication between the patient and provider. This scale is one of the most widely used measures of patient-provider communication in routine clinical practice in the United States [22].

Covariates examined descriptively in this study included (1) age (tertiles of ≤45, 46 - 55, or ≥56); (2) gender (ie, man, woman, transgender, and gender queer), (3) education (≤ high school, some college, or ≥ college graduate), (4) race or ethnicity (Asian/Pacific Islander, Black or African American, Hispanic/Latino, White, or Multiethnic), (5) marital status (single, separated, widowed, married, or partnered), (6) born in the United States (US nativity: yes or no), and (7) preferred language (English or Spanish). Covariates were chosen based on previous evidence of their potential impacts on communication outcomes, in core domains of socioeconomic status (eg, education), potential communication barriers (eg, language, nativity), social support (eg, marital status), and other demographics previously associated with communication scores (eg, gender, race or ethnicity).

Statistical Analysis

We first examined the unadjusted descriptive associations between various demographic measures and changes in CAHPS scores from baseline to exit surveys using paired two-tailed *t* tests. These bivariate comparisons examined changes in CAHPS score based on participant age, gender, preferred language, race or ethnicity, nativity, marital status, and education.

Further, our primary analysis examined the changes in CAHPS scores by RCT intervention arm. As the data were from a randomized controlled trial, we did not adjust for the covariates. We used mixed effects modeling in an intent-to-treat primary analysis to determine any significant differences in the communication outcome based on the intervention arm.

Results

t Test Findings

Between 2020 to 2022, 195 patients participated in the DIAMANTE trial. After taking account of participants who dropped out (either due to incomplete data or missing baseline or exit survey), 156 patients were included in the analysis. Participants' mean age was 48.3 (SD 12.2) years and 62.2% (97/195) were predominantly female. A total of 76.3% (n=119) participants were predominantly English-language preferring, were born in the United States (64.1%, n=100), and represented Asian/Pacific Islander (7.7%, n=12), Black or African American (17.3%, n=27), White (28.8%, n=45), Hispanic or Latino (36.5%, n=57), and Multiethnic (9.6%, n=15) racial and ethnic groups. A large portion of participants were single (44.2%, n=92) or married (27.6%, n=64), and had at least some college education and beyond (68.6%, n=107).

XSL•FO
RenderX

As shown in Table 1, there was an overall decrease in the CAHPS scores from baseline to follow-up for most demographic subgroups, indicating communication measures declined during the DIAMANTE trial, although these differences were not statistically significant. This included all categories of gender, age, English and Spanish language, race and ethnicity, US and non-US nativity status, marital status, and most educational attainment levels. A few patterns emerged as well when descriptively comparing groups; greater raw decreases in CAHPS scores were observed among Asian or Pacific Islander and Black or African American participants. In educational attainment levels and age groups, there were larger raw decreases in those who had some college education and less, and those who were younger, respectively.

Table . Baseline demographic and clinical characteristics.

Variables	Participants (N=156), n (%)	Baseline CAHPS[a] score (N=156)	Exit survey CAHPS score (N=156)	*P* value
Gender				
Man	55 (35.3)	85.5	84.4	.68
Woman	97 (62.2)	81.2	77.5	.11
Transgender and Gender Queer	4 (2.5)	68.8	68.8	≥.99
Age				
<45	62 (39.8)	80.7	76.1	.15
46 - 55	42 (26.9)	84.5	81.7	.31
>56	52 (33.3)	82.6	82.4	.96
Language				
English	119 (76.3)	80.7	77.9	.15
Spanish	37 (23.7)	87.7	85.6	.50
Race/Ethnicity				
Asian/Pacific Islander	12 (7.7)	79.7	71.9	.16
Black or African American	27 (17.3)	89.4	82.4	.13
White or Caucasian	45 (28.9)	76.4	74.7	.62
Hispanic/Latino	57 (36.5)	84.9	83.8	.68
Multiethnic	15 (9.6)	80.4	80.8	.93
Born in the United States				
Yes	100 (64.1)	80.4	77.5	.20
No	56 (35.9)	85.8	83.7	.33
Marital Status				
Single/Separated/Widowed	92 (59.0)	85.9	82.9	.15
Married/Partnered	64 (41.0)	77.8	75.4	.39
Education				
<High School	49 (31.4)	85.2	80.1	.13
Some College	40 (25.7)	86.3	81.7	.10
≥College Graduate	67 (42.9)	78.3	78.0	.91

[a]CAHPS: Consumer Assessment of Healthcare Providers and Systems.

Paired Samples Results

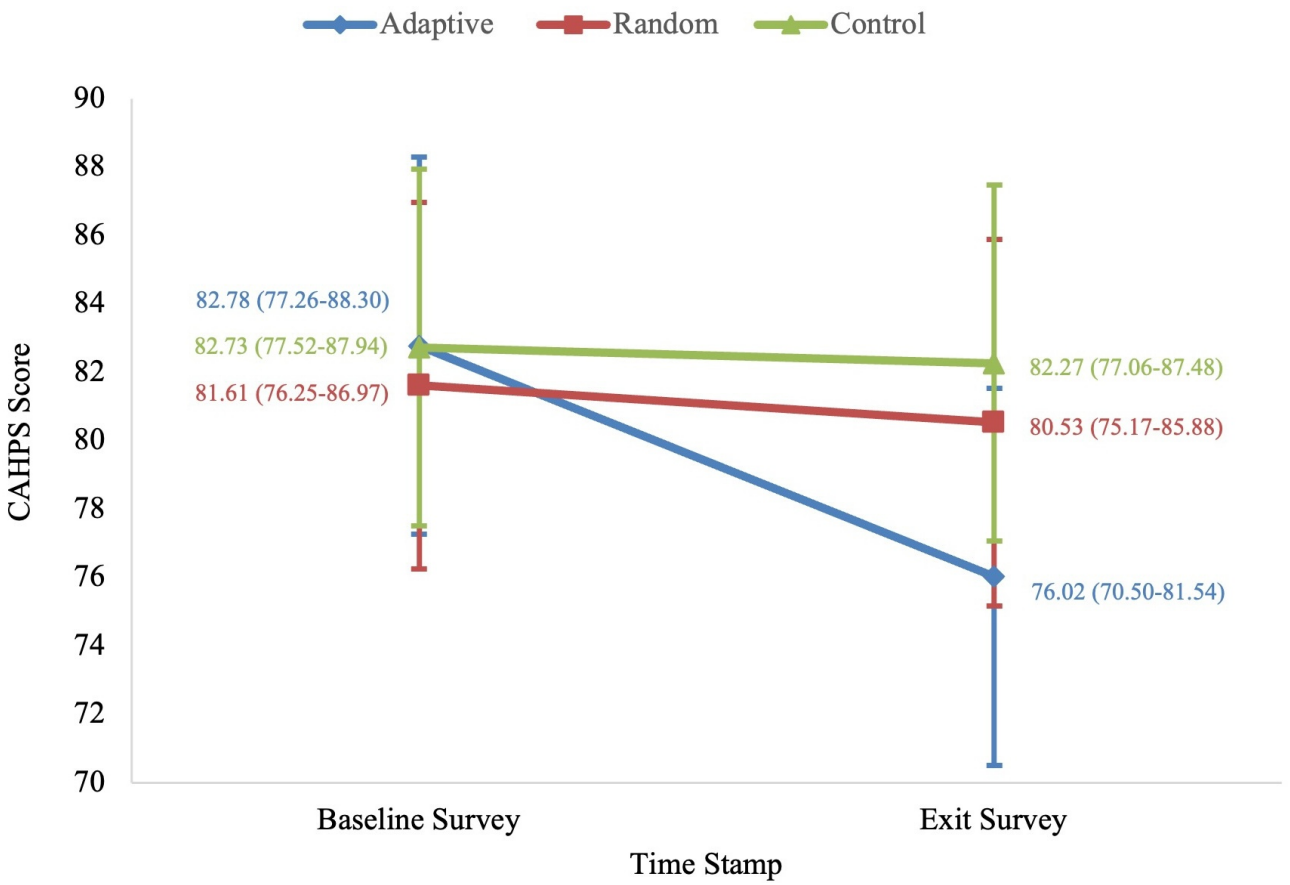
Overall, there was a marginal decrease in the average CAHPS score over the DIAMANTE clinical trial period of six months. A mean decrease of −2.6 points (95% CI −5.9 to 0.6; *P*=.11) was observed across the 156 patients who completed both the baseline and exit surveys, which was not statistically significant.

Mixed-Effects Modeling Results

Further, in the primary intent-to-treat analysis, there were no significant differences in CAHPS scores between the three intervention arms (Figure 1). While not statistically significant, the largest decrease in CAHPS score was seen in the adaptive arm (−6.3 95% CI −13.8 to 1.3; *P*=.11) compared to the control arm.

XSL•FO
RenderX

Figure 1. Mixed effects modeling to denote CAHPS score changes by intervention arms. CAHPS: Consumer Assessment of Healthcare Providers and Systems.



Discussion

Principal Findings

Our study examined patient-reported health care communication using CAHPS scores over the course of a 6-month RCT study (DIAMANTE). There was an overall marginal decrease in communication scores reported by patients in the study, without any significant differences by intervention arm, suggesting that the intervention did not alter patient-reported communication over time. Findings varied across different demographic subgroups; however, these trends in decreasing communication were similar in most subgroups.

Our findings are contrary to some previous literature which may suggest improvements in communication by leveraging digital technology. Previous studies show that initiating mHealth apps in the context of an already established patient-provider relationship is vital in building effective communication channels [14,23]. Additionally, increased touchpoints and motivational behaviors from providers throughout the usage of the mHealth app have been shown to improve patient-provider communication [24,25] and cultivate a greater sense of connectedness compared to traditional clinical encounters [26]. However, to our knowledge, this is the first evidence specific to RCTs using personalized digital communication.

While our study found no significant relationships between the digital intervention and patient reports of communication, it uncovered overall decreasing trends in patient-provider communication. One major reason for this finding may be related to the influence of the COVID-19 pandemic on overall patient-provider communication. While patient-provider communication ratings measured by CAHPS at a national level have been steadily increasing through 2019, there was a decrease in hospital CAHPS scores over 2020 - 2021 period [27]; other findings have been mixed. For example, one previous study found that hospital CAHPS scores decreased across 3381 US hospitals during COVID-19 [28], whereas other studies found that CAHPS scores did not differ between in-person and telehealth visits, with one comparing scores before and during COVID-19 and the second study comparing CAHPS score variations between the two services during the pandemic [29,30]. Thus, it is unclear how the COVID-19 pandemic, which limited in-person interactions between patients and their providers uniformly, impacted the foundational rapport and trust between patients and providers.

Lastly, it has been documented that the pandemic exacerbated burnout among clinicians [28]. This is particularly important given the need for clinicians to have sufficient support and capability to invest in relationship-centered care, and this was particularly challenging in the context of the pandemic with many competing demands (ie, both mental and physical) in health care settings [31]. Additional research has also demonstrated the rapidly increasing digital communication channels and resulting workload on clinicians in recent years

XSL•FO
RenderX

[32]. Therefore, there may be potential for mHealth tools, particularly those that are more personalized, to impact patients' perceptions of their relationship with their health care team via physician workload pathways. Moving forward, more research is needed on how the role of digital communication, including personalized communication with machine learning can support both patient perceptions of their ongoing care as well as the relevance of the technology to the well-being and workflows of clinicians. Further, since many mHealth tools are used for chronic care management or educational support outside the confines of traditional clinical communication, it is important to distinguish the role of these tools on ongoing provider relationships, particularly for patients who have longitudinal care from the same providers.

Limitations

This study has several limitations. The study sample is limited due to the nature of the original RCT. Although CAHPS is a validated source of assessing patient-provider trust and communication, more granular measures of patient-provider communication, such as assessments of specific types and frequency of communication for chronic illness management, may have identified different patterns or mechanisms related to patient-provider communication. In addition, it was difficult to tease out the role of the COVID-19 pandemic and the multiple recruitment strategies for the DIAMANTE trial, particularly related to variation in the connection between each patient and their primary care provider. However, to our knowledge, this is the first longitudinal assessment of CAHPS communication scores examined within a randomized trial involving a digital intervention, which is an important base for future research. With the surge of artificial intelligence and adaptive learning technologies, there is a need for ethical research on how these techniques affect communication and the trust of the communities whose confidence in the health care system was already fragmented.

Conclusions

As the use of adaptive learning–based digital health apps have increased across health care facilities, it is important for technology to extend, not replace strong in-person relationships built on trust and evidence–based communication strategies. In-person communication skills will remain critical even as the pace of digital and adaptive technology implementation accelerates.

Acknowledgments

This study was supported by grant number R01 HS25429 from the Agency for Healthcare Research and Quality. Dr. Lyles was also supported by the National Institute of Diabetes and Digestive and Kidney Disease (NIDDK P30DK092924). None of the funders had any role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript.

All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided APF support for the publication of this article.

Authors' Contributions

Conceptualization: AA, CRL, LL, MRAA
Formal analysis: LL
Supervision: AA, CRL, MRAA
Writing – original draft: LL
Writing – review & editing: AA, CRL, MRAA

Conflicts of Interest

AA owns the license to HealthySMS, which was used in the DIAMANTE study.

Checklist 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File, 1218 KB - mhealth_v13i1e64296_app1.pdf]

References

1. Fisher EB, Chan JCN, Nan H, Sartorius N, Oldenburg B. Co-occurrence of diabetes and depression: conceptual considerations for an emerging global health challenge. J Affect Disord 2012 Oct;142 Suppl:S56-S66. [doi: 10.1016/S0165-0327(12)70009-5] [Medline: 23062858]
2. Vos T, Abajobir AA, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet 2017 Sep;390(10100):1211-1259. [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28711585]

XSL•FO
RenderX

3. Chireh B, Li M, D'Arcy C. Diabetes increases the risk of depression: a systematic review, meta-analysis and estimates of population attributable fractions based on prospective studies. Prev Med Rep 2019 Jun;14:100822. [doi: 10.1016/j.pmedr.2019.100822] [Medline: 30815337]
4. Sartorius N. Depression and diabetes. Dialogues Clin Neurosci 2018 Mar;20(1):47-52. [doi: 10.31887/DCNS.2018.20.1/nsartorius] [Medline: 29946211]
5. Colon E, Giachello A, McIver L, Pacheco G, Vela L. Diabetes and depression in the Hispanic/Latino community. Clin Diabetes 2013 Jan 1;31(1):43-45. [doi: 10.2337/diaclin.31.1.43]
6. Thomas J, Jones G, Scarinci I, Brantley P. A descriptive and comparative study of the prevalence of depressive and anxiety disorders in low-income adults with type 2 diabetes and other chronic illnesses. Diabetes Care 2003 Aug;26(8):2311-2317. [doi: 10.2337/diacare.26.8.2311] [Medline: 12882854]
7. Wang Y, Min J, Khuri J, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400. [doi: 10.2196/15400] [Medline: 32343253]
8. Street RL Jr, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
9. MHealth programmes are sponsored by government. World Health Organization. URL: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/4774 [accessed 2023-11-26]
10. Aguilera A, Bruehlman-Senecal E, Demasi O, Avila P. Automated text messaging as an adjunct to cognitive behavioral therapy for depression: a clinical trial. J Med Internet Res 2017 May 8;19(5):e148. [doi: 10.2196/jmir.6914] [Medline: 28483742]
11. Watterson JL, Rodriguez HP, Shortell SM, Aguilera A. Improved diabetes care management through a text-message intervention for low-income patients: mixed-methods pilot study. JMIR Diabetes 2018 Oct 30;3(4):e15. [doi: 10.2196/diabetes.8645] [Medline: 30377141]
12. Anisha SA, Sen A, Bain C. Evaluating the potential and pitfalls of AI-powered conversational agents as humanlike virtual health carers in the remote management of noncommunicable diseases: scoping review. J Med Internet Res 2024 Jul 16;26:e56114. [doi: 10.2196/56114] [Medline: 39012688]
13. Nguyen MH, Hargittai E, Marler W. Digital inequality in communication during a time of physical distancing: the case of COVID-19. Comput Human Behav 2021 Jul;120:106717. [doi: 10.1016/j.chb.2021.106717]
14. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
15. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128. [doi: 10.2196/mhealth.8383] [Medline: 29914860]
16. Redfern J, Santo K, Coorey G, et al. Factors influencing engagement, perceived usefulness and behavioral mechanisms associated with a text message support program. PLOS ONE 2016;11(10):e0163929. [doi: 10.1371/journal.pone.0163929] [Medline: 27741244]
17. Piras EM, Miele F. On digital intimacy: redefining provider-patient relationships in remote monitoring. Sociol Health Illn 2019 Oct;41 Suppl 1:116-131. [doi: 10.1111/1467-9566.12947] [Medline: 31599992]
18. Aguilera A, Arévalo Avalos M, Xu J, et al. Effectiveness of a digital health intervention leveraging reinforcement learning: results from the Diabetes and Mental Health Adaptive Notification Tracking and Evaluation (DIAMANTE) randomized clinical trial. J Med Internet Res 2024 Oct 8;26:e60834. [doi: 10.2196/60834] [Medline: 39378080]
19. Aguilera A, Figueroa CA, Hernandez-Ramos R, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. BMJ Open 2020 Aug 20;10(8):e034723. [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]
20. Pathak LE, Aguilera A, Williams JJ, et al. Developing messaging content for a physical activity smartphone app tailored to low-income patients: user-centered design and crowdsourcing approach. JMIR Mhealth Uhealth 2021 May 19;9(5):e21177. [doi: 10.2196/21177] [Medline: 34009130]
21. Surveys and tools to advance patient-centered care. Agency Healthc Res Qual. URL: https://www.ahrq.gov/cahps/cahps-database/index.html [accessed 2023-03-10]
22. Quigley DD, Elliott MN, Qureshi N, Predmore Z, Hays RD. How the CAHPS clinician and group patient experience survey data have been used in research: a systematic review. J Patient Cent Res Rev 2024;11(2):88-96. [doi: 10.17294/2330-0698.2056] [Medline: 39044849]
23. Lu C, Hu Y, Xie J, et al. The use of mobile health applications to improve patient experience: cross-sectional study in Chinese public hospitals. JMIR Mhealth Uhealth 2018 May 23;6(5):e126. [doi: 10.2196/mhealth.9145] [Medline: 29792290]
24. Shaw Y, Courvoisier DS, Scherer A, et al. Impact of assessing patient-reported outcomes with mobile apps on patient-provider interaction. RMD Open 2021 Apr;7(1):e001566. [doi: 10.1136/rmdopen-2021-001566] [Medline: 33811177]

XSL•FO
RenderX

25. Van Baelen F, De Regge M, Larivière B, Verleye K, Schelfout S, Eeckloo K. Role of social and app-related factors in behavioral engagement with mhealth for improved well-being among chronically ill patients: scenario-based survey study. JMIR Mhealth Uhealth 2022 Aug 26;10(8):e33772. [doi: 10.2196/33772] [Medline: 36018618]
26. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019 Jun;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
27. National healthcare quality and disparities report. : AHRQ; 2023 URL: https://www.ahrq.gov/research/findings/nhqrdr/nhqdr23/index.html [accessed 2023-12-14]
28. Elliott MN, Beckett MK, Cohea CW, et al. Changes in patient experiences of hospital care during the COVID-19 pandemic. JAMA Health Forum 2023 Aug 4;4(8):e232766. [doi: 10.1001/jamahealthforum.2023.2766] [Medline: 37624612]
29. Hays RD, Skootsky SA. Patient experience with in-person and telehealth visits before and during the COVID-19 pandemic at a large integrated health system in the United States. J Gen Intern Med 2022 Mar;37(4):847-852. [doi: 10.1007/s11606-021-07196-4] [Medline: 34982370]
30. Gotthardt CJ, Haynes SC, Sharma S, Yellowlees PM, Luce MS, Marcin JP. Patient satisfaction with care providers during the COVID-19 pandemic: an analysis of consumer assessment of healthcare providers and systems survey scores for in-person and telehealth encounters at an academic medical center. Telemed J E Health 2023 Aug;29(8):1114-1126. [doi: 10.1089/tmj.2022.0460] [Medline: 36595515]
31. Boissy A. Getting to patient-centered care in a post-COVID-19 digital world: a proposal for novel surveys, methodology, and patient experience maturity assessment. NEJM Catal Innov Care Deliv 2020. [doi: 10.1056/CAT.19.1106]
32. Bressman E, Long JA, Honig K, et al. Evaluation of an automated text message-based program to reduce use of acute health care resources after hospital discharge. JAMA Netw Open 2022 Oct 3;5(10):e2238293. [doi: 10.1001/jamanetworkopen.2022.38293] [Medline: 36287564]

Abbreviations

CAHPS: Consumer Assessment of Healthcare Providers and Systems
DIAMANTE: Diabetes and Mental Health Adaptive Notification Tracking and Evaluation
HbA_{1c}: glycosylated hemoglobin
mHealth: mobile health
RCT: randomized controlled trial

Edited by L Buis; submitted 14.07.24; peer-reviewed by R Nooripour, S Babbott; revised version received 06.04.25; accepted 11.06.25; published 17.07.25.

Please cite as:
Leng L, Arévalo Avalos MR, Aguilera A, Lyles CR
The Impact of an Adaptive mHealth Intervention on Improving Patient-Provider Health Care Communication: Secondary Analysis of the DIAMANTE Trial
JMIR Mhealth Uhealth 2025;13:e64296
URL: https://mhealth.jmir.org/2025/1/e64296
doi:10.2196/64296

© Lynn Leng, Marvyn R Arévalo Avalos, Adrian Aguilera, Courtney R Lyles. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Validity, Accuracy, and Safety Assessment of an Aerobic Interval Training Using an App-Based Prehabilitation Program (PROTEGO MAXIMA Trial) Before Major Surgery: Prospective, Interventional Pilot Study

Sara Fatima Faqar Uz Zaman[1], MD; Svenja Sliwinski[1], MD; Lisa Mohr-Wetzel[2], MSci; Julia Dreilich[2], MSci; Natalie Filmann[3], Dr rer med; Charlotte Detemble[1,4], MD; Dora Zmuc[1], Dr med; Felix Chun[5], Prof Dr; Wojciech Derwich[1,6], Dr med; Waldemar Schreiner[1], Prof Dr; Wolf Bechstein[1], Prof Dr; Johannes Fleckenstein[2], Prof Dr; Andreas A Schnitzbauer[1,4], Prof Dr, MD

[1]Department of General, Visceral, Transplant and Thoracic Surgery, University Hospital Frankfurt, Goethe University Frankfurt, Frankfurt am Main, Germany

[2]Institute for Sports Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany

[3]Institute for Biostatistics and Mathematical Modelling, Goethe University Frankfurt, Frankfurt am Main, Germany

[4]Department of Surgery, Knappschaft Clinics, University Hospital Bochum, Ruhr University Bochum, Bochum, Germany

[5]Department of Urology, University Hospital Frankfurt, Goethe University Frankfurt, Frankfurt am Main, Germany

[6]Department of Vascular Surgery, University Hospital Frankfurt, Goethe University Frankfurt, Frankfurt am Main, Germany

Corresponding Author:
Andreas A Schnitzbauer, Prof Dr, MD
Department of General, Visceral, Transplant and Thoracic Surgery
University Hospital Frankfurt
Goethe University Frankfurt
Theodor-Stern-Kai 7
Frankfurt am Main, 60590
Germany
Phone: 49 69 6301
Email: andreas.schnitzbauer@rub.de

Abstract

Background: Major surgery is associated with significant morbidity and a reduced quality of life, particularly among older adults and individuals with frailty and impaired functional capacity. Multimodal prehabilitation can enhance functional recovery after surgery and reduce postoperative complications. Digital prehabilitation has the potential to be a resource-sparing and patient-empowering tool that improves patients' preoperative status; however, little remains known regarding their safety and accuracy as medical devices.

Objective: This study aims to test the accuracy and validity of a new software in comparison to the gold-standard electrocardiogram (ECG)-based heart rate measurement.

Methods: The PROTEGO MAXIMA trial was a prospective interventional pilot trial assessing the validity, accuracy, and safety of an app-based exercise program. The Prehab App calculates a personalized, risk-stratified aerobic interval training plan based on individual risk factors and utilizes wearables to monitor heart rate. Healthy students and patients undergoing major surgery were enrolled. A structured risk assessment was conducted, followed by a 6-minute walking test and a 37-minute supervised interval session. During the exercise, patients wore app-linked wearables for heart rate and distance measurements, which were compared with standard ECG and treadmill measurements. Safety, accuracy, and usability assessments included testing alarm signals, while the occurrence of adverse events served as the primary and secondary outcome measures.

Results: A total of 75 participants were included. The mean heart rate differences between wearables and standard ECG were ≤5 bpm (beats per minute) with a mean absolute percentage error of ≤5%. Regression analysis revealed a significant impact of the BMI (odds ratio 0.90, 95% CI 0.82-0.98, *P*=.02) and Timed Up and Go Test score (odds ratio 0.12, 95% CI 0.03-0.55, *P*=.006) on the accuracy of heart rate measurement; 29 (39%) patients experienced adverse events: pain (5/12, 42%), ECG electrode–related

XSL•FO
RenderX

skin irritations (2/42, 17%), dizziness (2/42, 17%), shortness of breath (2/42, 17%), and fatigue (1/42, 8%). No cardiovascular or serious adverse events were reported, and no serious device deficiency was detected. There were no indications of clinically meaningful overexertion based on laboratory values measured before and after the 6-minute walking test and exercise. The differences in means and ranges were as follows: lactate (mmol/l), mean 0.04 (range –3 to 6; *P*=.47); creatinine kinase (U/l), mean 12 (range –7 to 43; *P*<.001); and sodium (mmol/l), mean –2 (range –11 to 12; *P*<.001).

Conclusions: The interventional trial demonstrated the high safety of the exercise program and the accuracy of heart rate measurements using commercial wearables in patients before major surgery, paving the way for potential remote implementation in the future.

Trial Registration: German Clinical Trials Register DRKS00026985; https://drks.de/search/en/trial/DRKS00026985 and European Database on Medical Devices (EUDAMED) CIV-21-07-0307311.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2022-069394

(JMIR Mhealth Uhealth 2025;13:e55298) doi:10.2196/55298

KEYWORDS

digital health; prehab; major surgery; surgical oncology; smartwatches; safety and quality; surgery; surgical; oncology; validity; accuracy; safety management; management; aerobic; aerobic training; app; prehabilitation; pilot study; quality of life; medical device; wearable; wearables; heart rate

Introduction

Postoperative complications occur in 15%-40% of patients undergoing major surgery, potentially resulting in life-threatening conditions, a decline in quality of life, or reduced physical functioning [1-3]. Adverse events (AEs) associated with surgical procedures impose a significant financial burden due to additional costs from intensive care treatment, reoperations, or prolonged hospital stays [4,5]. Validated assessment tools, such as the Risk Analysis Index (RAI) score, the Eastern Cooperative Oncology Group (ECOG) Performance Status, and the Timed Up and Go Test (TUG), can accurately identify patients' individual risk factors and predict their surgical outcomes [6-9]. Importantly, the physical ability to recover from the physiological stress of surgery and the demands of an aggressive postoperative metabolism is strongly correlated with preoperative health status and physical functioning [10,11]. Accordingly, preoperative exercise interventions aim to enhance patients' aerobic capacity and muscle strength, significantly facilitating postoperative recovery.

Prehabilitation, an emerging field in perioperative care, addresses modifiable risk factors before surgery through interventions such as exercise training, nutritional support, and psychocognitive training [11]. Notably, moderate- to high-intensity programs combining aerobic and resistance training over 3-6 weeks have been shown in recent studies to significantly enhance cardiorespiratory reserve and the capacity to adapt to physical stress [11-13].

Digital health technologies have the potential to enhance health management and facilitate the large-scale implementation of personalized prehabilitation programs in clinical settings by offering cost-effectiveness, broad population accessibility, and high adherence to short-term exercise programs [14,15]. In particular, mobile health (mHealth) apps that leverage mobile wireless technologies and integrate with wearable devices are powerful tools for achieving health objectives through self- and remote monitoring, personalized goal setting, and gamification [16,17]. Wearable devices used for monitoring heart rate and activity are increasingly evaluated in accuracy validation studies [18,19]. The Prehab App has been developed as a platform linked to a wearable device, enabling individualized, risk-adjusted, home-based aerobic exercise training as prehabilitation before major surgery. The objective of the prospective interventional PROTEGO MAXIMA trial was to evaluate the accuracy of wearable measurements and the safety of the prehabilitation program compared with certified electrocardiogram (ECG) measurements in a supervised and controlled environment. Remote, independent, interventional, and risk-based exercising represents a novel, stand-alone approach in prehabilitation, necessitating regulated testing in compliance with the Medical Device Regulation (MDR) [20], the DIN (Deutsches Institut für Normung; German Industry Norm) ISO (International Organization for Standardization) 14155 [21], and the Medical Device Law Implementation Act (Medizinproduktedurchführungsgesetz; MPDG) [22]. This approach also holds first-mover status for real-time data capture. The hypothesis was that the software would demonstrate comparable accuracy and validity in heart rate measurement during a 6-minute walking test (6MWT) and a 47-minute interval training session, matching the performance of a certified ECG, which is considered the gold standard.

Methods

Study Design and Participants

The prospective interventional PROTEGO MAXIMA trial was designed to evaluate the validity, accuracy, and safety of a prehabilitation app. Reporting was conducted in accordance with the STARD (Standards for Reporting of Diagnostic Accuracy; Multimedia Appendix 1) reporting guidelines [23]. The trial included healthy students and patients scheduled for elective major surgery, conducted between March 25, 2022, and September 29, 2022, at the University Hospital Frankfurt [24]. The study protocol has been previously published [25]. In brief, the trial initially recruited 10 healthy students from the Sports Institute at Goethe University Frankfurt to establish a baseline data set from healthy volunteers. In the next phase, 65

XSL•FO
RenderX

patients (aged 18 years or older, capable of understanding and performing the endurance exercise program, and scheduled for elective major surgery) were recruited at the University Hospital Frankfurt. Eligible surgeries included gastrointestinal resections; resections of the hepatobiliary pancreatic system, endocrine glands, lung, or bronchus; splenectomy; abdominal wall hernia repairs; urological or gynecologic resections; or vascular surgeries excluding cardiovascular procedures. We did not exclude patients with chronic but compensated cardiovascular and pulmonary conditions, as these represented a key focus group for assessing risk. The main exclusion criteria were a history of pregnancy or breastfeeding, acute cardiovascular disease, or any acute noncardiopulmonary disorder that could affect or be exacerbated by exercise performance. All participants provided written informed consent. As this is a first-in-human-use pilot trial, we broadened the scope of data collection to include as many scenarios as possible, ranging from healthy students with anticipated textbook outcomes to normal and high-risk patients. At this stage, there was no control group as we needed to evaluate the validity and general safety of the interventional medical device, similar to a phase I study in drug development.

Ethical Considerations

The study adhered to the regulations of the DIN ISO 14155 [21], the MDR [20], and the MPDG [22]. The trial protocol was approved by the institutional review board of the University Hospital Frankfurt and by the Federal Institute for Pharmaceuticals and Medical Products (Bundesinstitut für Arzneimittel und Medizinprodukte; reference number 94.1.04-5660-13655) on February 7, 2022. The study was registered in the German Clinical Trial Register (DRKS00026985) on December 21, 2021, and in the European Database on Medical Devices (EUDAMED; CIV-21-07-0307311). Data from all patients were stored in an electronic case report form using the secuTrial system, a web-based data management application that complies with all required regulations. To assess the quality of the data and to audit the conduct of the trial, independent reviews were conducted by monitors following the monitoring plan of the clinical trial.

Aerobic Interval Training Using the Prehab App

The Prehab App is developed by Capreolos GmbH, a spin-off company from Goethe University, as a medical device class IIa according to the MDR. Currently, it is not yet CE-certified. The app is linked to a wrist-worn wearable (iOS- and Google Wear OS-based hardware) through a software app, providing an individualized, digitized prehabilitation program for patients undergoing major surgery to enhance their functional capacity using aerobic endurance exercise training. Based on the patient's risk factors (ECOG performance status, RAI score, TUG, hemoglobin values, and the presence of heart rate–changing medication), along with the resting heart rate, an individualized interval training program plan with a target heart rate range for moderate to vigorous training intensity (50%-80% of the maximum heart rate) is calculated using the Karvonen method (Multimedia Appendix 2) [26,27]. The predefined exercise program is then automatically uploaded to the patient's smartphone and is ready for use with a provided wearable. By measuring heart rate and distance during exercise, the app provides feedback on the patient's physical performance and potential alarms if the measured heart rate exceeds or falls below the suggested ranges (automated safety audit). Furthermore, the app used by the treating physician will be linked to the patient's app, allowing for supervision by receiving the same alarms as defined in the safety audit and derived from DIN ISO 14971–compliant risk management [28]. After completing the individual exercises, a patient-reported outcome form will collect symptoms during the exercise, which feeds into the safety audit in case of more than 2 reported symptoms. In the final intended use, patients will perform the training 3-4 times a week, selecting the type of exercise (eg, cycling, walking, running) at any time of the day, thereby facilitating a home-based, patient-empowering environment. A 6MWT at the beginning and end of the training period will measure baseline and endpoint fitness. The Quality-of-Life Questionnaire for patients with cancer, consisting of 30 questions (QLQC30), will assess the quality of life before and at the end of the training period.

A detailed overview of the risk assessment app interface and the flow diagram of the exercise training performed under controlled conditions is provided in Multimedia Appendices 2 and 3, respectively. Eligible participants underwent a structured risk assessment and were assigned to a predefined risk score (1-6) based on the ECOG performance status (ECOG 0=0 points, ECOG 1=1 point, and ECOG >1=2 points), TUG (0-7 seconds=1 point and >7 seconds=2 points), RAI-C score (≤25=0 points and >25=2 points), and hemoglobin values (>13 g/dl=0 points and ≤13 g/dl=2 points). Then, after a baseline 6MWT, participants performed 1 one-time endurance exercise program comprising a 47-minute interval training (5 minutes of warm-up, followed by alternating 2 minutes of high-intensity intervals and 3 minutes of low-intensity active rests for a total of 37 minutes, followed by a 5-minute cool down period) on a treadmill at the Institute of Sports Science of Goethe University Frankfurt. Participants wore 1 smartwatch on each wrist (1 iOS [Apple Inc.] and 1 Android [Google LLC/Alphabet Inc.] software combination), which was linked to the app on the complementary smartphone. Four different device combinations were assessed to extend the validity of the measurements: iPhone 13 + Apple Watch 7 (Apple 13), iPhone SE + Apple Watch 3 (Apple SE), Samsung A52 + Samsung Galaxy Watch4 (Samsung A52), and Google Pixel 6 + Samsung Galaxy Watch4 (Google 6). The rationale for choosing these wearables was straightforward: we aimed to provide the 2 major platforms used in smartphones and wearable technologies. As Google Wear OS was only recently introduced, the Samsung Galaxy Watch4 was the only compatible solution at the time of development. An extension to other wearable solutions can be easily performed later. Heart rates and distances were measured at different time points (6 time points during the 6MWT and 19 time points during the interval training) by the wearables and compared with a 12-lead gold-standard ECG (Custo diagnostic; Custo med GmbH) and the treadmill measurement. Participants' perceived intensity and pain during exercise were assessed using the Borg Scale and Visual and Numerical

XSL•FO
RenderX

Analogue Scale [29]. The maximum rate of oxygen consumption (VO_{2max}) in metabolic equivalent (MET) after the 6MWT was used to determine the baseline cardiorespiratory fitness [30,31]. Individual characteristics including wrist circumference, skin appendages, and skin humidity were assessed to evaluate their impact on the accuracy of measurements. As this is a fully regulated development process, the DIN ISO 13485 [28], the MDR [20], all applicable legislation such as the DIN ISO 27001 [32], and the recommendations for cybersecurity implemented by the Bundesamt für Sicherheit in der Informationstechnik (BSI [33]; Federal Office for Information Security) in Germany must be followed. This is done by a dedicated quality management program. Additionally, there is an external data safety and privacy officer within Capreolos GmbH who is responsible for monitoring this area. In brief, as the risk assessment is conducted by doctors on a separate app, all data are pseudonymized, and there are no personal data on the patient's app. There is only a study-specific number plus a log-in so that the patient can only be depseudonymized by the clinical institutions.

Safety Assessment of the Endurance Exercise Program

The aerobic exercise training was continuously supervised by a physician to ensure a safe environment. Secondary safety measures were based on blood samples—reported elsewhere—that were taken before and 15 minutes after the interval training (including creatine phosphokinase, creatine kinase-muscle-brain isoform, lactate dehydrogenase, glucose, sodium, potassium, chloride, C-reactive protein, blood carbon dioxide, lactate, and cell-free DNA). Moreover, the awareness of participants to specific alarm signals displayed on the wearable devices—by either color or haptic vibration—when exceeding or falling below the defined heart rate ranges was assessed. Symptoms, cardiac events, AEs, and serious AEs (SAEs) were assessed during the exercise and through structured telephone interviews or by visiting patients in the hospital (in case they had already undergone their planned surgery) on day 7 and 30 days after the exercise intervention [34].

Statistical Analysis

The sample size calculation was based on the assumption that a maximal irrelevant mean difference of ≤10 bpm (beats per minute) is regarded as acceptable [18,35]. An acceptable error rate using the mean absolute percentage error (MAPE) was defined to be +10% or −10% according to recommendations [36,37]. Considering a power of 80%, a 1-sided α of .05, and a dropout rate of 25%, a sample size of 75 patients was required. Categorical variables were summarized as numbers (percentages), and differences between groups were compared with the χ^2 test. Continuous variables were summarized with the Mann-Whitney *U* test or the Kruskal-Wallis test. Bland-Altman analysis, including calculations of mean differences and 95% limits of agreement (LoA), was used to compare heart rate and distance measurements. MAPE values were calculated as the error percentage between measurements. Outliers were not removed to prevent interference with assessing the accuracy of wearable measurements. Univariate and multivariate analyses were performed using a logistic regression model to assess the impact of different individual variables (eg, wrist circumference, skin humidity, skin appendices, risk scores) on the accuracy of heart rate measurement (odds ratio [OR] with or without 95% CI) and the impact of these variables on cardiorespiratory fitness was determined by VO_{2max}. Statistical tests were 2-sided, and *P* values ≤0.05 were deemed statistically significant. Statistical analyses were performed using SPSS software (version 28.0; IBM Corp.) and R (R Studio, version 4.2.2/R Foundation, version 2022.07.2).

Role of the Funding Source

The clinical trial was funded by the Else-Kroener-Fresenius Stiftung through the translational research program (number 2021_EKTP10). The funder had no role in study design, data collection, data analysis, data interpretation, or the writing of the manuscript.

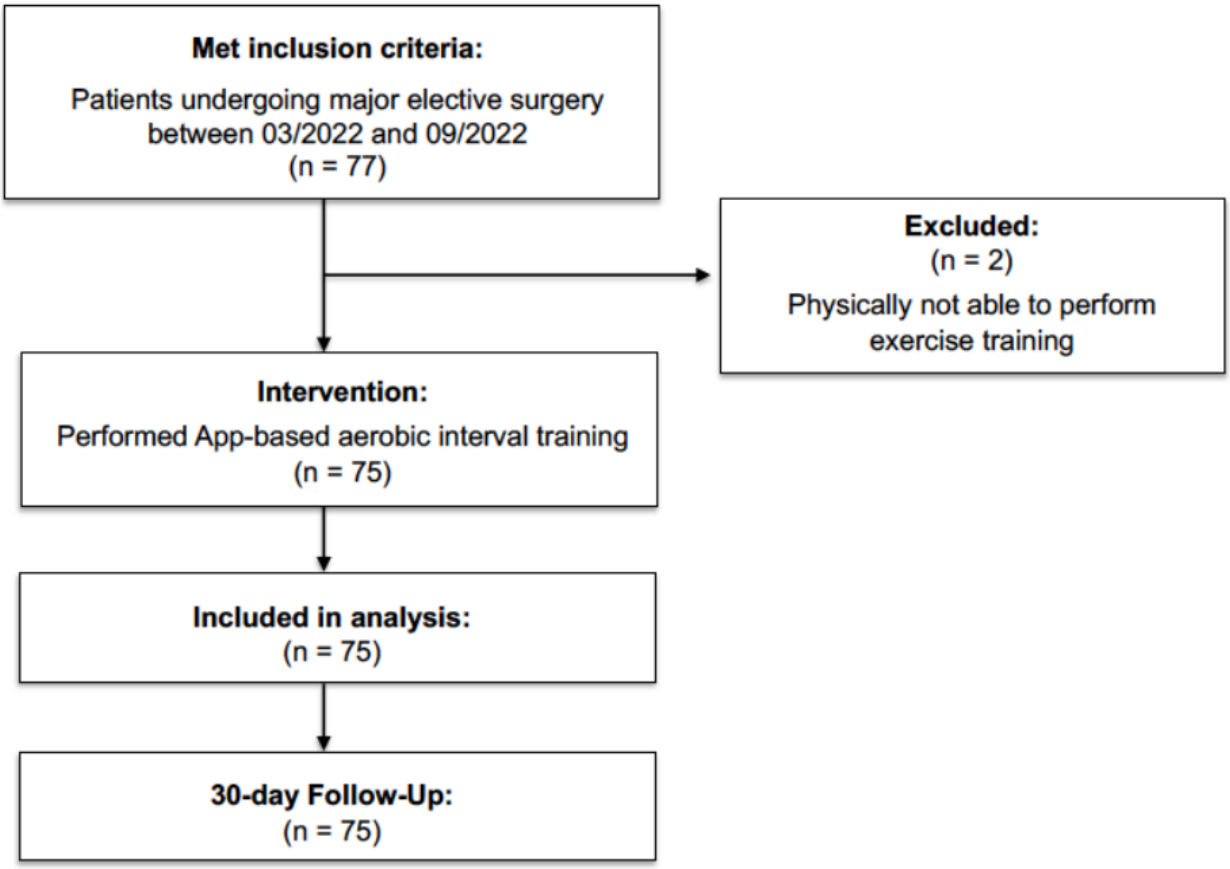
Results

Baseline Data and Risk Factors

Of the 77 participants recruited, 75 were included (Figure 1). Two individuals were excluded on the training day after double-checking the exclusion criteria because they gave misleading information during the first interview. Finally, a total of 75 participants with 150 wearables-smartphone combinations were analyzed, of whom no one was lost to follow-up. The study population was divided into 3 subgroups: 26 (35%) participants with a risk score of 0, 38 (51%) with a risk score between 1 and 3, and 11 (15%) with a risk score between 4 and 6 (Table 1). The median age was significantly higher in the risk score 4-6 group (63 vs 46 and 57, *P*=.01), consisting of significantly more male participants (9/11, 82%, for risk score 4-6, vs 12/26, 46%, for risk score 0, and 29/38, 76%, for risk score 1-3, *P*=.02; Multimedia Appendix 4 [38]). No differences were found in terms of BMI or smoking. Importantly, high- and medium-risk patients more often suffered from chronic cardiovascular or pulmonary disease compared with the other groups. High-risk patients were most commonly from the visceral (5/11, 45%) and thoracic surgical specialties (1/11, 9%) undergoing hepato-pancreato-biliary and lung resections, while urological patients (23/38, 61%) were more often assigned to the medium-risk group, scheduled for radical prostatectomy and nephrectomy. Low-risk patients predominantly comprised gynecological (4/26, 15%) and vascular surgical (3/26, 12%) patients undergoing salpingo-oophorectomy, hysterectomy, or carotid endarterectomy. Consistent with the cardiopulmonary comorbidities, the structured assessment of risk factors revealed a significantly higher percentage of RAI scores ≥26 in the high-risk group (5/11, 45%) compared with the other groups (7/38, 18%, and 0/26, 0%, for moderate- and low-risk groups, respectively; *P*<.001). Similarly, TUG scores of 7-11 seconds were more prevalent in the high-risk group (10/11, 91%) compared with the other groups (18/38, 47%, and 8/26, 31%, for moderate- and low-risk groups, respectively; *P*=.004). By contrast, no differences were observed regarding ECOG performance status and hemoglobin levels.

XSL•FO
RenderX

Figure 1. Trial profile for the interventional part of the PROTEGO MAXIMA trial: between March 25, 2022, and September 20, 2022, a total of 77 patients were enrolled, of whom 2 were excluded during exercise training. Finally, 75 patients were included in the statistical analysis.



XSL•FO
RenderX

Table 1. Baseline characteristics and risk profile of the study population by app-based risk scoring.

Variables	Overall (n=75)	Risk score 0 (n=26)	Risk score 1-3 (n=38)	Risk score 4-6 (n=11)	*P* value
Baseline characteristics					
Age (years); median (range)	56 (22-78)	46 (22-78)	57 (33-78)	63 (52-77)	.01
Sex, n (%)					.02
Female	25 (33)	14 (54)	9 (24)	2 (18)	
Male	50 (67)	12 (46)	29 (76)	9 (82)	
BMI (kg/m^2), median (range)	25 (16-40)	24 (20-40)	26 (16-35)	27 (21-40)	.12
Smoking, n (%)	15 (20)	3 (12)	7 (18)	5 (45)	.06
Medical department, n/N (%)					N/A[a]
Visceral surgery	17/64 (27)	3 (12)	9 (24)	5 (45)	
Urology	32/64 (50)	4 (15)	23 (61)	5 (45)	
Gynecology	8/64 (13)	4 (15)	4 (11)	0 (0)	
Thoracic surgery	3/64 (5)	1 (4)	1 (3)	1 (9)	
Vascular surgery	4/64 (6)	3 (12)	1 (3)	0 (0)	
Major surgery, n/N (%)					N/A
Radical prostatectomy	26/64 (41)	3/64 (5)	18/64 (28)	5/64 (8)	
Hepatobiliary pancreatic resections	9/64 (14)	0/64 (0)	1/64 (2)	4/64 (6)	
Partial or total nephrectomy	6/64 (9)	2/64 (3)	4/64 (6)	0/64 (0)	
Salpingo-oophorectomy or hysterectomy	6/64 (9)	3/64 (5)	3/64 (5)	0/64 (0)	
Carotid endarterectomy	4/64 (6)	3/64 (5)	1/64 (2)	0/64 (0)	
Lung resection	3/64 (5)	1/64 (2)	1/64 (2)	1/64 (2)	
Abdominal wall hernia repair	3/64 (5)	0/64 (0)	2/64 (3)	1/64 (2)	
Other	7/64 (11)	3/64 (5)	4/64 (6)	0/64 (0)	
Risk group					N/A
Healthy student	11 (15)	11 (42)	0 (0)	0 (0)	
No cardiopulmonary disease	53 (71)	13 (50)	31 (82)	9 (82)	
Chronic cardiovascular disease	7 (9)	1 (4)	5 (13)	1 (9)	
Chronic obstructive pulmonary disease	4 (5)	1 (4)	2 (5)	1 (9)	
Risk scoring					
Risk assessment					
Eastern Cooperative Oncology Group score, n (%)					N/A
	0	73 (97)	26 (100)	36 (95)	11 (100)
	1	2 (3)	0 (0)	2 (5)	0 (0)
Risk Analysis Index score, n (%)					<.001
	≤15	25 (33)	19 (73)	6 (16)	0 (0)
	16-25	38 (51)	7 (27)	25 (66)	6 (55)
	≥26	12 (16)	0 (0)	7 (18)	5 (45)
Timed Up and Go Test, seconds, n (%)					.004
	0-6	39 (52)	18 (69)	20 (53)	1 (9)
	7-11	36 (48)	8 (31)	18 (47)	10 (91)
Hemoglobin, g/dL, median (range)	14 (10-17)	14 (13-15)	14 (11-17)	14.0 (10-16)	.48
Functional capacity					

XSL•FO
RenderX

Variables	Overall (n=75)	Risk score 0 (n=26)	Risk score 1-3 (n=38)	Risk score 4-6 (n=11)	*P* value
VO_{2max}[b] (MET[c]), median (range)	3 (1-6)	3 (1-6)	3 (1-5)	2.4 (2-3)	.17

[a]N/A: not applicable.

[b]VO_{2max}: maximum rate of oxygen consumption.

[c]MET: metabolic equivalent.

Furthermore, by analyzing individual factors that impacted VO_{2max}, we found that age was a significant factor (*P*=.03) for cardiorespiratory fitness in the univariate analysis (see Multimedia Appendix 5). In general, the median VO_{2max} (in MET) for the 6MWT in all patients was 3.1 (range 1.2-6.0), indicating an overall low to moderate aerobic capacity.

Heart Rate Measurement Accuracy

The accuracy of the heart rate and distance measurement of The Prehab App software in combination with the wearables was compared with the measurement of the standard ECG and treadmill (Table 2). The mean heart rate differences for all devices were ≤5 bpm, ranging from 2.92 bpm (iPhone SE, LoA –21.0 to 26.8) to 4.48 bpm (Samsung A52, LoA –19.1 to 27.9) for the 6MWT and from 1.33 bpm (iPhone SE, LoA –19.3 to 22.0) to 2.61 bpm (iPhone 13, LoA –22.0 to 27.1) for the interval training. Consistently, MAPE values for all 4 devices were below the threshold of 10%, with lower values for the interval training (–0.3% [Google Pixel 6] to –1.8% [iPhone 13]) compared with the 6MWT (–1.6% [iPhone 13] to –2.9% [Samsung A52]; see Multimedia Appendix 6). Results showed MAPE values well below 5% and a mean heart rate difference under 5 bpm, indicating excellent reliability of iOS and Android heart rate measurements for The Prehab App software. Bland-Altman plots with 95% LoA are depicted in Figures 2 and 2. Interestingly, all 4 wearables exhibited a slight tendency to underestimate heart rates, which was generally greater for heart rate values over 100 bpm.

For a more detailed analysis regarding the accuracy of heart rate measurement, we subsequently assessed the impact of different individual characteristics, which was defined as an MAPE of <5% (Table 3). Multivariable analysis of individual variables assessed before the exercise training indicated that a low BMI (kg/m^2) (OR 0.897, 95% CI 0.822-0.980, *P*=.02) and a low TUG score (OR 0.122, 95% CI 0.027-0.546, *P*=.006) were independent predictive factors for high accuracy of the heart rate measurement.

XSL•FO
RenderX

Table 2. Comparison of heart rate and distance measurement between smartwatch versus ECG and treadmill.

Variables	iOS devices		Android devices	
	iPhone 13 + App Watch 7 (n=38)	iPhone SE + App Watch 3 (n=36)	Samsung A52 + Samsung Galaxy Watch4 (n=34)	Google Pixel 6 + Samsung Galaxy Watch4 (n=42)
Heart rate measurement				
6-minute walking test				
Mean (SD) (beats per minute)	105 (20)	101 (18)	106 (19)	100 (18)
Mean difference (beats per minute)	2.98	2.92	4.48	3.50
MAPE[a] (%)	–1.62	–1.76	–2.87	–2.24
LoA[b] (lower to upper; beats per minute)	–23.9 to 29.9	–21.0 to 26.8	–19.1 to 27.9	–21.1 to 28.1
Interval training				
Mean (SD) (beats per minute)	111 (23)	112 (26)	112 (21)	111 (23)
Mean difference (beats per minute)	2.61	1.33	2.16	1.54
MAPE (%)	–1.80	0.53	0.45	0.30
LoA (lower to upper; beats per minute)	–22.0 to 27.1	–19.3 to 22.0	–27.4 to 31.7	–19.6 to 22.7
Distance measurement				
6-minute walking test				
Mean (SD) (m)	494 (165)	537 (163)	519 (181)	510 (169)
Mean difference (m)	–29.22	–104.28	–31.67	–50.57
MAPE (%)	14.71	30.61	15.28	20.34
LoA (lower to upper; m)	–325.6 to 267.2	–361.8 to 153.2	–411.6 to 348.3	–373.1 to 272.0
Interval training				
Mean (SD) (m)	3650 (1596)	3813 (1327)	4069 (1562)	3612 (1517)
Mean difference (m)	27.51	–24.15	–12.27	208.15
MAPE (%)	0.05	–0.58	3.25	–6.95
LoA (lower to upper; m)	–2393.1 to 2448.1	–1824.4 to 1776.1	–2658.6 to 2634.0	–2003.6 to 2420.0

[a]MAPE: mean absolute percentage error.

[b]LoA: limits of agreement.

XSL•FO
RenderX

Figure 2. Bland-Altman plots for wearable measurement versus standard measurement: Bland-Altman plots for wearable heart rate measurements compared with electrocardiogram (ECG) measurements for 4 device combinations. The ECG heart rate, used as the reference, is shown on the x-axis, and the specified differences between ECG and the device are shown on the y-axis. The bold line corresponds to the mean difference, and the dashed lines indicate the 95% limits of agreement.

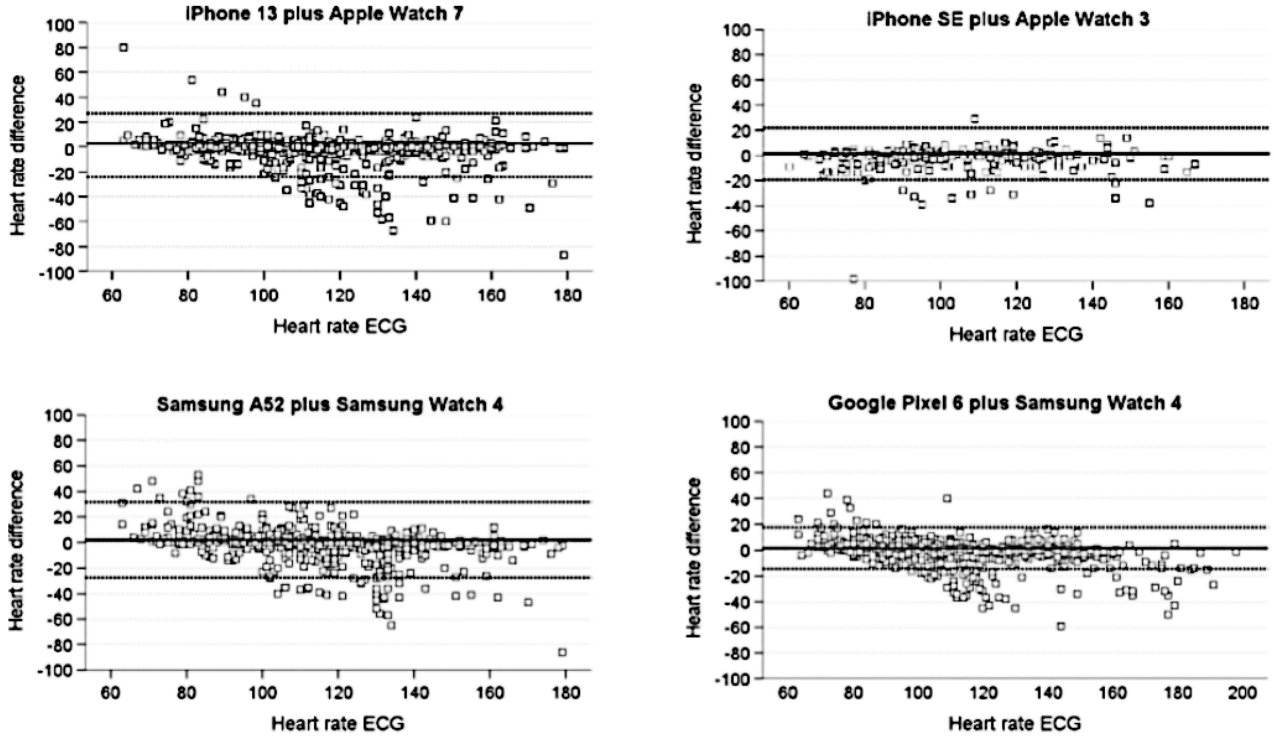
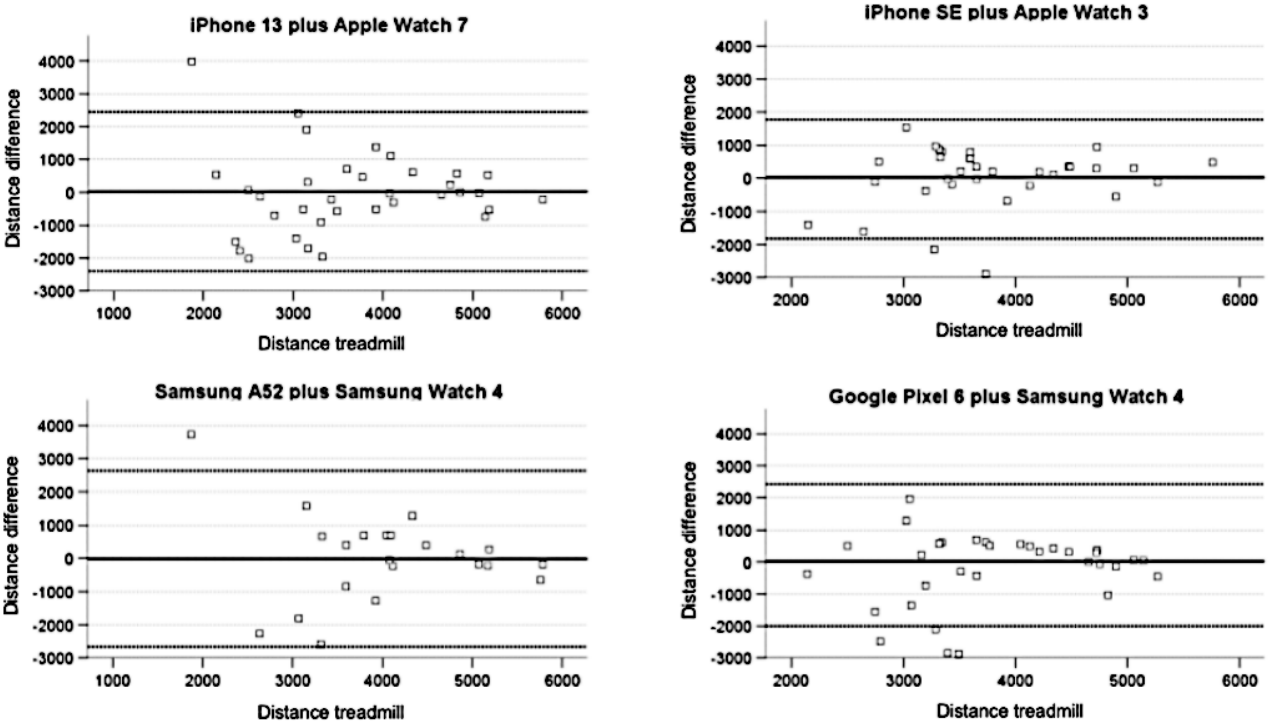


Figure 3. Bland-Altman plots for wearable measurement versus standard measurement: Bland-Altman plots for wearable distance measurements compared with treadmill measurements for 4 device combinations are depicted.



XSL•FO
RenderX

Table 3. Univariate and multivariate analyses for accuracy of heart rate measurement.

Variables		Univariate analysis		Multivariate analysis	
	Wearables (n=150)	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Baseline characteristics					
Age (years), median (range)	56 (22-78)	0.97 (0.94-0.99)	.03	N/A[a]	N/A
Sex, n (%)			.58	N/A	N/A
Female	50 (33)	1.26 (0.55-2.90)			
Male	100 (67)	Reference			
BMI (kg/m^2), median (range)	25 (16-40)	0.91 (0.84-0.99)	.03	0.90 (0.82-0.98)	.02
Risk assessment					
Risk Analysis Index score, n (%)			.22	N/A	N/A
≤15	50 (33)	Reference			
16-25	76 (51)	0.88 (0.39-2.11)			
≥26	24 (16)	3.74 (0.71-16.97)			
Timed Up and Go Test (seconds), n (%)			.009	N/A	.006
0-6	38 (25)	Reference	N/A	Reference	N/A
7-11	108 (72)	0.14 (0.03-0.61)	N/A	0.12 (0.03-0.55)	N/A
Hemoglobin (g/dl), median (range)	14 (9)	0.95 (0.73-1.24)	.70	N/A	N/A
Individual characteristics					
Resting heart rate[b] (bpm), median (range)	72 (53-118)	1.02 (0.99-1.05)	.30	N/A	N/A
Wrist circumference (cm), median (range)	17 (14-21)	0.77 (0.60-0.99)	.048	N/A	N/A
Skin humidity, n (%)			.08	N/A	N/A
Dry	90 (60)	Reference			
Humid	60 (40)	0.51 (0.23-1.10)			
Skin appendices, n (%)			.55	N/A	N/A
Blank	42 (28)	Reference			
Some hair	78 (52)	0.84 (0.33-2.16)			
Very hairy	30 (20)	0.55 (0.18-1.64)			
Borg scale, median (range)	10 (6-15)	0.82 (0.65-1.05)	.11	N/A	N/A
Numeric Rating Scale, median (range)	0 (0-9)	1.65 (0.80-3.41)	.18	N/A	N/A
VO_{2max}[c] (MET[d]), median (range)	3 (1-6)	1.54 (0.94-2.54)	.09	N/A	N/A

[a]N/A: not applicable.

[b]Accuracy of heart rate was defined as a mean absolute percentage error <5%.

[c]VO_{2max}: maximum rate of oxygen consumption.

[d]MET: metabolic equivalent.

Distance Measurements and Accuracy

Distance measurements of wearables compared with the treadmill revealed a higher variation, with mean differences ranging from –29.22 m (iPhone 13) to –104.28 m (iPhone SE) for the 6MWT (–31.67 m for Samsung A52 and –50.57 m for Google Pixel 6) and from –12.27 m (Samsung A52) to 208.15 m (Google Pixel 6) for the interval training (27.51 m for iPhone 13 and –24.15 m for iPhone SE). Consistently, the LoA were generally higher, with MAPE values exceeding 10% and a peak value of 30.61% for the iPhone SE.

Safety and Laboratory Assessments

The safety analysis of laboratory values, alarm settings of the app, and occurrence of AEs was conducted (Table 4). Lab analysis did not reveal any significant deviations in blood parameters, which remained within clinical limits. A detailed analysis from an exercise physiologist's view will be reported elsewhere. Awareness of alarm signals displayed on the wearables during the interval training was realized with a mean of 13.5 (SD 6.2) seconds. In total, 29 of 75 (39%) experienced an AE, of which 12 (16%) occurred during the interval training, comprising pain (5, 42%), skin irritation through ECG electrodes

XSL•FO
RenderX

(redness or vesicles; 2, 17%), dizziness (2, 17%), shortness of breath (2, 17%), and fatigue (1, 8%). During the follow-up at day 7, 3 of 15 (20%) participants still reported pain and 12 of 15 (80%) indicated skin irritations, which remained until the follow-up at day 30 in 2 cases. Notably, 14 events were related to the certified ECG (electrodes); however, these had nothing to do with the use of the software but had to be reported for regulatory reasons. There were no SAEs or device deficiencies that put the patient at an unacceptable risk. All AEs were periodically reported to the Bundesinstitut für Arzneimittel und Medizinprodukte according to the regulations of the MDR and the MPDG, and no AE lasted longer than 30 days. No SAE was reported during the trial, and no patient died until follow-up.

Table 4. Safety assessment of the app-based aerobic exercise training.

Variables	Overall (n=75)	*P* value
Laboratory values, differences median (range)[a]		
C-reactive protein (mg/dl)	0 (–2 to 0.2)	.41
Creatine kinase (U/l)	12 (–7 to 43)	<.001
Creatine kinase-muscle-brain isoform (U/l)	0 (–4 to 5)	.06
Lactate dehydrogenase (U/l)	12 (–132 to 69)	<.001
Glucose (mg/dl)	–4.5 (–121 to 26)	<.001
Sodium (mmol/l)	–2 (–11 to 12)	<.001
Potassium (mmol/l)	0 (–0.6 to 0.6)	.997
Chloride (mmol/l)	–2 (–9 to 9)	.01
Bicarbonate (mmol/l)	–0.35 (–11 to 7)	.29
Lactate (mmol/l)	0.04 (–3 to 7)	.47
Alarm setting of wearables		
Time of awareness seconds, mean (SD)	14 (6)	N/A[b]
Adverse events and overall survival, n (%)		
Overall adverse events	29 (39)	N/A
During interval training		
All events	12 (16)	N/A
Skin irritation through electrocardiogram electrodes	2 (17)	N/A
Pain	5 (42)	N/A
Dizziness	2 (17)	N/A
Shortness of breath	2 (17)	N/A
Fatigue	1 (8)	N/A
Day 7		
All events	15 (20)	N/A
Skin irritation through electrocardiogram electrodes	12 (80)	N/A
Pain	3 (20)	N/A
Day 30, n (%)		
All events	2 (3)	N/A
Skin irritation through electrocardiogram electrodes	2 (100)	N/A
Serious adverse event	0 (0)	N/A
Overall survival, n (%)	75 (100)	N/A

[a]Differences were calculated for blood samples taken before and after interval training (after-before).

[b]N/A: not applicable.

XSL•FO
RenderX

Discussion

Principal Findings

To the best of our knowledge, this is the first prospective trial report assessing the validity and safety of a risk-adjusted, app-based prehabilitation program using wearables for heart rate and physical activity monitoring. The analysis revealed a high accuracy of wearable heart rate measurement with a mean difference of ≤5 bpm and an MAPE of ≤5%, while distance measurement exhibited larger variations (mean difference ≤209 m and MAPE ≤31%), which can be explained through an impaired detection of patient's wrist movement during the exercise training on the treadmill and the lack of global positioning system trackers.

Prehabilitation in this context revealed a tremendous potential for improving modifiable factors in patients and thus increasing the safety and quality of surgical care. Perry et al [39] included 178 trials of heterogeneous quality on prehabilitation for patients undergoing major abdominal surgery, reporting prehabilitation to be superior to standard care in both length of hospital stay (1.81 days vs usual care) and postoperative pulmonary complications (risk ratio 0.55). This is consistent with the findings of the RCT by Barberan-Garcia et al [13] for postsurgical complications (31% vs 62%) in a cohort of 144 high-risk patients and with the meta-analysis of Moran et al [40] (OR 0.59, *P*=.03). Importantly, the study by Molenaar et al [12], which involved 251 patients undergoing resection of colorectal cancer, reported the benefit of a 4-week in-hospital multimodal prehabilitation program with significantly reduced severe postoperative complications (Comprehensive Complication Index>20: 17.1% vs 29.7%), medical complications (15.4% vs 27.3%), and enhanced functional capacity. At this stage, it seems that a resource-intensive prehabilitation program only works in clinical trials due to the enormous infrastructure requirements and costs associated with it, including patient travel over weeks, which seems hardly feasible. This suggests a real opportunity for a well-guided and safe remote approach that additionally fosters patient empowerment [41,42].

Studies on costs also suggest that an improvement in surgical outcomes, particularly the reduction in hospital length of stay, results in a significant reduction in hospital expenditures and the opportunity cost of surgery [4,5]. A more recent meta-analysis by Punnoose et al [43] of patients undergoing orthopedic surgery found a significant impact of prehabilitation on preoperative function, muscle strength, and reduction of pain. However, the current evidence on prehabilitation remains limited, characterized by broad heterogeneity between studies and a high risk of bias. Particularly, exercise regimens were not consistently reported in terms of the individual content of the exercises (ie, training type, intensity, duration) and a lack of compliance to high-intensity training substantially impairs the implementation of prehabilitation in clinical practice [11,15].

Recent studies evaluating digital prehabilitation models provide increasing evidence for their beneficial effects compared with conventional hospital-based in-person programs [14,15,44,45]. Moorthy et al [45] assessed the feasibility of a home-based supervised prehabilitation program involving 41 patients scheduled for esophago-gastric cancer surgery, revealing a higher program completion in the digital arm compared with the in-person cohort (84% vs 14%). Further, patients using the digital service exhibited increased cardiorespiratory fitness measured by the 30-second sit-to-stand test (*P*=.02) and the median heart rate recovery (10.5-15.5 bpm; *P*=.04). Similar positive effects on patients' adherence to digital exercise programs were found by Kadiri et al [46], who evaluated the efficacy of the "Fit 4 Surgery" app, which included 10 exercises for patients undergoing lung cancer surgery. Patients in the app cohort completed more exercise sessions (9 vs 2) in a shorter time (24 vs 45 days) before surgery and even during the postsurgical period (2 vs 0), compared with those attending hospital- or community-based classes. Importantly, a quality improvement study by Kimura et al [47] that assessed the impact of an online app-based prehabilitation program (SeamlessMD) on surgical outcomes after colorectal surgery demonstrated a significantly shorter length of hospital stay (adjusted estimates –1.15; *P*=.03), while there were no differences in terms of complication rates.

A remote prehabilitation program will empower patients' autonomy, alleviate pressure on health care workers and infrastructure, and eliminate unnecessary travel for patients. An oncologic diagnosis, coupled with the clear short-term goal of becoming the best version of oneself before a major surgical procedure, serves as a compelling motivator for patients and is a key driver of compliance and adherence to digital therapeutics [11]. Prehabilitation models incorporating remote monitoring with wearables for heart rate, physical activity, and progress tracking have been shown to support personalized goal setting, ensure safety through professional supervision, and enhance patient adherence to the program [15,17]. Various studies have assessed wearable heart rate accuracy by comparing them with the gold-standard reference method of ECG, indicating a low overall error, low mean difference SD, and acceptable agreement with ECG for consumer wearables (eg, Apple Watch or Fitbit) [18,48-50]. Wang et al [35] assessed the accuracy of 4 different wrist-worn wearables in 50 healthy students, showing median differences of ≤10 bpm, which were generally lower at rest compared with measurements during exercise. However, despite the acceptable accuracy in laboratory settings, real-time measurement in medical settings, including an assessment of Apple Watch 3 and Fitbit Charge 2 over 24 hours compared with ECG by Nelson et al [36], indicated that single heart rate observation could be inaccurate by significantly large margins, although aggregated values revealed high overall accuracy (<10% MAPE) [51].

The study has both strengths and weaknesses that will be addressed in the following. One of the study's strengths is that it follows a strict regulatory framework, supported by dedicated quality and risk management. All criteria assessed here were derived from risk management principles, explaining the stepwise approach that began with a first-in-human pilot study involving healthy students and patients within the intended use. The study was split into a pure usability part and an evaluation of the risk assessment, which will be reported elsewhere [26]. Another strength is the use of evidence-based risk calculation

XSL•FO
RenderX

tools. The RAI score emerged as the most suitable and user-friendly score, not requiring additional tools for assessment. It, combined with the TUG and ECOG scores, best met the requirements of a digital solution. Moreover, it is currently highlighted as the so-called surgical pause by the American College of Surgeons [52]. Other scores, such as the National Surgical Quality Improvement Program (NSQIP), require significantly more resources and have not demonstrated superiority in assessments, for example, within a German health care setting [53,54].

A weakness of the study may be that it is not a randomized controlled trial. However, given that we are establishing evidence in a young field of clinical medicine, safety and validity were the first steps in the stepwise approach. This was followed by a multicentric clinical trial that is currently underway. Comparability at this stage of remote prehabilitation is challenging due to the variety of methods used to implement it. van der Velde et al [55] performed a pilot randomized controlled trial in which they used a smartphone-based app to achieve behavioral changes before surgery. They found self-reported changes in alcohol use, nicotine cessation, and an increase in self-reported exercising. Wang and coworkers [14] developed an app to foster prehabilitation and assessed its usability. While they found promising usability results with their device, no efficacy data have been presented so far. Although the ideas of the 2 previously mentioned studies were only somewhat comparable to ours, the solutions lacked supervision of the exercises, making it unclear whether the patients engaged in light or moderate to vigorous exercise, which has significantly larger effects. Haveman et al [56] followed a similar approach to ours but differed by tracking individuals' activities and interpreting increases in heart rate and step counts 3 days before surgery as indicators of increased physical activity, without obtaining a specific readout from the assessed data. This shows that remote prehabilitation seems feasible, is being explored, and may be a hot topic for the next few years, as outlined in various arguments and discussion points previously [57].

Another challenge and critical endpoint measurement is compliance and adherence to the program provided. In this feasibility study and first-in-human use of a newly designed medical device, compliance was not an endpoint. However, the development of The Prehab App involved patients, doctors, and other experts from the very beginning [41]. The program was designed in a way that allows different risk groups to train within a general interval training framework, respecting their individual risk and adapting the individual thresholds of exercising to their profile [11,25,26]. During the randomized controlled trial, the Prehab App will reliably measure compliance using a traffic light system, as it documents every exercise in real time and provides precise information on how patients trained within their thresholds. Additionally, we will assess if the participants performed their exercises and if they took their supplementary nutrition using patient-reported outcomes. Reminders will be used to increase compliance and adherence.

All this together highlights the potential of digital prehabilitation interventions in achieving high levels of patient engagement, uptake, adherence, and usability satisfaction. App-based models provide a promising solution to overcome the limitations in scalability and sustainability seen with face-to-face programs in clinical practice. However, the application of digital prehabilitation models needs to be further assessed in multicenter randomized controlled settings to evaluate their effectiveness, particularly in terms of safety and efficacy in the final intended use, and to gain approval as a safe medical device [51,58].

Conclusions

This analysis presents the first pilot study assessing the validity and safety of a risk-stratified prehabilitation program for patients undergoing major elective surgery, utilizing an mHealth app linked to wearable devices for self- and remote monitoring of physical performance. The heart rate measurement was below the recommended threshold with an MAPE of ≤5%, demonstrating excellent accuracy. Laboratory analysis and assessment of AEs indicated that individually tailored exercise training can be safely performed in a supervised setting, justifying the remote testing of The Prehab App in a multicentric randomized controlled trial.

Use of Artificial Intelligence

No artificial intelligence has been used in the generation of the manuscript or any text presented here.

Acknowledgments

The Else Kroener Fresenius Foundation funded the clinical trial through the translational research program 2021_EKTP10, providing a total of €161,000 (US $168,821). JF was also supported by this grant for a different part of the clinical trial. The PROTEGO MAXIMA trial was a tremendous effort of many people. We must thank Mark Siller and his software development team at Seven D in Heilbronn, PD Dr. Georg Matheis for his support with medical device questions, and Jan von Hassel for his expertise and guidance in regulatory questions, including qualifying the whole team as MDR- and DIN ISO 14155-compliant investigators. Additionally, we extend our thanks to the Johner Institute team, especially Anne Arndt, for her support in MDR-compliant development and support. We also thank Claus Wuestenhagen for expert data safety and protection support. Our special thanks go to the Center for Clinical Trials at Frankfurt University Hospital, which supported us with independent monitoring conducted by Ina Müller. We also thank the Executive Department of Quality Management, led by Dr. Kyra Schneider, for their assistance with quality management questions. We are grateful to Professor Michael Behringer at the Institute for Sports Sciences at Goethe University for his cooperation. Finally, we are proud that the PROTEGO MAXIMA trial was the highest recruiting investigator-initiated trial in the Comprehensive Cancer Center Frankfurt/Main in 2022, with a total of 392 patients and participants across 3 modules.

XSL•FO
RenderX

Data Availability

Data can be accessed anonymously upon specific request and submission of the requesting group’s research question outline.

Authors' Contributions

Study conception and design were led by AAS, DZ, CD, JF, LM, JD, and NF. Data acquisition was conducted by SFFUZ, SS, AAS, JF, LM, CD, JD, FC, WD, and WS. Analysis and interpretation of data were carried out by SFFUZ, SS, AAS, NF, DZ, WOB, JF, FC, WD, and WS. Writing of the manuscript was done by SFFUZ, AAS, JF, LM, JD, and NF with input and final approval of all authors.

Conflicts of Interest

AAS, DZ, and CD are founders of Capreolos GmbH, a spin-off from Goethe University established to set up the regulatory framework for developing a medical device for prehabilitation. JF became a shareholder of Capreolos after the termination of the trial. All other authors declare that they have no direct or indirect conflicts of interest.

Multimedia Appendix 1
The STARD (Standards for Reporting of Diagnostic Accuracy) checklist.
[DOCX File , 26 KB - mhealth_v13i1e55298_app1.docx]

Multimedia Appendix 2
Display overview of the Patronus Prehab App: (A) overview of the structured risk assessment; (B) overview of the aerobic interval training.
[PNG File , 638 KB - mhealth_v13i1e55298_app2.png]

Multimedia Appendix 3
Flow diagram of the interventional PROTEGO MAXIMA trial. Structured risk assessment of patients undergoing major elective surgery using the Prehab App and calculation of an individual exercise program. 6MWT and interval training are performed on a treadmill in a supervised setting being connected to an ECG and wearing wearable devices. A follow-up after surgery is performed on days 7 and 30. 6MWT: 6-minute walking test; ECOG: Eastern Cooperative Oncology Group; Hb: hemoglobin; RAI: Risk Analysis Index; TUG: Timed Up and Go-Test.
[PNG File , 270 KB - mhealth_v13i1e55298_app3.png]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 255 KB - mhealth_v13i1e55298_app4.pdf]

Multimedia Appendix 5
Comparison of heart rate and distance measurement between smartwatch versus electrocardiogram (ECG) according to healthy students and patients.
[DOCX File , 15 KB - mhealth_v13i1e55298_app5.docx]

Multimedia Appendix 6 [PNG File , 93 KB - mhealth_v13i1e55298_app6.png]

References

1. Nepogodiev D, Martin J, Biccard B, Makupe A, Bhangu A, Nepogodiev D, et al. Global burden of postoperative death. The Lancet 2019 Feb;393(10170):401. [doi: 10.1016/s0140-6736(18)33139-8]
2. World Health Organization (WHO). WHO Guidelines for Safe Surgery: Safe Surgery Saves Lives. Geneva, Switzerland: World Health Organization; 2009.
3. Tevis S, Cobian A, Truong H, Craven M, Kennedy G. Implications of multiple complications on the postoperative recovery of general surgery patients. Ann Surg 2016;263(6):1213-1218. [doi: 10.1097/sla.0000000000001390]
4. Fuertes-Guiró F, Viteri Velasco E. The impact of frailty on the economic evaluation of geriatric surgery: hospital costs and opportunity costs based on meta-analysis. J Med Econ 2020 Aug 20;23(8):819-830 [FREE Full text] [doi: 10.1080/13696998.2020.1764965] [Medline: 32372679]
5. Barberan-Garcia A, Ubre M, Pascual-Argente N, Risco R, Faner J, Balust J, et al. Post-discharge impact and cost-consequence analysis of prehabilitation in high-risk patients undergoing major abdominal surgery: secondary results from a randomised controlled trial. Br J Anaesth 2019 Oct;123(4):450-456 [FREE Full text] [doi: 10.1016/j.bja.2019.05.032] [Medline: 31248644]

XSL•FO
RenderX

6. Carli F, Baldini G, Feldman LS. Redesigning the preoperative clinic: from risk stratification to risk modification. JAMA Surg 2021 Mar 01;156(2):191-192. [doi: 10.1001/jamasurg.2020.5550] [Medline: 33295945]
7. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5(6):649-656. [doi: 10.1097/00000421-198212000-00014]
8. Robinson TN, Wu DS, Sauaia A, Dunn CL, Stevens-Lapsley JE, Moss M, et al. Slower walking speed forecasts increased postoperative morbidity and 1-year mortality across surgical specialties. Ann Surg 2013 Oct;258(4):582-8; discussion 588 [FREE Full text] [doi: 10.1097/SLA.0b013e3182a4e96c] [Medline: 23979272]
9. Hall DE, Arya S, Schmid KK, Blaser C, Carlson MA, Bailey TL, et al. Development and initial validation of the Risk Analysis Index for measuring frailty in surgical populations. JAMA Surg 2017 Mar 01;152(2):175-182 [FREE Full text] [doi: 10.1001/jamasurg.2016.4202] [Medline: 27893030]
10. Carli F, Silver JK, Feldman LS, McKee A, Gilman S, Gillis C, et al. Surgical prehabilitation in patients with cancer: state-of-the-science and recommendations for future research from a panel of subject matter experts. Phys Med Rehabil Clin N Am 2017 Mar;28(1):49-64. [doi: 10.1016/j.pmr.2016.09.002] [Medline: 27913000]
11. Sliwinski S, Werneburg E, Faqar-Uz-Zaman SF, Detemble C, Dreilich J, Mohr L, et al. A toolbox for a structured risk-based prehabilitation program in major surgical oncology. Front Surg 2023 Jun 26;10:1186971-1186913 [FREE Full text] [doi: 10.3389/fsurg.2023.1186971] [Medline: 37435472]
12. Molenaar CJL, Minnella EM, Coca-Martinez M, Ten Cate DWG, Regis M, Awasthi R, PREHAB Study Group. Effect of multimodal prehabilitation on reducing postoperative complications and enhancing functional capacity following colorectal cancer surgery: the PREHAB randomized clinical trial. JAMA Surg 2023 Jun 01;158(6):572-581 [FREE Full text] [doi: 10.1001/jamasurg.2023.0198] [Medline: 36988937]
13. Barberan-Garcia A, Ubré M, Roca J. Personalised prehabilitation in high-risk patients undergoing elective major abdominal surgery: a randomized blinded controlled trial. Ann Surg 2018;267(1):A. [doi: 10.1097/sla.0000000000002293]
14. Wang T, Stanforth PR, Fleming RYD, Wolf JS, Stanforth D, Tanaka H. A mobile app with multimodality prehabilitation programs for patients awaiting elective surgery: development and usability study. JMIR Perioper Med 2021 Dec 30;4(2):e32575 [FREE Full text] [doi: 10.2196/32575] [Medline: 34967752]
15. Barberan-Garcia A, Cano I, Bongers BC, Seyfried S, Ganslandt T, Herrle F, et al. Digital support to multimodal community-based prehabilitation: looking for optimization of health value generation. Front Oncol 2021 Jun 17;11:662013-662017 [FREE Full text] [doi: 10.3389/fonc.2021.662013] [Medline: 34249698]
16. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
17. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020 Jan 13;3(1):4-6 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
18. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020 Jun 26;3(1):90-99 [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
19. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun 02;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]
20. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.). European Union. 2017. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32017R0745 [accessed 2024-12-27]
21. International Organization for Standardization (ISO). Clinical investigation of medical devices for human subjects — good clinical practice (ISO 14155:2020). ISO. 2020. URL: https://www.iso.org/standard/71690.html [accessed 2024-12-29]
22. Medical Device Law Implementation Act (Medizinproduktedurchführungsgesetz; MPDG). Medical Device Innovation Center. URL: https://www.mic-mainz.de/en/mpdg-2/ [accessed 2024-12-27]
23. Bossuyt P, Reitsma J, Bruns D, Gatsonis CA, Glasziou PP, Irwig L, STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015 Oct 28;351:h5527 [FREE Full text] [doi: 10.1136/bmj.h5527] [Medline: 26511519]
24. Martin D, Mantziari S, Demartines N, Hübner M, Study Group E. Defining major surgery: a Delphi consensus among European Surgical Association (ESA) members. World J Surg 2020 Jul 14;44(7):2211-2219 [FREE Full text] [doi: 10.1007/s00268-020-05476-4] [Medline: 32172309]
25. Faqar-Uz-Zaman SF, Sliwinski S, Detemble C, Filmann N, Zmuc D, Mohr L, et al. Study protocol for a pilot trial analysing the usability, validity and safety of an interventional health app programme for the structured prehabilitation of patients before major surgical interventions: the PROTEGO MAXIMA trial. BMJ Open 2023 Apr 05;13(4):e069394 [FREE Full text] [doi: 10.1136/bmjopen-2022-069394] [Medline: 37019492]
26. Sliwinski S, Faqar-Uz-Zaman SF, Heil J, Mohr L, Detemble C, Dreilich J, et al. Predictive value of a novel digital risk calculator to determine early patient outcomes after major surgery: a proof-of-concept pilot study. Patient Saf Surg 2024 Apr 12;18(1):13-12 [FREE Full text] [doi: 10.1186/s13037-024-00395-y] [Medline: 38610002]

XSL•FO
RenderX

27. Camarda SRDA, Tebexreni AS, Páfaro CN, Sasai FB, Tambeiro VL, Juliano Y, et al. Comparison of maximal heart rate using the prediction equations proposed by Karvonen and Tanaka. Arq Bras Cardiol 2008 Nov;91(5):311-314 [FREE Full text] [doi: 10.1590/s0066-782x2008001700005] [Medline: 19142375]
28. International Organization for Standardization (ISO). Medical devices—quality management systems—requirements for regulatory purposes (ISO 13485:2016). ISO. 2016. URL: https://www.iso.org/standard/59752.html [accessed 2024-12-27]
29. Borg GA. Psychophysical bases of perceived exertion. Medicine & Science in Sports & Exercise 1982;14(5):377???381. [doi: 10.1249/00005768-198205000-00012]
30. Amir M, Kabo P, Mappangara I, Djafar Z, Zainuddin AA, Warliani M, et al. Peak oxygen uptake and metabolic equivalents explained by six-minute walk test: a prospective observational study in predicting heart failure patient readmission. Ann Med Surg (Lond) 2022 May;77:103652 [FREE Full text] [doi: 10.1016/j.amsu.2022.103652] [Medline: 35638025]
31. Dourado V, Nishiaka R, Simões MSMP, Lauria V, Tanni S, Godoy I, et al. Classification of cardiorespiratory fitness using the six-minute walk test in adults: comparison with cardiopulmonary exercise testing. Pulmonology 2021 Nov;27(6):500-508 [FREE Full text] [doi: 10.1016/j.pulmoe.2021.03.006] [Medline: 33958319]
32. International Organization for Standardization (ISO). Information security, cybersecurity and privacy protection—information security management systems—requirements (ISO/IEC 27001:2022). ISO. 2022. URL: https://www.iso.org/standard/27001 [accessed 2024-12-27]
33. Bundesamt für Sicherheit in der Informationstechnik (BSI). URL: https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/informationen-und-empfehlungen_node.html [accessed 2024-12-27]
34. Faqar-Uz-Zaman S, Anantharajah L, Baumartz P, Sobotta P, Filmann N, Zmuc D, et al. The diagnostic efficacy of an app-based diagnostic health care application in the emergency room: eRadaR-trial. A prospective, double-blinded, observational study. Ann Surg 2022 Nov 01;276(5):935-942. [doi: 10.1097/SLA.0000000000005614] [Medline: 35925755]
35. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2017 Jan 01;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
36. Nelson MB, Kaminsky LA, Dickin DC, Montoye AHK. Validity of consumer-based physical activity monitors for specific activity types. Med Sci Sports Exerc 2016 Aug;48(8):1619-1628. [doi: 10.1249/MSS.0000000000000933] [Medline: 27015387]
37. Bunn J, Navalta JW, Fountaine CJ, Reece J. Current state of commercial wearable technology in physical activity monitoring 2015-2017. Int J Exerc Sci 2018;11(7):503-515 [FREE Full text] [Medline: 29541338]
38. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
39. Perry R, Herbert G, Atkinson C, England C, Northstone K, Baos S, et al. Pre-admission interventions (prehabilitation) to improve outcome after major elective surgery: a systematic review and meta-analysis. BMJ Open 2021 Sep 30;11(9):e050806 [FREE Full text] [doi: 10.1136/bmjopen-2021-050806] [Medline: 34593498]
40. Moran J, Wilson F, Guinan E, McCormick P, Hussey J, Moriarty J. Role of cardiopulmonary exercise testing as a risk-assessment method in patients undergoing intra-abdominal surgery: a systematic review. Br J Anaesth 2016 Mar;116(2):177-191 [FREE Full text] [doi: 10.1093/bja/aev454] [Medline: 26787788]
41. Schnitzbauer AA, Zmuc D, Fleckenstein J. Digital prehabilitation—a solution to resource shortages? The Lancet Digital Health 2024 Jan;6(1):e11. [doi: 10.1016/s2589-7500(23)00247-9]
42. Leitlinienprogramm O, Deutsche K, AWMF): P, Langversion 1. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF): Perioperatives Management bei gastrointestinalen Tumoren (POMGAT), Langversion 1.0, 2023, AWMF-Registernummer: 088-010OL. Office des Leitlinienprogrammes Onkologie. 2023. URL: https://www.leitlinienprogramm-onkologie.de/leitlinien/perioperatives-managementbei-gastrointestinalen-tumoren-pomgat/, [accessed 2024-12-25]
43. Punnoose A, Claydon-Mueller LS, Weiss O, Zhang J, Rushton A, Khanduja V. Prehabilitation for patients undergoing orthopedic surgery: a systematic review and meta-analysis. JAMA Netw Open 2023 Apr 03;6(4):e238050 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.8050] [Medline: 37052919]
44. Kimura F, Miyazaki M, Suwa T, Sugiura T, Shinoda T, Itoh H, et al. Evaluation of total hepatic vascular exclusion and pringle maneuver in liver resection. Hepatogastroenterology 2002;49(43):225-230. [Medline: 11941961]
45. Moorthy K, Halliday LJ, Noor N, Peters CJ, Wynter-Blyth V, Urch CE. Feasibility of implementation and the impact of a digital prehabilitation service in patients undergoing treatment for oesophago-gastric cancer. Curr Oncol 2023 Jan 30;30(2):1673-1682 [FREE Full text] [doi: 10.3390/curroncol30020128] [Medline: 36826089]
46. Kadiri SB, Kerr AP, Oswald NK, Budacan A, Flanagan S, Golby C, et al. Fit 4 surgery, a bespoke app with biofeedback delivers rehabilitation at home before and after elective lung resection. J Cardiothorac Surg 2019 Jul 05;14(1):132 [FREE Full text] [doi: 10.1186/s13019-019-0951-6] [Medline: 31277671]
47. Kimura CS, Bidwell S, Gurland B, Morris A, Shelton A, Kin C. Association of an online home-based prehabilitation program with outcomes after colorectal surgery. JAMA Surg 2023 Jan 01;158(1):100-102. [doi: 10.1001/jamasurg.2022.4485] [Medline: 36322070]

XSL•FO
RenderX

48. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
49. Boudreaux B, Hebert E, Hollander D, Williams BM, Cormier CL, Naquin MR, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
50. Stahl SE, An H, Dinkel DM, Noble JM, Lee J. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016 Apr 25;2(1):e000106 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
51. Martín-Escudero P, Cabanas AM, Dotor-Castilla ML, Galindo-Canales M, Miguel-Tobal F, Fernández-Pérez C, et al. Are activity wrist-worn devices accurate for determining heart rate during intense exercise? Bioengineering (Basel) 2023 Mar 15;10(2):254 [FREE Full text] [doi: 10.3390/bioengineering10020254] [Medline: 36829748]
52. Guidance for industry, investigating out-of-specification (OOS), test results for pharmaceutical production. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. 2006 Oct. URL: https://www.fda.gov/downloads/drugs/guidances/ucm070287.pdf [accessed 2024-12-25]
53. Hatipoglu E, Erginoz E, Askar A, Erguney S. Accuracy of the ACS NSQIP Surgical Risk Calculator for Predicting Postoperative Complications in Gastric Cancer Following Open Gastrectomy. The American Surgeon™ 2023 Oct 12:640-647. [doi: 10.1177/00031348231206581]
54. Höhn P, Runde F, Luu A, Fahlbusch T, Fein D, Klinger C, members of StuDoQ|Pancreas registry of the German Society for GeneralVisceral Surgery (DGAV), et al. Applicability of the surgical risk calculator by the American College of Surgeons in the setting of German patients undergoing complete pancreatectomy: multicentre study using data from the StuDoQ|Pancreas registry. BJS Open 2023 Mar 07;7(2):zrac164 [FREE Full text] [doi: 10.1093/bjsopen/zrac164] [Medline: 36882081]
55. van der Velde M, Valkenet K, Geleijn E, Kruisselbrink M, Marsman M, Janssen LM, et al. Usability and preliminary effectiveness of a preoperative mHealth app for people undergoing major surgery: pilot randomized controlled trial. JMIR Mhealth Uhealth 2021 Jan 07;9(1):e23402 [FREE Full text] [doi: 10.2196/23402] [Medline: 33410758]
56. Haveman ME, van Melzen R, El Moumni M, Schuurmann RCL, Hermens HJ, Tabak M, et al. Determining the reliable measurement period for preoperative baseline values with telemonitoring before major abdominal surgery: pilot cohort study. JMIR Perioper Med 2022 Nov 28;5(1):e40815 [FREE Full text] [doi: 10.2196/40815] [Medline: 36441586]
57. Blumenau Pedersen M, Saxton J, Birch S, Rasmussen Villumsen B, Bjerggaard Jensen J. The use of digital technologies to support home-based prehabilitation prior to major surgery: a systematic review. Surgeon 2023 Dec;21(6):e305-e315. [doi: 10.1016/j.surge.2023.05.006] [Medline: 37336649]
58. Wiley L, Terry N. Liability for mobile health and wearable technologies. Annals of Health Law 2016;25:62-97 [FREE Full text]

Abbreviations

6MWT: 6-minute walking test
AE: adverse event
bpm: beats per minute
BSI: Bundesamt für Sicherheit in der Informationstechnik
DIN: Deutsches Institut für Normung
ECG: electrocardiogram
ECOG: Eastern Cooperative Oncology Group
ISO: International Organization for Standardization
LoA: limits of agreement
MAPE: mean absolute percentage error
MDR: Medical Device Regulation
MET: metabolic equivalent
MPDG: Medizinproduktedurchführungsgesetz
NSQIP: National Surgical Quality Improvement Program
QLQC30: 30-item Quality-of-Life Questionnaire
RAI: Risk Analysis Index
SAE: serious adverse event
STARD: Standards for Reporting of Diagnostic Accuracy
TUG: Timed Up and Go Test
VO2max: maximum rate of oxygen consumption

XSL•FO
RenderX

Edited by M Sobolev; submitted 08.12.23; peer-reviewed by S Mitra, T Vilz; comments to author 05.05.24; revised version received 25.06.24; accepted 06.11.24; published 10.02.25.

Please cite as:
Faqar Uz Zaman SF, Sliwinski S, Mohr-Wetzel L, Dreilich J, Filmann N, Detemble C, Zmuc D, Chun F, Derwich W, Schreiner W, Bechstein W, Fleckenstein J, Schnitzbauer AA
Validity, Accuracy, and Safety Assessment of an Aerobic Interval Training Using an App-Based Prehabilitation Program (PROTEGO MAXIMA Trial) Before Major Surgery: Prospective, Interventional Pilot Study
JMIR Mhealth Uhealth 2025;13:e55298
URL: https://mhealth.jmir.org/2025/1/e55298
doi:10.2196/55298
PMID:

©Sara Fatima Faqar Uz Zaman, Svenja Sliwinski, Lisa Mohr-Wetzel, Julia Dreilich, Natalie Filmann, Charlotte Detemble, Dora Zmuc, Felix Chun, Wojciech Derwich, Waldemar Schreiner, Wolf Bechstein, Johannes Fleckenstein, Andreas A Schnitzbauer. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Integrating a Multimodal Digital Device for Continuous Perioperative Monitoring in Patients With Lung Cancer Undergoing Thoracic Surgery: Development and Usability Study

Runchen Wang[1*], MD; Jianqi Zheng[1*], MD; Wenwei Guo[1*], MD; Haiqi Huang[1*], MD; Qixia Wang[2], MD; Yihong Li[1], MD; Manwan Lin[1], MD; Linchong Huang[1], MD; Qing Zhang[1], MD; Kaishen Chen[3], BA; Zhiming Ye[1], MD; Hongsheng Deng[1], MD; Yu Jiang[1], MD; Yuechun Lin[1], MD; Yi Feng[1], MD; Ying Huang[1], MD; Ying Chen[1], MD; Jianxing He[1], MD, PhD; Hengrui Liang[1], MD, PhD

[1]Department of Thoracic Surgery, China State Key Laboratory of Respiratory Disease & National Clinical Research Center for Respiratory Disease, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

[2]Department of Respiratory and Critical Care Medicine, Guangzhou Institute of Respiratory Health, State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, National Center for Respiratory Medicine, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

[3]CinoCore AI department, Shanghai CinoCore Health Technology Co Ltd, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Hengrui Liang, MD, PhD
Department of Thoracic Surgery, China State Key Laboratory of Respiratory Disease & National Clinical Research Center for Respiratory Disease
the First Affiliated Hospital of Guangzhou Medical University
191 Yanjiang Road
Guangzhou, 510120
China
Phone: 86 02083062810
Email: hengrui_liang@163.com

Abstract

Background: Minimally invasive thoracic surgery has improved lung cancer outcomes but requires enhanced postoperative care. Traditionally, the episodic care model has limited timely and multidimensional monitoring of patients. Recent technological advances in multimodal digital devices, including wearable devices and electronic patient-reported outcomes (ePROs), offer a promising solution to these challenges. However, current studies focus on only a few parameters and limited application in thoracic surgery.

Objective: This study aims to propose a self-controlled study to evaluate the feasibility and reliability of multimodal digital devices, including wearables and ePROs, for continuous perioperative monitoring to enhance recovery after thoracic surgery.

Methods: We included 288 patients with non–small cell lung cancer from the Guangzhou Medical University cohort, which includes 2757 participants with various lung diseases. Digital data were collected during hospitalization using a commercial smartwatch combined with an ePROs questionnaire, while clinical data were obtained from electronic health records (EHRs). Agreement between the digital device and EHR was evaluated via Bland-Altman analysis. Time-series data were normalized for continuous outlier monitoring, and threshold analysis of ePROs scores were used to explore associations across different modules.

Results: Throughout hospitalization, digital devices provided a subjective overview of the patients' recovery trajectories. Results of Bland-Altman analysis demonstrated a high level of agreement between the digital device and the EHR. For body temperature, the analysis revealed a minimal bias of 0.02 °C (95% CI –0.01 °C to 0.05 °C), the agreement for heart rate showed a bias of 0.26 beats per minute (bpm; 95% CI –0.49 bpm to 1.01 bpm), and the bias for oxygen saturation was –0.06% (95% CI –0.27% to 0.15%), indicating close alignment between the 2 measurement methods. Meanwhile, wearable devices demonstrate significant potential in outlier detection compared to the episodic care model, offering accurate and sensitive monitoring of outliers between traditional measurement intervals. Using a thresholding method, we found that wearable metrics were correlated with the severity of ePROs.

XSL•FO
RenderX

Conclusions: These findings highlight the reliability and clinical potential of digital device–based multimodal systems within the enhanced recovery after surgery framework, offering a novel approach for continuous perioperative monitoring.

(JMIR Mhealth Uhealth 2025;13:e69512) doi:10.2196/69512

KEYWORDS

lung cancer; digital device; wearable device; patients reported outcomes; multi-modal; artificial intelligence; AI

Introduction

Background

Minimally invasive thoracic surgery (MITS) has significantly improved perioperative outcomes for patients with lung cancer, offering benefits such as reduced postoperative pain, shorter hospital stays, and quicker recovery times [1]. However, the success of MITS also demands higher standards for postoperative care, with a focus on early mobilization, increased physical activity, and personalized rehabilitation [2,3]. Building upon the foundation of smaller incisions, the enhanced recovery after surgery (ERAS) protocol further emphasizes both the speed and quality of recovery [4].

Currently, the commonly used perioperative monitoring model in clinical practice is episodic care, such as bedside monitoring, periodic assessment, and ward rounds. However, this care model limits patients' activities and timely feedback on subjective feelings. Moreover, some abnormal signs of intermittent attacks may be missed due to the discontinuity of monitoring. Therefore, a novel pattern to obtain multimodal, comprehensive, real-time, continuous, and reliable data of patient condition is urgently needed to achieve ERAS.

Recent technological advances in digital devices, including wearable devices and electronic patient-reported outcomes (ePROs) systems, offer a promising solution to these challenges [5,6]. Wearable devices, such as smartwatches and smart rings, can be used for continuous health monitoring, providing a more comprehensive assessment of an individual's well-being [7]. ePROs, in turn, can obtain real-time subjective feedback from patients. Previous studies have demonstrated that ePROs can reduce postoperative symptom burden and complications [8-10] and have shown their prognostic and therapeutic benefits [11].

Despite these advances, several challenges still remain. First, most studies focus on the continuous tracking of only a few parameters, predominantly heart rate (HR) and step count. Second, previous research has mainly concentrated on cardiovascular disease, as HR monitoring is relatively straightforward. Finally, there is a scarcity of studies combining wearable devices with ePROs for use in clinical practice, particularly in thoracic surgery.

Objective

To address the gaps mentioned earlier, we propose a prospective, self-controlled study to explore the feasibility and reliability of using multimodal digital devices, including wearable technology combined with ePROs, and provide a novel evidence-based approach to continuous perioperative monitoring for enhanced recovery after thoracic surgery.

Methods

Ethical Considerations

This prospective study received approval from the Ethics Committee of the National Center for Respiratory Medicine and the First Affiliated Hospital of Guangzhou Medical University (ES-2024-073-02) and was registered with the clinical trial registry (NCT06118229). All data incorporated in this study were deidentified. As compensation, participants in this study received regular follow-up visits and one-on-one counselling services provided by the medical team upon completion of the study. This study followed the Declaration of Helsinki.

Patients Selection

All patients were drawn from the Guangzhou Medical University cohort, a prospective, multicenter cohort designed to use multimodal data, including intelligent digital data, biospecimen data, and imaging artificial intelligence data from 2757 participants with various lung diseases, to assist in diagnosis, monitor treatment efficacy, and promote the rapid recovery of surgical patients. In this study, we prospectively and consecutively recruited patients with non–small cell lung cancer, one of the study arms within the Guangzhou Medical University cohort, aged ≥18 years, who were scheduled to undergo thoracic surgery between March 2023 and December 2023.

Detailed exclusion criteria were as follows: (1) patients unable to wear the wearable device or have poor compliance; (2) patients with a previous history of trauma, infection, or tumor; (3) patients without a smartphone equipped with an Android (Google LLC) or iOS (Apple Inc) operating system; (4) patients with other systemic inflammatory or neurological diseases; and (5) pregnant or lactating women.

All eligible patients provided informed consent after receiving detailed written information about the study, including the purpose of the study and use of wearable devices. Both verbal and written informed consent was obtained.

Wearable Device

The HUAWEI WATCH D (Huawei Device Co, Ltd) is a wrist-worn class II medical device approved by the Chinese National Medical Products Administration and certified under the EU Medical Device Regulation; it combines a micromechanical air pump, the TruBP pressure module, and temperature, photoplethysmography (PPG), and electrocardiography sensors to deliver continuous perioperative monitoring of body temperature, HR, oxygen saturation, and blood pressure. To limit motion- and wear-related artifacts, the watch uses an upgraded TruSeen 5.0+ optical system with 8 circumferential photodiodes and dual LEDs, a Fresnel film, and

XSL•FO
RenderX

a curved 2.5-D sapphire back with physical vapor deposition electrodes that improve skin contact; abnormal strap loosening or compression triggers an automatic hardware repeat-measurement and refusal protocol. Raw signals were processed in real time by a HiFi-Encoder deep learning pipeline that performs multichannel enhancement and concurrent time and frequency noise suppression, discarding any frame whose signal-quality index falls below 0.8; under vigorous movement, this yields 98% accuracy for HR deviations ≤10 beats per minute (bpm) and 95% for deviations ≤5 bpm, and the same signal-quality index filter is available to downstream analyses via Huawei Health Kit. Device accuracy had been confirmed against ANSI/AAMI/ISO 81060-2:2018 criteria, while the "Huawei Heart Study" (>3.2 million participants) demonstrated a 92.8% positive predictive value for atrial fibrillation (AF) detection, a result incorporated into the 2020 European Society of Cardiology AF guidelines, collectively supporting the WATCH D's capability to identify and suppress artifacts and to provide clinically reliable physiological data.

Collection of Wearable Data

The data recorded by the wearable device included body temperature, skin temperature, HR, oxygen saturation, sleep quality score, shallow sleep duration, deep sleep duration, dream duration, awake duration, daytime sleep duration, total sleep duration, step volume, distance movement, and calorie consumption. For safety considerations, the blood pressure measurement function on the wearable device was disabled by technicians from the backend.

Each participant was assigned a unique application account to access HUAWEI's developer application programming interface. Measurement data were transmitted to the application via Bluetooth, aggregated through the application programming interface backend, and deidentified before being stored on a secure network drive within the Tianhe-2 supercomputer, supported by the National Supercomputer Center in Guangzhou.

Collection of Electronic Health Record

Patients baseline characteristics, tumor characteristics, surgical information, and perioperative characteristics were included. Baseline characteristics comprised age, sex, BMI, smoking status, and consumption history. Tumor characteristics included pathological type, tumor location, tumor size, and tumor-node-metastasis stage. Surgical information encompassed American Society of Anesthesiologists grade, operation time, surgical procedures, anesthesia method, lymph node dissection, and intraoperative bleeding.

Postoperative characteristics included length of hospital stay, duration of chest tube placement, pleural drainage volume, body temperature, oxygen saturation, HR, amount of postoperative pain medication, frequency of pain medication, administration via pain medication, administration method, and visual analog score for pain and postoperative complications.

Collection of ePROs Data

ePROs included 9 symptoms, which is described in a former study [12], related on a scale from 0 (no symptom) to 10 (the most severe symptom). These symptoms were pain, cough, shortness of breath, restless sleep, fatigue, drowsiness, walking difficulties, and limitation of activities and distressed.

A link to the survey was sent each night at 7 PM during hospitalization. Automatic reminders were sent every 30 minutes thereafter if the questionnaires were not completed. If the survey was not completed by 10 PM, the investigator would remind the patients to complete it. All collected data were aggregated and stored using the same strategy as described earlier.

Time-Series Data Processing

Given the considerable variations in the duration of wearing the wearable device during the perioperative period among different patients, we normalized the time scale. Specifically, we set the start time of surgery as the origin time, with the preoperative period as negative time and the postoperative period as positive time.

A large amount of time-series data could be obtained using wearable devices, including body temperature, skin temperature, HR, and oxygen saturation. Multiple measurements could be recorded within a single minute to enhance the analyzability of the time-series data, and to further reduce the likelihood of potential false positives and false negatives arising from spurious signals, we averaged all measurements obtained within each minute and used these aggregated values in downstream analyses.

Statistical Analysis

Mean and SD or median and range were used to describe continuous data, while categorical data were presented as numbers and percentages. The agreement between the reference method and the wearable device was assessed using the Bland-Altman analysis [13]. To evaluate the feasibility and reliability between the wearable device and the reference method, Bland-Altman analysis was conducted, which is widely regarded as the gold standard for consistency assessment. This method quantifies the systematic bias and limits of agreement between 2 quantitative measurement methods, thereby providing statistical support for evaluating the reliability of the wearable device system. Subgroup analyses evaluating the reliability and feasibility of the digital device were conducted among patients with and without significant postoperative complications. Threshold analysis based on the ePROs score was used to evaluate relationships between different modules data.

Wearable devices can bridge gaps in patient monitoring within the traditional episodic care model. We analyzed the occurrences of outlier detections during periods lacking nursing supervision and visualized these occurrences as line graphs. After aggregating the raw high-frequency measurements into 1-minute averages, we defined outliers on the basis of these minute-averaged values: a minute-averaged body temperature >37 °C or <36 °C; a minute-averaged HR >90 bpm or<60 bpm; and a minute-averaged oxygen saturation <95%. Differences between groups were compared using Student *t* test (2-tailed), chi-squared test, or Fisher exact test.

All data preparation, analyses, and visualization were performed using R (version 4.0.5; The R Core Team, R Foundation for

XSL•FO
RenderX

Statistical Computing). We deemed *P* values <.05 were considered statistically significant.

Results

Participant Demographics and Baseline Data

A schematic overview of the study is presented in Figure 1. As presented in the flowchart diagram in Figure 2, 288 patients consented to participate in the study, including 174 female patients and 114 male patients with a median age of 56 (range 24-82) years. Baseline characteristics of the participants were presented in Table 1. At baseline, most patients underwent segmentectomy (n=203, 70.5%), while 85 (29.5%) underwent lobectomy. In total, 37 (12.8%) patients were former smokers, and 27 (9.4%) patients had a history of alcohol consumption. A total of 83 (28.8%) patients reported preoperative comorbidities, including hypertension, diabetes mellitus, chronic obstructive pulmonary disease, arrhythmia, coronary artery disease, and valvular heart disease. No intraoperative complications or surgical conversions were reported.

Figure 1. Schema of overall study design. A total of 288 patients were prospectively involved in this study. Multimodal data, including electronic health records (EHRs), wearable devices, and electronic patient-reported outcomes (ePROs), were collected, and multimethod statistical analysis was applied. ERAS: enhanced recovery after surgery. See Multimedia Appendix 1 for a high resolution version of this figure.

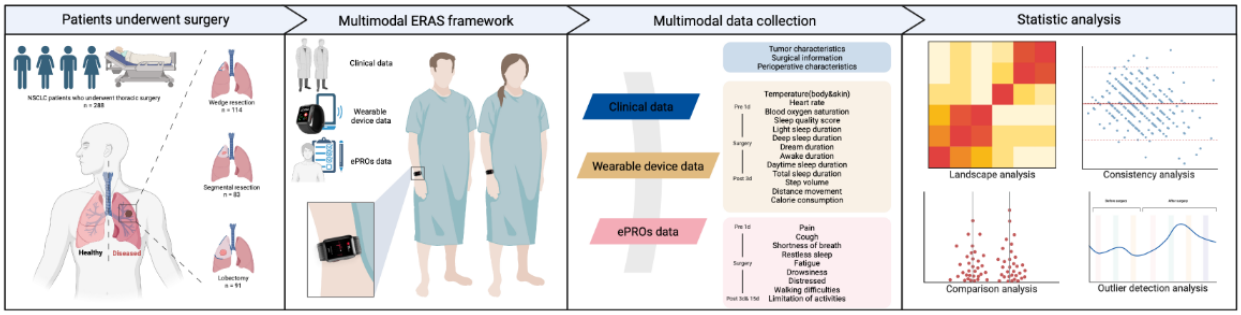
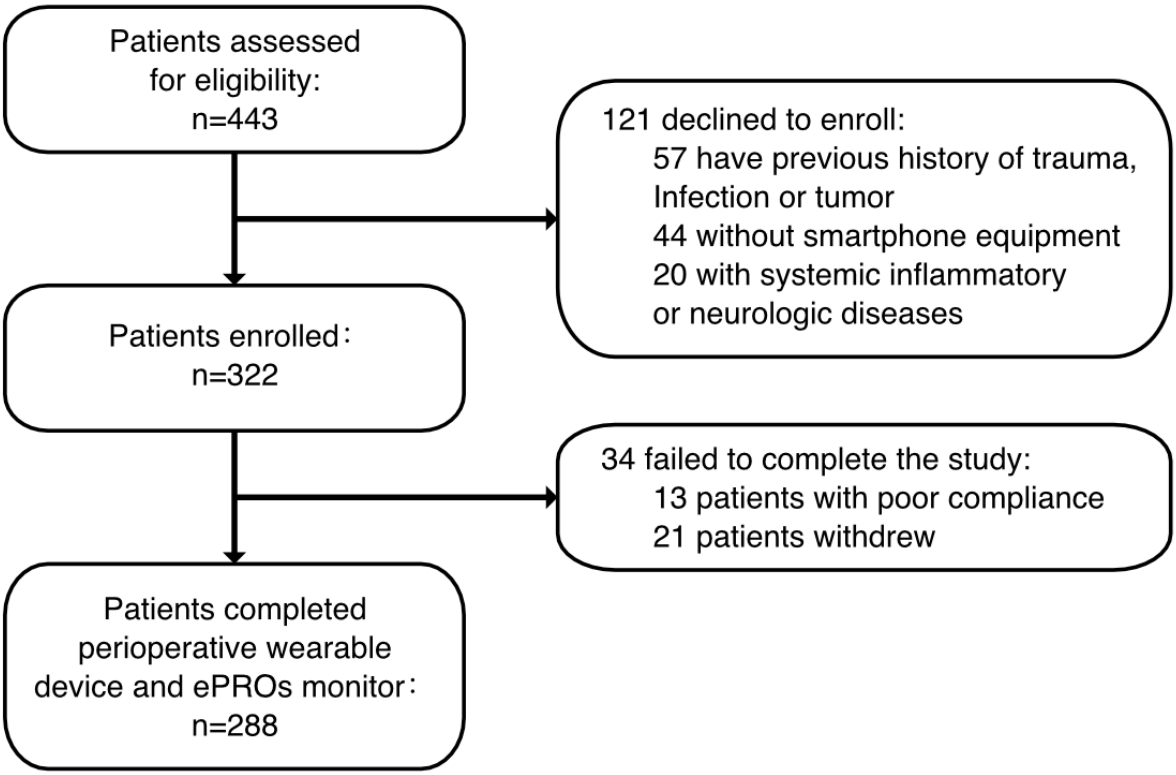


Figure 2. Patient recruitment. The flowchart illustrates the recruitment process and inclusion and exclusion criteria, and included participants underwent multimodal digital device monitoring. ePRO: electronic patient-reported outcome. See Multimedia Appendix 1 for a high resolution version of this figure.



XSL•FO
RenderX

Table 1. Baseline characteristics of enrolled participants (N=288).

Characteristic	Value
Sex, n (%)	
Female	174 (60.4)
Male	114 (39.6)
Age (y), median (range)	56 (24-82)
BMI (kg/m^2), n (%)	
<18.5	20 (6.9)
18.5-24.9	201 (69.8)
≥25	67 (23.3)
Smoking history, n (%)	
Former smoker	37 (12.8)
Never smoker	251 (87.2)
History of alcohol use, n (%)	
Current alcohol use	27 (9.4)
Never drank alcohol	261 (90.6)
Preoperative comorbidities, n (%)	
Hypertension	29 (10.1)
Diabetes mellitus	23 (8)
COPD[a]	17 (5.9)
Arrhythmia	7 (2.4)
CAD[b]	4 (1.4)
Valvular heart disease	3 (1)
Family history of cancer, n (%)	
Yes	46 (16)
No	242 (84)
Surgical position, n (%)	
LUL[c]	70 (24.3)
LLL[d]	33 (11.5)
RUL[e]	73 (25.3)
RML[f]	19 (6.6)
RLL[g]	49 (17)
LUL+LLL	17 (5.9)
RUL+RML	6 (2.1)
RUL+RLL	16 (5.6)
RML+RLL	3 (1)
RUL+RML+RLL	2 (0.7)
Surgical method, n (%)	
Lobectomy	85 (29.5)
Segmentectomy	203 (70.5)
Anesthesia method, n (%)	
Spontaneous ventilation	116 (40.3)

XSL•FO
RenderX

Characteristic	Value
Mechanical ventilation	172 (59.7)
Tumor stage, n (%)	
Tis	50 (17.4)
1	198 (68.8)
2	32 (11.1)
3	7 (2.4)
4	1 (0.3)
Node stage, n (%)	
Nx	14 (4.9)
0	254 (88.2)
1	11 (3.8)
2	8 (2.8)
3	1 (0.3)
Metastasis stage, n (%)	
0	285 (99)
1	3 (1)
ASA[h] grade, n (%)	
I	5 (1.7)
II	241 (83.7)
III	42 (14.6)
Surgical time (h), median (range)	1.50 (0.40-7.05)
Anesthesia time (h), median (range)	3.08 (1.35-8.15)
Intraoperative blood loss (mL), median (range)	10.00 (0-1300.00)

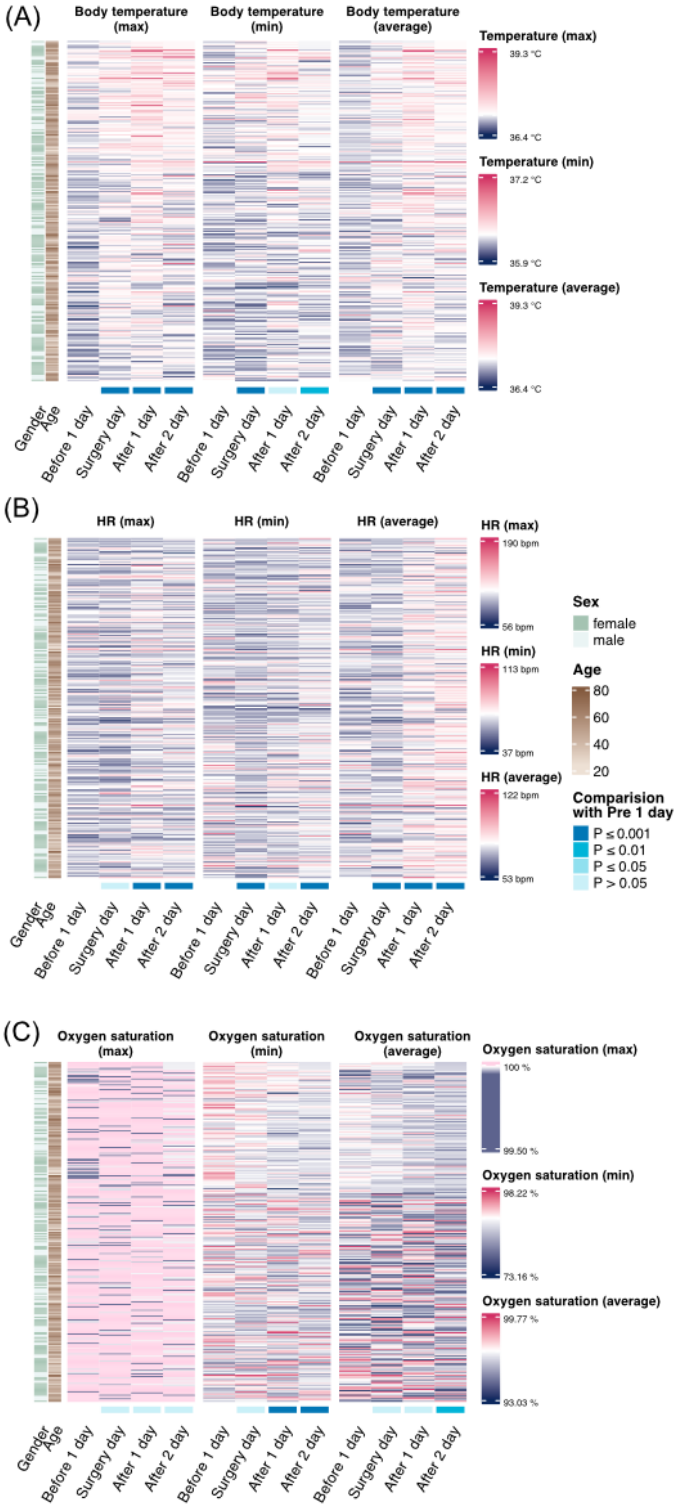
[a]COPD: chronic obstructive pulmonary disease.
[b]CAD: coronary artery disease.
[c]LUL: left upper lobe.
[d]LLL: left lower lobe.
[e]RUL: right upper lobe.
[f]RML: right middle lobe.
[g]RLL: right lower lobe.
[h]ASA: American Society of Anesthesiologists.

Mapping Patient Vital Status and Recovery During Perioperative Vital Signs and Recovery

Figures 3 and 4 demonstrate the utility of integrating wearable device data and ePROs to provide a comprehensive, real-time assessment of perioperative recovery. This analysis highlights the advantages of continuous monitoring in detecting critical physiological changes and tracking patient-reported symptoms during thoracic surgery recovery.

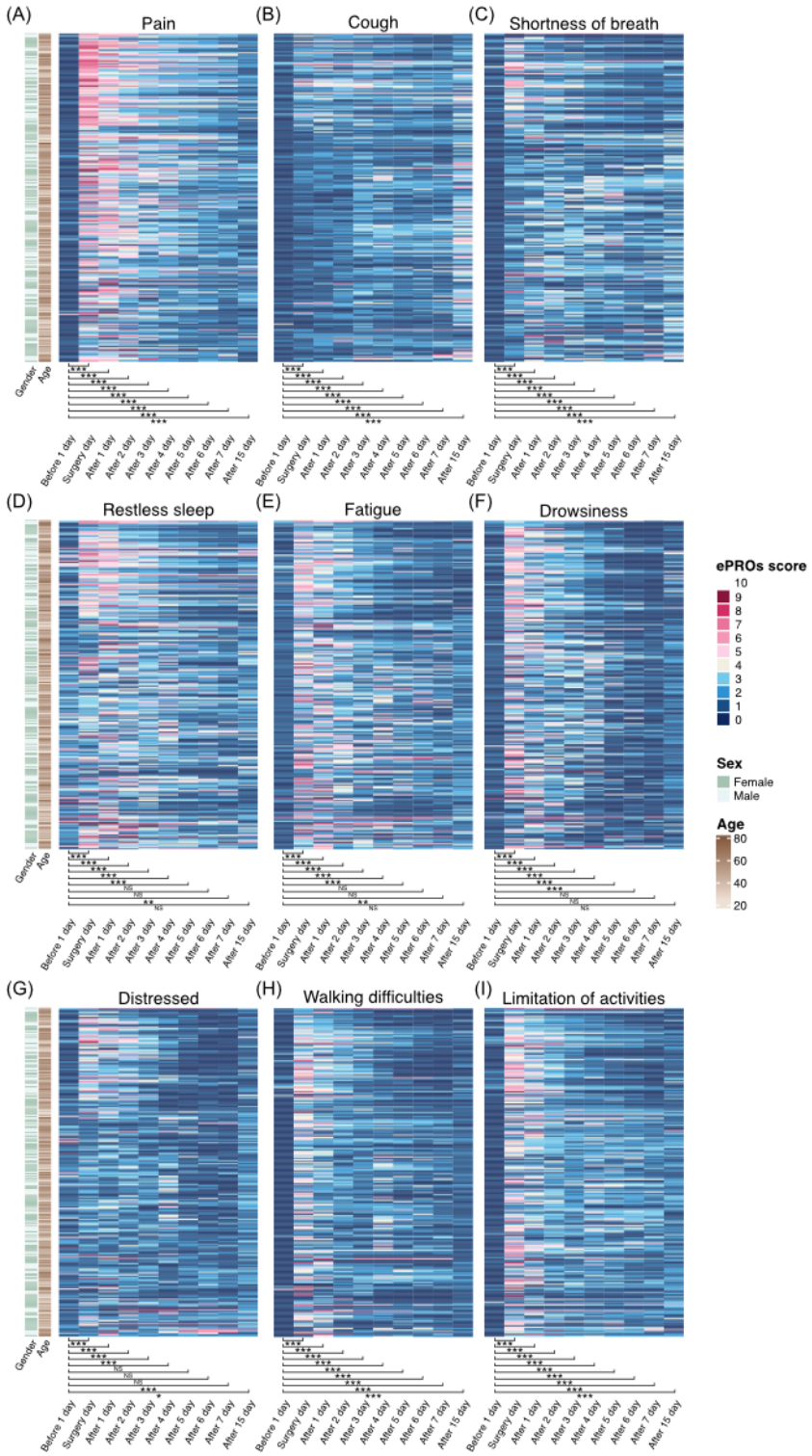
XSL•FO
RenderX

Figure 3. Heatmap illustrates the patients' perioperative body temperature, heart rate (HR), and oxygen saturation using wearable device data. (A) body temperature, (B) HR, and (C) oxygen saturation. Each parameter is displayed with its maximum, minimum, and average values. The sex and age of each patient are annotated and displayed alongside the heatmap. NS: not significant. See Multimedia Appendix 1 for a high resolution version of this figure.



XSL•FO
RenderX

Figure 4. The heat map illustrates the patients' perioperative recovery status using electronic patient-reported outcomes (eRPOs) data. (A) Pain, (B) cough, (C) shortness of breath, (D) restless sleep, (E) fatigue, (F) drowsiness, (G) distressed, (H) walking difficulties, and (I) limitation of activities. NS: not significant. See Multimedia Appendix 1 for a high resolution version of this figure.



In the analysis of wearable device data, both body and skin temperatures showed consistent increases after surgery, indicating inflammatory or healing responses (Figure 3A and Multimedia Appendix 2). HR demonstrated a clear upward trend, with statistically significant increases in maximum and average HR on the first and second postoperative days, highlighting the physical stress of surgery (Figure 3B). Oxygen saturation levels displayed a significant overall decline, especially in minimum and average values during the perioperative period, suggesting potential postoperative respiratory compromise that warrants close monitoring (Figure 3C). Early improvements in physical mobility, as evidenced by significant increases in step count, exercise distance, and calorie expenditure on the first postoperative day (Multimedia Appendix

XSL•FO
RenderX

2), underscore the rapid physical recovery promoted by ERAS protocols. Importantly, sleep quality significantly deteriorated on the day of surgery and the first postoperative day, reflecting the immediate postoperative impact on psychological recovery (Multimedia Appendix 2).

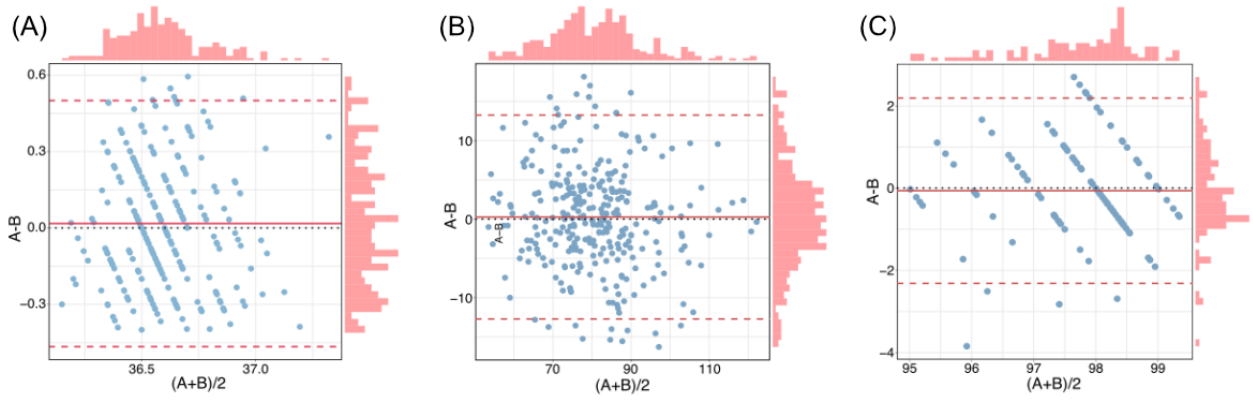
A total of 288 ePROs questionnaires were collected and analyzed. Figures 4A-4I showed that symptoms such as pain, cough, shortness of breath, and activity limitations peaked on the day of surgery and significantly improved within the first week and by day 15. This gradual symptom alleviation aligns with the expected recovery timeline, but the persistent challenges with sleep and fatigue underscore the need for tailored recovery interventions. The combination of wearable devices and ePROs provides a robust framework for perioperative monitoring, offering both objective physiological data and subjective patient feedback.

We also calculated the median and IQRs of ePRO scores during the perioperative period. The results suggested a pattern consistent with the objective trajectory of postoperative recovery. Taking pain as an example, the median pain score and IQR were 0 (IQR 0-0) on the day before surgery. The score peaked on the day of surgery, with a median of 5 (IQR 3-7), and gradually decreased starting from the second postoperative day, remaining at a low level. By the seventh postoperative day, the median pain score was 1 (IQR 1-2), and it further decreased by the 15th postoperative day, with a median of 1 (IQR 0-2; Multimedia Appendix 3).

Evaluation of Agreement Between Wearable Devices and Electronic Health Record

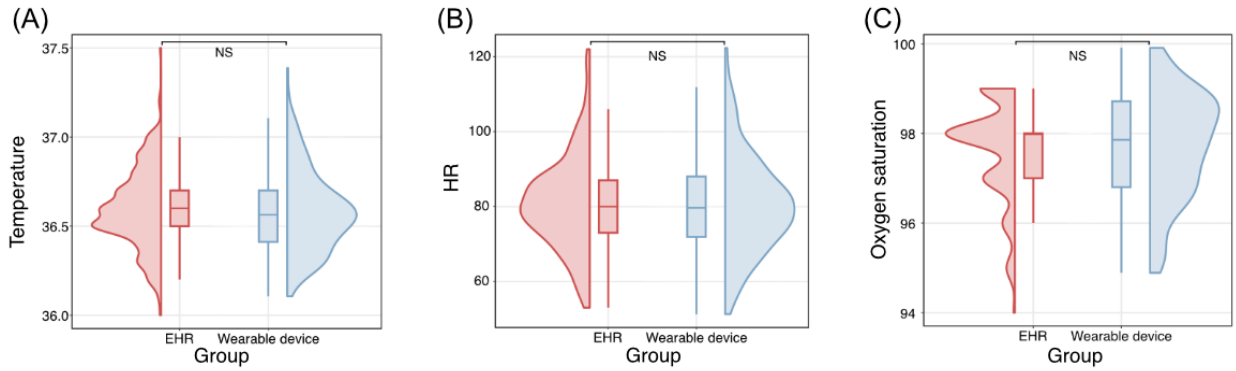
A Bland-Altman analysis was conducted to assess the agreement between wearable device data and electronic health record (EHR; episodic care model) for body temperature, HR, and oxygen saturation. For body temperature, the analysis revealed a minimal bias of 0.02 °C (95% CI –0.01 °C to 0.05 °C), with limits of agreement ranging from –0.47 °C (95% CI –0.52 °C to –0.41 °C) to 0.50 °C (95% CI 0.45 °C to 0.55 °C), indicating strong consistency between the 2 methods (Figure 5A). The agreement for HR showed a bias of 0.26 bpm (95% CI –0.49 bpm to 1.01 bpm), with limits of agreement between –12.71 bpm (95% CI –14.00 bpm to –11.43 bpm) and 13.24 bpm (95% CI 11.95 bpm to 14.53 bpm), reflecting slightly broader variability but overall reasonable alignment between the wearable device and EHR (Figure 5B). For oxygen saturation, the bias was –0.06% (95% CI –0.27% to 0.15%), with limits of agreement from –2.31% (95% CI –2.68% to –1.95%) to 2.19% (95% CI 1.83% to 2.56%), indicating close alignment between the 2 measurement methods (Figure 5C). Paired sample *t* tests confirmed a high level of agreement in the mean values for body temperature (Figure 6A), HR (Figure 6B), and oxygen saturation (Figure 6C), further supporting the reliability of wearable devices in perioperative monitoring. This analysis highlights that wearable devices provide accurate and clinically relevant measurements.

Figure 5. Bland-Altman plot shows the agreement between digital devices and electronic health records (EHRs) in all participants. (A) Body temperature, (B) heart rate (HR), and (C) oxygen saturation. See Multimedia Appendix 1 for a high resolution version of this figure.



XSL•FO
RenderX

Figure 6. Raincloud plot shows the comparison between digital devices and electronic health records (EHRs) in all participants. (A) Body temperature, (B) heart rate (HR), and (C) oxygen saturation. See Multimedia Appendix 1 for a high resolution version of this figure.



Continuous Vital Signs Monitoring: Bridging Gaps in Perioperative EHR Systems

Perioperative records in EHR systems often suffer from discontinuities, with significant temporal gaps in vital sign data. To address this issue, we conducted a systematic review to evaluate the benefits of wearable devices for continuous, 24-hour monitoring of vital signs. Table 2 summarizes the comparative performance of wearable devices in this context.

Table 2. Comparative performance and outliers between electronic health records (EHRs) and wearable devices.

	EHR	Wearable device	*P* value
Body temperature (times)			
Total number per patient	10.00	14,219.04	<.001
Total number per patient per day	3.08	4440.69	<.001
Outliers number per patient	0.23	2707.47	<.001
Outliers number per patient per day	0.07	845.56	<.001
He art rate (times)			
Total number per patient	10.19	4293.75	<.001
Total number per patient per day	2.65	1339.16	<.001
Outliers number per patient	2.08	964.96	<.001
Outliers number per patient per day	0.54	300.96	<.001
Oxygen saturation (times)			
Total number per patient	6.83	6924.67	<.001
Total number per patient per day	1.77	2057.24	<.001
Outliers number per patient	0.17	2644.67	<.001
Outliers number per patient per day	0.04	785.70	<.001

During hospitalization, body temperature, HR, and oxygen levels were measured an average of 3.08, 2.65, and 1.77 times per patient per day, respectively, during ward rounds, leading to 10.00, 10.19, and 6.83 measurements per patient over the course of their stay. In contrast, wearable devices recorded these parameters continuously, providing an average of 4440.69, 1339.16, and 2057.24 measurements per patient per day, with total readings reaching 14,219.04, 4293.75, and 6924.67 times.

The difference in detection sensitivity was particularly notable when comparing outlier detection. For body temperature, the ward rounds identified 0.07 times outliers per patient, per day, whereas the wearable devices detected significantly more outliers, averaging 845.56 times per patient per day. Similarly, HR outliers were recorded at a rate of 0.54 times per patient per day through ward rounds, compared to 300.96 times detected by the wearable devices. For oxygen saturation levels, ward rounds captured 0.04 times outliers per patient per day, while wearable devices identified 785.70 times outliers.

This comparison underscores the advantage of wearable devices in providing continuous, real-time vital sign monitoring, filling the gaps left by intermittent EHR-based records, and enabling earlier detection of critical physiological changes.

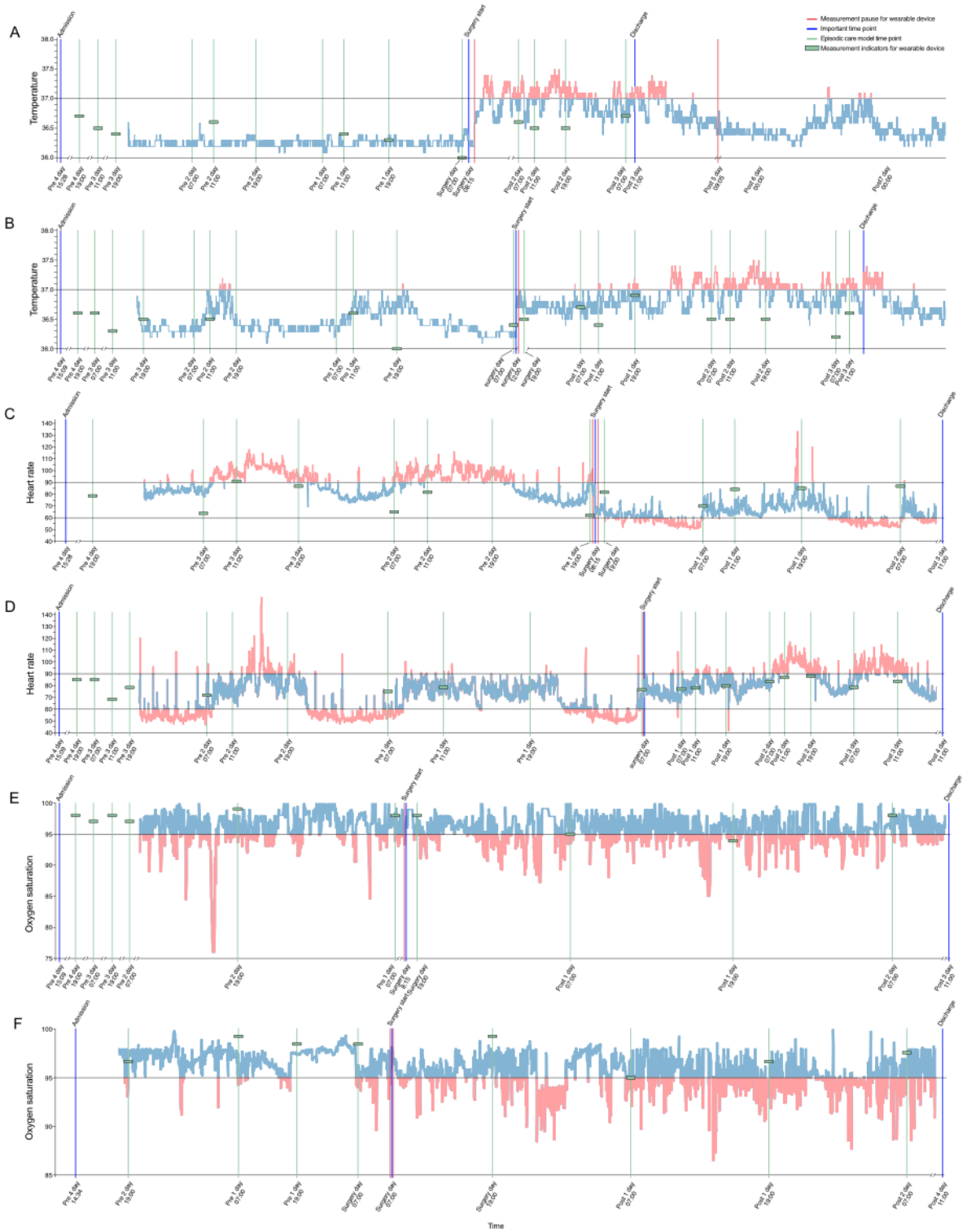
Typical Case Presentation and Comparison

To provide a clearer and more intuitive comparison between episodic care methods and digital devices, we sampled 2 typical cases and visualized their perioperative body temperature, HR, and oxygen saturation levels.

XSL•FO
RenderX

Figures 7A and 7B illustrated the perioperative body temperature trends for both patients. Preoperatively, the medical team assessed body temperature 11 and 17 times, respectively, for each patient, with all readings within normal ranges. Wearable device data closely mirrored these values, with most preoperative measurements also falling within normal limits. Notably, patient #1 maintained a more stable body temperature compared to patient #2, who experienced 2 transient spikes. However, the wearable device detected additional outliers that were missed by conventional monitoring, particularly in the intervals between routine checks.

Figure 7. Typical case presentation of perioperative continuous monitoring and comparison between episodic care method and digital device. (A) Perioperative body temperature trends of case 1, (B) perioperative body temperature trends of case 2, (C) perioperative heart rate (HR) trend of case 1, (D) perioperative HR trend of case 2, (E) perioperative oxygen saturation trend of case 1, and (F) perioperative oxygen saturation trend of case 2. See Multimedia Appendix 1 for a high resolution version of this figure.



XSL•FO
RenderX

Figures 7C and 7D depict the perioperative HR monitoring. The medical team recorded HR 12 and 16 times for each patient, respectively, and noted one outlier in the male patient. In contrast, the wearable device revealed time-specific fluctuations in HR that were not captured by conventional monitoring. Activities such as departmental visits occasionally led to missed measurements, an issue addressed by the continuous monitoring of the wearable device. For instance, in the selected female patient, measurements were missed at 11 AM and 7 PM on April 12, 2023, during which significant HR fluctuations occurred, as revealed by the wearable device.

Figures 7E and 7F show paired perioperative oxygen saturation monitoring. The medical team measured oxygen saturation levels 10 and 8 times for each patient, again noting an outlier in the male patient. The wearable device data showed high consistency with these traditional measurements while also providing continuous, uninterrupted data.

Assessing the Correlation Between Wearable Metrics and ePROs Scores

To explore the potential association between data from wearable devices and ePROs, we conducted a threshold analysis comparing wearable parameters between 2 subgroups, categorized based on the overall ePROs score and 9 subindexes.

Multimedia Appendix 4 summarizes the results of the threshold analysis based on the overall ePROs score. Notably, patients in the high-score group exhibited significantly higher extreme and mean values of HR, as well as lower minimum and mean oxygen saturation levels, compared to those in the low-score group. In addition, the high-score group showed elevated maximum body temperature and poorer sleep quality during the perioperative period. These findings suggest a strong association between elevated physiological stress, as detected by the wearable device, and higher self-reported symptom burden from ePROs.

Further analysis based on the 9 subindexes, detailed in Multimedia Appendix 5, revealed significant differences in extreme and mean values of body temperature, HR, and oxygen saturation levels between subgroups. These results underscore the association of wearable metrics in reflecting the severity of ePROs, providing valuable insights for more personalized perioperative management.

Reported Feasibility and Acceptance of Digital Device

Upon completing the study, all patients were interviewed to assess their experience with the ePROs and wearable devices. Table 3 presents the summarizations of their feedback. Overall, patients reported a positive experience, stating that both the wearable device and the ePRO questionnaires were easy to use, and the device was comfortable to wear and simple to maintain. The primary feedback centered on enhancing the functionality of the wearable device. Patients suggested improvements to the software interface for better usability and recommended adding a blood pressure monitoring feature to further expand the device's capabilities.

Table 3. Feedback regarding the digital device from patient and medical team interviews.

Themes	Interview response
Overall experience	• "It has a positive effect on my preoperative status assessment and postoperative recovery." • "It is non-invasive, I just need to wear the watch as usual and complete the electronic questionnaire." • "It is waterproof, I can even keep wearing it while taking a shower." • "I don't need to charge it frequently, once or twice a week is sufficient, which is very convenient." • "I'm satisfied, medical team can better understand my condition through the wearable device and ePROs, which makes me feel valued."
Device comfort	• "It looks like a watch for daily wear." • "There are different types of watchbands to choose, offering a comfortable, barely noticeable wear." • "It is very lightweight, and I wear it while sleeping without any impacts." • "I can easily fill the questionnaires on my phone, simply selecting options based on my condition. Only takes a few minutes to complete." • "The questionnaire comes with reminder notifications on my phone, ensuring that I don't forget to complete it."
Problems and improvements to the device	• "Plastic watchbands tend to cause swearing when worn outdoors in the summer." • "Is it possible to add a blood pressure monitoring feature?" • "Is it possible to use an infinitely adjustable watchband?" • "Is it possible to combine the wearable device's software with the questionnaire app into one? Download multiple apps can be quite frustrating."
Useful for the medical team?	• "Yes, it would greatly assist the medical team in supplementing critical data amidst their busy work schedules." • "Yes, it can be more objectively reflect the patient's perioperative vital status." • "Yes, it also serves as a motivation for patients. When patients wear the device and complete the ePROs questionnaires, it encourages them to be more proactive in order to achieve feedback and outcomes."

XSL•FO
RenderX

Evaluation of the Digital Device in Patients With and Without Postoperative Complications

Postoperative complications were recorded in 25 (8.7%) out of 288 patients. Of these 25 patients, 13 (52%) experienced air leaks, 6 (24%) developed sustained high fever, 5 (20%) had persistent pleural drainage, and 1 (4%) exhibited poor lung re-expansion. The remaining 263 (31.3%) patients did not report any recordable postoperative complications. The baseline characteristics of patients with postoperative complications were summarized in Multimedia Appendix 6.

Therefore, we performed the subgroup analyses based on patients with postoperative complications. The results of the Bland-Altman analysis demonstrated consistency with the overall findings. The body temperature revealed a minimal bias of –0.02 °C (95% CI –0.33 °C to 0.29 °C), with limits of agreement ranging from –0.33 °C (95% CI –0.44 °C to –0.22 °C) to 0.30 °C (95% CI 0.18 °C to 0.40 °C; Multimedia Appendix 7). The agreements for HR showed a bias of 1.65 bpm (95% CI –1.21 bpm to 4.51 bpm), with limits of agreement between –11.95 bpm (95% CI –16.66 bpm to –7.24 bpm) and 15.24 bpm (95% CI 10.53 bpm to 19.95 bpm; Multimedia Appendix 6). For oxygen saturation, the bias was 0.27% (–0.24% to 0.78%), with limits of agreement from –2.14% (95% CI –2.98% to –1.31%) to 2.68% (95% CI 1.85% to 3.52%; Multimedia Appendix 8). The results of the paired sample *t* test demonstrated a high level of consistency between digital device and EHR measurements in patients with postoperative complications (Multimedia Appendix 7).

We also conducted a subgroup analysis of patients without postoperative complications; their baseline characteristics were summarized in Multimedia Appendix 9. The results demonstrated a high degree of agreement between digital devices and EHR measurements (Multimedia Appendices 10 and 11).

In summary, the results of the subgroup analysis demonstrated a high level of consistency with the overall findings, thereby confirming the robustness of the study's main results. Furthermore, these findings reinforced that the digital device was capable of effectively capturing clinically relevant vital information even in complex clinical settings, supporting its broad feasibility and reliability for clinical application.

Discussion

Principal Findings

This is the first study to investigate the feasibility and reliability of using digital devices, including wearable devices and ePRO questionnaire, in the perioperative management of patients with lung cancer who underwent thoracic surgery. In this study, perioperative recovery was monitored through digital devices. Our results indicated that multimodal data could reflect patients' perioperative status objectively and provide rapid, sensitive outlier detection. This innovation can assist the medical team in optimizing recovery trajectories and achieving the goals of ERAS in the context of the MITS era.

Using prospectively collected data, we systematically mapped the perioperative recovery trajectory of patients with lung cancer who underwent thoracic surgery. Comparison across days revealed significant improvements in the vast majority of parameters, underscoring their critical role in recovery. Notably, the duration of shallow sleep, deep sleep, dreaming, total sleep, and day sleep, as recorded by the wearable device, showed no significant differences across days. However, the sleep quality score significantly improved during hospitalization. This discrepancy aligns with the restless sleep and drowsiness reported via ePROs, which reflect patients' subjective experiences, suggesting a higher level of accuracy [14,15]. Possible explanations include improved sleep structure, fewer sleep-cycle disruptions, enhanced subjective wakefulness and sleep experience, and gradual environmental adaptation. In addition, psychological changes after surgery, such as anxiety and distress, also play a role, as reflected in other ePROs-derived measures.

Although this is a noninterventional study, its potential impact on nursing care and therapeutic procedures holds significant clinical importance. Through the recovery trajectories monitored by wearable devices and ePROs, we were able to capture real-time vital signs and subjective experiences of patients, thereby providing more comprehensive information for clinical decision-making. As a self-reported tool, ePROs effectively reflect patients' subjective experiences, such as anxiety, pain, and activity limitations, and changes in their scores can serve as triggers for interventions. Specifically, when a score in a particular ePROs measure exceeds a set threshold, the nursing team can immediately implement corresponding interventions based on the patient's reported concerns. For example, if a patient reports high pain levels, the nursing team can provide a pain management plan according to the assessment results; if a patient's anxiety score exceeds the threshold, a psychological evaluation and consultation can be arranged. This precision-based intervention model using ePROs enables more individualized care, preventing both overintervention and undercare, thereby more effectively supporting the patient's recovery process. In addition, digital devices can promptly alert nursing staff when vital signs deviate from normal ranges, enabling timely interventions. The integration of ePROs and digital device monitoring data significantly enhances nursing efficiency, optimizes resource allocation, and reduces unnecessary interventions. Future interventional studies will further explore how these monitoring data can be integrated into routine clinical workflows to improve perioperative recovery outcomes and enhance nursing care quality.

Wearable devices are gaining attention in monitoring patients, and the health population has garnered significant attention. These devices are equipped with various sensors and chips designed to track a wide range of data, including HR sensors, accelerometers, gyroscopes, temperature sensors, oxygen saturation sensors, electrode sensors, and ambient light sensors. While their convenience is widely recognized, concerns regarding reliability and data security remain [16,17]. In this study, we conducted pairwise comparisons of body temperature, HR, and oxygen saturation obtained from EHR and wearable devices, respectively. Bland-Altman agreement analysis and

XSL•FO
RenderX

paired sample *t* test analysis demonstrated a high degree of consistency between the 2 modules' data, supporting the reliability, validity, and clinical utility of wearable devices. Subgroup analysis in patients with and without postoperative complications also showed consistency between overall results, further confirming the feasibility and reliability of the proposed digital device approach in this study. This is consistent with previous studies [18-21], including a systematic review by Germini et al [22], which confirmed the accuracy and acceptability of wrist-wearable activity trackers. This review included 65 studies, assessing 14 accuracy-related outcomes. Step count, HR, and energy expenditure were the most frequently evaluated. Fitbit Charge and Fitbit Charge HR (Fitbit Inc) showed the highest accuracy for step counts, while the Apple Watch (Apple Inc) performed better for HR. Acceptability, measured by data availability and wearing time, revealed that Fitbit Charge HR, Fitbit Flex 2, and Garmin Vivofit (Garmin Ltd) had the highest data availability, whereas GENEActiv (Activinsights Ltd) and FuelBank (Nike Inc) had the longest wearing time. Although previous studies, as mentioned earlier, have extensively demonstrated the accuracy and reliability of wearable devices, we acknowledge that in this study, we accepted the credibility of data collection provided by Huawei Technologies and did not conduct further independent validation of these data. Nevertheless, it is important to emphasize that quality assurances provided by manufacturers are not fully equivalent to rigorous independent data validation conducted in academic research. Therefore, the interpretation of our results should consider this specific limitation.

In this study, the wearable device estimated core body temperature by measuring peripheral skin temperature at the wrist. During the development phase, the research team established a mapping model by simultaneously collecting data on both core and skin temperatures. This model was further optimized using physiological and environmental parameters such as HR and ambient temperature. After the stability of the model was validated, it was integrated into the device's algorithm, enabling accurate real-time estimation of core body temperature based on skin temperature readings. In addition, the temperature sensor underwent calibration and performance testing under various environmental conditions to ensure measurement accuracy.

Continuous monitoring offers distinct advantages in perioperative management. In contrast to traditional EHRs, which capture vital signs at discrete time points, wearable devices provide real-time, continuous monitoring and reduce the likelihood of missing critical clinical changes. A key advancement in this area is the development of PPG technology, which has made HR monitoring a reliable and established feature of wearable devices [23,24]. A recent review in the *New England Journal of Medicine* summarized and provided an outlook on wearable digital health technologies for monitoring cardiovascular conditions. The gradual integration of wearable devices into cardiovascular care can support continuous monitoring of patients' vital signs and other physiological parameters, enabling rapid remote monitoring and management. This approach aims to facilitate early detection of disease deterioration, improve perioperative management, and reduce readmission rates [25]. Guo et al [26] evaluate the effectiveness of PPG technology integrated into wearable devices for detecting AF in a large population-based cohort across China. This study included 187,912 participants who monitored their pulse rhythm using a HUAWEI smartwatch for at least 14 days. Results showed that 0.23% (424/187,912) of the participants received a "suspected AF" notification, with 87% of cases confirmed as AF through clinical evaluation, yielding a positive predictive value of 91.6%. The study concluded that continuous home monitoring with PPG-based wearable devices is a feasible and effective method for early AF detection, enabling timely interventions to reduce AF-related complications. On the basis of these findings, wearable digital health technologies have been recommended for cardiovascular monitoring in the 2020 European Society of Cardiology guidelines [27].

In the perioperative management within the surgical community, wearable devices have shown promise across various surgical disciplines. Beqari et al [28] conducted a pilot study with 56 patients who had undergone cardiothoracic surgery to evaluate the effectiveness of the "NightSigal" algorithm. Patients were asked to wear a Fitbit watch for at least 1 week preoperatively and up to 90 days postoperatively to collect biometric data. The algorithm detected 17 out of 21 postoperative events, with a median detection time of 2 days before symptom onset. Sensitivity and specificity were 81% and 75%, respectively. Ghomrawi et al [29] assessed the use of Fitbit to monitor the postoperative complications in patients who have undergone pediatric appendectomy. The combined predictive ability for complex appendicitis was an area under the curve of 0.80, compared to 0.70 for simple appendicitis. Kane et al [30] examined activity monitors for predicting 30-day readmission by tracking activity levels before and after surgery. Readmitted patients showed lower return to baseline activity compared to nonreadmitted patients. After adjusting for cofounders, multivariable analysis demonstrated that returning to baseline activity was associated with a reduced risk of 30-day readmission. The model had a sensitivity of 75% and a specificity of 74%.

Despite the advances seen in other surgical fields, the utility of wearable devices in patients with lung cancer undergoing thoracic surgery remains underexplored. Our findings revealed that outliers in vital signs, such as fluctuations in body temperature, HR, oxygen saturation levels, and other key metrics, occurred between the time points of EHR monitoring. Implementing robust continuous monitoring and alarm mechanism could enable the medical team to better track patients' baseline vital signs, thereby facilitating ERAS protocols.

Effective outlier detection in this study relied on a 2-tier strategy that integrated device-embedded safeguards with protocol-level quality control. First, at the hardware–algorithm tier, the wrist-worn HUAWEI WATCH D combined (1) an upgraded TruSeen 5.0+ optical module—8 circumferential photodiodes, dual LEDs, a Fresnel film, and a curved 2.5-D sapphire or physical vapor deposition back—plus the TruBP pressure system [31], and (2) a HiFi-Encoder deep learning pipeline that discards frames whose signal-quality index is <0.8; together these

XSL•FO
RenderX

measures cut motion-related signal loss to one-quarter of that in the previous generation and yielded HR accuracies of 98% (≤10 bpm) and 95% (≤5 bpm) during vigorous movement. A dual-mode PPG-electrocardiography architecture driven by a high-performance chip was trained on large baseline datasets collected in rest, exercise, commuting, sleep, heat, and underwater conditions; it has already demonstrated a 92.8% positive predictive value for AF detection in >3.2 million participants and is cited in the 2020 European Society of Cardiology AF guidelines [27], while independent testing confirmed full compliance with ANSI/AAMI/ISO 81060-2:2018 [32]. The device is further supported by class II approval from China's National Medical Products Administration and European Union Medical Device Regulation certification. Second, at the study-design tier, intensive patient training and ward-level adherence checks led to exclusion of 13 poorly compliant participants; temperature, HR, and oxygen-saturation readings were cross-validated against EHR values to confirm accuracy; and continuous data were aggregated into 1-minute maxima, minima, and means, with outliers defined as minute-averaged temperature >37°C or <36°C, HR >90 or <60 bpm, or oxygen saturation <95%. This minute-level preprocessing attenuated sporadic spikes and focused analyses on sustained, clinically meaningful deviations. Collectively, the 6 elements—noise-reduction hardware, real-time AI filtering, extensive algorithm training, external validation and regulation, adherence management, and time-series preprocessing—provided a robust framework for distinguishing true physiological outliers from artifacts, thereby reinforcing the reliability and clinical applicability of our perioperative monitoring results.

The benefits of integrating EHRs with wearable devices are gradually becoming more apparent. Building on this foundation, we further incorporated ePROs to develop a multimodule perioperative monitoring network. Previous studies have demonstrated that patient reported outcomes (PROs) can effectively capture various aspects of a patient's perioperative status, such as sleep, fatigue, and drowsiness [33]. Dai et al [8] conducted a multicenter, randomized controlled trial demonstrating the feasibility and effectiveness of PROs-based symptom management in patients who underwent lung cancer surgery. In the early recovery phase, PROs-based groups experienced significantly fewer symptom threshold events at discharge within 4 weeks after discharge, with a lower incidence of complications, highlighting their potential to reduce symptom burden and complication rates. Earlier this year, Dai et al [9] updated the long-term follow-up results, showing that ePROs groups had a lower symptom burden and higher physical and emotional function scores over a follow-up period of 1 to 12 months compared to usual care.

In this study, we introduced a threshold analysis to stratify patients based on the ePROs overall score and subindexes. Most vital signs and physiological parameters measured by the wearable device showed statistically significant differences between the high and low ePROs score groups, suggesting a potential relevance between various digital device modules. This highlights the advantages of using multimodal data to comprehensively assess the perioperative recovery status of patients who underwent lung cancer surgery.

Our study also highlighted the potential economic benefits of the digital device–based continuous monitoring system. By enabling early detection of complications and timely interventions, the system may shorten hospital stays and reduce associated costs. Its integration into perioperative workflows can also alleviate nursing workload and improve resource efficiency. Furthermore, the use of the HUAWEI WATCH D, a certified medical-grade wearable device covered by the National Medical Products Administration, enhances the feasibility of large-scale clinical implementation.

The acceptability, usability, and comfort of digital devices are crucial determinants of their successful clinical implementation [25]. In this study, feedback from both patients and medical staff was largely positive, with participants acknowledging the ability of digital devices to gather detailed perioperative data without disrupting daily activities or routine care. Concerns were raised regarding the stability of the devices and the potential influence of external factors, such as physical movement, on data accuracy. Addressing these challenges through future technological advancements will be essential to ensuring the reliability of wearable devices in clinical practice.

To address these limitations and further enhance the robustness of digital perioperative monitoring, our study integrated ePROs as a complementary data modality alongside wearable devices. While wearable devices continuously captured objective physiological parameters such as HR and oxygen saturation, ePROs provided direct insights into patients' subjective symptoms and functional status, including pain, fatigue, and sleep quality. The design of combining sensor-based and self-reported data improved the overall reliability and interpretability of perioperative monitoring. Moreover, this integration supported seamless, minimally burdensome data acquisition and enabled more individualized recovery profiling. In future applications, this multimodal framework may also contribute to the development of more accurate and generalizable predictive models for surgical recovery.

There are several limitations in our study that warrant acknowledgment. First, this was a single-center, prospective, self-controlled study conducted among patients undergoing thoracic surgery, which limited the generalizability of the results to other populations. However, our findings provide a foundation for future multicenter clinical trials. Second, all patients enrolled were required to own and operate a smartphone and to have a basic understanding of the functions provided by the wearable device; while the research team made every effort to provide technical support and detailed assistance to all potentially eligible patients, we acknowledge that individuals with limited technological literacy, including some older adults, may have been excluded, which highlights the importance of future research into the usability and compliance of wearable technologies across diverse patient demographics. Third, although the 2-tier strategy of device-embedded safeguards coupled with protocol-level quality control used in this study could identify and manage most potential artifacts in the wearable-device data, we acknowledge that residual false

XSL•FO
RenderX

positive and false negative readings may still have influenced the study's results and conclusions. Fourth, although the wearable device used in this study was the first wrist-worn wearable device to receive class II medical device certification, the blood pressure measurement function required the patient to perform the measurement autonomously. This could potentially lead to frequent measurements that might affect blood circulation or result in missing or reduced data quality if measurements were not taken according to the study protocol. After careful consideration, to avoid interfering with the normal treatment process of the patients, we ultimately decided not to include blood pressure as one of the planned measurement outcomes in this study. Previous studies have demonstrated the safety and feasibility of using the HUAWEI Watch D for blood pressure measurement, which could compensate for the lack of blood pressure monitoring data in this study [21]. In the future, we will explore the clinical application of dynamic blood pressure monitoring for patients who underwent thoracic surgery using this device within a safer clinical framework. Fifth, individualized perioperative nursing interventions may have introduced outliers in the digital device–recorded data. For example, postoperative pain and the use of analgesics may have affected HR readings. Although this study used a multimodal monitoring approach to objectively reflect patient recovery from multiple perspectives and conducted a subgroup analysis to further verify the robustness of the findings, achieving real-time consistency and alignment between individualized perioperative clinical management and the parameters monitored by digital devices remains one of the key challenges for future research.

Conclusions

Within the perioperative management of patients with lung cancer under the ERAS framework, the digital device composed of wearable devices and ePROs demonstrated feasibility and reliability in the continuous perioperative monitoring of patients who underwent lung cancer surgery. Our study offered a novel approach for multimodal, continuous perioperative monitoring, thereby laying the foundation for future research.

Acknowledgments

The authors express their appreciation to all participants in this study; they also sincerely appreciate the technical staff from HUAWEI for generously answering their questions regarding the technical details of the devices. This study was supported by the R&D Program of Guangzhou National Laboratory (grant SRPG22-017). The writing and visualization in this study were entirely completed by the research team without the use of generative artificial intelligence for text or figure generation. ChatGPT was consulted for suggestions regarding code modification and error handling, and the final implementations were independently conducted based on the advice provided.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to institutional regulations concerning patient data but are available from the corresponding author upon reasonable request.

Authors' Contributions

JH is the co–corresponding author and can be reached at hejx@vip.163.com. RW, WG, and HL were responsible for the conceptualization. JZ, HH, WG, Y Li, ML, and QZ were responsible for data curation. RW, WG, Y Lin, and YF were responsible for the formal analysis. JH was responsible for funding acquisition. RW, HH, YC, and HL were responsible for the investigation. RW and YJ were responsible for the methodology. YH, JH, and HL were responsible for the project administration. KC was responsible for software. YC, JH, and HL were responsible for resources. WG, JH, and HL were responsible for the supervision. JZ, HH, and YF were responsible for validation. RW, LH, and ZY were responsible for visualization. RW and QW were responsible for writing—original draft. HD, YH, and HL were responsible for writing—reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
High resolution version of all figures.
[ZIP File (Zip Archive), 3258 KB - mhealth_v13i1e69512_app1.zip]

Multimedia Appendix 2
Heatmap illustrated the patients' perioperative skin temperature, activity, and sleep status using wearable device data. (A) skin temperature;(B) sleep assessment;(C) activity assessment. Each parameter was displayed with its maximum, minimum, and average values. The gender and age of each patient were annotated and displayed alongside the heatmap.
[PNG File , 3109 KB - mhealth_v13i1e69512_app2.png]

Multimedia Appendix 3

XSL•FO
RenderX

Median and Interquartile Range Distribution of electronic patients-report outcome.
[DOCX File , 29 KB - mhealth_v13i1e69512_app3.docx]

Multimedia Appendix 4
Results of threshold analysis based on the ePROs overall score.
[DOCX File , 39 KB - mhealth_v13i1e69512_app4.docx]

Multimedia Appendix 5
Results of threshold analysis based on the nine symptoms.
[XLSX File (Microsoft Excel File), 108 KB - mhealth_v13i1e69512_app5.xlsx]

Multimedia Appendix 6
Baseline characteristic of patients with postoperative complications.
[DOCX File , 25 KB - mhealth_v13i1e69512_app6.docx]

Multimedia Appendix 7
Bland-Altman plot showed the agreement between digital device and EHR in participants with postoperative complications. (A) body temperature; (B) HR; (C) oxygen saturation.
[PNG File , 1064 KB - mhealth_v13i1e69512_app7.png]

Multimedia Appendix 8
Raincloud plot showed the comparison between digital device and EHR in participants with postoperative complications. (A) body temperature; (B) HR; (C) oxygen saturation.
[PNG File , 1188 KB - mhealth_v13i1e69512_app8.png]

Multimedia Appendix 9
Baseline characteristic of patients without postoperative complications.
[DOCX File , 25 KB - mhealth_v13i1e69512_app9.docx]

Multimedia Appendix 10
Bland-Altman plot showed the agreement between digital device and EHR in participants without postoperative complications. (A) body temperature; (B) HR; (C) oxygen saturation.
[PNG File , 1553 KB - mhealth_v13i1e69512_app10.png]

Multimedia Appendix 11
Raincloud plot showed the comparison between digital device and EHR in participants without postoperative complications. (A) body temperature; (B) HR; (C) oxygen saturation.
[PNG File , 1221 KB - mhealth_v13i1e69512_app11.png]

References

1. Bendixen M, Jørgensen OD, Kronborg C, Andersen C, Licht PB. Postoperative pain and quality of life after lobectomy via video-assisted thoracoscopic surgery or anterolateral thoracotomy for early stage lung cancer: a randomised controlled trial. Lancet Oncol 2016 Jun;17(6):836-844. [doi: 10.1016/S1470-2045(16)00173-X] [Medline: 27160473]
2. Fan CJ, Chien HL, Weiss MJ, He J, Wolfgang CL, Cameron JL, et al. Minimally invasive versus open surgery in the Medicare population: a comparison of post-operative and economic outcomes. Surg Endosc 2018 Sep;32(9):3874-3880. [doi: 10.1007/s00464-018-6126-z] [Medline: 29484556]
3. Biere SS, Maas KW, Bonavina L, Garcia JR, van Berge Henegouwen MI, Rosman C, et al. Traditional invasive vs. minimally invasive esophagectomy: a multi-center, randomized trial (TIME-trial). BMC Surg 2011 Jan 12;11:2 [FREE Full text] [doi: 10.1186/1471-2482-11-2] [Medline: 21226918]
4. Nelson DB, Mehran RJ, Mena GE, Hofstetter WL, Vaporciyan AA, Antonoff MB, et al. Enhanced recovery after surgery improves postdischarge recovery after pulmonary lobectomy. J Thorac Cardiovasc Surg 2023 May;165(5):1731-40.e5 [FREE Full text] [doi: 10.1016/j.jtcvs.2022.09.064] [Medline: 36404141]
5. Wang C, Goel R, Noun M, Ghanta RK, Najafi B. Wearable sensor-based digital biomarker to estimate chest expansion during sit-to-stand transitions-a practical tool to improve sternal precautions in patients undergoing median sternotomy. IEEE Trans Neural Syst Rehabil Eng 2020 Jan;28(1):165-173 [FREE Full text] [doi: 10.1109/TNSRE.2019.2952076] [Medline: 31714229]

XSL•FO
RenderX

6. Michard F. Smartphones and e-tablets in perioperative medicine. Korean J Anesthesiol 2017 Oct;70(5):493-499 [FREE Full text] [doi: 10.4097/kjae.2017.70.5.493] [Medline: 29046768]
7. Boffa DJ, Kosinski AS, Furnary AP, Kim S, Onaitis MW, Tong BC, et al. Minimally invasive lung cancer surgery performed by thoracic surgeons as effective as thoracotomy. J Clin Oncol 2018 Aug 10;36(23):2378-2385. [doi: 10.1200/JCO.2018.77.8977] [Medline: 29791289]
8. Dai W, Feng W, Zhang Y, Wang XS, Liu Y, Pompili C, et al. Patient-reported outcome-based symptom management versus usual care after lung cancer surgery: a multicenter randomized controlled trial. J Clin Oncol 2022 Mar 20;40(9):988-996 [FREE Full text] [doi: 10.1200/JCO.21.01344] [Medline: 34995100]
9. Dai W, Wang Y, Liao J, Wei X, Dai Z, Xu W, et al. Electronic patient-reported outcome-based symptom management versus usual care after lung cancer surgery: long-term results of a multicenter, randomized, controlled trial. J Clin Oncol 2024 Jun 20;42(18):2126-2131 [FREE Full text] [doi: 10.1200/JCO.23.01854] [Medline: 38574304]
10. Basch E. Patient-reported outcomes - harnessing patients' voices to improve clinical care. N Engl J Med 2017 Jan 12;376(2):105-108. [doi: 10.1056/NEJMp1611252] [Medline: 28076708]
11. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
12. Lv C, Lu F, Zhou X, Li X, Yu W, Zhang C, et al. Efficacy of a smartphone application assisting home-based rehabilitation and symptom management for patients with lung cancer undergoing video-assisted thoracoscopic lobectomy: a prospective, single-blinded, randomised control trial (POPPER study). Int J Surg 2025 Jan 01;111(1):597-608. [doi: 10.1097/JS9.0000000000001845] [Medline: 38905505]
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
14. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
15. Sun V, Kim J. Symptom and functional recovery monitoring in thoracic surgery. J Thorac Dis 2020 Nov;12(11):6931-6939 [FREE Full text] [doi: 10.21037/jtd.2020.03.59] [Medline: 33282397]
16. Shei RJ, Holder IG, Oumsang AS, Paris BA, Paris HL. Wearable activity trackers-advanced technology or advanced marketing? Eur J Appl Physiol 2022 Sep 21;122(9):1975-1990 [FREE Full text] [doi: 10.1007/s00421-022-04951-1] [Medline: 35445837]
17. Kim J, Campbell AS, de Ávila BE, Wang J. Wearable biosensors for healthcare monitoring. Nat Biotechnol 2019 Apr;37(4):389-406 [FREE Full text] [doi: 10.1038/s41587-019-0045-y] [Medline: 30804534]
18. He Y, Chen Y, Tang L, Chen J, Tang J, Yang X, et al. Accuracy validation of a wearable IMU-based gait analysis in healthy female. BMC Sports Sci Med Rehabil 2024 Jan 02;16(1):2 [FREE Full text] [doi: 10.1186/s13102-023-00792-3] [Medline: 38167148]
19. van der Stam JA, Mestrom EH, Scheerhoorn J, Jacobs F, de Hingh IH, van Riel NA, et al. Accuracy of vital parameters measured by a wearable patch following major abdominal cancer surgery. Eur J Surg Oncol 2022 Apr;48(4):917-923 [FREE Full text] [doi: 10.1016/j.ejso.2021.10.034] [Medline: 34753618]
20. Cao R, Azimi I, Sarhaddi F, Niela-Vilen H, Axelin A, Liljeberg P, et al. Accuracy assessment of Oura ring nocturnal heart rate and heart rate variability in comparison with electrocardiography in time and frequency domains: comprehensive analysis. J Med Internet Res 2022 Jan 18;24(1):e27487 [FREE Full text] [doi: 10.2196/27487] [Medline: 35040799]
21. Zhang W, Zhou YN, Zhou Y, Wang JG. Validation of the watch-type HUAWEI WATCH D oscillometric wrist blood pressure monitor in adult Chinese. Blood Press Monit 2022 Oct 01;27(5):353-356. [doi: 10.1097/MBP.0000000000000608] [Medline: 35687029]
22. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022 Jan 21;24(1):e30791 [FREE Full text] [doi: 10.2196/30791] [Medline: 35060915]
23. Ray D, Collins T, Woolley S, Ponnapalli P. A review of wearable multi-wavelength photoplethysmography. IEEE Rev Biomed Eng 2023;16:136-151. [doi: 10.1109/rbme.2021.3121476]
24. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, et al. Detection of atrial fibrillation in a large population using wearable devices: the Fitbit heart study. Circulation 2022 Nov 08;146(19):1415-1424. [doi: 10.1161/circulationaha.122.060291]
25. Spatz ES, Ginsburg GS, Rumsfeld JS, Turakhia MP. Wearable digital health technologies for monitoring in cardiovascular medicine. N Engl J Med 2024 Jan 25;390(4):346-356. [doi: 10.1056/NEJMra2301903] [Medline: 38265646]
26. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, MAFA II Investigators. Mobile photoplethysmographic technology to detect atrial fibrillation. J Am Coll Cardiol 2019 Nov 12;74(19):2365-2375 [FREE Full text] [doi: 10.1016/j.jacc.2019.08.019] [Medline: 31487545]
27. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association

XSL•FO
RenderX

for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Feb 01;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
28. Beqari J, Powell J, Hurd J, Potter AL, McCarthy M, Srinivasan D, et al. A pilot study using machine-learning algorithms and wearable technology for the early detection of postoperative complications after cardiothoracic surgery. Ann Surg 2025 Mar 01;281(3):514-521. [doi: 10.1097/SLA.0000000000006263] [Medline: 38482684]
29. Ghomrawi HM, O'Brien MK, Carter M, Macaluso R, Khazanchi R, Fanton M, et al. Applying machine learning to consumer wearable data for the early detection of complications after pediatric appendectomy. NPJ Digit Med 2023 Aug 16;6(1):148 [FREE Full text] [doi: 10.1038/s41746-023-00890-z] [Medline: 37587211]
30. Kane WJ, Hassinger TE, Myers EL, Chu DL, Charles AN, Hoang SC, et al. Wearable technology and the association of perioperative activity level with 30-day readmission among patients undergoing major colorectal surgery. Surg Endosc 2022 Feb;36(2):1584-1592. [doi: 10.1007/s00464-021-08449-3] [Medline: 33782756]
31. Five years of evolution of HUAWEI TruSeen™ 5.0+, HUAWEI WATCH GT 3 achieves more accurate heart rate monitoring. Huawei. URL: https://consumer.huawei.com/sa-en/stories/wearables/news-gt3/ [accessed 2025-05-29]
32. Wang L, Xian H, Guo J, Li W, Wang J, Chen Q, et al. A novel blood pressure monitoring technique by smart HUAWEI WATCH: a validation study according to the ANSI/AAMI/ISO 81060-2:2018 guidelines. Front Cardiovasc Med 2022;9:923655 [FREE Full text] [doi: 10.3389/fcvm.2022.923655] [Medline: 36304535]
33. Wang XS, Gottumukkala V. Patient-reported outcomes: is this the missing link in patient-centered perioperative care? Best Pract Res Clin Anaesthesiol 2021 Dec;35(4):565-573 [FREE Full text] [doi: 10.1016/j.bpa.2020.10.006] [Medline: 34801218]

Abbreviations

AF: atrial fibrillation
bpm: beats per minute
EHR: electronic health record
ePRO: electronic patient-reported outcome
ERAS: enhanced recovery after surgery
HR: heart rate
MITS: minimally invasive thoracic surgery
PPG: photoplethysmography
PRO: patient-reported outcome

Edited by L Buis; submitted 02.12.24; peer-reviewed by H Yang, R Zhang, HT Bang; comments to author 18.03.25; revised version received 06.04.25; accepted 10.07.25; published 16.09.25.

Please cite as:
Wang R, Zheng J, Guo W, Huang H, Wang Q, Li Y, Lin M, Huang L, Zhang Q, Chen K, Ye Z, Deng H, Jiang Y, Lin Y, Feng Y, Huang Y, Chen Y, He J, Liang H
Integrating a Multimodal Digital Device for Continuous Perioperative Monitoring in Patients With Lung Cancer Undergoing Thoracic Surgery: Development and Usability Study
JMIR Mhealth Uhealth 2025;13:e69512
URL: https://mhealth.jmir.org/2025/1/e69512
doi:10.2196/69512
PMID:

©Runchen Wang, Jianqi Zheng, Wenwei Guo, Haiqi Huang, Qixia Wang, Yihong Li, Manwan Lin, Linchong Huang, Qing Zhang, Kaishen Chen, Zhiming Ye, Hongsheng Deng, Yu Jiang, Yuechun Lin, Yi Feng, Ying Huang, Ying Chen, Jianxing He, Hengrui Liang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Use and Application of mHealth Technologies in Perioperative Surgical Care: Narrative Review

Maria Camila Sierra[1*], MBBS, MCR; Henry To[1,2*], MBBS (Hons), MSurgEd, PhD; Wan Jun Song[3,4*], MBBS, MHM

[1]Department of Surgery, Werribee Mercy Hospital, Werribee, Melbourne, Australia

[2]Department of Medicine, Dentistry and Health Science, University of Melbourne, Melbourne, Australia

[3]Medical Director Hospital in the Home at Monash Health, Monash University, Melbourne, Australia

[4]Senior Researcher, Monash University, Melbourne, Australia

[*]all authors contributed equally

Corresponding Author:
Maria Camila Sierra, MBBS, MCR
Department of Surgery
Werribee Mercy Hospital
300 Princess Highway
Werribee, Melbourne, 3030
Australia
Phone: 61 03 8754 3000
Email: mcsizu@hotmail.com

Abstract

Background: Surgical procedures and their potential complications place substantial strain on patients, clinicians, and health care systems. These strains are driven by the anticipated morbidity and mortality, so that there is resource-intensive postoperative inpatient management. Given the concentration of surgical services within hospital settings, current standard levels of care have limitations such as communication gaps, time lapses before evaluation, and investment of resources, which limit accessibility and generate disparities in delivery of care. However, recent advances in digital health, including telemedicine platforms, mobile health (mHealth), and wearable technologies, present an opportunity to decentralize and extend perioperative care into community settings. This review explored how established mHealth technologies are being integrated into the perioperative pathway and their impact on surgical care delivery and outcomes. It also highlights possible emerging models of remote physician and patient interaction where benefits seem to be outweighing the risks.

Objective: The aim of this narrative review was to present collected evidence for the use of established mHealth technologies in the surgical pathway of patients and highlight their readiness and potential in models of standard care.

Methods: A comprehensive literature search was conducted across MEDLINE (via PubMed), Web of Science, and Scopus databases between October 2022 and May 2024. Additional sources were identified through reference list screening of relevant systematic reviews. Data were extracted and analyzed based on surgical specialty, type of mHealth intervention, cost-effectiveness, and ethical considerations. Findings were summarized in tables to illustrate key trends and variations across studies. The extracted data were tabulated and described qualitatively to highlight similarities, differences, and possible emerging trends across the studies.

Results: A total of 28 articles published between 2008 and 2022 were included for qualitative analysis, with most (n=21, 75%) originating from the United States, Germany, and the United Kingdom. The study designs were predominantly randomized controlled trials (n=9, 32%) and observational studies (n=8, 29%). Collectively, these studies involved 6344 patients undergoing mHealth-based perioperative interventions primarily in general surgery, orthopedics, and oncology. Interventions frequently used smartphones (n=10, 36%) and wearable devices, often in combination with other tracking and measuring systems. Applications included wound monitoring, postoperative follow-up, and patient education. Data collection was multimodal and typically conducted daily, yet only 36% (10/28) of the articles reported defined follow-up periods. Cost-effectiveness was rarely assessed, with only 4% (1/28) of the articles reporting per-patient savings. Overall, 64% (18/28) of the articles were rated as low quality due to methodological limitations.

Conclusions: mHealth- and telehealth-based interventions show promise in enhancing aspects of perioperative care by enabling remote monitoring, patient engagement, and improved care continuity. Future research should focus on scalable implementation,

XSL•FO
RenderX

true cost-effectiveness analysis, equitable access, and integration into clinical workflows to ensure broad applicability in current models of care.

(JMIR Mhealth Uhealth 2025;13:e52206) doi:10.2196/52206

KEYWORDS

surgical specialties; general surgery; postoperative care; perioperative care; ambulatory monitoring; telehealth; telemonitoring; remote consultation; mHealth

Introduction

Background

Surgical procedures and associated complications impose a significant burden on patients, treating teams, and health care systems that are related to the associated morbidity and mortality of the illness and treatment [1,2]. Consequently, there is a need for surgical management and follow-up to be conducted as inpatient care, where there is a wide availability of resources to attend to the circumstances in the patient surgical pathway. As procedures and aftercare are delivered at focal points (hospitals) where patients and resources are located, the number of patients assessed and treated is limited, and available resources vary greatly among institutions, cities, and countries, which places an important limitation on the standard of care worldwide [1-3].

The rapid progression of digital technology and artificial intelligence and the creation of new wireless wearable devices in recent years has initiated a transformation across numerous industries, and health care is now experiencing firsthand a transformation of its own [4]. Advances in mobile health (mHealth) technologies have allowed components of the care for patients to be adequately conducted away from the traditional hospital settings [5]. Telemedicine, mHealth, and similar technologies are a rapidly growing area where existing mobile technology is applied to aspects of standard care so that the same service can be provided in the community in a safe and tailored environment [5,6]. Despite innovations in digital technology, the surgical domain has understandably approached the integration and implementation of new and potentially revolutionary technologies with caution given the immediate and direct implications for patients' safety [7]. There is an increasing overwhelming need for community-based postsurgical care to be provided in settings where there is a higher level of patient comfort and convenience and be tailored to the individual's needs.

Given that the current standard of care for most surgical patients following hospital discharge does inherently present limitations such as communication gaps, time lapses before evaluation, and investment of resources, strategies in the perioperative setting can be analyzed at many levels to improve their efficiency and effectiveness. In this sense, mHealth strategies applied to the surgical patient pathway display great diversity in goal setting and applications, with hypothesized consequential improvements in patients' perioperative well-being and management of health changes [8-10]. In addition, the redistribution of the service burden can inherently impact the health economic budget distribution positively by redirecting inpatient hospital resources to those in most need [8,10].

There is already a great deal of technologies used in the community setting that can be accessed for health needs [4-6]. Data networks already exist to create, access, and securely transfer data remotely from individuals or care centers. The ubiquitous use of smartphones and wearable devices already offers the opportunity for medical data to be collected and stored in a more reliable, agile, and effective way [11]. These technological conditions paired with the good digital literacy in the community enable digital education, transition, and uptake rates to be more feasible. There is some available evidence that demonstrates that highly complex tasks within the hospital, such as performing a surgical procedure and monitoring patients with acute complex conditions, can now be performed remotely in a safe manner [12]. Mobile technologies can also provide the technological link to a remote group of medical experts, increasing the interconnectedness and globalization of health care as well as access to specialist care in remote areas [13].

Objectives

Given the readiness of the existing technologies and their potential, it is worth evaluating their application within the perioperative field. In this narrative review, we aimed to answer the questions of how already established mobile technologies (mHealth) can assist in the perioperative care of surgical patients, what their potential is for change in management, and what the different implementation possibilities are within the physician-patient interaction model. We collated research in which innovations were applied as interventions and outcomes were qualitatively analyzed.

Methods

Search Strategy

A comprehensive literature search was conducted between October 2022 and May 2024 across 3 electronic databases: MEDLINE (via PubMed), Web of Science, and Scopus. The search strategy was developed collaboratively by the study investigators and an academic research librarian to ensure methodological rigor.

The search combined controlled vocabulary terms (eg, MeSH [Medical Subject Headings]) and free-text keywords related to surgical specialties, general surgery, postoperative and preoperative care, ambulatory monitoring, telehealth, telemonitoring, telemedicine, remote consultation, mHealth, and digital health interventions. Synonyms and related terms were used in combination with Boolean operators (AND and OR) to broaden retrieval sensitivity. The search was limited to English-language publications from January 1, 2000, onward. The detailed search strategy for MEDLINE is provided in Multimedia Appendix 1. The search terms and syntax were

XSL•FO
RenderX

adapted to Scopus and Web of Science using database-specific indexing systems and search functions.

In addition to database searching, the reference lists of relevant systematic reviews were screened to identify additional studies meeting the inclusion criteria.

Study Selection

This narrative review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. After duplicate removal, titles and abstracts were independently screened by 2 reviewers (HT and MCS) to attain the final cohort of studies. Full-text articles were then retrieved and assessed against predefined inclusion and exclusion criteria (Textbox 1). In cases in which abstracts were unavailable or the available data were insufficient to determine relevance, the full article was reviewed and cross-checked against the research question and the predetermined eligibility criteria. Disagreements during the screening process were resolved through discussion between the 2 main researchers or consultation with a third reviewer (JS) when deemed relevant.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- English-language studies
- Human participant studies
- Original (new) data presentation and analysis

Exclusion criteria

- Studies that required the clinical team to be present in real time for the device to function
- Articles in languages that were not English
- Studies that involved patients with chronic illnesses
- Literature review articles
- Animal studies
- Studies on intraoperative interventions
- Study protocols
- Conference abstracts for which data could not be obtained
- Studies on telementoring
- Studies on remote monitoring by clinicians rather than remote monitoring performed by the patients

Eligibility Criteria

Eligible articles were required to be original research papers published in English reporting studies involving human participants and presenting primary data analysis. Studies were excluded when the publication was in a language other than English, involved populations with chronic illnesses, or based their research on animal models. Other types of evidence such as literature reviews, study protocols, or conference abstracts lacking accessible data or studies that focused mainly on intraoperative interventions were also not deemed appropriate for data analysis. Articles that described telementoring or studies that involved remote monitoring by clinicians rather than patients were also excluded. This review was restricted to publications made from the 2000s onward to capture advancements in mHealth technologies (Textbox 1).

Quality Appraisal

Although this is a narrative review, quality assessment was conducted to evaluate the strength and rigor of the evidence within the domains of mHealth and telemedicine. This step was deemed essential due to the emerging nature of these technologies and their growing relevance to contemporary models of care and intervention. Quality appraisal was undertaken independently by the researchers using criteria such as study design, risk of bias, potential confounding, allocation concealment, selective outcome reporting, loss to follow-up, early study termination, and imprecision of results. The appraisal was conducted under the framework and principles outlined in the article *Grading quality of evidence and strength of recommendations in clinical practice guidelines* [14]. Any discrepancies in quality rating were resolved through discussion. In accordance with the PRISMA guidelines and standards for reporting, transparency, and reproducibility, a summary of the quality assessment is provided in Multimedia Appendix 2 [10,15-41].

Data Extraction

Data were extracted from full-text articles using a standardized extraction form developed specifically for this review to ensure consistency throughout the data collection process. Extracted data included publication details, surgical subspecialty, study design, sample size, technology used, metrics assessed, data collection frequency, and follow-up duration. Additional data points included patient satisfaction, safety outcomes, complication rates, cost analyses, and ethical considerations.

Data extraction was carried out independently by the 2 primary reviewers (HT and MCS) using Microsoft Excel. Any

XSL•FO
RenderX

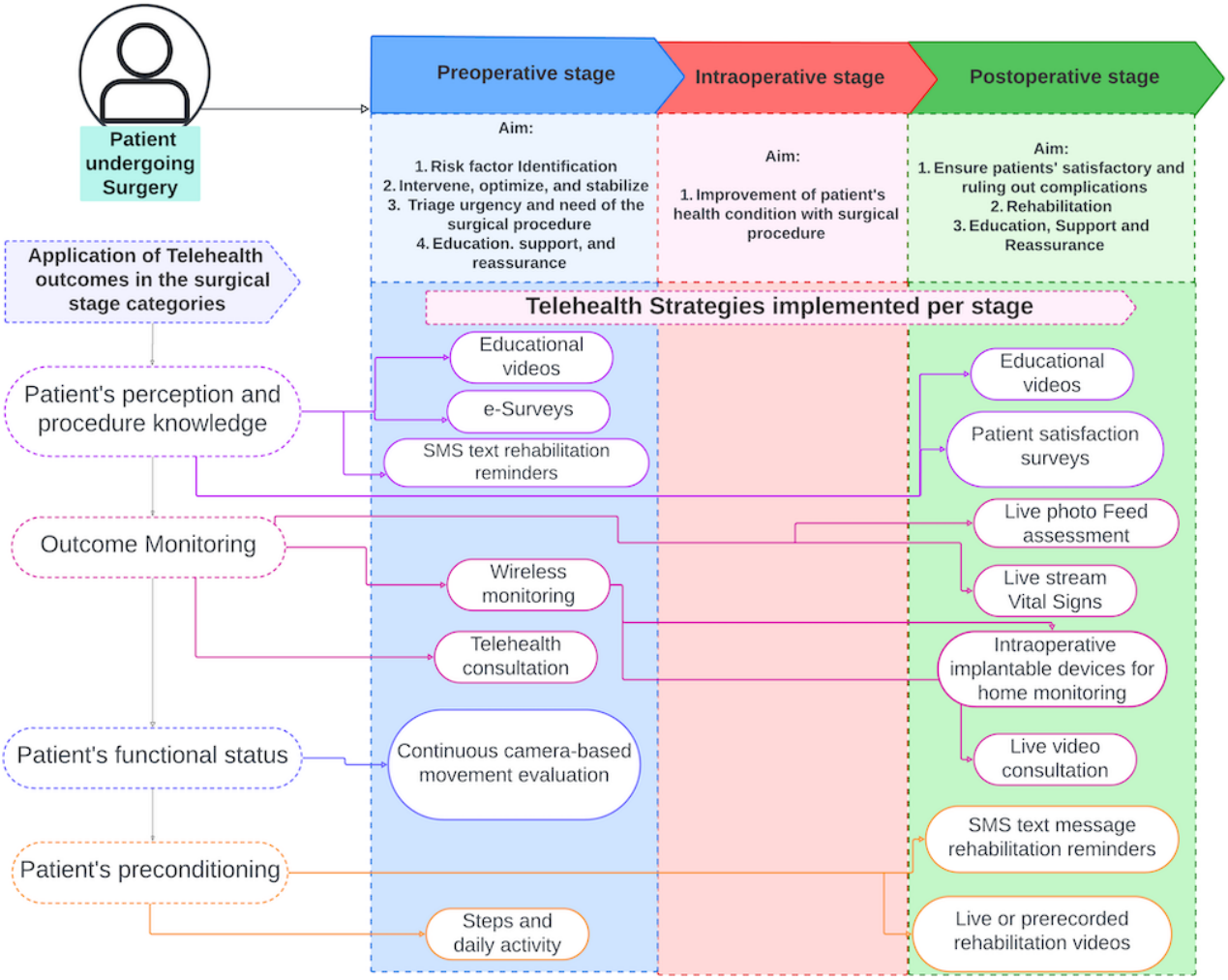
discrepancies were addressed through discussion. No automation tools were used for data extraction purposes.

Synthesis of Results

A narrative synthesis was undertaken to summarize the findings of the included studies. Results were grouped thematically based on surgical specialty, type of mHealth or telemonitoring technology used, cost-effectiveness, and ethical considerations. The extracted data were tabulated and described qualitatively to highlight similarities, differences, and possible emerging trends across the studies.

No meta-analysis was conducted due to the heterogeneity of study designs, outcome measures, and interventions. Where applicable, illustrative examples were provided to contextualize the findings and support interpretation (Figure 1).

Figure 1. Patient surgical pathway (preoperative stage, intraoperative stage, and postoperative stage).



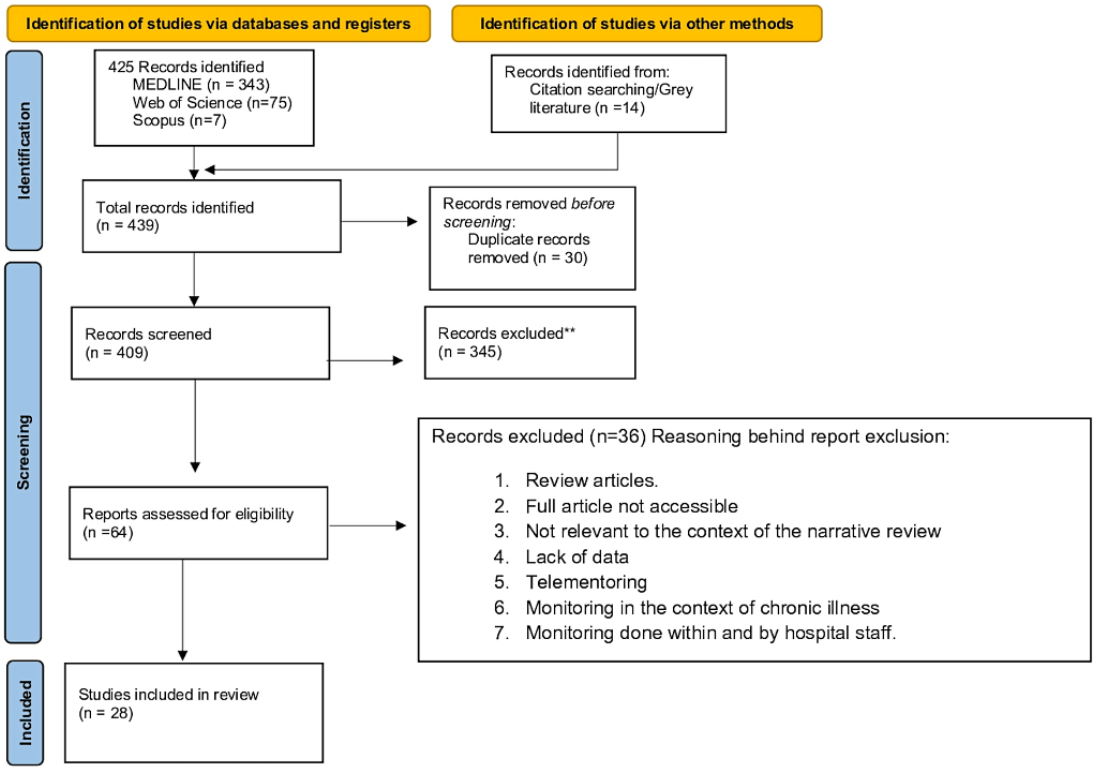
Results

Study Selection

A total of 439 records were initially identified through database searching (n=425, 96.8%) and gray literature (n=14, 3.2%). After removal of 6.8% (30/439) of duplicates, 409 articles remained for title and abstract screening, from which 64 (15.6%) publications were deemed potentially relevant and underwent full-text review, as exemplified in Figure 2. Following eligibility assessment, 56% (36/64) of the articles were excluded due to lack of relevance to the narrative review, lack of accessible full text, insufficient data, focus on telementoring, or focus on monitoring in the context of chronic illnesses. A final selection of 44% (28/64) of the full-text articles was retained for data extraction [10,15-41]. A table with full disclosure of the selected references can be found in Multimedia Appendix 3 [10,15-41].

XSL•FO
RenderX

Figure 2. PRISMA flowchart describing the screening methodology for the articles selected for this review. **No automation tools were used; all references were screened by title and abstract by the two main researchers (MCS and HT). These articles were not deemed relevant based on the title and abstract. Image adapted from Page et al [42].



Study Characteristics

The 28 included articles for data extraction were published between 2008 and 2022, and more than half (n=21, 75%) predominantly originated from the United States, Germany, and the United Kingdom. These countries had a publication upsurge from the year 2013 onward (Table 1).

Of the 28 included publications, 9 (32%) were randomized controlled trials, representing the highest level of evidence. Observational studies accounted for 29% (8/28) of the articles, and prospective and retrospective cohort studies each accounted for 4% (1/28) of the articles. Table 1 exhibits a trend toward higher levels of evidence in publications from 2018 onward when compared to previous years.

The included articles predominantly reported mHealth and telemonitoring strategies across 3 main specialties: general surgery, orthopedics, and oncology (Table 1). The most represented specialty was general surgery with 25% (7/28) of the studies, followed by orthopedics with 18% (5/28) of the studies and oncology with 11% (3/28) of the studies. Other specialties, including vascular surgery, obstetrics, procedural dermatology, and plastic surgery, each accounted for 7% (2/28) of the studies. Additional fields such as urology, transplant surgery, and nursing were each represented by 4% (1/28) of the studies.

XSL•FO
RenderX

Table 1. General characteristics of the included studies.

	Studies, n (%)
Studies published in 2008-2012 (n=5)	
Country of origin	
United States	3 (60)
Germany	1 (20)
Italy	1 (20)
Surgical subspecialty	
Nursing	1 (20)
Plastic surgery	1 (20)
Microsurgery	1 (20)
Oncology	1 (20)
Vascular surgery	1 (20)
Study design	
RCT[a]	2 (40)
Case-control	1 (20)
Observational study	1 (20)
Prospective cohort study	1 (20)
Type of journal	
Clinical	5 (100)
Studies published in 2013-2017 (n=13)	
Country of origin	
United States	9 (69)
Australia	1 (8)
Germany	1 (8)
United Kingdom	1 (8)
Spain	1 (8)
Surgical subspecialty	
Orthopedics	4 (31)
General surgery	3 (23)
Dermatology	1 (8)
Oncology	1 (8)
Transplant	1 (8)
Neurosurgery	1 (8)
Vascular surgery	1 (8)
Vascular and general surgery	1 (8)
Study design	
Observational study	6 (46)
Pilot study	2 (15)
Audit	1 (8)
Case series	1 (8)
Cross-sectional study	1 (8)
Retrospective cohort study	1 (8)
RCT	1 (8)

XSL•FO
RenderX

	Studies, n (%)
Type of journal	
Clinical	13 (100)
Studies published in 2018-2022 (n=10)	
Country of origin	
Germany	2 (20)
United Kingdom	2 (20)
United States	2 (20)
Canada	1 (10)
Iran	1 (10)
The Netherlands	1 (10)
The Philippines	1 (10)
Surgical subspecialty	
General surgery	4 (40)
Obstetrics	2 (20)
Orthopedics	1 (10)
Plastic surgery	1 (10)
Urology	1 (10)
Dermatology	1 (10)
Study design	
RCT	6 (60)
Case-control	1 (10)
Case series	1 (10)
Observational study	1 (10)
Systematic review	1 (10)
Type of journal	
Clinical	7 (70)
Technology	2 (20)
Global health	1 (10)
Total articles included (n=28)	
Country of origin	
United States	14 (50)
Germany	4 (14)
United Kingdom	3 (11)
Australia	1 (4)
Canada	1 (4)
Italy	1 (4)
Iran	1 (4)
The Netherlands	1 (4)
The Philippines	1 (4)
Spain	1 (4)
Surgical subspecialty	
General surgery	7 (25)
Orthopedics	5 (18)

XSL•FO
RenderX

	Studies, n (%)
Oncology	3 (11)
Vascular surgery	2 (7)
Dermatology	2 (7)
Obstetrics	2 (7)
Plastic surgery	2 (7)
Microsurgery	1 (4)
Nursing	1 (4)
Transplant	1 (4)
Urology	1 (4)
Vascular and general surgery	1 (4)
Study design	
RCT	9 (32)
Observational study	8 (29)
Case-control	2 (7)
Case series	2 (7)
Pilot study	2 (7)
Audit	1 (4)
Cross-sectional study	1 (4)
Prospective cohort study	1 (4)
Retrospective cohort study	1 (4)
Systematic review	1 (4)
Type of journal	
Clinical	25 (89)
Technology	2 (7)
Global health	1 (4)

[a]RCT: randomized controlled trial.

Study Populations and Sample Sizes

Across the 28 studies, a total of 6344 patients participated in mHealth-related surgical research. Study cohorts varied widely in size, ranging from ≤50 patients to ≥1000 patients. As shown in Table 2, most studies had small sample sizes, with 46% (13/28) including ≤50 patients. In contrast, only 7% (2/28) of the studies recruited sample sizes as large as ≥1000 patients. Studies with intermediate sample sizes (51-500 patients) were the second most common type encountered in the collected evidence, representing 36% (10/28) of the selected studies.

Of the 28 articles deemed relevant, 14 (50%) reported the use of multidevice strategies for patient monitoring. These strategies combined smartphones, computers, tablets, smartwatches, and other unique wearable devices for outcome measurement (Table 2). Figure 1 provides a visual summary of the strategies assessed and implemented across the patient surgical pathway.

XSL•FO
RenderX

Table 2. Data features from the included studies (N=28).

Feature	Studies, n (%)
Sample size	
≤50 patients	13 (46)
51-500 patients	10 (36)
501-1000 patients	3 (11)
≥1001 patients	2 (7)
Devices used for data collection	
Smartphone only	10 (36)
Tablet only	1 (4)
Computer only	0 (0)
Multidevice (combination of any of the aforementioned options or other external devices)	14 (50)
Not available	3 (11)
Parameters measured	
Vital signs	0 (0)
SSIs[a]	2 (7)
Patient satisfaction	3 (11)
Flap status and viability	2 (7)
Disability	0 (0)
Pain	0 (0)
Complications	0 (0)
Finances	1 (4)
Other (single option exclusive to a single study)	5 (18)
Multiple parameters	15 (54)
Metrics used	
Surveys or questionnaires	4 (14)
Photos	3 (11)
Scoring systems	2 (7)
Rating scales	0 (0)
Costs	1 (4)
Other (combination of any of the aforementioned options)	14 (50)
Not available	4 (14)
Follow-up period	
Before surgery	1 (4)
0-72 h after surgery	3 (11)
>72 h up to 1 mo after surgery	10 (36)
>1 mo-1 y after surgery	6 (21)
>1 y after surgery	0 (0)
Not available	8 (29)
Data collection frequency	
Daily or less	14 (50)
At least once a week or less	4 (14)
Consistently (fortnightly)	3 (11)
Not available	7 (25)

XSL•FO
RenderX

[a]SSI: surgical site infection.

Devices and Outcome Measurement

Several studies (5/28, 18%) focused on unique wearable devices for functional outcome measurement. Kayaalp et al [20], Chiang et al [26], and Calliess et al [32] validated new wearable-based gait kinematic analysis systems to assess functional recovery, implant longevity, and quality of life after total knee arthroplasty. Kayaalp et al [20] demonstrated that a new wearable system for knee function monitoring in healthy individuals showed comparable accuracy to that of motion capture systems, enabling postoperative mobility monitoring without restricting daily activities. Chiang et al [26] and Calliess et al [32] further supported the feasibility of wearable devices for continuous, objective assessment of functional outcomes.

The second most frequently reported mHealth strategy was the use of smartphones alone. Their use was reported in 36% (10/28) of the articles. Most smartphone-based interventions (10/28, 36%) revolved around web-based wound care and early detection of surgical site infections through photographic monitoring (3/28, 11%).

Henarejos et al [16] evaluated the effectiveness of an mHealth wound care education program in low- and middle-income countries using photographic monitoring to capture wound images during a 30-day follow-up. The images were shared with surgical teams via a digital exchange platform for timely intervention based on a color-coded classification system. The intervention resulted in a 19.1% reduction in surgical site infections despite study limitations associated with small sample size, loss to follow-up, and challenges related to health care infrastructure.

McLean et al [17] and Gunter et al [33] investigated smartphone-based models as assessment tools for early wound care, diagnosis, and intervention applied in the postoperative stage. McLean et al [17] focused on emergency abdominal surgery outcomes, including time of diagnosis, wound review attendance, and infection severity, whereas Gunter et al [33] focused on user satisfaction and validated a smartphone-based model as a wound monitoring system in vascular surgery. Both studies concluded that a smartphone-based model enabled postoperative follow-ups, early postoperative interventions, triage to appropriate care levels, and patient satisfaction.

Similarly, Hee Hwang and Mun [39] and Engel et al [15] explored smartphone-based models for remote free flap monitoring. Hee Hwang and Mun [39] demonstrated that smartphone-based flap assessment improved early diagnosis of flap compromise, increasing flap salvage rates from 96.2% to 100% and reducing reintervention times to 1.4 hours compared to a baseline of 4 hours. Engel et al [15] further validated the use of smartphone-based models as assessment systems and provided comparable accuracy rates, with shorter response times than those in in-person examinations.

Smartphones were also evaluated in the collection of data relevant to patient-centered outcomes, including rehabilitation, compliance, patient satisfaction, wound care reminders, and adherence to postsurgical rehabilitation programs. The studies by Hawkins et al [21], Segura-Sampedro et al [24], and Chee et al [28] appraised patient acceptance of smartphone-based models for follow-up. Segura-Sampedro et al [24] found that telemedicine-based follow-up services had a high sensitivity and specificity, supporting their feasibility and safety. Collectively, these studies reported reduced patient anxiety, improved postsurgical care understanding, and high acceptance of a self-monitoring program.

Although tablet-based interventions were less common, they were primarily used for patient satisfaction assessment and symptom tracking. Ertel et al [30] analyzed the feasibility of a video-based educational program and a telehealth monitoring model for post–liver transplantation care, including postoperative care management and readmission rates. The study concluded that the intervention was feasible, improved vital statistic maintenance, and enhanced patient awareness of preoperative and postoperative care management.

No articles identified in the literature search reported the use of artificial intelligence–based data capture, analysis, or monitoring.

Data Collection and Metrics

As shown in Table 2, the studies used a wide and variable range of data collection methods, including surveys, questionnaires, photos, rating scales, and scoring systems. In total, 50% (14/28) of the studies used a combination of these metrics to assess patient-reported outcomes, satisfaction levels, attitudes toward mHealth interventions, home monitoring, and pain or disability.

For instance, Khoshrounejad et al [25] investigated the impact of an SMS text message–based rehabilitation program after corrective hand surgery using questionnaires and rating scales to measure variables such as pain, disability, physiotherapy adherence, and patient satisfaction. The intervention demonstrated increased motivation for self-directed rehabilitation. Nevertheless, reduced rehabilitation session attendance was noted.

The studies by Schramm et al [22], Sun et al [27], and Card et al [38] all used combined metrics for patient-centered outcomes using clinical measures. The study conducted by Sun et al [27] demonstrated that wireless monitoring programs involving real-time symptom tracking and mobility and quality of life assessment could be applied effectively among patients with cancer undergoing major abdominal surgery. Furthermore, wireless monitoring programs facilitated early intervention, enhancing perioperative care when applied across the patient surgical pathway (Figure 1) by detecting real-time changes in symptom severity, mobility, and quality of life.

Data collection frequency varied, with 50% (14/28) of the articles reporting near-daily data collection (Table 2), whereas 11% (3/28) of the articles reported medium-term data collection (more frequently). Measurement duration variability was dependent on the parameters measured; for example, symptom burden, disability, rehabilitation progress, and patient satisfaction were commonly assessed over longer follow-up durations.

XSL•FO
RenderX

Despite frequent data collection being reported in most articles, 29% (8/28) of the articles did not report a follow-up period, leading to the assumption that data collection ended at patient discharge. Among the 36% (10/28) of the articles that specified a follow-up time frame, monitoring ranged from 72 hours to 1 month after surgery.

Economic Outcomes and Cost-Effectiveness

Only 7% (2/28) of the studies attempted to evaluate cost-related outcomes, with the study by Rosner et al [37] being the only one to systematically assess the financial impact of mHealth strategy application. The multicenter observational cohort study examined cost reductions in a 90-day period associated with an automated digital patient engagement platform in patients undergoing hip and knee arthroplasty. The intervention led to mean cost savings of US $656.52 per patient along with a relative reduction in 90-day readmission rates, hospitalization costs, and postoperative complications.

Barber et al [41] assessed telemetric intracranial pressure monitoring systems, which enabled home-based prolonged and safe intracranial pressure monitoring in neurosurgical patients. The system provided early reporting of combined outcomes across the preoperative and postoperative periods in the patient surgical pathway. Outcomes assessed included detection of clinical positive patient experiences, financial outcomes related to the outlay during pre- and postimplantation, and patient events. Clinical outcomes such as clinical events and associated complications were also deemed relevant for assessment. The system implementation yielded cost benefits by reducing unnecessary investigations.

Quality of Evidence

The quality assessment of the selected studies is detailed in Multimedia Appendix 2. This was conducted by the main researchers individually, and 64% (18/28) of the selected studies were classified as low-quality evidence. Common methodological limitations included study design issues, lack of blinding, publication and selection bias, and small sample sizes, which limited the strength of associations. mHealth research is concentrated within certain surgical specialties and typically limited to a single stage of the patient's pathway, mostly the postoperative period. As a result, its generalizability is limited, since findings cannot be easily applied across other specialties or the entire surgical pathway in different models of care.

Discussion

Principal Findings

Overview

Digital health care technologies, including mHealth strategies and telehealth, represent a rapidly evolving area where technology and health care intersect across the perioperative surgical care continuum. This narrative review highlights the significant opportunities that these technologies offer to enhance perioperative surgical care, particularly in the context of increasing health costs, limited inpatient resources, and the complexity of modern surgical procedures [43,44].

While telemedicine and remote consultation are already well integrated and show established benefits in clinical settings such as clinics, emergency departments, and rural trauma systems [15,45,46], this review demonstrated how successful digital health care technology implementation across the patient surgical pathway can be, impacting areas such as communication, patient convenience, and cost-effectiveness and care outcomes. The implementation of these technologies could streamline follow-up care and rehabilitation and significantly impact gaps in traditional surgical care, with the potential to substitute some components of inpatient care in a community setting.

Devices and Outcome Measurement

mHealth interventions often leverage common consumer technologies such as smartphones, tablets, smartwatches, and computers without requiring specialized or complex infrastructure. For instance, McGillicuddy et al [18] demonstrated how smartphones combined with wireless blood pressure monitors and medication trays supported postoperative care in patients undergoing transplants. Innovative approaches in the orthopedic field also showed promise, with 7% (2/28) of the studies developing and validating technologies for motion assessment in orthopedic care. These systems provided a cost-effective alternative to the gold-standard camera-based gait analysis systems using mobile technology, offering greater accessibility to patient condition assessment.

In the preoperative stage of the surgical patient journey, mHealth interventions improved new measuring strategies for patient engagement, assessment, and consistency and facilitated prehabilitation, as well as encouraging new modalities to provide informed consent and education. They also aided and positively influenced patients' behavior, expectations, and compliance, aspects that highly influence the postoperative stage, potentially leading to future complications or truncated recovery if not handled adequately. Hawkins et al [21] found that tailored educational videos reduced procedural anxiety and improved wound care understanding. Similarly, Migden et al [47] reported that preconsent video education enhanced consultation efficiency and patient satisfaction. Furthermore, preoperative SMS text messages were shown to support medication adherence and treatment continuity [48,49]. The implementation of these strategies demonstrates how digital tools can positively influence patient outcomes and impact surgical ones.

Postoperative applications focused on remote monitoring, rehabilitation, and complication detection. Studies demonstrated benefits in patient-led rehabilitation, early complication detection through virtual wound care, functional recovery tracking via wearables, enhanced symptom reporting and triage, assessment of return to function and quality of life, and supporting patient education and self-management.

Considering the potential for communication gaps between patients and care teams during the first 2 to 3 weeks after patient discharge, interventions that facilitate detection of complications within the patients' journey toward recovery, as well as early communication of rehabilitation modifications and recommencement of individual daily life activities, offer an opportunity for mHealth systems and groundbreaking

XSL•FO
RenderX

technologies to be explored and applied. Several studies (18/28, 64%) highlighted the effectiveness of mHealth in this matter [50,51]. For instance, Khoshrounejad et al [25] found that SMS text message–based rehabilitation systems improved short-term hand function following tendon repair, whereas Chen et al [52] reported improved rehabilitation adherence and range of motion in patients with frozen shoulder who received motivational SMS text messages.

Wearable devices were particularly useful when assessing functional monitoring during early rehabilitation by offering synchronous and objective data collection. This approach addressed the often poor correlation between subjective self-report and actual physical activity, laying the groundwork for mHealth interventions to assist in more accurate, patient-centered recovery monitoring [29].

Wound care has emerged as a promising area for virtual monitoring, leveraging mobile device use to detect early complications and escalate standard care practices accordingly. Studies have shown that smartphone-based imaging models for wound assessment are equivalent to in-person examination even for complex wounds such as pressure ulcers [53]. Data highlight benefits such as travel time reduction, time-efficient triage, and greater health care provider productivity [53-55]. These findings support the plausible and potentially efficient integration of digital photography into standard wound care protocols [33].

Data Collection and Metrics

Multidevice systems improve measurement accuracy and response timeliness of patients to clinicians. McGillion et al [34] found that, although remote automated monitoring technology did not increase *days alive at home* after surgery, it enabled early identification and correction of medication errors, reduced pain levels, and lowered readmission and emergency visits within 30 days. These outcomes underline the acceptability of technological implementation in the context of virtual care by patients and physicians, as well as the potential cost-effectiveness of remote automated monitoring technology in postoperative care [34].

Finally, mHealth system implementation has also proven to positively impact quality of life and return to baseline function. Basch et al [35] demonstrated that web-based symptom reporting in oncology patients led to more timely interventions, improved symptom control, fewer emergency visits, and enhanced survival. This emphasizes how mHealth may augment clinician awareness and symptom management.

Despite the many promising findings, several challenges persist within this area of research. Many studies lacked standardized outcome measures, long-term follow-ups, or cost-effectiveness analyses. For example, Goode et al [36] reported no significant effects of mHealth interventions on postprostatectomy incontinence. This highlights the need for further and more robust research on the role of digital interventions in specific clinical contexts, particularly where in-person services are unavailable.

Economic Outcomes and Cost-Effectiveness

In terms of the cost-effectiveness of mHealth systems in surgery, it is clear that there is limited direct analysis available. Most of the existing data stem from inferred associations linking the benefits of mHealth systems to improved surgical outcomes rather than providing concrete cost-effectiveness evaluations. Notably, the study by Rosner et al [37] stands out as the only one in this review to directly explore and quantify the cost-effectiveness of mHealth application within the surgical context. Their findings revealed that patients using an automated digital patient engagement platform following hip and knee arthroplasty experienced a 54.4% reduction in potentially avoidable complications. This had direct financial implications, with the study reporting an average saving of US $656.52 per patient at 90 days after surgery [37].

Quality Assessment and Evidence Strength

mHealth systems and implementation of new technologies have several challenges to overcome as universal access and targeted assistance become a pressing priority within the field. This system implementation should be able to facilitate the creation of protocols that respond to the needs of diverse population groups, such as people with less digital literacy than the average population and people with special needs or difficulties when handling new devices due to their lack of familiarity with new technology. In this era of interconnectedness, stringent data security in mHealth apps and telemedicine is imperative, protecting patient security and private data. The implementation of these systems requires new strategies for patient confidentiality and data safeguarding to be optimized and improved. The creation of tools, wireless wearable devices, and other elements for mHealth and telemedicine application requires the active participation of health care professionals to direct objective assessment and fully fulfill the aims of the strategies, where protocol compliance should be top priority. Inherently, a formal cost-utility analysis is necessary as reimbursement for non–face-to-face care continues to be problematic under traditional fee-for-service payment models that may not yet include virtual assessment implementation [31].

While conducting this review, some gaps in the literature were identified. These included limited discussion on ethical and safety considerations regarding mHealth system implementation in the patient surgical pathway and few articles discussing how to upscale and roll out their interventions, as well as the implementation in different models of care. Furthermore, the data did not provide evidence suggesting the integration of mHealth systems into standard hospital-level care and operations, which, in conjunction with lack of devices, standardization, and follow-up timing standardization, jeopardizes the immediate implementation of these devices at a larger scale within the surgical field.

Regardless of the vast number of publications on the use of telemedicine, wireless wearable devices, and mobile apps in surgical specialties, the available data collected in this narrative review still need to be read and evaluated with caution as the collective impact continues to be varied and there seem to be substantial questions regarding the quality appraisal of the

XSL•FO
RenderX

available evidence. When reading the available data, one must consider that studies are being conducted on a very small scale, compromising the possibility of establishing strong associations, which is a persistent limitation noted. Such limitations stress the need for multicenter trials and inclusive research that considers diverse health care contexts and infrastructure.

From the evidence gathered, none of the studies explored ethical considerations or system-level integration, making result generalizability an important aspect to consider in future research. Considering that the mHealth and telemedicine landscape is complex and varies from country to country, it is worth highlighting that most published articles and guidelines available are based on expert opinions, with evidence coming from high-income countries where strategies are applied to very specific scenarios and in highly controlled situations. Data then fail to provide insights on the challenges that may arise for low- and middle-income countries, where there is a substantial difference in circumstances, infrastructure, resource availability, and digital and educational literacy.

Conclusions

Considering the context of a changing health care delivery system with a shift to community-based care, mHealth is aiding more creative patient monitoring methods in the outpatient setting. mHealth and telemedicine can be effectively implemented to actively engage patient populations, making individuals active participants in the process of monitoring their own conditions and actively involving them in the decision-making processes, which overall improves the experience, efficiency, and outcomes of care.

There is an increasing body of evidence pointing toward mHealth and telemedicine demonstrating significant promise across the surgical care field, but this evidence should be interpreted with caution. Many interventions have been evaluated in small, highly controlled studies, which limits generalizability. In addition, further research is needed to address implementation barriers; validate cost-effectiveness; and develop scalable, standardized frameworks applicable across health care systems in multiple settings and contexts. Future work should prioritize inclusive design, robust ethical safeguards, and equitable access, ensuring that digital solutions are effective, acceptable, and sustainable across diverse patient populations and care environments.

Despite the highlighted limitations, digital health technologies, including mHealth and telehealth interventions, show a considerable potential to improve perioperative surgical care. These tools' strengths lie in enhancing patient education, enabling remote monitoring, supporting rehabilitation, and reducing unnecessary hospital visits, contributing to improved outcomes and system efficiency. Equally, there is evidence suggesting that mHealth technologies seem to have the potential to improve outcomes, reduce costs, and address gaps in perioperative care, particularly if future efforts prioritize equitable access, robust evaluation, and seamless integration into clinical workflows.

Acknowledgments

The authors would like to acknowledge the support of the Monash University and University of Melbourne library services for providing access to databases and literature search resources. They kindly thank Lorena Romero, librarian, for her assistance in providing material and guidance to support a well-structured research strategy, including input to refine the search strategy. The authors extend special thanks to the JMIR editorial team and assigned peer reviewers for all the input provided for this manuscript to be as complete as possible. During the preparation of this manuscript, generative artificial intelligence tools such as ChatGPT by OpenAI were used to assist with language refinement and formatting guidance for publication purposes. Threads have been provided in Multimedia Appendices 4 and 5. Despite the use of generative artificial intelligence tools, all outputs were critically reviewed, verified, and edited by the authors to ensure accuracy, clarity, and integrity. All authors take full responsibility for the content of this publication.

Data Availability

Data sharing is not applicable to this article as no data sets were generated during this study. All data analyzed during this study is included in this published article and can be found in Multimedia Appendix 6 [10,15-41].

Authors' Contributions

MCS and HT conceptualized and designed the narrative review. HT assisted in the process of article screening, quality appraisal, and data extraction. MCS led the process of article screening, quality appraisal, and data extraction and drafted the manuscript, with substantive revision from HT and JS. All authors contributed to the final version of the manuscript and approved it for final submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy used for data queries in MEDLINE (via PubMed).

XSL•FO
RenderX

[DOCX File , 16 KB - mhealth_v13i1e52206_app1.docx]

Multimedia Appendix 2
Quality appraisal table detailing the methodological evaluation of the included studies using the Grading of Recommendations Assessment, Development, and Evaluation framework.
[DOCX File , 15 KB - mhealth_v13i1e52206_app2.docx]

Multimedia Appendix 3
Summary table of the selected studies, outlining authors, publication year, country where the study originated, journal, and specialty.
[DOCX File , 24 KB - mhealth_v13i1e52206_app3.docx]

Multimedia Appendix 4
Evidence of ChatGPT (OpenAI) use during manuscript development, including prompts and artificial intelligence–generated content with author verification for language refinement purposes.
[DOCX File , 20 KB - mhealth_v13i1e52206_app4.docx]

Multimedia Appendix 5
Evidence of ChatGPT (OpenAI) use during manuscript development, including prompts and artificial intelligence–generated content with author verification for formatting purposes for publication.
[DOCX File , 25 KB - mhealth_v13i1e52206_app5.docx]

Multimedia Appendix 6
Supplementary summary tables created for the narrative review.
[XLSX File (Microsoft Excel File), 42 KB - mhealth_v13i1e52206_app6.xlsx]

References

1. Haidar S, Vazquez R, Medic G. Impact of surgical complications on hospital costs and revenues: retrospective database study of Medicare claims. J Comp Eff Res 2023 Jul;12(7):e230080 [FREE Full text] [doi: 10.57264/cer-2023-0080] [Medline: 37350467]
2. Patel AS, Bergman A, Moore BW, Haglund U. The economic burden of complications occurring in major surgical procedures: a systematic review. Appl Health Econ Health Policy 2013 Dec 29;11(6):577-592. [doi: 10.1007/s40258-013-0060-y] [Medline: 24166193]
3. Costabella F, Patel KB, Adepoju AV, Singh P, Attia Hussein Mahmoud H, Zafar A, et al. Healthcare cost and outcomes associated with surgical site infection and patient outcomes in low- and middle-income countries. Cureus 2023 Jul;15(7):e42493 [FREE Full text] [doi: 10.7759/cureus.42493] [Medline: 37637579]
4. Marwaha JS, Raza MM, Kvedar JC. The digital transformation of surgery. NPJ Digit Med 2023 May 31;6(1):103 [FREE Full text] [doi: 10.1038/s41746-023-00846-3] [Medline: 37258642]
5. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
6. Akhtar N, Khan N, Qayyum S, Qureshi MI, Hishan SS. Efficacy and pitfalls of digital technologies in healthcare services: a systematic review of two decades. Front Public Health 2022 Sep 16;10:869793 [FREE Full text] [doi: 10.3389/fpubh.2022.869793] [Medline: 36187628]
7. Riskin DJ, Longaker MT, Gertner M, Krummel TM. Innovation in surgery: a historical perspective. Ann Surg 2006 Nov;244(5):686-693. [doi: 10.1097/01.sla.0000242706.91771.ce] [Medline: 17060760]
8. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
9. Downey CL, Chapman S, Randell R, Brown JM, Jayne DG. The impact of continuous versus intermittent vital signs monitoring in hospitals: a systematic review and narrative synthesis. Int J Nurs Stud 2018 Aug;84:19-27 [FREE Full text] [doi: 10.1016/j.ijnurstu.2018.04.013] [Medline: 29729558]
10. Haveman ME, Jonker LT, Hermens HJ, Tabak M, de Vries JP. Effectiveness of current perioperative telemonitoring on postoperative outcome in patients undergoing major abdominal surgery: a systematic review of controlled trials. J Telemed Telecare 2021 Nov 01;30(2):215-229. [doi: 10.1177/1357633x211047710]
11. Amin T, Mobbs RJ, Mostafa N, Sy LW, Choy WJ. Wearable devices for patient monitoring in the early postoperative period: a literature review. Mhealth 2021;7:50 [FREE Full text] [doi: 10.21037/mhealth-20-131] [Medline: 34345627]

XSL•FO
RenderX

12. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
13. Gonçalves-Bradley DC, J Maria AR, Ricci-Cabello I, Villanueva G, Fønhus MS, Glenton C, et al. Mobile technologies to support healthcare provider to healthcare provider communication and management of care. Cochrane Database Syst Rev 2020 Aug 18;8(8):CD012927 [FREE Full text] [doi: 10.1002/14651858.CD012927.pub2] [Medline: 32813281]
14. Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. Allergy 2009 May;64(5):669-677. [doi: 10.1111/j.1398-9995.2009.01973.x] [Medline: 19210357]
15. Engel H, Huang JJ, Tsao CK, Lin CY, Chou PY, Brey EM, et al. Remote real-time monitoring of free flaps via smartphone photography and 3G wireless internet: a prospective study evidencing diagnostic accuracy. Microsurgery 2011 Nov 24;31(8):589-595. [doi: 10.1002/micr.20921] [Medline: 22072583]
16. Henarejos V, O'Connor K, Barrasa A, Villalonga A, Pastor C, Puyana JC, et al. Implementing a mHealth-based patient and nurse educational program to reduce wound infection in rural Philippines. Ann Glob Health 2022;88(1):76 [FREE Full text] [doi: 10.5334/aogh.3834] [Medline: 36118273]
17. McLean KA, Mountain KE, Shaw CA, Drake TM, Pius R, Knight SR, et al. Remote diagnosis of surgical-site infection using a mobile digital intervention: a randomised controlled trial in emergency surgery patients. NPJ Digit Med 2021 Nov 18;4(1):160 [FREE Full text] [doi: 10.1038/s41746-021-00526-0] [Medline: 34795398]
18. McGillicuddy JW, Weiland AK, Frenzel RM, Mueller M, Brunner-Jackson BM, Taber DJ, et al. Patient attitudes toward mobile phone-based health monitoring: questionnaire study among kidney transplant recipients. J Med Internet Res 2013 Jan 08;15(1):e6 [FREE Full text] [doi: 10.2196/jmir.2284] [Medline: 23305649]
19. Palombo D, Mugnai D, Mambrini S, Robaldo A, Rousas N, Mazzei R, et al. Role of interactive home telemedicine for early and protected discharge 1 day after carotid endarterectomy. Ann Vasc Surg 2009;23(1):76-80. [doi: 10.1016/j.avsg.2008.06.013] [Medline: 18809294]
20. Kayaalp ME, Agres AN, Reichmann J, Bashkuev M, Duda GN, Becker R. Validation of a novel device for the knee monitoring of orthopaedic patients. Sensors (Basel) 2019 Nov 27;19(23):5193 [FREE Full text] [doi: 10.3390/s19235193] [Medline: 31783551]
21. Hawkins SD, Koch SB, Williford PM, Feldman SR, Pearce DJ. Web app- and text message-based patient education in Mohs micrographic surgery-a randomized controlled trial. Dermatol Surg 2018 Jul;44(7):924-932. [doi: 10.1097/DSS.0000000000001489] [Medline: 29406486]
22. Schramm K, Lapert F, Nees J, Lempersz C, Oei SG, Haun MW, et al. Acceptance of a new non-invasive fetal monitoring system and attitude for telemedicine approaches in obstetrics: a case-control study. Arch Gynecol Obstet 2018 Dec;298(6):1085-1093. [doi: 10.1007/s00404-018-4918-y] [Medline: 30264201]
23. Cairns AE, Tucker KL, Leeson P, Mackillop LH, Santos M, Velardo C, et al. Self-management of postnatal hypertension: the SNAP-HT trial. Hypertension 2018 Aug;72(2):425-432. [doi: 10.1161/HYPERTENSIONAHA.118.10911] [Medline: 29967037]
24. Segura-Sampedro JJ, Rivero-Belenchón I, Pino-Díaz V, Rodríguez Sánchez MC, Pareja-Ciuró F, Padillo-Ruiz J, et al. Feasibility and safety of surgical wound remote follow-up by smart phone in appendectomy: a pilot study. Ann Med Surg (Lond) 2017 Sep;21:58-62 [FREE Full text] [doi: 10.1016/j.amsu.2017.07.040] [Medline: 28794868]
25. Khoshrounejad F, Tabesh H, Aalaei S, Khoshrounejad S, Moradi A, Eslami S. Effect of a text message-based support program on outcomes of patients after flexor tendon injury repair. J Hand Surg Am 2022 Nov;47(11):1085-1094. [doi: 10.1016/j.jhsa.2022.07.012] [Medline: 36064509]
26. Chiang CY, Chen KH, Liu KC, Hsu SJ, Chan CT. Data collection and analysis using wearable sensors for monitoring knee range of motion after total knee arthroplasty. Sensors (Basel) 2017 Feb 22;17(2):418 [FREE Full text] [doi: 10.3390/s17020418] [Medline: 28241434]
27. Sun V, Dumitra S, Ruel N, Lee B, Melstrom L, Melstrom K, et al. Wireless monitoring program of patient-centered outcomes and recovery before and after major abdominal cancer surgery. JAMA Surg 2017 Sep 01;152(9):852-859 [FREE Full text] [doi: 10.1001/jamasurg.2017.1519] [Medline: 28593266]
28. Chee SN, Lowe P, Lim A. Smartphone patient monitoring post-laser resurfacing. Australas J Dermatol 2017 Nov 10;58(4):e216-e222. [doi: 10.1111/ajd.12507] [Medline: 27282697]
29. Toogood PA, Abdel MP, Spear JA, Cook SM, Cook DJ, Taunton MJ. The monitoring of activity at home after total hip arthroplasty. Bone Joint J 2016 Nov;98-B(11):1450-1454. [doi: 10.1302/0301-620x.98b11.bjj-2016-0194.r1]
30. Ertel AE, Kaiser TE, Abbott DE, Shah SA. Use of video-based education and tele-health home monitoring after liver transplantation: results of a novel pilot study. Surgery 2016 Oct;160(4):869-876. [doi: 10.1016/j.surg.2016.06.016] [Medline: 27499142]
31. Kummerow Broman K, Oyefule OO, Phillips SE, Baucom RB, Holzman MD, Sharp KW, et al. Postoperative care using a secure online patient portal: changing the (inter)face of general surgery. J Am Coll Surg 2015 Dec;221(6):1057-1066 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2015.08.429] [Medline: 26453260]

XSL•FO
RenderX

32. Calliess T, Bocklage R, Karkosch R, Marschollek M, Windhagen H, Schulze M. Clinical evaluation of a mobile sensor-based gait analysis method for outcome measurement after knee arthroplasty. Sensors (Basel) 2014 Aug 28;14(9):15953-15964. [doi: 10.3390/s140915953] [Medline: 25171119]
33. Gunter R, Fernandes-Taylor S, Mahnke A, Awoyinka L, Schroeder C, Wiseman J, et al. Evaluating patient usability of an image-based mobile health platform for postoperative wound monitoring. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e113 [FREE Full text] [doi: 10.2196/mhealth.6023] [Medline: 27683059]
34. McGillion MH, Parlow J, Borges FK, Marcucci M, Jacka M, Adili A, et al. Post-discharge after surgery Virtual Care with Remote Automated Monitoring-1 (PVC-RAM-1) technology versus standard care: randomised controlled trial. BMJ 2021 Sep 30;374:n2209 [FREE Full text] [doi: 10.1136/bmj.n2209] [Medline: 34593374]
35. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
36. Goode PS, Johnson TM2, Newman DK, Vaughan CP, Echt KV, Markland AD, et al. Perioperative mobile telehealth program for post-prostatectomy incontinence: a randomized clinical trial. J Urol 2022 Aug;208(2):379-387. [doi: 10.1097/JU.0000000000002697] [Medline: 35389239]
37. Rosner BI, Gottlieb M, Anderson WN. Effectiveness of an automated digital remote guidance and telemonitoring platform on costs, readmissions, and complications after hip and knee arthroplasties. J Arthroplasty 2018 Apr;33(4):988-96.e4 [FREE Full text] [doi: 10.1016/j.arth.2017.11.036] [Medline: 29229238]
38. Card E, Nelson D, Jeskey M, Miller A, Michaels D, Hardeman W, et al. Early detection of a blood transfusion reaction utilizing a wireless remote monitoring device. Medsurg Nurs 2012;21(5):299-302, 308. [Medline: 23243788]
39. Hee Hwang J, Mun GH. An evolution of communication in postoperative free flap monitoring: using a smartphone and mobile messenger application. Plast Reconstr Surg 2012 Jul;130(1):125-129. [doi: 10.1097/PRS.0b013e318254b202] [Medline: 22743879]
40. Cleeland CS, Wang XS, Shi Q, Mendoza TR, Wright SL, Berry MD, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. J Clin Oncol 2011 Mar 10;29(8):994-1000. [doi: 10.1200/jco.2010.29.8315]
41. Barber JM, Pringle CJ, Raffalli-Ebezant H, Pathmanaban O, Ramirez R, Kamaly-Asl ID. Telemetric intra-cranial pressure monitoring: clinical and financial considerations. Br J Neurosurg 2017 Jun 19;31(3):300-306. [doi: 10.1080/02688697.2016.1229752] [Medline: 27644335]
42. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372 [FREE Full text] [doi: 10.1136/bmj.n71]
43. Naik N, Talyshinskii A, Rassweiler J, Hameed BM, Somani BK. Digital health innovations in urology: telemedicine, wearables, and mobile applications - a systematic review of literature. Curr Opin Urol 2024 Mar 01;34(2):116-127. [doi: 10.1097/MOU.0000000000001153] [Medline: 38038411]
44. Kaplan RS, Porter ME. How to solve the cost crisis in health care. Harv Bus Rev 2011 Sep;89(9):46-52, 54, 56. [Medline: 21939127]
45. Nepogodiev D, Martin J, Biccard B, Makupe A, Bhangu A, Nepogodiev D, et al. Global burden of postoperative death. The Lancet 2019 Feb;393(10170):401. [doi: 10.1016/s0140-6736(18)33139-8]
46. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. N Engl J Med 2009 Oct;361(14):1368-1375. [doi: 10.1056/nejmsa0903048]
47. Migden M, Chavez-Frazier A, Nguyen T. The use of high definition video modules for delivery of informed consent and wound care education in the Mohs Surgery Unit. Semin Cutan Med Surg 2008 Mar;27(1):89-93. [doi: 10.1016/j.sder.2008.02.001] [Medline: 18486030]
48. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
49. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
50. Fritsch CG, Ferreira PH, Prior JL, McLachlan AJ, Ferreira ML. Effects of using text message interventions for the management of musculoskeletal pain: a systematic review. Pain 2020 Nov 15;161(11):2462-2475. [doi: 10.1097/j.pain.0000000000001958] [Medline: 32569088]
51. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
52. Chen HC, Chuang TY, Lin PC, Lin YK, Chuang YH. Effects of messages delivered by mobile phone on increasing compliance with shoulder exercises among patients with a frozen shoulder. J Nurs Scholarsh 2017 Jul 20;49(4):429-437. [doi: 10.1111/jnu.12308] [Medline: 28632975]
53. Murphy RXJ, Bain MA, Wasser TE, Wilson E, Okunski WJ. The reliability of digital imaging in the remote assessment of wounds: defining a standard. Ann Plast Surg 2006 Apr;56(4):431-436. [doi: 10.1097/01.sap.0000202146.92893.6a] [Medline: 16557079]

XSL•FO
RenderX

54. Halstead LS, Dang T, Elrod M, Convit RJ, Rosen MJ, Woods S. Teleassessment compared with live assessment of pressure ulcers in a wound clinic: a pilot study. Adv Skin Wound Care 2003;16(2):91-96. [doi: 10.1097/00129334-200303000-00010] [Medline: 12690232]
55. Kim HM, Lowery JC, Hamill JB, Wilkins EG. Accuracy of a web-based system for monitoring chronic wounds. Telemed J E Health 2003;9(2):129-140 [FREE Full text] [doi: 10.1089/153056203766437471] [Medline: 12855036]

Abbreviations

MeSH: Medical Subject Headings
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 26.08.23; peer-reviewed by J Petch, S Azadnajafabad; comments to author 19.10.23; revised version received 31.05.24; accepted 20.05.25; published 10.10.25.

Please cite as:
Sierra MC, To H, Song WJ
Use and Application of mHealth Technologies in Perioperative Surgical Care: Narrative Review
JMIR Mhealth Uhealth 2025;13:e52206
URL: https://mhealth.jmir.org/2025/1/e52206
doi:10.2196/52206
PMID:

©Maria Camila Sierra, Henry To, Wan Jun Song. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Obtaining Patient-Reported Outcome Data via a Home Patient Monitoring App: Development, Implementation, and Validation of Novel Interface Terminology

Lucia Sacchi[1], PhD; Giordano Lanzola[1], PhD; Silvana Quaglini[1], PhD; Nicole Veggiotti[1], MSc; Silvia Panzarasa[1], BSc; Valentina Tibollo[2], MSc; Matteo Terzaghi[2], MSc; Itske Fraterman[3,4], PhD; Savannah Glaser[3,5,6], MSc; Manuel Ottaviano[7], PhD; Vadzim Khadakou[8], BA; Vitali Hisko[8], BA; Mor Peleg[9], PhD; Sofie Wilgenhof[3], MD; Henk Mallo[3], MSc; Alexandra Kogan[9], MSc; David Glasspool[10], PhD; Stephanie Medlock[5,11], PhD; Laura Del Campo[12], PhD; Matteo Gabetta[13], PhD; Mimma Rizzo[14], MD; Laura Deborah Locati[15,16], MD; Paola Gabanelli[17], MSc; Sara Demurtas[16], MD; Andrea Premoli[16], MD; Szymon Wilk[18], PhD

[1]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 5, Pavia, Italy

[10]Deontics Ltd, London, United Kingdom

[11]Amsterdam Public Health, Methodology & Aging & Later Life, Amsterdam, The Netherlands

[12]Associazione Italiana Malati di Cancro, Parenti e Amici (AIMAC), Roma, Italy

[13]Biomeris Srl, Pavia, Italy

[14]Medical Oncology Unit, Azienda Ospedaliera Universitaria Consorziale Policlinico di Bari, Bari, Italy

[15]Department of Internal Medicine and Medical Therapeutics, University of Pavia, Pavia, Italy

[16]Medical Oncology Unit, IRCCS Istituti Clinici Maugeri, Pavia, Italy

[17]Psychology Unit, IRCCS Istituti Clinici Maugeri, Pavia, Italy

[18]Institute of Computing Science, Poznan University of Technology, Poznan, Poland

[2]Medical Informatics and Artificial Intelligence Laboratory (LIM-IA), IRCCS Istituti Clinici Maugeri, Pavia, Italy

[3]The Netherlands Cancer Institute, Amsterdam, The Netherlands

[4]Department of Medical Oncology, University Medical Center Groningen, Groningen, The Netherlands

[5]Amsterdam UMC - University of Amsterdam, Medical Informatics, Amsterdam, The Netherlands

[6]Amsterdam Public Health, Methodology & Digital Health, Amsterdam, The Netherlands

[7]Life Supporting Technologies, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

[8]Bitsens JSC, Vilnius, Lithuania

[9]Department of Information Systems, University of Haifa, Haifa, Israel

Corresponding Author:
Lucia Sacchi, PhD
Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 5, Pavia, Italy

Abstract

Background: Adverse events (AEs) related to cancer treatment represent a valuable source of information that can be used to adjust therapy for individual patients. The National Institutes of Health developed the Common Terminology Criteria for Adverse Events (CTCAE), a comprehensive standardized terminology for health care providers to consistently report AEs during patient visits. Mobile health technologies, in principle, also allow AEs to be self-reported by patients in between visits; however, the terminology poses challenges for them, both in selecting the correct symptom to report and in rating its severity. The National Institutes of Health developed the Patient-Reported Outcomes–CTACE as the patient-oriented *companion* of the CTCAE. However, it shows some weaknesses in completeness and precision when used for continuous home patient monitoring and for decision support.

Objective: The aim of this paper is to propose a new terminology for reporting AEs that is easy for patients to use while also being clinically meaningful for health care providers and easily exploitable by decision support systems. Moreover, we aim to demonstrate its implementation and validation within the CAPABLE (Cancer Patients Better Life Experience) EU project.

Methods: The development of the new terminology starts from the CTCAE, which includes a comprehensive list of signs and symptoms along with guidance for accurately grading their severity. Through a multistep, participatory approach involving both patients and health care providers, we reduced and adapted the AE list for patient-oriented apps. During the CAPABLE project,

XSL•FO
RenderX

the proposed terminology was integrated into a mobile app and evaluated within a clinical pilot study involving 86 patients who were monitored through the app for at least 6 months, and a control cohort of 133 patients monitored using standard care practices.

Results: The final terminology includes 124 AEs, 49 expressed as "present or absent," and 77 associated with 4 description levels. A mapping between the description levels and the original CTCAE grades enables running the decision support system embedded in the CAPABLE app. The pilot study demonstrated that the majority of the patients used the symptoms reporting functionality, sharing also 24 unique AEs that are not present in the Patient-Reported Outcomes–CTCAE. Symptoms reported using the proposed terminology allowed the enactment of the clinical practice guidelines included in the CAPABLE decision support tool, triggering 11 distinct recommendations.

Conclusions: The results obtained from the clinical study support our claim regarding the need for a novel terminology for the self-reporting of AEs, characterized by ease of use, completeness, and clinical meaningfulness. Finally, by mapping our terminology to the CTCAE, we demonstrated that it is possible to exploit self-reported data to trigger decision support rules consistent with clinical practice guidelines.

(JMIR Mhealth Uhealth 2025;13:e65504) doi:10.2196/65504

KEYWORDS

adverse effects; cancer; clinical decision support system; decision-making; home patient; mHealth; mobile app; patient-reported; patient-reported outcome; patient with cancer; quality of life; self-reporting; symptom assessment

Introduction

During patients' follow-up, a complete and accurate recording of symptoms helps health care providers to make correct diagnoses and improve their decision-making. However, the collection of symptoms is almost exclusively done during patient visits. This makes it less accurate and complete, due to recall bias on the patient's side, and underreporting on the clinicians' side, caused by short visit duration, failure to ask questions systematically, and attribution bias [1,2]. Patient-reported outcomes (PROs) can mitigate these issues. They are defined as "'any report of the status of a patient's health condition that comes directly from the patient without interpretation of the patient's response by a clinician or anyone else." This definition comes with the claim by the Food and Drug Administration of their usefulness in assessing drug efficacy in clinical trials [3]. More recently, another regulatory body, the European Medicines Agency, launched the "Regulatory Science Strategy to 2025," one of the key goals of which is "advancing patient-centered access to medicines in partnership with healthcare systems." This includes integrating PROs and patients' preferences with additional data collected during the patient's clinical path to evaluate the drug risk profile [4]. Therefore, while the collection of PROs also presents challenges [5], it holds significant promise, and not only in the drug surveillance area [6].

PROs gained additional momentum with the widespread adoption of mobile health (mHealth) apps. As a matter of fact, several implementations of the so-called electronic patient-reported outcomes measures (ePROMs) have been developed [7]. The benefits of ePROMs include improved symptom control, physical function, quality of life (QoL), adherence to treatment, reduction in hospital admissions, improved survival, and communication between patients and health care providers [8-15].

Which terminologies and data collection approaches are suitable for ePROMs depend on the purpose of their use. In one of their seminal papers, Basch et al [16] describe a web application designed for patients with gynecological malignancies, who could access it on the day of their appointment using computer kiosks in hospital waiting rooms or their personal devices. Symptoms were collected through a questionnaire with 7 items adapted from the National Cancer Institute's (NCI) Common Terminology Criteria for Adverse Events (CTCAE) [17] plus a performance status item. Free text was available for entering additional information. Since then, several systems have been developed. PatientViewpoint is a web-based system, relying on questionnaires that patients fill out at regular time intervals upon system reminders [18]. Questionnaires, which can be selected according to patients' condition, were built on the PROMIS and Functional Assessment of Chronic Illness Therapy system [19]. STAR (Symptom Tracking and Reporting) is a web-based interface for patients with a high symptom burden, whose effectiveness was demonstrated in a randomized clinical trial on patients with advanced solid tumors [15]. It includes questions from the CTCAE, adapted for patient use, and pertaining to 12 common symptoms experienced during chemotherapy, graded on a 5-point scale (from 0 to 4) of increasing disability. Basch et al [20] describe PRO-Core, which includes a survey with questions from the NCI's Patient-Reported Outcomes–Common Terminology Criteria for Adverse Events (PRO-CTCAE) validated item library [21], selected based on prevalence across different types of advanced cancers, as well as questions about oral drug intake and performance status. Velikova et al [22] describe eRAPID, a system relying on internet-based questionnaires and managing the connection with the hospital electronic health record. The questionnaire items were identified according to different cancer types, patient interviews, and systematic reviews of PRO reporting for those groups.

A common feature of the systems described above is their use of questionnaires that reference predefined time periods (eg, the past week) for data collection. Moreover, the questionnaires contain only a subset of all the possible adverse events (AEs), tailored to the addressed population and context.

As a matter of fact, the real-time, longitudinal collection of symptom data directly from patients, up to the resolution of their condition, is inherently complex [23,24]. One of the

XSL•FO
RenderX

important challenges is precisely the definition of a suitable terminology to be used at the moment of symptom reporting. To address this challenge, in this paper, we aim at developing a terminology that, differently from the already available literature, is (1) general enough to be suitable for different patient populations; (2) appropriate for monitoring home patients, that is, patients receiving treatment at home, possibly having a home caregiver, and potentially undergoing extended follow-up periods; and (3) exploitable by a system that provides real-time support for AE management.

The terminology that we describe was developed as part of the European Horizon 2020 project CAPABLE (Cancer Patients Better Life Experience). Despite CAPABLE focusing on home patients with cancer, aiming to improve their QoL, the produced terminology is broadly applicable across patient populations.

The CAPABLE system consists of 2 front-end components and a back-end component [19]. The 2 front-end components are a mobile app for patients and a web-based dashboard for the health care providers.

The app supports patients in managing their condition. It includes several functionalities (educational material, mental well-being activities, suggestions for physical exercise, etc), but since a major issue for oncological patients is the occurrence of treatment-related AEs, its main goal is to provide easy and accurate reporting of AEs, in terms of signs and symptoms that patients may experience. It includes a symptom follow-up strategy to enhance monitoring compliance. Upon patients' consent, the app may also be used by home caregivers. This allows reporting those symptoms that, in certain conditions, a patient would not be able to report, for example, delirium. The dashboard is a web application through which health care providers can monitor patients' data remotely.

The back-end component [25] implements the CAPABLE decision support system (DSS) [26], which includes (1) a *virtual coach*, aimed at providing feedback to patients and their home caregivers through the app, and (2) a *physician DSS*, implemented as a set of computer-interpretable clinical guidelines providing evidence-based recommendations for doctors about the management of the reported AEs [27-31].

The objective of this paper is to introduce the AE-reporting terminology developed in the CAPABLE project. As mentioned, different from the terminologies already presented in the literature, we herein propose a highly flexible ePROs collection system, which allows patients to choose from an exhaustive list of signs and symptoms, enter data at any time, and manage the follow-up of the reported symptoms according to clinical practice guidelines. Finally, we present the evaluation of the terminology conducted as part of a clinical study involving patients who either used the app or received traditional follow-up care.

Methods

Overview

The process of defining a novel terminology began with an analysis of the 2 main existing AE terminologies, CTCAE and PRO-CTCAE, to identify their respective strengths and weaknesses. Insights from this analysis informed the strategy for creating the new terminology. In this section, we describe the whole process in detail.

Analysis of CTCAE and PRO-CTCAE

The CTCAE version 5 is a terminology developed by the NCI that allows standardized reporting of abnormal clinical findings, along with their grade of severity, which may range from grade 1 to grade 5. As shown in Table 1, grading may be based on the impact of the symptom on activities of daily living (ADLs, first row), or it may consider (also) objective descriptions of the symptom that represent increasing severity (eg, second row, where objectivity is based on the increased number of stools/day).

XSL•FO
RenderX

Table . Examples of the Common Terminology Criteria for Adverse Events (CTCAE) terminology, showing name, definition, and grading of 3 adverse events (source: National Cancer Institute [32]).

CTCAE term	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Ear pain[a]	Mild pain	Moderate pain; limiting instrumental ADLs[b]	Severe pain; limiting self-care ADLs	—[c]	—
Diarrhea[d]	Increase of <4 stools per day over baseline; mild increase in ostomy output compared to baseline	Increase of 4 - 6 stools per day over baseline; moderate increase in ostomy output compared to baseline	Increase of ≥7 stools per day over baseline; incontinence; hospitalization indicated; severe increase in ostomy output compared to baseline; limiting self-care ADLs	Life-threatening consequences; urgent intervention indicated	Death
Anemia[e]	Hemoglobin 10 g/dL—LLN[f]	Hemoglobin 8 - 10 g/dL	Hemoglobin <8 g/dL; transfusion indicated	Life-threatening consequences; urgent intervention indicated	Death

[a]A disorder characterized by a sensation of marked discomfort in the ear.

[b]ADL: activity of daily living.

[c]Not available (CTCAE defines the severity for this adverse event only in the range 1-3).

[d]A disorder characterized by frequent and watery bowel movements.

[e]A disorder characterized by a reduction in the amount of hemoglobin in 100 mL of blood. Signs and symptoms of anemia may include pallor of the skin and mucous membranes, shortness of breath, palpitations of the heart, soft systolic murmurs, lethargy, and fatigability.

[f]LLN: lower limit normal.

CTCAE is a de facto standard adopted by several scientific societies and in particular by the European Society of Medical Oncology, whose guidelines have been implemented in the CAPABLE DSS [27-31]. For example, the guideline on immune-related AEs [31] explicitly refers to CTCAE grades in its statements about the discontinuation of immune checkpoint inhibitor (ICI) therapy in case of immune-related neuropathy:

> *(1) for grade 1 symptoms, ICI treatment can be continued and the patient is monitored for deterioration; (2) for grade 2 symptoms, ICI treatment should be interrupted and oral or i.v. (methyl)prednisolone is initiated; and (3) For grade 3 and 4 symptoms, more intensive immune modulation may be required in addition to corticosteroids [...]*

The CTCAE includes more than 800 AEs, and it was designed to be used by health care providers; thus, it cannot be directly exploited as an interface terminology for a patient's app. One main reason is that several AEs are in fact medical diagnoses that patients or home caregivers cannot detect on their own because they require laboratory tests (see, eg, footnote "d" of Table 1 where anemia is defined).

Another challenge is the reliance on medically oriented technical language. For example, rash is described as "Macules/papules covering x% of the body surface…", where x determines the severity grade. Such terminology is clearly difficult for lay users to interpret.

The NCI also developed the PRO-CTCAE Measurement System, which is intended as the CTCAE companion terminology for patients [21]. As shown in Textbox 1, PRO-CTCAE allows reporting AEs either as present or absent or by specifying one or more of the following attributes: *frequency*, *severity*, *interference with daily activities*, and *amount*.

XSL•FO
RenderX

Textbox 1. Patient-Reported Outcomes–Common Terminology Criteria for Adverse Events approach to adverse event reporting.

Frequency

- In the last 7 days, how often did you have…?
  - Never
  - Rarely
  - Occasionally
  - Frequently
  - Almost constantly

Severity

- In the last 7 days, what was the severity of your…at its worst?
  - None
  - Mild
  - Moderate
  - Severe
  - Very severe

Interference

- In the last 7 days, how much did…interfere with your usual or daily activities?
  - Not at all
  - A little bit
  - Somewhat
  - Quite a bit
  - Very much

Amount

- In the last 7 days, did you have any…?
  - Not at all
  - A little bit
  - Somewhat
  - Quite a bit
  - Very much

Present or absent

- In the last 7 days, did you have any…?
  - No
  - Yes

PRO-CTCAE has been employed in several studies to collect AEs in patients with cancer and has been shown to enhance the quality of symptom reporting compared with standard reporting during routine visits [33]. It has also been associated with increased patient empowerment and improved QoL [34].

However, the PRO-CTCAE alone was not sufficient for our needs for the following reasons:

1. It includes only 80 AEs, while there are many more symptoms or signs included in the CTCAE, which are potentially detectable by patients or their home caregivers, as outlined by our analyses conducted with health care providers and patients.
2. The patient is asked to report on symptoms experienced *in the past 7 days*. First, this is not suitable for symptoms that need to be managed as soon as they arise, without having to wait for the next reporting window. Second, this might

XSL•FO
RenderX

force patients to fill in extra daily diaries to avoid forgetting some short-duration symptoms after a couple of days [34].
3. Symptom severity is only subjectively assessed, using a Likert scale measuring the impact on daily life activities, while it would be more appropriate to use (also) objective features, which could help health care providers to better understand the patient's condition.
4. As a consequence of (3), some of the PRO-CTCAE entries were hardly mappable to the CTCAE, even if efforts in this sense are ongoing [35]. Thus, in a system where patients enter data using the PRO-CTCAE and health care providers enter data using the CTCAE, there is an issue for data integration and their use in statistical analyses or DSSs (as mentioned, clinical practice guidelines refer to the CTCAE in their recommendations).
5. Different from the CTCAE, the PRO-CTCAE does not include any suggestions for symptom management. For example, in Table 1, we can see "hospitalization indicated" for Grade 3 diarrhea, which could be translated for a patient as "go to the emergency room" or "contact your physician."

In sum, the CTCAE, being developed for health care providers, is not appropriate for direct use by patients, whereas the PRO-CTCAE, despite being a validated and appropriate tool for a qualitative assessment of a patient's condition, has some drawbacks if included in a patient monitoring system that requires a complete, quantitative, precise, and timely assessment.

Development of the Novel CAPABLE Terminology

To overcome the mentioned limitations of the existing terminologies, we decided to start from the CTCAE and review it in depth. Figure 1 shows the high-level process schema. Oncologists and psychologists worked together to filter the CTCAE AEs list, based on the estimated patients' ability to recognize symptoms and on other features illustrated below, while psychologists and patients collaborated to replace complex terms with more accessible ones. When suitable, PRO-CTCAE terms were adopted.

More precisely, we went through the full list of AEs and performed the following steps:

1. Remove AEs that are not detectable by a patient or his or her home caregiver (eg, anemia).
2. Manage AEs, such as anxiety, depression, and sexual life, with dedicated and validated questionnaires, as suggested by our medical experts based on the observation that CTCAE terms for those conditions are not sufficiently detailed for monitoring purposes. These AEs are thus excluded from the terminology.
3. Remove AEs related to some vital signs, for example, weight gain or loss, or hypertension, which can be easily inferred from the multiple measures of body weight and blood pressure that patients periodically report in the app.
4. Remove AEs that are specializations of other AEs already informative enough for the oncologists (eg, hoarseness can be removed as it is covered by the general term voice change).

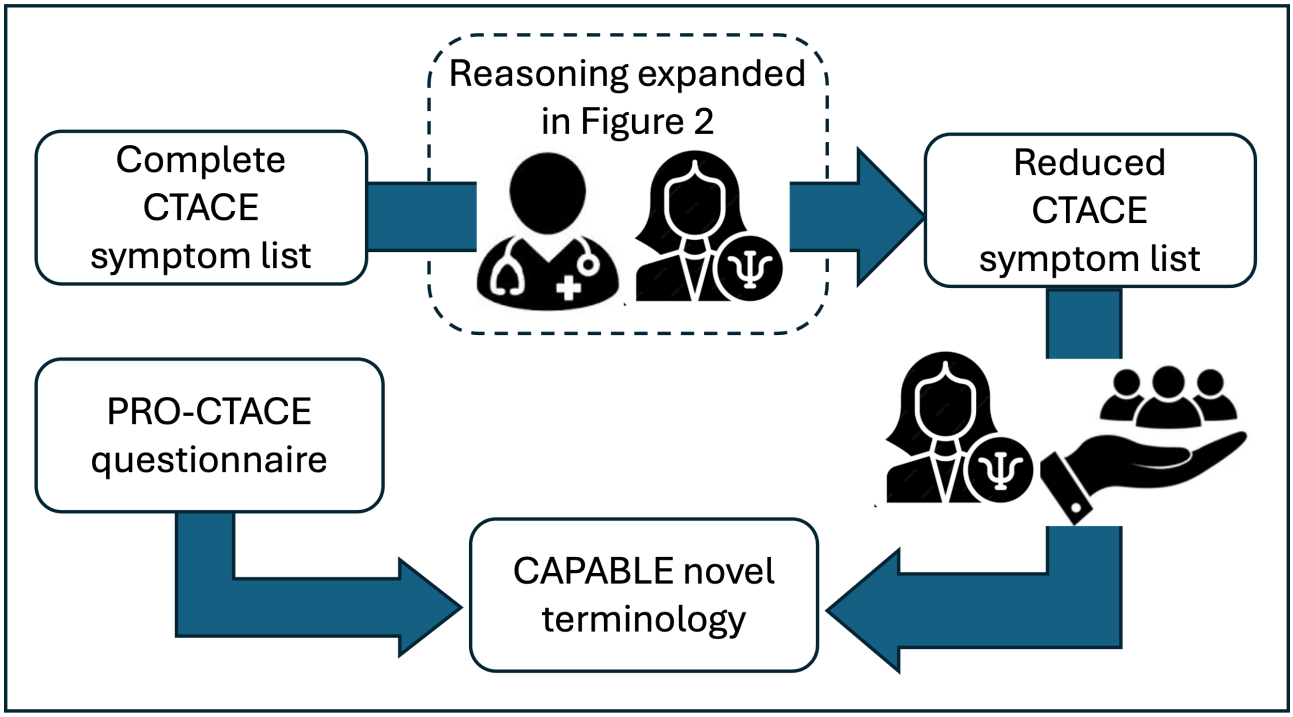
Once the reduced list was obtained, we performed the following additional steps:

1. Remove grade descriptions referring to death and life-threatening situations that are outside the CAPABLE scope and could be alarming for patients. This step was suggested by the psycho-oncologists working on the project.
2. Remove from the grade descriptions any statements referring to medical interventions, as these serve as guidance for clinicians rather than for patients.
3. Rephrase the description of AE names and grades using words that are easily understandable for patients. When possible, we exploited the PRO-CTCAE description. For some AEs, a CTCAE grade was split into 2 levels to allow patients to better describe their situation. For example, in the CTCAE, cough is described by 3 grades: mild, moderate, and severe symptoms. In our terminology, we provide 4 possible descriptions, splitting the *severe* level into (1) severe cough that interferes quite a bit with my usual or daily activities and (2) very severe cough that interferes with my daily self-care activities.
4. Employ a principle of the lowest risk of underestimation: if the distinction between grade descriptions of a symptom might not be clear for a patient, the symptom will be mapped to a higher severity level. For example, the CTCAE assesses the severity of pruritus based on whether the symptom is localized or generalized. This may be difficult for a patient to evaluate, so it was omitted from the descriptions, which measure the impact on ADLs instead. Thus, the report of a highly impacting itching would be mapped to CTCAE grade 2, instead of grade 1, even if localized. A physician can then assess the symptom and assign the correct CTCAE grade.

The last 2 points were carried out in collaboration with patients belonging to the Associazione Italiana Malati di Cancro patients' association (in Italy) and 2 expert nurses (in the Netherlands) to ensure the understandability of the new proposed phrasing. The terminology was developed in English, Italian, and Dutch, using the localized version of the available CTCAE and PRO-CTCAE terminologies as a basis.

XSL•FO
RenderX

Figure 1. Overall framework of the process used to develop the new terminology. CAPABLE: Cancer Patients Better Life Experience; CTACE: Common Terminology Criteria for Adverse Events; PRO-CTACE: Patient-Reported Outcome–CTACE.



Evaluation of the Proposed Terminology

CAPABLE was tested within a clinical pilot study that took place in 3 clinical centers, 2 in Italy and 1 in the Netherlands, from May 2023 to January 2024 (ClinicalTrials.gov NCT06161233 and NCT05827289) [36]. The system was used by 86 patients overall. In total, 2 main cancer types were addressed, metastatic renal cell carcinoma (mRCC; 30 patients: 20 males and 10 females—with a median age at enrollment of 62 y) and melanoma (31 patients: 12 males and 19 females—with a median age at enrollment of 65 y; one of these patients was lost at follow-up), and breast, head and neck, lung, stomach, ovary, and thyroid cancer were considered (26 patients overall: 5 males and 21 females, with a median age at enrollment of 50.4 y). An additional 133 patients were enrolled from May 2021 to March 2023 into a control cohort that did not use the CAPABLE app. The control cohort consisted of 77 patients with mRCC, 56 males and 21 females, with a median age at enrollment of 60 years, and 56 patients with melanoma, 22 males and 34 females, with a median age of 64 years.

Ethical Considerations

The studies were approved by the Ethical Committees of the clinical centers as follows: EC #2546 and EC #2741 ICS Maugeri—Pavia, EC #7558 Azienda Ospedaliero-Universitaria "Consorziale Policlinio"—Bari, and EC #22 - 981/NL81970.000.22, NedMec EC Amsterdam. Informed consent was obtained from all the patients participating in the studies. All patients also signed a data-processing consent form, and a Data Protection Impact Assessment was prepared and approved by each clinical center. The data used for the analyses presented in this paper were deidentified before being processed by the authors responsible for the analyses, who signed a data-processing agreement with the clinical centers. No compensation was provided to the study participants.

In addition to the clinical study, which allowed the collection of clinical data for a quantitative assessment of the intervention efficacy (the primary outcome of CAPABLE was QoL), a focus group was held near the conclusion of the study itself to gather qualitative data from the user experience. Moreover, at the end of the study, questionnaires on the ease and usefulness of the system were collected from patients and health care professionals. These questionnaires included 1 item for each functionality of the CAPABLE systems, including the symptom reporting, to help understand how patients and health care providers perceived the functionality, also in comparison with the others available.

Results

The Developed CAPABLE Terminology

As a result of the process summarized in Figure 1, we obtained a list of 124 AEs. The detailed symptoms selection process is described and the results of each step are depicted in Figure 2. Forty-nine symptoms are expressed as "present or absent," and 75 are associated with up to 4 possible "description levels."

XSL•FO
RenderX

Figure 2. Symptoms selection process. The Common Terminology Criteria for Adverse Events (CTCAE) revision steps and the number of symptoms selected at each step. Overlapping terms mean specialized adverse events that were subsumed by the more general ones.

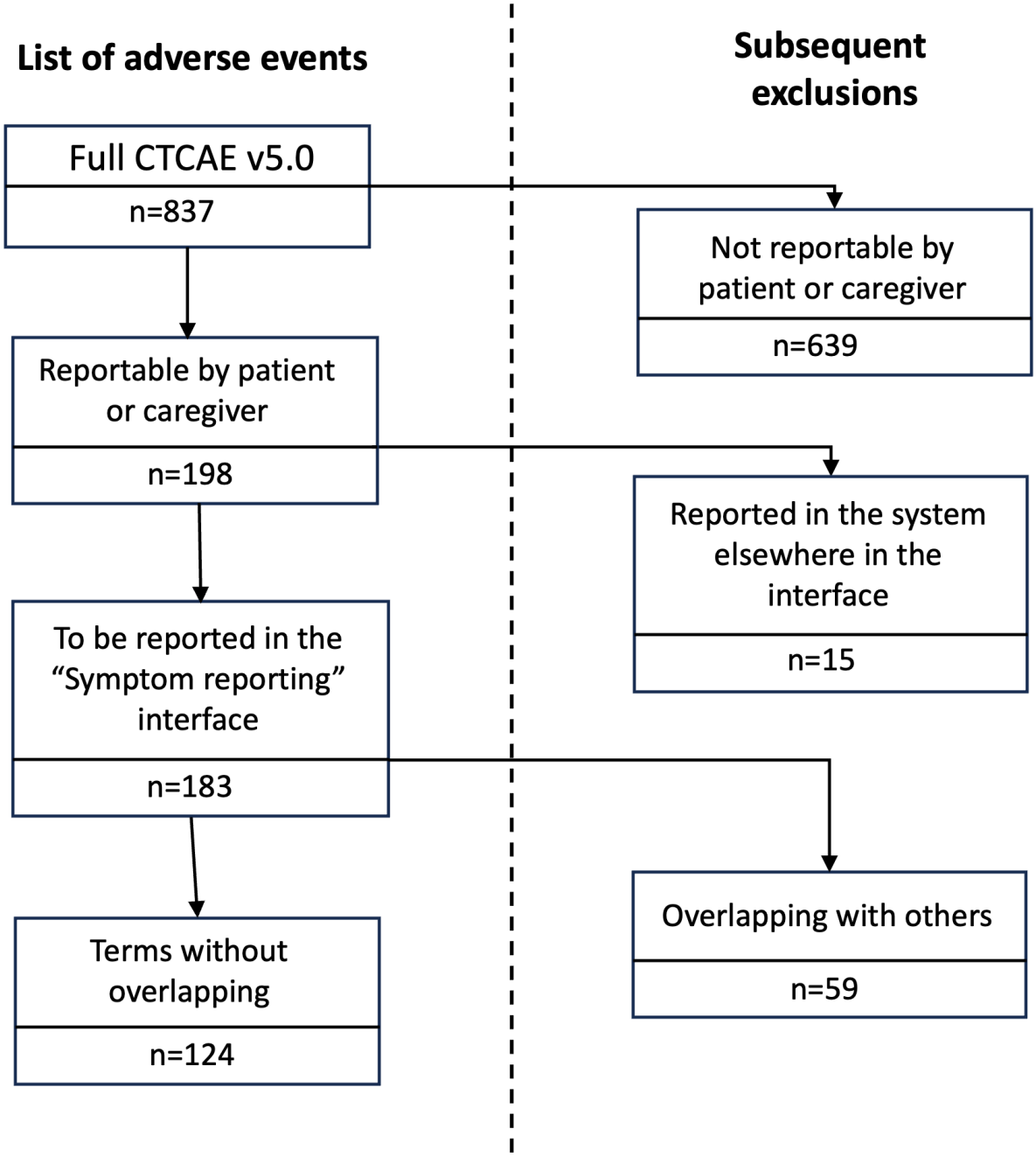


Table 2 provides an excerpt of the developed terminology, highlighting (1) the differences between the original CTCAE terms and their corresponding translations into lay language (column "CAPABLE term"), (2) the Systematized Nomenclature of Medicine code associated with each AE, (3) the levels that are shown to the patients in their app in order to grade the severity of the symptom (columns "Level 1" to "Level 4"), and (4) the mapping between the patient description levels and the original CTCAE grades (last column).

XSL•FO
RenderX

Table . A snapshot of the developed terminology.

CTCAE[a] term	CAPABLE[b] term	SNOMED[c] code	Level 1	Level 2	Level 3	Level 4	Map to CTCAE grade
Palmar-plantar erythrodysesthesia syndrome	Hand-foot syndrome (rash of the hands or feet that can cause cracking, peeling, redness, or pain)	403638003	Condition without pain that does not interfere or that interferes a little bit with my usual or daily activities (mild)	Condition with pain that interferes somewhat with my usual or daily activities (moderate)	Condition with pain that interferes quite a bit with my usual or daily activities (severe)	Condition with pain that interferes with my daily self-care activities (very severe)	• 1→1 • 2, 3 → 2 • 4 → 3
Skin hypopigmentation	Patches of skin that are lighter than your overall skin tone	23006000	Skin hypopigmentation with no psychosocial impact	Skin hypopigmentation with associated psychosocial impact	—[d]	—	• 1 → 1 • 2 → 2
Anosmia	Lost or changed sense of smell	44169009	Present	—	—	—	• 1 → 1
Anorexia	Decreased appetite	81492003	Decrease of appetite, but I eat or drink almost as usual (mild)	I cannot eat or drink as usual, but I have not lost weight (moderate)	I cannot eat or drink as usual, and I have lost weight (severe)	—	• 1 → 1 • 2 → 2 • 3 → 3
Diarrhea	Diarrhea (loose or watery stools)	62315008	Increase of <4 stools per day compared to usual amount of stools per day	Increase of 4 - 6 stools per day compared to usual amount of stools per day	Increase of ≥7 stools per day compared to usual amount of stools per day	—	• 1 → 1 • 2 → 2 • 3 → 3
Amenorrhea	Absence of menstrual periods for 3 consecutive menstrual cycles	14302001	Present	—	—	—	• 1 → 2

[a]CTCAE: Common Terminology Criteria for Adverse Events.

[b]CAPABLE: Cancer Patients Better Life Experience.

[c]SNOMED: Systematized Nomenclature of Medicine.

[d]Not applicable.

The mapping between the severity levels and the corresponding CTCAE grades is stored in the CAPABLE DSS and enables the conversion of the patient's selection into a CTCAE value, which is necessary for executing the guidelines. In addition, each CTCAE grade is mapped to the corresponding Systematized Nomenclature of Medicine–Clinical Terms code for interoperability purposes. The information regarding the symptoms and their severity is stored in the CAPABLE database, which is modeled according to the Observational Medical Outcomes Partnership common data model.

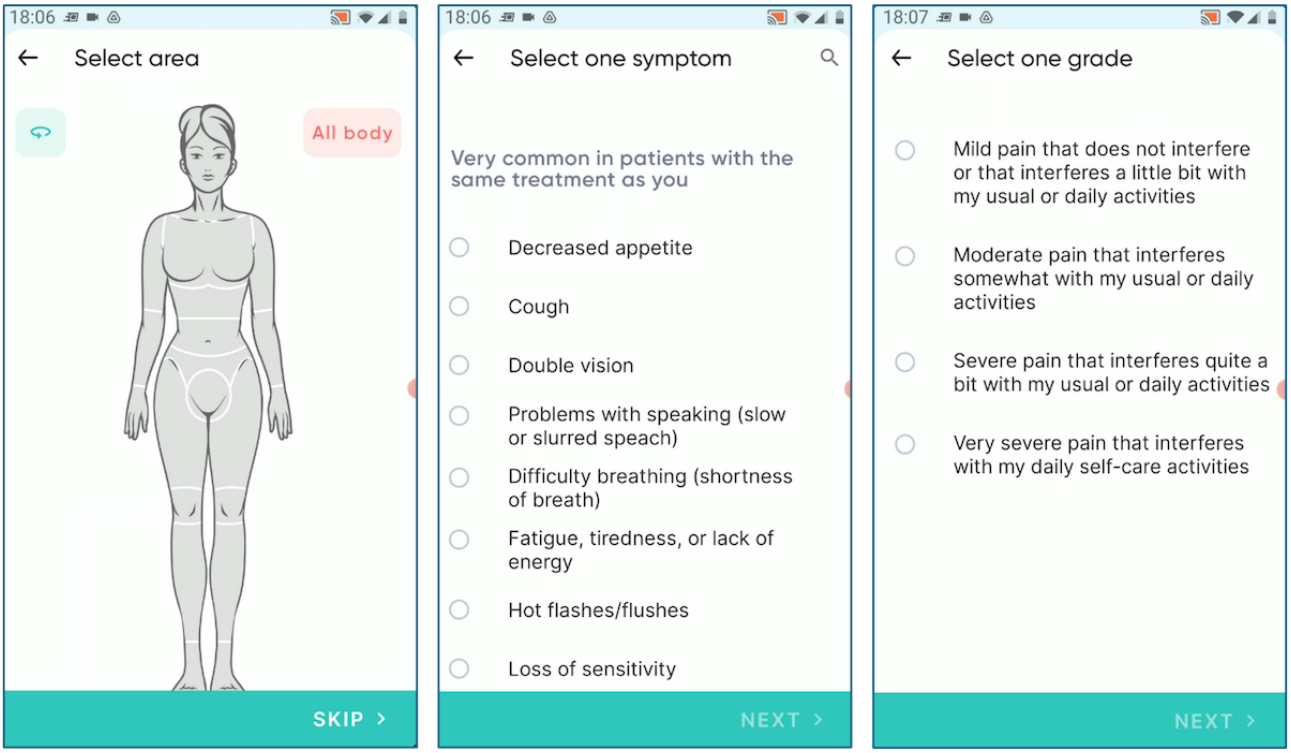
Implementation of the CAPABLE Terminology

Figure 3 depicts the integration of the proposed terminology within the user interface of the CAPABLE app for patients and their informal caregivers.

Within the physicians' dashboard, the CTCAE was implemented in its original form, without modifications, being a terminology specifically developed for clinical use (see the *Methods* section).

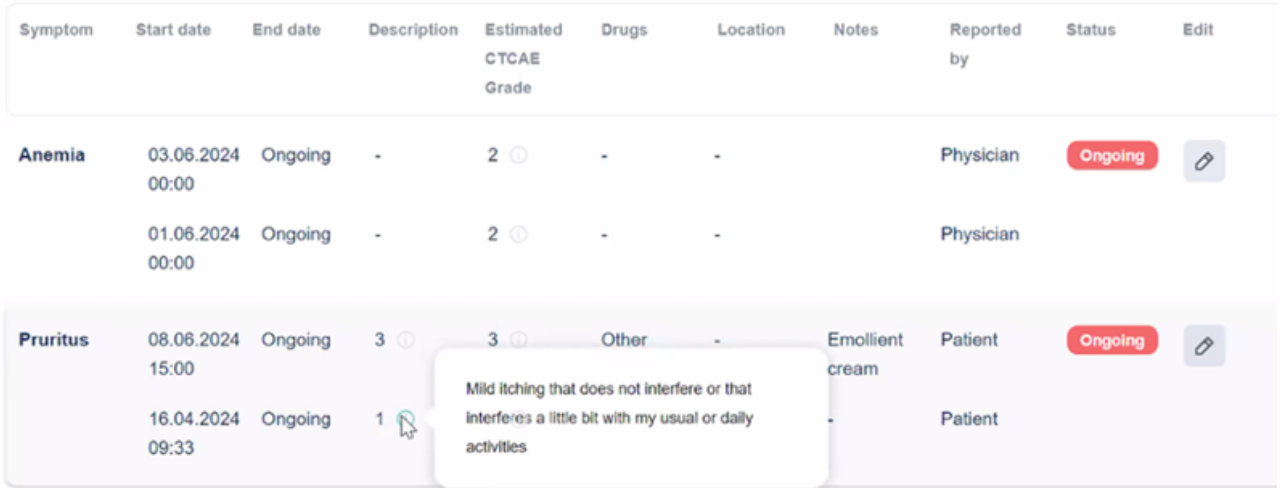
XSL•FO
RenderX

Figure 3. Symptom reporting in the CAPABLE (Cancer Patients Better Life Experience) patient app. From left to right: selecting the affected body part, list of symptoms (if a body part is selected, a filter is applied to the list), and description levels for the pain adverse event.



Patients eligible for home monitoring with the CAPABLE system are supposed to periodically go to the hospital for control visits. In between visits, doctors can monitor patients through the interface illustrated in Figure 4. As shown, they can see both the description level entered by the patient ("Description") and the corresponding CTCAE level ("Estimated CTCAE Grade"), in addition to the affected body part, if indicated by the patient ("Location"). During the visit, health care providers can update symptoms previously reported by patients (they also can change the level if needed) or add new ones. To this end, they have the full CTCAE terminology available in their graphical user interface. In Figure 4, pruritus has been entered by the patient, while anemia, which is a medical diagnosis, has been entered by the health care provider ("Reported By").

Figure 4. The CAPABLE (Cancer Patients Better Life Experience) physicians' dashboard. The graphical user interface dedicated to health care providers shows symptoms entered by patients, together with their severity level and the mapped Common Terminology Criteria for Adverse Events grade and those entered by doctors.



Evaluation Results

Use of the Symptom Reporting Functionality

During the evaluation, both qualitative and quantitative assessments were conducted. Among the qualitative evaluations, the focus group, held near the conclusion of the study, revealed that patients considered the app easy to use and effective and expressed disappointment about the service coming to an end [37,38].

After the end of the study, all patients were asked to score from 1 to 5 the level of ease of use of the key functionalities of the CAPABLE system, and 36 (80%) of the 45 patients who filled

XSL•FO
RenderX

in the usability questionnaire at the end of the study scored it very easy or easy to use. The level of usefulness, rated with the same scoring system, also received positive feedback from patients: noticeably, the highest rated functionality among those available in the app was the symptom reporting, as the patients perceived this collected information as key for the management of their disease during the treatment phase. A total of 11 health care professionals involved in the studies also assessed the level of ease of the patient app. The symptom reporting functionality was rated among the easiest and most useful tasks, together with the home functionalities and the introductory screens.

With respect to the quantitative results, the app was well exploited by the patients enrolled in the clinical pilot studies. In the Italian cohort, 51 (91%) of 56 patients entered at least 1 distinct symptom. In the Dutch cohort, this occurred for 18 (60%) of 30 patients, but it must be considered that 2 patients died because of rapidly progressive disease during the trial. The average number of new symptoms entered per patient each month is 2.1. Furthermore, after the initial entry, patients regularly provided updates on their symptoms.

To assess whether CAPABLE helped patients to report their symptoms, we analyzed data from patients with mRCC, as symptom information was also collected for the control cohort in this group. For the control patients, symptoms were collected only during control visits by health care providers. As illustrated in Table 3, the symptom variety was lower in the control group, likely due to recall bias and to physicians' tendency to overlook symptoms not perceived as related to cancer treatment. The higher incidence of fatigue in the control cohort can be explained by the fact that CAPABLE implemented a set of interventions, such as motivating patients to perform physical activity, able to mitigate that symptom.

Table . Symptoms reported in the 2 cohorts[a].

Adverse event	Occurrence, n
Control cohort	
Fatigue[b]	28
Mucositis	9
Palmar-plantar syndrome	7
Diarrhea	4
Nausea	3
CAPABLE[c] cohort	
Diarrhea	8
Muscle pain	5
Headache	5
Fever	5
Cough	5
Nausea	4
Fatigue	4
Mucositis	4
Palmar-plantar syndrome	3
Passing flatus	3
Backache	3

[a]Occurrences of symptoms entered by at least 3 patients were considered.

[b]Symptoms appearing in both cohorts are in italics.

[c]CAPABLE: Cancer Patients Better Life Experience.

Issues Overcome by the Novel Terminology

One of the issues that emerged from the initial analysis of the PRO-CTCAE terminology was that there were more AEs than those considered in the PRO-CTCAE that could be reported by patients or their caregivers. During the CAPABLE study, the patients were able to report 24 distinct symptoms that are not mentioned in the PRO-CTCAE, namely: *hand-foot syndrome, back pain, bone pain, burning sensation of skin, cramps, difficulty in walking, diplopia, dysarthria, edema of entire limb, neck pain, pain in limb, pain in the mouth, peripheral sensory neuropathy, red eye, thirst, malaise, influenza-like illness, toothache, hypersomnia, fever, edema of face, pain in face, tremor,* and *limitation of joint movement*. Note that PRO-CTCAE also includes "pain" but without the body site specification, which clinicians consider important. Thus, we believe that, also thanks to the friendly symptom descriptions and search strategy provided in the app, patients (possibly with their caregiver's help) were able to recognize and report a greater variety of AEs. The "Other" option for symptom reporting has

XSL•FO
RenderX

been used only 1 time, for reporting jaundice, which is currently not present either in the CTCAE or in the PRO-CTCAE. This suggests that patients managed most of the time to find the symptoms they were looking for, but also that such a term could be considered an additional term in an updated version of the terminology.

A second issue identified with the PRO-CTCAE was the absence of a direct mapping to CTCAE grades. This would prevent triggering those guidelines recommendations that consider CTCAE grades in their premise. To validate our terminology with respect to this issue, we counted how many different recommendations of that type have been fired during the pilot studies, and that would not be fired in case symptoms were reported in other ways. We found 11 distinct recommendations (related to diarrhea, fatigue, and skin problems) that physicians received, from one to several times, to treat 24 different patients. Without a proper mapping between the symptom reported by the patient and the CTCAE grade, it would not have been possible for the DSS to deliver such recommendations. To assess the level of satisfaction with the triggered recommendations, we collected impressions by 11 health care professionals about the ease and usefulness of the guideline access functionality, on a 1 - 5 Likert scale. Regarding ease of use, 7 (60%) of the participants judged the functionality easy to use, while 2 (20%) were neutral, with no participants giving negative feedback on the question. When asked about the usefulness, 4 (40%) of the health care professionals judged the functionality useful, 3 (30%) were neutral, while only 1 health care provider judged it not useful.

A third issue relates to questionnaires designed for administration at fixed intervals, such as on a weekly basis. For example, the PRO-CTCAE instructs: “For each question, please select the one response that best describes your experiences over the past 7 days.” This format is not suitable for a monitoring app, where symptoms need to be reported immediately upon occurrence. To support our statement, even if anecdotally, we illustrate the experience, within the CAPABLE study, of a patient who received early diagnosis of a life-threatening condition, thanks to this on-demand symptom reporting feature. In addition to cancer, the patient suffered from a psychiatric disorder for which he was being treated with medications. Through the dashboard, the oncologist noticed, for a few consecutive days, that the home caregiver reported dizziness and falls while walking. The caregiver annotated that this was probably due to the ongoing psychiatric treatment. However, for oncologists, those specific symptoms raised suspicion of brain metastases. Thus, the patient underwent an urgent brain computed tomography scan with contrast. Despite, unfortunately, the presence of metastasis being confirmed, the patient received a prompt diagnosis and started early treatment, also avoiding hospitalization that would have been necessary in case of late diagnosis.

Discussion

In this paper, we described the terminology adopted in the CAPABLE project for symptom reporting and its integration in graphical user interfaces. Both the terminology and the interfaces were designed by involving patients as the main actors of the requirement analysis, following an iterative codesign process including interviews where we were showing them, one after the other, the graphical user interface mockups [39]. As a result, although the CAPABLE system shares certain features with other systems described in the related work section, it also incorporates distinctive elements that position it beyond the current state of the art.

First, the CAPABLE does not propose a list of AEs that is filtered on the basis of the pathology and treatment. While this kind of filtering, used for example in [15,19,20,22], could represent a way to personalize the user interfaces, it could prevent entering AEs that are unexpected, for example, because the treatment has been introduced recently, and not enough data are available to exclude the possibility of new, still unknown AEs. Instead of shortening the AE list, to facilitate the AE selection for the patient, we use runtime strategies such as showing the most probable AEs at the top of the list (personalized according to the type of cancer and treatment), or filtering according to the body part selected through the graphical interface (Figure 2) [40]. This strategy makes our interface terminology generalizable and usable for different diseases.

Second, the CAPABLE allows a patient to enter symptoms as soon as they manifest, possibly even more frequently than once a day. This is different from the solutions currently available on the market [41] and in the literature [34,42-44] that usually require the patient to periodically fill out a questionnaire. In the CAPABLE paradigm, a patient just reports a symptom when it arises, choosing from different input modalities. This is useful when symptom management requires a timely intervention. On the other hand, if the patient does not experience any symptoms, he or she is not required to interact with the system. Questionnaires are also used in CAPABLE but just to collect data about psychological and nutritional status, with a frequency that is personalized according to the last entered data (the frequency of administration increases if critical values are reached in previous answers). The questionnaires were selected by the medical experts of the CAPABLE team to address the conditions that require more details than a single AE item might allow. The patient also receives a reminder in case any data are entered for more than a predefined period.

Finally, defining an extended interface terminology starting from the CTCAE can be beneficial for patients, giving them the potential to gradually learn to evaluate their own health condition more objectively. This learning process is also facilitated by the app, which provides timely and focused educational material at symptom entry: for each symptom, an info button provides an exhaustive description of the symptom itself.

Overall, the results show the efficacy of our design choices in collecting high-quality and timely data from home patients. However, not all patients used the app in the same way. The app was deployed in 2 countries with different patient populations and organizational settings. Italian patients showed greater engagement with the app compared to Dutch patients, likely due to the more consistent technical support provided in

XSL•FO
RenderX

Italy, where a dedicated bioengineering team was available. This difference was observed in overall app usage and was also reflected in the use of the symptom reporting feature. This highlights the importance of closely monitoring patients and promptly addressing their issues during the telemonitoring period.

As another limitation, it is important to point out that the proposed CAPABLE terminology is still lacking an external validation phase. As a matter of fact, the symptom reporting functionality was indirectly assessed through system usage during the CAPABLE clinical study, giving the encouraging results described above. On the other hand, we did not implement a specific post-deployment evaluation strategy specifically targeted to the developed terminology. As a future activity, a formal validation involving external stakeholders and other patient groups will be planned.

As a final consideration, we remark that it is hard to separate the development of a terminology from its intended use. The CAPABLE terminology is intended to be used by patients and their caregivers in mHealth apps dedicated to home patients' monitoring, and this has influenced some design choices. However, the obtained instrument is general enough to be exploited by different similar mHealth apps.

Funding

This work was supported by the EU H2020 program under GA #875052. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Data Availability

The CAPABLE (Cancer Patients Better Life Experience) terminology presented in this study is available as Multimedia Appendix 1. Italian and Dutch versions of the terminology are available from the authors upon request.

Authors' Contributions

Conceptualization: IF, LDL, LS, MR, NV, PG, SG, SP, SQ
Formal analysis: LS, SQ
Investigation: AP, HM, IF, LDL, MR, MT, S Wilgenhof, VT
Methodology: IF, LDL, LS, MR, NV, PG, SG, SP, SQ
Software: AK, DG, GL, MG, S Wilk, VH, VK
Supervision: LS, SQ
Validation: AP, LDC, MO, MP, SD, SM, S Wilk
Writing – original draft: LS, SP, SQ
Writing – review & editing: LS, SQ

Conflicts of Interest

LDL reports having the role of scientific consultant/advisory board participation with EISAI, MSD, Lilly, Novartis, Bayer, and Johnson & Johnson. MR reports advisory board participation with Merck Serono, MSD, AstraZeneca, Bristol Myers Squibb, Eisai, Gilead, and Astellas. All of these are unrelated to the present study. All other authors declare no conflicts of interest.

Multimedia Appendix 1
The CAPABLE (Cancer Patients Better Life Experience) terminology, including the 124 selected adverse events, with the corresponding Common Terminology Criteria for Adverse Events (CTCAE) terms, patient-oriented display of the term, description levels, and description of each level to the CTCAE grades.
[XLSX File, 95 KB - mhealth_v13i1e65504_app1.xlsx]

References

1. Fellowes D, Fallowfield LJ, Saunders CM, Houghton J. Tolerability of hormone therapies for breast cancer: how informative are documented symptom profiles in medical notes for "well-tolerated" treatments? Breast Cancer Res Treat 2001 Mar;66(1):73-81. [doi: 10.1023/a:1010684903199] [Medline: 11368413]
2. Redelmeier DA, Schull MJ, Hux JE, Tu JV, Ferris LE. Problems for clinical judgement: 1. Eliciting an insightful history of present illness. CMAJ 2001 Mar 6;164(5):647-651. [Medline: 11258213]
3. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006 Dec;4(1):79. [doi: 10.1186/1477-7525-4-79]
4. EMA regulatory science to 2025: strategic reflection. European Medicines Agency. 2020. URL: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf

XSL•FO
RenderX

5. Zini MLL, Banfi G. A narrative literature review of bias in collecting patient reported outcomes measures (PROMs). Int J Environ Res Public Health 2021 Nov 26;18(23):12445. [doi: 10.3390/ijerph182312445] [Medline: 34886170]
6. Puccini A, Viscardi G, Ciani O, et al. Patient-reported outcomes (PROs) in clinical trials and in clinical practice: report from the XXI national conference of the Italian Association of Medical Oncology (AIOM). BMJ Oncol 2025;4(1):e000783. [doi: 10.1136/bmjonc-2025-000783] [Medline: 40546927]
7. Di Maio M, Basch E, Denis F, et al. The role of patient-reported outcome measures in the continuum of cancer clinical care: ESMO Clinical Practice Guideline. Ann Oncol 2022 Sep;33(9):878-892. [doi: 10.1016/j.annonc.2022.04.007] [Medline: 35462007]
8. Warrington L, Absolom K, Conner M, et al. Electronic systems for patients to report and manage side effects of cancer treatment: systematic review. J Med Internet Res 2019 Jan 24;21(1):e10875. [doi: 10.2196/10875] [Medline: 30679145]
9. Basch E, Artz D, Iasonos A, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. J Am Med Inform Assoc 2007;14(3):264-268. [doi: 10.1197/jamia.M2177] [Medline: 17329732]
10. García Abejas A, Serra Trullás A, Sobral MA, Canelas D, Leite Costa F, Salvador Verges À. Improving the understanding and managing of the quality of life of patients with lung cancer with electronic patient-reported outcome measures: scoping review. J Med Internet Res 2023 May 3;25:e46259. [doi: 10.2196/46259] [Medline: 37021695]
11. Consolo L, Castellini G, Cilluffo S, Basile I, Lusignani M. Electronic patient-reported outcomes (e-PROMs) in palliative cancer care: a scoping review. J Patient Rep Outcomes 2022 Sep 23;6(1):102. [doi: 10.1186/s41687-022-00509-z] [Medline: 36138279]
12. Warnecke E, Salvador Comino M, Kocol D, et al. Electronic Patient-Reported Outcome Measures (ePROMs) improve the assessment of underrated physical and psychological symptom burden among oncological inpatients. Cancers (Basel) 2023;15(11):3029. [doi: 10.3390/cancers15113029]
13. Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197. [doi: 10.1001/jama.2017.7156]
14. Basch E, Schrag D, Jansen J, et al. Digital symptom monitoring with patient-reported outcomes in community oncology practices: a U.S. national cluster randomized trial. J Clin Oncol 2021 Dec 20;39(36_suppl):349527-349527. [doi: 10.1200/JCO.2021.39.36_suppl.349527]
15. Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
16. Basch E, Artz D, Dulko D, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. J Clin Oncol 2005 May 20;23(15):3552-3561. [doi: 10.1200/JCO.2005.04.275] [Medline: 15908666]
17. Common Terminology Criteria for Adverse Events (CTCAE) version 5. National Cancer Institute: Division of Cancer Treatment and Diagnosis. URL: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_5x7.pdf [accessed 2022-01-13]
18. Snyder CF, Blackford AL, Wolff AC, et al. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901. [doi: 10.1002/pon.3087] [Medline: 22544513]
19. Cella D, Choi SW, Condon DM, et al. PROMIS® adult health profiles: efficient short-form measures of seven health domains. Value Health 2019 May;22(5):537-544. [doi: 10.1016/j.jval.2019.02.004] [Medline: 31104731]
20. Basch E, Schrag D, Henson S, et al. Effect of electronic symptom monitoring on patient-reported outcomes among patients with metastatic cancer: a randomized clinical trial. JAMA 2022 Jun 28;327(24):2413-2422. [doi: 10.1001/jama.2022.9265] [Medline: 35661856]
21. Basch E, Reeve BB, Mitchell SA, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst 2014 Sep;106(9):1-11. [doi: 10.1093/jnci/dju244] [Medline: 25265940]
22. Velikova G, Absolom K, Hewison J, Holch P, Warrington L, Avery K, et al. Electronic Self-Reporting of Adverse Events for Patients Undergoing Cancer Treatment: The eRAPID Research Programme Including Two RCTs: NIHR Journals Library; 2022. [doi: 10.3310/FDDE8516]
23. Glenwright BG, Simmich J, Cottrell M, et al. Facilitators and barriers to implementing electronic patient-reported outcome and experience measures in a health care setting: a systematic review. J Patient Rep Outcomes 2023 Feb 14;7(1):13. [doi: 10.1186/s41687-023-00554-2] [Medline: 36786914]
24. Marco-Ruiz L, Bønes E, de la Asunción E, et al. Combining multivariate statistics and the think-aloud protocol to assess human-computer interaction barriers in symptom checkers. J Biomed Inform 2017 Oct;74:104-122. [doi: 10.1016/j.jbi.2017.09.002] [Medline: 28893671]
25. Lanzola G, Polce F, Parimbelli E, et al. The case manager: an agent controlling the activation of knowledge sources in a FHIR-based distributed reasoning environment. Appl Clin Inform 2023 Aug;14(4):725-734. [doi: 10.1055/a-2113-4443] [Medline: 37339683]
26. Glasspool D, Quaglini S, Kogan A, Leizer R, Wilk S. CAPABLE D59: final version of refined backend DSS including all components. Zenodo. 2024 Apr 2. URL: https://zenodo.org/records/10937793 [accessed 2025-11-26]

XSL•FO
RenderX

27. Bossi P, Antonuzzo A, Cherny NI, et al. Diarrhoea in adult cancer patients: ESMO Clinical Practice Guidelines. Ann Oncol 2018 Oct 1;29(Suppl 4):iv126-iv142. [doi: 10.1093/annonc/mdy145] [Medline: 29931177]
28. Peterson DE, Boers-Doets CB, Bensadoun RJ, Herrstedt J, ESMO Guidelines Committee. Management of oral and gastrointestinal mucosal injury: ESMO Clinical Practice Guidelines for diagnosis, treatment, and follow-up. Ann Oncol 2015 Sep;26 Suppl 5:v139-v151. [doi: 10.1093/annonc/mdv202] [Medline: 26142468]
29. Lacouture ME, Sibaud V, Gerber PA, et al. Prevention and management of dermatological toxicities related to anticancer agents: ESMO Clinical Practice Guidelines. Ann Oncol 2021 Feb;32(2):157-170. [doi: 10.1016/j.annonc.2020.11.005] [Medline: 33248228]
30. Fabi A, Bhargava R, Fatigoni S, Guglielmo M, Horneber M, Roila F, et al. ESMO Guidelines Committee. Electronic address: clinicalguidelines@ESMO.org. Cancer-related fatigue: ESMO Clinical Practice Guidelines for diagnosis and treatment. Ann Oncol 2020 Jun;31(6):713-723. [doi: 10.1016/j.annonc.2020.02.016]
31. Haanen J, Carbonnel F, Robert C, et al. Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2017 Jul;28:iv119-iv142. [doi: 10.1093/annonc/mdx225]
32. CTCAE and adverse event reporting: CTCAE. National Cancer Institute: Division of Cancer Diagnosis and Treatment. URL: https://dctd.cancer.gov/research/ctep-trials/for-sites/adverse-events#ctcae [accessed 2025-11-27]
33. Wang TQ, Samuel JN, Brown MC, et al. Routine surveillance of chemotherapy toxicities in cancer patients using the Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Oncol Ther 2018 Dec;6(2):189-201. [doi: 10.1007/s40487-018-0065-7] [Medline: 32700029]
34. Pradhan A, Dwivedi P, Pareek P, et al. Electronic-PRO Measures for adverse events Of Treatment In ONcology (E-PROMOTION). Int J Med Inform 2024 Feb;182:105305. [doi: 10.1016/j.ijmedinf.2023.105305] [Medline: 38061185]
35. Basch E, Becker C, Rogak LJ, et al. Composite grading algorithm for the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Clin Trials 2021 Feb;18(1):104-114. [doi: 10.1177/1740774520975120] [Medline: 33258687]
36. Fraterman I, Wollersheim BM, Tibollo V, et al. An eHealth App (CAPABLE) providing symptom monitoring, well-being interventions, and educational material for patients with melanoma treated with immune checkpoint inhibitors: protocol for an exploratory intervention trial. JMIR Res Protoc 2023 Oct 11;12:e49252. [doi: 10.2196/49252] [Medline: 37819691]
37. Locati LD, Rizzo M, Gabanelli P, et al. Electronic patient-reported outcome measures (E-Proms): results of the H2020 CAPABLE project for patients with metastatic renal cell cancer (mRCC). J Clin Oncol 2024 Jun 1;42(16_suppl):11045-11045. [doi: 10.1200/JCO.2024.42.16_suppl.11045]
38. Gabanelli P, Locati LD, Fiabane E, et al. The capable mhealth intervention: a qualitative study on patients' perspectives. SSRN . [doi: 10.2139/ssrn.4939228]
39. Glaser SLC, Fraterman I, van Brummelen N, et al. Usability and usefulness of a symptom management coaching system for patients with cancer treated with immune checkpoint inhibitors: comparative mixed methods study. JMIR Form Res 2025 Jan 23;9:e57659. [doi: 10.2196/57659] [Medline: 39847771]
40. Sacchi L, Veggiotti N, Girani E, Tibollo V, Quaglini S. Personalising symptoms reporting in telemonitoring applications for cancer patients. Stud Health Technol Inform 2022 May 25;294:900-904. [doi: 10.3233/SHTI220621] [Medline: 35612239]
41. Cureety. URL: https://www.cureety.com/ [accessed 2025-08-06]
42. Deutsch TM, Volmer LL, Feisst M, et al. ENABLE-app-based digital capture and intervention of patient-reported quality of life, adverse events, and treatment satisfaction in breast cancer: protocol for a randomized controlled trial. JMIR Res Protoc 2025 May 26;14:e69855. [doi: 10.2196/69855] [Medline: 40418799]
43. Xia Y, Guan X, Zhu W, Wang Y, Shi Z, He P. Effectiveness of symptom monitoring on electronic patient-reported outcomes (ePROs) among patients with lung cancer: a systematic review and meta-analysis. NPJ Digit Med 2025 Jul 3;8(1):399. [doi: 10.1038/s41746-025-01812-x] [Medline: 40604235]
44. Hassett MJ, Cronin C, Tsou TC, et al. eSyM: an electronic health record-integrated patient-reported outcomes-based cancer symptom management program used by six diverse health systems. JCO Clin Cancer Inform 2022 Jan;6:e2100137. [doi: 10.1200/CCI.21.00137] [Medline: 34985914]

Abbreviations

ADL: activity of daily living
AE: adverse event
CAPABLE: Cancer Patients Better Life Experience
CTCAE: Common Terminology Criteria for Adverse Events
DSS: decision support system
ePROMs: electronic patient-reported outcomes measures
ICI: immune checkpoint inhibitor
mHealth: mobile health
mRCC: metastatic renal cell carcinoma
NCI: National Cancer Institute

XSL•FO
RenderX

PRO: patient-reported outcome
QoL: quality of life
STAR: Symptom Tracking and Reporting

Edited by F Clemente; submitted 17.08.24; peer-reviewed by H Holmen, Y Zhu; revised version received 12.08.25; accepted 26.09.25; published 11.12.25.

Please cite as:
Sacchi L, Lanzola G, Quaglini S, Veggiotti N, Panzarasa S, Tibollo V, Terzaghi M, Fraterman I, Glaser S, Ottaviano M, Khadakou V, Hisko V, Peleg M, Wilgenhof S, Mallo H, Kogan A, Glasspool D, Medlock S, Del Campo L, Gabetta M, Rizzo M, Locati LD, Gabanelli P, Demurtas S, Premoli A, Wilk S
Obtaining Patient-Reported Outcome Data via a Home Patient Monitoring App: Development, Implementation, and Validation of Novel Interface Terminology
JMIR Mhealth Uhealth 2025;13:e65504
URL: https://mhealth.jmir.org/2025/1/e65504
doi:10.2196/65504

© Lucia Sacchi, Giordano Lanzola, Silvana Quaglini, Nicole Veggiotti, Silvia Panzarasa, Valentina Tibollo, Matteo Terzaghi, Itske Fraterman, Savannah Glaser, Manuel Ottaviano, Vadzim Khadakou, Vitali Hisko, Mor Peleg, Sofie Wilgenhof, Henk Mallo, Alexandra Kogan, David Glasspool, Stephanie Medlock, Laura Del Campo, Matteo Gabetta, Mimma Rizzo, Laura Deborah Locati, Paola Gabanelli, Sara Demurtas, Andrea Premoli, Szymon Wilk. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Dynamic Assessment of Fine Motor Control and Vocalization in Parkinson Disease Through a Smartphone App: Cross-Sectional Study of Time-Severity Interaction Effects

Youho Myong[1,2,3*], MD, PhD; Seo Jung Yun[2,4*], OT, MS; Kyudong Park[5,6*], PhD; Byung-Mo Oh[2,7,8], MD, PhD; Han Gil Seo[2,7], MD, PhD

[1]Department of Biomedical Engineering, Seoul National University College of Medicine, Seoul, Republic of Korea

[2]Department of Rehabilitation Medicine, Seoul National University Hospital, 101 Daehak-no, Jongno-gu, Seoul, Republic of Korea

[3]Medical Research Center, Seoul National University, Seoul, Republic of Korea

[4]Department of Human Systems Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

[5]Department of Artificial Intelligence Applications, Kwangwoon University, Seoul, Republic of Korea

[6]School of Information Convergence, Kwangwoon University, Seoul, Republic of Korea

[7]Department of Rehabilitation Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

[8]Institute on Aging, Seoul National University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Han Gil Seo, MD, PhD
Department of Rehabilitation Medicine, Seoul National University Hospital, 101 Daehak-no, Jongno-gu, Seoul, Republic of Korea

Abstract

Background: Parkinson disease (PD) is a progressive neurodegenerative disorder characterized by motor and nonmotor symptoms that worsen over time, significantly impacting quality of life. While clinical evaluations such as the Unified Parkinson's Disease Rating Scale (UPDRS) are standard for assessing disease severity, they offer somewhat limited temporal resolution and are susceptible to observer variability. Smartphone apps present a viable method for capturing detailed fluctuations in motor and vocal functions in real-world settings.

Objective: This study aimed to use a smartphone-based app to quantitatively evaluate the interaction effect between time and disease severity on motor and vocal symptoms in individuals with PD.

Methods: This was an exploratory, cross-sectional pilot study. Disease severity in persons with PD was assessed using the modified Hoehn & Yahr Scale, Voice Handicap Index, and UPDRS. We used a custom smartphone app to administer finger-tapping tasks, sustained phonation (/a/ and /i/), and rapid syllable repetition (/dadada/ and /pa-ta-ka/). The total tap counts, tap-to-tap variability, and vocal parameters (loudness, jitter, shimmer, repeat counts, and their variability) were analyzed. Each task was divided into 5 equal time frames to analyze performance changes over a short duration. Time-severity interactions were examined using linear mixed models.

Results: In total, 20 persons with PD and 20 healthy adults were included in this study. Persons with PD showed worse motor and vocal performance compared to healthy adults, with higher dysrhythmia; worse jitter, shimmer, and jitter and shimmer variability; and fewer repeat counts. During finger-tapping tasks, individuals with PD showed an earlier onset of dysrhythmia than their healthy counterparts. While a higher UPDRS part III score was associated with greater finger-tapping variability, there was no significant time-severity interaction for this motor task. However, linear mixed model analysis revealed significant time-severity interaction effects for vocal tasks, including /a/ loudness (P=.001), /a/ jitter (P=.01), /a/ shimmer (P=.001), /i/ loudness (P=.001), /i/ jitter (P<.001), /i/ shimmer (P<.001), and /pa-ta-ka/-variability (P=.04). This indicates that individuals with higher UPDRS part III scores experienced a more rapid decline in vocal control during the assessment period. All measured smartphone-based characteristics showed a significant correlation with UPDRS part III scores, with finger-tapping variability having the strongest correlation.

Conclusions: This study demonstrates that a smartphone-based assessment, conducted over just a few minutes, can detect subtle temporal changes in fine motor and vocal control. The app successfully captured the earlier onset of dysrhythmia in individuals with PD and, importantly, identified significant time-severity interaction effects in vocal performance. This suggests that such digital tools can provide sensitive, dynamic insights into symptom progression, potentially enabling more precise monitoring and timely clinical interventions for individuals with PD.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e69028) doi:10.2196/69028

KEYWORDS

Parkinson Disease; Mobile Applications; Digital Biomarkers; Voice; Motor Skills

Introduction

Background

Parkinson disease (PD) is a progressive neurodegenerative disorder characterized by the degeneration of dopaminergic neurons in the substantia nigra, a critical region of the basal ganglia [1,2]. This neuronal loss causes a marked decrease in dopamine, a neurotransmitter essential for regulating motor control [3]. Therefore, patients with PD experience motor symptoms such as bradykinesia, rigidity, resting tremor, and postural instability, progressively impairing their ability to perform daily activities. PD is further associated with a range of nonmotor symptoms that often precede motor symptoms. These include autonomic dysfunction [4], sleep disturbances [5], neuropsychiatric symptoms [6], sensory deficits [7], and gastrointestinal issues [8]. These symptoms highlight the broad impact of PD beyond large motor impairments that complicate disease management.

Of particular relevance to this study is the impact of PD on speech and fine motor control, which are closely linked to disruptions in basal ganglia–thalamocortical circuits [1,2,9]. As the disease progresses, patients often develop hypophonia and dysarthria that significantly impair their ability to communicate effectively [10]. These symptoms, combined with motor impairments, underscore the complexity of PD and highlight the need for comprehensive diagnostic and therapeutic strategies.

Finger tapping is a widely recognized assessment tool for evaluating motor dysfunction in PD [11]. The bradykinesia and decreased motor coordination associated with PD result in reduced total tap counts, increased variability in tap frequency, and impaired rhythmicity [12,13]. Recent studies have reported dysrhythmia or increased tap-to-tap variability as the main diagnostic features of PD using smartphone app [14-16]. Voice characteristics such as loudness, jitter (frequency variability), and shimmer (amplitude variability) are often altered in persons with PD owing to the rigidity and bradykinesia of the vocal muscles [10,17]. Vocal function tasks, such as producing sustained vowel sounds (/a/, /i/) and rapid syllable repetition (/dadada/, /pa-ta-ka/), can reveal subtle changes in voice control and coordination, making them valuable for assessing dysdiadochokinesis (difficulty with rapid alternating movements) and interval variability [18,19].

Traditionally, the diagnosis and assessment of disease progression of PD have relied on clinical observations of motor symptoms [2]. However, these methods can be subjective and may vary among clinicians, making episodic assessments during clinic visits inadequate for capturing day-to-day symptom variability, particularly in the early stages when symptoms may be subtle and fluctuating [19].

In addition to clinical observations, recent advancements in diagnostic methodologies have broadened the spectrum of tools available for the diagnosis of PD. Neuroimaging techniques such as dopamine transporter single-photon emission computed tomography and neuromelanin-sensitive magnetic resonance imaging have shown utility in supporting clinical diagnosis by visualizing nigrostriatal degeneration and related pathophysiology [1]. Furthermore, genetic testing is increasingly recognized as a valuable tool, particularly in patients with early-onset or familial PD. Mutations in genes such as *LRRK2*, *GBA*, and *SNCA* have been linked to specific subtypes of PD and may inform not only diagnosis but also prognosis and therapeutic selection. Recent commentary emphasizes the importance of integrating genetic testing into routine clinical care for patients with PD [2]. These evolving diagnostic modalities complement traditional clinical assessments and support the need for objective, technology-based tools for earlier and more precise characterization of PD symptomatology.

Such technological needs have paved the way for more accessible and continuous monitoring solutions, such as smartphone-based apps [20,21]. These apps leverage the sensors and processing capabilities of smartphones to perform real-time assessments of motor and vocal functions, offering a noninvasive, continuous monitoring tool that can be used in natural environments [17]. Smartphone-based assessments enable millisecond-level quantitative evaluation of dysrhythmia without observer biases [14]. Furthermore, the integration of smartphone-based monitoring into the diagnostic process for PD offers several advantages, including frequent symptom tracking in the patient's natural environment, providing a more comprehensive picture of disease progression [19,22]. This approach may improve diagnostic accuracy, particularly in the early stages of PD when traditional clinical assessments may miss subtle symptoms [23,24]. Furthermore, by enabling remote monitoring, these apps can enhance patient care by providing real-time feedback to clinicians and enabling timely adjustments to treatment plans [19].

Objectives

In this study, we developed a smartphone app to monitor PD-related motor and vocal symptoms. Using this app, we evaluated the finger-tapping and voice characteristics of individuals with PD and healthy adults. This study hypothesized that individuals with more severe PD would experience earlier onset of decline in vocal and fine motor control. Therefore, this study aimed to identify the differences in vocal and fine motor control function deterioration among persons with different levels of PD severity and confirm the time-severity interaction effect on finger-tapping and vocal assessments. In doing so, the authors expect that this study investigation would provide additional insights into the time-lapse perspective on PD diagnosis and severity assessment, leading to more precise diagnostic criteria and severity stratification.

XSL•FO
RenderX

Methods

Study Design and Participants

This was an exploratory, cross-sectional pilot study. The inclusion criteria for healthy adults were as follows: (1) age >19 years and (2) those who voluntarily consented in writing and agreed to participate in the study. The exclusion criteria were as follows: (1) history of neurological or orthopedic conditions that could impair upper limb function, (2) history of neurological or otolaryngological conditions that might impact voice and speech, (3) presence of facial paralysis, and (4) individuals considered unsuitable due to other concurrent conditions. The inclusion criteria for patients with PD were as follows: (1) age >19 years, (2) clinical diagnosis of idiopathic PD according to the UK Parkinson's Disease Society Brain Bank Diagnostic Criteria, and (3) modified Hoehn and Yahr (H&Y) scale stage 1 - 3. The exclusion criteria were as follows: (1) history of neurological or orthopedic diseases other than PD that could affect upper limb function; (2) history of neurological or otolaryngological diseases other than PD that could affect voice and speech; (3) presence of facial paralysis due to conditions other than PD; (4) a diagnosis of PD dementia based on the Korean Montreal Cognitive Assessment (K-MoCA), with the following cutoff scores [25]: <7 points: illiterate, <13 points: education duration of 0.5 to 3 years, <16 points: education duration of 4 to 6 years, <19 points: education duration of 7 to 9 years, and <20 points: education duration of 10 years or more; (5) presence of severe dyskinesia or on-off fluctuations; and (6) any comorbid conditions deemed to make study participation difficult [25]. Participants were recruited from an outpatient clinic at Seoul National University Hospital between January 2023 and January 2024, with the last follow-up date being January 23, 2024. Evaluations for patients with PD were conducted in the "on" state, representing the peak effect of PD medication. The "on" state refers to a self-perceived period of optimal physical function following antiparkinsonian medication. In this study, most assessments were performed approximately 1 to 1.5 hours after medication intake, when participants commonly reported being in their optimal functional state.

Ethical Considerations

The study protocol was approved by the institutional review board of Seoul National University Hospital (2211-048-1377; approved on December 16, 2022). Written informed consent was obtained from all participants. All study data were anonymized before analysis. This study was conducted in accordance with the principles of Good Clinical Practice and the Declaration of Helsinki. All participants received monetary compensation covering transportation to and from the study site. The participant pictured in this study provided written consent to publish the deidentified photograph for academic purposes.

Experimental Design

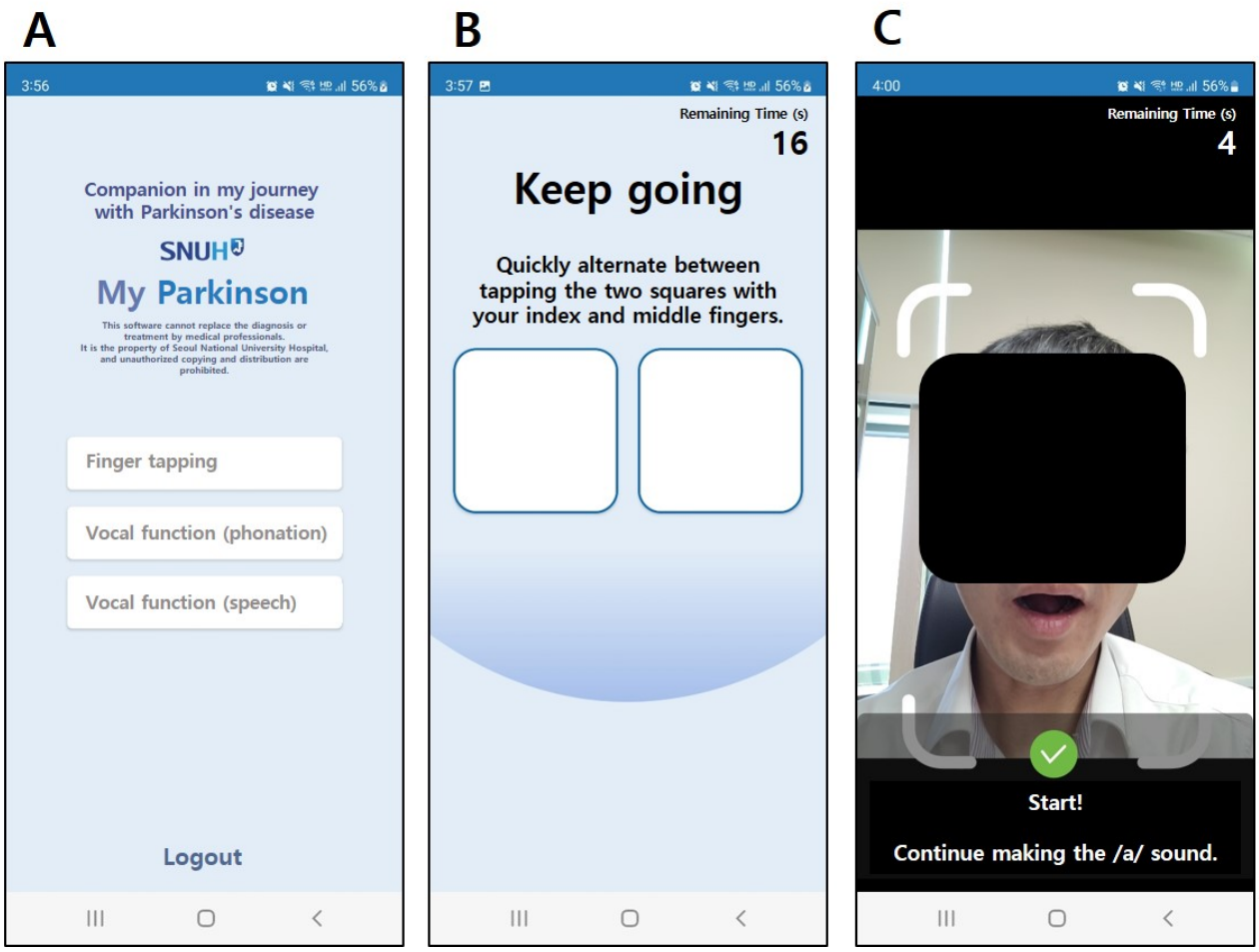
An in-house smartphone app was developed to quantify the fine motor control and vocal functions (Figure 1A). The app collected data on finger-tapping and various voice parameters using the built-in touch screen, microphone, and camera of an Android (Google LLC) smartphone. Access to the app was secured using an ID and a password. During registration, participant information such as name, date of birth, sex, and year of PD diagnosis was collected. Before each assessment session, participants were prompted by the app to input the time of their most recent antiparkinsonian medication intake. The app automatically recorded the start time of each assessment, allowing calculation of the elapsed time between medication intake and task performance.

For the finger-tapping assessment, participants were instructed to alternately tap 2 squares displayed on the smartphone screen as quickly as possible using the index and middle fingers for 20 seconds (Figure 1B). Separate menus were provided for the right and left hands. The voice assessment comprised 4 tasks: 2 phonation tasks and 2 speech tasks. In the phonation tasks, participants were required to sustain the vowel sounds /a/ and /i/ for 10 seconds without interruption. Each task was performed 3 times, with a 3-second rest between trials, and the transitions between the tasks and rest periods were automated. The speech assessment included alternating motion rate and sequential motion rate tasks. The alternating motion rate tested participants' ability to rapidly and repeatedly articulate the same phoneme, with instructions to say /dadada/ for 5 seconds. This was repeated 3 times with a 3-second interval between repetitions. The sequential motion rate evaluated the capacity to articulate different phonemes in sequence, where participants were instructed to say /pa-ta-ka/ for 5 seconds and repeat 3 times with 3-second breaks. Voice recordings were made using a smartphone microphone. To ensure consistent audio levels, the camera detected the participant's face and initiated recordings only when the face was within a specified distance from the device (Figure 1C). If the participant moved outside the predefined range either before or during the rest period, the test did not proceed.

The participants, including both healthy adults and persons with PD, underwent a single session of finger-tapping and vocal function evaluation using a smartphone app. All evaluations were conducted in a controlled hospital environment using a Galaxy S10 (Samsung) provided by the research team, with the investigator present to ensure standardized administration. Each participant completed all assessments twice. For healthy adults, only the age and sex were recorded. The baseline characteristics of persons with PD, including age, sex, disease duration, H&Y stage, Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) score, K-MoCA score, and Voice Handicap Index-10 (VHI-10), were collected. The MDS-UPDRS was developed to assess the impact of PD on activities of daily living and overall disease burden, allowing for the evaluation of both motor and nonmotor symptoms [26]. The MoCA assesses 8 cognitive domains to screen for cognitive impairment [27]. The VHI-10 is a self-assessment tool that measures the perceived severity of voice disorders and is categorized into functional, physical, and emotional domains [28]. It is widely used in adult patients with various types of voice disorders.

XSL•FO
RenderX

Figure 1. Screenshots of the smartphone app: (**A**) main menu, (**B**) finger-tapping task, and (**C**) phonation task. For the convenience of readers, the text in the figure was translated into English (originally in Korean).



Data Extraction

In the finger-tapping task, data were collected using a mobile app that recorded the frequency of participants' alternating button taps on the screen. Only taps in which the participants alternated between the left and right buttons were included in the analysis. Instances in which a single button was pressed consecutively without alternating were excluded. To assess the variability in the tapping frequency, the SD of the time intervals between consecutive taps was calculated.

For voice analysis, we used the OpenSMILE for Python toolkit version 2.5.0 (audEERING®) [29] to extract specific acoustic features from the eGeMAPSv02 feature set [30]. The extracted features included loudness, jitter, and shimmer, which are commonly used in speech analysis to quantify variations in voice quality. To assess the variability of jitter and shimmer, the coefficient of variation, a normalized measure of variability that represents the ratio of the SD to the mean, was calculated. This approach allowed quantification of the relative variability in these acoustic features, providing insights into the participants' voice modulation patterns. In addition, during the speech tasks in which participants produced the /dadada/ and /pa-ta-ka/ sounds, variability in the frequency of speech production was calculated as the SD of the time intervals between consecutive sounds. This measure helped to evaluate the consistency of sound production, offering further insight into the participants' speech motor control and temporal coordination during these tasks.

Statistical Analysis

Data are expressed as means and SDs for continuous variables and as frequencies (percentages) for categorical variables. Each recorded entry of finger-tapping (20 s on each side) and vocal tasks (10 s for phonation and 5 s for speech) was divided into 5 equal time frames (TFs), with the first TF being TF1, and the last being TF5 (ie, for finger tapping: TF1=0 s-4 s, TF2=4 s-8 s, TF3=8 s-12 s, TF4=12 s-16 s, and TF5=16 s-20 s; for phonation: TF1=0 s-2 s, TF2=2 s-4 s, TF3=4 s-6 s, TF4=6 s-8 s, and TF5=8 s-10 s; and for speech: TF1=0 s-1 s, TF2=1 s-2 s, TF3=2 s-3 s, TF4=3 s-4 s, and TF5=4 s-5 s). ANOVA was conducted to determine whether there were significant differences between the different TFs and groups [31]. A linear mixed model (LMM) was applied to examine the time-group and time-severity interaction effects within each data group [32]. Statistical significance was set at $P<.05$. All statistical analyses were performed using SPSS (version 26.0; IBM Corp).

Results

Study Participants

In total, 20 individuals with PD and 20 healthy adults without a history of neurological disorders were enrolled in this study.

XSL•FO
RenderX

The patient group comprised 85% (17/20) female participants with a mean age of 68.30 (SD 6.30) years. The H&Y scale was distributed as follows: 1.5 (4/20, 20%), 2 (5/20, 25%), 2.5 (7/20, 35%), and 3 (4/20, 20%). The mean MDS-UPDRS part III score was 22.10 (SD 10.64), the mean VHI-10 12.05 (SD 9.59), and the mean K-MoCA 25.25 (SD 2.40). The healthy adults group comprised 50% (10/20) female participants with a mean age of 33.55 (SD 7.85) years. The demographic information of the study participants is presented in Table 1.

Table . Demographic characteristics (N=20 for each group).

	Persons with PD[a]	Healthy adults
Age (years), mean (SD)	68.30 (6.30)	33.55 (7.85)
Sex (female), % (n/N)	85 (17/20)	50 (10/20)
H&Y[b] scale, % (n/N)		
1.5	20 (4/20)	N/A[c]
2	25 (5/20)	N/A
2.5	35 (7/20)	N/A
3	20 (4/20)	N/A
Disease duration (years), mean (SD)	8.90 (3.78)	N/A
UPDRS[d] total, mean (SD)	50.90 (16.00)	N/A
UPDRS part III, mean (SD)	22.10 (10.64)	N/A
VHI-10[e], mean (SD)	12.05 (9.59)	N/A
K-MoCA[f], mean (SD)	25.25 (2.40)	N/A

[a]PD: Parkinson disease.

[b]H&Y: Hoehn and Yahr.

[c]Not applicable.

[d]UPDRS: Unified Parkinson's Disease Rating Scale.

[e]VHI-10: Voice Handicap Index 10.

[f]K-MoCA: Korean Montreal Cognitive Assessment.

Overall Characteristics of Finger Tapping

The patient group demonstrated a lower total number of taps (mean 76.250, SD 35.932 vs mean 199.263, SD 32.600; *P*<.001) and a notably higher tap frequency variability (mean 0.158, SD 0.144 vs mean 0.0288, SD 0.0101; *P* <.001) than the healthy adult group. This phenomenon was consistently observed across all TFs when the total experimental time was evenly divided into 5 segments (TF1-TF5). The results of the mean total number of taps and tap-to-tap variability are presented in Table 2.

Table . Time frame (TF)–wise analysis of tap-to-tap variability.

	Healthy adults, mean (SD)	Persons with PD[a], mean (SD)	*t* value (df)	*P* value
Total taps	199.263 (32.600)	76.250 (35.932)	22.678 (38)	<.001
Mean variability	0.0288 (0.0101)	0.158 (0.144)	–8.017 (38)	<.001
TF 1	0.0238 (0.0100)	0.101 (0.0803)	–8.563 (38)	<.001
TF 2	0.0238 (0.0100)	0.113 (0.0744)	–10.657 (38)	<.001
TF 3	0.0258 (0.0109)	0.130 (0.0920)	–10.067 (38)	<.001
TF 4	0.0284 (0.0128)	0.134 (0.0860)	–10.897 (38)	<.001
TF 5	0.0323 (0.0124)	0.143 (0.0881)	–11.303 (38)	<.001

[a]PD: Parkinson disease.

Time-Group Interaction Effects on Finger Tapping

Both groups experienced increasing tap-to-tap variability as the TF progressed (*P*<.001). Furthermore, the LMM analysis revealed a significant time-by-group interaction during the finger-tapping task (*P*=.02), implying that the effect of time on increasing tap-to-tap variability was experienced differently by persons with PD and healthy adults (Table 3).

XSL•FO
RenderX

Table . Interaction effect of time and group on finger tapping.

	Coefficient (SE; 95% CI)	*t* value (df)	*P* value
Intercept	0.021 (0.012; –0.002 to 0.044)	1.815 (38)	.07
Time frame	0.072 (0.017; 0.040 to 0.105)	4.372 (38)	<.001
Group	0.002 (0.003; –0.003 to 0.007)	0.683 (38)	.50
Time frame: group	0.009 (0.004; 0.002 to 0.016)	2.388 (38)	.02

The positive coefficient in the TF:group row in Table 3 indicates that the patient group exhibited greater early-onset variability compared with the control group. This finding was further supported by the ANOVA and post hoc analysis of interval variability.

To assess whether the tapping intervals in both groups remained consistent or if fatigue was observed, an ANOVA was conducted on the interval variability across TF1 to TF5, which revealed a significant difference in the mean interval variability across the 5 TFs (P=.001 for healthy adults and P=.02 for persons with PD). Post hoc analyses found that while the interval variability of the control group significantly increased starting from TF3 (P=.01), that of the patient group did so earlier, starting from TF2 (P=.05). Together with the findings in Table 3, this implies that there was a meaningful interaction between time and group. More specifically, the patient group started to experience increasing variability earlier than the controls. The results of the ANOVA and post hoc analyses are described in (Tables S1-S4 in Multimedia Appendix 1).

Interaction of Time and Disease Severity on Finger Tapping in Persons With PD

While a higher UPDRS part III score indicated higher interval variability (P=.001), there was no significant interaction effect between time and disease severity (Table 4).

Table . Interaction effect between time and disease severity.

	Coefficient (SE; 95% CI)	*t* value (df)	*P* value
Intercept	0.0290 (0.021312; –0.0129 to 0.0709)	1.359 (19)	.18
Time frame	0.0112 (0.006426; –0.0015 to 0.0238)	1.739 (19)	.08
UPDRS[a] III	0.00291 (0.000869; 0.0012 to 0.0046)	3.352 (19)	.001
Time frame: UPDRS part III	–0.000376 (0.000262; –0.00055 to 0.00048)	–0.143 (19)	.89

[a]UPDRS part III: Unified Parkinson's Disease Rating Scale part III.

Overall Characteristics of Voice

The persons with PD group demonstrated significantly louder volumes (P<.001 for both /a/ and /i/) and higher mean jitter (P<.001 for /a/ and P=.006 for /i/), jitter variability (P<.001 for both /a/ and /i/), mean shimmer (P=.008 for /a/), and shimmer variability (P<.001 for both /a/ and /i/) in sustained phonation. For repeating speech tasks (/dadada/ and /pa-ta-ka/ sounds), the persons with PD group showed lower total repeats (P<.001 for both /dadada/ and /pa-ta-ka/), with elevated interval variability (P<.001 for /dadada/). Voice characteristics are presented in Table 5. There was no significant time-group interaction in the voice analysis (Table S5 in Multimedia Appendix 1).

XSL•FO
RenderX

Table . Group comparison on voice characteristics.

	Healthy adults, mean (SD)	Persons with PD[a], mean (SD)	*t* value (df)	*P* value
Sustained phonation				
Vowel /a/				
Loudness	1.573 (0.735)	2.051 (0.844)	–4.514 (38)	<.001
Mean jitter	0.0074 (0.00521)	0.151 (0.0174)	–4.647 (38)	<.001
Jitter variability	1.930 (1.583)	3.181 (1.959)	–5.180 (38)	<.001
Mean shimmer	0.477 (0.196)	0.578 (0.332)	–2.692 (38)	.008
Shimmer variability	0.737 (0.303)	0.971 (0.337)	–5.446 (38)	<.001
Vowel /i/				
Loudness	1.082 (0.503)	1.302 (0.519)	–3.225 (38)	.001
Mean jitter	0.0071 (0.00583)	0.0103 (0.0105)	–2.773 (38)	.006
Jitter variability	2.011 (1.960)	3.673 (2.420)	–5.608 (38)	<.001
Mean shimmer	0.402 (0.190)	0.417 (0.246)	–0.497 (38)	.62
Shimmer variability	0.766 (0.352)	1.071 (0.452)	–5.577 (38)	<.001
Speech				
/dadada/				
Repeat count	40.067 (4.623)	36.225 (5.329)	5.791 (38)	<.001
Interval variability	1.931 (0.874)	3.705 (2.755)	–6.325 (38)	<.001
/pa-ta-ka/				
Repeat count	41.010 (7.011)	37.442 (5.475)	4.089 (38)	<.001
Interval variability	4.970 (10.808)	5.087 (3.938)	–0.101 (38)	.91

[a]PD: Parkinson disease.

Interaction of Time and Disease Severity on Voice Characteristics in Persons With PD

A significant time-severity interaction effect was observed in many aspects of voice analysis. Specifically, time and UPDRS part III interaction effects were present in /a/ loudness (P=.001), /a/ jitter (P=.01), /a/ shimmer (P=.001), /i/ loudness (P=.001), /i/ jitter (P<.001), /i/ shimmer (P<.001), and /pa-ta-ka/-variability (P=.04). These results suggest that individuals with different disease severities exhibit significantly different performance trajectories over time. Participants with higher UPDRS part III scores showed greater declines in vocal and fine motor control over time, reflected by larger increases in jitter and shimmer and higher repeat variability in later time frames compared with those with lower scores. It is noteworthy that the time-disease severity interaction effect was significant in dependent variables where the independent time or independent disease severity effect was not statistically significant. This will be explored in greater depth in the Discussion section. The detailed results of the LMM analysis are described in Table 6.

XSL•FO
RenderX

Table . Interaction effect between time and Unified Parkinson's Disease Rating Scale part III (UPDRS part III) score in persons with Parkinson disease.

Variable	Coefficient (SE)	*t* value (df)	*P* value
Sustained phonation			
Vowel /a/			
Loudness			
Intercept	2.319 (0.190)	12.203 (19)	<.001
TF[a]	0.0825 (0.0573)	1.440 (19)	.15
UPDRS	0.00123 (0.00789)	0.155 (19)	.88
TF:UPDRS	–0.00830 (0.00238)	–3.489 (19)	.001
Jitter			
Intercept	0.00929 (0.00330)	2.820 (19)	.005
TF	–0.000456 (0.000994)	–0.459 (19)	.65
UPDRS	–0.000111 (0.000137)	–0.808 (19)	.42
TF:UPDRS	0.000106 (0.0000412)	2.578 (19)	.01
Shimmer			
Intercept	0.395 (0.0790)	5.001 (19)	<.001
TF	–0.0183 (0.0238)	–0.769 (19)	.44
UPDRS	–0.000815 (0.00328)	–0.249 (19)	.80
TF:UPDRS	0.00338 (0.000989)	3.416 (19)	.001
Vowel /i/			
Loudness			
Intercept	1.341 (0.118)	11.363 (19)	<.001
TF	0.0787 (0.0356)	2.211 (19)	.03
UPDRS	0.00238 (0.00481)	0.495 (19)	.62
TF:UPDRS	–0.00492 (0.00145)	–3.389 (19)	.001
Jitter			
Intercept	0.0112 (0.00285)	3.925 (19)	<.001
TF	–0.00232 (0.000860)	–2.700 (19)	.007
UPDRS	–0.000357 (0.000116)	–3.068 (19)	.002
TF:UPDRS	0.000196 (0.0000351)	5.602 (19)	<.001
Shimmer			
Intercept	0.393 (0.0674)	5.831 (19)	<.001
TF	–0.0470 (0.0203)	–2.310 (19)	.02
UPDRS	–0.00823 (0.00275)	–2.993 (19)	.003
TF:UPDRS	0.00498 (0.000829)	6.013 (19)	<.001
Speech			
/dadada/-variability			
Intercept	1.474 (0.441)	3.343 (19)	.001
TF	–0.0433 (0.133)	–0.326 (19)	.74
UPDRS	0.0462 (0.0180)	2.570 (19)	.01
TF:UPDRS	0.00307 (0.00542)	0.566 (19)	.57
/pataka/-variability			
Intercept	0.424 (0.746)	0.569 (19)	.57
TF	0.724 (0.225)	3.220 (19)	.001

XSL•FO
RenderX

Variable	Coefficient (SE)	*t* value (df)	*P* value
UPDRS	0.0961 (0.0304)	3.164 (19)	.002
TF:UPDRS	–0.0194 (0.00916)	–2.117 (19)	.04

[a]TF: time frame.

Correlation and Linear Regression Analyses of Disease Severity on Finger Tapping and Voice Characteristics

All voice characteristics and finger tapping were significantly correlated with the UPDRS part III scores. Linear regression analysis revealed that finger tapping had the strongest correlation with the UPDRS part III score (r=0.35; R^2=0.123; P<.001). The results of the correlation and linear regression analyses are presented in Table 7.

Table . Correlation and linear regression analysis of disease severity on finger tapping and voice characteristics.

	Correlation coefficient (*r*)	R^2 [a]	*P* value
Finger tapping	0.35	0.123	<.001
Sustained phonation			
Vowel /a/			
Loudness	–0.27	0.073	<.001
Jitter	0.14	0.02	<.001
Shimmer	0.25	0.065	<.001
Vowel /i/			
Loudness	–0.24	0.056	<.001
Jitter	0.18	0.032	<.001
Shimmer	0.21	0.046	<.001
Speech			
/dadada/-variability	0.28	0.081	<.001
/pa-ta-ka/-variability	0.12	0.014	.004

[a]R^2: coefficient of determination in linear regression analysis.

Discussion

Principal Findings

This study investigated chronological differences in vocal and fine motor control deterioration among individuals with varying levels of PD severity, with the goal of discovering the interaction effect between time and disease severity on finger-tapping and vocal assessments. Voice analysis demonstrated significant time-severity interaction effects in many aspects. Persons with higher UPDRS part III scores showed greater increases in /a/-loudness, /a/-jitter, /a/-shimmer, /e/-loudness, /e/-jitter, /e/-shimmer, and /pa-ta-ka/-variability. Such time-severity interaction effects were present even in the dependent variables where the independent time or independent disease severity effect was not statistically significant. Finger-tapping performance was significantly affected only by UPDRS part III scores, with no evident time-severity interaction.

In recent years, there has been growing interest in using smartphone-based apps to provide objective and continuous monitoring of PD symptoms [13,17,19-21,23,24]. Unlike traditional assessments, a smartphone app can continuously and quantitatively monitor finger-tapping performance and detect subtle changes in motor function that might not be evident during a physical examination [33]. By capturing tap-to-tap variability in finger tapping down to milliseconds and analyzing the data across 5 TFs, it became possible to examine both the time effect and the time-group interaction effect in individuals with PD and healthy adult groups. This continuous data collection allows for the detection of variability and subtle symptoms that might otherwise go unnoticed, thereby providing a more comprehensive understanding of a patient's condition over time [19]. The group effect between the patient and healthy adult groups was evident, as shown in Table 2. Further intriguing was the time-group interaction effect observed in finger tapping, as shown in Table 3 and Tables S1-S4 in Multimedia Appendix 1. This suggests that the temporal worsening of dysrhythmia develops differently, even among persons with PD, depending on the disease severity.

In addition to motor symptoms, PD also affects vocal function, presenting in symptoms such as hypophonia (reduced speech volume) and dysarthria (difficulty articulating speech) [6,17,34]. These symptoms are linked to the rigidity and bradykinesia of the vocal muscles, which are also controlled by the basal ganglia–thalamocortical circuits [35]. In this study, vocal characteristics such as loudness, jitter, and shimmer were significantly impaired in persons with PD compared to healthy adults. The increased interval variability observed during fast-alternating vocalization tasks (/dadada/ and /pa-ta-ka/)

XSL•FO
RenderX

further indicated impaired articulatory control in patients with PD [19].

Statistical analysis revealed significant time-disease severity interaction effects for nearly all the voice characteristics examined. Notably, this interaction effect was observed even when the individual effects of time or UPDRS part III scores on the voice characteristics (eg, loudness, jitter, and shimmer) were not significant. This phenomenon highlights an important insight often seen in statistical analysis: the main effects may be non-significant, yet the interaction effect can still be meaningful. In such scenarios, while the UPDRS part III scores may not show a significant impact on vocal function when averaged across all TFs, their effect may vary significantly across different periods. For instance, a patient might experience minimal vocal changes during sustained phonation in the early stages of the disease. However, as the condition progresses, the impact on vocal changes could become more pronounced [35]. This underscores the importance of considering interaction effects within the model, as they can reveal critical dynamics that are not captured by the main effects alone.

The time-disease severity interaction within the patient group further suggests that the progression of vocal symptoms varies according to disease severity over time, emphasizing the need for dynamic, longitudinal evaluations of vocal function in individuals with PD. As PD advances, the impact on vocal symptoms fluctuates depending on the severity at different stages, with those experiencing mild disease severity initially showing minimal changes, which may later intensify as the disease progresses [23,35,36]. This variability highlights the need for the regular monitoring of vocal function to allow timely adjustments in treatment strategies. Longitudinal assessments are essential for capturing the nuanced progression of symptoms and tailoring interventions to meet the evolving needs of patients [19,24,35].

The time-severity interaction effect observed in vocal function suggests that individuals with higher UPDRS part III scores experience greater vocal fatigue over time. Increased inhibition of the thalamus leads to a significant reduction in the ability of the motor cortex to generate and sustain precise, coordinated movements required for normal speech. Therefore, these individuals experience greater vocal fatigue, characterized by earlier and more severe reductions in vocal loudness and increased variability in pitch (jitter) and amplitude (shimmer). Over time, the continued loss of dopamine exacerbates these issues, resulting in a progressive decline in vocal function. While the correlation coefficients in Table 7 are modest, this may reflect the difference in temporal resolution and scope between app-based metrics and clinical rating scales. Smartphone measures capture specific, moment-to-moment performance, whereas clinical scores aggregate broader symptom impressions over time.

The earlier and greater exacerbation of dysrhythmia and vocal control deficits in individuals with higher disease severity reflects progressive degeneration of the basal ganglia–thalamocortical circuits [9,37]. Finger-tapping dysrhythmia is linked to the dysfunction of direct and indirect pathways, particularly those involving the degeneration of dopaminergic neurons in the substantia nigra pars compacta. As the substantia nigra pars compacta dopaminergic output to the striatum decreases with advancing disease, the inhibitory-excitatory balance within the striatonigral (direct) and striatopallidal (indirect) pathways is disrupted [16]. This imbalance exacerbates motor dysrhythmia, especially in higher-severity cases, due to increased subthalamic nucleus burst firing, which contributes to abnormal motor rhythms [14].

Conversely, vocal control deficits, which are more dependent on sustained muscle activity and coordination, are further compromised by advanced PD owing to greater degeneration in the pathways responsible for fine motor control, particularly the hyperdirect cortico-subthalamic pathway [22,24]. This pathway plays a crucial role in coordinating the speech muscles. In individuals with higher disease severity, the progressive loss of dopaminergic input leads to earlier fatigue and diminished motor precision, resulting in worsened vocal control [9].

Importantly, the strength of smartphone-based assessment lies not only in its accessibility but also in its capacity to quantify time-sensitive symptom trajectories with high granularity. Although this study did not observe a significant time-severity interaction in fine motor control (finger tapping), such effects were clearly evident in vocal characteristics. This discrepancy highlights the potential of vocal metrics as sensitive indicators of progression and suggests that certain domains may respond differently to disease burden over time. The ability to detect these nuanced changes through a smartphone app underscores its utility in capturing progression patterns that may elude conventional assessments, reaffirming the clinical value of mobile technology in PD monitoring.

Beyond the interpretation of statistical findings, this study contributes meaningfully toward broader clinical goals in the management of PD. By demonstrating that a smartphone-based assessment can detect time-severity interaction effects in both fine motor and vocal symptoms, this study supports the feasibility of dynamic, real-world monitoring of disease progression. Furthermore, the ability to capture subtle, early-onset symptom fluctuations using accessible digital tools aligns with the long-term objective of improving diagnostic accuracy, particularly in the early or prodromal stages of the disease. Ultimately, by facilitating more personalized, timely, and continuous monitoring, this approach holds the potential to enhance patient care through earlier interventions and more responsive treatment strategies.

Limitations

This study has several limitations. First, the healthy adult group and the persons with PD group were not adequately matched for age and sex, and the demographics of the control group did not accurately represent those of the general population with PD. Future studies should ensure better age-sex matching. However, it is important to note that the significant results observed in both the group effect and within-group time-disease severity analyses were obtained regardless of this mismatch. Second, although smartphone monitoring is useful for collecting real-world data, in this study, data collection through the app was conducted in a controlled research setting rather than in a natural environment. Future research should explore the use of

XSL•FO
RenderX

smartphone apps in home settings to better capture real-world data in persons with PD. Finally, the findings of this study warrant large-scale research to evaluate the diagnostic accuracy and reliability of smartphone apps. While this study focused only on finger-tapping and voice characteristics, future research should include other variables such as gait functions, blinking, and natural language analysis.

Conclusions

This study highlights the potential of smartphone-based time-lapse assessment of symptoms to discern disease severity in persons with PD. With only a few minutes of dynamic assessment, this approach can detect subtle temporal fluctuations in dysrhythmia and vocal control deficits, providing novel insights into the severity of PD.

Acknowledgments

We would like to thank Gyumin Lee for mobile application development, Heewon Min for partial contribution to voice analysis, and Hyeonsu Kim for mobile application design.

This research has been conducted by the Excellent researcher support project of Kwangwoon University in 2023 and was supported by a grant (03-2022-2210) from the Seoul National University Hospital Research Fund.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the sensitive nature of human data (eg, age, sex, and diagnosis) but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: YM (equal), SJY (equal), KP (equal), HGS (equal)
Data curation: YM (equal), SJY (equal)
Formal analysis: YM (equal), SJY (equal)
Funding acquisition: HGS
Investigation: HGS (lead), SJY (supporting)
Methodology: YM (equal), SJY (equal), BMO (supporting)
Project administration: HGS (lead), KP (supporting)
Resources: HGS (equal), KP (equal)
Supervision: HGS (lead), BMO (supporting)
Visualization: YM (equal), SJY (equal)
Validation: HGS
Writing - original draft: YM (lead), SJY (supporting), KP (supporting)
Writing - review & editing: YM (equal), SJY (equal), KP (equal), BMO (equal), HGS (equal)

Conflicts of Interest

None declared.

Multimedia Appendix 1
ANOVA and *post-hoc* analyses of voice characteristics and finger-tapping results.
[DOCX File, 27 KB - mhealth_v13i1e69028_app1.docx]

References

1. Waninger S, Berka C, Stevanovic Karic M, et al. Neurophysiological biomarkers of Parkinson's disease. J Parkinsons Dis 2020;10(2):471-480. [doi: 10.3233/JPD-191844] [Medline: 32116262]
2. Gan-Or Z. Clinical genetic testing in Parkinson's disease should become part of routine patient care. Brain (Bacau) 2024 Aug 1;147(8):2595-2597. [doi: 10.1093/brain/awae181] [Medline: 39087915]
3. Klein MO, Battagello DS, Cardoso AR, Hauser DN, Bittencourt JC, Correa RG. Dopamine: functions, signaling, and association with neurological diseases. Cell Mol Neurobiol 2019 Jan;39(1):31-59. [doi: 10.1007/s10571-018-0632-3] [Medline: 30446950]
4. Micieli G, Tosi P, Marcheselli S, Cavallini A. Autonomic dysfunction in Parkinson's disease. Neurol Sci 2003 May;24 Suppl 1:S32-S34. [doi: 10.1007/s100720300035] [Medline: 12774210]
5. Samizadeh MA, Fallah H, Toomarisahzabi M, et al. Parkinson's disease: a narrative review on potential molecular mechanisms of sleep disturbances, REM behavior disorder, and melatonin. Brain Sci 2023 Jun 6;13(6):914. [doi: 10.3390/brainsci13060914] [Medline: 37371392]

XSL•FO
RenderX

6. Hussein A, Guevara CA, Del Valle P, Gupta S, Benson DL, Huntley GW. Non-motor symptoms of Parkinson's disease: the neurobiology of early psychiatric and cognitive dysfunction. Neuroscientist 2023 Feb;29(1):97-116. [doi: 10.1177/10738584211011979] [Medline: 33966533]
7. Corrà MF, Vila-Chã N, Sardoeira A, et al. Peripheral neuropathy in Parkinson's disease: prevalence and functional impact on gait and balance. Brain (Bacau) 2023 Jan 5;146(1):225-236. [doi: 10.1093/brain/awac026] [Medline: 35088837]
8. Omotosho AO, Tajudeen YA, Oladipo HJ, et al. Parkinson's disease: are gut microbes involved? Brain Behav 2023 Aug;13(8):e3130. [doi: 10.1002/brb3.3130] [Medline: 37340511]
9. Zhou ZD, Yi LX, Wang DQ, Lim TM, Tan EK. Role of dopamine in the pathophysiology of Parkinson's disease. Transl Neurodegener 2023 Sep 18;12(1):44. [doi: 10.1186/s40035-023-00378-6] [Medline: 37718439]
10. Xiu N, Li W, Liu L, et al. A study on voice measures in patients with Parkinson's disease. J Voice 2024 Jun 17:S0892-1997(24)00168-1. [doi: 10.1016/j.jvoice.2024.05.018] [Medline: 38890016]
11. Li J, Zhu H, Wang H, et al. A three-dimensional finger-tapping framework for recognition of patients with mild Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng 2023;31:3331-3340. [doi: 10.1109/TNSRE.2023.3296883] [Medline: 37494164]
12. Yu T, Park KW, McKeown MJ, Wang ZJ. Clinically informed automated assessment of finger tapping videos in Parkinson's disease. Sensors (Basel) 2023 Nov 13;23(22):9149. [doi: 10.3390/s23229149] [Medline: 38005535]
13. Lee CY, Kang SJ, Hong SK, Ma HI, Lee U, Kim YJ. A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in Parkinson's disease. PLoS One 2016;11(7):e0158852. [doi: 10.1371/journal.pone.0158852] [Medline: 27467066]
14. Marsili L, Abanto J, Mahajan A, et al. Dysrhythmia as a prominent feature of Parkinson's disease: an app-based tapping test. J Neurol Sci 2024 Aug 15;463:123144. [doi: 10.1016/j.jns.2024.123144] [Medline: 39033737]
15. Tolleson CM, Dobolyi DG, Roman OC, et al. Dysrhythmia of timed movements in Parkinson's disease and freezing of gait. Brain Res 2015 Oct 22;1624:222-231. [doi: 10.1016/j.brainres.2015.07.041] [Medline: 26241766]
16. Lee LHN, Huang CS, Chuang HH, et al. An electrophysiological perspective on Parkinson's disease: symptomatic pathogenesis and therapeutic approaches. J Biomed Sci 2021 Dec 9;28(1):85. [doi: 10.1186/s12929-021-00781-z] [Medline: 34886870]
17. Ehsan MT, Pranto SI, Mamun KA. Real-time screening of Parkinson's disease based on speech analysis using smartphone. Presented at: 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER); May 4-6, 2021. [doi: 10.1109/NER49283.2021.9441145]
18. Kang K, Nunes AS, Sharma M, et al. Utilizing speech analysis to differentiate progressive supranuclear palsy from Parkinson's disease. Parkinsonism Relat Disord 2023 Oct;115:105835. [doi: 10.1016/j.parkreldis.2023.105835] [Medline: 37678101]
19. Omberg L, Chaibub Neto E, Perumal TM, et al. Remote smartphone monitoring of Parkinson's disease and individual response to therapy. Nat Biotechnol 2022 Apr;40(4):480-487. [doi: 10.1038/s41587-021-00974-9] [Medline: 34373643]
20. Adams JL, Kangarloo T, Tracey B, et al. Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study. NPJ Parkinsons Dis 2023 Apr 17;9(1):64. [doi: 10.1038/s41531-023-00497-x] [Medline: 37069193]
21. Reichmann H, Klingelhoefer L, Bendig J. The use of wearables for the diagnosis and treatment of Parkinson's disease. J Neural Transm (Vienna) 2023 Jun;130(6):783-791. [doi: 10.1007/s00702-022-02575-5] [Medline: 36609737]
22. Carrón J, Campos-Roca Y, Madruga M, Pérez CJ. A mobile-assisted voice condition analysis system for Parkinson's disease: assessment of usability conditions. Biomed Eng Online 2021 Nov 21;20(1):114. [doi: 10.1186/s12938-021-00951-y] [Medline: 34802448]
23. Rusz J, Hlavnicka J, Tykalova T, et al. Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng 2018 Aug;26(8):1495-1507. [doi: 10.1109/TNSRE.2018.2851787] [Medline: 29994713]
24. Singh S, Xu W. Robust detection of Parkinson's disease using harvested smartphone voice data: a telemedicine approach. Telemed J E Health 2020 Mar;26(3):327-334. [doi: 10.1089/tmj.2018.0271] [Medline: 31033397]
25. Kim JI, Sunwoo MK, Sohn YH, Lee PH, Hong JY. The MMSE and MoCA for screening cognitive impairment in less educated patients with Parkinson's disease. J Mov Disord 2016 Sep;9(3):152-159. [doi: 10.14802/jmd.16020] [Medline: 27667187]
26. Goetz CG, Tilley BC, Shaftman SR, et al. Movement disorder society-sponsored revision of the unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008 Nov 15;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
27. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
28. Yun YS, Kim H, Son YI, Choi HS. Validation of the Korean Voice Handicap Index (K-VHI) and the clinical usefulness of Korean VHI-10. Commun Sci Disord 2008;13:216-241 [FREE Full text]

XSL•FO
RenderX

29. Eyben F, Wöllmer M, Schuller B, editors. Opensmile: the Munich versatile and fast open-source audio feature extractor. Presented at: Proceedings of the 18th ACM international conference on Multimedia (MM '10); Oct 25-29, 2010. [doi: 10.1145/1873951.1874246]
30. Eyben F, Scherer KR, Schuller BW, et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. IEEE Trans Affective Comput 2015;7(2):190-202. [doi: 10.1109/TAFFC.2015.2457417]
31. St»hle L, Wold S. Analysis of variance (ANOVA). Chemometr Intell Lab Syst 1989 Nov;6(4):259-272. [doi: 10.1016/0169-7439(89)80095-4]
32. Bolker BM. Linear and generalized linear mixed models. Ecol Stat Contemp Theor Appl 2015:309-333. [doi: 10.1093/acprof:oso/9780199672547.003.0014]
33. Schalkamp AK, Peall KJ, Harrison NA, Sandor C. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. Nat Med 2023 Aug;29(8):2048-2056. [doi: 10.1038/s41591-023-02440-2] [Medline: 37400639]
34. Leite Silva ABR, Gonçalves de Oliveira RW, Diógenes GP, et al. Premotor, nonmotor and motor symptoms of Parkinson's disease: a new clinical state of the art. Ageing Res Rev 2023 Feb;84:101834. [doi: 10.1016/j.arr.2022.101834] [Medline: 36581178]
35. Watts CR, Zhang Y. Progression of self-perceived speech and swallowing impairment in early stage Parkinson's disease: longitudinal analysis of the Unified Parkinson's Disease Rating Scale. J Speech Lang Hear Res 2022 Jan 12;65(1):146-158. [doi: 10.1044/2021_JSLHR-21-00216] [Medline: 34851686]
36. Kuresan H, Samiappan D, Ghosh S, Gupta AS. Early diagnosis of Parkinson's disease based on non-motor symptoms: a descriptive and factor analysis. J Ambient Intell Human Comput 2021:1-15. [doi: 10.1007/s12652-021-02944-0]
37. Marino BLB, de Souza LR, Sousa KPA, et al. Parkinson's disease: a review from pathophysiology to treatment. Mini Rev Med Chem 2020;20(9):754-767. [doi: 10.2174/1389557519666191104110908] [Medline: 31686637]

Abbreviations

H&Y: Hoehn and Yahr
K-MoCA: Korean Montreal Cognitive Assessment
LMM: linear mixed model
MDS: Movement Disorder Society
PD: Parkinson disease
TF: time frame
UPDRS: Unified Parkinson's Disease Rating Scale
VHI-10: Voice Handicap Index-10

Edited by L Buis; submitted 20.11.24; peer-reviewed by ADO Andrade, J Lee, M Munikumar; revised version received 30.06.25; accepted 25.09.25; published 13.11.25.

Please cite as:
Myong Y, Yun SJ, Park K, Oh BM, Seo HG
Dynamic Assessment of Fine Motor Control and Vocalization in Parkinson Disease Through a Smartphone App: Cross-Sectional Study of Time-Severity Interaction Effects
JMIR Mhealth Uhealth 2025;13:e69028
URL: https://mhealth.jmir.org/2025/1/e69028
doi:10.2196/69028

© Youho Myong, Seo Jung Yun, Kyudong Park, Byung-Mo Oh, Han Gil Seo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Associations Between Daily Symptoms and Pain Flares in Rheumatoid Arthritis: Case-Crossover mHealth Study

Ting-Chen Chloe Hsu[1], PhD; Belay B Yimer[1], PhD; Pauline Whelan[2], PhD; Christopher J Armitage[3,4], PhD; Katie Druce[1], PhD; John McBeth[1,5,6], PhD

[1]Centre for Musculoskeletal Research, The University of Manchester, Oxford Road, Manchester, United Kingdom

[2]Centre for Health Informatics, Division of Informatics, Imaging & Data Sciences, The University of Manchester, Manchester, United Kingdom

[3]Manchester Centre for Health Psychology, The University of Manchester, Manchester, United Kingdom

[4]NIHR Greater Manchester Patient Safety Research Collaboration, The University of Manchester, Manchester, United Kingdom

[5]NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom

[6]School of Primary Care, Population Sciences and Medical Education, University of Southampton, Aldermoor Surgery, Aldermoor Close, Southampton, United Kingdom

Corresponding Author:
John McBeth, PhD
Centre for Musculoskeletal Research, The University of Manchester, Oxford Road, Manchester, United Kingdom

Abstract

Background: Mobile health (mHealth) technologies, such as smartphones and wearables, enable continuous assessments of individual health information. In chronic musculoskeletal conditions, pain flares, defined as periods of increased pain severity, often coincide with worsening disease activity and cause significant impacts on physical and emotional well-being. Using mHealth technologies can provide insights into individual pain patterns and associated factors.

Objective: This study aims to characterize pain flares and identify associated factors in rheumatoid arthritis (RA) by (1) describing the frequency and duration of pain flares using progressively stringent definitions based on pain severity, and (2) exploring associations between pain flares and temporal changes in symptoms across emotional, cognitive, and behavioral domains.

Methods: Our 30-day mHealth study collected daily pain severity and related symptoms (scores 1-5, higher are worse) via a smartphone app and passively recorded sleep and physical activity via a wrist-worn accelerometer. Pain flares were defined using a 5-point scale: (1) above average (AA): pain severity > personal median, (2) above threshold (AT): pain severity > 3, and (3) move above threshold (MAT): pain severity moves from 1, 2, 3 to 4 or 5. A case-crossover analysis compared within-person variations of daily symptoms across hazard (3 days before a pain flare) and control (3 days not preceding a pain flare) periods using mean and intraindividual standard deviation. Conditional logistic regression estimated the odds ratio (OR) for pain flare occurrence.

Results: A total of 195 participants (160/195, 82.1% females; mean age 57.2 years; average years with RA: 11.3) contributed 5290 days of data. Of these, 88.7% (173/195) experienced at least 1 AA flare (median monthly rate 4, IQR 2.1-5). Nearly half experienced at least 1 AT or MAT flare (median monthly rate 2, IQR 1-4). These pain flares lasted 2 days (IQR 2-3) on average across definitions, with some extending up to 12 days. Worsening mood over 3 days was associated with a 2-fold increase in the likelihood of AT flares the following day (OR 2.04, IQR 1.06-3.94; $P<.05$). Greater variability in anxiety over 3 days increased the likelihood of both AT (OR 1.67, IQR 1.01-2.78; $P<.05$) and MAT flares (OR 1.82, IQR 1.08-3.07; $P<.05$). Similarly, greater variability in sleepiness (OR 1.89, IQR 1.03-3.47; $P<.05$) also increased the likelihood of AT flares. Sedentary time (%) consistently showed almost no influence across all definitions. Similarly, the simplest definition of AA demonstrated no significant associations across all symptoms.

Conclusions: Pain flares were commonly observed in RA. Changes in sleep patterns and emotional distress were associated with pain flare occurrences. This analysis demonstrates the potential of daily mHealth data to identify pain flares, opening opportunities for timely monitoring and personalized management.

(JMIR Mhealth Uhealth 2025;13:e64889) doi:10.2196/64889

KEYWORDS

pain flares; mHealth; patient-generated data; rheumatoid arthritis; musculoskeletal; quality of life; smartphone; mobile health; wearable; sleep; physical activity

XSL•FO
RenderX

Introduction

Digital technology plays a crucial role in modern medicine. It not only improves the essential routines and workflows of health care but also helps minimize accessibility barriers and time and location constraints, as well as streamlining health data management. As digital technology advances and becomes widely adopted, it is transforming our approach to addressing individual health needs by moving beyond traditional self-report methods and enabling more dynamic and continuous assessments. We can now collect patient information from multiple sources with greater time-varying details over extended periods. Mobile health (mHealth) devices, such as smartphones and wearables, allow patients to log symptoms actively through numerical scales or visual tools [1-5]. They can also passively capture high-resolution streams of health-related and contextual information, such as sleep, physical activity, or weather with minimal burden [3,5]. This enhanced capability of characterizing temporal changes in symptom patterns can deepen our understanding of individual health trajectories at a more granular level and open opportunities to improve disease management through more personalized strategies.

Individuals with chronic musculoskeletal conditions, such as rheumatoid arthritis (RA), typically experience persistent pain that requires continuous treatment and long-term management. Daily fluctuations in pain are common [6,7]. While not all fluctuations require clinical attention, periods of significantly increased pain severity, known as pain flares, have a greater impact and present clinical challenges due to their variability and unpredictability [8-10]. Pain flares often occur concurrently with escalated disease activity [11-13], but not every occurrence involves inflamed disease activity, nor does inflamed disease activity always result in escalated pain [11]. Additionally, pain flares exhibit considerable heterogeneity both within and across individuals and are influenced by a complex interplay of biopsychosocial, behavioral, and environmental factors [3,14-17]. For example, depression was identified as a significant contributing factor for more severe pain and greater pain fluctuations in rheumatic diseases [16]. Longer time spent in bed and more severe fatigue were observed to heighten the risk of a pain flare in low back pain, characterized by a 2-point increase over the average pain [18,19]. These factors collectively contribute to a multitude of impacts, including functional disability, cognitive decline, emotional distress, diminished social interactions, and a lower quality of life [20,21].

Several studies on chronic pain have demonstrated the acceptability, feasibility, and usability of smartphone apps and wearables for remotely monitoring pain severity, disease activity, and associated factors such as sleep patterns, physical activity, mood, and weather conditions [1,3-5]. They have identified specific associations that influence fluctuations in pain severity. For example, certain weather conditions, including higher relative humidity and wind speed and lower atmospheric pressure, were associated with increased pain severity [3]. Worsening symptoms in the preceding week, characterized by higher mean scores and steeper slopes in physical activity, fatigue, sleep difficulty, physical and emotional well-being, and coping ability, were associated with a greater likelihood of self-reported flares [4]. In our exploratory study, we advanced previous research by leveraging daily patient-generated health data captured by mHealth devices to characterize pain flares and identify associated factors in a RA cohort. Our objectives included (1) describing the frequency and duration of pain flares using progressively stringent definitions based on self-reported pain severity, and (2) exploring associations between pain flares and temporal changes in symptoms across emotional, cognitive, and behavioral domains.

Methods

mHealth Cohort Study in RA

This secondary analysis used data from a prospective mHealth cohort study that investigated quality of life, sleep, and rheumatoid arthritis (QUASAR) [5]. The QUASAR study recruited 285 participants from the National Rheumatoid Arthritis Society (a UK-wide patient organization) between May 2017 and July 2018. Eligible participants were aged 18 years and older, with a self-reported clinical diagnosis of RA and receiving disease-modifying antirheumatic drugs [22], had access to an Android or iOS smartphone or tablet, and were not employed in shift work. Participants completed a paper-based baseline questionnaire including demographics, information about RA, and their health status. Over the 30-day study period, participants wore a validated triaxial accelerometer (MotionWatch 8 by CamNtech [23-25]) to measure sleep and physical activity. They also used a co-designed smartphone app (uMotif) to complete a daily sleep diary (The 9-item Consensus Sleep Diary, CSD [26]) and provide 10 daily symptoms (pain severity, illness impact, fatigue, mood, well-being, anxiety, disease control, challenge, sleepiness, and concentration) twice a day on a 5-point ordinal scale, once in the morning and once in the evening. Follow-up questionnaires were completed every 10 days after the baseline. A detailed description of the QUASAR study can be found in the protocol [27].

Data Preparation

Participants

A total of 266 participants gave consent for their data to be used for secondary purposes. To be included in the analysis, participants were required to (1) provide symptom data via the study app, (2) provide pain severity data for at least 7 consecutive days, and (3) achieve a completion rate of ≥70% for pain severity data.

Baseline Assessments

Demographic data included sex (male or female), age, ethnicity (White or non-White), marital status (single, married or with partner, or separated), occupations (employed, retired, voluntary, or seeking work), smoking status (never, past, or current), and average weekly alcohol consumption (none, moderate: 1 - 15 units, and heavy: ≥16 units).

Information about RA included disease duration, menopausal status, other associated rheumatic diseases (osteoarthritis, spondyloarthropathy or ankylosing spondylitis, fibromyalgia or chronic widespread pain, gout or other crystal arthritis, Sjögren syndrome, thyroid disorder, diabetes, multiple sclerosis,

XSL•FO
RenderX

and hypertension), sleep-related problems (restless leg syndrome and obstructive sleep apnea or snoring), current medications (sleep medicine and pain medicine), and disease activity, measured by Routine Assessment of Patient Index Data 3 (RAPID-3, score range: 0 - 30) [28]. Disease activity was categorized into near remission (≤3), low severity (3.1 - 6), moderate severity (6.1 - 12), and high severity (≥12).

Health status information consisted of sleep quality, assessed by the Pittsburgh Sleep Quality Index (score range 0 - 21, score of >5 indicates poor sleep quality) [29]; insomnia, evaluated by the Sleep Condition Indicator (score range 0 - 32, score of ≤16 indicates probable insomnia) [30]; and anxiety and depression, assessed using the Hospital Anxiety and Depression Scale (HADS, score range 0 - 21) [31]. HADS was categorized into normal (0 - 7), borderline case (8-10), and case (11-21).

Pain Severity and Illness Impact

Pain severity and illness impact were scored from 1 (no pain/impact) to 5 (very severe pain/impact) [32]. Higher scores represent greater severity. We used records submitted after 12 PM for data consistency and completeness. For multiple entries within the same day, we selected the final record.

Exposure Data

Exposure data consisted of 12 items, including 9 daily self-reported symptoms and 3 objective measurements. All self-reported scores were adjusted to follow the same direction, with higher scores being worse (score range 1 - 5). Similar to pain and impact data, only 1 record after 12 PM was included for each day. These daily self-reported symptoms were fatigue (no fatigue to very severe fatigue), mood (very happy to depressed), well-being (very well to very unwell), anxiety (very well to very anxious), disease control (very good control to no control), challenge (not challenging to severely challenging), sleepiness (not sleepy to very sleepy), concentration (excellent to poor), and sleep quality (very good to very poor). Sleep quality was measured by the CSD.

Objective measurements obtained from accelerometer data [33,34] included sleep efficiency (%), sedentary time (%), and time in bed (hour). Sleep efficiency was calculated as the ratio of actual sleep time to total time spent in bed. Sedentary time was defined as the ratio of time spent being sedentary during waking hours. The MotionWatch 8 was configured to sleep mode, capturing limb or bodily movements in 30-second epochs. The raw accelerometer data were processed with MotionWare, an actigraphy software developed by CamNtech. Physical activity during waking hours was calculated by excluding total sleep time and categorizing into 3 segments: sedentary, low, and moderate to vigorous behavior. The cut point of the MotionWatch 8 for sedentary behavior was ≤178.50 counts per minute, while moderate to vigorous activity was set at ≥562.50 counts per minute [35]. Any activity level between these boundaries was recorded as low activity.

Pain Flare Cycle

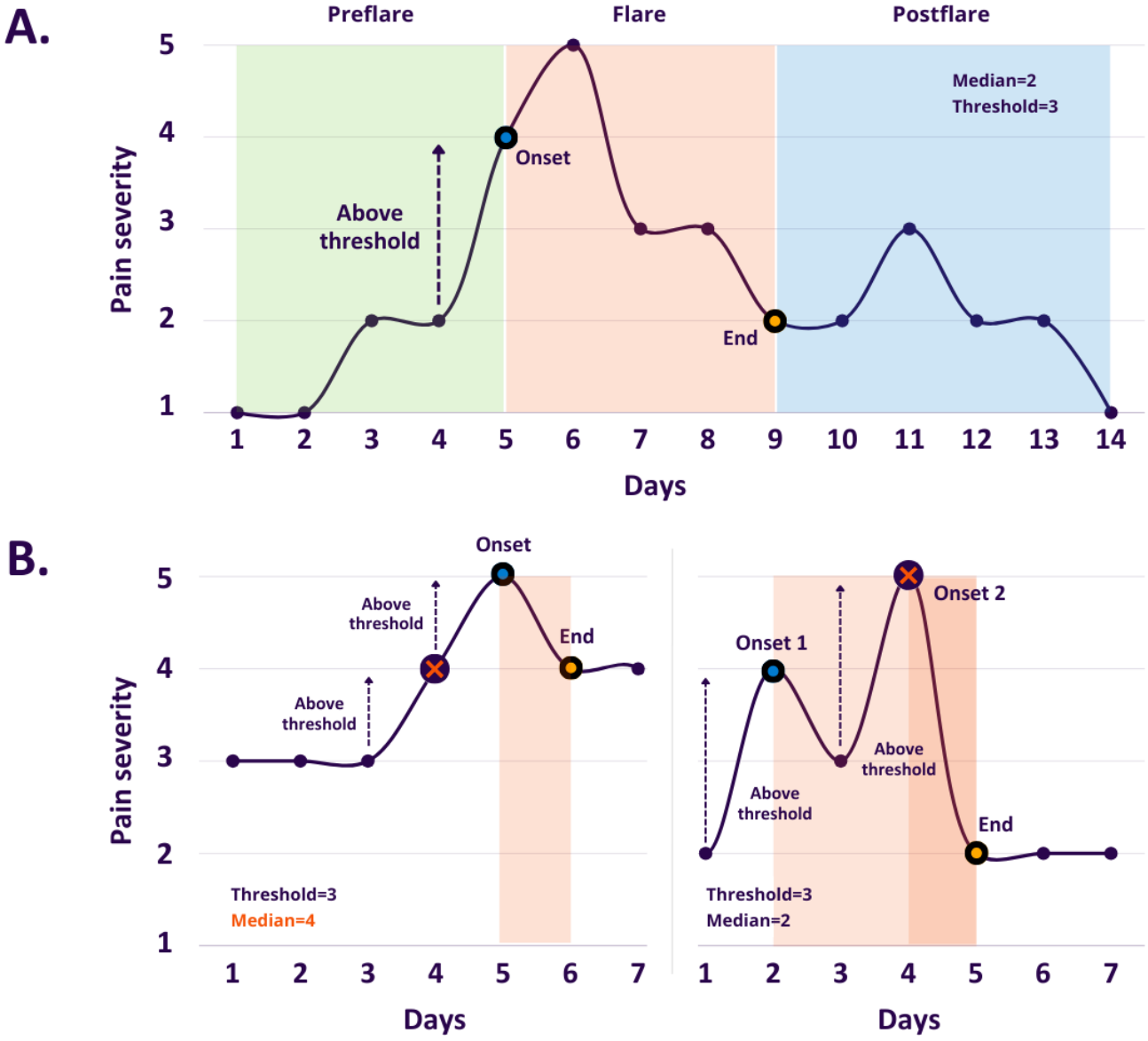
A pain flare cycle was divided into 3 phases: preflare, flare, and postflare. The flare was the interval from when pain severity increased until it subsided. Preflare and postflare represented the period of time before and after the flare, respectively.

The onset of a pain flare was defined using a 5-point scale [32,36], including (1) above average (AA): pain severity greater than personal median, (2) above threshold (AT): pain severity greater than 3, and (3) move above threshold (MAT): pain severity moves from 1, 2, or 3 yesterday to either 4 or 5 today. These definitions capture increasingly complex movements in pain levels and demonstrate at least a 25% change, meeting the minimum clinical importance difference [37].

When pain flares occur over multiple consecutive days, the first day of the sequence was identified as the onset of that flare. The end of a pain flare was marked by pain severity returning to the personal median or lower. All flare onsets must occur above the personal median pain scores to avoid conflicts with the criteria defining the end of a flare. When multiple flares end on the same date because of recurring onsets without any return in between, we remove onsets that occurred after the first one to avoid double counting. Figure 1 shows examples of pain flare identification.

XSL•FO
RenderX

Figure 1. Pain flare identification. The figure shows how pain flares are identified using the definition of above threshold. (A) Pain flare cycle with a median pain score of 2. It includes preflare (green), flare (red), and postflare (blue) phases. The duration of the flare (5 days) is marked between the onset (blue dot) and end (orange dot). (B) Two scenarios of exclusion. Left: the onset (day 4) meets the definition, but the pain severity is at the personal median pain level, conflicting with the criterion for the end of a pain flare. Right: 2 flare onsets (days 2 and 4) meet the definition but end on the same day (day 5). The second onset (day 4) is removed to avoid double counting.



Pain Flare Impact

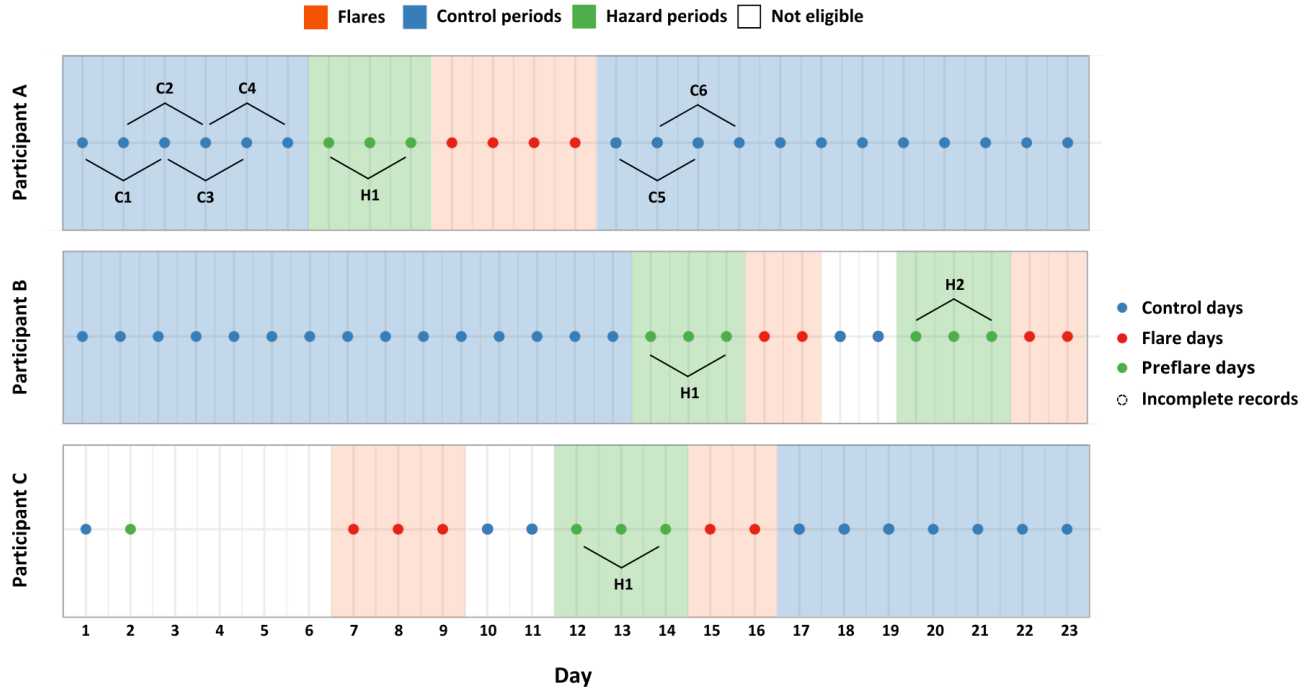
In addition to identifying pain flares by their severity, we examined their residual impact in the postflare phase. A carry-on effect was identified for those pain flares where, although pain severity had subsided, the overall illness impact remained elevated above personal median level. The measurement of carry-on effect began from the end of the flare and continued until the illness impact returned to the individual's median level or below.

Hazard and Control Periods

Hazard periods were selected as the 3 days leading up to each pain flare (preflare phase), excluding any days overlapping with other flares. As a result, not every pain flare was included in the analysis, as it might not have 3 nonoverlapping consecutive preflare days. Control periods were selected using a full stratum bidirectional approach [38] as any 3-day interval that neither preceded a pain flare nor overlapped with other flares or hazard periods, thereby fulfilling the exchangeability assumption for the case-crossover study design [39]. This ensured that there was no overlap between hazard and control periods, and that control periods served as referent windows that were sufficiently recent for comparison and could potentially have been pain flares [40]. Additionally, eligible hazard and control periods must have had complete data for 12 exposures. By using this design, participants serve as their own control, eliminating time-invariant confounders (eg, age, sex, and diagnosis). Each participant with at least 1 hazard and 1 control period formed a risk set. Figure 2 demonstrates the identification of hazard and control periods.

XSL•FO
RenderX

Figure 2. Identification of hazard and control periods. The figure illustrates how hazard and control periods are selected and how risk sets are formed. Hazard periods (green) are 3 days prior to flares (red). Control periods (blue) are 3 days that neither preceded flares (red) nor overlapped with hazard periods (green). A risk set is formed by at least 1 hazard period and 1 control period. Hazard periods (H1 and H2) are compared bidirectionally with control periods (eg, C1, C2, and C3). Ineligible days (white) are excluded from the analysis due to incomplete data.



Statistical Methods

Primary Analysis

All analyses were performed using R (version 4.3.1; R Foundation for Statistical Computing). Descriptive analysis was conducted to describe demographics, information about RA, and health status, as well as to characterize pain flares with 3 definitions. For each definition, pain flares were summarized by their frequency, duration, 30-day monthly rate, and the duration of impact.

Conditional logistic regression was used to estimate the odds ratio (OR) for pain flare occurrence. For each pain flare definition, we compared the changes over the 3 days prior to a flare (hazard) with the 3 days that did not precede a flare (control), across a total of 12 exposures. Mean and intraindividual standard deviation (iSD) were used to describe hazard and control periods [41]. The mean represents the average level of each exposure over 3 days. The iSD captures the magnitude of fluctuation over 3 days in each exposure, with higher values indicating increased day-to-day variability. This approach allowed us to determine whether variations in these exposures were associated with pain flares occurring.

Regression models of each definition were analyzed independently. The modeling consisted of 2 steps. First, each of the 12 exposures was analyzed separately using 2 metrics (mean and iSD) in univariable models. This resulted in 24 univariable models, with 2 metrics for each exposure. Subsequently, both metrics for each of the 12 exposures were included in multivariable models, resulting in 12 models. All models were fit using robust sandwich standard errors to account for within-participant correlation due to multiple hazard periods. The findings were presented as ORs with 95% CIs of the likelihood of a pain flare. A 95% CI that does not include the value of 1 indicates a statistically significant association, equivalent to achieving significance as determined by a *P* value.

Sensitivity Analysis

Sensitivity analysis was performed to assess the time overlap among control periods. In the primary analysis, control periods were selected as any 3-day interval not preceding a pain flare, during which time overlap was inherently allowed (Figure 2) and could lead to compounding effects of exposures [40,42]. To address this, we conducted a nonoverlap analysis, where control periods were selected as discrete, sequential 3-day intervals (eg, C1 and C4 in Figure 2). This eliminates time overlaps between control periods, thereby avoiding any compounding exposure effect.

Ethical Considerations

Ethical approval for this secondary analysis was not required. All participants provided consent for secondary use, and all data were fully anonymized. Approval was granted to the QUASAR study [5] by the National Research Ethics Service Committee North West-Liverpool Central Research Ethics Committee (reference: 17/NW/0217).

Results

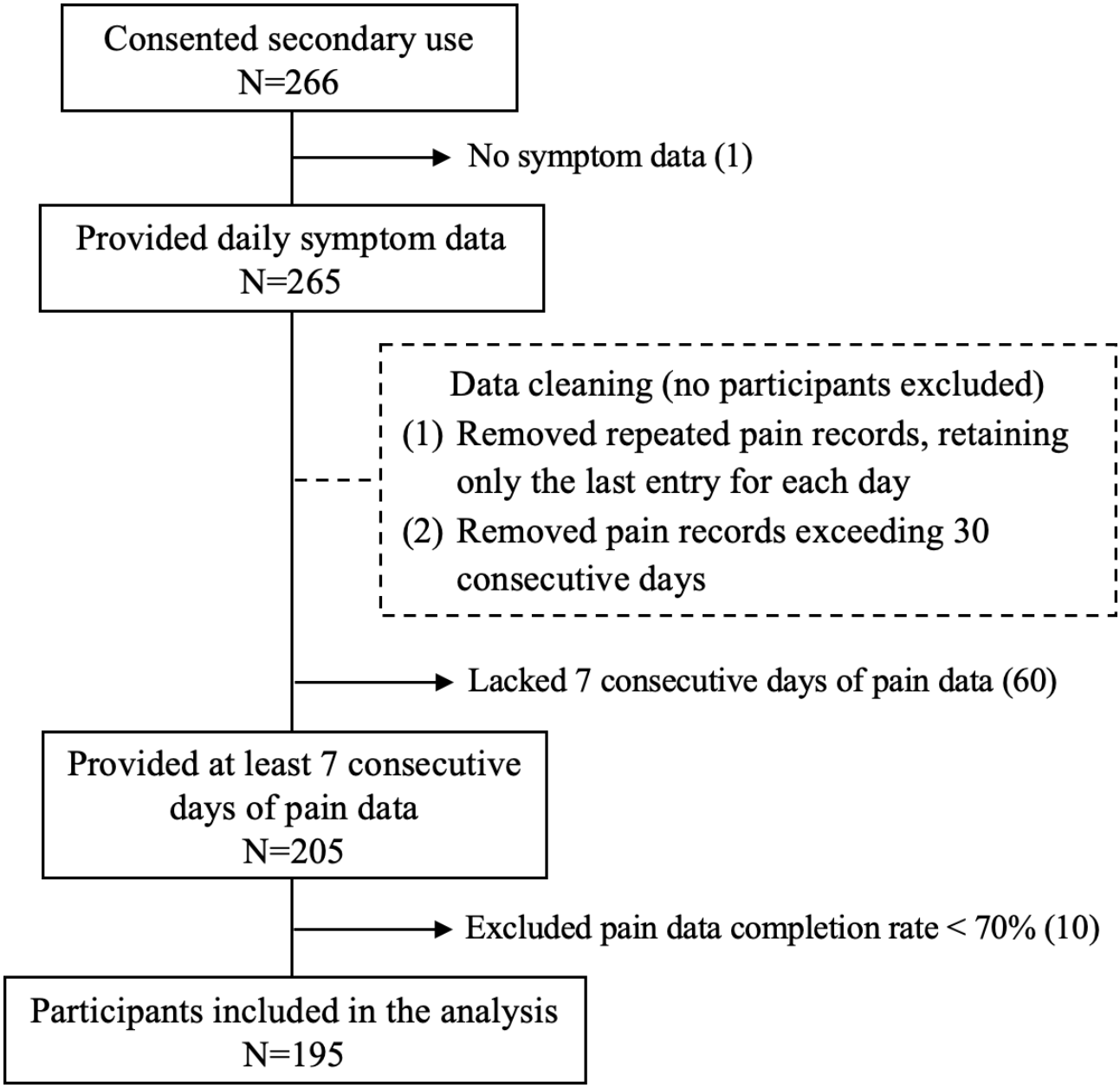
Overview

Of the 266 participants who consented for secondary use, 195 (73.3%) were included in the analysis (mean age 57.2 years; mean duration living with RA 11.3 years; 160/195, 82.1% females; 160/195, 82.1% White ethnicity; 92/195, 47.2% employed full time or part-time; and 148/195, 75.9% married or with partner). These participants contributed a total of 5290 person-days of pain data and 3832 person-days of complete

XSL•FO
RenderX

exposure data. Figure 3 shows the participant inclusion process. The reasons for exclusion were failing to provide symptom data (n=1) and achieve the required completion rate (n=70).

Figure 3. Flow diagram for participant inclusion. The figure shows the participant inclusion process based on three criteria: (1) provides symptom data, (2) provides pain data for at least 7 consecutive days, and (3) achieves a completion rate of ≥70%. The dash-lined box describes the approaches taken to remove duplicates and excessive data, without excluding any participant.



Baseline Health Characteristics

Besides RA, 60.5% (118/195) of the participants had one or more coexisting conditions. The most prevalent comorbid condition in this cohort was osteoarthritis (60/195, 30.8%), followed by hypertension (40/195, 20.5%), thyroid disorder (28/195, 14.4%), fibromyalgia (25/195, 12.8%), and Sjögren syndrome (24/195, 12.3%). Most females (119/160, 74.4%) had experienced menopause. Pain medication use was common (157/195, 80.5%). About one-third were also under medical treatment for sleep issues (54/195, 27.7%). More than half had never smoked (107/195, 54.9%) and were consuming moderate alcohol weekly (103/195, 52.8%).

High severity of RA (RAPID-3 median 12.7, IQR 7.5 - 17.5) was reported in more than half of the cohort (105/195, 53.8%). Baseline pain level was moderate (median 5.5, IQR 3 - 7) on a 11-point scale. Sleep disturbances were common, with 86.7% (169/195) experiencing poor sleep quality (Pittsburgh Sleep Quality Index median 10.3, IQR 7.5 - 13.3) and 37.4% (73/195) having probable insomnia (Sleep Condition Indicator median 19, IQR 13 - 23). Anxiety cases were reported in 34.4% (67/195) of the cohort (HADS anxiety median 9, IQR 7 - 11), whereas 18.5% (36/195) were considered depression cases (HADS depression median 7, IQR 4 - 10). Table S1 Supplementary 1in Multimedia Appendix 1 provides detailed baseline characteristics.

XSL•FO
RenderX

Pain Flare Characteristics

As shown in Table 1, the frequency of pain flares decreased when applying stricter definitions. About 88.7% (173/195) of the participants had at least 1 AA flare, with a median monthly rate of 4 flares (IQR 2.1 - 5). Nearly half of the participants experienced at least 1 AT (49.2%, 96/195) or MAT (45.6%, 89/195) flare, with a median monthly rate of 2 flares (IQR 1 - 4). The average duration of pain flares was consistent across definitions, with a median of 2 days (IQR 2 - 3). However, pain flares could last up to 12 days before returning to personal median level or lower.

Table . Pain flare characteristics.

	Above average	Above threshold	Move above threshold
Participants with ≥1 flare, n (%)	173 (88.7)	96 (49.2)	89 (45.6)
Flare counts, n	662	248	230
Flare duration (days)			
Median (IQR)	2 (2-3)	2 (2-3)	2 (2-3)
Minimum days	2	2	2
Maximum days	12	9	9
30-day monthly rate (flares)			
Median (IQR)	4 (2.1 - 5)	2 (1-4)	2 (1-4)
Minimum flares	1	1	1
Maximum flares	9	8	8

Pain Flare Impact

Analysis of pain flare impact found that fewer than 20% of pain flares across all definitions were affected by carry-on effect (Table 2). In other words, more than 80% of pain flares ended when both their pain severity and illness impact returned to personal median or lower on the same day. Among affected pain flares, a small number were excluded due to either missing impact data or insufficient data time frame. The average duration of these impact-affected pain flares was approximately 4 days but could last up to 12 days. The average duration of impact ranged from as short as 1 day to a week.

Table . Pain flare impact characteristics.

	Above average	Above threshold	Move above threshold
Total flares affected, n (%)	120 (18.1)	35 (14.1)	34 (14.8)
Total flares analyzed, n	111	34	33
Flare duration (days)			
Median (IQR)	4 (3-5)	4 (3-5)	4 (3-5)
Minimum days	3	3	3
Maximum days	12	11	9
Impact duration (days)			
Median (IQR)	1 (1-2)	1.5 (1 - 2.8)	1 (1-2)
Minimum days	1	1	1
Maximum days	7	6	5

Associations Between Daily Symptoms and Pain Flares

For all definitions, about half of the participants in each subgroup had both hazard and control periods for analysis (Supplementary 2Table S2 in Multimedia Appendix 1). As the definitions became stricter, we observed changes in the strength and direction of the relationships between exposures and pain flares. In both univariable and multivariable models, mood demonstrated consistent patterns across all definitions. Mood mean scores were positively associated with pain flare occurrence, whereas iSD scores were negatively associated. Meanwhile, the ORs for sedentary time consistently showed almost no effect, remaining near or at the null value.

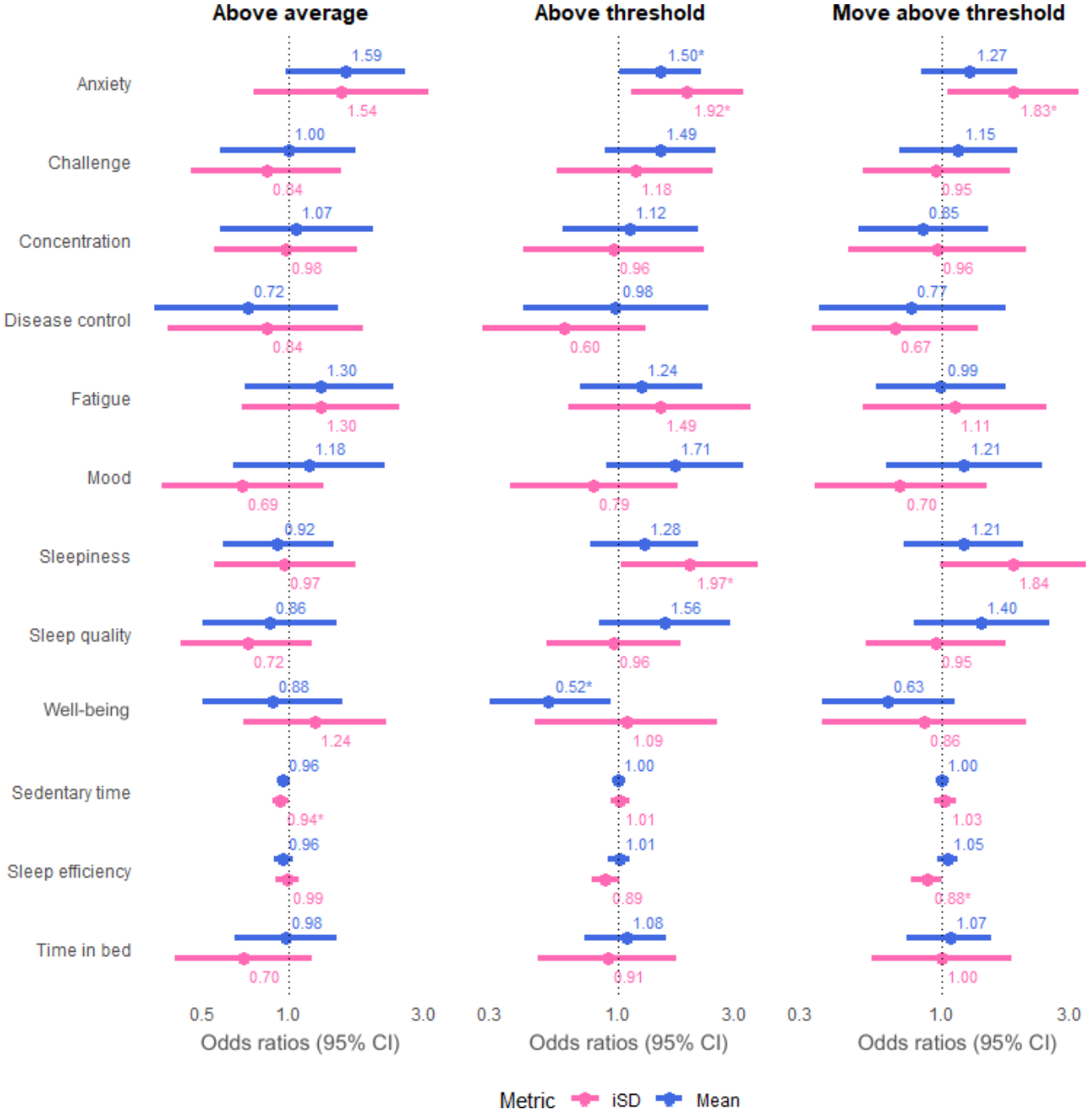
Patient-Reported Exposures

As shown in Figure 4, none of the 9 self-reported symptoms showed a statistically significant relationship with AA flares. However, significant associations were observed for AT and MAT flares. For example, a single unit increase in anxiety mean scores over 3 days was associated with a 1.5 times increased likelihood of an AT flare the following day (OR 1.5, 95% CI 1.02 - 2.2). Higher anxiety iSD scores over 3 days were associated with almost a 2-fold increase in the odds of an AT

XSL•FO
RenderX

flare the next day (OR 1.92, 95% CI 1.13 - 3.27) and a 1.83 times increased likelihood of a MAT flare (OR 1.83, 95% CI 1.05 - 3.21). Likewise, higher sleepiness iSD scores over 3 days nearly doubled the odds of an AT flare the next day (OR 1.97, 95% CI 1.03 - 3.74). Conversely, higher well-being mean scores reduced the odds of an AT flare by about half (OR 0.52, 95% CI 0.3 - 0.93).

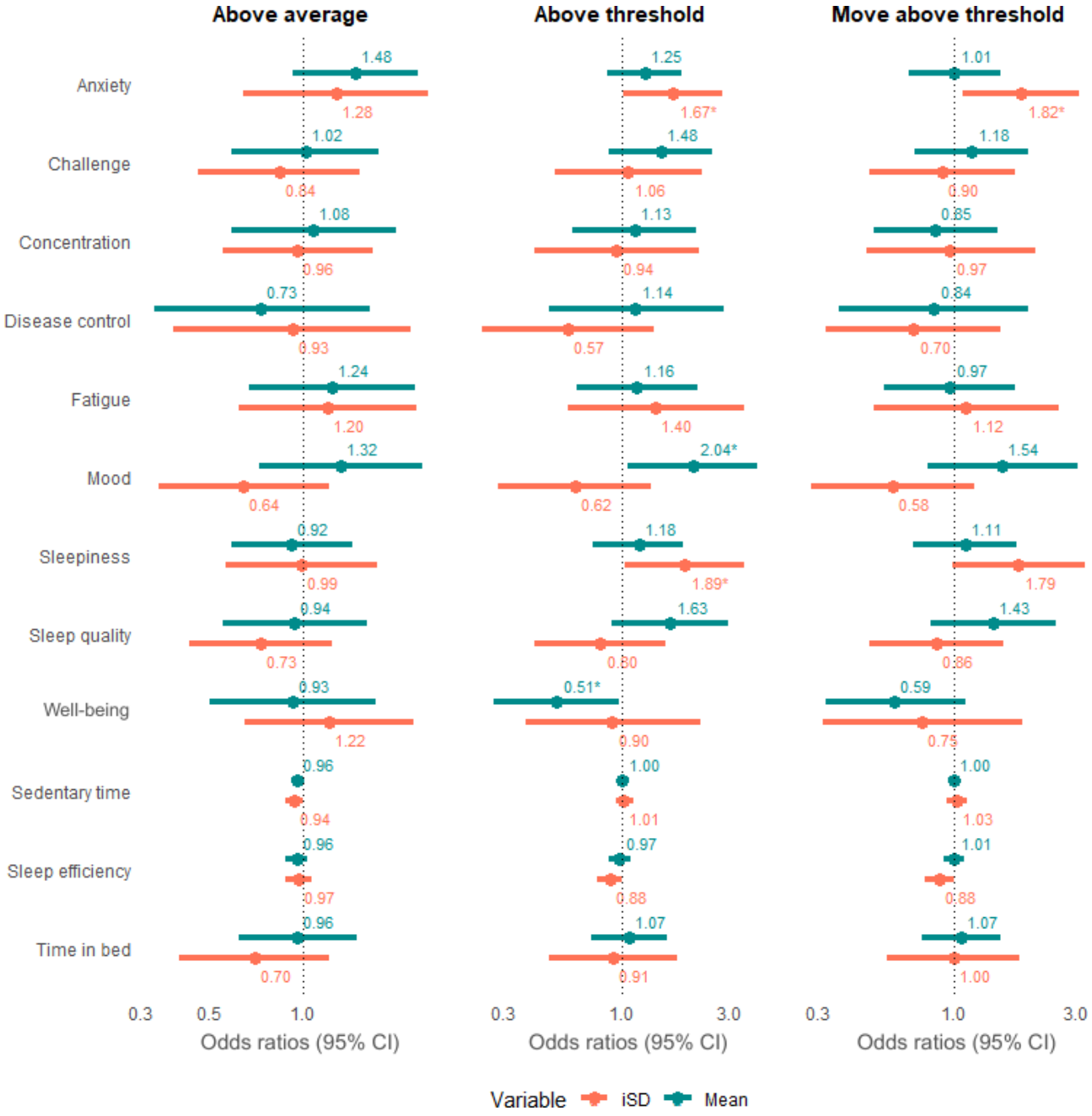
Figure 4. Univariable models across pain flare definitions. The figure shows the results of univariable models across all definitions. Mean (blue) and iSD (pink) were analyzed independently for each of the 12 exposures. *P<.05. iSD: intraindividual standard deviation.



In the multivariable models (Figure 5), higher mood mean scores over 3 days were more clearly associated with AT flares, showing a 2-fold increase in the odds after adjusting for the iSD effect (OR 2.04, 95% CI 1.06 - 3.94). Higher anxiety iSD scores over 3 days were associated with a 1.67 times increased likelihood of an AT flare the next day (OR 1.67, 95% CI 1.01 - 2.78), and a 1.82 times increased likelihood of a MAT flare (OR 1.82, 95% CI 1.08 - 3.07). These associations weakened slightly after adjusting for the mean effect. While anxiety mean scores continued to show a positive relationship with AT flares, this association was no longer significant. Well-being mean scores maintained a negative association with AT flares after adjusting for the iSD effect, reducing the odds by nearly half (OR 0.51, 95% CI 0.27 - 0.97). See full results in Tables S3 and S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 5. Multivariable models across pain flare definitions. The figure shows the results of multivariable models across all definitions. Both mean (green) and iSD (orange) were included in the analysis for each of the 12 exposures. *P<.05. iSD: intraindividual standard deviation.



. .

Accelerometer-Derived Exposures

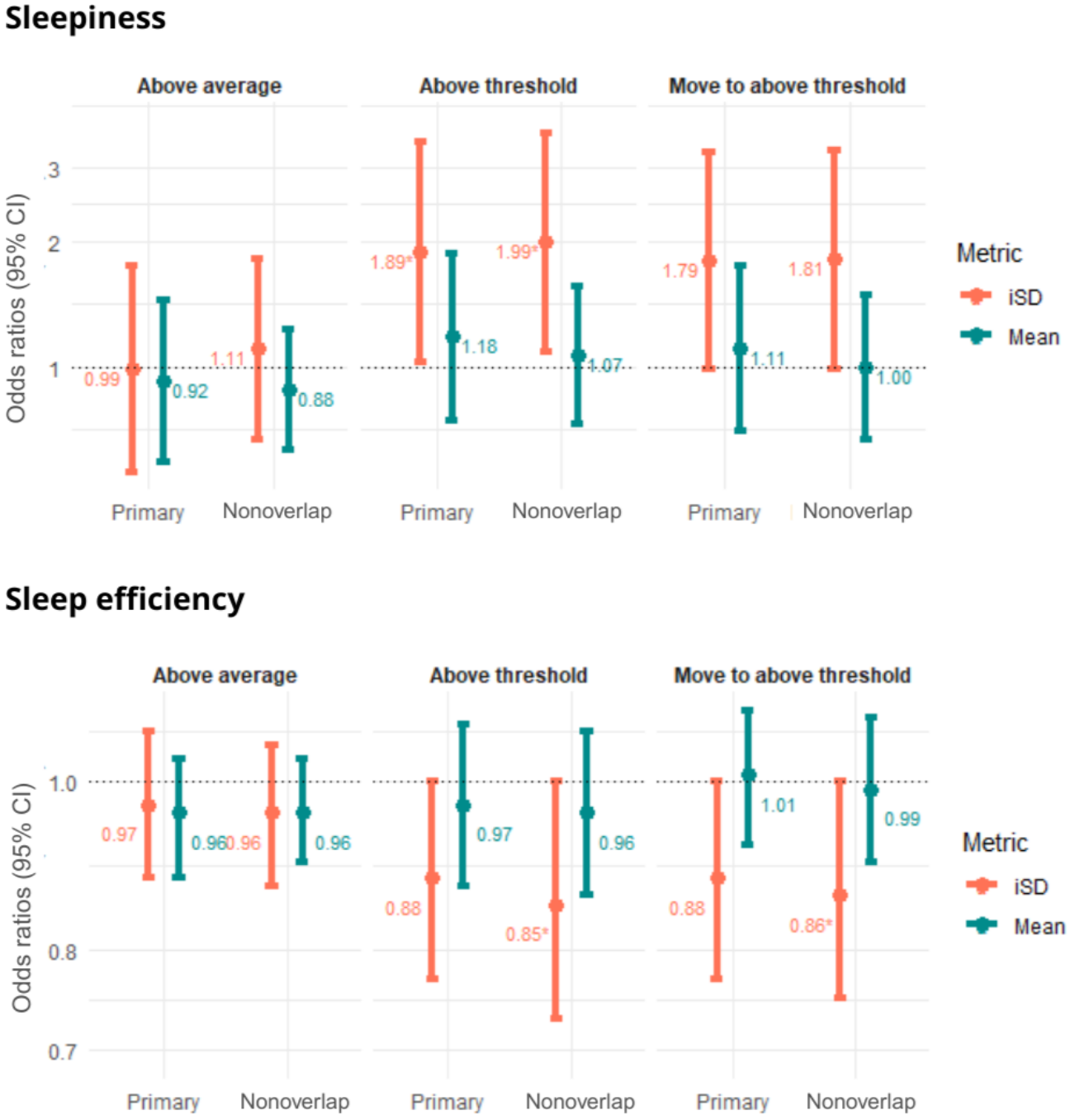
All objective measurements, including sedentary time, sleep efficiency, and time in bed, demonstrated no significant associations with pain flares across definitions in both univariable and multivariable models (Figures 4 and 5). The majority of associations were near or at an OR of 1, even after adjustments were made for the mean and iSD effects. Full results are shown in Tables S3 and S4 Supplementary 3 and 4in Multimedia Appendix 1.

Sensitivity Analysis

The results of the nonoverlap analysis were similar to those of the primary analysis, with most displaying narrower confidence intervals. However, variations in direction and strength of associations were observed. Across primary and nonoverlap analyses, no significant association was observed in AA flares. As shown in Figure 6, higher sleepiness iSD scores over 3 days were consistently associated with a 2-fold increase in the odds of an AT flare the next day, with an OR of 2.04 (95% CI 1.13 - 3.71) in the univariable model and 1.99 (95% CI 1.09 - 3.63) after adjusting for the mean effect.

XSL•FO
RenderX

Figure 6. Sensitivity analysis results for primary and nonoverlap models. The figure shows sensitivity analysis results for primary and nonoverlap models of sleepiness and sleep efficiency. The multivariable model includes mean (green) and iSD (orange). *P<.05. iSD: intraindividual standard deviation.



We observed a negative relationship between well-being mean scores and AT flares (OR 0.58, 95% CI 0.35 - 0.98), but this association did not persist after adjusting for the iSD effect. Similarly, higher sleep efficiency iSD scores reduced the likelihood of a MAT flare the following day (OR 0.87, 95% CI 0.76 - 0.99) but did not persist after adjusting the mean effect. Full results are shown in Figures S1 and S2 in Multimedia Appendix 1.

Discussion

Principal Findings

This exploratory study demonstrates the potential of mHealth technologies to improve our understanding of pain fluctuations in RA. By leveraging daily patient-generated health data, it was possible to describe the frequency and duration of pain flares and identify within-person associated factors. These findings would be difficult to obtain through traditional assessments. However, achieving this level of understanding also relied on sustained patient engagement with the technology, which is challenging due to various barriers such as user fatigue, technology literacy, or motivation [43]. The loss of engagement

XSL•FO
RenderX

can substantially impact the quality of data, leading to biases, misinterpretation, and incorrect conclusions. The QUASAR study, achieving near 90% engagement [5], provides a strong foundation for our analysis. The results led to 3 novel observations regarding pain flares and implications for future pain research using mHealth technologies. Additionally, we discuss limitations that need to be addressed for future improvement.

First, using 3 numerically defined pain flares, we found that the average duration consistently spans 2 days across all definitions. When considering their overall impact, the average duration extends to 4 days for all definitions. This suggests that, regardless of the sudden increase in pain severity, the initial surge typically diminishes within a few days; however, the reduction in pain severity does not necessarily indicate the end of its impact. Qualitative studies have documented the experience of flare-ups as highly individualized and complex [11,20]. Pain flares cannot be fully explained by pain levels alone; their broader impact must also be considered.

Second, emotional distress, as measured by mood and anxiety, is closely associated with pain flares. Feeling more depressed over 3 days appears to increase the likelihood of experiencing a pain flare the following day, consistent with previous research indicating that higher levels of depression are associated with increased pain [16]. Conversely, experiencing a higher degree of mood changes seems to reduce the likelihood of pain flare occurrence. This suggests that varied mood fluctuations could represent normal mood regulation, rather than being in a heightened depressive state. Contrary to previous findings indicating no predictive relationship between anxiety and pain variability [16], our findings show that feeling more anxious and experiencing more fluctuations in anxiety over 3 days increase the likelihood of pain flares.

Finally, patient-reported and accelerometer-derived measurements of sleep demonstrate varied relationships with pain flares. Perceived sleep quality shows no noticeable associations, regardless of the definition used. Perceived sleepiness shows a positive association with pain flares, except under the simplest definition where it has nearly no influence. A positive association was observed between longer time spent in bed and an increased likelihood of pain flares, a pattern similarly noted in low back pain [19]. However, none of these associations were statistically significant. Sleep efficiency demonstrates almost no influence on pain flares across definitions. It is worth noting that self-reported sleep patterns did not strongly correlate with actigraphy-derived sleep patterns in the QUASAR study [5]. Perceived and objective sleep measurements may capture different aspects of sleep issues. For example, perceived sleepiness likely reflects the cumulative impact of long-term poor sleep, whereas sleep efficiency could indicate short-term fluctuations in sleep patterns. The choice between subjective or objective sleep measures needs to be guided by their clinical relevance to pain management and tailored to the individual needs, a process that mHealth technologies can facilitate by integrating data from multiple sources to provide a more holistic view of sleep patterns.

The remaining symptoms, including perceived fatigue, concentration, disease control, and challenge, did not exhibit consistent patterns across definitions. Their temporal changes may not be directly associated with the occurrence of pain flares but might instead be more salient in their impact afterward. Sedentary time demonstrates minimal to no associations with pain flares. A similar pattern was also observed in prior research on low back pain, which found that longer sedentary behavior did not increase the odds of pain flares the next day [19]. An unexpected association was observed for well-being, where feeling more unwell appeared to lower the likelihood of pain flare occurrence. This counterintuitive finding needs further investigation, as changes in emotional state could alter pain management behaviors and potentially influence the pain experience [44].

Study Strengths and Limitations

The study strengths include using a large sample size with more than 15% of participants from diverse ethnic backgrounds, examining across multiple numerically defined pain flares, and employing case-crossover design, which collectively enhance the robustness of capturing within-person variations over time. Several limitations need to be highlighted.

First, the nature of RA as an inflammatory disease may complicate the definition of pain flares. Our approach using pain severity alone may overlook other factors contributing to increased pain, such as inflammation or medication changes. Second, despite using 3 definitions to encompass diverse pain fluctuations, participants with consistently mild or severe pain may not have been included in the analysis because their lack of fluctuations failed to meet any of the definitions. Third, we selected a 3-day window for hazard and control periods and adopted a full stratum bidirectional approach to maximize data availability. This approach was chosen to account for the constraints of the 30-day study period while minimizing overlaps between hazard periods and pain flares. However, there were overlaps between control periods and postflare phases, during which symptoms could persist and potentially intertwine with the carry-on effects of pain flares. Excluding postflare phases and using alternative options, such as a 7-day window, were considered but deemed unsuitable, given the data availability. Our sensitivity analysis supports the current approach as a robust method. Employing a rolling window to capture fluctuations in exposures reflects real-world dynamics. After adjusting for these overlaps, we observed the same patterns in exposure outcome associations with expected variations due to changes in statistical power. Fourth, physical activity was not quantified based on individual calibration, and the accelerometer was optimized to record sleep data, potentially compromising the sensitivity in detecting movements during daytime. Finally, days with missing data for pain or any of the 12 exposures were excluded. This likely introduced selection bias, reducing the number of participants included in the analysis and limiting the availability of hazard and control periods. Imputation was not used to avoid introducing bias, as the missing data were likely not missing at random.

XSL•FO
RenderX

Implications and Future Research

The study provides a novel narrative in understanding pain flares through daily patient-generated data. It shows the potential of mHealth technologies for multidimensional monitoring of symptom patterns. This approach enables a more nuanced understanding of individual pain experiences and associated factors, which offers health care providers and patients an opportunity to better predict and proactively manage pain flares [45]. Incorporating both patient-reported and accelerometer-derived measures allows for capturing a more complete view of a patient's condition and its progression. It also nudges active patient involvement in their own care, promoting self-monitoring to increase awareness. However, it is important to recognize that patient engagement with mHealth technologies remains a significant challenge in digital health research, with issues such as data missingness affecting the quality of the data or lack of transparency compromising the integrity of reporting. Digital exclusion, contributed by factors such as socioeconomic disparities or lack of digital skills, is another critical factor that could limit the generalizability of digital health research. The QUASAR study, for example, excluded individuals who did not own a personal smartphone or tablet. This restriction may have inadvertently overlooked certain populations and their unique characteristics.

For pain research specifically, future studies are recommended to include multiple pain-related symptoms and explore how symptoms vary during different stages of pain flares. This approach will enhance our understanding of their nuances and impact on patients, potentially leading to more timely monitoring and personalized management strategies. Furthermore, future studies also need to consider the impact of digital exclusion and strive for greater inclusivity to encourage wider participation.

Conclusions

Our study, which leveraged mHealth technologies and numerically defined pain severity, has concluded that pain flares are commonly observed in patients with RA. The analysis of daily patient-generated health data indicates that changes in sleep patterns and emotional distress over 3 days may be associated with the occurrence of pain flares the following day. This study represents an early example of identifying pain flares using daily data, opening opportunities for timely monitoring and personalized management. As digital technologies evolve, they hold significant potential to transform how we understand and manage chronic conditions. It is also crucial to address the challenges of patient engagement, the impact of digital exclusion, and the need for greater inclusivity in future research.

Acknowledgments

This work used the data from the quality of life, sleep, and rheumatoid arthritis (QUASAR) study, which was funded by Versus Arthritis (formerly Arthritis Research), United Kingdom (grant reference: 21188). The authors would like to thank the National Rheumatoid Arthritis Society; National Health Service (NHS) Clinical Research Networks of Greater Manchester, West Midlands, North West Coast, Yorkshire, and Humber; East Cheshire NHS Trust; North Cumbria University Hospitals NHS Trust; Kent Community Health NHS Foundation Trust; Solent NHS Trust; and Gateshead Health NHS Foundation Trust for their support with participant recruitment in the QUASAR study. The QUASAR study was also supported by infrastructure support from the Centre for Epidemiology Versus Arthritis (grant reference: 21755, formerly Arthritis Research UK Centre for Epidemiology [grant reference: 20380]). The Centre for Epidemiology Versus Arthritis had no further input into how the study was executed or in the analysis or interpretation of data. CJA is supported by NIHR Manchester Biomedical Research Centre and NIHR Greater Manchester Patient Safety Research Collaboration. The views expressed in this publication are those of the authors and not necessarily those of NIHR. Generative artificial intelligence was not used in producing the manuscript.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

TCCH and JMcB were responsible for the conceptualization and methodology of the study. TCCH handled data curation, formal analysis, investigation, project administration, and validation. JMcB was responsible for funding acquisition and supervision, along with PW and CJA. Relevant resources were provided by KLD and BBY. There was no software development involved in this study. TCCH contributed to the writing of the original draft, while all authors were involved in the review and editing of the manuscript.

Conflicts of Interest

PW is a director and shareholder of a for-profit digital mental health company CareLoop Health Ltd. BBY is an employee of Astellas Pharma Europe Ltd. His involvement in this study was completed before joining Astellas. The rest of the authors declare no conflict of interest.

Multimedia Appendix 1
Supplementary materials supporting the results.

XSL•FO
RenderX

[DOCX File, 72 KB - mhealth_v13i1e64889_app1.docx]

References

1. Ali SM, Lau WJ, McBeth J, Dixon WG, van der Veer SN. Digital manikins to self-report pain on a smartphone: a systematic review of mobile apps. Eur J Pain 2021 Feb;25(2):327-338. [doi: 10.1002/ejp.1688] [Medline: 33113241]
2. Austin L, Sharp CA, van der Veer SN, et al. Providing "the bigger picture": benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. Rheumatology (Oxford) 2020 Feb 1;59(2):367-378. [doi: 10.1093/rheumatology/kez207] [Medline: 31335942]
3. Dixon WG, Beukenhorst AL, Yimer BB, et al. How the weather affects the pain of citizen scientists using a smartphone app. NPJ Digit Med 2019;2(1):105. [doi: 10.1038/s41746-019-0180-3] [Medline: 31667359]
4. Gandrup J, Selby DA, van der Veer SN, Mcbeth J, Dixon WG. Using patient-reported data from a smartphone app to capture and characterize real-time patient-reported flares in rheumatoid arthritis. Rheumatol Adv Pract 2022;6(1):rkac021. [doi: 10.1093/rap/rkac021] [Medline: 35392426]
5. McBeth J, Dixon WG, Moore SM, et al. Sleep disturbance and quality of life in rheumatoid arthritis: prospective mHealth study. J Med Internet Res 2022 Apr 22;24(4):e32825. [doi: 10.2196/32825] [Medline: 35451978]
6. Allen KD, Coffman CJ, Golightly YM, Stechuchak KM, Keefe FJ. Daily pain variations among patients with hand, hip, and knee osteoarthritis. Osteoarthr Cartil 2009 Oct;17(10):1275-1282. [doi: 10.1016/j.joca.2009.03.021]
7. Peters ML, Sorbi MJ, Kruise DA, Kerssens JJ, Verhaak PFM, Bensing JM. Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. Pain 2000 Feb;84(2-3):181-192. [doi: 10.1016/s0304-3959(99)00206-7] [Medline: 10666523]
8. Coole C, Drummond A, Watson PJ, Radford K. What concerns workers with low back pain? Findings of a qualitative study of patients referred for rehabilitation. J Occup Rehabil 2010 Dec;20(4):472-480. [doi: 10.1007/s10926-010-9237-5] [Medline: 20373135]
9. Dunn KM, Jordan K, Croft PR. Characterizing the course of low back pain: a latent class analysis. Am J Epidemiol 2006 Apr 15;163(8):754-761. [doi: 10.1093/aje/kwj100] [Medline: 16495468]
10. Little CL, Schultz DM, House T, Dixon WG, McBeth J. Identifying weekly trajectories of pain severity using daily data from an mHealth Study: cluster analysis. JMIR Mhealth Uhealth 2024 Jul 19;12:e48582. [doi: 10.2196/48582] [Medline: 39028557]
11. McWilliams DF, Rahman S, James RJE, et al. Disease activity flares and pain flares in an early rheumatoid arthritis inception cohort; characteristics, antecedents and sequelae. BMC Rheumatol 2019;3(1):49. [doi: 10.1186/s41927-019-0100-9] [Medline: 31832600]
12. Murphy SL, Lyden AK, Kratz AL, et al. Characterizing pain flares from the perspective of individuals with symptomatic knee osteoarthritis. Arthritis Care Res (Hoboken) 2015 Aug;67(8):1103-1111. [doi: 10.1002/acr.22545] [Medline: 25580697]
13. Setchell J, Costa N, Ferreira M, Makovey J, Nielsen M, Hodges PW. What constitutes back pain flare? A cross sectional survey of individuals with low back pain. Scand J Pain 2017 Oct;17(1):294-301. [doi: 10.1016/j.sjpain.2017.08.003] [Medline: 28866119]
14. Davin S, Wilt J, Covington E, Scheman J. Variability in the relationship between sleep and pain in patients undergoing interdisciplinary rehabilitation for chronic pain. Pain Med 2014 Jun;15(6):1043-1051. [doi: 10.1111/pme.12438] [Medline: 24716856]
15. de Koning EJ, Timmermans EJ, van Schoor NM, et al. Within-Person pain variability and mental health in older adults with osteoarthritis: an analysis across 6 European cohorts. J Pain 2018 Jun;19(6):690-698. [doi: 10.1016/j.jpain.2018.02.006] [Medline: 29496636]
16. Schneider S, Junghaenel DU, Keefe FJ, Schwartz JE, Stone AA, Broderick JE. Individual differences in the day-to-day variability of pain, fatigue, and well-being in patients with rheumatic disease: associations with psychological variables. Pain 2012 Apr;153(4):813-822. [doi: 10.1016/j.pain.2012.01.001] [Medline: 22349917]
17. Stone AA, Broderick JE, Porter LS, Kaell AT. The experience of rheumatoid arthritis pain and fatigue: examining momentary reports and correlates over one week. Arthritis & Rheumatism 1997 Jun;10(3):185-193 [FREE Full text] [doi: 10.1002/art.1790100306]
18. Costa N, Smits E, Kasza J, et al. ISSLS PRIZE IN CLINICAL SCIENCE 2021: what are the risk factors for low back pain flares and does this depend on how flare is defined? Eur Spine J 2021 May;30(5):1089-1097. [doi: 10.1007/s00586-021-06730-6] [Medline: 33537911]
19. Costa N, Smits EJ, Kasza J, et al. Are objective measures of sleep and sedentary behaviours related to low back pain flares? Pain 2022 Sep 1;163(9):1829-1837. [doi: 10.1097/j.pain.0000000000002585] [Medline: 35984383]
20. Hewlett S, Sanderson T, May J, et al. "I'm hurting, I want to kill myself": rheumatoid arthritis flare is more than a high joint count—an international patient perspective on flare where medical help is sought. Rheumatology (Oxford) 2012 Jan;51(1):69-76. [doi: 10.1093/rheumatology/keq455] [Medline: 21565901]
21. Vincent A, Whipple MO, Rhudy LM. Fibromyalgia flares: a qualitative analysis. Pain Med 2016 Mar;17(3):463-468. [doi: 10.1111/pme.12676] [Medline: 25586303]

XSL•FO
RenderX

22. Muller S, Hider SL, Raza K, Stack RJ, Hayward RA, Mallen CD. An algorithm to identify rheumatoid arthritis in primary care: a Clinical Practice Research Datalink study. BMJ Open 2015 Dec 23;5(12):e009309. [doi: 10.1136/bmjopen-2015-009309] [Medline: 26700281]
23. Landry GJ, Best JR, Liu-Ambrose T. Measuring sleep quality in older adults: a comparison using subjective and objective methods. Front Aging Neurosci 2015;7:166. [doi: 10.3389/fnagi.2015.00166]
24. Resnick B, Boltz M, Galik E, Fix S, Zhu S. Feasibility, reliability, and validity of the MotionWatch 8 to evaluate physical activity among older adults with and without cognitive impairment in assisted living settings. J Aging Phys Act 2021 Jun 1;29(3):391-399. [doi: 10.1123/japa.2020-0198] [Medline: 33361499]
25. Yavuz-Kodat E, Reynaud E, Geoffray MM, et al. Validity of actigraphy compared to polysomnography for sleep assessment in children with autism spectrum disorder. Front Psychiatry 2019;10:551. [doi: 10.3389/fpsyt.2019.00551] [Medline: 31428003]
26. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb 1;35(2):287-302. [doi: 10.5665/sleep.1642] [Medline: 22294820]
27. Druce KL, Cordingley L, Short V, et al. Quality of life, sleep and rheumatoid arthritis (QUASAR): a protocol for a prospective UK mHealth study to investigate the relationship between sleep and quality of life in adults with rheumatoid arthritis. BMJ Open 2018 Jan 26;8(1):e018752. [doi: 10.1136/bmjopen-2017-018752] [Medline: 29374666]
28. Anderson J, Caplan L, Yazdany J, et al. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. Arthritis Care Res (Hoboken) 2012 May;64(5):640-647. [doi: 10.1002/acr.21649] [Medline: 22473918]
29. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. Sleep Med Rev 2016 Feb;25:52-73. [doi: 10.1016/j.smrv.2015.01.009] [Medline: 26163057]
30. Espie CA, Kyle SD, Hames P, Gardani M, Fleming L, Cape J. The Sleep Condition Indicator: a clinical screening tool to evaluate insomnia disorder. BMJ Open 2014 Mar 18;4(3):e004183. [doi: 10.1136/bmjopen-2013-004183] [Medline: 24643168]
31. Herrmann C. International experiences with the hospital anxiety and depression scale—a review of validation data and clinical results. J Psychosom Res 1997 Jan;42(1):17-41. [doi: 10.1016/s0022-3999(96)00216-4] [Medline: 9055211]
32. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. Ann Rheum Dis 1978 Aug;37(4):378-381. [doi: 10.1136/ard.37.4.378] [Medline: 686873]
33. Conley S, Knies A, Batten J, et al. Agreement between actigraphic and polysomnographic measures of sleep in adults with and without chronic conditions: a systematic review and meta-analysis. Sleep Med Rev 2019 Aug;46:151-160. [doi: 10.1016/j.smrv.2019.05.001] [Medline: 31154154]
34. Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013 Nov 1;36(11):1747-1755. [doi: 10.5665/sleep.3142] [Medline: 24179309]
35. Landry GJ, Falck RS, Beets MW, Liu-Ambrose T. Measuring physical activity in older adults: calibrating cut-points for the MotionWatch 8(©). Front Aging Neurosci 2015;7:165. [doi: 10.3389/fnagi.2015.00165] [Medline: 26379546]
36. Beukenhorst AL, McBeth J, Sergeant JC, Schultz DM, Dixon WG. OP0071 The prevalence of pain flares: impact of definition. Ann Rheum Dis 2018 Jun;77:86-87. [doi: 10.1136/annrheumdis-2018-eular.7398]
37. Olsen MF, Bjerre E, Hansen MD, et al. Pain relief that matters to patients: systematic review of empirical studies assessing the minimum clinically important difference in acute pain. BMC Med 2017 Feb 20;15(1):35. [doi: 10.1186/s12916-016-0775-3] [Medline: 28215182]
38. Wu Y, Li S, Guo Y. Space-time-stratified case-crossover design in environmental epidemiology study. Health Data Sci 2021;2021:9870798. [doi: 10.34133/2021/9870798] [Medline: 38487511]
39. Mittleman MA, Mostofsky E. Exchangeability in the case-crossover design. Int J Epidemiol 2014 Oct;43(5):1645-1655. [doi: 10.1093/ije/dyu081] [Medline: 24756878]
40. Lewer D, Petersen I, Maclure M. The case-crossover design for studying sudden events. BMJ Med 2022;1(1):e000214. [doi: 10.1136/bmjmed-2022-000214] [Medline: 36936574]
41. Mun CJ, Suk HW, Davis MC, et al. Investigating intraindividual pain variability: methods, applications, issues, and directions. Pain 2019 Nov;160(11):2415-2429. [doi: 10.1097/j.pain.0000000000001626] [Medline: 31145212]
42. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol 1991 Jan 15;133(2):144-153. [doi: 10.1093/oxfordjournals.aje.a115853] [Medline: 1985444]
43. Druce KL, Dixon WG, McBeth J. Maximizing Engagement in Mobile Health Studies. Rheum Dis Clin North Am 2019 May;45(2):159-172. [doi: 10.1016/j.rdc.2019.01.004]
44. Tang NKY, Salkovskis PM, Poplavskaya E, Wright KJ, Hanna M, Hester J. Increased use of safety-seeking behaviors in chronic back pain patients with high health anxiety. Behav Res Ther 2007 Dec;45(12):2821-2835. [doi: 10.1016/j.brat.2007.05.004] [Medline: 17588530]
45. Little CL, Druce KL, Dixon WG, Schultz DM, House T, McBeth J. What do people living with chronic pain want from a pain forecast? A research prioritization study. PLoS ONE 2023;18(10):e0292968. [doi: 10.1371/journal.pone.0292968] [Medline: 37824568]

XSL•FO
RenderX

Abbreviations

AA: above average
AT: above threshold
CSD: Consensus Sleep Diary
HADS: Hospital Anxiety and Depression Scale
iSD: intraindividual standard deviation
MAT: move above threshold
mHealth: mobile health
OR: odds ratio
QUASAR: quality of life, sleep, and rheumatoid arthritis
RA: rheumatoid arthritis
RAPID-3: Routine Assessment of Patient Index Data 3

Edited by L Buis; submitted 13.08.24; peer-reviewed by DF Mcwilliams, R Jamison; revised version received 28.03.25; accepted 01.04.25; published 21.07.25.

Please cite as:
Hsu TCC, Yimer BB, Whelan P, Armitage CJ, Druce K, McBeth J
Associations Between Daily Symptoms and Pain Flares in Rheumatoid Arthritis: Case-Crossover mHealth Study
JMIR Mhealth Uhealth 2025;13:e64889
URL: https://mhealth.jmir.org/2025/1/e64889
doi:10.2196/64889

© Ting-Chen Chloe Hsu, Belay B Yimer, Pauline Whelan, Christopher J Armitage, Katie Druce, John McBeth. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Digital Software Support Platform for Hyperthyroidism Management in South Korea: Markov Simulation Model-Based Cost-Effectiveness Analysis

Jung Hyun Kim[1*], MPH, PhD; Jaeyong Shin[2,3], MPH, MD, PhD; Man S Kim[4], PhD; Jae Hoon Moon[5,6,7*], MD, PhD

[1]Division of Tourism and Wellness, Hankuk University of Foreign Studies, 81 Oedae-ro, Mohyeon-eup, Cheoin-gu, Yongin, Republic of Korea

[2]Institute for Innovation in Digital Healthcare, Yonsei University Health System, Seoul, Republic of Korea

[3]Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Translational-Transdisciplinary Research Center, Clinical Research Institute, Kyung Hee University Hospital at Gangdong, School of Medicine, Kyung Hee University, Seoul, Republic of Korea

[5]Center for Artificial Intelligence in Healthcare, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[6]THYROSCOPE INC, Ulsan, Republic of Korea

[7]Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, 82, Gumi-ro 173 Beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jae Hoon Moon, MD, PhD
Center for Artificial Intelligence in Healthcare, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

Abstract

Background: The integration of wearable technology for heart rate monitoring offers potential advancements in managing hyperthyroidism by providing a feasible way to track thyroid function. Although digital health solutions are gaining traction in various chronic conditions, their cost-effectiveness in hyperthyroidism management requires deeper investigation.

Objective: This study aimed to evaluate the cost-effectiveness of a wearable or mobile-based thyroid function digital monitoring solution for hyperthyroidism management and to make a comparison with the existing standard approach within the South Korean health care context.

Methods: We developed a decision-analytic Markov microsimulation model to simulate the cost and effectiveness of digital monitoring in a cohort of 10,000 hypothetical hyperthyroidism patients aged 40 years. The analysis was conducted from the perspective of the health care system, with a 4.5% annual discount rate applied to costs and effectiveness and an inflation adjustment to 2022 values. Model inputs were sourced from clinical studies, publicly available datasets, and expert input, with outcomes measured in quality-adjusted life years (QALYs). Cost-effectiveness was evaluated through incremental cost-effectiveness ratios (ICERs) and net monetary benefits (NMB), with additional deterministic and probabilistic sensitivity analyses performed to address input uncertainties.

Results: Integrating digital monitoring yielded an additional 0.32 QALYs per patient at an incremental cost of US $3143, resulting in an ICER of US $9804.30 per QALY, significantly below the South Korean willingness-to-pay threshold of US $32,255/QALY. The digitally supported group exhibited improved rates of long-term remission (22.68%, 2268/10,000) and reduced postremission relapse (17.87%, 1787/10,000) compared to standard care (17.48%, 1748/10,000 and 26.37%, 2637/10,000, respectively). Probabilistic sensitivity analysis showed that digital intervention was the preferred cost-effective strategy in 64.4% (6440/10,000) of iterations. Subscription costs of the digital platform and the utility weight for thyroid-associated orbitopathy emerged as key factors affecting the ICER in sensitivity analyses.

Conclusions: The findings suggest that digital monitoring provides a cost-effective strategy for enhancing hyperthyroidism management, supporting sustained remission, and reducing relapse rates. As such, digital solutions could serve as a valuable adjunct to traditional care, with the cost-effectiveness analysis providing an economic basis for determining pricing and value-based reimbursement in health care systems. The study underscores the importance of integrating digital solutions in chronic disease management and suggests that further research should include societal costs, such as productivity, to capture economic benefits fully.

(JMIR Mhealth Uhealth 2025;13:e56738) doi:10.2196/56738

XSL•FO
RenderX

KEYWORDS

hyperthyroidism; digital monitoring solution; wearable devices; mobile-based monitoring; cost-effectiveness analysis; QALYs; hyperthyroidism management; apps; applications; digital health; digital technology; digital interventions; SDG3:good health and well-being; quality-adjusted life years

Introduction

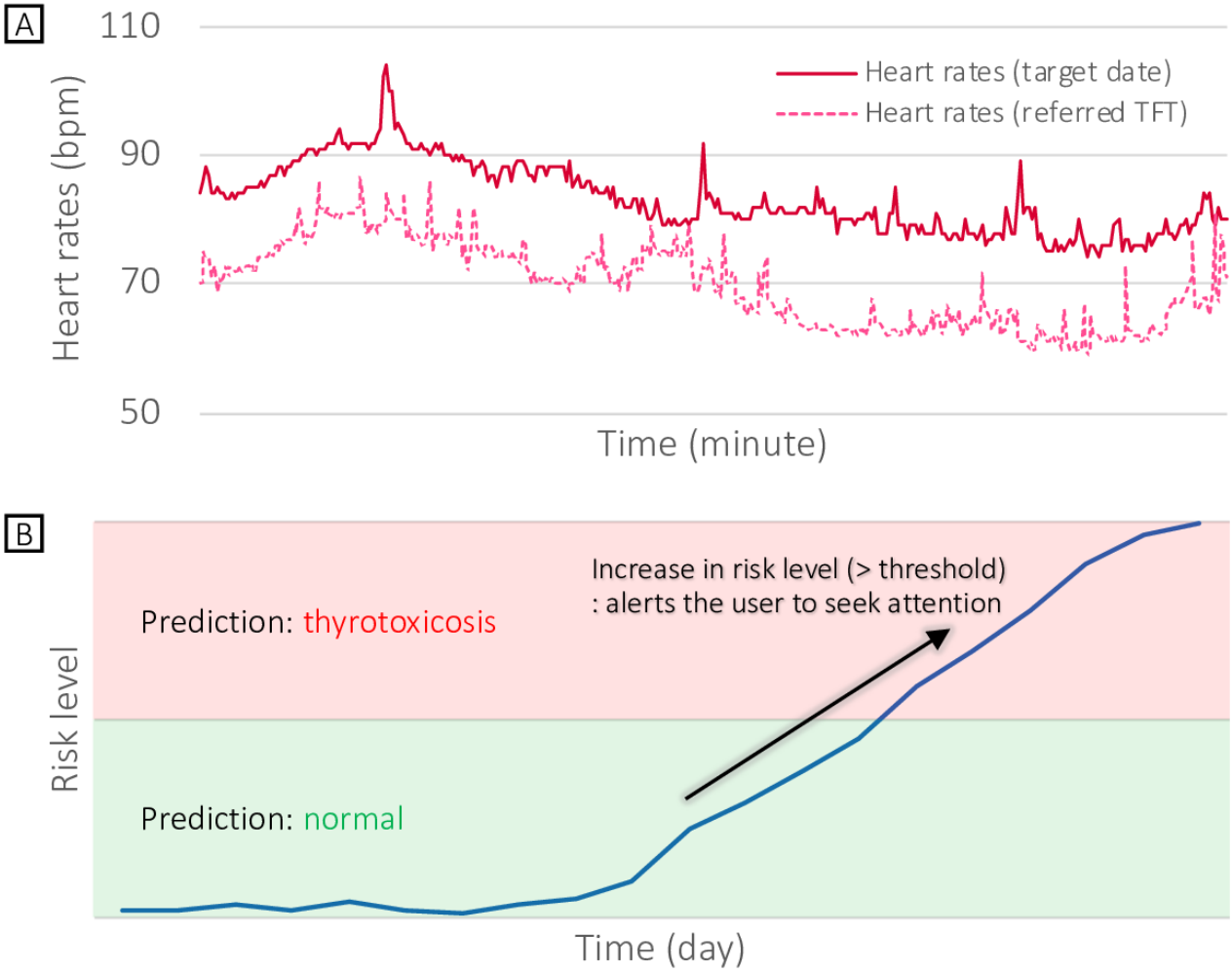
An overactive thyroid, medically termed hyperthyroidism or thyrotoxicosis, occurs when the thyroid gland generates an excessive amount of thyroid hormones. Graves disease (GD) is the most common etiology for hyperthyroidism, with an estimated annual incidence ranging from 20 to 50 cases per 100,000 persons [1]. In 2018, the prevalence of hyperthyroidism was 0.25%, and the age-standardized incidence in South Korea was 40.3 and 105.5 per 100,000 men and women, respectively [2]. Hyperthyroidism leads to a range of symptoms and signs due to thyroid hormones' impact on multiple organs. These include fatigue, anxiety, palpitations, excessive sweating, intolerance to heat, disrupted sleep, tremor, diarrhea, and weight loss [3]. Hyperthyroidism can be diagnosed through a thyroid function test (TFT) that measures serum thyroid hormone concentrations, including free thyroxine, thyrotropin, and other clinical factors. However, due to the nonspecific symptoms of hyperthyroidism, the timing of TFTs is often delayed, resulting in the diagnosis of hyperthyroidism occurring after it has significantly progressed. Therefore, having a straightforward and easily measurable parameter to predict thyroid function status would prove beneficial for the management of hyperthyroidism.

Photoplethysmography uses infrared light to monitor changes in blood circulation volume. This measurement offers valuable insights into the functioning of the cardiovascular system [4]. In recent times, the use of photoplethysmography technology for heart rate monitoring has gained popularity, primarily because it is easy to use and comfortable for wearers [5]. Various studies assessing the precision of wrist-worn heart rate monitors have been conducted, demonstrating relatively accurate heart rate measurements, especially for resting heart rate [6,7]. Collecting detailed longitudinal heart rate data and physical activity information using these wearable devices is straightforward, and such data can offer more insights than sporadic heart rate measurements. This capability enables the collection and analysis of more comprehensive and precise heart rate data, a crucial parameter that is affected by thyroid function status [8,9]. A medical software platform is under development that uses wearable technology to predict thyroid function state and assist in the management of thyroid dysfunction. It includes a mobile app and web solution for this purpose and is currently pending regulatory approval.

Despite the promising potential of wearable technology for managing thyroid dysfunction, there is limited research published on its actual impact concerning health care outcomes and overall costs. Therefore, we aimed to assess the cost-effectiveness of introducing a wearable or mobile-based thyroid function digital monitoring solution for hyperthyroidism management (Figure 1) driven by heart rate monitors commercialized as digital therapeutics and to compare it with the current standard approach.

XSL•FO
RenderX

Figure 1. Description of a wearable or mobile-based thyroid function digital monitoring solution designed for the effective management of hyperthyroidism. (A)Heart rates collected from a wearable device on each date (the date of referred thyroid function test and target date), and (B)the risk level of hyperthyroidism calculated for the target date is determined by comparing the heart rate distribution over N days, including the target date, with the heart rate distribution over N days, including the day of the referred thyroid function test. In cases where the risk level surpasses the designated threshold, a notification is sent to the user, prompting them to seek medical attention. TFT: thyroid function test.



Methods

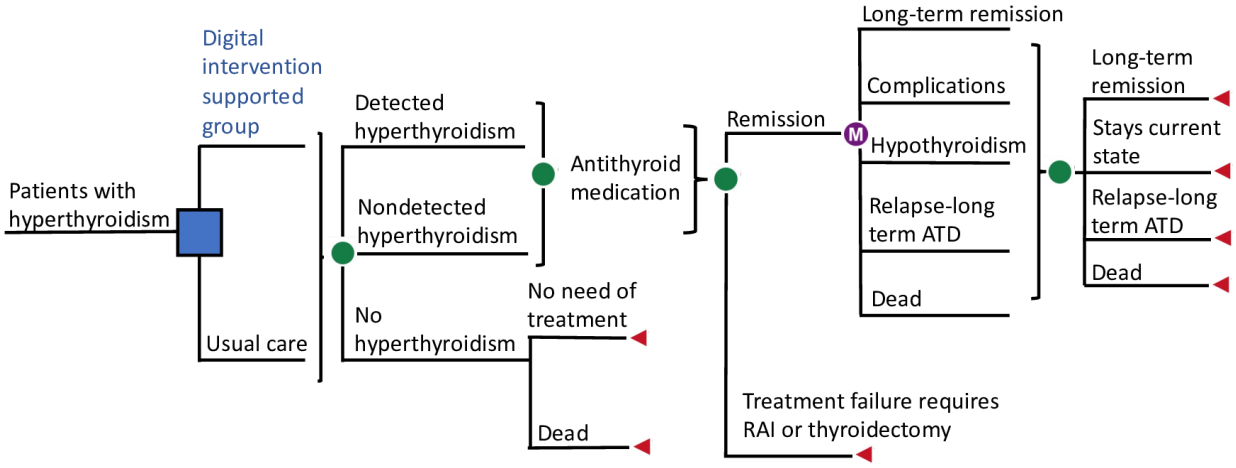
Model Overview

We constructed a decision-analytic Markov microsimulation model to simulate 10,000 iterations of the clinical course of patients with hyperthyroidism (Figure 2). While a strict definition distinguishes between hyperthyroidism and GD, this paper refers to hyperthyroidism and conducts data analysis under this terminology. However, in South Korea, the incidence of toxic nodules as the underlying cause of hyperthyroidism is exceedingly low, accounting for less than 0.5% of cases [10]. As a result, patient management in this study followed a protocol for GD. This model consists of several health states in which the simulated individual can be in and move between. The statistical definition of the model is a discrete-time stochastic process with the Markov property [11]. The baseline cohort was comprised of 40-year-old individual patients with hyperthyroidism in South Korea. To advance time in our model, we used an annual cycle length, and the simulation continued until death or 100 years of age. The simulation of the natural disease progression and the effect of screening in men and women requires data including prevalence, incidence, risk of events, morbidity, and mortality. The model captures the variability in individual patient responses to treatment and disease progression over time. Each iteration represents a single hypothetical patient moving through various health states based on probabilistic transitions, including remission, long-term remission, complications, hypothyroidism, relapse, long-term antithyroid drug (ATD) use, and death. Some proportion of patients remain in their current state (hypothyroid state), while others may transition to different health states, such as long-term remission or relapse requiring long-term ATD therapy to capture more realistic transitions between health states and better reflect the long-term impacts of treatments. Eventually, patients in any state could transition to an absorbing state, such as death. A lifetime Markov model was simulated over a total of 70 cycles with a one-year cycle length. Quality-adjusted life years (QALYs), which include both quantity (life-years gained) and quality (health-related quality of life incremental cost-effectiveness in utility value), were used as effectiveness variables. Incremental cost-effectiveness ratio (ICER) and incremental net monetary benefit were calculated to evaluate cost-effectiveness. By aggregating the outcomes from all 10,000

XSL•FO
RenderX

iterations, we estimate the overall clinical and economic impact of different strategies for hyperthyroidism. Figure 2 depicts the core model, and the concept of combining time variables, treatment eligibility, and clinical prognosis was adapted from the previous studies.

Figure 2. A basic description of the structure in the decision analytic Markov model. ATD: antithyroid drug; RAI: radioactive iodine.



National Health Insurance Service Claims Data

A retrospective, population-based cohort study was conducted to investigate complications associated with hyperthyroidism and the related costs in Korea. The study examined enrollment and claims data from individuals covered by the National Health Insurance Service (NHIS) between 2003 and 2015. Patients with hyperthyroidism were defined as those who had 2 or more diagnostic codes of E05 or E06 and had been prescribed ATDs for at least 6 months. These drugs included MMI (191801ATB and 191802ATB), carbimazole (471501ATB, 471502ATB, and 471503ATB), and PTU (220101ATB and 220102ATB). The final cohort consisted of 1171 patients, excluding those who had taken medications for less than 6 months, were younger than 20 years, or had missing information (Multimedia Appendix 1). The major complications considered were thyroid-associated orbitopathy (TAO), heart failure, atrial fibrillation, osteoporosis, fractures, and thyroid cancer. Direct medical costs, including physician and hospital reimbursements for treatment procedures, as well as the annual mean cost for each complication, were obtained from NHIS data (Table 1).

XSL•FO
RenderX

Table . Base case values and references of model input parameters.

Parameter	Estimation	Distribution	Source
Prevalence of hyperthyroidism	7.20% (range 5.76%-8.64%)	Beta	[12]
Failure of ATD[a]	5% (range 4%-6%) over 1.5 years	Beta	[13]
Hypothyroidism	2.9% (range 2.32%-3.48%) over 10.2 years	Beta	[14]
Relapse postremission with ATD	52.8% (range 42.24%-63.36%) over 3.73 years (reverts to 0 after 5 years)	Beta	[15]
Thyroid-associated orbitopathy	25% (range 20%-30%)	Beta	[16]
Atrial fibrillation	4.3% (range 3%-5%)	Beta	NHIS[b] [17],
Heart failure	1.6% (range 1.1% - 2.1%)	Beta	NHIS
Osteoporosis	1.59% (range 1.09% - 2.09%)	Beta	NHIS
Fracture	3.3% (range 2.8% - 3.8%)	Beta	NHIS
Thyroid cancer	1.23% (range 0.9% - 2%)	Beta	[18]
Hyperthyroidism sensitivity	87.13 (range 78.4 - 95.8)	Beta	GlandyTM, THYROSCOPE INC
Hyperthyroidism specificity	83.78 (range 75.4 - 92.15)	Beta	GlandyTM
T4[c] + TSH[d] sensitivity	100	Uniform	[19], Assumption
T4 + TSH specificity	100	Uniform	[19], Assumption
QALYs[e], estimate (range)			
Thyroid cancer	0.897 (0.852 - 0.941)	Beta	[20]
Hypothyroid	0.9479 (0.902 - 0.997)	Beta	[21]
Thyroid-associated orbitopathy	0.84 (0.798 - 0.882)	Beta	[22]
Hyperthyroidism with ATD	0.86 (t0.817 - 0.903)	Beta	[23]
Osteoporotic fracture	0.91 (0.8645 - 0.9555)	Beta	[24]
Fracture	0.83 (0.7885 - 0.8715)	Beta	[24]
Atrial fibrillation	0.81 (0.729 - 0.891)	Beta	[25]
Heart failure	0.78 (0.702 - 0.858)	Beta	[25]
Cost (US $), estimate (range)			
Acute care costs (initial year)	1125.15 (1012.64 - 1237.67)	Gamma	NHIS
Thyroid-associated orbitopathy (annual)	386.77 (367.44 - 406.11)	Gamma	NHIS
Atrial fibrillation (annual)	150.54 (143.02 - 158.07)	Gamma	NHIS
Heart failure (annual)	439.63 (417.65 - 461.61)	Gamma	NHIS
Osteoporosis (annual)	208.49 (198.07 - 218.91)	Gamma	NHIS
Fracture (annual)	347.41 (330.04 - 364.78)	Gamma	NHIS
Thyroid cancer (annual)	149.23 (141.77 - 156.69)	Gamma	NHIS
Hypothyroidism (annual)	250.37 (225.34 - 275.41)	Gamma	NHIS
Long-term ADT[f] (annual)	48.87 (43.98 - 53.76)	Gamma	NHIS
Additional cost			
Digital software aid use (subscription fee per month)	US $30 (40,000 KRW; range US $10-$50)[g]	Gamma	Assumption
TSH, free T4 test fee	US $92.44 (US $64.71 - $120.17)	Gamma	NHIS
Discount rates	4.5% (3%-5%)	—[h]	[26]

[a]ATD: antithyroid drug.

XSL•FO
RenderX

[b]NHIS: National Health Insurance Service.
[c]T4: thyroxine.
[d]TSH: thyroid-stimulating hormone.
[e]QALY: quality-adjusted life year.
[f]ADT: antithyroid drug therapy.
[g]KRW: Korean won.
[h]Not applicable.

Intervention and Comparators

To identify the potential effects of the digital monitoring solution used for hyperthyroidism management, we compared 2 strategies in this study. In the control group, suspected patients with hyperthyroidism were diagnosed and treated with their usual care without using the digital solution, while the digitally supported group was assisted by the digital monitoring solution in the management and treatment of hyperthyroidism. We hypothesize that having a daily check of hormone status using this digital solution may lead to reduced hospital visits and an increase in the probability of long-term remission.

Hyperthyroidism and Complications

In this analysis, we used the data from randomized controlled trial studies, the NHIS claims data [27], and the published literature for determining the probabilities of individuals with hyperthyroidism and its complications from the screening procedure. The primary objective of hyperthyroidism management was to limit the number of major complications, including TAO, heart failure, atrial fibrillation, osteoporosis, fractures, and thyroid cancer. Rates and probabilities of major complications were from a Korean setting were obtained from the NHIS and supplemented with data retrieved from the published literature based on the systematic review. Standard all-cause and disease-specific mortality based on age and risk for South Korea in 2021 [28] was used in the simulation model.

Quality-of-Life Adjustments and Costs

Utilities were estimated for all major events, including hyperthyroidism and related complications, which were obtained from published literature based on a systematic review and meta-analysis. Utilities were derived from a patient perspective to provide a more accurate representation of quality-of-life impacts by reflecting patients' lived experiences and to account for adaptation to chronic conditions over time. This context-specific approach yields more relevant data for specific treatments, especially for conditions with long-term quality-of-life implications, such as hyperthyroidism or hypothyroidism [29-32]. We adopted a health care system perspective according to the economic evaluation guidelines in Korea [26]. Direct medical care costs, including physician and hospital reimbursement for treatment procedures, were obtained from the NHIS. Total mean medical care costs for hyperthyroidism, TAO, heart failure, atrial fibrillation, osteoporosis, fractures, and thyroid cancer were obtained and confirmed from the published literature. Monthly subscription fees for the group with digital monitoring solution use were recurring annually and applied for a lifetime. Sensitivity analyses were performed to assess the association of cost reduction due to unnecessary hospital visits using the thyroid function digital monitoring solution (70% - 100%). A 4.5% discount rate was used for both costs and effects in the base-case scenario. Costs were adjusted to the year 2022 using consumer price indices for medical care [33] and converted to 2022 US dollars (US $1=1293.68 KRW) [34]. Table 1 presents the quality of life and utility decrements used in the model together with the other parameters.

Statistical Analysis

Our primary endpoints were QALYs, total costs (2022 USD), and ICERs. Future costs and QALYs were discounted at a rate of 4.5%. We used a willingness-to-pay (WTP) threshold of US $32,255 per QALY gained, reflecting 1-time the gross domestic product per capita in 2022 South Korea [26,35-38] to determine cost-effectiveness. A strategy was considered preferred if it resulted in the greatest increase in QALYs while being cost-effective. Overall, this study followed the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guideline (Checklist 1) [39].

To assess how the uncertainty in model inputs influences the outcomes, we conducted 1-way and probabilistic sensitivity analyses. In 1-way sensitivity analyses, we systematically changed 1 parameter at a time within a specified range of values while keeping all other input variables constant at their base-case values. The upper and lower bounds for these parameter ranges were determined based on either the minimum and maximum values available in relevant databases or the range of 90%-110% for each value.

Probabilistic sensitivity analyses were performed by repeatedly sampling model inputs simultaneously from defined probabilistic distributions. We used gamma distributions for costs and beta distributions for all other parameters. This process involved running the model 10,000 times, considering a cohort of 10,000 patients. We then calculated the percentage of instances in which each strategy was favored under the WTP thresholds.

Ethical Considerations

The procedures of this study underwent expedited review and received approval from the institutional review board of Severance Hospital, Yonsei University (IRB number: 4-2022-1525). In accordance with local legislation and institutional requirements, ethical approval was not mandatory for this study. Likewise, written informed consent for participation was not required, in accordance with national legislation and institutional requirements. All data have been deidentified.

XSL•FO
RenderX

Results

Base-Case Analysis

The results of base-case cost-effectiveness analysis are presented in Table 2. Patients with the implementation of the digital monitoring solution in managing hyperthyroidism were associated with increased costs and increased QALYs. The model predicted that the implementation of digital software in managing hyperthyroidism produced 11.48 QALYs and US $9593 expected costs compared to 11.16 QALYs and US $6445 generated by treatment as usual without digital software aid over a lifetime horizon. The introduction of the thyroid function digital monitoring solution would increase costs by US $3143 and prolong QALYs by 0.32. The ICER associated with hyperthyroidism with digital intervention management tool was US $9804.30 per QALY gained (Table 2), which was well below the hypothetical threshold value in Korea of US $32,255 per QALY gained.

Notably, the Markov probability analysis revealed that hyperthyroidism management with digital intervention increased the proportion of long-term remission and reduced the likelihood of postremission relapse of hyperthyroidism (Table 2 and Figure 3).

Table . Effectiveness and cost estimates for two strategies on a simulated cohort of patients with hyperthyroidism.

	Hyperthyroidism managed without digital software support	Hyperthyroidism managed with digital software support
Long-term remission, % (n/N)	17.48 (748/10,000)	22.68 (2268/10,000)
Complications, % (n/N)	15.17 (1517/10,000)	18.84 (1884/10,000)
Hypothyroidism, % (n/N)	5.40 (540/10,000)	5.02 (502/10,000)
Relapse postremission with ATD[a], % (n/N)	26.37 (2637/10,000)	17.87 (1787/10,000)
Discounted lifetime cost per person, US $	6449.77	9592.55
Discounted incremental cost per person, US $	—	3142.78
Discounted lifetime effectiveness per person, QALY[b]	11.16	11.48
Discounted incremental effectiveness per person, QALY	—	0.32
Discounted ICER[c], incremental costs per QALY gained	—	9804.30
Discounted NMB[d] in WTP[e] of US $32,255, US $	436,952,569.9	449,505,659.6
Discounted INMB[f] in WTP of US $32,255, US $	—	12,553,089.7

[a]ATD: antithyroid drug.

[b]QALY: quality-adjusted life year.

[c]ICER: incremental cost-effectiveness ratio.

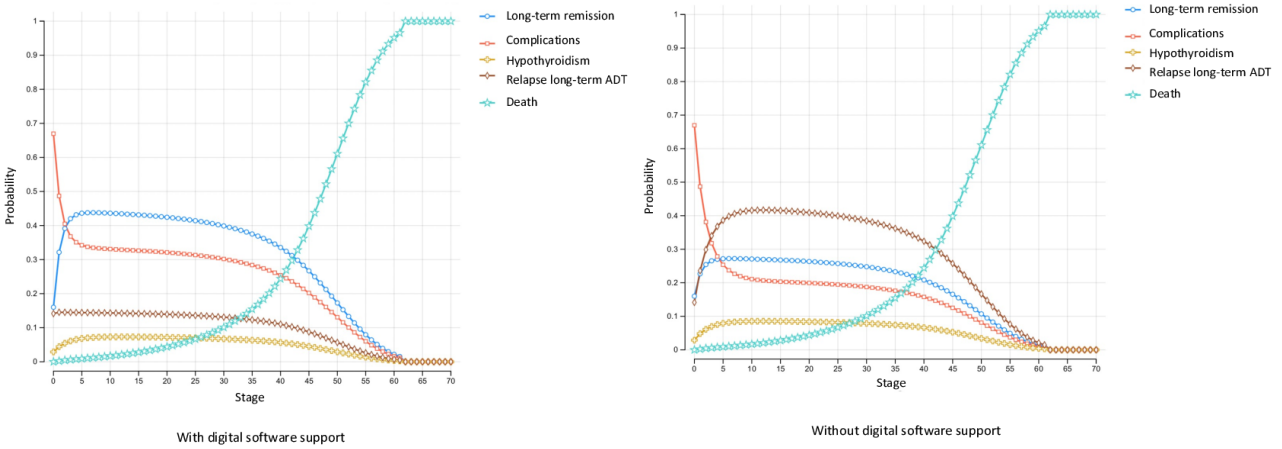
[d]NMB: net monetary benefit.

[e]WTP: willingness to pay.

[f]INMB: incremental net monetary benefit.

XSL•FO
RenderX

Figure 3. Markov probability analysis of each state with and without digital software support. ADT: antithyroid drug therapy.



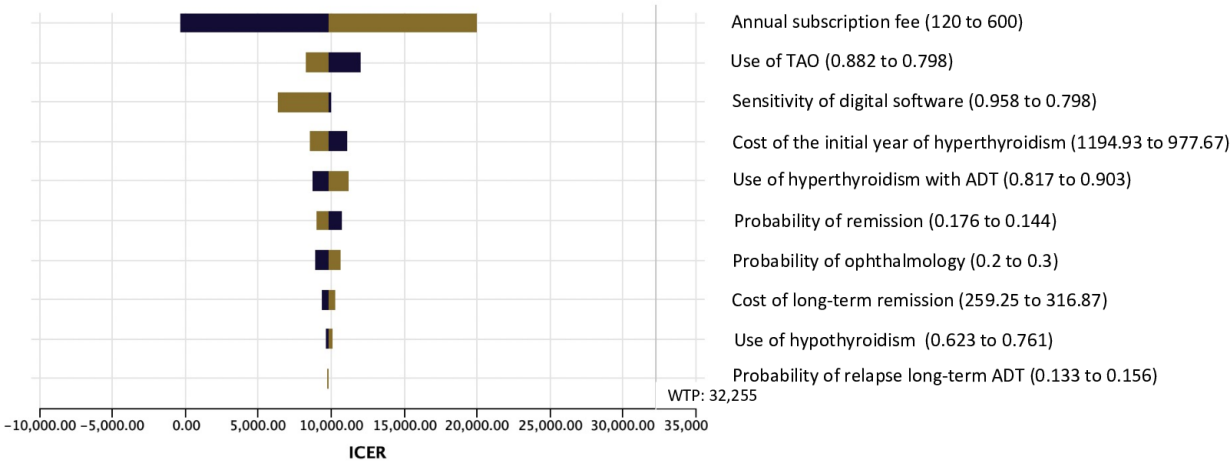
Sensitivity Analysis

We conducted a 1-way deterministic sensitivity analysis. The results of the 1-way deterministic sensitivity analyses with the tornado diagram are shown in Figure 4. The annual cost of the digital monitoring solution use, which ranged from US $120 to US $600, had the largest impact on the ICER value, followed by use of TAO and the sensitivity of the digital software. According to the one-way threshold analysis for the monthly digital support cost, the digital software aid arm became the dominant strategy when the cost was below US $10 per month (Multimedia Appendix 1 and Figure 2). Even the highest ICER value observed in this diagram was still below the threshold value of US $32,255 per QALY gained.

In Table 3, the ICERs were altered based on the assumption that hospital visits for TFT would decrease as hormone status is monitored and managed. This reduction would lead to lower costs associated with unnecessary hospital visits and tests.

For the Monte Carlo simulations, we presented the results of probabilistic sensitivity analysis as a cost-effectiveness acceptability curve, cost-effectiveness scatter plot, and incremental cost-effectiveness scatterplots. To study the uncertainty in our results, the statistical uncertainty was tested probabilistically 10,000 times. The result of the probabilistic analysis presented as acceptability curves in Figure 5 described the probabilities of being cost-effective when WTP changes. From the results of the Monte Carlo simulation, the digital intervention was the optimal cost-effective strategy with a probability of 64.4% represented in the form of an ICE scatterplot in Figure 5.

Figure 4. Tornado diagram for incremental cost-effectiveness ratio. ADT: antithyroid drug therapy; ICER: incremental cost-effectiveness ratio; TAO: thyroid-associated orbitopathy; WTP: willingness to pay.



XSL•FO
RenderX

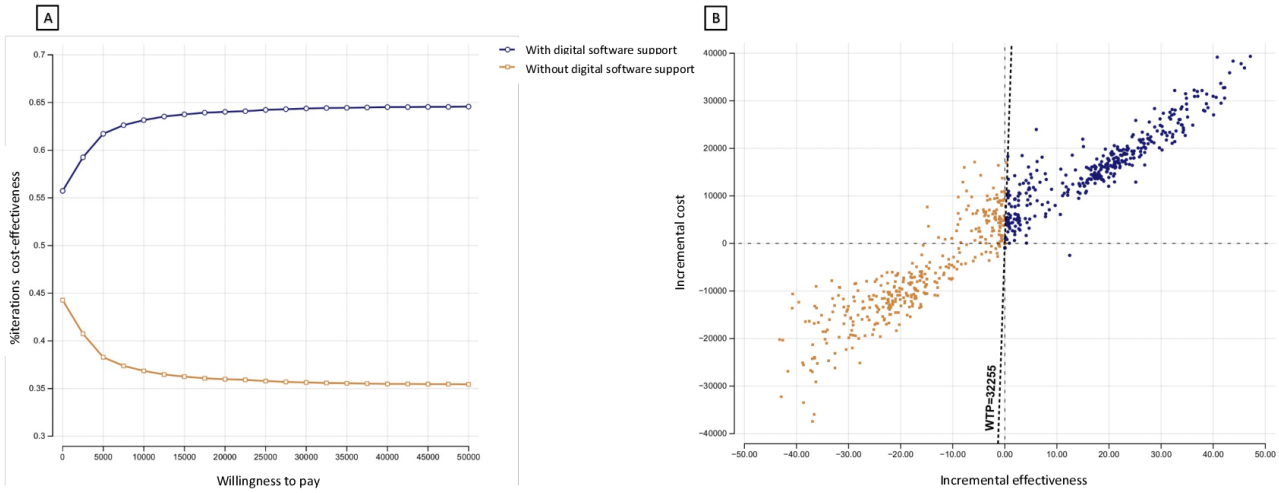
Table . One-way sensitivity analysis by reduction of cost for hospital visits.

Strategy	Cost (US $)	QALY[a] (years)	ICER[b] ($ per QALY)
Base case			
Treatment as usual	6449.77	11.16	—
Management with digital aid	9592.56	11.48	9804.31
10% cost reduction from unnecessary hospital visits			
Treatment as usual	6449.77	11.16	—
Management with digital aid	9541.50	11.48	9654.04
20% cost reduction from unnecessary hospital visits			
Treatment as usual	6449.77	11.16	—
Management with digital aid	9490.45	11.48	9485.76
30% cost reduction from unnecessary hospital visits			
Treatment as usual	6449.77	11.16	—
Management with digital aid	9439.39	11.48	9326.48

[a]QALY: quality-adjusted life year.

[b]ICER: incremental cost-effectiveness ratio.

Figure 5. Results of probabilistic sensitivity analysis: (**A**) cost-effectiveness acceptability curve and (**B**) incremental cost-effectiveness scatterplot. WTP: willingness to pay.



Discussion

Principal Findings

The application of artificial intelligence–driven wearable devices in cardiac disease management is operative; however, there is little evidence of their applications in thyroid disease and its cost-effectiveness. To the best of our knowledge, this is the first economic evaluation that considers the cost-effectiveness of a wearable or mobile-based thyroid function digital monitoring solution compared with standard care from a health care system perspective. The results demonstrated that this digital monitoring support would provide an incremental health gain per patient of 0.32 QALY with an ICER of US $9804.30 per QALY gained. A Markov microsimulation model was developed to optimize the use of data from the published studies based on a systematic review while accounting for the diversity and variability among patients with hyperthyroidism. The findings indicated that aiding patients with the digital monitoring solution resulted in slightly increased health care costs and improved health benefits compared to standard care. Therefore, the digital solution was considered highly cost-effective for functional outcomes in hyperthyroid management using a WTP threshold of US $32,255 per QALY gained.

Both deterministic and probabilistic sensitivity analyses demonstrated the robustness of our results. The cost-effectiveness was sensitive to the annual subscription fee for the hyperthyroidism management platform use, use of TAO, sensitivity of the digital device, initial treatment cost of hyperthyroidism, and use of long-term ATD. Despite these sensitivities, the digital intervention strategy maintained its advantageous position by including most of the parameters, with an ICER remaining below the threshold of US $32,255.

Our analysis revealed that the monthly cost of a digital monitoring solution for hyperthyroidism had the greatest impact on ICERs in the 1-way deterministic sensitivity analysis.

XSL•FO
RenderX

Although the model inputs and target disease were different, other studies pointed out a similar conclusion by evaluating the cost-effectiveness of digital therapeutics for low back pain [40] and hypertension [41]. The result from our model indicates that even minor adjustments in pricing can shift the balance between different strategies in this particular scenario, highlighting the sensitivity of the economic implications of these digital tools to various factors such as implementation details, settings, payer viewpoints, and assumed unnecessary hospital visit reduction. Further research exploring diverse payment approaches for digital intervention will be essential for conducting robust comparisons and reaching definitive conclusions regarding the health economic outcomes linked to digital technology. In Korea's single-payer system, the adoption of artificial intelligence–based systems is significantly influenced by the rate at which reimbursements are provided.

The major drawback of ATD therapy is the high recurrence rate of GD, the most common cause of hyperthyroidism, ranging from 20% to 70%, once the conventional 12 - 18 months of treatment is discontinued [42]. The serum concentration of thyrotropin receptor antibodies, a significant predictor of relapse in GD, can fluctuate or remain elevated in patients with GD, regardless of the continuous use of ATD. Therefore, there is a higher incidence of GD recurrence in clinical remission states following ATD treatment. The median duration of remission after ATD treatment is reported to be approximately 6.8 years [43], leading to suggestions of continuous, long-term ATD therapy [44-47]. A meta-analysis has shown that for each additional year of ATD treatment beyond 24 months, the remission rate increases by 16% [48]. These findings suggest that in cases where thyrotropin receptor antibody levels fail to return to normal after the conventional 12 - 18 months of ATD treatment, the possibility of considering long-term therapy should be explored [49]. Therefore, international medical societies are now considering long-term treatment as an option for managing Graves hyperthyroidism [42,50].

Significant adherence to pharmacological treatments is a multifaceted behavior influenced by various factors at different stages of a person's medication journey. Adherence to long-term therapy for chronic illnesses is estimated to reach only 50% on average [51]. At present, no single intervention has proven consistently effective in addressing long-term nonadherence. Digital interventions for medication adherence, however, offer a potential solution to enhance adherence by providing multifaceted interventions tailored to individual needs. Digital technological interventions have displayed initial promise in boosting medication adherence rates and offering health care professionals the ability to monitor and assess adherence [52-55]. Effectively managing hyperthyroidism requires patients to integrate it into their daily lives. By integrating digital monitoring solutions into the management of thyroid dysfunction, patients can regularly check their thyroid function status as part of their daily life, ensuring adherence to their medication schedule and promoting long-term health outcomes improvement.

In this research, we demonstrated the implementation of a digital intervention aid specifically designed to aid in the management of patients with hyperthyroidism. This innovation holds the potential to enhance health care outcomes and deliver cost-effectiveness on a large scale. Our sensitivity analyses have highlighted the robustness of our findings, even when considering variations in model parameters and assumptions. It is important to note that our study primarily adopts the perspective of the Korean national health care system and does not encompass a broader view of societal costs. Given the fact that hyperthyroidism is prevalent among individuals younger than 65 years, it is conceivable that the cost-effectiveness associated with hyperthyroidism could be even more substantial if we were to take into account factors such as productivity losses and a wider range of societal costs [12].

Limitations

This study comes with several limitations. We applied patient-derived uses in the study. Patients may underestimate the severity of certain health states due to adaptation or limited awareness of alternative health states, which can lead to higher-than-expected utility values. This effect is especially relevant in chronic conditions, where patients may grow accustomed to their symptoms over time; however, we adjusted these values by applying a decrement. While our analysis focused on the perspective of health care providers, it is essential to recognize that thyroid dysfunction has substantial health-related and economic implications not only for individual patients but also for society as a whole. The broader societal costs are expected to increase in the future, primarily due to the population with hyperthyroidism being an active working-age population in the society. In addition, because digital intervention–supported management has not been systematically integrated into health care practice in Korea and due to uncertainties surrounding the application of digital health frameworks, there remain ambiguities related to our estimates. Moreover, uncertainties are compounded by the possibility of negotiations influencing unit prices, which contribute to the overall expenses of the digital-based platform, especially if the intervention were to be introduced. We did not apply attrition rates in the model as we assumed that the patient with digital intervention would keep using the tool to manage hyperthyroidism, which could impact lifetime ICER. In addition, a limitation of this study is the potential discomfort associated with wearable or mobile-based monitoring solutions, particularly the burden of continuous use. This is especially relevant to older adult populations, who may face challenges with digital interventions, potentially adding a societal burden. Although capturing this discomfort in terms of costs or utility values is challenging, it may impact cost-effectiveness by affecting adherence and patient satisfaction. Future research should consider these factors, as they could influence the overall effectiveness and acceptability of digital health interventions.

Conclusions

In conclusion, the integration of a wearable or mobile-based thyroid function digital monitoring solution demonstrated good cost-effectiveness in hyperthyroidism management compared to ATD treatment as usual. Digital monitoring solution cost and its sensitivity influence cost-effectiveness, and we need to explore the balance among the digital monitoring support application cost, accuracy of hormonal level change status, and

XSL•FO
RenderX

effectiveness for target patients with hyperthyroidism in clinical practice. Digital therapeutics are currently in a pretest period to determine the extent of reimbursement they may receive from Korea's national universal health insurance system. Therefore, evaluating the cost-effectiveness of digital interventions, as illustrated in this study, is gaining significance in establishing suitable pricing for value-based reimbursement.

Acknowledgments

This work was supported by the Hankuk University of Foreign Studies Research Fund. It is also supported by the Technological Innovation R and D Program (RS-2023-00281098) funded by the Ministry of SMEs and Startups (MSS, Korea). In addition, this research was supported by the AI-based Medical System Digital Transformation Support Program through the National IT Industry Promotion Agency (NIPA), funded by the Ministry of Science and ICT (H0906-24-1006). A grant from the Korean Thyroid Association Clinical Research Award (2019) also contributed to this work. We express our gratitude to THYROSCOPE INC for providing information regarding digital intervention management for hyperthyroidism, which was critical to this study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JHK had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. JHM contributed clinical expertise to interpret the results and assisted with manuscript writing, with a focus on minimizing potential conflicts of interest in data analysis and result interpretation. JHK and JHM contributed equally as cocorresponding authors. JHK, JS, and JHM conceived and designed the study. JHK compiled the data and conducted the analyses. JHK and MSK developed the decision tree and Markov model. JHK and JHM drafted the manuscript. JHK, JS, MSK, and JHM critically revised the manuscript for important intellectual content.
JHK and JHM contributed equally as corresponding authors.

Conflicts of Interest

THYROSCOPE INC holds a patent for the wearable/mobile-based thyroid function digital monitoring solution. JHM is a stockholder and board member of THYROSCOPE INC. JHK, MSK, and JS declare no competing interests.

Multimedia Appendix 1
Additional files.
[DOCX File, 654 KB - mhealth_v13i1e56738_app1.docx]

Checklist 1
CHEERS (Consolidated Health Economic Evaluation Reporting Standards) 2022 checklist.
[DOCX File, 17 KB - mhealth_v13i1e56738_app2.docx]

References

1. Smith TJ, Hegedüs L. Graves' Disease. N Engl J Med 2016 Oct 20;375(16):1552-1565. [doi: 10.1056/NEJMra1510030] [Medline: 27797318]
2. Ahn HY, Cho SW, Lee MY, et al. Prevalence, treatment status, and comorbidities of hyperthyroidism in Korea from 2003 to 2018: a nationwide population study. Endocrinol Metab (Seoul) 2023 Aug;38(4):436-444. [doi: 10.3803/EnM.2023.1684] [Medline: 37435663]
3. Devereaux D, Tewelde SZ. Hyperthyroidism and thyrotoxicosis. Emerg Med Clin North Am 2014 May;32(2):277-292. [doi: 10.1016/j.emc.2013.12.001] [Medline: 24766932]
4. Tamura T, Maeda Y, Sekine M, et al. Wearable photoplethysmographic sensors—past and present. Electronics (Basel) 2014;3(2):282-302. [doi: 10.3390/electronics3020282]
5. Sviridova N, Sakai K. Human photoplethysmogram: new insight into chaotic characteristics. Chaos, Solitons & Fractals 2015 Aug;77:53-63. [doi: 10.1016/j.chaos.2015.05.005]
6. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34. [doi: 10.2196/mhealth.7043] [Medline: 28302596]
7. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2017 Jan 1;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]

XSL•FO
RenderX

8. Lee JE, Lee DH, Oh TJ, et al. Clinical feasibility of monitoring resting heart rate using a wearable activity tracker in patients with thyrotoxicosis: prospective longitudinal observational study. JMIR Mhealth Uhealth 2018 Jul 13;6(7):e159. [doi: 10.2196/mhealth.9884] [Medline: 30006328]
9. Kim KH, Lee J, Ahn CH, et al. Association between thyroid function and heart rate monitored by wearable devices in patients with hypothyroidism. Endocrinol Metab (Seoul) 2021 Oct;36(5):1121-1130. [doi: 10.3803/EnM.2021.1216] [Medline: 34674500]
10. Moon JH, Yi KH. The diagnosis and management of hyperthyroidism in Korea: consensus report of the Korean Thyroid Association. Endocrinol Metab (Seoul) 2013 Dec;28(4):275-279. [doi: 10.3803/EnM.2013.28.4.275] [Medline: 24396691]
11. Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the Economic Evaluation of Health Care Programmes: Oxford University Press; 2015.
12. Seo GH, Kim SW, Chung JH. Incidence & prevalence of hyperthyroidism and preference for therapeutic modalities in Korea. J Korean Thyroid Assoc 2013;6(1):56. [doi: 10.11106/jkta.2013.6.1.56]
13. In H, Pearce EN, Wong AK, et al. Treatment options for Graves disease: a cost-effectiveness analysis. J Am Coll Surg 2009 Aug;209(2):170-179. [doi: 10.1016/j.jamcollsurg.2009.03.025] [Medline: 19632593]
14. Azizi F, Ataie L, Hedayati M, et al. Effect of long-term continuous methimazole treatment of hyperthyroidism: comparison with radioiodine. Eur J Endocrinol 2005 May;152(5):695-701. [doi: 10.1530/eje.1.01904] [Medline: 15879354]
15. Sundaresh V, Brito JP, Wang Z, et al. Comparative effectiveness of therapies for Graves' hyperthyroidism: a systematic review and network meta-analysis. J Clin Endocrinol Metab 2013 Sep;98(9):3671-3677. [doi: 10.1210/jc.2013-1954] [Medline: 23824415]
16. Gontarz-Nowak K, Szychlińska M, Matuszewski W, et al. Current knowledge on Graves' orbitopathy. J Clin Med 2020 Dec 23;10(1):16. [doi: 10.3390/jcm10010016] [Medline: 33374706]
17. Franklyn JA, Sheppard MC, Maisonneuve P. Thyroid function and mortality in patients treated for hyperthyroidism. JAMA 2005 Jul 6;294(1):71-80. [doi: 10.1001/jama.294.1.71] [Medline: 15998893]
18. Yeh NC, Chou CW, Weng SF, et al. Hyperthyroidism and thyroid cancer risk: a population-based cohort study. Exp Clin Endocrinol Diabetes 2013 Jul;121(7):402-406. [doi: 10.1055/s-0033-1341474] [Medline: 23616188]
19. de los Santos ET, Starich GH, Mazzaferri EL. Sensitivity, specificity, and cost-effectiveness of the sensitive thyrotropin assay in the diagnosis of thyroid disease in ambulatory patients. Arch Intern Med 1989 Mar;149(3):526-532. [doi: 10.1001/archinte.149.3.526] [Medline: 2493228]
20. Mols F, Schoormans D, Smit JWA, et al. Age-related differences in health-related quality of life among thyroid cancer survivors compared with a normative sample: results from the PROFILES Registry. Head Neck 2018 Oct;40(10):2235-2245. [doi: 10.1002/hed.25325] [Medline: 29756245]
21. Pacini F, Ladenson PW, Schlumberger M, et al. Radioiodine ablation of thyroid remnants after preparation with recombinant human thyrotropin in differentiated thyroid carcinoma: results of an international, randomized, controlled study. J Clin Endocrinol Metab 2006 Mar;91(3):926-932. [doi: 10.1210/jc.2005-1651] [Medline: 16384850]
22. Brown GC, Brown MM, Chaudhry I, Stein JD. Opportunities to reduce potential bias in ophthalmic cost-utility analysis. JAMA Ophthalmol 2021 Apr 1;139(4):389-397. [doi: 10.1001/jamaophthalmol.2020.6591] [Medline: 33538789]
23. Yang F, Watanabe S, Hayashi K, et al. Cost-effectiveness analysis of antithyroid drug therapy, 131I therapy and subtotal thyroidectomy for Graves' disease. Radioisotopes 2007;56(2):65-76. [doi: 10.3769/radioisotopes.56.65] [Medline: 17766715]
24. Hiligsmann M, Ethgen O, Richy F, Reginster JY. Utility values associated with osteoporotic fracture: a systematic review of the literature. Calcif Tissue Int 2008 Apr;82(4):288-292. [doi: 10.1007/s00223-008-9117-6] [Medline: 18404243]
25. Wouters OJ, Naci H, Samani NJ. QALYs in cost-effectiveness analysis: an overview for cardiologists. Heart 2015 Dec;101(23):1868-1873. [doi: 10.1136/heartjnl-2015-308255] [Medline: 26395316]
26. Guidelines for economic evaluation of pharmaceuticals in Korea. HIRA. 2021. URL: https://repository.hira.or.kr/handle/2019.oak/2541 [accessed 2025-07-08]
27. Lee J, Lee JS, Park SH, et al. Cohort profile: the National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. Int J Epidemiol 2017 Apr 1;46(2):e15. [doi: 10.1093/ije/dyv319] [Medline: 26822938]
28. Data in number of deaths by region, sex, and age; 2021. Korean Statistical Information Service. 2021. URL: https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_2KAA811&conn_path=I3 [accessed 2025-05-23]
29. Dolan P. Whose preferences count? Med Decis Making 1999;19(4):482-486. [doi: 10.1177/0272989X9901900416] [Medline: 10520686]
30. Albrecht GL, Devlieger PJ. The disability paradox: high quality of life against all odds. Soc Sci Med 1999 Apr;48(8):977-988. [doi: 10.1016/s0277-9536(98)00411-0] [Medline: 10390038]
31. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: a call to reconsider current guidelines. Soc Sci Med 2016 Sep;165:66-74. [doi: 10.1016/j.socscimed.2016.07.043]
32. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ 1986 Mar;5(1):1-30. [doi: 10.1016/0167-6296(86)90020-2] [Medline: 10311607]
33. The average exchange rate in 2022. WooriBank. URL: https://spot.wooribank.com/pot/Dream?withyou=FXXRT0016 [accessed 2025-07-08]
34. KOSIS. Consumer Price Index: Korean Statistical Information Service; 2023.

XSL•FO
RenderX

35. Ahn JH, Kim YH, Shin SJ, et al. Research on methodologies for evidence-based healthcare decision-making process in korea. NECA. 2010. URL: https://www.neca.re.kr/lay1/program/S1T11C145/report/view.do?seq=54 [accessed 2025-07-08]
36. Goldman L, Garber AM, Grover SA, et al. 27th Bethesda Conference: matching the intensity of risk factor management with the hazard for coronary disease events. Task Force 6. Cost effectiveness of assessment and management of risk factors. J Am Coll Cardiol 1996 Apr;27(5):1020-1030. [doi: 10.1016/0735-1097(96)87734-5] [Medline: 8609317]
37. Shiroiwa T, Sung YK, Fukuda T, et al. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Econ 2010 Apr;19(4):422-437. [doi: 10.1002/hec.1481] [Medline: 19382128]
38. Towse A, Pritchard C, Devlin N. Cost-effectiveness thresholds: economic and ethical issues. OHE Monograph. 2002. URL: https://www.ohe.org/publications/cost-effectiveness-thresholds-economic-and-ethical-issues/ [accessed 2025-05-23]
39. Husereau D, Drummond M, Augustovski F, et al. Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Int J Technol Assess Health Care 2022 Jan 11;38(1):e13. [doi: 10.1017/S0266462321001732] [Medline: 35007499]
40. Lewkowicz D, Wohlbrandt AM, Bottinger E. Digital therapeutic care apps with decision-support interventions for people with low back pain in Germany: cost-effectiveness analysis. JMIR Mhealth Uhealth 2022 Feb 7;10(2):e35042. [doi: 10.2196/35042] [Medline: 35129454]
41. Nomura A, Tanigawa T, Kario K, et al. Cost-effectiveness of digital therapeutics for essential hypertension. Hypertens Res 2022 Oct;45(10):1538-1548. [doi: 10.1038/s41440-022-00952-x] [Medline: 35726085]
42. Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. Thyroid 2016 Oct;26(10):1343-1421. [doi: 10.1089/thy.2016.0229] [Medline: 27521067]
43. Bandai S, Okamura K, Fujikawa M, et al. The long-term follow-up of patients with thionamide-treated Graves' hyperthyroidism. Endocr J 2019 Jun 28;66(6):535-545. [doi: 10.1507/endocrj.EJ18-0418] [Medline: 30918165]
44. Azizi F, Yousefi V, Bahrainian A, et al. Long-term continuous methimazole or radioiodine treatment for hyperthyroidism. Arch Iran Med 2012 Aug;15(8):477-484. [Medline: 22827783]
45. Villagelin D, Romaldini JH, Santos RB, et al. Outcomes in relapsed Graves' disease patients following radioiodine or prolonged low dose of methimazole treatment. Thyroid 2015 Dec;25(12):1282-1290. [doi: 10.1089/thy.2015.0195] [Medline: 26414885]
46. Elbers L, Mourits M, Wiersinga W. Outcome of very long-term treatment with antithyroid drugs in Graves' hyperthyroidism associated with Graves' orbitopathy. Thyroid 2011 Mar;21(3):279-283. [doi: 10.1089/thy.2010.0181] [Medline: 21190446]
47. Chen DY, Jing J, Schneider PF, et al. Comparison of the long-term efficacy of low dose 131I versus antithyroid drugs in the treatment of hyperthyroidism. Nucl Med Commun 2009 Feb;30(2):160-168. [doi: 10.1097/MNM.0b013e3283134d4d] [Medline: 19194213]
48. Azizi F, Malboosbaf R. Long-term antithyroid drug treatment: a systematic review and meta-analysis. Thyroid 2017 Oct;27(10):1223-1231. [doi: 10.1089/thy.2016.0652] [Medline: 28699478]
49. Cooper DS. Long-term antithyroid drug treatment of patients with Graves' disease. Clin Thyroidol 2019 Jun;31(6):230-233. [doi: 10.1089/ct.2019;31.230-233]
50. Kahaly GJ, Bartalena L, Hegedüs L, et al. 2018 European Thyroid Association guideline for the management of Graves' hyperthyroidism. Eur Thyroid J 2018 Aug;7(4):167-186. [doi: 10.1159/000490384] [Medline: 30283735]
51. Adherence to long-term therapies:evidence for action. World Health Organization. 2003. URL: https://digicollections.net/medicinedocs/#d/s4883e [accessed 2024-06-28]
52. Etminani K, Göransson C, Galozy A, et al. Improving medication adherence through adaptive digital interventions (iMedA) in patients with hypertension: protocol for an interrupted time series study. JMIR Res Protoc 2021 May 12;10(5):e24494. [doi: 10.2196/24494] [Medline: 33978593]
53. Backes C, Moyano C, Rimaud C, et al. Digital medication adherence support: could healthcare providers recommend mobile health apps? Front Med Technol 2020;2:616242. [doi: 10.3389/fmedt.2020.616242] [Medline: 35047896]
54. Jácome C, Almeida R, Pereira AM, et al. Monitoring adherence to asthma inhalers using the InspirerMundi app: analysis of real-world, medium-term feasibility studies. Front Med Technol 2021;3:649506. [doi: 10.3389/fmedt.2021.649506] [Medline: 35047915]
55. Armitage LC, Kassavou A, Sutton S. Do mobile device apps designed to support medication adherence demonstrate efficacy? A systematic review of randomised controlled trials, with meta-analysis. BMJ Open 2020 Jan 30;10(1):e032045. [doi: 10.1136/bmjopen-2019-032045] [Medline: 32005778]

Abbreviations

ATD: antithyroid drug
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
GD: Graves disease
ICER: incremental cost-effectiveness ratio

XSL•FO
RenderX

NHIS: National Health Insurance Service
QALY: quality-adjusted life year
TAO: thyroid-associated orbitopathy
TFT: thyroid function test
WTP: willingness to pay

Edited by L Buis, TDA Cardoso; submitted 25.01.24; peer-reviewed by S Kim, W Zhou; revised version received 10.04.25; accepted 13.04.25; published 22.07.25.

Please cite as:
Kim JH, Shin J, Kim MS, Moon JH
A Digital Software Support Platform for Hyperthyroidism Management in South Korea: Markov Simulation Model-Based Cost-Effectiveness Analysis
JMIR Mhealth Uhealth 2025;13:e56738
URL: https://mhealth.jmir.org/2025/1/e56738
doi:10.2196/56738

© Jung Hyun Kim, Jaeyong Shin, Man S Kim, Jae Hoon Moon. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile App–Assisted Self-Monitoring of Blood Glucose in Type 2 Diabetes in Ningbo, China: 12-Month Retrospective Cohort Study

Xujia Ma[1*], BM; Kaushik Chattopadhyay[2*], PhD; Miao Xu[1], MD; Li Li[1], BM; Jialin Li[1], MM

[1]Department of Endocrinology and Metabolism, The First Affiliated Hospital of Ningbo University, Building 2, 22nd Fl., 59 Liuting Street, Ningbo, China

[2]Lifespan and Population Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Jialin Li, MM
Department of Endocrinology and Metabolism, The First Affiliated Hospital of Ningbo University, Building 2, 22nd Fl., 59 Liuting Street, Ningbo, China

Abstract

Background: Self-monitoring of blood glucose (SMBG) is recommended in clinical practice guidelines, including those in China, as part of patient education, self-management, and empowerment. With technological advancements, telecommunication technologies are now used for telemonitoring in health care. Mobile apps have become a practical tool for SMBG among patients with type 2 diabetes mellitus (T2DM). However, the long-term effectiveness of this approach in real-world practice requires further exploration.

Objective: The study aims to determine the effectiveness of mobile app–assisted SMBG in improving glycemic control in patients with T2DM at 12 months, in addition to standard care, in Ningbo, China.

Methods: In this retrospective cohort study, adults with T2DM who registered at the National Metabolic Management Center, Ningbo, for the first time between September 1, 2019, and June 30, 2022, and received standardized diabetes management were included. The study compared 2 groups: those who opted for mobile app–assisted SMBG and those who did not. Propensity score matching matched the mobile app–assisted SMBG group with the control group based on similar baseline characteristics. Glycemic control–related outcomes were compared at 12-month follow-up. Linear and logistic regression models were used to estimate mean differences and odds ratios (ORs) along with 95% CIs, respectively, and adjustments were made for baseline characteristics.

Results: A total of 160 patients (80 in each group) were included in the study. In the mobile app–assisted SMBG group, the median (IQR) frequency of blood glucose monitoring was 0 (0-2) times per week, with 28% (22/80) monitoring their blood glucose at least twice per week, and the app usage frequency was 1 (0-3) time per week, with 40% (32/80) logging in at least twice per week. There were no statistically significant differences observed between the mobile app–assisted SMBG group and the control group in glycemic control outcomes at 12 months. Specifically, the results showed no significant difference in (1) fasting blood glucose and glycosylated hemoglobin levels (mean difference −0.17 mmol/L, 95% CI −0.85 to 0.51 mmol/L; *P*=.62 and −0.12%, 95% CI −0.58% to 0.33%; *P*=.59, respectively) and (2) the proportion of patients achieving or maintaining fasting blood glucose at <7 mmol/L and glycosylated hemoglobin at <7% (OR 0.89, 95% CI 0.46-1.73; *P*=.74 and OR 0.91, 95% CI 0.44-1.88; *P*=.79, respectively).

Conclusions: In a real-world cohort of patients with T2DM in Ningbo, China, mobile app–assisted SMBG did not lead to statistically significant improvements in glycemic control at 12 months. This suggests that in a well-resourced setting, standard care alone may be relatively effective. However, opportunities for further improvement remain. The lack of observed benefit may be due to process-related issues, such as suboptimal engagement with the intervention. Addressing these challenges should be a focus of future research.

(JMIR Mhealth Uhealth 2025;13:e65919) doi:10.2196/65919

KEYWORDS

blood glucose; China; cohort study; mobile app; self-management; self-monitoring; type 2 diabetes

XSL•FO
RenderX

Introduction

Background

China has the highest prevalence of type 2 diabetes mellitus (T2DM) in the world, with approximately 130 million (11%) adults affected by T2DM [1]. T2DM is a complex metabolic disorder with significant implications for health, society, and the economy [2,3]. The persistent elevation of blood glucose levels in T2DM is associated with numerous complications, diminished quality of life, and heightened mortality rates [4-6]. T2DM clinical practice guidelines recommend adequate glycemic control, aiming for a fasting blood glucose level of <7 mmol/L and glycosylated hemoglobin (HbA_{1c}) level of <7% in most patients [7,8].

The role of self-monitoring of blood glucose (SMBG) in patients with T2DM, especially those not on insulin or at lower risk of hypoglycemia, remains debatable [9-13]. However, many clinical practice guidelines, including those in China, recommend SMBG as part of patient education, self-management, and empowerment [14,15]. The most common method for SMBG is the finger-stick capillary blood glucose test using a glucometer [14]. The frequency of SMBG depends on several factors, including the patient's blood glucose level, treatment regimen, and risk of hypoglycemia [14].

With technological advancements, various telecommunication technologies are now used for telemonitoring in health care [16]. Systematic reviews suggest that telemonitoring interventions may effectively improve glycemic control in patients with T2DM, although more primary studies are needed to evaluate the effectiveness of specific intervention delivery modes [17,18]. In China, mobile apps are used for SMBG in real clinical practice, including for patients with T2DM [19]. However, the effectiveness of this approach, especially in real-world practice and over the long term, required further exploration [20].

Although Ningbo is one of China's most economically developed cities [21] and had a T2DM prevalence of approximately 10% in 2018 [22], glycemic control among patients has remained suboptimal. For example, a community-based study of patients with T2DM in Ningbo reported an average HbA_{1c} level of 59.6 mmol/mol (7.6%) [23]. Similarly, a previous study conducted at our tertiary care diabetes center—before the implementation of standardized care under the National Metabolic Management Center (MMC) program—found that only half of the patients achieved the HbA_{1c} target of <7% after 6 months of treatment [24]. These findings reinforce that despite the availability of health care resources, glycemic targets are frequently unmet in routine clinical practice.

Objectives

This study aimed to evaluate whether the addition of mobile app–assisted SMBG to standard care could improve glycemic control over 12 months in patients with T2DM in Ningbo, China.

Methods

Study Design, Location, Population, and Baseline Characteristics

This retrospective cohort study used data from the MMC at The First Affiliated Hospital of Ningbo University. This MMC is the largest and main center in Ningbo and is one of the two provincial-level MMCs in the Zhejiang Province [25]. Adults (18-75 y) with T2DM who registered at this MMC for the first time between September 1, 2019, and June 30, 2022, and received standardized diabetes management were included. Baseline characteristics included sociodemographic information (age, sex, education, and household income) and clinical information (such as duration of T2DM [self-reported], blood glucose levels, and insulin regimen).

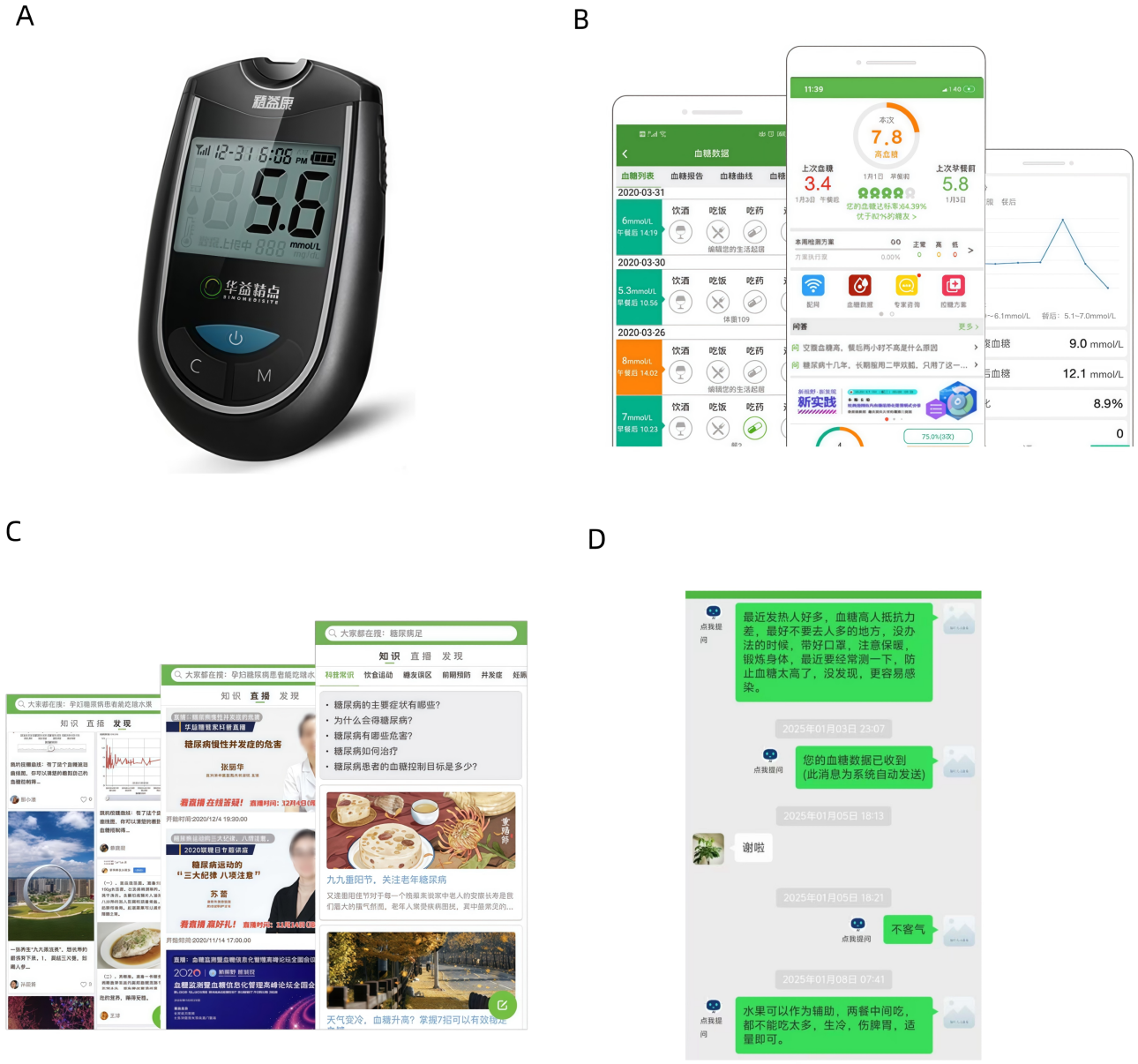
Exposure

All patients received standard care under the MMC program, which involved standardized diabetes management. SMBG and its importance were explained during the first visit, as recommended by the Chinese T2DM clinical practice guidelines [7]. Some patients voluntarily purchased a mobile app–assisted glucometer for self-monitoring (EZ-8 SIM Blood Glucose Meter with the Huayi Glucose Manager app, SINOMEDISITE, China), forming the mobile app–assisted SMBG group in the study. The device and app bundle cost approximately 200 RMB (about US $28), including the initial supply of test strips. The hospital had no financial or contractual ties to the manufacturer and did not offer direct technical support. Clinicians introduced the app during routine visits as a voluntary, supplementary tool. Those who did not opt for the app-assisted SMBG formed the control group.

The EZ-8 SIM Blood Glucose Meter is a compact device designed to measure capillary blood glucose levels, displaying results in millimoles per liter on a digital screen (see Figure 1A). It includes basic navigation buttons and connects to the Huayi Glucose Manager app via Bluetooth for data transmission.

XSL•FO
RenderX

Figure 1. (A) EZ-8 SIM blood glucose meter. (B) Huayi Glucose Manager app blood glucose data visualization interface. (C) App educational content interface. (D) App clinician communication interface.



The Huayi Glucose Manager app is a mobile-based telemonitoring platform intended to support SMBG in patients with T2DM. Blood glucose readings are automatically transmitted from the glucometer to the app and categorized as in-range (green), high (yellow), or low (red), with accompanying graphical displays (see Figure 1B). These features are designed to assist patients and clinicians in tracking glycemic trends in real time.

Patients can interact with the app by logging in multiple times per week. According to Chinese guidelines, SMBG is recommended twice weekly for patients on oral antidiabetic drugs (OADs), whereas those receiving insulin therapy are advised to monitor blood glucose levels more frequently throughout the day [7]. The app includes additional features intended to support self-management, such as prerecorded educational sessions (see Figure 1C) covering T2DM, complications associated with T2DM, and diet and lifestyle programs. Users can set lifestyle goals, track progress, and view changes in blood glucose levels corresponding to behavioral modifications. The app also allows direct communication between patients and clinicians (see Figure 1D).

Clinicians can access patient data and blood glucose readings via a QR code linked to the app, facilitating tailored advice, monitoring reminders, and secure communication. The app further provides a live-stream feature, enabling clinicians to conduct patient education sessions related to T2DM.

Outcomes

Outcomes at 12-month follow-up included fasting blood glucose levels, the proportion of patients achieving or maintaining a fasting blood glucose level of <7 mmol/L, HbA_{1c} levels, and the proportion of patients achieving or maintaining an HbA_{1c} level of <7%. Fasting blood glucose was assessed using the glucose oxidase method. HbA_{1c} was estimated using the high-performance liquid chromatography method with the D-10 Haemoglobin Analyzer (Bio-Rad).

XSL•FO
RenderX

Ethical Considerations

This study was approved by the Research Ethics Committee of The First Affiliated Hospital of Ningbo University (2019-R057). All patients, irrespective of future group assignment, provided written informed consent at the time of registration at the MMC, permitting the use of their routinely collected clinical data for research purposes. No additional consent was sought after the matching process, as this study involved retrospective analysis of anonymized data. To ensure patient privacy and confidentiality, all personal identifying information (such as names, hospital record numbers, and contact information) was removed from the dataset before analysis. The research team only had access to the fully anonymized data. Participants did not receive any compensation for their data being included in this study.

Statistics

Among patients with baseline and 12-month follow-up data on blood glucose levels (80 in the mobile app–assisted SMBG group; 1490 in the control group), propensity score matching was performed using R (version 4.1.2; *MatchIt* package, R Foundation for Statistical Computing). This process paired each patient in the mobile app–assisted SMBG group with a patient in the control group based on similar baseline characteristics [26]. Once a match was made, the patients were not reconsidered for further matching. Absolute standardized differences were calculated for baseline characteristics, with a value of <10% generally regarded as a marker of acceptable balance [27]. A multivariable regression model was used to calculate the propensity score. A 1:1 ratio matching algorithm with nearest-neighbor matching and no replacement was used, with a caliper width of 0.1 SD.

For categorical data, numbers and percentages were calculated. For continuous data, summary measures (mean or median and spread) were calculated depending on data distribution. Linear and logistic regression models were used to estimate unadjusted and adjusted mean differences (MDs) and odds ratios (ORs) along with 95% CIs, respectively. Adjustments were made for baseline characteristics. Data were analyzed using SPSS Statistics for Windows (version 26.0; IBM Corp).

Results

A total of 160 patients (80 in each group) were included in the matched cohort. At baseline, the mean (SD) age of patients was 47.13 (10.56) years, with 66 (41.3%) patients being female. The mean (SD) values for BMI, fasting blood glucose, and HbA_{1c} levels were 25.88 (4.06) kg/m², 9.17 (2.96) mmol/L, and 8.54% (2.36%), respectively. Regarding diabetes treatment, 71.9% (115/160) of the patients were treated with OADs alone, whereas 28.1% (45/160) of the patients were on insulin regimens, either alone or in combination with OADs. Diabetic microvascular complications were present in 53.8% (86/160) of the patients, whereas diabetic macrovascular complications were present in 4.4% (7/160) of the patients. The baseline characteristics of the 2 groups are shown in Table 1 (for characteristics before matching, see Table S1 in Multimedia Appendix 1). The 2 groups were similar in most baseline characteristics, with absolute standardized differences <10%, except for the duration of T2DM, which had a difference of 15.57%.

XSL•FO
RenderX

Table . Baseline characteristics of patients in the matched cohort (N=160).

	Mobile app–assisted SMBG[a] group (n=80)	Control group (n=80)	Absolute standardized difference (%)	*P* value
Sociodemographic characteristics				
Age (years), mean (SD)	46.79 (10.04)	47.48 (11.10)	6.85	.68
Sex, n (%)				>.99
Male	47 (58.75)	47 (58.75)	0	
Female	33 (41.25)	33 (41.25)	0	
Education, n (%)				>.99
<High school	34 (42.50)	34 (42.50)	0	
≥High school	46 (57.50)	46 (57.50)	0	
Household income (RMB[b]), n (%)				.46
≤100,000	29 (36.25)	31 (38.75)	5.20	
>100,000 - 300,000	30 (37.50)	29 (36.25)	2.58	
>300,000	18 (22.50)	20 (25.00)	5.99	
Unknown	3 (3.75)	0 (0)	19.74	
Clinical characteristics				
Duration of T2DM[c] (months), median (IQR)	38.50 (15.50-69.75)	52.50 (12.00-107.00)	15.57	.10
Fasting blood glucose (mmol/L), mean (SD)	9.20 (2.94)	9.14 (3.00)	2.06	.90
HbA_{1c}[d] (%), mean (SD)	8.57 (2.33)	8.50 (2.41)	3.01	.85
Insulin regimen, n (%)				.60
No	59 (73.75)	56 (70.00)	8.52	
Yes	21 (26.25)	24 (30.00)	8.52	

[a]SMBG: self-monitoring of blood glucose.

[b]RMB: Chinese Yuan (1 RMB=US $0.14).

[c]T2DM: type 2 diabetes mellitus.

[d]HbA_{1c}: glycosylated hemoglobin.

In the mobile app–assisted SMBG group, the median (IQR) frequency of blood glucose monitoring was 0 (0-2) times per week, with 28% (22/80) of patients monitoring their blood glucose at least twice per week. The median (IQR) app usage frequency was 1 (0-3) time per week, with 40% (32/80) of patients logging in at least twice per week.

Figure 2 shows unadjusted outcomes at 12 months. The mean (SD) fasting blood glucose in the mobile app–assisted SMBG and control groups was 7.09 (2.09) mmol/L and 7.31 (2.38) mmol/L, respectively (MD −0.21 mmol/L, 95% CI −0.91 to 0.49 mmol/L; *P*=.55). The proportion of patients achieving or maintaining a fasting blood glucose level of <7 mmol/L in the mobile app–assisted SMBG and control groups was 60% (48/80) and 61% (49/80), respectively (OR 0.95, 95% CI 0.50-1.79; *P*=.87). The mean (SD) of HbA_{1c} in the mobile app–assisted SMBG and control groups was 6.78% (1.51%) and 6.91% (1.61%), respectively (MD −0.13%, 95% CI −0.62% to 0.36%; *P*=.59). The proportion of patients achieving or maintaining an HbA_{1c} level of <7% in the mobile app–assisted SMBG and control groups was 68% (54/80) and 69% (55/80), respectively (OR 0.94, 95% CI 0.49-1.84; *P*=.87).

XSL•FO
RenderX

Figure 2. Unadjusted outcomes at 12 months. (A) FBG levels. (B) Proportion of patients achieving or maintaining an FBG level of <7 mmol/L. (C) HbA_{1c} levels. (D) Proportion of patients achieving or maintaining an HbA_{1c} level of <7%. FBG: fasting blood glucose; HbA_{1c}: glycosylated hemoglobin; SMBG: self-monitoring of blood glucose.

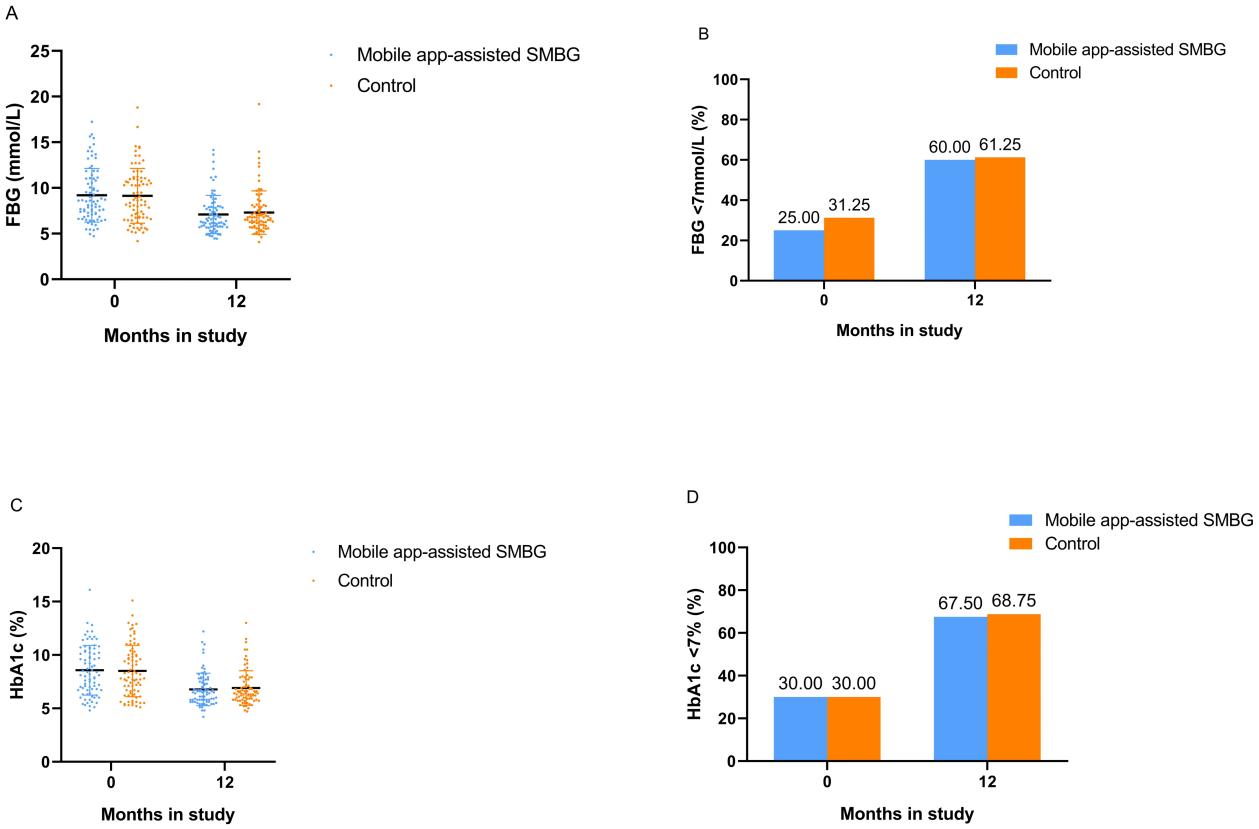


Table 2 reports adjusted outcomes at 12 months. There were no statistically significant differences observed between the mobile app–assisted SMBG group and the control group in glycemic control outcomes at 12 months. Specifically, the results showed (1) no significant difference in fasting blood glucose and HbA_{1c} levels (MD −0.17 mmol/L, 95% CI −0.85 to 0.51 mmol/L; *P*=.62 and −0.12%, 95% CI −0.58% to 0.33%; *P*=.59, respectively) and (2) no significant difference in the proportion of patients achieving or maintaining a fasting blood glucose level of <7 mmol/L and an HbA_{1c} level of <7% (OR 0.89, 95% CI 0.46-1.73; *P*=.74 and OR 0.91, 95% CI 0.44-1.88; *P*=.79, respectively). An additional data analysis at 6 months is provided in Table S2 in Multimedia Appendix 1. At 6 months, the mobile app–assisted SMBG group showed a reduction in HbA_{1c} compared with the control group (adjusted MD −0.43%, 95% CI −0.86% to 0%; *P*=.05), reaching borderline statistical significance. No significant differences were observed for fasting blood glucose levels or target achievement rates.

XSL•FO
RenderX

Table . Adjusted outcomes at 12 months[a].

	Adjusted MD[b]/OR[c] (95% CI)	*P* value
Fasting blood glucose (mmol/L)	−0.17 (−0.85 to 0.51)	.62
Fasting blood glucose <7 mmol/L		
Control	Ref	—[d]
Mobile app–assisted SMBG[e]	0.89 (0.46-1.73)	.74
HbA_{1c}[f] (%)	−0.12 (−0.58 to 0.33)	.59
HbA_{1c} <7%		
Control	Ref	—
Mobile app–assisted SMBG	0.91 (0.44-1.88)	.79

[a]Adjusted for age, sex, education, household income, duration of type 2 diabetes mellitus, baseline fasting blood glucose/HbA_{1c}, and insulin regimen.

[b]MD: mean difference.

[c]OR: odds ratio.

[d]Not applicable.

[e]SMBG: self-monitoring of blood glucose.

[f]HbA_{1c}: glycosylated hemoglobin.

Discussion

Our study assessed the long-term effectiveness of mobile app–assisted SMBG in improving glycemic control in patients with T2DM at 12 months, alongside standard care, in Ningbo, China. Although the mobile app–assisted SMBG group showed lower fasting blood glucose and HbA_{1c} levels compared with the control group, these differences were not statistically significant. Similarly, the proportion of patients achieving or maintaining a fasting blood glucose level of <7 mmol/L and an HbA_{1c} level of <7% was lower in the intervention group; however, these differences did not reach statistical significance.

Despite the lack of statistical significance, our findings offer valuable insights into the real-world application of mobile app–assisted SMBG. While standard care in this well-resourced setting has been shown to be effective [28-30], opportunities for further improvement remain. In our study, the supplementary mobile app–assisted SMBG did not yield clear benefits, likely due to process-related factors such as limited engagement, rather than a ceiling effect. Addressing these implementation challenges should be a key focus of future research. Specifically, previous studies have demonstrated that standard care at MMC effectively manages glycemic control in patients with T2DM [28-30]. Similarly, in our study, at baseline (ie, the point of registration at MMC for standardized diabetes care), the mean fasting blood glucose in both groups exceeded 9 mmol/L. By 12 months, this had decreased to approximately 7 mmol/L. Likewise, the mean HbA_{1c} at baseline was ≥8.5%, which declined to <7% at 12 months. These findings suggest that MMC's standardized care may be effective in improving glycemic control in individuals with T2DM. Patients in the intervention group monitored their blood glucose less frequently than the approximately 2 times per day typically associated with meaningful glycemic improvements [31], which may have limited the intervention's impact. The supplementary 6-month analysis showed a borderline significant reduction in HbA_{1c} in the intervention group, but this was not sustained at 12 months. This trend likely reflects a decline in engagement over time—a common challenge in digital health interventions—and aligns with previous studies reporting reduced effectiveness of mobile app–assisted SMBG as usage wanes [32-35]. Furthermore, given that our study participants had a median diabetes duration of 3 to 4 years, they were likely not newly diagnosed and may have already been familiar with SMBG through standard care, potentially reducing the added benefit of mobile app–assisted SMBG. Research indicates that mobile app–assisted SMBG is most beneficial for patients newly diagnosed with T2DM, particularly within the first 3 to 6 months, when treatment adjustments are most critical [34,36,37].

Systematic reviews examining factors associated with the uptake, engagement, and retention of mobile apps, including those used for self-management of chronic conditions such as diabetes, have identified a range of influences aligned with the Capability, Opportunity, Motivation for Behavior model. These include challenges such as usability challenges, technical difficulties, lack of customizable features, absence of supportive functionalities (eg, reminders or feedback), and limited perceived usefulness of the apps [38-40]. In addition, a previous study evaluating Chinese mobile apps for diabetes self-management identified concerns related to app quality, functionality, and features and recommended improvements through co-design involving researchers, health care professionals, and target users [41]. While it is possible that the developers of the mobile app used in our study received some informal feedback from key stakeholders, we are not aware of any systematic co-development approach undertaken.

To the best of our knowledge, this is the first study to assess the long-term (12 mo) effectiveness of mobile app–assisted SMBG in a real-world cohort of patients with T2DM in Ningbo, China. Unlike randomized controlled trials, which are rigorously

XSL•FO
RenderX

controlled but may not fully reflect the complexities of real-world clinical settings [42,43], our study leveraged real-world data from routine clinical practice within a high-quality treatment environment. Propensity score matching was used to create a balanced control group and address potential confounders, with known confounders further adjusted for in the analysis. Despite this, the absolute standardized difference in diabetes duration between groups remained above the accepted 10% threshold, with the intervention group having a shorter average duration. This imbalance may have conferred an advantage, as shorter diabetes duration is typically associated with better glycemic control [24], and mobile app interventions tend to be more effective in individuals with more recent diagnoses [44]. Nevertheless, we did not observe statistically significant improvements in glycemic control in the intervention group, even with this potential bias in its favor.

Several other limitations should also be acknowledged. First, the small sample size reduced statistical power, which may have limited our ability to detect subtle differences. As this was a retrospective analysis of existing data, no formal sample size calculation was performed. Instead, we included all eligible patients (n=1570) and applied propensity score matching to design this retrospective cohort study (n=160). Because patients voluntarily purchased the mobile app–assisted glucometers, the intervention group may have differed from the control group, introducing potential confounding factors. Apart from the app development and engagement issues discussed earlier, the study encountered additional process-related limitations due to restricted dataset access granted by the app developers for research purposes. First, although we reported the overall frequency of app engagement by patients, including SMBG, the dataset available to us did not include longitudinal data needed to evaluate usage patterns over time. This constrained our ability to assess whether declining engagement may have contributed to the waning effectiveness of the intervention. In addition, we did not have access to data on how clinicians and patients interacted with specific app features—such as doctor-patient communication, delivery of clinical advice, or patient self-management—limiting insights into the mechanisms of engagement. Safety data were also unavailable in this retrospective study, as they were not routinely collected, limiting the comprehensive evaluation of the intervention. Because the study was conducted in a highly specialized, resource-intensive setting, the generalizability of our findings to lower-resource environments may be limited. There are inherent trade-offs in real-world observational studies, which prioritize practical applicability and relevance over the controlled conditions typically found in randomized controlled trials. Future large-scale, prospective studies are needed to evaluate high-quality mobile app–assisted SMBG interventions, incorporating strategies to optimize engagement and focusing on both long-term effectiveness and safety. Causal inference frameworks such as the Bradford Hill criteria or the 9 principles for determining causality in epidemiology should be applied to strengthen the evidence base [45].

In conclusion, in a real-world cohort of patients with T2DM in Ningbo, China, mobile app–assisted SMBG did not lead to statistically significant improvements in glycemic control at 12 months. This suggests that in a well-resourced setting, standard care alone may be relatively effective. However, opportunities for further improvement remain. The lack of observed benefit may be due to process-related issues, such as suboptimal engagement with the intervention. Addressing these challenges should be a focus of future research.

Acknowledgments

The authors thank Yanshu Chen and Miao Chen for managing the Metabolic Management Center database. This work was supported by the Medical Health Science and Technology Project of Zhejiang Province (grant 2024KY1506) and the Major Science and Technology Projects for Health of Zhejiang Province (grant WKJ-ZJ-2216).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JL and KC designed the study. XM extracted and cleaned the data. XM and KC analyzed the data and wrote the first draft of the manuscript. All the authors revised it critically for important intellectual content and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables reporting patients' baseline characteristics and their unadjusted and adjusted outcomes at 6 months.
[DOCX File, 21 KB - mhealth_v13i1e65919_app1.docx]

References

XSL•FO
RenderX

1. Li Y, Teng D, Shi X, et al. Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study. BMJ 2020 Apr 28;369:m997. [doi: 10.1136/bmj.m997] [Medline: 32345662]
2. Liu X, Zhang L, Chen W. Trends in economic burden of type 2 diabetes in China: based on longitudinal claim data. Front Public Health 2023 Apr;11:1062903. [doi: 10.3389/fpubh.2023.1062903]
3. Magliano DJ, Boyko EJ, IDF Diabetes Atlas 10th Edition Scientific Committee. IDF Diabetes Atlas, 10th edition: International Diabetes Federation; 2021. URL: https://www.ncbi.nlm.nih.gov/books/NBK581934/ [accessed 2025-08-11]
4. Li J, Chattopadhyay K, Xu M, et al. Prevalence and associated factors of vascular complications among inpatients with type 2 diabetes: a retrospective database study at a tertiary care department, Ningbo, China. PLoS One 2020;15(6):e0235161. [doi: 10.1371/journal.pone.0235161]
5. Carter N, Li J, Xu M, et al. Health-related quality of life of people with type 2 diabetes and its associated factors at a tertiary care clinic in Ningbo, China: a cross-sectional study. Endocrinol Diabetes Metab 2022 Sep;5(5):e353. [doi: 10.1002/edm2.353] [Medline: 35780497]
6. Chan JCN, Lim LL, Wareham NJ, et al. The Lancet Commission on diabetes: using data to transform diabetes care and patient lives. Lancet 2021 Dec 19;396(10267):2019-2082. [doi: 10.1016/S0140-6736(20)32374-6] [Medline: 33189186]
7. Chinese Diabetes Society. Guideline for the prevention and treatment of diabetes mellitus in China (2024 edition). Chin J Diabetes Mellitus 2025 Jan;17(1):16-139. [doi: 10.3760/cma.j.cn115791-20241203-00705]
8. Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2022 Nov 1;45(11):2753-2786. [doi: 10.2337/dci22-0034] [Medline: 36148880]
9. Type 2 diabetes in adults: management. National Institute for Health and Care Excellence. 2022 Jun 29. URL: https://www.nice.org.uk/guidance/ng28 [accessed 2025-08-11]
10. Chircop J, Sheffield D, Kotera Y. Systematic review of self-monitoring of blood glucose in patients with type 2 diabetes. Nurs Res 2021;70(6):487-497. [doi: 10.1097/NNR.0000000000000542] [Medline: 34292228]
11. Holmes-Truscott E, Baptista S, Ling M, et al. The impact of structured self-monitoring of blood glucose on clinical, behavioral, and psychosocial outcomes among adults with non-insulin-treated type 2 diabetes: a systematic review and meta-analysis. Front Clin Diabetes Healthc 2023;4:1177030. [doi: 10.3389/fcdhc.2023.1177030] [Medline: 37153750]
12. Zou Y, Zhao S, Li G, Zhang C. The efficacy and frequency of self-monitoring of blood glucose in non-insulin-treated T2D patients: a systematic review and meta-analysis. J Gen Intern Med 2023 Feb;38(3):755-764. [doi: 10.1007/s11606-022-07864-z] [Medline: 36403159]
13. Machry RV, Rados DV, de Gregório GR, Rodrigues TC. Self-monitoring blood glucose improves glycemic control in type 2 diabetes without intensive treatment: a systematic review and meta-analysis. Diabetes Res Clin Pract 2018 Aug;142:173-187. [doi: 10.1016/j.diabres.2018.05.037] [Medline: 29857093]
14. Bao Y, Zhu D, Chinese Diabetes Society. Clinical application guidelines for blood glucose monitoring in China (2022 edition). Diabetes Metab Res Rev 2022 Nov;38(8):e3581. [doi: 10.1002/dmrr.3581] [Medline: 36251516]
15. American Diabetes Association Professional Practice Committee. 7. Diabetes technology: standards of care in diabetes-2024. Diabetes Care 2024 Jan 1;47(Suppl 1):S126-S144. [doi: 10.2337/dc24-S007] [Medline: 38078575]
16. Peralta-Ochoa AM, Chaca-Asmal PA, Guerrero-Vásquez LF, Ordoñez-Ordoñez JO, Coronel-González EJ. Smart healthcare applications over 5G networks: a systematic review. Appl Sci (Basel) 2023 Jan;13(3):1469. [doi: 10.3390/app13031469]
17. Kim Y, Park JE, Lee BW, Jung CH, Park DA. Comparative effectiveness of telemonitoring versus usual care for type 2 diabetes: a systematic review and meta-analysis. J Telemed Telecare 2019 Dec;25(10):587-601. [doi: 10.1177/1357633X18782599] [Medline: 30012042]
18. Emonena H, Ojo O. The efficacy of tele-monitoring in maintaining glycated haemoglobin levels in patients with type 2 diabetes mellitus: a systematic review. Int J Environ Res Public Health 2022 Dec 13;19(24):16722. [doi: 10.3390/ijerph192416722] [Medline: 36554601]
19. Hsu J, Liu D, Yu YM, et al. The top Chinese mobile health apps: a systematic investigation. J Med Internet Res 2016 Aug 29;18(8):e222. [doi: 10.2196/jmir.5955] [Medline: 27573724]
20. Han CY, Zhang J, Ye XM, et al. Telemedicine-assisted structured self-monitoring of blood glucose in management of T2DM results of a randomized clinical trial. BMC Med Inform Decis Mak 2023 Sep 14;23(1):182. [doi: 10.1186/s12911-023-02283-4] [Medline: 37710226]
21. National data: annual data for major cities. National Bureau of Statistics of China. 2022. URL: https://data.stats.gov.cn/easyquery.htm?cn=E0105 [accessed 2024-08-25]
22. Du F, Feng W, Wang C. Epidemiological status and influence factors of diabetes mellitus in Fenghua, Ningbo. Chinese Rural Health Service Administration 2018 Mar;38(1):40-42. [doi: 10.3969/j.issn.1005-5916.2018.01.015]
23. Li QQ, Liang JY, Wang JM, et al. Applications of the NDR and DIAL models for risk prediction on cardiovascular disease in patients with type 2 diabetes in Ningbo. Zhonghua Liu Xing Bing Xue Za Zhi 2022 Jun 10;43(6):945-952. [doi: 10.3760/cma.j.cn112338-20211116-00891] [Medline: 35725354]

XSL•FO
RenderX

24. Li J, Chattopadhyay K, Xu M, et al. Glycaemic control in type 2 diabetes patients and its predictors: a retrospective database study at a tertiary care diabetes centre in Ningbo, China. BMJ Open 2018 Mar 25;8(3):e019697. [doi: 10.1136/bmjopen-2017-019697] [Medline: 29581203]
25. Our hospital was awarded the honorary titles of “provincial management center for standardized metabolic disease management center” and “provincial demonstration center”. The First Affiliated Hospital of Ningbo University. 2021 Nov 4. URL: http://www.nbdyyy.com/art/2021/11/4/art_80_633223.html [accessed 2024-08-25]
26. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May;46(3):399-424. [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]
27. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009 Nov 10;28(25):3083-3107. [doi: 10.1002/sim.3697] [Medline: 19757444]
28. Wang X, Zhao W, Ke J, Zhao D. Comparison and analyses of therapeutic effects between metabolic management center (MMC) and conventional management modes. Acta Diabetol 2020 Sep;57(9):1049-1056. [doi: 10.1007/s00592-020-01518-4] [Medline: 32248347]
29. Li X, Bai X, Wang J, et al. Metabolic outcomes in patients with type 2 diabetes mellitus after metabolic management center. Medicine (Abingdon) 2022;101(45):e31342. [doi: 10.1097/MD.0000000000031342]
30. Yang J, Gong Y, Lai M, Li N, Zhang A, Gu L, et al. Metabolic outcomes of type 2 diabetes patients with different diabetic durations under the standardized metabolic disease management model. Chin J Endocrinol Metab 2021 Aug;37(2):106-113. [doi: 10.3760/cma.j.cn311282-20200315-00168]
31. Montero AR, Toro-Tobon D, Gann K, Nassar CM, Youssef GA, Magee MF. Implications of remote monitoring technology in optimizing traditional self-monitoring of blood glucose in adults with T2DM in primary care. BMC Endocr Disord 2021 Nov 10;21(1):222. [doi: 10.1186/s12902-021-00884-6] [Medline: 34758807]
32. Nahum-Shani I, Yoon C. Towards the science of engagement with digital interventions. Curr Dir Psychol Sci 2024 Aug;33(4):239-246. [doi: 10.1177/09637214241254328] [Medline: 39552747]
33. Li J, Sun L, Wang Y, et al. A mobile-based intervention for glycemic control in patients with type 2 diabetes: retrospective, propensity score-matched cohort study. JMIR Mhealth Uhealth 2020 Mar 11;8(3):e15390. [doi: 10.2196/15390] [Medline: 32159518]
34. Young LA, Buse JB, Weaver MA, et al. Glucose self-monitoring in non-insulin-treated patients with type 2 diabetes in primary care settings: a randomized trial. JAMA Intern Med 2017 Jul 1;177(7):920-929. [doi: 10.1001/jamainternmed.2017.1233] [Medline: 28600913]
35. Lee EY, Cha SA, Yun JS, et al. Efficacy of personalized diabetes self-care using an electronic medical record-integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial. J Med Internet Res 2022 Jul 28;24(7):e37430. [doi: 10.2196/37430] [Medline: 35900817]
36. Yang Y, Lee EY, Kim HS, Lee SH, Yoon KH, Cho JH. Effect of a mobile phone-based glucose-monitoring and feedback system for type 2 diabetes management in multiple primary care clinic settings: cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Feb 26;8(2):e16266. [doi: 10.2196/16266] [Medline: 32130172]
37. Johari SM, Razalli NH, Chua KJ, Shahar S. The efficacy of self-monitoring of blood glucose (SMBG) intervention package through a subscription model among type-2 diabetes mellitus in Malaysia: a preliminary trial. Diabetol Metab Syndr 2024 Jun 21;16(1):135. [doi: 10.1186/s13098-024-01379-9] [Medline: 38902819]
38. Patail K, Pai HH, Greenfield G, et al. What are the perceptions and experiences of adults using mobile applications for self-management in diabetes? A systematic review. BMJ Open 2025 Jan 20;15(1):e086671. [doi: 10.1136/bmjopen-2024-086671] [Medline: 39833008]
39. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
40. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572. [doi: 10.2196/17572] [Medline: 32348255]
41. Gong E, Zhang Z, Jin X, et al. Quality, functionality, and features of Chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. JMIR Mhealth Uhealth 2020 Apr 7;8(4):e14836. [doi: 10.2196/14836] [Medline: 32255432]
42. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000 Jun 22;342(25):1887-1892. [doi: 10.1056/NEJM200006223422507] [Medline: 10861325]
43. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med 2018 Aug;210:2-21. [doi: 10.1016/j.socscimed.2017.12.005] [Medline: 29331519]
44. Chang YT, Tu YZ, Chiou HY, Lai K, Yu NC. Real-world benefits of diabetes management app use and self-monitoring of blood glucose on glycemic control: retrospective analyses. JMIR Mhealth Uhealth 2022 Jun 15;10(6):e31764. [doi: 10.2196/31764] [Medline: 35704371]
45. Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965 May;58(5):295-300. [doi: 10.1177/003591576505800503] [Medline: 14283879]

XSL•FO
RenderX

Abbreviations

HbA_{1c}: glycosylated hemoglobin
MD: mean difference
MMC: Metabolic Management Center
OAD: oral antidiabetic drug
OR: odds ratio
SMBG: self-monitoring of blood glucose
T2DM: type 2 diabetes mellitus

Edited by M DeJonckheere; submitted 30.08.24; peer-reviewed by A Wägner, Z Qin; revised version received 08.06.25; accepted 22.07.25; published 02.09.25.

Please cite as:
Ma X, Chattopadhyay K, Xu M, Li L, Li J
Mobile App–Assisted Self-Monitoring of Blood Glucose in Type 2 Diabetes in Ningbo, China: 12-Month Retrospective Cohort Study
JMIR Mhealth Uhealth 2025;13:e65919
URL: https://mhealth.jmir.org/2025/1/e65919
doi:10.2196/65919

© Xujia Ma, Kaushik Chattopadhyay, Miao Xu, Li Li, Jialin Li. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of a 24-Week Mobile App–Based Human Coaching Program on Body Composition and Lipid Metabolism in Breast Cancer Survivors With Overweight or Obesity: Single-Arm Prospective Cohort Study

Eun-Gyeong Lee[1,2], MD, MSc; Dong-Eun Lee[3], PhD; Jeongseon Kim[1], PhD; Jaihong Han[2], MD; Seeyoun Lee[2], MD, PhD; Han-Sung Kang[2], MD, PhD; Eun Sook Lee[2], MD, PhD; Heejung Chae[2], MD, PhD; Sung Hoon Sim[2], MD, PhD; Keun Seok Lee[2], MD, PhD; Jungeun Lee[4], PhD; Hyun Jeong Lee[5], MD, PhD; Ji Sung Yoo[6], MD, PhD; Gyung Ah Wie[7], PhD; So-Youn Jung[1,2], MD, PhD

[1]Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, Republic of Korea

[2]Center for Breast Cancer, National Cancer Center, 323 Ilsan-ro, Ilsandong-gu, Goyang, Republic of Korea

[3]Biostatistics Collaboration Team, Research Core Center, Research Institute of National Cancer Center, Goyang, Republic of Korea

[4]Noom, Inc., New York, NY, United States

[5]Department of Psychiatry and Behavioral Science, National Cancer Center, Goyang, Republic of Korea

[6]Department of Rehabilitation Medicine, Research Institute and Hospital, National Cancer Center, Goyang, Republic of Korea

[7]Department of Clinical Nutrition, National Cancer Center, Goyang, Republic of Korea

Corresponding Author:
So-Youn Jung, MD, PhD
Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, Republic of Korea

Abstract

Background: Overweight or obesity is a prognostic factor for breast cancer recurrence and breast cancer–related deaths. However, weight control is difficult for breast cancer survivors because of menopause, chemotherapy, antihormonal therapy, and psychological issues.

Objective: This study aimed to develop a 24-week mobile app–based human coaching program using Noom and evaluate its efficacy in breast cancer survivors who are excessively overweight or with obesity, including those who successfully used the program.

Methods: In this single-arm prospective cohort study, 130 breast cancer survivors with BMI ≥25 were enrolled and received a 24-week program, including diet-, exercise-, and psychology-based content with the trained human coach in Noom between 2019 and 2021. For a hyperactive group who joined for more than 16 weeks, we evaluated weight, BMI, lipid level, bioimpedance, and quality of life at baseline, 6-month, and 12-month follow-up.

Results: Among 130 breast cancer survivors, 101 (77.7%) and 93 (71.5%) completed the 6-month and 12-month follow-ups, respectively. The mean age of all participants was 54.90 (SD 7.42) years. At baseline, the median BMI was 27.14 (IQR 25.20 - 35.36) for the hyperactive group and 27.50 (IQR 25.20 - 35.50) for the active and inactive group. In the hyperactive group (68/101, 67%), body weight and BMI significantly reduced (mean difference −1.97, 95% CI −2.65 to −1.26 kg; P<.01 and mean difference −0.86, 95% CI −1.15 to −0.56; P<.01, respectively) at 6 months and were maintained at 12 months without the yo-yo effect. Among the lipid panel, triglyceride levels decreased significantly (−34.13, 95% CI −58.09 to−10.17; P<.01) and were maintained at 12 months. With respect to bioimpedance components, skeletal muscle mass (kg), body fat mass (kg), percent body fat (%), waist-to-hip ratio, and visceral fat area (cm^2) improved in the first 6 months. However, waist-to-hip ratio and visceral fat area increased during the next 6 months. Based on the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ C30) and Breast Cancer Module (23 items), nausea or vomiting, constipation, body image, and arm and breast symptoms significantly improved during the first 6 months.

Conclusions: This study demonstrated that a 24-week mobile app–based human coaching program is beneficial for controlling body weight, BMI, triglyceride, and body composition in terms of bioimpedance for breast cancer survivors who are excessively overweight or have obesity.

Trial Registration: ClinicalTrials.gov NCT 05506189; https://clinicaltrials.gov/ct2/show/NCT05506189

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e64846) doi:10.2196/64846

KEYWORDS

mobile app; body weight; survivor; breast cancer; obesity

Introduction

Obesity is associated with the risk of breast cancer and its recurrence. This relationship involves altered fatty acid metabolism, extracellular matrix remodeling, secretion of adipokines and anabolic and sex hormones, immune dysregulation, and chronic inflammation [1]. A previous systematic meta-analysis showed that women with obesity and breast cancer had poorer survival rates than women without obesity [2].

This increase in the number of breast cancer survivors is largely attributable to advancements in early detection and treatment. An estimated 297,790 new cases of invasive breast cancer are expected in women in the United States, along with 55,720 new cases of noninvasive breast cancer. There are over 3.8 million breast cancer survivors in the United States. The 5-year relative survival rate for patients with breast cancer is 91% [3]. Breast cancer survivors face significant challenges in managing their weight because of a confluence of factors, including hormonal changes induced by menopause, side effects of chemotherapy and antihormonal treatments, and psychological stressors related to their condition. It is crucial to reduce BMI and maintain an optimal BMI in women with obesity and breast cancer.

Exercise interventions have gained increasing attention globally for their role in improving body composition, physical fitness, and overall health outcomes among individuals with overweight and obesity, including breast cancer survivors [4]. Among various modalities, combined aerobic and resistance training is highly effective in enhancing body composition, lipid metabolism, inflammation, and overall quality of life in this population [5]. In addition, high-intensity interval training and combined aerobic and resistance training have been shown to provide both physiological and psychological benefits, including improvements in cardiometabolic markers, mood, and exercise adherence. Furthermore, these structured exercise programs are safe and effective in populations with metabolic diseases, reinforcing their role in weight management strategies [6,7].

Noom is a digital app dedicated to weight management that features functionalities for monitoring caloric intake and physical activity, facilitating collective encouragement through group involvement, and crucially providing individualized coaching. Unlike programs that prioritize rapid weight reduction, this platform educates users on acquiring new competencies and establishing sustainable healthy practices. Such mobile apps are increasingly used in the health care sector. Comparing the dietary data collected on energy and macronutrients, Noom's estimates for mean daily fat intake and percent total energy from carbohydrates are comparable to those of conventional dietary assessment tools. Noom's estimates for daily energy, protein, and carbohydrate intakes are significantly higher. The Noom could be useful for monitoring dietary intake, although more research is needed to determine its accuracy for micronutrients and other dietary components [8]. A study of 35,921 Noom Coach app users found that 77.9% experienced weight loss, with dinner input frequency being the most significant factor (odds ratio [OR] 10.69, 95% CI 6.20-19.53; P<.001), and frequent weight input reducing the yo-yo effect (OR 0.59, 95% CI 0.39-0.89; P<.001). This study demonstrates the app's clinical utility for weight reduction, especially for users who consistently monitor their weight and diet [9].

This study aimed to develop a 24-week mobile app–based human coaching program using Noom and evaluate its efficacy in breast cancer survivors who are excessively overweight or with obesity.

Methods

Trial Design and Study Population

We conducted a single-arm prospective cohort study between 2019 and 2021 at the National Cancer Center in the Republic of Korea. We provided an informational paper tailored to breast cancer survivors at the National Cancer Center within the existing Noom program. Experts in breast cancer, rehabilitation, nutrition, and psychiatry contributed to the writing of this paper. We planned to enroll breast cancer survivors who were at an excessive weight or with obesity. They were enrolled and joined a 24-week mobile app–based human coaching program during the first 6 months. They were followed up for 12 months without further application of the Noom program. Participants were recruited via physician referrals during routine visits to the outpatient breast cancer center.

The inclusion criteria were breast cancer survivors who had completed their primary treatment and were either scheduled for or currently undergoing follow-up examinations; aged between 18 and 70 years; diagnosed with breast cancer at stages I, II, or III; with a BMI≥25 kg/m² and <40 kg/m²; and capable of using a mobile app, as this was likely a component of the study methodology or follow-up process. All participants were required to sign a research consent form, acknowledge their voluntary participation, and understand the study's aims and procedures.

Individuals were excluded from the study if they: (1) were diagnosed with stage IV breast cancer or carcinoma in situ; (2) did not use smartphones; (3) had previous treatment for any cancer, including breast cancer; (4) had multiple organ tumors and recurrent or metastatic cancers; (5) had 2 or more uncontrolled chronic conditions, such as stroke, uncontrolled diabetes mellitus, uncontrolled hypertension, uncontrolled hypercholesterolemia, or uncontrolled psychological disorders; (6) declined to participate in the study; and (7) had communication difficulties impeding participation in the study.

We categorized participants into 3 groups based on the duration of their activities. These groups were defined as hyperactive, active, or inactive. Participants in the hyperactive group engaged

XSL•FO
RenderX

in activities for more than 16 weeks. Those in the active group participated in activities for 8 - 16 weeks. Finally, the inactive group included participants with less than 8 weeks of activity.

Procedures

For this study, we developed and implemented a 24-week program, including 5 categorized self-developed educational materials about diet-, exercise-, disease-, lifestyle-, and psychology-related content in the Noom app. The participants were instructed to use the Noom app, a mobile app–based human coaching program. Once familiar with the operation of Noom, the participants were required to log their daily dietary intake and lifestyle habits into the app for 24 weeks. An assigned Noom coach analyzed the data and provided individualized coaching through in-app messages at least once a week, offering feedback on lifestyle habits and dietary modifications [10].

To assess the effectiveness of the mobile app–based human coaching program in managing nutrition and lifestyle habits, the participants underwent physical measurements, clinical laboratory tests, and quality of life (QOL) surveys at the onset of the study and at the 6- and 12-month follow-ups. Anthropometric measurements were assessed using BMI and InBody analysis. For the QOL assessment, we used the EORTC Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30) and QLQ-BR23. These tools were scored according to the European Organisation for Research and Treatment of Cancer (EORTC) scoring manual [11].

Outcomes

The primary outcome was BMI reduction in the hyperactive group with a 24-week mobile app–based human coaching program through personalized nutrition and lifestyle modifications. Secondary outcomes were weight reduction and improvement in body composition, which were assessed using blood tests, bioimpedance, and QOL.

Sample Size Calculation and Statistical Analysis

This single-arm study was conducted over 24 weeks. It aimed for a reduction in BMI of at least mean 0.8 (SD 1.45) kg/m^2 in the hyperactive group, with an α of .05 and 90% power [12]. Considering a dropout rate of 50% after registration to the group of interest and 60% from the group of interest to the hyperactive group, 130 participants were required for this study.

The baseline characteristics were compared among the hyperactive, inactive, and active groups. Comparisons between the 2 groups were performed using the Chi-square test or Fisher exact test and the 2-tailed sample *t* test, depending on the type of variable. Changes in body weight, BMI, and body composition from baseline to 6 months in the hyperactive group were compared using 2-tailed paired *t* tests. One-way ANOVA was performed to determine changes over time for patients with physical measurements and QOL completed in 12 months, and changes at each time point were compared using paired *t* tests. Categorical variables were summarized as frequencies and percentages, continuous variables as means and SDs, and mean differences as difference values and 95% CI. Statistical significance was set at $P<.05$, and the results were analyzed using R version 4.1.2 (R Foundation for Statistical Computing).

Ethical Considerations

This study was approved by the institutional review board of the National Cancer Center of Korea (NCC2019-0098). All participants provided written informed consent prior to enrollment. The original informed consent obtained for primary data collection included provisions for secondary analysis without requiring additional consent. All experimental protocols were reviewed and approved by the same institutional review board, and all methods were conducted in accordance with relevant ethical guidelines and regulations [13,14]. Study data were anonymized to protect participants' privacy and confidentiality. During the study period, participants received a small token of appreciation, including partial reimbursement for test-related expenses and access to the paid mobile app.

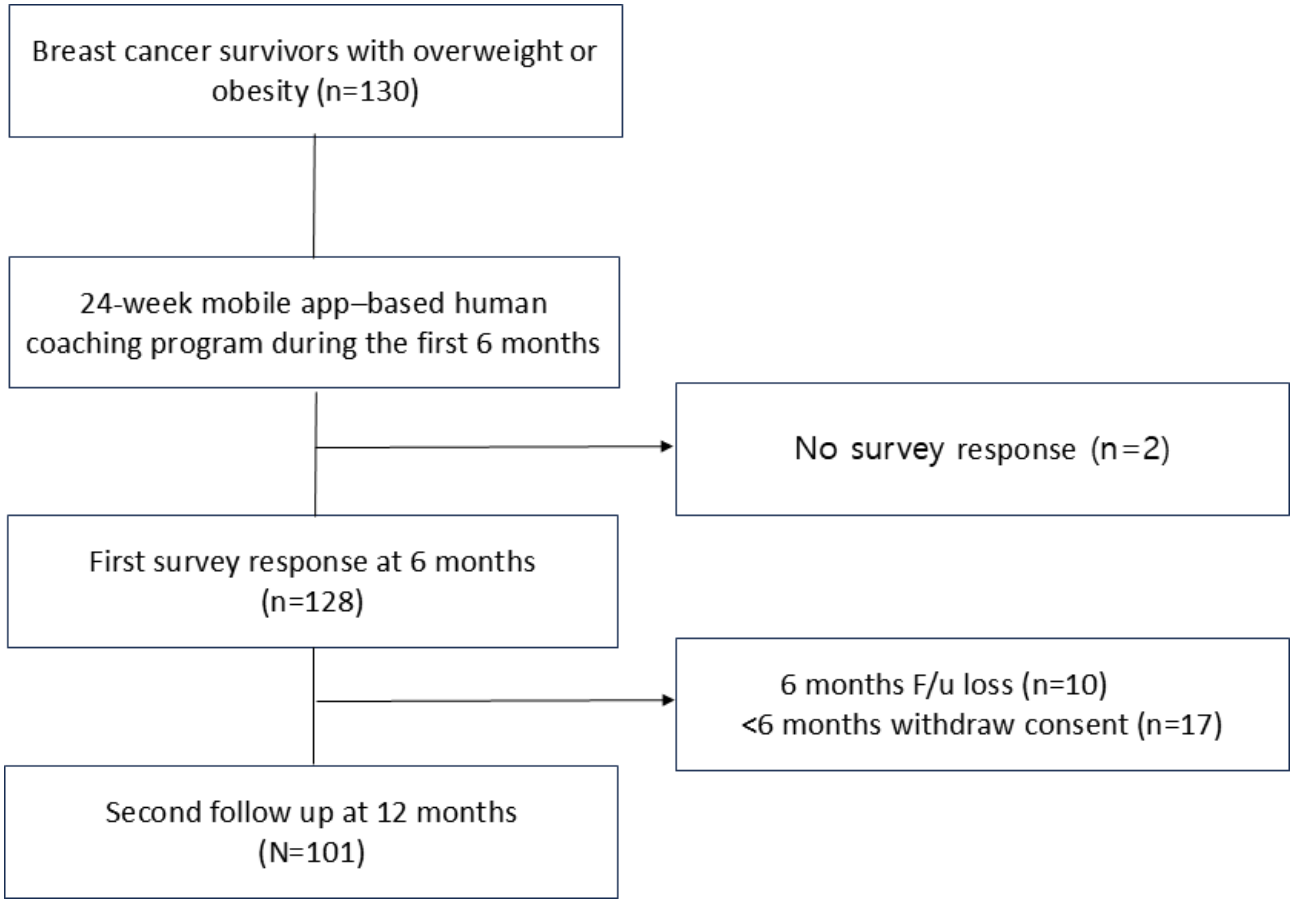
Results

Between May 2019 and December 2021, 130 participants who met the inclusion criteria were enrolled. A total of 77.5% (101 of the participants completed the study at 6 months follow-up using the 24-week program, and 71.5% (n=93) completed all study 12-month follow-up visits (Figure 1).

The mean age of all participants was 54.90 (SD 7.42) years (54.24, SD 7.84 years for the hyperactive group and 56.27, SD 6.38 years for the inactive and active group). In Baseline, the median BMI for the hyperactive group was 27.14 (IQR 25.20 - 35.36), while for the active and inactive groups, it was 27.50 (IQR 25.20 - 35.50). Among the participants, 85 (84.2%/101) individuals had a BMI in the range of 25 - 29.9, while 16 (15.8%/101) individuals had a BMI of 30 or higher. At baseline, no differences were observed between the groups in terms of BMI, occupational status, clinicopathologic characteristics, and underlying diseases, including hypertension, diabetes mellitus, and hyperlipidemia. Among the participants, 67% (68/101) were in the hyperactive group, and 33% (33/101) were in the inactive and active groups (Table 1).

XSL•FO
RenderX

Figure 1. Trials diagram of recruitment and completion.



XSL•FO
RenderX

Table . Baseline characteristic of study population.

Variables	Total (N=101)	Hyperactive (n=68)	Inactive and active (n=33)	*P* value
Age (year), mean (SD)	54.90 (7.42)	54.24 (7.84)	56.27 (6.38)	.20[a]
Age group (year), n (%)				.47[b]
≤40	5 (5.0)	5 (7)	0 (0)	
41-50	26 (25.7)	18 (26)	8 (24)	
51-60	43 (42.6)	27 (40)	16 (49)	
61-70	27 (26.7)	18 (27)	9 (27)	
BMI, n (%)				.87[c]
25-29.9	85 (84.2)	58 (85)	27 (82)	
≥30	16 (15.8)	10 (15)	6 (18)	
Education, n (%)				.16[b]
Less than a high school diploma	14 (13.9)	10 (15)	4 (12)	
High School Diploma	48 (47.5)	28 (41)	20 (61)	
College degree or higher	39 (38.6)	30 (44)	9 (27)	
Occupation, n (%)				.93[c]
Unemployment	56 (55.5)	37 (54)	19 (58)	
Employment	45 (44.5)	31 (46)	14 (42)	
T[d] stage, n (%)				.24[b]
1	67 (66.3)	46 (68)	21 (64)	
2	30 (29.7)	21 (31)	9 (27)	
3	4 (4.0)	1 (1)	3 (9)	
N[e] stage, n (%)				.06[b]
0	68 (67.3)	51 (7)	17 (52)	
1	29 (28.7)	15 (22)	14 (42)	
2	2 (2.0)	1 (1)	1 (3)	
3	2 (2.0)	1 (1)	1 (3)	
Stage, n (%)				.43[b]
I	53 (52.5)	37 (54)	16 (49)	
II	43 (42.5)	29 (43)	14 (42)	
III	5 (5.0)	2 (3)	3 (9)	
ER[f], n (%)				.77[b]
Negative	15 (14.8)	11 (16)	4 (12)	
Positive	86 (85.2)	57 (84)	29 (88)	
PR[g], n (%)				.74[c]
Negative	24 (23.8)	15 (22)	9 (27)	
Positive	77 (76.2)	53 (78)	24 (73)	
HER2[h], n (%)				.77[b]
Negative	86 (85.1)	57 (84)	29 (88)	
Positive	15 (14.9)	11 (16)	4 (12)	
Surgery, n (%)				>.99[b]

XSL•FO
RenderX

Variables	Total (N=101)	Hyperactive (n=68)	Inactive and active (n=33)	*P* value
BCS[i]	87 (86.1)	58 (85)	29 (88)	
Mastectomy	14 (13.9)	10 (15)	4 (12)	
Chemotherapy, n (%)				.53[c]
Done	49 (48.5)	31 (46)	18 (55)	
Not done	52 (51.5)	37 (54)	15 (45)	
Radiotherapy, n (%)				.71[b]
Done	92 (91.1)	61 (90)	31 (94)	
Not done	9 (8.9)	7 (10)	2 (6)	
Antihormonal therapy, n (%)				.54[b]
Done	87 (86.1)	57 (84)	30 (91)	
Not done	14 (13.9)	11 (16)	3 (9)	
HTN[j], n (%)				.11[c]
Yes	28 (27.7)	15 (22)	13 (39)	
No	73 (72.3)	53 (78)	20 (61)	
DM[k], n (%)				.54[b]
Yes	13 (12.9)	10 (15)	3 (9)	
No	88 (87.1)	58 (85)	30 (91)	
Hyperlipidemia, n (%)				.70[c]
Yes	19 (18.8)	14 (21)	5 (15)	
No	82 (81.2)	54 (79)	28 (85)	

[a]*t* test.
[b]Fisher exact test, *P*<.05: significance levels.
[c]χ^2 test.
[d]T: tumor.
[e]N: node.
[f]ER: estrogen receptor.
[g]PR: progesterone receptor.
[h]HER2: human epidermal growth factor receptor 2.
[i]BCS: breast-conserving surgery.
[j]HTN: hypertension.
[k]DM: diabetes mellitus.

At 6 months, 68 patients in the hyperactive group achieved an average weight loss of 1.96 kg and a mean decrease in BMI of 0.86 (SD 1.22) kg/m^2. In addition, on average, skeletal muscle mass increased by 0.75 (SD 2.6) kg, and body fat percentage decreased by 3.58% (SD 6.88). Furthermore, the waist-hip ratio decreased by 0.04, and the visceral fat area decreased by 23 cm² (Table 2).

When comparing the hyperactive group with the active and inactive groups over 6 months, there was a statistically significant decrease in weight, BMI, total cholesterol level, LDL-cholesterol level, body fat mass (BFM, kg), percent body fat (PBF, %), and visceral fat area (VFA; Table S1 in Multimedia Appendix 1).

At 12 months, 61 participants in the hyperactivity group had completed the program. Compared with baseline, there were statistically significant decreases in weight, BMI, fasting blood sugar (FBS) level, triglyceride level, skeletal muscle mass, BFM, PBF, waist-to-hip ratio, and VFA at 6 months. Furthermore, at 12 months, there were significant decreases in weight, BMI, FBS, triglyceride, BFM, PBF, and VFA. However, when examining the differences in changes between 6 and 12 months for this hyperactive group, the weight change decreased by 0.31 kg and the BMI change increased by 0.2 kg/m^2 without statistical significance. Furthermore, the differences in waist-to-hip ratio and VFA showed statistically significant increases of 0.04 (*P*<.01) and 14.33 cm² (*P*=.01), respectively (Figure 2, Table 3).

XSL•FO
RenderX

When comparing QOL using the EORTC QLQ-C30 and EORTC QLQ-BR23 questionnaires among those in the hyperactive group who completed the 12-month program, there was a statistically significant improvement in symptoms related to nausea and vomiting among the symptom scales. During the 6-month program, there were significant improvements in nausea and vomiting, constipation, body image, arm symptoms, and breast symptoms. However, at 12 months, only nausea and vomiting had significantly improved (Table 4).

Table . Change of body weight, BMI, and body composition of hyperactive group during 6 months.

Variables	Hyperactive group (n=68)				
	Baseline, mean (SD)	6 months, mean (SD)	Mean difference (SD; 95% CI)	*t* test (*df*=67)	*P* value[a]
Weight (kg)	68.71 (7.83)	66.75 (8.60)	–1.96 (2.87; –2.65 to –1.26)	–5.61	<.01
BMI (kg/m^2)	27.77 (2.38)	26.91 (2.69)	–0.86 (1.22; –1.15 to –0.56)	–5.8	<.01
SBP[b] (mmHg)	132.65 (13.07)	129.66 (12.85)	–2.99 (14.31; –6.45 to 0.48)	–1.72	.09
DBP[c] (mmHg)	78.43 (9.99)	76.57 (8.41)	–1.85 (9.76; –4.22 to 0.51)	–1.57	.12
FBS[d] (mg/dL)	114.44 (29.53)	105.97 (16.67)	–8.47 (26.09; –14.79 to –2.16)	–2.68	.09
HbA_{1c}[e] (%)	6.07 (0.68)	6.01 (0.54)	–0.06 (0.44; –0.15 to 0.04)	–10.97	.23
Total cholesterol (mg/dL)	192.78 (42.33)	191.9 (38.42)	–0.88 (28.8; –7.85 to 6.09)	–0.25	.80
Triglycerides (mg/dL)	172.68 (93.92)	138.54 (99.29)	–34.13 (98.98; –58.09 to –10.17)	–2.84	.01
HDL[f]-cholesterol (mg/dL)	56.18 (13.90)	57.76 (13.62)	1.59 (9.83; –0.79 to 3.97)	1.33	.19
LDL[g]-cholesterol (mg/dL)	110.49 (34.86)	109.13 (32.45)	–1.35 (25.84; –7.61 to 4.90)	–0.43	.67
Skeletal muscle mass (kg)	23.85 (3.03)	24.6 (4.00)	0.75 (2.6; 0.12 to 1.38)	2.38	.02
Body fat mass (kg)	25.18 (5.23)	22.17 (6.90)	–3.01 (4.94; –4.20 to –1.81)	–5.02	<.01
Percent body fat (%)	36.47 (4.64)	32.89 (7.97)	–3.58 (6.88; –5.25 to –1.92)	–4.3	<.01
Waist-hip ratio	0.88 (0.07)	0.84 (0.14)	–0.04 (0.15; –0.07 to –0.002)	–2.13	.04
VFA[h] (cm^2)	113.1 (34.57)	90.09 (41.3)	–23.00 (40.71; –32.86 to –13.15)	–4.66	<.01

[a]2-tailed paired *t* test, *P*<.05: significance levels.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]FBS: Fasting blood sugar.

[e]HbA_{1c}: glycosylated hemoglobin A_{1c}.

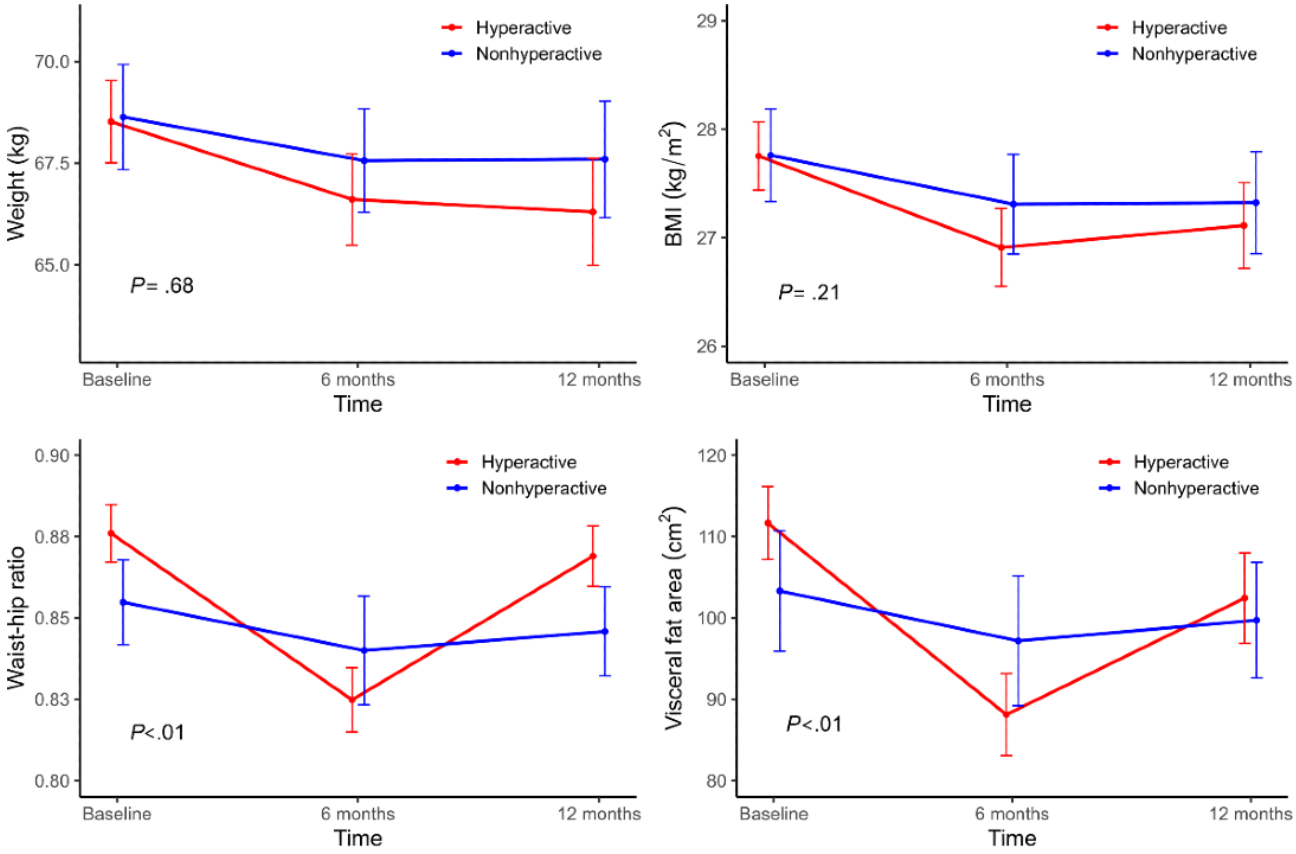
[f]HDL: high-density lipoprotein.

[g]LDL: low-density lipoprotein.

[h]VFA: visceral fat area.

XSL•FO
RenderX

Figure 2. Weight, BMI, and body composition change of hyperactive versus nonhyperactive (inactive and active) group at 6 months and 12 months.



XSL•FO
RenderX

Table . Change of body weight, BMI, and body composition of the hyperactive group at 6 and 12 months.

Variables	Hyperactive (n=61)			F (*df*=2, 180)	*P*[a] value	6 months-Baseline			12 months-Baseline			12 months-6 months		
	Baseline	6 months	12 months			Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value
Weight (kg)	68.52 (7.88)	66.61 (8.78)	66.3 (10.25)	1.52	.22	–1.91 (2.91; –2.65 to –1.17)	–5.13	<.01	–2.22 (5.68; –3.67 to –0.77)	–3.05	<.01	–0.31 (5.8; –1.79 to 1.17)	–0.42	.68
BMI (kg/m^2)	27.75 (2.45)	26.91 (2.8)	27.11 (3.08)	1.08	.34	–0.84 (1.24; –1.16 to –0.53)	–5.31	<.01	–0.64 (1.59; –1.05 to –0.23)	–3.14	<.01	0.2 (1.27; –0.12 to 0.53)	1.25	.22
SBP[c] (mmHg)	132.18 (13.43)	129.44 (13.22)	131.98 (11.87)	0.86	.42	–2.74 (14.76; –6.52 to 1.04)	–1.45	.15	–0.2 (14.87; −4.01 to 3.61)	–0.1	.92	2.54 (12.81; –0.74 to 5.82)	1.55	.13
DBP[d] (mmHg)	78.64 (9.92)	76.69 (8.62)	76.75 (8.63)	0.91	.40	–1.95 (10.1; –4.54 to 0.64)	–1.51	.14	–1.89 (10.71; –4.63 to 0.86)	–1.37	.17	0.07 (8.71; –2.16 to 2.3)	0.06	.95
FBS[e] (mg/dL)	114.08 (30.44)	106.11 (16.69)	105.59 (21.21)	2.5	.08	–7.97 (26.75; –14.82 to –1.12)	–2.33	.02	–8.49 (25.33; –14.98 to –2)	–2.62	.01	–0.52 (17.4; –4.98 to 3.93)	–0.24	.81
HbA_{1c}[f] (%)	6.04 (0.69)	5.99 (0.54)	6.10 (0.70)	0.48	.62	–0.05 (0.41; –0.15 to 0.05)	–0.98	.33	0.06 (0.63; –0.1 to 0.22)	0.8	.43	0.11 (0.49; –0.01 to 0.24)	1.83	.07
Total cholesterol (mg/dL)	190.98 (41.99)	190.51 (38.55)	186.3 (39.55)	0.25	.78	–0.48 (28.94; –7.89 to 6.94)	–0.13	.90	–4.69 (35.15; –13.69 to 4.31)	–1.04	.30	–4.21 (23.6; –10.26 to 1.83)	–1.39	.17
Triglycerides (mg/dL)	174.02 (92.65)	132.02 (69.74)	135.72 (65.26)	5.59	<.01	–42 (84.62; –63.67 to –20.33)	–3.88	<.01	–38.3 (79.04; –58.54 to –18.05)	–3.78	<.01	3.7 (60.89; –11.89 to 19.3)	0.48	.64
HDL[g]-cholesterol (mg/dL)	55.77 (13.49)	57.79 (13.03)	57.57 (14.04)	0.41	.66	2.02 (9.99; –0.54 to 4.58)	1.58	.12	1.8 (10.97; –1.01 to 4.61)	1.28	.20	–0.21 (8.98; –2.51 to 2.09)	–0.19	.85
LDL[h]-cholesterol (mg/dL)	109.64 (35.16)	109.02 (32.92)	107.8 (34.99)	0.05	.96	–0.62 (26.54; –7.42 to 6.17)	–0.18	.86	–1.84 (30.98; –9.77 to 6.1)	–0.46	.65	–1.21 (21.23; –6.65 to 4.22)	–0.45	.66
Skeletal muscle mass (kg)	23.81 (3.08)	24.73 (3.99)	24.23 (3.21)	1.09	.34	0.92 (2.52; 0.28 to 1.56)	2.85	.01	0.42 (1.97; –0.09 to 0.92)	1.65	.10	–0.5 (2.85; –1.23 to 0.23)	–1.37	.17

XSL•FO
RenderX

Variables	Hyperactive (n=61)			F (*df*=2, 180)	*P*[a] value	6 months-Baseline			12 months-Baseline			12 months-6 months		
	Baseline	6 months	12 months			Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value
Body fat mass (kg)	25.06 (5.28)	21.79 (6.75)	23.04 (7.2)	3.98	.02	–3.27 (4.8; –4.5 to –2.04)	–5.33	<.01	–2.02 (4.5; –3.17 to –0.87)	–3.51	<.01	1.25 (5.65; –0.19 to 2.7)	1.73	.09
Percent body fat (%)	36.39 (4.72)	32.36 (7.53)	33.78 (6.85)	6.07	<.01	–4.03 (6.53; –5.7 to –2.36)	–4.82	<.01	–2.61 (5.75; –4.08 to –1.13)	–3.54	<.01	1.42 (7.49; –0.5 to 3.34)	1.48	.14
Waist-hip ratio	0.88 (0.07)	0.82 (0.08)	0.87 (0.07)	8.58	<.01	–0.05 (0.09; –0.08 to –0.03)	–4.27	<.01	–0.01 (0.06; –0.02 to 0.01)	–0.85	.40	0.04 (0.1; 0.02 to 0.07)	3.24	<.01
VFA[i] (cm^2)	111.68 (35.05)	88.12 (39.47)	102.45 (43.44)	5.52	<.01	–23.56 (39.28; –33.62 to –13.5)	–4.68	<.01	–9.23 (29.27; –16.72 to –1.73)	–2.46	.02	14.33 (42.95; 3.33 to 25.33)	2.61	.01

[a]*P*: one way ANOVA.

[b]*P*: 2-tailed paired T-test, *P*<.05: significance levels.

[c]SBP: systolic blood pressure.

[d] DBP: diastolic blood pressure.

[e]FBS: fasting blood sugar.

[f]HbA_{1c}: glycosylated hemoglobin A_{1c}.

[g]HDL: high-density lipoprotein.

[h]LDL: low-density lipoprotein.

[i]VFA: visceral fat area.

XSL•FO
RenderX

Table . Comparison of quality of life in hyperactive group (n=61) through 12 months.

Variables	Hyperactive (n=61)			F (*df*=2, 180)	*P*[a] value	Hyperactive								
	Baseline	6 months	12 months			6 months-Baseline			12 months-Baseline			12 months-6 months		
	mean (SD)	mean (SD)	mean (SD)			Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value
EORTC QLQ-C30														
Global health status	66.94 (19.06)	67.49 (18.43)	70.9 (18.92)	0.8	.45	0.55 (19.12; –4.35 to 5.44)	0.22	.82	3.96 (21.5; –1.54 to 9.47)	1.44	.16	3.42 (21.96; –2.21 to 9.04)	1.21	.23
Functional scales														
Physical functioning	82.62 (10.91)	84.04 (9.98)	83.39 (13.31)	0.23	.79	1.42 (10.26; –1.21 to 4.05)	1.08	.28	0.77 (12.48; –2.43 to 3.96)	0.48	.63	–0.66 (12.45; –3.85 to 2.53)	–0.41	.68
Role functioning	84.43 (19.92)	88.8 (17.41)	89.34 (15.82)	1.4	.25	4.37 (17.72; –0.17 to 8.91)	1.93	.06	4.92 (18.6; 0.16 to 9.68)	2.07	.04	0.55 (19.48; –4.44 to 5.53)	0.22	.83
Emotional functioning	82.1 (17.54)	78.96 (21.87)	80.87 (16.41)	0.43	.65	–3.14 (20.48; –8.39 to 2.1)	–1.2	.24	–1.23 (14.01; –4.82 to 2.36)	–0.69	.50	1.91 (20.21; –3.26 to 7.09)	0.74	.46
Cognitive functioning	79.51 (17.58)	80.6 (15.57)	79.51 (18.61)	0.08	.92	1.09 (16.35; –3.09 to 5.28)	0.52	.60	0 (12.91; –3.31 to 3.31)	0	>.99	1.09 (16.35; –5.28 to 3.09)	–0.52	.60
Social functioning	89.07 (15.48)	92.35 (15.39)	90.44 (16.52)	0.66	.52	3.28 (16.05; –0.83 to 7.39)	1.6	.12	1.37 (18.58; –3.39 to 6.13)	0.57	.57	–1.91 (20.43; –7.15 to 3.32)	–0.73	.47
Symptom scales														
Fatigue	30.42 (20.17)	28.42 (17.75)	27.69 (16.31)	0.37	.69	–2 (17.63; –6.52 to 2.51)	–0.89	.38	–2.73 (14.72; –6.5 to 1.04)	–1.45	.15	–0.73 (15.43; –4.68 to 3.22)	–0.37	.71
Nausea / Vomiting	7.38 (13.45)	2.73 (6.93)	3.01 (9.38)	3.92	.02	–4.64 (13.99; –8.23 to –1.06)	–2.59	.01	–4.37 (14.24; –8.02 to –0.72)	–2.4	.02	0.27 (8.33; –1.86 to 2.41)	0.26	.80
Pain	19.67 (24.25)	19.95 (17.7)	18.03 (22.63)	0.14	.87	0.27 (18.88; –4.56 to 5.11)	0.11	.91	–1.64 (21.45; –7.13 to 3.86)	–0.6	.55	–1.91 (22.38-764 to 3.82)	–0.67	.51
Dyspnea	16.39 (23.27)	14.21 (19.68)	16.39 (22.46)	0.2	.82	–2.19 (27.13; –9.13 to 4.76)	–0.63	.53	0 (26.53; –6.79 to 6.79)	0	>.99	2.19 (21.83; –3.41 to 7.78)	0.78	.44

XSL•FO
RenderX

Variables	Hyperactive (n=61)			F (*df*=2, 180)	*P*[a] value	Hyperactive								
	Baseline	6 months	12 months			6 months-Baseline			12 months-Baseline			12 months-6 months		
	mean (SD)	mean (SD)	mean (SD)			Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value
Insomnia	33.88 (30.73)	39.34 (33.61)	37.16 (32.26)	0.44	.64	5.46 (32.87; –2.96 to 13.88)	1.3	.20	3.28 (27.02; –3.64 to 10.2)	0.95	.35	–2.19 (30.35; –9.96 to 5.59)	–0.56	.58
Appetite loss	7.65 (21.42)	8.2 (17.9)	6.01 (15.53)	0.23	.79	0.55 (26.17; –6.16 to 7.25)	0.16	.87	–1.64 (25.41; –8.15 to 4.87)	–0.5	.62	–2.19 (20.97; –7.56 to 3.18)	–0.81	.42
Constipation	23.5 (28.77)	14.21 (20.6)	19.67 (25.37)	0.72	.49	–9.29 (22.88; –15.15 to –3.43)	–3.17	<.01	–3.83 (25.16; –10.27 to 2.62)	–1.19	.24	5.46 (24.48; –0.81 to 11.73)	1.74	.09
Diarrhea	9.29 (16.25)	9.84 (17.58)	6.56 (14.68)	2.1	.12	0.55 (17.73; –4 to 5.09)	0.24	.81	–2.73 (15.27; –6.64 to 1.18)	–1.4	.17	–3.28 (18.96; –8.13 to 1.58)	–1.35	.18
Financial problems	7.1 (17.34)	6.01 (16.68)	7.65 (16.55)	0.15	.86	–1.09 (20.15; –6.25 to 4.07)	–0.42	.67	0.55 (16.66-37.2 to 4.81)	0.26	.80	1.64 (18.68; –3.15 to 6.42)	0.69	.50
QLQ-C30 Summary Score	82.27 (10.59)	83.68 (9.14)	83.77 (9.69)	0.45	.64	1.41 (9.15; –0.93 to 3.75)	1.2	.23	1.5 (8.33; –0.63 to 3.63)	1.41	.17	0.09 (9.97; –2.46 to 2.64)	0.07	.94
EORTC QLQ-BR23														
Functional scales														
Body image	67.35 (27.19)	75 (25.69)	71.45 (25.43)	1.31	.27	7.65 (23.49; 1.64 to 13.67)	2.54	.01	4.1 (24.04; –2.06 to 10.25)	1.33	.19	–3.55 (20.61; –8.83 to 1.73)	–1.35	.18
Future Perspective	50.82 (28.94)	58.47 (27.66)	56.28 (28.25)	1.18	.31	7.65 (33.55; –0.94 to 16.24)	1.78	.08	5.46 (32.31; –2.81 to 13.74)	1.32	.19	–2.19 (27.8; –9.31 to 4.93)	–0.61	.54
Symptom scales														
Arm symptoms	29.69 (21.63)	22.22 (19.88)	26.78 (20.92)	1.99	.14	–7.47 (17.77; –12.02 to –2.92)	–3.28	.00	–2.91 (23.03; –8.81 to 2.98)	–0.99	.33	4.55 (23.07; –1.36 to 10.46)	1.54	.13

XSL•FO
RenderX

Variables	Hyperactive (n=61)			F (*df*=2, 180)	*P*[a] value	Hyperactive								
	Baseline	6 months	12 months			6 months-Baseline			12 months-Baseline			12 months-6 months		
	mean (SD)	mean (SD)	mean (SD)			Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value	Mean difference (SD; 95% CI)	*t* test (*df*=60)	*P*[b] value
Breast symptoms	14.48 (15.73)	10.79 (10.69)	12.57 (12.42)	1.21	.30	–3.69 (14.23; –7.33 to –0.04)	–2.02	.05	–1.91 (13.38; –5.34 to 1.52)	–1.12	.27	1.78 (10.33; –0.87 to 4.42)	1.34	.18
Systemic therapy	23.26 (13.53)	22.56 (12.83)	23.11 (12.83)	0.05	.95	–0.7 (11.66; –3.69 to 2.28)	–0.47	.64	–0.16 (10.06; –2.73 to 2.42)	–0.12	.90	0.55 (9.92; –1.99 to 3.09)	0.43	.67

[a]*P*: one-way ANOVA.

[b]*P*: Paired *t* test, *P*<.05: significance levels.

Discussion

Principal Findings

This study showed that a 24-week mobile app–based human coaching program for breast cancer survivors was effective in reducing excessive weight and obesity. The hyperactive group lost 1.96 kg and 0.86 kg/m^2 at 6 months. Also, the hyperactive group showed improved QOL at 6 months. At the 12-month follow-up, weight and BMI tended to return toward baseline levels, suggesting that the intervention period may need to be extended for sustained effects. This highlights the importance of considering long-term strategies to maintain the benefits observed during the intervention. The significant impact of structured exercise programs is evident in weight management, body composition improvements, and overall physical and mental well-being.

A study found that combined aerobic and resistance training was the most effective exercise modality for improving cardiometabolic health markers. The findings suggest a synergistic effect when integrating multiple exercise types, leading to enhanced body composition and metabolic health outcomes [15]. Another review study found that regular, moderate-to-high-intensity aerobic exercise, especially when combined with weight loss and dietary modifications, can lead to sustained improvements in HDL-C levels. The authors recommend that individuals with overweight and obesity engage in such exercise regimens to enhance their HDL-C levels and overall health [16]. Other exercises, including yoga and Pilates, led to significant improvements in anthropometric measures, body composition, glucose and lipid metabolism, and blood pressure among women who were sedentary, overweight, or had obesity [17-19]. In our study, we observed that participants who actively engaged with the Noom platform experienced significant improvements in body composition, including reductions in body weight and BMI.

Maintaining a healthy weight through diet and exercise can help regulate hormone levels and improve overall health, potentially enhancing survival rates and reducing recurrence. In a previous study, a 16-week monitored program of aerobic and resistance exercises designed to target metabolic syndrome significantly improved QOL, depression symptoms, fatigue levels, and physical fitness in ethnically diverse, inactive, overweight survivors with obesity or breast cancer. These improvements remained evident at the follow-up 3 months later [20].

Our study showed that body weight, FBS, and triglyceride levels decreased over 6 months, which was maintained for up to 12 months; however, no further change was observed between the 6- and 12-month marks. Regarding BMI, an improvement was observed during the initial 6 months of using the Noom app. However, when Noom was discontinued, the following 6 - 12 months showed a reversal to baseline levels. The limitations in maintaining a diet over an extended period may be influenced by various factors, including individual dietary habits, social and cultural influences, psychological factors, and environmental barriers that often make long-term diet adherence challenging.

Being excessively overweight, or people with obesity, is linked to a higher risk of developing breast cancer, especially in postmenopausal women [21]. A study showed significantly poorer outcomes, including overall survival (OS) and disease-free survival (DFS), in patients with severe obesity (BMI≥40 kg/m²). This contrast was not observed in patients who have moderate obesity (BMI 35.0 - 39.9 kg/m²), those with slight obesity (BMI 30.0 - 34.9 kg/m²), or those with excessive weight (BMI 25.0 - 29.9 kg/m²) when compared with those who are underweight or of normal weight (BMI<25.0 kg/m²) with early breast cancer [22]. A total of 301 menopausal women with breast cancer were studied. The findings showed an adjusted OR of 1.37 (95% CI 0.73-2.56) for patients with a BMI of 27 kg/m^2 or more compared to those with a BMI under 27 kg/m^2. They suggest that patients with obesity had poorer survival rates [23]. A meta-analysis of 43 other studies indicated

XSL•FO
RenderX

that women with obesity and breast cancer had lower survival rates than women without obesity. This was consistent for OS (hazard ratio 1.33 95% CI 1.21-1.47) and breast cancer–specific survival (hazard ratio 1.33; 95% CI 1.19-1.50) [2].

Breast cancer survivors frequently experience cancer-related symptoms, including cognitive difficulties, pain, insomnia, and urinary incontinence, with 65% experiencing ongoing pain [24-26]. Weight management improves the QOL of breast cancer survivors by alleviating side effects, improving mobility, and boosting mental health. A comprehensive randomized controlled trial involving a diverse cohort of breast cancer survivors who are excessively overweight or people with obesity found that, while an intensive, group-based weight loss intervention yielded positive results in reducing weight, it led to only marginal improvements in vitality and transient enhancements in physical functioning and symptomatology [27]. Our study yielded similar results. In the hyperactive group, various physical symptoms, including body image, improved during the first 6 months. However, only the alleviation of nausea and vomiting was sustained for more than 12 months.

Mobile interventions for weight management programs are widely used in various diseases, such as diabetes, heart disease, and obesity, especially those where weight management is crucial [9,10,12]. Among patients with type 2 diabetes, those who completed the program experienced a significant weight reduction of 5.6% (with a SE of 0.81; $P<.001$) after 6 months. They maintained a weight loss of 4.7% (with an SE of 0.88; $P<.001$) at the 12-month mark. In contrast, the control group showed a negligible weight loss of 0.15% at 6 months (SE 0.64; $P=.85$) and a slight weight gain of 0.33% (SE 0.70; $P=.63$) at 12 months [28]. In the systematic review, 7 papers [29, 30, 31, 32, 33, 34, 35] met the inclusion criteria. The most commonly observed measure in these studies was the change in participants' weight, which was noted in 57% of the studies. In addition, most studies (71%) found statistically significant improvements in at least one area, including weight loss, increased physical activity, dietary habits, reduction in BMI, reduced waist circumference, decreased intake of sugar-sweetened beverages, reduced screen time, and measures of satisfaction or acceptability [36]. In our study as well, we observed improvements in weight management, physical activity, and QOL in the hyperactive group when the device was applied. The use of these devices is clinically significant for patients with breast cancer. Many of these patients consistently focus on dietary habits and weight management. Therefore, smartphone-based interventions that offer enhanced accessibility could improve patient care and reduce the risk of obesity. Another study showed the impact of the Noom mobile behavior change program on weight loss among breast cancer survivors, reporting a mean weight loss of 4.8 (SD 4.4) kg (5.6% of initial body weight) over 26 weeks, with increased physical activity levels and improved body image. Their study primarily focused on self-reported weight changes and app engagement [37]. In contrast, our study extends beyond weight loss by assessing changes in body composition, including metabolic health markers, rather than focusing solely on BMI or weight. In addition, a key strength of our study is the postintervention follow-up, which examines whether the effects of app usage are sustained after the cessation of the program. This provides valuable insights into the long-term effectiveness of mobile health interventions in breast cancer survivors.

Limitation

However, this study has a few limitations. First, this study regards the lack of information to identify the association of this program with OS or DFS in the participants. Further follow-up research is necessary to investigate the mobile health intervention value for OS and DFS. Second, this study has a relatively small number of participants with obesity, defined as BMI >30. Owing to the potential for different patterns of change depending on the degree of obesity, future studies may need to be stratified by similar proportions to address this variable. This study was a preliminary, single-arm prospective study designed to evaluate the efficacy of an application program in modifying body composition. However, as the analysis was conducted only on patients with hyperactivity, there is a potential for selection bias. Further research, including a randomized controlled trial, is needed to validate these findings.

Implications

This study demonstrates a shift from the traditional, resource-intensive approaches to nutritional interventions and lifestyle modifications. By using the mobile app–based Noom platform, this study provided a structured and accessible method for dietary tracking and lifestyle management. Furthermore, it implemented personalized nutrition and lifestyle adjustments through human coaching. This study was able to realize improvements in nutrition and lifestyle habits, ultimately contributing to improvement in QOL among breast cancer survivors.

Conclusions

Combating obesity and encouraging healthy lifestyle habits are essential components of preventive and therapeutic strategies for breast cancer. Mobile health intervention strategies could prove beneficial in various programs by providing information and support for the self-monitoring of survivors within supervised settings.

Acknowledgments

We would like to express our gratitude to all participants who contributed to this study. We also acknowledge the support of the research staff and collaborators at the National Cancer Center of Korea for their valuable assistance. This research was funded by the National Cancer Center (grant NCC 1910241-3), the Korean Cancer Survivors Healthcare Research and Development Project through the National Cancer Center, supported by the Ministry of Health and Welfare, Republic of Korea (grant NCC

XSL•FO
RenderX

23F1940). The funder had no role in the study design; collection, analysis, and interpretation of data; or writing of the reports. The authors declare that no generative artificial intelligence was used in the creation of this manuscript.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: SYJ and JL
Data curation: SYJ, EGL, SL, JHH, HSK, and ESL
Formal analysis: DEL
Investigation: SYJ
Methodology: DEL, EGL, and SYJ
Project administration: SYJ
Software: DEL
Funding acquisition: SYJ
Supervision: SYJ
Writing - original draft: EGL
Writing - review & editing: EGL, DEL, JK, JHH, SL, HSK, ESL, HC, SHS, KSL, JL, HJL, JSY, GAW, and SYJ.

Conflicts of Interest

JL is employed by Noom, Inc., and receives a salary and stock options. The company had no role in this study design, data collection, analysis and article preparation or publication.

Multimedia Appendix 1
Comparison between the hyperactive group and the combined active and inactive groups over 6 months.
[DOCX File, 18 KB - mhealth_v13i1e64846_app1.docx]

References

1. Pati S, Irfan W, Jameel A, Ahmed S, Shahid RK. Obesity and cancer: a current overview of epidemiology, pathogenesis, outcomes, and management. Cancers (Basel) 2023 Jan 12;15(2):485. [doi: 10.3390/cancers15020485] [Medline: 36672434]
2. Protani M, Coory M, Martin JH. Effect of obesity on survival of women with breast cancer: systematic review and meta-analysis. Breast Cancer Res Treat 2010 Oct;123(3):627-635. [doi: 10.1007/s10549-010-0990-0] [Medline: 20571870]
3. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023 Jan;73(1):17-48. [doi: 10.3322/caac.21763] [Medline: 36633525]
4. Newsome AM, Reed R, Sansone J, Batrakoulis A, McAvoy C, W. Parrott M. 2024 ACSM worldwide fitness trends: future directions of the health and fitness industry. ACSM's Health and Fitness Journal 2024;28(1):14-26. [doi: 10.1249/FIT.0000000000000933]
5. Al-Mhanna SB, Batrakoulis A, Norhayati MN, et al. Combined aerobic and resistance training improves body composition, alters cardiometabolic risk, and ameliorates cancer-related indicators in breast cancer patients and survivors with overweight/obesity: a systematic review and meta-analysis of randomized controlled trials. J Sports Sci Med 2024 Jun;23(2):366-395. [doi: 10.52082/jssm.2024.366] [Medline: 38841642]
6. Batrakoulis A, Fatouros IG. Psychological adaptations to high-intensity interval training in overweight and obese adults: a topical review. Sports (Basel) 2022 Apr 22;10(5):64. [doi: 10.3390/sports10050064] [Medline: 35622474]
7. Batrakoulis A, Jamurtas AZ, Fatouros IG. High-intensity interval training in metabolic diseases: physiological adaptations. ACSMs Health Fit J 2021;25(5):54-59. [doi: 10.1249/FIT.0000000000000703]
8. Kim J, Kim H, Lee J, et al. Comparison of energy and macronutrients between a mobile application and a conventional dietary assessment method in Korea. J Acad Nutr Diet 2022 Nov;122(11):2127-2133. [doi: 10.1016/j.jand.2022.04.007] [Medline: 35430416]
9. Chin SO, Keum C, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 7;6:34563. [doi: 10.1038/srep34563] [Medline: 27819345]
10. Kim Y, Oh B, Shin HY. Effect of mHealth with offline antiobesity treatment in a community-based weight management program: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 21;8(1):e13273. [doi: 10.2196/13273] [Medline: 31961335]
11. Fayers P, Aaronson NK, Bjordal K, Grønvold M, Curran D, Bottomley A. EORTC QLQ-C30 Scoring Manual: European Organisation for Research and Treatment of Cancer: European Organisation for Research and Treatment of Cancer; 2001.
12. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile Diabetes Prevention Program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016;4(1):e000264. [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]

XSL•FO
RenderX

13. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053]
14. van Delden JJM, van der Graaf R. Revised CIOMS international ethical guidelines for health-related research involving humans. JAMA 2017 Jan 10;317(2):135-136. [doi: 10.1001/jama.2016.18977] [Medline: 27923072]
15. Batrakoulis A, Jamurtas AZ, Metsios GS, et al. Comparative efficacy of 5 exercise types on cardiometabolic health in overweight and obese adults: a systematic review and network Meta-analysis of 81 randomized controlled trials. Circ Cardiovasc Qual Outcomes 2022 Jun;15(6):e008243. [doi: 10.1161/CIRCOUTCOMES.121.008243] [Medline: 35477256]
16. Badri Al-mhanna S, Leão C, Wan Ghazali WS, et al. Impact of exercise on high-density lipoprotein cholesterol in adults with overweight and obesity: a narrative review. Ann Appl Sport Sci 2024;12(2):0-0. [doi: 10.61186/aassjournal.1300]
17. Batrakoulis A. Psychophysiological adaptations to pilates training in overweight and obese individuals: a topical review. Diseases 2022 Sep 29;10(4):71. [doi: 10.3390/diseases10040071] [Medline: 36278570]
18. Batrakoulis A. Role of mind–body fitness in obesity. Diseases 2022 Dec 21;11(1):1. [doi: 10.3390/diseases11010001] [Medline: 36648866]
19. Batrakoulis A. Psychophysiological adaptations to yoga practice in overweight and obese individuals: a topical review. Diseases 2022 Nov 17;10(4):107. [doi: 10.3390/diseases10040107] [Medline: 36412601]
20. Dieli-Conwright CM, Courneya KS, Demark-Wahnefried W, et al. Aerobic and resistance exercise improves physical fitness, bone health, and quality of life in overweight and obese breast cancer survivors: a randomized controlled trial. Breast Cancer Res 2018 Oct 19;20(1):124. [doi: 10.1186/s13058-018-1051-6] [Medline: 30340503]
21. Chlebowski RT, Aragaki AK, Anderson GL, et al. Dietary modification and breast cancer mortality: long-term follow-up of the Women's Health Initiative randomized trial. J Clin Oncol 2020 May 1;38(13):1419-1428. [doi: 10.1200/JCO.19.00435] [Medline: 32031879]
22. Widschwendter P, Friedl TW, Schwentner L, et al. The influence of obesity on survival in early, high-risk breast cancer: results from the randomized SUCCESS A trial. Breast Cancer Res 2015 Sep 18;17(1):129. [doi: 10.1186/s13058-015-0639-3] [Medline: 26385214]
23. Katoh A, Watzlaf VJ, D'Amico F. An examination of obesity and breast cancer survival in post-menopausal women. Br J Cancer 1994 Nov;70(5):928-933. [doi: 10.1038/bjc.1994.422] [Medline: 7947099]
24. Bower JE. Behavioral symptoms in patients with breast cancer and survivors. J Clin Oncol 2008 Feb 10;26(5):768-777. [doi: 10.1200/JCO.2007.14.3248] [Medline: 18258985]
25. Gärtner R, Jensen MB, Nielsen J, Ewertz M, Kroman N, Kehlet H. Prevalence of and factors associated with persistent pain following breast cancer surgery. JAMA 2009 Nov 11;302(18):1985-1992. [doi: 10.1001/jama.2009.1568] [Medline: 19903919]
26. Mejdahl MK, Andersen KG, Gärtner R, Kroman N, Kehlet H. Persistent pain and sensory disturbances after treatment for breast cancer: six year nationwide follow-up study. BMJ 2013 Apr 11;346:f1865. [doi: 10.1136/bmj.f1865] [Medline: 23580693]
27. Demark-Wahnefried W, Colditz GA, Rock CL, et al. Quality of life outcomes from the Exercise and Nutrition Enhance Recovery and Good Health for You (ENERGY)-randomized weight loss trial among breast cancer survivors. Breast Cancer Res Treat 2015 Nov;154(2):329-337. [doi: 10.1007/s10549-015-3627-5] [Medline: 26518022]
28. Toro-Ramos T, Michaelides A, Anton M, et al. Mobile delivery of the diabetes prevention program in people with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 8;8(7):e17842. [doi: 10.2196/17842] [Medline: 32459631]
29. Gasser R, Brodbeck D, Degen M, Luthiger J, Wyss R, Reichlin S. Persuasiveness of a mobile lifestyle coaching application using social facilitation was evaluated to determine its potential to improve user engagement and promote sustained behavior change. In: IJsselsteijn WA, De Kort YAW, Midden C, van den Hoven E, editors. PERSUASIVE Technology PERSUASIVE 2006 Lecture Notes in Computer Science: Springer; 2006, Vol. 3962:27-38. [doi: 10.1007/11755494]
30. Gerber BS, Stolley MR, Thompson AL, Sharp LK, Fitzgibbon ML. Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. Health Informatics J 2009 Mar;15(1):17-25. [doi: 10.1177/1460458208099865] [Medline: 19218309]
31. Haapala I, Barengo NC, Biggs S, Surakka L, Manninen P. Weight loss by mobile phone: a 1-year effectiveness study. Public Health Nutr 2009 Dec;12(12):2382-2391. [doi: 10.1017/S1368980009005230] [Medline: 19323865]
32. Joo NS, Kim BT. Mobile phone short message service messaging for behaviour modification in a community-based weight control programme in Korea. J Telemed Telecare 2007;13(8):416-420. [doi: 10.1258/135763307783064331] [Medline: 18078554]
33. Lee W, Chae YM, Kim S, Ho SH, Choi I. Evaluation of a mobile phone-based diet game for weight control. J Telemed Telecare 2010;16(5):270-275. [doi: 10.1258/jtt.2010.090913] [Medline: 20558620]
34. Patrick K, Raab F, Adams MA, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009 Jan 13;11(1):e1. [doi: 10.2196/jmir.1100] [Medline: 19141433]
35. Shapiro JR, Bauer S, Hamer RM, Kordy H, Ward D, Bulik CM. Use of text messaging for monitoring sugar-sweetened beverages, physical activity, and screen time in children: a pilot study. J Nutr Educ Behav 2008;40(6):385-391. [doi: 10.1016/j.jneb.2007.09.014] [Medline: 18984496]

XSL•FO
RenderX

36. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329. [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
37. Shen S, Salehi E, White C, Chen Y, Iyengar NM. Evaluation of a mobile behavior change program for weight loss in breast cancer survivors. NPJ Breast Cancer 2024 Jun 29;10(1):53. [doi: 10.1038/s41523-024-00659-x] [Medline: 38951532]

Abbreviations

BFM: body fat mass
BR 23: Breast Cancer Module (23 items)
DFS: disease-free survival
EORTC: European Organisation for Research and Treatment of Cancer
EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
FBS: fasting blood sugar
OR: odds ratio
OS: overall survival
PBF: percent body fat
QOL: quality of life
VFA: visceral fat area

Edited by L Buis; submitted 28.07.24; peer-reviewed by A Batrakoulis, K Haas, S Bulduk; revised version received 13.07.25; accepted 25.07.25; published 24.09.25.

Please cite as:
Lee EG, Lee DE, Kim J, Han J, Lee S, Kang HS, Lee ES, Chae H, Sim SH, Lee KS, Lee J, Lee HJ, Yoo JS, Wie GA, Jung SY
Impact of a 24-Week Mobile App–Based Human Coaching Program on Body Composition and Lipid Metabolism in Breast Cancer Survivors With Overweight or Obesity: Single-Arm Prospective Cohort Study
JMIR Mhealth Uhealth 2025;13:e64846
URL: https://mhealth.jmir.org/2025/1/e64846
doi:10.2196/64846

© Eun-Gyeong Lee, Dong-Eun Lee, Jeongseon Kim, Jaihong Han, Seeyoun Lee, Han-Sung Kang, Eun Sook Lee, Heejung Chae, Sung Hoon Sim, Keun Seok Lee, Jungeun Lee, Hyun Jeong Lee, Ji Sung Yoo, Gyung Ah Wie, So-Youn Jung. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Providing 2 Types of mHealth Interventions to Support Self-Management Among People Living With HIV: Randomized Clinical Trial

Gwang Suk Kim[1], RN, PhD; Layoung Kim[1,2], RN, PhD; Seoyoung Baek[3], RN, PhD; Sooyoung Kwon[4], RN, MSN; Ji Min Kim[5], RN, BSN; Jun Yong Choi[6], MD, PhD; Jae-Phil Choi[7], MD, PhD

[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, 50-1, Yonsei-Ro, Seodaemun-gu, Seoul, Republic of Korea

[2]College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea

[3]Korea Armed Forces Nursing Academy, Daejeon, Republic of Korea

[4]Department of Nursing, Graduate School of Yonsei University, Seoul, Republic of Korea

[5]Division of Infectious Disease, Yonsei University Health System, Seoul, Republic of Korea

[6]Department of Internal Medicine, College of Medicine, Yonsei University, Seoul, Republic of Korea

[7]Division of Infectious Diseases, Seoul Medical Center, Seoul, Republic of Korea

Corresponding Author:
Layoung Kim, RN, PhD
Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, 50-1, Yonsei-Ro, Seodaemun-gu, Seoul, Republic of Korea

Abstract

Background: Mobile health (mHealth) has been continuously developed to support the HIV care continuum for people living with HIV. Considering the practical needs and acceptability of digital health solutions, it is essential to explore effective content and diverse delivery methods for self-management support.

Objective: This study aimed to assess the effectiveness of 2 non–face-to-face mHealth interventions for people living with HIV. We compared the impact on HIV self-management of (1) a link group, which received access to information via mobile link, and (2) an app group, which used a mobile app enabling information exploration and self-recording of health outcomes, including medication adherence, symptoms, mental health score, and sexual safety.

Methods: A 2-arm, prospective, randomized clinical trial was conducted, involving 83 people living with HIV aged 19 years or older, who were assigned to the app group (n=42) or link group (n=41). The primary outcome was self-reported self-efficacy for HIV management (HIV-SE), which comprised 6 domains: managing depression or mood, medication, symptoms, and fatigue; communicating with health care providers; and getting support or help. A paired *t* test and generalized estimating equation were used to analyze the outcomes at baseline, 4 weeks postintervention, and 8 weeks after an additional 4-week voluntary use period.

Results: Both groups demonstrated improvements in total HIV-SE scores at 4 weeks compared with baseline. All domain scores improved in the app group, with a significant increase in total HIV-SE and managing fatigue. The link group significantly improved in managing depression or mood, fatigue, and getting support or help domains. The generalized estimating equation analysis indicated that, compared with the link group, the app group had significant group-by-time interaction with a positive effect on managing symptoms at 4 weeks (β=0.635, 95% CI 0.023 to 1.247; *P*=.04) but a negative effect on managing depression or mood at 8 weeks (β=−0.824, 95% CI −1.448 to −0.200; *P*=.01). Only 9.5% (4/42) of app group participants maintained daily visits during the voluntary use period of 4 to 8 weeks.

Conclusions: Both types of informational mHealth interventions, through mobile apps or link access, contributed to improving HIV-SE. Delivering information via direct text message links could be suitable for individuals who are hesitant to use HIV-related apps. While mobile apps promote self-monitoring and symptom management through self-recording and reflection, strategies are needed to sustain long-term app engagement. In addition, user-customized psychiatric content beyond mental health recordings has been suggested for managing depressed moods in mHealth interventions.

Trial Registration: Clinical Research Information Service KCT0008014; https://trialsearch.who.int/Trial2.aspx?TrialID=KCT0008014

(JMIR Mhealth Uhealth 2025;13:e60905) doi:10.2196/60905

KEYWORDS

telemedicine; HIV; self-management; mHealth; randomized clinical trial; mobile health

XSL•FO
RenderX

Introduction

People living with HIV who adhere to antiretroviral therapy (ART) and achieve viral suppression have not been reported to sexually transmit the virus to others [1]. According to a cohort study by the Korea Disease Control and Prevention Agency, the ART adherence rate in Korea, defined as taking medication without interruption for 2 consecutive weeks, was 87.6% [2,3]. Additionally, national HIV epidemiology data reported that the ART treatment rate, based on prescription records in Korea, was 95.5%, and the viral suppression rate was 96%, indicating that most people living with HIV belong to the retention in care [4]. While scientific knowledge of "undetectable equals untransmittable (U=U)" has contributed to transforming HIV into a chronic illness, HIV-related discrimination with stigma remains prevalent in health care settings [5]. People living with HIV rely on infectious disease hospital-based resources for HIV-related information and consultation [6]. Several mobile-based interventions have been developed to address barriers to taking medications or visiting clinics, offering features such as adherence tracking and text message reminders [7,8]. However, existing mobile health (mHealth) solutions primarily focus on improving ART adherence rather than addressing the needs of the majority of people living with HIV who are already adherent to ART and retained in care. Moreover, studies have reported that the adoption and use of apps account for only a small proportion [9,10]. Despite the increasing availability of digital health information, this population continues to express a need for evidence-based, reliable information in the mHealth needs survey [11]. These findings highlight the ongoing demand for informational mHealth solutions for community-dwelling people living with HIV.

Given that many people living with HIV achieve ART adherence, integrated care to empower them to play a role in the HIV care continuum is increasingly emphasized [12,13]. Similar to other chronic illnesses, they require self-efficacy to manage both their HIV condition and holistic health beyond HIV disease in a long-term manner [13]. Self-efficacy for HIV disease management is a perceived behavior control indicator that practices health care behaviors along with compliance with treatment. Therefore, it is often used to evaluate the outcome of integrated chronic disease management, beyond clinical parameters such as viral suppression [14]. Integrated health services should encompass not only ART management but also engaging in social life, family planning, maintaining a safe sexual life, and managing psychological health, aligning with the vision of the HIV care continuum. In this study, informational mHealth interventions comprised 6 domains—HIV management, ART care, psychological health care, sexual health care, health promotion, and family and social life—that could provide knowledge and information regarding integrated health services [11,12]. Informational mHealth was designed to be delivered through either a mobile app or direct links via text messages. While mobile apps offer a structured platform to improve self-efficacy in HIV management, concerns over privacy and fear of disclosing HIV status may pose barriers to adoption. The coexistence of concealment and exposure [15] should be ensured and understood in the context of mHealth interventions. Considering differences in user preferences, whether receiving information solely through text message links is effective needs to be evaluated. Furthermore, the potential advantages of mobile apps—not only as information sources but also for their self-recording functionalities—should be explored. This study evaluates the effectiveness of informational mHealth in improving self-efficacy for HIV management (HIV-SE). We assess the impact of 2 mHealth delivery methods—a mobile app and links in text messages—through a prepost study design and compare their effectiveness in improving self-efficacy among people living with HIV.

Methods

Ethical Considerations

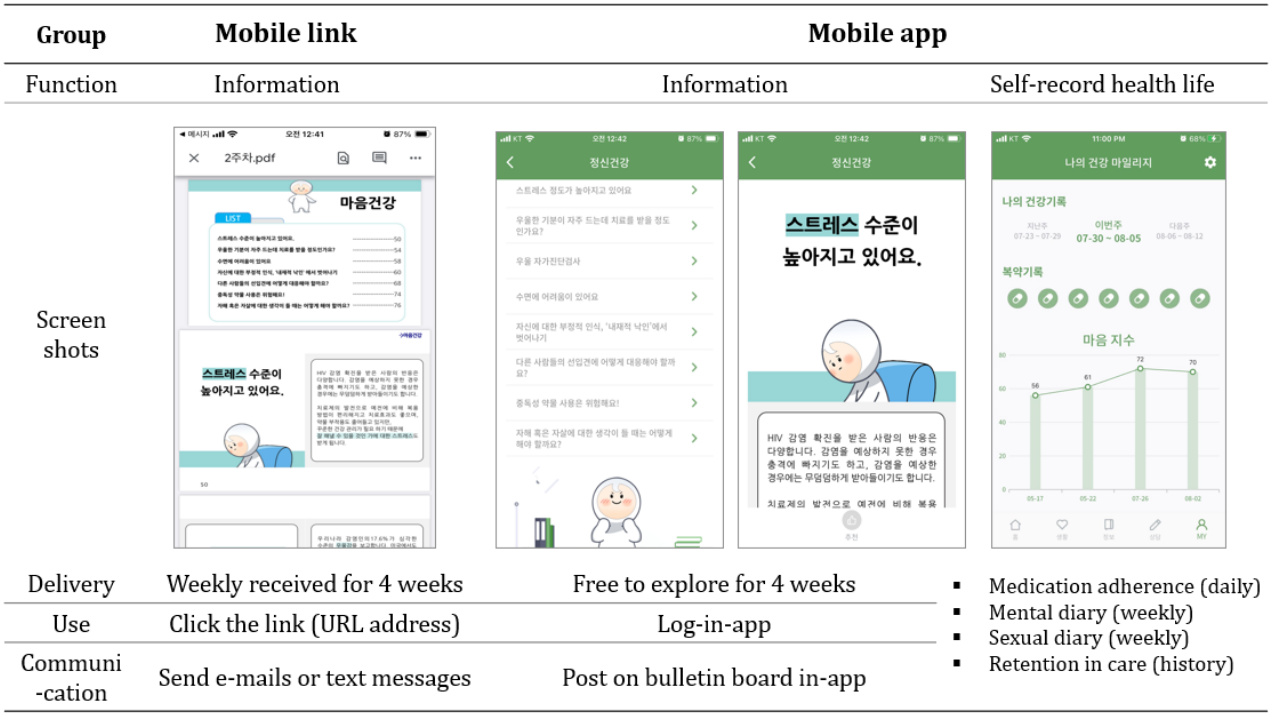
This study was approved by the Institutional Review Board of the first author's affiliation (4-2022-1192), and the protocol was registered with the Clinical Research Information Service (KCT0008014). All interventions and assessments were conducted through non–face-to-face contact, maintaining the privacy and confidentiality of the participants. Self-reported assessments were administered only after participants read and clicked on the informed consent via a web-based platform. Participants self-downloaded the mobile app from Google Play or the App Store and registered it on their mobile device. The app was available for both iOS and Android platforms, ensuring accessibility and minimizing potential selection bias in participation. Participants took part anonymously and could withdraw from the study at any time by simply choosing not to respond. To protect confidentiality, the app used a secure user authentication system that did not require personally identifiable information, using an anonymized ID and password. All self-recorded data, including diary entries and symptom tracking, were stored in a deidentified format. Users also had the option to delete their records at any time, giving them full control over their data. After completing the interventions, participants were compensated with gift coupons worth US$ 70 (KRW ₩100,000) for the app group and US$ 35 (KRW ₩50,000) for the link group.

Trial Design

This study was a 2-arm randomized clinical trial. We compared the link group, which received text messages of URL links for accessing information, with the app group, which used a mobile app for exploring information and self-recording health outcomes (Figure 1). The 2 interventional arms were implemented using participants' smartphones. The contents of the information provided were the same in both groups; however, the delivery methods were different. The link group clicked on the URL address of the weekly received text messages, and the app group logged into the app to view the content posted at all times.

XSL•FO
RenderX

Figure 1. Overview of mobile intervention methods: mobile link versus mobile app.



A priori sample size was estimated using G*power 3.1.9.2 software. Based on a previous study evaluating a similar mHealth intervention [16], a sample size of 60 participants was required to detect an effect size of 0.44, with a type 1 error rate of 5% and a power of 0.80. Considering an expected dropout rate of approximately 30%, the target sample size was set at 86 participants. Block randomization (block size of 4) was used to ensure 1:1 balanced enrollment. Research staff who generated the random numbers and block sequences (ie, AABB, ABAB, BAAB, BABA, BBAA, and ABBA) using Microsoft Excel and those who registered participants were different. However, the research staff could not be blinded after the arm assignment to be involved in data control. Participants were unable to blind the assignment in the app-based trial. The research staff and participants did not know each other before and had never met. All contacts from the research team were made via text messages or email, and a phone number was provided for participants to call 24/7 if needed. The period of the mHealth intervention was 4 weeks, with assessments at baseline (T0), 4 weeks postintervention (T1), and a follow-up at 8 weeks, after an additional 4-week voluntary use period (T2).

Participant Recruitment

The participants were recruited between July 19 and October 18, 2023. The eligibility criteria were as follows: (1) people diagnosed with HIV, aged 19 years and older, (2) individuals who owned and used a smartphone, and (3) those who had no difficulty reading and responding to the questionnaire via smartphone. Participants were excluded if they had a medical diagnosis that required immediate treatment and counseling, other than HIV infection. Participants were recruited using an online access method. A prior study to identify the mHealth needs of people living with HIV was conducted at 5 hospitals, and about half of the participants expressed their willingness to participate in further mHealth interventions. A recruitment poster was emailed to 138 individuals who had left their email addresses and expressed interest in participating in the intervention study based on a previous service-needs assessment conducted in 2020 - 2021. In addition, 340 small leaflets that fit into a pocket were distributed to visiting outpatients “people living with HIV” by medical staff at the 2 hospitals. The poster and leaflets specified the study purpose and participation method, and included a QR code that could access the web-based baseline survey. Participants who completed the baseline survey were randomly assigned to the app or link groups. The research staff sent the app group a text message explaining how to download the app from the App Store or Google Play. The app allowed participants to log in only after receiving approval from the research team. The link group received a text message mentioning that they would receive a weekly text message including a link to access the information.

Interventions

Overview

Both groups received text messages from the research team once a week to encourage their engagement with mHealth during the 4 weeks of intervention. The app group received a text message that encouraged the use of app functions, including the information screen, while the link group was encouraged to read the information material. After completing the intervention period at 4 weeks (T1), the participants were informed that the mandatory use period ended and that they could use the app or link address autonomously for the next 4 weeks (T2). During the 4 weeks of voluntary use, the research team did not contact the participants.

XSL•FO
RenderX

App Group

Overview

This group used a mobile app to support their HIV self-management. The app, namely, "ESSC (Excellent Self-Supervised HIV Care)," was developed using 3 iterative cycles of the Design Science Research framework [11]. Three cycles comprise relevance, derived from users' requirement need survey; rigor, organized based on the information–motivation–behavioral skills model with validity being evaluated by experts; and design, which builds artifacts including functions, contents, and features. Formative details of the development process are described in a previous study [11]. The major functions of the ESSC app are as follows.

Information

Overall, 47 health information components were provided in 6 domains: HIV management, ART care, psychological health care, sexual health care, health promotion, and family and social life. Integrated information was created to provide essential information about leading a healthy life with HIV. Each piece of information includes a literature source for accuracy and trustworthiness. To deliver acceptable and intensive information, all content was designed with brief descriptions and images, such as poster presentation templates. The title of each was framed as a question (eg, Would it be a problem if I do not notify others that I am infected?).

Health Life Records

Users record their health in 4 categories: medication adherence, mental diary, sexual diary, and retention in care. They checked the boxes of daily ART medication adherence and uncomfortable symptoms. In the mental diary, users could choose their strengths weekly, which were presented as 24 items by Peterson and Seligman [17]. After being aware of the strength, negative and positive emotions experienced last week were recorded and scored as an overall emotional score (−100=as sad as I can imagine, 0=not sad at all, 100=as happy as I can imagine). In the sexual diary, users could record sexual activity within the past week and check boxes for any difficulties they experienced in their sexual life (disclosing HIV to a partner, using condoms, fear of viral detection, believing that people living with HIV cannot have sex, etc). Then, they identified safe sexual guidelines for people living with HIV. For retention in care, the user was asked to record clinic visits and the progress of laboratory results (CD4+T lymphocyte [CD4+T] cell count and viral load).

Others

Users' health records are demonstrated by graphical and descriptive presentations on "My page." Self-reflection on one's own health records enables individuals to be aware of meaningful changes and motivates them to engage in healthy behaviors [18]. After completing the mental and sexual diaries, a pop-up message prompted users to communicate with the health care provider (HCP). When users left various questions about information, personal concerns, and app features, a nurse with more than 10 years of experience in HIV research responded within 24 hours.

Link Group

These participants received text messages, including the link to access information, once a week for the 4-week intervention period. When participants click the URL address in the text message, the screen is linked to the educational material, a PDF file. This information is the same as in the poster-templated content of the ESSC app. The 6 domains were categorized into 4 sessions: HIV management, psychological and sexual health care, ART care, and health promotion and family and social life. When participants had questions about information or personal concerns, they could send text messages or emails to the research team.

Measurement

Overview

Participants' basic information was measured at baseline (T0), and primary and secondary outcomes were assessed 3 times (T0, T1, and T2). The usability of the ESSC app was evaluated at 4 weeks (T1). All assessments were self-reported using a web-based platform (Google Survey).

Primary Outcome: Self-Efficacy for HIV Management

The Korean version of the HIV-SE scale [14] was used. The 33 items comprised 6 domains: managing depression or mood (9), medications (6), symptoms (5), and fatigue (4); communicating with the HCP (4); and getting support or help (5). The participants rated confidence in their ability to perform specific behaviors on a 10-point Likert scale (1 = "not sure at all" to 10 = "totally sure"). The original scales developed by Shively et al [19] were translated into Korean and validated the conceptual domains as a 6-factor model. In previous studies [20-23], the 6 domains were used to identify specific subconcepts of self-efficacy and analyze the multivariate factors in predicting self-management. Cronbach α for the Korean scale was 0.91 [14] and 0.93 in this study.

Secondary Outcomes

Medication adherence was measured using a self-reported visual analogue scale (VAS), where participants were asked to rate their adherence to prescribed medication over the past month on a scale from 0 to 100 (0=not taking the prescribed medications, 100=taking every single dose of prescribed medications). This method has been used in previous adherence studies [24,25] and has shown a strong correlation with objective adherence measures such as electronic monitoring [25]. A single-item VAS was selected to avoid redundancy, as adherence behaviors were already assessed through the HIV-SE domain "managing medications," which includes taking medications on time, daily, and according to prescribed instructions.

Self-efficacy regarding condom use was assessed using a single item on perceived behavioral control toward safe sexual intercourse. "If I have sex with my partner, I am confident that I always use a condom," the participants responded on a 5-point Likert scale (1 = "not at all" to 5 = "very much").

Health-related quality of life was assessed using the EuroQoL group's VAS, which measures perceived health on a particular day on a scale ranging from 0 to 100 (0=the worst health imaginable, 100=the best health imaginable) [23].

XSL•FO
RenderX

Usability Tests

Usability tests were conducted only for the app group. User records accumulated in the app were used to quantitatively assess usability. The Health Information Technology Usability Evaluation Scale (Health-ITUES) was used to evaluate the quality of the ESSC application. Schnall et al [26] developed 20 items of Health-ITUES to evaluate the usability of a mobile app for symptom self-management of community-dwelling people living with HIV on a 5-point Likert scale (from 1 = "strongly disagree" to 5 = "strongly agree"). The Korean version [27] was adapted for HIV disease self-management and includes 4 subscales: impact, perceived usefulness, ease of use, and user control. Finally, the participants were free to write about any problems they experienced while using the app or anything that needed to be corrected or supplemented. This qualitative feedback was also conducted in the link group.

Basic Information

Sociodemographic variables included age, sex, educational level, economic status, marital status, residence, employment status, and gender orientation. Use and need for mHealth were assessed by responses to the time spent using mHealth per day and their requirements for HIV health apps on a 10-point Likert scale (1 = "not at all" to 10 = "totally need"). The Short-Form Health Literacy Scale (SF-HLS) was used to measure participants' literacy, including access to and understanding of health care, disease management, and health promotion. The Korean version [28] that verified the reliability and validity of the original 12 items [29] was used.

For personal HIV status, data on the year of diagnosis, HIV status with symptoms, transmission mode of infection, and experience with self-help participation were collected. The Korean version [30] of the Short Form of the World Health Organization Quality of Life–HIV questionnaire (WHOQoL-HIV-BREF) [31] was partially used to assess the psychological domain. Laboratory results for viral load and CD4+T cells were self-reported based on participants' most recent hospital visit. Viral load was classified as undetectable, detectable, or unknown, and CD4+T cell counts were classified into ≥500, 200 - 499, <200 cells/mm^3, and unknown [32]. In addition, participants were asked whether they had taken medication without interruption for 2 consecutive weeks, using the same approach as the Korea Disease Control and Prevention Agency cohort study [3], to allow for comparison of baseline adherence levels with the general population of people living with HIV.

Statistical Analysis

Basic information variables at baseline are presented by the groups using descriptive analysis (counts with percentages or mean values with SDs). Changes in primary and secondary outcome variables between baseline and postintervention were analyzed using paired 2-sample *t* tests, which examined the differences in the mean values of outcomes by group during the intervention. The primary outcome variables were compared between the app and link groups using a generalized estimating equation (GEE). It is an appropriate statistical method that analyzes repeated measures data reflecting the interaction between group and time and performs intention-to-treat analysis. To compare the effectiveness of group differences across time, GEE analysis was performed, including the results of the 3 assessment points (T0, T1, and T2). Here, the GEE analysis was adjusted for 3 covariates: number of years after HIV diagnosis, psychological domain of the WHOQoL-HIV BREF, and SF-HLS. Previous studies have shown that the HIV self-management ability was affected by the duration of HIV diagnosis, psychological status, and literacy [33-36]. In addition, health literacy scores were significantly different between the 2 groups in the homogeneity test of basic information using the chi-square test or independent *t* test. Therefore, the research team acknowledged that these 3 covariates should be included in the evaluation of HIV-SE effectiveness. For all analyses, the threshold for statistical significance was a *P* value of .05. According to the statistical method, the GEE analysis included all data of the withdrawn participants, but the paired *t* test could not include withdrawn participants. Additionally, open-ended qualitative feedback collected during the usability test was analyzed using a descriptive content analysis approach. Two researchers independently reviewed the responses, grouped them into common themes, and categorized the feedback as either positive or negative from the user perspective. Key patterns were summarized narratively to complement the quantitative usability findings.

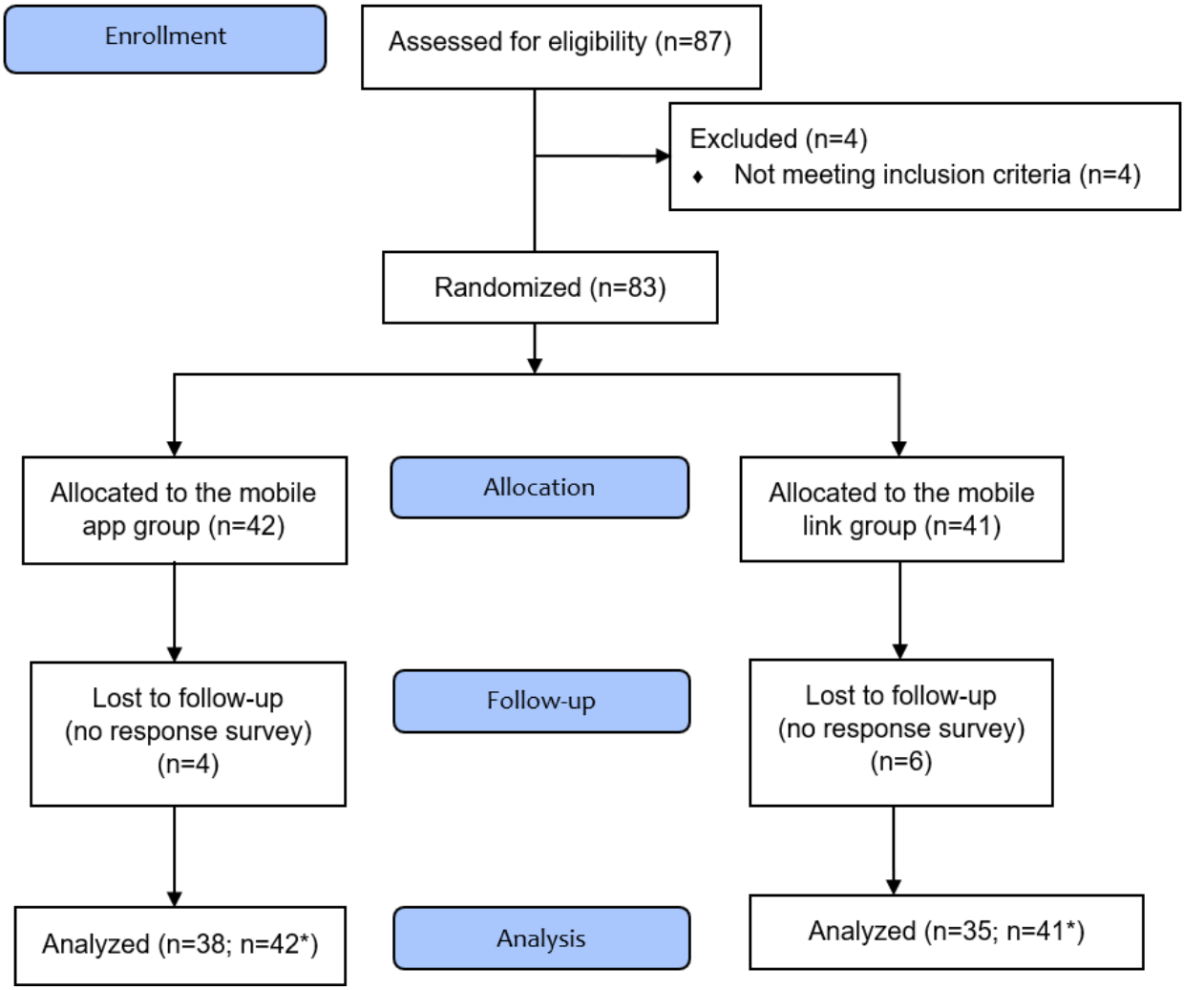
Results

Participants

Figure 2 illustrates a flow diagram for trial recruitment. Overall, 87 individuals accessed online sites to participate. After responding to the baseline survey, 4 were excluded as they met the exclusion criteria: had a psychiatric diagnosis (ie, bipolar disorder or depression) with symptoms, and were under treatment. The remaining 83 people living with HIV were assigned to the 2 groups, of which 73 completed the 8-week follow-up. When required to self-report assessments after completing the intervention at 4 weeks, 12% (10/83) of participants did not respond to the research team contact.

XSL•FO
RenderX

Figure 2. Consolidated Standards of Reporting Trials (CONSORT) flow diagram for trial recruitment. *Data were analyzed using a generalized estimating equation.



The baseline characteristics are shown in Table 1. The mean age of the participants was 39.42 years (range: 21 - 69 years), 96.4% (80/83) men, and the majority, 85.5% (71/83), were men who had sex with men, including homosexual and bisexual gender orientation. A total of 9 in 10 participants had never been married, and half lived alone. There was an average of 9.51 years after HIV diagnosis, and 6 in 7 participants had viral suppression and no HIV-related symptoms. Only 7 perceived themselves as having an AIDS status. The results of the homogeneity test showed that the SF-HLS scores ($t(81)=2.667$, $P=.009$) were significantly different between the 2 groups, while the other variables showed no significant differences.

XSL•FO
RenderX

Table . Baseline characteristics of the study participants.

Variable	Total (n=83)	App group (n=42)	Link group (n=41)
Age (years), mean (SD)	39.42 (10.38)	37.67 (10.72)	41.22 (9.82)
Sex, n (%)			
Male	80 (96.4)	40 (95.2)	40 (97.6)
Female	3 (3.6)	3 (4.8)	1 (2.4)
Education level, n (%)			
≤High school	32 (38.6)	20 (47.6)	12 (29.3)
≥College	51 (61.4)	22 (52.4)	29 (70.7)
Economic status, n (%)			
High	5 (6.0)	3 (7.1)	2 (4.9)
Middle	47 (56.6)	21 (50.0)	26 (63.4)
Low	31 (37.4)	18 (42.9)	13 (31.7)
Marital status, n (%)			
Never married	74 (89.2)	39 (92.9)	35 (85.3)
Married	7 (8.4)	3 (7.1)	4 (9.8)
Divorce	2 (2.4)	0 (0.0)	2 (4.9)
Residence, n (%)			
Alone	45 (54.2)	26 (61.9)	19 (46.3)
With family	26 (31.3)	13 (31.0)	13 (31.7)
With friend	12 (14.5)	3 (7.1)	9 (22.0)
Employment, n (%)			
Not working	17 (20.5)	10 (23.8)	7 (17.1)
Employed (temporary)	18 (21.7)	9 (21.4)	9 (21.9)
Employed (regular)	48 (57.8)	23 (54.8)	25 (61.0)
Gender orientation, n (%)			
Homosexual	62 (74.7)	30 (71.4)	32 (78.1)
Heterosexual	12 (14.5)	6 (14.3)	6 (14.6)
Bisexual	9 (10.8)	6 (14.3)	3 (7.3)
Mobile health, mean (SD)			
Need (out of 10 points)	7.04 (2.43)	7.17 (2.58)	6.90 (2.30)
Use (hour/day)	4.28 (3.41)	4.30 (4.09)	4.27 (2.58)
Short-form health literacy, mean (SD)			
Scale (out of 4 points)	3.04 (0.62)	3.22 (0.55)	2.87 (0.65)
Years after diagnosis[a]	9.51 (7.04)	8.83 (7.45)	10.27 (7.68)
HIV status, n (%)			
No symptoms	72 (86.8)	37 (88.1)	35 (85.4)
Have symptoms	4 (4.8)	2 (4.8)	2 (4.9)
AIDS	7 (8.4)	3 (7.1)	4 (9.7)
Visit a hospital within 6 months, n (%)	79 (95.2)	40 (95.2)	39 (95.1)
2-week uninterrupted ART[b], n (%)	65 (78.3)	36 (85.7)	29 (70.7)
CD4+T[c] cell count, n (%)			

XSL•FO
RenderX

Variable	Total (n=83)	App group (n=42)	Link group (n=41)
<200	5 (6.0)	2 (4.8)	3 (7.3)
200~499	13 (15.7)	9 (21.4)	4 (9.8)
500 or over	57 (68.7)	25 (59.5)	32 (78.0)
Unknown	8 (9.6)	6 (14.3)	2 (4.9)
Viral load, n (%)			
Undetectable	71 (85.6)	37 (88.1)	34 (82.9)
Detectable	6 (7.2)	1 (2.4)	5 (12.2)
Unknown	6 (7.2)	4 (9.5)	2 (4.9)
WHOQoL-HIV BREF[d], mean (SD)			
Psychological domain	61.30 (14.04)	64.19 (14.00)	58.34 (13.62)
Transmission mode, n (%)			
Heterosexual intercourse	7 (8.4)	2 (4.7)	5 (12.2)
Homosexual	58 (69.9)	33 (78.6)	25 (61.0)
Injection of blood, etc	18 (21.7)	7 (16.7)	11 (26.8)
Self-help group participation, n (%)	19 (22.9)	12 (28.6)	7 (17.1)

[a]In the mobile link group, 4 data points were missing because the participants did not remember their diagnosis year.

[b]ART: antiretroviral therapy.

[c]CD4+ T: CD4 T lymphocyte.

[d]WHOQoL-HIV BREF: World Health Organization Quality of Life–HIV questionnaire.

Outcomes

The difference in mean values of each group of primary and secondary outcomes between baseline (T0) and postintervention (T1) is shown in Table 2. During the intervention period, the overall outcomes improved in the app group, except for medication adherence. The results of the paired *t* test showed that the total score and the managing fatigue domain of the HIV-SE significantly improved in the app group. The link group also showed significant improvements in 3 domains: managing depression or mood, managing fatigue, and getting support or help. However, there was no significant improvement in secondary outcomes in either group. At 8 weeks (T2), the voluntary use for 4 weeks after completing the intervention, the overall primary outcomes decreased from the measured score at T1 (Table 3), but managing depression or mood in the link group increased over 8 weeks.

XSL•FO
RenderX

Table . Differences in the mean of outcomes during the intervention period.

Variables	App group (n=38), mean (SD)			Link group (n=35), mean (SD)		
	T0[a]	T1[a]	*P* value[b]	T0	T1	*P* value[b]
Self-efficacy for HIV self-management	6.82 (1.51)	7.19 (1.37)	.03	6.37 (1.27)	6.61 (1.31)	.07
Managing depression or mood	6.47 (1.97)	6.72 (1.78)	.36	5.57 (1.49)	6.03 (1.63)	.04
Managing medication	8.61 (1.80)	8.74 (1.45)	.65	8.67 (1.25)	8.81 (1.10)	.28
Managing symptoms	7.40 (2.28)	7.94 (1.52)	.08	7.24 (1.59)	7.21 (1.86)	.47
Managing fatigue	6.60 (2.31)	7.32 (1.92)	.03	5.74 (2.03)	6.21 (2.00)	.02
Communication with HCP[c]	7.92 (2.18)	8.03 (2.11)	.95	7.76 (2.35)	7.53 (2.14)	.35
Getting support or help	4.02 (2.13)	4.64 (2.55)	.10	3.56 (1.82)	3.97 (1.98)	.03
Medication adherence	95.50 (16.31)	94.71 (16.97)	.58	97.17 (5.26)	97.23 (5.77)	.56
Self-efficacy about condom use	3.64 (1.41)	3.97 (1.17)	.24	3.41 (1.18)	3.46 (1.22)	.40
Health-related quality of life	72.69 (15.31)	76.11 (18.89)	.16	73.37 (13.38)	73.77 (20.35)	.93

[a]T0=baseline; T1=4 weeks follow up.

[b]Results based on paired *t* test.

[c]HCP: health care provider.

XSL•FO
RenderX

Table . Comparison of self-efficacy for HIV self-management between the mobile app and link group.

Variable	App group (n=42), mean (SD)	Link group (n=41), mean (SD)	App group*Time interaction effects	
			β (95%CI)	*P* value[a]
Self-efficacy for HIV self-management				
T0[b]	6.82 (1.51)	6.37 (1.27)	—[c]	—
T1[b]	7.19 (1.37)	6.61 (1.31)	0.029 (−0.301 to 0.358)	.86
T2[b]	6.83 (1.42)	6.53 (1.38)	−0.264 (−0.649 to 0.120)	.18
Managing depression or mood				
T0	6.47 (1.97)	5.57 (1.49)	—	—
T1	6.72 (1.78)	6.03 (1.63)	−0.204 (−0.755 to 0.347)	.47
T2	6.38 (1.95)	6.30 (1.55)	−0.824 (−1.448 to −0.200)	.01
Managing medication				
T0	8.61 (1.80)	8.67 (1.25)	—	—
T1	8.74 (1.45)	8.81 (1.10)	−0.219 (−0.612 to 0.174)	.27
T2	8.75 (1.51)	8.49 (1.32)	0.051 (−0.433 to 0.536)	.84
Managing symptoms				
T0	7.40 (2.28)	7.24 (1.59)	—	—
T1	7.94 (1.52)	7.21 (1.86)	0.635 (0.023 to 1.247)	.04
T2	7.25 (2.26)	6.82 (2.09)	0.296 (−0.403 to 0.996)	.41
Managing fatigue				
T0	6.60 (2.31)	5.74 (2.03)	—	—
T1	7.32 (1.92)	6.21 (2.00)	0.073 (−0.531 to 0.678)	.81
T2	6.55 (2.52)	5.95 (2.06)	−0.346 (−1.110 to 0.417)	.37
Communication with HCP[d]				
T0	7.92 (2.18)	7.76 (2.35)	—	—
T1	8.03 (2.11)	7.53 (2.14)	0.099 (0.570 to 0.768)	.77
T2	8.13 (1.82)	7.70 (2.21)	−0.060 (−0.752 to 0.632)	.87
Getting support or help				
T0	4.02 (2.13)	3.56 (1.82)	—	—
T1	4.64 (2.55)	3.97 (1.98)	0.047 (−0.719 to 0.814)	.90
T2	4.24 (2.34)	3.83 (1.95)	−0.166 (−0.790 to 0.458)	.60

[a]Results based on generalized estimating equations adjusted for the duration of HIV infection, short-form health literacy scale, and psychological domain for the World Health Organization Quality of Life–HIV questionnaire (group: mobile link group reference; time points: T1 and T2, with T0 as reference).

[b]T0=baseline; T1=4 weeks follow up; T2=8 weeks follow up.

[c]Not applicable.

[d]HCP: health care provider.

GEE was used for group comparisons of differential changes from T0 at T1 and T2. The effects of the app group compared with the changes in the link group over time points, as a reference at T0 in self-efficacy for HIV self-management, are shown in Table 3. Regarding symptom management, the app group showed a significant improvement at T1, but this effect was not observed at T2. On the other hand, a significant reduction was observed at T2 in scores for managing depression or mood compared with the link group. There were no statistically significant findings for the other outcome variables in the GEE analysis.

Usability Test of the Mobile App

Regarding the quantitative usability of the ESSC app during the 4-week intervention period, 42.9% (18/42) of app group participants completed the in-app logs fully, with medication

XSL•FO
RenderX

adherence at 54.8% (23/42), mental diary at 54.8% (23/42), sexual diary at 64.3% (27/42), and treatment retention at 97.6% (41/42) (Table 4). Only 9.5% (4/42) maintained daily medication records from the completion of the intervention to the voluntary use period of 4 - 8 weeks. The periods of app use per participant ID are presented in Figure 3. Individuals who recorded daily medication adherence tended to record weekly mental and sexual diaries.

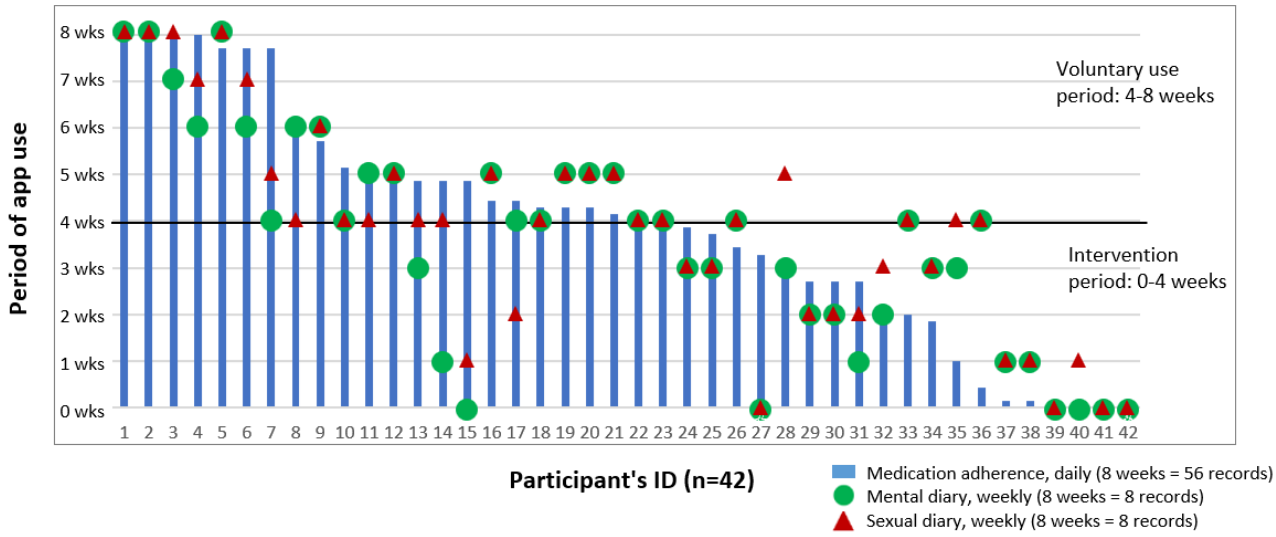
Table . Number of users and records of mobile app group during the intervention period (n=42).

Categories	Users, n (%)	Completers, n (%)	Records, n (%)
Medication adherence (daily)	40 (95.2)	23 (54.8)	877 (100)
Took ART[a]	40 (95.2)		855 (97.5)
Have symptoms	8 (19.0)		40 (4.6)
Sexual life (weekly)	38 (90.5)	27 (64.3)	128 (100)
Have sexual intercourse	19 (45.2)		31 (24.2)
If yes, condom usage	10 (23.8)		14 (10.9)
Mental diary (weekly)	36 (85.7)	23 (54.8)	120 (100)
Feeling sad	7 (16.7)		12 (10.0)
Retention in care (recent history)	41 (97.6)	41 (97.6)	45 (100)
Visit a hospital within 6 months	39 (92.9)		44 (97.8)
CD4+T[b] cell (500 or over)	27 (64.3)		30 (66.7)
Viral load (undetectable)	33 (78.6)		36 (80.0)

[a]ART: antiretroviral therapy.

[b]CD4+T: CD4+T lymphocyte.

Figure 3. Graph of the app usage by mobile app participant ID over 8 weeks.



The mean scale of Health-ITUES after the intervention period was 3.98 (SD 0.75) out of 5 points. The 4 subscale scores were as follows: impact (3.93, SD 0.95), perceived usefulness (3.86, SD 0.99), perceived ease of use (4.32, SD 0.71), and user control (3.95, SD 0.80). While it was easy for the participants to become skillful users of the ESSC app, the impact on a healthy life and the usefulness of supporting HIV self-management did not exceed 4 points. Qualitative feedback regarding the ESSC app or the link is shown in Textbox 1. Participants believed that the information was easy to understand and important; they wanted to regularly receive updated information, such as up-to-date treatment (eg, long-acting injectable HIV ART).

XSL•FO
RenderX

Textbox 1. Qualitative feedback after using the Excellent Self-Supervised HIV Care app or the link.

The Excellent Self-Supervised HIV Care app in the app group:

A. Information

- Positive: “Easy to understand without difficult terms” and “Essential and comprehensive content”
- Negative: “Lack of up-to-date information (eg, latest treatment development trend, case of cure, and clinical study results),” “Lack of details and diversity (eg, specific opportunistic infections such as condyloma),” and “Need to upload new content each week”

B. Self-record health life

- Positive: “Care more about taking medicine than before”
- Negative: “Unnecessary for smart patients with good medication adherence” and “Mind diary only suitable for depressed people”

C. Engagement

- Positive: “A cute character” and “Appropriate for the initial stage”
- Negative: “Insufficient motivation to use” and “Need for community features to communicate between users”

Accessing information in the link group:

A. Information

- Positive: “Well-organized information” and “Accurate, reliable information”
- Negative: "Needs to be constantly updated," "Pursue more details," and "Enlarge health information beyond HIV treatment"

B. Digital resource

- Positive: “An opportunity to read and study on my own, unlike consulting with medical staff” and “Easy to remember like an ebook”
- Negative: “Too large volume to view on a mobile phone”

Discussion

Principal Results

This study is the first trial to conduct fully non–face-to-face digital interventions for community-dwelling people living with HIV in Korea. These digital interventions have been designed to expand access to health information resources that are limited to hospital settings. Two-arm interventions were performed using a mobile app or link in text messages, which are popular implementation methods for mHealth. Comprehensive information that helps people living with HIV self-manage their HIV disease and a healthy life was provided by logging into the app or accessing links in the text messages. While the mobile link provided only information, the mobile app had functions to self-record and view their recorded health data. After 4 weeks, the total self-efficacy score for HIV management increased from the baseline in both groups, and a statistically significant improvement was found in the app group. Although the GEE analysis revealed no significant between-group difference in total self-efficacy score over 8 weeks, domain-specific comparisons indicated that the app group experienced a significant positive effect in managing symptoms at 4 weeks and a negative effect in managing depression or mood at 8 weeks. Our findings suggest that the information-driven mobile app helps improve self-efficacy for HIV disease management and that the function of self-recording health life has benefits for symptom management.

Informational mHealth Intervention Implementation

Informational support is the most frequently requested and provided form of assistance within online support groups [37]. In this study, 9 in 10 participants were single men, and half lived alone; they spent more than 4 hours a day using the app or accessing the internet via smartphone. Despite being well-positioned to engage in online support groups, only a quarter of participants had experienced participating in self-help groups. Previous studies on online-based relationships among people living with HIV have identified challenges such as difficulty maintaining relationships, concerns about HIV disclosure, and the spread of inaccurate information [38]. In line with a prior mHealth needs survey, it was found that people living with HIV desired evidence-based and trusted information first, and easy-to-understand content [11]. They reported that the need to communicate with other people living with HIV was low, but the need for HCP was high [11]. In this study, mHealth interventions through a mobile app or link provided validated information from HCP across 6 domains: HIV management, ART care, psychological health care, sexual health care, health promotion, and family and social life, which covered a holistic approach to healthy life. While both groups demonstrated improvements in total self-efficacy scores, no improvements were observed in the domains of managing medication or communication with HCP. Similarly, medication adherence did not improve, which aligns with previous studies showing that digital interventions often have limited impact on adherence among people living with HIV who are already taking ART consistently [39,40]. Most participants had maintained HIV treatment and viral suppression for years, with an average

XSL•FO
RenderX

of 10 years since HIV diagnosis, reflecting long-term retention in care. While therapeutic communication with HCPs has been emphasized for individuals newly diagnosed with HIV [41], information provision alone may not enhance communication self-efficacy among those already familiar with their HCPs. Furthermore, although the informational content addressed diverse domains, such as sexual health care and social life, it did not improve self-efficacy in condom usage and health-related quality of life.

Effectiveness of Self-Record Health Life in the App Group

The app group showed a significant improvement in self-efficacy for managing symptoms at 4 weeks compared with the link group. According to self-recorded symptoms on the ESSC app, a quarter of participants experienced symptoms that lasted from a minimum of one day up to 19 days. As symptom experiences are adverse signs of HIV progression and strongly related to the quality of life of people living with HIV, previous studies have developed mobile apps aimed at symptom management [42,43]. Similar to our study, a 4-week intervention using an app with educational content, self-assessment, and health tracking was shown to enhance symptom management knowledge [43]. Unlike that study, which included weekly group sessions on symptoms, our intervention did not provide customized interactive sessions. Instead, participants engaged in self-reflection through the “My Page” feature, where they could review their own health data. Interestingly, in another intervention study, web-based interactive sessions for patients did not significantly improve symptom management [44]. The authors explained that skill-learning processes may not effectively enhance self-efficacy in patients with high treatment adherence. Despite the absence of structured interactions, self-efficacy for managing symptoms improved, suggesting that self-recording alone may contribute to symptom self-management. Meanwhile, recording the mental diary of the ESSC app had no significant effect on managing depression or mood compared with the mobile link group. Rather, improvement was seen at 8 weeks in the link group that provided only psychological health care information. However, it is difficult to determine whether the mental diary had an adverse effect on mood management because the number of records was low between 4 and 8 weeks. Other apps specializing in symptom management [42,43], including psychological coping strategies for depression and anxiety severity, have been shown to help reduce psychological distress. User-customized psychiatric content beyond a mental diary is suggested for managing depressed moods.

Challenges of App Improvement

Only one in 10 participants maintained their daily visits after completion of the intervention, during the voluntary use period between 4 and 8 weeks. In other app interventions requiring a query response for medication adherence, the response rate was 47.7% over 2 years [45]. The app provides weekly summary reports, and its most popular feature is the community message board, which allows users to share and interact with other users on social networks. The ESSC app has features for viewing summary reports at any time on “My Page,” but there is no community board for user interaction. This reflected the low need for peer communication among people living with HIV in the mHealth needs survey [11]. However, participants in this study gave feedback that they lacked motivation to continue using the app and expressed a desire for a community feature. By sharing experiences and common issues, people living with HIV can tap opportunities to perceive the usefulness of support groups and reduce stigma [46]. The community board should be operated in a safe space so that users do not encounter interpersonal conflicts [37], which will be refined in the next version of the ESSC app improvement. In addition, to maintain engagement with the app, participants responded that information should be updated regularly and interesting content should be provided to people living with HIV, such as a case of cure and the latest treatment development. HIV treatments have continued to develop in the evolution of ARTs, including injectable medication, and this progress has extended life expectancy and improved the quality of life [47]. Accordingly, it is natural that people living with HIV desire to stay up-to-date on the treatment, and the ESSC app should incorporate a dedicated news section with the latest information.

Implications for mHealth Adoption

The link group, which accessed digital information via individual text messages without signing up for the app, showed significantly improved self-efficacy in managing depression or mood, fatigue, and getting support or help. While the mHealth app is an emerging technology for empowering patients’ self-management of chronic illnesses, it must also address potential risks to confidentiality and ensure patient autonomy [48]. Some users in this study perceived the ESSC app as an unnecessary device for good patients. The study by Jacomet et al [10] of people living with HIV and their physicians’ use patterns, benefits, and perceived barriers to different aspects of eHealth reported that 26% of people living with HIV and 43% of physicians were enthusiastic about eHealth, whereas 31% and 37%, respectively, expressed skepticism. To adopt mHealth for patients, multiple factors should be considered, including user characteristics, literacy, system utility, technology infrastructure, collaboration, and regulatory settings [49]. Diversifying the range of mHealth services available to patients can reduce barriers to accessing digital health resources and contribute to health equity. Link-based access can be proposed as an alternative mHealth strategy to bridge the gap between people living with HIV who are hesitant to use apps and those who use app in the context of digital information technology.

Limitations

There are several limitations in interpreting the results. First, the inclusion criteria for smartphone users may not be representative of the generalized findings for people living with HIV. To ensure representativeness and health equity, further studies are needed to enhance community-based enrollment without limiting smartphone literacy. Second, although the final sample size exceeded the minimum required (n=60) based on a priori power analysis, it did not reach the target of 86 participants. While the final sample size was sufficient to detect moderate effect sizes, the reduced sample size may have limited the ability to identify smaller differences between groups. This

XSL•FO
RenderX

limitation should be considered when interpreting the findings. Third, the 4-week intervention period may have been too short to fully assess the effectiveness and usability of the app. To observe sustained engagement patterns, an additional 4-week voluntary-use period was provided. However, voluntary use during the follow-up period was notably low. We designed a relatively short intervention period, as the primary aim of this study was to evaluate the impact of 2 types of fully non-face-to-face, information-driven mHealth delivery formats. To enhance usability and engagement for longer-term interventions, future versions of the app should incorporate additional features, such as user-interactive community boards, that go beyond information provision and self-recorded healthy living. To determine the practical usefulness of the ESSC app, it is necessary to evaluate its long-term effectiveness following future app improvements.

Conclusions

The provision of integrated information using mobile devices improves HIV-SE among people living with HIV. Both mobile apps and link accessing in text messages are useful methods for providing non-face-to-face information. People living with HIV who have already been retained in care may not prefer to become online members of HIV-related apps, and mobile-based interventions should be applied according to individual preferences. While an informational mobile app that allows users to self-record and reflect on their healthy life improved self-efficacy for managing symptoms compared with link accessing in text messages, the app has challenges in maintaining long-term use. Usability findings suggest that up-to-date news sections and user-to-user community features are required for long-term engagement with informational mobile apps.

Acknowledgments

The authors express their deep gratitude to the people living with HIV who participated in the study. We are grateful for the support of Choazsoft Corporation in the process of operating the mobile app. This work was supported by the National Research Foundation of Korea grant funded by the South Korean Government (Ministry of Science and ICT; number 2020R1A2C101081713; principal investigator: GSK) and a Faculty Research Grant from Yonsei University College of Nursing (grant number 6-2022-0034). The corresponding author received scholarships from the Brain Korea 21 FOUR Project, funded by the National Research Foundation of Korea, Yonsei University College of Nursing.

Conflicts of Interest

None declared.

Checklist 1
CONSORT-EHEALTH V1.6 check list. CONSORT: Consolidated Standards of Reporting Trails.
[PDF File, 10009 KB - mhealth_v13i1e60905_app1.pdf]

Checklist 2
CONSORT 2010 checklist. CONSORT: Consolidated Standards of Reporting Trails.
[DOC File, 223 KB - mhealth_v13i1e60905_app2.doc]

References

1. U=U taking off in 2017. Lancet HIV 2017 Nov;4(11):e475. [doi: 10.1016/S2352-3018(17)30183-2]
2. Oh KS, Lee JS, Kim HC, Kang HY, Lee JY, Han E. Effects of depression on medication adherence in HIV/AIDS patients: Korea HIV/AIDS cohort study. J Infect Public Health 2023 Oct;16(10):1598-1605. [doi: 10.1016/j.jiph.2023.07.018] [Medline: 37573850]
3. Oh KS, Lee JS, Han E. Risk factors associated with medication adherence in HIV/AIDS patients. Korean J Clin Pharm 2023 Dec 31;33(4):254-260. [doi: 10.24304/kjcp.2023.33.4.254]
4. HIV/AIDS notifications in the Republic of Korea. Korea Disease Control and Prevention Agency. 2021. URL: www.kdca.go.kr/contents.es?mid=a20301070504 [accessed 2024-01-15]
5. Choi JP, Seo BK. HIV-related stigma reduction in the era of undetectable equals untransmittable: the South Korean perspective. Infect Chemother 2021 Dec;53(4):661-675. [doi: 10.3947/ic.2021.0127] [Medline: 34979602]
6. Choi JP, Lee JH, An JM, Kim J, Won N, Choi YH. The story and implications of the Korean health care facility counseling project on people living with HIV. Infect Chemother 2023 Jun;55(2):167-178. [doi: 10.3947/ic.2023.0024] [Medline: 37272236]
7. Shah R, Watson J, Free C. A systematic review and meta-analysis in the effectiveness of mobile phone interventions used to improve adherence to antiretroviral therapy in HIV infection. BMC Public Health 2019 Jul 9;19(1):915. [doi: 10.1186/s12889-019-6899-6] [Medline: 31288772]
8. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management In HIV: a systematic review. Open AIDS J 2017;11(119-32):119-132. [doi: 10.2174/1874613601711010119] [Medline: 29290888]

XSL•FO
RenderX

9. Marent B, Henwood F, Darking M, EmERGE Consortium. Development of an mHealth platform for HIV care: gathering user perspectives through co-design workshops and interviews. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e184. [doi: 10.2196/mhealth.9856] [Medline: 30339132]
10. Jacomet C, Ologeanu-Taddei R, Prouteau J, et al. Adoption and attitudes of eHealth among people living with HIV and their physicians: online multicenter questionnaire study. JMIR Mhealth Uhealth 2020 Apr 15;8(4):e16140. [doi: 10.2196/16140] [Medline: 32293581]
11. Kim GS, Kim L, Baek S, et al. Three cycles of mobile app design to improve HIV self-management: a development and evaluation study. Digit HEALTH 2024;10:20552076241249294. [doi: 10.1177/20552076241249294] [Medline: 38680735]
12. Kim GS, Baek S, Kim N, et al. Network visualization to interpret which healthcare services are central to people living with HIV. J Adv Nurs 2024 Oct;80(10):4135-4146. [doi: 10.1111/jan.16137] [Medline: 38444110]
13. Continuity and Coordination of Care: A Practice Brief to Support Implementation of the WHO Framework on Integrated People-Centred Health Services: World Health Organization; 2018. URL: https://iris.who.int/handle/10665/274628 [accessed 2025-05-17]
14. Kim GS, Kim L, Shim MS, et al. Psychometric properties of the Korean version of self-efficacy for HIV disease management skills. J Korean Acad Nurs 2023 Jun;53(3):295-308. [doi: 10.4040/jkan.23016] [Medline: 37435761]
15. Lee SA, Kim EY, Chang SO. A life where concealment and exposure coexist-the antiretroviral therapy adherence of people living with HIV infection: a phenomenological study. J Adv Nurs 2020 Apr;76(4):1027-1036. [doi: 10.1111/jan.14309] [Medline: 32003040]
16. Kim K, Park W. Effects of mobile navigation program in colorectal cancer patients based on uncertainty theory. J Korean Acad Nurs 2019 Jun;49(3):274-285. [doi: 10.4040/jkan.2019.49.3.274] [Medline: 31266924]
17. Peterson C, Seligman ME. Character Strengths and Virtues: A Handbook and Classification: Oxford University Press; 2004.
18. Nurmi J, Knittle K, Ginchev T, et al. Engaging users in the behavior change process with digitalized motivational interviewing and gamification: development and feasibility testing of the precious app. JMIR Mhealth Uhealth 2020 Jan 30;8(1):e12884. [doi: 10.2196/12884] [Medline: 32003750]
19. Shively M, Smith TL, Bormann J, Gifford AL. Evaluating self-efficacy for HIV disease management skills. AIDS Behav 2002;6(4):371-379. [doi: 10.1023/A:1021156914683]
20. Sharma N, Sokhey G. Role of self-efficacy in quality of life of people living with HIV/AIDS. Indian J Health Wellbeing 2013;4(1).
21. Huang L, Li L, Zhang Y, Li H, Li X, Wang H. Self-efficacy, medication adherence, and quality of life among people living with HIV in Hunan Province of China: a questionnaire survey. J Assoc Nurses AIDS Care 2013;24(2):145-153. [doi: 10.1016/j.jana.2012.04.006] [Medline: 22770801]
22. Li X, Huang L, Wang H, Fennie KP, He G, Williams AB. Stigma mediates the relationship between self-efficacy, medication adherence, and quality of life among people living with HIV/AIDS in China. AIDS Patient Care STDS 2011 Nov;25(11):665-671. [doi: 10.1089/apc.2011.0174] [Medline: 22023316]
23. Kim GS, Kim L, Lee S, Shim MS, Lee Y, Baek S. Factors associated with health-related quality of life among people living with HIV in South Korea: Tobit regression analysis. PLOS ONE 2024;19(5):e0303568. [doi: 10.1371/journal.pone.0303568] [Medline: 38753733]
24. Simoni JM, Kurth AE, Pearson CR, Pantalone DW, Merrill JO, Frick PA. Self-report measures of antiretroviral therapy adherence: a review with recommendations for HIV research and clinical management. AIDS Behav 2006 May;10(3):227-245. [doi: 10.1007/s10461-006-9078-6] [Medline: 16783535]
25. Walsh JC, Mandalia S, Gazzard BG. Responses to a 1 month self-report on adherence to antiretroviral therapy are consistent with electronic data and virological treatment outcome. AIDS 2002 Jan 25;16(2):269-277. [doi: 10.1097/00002030-200201250-00017] [Medline: 11807312]
26. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 5;6(1):e4. [doi: 10.2196/mhealth.8851] [Medline: 29305343]
27. Lee J, Schnall R. Validity and reliability of the Korean version of the health information technology usability evaluation scale: psychometric evaluation. JMIR Med Inform 2022 Jan 24;10(1):e28621. [doi: 10.2196/28621] [Medline: 35072630]
28. Seo YJ, Kwak EM, Jo M, Ko AR, Kim SH, Oh H. Reliability and validity of the Korean version of short-form health literacy scale for adults. J Korean Acad Community Health Nurs 2020;31(4):416. [doi: 10.12799/jkachn.2020.31.4.416]
29. Duong TV, Chang PW, Yang SH, et al. A new comprehensive short-form health literacy survey tool for patients in general. Asian Nurs Res (Korean Soc Nurs Sci) 2017 Mar;11(1):30-35. [doi: 10.1016/j.anr.2017.02.001] [Medline: 28388977]
30. Lee WK, Kim SW, Kim HI, et al. Development of quality of life with WHOQOL-HIV BREF Korean version among HIV patients in Korea. JKDISS 2014 Mar 31;25(2):337-347. [doi: 10.7465/jkdi.2014.25.2.337]
31. WHOQOL-HIV bref, 2012 revision. : World Health Organization; 2002.
32. Selik RM, Mokotoff ED, Branson B, Owen SM, Whitmore S, Hall HI. Revised surveillance case definition for HIV infection—United States, 2014. : Morbidity and Mortality Weekly Report (MMWR); 2014 URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr6303a1.htm [accessed 2025-05-23]

XSL•FO
RenderX

33. Rodkjaer L, Chesney MA, Lomborg K, Ostergaard L, Laursen T, Sodemann M. HIV-infected individuals with high coping self-efficacy are less likely to report depressive symptoms: a cross-sectional study from Denmark. Int J Infect Dis 2014 May;22:67-72. [doi: 10.1016/j.ijid.2013.12.008] [Medline: 24657129]
34. Webel AR, Sattar A, Schreiner N, Kinley B, Moore SM, Salata RA. The impact of mental wellness on HIV self-management. J Assoc Nurses AIDS Care 2016;27(4):468-475. [doi: 10.1016/j.jana.2016.03.002] [Medline: 27066751]
35. Wolf MS, Davis TC, Osborn CY, Skripkauskas S, Bennett CL, Makoul G. Literacy, self-efficacy, and HIV medication adherence. Patient Educ Couns 2007 Feb;65(2):253-260. [doi: 10.1016/j.pec.2006.08.006] [Medline: 17118617]
36. Fauzi A, Anggraini N, Fatkhurohman N. Self-management: a comprehensive approach to improve quality of life among people living with HIV in Indonesia. Belitung Nurs J 2021;7(5):395-401. [doi: 10.33546/bnj.1554] [Medline: 37496503]
37. Coulson NS, Buchanan H. The role of online support groups in helping individuals affected by HIV and AIDS: scoping review of the literature. J Med Internet Res 2022 Jul 26;24(7):e27648. [doi: 10.2196/27648] [Medline: 35881456]
38. Mo PKH, Coulson NS. Are online support groups always beneficial? A qualitative exploration of the empowering and disempowering processes of participation within HIV/AIDS-related online support groups. Int J Nurs Stud 2014 Jul;51(7):983-993. [doi: 10.1016/j.ijnurstu.2013.11.006] [Medline: 24342449]
39. Arayasirikul S, Turner C, Trujillo D, Le V, Wilson EC. Efficacy and impact of digital HIV care navigation in young people living with HIV in San Francisco, California: prospective study. JMIR Mhealth Uhealth 2020 May 8;8(5):e18597. [doi: 10.2196/18597] [Medline: 32383680]
40. Shim MS, Kim S, Choi M, Choi JY, Park CG, Kim GS. Developing an app-based self-management program for people living with HIV: a randomized controlled pilot study during the COVID-19 pandemic. Sci Rep 2022 Nov 12;12(1):19401. [doi: 10.1038/s41598-022-19238-w] [Medline: 36371442]
41. Kim GS, Shim MS, Choi JY, Lee JI, Kim JM. Testing a question prompt intervention to improve communication between patients with HIV and healthcare providers: a pilot study. J Community Health Nurs 2020;37(3):153-163. [doi: 10.1080/07370016.2020.1780045] [Medline: 32820975]
42. Schnall R, Cho H, Mangone A, Pichon A, Jia H. Mobile health technology for improving symptom management in low income persons living with HIV. AIDS Behav 2018 Oct;22(10):3373-3383. [doi: 10.1007/s10461-017-2014-0] [Medline: 29299790]
43. Han S, Pei Y, Zhao R, et al. Effects of a symptom management intervention based on group sessions combined with a mobile health application for persons living with HIV in China: a randomized controlled trial. Int J Nurs Sci 2021 Oct 10;8(4):370-379. [doi: 10.1016/j.ijnss.2021.07.002] [Medline: 34631986]
44. Côté J, Godin G, Ramirez-Garcia P, et al. Virtual intervention to support self-management of antiretroviral therapy among people living with HIV. J Med Internet Res 2015 Jan 6;17(1):e6. [doi: 10.2196/jmir.3264] [Medline: 25563775]
45. Laurence C, Wispelwey E, Flickinger TE, et al. Development of PositiveLinks: a mobile phone app to promote linkage and retention in care for people with HIV. JMIR Form Res 2019 Mar 20;3(1):e11578. [doi: 10.2196/11578] [Medline: 30892269]
46. Gadgil G, Prybutok G, Prybutok V. Qualitative investigation of the role of quality in online community support for people living with HIV and AIDS. Qual Manag J 2018 Oct 2;25(4):171-185. [doi: 10.1080/10686967.2018.1515524]
47. Haris M, Abbas R. Four decades of HIV: global trends, testing assays, treatment, and challenges. Zoonoses 2024;4(1):997. [doi: 10.15212/ZOONOSES-2023-0039]
48. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817] [Medline: 31293246]
49. Addotey-Delove M, Scott RE, Mars M. Review of patients' perspectives of m-health adoption factors in the developing world. development of a proposed conceptual framework. Inf Med Unlocked 2020;21:100460. [doi: 10.1016/j.imu.2020.100460]

Abbreviations

ART: antiretroviral therapy
CD4+T: CD4+T lymphocyte
ESSC: Excellent Self-Supervised HIV Care
GEE: generalized estimating equation
HCP: health care provider
Health-ITUES: Health Information Technology Usability Evaluation Scale
HIV-SE: self-efficacy for HIV management
mHealth: mobile health
SF-HLS: Short-Form Health Literacy Scale
VAS: visual analogue scale
WHOQoL-HIV BREF: World Health Organization Quality of Life–HIV questionnaire

XSL•FO
RenderX

Edited by L Buis; submitted 24.05.24; peer-reviewed by S Adeyemi, Y Pang, Y Li; revised version received 22.04.25; accepted 23.04.25; published 29.05.25.

Please cite as:
Kim GS, Kim L, Baek S, Kwon S, Kim JM, Choi JY, Choi JP
Providing 2 Types of mHealth Interventions to Support Self-Management Among People Living With HIV: Randomized Clinical Trial
JMIR Mhealth Uhealth 2025;13:e60905
URL: https://mhealth.jmir.org/2025/1/e60905
doi:10.2196/60905

© Gwang Suk Kim, Layoung Kim, Seoyoung Baek, Sooyoung Kwon, Ji Min Kim, Jun Yong Choi, Jae-Phil Choi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Preferences for Mobile App Features to Support People Living With Chronic Heart Diseases: Discrete Choice Study

Sumudu Avanthi Hewage[1], MBBS, MSc, MD; Sameera Senanayake[1,2,3], MBBS, MSc, MD, PhD; David Brain[1], PhD; Michelle J Allen[1], PhD; Steven M McPhail[1,4], PhD; William Parsonage[1,5,6], MD; Tomos Walters[6,7], MD, PhD; Sanjeewa Kularatna[1,2,3], MSc, DDS, MD, PhD

[1]Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, School of Public Health and Social Work, Queensland University of Technology, 61 Musk Avenue, Brisbane, Australia

[2]Health services and systems research, Duke-NUS Medical School, 8, College Road, Singapore

[3]National Heart Research Institute Singapore, National Heart Centre, 5, Hospital Drive, Singapore

[4]Digital Health and Informatics Directorate, Metro South Health, Queensland, Australia

[5]Department of Cardiology, Royal Brisbane and Women's Hospital, Queensland, Australia

[6]Queensland Cardiovascular Group, Queensland, Australia

[7]Faculty of Medicine, The University of Queensland, Queensland, Australia

Corresponding Author:
Sumudu Avanthi Hewage, MBBS, MSc, MD
Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, School of Public Health and Social Work, Queensland University of Technology, 61 Musk Avenue, Brisbane, Australia

Abstract

Background: Using digital health technologies to aid individuals in managing chronic diseases offers a promising solution to overcome health service barriers such as access and affordability. However, their effectiveness depends on adoption and sustained use, influenced by user preferences.

Objectives: This study quantifies the preferences of individuals with chronic heart disease (CHD) for features of a mobile health app to self-navigate their disease condition.

Methods: We conducted an unlabeled web-based choice survey among adults older than 18 years with CHD living in Australia, recruited via a web-based survey platform. Four app attributes—ease of navigation, monitoring of blood pressure and heart rhythm, health education, and symptom diary maintenance—were systematically chosen through a multistage process. This process involved a literature review, stakeholder interviews, and expert panel discussions. Participants chose a preferred mobile app out of 3 alternatives: app A, app B, or neither. A D-optimal design was developed using Ngene software, informed by Bayesian priors derived from pilot survey data. Latent class model analysis was conducted using Nlogit software (Econometric Software, Inc). We also estimated attribute importance and anticipated adoption rates for 3 app versions.

Results: Our sample included 302 participants with a mean age of 50.5 (SD 18.2) years. Latent class model identified 2 classes. Older respondents with education beyond high school, prior experience with mobile health apps, and a positive perception of app usefulness were more likely to be in class 1 (257/303, 85%) than in class 2 (45/303, 15%). Class 1 membership preferred adopting a mobile app (app A: β coefficient 0.74, 95% uncertainty interval (UI) 0.41-1.06; app B: β coefficient 0.53, 95% UI 0.22-0.85). Participants favored apps providing postmonitoring recommendations (β coefficient 1.45, 95% UI 1.26-1.64), tailored health education (β coefficient 0.50, 95% UI 0.36-0.64), and unrestricted symptom diary entry (β coefficient 0.58, 95% UI 0.41-0.76). Class 2 showed no preference for app adoption (app A: β coefficient −1.18, 95% UI −2.36 to 0.006; app B: β coefficient −0.78, 95% UI −1.99 to 0.42) or any specific attribute levels. Vital sign monitoring was the most influential attribute among the 4. Scenario analysis revealed an 84% probability of app adoption with basic features, rising to 92% when app features aligned with respondents' preferences.

Conclusions: The study's findings suggest that designing preference-informed mobile health apps could significantly enhance adoption rates and engagement among individuals with CHD, potentially leading to improved clinical outcomes. Adoption rates were notably higher when app attributes included easy navigation, vital sign monitoring, feedback provision, personalized health education, and flexible data entry for symptom diary maintenance. Future research to explore factors influencing app adoption among different groups of patients is warranted.

(JMIR Mhealth Uhealth 2025;13:e58556) doi:10.2196/58556

XSL•FO
RenderX

KEYWORDS

digital health technologies; user preferences; latent class model; monitoring vital signs; adoption rates; app; chronic heart disease; heart disease; digital health; effectiveness; user; mobile health app; self-navigate; health education; symptom; monitoring; adoption

Introduction

Chronic heart diseases (CHDs) pose a significant global health challenge, with cases doubling from 271 million in 1990 to 523 million in 2019, leading to a rise in related deaths and disability-adjusted life years, particularly in regions where rates had previously declined [1]. Addressing this global scenario necessitates urgent attention to implementing currently available policies and interventions. However, widespread implementation of interventions that are effective in prevention and management of CHD is hindered by factors such as limited accessibility and affordability, demanding innovative solutions [2,3]. Another barrier to effective management of CHD is the lack of personalized care planning tailored to patient priorities and social contexts, which are vital in providing high quality, patient-centered care [4,5].

Digital health technologies offer promising solutions to overcome some of these barriers [6-8], enhancing clinical outcomes and inducing behavioral changes among individuals with CHD [9-11]. Recent evidence suggests favorable cost-effectiveness outcomes for interventions using digital health technologies, which may aide in the optimization of health care resource usage [12]. Given the necessity for users to take an active role, the widespread adoption and sustained usage of digital health interventions may pose challenges [13,14]. Aligning technology with user preferences increases the probability that intended populations adopt and enjoy its use [15]. Therefore, understanding user preferences is imperative for the successful implementation and sustained use of digital health interventions.

There are numerous methods used to elicit preferences in health preference research. Among these, choice-based methods such as discrete choice experiments (DCEs) are arguably the best known and most commonly used, offering valuable means to systematically analyze user preferences [15,16]. With the rise in the focus toward patient-centered care, DCEs are increasingly used in a range of health policy, planning and resource allocation decisions across disease prevention, diagnosis and treatment, access to services, and health care employment [17,18]. In the framework of a DCE, respondents evaluate alternatives characterized by attribute-level combinations, selecting their preferred option [16]. Choice modeling analysis, rooted in "Random Utility Theory" [19], assumes that this preferred option presents the highest utility for the respondent and can, therefore, quantify preferences and discern overall inclinations toward attribute levels [19].

Our study aimed to investigate the preferences of Australians living with CHD regarding the features of a mobile health app, designed to assist them. We believe that our findings can inform the development of preference-informed mobile apps, enhancing adoption and sustained usage and ultimately improving health outcomes.

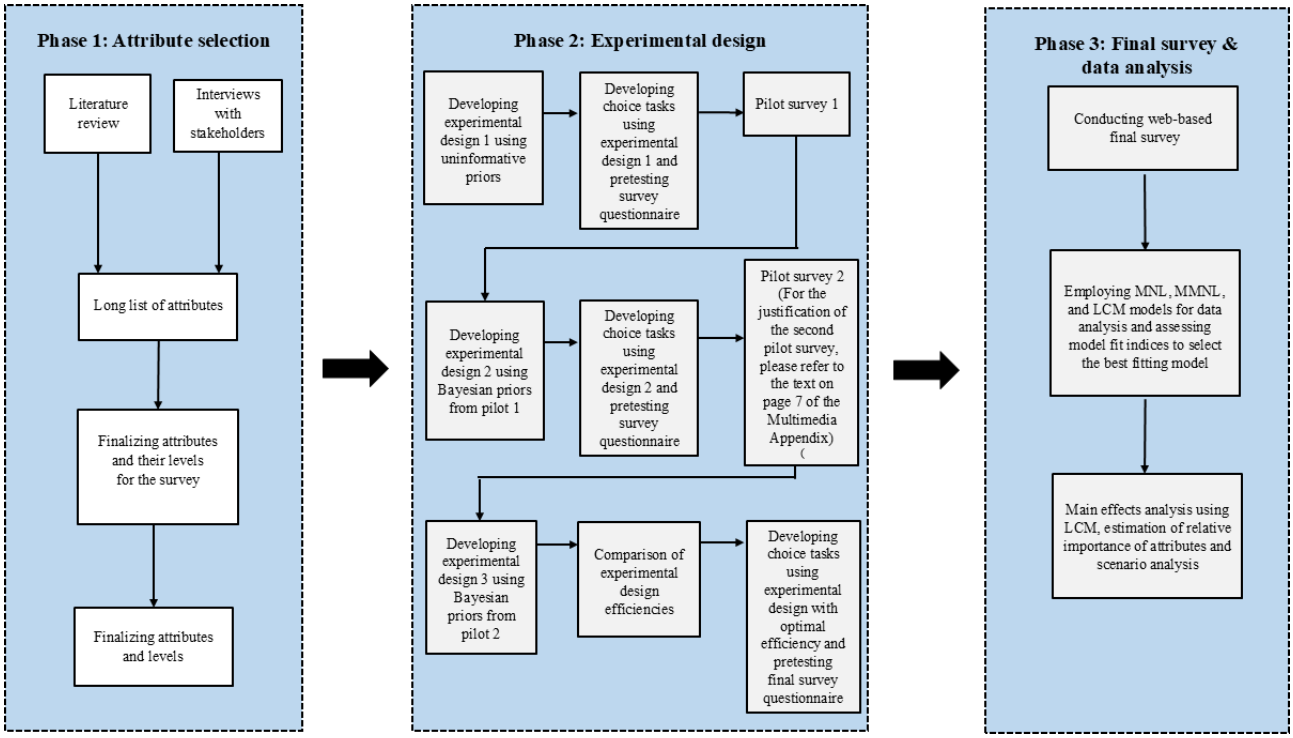
Methods

Study Design

The manuscript was prepared in accordance with the DIRECT (Discrete Choice Experiment Reporting Checklist) checklist for DCEs, as detailed in Table S1 in Multimedia Appendix 1. We conducted the study using three phases: (1) attribute selection, (2) experimental design, and (3) final survey and data analysis, as illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Overview of the study methods. MNL: multinomial logit; MMNL: mixed multinomial logit; LCM: latent class model.



Phase 1: Attribute Selection

Literature Review to Explore Key Attributes

As an attribute-based experiment, the validity of a DCE heavily relies on appropriately specifying attributes and their levels [17]. However, there is no standard process for identifying and selecting attributes in DCE that may influence the decision of interest [20]. In stage 1, we reviewed existing literature to identify the factors that may affect the use of mobile apps developed for various diseases. The search strategy is shown in Table S2 in Multimedia Appendix 1. The search resulted in 32 papers published between 2012 and 2023, with 23 selected for our final analysis. We identified 38 attributes related to mobile apps from these papers. Figure S1 and Table S3 in Multimedia Appendix 1 provide the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram for the review and the list of attributes identified from the review, respectively.

Interviews With Stakeholders

In alignment with previously published recommendations, qualitative work was conducted during attribute development [21]. We interviewed 7 individuals with CHD and 8 health care professionals on the web or via telephone. The discussions yielded themes centered on the user-friendly nature of the app, the capacity of the app to assist in self-monitoring of disease conditions, the need for personalized health education, concerns about data security, and considerations regarding subscription charges. Finally, 2 more attributes were added to the list generated from the stage 1 literature review, bringing the total to 40 attributes.

Finalizing the Attributes and Their Levels

Including all attributes identified through literature review and talking to stake holders are impractical due to respondent cognitive burden and sample size limitations [20]. Therefore, authors SK and SH, with backgrounds in health economics and medicine, respectively, reviewed and condensed the list to 7 relevant and suitable attributes by excluding irrelevant ones and merging related ones. A panel of experts including a general cardiologist, a cardiac electrophysiologist, 3 health economists, and an implementation scientist evaluated the proposed 7 attributes and their levels. They determined the significance and applicability of these attributes, offering feedback on the proposed levels through several rounds of evaluation. Four attributes with the highest total scores were selected through group deliberation, each refined with 3 levels based on feedback. Table 1 details these attributes, definitions, levels, and expected preferences.

XSL•FO
RenderX

Table . Attributes and their levels selected for the study.

Attribute	Working definition	Levels	Expected direction[a]
Training	The level of training required to use the mobile app for the first time.	• Easy to use and requires no training.	Reference
		• Usable after a basic training for 15 minutes.	−
		• Usable after an advanced training for 30 minutes.	−−
Monitoring	Ability of the mobile app to monitor blood pressure and heart rhythm and provide recommendations on action.	• The app cannot measure your blood pressure or heart rhythm.	Reference
		• The app can measure your blood pressure and heart rhythm but does not provide recommendations on what you should do next.	+
		• The app can measure your blood pressure or heart rhythm and provide recommendations on what you should do next.	++
Health education	Availability and nature of health education messages within the mobile app.	• Health educational messages are not available in the app.	Reference
		• Health educational messages in the app are generalized (not tailored to your individual needs).	+
		• Health educational messages in the app are tailored to your individual needs.	++
Symptom diary	Ability of the mobile app to function as a diary to record symptoms or signs by the user.	• Keeping a diary of symptoms over time cannot be done in the app.	Reference
		• Keeping a diary of symptoms over time can be done, but it is limited to specific questions in the app.	+
		• Keeping a diary of symptoms over time can be done in the app without any restrictions.	++

[a]This column shows the expected direction for participants' preferences for each attribute level. The "+" and "++" symbols indicate positive direction, while the negative symbols in the cells above indicate negative direction. The number of symbols is an indication of its strength. For example, ++ indicates "strongly positive" while + indicates "positive."

Phase 2: Experimental Design

Developing Experimental Design 1 Using Uninformative Priors

In the context of DCE, experimental designs pertain to how options, comprising attributes and their respective levels, are presented to participants [22]. While full factorial designs encompass all possible combinations of attribute levels [23], they can be extensive, necessitating large sample sizes or many choice questions per respondent. In studies with a high number of possible choice question combinations, fractional factorial designs, a subset of attribute-level combinations, are recommended [23] and were selected for this study. Our study comprised a 2-alternative design using 4 attributes, each with 3 levels, yielding 81 possible profiles (3^4) and 3240 combinations of choice questions [$3^4\times(3^4-1)/2$] [24]. We chose a D-optimal design over an orthogonal design for its capacity to produce accurate parameter estimates with a smaller sample size [25].

XSL•FO
RenderX

One important consideration in D-optimal design is specifying the number of choice tasks or the design size [26]. According to the formula "number of attribute levels/(number of alternatives-1)," a minimum design size of 6 per block was required [27]. While larger design sizes typically improve statistical efficiency, they can compromise response efficiency [23]. We assessed the normalized D-error for various design sizes beyond the minimum requirement and chose a design size of 16 rows based on the percentage reduction in the normalized D-error (Table S4 in Multimedia Appendix 1). In addition, we used 2 blocks, presenting only 8 choice tasks per respondent, to further enhance response efficiency.

D-optimal designs require specifying parameter priors for each attribute level [28]. While informed priors generally lead to more efficient designs, applying incorrect priors may compromise the expected efficiency compared with uninformative priors [29]. To address this, we generated informative priors from a pilot survey among 67 respondents, using small directional prior values in the prepilot design. The last column of Table 1 indicates the assumed direction for all attribute levels, and small directional prior values used in this exercise are available in Table S5 in Multimedia Appendix 1.

Dummy coding was used to code attribute levels categorically, with the most basic level serving as the reference to interpret results [30]. The position of the opt-out alternative in the choice task was randomly varied so that all 3 alternatives had an equal chance of appearing in different positions. This was done to prevent order effects dependent on position [31].

We used Ngene software (Econometric Software, Inc) to generate our pilot experimental design [32]. The selected design was evaluated for attribute-level balance and minimal overlap, resulting in the experimental design for the first pilot survey. A design is considered balanced when each level of an attribute appears an equal number of times [33], and overlap refers to the repetition of specific attribute levels across a set of alternatives [22]. The Ngene code for our prepilot design is shown in Figure S2 in Multimedia Appendix 1, and an illustration of the Ngene design is shown in Figure S3 in Multimedia Appendix 1.

Developing and Pretesting the Survey Questionnaire

The selected attribute levels were then transformed to ensure meaningful presentation to participants. The survey questionnaire was pretested to ensure clarity and meaningful presentation to participants. Purposeful sampling was used to select nonacademic staff members from the research team's department for in-person pretesting, enabling direct feedback on paper copies of the experiment. In response to feedback, we refined the wording of specific levels to ensure clear communication of the intended meaning of attribute levels, as illustrated in Figure 2.

Figure 2. A choice task presented to respondents in this study.

Scenario:
Your clinician introduced you to a new mobile health app to help manage your heart condition. You have 2 versions to choose from, app A and app B, each with different features listed below. You can also choose not to use the app at all.
Which option would you prefer? Please click the circle at the end of the column to make your choice.

	app A	app B	Neither
This app is	easy to use and requires no training.	usable after a basic training of 15 mins.	I do not want to use app A or app B with these features.
This app can	not measure your blood pressure or heart rhythm.	measure your blood pressure and heart rhythm and give recommendations on what you should do next.	
Health education messages in the app are	Tailored to your personal needs.	Generic (not tailored to your individual needs).	
Keeping a diary of symptoms over time	Can be done but it is limited to specific questions in the app.	Can be done but it is limited to specific questions in the app.	
Which option do you prefer most?	○	○	○

Subsequently, a web-based pilot test was conducted among adults aged 18 years and older with CHD, recruited through the PureProfile platform, to estimate Bayesian priors for the final survey design. Each choice task began with an introductory scenario, prompting respondents to envision a mobile app offering to assist them with their CHD. In this unlabeled experiment, the alternatives were referred to as mobile app A and B. Following best practice recommendations in health DCEs, respondents were allowed to opt out if the presented combinations of attribute levels for either app did not align with their preferences [16,34].

While "Random Utility Theory" suggests that respondents opt out only when presented with less attractive alternatives, research indicates that decisions are influenced by motives beyond maximizing personal utility [35-38]. Hence, we adopted a dual response design: if respondents chose not to use any app with the presented features, they were then presented with the same choice task without the "no mobile app" alternative and forced to make a choice. This design mitigates potential power loss and minimizes opt-outs for reasons other than seeking the highest personal utility [34]. The survey design with the forced choice task is depicted in Figure S4 in Multimedia Appendix 1.

Pilot Surveys and Selection of the Experimental Design for the Final Survey

The pilot survey aimed to gather informative priors for the subsequent experimental design and to pretest the questionnaire for user-friendliness and clarity. Table S7 in Multimedia

XSL•FO
RenderX

Appendix 1 presents the Bayesian priors derived from the pilot survey.

The study focused on adults aged 18 years and older with CHD, as the mobile app was specifically designed for this target population. These criteria also accounted for differences in disease management strategies for younger patients younger than 18 years, making certain features of the app less applicable to them. Participants were recruited through PureProfile [39], an Australian based secure web application frequently used by research academics for web-based surveys that sources its members through diverse online and offline channels. These channels include internal referral programs, paid acquisition, social media, public relations, search engine marketing, offline marketing, and location-based registration [40].

In the absence of specific guidelines for pilot survey sample size estimations, we collected data from 32 participants based on our prior DCE project experience [41]. Pilot survey data were analyzed using the multinomial logit (MNL) model, recognized as the fundamental choice model [23]. We used Nlogit, a widely used commercial software for choice modeling [42], for the analyses. Detailed analysis methods and results are shown in Table S6 and Figure S5 in Multimedia Appendix 1.

Phase 3: Final Survey and Statistical Analysis

Determining the right sample size relies on factors such as question format, task complexity, result precision, population diversity, participant availability, and the need for subgroup analysis [33,43]. Orme [44] recommended a minimum sample size of 300. Marshall et al [45] observed that the average sample size for health care DCE published from 2005 to 2008 was 259, with nearly 40% ranging between 100 and 300 respondents. Another review studying methods of DCE studies conducted among primary health care professionals found a median sample size of 294 across 34 studies [46]. Accordingly, we opted for a sample size of 300, which also satisfied the minimum sample size estimated using efficiency parameters of the experimental design (highest Sb mean estimate × number of blocks) [25] except for 1 attribute level. Participants were recruited via PureProfile, adhering to eligibility criteria consistent with the pilot test, which included adults aged 18 years and older diagnosed with CHD. Data collection via the web-based survey was completed over a period of 21 days.

Main Effects Analysis

Our survey design, prompting respondents who initially selected the “neither” alternative to subsequently choose either app A or B, yielded 2 distinct datasets: unconditional data, capturing free choice, and conditional data, reflecting forced choices. In this study, we considered combining the conditional and unconditional datasets to be inappropriate due to observed differences in participant decision-making processes. This may contradict the principles of random utility theory essential to DCE [47]. Combining conditional and unconditional datasets also raises analytical concerns regarding the potential for biased parameter estimates [48], skewed demand modeling, and inconsistencies in reference data values. Moreover, in scenarios where opting out is a realistic market option and predicting uptake is critical, such as in our study, an unconditional demand model is recommended [47]. It is for these reasons that our main analysis was restricted to the unconditional data. However, results for conditional and combined data are available in Figures S6 and S7 in Multimedia Appendix 1.

In our primary analysis, we used the MNL, mixed multinomial logit model, and latent class model (LCM), assessing model fit using Log Likelihood Ratio and Akaike information criterion per observation to identify the most suitable model. We selected LCM for the main analysis as it had the best model fit indices [27]. Table S8 in Multimedia Appendix 1 presents the outcomes of this comparison. Latent class model assumes that parameter coefficients are distributed among individuals with a discrete distribution, leading to a finite number of classes, each with specific behavioral implications. While respondents are not deterministically assigned to any particular class, they display a probability, known as class assignment probability, of belonging to each class based on sociodemographic characteristics [27]. Subsequently, within each defined latent class, an MNL model was applied.

In the LCM analysis, we explored various model configurations with different numbers of classes (2 and 3) and sociodemographic covariates for class assignments. We initially selected essential sociodemographic covariates for integration into the model through consensus. Subsequently, their retention in the model was determined based on statistical significance [49]. The final model selection was ultimately guided by model fitness indices and logical coherence. The Nlogit code for the chosen LCM analysis is shown in Figure S8 in Multimedia Appendix 1.

The final LCM estimated the importance that participants placed on each attribute level compared with its reference level (part-worth) [30]. To understand the importance of each attribute on the total utility of the mobile app, we estimated relative importance of attributes [50], using parameter coefficients from the base MNL model. This calculation is shown in Table S9 and Figure S9 in Multimedia Appendix 1.

Scenario Analysis

Our scenario analysis explored the potential adoption of 3 mobile apps using the base MNL model. The basic app featured the most rudimentary levels for all attributes. The second scenario depicted a mobile app currently used to enhance cardiac rehabilitation in specific private cardiac clinics in Queensland, Australia [51]. This app necessitated basic training, offered facilities for vital sign recording, provided generalized health education, and allowed users to input information into the app as a symptom diary without any restrictions. The third scenario envisioned an advanced app capable of providing the highest levels for each attribute. Analyses were carried out using Nlogit version 6.0 [42], and results are presented as percentage changes from the base share for each scenario. Nlogit code for the scenario analyses is available in Figure S10 in Multimedia Appendix 1.

Ethical Considerations

This study was approved by the university human research ethics committee of Queensland University of Technology, Australia (reference no. 5732). Respondents were provided with

XSL•FO
RenderX

a participant information sheet within the web-based survey, and written consent was obtained prior to engagement with the choice tasks. Access to the survey was restricted to individuals who provided consent. Participants were monetarily compensated for their time in accordance with the terms and conditions of the survey platform, PureProfile [39]. Confidentiality of participants was maintained by anonymizing the data and presenting findings in an aggregated format.

Results

Participant Characteristics

Our sample of 302 participants had a mean age of 50.5 (SD 18.2) years, with a slight majority of males (169/302, 56.0%). Most participants (181/302, 59.9%) had CHD for more than 2 years, and 34.1% (103/302) used a mobile app for their condition. While 45.0% (136/302) expressed interest in future app use, 14% (42/302) did not. Geographically, participants were from all Australian states, with the highest representation coming from New South Wales (86/302, 28%) and Victoria (82/302, 27%), the 2 most populous states in Australia [52]. Median survey completion time was 6.8 (IQR 4.6-10.3) minutes. More sociodemographic details are shown in Table 2.

XSL•FO
RenderX

Table . Sociodemographic and clinical characteristics of participants (N=302).

Characteristics	Number	Frequency (%)
Age (years; mean 50.5, SD 18.2 years)		
18 - 24	16	5
25 - 34	61	20
35 - 44	58	19
45 - 54	36	12
55 - 64	42	14
65 - 74	54	18
>75	35	12
Sex		
Male	169	56
Female	133	44
Level of education		
High school not completed	21	7
High school completed	74	24
Undergraduate	132	44
Postgraduate	75	25
Employment		
Full-time employed	147	49
Part-time employed	53	17
Unemployed	7	2
Disability pension	14	5
Retired	77	25
Other	3	1
Prefer not to say	1	0.3
Annual gross income		
Less than US $12,500 (Aus $20,000)	18	6
US $20,000-US $28,210 (Aus $20,000-Aus $45,000)	71	23
US $28,210-US $37,614 (Aus $45,001-Aus $60,000)	43	14
US $37,614-US $56,421 (Aus $60,001-Aus $90,000)	67	22
US $56,421-US $75,228 (Aus $90,001-Aus $120,000)	44	15
US $75,228-US $94,035 (Aus $120,001-Aus $150,000)	38	13
More than US $94,035 (Aus $150,000)	21	7
State/territory		
Australian Capital Territory	8	3
New South Wales	86	28
Northern Territory	1	0.3
South Australia	22	7
Victoria	82	27
Queensland	65	21

XSL•FO
RenderX

Characteristics	Number	Frequency (%)
Western Australia	34	11
Tasmania	4	1
Type of heart disease		
Heart rhythm abnormality/pacemaker insertion	60	20
Ischemic heart disease/blocked arteries	71	23
Heart failure/heart weakness	40	13
Cardiomyopathy/heart muscle disease	46	15
Heart valve disease/valve replacement	29	10
Other	56	18
Duration since the diagnosis of heart disease (years)		
<1	46	15
1 - 2	75	25
>2	181	60
Previous use of a mobile health app		
I have used a mobile health app before, and I find it useful.	103	34
I have used a mobile health app before, and I did not find it useful.	21	7
I have not used a mobile health app before, but I would like to use one.	136	45
I have not used a mobile health app before, and I do not think I will use one in the near future.	42	14

Main Effects Analysis

All respondents completed a minimum of 8 choice tasks (unconditional data). Due to the nature of the survey, the total number of choice tasks completed varied among respondents. In total, there were 2803 choice observations from the 302 participants. Most respondents (172/302, 56.9%) answered only 8 choice tasks, indicating that they never selected the “neither” alternative throughout the survey. Conversely, 5.6% (17/302) of participants chose the ’neither’ alternative in all 8 primary choice tasks, resulting in 16 choice tasks for each. The distribution of “neither” alternative for the study sample is shown in Table S10 in Multimedia Appendix 1.

Table 3 presents the outcomes of the LCM, illustrating parameter estimates for the class assignment model and coefficients for each attribute level. Our results identified 2 latent classes within our study sample with 2 distinct preference behavior patterns. Conceptually, LCM operates assuming that preferences are shaped by both observable attributes and unobservable, or latent, heterogeneity [27]. This latent heterogeneity is presumed to represent distinct “preference groups” or “classes” within the sample, with individuals probabilistically assigned to these classes. The model delineated 2 latent classes based on four sociodemographic variables: (1) age, (2) level of education, (3) previous usage of mobile health apps, and (4) perception of the usefulness of mobile health apps. Participants with a higher number of “neither” selections did not exhibit a higher probability of belonging to any specific class identified by LCM analysis (Figure S11 in Multimedia Appendix 1).

Class 1 comprised a higher probability of 85.3% (257/302) for respondents. At the population level, older respondents with education beyond high school, prior experience with mobile health apps, and a positive perception of their utility are more likely to be classified into class 1 than class 2. This probability is predominantly influenced by the individual’s perception of the usefulness of a mobile app (β coefficient 2.9), followed by the level of education (β coefficients 1.2) and previous experience with mobile apps (β coefficients 1.2).

For instance, a 65-year-old individual living with CHD, possessing education beyond high school, prior experience with health apps, and perceiving them as useful, exhibits a 99.7% probability of belonging to class 1. In contrast, a 65-year-old patient with CHD with only high school education, no prior app experience, and a negative perception of app utility has a 60.3% probability of belonging to class 1. Conversely, a 32-year-old patient with CHD with high school education, no prior app experience, and a negative perception of app utility has a 63.9% probability of belonging to class 1. Detailed calculations of class-specific utility and probabilities are shown in Table S11 in Multimedia Appendix 1. On the contrary, membership in class 2 showed no discernible preference for either adopting an

XSL•FO
RenderX

app or abstaining from it (app A: −1.18, 95% uncertainty interval [UI] −2.36 to 0.006; app B: −0.78, 95% UI −1.99 to 0.42).

Parameters reported in Table 3 indicate the preferences at the population level, given the 2 classes identified. Respondents in class 1 preferred adopting a mobile app (β coefficient for app A 0.74, 95% UI 0.41-1.06; β coefficient for app B 0.53, 95% UI 0.22-0.85). For them, all attributes contributed positively to the utility of using a mobile app except for the training required before using the app. As anticipated, the preference for advanced training was lower (β coefficient −0.48, 95% UI −0.61 to −0.36) than basic training. Notably, the preference between having no training and undergoing basic training did not reach statistical significance. Respondents also preferred an app capable of providing recommendations on their next steps after monitoring vital signs (β coefficient 1.45, 95% UI 1.26-1.64). Even without recommendations, the ability to monitor blood pressure and heart rhythm retained a significant preference (β coefficient 1.07, 95% UI 0.88-1.26), surpassing the preference for an app that cannot monitor vital signs. These respondents also preferred to receive health education messages tailored to their individual needs (β coefficient 0.50, 95% UI 0.36-0.64), followed by receiving generalized health education messages (β coefficient 0.29, 95% UI 0.13-0.44) compared with not receiving health education messages at all. The ability to use the mobile app as a symptom diary without any restrictions (β coefficient 0.58, 95% UI 0.41-0.76) was preferred compared with limiting it to app-generated specific prompts (β coefficient 0.23, 95% UI 0.06-0.41) or not being able to use the app as a symptom diary. In contrast, individuals in class 2 exhibited no particular inclination toward any of the presented attribute levels.

XSL•FO
RenderX

Table . Results of the selected latent class model (log-likelihood function = −2050.644, Akaike information criterion/N=1.718).

	Class 1		Class 2	
	β Coefficient (95% UI[a])	SE	β Coefficient (95% UI)	SE
Class properties				
Class membership	85.3%	—[b]	14.7%	—
Constant	−1.53 (−3.36 to 0.31)	0.94	Reference	—
Age (years)	0.03[c] (0.002 to 0.06)	0.01	—	—
Education above high school	1.23[c] (0.27 to 2.19)	0.49	—	—
Having used a mobile health app before	1.16[c] (0.12 to 2.21)	0.53	—	—
Positive perception of usefulness of mobile health apps	2.92[d] (1.88 to 3.95)	0.53	—	—
Alternative specific constant				
Neither	Reference	—	Reference	—
Mobile app A	0.74[d] (0.41 to 1.06)	0.16	−1.18[e] (−2.36 to 0.006)	0.60
Mobile app B	0.53[d] (0.22 to 0.85)	0.16	−0.78 (−1.99 to 0.42)	0.61
Training				
No training	Reference	—	—	—
Basic training	0.008 (−0.13 to 0.14)	0.07	0.30 (−0.27 to 0.88)	0.29
Advanced training	−0.49[d] (−0.61 to −0.36)	0.06	−0.06 (−0.99 to 0.86)	0.47
Monitoring of vital signs				
No Monitoring	Reference	—	—	—
Monitor without recommendations	1.07[d] (0.88 to 1.26)	0.09	−0.08 (−1.15 to 0.98)	0.54
Monitor with recommendations	1.45[d] (1.26 to 1.64)	0.10	−0.91 (−2.00 to 0.18)	0.55
Health education				
No health education	Reference	—	—	—
Generalized health education	0.29[d] (0.14 to 0.44)	0.08	−0.27 (−1.16 to 0.61)	0.45
Individualized health education	0.50[d] (0.36 to 0.64)	0.07	−0.57 (−1.37 to 0.23)	0.41
Maintaining a symptom diary				
Not possible	Reference	—	—	—
Possible but restricted to app questions	0.23[d] (0.06 to 0.41)	0.09	−0.30 (−1.10 to 0.50)	0.41
Possible without any restrictions	0.58[d] (0.41 to 0.76)	0.09	−0.28 (−1.21 to 0.66)	0.48

[a]UI: uncertainty interval.

[b]Not applicable.

[c]Significance at 5% level.

[d]Significance at 1% level.

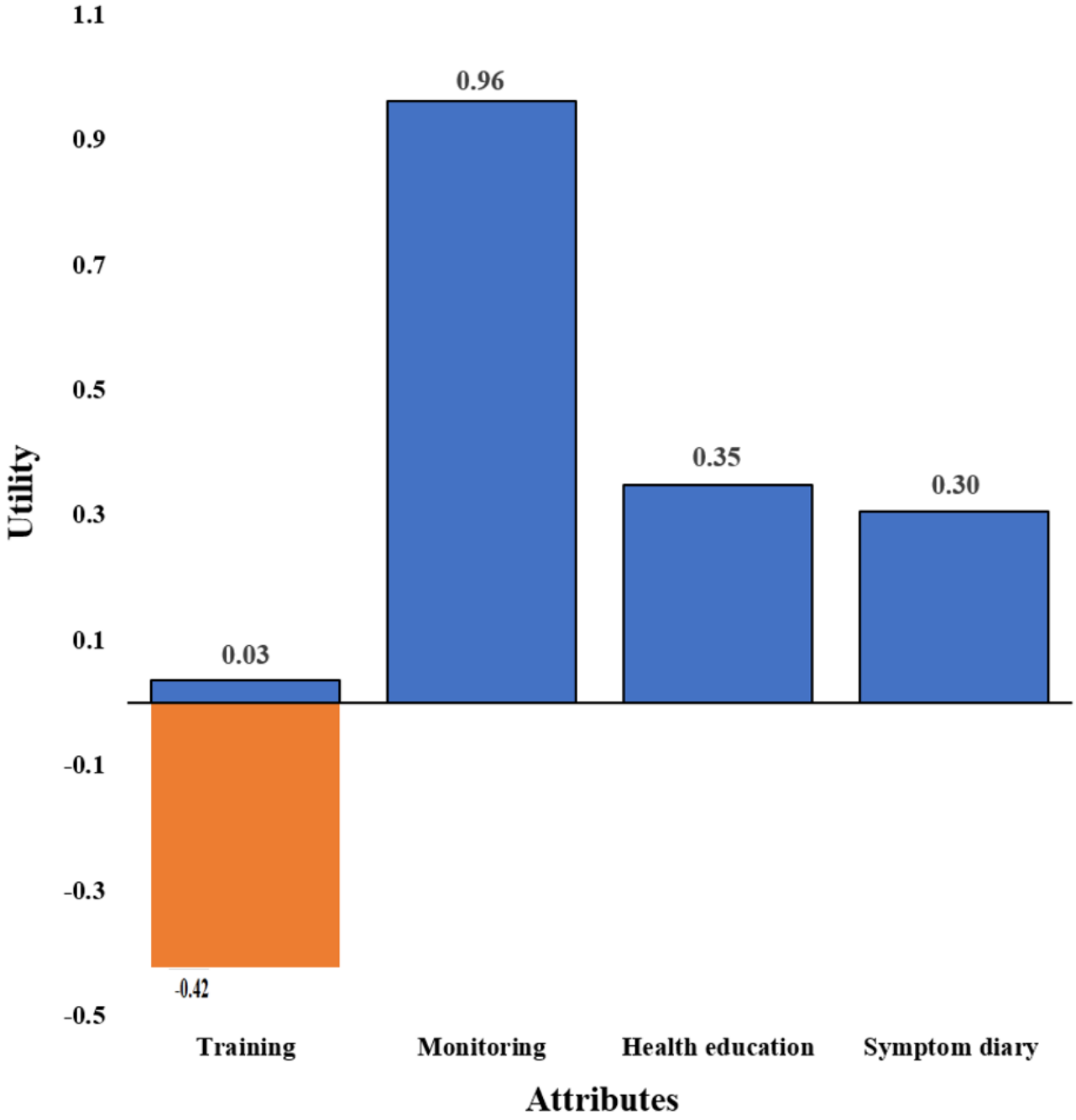
[e]Significance at 10% level.

To identify observable heterogeneity across different types of heart diseases, a subgroup analysis was conducted for each disease type. The analysis revealed that participants with most heart disease types, except heart failure and cardiomyopathy, generally disliked advanced training. In contrast, individuals with heart failure did not show a preference for facilities to

XSL•FO
RenderX

monitor vital signs. Preferences for other attributes varied across disease types, as detailed in Table S12 in Multimedia Appendix 1. However, these results should be interpreted with caution, as the study was not powered for post hoc subgroup analyses.

Figure 3. Utility ranges for attributes.



Relative Importance of Attributes

As shown in Figure 3, utility ranges were positive for all attributes except for training.

We estimated what app features were most valued by the survey participants by calculating the relative importance of attributes. Accordingly, the most influential feature affecting participants' decision to adopt a mobile app was its ability to monitor vital signs (relative importance of 46.4%). Participants next weighed the level of training required to navigate the app (relative importance of 22.1%). The delivery method of health education (relative importance of 16.8%) and app's ability to function as a symptom diary (relative importance of 14.7%) were considered less in their decision-making process. However, it is essential to note that the magnitude of relative importance is not directional. A higher relative importance does not necessarily indicate that respondents preferred it, but merely that they considered that attribute important. Calculation of relative importance of attributes is detailed in Table S9 in Multimedia Appendix 1.

Scenario Analysis

We assessed the uptake of 3 versions of a mobile health app via scenario analyses (Table 4). The probability of participants

XSL•FO
RenderX

adopting an app with basic features (scenario 1) was 84%. Upon upgrading the app features to those delineated in scenario 2, the adoption rate increased by 8.1%. However, with further enhancement of attributes to create an advanced app (scenario 3), the adoption rate increased only by 7.9%, which is a marginal drop compared with scenario 2. This slight decline could be due to the requirement for advanced training in the app in scenario 3. Training was estimated to hold the second-highest relative importance among all attributes, with advanced training being unfavorably viewed by respondents in class 1, who comprised a higher probability (85%) of the sample.

Table . Results of the scenario analysis.

	Scenario 1 (basic app)	Scenario 2 (resembles a current app in use)[a]	Scenario 3 (advanced app)
Training	No training required	Basic training required	Advanced training required
Monitoring	No monitoring of vital signs	Monitors vital signs but provides no recommendations	Monitors and provides recommendations
Health education	No health education	Generalized health education	Personalized health education
Maintaining a symptom diary	Cannot enter information	Can enter restricted information	Can enter information without any restrictions
Base share (%)	84	84	84
Scenario share (%)	84	92.1	91.9
Change of share from base to scenario (%)	N/A[b]	+8.1	+7.9

[a]A mobile app currently used to enhance cardiac rehabilitation in a specific private cardiac clinic in Queensland, Australia [51].

[b]N/A: not applicable.

Discussion

Principal Findings

Our LCM analysis uncovered 2 distinct latent groups among our survey respondents. Class 1 members, who are typically well-educated older adults with prior experience in app usage and a positive perception of app utility, expressed a preference for apps that are easy to navigate with minimal or no training. In addition, they favored features such as vital sign monitoring, feedback provision, personalized health education, and symptom diary functionality. On the other hand, class 2 members did not show a clear preference for adopting or rejecting mobile apps based on the attributes outlined in our survey; their preferences for attribute levels were indifferent. Among the presented app features, the ability to monitor vital signs and provide feedback primarily influenced the decision to adopt an app.

Comparison With Previous Work

As this was the first study investigating a population of individuals living with cardiac diseases, we were unable to compare our findings with studies from similar cohorts of patients. Our study found an encouraging 84% potential adoption rate for a mobile health app assisting individuals with CHD, even with basic features. Preferences for adopting mobile apps to assist with self-management have been demonstrated in various other health-related contexts, such as depression and anxiety [53], interventions for alcohol [54], diabetes mellitus [55], and smoking cessation [56].

The app feature of monitoring blood pressure and heart rhythm garnered strong preference among the majority, particularly when it provided recommendations. This attribute also exhibited the highest relative importance, indicating the significance individuals living with CHD attribute to monitoring their vital signs. Similar preferences for feedback and suggestions in mobile health apps have been reported in research on other chronic conditions such as HIV [57], cancer [58], and metabolic syndrome [59].

Most participants expressed indifference toward basic training versus no training, but they opposed mobile apps requiring advanced training, indicating a preference for apps that are user-friendly and straightforward to navigate. This finding was further confirmed by participants assigning significant importance to the level of training required, ranking it second highest in relative importance. Recent evidence on mobile health interventions for chronic diseases has underscored simplicity and ease of navigation as crucial factors in determining the effective use of the intervention [11,60]. In addition, ease of use has been recognized as a significant influence on the decision to purchase mobile health apps [61]. Majority of respondents also favored personalized health education over general information, aligning with previous studies emphasizing user preferences for personalization of app functionalities [56,59,62-65].

It is widely recognized that user characteristics have a great influence on sustained use of technologies designed for behavior change [66,67]. In addition to preference toward app features, our findings shed light on the influence of user characteristics on technology adoption. Class 1 membership, characterized by well-educated older adults with prior app experience and a positive perception of app utility, demonstrated a strong inclination toward adopting mobile apps for CHD management. A recent longitudinal study identified 4 dimensions that could influence the adoption and sustained use of mobile health apps, including the “user’s assessment of mobile health apps” [60]. Our findings on class assignment demographics support this conclusion, indicating that individuals with previous mobile

XSL•FO
RenderX

app experience and positive perceptions of app utility were more inclined to adopt the app. The identification of older adults with class 1, rather than younger individuals, was an unexpected finding, as young adults are generally presumed to be more receptive to and adopt digital health interventions [68,69].

Individuals more likely to be classified into class 2—often younger, with a high school education or below, lacking familiarity with mobile apps, and perceiving them as not useful—showed no clear preference for either adopting or abstaining from using an app based on the attributes presented in our survey. For this demographic, the features outlined in our study may not be pertinent or adequate to stimulate app adoption. Future research investigating populations with similar characteristics would be beneficial in elucidating the reasons for mobile app nonadoption, which may encompass specific barriers such as unfamiliarity with mobile technology or skepticism regarding its utility.

Limitations

While we aimed to enhance internal validity, our study has inherent limitations. It focused on 302 Australian patients with CHD, relying on self-reported data in a web-based survey. While our sample encompassed a diverse range of ages, types of CHD, and nearly equal representation of both males and females, our findings may not generalize well to dissimilar populations, including other diseases groups, cultural contexts or people with dissimilar digital literacy, or rates of smartphone ownership. Further research on varied populations could enrich the understanding of mobile health app acceptance and feature preferences.

Selecting attributes and specifying attribute levels are inherently subjective and may not comprehensively capture the spectrum of factors influencing individuals' preferences for mobile health apps. Despite our efforts to enhance this process by reviewing literature and consulting stakeholders for contextual insights, it is possible that certain attributes or levels important to patients may have been overlooked or insufficiently represented in our study design. Nevertheless, we believe that the attributes outlined in our study could still exert a substantial influence on real-world app adoption and usage. We recommend ensuring patient representation within the expert panel in future research endeavors.

Although we pretested the survey questionnaire to enhance understandability, biases linked to respondents' interpretation of choice tasks may persist, influencing their choice preferences. In addition, in practice, individuals may consider a broader array of factors and trade-offs, such as technological literacy, access to health care services, and socioeconomic status, when adopting mobile apps. In addition, some participants may not have considered all presented attributes in their decision-making process (attribute nonattendance). This may result in biased coefficient estimates and a skewed understanding of respondent preferences, an inherent bias in DCE [70,71].

Conclusions

The majority of respondents expressed a preference for adopting an app, even with basic features. Adoption rates were further boosted when app attributes included easy navigation, vital sign monitoring, feedback provision, personalized health education, and flexible data entry for symptom diary maintenance. However, it seems that these adoption rates may vary based on population demographics, with a minority showing reluctance to adopt apps with the features outlined in our study. Future research, encompassing both quantitative and qualitative approaches to explore the factors influencing app adoption among the demographics identified in our study, which are less receptive to mobile apps, is likely to contribute significantly to advancing this field.

Given that the majority of individuals living with CHD are inclined to adopt mobile health apps to manage their condition, we are optimistic that our findings will provide valuable insights in designing preference-informed mobile health apps. This, in turn, has the capacity to enhance adoption rates and promote sustained engagement with mobile apps among individuals living with CHD, thereby potentially contributing to improvements in clinical outcomes.

Acknowledgments

The authors wish to extend their sincere gratitude to Prof John Rose, Prof Michiel Bliemer, and Prof David Hensher of the University of Sydney for their invaluable insights and guidance in resolving their methodological and analytical queries. This study was conducted for the degree of doctor of philosophy of the first author, SH. She receives the postgraduate research award (international) from the Queensland University of Technology, Australia, and the Queensland Cardiovascular Research Network (QCVRN) Top Up Scholarship.

Data Availability

The dataset can be publicly accessed on the Queensland University of Technology website [72].

Authors' Contributions

SH was involved in study design, participant management, data analysis, data interpretation, in all phases of the study, and in manuscript preparation. SS was involved in the study design, data analysis, data interpretation in phases 2 and 3, and manuscript preparation. MA was involved in study design, participant management, data analysis and interpretation in phase 1, and manuscript preparation. SM was involved in data interpretation in phase 3 and manuscript preparation. WP was involved in phase 1 study design, data interpretation, and manuscript preparation. TW was involved in phase 1 data interpretation and manuscript preparation.

XSL•FO
RenderX

DB was involved in data interpretation in phase 3 and manuscript preparation. SK was involved in study design, participant management, data analysis, data interpretation, in all phases of the study, and in manuscript preparation. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for the review; long list of attributes; directional priors for the prepilot test; Ngene and Nlogit codes for prepilot, pilot, and final surveys; model efficiency parameters; relative importance of attribute calculations; and Nlogit code for scenario analysis and class-specific profiling calculations.
[PDF File, 850 KB - mhealth_v13i1e58556_app1.pdf]

References

1. Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol 2020 Dec 22;76(25):2982-3021. [doi: 10.1016/j.jacc.2020.11.010] [Medline: 33309175]
2. Zurynski Y, Ansell J, Ellis LA, et al. Accessible and affordable healthcare? Views of Australians with and without chronic conditions. Intern Med J 2021 Jul;51(7):1060-1067. [doi: 10.1111/imj.15172] [Medline: 33350562]
3. Free C, Phillips G, Galli L, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362. [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
4. Edwards ST, Dorr DA, Landon BE. Can personalized care planning improve primary care? JAMA 2017 Jul 4;318(1):25-26. [doi: 10.1001/jama.2017.6953] [Medline: 28594991]
5. Sullivan SS, Mistretta F, Casucci S, Hewner S. Integrating social context into comprehensive shared care plans: a scoping review. Nurs Outlook 2017;65(5):597-606. [doi: 10.1016/j.outlook.2017.01.014] [Medline: 28237357]
6. World Health Organization. MHealth: new horizons for health through mobile technologies. 2011. URL: https://iris.who.int/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2025-03-21]
7. Neves AL, Burgers J. Digital technologies in primary care: implications for patient care and future research. Eur J Gen Pract 2022 Dec;28(1):203-208. [doi: 10.1080/13814788.2022.2052041] [Medline: 35815445]
8. Chén OY, Roberts B. Personalized health care and public health in the digital age. Front Digit Health 2021;3:595704. [doi: 10.3389/fdgth.2021.595704] [Medline: 34713084]
9. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480. [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
10. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
11. Pfaeffli Dale L, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: a systematic review. Eur J Prev Cardiolog 2016 May;23(8):801-817. [doi: 10.1177/2047487315613462]
12. Gentili A, Failla G, Melnyk A, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022;10:787135. [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
13. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers. J Am Pharm Assoc (2003) 2013;53(2):172-181. [doi: 10.1331/JAPhA.2013.12202] [Medline: 23571625]
14. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120. [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
15. Ostermann J, Brown DS, de Bekker-Grob EW, Mühlbacher AC, Reed SD. Preferences for health interventions: improving uptake, adherence, and efficiency. Patient 2017 Aug;10(4):511-514. [doi: 10.1007/s40271-017-0251-y] [Medline: 28597375]
16. Lancsar E, Fiebig DG, Hole AR. Discrete choice experiments: a guide to model specification, estimation and software. Pharmacoeconomics 2017 Jul;35(7):697-716. [doi: 10.1007/s40273-017-0506-4] [Medline: 28374325]
17. Mangham LJ, Hanson K, McPake B. How to do (or not to do)... Designing a discrete choice experiment for application in a low-income country. Health Policy Plan 2009 Mar;24(2):151-158. [doi: 10.1093/heapol/czn047] [Medline: 19112071]
18. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 2008;26(8):661-677. [doi: 10.2165/00019053-200826080-00004] [Medline: 18620460]

XSL•FO
RenderX

19. McFadden D. The choice theory approach to market research. Mark Sci 1986 Nov;5(4):275-297. [doi: 10.1287/mksc.5.4.275]
20. De Brún A, Flynn D, Ternent L, et al. A novel design process for selection of attributes for inclusion in discrete choice experiments: case study exploring variation in clinical decision-making about thrombolysis in the treatment of acute ischaemic stroke. BMC Health Serv Res 2018 Jun 22;18(1):483. [doi: 10.1186/s12913-018-3305-5] [Medline: 29929523]
21. Louviere JJ, Hensher DA, JD S. Stated Choice Methods: Analysis and Application: Cambridge University Press; 2000.
22. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16(1):3-13. [doi: 10.1016/j.jval.2012.08.2223] [Medline: 23337210]
23. Hensher DA, Rose JM, Greene WH. Applied Choice Analysis, 2nd edition: Cambridge University Press; 2015.
24. Friedel JE, Foreman AM, Wirth O. An introduction to "discrete choice experiments" for behavior analysts. Behav Processes 2022 May;198:104628. [doi: 10.1016/j.beproc.2022.104628] [Medline: 35354088]
25. Rose JM, Bliemer MCJ. Sample size requirements for stated choice experiments. Transportation (Amst) 2013 Sep;40(5):1021-1041. [doi: 10.1007/s11116-013-9451-z]
26. Rose JM, Bliemer MCJ. Constructing efficient stated choice experimental designs. Transp Rev 2009 Sep;29(5):587-617. [doi: 10.1080/01441640902827623]
27. Hensher DA, Rose JM, Greene WH. Applied Choice Analysis: A Primer: Cambridge University Press; 2005.
28. Alamri AS, Georgiou S, Stylianou S. Discrete choice experiments: an overview on constructing D-optimal and near-optimal choice sets. Heliyon 2023 Jul;9(7):e18256. [doi: 10.1016/j.heliyon.2023.e18256] [Medline: 37539251]
29. Carlsson F, Martinsson P. Design techniques for stated preference methods in health economics. Health Econ 2003 Apr;12(4):281-294. [doi: 10.1002/hec.729] [Medline: 12652515]
30. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Health 2016 Jun;19(4):300-315. [doi: 10.1016/j.jval.2016.04.004] [Medline: 27325321]
31. Day B, Bateman IJ, Carson RT, et al. Ordering effects and choice set awareness in repeat-response stated preference studies. J Environ Econ Manage 2012 Jan;63(1):73-91. [doi: 10.1016/j.jeem.2011.09.001]
32. van Cranenburgh S, Collins AT. New software tools for creating stated choice experimental designs efficient for regret minimisation and utility maximisation decision rules. J Choice Model 2019 Jun;31:104-123. [doi: 10.1016/j.jocm.2019.04.002]
33. Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011 Jun;14(4):403-413. [doi: 10.1016/j.jval.2010.11.013] [Medline: 21669364]
34. Veldwijk J, Lambooij MS, de Bekker-Grob EW, Smit HA, de Wit GA. The effect of including an opt-out option in discrete choice experiments. PLoS One 2014;9(11):e111805. [doi: 10.1371/journal.pone.0111805] [Medline: 25365169]
35. Ritov I, Baron J. Status-quo and omission biases. J Risk Uncertainty 1992 Feb;5(1):49-61. [doi: 10.1007/BF00208786]
36. Boxall P, Adamowicz W(, Moon A. Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement*. Aus J Agric Resour Econ 2009 Oct;53(4):503-519. [doi: 10.1111/j.1467-8489.2009.00469.x]
37. Luce MF, Payne JW, Bettman JR. Emotional trade-off difficulty and choice. J Mark Res 1999 May;36(2):143-159. [doi: 10.1177/002224379903600201]
38. Dhar R. Consumer preference for a no‐choice option. J Consum Res 1997 Sep;24(2):215-231. [doi: 10.1086/209506]
39. PureProfile. URL: https://www.pureprofile.com/home/ [accessed 2024-06-06]
40. Pureprofile. ESOMAR 37 questions to help researcher buyers. 2023. URL: https://business.pureprofile.com/esomar/ [accessed 2025-04-09]
41. Kularatna S, Allen M, Hettiarachchi RM, et al. Cancer survivor preferences for models of breast cancer follow-up care: selecting attributes for inclusion in a discrete choice experiment. Patient 2023 Jul;16(4):371-383. [doi: 10.1007/s40271-023-00631-0] [Medline: 37213062]
42. Green WH. NLOGIT Reference Guide: Version 50: Econometric Software, Inc; 2002.
43. Louviere JJ, Hensher DA, Swait JD. Combining sources of preference data. In: Stated Choice Methods: Analysis and Applications: Cambridge University Press; 2000:227-251.
44. Orme BK. Getting Started With Conjoint Analysis: Strategies for Product Design and Pricing Research: Research Publishers; 2010.
45. Marshall D, Bridges JFP, Hauber B, et al. Conjoint analysis applications in health—how are studies being designed and reported?: an update on current practice in the published literature between 2005 and 2008. Patient 2010 Dec 1;3(4):249-256. [doi: 10.2165/11539650-000000000-00000] [Medline: 22273432]
46. Merlo G, van Driel M, Hall L. Systematic review and validity assessment of methods used in discrete choice experiments of primary healthcare professionals. Health Econ Rev 2020 Dec 9;10(1):39. [doi: 10.1186/s13561-020-00295-8] [Medline: 33296066]
47. Whitty JA, Lancsar E, De Abreu Lourenco R, Howard K, Stolk EA. Putting the choice in choice tasks: incorporating preference elicitation tasks in health preference research. Patient 2024 May 14. [doi: 10.1007/s40271-024-00696-5] [Medline: 38744798]

XSL•FO
RenderX

48. Ryan M, Skåtun D. Modelling non-demanders in choice experiments. Health Econ 2004 Apr;13(4):397-402. [doi: 10.1002/hec.821] [Medline: 15067675]
49. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis: Springer; 2015.
50. Gonzalez JM. A guide to measuring and interpreting attribute importance. Patient 2019 Jun;12(3):287-295. [doi: 10.1007/s40271-019-00360-3] [Medline: 30906968]
51. Cardihab Pvt Ltd. Digital cardiac rehabilitation. URL: https://cardihab.com/ [accessed 2024-02-20]
52. Australian Bureau of Statistics. National, state and territory population. 2024 Sep. URL: https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/sep-2024 [accessed 2025-03-31]
53. Lipschitz J, Miller CJ, Hogan TP, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334. [doi: 10.2196/11334] [Medline: 30681968]
54. Borges A, Caviness C, Abrantes AM, et al. User-centered preferences for a gait-informed alcohol intoxication app. Mhealth 2023;9:6. [doi: 10.21037/mhealth-21-55] [Medline: 36760789]
55. Conway N, Campbell I, Forbes P, Cunningham S, Wake D. mHealth applications for diabetes: user preference and implications for app development. Health Informatics J 2016 Dec;22(4):1111-1120. [doi: 10.1177/1460458215616265] [Medline: 26635324]
56. McClure JB, Hartzler AL, Catz SL. Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers. JMIR Mhealth Uhealth 2016 Feb 12;4(1):e17. [doi: 10.2196/mhealth.5181] [Medline: 26872940]
57. Ramanathan N, Swendeman D, Comulada WS, Estrin D, Rotheram-Borus MJ. Identifying preferences for mobile health applications for self-monitoring and self-management: focus group findings from HIV-positive persons and young mothers. Int J Med Inform 2013 Apr;82(4):e38-e46. [doi: 10.1016/j.ijmedinf.2012.05.009] [Medline: 22704234]
58. Cho Y, Zhang H, Harris MR, Gong Y, Smith EL, Jiang Y. Acceptance and use of home-based electronic symptom self-reporting systems in patients with cancer: systematic review. J Med Internet Res 2021 Mar 12;23(3):e24638. [doi: 10.2196/24638] [Medline: 33709929]
59. Joshi A, Amadi C, Schumer H, Galitzdorfer L, Gaba A. A human centered approach to design a diet app for patients with metabolic syndrome. Mhealth 2019;5:43. [doi: 10.21037/mhealth.2019.08.13] [Medline: 31620470]
60. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983. [doi: 10.2196/12983] [Medline: 31469081]
61. Xie Z, Or CK. Consumers' preferences for purchasing mHealth apps: discrete choice experiment. JMIR Mhealth Uhealth 2023 Sep 13;11:e25908. [doi: 10.2196/25908] [Medline: 37707310]
62. Nicol GE, Ricchio AR, Metts CL, et al. A smartphone-based technique to detect dynamic user preferences for tailoring behavioral interventions: observational utility study of ecological daily needs assessment. JMIR Mhealth Uhealth 2020 Nov 13;8(11):e18609. [doi: 10.2196/18609] [Medline: 33055063]
63. Houwen T, Vugts MAP, Lansink KWW, et al. Developing mHealth to the context and valuation of injured patients and professionals in hospital trauma care: qualitative and quantitative formative evaluations. JMIR Hum Factors 2022 Jun 20;9(2):e35342. [doi: 10.2196/35342] [Medline: 35723928]
64. Nittas V, Mütsch M, Braun J, Puhan MA. Self-monitoring app preferences for sun protection: discrete choice experiment survey analysis. J Med Internet Res 2020 Nov 27;22(11):e18889. [doi: 10.2196/18889] [Medline: 33245282]
65. Olsen SH, Saperstein SL, Gold RS. Content and feature preferences for a physical activity app for adults with physical disabilities: focus group study. JMIR Mhealth Uhealth 2019 Oct 11;7(10):e15019. [doi: 10.2196/15019] [Medline: 31605518]
66. Hekler EB, Klasnja P, Froehlich JE, Buman MP. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 27, 2013; Paris, France URL: https://doi.org/10.1145/2470654.2466452 [accessed 2025-04-09] [doi: 10.1145/2470654.2466452]
67. Mlekus L, Bentler D, Paruzel A, Kato-Beiderwieden AL, Maier GW. How to raise technology acceptance: user experience characteristics as technology-inherent determinants. Gr Interakt Org 2020 Sep;51(3):273-283. [doi: 10.1007/s11612-020-00529-7]
68. Ferretti A, Hubbs S, Vayena E. Global youth perspectives on digital health promotion: a scoping review. BMC Digit Health 2023;1(1):25. [doi: 10.1186/s44247-023-00025-0]
69. Malloy J, Partridge SR, Kemper JA, Braakhuis A, Roy R. Co-design of digital health interventions with young people: a scoping review. Digit Health 2023;9:20552076231219117. [doi: 10.1177/20552076231219117] [Medline: 38107986]
70. Heidenreich S, Watson V, Ryan M, Phimister E. Decision heuristic or preference? Attribute non-attendance in discrete choice problems. Health Econ 2018 Jan;27(1):157-171. [doi: 10.1002/hec.3524] [Medline: 28620975]
71. Nguyen TC, Robinson J, Whitty JA, Kaneko S, Nguyen TC. Attribute non-attendance in discrete choice experiments: a case study in a developing country. Econ Anal Policy 2015 Sep;47:22-33. [doi: 10.1016/j.eap.2015.06.002]

XSL•FO
RenderX

72. Integrating a mobile application to enhance atrial fibrillation care: key insights from an implementation study guided by the consolidated framework for implementation research (CFIR). Queensland University of Technology. URL: https://doi.org/10.25912/RDF_1743047027785 [accessed 2025-04-09]

Abbreviations

CHD: chronic heart disease
DCE: discrete choice experiment
DIRECT: Discrete Choice Experiment Reporting Checklist
LCM: latent class model
MNL: multinomial logit
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UI: uncertainty interval

Edited by L Buis; submitted 19.03.24; peer-reviewed by F Montano-Campos, JV Johansson; revised version received 25.11.24; accepted 26.02.25; published 25.04.25.

Please cite as:
Hewage SA, Senanayake S, Brain D, Allen MJ, McPhail SM, Parsonage W, Walters T, Kularatna S
Preferences for Mobile App Features to Support People Living With Chronic Heart Diseases: Discrete Choice Study
JMIR Mhealth Uhealth 2025;13:e58556
URL: https://mhealth.jmir.org/2025/1/e58556
doi:10.2196/58556

© Sumudu Avanthi Hewage, Sameera Senanayake, David Brain, Michelle Allen, Steven M McPhail, William Parsonage, Tomos Walters, Sanjeewa Kularatna. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

mHealth-Augmented Care for Reducing Depression Symptom Severity Among Patients With Chronic Pain: Exploratory, Retrospective Cohort Study

Dan Holley[1], PhD; Amanda Brooks[1], AM, LCSW; Matthew Hartz[2], BS; Sudhir Rao[3], MD; Thomas Zaubler[1], MD

[1]Clinical Operations, NeuroFlow, Inc, 1601 Market St, Suite 1500, Philadelphia, PA, United States

[2]Data and Engineering, NeuroFlow, Inc, Philadelphia, PA, United States

[3]Pain and Spine Specialists, Mt. Airy, MD, United States

Corresponding Author:
Dan Holley, PhD
Clinical Operations, NeuroFlow, Inc, 1601 Market St, Suite 1500, Philadelphia, PA, United States

Abstract

Background: Depression and chronic pain are commonly comorbid, mutually reinforcing, and debilitating. Emerging approaches to mobile behavioral health care (mHealth) promise to improve outcomes for patients with comorbid depression and chronic pain by integrating with existing care models to bolster support and continuity between clinical visits; however, the evidence base supporting the use of mHealth to augment care for this patient population is limited.

Objective: To develop an evidence base that sets the stage for future research, we aimed to explore the associations between changes in depression severity and various integrated care models, with and without mHealth augmentation, among patients with comorbid depression and nonmalignant chronic pain.

Methods: Our team leveraged retrospective, real-world data from 3837 patients with comorbid depression and nonmalignant chronic pain who received integrated behavioral health care (IBH) at a subspecialty pain clinic. We analyzed one IBH-only, non-mHealth cohort (n=2765), an mHealth-augmented cohort (n=844), and a collaborative care (CoCM)+mHealth cohort (n=136), which were supported by the NeuroFlow mHealth platform, and a pre-CoCM mHealth cohort (n=92), which was supported by the mHealth platform for 3 months prior to beginning the chronic pain treatment. We evaluated changes in depression severity between treatment cohorts via longitudinal analyses of both clinician- and mHealth-administered Patient Health Questionnaire-9 (PHQ-9) assessments.

Results: mHealth-augmented integrated care led to significantly greater proportions of patients reaching clinical benchmarks for reduction (725/844, 86% vs 2112/2765, 76%), response (689/844, 82% vs 2027/2765, 73%), and remission (629/844, 75% vs 1919/2765, 69%) compared with integrated care alone. Furthermore, hierarchical regression modeling revealed that patients who received mHealth-augmented psychiatric CoCM experienced the greatest sustained reductions in on-average depression severity compared with other cohorts, irrespective of clinical benchmarks. In addition, patients who engaged with an mHealth platform before entering CoCM experienced a 7.2% reduction in average depression severity before starting CoCM treatment.

Conclusions: Our findings suggest that mHealth platforms have the potential to improve treatment outcomes for patients with comorbid chronic pain and depression by providing remote measurement–based care, tailored interventions, and improved continuity between appointments. Moreover, our study set the stage for further research, including randomized controlled trials to evaluate causal relationships between mHealth engagement and treatment outcomes in integrated care settings.

(JMIR Mhealth Uhealth 2025;13:e52764) doi:10.2196/52764

KEYWORDS

mHealth; mobile health; app; behavioral health care; depression; mental health; screening; pain; chronic pain; psychiatric screenings; digital health care; psychiatry; psychiatric; longitudinal; assessment; behavioral; self-help; integrated; comorbidity; augmented care

Introduction

America's psychiatric crisis impairs our health, happiness, and productivity on an industrial scale. Half of all Americans will experience a behavioral health condition at some point in life, and 1 in 5 will experience a condition so severe that it is debilitating [1,2]. Despite the ubiquitous nature of these conditions and their sprawling impacts, psychiatric care is underused, and many individuals resist treatment up to the point of emergency crises [3,4].

XSL•FO
RenderX

Exacerbating this, behavioral health conditions are frequently accompanied by physical health complications. For instance, depression and chronic pain are commonly comorbid [5-10] and feature prominently atop leading causes of disability worldwide [1]. Along with 23% to 37% higher annual health care costs [11], individuals who live with chronic pain experience increased risk of suicidality [12,13], reduced quality of life [14-16], poor occupational fulfillment [17], and disrupted family dynamics [18,19]. While chronic pain's causal contributions to depression are intuitive and well-established, emerging neuroscientific research indicates that this relationship is bidirectional: depression can sensitize the body to pain via multiple mechanisms across a shared neural substrate [20-22]. Thus, treating either depression or chronic pain alone may be less effective than treating both conditions in an integrated care setting. Despite the biobehavioral complexities of comorbid depression and pain, treatments are often siloed, which contributes to second-order crises. The opioid epidemic, for example, is enabled by a longstanding reliance on pharmacological treatment for chronic pain [23-25].

Traditional treatment models that focus on either behavioral or physical health conditions in isolation often fail to address the complex interplay between comorbid conditions like depression and chronic pain. In the absence of integrated care approaches, patients face fragmented care pathways, which can exacerbate symptoms, lead to suboptimal outcomes, and contribute to treatment resistance [5,23,26,27]. Without mHealth augmentation, health care providers are limited in their ability to monitor patients between visits, track real-time changes in health status, and intervene before crises occur [28-31]. These challenges are particularly pronounced in patients with chronic pain and comorbid depression, who may be at higher risk of falling through the gaps of episodic care models.

Integrated treatment models, notably psychiatric collaborative care (CoCM), have emerged as sophisticated approaches in the treatment of complex patients who present with behavioral and physical health comorbidities [26,27,32]. Under these models, multifaceted care teams provide integrated, population-based, patient-centered behavioral and physical health care. In CoCM, for example, medical providers, behavioral health care managers, caseload supervising psychiatrists, and patients share records and work together as one team to establish and achieve goals, thereby enabling the coordinated treatment of comorbid behavioral and physical health issues, such as depression and chronic pain. While integrated approaches are not widely used, an extensive evidence base attests to their effectiveness and efficiency [26,27,33,34].

Nevertheless, questions remain about whether and how emerging technologies, such as mobile behavioral health care (mHealth) platforms [3,35,36], can augment and scale integrated care models. Despite the surge in the implementation and broadening of mHealth during the pandemic, the utilization of such technologies is still not at its fullest. Stand-alone self-management tools, such as mental health apps, offer limited proven benefits. However, technology-driven solutions designed to complement and expand the mental health workforce hold boundless promise [28]. mHealth is an especially attractive facet of integrated care that promises to improve personalized treatment and enhance patient continuity via remote measurement–based care and tailored, real-time digital interventions. Critically, mHealth platforms can provide scalable, location-agnostic support that reaches patients where they are, when they need it, thereby helping to overcome barriers, such as provider scarcity and treatment stigma [29-31]. They also help to elevate and standardize the quality of behavioral care by providing clinical decision support for providers and digital self-management tools for patients, all of which facilitate compliance with evidence-based care.

While much has been written about the advantages of mHealth [28-31,35,36], our study is among the first to empirically investigate its use among integrated behavioral health care (IBH) patients with comorbid chronic pain and depression, in a real-world setting. To better understand the relationships between mHealth-augmented treatment approaches and patient outcomes in cases of comorbid behavioral and physical health conditions, we conducted an exploratory, retrospective cohort study of behavioral health outcomes in 3837 patients who received IBH at a subspecialty pain clinic. We analyzed data collected during routine health care operations to evaluate changes in depressive symptomatology over time among cohorts defined by the type of IBH administered and the availability and timing of mHealth support. Our objective was to evaluate the associations between mHealth support and behavioral health outcomes among patients with chronic pain, and to provide new insights into the potential for mHealth to augment IBH.

Methods

Study Design

We used a retrospective-cohort design to explore the associations between treatment approaches and changes in depression severity among 3837 patients aged 18 years and over who showed signs of clinical depression and received treatment for chronic nonmalignant pain at a subspecialty pain clinic between April 2019 and April 2023.

Patient Cohorts

To understand the links between different treatment approaches and changes in depression severity, we organized our retrospective patient population into 4 distinct treatment cohorts, further described in Table 1. Because this was a retrospective study of real-world data, patients were not assigned to cohorts, and we have no insight into the process by which patients arrived in their cohorts.

XSL•FO
RenderX

Table . Treatment cohorts.

	Cohort				
	IBH[a] (n=2765)	IBH+mHealth[b] (n=844)	CoCM[c]+mHealth (n=136)	Pre-CoCM mHealth (n=92)	Total: all cohorts (N=3837)
Sex, n (%)					
Female	1824 (66)	610 (72.3)	102 (75)	67 (72.8)	2603 (67.9)
Male	940 (34)	234 (27.7)	34 (25)	25 (27.2)	1233 (32.1)
Age (years), mean (SD)	55.95 (13.61)	55.60 (13.81)	60.25 (13.06)	50.39 (13.47)	55.71 (13.73)
Treatment description	Patients were regularly evaluated for symptoms of depression using the PHQ-9[d] during scheduled office visits. Elevated PHQ-9 assessments were reviewed by either an MD or BH[e] specialist. All patients were provided with stress reduction techniques, including meditation, mindfulness, and deep breathing exercises. Stress-reduction interventions were provided by either an MD or on-site BH specialist.	Patients received IBH as described above, augmented by use of the mHealth app.	Patients received highly integrated medical and BH care under the CoCM, with access to the mHealth app at the start of CoCM care.	A unique group of CoCM patients who had at least 3 months of advance access to the mHealth app prior to the start of CoCM treatment, and who completed at least one mHealth activity per month in those 3 months.	—[f]
Notes	—	—	Because all CoCM patients' treatment was augmented by mHealth, our study design lacks a CoCM-without-mHealth group for comparison.	—	Between-group age differences were not statistically significant (one-way ANOVA: $F_{3,3833}$=1.23; P=.30).

[a]IBH: integrated behavioral health care.

[b]mHealth: mobile behavioral health care

[c]CoCM: collaborative care.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]BH: behavioral health.

[f]Not available.

Measuring Depression Severity

To measure changes in depression severity over time, we evaluated real-world data from clinic- and mobile-administered Patient Health Questionnaire-9 (PHQ-9) assessments [37]. The PHQ-9 is a well-validated and reliable measure for the detection and severity stratification of depressive disorders and is regarded as a suitable tool for clinical depression research. Its 9 questions address the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision* components of major depressive disorder on a 0 ("not at all") to 3 ("nearly every day") scale. Scores range from 0 to 27, with the following ranges commonly regarded as qualitative differences in depression severity: 0 - 4, not depressed; 5 - 9, mild depression; 10 - 14, moderate depression; 15 - 19, moderate-to-severe depression; and 20 - 27, severe depression.

Mobile Behavioral Health Platform

Our study used the NeuroFlow digital behavioral health platform (NeuroFlow, Inc.) to document, monitor, and share updates on patients' emotional state, sleep patterns, and stress levels, and to perform routine assessments, including the PHQ-9. In addition, the app offers informative videos tailored to users' self-reported scores and evolving symptoms. Health care providers use the mHealth platform to enhance continuity between visits and track patient-reported outcomes to garner unique insights into the patients' physical and behavioral well-being.

XSL•FO
RenderX

Statistical and Computational Methods

To compare initial depression severity of our cohorts, we first established a clinical index for each patient based on PHQ-9 scores. Consistent with Kroenke et al [37], we defined our clinical index as the first occurrence of a PHQ-9 score of ≥10, which indicates moderate (or worse) depression. We then used a one-way ANOVA and subsequent post hoc tests to evaluate our cohorts for mean differences in initial severity. We also computed a one-way ANOVA to test for significant differences in our cohorts' mean ages at the outset of the study.

We used a series of chi-square tests to compare the IBH and IBH+mHealth cohorts based on the proportion of patients reaching established clinical benchmarks for reduction (ie, a ≥5-point drop in PHQ-9 score recorded any time after clinical indexing), response (ie, a ≥50% drop following clinical indexing), and remission (ie, a score of <5 following clinical indexing). Furthermore, we computed a series of independent-samples *t* tests (2-tailed) to explore differences in the speeds with which each cohort reached these benchmarks.

In addition, we leveraged hierarchical (mixed-effects) regression modeling to examine the relationships between different treatment approaches and the temporal dynamics of depression severity from the time of clinical indexing onward, irrespective of clinical benchmarks. This model accounts for repeated measurements of PHQ-9 scores over time within each patient, treating patients as random effects and treatment cohorts as fixed effects. We calculated the slope and standard error of evolving PHQ-9 scores for each patient and incorporated time as a continuous variable to assess changes in depression severity over time. The model included the fixed effect of cohort, with the IBH, IBH+mHealth, and CoCM+mHealth cohorts treated as categorical variables using dummy coding. We included patient-specific random intercepts and slopes to account for individual variability in depression severity trajectories.

We instantiated this model using the statsmodels library's *mixedlm()* function in Python 3.8.3. The analysis was implemented as a mixed-effects linear regression with the formula PHQ_SCORE ~ TIME + C(COHORT), where PHQ_SCORE was the dependent variable, TIME represented the temporal evolution of scores, and C(COHORT) specified the treatment cohorts as categorical predictors. We also included the interaction term between time and cohort to evaluate whether changes in PHQ-9 scores over time differed across cohorts. Random effects for both the intercept and the slope of time were included to account for the nested structure of the data, with repeated measures for each patient. This approach allowed us to account for both within- and between-subject variance. The random-effects structure assumed a normal distribution for both random intercepts and slopes, and we fit the model using the Restricted Maximum Likelihood estimation method, ensuring convergence by using the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm.

Other model assumptions included:

1. Random effects for patients: we assumed that individual patients had their own baseline PHQ-9 scores and their own trajectories of change over time.
2. Fixed effects for treatment cohorts: the treatment cohorts were treated as fixed, meaning we expected consistent effects of each cohort across patients.
3. Normality of residuals: we assumed that the residuals of the model followed a normal distribution, and that the relationship between time and PHQ-9 scores was linear.
4. Time as a continuous variable: we assumed that changes in depression severity occurred continuously over time, rather than at specific time points.

Finally, we performed a paired-samples *t* test (2-tailed) to compare mean differences in depression severity among pre-CoCM mHealth patients at two time points: PHQ-9–derived clinical index scores recorded before entering CoCM, and PHQ-9 scores associated with the start of CoCM treatment.

Because our analyses were exploratory, all hypothesis tests were 2-tailed and thresholded at $P<.05$. Results reported here are uncorrected for multiple comparisons. All statistical tests were performed in Python v3.8.3 (Python Software Foundation) using standard statistical libraries (eg, scipy.stats v1.10.1) and functions (eg, *stats.ttest_rel()* for paired-sample *t* tests, *stats.f_oneway()* for one-way ANOVA, etc).

Organizations Involved

Pain and Spine Specialists is a pain subspecialty clinic with locations in Maryland, Pennsylvania, and Virginia. Pain and Spine Specialists' network of board-certified clinicians provides chronic pain management and interventional pain treatment in integrated care settings. NeuroFlow, Inc., is a private, for-profit digital behavioral health company headquartered in Philadelphia, Pennsylvania. Its mHealth platform is a Health Insurance Portability and Accountability Act (HIPAA)–compliant, US Food and Drug Administration (FDA)–exempt, cloud-based tool that facilitates behavioral health care access, engagement, and remote measurement–based care.

Ethical Considerations

All patient data were collected as a facet of routine health care operations and were anonymized prior to analysis. Our study was exempted from institutional review board oversight by Advarra's Center for IRB Intelligence (protocol number 00069060, dated January 23, 2023), which covered both the collection of data and the waiver of informed consent for our analyses. Data are stored on access-controlled, firewall-protected servers within NeuroFlow, Inc.

Results

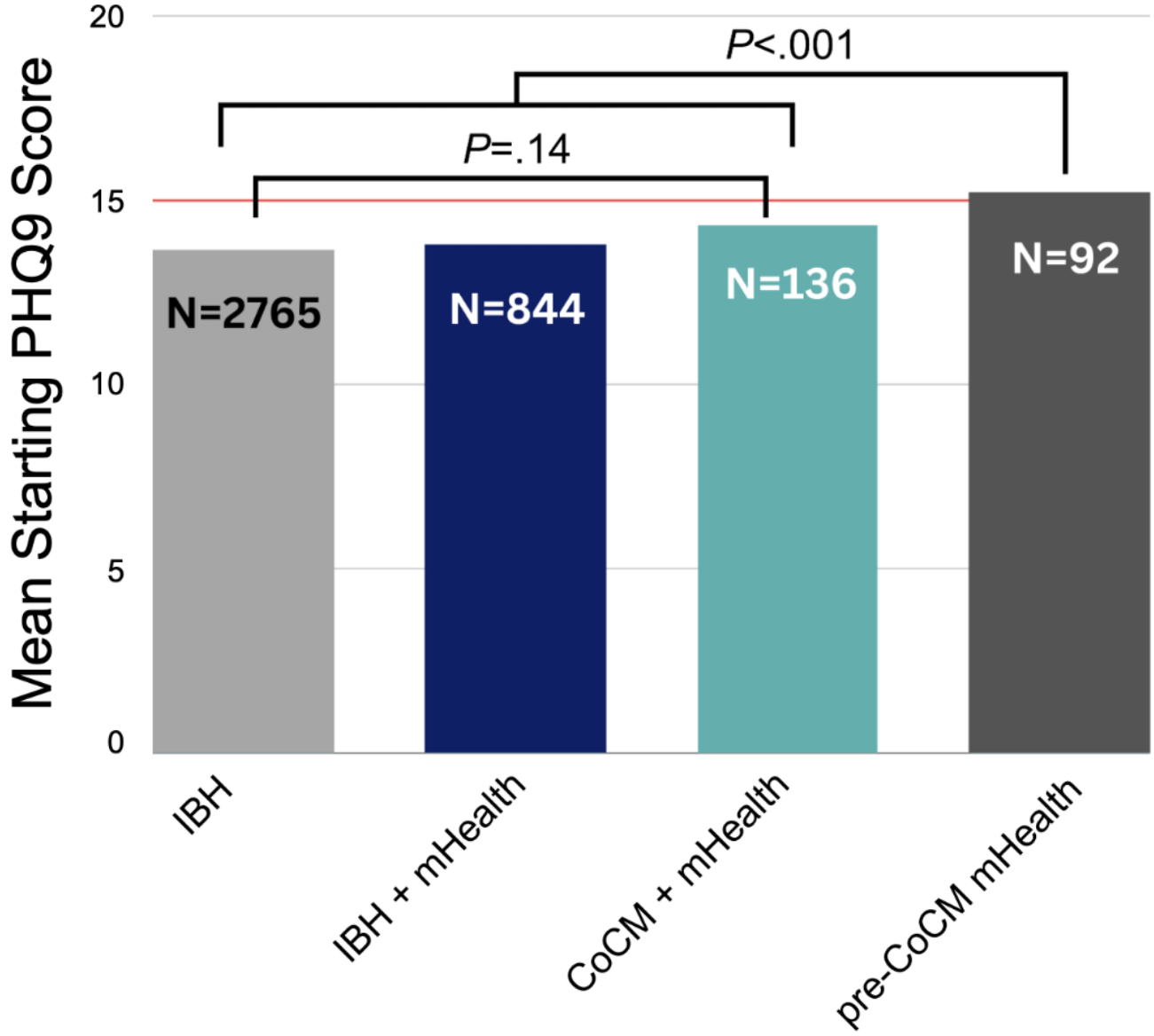
Comparing Initial Depression Symptoms Between Cohorts

As a first step in evaluating our cohorts, we compared each group's mean clinical indices to determine whether starting depression severity significantly differed across all groups. Although we expected no significant differences by cohort, that was not always the case (Figure 1). We found the average starting severity for the IBH (13.64, SD 3.91), IBH+mHealth (13.78, SD 3.70), and CoCM+mHealth (14.32, SD 3.69) cohorts to be both categorically (ie, "moderate depression") and statistically similar (one-way ANOVA, $F_{2,3742}=1.97$; $P=.14$).

XSL•FO
RenderX

However, average starting severity for the pre-CoCM+mHealth cohort (15.20, SD 4.16) was both categorically (moderate-to-severe depression) and statistically higher than the other cohorts ($F_{3,3833}$=5.77; P<.001). Given the significant difference in baseline severity, the variability in pretreatment mHealth engagement duration, and the smaller sample size within this cohort, we have excluded it from subsequent between-group analyses to avoid confounding the results.

Figure 1. Measuring mean starting depression severity revealed no significant differences by cohort, except for significantly elevated initial depression severity in the pre-CoCM mHealth cohort. The red line illustrates the categorical cutoff for moderate-to-severe depression. This finding informed our downstream analyses by checking against a priori differences between cohorts and resulted in our exclusion of the pre-CoCM mHealth cohort from subsequent groupwise analyses. CoCM: collaborative care; IBH: integrated behavioral health care; mHealth: mobile behavioral health care; PHQ-9: Patient Health Questionnaire-9.



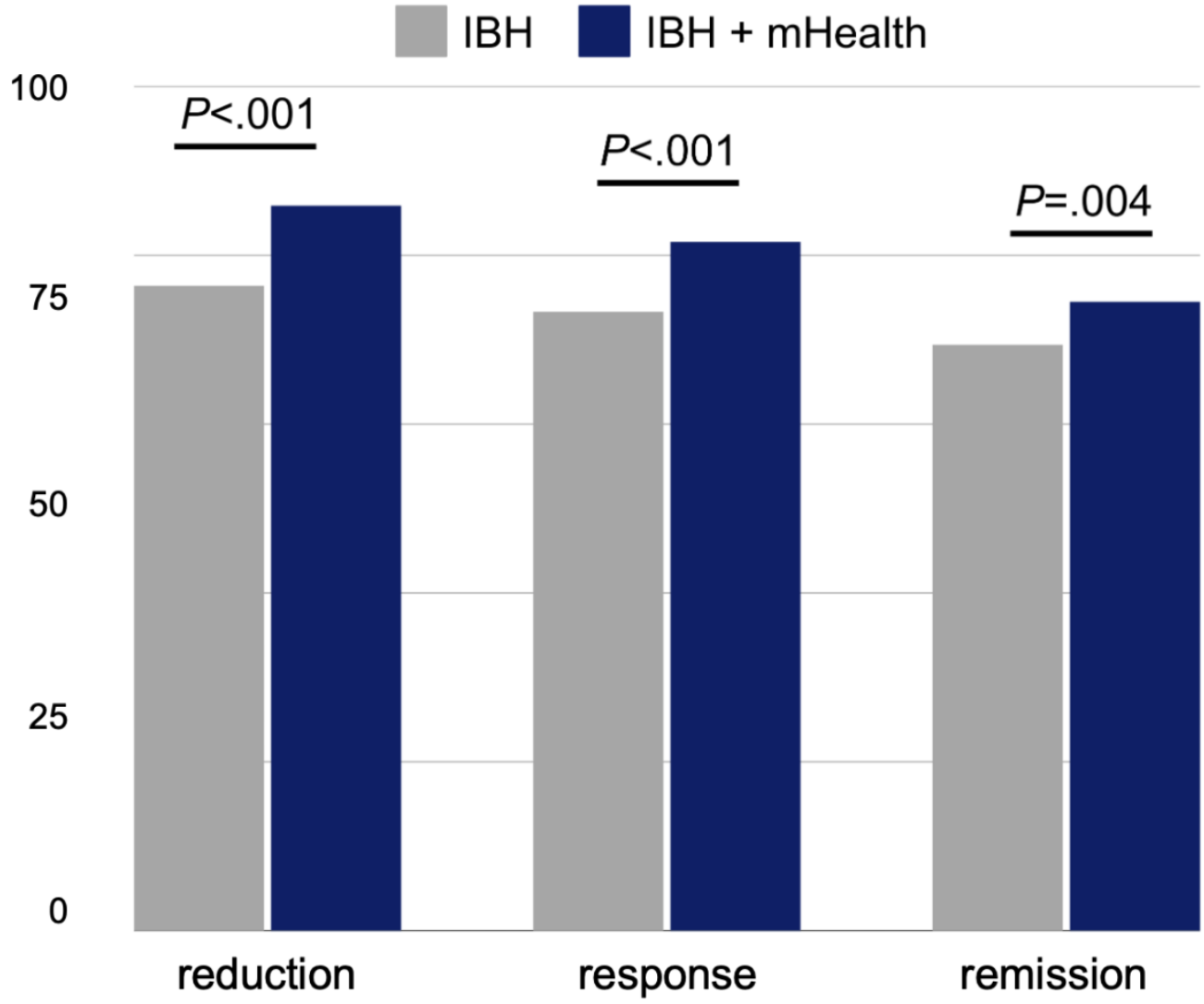
IBH+mHealth Is Linked to Proportionately Greater Improvements Than IBH on Its Own

Next, we aimed to dissociate the contribution of mHealth to relationships between treatment types and patient outcomes. To do this, we analyzed the IBH (n=2765) and IBH+mHealth (n=844) cohorts. Patients in these groups received equivalent treatment, with one important exception: the latter also received mHealth-administered content, measurement, and support (see *Methods*). Having already established that these cohorts did not significantly differ in their starting depression severity, we proceeded to test the groups for differences in the proportion of patients who reached the clinical benchmarks for reduction, response, and remission (see Methods). We found that significantly more patients in the IBH+mHealth cohort reached reduction, response, and remission compared with patients engaged in IBH without mHealth support (Figure 2; reduction: χ_1^2=111.45, P<.001; response: χ_1^2=28.82, P<.001; remission: χ_1^2=8.04, P=.004). In total, 85.9% (725/844), 81.6% (689/844), and 74.5% (629/844) of IBH+mHealth patients reached reduction, response, and remission, respectively, compared with 76.4% (2112/2765), 73.3% (2027/2765), and 69.4% (1919/2765) of IBH patients. Surprisingly, however, IBH without mHealth was associated with significantly faster mean times to response (116 d vs 96.1 d; independent-samples t test: t_{2714}=2.08; P=.04)

XSL•FO
RenderX

and remission (134.9 d vs 113.5 d; t_{2546}=6.32; *P*<.001) benchmarks compared with integrated care+mHealth. Subsequent hierarchical regression modeling (described below) suggests that this finding may be a byproduct of day-to-day fluctuations in PHQ-9 scoring, and less suggestive of treatment effectiveness, per se. In addition, this finding may reflect the tendency of patients to be more forthright in virtually administered PHQ-9 screenings, which afford greater privacy than in-person screenings [38]. Nevertheless, this finding warrants additional research.

Figure 2. We compared the percentage of patients that reached clinical benchmarks for reduction, response, and remission between our two largest cohorts to provide insight into the links between mHealth augmentation and enhanced patient outcomes. Integrated behavioral health care+mHealth (n=844) was associated with significantly greater proportions of patients reaching clinical benchmarks of reduction (χ_1^2=111.45; *P*<.001), response (χ_1^2=28.82; *P*<.001), and remission (χ_1^2=8.04; *P*=.004) compared with IBH on its own (n=2765). IBH: integrated behavioral health care; mHealth: mobile behavioral health care.



mHealth-Augmented CoCM Is Linked to Robust Decreases in Depression Severity, Irrespective of Clinical Benchmarks

Although analyzing clinical benchmarks of reduction, response, and remission can be an informative measure of patient outcomes, additional approaches can be useful to account for day-to-day variation in symptoms. PHQ-9 scores fluctuate, so examining overall trends in symptomatology, including scores recorded after reaching a clinical benchmark, provides a more complete picture of patients' changing depression severity. Because of this, we sought to measure severity trends in each cohort, irrespective of when or whether patients reached clinical benchmarks. To do this, we built a hierarchical regression model with fixed effects of cohort and random effects of individual patients to evaluate the linear associations between treatment approaches and changes in depression severity over time. We calculated within-group regression coefficients and standard errors for each patient and collapsed those values by cohort to compare linear trends between groups using the approach described by Aiken and West [39]. While depression severity among all cohorts improved over time, we found that the CoCM+mHealth cohort was associated with significantly faster, sustained symptom reduction than the IBH (z=2.47; *P*=.01) or IBH+mHealth (z=4.43; *P*<.001) cohort. In CoCM+mHealth, every 8.7 days of care was associated with a sustained, groupwise 1-point decrease in average PHQ-9 score. By comparison, in the IBH cohort, a 1-point decrease occurred

XSL•FO
RenderX

every 19.8 days, on average, and, in the IBH+mHealth cohort, this association was 1 point every 43.5 days. Despite the clinically significant difference in trajectories between the IBH and IBH+mHealth cohorts, these groups' differences were not statistically significant (z=1.25; *P*=.21). Like the clinical benchmark analyses discussed in the preceding section, this finding should be interpreted cautiously in light of established differences between in-person and digitally administered PHQ-9 scores [38].

Pre-CoCM mHealth Is Associated With Significant Improvements in Depression Severity

A subsection of 92 of our patients had early access to the mHealth app and engaged with mHealth modules at least once per month in the 3 months preceding CoCM treatment. As shown in Figure 1, these patients, on average, had the most severe starting depression of the cohorts we examined, and comprised the only cohort that started with "moderate-to-severe" depression. We hypothesized that this group would see reduced depression severity prior to the start of CoCM. To test this hypothesis, we compared differences in mean clinical index scores (ie, the patients' first ever PHQ-9 scores of ≥10) to their CoCM index scores (ie, PHQ-9 scores of ≥10 circa commencement of CoCM treatment). We found that pre-CoCM mHealth engagement was associated with a statistically significant reduction in PHQ-9 scores, which fell 7.20% from an average of 15.20 at the time of clinical indexing to 14.10 circa the start of CoCM (paired-samples *t* test: *t*=2.44, *P*=.0). This reduction brought the pre-CoCM cohort from an average starting severity of "moderate-to-severe" to "moderate," which was then similar to that of the other groups.

Discussion

Primary Findings

Here, we have presented the results of an exploratory, retrospective cohort study of real-world data collected from 3837 patients experiencing depression and receiving various levels of integrated care, with and without mHealth support, at a subspecialty pain clinic. While much of the existing literature has focused on the theoretical benefits of mHealth [28,36], our study is one of the first to empirically examine its impact within a real-world integrated care setting among patients with comorbid chronic pain and depression. Our findings contribute to the growing literature on the potential for mHealth to support complex patients between clinical visits, for instance, through the administration of remote assessments [3,35,40]. Moreover, this study provides new evidence linking mHealth support to enhanced behavioral health outcomes, underscoring the need for further research to establish causality and optimize its clinical implementation.

Augmenting integrated care with mHealth technology offers several benefits, including remote automated measurement, real-time digital support, and enhanced engagement between visits. In addition, mHealth-augmented approaches empower providers with nonpharmacological treatment options that could potentially reduce opioid prescribing and, by extension, misuse [41]. Our study suggests that mHealth support can be a valuable tool to reduce depression symptoms in patients who suffer from comorbid chronic pain and depression. Indeed, our CoCM+mHealth cohort saw the most rapid and robust improvements in depression severity, irrespective of clinical benchmarks. While further research is needed, these findings add to the literature evidencing CoCM's effectiveness [26,27] while hinting at the intriguing possibility that mHealth-augmented CoCM may outperform CoCM on its own. Meanwhile, between the two cohorts in which mHealth was dissociable (ie, IBH and IBH+mHealth, which together comprised 94% of our total sample; Figure 2), a significantly greater proportion of patients achieved clinical reduction, response, and remission when their integrated care was supplemented by the mHealth app. We additionally found that mHealth engagement was associated with a statistically significant decrease in depression severity within a cross-section of CoCM patients who had early access to the mHealth platform.

Interpreting Our Findings in the Context of the Existing Literature

Our study adds to the growing evidence that mHealth interventions can improve depression outcomes, particularly in integrated care settings. Similar to Kim and Lee [36], who demonstrated that mHealth tools effectively managed depression in real-world populations, we found that mHealth support enhanced outcomes over time. Likewise, Torous et al [28] highlighted the critical role of mHealth in expanding access to mental health care during the COVID-19 pandemic. Importantly, our study suggests that mHealth is more effective when used as part of a coordinated care framework, rather than as a stand-alone tool. This finding aligns with Unützer et al [26], who emphasized that CoCM models are essential for managing patients with comorbid behavioral and physical health conditions, such as depression and chronic pain. Recent studies [42,43] have further confirmed that mHealth's impact on mental health outcomes is maximized when integrated with comprehensive care systems, supporting our approach.

In contrast to the robust effects seen with integrated mHealth approaches, challenges around patient engagement remain a significant barrier: promoting adherence to mHealth interventions is an ongoing challenge [40] and can be especially difficult among patients with multiple comorbidities [44,45]. This may have been reflected in our pre-CoCM cohort, where variability in engagement potentially influenced the effectiveness of mHealth support (although we hesitate to strongly interpret this cohort's outcomes; see Results for rationale). While our findings echo the benefits reported in studies like Ye et al's [46] meta-analysis on internet-based cognitive behavioral therapy, they also underscore the complexities of delivering mHealth interventions in real-world clinical settings, particularly among patients with multiple comorbidities, and reinforce the need for tailored engagement strategies.

Study Limitations

Several limitations should be noted when interpreting our results. First, our retrospective cohort study lacked a true control group, and we were therefore unable to compare our treatment cohorts to placebo or untreated controls. Because of this, our study was

XSL•FO
RenderX

not designed to investigate the potential adverse impact of untreated depression over time, which may be especially relevant for patients with chronic pain. Future prospective cohort studies that do so would help contextualize our findings. We also lacked access to robust electronic health record data, including nondepression comorbidities and social determinants of health–related information, which could have affected our observations. However, the large sample sizes in the IBH and IBH+mHealth cohorts and the a posteriori randomization of our retrospective design likely mitigated the impact of these unmeasured confounds. Furthermore, because we collected PHQ-9 scores but not self-reported pain severity, we are unable to evaluate the associations between treatment cohorts and changing pain symptomatology; nor did we have information on prescription pain medication or opioid use. Finally, a majority (86%) of patients declined to provide optional demographic data (eg, race and ethnicity) during patient intake, which precludes the examination of demographic relationships to outcomes. Nevertheless, our study provides early evidence on the associations between mHealth-augmented integrated care and patient outcomes within a large sample of complex patients.

Future Directions

Our findings set the stage for future research into the causal effects of mHealth engagement and its potential to augment treatment across the integrated care continuum based on the needs, motivations, and readiness of the patient and their health care system. Patients who present with depression and comorbid chronic pain are ideally suited for future randomized controlled studies examining the impact of mHealth interventions on complex patient outcomes in CoCM settings. Such studies would provide key insights into the potential for evidence-based, depression-centered mHealth content (eg, digital cognitive behavioral therapy modules [46-48]) and remote measurements to reduce physical pain symptoms by targeting one pillar of these mutually reinforcing ailments [6-9,22].

Future research should also explore how mHealth interventions can be systematically integrated into existing care models such as CoCM, primary care, and specialty care settings. Process studies will be crucial in identifying the operational challenges associated with embedding mHealth tools into standard care workflows, such as aligning mHealth use with clinical visits, ensuring data interoperability across platforms, and training care teams to use mHealth insights effectively. These studies could reveal barriers to implementation, such as limited clinician buy-in, technology literacy issues among patients, and regulatory or privacy concerns—any of which may impact the scalability of mHealth solutions. Understanding how mHealth can complement, rather than disrupt, existing workflows will be critical to successful integration.

In parallel, future research should focus on understanding how specific components of mHealth, such as tailored intervention timing and personalized engagement strategies, influence outcomes across different patient populations. Investigating the optimal duration and frequency of mHealth augmentation will also help to establish the minimum effective dose for improving both depression and pain-related outcomes. Studies evaluating long-term outcomes, such as sustained symptom reduction and quality-of-life improvements, would offer valuable data on the durability of mHealth effects.

In addition, cost-of-care analyses should be conducted to explore how mHealth integrates with existing care frameworks and whether it offers cost-effective solutions for managing depression and chronic pain [27,33,34]. Another key area of research is the potential for mHealth to reduce opioid prescribing and misuse, particularly in patients with chronic pain and psychiatric comorbidities [24,41]. Investigating the relationship between mHealth usage, opioid reduction, and substance use disorder prevention could have profound implications for public health policy and pain management strategies.

Acknowledgments

The authors thank Osman Zuberi and Katie Finkbeiner for liaising between organizations involved in the study, which was instrumental to the timely provision of datasets used. We thank Carol Sarnoski for insights into integrated behavioral health care processes, which were crucial to our understanding of how in-clinic care was administered during the study. The authors also thank Samuel Kampa for his mentorship and analytical insights, which significantly improved this manuscript.

Data Availability

The datasets generated during and/or analyzed during the current study may be made available by the corresponding author, DH, upon reasonable request.

Authors' Contributions

DH contributed to formal analysis, methodology, visualization, and writing – original draft. DH, AB, and TZ contributed to methodology, conceptualization, and writing – review and editing. DH, AB, and SR contributed to project administration. MH and SR performed data collection, data curation, and investigation. AB, TZ, and SR contributed to resources and supervision.

Conflicts of Interest

DH, AB, MH, and TZ were salaried employees of NeuroFlow, Inc, during this study. SR was employed as the medical director of Pain and Spine Specialists during this study. This study received no outside funding or grants.

XSL•FO
RenderX

References

1. Vos T, Abajobir AA, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet 2017 Sep;390(10100):1211-1259. [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28711585]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Pardes A, Rene R, Chun P, Cherson M. Technology-enabled behavioral health integration decreases emergency department utilization. Clin Pract Epidemiol Ment Health 2022;18:e174501792208150. [doi: 10.2174/17450179-v18-e2208150] [Medline: 37274856]
4. Lake J, Turner MS. Urgent need for improved mental health care and a more collaborative model of care. Perm J 2017;21:17-024. [doi: 10.7812/TPP/17-024] [Medline: 28898197]
5. Arnow BA, Hunkeler EM, Blasey CM, et al. Comorbid depression, chronic pain, and disability in primary care. Psychosom Med 2006;68(2):262-268. [doi: 10.1097/01.psy.0000204851.15499.fc] [Medline: 16554392]
6. Banks SM, Kerns RD. Explaining high rates of depression in chronic pain: a diathesis-stress framework. Psychol Bull 1996;119(1):95-110. [doi: 10.1037//0033-2909.119.1.95]
7. Geisser ME, Roth RS, Theisen ME, Robinson ME, Riley JL. Negative affect, self-report of depressive symptoms, and clinical depression: relation to the experience of chronic pain. Clin J Pain 2000 Jun;16(2):110-120. [doi: 10.1097/00002508-200006000-00004] [Medline: 10870723]
8. de Heer EW, Gerrits M, Beekman ATF, et al. The association of depression and anxiety with pain: a study from NESDA. PLoS One 2014;9(10):e106907. [doi: 10.1371/journal.pone.0106907] [Medline: 25330004]
9. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: A literature review. Arch Intern Med 2003 Nov 10;163(20):2433-2445. [doi: 10.1001/archinte.163.20.2433] [Medline: 14609780]
10. Williams LS, Jones WJ, Shen J, Robinson RL, Weinberger M, Kroenke K. Prevalence and impact of depression and pain in neurology outpatients. J Neurol Neurosurg Psychiatry 2003 Nov;74(11):1587-1589. [doi: 10.1136/jnnp.74.11.1587] [Medline: 14617727]
11. Von Korff M, Scher AI, Helmick C, et al. United States National Pain Strategy for Population Research: concepts, definitions, and pilot data. J Pain 2016 Oct;17(10):1068-1080. [doi: 10.1016/j.jpain.2016.06.009] [Medline: 27377620]
12. Calati R, Laglaoui Bakhiyi C, Artero S, Ilgen M, Courtet P. The impact of physical pain on suicidal thoughts and behaviors: meta-analyses. J Psychiatr Res 2015 Dec;71:16-32. [doi: 10.1016/j.jpsychires.2015.09.004] [Medline: 26522868]
13. Costanza A, Chytas V, Piguet V, et al. Meaning in life among patients with chronic pain and suicidal ideation: mixed methods study. JMIR Form Res 2021 Jun 4;5(6):e29365. [doi: 10.2196/29365] [Medline: 34003136]
14. Kerr S, Fairbrother G, Crawford M, Hogg M, Fairbrother D, Khor KE. Patient characteristics and quality of life among a sample of Australian chronic pain clinic attendees. Intern Med J 2004 Jul;34(7):403-409. [doi: 10.1111/j.1444-0903.2004.00627.x] [Medline: 15271174]
15. Sprangers MA, de Regt EB, Andries F, et al. Which chronic conditions are associated with better or poorer quality of life? J Clin Epidemiol 2000 Sep;53(9):895-907. [doi: 10.1016/s0895-4356(00)00204-3] [Medline: 11004416]
16. Smith BH, Torrance N, Bennett MI, Lee AJ. Health and quality of life associated with chronic pain of predominantly neuropathic origin in the community. Clin J Pain 2007 Feb;23(2):143-149. [doi: 10.1097/01.ajp.0000210956.31997.89] [Medline: 17237663]
17. Adams G, Salomons TV. Attending work with chronic pain is associated with higher levels of psychosocial stress. Can J Pain 2021 May 18;5(1):107-116. [doi: 10.1080/24740527.2021.1889925] [Medline: 34189394]
18. Schwartz L, Slater MA, Birchler GR. The role of pain behaviors in the modulation of marital conflict in chronic pain couples. Pain 1996;65(2-3):227-233. [doi: 10.1016/0304-3959(95)00211-1] [Medline: 8826511]
19. Kemler MA, Furnée CA. The impact of chronic pain on life in the household. J Pain Symptom Manage 2002 May;23(5):433-441. [doi: 10.1016/s0885-3924(02)00386-x] [Medline: 12007761]
20. Graham-Engeland J, DeMeo NN, Jones DR, Mathur A, Smyth JM, Sliwinski MJ, et al. Individuals with both higher recent negative affect and physical pain have higher levels of C-reactive protein. Brain Behav Immun Health 2022;21:100431. [doi: 10.1016/j.bbih.2021.100431] [Medline: 35243409]
21. Raver C, Uddin O, Ji Y, Li Y, Cramer N, Jenne C, et al. An amygdalo-parabrachial pathway regulates pain perception and chronic pain. J Neurosci 2020;40(17):3424-3442. [doi: 10.1523/JNEUROSCI.2012-19.2020] [Medline: 32217613]
22. Wilson TD, Valdivia S, Khan A, et al. Dual and opposing functions of the central amygdala in the modulation of pain. Cell Rep 2019 Oct 8;29(2):332-346. [doi: 10.1016/j.celrep.2019.09.011] [Medline: 31597095]
23. Hedegaard H, Bastian BA, Trinidad JP, Spencer M, Warner M. Drugs most frequently involved in drug overdose deaths: United States, 2011-2016. Natl Vital Stat Rep 2018 Dec;67(9):1-14. [Medline: 30707673]
24. Bonnie RJ, Schumacher MA, Clark JD, Kesselheim AS. Pain management and opioid regulation: continuing public health challenges. Am J Public Health 2019 Jan;109(1):31-34. [doi: 10.2105/AJPH.2018.304881] [Medline: 32941766]

XSL•FO
RenderX

25. Finnerup NB, Attal N, Haroutounian S, McNicol E, Baron R, Dworkin RH, et al. Pharmacotherapy for neuropathic pain in adults: Systematic review, meta-analysis, and updated NeuPSIG recommendations. Lancet Neurol 2015 Feb;14(2):162-173. [doi: 10.1016/S1474-4422(14)70251-0] [Medline: 25575710]
26. Unützer J, Carlo AD, Collins PY. Leveraging collaborative care to improve access to mental health care on a global scale. World Psychiatry 2020 Feb;19(1):36-37. [doi: 10.1002/wps.20696] [Medline: 31922696]
27. Unützer J, Katon W, Callahan CM, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. JAMA 2002 Dec 11;288(22):2836-2845. [doi: 10.1001/jama.288.22.2836] [Medline: 12472325]
28. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848. [doi: 10.2196/18848] [Medline: 32213476]
29. Bolin JN, Bellamy GR, Ferdinand AO, et al. Rural Healthy People 2020: new decade, same challenges. J Rural Health 2015;31(3):326-333. [doi: 10.1111/jrh.12116] [Medline: 25953431]
30. Morales DA, Barksdale CL, Beckel-Mitchener AC. A call to action to address rural mental health disparities. J Clin Transl Sci 2020 May 4;4(5):463-467. [doi: 10.1017/cts.2020.42] [Medline: 33244437]
31. Knaak S, Mantler E, Szeto A. Mental illness-related stigma in healthcare. Healthc Manage Forum 2017 Mar;30(2):111-116. [doi: 10.1177/0840470416679413]
32. Ramanuj P, Ferenchik E, Docherty M, Spaeth-Rublee B, Pincus HA. Evolving models of integrated behavioral health and primary care. Curr Psychiatry Rep 2019 Jan 19;21(1):4. [doi: 10.1007/s11920-019-0991-0] [Medline: 30656618]
33. Carlo AD, Unützer J, Ratzliff ADH, Cerimele JM. Financing for collaborative care – a narrative review. Curr Treat Options Psychiatry 2018 Sep;5(3):334-344. [doi: 10.1007/s11920-019-0985-4] [Medline: 30661126]
34. Rocks S, Berntson D, Gil-Salmerón A, et al. Cost and effects of integrated care: a systematic literature review and meta-analysis. Eur J Health Econ 2020 Nov;21(8):1211-1221. [doi: 10.1007/s10198-020-01217-5] [Medline: 32632820]
35. Pardes A, Lynch W, Miclette M, McGeoch E, Daly BP. Use of a mobile health (mHealth) platform for remote assessment of suicidal ideation, depression, and anxiety: a longitudinal retrospective study. Innov Dig Health, Diagn, Biomark 2022 Jan 1;2(2022):8-15. [doi: 10.36401/IDDB-21-03]
36. Kim S, Lee K. Screening for depression in mobile devices using Patient Health Questionnaire-9 (PHQ-9) data: a diagnostic meta-analysis via machine learning methods. Neuropsychiatr Dis Treat 2021;17:3415-3430. [doi: 10.2147/NDT.S339412] [Medline: 34848962]
37. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
38. Miller DP Jr, Foley KL, Bundy R, et al. Universal screening in primary care practices by self-administered tablet vs nursing staff. JAMA Netw Open 2022 Mar 1;5(3):e221480. [doi: 10.1001/jamanetworkopen.2022.1480] [Medline: 35258581]
39. Aiken LS, West SG. Multiple Regression: Testing and Interpreting Interactions: SAGE Publications; 1991.
40. Ramos RM, Cheng PGF, Jonas SM. Validation of an mHealth app for depression screening and monitoring (psychologist in a pocket): correlational study and concurrence analysis. JMIR Mhealth Uhealth 2019 Sep 16;7(9):e12051. [doi: 10.2196/12051] [Medline: 31538946]
41. Sandhu HK, Booth K, Furlan AD, et al. Reducing opioid use for chronic pain with a group-based intervention: a randomized clinical trial. JAMA 2023 May 23;329(20):1745-1756. [doi: 10.1001/jama.2023.6454] [Medline: 37219554]
42. Lu SC, Xu M, Wang M, et al. Effectiveness and minimum effective dose of app-based mobile health interventions for anxiety and depression symptom reduction: systematic review and meta-analysis. JMIR Ment Health 2022 Sep 7;9(9):e39454. [doi: 10.2196/39454] [Medline: 36069841]
43. Leong QY, Sridhar S, Blasiak A, et al. Characteristics of mobile health platforms for depression and anxiety: content analysis through a systematic review of the literature and systematic search of two app stores. J Med Internet Res 2022 Feb 4;24(2):e27388. [doi: 10.2196/27388] [Medline: 35119370]
44. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
45. Philippe TJ, Sikder N, Jackson A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022 May 12;9(5):e35159. [doi: 10.2196/35159] [Medline: 35551058]
46. Ye YY, Zhang YF, Chen J, et al. Internet-based cognitive behavioral therapy for insomnia (ICBT-i) improves comorbid anxiety and depression-a meta-analysis of randomized controlled trials. PLoS One 2015;10(11):e0142258. [doi: 10.1371/journal.pone.0142258] [Medline: 26581107]
47. Luik AI, Bostock S, Chisnall L, et al. Treating depression and anxiety with digital cognitive behavioural therapy for insomnia: a real world NHS evaluation using standardized outcome measures. Behav Cogn Psychother 2017 Jan;45(1):91-96. [doi: 10.1017/S1352465816000369] [Medline: 27456542]
48. Holley D, Lubkin S, Brooks A, Zaubler T, Carney C. Digital CBT interventions predict robust improvements in anxiety and depression symptoms: a retrospective database study. Innov Dig Health, Diagn, Biomark 2024 Jan 1;4(2024):53-55. [doi: 10.36401/IDDB-24-4]

XSL•FO
RenderX

Abbreviations

BH: behavioral health
CoCM: collaborative care
FDA: US Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
IBH: integrated behavioral health care
mHealth: mobile behavioral health care
PHQ-9: Patient Health Questionnaire-9

Edited by L Buis; submitted 14.09.23; peer-reviewed by C Budhathoki, J López-Marcos, R Koli; revised version received 02.10.24; accepted 18.11.24; published 10.01.25.

Please cite as:
Holley D, Brooks A, Hartz M, Rao S, Zaubler T
mHealth-Augmented Care for Reducing Depression Symptom Severity Among Patients With Chronic Pain: Exploratory, Retrospective Cohort Study
JMIR Mhealth Uhealth 2025;13:e52764
URL: https://mhealth.jmir.org/2025/1/e52764
doi:10.2196/52764

© Dan Holley, Amanda Brooks, Matthew Hartz, Sudhir Rao, Thomas Zaubler. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Effectiveness of a Mobile App for Breast Cancer Self-Management on Self-Efficacy: Nonrandomized Intervention Trial

Sun Mi Kim[1], MD, PhD; Da Seul Kim[1], PhD; Yoonsung Jang[1], BA; Min Kyoon Kim[2], MD, PhD; Eun-Seung Yu[3], PhD; Doug Hyun Han[1], MD, PhD; Hee Jun Kim[4], MD, PhD

[1]Department of Psychiatry, College of Medicine, Chung-Ang University, Seoul, Republic of Korea

[2]Department of Surgery, College of Medicine, Chung-Ang University, Seoul, Republic of Korea

[3]Department of Counseling Psychology, The Cyber University of Korea, Seoul, Republic of Korea

[4]Division of Medical Oncology, Department of Internal Medicine, College of Medicine, Chung-Ang University, Seoul, Republic of Korea

Corresponding Author:
Hee Jun Kim, MD, PhD
Division of Medical Oncology
Department of Internal Medicine, College of Medicine
Chung-Ang University
84 Heukseok-ro
Dongjack-gu
Seoul, 06974
Republic of Korea
Phone: 82 262993159
Fax: 82 262992017
Email: heejun.dino11@gmail.com

Abstract

Background: Numerous mobile apps have been developed for patients with cancer. However, there is still no comprehensive app for patients with breast cancer that integrates evidence-based medical information, psychological support, and schedule management through a multidisciplinary medical approach.

Objective: We aimed to investigate whether a mobile app designed to assist in the self-management of patients with breast cancer is feasible and positively affects their self-efficacy and other psychological aspects.

Methods: The Cancer Manager (CAMA) app was developed to assist in the self-management of patients with breast cancer and survivors of cancer according to cancer trajectory. Its functionalities include providing evidence-based digitalized information created by experts, managing patients' medication and medical appointment schedules, and providing a delayed question and answer system for patients to query health care professionals. In this nonrandomized intervention trial, we analyzed data from 66 patients with breast cancer, divided into experimental (CAMA: n=34, 52%) and control (treatment as usual: n=32, 48%) groups. Group allocation was determined based on the patient's willingness to use the app and access to compatible smartphones. Outcome measures included the Korean version of the Cancer Survivor Self-Efficacy Scale, the Korean version of the Mini-Mental Adjustment to Cancer (K-Mini-MAC) Scale, the World Health Organization Quality of Life Brief Version, Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7 (GAD-7), and Menopause Emotional Symptoms Questionnaire (MESQ). A user satisfaction survey was also conducted.

Results: Throughout the intervention period, the CAMA group (vs treatment as usual group) demonstrated significant improvements in the seeking help and support subscale of the Korean version of the Cancer Survivor Self-Efficacy Scale ($F_{1,64}$=5.09; P=.03), the psychological well-being subscale of the World Health Organization Quality of Life Brief Version ($F_{1,64}$=5.48; P=.02), the anxious preoccupation subscale ($F_{1,64}$=5.49; P=.02) and positive attitude subscale ($F_{1,64}$=5.44; P=.02) of the K-Mini-MAC Scale, PHQ-9 ($F_{1,64}$=4.83; P=.03), GAD-7 ($F_{1,64}$=5.48; P=.02), and MESQ ($F_{1,64}$=4.30; P=.04). Changes in the anxious preoccupation subscale of the K-Mini-MAC Scale scores were positively correlated with changes in the PHQ-9 (r=0.46; P=.007) and GAD-7 (r=0.41; P=.02) scores and negatively correlated with changes in the positive attitude subscale of the K-Mini-MAC Scale scores (r=–0.36; P=.04). Changes in the PHQ-9 scores were positively correlated with changes in the

XSL•FO
RenderX

GAD-7 (r=0.66; P<.001) and MESQ (r=0.35; P=.04) scores. The user satisfaction survey offered insights into the CAMA app's positive impact; trust-building outcomes; and opportunities for enhancement, such as the inclusion of communication tools and continued content enrichment.

Conclusions: The mobile app for breast cancer self-management, CAMA, was deemed feasible and showed promise in improving the patients' self-efficacy regarding seeking help and support, positive attitude toward cancer, and psychological well-being. In addition, its use might help reduce anxious preoccupation with cancer, depressive mood, anxiety, and menopausal emotional symptoms.

Trial Registration: Clinical Research Information Service KCT0007917; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=23348

(JMIR Mhealth Uhealth 2025;13:e63989) doi:10.2196/63989

KEYWORDS

breast cancer; mobile health; mHealth; health education; self-efficacy; psychological adjustments; mobile phone

Introduction

Background

Breast cancer is the most prevalent cancer among women worldwide, accounting for approximately 1 out of every 4 cancer diagnoses [1]. Despite advancements in breast cancer treatment, patients continue to experience significant psychological and physical burdens [2,3]. The wide range of subtypes of breast cancer adds to its complexity, as each has its distinct biological and genetic characteristics. This complexity leads to the need for highly individualized chemotherapy regimens tailored to factors such as hormone receptor status, human epidermal growth factor receptor 2 status, and genetic mutations [4,5]. Research also shows that patients with breast cancer often face, throughout the cancer journey, various unmet needs, particularly in areas such as physical and functional challenges, psychological support, and access to clear medical information [6-8]. More specifically, common issues include fatigue, pain, anxiety, fear of recurrence, difficulties in daily activity management, and a lack of coordinated care [9]. The needs of patients with breast cancer are also often more pronounced in younger patients and those with advanced cancer, especially during the transition from active treatment to survivorship [3,7].

Among patients with cancer, unmet needs have been linked to increased long-term distress and decreased quality of life (QOL) [10,11]. The symptom burden that arises from these unmet needs significantly influences the physical and psychological QOL of survivors of cancer [12]. Therefore, interventions to address these unmet needs and support the physical and psychological health of survivors of cancer are crucial at every stage of the cancer journey. However, in clinical settings, there can be physical constraints (eg, associated with staffing, space, and time limitations) to realize needs of all survivors of cancer, and attempts to address these needs may impose additional burdens on health care providers and systems. Given this reality, it is unsurprising that interest in using mobile apps in medical settings has been growing, enabling stakeholders to overcome challenges in addressing the needs of patients with cancer [13].

Prior Work

Given that breast cancer is the most common cancer (vs other cancer types) among women with a higher prevalence of unmet needs, many cancer-related apps are primarily targeted at this population [14]. Mobile health (mHealth) apps for patients with breast cancer have also become increasingly sophisticated, integrating various functionalities to support patients throughout their treatment and survivorship phases. Indeed, the extant apps for this type of cancer address a range of needs, including information on survivorship, upper limb dysfunction, sleep disturbances, nutrition, general health education, emotional support, energy balance challenges, and motivation [15]. Furthermore, they facilitate continuous monitoring, real-time feedback, and resource provision, which are essential in managing the complexities of breast cancer treatment [15]. mHealth apps have been reported to provide significant benefits for symptom management, patient engagement, and self-efficacy, which in turn are all critical for successful long-term outcomes [16,17]. mHealth-based intervention implementation has also been shown to positively influence QOL, stress, and weight management of survivors of breast cancer [17]. Particularly, these interventions enable patients with breast cancer to report symptoms and treatment-related side effects, promoting effective self-management [18].

Despite the advances, current research on mHealth apps for patients with breast cancer features several gaps. For instance, most research-based apps focus on only 1 or 2 functions (eg, information delivery, adherence management, lifestyle enhancement, psychological support, or referral and linkage to resources) and thus fail to integrate all these features into a single app. In addition, most apps lack personalized recommendations based on individual patient data and face usability challenges, particularly for diverse populations [19,20]. Researchers have also posited that the effectiveness of the psychological support features of mHealth apps is subpar and that the specific needs of different breast cancer subtypes are underexplored [21]. Studies have traditionally focused on improving QOL [22], often at the expense of investigating other critical clinical outcomes and the cost-effectiveness of these technologies in health care settings [23].

Theory

The development of our app, especially its structure and functionality, was guided by 2 pivotal theoretical frameworks, namely, the chronic care model (CCM) and the health belief model (HBM). The CCM emphasizes the importance of an integrated approach to managing chronic diseases such as breast

XSL•FO
RenderX

cancer [24,25]. It advocates for a patient-centered approach that not only focuses on treating the disease but also on improving patients' self-management capabilities and fostering long-term wellness [25]. The CCM informed our app's design by its emphasis on continuous and integrative care that incorporates medical, psychological, and social support. For example, features such as real-time symptom tracking, personalized health education, and coordinated care planning were developed to align with the CCM's core elements, ensuring that the app supports patients in managing their health proactively and holistically [26,27]. Meanwhile, the HBM was essential in addressing the psychological and behavioral aspects of health management within the app. The HBM posits that individuals' health-related behaviors are influenced by their perceptions of the severity of a health condition, their susceptibility to the condition, the benefits of taking preventive action, and the barriers to doing so [28]. This model was particularly influential in the development of the app's psychological support module, which aims to enhance patients' motivation and confidence in condition management [29]. By incorporating tools that address perceived barriers (eg, educational resource provision) and highlighting the benefits of adherence to treatment and self-care routines, the app was designed to empower patients to take an active role in managing their health [30,31].

These frameworks provided a structured approach for the development of the app, ensuring that it not only meets the immediate medical needs but also supports long-term behavioral change and self-efficacy improvement among patients with breast cancer. The integration of CCM and HBM into the app enabled the delivery of proactive and personalized patient-centered care that is supportive of the complex psychological and physical needs of survivors of breast cancer.

Study Aim

This study aimed to develop a self-management support app tailored to the needs of patients with breast cancer and verify its feasibility. Furthermore, we investigated whether the app would have a positive effect on patient self-efficacy and other psychological aspects compared to a control group—hereinafter treatment as usual (TAU) group—that did not use the app. In this study, we chose self-efficacy as the primary outcome based on the self-efficacy theory by Bandura [32], which posits that individuals with higher self-efficacy are more likely to engage in self-management behaviors, leading to better health outcomes [33]. This theory is particularly relevant in chronic disease management, where patient engagement in treatment and self-care practices is crucial for long-term treatment success. Enhancing self-efficacy can empower patients to better cope with the physical and psychological burdens associated with breast cancer, thus improving their overall QOL.

Our research differs from previous work because it involved the incorporation of a broader range of features into a single app, providing a more holistic and culturally sensitive tool (vs other similar apps) tailored to the specific needs of patients with breast cancer. By addressing the identified shortcomings in the literature and using a multidisciplinary approach, we aimed to offer a comprehensive mHealth solution where other similar apps have focused on fewer aspects of care. Consequently, we developed 4 modules: an evidence-based medical information module, a psychological support module, a medication and appointment management module, and a patient inquiry module. On the basis of these aims, we proposed the following hypotheses:

- Hypothesis 1: app use will significantly improve self-efficacy among patients with breast cancer compared to routine care.
- Hypothesis 2: app use will significantly improve QOL; mental adjustment to cancer; and psychological symptoms such as depression, anxiety, and menopausal emotional symptoms.
- Hypothesis 3: there will be a statistically significant association between the amount of change in the psychological variables that improved after app use.

Methods

Study Design

This study was a 12-week nonrandomized controlled intervention trial. This study used a nonrandomized design, selected mainly for convenience and logistical considerations. Randomized group assignment was deemed infeasible within the scope of this study due to resource and time constraints, as implementing randomization would have required additional infrastructure and staffing. While this design choice imposed certain limitations, it allowed for efficient recruitment and management of participants in real-world clinical settings. This study was reported in accordance with the TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) statement, and a completed TREND checklist is provided in Multimedia Appendix 1.

Ethical Considerations

The study protocol was reviewed and approved by the Institutional Review Board of Chung-Ang University Hospital (2201-025-494). All participants provided written informed consent before enrollment in the study. Participant data were anonymized and stored on a secure server accessible only to the research team. Participants received approximately US $14 as compensation for each of their baseline and follow-up visits. The clinical trial registration was approved by the Clinical Research Information Service (KCT0007917).

Study Participants and Recruitment

Using convenience sampling, 80 patients with pathologically proven clinical stage 0 to 4 breast cancer were recruited through an advertisement in hospital bulletins of the Digital Cancer Center at Chung-Ang University Hospital, located in Seoul, Korea, from January 2023 to June 2024. We focused on patients who voluntarily indicated their interest in participating in the study, and those who agreed to participate in the study were screened according to the following inclusion criteria: an adult female who was either undergoing active treatment for breast cancer or was a survivor of breast cancer. The exclusion criteria were as follows: current or history of uncontrolled medical diseases other than cancer; current or history of uncontrolled psychiatric diagnosis, including severe mood disorders, alcohol or substance use disorders, and intellectual disabilities; inability

XSL•FO
RenderX

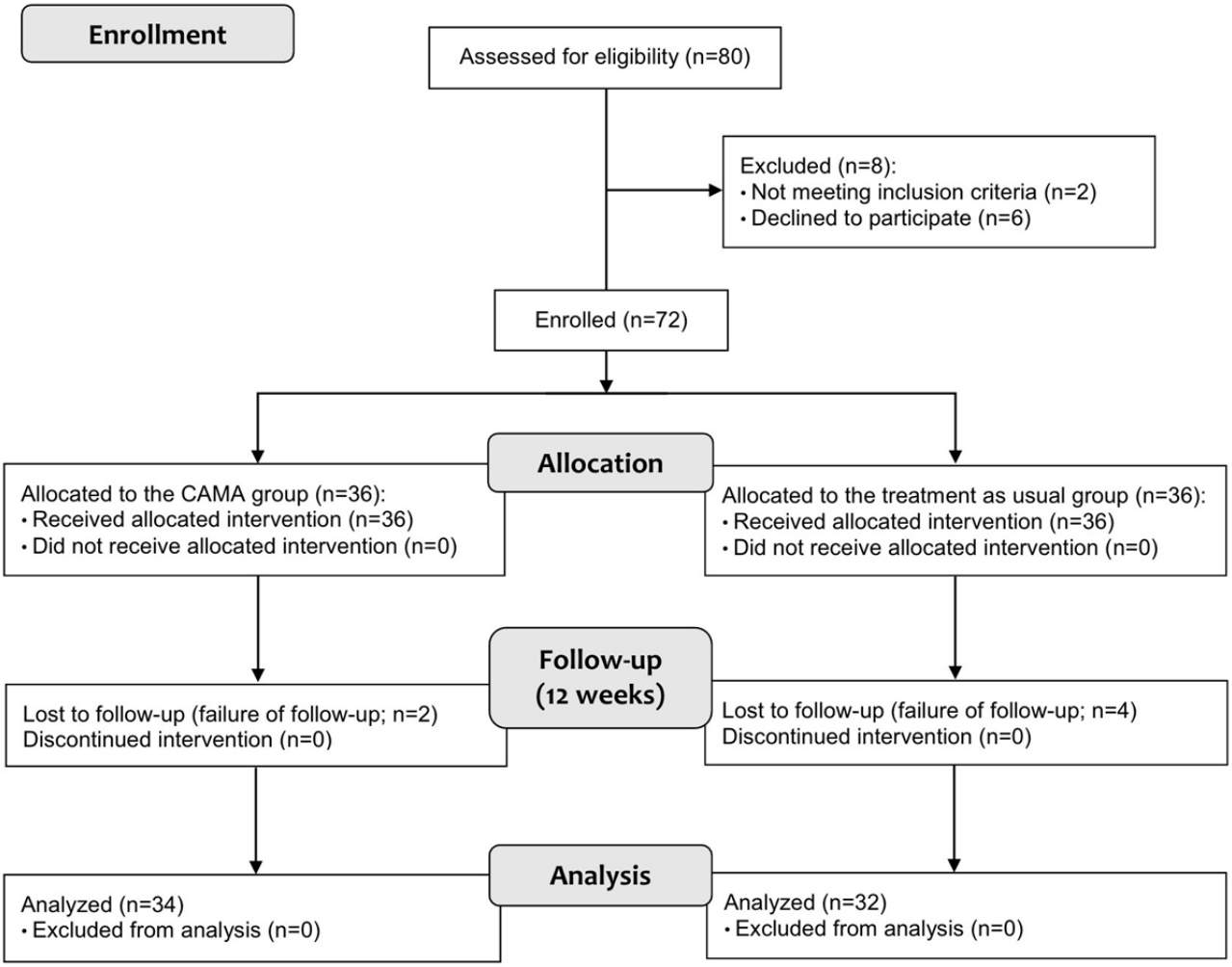
to engage in linguistic communication; and inability to use mobile apps on the phone.

Participants were assigned to the Cancer Manager (CAMA) group (ie, the experimental group) if they indicated a willingness to use the app and had access to a compatible smartphone, while those who did not meet these criteria were assigned to the TAU group. CAMA group participants used a mobile app, CAMA (HuDIT), designed to assist in the self-management of patients with breast cancer, and received routine care and education in the hospital; TAU group participants only received routine care and education. Routine care for patients with cancer aims to deliver comprehensive health care management services; does not encompass additional digital interventions; and typically includes regular medical follow-ups, symptom management, psychological support, nutrition counseling, palliative care, patient education, care coordination, monitoring and testing, access to emergency care, and social support services.

The eligibility screening and informed consent procedures were conducted by physicians (SMK, MKK, or HJK) or researchers (DSK or YJ). After the researchers provided explanations about the app, the app was installed on the patient's device either by the researchers or by the patients themselves. Then, the researchers provided participants with appropriate instructions about the intervention, including how to use the app, the assessment schedule, and how to contact the *Manager* of the CAMA or research team. Patients were instructed to access the app daily and use it for at least 10 minutes. We administered the baseline and follow-up questionnaires from January 2023 to June 2024 via in-person visits or phone calls.

In total, 80 patients with breast cancer underwent eligibility assessment (Figure 1), and some were excluded because of the following reasons: 6 (8%) declined to participate before enrollment, 1 (1%) had a major depressive disorder with agitation and suicidal ideation, and 1 (1%) was diagnosed with major neurocognitive disorder due to Alzheimer disease.

Figure 1. Study flow diagram. CAMA: Cancer Manager.



Allocation

We enrolled 72 participants based on the inclusion and exclusion criteria, with 36 (50%) participants each in the experimental (CAMA) and TAU groups (Figure 1). As aforementioned, group allocation was determined based on patients' willingness to use the app and their access to compatible smartphones. This study did not use matching techniques for the nonrandomized assignment procedures. The CAMA group used the CAMA mobile app for 12 weeks, whereas the TAU group received no intervention during that period. Both groups maintained their usual treatment regimens throughout the 12 weeks. Clinical assessments of self-efficacy and other psychological aspects were conducted immediately after the beginning of the

XSL•FO
RenderX

experiment (eg, right before patients downloaded the app and started using it) and immediately after the end of the experiment (eg, 1 day after the formal ending of the experiment).

App Development: CAMA

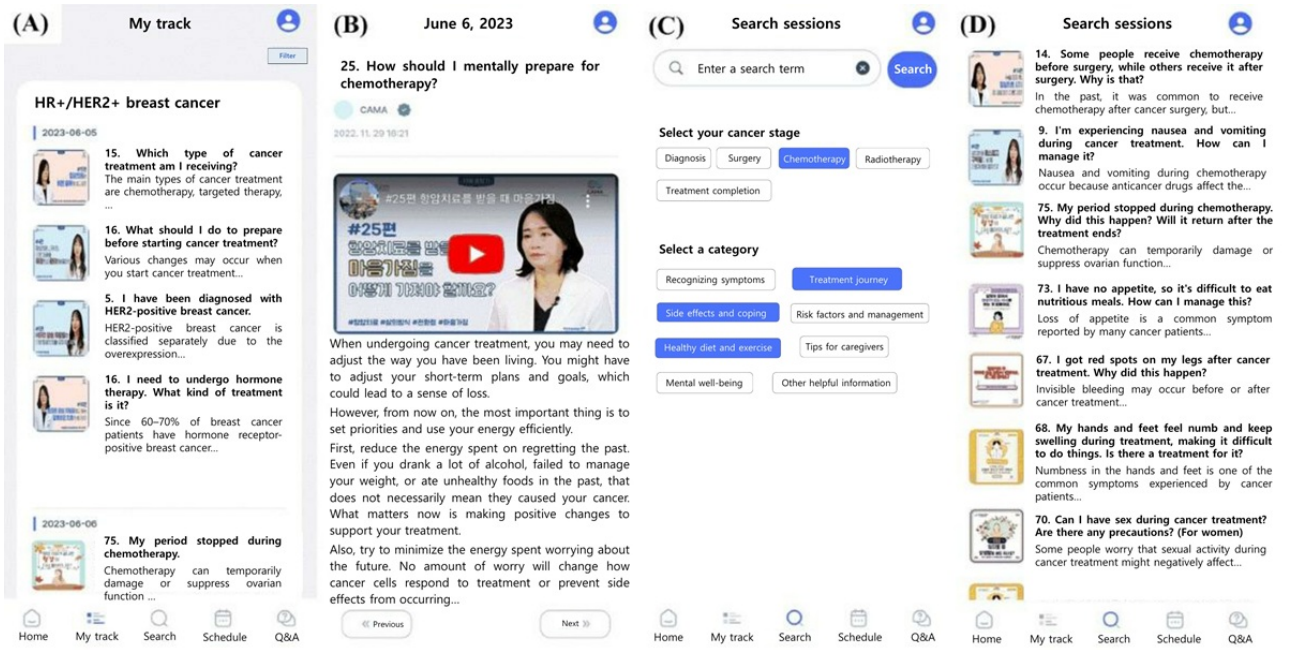
Overview

The CAMA app was developed to support patients with breast cancer in their self-management across different phases of the cancer journey, including at the time of diagnosis, during active treatment, and after treatment completion (ie, during the cancer survivorship phase). The CAMA app had three main functionalities: (1) it provided evidence-based digitalized information created by experts, (2) it managed patients' medication and medical appointment schedules, and (3) it delivered a delayed question and answer (Q and A) system whereby patients can ask questions to health care professionals.

Evidence-Based Digitalized Information

First, oncologists, surgeons, psychiatrists, and psychologists were recruited to create informative content tailored to each phase of the cancer journey (Figure 2). These digitalized contents, including videos, articles, images, and infographics, were registered on the website of the administrator of the CAMA app. The CAMA app provides personalized information to the patient depending on whether one is undergoing chemotherapy, surgery, or radiation therapy or is in the cancer survivorship phase. The digital contents covered topics such as information about breast cancer diagnosis, symptom management, preparation for treatment, management of treatment side effects, diet and nutrition, rehabilitation and exercise, psychological support, and information for caregivers.

Figure 2. Representative screens of the My care track and information content search features of the Cancer Manager app. This figure was created by translating the original Korean text from the app's screenshots into English. (A) My care track feature: this function allows patients to receive personalized information relevant to their breast cancer type (eg, hormone-positive and human epidermal growth factor receptor 2-positive) and treatment phase (eg, chemotherapy) through daily alarms. (B) Information content: clicking on the alarm window provides information to read on that day in the form of videos, infographics, and articles. All information was authored by experts in their respective fields (eg, hematologic oncology, breast surgery, psychiatry, and psychology). (C) Information content search feature: this function allows patients to directly search for the information they need. They can either enter search terms directly through the search box (at the top) or click on keywords for specific cancer trajectories and areas (at the bottom) before clicking the search button. (D) Information content search results: after entering search terms or clicking on keywords, pressing the search button displays a list of relevant contents, and a subsequent click on an item on the list reveals its details.



Physicians and a nurse, a social worker, or a psychologist (referred to as the *Manager*) define the most relevant content according to the cancer-related condition (referred to as *My Care Track*), and the *Manager* registers the content on the administrator's website. Alternatively, patients can search for and access information based on their requirements.

The *My Care Track* provides tailored content on a patient-specific basis, delivering helpful information regarding the patient's cancer treatment stage. The contents are listed by date, and the app sends reminders and provides recommended content daily. For instance, the app considers factors such as the patient's breast cancer subtype, time since cancer diagnosis, ongoing chemotherapy, or upcoming radiation therapy, providing essential daily guidance in concise, manageable pieces.

The *Manager* (referring to the nurse, social worker, or psychologist here) of the CAMA app is also responsible for coordinating the interactions between physicians and patients and the app and the offline treatment process. On the basis of previous research findings, barriers to the effectiveness of health care delivery through apps include a lack of feedback and support, a perception of the app as impersonal and cold, excessive or complicated information or intervention, discrepancies between the current treatment status and the information delivered, insufficient understanding of program use methods, and a lack of motivation [34]. The *Manager* of

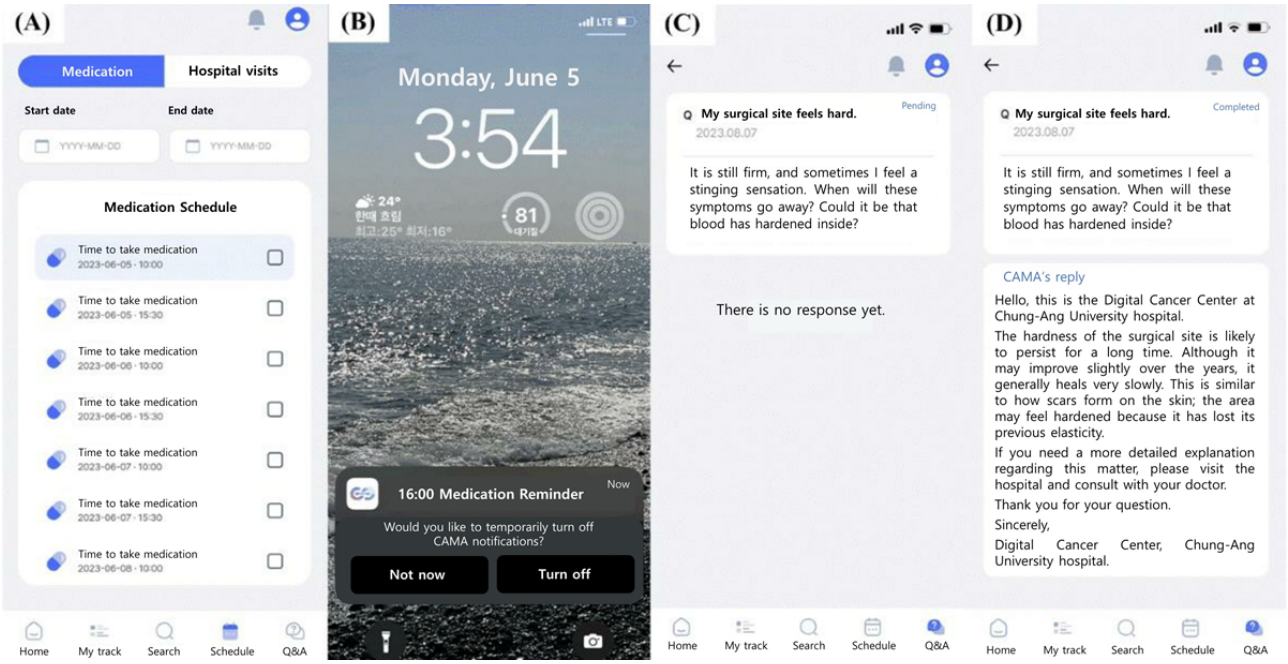
XSL•FO
RenderX

the CAMA app served to overcome these challenges, providing feedback and support, serving as a human communication channel, delivering tailored information or interventions, offering technical support for app use, and reinforcing motivation for app use. The *Manager* provided a communication channel between patients and health care professionals that was available during working hours (9 AM to 5 PM, Monday to Friday). In addition, every 2 weeks, the *Manager* texted participants to encourage the consistent use of the app. Every 4 weeks, the *Manager* called participants to check their frequency of use and how many days per week they used the app. Throughout the study, participants in the CAMA group interacted with health care professionals via the app an average of 0.44 (SD 0.93) times, with 15 interactions observed among the 34 participants in the CAMA group. Among the participants who completed the study, all met the minimum requirement of using the app for at least 10 minutes daily, resulting in a 100% engagement rate among these participants throughout the study procedures. Of 34 participants, 2 (6%) in the CAMA group who failed to meet this requirement or withdrew from the study were not included in the final analysis.

Patients' Medication and Appointment Schedules

In the CAMA app, the *Manager* can register the patient's medication and medical appointment schedules through the website, or the patient can directly input these pieces of information into the app (Figure 3). An alert is triggered when the scheduled time arrives, and once the patient completes their medication or appointment, they can mark it as completed in the app.

Figure 3. Representative screens for the medication and appointment reminder and delayed question and answer system features of the Cancer Manager app. This figure was created by translating the original Korean text from the app's screenshots into English. (A) and (B) Medication and appointment reminder features: this feature provides alarms for individual medication and medical appointment schedules, enabling patients to mark them as completed and supporting treatment adherence and monitoring. (C) and (D) Delayed question and answer system feature: patients can post questions on the app regarding cancer-related issues about which they are curious. After the Manager gathers these questions, they are forwarded to a relevant professional (eg, a designated physician or psychologist) to provide answers accordingly, who sends the answers back to the Manager within a few days. Subsequently, the Manager organizes the information and responds to the patients through the app.



Delayed Q and A System

Patients can choose whether they would like to have face-to-face or remote consultations with the *Manager* and can use the app to ask questions to health care professionals, including attending physicians. When a patient asks a question, the health care professionals must review it and respond within a few days. This delayed Q and A system—not a live one—was introduced to alleviate the burden on health care professionals and ensure the sustainability of the app. Throughout the study, an average of 1.41 questions were asked per participant in the CAMA group through the delayed Q and A system (48 times among the 34 CAMA group participants).

Data Collection

Sociodemographic and Clinical Characteristics

This study collected data on the following: sociodemographic characteristics (eg, age, years of education, and monthly family income); clinical characteristics, including cancer stage at enrollment, whether initial or recurring cancer at enrollment, performance status, pathological characteristics, receptor expression status (eg, hormone receptors such as estrogen receptor or progesterone receptor and human epidermal growth factor receptor 2), and triple-negative phenotype; cancer treatment types being received (eg, chemotherapy, surgery, and radiation therapy); and whether the patient is in the cancer survivorship phase.

XSL•FO
RenderX

Psychological Characteristics and User Satisfaction

Participants' self-efficacy, QOL, mental adjustment to cancer, and psychological symptoms (ie, depression and anxiety) were assessed using validated scales at baseline and after the 12-week intervention. A satisfaction survey was also conducted among the participants who completed 12 weeks of CAMA use.

Measurements

The Eastern Cooperative Oncology Group scale [35] was used to assess patient performance status based on physical activity. It was measured by health care providers and comprised 5 grades ranging from 0 (no symptoms) to 4 (completely bedridden throughout the day).

To evaluate self-efficacy, we used the 10-item Korean version of the Cancer Survivor Self-Efficacy Scale (CS-SES-K) [36,37]. This self-reported tool comprises 2 subscales and 10 items, namely, managing health problems (5 items) and seeking help and support (5 items). The items were rated on a scale ranging from 1 (not at all confident) to 10 (totally confident), and scores were calculated by adding the scores for each item, with higher scores indicating greater self-efficacy. The CS-SES-K demonstrated high reliability and validity when applied to Korean cancer survivors [37], and the Cronbach α was 0.92 for the original tool [36] and the Korean version [37]. In our study, the internal consistency of the CS-SES-K was also high, with a Cronbach α of 0.93.

To assess mental adjustment to cancer, we used the 29-item Korean version of the Mini-Mental Adjustment to Cancer (K-Mini-MAC) Scale. This 29-item self-reported tool categorizes cognitive and behavioral responses to cancer into 4 types: helplessness or hopelessness (8 items), anxious preoccupation (8 items), cognitive avoidance (4 items), and positive attitude (fatalism and fighting spirit; 9 items) [38,39]. Items were rated on a scale ranging from 1 (definitely does not apply to me) to 4 (definitely applies to me), with higher scores indicating a more pronounced inclination to exhibit a specific response. This scale demonstrated acceptable reliability and validity when applied to Korean survivors of cancer [38]. The Cronbach α range was 0.62 to 0.88 for the original tool [39] and 0.72 to 0.86 for the Korean version [38]. In our study, the Cronbach α ranged from 0.74 to 0.89.

To evaluate QOL, we used the World Health Organization Quality of Life Brief Version (WHOQOL-BREF) [40,41]. This 26-item self-reported tool comprises 2 items measuring overall QOL and general health, along with 24 items divided into 4 subscales: physical health (7 items), psychological well-being (6 items), social relationships (3 items), and environmental aspects (8 items). Each item was scored on a scale ranging from 1 (not at all or very dissatisfied) to 5 (completely or very satisfied) with reverse-scoring items. Scoring was based on a specific formula, and higher scores on each subscale indicated a higher QOL in that respective domain [41]. This scale has demonstrated good validity, internal consistency, and test-retest reliability across various populations [40,41]. The Cronbach α range was 0.66 to 0.82 for the original tool [40] and 0.90 for the Korean version [41]. In this study, the Cronbach α range was 0.67 to 0.84, indicating high internal consistency.

The 9-item Patient Health Questionnaire-9 (PHQ-9) was used to assess depression and depressive symptom severity [42,43]. This self-reported tool was responded to on a 4-point Likert scale ranging from 0 (none) to 3 (almost every day), and each item assessed how often a responder had experienced depressive symptoms in the past 2 weeks. The scores were calculated by adding the scores for each item, with higher scores indicating more severe depression. A previous study conducted with Korean psychiatric outpatients showed an internal consistency reliability of 0.81 and a test-retest reliability of 0.89, with significant concurrent validity [43]. Moreover, the Cronbach α was 0.89 for the original tool [42] and 0.84 in this study.

To assess the degree of anxiety and anxiety symptoms, we used the 7-item Generalized Anxiety Disorder-7 (GAD-7) [44,45]. This self-reported tool was responded to on a 4-point Likert scale ranging from 0 (none) to 3 (almost every day), and each item assessed how often one had experienced anxiety symptoms in the past 2 weeks. The scores were calculated by adding the scores for each item, with higher scores indicating more severe anxiety. The GAD-7 demonstrated high validity and reliability for assessing anxiety in Korean psychiatric outpatients [45]. The Cronbach α was 0.92 for the original tool [44] and 0.88 for the Korean version [45]. In this study, Cronbach α was 0.90, demonstrating high internal consistency.

The Menopause Emotional Symptoms Questionnaire (MESQ) measures the emotional symptoms experienced by menopausal women. This self-reported scale comprises 13 items rated on a 5-point scale ranging from 0 (never experienced) to 4 (experienced very frequently). The MESQ is a valid and reliable tool for screening and measuring emotional symptoms in Korean menopausal women [46]. This scale was included because many patients with breast cancer had hormone-positive breast cancer and experienced menopause-like feelings while undergoing antihormonal therapy. In this study, the Cronbach α of 0.94 suggested a high internal consistency for the scale.

To assess satisfaction with CAMA use, we created a 17-item satisfaction questionnaire based on questionnaires used in previous studies [47]. Each self-reported item was rated on a 5-point Likert scale ranging from 0 (strongly disagree) to 5 (strongly agree). An additional open-ended question was included for responders to provide comments to app developers and the medical team.

Sample Size Calculation

The target sample size was calculated using G*Power (version 3.1.9.7; Faul, Universität Kiel). A Wilcoxon signed rank test was considered necessary to compare the impact of CAMA use in a group of survivors of cancer. At the time of this study, no previous research had used the CS-SES-K as the primary outcome. Considering these descriptions and based on the effect size in a previous study evaluating the impact of a digital intervention on the self-efficacy of patients with cancer [48], a significance level of 0.05, power of 0.85, and an effect size of 0.20 (medium), the sample size required was 60 participants. Given a 15% dropout rate, the goal was to recruit 72 participants, with 36 (50%) participants in each group, a number this study could reach.

XSL•FO
RenderX

Statistical Analyses

All statistical analyses were performed using SPSS for Windows (version 28.0; IBM Corp). The analysis was restricted to data from participants who had no missing values and completed both the pre- and postintervention assessments, which means that a per-protocol analysis approach was used. Independent samples 2-tailed *t* tests and chi-square tests were used to compare baseline characteristics between the CAMA and TAU groups. To evaluate the effectiveness of the CAMA app on primary and secondary outcomes, a mixed-effects ANOVA was conducted, with group (CAMA vs TAU) as a between-subject factor and time (baseline vs follow-up) as a within-subject factor. The primary outcome was self-efficacy (CS-SES-K scores), while the secondary outcomes included psychological well-being (WHOQOL-BREF scores), mental adjustment to cancer (K-Mini-MAC Scale scores), depressive symptoms (PHQ-9 scores), anxiety (GAD-7 scores), and menopausal emotional symptoms (MESQ scores). Pearson correlation analyses were conducted within the CAMA group to explore the relationships between changes in psychological outcomes. Moreover, we performed Pearson correlation analyses for scale scores that demonstrated a statistically significant improvement from baseline to follow-up.

Results

Participant Allocation and Baseline Characteristics

As mentioned in the *Methods* section, 80 patients with breast cancer underwent an eligibility assessment (Figure 1), but only 72 (90%) were enrolled, and 66 (82%) completed the study (completion rate: 66/72, 92%). In the CAMA group, discontinuation resulted from noncompliance with CAMA use in 2 participants; in the TAU group, 4 patients did not undergo any follow-up assessments (discontinuation rates: 2/36, 6% in the CAMA group and 4/36, 11% in the TAU group). Throughout the study period, no adverse events associated with the CAMA intervention were observed.

Results of the independent samples 2-tailed *t* test and chi-square test indicated no significant differences in the demographic factors between the CAMA and TAU groups at baseline (Table 1). Similarly, the CAMA and TAU groups exhibited no significant differences at baseline in the subscale and total scores for the CS-SES-K; the subscale scores for the WHOQOL-BREF and K-Mini-MAC Scale; and the total scores for the PHQ-9, GAD-7, and MESQ (Table 1).

XSL•FO
RenderX

Table 1. Sociodemographic and clinical characteristics at baseline.

Variables	CAMA[a] group (n=34)	TAU[b] group (n=32)	*t* test (*df*=64)	χ^2 (*df*)	*P* value
Sociodemographic variables					
Age (y), mean (SD)	55.2 (7.8)	52.3 (8.8)	1.41	—[c]	.16
Years of education, mean (SD)	13.5 (2.3)	14.0 (2.9)	–0.73	—	.47
Living with a partner, n (%)			—	0.51 (21)	.48
Yes	27 (79)	23 (72)			
No	7 (21)	9 (28)			
Monthly family income (USD), mean (SD)	3329.0 (3851.5)	4686.1 (6735.0)	–1.01	—	.32
Job, n (%)			—	0.06 (21)	.80
Yes	17 (50)	17 (53)			
No	17 (50)	15 (47)			
Religion, n (%)			—	2.22 (21)	.14
Yes	26 (76)	19 (59)			
No	8 (24)	13 (41)			
Clinical characteristics					
Duration from breast cancer diagnosis (mo), mean (SD)	40.7 (68.9)	38.3 (47.4)	0.17	—	.87
Performance status (ECOG[d]), n (%)			—	0.91 (21)	.34
0	10 (29)	13 (41)			
1	24 (71)	19 (59)			
Histology, n (%)			—	0.50 (21)	.48
Invasive ductal carcinoma	21 (62)	17 (53)			
Others	13 (38)	15 (47)			
ER[e], n (%)			—	0.91 (21)	.34
Negative	10 (29)	13 (41)			
Positive	24 (71)	19 (59)			
PR[f], n (%)			—	2.15 (21)	.14
Negative	13 (38)	18 (56)			
Positive	21 (62)	14 (44)			
HER2[g], n (%)			—	<0.001 (21)	.99
Negative	18 (53)	17 (53)			
Positive	16 (47)	15 (47)			
Triple-negative phenotype, n (%)			—	0.53 (21)	.47
Nontriple negative	28 (82)	24 (75)			
Triple negative	6 (18)	8 (25)			
Cancer stage at enrollment, n (%)			—	3.12 (24)	.54
0	4 (12)	7 (22)			
1	3 (9)	5 (16)			
2	7 (21)	3 (9)			
3	1 (3)	1 (3)			
4	19 (56)	16 (50)			
Cancer recurrence, n (%)			—	0.53 (21)	.47

XSL•FO
RenderX

Variables	CAMA[a] group (n=34)	TAU[b] group (n=32)	*t* test (*df*=64)	χ^2 (*df*)	*P* value
Initial	28 (82)	24 (75)			
Relapse	6 (18)	8 (25)			
Metastatic location, n (%)			—	1.39 (22)	.50
No metastasis	18 (53)	21 (66)			
Nonvisceral lesion	10 (29)	8 (25)			
Visceral lesion	6 (18)	3 (9)			
Treatment phase at enrollment, n (%)			—	3.94 (24)	.27
Diagnosis	0 (0)	0 (0)			
Surgery	1 (3)	5 (16)			
Chemotherapy	26 (76)	23 (72)			
Radiotherapy	3 (9)	1 (3)			
Survivor	4 (12)	3 (9)			

[a]CAMA: Cancer Manager.

[b]TAU: treatment as usual.

[c]Not applicable.

[d]ECOG: Eastern Cooperative Oncology Group.

[e]ER: estrogen receptor.

[f]PR: progesterone receptor.

[g]HER2: human epidermal growth factor receptor 2.

Comparison of Changes in Clinical Scale Scores Between the CAMA and TAU Groups

Throughout the intervention period, the CAMA group (vs the TAU group) demonstrated observable improvements in the following scale scores: the seeking help and support subscale of CS-SES-K ($F_{1,64}$=5.09; *P*=.03), the psychological well-being subscale of WHOQOL-BREF ($F_{1,64}$=5.48; *P*=.02), the anxious preoccupation subscale ($F_{1,64}$=5.49; *P*=.02) and the positive attitude subscale ($F_{1,64}$=5.44; *P*=.02; Table 2) of the K-Mini-MAC Scale, PHQ-9 ($F_{1,64}$=4.83; *P*=.03), GAD-7 ($F_{1,64}$=5.48; *P*=.02), and MESQ ($F_{1,64}$=4.30; *P*=.04; Table 3). No significant differences were observed between the 2 groups in total or subscale scores for the other scales.

XSL•FO
RenderX

Table 2. Differences in the mean scores for the self-efficacy and psychological scales at baseline.

Psychological scales and variables	CAMA[a] group (n=34), mean (SD)	TAU[b] group (n=32), mean (SD)	*t* test (*df*=64)	*P* value
CS-SES-K[c]				
Managing health problems	33.4 (10.2)	32.7 (11.8)	0.23	.82
Seeking help and support	38.2 (7.8)	37.2 (9.8)	0.47	.64
Total	71.6 (16.9)	69.9 (20.2)	0.36	.72
WHOQOL-BREF[d]				
Physical health	12.6 (3.1)	13.4 (2.0)	–1.26	.21
Psychological well-being	13.1 (3.7)	14.0 (3.4)	–1.04	.30
Social relationships	13.9 (2.7)	14.0 (2.6)	–0.09	.93
Environmental aspects	13.4 (2.8)	14.6 (2.8)	–1.79	.08
K-Mini-MAC[e] Scale				
Helplessness or hopelessness	13.4 (4.6)	13.2 (5.4)	0.18	.86
Anxious preoccupation	22.2 (5.0)	20.4 (5.8)	1.36	.18
Cognitive avoidance	11.1 (2.6)	10.1 (2.9)	1.61	.11
Positive attitude (fatalism and fighting spirit)	27.3 (4.7)	27.2 (7.2)	0.09	.93
PHQ-9[f]	8.3 (4.3)	7.0 (5.4)	1.09	.28
GAD-7[g]	6.7 (4.4)	4.7 (4.0)	1.91	.06
MESQ[h]	23.7 (10.4)	19.5 (9.8)	1.69	.10

[a]CAMA: Cancer Manager.

[b]TAU: treatment as usual.

[c]CS-SES-K: Korean version of the Cancer Survivor Self-Efficacy Scale.

[d]WHOQOL-BREF: World Health Organization Quality of Life Brief Version.

[e]K-Mini-MAC: Korean version of the Mini-Mental Adjustment to Cancer.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]GAD-7: Generalized Anxiety Disorder-7.

[h]MESQ: Menopause Emotional Symptoms Scale.

XSL•FO
RenderX

Table 3. Comparison of score changes in the self-efficacy and psychological scales between the Cancer Manager (CAMA) and treatment as usual (TAU) groups.

Variables	CAMA group (n=34), mean (SD)		TAU group (n=32), mean (SD)		Group × time effect (mixed ANOVA)		
	Baseline	Follow-up	Baseline	Follow-up	*F* test (*df*=1,64)	*P* value	Partial η^2
CS-SES-K[a]							
Managing health problems	33.4 (10.2)	33.1 (9.9)	32.7 (11.8)	29.8 (11.0)	1.15	.29	0.018
Seeking help and support	38.2 (7.8)	39.9 (6.9)	37.2 (9.8)	34.5 (10.6)	5.09	.03	0.074
Total	71.6 (16.9)	72.9 (15.2)	69.9 (20.2)	64.3 (20.2)	3.23	.08	0.048
WHOQOL-BREF[b]							
Physical health	12.6 (3.1)	12.0 (2.8)	13.4 (2.0)	12.6 (3.2)	0.03	.86	<0.001
Psychological well-being	13.1 (3.7)	13.8 (2.7)	14.0 (3.4)	12.7 (3.6)	5.48	.02	0.079
Social relationships	13.9 (2.7)	13.3 (2.4)	14.0 (2.6)	12.6 (3.5)	1.56	.22	0.024
Environmental aspects	13.4 (2.8)	13.8 (2.4)	14.6 (2.8)	14.0 (2.5)	2.32	.13	0.035
K-Mini-MAC[c] Scale							
Helplessness or hopelessness	13.4 (4.6)	13.3 (4.6)	13.2 (5.4)	13.5 (4.6)	0.17	.68	0.003
Anxious preoccupation	22.2 (5.0)	19.2 (4.8)	20.4 (5.8)	19.9 (5.3)	5.49	.02	0.079
Cognitive avoidance	11.1 (2.6)	11.1 (2.6)	10.1 (2.9)	10.2 (2.7)	0.12	.73	0.002
Positive attitude (fatalism and fighting spirit)	27.3 (4.7)	27.8 (4.3)	27.2 (7.2)	25.3 (5.3)	5.44	.02	0.078
PHQ-9[d]	8.3 (4.3)	6.6 (4.0)	7.0 (5.4)	8.0 (6.4)	4.83	.03	0.070
GAD-7[e]	6.7 (4.4)	5.5 (4.4)	4.7 (4.0)	5.7 (5.0)	5.48	.02	0.079
MESQ[f]	23.7 (10.4)	17.4 (9.4)	19.5 (9.8)	19.3 (12.7)	4.30	.04	0.063

[a]CS-SES-K: Korean version of the Cancer Survivor Self-Efficacy Scale.

[b]WHOQOL-BREF: World Health Organization Quality of Life Brief Version.

[c]K-Mini-MAC: Korean version of the Mini-Mental Adjustment to Cancer.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]MESQ: Menopause Emotional Symptoms Questionnaire.

Correlations Among Score Changes in Self-Efficacy and Psychological Scales Within the CAMA Group

Changes in the scores for the anxious preoccupation subscale of the K-Mini-MAC Scale were positively correlated with those for the PHQ-9 (r=0.46; P=.007) and GAD-7 (r=0.41; P=.02) scores and negatively correlated with those for the positive attitude subscale of the K-Mini-MAC Scale (r=–0.36; P=.04; Table 4; Figure 4). Changes in the scores for the PHQ-9 were positively correlated with those for the GAD-7 (r=0.66; P<.001) and MESQ (r=0.35; P=.04).

XSL•FO
RenderX

Table 4. Correlations among score changes for clinical scales within the Cancer Manager group (N=34).

Variables	Δ[a] CS-SES-K-seeking[b]	Δ WHOQOL-BREF-psy[c]	Δ K-Mini-MAC-AP[d]	Δ K-Mini-MAC-PA[e]	Δ PHQ-9[f]	Δ GAD-7[g]	Δ MESQ[h]
Δ CS-SES-K-seeking							
r	1	0.195	0.026	0.042	–0.214	–0.022	–0.111
P value	—[i]	.27	.89	.81	.22	.90	.53
Δ WHOQOL-BREF-psy							
R	0.195	1	–0.318	–0.116	–0.039	–0.166	–0.053
P value	.27	—	.07	.51	.83	.35	.77
Δ K-Mini-MAC-AP Scale							
r	0.026	–0.318	1	–0.363	0.456	0.406	0.261
P value	.89	.07	—	.04	.007	.02	.14
Δ K-Mini-MAC-PA Scale							
r	0.042	–0.116	–0.363	1	–0.111	–0.094	0.008
P value	.81	.51	.04	—	.53	.60	.96
Δ PHQ-9							
r	–0.214	–0.039	0.456	–0.111	1	0.656	0.354*
P value	.22	.83	.007	.53	—	<.001	.04
Δ GAD-7							
r	–0.022	–0.166	0.406	–0.094	0.656	1	0.240
P value	.90	.35	.02	.60	<.001	—	.17
Δ MESQ							
r	–0.111	–0.053	0.261	0.008	0.354	0.240	1
P value	.53	.77	.14	.96	.04	.17	—

[a]Δ: the differences between baseline and follow-up assessments.

[b]CS-SES-K-seeking: seeking help and support subscale of the Korean version of the Cancer Survivor Self-Efficacy Scale.

[c]WHOQOL-BREF-psy: psychological health subscale of the World Health Organization Quality of Life Brief Version.

[d]K-Mini-MAC-AP: anxious preoccupation subscale of the Korean version of the Mini-Mental Adjustment to Cancer.

[e]K-Mini-MAC-PA: positive attitude (fatalism and fighting spirit) subscale of the Korean version of the Mini-Mental Adjustment to Cancer.

[f]PHQ-9: Patient Health Questionnaire-9.

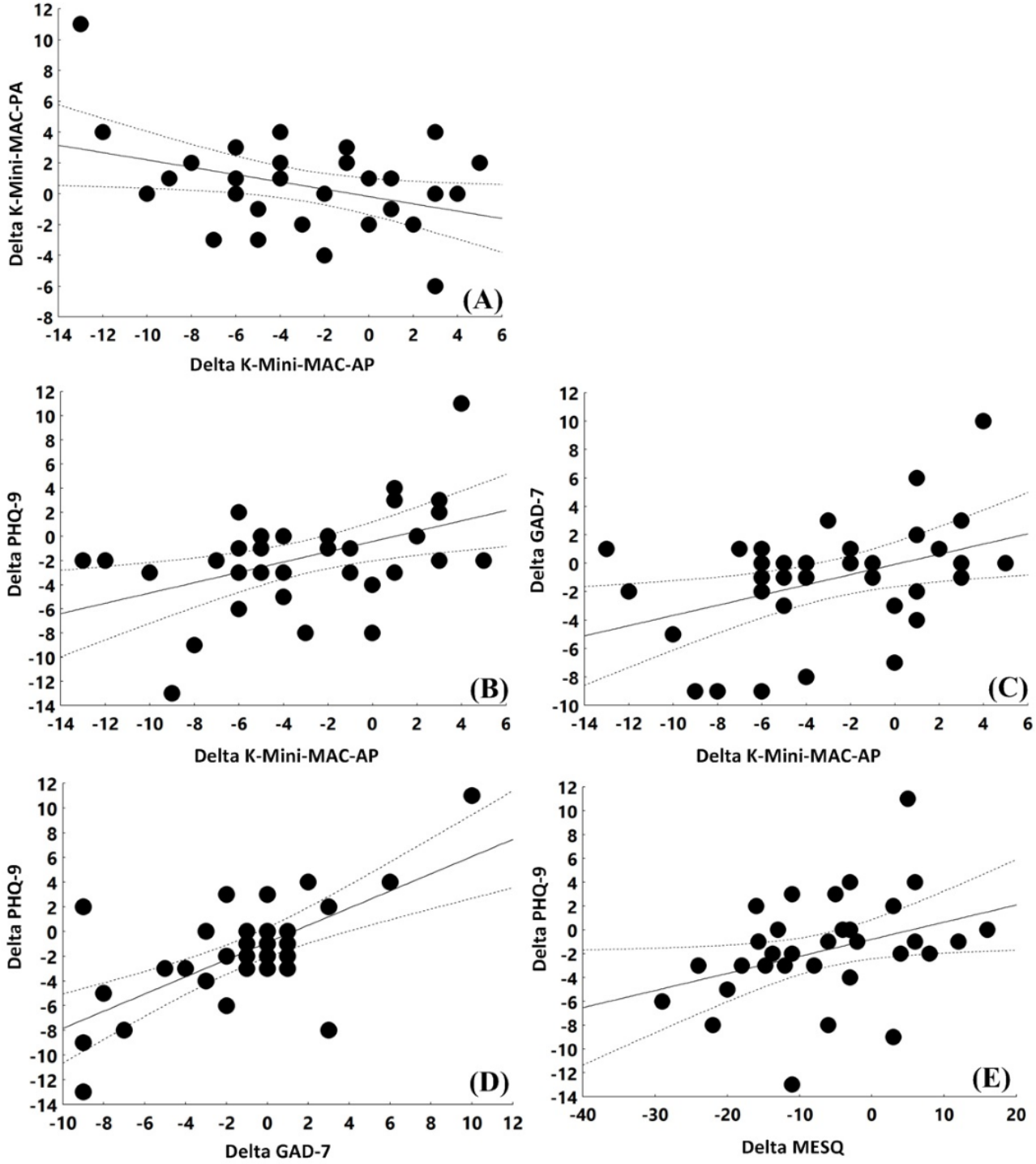
[g]GAD-7: Generalized Anxiety Disorder-7.

[h]MESQ: Menopause Emotional Symptoms Questionnaire.

[i]Not applicable.

XSL•FO
RenderX

Figure 4. Correlations between score changes for the psychological scales. (A) Changes in the scores for the anxious preoccupation subscale of the Korean version of the Mini-Mental Adjustment to Cancer Scale (K-Mini-MAC-AP) versus those for the positive attitude subscale of the Korean version of the Mini-Mental Adjustment to Cancer Scale (K-Mini-Mac-PA); the statistical results are r=–0.36 and *P*=.04. (B) Changes in the scores for the K-Mini-MAC-AP versus those for the Patient Health Questionnaire-9 (PHQ-9); the statistical results are r=0.46 and *P*=.007. (C) Changes in the scores for the K-Mini-MAC-AP versus those for the Generalized Anxiety Disorder-7 (GAD-7); the statistical results are r=0.41 and *P*=.02. (D) Changes in the scores for PHQ-9 versus those for the GAD-7; the statistical results are r=0.66 and *P*<.001. (E) Changes in the scores for PHQ-9 versus those for the Menopause Emotional Symptoms Questionnaire (MESQ); the statistical results are r=0.35 and *P*=.04.



User Satisfaction

Regarding the mean scores for all 17 items of the satisfaction survey, they were above an average of 4 points (Table 5), and 26 (76%) out of 34 participants responded with *strongly agree* or *agree* to the question about satisfaction with CAMA use. More than 90% of the participants also indicated the following: CAMA use (32/34, 94%) and the information it provides (33/34, 97%) are easy to understand, CAMA benefits physical health (31/34, 91%) and psychological and emotional health (31/34, 91%) management, the intention to use digital health care apps such as CAMA again (31/34, 91%).

XSL•FO
RenderX

Table 5. Cancer Manager (CAMA) user satisfaction survey results (N=34).

Questions	Values, mean (SD)	Values, n (%)[a]
1. Overall, I am satisfied with the use of CAMA.	4.2 (0.9)	26 (76)
2. It is easy to learn how to use CAMA.	4.5 (0.6)	32 (94)
3. The information content introduced by CAMA is helpful.	4.5 (0.7)	30 (88)
4. The information content of CAMA is easy to understand.	4.7 (0.5)	33 (97)
5. The information content of CAMA is interesting.	4.4 (0.8)	30 (88)
6. CAMA helps me manage my physical health.	4.5 (0.7)	31 (91)
7. CAMA helps me understand my cancer treatment process.	4.3 (0.8)	28 (82)
8. CAMA helps me know symptoms caused by cancer.	4.5 (0.7)	29 (85)
9. CAMA helps me control symptoms caused by cancer.	4.4 (0.7)	29 (85)
10. CAMA helps me know side effects of treatment.	4.3 (0.8)	28 (82)
11. CAMA helps me cope with side effects of treatment.	4.3 (0.8)	27 (79)
12. CAMA helps me with my psychological and emotional health.	4.4 (0.7)	31 (91)
13. CAMA helps me understand psychological difficulties related to cancer.	4.4 (0.7)	29 (85)
14. CAMA helps me control psychological difficulties associated with cancer.	4.3 (0.7)	29 (85)
15. CAMA helps me manage my daily life.	4.4 (0.7)	29 (85)
16. I want to use a digital health care app like CAMA again.	4.5 (0.7)	31 (91)
17. I want to introduce CAMA to other patients with breast cancer.	4.5 (0.7)	30 (88)

[a]Percentage of participants who responded with *strongly agree* or *agree* to each question.

As mentioned in the Methods section, we also provided participants with an open-ended item where they could provide feedback on satisfaction and areas for improvement regarding the app, which yielded the following insights. Participants were reportedly satisfied with the ease with which they could access reliable information through the app and appreciated its convenience. They pointed out issues with internet searches, as some answers found through the search engines stemmed from nonexperts and introduced bias. They praised the CAMA app for presenting the opinions of medical professionals, which helped build their trust. Participants emphasized the value of the CAMA app through its provision of trustworthy information and support, as these functions helped reduce their impulsive information-seeking behaviors driven by fear. They mentioned that the collaborative efforts of specialists in hematology, breast surgery, and psychiatry afforded them beneficial, integrated therapeutic information. Some suggested improvements included incorporating patient communication features and consistently updating the platform with more comprehensive and in-depth information.

Discussion

Principal Findings

The CAMA app was developed to provide evidence-based, expert-authored, personalized digital information for patients with breast cancer according to their breast cancer type and cancer journey phase. In this study, we explored the potential benefits of the CAMA app, observing patterns that suggest possible efficacy in improving self-efficacy, psychological well-being, and mental adjustment to cancer, although the nonrandomized design limited the robustness of these findings and necessitated cautious interpretation. Compared to the TAU group, those in the CAMA group showed the following: significant increases in their self-efficacy for seeking help and support, psychological well-being, and positive attitude toward cancer; significant reductions in their anxious preoccupation with cancer; and decreased levels of depression, anxiety, and menopausal emotional symptoms. Furthermore, score changes for anxious preoccupation with cancer were positively correlated with those in depression and anxiety and negatively correlated with score changes in positive attitude toward cancer. Notably, score changes in depression were positively correlated with anxiety and menopausal emotional symptoms. The feasibility of the app for the targeted population was generally considered good. These findings offer insights into the app's positive impact, trust-building outcomes, and opportunities for enhancement (eg, introduction of communication tools and ongoing content enrichment). However, these findings are preliminary and should be validated through future randomized controlled trials.

The CAMA app developed and validated in this study offers a unique and comprehensive approach to breast cancer care, distinguishing it from existing mHealth apps. Unlike many current apps that focus on singular aspects such as symptom management, lifestyle changes, and psychological support [15,49], the CAMA app integrates a wide range of functionalities into a single platform. This includes adherence management, tailored medical information, psychological modules, and logistical features such as scheduling and reminders, addressing

XSL•FO
RenderX

the multidisciplinary needs of patients with breast cancer. Furthermore, the CAMA app provides personalized information that dynamically adapts to the user's breast cancer subtype, ongoing or planned treatments (eg, chemotherapy, surgery, and radiation therapy), and recovery stage. This level of customization ensures that patients receive relevant and actionable content tailored to their specific circumstances, overcoming the limitations of generic mHealth apps, which often fail to address unique patient needs [20,49,50]. In addition, a key innovation of the CAMA app is its *My Care Track* feature, which delivers daily personalized content. This contrasts with many existing apps that require users to actively search for information, making the process time consuming and less engaging [50,51]. By offering concise, context-specific updates, the CAMA app ensures timely support while enhancing user engagement [34,52]. The app's development was also informed by a multidisciplinary team of health care professionals, including oncologists, surgeons, psychiatrists, and psychologists. Research has highlighted the importance of such collaborative design in creating effective and credible mHealth interventions [53,54]. In addition, the CAMA app addresses usability barriers often cited as challenges in mHealth adoption by simplifying complex medical and psychological information into user-friendly formats and providing interactive tools, such as reminders and visual aids [34,51]. By integrating these features, the CAMA app not only meets the diverse needs of patients with breast cancer but also addresses critical gaps in existing mHealth solutions, particularly the lack of comprehensive, personalized, and evidence-based approaches. Its holistic and patient-centric design demonstrates significant potential to improve both the physical and psychological outcomes of individuals navigating the complexities of breast cancer care [21,54].

In this study, the CAMA group demonstrated significant improvements in self-efficacy for seeking help and support compared to the TAU group after the intervention. The items for this variable included queries about confidence in accessing information, reaching out to individuals for assistance and support, independently handling problems, contacting physicians, and receiving help to address issues [37]. Considering the contents in the items, one can infer that using the CAMA app led participants to experience a heightened sense of confidence in accessing the information they needed whenever necessary, identifying human resources for support (eg, medical professionals), and feeling empowered to tackle challenges independently or with others' assistance. In a similar context and with another sample of patients with breast cancer, a study applied a web-based cancer management system and found that the score for self-efficacy in coping with cancer-related stress increased significantly over time after system implementation [55]. Moreover, researchers have reported on the self-efficacy–enhancing outcomes of web- or app-based interventions for survivors of breast cancer who completed primary treatment and were undergoing chemotherapy [56,57]. A systematic review and meta-analysis revealed that eHealth-based self-management interventions effectively addressed fatigue and enhanced self-efficacy in adult patients with cancer [58]. Thus, our findings corroborate the crucial role observed in previous research of digital interventions in improving self-efficacy among patients with cancer and survivors of cancer.

The hypothesis established before the study onset was that the scores for the CS-SES-K and its 2 subscales would increase significantly after the intervention. However, the total score and the score for the subscale of managing health problems did not increase significantly. This may be explained by a significant deterioration, possibly caused by factors external to the intervention, in the participants' objective health status upon the completion of the 12-week intervention compared to the health status at enrollment. Specifically, of the 66 total study participants, some (26/34, 76%) of CAMA group and some (23/32, 72%) were either approaching or currently undergoing chemotherapy at the time of enrollment. At this point, one may infer that the health of the 8 participants of CAMA group and 5 participants of TAU group who were exposed to the side effects of these concoctions of treatments (eg, nausea, fatigue, appetite loss, diarrhea, constipation, skin and nail deformity, and hair loss, particularly during chemotherapy) worsened significantly throughout the study period, which may have influenced their assessments after the intervention. In particular, the managing health problems subscale contains items that query the respondent's confidence in managing fatigue, physical discomfort or pain, emotional distress, other symptoms, health problems, and performing various tasks and activities. Thus, the deterioration in participants' physical conditions during and because of cancer treatment may have hindered the potential improvements the intervention could have yielded for the managing health problems subscale. A pilot study showed that an intervention encompassing web-based self-management exercises and diet intervention for survivors of breast cancer improved self-efficacy [59]; however, various studies showed results similar to ours (ie, self-efficacy did not improve). In a previous study conducted in Korea, the level of self-efficacy improved immediately after a 7-week intervention involving education and skill training through telephone counseling for patients with breast cancer. Nonetheless, this effect did not persist and was not observed at the 20-week follow-up [60]. For survivors of incurable cancer using eHealth self-management apps, there was no general enhancement in self-efficacy [61]. Similarly, survivors of cancer experiencing fatigue who were exposed to a web-based intervention aimed at supporting self-management of cancer-related fatigue showed no changes in self-efficacy [49].

The intervention of this study significantly reduced anxious preoccupation and increased positive attitude and mental adjustment to cancer. Mental adjustment to cancer has been defined as an individual's cognitive and behavioral reactions following a cancer diagnosis [52]. Four types of mental adjustment to cancer have been suggested: (1) helplessness or hopelessness, namely, the tendency to feel overwhelmed and resigned because of cancer and experience feelings of powerlessness and hopelessness; (2) anxious preoccupation, which describes consistent cancer-related anxiety that often leads to the experience of high psychological distress; (3) cognitive avoidance, referring to a tendency to avoid thinking about the condition and its implications and to use distractions as a coping strategy; and (4) positive attitude (fatalism and

XSL•FO
RenderX

fighting spirit), describing the adoption of an optimistic perspective toward cancer and the process of seeking growth and meaning in the experience. Even survivors of cancer diagnosed 5 to 6 years ago have been observed to exhibit these specific coping patterns related to their cancer experience [62]. Thus, one's mental adjustment to cancer encompasses the evaluation of cancer implications (ie, one's understanding of its effects), subsequent responses (ie, thoughts and actions aimed at minimizing the threat), and involuntary emotional reactions to cancer [52]. Having a fighting spirit seemed to benefit psychological adjustment [39,50,54,63], whereas helplessness or hopelessness and anxious preoccupation consistently correlated with elevated levels of adjustment difficulties or low levels of health-related QOL [51,62]. Overall, the decrease in anxious preoccupation scores and the increase in positive attitude scores after using CAMA suggest potential benefits of the app, aligning with findings from previous research.

Furthermore, a study involving an early, structured psychoeducational group intervention for patients with breast cancer showed a significant reduction in anxious preoccupation after the intervention [64]. Another study on a cognitive existential group therapy intervention for women with early-stage breast cancer demonstrated a trend-level reduction in anxious preoccupation [65]. In a Korean study focused on the effects of an integrated educational program on self-efficacy and coping strategies immediately after a breast cancer diagnosis, significant reductions were observed in anxious preoccupation [66]. However, there is a lack of research on positive attitudes and related improvements following interventions for patients with cancer. The *fatalism* referred to in the subdimension of the K-Mini-MAC Scale has been reported to be more associated with active engagement in spirituality and religious practices rather than passive acceptance of the inevitable [50]. This positive and active form of fatalism is associated with a fighting spirit, a sense of personal control, and patient determination, in which patients take active steps to cure their illness or ameliorate its effects [50]. In this study, changes in anxious preoccupation were negatively correlated with those in positive attitudes. This allows us to assume cautiously that the app reduces anxiety by affording users appropriate psychoeducation along with credible treatment suggestions, which, in turn, increases positive attitudes, including fighting spirit.

Moreover, changes in anxious preoccupation were positively correlated with changes in depression and anxiety. These findings are consistent with evidence depicting fighting spirit as having a consistent correlation with lower levels of anxiety and depression and helplessness, hopelessness, or anxious preoccupation as being more likely associated with depression and anxiety [67,68]. In addition, higher depression levels were correlated with higher anxious preoccupation, helplessness, or hopelessness and lower fighting spirit levels in past literature [69-71].

As discussed in the previous paragraphs, various studies have been devoted to exploring the effect of various types of interventions on the anxious preoccupation of patients with cancer, and they have observed a significant decrease in such preoccupation. However, they seem to have not thoroughly pondered about or explained why there was such a reduction in this single dimension and no reduction in other dimensions of mental adjustment to cancer. Upon careful consideration, we speculated that mental adjustment and coping strategy changes may necessitate cognitive or behavioral shifts, which are more likely to occur over a longer period, requiring a more extended follow-up procedure to be observed. These findings indicate the potential need to complement stress management content intervention strategies with direct intervention strategies, such as coping skills training. For instance, a psychological education program, which included health education, stress management, and coping skills training, was conducted with patients with melanoma immediately after diagnosis, yielding positive results (ie, reduction in passive resignation coping strategies) in the experimental group at the 3-month follow-up [72].

The CAMA intervention also significantly reduced depression, anxiety, and menopausal emotional symptom levels, with changes in depression levels being significantly correlated with those in anxiety and menopausal emotional symptoms. Individuals diagnosed with breast cancer often face elevated psychological challenges (eg, feelings of anxiety and depression) that can adversely influence their QOL and chances of survival [73,74]. Accordingly, scholars have recognized the crucial role of psychological support in breast cancer treatment and established various related interventions [75], including psychoeducational interventions that encompass training to handle emotions, receive information, express feelings, discuss concerns, solve problems, receive support, and develop coping strategies [76]. A systematic review revealed that psychoeducational interventions significantly reduced anxiety and depressive symptoms among women diagnosed with breast cancer [75]. Another systematic review and meta-analysis study focused on internet-based psychoeducational interventions further highlighted a significant effect on decreasing depression and fatigue in patients with cancer [77]. Following these pieces of evidence, we have incorporated breast cancer–related medical and surgical treatment information and an equivalent amount of psychoeducational information into the CAMA app's content, including the following: introductory information on the psychological responses patients may experience during different treatment stages, coping strategies, adaptive or maladaptive symptoms, behavioral therapies, potential psychological reactions of patients and caregivers, and psychological information caregivers should be aware of. These different pieces of information were included based on the belief that they would exert alleviating effects on patients' depression and anxiety.

In this study, the CAMA group (vs the TAU group) demonstrated significant improvements in psychological well-being. Previous studies' results on the impact of psychoeducational interventions on the QOL of patients with cancer are consistent with those of this study. For instance, a meta-analysis demonstrated that internet-based psychoeducational interventions significantly decreased depression and fatigue and the insufficiency of evidence to support their effects on psychological distress and QOL in individuals with cancer [77]. A recent systematic review of studies conducted with survivors of breast cancer suggested

XSL•FO
RenderX

that psychoeducation significantly improved QOL and alleviated anxiety symptoms but not depression [73]. Our study did not show significant results for any QOL subscale other than psychological well-being; we speculate that the nature of the items within the WHOQOL-BREF may have contributed to these results or their lack thereof. For instance, items within the physical health subscale (eg, energy levels, physical functioning, sleep, and occupational concerns) encompass variables heavily influenced by the medical cancer treatment process (ie, chemotherapy, surgery, or radiation therapy). Thus, similar to the reasoning presented earlier for the nonsignificant changes regarding self-efficacy (as measured by the CS-SES-K), the deterioration in participants' actual physical conditions caused by cancer treatment may have hindered the intervention's influence on QOL.

The findings of this study can be interpreted through the lens of the CCM and the HBM, both of which were key to the development of the CAMA app. The CCM emphasizes the importance of providing ongoing, coordinated, and patient-centered care, particularly when managing chronic diseases such as breast cancer [24,25]. In line with this model, the CAMA app facilitated continuous support and empowered patients to take an active role in their health management by integrating medical, psychological, and social resources. This was reflected in the significant improvements observed in self-efficacy, particularly in seeking help and support, and in the reduction of psychological distress among users. Analyzing these findings from the perspective of the HBM further enriches the discussion, as this model focuses on the cognitive and emotional factors that influence health behaviors [29]. The CAMA app was designed to enhance patients' perceived control over their health by providing reliable information, managing treatment schedules, and offering psychological support. At this point, considering the reduction in anxiety, depression, and menopausal emotional symptoms observed in the CAMA group, we suggest that our findings align with the HBM's premise, namely, that addressing perceived barriers to and emphasizing the benefits of proactive health behaviors can lead to better psychological outcomes [31,78,79]. The app's success in fostering a positive attitude and reducing anxious preoccupation among patients with breast cancer suggests that it effectively addressed users' concerns and motivated them to engage in healthier behaviors. These suggestions align with the principles of both the CCM and HBM.

Limitations

This study had some limitations. First, it used a nonrandomized design, chosen for pragmatic reasons such as feasibility and logistical constraints. While this facilitated implementation in a real-world clinical setting, it also introduced limitations, including potential selection bias and reduced generalizability. Baseline comparisons between the intervention and control groups showed no significant differences, reducing the need for adjustments in analyses. However, the intervention group included participants willing to use the app and with access to compatible smartphones, which may have influenced the outcomes. Thus, differences between groups could partly reflect these pre-existing characteristics rather than the app intervention alone. As the lack of randomization prevents the exclusion of confounding variables, the findings should be interpreted as exploratory with caution. Future randomized controlled trials are necessary to confirm these preliminary results. Second, the sample size was small, and the follow-up had a short duration. However, considering the nature of this feasibility study, its aim of rapidly assessing the intervention's usability, and providing empirical evidence for better-designed future research, these limitations can be deemed acceptable. Furthermore, the power calculations demonstrated that the sample size, albeit small, was sufficient for effect validation. Third, this study included participants with significant divergences regarding the time since breast cancer diagnosis (ie, ranging from a few months to 20 years), potentially limiting the findings owing to variations in cancer experiences. Cancer-related self-efficacy and adjustment styles tend to change throughout the cancer journey as they are influenced by condition controllability [80]. Thus, greater sample homogeneity could lead to differential results, implying that future long-term randomized controlled trials must use larger samples and conduct subgroup analyses based on patients' medical characteristics, cancer journey phase, and treatment types. Fourth, this study did not directly assess physical health and health care use–related indicators. Future researchers should consider direct physical health indicators and health care use–related indicators, such as the following: blood parameters, adverse events and postoperative complications, treatment adherence (medication and outpatient visits), screening adherence (cancer-related and routine checkups), health care costs, or unplanned emergency department and outpatient visits. Finally, app effectiveness was assessed using only subjective measures. Future researchers should collect and analyze data on variables such as medication adherence, compliance with outpatient visits, and objectively measurable health indicators (eg, heart rate variability and blood parameters) to evaluate app efficacy more comprehensively.

Conclusions

The mobile app for assisting in the self-management of patients with breast cancer, CAMA, was deemed feasible and shows potential for enhancing the patients' self-efficacy regarding seeking help and support, positive attitude toward cancer, and psychological well-being. In addition, its use may help reduce their anxious preoccupation with cancer, depressive mood, anxiety, and menopausal emotional symptoms. However, the nonrandomized study design highlights the need for further randomized controlled trials to confirm these preliminary findings.

Acknowledgments

This research was supported by a grant from the Korean Cancer Survivors Healthcare R&D Project through the National Cancer Center, funded by the Ministry of Health and Welfare, Republic of Korea (RS-2023-CC139876).

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) checklist.
[DOCX File , 40 KB - mhealth_v13i1e63989_app1.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Dinapoli L, Colloca G, Di Capua B, Valentini V. Psychological aspects to consider in breast cancer diagnosis and treatment. Curr Oncol Rep 2021 Mar 11;23(3):38 [FREE Full text] [doi: 10.1007/s11912-021-01049-3] [Medline: 33709235]
3. Parker PD, McSweeney JC, McQueen A, Jin J, Bryant-Smith G, Henry-Tillman R. "I'm not fighting anymore so what do I do now?" Young women's challenges while transitioning out of active breast cancer treatment and into survivorship. Inquiry 2023;60:469580231164230 [FREE Full text] [doi: 10.1177/00469580231164230] [Medline: 37039391]
4. Johnson KS, Conant EF, Soo MS. Molecular subtypes of breast cancer: a review for breast radiologists. J Breast Imaging 2021 Jan 26;3(1):12-24. [doi: 10.1093/jbi/wbaa110] [Medline: 38424845]
5. Shao J, Rodrigues M, Corter AL, Baxter NN. Multidisciplinary care of breast cancer patients: a scoping review of multidisciplinary styles, processes, and outcomes. Curr Oncol 2019 Jun;26(3):e385-e397 [FREE Full text] [doi: 10.3747/co.26.4713] [Medline: 31285683]
6. Stovall E, Greenfield S, Hewitt M. From cancer patient to cancer survivor: lost in transition. National Academy of Medicine. URL: https://nap.nationalacademies.org/catalog/11468/from-cancer-patient-to-cancer-survivor-lost-in-transition [accessed 2024-04-29]
7. Gast KC, Allen SV, Ruddy KJ, Haddad TC. Novel approaches to support breast cancer survivorship care models. Breast 2017 Dec;36:1-13. [doi: 10.1016/j.breast.2017.08.004] [Medline: 28843976]
8. Burg MA, Adorno G, Lopez ED, Loerzel V, Stein K, Wallace C, et al. Current unmet needs of cancer survivors: analysis of open-ended responses to the American Cancer Society Study of Cancer Survivors II. Cancer 2015 Feb 15;121(4):623-630 [FREE Full text] [doi: 10.1002/cncr.28951] [Medline: 25581252]
9. Fan R, Wang L, Bu X, Wang W, Zhu J. Unmet supportive care needs of breast cancer survivors: a systematic scoping review. BMC Cancer 2023 Jun 26;23(1):587 [FREE Full text] [doi: 10.1186/s12885-023-11087-8] [Medline: 37365504]
10. Ellegaard MB, Grau C, Zachariae R, Bonde Jensen A. Fear of cancer recurrence and unmet needs among breast cancer survivors in the first five years. A cross-sectional study. Acta Oncol 2017 Feb;56(2):314-320. [doi: 10.1080/0284186X.2016.1268714] [Medline: 28093034]
11. Hansen DG, Larsen PV, Holm LV, Rottmann N, Bergholdt SH, Søndergaard J. Association between unmet needs and quality of life of cancer patients: a population-based study. Acta Oncol 2013 Feb;52(2):391-399. [doi: 10.3109/0284186X.2012.742204] [Medline: 23244672]
12. Cheng KK, Wong WH, Koh C. Unmet needs mediate the relationship between symptoms and quality of life in breast cancer survivors. Support Care Cancer 2016 May;24(5):2025-2033. [doi: 10.1007/s00520-015-2994-0] [Medline: 26530229]
13. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
14. Charbonneau DH, Hightower S, Katz A, Zhang K, Abrams J, Senft N, et al. Smartphone apps for cancer: a content analysis of the digital health marketplace. Digit Health 2020;6:2055207620905413 [FREE Full text] [doi: 10.1177/2055207620905413] [Medline: 32110428]
15. Cai T, Huang Y, Zhang Y, Lu Z, Huang Q, Yuan C. Mobile health applications for the care of patients with breast cancer: a scoping review. Int J Nurs Sci 2021 Oct 10;8(4):470-476 [FREE Full text] [doi: 10.1016/j.ijnss.2021.07.003] [Medline: 34631997]
16. Singleton AC, Raeside R, Hyun KK, Partridge SR, Di Tanna GL, Hafiz N, et al. Electronic health interventions for patients with breast cancer: systematic review and meta-analyses. J Clin Oncol 2022 Jul 10;40(20):2257-2270 [FREE Full text] [doi: 10.1200/JCO.21.01171] [Medline: 35500200]
17. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-tested mobile apps for breast cancer care: systematic review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
18. Cruz FO, Vilela RA, Ferreira EB, Melo NS, Reis PE. Evidence on the use of mobile apps during the treatment of breast cancer: systematic review. JMIR Mhealth Uhealth 2019 Aug 27;7(8):e13245 [FREE Full text] [doi: 10.2196/13245] [Medline: 31456578]
19. Kalke K, Ginossar T, Bentley JM, Carver H, Shah SF, Kinney AY. Use of evidence-based best practices and behavior change techniques in breast cancer apps: systematic analysis. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e14082 [FREE Full text] [doi: 10.2196/14082] [Medline: 32012084]

XSL•FO
RenderX

20. Kirsch EP, Kunte SA, Wu KA, Kaplan S, Hwang ES, Plichta JK, et al. Digital health platforms for breast cancer care: a scoping review. J Clin Med 2024 Mar 27;13(7):1937 [FREE Full text] [doi: 10.3390/jcm13071937] [Medline: 38610702]
21. Horn A, Jírů-Hillmann S, Widmann J, Montellano FA, Salmen J, Pryss R, et al. Systematic review on the effectiveness of mobile health applications on mental health of breast cancer survivors. J Cancer Surviv 2025 Feb;19(1):1-17. [doi: 10.1007/s11764-023-01470-6] [Medline: 37906420]
22. Saevarsdottir SR, Gudmundsdottir SL. Mobile apps and quality of life in patients with breast cancer and survivors: systematic literature review. J Med Internet Res 2023 Jul 26;25:e42852 [FREE Full text] [doi: 10.2196/42852] [Medline: 37494111]
23. Chan RJ, Howell D, Lustberg MB, Mustian K, Koczwara B, Ng CC, et al. Advances and future directions in the use of mobile health in supportive cancer care: proceedings of the 2019 MASCC Annual Meeting symposium. Support Care Cancer 2020 Sep 14;28(9):4059-4067. [doi: 10.1007/s00520-020-05513-x] [Medline: 32405966]
24. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-544. [Medline: 8941260]
25. Coleman K, Austin BT, Brach C, Wagner EH. Evidence on the chronic care model in the new millennium. Health Aff (Millwood) 2009;28(1):75-85 [FREE Full text] [doi: 10.1377/hlthaff.28.1.75] [Medline: 19124857]
26. Nundy S, Dick JJ, Goddu AP, Hogan P, Lu CE, Solomon MC, et al. Using mobile health to support the chronic care model: developing an institutional initiative. Int J Telemed Appl 2012;2012:871925 [FREE Full text] [doi: 10.1155/2012/871925] [Medline: 23304135]
27. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LM. The eHealth enhanced chronic care model: a theory derivation approach. J Med Internet Res 2015 Apr 01;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
28. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
29. Green EC, Murphy EM, Gryboski K. The health belief model. In: Gulliver SB, Cohen LM, editors. The Wiley Encyclopedia of Health Psychology. Hoboken, NJ: John Wiley & Sons; 2020:211-214.
30. Nasution A, Yusuf A, Lean Keng S, Rasudin NS, P Iskandar Y, Ab Hadi IS. Development of mobile app for breast examination awareness using health belief model: a qualitative study. Asian Pac J Cancer Prev 2021 Oct 01;22(10):3151-3163. [doi: 10.31557/apjcp.2021.22.10.3151]
31. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec;25(8):661-669. [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
32. Bandura A. Self Efficacy: The Exercise of Control. New York, NY: Worth Publishers; 1997.
33. Foster C, Fenlon D. Recovery and self-management support following primary cancer treatment. Br J Cancer 2011 Nov 08;105 Suppl 1(Suppl 1):S21-S28 [FREE Full text] [doi: 10.1038/bjc.2011.419] [Medline: 22048029]
34. Bradbury K, Steele M, Corbett T, Geraghty AW, Krusche A, Heber E, et al. Developing a digital intervention for cancer survivors: an evidence-, theory- and person-based approach. NPJ Digit Med 2019;2:85 [FREE Full text] [doi: 10.1038/s41746-019-0163-4] [Medline: 31508496]
35. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology group. Am J Clin Oncol 1982 Dec;5(6):649-655. [Medline: 7165009]
36. Foster C, Breckons M, Cotterell P, Barbosa D, Calman L, Corner J, et al. Cancer survivors' self-efficacy to self-manage in the year following primary treatment. J Cancer Surviv 2015 Mar;9(1):11-19 [FREE Full text] [doi: 10.1007/s11764-014-0384-0] [Medline: 25028218]
37. Kim SH, Lee BG, Lee JY, Kim SJ, Hur MH, Lee MH, et al. Psychometric evaluation of a Korean version of the cancer survivors' self-efficacy scale. Cancer Nurs 2019;42(6):509-514. [doi: 10.1097/NCC.0000000000000691] [Medline: 30601268]
38. Kang JI, Chung HC, Kim SJ, Choi HJ, Ahn JB, Jeung H, et al. Standardization of the Korean version of mini-mental adjustment to cancer (K-Mini-MAC) scale: factor structure, reliability and validity. Psychooncology 2008 Jun;17(6):592-597. [doi: 10.1002/pon.1277] [Medline: 17957732]
39. Watson M, Law MG, Santos MD, Greer S, Baruch J, Bliss J. The Mini-MAC: further development of the mental adjustment to cancer scale. J Psychosoc Oncol 1994 Oct 31;12(3):33-46. [doi: 10.1300/J077V12N03_03]
40. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
41. Min S, Lee CI, Kim KI, Suh SY, Kim DK. Development of Korean version of WHO quality of life scale abbreviated version (WHOQOL-BREF). J Korean Neuropsychiatr Assoc 2000;39(3):571-579 [FREE Full text]
42. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
43. Kim M, Jung S, Park JE, Sohn JH, Seong SJ, Kim BS, et al. Validation of the patient health Questionnaire-9 and patient health Questionnaire-2 in the general Korean population. Psychiatry Investig 2023 Sep;20(9):853-860 [FREE Full text] [doi: 10.30773/pi.2023.0100] [Medline: 37794667]
44. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

XSL•FO
RenderX

45. Lee SH, Shin C, Kim H, Jeon SW, Yoon HK, Ko YH, et al. Validation of the Korean version of the Generalized Anxiety Disorder 7 self-rating scale. Asia Pac Psychiatry 2022 Mar 06;14(1):e12421. [doi: 10.1111/appy.12421] [Medline: 32893471]
46. Kim NY, Yoo SK, Jin JC, Han S, Hong JS, Kim SM. Development and verification of the Menopause Emotional Symptom questionnaire. Menopause 2023 Sep 01;30(9):961-968. [doi: 10.1097/GME.0000000000002224] [Medline: 37490658]
47. Kim SM, Kim HJ, Hwang HC, Hong JS, Bae S, Min KJ, et al. The effects of a serious game on depressive symptoms and anxiety in breast cancer patients with depression: a pilot study using functional magnetic resonance imaging. Games Health J 2018 Dec;7(6):409-417. [doi: 10.1089/g4h.2017.0183] [Medline: 30383458]
48. Børøsund E, Ehlers SL, Clark MM, Andrykowski MA, Cvancarova Småstuen M, Solberg Nes L. Digital stress management in cancer: testing StressProffen in a 12-month randomized controlled trial. Cancer 2022 Apr 01;128(7):1503-1512 [FREE Full text] [doi: 10.1002/cncr.34046] [Medline: 34855212]
49. Foster C, Grimmett C, May CM, Ewings S, Myall M, Hulme C, et al. A web-based intervention (RESTORE) to support self-management of cancer-related fatigue following primary cancer treatment: a multi-centre proof of concept randomised controlled trial. Support Care Cancer 2016 Jun;24(6):2445-2453 [FREE Full text] [doi: 10.1007/s00520-015-3044-7] [Medline: 26643072]
50. Ho SM, Fung WK, Chan CL, Watson M, Tsui YK. Psychometric properties of the Chinese version of the Mini-Mental Adjustment to Cancer (MINI-MAC) scale. Psychooncology 2003 Sep;12(6):547-556. [doi: 10.1002/pon.672] [Medline: 12923795]
51. Johansson M, Rydén A, Finizia C. Mental adjustment to cancer and its relation to anxiety, depression, HRQL and survival in patients with laryngeal cancer - a longitudinal study. BMC Cancer 2011 Jun 30;11:283 [FREE Full text] [doi: 10.1186/1471-2407-11-283] [Medline: 21718478]
52. Greer S, Watson M. Mental adjustment to cancer: its measurement and prognostic importance. Cancer Surv 1987;6(3):439-453. [Medline: 3326657]
53. Kapoor A, Nambisan P, Baker E. Mobile applications for breast cancer survivorship and self-management: a systematic review. Health Informatics J 2020 Dec 25;26(4):2892-2905 [FREE Full text] [doi: 10.1177/1460458220950853] [Medline: 32842830]
54. Grassi L, Buda P, Cavana L, Annunziata MA, Torta R, Varetto A. Styles of coping with cancer: the Italian version of the Mini-Mental Adjustment to Cancer (Mini-MAC) scale. Psychooncology 2005 Feb;14(2):115-124. [doi: 10.1002/pon.826] [Medline: 15386782]
55. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014 Dec 18;16(12):e295 [FREE Full text] [doi: 10.2196/jmir.3348] [Medline: 25525672]
56. van den Berg SW, Gielissen MF, Custers JA, van der Graaf WT, Ottevanger PB, Prins JB. BREATH: web-based self-management for psychological adjustment after primary breast cancer-results of a multicenter randomized controlled trial. J Clin Oncol 2015 Sep 01;33(25):2763-2771. [doi: 10.1200/JCO.2013.54.9386] [Medline: 26169621]
57. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]
58. Xu A, Wang Y, Wu X. Effectiveness of e-health based self-management to improve cancer-related fatigue, self-efficacy and quality of life in cancer patients: systematic review and meta-analysis. J Adv Nurs 2019 Dec;75(12):3434-3447. [doi: 10.1111/jan.14197] [Medline: 31566769]
59. Lee MK, Yun YH, Park HA, Lee ES, Jung KH, Noh D. A web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
60. Kim SH, Choe YH, Cho YU, Park S, Lee MH. Effects of a partnership-based, needs-tailored self-management support intervention for post-treatment breast cancer survivors: a randomized controlled trial. Psychooncology 2022 Mar;31(3):460-469. [doi: 10.1002/pon.5828] [Medline: 34549864]
61. Schuit AS, Holtmaat K, Lissenberg-Witte BI, Eerenstein SE, Zijlstra JM, Eeltink C, et al. Efficacy of the eHealth application Oncokompas, facilitating incurably ill cancer patients to self-manage their palliative care needs: a randomized controlled trial. Lancet Reg Health Eur 2022 Jul;18:100390 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100390] [Medline: 35496496]
62. Zucca AC, Boyes AW, Lecathelinais C, Girgis A. Life is precious and I'm making the best of it: coping strategies of long-term cancer survivors. Psychooncology 2010 Dec;19(12):1268-1276. [doi: 10.1002/pon.1686] [Medline: 20151408]
63. Alcalar N, Ozkan S, Kucucuk S, Aslay I, Ozkan M. Association of coping style, cognitive errors and cancer-related variables with depression in women treated for breast cancer. Jpn J Clin Oncol 2012 Oct;42(10):940-947. [doi: 10.1093/jjco/hys119] [Medline: 22859828]
64. Capozzo MA, Martinis E, Pellis G, Giraldi T. An early structured psychoeducational intervention in patients with breast cancer: results from a feasibility study. Cancer Nurs 2010;33(3):228-234. [doi: 10.1097/NCC.0b013e3181c1acd6] [Medline: 20357661]

XSL•FO
RenderX

65. Kissane DW, Bloch S, Smith GC, Miach P, Clarke DM, Ikin J, et al. Cognitive-existential group psychotherapy for women with primary breast cancer: a randomised controlled trial. Psychooncology 2003 Sep;12(6):532-546. [doi: 10.1002/pon.683] [Medline: 12923794]
66. Lee KS, Lee R, Kim DM, Kim SH. The effects of a comprehensive education program on knowledge, self-efficacy, and coping style among newly diagnosed patients with breast cancer. Asian Oncol Nurs 2012;12(1):35. [doi: 10.5388/aon.2012.12.1.35]
67. Osborne RH, Elsworth GR, Kissane DW, Burke SA, Hopper JL. The Mental Adjustment to Cancer (MAC) scale: replication and refinement in 632 breast cancer patients. Psychol Med 1999 Nov 01;29(6):1335-1345. [doi: 10.1017/s0033291799001142] [Medline: 10616939]
68. O'Brien CW, Moorey S. Outlook and adaptation in advanced cancer: a systematic review. Psychooncology 2010 Dec;19(12):1239-1249. [doi: 10.1002/pon.1704] [Medline: 20200856]
69. Seok JH, Kim LS, Hong N, Hong HJ, Kim SJ, Kang HJ, et al. Psychological and neuroendocrinological characteristics associated with depressive symptoms in breast cancer patients at the initial cancer diagnosis. Gen Hosp Psychiatry 2010;32(5):503-508. [doi: 10.1016/j.genhosppsych.2010.06.001] [Medline: 20851271]
70. Gibbons A, Groarke A, Sweeney K. Predicting general and cancer-related distress in women with newly diagnosed breast cancer. BMC Cancer 2016 Dec 03;16(1):935 [FREE Full text] [doi: 10.1186/s12885-016-2964-z] [Medline: 27914469]
71. Tojal C, Costa R. Depressive symptoms and mental adjustment in women with breast cancer. Psychooncology 2015 Sep;24(9):1060-1065. [doi: 10.1002/pon.3765] [Medline: 25645194]
72. Fawzy NW. A psychoeducational nursing intervention to enhance coping and affective state in newly diagnosed malignant melanoma patients. Cancer Nurs 1995 Dec;18(6):427-438. [Medline: 8564938]
73. Setyowibowo H, Yudiana W, Hunfeld JA, Iskandarsyah A, Passchier J, Arzomand H, et al. Psychoeducation for breast cancer: a systematic review and meta-analysis. Breast 2022 Apr;62:36-51 [FREE Full text] [doi: 10.1016/j.breast.2022.01.005] [Medline: 35121502]
74. Shim EJ, Lee JW, Cho J, Jung HK, Kim NH, Lee JE, et al. Association of depression and anxiety disorder with the risk of mortality in breast cancer: a national health insurance service study in Korea. Breast Cancer Res Treat 2020 Jan;179(2):491-498. [doi: 10.1007/s10549-019-05479-3] [Medline: 31673880]
75. Al-Alawi KS, Al-Azri M, Al-Fahdi A, Chan MF. Effect of psycho-educational intervention to reduce anxiety and depression at postintervention and follow-up in women with breast cancer: a systematic review and meta-analysis. Semin Oncol Nurs 2022 Dec;38(6):151315. [doi: 10.1016/j.soncn.2022.151315] [Medline: 35879186]
76. Cipolletta S, Simonato C, Faccio E. The effectiveness of psychoeducational support groups for women with breast cancer and their caregivers: a mixed methods study. Front Psychol 2019;10:288 [FREE Full text] [doi: 10.3389/fpsyg.2019.00288] [Medline: 30833920]
77. Wang Y, Lin Y, Chen J, Wang C, Hu R, Wu Y. Effects of internet-based psycho-educational interventions on mental health and quality of life among cancer patients: a systematic review and meta-analysis. Support Care Cancer 2020 Jun;28(6):2541-2552. [doi: 10.1007/s00520-020-05383-3] [Medline: 32179998]
78. Darvishpour A, Vajari SM, Noroozi S. Can health belief model predict breast cancer screening behaviors? Open Access Maced J Med Sci 2018 May 20;6(5):949-953 [FREE Full text] [doi: 10.3889/oamjms.2018.183] [Medline: 29875878]
79. Jones CJ, Smith H, Llewellyn C. Evaluating the effectiveness of health belief model interventions in improving adherence: a systematic review. Health Psychol Rev 2014;8(3):253-269. [doi: 10.1080/17437199.2013.802623] [Medline: 25053213]
80. Rottmann N, Dalton SO, Christensen J, Frederiksen K, Johansen C. Self-efficacy, adjustment style and well-being in breast cancer patients: a longitudinal study. Qual Life Res 2010 Aug;19(6):827-836. [doi: 10.1007/s11136-010-9653-1] [Medline: 20401551]

Abbreviations

CAMA: Cancer Manager
CCM: chronic care model
CS-SES-K: Korean version of the Cancer Survivor Self-Efficacy Scale
GAD-7: Generalized Anxiety Disorder-7
HBM: health belief model
K-Mini-MAC: Korean version of the Mini-Mental Adjustment to Cancer
MESQ: Menopause Emotional Symptoms Questionnaire
mHealth: mobile health
PHQ-9: Patient Health Questionnaire-9
Q and A: question and answer
QOL: quality of life
TAU: treatment as usual
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs
WHOQOL-BREF: World Health Organization Quality of Life Brief Version

XSL•FO
RenderX

Edited by L Buis; submitted 07.07.24; peer-reviewed by CH Chien, AKC Wong; comments to author 05.08.24; revised version received 26.10.24; accepted 13.02.25; published 26.03.25.

Please cite as:
Kim SM, Kim DS, Jang Y, Kim MK, Yu ES, Han DH, Kim HJ
Evaluating the Effectiveness of a Mobile App for Breast Cancer Self-Management on Self-Efficacy: Nonrandomized Intervention Trial
JMIR Mhealth Uhealth 2025;13:e63989
URL: https://mhealth.jmir.org/2025/1/e63989
doi:10.2196/63989
PMID:

©Sun Mi Kim, Da Seul Kim, Yoonsung Jang, Min Kyoon Kim, Eun-Seung Yu, Doug Hyun Han, Hee Jun Kim. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Smartphone App Self-Management Program for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial of Clinical Outcomes

Lisa Glynn[1,2*], BSc, MSc, PGD, PhD; Eddie Moloney[2,3*], MD; Stephen Lane[2,3], BCh, BAO, MB, PhD; Emma McNally[2,3], BCh, BAO, MB; Carol Buckley[2], BSc, HDip; Margaret McCann[3], BNS, MSc, PhD; Catherine McCabe[3*], BNS, MSc, PhD

[1]School of Nursing and Midwifery, University of Galway, Galway, Ireland

[2]Tallaght University Hospital, Dublin, Ireland

[3]Trinity College, Dublin, Ireland

[*]these authors contributed equally

Corresponding Author:
Lisa Glynn, BSc, MSc, PGD, PhD
School of Nursing and Midwifery
University of Galway
University Road
Galway, H91 TK33
Ireland
Phone: 353 91 492345
Email: lisa.l.glynn@universityofgalway.ie

Abstract

Background: Chronic obstructive pulmonary disease (COPD) negatively impacts clinical health outcomes, resulting in frequent exacerbations, increased hospitalizations, reduced physical activity, deteriorated quality of life, and diminished self-efficacy. Previous studies demonstrated that a self-management program tailored for adults with COPD improves self-management decisions, resulting in a positive effect on clinical health outcomes. Limitations of these studies include issues regarding heterogeneity among interventions used, patient population characteristics, outcome measures, and longitudinal studies. Limited studies focused on the use of a comprehensive self-management program using a smartphone app for adults with COPD over 12 months.

Objective: This study aimed to explore the effectiveness of a smartphone app self-management program and monthly phone calls compared with standard respiratory outpatient care on clinical health outcomes in adults with COPD.

Methods: This was a 3-arm parallel pilot randomized controlled trial (RCT) that included 92 participants. Participants were randomized into intervention arm 1, which included a self-management smartphone app and monthly phone calls (n=31); intervention arm 2, which included a self-management smartphone app (n=31); and arm 3, which was standard respiratory outpatient care (n=30). All arms received standard respiratory outpatient care. The primary outcome was a binary indicator equal to 1 if participants reported attendance to a general practitioner (GP) and or a hospital setting as a result of an exacerbation and 0 otherwise. This indicator was recorded at 6 months and 12 months from the baseline. Secondary outcomes included engagement, breathlessness, physical activity, health-related quality of life, and self-efficacy.

Results: There was a statistically significant difference (P=.03), indicating fewer exacerbations in the intervention arm 2 compared with the control arm at 6 months in the hospital setting. The intervention arms had a statistically significant difference indicating a lower risk of developing an exacerbation at 6 months in both the GP (P=.01) and hospital setting (P=.006) compared to the control arm. Furthermore, intervention arm 1 demonstrated a statistically significant difference in exercise capacity at 6 and 12 months (P=.02 and P=.03). The intervention arm 2 illustrated a statistically significant difference in step count (P=.009) compared to the control arm. The majority of participants (60%, 33/55) used the app over the 12-month period.

Conclusions: This study demonstrated that a smartphone app self-management program had a positive effect on clinical health outcomes for participants with COPD in comparison to standard respiratory outpatient care. This study illustrated benefits such as reduced exacerbations resulting in fewer hospitalizations, improved exercise capacity, and physical activity among the

XSL•FO
RenderX

intervention arms. This was a single-center study, which was limited in power to demonstrate significant effects on all measured outcomes but paves the way for a larger, fully powered multicenter trial exploring the effect of a smartphone app self-management program on clinical health outcomes in adults with COPD.

Trial Registration: ClinicalTrials.gov NCT05061810; https://clinicaltrials.gov/study/NCT05061810

(JMIR Mhealth Uhealth 2025;13:e56318) doi:10.2196/56318

KEYWORDS

COPD; self-management; smartphone application; clinical health outcomes; chronic obstructive pulmonary disease; clinical health; patient; patients; randomised controlled trial; RCT; smartphone; app; apps; application; applications; mhealth; mobile health; effectiveness; intervention; interventions; hospital setting; quality of life; self-efficacy; hospital; hospitals; mobile phone

Introduction

Background

Chronic obstructive pulmonary disease (COPD) is a progressive lung disease resulting in persistent respiratory symptoms, such as a chronic productive cough, breathlessness, wheezing, and airflow limitation primarily caused by cigarette smoking and other noxious gases. It is characterized by a chronic decline in lung function with irreversible airflow obstruction and systemic manifestations resulting in frequency and severity of exacerbations. COPD exacerbations are defined as acute events described by a worsening of respiratory symptoms that are beyond normal day-to-day variations [1,2]. COPD exacerbations are the most frequent presentation in the hospital setting, both nationally and internationally. This accounts for a substantial economic burden [2,3]; for instance, within the European Union, COPD accounts for 56% of the overall respiratory costs annually [4]. COPD can negatively impact clinical health outcomes, resulting in frequent exacerbations, increased hospitalizations, reduced physical activity, deteriorating quality of life, and a diminished sense of self-efficacy [1,2,5-9].

Self-management programs have been shown to improve individual's knowledge, confidence, and skills to self-manage their chronic illness [1,10]. This results in improved clinical health outcomes in terms of enhanced quality of life, improved physical activity, and fewer exacerbations, resulting in reduced hospitalizations, morbidities, and premature death among this cohort [1,7-9]. Successful self-management interventions among adults with COPD result in cost-effectiveness for the health care service [10-14]. Previous studies using self-management programs through a smartphone app have shown a positive effect on clinical health outcomes among this cohort [15-23]. However, these studies [15-28] have issues with heterogeneity among interventions used, consistency of their application, patient population specifics, duration of studies, and outcome measures. In addition, studies have suggested that support from a third party, such as a health care professional may improve engagement levels [16,21,23,24,28]. However, there is no conclusive evidence to support that a third party, such as a health care professional, has a statistically significant impact on engagement levels among this cohort. Therefore, there is a strong need for more research surrounding smartphone apps delivering a comprehensive self-management program for adults with COPD to better understand its role in health care. Furthermore, there is a need to explore whether the use of a third-party involvement, such as a health care professional, to support trial participants in using the app, has any impact on engagement levels.

The use of smartphone apps, remote monitoring, and telemedicine are more frequently used in the delivery of health care both nationally and internationally than before the COVID-19 pandemic [29,30]. It has also been reported that the COVID-19 pandemic has positively influenced the older population's views on technology, resulting in increased use of smartphones following the pandemic [30,31]. Readily available access to educational resources can be a challenge for participants with COPD due to cost, environmental barriers, timing not being suitable, or a lack of transport to travel to these educational sessions [32]. Smartphones have many benefits, such as providing convenient communication with the patient and health care professional; they offer portability, Bluetooth, and internet connection, allowing for the use of various smartphone apps to work anywhere at any time. In addition, smartphone apps support behavior changes by providing education, interactive feedback, motivational messages, and access to internet-based resources available at any time. Also, they are generally available at a lower cost in comparison to other digital technologies, such as computers or tablets [13,33,34]. However, technology is still evolving [35], and the best practices in relation to smartphone apps supporting a comprehensive self-management program are not yet well established.

We conducted a single-center, 3-arm parallel pilot randomized controlled trial (RCT) to explore the effect of a smartphone app self-management program on clinical health outcomes in adults with COPD on a longitudinal basis of 12 months. The main components of this self-management program were monthly education, symptom tracking, communication with a health care professional, goal setting, and weekly motivational messages from the app. Symptom tracking included monitoring physical activity (step count), breathlessness score (modified Medical Research Council dyspnea scale [mMRC]), and recording of clinical parameters (lung function [forced expiratory volume at 1 second, FEV1] and oxygen saturation [SpO_2]) using devices such as a spirometer and pulse oximeter. The measurements obtained from the pulse oximeter and spirometer were downloaded through Bluetooth to the self-management app on the participants' smartphones. This paper provides data about exacerbation rates, engagement, physical activity, breathlessness, health-related quality of life (HRQoL), and self-efficacy.

XSL•FO
RenderX

Objective

The objective of this study was to explore the effectiveness of a smartphone app self-management program and monthly phone calls compared with standard respiratory outpatient care on clinical health outcomes in adults with COPD.

Methods

Overview

As the concept of this trial was the first of its kind in Ireland, it was important to assess its feasibility; therefore, a single-center pilot trial was conducted. Pilot trials are typically not powered studies as they explore a new concept or intervention or, indeed, a trial design where more data is required before progressing to a larger study [36]. A statistical package [37] was used to calculate the number of participants required for a larger national study that resulted in 1888 participants. A CONSORT (Consolidated Standards of Reporting Trials) checklist was completed to guide the design, analysis, and reporting of trial findings (see Multimedia Appendix 1). This trial compared patient outcomes across three arms of the intervention (all in addition to standard respiratory outpatient care): (1) a smartphone app self-management program and monthly phone calls, (2) the smartphone app self-management program alone, and (3) no additional outpatient care. In addition, this trial provides information on the effect size, refusal rates, and attrition rates that will aid in recalculation of the sample size required for the larger multicenter trial.

Recruitment

Participants who attended the respiratory outpatient department (OPD) and met the inclusion criteria were invited to participate in this study by telephone.

Participants were eligible if they (1) were aged 18 years or older, (2) had a confirmed COPD diagnosis defined as the presence of post-bronchodilator FEV1/FVC<0.70 [1], (3) had COPD whose severity was defined by the Global Initiative for Chronic Obstructive Lung Disease guidelines (2023) [1] were included, (4) were able to give informed consent, (5) had a smartphone and were able to use it, and (6) had good dexterity to use devices such as a handheld spirometer and pulse oximeter.

Eligible participants were provided with a hard copy of the study information pack, which included a cover letter, an information leaflet outlining the study, an informed consent form, and a returning stamped envelope. Participation in this study was completely voluntary. No financial incentives were offered to participate in this study.

Once written consent was received, a trial register form was completed, and the participant was assigned an ID number. Participants were randomized into each arm using a random allocation computer software package called Random Allocation Software 2.0 [38]. Given the nature of the trial, it was not feasible to blind participants and the researchers involved, given the type of intervention, a smartphone app. However, the research team was blinded to the allocation of participants to each arm by using allocation concealment. Allocation concealment was achieved using an independent health care worker who retained the random allocation sequence. This guaranteed adequate allocation concealment by preventing those recruiting and entering participants into the study from knowing the next assignment. The researcher gave the independent health care worker the participant ID, which was then recorded to the next randomization sequence number. Permuted block randomization, block sizes of three (allocation ratio of 1:1:1), was used in this study to ensure equilibrium. Participants were randomized into intervention arm 1, which included the smartphone app self-management program and monthly phone calls (n=31), intervention arm 2 included the smartphone app self-management program (n=31), and arm 3 was standard respiratory outpatient care (n=30). Furthermore, outcome assessors were blinded using an independent advanced nurse practitioner who reviewed and confirmed whether the outcome assessment met the primary case definition.

Procedures

The intervention in this trial was a comprehensive self-management program through a smartphone app (Figure 1). The main components of this self-management program were monthly education, symptom tracking, communication with a health care professional, goal setting, and weekly motivational messages from the app. Symptom tracking included monitoring physical activity (step count), breathlessness score (mMRC), recording of clinical parameters (lung function [FEV1]), and oxygen saturation (SpO_2) using devices such as a Spirobank smart spirometer (Figure 2), and a Nonin pulse oximeter (Figure 3) that downloaded via Bluetooth to the self-management program app on the participants' smartphone. In addition, 2 educational videos were uploaded to the self-management app each month for 12 months. Educational videos included features relating to self-management, maintaining a healthy lifestyle, and national support available to participants with COPD.

Participants in the intervention arms were sent an email inviting them to download and activate the self-management program app on their smartphones. This email provided participants with a unique username and password. Education in using the app and devices was provided by the research team and the company patientMpower (pMp). Participants in the intervention arms were provided with a hardback intervention study pack that reiterated the information on how to use the smartphone self-management app and the devices (see Multimedia Appendix 2). Technical support was provided 5 days a week by pMp, and their contact details were displayed in the app and were provided in the intervention study pack. Participants in the intervention arms were advised to use the app by viewing the educational content, input their mMRC score, and use the devices weekly or more times if they wished for 12 months. The smartphone app self-management program sent a motivational message to the participants twice a week, prompting them to use the app. Participants were informed at the recruitment stage that data were collected retrospectively and was not monitored daily. Data were reviewed during the scheduled visits at 6 and 12 months. Participants were informed that the smartphone self-management app was not a replacement for their standard medical care, and in the event of a deterioration in their health status, they should contact their general practitioner (GP) or

XSL•FO
RenderX

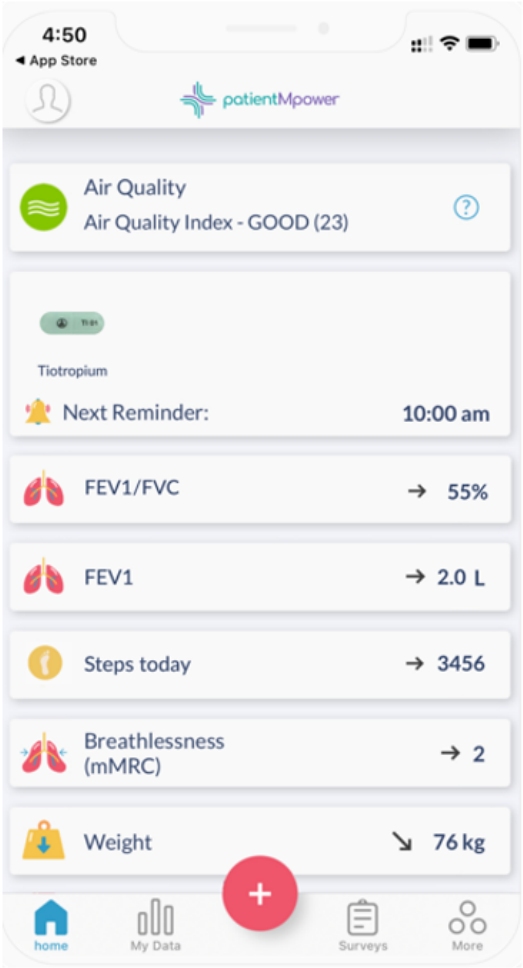
hospital for medical attention. On completion of the trial, app usage was analyzed to determine participants' engagement with the app over the 12 months.

Participants allocated to arm 1 received the self-management app along with monthly phone calls from a health care professional and standard respiratory outpatient care. The aim of the monthly phone calls was to provide support to participants using the app. Participants assigned to arm 2 received the self-management app and standard respiratory outpatient care. Participants in arm 3 received standard respiratory outpatient care.

Participants in all the arms of this trial received standard respiratory outpatient care, which consisted of routine visits by telephone at 6 and 12 months from the respiratory OPD at the research site. During these visits, participants informed the research team of any GP and or hospital visits as a result of an exacerbation of COPD. In addition, they informed the team of their physical activity (exercise behavior, capacity, and step count), breathlessness score (mMRC), nutritional intake, pulmonary rehabilitation, medication adherence, smoking status, vaccinations, and if they had contracted the COVID-19 virus. Finally, they completed questionnaires on the burden of symptoms, quality of life, self-efficacy, and user engagement scale (arm 1 and arm 2 only). Participants were also given self-management education and advice during these visits.

It was anticipated that using the self-management app that provided monthly education, motivation, support from a health care professional, and objective data would enhance participants' knowledge and self-care skills to better manage their chronic illness. This may improve clinical health outcomes, reduce hospitalizations and hospital-associated complications, thereby reducing morbidity, mortality, and overall health care costs among this cohort. The clinical outcomes evaluated in this trial were exacerbation rates, levels of engagement, physical activity, breathlessness, quality of life, and self-efficacy.

Figure 1. Smartphone self-management app. FEV1: forced expiratory volume at 1 second; FVC: forced vital capacity; mMRC: modified Medical Research Council Dyspnea Scale.



XSL•FO
RenderX

Figure 2. Spirobank smart spirometer.

Figure 3. Nonin pulse oximeter.

Primary Outcomes

The primary outcome was a binary indicator equal to 1 if participants reported attendance to a GP and or a hospital setting as a result of an exacerbation and 0 otherwise. This indicator was recorded at 6 months and 12 months from the baseline.

Secondary Outcomes

The first secondary outcome was engagement, which measured viewing the activation rates of the app and devices over 12 months and through the use of a user engagement scale. Also, participants were contacted by phone at 3, 6, and 12 months to complete the refined user engagement scale [39].

Second, physical activity was measured at baseline, 6, and 12 months as the participant reported exercise capacity (participant reported) and step count.

Third, breathlessness was measured using the mMRC scale [40] at baseline, 6, and 12 months. Higher scores represent the severity of breathlessness.

Fourth, the chronic obstructive pulmonary disease assessment test (CAT) [41] score was measured at baseline, 6, and 12 months. Higher scores signified the greater burden of COPD had on the participants' overall health and well-being.

Fifth, the HRQoL was measured using the clinical COPD questionnaire [42] at baseline, 6, and 12 months. Higher scores illustrated poorer HRQoL.

Sixth, self-efficacy was measured using the Self-Efficacy for Managing Chronic Disease 6-Item Scale at baseline, 6, and 12 months [43]. Higher scores resulted in greater self-efficacy.

Statistical Analysis

The change in an outcome was considered statistically significant when $P<.05$. All data were analyzed using an intention-to-treat analysis (ITT), resulting in all participants in their allocated arm being followed up to trial completion (12 months). To correlate with the principle of ITT analysis, all participants in this trial were kept in their assigned arms, and sensitivity analysis was conducted. Sensitivity analysis was performed by using the negative outcome for participant attrition

XSL•FO
RenderX

(loss of participants to follow up) in the intervention arms (exacerbation of COPD) and participant attrition in the control arm to have the best outcome (no exacerbation of COPD), and this was repeated vice versa. Sensitivity analysis results were compared with the original analysis. This assessed the probable impact of participant attrition, resulting in an unbiased estimate of the true effect of the intervention.

Summary statistics were calculated, with the frequency and the percentages being reported for categorical variables. Continuous variables are presented as mean (SD). Finally, for ordinal variables, median and IQR were reported since they were not normally distributed as determined by the Shapiro-Wilk test. To test for significance between baseline and intervention effectiveness using categorical data, the chi-square, and Fisher exact test were used. Fisher exact test is often used for smaller sample sizes. Independent *t* tests were used to compare the mean difference of continuous outcomes between the intervention (arms 1 and 2) versus the control arm (arm 3). Furthermore, the relative risk and absolute risk reduction, along with the associated 95% CI, were used on all outcomes of this study. The ANOVA test was used for continuous variables (dependent variable) to investigate for a significant difference. In relation to the ordinal variables, the Mann-Whitney *U* test was completed to compare differences between arm 1 and arm 2. The Kruskal Wallis test was completed for ordinal variables to determine a statistically significant difference between the three arms, arm 1, arm 2, and arm 3. The logistic model was used to explore the disparities between the three arms in relation to the primary outcome over time. This provided estimates of the probability of developing a COPD exacerbation over time in each arm, resulting in a GP or hospital visit. The possible influencing factors in developing an exacerbation were included, such as age, gender, time, severity of disease, comorbidities, and smoking history. This data was presented as both adjusted and unadjusted odds ratios (OR) and their corresponding 95% CIs for each variable. In addition, the generalized linear mixed model (GLMM) was used as participants were measured at two different time points, 6 and 12 months. This model is an extension of the logistic regression model, which identifies the correlation exhibited by longitudinal data. The GLMM was used because there were categorical outcomes measured over time resulting in correlated data. The logistic regression model ignores the fact that this data is correlated, thereby ignoring this correlation.

Ethical Considerations

This study was approved by the St James' Hospital and Tallaght University Hospital Joint Research Ethics Committee, Dublin, Ireland on July 02, 2021. This trial adhered to the rights of the persons, and each participant had the right to make an informed, voluntary decision to participate in this trial. Eligible participants were provided with verbal and written study information packs. Participants had the right to withdraw from this trial at any stage. To adhere to privacy and confidentiality, each participant was assigned a trial identification number and this number was used on all trial documentation. The pseudoanonymized data were stored in a password-protected shared folder at the study center, which was only accessed by the research team. In addition, monetary rewards were not offered for participation in the trial. The ClinicalTrial.gov ID assigned to this study was NCT05061810.

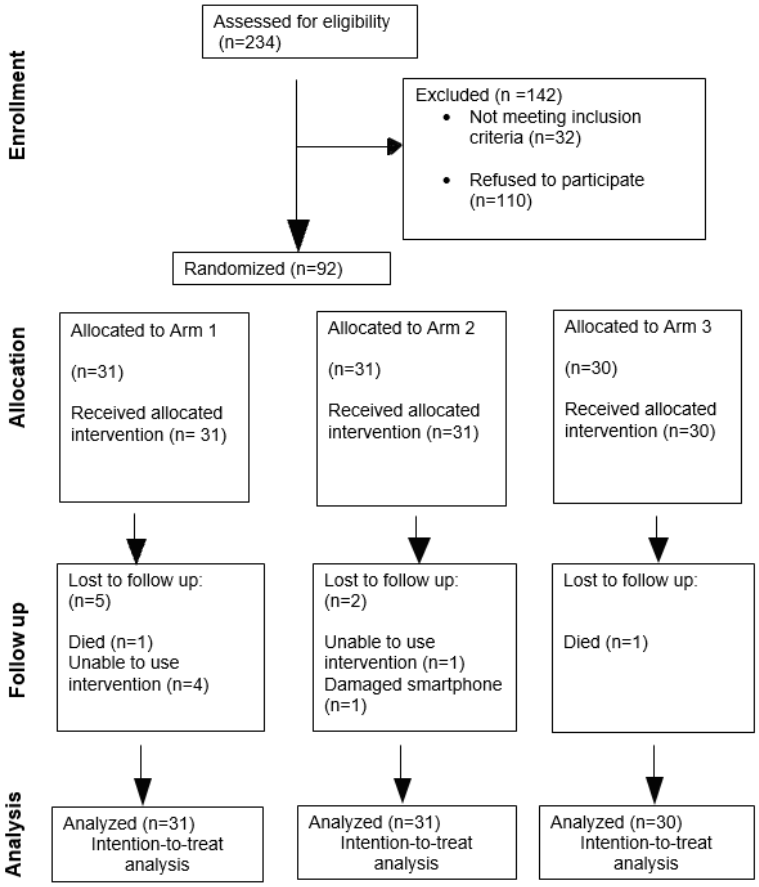
Results

A total of 234 participants were assessed for eligibility from August 2021 to February 2022. From this sample, 202 participants were eligible, and 32 participants were not eligible as they did not meet the inclusion criteria. The main reasons for ineligibility were no smartphone, digital literacy, reduced dexterity, therefore, unable to use the devices, dementia, and blindness. From the sample of 202 eligible participants, 110 participants refused to participate in the trial for various reasons such as “not interested,” “had no time” (n=90), or “ongoing medical investigations” (n=20; see Figure 4 for the CONSORT flow diagram).

A total of 92 participants were randomized to each arm, 31 participants were allocated to arm 1, 31 participants were allocated to arm 2, and 30 participants were allocated to arm 3 (Figure 4). There were 8 participants who were lost to follow-up. Data were collected at the following time points: baseline, 3 (user engagement only), 6, and 12 months. The mean age of participants was 66.8 (SD 7.9) years, ranging from 45 to 81 years of age. There were 57% (52/92) females in the study, with most participants having moderate airway obstruction (43/92, 47%), see Table 1. There were no significant differences in the distribution of baseline characteristics across the three arms except for the variable pulmonary rehabilitation (Table 2). A substantial proportion (84/92, 91%) of the participants reported not engaging with pulmonary rehabilitation at baseline. As this study was conducted during a respiratory pandemic, COVID-19, routine face-to-face pulmonary rehabilitation was not operating due to the Government imposed lockdown restrictions. There was a statistically significant difference, P=.01, among the three arms concerning the participants who underwent pulmonary rehabilitation engagement and those who did not (Table 2). Post hoc pairwise comparisons with Bonferroni correction indicated that arm 1 and arm 2 were different from each other with a P=.02, with more participants in arm 2 participating in pulmonary rehab in comparison with arm 1. No adverse events were reported during this trial.

XSL•FO
RenderX

Figure 4. CONSORT flow diagram. ITT: intention-to-treat analysis.



XSL•FO
RenderX

Table 1. Baseline characteristics of trial participants (N=92).

Variables	Values
Sex, n (%)	
Male	40 (43)
Female	52 (57)
Etiology, n (%)	
Cigarette smoking	90 (98)
Recreational drug abuse	2 (2)
Comorbidities (CCI[a]), n (%)	
Mild	56 (61)
Moderate	28 (30)
Severe	8 (9)
Exercise behavior, n (%)	
Engaging	13 (14)
Not engaging	78 (85)
Other	1 (1)
Nutritional intake, n (%)	
Yes	79 (86)
No	13 (14)
CAT score[b] , n (%)	
Low	16 (17)
Medium	29 (32)
High	35 (38)
Very high	12 (13)
Severity, n (%)	
Mild	25 (27)
Moderate	43 (47)
Severe	17 (19)
Very severe	7 (7)
Medication, n (%)	
No therapy	3 (3)
Single therapy	5 (5)
Dual therapy	24 (26)
Triple therapy	4 (65)
Age, mean (SD)	66.8 (7.9)
Flu vaccine, n (%)	
Yes	53 (58)
No	39 (42)
Pulmonary rehabilitation, n (%)	
Yes	8 (9)
No	84 (91)
Medication adherence, n (%)	
Yes	83 (90)

XSL•FO
RenderX

Variables	Values
No	7 (8)
Not applicable	2 (2)
mMRC[c] , n (%)	
0	5 (5)
1	18 (20)
2	44 (48)
3	18 (20)
4	7 (7)
Quit smoking, n (%)	
Yes	45 (49)
No	47 (51)
Baseline COVID-19 vaccine, n (%)	
Yes	89 (97)
No	3 (3)
COVID-19 vaccine at 6 months, n (%)	
Yes	86 (99)
No	1 (1)
COVID-19 vaccine at 12 months	
Yes	84 (99)
No	1 (1)
Contracted COVID-19 at 6 months, n (%)	
Yes	33 (38)
No	54 (62)
Contracted COVID-19 at 12 months, n (%)	
Yes	48 (56)
N	37 (44)

[a]CCI: Charlson Comorbidity Index.

[b]CAT: chronic obstructive pulmonary disease assessment test.

[c]mMRC: modified Medical Research Council Dyspnea Scale.

XSL•FO
RenderX

Table 2. Baseline characteristics by the randomized arm.

Variable	Arm 1 (n=31)[a]	Arm 2 (n=31)[b]	Arm 3 (n=30)[c]	*P* value
Sex, n (%)				.63
Male	15 (48)	14 (45)	11 (37)	
Female	16 (52)	17 (55)	19 (63)	
Etiology, n (%)				.99
Cigarette smoking	30 (97)	30 (97)	30 (100)	
Recreational drug abuse	1 (3)	1 (3)	0 (0)	
Co-morbidities (CCI[d]), n (%)				.18
Mild	16 (51)	18 (58)	22 (73)	
Moderate	12 (39)	9 (29)	7 (23)	
Severe	3 (10)	4 (13)	1 (3)	
Exercise behavior, n (%)				.65
Engaging	6 (19)	4 (13)	3 (10)	
Not engaging	25 (81)	26 (84)	27 (90)	
Other	0 (0)	1 (3)	0 (0)	
Quit smoking, n (%)				.51
Yes	13 (42)	15 (48.4)	17 (57)	
No	18 (58)	16 (51.6)	13 (43)	
COVID-19 vaccine, n (%)				.31
Yes	31 (100)	30 (97)	28 (93)	
No	0 (0)	1 (3)	2 (7)	
Flu vaccine, n (%)				.81
Yes	19 (61)	18 (58)	16 (53)	
No	12 (39)	13 (42)	14 (47)	
Pulmonary rehabilitation, n (%)				.01
Yes	0 (0)	6 (19)	2 (7)	
No	31 (100)	25 (81)	28 (93)	
Medicine adherence, n (%)				.22
Yes	30 (97)	25 (81)	28 (93)	
No	1 (3.2)	4 (13)	2 (7)	
Not applicable	0 (0)	2 (7)	0 (0)	
Nutritional intake, n (%)				.86
Yes	27 (87)	27 (87)	25 (83)	
No	4 (13)	4 (13)	5 (17)	
mMRC score[e] , n (%)				.74
0	2 (6)	2 (6)	1 (3)	
1	5 (16)	8 (26)	5 (17)	
2	16 (52)	16 (52)	12 (40)	
3	6 (19)	4 (13)	8 (27)	
4	2 (6)	1 (3)	4 (13)	
CAT[f], mean (SD)	3.0 (1.0)	2.0 (1.5)	3.0 (1.0)	.35
Medication , n (%)				.06

XSL•FO
RenderX

Variable	Arm 1 (n=31)[a]	Arm 2 (n=31)[b]	Arm 3 (n=30)[c]	*P* value
No therapy	0 (0)	3 (10)	0 (0)	
Single therapy	4 (13)	0 (0)	1 (3)	
Dual therapy	10 (32)	8 (26)	6 (20)	
Triple therapy	17 (55)	20 (64)	23 (77)	
Severity, mean (SD)	2.0 (1.0)	2.0 (1.0)	2.0 (2.0)	.58
Age (years), mean (SD)	65.7 (7.7)	66.5 (8.5)	68.2 (7.3)	.45

[a]Arm 1 includes a self-management app, monthly phone calls, and standard respiratory outpatient care.

[b]Arm 2 is the self-management app and standard respiratory outpatient care.

[c]Arm 3 is standard respiratory outpatient care, the control group.

[d]CCI: Charlson Comorbidity Index.

[e]mMRC: modified Medical Research Council Dyspnea Scale.

[f]CAT: chronic obstructive pulmonary disease.

COPD Exacerbations Reviewed in the GP and Hospital Settings

The proportion of participants in each arm that experienced exacerbations in the hospital and GP setting is reported in Table 3. There was a statistically significant difference in the proportion of participants who had a COPD exacerbation versus those who did not have an exacerbation at 6 months in the hospital setting across the three arms, *P*=.03 (Table 3). Post hoc pairwise comparisons (*P*=.02) revealed that intervention arm 2 experienced fewer exacerbations reviewed in the hospital setting than those in the control arm. The intervention arms had significantly lower odds of developing a COPD exacerbation in both the GP (with an adjusted OR of 0.20 (95% CI 0.06-0.61; *P*=.005) and hospital (OR 0.26, 95% CI 0.07-0.99; *P*=.049 and OR 0.18, 95% CI 0.04-0.78, *P*=.02) setting compared with the control arm (Table 4). The probability of developing a COPD exacerbation for participants with severe severity is 6.55 times significantly higher than for participants with moderate severity, with *P*=.009 (Table 4). There was a statistically significant difference (*P*=.01) at 6 months in the GP setting, indicating that the intervention arm had a lower risk of developing an exacerbation than the control arm (Table 5). The control arm had the highest estimated probability of having a COPD exacerbation over time, see Figure 5. At 6 months, in the hospital setting, there was a statistically significant difference (*P*=.01 and *P*=.006), indicating the intervention arms had a significantly lower risk of developing a COPD exacerbation compared with the control arm (Table 6). However, after 6 months of follow-up, there was a sharp increase in all the arms, with the control arm having the highest and the intervention arms having the lowest estimated probabilities of developing an exacerbation (Figure 6). The risk of developing an exacerbation was 0.04 times lower for participants with mild severity compared with participants with severe severity, with a statistically significant (*P*=.02; Table 6). Sensitivity analysis indicated a significant difference at 12 months (*P*=.02), where the intervention arms experienced fewer exacerbations than the control arm at 12 months (see Multimedia Appendix 3). Participants with missing data in the control arm were assumed to have experienced an exacerbation (negative outcome), while participants with missing data in the intervention arms (arms 1 and 2) were considered to have experienced no exacerbation (best outcome). However, these results may not accurately reflect the true outcomes of participants with missing data, leading to differences in statistical significance.

XSL•FO
RenderX

Table 3. Comparison of proportions of participants with and without an exacerbation in the general practitioner (GP) and hospital settings across arms.

Variable	Arm 1 (n=31)	Arm 2 (n=31)	Arm 3 (n=30)	*P* value
Hospital				
Baseline				.08
Yes, n (%)	10 (32)	10 (32)	17 (57)	
No, n (%)	21 (68)	21 (68)	13 (43)	
Missing, n	0	0	0	
6 months				.03
Yes, n (%)	3 (11)	2 (6)	9 (31)	
No, n (%)	24 (89)	29 (94)	20 (69)	
Missing, n	4	0	1	
12 months				.88
Yes, n (%)	5 (19)	6 (21)	7 (24)	
No, n (%)	21 (81)	23 (79)	22 (76)	
Missing, n	5	2	1	
General practitioner				
Baseline				.50
Yes, n (%)	28 (90)	25 (81)	24 (80)	
No, n (%)	3 (10)	6 (19)	6 (20)	
Missing, n	0	0	0	
6 months				.10
Yes, n (%)	12 (44)	8 (26)	15 (52)	
No, n (%)	15 (56)	23 (74)	14 (48)	
Missing, n	4	0	1	
12 months				.10
Yes, n (%)	11 (42)	11 (38)	19 (66)	
No, n (%)	15 (58)	18 (62)	10 (34)	
Missing, n	5	2	1	

XSL•FO
RenderX

Table 4. Generalized linear mixed model (GLMM) results for general practitioner (GP) and hospital settings.

Variable	General practitioner			Hospital		
	Unadjusted, OR[a] (95% CI)	Adjusted, OR (95% CI)	*P* value	Unadjusted, OR (95% CI)	Adjusted, OR (95% CI)	*P* value
Arm 1[b]	0.45 (0.15-1.34)	0.41 (0.14-1.22)	.11	0.37 (0.09-1.52)	0.26 (0.07-0.99)	.04
Arm 2[c]	0.25 (0.08-0.76)	0.20 (0.06-0.61)	.005	0.28 (0.07-1.20)	0.18 (0.04-0.78)	.02
Severity						
Moderate	3.70 (1.15-11.86)	4.28 (1.37-13.34)	.01	5.54 (1.00-30.61)	__[d]	—
Severe	5.28 (1.26-22.11)	6.55 (1.60-26.90)	.009	5.68 (0.79-40.99)	—	—
Very Severe	3.79 (0.63-22.70)	2.79 (0.51-15.38)	.23	9.27 (0.84-101.95)	—	—
Comorbidities						
Mild	4.97 (0.58-42.97)	—	—	0.24 (0.03-2.12)	0.10 (0.01-0.98)	.048
Moderate	4.65 (0.50-43.20)	—	—	1.04 (0.12-9.29)	0.58 (0.07-4.77)	.60
Baseline smoking	1.54 (0.64-3.71)	—	—	1.31 (0.42-4.04)	—	—
Age	0.99 (0.93-1.04)	—	—	1.05 (0.97-1.14)	—	—
Gender	0.83 (0.35-2.02)	—	—	2.06 (0.63-6.75)	—	—
Time	1.69 (0.82-3.46)	—	—	1.59 (0.64-3.95)	—	—

[a]OR: odds ratio.

[b]Arm 1 includes a self-management app, monthly phone calls, and standard respiratory outpatient care.

[c]Arm 2 is the self-management app and standard respiratory outpatient care.

[d]Not applicable.

Table 5. Logistic model results for general practitioner (GP) setting.

Variable	Baseline		6 months		12 months	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Arm 1[a]	2.14 (0.44-12.28)	.35	0.59 (0.16-2.02)	.40	0.29 (0.08-0.96)	.04
Arm 2[b]	0.77 (0.17-3.29)	.71	0.21 (0.05-0.71)	.01	0.29 (0.08-0.95)	.04
Comorbidities: mild	1.55 (0.17-10.54)	.66	1.76 (0.16-43.09)	.66	3.56 (0.38-81.90)	.31
Comorbidities: moderate	2.59 (0.24-24.63)	.40	1.99 (0.18-49.12)	.60	1.98 (0.19-46.42)	.59
Severity: mild	2.59 (0.37-17.87)	.32	0.24 (0.03-1.85)	.16	0.44 (0.06-2.83)	.38
Severity: moderate	5.98 (0.83-43.99)	.06	2.10 (0.36-13.83)	.41	1.02 (0.16-6.19)	.98
Severity: severe	16.26 (1.38-433.05)	.04	3.27 (0.47-26.91)	.24	1.00 (0.13-7.09)	.99
Age	0.97 (0.88-1.06)	.49	0.98 (0.91-1.05)	.55	1.01 (0.94-1.08)	.75
Smoking status: yes	0.98 (0.25-3.74)	.97	1.30 (0.44-3.97)	.63	0.33 (0.11-0.95)	.04
Sex: female	1.97 (0.56-7.34)	.29	1.00 (0.35-2.83)	.99	0.39 (0.13-1.06)	.07

[a]Arm 1 includes a self-management app, monthly phone calls, and standard respiratory outpatient care.

[b]Arm 2 is the self-management app and standard respiratory outpatient care.

XSL•FO
RenderX

Figure 5. Estimated probabilities of an exacerbation over time in the GP setting.

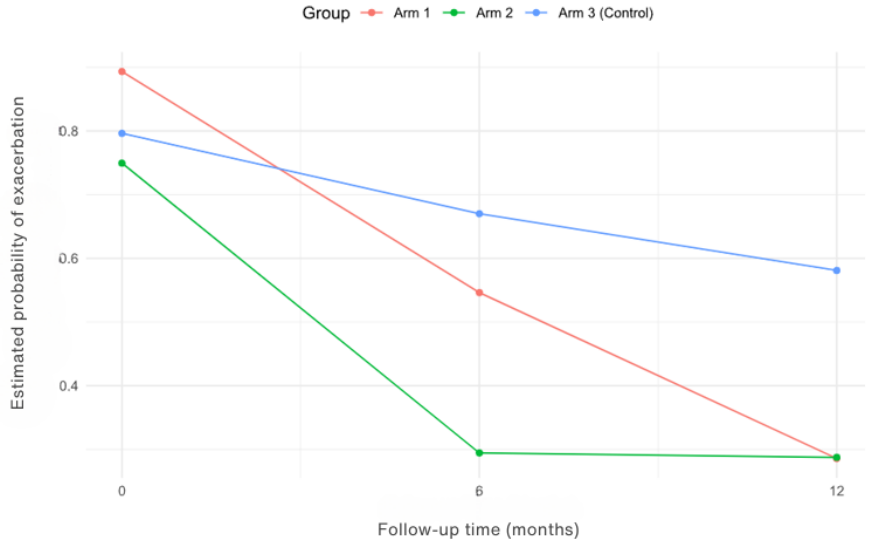


Table 6. Logistic model results for the hospital setting.

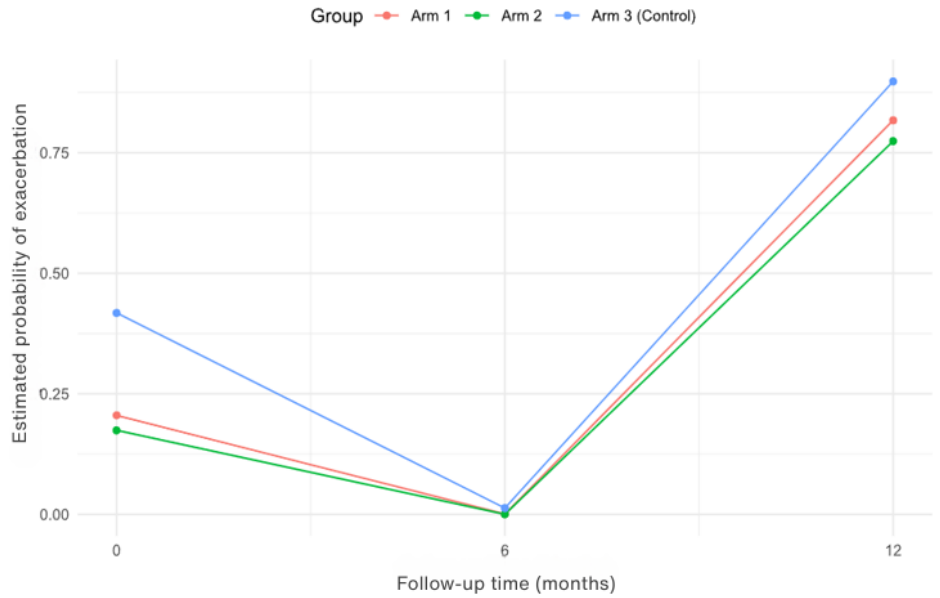
Variable	Baseline		6 months		12 months	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Arm 1[a]	0.36 (0.11-1.09)	.07	0.06 (0.0046-0.45)	.01	0.51 (0.09-2.37)	.39
Arm 2[b]	0.29 (0.09-0.89)	.03	0.02 (0.001-0.24)	.006	0.39 (0.08-1.76)	.23
Comorbidities: mild	0.85 (0.15-5.30)	.85	0.04 (0.001-1.84)	.08	0.06 (0.004-0.80)	.03
Comorbidities: moderate	0.79 (0.13-5.29)	.80	0.66 (0.03-22.11)	.79	0.28 (0.03-3.23)	.28
Severity: mild	0.62 (0.09-4.19)	.61	0.26 (0.01-8.12)	.38	0.04 (0.001-0.49)	.02
Severity: moderate	1.29 (0.22-8.25)	.77	1.79 (0.18-42.16)	.64	0.26 (0.04-1.79)	.16
Severity: severe	2.64 (0.39-19.99)	.32	0.99 (0.06-27.47)	.99	0.74 (0.10-5.69)	.76
Age	1.01 (0.95-1.08)	.69	1.09 (0.97-1.25)	.16	0.99 (0.91-1.08)	.88
Smoking: yes	1.01 (0.38-2.67)	.98	0.44 (0.08-2.05)	.30	0.97 (0.26-3.61)	.96
Sex: female	0.72 (0.28-1.82)	.48	2.59 (0.52-16.77)	.26	3.38 (0.89-1.56)	.09

[a]Arm 1 includes a self-management app, monthly phone calls, and standard respiratory outpatient care.

[b]Arm 2 is the self-management app and standard respiratory outpatient care.

XSL•FO
RenderX

Figure 6. Estimated probabilities of an exacerbation over time in the hospital setting.



Engagement

There was a statistically significant difference in the distribution of the self-reported user engagement scale across the intervention arms at 6 months and 12 months of follow-up (*P*=.02 and *P*=.03, respectively; Table 7). Arm 1 reported better engagement than arm 2 at 6 and 12 months. The Friedman rank sum test was used to analyze the potential differences in engagement using the app across the three time points. The test yielded a Friedman chi-squared statistic of 1.273 (*P*=.52). There were no statistically significant differences in the median of user app engagement across the three time points, which indicates that the monthly phone calls in arm 1 did not have a statistically significant effect on engagement at 3, 6, and 12 months of follow-up. Engagement was sustained across the 12 months among participants in this trial. The average adherence among participants using the pulse oximeter was 66.8 (SD 33.5) and 51.7 (SD 31.7) using the spirometry device over 52 weeks (Table 8). App usage determined by app logins was stable among the intervention arms from week 1 to week 46. Following week 46 until week 52, there was a gradual decline in using the app and devices noted among 40% (n=22) of participants in the intervention arms (Figure 7). Participants in the intervention arms were sent a letter 6 weeks in advance of the 12-month appointment informing them that the trial was due to end. This may have contributed to the reduced engagement from week 46 until 52, as participants were aware the trial was coming to an end. However, participants engaged with the app once or more times a week despite age, gender, or severity of disease (Figure 7). One participant withdrew from this study due to unresolved technical issues.

Table 7. User engagement for the three follow-up periods.

Variable	Arm 1 (n=31)[a], n (%)	Arm 2 (n=31)[b], n (%)	*P* value
3 months			.60
Average engagement	4 (13.3)	3 (9.7)	
Very good	21 (70.0)	26 (83.9)	
Excellent	5 (16.7)	2 (6.5)	
6 months			.02
Average engagement	2 (7.4)	8 (25.8)	
Very good	19 (70.4)	21 (67.7)	
Excellent	6 (22.2)	2 (6.5)	
12 months			.04
Average engagement	1 (3.9)	5 (16.7)	
Very good	18 (69.2)	22 (73.3)	
Excellent	7 (26.9)	3 (10)	

[a]Arm 1 includes a self-management app, monthly phone calls, and standard respiratory outpatient care.

[b]Arm 2 is the self-management app and standard respiratory outpatient care.

XSL•FO
RenderX

Table 8. Adherence to using devices.

Variable	Arm 1[a], mean (SD)	Arm 2[b], mean (SD)	Arm 3[c], mean (SD)	*P* value
Adherence to the pulse oximeter	67.0 (33.8)	66.7 (33.7)	—[d]	.97
Adherence with spirometry	52.1 (34.6)	51.3 (29.1)	—	.91

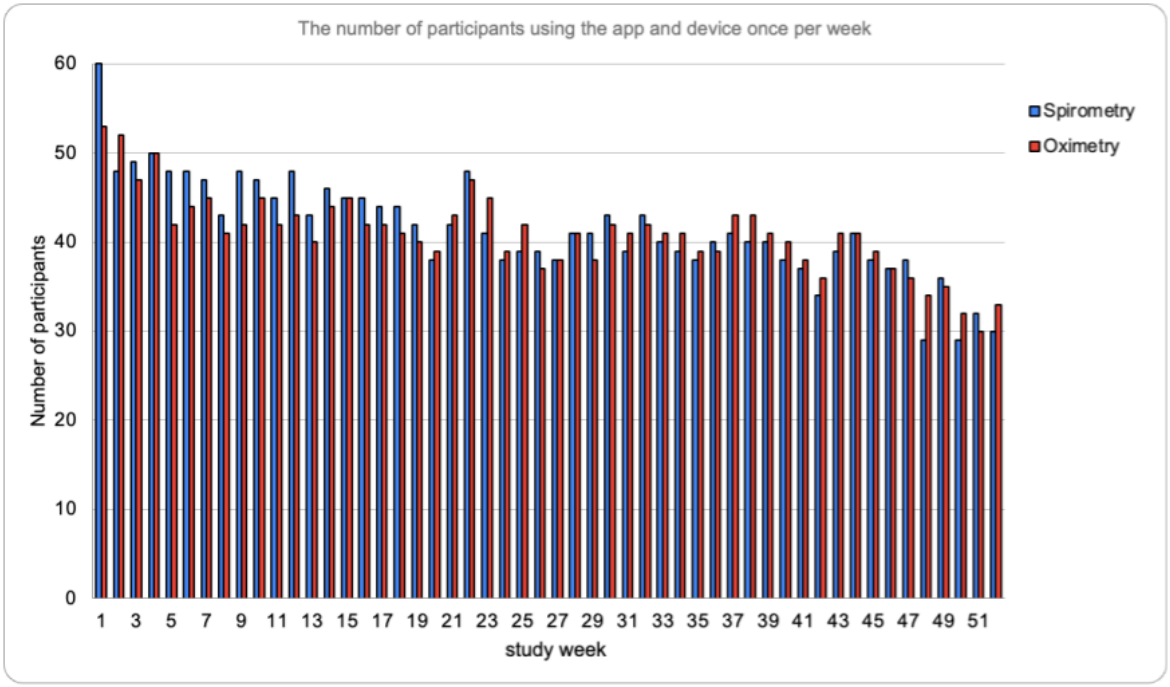
[a]Arm 1 includes a self-management app, monthly phone calls, and standard respiratory outpatient care.

[b]Arm 2 is the self-management app and standard respiratory outpatient care.

[c]Arm 3 is the standard respiratory outpatient care, the control group.

[d]Not applicable.

Figure 7. App logins and use of devices over 12 months.



Physical Activity

Most participants (n=84, 91%) reported not engaging with pulmonary rehabilitation at baseline (Table 1). There was a statistically significant difference (*P*=.01) among the 3 arms concerning the participants who participated in pulmonary rehabilitation and those who did not (Table 2). Post hoc pairwise comparisons with Bonferroni correction indicated that arm 1 and arm 2 were different from each other (*P*=.02) with more participants in arm 2 participating in pulmonary rehab in comparison to arm 1. Exercise capacity at both 6 and 12 months of follow-up was statistically significant (*P*=.02 and *P*=.03, respectively; Table 9). Post hoc analysis revealed that the intervention arm 1 illustrated an improved exercise capacity than the control group (*P*=.03).

Also, there was a statistically significant difference in the mean step count (*P*=.009; Table 9). Post hoc pairwise comparisons with Bonferroni correction revealed that arm 2 had an increased step count in comparison to the control group (*P*=.01).

XSL•FO
RenderX

Table 9. Summary statistics for secondary outcomes in randomized groups.

Variable	Arm 1[a]	Arm 2[b]	Arm 3[c]	*P* value
Baseline				
Exercise capacity, mean (SD)	2.0 (1.0)	3.0 (2.0)	3.0 (2.0)	.24
HRQoL[d], mean (SD)	3.0 (1.0)	3.0 (1.0)	3.0 (1.0)	.98
Self-efficacy, mean (SD)	3.0 (1.5)	3.0 (2.0)	3.0 (2.0)	.71
Step count, mean (SD)	5250 (5764)	3500 (4253)	3500 (3000)	.18
mMRC score[e], n (%)				.74
0	2 (6)	2 (6)	1 (3)	
1	5 (16)	8 (26)	5 (17)	
2	16 (52)	16 (52)	12 (40)	
3	6 (19)	4 (13)	8 (27)	
4	2 (6)	1 (3)	4 (13)	
CAT score[f], mean (SD)	3.0 (1.0)	2.0 (1.5)	3.0 (1.0)	.35
6 months				
Exercise capacity, mean (SD)	2.0 (1.0)	2.0 (1.0)	3.0 (2.0)	.02
HRQoL, mean (SD)	2.0 (1.0)	2.0 (1.0)	2.0 (1.0)	.71
Self-efficacy, mean (SD)	4.0 (1.5)	3.0 (1.5)	3.0 (2.0)	.97
Step count, mean (SD)	4650 (6863.3)	3900 (4100)	3500 (6000)	.28
mMRC score, n (%)				.77
0	4 (15)	5 (16)	1 (3)	
1	10 (37)	11 (35)	12 (41)	
2	7 (26)	9 (29)	7 (24)	
3	4 (15)	5 (16)	5 (17)	
4	2 (7)	1 (3)	4 (14)	
CAT score, mean (SD)	2.0 (2.0)	2.0 (0.5)	2.0 (1.0)	.65
12 months				
Exercise capacity, mean (SD)	2.0 (1.0)	2.0 (1.0)	3.0 (1.0)	.03
HRQoL, mean (SD)	2.0 (0.75)	2.0 (0.0)	2.0 (1.0)	.16
Self-efficacy, mean (SD)	4.0 (2.0)	4.0 (2.0)	3.0 (1.0)	.54
Step count, mean (SD)	4500 (6001)	4800 (6315.5)	999 (3251)	.009
mMRC score, n (%)				.14
0	2 (8)	1 (3)	3 (10)	
1	11 (42)	15 (50)	6 (21)	
2	8 (31)	6 (20)	7 (24)	
3	2 (8)	7 (23)	10 (34)	
4	3 (12)	1 (3)	3 (10)	
CAT, mean (SD)	2.0 (1.0)	2.0 (1.0)	2 (1.0)	.08

[a]Arm 1 includes a self-management app, monthly phone calls, and standard respiratory outpatient care.

[b]Arm 2 is the self-management app and standard respiratory outpatient care.

[c]Arm 3 is the standard respiratory outpatient care, the control group.

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

[e]mMRC: modified Medical Research Council Dyspnea Scale.

[f]CAT: chronic obstructive pulmonary disease assessment test.

XSL•FO
RenderX

Assessment of HRQoL

Participants with poor control of HRQoL decreased from 53 participants at baseline to 29 and 20 participants at 6 and 12 months, respectively (Multimedia Appendix 4 shows the summary statistics for secondary outcomes). Although there was no statistically significant difference in terms of HRQoL across the 3 arms during the various time points, participants in the intervention arms were at a lower risk of developing an exacerbation in comparison to the control arm, which impacts HRQoL. Furthermore, in the intervention arms, the relative risk of 0.55 (95% CI 0.30-0.99) and 0.45 (95% CI 0.24-0.85) at baseline signified a 45% significantly lower risk of developing a high CAT score indicating uncontrolled and symptomatic COPD in comparison to the control arm (Table 10).

XSL•FO
RenderX

Table 10. Relative risk (RR), absolute risk reduction (ARR), and 95% CI for the secondary outcomes.

Variables and arms	Baseline	6 months	12 months
RR[a] (95% CI)			
Exercise capacity			
Arm 1	0.45 (0.24-0.85)	0.33 (0.16-0.71)	0.35 (0.16-0.74)
Arm 2	0.52 (0.28-0.94)	0.42 (0.22-0.80)	0.40 (0.20-0.78)
HRQoL[b]			
Arm1	0.68 (0.39-1.18)	0.44 (0.23-0.88)	0.27 (0.12-0.62)
Arm 2	0.71 (0.41-1.22)	0.32 (0.16-0.66)	0.23 (0.10-0.53)
Self-efficacy			
Arm 1	0.61 (0.35-1.08)	0.44 (0.23-0.87)	0.42 (0.21-0.86)
Arm 2	0.58 (0.32-1.04)	0.52 (0.28-0.94)	0.41 (0.21-0.81)
CAT[c] score			
Arm 1	0.55 (0.30-0.99)	0.30 (0.13-0.65)	0.23 (0.09-0.56)
Arm 2	0.45 (0.24-0.85)	0.19 (0.08-0.46)	0.23 (0.10-0.53)
mMRC[d] score			
Arm 1	0.26 (0.12-0.56)	0.22 (0.09-0.54)	0.19 (0.07-0.50)
Arm 2	0.16 (0.06-0.41)	0.19 (0.08-0.46)	0.27 (0.12-0.58)
ARR[e] (95% CI)			
Exercise capacity			
Arm 1	–0.25 (–0.42 to –0.07)	–0.32 (–0.50 to –0.14)	–0.34 (–0.53 to –0.16
Arm 2	–0.18 (–0.36 to –0.01)	–0.24 (–0.41 to –0.06)	–0.29 (–0.46 to -0.11)
HRQoL			
Arm 1	–0.02 (–0.19 to 0.14)	0.09 (–0.09 to 0.29)	–0.14 (–0.32 to 0.03)
Arm 2	0.01 (–0.15 to 0.17)	–0.02 (–0.19 to 0.14)	–0.18 (–0.33 to -0.03)
Self-efficacy			
Arm 1	–0.09 (–0.26 to 0.084)	–0.07 (–0.26 to 0.11)	–0.09 (–0.28 to 0.10)
Arm 2	–0.12 (–0.29 to 0.05)	–0.001 (–0.18 to 0.17)	–0.10 (–0.28 to 0.08)
CAT score			
Arm 1	0.02 (–0.16 to 0.19)	–0.05 (–0.22 to 0.12)	–0.18 (–0.34 to –0.02)
Arm 2	–0.08 (–0.26 to 0.09)	–0.15 (–0.29 to –0.01)	–0.18 (–0.33 to –0.03)
mMRC score			
Arm 1	–0.14 (–0.30 to 0.01)	–0.09 (–0.24 to 0.07)	–0.26 (–0.41 to –0.10)
Arm 2	–0.24 (–0.37 to –0.11)	–0.12 (–0.26 to 0.02)	0.18 (–0.34 to –0.02)

[a]RR: relative risk.

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

[c]CAT: chronic obstructive pulmonary disease assessment test.

[d]mMRC: modified Medical Research Council Dyspnea Scale.

[e]ARR: absolute risk reduction.

Breathlessness

The calculated relative risk relating to the mMRC within the intervention arms was 0.26 (95% CI 0.12-0.56) and 0.16 (95% CI 0.06-0.41). This finding suggests that participants in the intervention arms had a substantially lower significant risk of experiencing high levels of breathlessness compared with those in the control arm at baseline, 6 months, and 12 months (Table 10).

XSL•FO
RenderX

Self-Efficacy

There was no statistically significant difference across the three arms during the various time points relating to self-efficacy (Table 9). However, it was evident that the intervention arms had a lower risk of poor self-efficacy at 6 and 12 months compared with the control arms (Table 10).

Discussion

Principal Findings

This pilot RCT explored the effect of a smartphone app self-management program on clinical health outcomes among COPD participants over 12 months. This study demonstrated statistical significance in the intervention arms whereby there were fewer exacerbations presenting in the hospital setting and improved exercise capacity and step count in comparison with the standard respiratory outpatient care.

It was apparent that there were significantly fewer exacerbations (P=.03) at 6 months in the hospital setting in comparison with the control arm. Furthermore, the odds of developing an exacerbation were lower in the intervention arms in comparison to the control arm. In this trial, participants in the intervention arms had access to educational videos through the app pertaining to recognizing symptoms of an exacerbation and how to self-manage their symptoms or seek help early. In addition, these videos provided education related to the importance of medication adherence and guidance on how to engage with a self-management plan. The main difference between the arms was the use of a smartphone app self-management program, which perhaps resulted in increased knowledge, awareness, and confidence, empowering participants to self-manage their chronic illness, resulting in reduced COPD exacerbation rates and hospitalizations among the intervention arms. This finding was similar in other studies [17,18,22]. Reducing exacerbation rates reduces hospitalizations and morbidities and prevents premature death among this cohort [1,16,44-48]. Also, reducing hospital admissions among this cohort may demonstrate cost effectiveness resulting in reducing the economic burden for healthcare providers [10-12].

There was an increase noted in the estimated probability of developing an exacerbation after 6 months in the hospital setting, which may have resulted in 48% (n=56) of participants reporting contracting COVID-19 at the 12-month follow-up. This may have resulted in increased hospital visits. Also, the 12 months of follow-up included the winter season in Ireland, resulting in high rates of influenza, COVID-19, and pneumonia that would have impacted participants in this trial who were susceptible to these respiratory illnesses due to their underlying respiratory condition. However, the control arm had the highest probability of developing a COPD exacerbation following 6 months in the hospital setting. The findings from this trial illustrated that the risk of developing a COPD exacerbation was 0.04 times lower for participants with a mild level of disease compared to participants with a higher severity of disease, and this is congruent with the literature [1-3]. The majority of participants (73%, 67/92) in this trial had moderate to severe COPD, resulting in a higher proportion of participants at risk of developing an exacerbation in comparison to those with mild severity (n=25). As this study was conducted during the COVID-19 pandemic, routine face-to-face pulmonary rehabilitation was not operating due to the Government-imposed lockdown restrictions. The main benefits of pulmonary rehabilitation include reduced exacerbations resulting in fewer hospitalizations, enhanced physical activity, improved respiratory symptoms, and quality of life [1,2,9,14]. Most participants in this study did not engage in pulmonary rehabilitation at baseline, which may have also affected the exacerbation rate in the GP and hospital setting.

There was a statistically significant improvement in exercise capacity at both 6 and 12 months in the intervention arm in comparison with the control arm (P=.02 and P=.03). Also, the intervention arm demonstrated improved physical activity in the form of step count in comparison to the control arm (P=.009). The reason for this may be that participants in the intervention arms were encouraged to monitor and track goals in relation to their step count. In addition, they had access to educational videos that emphasized the importance of physical activity in maintaining self-control of their chronic illness, COPD. Previous studies have highlighted that including physical activity in a self-management program improves their exercise capacity, which positively affects other clinical health outcomes such as HRQoL, symptom control, and self-efficacy in patients with COPD [13-16,20].

All participants in this study engaged with the app despite age, gender, or severity of disease, and this was similar to other studies [17,25,26]. Interestingly, the monthly phone calls in arm 1 did not have a statistically significant effect on app engagement during this study. This suggests that factors other than the frequency of phone calls contributed to the observed changes in engagement levels over time. Technical issues did not cause harm but were inconvenient to participants, which may have resulted in reduced engagement. Technical issues such as loss of functionality of the app, inability to pair devices with Bluetooth on the smartphone, duplicate recordings, and updating and advancing the app occurred during the study. This is a common problem when using smartphone apps and has been frequently reported in the literature [12,15,16,23-28]. Given that technology is advancing rapidly [49], it is envisioned that smartphone apps will become a natural complement within clinical practice [29,50,51]. Therefore, there is a need for investment in technical advances and support to be readily available for participants to support engagement levels using self-management apps. Previous studies reported sustained engagement for durations of 3 months [17], 8 weeks [26], and 6 weeks [24,25]; however, they noted a gradual decline toward the end of the study [17,24-26]. In this trial, there was a gradual decline in using the app and devices noted among 40% (n=22/55) of participants in the intervention arms, which was unrelated to technical issues. Participants in the intervention arms were sent a letter 6 weeks in advance of the 12-month appointment informing them that the trial was due to end. This may have contributed to the reduced engagement from week 46 until 52 as participants were aware that the trial was coming to an end, the Hawthorne effect commonly reported in RCTs. Despite this, 60% (33/55) of participants in the intervention arms continued to use the app and devices for 12 months. As

XSL•FO
RenderX

outlined in previous studies [17,18,22,52], perhaps participants in this study were motivated to engage with the app as they experienced benefits such as reduced hospital visits due to an exacerbation, improved physical activity, knowledge, awareness, and confidence in self-managing their chronic illness. Furthermore, as noted in the literature, engaging with self-management apps empowers participants to discuss aspects of their chronic illness at their next health care appointment, thereby enhancing the rapport between patients and health care practitioners [12,53].

Strengths and Limitations

The strength of this trial was the use of a comprehensive self-management program comprised of monthly education, symptom tracking, goal setting, communication with a health care professional, and weekly motivational messages. However, smartphone technology investigated in previous studies among participants with COPD only focused on certain aspects of a self-management program, resulting in issues with heterogeneity among interventions used [15-28]. A self-management program needs to address a broader number of relevant issues in order to increase the user's knowledge, confidence, and skills to self-manage their chronic illness [1,47]. This is the first national pilot RCT exploring the effect of a smartphone app self-management program on clinical health outcomes in adults with COPD over 12 months. Only a few international longitudinal studies have been published on this topic [18,20]. Therefore, this study will contribute to the body of existing knowledge on smartphone apps supporting a self-management program among individuals with COPD. The results of this trial demonstrated significant benefits in clinical health outcomes, such as reduced exacerbation rates in the hospital setting, improved exercise capacity, and increased step count among the intervention arms in comparison with standard care. These benefits may have occurred due to readily available education on their smartphone pertaining to self-management, symptom tracking, motivation, and goal setting that resulted in enhanced self-monitoring skills among participants in the intervention arms. Finally, this study will inform a larger trial in terms of methodology, eligibility criteria, recruitment process, sample size, randomization procedures, data collection, and data analysis process. Furthermore, this trial will allow for a review of case definitions of primary outcomes, an overview of retention rates among trial participants, and unanticipated challenges that occurred, which can be addressed when preparing for a larger trial.

This was a single-center trial, which is limiting. Although the sample obtained was from a large academic teaching hospital in the Republic of Ireland, and the findings may be reflective of sites of similar context, they are not generalizable to all hospital settings. In addition, this trial is limited by design and, due to the small sample size, lacks adequate power to demonstrate significant effects on all measured outcomes. This study made multiple comparisons between arms, which can lead to a false inference. However, this was a pilot study, and the results will not be used to make treatment decisions but can be used to generate hypotheses in subsequent, adequately powered trials. This study will provide valuable data for expanding to a multicenter trial. Another limitation is that technical issues experienced during the trial may have limited engagement levels among participants. This is a common problem when using smartphone apps and has been frequently reported in the literature [12,15-28]. There is a need for investment in technical advances and support to be readily available for participants to support engagement levels using self-management apps.

Future Research

It was evident from this trial, similar to the literature, that self-management apps via a smartphone only appeal to a certain cohort of participants for many reasons, such as digital literacy, not owning a smartphone, lack of internet connectivity, poor dexterity to use a smartphone or simply not interested in self-managing their COPD [16-20,22,27,28,35]. It is worth noting that this trial and other studies were conducted before and during the COVID-19 pandemic [16-20,22-28]. Since the COVID-19 pandemic, the use of smartphone apps has increased the delivery of health care both nationally and internationally [29,30]. Given the increased use of digital technology in the delivery of health care currently, further research is required to ascertain if the refusal rate to participate in similar studies remains high. In addition, larger future studies could explore the impact of monthly phone calls relating to the content of the phone calls, user preferences, or external circumstances that may influence app engagement levels among this population. Finally, there is a need for a larger multicenter trial exploring the effect of a smartphone app self-management program on clinical health outcomes in adults with COPD.

Conclusion

This study demonstrated that a smartphone app self-management program had a positive effect on clinical health outcomes for participants with COPD in comparison to standard care. Furthermore, this study demonstrated statistical significance in the intervention arms whereby there were fewer exacerbations presenting in the hospital setting and improved exercise capacity and physical activity in comparison to the standard respiratory outpatient care. The majority of participants engaged with the app for 12 months. There is a need for more longitudinal multicenter studies with fully powered samples to confirm sustained benefits and perhaps identify more statistically significant events.

Acknowledgments

We would like to acknowledge the Centre for Support and Training in Analysis and Research (CSTAR), University College Dublin, Ireland, for providing statistical advice and support for this research. In addition, we would like to acknowledge the Health Service Executive, Steeven's Lane, Dublin 8, Ireland; Tallaght University Hospital, Dublin 24, Ireland; Meath Foundation

XSL•FO
RenderX

& Adelaide Health Foundation, Tallaght University Hospital, Dublin 24, Ireland, for funding this trial. The funders played no role in the study design, data collection, analysis, and interpretation of data or the writing of this paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. All relevant data are included in this article.

Authors' Contributions

LG contributed to conceptualization, methodology, recruitment, acquisition of data, software, formal analysis, investigation, project administration, visualization, writing–original draft, and writing–review and editing. E Moloney contributed to conceptualization, methodology, recruitment, acquisition of data, investigation, visualization, and writing–review and editing. SL contributed to methodology, recruitment, review, and editing. E McNally contributed to methodology, recruitment, data management, acquisition of data, software, and writing–review and editing. CB contributed to methodology, recruitment, data management, acquisition of data, software, and writing–review and editing. MM contributed to conceptualization, methodology, software, formal analysis, investigation, visualization, and writing–review and editing. CM contributed to conceptualization, methodology, software, validation, formal analysis, investigation, visualization, and writing–review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1029 KB - mhealth_v13i1e56318_app1.pdf]

Multimedia Appendix 2
Intervention study pack.
[DOCX File , 396 KB - mhealth_v13i1e56318_app2.docx]

Multimedia Appendix 3
Sensitivity analysis.
[DOCX File , 17 KB - mhealth_v13i1e56318_app3.docx]

Multimedia Appendix 4
Summary statistics for secondary outcomes.
[DOCX File , 17 KB - mhealth_v13i1e56318_app4.docx]

References

1. Gold report. Global Initiative for Chronic Obstructive Lung Disease. 2025. URL: https://goldcopd.org/2025-gold-report/ [accessed 2025-03-01]
2. Global initiative for chronic obstructive lung disease. Irish Thoracic Society. 2018. URL: https://irishthoracicsociety.com/respiratory-health-of-the-nation-2018/ [accessed 2025-01-29]
3. Snell N, Strachan D, Hubbard R, Gibson J, Gruffydd-Jones K, Jarrold I. S32 epidemiology of chronic obstructive pulmonary disease (COPD) in the uk: findings from the british lung foundation's 'respiratory health of the nation' project. Thorax 2016;71(Suppl 3):A20.1-A2A20 [FREE Full text] [doi: 10.1136/thoraxjnl-2016-209333.38]
4. Global Impact of Respiratory Disease - Firs. URL: https://firsnet.org/wp-content/uploads/2025/01/FIRS_Master_09202021.pdf [accessed 2025-03-29]
5. Lundell S, Holmner, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir Med 2015;109(1):11-26 [FREE Full text] [doi: 10.1016/j.rmed.2014.10.008] [Medline: 25464906]
6. Yang F, Xiong ZF, Yang C, Li L, Qiao G, Wang Y, et al. Continuity of care to prevent readmissions for patients with chronic obstructive pulmonary disease: A systematic review and meta-analysis. COPD 2017;14(2):251-261 [FREE Full text] [doi: 10.1080/15412555.2016.1256384] [Medline: 28326901]
7. Shaw G, Whelan ME, Armitage LC, Roberts N, Farmer AJ. Are COPD self-management mobile applications effective? A systematic review and meta-analysis. NPJ Prim Care Respir Med 2020;30(1):11 [FREE Full text] [doi: 10.1038/s41533-020-0167-1] [Medline: 32238810]

XSL•FO
RenderX

8. Kaptein A, Fischer M, Scharloo M. Self-management in patients with COPD: theoretical context, content, outcomes, and integration into clinical care. Int J Chron Obstruct Pulmon Dis 2014;9:907-917 [FREE Full text] [doi: 10.2147/COPD.S49622] [Medline: 25214777]
9. Lenferink A, Brusse-Keizer M, van der Valk PD, Frith PA, Zwerink M, Monninkhof EM, et al. Self-management interventions including action plans for exacerbations versus usual care in patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017;8(8):CD011682 [FREE Full text] [doi: 10.1002/14651858.CD011682.pub2] [Medline: 28777450]
10. Khdour MR, Agus AM, Kidney JC, Smyth BM, McElnay JC, Crealey GE. Cost-utility analysis of a pharmacy-led self-management programme for patients with COPD. Int J Clin Pharm 2011;33(4):665-673. [doi: 10.1007/s11096-011-9524-z] [Medline: 21643784]
11. Zwerink M, Brusse-Keizer M, van der Valk PDLPM, Zielhuis GA, Monninkhof EM, van der Palen J, et al. Self management for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2014;2014(3):CD002990 [FREE Full text] [doi: 10.1002/14651858.CD002990.pub3] [Medline: 24665053]
12. de Batlle J, Massip M, Vargiu E, Nadal N, Fuentes A, Ortega Bravo M, CONNECARE-Lleida Group. Implementing mobile health-enabled integrated care for complex chronic patients: intervention effectiveness and cost-effectiveness study. JMIR Mhealth Uhealth 2021;9(1):e22135 [FREE Full text] [doi: 10.2196/22135] [Medline: 33443486]
13. Alwashmi M, Hawboldt J, Davis E, Marra C, Gamble J, Abu Ashour W. The effect of smartphone interventions on patients with chronic obstructive pulmonary disease exacerbations: A systematic review and meta-analysis. JMIR Mhealth Uhealth 2016;4(3):e105 [FREE Full text] [doi: 10.2196/mhealth.5921] [Medline: 27589898]
14. Demeyer H, Louvaris Z, Frei A, Rabinovich RA, de Jong C, Gimeno-Santos E, Mr Papp PROactive study groupthe PROactive consortium. Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. Thorax 2017;72(5):415-423 [FREE Full text] [doi: 10.1136/thoraxjnl-2016-209026] [Medline: 28137918]
15. Rassouli F, Boutellier D, Duss J, Huber S, Brutsche MH. Digitalizing multidisciplinary pulmonary rehabilitation in COPD with a smartphone application: an international observational pilot study. Int J Chron Obstruct Pulmon Dis 2018;13:3831-3836 [FREE Full text] [doi: 10.2147/COPD.S182880] [Medline: 30538444]
16. Park SK, Bang CH, Lee SH. Evaluating the effect of a smartphone app-based self-management program for people with COPD: A randomized controlled trial. Appl Nurs Res 2020;52:151231. [doi: 10.1016/j.apnr.2020.151231] [Medline: 31955942]
17. North M, Bourne S, Green B, Chauhan AJ, Brown T, Winter J, et al. A randomised controlled feasibility trial of E-health application supported care vs usual care after exacerbation of COPD: the RESCUE trial. NPJ Digit Med 2020;3:145 [FREE Full text] [doi: 10.1038/s41746-020-00347-7] [Medline: 33145441]
18. van Buul AR, Derksen C, Hoedemaker O, van Dijk O, Chavannes NH, Kasteleyn MJ. eHealth program to reduce hospitalizations due to acute exacerbation of chronic obstructive pulmonary disease: retrospective study. JMIR Form Res 2021;5(3):e24726 [FREE Full text] [doi: 10.2196/24726] [Medline: 33734091]
19. Spielmanns M, Gloeckl R, Jarosch I, Leitl D, Schneeberger T, Boeselt T, et al. Using a smartphone application maintains physical activity following pulmonary rehabilitation in patients with COPD: a randomised controlled trial. Thorax 2023;78(5):442-450 [FREE Full text] [doi: 10.1136/thoraxjnl-2021-218338] [Medline: 35450945]
20. Wang L, Guo Y, Wang M, Zhao Y. A mobile health application to support self-management in patients with chronic obstructive pulmonary disease: a randomised controlled trial. Clin Rehabil 2021;35(1):90-101. [doi: 10.1177/0269215520946931] [Medline: 32907384]
21. Kooij L, Vos PJE, Dijkstra A, van Harten WH. Effectiveness of a mobile health and self-management app for high-risk patients with chronic obstructive pulmonary disease in daily clinical practice: mixed methods evaluation study. JMIR Mhealth Uhealth 2021;9(2):e21977 [FREE Full text] [doi: 10.2196/21977] [Medline: 33538699]
22. Alharbey R, Chatterjee S. An mHealth assistive system "MyLung" to empower patients with chronic obstructive pulmonary disease: design science research. JMIR Form Res 2019;3(1):e12489 [FREE Full text] [doi: 10.2196/12489] [Medline: 30888329]
23. Rodríguez Hermosa JL, Fuster Gomila A, Puente Maestu L, Amado Diago CA, Callejas González FJ, Malo De Molina Ruiz R, et al. Assessing the usefulness of the prevexair smartphone application in the follow-up high-risk patients with COPD. Int J Chron Obstruct Pulmon Dis 2021;16:53-65 [FREE Full text] [doi: 10.2147/COPD.S279394] [Medline: 33447026]
24. Knox L, Gemine R, Rees S, Bowen S, Groom P, Taylor D, et al. Assessing the uptake, engagement, and safety of a self-management app, COPD.Pal®, for chronic obstructive pulmonary disease: a pilot study. Health Technol 2021;11(3):557-562. [doi: 10.1007/s12553-021-00534-w]
25. Miller S, Teufel R, Nichols M, Davenport P, Mueller M, Silverman E, et al. Feasibility of RESP-FIT: technology-enhanced self-management intervention for adults with COPD. Int J Chron Obstruct Pulmon Dis 2021;16:3263-3273 [FREE Full text] [doi: 10.2147/COPD.S326675] [Medline: 34887656]
26. Gelbman BD, Reed CR. An integrated, multimodal, digital health solution for chronic obstructive pulmonary disease: prospective observational pilot study. JMIR Form Res 2022;6(3):e34758 [FREE Full text] [doi: 10.2196/34758] [Medline: 35142291]

XSL•FO
RenderX

27. Bowler R, Allinder M, Jacobson S, Miller A, Miller B, Tal-Singer R, et al. Real-world use of rescue inhaler sensors, electronic symptom questionnaires and physical activity monitors in COPD. BMJ Open Respir Res 2019;6(1):e000350 [FREE Full text] [doi: 10.1136/bmjresp-2018-000350] [Medline: 30956796]
28. Crooks MG, Elkes J, Storrar W, Roy K, North M, Blythin A, et al. Evidence generation for the clinical impact of myCOPD in patients with mild, moderate and newly diagnosed COPD: a randomised controlled trial. ERJ Open Res 2020;6(4):00460-02020 [FREE Full text] [doi: 10.1183/23120541.00460-2020] [Medline: 33263052]
29. The Future of Telemedicine After COVID-19 - OECD. 2023. URL: https://www.oecd.org/coronavirus/policy-responses/the-future-of-telemedicine-after-covid-19-d46e9a02/ [accessed 2025-03-29]
30. Digital forcare: A digital health framework for Ireland 2024-2030. Department of Health. 2024. URL: https://www.gov.ie/en/publication/0d21e-digital-for-care-a-digital-health-framework-for-ireland-2024-2030/ [accessed 2024-05-28]
31. Tech usage among older adults skyrockets during pandemic. 2023. URL: https://press.aarp.org/2021-4-21-Tech-Usage-Among-Older-Adults-Skyrockets-During-Pandemic [accessed 2025-01-10]
32. Cox NS, Oliveira CC, Lahham A, Holland AE. Pulmonary rehabilitation referral and participation are commonly influenced by environment, knowledge, and beliefs about consequences: a systematic review using the theoretical domains framework. J Physiother 2017;63(2):84-93 [FREE Full text] [doi: 10.1016/j.jphys.2017.02.002] [Medline: 28433238]
33. Licskai CJ, Sands TW, Ferrone M. Development and pilot testing of a mobile health solution for asthma self-management: asthma action plan smartphone application pilot study. Can Respir J 2013;20(4):301-306 [FREE Full text] [doi: 10.1155/2013/906710] [Medline: 23936890]
34. Gunawardena KC, Jackson R, Robinett I, Dhaniska L, Jayamanne S, Kalpani S, et al. The influence of the smart glucose manager mobile application on diabetes management. J Diabetes Sci Technol 2019;13(1):75-81 [FREE Full text] [doi: 10.1177/1932296818804522] [Medline: 30264583]
35. Byrne T, Murray N, McDonnell-Naughton M, Rowan NJ. Perceived factors informing the pre-acceptability of digital health innovation by aging respiratory patients: a case study from the republic of Ireland. Front Public Health 2023;11:1203937 [FREE Full text] [doi: 10.3389/fpubh.2023.1203937] [Medline: 37942252]
36. Bond C, Lancaster GA, Campbell M, Chan C, Eddy S, Hopewell S, et al. Pilot and feasibility studies: extending the conceptual framework. Pilot Feasibility Stud 2023;9(1):24 [FREE Full text] [doi: 10.1186/s40814-023-01233-1] [Medline: 36759879]
37. Sealed Envelope Power Calculators. URL: https://www.sealedenvelope.com/power/ [accessed 2022-08-10]
38. Random Allocation Software. URL: https://en.freedownloadmanager.org/Windows-PC/Random-Allocation-Software-FREE.html [accessed 2022-08-20]
39. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. International Journal of Human-Computer Studies 2018;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
40. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest 1988;93(3):580-586 [FREE Full text] [doi: 10.1378/chest.93.3.580] [Medline: 3342669]
41. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD assessment test. Eur Respir J 2009;34(3):648-654 [FREE Full text] [doi: 10.1183/09031936.00102509] [Medline: 19720809]
42. van der Molen T, Willemse BWM, Schokker S, ten Hacken NHT, Postma DS, Juniper EF. Development, validity and responsiveness of the Clinical COPD Questionnaire. Health Qual Life Outcomes 2003;1:13 [FREE Full text] [doi: 10.1186/1477-7525-1-13] [Medline: 12773199]
43. Ritter PL, Lorig K. The english and spanish self-efficacy to manage chronic disease scale measures were validated using multiple studies. J Clin Epidemiol 2014;67(11):1265-1273. [doi: 10.1016/j.jclinepi.2014.06.009] [Medline: 25091546]
44. Pedone C, Lelli D. Systematic review of telemonitoring in COPD: an update. Pneumonol Alergol Pol 2015;83(6):476-484 [FREE Full text] [doi: 10.5603/PiAP.2015.0077] [Medline: 26559801]
45. MacNab M, Lee SH, McCloughan L, Hanley J, McKinstry B, Pinnock H. Oximetry-supported self-management for chronic obstructive pulmonary disease: mixed method feasibility pilot project. BMC Health Serv Res 2015;15:485 [FREE Full text] [doi: 10.1186/s12913-015-1135-2] [Medline: 26503028]
46. Cannon D, Buys N, Sriram KB, Sharma S, Morris N, Sun J. The effects of chronic obstructive pulmonary disease self-management interventions on improvement of quality of life in COPD patients: A meta-analysis. Respir Med 2016;121:81-90 [FREE Full text] [doi: 10.1016/j.rmed.2016.11.005] [Medline: 27888996]
47. Newham JJ, Presseau J, Heslop-Marshall K, Russell S, Ogunbayo OJ, Netts P, et al. Features of self-management interventions for people with COPD associated with improved health-related quality of life and reduced emergency department visits: a systematic review and meta-analysis. Int J Chron Obstruct Pulmon Dis 2017;12:1705-1720 [FREE Full text] [doi: 10.2147/COPD.S133317] [Medline: 28652723]
48. Al Rajeh A, Hurst J. Monitoring of physiological parameters to predict exacerbations of chronic obstructive pulmonary disease (COPD): A systematic review. J Clin Med 2016;5(12):108 [FREE Full text] [doi: 10.3390/jcm5120108] [Medline: 27897995]

XSL•FO
RenderX

49. Chakravorti N. Digital transformation overview. Digital Transformation.: Productivity Press; 2022. URL: https://www.taylorfrancis.com/chapters/mono/10.4324/9781003270904-1/digital-transformation-overview-nirjhar-chakravorti [accessed 2025-04-01]
50. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
51. Ryu S. Book Review: mHealth: new horizons for health through mobile technologies: based on the findings of the second global survey on eHealth (Global Observatory for eHealth Series, Volume 3). Healthc Inform Res 2012;18(3):231. [doi: 10.4258/hir.2012.18.3.231]
52. Doyle J, Murphy E, Gavin S, Pascale A, Deparis S, Tommasi P, et al. A digital platform to support self-management of multiple chronic conditions (ProACT): findings in relation to engagement during a one-year proof-of-concept trial. J Med Internet Res 2021;23(12):e22672 [FREE Full text] [doi: 10.2196/22672] [Medline: 34914612]
53. Nunes F, Verdezoto N, Fitzpatrick G, Kyng M, Grönvall E, Storni C. Self-Care technologies in HCI. ACM Trans Comput-Hum Interact 2015;22(6):1-45. [doi: 10.1145/2803173]

Abbreviations

CAT: chronic obstructive pulmonary disease assessment test
CONSORT: Consolidated Standards of Reporting Trials
COPD: chronic obstructive pulmonary disease
FEV1: forced expiratory volume at 1 second
GLMM: generalized linear mixed model
GP: general practitioner
HRQoL: health-related quality of life
ITT: intention-to-treat analysis
mMRC: modified Medical Research Council Dyspnea Scale
OPD: outpatient department
OR: odds ratio
pMp: PatientMpower
RCT: randomized controlled trial
SpO2: peripheral oxygen saturation

Edited by L Buis; submitted 15.01.24; peer-reviewed by R Salari, J Matheson, J Chen; comments to author 01.05.24; revised version received 05.06.24; accepted 22.02.25; published 23.04.25.

Please cite as:
Glynn L, Moloney E, Lane S, McNally E, Buckley C, McCann M, McCabe C
A Smartphone App Self-Management Program for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial of Clinical Outcomes
JMIR Mhealth Uhealth 2025;13:e56318
URL: https://mhealth.jmir.org/2025/1/e56318
doi:10.2196/56318
PMID:

©Lisa Glynn, Eddie Moloney, Stephen Lane, Emma McNally, Carol Buckley, Margaret McCann, Catherine McCabe. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinical Impact of Personalized Physician's Education and Remote Feedback Via a Digital Platform on Glycemic Control: Pilot Randomized Controlled Trial

Jin Yu[1], MD, PhD; Joonyub Lee[1], MD, PhD; Yeoree Yang[1,2], MD, PhD; Eun Young Lee[1], MD, PhD; Seung-Hwan Lee[1,2], MD, PhD; Jae-Hyoung Cho[1,2], MD, PhD

[1]Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[2]Catholic Smart Health Care Center, The Catholic University of Korea, Seoul, Republic of Korea

Corresponding Author:
Jae-Hyoung Cho, MD, PhD
Division of Endocrinology and Metabolism, Department of Internal Medicine
Seoul St. Mary's Hospital, College of Medicine
The Catholic University of Korea
Banpodaero 222
Seoul, 06591
Republic of Korea
Phone: 82 222586007
Email: drhopper@ikoob.com

Abstract

Background: The digital education platform Doctorvice (iKooB Inc.) offers face-to-face physician-patient education during outpatient clinic visits, remote glucose monitoring, and the delivery of educational messages, and is expected to be effective for personalized diabetes care.

Objective: This study aims to evaluate the effectiveness of the digital education platform for diabetes care by comparing cases that included both face-to-face education and remote monitoring with those that included only face-to-face education.

Methods: This was a randomized clinical study conducted at the Diabetes Center of Seoul St. Mary's Hospital. Participants were aged ≥19 years and had glycated hemoglobin (HbA_{1c}) levels between 7.5% and 9.5%. In the intervention group, physicians used the digital education platform to provide face-to-face education at enrollment and at the 3- and 6-month visits, along with remote monitoring during the first 3 months of the 6-month study period. The control group received conventional outpatient education. Both groups completed questionnaires—assessing satisfaction with diabetes treatment, diabetes-related stress, and adherence to diabetes medication—at the beginning and end of the study. The primary endpoint was the change in HbA_{1c} levels.

Results: A total of 66 participants were enrolled between August 1, 2022, and August 31, 2023. Of these, 26 in the intervention group and 30 in the control group were analyzed, excluding 10 participants who dropped out of the study. The mean baseline HbA_{1c} levels were 8.3% (SD 0.6%) in the intervention group and 8.0% (SD 0.5%) in the control group. At the 3-month follow-up, mean HbA_{1c} decreased by 0.5%-7.8% (SD 0.9%; *P*=.01) in the intervention group and by 0.2%-7.8% (SD 0.7%) in the control group. HbA_{1c} levels substantially improved during the first 3 months with both face-to-face education and remote glucose monitoring. However, HbA_{1c} tended to increase during the 3- to 6-month follow-up in the intervention group without the remote monitoring service. Satisfaction with diabetes treatment significantly improved at the end of the study compared with baseline in the intervention group (mean change +3.6 points; *P*=.006). Medication adherence improved in both groups, with no significant difference at 6 months (*P*=.59), although the intervention group showed a greater increase from baseline. Subgroup analysis indicated that the reduction in HbA_{1c} was greater for patients with baseline HbA_{1c} levels ≥8.0%, those aged ≥65 years, smokers, drinkers, and those with obesity in the intervention group.

Conclusions: The digital education platform for personalized diabetes management may be beneficial for glycemic control in type 2 diabetes mellitus. Its effectiveness appears to be enhanced when physicians provide personalized face-to-face education combined with remote feedback.

XSL•FO
RenderX

Trial Registration: Clinical Research Information Service (CRiS) of Republic of Korea KCT0007953; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=23507&search_page=L

(JMIR Mhealth Uhealth 2025;13:e67151) doi:10.2196/67151

KEYWORDS

type 2 diabetes mellitus; digital health; education, distance; remote patient monitoring; distance counseling

Introduction

Diabetes is a global health challenge, affecting 540 million people worldwide as of 2021—a number projected to reach 783 million by 2045 [1]. In Korea, the prevalence of diabetes has steadily increased over the past 9 years, reaching 16.7% in 2020; however, only 25% of these individuals achieved a glycated hemoglobin (HbA_{1c}) level of <6.5% [2].

Effective diabetes management is crucial, as patients with diabetes are responsible for making most of their own health decisions, including dietary choices, physical activity, blood glucose monitoring, carbohydrate counting, and insulin dosage adjustments [3]. Recognizing the complexity of diabetes management, the American Diabetes Association (ADA) guidelines emphasize "diabetes self-management education and support" (DSMES) as an essential component of comprehensive diabetes treatment [4]. Studies conducted in the United Kingdom, such as the patient-centered, group-based self-management program (X-PERT) and the "Diabetes Education and Self-Management for Ongoing and Newly Diagnosed" (DESMOND) program [5,6], along with the "Structured Intensive Diabetes Education Program" (SIDEP) study in Korea [7], have demonstrated that DSMES significantly improves glycemic control. Additionally, DSMES has been shown to reduce HbA_{1c} levels, enhance quality of life, and decrease hospitalizations and health care costs [3,8].

Patients with type 2 diabetes mellitus (T2DM) present with diverse characteristics, including age, sex, BMI, and lifestyle habits. These factors—along with differences in insulin resistance, insulin secretion capacity, and the presence of comorbidities—contribute to the heterogeneity of diabetes and its related complications [9-12]. Therefore, it is crucial to provide individualized education tailored to patient-specific characteristics rather than relying solely on general education.

Given the chronic nature of diabetes, maintaining consistent blood glucose control over the long term is particularly challenging [13]. Therefore, in addition to encouraging patient self-monitoring, effective diabetes management requires medical professionals to provide continuous, appropriate, and patient-specific feedback through education. Considering patients' diverse characteristics and the need for ongoing professional oversight, it is essential that physicians—who best understand their patients—deliver in-person education tailored to individual needs. Moreover, providing blood glucose monitoring to patients may enhance diabetes management.

Recent advancements in information technology have led to the development of various digital health care systems. In line with these advancements, the ADA now recommends that DSMES can be effectively delivered through telehealth and digital health care platforms [4]. Studies have explored glucose control via apps, web-based platforms, and message-based diabetes education programs [14-16]. However, these studies primarily focus on patients who measure their own blood glucose levels and receive standardized feedback, falling short of achieving individualized treatment and management. More effective diabetes treatment and management can be achieved when patients monitor their blood glucose levels under physician guidance and receive immediate, appropriate feedback from their physicians.

In response to this need, a digital-based smart education platform for physicians and patients, called Doctorvice (iKooB Inc.), has been developed [17]. This system provides a comprehensive web interface, Doctorvice Clinic, enabling physicians to quickly access patients' clinical information and deliver educational messages. Physicians can conduct face-to-face education by selecting appropriate content from a web-based library containing over 3000 pages of educational materials or by customizing additional instructions on a digital canvas. The educational content can be transmitted to patients via SMS text messages or the Doctorvice app, allowing them to receive and review the materials repeatedly. Patients can upload self-measured blood glucose (SMBG) data through the app, which physicians can access via the web at any time to provide additional educational content on topics such as healthy eating habits, physical activity, and diabetes self-care. This bidirectional communication supports an educational approach that combines face-to-face sessions with continuous remote monitoring and feedback.

In this study, we aimed to evaluate the effectiveness of a physician-led digital education platform on glycemic control in patients with T2DM. Specifically, we hypothesized that patients receiving both face-to-face education and continuous remote monitoring and feedback from physicians would show greater improvements in glycemic control compared with those receiving only face-to-face education during outpatient visits, without subsequent feedback.

Methods

Study Aim

This study was a 6-month, nonblinded randomized controlled trial (KCT0007953) involving patients with T2DM who visited the Diabetes Clinic at Seoul St. Mary's Hospital between August 1, 2022, and August 31, 2023. The study utilized a digital education platform to assess its effect on the diabetes care model, where physicians provided individualized face-to-face education based on digital content. Patients uploaded their blood glucose levels and received remote feedback from the physician. Moreover, we examined the difference between providing

XSL•FO
RenderX

face-to-face education with continuous remote feedback and providing face-to-face education alone.

Study Design

Participants

Patients with T2DM aged ≥19 years and with an HbA_{1c} level of 7.5%-9.5% were enrolled. Exclusion criteria were patients with type 1 diabetes mellitus, latent autoimmune diabetes in adults, gestational diabetes, or those using insulin. Additionally, patients who required immediate treatment for acute diabetic complications, had serious medical conditions, were infected, underwent surgery, or had a serious injury were excluded. Further, individuals with cognitive decline, illiteracy, or no access to a smartphone were excluded from the study due to difficulties using the digital smart platform. People with no prior history of accessing the app were classified as nonusers and excluded.

Intervention and Control Groups

The participants were randomly assigned to the intervention and control groups in a 1:1 ratio using a random allocation table created by the Clinical Trial Support Center. Both groups received basic education on SMBG, blood pressure (BP) measurement, and lifestyle modification regimens in accordance with the clinical practice guidelines of the Korean Diabetes Association [2]. Outpatient visits were conducted at 3-month intervals over a 6-month period, with questionnaires administered at the beginning and end of the study, and blood tests performed at 0, 3, and 6 months. The blood tests included measurements of blood glucose, HbA_{1c}, blood urea nitrogen, creatinine, liver function tests (aspartate aminotransferase and alanine aminotransferase), and lipid profiles (total cholesterol, triglycerides, high-density lipoprotein cholesterol, and low-density lipoprotein cholesterol). BP, weight, and height were assessed during each outpatient visit.

Alcohol consumption was categorized into 2 groups: nondrinkers and drinkers. Nondrinkers were defined as individuals who did not consume alcohol, while drinkers were those who consumed alcohol, regardless of the amount. Obesity was classified into 5 stages based on BMI values: underweight (<18.5 kg/m^2), normal weight (18.5-22.9 kg/m^2), overweight (23.0-24.9 kg/m^2), obesity class I (25.0-29.9 kg/m^2), and obesity class II (≥30 kg/m^2).

Control Group—Usual Care Group

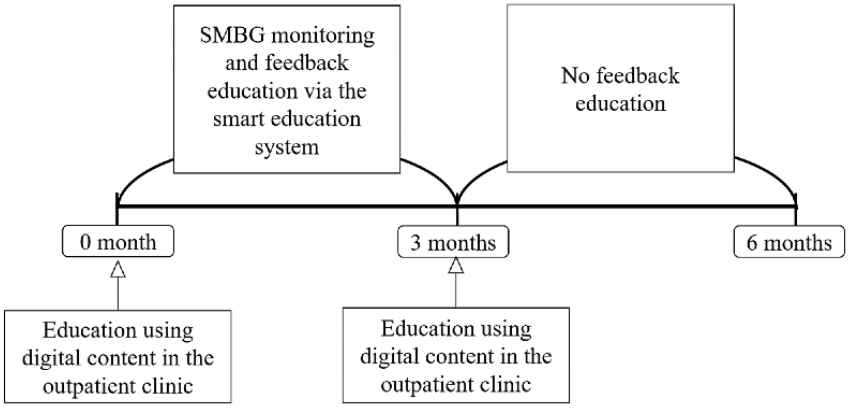
The control group received standard face-to-face care at the clinic, with diabetes education provided at registration, at the 3-month follow-up, and at the end of the study. Participants were trained on SMBG levels at enrollment and were asked to record these levels in their own notebooks. They did not use the digital education platform, Doctorvice, and therefore did not receive feedback from the physician.

Intervention Group—The Digital Education Platform User Group

In the intervention group, we aimed to assess the effects of a management model that integrates digital content–based education delivered through a platform during face-to-face care at outpatient clinics, combined with remote monitoring and feedback provided by the physician at home. Digital content–based education was provided at the time of patient registration, and the educational content was delivered through the patient's app (Doctorvice) or via SMS text messages. For the first 3 months, patients uploaded their SMBG data via the app, and the physician regularly provided appropriate feedback, including messages and content. Feedback was given at 2-3-week intervals, depending on the patient's SMBG levels. The feedback included newly created educational content tailored to each patient based on their physician's recommendations, covering topics such as diet control, physical activity, and diabetes self-care.

The platform allowed physicians to view patients' basic information at a glance, including BMI, medical history, family history, and alcohol and smoking history, as well as SMBG data, which facilitated personalized education. From months 3 to 6 (end of the study), patients were instructed to measure their own blood glucose levels and upload the values via the app; however, physicians were not allowed to provide feedback during this period. Thus, only self-management was maintained from months 3 to 6 (Figure 1). By structuring the study in this manner, we were able to evaluate the effects of remote education.

Figure 1. Study design in the intervention group. SMBG: self-measured blood glucose.



XSL•FO
RenderX

Digital Education Platform

The digital education platform (Doctorvice) serves as a medical consulting and support system that connects physicians and patients. It includes Doctorvice Clinic, a web-based platform accessible to physicians during outpatient clinic visits, and the Doctorvice app, which patients can use on their mobile devices. Physicians can use Doctorvice Clinic to review a patient's clinical information (eg, age, sex, medical history, family history, and smoking and alcohol history) at a glance, as well as monitor blood glucose and BP data uploaded by the patient (see Figure S1 in Multimedia Appendix 1).

The platform includes a digital content library containing approximately 3000 pieces of content related to diabetes and hypertension education. The content is categorized by disease and symptoms, diagnosis, treatment, side effects, medications, complications, prevention, and management. Physicians can use these contents to provide personalized education based on each patient's characteristics. Additionally, they can utilize the web canvas for further customization. Physicians can select content relevant to the patient's specific situation (eg, food intake, exercise, moderation in alcohol consumption, smoking cessation, and medication adherence) and place it on the canvas. They can also edit the content on the canvas by drawing pictures or writing text using the editing tool (see Figure S2 in Multimedia Appendix 1). This allows the creation of personalized educational content tailored to the patient, which can be sent via SMS text messages or the Doctorvice app. Patients can then access and review this content repeatedly on their smartphones.

Patients can upload their SMBG or BP data through the Doctorvice app (see Figure S3 in Multimedia Appendix 1). Physicians can monitor these uploaded data in real time via Doctorvice Clinic and provide feedback through messages and personalized content at any time. This platform fosters a continuous connection between physicians and patients by sharing content and data, facilitating both face-to-face and remote interactions as needed.

Questionnaires

This study selected 3 questionnaires to assess the clinical impact of the smart education platform in strengthening diabetes self-management behavior and measuring patient management patterns. The questionnaires included satisfaction with diabetes treatment, diabetes stress measurement, and adherence to diabetes medication (see Tables S1-S3 in Multimedia Appendix 1) [18-20]. Satisfaction with diabetes treatment was measured using 8 questions, with a maximum score of 48 points, where higher scores indicate greater satisfaction. Diabetes stress was assessed with 17 questions, with a maximum score of 85 points; higher scores reflect higher stress levels. Adherence to diabetes medication was evaluated with 8 questions, where a "no" answer received 1 point. The final question, "How often do you have difficulty remembering to take all your diabetes medications?" was scored from 4 (never) to 0 (always), for a total possible score of 11 points. Higher scores indicated greater adherence. These 3 surveys were administered at enrollment and again at the end of the study (6 months later) by the same research nurse to minimize investigator bias.

Outcomes

The primary outcome measure was the difference in HbA_{1c} levels between the control and intervention groups. Changes in HbA_{1c} levels from baseline to 3 months and from 3 to 6 months were analyzed. Subgroup analyses were performed based on SMBG upload frequency, age, baseline HbA_{1c} level, smoking status, alcohol consumption, and BMI. The secondary outcome was the difference in questionnaire results before and after the study.

Statistical Analysis

Statistical analyses were performed using SPSS (version 27.0; IBM Corp.) and Rex-Pro (version 3.6.0; Rexsoft, Co. Ltd.). The Kolmogorov-Smirnov test or the Q-Q plot was used to assess the normality of the data. Descriptive statistics are presented as means and SDs for continuous variables, and as frequencies with proportions for categorical variables. Two-tailed independent sample paired *t* tests and ANOVA were used to compare differences between groups for normally and nonnormally distributed data, respectively. Paired *t* tests and Wilcoxon tests were used to evaluate pre- and postintervention differences for normally and nonnormally distributed data, respectively. The Pearson correlation coefficient was applied to assess relationships between variables.

Ethical Considerations

This study was approved by the institutional review board at The Catholic University of Korea (approval number KC19FESI0604) in September 2019. Informed consent was obtained from all participants. To protect their privacy, an independent anonymization officer, who was not involved in the study intervention, was assigned to handle the anonymization process. No monetary compensation was provided to the participants.

Results

Baseline Characteristics

A total of 66 participants were initially enrolled in the study. After excluding 10 individuals who did not attend the outpatient clinic at 3 or 6 months, the final analysis included 26 (46%) participants in the intervention group and 30 (54%) participants in the control group. The mean age at baseline for these 56 participants was 60.3 years, and 42 (75%) were male (Table 1). The mean HbA_{1c} levels were 8.1% (SD 0.5%) overall, 8.3% (SD 0.6%) in the intervention group, and 8.0% (SD 0.5%) in the control group. The average BMI was 26.5 (SD 4.6) kg/m^2 in the intervention group and 25.5 (SD 3.4) kg/m^2 in the control group. There were no significant differences in the proportions of nonsmokers, ex-smokers, and current smokers (P=.28), nor in alcohol consumption (P=.28). The average number of diabetes medications was 2.9 (SD 0.7) in the intervention group and 2.8 (SD 0.8) in the control group. Of the 56 participants, 47 completed the surveys at both the beginning and end of the study, including 20 from the intervention group and 27 from the control group. Figure 2 presents the CONSORT (Consolidated Standards of Reporting Trials) flowchart (also see Multimedia Appendix 2 for the CONSORT checklist).

XSL•FO
RenderX

Table 1. Baseline clinical characteristics of the study participants.

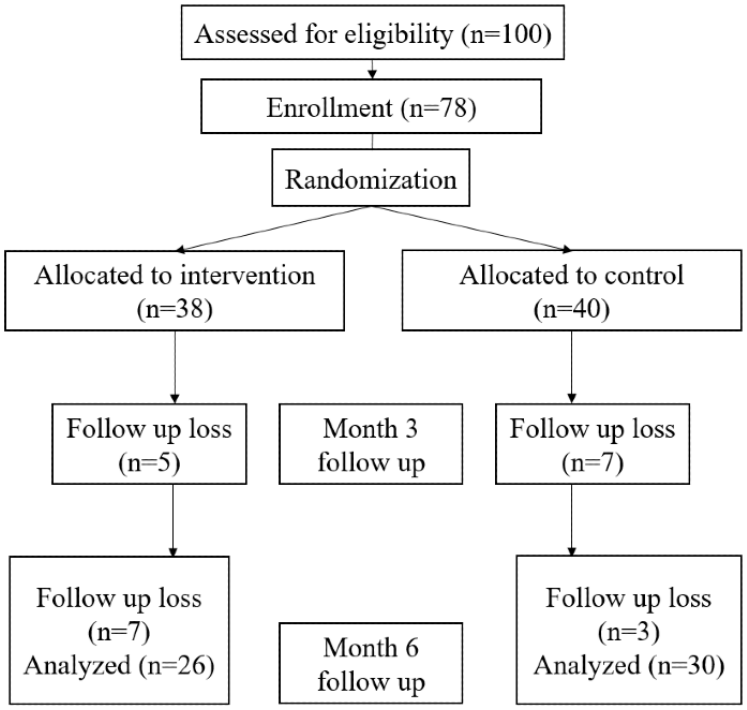
Characteristics	Total (N=56)	Intervention group (n=26)	Control group (n=30)	*P* value
Age (years), mean (SD)	60.3 (9.1)	59.0 (10.8)	61.3 (7.3)	.37
Male sex, n (%)	42 (75)	22 (85)	20 (67)	.22
Glucose (mg/dL), mean (SD)	154 (30)	159 (34)	150 (25)	.26
Glycated hemoglobin (HbA_{1c}[a]; %), mean (SD)	8.1 (0.5)	8.3 (0.6)	8.0 (0.5)	.10
Diabetes mellitus duration (years), mean (SD)	12.4 (7.0)	12.4 (7.5)	12.5 (6.6)	.94
Systolic blood pressure (mmHg), mean (SD)	124 (13)	126 (7)	123 (16)	.47
Diastolic blood pressure (mmHg), mean (SD)	73 (11)	76 (10)	70 (12)	.046
Urea nitrogen (mg/dL), mean (SD)	18.1 (5.5)	18.0 (6.2)	18.1 (4.9)	.95
Creatinine (mg/dL), mean (SD)	0.89 (0.27)	0.95 (0.30)	0.84 (0.25)	.15
Estimated glomerular filtration rate (mL/min/1.73 m^2), mean (SD)	87.6 (18.7)	86.1 (20.7)	88.8 (16.4)	.59
Aspartate aminotransferase (U/L), mean (SD)	27 (13)	30 (14)	24 (11)	.07
Alanine aminotransferase (U/L), mean (SD)	39 (57)	33 (19)	43 (77)	.48
Gamma-glutamyl transferase (U/L), mean (SD)	56 (110)	62 (88)	51 (129)	.70
Total cholesterol (mg/dL), mean (SD)	142 (31)	144 (34)	140 (28)	.62
Triglycerides (mg/dL), mean (SD)	139 (78)	156 (87)	124 (68)	.15
High-density lipoprotein cholesterol (mg/dL), mean (SD)	45 (11)	46 (12)	45 (10)	.60
Low-density lipoprotein cholesterol (mg/dL), mean (SD)	70 (24)	71 (25)	70 (25)	.88
Body weight (kg), mean (SD)	72.1 (14.5)	74.5 (17.3)	70.1 (11.5)	.28
BMI (kg/m^2), mean (SD)	26.0 (4.0)	26.5 (4.6)	25.5 (3.4)	.37
Number of oral hypoglycemic agents, mean (SD)	2.8 (0.7)	2.9 (0.7)	2.8 (0.8)	.95
Smoking, n (%)				.28
None	25 (44.6)	9 (35)	16 (53)	
Ex	22 (39.3)	13 (50)	9 (30)	
Current	9 (16.1)	4 (15)	5 (17)	
Alcohol consumption, n (%)	29 (52)	16 (62)	13 (43)	.73
Total education time (minutes), mean (SD)	N/A[b]	22.6 (9.7)	N/A	N/A
Number of education sessions, mean (SD)	N/A	4.4 (1.1)	N/A	N/A
Education time per session (minutes), mean (SD)	N/A	5.2 (2.1)	N/A	N/A
Number of education contents sent to patients, mean (SD)	N/A	24.8 (7.8)	N/A	N/A

[a]HbA_{1c}: glycated hemoglobin.

[b]N/A: not applicable.

XSL•FO
RenderX

Figure 2. CONSORT flowchart.



In the intervention group, the average duration of each educational session using the platform was 22.6 (SD 9.7) minutes, with an average of 4.4 (SD 1.1) sessions held over the first 3 months. Additionally, a mean of 24.8 (SD 7.8) pieces of educational content was sent to participants during this period, resulting in an average of 5.2 minutes of education per participant.

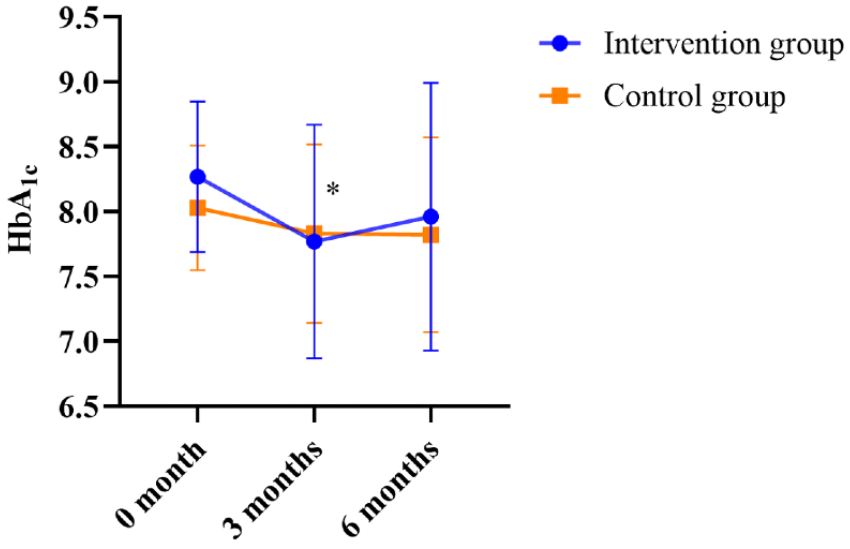
Although it was not possible to precisely measure the education time in the control group, given the unique characteristics of the outpatient care environment in Korea—where the average consultation time per patient typically ranges from 3 to 5 minutes—it can be assumed that the education time in the control group was similar to this range.

Change in HbA^{1c}

Changes in HbA^{1c} Over the First 3 Months

During the first 3 months, the intervention group experienced a significant reduction in HbA_{1c} levels, decreasing from a mean baseline level of 8.3% (SD 0.6%) to 7.8% (SD 0.9%), a change of –0.5% (*P*=.01; Figure 3). By contrast, the control group showed a nonsignificant decrease in mean HbA_{1c} levels, from 8.0% (SD 0.5%) to 7.8% (SD 0.7%), a change of –0.2% (*P*=.13).

Figure 3. Changes in glycated hemoglobin (HbA_{1c}) after education using the digital platform. *The change in HbA_{1c} from baseline to 3 months was significant in the intervention group (*P*=.01).



XSL•FO
RenderX

Changes in HbA^{1c} Between 3 and 6 Months

In the intervention group, HbA_{1c} levels showed a slight increase of 0.2% (from a mean of 7.8%, SD 0.9%, at 3 months to 8.0%, SD 1.0%, at 6 months), although this change was not statistically significant (*P*=.28). By contrast, the control group exhibited minimal change, with no significant difference in HbA_{1c} levels, remaining at a mean of 7.8% (SD 0.7%) at 3 months and a mean of 7.8% (SD 0.8%) at 6 months (*P*=.92).

Changes in HbA^{1c} Over 6 Months

At the 6-month follow-up, the HbA_{1c} levels in the intervention group decreased by 0.3% from baseline (*P*=.12), while in the control group, they decreased by 0.2% (*P*=.18). However, neither group showed a statistically significant reduction compared with baseline.

Analysis of Questionnaires

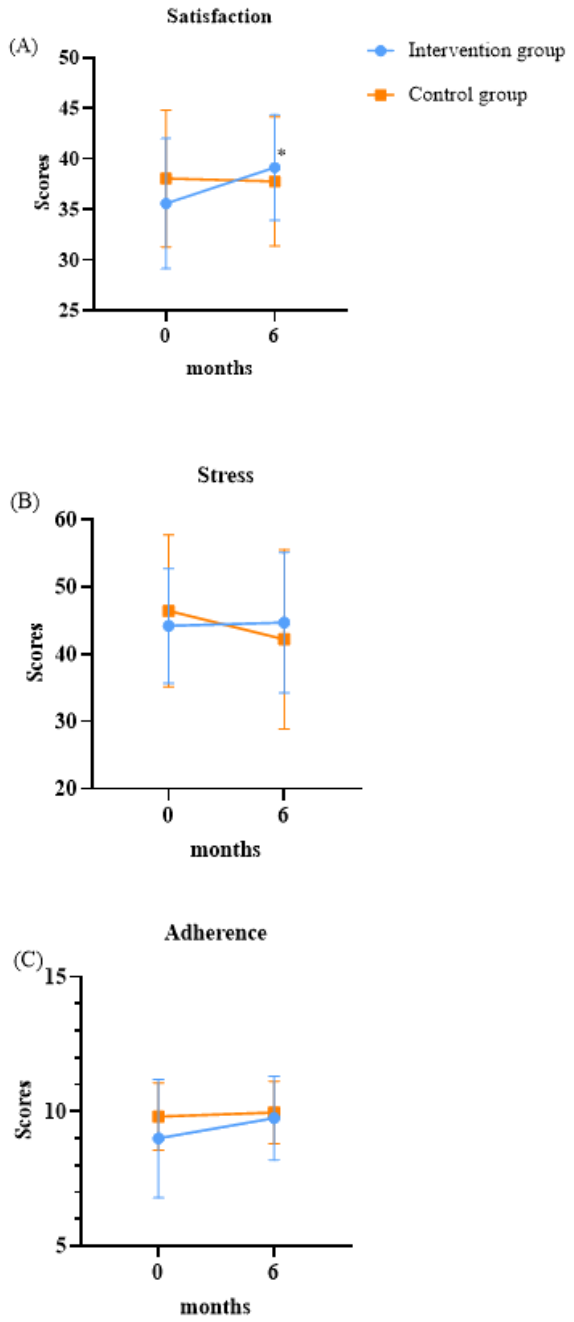
Satisfaction With Diabetes Treatment

The baseline Cronbach α for both the intervention and control groups was 0.8, indicating a high level of reliability. At the 6-month follow-up, the Cronbach α decreased to 0.6 for both groups, maintaining a moderate level of reliability.

The mean baseline satisfaction score for diabetes treatment was 35.6 (SD 6.4) points in the intervention group and 38.1 (SD 6.8) points in the control group (Figure 4A). At 6 months, the intervention group showed a significant increase of 3.6 points, reaching a mean of 39.2 (SD 5.2) points, while the control group experienced a slight decrease of 0.3 points, reaching a mean of 37.8 (SD 6.4) points. The intervention group demonstrated a significant improvement in satisfaction with diabetes treatment compared with the baseline (*P*=.006).

XSL•FO
RenderX

Figure 4. (A) Changes in the "satisfaction with diabetes treatment" score. (B) Changes in the "diabetes stress measurement" score. (C) Changes in the "adherence of diabetes medication" score. *The change in the satisfaction score was significant in the intervention group (*P*=.006). Vertical error bars indicate SD.



Diabetes Stress Measurement

The baseline Cronbach α was 0.7 for the intervention group and 0.9 for the control group, indicating a high level of reliability. At the 6-month follow-up, the Cronbach α increased to 0.8 in the intervention group, while it remained at 0.9 in the control group, maintaining a high level of reliability.

The mean baseline diabetes stress score was 44.2 (SD 8.5) points in the intervention group and 46.4 (SD 11.3) points in the control group (Figure 4B). At 6 months, the mean score in the intervention group was 44.7 (SD 10.4) points, while the control group had a mean score of 42.2 (SD 13.3) points. The change in stress scores differed significantly; the intervention group experienced an average change of 0.5 (SD 9.7), whereas the control group showed a decrease of –4.2 (SD 6.8). This indicates that stress levels improved more in the control group compared with the intervention group, with the difference approaching significance (*P*=.06).

Adherence to Diabetes Medication

The baseline Cronbach α was 0.6 for the intervention group and 0.4 for the control group. At the 6-month follow-up, Cronbach α decreased to 0.4 in the intervention group and 0.3 in the control group.

The mean baseline adherence score was 9.0 (SD 2.2) points in the intervention group and 9.8 (SD 1.2) points in the control group, with no significant difference between the 2 groups (*P*=.11). However, individuals in the control group exhibited

XSL•FO
RenderX

better adherence at the time of enrollment (Figure 4C). At the 6-month follow-up, the intervention group saw an increase of 0.8 points, reaching a mean of 9.8 (SD 1.6) points, while the control group's score increased by 0.2 points to a mean of 10.0 (SD 1.2) points. There was no significant difference in adherence scores between the 2 groups (*P*=.59), although adherence scores in the intervention group showed a greater tendency for improvement compared with the control group.

Correlation Analysis Between Questionnaires' Score and HbA^{1c}

Correlation analysis was conducted to explore the relationship between changes in HbA_{1c} levels and the scores from the survey results over the course of the study.

In the intervention group, the change in HbA_{1c} levels during the first 3 months showed a strong and significant (*P*=.17) inverse relationship with the change in HbA_{1c} levels during the 3-6 months (–0.7095), indicating that improvement in HbA_{1c} levels during the first 3 months was associated with deterioration in HbA_{1c} levels during the subsequent 3 months (Table 2). The change in stress scores was associated with the change in HbA_{1c} levels during the first 3 months (–0.3339), the entire 6-month study period (0.4372), and the 3-6-month period (0.6294). This suggests that during the first 3 months, improvement in HbA_{1c} levels was moderately associated with reduced stress. However, improvement in HbA_{1c} levels over the full 6 months and during the 3-6-month period was associated with increased stress. Additionally, although the correlation between HbA_{1c} improvement and adherence scores was not statistically significant (*P*=.17), there was a tendency for better adherence to be associated with improved HbA_{1c} levels in the intervention group during the first 3 months and over the entire 6-month period.

In the control group, the Pearson correlation between the degree of HbA_{1c} improvement over the 6-month period and the degree of stress improvement was 0.4475 (Table 3).

Table 2. The Pearson correlation in the intervention group.

Variable	HbA_{1c}[a] 0-3[b] (*P* value)	HbA_{1c} 0-6[c] (*P* value)	HbA_{1c} 3-6[d] (*P* value)	Satisfaction 0-6[e] (*P* value)	Stress 0-6[f] (*P* value)	Adherence 0-6[g] (*P* value)
HbA_{1c} 0-3	—[h]	0.2881 (.21)	–0.7095 (<.001)	–0.0097 (.97)	–0.3339 (.15)	–0.3160 (.17)
HbA_{1c} 0-6	0.2881 (.22)	—	0.4704 (.04)	0.0051 (.98)	0.4372 (.05)	–0.3229 (.17)
HbA_{1c} 3-6	–0.7095 (<.001)	0.4704 (.04)	—	0.0126 (.96)	0.6294 (.003)	0.0536 (.82)
Satisfaction 0-6	–0.0097 (.97)	0.0051 (.98)	0.0126 (.96)	—	–0.3945 (.09)	0.5198 (.02)
Stress 0-6	–0.3339 (.15)	0.4372 (.05)	0.6294 (.003)	–0.3945 (.09)	—	–0.3795 (.10)
Adherence 0-6	–0.3160 (.17)	–0.3229 (.17)	0.0536 (.82)	0.5198(.02)	–0.3795 (.10)	—

[a]HbA_{1c}: glycated hemoglobin.

[b]HbA_{1c} 0-3: changes in HbA_{1c} during the first 3 months.

[c]HbA_{1c} 0-6: changes in HbA_{1c} during the first 6 months.

[d]HbA_{1c} 3-6: changes in HbA_{1c} during 3-6 months.

[e]Satisfaction 0-6: changes in satisfaction with diabetes treatment over 6 months.

[f]Stress 0-6: changes in diabetes stress measurement over 6 months.

[g]Adherence 0-6: changes in adherence to diabetes medications over 6 months.

[h]Not applicable.

XSL•FO
RenderX

Table 3. The Pearson correlation in the control group.

Variable	HbA_{1c}[a] 0-3[b] (*P* value)	HbA_{1c} 0-6[c] (*P* value)	HbA_{1c} 3-6[d] (*P* value)	Satisfaction 0-6[e] (*P* value)	Stress 0-6[f] (*P* value)	Adherence 0-6[g] (*P* value)
HbA_{1c} 0-3	—[h]	0.7331 (<.001)	–0.1907 (.34)	0.0133 (.95)	0.3083 (.12)	0.0310 (.88)
HbA_{1c} 0-6	0.7331 (<.001)	—	0.5278 (.005)	–0.1690 (.40)	0.4475 (.02)	0.0571 (.78)
HbA_{1c} 3-6	–0.1907 (.34)	0.5278 (.005)	—	–0.2605 (.19)	0.2608 (.19)	0.0437 (.83)
Satisfaction 0-6	0.0133 (.95)	–0.1690 (.40)	–0.2605 (.19)	—	–0.1907 (.34)	0.1357 (.50)
Stress 0-6	0.3083 (.12)	0.4475 (.02)	0.2608 (.19)	–0.1907 (.34)	—	0.0051 (.98)
Adherence 0-6	0.0310 (.88)	0.0571 (.78)	0.0437 (.83)	0.1357 (.50)	0.0051 (.98)	—

[a]HbA_{1c}: glycated hemoglobin.

[b]HbA_{1c} 0-3: changes in HbA_{1c} during the first 3 months.

[c]HbA_{1c} 0-6: changes in HbA_{1c} during the first 6 months.

[d]HbA_{1c} 3-6: changes in HbA_{1c} during 3-6 months.

[e]Satisfaction 0-6: changes in satisfaction with diabetes treatment over 6 months.

[f]Stress 0-6: changes in diabetes stress measurement over 6 months.

[g]Adherence 0-6: changes in adherence to diabetes medications over 6 months.

[h]Not applicable.

Subgroup Analysis

Change in HbA[1c] According to SMBG Upload Frequency

Participants who uploaded their blood glucose levels via the app more than once a week were classified as “high,” whereas those who uploaded less frequently were classified as “low.” The intervention group was subdivided based on their upload frequency in the first and last 3 months: group 1 consisted of participants classified as “high” for both periods, group 2 as “high” for the first 3 months and “low” for the last 3 months, group 3 as “low” for the first 3 months and “high” for the last 3 months, and group 4 as “low” for both periods. In the intervention group, there were 4 patients in group 1, 6 in group 2, 4 in group 3, and 12 in group 4. Compared with baseline, HbA_{1c} levels improved by –1.3% in group 1, –0.5% in group 2, and –0.5% in group 4 at 3 months, while HbA_{1c} increased by 0.2% in group 3. Notably, patients classified as “high” throughout the 6 months (group 1) had a significantly greater reduction in HbA_{1c} compared with those classified as “low” for the entire 6-month period (group 4; post hoc analysis *P*=.02; Table 4, Figure 5). The control group could not be analyzed for this metric due to the absence of SMBG uploads via the app.

Table 4. Changes in HbA_{1c}[a] according to SMBG[b] upload frequency.

Group	Baseline, mean (SD)	3 months, mean (SD)	*P* value	Δ[c] (baseline to 3 months) (SD)	6 months, mean (SD)	*P* value	Δ (3-6 months) (SD)
Group 1 (n=4)	8.4 (0.7)	7.1 (0.9)	.14	–1.3 (1.3)	6.9 (0.8)	.49	–0.2 (0.6)
Group 2 (n=6)	8.4 (0.6)	7.9 (1.3)	.33	–0.5 (1.1)	8.1 (1.2)	.79	0.1 (1.0)
Group 3 (n=4)	8.2 (0.4)	8.3 (0.4)	.64	0.2 (0.6)	8.1 (0.5)	.43	–0.3 (0.5)
Group 4 (n=12)	8.2 (0.6)	7.7 (0.7)	.03	–0.5 (0.7)	8.3 (1.0)	.08	0.5 (0.9)

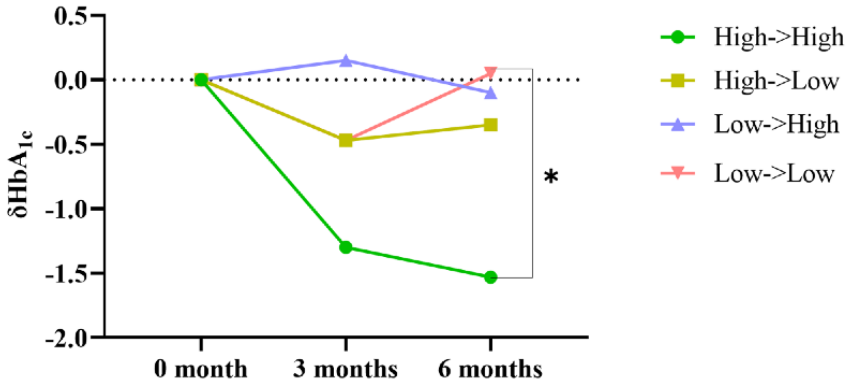
[a]HbA_{1c}: glycated hemoglobin.

[b]SMBG: self-measured blood glucose.

[c]Δ: difference in HbA_{1c}.

XSL•FO
RenderX

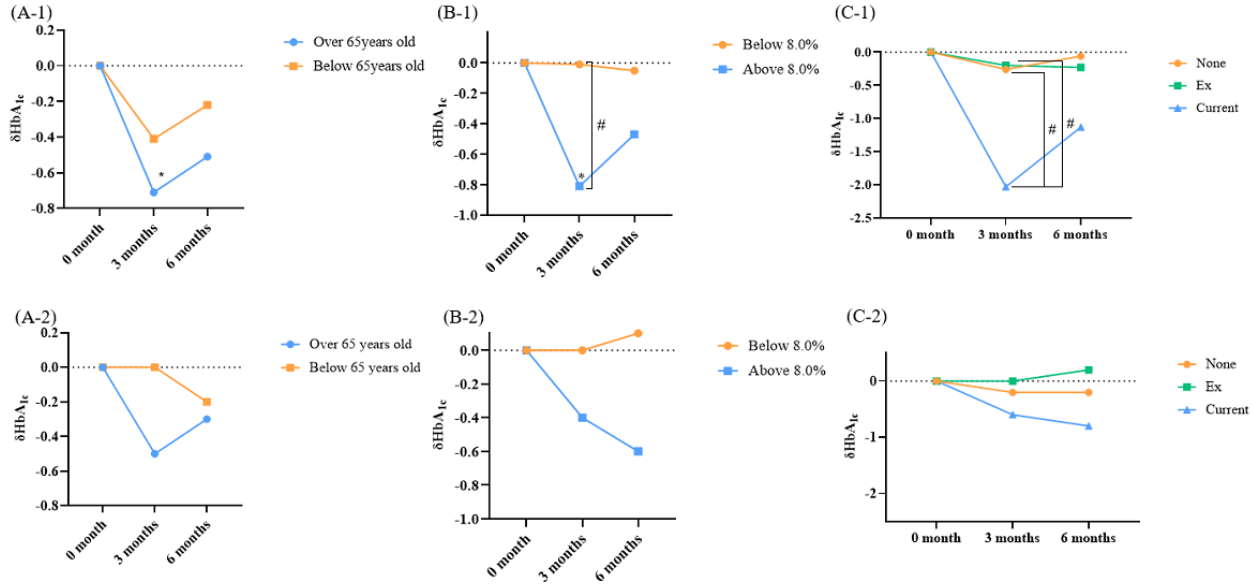
Figure 5. Changes in glycated hemoglobin (HbA_{1c}) according to self-measured blood glucose (SMBG) upload frequency. *Group 1 had a significantly greater reduction in HbA_{1c} compared with group 4 (post hoc analysis *P*=.02).



Change in HbA^{1c} According to Age

The degree of improvement in HbA_{1c} levels was assessed by classifying patients into 2 age groups: those aged <65 years and ≥65 years. In the intervention group, there were 18 individuals aged <65 years, with a mean baseline HbA_{1c} level of 8.1% (SD 0.6%; Figure 6A-1). At 3 months, HbA_{1c} improved by –0.4%, reaching a mean of 7.7% (SD 1.0%; *P*=.10). In the intervention group, there were 8 individuals aged ≥65 years, with a mean baseline HbA_{1c} level of 8.6% (SD 0.5%). At 3 months, their HbA_{1c} improved by –0.7%, reaching a mean of 7.9% (SD 0.7%; *P*=.03). In the control group, there were 19 individuals aged <65 years, with a mean baseline HbA_{1c} level of 8.0% (SD 0.5%). Additionally, there were 11 individuals aged ≥65 years, with a mean baseline HbA_{1c} level of 8.0% (SD 0.6%). There was no significant difference in HbA_{1c} improvement between the age groups in the control group (*P*=.49; Figure 6A-2).

Figure 6. (A-1) Changes in HbA_{1c} according to age (≥65 years vs <65 years) in the intervention group. (A-2) Changes in HbA_{1c} according to age (≥65 years vs <65 years) in the control group. (B-1) Changes in HbA_{1c} according to baseline HbA_{1c} (8.0%) in the intervention group. (B-2) Changes in HbA_{1c} according to baseline HbA_{1c} (8.0%) in the control group. (C-1) According to smoking status in the intervention group. (C-2) Changes in HbA_{1c} according to smoking status in the control group. (D-1) Changes in HbA_{1c} according to drinking status in the intervention group. (D-2) Changes in HbA_{1c} according to drinking status in the control group. (E-1) Changes in HbA_{1c} according to BMI level in the intervention group. (E-2) Changes in HbA_{1c} according to BMI level in the control group. *Compared with baseline, change in HbA_{1c} at 3 months was significant in the intervention group. #There was a significant difference in the degree of HbA_{1c} change between the 2 groups. HbA_{1c}: glycated hemoglobin.



Change in HbA^{1c} According to Baseline HbA^{1c}

In the intervention group, 16 participants had an HbA_{1c} level ≥8.0%, while 10 participants had an HbA_{1c} level <8.0%. At 3 months, the group with HbA_{1c} ≥8.0% showed an improvement of –0.8% (*P*=.006) compared with baseline, while the group with HbA_{1c} <8.0% showed only a slight change (*P*=.95) compared with baseline. A significant difference was observed between the 2 groups (*P*=.03; Figure 6B-1). In the control group, 13 patients had an HbA_{1c} level ≥8.0%, while 17 patients had an HbA_{1c} level <8.0%. No significant difference was observed between the 2 groups (*P*=.30; Figure 6B-2).

XSL•FO
RenderX

Change in HbA^{1c} According to the Smoking Status

In the intervention group, 9 participants were nonsmokers, 13 were ex-smokers, and 4 were current smokers. At 3 months, the mean improvement in HbA_{1c} compared with baseline was –0.3% (SD 0.5%) in nonsmokers, –0.2% (SD 0.6%) in ex-smokers, and –2.0% (SD 1.0%) in current smokers (Figure 6C-1). Post hoc analysis revealed significant differences between current smokers and both nonsmokers (P<.001) and ex-smokers (P<.001), but no significant difference between nonsmokers and ex-smokers (P=.98)

In the control group, 16 participants were nonsmokers, 9 were ex-smokers, and 5 were current smokers. Changes in HbA_{1c} did not differ significantly among the smoking status groups (post hoc analysis: noncurrent, P=.52; ex-current, P=.30; and non-ex, P=.76; Figure 6C-2).

Change in HbA^{1c} According to Alcohol Consumption

In the intervention group, there were 10 nondrinkers and 16 alcohol drinkers. At 3 months, the improvement in HbA_{1c} compared with baseline was –0.1% in nondrinkers and –0.8% in alcohol drinkers (P=.06; Figure 6D-1). The control group included 17 nondrinkers and 13 drinkers. No significant difference was observed between the 2 groups (P=.51; Figure 6D-2).

Change in HbA^{1c} According to the Baseline BMI Level

Based on baseline BMI, participants were classified into 4 groups: group 1 (normal weight), group 2 (preobesity), group 3 (obesity stage 1), and group 4 (obesity stages 2 and 3), with no participants classified as underweight in any group. In the intervention group, there were 6 participants in group 1, 5 in group 2, 11 in group 3, and 4 in group 4. At 3 months, HbA_{1c} improvement was more pronounced in group 4 compared with the other groups, though this difference was not statistically significant (post hoc analysis: groups 1-2, P=.66; groups 1-3, P=.80; groups 1-4, P=.35; groups 2-3, P=.97; groups 2-4, P=.06; and groups, 3-4, P=.06; Figure 6E-1). In the control group, with 7 participants in group 1, 10 in group 2, 10 in group 3, and 3 in group 4, no significant differences in HbA_{1c} improvement were observed among the groups (P=.37; Figure 6E-2).

Discussion

In this pilot randomized controlled trial involving patients with T2DM, a digital education platform combining face-to-face education with remote feedback over a 3-month period resulted in a significantly greater reduction in HbA_{1c} (–0.5%) compared with conventional education (–0.2%). However, HbA_{1c} levels in the intervention group slightly increased after remote feedback ended. Treatment satisfaction improved significantly in the intervention group, while stress decreased more in the control group. Medication adherence improved modestly in both groups, with no significant between-group differences. Greater HbA_{1c} improvement was associated with more frequent SMBG uploads and was more pronounced in participants with higher baseline HbA_{1c}, age ≥65 years, smoking, drinking, or obesity.

This study demonstrated that the digital education platform for personalized diabetes management was beneficial for glycemic control in T2DM and was fairly effective when physicians provided both personalized face-to-face education and remote monitoring and feedback. A significant decrease in HbA_{1c} was observed during the first 3 months, when face-to-face education combined with remote monitoring and feedback was provided through this platform. However, during the 3-6-month period, when education was limited to face-to-face sessions at the outpatient clinic without further feedback from physicians, HbA_{1c} levels tended to increase slightly. Even in the control group, which received only regular face-to-face education at the outpatient clinic, HbA_{1c} levels decreased over the first 3 months, suggesting that initial education may be effective in diabetes management. Nevertheless, the more significant reduction in HbA_{1c} levels observed in the intervention group can be attributed to the additional effect of SMBG uploads and feedback from physicians. The effectiveness of education was enhanced for the intervention group by enabling more continuous learning, as they received digital education and educational materials via SMS text messages or the Doctorvice app. This contributed to a more significant reduction in HbA_{1c} levels. One study has noted that fostering user engagement is essential for diabetes self-management apps to effectively complement clinical care [21]. Another study emphasized that outcomes in diabetes management varied depending on the level of message-based interaction [22]. These findings align with the results of our study, underscoring the importance of patient-centered, personalized, and sustained interaction for optimal diabetes care. The slight increase in HbA_{1c} levels following the discontinuation of remote feedback in our intervention group further reinforces the critical role of ongoing feedback in supporting long-term glycemic control.

In this study, at 3 months, the difference in HbA_{1c} levels between the 2 groups was approximately –0.3%, suggesting that the additional education on diabetes management provided during this period produced results comparable to those of other studies. In a systematic review and meta-analysis of diabetes management using mobile phone apps [23], the difference in HbA_{1c} between the intervention and control groups ranged from –0.69% to –0.28%. Additionally, another meta-analysis indicated that improvements in HbA_{1c} levels occurred only when clinical interventions, such as medication adjustments, were included [24]. Although no medication adjustments were made during this study, the observed improvement in HbA_{1c} levels aligns with findings from other studies that emphasize the importance of physician involvement in diabetes management. In the intervention group, HbA_{1c} levels tended to increase again during the 3- to 6-month period, when remote feedback was no longer provided. This trend aligns with findings from other studies [25,26], which reported the greatest effects during short-term study periods. Furthermore, these studies suggested that the most significant improvements are typically observed around 6 months. The decline in educational effect observed after discontinuing the intervention highlights the importance of sustained support. These findings suggest that

XSL•FO
RenderX

health care providers should offer long-term remote education to maintain and reinforce behavioral changes in diabetes care.

Survey results indicated that satisfaction with diabetes treatment and adherence to diabetes medications tended to improve in the intervention group compared with the control group. The non–face-to-face education provided outside of the outpatient clinic served as an additional treatment component, tailored to each patient, which likely contributed to these improvements. However, the reliability of the diabetes medication adherence survey was low, which may be attributed to the small number of items. Additionally, the use of a dichotomous response format (yes/no) may have contributed to the lower reliability. In the diabetes stress measurement, stress scores significantly improved in the control group. The Pearson correlation analysis revealed a negative correlation between changes in HbA_{1c} levels during the first 3 months and stress in the intervention group. The diabetes stress measurement included items on the emotional burden, physician-related distress, regimen-related distress, and diabetes-related interpersonal distress. Although analyzing the changes in scores for each category would have been beneficial, this was not conducted. In the intervention group, the increase in stress may have been influenced by continuous contact, which could have heightened the emotional burden and exacerbated regimen-related distress due to the discrepancy between personal lifestyle and educational content. By contrast, the control group may have perceived the overall experience more positively. Nevertheless, improvements in treatment satisfaction and adherence suggest that the digital intervention still provided meaningful benefits, consistent with previous studies reporting positive effects of digital health interventions on the quality of life in patients with type 2 diabetes [27,28]. A more detailed analysis of individual item scores could have offered a clearer explanation of these findings.

SMBG upload frequency, inferred from blood glucose level uploads via the app, indicated that participants who consistently uploaded their SMBG data over 6 months experienced better HbA_{1c} improvement compared with those with infrequent uploads.

In addition, the effect of the digital education platform on reducing HbA_{1c} levels was notably greater for individuals with HbA_{1c} levels ≥8.0% and those aged ≥65 years. However, some studies suggest that the effect of digital education is more pronounced in patients with higher baseline HbA_{1c} levels, younger age, and shorter disease duration [29], while others indicate that it is effective in older adults regardless of baseline HbA_{1c} levels [30]. The optimal age group for the effect of the digital education platform has not yet been clearly established. Our research suggests that simultaneous face-to-face education and remote feedback can enhance effectiveness, even in older adults. As widely recognized, higher baseline HbA_{1c} levels are associated with greater educational effects. Similarly, our study demonstrated that the effectiveness of the digital education platform was more pronounced in individuals with higher baseline HbA_{1c} levels.

Interestingly, current smokers in the intervention group experienced a more significant reduction in HbA_{1c} levels compared with nonsmokers and ex-smokers. Additionally, there was a trend toward decreased HbA_{1c} levels in alcohol drinkers compared with nondrinkers. Furthermore, the HbA_{1c} reduction in the intervention group was more pronounced among those with a BMI ≥30 kg/m^2, suggesting that the digital education system may have been particularly effective in encouraging lifestyle improvements through continuous education, especially for individuals with less healthy habits.

The strength of our study lies in the additional survey conducted in the outpatient clinic, which allowed us to confirm the use of the digital education system and assess patients' health status, behaviors, and adherence from multiple perspectives. This comprehensive approach facilitates a thorough analysis of patient self-management. Unlike previous studies, which did not establish the effect of diabetes self-management through digital education on smoking or other lifestyle groups, our study conducted detailed subgroup analyses to identify specific groups where digital education could be effective. As the age group of smartphone users gradually expands, it is crucial for seniors to actively engage with digital health care. Notably, our study included participants aged ≥65 years, addressing a gap in the existing research. The main strength of this study lies in its attempt to examine the effectiveness of the digital education system across various factors, such as baseline HbA_{1c}, age, and lifestyle habits.

However, our study had limitations. The sample size was smaller than initially targeted, with fewer than 30 participants in each group after accounting for dropouts. This small sample size limited the ability to conduct more detailed subgroup analyses and reduced the statistical power of our findings. Additionally, because the patient intervention period was only 6 months—relatively short for evaluating changes in lifestyle habits—only blood glucose control was assessed in diabetes management. Besides, the exact number of SMBGs performed by the control group was unknown, complicating the analysis of improvements in blood glucose control. Another limitation was the insufficient assessment of the digital education platform itself; collecting feedback on satisfaction, needs, and evaluation of the educational content through surveys would have been beneficial for system development. Conducting a survey in the third month would have allowed for a more accurate comparison of the platform's functions and survey results. Additionally, it would have been beneficial to include not only HbA_{1c} and the 3 survey results but also lifestyle improvements. Further analysis of changes in individual items within each questionnaire would have provided a more detailed understanding of the results. Moreover, analyzing the data according to the categories of educational content would have strengthened the findings.

In conclusion, personalized diabetes management using a digital education system appears to be beneficial for glycemic control in patients with T2DM. The system was particularly effective for individuals who had not achieved optimal glycemic control, were older, or had poorer lifestyle habits. Future research should focus on evaluating the efficacy of this digital education platform in individuals with diverse clinical profiles and conducting longer-term follow-up studies with larger participant groups.

XSL•FO
RenderX

Acknowledgments

This study was selected as the “first research support project” supported by the Research Fund of Seoul St. Mary’s Hospital, The Catholic University of Korea from October 1, 2022, to September 30, 2023. Further, this research was supported by a grant from the Korean Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant RS-RS-2023-KH135944). Yuhan Inc. provided the research funding in August 2019. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JY and JHC conceived and designed the study. Data were collected by JY, JL, YY, EYL, SHL, and JHC, while JY curated and prepared the data for analysis. JY and JHC conducted the statistical analyses and interpreted the results. JY created the figures and visualizations used in the manuscript and also drafted the original manuscript, which was subsequently reviewed and edited by both JY and JHC.

Conflicts of Interest

JHC is the CEO of iKoob Inc. The other authors have nothing to declare.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 801 KB - mhealth_v13i1e67151_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2727 KB - mhealth_v13i1e67151_app2.pdf]

References

1. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Research and Clinical Practice 2022 Jan;183:109119 [FREE Full text] [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
2. Choi JH, Lee KA, Moon JH, Chon S, Kim DJ, Kim HJ, et al. 2023 clinical practice guidelines for diabetes mellitus of the Korean Diabetes Association. Diabetes Metab J 2023 Sep;47(5):575-594 [FREE Full text] [doi: 10.4093/dmj.2023.0282] [Medline: 37793979]
3. Chatterjee S, Davies MJ, Heller S, Speight J, Snoek FJ, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Feb;6(2):130-142. [doi: 10.1016/S2213-8587(17)30239-5] [Medline: 28970034]
4. American Diabetes Association Professional Practice Committee. 5. Facilitating positive health behaviors and well-being to improve health outcomes: standards of care in diabetes-2024. Diabetes Care 2024 Jan 01;47(Suppl 1):S77-S110. [doi: 10.2337/dc24-S005] [Medline: 38078584]
5. Deakin TA, Cade JE, Williams R, Greenwood DC. Structured patient education: the diabetes X-PERT Programme makes a difference. Diabet Med 2006 Sep;23(9):944-954 [FREE Full text] [doi: 10.1111/j.1464-5491.2006.01906.x] [Medline: 16922700]
6. Skinner TC, Carey ME, Cradock S, Daly H, Davies MJ, Doherty Y, DESMOND Collaborative. Diabetes Education and Self-Management for Ongoing and Newly Diagnosed (DESMOND): process modelling of pilot study. Patient Educ Couns 2006 Dec;64(1-3):369-377. [doi: 10.1016/j.pec.2006.04.007] [Medline: 17011154]
7. Ko S, Song K, Kim S, Lee J, Kim J, Shin J, et al. Long-term effects of a structured intensive diabetes education programme (SIDEP) in patients with type 2 diabetes mellitus--a 4-year follow-up study. Diabet Med 2007 Jan 10;24(1):55-62. [doi: 10.1111/j.1464-5491.2007.02013.x] [Medline: 17227325]
8. Gershkowitz B, Hillert C, Crotty B. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021 Mar 25;106(4):e1513-e1520. [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]
9. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 2018 May;6(5):361-369. [doi: 10.1016/S2213-8587(18)30051-2] [Medline: 29503172]

XSL•FO
RenderX

10. Pang M, Li Y, Gu W, Sun Z, Wang Z, Li L. Recent advances in epigenetics of macrovascular complications in diabetes mellitus. Heart Lung Circ 2021 Feb;30(2):186-196. [doi: 10.1016/j.hlc.2020.07.015] [Medline: 32873490]
11. Gupta A, Behl T, Sehgal A, Bhardwaj S, Singh S, Sharma N, et al. Exploring the recent molecular targets for diabetes and associated complications. Mol Biol Rep 2021 Mar 24;48(3):2863-2879. [doi: 10.1007/s11033-021-06294-0] [Medline: 33763776]
12. Harding JL, Pavkov ME, Magliano DJ, Shaw JE, Gregg EW. Global trends in diabetes complications: a review of current evidence. Diabetologia 2019 Jan;62(1):3-16 [FREE Full text] [doi: 10.1007/s00125-018-4711-2] [Medline: 30171279]
13. Patton SR. Adherence to glycemic monitoring in diabetes. J Diabetes Sci Technol 2015 May;9(3):668-675 [FREE Full text] [doi: 10.1177/1932296814567709] [Medline: 25591853]
14. Maharaj A, Lim D, Murphy R, Serlachius A. Comparing two commercially available diabetes apps to explore challenges in user engagement: randomized controlled feasibility study. JMIR Form Res 2021 Jun 16;5(6):e25151. [doi: 10.2196/25151] [Medline: 34132640]
15. Al Omar M, Hasan S, Palaian S, Mahameed S. The impact of a self-management educational program coordinated through WhatsApp on diabetes control. Pharm Pract (Granada) 2020 May 03;18(2):1841. [doi: 10.18549/pharmpract.2020.2.1841]
16. Xia S, Maitiniyazi G, Chen Y, Wu X, Zhang Y, Zhang X, et al. Web-based TangPlan and WeChat combination to support self-management for patients with type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2022 Mar 30;10(3):e30571. [doi: 10.2196/30571] [Medline: 35353055]
17. iKooB. Introducing Doctorvice Clinic. iKooB. URL: https://tinyurl.com/2tupwp29 [accessed 2025-04-23]
18. Bradley C. Diabetes treatment satisfaction questionnaire. Change version for use alongside status version provides appropriate solution where ceiling effects occur. Diabetes Care 1999 Mar;22(3):530-532. [doi: 10.2337/diacare.22.3.530] [Medline: 10097946]
19. Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J, et al. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. Diabetes Care 2005 Mar 25;28(3):626-631. [doi: 10.2337/diacare.28.3.626] [Medline: 15735199]
20. Lee W, Ahn J, Kim J, Hong Y, Hong SK, Kim YT, et al. Reliability and validity of a self-reported measure of medication adherence in patients with type 2 diabetes mellitus in Korea. J Int Med Res 2013 Aug 16;41(4):1098-1110 [FREE Full text] [doi: 10.1177/0300060513484433] [Medline: 23860015]
21. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. User experiences with a type 2 diabetes coaching app: qualitative study. JMIR Diabetes 2020 Jul 17;5(3):e16692 [FREE Full text] [doi: 10.2196/16692] [Medline: 32706649]
22. Miele F, Clementi S, Gennaro R, Nicolao I, Romanelli T, Speese K, et al. Text messaging and type 1 diabetes management: qualitative study exploring interactions among patients and health care professionals. JMIR Diabetes 2019 May 10;4(2):e11343. [doi: 10.2196/11343] [Medline: 31094332]
23. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
24. Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, Alberta Kidney Disease Network. Effect of telemedicine on glycated hemoglobin in diabetes: a systematic review and meta-analysis of randomized trials. CMAJ 2017 Mar 06;189(9):E341-E364 [FREE Full text] [doi: 10.1503/cmaj.150885] [Medline: 27799615]
25. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun 22;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]
26. Minet L, Møller S, Vach W, Wagner L, Henriksen JE. Mediating the effect of self-care management intervention in type 2 diabetes: a meta-analysis of 47 randomised controlled trials. Patient Education and Counseling 2010 Jul;80(1):29-41. [doi: 10.1016/j.pec.2009.09.033] [Medline: 19906503]
27. Skovlund SE, Nicolucci A, Balk-Møller N, Berthelsen DB, Glümer C, Perrild H, et al. Perceived benefits, barriers, and facilitators of a digital patient-reported outcomes tool for routine diabetes care: protocol for a national, multicenter, mixed methods implementation study. JMIR Res Protoc 2021 Sep 03;10(9):e28391 [FREE Full text] [doi: 10.2196/28391] [Medline: 34477563]
28. Abdelhameed F, Pearson E, Parsons N, Barber TM, Panesar A, Summers C, et al. Health outcomes following engagement with a digital health tool among people with prediabetes and type 2 diabetes: prospective evaluation study. JMIR Diabetes 2023 Dec 28;8:e47224. [doi: 10.2196/47224] [Medline: 38016426]
29. Yang Y, Lee EY, Kim H, Lee S, Yoon K, Cho J. Effect of a mobile phone-based glucose-monitoring and feedback system for type 2 diabetes management in multiple primary care clinic settings: cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Feb 26;8(2):e16266 [FREE Full text] [doi: 10.2196/16266] [Medline: 32130172]
30. Quinn CC, Shardell MD, Terrin ML, Barr EA, Park D, Shaikh F, et al. Mobile diabetes intervention for glycemic control in 45- to 64-year-old persons with type 2 diabetes. J Appl Gerontol 2016 Feb;35(2):227-243. [doi: 10.1177/0733464814542611] [Medline: 25098253]

XSL•FO
RenderX

Abbreviations

ADA: The American Diabetes Association
BP: blood pressure
CONSORT: Consolidated Standards of Reporting Trials
DESMOND: Diabetes Education and Self-Management for Ongoing and Newly Diagnosed
DSMES: diabetes self-management education and support
HbA_{1c}: glycated hemoglobin
SIDEP: Structured Intensive Diabetes Education Program
SMBG: self-measured blood glucose
T2DM: type 2 diabetes mellitus

Edited by L Buis; submitted 04.10.24; peer-reviewed by SD Huisman, MA Virtanen; comments to author 18.12.24; revised version received 11.02.25; accepted 14.04.25; published 01.05.25.

Please cite as:
Yu J, Lee J, Yang Y, Lee EY, Lee SH, Cho JH
Clinical Impact of Personalized Physician's Education and Remote Feedback Via a Digital Platform on Glycemic Control: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e67151
URL: https://mhealth.jmir.org/2025/1/e67151
doi:10.2196/67151
PMID:40310669

©Jin Yu, Joonyub Lee, Yeoree Yang, Eun Young Lee, Seung-Hwan Lee, Jae-Hyoung Cho. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 01.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effect of a Mobile App (eMOM) on Self-Discovery and Psychological Factors in Persons With Gestational Diabetes: Mixed Methods Study

Sini Määttänen[1], MD; Saila Koivusalo[2,3], MD, PhD; Hanna Ylinen[4], MSc; Seppo Heinonen[2], MD, Prof Dr Med; Mikko Kytö[5,6], PhD

[1]Faculty of Medicine, University of Helsinki, Helsinki, Finland

[2]Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland

[3]Shared Group Services, Helsinki University Hospital, Helsinki, Finland

[4]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

[5]Development and Strategy Unit, Helsinki University Hospital, Helsinki, Finland

[6]Department of Computer Science, University of Helsinki, Helsinki, Finland

Corresponding Author:
Sini Määttänen, MD
Faculty of Medicine
University of Helsinki
Haartmaninkatu 8
Helsinki, 00014
Finland
Phone: 358 407594806
Email: sini.maattanen@helsinki.fi

Abstract

Background: Gestational diabetes is a type of diabetes that develops during pregnancy and increases the risk of developing type 2 diabetes later in life. The rising prevalence of gestational diabetes mellitus (GDM) highlights the need for more comprehensive treatment strategies, with a particular emphasis on supporting maternal self-management. We showed recently that a mobile app, eMOM, where glucose, nutrition, and physical activity are combined within a single app, significantly improves multiple clinical outcomes among persons with gestational diabetes.

Objective: This study aims to explore the effects of the eMOM on maternal self-discovery and learning, autonomous motivation to manage GDM, and psychological well-being. Additionally, we examine the correlation between improved maternal clinical outcomes and change in autonomous motivation. We also assess the acceptance and usability of the eMOM app.

Methods: Building upon the original randomized controlled trial (RCT), in which the intervention arm used a mobile app (eMOM), we conducted a mixed methods study that included an investigation of eMOM log files, semistructured interviews on self-discovery, and an examination of questionnaires assessing motivation (Treatment Self-Regulation Questionnaire and Perceived Competence Scale), depression (Edinburgh Postnatal Depression Scale), technology use and acceptance (Unified Theory of Acceptance of Use of Technology questionnaire), and usability (modified Software Usability Measurement Inventory). Additionally, we monitored participants' stress levels using wearable electrocardiographic devices (FirstBeat Bodyguard 2). A total of 148 individuals participated in the original RCTs, with 76 in the intervention arm and 72 in the control arm. From the intervention arm, 18 participants were randomly selected for interviews in this study.

Results: Results show that the use rate of eMOM was high, and novel visualization supported self-discovery in persons with GDM. Most participants (17/18, 94%) indicated that the eMOM app helped to find the associations between their daily activities and glucose levels. Especially having nutrition visualized together with glucose was highly appreciated. Participants also reported learning about the associations between physical activity and glucose levels. No differences were observed between the intervention and control arms in autonomous motivation, depression, or stress. Furthermore, there were no correlations between improved clinical outcomes and changes in motivation. Accessibility and usability ratings were consistently high throughout the intervention.

Conclusions: The eMOM mobile app combining data from continuous glucose monitor, food diary, and physical activity tracker supports maternal self-discovery related to GDM without contributing to depression or adding extra stress. This encourages the

XSL•FO
RenderX

use of such mobile apps in maternity care. Notably, motivational factors did not correlate with the positive outcomes observed in our prior RCT, suggesting that self-discovery has a greater impact on clinical results.

Trial Registration: ClinicalTrials.gov NCT04714762; https://www.clinicaltrials.gov/study/NCT04714762

(JMIR Mhealth Uhealth 2025;13:e60855) doi:10.2196/60855

KEYWORDS

self-discovery; self-management; motivation; self-determination theory; self-tracking; gestational diabetes; mobile apps; mixed methods study; diabetes; support; women; pregnancy; maternal; apps; glucose; nutrition; physical activity; psychological; well-being; interview; questionnaire; usability; depression; stress

Introduction

Gestational diabetes mellitus (GDM) is an increasing issue in maternal health care as it affects approximately 14% of pregnant individuals globally [1]. GDM increases the risk of developing type 2 diabetes and cardiovascular diseases [2,3], and also predisposes the child to adulthood obesity and type 2 diabetes [4].

In order to respond to the increasing demand for treatment, digital health is a topic of growing interest, as it provides innovative approaches to support the clinical management of various conditions. Mobile apps and wearable devices provide an opportunity for continuous tracking of numerous parameters for self-management of noncommunicable diseases. Diabetes apps have been shown to improve patient self-management in type 1 and 2 diabetes [5-8], while also being cost-effective [9]. Recent studies have provided promising findings on telemedicine in the management of GDM [10-16], but the effective interventions have involved significant effort from health care professionals regularly, which is not an optimal use of digital tools.

An emerging area of interest within digital health is the concept of self-discovery through self-tracking data. In diabetes management, this involves individuals using data on daily activities—such as nutrition, physical activity, sleep, and stress—to understand how these behaviors affect glucose levels. This concept, which has gained attention in the field of human-computer interaction [17-19], provides a valuable framework for understanding how self-tracking can enhance the self-management of diabetes.

To enhance the effectiveness of mobile apps designed for GDM management without the involvement of health care professionals, we developed an easy-to-use eMOM app that differs from the prior diabetes mobile apps in several aspects. First, it visualizes data from a continuous glucose monitor (CGM), physical activity sensor, and digital food diary in real time within the same app [20]. Second, the idea of eMOM is to teach a patient how their own lifestyle choices affect their glucose levels without extra help from health care personnel. Third, eMOM is used periodically, 1 week a month, to avoid getting tired of digital devices. The recent results of our randomized controlled trial (RCT) demonstrated that individuals with GDM who used eMOM improved their fasting glucose levels, enhanced physical activity, increased vegetable intake, and resulted in less weight gain during pregnancy [21]. Additionally, the incidence of newborns with macrosomia was lower among the individuals using eMOM [21]. Many of these improved clinical outcomes were associated with the use of eMOM, suggesting that eMOM supported self-management of persons with GDM.

The aim of this study is to evaluate how the use of eMOM affects the underlying maternal learning processes, that is, self-discovery, in the context of GDM self-management. We will also examine whether eMOM affects autonomous motivation for GDM self-management or psychological well-being. Additionally, we will assess whether there is an association between motivational factors and improved clinical findings in the RCT study. Finally, we will evaluate the acceptance and usability of the eMOM app.

Methods

Design

This study is a secondary analysis of eMOM GDM RCT [20], where the intervention arm used the eMOM app with a CGM and activity bracelet for 1 week per month and a digital food diary for 3 days of this week (referred to as the ‘app week’ later in this paper), from gestational weeks 24-28 to delivery. The eMOM GDM study was conducted between February 2021 and December 2022 in Helsinki, Finland. Patients were recruited from antenatal clinics in the Helsinki metropolitan area by research nurses.

Participants eligible for the study were persons aged 18-45 years with a diagnosis of gestational diabetes at gestational weeks 24-28. Individuals were excluded from the study if they had preexisting type 1 or type 2 diabetes, were using medications that affect glucose metabolism (such as continuous corticosteroid therapy or metformin), had multiple pregnancies, physical disabilities, current substance abuse, severe psychiatric disorders that could complicate participation, or significant difficulties in cooperating, such as inadequate Finnish language skills.

Participants were randomized 1:1 into intervention and control arms. The intervention arm received standard clinical antenatal care in addition to the eMOM app, while the control arm received only standard clinical care. Standard clinical care consisted of periodic visits with a nurse and a doctor in municipal maternal clinics, as offered for all pregnant individuals in Finland. A regular care protocol for low-risk pregnancies includes 9-10 visits with a nurse and 2 visits with a doctor. After being diagnosed with GDM, individuals receive guidance on diet, physical activity, and self-monitoring of blood glucose with electronic capillary glucose meters [22].

XSL•FO
RenderX

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and received approval from the ethics committee of Helsinki University Hospital (approval no. HUS/2165/2018). All participants enrolled voluntarily, provided written informed consent, and were informed of their right to withdraw from the study at any time without consequence. No financial compensation was offered for participation. All data were pseudonymized and nonidentifiable to protect participant confidentiality.

eMOM App

The eMOM mobile app was developed to visually represent data from a tissue CGM (Medtronic Guardian Connect with Enlite Sensor) and an activity bracelet (Garmin Vivosmart 3) together with a digital food diary. The information from the sensors and food diary is updated every 10 minutes to eMOM. In our previous study [23], physical activity was measured with 3 different devices (activity bracelet, hip-worn sensor, and electrocardiography sensor). Most participants preferred the bracelet to hip sensor, so it was selected as an activity measurement tool for this study. Screenshots from the eMOM app are presented in Figure 1. For displaying the data, the eMOM app has 2 main views: a day view (Figures 1A and 1D) and a week view (Figure 1E). Nutrition intake from each energy-yielding nutrient is visualized as stacked bars (Figure 1A). More detailed information (recorded food items and nutrient intake in grams) can be accessed (Figure 1B) by tapping the stacked bar (Figure 1A). The eMOM app shows detailed glucose values (Figure 1C) when tapping the glucose curve on the screen (Figures 1A and 1D).

Figure 1. (A) Day view with nutrition and glucose filters selected, (B) detailed nutrition view, (C) detailed glucose view, (D) day view with physical activity (steps), and (E) week view with glucose and physical activity (minutes) selected. The following outcomes were visible to the user through the eMOM: (1) CGM glucose time in range (3.9-7.8 mmol/l), (2) postprandial increase of continuous glucose in 2 hours (mmol/l), (3) energy intake (kJ/day), (4) intake of carbohydrate (g/day), (5) intake of total fat (g/day), (6) intake of saturated fat (g/day), (7) intake of fiber (g/day), (8) intake of protein (g/day), (9) intake of sucrose (g/day), (10) number of walking steps (steps/day), and (11) duration of physical activity (minutes/day). CGM: continuous glucose monitor. Adapted from Kytö et al [21]. Screenshots (A-E) from the eMOM (Copyright: Fujitsu Finland).

Interactions With eMOM

The interactions with eMOM were collected in log files to determine how the participants used eMOM during the study. Each row in the log file included a pseudonymized user ID, timestamp, used view (day view, week view, detailed nutrition view, detailed glucose view, or info view), and filter used in day view (glucose, nutrition, and physical activity) or in week view (glucose, nutrition, physical activity, weight, and sleep).

Questionnaires and Interviews

The background questionnaire covered experience with self-tracking at baseline (gestational weeks 24-28). Other background information, such as age, BMI, parity, and previous GDM status, were collected from hospital registries.

Maternal learning and self-discovery were assessed through interviews conducted with randomly selected participants from the intervention arm. These interviews took place after participants had been using the app for approximately 2 months. Interviews were semistructured and carried out remotely over videoconferencing software. They were audio or video recorded

XSL•FO
RenderX

depending on the participant's preference. They lasted from 30 to 56 minutes; the mean duration was 40 (SD 8.5) minutes. The main interview questions were, "Have you learned something while using the eMOM app? If yes, what?" "Has the eMOM helped you to make changes in your lifestyle? If yes, how?" "What features of the eMOM were most useful for you?" "How would you develop the app?" The interview script is provided in Multimedia Appendix 1.

To assess maternal motivation, we used 2 questionnaires based on the self-determination theory: the Treatment Self-Regulation Questionnaire (TSRQ) and the Perceived Competence Scale (PCS) for diabetes [24]. The TSRQ questionnaire assesses levels of autonomous versus controlled motivation, while the PCS measures perceived competence. Both questionnaires were administered at baseline (gestational weeks 24-28) and again between gestational weeks 35-37. Details of the questionnaires and scoring methods are available in Multimedia Appendix 2.

The maternal depression was measured with the Edinburgh Postnatal Depression Scale [25] at baseline (gestational weeks 24-28) and again at gestational weeks 35-37. The scale consists of 10 items, each scored from 0 to 3, yielding a total score range of 0 to 30. A score of 10 or higher is commonly used as the threshold to indicate possible depression [25].

The acceptance of eMOM was measured with the Unified Theory of Acceptance and Use of Technology questionnaire [26] after each app week (once a month), and the usability of the eMOM app was measured once, 4 weeks after taking the app into use with a questionnaire adapted from the Software Usability Measurement Inventory questionnaire [27].

Stress Measurement

The Firstbeat Bodyguard 2, a chest-mounted sensor by Firstbeat Technologies, tracks heart rate variability (HRV) and physical activity using 3D acceleration over a continuous 3-day period. It uses 2 clinical-grade electrocardiographic electrodes for attachment and boasts a high accuracy in HRV measurement, with less than a 3-millisecond error and over 99.9% detection accuracy compared to standard electrocardiographic devices [28]. Additionally, Firstbeat's Lifestyle Assessment software analyzes the HRV and motion data to evaluate stress and recovery levels, a methodology supported by extensive research validation [29,30]. Both the intervention and control arms used the Firstbeat sensor during the baseline period (gestational weeks 24-28) and again in the third trimester (gestational weeks 35-37). The sensor was worn continuously for 3 days during each phase.

Analysis

The interviews underwent transcription, and 2 researchers (HY and SM) familiarized themselves with the content by reviewing the transcripts. The subsequent analysis followed the framework method, a recommended approach for multidisciplinary health research [31]. We used a combination of deductive and inductive approaches, where we had a set of initial codes (self-tracking of glucose, self-tracking of nutrition, and self-tracking of physical activity) at the beginning of the coding and used new codes when emerged from the gathered data. These new codes were combined into broader categories. The coding was performed by HY using ATLAS.ti software, which enabled charting of resulting themes and participants' responses. This matrix presentation enabled us to analyze patterns, similarities, and differences in responses within each theme. Coding and the resulting themes were reviewed by 2 additional researchers (SM and MK). Quotes featured in the results were translated into English using an intelligent verbatim technique, which involves the removal of filler words like "er" during the translation process.

The log files were examined by computing the frequency and timing of interactions with the eMOM platform. Finally, we used triangulation across these various data sources (log files from eMOM, interviews, and questionnaires) to gain a comprehensive understanding of how self-tracking with eMOM supported self-discovery.

For comparing the difference between the arms, we performed a 2-sided *t* test for continuous outcomes, setting the significance threshold at 5%. For binary outcomes, we applied a chi-square test, resorting to the Fisher exact test when expected frequencies fell below 10. The difference between app weeks in the acceptance of eMOM was evaluated using repeated-measures ANOVA.

In analyzing the correlation between motivation and primary outcome from Kytö et al [21], we examined the linear relationship (Pearson correlation). For the binary outcomes, we used logistic regression.

Results

Participants

In total, 148 participants with GDM participated in the study. Out of 76 persons in the intervention arm who used the eMOM app, 63 continued until delivery. The main reason for dropout was the need to start GDM medication. In total, 18 participants from the intervention arm were randomly selected for user experience interviews for this study. The log files were collected from all the participants in the intervention arm. Baseline characteristics are presented in Table 1. Participants were familiar with using mobile apps and activity trackers and had a good conception of the significance of different nutrients for the diet, as shown in Table 1. No significant differences were observed between the arms at the baseline.

XSL•FO
RenderX

Table 1. Background characteristics.

	Interviews (n=18)	Intervention arm (n=76)	Control arm (n=72)	Difference between intervention and control arm (95% CI)
Age (years), mean (SD)	32.8 (3.7)	34.2 (4.1)	34 (3.9)	0.2 (–1.1 to 1.5)
BMI, mean (SD)	27.5 (5.1)	27.5 (5.5)	26.7 (4.5)	0.8 (–0.8 to 2.4)
Parity—primiparous, n (%)	10 (56)	38 (50)	37 (51)	–1% (–18% to 15%)
Early GDM[a], n (%)	7 (38)	37 (49)	35 (49)	0% (–16% to 16%)
GDM status (previous GDM), n (%)	2 (11)	17 (22)	18 (25)[b]	–3% (–17% to 11%)
Gestational week at enrollment, mean (SD)	27 (1.8)	26.5 (1.7)	26.3 (1.7)	0.2 (0.4 to –0.7)
Weeks in the study at the interview, mean (SD)	9.3 (2.6)	—[c]	—	—
"I am used to use mobile apps," n (%)[d]				
1	0 (0)	1 (2)	4 (6)	—
2	0 (0)	0 (0)	1 (2)	—
3	0 (0)	0 (0)	0 (0)	—
4	2 (15)	8 (13)	5 (8)	—
5	11 (85)	55 (86)	54 (84)	—
"I am used to use activity trackers (like Fitbit and Polar)," n (%)[d]				
1	1 (8)	11 (17)	11 (18)	—
2	0 (0)	3 (5)	4 (6)	—
3	1 (8)	4 (6)	1 (2)	—
4	3 (23)	18 (28)	20 (32)	—
5	6 (46)	28 (44)	27 (43)	—
"I am familiar with measuring my blood glucose," n (%)[d]				
1	2 (15)	10 (16)	3 (5)	—
2	3 (23)	3 (5)	6 (10)	—
3	2 (15)	4 (6)	1 (2)	—
4	3 (31)	21 (33)	19 (30)	—
5	3 (38)	25 (41)	34 (54)	—
"I am familiar with keeping a food diary," n (%)[d]				
1	1 (8)	7 (11)	5 (8)	—
2	0 (0)	8 (13)	5 (8)	—
3	0 (0)	4 (6)	2 (3)	—
4	8 (62)	48 (48)	34 (54)	—
5	4 (31)	14 (22)	17 (27)	—
"I know the significance of different nutrients, fiber, and fat for the diet," n (%)[d]				
1	0 (0)	2 (3)	1 (2)	—
2	0 (0)	0 (0)	3 (5)	—
3	0 (0)	2 (3)	2 (3)	—
4	6 (46)	28 (44)	25 (40)	—
5	7 (54)	32 (50)	32 (51)	—
Autonomous motivation (TSRQ[e]), mean (SD)[f]	4.86 (0.51)	4.85 (0.51)[b]	4.68 (0.52)[g]	0.17 (–0.01 to 0.34)
Perceived competence (PCS[h]), mean (SD)[f]	5.71 (1.18)	5.60 (0.97)[b]	5.49 (1.18)[g]	0.11 (–0.25 to 0.47)

XSL•FO
RenderX

	Interviews (n=18)	Intervention arm (n=76)	Control arm (n=72)	Difference between intervention and control arm (95% CI)
Elevated depressive symptoms (EPDS[i] points>10), n (%)	3 (17)	12/71 (17)[j]	10 (14)	3% (–8.8% to 14.8%)
Stress percentage	53.5 (13.2)[j]	56.1 (13)[k]	57.5 (11.7)[l]	–1.4 (–6 to 3.2)

[a]GDM: gestational diabetes mellitus.

[b]Result missing from 1 participant.

[c]Not applicable.

[d]The Likert scale was from 1=strongly disagree to 5=strongly agree.

[e]TSRQ: Treatment Self-Regulation Questionnaire.

[f]The scale was from 1=low to 7=high.

[g]Result missing from 4 participants.

[h]PCS: Perceived Competence Scale.

[i]EPDS: Edinburgh Postnatal Depression Scale.

[j]Result missing from 1 participant.

[k]Result missing from 2 participants.

[l]Result missing from 3 participants.

eMOM Use

Users interacted with the eMOM app most frequently during the morning hours (7-10 AM) and the evening hours (7-10 PM). Interaction levels were lowest at night (12-5 AM), increased sharply around 6 AM, peaked at 9 AM, and remained relatively steady throughout the day with a secondary rise in the evening.

During the first week of use, participants interacted with the app an average of 18.7 (95% CI 17-20.4) times per day. In the following application weeks, daily interactions ranged between 10 and 12 times. Between active usage weeks, the interaction frequency dropped significantly to 1-3 times per day.

Regarding app usage patterns, the day view was by far the most commonly used feature, accounting for 65% of total interactions (Table 2). This was followed by the week view (18%), detailed glucose view (10%), information section (4%), and detailed nutrition view (3%). When viewing data, users most frequently applied the glucose filter (65% of the time), followed by the nutrition and physical activity filters (each 15%), while the weight and sleep filters were used only 3% of the time (Table 3). Glucose was mostly viewed together with nutrition and physical activity, as shown in Figure 2.

Table 2. Usage of different views.

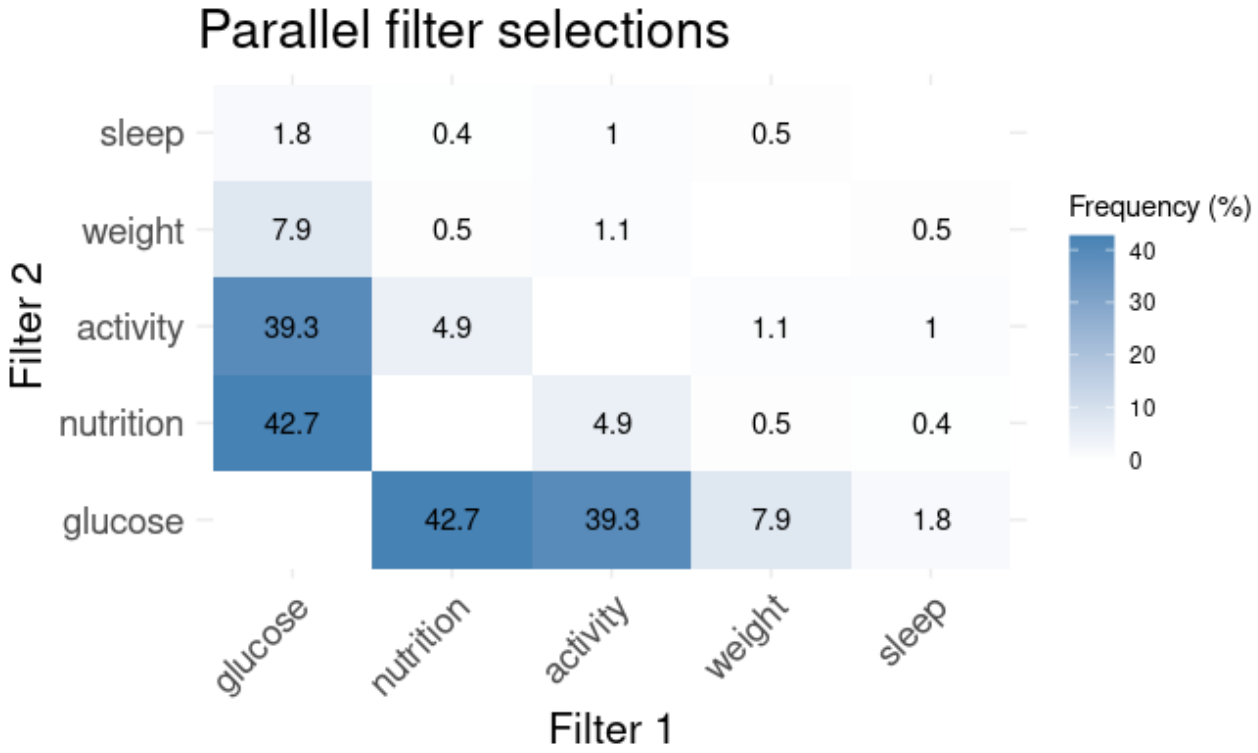
View	Frequency (%)
Day view	65
Week view	18
Detailed glucose view	10
Information section	4
Detailed nutrition view	3

Table 3. Usage of different filters.

Filter	Frequency (%)
Glucose	65
Nutrition	15
Physical activity	15
Weight	3
Sleep	3

XSL•FO
RenderX

Figure 2. Parallel filter selections.



Self-Discovery

The majority of interviewed participants (17/18, 94%) reported that the eMOM app facilitated their understanding of the connections between their daily activities and glucose levels. Primarily, participants found their learning experiences centered around nutrition (15/18, 83%) or physical activity (8/18, 44%). Seven (39%) participants acknowledged the combined influence of both nutrition and activity on their glucose levels. The interviews highlighted various methods of learning, such as experimentation, comparison, and observation.

Regarding self-discovery, 335 codes and 22 categories emerged from the interview data. These categories were then combined into main themes, which we present in a condensed form in the following sections. We have gathered all the supportive features of eMOM discussed in the interviews, along with direct comments from participants, and organized them into Table 4.

XSL•FO
RenderX

Table 4. Supporting features of eMOM identified through interviews.

Feature	How eMOM supports	Comments from the interviews
Nutrition		
Experimenting with different foods	The detailed nutrition view displays the specific foods in each meal and provides the breakdown of various nutrients, including the amount of carbohydrates, proteins, and fats, in grams.	• "I switched from refined grains to whole grains, and it [glucose] has stayed pretty stable with that change." [Participant 15] • "If you consume, for example, a lot of carbohydrates in a meal, the effects show up immediately." [Participant 18] • "—It raised my blood sugar too high, or very close to the limit. So, I started to reduce my carbohydrate intake, which helped me find a balance. I realized that I should eat a small, protein-rich breakfast to kickstart my day. Then, after a couple of hours, I can have another meal with carbohydrates, and that keeps my blood sugar stable." [Participant 8] • "I learned that I can't eat fruits in the morning because they're too high in sugar after a night's fasting." [Participant 3]
Meal rhythm	The nutrition filter displays meals as bar graphs on a daily timeline, allowing users to view them alongside the glucose curve.	• "I hadn't realized that I was going way too long between meals. Like, you should eat more often. And by looking from the app I have noticed that if I don't eat regularly enough, it [glucose] starts to rise." [Participant 10] • "Having a proper meal later in the evening has the biggest effect. I mean, if I haven't eaten enough, my glucose levels spike." [Participant 11] • "I experimented with a variety of foods and discovered that eating a chocolate bar alone in the middle of the day spikes my blood sugar, while having it as a dessert after lunch does not necessarily have the same effect." [Participant 12]
Portion sizes	The nutrition filter displays the calorie count for each meal. Portion sizes are visually represented with a bar graph, making it easier to see the proportion of meals throughout the day.	• "Some restaurant meals etc. are so big that the blood sugar rises very high for a very long time. It was a very useful fact for me, as I immediately quit eating those." [Participant 12] • "I gained quite a good understanding of the number of calories in my meals and their variation. Since it is presented as bar graphs, it helps to get some insight into what changes I should make to my meal habits in the future." [Participant 17]
Activity		
General activity level	Activity filter shows the number of steps and active minutes and can be compared to the glucose curve.	• "On February 19th, I was quite active, which kept my glucose levels within their normal range. However, on February 20th, I wasn't very active, and my glucose levels immediately rose above the threshold values. I've been comparing them in this way." [Participant 2] • "I have noticed that when I'm more active, the glucose level decreases faster." [Participant 16] • "By selecting both the activity and glucose filters, you can see the number of steps, and it becomes clear that, for example, taking a longer walk has an immediate effect on glucose levels." [Participant 1]
Physical activity after a meal	The activity filter displays the number of steps per hour using bar graphs. These data can be viewed alongside the glucose curve, allowing users to observe the effect of physical activity on glucose levels on an hourly basis.	• "The other day I took a longer walk on a forest track after a meal and noticed that the glucose level didn´t rise as high as it normally would after a meal." [Participant 3] • "I learned very concretely that if I take kids to football practice right after a meal, glucose starts to decrease even after a bigger meal. But if I sit at a brunch table with my friends, it [glucose] rises very high very easily. This was very useful information, I think." [Participant 12] • "I noticed that if I, for example, take a short walk after eating, the glucose curve starts decreasing more." [Participant 8]
Other supporting features of data visualization	eMOM integrates nutrition and activity data in one place, offering a clearer view of glucose variation, and displaying target glucose values, which helps users track and manage their condition more effectively.	• "The eMOM app gathered both nutrition and food data in the same place, so if you wanted to do some kind of learning analysis about your nutrition or activity, the eMOM app was definitely better [versus only CGM/Medtronic] since it shows the combined data." [Participant 3] • "CGM/Medtronic does not display the glucose target, like upper and lower bounds. I think that, for example, the variation in glucose level is more clearly presented in the eMOM app." [Participant 6] • "Having target values on glucose helps to see how much you exceed the target." [Participant 8]

Learning Between Nutrition and Glucose

The majority of the participants (15/18, 83%) discovered the patterns of association between nutrition and glucose. Nutrition information was presented through stacked bar charts, with a possibility to be integrated with the glucose curve, enabling direct comparisons of how various meals impacted glucose levels. Participants engaged in experimentation and comparison,

XSL•FO
RenderX

evaluating different foods and nutrients, mealtimes, and portion sizes.

As shown in Figure 2, glucose was mainly viewed alongside nutrition (43% of the time). Table 4 illustrates user comments on how nutrition data has contributed to their self-discovery.

Learning Between Physical Activity and Glucose

Although the activity was viewed with similar frequency as nutrition (see Table 3) and nearly as frequently in parallel with glucose visualization (39% of the time; Figure 2), it contributed less to self-discovery. However, almost half of the participants (8/18, 44%) noted a decrease in their glucose levels following exercise. Participants also assessed their activity levels on different days to explore the relationship between their activity—both active and inactive days—and their glucose levels on those days. Table 4 summarizes user comments on how activity data contributed to their self-discovery process.

Learning Between Sleep and Glucose

Sleep data were found to be challenging to interpret, with a mere 3% viewing frequency (Table 3). Four participants reported discrepancies between the sleep data provided by the device and their own assessments of sleep duration and quality. One participant observed an association between a poorly slept night and elevated morning glucose level.

Supporting Features of Learning in the eMOM App

The visualization of measurements emerged as a pivotal aspect of the learning process. Through observing the glucose curve, participants noticed fluctuations in glucose levels. When various filters (nutrition, activity, and sleep) were applied to the same graph alongside the glucose curve, participants could observe how different factors influenced glucose levels. The ability to view nutrition information in a quantitative format was also valued, and having a specific glucose target range was found to be important. Participants found the eMOM app more informative than Medtronic's CGM in evaluating the overall picture of glucose balance. Table 4 summarizes the primary features used by individuals with GDM in the self-discovery process and how the eMOM app supported the identification of associations between these features and glucose levels.

Temporal Progression of Learning

In the evenings, eMOM served as a tool for obtaining a comprehensive overview of the day's events, as 1 participant described the following.

> *And then, at the end of the day, I looked at the eMOM application to see how the day had gone overall. I used it to get a summary.* [Participant 9]

Conversely, during the day, participants tended to rely on Medtronic for checking glucose levels, as it offered easier and quicker measurements due to its lower latency, updating every 5 minutes compared to eMOM's 10-minute intervals.

For most participants (n=10, 56%), the first week of using the app was found to be the most significant in terms of learning. Two participants expressed.

> *The most useful was the first application week; that's when I learned the most.* [Participant 14]

> *The most benefit came during those initial weeks, when I could really see how my diet affected me. Based on that, I was able to adjust it to suit myself better.* [Participant 11]

This reflects their strong interest in monitoring glucose levels during this period. As a result, interactions with eMOM were most frequent during this initial learning phase, as illustrated in Figure 2.

Periodic use of the eMOM app was found to be effective in maintaining motivation and sustaining interest in tracking activities. As 1 participant noted.

> *If it was available all the time, it probably wouldn't be used as much. Like, at the beginning, it is probably the most interesting, but when you learn to know how different things affect, then you no longer are interested in observing it.* [Participant 15]

Inhibiting Features and Improvement Suggestions

Some participants (5/18, 28%) expressed a desire for enhanced informativeness by having all three filters (glucose, nutrition, and activity) integrated into the same graph.

A desire for additional support in interpreting data was expressed by 11 (61%) participants, particularly in understanding the impact of sleep data on glucose levels. Moreover, personalized recommendations regarding nutrition and more specific daily goals for macronutrients were requested. Preferences regarding support varied. While 1 participant mentioned seeking guidance from a nurse to interpret glucose data and 2 participants wished for assistance from health care personnel, others (8/18, 44%) would have been satisfied with suggestions or notifications from the app.

In this intervention, the activity bracelet tracked only steps and overall daily activity. Participants desired a broader range of sports tracking capabilities. For example, swimming is a preferred activity among pregnant persons, but Vivosmart 3 was not able to track swimming. Issues with activity tracking, such as inaccurate step counting during activities like walking with a stroller or counting knitting as steps, led to a sense of distrust in the accuracy of activity data.

One participant expressed a desire for guidance on physical activity, including recommendations on duration and types of exercises to meet daily activity goals. Some also mentioned that motivation toward physical activity decreased as pregnancy progressed and exercise levels decreased.

Furthermore, some participants noted that when their glucose levels remained within target ranges, their motivation to use the app and adhere to a healthy diet diminished. As 1 participant articulated.

> *My glucose values have stayed quite good, but if I would have had more variation in glucose values it might have been easier to see associations between things. But since my glucose values have been on target, I haven't really made any changes to my diet*

XSL•FO
RenderX

or other things I otherwise might have done. [Participant 14]

The Effect of eMOM on Motivation, Depression, and Stress

Autonomous motivation (measured with TSRQ) decreased in both arms: –0.11 (SD 0.38) in the control arm and –0.18 (SD 0.32) in the intervention arm between the baseline and gestational weeks 35-37. The between-arm difference was not statistically significant for the change of TSRQ (difference: –0.09; t_{97}=1.15; *P*=.25) nor for the change of perceived competence (difference: –0.27; t_{97}=1.21; *P*=.23). There was no statistically significant difference between the arms in the incidence of Edinburgh Postnatal Depression Scale scores greater than 10 (χ^2_1=0.10; *P*=.75). The summary of the results on the motivation is presented in Table 5.

Table 5. The effect of eMOM on autonomous motivation, competence, EPDS[a], and stress percentages.

	Control arm		Intervention arm		Statistics		
	n	Mean (SD)	n	Mean (SD)	Difference	*t* test (2-tailed)	*P* value
Change in autonomous motivation (TSRQ[b])	48	–0.10 (0.41)	51	–0.19 (0.36)	–0.09	1.15 (97)	.25
Change in perceived competence (PCS[c])	48	0.16 (1.03)	51	–0.11 (1.17)	–0.27	1.21 (97)	.23
Incidence of EPDS>10	52	5 (10)[d]	52	6 (12)[d]	1 (2)[d]	0.10 (1)[e]	.75
Change in stress percentage	51	1.4 (12.5)	58	4.1 (11.9)	2.7	1.18 (107)	.24

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]TSRQ: Treatment Self-Regulation Questionnaire.

[c]PCS: Perceived Competence Scale.

[d]n (%) values.

[e]Chi-square test.

Correlations Between Motivation and Improved Clinical Outcomes

Correlations between autonomous motivation (TSRQ) and perceived competence (PCS) and improved primary and secondary clinical outcomes [21] in the intervention arm are shown in Table 6. No significant correlations were found in change of motivational factors and improved clinical outcomes.

Table 6. Correlations between autonomous motivation (TSRQ[a]) and perceived competence (PCS[b]) on clinical outcomes.

Outcome where the effect of the intervention was observed in Kytö [21]	Change in autonomous motivation			Change in perceived competence		
	n	Pearson coefficient (*r*)	*P* value	n	Pearson coefficient (*r*)	*P* value
Primary outcome: Change in fasting glucose from the baseline to 35-37 gestational weeks (mmol/L)	49	–0.02	.88	49	–0.05	.75
Mean morning capillary fasting glucose (mmol/L)	35	–0.11	.52	35	0.26	.12
Timing of medication start[c]	10	–0.20	.54	10	0.25	.43
Gestational weight gain from the baseline to delivery (kg)	49	0.07	.61	49	0.06	.67
Change in intake of vegetables (g/MJ)	36	–0.04	.79	36	0.24	.15
Change in duration of sedentary behavior (minute/day)	46	–0.09	.54	46	–0.14	.35
Change in duration of light-physical-activity (minute/day)	46	0.05	.74	46	0.06	.67
Macrosomia (>4 kg)[d]	49	1.26[e]	.44	49	–0.35[e]	.48

[a]TSRQ: Treatment Self-Regulation Questionnaire.

[b]PCS: Perceived Competence Scale.

[c]Only the baseline measurement of motivation available.

[d]Logistic regression.

[e]β coefficient.

The Acceptance and Usability of eMOM

The acceptance of eMOM and wearable sensors was high over the whole intervention (see Table 7 for a subset of statements and Multimedia Appendix 3 for responses to all the statements). For example, participants found the eMOM and the sensors useful for supporting lifestyle changes and they perceived eMOM easy to use.

XSL•FO
RenderX

Table 7. Responses to a subset of statements in the UTAUT[a] questionnaire[b].

	App week 1 (n=67)	App week 2 (n=47)	App week 3 (n=26)	App week 4 (n=10)
	Somewhat agreed, agreed, or totally agreed (Likert scale 5-7), n (%)	Somewhat agreed, agreed, or totally agreed (Likert scale 5-7), n (%)	Somewhat agreed, agreed, or totally agreed (Likert scale 5-7), n (%)	Somewhat agreed, agreed, or totally agreed (Likert scale 5-7), n (%)
Perceived usefulness				
I would find the eMOM app with the sensors useful for supporting healthier lifestyle.	65 (94)	45 (94)	26 (100)	10 (100)
Using the eMOM app with the sensors enables me to improve my lifestyle more quickly.	64 (93)	45 (94)	25 (96)	10 (100)
If I use the sensor, I will increase my chances of getting a concrete lifestyle improvement.	65 (94)	44 (92)	24 (92)	10 (100)
Perceived ease of use				
My interaction with the sensor is clear and understandable.	62 (90)	44 (92)	26 (100)	10 (100)
I find the eMOM app with the sensors easy to use.	61 (88)	46 (96)	24 (92)	10 (100)
Attitude toward behavior and use				
Using the eMOM app with the sensors is a good idea.	68 (99)	46 (96)	24 (92)	10 (100)
The eMOM app and the sensors make changing lifestyle more interesting.	67 (97)	46 (96)	25 (96)	10 (100)
Intrinsic motivation				
Using the eMOM app and the sensors is fun.	53 (77)	31 (65)	20 (77)	9 (90)
I like using the eMOM app and the sensors.	59 (86)	40 (83)	23 (88)	9 (90)
Extrinsic motivation				
People who influence my behavior think that I should use the eMOM app and the sensors.	11 (16)	7 (15)	6 (23)	3 (30)
Facilitation				
I have the resources necessary to use the eMOM app and the sensors.	65 (96)	44 (94)	26 (100)	10 (100)
I have the knowledge necessary to use the eMOM app and the sensors.	65 (96)	46 (98)	26 (100)	10 (100)
I could improve my lifestyle using the eMOM app and the sensors If there was no one around to help me with eMOM and sensors.	54 (81)	39 (83)	22 (85)	10 (100)
Anxiety				
I feel apprehensive about using the system.	3 (4)	2 (4)	2 (8)	1 (10)
The eMOM app and the sensors are somewhat intimidating to me.	1 (1)	0 (0)	0 (0)	0 (0)

[a]UTAUT: Unified Theory of Acceptance and Use of Technology.

[b]Participants responded to each statement in UTAUT questionnaire using a 7-point Likert scale ranging from 1=totally disagree to 7=totally agree.

Responses to the usability questionnaire (modified from the Software Usability Measurement Inventory) were in line with responses to the Unified Theory of Acceptance and Use of Technology. For example, 86% (38/44) of participants disagreed with the statement “Learning the functions of the eMOM GDM application takes too long,” and 66% (29/44) agreed with the

XSL•FO
RenderX

statement "I understand and know how to act based on the information provided by the application." 59% (26/44) of participants agreed with the statement "eMOM answers too slowly to inputs," which may have influenced the preference for Medtronic, which updates more quickly, during the day and eMOM more frequently in the evening. See Figure 3 for all the results of the usability questionnaire.

Figure 3. Responses to modified SUMI questionnaire (N=44). SUMI: Software Usability Measurement Inventory.

Discussion

Principal Findings

We found that a mobile app combining CGM, nutrition, and physical activity within the same app was highly used and supported self-discovery among persons with gestational diabetes. Providing pregnant persons with the opportunity to compare the effects of nutrition and physical activity on glucose levels facilitated their understanding of cause and effect. Almost all participants using the mobile app discovered associations between either nutrition or physical activity, and glucose levels. The majority found associations between nutrition and glucose, and nearly half of the participants identified associations between physical activity and glucose. There were no differences between the intervention and control arms in the change of perceived competence to manage GDM, autonomous motivation, depression, or stress scores. Additionally, autonomous motivation did not influence the improved clinical outcomes observed in our previous RCT [21]. Acceptance and usability remained high throughout the entire study period [21].

Self-Discovery

The impact of an app that comprehensively integrates data in a single platform (including the automatic collection of activity and continuous glucose data) on self-discovery has not been previously studied in the context of gestational diabetes. Our study demonstrates that participants gained valuable insights into the associations between nutrition and glucose levels, facilitated by the eMOM app's integrated data display. This comprehensive platform, which automatically collects activity and continuous glucose data, enabled users to observe the real-time impacts of their dietary choices on glucose levels. The ability to view glucose curves alongside nutritional details in a

XSL•FO
RenderX

single window simplified the interpretation of meal timing and its effects, helping users better regulate their meal rhythms. Participants found value in comparing different foods and portion sizes, highlighting the importance of detailed nutritional information over simple carbohydrate tracking. These findings align with our previous work with various sensors, but without the eMOM app [23], where participants noted causal relationships between nutrition and glucose levels.

Learning the associations between physical activity and glucose proved to be somewhat more challenging than understanding the associations between nutrition and glucose. However, in this study, the associations between physical activity and glucose were more prominent compared to our study without the eMOM app [23]. The app's parallel visualization of activity and glucose data may explain this improvement, as users could more easily observe associations between their physical activity and glucose changes. However, selecting features related to physical activity is challenging during pregnancy, as it complicates activity tracking. Activity bracelets typically track activity based on heart rate, which physiologically increases during pregnancy [32]. Currently, there are no activity bracelets designed specifically for pregnancy, which would account for these physical changes. While walking is a common form of physical activity during pregnancy [33,34], it may become difficult in the later stages. Swimming, water running, and weightlifting are preferred alternatives among pregnant individuals [33]. Therefore, it is essential to be able to track these alternative forms of physical activity. It is also worth noting that as pregnancy progresses, there is a general decline in physical activity levels [35,36]. Consequently, the relevance of physical activity in the management of GDM diminishes as pregnancy approaches its later stages. Similar challenges in physical activity tracking emerged in a study by Årsand et al [37] on type 2 diabetes, where patients expressed a desire for a broader range of trackable sports beyond just counting steps.

The eMOM app also provided data on sleep patterns, although interpreting this information proved challenging. Only 1 participant reported observing a potential connection between sleep and glucose levels, noting that during the flu accompanied by poor sleep, her glucose levels were higher. However, this association remains uncertain, as the flu itself can affect glucose levels. These results highlight the difficulty in finding associations between sleep patterns and glucose levels. Additionally, there is limited evidence on the effectiveness of wearable sensors on sleep during pregnancy [38]. Further investigation is needed in this area.

Recent research has explored ways to enhance the use of self-tracking data in diabetes self-management, although focusing primarily on type 1 and 2 diabetes [18,39-43]. However, self-management support is also crucial for individuals diagnosed with GDM, especially immediately following diagnosis, when adapting to the condition can be challenging [44,45]. Many individuals with GDM rely on trial-and-error learning [44-48], as seen in studies like that of Ibrahim et al [44], where participants expressed a desire to have all their self-monitoring tools gathered in one place to aid in sense-making. The recent review by Safiee et al [49] further supports this, showing that telemedicine interventions in GDM management improved self-management skills, particularly around understanding nutrition's impact on glucose. However, all prior reviewed interventions lacked CGM. Capillary glucose measurements capture only single-point readings, missing the full-day glucose fluctuations, including nocturnal hyperglycemia, which has been linked to a higher likelihood of large-for-gestational-age babies [50]. In contrast, the eMOM app provided a more comprehensive view of glucose levels, allowing participants to track changes over time. It is, however, worth noting that although CGMs have been found to be acceptable among persons with GDM [51-54], their use alone may not lead to improved glycemic control [54,55] or a reduction in macrosomia (birth weight > 4 kg) [53]. One contributing factor is the lack of clarity regarding the cause-and-effect relationships between lifestyle choices and glucose levels among persons diagnosed with GDM [45,47,56-58]. eMOM provides a tool for observing these causal relationships, thus enhancing the self-discovery process in managing GDM.

These findings emphasize the value of comprehensive data visualization tools in facilitating meaningful self-discovery in GDM management, offering both immediate feedback and clearer connections between daily behaviors and glucose outcomes.

Motivation

While previous studies indicate a strong motivation among individuals with GDM to manage their condition [45,48,59], our study aimed to understand how eMOM influences motivation, with a particular focus on autonomous motivation and perceived competence—both key components of self-determination theory. Contrary to the findings of Williams et al [60], who demonstrated a significant correlation between improved glucose levels and autonomous motivation among type 2 diabetes patients, our findings suggest that the eMOM app did not lead to improved clinical outcomes through increased autonomous motivation. At the start of the intervention, our study participants in both the intervention and control arms were already highly motivated, leaving little room for further increases in autonomous motivation. Additionally, those using eMOM did not increase their perceived competence to manage GDM compared with the control arm. These findings suggest that the improved maternal clinical outcomes reported earlier [21] are primarily attributed to self-discovery rather than increased motivation. Thus, motivational interventions that are effective in type 2 diabetes management may not directly apply to GDM. Instead, educational approaches, such as the eMOM app, play a crucial role in helping persons with GDM learn how to manage their condition effectively.

Strengths and Limitations

This study is the first to combine qualitative and quantitative approaches to investigate how comprehensive self-tracking influences self-discovery and motivation in managing GDM. The eMOM app offers a novel approach to GDM self-management by integrating CGM, nutrition tracking, and physical activity monitoring in real-time within a single app, without guidance from health care personnel, thus addressing a gap in previous GDM apps [61]. This integrated approach was

XSL•FO
RenderX

also requested in our previous study, where participants used the same variable sensors and meters (CGM, physical activity, food diary, and sleep) as in this study, but separately in different apps and devices [23].

While some previous GDM apps have aimed to reduce reliance on health care personnel [15,16,62,63], none have incorporated the level of comprehensive and automated data tracking offered by the eMOM app. To our knowledge, only 1 earlier app, Habits-GDM by Yew et al [15], combined glucose data, a food diary, and physical activity tracking. It used capillary glucose measurements and monitored physical activity solely through step counts using a phone pedometer. The app also displayed only total calories and carbs, lacking detailed nutritional data. While Habits-GDM resulted in lower proportions of glucose levels above target and reduction in composed neonatal outcomes, none of the individual neonatal outcomes were statistically significant, and the intervention did not reduce gestational weight gain. Habits-GDM also included a messaging feature, allowing users to consult health care personnel and receive responses within 24 hours. However, the study did not report information on the app's usage rates.

Similarly, MobiGuide by Peleg et al [62] combined capillary glucose measurements with step counts but lacked nutrition tracking. While it showed a significant reduction in blood pressure, it did not demonstrate any significant differences in glucose levels, insulin use, or neonatal outcomes. The Pregnant+ app by Borgen et al [63] allowed users to manually log physical activity, but none did, highlighting that manual logging alone was insufficient for meeting users' needs.

As reported in the RCT study [21], eMOM use was consistently high. Most of the learning occurred, however, during the first application week. In the following application weeks, eMOM was mainly used to check if the glucose stayed within the limits. As noted by the participants, continuous availability of the sensors might lead to decreased motivation for tracking. Due to this periodicity, engagement with eMOM remained high during the application weeks, in contrast to the declining engagement observed in automatic self-tracking among people with type 2 and type 1 diabetes [64]. Participants continued to interact with eMOM even outside the designated application weeks, even when they did not have access to the activity bracelet or CGM. However, these interactions were significantly less compared to the application weeks. Throughout this study, the acceptance of eMOM and wearable sensors remained high, as indicated by questionnaires.

Previous research has suggested that extensive self-monitoring could lead to adverse psychological outcomes [65,66]. We found no evidence of an effect on depression or stress, as measured by Edinburgh Postnatal Depression Scale questionnaires and Firstbeat stress scores. This suggests that comprehensive tracking does not have a detrimental effect on maternal psychological well-being.

The most significant limitation concerning the self-discovery process was the inability to display nutrition, physical activity, and glucose data together in 1 graph. Participants expressed a need for more support in interpreting data and making informed choices. For example, the app could send notifications when glucose values exceed the target range and instruct users to review their recent meals.

Self-discovery's impact on clinical outcomes was not objectively measured. However, the results of the interviews suggest that self-discovery had a positive effect on clinical outcomes. Another limitation is that individuals who required medication for GDM were excluded from the study. This may have introduced bias into the results of motivation, depression, and stress levels, as the analyses only included patients whose condition was managed well enough to avoid medication.

Insights, Recommendations, and Future Directions

Implementing digital tools in clinical care requires patients to have health literacy. Low health literacy in patients with type 2 diabetes has been linked to lower engagement with mobile-based health interventions [67]. To be effective, information provided by these apps should be tailored to the health literacy levels of the target population [68]. In future research, eMOM should be evaluated in populations with varying levels of health literacy. Artificial intelligence–based solutions could assist users in understanding lifestyle improvements based on their health data.

Our findings suggest that an app presenting features in a quantifiable format is well-suited for the self-discovery process in GDM. Additionally, the limited number of features in GDM and the clear causal relationships between glucose and these features make this process easier to learn. Self-discovery in GDM exemplifies goal-driven tracking [69], particularly for monitoring nutrition and physical activity to maintain glucose levels within the recommended limits set by health care providers. We believe that our approach of integrating data from various sources in real-time could be beneficial for other noncommunicable diseases, such as type 2 diabetes, where the relationships between features and outcomes are relatively clear and learnable. Especially after diagnosis, when individuals are still in the learning phase of managing their condition, patients could benefit from a solution like ours.

Beyond the short-term effects on pregnancy outcomes, gestational diabetes also has long-lasting implications for maternal health [2,3]. Future research should investigate the extended impact of using the eMOM app, particularly in the postpartum period. Examining how this tool influences long-term health behaviors, psychological well-being, and the prevention of type 2 diabetes could offer valuable insights into its role in promoting sustained self-management beyond pregnancy.

Conclusions

By visualizing CGM, nutrition, and activity in a single app in real-time, the eMOM app provides a promising tool to enhance self-discovery in GDM, without additional support from health care personnel. Our findings also suggest that eMOM's impact is more related to self-discovery than to increased motivation. To support learning further, future work involves visualization of more than 1 feature impacting glucose at the same time and the use of artificial intelligence–based recommendations. In addition to GDM, this kind of mobile health solution could be

XSL•FO
RenderX

used to support learning in other noncommunicable diseases, such as type 2 diabetes.

Acknowledgments

The authors wish to thank all the persons with gestational diabetes mellitus (GDM) in Finland for taking part in the research and the study nurses Jaana Palukka, Sanni Lampi, and Milla Tuhkanen. This work was funded by the Business Finland eMOM GDM project.

Authors' Contributions

SM analyzed the qualitative data and contributed to the original draft. SK conceptualized the eMOM app, secured funding, planned the study design, and cowrote the original draft with SM, MK, and HY. HY analyzed the qualitative data and cowrote the original draft alongside SM, SK, and MK. SH conceptualized the eMOM app, acquired funding, and planned the study design. MK contributed to the conceptualization of the eMOM app, planned the study design and data analysis, conducted participant interviews, analyzed the quantitative data, and coauthored the original draft with SM, SK, and HY. All authors reviewed and edited the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview questions.
[DOCX File , 20 KB - mhealth_v13i1e60855_app1.docx]

Multimedia Appendix 2
TSRQ and PCS questionnaires. TSRQ: Treatment Self-Regulation Questionnaire; PCS: Perceived Competence Scale.
[DOCX File , 25 KB - mhealth_v13i1e60855_app2.docx]

Multimedia Appendix 3
Responses to the UTAUT (Unified Theory of Acceptance and Use of Technology) questionnaire.
[XLSX File (Microsoft Excel File), 13 KB - mhealth_v13i1e60855_app3.xlsx]

References

1. Wang H, Li N, Chivese T, Werfalli M, Sun H, Yuen L, IDF Diabetes Atlas Committee Hyperglycaemia in Pregnancy Special Interest Group. IDF Diabetes Atlas: estimation of global and regional gestational diabetes mellitus prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's criteria. Diabetes Res Clin Pract 2022 Jan;183:109050. [doi: 10.1016/j.diabres.2021.109050] [Medline: 34883186]
2. Bellamy LP, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet 2009 May 23;373(9677):1773-1779. [doi: 10.1016/S0140-6736(09)60731-5] [Medline: 19465232]
3. Kim C. Gestational diabetes: risks, management, and treatment options. Int J Womens Health 2010 Oct 07;2:339-351 [FREE Full text] [doi: 10.2147/IJWH.S13333] [Medline: 21151681]
4. Damm P, Houshmand-Oeregaard A, Kelstrup L, Lauenborg J, Mathiesen ER, Clausen TD. Gestational diabetes mellitus and long-term consequences for mother and offspring: a view from Denmark. Diabetologia 2016 Jul;59(7):1396-1399. [doi: 10.1007/s00125-016-3985-5] [Medline: 27174368]
5. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
6. Moschonis G, Siopis G, Jung J, Eweka E, Willems R, Kwasnicka D, DigiCare4You Consortium. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e125-e143 [FREE Full text] [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]
7. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov 13;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
8. Wu IXY, Kee JCY, Threapleton DE, Ma RCW, Lam VCK, Lee EKP, et al. Effectiveness of smartphone technologies on glycaemic control in patients with type 2 diabetes: systematic review with meta-analysis of 17 trials. Obes Rev 2018 Jun;19(6):825-838. [doi: 10.1111/obr.12669] [Medline: 29345109]

XSL•FO
RenderX

9. Chaugule S, Graham C. Cost-effectiveness of G5 Mobile continuous glucose monitoring device compared to self-monitoring of blood glucose alone for people with type 1 diabetes from the Canadian societal perspective. J Med Econ 2017 Nov;20(11):1128-1135 [FREE Full text] [doi: 10.1080/13696998.2017.1360312] [Medline: 28745578]
10. Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. Am J Obstet Gynecol 2018 Apr;218(4):453.e1-453.e7. [doi: 10.1016/j.ajog.2018.01.044] [Medline: 29425836]
11. Adesina N, Dogan H, Green S, Tsofliou F. Effectiveness and usability of digital tools to support dietary self-management of gestational diabetes mellitus: a systematic review. Nutrients 2021 Dec 21;14(1):10 [FREE Full text] [doi: 10.3390/nu14010010] [Medline: 35010884]
12. Xie W, Dai P, Qin Y, Wu M, Yang B, Yu X. Effectiveness of telemedicine for pregnant women with gestational diabetes mellitus: an updated meta-analysis of 32 randomized controlled trials with trial sequential analysis. BMC Pregnancy Childbirth 2020 Apr 06;20(1):198 [FREE Full text] [doi: 10.1186/s12884-020-02892-1] [Medline: 32252676]
13. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 09;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
14. Guo H, Zhang Y, Li P, Zhou P, Chen LM, Li SY. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
15. Yew TW, Chi C, Chan SY, van Dam RM, Whitton C, Lim CS, et al. A randomized controlled trial to evaluate the effects of a smartphone application-based lifestyle coaching program on gestational weight gain, glycemic control, and maternal and neonatal outcomes in women with gestational diabetes mellitus: the SMART-GDM study. Diabetes Care 2021 Feb;44(2):456-463 [FREE Full text] [doi: 10.2337/dc20-1216] [Medline: 33184151]
16. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]
17. Mamykina L, Heitkemper EM, Smaldone AM, Kukafka R, Cole-Lewis HJ, Davidson PG, et al. Personal discovery in diabetes self-management: discovering cause and effect using self-monitoring data. J Biomed Inform 2017 Dec;76:1-8 [FREE Full text] [doi: 10.1016/j.jbi.2017.09.013] [Medline: 28974460]
18. Raj S, Toporski K, Garrity A, Lee JM, Newman MW. "My blood sugar is higher on the weekends": finding a role for context and context-awareness in the design of health self-management technology. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300349]
19. Karkar R, Zia J, Vilardaga R, Mishra SR, Fogarty J, Munson SA, et al. A framework for self-experimentation in personalized health. J Am Med Inform Assoc 2016 May;23(3):440-448 [FREE Full text] [doi: 10.1093/jamia/ocv150] [Medline: 26644399]
20. Kytö M, Markussen LT, Marttinen P, Jacucci G, Niinistö S, Virtanen SM, et al. Comprehensive self-tracking of blood glucose and lifestyle with a mobile application in the management of gestational diabetes: a study protocol for a randomised controlled trial (eMOM GDM study). BMJ Open 2022 Nov 07;12(11):e066292 [FREE Full text] [doi: 10.1136/bmjopen-2022-066292] [Medline: 36344008]
21. Kytö M, Hotta S, Niinistö S, Marttinen P, Korhonen TE, Markussen LT, et al. Periodic mobile application (eMOM) with self-tracking of glucose and lifestyle improves treatment of diet-controlled gestational diabetes without human guidance: a randomized controlled trial. Am J Obstet Gynecol 2024 Nov;231(5):541.e1-541.e16. [doi: 10.1016/j.ajog.2024.02.303] [Medline: 38432415]
22. Gestational diabetes. Current care guidelines.: The Finnish Medical Society Duodecim; 2022. URL: https://www.kaypahoito.fi/hoi50068 [accessed 2023-11-23]
23. Kytö M, Koivusalo S, Tuomonen H, Strömberg L, Ruonala A, Marttinen P, et al. Supporting the management of gestational diabetes mellitus with comprehensive self-tracking: mixed methods study of wearable sensors. JMIR Diabetes 2023 Oct 31;8:e43979 [FREE Full text] [doi: 10.2196/43979] [Medline: 37906216]
24. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes Care 1998 Oct;21(10):1644-1651. [doi: 10.2337/diacare.21.10.1644] [Medline: 9773724]
25. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
26. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
27. Kirakowski J, Corbett M. SUMI: the Software Usability Measurement Inventory. Brit J Educational Tech 2006 Oct 27;24(3):210-212. [doi: 10.1111/j.1467-8535.1993.tb00076.x]
28. Parak J, Korhonen I. Accuracy of Firstbeat Bodyguard 2 beat-to-beat heart rate monitor. Firstbeat. 2015. URL: https://assets.firstbeat.com/firstbeat/uploads/2015/10/white_paper_bodyguard2_final.pdf [accessed 2024-03-10]

XSL•FO
RenderX

29. Kuula L, Pesonen AK. Heart rate variability and Firstbeat method for detecting sleep stages in healthy young adults: feasibility study. JMIR Mhealth Uhealth 2021 Feb 03;9(2):e24704 [FREE Full text] [doi: 10.2196/24704] [Medline: 33533726]
30. Föhr T, Tolvanen A, Myllymäki T, Järvelä-Reijonen E, Peuhkuri K, Rantala S, et al. Physical activity, heart rate variability-based stress and recovery, and subjective stress during a 9-month study period. Scand J Med Sci Sports 2017 Jun;27(6):612-621. [doi: 10.1111/sms.12683] [Medline: 27037904]
31. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
32. Kazma JM, van den Anker J, Allegaert K, Dallmann A, Ahmadzia HK. Anatomical and physiological alterations of pregnancy. J Pharmacokinet Pharmacodyn 2020 Aug;47(4):271-285 [FREE Full text] [doi: 10.1007/s10928-020-09677-1] [Medline: 32026239]
33. Evenson KR, Savitz DA, Huston SL. Leisure-time physical activity among pregnant women in the US. Paediatr Perinat Epidemiol 2004 Nov;18(6):400-407. [doi: 10.1111/j.1365-3016.2004.00595.x] [Medline: 15535815]
34. Mottola MF, Campbell MK. Activity patterns during pregnancy. Can J Appl Physiol 2003 Aug;28(4):642-653. [doi: 10.1139/h03-049] [Medline: 12904639]
35. Borodulin K, Evenson KR, Herring AH. Physical activity patterns during pregnancy through postpartum. BMC Womens Health 2009 Nov 19;9:32 [FREE Full text] [doi: 10.1186/1472-6874-9-32] [Medline: 19925650]
36. Pereira MA, Rifas-Shiman SL, Kleinman KP, Rich-Edwards JW, Peterson KE, Gillman MW. Predictors of change in physical activity during and after pregnancy: Project Viva. Am J Prev Med 2007 Apr;32(4):312-319 [FREE Full text] [doi: 10.1016/j.amepre.2006.12.017] [Medline: 17383562]
37. Årsand E, Tatara N, Østengen G, Hartvigsen G. Mobile phone-based self-management tools for type 2 diabetes: the few touch application. J Diabetes Sci Technol 2010 Mar 01;4(2):328-336 [FREE Full text] [doi: 10.1177/193229681000400213] [Medline: 20307393]
38. Hsiao WH, Paterno MT, Iradukunda F, Hawkins' M. The preliminary efficacy of a sleep self-management intervention using a personalized health monitoring device during pregnancy. Behav Sleep Med 2021;19(6):705-716 [FREE Full text] [doi: 10.1080/15402002.2020.1851230] [Medline: 33245245]
39. Desai PM, Levine ME, Albers DJ, Mamykina L. Pictures worth a thousand words: reflections on visualizing personal blood glucose forecasts for individuals with Type 2 Diabetes. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing System; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3174112]
40. Coşkun A, Karahanoğlu A. Data sensemaking in self-tracking: towards a new generation of self-tracking tools. Int J Hum Comput Interact 2022 May 26;39(12):2339-2360. [doi: 10.1080/10447318.2022.2075637]
41. Katz DS, Price BA, Holland S, Dalton NS. Data, data everywhere, and still too hard to link: insights from user interactions with diabetes apps. 2018 Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3174077]
42. Katz DS, Price BA, Holland S, Dalton NS. Designing for diabetes decision support systems with fluid contextual reasoning. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3174199]
43. Mamykina L, Heitkemper EM, Smaldone AM, Kukafka R, Cole-Lewis HJ, Davidson PG, et al. Personal discovery in diabetes self-management: discovering cause and effect using self-monitoring data. J Biomed Inform 2017 Dec;76:1-8 [FREE Full text] [doi: 10.1016/j.jbi.2017.09.013] [Medline: 28974460]
44. Ibrahim Z, Caldeira C, Chung CF. Supporting experiential learning in people with gestational diabetes mellitus. 2024 Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, HI. [doi: 10.1145/3613904.3642674]
45. Carolan M. Women's experiences of gestational diabetes self-management: a qualitative study. Midwifery 2013 Jun;29(6):637-645. [doi: 10.1016/j.midw.2012.05.013] [Medline: 22877761]
46. Carolan M, Gill GK, Steele C. Women's experiences of factors that facilitate or inhibit gestational diabetes self-management. BMC Pregnancy Childbirth 2012 Sep 18;12:99 [FREE Full text] [doi: 10.1186/1471-2393-12-99] [Medline: 22988897]
47. Parsons J, Ismail K, Amiel S, Forbes A. Perceptions among women with gestational diabetes. Qual Health Res 2014 Apr;24(4):575-585. [doi: 10.1177/1049732314524636] [Medline: 24682021]
48. Draffin CR, Alderdice FA, McCance DR, Maresh M, Harper Md Consultant Physician R, McSorley O, et al. Exploring the needs, concerns and knowledge of women diagnosed with gestational diabetes: a qualitative study. Midwifery 2016 Sep;40:141-147. [doi: 10.1016/j.midw.2016.06.019] [Medline: 27553869]
49. Safiee L, Rough DJ, Whitford H. Barriers to and facilitators of using eHealth to support gestational diabetes mellitus self-management: systematic literature review of perceptions of health care professionals and women with gestational diabetes mellitus. J Med Internet Res 2022 Oct 27;24(10):e39689 [FREE Full text] [doi: 10.2196/39689] [Medline: 36301613]
50. Law GR, Alnaji A, Alrefaii L, Endersby D, Cartland SJ, Gilbey SG, et al. Suboptimal nocturnal glucose control is associated with large for gestational age in treated gestational diabetes mellitus. Diabetes Care 2019 May;42(5):810-815 [FREE Full text] [doi: 10.2337/dc18-2212] [Medline: 30765428]

XSL•FO
RenderX

51. Jovanovic L. The role of continuous glucose monitoring in gestational diabetes mellitus. Diabetes Technol Ther 2000;2(Suppl 1):S67-S71. [doi: 10.1089/15209150050214159] [Medline: 11469635]
52. Scott EM, Bilous RW, Kautzky-Willer A. Accuracy, user acceptability, and safety evaluation for the FreeStyle Libre Flash Glucose Monitoring System when used by pregnant women with diabetes. Diabetes Technol Ther 2018 Mar;20(3):180-188 [FREE Full text] [doi: 10.1089/dia.2017.0386] [Medline: 29470094]
53. Voormolen DN, DeVries JH, Sanson RME, Heringa MP, de Valk HW, Kok M, et al. Continuous glucose monitoring during diabetic pregnancy (GlucoMOMS): a multicentre randomized controlled trial. Diabetes Obes Metab 2018 Aug;20(8):1894-1902. [doi: 10.1111/dom.13310] [Medline: 29603547]
54. Lane AS, Mlynarczyk MA, de Veciana M, Green LM, Baraki DI, Abuhamad AZ. Real-time continuous glucose monitoring in gestational diabetes: a randomized controlled trial. Am J Perinatol 2019 Jul;36(9):891-897. [doi: 10.1055/s-0039-1678733] [Medline: 30818406]
55. Wei Q, Sun Z, Yang Y, Yu H, Ding H, Wang S. Effect of a CGMS and SMBG on maternal and neonatal outcomes in gestational diabetes mellitus: a randomized controlled trial. Sci Rep 2016 Jan 27;6:19920. [doi: 10.1038/srep19920] [Medline: 26814139]
56. Persson M, Winkvist A, Mogren I. 'From stun to gradual balance': women's experiences of living with gestational diabetes mellitus. Scand J Caring Sci 2010 Sep;24(3):454-462. [doi: 10.1111/j.1471-6712.2009.00735.x] [Medline: 19878489]
57. Catalano PM, Huston L, Amini SB, Kalhan SC. Longitudinal changes in glucose metabolism during pregnancy in obese women with normal glucose tolerance and gestational diabetes mellitus. Am J Obstet Gynecol 1999 Apr;180(4):903-916. [doi: 10.1016/s0002-9378(99)70662-9] [Medline: 10203659]
58. Mamykina L, Mynatt E, Davidson P, Greenblatt D. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. 2008 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 5-10, 2008; Florence, Italy. [doi: 10.1145/1357054.1357131]
59. Evans MK, O'Brien B. Gestational diabetes: the meaning of an at-risk pregnancy. Qual Health Res 2005 Jan;15(1):66-81. [doi: 10.1177/1049732304270825] [Medline: 15574716]
60. Williams GC, McGregor HA, Zeldman A, Freedman ZR, Deci EL. Testing a self-determination theory process model for promoting glycemic control through diabetes self-management. Health Psychol 2004 Jan;23(1):58-66. [doi: 10.1037/0278-6133.23.1.58] [Medline: 14756604]
61. Kytö M, Koivusalo S, Ruonala A, Strömberg L, Tuomonen H, Heinonen S, et al. Behavior change app for self-management of gestational diabetes: design and evaluation of desirable features. JMIR Hum Factors 2022 Oct 12;9(4):e36987. [doi: 10.2196/36987] [Medline: 36222806]
62. Peleg M, Shahar Y, Quaglini S, Fux A, García-Sáez G, Goldstein A, et al. MobiGuide: a personalized and patient-centric decision-support system and its evaluation in the atrial fibrillation and gestational diabetes domains. User Model User-Adap Inter 2017 Mar 11;27(2):159-213. [doi: 10.1007/s11257-017-9190-5]
63. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in Norway. BMJ Open 2019 Nov 11;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
64. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020 Nov 06;8(11):e22212 [FREE Full text] [doi: 10.2196/22212] [Medline: 32975198]
65. Franciosi M, Pellegrini F, De Berardis G, Belfiglio M, Cavaliere D, Di Nardo B, QuED Study Group. The impact of blood glucose self-monitoring on metabolic control and quality of life in type 2 diabetic patients: an urgent need for better educational strategies. Diabetes Care 2001 Nov;24(11):1870-1877. [doi: 10.2337/diacare.24.11.1870] [Medline: 11679449]
66. O'Kane MJ, Bunting B, Copeland M, Coates VE, ESMON Study Group. Efficacy of self monitoring of blood glucose in patients with newly diagnosed type 2 diabetes (ESMON Study): randomised controlled trial. BMJ 2008 May 24;336(7654):1174-1177 [FREE Full text] [doi: 10.1136/bmj.39534.571644.BE] [Medline: 18420662]
67. Al Sayah F, Majumdar SR, Williams B, Robertson S, Johnson JA. Health literacy and health outcomes in diabetes: a systematic review. J Gen Intern Med 2013 Mar;28(3):444-452 [FREE Full text] [doi: 10.1007/s11606-012-2241-z] [Medline: 23065575]
68. Chen Q, Carbone ET. Functionality, implementation, impact, and the role of health literacy in mobile phone apps for gestational diabetes: scoping review. JMIR Diabetes 2017 Oct 04;2(2):e25 [FREE Full text] [doi: 10.2196/diabetes.8045] [Medline: 30291088]
69. Rooksby J, Rost M, Morrison A, Chalmers M. Personal tracking as lived informatics. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; May 1, 2014; Toronto, ON. [doi: 10.1145/2556288.2557039]

Abbreviations

CGM: continuous glucose monitor
GDM: gestational diabetes mellitus
HRV: heart rate variability

XSL•FO
RenderX

PCS: Perceived Competence Scale
RCT: randomized controlled trial
TSRQ: Treatment Self-Regulation Questionnaire

Edited by M Waring; submitted 28.05.24; peer-reviewed by O Taiwo, KR Konnaiyan, E Pustozerov; comments to author 01.10.24; revised version received 23.11.24; accepted 26.03.25; published 04.06.25.

Please cite as:
Määttänen S, Koivusalo S, Ylinen H, Heinonen S, Kytö M
The Effect of a Mobile App (eMOM) on Self-Discovery and Psychological Factors in Persons With Gestational Diabetes: Mixed Methods Study
JMIR Mhealth Uhealth 2025;13:e60855
URL: https://mhealth.jmir.org/2025/1/e60855
doi:10.2196/60855
PMID:40466096

©Sini Määttänen, Saila Koivusalo, Hanna Ylinen, Seppo Heinonen, Mikko Kytö. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 04.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Culturally and Linguistically Tailored Intervention to Improve Diabetes-Related Outcomes in Chinese Americans With Type 2 Diabetes: Pilot Randomized Controlled Trial

Jing Liu[1*], PhD; Jiepin Cao[1*], PhD; Yun Shi[1*], PhD; Mary Ann Sevick[1], ScD; Nadia Islam[1], PhD; Naumi Feldman[2], DrPH; Huilin Li[1], PhD; Chan Wang[1], PhD; Yanan Zhao[1], PhD; Kosuke Tamura[3], PhD; Natalie Levy[4], MD; Nan Jiang[1], PhD; Ziqiang Zhu[5], MD; Yulin Wang[6], MD; Jia Hong[7], MD; Lu Hu[1], PhD

[1]Department of Population Health, NYU Langone Health, New York City, NY, United States

[2]Charles B. Wang Community Health Center, New York City, NY, United States

[3]Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States

[4]Department of Medicine, NYU Langone Health, New York City, NY, United States

[5]Wellsure Medical Practice, New York City, NY, United States

[6]Yulin Medicine PLLC, New York City, NY, United States

[7]Hong Jia Medical PC, New York City, NY, United States

[*]these authors contributed equally

Corresponding Author:
Lu Hu, PhD
Department of Population Health
NYU Langone Health
180 Madison Avenue
New York City, NY, 10016
United States
Phone: 1 6465013438
Email: lu.hu@nyulangone.org

Abstract

Background: Chinese Americans with type 2 diabetes (T2D) face many linguistic, cultural, and social determinants of health–related barriers to accessing evidence-based diabetes interventions. Our team developed the culturally and linguistically tailored Chinese American Research and Education (CARE) program to provide evidence-based diabetes education and support to this group and demonstrated the feasibility, acceptability, and potential efficacy of the intervention on improving hemoglobin A_{1c} levels. However, it remains unclear whether the CARE program also improves diabetes self-efficacy and psychosocial outcomes in the same study sample.

Objective: This is a secondary analysis to examine the potential efficacy of the CARE program on secondary outcomes, including diabetes self-efficacy, self-care activities, beliefs in diabetes self-care activities, and diabetes distress among Chinese Americans with T2D.

Methods: A 2-arm, pilot randomized controlled trial was conducted to evaluate the CARE program between March 1, 2021, and April 21, 2023. The trial included 60 Chinese Americans aged 18 to 70 years who had a diagnosis of T2D and a baseline hemoglobin A_{1c} level of 7% or higher. Participants were recruited from various health care settings in New York City, including community health centers, private primary care providers, and NYU Langone Health and its affiliates, and were randomly assigned to either the CARE intervention group (n=30) or a waitlist control group (n=30). The intervention consisted of 2 culturally and linguistically tailored educational videos per week for 12 weeks, covering diabetes self-care topics such as healthy eating, physical activity, and medication adherence. These videos were delivered via the WeChat app. In addition, community health workers provided support calls to assist them in setting goals, problem-solving, and addressing social determinants of health barriers every 2 weeks. Secondary outcomes included patient self-reported diabetes self-efficacy, self-care activities, beliefs in diabetes self-care activities, and diabetes distress. Outcomes were assessed at baseline, 3 months, and 6 months.

XSL•FO
RenderX

Results: Participants had a mean age of 54.3 (SD 11.5) years and 62% (37/60) were male, 78% (47/60) were married, 58% (35/60) were employed, 70% (42/60) had a high school education or lower, and 88% (53/60) reported limited English proficiency. Intervention participants demonstrated statistically significant improvements in self-efficacy at 3 months (estimated difference in change: 8.47; 95% CI 2.44-14.5; adjusted *P*=.02), diabetes distress at 6 months (estimated difference in change: –0.43; 95% CI –0.71 to –0.15; adjusted *P*=.009), and adherence to a healthy diet at both 3 months (estimated difference in change: 1.61; 95% CI 0.46-2.75; adjusted *P*=.02) and 6 months (estimated difference in change: 1.64; 95% CI 0.48-2.81; adjusted *P*=.02).

Conclusions: The culturally and linguistically tailored intervention showed promise in improving self-efficacy and diabetes self-care activities among Chinese Americans with T2D, warranting validation through a large-scale randomized controlled trial.

Trial Registration: ClinicalTrials.gov NCT03557697; https://clinicaltrials.gov/study/NCT03557697.

(JMIR Mhealth Uhealth 2025;13:e78036) doi:10.2196/78036

KEYWORDS

type 2 diabetes; diabetes self-management; health equity; minority health; immigrant health; mobile health; social media; community health worker; social needs; social determinants of health

Introduction

Chinese Americans are among the fastest-growing racial and ethnic minority groups in the United States and face a significant burden of type 2 diabetes (T2D). The prevalence of T2D in this population is remarkably higher than in that of non-Hispanic White Americans [1,2] despite generally lower BMI levels [3,4]. Beyond the high prevalence, Chinese Americans often experience worse outcomes and complications due to social determinants of health (SDOH) barriers, such as limited access to health care services, financial constraints, limited English proficiency, and cultural barriers.All these factors can impede adherence to recommended T2D self-care practices to achieve optimal glycemic control.

Self-care behaviors are essential components of effective T2D management, such as medication adherence, healthy eating, regular physical activity, and routine blood glucose monitoring [5,6]. Self-efficacy, referred to as an individual's confidence in their ability to manage the illness [6,7], is a well-established predictor of engagement in these self-care practices and diabetes distress [8,9]. Higher levels of self-care behaviors and self-efficacy have been associated with improved glycemic control [5,9]. Notably, a prior study of 3067 older participants from the 2009 California Health Interview Survey demonstrated that Asian individuals with diabetes exhibit the lowest levels of diabetes self-efficacy when compared to non-Hispanic White, American Indian or Alaska Native, African American, and Hispanic or Latino populations [10]. Asian individuals were also significantly less likely than White individuals to regularly test their blood glucose or receive foot examinations in accordance with clinical guidelines [11,12]. For Chinese Americans, the largest Asian subgroup, enhancing self-efficacy and self-care practices may be particularly impactful, as this population often faces significant cultural and linguistic barriers to accessing diabetes services. These disparities underscore the pressing need for easily accessible, tailored interventions to strengthen self-efficacy and promote sustained self-care behaviors among Chinese Americans.

Diabetes self-management education and support (DSMES) programs have been shown to be effective in promoting diabetes self-efficacy and self-care behaviors and improving glycemic control among patients with T2D [13,14]. Yet, these evidence-based programs are underused by Chinese Americans. Existing DSMES programs often lack cultural relevance, making it difficult for Chinese Americans to apply the recommendations to their diet [15,16]. Limited English proficiency also poses a substantial communication barrier between Chinese American patients and health care providers [17]. Furthermore, the requirement for in-person visits in traditional DSMES programs can be impractical for many Chinese Americans, who often face transportation challenges and work long hours [18,19]. Therefore, supportive interventions that are easily accessible and address the cultural and linguistic needs of this population are required.

To address these challenges, we developed the Chinese American Research and Education (CARE) program by culturally and linguistically adapting the DSMES intervention specifically for Chinese Americans with T2D [20]. The CARE adaptation was guided by both the Cultural Adaptation Model [21] and the Ecological Validity Model [22]. The CARE program is a 12-week intervention grounded in Social Cognitive Theory, which posits that self-efficacy is a crucial determinant of behavior change [23]. It includes 24 brief videos covering key topics in T2D self-management, such as diabetes basics, healthy eating, physical activity, medication adherence, glucose monitoring, and behavioral techniques (eg, goal-setting and problem-solving). Examples of the CARE videos are presented in Multimedia Appendix 1. Additionally, community health workers provide supportive phone calls every 2 weeks to participants to assist them in setting goals and problem-solving and address specific SDOH challenges encountered in their self-management efforts.

The CARE program was previously evaluated through a pilot randomized controlled trial (RCT) to assess its feasibility, acceptability, and potential efficacy on primary outcomes, including hemoglobin A_{1c} (HbA_{1c}) levels, BMI, physical activities, and diet, as reported elsewhere [24]. The present study is a secondary analysis of the CARE trial, focusing on its potential impact on diabetes self-efficacy, self-care activities, beliefs in diabetes self-care activities, and diabetes distress among Chinese immigrants with T2D. We hypothesize that participants in the CARE intervention group will demonstrate significantly greater improvements in diabetes self-efficacy,

XSL•FO
RenderX

self-care activities, and beliefs in diabetes self-care activities and reductions in diabetes distress compared to those in the waitlist control group at both the 3- and 6-month follow-ups.

Methods

Design

The current study is a secondary analysis of the pilot RCT that examined the potential efficacy of the CARE program on HbA_{1c} levels. A total of 60 Chinese American participants with T2D were enrolled and randomly assigned in a 1:1 ratio to either the CARE intervention group or a waitlist control group. This study follows the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Multimedia Appendix 2) [25]. Detailed descriptions of the pilot RCT have been published elsewhere [20,24].

Participants and Recruitment

Participants were recruited from March 2021 to March 2022 from various New York City–based health care facilities, including the Charles B Wang Community Health Center, private primary care providers, and NYU Langone Health and its affiliates. Recruitment strategies included (1) distributing study flyers, (2) direct referrals from health care providers, and (3) NYU Langone Health electronic medical records. Study staff members screened interested individuals by telephone to confirm eligibility.

Participants were eligible if they met the following criteria: (1) self-identified as Chinese immigrants or Chinese Americans; (2) were between the ages of 18 and 70 years; (3) were proficient in Mandarin, one of the most widely spoken Chinese languages; (4) had a diagnosis of T2D with a most recent HbA_{1c} level of 7% or higher at baseline; (5) used WeChat, a free social media app widely used among Chinese Americans; (6) owned a smartphone or tablet or agreed to use a study-provided device; and (7) were interested in receiving T2D-related videos via WeChat.

Exclusion criteria included significant uncorrected vision or hearing impairments, reluctance to comply with random assignments, pregnancy or plans to become pregnant during the study, breastfeeding, and residing in environments impeding T2D self-management, such as nursing homes.

Intervention

CARE Intervention Group

Based on Social Cognitive Theory [23], the CARE intervention aims to enhance self-efficacy in T2D self-management by addressing mastery experiences, social modeling, verbal persuasion, and physiological states. Participants in the intervention group received 2 culturally and linguistically tailored DSMES videos per week for 12 weeks (3 months) through WeChat. These videos covered various T2D self-care topics, including diabetes basics, healthy eating, physical activity, medication adherence, glucose monitoring, and behavioral techniques such as goal-setting and problem-solving. These videos were delivered in Mandarin. Additionally, participants received support calls from trained community health workers every 2 weeks, who assisted them in setting goals, problem-solving, and addressing SDOH barriers. After 12 weeks of video education and support calls, participants in the intervention group received usual care (3-6 months).

Waitlist Control Group

Waitlist control group participants received usual care. After the study ended (at 6 months), they were given access to the same DSMES educational videos via WeChat as an incentive.

Blinding

The blinding of participants and community health workers who delivered the intervention was not feasible due to the nature of the intervention. However, the data management and statistical analyses were conducted by analysts who were blinded to the allocation.

Measures

Demographic Information

Information on age, gender, marital status, education, annual income, employment status, English proficiency, and years of residency in the United States were collected at baseline.

Self-Efficacy for Diabetes

The Stanford 8-item Self-Efficacy for Diabetes scale was used to evaluate participants' confidence in managing their diabetes through various self-care behaviors, including diet, physical activity, blood glucose monitoring, and self-control [26]. Participants rated their confidence on a 10-point Likert scale, ranging from 1 (not at all confident) to 10 (totally confident). The total score was calculated, with higher scores indicating greater self-efficacy.

Diabetes Self-Care Activities

The Summary of Diabetes Self-Care Activities was used to assess specific diabetes self-care activities across general diet (ie, following a healthful eating plan), specific diet (ie, intake of fruits and vegetables and high-fat foods), exercise, self-monitoring of blood glucose, foot care, and medication adherence [27]. Participants indicated how many days in the past week they engaged in these specified behaviors. The mean number of days adherent to each type of diabetes self-care activity per week was calculated.

Beliefs in Diabetes Self-Care Activities

Participants rated the importance of 5 self-care behaviors related to diabetes-following a diabetic diet, exercising, self-monitoring blood glucose, taking medication, and checking their feet. Each behavior was rated on a 5-point Likert scale, where 1 indicated "not important" and 5 indicated "very important" [28].

Diabetes Distress

The Diabetes Distress Scale was used to evaluate 4 domains of diabetes-related distress: emotional burden, physician-related distress, regimen-related distress, and interpersonal distress [29]. This 17-item scale required participants to rate the extent to which each item was problematic for them in the past month using a 6-point Likert scale ranging from 1 (no problem) to 6

XSL•FO
RenderX

(serious problem). Mean item scores were calculated, with higher scores indicating a greater level of distress.

Sample Size Calculation

This study is a pilot RCT primarily aimed at establishing the feasibility, acceptability, and potential efficacy of the CARE program on HbA_{1c} levels. The sample size of this pilot RCT was determined based on HbA_{1c} levels, which is the most commonly evaluated primary outcome when testing DSMES programs. According to Whitehead et al [30], in order to detect a small group difference of 0.2% in HbA_{1c} levels with 80% power and 5% type I error in a future full-scale RCT, we would need to recruit 50 participants in a pilot RCT. To account for a 20-25% attrition rate observed in previous studies with Chinese Americans [31-34], we aimed to enroll 60 participants, with 30 in each of the intervention and control groups.

Data Analysis

We followed the intention-to-treat principle in data analysis. We used chi-square tests and independent 2-sample *t* tests to compare variables between the intervention and control groups. We applied piecewise linear mixed models to compare changes in outcomes over 0-6 months between the intervention and control groups. In these models, the 2 time periods (0-3 months and 3-6 months), the intervention group, and the intervention × time period interactions were modeled as fixed effects. We considered the participant as a random effect and included variables that showed statistically significant differences as covariates for adjustment. We applied the Tukey method to adjust *P* values for differences across time points within each group and the Bonferroni method to adjust *P* values for differences in changes over time between the 2 groups. We used R software (version 4.4.2; R Foundation for Statistical Computing) for statistical analyses and set the significance level at .05 for 2-tailed tests.

In addition to statistical significance, this study used the minimal clinically important difference (MCID) to assess the clinical significance of between-group differences. In this study, an MCID of 0.25 was considered clinically significant for the Diabetes Distress Scale [35]. As no established MCIDs are available for the Stanford 8-item Self-Efficacy for Diabetes Scale, the Summary of Diabetes Self-Care Activities, or the Beliefs in Diabetes Self-Care Activities, a threshold of 6% to 10% of the total score was used for these scales, as recommended by previous research [36]. Therefore, an MCID of 4.8-8.0 points on the Self-Efficacy for Diabetes Scale (total score: 80), 0.42-0.7 on the Summary of Diabetes Self-Care Activities (total score: 7), and 0.3-0.5 on the Beliefs in Diabetes Self-Care Activities (total score: 5) were used to evaluate the clinical significance of these outcomes. Clinical significance was determined by comparing the 95% CI of the observed differences to the MCID threshold for each scale: differences were considered “clinically significant” if the lower bound of the 95% CI exceeded the threshold, “may be significant” if the 95% CI included the threshold, and “not clinically significant” if the upper bound of the 95% CI was below the threshold [37].

Ethical Considerations

This study was conducted in accordance with the principles of the Declaration of Helsinki. The study was reviewed and approved by the NYU Grossman School of Medicine Institutional Review Board (IRB# s18-00609). All participants took part in the study voluntarily. Informed consent was obtained from all participants prior to their enrollment in the study. Participants’ identifiable information was deidentified before data analysis to ensure their privacy and confidentiality. Participants received US $90 for completing the baseline, 3-month, and 6-month follow-ups. Participants from both groups were also compensated US $5 per week (up to US $60) to offset mobile data costs related to viewing intervention videos. The compensation was provided via a ClinCard.

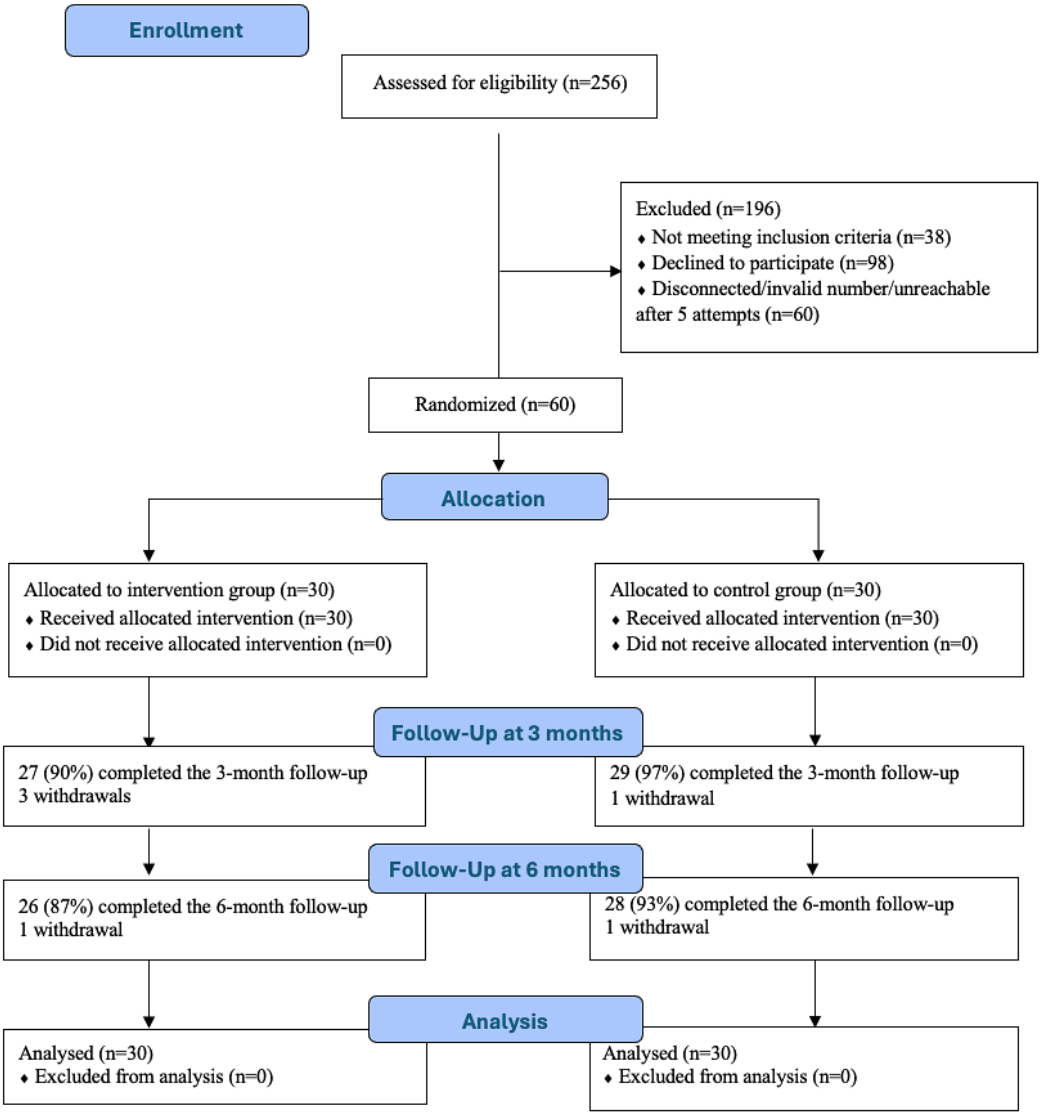
Results

Sample Characteristics

A CONSORT flow diagram is presented in Figure 1. The mean age of participants was 54.3 (SD 11.5) years, with 62% (37/60) self-identifying as male, 78% (47/60) being married, 58% (35/60) being currently employed, 70% (42/60) having a high school education or lower, 75% (45/60) reporting an annual household income under $55,000, and 88% (53/60) indicating limited English proficiency. At baseline, the mean HbA_{1c} level was 8.2% (SD 1.2%) and the mean BMI was 26.5 (SD 5.1) kg/m^2 (Asian individuals with a BMI of 23 kg/m^2 or greater are considered to be overweight). There were no significant differences in these demographic characteristics between the intervention and control groups. See Table 1 for details.

XSL•FO
RenderX

Figure 1. The CONSORT flow diagram.



XSL•FO
RenderX

Table 1. Sample characteristics (N=60).

Characteristics	Total (N=60)	CARE[a] group (n=30)	Control group (n=30)	*P* value
Age (years), mean (SD)	54.3 (11.5)	55.6 (11.5)	52.9 (11.4)	.36
Gender, n (%)				>.99
Men	37 (62)	18 (60)	19 (63)	
Women	23 (38)	12 (40)	11 (37)	
Being married, n (%)	47 (78)	23 (77)	24 (80)	>.99
Education, n (%)				.40
≤ high school education	42 (70)	19 (63)	23 (77)	
> high school education	18 (30)	11 (37)	7 (23)	
Annual household income (US $), n (%)				.80
≤55,000	45 (75)	23 (77)	22 (73)	
>55,000	12 (20)	5 (17)	7 (23)	
Missing	3 (5)	2 (7)	1 (3)	
Employment status, n (%)				.30
Employed	35 (58)	14 (47)	21 (70)	
Unemployed	10 (17)	7 (23)	3 (10)	
Retired	13 (22)	8 (27)	5 (17)	
Other	2 (3)	1 (3)	1 (3)	
English proficiency, n (%)				.64
Very well	6 (10)	2 (7)	4 (13)	
Limited English proficiency	53 (88)	28 (93)	25 (83)	
Missing	1 (2)	0 (0)	1 (3)	
Duration of residency (years), mean (SD)	19.2 (10.8)	20.4 (11.4)	17.9 (10.2)	.38
Missing duration of residency data, n (%)	2 (3)	1(3)	1(3)	—[b]
HbA_{1c} (%), mean (SD)	8.24 (1.24)	8.35 (1.30)	8.12 (1.19)	.46
Body weight (lbs), mean (SD)	160 (42.9)	154 (34.0)	166 (50.3)	.31
BMI (kg/m²), mean (SD)	26.5 (5.1)	26.0 (4.3)	27.1 (5.8)	.45
Self-efficacy for diabetes, mean (SD)	55.1 (14.3)	53.3 (14.2)	56.9 (14.5)	.35
Missing self-efficacy for diabetes data, n (%)	2 (3)	1 (3)	1 (3)	—
Self-care activities, mean (SD)				
General diet	3.91 (2.59)	3.60 (2.67)	4.22 (2.51)	.36
Specific diet	2.69 (1.72)	2.62 (1.84)	2.77 (1.63)	.74
Exercise	2.88 (2.46)	2.53 (2.45)	3.23 (2.45)	.27
Blood glucose testing	2.52 (2.72)	2.70 (2.87)	2.33 (2.59)	.66
Missing blood glucose testing data, n (%)	16 (27)	7 (23)	9 (3)	—
Foot Care	1.30 (1.81)	1.64 (2.00)	0.967 (1.56)	.16
Missing foot care data, n (%)	1 (2)	1 (3)	0	—
Medication	6.70 (0.91)	6.79 (0.82)	6.61 (0.99)	.44
Missing medication data, n (%)	2 (3)	1 (3)	1 (3)	—
Beliefs in diabetes self-care activities, mean (SD)	4.56 (0.56)	4.57 (0.58)	4.54 (0.55)	.85
Missing beliefs in diabetes self-care activities data, n (%)	2 (3)	1 (3)	1 (3)	—
Diabetes distress, mean (SD)	1.74 (0.84)	2.09 (1.01)	1.40 (0.43)	.002

XSL•FO
RenderX

Characteristics	Total (N=60)	CARE[a] group (n=30)	Control group (n=30)	*P* value
Missing diabetes distress data, n (%)	1 (2)	1 (3)	0	—

[a]CARE: Chinese American Research and Education.

[b]Not applicable.

Self-Efficacy for Diabetes

Table 2 presents the changes in participants' outcomes at baseline, 3 months, and 6 months. At baseline, the intervention group had a self-efficacy score of 53.3 (SD 14.2), while the control group scored 56.9 (SD 14.5). This difference was not statistically significant (*P*=.35).

Compared to the control group, participants in the intervention group demonstrated a statistically significantly greater increase in self-efficacy for diabetes management at 3 months (estimated difference in change: 8.47; 95% CI 2.44 to 14.5; adjusted *P*=.02). The observed between-group difference in self-efficacy at the 3-month follow-up may indicate a clinically meaningful improvement. However, at 6 months, the increase was not statistically or clinically significant (estimated difference in change: 5.87; 95% CI –0.23 to 11.98; adjusted *P*=.18).

XSL•FO
RenderX

Table 2. Outcomes of interests at baseline, 3 months, and 6 months.

Outcomes by time period (months)	CARE[a]				Control				Difference of change (CARE – control)			
	Est[b]	95% CI	Raw *P* value[c]	Adj *P* value[d]	Est	95% CI	Raw *P* value[c]	Adj *P* value[d]	Est	95% CI	Raw *P* value[c]	Adj *P* value[e]
Self-efficacy for diabetes												
0-3	5.76	(1.45 to 10.06)	.009	.03	–2.72	(–6.93 to 1.5)	.21	.41	8.47	(2.44 14.5)	.006	.02
3-6	0.32	(–4.07 to 4.72)	.88	.99	2.92	(–1.31 to 7.16)	.17	.36	–2.6	(–8.7 to 3.5)	.40	>.99
0-6	6.08	(1.72 to 10.44)	.007	.02	0.21	(–4.07 to 4.48)	.92	.99	5.87	(–0.23 to 11.98)	.06	.18
Self-care activities: general diet												
0-3	2.04	(1.21 to 2.86)	< .001	< .001	0.43	(–0.37 to 1.23)	.29	.54	1.61	(0.46 to 2.75)	.007	.02
3-6	–0.3	(–1.15 to 0.54)	.48	.76	–0.34	(–1.15 to 0.48)	.41	.70	0.04	(–1.14 to 1.21)	.95	>.99
0-6	1.74	(0.9 to 2.57)	< .001	< .001	0.09	(–0.72 to 0.91)	.82	.97	1.64	(0.48 to 2.81)	.006	.02
Self-care activities: specific diet												
0-3	0.17	(–0.47 to 0.81)	.60	.86	–0.31	(–0.93 to 0.32)	.33	.59	0.48	(–0.42 to 1.38)	.29	.88
3-6	0.42	(–0.24 to 1.08)	.21	.42	0.51	(–0.12 to 1.15)	.11	.25	–0.09	(–1.01 to 0.82)	.84	>.99
0-6	0.59	(–0.06 to 1.24)	.07	.17	0.2	(–0.43 to 0.84)	.53	.80	0.39	(–0.52 to 1.29)	.40	>.99
Self-care activities: exercise												
0-3	1.26	(0.43 to 2.1)	.003	.009	0.2	(–0.61 to 1.01)	.63	.88	1.07	(–0.1 to 2.23)	.07	.22
3-6	0.29	(–0.56 to 1.15)	.50	.76	0.69	(–0.13 to 1.52)	.10	.22	–0.4	(–1.59 to 0.79)	.51	>.99
0-6	1.56	(0.71 to 2.4)	< .001	.001	0.89	(0.07 to 1.72)	.03	.08	0.67	(–0.51 to 1.84)	.27	.80
Self-care activities: blood sugar testing												
0-3	0.47	(–0.52 to 1.46)	.35	.61	–0.7	(–1.68 to 0.28)	.16	.33	1.17	(–0.22 to 2.56)	.09	.30
3-6	–0.36	(–1.39 to 0.66)	.48	.76	0.56	(–0.45 to 1.58)	.27	.51	–0.93	(–2.37 to 0.51)	.20	.61
0-6	0.1	(–0.9 to 1.11)	.84	.98	–0.14	(–1.13 to 0.86)	.79	.96	0.24	(–1.18 to 1.66)	.74	>.99
Self-care activities: foot care												
0-3	–0.22	(–0.88 to 0.44)	.51	.78	0	(–0.63 to 0.63)	.99	>.99	–0.22	(–1.13 to 0.69)	.64	>.99
3-6	0.29	(–0.38 to 0.95)	.40	.67	–0.17	(–0.81 to 0.47)	.61	.86	0.45	(–0.47 to 1.38)	.33	>.99
0-6	0.07	(–0.6 to 0.73)	.85	.98	–0.17	(–0.81 to 0.47)	.60	.86	0.24	(–0.69 to 1.16)	.61	>.99
Self-care activities: medication												
0-3	–0.03	(–0.59 to 0.53)	.92	.99	0.06	(–0.49 to 0.62)	.82	.97	–0.09	(–0.88 to 0.7)	.82	>.99
3-6	–0.44	(–1.01 to 0.14)	.14	.30	–0.12	(–0.69 to 0.45)	.67	.91	–0.31	(–1.12 to 0.5)	.45	>.99

XSL•FO
RenderX

Outcomes by time period (months)	CARE[a]				Control				Difference of change (CARE – control)			
	Est[b]	95% CI	Raw *P* value[c]	Adj *P* value[d]	Est	95% CI	Raw *P* value[c]	Adj *P* value[d]	Est	95% CI	Raw *P* value[c]	Adj *P* value[e]
0-6	–0.46	(–1.03 to 0.11)	.11	.24	–0.06	(–0.63 to 0.51)	.84	.98	–0.41	(–1.21 to 0.4)	.32	.96
Beliefs in diabetes self-care activities												
0-3	0.05	(–0.15 to 0.24)	.64	.88	0.21	(0.02 to 0.4)	.03	.08	–0.17	(–0.44 to 0.11)	.24	.70
3-6	0.1	(–0.1 to 0.3)	.35	.61	0	(–0.2 to 0.19)	.98	>.99	0.1	(–0.18 to 0.38)	.49	>.99
0-6	0.14	(–0.06 to 0.34)	.16	.33	0.21	(0.02 to 0.4)	.04	.09	–0.07	(–0.35 to 0.21)	.63	>.99
Diabetes distress												
0-3	–0.24	(–0.44 to –0.04)	.02	.047	0.01	(–0.18 to 0.2)	.93	.99	–0.25	(–0.52 to 0.03)	.08	.23
3-6	–0.13	(–0.34 to 0.07)	.19	.39	0.05	(–0.15 to 0.25)	.62	.87	–0.18	(–0.47 to 0.1)	.2	.60
0-6	–0.37	(–0.58 to –0.17)	< .001	< .001	0.06	(–0.13 to 0.25)	.55	.82	–0.43	(–0.71 to –0.15)	.003	.009

[a]CARE: Chinese American Research and Education.

[b]Est: estimated difference in change.

[c]Initial *P* values without adjustment.

[d]Adjusted *P* values using the Tukey method.

[e]Adjusted *P* values using the Bonferroni method.

Self-Care Activities

At baseline, participants in the intervention group and control group did not differ in self-care activities (all *P*>.05). Specifically, participants in the intervention group followed their healthful eating plan for an average of 3.60 (SD 2.67) days per week, compared to 4.22 (SD 2.51) days per week in the control group, with no statistically significant difference (*P*=.36). For the specific diet, the intervention group followed it for 2.62 (SD 1.84) days per week, while the control group followed it for 2.77 (SD 1.63) days per week, which was not statistically significant (*P*=.74). Regarding exercise, the intervention group exercised for 2.53 (SD 2.45) days per week, compared to 3.23 (SD 2.45) days per week in the control group, with no significant difference (*P*=.27). Foot care was performed by the intervention group for 1.64 (SD 2.00) days per week, compared to 0.967 (SD 1.56) days per week in the control group, with no significant difference (*P*=.16). Finally, both groups adhered to their medication regimens similarly, with the intervention group adhering for 6.79 (SD 0.82) days per week and the control group for 6.61 (SD 0.99) days per week, with no statistically significant difference (*P*=.44).

Participants in the intervention group were more adherent to a healthy diet at 3 months (estimated difference in change: 1.61; 95% CI 0.46 to 2.75; adjusted *P*=.02) and 6 months (estimated difference in change: 1.64; 95% CI 0.48 to 2.81; adjusted *P*=.02) compared to the control group. The observed between-group differences in healthy diet at both the 3- and 6- month follow-ups may indicate a clinically meaningful improvement. There was no statistical or clinical significance in the between-group differences in other self-care activities at 3 months or 6 months, including exercise, blood glucose testing, foot care, and medication adherence (all *P*>.05).

Beliefs in Diabetes Self-Care Activities

At baseline, both the intervention group (mean 4.57, SD 0.58) and the control group (mean 4.54, SD 0.55) recognized the importance of self-management activities for their diabetes, including following a healthy diet, exercising, self-monitoring blood glucose, taking medication, and checking their feet. The difference was not statistically significant (*P*=.85). There were no statistically or clinically significant differences in beliefs about self-management at 3 months (adjusted *P*=.70) or 6 months (adjusted *P*>.99).

Diabetes Distress

At baseline, both the intervention group (mean 2.09, SD 1.01) and the control group (mean 1.40, SD 0.43) reported minimal levels of diabetes distress. However, participants in the intervention group reported significantly higher diabetes distress compared to those in the control group (*P*=.002).

Compared to the control group, participants in the intervention group reported a significant improvement in diabetes distress at 6 months (estimated difference in change: –0.43; 95% CI: –0.71 to –0.15; adjusted *P*=.009). The observed between-group differences in diabetes distress at the 6-month follow-up may indicate a clinically meaningful improvement. No statistically or clinically significant differences in diabetes-related distress

XSL•FO
RenderX

between the intervention and control groups were observed at 3 months (adjusted *P*=.23).

Discussion

Principal Findings

This study found that, compared to usual care, our 12-week CARE intervention, which included weekly diabetes education videos and biweekly support calls from community health workers, demonstrated statistically significant improvements in self-efficacy at 3 months, adherence to a healthy diet at both 3 and 6 months, and diabetes distress at 6 months. These between-group improvements may be clinically significant. These findings highlight the potential benefits of the CARE program, a culturally and linguistically tailored, social media–based intervention, in improving self-efficacy and diabetes self-care activities among Chinese Americans with T2D. The results suggest that culturally and linguistically tailored interventions may be a promising strategy for addressing the unique challenges faced by this population in managing T2D and improving diabetes-related health outcomes.

The intervention group demonstrated a significant increase in self-efficacy in managing diabetes compared to the control group, and this increase may have reached the threshold for clinical significance. Our findings contribute to the growing body of literature on the effectiveness of mobile health–based interventions aimed at diabetes education and self-management in enhancing self-efficacy, as evidenced in a recent systematic review [38]. Notably, none of the US-based interventions included in the review were specifically tailored for Chinese immigrants, an underserved population with a high burden of T2D and unique barriers to accessing these evidence-based programs [1-4,15-19]. By tailoring the intervention to the cultural and linguistic needs of the Chinese immigrants, the CARE program was able to engage participants in a more meaningful and effective way, which likely increased their confidence in adhering to self-care practices, a key determinant of change in other self-management behaviors [23].

The significant improvement in adherence to a healthy diet in the intervention group compared to the control group further supports the potential of the CARE intervention, particularly in addressing dietary changes, one of the most challenging aspects of T2D management for Chinese immigrants [15]. This improvement may represent a clinically significant change. This group typically follows a traditional Chinese diet [39,40], which is often not reflected in Western dietary guidelines commonly recommended in existing DSMES programs [16]. To address these challenges, the CARE program provided information on culturally appropriate food alternatives for Chinese Americans with T2D. It included video tutorials on reading food labels in local ethnic grocery stores. These strategies could facilitate the adoption of healthier dietary practices in this group. These findings underscore the importance of culturally tailored interventions in diabetes management for this immigrant population.

The intervention group demonstrated a significant improvement in managing diabetes distress compared to the control group. This improvement may be clinically significant. This finding aligns with previous interventional studies, which have reported that diabetes education interventions effectively reduce diabetes distress and offer emotional support to Chinese Americans [31,34]. The improvement in diabetes distress may be attributed to the multifaceted components of the CARE intervention, which integrates diabetes-related knowledge, behavioral techniques, and biweekly phone calls from community health workers to address SDOH concerns. These elements may collectively enhance participants' self-efficacy in managing diabetes, subsequently alleviating their disease-related distress and other concerns throughout their diabetes journey.

We did not observe significant improvements in other aspects of self-care activities and beliefs in diabetes self-care activities. Several factors may have contributed to these null results. First, the small sample size likely resulted in insufficient statistical power to detect between-group differences in these outcomes. Secondly, the relatively short duration of our intervention (12 weeks) compared to some other DSMES programs (24 weeks) may not have been enough time to produce observable changes in these behavioral aspects of diabetes self-management. A larger-scale RCT with sufficient power is necessary to examine the effectiveness of the intervention across a broader range of diabetes-related outcomes. These findings will also inform the sample size for a future, fully powered RCT to assess CARE's effectiveness and its potential to generalize these promising results to a broader population.

Several limitations should be noted. First, the sample size for this study was powered for HbA_{1c} levels; therefore, the self-efficacy and self-care activities reported in the current study may not have sufficient statistical power. The small sample size may limit the ability to detect significant differences in some outcomes. However, the primary goal of this pilot RCT was to assess the potential efficacy of the intervention and to generate pilot data for a future large-scale RCT. Second, all outcomes were self-reported and collected by community health workers over the phone and thus are subject to potential recall bias and social desirability bias. Additionally, our findings may not be generalizable to all Chinese Americans, particularly those with different sociodemographic characteristics. For instance, the majority of participants in our study had limited English proficiency, so the results may not apply to more acculturated counterparts who are fluent in English. Despite these limitations, this study is important as it serves as the first step in exploring scalable strategies to disseminate evidence-based diabetes interventions for underserved and minoritized populations. It also has the potential to serve as a care model for delivering culturally and linguistically tailored interventions to other immigrant or minoritized communities facing similar challenges in managing T2D and to lay the groundwork for a larger RCT.

Conclusions

This study provides promising evidence that culturally and linguistically tailored interventions may improve self-efficacy, dietary adherence, and diabetes distress among Chinese Americans. The use of accessible platforms like social media for health education shows promise in supporting diabetes management. Future research with adequate statistical power

XSL•FO
RenderX

is needed to establish the effectiveness of our intervention. By addressing the cultural and linguistic needs of immigrant populations, tailored interventions have the potential to reduce health disparities and improve diabetes-related health outcomes for immigrant or minoritized communities with T2D.

Acknowledgments

We thank our research participants for their time and participation. LH received support from the following organizations: National Institutes of Health (NIH; grants R01MD017579-01A1, K99MD012811, R00MD012811, P50MD017356 pilot award, U54MD000538-15 pilot award, and P30DK111022 pilot award), Agency for Healthcare Research and Quality (AHRQ; grant R18HS029813), and the American Diabetes Association (ADA; grant 7-22-ICTSN-08). The article processing charge was funded by LH's discretionary fund. The Socio-Spatial Determinants of Health Laboratory is supported by the Division of Intramural Research at the National Institute on Minority Health and Health Disparities (NIMHD) of the National Institutes of Health (NIH) (grant ZIA MD000020).

The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript. This research was supported, in part, by the Intramural Research Program of the NIH. The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the US Department of Health and Human Services.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JC, MAS, NI, NL, and LH
Methodology: MAS, NI, HL, KT, NL, and LH
Formal analysis: CW, YZ, and HL
Data curation: ZZ, YW, JH, and YS
Writing—original draft: JC and JL
Writing—review and editing: JC, YS, JL, MAS, NI, NF, HL, CW, YZ, KT, NL, NJ, ZZ, YW, JH, and LH
Funding acquisition: LH

Conflicts of Interest

LH reports holding stock in Tencent Holdings Limited. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. All of the other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Screenshots of the Chinese American Research and Education (CARE) intervention.
[PNG File , 4943 KB - mhealth_v13i1e78036_app1.png]

Multimedia Appendix 2
CONSORT 2010 checklist.
[PDF File (Adobe PDF File), 163 KB - mhealth_v13i1e78036_app2.pdf]

References

1. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. JAMA 2015 Sep 08;314(10):1021-1029. [doi: 10.1001/jama.2015.10029] [Medline: 26348752]
2. Rajpathak SN, Wylie-Rosett J. High prevalence of diabetes and impaired fasting glucose among Chinese immigrants in New York City. J Immigr Minor Health 2011 Feb;13(1):181-183 [FREE Full text] [doi: 10.1007/s10903-010-9356-2] [Medline: 20533090]
3. Yeh M, Fahs M, Shelley D, Yerneni R, Parikh NS, Burton D. Body weight and length of residence in the US among Chinese Americans. J Immigr Minor Health 2009 Oct;11(5):422-427 [FREE Full text] [doi: 10.1007/s10903-007-9113-3] [Medline: 18085438]

XSL•FO
RenderX

4. Lee JWR, Brancati FL, Yeh H. Trends in the prevalence of type 2 diabetes in Asians versus whites: results from the United States National Health Interview Survey, 1997-2008. Diabetes Care 2011 Feb;34(2):353-357 [FREE Full text] [doi: 10.2337/dc10-0746] [Medline: 21216863]
5. Ahmad F, Joshi SH. Self-care practices and their role in the control of diabetes: a narrative review. Cureus 2023 Jul;15(7):e41409 [FREE Full text] [doi: 10.7759/cureus.41409] [Medline: 37546053]
6. D'Souza MS, Karkada SN, Parahoo K, Venkatesaperumal R, Achora S, Cayaban ARR. Self-efficacy and self-care behaviours among adults with type 2 diabetes. Appl Nurs Res 2017 Aug;36:25-32. [doi: 10.1016/j.apnr.2017.05.004] [Medline: 28720235]
7. Bandura A, Freeman WH, Lightsey R. Self-efficacy: the exercise of control. J Cogn Psychother 1999 Jan 01;13(2):158-166. [doi: 10.1891/0889-8391.13.2.158]
8. Lin P, Lee T, Liu C, Lee Y. The effect of self-efficacy in self-management on diabetes distress in young reople with type 2 diabetes. Healthcare (Basel) 2021 Dec 15;9(12):1736 [FREE Full text] [doi: 10.3390/healthcare9121736] [Medline: 34946462]
9. Amer FA, Mohamed MS, Elbur AI, Abdelaziz SI, Elrayah ZA. Influence of self-efficacy management on adherence to self-care activities and treatment outcome among diabetes mellitus type 2 Sudanese patients. Pharm Pract (Granada) 2018 Dec 31;16(4):1274. [doi: 10.18549/pharmpract.2018.04.1274]
10. Kim G, Shim R, Ford KL, Baker TA. The relation between diabetes self-efficacy and psychological distress among older adults: do racial and ethnic differences exist? J Aging Health 2015 Mar;27(2):320-333 [FREE Full text] [doi: 10.1177/0898264314549662] [Medline: 25231883]
11. Littman A, Knott C, Boyko E, Hawes S. Associations between racial and ethnic groups and foot self-inspection in people with diabetes. Diabetes Care 2020 May;43(5):956-963 [FREE Full text] [doi: 10.2337/dc19-1486] [Medline: 32132006]
12. Kim G, Ford KL, Chiriboga DA, Sorkin DH. Racial and ethnic disparities in healthcare use, delayed care, and management of diabetes mellitus in older adults in California. J Am Geriatr Soc 2012 Dec;60(12):2319-2325. [doi: 10.1111/jgs.12003] [Medline: 23194086]
13. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. Clin Diabetes 2016 Apr;34(2):70-80 [FREE Full text] [doi: 10.2337/diaclin.34.2.70] [Medline: 27092016]
14. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016 Jun;99(6):926-943 [FREE Full text] [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]
15. Li-Geng T, Kilham J, McLeod KM. Cultural influences on dietary self-management of type 2 diabetes in East Asian Americans: a mixed-methods systematic review. Health Equity 2020;4(1):31-42 [FREE Full text] [doi: 10.1089/heq.2019.0087] [Medline: 32195450]
16. Chesla CA, Chun KM, Kwan CM. Cultural and family challenges to managing type 2 diabetes in immigrant Chinese Americans. Diabetes Care 2009 Oct;32(10):1812-1816 [FREE Full text] [doi: 10.2337/dc09-0278] [Medline: 19628812]
17. Ho EY, Chesla CA, Chun KM. Health communication with Chinese Americans about type 2 diabetes. Diabetes Educ 2012;38(1):67-76. [doi: 10.1177/0145721711428774] [Medline: 22127677]
18. Liang Z, Zhou B. The effects of legal status on employment and health outcomes among low-skilled Chinese immigrants in New York City. Ann Am Acad Pol Soc Sci 2016 Jun 14;666(1):150-163 [FREE Full text] [doi: 10.1177/0002716216650632]
19. Zhou M, Nordquist R. Work and its place in the lives of immigrant women: garment workers in New York City's Chinatown. Appl Behav Sci Rev 1994 Jan;2(2):187-211. [doi: 10.1016/1068-8595(94)90013-2]
20. Hu L, Islam N, Zhang Y, Shi Y, Li H, Wang C, et al. Leveraging social media to increase access to an evidence-based diabetes intervention among low-income Chinese immigrants: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2022 Oct 28;11(10):e42554 [FREE Full text] [doi: 10.2196/42554] [Medline: 36306161]
21. Domenech Rodríguez MM, Baumann AA, Schwartz AL. Cultural adaptation of an evidence based intervention: from theory to practice in a Latino/a community context. Am J Community Psychol 2011 Mar;47(1-2):170-186. [doi: 10.1007/s10464-010-9371-4] [Medline: 21116707]
22. Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: issues for the cultural adaptation and development of psychosocial treatments with Hispanics. J Abnorm Child Psychol 1995 Feb;23(1):67-82. [doi: 10.1007/bf01447045]
23. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
24. Shi Y, Sevick MA, Tang H, Wang C, Zhao Y, Yoon S, et al. Culturally tailored social media intervention improves health outcomes in Chinese Americans with type 2 diabetes: preliminary evidence from a pilot RCT. J Gen Intern Med 2025 Jul;40(10):2319-2329. [doi: 10.1007/s11606-025-09430-9] [Medline: 40016380]
25. Merkow RP, Kaji AH, Itani KMF. The CONSORT framework. JAMA Surg 2021 Sep 01;156(9):877-878 [FREE Full text] [doi: 10.1001/jamasurg.2021.0549] [Medline: 33825818]

XSL•FO
RenderX

26. Lorig K, Ritter PL, Villa FJ, Armas J. Community-based peer-led diabetes self-management: a randomized trial. Diabetes Educ 2009;35(4):641-651. [doi: 10.1177/0145721709335006] [Medline: 19407333]
27. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
28. Xu Y, Savage C, Toobert D, Pan W, Whitmer K. Adaptation and testing of instruments to measure diabetes self-management in people with type 2 diabetes in mainland China. J Transcult Nurs 2008 Jul;19(3):234-242. [doi: 10.1177/1043659608319239] [Medline: 18579863]
29. Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J, et al. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. Diabetes Care 2005 Mar;28(3):626-631. [doi: 10.2337/diacare.28.3.626] [Medline: 15735199]
30. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
31. Chesla CA, Chun KM, Kwan CM, Mullan JT, Kwong Y, Hsu L, et al. Testing the efficacy of culturally adapted coping skills training for Chinese American immigrants with type 2 diabetes using community-based participatory research. Res Nurs Health 2013 Aug;36(4):359-372. [doi: 10.1002/nur.21543] [Medline: 23606271]
32. Ivey S, Tseng W, Kurtovich E, Lui B, Weir RC, Liu J, et al. Evaluating a culturally competent health coach intervention for Chinese American patients with diabetes. Diabetes Spectr 2012 May 01;25(2):93 [FREE Full text] [doi: 10.2337/diaspect.25.2.93] [Medline: 24311938]
33. Wang C, Chan SMA. Culturally tailored diabetes education program for Chinese Americans: a pilot study. Nurs Res 2005;54(5):347-353. [doi: 10.1097/00006199-200509000-00009] [Medline: 16224321]
34. Sun AC, Tsoh JY, Saw A, Chan JL, Cheng JW. Effectiveness of a culturally tailored diabetes self-management program for Chinese Americans. Diabetes Educ 2012;38(5):685-694 [FREE Full text] [doi: 10.1177/0145721712450922] [Medline: 22722610]
35. Banks J, Amspoker AB, Vaughan EM, Woodard L, Naik AD. Ascertainment of minimal clinically important differences in the diabetes distress scale-17: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2023 Nov 01;6(11):e2342950 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.42950] [Medline: 37966840]
36. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. J Clin Epidemiol 2017 Feb;82:128-136 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.11.016] [Medline: 27986610]
37. Schober P, Bossers S, Schwarte L. Statistical significance versus clinical importance of observed effect sizes: what do P values and confidence intervals really represent? Anesth Analg 2018;126(3):1068-1072. [doi: 10.1213/ane.0000000000002798]
38. Aminuddin HB, Jiao N, Jiang Y, Hong J, Wang W. Effectiveness of smartphone-based self-management interventions on self-efficacy, self-care activities, health-related quality of life and clinical outcomes in patients with type 2 diabetes: a systematic review and meta-analysis. Int J Nurs Stud 2021 Apr;116:103286. [doi: 10.1016/j.ijnurstu.2019.02.003] [Medline: 30827741]
39. Zhuang P, Mao L, Wu F, Wang J, Jiao J, Zhang Y. Cooking oil consumption is positively associated with risk of type 2 diabetes in a Chinese nationwide cohort study. J Nutr 2020 Jul 01;150(7):1799-1807 [FREE Full text] [doi: 10.1093/jn/nxaa103] [Medline: 32364230]
40. Beasley JM, Yi SS, Ahn J, Kwon SC, Wylie-Rosett J. Dietary patterns in Chinese Americans are associated with cardiovascular disease risk factors, the Chinese American Cardiovascular Health Assessment (CHA CHA). J Immigr Minor Health 2019 Oct;21(5):1061-1069 [FREE Full text] [doi: 10.1007/s10903-018-0800-z] [Medline: 30097786]

Abbreviations

CARE: Chinese American Research and Education
CONSORT: Consolidated Standards of Reporting Trials
DSMES: Diabetes self-management education and support
HbA_{1c}: hemoglobin A_{1c}
MCID: minimal clinically important difference
RCT: randomized controlled trial
SDOH: social determinants of health

XSL•FO
RenderX

Edited by L Buis; submitted 25.05.25; peer-reviewed by L Hartling; comments to author 20.06.25; revised version received 08.08.25; accepted 09.09.25; published 27.10.25.

Please cite as:
Liu J, Cao J, Shi Y, Sevick MA, Islam N, Feldman N, Li H, Wang C, Zhao Y, Tamura K, Levy N, Jiang N, Zhu Z, Wang Y, Hong J, Hu L
A Culturally and Linguistically Tailored Intervention to Improve Diabetes-Related Outcomes in Chinese Americans With Type 2 Diabetes: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e78036
URL: https://mhealth.jmir.org/2025/1/e78036
doi:10.2196/78036
PMID:

©Jing Liu, Jiepin Cao, Yun Shi, Mary Ann Sevick, Nadia Islam, Naumi Feldman, Huilin Li, Chan Wang, Yanan Zhao, Kosuke Tamura, Natalie Levy, Nan Jiang, Ziqiang Zhu, Yulin Wang, Jia Hong, Lu Hu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of a Transtheoretical Model–Based mHealth Intervention on Transition Readiness in Adolescents With Epilepsy: Quasi-Experimental Study

Qing Xia[1], MM; Shuangzi Li[1], MM; Ting Wang[1], BMed; Mingping Fan[1], BMed; Jie Xia[1], BMed; Lingling Xie[1], MD; Huaying Yin[2], BMed

[1]Department of Neurology, Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Child Neurodevelopment and Cognitive Disorders, Chongqing, China

[2]Department of Child Health Care, Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Child Neurodevelopment and Cognitive Disorders, Chongqing, China

Corresponding Author:
Huaying Yin, BMed
Department of Child Health Care
Children's Hospital of Chongqing Medical University
National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Child Neurodevelopment and Cognitive Disorders
136 Zhongshan Second Road, Yuzhong District
Chongqing, 400010
China
Phone: 86 13983686874
Fax: 86 13983686874
Email: sarah6524@126.com

Abstract

Background: Enhancing self-management and transition readiness in adolescents with epilepsy is essential for successful transition to adult care. The combination of the transtheoretical model (TTM) and mobile health (mHealth) management provides a framework for reducing intervention costs while personalizing care.

Objective: This quasi-experimental study evaluates the feasibility of TTM-based mHealth management for improving transition services in adolescents with epilepsy.

Methods: A total of 98 adolescent patients with epilepsy aged 12-18 years were recruited. Using a nonrandomized design based on treatment locations, they were allocated into either the intervention group (n=49) or the control group (n=49). The intervention group received a TTM-based mHealth management program, which included phase-specific group sessions led by a multidisciplinary team and conducted via Tencent Meeting every 2 weeks or monthly (biweekly for the precontemplation, contemplation, and preparation, and monthly for the action and maintenance). The sessions involved lectures, discussions, and a mini-program that provided disease management support, motivational strategies, and digital reminders tailored to each stage. The control group received conventional remote extended care, consisting of biweekly group lectures and discussions for all patients and their families via Tencent Meeting, supplemented by regular health education materials delivered through a WeChat group. Telephone follow-ups were conducted at the third and sixth months. The total intervention duration was 6 months for both groups. Outcomes were assessed after 6 months using the self-management stage, Self-Management and Transition to Adulthood with Rx=Treatment questionnaire, and a self-developed program acceptability questionnaire.

Results: Postintervention, the intervention group demonstrated significantly better self-management behavior stages compared with controls. At the end of 6 months of intervention, the majority of participants in the intervention group reached the action stage (16/49, 32.65%) and maintenance stage (14/49, 28.57%), whereas most controls remained in precontemplation (12/49, 24.49%) and contemplation stages (13/49, 26.53%). Both groups showed significant improvements from baseline in medication management, health care participation, disease knowledge, doctor-patient communication, and transition readiness total scores at 6-month follow-up (all $P<.05$). Notably, the intervention group achieved additional incremental benefits versus controls (medication management: 3.81, 95% CI 1.26-6.36; health care engagement: 2.77, 95% CI 0.52-5.02; disease knowledge: 1.30,

XSL•FO
RenderX

95% CI 0.28-2.31; provider communication: 3.42, 95% CI 1.62-5.22; transition readiness: 11.30, 95% CI 5.70-6.89; effect sizes [Cohen *d*] ranged from 0.527 to 0.864, indicating moderate-to-large clinical effects). The overall satisfaction scores were 4.43 (SD 0.50) for patients and 4.16 (SD 0.82) for health care providers.

Conclusions: The TTM-based mHealth management program may effectively improve self-management behavior changes and enhance readiness for transition among adolescents with epilepsy, thereby facilitating a smooth transition to adult health care. The program demonstrated high acceptability, providing a reference for establishing clinical transition service protocols. However, this study was a single-center, quasi-experimental trial with a small sample size and short intervention duration. The findings need to be confirmed by larger-scale randomized controlled trials to verify efficacy.

(JMIR Mhealth Uhealth 2025;13:e70085) doi:10.2196/70085

KEYWORDS

epilepsy; adolescence; transitional services; trans-theoretical model; mobile health

Introduction

Epilepsy is a common chronic childhood disorder, and approximately 50% of children with epilepsy have seizures that continue into adulthood [1]. This subset of children with epilepsy will need to be transitioned to adult medical care, and experts recommend that the medical transition should begin in early adolescence (11-15 years of age) and extend through to adulthood [2]. If the health care transition for adolescents to adults with epilepsy does not go well, it may lead to treatment interruption–induced seizures, exacerbation [3], emotional problems [4], and social problems such as lack of education and social isolation [5]. Transition readiness is one of the important elements influencing the successful health care transition of chronically ill adolescents [6], and is the ability of chronically ill adolescents to begin to complete the transition process with their families and health care support systems [7]. Early preparation for transition, communicating transition information to chronically ill adolescents, being patient-centered, and assisting patients in acquiring self-management knowledge and skills are important principles to follow when providing transition services [2,8]. Ideal transition service plans must encourage patients to take control of their lives [5]. Therefore, developing self-management behaviors and improving transition readiness in adolescents with epilepsy are critical steps in successful medical transition.

Currently, there are various intervention measures for self-management of patients with epilepsy, and the outcome indicators are highly heterogeneous [9]. Among common intervention methods, group-based interventions have broader coverage and can reduce costs, but they often fail to adequately account for individual differences. In contrast, personalized interventions focus on addressing individual needs but require greater human and time resources, which face constraints due to China's current limited medical resources [10]. Therefore, how to reduce intervention costs while addressing personalized needs remains an unresolved challenge.

The transtheoretical model (TTM) is a staged model for promoting behavioral change, and the TTM emphasizes the use of stage-specific behavior change strategies tailored to individuals at different stages, while prioritizing the development of self-efficacy to facilitate active engagement in behavioral modification [11], and previous studies have demonstrated the model's positive significance in the transformation of personal health behaviors in patients with diabetes, chronic kidney disease, and cardiovascular disease [12-14], that provides a theoretical basis for personalized collective intervention from the behavioral stage. In addition, the mobile health (mHealth) management model has been widely noticed and explored in recent years [15,16], with notable applications in self-management for patients with epilepsy and transition care for adolescents with epilepsy transitioning to adulthood [17-19], demonstrating promising potential in epilepsy health management. mHealth's quick and convenient features can reduce labor costs and time spent on medical care, providing a new perspective for adolescents with epilepsy to achieve transitional services. This study aims to develop an intervention program based on the TTM and the mHealth model, followed by empirical validation. It is anticipated that this program will improve self-management behaviors and transition readiness among adolescents with epilepsy, while also gaining acceptance from both patients and health care providers. The details are reported as follows.

Methods

Study Design

This quasi-experimental study with a 2-arm cluster allocation design was conducted and reported in accordance with the Transparent Reporting of Evaluations with Nonrandomized Designs statement [20].

Ethical Considerations

This study was approved by the Ethics Committee of Children's Hospital of Chongqing Medical University (approval number: 2021, Ethical Review [Research] number 146). All procedures involving human participants were conducted in accordance with the ethical standards of the institutional and national research committees and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. As the study involved adolescents with epilepsy and the collection of their medical and questionnaire data, a full ethical review was mandatory and obtained prior to the commencement of the study. Written informed consent was obtained from all parents or legal guardians of the participating adolescents. Additionally, written assent was provided by all adolescent participants themselves. They were thoroughly informed about the study's purpose, procedures, potential risks and benefits, and their right to withdraw from the study at any time without any detriment

XSL•FO
RenderX

to their medical care. For the secondary analysis of data generated within the study's own mini-program, the original informed consent covered the use of this data for research purposes, as confirmed by the ethics committee, thereby waiving the need for additional consent for analysis. All participant data were anonymized and deidentified prior to analysis. Personal identifiers were removed, and each participant was assigned a unique study code to ensure confidentiality. Data were stored on encrypted, password-protected servers with access restricted to the principal investigator and authorized research personnel only. The mini-program development company signed a legally binding patient information confidentiality agreement to ensure the protection of all participant data. No financial compensation was provided to participants. However, all participants in both groups received the standard care and additional remote support (either the TTM-based program or conventional extended care) at no cost throughout the study period. This included access to educational materials, remote consultations, and continuous support from the health care team.

Objects of Study

Adolescents with epilepsy (12-18 years old) who visited the outpatient department of a tertiary children's specialty hospital in Chongqing between August 2021 and February 2022 were screened for eligibility. Inclusion criteria were diagnosis of nonrefractory epilepsy for 6 months or more and seizure-free for 30 days; epilepsy diagnostic criteria were the 2014 International League Against Epilepsy's definition of epilepsy [21], and categorization criteria were the revised version of the 2017 International League Against Epilepsy [22]; exclusion criteria were (1) impaired use of smartphones, (2) communication disorders, and (3) comorbidities with other systemic chronic or malignant diseases or psychiatric disorders. Written informed consent and parental assent were obtained.

Sample Size Calculation

An a priori sample size calculation was performed using G*Power software (version 3.1.9.7; Heinrich-Heine-University Düsseldorf). The calculation was based on the primary outcome of transition readiness, as measured by the total score of the Chinese version of the Self-Management and Transition to Adulthood with Rx=Treatment (STARx) questionnaire [23].

The sample size was estimated using the formula ( ). The parameters for the calculation were specified as follows: a 2-tailed significance level (α) of 0.05, a statistical power (1-β) of 0.90, and an allocation ratio (intervention: control) of 1:1 (ie, k=1). The effect size (Cohen *d*) was set at 0.714. This effect size was derived from a pilot study we conducted prior to the main trial, which included 15 adolescents with epilepsy in each group (30 total). The pilot results indicated a mean difference of 5.0 points in the change of STARx total scores between the intervention and control groups, with a pooled SD of 7.0 points (d=5.0/7.0≈0.714), which represents a moderate-to-large effect. Based on these parameters, the power analysis indicated that a minimum of 42 participants per group was required. To account for a potential dropout rate of 10% during the 6-month intervention period, the target sample size was increased to 49 participants per group, resulting in a total required sample of 98 participants.

Clinical Data Collection

The clinical data collected encompassed the following categories.

Baseline Demographic and Clinical Characteristics

These included age, sex, duration of epilepsy, current antiepileptic medications, seizure frequency in the past 6 months, and any comorbid conditions. This information was primarily collected on-site during the initial enrollment interview using a standardized case report form.

Primary and Secondary Outcome Measures

This referred to the data from the validated questionnaires, which were collected using Questionnaire Star at both baseline (T0) and the 6-month follow-up (T1). The specific measures included the following: the primary outcome was the total score and domain scores (medication management, health-care engagement, disease knowledge, and provider communication) of the Chinese version of the STARx questionnaire [23], while the secondary outcomes were the self-management behavior stage assessment form and the self-developed program acceptability questionnaire (for the intervention group only at T1).

Process Data (For the Intervention Group)

Data generated within the mini-program, such as medication adherence records (from reminder logs), self-recorded seizure frequency and mood diaries, and engagement metrics (eg, frequency of accessing educational materials), were automatically collected by the platform. A combination of on-site (for baseline characteristics) and Questionnaire Star (for outcome measures) was used for data collection. All data were organized using Microsoft Excel software and were independently checked by 2 researchers to ensure accuracy.

Evaluation Indicators

Stages of Self-Management Behavior

This study incorporates terminological modifications to the definitions of the behavioral stages in conjunction with the TTM to ensure that the meanings of the scale questions are clear. Self-management behaviors in adolescents with epilepsy are listed according to the comprehensive definition of "self-management" in children with epilepsy by Wagner et al [9]. Finally, the assessment form was finalized after expert consultation. That is, "Based on the self-management performance listed above, have you started to self-manage?" patients were asked to select the most appropriate answer, including the preintentional ("never thought about self-management or thought about it 6 months later"), the intentional ("didn't start to self-manage, but I started to think about it and am going to start in 6 months"), the preparatory ("didn't start to self-manage, but I'm going to start in 30 days"), the action ("I've already started to self-manage, but in less than 6 months"), and the maintenance ("I have been self-managing for more than 6 months"). In order from the preintentional, intentional, preparation, action to the maintenance, the closer the individual is to the maintenance of behavior, the more conducive to successful behavior change. This self-management behavior stage assessment form is an investigator-developed

XSL•FO
RenderX

instrument, and its design strictly followed scientific standards: (1) taking the TTM theory as the core framework, and selecting assessment dimensions based on the definition of self-management in children with epilepsy proposed by Wagner et al [9]; (2) inviting a panel of 3 chief physicians of child neurology, 2 pediatric nursing experts, and 2 psychology professors to conduct content validity testing on the assessment items, with the expert content validity index reaching 0.92; (3) a pilot test was conducted on 20 adolescents meeting the inclusion criteria before the formal study, and the results showed that the Kappa coefficient of the assessment form was 0.83 (P<.001), indicating good interrater reliability.

Transition Preparation

The STARx questionnaire was used, which was originally developed by American scholars Ferris et al [23] to measure chronic disease. Adolescent self-management and readiness for health care transition were validated among adolescents aged 12 to 25 years with chronic illnesses and showed adequate internal reliability, with a Cronbach α coefficient of 0.80 for the questionnaire. The research team from the School of Nursing at Shanghai Jiao Tong University translated and culturally adapted the questionnaire. The Chinese version of the STARx questionnaire demonstrated good internal consistency, with a Cronbach α coefficient of 0.812. The instrument comprises 4 domains as follows: medication management, health care engagement, disease knowledge, and provider communication, totaling 18 items [24]. For scoring, medication-related items (with the option “I am not currently taking medications” scored as 6 points) range from 1 to 6 points. All other items are scored from 1 to 5 points. Higher total scores indicate better transition readiness levels.

Acceptance of the Intervention Program in the Experimental Group

Participating study health care professionals and adolescents with epilepsy in the experimental group were given a questionnaire at the end of the intervention to assess the acceptability of the program, which was a self-administered questionnaire with appropriate questions and scores on the operability, usefulness, convenience, affordability, and clinical feasibility of the program. The questionnaire used a Likert 5-point scale with “5” being strongly agree and “1” being strongly disagree. This intervention program acceptability questionnaire is an investigator-developed instrument, and its development process was as follows: (1) core dimensions affecting intervention acceptability (operability, usefulness, and convenience) were identified through literature review; (2) initial items were designed with reference to published medical intervention acceptability assessment tools (eg, the mHealth intervention satisfaction questionnaire developed by Yamamoto et al [25]), resulting in 15 candidate items; (3) 3 ambiguous items were deleted after review by the aforementioned expert panel, and 12 formal items were finally determined; (4) pilot test results showed that the overall Cronbach α coefficient of the questionnaire was 0.84, and the Cronbach α coefficients of each dimension ranged from 0.78 to 0.86, indicating good internal consistency reliability.

Intervention Methods

Experimental Group Intervention Program

Based on the literature review, the research team developed a draft intervention protocol, which was finalized through expert consultation and pilot testing.

Self-Management Applet Design and Development

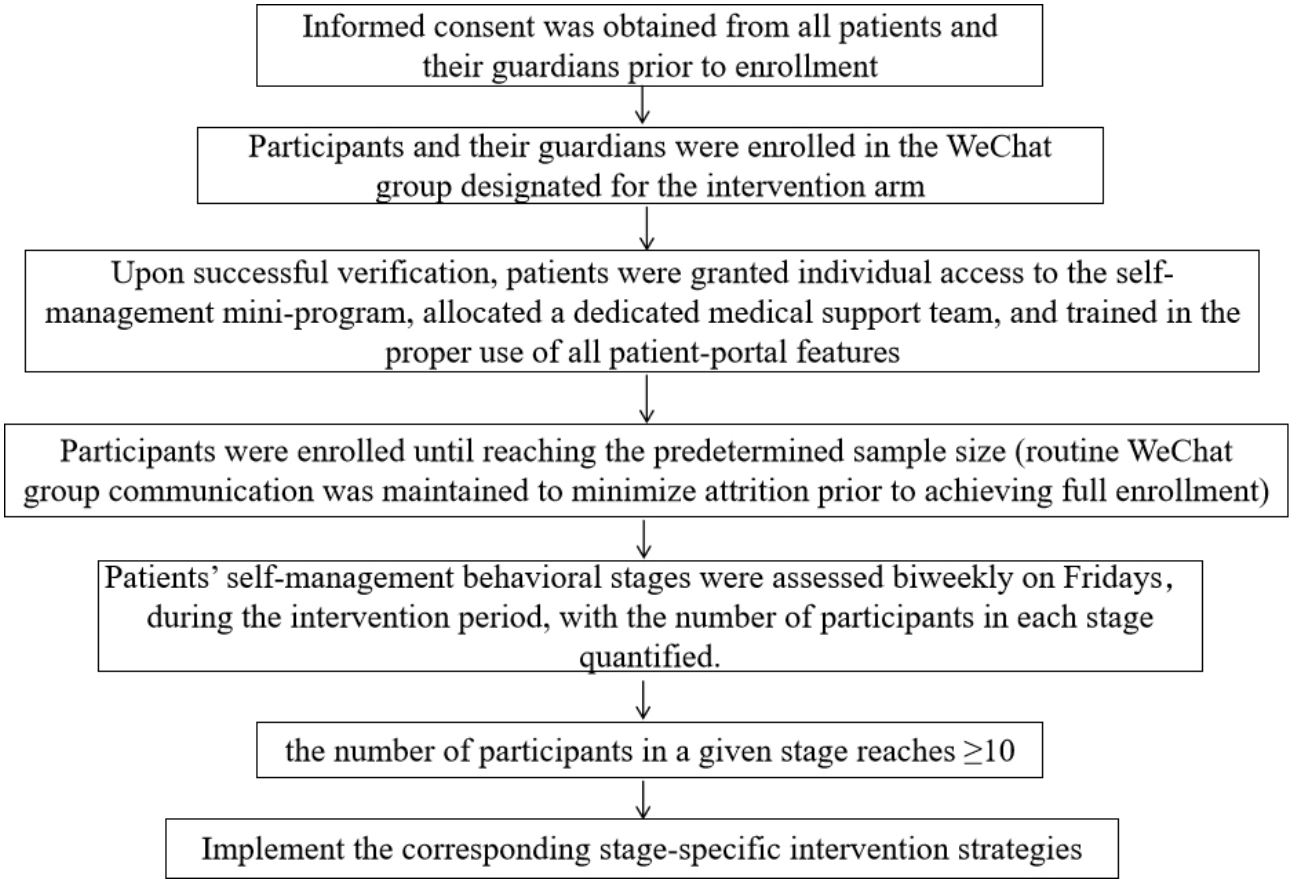
The design team includes 2 medical and nursing specialists, each in child neurology, and one counselor II. The R&D team is a Chengdu Technology Limited Liability company. The applet is used to help adolescents with epilepsy recognize the disease and manage it, and it is set up with a patient side and a health care side. Functions on the patient’s side include learning about epilepsy and self-management, medication reminders, epilepsy logs (based on the pediatric version recommended by the Chinese Association Against Epilepsy, including mood, medication, and seizure records), and data export, as well as patient-physician communication and interaction. Functions on the provider’s side include managing patients’ information, disseminating health education knowledge, viewing patients’ epilepsy logs, and patient-physician communication and interaction. Prior to formal use, adolescents with epilepsy were selected to try out the program until their ease of use, utility, and satisfaction with the applet were all 80% or above. Before the official use of the mini-program, a patient information confidentiality agreement was signed with the mini-program development company.

Intervention Implementation Process

The intervention lasted for 6 months (24 weeks) and was conducted through a phased, group-based online intervention combining a WeChat mini-program, WeChat, and Tencent Meeting. The implementation process involved 3 main stages. (1) Explain the research purpose and protocol to patients and their guardians. After obtaining written informed consent, patients and parents join the WeChat group for the experimental group. (2) Patients individually access the mini-program to complete their personal information, and then patients are assigned a medical team consisting of “one doctor, one nurse, and one psychological counselor.” This assigned medical team was subsequently responsible for delivering the phase-specific group interventions to their assigned patients. The team explains the mini-program’s functions and use, including how to access health education materials and log medication details, seizure records, and mood diaries in the mini-program. Patients can leave messages in the communication area for special circumstances. (3) Every 2 weeks (on Friday), each patient’s behavioral stage is assessed. When the number of patients in a particular stage reaches ≥10, the corresponding stage-based intervention strategy is implemented, with both patients and parents participating. Since individuals in the action and maintenance demonstrate higher levels of information, motivation, and behavioral skills [11], the intervention frequency is reduced from biweekly (for other stages) to monthly for these 2 stages. The implementation process is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Intervention implementation flowchart.



Implementation Strategy

The mHealth intervention program, based on the TTM, was delivered in stages through group sessions held every 2 weeks or monthly (biweekly for the precontemplation, contemplation, and preparation; monthly for the action and maintenance). These sessions included structured lectures and discussions conducted via Tencent Meeting, combined with a mini-program that provided stage-appropriate disease management support, motivational strategies, and automated reminders. The group sessions were facilitated by a dedicated multidisciplinary transition team, which included a child neurologist, a pediatric epilepsy specialist nurse, and a psychological counselor. The neurologist and nurse were responsible for delivering the stage-specific educational lectures on topics such as epilepsy pathophysiology, medication adherence, and seizure first-aid. The psychological counselor led the motivational interviewing and discussion sessions, using techniques tailored to each TTM stage (eg, exploring ambivalence in the contemplation stage, strengthening commitment in the preparation stage). All facilitators received prior training on TTM principles and the intervention protocol to ensure consistency. Details are presented in Table 1.

XSL•FO
RenderX

Table 1. Forms of intervention and intervention strategies for different self-management behavioral stages in adolescents with epilepsy. The group lecture and interactive intervention strategy were delivered via Tencent Meeting.

Plateau and forms of intervention	Intervention strategies	Intervention frequency
Precontemplation		1 per fortnight
Group lectures	Lectures were given on the long-term management of epilepsy, so that patients and their families can recognize the treatment and prognosis of epilepsy, and the importance of preparing for the transition from adolescence to adulthood and self-management.	
Collective exchange	After the lecture, invite 2-3 patients to share their current concerns, and have the speaker address them.	
Guided learning	The facilitator will guide patients to access the mini-program's education section to read stage-appropriate articles: "Epilepsy Treatment & Prognosis," "Long-term Epilepsy Management," and "Understanding Healthcare Transition."	
Contemplation		1 per fortnight
Group lectures	The lectures are identical to the preconsciousness.	
Collective exchange	1. During the postlecture sharing session, a volunteer parent may discuss 3 caregiving aspects: receiving the diagnosis, managing uncontrolled symptoms, and coping with caregiver dependency. 2. The speaker will present a case example of a patient with excellent self-management from our outpatient follow-up cohort, to enhance participants' perception of treatment benefits.	
Guided learning	The facilitator will guide patients to the mini-program's education section to study the phase-appropriate article "Top 10 Q&A on Epilepsy," enabling them to acquire accurate health care knowledge and dispel disease-related misconceptions and unnecessary lifestyle restrictions.	
Preparation		1 per fortnight
Group lectures	Conducting lectures on daily safety management, medication management, and follow-up management of patients with epilepsy, so that patients can learn how to properly self-manage and how parents can collaborate, and learn about available medical resources.	
Collective exchange	The host medical staff guides each patient to share their current preparations for self-management in terms of material and information resources, and to learn from each other.	
Sign the pledge form	Distribute a self-management e-commitment form and ask each patient to sign it electronically.	
Prepare a planner	Patients are required to complete a self-management plan form. The session facilitator will collect the submissions and provide guidance for clearly inappropriate plans.	
Guided learning	The facilitator will guide patients to access the mini-program's education section to read stage-specific health articles: "Emergency Seizure Management at Home," "Medication Adherence Guidelines," "Optimal Follow-up Practices," and "Preparing for Adult Healthcare Transition."	
Action		1 per month
Group lectures	Lecture: Continuing Management of Chronic Diseases in the Information Age, so that patients and parents can master the management of diseases with the help of information technology, such as internet hospitals, medication management applets, and so on.	
Collective exchange	1. The facilitator will guide patients and parents to articulate challenges and barriers encountered during implementation, followed by group discussions to develop solutions, enabling mutual learning and experience sharing. 2. The facilitator will specifically emphasize reducing parental overinvolvement in patient affairs and encourage independent decision-making to enhance the patient's self-efficacy.	

XSL•FO
RenderX

Plateau and forms of intervention	Intervention strategies	Intervention frequency
Record in the epilepsy diary	All patients are required to complete an epilepsy diary in the mini-program (including seizure episodes, medication intake, and mood records) at least once weekly. Health care providers will review the submitted seizure logs, medication adherence, and mood entries every Friday, with timely interventions for any recording errors or concerning mood indicators.	
Guided learning	The facilitator will guide patients to access the mini-program's education section to read phase-specific articles: "When Should I Transition to Adult Healthcare?" and "How to Be Fully Prepared for Transition?"	
Enable reminders	The mini-program activated medication reminder alerts to reinforce patient adherence behaviors.	
Engage reward	A reward mechanism is established where patients earn stars for completing health education modules or epilepsy diary entries at preset frequencies. Accumulated stars can be redeemed for material rewards upon reaching specified thresholds.	
Maintenance		1 per month
Group lectures	Conducting lectures on the mental health and social adaptation of adolescents with epilepsy; in addition, collecting opinions from parents to conduct targeted lectures.	
Collective exchange	The presiding medical staff guided each patient to share the good methods and problems encountered in the operation, so that we could learn from each other.	
Enable reminders	Same action.	
Engage reward	Same action.	
Guided learning	The facilitator will guide patients to review the stage-appropriate educational article "Coping with Emotional Regulation Challenges" in the mini-program's learning section. Subsequent push notifications will be tailored based on real-time feedback from patients and parents regarding emerging concerns.	

Control Group Intervention Program

The control group used conventional continuing care management, instructed patients to record paper epilepsy logs, and established a WeChat group to facilitate communication. After reaching the predetermined sample size, collective lectures will be conducted via Tencent Meeting for all patients and their parents on Sundays of the 2nd, 4th, 6th, 8th, and 10th weeks postenrollment. The lectures will follow the content sequence of the intervention protocol for the experimental group, covering stages from precontemplation to maintenance. Each lecture will be followed by a discussion and questions and answers session. Additionally, one educational article per week will be uploaded to the WeChat group for patients and parents to study. These articles will align with the educational push notifications from the mini-program corresponding to the experimental group's stages (from precontemplation to maintenance). Telephone follow-ups will be conducted at the 3rd and 6th months.

Reporting Guidelines

The authors have completed the Transparent Reporting of Evaluations with Nonrandomized Designs checklist for reporting of nonrandomized trials. The completed checklist is available as Multimedia Appendix 1.

Statistical Methods

Statistical analysis was performed using SAS (version 9.4; SAS Institute Inc) software. Baseline balance between the experimental and control groups was assessed using standardized mean differences (SMDs), with an SMD <0.1 indicating good balance. Categorical variables were described using counts and percentages, and comparisons between groups were conducted using the chi-square test. The normality of continuous variables was evaluated using quantile-quantile plots. Normally distributed continuous variables were expressed as mean (SD), and between-group comparisons were performed using *t* tests. The between-group differences, 95% CIs, and Cohen *d* were calculated using analysis of covariance adjusted for baseline covariates to describe effect sizes. For ordinal categorical variables, between-group comparisons were conducted using the Mann-Whitney *U* test. For stage-transition ordinal data, between-group differences were analyzed using an ordinal logistic regression model adjusted for baseline covariates, with odds ratios (ORs) and 95% CIs reported to describe effect sizes. Within-group prepost changes were assessed using the Wilcoxon signed-rank test. A 2-sided *P* value <.05 was considered statistically significant.

Results

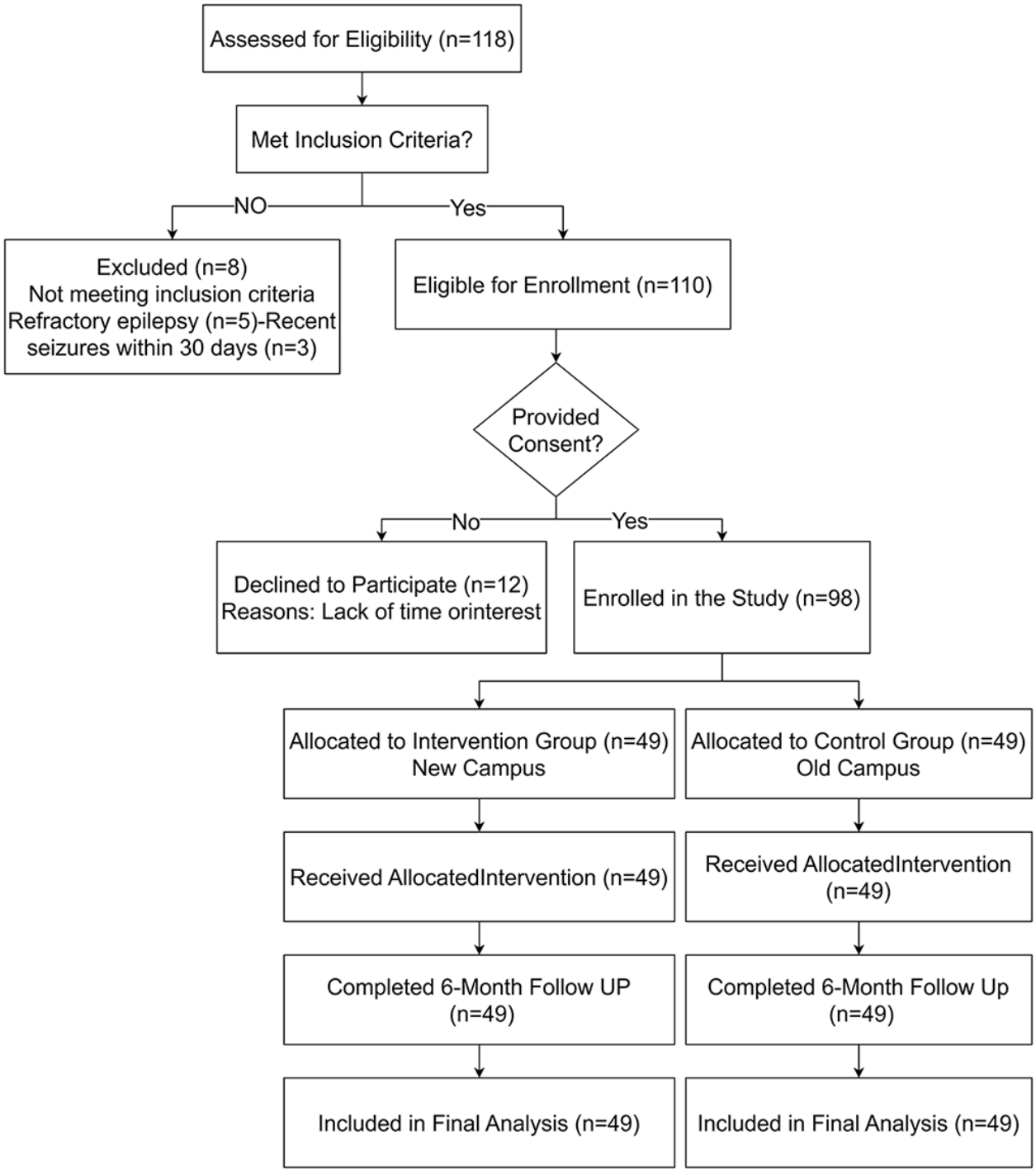
General Information

A total of 118 adolescents were assessed for eligibility during the study period. Of these, 20 were excluded: 8 did not meet the inclusion criteria (5 due to refractory epilepsy, 3 due to recent seizures within 30 days), and 12 declined to participate (citing reasons such as lack of time or interest). The remaining

XSL•FO
RenderX

98 eligible adolescents who provided written informed consent (and parental assent) were enrolled in the study. To avoid contamination between study groups, participants recruited from the new campus were assigned to the intervention group (n=49), while those from the old campus served as the control group (n=49). A flowchart detailing the participant screening, enrollment, and allocation process is provided in Figure 2. In this study, 52 cases were enrolled in the experimental group and 52 cases in the control group, among which 1 case in the experimental group withdrew from the study due to diagnosis of leukemia during the intervention period, and 2 cases voluntarily chose to withdraw from the study in the middle of the study; 2 cases in the control group refused to withdraw from the follow-up visit, and 1 case was lost to the follow-up visit; finally, 49 cases in each of the experimental group and the control group completed the study. The 2 groups showed baseline differences (SMD >0.1) in gender, only-child status, disease duration, caregiver type, caregiver education level, and residence (Table 2).

Figure 2. Flow diagram of participant enrollment, allocation, and analysis.



XSL•FO
RenderX

Table 2. Comparison of the general information of the 2 groups of patients.

Projects	Experimental group (n=49)	Control group (n=49)	Chi-square (*df*)	*P* value	Standardized mean difference
Sex, n (%)			1.475 (1)	.23	0.247
Male	20 (40.8)	26 (53.1)			
Female	29 (59.2)	23 (46.9)			
Age (years), n (%)			0.044 (1)	.83	0.043
12-15	32 (65.3)	31 (63.3)			
15.1-18	17 (34.7)	18 (36.7)			
An only child, n (%)			1.111 (1)	.29	0.214
Yes	20 (40.8)	15 (30.6)			
No	29 (59.2)	34 (69.4)			
Duration illness (year), n (%)			1.449 (3)	.65	0.160
≤1	2 (28.57)	5 (71.43)			
1-3	12 (50.00)	12 (50.00)			
3-5	21 (51.22)	20 (47.78)			
>5	14 (53.85)	12 (46.15)			
Caregiver, n (%)			1.576 (2)	.46	0.252
Oneself	6 (12.2)	5 (10.2)			
Grandparents	9 (18.4)	5 (10.2)			
Father/Mother	34 (69.4)	39 (79.6)			
Caregiver literacy, n (%)			0.379 (2)	.83	0.112
Junior high school and below	21 (42.9)	19 (38.8)			
High school or higher vocational	10 (20.4)	9 (18.4)			
College and above	18 (36.7)	21 (42.9)			
Current address, n (%)			0.378 (2)	.83	0.122
Township	12 (24.5)	11 (22.4)			
District and county	20 (40.8)	23 (46.9)			
Municipal and above	17 (34.7)	15 (30.6)			
Monthly household income (thousands of dollars), n (%)			2.895 (3)	.41	0.091
≤2	3 (6.1)	1 (2.0)			
2-4	15 (30.6)	17 (34.7)			
4-8	18 (36.7)	23 (46.9)			
>8	13 (26.5)	8 (16.3)			

Comparison of Self-Management Behavioral Stages Between the 2 Groups Before and After the Intervention

There was no statistical difference in the preintervention self-management behavioral stage between the 2 groups (*P*>.05). At the end of 6 months of intervention, the intervention group demonstrated significantly better self-management behavior stages compared with controls (OR 2.543, 95% CI 1.207-5.357, *P*=.01). At the end of 6 months of intervention, the majority of participants in the intervention group reached the action stage (16/49, 32.65%) and maintenance stage (14/49, 28.57%), whereas most controls remained in precontemplation (12/49, 24.49%) and contemplation stages (13/49, 26.53%), as shown in Table 3.

XSL•FO
RenderX

Table 3. Comparison of self-management behavioral stages between the 2 groups before and after intervention examples (N=49).

Time group	Precontemplation, n (%)	Contemplation, n (%)	Preparatory, n (%)	Action, n (%)	Maintenance, n (%)	Wald chi-square (*df*)[a]	*P* value[a]	Odds ratio (95%CI)[a]
Preintervention						0.176 (1)	.68	1.179 (0.546-2.546)
Control group	25 (51.02)	7 (14.29)	4 (8.16)	9 (18.37)	4 (8.16)			
Experimental group	22 (44.90)	9 (18.37)	3 (6.12)	10 (20.41)	5 (10.20)			
Postintervention						6.024 (1)	.01	2.543 (1.207-5.357)
Control group	12 (24.49)	13 (26.53)	8 (16.33)	9 (18.37)	7 (14.29)			
Experimental group	6 (12.24)	8 (16.33)	5 (10.20)	16 (32.65)	14 (28.57)			

[a]Ordinal logistic model was adjusted for gender, age, only-child status, disease duration, caregiver type, caregiver's education level, residence location, and monthly household income. The Wilcoxon signed-rank test revealed a statistically significant difference in the experimental group between baseline and the 6-month postintervention assessment (Z=5.443, *P*<.001). The control group showed statistically significant differences between baseline and the 6-month postintervention assessment (Z=5.000, *P*<.001).

Comparison of Transition Readiness Between the 2 Groups Before and After the Intervention

There was no statistical difference between the 2 groups in terms of total transition readiness scores and scores on the dimensions before the intervention (*P*>.05). At the end of 6 months of intervention, the intervention group demonstrated significantly higher scores than the control group in health care engagement (mean difference=2.28, 95% CI 0.57-3.99), disease knowledge (1.16, 95% CI 0.40-1.93), provider communication (2.08, 95% CI 0.77-3.40), transition readiness score (7.96, 95% CI 3.54-12.37), and effect sizes (Cohen *d*) ranged from 0.571 to 0.772. Both groups showed significant improvements from baseline in medication management, health care participation, disease knowledge, doctor-patient communication, and transition readiness total scores at 6-month follow-up (all *P*<.05). Notably, the intervention group achieved additional incremental benefits versus controls (medication management: 3.81, 95% CI 1.26-6.36; health care engagement: 2.77, 95% CI 0.52-5.02; disease knowledge: 1.30, 95% CI 0.28-2.31; provider communication: 3.42, 95% CI 1.62-5.22; transition readiness: 11.30, 95% CI 5.70-6.89; effect sizes [Cohen *d*] ranged from 0.527 to 0.864, indicating moderate-to-large clinical effects, as shown in Table 4).

XSL•FO
RenderX

Table 4. Comparison of transition readiness between the 2 groups of patients before and after the intervention.

Time and group	Medication management	Health care engagement	Diseases knowledge	Provider communication	Transition readiness score
Preintervention					
Control group (n=52), mean (SD)	20.59 (6.23)	11.33 (3.63)	5.78 (2.19)	15.29 (3.52)	52.98 (9.89)
Experimental group (n=52), mean (SD)	19.51 (5.47)	11.45 (5.02)	5.76 (1.94)	14.43 (4.40)	51.14 (13.79)
Between-group differences, 95% CI[a]	–1.38 (–3.86 to 1.10)	–0.49 (–2.39 to 1.40)	–0.13 (–1.02 to 0.75)	–1.33 (–3.05 to 0.39)	–3.34 (–8.67 to 2.00)
t[a]	–1.108	–0.518	–0.294	–1.542	–1.244
P value[a]	.271	.606	.769	.127	.217
Postintervention					
Control group (n=49), mean (SD)	22.80 (6.18)	13.33 (3.80)	7.35 (1.75)	16.02 (3.14)	59.49 (9.99)
Experimental group (n=49), mean (SD)	25.31 (5.32)	16.00 (4.12)	8.57 (1.80)	18.47 (3.08)	68.35 (10.51)
Between-group differences, 95% CI[a]	2.43 (–0.03 to 4.89)	2.28 (0.57 to 3.99)	1.16 (0.40 to 1.93)	2.08 (0.77 to 3.40)	7.96 (3.54 to 12.37)
t[a]	1.963	2.653	3.021	3.152	3.587
P value[a]	.053	.010	.003	.002	.001
Cohen *d*[a]	0.422	0.571	0.650	0.678	0.772
Postintervention changes					
Control group (n=49), mean (SD)	2.20 (4.99)[b]	2.00 (2.45)[b]	1.57 (2.25)[b]	0.73 (2.35)[b]	6.51 (7.89)[b]
Experimental group (n=49), mean (SD)	5.80 (6.28)[b]	4.55 (6.57)[b]	2.82 (2.55)[b]	4.04 (5.44)[b]	17.20 (15.78)[b]
Between-group differences, 95% CI[a]	3.81 (1.26-6.36)	2.77 (0.52-5.02)	1.30 (0.28-2.31)	3.42 (1.62-5.22)	11.30 (5.70-16.89)
t[a]	2.973	2.452	2.533	3.780	4.018
P value[a]	.004	.016	.013	<.001	<.001
Cohen *d*[a]	0.639	0.527	0.545	0.813	0.864

[a]Analysis of covariance (ANCOVA) was performed with adjustments for gender, age, only-child status, disease duration, caregiver type, caregiver's education level, residence, and monthly household income.

[b]The paired *t* test showed a statistically significant result ($P<.05$)

Evaluation of the Acceptability of the Intervention Program in the Experimental Group

Adolescents with epilepsy rated the program as low as "This approach solves some of my emotional problems" (3.45, SD 0.82 points), and medical staff rated it as low as "This approach solves personalized problems" (3.70, SD 0.48 points). The overall satisfaction scores were 4.43 (SD 0.50) for patients and 4.16 (SD 0.82) for health care providers, as shown in Table 5.

XSL•FO
RenderX

Table 5. Acceptance ratings of transtheoretical model–based mobile health (mHealth) management program.

Item title	Score, mean (SD)
Evaluation of the program by adolescents with epilepsy	
1. I like this remote group intervention approach	4.43 (0.79)
2. Appropriate frequency of interventions	3.96 (0.71)
3. Parental involvement is helpful to me	3.67 (0.69)
4. The words and actions of other patients in the discussion had an impact on me	3.94 (0.77)
5. This approach allows for normal communication	4.02 (0.69)
6. This way I can learn about diseases	4.43 (0.65)
7. This approach teaches me self-management	4.16 (0.59)
8. This approach can solve some of my emotional problems.	3.45 (0.82)
9. This approach saves time and energy	4.16 (0.72)
10. This approach saves spending on travelling to the site	4.63 (0.49)
11. I think it could be promoted to other young people with epilepsy who need it	4.18 (0.57)
12. General feeling of satisfaction	4.43 (0.50)
Evaluation of the program by medical staff	
1. This approach helps patients	4.60 (0.52)
2. This approach allows for normal communication with the patient.	4.10 (0.74)
3. This approach addresses personalization	3.70 (0.48)
4. This approach saves time and effort relative to individualized interventions	4.50 (0.53)
5. I think it can be used in clinical practice	4.10 (0.64)
6. General feeling of satisfaction	4.16 (0.82)

Discussion

The TTM-Based mHealth Intervention May Facilitate the Establishment of Self-Management Behaviors in Adolescents With Epilepsy

The survey found that patients with chronic diseases have different information needs at various stages of behavior change [26]. The greatest advantage of TTM is that it recognizes the differences between individuals at different stages of behavior. Stage-matched behavior change strategies were implemented by researchers, which enhanced participation rates and promoted patients' progression toward the action and maintenance stages [11]. At the end of 6 months of intervention, the intervention group demonstrated significantly better self-management behavior stages compared with controls (OR 2.543, 95% CI 1.207-5.357, *P*=.01). At the end of 6 months of intervention, the majority of participants in the intervention group reached the action stage (16/49, 32.65%) and maintenance stage (14/49, 28.57%), whereas most controls remained in precontemplation (12/49, 24.49%) and contemplation stages (13/49, 26.53%). Evidently, adolescent patients with epilepsy who received stage-matched behavioral interventions demonstrated favorable self-management behavior modifications, with the majority sustaining these changes as habitual practices. This study protocol used individualized strategies tailored to participants' specific stages of behavioral change, effectively addressing unique needs while enhancing intrinsic motivation for engagement. The technology-delivered remote intervention aligned well with adolescents' social habits, potentially improving participation rates and thereby better supporting behavioral transitions. These coordinated approaches successfully promoted the establishment of healthy self-management behaviors among youth with epilepsy, including medication adherence, regular follow-ups, and maintenance of healthy sleep routines. Furthermore, while individual behaviors represent manifestations of personal volition and psychological factors, they are profoundly influenced by external environmental contexts [27]. When patients at identical behavioral stages and their caregivers participated in group educational sessions with peer discussions, observed peer behaviors, and shared experiences these factors potentially induced social conformity effects, thereby facilitating positive behavioral modifications. The self-management applet gives patients a convenient and informative reminder service that stimulates and reinforces self-management behaviors. As self-management behaviors in the adolescent-to-adult transition with epilepsy shift to the action and retention phases, self-management skills improve accordingly, facilitating the transition of adolescents with epilepsy to adult medical care. Our findings are consistent with previous research applying the TTM to other chronic conditions. For instance, a study on

XSL•FO
RenderX

diabetes self-management demonstrated that stage-matched interventions significantly improved participants' progression to action and maintenance stages, leading to better glycemic control [28]. Similarly, research in heart failure patients showed that TTM-based counseling was effective in promoting medication adherence and healthy lifestyles [29]. This convergence of evidence across different diseases strengthens the theoretical robustness of the TTM and suggests its broad applicability in chronic disease management, including epilepsy.

Notably, the establishment of self-management behaviors in patients with chronic conditions constitutes a complex and protracted process [30]. This necessitates sustained clinical support from health care providers. Consequently, health care professionals should first acknowledge the longitudinal nature and potential challenges of behavior formation, which fosters patient understanding and trust while facilitating patient-centered support strategies. Furthermore, to promote long-term self-management behaviors among adolescents with epilepsy, clinicians can enhance self-efficacy through collaborative goal-setting, individualized education, and encouragement of self-monitoring and decision-making. Improved self-efficacy and self-management capacity exhibit bidirectional positive reinforcement [31,32], thereby contributing to sustainable health behavior establishment. Additionally, leveraging modern mHealth technologies enables convenient and continuous therapeutic alliances. These digital solutions eliminate temporal and spatial barriers, particularly engaging tech-savvy adolescents through accessible communication channels. Integrated functions such as automated monitoring, seamless communication, and reminder systems may further optimize patient engagement. Finally, family members and caregivers should be actively involved in the health management process, as their provision of emotional support and practical assistance reinforces patients' motivation for self-management.

The TTM-Based mHealth Intervention May Enhance Transition Readiness Among Adolescents With Epilepsy

This study found that at baseline, adolescents with epilepsy demonstrated moderate transition readiness levels, with mean scores of 50.75 (SD 14.02) (intervention group) and 53.10 (SD 9.76) (control group). At the end of 6 months of intervention. Both groups showed significant improvements from baseline in medication management, health care participation, disease knowledge, doctor-patient communication, and transition readiness total scores at 6-month follow-up (all $P<.05$). Notably, the intervention group achieved additional incremental benefits versus controls (medication management: 3.81, 95% CI 1.26-6.36; health care engagement: 2.77, 95% CI 0.52-5.02; disease knowledge: 1.30, 95% CI 0.28-2.31; provider communication: 3.42, 95% CI 1.62-5.22; transition readiness: 11.30, 95% CI 5.70-6.89; effect sizes [Cohen *d*] ranged from 0.527 to 0.864, indicating moderate-to-large clinical effects). It is evident that the TTM-based mHealth intervention may comprehensively enhance transition readiness in adolescents with epilepsy through multiple mechanisms. First, stage-matched behavioral strategies address individualized patient needs, thereby improving self-efficacy and engagement in self-management practices. This personalized approach motivates active learning of disease knowledge and self-care skills. The mobile-delivered remote intervention reduces psychological barriers by leveraging adolescents' preference for digital communication, creating a more relaxed interaction environment compared with face-to-face encounters. The program's exclusive access for patients (excluding parents) fosters independent disease management, while strengthened patient-provider collaboration builds problem-solving confidence through shared decision-making. During the action phase, the intervention systematically reduces parental over-management of medical tasks while establishing collaborative parent-adolescent partnerships, thereby enhancing autonomous self-management and strengthening health care participation and health responsibility. The mHealth platform sustains behavioral engagement through tailored health education delivery, medication adherence reminders, and incentivized tracking mechanisms, which collectively facilitate disease-specific knowledge acquisition, proper pharmacotherapy management, and ultimately the internalization of stable health behaviors. While a predominant focus of many existing mHealth solutions in epilepsy has been on fundamental features such as seizure tracking and medication reminders, the innovative core of our program lies in its strong theoretical foundation in the TTM, which is specifically designed to drive purposeful and progressive behavioral change. This represents a significant evolution from passive monitoring tools towards a more dynamic system of active, staged coaching. It is widely recognized that a common limitation in the field of digital health is the development of apps that provide informational support but lack a robust theoretical framework to guide behavioral modification. Our study directly addresses this gap. The results demonstrate that a theory-driven, behaviorally-focused mHealth intervention can effectively target the multifaceted challenge of transition readiness, which encompasses not only the acquisition of knowledge but also the critical development of skills, motivation, and self-efficacy. Current surveys indicate that adolescents predominantly use mobile platforms during leisure time to access science education information [33]. Mobile-delivered interventions enable precise and concise health education, allowing patients to engage in learning anytime and anywhere during fragmented free moments. This approach not only enhances participation rates but also facilitates timely identification and intervention by health care providers, ultimately supporting adolescents with epilepsy in achieving optimal transition readiness and successful transfer to adult health care systems.

The intervention demonstrated high acceptability among both adolescents with epilepsy and participating health care providers. Existing studies on the evaluation of services for the transition from chronically ill adolescents to adults have focused on indicators of care improvement, disease outcomes, quality of life, and skills improvement, with few evaluating the acceptability of the intervention program [8]. The overall acceptance of the program in this study was good. Adolescents with epilepsy rated the lowest score as "This approach can solve some of my emotional problems," which may be related to the complexity of psychological or emotional problems in people with epilepsy and the need to intervene in multiple ways [34],

XSL•FO
RenderX

suggesting that health professionals need to develop the most appropriate intervention strategies based on a comprehensive assessment, taking into account the environmental and social needs of the individual and the family, when providing psychological interventions. The most appropriate intervention strategies. For example, while common psychological problems can be addressed through efficient and cost-effective group interventions, attention needs to be paid to assessing the psychological needs of specific individuals and targeting interventions. In addition, health care professionals believe that more individualized interventions save time and effort but may be less effective in addressing individualized problems [35]. Therefore, given the limited human resources in health care, group interventions should be maximized, while taking into account the need to identify specific individual problems and individualized interventions.

Conclusions

The TTM-based mHealth intervention demonstrated good feasibility for transitional care in adolescents with epilepsy. This program enabled personalized group interventions while making health management more accessible and continuous. The approach may facilitate healthy behavior establishment and enhance transition readiness, thereby supporting successful transfer to adult health care. Both patients and health care providers reported high acceptability, suggesting good clinical translatability.

Because allocation was by hospital district, the sample was small and self-reported, and follow-up was limited to 6 months. Furthermore, the participants were restricted to adolescents with nonrefractory epilepsy and without cognitive impairment, excluding those with more complex disease characteristics. The results should be interpreted with caution. Future studies should expand the sample size, extend the intervention and follow-up periods, and use randomized controlled trial designs to further validate the intervention's efficacy. Additionally, research should focus on transitional care for adolescents with comorbid cognitive impairments or poorly controlled seizures. This study assessed acceptability using descriptive methods; future work should explore standardized or quantifiable evaluation approaches.

Acknowledgments

This work was supported by a research project at the hospital level of nursing discipline at Children's Hospital of Chongqing Medical University (No.CHCQMU2020.14).

Conflicts of Interest

None declared.

Multimedia Appendix 1
TREND checklist.
[DOCX File , 19 KB - mhealth_v13i1e70085_app1.docx]

References

1. Andrade DM, Bassett AS, Bercovici E, Borlot F, Bui E, Camfield P, et al. Epilepsy: transition from pediatric to adult care. recommendations of the Ontario epilepsy implementation task force. Epilepsia 2017;58(9):1502-1517. [doi: 10.1111/epi.13832] [Medline: 28681381]
2. Betz CL. SPN position statement: transition of pediatric patients into adult care. J Pediatr Nurs 2017;35:160-164. [doi: 10.1016/j.pedn.2017.05.003] [Medline: 28576441]
3. Camfield P, Camfield C. Transition to adult care for children with chronic neurological disorders. Ann Neurol 2011;69(3):437-444. [doi: 10.1002/ana.22393] [Medline: 21391239]
4. Geerlings RPJ, Aldenkamp AP, de With PHN, Zinger S, Gottmer-Welschen LMC, de Louw AJA. Transition to adult medical care for adolescents with epilepsy. Epilepsy Behav 2015;44:127-135. [doi: 10.1016/j.yebeh.2014.12.041] [Medline: 25679495]
5. Camfield PR, Andrade D, Camfield CS, Carrizosa-Moog J, Appleton R, Baulac M, et al. How can transition to adult care be best orchestrated for adolescents with epilepsy? Epilepsy Behav 2019;93:138-147. [doi: 10.1016/j.yebeh.2018.12.015] [Medline: 30642688]
6. Mazur A, Dembinski L, Schrier L, Hadjipanayis A, Michaud P. European academy of paediatric consensus statement on successful transition from paediatric to adult care for adolescents with chronic conditions. Acta Paediatr 2017;106(8):1354-1357. [doi: 10.1111/apa.13901] [Medline: 28471516]
7. Fredericks EM. Transition readiness assessment: the importance of the adolescent perspective. Pediatr Transplant 2017;21(3). [doi: 10.1111/petr.12921] [Medline: 28370888]
8. White PH, Cooley WC, Transitions Clinical Report Authoring Group, American Academy of Pediatrics, American Academy of Family Physicians, American College of Physicians. Supporting the health care transition from adolescence to adulthood in the medical home. Pediatrics 2018;142(5):e20182587. [doi: 10.1542/peds.2018-2587] [Medline: 30348754]
9. Wagner JL, Modi AC, Johnson EK, Shegog R, Escoffery C, Bamps Y, et al. Self-management interventions in pediatric epilepsy: what is the level of evidence? Epilepsia 2017;58(5):743-754. [doi: 10.1111/epi.13711] [Medline: 28233908]

XSL•FO
RenderX

10. Chai K, Zhang Y, Chang K. Regional disparity of medical resources and its effect on mortality rates in China. Front Public Health 2020;8:8. [doi: 10.3389/fpubh.2020.00008] [Medline: 32117848]
11. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
12. Zare M, Tarighat-Esfanjani A, Rafraf M, Shaghaghi A, Asghari-Jafarabadi M, Shamshiri M. The barriers and facilitators of self-management among adults with type 2 diabetes mellitus: a trans theoretical model (TTM)-based mixed method study in Iran. Diabetes Metab Syndr Obes 2020;13:2687-2699 [FREE Full text] [doi: 10.2147/DMSO.S230083] [Medline: 32821141]
13. Teng H, Yen M, Fetzer S, Sung J, Hung S. Tailoring health-promoting programs for patients with chronic kidney disease: randomized controlled trial. West J Nurs Res 2021;43(2):138-150. [doi: 10.1177/0193945920942487] [Medline: 32698694]
14. Liu YS, Ren HY. The application progress of trans-theoretical model on changing the healthy behavior of patients with coronary heart disease. Zhonghua Xin Xue Guan Bing Za Zhi 2019;47(4):331-334. [doi: 10.3760/cma.j.issn.0253-3758.2019.04.013] [Medline: 31060195]
15. Cruz-Ramos NA, Alor-Hernández G, Colombo-Mendoza LO, Sánchez-Cervantes JL, Rodríguez-Mazahua L, Guarneros-Nolasco LR. mHealth apps for self-management of cardiovascular diseases: a scoping review. Healthcare (Basel) 2022;10(2):322. [doi: 10.3390/healthcare10020322] [Medline: 35206936]
16. Bene BA, O'Connor S, Mastellos N, Majeed A, Fadahunsi KP, O'Donoghue J. Impact of mobile health applications on self-management in patients with type 2 diabetes mellitus: protocol of a systematic review. BMJ Open 2019;9(6):e025714. [doi: 10.1136/bmjopen-2018-025714] [Medline: 31243029]
17. Alzamanan MZ, Lim K, Akmar Ismail M, Abdul Ghani N. Self-management apps for people with epilepsy: systematic analysis. JMIR Mhealth Uhealth 2021;9(5):e22489 [FREE Full text] [doi: 10.2196/22489] [Medline: 34047709]
18. Khoshkangin A, Agha Seyyed Esmaeil Amiri FS, Ghaddaripouri K, Noroozi N, Mazaheri Habibi MR. Investigating the role of mobile health in epilepsy management: a systematic review. J Educ Health Promot 2023;12:304. [doi: 10.4103/jehp.jehp_1188_22] [Medline: 38023071]
19. Virella Pérez YI, Medlow S, Ho J, Steinbeck K. Mobile and web-based apps that support self-management and transition in young people with chronic illness: systematic review. J Med Internet Res 2019;21(11):e13579. [doi: 10.2196/13579] [Medline: 31746773]
20. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
21. Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, et al. ILAE official report: a practical clinical definition of epilepsy. Epilepsia 2014;55(4):475-482. [doi: 10.1111/epi.12550] [Medline: 24730690]
22. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: position paper of the ILAE commission for classification and terminology. Epilepsia 2017;58(4):512-521 [FREE Full text] [doi: 10.1111/epi.13709] [Medline: 28276062]
23. Ferris M, Cohen S, Haberman C, Javalkar K, Massengill S, Mahan JD, et al. Self-management and transition readiness assessment: development, reliability, and factor structure of the StarX questionnaire. J Pediatr Nurs 2015;30(5):691-699. [doi: 10.1016/j.pedn.2015.05.009] [Medline: 26209873]
24. Ma J, Yu Q, Ding W, Zhang T, Zhang Y. Psychometric properties of the 'Self-Management and Transition to Adulthood with R = Treatment Questionnaire' in Chinese children and young people with chronic diseases. Int J Nurs Pract 2021;27(2):e12880. [doi: 10.1111/ijn.12880] [Medline: 32935431]
25. Yamamoto K, Ito M, Sakata M, Koizumi S, Hashisako M, Sato M, et al. Japanese version of the mobile app rating scale (MARS): development and validation. JMIR Mhealth Uhealth 2022;10(4):e33725. [doi: 10.2196/33725] [Medline: 35197241]
26. Wood AR, Ross L, Wood RJ. Motivational interviewing and chronic care management using the transtheoretical model of change. Health Soc Work 2023;48(4):271-276. [doi: 10.1093/hsw/hlad020] [Medline: 37615973]
27. Subotnik RF, Olszewski-Kubilius P, Worrell FC. Environmental factors and personal characteristics interact to yield high performance in domains. Front Psychol 2019;10:2804. [doi: 10.3389/fpsyg.2019.02804] [Medline: 31920847]
28. Dunkel A, von Storch K, Hochheim M, Zank S, Polidori MC, Woopen C. Long-term effects of transtheoretical model-based lifestyle intervention on self-efficacy and self-management in patients with type 2 diabetes - randomised controlled trial. Int J Behav Med 2025;32(1):45-57. [doi: 10.1007/s12529-024-10323-0] [Medline: 39349794]
29. Yuan D, Xue Y, Zhou Y. Improving nutritional status in chronic heart failure patients: effectiveness of a transtheoretical model-based stepwise nutritional management program. Risk Manag Healthc Policy 2025;18:1683-1695. [doi: 10.2147/RMHP.S509402] [Medline: 40416832]
30. Ouyang W, Chen H, Xu X, Zhang X, Fu L, Tang F, et al. Self-management program for patients with chronic kidney disease (SMP-CKD) in Southern China: protocol for an ambispective cohort study. BMC Nephrol 2022;23(1):93 [FREE Full text] [doi: 10.1186/s12882-022-02700-2] [Medline: 35247988]
31. Wang Y, Masingboon K, Wacharasin C. Mediating role of self-efficacy in the relationship between family functioning and self-management behaviors in patients with coronary heart disease: a cross-sectional study in Jiangsu, China. Belitung Nurs J 2025;11(1):59-66. [doi: 10.33546/bnj.3638] [Medline: 39877215]

XSL•FO
RenderX

32. Zhang R, Li X, Luo H, Niu J, Zhang H. Effect of self-efficacy, disease perception, social support, anxiety, and depression on self-management in young patients with stroke. J Neurosci Nurs 2025;57(2):68-73. [doi: 10.1097/JNN.0000000000000813] [Medline: 39625809]
33. Goodyear VA, Armour KM, Wood H. Young people and their engagement with health-related social media: new perspectives. Sport Educ Soc 2019;24(7):673-688. [doi: 10.1080/13573322.2017.1423464] [Medline: 31814804]
34. Goselink RJM, Olsson I, Malmgren K, Reilly C. Transition to adult care in epilepsy: a systematic review. Seizure 2022 Oct;101:52-59. [doi: 10.1016/j.seizure.2022.07.006] [Medline: 35901664]
35. Michaelis R, Tang V, Goldstein LH, Reuber M, LaFrance WC, Lundgren T, et al. Psychological treatments for adults and children with epilepsy: evidence-based recommendations by the international league against epilepsy psychology task force. Epilepsia 2018;59(7):1282-1302. [doi: 10.1111/epi.14444] [Medline: 29917225]

Abbreviations

mHealth: mobile health
OR: odds ratio
SMD: standardized mean difference
STARx: Self-Management and Transition to Adulthood with Rx=Treatment
TTM: transtheoretical model

Edited by L Buis; submitted 15.12.24; peer-reviewed by Y Pan, Y Wang, DF Borges; comments to author 30.06.25; revised version received 01.08.25; accepted 14.10.25; published 11.11.25.

Please cite as:
Xia Q, Li S, Wang T, Fan M, Xia J, Xie L, Yin H
Effects of a Transtheoretical Model–Based mHealth Intervention on Transition Readiness in Adolescents With Epilepsy: Quasi-Experimental Study
JMIR Mhealth Uhealth 2025;13:e70085
URL: https://mhealth.jmir.org/2025/1/e70085
doi:10.2196/70085
PMID:41217803

©Qing Xia, Shuangzi Li, Ting Wang, Mingping Fan, Jie Xia, Lingling Xie, Huaying Yin. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Transdiagnostic Cognitive Control Training for Patients Waiting for Outpatient Psychotherapy: Randomized Clinical Trial

Maximilian Blomberg[1,2], PhD; Lisa Oberender[2], MSc; Ernst Koster[3], Prof Dr; Timo Brockmeyer[1,2], Prof Dr

[1]Department of Clinical Psychology and Translational Psychotherapy, Institute of Psychology, University of Münster, Fliednerstrasse 21, Münster, Germany

[2]Department of Clinical Psychology and Psychotherapy, Institute of Psychology, University of Göttingen, Göttingen, Germany

[3]Psychopathology and Affective Neuroscience Lab, Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

Corresponding Author:
Maximilian Blomberg, PhD
Department of Clinical Psychology and Translational Psychotherapy, Institute of Psychology, University of Münster, Fliednerstrasse 21, Münster, Germany

Abstract

Background: Various mental disorders are associated with impaired cognitive control, which is crucial for effective emotion regulation. Cognitive control training has demonstrated promise in enhancing emotion regulation and alleviating distress in disorders characterized by repetitive negative thinking, such as depression and anxiety.

Objective: Given the importance of cognitive control and emotion regulation across mental disorders, this study investigates the efficacy of a mobile cognitive control training in a transdiagnostic outpatient sample awaiting psychotherapy.

Methods: In this randomized clinical superiority trial with 2 parallel arms, 80 patients with various mental disorders from an outpatient waiting list received either 10 sessions of mobile cognitive control training using the Paced Auditory Serial Addition Test (PASAT) or an active control training using a speed of response task. The primary outcome was mental distress, measured by the Hopkins Symptom Checklist-11 (HSCL-11). Secondary outcomes included measures of cognitive control, rumination, repetitive negative thinking, difficulties in emotion regulation, cognitive emotion regulation, and disorder-specific symptoms. Outcomes were measured at baseline, post training, and at 3-month and 6-month follow-up.

Results: Contrary to our primary hypothesis, cognitive control training was not superior in improving global mental distress directly after training (B=−.03, 95% CI –0.21, 0.16; $t_{179.60}$=–0.26; P=.80; d=−0.04, 95% CI –0.35, 0.28); however, it led to greater improvements in cognitive control (B=−0.56, 95% CI –0.59,–0.54; z=−18.02; P<.001; d=−1.23, 95% CI −1.30,–1.20). This effect was similar at the 3-month and 6-month follow-up. Furthermore, at 3-month follow-up, cognitive control training resulted in fewer difficulties in emotion regulation (B=4.73, 95% CI 0.52, 9.12; $t_{177.99}$=2.09; P=.04; d=0.34, 95% CI 0.04, 0.65), and anxiety symptoms (B=2.94, 95% CI 0.38, 5.82; $t_{66.51}$=2.09; P=.04; d=0.70, 95% CI 0.09, 1.38), although the latter refers to a small subsample of patients with anxiety disorders. At 6-month follow-up, cognitive control training led to more adaptive cognitive emotion regulation (B=−5.18, 95% CI −9.74,–0.41; $t_{180.90}$=−2.16; P=.03; d=−0.40, 95% CI −0.75,−0.03), and less social anxiety (B=2.00, 95% CI 0.14, 3.81; $t_{43.43}$=2.08; P=.04; d=0.66, 95% CI 0.05, 1.24). The groups did not differ in any other outcome at any point in time.

Conclusions: This study is the first to assess the efficacy of a mobile cognitive control training using the PASAT in a transdiagnostic outpatient sample. There was no evidence for the training's efficacy on global mental distress and only weak evidence for the superiority in measures of emotion regulation and anxiety at follow-ups. Potential mediating pathways and moderating factors, such as the number of training sessions, should be investigated in larger studies.

Trial Registration: German Clinical Trials Register DRKS00024408; https://drks.de/search/en/trial/DRKS00024408

(JMIR Mhealth Uhealth 2025;13:e65867) doi:10.2196/65867

KEYWORDS

cognitive control; emotion regulation; mHealth; eHealth; transdiagnostic approach

XSL•FO
RenderX

Introduction

Background

Mental disorders are a major contributor to global disability and are associated with increased mortality, somatic comorbidities, and substantial socioeconomic costs [1,2]. While meta-analyses confirm the efficacy of psychotherapy and pharmacotherapy [3,4], effect sizes remain modest and relapse rates are high [5], underscoring the need for more effective treatments.

Frameworks such as the Research Domain Criteria (RDoC) propose a transdiagnostic approach to investigating shared risk factors among mental disorders [6,7]. Emotion regulation is one such factor implicated in the onset, maintenance, and relapse of various disorders [8-10]. Emotion regulation refers to the process by which individuals modulate the quality, intensity, and timing of their emotional responses [11]. Crucially, emotion regulation skills predict symptom improvement across disorders and treatments [12-17], and emotion regulation is targeted by many psychotherapeutic interventions [18-20].

Since Ochsner and Gross's foundational work [21,22], research has increasingly focused on to what extent emotion regulation depends on cognitive control [23-25]. Cognitive control refers to a set of interrelated processes, largely mediated by the prefrontal cortex, that support goal-directed behavior [26,27]. These processes include shifting (flexibly switching between tasks or mental sets), inhibition (suppressing automatic or prepotent responses), and updating (maintaining and renewing relevant information in the working memory while discarding irrelevant material) [28]. A multitude of meta-analyses have shown that numerous mental disorders are associated with reduced cognitive control, covering all domains of shifting, inhibition, and updating (see [29] for a review of meta-analyses).

Notably, impairments in cognitive control, especially difficulties with inhibition and working memory updating, have been associated with maladaptive emotion regulation strategies [23,30-34]. As outlined in the impaired disengagement hypothesis [35], rumination serves as an illustrative case of how cognitive control impairments contribute to maladaptive emotion regulation. According to this hypothesis, individuals with impaired cognitive control struggle to divert their attention and focus from negative (self-referential) information in working memory. This prolongs a ruminative inward focus on the negative material that increases negative affect and thus lowers the threshold for additional negative information to enter working memory [35]. This cycle of increased rumination, qualified by impaired cognitive control, has been supported by meta-analytic findings on the links between rumination and difficulties in discarding no longer relevant material from working memory [36].

Given the links between impaired cognitive control and maladaptive emotion regulation, interventions that enhance cognitive control may offer a promising avenue for alleviating mental distress rooted in such dysfunctional emotion regulation. One commonly used approach to improve cognitive control is based on training using the Paced Auditory Serial Addition Test (PASAT) [37,38]. In the PASAT, participants hear a continuous stream of single-digit numbers ranging from 1 to 9. They are instructed to continuously add and respond to the sum of the last 2 numbers they heard. For example, if they last heard the numbers 2 and 6, the correct answer is 8. If they hear the number 9 next, the correct answer then is 15 (because 6+9=15). This requires them to constantly update their working memory by replacing the next-to-last heard number with the most recently heard number to compute the updated sum. Furthermore, the PASAT requires inhibition, since participants have to suppress responses to the no longer relevant numbers and have to inhibit immediate responses to the current number (before summing the numbers up). As such, the PASAT engages cognitive control processes such as working memory updating and inhibition [28].

PASAT-based cognitive control training has shown efficacy in reducing symptoms of depression and maladaptive emotion regulation, such as rumination, in individuals with current depressive episodes [39-43], those at elevated risk [44], and patients in remission [45-47]. In these intervention studies, the PASAT has frequently been compared to an active control condition that focused on speed of responses (ie, respond immediately to the last heard number) and thus does not target core cognitive control processes such as updating of working memory [48]. However, the benefits of the training have not been observed in nonclinical samples of healthy students [49-51], older adults [52], or high trait ruminators [53]. Moreover, PASAT-based training has not shown an additive benefit when combined with behavioral activation [54] or when used as an adjunct to inpatient treatment [55]. Thus, while promising for clinical populations, its effectiveness appears limited in nonclinical samples and may be limited when combined with other treatments.

Objectives

Most intervention studies using the PASAT have focused primarily on individuals with depression or an increased vulnerability for depression. This may have overlooked the broader applicability of cognitive control training across the spectrum of mental disorders. Given the well-documented impairments in cognitive control and maladaptive emotion regulation across various mental disorders [8,29], this study investigates the efficacy of PASAT-based cognitive control training as a transdiagnostic intervention in a heterogeneous clinical sample waiting for psychotherapy in an outpatient clinic. Our primary hypothesis is that participants receiving PASAT-based cognitive control training will show a significantly greater reduction in mental distress from baseline to post training (primary end point). This effect is expected to exceed that of an active control condition, which emphasizes response speed but does not target cognitive control processes. Based on previous evidence for the intervention's efficacy in reducing rumination in individuals with depression [48], we further hypothesize that cognitive control training will lead to greater improvements in emotion regulation compared to the control condition. However, as a narrow focus on rumination may miss broader emotion regulation impairments across diagnoses [56], we will assess this outcome using multiple measures, including a repetitive negative thinking style, general

XSL•FO
RenderX

difficulties in emotion regulation, and cognitive emotion regulation strategies. To explore the broader functional significance of training-induced improvements, we will also examine effects on work and social adjustment and quality of life. Finally, to capture potential improvements at the symptom level, we additionally examine changes in disorder-specific symptoms. To assess potential longer-term effects of the training, follow-up assessments are conducted not only post training but also at 3 and 6 months.

Methods

Design

Participants were randomly assigned (1:1 ratio) to receive either a cognitive control training or an active control training (ie, a training that focused on response speed). Randomization (without stratification) was carried out automatically with the *psych* package, version 2.2.9 [57] directly before the first training session. We used the online survey framework formr [58] throughout the entire study to conduct the assessments and deliver the interventions. The computerized allocation to training conditions was concealed from the study personnel. Only after a phone call at the time when half of the training sessions were completed (intended to enhance engagement following the first 5 training sessions, see section "Treatments") could participants reveal their condition to the study personnel. Importantly, the study personnel never inquired about details of the participants' training procedures that could disclose their conditions. Upon recruitment, participants were blind to the nature of the different training conditions. They were informed that during the training, they would engage in a task that demands their concentration. After randomization, they only received specific information about their respective condition. All instructions were written and standardized. All outcomes were either self-report measures or computerized tasks, so that no masking of assessors was required. Assessments of baseline, post training, 3-month, and 6-month follow-up data of all self-report questionnaires were conducted independently by the patients through formr. Training and assessment tasks were programmed using JsPsych [59] and JATOS [60]. This trial was conducted in line with the CONSORT (Consolidated Standards of Reporting Trials) guidelines (see Checklist 1).

Participants

Between May 2021 and November 2023, a total of 80 patients were recruited from the waiting list of the outpatient psychotherapy clinic at the Institute of Psychology of the University of Göttingen, Germany, which offers cognitive behavioral therapy. Forty patients (50%) were randomly assigned to the cognitive control training and 40 patients to the speed of response training (active control). Patients were initially made aware of the study with the help of leaflets upon registering for the waiting list of the treatment facility. Patients were eligible for the study if they were between 18 and 65 years old and met the diagnostic criteria for one or more of the following mental disorders that had been associated with cognitive control impairments [29]: major depressive disorder (including recurrent or persistent depressive disorder), agoraphobia, panic disorder, social anxiety disorder (SAD), specific phobia, generalized anxiety disorder, obsessive-compulsive disorder, posttraumatic stress disorder, somatic symptom disorder, eating disorder (anorexia nervosa, bulimia nervosa, and binge-eating disorder), substance use disorder (abstinent for at least 3 months), borderline personality disorder, or adjustment disorder. To maximize external validity, exclusion criteria were kept at a minimum: nonabstinent substance use disorder, severe psychiatric comorbidity (eg, schizophrenia), medical instability, dementia, acute suicidality, or insufficient knowledge of the German language to understand the questionnaires and instructions. Diagnostic eligibility was assessed according to the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) [61], using the *Diagnostic Interview for Mental Disorders* [62]. Screening of adjustment disorder was conducted using the relevant section from the *Structured Clinical Interview for DSM-5 Disorders* [63], since this diagnosis is not covered by the *Diagnostic Interview for Mental Disorders*. Interviews were conducted by clinical psychology graduate students with master's degrees, trained and supervised by a licensed psychotherapist with extensive experience in both interview formats.

Sample size was determined by an a priori power analysis using G*Power [64]. Based on previous findings in a similar study [45], this study was powered to detect a moderate within-between (time x training group) interaction effect of f=0.25 in a repeated-measures design for the primary hypothesis that the cognitive control training would reduce global mental distress from baseline to posttraining assessment significantly more than the speed of response training. With α=.05, statistical power 1–β=.95, and while accounting for an approximate 25% of data loss, a minimum sample size of 72 would have been sufficient, which was conservatively rounded up to 80, following recommendations for pilot trials [65].

Treatments

Overview

Participants performed 10 online training sessions of either the cognitive control training or the speed of response training over the course of 2 weeks. They conducted these sessions independently using their smartphones via a web browser, enabling a flexible and remote completion of the intervention. According to meta-analytic findings, conducting cognitive control training remotely instead of in a laboratory setting does not negatively affect its efficacy [66]. Additionally, remote training provides a more ecologically valid context. Immediately before the first training session, both groups watched a standardized psychoeducative video that was specifically created for this study and that aimed to increase task engagement [41]. For both conditions, the video highlighted that participating in a cognitive training with a demanding task necessitates sustained attention and focus even if frustrating negative thoughts may emerge. The video and its transcript can be found on the project's Open Science Framework page [67]. After completing the first 5 training sessions, participants received a short phone call from the study personnel with the purpose of motivating participants to complete the remaining 5 sessions. During the call, participants' completion of the first half was positively acknowledged, and they were encouraged to finish the second

XSL•FO
RenderX

half. The call also included a discussion on overcoming potential obstacles and improving conditions for training completion. For example, advice was given on training in a private, quiet environment and using headphones to minimize any distraction from the task.

Cognitive Control Training

Participants in the cognitive control training group underwent 10 training sessions, each lasting 20 minutes, using an adaptive version of the PASAT (aPASAT) that has been used in cognitive control training studies before [45,49]. During training, participants hear a continuous stream of single digits (1-9) and must calculate and respond with the sum of the 2 most recently presented digits, selecting the correct answer from options on the screen ranging from 1 to 18. To tailor the training to the individual's performance level, the speed of digit presentation was adjusted dynamically. The interstimulus interval (ISI) between 2 digits began at 3000 milliseconds for each training session and was decreased or increased by 100 milliseconds after every sequence of 4 consecutive correct or incorrect responses, respectively. Besides an initial set of 8 practice trials to familiarize participants with the task, each training session consisted of 400 trials. Consistent with prior research, training progress was monitored using the median ISI levels achieved in each session [45,49].

Speed of Response Training

Participants in the speed of response training group received the same number of training sessions and trials as participants in the cognitive control training group. Contrary to the cognitive control training, patients receiving the control training based on the Speed of Response Task (SOR) [49], in which participants just respond as fast as possible to the last heard number. Therefore, this training does not target the same cognitive processes as the adaptive PASAT but captures response speed and sustained attention. The same criteria for the starting ISI and changes according to performance are applied.

Outcomes

Timeline for Assessments

All outcomes were assessed at baseline, directly following the training, and then again at 3-month and 6-month follow-up. In addition to assessing immediate training effects, we selected these follow-up time points commonly used in cognitive control training studies, as they may capture longer-term effects that only emerge over time [48]. Unlike the other measures, quality of life and work and social adjustment were only assessed at baseline and at 3-month and 6-month follow-up to reduce the time burden for participants and because training was not expected to lead to meaningful changes in these outcomes within such a short period of time.

Primary End Point: Global Mental Distress

The level of global mental symptom distress was assessed using the Hopkins Symptom Checklist-11 (HSCL-11) [68], a self-administered questionnaire with robust psychometric characteristics that measures a range of mental symptoms, such as anxiety and depression. Each of the 11 items is rated on a Likert scale from 1 to 4 with higher scores indicating greater symptom distress (mean scores for the total scale thus range between 1 and 4). At baseline assessment, Cronbach α for the entire sample was 0.84. Given the transdiagnostic approach of the trial, the HSCL-11 was selected for its ability to assess mental distress that is not confined to the specific symptom pattern of a particular mental disorder but instead measures symptoms that are common across various mental disorders [68].

Secondary End Points

Rumination

Rumination was assessed using the 10-item version of the Response Style Questionnaire (RSQ-10) [69]. The RSQ-10 is a self-report questionnaire that measures rumination on a 4-point Likert scale from 1 to 4, with higher scores indicating a stronger tendency to ruminate (total score range: 10 - 40). Reliability and validity of the German version have been demonstrated (RSQ-10D) [70]. Importantly, the brooding and reflection subscales of the RSQ-10 measure rumination above and beyond what can be explained by symptoms of depression [69]. At baseline, Cronbach α was 0.73. Following suggestions by the authors of the RSQ-10, we also conducted exploratory subanalyses focusing on the 5-item brooding subscale of the questionnaire, since brooding depicts the more dysfunctional component of rumination that is associated with mental disorders [69,71].

Repetitive Negative Thinking

Given the transdiagnostic nature of repetitive negative thinking and difficulties with disengagement of (content- and disorder-independent) negative thoughts [72], repetitive negative thinking was measured alongside rumination to assess this specific style of thinking. We used the Perseverative Thinking Questionnaire (PTQ) [73]. Each of the 15 items of the PTQ (eg, The same thoughts keep going through my mind again and again.) is scored on a Likert scale from 0 to 4 with higher scores displaying more repetitive negative thinking (total score range: 0 - 60). At baseline, Cronbach α was 0.93.

Difficulties in Emotion Regulation

Difficulties in emotion regulation were assessed using the 16-item version of the Difficulties in Emotion Regulation Scale (DERS-16) [74]. The DERS-16 is a short form of the original 36-item DERS [75] which retains the good reliability and validity of the original scale. Difficulties in emotion regulation are assessed on a Likert scale from 1 to 5 with higher scores indicating greater difficulties in regulating emotions (total score range: 16 - 80). At baseline, Cronbach α was 0.91.

Cognitive Emotion Regulation

Due to the expected specific nature of the challenges in emotion regulation linked to impairments in cognitive control, we complemented the evaluation of general difficulties in emotion regulation with a more targeted assessment of habitual cognitive emotion regulation using the Cognitive Emotion Regulation Questionnaire (CERQ) [76]. With 36 items which are scored on a Likert scale from 1 to 5, the CERQ measures cognitive emotion regulation on 9 subscales. Consistent with an earlier

XSL•FO
RenderX

study on cognitive control training [45], we calculated compound scores for adaptive cognitive emotion regulation using items from the subscales "positive refocusing," "planning," "positive reappraisal," "putting into perspective," and "acceptance" (total score range: 20 - 100) and for maladaptive cognitive emotion regulation using the items from the subscales "self-blame," "other-blame," "rumination," "catastrophizing" (total score range: 16 - 80). A German translation has been psychometrically validated [77]. At baseline, Cronbach α was 0.78 and 0.70 for the adaptive and maladaptive cognitive emotion regulation compound score, respectively.

Cognitive Control

Cognitive control was measured using a nonadaptive assessment version of the PASAT [37] during which participants performed 3 blocks of increasing difficulty (ISI block 1=3000 ms; ISI block 2=2000 ms; ISI block 3=1500 ms). The assessment consisted of a total of 180 trials, with 60 trials per block. Additionally, participants performed 8 practice trials beforehand. As a performance indicator, we analyzed trial data and fitted a model using a binomial error distribution on the probability of a correct response for an individual trial (refer to the section "Statistical Analysis"). This can also be interpreted as the expected ratio of correct trials per assessment (minimum 0, maximum 1).

Quality of Life

Quality of life was assessed using the World Health Organization Quality of Life Brief Form (WHOQOL-BREF) [78]. The WHOQOL-BREF is a transdiagnostic self-report measure to assess the health condition and impact of illnesses on daily activities and behavior in 4 domains (physical, psychological, social, and environmental) on a Likert scale from 1 to 5. Scores for each domain are converted to a range between 0 (worst) and 100 (best quality of life). Psychometric quality and culture-sensitive validity have been shown for the WHOQOL-BREF in several studies (eg, [78]). Cronbach α at baseline for the 4 domains ranged between 0.70 (social) and 0.83 (psychological).

Work and Social Adjustment

The Work and Social Adjustment Scale (WSAS) [79] is a 5-item self-report measure of functional impairment in work-related and social domains attributable to an identified problem. Items are rated on a Likert scale from 0 (not at all impaired) to 8 (very severely impaired). Total scores range between 0 and 40. Psychometric quality and validity of the WSAS for various disorders have been reported [79]. At baseline, Cronbach α was 0.74.

Disorder-Specific End Points

Depending on their diagnoses, participants completed additional disorder-specific self-report questionnaires at baseline, post training, and at the 3- and 6-month follow-ups. Patients who met *DSM-5* criteria for major depressive disorder completed the Patient Health Questionnaire-9 (PHQ-9) [80]. Anxiety symptoms in patients with an anxiety disorder (excluding SAD) were assessed using the Generalized Anxiety Disorder Questionnaire (GAD-7) [81]. Participants diagnosed with SAD completed the Mini-Social Phobia Inventory (mini-SPIN) [82]. Other diagnoses were less common in our sample (each <5, see Table 1). Further details on disorder-specific outcome measures are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Sample characteristics.[a]

Variable	aPASAT[b] training group (n=40)	SOR[c] training group (n=40)	Total sample (N=80)
Age (years), mean (SD)	26.1 (6.3)	25.6 (3.9)	25.8 (5.2)
Women, n (%)	23 (57)	22 (55)	45 (56)
Psychotropic medication, n (%)	11 (28)	11 (28)	22 (28)
Education (years), mean (SD)	17.3 (3.4)	17.6 (3.3)	17.5 (3.3)
Diagnosis according to the *DSM-5*[d], n (%)			
Major depressive disorder	31 (78)	34 (85)	65 (81)
Anxiety disorder	15 (38)	16 (40)	31 (39)
Posttraumatic stress disorder	2 (5)	1 (2)	3 (4)
Somatic symptom disorder	0 (0)	3 (8)	3 (4)
Social anxiety disorder	10 (25)	11 (28)	21 (26)
Borderline personality disorder	2 (5)	2 (5)	4 (5)
Eating disorder	0 (0)	0 (0)	0 (0)
Obsessive-compulsive disorder	2 (5)	0 (0)	2 (2)
Substance abuse disorder	0 (0)	0 (0)	0 (0)
Adjustment disorder	0 (0)	0 (0)	0 (0)

[a]Data are mean (SD) or n (%); anxiety disorder encompasses panic disorder, agoraphobia, specific phobia, and generalized anxiety disorder; eating disorder encompasses anorexia nervosa, bulimia nervosa, and binge eating disorder.

[b]aPASAT: adaptive Paced Auditory Serial Addition Test.

[c]SOR: Speed of Response Task.

[d]*DSM-5*: *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition).

Additional Measures

We measured treatment credibility and expectancy using the Credibility Expectancy Questionnaire (CEQ) [83] at baseline and post training. Additionally, we assessed the intake of psychotropic medication and monitored the commencement of psychotherapy between baseline and 6-month follow-up.

Procedure

Participants were instructed to complete all study assessments and training on their smartphones. After screening for eligibility, patients returned home and completed the baseline assessment on the next day. They were instructed to complete 10 online training sessions within a 2-week period. Patients could only complete 1 training session per day. To increase adherence, depending on their preferences, participants received automated daily reminders either via email or SMS. We monitored training adherence by regularly checking the completion of training sessions remotely and phoned participants in case of strong deviations from the training schedule (ie, multiple days without completing a training session). Participants returned to the laboratory for the posttraining assessment, where they conducted all self-report questionnaires and the PASAT assessment on their smartphones. At this time, they also received instructions for the upcoming follow-up assessments and engaged in a short qualitative interview on their experience with the training. The results of these qualitative interviews will be reported elsewhere since they are beyond the scope of this quantitative assessment of the cognitive control training. Follow-ups were conducted remotely, starting after the final training session. Patients were debriefed after the last follow-up.

Statistical Analysis

Primary analyses followed an intention-to-treat approach. Outliers (|z|>3) across outcomes were adjusted using winsorization to reduce the influence of extreme values on the analysis. However, results were cross-validated with unadjusted data to ensure robustness. All analyses were conducted using linear mixed-effects models [84] with the *lme4 package* version 1.1 - 31 [85] in R (version 4.5.0; R Core Team) [86].

We compared the average outcome score trajectories between the 2 training conditions using random intercept models and fixed effects for time of assessment, condition, and their interaction. Time of assessment was modeled as a repeated-measures factor with 4 levels (baseline, post training, 3-month, and 6-month follow-up), while training condition was a between-subjects factor (aPASAT vs SOR). For quality-of-life outcomes, time had 3 levels: baseline, 3-month follow-up, and 6-month follow-up. We fitted a model using a binomial error distribution to analyze the data from the PASAT assessment version to measure cognitive control, since PASAT performance is bound between 0 and 1 (portion of correct trials) and therefore not normally distributed. This yields a z-statistic. As performance is expected to change given different ISI, we included the current ISI as a covariate.

As a first step, we compared a full model including time of assessment, training condition, and their interaction to a reduced model with time of assessment as the sole predictor. This

XSL•FO
RenderX

approach guards against "cryptic multiple testing," that is, inflated false positives from interpreting multiple time point–specific effects without first establishing a significant overall interaction [87]. We then examined specific interaction effects between training condition and time of assessment for each interval: from baseline to post training, and to the 3- and 6-month follow-up (3 group-by-time contrasts in total).

To standardize interpretation across outcomes, where lower scores typically indicate improvement (eg, reduced mental distress), we set the aPASAT group as the reference. In this coding scheme, positive interaction effects indicate greater improvement in the aPASAT group relative to the SOR group. For outcomes where higher scores reflect improvement (eg, cognitive control and adaptive emotion regulation), the same reference coding applies, and negative interaction effects similarly indicate greater improvement in the aPASAT group. Two-tailed *P* values for fixed effects were calculated using the *lmerTest* package (version 3.1‐3) [88], applying restricted maximum likelihood estimation and Satterthwaite approximation for degrees of freedom [89]. Effect sizes for fixed effects were computed by dividing the beta coefficient by the SD of outcome scores post training in the SOR condition [90]. CIs (95%) were derived using nonparametric bootstrapping with 1000 resamples [91].

To evaluate the robustness of results against missing data, we used a pattern-mixture model approach [92]. To this end, we compared pattern-mixture models including missingness as an additional fixed effect to standard models via likelihood ratio tests.

To explore potential moderators, we tested 3-way interactions between each moderator, time of assessment, and condition. Moderators included training progress (measured as change in median ISI from first to last training session [41]), number of completed training sessions, and participant age (given known age-related decline in cognitive control; see [93]). All moderators were z-standardized.

Finally, we conducted a per-protocol analysis including only participants who completed at least 7 or more (ie, ≥70%) training sessions and computed proportions of clinically significant change following the criteria by Jacobson and Truax [94].

Ethical Considerations

This monocenter randomized controlled superiority trial with 2 parallel arms was approved by the ethics committee of the Georg-Elias-Müller-Institute of Psychology, University of Göttingen, Germany (case number #205) and registered at the German Clinical Trials Register [95]. The study procedures adhered to the ethical standards of the institutional and national research committees and to the principles of the Declaration of Helsinki. All participants provided written informed consent and received financial compensation (€50 [approximately US $55]) for their participation. All data were deidentified prior to analysis to ensure participant privacy and confidentiality. Deidentified data, analysis code, and source code of the mobile trainings and assessments are available on the project's Open Science Framework page [67].

Results

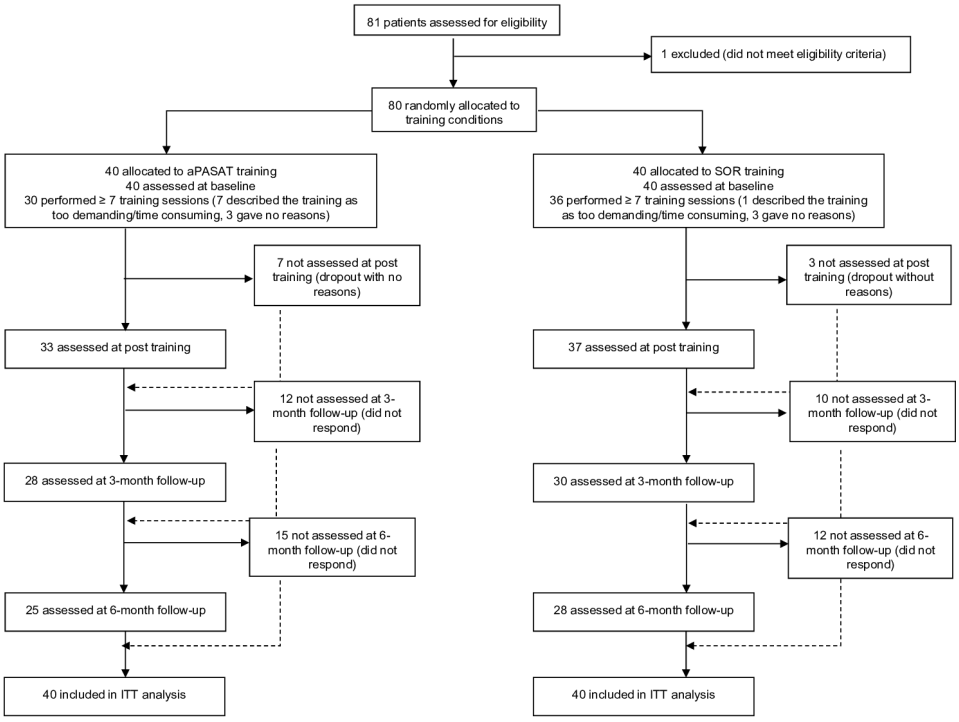
Participant Overview

Sample characteristics including the distribution of diagnoses across conditions and the total sample can be seen in Table 1. Training groups did not differ regarding any of these characteristics (all $P>.24$).

Participant flow is displayed in Figure 1. Fifteen (38%) patients in the cognitive control training group did not complete the 6-month follow-up, and 12 (30%) in the speed of response training group. Rates of complete and incomplete assessments did not differ ($\chi^2_1=0.81$; $P=.37$).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart. aPASAT: adaptive version of the Paced Auditory Serial Addition Test; ITT: intention-to-treat; SOR: Speed of Response Task.



Participants in the cognitive control training group performed 8.1 training sessions on average (SD 2.9). Participants in the speed of response training condition performed 9.1 (SD 1.8) training sessions on average. Training dosage did not differ between groups ($t_{62.15}$=1.73; *P*=.09). Thirty (75%) participants in the cognitive control group completed the training per protocol (ie, at least 7 training sessions), and 36 (90%) in the speed of response training group. Rates of training completion per protocol again did not differ (χ^2_1=2.16; *P*=.14).

On average, participants in the cognitive control training group spent 14.0 (SD 9.9) days completing the training sessions, while participants in the speed of response training group spent 15.5 (SD 7.7) days. Training groups did not differ in this regard ($t_{71.61}$=0.76; *P*=.45).

At posttraining assessment, 8 patients in the aPASAT training condition had started outpatient therapy, while 5 did so in the SOR training group. Furthermore, patients in the aPASAT training group received, on average, 3.1 (SD 1.9) therapy sessions until posttraining assessment, and patients in the SOR group received 2.8 (SD 1.3) sessions. At the 3-month follow-up, 26 patients in the aPASAT and 35 in the SOR training condition commenced psychotherapy. Patients in the aPASAT group received, on average, 7.2 (SD 3.0) sessions until 3-month follow-up, while patients in the SOR group received 7.5 (SD 3.1) sessions. Finally, at the 6-month follow-up, nearly all participants have started the awaited outpatient psychotherapy, with 29 in the aPASAT and 36 in the SOR training group. Patients in the aPASAT group received 11.0 (SD 6.4) therapy sessions until 6-month follow-up, while patients in the SOR group received 11.5 (SD 4.5) sessions. Training groups did not differ in therapy commencement (all *P*>.46), nor in the number of therapy sessions at any assessment (all *P*>.72).

Means, SDs, and number of participants for primary, secondary, and disorder-specific outcomes of depression and (social) anxiety across all assessments can be found in Table 2. Due to the small number of patients meeting criteria for a diagnosis of posttraumatic stress disorder, somatic symptom disorder, borderline personality disorder, an eating disorder, obsessive-compulsive disorder, substance use disorder, or adjustment disorder (each <5, see Table 1), no statistical models on disorder-specific measures have been conducted for these subsamples. Still, means, SDs, and number of participants for these outcomes together with quality of life and work and social adjustment across all assessments are available in Table S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Means and SDs at baseline, post training, and follow-ups.

Outcome	Baseline		Post training		Three-month follow-up		Six-month follow-up	
	aPASAT[a]	SOR[b]	aPASAT	SOR	aPASAT	SOR	aPASAT	SOR
HSCL-11[c]								
n	40	40	33	37	28	30	25	28
Mean (SD)	2.07 (0.57)	2.00 (0.50)	2.03 (0.56)	1.87 (0.59)	1.98 (0.53)	1.79 (0.51)	1.92 (0.51)	1.86 (0.52)
RSQ-10D[d]								
n	40	40	32	37	28	30	25	28
Mean (SD)	23.27 (5.30)	23.30 (4.88)	23.50 (4.33)	21.95 (5.47)	23.36 (5.21)	20.77 (4.92)	22.48 (5.61)	21.11 (5.43)
PTQ[e]								
n	40	40	32	37	28	30	25	28
Mean (SD)	39.55 (9.75)	36.17 (10.05)	39.19 (7.91)	31.78 (11.84)	35.37 (8.83)	31.30 (11.03)	35.40 (10.92)	32.50 (12.85)
DERS-16[f]								
n	40	40	32	37	28	30	25	28
Mean (SD)	51.00 (13.50)	44.88 (11.61)	49.66 (11.24)	45.16 (14.00)	46.07 (11.13)	41.93 (12.54)	48.24 (12.29)	42.93 (14.90)
CERQ[g] (adaptive)								
n	40	40	32	37	28	30	25	28
Mean (SD)	49.55 (9.07)	55.90 (11.95)	51.47 (9.39)	56.19 (12.94)	51.14 (9.10)	53.67 (12.33)	50.56 (8.05)	51.82 (12.33)
CERQ[h] (maladaptive)								
n	40	40	32	37	28	30	25	28
Mean (SD)	39.64 (9.31)	38.80 (8.71)	40.44 (9.10)	37.78 (10.52)	39.11 (7.71)	36.67 (10.38)	39.60 (9.83)	36.75 (12.72)
PASAT[i]								
n	37	36	33	36	25	27	24	23
Mean (SD)	0.58 (0.18)	0.56 (0.16)	0.88 (0.12)	0.71 (0.15)	0.87 (0.09)	0.73 (0.14)	0.90 (0.06)	0.79 (0.13)
PHQ-9[j]								
n	31	34	25	31	22	26	18	24
Mean (SD)	12.94 (5.08)	11.24 (5.07)	11.08 (4.66)	9.50 (5.25)	11.32 (3.93)	8.92 (4.17)	10.78 (4.10)	9.50 (4.42)
GAD-7[k]								
n	15	16	12	15	10	12	10	12
Mean (SD)	10.07 (5.13)	8.62 (4.81)	9.92 (4.19)	7.93 (4.22)	6.10 (3.75)	6.83 (4.09)	8.50 (4.28)	8.75 (5.12)
mini-SPIN[l]								
n	10	11	7	11	9	8	8	6
Mean (SD)	9.80 (2.39)	7.09 (2.81)	10.29 (1.80)	6.18 (3.06)	8.56 (2.79)	7.25 (2.60)	9.00 (2.73)	7.83 (3.06)
RSQ-10D[m] (brooding)								
n	40	40	32	37	28	30	25	28
Mean (SD)	11.15 (2.82)	11.07 (2.64)	11.16 (2.13)	10.62 (2.86)	11.04 (2.74)	10.10 (2.62)	10.84 (2.82)	10.04 (2.81)
RSQ-10D[n] (reflection)								
n	40	40	32	37	28	30	25	28
Mean (SD)	12.12 (3.17)	12.22 (2.84)	12.34 (2.60)	11.32 (3.00)	12.32 (2.87)	10.67 (2.71)	11.64 (2.97)	11.07 (2.96)

XSL•FO
RenderX

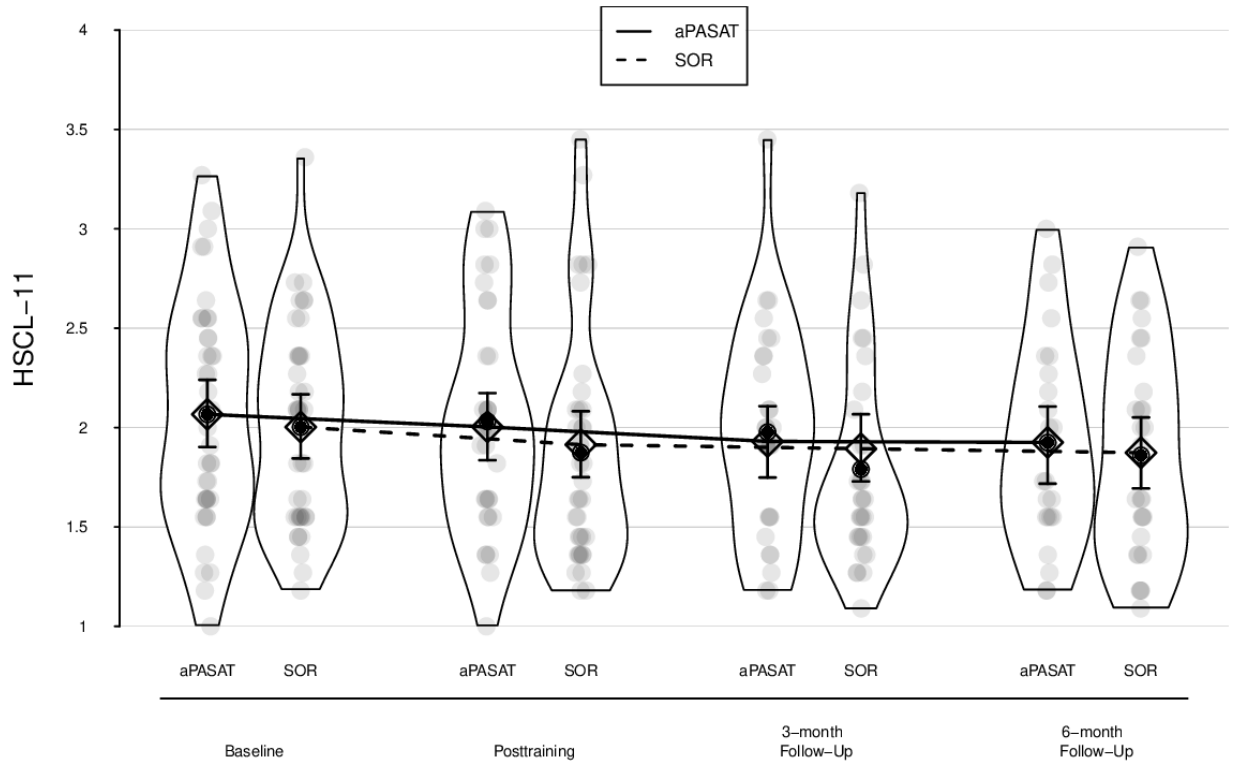
[a]aPASAT: adaptive version of Paced Auditory Serial Addition Test.
[b]SOR: Speed of Response Task.
[c]HSCL-11: Hopkins Symptom Checklist-11.
[d]RSQ-10D: Response Style Questionnaire (10-item version).
[e]PTQ: Perseverative Thinking Questionnaire.
[f]DERS-16: Difficulties in Emotion Regulation Questionnaire (16-item version).
[g]CERQ (adaptive): Cognitive Emotion Regulation Questionnaire (adaptive subscale).
[h]CERQ (maladaptive): Cognitive Emotion Regulation Questionnaire (maladaptive subscale).
[i]PASAT: Paced Auditory Serial Addition Test.
[j]PHQ-9: Patient Health Questionnaire 9.
[k]GAD-7: Generalized Anxiety Disorder Scale 7.
[l]mini-SPIN: Mini-form of the Social Phobia Inventory.
[m]RSQ-10D (brooding): Response Style Questionnaire (brooding subfacet).
[n]RSQ-10D (reflection): Response Style Questionnaire (reflection subfacet).

Intention-to-Treat Analysis

Global Mental Distress

The full-reduced model comparison for the primary outcome indicated that including training condition and its interaction with time of assessment did not significantly improve the model fit for global mental distress (χ^2_4=0.60; *P*=.96). Consistent with the full-reduced model comparison, the individual effect for the interaction between training condition and time until postassessment (primary end point) was also not significant (B=−0.03, 95% CI −0.21, 0.16; $t_{179.60}$=−0.26; *P*=.80; *d*=−0.04, 95% CI−0.35, 0.28). The same applied to the interaction effect of training condition and time between baseline and 3-month as well as 6-month follow-up assessment (secondary end point; B=0.03, 95% CI −0.17, 0.23; $t_{181.34}$=0.25; *P*=.80; *d*=0.05, 95% CI −0.28, 0.39; and B=0.01, 95% CI −0.21, 0.22; $t_{181.69}$=0.11; *P*=.91; *d*=0.02, 95% CI −0.35, 0.37, respectively). Taken together, this indicates that the change in global mental distress over time was not different between training groups, neither in the short term nor in the long term. Figure 2 displays the change trajectories for HSCL-11 scores by group, including both raw data and model-based estimates. In addition, Figure S3 in the Multimedia Appendix 1 depicts individual patient-level changes in global mental distress as measured with the HSCL-11 from baseline to 6-month follow-up.

Figure 2. Model-based change trajectories in global mental distress. Gray dots = individual data points; black dots = empirical means; diamonds = model−based estimates (95% CIs). Bean width reflects distribution. aPASAT: adaptive Paced Auditory Serial Addition Test training; HSCL−11: Hopkins symptom checklist 11; SOR: Speed of response training.



XSL•FO
RenderX

Cognitive Control

The full-reduced model comparison indicated that including training condition and its interaction with time of assessment significantly improved model fit for the PASAT assessment outcome (χ^2_4=389.04; *P*<.001). Consistent with this, the individual effect for the interaction between training condition and time until posttraining assessment was significant for PASAT performance (B=−0.56, 95% CI −0.59, - 0.54; *z*=–18.02; *P*<.001; *d*=−1.23, 95% CI −1.30,−1.20). The same was true for the interaction between training condition and time until 3- and 6-month follow-up (B=−0.42, 95% CI −0.48,−0.40; *z*=−11.62; *P*<.001; *d*=−0.92, 95% CI −1.06,−0.88; and B=−0.44, 95% CI −0.50,−0.42; *z*=−11.22; *P*<.001; *d*=−0.97, 95% CI −1.10,−0.92, respectively). In summary, the improvement in cognitive controls as measured by the assessment version of the PASAT was more pronounced in the cognitive control training group across all points in time.

Emotion Regulation

None of the full-reduced model comparisons for the outcomes measuring rumination, repetitive negative thinking, difficulties in emotion regulation, or cognitive emotion regulation were significant (all *P* >.06). This indicates that overall changes over time were not different between training conditions for these measures. Consequently, significant effects at the level of individual interactions (eg, the interaction between training condition and time from baseline to posttraining assessment) should be interpreted with caution, as we did not specify hypotheses for every outcome at each assessment, and the effects were not strong enough to substantially improve model fit beyond what was explained by time alone considering all assessments.

However, the analysis of time point–specific interaction effects indicated that cognitive control training was superior to speed of response training in improving difficulties in emotion regulation (B=4.73, 95% CI 0.52, 9.12; $t_{177.99}$=2.09; *P*=.04; *d*=0.34, 95% CI 0.04, 0.65) and adaptive cognitive emotion regulation (B=−3.81, 95% CI −8.55, 0.67; $t_{180.51}$=–1.64; *P*=.10; *d*=−0.29, 95% CI −0.66, 0.05) from baseline to 6-month follow-up. In contrast, cognitive control training appeared less efficacious than the speed of response training in reducing rumination over the same period (B=−2.35, 95% CI −4.49,–0.19; $t_{182.49}$=−2.16; *P*=.03; *d*=−0.43, 95% CI −0.82,−0.03). Details on estimates and effect sizes for other nonsignificant comparisons as well as conditional main effects of time (effects of time in the cognitive control training group) can be found in Table S3 in Multimedia Appendix 1.

Disorder-Specific Outcomes

Depression

Most patients (81%) met criteria for a diagnosis of major depressive disorder (see Table 1). As indicated by the nonsignificant full-reduced model comparison, training groups did not differ regarding changes in symptoms of depression as measured with the PHQ-9 (χ^2_4=3.08; *P*=.54). Consistent with that, symptoms of depression did not change differently between training groups from baseline to posttraining assessment (B=0.85, 95% CI −0.92, 2.74; $t_{142.00}$=0.87; *P*=.39; *d*=0.16, 95% CI −0.18, 0.52), 3-month follow-up (B=1.26, 95% CI −0.53, 3.14; $t_{143.30}$=1.21; *P*=.23; *d*=0.24, 95% CI −0.10, 0.60), or 6-month follow-up (B=1.39, 95% CI −0.65, 3.57; $t_{143.72}$=1.27; *P*=.21; *d*=0.26, 95% CI −0.12, 0.68).

Social Anxiety

The full-reduced model comparisons revealed a significant interaction of time and training condition for symptoms of social anxiety as measured by the mini-SPIN (χ^2_4=13.32; *P*=.01). While symptoms of social anxiety did not change differently between training groups from baseline to posttraining assessment (B=−0.32, 95% CI −1.93, 1.24; $t_{43.04}$=−0.37; *P*=.72; *d*=−0.10, 95% CI −0.63, 0.41), the interaction was significant from baseline to 3-month follow-up (B=1.93, 95% CI 0.28, 3.72; $t_{43.26}$=2.16; *P*=.04; *d*=0.63, 95% CI 0.09, 1.21), and from baseline to 6-month follow-up (B=2.00, 95% CI 0.14, 3.81; $t_{43.43}$=2.08; *P*=.04; *d*=0.66, 95% CI 0.05, 1.24). However, these findings are based on the small number of patients who met the diagnostic criteria for SAD and who had also completed the 3-month follow-up (n=17) or 6-month follow-up (n=14), respectively (see Table S6 in Multimedia Appendix 1 for the number of measurements per outcome and assessment). As such, they should be interpreted with caution.

Anxiety

The full-reduced model comparison showed that the interaction between training condition and time did not improve the model fit for symptoms of anxiety other than social phobia beyond effects of time (χ^2_4=5.43; *P*=.25). In line with the full-reduced model comparison, symptoms of anxiety did not change differently between groups from baseline to posttraining assessment (B=0.45, 95% CI −1.90, 2.99; $t_{66.00}$=0.34; *P*=.73; *d*=0.11, 95% CI −0.45, 0.71). However, the interaction of training condition and time was significant between baseline and 3-month follow-up (B=2.94, 95% CI 0.38, 5.82; $t_{66.51}$=2.09; *P*=.04; *d*=0.70, 95% CI 0.09, 1.38), but this effect was not maintained to the 6-month follow-up (B=1.65, 95% CI −0.91, 4.32; $t_{66.46}$=1.17; *P*=.25; *d*=0.39, 95% CI −0.22, 1.03). Like the analysis for symptoms of social anxiety, it was only a small number of patients (n=22) who met the diagnostic criteria for an anxiety disorder other than SAD and who also completed the 3-month follow-up (see Table S6 in Multimedia Appendix 1). As with the analysis of emotion regulation, the nonsignificant full-reduced model comparison advises caution in interpreting single interaction effects.

All other disorder-specific outcomes were not analyzed in detail due to the small number of patients (each n<5) meeting diagnostic criteria of the respective disorder. However, means and SDs for these outcomes are provided in Table S4 in Multimedia Appendix 1.

Quality of Life and Work and Social Adjustment

All full-reduced model comparisons showed that adding training condition and its interaction with time did not change the model fit for any quality of life and work and social adjustment outcome, including all WHOQOL-BREF subscales and the

XSL•FO
RenderX

WSAS, all P>.22. Estimates and effect sizes can be found in Table S3 in Multimedia Appendix 1.

Additional Tables and Figures

The number of measurements and participants in the linear mixed-effect models can be found in Table S5 in Multimedia Appendix 1. Change trajectories for other outcomes, including disorder-specific outcomes and quality of life and work and social adjustment, are illustrated in Figures S4-S17 in Multimedia Appendix 1.

Credibility and Expectancy

Credibility and expectancy for the training's efficacy both decreased between baseline and posttraining assessment (B=−0.82, $t_{77.75}$=–2.79, P=.007; and B=–0.63, $t_{74.68}$=–2.34, P=.02, respectively). Groups did not differ in this regard (all P>.34).

Clinical Significance

Only 1 patient experienced a clinically significant change in the primary outcome (decrease in HSCL-11 total score of at least 0.62 and final HSCL-11 score lower than 1.91) [68] from baseline to 6-month follow-up. No clinically significant changes occurred from baseline to the posttraining or 3-month follow-up assessment. Additionally, there were no patients who experienced a clinically significant change in symptoms of depression (decrease in PHQ-9 of at least 6.76 and final PHQ-9 score below 8.60) [80] from baseline to post training, 3-month, or 6-month follow-up assessment.

Missing Data Analysis

For all measures including global mental distress, emotion regulation, and cognitive control, likelihood ratio tests of the pattern mixture models indicated that the results did not change when including missingness into the models (all P>.28, see Table S6 in Multimedia Appendix 1). This indicates that missing values had no influence on the above-reported results.

Per-Protocol Analysis

The analysis of the data of participants that finished at least two-thirds of the training sessions revealed the same pattern of results with one exception: for repetitive negative thinking, the full-reduced model comparison was significant, P=.05. However, none of the individual interaction effects between training condition and time of assessment between baseline, post training, 3-, or 6-month follow-up were significant (all P>.19). Therefore, the significant full-reduced model comparison likely reveals a general difference between training conditions in repetitive negative thinking, which did not change across time of assessments.

For detailed estimates and effect sizes for direct comparisons between training groups at all assessments in the per-protocol analysis, see Table S7 and Table S8 for the number of measurements and participants in Multimedia Appendix 1. Figures S18-S24 in Multimedia Appendix 1 depict change trajectories for global mental distress, emotion regulation, and cognitive control across conditions in the per-protocol analysis.

Moderation Analysis

Training Progress

Training performance as measured by the median ISI in a training session improved in the cognitive control training condition (B=–64.70; $t_{69.28}$=–7.60; P<.001). A significant interaction effect showed that the training progress was stronger in the cognitive control compared to the speed of response training group (B=57.31; $t_{68.56}$=4.94; P<.001).

Training progress as measured by change in median ISI significantly moderated the interaction effect between training condition and time for the performance in the assessment version of the PASAT (B=–0.08; z=–5.65; P<.001). This simply indicates that cognitive control training was more efficacious in improving near-transfer cognitive control for those patients with more training progress. All other moderation effects by training progress were not significant (all P>.12).

Age

The 3-way interaction between age, training condition, and assessment was not significant for any primary or secondary outcome (all P>.09), indicating that differing changes in outcomes between training groups over time were not moderated by the age of participants.

Number of Training Sessions

Although the number of completed training sessions was not different between training conditions (see above), the number of completed training sessions significantly moderated the interaction effect of time and training condition. Specifically, completing more training sessions led to greater reductions in global mental distress in the speed of response compared to the cognitive control training (B=–0.16; $t_{204.86}$=–3.01; P=.003). The same was true for improvements in rumination (B=–1.64; $t_{208.79}$=–3.02; P=.003) and cognitive control (B=–0.18; z=–7.68; P<.001). In other words, participants who received the speed of response training showed greater improvements in these outcomes than those in the cognitive control group when they completed more training sessions.

Discussion

Principal Results

To our knowledge, this study is the first to examine the efficacy of mobile cognitive control training across a broad spectrum of mental disorders in individuals waiting for outpatient psychotherapy. Contrary to our primary hypothesis, cognitive control training, using the adaptive PASAT, was not superior to speed of response training in reducing global mental distress, neither in the short nor in the long term.

Comparison With Prior Work

These findings diverge from earlier research on adaptive PASAT training in individuals with major depression [39-43]. Four of these 5 studies focusing on patients with current major depression combined cognitive control training with transcranial direct current stimulation [39,40,42,43]. Notably, control groups in these studies receiving cognitive control training paired with

XSL•FO
RenderX

sham transcranial direct current stimulation improved as well, suggesting that cognitive control training might be similarly effective without any parallel brain stimulation. However, these studies were characterized by relatively small sample sizes, with the largest n=43 [41]. This limitation is notable, as only large effects could be detected in such small samples, unlike the smaller overall effect (Hedges g=0.29) reported for PASAT-based cognitive control training in a meta-analysis [48]. Consequently, studies [39] and [42], which reported large postassessment training effects greater than 1.00, may be considered outliers given their small sample sizes. When planning our study, we based our sample size on the most comparable study using the same intervention protocol, which reported a medium-sized effect on the primary outcome [45]. However, considering the meta-analysis published after our data collection was completed, future studies should aim for larger sample sizes. Our results are consistent with studies that examined cognitive control training as an adjunct to inpatient treatment [55] or in combination with behavioral activation for outpatients [54], which found no added benefit from the training. Given the efficacy of cognitive control training in studies with larger samples of individuals with vulnerability to or in remission from major depression [45,96], it can be tentatively concluded that the intervention may be more suitable as a (relapse) prevention for depression and less for people with an acute full-syndrome mental disorder.

Implications

The cognitive control training led to significant improvements in the nonadaptive assessment version of the PASAT, showing a near-transfer effect. This outcome is expected given that the training and the assessment version of the task are almost identical. Cognitive control training often involves assessment tasks closely related to those used in the intervention, and studies typically report limited far-transfer effects to other cognitive tasks [53,55,97]. Conversely, one study found that adaptive PASAT training enhanced performance in an n-back task, which targets similar cognitive functions, highlighting some degree of transfer [49]. Given the already challenging nature of the PASAT [98,99] and the number of outcome measures used to cover a broad spectrum of clinical parameters and emotion regulation, we did not assess transfer effects to other cognitive control tasks. Despite several months without further training, the cognitive control training group maintained a higher performance on the assessment version of the PASAT compared to the speed of response group at follow-ups, indicating the potential for a lasting impact of the training. Future research should explore far-transfer effects using various cognitive measures, ideally with minimal overlap in processes not related to cognitive control, such as auditory or spatial processing.

The analysis of individual effects indicated superiority of the cognitive control training in improving difficulties in emotion regulation from baseline to 3-month follow-up and in cognitive emotion regulation from baseline to 6-month follow-up, respectively. These results align with prior research on patients remitted from depression [45]. However, contrary to the findings in the previous study that also used the CERQ, our results showed an interaction effect in the compound score for adaptive cognitive emotion regulation. We did not find any differences for the compound score of maladaptive cognition and emotion regulation. Additionally, a significant interaction effect for rumination from baseline to 3-month follow-up showed that the improvement was stronger for the speed of response training group. This result, which indicates the inferiority of the PASAT training condition regarding rumination, is in stark contrast to the meta-analytic findings in clinically depressed samples [48]. However, for all outcomes related to emotion regulation, the full-reduced model comparison was not significant. Therefore, these single effects were not substantial enough considering all assessments. As we did not formulate hypotheses for each of these secondary outcomes for each given time interval, there is a risk for these individual effects to be statistical artifacts, and they should be interpreted with caution.

The findings on disorder-specific symptoms tentatively suggest that the training may be able to reduce anxiety symptoms, especially social anxiety, in the long term. However, it is important to consider the limited number of patients with the respective diagnosis who provided data at follow-up assessments (see Table S5 in Multimedia Appendix 1). Thus, these findings should be treated with caution until replication in larger samples.

Any improvements in outcomes that were not significantly different between training conditions may be attributed to factors that applied to both groups equally, such as the commencement of or number of psychotherapy sessions during follow-up periods (which were not different between training groups), spontaneous symptom improvements, regression to the mean, demand effects, or a combination of those. It also cannot be ruled out that shared but unknown mechanisms of both cognitive control and speed of response training contributed to these improvements. Although the speed of response task may not train the updating of working memory content, it does necessitate sustained attention to effectively perform the task, thus making it an active control condition [45].

Finally, the optimal dosage and frequency of cognitive control training need to be established. A recent study suggests that at least 10 training sessions are necessary to achieve a medium-sized effect on symptoms of depression [100]. Consistent with our approach, most studies using the PASAT have implemented approximately 10 training sessions [48]. Contrary to our expectations, the moderation analysis suggests that cognitive control training was relatively less efficacious than speed of response training in alleviating global mental distress, reducing rumination, and improving cognitive control among participants who completed a high number of training sessions. However, this was an exploratory analysis, and identifying moderators in 3-way interactions typically requires sample sizes in the several hundred [101], far exceeding that of this study. Moreover, the number of training sessions was limited, as participants could complete 10 training sessions at most. Additionally, most participants completed 7 or more training sessions (see “Participant Overview”). This limits the variability needed to meaningfully assess the moderating role of differing numbers of training sessions.

Beyond determining the optimal dosage for posttraining effects, future research should address whether booster sessions during follow-up periods could maintain or enhance long-term

XSL•FO
RenderX

outcomes. Furthermore, it would be similarly important to investigate the ideal timing of training sessions, for instance, in individuals at risk for depression, training during periods of increased rumination may amplify its efficacy [102]. In this regard, mobile health interventions, such as the mobile cognitive control training studied here, might serve as suitable just-in-time adaptive interventions [103].

Limitations

The findings of this study should be interpreted considering its methodological strengths and limitations. Despite our efforts to reach out to patients to encourage the completion of assessments, nearly half of the participants had missing data at one of the 4 assessments, primarily the 3-month and 6-month follow-up (see Figure 1). Notably, however, attrition rates did not differ between groups, and the vast number of participants finished the training per protocol, that is, they completed at least 7 of the 10 scheduled training sessions. This indicates that factors related to attrition at follow-up are likely unrelated to the specifics of the training protocol. Moreover, our follow-up sample sizes are comparable with those in other cognitive control training studies [48].

While a meta-analysis suggests that PASAT-based cognitive control training is more effective in reducing depressive symptoms immediately after the training than at follow-up [48], the lack of follow-up assessments in most studies limits conclusions about long-term effects. Small samples and highly variable follow-up intervals (ranging from weeks to a year) further reduce power and comparability. Some evidence points to longer-term benefits for outcomes such as rumination and emotion regulation [45]. These effects may unfold gradually, as participants need time to apply trained strategies such as filtering irrelevant (negative) information from working memory in daily life. Future studies should extend follow-up periods up to 12 months, as done in cognitive bias modification research [104], and explore whether training-induced gains in cognitive control enhance responsiveness to subsequent psychotherapy [105].

Our sample, with a mean age of 25.8 (SD 5.2) years, was relatively young compared to other cognitive control training studies in depression [48]. This might be since the treatment facility is situated on a university campus, and therefore often frequented by students with mental health problems. Cognitive control tends to decline with age [93], but we found no evidence that age moderated training effects in our relatively young sample. Future studies should include more diverse community samples in terms of age and education. Since mental disorders frequently co-occur, some participants met criteria for more than one disorder (see Table 1). This raises the possibility of overlap in disorder-specific analyses, as individuals may influence multiple outcomes. However, our primary outcome was global mental distress, not tied to any single diagnosis. Given the high prevalence of comorbidity in clinical populations [106], we chose to include patients with multiple diagnoses to enhance ecological validity. Finally, due to the monocentric design and recruitment from a Western European outpatient setting, the generalizability of our findings to other clinical or cultural contexts may be limited.

Notably, almost no patients showed clinically significant improvements in global mental distress or depressive symptoms at the 6-month follow-up, which is unexpected given that most began psychotherapy during this period. However, our sample reported relatively low baseline distress, leaving limited room for substantial improvement. For example, patients with major depression had an average baseline PHQ-9 of 12.0 (see Table 2). Given the PHQ-9's reliable change index of 6.8 points [80], a clinically significant improvement would require a follow-up score below 6, which is typical of healthy individuals [107]. Moreover, such changes generally do not occur within the first few sessions of outpatient therapy [108]. That most participants began psychotherapy during the follow-up is unsurprising given the recruitment from a waiting list. While this could confound observed improvements, clinically significant changes were rare overall, and psychotherapy uptake did not differ between training groups, nor did the number of therapy sessions between groups. Thus, any confounding effect by psychotherapy is either likely minimal or evenly distributed between groups.

Conclusions

This study examined the efficacy of mobile cognitive control training as a transdiagnostic intervention for individuals awaiting outpatient psychotherapy. Despite its limitations, the study's design allowed for an ecologically valid training context. In summary, we did not find evidence for the superiority of the cognitive control training over an active control in reducing symptoms of mental distress. Based on our results, we cannot generally suggest cognitive control training in its current form as an intervention for patients with an acute, full-blown mental disorder. Furthermore, examining the phase of illness, potential working mechanisms, and far-transfer effects in large samples is necessary to further assess the potential of this intervention.

Acknowledgments

The authors would like to acknowledge the assistance provided by the staff of the outpatient psychotherapy clinic at the Institute of Psychology of the University of Göttingen, Germany, and by student assistant Nina Vogel for their help with data collection. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The study was supported by internal resources from the Department of Clinical Psychology and Psychotherapy, Institute of Psychology, University of Göttingen, Germany.

XSL•FO
RenderX

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Authors' Contributions

TB and EK contributed to the conceptualization of the study. MB was responsible for data curation and formal analysis. TB acquired funding for the project. MB and LO conducted the investigation. TB, EK, and MB contributed to the development of the methodology. TB and MB were responsible for project administration and resource provision, with support from LO. MB developed the software. TB and EK supervised the research activity. MB and TB contributed to validation and visualization. MB prepared the original draft of the manuscript. TB, EK, MB, and LO reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional material including tables, figures, and details for disorder-specific outcomes.
[PDF File, 1963 KB - mhealth_v13i1e65867_app1.pdf]

Checklist 1
CONSORT checklist.
[PDF File, 246 KB - mhealth_v13i1e65867_app2.pdf]

References

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2]
2. Smith K. Trillion-dollar brain drain. Nature New Biol 2011 Oct 6;478(7367):15-15. [doi: 10.1038/478015a]
3. Huhn M, Tardy M, Spineli LM, et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. JAMA Psychiatry 2014 Jun;71(6):706-715. [doi: 10.1001/jamapsychiatry.2014.112] [Medline: 24789675]
4. Leichsenring F, Steinert C, Rabung S, Ioannidis JPA. The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. World Psychiatry 2022 Feb;21(1):133-145. [doi: 10.1002/wps.20941] [Medline: 35015359]
5. Steinert C, Hofmann M, Kruse J, Leichsenring F. Relapse rates after psychotherapy for depression – stable long-term effects? A meta-analysis. J Affect Disord 2014 Oct;168:107-118. [doi: 10.1016/j.jad.2014.06.043]
6. Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. Am J Psychiatry 2014 Apr;171(4):395-397. [doi: 10.1176/appi.ajp.2014.14020138] [Medline: 24687194]
7. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 2010 Jul;167(7):748-751. [doi: 10.1176/appi.ajp.2010.09091379] [Medline: 20595427]
8. Aldao A, Nolen-Hoeksema S, Schweizer S. Emotion-regulation strategies across psychopathology: a meta-analytic review. Clin Psychol Rev 2010 Mar;30(2):217-237. [doi: 10.1016/j.cpr.2009.11.004] [Medline: 20015584]
9. Hu T, Zhang D, Wang J, Mistry R, Ran G, Wang X. Relation between emotion regulation and mental health: a meta-analysis review. Psychol Rep 2014 Apr;114(2):341-362. [doi: 10.2466/03.20.PR0.114k22w4] [Medline: 24897894]
10. Sloan E, Hall K, Moulding R, Bryce S, Mildred H, Staiger PK. Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: a systematic review. Clin Psychol Rev 2017 Nov;57:141-163. [doi: 10.1016/j.cpr.2017.09.002] [Medline: 28941927]
11. Gross JJ. The emerging field of emotion regulation: an integrative review. Rev Gen Psychol 1998 Sep;2(3):271-299. [doi: 10.1037/1089-2680.2.3.271]
12. Berking M, Ebert D, Cuijpers P, Hofmann SG. Emotion regulation skills training enhances the efficacy of inpatient cognitive behavioral therapy for major depressive disorder: a randomized controlled trial. Psychother Psychosom 2013;82(4):234-245. [doi: 10.1159/000348448] [Medline: 23712210]
13. Fehlinger T, Stumpenhorst M, Stenzel N, Rief W. Emotion regulation is the essential skill for improving depressive symptoms. J Affect Disord 2013 Jan 10;144(1-2):116-122. [doi: 10.1016/j.jad.2012.06.015] [Medline: 22939389]
14. Kleinstäuber M, Allwang C, Bailer J, et al. Cognitive behaviour therapy complemented with emotion regulation training for patients with persistent physical symptoms: a randomised clinical trial. Psychother Psychosom 2019;88(5):287-299. [doi: 10.1159/000501621] [Medline: 31430755]

XSL•FO
RenderX

15. Mennin DS, Fresco DM, O'Toole MS, Heimberg RG. A randomized controlled trial of emotion regulation therapy for generalized anxiety disorder with and without co-occurring depression. J Consult Clin Psychol 2018 Mar;86(3):268-281. [doi: 10.1037/ccp0000289] [Medline: 29504794]
16. O'Toole MS, Renna ME, Mennin DS, Fresco DM. Changes in decentering and reappraisal temporally precede symptom reduction during emotion regulation therapy for generalized anxiety disorder with and without co-occurring depression. Behav Ther 2019 Nov;50(6):1042-1052. [doi: 10.1016/j.beth.2018.12.005] [Medline: 31735240]
17. Radkovsky A, McArdle JJ, Bockting CLH, Berking M. Successful emotion regulation skills application predicts subsequent reduction of symptom severity during treatment of major depressive disorder. J Consult Clin Psychol 2014 Apr;82(2):248-262. [doi: 10.1037/a0035828] [Medline: 24564219]
18. Berking M, Wupperman P, Reichardt A, Pejic T, Dippel A, Znoj H. Emotion-regulation skills as a treatment target in psychotherapy. Behav Res Ther 2008 Nov;46(11):1230-1237. [doi: 10.1016/j.brat.2008.08.005] [Medline: 18835479]
19. Gratz KL, Weiss NH, Tull MT. Examining emotion regulation as an outcome, mechanism, or target of psychological treatments. Curr Opin Psychol 2015 Jun 1;3:85-90. [doi: 10.1016/j.copsyc.2015.02.010] [Medline: 25859561]
20. Rottenberg J, Gross JJ. Emotion and emotion regulation: a map for psychotherapy researchers. Clin Psychol Sci Pract 2007;14(4):323-328. [doi: 10.1111/j.1468-2850.2007.00093.x]
21. Ochsner K, Gross J. The cognitive control of emotion. Trends Cogn Sci (Regul Ed) 2005 May;9(5):242-249. [doi: 10.1016/j.tics.2005.03.010]
22. Ochsner KN, Silvers JA, Buhle JT. Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. Ann N Y Acad Sci 2012 Mar;1251(1). [doi: 10.1111/j.1749-6632.2012.06751.x]
23. Joormann J, Yoon KL, Zetsche U. Cognitive inhibition in depression. Appl Prev Psychol 2007 Dec;12(3):128-139. [doi: 10.1016/j.appsy.2007.09.002]
24. Joormann J, D'Avanzato C. Emotion regulation in depression: examining the role of cognitive processes. Cogn Emot 2010 Sep;24(6):913-939. [doi: 10.1080/02699931003784939]
25. Pruessner L, Barnow S, Holt DV, Joormann J, Schulze K. A cognitive control framework for understanding emotion regulation flexibility. Emotion 2020 Feb;20(1):21-29. [doi: 10.1037/emo0000658] [Medline: 31961173]
26. Miller EK. The prefrontal cortex and cognitive control. Nat Rev Neurosci 2000 Oct;1(1):59-65. [doi: 10.1038/35036228] [Medline: 11252769]
27. Miyake A, Friedman NP. The nature and organization of individual differences in executive functions: four general conclusions. Curr Dir Psychol Sci 2012 Feb;21(1):8-14. [doi: 10.1177/0963721411429458] [Medline: 22773897]
28. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cogn Psychol 2000 Aug;41(1):49-100. [doi: 10.1006/cogp.1999.0734] [Medline: 10945922]
29. Snyder HR, Miyake A, Hankin BL. Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. Front Psychol 2015;6:328. [doi: 10.3389/fpsyg.2015.00328] [Medline: 25859234]
30. De Lissnyder E, Koster EHW, Derakshan N, De Raedt R. The association between depressive symptoms and executive control impairments in response to emotional and non-emotional information. Cogn Emot 2010 Feb;24(2):264-280. [doi: 10.1080/02699930903378354]
31. De Lissnyder E, Koster EHW, De Raedt R. Emotional interference in working memory is related to rumination. Cogn Ther Res 2012 Aug;36(4):348-357. [doi: 10.1007/s10608-011-9352-4]
32. Joormann J. Cognitive inhibition and emotion regulation in depression. Curr Dir Psychol Sci 2010 Jun;19(3):161-166. [doi: 10.1177/0963721410370293]
33. Whitmer AJ, Gotlib IH. An attentional scope model of rumination. Psychol Bull 2013 Sep;139(5):1036-1061. [doi: 10.1037/a0030923] [Medline: 23244316]
34. Zetsche U, D'Avanzato C, Joormann J. Depression and rumination: relation to components of inhibition. Cogn Emot 2012 Jun;26(4):758-767. [doi: 10.1080/02699931.2011.613919]
35. Koster EHW, De Lissnyder E, Derakshan N, De Raedt R. Understanding depressive rumination from a cognitive science perspective: the impaired disengagement hypothesis. Clin Psychol Rev 2011 Feb;31(1):138-145. [doi: 10.1016/j.cpr.2010.08.005] [Medline: 20817334]
36. Zetsche U, Bürkner PC, Schulze L. Shedding light on the association between repetitive negative thinking and deficits in cognitive control - a meta-analysis. Clin Psychol Rev 2018 Jul;63:56-65. [doi: 10.1016/j.cpr.2018.06.001] [Medline: 29913351]
37. Gronwall DMA. Paced auditory serial-addition task: a measure of recovery from concussion. Percept Mot Skills 1977 Apr;44(2):367-373. [doi: 10.2466/pms.1977.44.2.367]
38. Siegle GJ, Ghinassi F, Thase ME. Neurobehavioral therapies in the 21st century: summary of an emerging field and an extended example of cognitive control training for depression. Cogn Ther Res 2007 Apr;31(2):235-262. [doi: 10.1007/s10608-006-9118-6]

XSL•FO
RenderX

39. Brunoni AR, Boggio PS, De Raedt R, et al. Cognitive control therapy and transcranial direct current stimulation for depression: a randomized, double-blinded, controlled trial. J Affect Disord 2014 Jun;162:43-49. [doi: 10.1016/j.jad.2014.03.026] [Medline: 24767004]
40. Segrave RA, Arnold S, Hoy K, Fitzgerald PB. Concurrent cognitive control training augments the antidepressant efficacy of tDCS: a pilot study. Brain Stimul 2014;7(2):325-331. [doi: 10.1016/j.brs.2013.12.008] [Medline: 24486425]
41. Siegle GJ, Price RB, Jones NP, Ghinassi F, Painter T, Thase ME. You gotta work at it: pupillary indices of task focus are prognostic for response to a neurocognitive intervention for rumination in depression. Clin Psychol Sci 2014 Jul;2(4):455-471. [doi: 10.1177/2167702614536160]
42. Sommer A, Plewnia C. Depression treatment by tDCS-enhanced cognitive control training: a test of two stimulation intensities. Brain Stimul 2021;14(4):987-989. [doi: 10.1016/j.brs.2021.06.004] [Medline: 34147374]
43. Vanderhasselt MA, De Raedt R, Namur V, et al. Transcranial electric stimulation and neurocognitive training in clinically depressed patients: a pilot study of the effects on rumination. Prog Neuropsychopharmacol Biol Psychiatry 2015 Mar 3;57:93-99. [doi: 10.1016/j.pnpbp.2014.09.015] [Medline: 25455589]
44. Calkins AW, McMorran KE, Siegle GJ, Otto MW. The effects of computerized cognitive control training on community adults with depressed mood. Behav Cogn Psychother 2015 Sep;43(5):578-589. [doi: 10.1017/S1352465814000046] [Medline: 24589123]
45. Hoorelbeke K, Koster EHW. Internet-delivered cognitive control training as a preventive intervention for remitted depressed patients: evidence from a double-blind randomized controlled trial study. J Consult Clin Psychol 2017 Feb;85(2):135-146. [doi: 10.1037/ccp0000128] [Medline: 27362792]
46. Hoorelbeke K, Van den Bergh N, De Raedt R, Wichers M, Koster EHW. Preventing recurrence of depression: long-term effects of a randomized controlled trial on cognitive control training for remitted depressed patients. Clin Psychol Sci 2021 Jul;9(4):615-633. [doi: 10.1177/2167702620979775]
47. Hoorelbeke K, Van den Bergh N, De Raedt R, Wichers M, Albers CJ, Koster EHW. Regaining control of your emotions? Investigating the mechanisms underlying effects of cognitive control training for remitted depressed patients. Emotion 2023 Feb;23(1):194-213. [doi: 10.1037/emo0001067]
48. Vander Zwalmen Y, Liebaert E, Hoorelbeke K, et al. Treatment response following adaptive PASAT training for depression vulnerability: a systematic review and meta-analysis. Neuropsychol Rev 2024 Mar;34(1):232-249. [doi: 10.1007/s11065-023-09581-8] [Medline: 36853526]
49. Hoorelbeke K, Koster EHW, Demeyer I, Loeys T, Vanderhasselt MA. Effects of cognitive control training on the dynamics of (mal)adaptive emotion regulation in daily life. Emotion 2016 Oct;16(7):945-956. [doi: 10.1037/emo0000169]
50. Van den Bergh N, Vermeersch S, Hoorelbeke K, Vervaeke J, De Raedt R, Koster EHW. Cognitive control training as an augmentation strategy to CBT in the treatment of fear of failure in undergraduates. Cogn Ther Res 2020 Dec;44(6):1199-1212. [doi: 10.1007/s10608-020-10129-w]
51. Vervaeke J, Hoorelbeke K, Baeken C, Van Looy J, Koster EHW. Transfer and motivation after cognitive control training for remitted depression in healthy sample. J Cogn Enhanc 2020 Mar;4(1):49-61. [doi: 10.1007/s41465-019-00135-6]
52. Vanderhasselt MA, Demeyer I, Van Imschoot L, Hoorelbeke K, De Raedt R. Cognitive control training in healthy older adults: a proof of concept study on the effects on cognitive functioning, emotion regulation and affect. Cogn Ther Res 2021 Oct;45(5):959-968. [doi: 10.1007/s10608-020-10154-9]
53. Hoorelbeke K, Koster EHW, Vanderhasselt MA, Callewaert S, Demeyer I. The influence of cognitive control training on stress reactivity and rumination in response to a lab stressor and naturalistic stress. Behav Res Ther 2015 Jun;69:1-10. [doi: 10.1016/j.brat.2015.03.010] [Medline: 25841177]
54. Moshier SJ, Otto MW. Behavioral activation treatment for major depression: a randomized trial of the efficacy of augmentation with cognitive control training. J Affect Disord 2017 Mar 1;210:265-268. [doi: 10.1016/j.jad.2017.01.003] [Medline: 28068613]
55. Ferrari GRA, Vanderhasselt MA, Rinck M, et al. A cognitive control training as add-on treatment to usual care for depressed inpatients. Cogn Ther Res 2021 Oct;45(5):929-943. [doi: 10.1007/s10608-020-10197-y]
56. Aldao A, Nolen-Hoeksema S. Specificity of cognitive emotion regulation strategies: a transdiagnostic examination. Behav Res Ther 2010 Oct;48(10):974-983. [doi: 10.1016/j.brat.2010.06.002] [Medline: 20591413]
57. Revelle W. Psych: procedures for psychological, psychometric, and personality research. Comprehensive R Archive Network (CRAN). 2015. URL: https://doi.org/10.32614/CRAN.package.psych [accessed 2025-11-16]
58. Arslan RC, Walther MP, Tata CS. formr: a study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. Behav Res Methods 2020 Feb;52(1):376-387. [doi: 10.3758/s13428-019-01236-y] [Medline: 30937847]
59. de Leeuw JR. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. Behav Res Methods 2015 Mar;47(1):1-12. [doi: 10.3758/s13428-014-0458-y] [Medline: 24683129]
60. Lange K, Kühn S, Filevich E. "Just Another Tool for Online Studies" (JATOS): an easy solution for setup and management of web servers supporting online studies. PLoS One 2015;10(6):e0130834. [doi: 10.1371/journal.pone.0130834] [Medline: 26114751]

XSL•FO
RenderX

61. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Association; 2013.
62. Margraf J, Cwik JC, von BR, Suppiger A, Schneider S. DIPS Open Access 12: Diagnostisches Interview Bei Psychischen Störungen: OMP Ruhr-Universität Bochum; 2022.
63. First MB. Structured clinical interview for the DSM (SCID). In: The Encyclopedia of Clinical Psychology: John Wiley & Sons, Ltd; 2015:1-6. [doi: 10.1002/9781118625392.wbecp351]
64. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
65. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073. [doi: 10.1177/0962280215588241] [Medline: 26092476]
66. Au J, Sheehan E, Tsai N, Duncan GJ, Buschkuehl M, Jaeggi SM. Improving fluid intelligence with training on working memory: a meta-analysis. Psychon Bull Rev 2015 Apr;22(2):366-377. [doi: 10.3758/s13423-014-0699-x]
67. Transdiagnostic cognitive control training. OSF. URL: https://osf.io/5cah4/ [accessed 2025-11-09]
68. Lutz W, Tholen S, Schürch E, Berking M. The development, validation, and reliability of short-forms of current instruments for the evaluation of therapeutic progress in psychotherapy and psychiatry.. Diagnostica 2006 Jan;52(1):11-25. [doi: 10.1026/0012-1924.52.1.11]
69. Treynor W, Gonzalez R, Nolen-Hoeksema S. Rumination reconsidered: a psychometric analysis. Cognit Ther Res 2003 Jun;27(3):247-259. [doi: 10.1023/A:1023910315561]
70. Huffziger S, Kühner C. The rumination components broading and reflection. Zeitschrift für Klinische Psychologie und Psychotherapie 2012 Jan;41(1):38-46. [doi: 10.1026/1616-3443/a000118]
71. Watkins ER. Constructive and unconstructive repetitive thought. Psychol Bull 2008 Mar;134(2):163-206. [doi: 10.1037/0033-2909.134.2.163] [Medline: 18298268]
72. Ehring T, Watkins ER. Repetitive negative thinking as a transdiagnostic process. Int J Cogn Ther 2008 Sep;1(3):192-205. [doi: 10.1521/ijct.2008.1.3.192]
73. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. J Behav Ther Exp Psychiatry 2011 Jun;42(2):225-232. [doi: 10.1016/j.jbtep.2010.12.003]
74. Bjureberg J, Ljótsson B, Tull MT, et al. Development and Validation of a Brief Version of the Difficulties in Emotion Regulation Scale: the DERS-16. J Psychopathol Behav Assess 2016 Jun;38(2):284-296. [doi: 10.1007/s10862-015-9514-x] [Medline: 27239096]
75. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. J Psychopathol Behav Assess 2004 Mar;26(1):41-54. [doi: 10.1023/B:JOBA.0000007455.08539.94]
76. Garnefski N, Kraaij V, Spinhoven P. Negative life events, cognitive emotion regulation and emotional problems. Pers Individ Dif 2001 Jun;30(8):1311-1327. [doi: 10.1016/S0191-8869(00)00113-6]
77. Loch N, Hiller W, Witthöft M. The Cognitive Emotion Regulation Questionnaire (CERQ): psychometric evaluation of a German adaptation. Zeitschrift für Klinische Psychologie und Psychotherapie 2011 Apr;40(2):94-106. [doi: 10.1026/1616-3443/a000079]
78. Skevington SM, Lotfy M, O'Connell KA, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004 Mar;13(2):299-310. [doi: 10.1023/B:QURE.0000018486.91360.00] [Medline: 15085902]
79. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2002 May;180(5):461-464. [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]
80. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
81. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing Generalized Anxiety Disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
82. Wiltink J, Kliem S, Michal M, et al. Mini - Social Phobia Inventory (mini-SPIN): psychometric properties and population based norms of the German version. BMC Psychiatry 2017 Nov 25;17(1):377. [doi: 10.1186/s12888-017-1545-2] [Medline: 29178857]
83. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4] [Medline: 11132119]
84. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang 2008 Nov;59(4):390-412. [doi: 10.1016/j.jml.2007.12.005]
85. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67:1-48. [doi: 10.18637/jss.v067.i01]
86. Team RC. R: a language and environment for statistical computing. : R Foundation for Statistical Computing; 2020.

XSL•FO
RenderX

87. Forstmeier W, Schielzeth H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. Behav Ecol Sociobiol 2011 Jan;65(1):47-55. [doi: 10.1007/s00265-010-1038-5]
88. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Soft 2017 Dec;82(13):1-26. [doi: 10.18637/jss.v082.i13]
89. Luke SG. Evaluating significance in linear mixed-effects models in R. Behav Res Methods 2017 Aug;49(4):1494-1502. [doi: 10.3758/s13428-016-0809-y] [Medline: 27620283]
90. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009 Mar;14(1):43-53. [doi: 10.1037/a0014699] [Medline: 19271847]
91. Davison AC, Hinkley DV. Bootstrap Methods and Their Application: Cambridge University Press; 1997.
92. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol Methods 1997 Mar;2(1):64-78. [doi: 10.1037/1082-989X.2.1.64]
93. Campbell KL, Grady CL, Ng C, Hasher L. Age differences in the frontoparietal cognitive control network: implications for distractibility. Neuropsychologia 2012 Jul;50(9):2212-2223. [doi: 10.1016/j.neuropsychologia.2012.05.025] [Medline: 22659108]
94. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037//0022-006X.59.1.12]
95. International clinical trials registry platform. World Health Organization. URL: https://trialsearch.who.int/Trial2.aspx?TrialID=DRKS00024408 [accessed 2025-11-09]
96. Hoorelbeke K, Vervaeke J, Siegle GJ, Baeken C, Koster EHW. Individual differences associated with treatment adherence and transfer effects following gamified web-based cognitive control training for repetitive negative thinking. Internet Interv 2022 Mar;27:100507. [doi: 10.1016/j.invent.2022.100507] [Medline: 35242588]
97. Koster EHW, Hoorelbeke K, Onraedt T, Owens M, Derakshan N. Cognitive control interventions for depression: a systematic review of findings from training studies. Clin Psychol Rev 2017 Apr;53:79-92. [doi: 10.1016/j.cpr.2017.02.002] [Medline: 28273486]
98. Lejuez CW, Kahler CW, Brown RA. A modified computer version of the Paced Auditory Serial Addition Task (PASAT) as a laboratory-based stressor. Behav Ther 2003;26(4):290-293 [FREE Full text]
99. Tombaugh TN. A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). Arch Clin Neuropsychol 2006 Jan;21(1):53-76. [doi: 10.1016/j.acn.2005.07.006] [Medline: 16290063]
100. Vander Zwalmen Y, Demeester D, Hoorelbeke K, Verhaeghe N, Baeken C, Koster EHW. The more, the merrier? Establishing a dose-response relationship for the effects of cognitive control training on depressive symptomatology. J Consult Clin Psychol 2025 Mar;93(3):161-175. [doi: 10.1037/ccp0000945] [Medline: 40014506]
101. Dawson JF, Richter AW. Probing three-way interactions in moderated multiple regression: development and application of a slope difference test. J Appl Psychol 2006 Jul;91(4):917-926. [doi: 10.1037/0021-9010.91.4.917] [Medline: 16834514]
102. Michalak J, Hölz A, Teismann T. Rumination as a predictor of relapse in mindfulness‐based cognitive therapy for depression. Psychol Psychother 2011 Jun;84(2):230-236. [doi: 10.1348/147608310X520166]
103. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8]
104. Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. Psychol Sci 2011 Apr;22(4):490-497. [doi: 10.1177/0956797611400615] [Medline: 21389338]
105. Mohlman J, Gorman JM. The role of executive functioning in CBT: a pilot study with anxious older adults. Behav Res Ther 2005 Apr;43(4):447-465. [doi: 10.1016/j.brat.2004.03.007]
106. Kessler RC, Ormel J, Petukhova M, et al. Development of lifetime comorbidity in the World Health Organization world mental health surveys. Arch Gen Psychiatry 2011 Jan;68(1):90-100. [doi: 10.1001/archgenpsychiatry.2010.180] [Medline: 21199968]
107. Gräfe K, Zipfel S, Herzog W, Löwe B. Screening for psychiatric disorders with the Patient Health Questionnaire (PHQ-D)". Diagnostica 2004 Oct;50(4):171-181. [doi: 10.1026/0012-1924.50.4.171]
108. Anderson EM, Lambert MJ. A survival analysis of clinically significant change in outpatient psychotherapy. J Clin Psychol 2001 Jul;57(7):875-888. [doi: 10.1002/jclp.1056] [Medline: 11406801]

Edited by L Buis; submitted 28.08.24; peer-reviewed by U Zetsche, YC Kato-Lin; accepted 22.10.25; published 26.11.25.

Please cite as:
Blomberg M, Oberender L, Koster E, Brockmeyer T
Transdiagnostic Cognitive Control Training for Patients Waiting for Outpatient Psychotherapy: Randomized Clinical Trial
JMIR Mhealth Uhealth 2025;13:e65867
URL: https://mhealth.jmir.org/2025/1/e65867
doi:10.2196/65867

XSL•FO
RenderX

© Maximilian Blomberg, Lisa Oberender, Ernst Koster, Timo Brockmeyer. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone-Based Approach-Avoidance Bias Modification Training for Depression: Randomized Clinical Trial

Maximilian Blomberg[1,2], PhD; Hilmar Gero Zech[3,4], PhD; Maximilian Kluge[2], MSc; Nico Böhmert[2], MSc; Helmut Platte[5], MD; Timo Brockmeyer[1,2], Prof Dr

[1]Department of Clinical Psychology and Translational Psychotherapy, Institute of Psychology, University of Münster, Fliednerstr. 21, Münster, Germany

[2]Department of Clinical Psychology and Psychotherapy, Institute of Psychology, University of Goettingen, Göttingen, Germany

[3]Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Dresden, Germany

[4]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Centre of Mental Health, University of Würzburg, Würzburg, Germany

[5]Paracelsus Roswithaklinik Bad Gandersheim, Bad Gandersheim, Germany

Corresponding Author:
Maximilian Blomberg, PhD
Department of Clinical Psychology and Translational Psychotherapy, Institute of Psychology, University of Münster, Fliednerstr. 21, Münster, Germany

Abstract

Background: Effective treatments for depression are available, yet many patients do not respond to treatment or experience relapse. Cognitive bias modification aims to ameliorate cognitive biases that contribute to the development and maintenance of the disorder.

Objective: This study examines the efficacy of a novel mobile approach-avoidance bias modification training with socioemotional cues for depression.

Methods: In this randomized clinical superiority trial, 75 inpatients with depression underwent 6 sessions of either active or sham approach-avoidance bias modification training with socioemotional cues over the course of 2 weeks alongside inpatient treatment as usual. The primary outcome was self-reported depressive symptoms, and the secondary outcomes included approach-avoidance bias based on reaction time and response force, anhedonia, and positivity. Outcomes were assessed before and after the training, and again at 2-week and 6-month follow-ups. The primary hypothesis was that active training would lead to a stronger decrease in symptoms of depression at the end of training.

Results: Both groups improved in depressive symptoms from baseline to the end-of-training assessment but did not differ in this regard (B=−1.14, 95% CI −5.65-3.41; $t_{188.61}$=−0.47; P=.64; d=−0.09, 95% CI −0.46-0.28). Changes in anhedonia, positivity, and approach-avoidance bias were also not different between training groups, neither at end of training (P=.16) nor at the 2-week follow-up (P=.69). However, the active training group showed a significantly greater reduction in depressive symptoms from the baseline assessment to the 6-month follow-up (B=7.26, 95% CI 2.53-11.93; $t_{190.54}$=2.95; P=.004; d=0.58, 95% CI 0.20-0.96). Permuted split-half reliability of the mobile assessment of approach-avoidance bias ranged from 0.77 to 0.94 for reaction times and from 0.81 to 0.93 for response force. Approach-avoidance bias was not altered by the training and did not mediate the training effects.

Conclusions: Mobile approach-avoidance bias modification training with socioemotional cues did not reduce depressive symptoms in the short term but did in the long term. Mobile training and assessment versions may be more feasible in the future, as they require no joystick setup and can be conducted on patients' smartphones. Future work needs to further examine short- and long-term efficacy and the mechanisms driving long-term symptom change in larger multicenter trials.

Trial Registration: German Clinical Trials Register DRKS00022447; https://drks.de/search/en/trial/DRKS00022447

(JMIR Mhealth Uhealth 2025;13:e69033) doi:10.2196/69033

KEYWORDS

affective disorders; approach bias; cognitive bias modification; eHealth; mHealth; smartphone intervention; mobile health

XSL•FO
RenderX

Introduction

Major depressive disorder (MDD) is among the most prevalent mental disorders and a leading contributor to global disability [1] with patients facing a higher mortality rate [2], suffering from numerous somatic sequelae [3] and considerable psychosocial and socioeconomic challenges [4]. Besides medication, cognitive behavioral therapy (CBT) is considered the most well-established and empirically supported effective treatment for depression. However, as indicated by a comprehensive meta-analysis, CBT still yields only a 42% response rate and 36% remission following treatment [5].

The limited efficacy of treatment may in part be due to persisting cognitive biases that contribute to the development and maintenance of depression [6,7]. Cognitive biases are systematic biases in information processing [7], with memory [8], attention [9], and interpretation biases [10] most frequently studied. Furthermore, approach-avoidance biases describe automatic action tendencies following the evaluation of stimuli as either positive or negative [11,12]. Importantly, depression is characterized by reduced approach motivation toward positive cues [13]. For instance, contrary to healthy controls, patients show less approach behavior toward positive social-emotional cues (eg, smiling faces) and less avoidance of negative cues (eg, angry faces) [14].

Cognitive bias modification (CBM) aims to alter biased information processing by implementing contingencies into tasks traditionally used to assess these biases. For instance, approach-avoidance biases are often measured with a joystick task, where participants push or pull stimuli away or toward themselves [15]. To modify approach-avoidance tendencies, participants engage in a learning paradigm to consistently show approach or avoidance behavior in response to certain stimuli. In alcohol use disorder, patients repeatedly perform avoidance movements in response to alcohol-related cues and approach movements to nonalcoholic cues [16]. This active training is often compared to a sham version in which positive and negative stimuli are equally often approached and avoided [17].

Although approach-avoidance bias modification has shown promise in disorders such as addiction [18], eating disorders [19], and anxiety [20,21], evidence in depression remains limited. At least 2 randomized controlled trials have examined multisession training in inpatients with depressive symptoms. Vrijsen et al [22] reported that 4 biweekly sessions reduced depressive symptoms in patients with MDD compared to sham training, although the targeted bias was unchanged. Becker et al [23] found comparable effects in a mixed inpatient sample, but benefits were restricted to those with moderate-to-severe depression. In both studies, active training paired approach movements with positive stimuli and avoidance with neutral ones, while in the sham version, no systematic pairing of movements and stimuli was trained. Additionally, a small uncontrolled study in patients with depression found that combining 4 sessions of approach-avoidance training with behavioral activation was associated with improvements in depressive and anxiety symptoms as well as social connectedness [24]. Finally, a neuroimaging study using a single training session reported increased activation in reward-related regions (medial prefrontal cortex, caudate) during social reward anticipation, despite no change in bias [25].

All of these studies used a joystick task but varied in stimulus selection and training design. The first 2 studies utilized a broad spectrum of stimuli (eg, animals, people, objects) taken from the *International Affective Picture System* [26] with the aim to target general approach-avoidance biases [22,23], whereas the latter 2 focused on stimuli with positive and negative emotional facial expressions [24,25]. This emphasis on faces may be especially relevant, as depression is associated with altered approach-avoidance behavior toward emotional expressions [14] and less neural activation in reward circuitry during approach to social cues [27]. Training protocols also differed in the affective variety of stimuli included. In the first 2 studies, participants approached positive and avoided neutral stimuli [22,23], while the latter 2 combined an approach toward positive stimuli with lateral movements for neutral ones [24,25]. Evidence on working mechanisms of approach-avoidance training suggests that including avoidance of negative stimuli may further reduce dysfunctional biases [28]. Thus, paradigms integrating both approaches to positive and avoidance of negative stimuli may prove more effective than those focusing on only one of these.

Taken together, existing studies suggest that approach-avoidance training may serve as a useful add-on treatment for depression. However, evidence is currently limited by small samples [24,25]; missing control groups [24]; absent follow-ups [22,23,25]; mixed clinical samples [23]; single-session designs [25]; and issues of attrition, data loss, and variable intervals between baseline and end-of-training assessment [23].

This study seeks to evaluate a novel smartphone-based intervention derived from the *Mobile Approach-Avoidance Task* (mobile AAT) [29]. In this task, approach behavior is mapped as moving the smartphone (displaying the stimuli) toward the body. Correspondingly, avoidance behavior is mapped as moving the smartphone or the stimuli away from one's own body. Compared to traditional joystick setups, the mobile format offers greater ecological validity, as it does not rely on zooming effects to simulate changes in physical distance to a stimulus [15] and as it allows training to be conducted flexibly in daily life. Furthermore, assessments of the mobile AAT have demonstrated adequate reliability of measurement using face stimuli [29,30]. In addition to reaction times, the mobile AAT also records response force in hand or arm movements. Reaction time and response forced approach-avoidance bias may capture distinct processes, as they were uncorrelated in earlier studies [29]. Given the absence of mediation effects for reaction time measures in earlier trials [22,23], exploring alternative indices such as response force may provide new insights into underlying mechanisms of approach-avoidance training.

Aside from changes in approach-avoidance tendencies, few studies have examined mediators that explain why depression symptoms may decrease after training. One candidate is anhedonia, a core feature of depression characterized by reduced motivation and diminished capacity to experience pleasure. Anhedonia predicts psychosocial functioning, depression

XSL•FO
RenderX

severity, and treatment response [31,32] and has been proposed as a potential target of change in approach-avoidance bias modification [23]. Notably, it is associated with reduced approach motivation toward socioemotional cues [33]. We will therefore examine the effects of approach-avoidance training on anhedonia and its potential mediating role in symptom reduction. Conversely, positivity reflects a general tendency to view life, oneself, experiences, and the future positively. It is inversely related to depression [34,35] and aligns with behavioral activation approaches in social contexts [36]. Because depression is marked by reduced approach toward positive socioemotional cues [14], and the training targets this approach behavior, we will also examine its effects on positivity and its potential mediating role in symptom reduction.

The primary hypothesis in this study was that, in addition to CBT-based inpatient treatment as usual (TAU), 6 sessions of active approach-avoidance bias modification training would lead to a greater reduction in symptoms of depression than a sham training from baseline to end-of-training assessment. Additionally, we expected the active training group to be superior in symptom reduction from baseline to 2-week and 6-month follow-ups. Regarding secondary outcomes, we expected a stronger increase in approach-avoidance bias (ie, stronger approach of positive and stronger avoidance of negative stimuli) and positivity as well as a greater decrease in anhedonia in the active training group. Furthermore, we will test if a change in approach-avoidance bias from baseline to the end of training mediates the reduction in depressive symptoms. As exploratory aims, we examined the mediating roles of anhedonia and positivity and assessed whether training outcomes would be moderated by baseline characteristics.

Methods

Ethical Considerations

This monocenter randomized controlled superiority trial with 2 parallel arms was approved by the ethics committee of the Georg-Elias-Müller-Institute, University of Göttingen (#205) and registered at the German Clinical Trials Register (DRKS00022447). The study procedures adhered to the ethical standards of the institutional and national research committees and to the principles of the Declaration of Helsinki. All participants provided written informed consent and received financial compensation (€40 [US $45]) for their participation. All data were deidentified prior to analysis to ensure participant privacy and confidentiality.

Design

Participants were randomly assigned (1:1 ratio) to receive either active or sham approach-avoidance training as an adjunct to inpatient TAU. Randomization was carried out with the *psych* package, version 2.2.9 [37]. In both conditions, participants were instructed that during the training, they would see facial expressions of varying emotions and would need to respond by either pulling the smartphone toward them or pushing it away. They were also told that the study would compare different training conditions and that they would receive more specific instructions on their respective condition after randomization. Depending on their allocated condition, one of two keywords ("depression1" or "depression2") was entered by participants into the mobile AAT app, which was set up together with study personnel after screening for study eligibility. Participants were blinded to the keyword-condition mapping. However, the study personnel were not blinded to the keyword-condition mapping. All outcomes were either self-report measures or computerized tasks that were conducted independently and remotely by the patients without contact with the study personnel. Baseline, end-of-training, and 2-week follow-up data were collected through the mobile AAT app. Six-month follow-up assessments were conducted through the online survey framework *formr* [38]. Follow-up periods commenced after the end-of-training assessment. Patients were debriefed after the 6-month follow-up.

Participants

Between October 2020 and December 2021, patients were recruited from the Paracelsus Roswithaklinik Bad Gandersheim, Germany, an inpatient rehabilitation center for mental and psychosomatic disorders with a focus on CBT. We collected data from a total of 75 participants. A total of 38 (51%) patients were randomly assigned to the active training and 37 (49%) to the sham training. Patients were initially made aware of the study with the help of leaflets, information events, and by their therapists at the end of group therapy sessions. Patients were eligible for the study if they were 18 years or older and met the diagnostic criteria for either (recurring) MDD, persistent depressive disorder, or both according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) [39]. Study assessors screened patients using the *Structured Clinical Interview for DSM-5 Disorders* [40]. To maximize external validity, exclusion criteria were kept to a minimum of a remaining therapy duration of less than 14 days (to complete the training in addition to TAU), substance use disorder (nonabstinent), severe psychiatric comorbidity (eg, schizophrenia, bipolar disorder), medical instability, or acute suicidality. Data on ethnicity and socioeconomic status were not collected following principles of purpose limitation to the research questions and data minimization according to article 5 of the *General Data Protection Regulation* [41]. However, the vast majority of patients were White and stemmed from a Western European country.

Sample size was determined by an a priori power analysis using *G*Power* [42]. The study was powered for the primary hypothesis that the active training would lead to a stronger reduction in symptoms of depression from baseline to end-of-training assessment compared to the sham training. Based on an earlier study with a similar sample and a similar intervention [22], we expected a moderate within-between (time × group) interaction effect of f=.25 in a repeated-measures design, with α=.05 and statistical power 1-β=.95. While accounting for an approximate 20% of data loss (eg, due to technical errors), a minimum sample size of 65 would have been necessary, which was conservatively rounded up to 75, following recommendations for pilot trials [43].

XSL•FO
RenderX

Treatments

Treatment as Usual

All participants received multimodal CBT-based TAU in an inpatient setting for a duration of 6 weeks. TAU comprised typical elements of inpatient psychotherapy such as group psychotherapy, individual psychotherapy, and involvement of caregivers as well as movement, occupational, and art therapy. Within these elements, CBT-oriented interventions were applied, such as psychoeducation, cognitive restructuring, behavioral activation, social skills and emotion regulation training, and relapse prevention. Testing the additive efficacy of CBM adjunct to TAU is a common design in this field of research [16,22,23,44].

Active Approach-Avoidance Training

Participants in the active training group received 6 sessions of an approach-avoidance training (20 min each) using an adapted version of the mobile AAT [29]. During the training, pictures of angry and happy faces taken from the *Radboud Faces Database* [45] appeared on the screen of the participants' smartphones. A small cross or circle (approximately on the forehead) was superimposed on the face images. However, the superimposing was small enough so that the portrayed emotion could not be overlooked (see the study's OSF repository [46] for an example). Participants were instructed to always pull the smartphone toward them if an image contained a cross and to always push their smartphone in response to images with circles. Each stimulus was preceded by a fixation cross displayed for 1.5 seconds [29]. Images were then presented for 2 seconds or until a response occurred.

In the active training group, all pictures of happy faces were paired with crosses and all angry faces with circles. Thus, participants learned to consistently approach happy faces and avoid angry ones (ie, to show valence-congruent movements). Each training session involved 40 images and 200 trials (10 female and 10 male faces, both in their respective happy and angry version, each repeated 5 times). Sessions were subdivided into 10 sets of 20 trials each. Before each training, participants completed 8 practice trials with different images to get accustomed to the movements and instructions.

Sham Approach-Avoidance Training

Participants in the sham training group received the same dosage and duration of training as participants in the active training group. The only difference was that in the sham training, the superimposing of crosses and circles on images was balanced, leading to 50% of trials demanding valence-congruent movements (approach happy or avoid angry faces) and the other half demanding valence-incongruent movements (approach angry or avoid happy faces).

Training Schedule

In both training conditions, participants were instructed to complete all sessions in the mobile AAT app (including baseline assessment, 6 training sessions, and end-of-training assessment) over the course of 2 weeks, with completing at most 1 session per day. We monitored adherence by regularly checking the completion of training sessions remotely and phoning participants in case of strong deviations from the training schedule (eg, multiple days without training session completion).

Outcomes

Primary Outcome: Symptoms of Depression

Depressive symptoms were measured at baseline, the end of training, and 2-week and 6-month follow-ups by the *Beck Depression Inventory Second Edition* (BDI-II) [47], a very commonly used self-report questionnaire with good psychometric properties [48]. It comprises 21 items that assess the presence and severity of depressive symptoms over the last 2 weeks. Responses are given on a 4-point scale. A total score of 0 - 13 represents minimal, 14 - 19 mild, 20 - 28 moderate, and 29 - 63 severe depression. Values for Cronbach α at each assessment ranged from 0.90 to 0.94.

Secondary Outcomes

Approach-Avoidance Bias

Approach-avoidance bias was measured with the mobile AAT [29] at baseline, the end of training, and the 2-week follow-up. The source code of the mobile AAT is available on OSF [49]. For each assessment, trials were subdivided into 4 cells, corresponding to the combinations of the movement direction (approach vs avoid) and stimulus valence (happy vs angry faces). Following previous studies [22,23], scores for approach-avoidance bias were calculated as a compound measure by subtracting the median reaction times for valence-congruent movements from incongruent ones. According to Kahveci et al [50], this type of bias score should be called the double median difference score:

$$Bias(RT) = (avoid\,positive + approach\,negative) - (approach\,positive + avoid\,negative)$$

Thus, a positive score indicates a relatively faster approach behavior toward happy faces and faster avoidance behavior in response to angry faces. Details on preprocessing the mobile AAT data can be found in Multimedia Appendix 1. In calculating the double median difference scores for response force, the equation above was changed to ensure that higher scores mirror a more pronounced reaction. This was achieved by subtracting the force of valence-incongruent movements from that of valence-congruent movements:

$$Bias(Force) = (approach\,positive + avoid\,negative) - (avoid\,positive + approach\,negative)$$

Consequently, positive values indicate stronger responses (ie, more energetic arm movements) consistent with a "healthy" approach in response to happy faces and stronger avoidance movements in response to angry faces. In each assessment session, 60 pictures (15 female, 15 male faces, both in an angry and a happy version) from the *Radboud Faces Database* [45] were used. Assessment sessions consisted of 120 trials, divided into 2 blocks. In 1 block, participants were asked to perform valence-congruent movements (approach happy or avoid angry faces), while in the other block, they had to perform incongruent movements (approach angry or avoid happy faces). The order of blocks was randomized per participant and kept constant for the baseline and end-of-training assessments to prevent skewed bias measure changes due to habituation effects in the active training group. Thus, the assessment version of the AAT followed an explicit or feature-relevant design, instructing

XSL•FO
RenderX

participants to approach or avoid faces depending on which emotion they express. Conversely, the training version of the AAT followed an implicit or feature-irrelevant design, instructing participants to show approach or avoidance movements depending on what symbol was superimposed on the face images.

Anhedonia

The *Dimensional Anhedonia Rating Scale* (DARS-17) [51] was used to assess anhedonia at baseline and end of training. The DARS-17 is a self-report measure consisting of 4 subscales measuring the enjoyment of pastimes and hobbies, food and drink, social activities, and sensory experiences. Each of the 17 items is rated on a 5-point Likert scale. Lower scores indicate stronger anhedonia. The reliability and validity of the German version have been demonstrated. Additionally, stronger anhedonia correlates with more severe symptoms of depression [51]. Values for Cronbach α at each assessment ranged from 0.86 to 0.94.

Positivity

Positivity was assessed by the *Positivity Scale* (P Scale) [34] at baseline and the end of training. The P Scale is an 8-item self-report questionnaire that measures the tendency to view oneself, life, and the future with a positive outlook. Items (eg, *I have great faith in the future*) are scored on a 5-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The total sum score correlates negatively with depression [34]. Values for Cronbach α at each assessment ranged from 0.85 to 0.90.

Credibility

To check how credible participants rated the training, we asked them at the 2-week follow-up to indicate on a 5-point Likert scale how efficacious they felt the intervention was, ranging from 1 (*very efficacious*) to 5 (*very inefficacious*), and how confident they were in their judgment, ranging from 1 (*very sure*) to 5 (*very unsure*).

Statistical Analysis

Primary analysis followed an intention-to-treat approach. Outliers ($|z|>3$) were winsorized. Linear mixed-effects models were fitted with random intercepts for subjects and fixed effects of time, training condition, and their interaction [52]. Time was considered a repeated measures factor with either 4 (BDI-II), 3 (approach-avoidance bias), or 2 (DARS-17, P Scale) levels.

Interaction effects were tested using full versus reduced model comparisons [53], followed by post hoc comparisons per time interval (ie, baseline to end-of-training assessment, baseline to 2-wk follow-up, and baseline to 6-mo follow-up). Fixed effects were evaluated using restricted maximum likelihood and Satterthwaite approximation [54]. Effect sizes were calculated by dividing beta coefficients by the end-of-training SD in the sham training group [55]. Nonparametric bootstrapped 95% CIs were reported. Missing data were assessed using a formal test for missing completely at random and pattern-mixture models including missingness indicators [56].

To test potential moderators of training effects, we examined whether baseline outcomes influenced the interaction between time and training group. Specifically, we compared reduced models including time, training group, and their 2-way interactions with baseline outcomes with full models that additionally included the 3-way interaction between baseline outcome, training condition, and time. Additionally, we explored whether the effect of training condition on depressive symptoms at end-of-training, 2-week, and 6-month follow-ups was mediated by approach-avoidance bias, anhedonia, or positivity using path analysis. Baseline scores of the mediator and outcome were included as covariates [57]. Details on model specifications and used packages can be found in Multimedia Appendix 1.

Reaction times and response force were also analyzed at the trial level with maximal random effects structures [58]. Additionally, we conducted a complete case analysis and assessed permuted split-half reliability of the bias scores [59]. Lastly, we calculated proportions of clinically significant changes from the baseline assessment to the 6-month follow-up in the primary outcome between the groups (improvement in BDI-II scores ≥8 and final BDI-II at the 6-mo follow-up <19) [47,60].

Results

Sample Characteristics

Sample characteristics across conditions can be seen in Table 1. The training groups did not differ in any of these sample characteristics (P=.10).

XSL•FO
RenderX

Table . Sample characteristics at baseline (N=75).

Patient characteristic	Active training (n=38)	Sham training (n=37)	Total sample
Age (y), mean (SD)	49 (8.7)	48.8 (10.9)	48.9 (9.8)
Gender (women), n (%)	24 (63)	25 (68)	49 (65)
MDD[a] (diagnosis), n (%)	32 (84)	32 (86)	64 (85)
Recurrent MDD	11 (29)	13 (35)	24 (32)
Double depression[b]	17 (45)	11 (30)	28 (37)
Persistent depressive disorder (diagnosis), n (%)	6 (16)	5 (14)	11 (15)
Psychotropic medication, n (%)	32 (84)	24 (65)	56 (75)
Inpatient treatment duration (d), mean (SD)	14.7 (5)	15.5 (8.1)	15.1 (6.7)

[a]MDD: major depressive disorder.

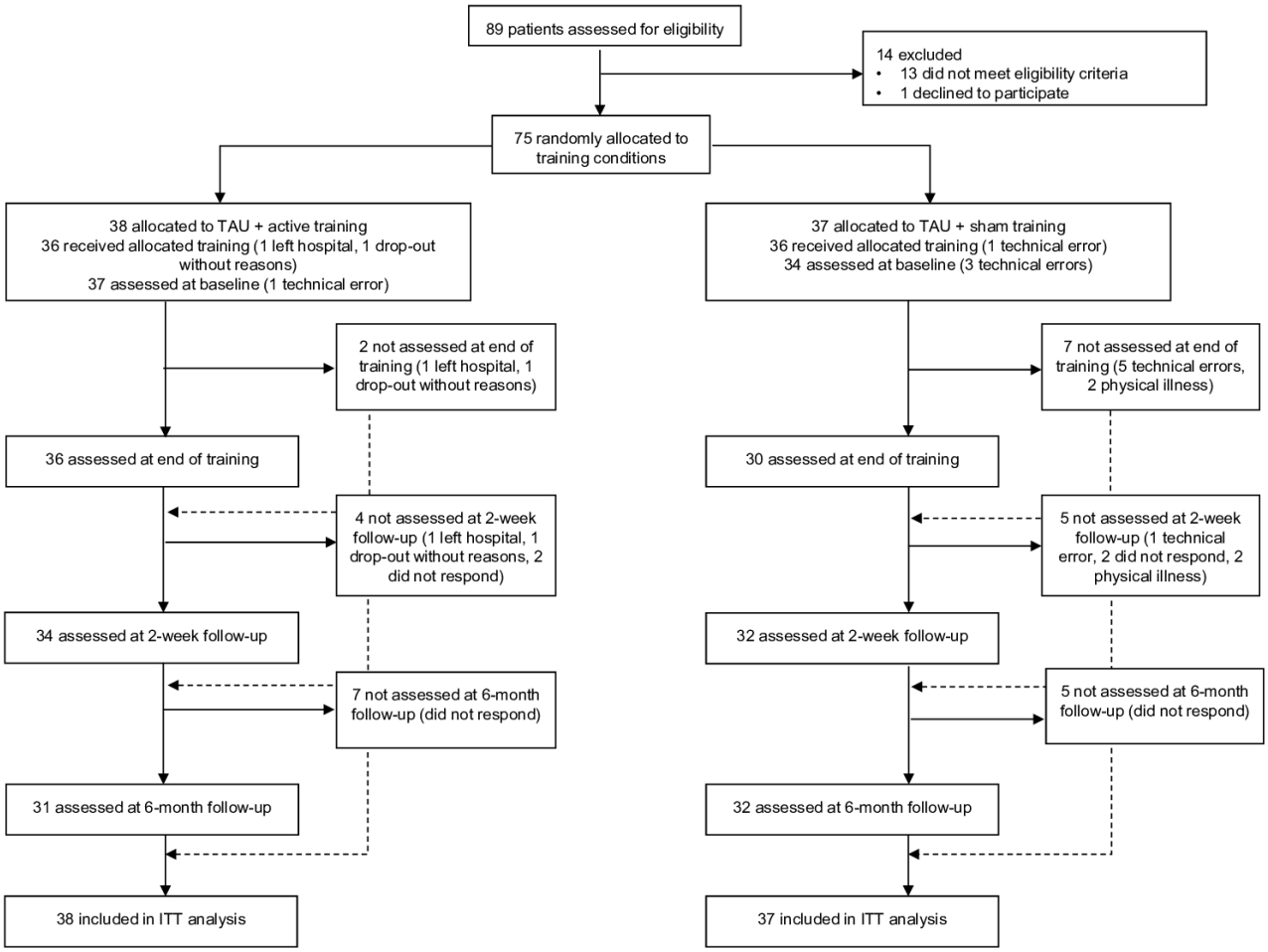
[b]Double depression: concurrent persistent depressive disorder and major depressive disorder.

Means, SD, and number of participants for primary and secondary outcomes at each assessment can be found in Table S1 in Multimedia Appendix 1. The training groups did not differ in any outcome at the baseline assessment (P=.22).

Figure 1 shows the participant flow. Participants in the active training condition received 5.8 (SD 0.6) training sessions on average. Likewise, participants in the sham training condition received 5.8 (SD 0.7) training sessions on average. Training dosage did not differ between groups ($t_{59.86}$=0.15; P=.88). The training groups also did not differ in time spent on completing sessions ($t_{54.93}$=−1.77; P=.08; active training: mean 10.3, SD 3.7 d; sham training: mean 12.2, SD 4.6 d). Reasons for training discontinuance were physical illness, early discharge, and technical errors using the smartphone app. At the 6-month follow-up, 7 (18%) patients were lost in the active training group and 5 (14%) in the sham training group. A total of 29 (76%) participants in the active training group completed all assessments, and 24 (65%) did so in the sham training group. Rates of complete and incomplete assessments did not differ between the groups ($\chi^2_{1.00}$=0.70; P=.40). Among the 62 participants who provided the respective information at the 6-month follow-up, 2 participants in the active training group and 3 participants in the sham training group had started another inpatient treatment after discharge. A total of 24 participants in the active training group and 26 participants in the sham training group had started outpatient psychotherapy, and 15 patients in each training group changed psychotropic medication. Groups did not differ in any of these aspects (P=.75).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart. A patient may miss an earlier assessment while still completing a later one (see dashed lines). ITT: intention to treat; TAU: treatment as usual.

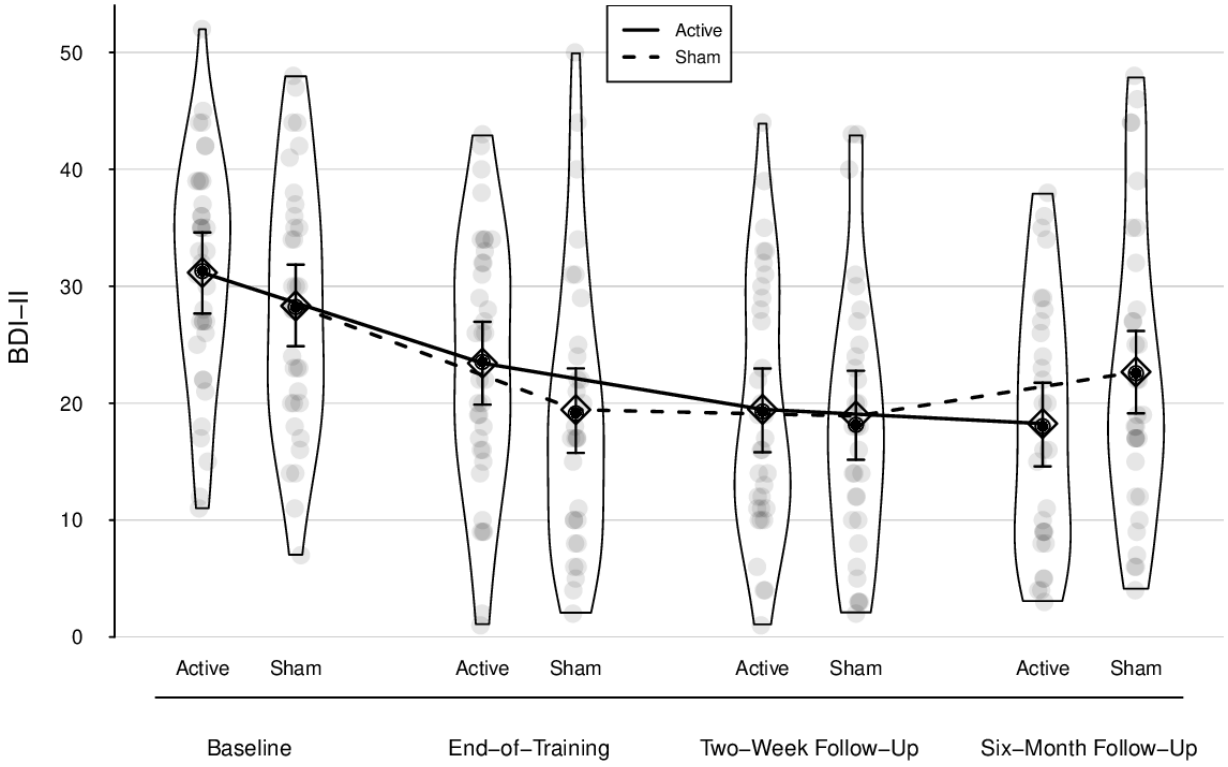


Intention-to-Treat Analysis

The full-reduced model comparison indicated that including training condition and its interaction with time significantly improved the model fit for symptoms of depression (χ^2_4=13.41; *P*=.009). Symptoms of depression decreased from baseline to the end-of-training assessment in the active training group, as indicated by a significant conditional main effect of time (*B*=−7.75, 95% CI −10.76 to −4.65; $t_{188.72}$=−4.71; *P*<.001; *d*=−0.63, 95% CI −0.88 to −0.38). However, the interaction term for condition and time until the end-of-training assessment was not significant (*B*=−1.14, 95% CI −5.65 to 3.41; $t_{188.61}$=−0.47; *P*=.64; *d*=−0.09, 95% CI −0.46 to 0.28). This means that the decrease in symptoms of depression was not different between the groups for the time period between baseline and the end of training. The same was true for the time between baseline and the 2-week follow-up assessment. Symptoms of depression decreased in the active training condition (*B*=−11.70, 95% CI −14.73 to −8.48; $t_{189.53}$=−6.98; *P*<.001; *d*=−1.04, 95% CI −1.30 to −0.75), but this effect was not different from that in the sham training condition, as indicated by a nonsignificant interaction (*B*=2.27, 95% CI −2.40 to 6.88; $t_{190.15}$=0.94; *P*=.35; *d*=0.20, 95% CI −0.21 to 0.61). At the 6-month follow-up, a conditional main effect indicated that symptoms of depression had decreased since the baseline assessment in the active training condition (*B*=−12.91, 95% CI −16.17 to −9.64; $t_{190.37}$=−7.47; *P*<.001; *d*=−1.04, 95% CI −1.30 to −0.77). Furthermore, the interaction between condition and time of assessment was significant for this time period (*B*=7.26, 95% CI 2.53-11.93; $t_{190.54}$=2.95; *P*=.004; *d*=0.58, 95% CI 0.20-0.96). This means that the decrease in symptoms of depression was significantly stronger for the active training group compared to the sham training group for the time between the baseline assessment and 6-month follow-up. Figure 2 displays the change trajectories for BDI-II scores by group, including both raw data and model-based estimates. Patient-level changes in symptoms of depression from baseline to 6-month follow-up are depicted in Figure S1 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 2. Model-based change trajectories in symptoms of depression. Gray dots represent individual data points. Solid black dots show empirical means. Diamonds show model-based estimates with 95% CIs. Bean width reflects distribution. BDI-II: Beck Depression Inventory II.



None of the full-reduced model comparisons for the secondary outcomes assessing approach-avoidance bias, general anhedonia, and "social activities" subscale as well as positivity were significant (P=.33). This indicates that changes over time were not different between training conditions for these outcomes. The pattern of the results remained consistent across other indices of approach-avoidance tendencies. This includes simple biases (differences in reaction time or response force during approach or avoidance of happy stimuli and differences in reaction time or response force during approach or avoidance of angry stimuli) and straightforward comparisons involving only reaction times and response forces for approaching happy stimuli and for avoiding angry stimuli. A significant conditional main effect indicated an increase in positivity from baseline to the end-of-training assessment in the active training group (B=1.94, 95% CI 0.65-3.21; $t_{64.11}$=2.99; P=.004; d=0.32, 95% CI 0.11-0.53). However, this effect did not differ significantly from that observed in the sham training group (B=1.03, 95% CI −0.83 to 2.95; $t_{64.57}$=1.07; P=.29; d=0.17, 95% CI −0.14 to 0.49). For feelings of anhedonia, neither the conditional main effect of time nor its interaction with the training group reached significance (all P values >.16). Similarly, for approach-avoidance bias, there was no significant conditional main effect of time in the active training group and no significant interaction with the training group, for either reaction times (P=.19) or response force (P=.23).

Table S2 in Multimedia Appendix 1 contains estimates and effect sizes for direct comparisons for all outcomes and assessment intervals, including conditional main effects of time in the active training condition and interaction effects. The number of measurements and participants in the random intercept models can be found in Table S3 in Multimedia Appendix 1. Figures S2-S5 in Multimedia Appendix 1 depict model-based change trajectories for secondary outcomes.

Clinical Significance

Overall, 23 (31%) patients experienced a clinically significant change in BDI-II scores from baseline to the 6-month follow-up. In the active group, 15 (39%) patients showed a clinically significant improvement. In the sham training group, 8 (41%) patients did so, which corresponds to a number needed to treat [61] of 5.6 or an area under the receiver operating characteristic curve of 0.59% [62]. However, the training groups did not differ significantly in clinical significant change ($\chi^2_{1.00}$=2.03; P=.15).

Reliability of Bias Assessment

The Spearman-Brown permuted split-half reliabilities for measuring approach-avoidance bias at different assessments with the mobile AAT ranged from 0.77 to 0.94 for reaction times and from 0.81 to 0.93 for response force. This indicates acceptable-to-excellent reliability across the 2 forms of measuring approach-avoidance bias across time [63].

Credibility

At the 2-week follow-up, the training groups did not differ in how they perceived the effectiveness of the intervention ($t_{63.65}$=1.15; P=.26; active training: mean 3.4, SD 1.0; sham training: mean 3.1, SD 0.9) nor in the confidence of their judgment of the efficacy of the intervention ($t_{63.67}$=–0.22; P=.83; active training: mean 3.3, SD 1.0; sham training: mean 3.4, SD 0.9).

XSL•FO
RenderX

Moderation Analysis

Full-reduced model comparisons showed that baseline BDI-II levels significantly moderated the interaction between training condition and time on the approach-avoidance bias based on response force. Specifically, the bias increased more strongly in the active training condition among participants with higher baseline depression. However, this effect did not remain significant after Bonferroni correction. No other moderation effects were observed. The results of the moderation analysis can be found in Table S7 in Multimedia Appendix 1.

Mediation Analysis

The analysis of the effect of training condition on symptoms of depression at the 6-month follow-up via approach-avoidance bias based on reaction time at the end of training including the baseline levels of symptoms of depression and the bias as covariates in a path analysis with 10,000 bootstrapped resamples revealed no significant mediation effect (indirect effect −0.06, 95% CI −1.71 to 1.22; *P*=.91). The same was true for any other potential mediator including anhedonia, positivity, and approach-avoidance bias as response force at the end-of-training assessment and follow-ups (*P*=.17).

Further Analysis

The results of maximum random effects models, missing data analysis, and complete case analyses all resembled the main findings and can be found in Multimedia Appendix 1.

Discussion

To the best of our knowledge, this is the first clinical trial to examine mobile approach-avoidance training adjunct to inpatient TAU in depression. In contrast to our primary hypothesis, the active training was not superior to the sham version in short-term reductions in depressive symptoms (ie, at the end of training and 2-wk follow-up). However, active approach-avoidance training was superior to sham training regarding long-term reductions in depressive symptoms (ie, at the 6-mo follow-up). In terms of clinically significant change, when comparing a patient who received active training against a patient who received sham training, 59% of the time, the patient who received the active training would have a better response (ie, a clinically significant change). This resembles a small-to-moderate effect size [62], although this difference was not significant, potentially due to the sample size.

The lack of short-term superiority of the intervention is in contrast to a previous study on approach-avoidance training in inpatients with depression [22] and a similar trial in a transdiagnostic inpatient sample with elevated symptoms of depression [23]. These studies used joystick-based training instead of a mobile version. However, it does not seem likely that this would lead to faster clinical improvements. A simple reason could be that the short-term effects became visible in the larger samples of these 2 previous studies. In contrast, the result of the long-term efficacy of the training in the present study extends findings from an uncontrolled proof-of-concept study [24], in which clinical improvements were observed 3 months after approach-avoidance training combined with behavioral activation. Furthermore, the long-term efficacy of the training in depression aligns with the positive outcomes of the intervention observed in alcohol-use disorder, where studies have repeatedly demonstrated significant improvements at 1-year follow-ups for individuals undergoing approach-avoidance training [16,18,44]. Given that previous studies on approach-avoidance training in inpatients with depression did not assess long-term effects [22,23], this study offers valuable new insights into the efficacy of this intervention over an extended time period. Furthermore, this study is the first to demonstrate the feasibility of a mobile-based approach-avoidance training that can be easily integrated into patients' daily lives, requiring only a smartphone and about 15 minutes per session.

This study also examined potential mechanisms of approach-avoidance training. However, the findings on secondary outcomes showed no differential effects of active training on anhedonia, positivity, or reaction-time and response-force biases, where both training groups improved similarly. This aligns with previous studies of approach-avoidance training in depression and other mental disorders, including addiction [16,22,23]. Likewise, the reduction in depressive symptoms was not mediated by changes in bias. Thus, the underlying mechanisms of the intervention remain to be elucidated in future research. In this respect, approach-avoidance training aligns with other interventions, such as antidepressants and CBT, for which the exact mechanisms are still not fully understood [64].

However, several methodological aspects must be considered when interpreting our results. In the bias assessment, participants were instructed to approach or avoid faces based on their emotional expression (feature-relevant). In contrast, during training, responses were directed by superimposed symbols, rendering the task feature-irrelevant to the depicted emotion, although the emotional content remained visible. This implicit design was chosen to align with prior clinical intervention studies [16,22,23]. However, approach-avoidance behavior may differ between explicit, feature-relevant tasks and implicit, feature-irrelevant tasks [65]. Developing a feature-relevant training version in the future could therefore be informative, although creating a comparable control condition would be challenging, as it would require varying task instructions throughout the training. Finally, we retained a feature-relevant assessment version to maximize reliability and comparability with earlier mobile AAT studies [29,30]. Indeed, evidence indicates that feature-relevant instructions yield more reliable and valid measures than feature-irrelevant paradigms [50,66].

What are the possible mechanisms of the training that could lead to a long-term reduction in symptoms of depression? Changes in implicit approach tendencies, particularly toward socioemotional cues, may gradually translate into explicit social approach behaviors. Over time, such changes can reduce depressive symptoms by fostering rewarding social experiences. Dysfunctional processing of socioemotional stimuli often leads to social withdrawal, limiting positive reinforcement and thereby maintaining depression [13,67,68]. Consistent with this, the quality of social relationships and behavioral activation are key determinants of symptom severity and chronicity [69,70]. An encouraging social approach is therefore a central component

XSL•FO
RenderX

of contemporary psychotherapy [71]. Our study extends previous approach-avoidance training in depression [22,23], which typically used heterogeneous image sets [26], by employing standardized facial expressions. This design specifically targets impaired approach behavior in social contexts [14,65]. Furthermore, functional magnetic resonance imaging findings suggest that approach-avoidance training can normalize aberrant reward sensitivity to positive social cues [25,27]. Training approach behavior toward positive stimuli could also strengthen mood-incongruent processing, which is diminished in depression [72,73], yet has been shown to repair negative mood states [74] and predict more favorable outcomes [75]. Individuals may need time and opportunities after inpatient treatment to translate implicit training gains in social approach behavior into real-world social interactions for these effects to unfold gradually. Future research should therefore examine downstream effects of approach-avoidance training on social connectedness, relationship functioning, and behavioral activation as potential mediators of long-term symptom reduction [36].

Patients with depression frequently relapse or experience symptom worsening after treatment [5]. Consistent with this, depressive symptoms in the sham training group increased from the 2-week to 6-month follow-up. Such long-term limitations of treatment may partly reflect persistent cognitive biases that contribute to the maintenance of depression [6,7] and were directly targeted in the active training condition. Thus, our findings may also suggest that active approach-avoidance training helped prevent symptom worsening from mid- to long-term. Future studies should include longer follow-ups (eg, up to 12 mo, as is common in addiction research [16]) to clarify the trajectory of symptom change over time.

Our study's findings should be considered in light of its strengths and limitations. A strength of the mobile AAT is that it enables assessment of both reaction times and response force, adding a dimension not available in traditional joystick paradigms. However, mobile AAT paradigms are still relatively new [29], and prior studies have struggled to reliably detect force-related approach-avoidance biases [76,77]. More basic research is needed to clarify the nature of force-based biases and their relation to reaction times. Additionally, we were able to assess the vast majority (n=63, 84%) of patients at the 6-month follow-up. Dropout and attrition rates were considerably low and did not differ between conditions, but still, some data were lost due to technical errors. A large number (n=50) of the participants had started outpatient psychotherapy by the 6-month follow-up. Importantly, this rate did not differ between the training groups. This aligns with the general logic of randomized clinical trials, in which randomization is assumed to balance potential confounding factors (eg, concurrent outpatient therapy) across training conditions [78]. Due to the monocentric approach in a clinic in a Western European country, the generalizability of the findings to other treatment settings and differing cultural contexts is limited. Patients completed the mobile AAT independently on their smartphones, and we could not fully verify whether prototypical approach-avoidance movements were performed throughout. We monitored adherence by tracking session completion online and phoning participants if deviations occurred. Although patients could technically complete multiple training sessions per day, data indicate that they generally followed the 2-week protocol, and training duration did not differ between the groups. Still, future training versions may technically be restricted to 1 session per day. The results of the credibility checks do not suggest that patients in the 2 training conditions differed in their expectations regarding the efficacy of the training 2 weeks after completion of the intervention. The clinical outcomes of our study relied on self-reported measures. Future research may also include clinician-rated assessments of depressive symptoms to support the self-assessment. Furthermore, our study utilized a limited range of highly standardized stimuli with facial expressions of emotions from a well-established image database [45]. While using stimuli of friendly and angry faces might appear overly simplistic, customizing the training with individualized stimuli for each patient could hinder the standardization of the intervention and introduce confounding variables. This needs to be considered when assessing the reliability of the approach-avoidance task. Finally, our sample size was sufficient to detect moderate interaction effects, but it lacked power to identify smaller treatment effects. A larger sample would increase sensitivity to subtle short-term effects, yield more precise estimates of change over time, and allow adequately powered moderation and mediation analyses, which in this study were only exploratory.

This study investigated the efficacy of a mobile approach-avoidance training as an add-on to inpatient treatment for individuals with depression. Despite its limitations, the study used novel methodological approaches that could pave the way for more flexible research designs for CBM studies in the future. In summary, we found evidence for long-term but not short-term superiority of the active training compared to a sham version in terms of the primary outcome, symptoms of depression. This effect, however, cannot yet be explained by a change in approach-avoidance bias, neither when measured via reaction times nor when measured via response force. Future research should aim to elucidate the mechanisms at work of this intervention and replicate the findings in larger samples in multicenter trials.

Acknowledgments

We would like to thank Lena Temmen and Julia Koch for their help with the data collection.

Funding

The work of HGZ was partially supported by a grant from the German Research Foundation (DFG Project number 40217046).

XSL•FO
RenderX

Data Availability

All anonymized data and analysis code are available at the project's Open Science Framework (OSF) web page [46]. The source code of the Mobile Approach-Avoidance Task is available on OSF [49].

Authors' Contributions

Conceptualization: TB
Data curation: HGZ, MB
Formal analysis: MB
Funding acquisition: TB
Investigation: MB, MK, NB
Methodology: MB, TB
Project administration: MB, TB
Resources: HP
Software: HGZ, MB
Supervision: TB
Validation: MB, TB
Visualization: MB
Writing – original draft: MB
Writing – review and editing: HP, HGZ, MB, MK, NB, TB

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material including additional tables, figures, and details of statistical analyses.
[PDF File, 1510 KB - mhealth_v13i1e69033_app1.pdf]

Checklist 1
CONSORT-EHEALTH checklist (V1.6.1).
[PDF File, 1103 KB - mhealth_v13i1e69033_app2.pdf]

References

1. Rehm J, Shield KD. Global burden of disease and the impact of mental and addictive disorders. Curr Psychiatry Rep 2019 Feb 7;21(2):1-7. [doi: 10.1007/s11920-019-0997-0] [Medline: 30729322]
2. Cuijpers P, Smit F. Excess mortality in depression: a meta-analysis of community studies. J Affect Disord 2002 Dec;72(3):227-236. [doi: 10.1016/s0165-0327(01)00413-x] [Medline: 12450639]
3. Penninx BWJH, Milaneschi Y, Lamers F, Vogelzangs N. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. BMC Med 2013 May 15;11(1):129. [doi: 10.1186/1741-7015-11-129] [Medline: 23672628]
4. Kessler RC. The costs of depression. Psychiatr Clin North Am 2012 Mar;35(1):1-14. [doi: 10.1016/j.psc.2011.11.005] [Medline: 22370487]
5. Cuijpers P, Miguel C, Harrer M, et al. Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. World Psychiatry 2023 Feb;22(1):105-115. [doi: 10.1002/wps.21069] [Medline: 36640411]
6. Everaert J, Koster EHW, Derakshan N. The combined cognitive bias hypothesis in depression. Clin Psychol Rev 2012 Jul;32(5):413-424. [doi: 10.1016/j.cpr.2012.04.003] [Medline: 22681914]
7. Mathews A, MacLeod C. Cognitive vulnerability to emotional disorders. Annu Rev Clin Psychol 2005;1:167-195. [doi: 10.1146/annurev.clinpsy.1.102803.143916] [Medline: 17716086]
8. Everaert J, Vrijsen JN, Martin-Willett R, van de Kraats L, Joormann J. A meta-analytic review of the relationship between explicit memory bias and depression: depression features an explicit memory bias that persists beyond a depressive episode. Psychol Bull 2022 May;148(5-6):435-463. [doi: 10.1037/bul0000367]
9. Peckham AD, McHugh RK, Otto MW. A meta-analysis of the magnitude of biased attention in depression. Depress Anxiety 2010 Dec;27(12):1135-1142. [doi: 10.1002/da.20755] [Medline: 21049527]
10. Everaert J, Podina IR, Koster EHW. A comprehensive meta-analysis of interpretation biases in depression. Clin Psychol Rev 2017 Dec;58:33-48. [doi: 10.1016/j.cpr.2017.09.005] [Medline: 28974339]
11. Cacioppo JT, Priester JR, Berntson GG. Rudimentary determinants of attitudes. II: arm flexion and extension have differential effects on attitudes. J Pers Soc Psychol 1993 Jul;65(1):5-17. [doi: 10.1037//0022-3514.65.1.5] [Medline: 8355142]

XSL•FO
RenderX

12. Chen M, Bargh JA. Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. Pers Soc Psychol Bull 1999 Feb;25(2):215-224. [doi: 10.1177/0146167299025002007]
13. Trew JL. Exploring the roles of approach and avoidance in depression: an integrative model. Clin Psychol Rev 2011 Nov;31(7):1156-1168. [doi: 10.1016/j.cpr.2011.07.007] [Medline: 21855826]
14. Radke S, Güths F, André JA, Müller BW, de Bruijn ERA. In action or inaction? Social approach-avoidance tendencies in major depression. Psychiatry Res 2014 Nov 30;219(3):513-517. [doi: 10.1016/j.psychres.2014.07.011] [Medline: 25060832]
15. Rinck M, Becker ES. Approach and avoidance in fear of spiders. J Behav Ther Exp Psychiatry 2007 Jun;38(2):105-120. [doi: 10.1016/j.jbtep.2006.10.001] [Medline: 17126289]
16. Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. Psychol Sci 2011 Apr;22(4):490-497. [doi: 10.1177/0956797611400615] [Medline: 21389338]
17. Loijen A, Vrijsen JN, Egger JIM, Becker ES, Rinck M. Biased approach-avoidance tendencies in psychopathology: a systematic review of their assessment and modification. Clin Psychol Rev 2020 Apr;77:101825. [doi: 10.1016/j.cpr.2020.101825] [Medline: 32143108]
18. Kakoschke N, Kemps E, Tiggemann M. Approach bias modification training and consumption: a review of the literature. Addict Behav 2017 Jan;64:21-28. [doi: 10.1016/j.addbeh.2016.08.007] [Medline: 27538198]
19. Brockmeyer T, Friederich HC, Küppers C, et al. Approach bias modification training in bulimia nervosa and binge-eating disorder: a pilot randomized controlled trial. Int J Eat Disord 2019 May;52(5):520-529. [doi: 10.1002/eat.23024] [Medline: 30689229]
20. Beard C. Cognitive bias modification for anxiety: current evidence and future directions. Expert Rev Neurother 2011 Feb;11(2):299-311. [doi: 10.1586/ern.10.194] [Medline: 21306216]
21. MacLeod C, Mathews A. Cognitive bias modification approaches to anxiety. Annu Rev Clin Psychol 2012;8(1):189-217. [doi: 10.1146/annurev-clinpsy-032511-143052] [Medline: 22035241]
22. Vrijsen JN, Fischer VS, Müller BW, et al. Cognitive bias modification as an add-on treatment in clinical depression: results from a placebo-controlled, single-blinded randomized control trial. J Affect Disord 2018 Oct 1;238:342-350. [doi: 10.1016/j.jad.2018.06.025] [Medline: 29908472]
23. Becker ES, Barth A, Smits JAJ, Beisel S, Lindenmeyer J, Rinck M. Positivity-approach training for depressive symptoms: a randomized controlled trial. J Affect Disord 2019 Feb 15;245:297-304. [doi: 10.1016/j.jad.2018.11.042] [Medline: 30439675]
24. Sweet AM, Pearlstein SL, Paulus MP, Stein MB, Taylor CT. Computer-delivered behavioural activation and approach-avoidance training in major depression: proof of concept and initial outcomes. Br J Clin Psychol 2021 Sep;60(3):357-374. [doi: 10.1111/bjc.12287] [Medline: 33772806]
25. Bomyea J, Choi SH, Sweet A, Stein M, Paulus M, Taylor C. Neural changes in reward processing following approach avoidance training for depression. Soc Cogn Affect Neurosci 2021 Oct 13;17(3):336-349. [doi: 10.1093/scan/nsab107] [Medline: 34643736]
26. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical Report A-8. : University of Florida; 2008.
27. Derntl B, Seidel EM, Eickhoff SB, et al. Neural correlates of social approach and withdrawal in patients with major depression. Soc Neurosci 2011;6(5-6):482-501. [doi: 10.1080/17470919.2011.579800] [Medline: 21777105]
28. Ferrari GRA, Möbius M, Becker ES, Spijker J, Rinck M. Working mechanisms of a general positivity approach-avoidance training: effects on action tendencies as well as on subjective and physiological stress responses. J Behav Ther Exp Psychiatry 2018 Jun;59:134-141. [doi: 10.1016/j.jbtep.2018.01.005] [Medline: 29408051]
29. Zech HG, Rotteveel M, van Dijk WW, van Dillen LF. A mobile approach-avoidance task. Behav Res Methods 2020 Oct;52(5):2085-2097. [doi: 10.3758/s13428-020-01379-3] [Medline: 32180179]
30. Zech HG, Gable P, van Dijk WW, van Dillen LF. Test-retest reliability of a smartphone-based approach-avoidance task: effects of retest period, stimulus type, and demographics. Behav Res Methods 2023 Aug;55(5):2652-2668. [doi: 10.3758/s13428-022-01920-6] [Medline: 35915356]
31. Rizvi SJ, Quilty LC, Sproule BA, Cyriac A, Michael Bagby R, Kennedy SH. Development and validation of the Dimensional Anhedonia Rating Scale (DARS) in a community sample and individuals with major depression. Psychiatry Res 2015 Sep 30;229(1-2):109-119. [doi: 10.1016/j.psychres.2015.07.062] [Medline: 26250147]
32. Vinckier F, Gourion D, Mouchabac S. Anhedonia predicts poor psychosocial functioning: results from a large cohort of patients treated for major depressive disorder by general practitioners. Eur Psychiatry 2017 Jul;44:1-8. [doi: 10.1016/j.eurpsy.2017.02.485] [Medline: 28535406]
33. Grèzes J, Risch N, Courtet P, Olié E, Mennella R. Depression and approach-avoidance decisions to emotional displays: the role of anhedonia. Behav Res Ther 2023 May;164:104306. [doi: 10.1016/j.brat.2023.104306] [Medline: 37043847]
34. Caprara GV, Alessandri G, Eisenberg N, et al. The positivity scale. Psychol Assess 2012 Sep;24(3):701-712. [doi: 10.1037/a0026681] [Medline: 22250591]
35. Heikamp T, Alessandri G, Laguna M, Petrovic V, Caprara MG, Trommsdorff G. Cross-cultural validation of the positivity-scale in five European countries. Pers Individ Dif 2014 Dec;71:140-145. [doi: 10.1016/j.paid.2014.07.012]

XSL•FO
RenderX

36. Taylor CT, Pearlstein SL, Kakaria S, Lyubomirsky S, Stein MB. Enhancing social connectedness in anxiety and depression through amplification of positivity: preliminary treatment outcomes and process of change. Cognit Ther Res 2020 Aug;44(4):788-800. [doi: 10.1007/s10608-020-10102-7] [Medline: 32661447]
37. Revelle W. Package “psych". : The Comprehensive R Archive Network; 2015. [doi: 10.32614/CRAN.package.psych]
38. Arslan RC, Walther MP, Tata CS. formr: a study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. Behav Res Methods 2020 Feb;52(1):376-387. [doi: 10.3758/s13428-019-01236-y] [Medline: 30937847]
39. Diagnostic and Statistical Manual of Mental Disorders: DSM-5: American Psychiatric Association; 2013, Vol. 5. URL: https://psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596 [accessed 2025-11-14]
40. First MB. Structured clinical interview for the DSM (SCID). In: The Encyclopedia of Clinical Psychology: Wiley; 2014:1-6. [doi: 10.1002/9781118625392]
41. European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council. EUR-Lex. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng [accessed 2025-11-19]
42. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
43. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073. [doi: 10.1177/0962280215588241] [Medline: 26092476]
44. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? Dev Cogn Neurosci 2013 Apr;4:38-51. [doi: 10.1016/j.dcn.2012.11.002] [Medline: 23218805]
45. Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. Presentation and validation of the Radboud Faces Database. Cogn Emot 2010 Dec;24(8):1377-1388. [doi: 10.1080/02699930903485076]
46. Smartphone-based approach-avoidance bias modification training for depression. Open Science Framework. URL: https://osf.io/u5zj8/ [accessed 2025-11-14]
47. Hautzinger M, Keller F, Kühner C. BDI-II Beck-Depressions-Inventar Revision - Manual: Harcourt Test Services; 2006.
48. Kühner C, Bürger C, Keller F, Hautzinger M. Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II): Befunde aus deutschsprachigen Stichproben. Nervenarzt 2007 Jun;78(6):651-656. [doi: 10.1007/s00115-006-2098-7]
49. Smartphone-based approach-avoidance bias modification training for depression. Open Science Framework. URL: https://osf.io/y5b32/ [accessed 2025-11-14]
50. Kahveci S, Rinck M, van Alebeek H, Blechert J. How pre-processing decisions affect the reliability and validity of the approach-avoidance task: evidence from simulations and multiverse analyses with six datasets. Behav Res Methods 2024 Mar;56(3):1551-1582. [doi: 10.3758/s13428-023-02109-1] [Medline: 37221345]
51. Wellan SA, Daniels A, Walter H. State anhedonia in young healthy adults: psychometric properties of the German Dimensional Anhedonia Rating Scale (DARS) and effects of the COVID-19 pandemic. Front Psychol 2021;12:682824. [doi: 10.3389/fpsyg.2021.682824] [Medline: 34248783]
52. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang 2008 Nov;59(4):390-412. [doi: 10.1016/j.jml.2007.12.005]
53. Forstmeier W, Schielzeth H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner’s curse. Behav Ecol Sociobiol 2011 Jan;65(1):47-55. [doi: 10.1007/s00265-010-1038-5] [Medline: 21297852]
54. Luke SG. Evaluating significance in linear mixed-effects models in R. Behav Res Methods 2017 Aug;49(4):1494-1502. [doi: 10.3758/s13428-016-0809-y] [Medline: 27620283]
55. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009 Mar;14(1):43-53. [doi: 10.1037/a0014699] [Medline: 19271847]
56. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol Methods 1997;2(1):64-78. [doi: 10.1037//1082-989X.2.1.64]
57. Hayes AF, Rockwood NJ. Regression-based statistical mediation and moderation analysis in clinical research: observations, recommendations, and implementation. Behav Res Ther 2017 Nov;98:39-57. [doi: 10.1016/j.brat.2016.11.001] [Medline: 27865431]
58. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 2013 Apr;68(3):255-278. [doi: 10.1016/j.jml.2012.11.001] [Medline: 24403724]
59. Kahveci S, Bathke AC, Blechert J. Reaction-time task reliability is more accurately computed with permutation-based split-half correlations than with Cronbach’s alpha. Psychon Bull Rev 2025 Apr;32(2):652-673. [doi: 10.3758/s13423-024-02597-y] [Medline: 39443394]
60. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037/0022-006X.59.1.12] [Medline: 2002127]
61. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988 Jun 30;318(26):1728-1733. [doi: 10.1056/NEJM198806303182605] [Medline: 3374545]

XSL•FO
RenderX

62. Kraemer HC, Morgan GA, Leech NL, Gliner JA, Vaske JJ, Harmon RJ. Measures of clinical significance. J Am Acad Child Adolesc Psychiatry 2003 Dec;42(12):1524-1529. [doi: 10.1097/00004583-200312000-00022] [Medline: 14627890]
63. Nunnally JC, Bernstein IH. Psychometric Theory: McGraw-Hill; 1994.
64. Ma Y. Neuropsychological mechanism underlying antidepressant effect: a systematic meta-analysis. Mol Psychiatry 2015 Mar;20(3):311-319. [doi: 10.1038/mp.2014.24] [Medline: 24662929]
65. Seidel EM, Habel U, Finkelmeyer A, Schneider F, Gur RC, Derntl B. Implicit and explicit behavioral tendencies in male and female depression. Psychiatry Res 2010 May 15;177(1-2):124-130. [doi: 10.1016/j.psychres.2010.02.001] [Medline: 20199811]
66. Lender A, Meule A, Rinck M, Brockmeyer T, Blechert J. Measurement of food-related approach-avoidance biases: larger biases when food stimuli are task relevant. Appetite 2018 Jun 1;125:42-47. [doi: 10.1016/j.appet.2018.01.032] [Medline: 29407526]
67. Kamenov K, Caballero FF, Miret M, et al. Which are the most burdensome functioning areas in depression? A cross-national study. Front Psychol 2016;7:1342. [doi: 10.3389/fpsyg.2016.01342] [Medline: 27630609]
68. Brockmeyer T, Holtforth MG, Krieger T, et al. Preliminary evidence for a nexus between rumination, behavioural avoidance, motive satisfaction and depression. Clin Psychol Psychother 2015;22(3):232-239. [doi: 10.1002/cpp.1885] [Medline: 24464405]
69. Åhlin JK, Rajaleid K, Jansson-Fröjmark M, Westerlund H, Magnusson Hanson LL. Job demands, control and social support as predictors of trajectories of depressive symptoms. J Affect Disord 2018 Aug 1;235:535-543. [doi: 10.1016/j.jad.2018.04.067] [Medline: 29689506]
70. Brockmeyer T, Kulessa D, Hautzinger M, Bents H, Backenstrass M. Differentiating early-onset chronic depression from episodic depression in terms of cognitive-behavioral and emotional avoidance. J Affect Disord 2015 Apr 1;175:418-423. [doi: 10.1016/j.jad.2015.01.045] [Medline: 25679196]
71. Jacobson NS, Martell CR, Dimidjian S. Behavioral activation treatment for depression: returning to contextual roots. Clin Psychol Sci Pract 2001;8(3):255-270. [doi: 10.1093/clipsy.8.3.255]
72. Capecelatro MR, Sacchet MD, Hitchcock PF, Miller SM, Britton WB. Major depression duration reduces appetitive word use: an elaborated verbal recall of emotional photographs. J Psychiatr Res 2013 Jun;47(6):809-815. [doi: 10.1016/j.jpsychires.2013.01.022] [Medline: 23510497]
73. Joormann J, Siemer M, Gotlib IH. Mood regulation in depression: differential effects of distraction and recall of happy memories on sad mood. J Abnorm Psychol 2007 Aug;116(3):484-490. [doi: 10.1037/0021-843X.116.3.484]
74. Forgas JP, Ciarrochi JV. On managing moods: evidence for the role of homeostatic cognitive strategies in affect regulation. Pers Soc Psychol Bull 2002 Mar;28(3):336-345. [doi: 10.1177/0146167202286005]
75. Brockmeyer T, Kulessa D, Hautzinger M, Bents H, Backenstrass M. Mood-incongruent processing during the recall of a sad life event predicts the course and severity of depression. J Affect Disord 2015 Nov 15;187:91-96. [doi: 10.1016/j.jad.2015.08.010] [Medline: 26322713]
76. Puca RM, Rinkenauer G, Breidenstein C. Individual differences in approach and avoidance movements: how the avoidance motive influences response force. J Pers 2006 Aug;74(4):979-1014. [doi: 10.1111/j.1467-6494.2006.00400.x] [Medline: 16787426]
77. Rotteveel M, Phaf RH. Automatic affective evaluation does not automatically predispose for arm flexion and extension. Emotion 2004 Jun;4(2):156-172. [doi: 10.1037/1528-3542.4.2.156] [Medline: 15222853]
78. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340(mar23 1):c869. [doi: 10.1136/bmj.c869] [Medline: 20332511]

Edited by L Buis; submitted 20.11.24; peer-reviewed by B Kraft, J Blechert; accepted 22.10.25; published 26.11.25.

Please cite as:
Blomberg M, Zech HG, Kluge M, Böhmert N, Platte H, Brockmeyer T
Smartphone-Based Approach-Avoidance Bias Modification Training for Depression: Randomized Clinical Trial
JMIR Mhealth Uhealth 2025;13:e69033
URL: https://mhealth.jmir.org/2025/1/e69033
doi:10.2196/69033

© Maximilian Blomberg, Hilmar Gero Zech, Maximilian Kluge, Nico Böhmert, Helmut Platte, Timo Brockmeyer. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

How Engagement Changes Over Time in a Digital Eating Disorder App: Observational Study

Rachael E Flatt[1], PhD; Laura M Thornton[1], PhD; Jenna Tregarthen[2], PGDip; Stuart Argue[2], BS; Cynthia M Bulik[1,3,4], PhD

[1]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Recovery Record (United States), Palo Alto, CA, United States

[3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels väg 12A, Stockholm, Sweden

[4]Department of Nutrition, University of North Carolina at Chapel Hill, 101 Manning Drive, Chapel Hill, NC, United States

Corresponding Author:
Cynthia M Bulik, PhD
Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Abstract

Background: Engagement with digital mental health interventions is often measured as a summary-level variable and remains underresearched despite its importance for meaningful symptom change. This study deepens understanding of engagement in a digital eating disorder intervention, recovery record, by measuring engagement with unique components of the app, on 2 different devices (phone and watch), and at a summary level.

Objective: This study described and modeled how individuals engaged with the app across a variety of measures of engagement and identified baseline predictors of engagement.

Methods: Participants with current binge-eating behavior were recruited as part of the Binge Eating Genetics Initiative study to use a digital eating disorder intervention for 4 weeks. Demographic and severity of illness variables were captured in the baseline survey at enrollment, and engagement data were captured through both an iPhone and Apple Watch version of the intervention. Engagement was characterized by log type (urge, behavior, mood, or meal), device type (logs on phone or watch), and overall usage (total logs) and averaged each week for 4 weeks. Descriptives were tabulated for demographic and engagement variables, and multilevel growth models were conducted for each measure of engagement with baseline characteristics and time as predictors.

Results: Participants (N=893) self-reported as primarily White (743/871, 85%), non-Hispanic (801/893, 90%), females (772/893, 87%) with a mean age of 29.6 (SD 7.4) years and mean current BMI of 32.5 (SD 9.8) kg/m^2 and used the app for a mean of 24 days. Most logs were captured on phones (217,143/225,927; 96%), and mood logs were the most used app component (174,818/282,136; 62% of logs). All measures of engagement declined over time, as illustrated by the visualizations, but each measure of engagement illustrated unique participant trajectories over time. Time was a significant negative predictor in every multilevel model. Sex and ethnicity were also significant predictors across several measures of engagement, with female and Hispanic participants demonstrating greater engagement than male and non-Hispanic counterparts. Other baseline characteristics (age, current BMI, and binge episodes in the past 28 days) were significant predictors of 1 measure of engagement each.

Conclusions: This study highlighted that engagement is far more complex and nuanced than is typically described in research, and that specific components and mode of delivery may have unique engagement profiles and predictors. Future work would benefit from developing early engagement models informed by baseline characteristics to predict intervention outcomes, thereby tailoring digital eating disorder interventions at the individual level.

Trial Registration: ClinicalTrials.gov NCT04162574; https://tinyurl.com/4jjfhemf

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-020-02698-7

(JMIR Mhealth Uhealth 2025;13:e68824) doi:10.2196/68824

KEYWORDS

eating disorders; wearable technology; mobile application; engagement; baseline

XSL•FO
RenderX

Introduction

Background

Digital interventions hold substantial promise for addressing mental health disorders, including eating disorders (EDs). Traditional face-to-face ED treatment is often cost-prohibitive, especially for the uninsured, difficult to access, not scalable, retrospective in nature, and unable to offer support in real time when patients need it most [1-3]. Digital ED interventions address these obstacles by offering evidence-based and in-the-moment treatment options that are affordable and accessible [4-7]. A crucial aspect of digital interventions is engagement [8,9], which parallels measures of retention and adherence in face-to-face treatments. Just like individuals attend sessions, complete self-monitoring forms for home practice, and apply learned skills outside of sessions, digital interventions for EDs can measure similar and expanded forms of engagement that may be key to achieving and maintaining treatment gains.

Similar to observations in the broader digital mental health tool literature, engagement in digital ED interventions is low [6], the definitions used to describe engagement are heterogeneous and mostly limited to static or summary-level variables, and predictors of engagement are not well researched. Thus, this study analyzes engagement in a digital ED intervention, recovery record, over the course of a month for individuals with binge-type EDs. Specifically, we model longitudinal growth curves to examine the dynamic trajectories of several measures of engagement and subsequently identify baseline predictors.

Binge-Type Eating Disorders

Bulimia nervosa (BN) and binge eating disorder (BED) affect between 1.0% - 3.5% of the population [10,11], disproportionately affect females, and have a typical age of onset in the late teens and early 20s [12]. If left untreated, both disorders carry significant psychological and medical implications, high health care use, and high relapse rates [13,14]. Given that the courses of these disorders are dynamic in nature (eg, diagnostic shift and changes in temporal patterns of binge eating and compensatory behaviors), digital ED interventions need to be flexible such that individuals can engage with them when they need them most in the moment. Digital ED interventions can help address critical gaps in accessing care in a timely fashion; however, it is essential that we begin to refine our understanding and conceptualizations of how individuals with EDs engage with these tools to support future work that facilitates uptake, usability, acceptability, and positive clinical outcomes.

Engagement in Digital Eating Disorder Interventions

Engagement, broadly defined in this paper as the ways and extent to which an individual uses a digital intervention, is frequently cited as a key challenge in digital mental health tools [8,15,16]. Although different terms are used to describe how an individual uses a digital mental health app (engagement, app usage, adherence, etc), the term engagement is used throughout this paper for consistency. Across studies reporting on engagement in digital ED interventions, barriers to initiating engagement identified across qualitative and quantitative studies of digital ED interventions mirrored those of other digital mental health interventions and included logistical constraints such as cost, accessibility, usability or functionality [17], time [18], and privacy of personal health information [19]. Perceived barriers include treatment credibility and expectancy [20], motivation, accountability, content and feature preferences [17], severity of illness, trust of the intervention [18], satisfaction, intervention personalization, and ease of use [19]. Once individuals initiate engagement with an app, several papers also cite maintaining engagement as a significant challenge: the median percentage of users in a review of treatment studies of digital ED interventions where all prescribed modules or activities were completed was only 36% [6]. Greater engagement is shown to be associated with better treatment outcomes [21,22], suggesting that maintaining engagement is important for users seeking meaningful change in their ED symptoms. Notably, most of these definitions only captured summary-level or endpoint measurements of engagement, so little is known about how engagement changes over time and if changes in engagement contribute to treatment gains.

Researchers also contend with methodological challenges, namely that there are seemingly endless conceptualizations and both qualitative and quantitative measurements of engagement [16,23,24], and most studies evaluating engagement do not capture multiple data streams to illustrate dynamics. The types of engagement data collected also vary widely depending on the features and content delivered to users through the intervention, the data users consent to provide, and the mode of delivery (eg, through a smartphone app, wearable technology, and web-based browser). In the context of digital ED interventions, the definitions of engagement vary widely: for example, studies defined engagement as number of modules completed [21]; total number of logs completed, total number of days active with the app, and length of time using the app [22]; interactivity and usability [25]; and total app views and total number of meal logs [26]; among others. Relatedly, apps and devices can limit or bias the engagement data that are reported due to differences in data capture. As an example, apps may capture the number of logins without adequately capturing a logout time, thus biasing the data toward extended periods of time where it appears that users are logged on and engaged. In research studies, investigative teams also typically use strategies such as reminders, phone calls, and compensation to boost user engagement [6], which may inflate overall engagement results compared with naturalistic app use. As a result, the field has yet to elucidate common themes about temporal changes in engagement beyond summary-level variables.

Recovery Record Engagement Research

This study uses data collected through recovery record [3], a widely used evidence-based ED app; however, only 2 studies assessed participant engagement in recovery record to date. The first qualitative study explored engagement in a Danish-translated version of Recovery Record in participants with anorexia nervosa (AN) or BN [27], the majority of whom used the app between 1 and 4 months. Participants reported that engaging with the app helped them confront the ED or log meals more constructively, leading to less concern with caloric intake, but also reported that engagement could be obstructive by

XSL•FO
RenderX

increasing obsessions with logging or by giving participants ideas about other compensatory behaviors (eg, participant sees "excessive exercise" as an option to log and begins to feel an urge to exercise). The second study [22] evaluated participants' engagement with recovery record in a sample of recovery record users as part of a larger randomized controlled trial testing efficacy of the app. The total number of meal logs and total number of days the app was used significantly and positively mediated the treatment effect on clinical response 8 weeks later, indicating that this greater engagement could lead to more positive treatment outcomes. However, only 3 measurements of engagement were sufficiently tested as mediators, and the study did not describe or evaluate changes in engagement over time.

Despite the multitude of interactive features that most digital ED apps offer, including recovery record, most digital ED research has focused on heterogenous summary-level measurements of engagement (eg, number of days the app was used), and the analyses do not reflect the dynamic nature of engagement. In addition, baseline contextual factors that may influence engagement trajectories have not yet been thoroughly explored with digital ED interventions. As a result, valuable information pertaining to engagement trajectories is mostly unavailable, and we may fail to capture how different types of engagement change over time and key participant characteristics that are associated with engagement. This study addresses these issues by defining several types of engagement based on the available interactive features in recovery record, modeling trajectories for each measure of engagement, and identifying key baseline predictors of engagement.

Aims

The aims of this study were to: (1) describe characteristics of the sample and how individuals engaged with the app across a variety of measures of engagement, (2) model the trajectories of engagement over the 30-day course of the study, and (3) identify baseline demographic and ED symptom predictors of engagement. Based on previously published literature, we hypothesized that individuals will generally demonstrate downward trajectories of engagement across the study period. However, the dearth of evidence on engagement in digital ED interventions and in recovery record does not support more specific a priori hypotheses for baseline predictors of engagement.

Methods

Participants

Participants were recruited as part of a larger parent case-only trial, the Binge Eating Genetics Initiative (BEGIN) study. The full study protocol for the parent trial is available elsewhere [28]. Briefly, the BEGIN study sought to integrate genetic, microbiome, phenotypic, and behavioral data for individuals with binge-type EDs. Pertinent to this paper, participants used recovery record through an iPhone and Apple Watch to collect actively logged behavioral and affective data in addition to passive data on heart rate and steps as foundations for just-in-time adaptive interventions (refer to Bulik et al [28] and Flatt et al [29] for more details on the technology component of the study as well as photos of the app on the Apple Watch). Inclusion criteria for this study included (1) current binge eating, (2) lifetime diagnosis of either BN or binge-eating disorder, (3) US resident, (4) between 18 and 45 years old, (5) reads and speaks English, (6) current iPhone user, (7) ambulatory, and (8) willing and able to participate in the study, wear an Apple Watch, and use recovery record. One additional criterion for analysis in this study was completion of at least 1 log on Recovery Record.

Exclusion criteria included (1) currently pregnant or breastfeeding, (2) history of bariatric surgery, (3) current use of hormone therapy, (4) inpatient treatment or hospitalization for ED in the 2 weeks before enrollment, (5) current suicidality, and (6) antibiotic or probiotic use at enrollment. Of note, some exclusion criteria were related to other aspects of the parent trial (eg, microbiome testing).

Procedures

Participants were recruited primarily through recovery record, social media posts on Facebook and X (formerly Twitter), and emails through a University of North Carolina participant registry listserve. After completing three logs on the recovery record and demonstrating initial engagement with the app, individuals were sent a message from the app with a brief description of the study and were invited to complete online consent forms followed by a screener for lifetime ED diagnosis using the ED100Kv2 [30,31]. Those who met all inclusion criteria were offered a second consent and the option to participate in the study. They were subsequently asked to complete a baseline questionnaire consisting of several demographic questions and measures assessing ED and general psychopathology (depression, anxiety, and attention-deficit or hyperactivity disorder screeners). Packages containing an Apple Watch (if they did not already have one) and sampling kits (for genetic and microbiome sampling) were then sent to participants within the first few days of enrollment and included a set of instructions to set up and configure recovery record on their Apple Watch.

Participants were asked to use a recovery record, a cognitive behavioral therapy-based application designed to support individuals with an ED, through an iPhone and an Apple Watch (first generation) with a version of the recovery record app designed specifically for the parent trial. Participants were asked to log ED urges and behaviors including binge eating and compensatory behaviors (vomiting, diuretic and laxative misuse, excessive exercise, and fasting) and their mood through the recovery record app on the Apple Watch, although these logs could also be completed on the iPhone app. In addition, participants logged their meals on the iPhone app rather than the Apple Watch given the larger screen. Skills (eg, distraction, mindful breathing, emotion regulation, and challenging negative thoughts) were also available for participants to use. Individuals who were already working with a clinician outside of the study could connect with them through the recovery record app; however, this was not included as a part of engagement data collection. If participants had technical difficulties with the app, they were instructed to reach out to the study team. Participants also had the ability to personalize the app and enable push

XSL•FO
RenderX

notifications or daily reminders; however, they were not required to enable these notifications as part of the study. All app content was accessible to participants throughout the duration of the study. Midpoint and endpoint questionnaires assessing ED, mood, and anxiety symptoms were administered 15 and 30 days, respectively, after enrollment.

Measures

Demographic information was collected via a questionnaire administered at baseline on age, gender, race, and ethnicity. Biological sex at birth was determined via saliva sample for the genetic testing component of the parent study [32]. Current BMI was calculated at baseline with self-reported height and weight. Lifetime ED diagnosis was determined by algorithm using items from the ED100Kv2 [30,31].

Information on current ED symptomatology was collected through the Eating Disorders Examination Questionnaire (EDE-Q) [33] administered in the baseline, midpoint, and endpoint questionnaires. The EDE-Q is a widely used self-report ED questionnaire and has demonstrated good validity and reliability in community samples [34]. In total, 28 items cover various aspects of ED pathology including weight, shape, and eating concerns, current BMI, and ED behaviors including binge eating and compensatory behavior frequency (vomiting, fasting, excessive exercise, diuretic, and laxative misuse) over the past 28 days. The EDE-Q global score [35,36] was calculated from the 22 Likert-scale items (0=no days, 6=every day for items assessing frequency OR 0=not at all, 6=markedly for items assessing distress or impairment).

Engagement

To extend the focus of existing literature on summary-level variables and broaden the types of engagement measured, we assessed the interactive features Recovery Record offers. In addition, since a central component of evidence-based treatment for binge-type EDs is self-monitoring of meals, mood, urges, and ED behaviors to help patients identify triggers and maintaining factors of the ED [33], participants were explicitly instructed to focus on logging these aspects through the app. All data used to describe types of engagement were collected through the recovery record app on both the iPhone and the Apple Watch. As such, the types and definitions of engagement used for this study, presented in Table 1, focus on use of meal, mood, and behavioral logs, overall usage, and through which mode of delivery (ie, iPhone or Apple Watch). All measures of engagement defined in Table 1 take an approach of mean usage (ie, the average number of times per day in 1 week that an individual used part of the app) to help capture change in engagement over time. To characterize how participants engaged with the app, each type of engagement was tabulated over week-long periods, thus participants can have up to 4 repeated measures of engagement for weeks 1-4. However, since participants received their Apple Watch devices approximately 1 week into the study, we only include data from weeks 2 - 4; therefore, participants have up to 3 repeated measures of engagement in the watch log models only. Week 1 data collection began the day after enrollment.

Table . Engagement terms and descriptions of engagement definitions, each used as the dependent variable in separate multilevel models.

Engagement measure	Definition of engagement
Log type	
Mean urge logs	Number of times a participant logged an urge during a day, averaged over a 7-day timespan
Mean eating behavior logs	Number of times a participant logged an eating behavior during a day, averaged over a 7-day timespan
Mean mood logs	Number of times a participant logged a mood during a day, averaged over a 7-day timespan
Mean meal logs	Number of times a participant logged a meal during a day, averaged over a 7-day timespan
Device type	
Mean phone logs	Number of times the app was used on the iPhone in any capacity during a day, averaged over a 7-day timespan
Mean watch logs[a]	Number of times the app was used on the Apple Watch in any capacity during a day, averaged over a 7-day timespan
Mean use	The number of times the app was used in any capacity on the iPhone or Watch during a day, averaged over a 7-day timespan

[a]Due to participants receiving their Apple Watch devices ~1 week into the study, we only report data from weeks 2 - 4.

Data Analysis

All data preparation and analyses were conducted using SAS 9.4 (SAS Institute Inc) [37]. To prepare data for analysis, we first screened for unrealistic values of binge eating and compensatory behavior episodes reported at baseline (ie, 500 episodes reported in the past 28 days), and winsorized those values. For model estimation and interpretation purposes and to protect the privacy of participants with demographic characteristics with cell sizes <5, we included male and female participants who reported data for all baseline predictor variables (age, sex, ethnicity, current BMI, and baseline binge-eating episodes). For the engagement data, we screened for and

XSL•FO
RenderX

removed duplicates and impossible or improbable measures of engagement (ie, future-dated timepoints, 1000 logs of the same event in 1 day); no imputation methods were used, since lack of engagement data at a given time point was not necessarily indicative of missingness.

To address aim 1, we first characterized the sample at baseline. Descriptive statistics (n's, percentages, means, and SDs as appropriate depending on variable type) were provided on demographic variables (age, gender, sex, race, ethnicity, and current BMI), ED diagnosis, baseline ED psychopathology (EDE-Q global scores and number of binge-eating and compensatory behavior episodes in the past 28 days as measured by the EDE-Q), and for each measure of engagement listed in Table 1. We also performed a Poisson regression of the total number of days participants used recovery record in any capacity. Baseline predictors (age, sex, ethnicity, current BMI, and number of binge-eating episodes at baseline) were used as independent variables. Race was not used due to small cell sizes.

For aim 2, we analyzed the engagement data using multilevel growth models due to the nested structure of the data (up to 4 repeated measures within individuals for all measures of engagement except for mean watch logs, which is up to 3 repeated measures of engagement due to participants receiving their devices by the end of week 1), using the types of engagement included in Table 1 as the dependent variables. First, we visualized the data by plotting each measure of engagement over time (measured in weeks) using spaghetti plots to identify what type of functional form should be used (ie, linear and piecewise) for each model. For each measure of engagement, we began with unconditional multilevel models using the multilevel model PROC MIXED function. If the spaghetti plots were unclear as to what functional form should be used, we compared the intraclass correlation (ICC), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values in the initial unconditional models to determine which functional form was most appropriate. Findings from the spaghetti plots and unconditional models were qualitatively summarized in the text to highlight relevant themes.

For aim 3, we expanded analyses to conditional models of engagement by including time (level 1 predictor measured in weeks) and the following baseline demographic predictors as time-invariant covariates (level 2 predictors): age, sex, ethnicity, and current BMI. Given that the parent sample recruited individuals with current binge eating, we then added baseline number of binge-eating episodes in the past 28 days from the EDE-Q as a measure of severity of illness to use as a level 2 time-invariant predictor of engagement. We mean centered age, current BMI, and baseline binge-eating episodes for ease of interpretation. For the categorical predictors, the reference variables were set to week 1 for time, participants who were categorized as male based on genotype for sex, and participants who self-identified as Hispanic for ethnicity. Predictors with significant fixed effects estimates were reported for each model, and themes are summarized in the text.

All multilevel models were fit using restricted maximum likelihood, the α level for fixed effects was set to 0.05, and we used Satterthwaite degrees of freedom approximations to reduce type 1 error rates [38]. We limited the models to fixed effects to help with estimation that was consistent across models, especially given that the predictors of interest were primarily level 2 time-invariant predictors within the context of this study and plan to further explore random effects in future analyses.

Ethical Considerations

All study procedures were approved by the University of North Carolina Biomedical institutional review board (IRB# 17-0242 and 20-3229). All research activities were conducted in compliance with ethical standards, ensuring the privacy and confidentiality of participants' data; all data were deidentified prior to analysis. Participants provided informed consent before initiating the study. They did not receive compensation for their participation, however, those who were sent an Apple Watch were able to keep the devices at the end of the study.

Results

Demographics and Sample Description

A total of 893 participants engaged with recovery record at least once during the study, had complete data for all baseline predictors, and were subsequently included in analyses. Participants in this study represented approximately three-fourths of the full sample recruited for the parent BEGIN study (n=1166). Of the 893 participants, 772 (86.5%) were assigned as female based on genotype and 121 (13.5%) as male; 680 (84.1%) self-identified as women, 120 (14.8%) as men, and 9 (1.1%) as nonbinary or third gender (n=84 did not report gender). The sample mostly identified as White (n=743/871, 85.3%), followed by more than 1 race (54/871, 6.2%), African American (35/871, 4.0%), Asian (33/871, 3.7%), and Native American or American Indian (6/871, 0.7%); and 22 did not report race. In addition, 92 of the 893 (10.3%) participants identified as Hispanic. The mean age of the sample was 29.6 (SD 7.4) years, and the mean current BMI was 32.5 (SD 9.8) kg/m^2.

Across eating disorder characteristics, 78.3% (699/893) met ED100Kv2 criteria for lifetime BED, 26.9% (240/893) for lifetime BN, and 18.7% (167/893) for lifetime AN (participants could meet criteria for more than 1 ED diagnosis). The mean EDE-Q global score was 3.93 (SD 1.01). The mean EDE-Q subscale scores were 3.65 (SD 1.25) for eating concern, 2.90 (SD 1.58) for restraint, 4.71 (SD 1.12) for shape concern, and 4.44 (SD 1.12) for weight concern. At baseline, participants reported a mean of 12.92 (SD 9.57) binge episodes, 3.00 (SD 9.46) vomiting episodes, 0.89 (SD 0.43) laxative or diuretic misuse episodes, and 4.16 (SD 6.71) compulsive exercise episodes in the past 28 days as captured by the EDE-Q.

For measures of engagement, the sample used the recovery record app for an average of 24.09 (SD 7.18) days out of 30, with (209/893, 23.4%) participants using the app all 30 days. Across all participants included in this study, a total of 225,927 recovery record logs were captured over 4 weeks. Refer to Table 2 for descriptive data on the total sum of logs for each engagement variable over 4 weeks as well as the mean number of logs per day for each engagement variable over 4 weeks and by week. Notably, the majority of logs were completed on the

XSL•FO
RenderX

iPhone (n=217,143/225,927 logs, 96.1%) rather than on the Apple Watch. After centering age, current BMI, and number of binge episodes at baseline, age (ß=.00, χ^2_1=5.74, false discovery rate [FDR]–adjusted *P*=.03) and sex (ß=.06, χ^2_1=8.18, FDR-adjusted *P*<.01) were significant predictors of the total number of days participants used the recovery record app. Specifically, participants who were older and female used the app a greater number of days.

Table . Total sum, median, and mean number of logs per day across 4 weeks and by week for each engagement variable.

Engagement variable	Total sum of logs over all 4 weeks, n	Logs per day over all 4 weeks, mean (SD)	Logs per day overall 4 weeks, median (IQR)	Logs per day during week 1 (n=893), mean (SD)	Logs per day during week 2 (n=835), mean (SD)	Logs per day during week 3 (n=784), mean (SD)	Logs per day during week 4 (n=681), mean (SD)	Percent change from week 1 to week 4, %
Log type								
Urge logs	14,855	0.7 (1.1)	0.3 (0.0-0.9)	0.8 (1.8)	0.8 (1.2)	0.6 (1.1)	0.4 (1.0)	−47.5
Behavior logs	25,420	1.1 (1.3)	0.7 (0.3-1.6)	1.5 (1.4)	1.3 (1.4)	0.1 (1.2)	0.7 (1.0)	−50.0
Mood logs	174,818	7.8 (6.4)	6.0 (2.9-9.6)	9.8 (6.5)	8.3 (6.5)	7.0 (6.1)	5.7 (5.5)	−41.6
Meal logs	67,043	3.0 (1.8)	3.0 (1.4-4.6)	3.6 (1.6)	3.1 (1.8)	2.8 (1.8)	2.4 (1.8)	−34.4
Device type								
Phone logs	217,143	9.7 (8.2)	8.0 (3.7-13.1)	12.4 (8.6)	10.2 (8.3)	8.5 (7.6)	7.0 (6.9)	−43.7
Watch logs[a]	6977	0.4 (0.9)	0.0 (0.0-0.4)	—[b]	0.7 (1.1)	0.4 (0.8)	0.2 (0.6)	−67.7
All logs	225,927	10.1 (8.4)	8.3 (3.9-13.7)	12.7 (8.8)	10.9 (8.6)	8.9 (7.8)	7.2 (7.1)	−43.3

[a]Due to participants receiving their Apple Watch devices ~1 week into the study, we only report watch log data from weeks 2 - 4.

[b]Not available.

Characterization of Engagement

To illustrate participant engagement over the course of 4 weeks (3 weeks for the watch logs) and to evaluate the functional forms to be used for subsequent conditional multilevel models, spaghetti plots for each measure of engagement were created using the mean number of logs over each week, seen in Multimedia Appendix 1. Across each measure, there was a general downward and linear trajectory of engagement across time. The mean meal logs plot had the most variability in individual trajectories and was initially difficult to discern a clear functional form. However, after visualizing the data via spaghetti plots in smaller groups of participants (ie, n=100) in combination with the data from Table 2, both linear and quadratic functional forms were tested in unconditional models. After comparing the ICC, AIC, and BIC, a linear functional form was determined to be the best fit for each measure of engagement and was used in the subsequent multilevel models detailed in the next section. Model fit statistics (ICC, AIC, and BIC) can be found for each model in Table S1 in Multimedia Appendix 2.

Predictors of Engagement

Table 3 presents the results from the conditional models evaluating which baseline characteristics were significant predictors of engagement. The first set of models used demographics as predictors; the second set of models added an additional predictor, baseline binge episodes, which was used as a measure of illness severity. Age and current BMI regression coefficients and standard errors were calculated based on 5-year and 5 kg/m^2 differences, respectively, to facilitate interpretation for more meaningful differences in engagement. However, baseline binge episodes were not altered given previous research that illustrates 1 binge episode per week results in a meaningful clinical difference in psychopathology for individuals with binge-type eating disorders [39].

Time, measured in weeks, was a significant negative predictor in nearly every instance, indicating that the number of weeks into the study consistently predicted a decline in engagement, regardless of how it was measured. Across demographic characteristics, sex and ethnicity were significant, positive predictors of mean mood, meal, and phone logs, as well as mean use. Mean-centered age was a significant positive predictor of mean meal log engagement such that those who were older logged more meals, and mean-centered current BMI was a significant negative predictor of mean urge logs such that those with higher BMIs at baseline logged fewer urges.

In the second set of models, the number of binge episodes at baseline was a significant, positive predictor of mean behavior logs. However, the number of binge episodes at baseline was not a significant predictor of any other measure of engagement, and there were no changes to significance for any other predictors when this variable was added to each engagement model. Notably, the model fit statistics and parameter estimates demonstrated negligible changes between models 1 and 2 for all measures of engagement.

XSL•FO
RenderX

Table . Multilevel model summaries for baseline demographic and severity of illness variables predicting engagement measures. The reference values for categorical variables were set to week 1, males, and individuals who identified as Hispanic. The watch log models only use data aggregated from weeks 2, 3, and 4 since most participants received their Apple Watch devices by the end of week 1. *P* values are presented as false discovery rate-corrected *P* values. Bolded values are significant.

Engagement variable	Week			Age[a]	Sex (female)	Ethnicity (Hispanic)	Current BMI[a]	Binge episodes
	2 b (95% CI)	3 b (95% CI)	4 b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)
Mean urge logs								
M1[b]	0.0 (−0.1 to 0.1)	−0.3[d] (−0.3 to −0.2)	−0.5 (−0.5 to −0.4)	0.0 (0.0 to 0.0)	0.1 (−0.1 to 0.3)	0.0 (−0.2 to 0.2)	−0.1[d] (−0.1 to 0.0)	—[e]
M2[c]	0.0 (−0.1 to 0.1)	−0.3[d] (−0.3 to −0.2)	−0.5[d] (−0.5 to −0.4)	0.0 (0.0 to 0.0)	0.1 (−0.1 to 0.3)	0.0 (−0.2 to 0.2)	−0.1[d] (0.0 to 0.0)	0.0 (0.0 to 0.0)
Mean behavior logs								
M1	−0.3[d] (−0.3 to −0.2)	−0.6[d] (−0.7 to −0.5)	−0.9[d] (−1.0 to −0.8)	0.0 (0.0 to 0.1)	0.0 (−0.2 to 0.2)	0.2 (−0.1 to 0.4)	0.0 (−0.1 to 0.0)	—
M2	−0.3 (−0.3 to −0.2)	−0.6 (−0.7 to −0.5)	−0.9 (−1.0 to −0.8)	0.0 (0.0 to 0.1)	0.0 (−0.2 to 0.2)	0.2 (−0.1 to 0.4)	0.0 (−0.1 to 0.0)	0.0[d] (0.0 to 0.0)
Mean mood logs								
M1	−1.5[d] (−1.7 to −1.2)	−2.8[d] (−3.1 to −2.6)	−4.3[d] (−4.6 to −4.1)	0.2 (0.0 to 0.4)	1.9[d] (1.0 to 2.8)	1.6[d] (0.6 to 2.6)	−0.1 (−0.3 to 0.0)	—
M2	−1.5[d] (−1.7 to −1.2)	−2.8[d] (−3.1 to −2.6)	−4.3[d] (−4.6 to −4.1)	0.2 (0.0 to 0.4)	1.9[d] (1.0 to 2.8)	1.6[d] (0.6 to 2.6)	−0.1 (−0.3 to 0.0)	0.0 (0.0 to 0.0)
Mean meal logs								
M1	−0.6[d] (−0.7 to −0.5)	−1.0[d] (−1.1, −0.9)	−1.6[d] (−1.7 to −1.5)	0.1[d] (0.0 to 0.2)	0.6[d] (0.3 to 0.9)	0.6[d] (0.3 to 0.9)	0.0 (−0.1 to 0.0)	—
M2	−0.6[d] (−0.7 to −0.5)	−1.0[d] (−1.1 to −0.9)	−1.6[d] (−1.7 to −1.5)	0.1[d] (0.0 to 0.2)	0.6[d] (0.3 to 0.9)	0.6[d] (0.3 to 0.9)	0.0 (−0.1 to 0.0)	0.0 (0.0 to 0.0)
Mean phone logs								
M1	−2.5[d] (−2.9 to −2.1)	−4.6[d] (−5.0 to −4.2)	−6.7[d] (−7.1 to −6.3)	0.3 (0.1 to 0.6)	2.8[d] (1.5 to 4.2)	2.2[d] (0.7 to 3.8)	−0.2 (−0.5 to 0.0)	—
M2	−2.5[d] (−2.9 to −2.1)	−4.6[d] (−5.0 to −4.2)	-6.7[d] (−7.1 to −6.3)	0.3 (0.1 to 0.6)	2.8[d] (1.4 to 4.2)	2.2[d] (0.7 to 3.8)	−0.2 (−0.5 to 0.0)	0.0 (0.0 to 0.1)
Mean watch logs								
M1	—	−0.3[d] (−0.3 to −0.2)	−0.5[d] (−0.5 to −0.4)	0.0 (0.0 to 0.1)	−0.1 (−0.2 to 0.1)	−0.1 (−0.3 to 0.1)	0.0 (0.0 to 0.0)	—
M2	—	−0.3[d] (−0.3 to −0.2)	−0.5[d] (−0.5 to −0.4)	0.0 (0.0 to 0.1)	−0.1 (−0.2 to 0.1)	−0.1 (−0.3 to 0.0)	0.0 (0.0 to 0.0)	0.0 (0.0 to 0.0)
Mean use								
M1	−2.2[d] (−2.6 to 1.8)	4.5[d] (−4.9 to −4.1)	−6.8[d] (−7.2 to −6.4)	0.3 (0.0 to 0.6)	2.8[d] (1.4 to 4.2)	2.2[d] (0.6 to 3.8)	−0.2 (−0.5 to 0.0)	—

XSL•FO
RenderX

Engagement variable	Week			Age[a]	Sex (female)	Ethnicity (Hispanic)	Current BMI[a]	Binge episodes
	2 b (95% CI)	3 b (95% CI)	4 b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)
M2	−2.2[d] (−2.6 to 1.8)	4.5[d] (−4.9 to −4.1)	−6.8[d] (−7.3 to −6.4)	0.3 (0.0 to 0.6)	2.8[d] (1.3 to 4.2)	2.2[d] (0.6 to 3.8)	−0.2 (−0.5 to 0.0)	.03 (−0.02 to 0.08)

[a]Age and current BMI regression coefficients and standard errors were calculated based on 5-year and 5kg/m^2 differences, respectively, to facilitate interpretation for more meaningful differences in engagement.

[b]M1: model 1 presents the conditional models using demographic variables as predictors.

[c]M2: model 2 presents conditional models with the demographic variables and baseline binge episodes as predictors.

[d]Indicates values that are significant.

[e]Not applicable.

Discussion

Principal Findings

This study described various measures of engagement with a digital ED app, recovery record, and deepened our understanding of how individuals with binge eating use different components and delivery methods of the app. All measures of engagement declined over the course of the study, consistent with trends observed in other digital ED and mental health interventions; however, participants engaged with the app for an average of 3.5 weeks, which was greater than expected given that most digital mental health interventions observe engagement over fewer days [40-42]. Several baseline variables emerged as significant predictors of unique measures of engagement, highlighting the importance of more nuanced assessments of engagement with digital ED interventions. Findings for each aim are discussed in turn below, followed by notable discussion points for measures of engagement and study limitations.

Although every measure of engagement declined, the percentage of participants in this study that were still using recovery record at the end of 4 weeks (209/893, 23%) was substantially greater than that observed in other self-monitoring apps (6%) [40]. Across log types, mean behavior and urge logs had the largest percent reduction over 4 weeks, which could reflect both the overall decline in engagement in combination with decreased ED symptomatology as observed in the BEGIN feasibility study [29]. In addition, participants logged an average of 3 meals/day, and mean meal logs had the smallest percent reduction (1-[2.35/3.58], 34%) of any measure of engagement over the course of 4 weeks, illustrating that most participants were on track with a regular eating treatment target in cognitive behavioral ED treatments [33].

Mood logs were the most frequently used logs, accounting for 62% of all logs, which is in part explained by participants having the ability to create a unique mood log for each emotion as part of meal or behavior log, or as a separate mood log altogether. The greater number of opportunities to log moods in addition to the more limited nature of when meals and ED behaviors or urges occur may in large part explain the vast difference in sample sizes between log types. When measuring engagement across device type, participants completed substantially more phone logs than watch logs, which could partially be explained by meal logs only being offered on the phone. However, the percent reduction in watch logs was 35% greater than the percent reduction of phone logs over the last 3 weeks, demonstrating that the decline in watch engagement was much steeper than the phone engagement. Although the watch app was designed to improve discreetness of completing logs while simultaneously enhancing the usability of and engagement with the app, qualitative research may be necessary to understand what maintained participants' engagement on their phones more than on the watch.

Second, the engagement trajectory visualizations illustrated that most of the engagement measures had similar overall downward trajectories. In conjunction with weekly summary data presented in Table 2, the spaghetti plots exposed the individual variability in engagement trajectories and underscored that not all components of the app were used the same over the course of a month. Although each measure of engagement used a linear functional form for subsequent multilevel models, the most variation in individual trajectories was observed in mean meal logs, which could vary tremendously on an individual's availability to complete a more time-intensive log. As an extension of this aim in future work, engagement visualizations may serve useful in characterizing subgroups of engagement profiles through a repeated measures latent profile analysis and can subsequently be used to identify unique engagement and symptom profiles associated with positive intervention outcomes (Peiper et al [43] for an example in treatment-resistant depression). Subsequent studies may also seek to visualize digital intervention usage at more granular levels in terms of time and within individuals, which would be essential to tailor interventions to individual users and to provide just-in-time adaptive interventions that are responsive to engagement.

Third, the multilevel models yielded results that primarily illustrated that time was the best predictor of engagement. Time, measured in weeks, was a significant predictor in every model, and the addition of other predictors to the models typically did not improve model fit. This result was unsurprising because a consistent theme in digital intervention literature is that engagement drops as time progresses [40,44]. Beyond time, sex and ethnicity were the most common significant baseline predictors of engagement. Those who were assigned as female based on genotyping or self-identified as Hispanic were more engaged than their male and non-Hispanic counterparts, respectively, for 4 measures of engagement: mean meal logs,

XSL•FO
RenderX

mean mood logs, mean phone logs, and mean use. Although there is little research evaluating baseline predictors of engagement in digital ED interventions, 1 recent study found that sex and ethnicity did not predict engagement when defined as a dichotomous variable of greater or less than 10 minutes of app usage [45]. However, there is some evidence to suggest sex and ethnicity are significant predictors of engagement in other digital mental health interventions [eg, 46,47]. A key point to consider is how the design of the app may have biased engagement toward demographic subgroups (eg, cultural adaptations may increase engagement of racial or ethnic groups). Even though acceptance-facilitating interventions can increase participants' acceptance, motivations, and positive attitudes toward digital ED interventions regardless of demographic characteristics, initial work in this area highlights that engagement does not improve [48], underscoring the importance of using user-centered design principles for improving engagement for target demographics [49].

A second set of baseline characteristics included significant predictors of only 1 measure of engagement each. Current BMI was a significant negative predictor of mean urge logs, and this result could be interpreted as those with higher BMIs may not have experienced as many urges or did not log their urges as often as those with lower BMIs. Similarly, the number of baseline binge episodes was a significant predictor of the mean behaviors logged, possibly reflecting that those who had greater mean behavior logs over time were more engaged in the app. However, an alternative explanation is that participants with a higher number of baseline binge episodes had more opportunities to log behaviors. To disentangle these results in future studies, it will be essential to assess and compare retrospective self-reported urges and behaviors over a given interval to the behaviors logged in the moment via digital intervention. Notably, the addition of binge episodes at baseline to the second set of engagement models did not improve fit statistics, illustrating little to no contribution to improvement in engagement prediction.

Finally, age was only a significant predictor of mean meal logs, with older individuals logging more meals than younger counterparts, which was somewhat surprising given the technological literacy of younger generations. Although age, current BMI, and binge episodes at baseline were only significant predictors in 1 engagement model each, these results support findings from machine learning algorithms recently identifying age and measures of ED symptom severity as significant predictors of engagement with and drop out from another digital ED intervention [50]. An important caveat to these 3 results is that the parameter estimates for age, current BMI, and number of binge episodes at baseline hovered around 0, warranting future research as to their clinical and practical use.

Three additional points are worth noting. First, participants used the phone to engage with Recovery Record far more than the watch. Despite being confounded by the recruitment of previous recovery record users who only had access to the phone app, the percent change and more consistent engagement with the phone illustrate that participants were more inclined to use this method of delivery. It is possible that participants may have been less used to wearing or using wearable technology, thus turning to the iPhone version of the app more often. Relatedly, more app content (ie, meal logs) was available through the iPhone app, which also may have contributed to the engagement patterns observed across devices. The models of mean phone and watch logs also demonstrated that 2 baseline characteristics (sex and ethnicity) may be used to help differentiate subsequent engagement depending on the delivery platform, suggesting that design is important for delivering ED interventions through wearables and smartphones. Second, meal logs were unique compared with the other measures of engagement for a few reasons: this measure had the smallest percent reduction over 4 weeks, demonstrated the most variability of participant trajectories as evidenced by the spaghetti plots, and was significantly predicted by 3 baseline demographic predictors (age, sex, and ethnicity). Although this study is an exploratory investigation into expanding how we define engagement, this collection of findings warrants future replication and qualitative studies in other samples with ED psychopathology to evaluate if and why meal logs are consistently and proportionally more used than other components. Finally, an important point about mean use is that several multilevel models had the same significant baseline predictors, suggesting that overall engagement results may be reflecting groups of individuals with similar characteristics engaging with specific digital intervention functions in similar capacities. This may also mean that using overall measures of engagement obscures other results that are more sensitive or have less power. Taken together, our findings underscore that engagement with a digital ED intervention is more nuanced and complex than this research often presents when describing engagement through a single measure and is worth deeper exploration to optimize what individuals gain from using a digital ED intervention.

Limitations

Overall engagement with various functions of the app was high, particularly the number of days used. A key limitation that could partially explain this observation is that many participants who entered the study had already used the app, and they were required to complete 3 logs in Recovery Record before enrolling in the study. Therefore, the baseline level of engagement may be higher than what would be observed outside of a research study. In addition, participants' greater likelihood of engaging could be due to the convenience and accessibility of using the app on both the phone and the watch coupled with the relative ease of completing simple functions compared with other digital eating disorder interventions (ie, logging a behavior is easier and quicker than completing a guided self-help session in an app). Future studies may consider comparing the same granularity of engagement with new and existing users of digital interventions. In addition, the measures of engagement had significant overlap in the data used to test the multilevel models, so it was unsurprising to observe patterns across the significance of baseline predictors. To address this in future work, it may be useful to conduct split-half studies where the discovery sample would identify significant predictors and these models would be tested with the replication sample. Finally, another limitation is that the 4 measures of engagement where sex and ethnicity were significant predictors of engagement had the largest sample

XSL•FO
RenderX

size of logs, which could indicate that these models were overpowered. An alternative explanation is that enough data were acquired to consistently detect sex and ethnicity as significant predictors, and the amount of data collected in the other 3 models was insufficient or lacked sufficient variability to detect differences. Future studies may seek to replicate the current findings before evaluating thresholds to determine what baseline characteristics are clinically useful in identifying meaningful changes in engagement. Finally, the findings of this study should be considered in the context that the sample was primarily White, non-Hispanic women with binge-type EDs, so generalizability may be limited to broader demographics and other EDs. Relatedly, engagement associated with gender was not evaluated in this study; however, understanding the engagement patterns of gender-diverse populations will be critical to tailoring content that reflects their experiences and facilitating acceptance of the intervention [51].

Conclusion

This study provided a novel view of engagement that characterized participants' usage of different functions, method of delivery, and overall usage of a digital ED intervention. Key predictors, time, and 2 demographic predictors (sex and ethnicity) were consistently significant despite unique measurements of engagement across log type, device type, and overall engagement. Other baseline demographic and severity of illness predictors were significant in only 1 measure of engagement each, highlighting opportunities to tease out more complex and nuanced understanding of the use of and engagement with different functions of a digital ED app. Future work may consider mixed qualitative and quantitative approaches to better understand and enhance engagement. In addition, identifying unique engagement profiles that, in combination with baseline characteristics, can be used to predict intervention outcomes may allow researchers and clinicians to intervene earlier on engagement and harmful eating behaviors. Considering the importance of consistent engagement in traditional psychotherapy, which is required for meaningful symptom change, this study sets the stage for understanding what types of engagement with digital ED interventions may be most helpful and can reliably predict change in symptoms.

Acknowledgments

We are deeply grateful to all the individuals who, through their participation, made BEGIN a success. Funding was provided by the National Science Foundation (DGE-1650116; REF, principal investigator); Foundation of Hope, Raleigh, North Carolina (CMB, principal investigator); National Eating Disorders Association (CMB and JT, principal investigators); Brain and Behavior Research Foundation (NARSAD Distinguished Investigator Grant; CMB, principal investigator); National Institute of Mental Health (R01MH119084, CMB/Butner, MPIs; U01 MH109528, Sullivan principal investigator, CMB Co-investigator). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Apple Watch devices were provided to the University of North Carolina at Chapel Hill by Apple Inc. at no cost to the study. No funding bodies were involved in the design of the study and collection, analysis, interpretation of data, or writing the manuscript.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to their proprietary nature, as designated by Recovery Record. Data may be obtained from the corresponding author on reasonable request.

Authors' Contributions

REF contributed to conceptualization, data curation, formal analysis, investigation, project administration, resources, visualization, writing – original draft, and writing – review & editing. LMT managed conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, writing – original draft, and writing – review & editing. JT handled conceptualization, resources, and writing – review & editing. SA contributed conceptualization, resources, and writing – review & editing. CMB handled conceptualization, formal analysis, funding acquisition, methodology, supervision, writing – original draft, and writing – review & editing.

Conflicts of Interest

CM Bulik reports: Pearson (author, royalty recipient). J Tregarthen and S Argue report: Recovery Record (shareholders, employees). These relationships have been disclosed to and are under management by UNC-Chapel Hill. Tregarthen and Argue were not involved in the study design or data analysis. All participants were notified of their involvement in the study in the consent forms: "In addition, Jenna Tregarthen, the external site PI on this study, and her spouse, Stuart Argue, the external study coordinator, have ownership interest in the Recovery Record technology being evaluated and utilized in this research and have ownership interest in Recovery Record, the company developing the technology platform. If this technology is successful at some point in the future, Jenna Tregarthen and Stuart Argue may receive financial benefits."

Multimedia Appendix 1
Spaghetti plots of each engagement measure across 4 weeks.

XSL•FO
RenderX

[DOCX File, 2813 KB - mhealth_v13i1e68824_app1.docx]

Multimedia Appendix 2
Model fit statistics for baseline predictors of engagement.
[DOCX File, 16 KB - mhealth_v13i1e68824_app2.docx]

References

1. Kazdin AE, Fitzsimmons-Craft EE, Wilfley DE. Addressing critical gaps in the treatment of eating disorders. Int J Eat Disord 2017 Mar;50(3):170-189. [doi: 10.1002/eat.22670] [Medline: 28102908]
2. Hart LM, Granillo MT, Jorm AF, Paxton SJ. Unmet need for treatment in the eating disorders: a systematic review of eating disorder specific treatment seeking among community cases. Clin Psychol Rev 2011 Jul;31(5):727-735. [doi: 10.1016/j.cpr.2011.03.004] [Medline: 21501580]
3. Tregarthen JP, Lock J, Darcy AM. Development of a smartphone application for eating disorder self-monitoring. Int J Eat Disord 2015 Nov;48(7):972-982. [doi: 10.1002/eat.22386] [Medline: 26213130]
4. Kass AE, Balantekin KN, Fitzsimmons-Craft EE, Jacobi C, Wilfley DE, Taylor CB. The economic case for digital interventions for eating disorders among United States college students. Int J Eat Disord 2017 Mar;50(3):250-258. [doi: 10.1002/eat.22680] [Medline: 28152203]
5. Smith KE, Juarascio A. From ecological momentary assessment (EMA) to ecological momentary intervention (EMI): past and future directions for ambulatory assessment and interventions in eating disorders. Curr Psychiatry Rep 2019 Jun 4;21(7):53. [doi: 10.1007/s11920-019-1046-8] [Medline: 31161276]
6. Linardon J, Shatte A, Messer M, Firth J, Fuller-Tyszkiewicz M. E-mental health interventions for the treatment and prevention of eating disorders: An updated systematic review and meta-analysis. J Consult Clin Psychol 2020 Nov;88(11):994-1007. [doi: 10.1037/ccp0000575] [Medline: 32852971]
7. Taylor CB, Graham AK, Flatt RE, Waldherr K, Fitzsimmons-Craft EE. Current state of scientific evidence on Internet-based interventions for the treatment of depression, anxiety, eating disorders and substance abuse: an overview of systematic reviews and meta-analyses. Eur J Public Health 2021 Jul 7;31(31 Suppl 1):i3-i10. [doi: 10.1093/eurpub/ckz208] [Medline: 32918448]
8. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
9. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766. [doi: 10.2196/30766] [Medline: 35099398]
10. Hudson JI, Hiripi E, Pope HG Jr, Kessler RC. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication [published correction appears in Biol Psychiatry. 2012 Jul 15;72(2):164]. Biol Psychiatry 2012 Jul 15;61(3):348-358. [doi: 10.1016/j.biopsych.2006.03.040]
11. Smink FRE, van Hoeken D, Hoek HW. Epidemiology of eating disorders: incidence, prevalence and mortality rates. Curr Psychiatry Rep 2012 Aug;14(4):406-414. [doi: 10.1007/s11920-012-0282-y] [Medline: 22644309]
12. Kessler RC, Berglund PA, Chiu WT, et al. The prevalence and correlates of binge eating disorder in the World Health Organization World Mental Health Surveys. Biol Psychiatry 2013 May 1;73(9):904-914. [doi: 10.1016/j.biopsych.2012.11.020] [Medline: 23290497]
13. Welch E, Jangmo A, Thornton LM, et al. Treatment-seeking patients with binge-eating disorder in the Swedish national registers: clinical course and psychiatric comorbidity. BMC Psychiatry 2016 May 26;16:163. [doi: 10.1186/s12888-016-0840-7] [Medline: 27230675]
14. Berkman ND, Lohr KN, Bulik CM. Outcomes of eating disorders: a systematic review of the literature. Int J Eat Disord 2007 May;40(4):293-309. [doi: 10.1002/eat.20369] [Medline: 17370291]
15. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37. [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
16. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
17. Venkatesh A, Chang A, Green EA, et al. Perceived facilitators and barriers to engaging with a digital intervention among those with food insecurity, binge eating, and obesity. Nutrients 2021 Jul 19;13(7):2458. [doi: 10.3390/nu13072458] [Medline: 34371967]
18. Nitsch M, Dimopoulos CN, Flaschberger E, et al. A guided online and mobile self-help program for individuals with eating disorders: an iterative engagement and usability study. J Med Internet Res 2016 Jan 11;18(1):e7. [doi: 10.2196/jmir.4972] [Medline: 26753539]

XSL•FO
RenderX

19. Anastasiadou D, Folkvord F, Serrano-Troncoso E, Lupiañez-Villanueva F. Mobile health adoption in mental health: user experience of a mobile health app for patients with an eating disorder. JMIR Mhealth Uhealth 2019 May 31;7(6):e12920. [doi: 10.2196/12920] [Medline: 31199329]
20. Watson HJ, Levine MD, Zerwas SC, et al. Predictors of dropout in face-to-face and internet-based cognitive-behavioral therapy for bulimia nervosa in a randomized controlled trial. Int J Eat Disord 2017 May;50(5):569-577. [doi: 10.1002/eat.22644] [Medline: 27862108]
21. Fitzsimmons-Craft EE, Taylor CB, Graham AK, et al. Effectiveness of a digital cognitive behavior therapy-guided self-help intervention for eating disorders in college women: a cluster randomized clinical trial. JAMA Netw Open 2020 Aug 3;3(8):e2015633. [doi: 10.1001/jamanetworkopen.2020.15633] [Medline: 32865576]
22. Kim JP, Sadeh-Sharvit S, Welch HA, Neri E, Tregarthen J, Lock J. Eating disorders early app use mediates treatment effect on clinical improvement. Int J Eat Disord 2022 Mar;55(3):382-387. [doi: 10.1002/eat.23652] [Medline: 34904745]
23. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
24. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 1;70(7):538-544. [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
25. Linardon J, King T, Shatte A, Fuller-Tyszkiewicz M. Usability evaluation of a cognitive-behavioral app-based Intervention for Binge Eating and Related psychopathology: a qualitative study. Behav Modif 2022 Sep;46(5):1002-1020. [doi: 10.1177/01454455211021764] [Medline: 34075803]
26. Linardon J, Messer M, Shatte A, et al. Does the method of content delivery matter? Randomized controlled comparison of an internet-based intervention for eating disorder symptoms with and without interactive functionality. Behav Ther 2022 May;53(3):508-520. [doi: 10.1016/j.beth.2021.12.001] [Medline: 35473653]
27. Lindgreen P, Lomborg K, Clausen L. Patient experiences using a self-monitoring app in eating disorder treatment: qualitative study. JMIR Mhealth Uhealth 2018 Jun 22;6(6):e10253. [doi: 10.2196/10253] [Medline: 29934285]
28. Bulik CM, Butner JE, Tregarthen J, et al. The Binge Eating Genetics Initiative (BEGIN): study protocol. BMC Psychiatry 2020 Jun 16;20(1):307. [doi: 10.1186/s12888-020-02698-7] [Medline: 32546136]
29. Flatt RE, Thornton LM, Smith T, et al. Retention, engagement, and binge-eating outcomes: Evaluating feasibility of the Binge-Eating Genetics Initiative study. Int J Eat Disord 2022 Aug;55(8):1031-1041. [doi: 10.1002/eat.23726] [Medline: 35502471]
30. Watson HJ, Yilmaz Z, Thornton LM, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat Genet 2019 Aug;51(8):1207-1214. [doi: 10.1038/s41588-019-0439-2] [Medline: 31308545]
31. Thornton LM, Munn-Chernoff MA, Baker JH, et al. The Anorexia Nervosa Genetics Initiative (ANGI): Overview and methods. Contemp Clin Trials 2018 Nov;74:61-69. [doi: 10.1016/j.cct.2018.09.015] [Medline: 30287268]
32. Igudesman D, Abbaspour A, Reed KK, et al. Laxative abuse is associated with a depleted gut microbial community structure among women and men with binge-eating disorder or bulimia nervosa: the Binge Eating Genetics Initiative. Psychosom Med 2023 Oct 1;85(8):727-735. [doi: 10.1097/PSY.0000000000001226] [Medline: 37363967]
33. Fairburn CG. Cognitive Behavior Therapy and Eating Disorders: Guilford Press; 2008.
34. Mond JM, Hay PJ, Rodgers B, Owen C, Beumont PJV. Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. Behav Res Ther 2004 May;42(5):551-567. [doi: 10.1016/S0005-7967(03)00161-X] [Medline: 15033501]
35. Carey M, Kupeli N, Knight R, Troop NA, Jenkinson PM, Preston C. Eating Disorder Examination Questionnaire (EDE-Q): Norms and psychometric properties in U.K. females and males. Psychol Assess 2019 Jul;31(7):839-850. [doi: 10.1037/pas0000703] [Medline: 30802119]
36. Allen KL, Byrne SM, Lampard A, Watson H, Fursland A. Confirmatory factor analysis of the Eating Disorder Examination-Questionnaire (EDE-Q). Eat Behav 2011 Apr;12(2):143-151. [doi: 10.1016/j.eatbeh.2011.01.005] [Medline: 21385645]
37. SAS/STAT® software: version 94. : SAS Institute, Inc; 2014.
38. Luke SG. Evaluating significance in linear mixed-effects models in R. Behav Res Methods 2017 Aug;49(4):1494-1502. [doi: 10.3758/s13428-016-0809-y] [Medline: 27620283]
39. Trace SE, Thornton LM, Root TL, et al. Effects of reducing the frequency and duration criteria for binge eating on lifetime prevalence of bulimia nervosa and binge eating disorder: implications for DSM-5. Int J Eat Disord 2012 May;45(4):531-536. [doi: 10.1002/eat.20955] [Medline: 21882218]
40. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement With mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
41. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338. [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]

XSL•FO
RenderX

42. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869. [doi: 10.2196/12869] [Medline: 31333198]
43. Peiper NC, Nelson BW, Aschbacher K, Forman-Hoffman VL. Trajectories of depression symptoms in a therapist-supported digital mental health intervention: a repeated measures latent profile analysis. Soc Psychiatry Psychiatr Epidemiol 2023 Aug;58(8):1237-1246. [doi: 10.1007/s00127-022-02402-y] [Medline: 36651947]
44. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019;2:54. [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
45. Kasson E, Vázquez MM, Li X, et al. A pilot study of factors influencing engagement with an mHealth intervention among teens with eating disorder symptoms. J Technol Behav Sci 2025 Jun;10(2):424-434. [doi: 10.1007/s41347-024-00444-4] [Medline: 40895234]
46. Marinova N, Rogers T, MacBeth A. Predictors of adolescent engagement and outcomes - A cross-sectional study using the togetherall (formerly Big White Wall) digital mental health platform. J Affect Disord 2022 Aug 15;311:284-293. [doi: 10.1016/j.jad.2022.05.058] [Medline: 35588912]
47. Arnold C, Farhall J, Villagonzalo KA, Sharma K, Thomas N. Engagement with online psychosocial interventions for psychosis: A review and synthesis of relevant factors. Internet Interv 2021 Sep;25:100411. [doi: 10.1016/j.invent.2021.100411] [Medline: 34401370]
48. Linardon J, Anderson C, Chapneviss T, Hants E, Shatte A, Fuller-Tyszkiewicz M. Effects of an acceptance-facilitating intervention on acceptance and usage of digital interventions for binge eating. Psychiatr Serv 2022 Oct 1;73(10):1173-1176. [doi: 10.1176/appi.ps.202100616] [Medline: 35354324]
49. Graham AK, Kosmas JA, Massion TA. Designing digital interventions for eating disorders. Curr Psychiatry Rep 2023 Apr;25(4):125-138. [doi: 10.1007/s11920-023-01415-x] [Medline: 36928767]
50. Linardon J, Fuller-Tyszkiewicz M, Shatte A, Greenwood CJ. An exploratory application of machine learning methods to optimize prediction of responsiveness to digital interventions for eating disorder symptoms. Int J Eat Disord 2022 Jun;55(6):845-850. [doi: 10.1002/eat.23733] [Medline: 35560256]
51. Fowler JA, Buckley L, Muir M, et al. Digital mental health interventions: A narrative review of what is important from the perspective of LGBTQIA+ people. J Clin Psychol 2023 Nov;79(11):2685-2713. [doi: 10.1002/jclp.23571] [Medline: 37528773]

Abbreviations

AIC: Akaike Information Criterion
AN: anorexia nervosa
BED: binge-eating disorder
BEGIN: Binge Eating Genetics Initiative
BIC: Bayesian Information Criterion
BN: bulimia nervosa
ED: eating disorder
EDE-Q: Eating Disorder Examination Questionnaire v6
FDR: False Discovery Rate
ICC: intraclass correlation

Edited by M Waring; submitted 14.11.24; peer-reviewed by E Soderstrom, H Szlyk; revised version received 09.07.25; accepted 11.07.25; published 30.09.25.

Please cite as:
Flatt RE, Thornton LM, Tregarthen J, Argue S, Bulik CM
How Engagement Changes Over Time in a Digital Eating Disorder App: Observational Study
JMIR Mhealth Uhealth 2025;13:e68824
URL: https://mhealth.jmir.org/2025/1/e68824
doi:10.2196/68824

© Rachael E Flatt, Laura M Thornton, Jenna Tregarthen, Stuart Argue, Cynthia M Bulik. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Hybrid Health IT and Telehealth–Delivered Behavioral Weight Loss Services for Primary Care Patients With Cardiovascular Risk Factors: Intervention Component Design and Pragmatic Randomized Feasibility Trial

Ronald T Ackermann[1,2,3], MD, MPH; Kenzie A Cameron[1,3,4], PhD; David T Liss[1,3], PhD; Nancy Dolan[1,2], MD; Cassandra Aikman[3], MPH; Amy R Carson[5], RD, CPT; Sterling A Harris[3], BA; Kathryn Doyle[2], MS; Andrew J Cooper[1,3], MS, MPH; Brian Hitsman[3,4], PhD

[1]Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Northwestern Medicine, Chicago, IL, United States

[3]Institute for Public Health and Medicine, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Drive, Suite 609, Chicago, IL, United States

[4]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[5]Fitness Formula Clubs, Chicago, IL, United States

Corresponding Author:
Ronald T Ackermann, MD, MPH
Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Abstract

Background: Intensive lifestyle interventions (ILI) improve weight loss and cardiovascular risk factors, but health care systems face challenges in implementing them. We engaged stakeholders to cocreate and evaluate primary care implementation strategies for ILI components.

Objective: This study aimed to describe the design of intervention components and implementation strategies and to evaluate the feasibility of pragmatic trial enrollment and randomization procedures, as well as the acceptability and preliminary effectiveness of the interventions.

Methods: The study setting was a single, urban primary care office. Patients with a BMI ≥27 kg/m² and ≥1 cardiovascular risk factor were sent a single electronic health record (EHR) message between December 2019 and January 2020 offering services to support a weight loss goal of 10 pounds in 10 weeks. All patients who affirmed interest were pragmatically enrolled in a trial offering basic lifestyle support (BLS), which provided a scale that transmits weight data to the EHR using cellular networks, a coupon to enroll in lifestyle coaching resources through a partnering fitness organization, and periodic EHR messages encouraging use of those resources. About half (n=42) of participants were randomized by an EHR algorithm to also receive customized lifestyle support (CLS), including weekly email messages adaptive to weight loss progress and telephonic coaching by a nurse for those facing challenges. Interventions and assessments spanned from January to July 2020, with disruption by the COVID-19 pandemic. Weight data were collected from administrative sources. Qualitative analysis of stakeholder recommendations and patient interviews assessed intervention acceptability, appropriateness, and sustainability.

Results: Over 6 weeks, 426 patients were sent the EHR invitation message, and 80 (18.8%) patients affirmed interest in the weight loss goal and were enrolled. Overall, 48 of 80 (60%) trial participants lost weight at 6 months; 12 (15%) exhibited weight loss ≥5%, with no significant difference between CLS and BLS arms (*P*=.85). During the 12 weeks of adaptive MyChart (Epic Systems) messaging, 18 (43%) CLS patients and 8 (21%) BLS patients performed daily self-weighing (*P*=.06), and 22 (52%) CLS patients and 14 (37%) BLS patients enrolled in referral-based lifestyle resources (*P*=.18).

Conclusions: Pragmatic enrollment, randomization, and data collection procedures proved feasible, and interventions showed preliminary effectiveness warranting further study in a larger trial.

Trial Registration: ClinicalTrials.gov NCT03998046; https://www.clinicaltrials.gov/study/NCT03998046

(JMIR Mhealth Uhealth 2025;13:e58722) doi:10.2196/58722

XSL•FO
RenderX

KEYWORDS

mHealth; mobile health; mobile application; app; application; obesity; overweight; obese; weight loss; cardiovascular; cardiovascular risk factor; risk factor; cardiovascular disease; CVD; cardiovascular illness; diabetes; primary prevention; behavioral intervention; lifestyle intervention; implementation; health care delivery; primary care; feasibility

Introduction

Moderate-intensity physical activity and just 10 to 15 pounds of weight loss can improve blood glucose, blood pressure, and cholesterol and reduce the need for medications to control those cardiovascular risk factors [1-6]. Achieving and maintaining weight loss is difficult, particularly if the environment surrounding individuals is not supportive of healthy eating and physical activity [7]. Adults are more likely to achieve weight loss and activity goals when supported by behavioral interventions involving recurring contact with a behavioral coach over months to years [1,2]. Such "intensive lifestyle interventions" (ILIs) improve health-related quality of life, reduce work absence, and lower future health care usage and costs [8,9].

US health care systems are a natural partner for ILI implementation because 5 in 6 adults complete office visits each year [10], during which cardiovascular risk factors such as body weight, blood pressure, glucose, and cholesterol are routinely assessed. Health care systems also have incentives to intervene before cardiovascular risk factors cause symptoms, functional impairments, or costly health complications [9]. However, health care systems face challenges in implementing ILI [11,12], which require dedicated personnel, space, supplies, and technologies that generally do not align well with the existing structure, workforce, and routines of health care delivery.

Past research has investigated strategies to simplify the role of health care systems in ILI delivery. Examples include reducing the frequency or duration of behavioral coaching contacts or shifting the source of coaching to outside health care settings, such as via wellness professionals employed by community organizations or distance coaching delivered using a computer or smartphone [13-17]. These strategies have lower effectiveness than more intensive interventions studied in randomized controlled trials [18], but also cost less and reach more people [8,19,20]. Despite increasing numbers of ILI programs nationally, clinicians refer very few patients to ILIs, and most people who may benefit remain unaware [19]. There is an immense need for research that identifies feasible and acceptable strategies for health care systems to implement ILIs.

One approach used to adapt and improve the implementation of multicomponent behavioral interventions without eroding their fidelity is to (1) distill them to their core components and (2) devise simpler implementation strategies for each component that improve their feasibility and acceptability within health care system delivery settings [21,22]. When there are multiple feasible implementation strategies for each core component, this approach also offers opportunities for health care systems to customize intervention delivery based on individual patient preferences (eg, letting them choose between formats) or their responsiveness (eg, offering "option A" but stepping up to offer more intensive "option B" if patients do not engage or improve). This approach holds promise for expanding the reach of high-fidelity intervention delivery to many more patients, by offering less intensive ways to get started for those who are initially unable or unwilling to enroll when referred to a full-scale program offered in a single format or location, as well as by better aligning opportunities for engagement in each component with different forms of motivation and ability. Extensive previous research over the past 2 decades has helped to define the core components of ILIs to include support for: an initial weight loss goal of at least 5 to 10 kg (11 to 22 pounds) at a pace of about 0.5 to 2 pounds per week [1]; resources and training for individuals to self-monitor dietary, activity, and weight changes [23]; and longitudinal coaching (typically with contact every 1 to 4 weeks). Coaching serves to create accountability [24], develop self-regulation and problem-solving skills [25,26], suggest refinements to behavioral strategies, and build self-efficacy and mastery over behaviors [23]; coaching can be effective when offered face-to-face or via a technology platform, with individual participants or groups [1].

Over 3 years, we engaged with health care system and fitness industry leaders, managers, and service professionals to cocreate and evaluate the feasibility of an array of practical implementation strategies for health care systems to support ILI core components. To our knowledge, this is the first such effort in which all intervention components are embedded within or coordinated directly by an electronic health record (EHR) management system. This paper aims to describe the collaborative approach used for the creation of the ILI implementation strategies and to report preliminary data for the feasibility, acceptability, and potential effectiveness of this approach.

Methods

Research Approaches

Our research involved both stakeholder-engaged "cocreation" of implementation strategies using collaborative design principles and a small, parallel, 2-arm, pilot trial with 1:1 individual treatment assignment to demonstrate feasibility and acceptability of the strategies in a single general internal medicine practice setting in downtown Chicago, Illinois, United States. The trial involved pragmatic recruitment, enrollment, and data collection strategies [27]. Eligibility determination and intervention allocation were automated within the EHR; trial enrollment proceeded with an Institutional Review Board (IRB)-approved waiver of written informed consent, and most outcomes were ascertained from administrative data systems. Each of these approaches minimized contact between study patients and the research team and enabled study patients and the interventions they received to approximate "real-world" conditions. Eligible patients were enrolled from December 2019 to January 2020; data were collected through January 2021. The experimental services were offered to patients beginning in

XSL•FO
RenderX

January 2020, and access to some services (described below in the "Patient Engagement in Intervention Components" subheading) was interrupted by the COVID-19 pandemic stay-at-home order in March 2020. All study procedures, including pragmatic enrollment and randomization under a waiver of written informed consent for research, were approved by the Northwestern University IRB, and the pilot trial was registered and posted on June 25, 2019, at ClinicalTrials.gov (NCT03998046). This paper was structured to comply with the CONSORT-EHEALTH (Consolidated Standards Of Reporting Trials for Electronic Health) Interventions reporting guidelines [28].

Stakeholder Engagement in the Research

From September 2016 to December 2019, we engaged stakeholders to consider research evidence for ILI core components [1,2,23] and cocreate implementation strategies for sustainable delivery of each core component in busy primary care practice settings. Recurring meetings of a "design team" included behavioral and social scientists, health care system nursing supervisors, a practicing primary care physician, 2 wellness program leaders from the partnering fitness club network, and health care system employees who develop and manage new tools and applications within the EHR system. Individual meetings were held with additional stakeholders, including health care system leaders, medical group and practice leaders, and the CEO and wellness program leaders from a regional fitness club network. Research staff members organized meetings, kept detailed notes, verified recommendations and action steps with participants, and integrated ideas into an array of design themes and concepts that informed intervention prototypes, pretests, and refinements [11].

Eligibility Criteria and Recruitment of Feasibility Trial Participants

A clinical data manager used EHR reporting tools to generate lists of eligible patients. Inclusion criteria were: age 18 - 75 years; ≥1 log-in within the past 6 months to MyChart (ie, a secure EHR access portal for patients to view their chart and exchange information with care team members; Epic Systems); completion of ≥1 office visit in the past 6 months at the participating primary care practice site; measured weight and height at the most recent office visit indicating BMI ≥27.0 kg/m^2; and ≥1 cardiometabolic risk factor (high blood pressure, abnormal blood cholesterol, prediabetes, or type 2 diabetes). Patients were ineligible if their last blood pressure was ≥180/105 mm Hg or their last hemoglobin A1c was ≥10.9%. Patients were also ineligible if they received cancer treatment in the past 2 years, were hospitalized in the past 3 months, were pregnant, or had a personal history of an eating disorder or serious mental illness.

The data manager used EHR messaging tools to send each physician a list of their own eligible patients, asking them to reply within 1 week if they wished to remove patients for whom they believed offering support for lifestyle change would be inappropriate. After 1 week, the data manager used the EHR's batch patient messaging tool to transmit a standardized outreach "goal setting" message to each patient remaining on the list. The message encouraged healthy eating and physical activity changes with a weight loss goal, and it offered free-of-charge access to technologies and coaching support to help recipients achieve the goal. The message also prompted patients to click an embedded link if they wished to adopt a weight loss goal and receive support. Because limited resources were available to support intervention costs during feasibility testing, we sent MyChart messages to only 100 - 200 patients every 2 weeks until the target sample size of 80 patients had affirmed interest in receiving the intervention services. This sample size was believed sufficient to demonstrate the feasibility and fidelity of pragmatic randomization and data collection procedures, to test the deployment of all variations of the patient messages, and to collect robust qualitative data about intervention perceptions. All 80 patients were included in the pragmatic trial evaluation; all received a bundle of basic lifestyle support (BLS) services (below in the "Procedures for Designing Intervention Components and Implementation Strategies" subheading), and about half (n=42) of patients were randomly assigned by a "silent" EHR algorithm to also receive additional, more customized lifestyle support (CLS) services (below in the "Procedures for Designing Intervention Components and Implementation Strategies" subheading). Both interventions are inherently unblinded.

After the trial enrollment target was reached, study patients were randomly selected for recruitment to complete semistructured telephone interviews to assess perceptions toward the intervention components. A research assistant emailed and then phoned 30 study patients before a targeted sample size of 15 interviews had been completed.

Procedures for Designing Intervention Components and Implementation Strategies

Design goals were to cocreate ILI intervention components and implementation strategies that are evidence-based [1,2] and perceived by stakeholders as technically feasible, acceptable, and financially sustainable [29]. This approach followed collaborative design principles [11] and produced ILI implementation strategies for support of weight loss goal setting, self-weighing, and recurring coaching support. Stakeholders advised using the EHR to identify eligible patients, to message them to encourage a weight loss goal, inspire accountability, and increase engagement in coaching support services, as well as to monitor progress metrics as a means to offer adaptive, higher intensity services to patients facing challenges with automated support alone (Table S1 in Multimedia Appendix 1). Stakeholders cautioned that encouragement of a weight loss goal in a MyChart message could be perceived as inappropriate by some patients; they advised framing the goal and offering resources instead as a means to improve cardiovascular risk factor control. They also advised recommending a relatively simple initial goal of "about 10 pounds over 10 weeks," believing this would be acceptable, appropriate, and achievable for most patients. This initial goal aligns with the large evidence base for effective ILIs, which typically support an initial weight loss goal of at least 5 kg (11 pounds) at a pace of about 0.5 to 2 pounds per week [1]. Stakeholders advised exploring the feasibility of receiving data from a wireless scale and analyzing that data directly within the EHR as a means to inform more personalized feedback about self-weighing and to identify and

XSL•FO
RenderX

target patients with more adaptive forms of intervention support. They advised providing all patients with free-of-charge access to coaching services offered by a partnering fitness organization and assigning a practice nurse to provide "step-up" telephonic coaching to patients who do not make progress toward their weight goal. They emphasized that we should evaluate the quality of all new resources, whether patients perceive them as appropriate, and whether their costs are sustainable.

The resulting intervention components, which are described in detail in Multimedia Appendix 1, were all deployed or coordinated by the EHR. Interventions commenced when a patient received the introductory MyChart goal-setting message (above in the "Eligibility Criteria and Recruitment of Feasibility Trial Participants" subheading) and clicked the link to affirm interest in a weight loss goal. All 80 patients received, free of charge, a bundle of BLS services. Specifically, they were shipped an electronic weight scale (e-Scale), which had been programmed to transmit weight data back to their electronic chart using cellular networks, at no cost to the patient. Each patient received a follow-up MyChart message with simple instructions for using the scale to weigh themselves daily, as well as a "coupon code" to enroll free of charge in longitudinal lifestyle coaching and support resources offered by a regional fitness organization partner. Patients could choose to access the referral services in 2 formats: face-to-face access to a dietitian and fitness coach at 11 facility locations or fully remote access to the same professionals via video chat with a smartphone, tablet, or personal computer. The face-to-face option provided full health club facility access, and the remote delivery option provided a suite of healthy food preparation and physical activity tips, tracking tools, and videos. Each patient also received instructions for messaging a dedicated nurse-coach at the primary care practice if they had problems using the support services or if they wanted more personalized support for reaching dietary, physical activity, or weight loss goals.

Anticipating that many patients may benefit from more frequent nudging and greater personalization of the behavioral support, the design team recommended 2 additional intervention components with these goals. Because more intensive services introduce additional implementation challenges and cost, we evaluated these components by offering them only to about half (n=42) of patients who had been randomized to receive, free of charge, CLS services. Specifically, for the first 12 weeks, each patient in the CLS arm was sent an automated weekly MyChart message with coaching tips customized to their use of the e-Scale in the previous week, whether they had made weight loss progress (defined as ≥0.5 pounds of weight loss from the first to last weight measure during the previous week), and whether they had enrolled in referral-based services offered by the fitness partner. A practice nurse used an EHR dashboard to monitor the entire CLS group and attempted to call patients to offer customized "step-up" telephone coaching if either (1) they had not used their scale in the previous week or (2) for 2 consecutive weeks they had not achieved ≥0.5 pounds of weight loss.

Study Outcomes

The primary quantitative outcome was the percentage of patients achieving ≥5% weight loss from baseline at 6±3 months. Additional outcomes included the feasibility of automated trial processes (eligibility determination, enrollment, and randomization); the feasibility of pragmatic weight data capture from office visits and e-Scales; usage of intervention components; health care system costs to implement each intervention component; and patient perceptions regarding appropriateness, ease of access, ease of use, acceptability, desirability, credibility, and usefulness.

Data Sources

EHR data spanning ±15 months of the launch of the trial provided information about patient baseline characteristics, service usage patterns, and metabolic and biometric outcomes, including body weight and height measured at office visits. BodyTrace e-Scales given to all study patients provided a secondary source of body weight data when weight data were missing in the EHR. The scale manufacturer reports accuracy within ±0.1 kg to a weight limit of about 150 kg (330 lbs) [30].

Multiple data sources were used to describe the usage of the different intervention components. Input from health care system leaders and study team members identified the specific personnel, supplies, contract services, and facility and administrative inputs needed to implement each intervention component. e-Scales provided data on the frequency of weight self-monitoring. Administrative data systems captured the number and type of contacts made between nurses and study participants. The partnering fitness network also provided data about patient participation in the 2 coaching support options and facility use.

Multiple data sources were also used to inform health care system costs to implement each intervention component, including personnel, supplies, technology, and other facility and administrative expenses. Staff members employed by both the health care system and the partnering fitness network were interviewed to estimate the amount of time they spent supporting different intervention activities over 6 and 12 months. Hourly salary costs of different personnel were collected from reports by the US Bureau of Labor Statistics. Costs of e-Scales, community intervention program fees, and development costs for health IT components were determined directly from invoice amounts. Maintenance costs for IT support were estimated from interviews with IT staff members, who were asked to estimate hours per week spent supporting the intervention components.

Patient-reported data were collected by semistructured telephone interviews with 15 randomly selected patients who were recruited beginning approximately 12 weeks after their enrollment in the feasibility trial (refer to the "Eligibility Criteria and Recruitment of Feasibility Trial Participants" subheading). Interviews assessed perceptions regarding the appropriateness, ease of access, ease of use, acceptability, desirability, credibility, and usefulness of the intervention components in helping the patient achieve health-related goals [11,31].

XSL•FO
RenderX

Variable Construction for the Weight Loss Outcome

Analysis of body weight changes used data from both the EHR and e-Scales. The last EHR weight measure before each patient's randomization date served as the baseline body weight, and the EHR weight measure closest to the date of randomization plus 6±3 months served as the follow-up weight. For patients who were missing such a follow-up EHR weight value, we estimated the 6-month value as one-half the difference between the baseline value and the EHR weight value closest to 12±3 months. For participants missing both a 6- and 12-month weight value in the EHR, we used the e-Scale weight value closest to 6±3 months. The percentage of weight loss at 6±3 months was calculated as: [baseline body mass (kg) – 6-month body mass (kg)] ÷ baseline body mass (kg). The dichotomous outcome of achieving ≥5% weight loss at 6±3 months was then determined for each patient. For patients missing all follow-up weight data, we assumed conservatively that they failed to reach 5% weight loss.

Analysis of the Feasibility of Automated Enrollment and Randomization Procedures

Univariate and bivariate descriptive statistics were used to examine the time needed to accrue the targeted sample size and to assess representativeness by comparing the demographic and clinical characteristics of patients enrolled with those of the entire eligible patient population. We also compared the sample allocation and balance of demographic and clinical characteristics of patients in each treatment arm following the automated allocation procedure.

Analysis of the Feasibility of Pragmatic Weight Data Capture

We evaluated the feasibility of weight outcome data capture from the pragmatic sources by examining proportions of patients with sufficient EHR or e-Scale data capture at the different time points of interest (baseline, 6±3 months, and 12±3 months).

Analysis of Preliminary Data for Weight Loss Effects

We used chi-square tests to compare all patients randomized to CLS and BLS arms on the proportion that reached the dichotomous outcomes of any weight loss at 6 months and the intervention goal threshold of ≥5% weight loss at 6 months.

Analysis of Patient Engagement in Intervention Components

We used descriptive statistics to summarize proportions of study patients who engaged in different intervention components (ie, e-Scale self-weighing, referral-based coaching resources, step-up telephone coaching by health care system nurses), as well as their volumes of use and timing of use relative to both study enrollment and the sudden start of the COVID-19 pandemic

Analysis of Intervention Delivery Costs

We summarized the mean per-person intervention cost for each study arm, both overall and across categories of personnel, materials and supplies, and facility and administrative costs.

Analysis of Patient Perceptions Toward Intervention Components

We conducted qualitative analysis on semistructured interview transcripts capturing patient perceptions regarding the appropriateness, ease of access, ease of use, acceptability, desirability, credibility, and usefulness of intervention practice components (Table S2 in Multimedia Appendix 1). Example themes and representative quotes were selected by group deliberation and consensus and are summarized in narrative format.

Ethical Considerations

All study procedures were performed in accordance with the Declaration of Helsinki. Study procedures, including pragmatic enrollment and randomization under a waiver of written informed consent for research, were approved by the Northwestern University IRB (STU00207153). The pilot trial was registered at ClinicalTrials.gov (NCT03998046) and posted on June 25, 2019.

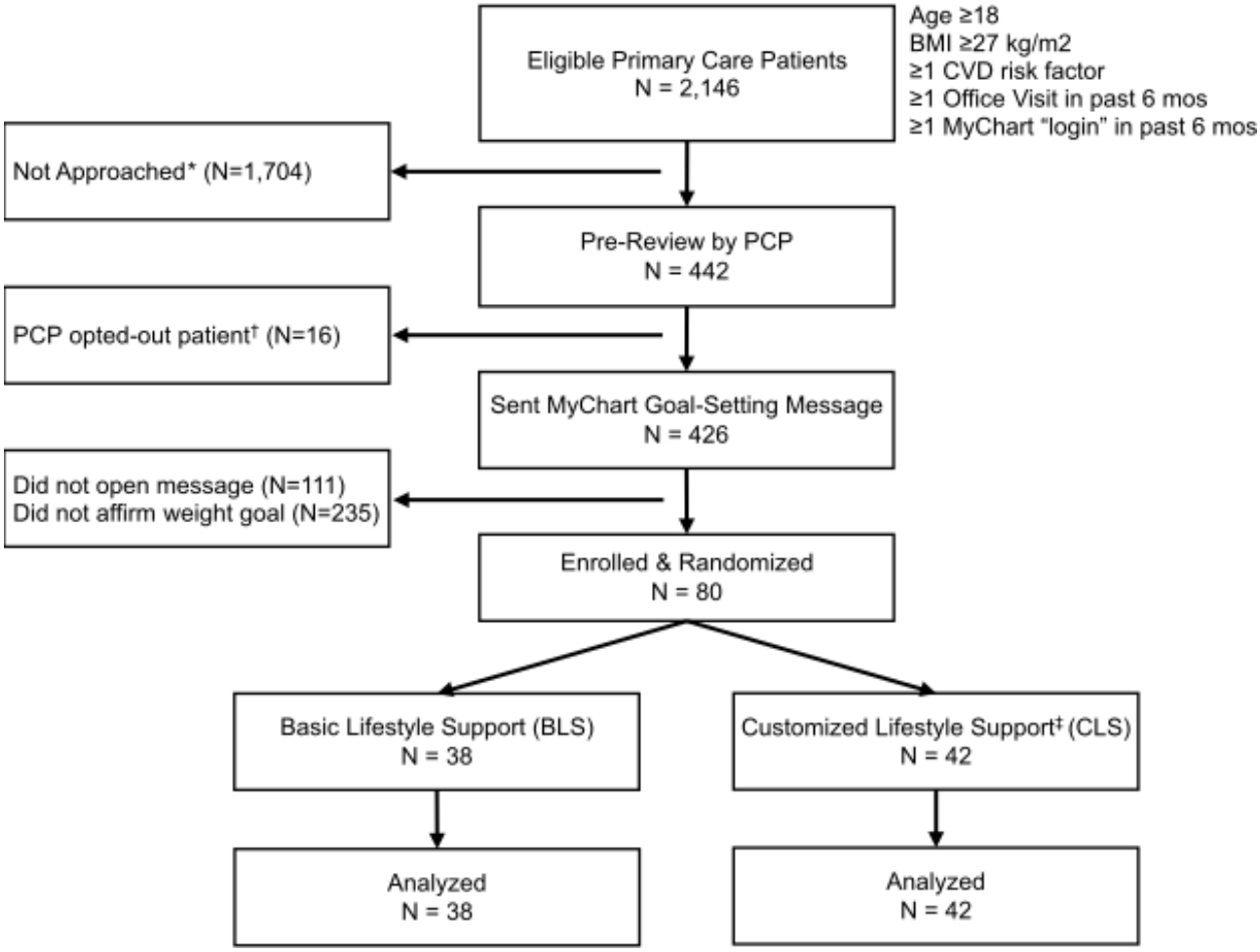
Results

Feasibility of Automated Trial Enrollment and Randomization Procedures

The eligibility list for the feasibility trial included 2146 patients with a mean age of 58.3 years and mean BMI of 33.5 kg/m^2; 1295 (60%) were women, 783 (36%) were aged ≥65 years, 617 (29%) were non-Hispanic Black, 1089 (51%) were non-Hispanic White, and 184 (9%) were Hispanic. Over 6 weeks, the clinical data manager sent names of 442 randomly selected eligible patients to 34 primary care clinicians for prereview; clinicians asked that only 16 (4%) patients not be contacted. A single outreach MyChart message was sent to each of the remaining 426 patients, of whom 315 (74%) opened the message within 4 weeks, and 80 (19%) affirmed interest in the weight loss goal (Figure 1).

XSL•FO
RenderX

Figure 1. Feasibility trial flow. CVD: cardiovascular disease; PCP: primary care physician. * Batches of 100-200 patients were randomly selected from the eligibility list for pre-review by their physician before being sent a single MyChart message offering weight loss support; this procedure was repeated every 2 weeks until 80 patients had accepted the services; patients remaining on the list were not approached; † Prior to sending a MyChart message to eligible patients, PCPs asked to opt-out 16 (3.6%) patients from being offered services; ‡ Study patients who were randomized to receive "Customized Lifestyle Support" services received additional weekly email messages adapted to individual weight loss progress and step-up telephonic coaching by a nurse when they faced challenges (see text for details).



Demographic characteristics of the 80 participants (n=50, 63% women; n=24, 30% aged ≥65 years; n=23, 29% non-Hispanic Black; n=34, 43% non-Hispanic White; n=11, 14% Hispanic or Latino) approximated the characteristics of the overall target population, with slight over-representation of those identifying as Hispanic or Latino, non-Hispanic Black, or non-Hispanic Asian (Table 1). In the setting of relatively small sample sizes, trial patients randomized to CLS were more often than BLS patients to be older than 45 years, male, and Hispanic or Latino. CLS patients were also more likely to have prediabetes and a higher Charlson comorbidity score [32].

XSL•FO
RenderX

Table . Baseline characteristics of study patients.

Baseline characteristic[a]	Eligible, approached (n=426)	Affirmed weight goal, enrolled (n=80)	Basic lifestyle support (BLS; n=38)	Customized lifestyle support (CLS; n=42)	Approached and completed interview[b] (n=15)
Age group (years), n (%)					
18 - 44	61 (14)	12 (15)	8 (21)	4 (10)	0 (0)
45 - 64	230 (54)	44 (55)	18 (47)	26 (62)	12 (80)
Age 65+	135 (32)	24 (30)	12 (32)	12 (29)	3 (20)
Sex, n (%)					
Female	234 (55)	50 (63)	26 (68)	24 (57)	10 (67)
Male	192 (45)	30 (38)	12 (32)	18 (43)	5 (33)
Race or ethnicity, n (%)					
Hispanic or Latino	44 (10)	11 (14)	3 (8)	8 (19)	4 (27)
Non-Hispanic Black	113 (27)	23 (29)	12 (32)	11 (26)	6 (40)
Non-Hispanic Asian	14 (3)	6 (8)	2 (5)	4 (10)	0 (0)
Non-Hispanic White	217 (51)	34 (43)	16 (42)	18 (43)	4 (27)
Other	38 (9)	6 (7)	5 (13)	1 (2)	1 (7)
Qualifying diagnoses, n (%)					
Prediabetes	128 (30)	30 (38)	19 (50)	11 (26)	6 (40)
Diabetes	138 (32)	19 (24)	6 (16)	13 (31)	5 (33)
High blood pressure	324 (76)	62 (78)	29 (76)	33 (79)	13 (87)
Dyslipidemia	341 (80)	61 (76)	27 (71)	34 (81)	11 (73)
Charlson score[c], mean (SD)	1.34 (1.79)	1.33 (1.97)	1.13 (1.70)	1.50 (2.19)	1.13 (1.46)
Baseline risk factor levels, mean (SD)[d]					
System International (SI) Units					
Mean weight, kg	95.0 (15.8)	94.3 (17.2)	95.1 (18.4)	93.6 (16.2)	95.6 (18.5)
Mean BMI, kg/m^2	33.2 (5.6)	33.6 (6.1)	34.1 (7.1)	33.1 (4.9)	34.7 (8.5)
Hemoglobin A1c, mmol/mol	44.5 (11.5)	43.8 (12.6)	42.4 (9.8)	45.2 (14.8)	42.5 (6.7)
Systolic blood pressure, mm Hg	131 (15)	135 (18)	137 (18)	133 (18)	135 (20)
Total cholesterol, mmol/L	4.69 (1.15)	4.79 (1.17)	4.74 (1.06)	4.84 (1.27)	5.20 (1.18)
HDL[e] cholesterol, mmol/L	1.33 (0.36)	1.42 (0.35)	1.45 (0.40)	1.39 (0.31)	1.38 (0.30)
Non-HDL cholesterol, mmol/L	3.36 (1.07)	3.37 (1.06)	3.29 (0.91)	3.45 (1.18)	3.82 (1.19)
Conventional units					

XSL•FO
RenderX

Baseline characteristic[a]	Eligible, approached (n=426)	Affirmed weight goal, enrolled (n=80)	Basic lifestyle support (BLS; n=38)	Customized lifestyle support (CLS; n=42)	Approached and completed interview[b] (n=15)
Mean weight, lbs	209.0 (34.8)	207.5 (37.8)	209.2 (40.5)	205.9 (35.6)	210.4 (40.8)
Hemoglobin A1c, %	6.2 (1.1)	6.2 (1.2)	6.0 (0.9)	6.3 (1.4)	6.0 (0.6)
Total cholesterol, mg/dL	181.3 (44.3)	185.4 (45.3)	183.2 (41.0)	187.3 (49.2)	201.1 (45.4)
HDL cholesterol, mg/dL	51.3 (14.1)	54.9 (13.7)	56.1 (15.5)	53.8 (12.0)	53.4 (11.7)
Non-HDL cholesterol, mg/dL	130.0 (41.4)	130.4 (40.9)	127.1 (35.2)	133.5 (45.6)	147.8 (45.9)

[a]Sex and race or ethnicity are based on self-reported values stored in each patient's electronic chart. Weight, blood pressure, hemoglobin A1c, and cholesterol values are based on the last recorded value within the past 12 months in each patient's electronic chart. BMI is based on the last recorded weight and height in each patient's electronic chart. Some column percentages may not sum to 100% due to rounding.

[b]Thirty of the 80 trial participants were randomly selected and approached for interviews to recruit 15 respondents.

[c]Charlson comorbidity score is a weighted score based on the number of comorbid diagnoses.

[d]Values are based on electronic health records. One hundred percent of patients had a weight and systolic blood pressure value in the past 12 months. A1c values were available for 338, 67, 33, 34, and 13 patients, and cholesterol values were available for 418, 78, 37, 41, and 14 patients in the eligible, enrolled, BLS, CLS, and interviewed groups, respectively.

[e]HDL: high-density lipoprotein.

Feasibility of Pragmatic Weight Data Capture

Despite lower-than-usual office visit rates during the early phase of the COVID-19 pandemic, 44 (55%) of feasibility trial patients had an office weight measure recorded in the EHR at 6±3 months from their trial enrollment. Among the remaining 36 patients, 26 (72%) had a weight recorded at 12±3 months, and of the remaining 10 patients, 7 (70%) had an e-Scale weight received at 6±3 months. Thus, our pragmatic data capture strategy enabled estimation of a 6-month weight change for 77 (96%) of all trial participants. The average time between trial enrollment and the weight used for the 6-month data analysis was 184.5 days.

Analysis of Preliminary Weight Loss Effects

Overall, 48 (60%) of 80 participants had lower weight at follow-up than at baseline; 12 (15%) exhibited a weight loss of ≥5% of baseline weight. The sample size for this feasibility study was not designed to test differences in weight loss goal achievement between randomized treatment arms, and there was no statistically significant difference between CLS and BLS (P=.85).

Using e-Scale data and assuming that 11 of the 80 patients who did not self-weigh also did not lose weight, 28 (35%) of all participants achieved ≥10 lbs of weight loss (ie, the initial goal recommended by the MyChart message), and 24 (30%) achieved ≥5% of weight loss at some point during the 12-week automated messaging phase. Thus, half of the patients who achieved a 5% weight loss during the 12 weeks of active support maintained the loss at 6 months.

Patient Engagement in Intervention Components

All except 3 (4%) patients were able to initialize their e-Scale and register a baseline weight. Over the 12-week MyChart messaging phase of the intervention, 11 (14%) study patients did not use their scales for self-weighing, 7 (9%) participated in weighing in some but not all weeks, and 62 (78%) engaged in self-weighing during all 12 weeks. Self-weighing daily (5 to 7 days per week) was performed by 26 (33%) patients during all 12 weeks; more patients assigned to CLS (customized weekly messaging) performed daily self-weighing in all weeks (18/42, 43% patients) compared with those in BLS (8/38, 21%; P=.06).

Overall, 36 (45%) patients enrolled in one of the two referral-based coaching platforms offered by the fitness industry partner, 14 (37%) in the BLS arm and 22 (52%) in the CLS arm (P=.18). Among all who enrolled, 22 (28%) elected the face-to-face option and 14 (18%) elected the fully remote delivery option. Patients enrolled in these services between 15 and 77 days after receiving the initial outreach MyChart message. Access to fitness facilities (ie, the face-to-face delivery option) was disrupted by a COVID-19 pandemic stay-at-home order beginning the third week of March 2020. At that time, patients had been enrolled in the referral-based services for 19 to 75 days; the 22 patients in face-to-face support had completed an average of 3 (range 1-6) individual coaching sessions and attended a partnering fitness facility 0 to 24 times. The 14 patients receiving fully remote support had completed an average of 4 (range 2-5) individual coaching sessions before the pandemic and an average of 3 additional (range 0-5) sessions afterward. Among patients randomized to CLS, 41 (98%) received at least one step-up telephonic nurse coaching call; 19 (45%) received ≥5 nurse calls (maximum 11 calls received) over the 12-week intensive support phase.

Intervention Delivery Costs

Direct medical costs associated with offering each intervention component are summarized in Table S3 in Multimedia Appendix

XSL•FO
RenderX

1. Mean health care system costs to offer the intervention as delivered during the feasibility trial were estimated at US $335 per person over the first 6 months (US $284 for BLS patients; US $382 for CLS patients) and US $150 per person over the second 6 months (US $124 for BLS; US $174 for CLS patients).

Patient Perceptions Toward Intervention Components

Among the 80 feasibility trial participants, 30 were randomly selected and approached, of whom 15 agreed to complete a telephone interview; 11 were unreachable, 5 declined participation, and 2 asked to be recontacted at a future time. Respondents included 5 men and 10 women, with demographic and clinical characteristics generally representative of the overall target population (Table 1). Patient perceptions are summarized below; representative quotes and considerations for design refinements are organized in Table S2 in Multimedia Appendix 1.

Most patients described the MyChart messages and website resources as generally easy to understand and navigate. However, some reported feeling frustrated or less engaged by not having a clear understanding upfront for how the different intervention components were coordinated or should be used. Trust in the person or organization providing lifestyle support services emerged as an important theme. Some patients expressed being distrustful, in general, when offered resources for "free" from another organization, recommending that it be made very clear that their doctor or the health care system was recommending and coordinating the activities. Others indicated that they had not initially realized they could access dietary and physical activity coaching at the partnering fitness organization and were unsure how to engage with the health care system nurse if they had problems.

Nearly all patients reported that an initial weight loss goal of about 10 pounds in 10 weeks was appropriate and achievable. Some expressed interest in larger weight loss goals and access to services beyond 10 to 12 weeks. Many patients reported that their weight loss goal was motivated by other health or social goals, such as reducing the need for more pills to treat high blood pressure.

Several patients highlighted the importance of supportive accountability in successfully changing behavior. They emphasized the practice nurse, registered dietitian, and physical trainers as key sources of supportive accountability, more so than the automated messages. Patients expressed strengths and limitations of the automated messages, highlighting that the messages were a helpful reminder or source of encouragement when making progress, but were insufficient or even demoralizing when not making progress.

Many patients viewed self-weighing as an additional source of accountability. Some expressed interest in more elaborate e-Scales or integration with other technologies, such as smartphone apps that provide immediate and customized feedback. Other patients indicated that they preferred support from a member of the health care or fitness center team. Generally, patients expressed frustration when they viewed e-Scales (and self-weighing) as a direct source of support**,** rather than as a means for health care providers and coaches to track and customize their support. Importantly, some viewed daily weighing as a source of stress and felt that, no matter how closely they followed diet and physical activity plans, their daily weight remained unpredictable. Several patients highlighted low expectations for success stemming from negative experiences with previous weight loss attempts. Those patients recommended a greater focus on recognizing and addressing such expectations, as well as de-emphasizing a focus on the daily weight readings as a sign of success or failure, while offering more immediate support beyond automated messaging. Relatedly, some patients highlighted that emotions have a larger influence on dietary behaviors than reminders to eat healthy. Those patients recommended a greater focus on acknowledging and addressing emotions both as drivers and consequences of attempts to lose weight and change diet.

Finally, all patients indicated that the abrupt start of the COVID-19 pandemic presented a profound barrier to healthy behaviors and weight loss. Many indicated frustration with becoming more sedentary, adopting "stress eating" behaviors, and gaining weight. The pandemic restrictions eliminated access to fitness facilities, which further contributed to frustration.

Discussion

Principal Findings

This preliminary study was conducted to demonstrate the feasibility of several pragmatic practice components and implementation strategies for primary care offices to support the core components of ILIs. We also sought to demonstrate the feasibility of capturing and analyzing weight data from 2 secondary sources, the EHR and e-Scales, as an important biometric outcome for future research. After a single outbound MyChart message was sent to encourage a weight loss goal by 426 overweight or obese adults with cardiovascular risk factors, 315 (74%) opened the message, and 80 (25%) of those who opened it affirmed interest in receiving support services to lose weight. Among those 80 patients, 24 (30%) achieved ≥5% weight loss in the first 12 weeks and, despite the disruptive impacts of a pandemic, 12 (15%) still exhibited ≥5% weight loss at 6 months, at a time in history when weight gain was common [33-35].

Weekly "nudging" of patients with a customized MyChart message plus step-up telephone coaching by a practice nurse increased participation in daily self-weighing (43% vs 21% of patients through 12 weeks) and enrollment in referral-based lifestyle support resources offered by a partnering fitness organization (52% vs 37%). Although this small feasibility trial did not demonstrate whether patients are more likely to reach weight loss goals when provided with these more customized services, this approach warrants further study.

Relationship to Other Past Research

Our work is consistent with other recent studies of pragmatic, technology-facilitated lifestyle interventions in primary care settings [13-17], but we are unaware of other examples where the intervention components have been fully embedded within an EHR system. Our approach used design thinking with service professionals, health care system leaders, health payers, and

XSL•FO
RenderX

researchers to create practice components that not only are evidence-based but also desirable to implementers, technically feasible, and financially sustainable [11]. Design prototypes were iteratively pretested and refined with stakeholder collaboration, and a small feasibility trial demonstrated strong promise for both effectiveness and sustainability.

Our work also builds on past research evaluating referral linkages between health care systems and community-delivered ILIs [36-39]. However, most past research focused on referrals to a comprehensive program bundle, such as the National Diabetes Prevention Program, often offered in a single format or location [18,40]. Offering a multicomponent program such as the National Diabetes Prevention Program requires considerable community organizational capacity and financial investment, which imposes implementation barriers [41,42], particularly when enrollment is low or third-party payment is limited. A unique aspect of our approach was to distill ILI programs down to core components and to develop implementation strategies for each component to minimize demands placed on the implementing organization [22]. The resulting approaches were designed to help the partnering health care systems achieve existing goals relating to US Preventive Services Task Force recommendations for linking overweight or obese adults with cardiovascular risk factors with evidence-based behavioral interventions, while minimizing interference with existing, busy clinical workflows or placing more work on clinicians. We also considered how to offer the resulting services using an array of different channels (eg, MyChart, telephone, smartphone application, and fitness facility) to maximize reach by appealing to different segments of the target population. By leveraging existing clinical personnel with skills and incentives to support behavioral support services, linkages with extant service organizations that offer ILI resources, and EHR technologies that are now available in most health care settings, these pragmatic approaches, if effective, could have high potential for reproducibility in other settings.

Implications for Practice and Further Research

Conceptually, the development of discrete implementation strategies for individual ILI components has appeal if they successfully minimize workflow disruptions, distribute costs of the new activities across different partners, and map costs to existing pathways for health payer reimbursement. Possible disadvantages are that the individual components may not prove as effective in isolation, and additional strategies must be developed to coordinate staff roles, technologies, and patient transitions among the practice components. Some patients may still face access barriers if, for example, referral-based resources are limited or health payers do not provide reimbursement for remote weight monitoring or lifestyle counseling activities by nonphysicians [43,44]. Although our approach offers lower-technology intervention options, such as nurse-delivered telephone counseling and facility-based resources offered by a partnering fitness organization, these interventions may still have different effects based on technology preferences and literacy. It will be important for future research to evaluate these dimensions of implementation in the context of a larger and more definitive effectiveness trial.

Conclusion

This preliminary research indicates that strategies to implement ILI core components are feasible for primary care practices and generally acceptable to patients. Further research should seek to optimize implementation strategies to maximize feasibility, acceptability, sustainability, and reach, as well as to demonstrate effectiveness and cost-effectiveness compared with alternative service models. Engaging stakeholders in cocreation of these strategies holds promise to improve both implementation and reach of ILI components in new ways that achieve population health and health equity for millions of Americans who are currently overweight or obese and already engaged by the health care sector.

Acknowledgments

We would like to acknowledge Northwestern Medical Group Primary Care, Northwestern Memorial Healthcare, and the Fitness Formula Clubs for collaborating to develop and implement the practice interventions evaluated by this research. We are particularly grateful to clinicians and staff of both partnering organizations for contributing the ideas and energy that made this research possible. This study was funded by the US National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, under award R34DK114773.

Data Availability

The datasets generated or analyzed during this study are not publicly available because they are held and protected by the partnering health care system (Northwestern Memorial Healthcare). They are available from the corresponding author on reasonable request and with permission from Northwestern Memorial Healthcare.

Authors' Contributions

RTA was responsible for conceptualization, methodology, funding acquisition, investigation, formal analysis, visualization, supervision, and writing (original draft and editing). KAC, DTL, and BH contributed to conceptualization, methodology, funding acquisition, investigation, formal analysis, and writing (review and editing). ND, ARC, and KD contributed to conceptualization, resources, and writing (review and editing). AJC was responsible for methodology, resources, software, data curation, formal

XSL•FO
RenderX

analysis, validation, visualization, and writing (review and editing). CA and SAH contributed to the investigation, project administration, verification, and writing (review and editing).

Conflicts of Interest

RTA received salary support from grant funding for this research and provides part-time clinical services as an affiliated physician at Northwestern Medicine. KAC, DTL, CA, SAH, AJC, and BH received salary support from grant funding for this research. ND is employed by Northwestern Medicine as a primary care physician at the participating primary care practice. KD is employed by Northwestern Medicine and serves as vice president of operations. ARC is employed by Fitness Formula Clubs as a registered dietitian coordinator. The authors have no other conflicts of interest to declare.

Multimedia Appendix 1
Intervention design details.
[DOCX File, 89 KB - mhealth_v13i1e58722_app1.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1349 KB - mhealth_v13i1e58722_app2.pdf]

References

1. LeBlanc ES, Patnode CD, Webber EM, Redmond N, Rushkin M, O'Connor EA. Behavioral and pharmacotherapy weight loss interventions to prevent obesity-related morbidity and mortality in adults: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2018 Sep 18;320(11):1172-1191. [doi: 10.1001/jama.2018.7777] [Medline: 30326501]
2. O'Connor EA, Evans CV, Rushkin MC, Redmond N, Lin JS. Behavioral counseling to promote a healthy diet and physical activity for cardiovascular disease prevention in adults with cardiovascular risk factors: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2020 Nov 24;324(20):2076-2094. [doi: 10.1001/jama.2020.17108] [Medline: 33231669]
3. Ratner R, Goldberg R, Haffner S, et al. Impact of intensive lifestyle and metformin therapy on cardiovascular disease risk factors in the Diabetes Prevention Program. Diabetes Care 2005 Apr;28(4):888-894. [doi: 10.2337/diacare.28.4.888] [Medline: 15793191]
4. The Look AHEAD Research Group, Pi-Sunyer X, Blackburn G, et al. Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes. Diabetes Care 2007 Jun 1;30(6):1374-1383. [doi: 10.2337/dc07-0048] [Medline: 17259515]
5. Hamman RF, Wing RR, Edelstein SL, et al. Effect of weight loss with lifestyle intervention on risk of diabetes. Diabetes Care 2006 Sep;29(9):2102-2107. [doi: 10.2337/dc06-0560] [Medline: 16936160]
6. Mudaliar U, Zabetian A, Goodman M, et al. Cardiometabolic risk factor changes observed in diabetes prevention programs in US settings: a systematic review and meta-analysis. PLoS Med 2016 Jul;13(7):e1002095. [doi: 10.1371/journal.pmed.1002095] [Medline: 27459705]
7. Warren M, Beck S, Lieberman D, Ilakkuvan V, Lipson SK. The state of obesity: better policies for a healthier America. : Trust For America's Health (TFAH); 2021 Sep URL: https://www.tfah.org/wp-content/uploads/2021/09/2021ObesityReport_Fnl.pdf [accessed 2025-09-27]
8. Herman WH, Brandle M, Zhang P. Costs associated with the primary prevention of type 2 diabetes mellitus in the Diabetes Prevention Program. Diabetes Care 2003 Jan 1;26(1):36-47. [doi: 10.2337/diacare.26.1.36]
9. Li R, Qu S, Zhang P, et al. Economic evaluation of combined diet and physical activity promotion programs to prevent type 2 diabetes among persons at increased risk: a systematic review for the Community Preventive Services Task Force. Ann Intern Med 2015 Sep 15;163(6):452-460. [doi: 10.7326/M15-0469] [Medline: 26167962]
10. Interactive summary health statistics for adults. Centers for Disease Control and Prevention. URL: https://wwwn.cdc.gov/NHISDataQueryTool/SHS_adult/index.html [accessed 2025-09-27]
11. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016 Jun;23(2):180-200. [doi: 10.1111/cpsp.12154] [Medline: 29456295]
12. Yarnall KSH, Pollak KI, Østbye T, Krause KM, Michener JL. Primary care: Is there enough time for prevention? Am J Public Health 2003 Apr;93(4):635-641. [doi: 10.2105/ajph.93.4.635] [Medline: 12660210]
13. Aveyard P, Lewis A, Tearne S, et al. Screening and brief intervention for obesity in primary care: a parallel, two-arm, randomised trial. The Lancet 2016 Nov;388(10059):2492-2500. [doi: 10.1016/S0140-6736(16)31893-1]
14. Bennett GG, Steinberg D, Askew S, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018 Dec;55(6):777-786. [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]

XSL•FO
RenderX

15. Foley P, Steinberg D, Levine E, et al. Track: a randomized controlled trial of a digital health obesity treatment intervention for medically vulnerable primary care patients. Contemp Clin Trials 2016 May;48:12-20. [doi: 10.1016/j.cct.2016.03.006] [Medline: 26995281]
16. Little P, Stuart B, Hobbs FR, et al. An internet-based intervention with brief nurse support to manage obesity in primary care (POWeR+): a pragmatic, parallel-group, randomised controlled trial. The Lancet Diabetes & Endocrinology 2016 Oct;4(10):821-828. [doi: 10.1016/S2213-8587(16)30099-7]
17. McVay M, Steinberg D, Askew S, Bennett GG. Provider counseling and weight loss outcomes in a primary care-based digital obesity treatment. J Gen Intern Med 2019 Jun;34(6):992-998. [doi: 10.1007/s11606-019-04944-5] [Medline: 30891688]
18. Ali MK, Echouffo-Tcheugui JB, Williamson DF. How effective were lifestyle interventions in real-world settings that were modeled on the diabetes prevention program? Health Aff (Millwood) 2012 Jan;31(1):67-75. [doi: 10.1377/hlthaff.2011.1009] [Medline: 22232096]
19. Ali MK, McKeever Bullard K, Imperatore G, et al. Reach and use of diabetes prevention services in the United States, 2016-2017. JAMA Netw Open 2019 May 3;2(5):e193160. [doi: 10.1001/jamanetworkopen.2019.3160] [Medline: 31074808]
20. Golovaty I, Wadhwa S, Fisher L, et al. Reach, engagement and effectiveness of in-person and online lifestyle change programs to prevent diabetes. BMC Public Health 2021 Dec;21(1):1314. [doi: 10.1186/s12889-021-11378-4]
21. Chorpita BF, Becker KD, Daleiden EL. Understanding the common elements of evidence-based practice: misconceptions and clinical examples. J Am Acad Child Adolesc Psychiatry 2007 May;46(5):647-652. [doi: 10.1097/chi.0b013e318033ff71] [Medline: 17450056]
22. Chorpita BF, Daleiden EL, Weisz JR. Identifying and selecting the common elements of evidence based interventions: a distillation and matching model. Ment Health Serv Res 2005 Mar;7(1):5-20. [doi: 10.1007/s11020-005-1962-6] [Medline: 15832690]
23. Venditti EM, Kramer MK. Necessary components for lifestyle modification interventions to reduce diabetes risk. Curr Diab Rep 2012 Apr;12(2):138-146. [doi: 10.1007/s11892-012-0256-9] [Medline: 22350807]
24. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30. [doi: 10.2196/jmir.1602] [Medline: 21393123]
25. Madigan CD, Daley AJ, Lewis AL, Aveyard P, Jolly K. Is self-weighing an effective tool for weight loss: a systematic literature review and meta-analysis. Int J Behav Nutr Phys Act 2015 Aug 21;12:104. [doi: 10.1186/s12966-015-0267-4] [Medline: 26293454]
26. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265 [FREE Full text] [doi: 10.1002/oby.20946] [Medline: 25521523]
27. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015 May 8;350:h2147. [doi: 10.1136/bmj.h2147] [Medline: 25956159]
28. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
29. What is human-centered design? Design Kit. 2015. URL: https://www.designkit.org/ [accessed 2022-05-09]
30. BodyTrace I. BodyTrace. URL: https://www.bodytrace.com/medical/ [accessed 2024-01-18]
31. Weiner BJ, Lewis CC, Stanick C, et al. Psychometric assessment of three newly developed implementation outcome measures. Implementation Sci 2017 Dec;12(1):108. [doi: 10.1186/s13012-017-0635-3]
32. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992 Jun;45(6):613-619. [doi: 10.1016/0895-4356(92)90133-8] [Medline: 1607900]
33. Freedman DS, Kompaniyets L, Daymont C, Zhao L, Blanck HM. Weight gain among US adults during the COVID-19 pandemic through May 2021. Obesity (Silver Spring) 2022 Oct;30(10):2064-2070. [doi: 10.1002/oby.23511] [Medline: 35822832]
34. Lin AL, Vittinghoff E, Olgin JE, Pletcher MJ, Marcus GM. Body weight changes during pandemic-related shelter-in-place in a longitudinal cohort study. JAMA Netw Open 2021 Mar 1;4(3):e212536. [doi: 10.1001/jamanetworkopen.2021.2536] [Medline: 33749764]
35. Seal A, Schaffner A, Phelan S, et al. COVID‐19 pandemic and stay‐at‐home mandates promote weight gain in US adults. Obesity (Silver Spring) 2022 Jan;30(1):240-248. [doi: 10.1002/oby.23293]
36. Ackermann RT. Description of an integrated framework for building linkages among primary care clinics and community organizations for the prevention of type 2 diabetes: emerging themes from the CC-Link study. Chronic Illn 2010 Jun;6(2):89-100. [doi: 10.1177/1742395310364857] [Medline: 20484325]
37. Community-clinical linkages for the prevention and control of chronic diseases: a practitioner's guide. : Centers for Disease Control and Prevention (CDC); 2016 URL: www.cdc.gov/dhdsp/pubs/docs/ccl-practitioners-guide.pdf [accessed 2025-09-28]
38. Etz RS, Cohen DJ, Woolf SH, et al. Bridging primary care practices and communities to promote healthy behaviors. Am J Prev Med 2008 Nov;35(5 Suppl):S390-S397. [doi: 10.1016/j.amepre.2008.08.008] [Medline: 18929986]

XSL•FO
RenderX

39. Porterfield DS, Hinnant LW, Kane H, Horne J, McAleer K, Roussel A. Linkages between clinical practices and community organizations for prevention: a literature review and environmental scan. Am J Public Health 2012 Jun;102 Suppl 3(Suppl 3):S375-S382. [doi: 10.2105/AJPH.2012.300692] [Medline: 22690974]
40. Vojta D, Koehler TB, Longjohn M, Lever JA, Caputo NF. A coordinated national model for diabetes prevention: linking health systems to an evidence-based community program. Am J Prev Med 2013 Apr;44(4 Suppl 4):S301-S306. [doi: 10.1016/j.amepre.2012.12.018] [Medline: 23498291]
41. Madrigal L, Manders OC, Kegler M, et al. Inner and outer setting factors that influence the implementation of the National Diabetes Prevention Program (National DPP) using the Consolidated Framework for Implementation Research (CFIR): a qualitative study. Implement Sci Commun 2022 Oct 1;3(1):104. [doi: 10.1186/s43058-022-00350-x]
42. Medicare Diabetes Prevention Program: where are the suppliers? Am J Manag Care 2020 Jun 1;26(6):e198-e201. [doi: 10.37765/ajmc.2020.43496]
43. Hayes CJ, Dawson L, McCoy H, et al. Utilization of remote patient monitoring within the United States health care system: a scoping review. Telemed J E Health 2023 Mar;29(3):384-394. [doi: 10.1089/tmj.2022.0111] [Medline: 35819861]
44. Freeman KJ, Grega ML, Friedman SM, et al. Lifestyle medicine reimbursement: a proposal for policy priorities informed by a cross-sectional survey of lifestyle medicine practitioners. Int J Environ Res Public Health 2021 Nov 5;18(21):11632. [doi: 10.3390/ijerph182111632] [Medline: 34770148]

Abbreviations

BLS: basic lifestyle support
CLS: customized lifestyle support
CONSORT-EHEALTH: CONsolidated Standards Of Reporting Trials for Electronic Health
e-Scale: electronic scale
EHR: electronic health record
ILI: intensive lifestyle intervention
IRB: Institutional Review Board

Edited by M Dorsch; submitted 22.03.24; peer-reviewed by A Woodward, CA Brunette; revised version received 23.06.25; accepted 02.07.25; published 22.10.25.

Please cite as:
Ackermann RT, Cameron KA, Liss DT, Dolan N, Aikman C, Carson AR, Harris SA, Doyle K, Cooper AJ, Hitsman B
Hybrid Health IT and Telehealth–Delivered Behavioral Weight Loss Services for Primary Care Patients With Cardiovascular Risk Factors: Intervention Component Design and Pragmatic Randomized Feasibility Trial
JMIR Mhealth Uhealth 2025;13:e58722
URL: https://mhealth.jmir.org/2025/1/e58722
doi:10.2196/58722

© Ronald T Ackermann, Kenzie A Cameron, David T Liss, Nancy Dolan, Cassandra Aikman, Amy R Carson, Sterling A Harris, Kathryn Doyle, Andrew J Cooper, Brian Hitsman. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating a Mobile Digital Therapeutic for Vasomotor and Behavioral Health Symptoms Among Women in Midlife: Randomized Controlled Trial

Jennifer Duffecy[1], PhD; Arfa Rehman[2], MBA; Scott Gorman[2], MBA, JD; Yong Lin Huang[2], PhD; Heide Klumpp[1], PhD

[1]Department of Psychiatry, University of Illinois-Chicago, 912 S Wood, Chicago, IL, United States

[2]Chorus Health, San Francisco, CA, United States

Corresponding Author:
Jennifer Duffecy, PhD
Department of Psychiatry, University of Illinois-Chicago, 912 S Wood, Chicago, IL, United States

Abstract

Background: Perimenopausal and menopausal symptoms affect many women's well-being and quality of life. Digital technologies, especially smartphones, allow self-management interventions for menopausal symptoms, but they are understudied. **Objective:** We evaluated whether a novel digital care app, Caria, effectively reduced vasomotor and behavioral health symptoms of menopause. **Methods:** We enrolled 149 women for a 6-week randomized controlled trial (app treatment: n=112; web-based educational control: n=37). Enrolled participants had problematic vasomotor symptoms and at least one elevated behavioral health symptom (depression, anxiety, or sleep issues). Web-based self-reported assessments (Hot Flush Rating Scale [HFRS], Patient Health Questionnaire Depression Scale-8 [PHQ-8], Generalized Anxiety Disorder-7, and Pittsburgh Sleep Quality Index [PSQI]) were conducted at baseline, 3 weeks, and 6 weeks. **Results:** For hot flash severity (HFRS; treatment baseline mean 16.4, SD 6.7 to 6-wk mean 13.6, SD 6.6; control: baseline mean 19.1, SD 7.3 to 6-wk mean 17.8, SD 7.2), a repeated-measures ANOVA revealed main effects for time ($F_{2,262}$=9.82; P<.001) and treatment arm ($F_{1,131}$=6.08; P=.01) and a significant time × treatment arm interaction ($F_{2,262}$=3.23; P=.04); the treatment arm showed lower hot flash severity than the control arm (t_{147}=2.72; P=.007). For depression scores (PHQ-8; treatment baseline mean 14.0, SD 3.8 to 6-wk mean 11.2, SD 5.3; control baseline mean 15.0, SD 3.7 to 6-wk mean 13.4, SD 4.1), a repeated-measures ANOVA showed a main effect of time in the treatment arm ($F_{2,96}$=15.2; P<.001) but not the control arm ($F_{2,40}$=2.0; P=.15). Follow-up 2-tailed paired t tests in the treatment arm showed depression decreased from baseline to week 3 (t_{49}=3.3; P=.002) and from weeks 3 to 6 (t_{48}=2.3; P=.02). For sleep quality scores (PSQI; treatment baseline mean 10.7, SD 3.1 to 6-wk mean 10.0, SD 3.5; control baseline mean 11.5, SD 3.7 to 6-wk mean 11.0, SD 3.7), the repeated-measures ANOVA showed a main effect of time in the treatment arm ($F_{2,186}$=7.8; P=.001) but not the control arm ($F_{2,62}$=1.3; P=.28). Follow-up 2-tailed paired t tests in the treatment arm showed a significant decrease in sleep issues from baseline to week 3 (t_{95}=3.9; P<.001) but no change from weeks 3 to 6 (t_{93}=0.2; P=.81). Participants with elevated anxiety symptoms showed decreased symptoms in both the treatment and control groups. App engagement was high (average logins over 6 weeks: 53.2). **Conclusions:** The findings highlight the potential of digital interventions for mitigating menopausal vasomotor and behavioral health symptoms. Significant improvements in the intervention group underscore the app's effectiveness in providing relief from some of the most challenging aspects of menopause. This study contributes to the evidence supporting digital health interventions in managing menopausal symptoms, presenting a promising avenue for accessible and scalable solutions for women in midlife.

Trial Registration: ClinicalTrials.gov NCT04882982; https://clinicaltrials.gov/study/NCT04882982

(JMIR Mhealth Uhealth 2025;13:e58204) doi:10.2196/58204

KEYWORDS

randomized controlled trial; digital therapeutic; women's health; behavioral health symptoms; behavioral; women; woman; physical symptoms; psychological symptoms; menopause; well-being; quality of life; digital technology; self-management; menopausal symptoms; efficacy; digital care; digital care program; app; vasomotor; depression; anxiety; sleep; mobile phone

XSL•FO
RenderX

Introduction

Menopause, defined as the permanent cessation of menstrual cycle, marks a major transition in a woman's life [1]. It is typically preceded by perimenopause, a stage characterized by fluctuating hormone levels and the onset of various physiological changes [2]. These can include vasomotor symptoms (VMS) such as hot flashes (sudden, temporary onset of warmth in the body, flushing, and sweating) and night sweats (episodes of excessive sweating that happen during sleep), as well as depression, anxiety, and sleep disturbances [3]. In the United States, approximately 6000 women transition into menopause each day, many of whom face these challenges for years before and after menopause [4]. In recent years, research has begun to unveil the profound effects of these symptoms on women's well-being, work performance, and careers, as well as the associated economic burden [5,6]. For example, sleep disturbances due to hot flashes, night sweats, and insomnia often result in daytime fatigue, impairing cognitive function, and productivity. Studies have found that menopausal symptoms can lead to women taking more sick leave, reducing working hours, changing roles, or even leaving the workforce altogether [7,8].

Given that menopause symptoms can persist for several years, the long-term impact of untreated symptoms on women's quality of life and work performance is significant, resulting in substantial costs due to increased health care utilization and productivity losses [5,9]. The annual cost of menopausal symptoms in the United States alone is estimated to be US $26.6 billion, including US $1.8 billion due to productivity losses [10]. As the life expectancy of women continues to rise and midlife and older women make up one of the fastest-growing employment groups, it becomes increasingly important to address the considerable impact of menopausal symptoms through effective solutions [9,10].

Despite the impact of menopausal symptoms, there has been a notable lack of research on effective and accessible interventions. Common treatments for vasomotor and behavioral symptoms among women in midlife include pharmaceuticals such as menopausal hormone therapy and selective serotonin reuptake inhibitors [11-13]. However, menopausal hormone therapy is contraindicated for some women (such as survivors of breast cancer) [13], or some women may seek nonpharmacological interventions [12]. Behavioral interventions, such as cognitive behavioral therapy (CBT), have been recommended as a potential treatment for menopausal symptoms [14], with The Menopause Society rating the support for CBT to address menopause symptoms as level 1 (good and consistent evidence) [15]. CBT interventions focus on changing negative thought patterns and behaviors, promoting adaptive coping strategies, and enhancing emotional regulation [16,17]. Such interventions have been shown to improve psychological symptoms associated with health conditions such as insomnia, depression, cancer, chronic pain, and premenstrual syndrome [18-21], but would benefit from additional study in the context of menopause.

Recent studies have explored the efficacy of behavioral health interventions, such as CBT and mindfulness, which involve nonjudgmental awareness of internal state and surroundings [22], in improving health outcomes among menopausal women with promising results [23,24]. In a single-blind randomized controlled trial (RCT) by Greene et al [25], group-based, in-person CBT showed a reduction in VMS interference and bothersomeness, as well as depressive symptoms and sleep difficulties. In a 2018 study of a self-help unguided CBT booklet, results showed a reduction in problem rating of hot flashes and night sweats, in addition to improvements in sleep and work impairment [24].

Clinical guidelines recommend evidence-based psychotherapies such as CBT as the primary treatment for depressive symptoms during the menopause transition [25,26]. However, in-person or clinician-administered behavioral interventions can be costly and difficult to access for patients [27,28]. Digital health technology, such as smartphone apps, provides the opportunity to deliver effective behavioral treatments for individuals experiencing difficulties with menopause (eg, distress related to hot flashes, depression, and anxiety) in more scalable and cost-effective ways [29-31]. The increased convenience and privacy of digital approaches also have the potential to address health equity gaps by reducing barriers due to travel, logistics, and stigma. They also provide an additional treatment option for women who prefer alternatives to traditional treatments for menopausal symptoms, such as menopause hormone therapy, which is associated with an increased risk of certain health conditions [32].

In recent years, several studies have evaluated the efficacy of smartphone-based interventions for anxiety, depression, and insomnia (among others) [33-36]. However, there is little research on mobile-based behavioral interventions tailored for menopause. Meta-analyses have found few digital solutions for menopause, and most were focused on decision-support for menopausal treatments or content related to education and clinical guidelines, but few interventions for symptom relief [37,38]. Another 2019 study by Atema et al [39] investigated the efficacy of internet-based CBT and found significant reductions in the perceived impact of hot flashes and improved sleep quality. However, the sample was limited to survivors of breast cancer and only enabled the investigation of effects on treatment-induced menopausal symptoms. These studies highlight the dearth of mobile-enabled, self-guided, evidence-based behavioral health interventions for managing menopausal symptoms.

The objective of this study was to assess the efficacy of a novel mobile-based digital care program (DCP) called Caria (Chorus Health Inc) in reducing vasomotor and behavioral health symptoms (eg, hot flashes, depression, anxiety, and poor sleep quality) and explore treatment variables including length of treatment and treatment engagement. Based on evidence from prior research about the efficacy of behavioral interventions for treating menopausal symptoms, we hypothesized that the DCP would reduce VMS-related distress compared to the control group, as measured by the Hot Flush Problem Rating Scale (primary outcome). We also investigated the impact on depression, anxiety, and sleep quality (exploratory outcomes).

XSL•FO
RenderX

Methods

Study Design

The study was a 2-armed, randomized, controlled trial of participants with symptoms of menopause. Participants were recruited via digital channels, provided informed consent, and completed the intervention fully remotely. We followed CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting this trial.

Ethical Considerations

The study was approved by the Institutional Review Board at the University of Illinois Chicago (STUDY2021-0282) and registered with clinicaltrials.gov (NCT04882982) on May 6, 2021. Participants completed an informed consent and had the ability to discontinue participation at any time. All participant data were deidentified prior to analysis. Participants were financially compensated with US $10 for the baseline, US $15 for the 3-week, and US $20 Amazon gift cards for the 6-week assessments. Those who were screened via MTurk received an additional US $1 for completion of the screening survey.

Recruitment

People were recruited between February and October 2022 via advertisements placed on Facebook, Instagram, and Amazon Turk describing a "menopausal symptoms research study." After clicking on the advertisement, potential participants were given additional study information, as well as the link to the screening questionnaire on Qualtrics, hosted by the University of Illinois at Chicago. The screening questionnaire assessed demographic characteristics and inclusion criteria for the study. People were eligible if they had (1) variable menstrual cycles ("If you have had a period in the last 12 mo, has it been irregular;" menopausal transition) or were more than 1 year from their last menstrual period (postmenopause); (2) hot flash or night sweats experienced as problematic (as indicated by an average score of two or higher on three items of the Hot Flush Rating Scale [HFRS]); (3) elevated score on at least one measure of distress—either depression (Patient Health Questionnaire Depression Scale-8 [PHQ-8] of 10 or greater), sleep (Pittsburgh Sleep Quality Index [PSQI] of 5 or greater), or anxiety (Generalized Anxiety Disorder-7 [GAD-7] of 10 or greater); (4) own an apple device able to run the Caria App; (5) English speaking; and (6) aged 30 - 60 years. Menopause status was not fully characterized as the study was seeking to enroll a general population. The presence of other health conditions, such as polycystic ovary syndrome, was not assessed. Internet and mobile or computer literacy were implicit criteria, as participants must be proficient in all three to enroll and access the intervention. Participants were quasi-anonymous, as emails were collected and used as a means of communication. Technical measures (eg, disallowing the same device, IP, or location coordinates to enroll more than once) were used to prevent and detect the use of multiple identities.

A total of 149 people consented to participate in the study and were randomized to receive either the Caria DCP for 6 weeks (experimental group) or to receive a link to the publicly available The Menopause Society website with educational content about menopause symptom management (control group). Participants were followed up at the 3-week and 6-week mark.

Allocation Strategy or Randomization

Participants were randomized into the trial via an algorithm that randomly seeded and shuffled participants into their groups using a treatment-to-control ratio of approximately 3:1, respectively. To prevent selection bias, the researchers had no way of knowing which participants would be allocated to the DCP or control group (allocation concealment). This randomization allocation was chosen as the aims of the trial included exploring the dimensions of treatment length, usability, and engagement. The utilization of unequal allocation is an appropriate method for evaluating clinical utility [40]. After participants were randomized, the intervention group received an email with written and video instructions on how to download and use the Caria app. Control participants received an email with The Menopause Society website link [41]. Neither the researchers nor the participants were blinded to the group allocation due to the design of the study.

Procedures or Study Flow

Participants were first sent a link via email to fill out a baseline questionnaire before starting their respective interventions (DCP vs control). Those in the DCP arm were provided with written and video instructions on app installation and use. Participants were told that they had access to DCP for 6 weeks, but were not explicitly told how often or when to use it. Those in the control group were told that they had access to a website with resources on menopause for 6 weeks. Follow-up surveys were sent to both groups at the 3-week and 6-week mark. At the end of the study, all participants were debriefed via email, and both groups were provided with a link to the Caria app and The Menopause Society Website as resources.

Measures

All measures were self-assessed via web-based questionnaires at 3 time points—baseline, 3 weeks, and 6 weeks. Participants also completed a screening assessment. The selected measures are all commonly administered digitally.

The HFRS is a 5-item self-reported assessment that measures the frequency of hot flushes (in days or weeks) and hot flush problem rating (sum of 3 items ranging from 1 to 10 each) [42]. The Hot Flush Problem Rating assesses the perceived distress associated with hot flushes and their impact on various aspects of psychological well-being, daily functioning, and social interactions. An average score of 2 or higher on 3 items of the HFRS is considered problematic. Prior studies using the HFRS have demonstrated reasonable test-retest reliability and concurrent validity [42,43]. Using Cronbach α, internal consistency among our sample was excellent (α=0.94).

The PHQ-8 is an 8-item self-reported questionnaire that measures current depression [44]. It is well validated as a diagnostic measure across clinical studies. Scores for each item range from 0 to 3, and total scores are calculated by summing the 8 items and can range from 0 to 24. Higher scores equate to higher depression. Scores of 10 or greater are considered clinically significant. The PHQ-8 is considered to be a reliable

XSL•FO
RenderX

and valid measure for screening depression in the general population [45]. Internal consistency among our sample was good (Cronbach α=0.85).

The PSQI is a 19-item self-reported questionnaire that measures sleep quality [46]. A global PSQI score is calculated by summing 7 component scores (comprised of: subjective sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbance, use of sleep medication, and daytime dysfunction). Scores of 5 and above indicate sleep problems. Reliability and validity of the PSQI are considered acceptable [47,48]. Internal consistency among our sample was acceptable (Cronbach α=0.60).

The GAD-7 is a brief 7-item self-reported measure of generalized anxiety [49]. Scores for each item range from 0 to 3, and total scores are calculated by summing the 7 items and can range from 0 to 27. Higher scores equate to higher levels of anxiety. Scores of 10 or greater are considered clinically significant. The GAD-7 has shown high reliability and validity for screening generalized anxiety in the general population [50]. Internal consistency among our sample was excellent (Cronbach α=0.92).

Adherence to the intervention was measured using the number of logins as the primary marker. Data on app persistence (defined in days, from account creation on the app to the end of 6 wk intervention) and the total time spent on the app in that period were also collected (Table 1).

The Usefulness, Satisfaction, and Ease Questionnaire (USE) is a widely accepted instrument for usability that measures satisfaction, usefulness, ease of use, and ease of learning using a 1 - 7 Likert scale [51]. Higher numbers indicate greater usability and satisfaction. It has demonstrated validity for evaluating the subjective usability of software [52]. In this study, a brief version of the measure (5 items) covering Usefulness (1 item), Ease of Use (1 item), Ease of Learning (1 item), Satisfaction (2 items), and Total Scale was used to measure overall usability of and satisfaction with the intervention. Internal consistency of the USE among our sample was high (Cronbach α=0.96). USE ratings are listed in Table 2.

XSL•FO
RenderX

Table . Participant demographics.

Demographics	Overall	Intervention: Caria	Control: TMS[a] website	Tests of independence (*P* value)
Age (years)				.48
Mean (SD)	47.72 (7.68)	47.78 (7.78)	48.46 (7.45)	
Range	30-60	30-60	30-60	
Menopause status, n (%)				.81
Perimenopause	86 (57.7)	48 (42.9)	15 (40.5)	
Menopause	63 (42.3)	64 (57.1)	22 (59.5)	
Ethnicity, n (%)				.04
Hispanic	12 (8.1)	12 (10.7)	0 (0)	
Non-Hispanic	137 (91.9)	100 (89.3)	37 (100)	
Race, n (%)				.51
American Indian/Alaskan Native	2 (1.3)	2 (1.8)	0 (0)	
Asian	7 (4.7)	6 (5.4)	1 (2.7)	
Black or African American	13 (8.7)	8 (7.1)	5 (13.5)	
More than one race	9 (6)	8 (7.1)	1 (2.7)	
Native Hawaiian/Pacific Islander	0 (0)	0 (0)	0 (0)	
White	118 (79.2)	88 (78.6)	30 (81.1)	
Employment status, n (%)				.09
Full-time (>32 h/wk)	84 (56.4)	64 (57.1)	20 (54.1)	
Part-time (<32 h/wk)	29 (19.5)	18 (16.1)	11 (29.7)	
Not employed, seeking	4 (2.7)	4 (3.6)	0 (0)	
Not employed, not seeking	17 (11.4)	16 (14.3)	1 (2.7)	
Retired	5 (3.4)	2 (1.8)	3 (8.1)	
Disabled	8 (5.4)	6 (5.4)	2 (5.4)	
Student	1 (0.7)	1 (0.9)	0 (0)	
Declined/Unknown	1 (0.7)	1 (0.9)	0 (0)	
Education level, n (%)				.41
<12th grade	1 (.7)	1 (.9)	0 (0)	
High school graduate or GED[b]	12 (8.1)	11 (9.8)	1 (2.7)	
Some college or associate's (2 y) degree	47 (31.5)	32 (28.6)	15 (40.5)	
Bachelor's (4 y degree)	53 (35.6)	39 (34.8)	14 (37.8)	
Graduate degree or more	36 (24.2)	29 (25.9)	7 (18.9)	
Income level (in US $), n (%)				.81
<10,000	2 (1.3)	1 (.9)	1 (2.7)	
10,000-25,000	11 (7.4)	7 (6.3)	4 (10.8)	
26,000-50,000	23 (15.4)	18 (16.1)	5 (13.5)	
51,000-75,000	34 (22.8)	27 (24.1)	7 (18.9)	

XSL•FO
RenderX

Demographics	Overall	Intervention: Caria	Control: TMS[a] website	Tests of independence (*P* value)
76,000-100,000	32 (21.5)	25 (22.3)	7 (18.9)	
>100,000	46 (31.5)	34 (30.4)	13 (35.1)	
Marital status, n (%)				.24
Single, never married	29 (19.5)	20 (17.9)	9 (24.3)	
Married	87 (58.4)	65 (58.0)	22 (59.5)	
Separated	6 (4)	6 (5.4)	5 (13.5)	
Divorced	23 (15.4)	18 (16.1)	1 (2.7)	
Widowed	1 (.7)	0 (0)	1 (2.7)	
Other	3 (2)	3 (2.7)	0 (0)	

[a]TMS: The Menopause Society .

[b]GED: General Education Development.

Table . App engagement and usefulness ratings (Caria-group only; n=100).

Measure	Value (n=100)
Number of sessions	
Mean (SD)	53.24 (66.97)
Range	1 - 398
Time on app (hours)	
Mean (SD)	3.77 (15.57)
Range	0.01 - 157
App persistence (days)	
Mean (SD)	30 (15.73)
Range	0 - 42
USE[a]—Usefulness of app	
Mean (SD)	4.96 (1.86)
Range	1-7
USE—Ease of use	
Mean (SD)	5.25 (2.06)
Range	1-7
USE—Ease of learning app	
Mean (SD)	5.31 (2.02)
Range	1-7
USE—Satisfaction with app	
Mean (SD)	5.04 (2.02)
Range	1-7
USE Total Score, mean (SD)	20.6 (7.5)

[a]USE: Usefulness, Satisfaction, and Ease Questionnaire.

Interventions

Control Condition

Participants randomized to the control group received a link to educational material on menopause symptoms from The Menopause Society website, a well-respected menopause organization. The Menopause Society website contains hyperlinks to internal and external resources and averages less than one page of text per web page.

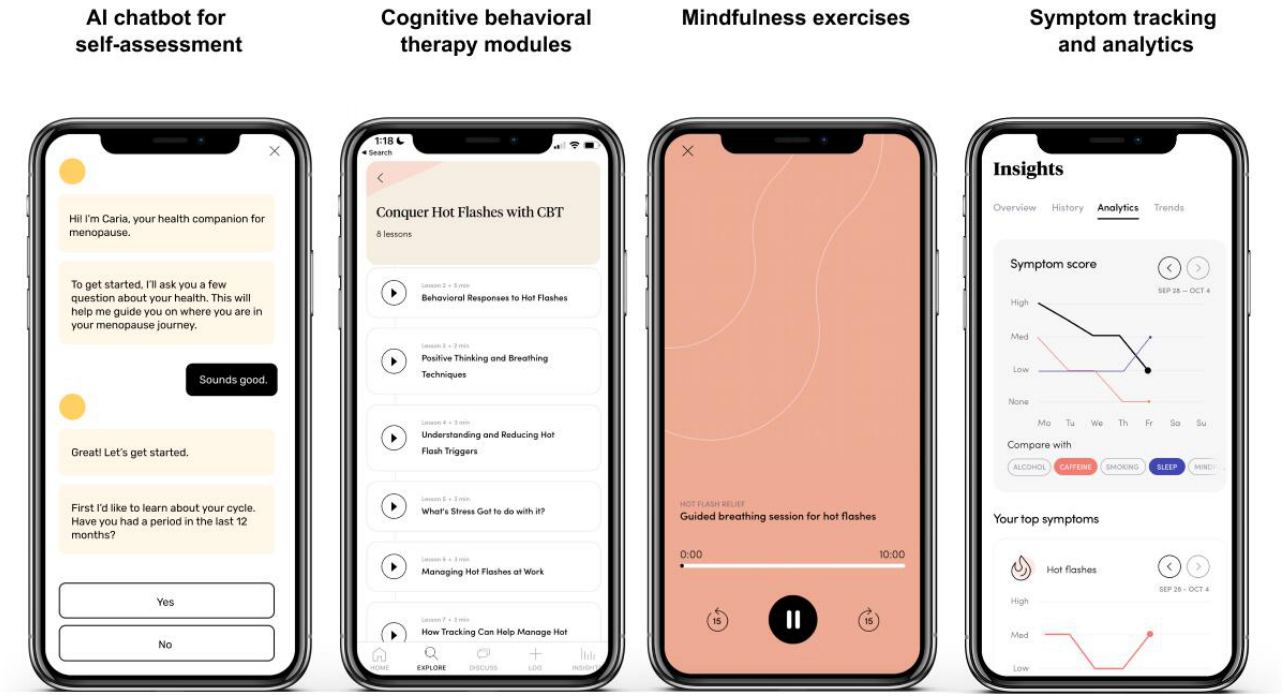
XSL•FO
RenderX

Active Condition—Caria DCP

Participants randomized to the active intervention group (ie, Caria DCP) received a link to download Caria (version 1.9.12), a smartphone-based app that includes a self-guided DCP designed for menopause symptoms. No additional cointerventions were provided. Participants were provided free access to the app (available on iOS) for the duration of the study, along with written and video instructions on app installation and use. There were no bug fixes or content changes between the study commencement and the study conclusion.

The Caria app includes an artificial intelligence (AI)-chatbot-assisted menopause symptom assessment and a 6-week multimodal, self-guided DCP. For each day of the program, participants receive a number of modules and tasks to complete, including a CBT lesson, a mindfulness-based exercise in the form of a guided audio session, a mobility exercise video, and a community discussion topic where participants can receive or provide peer support asynchronously with other app users. Video content on the app is approximately 2 - 10 minutes in duration, and the text content is approximately 800 - 1000 words (roughly a 2 - 5 min read). Figure 1 shows examples of app features. The program design was informed by prior research and care guidelines based on the efficacy of approaches such as CBT, mindfulness, and mobility for menopausal symptoms [25,27,53,54].

Figure 1. Examples of Caria app features and digital care program content.



The DCP is divided into 6 weeks, covering themes such as understanding the menopausal transition (wk 1) and learning behavioral approaches to managing common symptoms, with a particular focus on anxiety and depression (wk 2), hot flashes and night sweats (wk 3), and sleep (wk 4). The DCP also covers behavioral techniques for improving mental wellness during the menopause transition (wk 5) and managing age-related health conditions and comorbidities, such as heart health and cognitive health (wk 6). All modules are available to users at any time and can be accessed in any order. Users are not required to complete one module before proceeding to the next. This flexible approach was designed to accommodate individual preferences and needs, allowing users to engage with the content that is most relevant to them at any given time. Other features of the app include a daily symptom tracker, data-driven health analytics, and on-demand lessons for a variety of topics such as managing the impact of menopause on work and personal relationships. Personalized health tips and recommendations are provided to participants based on their individual symptom logs. Participants can opt in to daily reminders and notifications to encourage regular engagement with the program.

Statistical Analysis

Power Analysis

The sample size was determined based on a previous study examining hot flash severity as the primary outcome in a self-guided CBT intervention for menopausal symptoms in a similar general (not treatment-induced menopause) population [26]. G*power was used to calculate a sample size for a 3:1 allocation ratio with an effect size of 0.75, resulting in a sample size of 126 (94:32), therefore, a sample of 145 after accounting for 15% attrition would be adequately powered to test hypotheses [55].

Treatment Outcomes

Average hot flash severity (HFRS total score) served as the primary outcome measure and was submitted to a repeated measures ANOVA where time (T1, T2, and T3) was the within-subjects factor and treatment arm (Caria, control) was the between-subjects factor.

Exploratory outcomes were examined by subgroups (ie, participants with cutoffs reflective of depression, anxiety, or

XSL•FO
RenderX

poor sleep quality in the clinical range). This was done to examine the intervention impact on those with symptom levels indicative of a need for treatment. Since analyses based on subgroups reduced the power to detect differences between treatment arms, a planned comparison approach was used. Specifically, repeated measures ANOVA, where time (T1, T2, and T3) was the within-subjects factor, was performed within each treatment arm. Significant main effects of time were followed up with 2-tailed paired *t* tests to aid in the interpretation of results. We also evaluated whether participants who completed all 6 weeks of treatment significantly differed from noncompleters in terms of clinical characteristics (eg, hot flush severity). We did not expect differences; however, it is important to evaluate to determine whether premature discontinuation was driven by differences in clinical severity.

All analysis was 2-tailed with the α level set to .05. Participants with incomplete data were omitted from the analysis. Statistical parameters include central tendency (ie, means) and SD for variance. Significant omnibus findings (eg, ANOVA) were followed-up with simple effects analysis (ie, 2-tailed independent *t* tests and paired *t* tests). For the primary outcome measure, hot flash severity, Bonferroni was used to adjust for multiple comparisons. Hedges *g* (mean baseline minus mean posttreatment or SD pooled) was used as the effect size for primary and exploratory outcome measures. Analysis was performed in SPSS (version 27; IBM Corp).

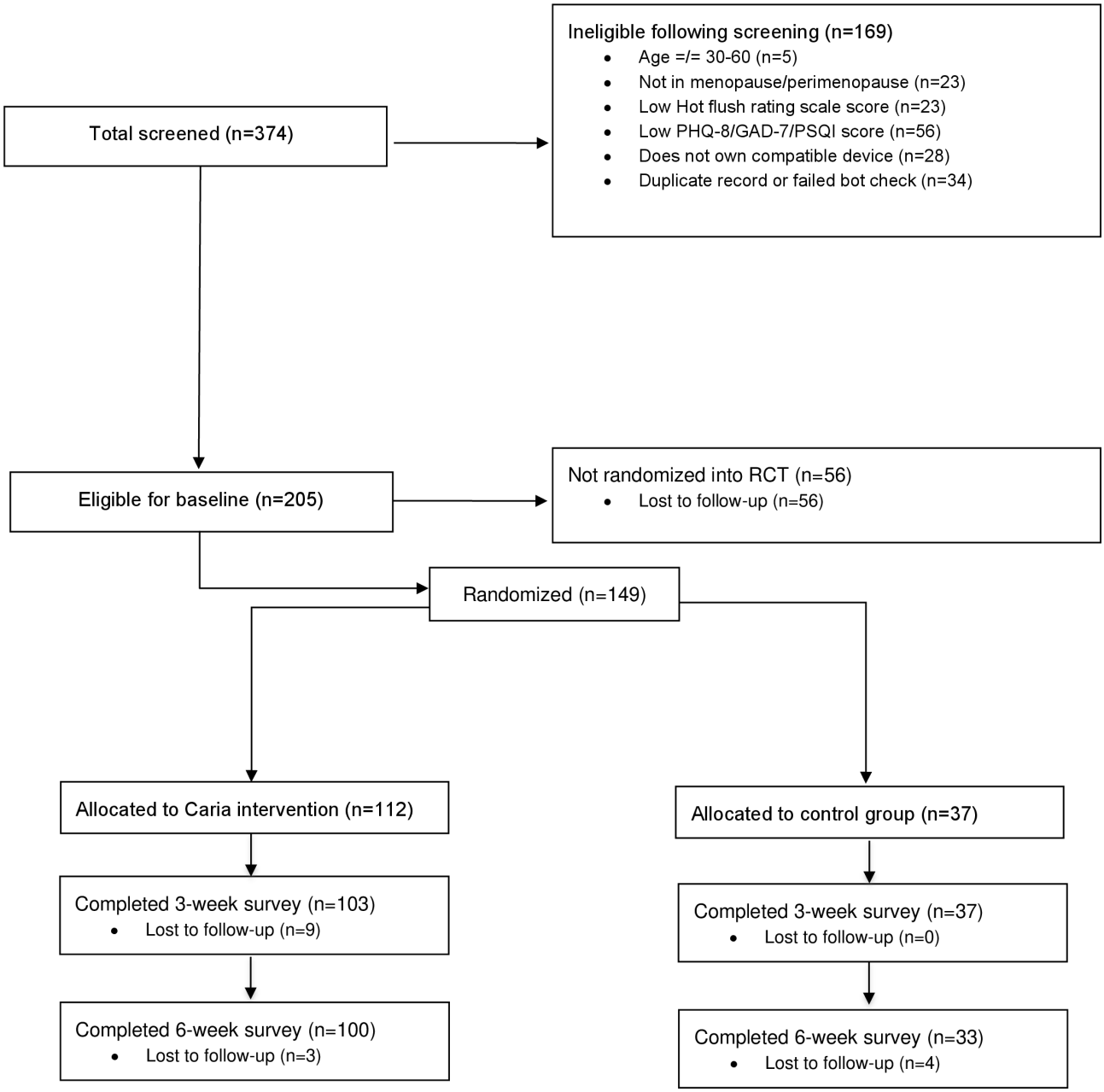
Results

Overview

Eligibility screening was conducted for 374 participants. A total of 169 participants were not eligible to participate, and 56 participants were lost to follow-up. In total, 149 participants were randomly assigned to either the active DCP (n=112) or the control group (n=37). Figure 2 shows the consort diagram with details of the group allocation and follow-up.

XSL•FO
RenderX

Figure 2. Study flowchart according to CONSORT. CONSORT: Consolidated Standards of Reporting Trials; GAD-7: Generalized Anxiety Disorder-7; PHQ-8: Patient Health Questionnaire Depression Scale-8; PSQI: Pittsburgh Sleep Quality Index; RCT: randomized controlled trial.



User Statistics

Participants were recruited from the United States and were between the ages of 30 and 60 (mean 47.68, SD 7.69) years. Among the sample, 86 (57.7%) participants were in perimenopause (as defined by having a period in the last 12 mo or irregular periods) and the remaining 63 (42.3%) were postmenopausal. The racial breakdown of the participants was as follows: 118 (79.2%) White, 13 (8.8%) Black/African American, 9 (6%) Mixed, 7 (4.7%) Asian, and 2 (1.3%) American Indian/Alaskan Native, out of a total of 149 participants. Racial classification terms were provided by researchers, and responses were selected by participants. Table 1 shows the participant demographics.

Baseline Characteristics

Participants randomized to the active Caria arm (n=112) had marginally lower hot flash severity than those assigned to the control arm (n=37; t_{147}=1.97; P=.05); frequency of hot flashes did not differ (t_{147}=1.2; P=.23) nor did depression, anxiety, or sleep quality (lowest P=.34) between treatment conditions. Regarding demographic characteristics, participants in the Caria versus control arm did not differ in age (t_{147}=.67; P=.97), level of education in years (t_{147}=.11; P=.91), or race (χ^2_4=3.3; P=.51).

When evaluating clinical depression (ie, participants with PHQ-8 total score ≥10), average depression did not differ between participants in the Caria condition (n=56 mean 14.4, SD 3.9) and control condition (n=21; mean 15.0, SD 3.7; t_{75}=0.6; P=.50). For clinical anxiety (ie, participants with GAD-7 total score ≥10), average anxiety did not differ between participants in the

XSL•FO
RenderX

Caria condition (n=39; mean 14.9, SD 3.2) and control condition (n=15; mean 16.0, SD =3.8; t_{52}=1.1; *P*=.29). Concerning clinical poor sleep quality (ie, participants with PSQI global score ≥5), average sleep quality did not differ between participants in the Caria condition (n=102; mean 11.0, SD 2.9) and control condition (n=35; mean 11.5, SD 3.7*;* t_{135}=0.9; *P*=.37).

Engagement

Engagement with the DCP was tracked by measuring the number of unique sessions and time spent on the app. Over the course of the 6-week intervention, participants logged into the Caria app an average of 53.2 times and spent 3.8 hours on the app. On average, the time between the first and last login was 30 days. Data were not collected on access to the control educational materials. Table 2 provides more information.

Treatment Outcome

Participants who completed all 6 weeks of treatment (133/149, 89%) were similar to noncompleters (16/149, 10.7%) in terms of baseline hot flash severity (t_{147}=1.18; *P*=.23), depression (t_{146}=.38; *P*=.69), anxiety (t_{147}=.04; *P*=.96), and sleep quality (t_{147}=1.45; *P*=.14). Regarding hot flash frequency, baseline frequency was marginally lower for completers relative to noncompleters (t_{147}=1.9; *P*=.05). Finally, there was no difference with regard to distribution of completers or noncompleters between treatment conditions (χ^2_1=0.0; *P*=.98).

Primary Outcome Measure

Repeated measures ANOVA results for evaluating change in hot flash severity revealed main effects for time ($F_{2,262}$=9.82; *P*<.001) and treatment arm ($F_{1,131}$=6.08; *P*=.01), and there was a significant time × treatment arm interaction ($F_{2,262}$=3.23; *P*=.04). To evaluate the main effect of time, treatment arm was collapsed and 2-tailed paired *t* tests showed hot flash severity significantly decreased from baseline to week 3 (t_{139}=4.67; *P*<.001) but not from weeks 3 to 6 (t_{132}=1.25; *P*=.21). The significant finding was significant when adjusting for multiple comparisons (*P*=.02). To examine the main effect of the treatment arm, time was collapsed, and 2-tailed independent *t* tests revealed participants in the Caria treatment arm had less hot flash severity than participants in the control arm (t_{147}=2.72; *P*=.007). The significant finding was significant when adjusting for multiple comparisons (*P*=.02).

To explore the interaction, 2-tailed paired *t* tests were performed within each treatment arm. Participants in the Caria arm showed a significant decrease in hot flash severity from baseline to week 3 (t_{102}=3.54; *P*=.001) and from weeks 3 to 6 (t_{99}=2.90; *P*=.005). Findings were significant when adjusting for multiple comparisons (*P*=.02). Results for the control arm revealed a significant decrease in hot flash severity from baseline to week 3 (t_{36}=3.33; *P*=.002) but not from weeks 3 to 6 (t_{32}=1.19; *P*=.24). The significant finding was significant when adjusting for multiple comparisons (*P*=.02). Two-tailed independent *t* test analysis for each time point showed the marginally lower baseline hot flash severity for participants eventually assigned to the Caria relative to control arm reported previously (t_{147}=1.97; *P*=.05) and no significant treatment arm effects were observed at week 3 (t_{138}=1.70; *P*=.09). However, after completing treatment (ie, week 6), participants in the active treatment arm exhibited less hot flash severity than participants in the control arm (t_{131}=3.21; *P*=.002). The significant week 6 finding was significant when adjusting for multiple comparisons (*P*=.02).

In summary, participants randomized to the Caria treatment arm exhibited significantly greater improvement in hot flashes after completing 6 weeks of treatment relative to participants randomized to the control arm.

Exploratory Outcome Measures

Depression

For participants with a PHQ-8 total score ≥10 (77/149, 51.6%), the repeated measures ANOVA of change in PHQ-8 scores showed a main effect of time for participants in the Caria treatment arm ($F_{2,96}$=15.2; *P*<.001) but not control arm ($F_{2,40}$=2.0; *P*=.15). Follow-up paired *t* tests for participants in the Caria treatment arm showed depression decreased from baseline to week 3 (t_{49}=3.3; *P*=.002). Depression also decreased from weeks 3 to 6 (t_{48}=2.3; *P*=.02).

Sleep Quality

For participants with a PSQI global score ≥5 (137/149, 91.9%), the repeated measures ANOVA of PSQI scores showed a main effect of time for participants in the Caria treatment arm ($F_{2,186}$=7.8; *P*=.001) but not control arm ($F_{2,62}$=1.3; *P*=.28). Follow-up paired *t* tests for the Caria condition showed poor sleep quality significantly decreased from baseline to week 3 (t_{95}=3.9; *P*<.001). Yet, there was no change in sleep quality from weeks 3 to 6 (t_{93}=0.2; *P*=.81).

Anxiety

For participants with a GAD-7 total score ≥10 (54/149, 36.2%), the repeated measures ANOVA of change in GAD-7 scores showed a main effect of time for participants in the Caria treatment arm ($F_{2,68}$=5.0; *P*=.01) and control arm ($F_{2,26}$=3.8; *P*=.03). In the Caria treatment arm, follow-up paired *t* tests showed anxiety significantly decreased from baseline to week 3 (t_{34}=2.4; *P*=.02). Yet, there was no change in anxiety from weeks 3 to 6 (t_{34}=0.9; *P*=.36).

Results were similar in the control arm. Specifically, follow-up paired *t* tests showed anxiety significantly decreased from baseline to week 3 (t_{14}=2.6; *P*=.02). However, there was no change in anxiety from weeks 3 to 6 (t_{13}=1.6; *P*=.13).

In summary, in the clinically depressed cohort, depression decreased across all time points within the Caria condition but not the control condition. In contrast, for clinically anxious participants, anxiety decreased during the early, but not later, phase of treatment, regardless of treatment arm. Finally, for participants with poor clinical sleep quality, sleep improved within the active condition but not the control condition. In addition, the significant improvement in the active condition was driven by the change in the early phase of treatment. Table 3 provides the details for all measures at all time points for each treatment arm.

XSL•FO
RenderX

Table . Group means, SDs, and *P* values comparing timepoints with effect sizes of participants randomized to use the DCP[a] versus a control.

Outcome	Group					
	Baseline, mean (SD)	3 weeks, mean (SD)	*P* value (comparing baseline to 3 weeks)	6 weeks, mean (SD)	*P* value (comparing 3 weeks to 6 weeks)	Effect Size (Hedges *g*; baseline to 6 weeks)
Primary outcomes						
HFRS[b]—Problem						0.64
Caria DCP (n=103)	16.4 (6.7)	14.7 (6.3)[c]	.001	13.6 (6.6)[c]	.005	
Control (n=37)	19.1 (7.3)	16.9 (7.4)[c]	.002	17.8 (7.2)	.24	
HFRS—Frequency						0.02
Caria DCP (n=103)	2.8 (2.7)	2.5 (4.1)	.50	2.9 (5.4)	.61	
Control (n=37)	2.3 (3.2)	2.2 (2.7)	.94	2.7 (2.7)	.30	
Exploratory outcomes						
Depression (PHQ-8)						0.45
Caria DCP (n=50)[d]	14.0 (3.8)	12.2 (4.7)[c]	.002	11.2(5.3)[c]	.02	
Control (n=21)	15.0 (3.7)	13.3 (4.5)	.12	13.4 (4.1)	.93	
Sleep Disturbance (PSQI)						0.25
Caria DCP (n=100)[e]	10.7 (3.1)	10.0 (3.4)[c]	<.001	10.0 (3.5)	.81	
Control (n=35)[f]	11.5 (3.7)	10.9 (4.0)	.10	11.0 (3.7)	.82	
Anxiety (GAD-7)						0.33
Caria DCP (n=35)	14.8 (3.3)	13.3 (4.7)[c]	.02	12.7 (5.5)	.36	
Control (n=15)[g]	16.0 (3.8)	12.9 (5.4)[c]	.02	14.5 (4.5)	.13	

[a]DCP: digital care program.
[b]HFRS: Hot Flush Rating Scale.
[c]*P*<.05
[d]For 3 and 6 weeks, n=49.
[e]For 3 and 6 weeks, n=97.
[f]For 3 and 6 weeks, n=32.
[g]For 3 and 6 weeks, n=14.

Discussion

Principal Results

This study is the first evaluation of a novel smartphone-delivered DCP, Caria, in women with symptoms of menopause. Results from this RCT supported the primary hypothesis as distress related to hot flashes significantly decreased in participants randomized to the Caria arm compared to the control arm. Exploratory clinical measures of depression and sleep problems were also shown to improve in participants in the DCP arm but not in the control arm. The depression effect size was comparable to therapist-delivered interventions for depression in similar mildly distressed populations [56,57], suggesting that the digital behavioral intervention that uses CBT and mindfulness strategies may be an effective treatment for individuals experiencing symptoms related to perimenopause or menopause. This supports a digital, self-guided model of menopause-related care.

Specifically, symptoms of distress from hot flashes and depression demonstrated improvement in the DCP group over 6 weeks but not in the control group. Sleep disturbance for those in the DCP group improved over the first 3 weeks and remained stable through treatment, while there was no change in the control group. For those who used the DCP, distress from hot flashes decreased by 17.1%, depression symptoms decreased by 12.3%, anxiety symptoms decreased by 14.3%, and sleep problems decreased by 5.9%. Those with elevated anxiety symptoms who used Caria also showed a decrease in symptoms, but so did those in the control group. While there was a change in anxiety symptoms and a small to moderate effect size

XSL•FO
RenderX

(g=0.33) for DCP use, the small sample size of those with elevated symptoms makes drawing meaningful conclusions difficult.

Clinical measures were examined from baseline to 3 weeks (midpoint) and 3 weeks to 6 weeks (posttreatment) in order to explore the length of treatment needed and change over time. In this sample, participants in the control group did experience improvement in hot flash symptoms and anxiety in the first 3 weeks of the study, suggesting there may have been value in the web-based educational material related to menopause symptoms (the control intervention), as it validates the symptoms experienced. It could also be a placebo response to participation in a research trial. However, those who used Caria showed additional improvement in hot flash distress and depression in the second half of the intervention (3-6 wk), while there was no change among participants in the control group. In addition, a significant improvement in sleep among those in the Caria group occurred during the first 3 weeks of intervention, while no change occurred in the control group. Overall, findings support that the DCP offers valuable intervention beyond web-based educational material and that 6 weeks of intervention provides more benefit than 3 weeks.

Notably, the frequency of hot flashes did not change significantly in this study; however, there was a reduction in the distress associated with these symptoms. This is a significant finding, given that the level of perceived distress caused by VMS is linked to quality of life and the likelihood of seeking health care services, rather than the frequency [58]. This study demonstrates that the DCP, despite not directly impacting the number of hot flashes experienced, effectively alleviates distress and consequently offers a promising avenue for enhancing overall quality of life and impacting health care utilization. It provides an option for women who prefer nonpharmacological interventions [14].

With regard to depression, depressive symptoms typically increase throughout the menopause transition, and depression rates are higher in women with VMS [59]. Clinical guidelines recommend evidence-based psychotherapies such as CBT as the primary treatment for depressive symptoms during the menopause transition [28]. However, long waitlists, high costs of care, and difficulty finding mental health providers with expertise in menopause limit women's ability to receive adequate treatment. Given that this is a high-risk period in a woman's life, a low-intensity intervention that can be easily disseminated has the potential for significant impact.

The effect size of Caria for depressive symptoms (effect size, g=0.45) was comparable to therapist-delivered interventions for a similar level of mild depressive symptoms (effect size, d=0.42) and better than many unguided digital interventions for depression (effect size, g=0.27) [56,57]. Intervening when depressive symptoms are mild may prevent the development of more severe episodes, yielding substantial benefits for patients, families, and the health care system. Individuals with mild depressive symptoms have been found to derive benefit from unguided interventions [60], and our data further supports the value of a self-guided DCP for menopause-related behavioral health that can be easily accessed.

As for sleep, approximately 80% of women who experience hot flashes have difficulty sleeping, and more severe hot flashes make it difficult to maintain sleep [61]. For those in the Caria group, improvement in sleep occurred over the first 3 weeks of treatment, and then, improvements remained stable during the second half of the intervention. The premise of CBT is that changing how one perceives an experience may change how one feels or behaves in response. Thus, it is possible that decreased distress about hot flashes (ie, improved coping related to hot flashes) demonstrated by Caria participants may lead to improved sleep through a better ability to respond to the experience of hot flashes. However, further research is needed to test this proposed mechanism of action.

It is also important to note that comorbidities and concurrent symptoms are a uniquely challenging aspect of menopause for women in midlife, with a significant impact on women's well-being and productivity. It is the presence of concurrent menopausal symptoms that has been linked to productivity losses in the workplace, rather than a specific symptom [5]. In this sample, all participants had elevated levels of perceived distress from hot flashes, as well as one additional elevated mental health symptom. However, the majority of our participants had considerable levels of comorbidity; 33% (49/149) of participants had elevated levels of sleep difficulties, depression, and anxiety; 19% (28/149) had both elevated sleep difficulties and depression; and 3% (4/149) had both sleep difficulties and anxiety. The findings of this study and improvement in multiple symptoms point to the importance of digital interventions for menopause, taking a comprehensive approach to symptom management and targeting multiple concerns.

Sleep disturbances, in particular, have far-reaching effects as they have been associated with mood fluctuations, memory problems, obesity, and other cardiovascular risk factors, which are prevalent among women in midlife [62]. The noteworthy results of this study in demonstrating improvement in sleep-related symptoms hold significant promise, as addressing sleep issues may offer a potential pathway for enhancing overall health outcomes in menopausal women, including the management of other health conditions.

In addition to symptom improvement, participants in the DCP arm demonstrated consistent and high levels of engagement while reporting favorable levels of usability. Participants rated ease of use, ease of learning, usefulness, and satisfaction around 5 out of a 7-point scale (with 7 being most favorable). Participants engaged with the DCP for an average of 30 days in the 42-day trial, with a mean of 53.2 (SD 66.97) logins. Engagement and use of the Caria app are notable, particularly given the low rates of engagement typically found in health apps. A 2019 study examining mental health apps with similar components to Caria (mindfulness, peer support, and tracking) found 30-day retention rates of only 3.3% [63].

Comparison With Prior Work

Although there are a few other self-guided digital interventions for menopause symptoms, the DCP performs favorably in comparison, with a higher effect for hot flash distress, and a similar impact on sleep [39]. The small to moderate range of

XSL•FO
RenderX

symptom reduction in this study may be attributed to the relatively low baseline symptom severity reported by the majority of participants in the sample, resulting in less room for substantial improvement in symptom outcomes. While the effect sizes observed in the study were in the small to moderate range, it is worth noting the potential impact. By effectively addressing mild symptoms, self-guided DCPs like the one evaluated in this study have the potential to prevent symptom exacerbation into more severe and costly conditions and play a role in alleviating the burden on outpatient health services and associated health care costs [64-66].

Unlike many existing interventions that rely on evidence-informed content without rigorous testing, this study assessed the efficacy of the DCP using the gold-standard RCT methodology. Another significant strength of this study is the highly engaging design and self-guided delivery of the intervention. Participants demonstrated a high level of adherence and engagement with the intervention without the need for external support or coaching, suggesting the potential for high scalability. Furthermore, this study examined a range of symptoms that commonly impact women in midlife. Given the complexity of the menopausal transition, the comprehensive assessment of various aspects of physical and psychological health in this study provides a more holistic understanding of the intervention's effectiveness in addressing the multifaceted challenges faced by women in midlife.

Limitations

This trial has certain limitations that warrant consideration. First, participants in this study exhibited relatively low levels of behavioral health problems (anxiety, depression, and sleep disturbance) at baseline. Testing the DCP in a population with higher levels of symptoms would provide valuable insights into its effectiveness in addressing more severe symptoms and determining its suitability for a broader range of individuals. The predominantly White (118/149, 79%) and highly educated (89/149, 60% with college education or higher) sample may limit the generalizability of the findings. Future studies should aim to test the DCP in a more diverse sample to better capture the experiences of underrepresented populations. Recruitment exclusively through digital advertisements may have introduced a bias toward participants who are comfortable using technology and open to receiving health care in a digital format. As a result, the findings may not fully generalize to a broader population that may have varying levels of digital literacy or preferences for alternative modes of health care delivery. Additionally, the app's infrastructure did not allow for monitoring the completion of specific modules, therefore, we were unable to measure adherence to individual components. Future research can benefit from more granular tracking features on the app. Neither participants nor researchers were blinded to the treatment condition. This can lead to overestimation of treatment effects. Finally, the use of self-report measures and the lack of long-term follow-up are limitations of the study. Future studies should incorporate posttreatment follow-up assessments with clinician-administered assessments to evaluate the sustained effects of the intervention over an extended period.

Further Research

Future research can also build upon these findings and explore additional enhancements to the DCP, such as the incorporation of wearables and AI. Wearables like smartwatches or fitness trackers can monitor symptoms such as sleep disturbances, heart rate, and body temperature, and have the potential to provide objective and continuous data on vasomotor and behavioral health symptoms [67,68]. Combined with AI and software apps, they can also facilitate timely interventions and personalized recommendations. For instance, AI-powered systems can identify specific triggers or patterns that exacerbate symptoms and provide tailored strategies for managing them. This approach opens up new possibilities for targeted interventions and improving health outcomes.

Conclusions

In conclusion, this study evaluated the efficacy of a 6-week DCP for common symptoms of menopause and found significant improvements in distress from VMS, depression, and sleep problems. The findings reinforce the significant impact of menopausal symptoms and validate the potential of digital interventions to address these challenges effectively. By leveraging the advantages of mobile technology and evidence-based interventions, the DCP provided a more scalable and accessible approach to support women in midlife experiencing symptoms of menopause. The improvements observed in this study contribute to the growing body of evidence supporting the effectiveness of digital behavioral interventions and fill a gap in the research on digital interventions targeted at women in midlife.

Note on Terminology

This paper uses the term "women" in concordance with the World Health Organization. [69] However, it is important to acknowledge that people who identify as trans and nonbinary also experience menopause. There is currently a scarcity of readily available data on these populations. Further research efforts are required to better understand and address the unique experiences and challenges faced by nonbinary and transgender individuals in the context of menopause.

Acknowledgments

We would like to thank all participants and organizations that took part in the study. The Caria app was developed by Chorus Health Inc, who also sponsored the study.

XSL•FO
RenderX

Authors' Contributions

JD, AR, and SG conceived of the study and designed it. YH oversaw participant recruitment and data collection. HK conducted the statistical analyses and provided an interpretation of the data and results. JD, AR, SG, and HK drafted the manuscript. All authors revised the manuscript and approved the final content.

Conflicts of Interest

JD is an independent scientific and clinical consultant who received an honorarium from Chorus Health, the sponsor of this study. JD has also received consulting fees from My Alloy, a menopause company unrelated to this research. AR and SG are employed by and hold equity in Chorus Health. YLH received compensation from Chorus Health as a paid contractor.

Checklist 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File, 2879 KB - mhealth_v13i1e58204_app1.pdf]

References

1. Peacock K, Carlson K. Menopause: StatPearls Publishing; 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK507826/ [accessed 2025-06-02]
2. Delamater L, Santoro N. Management of the perimenopause. Clin Obstet Gynecol 2018 Sep;61(3):419-432. [doi: 10.1097/GRF.0000000000000389] [Medline: 29952797]
3. Thurston RC, Joffe H. Vasomotor symptoms and menopause: findings from the study of women's health across the nation. Obstet Gynecol Clin North Am 2011 Sep;38(3):489-501. [doi: 10.1016/j.ogc.2011.05.006] [Medline: 21961716]
4. Takahashi TA, Johnson KM. Menopause. Med Clin North Am 2015 May;99(3):521-534. [doi: 10.1016/j.mcna.2015.01.006] [Medline: 25841598]
5. Whiteley J, DiBonaventura MD, Wagner JS, Alvir J, Shah S. The impact of menopausal symptoms on quality of life, productivity, and economic outcomes. J Womens Health (Larchmt) 2013 Nov;22(11):983-990. [doi: 10.1089/jwh.2012.3719] [Medline: 24083674]
6. OlaOlorun FM, Shen W. Menopause: Oxford Research Encyclopedia of Global Public Health; 2020. [doi: 10.1093/acrefore/9780190632366.013.176]
7. DePree B, Shiozawa A, King D, et al. Association of menopausal vasomotor symptom severity with sleep and work impairments: a US survey. Menopause 2023 Sep 1;30(9):887-897. [doi: 10.1097/GME.0000000000002237] [Medline: 37625086]
8. Evandrou M, Falkingham J, Qin M, Vlachantoni A. Menopausal transition and change in employment: evidence from the National Child Development Study. Maturitas 2021 Jan;143(96-104):96-104. [doi: 10.1016/j.maturitas.2020.09.010] [Medline: 33308643]
9. Berecki-Gisolf J, Begum N, Dobson AJ. Symptoms reported by women in midlife. Menopause 2009;16(5):1021-1029. [doi: 10.1097/gme.0b013e3181a8c49f]
10. Faubion SS, Enders F, Hedges MS, et al. Impact of menopause symptoms on women in the workplace. Mayo Clin Proc 2023 Jun;98(6):833-845. [doi: 10.1016/j.mayocp.2023.02.025] [Medline: 37115119]
11. Manson JE, Crandall CJ, Rossouw JE, et al. The women's health initiative randomizedtrials and clinical practice. JAMA 2024 May 28;331(20):1748-1760. [doi: 10.1001/jama.2024.6542] [Medline: 38691368]
12. DePree B, Houghton K, Shiozawa A, et al. Treatment and resource utilization for menopausal symptoms in the United States: a retrospective review of real-world evidence from US electronic health records. Menopause 2023 Jan 1;30(1):70-79. [doi: 10.1097/GME.0000000000002095] [Medline: 36576444]
13. Szabo RA, Marino JL, Hickey M. Managing menopausal symptoms after cancer. Climacteric 2019 Dec;22(6):572-578. [doi: 10.1080/13697137.2019.1646718] [Medline: 31433675]
14. Shufelt CL, Brown V, Carpenter JS. The 2023 nonhormone therapy position statement of The North American Menopause Society. Menopause 2023;30(6):573-590. [doi: 10.1097/GME.0000000000002200]
15. Faubion SS, Crandall CJ, Davis L. The 2022 hormone therapy position statement of The North American Menopause Society. Menopause 2022;29(7):767-794. [doi: 10.1097/GME.0000000000002028] [Medline: 36040433]
16. Ye M, Shou M, Zhang J, et al. Efficacy of cognitive therapy and behavior therapy for menopausal symptoms: a systematic review and meta-analysis. Psychol Med 2022 Feb;52(3):433-445. [doi: 10.1017/S0033291721005407] [Medline: 35199638]
17. Fenn K, Byrne M. The key principles of cognitive behavioural therapy. InnovAiT: Education and inspiration for general practice 2013 Sep;6(9):579-585. [doi: 10.1177/1755738012471029]
18. Hofmann SG, Gómez AF. Mindfulness-based interventions for anxiety and depression. Psychiatr Clin North Am 2017 Dec;40(4):739-749. [doi: 10.1016/j.psc.2017.08.008] [Medline: 29080597]
19. Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440. [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]

XSL•FO
RenderX

20. Lim JA, Choi SH, Lee WJ, et al. Cognitive-behavioral therapy for patients with chronic pain. Medicine (Baltimore) 2018;97(23):e10867. [doi: 10.1097/MD.0000000000010867]
21. Fordham B, Sugavanam T, Edwards K, et al. The evidence for cognitive behavioural therapy in any condition, population or context: a meta-review of systematic reviews and panoramic meta-analysis. Psychol Med 2021 Jan;51(1):21-29. [doi: 10.1017/S0033291720005292] [Medline: 33455594]
22. Keng SL, Smoski MJ, Robins CJ. Effects of mindfulness on psychological health: a review of empirical studies. Clin Psychol Rev 2011 Aug;31(6):1041-1056. [doi: 10.1016/j.cpr.2011.04.006] [Medline: 21802619]
23. Maddineshat M, Keyvanloo S, Lashkardoost H, Arki M, Tabatabaeichehr M. Effectiveness of group cognitive-behavioral therapy on symptoms of premenstrual syndrome (PMS) . Iran J Psychiatry 2016 Jan;11(1):30-36. [Medline: 27252766]
24. Hunter MS, Chilcot J. Is cognitive behaviour therapy an effective option for women who have troublesome menopausal symptoms? Br J Health Psychol 2021 Sep;26(3):697-708. [doi: 10.1111/bjhp.12543] [Medline: 34101946]
25. Green SM, Donegan E, Frey BN, et al. Cognitive behavior therapy for menopausal symptoms (CBT-Meno): a randomized controlled trial. Menopause 2019 Sep;26(9):972-980. [doi: 10.1097/GME.0000000000001363] [Medline: 31453958]
26. Hardy C, Griffiths A, Norton S, Hunter MS. Self-help cognitive behavior therapy for working women with problematic hot flushes and night sweats (MENOS@Work): a multicenter randomized controlled trial. Menopause 2018 May;25(5):508-519. [doi: 10.1097/GME.0000000000001048] [Medline: 29315132]
27. NICE guideline: menopause diagnosis and management. National Institute for Health and Care Excellence. 2015. URL: https://www.nice.org.uk/guidance/ng23 [accessed 2024-02-03]
28. Maki PM, Kornstein SG, Joffe H, et al. Guidelines for the evaluation and treatment of perimenopausal depression: summary and recommendations. J Womens Health (Larchmt) 2019 Feb;28(2):117-134. [doi: 10.1089/jwh.2018.27099.mensocrec] [Medline: 30182804]
29. Thase ME, McCrone P, Barrett MS, et al. Improving cost-effectiveness and access to cognitive behavior therapy for depression: providing remote-ready, computer-assisted psychotherapy in times of crisis and beyond. Psychother Psychosom 2020;89(5):307-313. [doi: 10.1159/000508143] [Medline: 32396917]
30. Mitchell LM, Joshi U, Patel V, Lu C, Naslund JA. Economic evaluations of internet-based psychological interventions for anxiety disorders and depression: a systematic review. J Affect Disord 2021 Apr 1;284:157-182. [doi: 10.1016/j.jad.2021.01.092] [Medline: 33601245]
31. Darden M, Espie CA, Carl JR, et al. Cost-effectiveness of digital cognitive behavioral therapy (Sleepio) for insomnia: a Markov simulation model in the United States. Sleep 2021 Apr 9;44(4):zsaa223. [doi: 10.1093/sleep/zsaa223] [Medline: 33151330]
32. Prentice RL, Aragaki AK, Chlebowski RT, et al. Randomized trial evaluation of the benefits and risks of menopausal hormone therapy among women 50-59 years of age. Am J Epidemiol 2021 Feb 1;190(3):365-375. [doi: 10.1093/aje/kwaa210] [Medline: 33025002]
33. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health—a systematic review and meta-analysis. NPJ Digit Med 2019;2(118):118. [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
34. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336. [doi: 10.1002/wps.20673] [Medline: 31496095]
35. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug;218:15-22. [doi: 10.1016/j.jad.2017.04.046]
36. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
37. Im EO, Lee Y, Chee E, Chee W. Web-based interventions for menopause: a systematic integrated literature review. Maturitas 2017 Jan;95:24-30. [doi: 10.1016/j.maturitas.2016.10.009] [Medline: 27889049]
38. Gkrozou F, Tsonis O, Godden M, Siafaka V, Paschopoulos M. Mobile health (mHealth) apps focused on menopause: are they any good? Post Reprod Health 2019 Dec;25(4):191-198. [doi: 10.1177/2053369119878978] [Medline: 31679433]
39. Atema V, van Leeuwen M, Kieffer JM, et al. Efficacy of internet-based cognitive behavioral therapy for treatment-induced menopausal symptoms in breast cancer survivors: results of a randomized controlled trial. J Clin Oncol 2019 Apr 1;37(10):809-822. [doi: 10.1200/JCO.18.00655] [Medline: 30763176]
40. Hey SP, Kimmelman J. The questionable use of unequal allocation in confirmatory trials. Neurology (ECronicon) 2014 Jan 7;82(1):77-79. [doi: 10.1212/01.wnl.0000438226.10353.1c] [Medline: 24306005]
41. The Menopause Society. URL: https://www.menopause.org/ [accessed 2024-02-03]
42. Hunter MS, Liao KL. A psychological analysis of menopausal hot flushes. Br J Clin Psychol 1995 Nov;34(4):589-599. [doi: 10.1111/j.2044-8260.1995.tb01493.x] [Medline: 8563666]
43. Hunter MS, Haqqani JR. An investigation of discordance between subjective and physiological measures of vasomotor symptoms. Climacteric 2011 Feb;14(1):146-151. [doi: 10.3109/13697131003735585]

XSL•FO
RenderX

44. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026]
45. Shin C, Lee SH, Han KM, Yoon HK, Han C. Comparison of the usefulness of the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of psychiatric outpatient data. Psychiatry Investig 2019 Apr;16(4):300-305. [doi: 10.30773/pi.2019.02.01] [Medline: 31042692]
46. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
47. Beaudreau SA, Spira AP, Stewart A, et al. Validation of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale in older Black and White women. Sleep Med 2012 Jan;13(1):36-42. [doi: 10.1016/j.sleep.2011.04.005] [Medline: 22033120]
48. Hinz A, Glaesmer H, Brähler E, et al. Sleep quality in the general population: psychometric properties of the Pittsburgh Sleep Quality Index, derived from a German community sample of 9284 people. Sleep Med 2017 Feb;30:57-63. [doi: 10.1016/j.sleep.2016.03.008] [Medline: 28215264]
49. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
50. Löwe B, Decker O, Müller S, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
51. Lund AM. Measuring Usability with the USE Questionnaire. Usability Interface 2001;8:3-6 [FREE Full text]
52. Duffecy J, Grekin R, Long JD, Mills JA, O'Hara M. Randomized controlled trial of sunnyside: individual versus group-based online interventions to prevent postpartum depression. J Affect Disord 2022 Aug 15;311:538-547. [doi: 10.1016/j.jad.2022.05.123] [Medline: 35654284]
53. Kai Y, Nagamatsu T, Kitabatake Y, Sensui H. Effects of stretching on menopausal and depressive symptoms in middle-aged women: a randomized controlled trial. Menopause 2016 Aug;23(8):827-832. [doi: 10.1097/GME.0000000000000651] [Medline: 27300113]
54. Wong C, Yip BHK, Gao T, et al. Mindfulness-based stress reduction (MBSR) or psychoeducation for the reduction of menopausal symptoms: a randomized, controlled clinical trial. Sci Rep 2018 Apr 26;8(1):6609. [doi: 10.1038/s41598-018-24945-4] [Medline: 29700350]
55. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
56. Cuijpers P, Smit F, van Straten A. Psychological treatments of subthreshold depression: a meta-analytic review. Acta Psychiatr Scand 2007 Jun;115(6):434-441. [doi: 10.1111/j.1600-0447.2007.00998.x] [Medline: 17498154]
57. Karyotaki E, Riper H, Twisk J, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
58. Hunter MS, Nuttall J, Fenlon D. A comparison of three outcome measures of the impact of vasomotor symptoms on women's lives. Climacteric 2019 Aug;22(4):419-423. [doi: 10.1080/13697137.2019.1580258] [Medline: 30905177]
59. Cohen LS, Soares CN, Vitonis AF, Otto MW, Harlow BL. Risk for new onset of depression during the menopausal transition. Arch Gen Psychiatry 2006 Apr 1;63(4):385. [doi: 10.1001/archpsyc.63.4.385]
60. Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioraltherapy for depression. JAMA Psychiatry 2021 Apr 1;78(4):361. [doi: 10.1001/jamapsychiatry.2020.4364]
61. Thurston RC, Chang Y, Buysse DJ, Hall MH, Matthews KA. Hot flashes and awakenings among midlife women. Sleep 2019 Sep 6;42(9):zsz131. [doi: 10.1093/sleep/zsz131] [Medline: 31152182]
62. Medic G, Wille M, Hemels ME. Short- and long-term health consequences of sleep disruption. Nat Sci Sleep 2017;9:151-161. [doi: 10.2147/NSS.S134864] [Medline: 28579842]
63. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
64. Gentili A, Failla G, Melnyk A, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022;10(787135):787135. [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
65. McCrae CS, Bramoweth AD, Williams J, Roth A, Mosti C. Impact of brief cognitive behavioral treatment for insomnia on health care utilization and costs. J Clin Sleep Med 2014 Feb 15;10(2):127-135. [doi: 10.5664/jcsm.3436] [Medline: 24532995]
66. Kraepelien M, Mattsson S, Hedman-Lagerlöf E, et al. Cost-effectiveness of internet-based cognitive–behavioural therapy and physical exercise for depression. BJPsych open 2018 Jul;4(4):265-273. [doi: 10.1192/bjo.2018.38]
67. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc 2019 Jul;51(7):1538-1557. [doi: 10.1249/MSS.0000000000001947] [Medline: 30789439]
68. Abd-Alrazaq A, AlSaad R, Aziz S, et al. Wearable artificial intelligence for anxiety and depression: scoping review. J Med Internet Res 2023 Jan 19;25:e42672. [doi: 10.2196/42672] [Medline: 36656625]
69. Menopause. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/menopause [accessed 2024-02-03]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
DCP: digital care program
GAD-7: Generalized Anxiety Disorder-7
HFRS: Hot Flush Rating Scale
PHQ-8: Patient Health Questionnaire Depression Scale-8
PSQI: Pittsburgh Sleep Quality Index
RCT: randomized controlled trial
USE: Usefulness, Satisfaction, and Ease Questionnaire
VMS: vasomotor symptoms

Edited by L Buis, TDA Cardoso; submitted 09.03.24; peer-reviewed by CJ Gibson, L Yang, R Meng; revised version received 11.04.25; accepted 14.04.25; published 17.06.25.

Please cite as:
Duffecy J, Rehman A, Gorman S, Huang YL, Klumpp H
Evaluating a Mobile Digital Therapeutic for Vasomotor and Behavioral Health Symptoms Among Women in Midlife: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e58204
URL: https://mhealth.jmir.org/2025/1/e58204
doi:10.2196/58204

© Jennifer Duffecy, Arfa Rehman, Scott Gorman, Yong Lin Huang, Heide Klumpp. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

An App-Based Cognitive Behavioral Therapy Program Tailored for College Students: Randomized Controlled Trial

Min Hee Kim[1], MA; Yeon-kug Moon[2], PhD; Kyong-Mee Chung[1], PhD

[1]Department of Psychology, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

[2]Department of Artificial Intelligence and Data Science, Sejong University, Seoul, Republic of Korea

Corresponding Author:
Kyong-Mee Chung, PhD
Department of Psychology, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

Abstract

Background: Technology-based cognitive behavioral therapy programs are accessible interventions to address mental health challenges, particularly among college students. Despite their potential, low adherence rates remain a key challenge.

Objective: This study aims to assess the effects of the tailored content and gamification elements of the Mind Booster Green program on reducing depressive symptoms and improving college life adjustment.

Methods: A randomized, unblinded controlled trial was conducted among college students. All study procedures were conducted remotely using web-based tools. The participants were randomly assigned to the intervention or waitlist control groups. The intervention group used Mind Booster Green, an app-based self-help cognitive behavioral therapy program for 1 month. The program included tailored case stories and gamification elements, such as point and level systems, to enhance user engagement and adherence. Outcomes were self-assessed using web-based questionnaires and included changes in depressive symptoms, college life adjustment, and negative and positive automatic thoughts. The usability of the app was also evaluated. Outcomes were measured at 3 time points: preintervention, postintervention, and at a 2-month follow-up (F/U), using validated and standardized tools.

Results: A total of 170 individuals (mean age 22.60, SD 3.37 years; 136/170, 80% female) participated in the study. A chi-square analysis revealed no significant differences between the two groups at baseline in terms of age, sex, or class year (P>.05). A generalized estimating equation analysis revealed significant time×group interactions for all variables. Compared to the control group, the intervention group showed greater improvements across all outcomes, with between-group effect sizes ranging from -0.78 to 0.49. For derpessive symptoms, large within-group effect size were observed (Patient Health Questionnaire-9: pre to post, Cohen d=1.12; pre to F/U, Cohen d=1.15; Beck Depression Inventory-II: pre to post, Cohen d=0.90; pre to F/U, Cohen d=1.04). Large within-group effect size was also found for adjustment to college life (Student Adaptation to College Questionnaire–Revised: pre to post, Cohen d=–0.87; pre to F/U, Cohen d=-0.85), and moderate effect for negative automatic thoughts (Automatic Thought Questionnaire–Negative, Short Form: pre to post, Cohen d=0.36; pre to F/U, Cohen d=0.58) and positive automatic thoughts (Automatic Thought Questionnaire—Positive, Short Form: pre to post, Cohen d=–0.45; pre to F/U, Cohen d=-0.44). Adherence rates were 89% and 99% for the intervention and control groups, respectively. The usability test results, assessed using the Mobile App Rating Scale, showed an overall score of 3.88, with scores above the medium level in the engagement, functionality, aesthetics, and information quality categories.

Conclusions: Mind Booster Green demonstrated substantial potential as a complementary interventio to traditional psychological services for college students, providing a cost-effective and scalable solution for mental health issues. Future research should explore the applicability of this program in diverse populations.

Trial Registration: Korean Clinical Trial Registry (Clinical Research Information Service) KCT0009758 (retrospectively registered); https://cris.nih.go.kr/cris/search/detailSearch.do?seq=27868&search_page=L

(JMIR Mhealth Uhealth 2025;13:e50006) doi:10.2196/50006

KEYWORDS

t-CBT; college students; depressive symptoms; randomized controlled trial; RCT; app-based; cognitive behavioral therapy; CBT; self-help; technology-based cognitive behavioral therapy; mobile phone

XSL•FO
RenderX

Introduction

Cognitive behavioral therapy (CBT) is an evidence-based treatment with substantial evidence supporting its effectiveness in alleviating depressive symptoms [1]. However, delivery constraints, such as requiring 12 - 16 weeks of face-to-face treatment by a professional, have hindered its widespread dissemination [2]. Recent advances in digital technology, including smartphones, have led to the development of technology-based CBT (t-CBT) programs that use digital tools. Several studies have demonstrated the effectiveness of these programs in reducing depressive symptoms [3-6]. With its advantages, such as high accessibility, low cost, and minimal stigmatization, t-CBT is gaining attention as an innovative treatment delivery method that can supplement or even provide an alternative to traditional face-to-face CBT [7].

However, participation in t-CBT interventions relies on users' willingness and motivation, resulting in the need to address issues such as low engagement and high dropout rates. For example, meta-analyses have reported user adherence rates of only 17% - 26% [8,9], and usage surveys of real-world open-access unguided programs have reported much lower rates of approximately 4% [10]. The degree of adherence to an intervention program can be linked to its effectiveness [11]; therefore, the need to consider and research strategies to improve usability, user engagement, and adherence is increasing [12].

Tailoring has recently been proposed as a strategy for enhancing user engagement and adherence. This strategy involves personalizing an intervention's content and layout based on users' demographic and psychological characteristics and use context [13]. Tailoring is particularly suitable for diagnostic groups with heterogeneous symptoms or demographic factors, such as individuals with depression, and is achieved by clearly defining the treatment target group based on specific factors (eg, age, occupation, socioeconomic status, and ethnicity) and developing appropriate treatment strategies accordingly [14,15]. Tailored content can provide personally relevant information directly to users, thereby increasing the initial uptake and use of an app and facilitating changes in user attitudes or behaviors [16]. In the context of t-CBT, tailoring strategies can be approached in several ways, including (1) providing content that reflects the characteristics and experiences of the target population [17-20], (2) developing personalized treatment plans according to individuals' diagnoses or symptoms [21,22], and (3) enabling users to customize features such as menus, tools, and notifications [23]. Tailored mental health intervention programs have been generally reported to positively affect user adherence, thereby enhancing program effectiveness and user satisfaction [16,24,25].

Among the aspects of tailoring, providing content that reflects the characteristics and experiences of the target population is suggested to facilitate users' acquisition of emotional, cognitive, and behavioral skills without expert assistance [26,27]. In practice, t-CBT uses targeted audience-specific examples, language, and various multimedia formats (eg, text, video, and pictures) to deliver psychoeducational content, providing the necessary information and modeling of skills they need to learn to be delivered to users more effectively [24]. For example, Strecher et al [20] found that providing tailored fictional success stories reflecting user demographics such as sex, age, race, marital status, and smoking-related factors increased long-term engagement in a smoking cessation program, which was linked to program effectiveness.

Gamification, which involves applying game elements to nongame digital platforms (such as web- or app-based systems) [28], is another strategy for encouraging user engagement and adherence. Gamification has been reported to increase user motivation and engagement, make interactions with app-based programs enjoyable, and positively influence behavioral changes related to physical health and psychological well-being [29,30]. In technology-based mental health interventions, gamification is suggested to play a role in increasing app use and user engagement and supporting the delivery of the intervention's active ingredients, thereby enhancing therapeutic outcomes [31]. In a previous meta-analysis, Cheng and Ebrahimi [32] showed that gamified interventions are more effective than nongamified ones in reducing depressive and anxiety symptoms. In the context of t-CBT for mental health, gamification has been implemented using avatars [33,34], storytelling, and reward systems [35,36]. Specifically, several self-management program studies have shown that positive reinforcement such as points, badges, digital rewards, and tangible prizes have been shown to effectively promote health-related target behaviors [37]. These strategies must be considered for active applications in t-CBT to encourage users to log in regularly, use the app, and complete in-app activities [38].

Although t-CBT has primarily been developed and deployed in populations with specific psychiatric conditions, interest in its application in younger populations, particularly college students, is growing. Depression is a common mental health concern among college students [39,40] and can interfere with academic retention [41], career development [42], and social relationships [43], ultimately affecting their overall quality of life [44]. The increasing prevalence of depression among college students has become a social concern for clinicians and mental health professionals [45,46]. Institutional efforts, such as student counseling services, aim to address students' mental health needs; however, actual usage rates are low because of long wait times and concerns about social stigma [47-50]. In this context, t-CBT may serve as a useful means of addressing such difficulties. Given college students' high smartphone penetration and favorable attitudes toward app-based programs [51,52], t-CBT holds significant potential for improving access to mental health care and addressing the unmet needs of this population.

Building on the existing research, two strategies can be applied to t-CBT programs aimed at reducing depressive symptoms among college students. First, programs should be tailored to the cognitive, behavioral, and emotional characteristics of depressive symptoms experienced by college students, as well as stressors related to college life. Second, positive reinforcement strategies should be incorporated through gamification to encourage college students to use t-CBT programs voluntarily. This is based on research findings that younger digital natives [53], such as current college students, are more likely to find gamified products useful and enjoyable

XSL•FO
RenderX

[54]; moreover, achievement-related gamification features, including badges and upgrades, have a more positive impact on this group than on older groups [55].

Despite its potential, relatively few studies have focused on developing and validating t-CBT programs that incorporate tailoring or gamification to alleviate depression among college students. While several studies have adapted interventions originally designed for the general adult populations [56-58], such efforts often lack sufficient consideration for the distinctive psychological and contextual needs of college students. Nevertheless, some studies have indicated the potential of tailored t-CBT programs for this population. Mullin et al [18] demonstrated significant reductions in depressive and anxiety symptoms among college students using a tailored therapist-guided CBT program (UniWellbeing). Similarly, Lattie et al [59] reported increased mental health literacy and high use rates among students who used a tailored, unguided CBT program (IntelliCare). Recent evidence from an Asian context supports the effectiveness of internet-based CBT programs tailored for local college students, further suggesting their broader applicability [60]. In Korea, some CBT programs originally developed for adults have been evaluated among college students [61]; however, such studies remain limited. Furthermore, studies incorporating gamification elements that are explicitly designed to enhance voluntary engagement and sustained use are scarce, indicating a notable gap and underscoring the need for additional research.

To address this gap, this study developed and evaluated the effectiveness of Mind Booster Green, a tailored, app-based self-help CBT program for reducing depressive symptoms among college students. Changes in depressive symptoms and adaptation to college life were assessed to verify the effectiveness of the program. Additionally, changes in negative and positive automatic thoughts were examined to verify the applicability of CBT mechanisms to students. App usability was evaluated using standardized measures of engagement, functionality, aesthetics, and information quality.

Methods

Recruitment

Participants were recruited through advertisements on social media (eg, Facebook and Naver) and web-based communities popular among college students and young adults (eg, Everytime and job search cafés). Individuals had to meet the following inclusion criteria: (1) undergraduate or graduate students; (2) native Korean speakers with no difficulty using smartphones; (3) used a smartphone with iOS 12 or higher or Android 8 or higher; (4) scored 5 or higher on the Patient Health Questionnaire-9 (PHQ-9), indicating mild depression, and less than 1 on the suicide item, indicating a low risk of suicide; and (5) were not currently receiving other mental health services (eg, medication, psychotherapy, or counseling). The participants received an explanation of the study through a video posted on the study website and signed an electronic consent form. Additionally, under institutional review board (IRB) approval, they submitted a copy of their ID to verify their identity.

Study Setting

The study was conducted remotely using web-based tools in South Korea from June 2021 to March 2022. All study procedures, including recruitment, screening, consent, intervention delivery, and outcome assessments, were conducted remotely owing to the impact of COVID-19. Data were collected through Qualtrics (Qualtrics), and survey links were delivered individually via a dedicated social networking service (SNS) channel. All participant communications, including survey timing notifications, survey links, reminders, inquiries, and risk issues, were managed through this channel. All procedures were performed by trained researchers as listed in the IRB-approved protocol.

Measures

PHQ-9

The Korean version of the PHQ-9, developed by Kroenke et al [62] and validated by Park et al [63], was used to screen participants and measure changes in depression severity. The PHQ-9 consists of 9 items that assess depressive symptoms experienced over the past 2 weeks, rated on a 4-point Likert scale from 0=not at all to 3=nearly every day. It was administered during the screening to identify participants with mild or higher levels of depression (PHQ-9≥5) and no moderate or higher suicidal ideation (PHQ-9 item 9≤1). Given the time gap between screening and baseline assessments, the PHQ-9 was also included in the pre, post, and follow-up assessments. The internal consistency (Cronbach α) in the Korean standardization study [64] was 0.95, and in this study, it was 0.79.

Korean Beck Depression Inventory-II

The Korean version of the Beck Depression Inventory-II (BDI-II), developed by Beck et al [65] and validated by Kim et al [66], was used to assess changes in depression severity. The BDI-II is a 21-item standardized questionnaire that is widely used to evaluate depressive symptoms. Participants rated their level of difficulty with depressive symptoms such as sadness, lethargy, feelings of worthlessness, and changes in sleep and appetite over the past 2 weeks on a 4-point Likert scale ranging from 0=not at all to 3=severely. Regarding internal consistency, Cronbach α was 0.83 in the Korean standardization study [67] and 0.89 in this study.

Automatic Thought Questionnaire—Negative/Positive, Short Form

The Automatic Thought Questionnaire—Negative, Short Form (ATQN-SF) and the Automatic Thought Questionnaire—Positive, Short Form (ATQP-SF) were used to assess participants' patterns of automatic thoughts—both negative and positive. The ATQN-SF, developed by Hollon and Kendall [68], adapted into Korean by Kwon and Yoon [69], and further refined by Heo and Kim [70], comprises 9 items rated on a 5-point Likert scale (0=not at all to 4=always). It measures negative automatic thoughts associated with depression and assesses changes in thought patterns in alignment with CBT mechanisms. The ATQP-SF, a 10-item scale adapted and validated in Korean by Heo and Kim [71] from the original

XSL•FO
RenderX

Automatic Thought Questionnaire-Revised by Kendall et al [72], evaluates the frequency of positive automatic thoughts, which are often lacking in individuals vulnerable to depression. Both subscales demonstrated excellent internal consistency, with Cronbach α values of 0.96 (ATQN-SF) and 0.93 (ATQP-SF) reported in previous studies and 0.89 (ATQN-SF) and 0.91 (ATQP-SF) in this study.

Student Adaptation to College Questionnaire—Revised

The Student Adaptation to College Questionnaire—Revised (SACQ-R) was used to measure participants' adaptation to college life. Developed by Baker and Siryk [73] and revised and adapted into Korean by Lee [74], the SACQ-R assesses academic, social, emotional, and physical adjustment, as well as attachment to the institution. The 25-item measure is rated on a 5-point Likert scale ranging from 1=not at all to 5=very much. Regarding internal consistency, Cronbach α was 0.85 in previous research [74] and in this study.

Mobile App Rating Scale

To evaluate the usability of the app, the Mobile App Rating Scale (MARS) developed by Stoyanov et al [75] was translated into Korean and used according to the guidelines of the International Test Commission [76]. The scale comprises 33 items across five factors: engagement, functionality, aesthetics, information quality, and app-specific items (subjective satisfaction). Participants rated each item on a 5-point Likert scale ranging from 1=not at all to 5=very much. Additionally, an open-ended question was included to gather unstructured subjective feedback: "What did you like about using the app?" Usability was assessed postintervention in the intervention group (n=80).

Design and Procedure

Overview

All study procedures were conducted in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines. This was a two-arm, parallel-group, randomized controlled trial with a waitlist control (WLC) design. An a priori power analysis was conducted using a linear multiple regression model (*F* test) in G*Power (Heinrich Heine University Düsseldorf), as no standardized procedure is available for power analysis in generalized estimating equation (GEE). The analysis indicated that a minimum sample size of 42 participants was required to detect a medium effect size (F_2=0.25) with *P*=.05 and power=0.80, assuming two predictors (df1=2, df2=39), based on prior t-CBT efficacy studies.

Participants who provided informed consent were randomly assigned to the intervention or WLC group. Randomization was conducted using a pregenerated random number table created in Excel (Microsoft Corp) by a research assistant registered in the IRB-approved protocol. The assistant used the =RAND() function to generate random values, which were sorted to assign the participants to groups. As this was an open-label trial, both the participants and investigators were aware of the group assignments.

Participants who met the inclusion criteria underwent the following steps: (1) preintervention assessment, (2) intervention or waiting period depending on group assignment, (3) postintervention assessment, and (4) follow-up assessment. The preintervention assessment was conducted immediately after participants provided informed consent. The postintervention assessment was conducted after the intervention group completed the 28-session Mind Booster Green training and 30 days postbaseline for the WLC group. A follow-up assessment was conducted 2 months after the postintervention assessment. Both groups completed the PHQ-9, BDI-II, SACQ-R, ATQN-SF, and ATQP-SF at all the assessment points. Additionally, the MARS was administered to the intervention group during the postintervention assessment for usability evaluation.

After completing the 2-month follow-up assessment, participants in the WLC group were granted access to the app. App access was provided only after all required assessments were completed. No symptom data or app use records were collected from the WLC group following the intervention, and postintervention data were not included in the outcome analyses.

Intervention Condition: Mind Booster Green Group

Participants assigned to the intervention group downloaded and installed the app from Apple's app store (iOS) or the Google Play store (Android) via a URL provided by the researchers. Upon registration, the researchers approved the use of the app on the admin page. The participants began training independently without additional guidance or support and were advised to complete one session per day, with the researchers monitoring their progress via the admin page. Participants who did not log in or complete sessions for more than 3 days received prompt SMS text messages, and those who did not engage for more than 10 days were considered withdrawn. Other than the SMS text message prompts to resume the use of the app, participants received no feedback on their performance from the researchers, who were only available to answer questions about the study via an SNS channel.

Mind Booster Green

Mind Booster Green is a self-help CBT app adapted from "HaruToday," a CBT app initially developed for patients with cancer [17,77] with the support of the Korean Ministry of Health and Welfare (HA16C0021). HaruToday has three modules (depression or anxiety, sleep, and pain issues) and has been validated for its effectiveness on depression or anxiety [17] and sleep [77]. The Mind Booster Green program was modified to address depression in undergraduate and graduate students by reducing the number of sessions from 48 to 28 based on user feedback. Tailored CBT content was developed through focus group interviews with undergraduate and graduate students to gather specific case stories related to their experiences with depression. A pilot test involving CBT experts and college students was conducted, and the final version was completed through revision and supplementation. The program used in this study was not updated during the experimental period [78].

The main training program comprises 5 zones with 28 sessions targeting effective CBT techniques for managing depressive

XSL•FO
RenderX

symptoms: (1) psychoeducation, (2) cognitive restructuring, (3) behavioral activation, (4) relaxation training, and (5) problem-solving. Each zone includes 4 - 10 sessions, each lasting approximately 10 - 15 minutes and consisting of 50 - 70 illustrated slides with pages where users can input answers, complete quizzes, or fill out worksheets. Audio narration is available for each slide and can be customized by the user.

Six characters representing college students are featured in the story and the text and illustrations detail how they work with experts to explore their depressive symptoms, learn cognitive-behavioral skills to manage their symptoms, and overcome challenges (Figure 1). Users also engage in self-guided tasks such as quizzes and activity planning to apply learned skills. Table 1 summarizes the specific characteristics of each zone. The app includes mood-monitoring features, allowing users to track their mood visually and providing thought records and relaxation audio or video to apply the learned techniques in real life.

Figure 1. Example screen from Mind Booster Green.

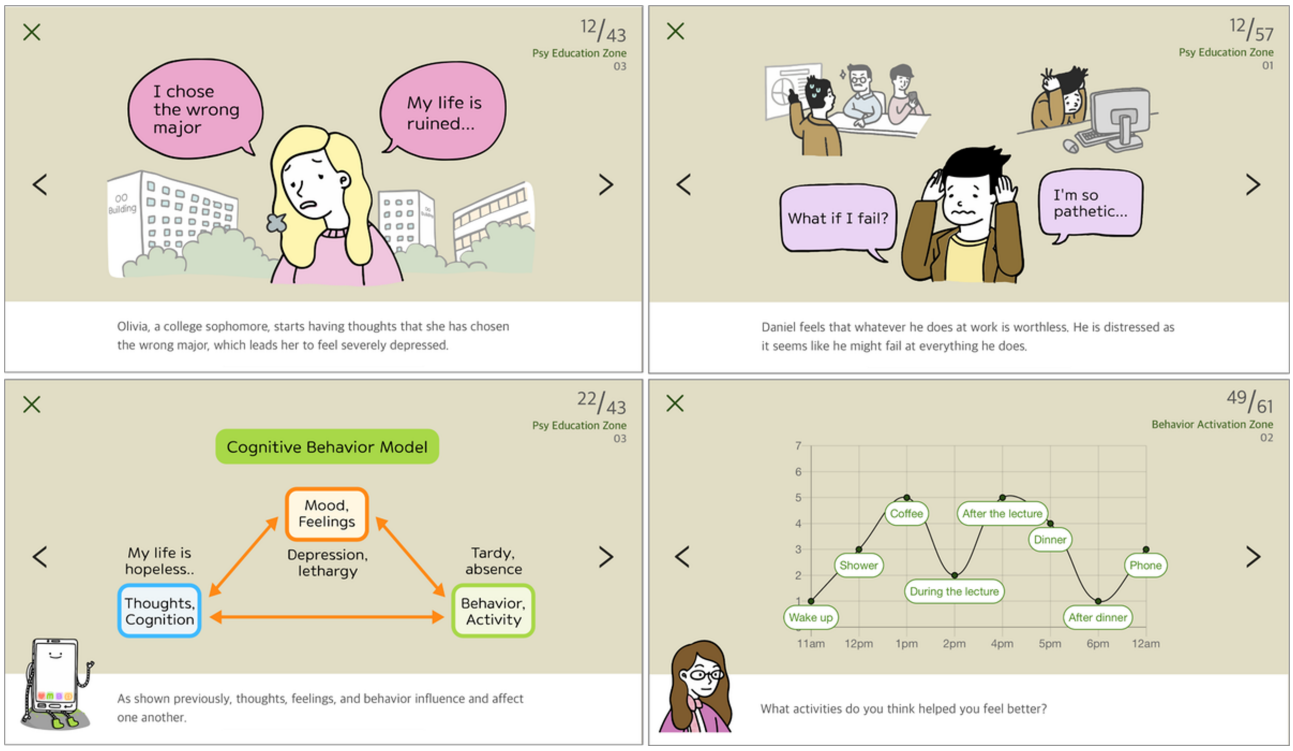


Table . Session outline of Mind Booster Green.

Training zone	Number of sessions	Content
Psychoeducation	4	• Characteristics of depressive symptoms • Core principles of CBT[a] • Self-monitoring techniques
Behavioral activation	5	• Pleasant activities exploration • Pleasant activities planning
Cognitive restructuring	10	• Automatic thoughts and cognitive distortions • Cognitive restructuring techniques • Thought record
Problem-solving	5	• Five-step problem-solving approach • Time management • Assertiveness skills • Anger management strategies
Relaxation	4	• Progressive muscle relaxation • Guided imagery techniques

[a]CBT: cognitive behavioral therapy.

XSL•FO
RenderX

Mind Booster Green uses point-and-level systems based on positive reinforcement. Participants earned points for attendance and completion of mood checks, sessions, training zones, and thought records. Accumulating points led to leveling, with 5 levels requiring progressively more points. As the participants leveled up, the main character's potted plant grew, providing encouragement and emotional support.

WLC Condition

Participants in the WLC group were instructed to maintain their usual activities for 30 days without additional guidance or support from the researchers. Similar to the intervention group, they were able to contact the research team through the SNS channel for study-related inquiries.

Statistical Analysis

The analyses included data from 91 participants in the intervention group, 79 participants in the control group for the intention-to-treat (ITT) analysis, and 78 participants in both groups for the per-protocol (PP) analysis. One-way ANOVAs and chi-square analyses were conducted to examine demographic characteristics and ensure group homogeneity in the preassessment. To evaluate the effectiveness of the intervention, an ITT analysis [79] was performed using the entire dataset, including dropouts, and a PP analysis was conducted with participants who completed all procedures. A GEE was used to test for group differences in changes over time [80]. As all dependent variables were nonnormally distributed, the GEE analysis used an unstructured working correlation structure and gamma distribution with a log-link response scale. Time and group were the predictor variables, and the dependent variables included the PHQ-9, BDI-II, SACQ-R, ATQN-SF, and ATQP-SF, with preassessment scores as covariates. Statistical analyses were performed using SPSS (version 27; IBM Corp). Within- and between-group effect sizes (Cohen *d*) were calculated using absolute scores at each time point.

Ethical Considerations

This study was reviewed and approved by the IRB of Yonsei University (IRB 7001988 - 202202HR-1127-10) in accordance with the institutional and national ethical guidelines for research involving human participants. All participants were fully informed about the purpose, procedures, and voluntary nature of the study through a video explanation posted on the study website and provided written informed consent by electronically signing a digital consent form prior to participation. They were also informed of their right to withdraw from the study at any time, without penalty. All collected data were deidentified to ensure the participants' privacy and confidentiality. No personally identifiable information was stored, and all data were securely managed on encrypted servers with access limited to authorized research staff. Participants received monetary compensation equivalent to 50,000 KRW (approximately US $40) after completing the follow-up assessment, with prorated compensation for partial participation. Participants in the WLC group were granted access to the app after the follow-up assessment, along with compensation. These measures were implemented to ensure fairness and acknowledge the participants' time and contributions to the study.

Trial Registration

This study was approved by the IRB prior to its initiation. However, the clinical trial was retrospectively registered in the Korean Clinical Trial Registry (Clinical Research Information Service) under registration number KCT0009758. At the time of the study, the necessity of prospective clinical trial registration was not fully recognized, particularly because trial registration is not a legal requirement in Korea. However, we promptly addressed this oversight and retrospectively completed the registration. The research procedures reported in the registration are fully consistent with the protocol approved by the IRB. It was ensured that all research procedures adhered to the ethical standards approved by the IRB and that retrospective registration did not impact the integrity or outcomes of the study.

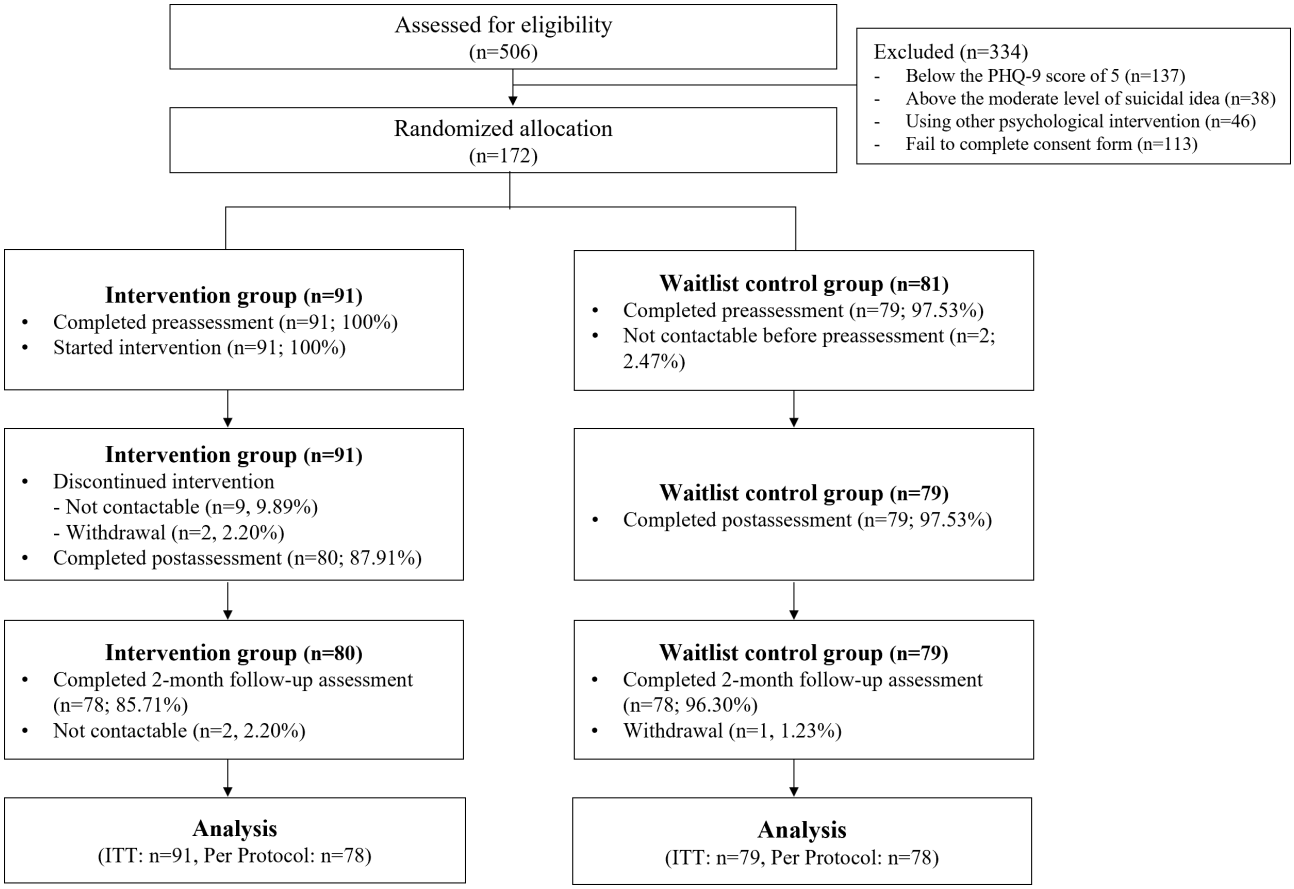
Results

Participant Flow and Sample Description

Of the 506 undergraduate and graduate students who volunteered to participate, 172 met the inclusion criteria and were enrolled. A pregenerated random number table in Excel was used to randomly assign participants to the intervention (n=91) or WLC (n=81) group. Before the baseline assessment, 2 participants in the WLC group ceased contact and were excluded from the study. During the intervention period, 11 participants were excluded owing to loss of contact (intervention group, n=9) or voluntary withdrawal (intervention group, n=2). Between the postintervention and follow-up assessments, contact with 2 participants from the intervention group was lost, and 1 participant from the WLC group was excluded because they used other psychological services. Consequently, 156 participants (78 in each group) completed all three assessments. For the ITT analysis, data from 91 and 79 participants in the intervention and control groups, respectively, were included. For the PP analysis, data from 78 participants in both groups were included. Figure 2 presents a detailed flowchart of the participant selection process.

XSL•FO
RenderX

Figure 2. CONSORT diagram of participant flow. CONSORT: Consolidated Standards of Reporting Trials; ITT: intention-to-treat; PHQ-9: Patient Health Questionnaire-9.



The mean age of the study participants (n=170) was 22.60 (SD 3.37, median 22, IQR 18 - 38) years, with 80% (136/170) female and 20% (34/170) male participants. Of the total participants, 17.1% (29/170) were freshmen, 17.6% (30/170) sophomores, 24.1% (41/170) juniors, 22.9% (39/170) seniors or above, and 18.2% (31/170) graduate students. Table 2 shows the demographic characteristics of participants in both groups. The chi-square analysis showed no differences between the intervention and WLC groups in terms of age, sex, or class year (Table 2; all P>.05).

Table . Participant characteristics by group.

Characteristics	Intervention (n=91)	Control (n=79)	Chi-square (*df*)	*P* value
Age (years), mean (SD)	22.31 (3.34)	22.91 (3.41)	1.31[a] (1, 168)	.25
Sex, n (%)			0.0 (1)	.94
Female	73 (80.2)	63 (79.7)		
Male	18 (19.8)	16 (20.3)		
Year, n (%)			3.3 (4)	.51
Freshman	14 (15.4)	15 (19.0)		
Sophomore	20 (22.0)	10 (12.7)		
Junior	19 (20.9)	22 (27.8)		
Senior level or above	21 (23.1)	18 (22.8)		
Graduate student	17 (18.7)	14 (17.7)		

[a]For age, *F* test value is provided.

Table 3 presents the observed means and standard deviations of the PHQ-9, BDI-II, SACQ-R, ATQP-SF, and ATQN-SF scores by group at preassessment and the results of the between-group analyses of variance. No significant differences were observed between the groups for any variable at preassessment (P>.05).

XSL•FO
RenderX

Table . Observed means and SDs for scores of each variable by group.

Measures and time	Intervention (Pre n=91; Post n=80; F/U[a] n=78), mean (SD)	Control (Pre n=79; Post n=79; F/U n=78), mean (SD)	*F* test (*df*)	*P* value
PHQ-9[b]			0.98 (1, 168)	.40
Pre	10.75 (4.12)	10.13 (4.00)		
Post	6.09 (4.19)	8.51 (4.57)		
F/U	5.81 (4.43)	7.46 (4.68)		
BDI-II[c]			2.55 (1, 168)	.14
Pre	22.59 (9.51)	20.23 (9.01)		
Post	13.38 (10.80)	17.90 (9.73)		
F/U	12.26 (10.65)	16.21 (10.62)		
SACQ-R[d]			0.84 (1, 168)	.33
	68.13 (12.40)	69.91 (12.97)		
Pre				
Post	79.16 (13.05)	71.76 (16.98)		
F/U	79.38 (14.17)	73.27 (16.83)		
ATQP-SF[e]			1.40 (1, 168)	.40
Pre	12.67 (6.23)	13.85 (6.76)		
Post	15.85 (7.83)	14.28 (7.56)		
F/U	16.18 (9.52)	14.71 (9.36)		
ATQN-SF[f]			2.49 (1, 168)	.10
Pre	11.51 (7.77)	9.75 (6.60)		
Post	8.66 (7.89)	9.75 (8.48)		
F/U	7.24 (6.99)	9.42 (7.98)		

[a]F/U: follow-up.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]BDI-II: Beck Depression Inventory-II.

[d]SACQ-R: Student Adaptation to College Questionnaire—Revised.

[e]ATQP-SF: Automatic Thought Questionnaire—Positive, Short Form.

[f]ATQN-SF: Automatic Thought Questionnaire—Negative, Short Form.

Effectiveness of Mind Booster Green

ITT Analysis

The estimated marginal means and SEs of the PHQ-9, BDI-II, SACQ-R, ATQP-SF, and ATQN-SF scores at the pre, post, and follow-up assessments are presented in Table 4, and the trends in score changes across time points are illustrated in Figure 3. The GEE analysis revealed a significant time×group interaction for all variables, and the main effects of group and time were significant for all 5 variables. These findings indicated significant differences between the intervention and control groups in terms of changes in depressive symptoms, adjustment to college life, and positive and negative automatic thoughts over time. The paired comparisons showed that the intervention group had lower scores for depressive symptoms (PHQ-9 and BDI-II) and negative automatic thoughts (ATQN-SF) and higher scores for adjustment to college life (SACQ-R) and positive automatic thoughts (ATQP-SF) after the intervention than before the use of Mind Booster Green. An examination of the effect size of the change from pre- to postassessment and follow-up assessment scores in the intervention group showed a large decrease in depressive symptom scores after using the app (pre to post Cohen *d*: PHQ-9 1.12, BDI-II 0.90), followed by a small but sustained decrease in scores up to 2 months after the postassessment (pre to F/U Cohen *d*: PHQ-9 1.15, BDI-II 1.04). Negative automatic thoughts followed a pattern similar to that of depressive symptoms; however, the magnitude of change was smaller (pre to post Cohen *d*=0.36; pre to F/U Cohen *d*=0.58). Although scores for adjustment to college life and positive automatic thoughts increased at postassessment compared with preassessment (pre to post Cohen *d*: SACQR –0.87, ATQP-SF –0.45), little difference was found in these scores between the post and follow-up assessments (pre to F/U Cohen *d*: SACQ-R –0.85, ATQP-SF –0.44). Table 5 presents the effect sizes of the mean score changes at each time point.

XSL•FO
RenderX

Table . Results of the GEE[a] analysis with covariate adjustment for the preassessment scores of each variable[b].

Measures	Intention-to-treat		Per protocol	
	Wald chi-square (*df*)	*P* value	Wald chi-square (*df*)	*P* value
PHQ-9[c]				
Group×Time	28.7 (2)	<.001	29.5 (2)	<.001
Group	17.0 (1)	<.001	18.1 (1)	<.001
Time	118.1 (2)	<.001	118.0 (2)	<.001
BDI-II[d]				
Group×Time	28.5 (2)	<.001	35.8 (2)	<.001
Group	23.1 (1)	<.001	30.6 (1)	<.001
Time	81.8 (2)	<.001	96.0 (2)	<.001
SACQ-R[e]				
Group×Time	30.9 (2)	<.001	32.0 (2)	<.001
Group	26.2 (1)	<.001	27.5 (1)	<.001
Time	93.3 (2)	<.001	94.9 (2)	<.001
ATQP-SF[f]				
Group×Time	7.6 (2)	.02	7.8 (2)	.02
Group	6.0 (1)	.02	5.9 (1)	.02
Time	35.9 (2)	<.001	35.8 (2)	<.001
ATQN-SF[g]				
Group×Time	9.3 (2)	.01	9.0 (2)	.01
Group	10.3 (1)	.001	10.0 (1)	.002
Time	2.8 (2)	.24	2.8 (2)	.25

[a]GEE: generalized estimating equation.

[b]Intervention pre n=91, post n=80, F/U n=78; Control pre n=79, post n=79, F/U n=78.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]BDI-II: Beck Depression Inventory-II.

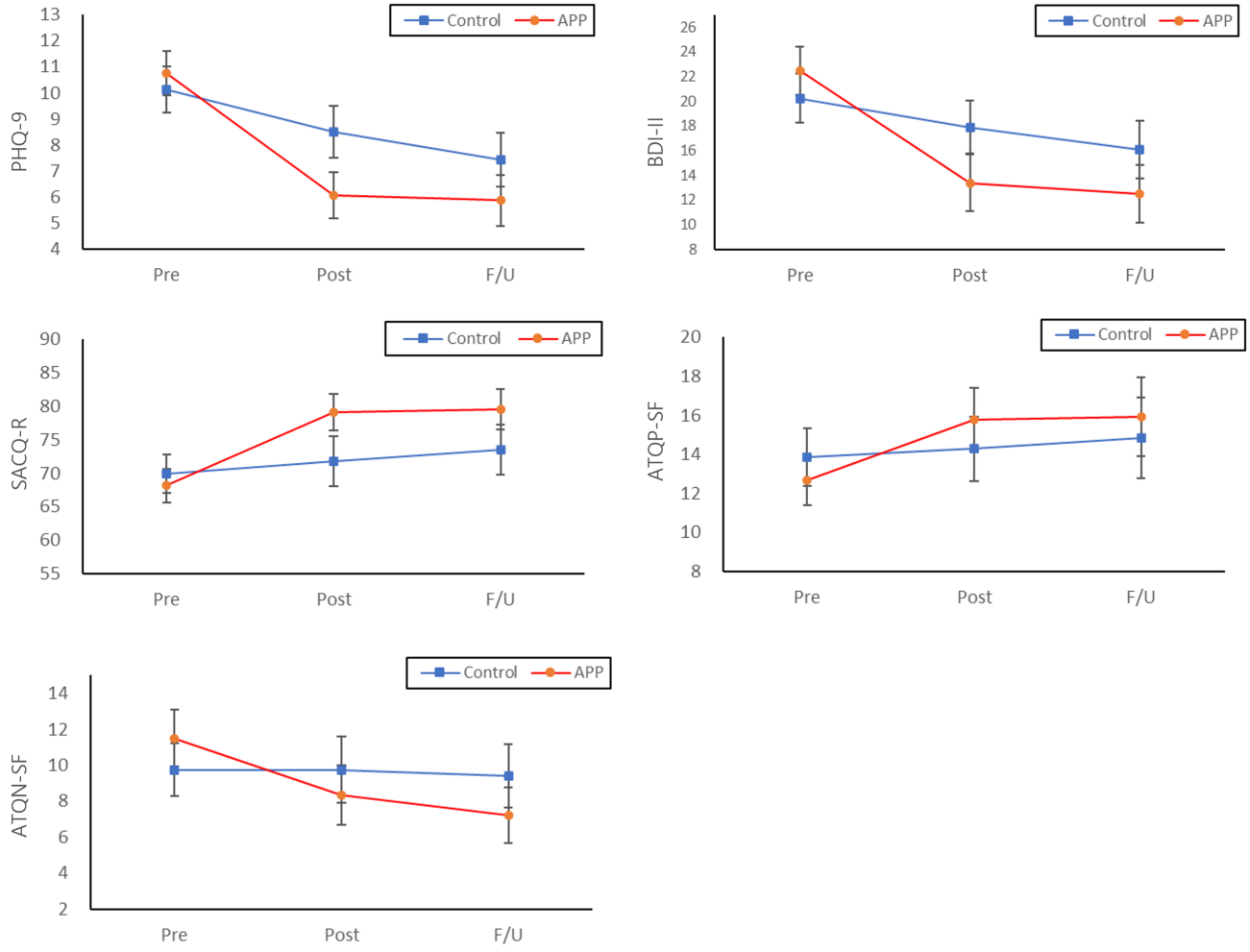
[e]SACQ-R: Student Adaptation to College Questionnaire—Revised.

[f]ATQP-SF: Automatic Thought Questionnaire—Positive, Short Form.

[g]ATQN-SF: Automatic Thought Questionnaire—Negative, Short Form.

XSL•FO
RenderX

Figure 3. Means and 95% intervals for the groups across the assessment time points. APP: ____; ATQN-SF: Automatic Thought Questionnaire—Negative, Short Form; ATQP-SF: Automatic Thought Questionnaire—Positive, Short Form; BDI-II: Beck Depression Inventory-II; PHQ-9: Patient Health Questionnaire-9; SACQ-R: Student Adaptation to College Questionnaire—Revised.



XSL•FO
RenderX

Table . Estimated marginal means and effect size.

	Preintervention (n=91), Control (n=79)		Postintervention (n=80), Control (n=79)		F/U[a] intervention (n=78), Control (n=78)		Within-group effect size from prescores				Between-group effect size			
							Post		F/U		Post		F/U	
	Mean (SE[b])	95% CI[c]	Mean (SE)	95% CI	Mean (SE)	95% CI	Cohen *d*	95% CI	Cohen *d*	95% CI	Cohen *d*	95% CI	Cohen *d*	95% CI
PHQ-9[d]											–0.78	–1.22 to –0.34	–0.36	–0.68 to –0.05
Intervention	9.84 (0.08)	9.68 to 10.00	5.68 (0.32)	5.09 to 6.34	5.96 (0.49)	5.08 to 7.00	1.12	0.82 to 1.42	1.15	0.85 to 1.46				
Control	9.71 (0.10)	9.51 to 9.91	8.32 (0.42)	7.53 to 9.20	7.38 (0.41)	6.61 to 8.23	0.38	0.07 to 0.69	0.61	0.30 to 0.93				
BDI-II[e]											–0.44	–0.75 to –0.13	–0.37	–0.69 to –0.06
Interventio	19.87 (0.18)	19.52 to 20.23	11.87 (0.80)	10.4 to 13.55	11.18 (0.89)	9.57 to 13.06	0.90	0.60 to 1.2	1.04	0.71 to 1.37				
Control	19.66 (0.32)	19.04 to 20.29	17.62 (0.75)	16.21 to 19.16	16.26 (0.90)	14.60 to 18.12	0.25	–0.06 to 0.56	0.41	0.10 to 0.72				
SACQ-R[f]											0.49	0.18 to 0.80	0.39	0.08 to 0.71
Intervention	67.64 (0.28)	67.09 to 68.20	79.54 (1.31)	77.01 to 82.14	80.21 (1.57)	77.2 to 83.34	–0.87	–1.17 to –0.57	–0.85	–1.15 to –0.54				
Control	67.82 (0.33)	67.17 to 68.48	70.68 (0.93)	68.87 to 72.53	71.94 (1.55)	68.96 to 75.05	–0.12	–0.43 to 0.19	–0.22	–0.54 to 0.09				
ATQP-SF[g]											0.20	–0.11 to 0.51	0.16	–0.16 to 0.47
Intervention	11.81 (0.22)	11.39 to 12.25	15.68 (0.68)	14.4 to 17.08	16.04 (0.86)	14.43 to 17.81	–0.45	–0.75 to –0.15	–0.44	–0.74 to –0.13				
Control	11.83 (0.21)	11.42 to 12.24	13.00 (0.73)	11.64 to 14.52	13.52 (0.82)	12.01 to 15.21	–0.06	–0.37 to 0.25	–0.10	–0.42 to 0.21				
ATQN-SF[h]											–0.13	–0.44 to 0.18	–0.29	–0.60 to 0.02
Intervention	9.01 (0.26)	8.51 to 9.55	7.56 (0.55)	6.55 to 8.71	6.86 (0.64)	5.72 to 8.23	0.36	0.06 to 0.66	0.58	0.28 to 0.88				
Control	9.04 (0.27)	8.53 to 9.58	9.41 (0.59)	8.33 to 10.63	9.85 (0.76)	8.46 to 11.47	0.00	–0.31 to 0.31	0.04	–0.27 to 0.36				

[a]F/U: follow-up.

[b]SE: Standard Error

[c]CI: Confidence Interval

[d]PHQ-9: Patient Health Questionnaire-9.

[e]BDI-II: Beck Depression Inventory-II.

XSL•FO
RenderX

[f]SACQ-R: Student Adaptation to College Questionnaire–Revised.

[g]ATQP-SF: Automatic Thought Questionnaire—Positive, Short Form.

[h]ATQN-SF: Automatic Thought Questionnaire–Negative, Short Form.

The control group showed a decrease in depressive symptoms from pre- to postassessment and follow-up; however, the magnitude was smaller than that in the intervention group, and the group mean scores were higher than those in the intervention group when comparing the PHQ-9 and BDI-II scores for the post and follow-up assessments (Table 5). When comparing scores for the pre, post, and follow-up assessments, the effect sizes (Cohen *d*) for changes in adjustment to college life and positive and negative automatic thoughts ranged from −0.22 to 0.04, indicating negligible or a small change.

PP Analysis

A PP analysis was conducted on the data of participants who completed all 28 sessions of the Mind Booster Green training program and all three assessments. Adherence rates were 89% and 99% in the intervention and WLC groups, respectively. The analysis was conducted using the GEE, as in the ITT analysis. The results showed significant time×group interactions for the PHQ-9, BDI-II, SACQ-R, ATQP-SF, and ATQN-SF scores and significant main effects of group and time for all variables. These findings were consistent with those of the ITT analysis. The results of the PP analysis are presented in Table 4.

Usability Test Results

The usability test was conducted using the MARS in the intervention group (n=80). The average usability score for Mind Booster Green was 3.88. Across domains, the mean scores were 3.57 (SD 0.66) for engagement, 4.17 (SD 0.76) for functionality, 3.40 (SD 0.75) for aesthetic quality, 4.36 (SD 0.56) for information quality, and 3.92 (SD 0.67) for app-specific features. All domain scores were above the medium level.

Discussion

Principal Findings

Mind Booster Green is an app-based CBT program designed to alleviate depressive symptoms in college students. To enhance the program's effectiveness and encourage continuous and independent use, the following strategies were applied: (1) tailoring the content using case stories reflecting the experiences college students are likely to have and (2) incorporating gamification elements, such as point and level systems, to promote app use. To verify the effectiveness of the program, we examined the reduction in depressive symptoms and improvement in college life adjustment and assessed the app's usability using a standardized measure. The key findings and implications of this study were as follows.

First, the self-help t-CBT program Mind Booster Green was found to be effective in alleviating depressive symptoms in college students immediately after completing the training provided by the app, and the effect was sustained for at least 2 months postintervention. Additionally, the intervention group that used Mind Booster Green reported improved academic, social, emotional, and health-related adjustment to college life. These results are encouraging as they demonstrate that app-based CBT programs can effectively reduce depressive symptoms through cognitive-behavioral skills training and extend these improvements to students' real-life functioning. The results of this study suggest that Mind Booster Green can complement or serve as an alternative to existing psychological counseling services, providing a cost-effective and efficient solution to mental health issues among college students.

Second, based on a detailed examination of the program's effectiveness, the intervention group exhibited reductions in depressive symptoms and changes in automatic thoughts, specifically a decrease in negative automatic thoughts and an increase in positive thoughts. Although this study did not conduct a formal mediation analysis to determine whether changes in thought patterns directly mediated symptom improvement, these findings are consistent with the theoretical underpinnings of CBT that emphasize the role of cognitive restructuring in alleviating emotional distress. CBT assumes that emotions and behaviors are influenced by individuals' cognitive appraisals and seeks to reduce psychological symptoms by identifying and modifying negatively biased thoughts through cognitive and behavioral strategies [81]. Acquiring skills to challenge and replace maladaptive automatic thoughts is considered a core component of CBT treatment [82,83]. Researchers have emphasized the importance of embedding these core mechanisms into digital CBT interventions and delivering them in ways that maintain therapeutic integrity [12]. The observed changes in automatic thoughts suggest that the program successfully incorporated the key elements of CBT and may have contributed to symptom improvement. Although relatively few studies have directly examined therapeutic mechanisms in t-CBT interventions, existing evidence reflects cognitive changes that are in line with CBT theory, suggesting potential applicability across different interventions [84-87]. These findings support the fundamental theoretical model of CBT and demonstrate its effectiveness across digital formats. They also provide empirical support for the active implementation of t-CBT as a practical and scalable approach for treating depression.

Third, despite being a self-help program, Mind Booster Green showed a relatively high adherence rate. Mind Booster Green was designed for students to use independently without the help of therapists or experts. Although self-help programs maximize the economic and accessibility advantages of t-CBT programs, low adherence rates remain a challenge. The adherence rate of 89% in the intervention group was significantly higher than the reported adherence rate of approximately 26% for technology-based self-help interventions but similar to the 85% adherence rate reported for traditional face-to-face interventions [9]. Although adherence rates in randomized controlled trials are generally higher than those in observational studies [88,89], the rate observed in this study is notable. Enhancing user engagement and participation in intervention programs is closely linked to effectiveness [11], making adherence an important issue to address when developing self-help t-CBT programs.

XSL•FO
RenderX

Thus, this study actively applied two strategies—tailoring and gamification—and the high adherence rate indirectly suggests their effectiveness.

The tailoring strategy in Mind Booster Green was designed based on social learning theory and intended to effectively model the cognitive and behavioral strategies of CBT. Although tailoring can be broadly applied to provide information that fits user characteristics and needs, the tailoring strategy in Mind Booster Green focuses on program content composition. The program features college students who share their experiences related to depression and learn CBT skills to overcome their problems, thereby modeling how users in similar situations can apply these skills in their lives. Specific narratives reflecting the characteristics of a particular population facilitate effective modeling, a technique widely used in various psychological therapy contexts [90] and reported to enhance user engagement [91]. The changes in automatic thoughts observed alongside the reduction in depressive symptoms in this study indicated that the cognitive skills targeted by the program were appropriately conveyed to the participants. Additionally, approximately 20% of the study participants mentioned that "real college student cases" in the content helped them relate to the program and learn CBT skills.

The gamification strategy in Mind Booster Green was designed based on behavioral theory to ensure that users continuously engaged in and completed the app's treatment activities. Mind Booster Green clarified the tasks that the participants needed to accomplish each day (goal setting), provided feedback on progress (progress and feedback), and implemented point and level systems (rewards) based on task completion. These are the key gamification strategies used in technology-based mental health interventions [92]. Notably, the gamification strategy used in this study was based on behavioral theory and aimed to help users learn and engage in new behaviors. Specifically, we analyzed the tasks that users needed to complete in Mind Booster Green and designed positive reinforcement strategies to increase the performance of those tasks. Points were awarded for engaging in and completing the training tasks, and a level system was introduced that showed the main character evolving as the points accumulated. Behavioral theory scientifically explores the factors that elicit and maintain human behavior, and strategies based on this theory have been reported to be effective in increasing app use behaviors [30,93]. As demonstrated in this study, using task analysis and reinforcement schedules based on established behavioral theories in app-based mental health programs can effectively encourage and teach desired behaviors.

The usability evaluation results for Mind Booster Green also reflected the effectiveness of these strategies. In this study, the MARS [75], a standardized assessment tool, was used to measure usability, allowing for comparisons with other programs. Lau et al [94] used MARS to assess the released app-based mental health interventions and reported average overall scores of 3.52 (SD 0.71). The overall score for Mind Booster Green was 3.88, which was slightly higher than the previously reported results obtained using MARS, with notably higher scores for functionality (4.17) and information quality (4.36). These positive user evaluations suggested that the intervention program development strategy, which considers user characteristics and preferences, positively influenced the user experience. This study clearly defined the target behaviors that users need to acquire to alleviate depressive symptoms and guided these behaviors by developing specific elements of tailoring and gamification based on psychological theory. The effectiveness of Mind Booster Green, the high adherence rate, and positive usability evaluation results highlight the importance of developing theoretically grounded strategies for therapeutic mechanisms and app-use behavior.

Limitations and Future Directions

This study was limited to one country, the Republic of Korea, and primarily included students from 4-year universities in Seoul, despite recruiting participants from various universities and graduate schools nationwide. The participant demographics showed a sex imbalance, with significantly fewer male participants than female participants. Additionally, the study included a broad range of participants, from freshmen to graduate students. Future studies should include a more diverse range of school types and recruit a larger number of participants of different sexes and class-year groups for repeated verification. Expanding the verification to an international scope to explore whether the program's effectiveness can be generalized to college students in other countries could also provide valuable data for developing tailored t-CBT programs for the target population.

This study applied tailoring and gamification features based on behavioral theories such as social learning and positive reinforcement to a t-CBT program and showed positive results in reducing depressive symptoms and improving adherence rates and usability. However, future research must investigate how the individual factors of tailoring and gamification mediate these positive outcomes. Existing studies analyzing the impact of these features on program effectiveness and adherence have produced conflicting results [95] owing to the limited number of studies and differences in how these features are defined. Thus, the continuous development of t-CBT programs incorporating tailoring and gamification is needed, along with research on how these effects are mediated by the specific designs and theoretical underpinnings of these strategies.

Conclusions

This study found that the self-help t-CBT program Mind Booster Green can be effective in reducing depressive symptoms and improving adjustment to college life among college students. These results suggested the potential of self-help t-CBT programs in alleviating and preventing depression in this target group. Future research should test the effectiveness and usability of the strategies applied in this study, including tailoring and gamification, across diverse populations.

XSL•FO
RenderX

Acknowledgments

This work was supported by the Technology Innovation Program (RS-2024 - 00487049, Development and Demonstration of Complex Emotion Recognition On-device AI Technology for In-vehicle Driver Emotional Services) funded by the Ministry of Trade, Industry and Energy (Korea) and by a Yonsei University Research Grant (2021).

Conflicts of Interest

KMC is the founder of the company that commercializes the intervention described in this manuscript.

Editorial Notice

This randomized study was only retrospectively registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Checklist 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File, 831 KB - mhealth_v13i1e50006_app1.pdf]

References

1. Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive–behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. Br J Psychiatry 2010 Mar;196(3):173-178. [doi: 10.1192/bjp.bp.109.066001]
2. Williams C, Martinez R. Increasing access to CBT: stepped care and CBT self-help models in practice. Behav Cogn Psychother 2008 Nov;36(6):675-683. [doi: 10.1017/S1352465808004864]
3. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
4. Karyotaki E, Riper H, Twisk J, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms. JAMA Psychiatry 2017 Apr 1;74(4):351. [doi: 10.1001/jamapsychiatry.2017.0044]
5. Saad A, Bruno D, Camara B, D'Agostino J, Bolea-Alamanac B. Self-directed technology-based therapeutic methods for adult patients receiving mental health services: systematic review. JMIR Ment Health 2021 Nov 26;8(11):e27404. [doi: 10.2196/27404] [Medline: 34842556]
6. Edge D, Watkins ER, Limond J, Mugadza J. The efficacy of self-guided internet and mobile-based interventions for preventing anxiety and depression—a systematic review and meta-analysis. Behav Res Ther 2023 May;164:104292. [doi: 10.1016/j.brat.2023.104292] [Medline: 37003138]
7. Webb CA, Rosso IM, Rauch SL. Internet-based cognitive-behavioral therapy for depression: current progress and future directions. Harv Rev Psychiatry 2017;25(3):114-122. [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
8. Karyotaki E, Kleiboer A, Smit F, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an "individual patient data" meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
9. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
10. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
11. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
12. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338. [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
13. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digital Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
14. Cuijpers P, Karyotaki E, Eckshtain D, et al. Psychotherapy for depression across different age groups: a systematic review and meta-analysis. JAMA Psychiatry 2020 Jul 1;77(7):694-702. [doi: 10.1001/jamapsychiatry.2020.0164] [Medline: 32186668]

XSL•FO
RenderX

15. Toyomoto R, Sakata M, Yoshida K, et al. Prognostic factors and effect modifiers for personalisation of internet-based cognitive behavioural therapy among university students with subthreshold depression: a secondary analysis of a factorial trial. J Affect Disord 2023 Feb 1;322:156-162. [doi: 10.1016/j.jad.2022.11.024] [Medline: 36379323]
16. Morrison LG. Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: a review. Digital Health 2015;1:2055207615595335. [doi: 10.1177/2055207615595335] [Medline: 29942544]
17. Ham K, Chin S, Suh YJ, et al. Preliminary results from a randomized controlled study for an app-based cognitive behavioral therapy program for depression and anxiety in cancer patients. Front Psychol 2019;10:1592. [doi: 10.3389/fpsyg.2019.01592] [Medline: 31402881]
18. Mullin A, Dear BF, Karin E, et al. The UniWellbeing course: a randomised controlled trial of a transdiagnostic internet-delivered cognitive behavioural therapy (CBT) programme for university students with symptoms of anxiety and depression. Internet Interventions 2015 May;2(2):128-136. [doi: 10.1016/j.invent.2015.02.002]
19. Weaver A, Zhang A, Xiang X, Felsman P, Fischer DJ, Himle JA. Entertain Me Well: an entertaining, tailorable, online platform delivering CBT for depression. Cogn Behav Pract 2023 Feb;30(1):96-115. [doi: 10.1016/j.cbpra.2021.09.003] [Medline: 36741472]
20. Strecher VJ, McClure JB, Alexander GL, et al. Web-based smoking-cessation programs: results of a randomized trial. Am J Prev Med 2008 May;34(5):373-381. [doi: 10.1016/j.amepre.2007.12.024] [Medline: 18407003]
21. Karyotaki E, Klein AM, Ciharova M, et al. Guided internet-based transdiagnostic individually tailored cognitive behavioral therapy for symptoms of depression and/or anxiety in college students: a randomized controlled trial. Behav Res Ther 2022 Mar;150:104028. [doi: 10.1016/j.brat.2021.104028] [Medline: 35066365]
22. Batterham PJ, Calear AL, Farrer L, McCallum SM, Cheng VWS. FitMindKit: randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interventions 2018 Jun;12:91-99. [doi: 10.1016/j.invent.2017.08.002] [Medline: 30135773]
23. Balaskas A, Schueller SM, Cox AL, Doherty G. The functionality of mobile apps for anxiety: systematic search and analysis of engagement and tailoring features. JMIR Mhealth Uhealth 2021 Oct 6;9(10):e26712. [doi: 10.2196/26712] [Medline: 34612833]
24. Mukhiya SK, Wake JD, Inal Y, Pun KI, Lamo Y. Adaptive elements in internet-delivered psychological treatment systems: systematic review. J Med Internet Res 2020 Nov 27;22(11):e21066. [doi: 10.2196/21066] [Medline: 33245285]
25. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 6;20(6):e10120. [doi: 10.2196/10120] [Medline: 29875087]
26. Cuijpers P, Muñoz RF, Clarke GN, Lewinsohn PM. Psychoeducational treatment and prevention of depression: the “Coping with Depression” course thirty years later. Clin Psychol Rev 2009 Jul;29(5):449-458. [doi: 10.1016/j.cpr.2009.04.005] [Medline: 19450912]
27. Rigabert A, Motrico E, Moreno-Peral P, et al. Effectiveness of online psychological and psychoeducational interventions to prevent depression: systematic review and meta-analysis of randomized controlled trials. Clin Psychol Rev 2020 Dec;82:101931. [doi: 10.1016/j.cpr.2020.101931] [Medline: 33137611]
28. Deterding S, Dixon D, Khaled R, Nacke L, editors. From game design elements to gamefulness: defining “gamification". : Association for Computing Machinery; 2011 Presented at: The 15th International Academic MindTrek Conference: Envisioning Future Media Environments; Sep 28-30, 2011 p. 9-15. [doi: 10.1145/2181037.2181040] [Medline: 27274957]
29. Cheng VWS, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717. [doi: 10.2196/13717] [Medline: 31244479]
30. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
31. Cheng VWS. Recommendations for implementing gamification for mental health and wellbeing. Front Psychol 2020;11:586379. [doi: 10.3389/fpsyg.2020.586379] [Medline: 33365001]
32. Cheng C, Ebrahimi OV. A meta-analytic review of gamified interventions in mental health enhancement. Comput Human Behav 2023 Apr;141:107621. [doi: 10.1016/j.chb.2022.107621]
33. Six SG, Byrne KA, Aly H, Harris MW. The effect of mental health app customization on depressive symptoms in college students: randomized controlled trial. JMIR Ment Health 2022 Aug 9;9(8):e39516. [doi: 10.2196/39516] [Medline: 35943788]
34. Fleming TM, Stasiak K, Moselen E, et al. Revising computerized therapy for wider appeal among adolescents: youth perspectives on a revised version of SPARX. Front Psychiatry 2019;10:802. [doi: 10.3389/fpsyt.2019.00802] [Medline: 31824344]
35. Bakker D, Kazantzis N, Rickwood D, Rickard N. Development and pilot evaluation of smartphone-delivered cognitive behavior therapy strategies for mood- and anxiety-related problems: MoodMission. Cogn Behav Pract 2018 Nov;25(4):496-514. [doi: 10.1016/j.cbpra.2018.07.002]
36. Miloff A, Marklund A, Carlbring P. The challenger app for social anxiety disorder: new advances in mobile psychological treatment. Internet Interventions 2015 Nov;2(4):382-391. [doi: 10.1016/j.invent.2015.08.001]

XSL•FO
RenderX

37. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e13080. [doi: 10.2196/13080] [Medline: 30747715]
38. Mullins JK, Sabherwal R. Gamification: a cognitive-emotional view. J Bus Res 2020 Jan;106:304-314. [doi: 10.1016/j.jbusres.2018.09.023]
39. Auerbach RP, Mortier P, Bruffaerts R, et al. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638. [doi: 10.1037/abn0000362] [Medline: 30211576]
40. Chang JJ, Ji Y, Li YH, Pan HF, Su PY. Prevalence of anxiety symptom and depressive symptom among college students during COVID-19 pandemic: a meta-analysis. J Affect Disord 2021 Sep 1;292:242-254. [doi: 10.1016/j.jad.2021.05.109] [Medline: 34134022]
41. Awadalla S, Davies EB, Glazebrook C. A longitudinal cohort study to explore the relationship between depression, anxiety and academic performance among Emirati university students. BMC Psychiatry 2020 Dec;20(1):1-10. [doi: 10.1186/s12888-020-02854-z]
42. Saunders DE, Peterson GW, Sampson JP, Reardon RC. Relation of depression and dysfunctional career thinking to career indecision. J Vocat Behav 2000 Apr;56(2):288-298. [doi: 10.1006/jvbe.1999.1715]
43. Whitton SW, Whisman MA. Relationship satisfaction instability and depression. J Fam Psychol 2010;24(6):791-794. [doi: 10.1037/a0021734]
44. Mahmoud JSR, Staten RT, Hall LA, Lennie TA. The relationship among young adult college students' depression, anxiety, stress, demographics, life satisfaction, and coping styles. Issues Ment Health Nurs 2012 Feb 28;33(3):149-156. [doi: 10.3109/01612840.2011.632708]
45. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019 Jan 1;70(1):60-63. [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
46. Copeland WE, McGinnis E, Bai Y, et al. Impact of COVID-19 pandemic on college student mental health and wellness. J Am Acad Child Adolesc Psychiatry 2021 Jan;60(1):134-141. [doi: 10.1016/j.jaac.2020.08.466] [Medline: 33091568]
47. Auerbach RP, Alonso J, Axinn WG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970. [doi: 10.1017/S0033291716001665] [Medline: 27484622]
48. Xiao H, Carney DM, Youn SJ, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017 Nov;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
49. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]
50. Raunic A, Xenos S. University counselling service utilisation by local and international students and user characteristics: a review. Int J Adv Counselling 2008 Dec;30(4):262-267. [doi: 10.1007/s10447-008-9062-0]
51. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. : Pew Research Center; 2016 URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies [accessed 2025-05-25]
52. Batterham PJ, Calear AL. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26. [doi: 10.2196/mental.7722] [Medline: 28666976]
53. Creighton TB. Digital natives, digital immigrants, digital learners: an international empirical integrative review of the literature. Educ Leadersh Rev 2018;19(1):132-140 [FREE Full text]
54. Bittner JV, Shipper J. Motivational effects and age differences of gamification in product advertising. J Consum Mark 2014 Aug 11;31(5):391-400. [doi: 10.1108/JCM-04-2014-0945]
55. Zhang L, Shao Z, Li X, Feng Y. Gamification and online impulse buying: the moderating effect of gender and age. Int J Inf Manage 2021 Dec;61:102267. [doi: 10.1016/j.ijinfomgt.2020.102267]
56. Powell J, Hamborg T, Stallard N, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. J Med Internet Res 2013;15(1):e2. [doi: 10.2196/jmir.2240]
57. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
58. Twomey C, O'Reilly G. Effectiveness of a freely available computerised cognitive behavioural therapy programme (MoodGYM) for depression: meta-analysis. Aust N Z J Psychiatry 2017 Mar;51(3):260-269. [doi: 10.1177/0004867416656258] [Medline: 27384752]
59. Lattie E, Cohen KA, Winquist N, Mohr DC. Examining an app-based mental health self-care program, intellicare for college students: single-arm pilot study. JMIR Ment Health 2020 Oct 10;7(10):e21075. [doi: 10.2196/21075] [Medline: 33037874]
60. Pan JY, Carlbring P, Lu L. Efficacy of internet-based cognitive behavioral therapy for Hong Kong university students: a randomized controlled trial. Res Soc Work Pract ;2024:10497315241252054. [doi: 10.1177/10497315241252054]
61. Ha SW, Kim J. Designing a scalable, accessible, and effective mobile app based solution for common mental health problems. Int J Hum-Comput Interact 2020 Aug 26;36(14):1354-1367. [doi: 10.1080/10447318.2020.1750792]

XSL•FO
RenderX

62. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
63. Park SJ, Choi HR, Choi JH, Kim KW, Hong JP. Reliability and validity of the Korean version of the Patient Health Questionnaire-9 (PHQ-9). Anxiety Mood 2010;6(2):119-124 [FREE Full text]
64. An JY, Seo ER, Lim KH, Shin JH, Kim JB. Standardization of the Korean version of screening tool for depression (Patient Health Questionnaire-9, PHQ-9). J Korean Soc Biol Ther Psychiatry 2013;19(1):47-56 [FREE Full text]
65. Beck AT, Steer RA, Depression B. Inventory—2nd Edition Manual: The Psychological Corporation; 1996.
66. Kim JH, Lee EH, Hwang ST, Hong SH. Korean Version of the Beck Depression Inventory-II Manual: Korean Psychology Co., Ltd; 2015.
67. Sung HM, Kim JB, Park YN, Bae DS, Lee SH, Ahn HY. A study on the reliability and the validity of Korean version of the Beck Depression Inventory-II (BDI-II). J Korean Soc Biol Psychiatry 2008;14(2):201-212 [FREE Full text]
68. Hollon SD, Kendall PC. Cognitive self-statements in depression: development of an automatic thoughts questionnaire. Cogn Ther Res 1980 Dec;4(4):383-395. [doi: 10.1007/BF01178214]
69. Kwon SM, Yoon HK. Development and application of the Korean version of the Automatic Thoughts Questionnaire. Stud Res 1994;29(1):10-25.
70. Heo S, Kim J. Development and validation of short form of the Korean version of the Automatic Thoughts Questionnaire-Negative. HSS21 2020 Jun 30;11(3):783-796. [doi: 10.22143/HSS21.11.3.56]
71. Heo S, Kim J. A validation of the Korean Version of the Positive Subscale of the Automatic Thoughts Questionnaire-Revised (K-ATQ-RP). HSS21 2020 Oct 30;11(5):751-764. [doi: 10.22143/HSS21.11.5.55]
72. Kendall PC, Howard BL, Hays RC. Self-referent speech and psychopathology: the balance of positive and negative thinking. Cogn Ther Res 1989 Dec;13(6):583-598. [doi: 10.1007/BF01176069]
73. Baker RW, Siryk B. Measuring adjustment to college. J Couns Psychol 1984;31(2):179-189. [doi: 10.1037/0022-0167.31.2.179]
74. Lee YJ. The Non-Residential Students' Adaptation to College Life and Career Plans: Ewha Womans University; 1999.
75. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015;3(1):e27. [doi: 10.2196/mhealth.3422]
76. Gregoire J. ITC guidelines for translating and adapting tests (second edition). Int J Test 2018 Apr 3;18(2):101-134. [doi: 10.1080/15305058.2017.1398166]
77. Chung KM, Suh YJ, Chin S, et al. A pilot study testing the efficacy of dCBT in patients with cancer experiencing sleep problems. Front Psychol 2022;13:699168. [doi: 10.3389/fpsyg.2022.699168] [Medline: 35756310]
78. Mind booster green. iApp Store. URL: https://apps.apple.com/kr/app/id1557176314 [accessed 2025-05-20]
79. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
80. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology 2010 Jul;21(4):467-474. [doi: 10.1097/EDE.0b013e3181caeb90] [Medline: 20220526]
81. Beck JS. Cognitive Behavior Therapy: Basics and Beyond: Guilford Press; 2020.
82. Cristea IA, Kok RN, Cuijpers P. Efficacy of cognitive bias modification interventions in anxiety and depression: meta-analysis. Br J Psychiatry 2015 Jan;206(1):7-16. [doi: 10.1192/bjp.bp.114.146761]
83. Forand NR, Barnett JG, Strunk DR, Hindiyeh MU, Feinberg JE, Keefe JR. Efficacy of guided iCBT for depression and mediation of change by cognitive skill acquisition. Behav Ther 2018 Mar;49(2):295-307. [doi: 10.1016/j.beth.2017.04.004] [Medline: 29530267]
84. Mogoaşe C, David D, Koster EHW. Clinical efficacy of attentional bias modification procedures: an updated meta-analysis. J Clin Psychol 2014 Dec;70(12):1133-1157. [doi: 10.1002/jclp.22081] [Medline: 24652823]
85. Wu A, Scult MA, Barnes ED, Betancourt JA, Falk A, Gunning FM. Smartphone apps for depression and anxiety: a systematic review and meta-analysis of techniques to increase engagement. NPJ Digital Med 2021 Feb;4(1). [doi: 10.1038/s41746-021-00386-8]
86. Domhardt M, Steubl L, Boettcher J, et al. Mediators and mechanisms of change in internet- and mobile-based interventions for depression: a systematic review. Clin Psychol Rev 2021 Feb;83:101953. [doi: 10.1016/j.cpr.2020.101953]
87. Mogoa e C, Cobeanu O, David O, Giosan C, Szentagotai A. Internet-based psychotherapy for adult depression: what about the mechanisms of change? J Clin Psychol 2017 Jan;73(1):5-64. [doi: 10.1002/jclp.22326] [Medline: 27684405]
88. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13. [doi: 10.2196/jmir.1194] [Medline: 19403466]
89. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152. [doi: 10.2196/jmir.2104] [Medline: 23151820]
90. Hamid A, Arshad R, Shahid S. What are you thinking? using CBT and storytelling to improve mental health among college students. 2022 Apr 29 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 30 to May 5, 2022; New Orleans, USA. [doi: 10.1145/3491102.3517603]

XSL•FO
RenderX

91. Busselle R, Bilandzic H. Measuring narrative engagement. Media Psychol 2009 Nov 23;12(4):321-347. [doi: 10.1080/15213260903287259]
92. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016 Aug 24;3(3):e39. [doi: 10.2196/mental.5710] [Medline: 27558893]
93. Xu L, Shi H, Shen M, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 3;10(2):e27794. [doi: 10.2196/27794] [Medline: 35113034]
94. Lau N, O'Daffer A, Yi-Frazier JP, Rosenberg AR. Popular evidence-based commercial mental health apps: analysis of engagement, functionality, aesthetics, and information quality. JMIR Mhealth Uhealth 2021 Jul 14;9(7):e29689. [doi: 10.2196/29689] [Medline: 34259639]
95. Six SG, Byrne KA, Tibbett TP, Pericot-Valverde I. Examining the effectiveness of gamification in mental health apps for depression: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 29;8(11):e32199. [doi: 10.2196/32199] [Medline: 34847058]

Abbreviations

ATQN-SF: Automatic Though Questionnaire—Negative, Short Form
ATQP-SF: Automatic Though Questionnaire—Positive, Short Form
BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GEE: generalized estimating equation
IRB: institutional review board
ITT: intention-to-treat
MARS: Mobile App Rating Scale
PHQ-9: Patient Health Questionnaire-9
PP: per-protocol
SACQ-R: Student Adaptation to College Questionnaire-Revised
SNS: social networking service
t-CBT: technology-based cognitive behavioral therapy
WLC: waitlist control

Edited by L Buis; submitted 12.09.24; peer-reviewed by JY Pan; revised version received 15.04.25; accepted 29.04.25; published 19.06.25.

Please cite as:
Kim MH, Moon YK, Chung KM
An App-Based Cognitive Behavioral Therapy Program Tailored for College Students: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e50006
URL: https://mhealth.jmir.org/2025/1/e50006
doi:10.2196/50006

©Min Hee Kim, Yeon-kug Moon, Kyong-Mee Chung. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing the Efficacy of the INTELLECT Cognitive Behavioral Therapy Mobile App for Anxiety and Depressive Symptoms Among At-Risk Japanese Employees: Randomized Controlled Trial

Kengo Yokomitsu[1], PhD; Riki Oimatsu[2], MA; Sean Han Yang Toh[3], BSocSci; Oliver Sündermann[4], PhD

[1]School of Psychological Sciences, University of Human Environments, 9-12, Dougo-himata, Matsuyama, Ehime, Japan

[2]Intellect Japan Co, Ltd, Tokyo, Japan

[3]Department of Psychology, Harvard University, Cambridge, MA, United States

[4]Intellect Pte Ltd, Singapore, Singapore

Corresponding Author:
Kengo Yokomitsu, PhD
School of Psychological Sciences, University of Human Environments, 9-12, Dougo-himata, Matsuyama, Ehime, Japan

Abstract

Background: In Japan, the prevalence of anxiety and depressive symptoms within the working population has risen. This has been accentuated by the economic repercussions of the COVID-19 pandemic and the social isolation resulting from remote work setups. Mobile health apps, particularly those incorporating cognitive behavioral therapy (CBT) features, have shown potential in addressing these symptoms. These self-guided CBT interventions hold promise in alleviating the heightened depressive and anxiety symptoms often observed among Japanese employees.

Objective: Using a randomized controlled trial, we compared the efficacy of the "INTELLECT" app against a no-treatment control group in improving depressive symptoms and CBT skills among Japanese full-time employees at postintervention and 2-month follow-up.

Method: A total of 123 full-time Japanese employees were randomly allocated to either the intervention group (INTELLECT), where they engaged with self-help CBT features, or to a control group receiving no treatment. Intervention participants were required to engage with these features for at least 20 minutes per week over a span of 4 weeks. Weekly self-reported assessments were collected from all participants starting from baseline and continuing until the end of the 4-week intervention period. Subsequent assessments were conducted at 1-month and 2-month follow-up intervals. Linear mixed models were used to evaluate any effects of the self-guided intervention on depressive symptoms, as measured by the Patient Health Questionnaire-4, and cognitive behavioral skills, as measured by the Cognitive Behavioral Therapy Skills Scale. The app's feasibility, usability, and acceptability ratings were also examined using the Implementation Outcome Scales for Digital Mental Health (iOSDMH).

Results: The final sample (n=73) consisted of 46 (63%) participants who were female, 23 (32%) participants who were male, and 4 (6%) participants who identified as other genders, with a mean age of 40.4 (SD 10.7) years. Significant time × group interactions were found at postintervention and 2-month follow-up, with the intervention group (n=34) reporting significantly lower depressive symptoms than the control group (n=38) at postintervention ($t_{364.7426}$=−2.243; P=.03; Cohen d=−0.57, 95% CI −1.07 to −0.06) and 2-month follow-up ($t_{364.6948}$=−3.284; P<.001; Cohen d=−0.85, 95% CI −1.38 to −0.32). In addition, intervention participants reported significantly greater improvements in self-monitoring cognitive skills than control participants at postintervention ($t_{120.7526}$=2.672; P=.01; Cohen d=0.68, 95% CI 0.17 to 1.18) but not follow-up ($t_{121.5475}$=1.947; P=.05; Cohen d=0.50, 95% CI −0.01 to 1.02).

Conclusions: This study provides evidence that CBT features on the INTELLECT app are effective in improving depressive symptoms and self-monitoring cognitive skills.

Trial Registration: University Hospital Medical Information Network Center UMIN000051354; https://tinyurl.com/4euwymyb

(JMIR Mhealth Uhealth 2025;13:e60871) doi:10.2196/60871

KEYWORDS

depressive symptom; internet cognitive behavioral therapy; randomized controlled trial; depression; behavior; cognitive; mobile app; anxiety; Japan; mobile health; employee; linear mixed model; questionnaire; feasibility; acceptability; self-monitoring usability

XSL•FO
RenderX

Introduction

Overview

Depressive and anxiety disorders wield substantial global influence. The Global Burden of Diseases, Injuries, and Risk Factors Study, for instance, identified depression and anxiety disorders as 2 of the most disabling mental disorders, ranking among the top 25 leading causes of global burden in 2019 [1]. The COVID-19 pandemic has also seen a notable surge of 27.6% in major depressive disorder cases and a 25.6% increase in anxiety disorder cases in 2020 [2]. Similar figures were reported in Japan, where a previous study [3] conducted a nationwide survey of over 8000 Japanese citizens and found that approximately one-fifth of respondents experienced clinically severe anxiety and depressive symptoms. Numerous studies also found that significant changes in the economic and working conditions caused by the pandemic contributed to a marked increase in symptoms of anxiety and depression among Japanese employees [4-6].

Despite the increasing prevalence of these symptoms, many Japanese employees primarily rely on informal counseling with family members or non-evidence-based self-management strategies for coping [7]. Concurrently, there is a growing demand for alternative, accessible mental health care options. Mobile health (mHealth) apps hold significant promise in delivering cost-effective, evidence-based solutions for mental disorders and in facilitating outcome monitoring. These services are especially valuable for socially disadvantaged groups, such as individuals who face barriers in accessing face-to-face clinical services [8]. Such services are also timely given the increasing ownership of mobile devices [9,10], which has facilitated access to mental health care in daily life, overcoming geographical and temporal limitations [11-14]. The increase in mental health apps in Japan is evident from a systematic review conducted between June 4 and June 11, 2021, which identified 172 mental health apps available on Japan's Google Play Store and Apple App Store [15]. By September 2022, 53% of Japanese smartphone users were engaging with health care apps, indicating a remarkable 21% point increase from September 2019 [16]. The popularity of digital health care solutions seems to offer the potential for digital interventions to alleviate the declining mental health of Japanese employees. Several studies [17,18] have implemented digital interventions targeting Japanese workers and have reported reductions in depressive symptoms. However, these studies have primarily focused on employees with subthreshold depressive symptoms, and research involving higher-risk Japanese working populations remains limited. In Japan, digital mental health interventions are predominantly implemented in universally accessible formats designed to accommodate a broad spectrum of users. Prior research on such universally designed digital interventions has primarily concentrated on the prevention of major depressive episodes [19] or targeted individuals exhibiting subthreshold depressive symptoms [17,18]. Given the relatively high psychological barriers to seeking psychiatric services in Japan, there is substantial interest in anonymously accessible self-help interventions. Accordingly, evaluating the efficacy of universally designed programs in alleviating depressive symptoms among high-risk workers in Japan holds particular significance, not only for guiding future clinical applications but also for contributing to advancing research in this domain.

Moreover, it is now well established in the field that the effectiveness of digital cognitive behavioral therapy (CBT) in reducing depressive symptoms is supported by a growing body of evidence. Indeed, the previous study [20] found significantly larger effect sizes for self-help interventions involving CBT principles in improving anxiety and depression levels compared with no-treatment control groups. In some studies, these self-help CBT features even demonstrated comparable effectiveness and adherence rates to traditional face-to-face therapy [21,22]. Additionally, these features have reduced depressive symptoms in Japanese adults [23]. Given the importance of incorporating behavior change mechanisms to promote positive health behaviors, CBT is regarded as a fundamental therapeutic component in the delivery of mHealth apps for individuals experiencing mental health challenges. In this context, Tudor-Sfetea et al [24] found that compared with non-CBT-based interventions, CBT-based digital interventions are associated with greater user engagement and sustained app use, both of which contributed to facilitating behavioral change. Furthermore, Lin et al [25] highlighted that psychological empowerment and well-being in their life are key elements that can enhance the effectiveness of CBT-based mHealth interventions. Building on this perspective, this study explores how user engagement is also observable among Japanese working adults using a CBT-based intervention. Despite some evidence of these CBT self-help features, little is known about their effectiveness for clinically at-risk working adults with depressive and anxiety symptoms in Asia.

This Study

The primary objective of the study is to examine whether self-help CBT-based features can improve depressive and anxiety symptoms among Japanese full-time employees experiencing mild, moderate, or severe levels of these symptoms. In selecting evidence-based features, the study uses "INTELLECT," a mental health app that provides free access to a variety of evidence-based self-help features aimed at enhancing mental well-being for employees and consumers across Asia-Pacific countries. The efficacies for some of these features have been demonstrated in previous randomized controlled trials (RCTs) [26-28] and longitudinal studies [29,30]. In particular, the CBT-based features within INTELLECT have shown significant improvements in positive (ie, psychological resilience, body image perceptions, and self-compassion) [27,29] and negative mental well-being (ie, symptoms of anxiety and stress) [26,28] when compared with active control groups among working adults. However, no studies to our knowledge have tested the effectiveness of these features in clinically at-risk working populations. In an RCT setting, our primary hypothesis posits that the CBT features within the INTELLECT app will demonstrate greater efficacy in reducing symptoms of anxiety and depression among Japanese employees experiencing mild, moderate, or severe levels of these symptoms, compared with a waitlist control group. Our secondary objectives encompass evaluating whether these features also lead to improvements in

XSL•FO
RenderX

CBT skills and assessing the acceptability, feasibility, and satisfaction levels among its users.

Methods

Trial Design

This study was an RCT with two groups: (1) a group engaging with self-help CBT features on an INTELLECT app for 3 weeks (intervention group) and (2) a no-treatment and assessment-only (AO) group (control group).

Outcome measures were assessed from baseline and continued until the end of the 4-week intervention period, with subsequent assessments conducted at 1-month and 2-month follow-up intervals. The primary dependent variable measured was depression level, while secondary dependent variables included cognitive behavioral skills. The app's feasibility, usability, and acceptability ratings were also assessed using the Implementation Outcome Scales for Digital Mental Health (iOSDMH).

Ethical Considerations

Ethics approval was given to the institutional review board of the first author's affiliation (Ethics Review Committee, School of Psychological Sciences, University of Human Environments: 2023D-001). This trial was registered with the University Hospital Medical Information Network (UMIN) Center on June 16, 2023 (registration number: UMIN000051354). All participants provided informed consent before beginning the study. Consent was obtained electronically through a secure Google Survey platform after participants reviewed the study information sheet. Participation was entirely voluntary, and individuals could withdraw at any time without consequence. All data collected were anonymized prior to analysis. Personally identifiable information was neither collected nor stored. Data were stored on encrypted servers with restricted access to protect participant privacy and confidentiality.

Participants received compensation for their participation in the form of Amazon Gift Cards totaling ¥9000 Japanese yen (approximately US $64). Reimbursements were distributed in 3 stages: baseline (¥1000; approximately US $7), postintervention (¥3000; approximately US $21), and 2-month follow-up (¥5000; approximately US $36). All currency conversions to US dollars are based on the exchange rate at the time of the study (¥1=US $0.00713, as of June 16, 2023).

Participants and Recruitment

Participants were included in the trial if they met the following inclusion criteria: (1) mild to severe depressive symptoms on the Japanese version of Patient Health Questionnaire-9 (PHQ-9) [31] with scores between 10 and 27, (2) aged 18 years or older, and (3) owning and able to use a smartphone/tablet device.

Participants were initially introduced to the study through web-based advertisements disseminated from June 16, 2023, to October 31, 2023. These advertisements directed potential participants to recruitment sites where participation information sheets and informed consent were provided. A total of 203 participants initially responded to online advertisements for recruitment. Of these, 130 were excluded from the study due to various reasons, including voluntary withdrawal, failure to complete the questionnaire, or not meeting the study's inclusion criteria. The final sample (n=73) included 46 (63%) female participants, 23 (32%) male participants, and 4 (6%) participants identifying as other genders. The participants had a mean age of 40.4 (SD 10.7) years and were all full-time Japanese employees who owned a smartphone. Participants received monetary reimbursement for their participation.

Shortly after the trial began, the eligibility criteria were amended to eliminate the exclusion criterion related to suicidal ideation as assessed by item 9 of the PHQ-9. This revision was necessitated by the fact that all initial applicants were being excluded under this criterion, thereby rendering the continuation of the trial infeasible.

Procedure

Participants who were directed to a dedicated recruitment website via online advertisements followed a link to a secured Google Survey, where participants completed baseline measures on the severity of depression (PHQ-9), and demographic information such as gender, age, occupation, education level, annual income, marital status, types of smartphones owned, and history of treatment for depression and structured psychotherapy for mental disorders to establish baseline ratings. Only full-time Japanese employees aged 18 - 65 years with over mild levels of depressive symptoms (n=123) were then automatically randomized to one of 2 conditions—intervention (n=59) or active waitlist control (n=62)—using a computer-generated random sequence. Upon consenting to participate in the study, participants completed the primary outcome measure on the severity of depression (PHQ-4) and the secondary outcome measures on the skills related to cognitive behavioral therapy (Cognitive Behavioral Therapy Skills Scale [CBTSS]). In this study, they were not informed about the conditions they were allocated to and the real nature of the study. Instead, they were informed that the study would examine the efficacy of the mHealth app (INTELLECT) in promoting well-being.

Participants in the intervention group were then directed to download the INTELLECT mHealth app and provided with a unique code to register and access a research version of the app containing only the assigned program. This step was taken to ensure that any observed treatment effects were solely attributable to the program, rather than other features available in the app. It was verified whether participants had downloaded and registered the app through the backend system. Intervention group participants engaged with the "INTELLECT" app for a minimum of 20 minutes per week over a 4-week period, while active waitlist control group participants accessed the app once the intervention participants had completed the 4-week intervention. Engagement of the app for the intervention group was monitored on a weekly basis, and participants were informed when the app engagement was low. Five participants dropped out of the research, either due to low engagement with the app or loss of contact. Additionally, one participant's data will not be included in the analysis due to excessive use of the app after 4 weeks.

Both intervention and control participants completed a series of self-report measures at 7 timepoints: baseline (PHQ-4,

XSL•FO
RenderX

CBT-SS), week 1 (PHQ-4), week 2 (PHQ-4), week 3 (PHQ-4), postintervention (PHQ-4, CBT-SS, iOSDMH), 1-month (PHQ-4), and 2-month follow-up (PHQ-4, CBT-SS). Upon full participation, each participant received 9000 Japanese yen (approximately US $58) in Amazon Gift Cards as a token of appreciation. Reimbursements were distributed at 3 different timepoints as follows: Baseline (¥1000 Japanese yen [approximately US $6] Amazon Gift Card), Postintervention (¥3000 Japanese yen [approximately US $18] Amazon Gift Card), 2-month follow-up (¥3000 Japanese yen [approximately US $18] Amazon Gift Card).

Intervention

INTELLECT is a consumer-based mental health app that offers users access to a variety of self-guided, evidence-based features, some of which have been validated in previous RCTs [26-28] and longitudinal studies [29,30]. The app's "Home" tab provides access to 4 main self-care features as follows: "Learning Paths," "Rescue Sessions," "Wellbeing Check-in," and "Guided Journals." "Learning Paths" are designed to educate participants using evidence-based content to deepen their understanding and improve their self-management of mental health. For example, the "Feeling Depressed Learning Path" offers psychoeducation on cognitive distortions linked to low mood and depression, helping users learn how to reframe their thoughts. In other studies, this intervention has been shown to effectively aid users in altering their negative thoughts and emotions. "Rescue Sessions" are self-guided interventions that focus on addressing specific themes of adversity commonly faced by individuals. These themes include "overthinking," "motivation," "stress management," "burnout," and "relationships". Each session provides targeted support using CBT-based strategies to overcome these challenges. The interventions are firmly rooted in CBT principles to identify and change negative thought patterns and behaviors. "Rescue Sessions" provides a comprehensive toolkit that enables individuals to strengthen their mental resilience and well-being by selecting the session most relevant to their current struggles, mindfulness, self-compassion, and CBT. The "Wellbeing Check-in" feature allows users to evaluate their feelings and stress levels by selecting emojis that best represent their current state. This daily self-monitoring tool encourages participants to track their emotions, fostering greater self-awareness and providing opportunities to manage negative emotions. "Guided Journals" offer various themes such as "reflection," "problem-solving," "goal-setting," "sleep," or "self-esteem." Each journal prompts participants to write entries related to these themes, aiding in the identification and restructuring of cognitive patterns. For example, the "gratitude" journal encourages users to identify and record positive instances of the day, challenging momentary negative thought patterns and promoting cognitive restructuring skills. This practice helps foster a habit of considering alternative perspectives.

Randomization, Allocation Concealment, and Blinding

Regarding assignment to both participant groups, INTELLECT Japan's researcher conducted simple randomization using a computer-generated random sequence developed by the researcher. No blocking strategy was used during the allocation process. Stratification was used for the previous management of depression and structured psychotherapy for mental disorders. To implement the random allocation sequence, a roster of 73 individuals who consented to participate in the study was allocated to either the INTELLECT or AO groups according to the pregenerated random sequence described above. Additionally, all outcome measures were collected through self-administered online questionnaires, and the data collection process was designed such that individual participant responses could not influence the behavior or decisions of the study staff. These procedures were implemented to minimize bias during group allocation. However, complete allocation concealment may not have been fully ensured due to the potential predictability of the assignment sequence. Furthermore, owing to the nature of the study design, no efforts were made to match the intervention and control conditions in terms of appearance or delivery. Participants in the intervention group were granted immediate access to the app, whereas those assigned to the waitlist control group received no intervention during the trial period. Consequently, the 2 groups differed in both the content they received and their level of engagement. To minimize expectancy effects, all participants were informed that access to the app would be provided either immediately or following the completion of the study.

Measures

No changes were made to the primary or secondary outcomes after the trial commenced.

Primary Outcome Measure

The Japanese version of Patient Health Questionnaire-4 (PHQ-4) [32] is a 4-item self-reported questionnaire for depression symptoms. Items are scored on a 4-point Likert scale (ranging from 0="not at all" to 3="nearly every day"), with higher scores indicating more depressive symptoms. In an earlier study [31], this scale demonstrated good internal consistency (α=.93) and good convergent validity (correlation coefficient with the Kessler Psychological Distress Scale: r=0.81) in a Japanese clinical population.

Secondary Outcome Measures

The CBTSS is a 32-items self-reported questionnaire that assesses the frequency of CBT skills utilization [33]. The CBTSS was included in this study to examine whether using the INTELLECT app improves cognitive behavioral therapy skills. Items are scored on a 4-point Likert-type scale (ranging from 1="almost none of the above" to 4="quite true"), with higher scores indicating frequent CBT skills utilization. This scale has a 5-factor structure (cognitive restructuring, behavioral activation, assertiveness communication, self-monitoring, and problem-solving). The high reliability and validity of this scale have been reported in previous studies involving Japanese workforce populations [33].

The Implementation Outcome Scales for Digital Mental Health (iOSDMH measures the feasibility, appropriateness, and acceptability of digital mental health interventions. This is a 19-items self-report measure using a 4-point Likert-type scale (ranging from 1="disagree" to 4="agree"). Internal consistency

XSL•FO
RenderX

and validity were confirmed in the scale development study [34].

Analytic Approach

Descriptive statistics were reported as means and SDs for continuous variables and frequencies for dichotomous or categorical variables. We used linear mixed modeling (LMM) to analyze the study's primary outcome measures. LMM was selected because of its strength in accommodating missing data and its ability to incorporate random effects into analyses. In the primary analysis, the dependent variable was the PHQ-4 score, and the independent variables were assignment (categorical variables: INTELLECT, AO) and time (categorical variables: baseline [T1], and each follow-up [T2-T7]), with the interaction of assignment and time as a fixed-effect variable and participants as a random-effects variable. Secondary outcomes (secondary outcome measures) were analyzed in the same manner as the primary outcome. Moreover, the standardized effect sizes and 95% CI were estimated. Cohen *d* was the effect size reported, whereby 0.20 to 0.49 indicates a small effect, 0.50 to 0.79 indicates a moderate effect, and 0.80 indicates a large effect. For all analyses, $P<.05$ was considered statistically significant. All statistical analyses were performed using R software (version 4.0.2; R Foundation for Statistical Computing). The "lme4" package was used for LMM.

Sample Size

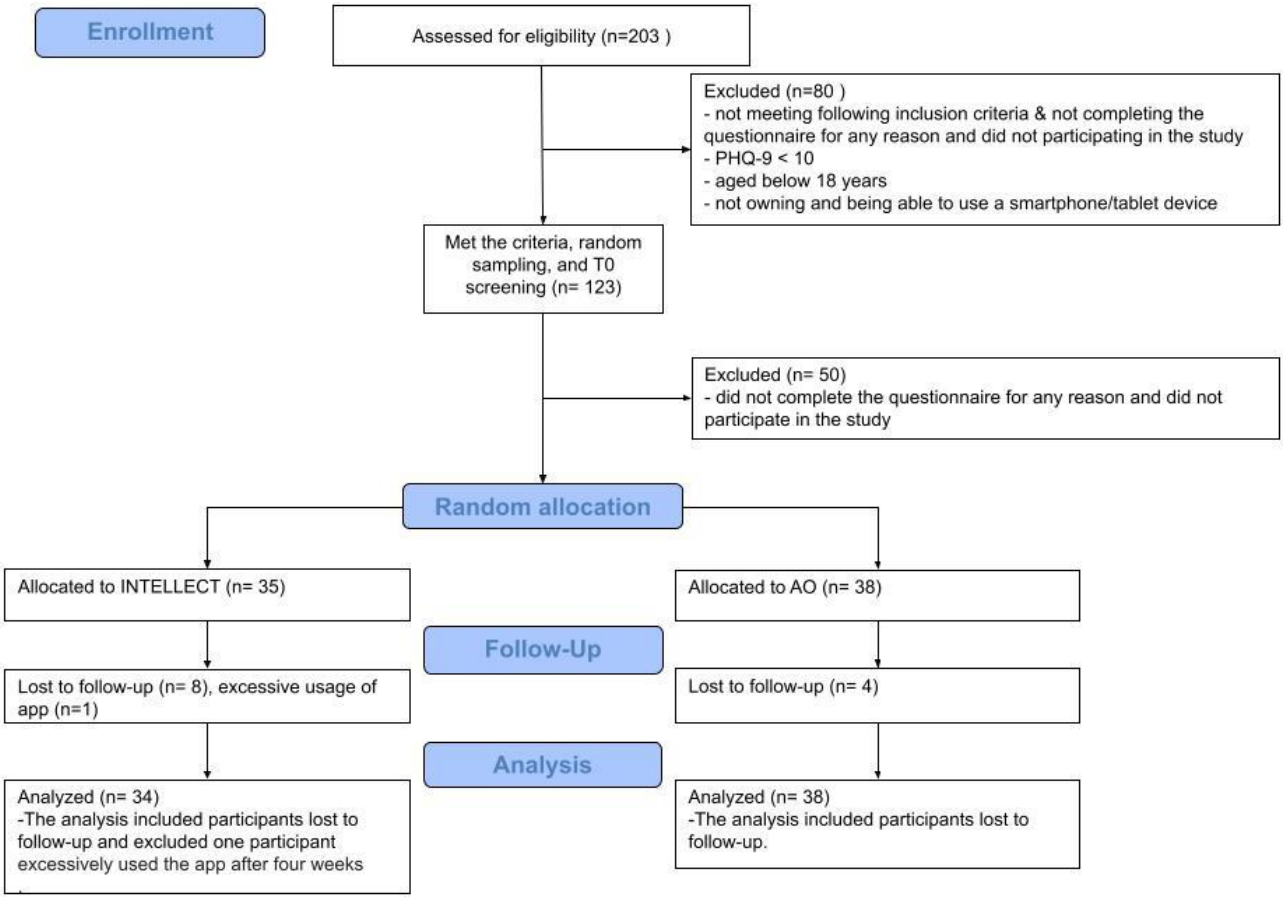
In this study, no formal sample size calculation was performed prior to trial initiation due to the exploratory nature of the research. Recruitment was discontinued after enrolling 73 participants owing to research funding constraints. No interim analyses were planned or conducted, and no stopping guidelines were established.

Results

Participant Flow, Characteristics, and Recruitment

Of the 73 participants, 48% (n=35, mean age 38.91, SD 10.03 years) and 52% (n=38, mean age 41.82, SD 10.80 years) were assigned to INTELLECT and AO groups, respectively. Of the 35 participants allocated to the INTELLECT group, 8 were nonresponsive at the 2-month follow-up, as were 4 of the 38 participants in the AO group (Figure 1). The study period spanned between June 16, 2023, and March 31, 2024. Table 1 shows participants' demographic data.

Figure 1. CONSORT (Consolidated Standards of Reporting Trails) flow diagram. AO: assessment-only; PHQ: Patient Health Questionnaire.



XSL•FO
RenderX

Table . Participant demographics and clinical characteristics (n=73).

	Total (n=73), n (%)	Intervention group (n=35), n (%)	Control group (n=38), n (%)
Gender			
Men	23 (31.5)	13 (37.1)	10 (26.3)
Women	46 (63.0)	20 (57.1)	26 (68.4)
Other	3 (4.1)	1 (2.9)	2 (5.3)
No response	1 (1.4)	1 (2.9)	0 (0)
Age (years)	40.42 (10.7)	38.91 (10.03)	41.82 (10.80)
Occupation			
Managing post	4 (5.5)	2 (5.7)	2 (5.3)
Technical job	16 (21.9)	6 (17.1)	10 (26.3)
Clerical worker	11 (15.1)	6 (17.1)	5 (13.2)
Sales staff	3 (4.1)	2 (5.7)	1 (2.6)
Service industry	15 (20.5)	9 (25.7)	6 (15.8)
Security worker	1 (1.4)	1 (2.9)	0 (0)
Production process worker	1 (1.4)	0 (0)	1 (2.6)
Transport and machine operation worker	3 (4.1)	1 (2.9)	2 (5.3)
Construction and mining worker	3 (4.1)	2 (57.0)	1 (2.6)
Carrying, cleaning, packaging, and related worker	2 (2.7)	0 (0)	2 (5.3)
Full-time homemaker	5 (6.8)	2 (5.7)	3 (7.9)
Family care/own illness treatment	3 (4.1)	2 (5.7)	1 (2.6)
Temporary retirement	5 (6.8)	2 (5.7)	3 (7.9)
Others	1 (4.1)	0 (0)	1 (2.6)
Education			
Elementary and junior high school	7 (9.6)	3 (8.6)	4 (10.5)
High school (equivalent test)	23 (31.5)	10 (28.6)	13 (34.2)
2-year and career college	8 (11.0)	6 (17.1)	2 (5.3)
4-year college	31 (42.5)	13 (37.1)	18 (47.4)
Graduate school	4 (5.5)	3 (8.6)	1 (2.6)
Income (monthly) (Japanese ¥)[a]			
<100,000	9 (12.3)	4 (11.4)	5 (13.2)
100,000 - 199,999	11 (15.1)	4 (11.4)	7 (18.4)
200,000 - 299,999	9 (12.3)	7 (20.0)	2 (5.3)
300,000 - 399,999	23 (31.5)	10 (28.6)	13 (34.2)
400,000 - 499,999	7 (9.6)	4 (11.4)	3 (7.9)
500,000 - 599,999	6 (8.2)	3 (8.6)	3 (7.9)
600,000 - 699,999	2 (2.7)	1 (2.9)	1 (2.6)
700,000 - 799,999	3 (4.1)	1 (2.9)	2 (5.3)
800,000 - 899,999	1 (1.4)	1 (2.9)	0 (0)
900,000 - 999,999	1 (1.4)	0 (0)	1 (2.6)
1000,000 - 1500,000	0 (0)	0 (0)	0 (0)
>1500,000	1 (1.4)	0 (0)	1 (2.6)

XSL•FO
RenderX

	Total (n=73), n (%)	Intervention group (n=35), n (%)	Control group (n=38), n (%)
Marital status			
Married	33 (45.2)	15 (42.9)	18 (47.4)
Single	30 (41.1)	17 (48.6)	13 (34.2)
Divorced	10 (13.7)	3 (8.6)	7 (18.4)
Smartphone they are using			
Android	25 (34.2)	14 (40.0)	11 (28.9)
iPhone	46 (63.0)	20 (57.1)	26 (68.4)
Others	2 (2.7)	1 (2.9)	1 (2.6)
Previous management of depression	37 (50.7)	19 (54.3)	18 (47.4)
Previous structured psychotherapy for mental disorders other than depression	6 (8.2)	3 (8.6)	3 (7.9)
Habitual use of alcohol	8 (11.0)	6 (17.1)	2 (5.3)
Habitual use of tobacco	16 (21.9)	7 (20.0)	9 (23.7)

[a] ¥1=US $0.00713, as of June 16, 2023.

Group Differences for the Primary Outcome Variable

72 participants were included in the analysis. The results showed that participants' PHQ-4 scores significantly decreased from baseline at both postintervention and 2-month follow-up. There was a significant interaction between group and time at postintervention ($t_{364.7426}$=−2.243, P=.03) and 2-month follow-up ($t_{364.6948}$=−3.284, P<.001). As shown in Table 2, the effect sizes were moderate and large at postintervention and follow-up, respectively at postintervention (Cohen d=−0.57, 95% CI −1.07 to −0.06) and at 2 month follow-up (Cohen d=−0.85, 95% CI −1.38 to −0.32). A post hoc power analysis was conducted for the primary outcome based on a sample size of 72, an observed effect size of Cohen d=0.85, and a significance level of 0.05. The analysis indicated a power of approximately 99%, suggesting the study had sufficient statistical power to detect the observed effect.

XSL•FO
RenderX

Table . Results of depressive symptoms.

Group	Participants, n	Mean (SD)	Cohen *d*	95% CI
Time 1			–0.07	–0.53 to 0.39
Intervention	34	8.18 (3.19)		
Control	38	8.39 (2.97)		
Time 2			–0.19	–0.69 to 0.31
Intervention	30	7.23 (3.35)		
Control	33	8.00 (2.73)		
Time 3			–0.30	–0.81 to 0.22
Intervention	26	6.15 (3.71)		
Control	34	7.53 (3.09)		
Time 4			–0.14	–0.64 to 0.37
Intervention	27	6.52 (3.65)		
Control	34	6.97 (2.97)		
Time 5			–0.57	–1.07 to –0.06
Intervention	29	5.62 (3.53)		
Control	34	7.74 (3.12)		
Time 6			–0.57	–1.09 to –0.04
Intervention	26	4.62 (3.71)		
Control	33	6.91 (3.39)		
Time 7			–0.85	–1.38 to –0.32
Intervention	26	4.35 (2.37)		
Control	35	7.03 (3.55)		

Group Differences for the Secondary Variable

As shown in Table 3, the results showed that participants' CBTSS scores increased postintervention and follow-up. The intervention group showed a significant improvement in CBTSS's self-monitoring score only at postintervention compared with the control group ($t_{120.7526}$=2.672, *P*=.01). The effect size at postintervention was moderate (Cohen *d*=0.68, 95% CI 0.17 to 1.18). However, the intervention participants did not significantly improve self-monitoring cognitive skills than control participants at follow-up ($t_{121.5475}$=1.947, *P*=.05, Cohen *d*=0.50, 95% CI −0.05 to 1.02).

XSL•FO
RenderX

Table . Results of the Cognitive Behavioral Therapy Skills Scale scores.

Group	Time 1				Time 2				Time 3			
	Participants, n	Mean (SD)	Cohen *d*	95% CI	Participants, n	Mean (SD)	Cohen *d*	95% CI	Participants, n	Mean (SD)	Cohen *d*	95% CI
Cognitive restructuring			–0.08	–0.54 to 0.38			0.42	–0.08 to 0.92			0.42	–0.09 to 0.93
Intervention	34	6.18 (4.37)			29	8.79 (4.69)			26	8.92 (4.22)		
Control	38	6.50 (3.52)			34	7.26 (4.21)			35	7.43 (4.19)		
Behavioral activation			0.08	–0.39 to 0.54			0.49	–0.02 to 0.99			0.32	–0.19 to 0.83
Intervention	34	4.85 (3.17)			29	7.07 (3.70)			26	7.69 (3.67)		
Control	38	4.58 (3.75)			34	4.71 (3.91)			35	6.09 (4.28)		
Assertiveness communication			0.09	–0.37 to 0.55			0.23	–0.27 to 0.72			0.25	–0.26 to 0.76
Intervention	34	9.15 (3.65)			29	10.21 (3.52)			26	10.92 (3.38)		
Control	38	8.84 (3.37)			34	9.35 (3.61)			35	9.80 (3.01)		
Self-monitoring			–0.13	–0.59 to 0.34			0.68	0.17 to 1.18			0.50	–0.05 to 1.02
Intervention	34	7.35 (4.55)			29	11.97 (4.85)			26	11.81 (4.17)		
Control	38	7.89 (4.09)			34	8.76 (4.06)			35	9.51 (5.06)		
Problem solving			0.24	–0.23 to 0.70			0.34	–0.16 to 0.83			0.48	–0.03 to 1.00
Intervention	34	6.65 (3.99)			29	8.83 (3.86)			26	9.81 (4.25)		
Control	38	5.76 (3.50)			34	6.56 (3.74)			35	6.86 (4.07)		

App Engagement Among the Intervention Group

As shown in Table 4, the average degree of satisfaction measured by the iOSDMH among users of the INTELLECT app was notably high. Among the 29 intervention participants, 16 - 24 (55% - 83%) participants answered “agree” or “relatively agree” on each item of acceptability (n=16 - 24, 55% - 83%), appropriateness (n=19 - 24, 66% - 83%), and feasibility (n=7 - 23, 24% - 79%). The highest proportions of positive responses in each of the 3 aspects were “ this program is acceptable for me”(24 out of 29 participants, 83%) in acceptability, “the content of the program is appropriate” (24 out of 29 participants, 83%) in appropriateness, and “the length of 1 content is implementable; the program is easy to understand” (23 out of 29 participants, 79%) in feasibility, respectively. Regarding overall satisfaction with INTELLECT app, 23 out of 29 participants (79%) answered “agree” or “relatively agree.” Regarding harms, 2 - 8 (7% - 28%) participants answered “agree” or “relatively agree,” except for “excessive pressure on learning this program regularly” (12 out of 29 participants, 41%). Over the 4-week period, the “Wellbeing Check-in” module was the most frequently used, followed by “Rescue Sessions” and “Guided Journals.” Notably, the “Learning Path” component required the most time investment from users, likely explaining the discrepancy between its number of initial engagements and lower completion rate.

XSL•FO
RenderX

Table . Results of the INTELLECT app engagement.

	Mean (SD)	Zeroth quartile	First quartile	Median	Third quartile	Fourth quartile
iOSDMH[a]						
Acceptability	8.59 (2.40)	3.00	6.50	9.00	10.00	12.00
Appropriateness	11.59 (2.61)	4.00	10.50	12.00	13.00	16.00
Feasibility	16.52 (2.85)	10.00	15.00	17.00	18.00	21.00
Satisfaction	2.86 (0.74)	1.00	3.00	3.00	3.00	4.00
Harms	8.52 (3.30)	5.00	6.00	8.00	10.00	18.00
Used components						
Starting learning paths	5.19 (3.03)	2	3	4	5	13
Completed learning paths	3.63 (2.72)	1	2	3	4	11
Rescue sessions	7.67 (5.28)	0	4	6	11	20
Wellbeing check-in	15.11 (8.31)	1	9	15	21	33
Guided journals	7.19 (5.68)	1	2	6	11	22

[a]iOSDMH: Implementation Outcome Scales for Digital Mental Health.

Discussion

Principal Results

The primary aim of this study was to compare the efficacy of INTELLECT's CBT-based features in improving anxiety and depressive symptoms among Japanese employees experiencing mild, moderate, or severe levels of these symptoms, compared with a waitlist control group. In addition, the study examined whether these features resulted in significant improvements in CBT-related life skills and assessed the acceptability, feasibility, and satisfaction levels among users.

Our primary hypothesis was supported, with the intervention group showing significantly greater improvements in depression and anxiety levels at postintervention and 2-month follow-up. Effect sizes of these improvements at each timepoint were considerably larger than those reported in similar studies reviewed by a recent meta-analysis [35]. These findings add to the scant literature by showing that self-guided CBT features can be highly effective in reducing anxiety and depressive symptoms among clinically at-risk working adults in Japan. This aligns with previous research showing comparable efficacy of brief CBT-based mobile features, or computerized CBT in alleviating depressive symptoms among university students experiencing baseline depressive symptoms in Western societies [36-38].

Our secondary hypothesis was partially supported. The self-guided CBT features also demonstrated significantly greater improvements in the self-monitoring aspect of cognitive behavioral skills than the waitlist control group, as assessed through the CBTSS. The effect size of this improvement at postintervention was moderate (Cohen d=0.68, 95% CI 0.17 to 1.18). Self-monitoring questions are prominently featured throughout INTELLECT's self-guided features. For example, thought records are embedded within all of INTELLECT's "Learning Paths," "Rescue Sessions", and "Guided Journals," consistently prompting each participant to analyze their own thought patterns. It is probable that the decrease in anxiety and depressive symptoms among our participants was related to the increase in self-monitoring. This technique alone has previously shown efficacy in alleviating anxiety and depressive symptoms in adults [39,40] and youths diagnosed with major depression [41]. However, our study did not conduct longitudinal mediation analyses to confirm this hypothesis. Future research could use more rigorous statistical methods, such as multilevel mediation analyses, to support this claim. In contrast, for other skills assessed by the CBTSS (ie, Behavioral Activation, Assertiveness Communication, Problem Solving), only specific self-guided features within INTELLECT were tailored to enhance these skills. For example, within INTELLECT's suite of Learning Paths, approximately half of them concentrate on modifying behavioral patterns, while only 3 focus on modifying communication styles. Researchers have also proposed that the anticipated type of improvements resulting from the interventions correlate with the type of content offered within the intervention [42]. Future researchers are encouraged to isolate various types of self-guided interventions and assess the efficacy of each intervention using a similar RCT design to ours. In comparison to the control group, the intervention group also did not exhibit significantly greater improvements in the cognitive restructuring aspect of CBT skills. There are a couple of possible reasons. First, cognitive restructuring has been identified as one of the most challenging CBT skills to impart, even among trained counselors and psychotherapists with many years of experience [43]. Furthermore, cognitive restructuring is widely recognized as a difficult skill to master independently, both within and outside of therapeutic contexts [44,45]. Further research is needed to explore suitable methodologies for effectively teaching cognitive restructuring skills within self-help modalities, enabling users to learn autonomously. Finally, it is important to note that participants in both the

XSL•FO
RenderX

intervention and control groups demonstrated significant improvements in CBT skills over time (Table 3). This particular control group may have masked the intervention effects on CBT skills. Future researchers can more effectively assess the efficacy of INTELLECT's self-guided CBT interventions by using a 3-armed RCT, wherein the efficacy of these CBT features is compared against those of an active control group and a waitlist control group.

Finally, it is noteworthy that a sizable percentage of our participants expressed satisfaction with using INTELLECT, rating it highly in terms of acceptability, appropriateness for improving mental health, and ease of understanding. The self-help features of INTELLECT, which use plain language, illustrations, videos, and audio narration to explain evidence-based content, may contribute to these positive implementation outcomes. Given the substantial effect sizes and user-friendly design of this app, we recommend broader dissemination of the INTELLECT app to alleviate anxiety and depressive symptoms among employees in need. Lin et al [25] highlighted that psychological empowerment and inspiration are key elements that can enhance the effectiveness of CBT-based mHealth interventions. Future studies must conduct a more detailed analysis of the role of user engagement in enhancing the effectiveness of CBT-based interventions.

Limitation

This study has 3 limitations. First, although INTELLECT's self-help interventions significantly reduced depressive symptoms during the study period and at the 2-month follow-up, it is uncertain if these improvements would persist beyond 2 months. Previous research on similar internet-based CBT interventions suggests a potential decline in efficacy over extended durations. Hence, it is important for future studies to evaluate the longitudinal effects of mobile-based CBT interventions. Second, one methodological limitation of this study concerns the potential predictability of the group assignment sequence. Group allocation was conducted using simple randomization based on a computer-generated sequence developed in advance by the researcher, without the use of blocking. Although stratification was applied for participants' history of depression management and structured psychotherapy, and the researcher had no access to other participant data, the fixed nature of the sequence may have allowed for some degree of predictability. Consequently, full allocation concealment could not be ensured. While there is no indication that this affected the integrity of the assignment process, this limitation should be taken into account when interpreting the internal validity of the study findings. Third, this intervention study primarily focused on assessing improvements in depressive symptoms and user engagement with the app, overlooking potential adverse effects associated with its use. A thorough examination of postintervention side effects is essential to ensure the app's safety across diverse user demographics. Last, while the INTELLECT app effectively enhanced self-monitoring capabilities, it did not significantly influence the development of other crucial CBT-related skills, such as cognitive restructuring, behavioral activation, communication, or problem-solving. Future studies may explore strategies for enhancing the app to facilitate a more comprehensive development of these skills, which could augment INTELLECT's efficacy in addressing anxiety and depressive symptoms.

Conclusions

This study provides the first evidence of the efficacy of the INTELLECT app, including all components, in reducing depressive symptoms among Japanese adult workers. The use of the INTELLECT app significantly enhanced self-monitoring skills, a key component of the observed improvement. However, the absence of postintervention effects on other CBT skills underscores the necessity for further investigation in these areas.

Acknowledgments

The authors wish to thank all study participants and the staff at the University of Human Environments, Japan, for their invaluable support throughout this project. Portions of this manuscript were prepared with the assistance of ChatGPT, a language model developed by OpenAI. Its use was limited to language editing and text refinement and did not influence the research design, data analysis, or conclusions. The author, a native Japanese speaker, used ChatGPT to enhance the clarity and fluency of the English text. ChatGPT was also used in the preparation of responses to reviewer comments to ensure clarity in communication.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors were involved in the conceptualization of the study. KY and RO developed the methodology, curated the data, and conducted the formal analysis. They also carried out the data collection and prepared the original draft of the manuscript. SHYT reviewed and edited all sections of the manuscript and finalized the draft with KY's approval. OS supervised the overall procedures of the study. All authors approved the final version of the manuscript.

Conflicts of Interest

Among the authors, RO and OS are employees of INTELLECT, the company that developed the app used in this study. SHYT is a research associate in INTELLECT's clinical team. Although KY has signed a nondisclosure agreement (NDA) with

XSL•FO
RenderX

INTELLECT, they did not receive any financial or other competing interests related to the conduct of this study. However, KY received honoraria and travel support for giving a lecture at a workshop organized by INTELLECT.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1). CONSORT: Consolidated Standards of Reporting Trails.
[PDF File, 3193 KB - mhealth_v13i1e60871_app1.pdf]

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet 2020 Oct 17;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
2. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 6;398(10312):1700-1712. [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
3. Midorikawa H, Tachikawa H, Taguchi T, et al. Demographics associated with stress, severe mental distress, and anxiety symptoms during the COVID-19 pandemic in Japan: nationwide cross-sectional web-based survey. JMIR Public Health Surveill 2021 Nov 22;7(11):e29970. [doi: 10.2196/29970] [Medline: 34653018]
4. Miyake F, Odgerel CO, Hino A, et al. Job stress and loneliness among desk workers during the COVID-19 pandemic in Japan: focus on remote working. Environ Health Prev Med 2022;27:33. [doi: 10.1265/ehpm.22-00107] [Medline: 35965100]
5. Ueda M, Nordström R, Matsubayashi T. Suicide and mental health during the COVID-19 pandemic in Japan. J Public Health (Oxf) 2022 Aug 25;44(3):541-548. [doi: 10.1093/pubmed/fdab113] [Medline: 33855451]
6. Matsubayashi T, Ishikawa Y, Ueda M. Economic crisis and mental health during the COVID-19 pandemic in Japan. J Affect Disord 2022 Jun 1;306(306):28-31. [doi: 10.1016/j.jad.2022.03.037] [Medline: 35306120]
7. Survey on mental health for workers and future issues. AIG Institute. URL: https://www-510.aig.co.jp/assets/documents/institute/insight/institute-insight-03-en.pdf [accessed 2024-03-20]
8. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: in review, in use, and into the future. Prof Psychol Res Pract ;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
9. Consumer health apps and digital health tools proliferate, improving quality and health outcomes for patients, says new report from IQVIA institute. IQVIA. URL: https://www.iqvia.com/newsroom/2021/07/consumer-health-apps-and-digital-health-tools-proliferate-improving-quality-and-health-outcomes-for [accessed 2024-03-20]
10. The growing value of digital health: evidence and impact on human health and the healthcare system institute report. : IQVIA; 2017 URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/the-growing-value-of-digital-health [accessed 2024-03-20]
11. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1(90):12. [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
12. Akobeng AK. Understanding randomised controlled trials. Arch Dis Child 2005 Aug;90(8):840-844. [doi: 10.1136/adc.2004.058222] [Medline: 16040885]
13. Huguet A, Rao S, McGrath PJ, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS ONE 2016;11(5):e0154248. [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
14. Shen N, Levitan MJ, Johnson A, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16. [doi: 10.2196/mhealth.3713] [Medline: 25689790]
15. Yamamoto K, Ito M, Sakata M, et al. Japanese version of the mobile app rating scale (MARS): development and validation. JMIR Mhealth Uhealth 2022 ;10(4):e33725. [doi: 10.2196/33725]
16. Health care app market report. : Fuller, Inc; 2022 URL: https://go.appa.pe/l/463132/2022-12-07/23dj54f/463132/1694579631kR8d4CRc/healthcare_app_marketreport_2022.pdf [accessed 2024-03-20]
17. Imamura K, Kawakami N, Furukawa TA, et al. Effects of an internet-based cognitive behavioral therapy (iCBT) program in manga format on improving subthreshold depressive symptoms among healthy workers: a randomized controlled trial. PLoS ONE 2014;9(5):e97167. [doi: 10.1371/journal.pone.0097167] [Medline: 24844530]
18. Yasukawa S, Tanaka T, Yamane K, et al. A chatbot to improve adherence to internet-based cognitive-behavioural therapy among workers with subthreshold depression: a randomised controlled trial. BMJ Ment Health 2024 Jan 10;27(1):e300881. [doi: 10.1136/bmjment-2023-300881] [Medline: 38199786]
19. Imamura K, Kawakami N, Furukawa TA, et al. Does internet-based cognitive behavioral therapy (iCBT) prevent major depressive episode for workers? a 12-month follow-up of a randomized controlled trial. Psychol Med 2015 Jul;45(9):1907-1917. [doi: 10.1017/S0033291714003006] [Medline: 25562115]
20. Linardon J, Gleeson J, Yap K, Murphy K, Brennan L. Meta-analysis of the effects of third-wave behavioural interventions on disordered eating and body image concerns: implications for eating disorder prevention. Cogn Behav Ther 2019 Jan 2;48(1):15-38. [doi: 10.1080/16506073.2018.1517389]

XSL•FO
RenderX

21. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
22. van Ballegooijen W, Cuijpers P, van Straten A, et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS ONE 2014;9(7):e100674. [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
23. Karyotaki E, Riper H, Twisk J, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
24. Tudor-Sfetea C, Rabee R, Najim M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions (preprint). JMIR Mhealth Uhealth. Preprint posted online on 2018. [doi: 10.2196/preprints.9405]
25. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective behavioral changes through a digital mHealth app: exploring the impact of hedonic well-being, psychological empowerment and inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024. [doi: 10.2196/10024] [Medline: 29907557]
26. Kosasih FR, Yee VTS, Toh SHY, Sündermann O. Efficacy of Intellect's self-guided anxiety and worry mobile health programme: a randomized controlled trial with an active control and a 2-week follow-up. PLOS Digit Health 2023 May;2(5):e0000095. [doi: 10.1371/journal.pdig.0000095] [Medline: 37224139]
27. Ong WY, Sündermann O. Efficacy of the mental health app "Intellect" to improve body image and self-compassion in young adults: a randomized controlled trial with a 4-week follow-up. JMIR Mhealth Uhealth 2022 Nov 21;10(11):e41800. [doi: 10.2196/41800] [Medline: 36409542]
28. Toh SHY, Tan JHY, Kosasih FR, Sündermann O. Efficacy of the mental health app intellect to reduce stress: randomized controlled trial with a 1-month follow-up. JMIR Form Res 2022 Dec 14;6(12):e40723. [doi: 10.2196/40723] [Medline: 36515984]
29. Yang Toh SH, Lee SC, Kosasih FR, Lim JW, Sündermann O. Preliminary effectiveness of an evidence-based mobile application to promote resilience among working adults in Singapore and Hong Kong: intensive longitudinal study. Digit Health 2023;9(9):20552076231178616. [doi: 10.1177/20552076231178616] [Medline: 37274370]
30. Toh SHY, Lee SC, Sündermann O. Mobile behavioral health coaching as a preventive intervention for occupational public health: retrospective longitudinal study. JMIR Form Res 2023 Oct 20;7:e45678. [doi: 10.2196/45678] [Medline: 37862086]
31. Doi S, Ito M, Takebayashi Y, Muramatsu K, Horikoshi M. Factorial validity and invariance of the patient health questionnaire (PHQ)-9 among clinical and non-clinical populations. PLoS ONE ;13(7):e0199235. [doi: 10.1371/journal.pone.0199235]
32. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
33. Sakata M, Toyomoto R, Yoshida K, et al. Development and validation of the cognitive behavioural therapy skills scale among college students. Evid Based Ment Health 2021 May;24(2):70-76. [doi: 10.1136/ebmental-2020-300217] [Medline: 33402377]
34. Sasaki N, Obikane E, Vedanthan R, et al. Implementation outcome scales for digital mental health (iOSDMH): scale development and cross-sectional study. JMIR Form Res 2021 Nov 23;5(11):e24332. [doi: 10.2196/24332] [Medline: 34817391]
35. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
36. Mantani A, Kato T, Furukawa TA, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J Med Internet Res 2017 Nov 3;19(11):e373. [doi: 10.2196/jmir.8602] [Medline: 29101095]
37. Bhuyan SS, Lu N, Chandak A, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
38. Six SG, Byrne KA, Aly H, Harris MW. The effect of mental health app customization on depressive symptoms in college students: randomized controlled trial. JMIR Ment Health 2022 Aug 9;9(8):e39516. [doi: 10.2196/39516] [Medline: 35943788]
39. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health 2016 Nov 23;3(4):e49. [doi: 10.2196/mental.6202] [Medline: 27881358]
40. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
41. Hetrick SE, Robinson J, Burge E, et al. Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. JMIR Ment Health 2018 Jan 23;5(1):e9. [doi: 10.2196/mental.9041] [Medline: 29362208]

XSL•FO
RenderX

42. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? a review using techniques from critical interpretive synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
43. Stark KD, Arora P, Funk CL. Training school psychologists to conduct evidence‐based treatments for depression. Psychol Sch 2011 Mar;48(3):272-282. [doi: 10.1002/pits.20551]
44. Willner P, Brace N, Phillips J. Assessment of anger coping skills in individuals with intellectual disabilities. J Intellect Disabil Res 2005 May;49(Pt 5):329-339. [doi: 10.1111/j.1365-2788.2005.00668.x] [Medline: 15817050]
45. Neal-Barnett A, Stadulis R, Ellzey D, et al. Evaluation of the effectiveness of a musical cognitive restructuring app for black inner-city girls: survey, usage, and focus group evaluation. JMIR Mhealth Uhealth 2019 Jun 27;7(6):e11310. [doi: 10.2196/11310] [Medline: 31188130]

Abbreviations

AO: assessment-only
CBT: cognitive behavioral therapy
CBTSS: Cognitive Behavioral Therapy Skills Scale
iOSDMH: Implementation Outcome Scales for Digital Mental Health
LMM: linear mixed model
mHealth: mobile health
PHQ: Patient Health Questionnaire
RCT: randomized controlled trial
UMIN: University Hospital Medical Information Network

Edited by L Buis; submitted 29.05.24; peer-reviewed by A Eisingerich, N Sasaki; revised version received 11.04.25; accepted 11.04.25; published 24.06.25.

Please cite as:
Yokomitsu K, Oimatsu R, Toh SHY, Sündermann O
Assessing the Efficacy of the INTELLECT Cognitive Behavioral Therapy Mobile App for Anxiety and Depressive Symptoms Among At-Risk Japanese Employees: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e60871
URL: https://mhealth.jmir.org/2025/1/e60871
doi:10.2196/60871

© Kengo Yokomitsu, Riki Oimatsu, Sean Han Yang Toh, Oliver Sündermann. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effect of a Cognitive Behavioral Therapy–Based AI Chatbot on Depression and Loneliness in Chinese University Students: Randomized Controlled Trial With Financial Stress Moderation

Yahui Wang[1], MSc; Xuhong Li[1], MSW; Qiaochu Zhang[1], PhD; Dannii Yeung[1], PhD; Yihan Wu[2], PhD

[1]Department of Social and Behavioural Sciences, City University of Hong Kong, Hong Kong, China

[2]Department of Sociology and Social Security, Ginling College, Nanjing Normal University, 1 Wenyuan Rd, Qixia District, Nanjing, Jiangsu, China

Corresponding Author:
Yihan Wu, PhD
Department of Sociology and Social Security, Ginling College, Nanjing Normal University, 1 Wenyuan Rd, Qixia District, Nanjing, Jiangsu, China

Abstract

Background: Mental health concerns are prevalent among university students, with financial stress further compounding these issues. While cognitive behavioral therapy (CBT) is effective for these conditions, its delivery through artificial intelligence (AI) chatbots represents a promising approach, especially in non-Western contexts.

Objective: This study aims to investigate the efficacy of a culturally adapted, CBT-based AI chatbot for improving the well-being of Chinese university students and to examine whether financial stress moderates its effectiveness.

Methods: In this randomized controlled trial, 100 university students (mean age 20.8, SD 2.2 years; 62/100, 62% female) were allocated to either an intervention (n=50) or a waitlist control group (n=50). The intervention group interacted with a CBT-based AI chatbot for 7 consecutive days. Depression (Center for Epidemiologic Studies Depression Scale), anxiety (Generalized Anxiety Disorder-7 scale), and loneliness (UCLA Loneliness Scale) were assessed at baseline, day 3, and day 7. Financial stress was measured using the Psychological Inventory of Financial Scarcity.

Results: Significant group×time interactions were found for depression ($F_{2,196}$=8.63; P<.001; η^2_p=.08) and loneliness ($F_{2,196}$=5.57; P=.004; η^2_p=.05), but not for anxiety ($F_{2,196}$=1.31; P=.27; η^2_p=.01). Post hoc comparisons showed significant reductions in both depression (t=3.85; P<.001) and loneliness (t=4.28; P<.001) from baseline to postintervention in the intervention group, with corresponding effect sizes of Cohen d=0.71 (95% CI 0.30 - 1.12) and Cohen d=0.60 (95% CI 0.20 - 1.00), respectively. No significant changes were observed in the waitlist control group. Exploratory subgroup analyses revealed that participants with high financial stress demonstrated significantly greater improvements in depression ($F_{2,52}$=11.56; P<.001; η^2_p=.31) and loneliness ($F_{2,52}$=11.18; P<.001; η^2_p=.30) compared to those with low financial stress.

Conclusions: The culturally adapted, CBT-based AI chatbot effectively reduced depression and loneliness in Chinese university students, with stronger effects among those experiencing high financial stress. These findings highlight the potential of AI-driven interventions to provide accessible mental health support, particularly for financially stressed students.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2500100868; https://www.chictr.org.cn/showproj.html?proj=253905

(JMIR Mhealth Uhealth 2025;13:e63806) doi:10.2196/63806

KEYWORDS

chatbot; cognitive behavioral therapy; financial stress; university students; artificial intelligence; depression; loneliness; randomized controlled trial

Introduction

Mental Health Challenges Among University Students and Financial Stress

University students face numerous stressors that can impact their mental health and well-being. Depression, anxiety, and loneliness are prevalent concerns in this population [1], with global estimates indicating that over 30% of tertiary students experience clinically significant symptoms [2]. Compared to primary and secondary school students, university students are at a higher risk of developing mental health issues due to the unique challenges they face, such as increased academic pressure, financial burden, and the transition to adulthood [3]. Existing studies consistently illustrate that anxiety and depression lead to a variety of negative outcomes among university students, such as poor academic performance, reduced social networks, low quality of life, and sleep problems [4,5].

XSL•FO
RenderX

In the postpandemic era, financial stress has significantly worsened the psychological health of university students in various countries, including the United States, United Kingdom, Germany, and China [6-8]. The rising costs of higher education, coupled with limited financial resources, place a significant burden on many students, leading to increased vulnerability to mental health problems [9]. Research consistently suggests that financial hardship is associated with higher levels of stress, anxiety, and depression among university students [10,11]. This relationship appears particularly pronounced in the postpandemic context, where economic uncertainties have intensified financial pressures on students [8]. Addressing the impact of financial stress on university students' mental health is, therefore, increasingly crucial.

Underuse of Mental Health Services

Despite the high prevalence of mental health issues, most university students seldom seek professional mental health services provided by their institutions or external providers [4,12]. A systematic review and meta-analysis study reported that 35% (95% CI 22% - 50%) of university students used mental health services, with individual study estimates ranging from 13.7% to 68.6% [13]. This substantial heterogeneity reflects differences in health care systems, cultural contexts, service definitions, and accessibility of services across countries and institutions. Notably, the same review also examined specific subgroups and service types, finding that overall outpatient use rates were particularly low among students in non-Western contexts, such as China (5.1%) and Bangladesh (7.1%), compared to the average rates in Western settings [13]. This East-West disparity highlights the particular challenges faced by students in Asian contexts.

The underuse of mental health services can be attributed to various barriers, such as stigma, time constraints, unaffordability, and limited accessibility [14-16]. These barriers are often more pronounced in collectivistic cultures like China, where mental health stigma is particularly strong and seeking help may be viewed as bringing shame to one's family [17]. The combination of high need and low service use—especially pronounced in Chinese contexts—underscores the importance of developing innovative, culturally sensitive interventions that can overcome these substantial barriers to care.

Digital Interventions and AI Chatbots as a Promising Solution

Digital mental health interventions, such as mobile apps and web-based self-help programs, show promise in overcoming treatment and access barriers [18]. These interventions can meet the high demand for campus counseling services and reduce access barriers by providing convenient, anonymous, and cost-effective support to students. A meta-analysis by Harrer et al [19] found that internet-based interventions were effective in reducing symptoms of depression, anxiety, and stress among university students, with effect sizes being comparable to those of face-to-face interventions. However, many digital interventions still face challenges with user engagement, personalization, and cultural relevance [20-22].

Recent advancements in artificial intelligence (AI) and natural language processing have paved the way for the development of AI-powered chatbots that can deliver psychological support through interactive conversations [23]. These AI-powered chatbots offer several advantages over traditional digital interventions: they provide 24/7 access to support, reduce stigma associated with help-seeking, and offer personalized, engaging experiences [24]. Emerging research supports their efficacy; for example, Fitzpatrick et al [25] found that a CBT-based conversational agent effectively reduced symptoms of depression and anxiety among young adults, while D'Alfonso [26] reported promising results using an AI chatbot for university students experiencing financial difficulties.

AI-powered chatbots can be particularly beneficial for university students, who are digital natives and may prefer seeking help through technology-based interventions [18]. The COVID-19 pandemic has further highlighted the importance of digital mental health interventions, as social distancing measures have limited access to in-person mental health services [27]. The pandemic has led to widespread social isolation and disruption of daily routines, which may exacerbate the risk of mental health problems among students [28]. Chatbots can provide a convenient and accessible source of support for students who may have smaller social networks or feel hesitant to share their problems with others due to the pandemic. Furthermore, the anonymous nature of chatbot interactions may encourage students who are hesitant to share their problems with parents, friends, or teachers to seek help [29]. Chatbot interventions can provide personalized support through interactive conversations, which may include psychoeducation, reflection prompts, in-session exercises, and personalized feedback based on users' responses [25]. These interactive features can enhance user engagement, promote the acquisition and application of coping skills, and provide a sense of social support, which may be particularly valuable for students experiencing social isolation or loneliness [30].

Cultural Adaptation of CBT: A Necessity, Not an Option

Cognitive-behavioral therapy (CBT) has demonstrated effectiveness for treating common mental health problems among university students [31]. However, cultural context significantly influences how mental health problems are experienced, expressed, and addressed, making cultural adaptation of psychological interventions not merely beneficial but essential for maximizing effectiveness.

The need for cultural adaptation is particularly clear in Chinese contexts, where mental health experiences differ markedly from Western norms [17]. In Chinese culture, mental health problems are often viewed through a collectivistic lens where individual psychological issues can threaten family reputation ("face"), creating a "courtesy stigma" that extends beyond the individual to their entire family unit, which can intensify stigma and discourage help-seeking [32]. This family-oriented stigma creates stronger barriers to help-seeking compared to more individualistic Western societies. In addition, traditional Chinese values emphasize emotional restraint and self-reliance and the importance of maintaining social harmony, which may conflict

XSL•FO
RenderX

with Western therapeutic approaches that encourage emotional expression and seeking external help [33]. Mental health issues in Chinese contexts are often conceptualized in somatic terms rather than psychological ones, with cultural idioms of distress frequently manifesting as physical complaints [34].

These cultural differences are not merely theoretical but have demonstrated practical implications for intervention effectiveness. Several studies have found that standard Western CBT interventions may have lower engagement and effectiveness when applied directly to Chinese populations without cultural adaptation [35,36]. For example, Hwang et al [36] demonstrated that culturally adapted CBT for Chinese Americans showed not only significantly better clinical outcomes but also substantially lower dropout rates compared to standard CBT. Similarly, Choi et al [37] found that cultural adaptation of CBT significantly improved treatment adherence and outcomes among Chinese Australians.

A recent meta-analysis by Li et al [38] examining culturally adapted interventions for common mental disorders in people of Chinese descent found these adaptations beneficial, though the efficacy did not differ significantly between culturally modified interventions and culturally specific interventions. Rathod et al [39] also reviewed culturally adapted psychological interventions and found varying levels of effectiveness across different cultural groups and adaptation approaches. These highlight that thoughtful adaptation, whether modifying existing approaches or developing culturally specific ones, can improve outcomes when grounded in an understanding of core cultural concepts and values while maintaining essential therapeutic mechanisms.

Research Gaps and This Study

Despite the promising potential of AI-powered chatbots and the importance of cultural adaptation, significant research gaps remain. First, most studies evaluating digital mental health interventions have been conducted with Western samples, limiting generalizability to Chinese university students [40,41]. Second, as revealed by a systematic review of digital mental health in China [42], Chinese studies rarely adopt co-production approaches in developing digital health technologies and often demonstrate less rigorous methodological quality. Third, few studies have examined whether digital mental health interventions are equally effective across different levels of financial stress, which is an increasingly important consideration given growing economic pressures on students.

Most critically, despite the theoretical promise of combining AI chatbot technology with culturally adapted CBT content, very few randomized controlled trials have evaluated such interventions specifically for Chinese university students experiencing financial stress. This represents a significant missed opportunity to address an urgent public health need through innovative, scalable approaches that could potentially overcome cultural and structural barriers to care.

This study addresses these gaps by evaluating the efficacy of a custom-developed, culturally adapted, CBT-based AI chatbot in reducing depression, anxiety, and loneliness among Chinese university students, with particular attention to the moderating role of financial stress. Using a randomized controlled trial design, we tested two primary hypotheses:

- Hypothesis 1: Students who engaged with the culturally adapted chatbot would show greater improvements in depression, anxiety, and loneliness compared to those in the waitlist control group.
- Hypothesis 2: The intervention's effectiveness would be moderated by participants' baseline levels of financial stress, with those experiencing higher financial stress benefiting more from the intervention compared to those with lower financial stress.

By examining both the overall efficacy of the chatbot and the moderating effect of financial stress, this study aims to provide valuable insights into how AI-driven, culturally adapted interventions might be optimally deployed to improve the well-being of vulnerable student populations.

Methods

Study Design and Participants

This study used a parallel-group, randomized controlled trial design. Participants were recruited from multiple public universities through email advertisements and campus flyers. The inclusion criteria were (1) being a currently enrolled full-time undergraduate or graduate student aged 18 - 24 years, (2) owning an Android smartphone, and (3) being willing to engage with the chatbot daily for 7 days. The exclusion criteria included (1) current engagement in any type of psychotherapy, (2) having a diagnosis of any severe mental illness in history (eg, schizophrenia, bipolar disorder, etc), and (3) current use of psychiatric medications. Initial screening was conducted via online questionnaire (Wenjuanxing), and eligible participants were invited for further screening interviews conducted by trained research staff to confirm eligibility and obtain informed consent.

Ethical Considerations

The study received ethical approval from the Institutional Review Board of Nanjing Normal University (NNU202301018). All participants provided written informed consent before enrollment. The consent process explicitly detailed the study purpose, procedures, potential risks and benefits, data handling protocols, and participants' right to withdraw at any time without penalty. Data were anonymized using participant identification codes, and all information was stored on encrypted, password-protected servers compliant with institutional data security standards. Participants who completed the study received monetary compensation for their time and contribution. The study was registered with the Chinese Clinical Trial Registry (ChiCTR2500100868).

Sample Size Determination and Randomization

A priori power analysis using G*Power 3.1 (Heinrich-Heine-Universität Düsseldorf) was conducted to determine the necessary sample size for detecting a small effect size (Cohen f=0.2) with 80% power at a significance level of α=.05. The anticipated effect size was based on meta-analytic findings by Firth et al [43], who reported effect sizes for

XSL•FO
RenderX

smartphone-based mental health interventions targeting depression ranging from g=0.22 to g=0.56, depending on control condition type. We selected f=0.2 as a conservative estimate appropriate for our brief 7-day intervention with waitlist control, representing a clinically meaningful effect size within the empirically derived range.

The analysis was based on a 2 (group: intervention vs control) × 3 (time: T1, T2, T3) mixed ANOVA design, considering the interaction between within-subjects and between-subjects factors. The power analysis indicated that a total sample size of 42 participants (21 per group) was required. Adjusting for an anticipated 31% attrition rate (the weighted average for internet-based treatment programs according to a systematic review by Melville et al [44]), the adjusted total sample size was 60. However, to ensure sufficient power and account for potential dropouts, we aimed for a larger sample of 100 participants, which would maintain adequate statistical power even with higher-than-expected attrition. Detailed power analysis parameters and post hoc calculations for moderation analyses are provided in Multimedia Appendix 1.

A total of 120 students were recruited, of which 100 met the inclusion criteria and were enrolled in the study. Participants were then randomly allocated in a 1:1 ratio to either the intervention group (n=50) or the waitlist control group (n=50) using a computer-generated block randomization sequence created by an independent statistician not involved in recruitment. Group allocation was concealed through a digital randomization system managed by a researcher not involved in participant recruitment or assessment. The system revealed allocations only after participants completed the baseline assessments, ensuring that neither participants nor researchers could predict group assignments.

Intervention

CBT-Based AI Chatbot Development

The CBT-based AI chatbot “Psy-Bot” was developed through a systematic, multistage process by our interdisciplinary research team consisting of clinical psychologists, AI experts, cultural adaptation specialists, and user experience designers. The development followed established frameworks for digital mental health interventions [45] and cultural adaptation of psychological treatments [36,46,47].

Content Development and Theoretical Foundation

The content was grounded in empirically supported CBT principles [48], with specific adaptation for university students and the Chinese cultural context. The development process included 3 key phases.

First, a comprehensive review of CBT protocols effective for depression, anxiety, and loneliness in university students was conducted, with particular focus on studies concerning financial stress. Key therapeutic components were identified from meta-analyses [49] and clinical practice guidelines.

Second, core therapeutic components were implemented based on Beck’s cognitive model [48] and included cognitive restructuring (identifying and challenging negative automatic thoughts), behavioral activation (scheduling pleasant activities and reducing avoidance), problem-solving training (systematic approach to addressing stressors), mindfulness exercises (present-focused awareness training), and social skills development (communication and assertiveness training).

Third, content was systematically adapted using Bernal’s framework for cultural adaptation [46], which addresses 8 dimensions: language, persons, metaphors, content, concepts, goals, methods, and context. This process involved consultations with experts in Chinese mental health (n=3) and focus groups with Chinese university students (n=10), review by bilingual psychologists to ensure linguistic and conceptual equivalence, and pilot-testing with the target population.

The 7-Day Module Structure

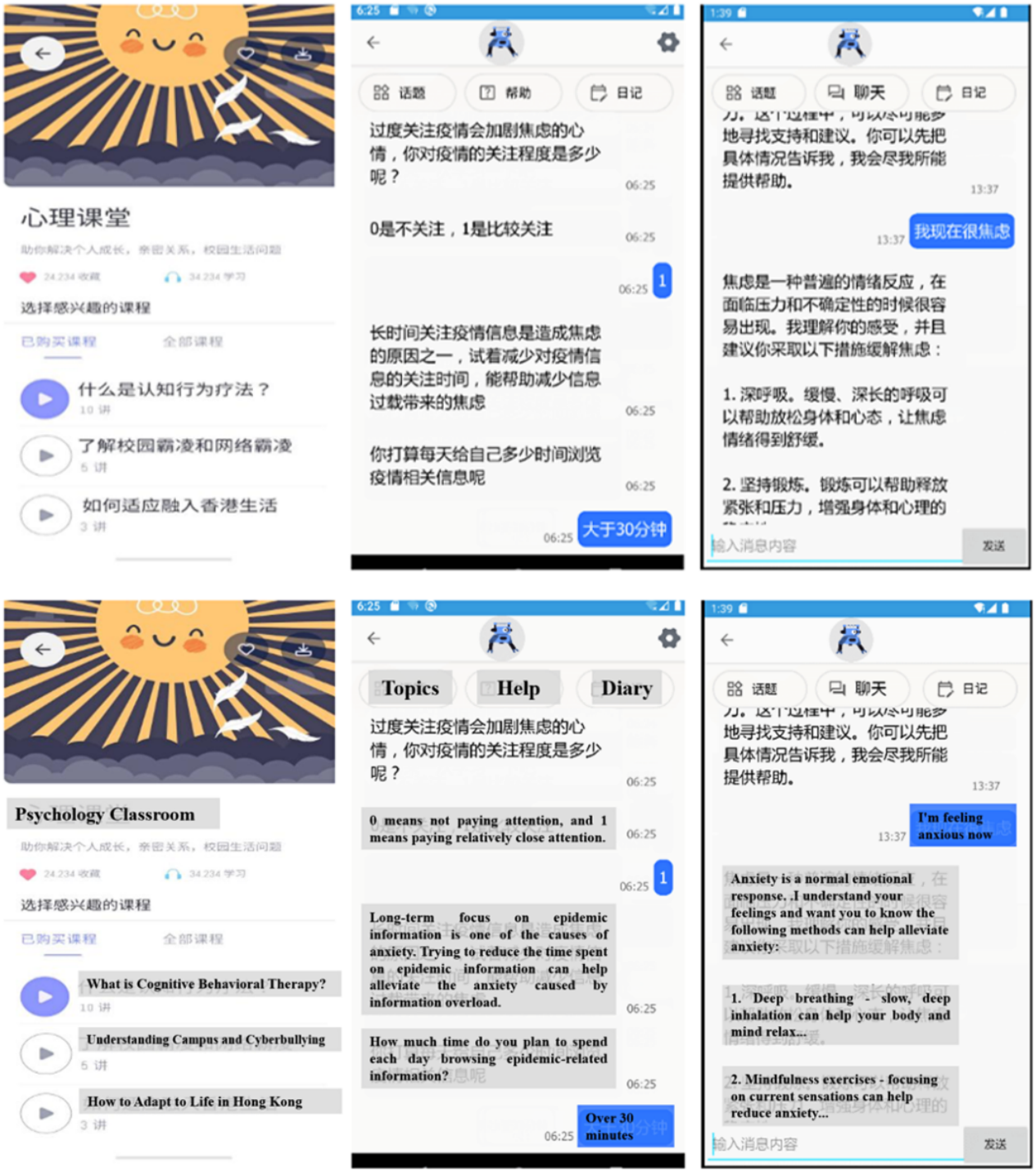
The chatbot content was structured as a progressive seven-day CBT-based intervention, with each daily module building on previously learned skills:

Day 1: Introduction to CBT principles established the cognitive-behavioral model and foundational skills for identifying thoughts, emotions, and behaviors.

Day 2: Coping with campus-related stressors addressed interpersonal challenges including campus bullying (as shown in Figure 1A), using cognitive restructuring and assertiveness techniques.

XSL•FO
RenderX

Figure 1. App interface screenshots with English translations. (**A**) Module selection: The app provided various modules, including an introduction to CBT, coping with campus bullying, and adapting to new environments. (**B**) Chat interface: Users could select a topic, initiate a conversation with the chatbot, and access a personal diary feature. The displayed dialog demonstrates how the chatbot helps users relieve anxiety by explaining its nature and providing coping methods. (**C**) Pandemic-related anxiety: This section features an interactive conversation about pandemic-related anxiety, where the chatbot provides personalized responses and guidance based on the user's input. Chinese interface elements are shown with English translations below. CBT: cognitive behavioral therapy.



Day 3: Adapting to new environments helped students navigate transitions and adjustment difficulties (specifically displayed in the app interface; see Figure 1A), using behavioral experiments and gradual exposure.

Day 4: Managing anxiety incorporated both general anxiety management and specific contexts such as pandemic-related anxiety (as shown in Figure 1C), using cognitive reframing and relaxation strategies.

Day 5: Financial stress management addressed cognitive patterns related to financial scarcity, problem-solving for financial challenges, and adaptive coping strategies, developed based on research on financial stress interventions.

XSL•FO
RenderX

Day 6: Emotion regulation and mood improvement focused on behavioral activation and mindfulness exercises to address depressive symptoms and enhance positive emotions.

Day 7: Building resilience and relapse prevention, consolidated learning and established maintenance strategies for continued well-being after the intervention period.

Each module was designed to be completed in approximately 20 - 30 minutes. The complete module content and structure are presented in Table S1 in Multimedia Appendix 2.

Cultural Adaptation

As noted above, all content was systematically adapted following Bernal et al's [46] empirically validated framework for cultural adaptation. This framework has been successfully applied in previous cultural adaptations of psychological interventions for Asian populations [35,36].

Key cultural adaptations included using indirect communication patterns aligned with Chinese preferences for emotional expression, incorporating collectivistic values emphasizing family harmony, addressing "face" concerns related to seeking help and financial difficulties, and integrating culturally resonant metaphors and Chinese idioms. Expert ratings indicated strong cultural appropriateness (mean rating 4.1/5 for cultural relevance). Detailed examples of cultural adaptations across all 8 dimensions of Bernal and colleague's [46] framework are provided in Table S2 in Multimedia Appendix 2.

Technical Architecture and User Experience

The chatbot used a hybrid approach to response generation, combining rule-based responses with natural language processing to ensure clinical validity and safety. The system architecture included a dialog management system based on the RASA open-source framework (Rasa Technologies Inc) [50], with custom extensions for therapeutic dialog; natural language understanding implemented using bidirectional encoder representations from transformers–based models [51], fine-tuned on therapeutic conversations in Chinese; response generation combining pre-scripted therapeutic responses (70%), template-based dynamic responses (20%), and contextually generated responses (10%); and safety monitoring with automated detection of severe distress or risk expressions.

As shown in Figure 1A, the app interface allowed users to browse modules, mark favorites, and track popularity. The chatbot delivered content through conversational interactions (see Figure 1B), explaining psychological concepts, providing coping strategies, and engaging users in interactive therapeutic exercises. For specific concerns like pandemic-related anxiety, the chatbot provided tailored guidance based on individual expressions of distress (see Figure 1C).

The app incorporated gamification elements, such as progress tracking and virtual rewards, to enhance user engagement. It also sent gentle reminders if a session was missed. User interactions, including session completion and duration, were automatically logged in a secure Firebase database.

Validation Process

The chatbot underwent a rigorous 3-stage validation process:

1. Content validation: A total of 3 experts in CBT, digital mental health, and Chinese cultural psychology independently rated each module using standardized assessment forms. Mean ratings showed strong adherence to CBT principles (mean 4.2, SD 0.15, out of 5), cultural relevance (mean 4.1, SD 0.14, out of 5), and anticipated engagement (mean 3.9, SD 0.28, out of 5).
2. Technical validation: Comprehensive testing of system functionality, response accuracy, data security, and user experience was conducted by both technical staff and independent testers. This included simulating dialog turns and stress-testing the system under various input conditions.
3. Pilot-testing: A preliminary pilot test with 10 university students (separate from the main study sample) was conducted to gather feedback through structured questionnaires. The System Usability Scale [52] yielded a mean score of 82.3 (SD 3.83) out of 100 (range 76 - 89), indicating excellent usability according to established benchmarks [53]. User experience was further assessed using a modified version of the Mobile Application Rating Scale [54], which showed good comprehension of therapeutic content (mean rating 4.4, SD 0.20, out of 5; range 3.9 - 4.7).

Based on validation feedback, refinements were made to the interface and conversational flow before the main trial. Detailed validation metrics are presented in Table S3 in Multimedia Appendix 2, and sample dialog exchanges demonstrating the application of CBT techniques with cultural adaptations are provided in Table S4 in Multimedia Appendix 2.

Module Delivery and User Interaction

Participants in the intervention group accessed the chatbot through a mobile app installed on their Android smartphones via an APK file. The app featured a user-friendly interface with a chat window, notification settings, and progress tracking (see Figure 1). The chatbot delivered the modules through a conversational interface, using a combination of predefined scripts and dynamic response generation. It also incorporated multimedia elements, such as images and audio clips, to enhance user engagement.

The app was designed to unlock 1 module per day to ensure that participants followed the intended learning schedule. Participants were encouraged to complete 1 module per day, with the app sending gentle reminders if a session was missed. All materials used in the chatbot were either original content developed by the research team or sourced from public domain resources with proper attribution.

Waitlist Control Group

Participants in the waitlist control group did not receive access to the chatbot during the 7-day study period. They were informed that they would receive access to the app after completing the postintervention assessment. To minimize attrition, the waitlist control group participants received daily check-in messages expressing appreciation for their continued participation and reminding them of the upcoming assessments. The participants in the waitlist control group were asked to complete the same set of questionnaires as the intervention

XSL•FO
RenderX

group at the 3 time points of assessment (baseline, midintervention, and postintervention).

Measures

Online self-report questionnaires were administered at 3 time points: baseline (T1, before the intervention on day 1), midintervention at day 3 (T2, day 3), and postintervention at day 7 (T3, day 7). The primary outcomes were as follows:

- Depression: Measured using the Chinese version of the 10-item Center for Epidemiologic Studies Depression Scale (CES-D-10) [55], a 10-item scale assessing the frequency of depressive symptoms over the past week on a 4-point Likert scale. Total scores range from 0 to 30, with higher scores indicating more severe depressive symptoms. In this study, the Cronbach α for the CES-D-10 were 0.89, 0.89, and 0.91 at T1, T2, and T3, respectively.
- Anxiety: Assessed using the Generalized Anxiety Disorder-7 scale (GAD-7) [56]), a 7-item measure of anxiety symptoms over the past 2 weeks, rated on a 4-point Likert scale. Higher scores indicate greater levels of anxiety (range: 0 - 21). In this study, the Cronbach α was 0.92 at T1, 0.93 at T2, and 0.94 at T3.
- Loneliness: Measured using the UCLA Loneliness Scale (version 3) [57], a 20-item scale assessing subjective feelings of loneliness and social isolation on a 4-point Likert scale. Higher scores indicate greater loneliness (range: 20 - 80). In this study, the Cronbach α was 0.93 at T1, 0.94 at T2, and 0.95 at T3.
- Financial stress: Participants also completed the Psychological Inventory of Financial Scarcity (PIFS) [58] at baseline, which is a 12-item scale measuring subjective perceptions of economic hardship on a 7-point Likert scale. Higher scores indicate greater financial stress. In this study, Cronbach α was 0.93 at T1, 0.94 at T2, and 0.95 at T3.

Statistical Analysis

Quantitative data were analyzed using R software (version 4.3.0; R Core Team). Before the main analyses, data were screened for outliers and violations of statistical assumptions, with no significant issues detected. Our screening process involved visual inspection of boxplots and calculation of *z* scores, with values exceeding ±3.29 considered potential outliers. No extreme outliers requiring special handling were identified in the dataset. Missing data (attrition rate 11%, 11/100) were handled using the last observation carried forward (LOCF) method, maintaining the intention-to-treat principle while considering the study's short duration (7 d).

Baseline group differences in demographic characteristics and outcome measures were examined using independent *t* tests for continuous variables and χ^2 tests for categorical variables. To assess intervention effects, 2 (group: intervention vs control) × 3 (time: T1, T2, T3) mixed ANOVAs were conducted for each outcome variable, with group as the between-subjects factor and time as the within-subjects factor. Significant group×time interactions were probed with pairwise comparisons using Bonferroni correction to control for multiple testing. Effect sizes were reported using partial eta-squared (η^2_p) to provide a measure of the magnitude of the intervention effects. For outcomes with significant group×time interactions, Bonferroni-corrected pairwise comparisons were conducted to identify specific patterns of change. In addition, Cohen *d* effect sizes were calculated to quantify the magnitude of within-group changes from baseline to postintervention (T1 to T3) and between-group differences in change scores.

To examine the potential moderating role of financial stress on the intervention's effectiveness, a moderation analysis was first conducted on the entire sample using mixed linear models. Exploratory subgroup analyses were conducted to further examine potential differential intervention effects based on participants' baseline financial stress levels. The sample was median-split into high and low financial stress subgroups based on baseline PIFS scores. Repeated measures ANOVAs were then conducted separately for each subgroup and outcome variable, with time as the within-subjects factor. This approach allowed us to examine whether the intervention had differential effects depending on participants' level of financial stress.

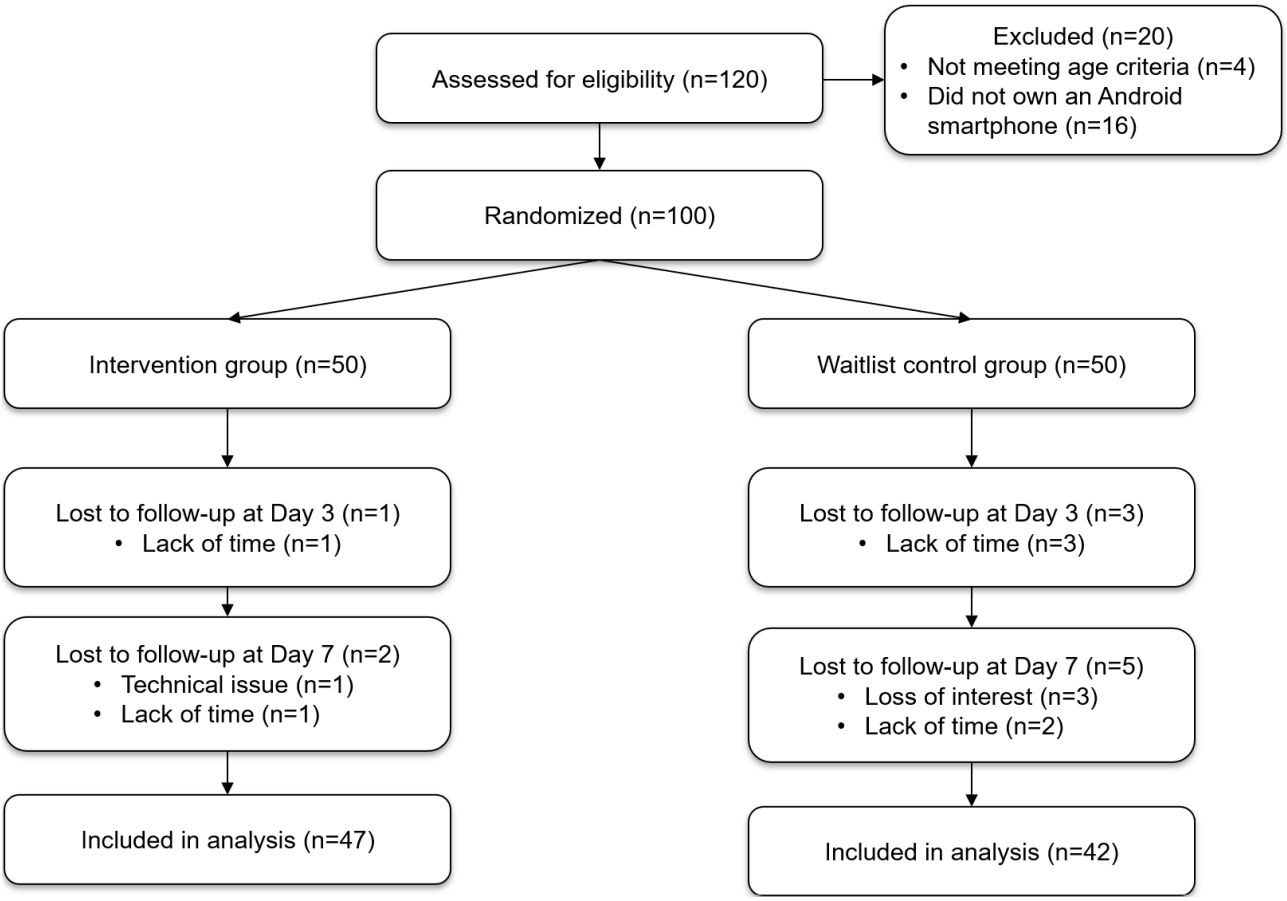
Results

Participant Characteristics

The CONSORT (Consolidated Standards of Reporting Trials) flow diagram (Figure 2) illustrates the participant flow through each stage of the study.

XSL•FO
RenderX

Figure 2. The CONSORT flow diagram. CONSORT: Consolidated Standards of Reporting Trials.



The final sample comprised 100 university students (mean age 20.8, SD 2.24 y; 62/100, 62% female) who were included in the intention-to-treat analysis using LOCF imputation. There were no significant baseline differences between the intervention and waitlist control groups on any demographic or outcome measures (P>.05). Detailed sample characteristics are presented in Table 1.

Baseline group differences were examined using independent 1-tailed t tests for continuous variables (age, CES-D, GAD-7, UCLA Loneliness Scale, and PIFS) and χ^2 tests for categorical variables (sex). There were no significant differences in age ($t_{80.91}$=0.80; P=.43), sex (χ^2_1=0.04; P=.84), CES-D scores ($t_{97.81}$=0.59; P=.56), GAD-7 scores ($t_{98.00}$=0.98; P=.33), or UCLA Loneliness scores ($t_{97.88}$=1.09; P=.28) between the intervention and waitlist control groups at baseline. Regarding financial circumstances, there was a marginally significant difference between groups in PIFS scores ($t_{96.05}$=1.81; P=.073), with the intervention group reporting slightly higher financial stress (mean 49.60, SD 14.67) than the control group (mean 44.10, SD 16.24), though this difference did not reach statistical significance at the conventional alpha level.

XSL•FO
RenderX

Table . Baseline demographic characteristics of participants (N=100).

Variable	Intervention (n=50)	Control (n=50)	Test statistic		*P* value
			t test (*df*)	Chi-square (*df*)	
Age, mean (SD)	20.96 (2.72)	20.60 (1.65)	0.80 (80.91)	—[a]	.43
Sex, n (%)			—	0.04 (1.00)	.84
Female	32 (64)	30 (60)			
Male	18 (36)	20 (40)			
CES-D[b], mean (SD)	14.00 (5.98)	13.28 (6.25)	0.59 (97.81)	—	.56
GAD-7[c], mean (SD)	13.62 (4.91)	12.66 (4.92)	0.98 (98.00)	—	.33
UCLA[d], mean (SD)	47.14(11.67)	44.64 (11.26)	1.09 (97.88)	—	.28
PIFS[e], mean (SD)	49.60 (14.67)	44.10 (16.24)	1.81 (96.05)	—	.07

[a]Not available.

[b]CES-D: Center for Epidemiologic Studies Depression Scale.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]UCLA: UCLA Loneliness Scale.

[e]PIFS: Psychological Inventory of Financial Scarcity Scale.

Attrition and Engagement

The overall attrition rate was 11% (11/100), with no significant difference between the intervention (6%, 3/50) and control (16%, 8/50) groups. In the intervention group, 47 participants completed all seven sessions, while 3 participants did not fully complete the intervention. In the waitlist control group, 3 participants did not complete the midintervention assessment on Day 3, and 5 participants did not provide valid responses to the postintervention assessment on Day 7. Reasons for dropout included lack of time, loss of interest, and technical difficulties with the app.

Intervention Effects on Mental Health Outcomes

Mixed-design ANOVAs were conducted to assess the intervention effects on each outcome variable (CES-D, GAD-7, and UCLA Loneliness Scale), with group (intervention vs control) as the between-subjects factor and time (T1, T2, and T3) as the within-subjects factor. The results are presented in Table 2.

Table . Mixed-design ANOVA results for intervention effects on mental health outcomes.

Outcome	*F* test (*df*)	*P* value	η^2_p
CES-D[a]			
Group	0.08 (1, 98)	.78	<.001
Time	7.18 (2, 196)	<.001	.07
Group × time	8.63 (2, 196)	<.001	.08
GAD-7[b]			
Group	0.74 (1, 98)	.39	.008
Time	0.70 (2, 196)	.50	.007
Group × time	1.31 (2, 196)	.27	.01
UCLA Loneliness Scale			
Group	0.26 (1, 98)	.61	.003
Time	11.13 (2, 196)	<.001	.10
Group × time	5.57 (2, 196)	.004	.05

[a]CES-D: Center for Epidemiologic Studies Depression Scale.

[b]GAD-7: Generalized Anxiety Disorder-7.

Significant group × time interactions were found for CES-D ($F_{2,\ 196}$=8.63; *P*<.001; η^2_p=.08) and UCLA Loneliness Scale scores ($F_{2,196}$=5.57; *P*=.004; η^2_p=.05), indicating differential changes in depression and loneliness between the intervention and waitlist control groups over time. These effect sizes represent medium effects according to Cohen benchmarks. No significant interaction effect was found for GAD-7 scores ($F_{2,196}$=1.31; *P*=.28; η^2_p=.01).

XSL•FO
RenderX

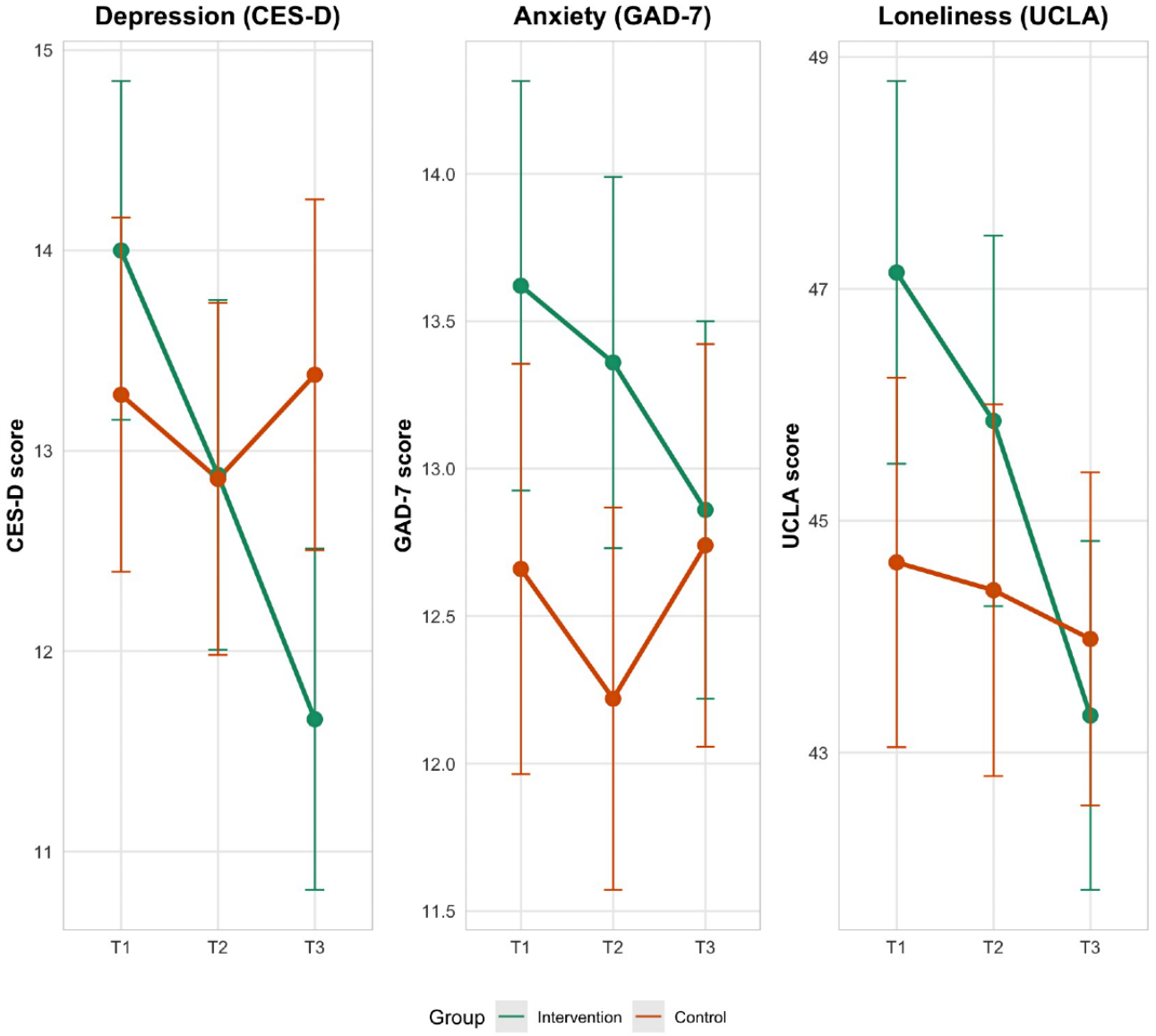
For CES-D scores, post hoc comparisons with Bonferroni correction revealed that the intervention group showed significant reductions from baseline to post-intervention (T1 to T3: t_{49}=3.85; P<.001; P_{adj}=.003), as well as from midintervention to post-intervention (T2 to T3: t_{49}=2.94; P=.025; P_{adj}=.026). In contrast, the waitlist control group showed no significant change from baseline to postintervention (all P_{adj}>.05).

Similarly, for UCLA Loneliness scores, the intervention group demonstrated a significant reduction from baseline to postintervention (T1 to T3: t_{49}=4.28; P<.001; P_{adj}<.001) and from midintervention to postintervention (T2 to T3: t_{49}=3.44; P=.001; P_{adj}=.01). No significant changes were observed in the waitlist control group (all P_{adj}>.05).

To quantify the magnitude of these effects, we calculated Cohen d for the change scores (T1 to T3) between the intervention and waitlist control groups. For CES-D, this analysis yielded Cohen d=0.71 (95% CI 0.30 - 1.17), and for UCLA Loneliness Scale, Cohen d=0.60 (95% CI 0.20 - 1.00), representing medium to large effects. These effect sizes based on change scores better reflect the differential patterns of improvement indicated by the significant group×time interactions.

Figure 3 illustrates the mean scores for CES-D, GAD-7, and UCLA Loneliness Scale across time points for the intervention and waitlist control groups.

Figure 3. Mean scores for the CES-D, GAD-7, and UCLA Loneliness Scale across time points for the intervention and waitlist control groups. CES-D: Center for Epidemiologic Studies Depression Scale; GAD-7: Generalized Anxiety Disorder-7 scale.



The plots in Figure 3 illustrate the differential changes in CES-D and UCLA Loneliness Scale scores between the intervention and waitlist control groups over time. For both outcomes, the intervention group (represented by the green line) shows a steeper decline in scores compared to the waitlist control group (represented by the orange line), suggesting a greater reduction in depressive symptoms and loneliness.

Moderation Analysis

To examine the potential moderating role of financial stress on the intervention's effectiveness, a moderation analysis was first

XSL•FO
RenderX

conducted on the entire sample using mixed linear models. The models included financial stress, time, intervention, and their interactions as fixed effects, and subject as a random effect.

For CES-D scores, the analysis revealed significant main effects of time ($F_{2,192}$=6.96, P=.001) and financial stress ($F_{1,96}$=12.96; P<.001), as well as significant Group×Time ($F_{2,192}$=7.92; P<.001) and Time×Financial stress ($F_{2,192}$=5.90; P=.003) interactions. Most importantly, a marginally significant 3-way interaction between group, time, and financial stress was observed ($F_{2,192}$=2.86; P=.06), suggesting that the effectiveness of the intervention on depressive symptoms might vary based on participants' financial stress levels. A post hoc power analysis indicated that with our sample size of 100 participants, α set at .05, and the corresponding effect size (f=0.17), the achieved power was 96.7% (see Figure S2 in Multimedia Appendix 1). This suggests adequate power for detecting the interaction effect in our primary analysis.

For UCLA Loneliness Scale scores, significant main effects were found for time ($F_{2,192}$=10.54; P<.001) and financial stress ($F_{1,96}$=23.08; P<.001). A significant Group×Time interaction was also observed ($F_{2,192}$=5.19; P=.006), indicating differential changes in loneliness between the intervention and waitlist control groups over time. However, the 3-way interaction between group, time, and financial stress did not reach statistical significance ($F_{2,192}$=2.32; P=.101), suggesting that the intervention's effect on loneliness was not significantly moderated by financial stress.

To further explore potential differential effects based on financial stress levels, we conducted exploratory subgroup analyses for depression and loneliness. Participants in the intervention group were divided into high (n=27) and low (n=23) financial stress subgroups based on a median split of baseline PIFS scores (median=51). Separate repeated measures ANOVAs were conducted for each subgroup. The results are presented in Table 3.

For the high financial stress, significant time effects were found for CES-D ($F_{2,52}$=11.56, P<.001, η^2_p=.31) and UCLA Loneliness Scale scores ($F_{2,52}$=11.18, P<.001, η^2_p=.30), representing large effects. In contrast, for the low financial stress subgroup (n=23), no significant time effects were found for CES-D ($F_{2,44}$=1.22, P=.31, η^2_p=.05) and only a marginally significant effect was observed for UCLA Loneliness in this subgroup ($F_{2,44}$=3.44, P=.04, η^2_p=.14) with a substantially smaller effect size.

These results suggest that the intervention was particularly effective for participants with high financial stress, demonstrating large effects on both depression and loneliness in this subgroup, while showing limited or no significant effects for those with lower financial stress.

Table . Repeated measures ANOVA results for high and low financial stress subgroups.

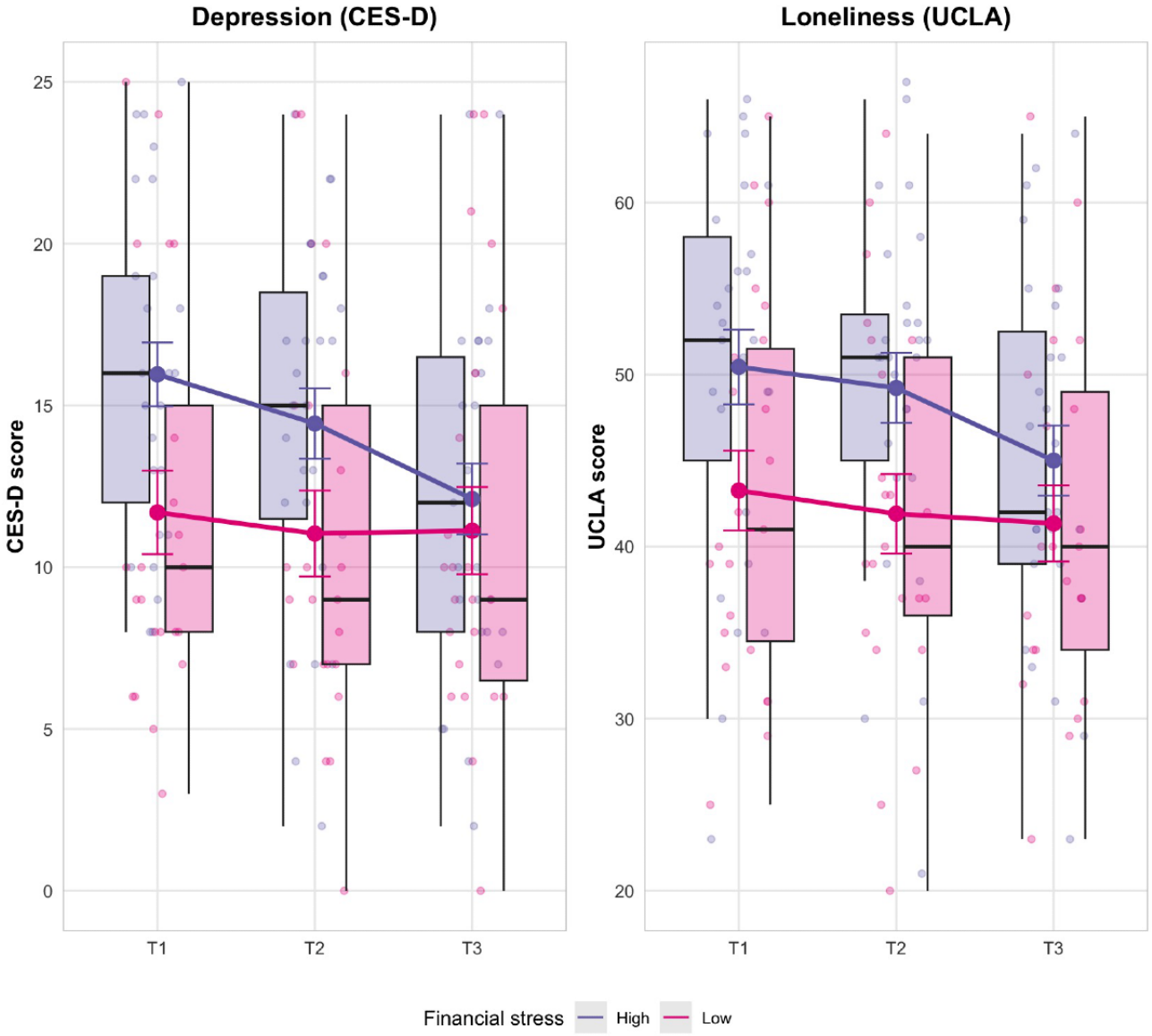
Outcome	Effect	F test (df)	P value	η^2_p
CES-D[a]				
High stress	Time	11.56 (2, 52)	<.001	.31
Low stress	Time	1.22 (2, 44)	.31	.005
UCLA Loneliness Scale				
High stress	Time	11.18 (2, 52)	<.001	.30
Low stress	Time	3.44 (2, 44)	.04	.14

[a]CES-D: Center for Epidemiologic Studies Depression Scale.

Figure 4 illustrates these differential patterns, with more pronounced reductions in both depression and loneliness scores among participants with high financial stress compared to those with low financial stress. For both outcomes, participants with high financial stress show a steeper decline in scores over time, indicating greater improvements in depressive symptoms and loneliness. These visualizations provide additional support for the differential effectiveness of the intervention based on participants' level of financial stress, with those experiencing high financial stress benefiting more from the intervention in terms of reducing depressive symptoms and loneliness.

XSL•FO
RenderX

Figure 4. Boxplot for CES-D and UCLA Loneliness Scale across time points for high and low financial stress subgroups in the intervention group. CES-D: Center for Epidemiologic Studies Depression Scale.



Discussion

Summary of Key Findings

This study demonstrated the efficacy of a culturally adapted, CBT-based AI chatbot intervention in reducing depression and loneliness among Chinese university students, with more pronounced effects observed in students experiencing higher levels of financial stress. The intervention significantly reduced depressive symptoms and feelings of loneliness compared to a waitlist control condition, with moderate to large effect sizes (T1 to T3: Cohen *d*=0.71 and 0.60, respectively). Notably, the chatbot did not produce significant improvements in anxiety symptoms, suggesting differential effectiveness across mental health domains. Furthermore, exploratory analyses revealed that students with high financial stress benefited substantially more from the intervention than those with lower financial stress, highlighting the potential usability of this approach for addressing mental health disparities related to economic factors.

Differential Effectiveness for Depression, Loneliness, and Anxiety

The significant reduction in depressive symptoms and loneliness among participants who engaged with the chatbot is consistent with previous studies on the effectiveness of AI-driven interventions for mental health. For example, a randomized trial of a CBT-based conversational agent (Woebot; Woebot Health) reported significant decreases in depression among young adults [25]. A scoping review likewise noted that several chatbot studies report lower loneliness, attributing this to enhanced perceived social support and engagement [24].

However, we found that the intervention did not significantly improve anxiety symptoms, which warrants further consideration, though it aligns with patterns observed in other digital mental health research. Meta-analytic data show systematically larger and more reliable effects for depression (Hedges *g*=0.56, 95% CI 0.38 - 0.74) than anxiety (Hedges *g*=0.45, 95% CI 0.30 - 0.6) across smartphone-based interventions [43].

XSL•FO
RenderX

Several factors may explain this differential effectiveness. First, anxiety disorders are characterized by physiological arousal components (eg, increased heart rate, muscle tension, and hypervigilance), which may be less responsive to purely cognitive-behavioral text-based interventions without complementary somatic management techniques or in-person guidance. While our chatbot included some relaxation exercises, these may have been insufficient to address the physiological aspects of anxiety effectively through a digital-only medium.

Second, the content of our chatbot focused more heavily on cognitive restructuring, targeting negative thought patterns about the self and future (which are central to depression) and behavioral activation to increase social connection (addressing loneliness), while potentially giving less attention to anxiety-specific techniques such as systematic desensitization or detailed safety behavior reduction. This content imbalance is consistent with the finding that many digital interventions emphasize depression-focused content over anxiety-specific components [18].

Third, anxiety disorders often feature avoidance behaviors that may require more intensive clinical intervention and monitoring to address effectively. Chatbot interactions may be less effective at detecting subtle avoidance patterns compared to in-person therapy. This aligns with findings from a meta-analysis, which reported stronger effects of digital interventions for depression compared to anxiety disorders, particularly when human guidance was minimal [59].

Fourth, the temporal dynamics of symptom improvement may differ between depression and anxiety. In symptom-trajectory studies, improvements in anxiety symptoms often emerge later in treatment compared to depressive symptoms. The study found that technology-based interventions for anxiety typically required a longer duration to demonstrate substantial effects [60]. Our relatively brief 7-day intervention period may have been sufficient to initiate changes in depressive symptoms but too short to observe meaningful improvements in anxiety.

Finally, cultural factors may have influenced the differential responsiveness of depression versus anxiety symptoms. In Chinese cultural contexts, expressions of depression may be more acceptable to acknowledge and address than anxiety, which is sometimes normalized as a necessary component of academic and professional achievement [32]. Cultural adaptations that more specifically target culturally shaped manifestations of anxiety might be necessary to improve the effectiveness of the intervention for anxiety symptoms in this population [36].

The Moderating Role of Financial Stress

The exploratory subgroup analyses revealed that students experiencing high financial stress showed significant improvements in depression and loneliness, while those with low financial stress demonstrated more modest changes. This pattern, though based on exploratory analyses given the marginally significant 3-way interaction, is consistent with emerging research on the relationship between financial stress and mental health interventions.

Students with high financial stress may face additional challenges, which make them more receptive to intervention. Longitudinal evidence shows that financial difficulties predict subsequent mental health problems in university students through increased stress and reduced coping resources [10]. Our chatbot specifically addressed financial concerns through dedicated modules on cognitive restructuring for financial worries and problem-solving for financial challenges, making it particularly relevant for financially stressed students.

Financial barriers often prevent students from accessing traditional mental health services. A review found that cost concerns were among the top barriers to help-seeking among university students, with those experiencing financial hardship particularly unlikely to access in-person services [61]. The chatbot's accessibility may have been especially valuable for these students, offering support without a financial burden. This aligns with a recent work demonstrating that digital interventions can help bridge the treatment gap for economically disadvantaged students [62].

Furthermore, the anonymity of digital interventions may be particularly important for students concerned about the stigma of help-seeking. A meta-review found that stigma concerns were heightened among individuals from lower socioeconomic backgrounds, who reported greater fears of judgment when seeking mental health support [63]. Our chatbot's private, anonymous nature may have circumvented these concerns.

This study found that financial stress–moderated intervention effectiveness adds to emerging evidence that economic factors play a crucial role in treatment response. For instance, a web-based CBT program ("Space from Money Worries") improved mental health and financial well-being outcomes most in participants reporting severe money concerns [64]. This suggests that comprehensive approaches that consider socioeconomic factors may enhance intervention effectiveness for vulnerable populations.

Theoretical and Practical Implications

This study has several theoretical and practical implications. From a theoretical perspective, it supports the feasibility of adapting evidence-based CBT principles for delivery through AI chatbots, extending the reach of these interventions to populations who may not access traditional services. The results align with the unified protocol model of transdiagnostic CBT [65], which suggests that core therapeutic elements can effectively address common underlying processes across different psychological symptoms.

The differential effectiveness for various mental health outcomes highlights the importance of considering the specific mechanisms of change for different psychological symptoms and adapting digital interventions accordingly. This idea echoes the behavioral-intervention-technology model, which emphasizes the need to match intervention components to specific therapeutic targets and mechanisms [66].

Practically, this study supports the integration of AI-driven interventions into existing university mental health support systems. The findings suggest that such interventions can be particularly beneficial for financially stressed students, providing a way to overcome barriers to traditional mental health services. Universities could consider implementing similar chatbot

XSL•FO
RenderX

interventions as part of a stepped care approach, potentially using them as an initial intervention or as a supplement to traditional services, as recommended by Lattie et al [20] in their framework for digital mental health implementation in higher education settings.

Furthermore, the cultural adaptation process demonstrated in this study provides a framework for developing culturally sensitive digital mental health interventions. The incorporation of cultural elements, such as collectivistic values, indirect communication styles, and culturally relevant metaphors, may have enhanced the acceptability and effectiveness of the intervention for Chinese students. This approach aligns with the framework for cultural adaptation of psychological interventions and could be applied to adapt digital interventions for other cultural contexts, potentially improving their global applicability.

Limitations and Future Directions

Our sample size (N=100) was adequate for detecting primary effects according to our a priori power analysis (which determined a minimum sample of 52 participants was needed), but limited the power of subgroup and moderation analyses. The exploratory nature of our subgroup analyses, with relatively small sample sizes in each financial stress category (27 high stress and 23 low stress in the intervention group), limits the robustness of our moderation findings. Future research with larger samples specifically powered for moderation analyses is needed to confirm the differential effectiveness of the intervention based on financial stress levels.

In addition, our exclusion criteria (current psychotherapy, psychiatric-medication use, or a history of severe mental illness) restrict generalizability to the wider student population, where complex mental health needs are common. Although these criteria were necessary to isolate intervention effects and protect participants in this initial evaluation, future work should test the chatbot as an adjunct to conventional treatments in more clinically diverse samples, in keeping with recommendations that digital tools complement rather than replace routine care [67].

The use of a waitlist control group, while providing a baseline for comparison, has inherent limitations recognized in digital mental health research [45]. Waitlist controls may not account for nonspecific factors such as expectancy effects or attention. Participants in waitlist conditions may also postpone seeking help or making changes while waiting for the intervention, potentially exaggerating between-group differences. Future studies should consider active control conditions, such as psychoeducational resources or attention-matched control interventions, to better isolate the specific effects of the chatbot's therapeutic components.

Our reliance on self-report measures may have introduced reporting biases. Cultural factors might influence how Chinese students report mental health symptoms, potentially affecting the accuracy of our assessments. Chinese students, for example, tend to emphasize somatic rather than psychological complaints, which may have affected our measurement of mental health outcomes. Future studies should incorporate multiple assessment methods, including clinician-rated measures, behavioral indicators, or physiological markers, to provide a more comprehensive evaluation of intervention effects.

The relatively brief 7-day intervention period may have been insufficient to produce meaningful changes in anxiety symptoms or to establish lasting improvements in depression and loneliness. Meta-analytic evidence indicates stronger effects for digital CBT delivered over a long period [68]. Future studies should investigate longer intervention durations and include follow-up assessments to evaluate the sustainability of effects over time.

Implementation Considerations

The high adherence rate observed in our study (47/50, 94% in the intervention group) exceeds typical rates reported in digital intervention research, where adherence rates of 40% - 60% are common. While this speaks to the engaging nature of the chatbot, it may have been influenced by the monetary compensation provided to participants, particularly relevant for financially stressed students. In real-world implementations without financial incentives, adherence rates might be lower. Future research should explore sustainable engagement strategies that do not rely on monetary compensation, such as gamification elements, personalized messaging, or integration with existing university systems, which have been proposed [69] and should be explored.

While our chatbot incorporated several advanced features, it still had technical limitations. The rule-based and template-based responses (comprising 90% of interactions) may not fully capture the nuanced needs of users compared to more sophisticated AI systems. As natural language processing technology advances, future iterations could incorporate more adaptive and personalized interaction capabilities while maintaining clinical safety and appropriateness.

Finally, despite substantial efforts we made to culturally adapt the chatbot, there may be deeper cultural nuances that were not fully addressed. Future research could use more comprehensive cultural adaptation frameworks and involve larger, more diverse groups of cultural consultants and end users in the development process to enhance cultural relevance and acceptability.

Conclusions

This study demonstrated that a culturally adapted, CBT-based AI chatbot can effectively reduce depression and loneliness symptoms among Chinese university students, with particularly strong effects for those experiencing high financial stress. These findings highlight the potential of digital mental health interventions to address mental health disparities related to economic factors and expand access to support for underserved populations.

The differential effectiveness across mental health domains, with significant improvements in depression and loneliness but not anxiety, underscores the importance of tailoring digital interventions to specific psychological mechanisms and cultural contexts. Future developments should incorporate stronger anxiety-specific components and longer intervention periods to

XSL•FO
RenderX

potentially enhance effects across a broader range of mental health outcomes.

As universities worldwide face increasing demand for mental health services amid resource constraints, culturally adapted digital interventions offer a promising approach to extend support to students who might otherwise go untreated. The moderating effect of financial stress suggests that such interventions may be particularly valuable for economically disadvantaged students, potentially helping to address socioeconomic disparities in mental health treatment access and outcomes.

Future research should focus on refining these interventions to address anxiety more effectively, investigating their long-term effects, and exploring their implementation in diverse real-world settings. In addition, studies examining the specific mechanisms through which financial stress impacts mental health and how digital interventions can best address these mechanisms would further enhance our understanding and improve intervention design for vulnerable student populations.

Acknowledgments

This research was funded by the Humanities and Social Sciences Foundation of the Ministry of Education of China (grant 23YJC840022), the Social Science Foundation of Jiangsu Province (grant 23SHC009), and the General Project of Philosophy and Social Science Research in Universities in Jiangsu Province (grant 2023SJYB0243)

Data Availability

The datasets generated or analyzed during this research are available from the corresponding author on reasonable request.

Authors' Contributions

Y Wang wrote the manuscript and developed the methodology. XL conceptualized the project and performed data curation and analysis. QZ reviewed and edited the manuscript. DY reviewed and revised the draft. Y Wu oversaw project administration, supervision, and funding acquisition. All authors approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Power analysis.
[DOCX File, 499 KB - mhealth_v13i1e63806_app1.docx]

Multimedia Appendix 2
The detailed cognitive behavioral therapy (CBT) module structure, cultural adaptation framework, validation metrics, and sample chatbot dialogue examples.
[DOCX File, 25 KB - mhealth_v13i1e63806_app2.docx]

Checklist 1
CONSORT-EHEALTH Checklist (V 1.6.1).
[PDF File, 1264 KB - mhealth_v13i1e63806_app3.pdf]

References

1. Ibrahim AK, Kelly SJ, Adams CE, Glazebrook C. A systematic review of studies of depression prevalence in university students. J Psychiatr Res 2013 Mar;47(3):391-400. [doi: 10.1016/j.jpsychires.2012.11.015] [Medline: 23260171]
2. Auerbach RP, Mortier P, Bruffaerts R, et al. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638. [doi: 10.1037/abn0000362] [Medline: 30211576]
3. Park SY, Andalibi N, Zou Y, Ambulkar S, Huh-Yoo J. Understanding students' mental well-being challenges on a university campus: interview study. JMIR Form Res 2020 Mar 5;4(3):e15962. [doi: 10.2196/15962] [Medline: 32134393]
4. Lee J, Jeong HJ, Kim S. Stress, anxiety, and depression among undergraduate students during the COVID-19 pandemic and their use of mental health services. Innov High Educ 2021;46(5):519-538. [doi: 10.1007/s10755-021-09552-y] [Medline: 33907351]
5. Lun KW, Chan C, Ip PK, et al. Depression and anxiety among university students in Hong Kong. Hong Kong Med J 2018 Oct;24(5). [doi: 10.12809/hkmj176915]

XSL•FO
RenderX

6. Negash S, Kartschmit N, Mikolajczyk RT, et al. Worsened financial situation during the COVID-19 pandemic was associated with depressive symptomatology among university students in Germany: results of the COVID-19 international student well-being study. Front Psychiatry 2021;12(743158):743158. [doi: 10.3389/fpsyt.2021.743158] [Medline: 34975561]
7. Wang Z, Zou Q. Prevalence and associated factors of depressive symptoms among the young adults during the post-epidemic period - evidence from the first wave of COVID-19 in Hubei Province, China. Acta Psychol (Amst) 2022 Jun;226:103577. [doi: 10.1016/j.actpsy.2022.103577] [Medline: 35349926]
8. Bennett J, Heron J, Kidger J, Linton MJ. Investigating change in student financial stress at a UK university: multi-year survey analysis across a global pandemic and recession. Education Sciences 2023 Dec;13(12):1175. [doi: 10.3390/educsci13121175]
9. Adams DR, Meyers SA, Beidas RS. The relationship between financial strain, perceived stress, psychological symptoms, and academic and social integration in undergraduate students. J Am Coll Health 2016 Jul 3;64(5):362-370. [doi: 10.1080/07448481.2016.1154559]
10. Richardson T, Elliott P, Roberts R, Jansen M. A longitudinal study of financial difficulties and mental health in a national sample of British undergraduate students. Community Ment Health J 2017 Apr;53(3):344-352. [doi: 10.1007/s10597-016-0052-0] [Medline: 27473685]
11. Farrer LM, Gulliver A, Bennett K, Fassnacht DB, Griffiths KM. Demographic and psychosocial predictors of major depression and generalised anxiety disorder in Australian university students. BMC Psychiatry 2016 Jul 15;16(241):241. [doi: 10.1186/s12888-016-0961-z] [Medline: 27422284]
12. Auerbach RP, Alonso J, Axinn WG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970. [doi: 10.1017/S0033291716001665] [Medline: 27484622]
13. Osborn TG, Li S, Saunders R, Fonagy P. University students' use of mental health services: a systematic review and meta-analysis. Int J Ment Health Syst 2022 Dec 17;16(1):57. [doi: 10.1186/s13033-022-00569-0] [Medline: 36527036]
14. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306. [doi: 10.2196/jmir.6588] [Medline: 28916506]
15. Guenthner L, Baldofski S, Kohls E, Schuhr J, Brock T, Rummel-Kluge C. Differences in help-seeking behavior among university students during the COVID-19 pandemic depending on mental health status: results from a cross-sectional survey. Behav Sci (Basel) 2023 Oct 25;13(11):885. [doi: 10.3390/bs13110885] [Medline: 37998632]
16. Kearns M, Muldoon OT, Msetfi RM, Surgenor PWG. Understanding help-seeking amongst university students: the role of group identity, stigma, and exposure to suicide and help-seeking. Front Psychol 2015;6:1462. [doi: 10.3389/fpsyg.2015.01462] [Medline: 26483722]
17. Yang LH, Purdie-Vaughns V, Kotabe H, et al. Culture, threat, and mental illness stigma: identifying culture-specific threat among Chinese-American groups. Soc Sci Med 2013 Jul;88:56-67. [doi: 10.1016/j.socscimed.2013.03.036]
18. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869. [doi: 10.2196/12869] [Medline: 31333198]
19. Harrer M, Apolinário-Hagen J, Fritsche L, et al. Effect of an internet- and app-based stress intervention compared to online psychoeducation in university students with depressive symptoms: results of a randomized controlled trial. Internet Interv 2021 Apr;24:100374. [doi: 10.1016/j.invent.2021.100374] [Medline: 33718001]
20. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Feb;1(2):87-100. [doi: 10.1038/s44159-021-00003-1]
21. Lipschitz JM, Van Boxtel R, Torous J, et al. Digital mental health interventions for depression: scoping review of user engagement. J Med Internet Res 2022 Oct 14;24(10):e39204. [doi: 10.2196/39204] [Medline: 36240001]
22. Fisher A, Corrigan E, Cross S, et al. Decision-making about uptake and engagement among digital mental health service users: a qualitative exploration of therapist perspectives. Clin Psychol (Aust Psychol Soc) 2023 May 4;27(2):171-185. [doi: 10.1080/13284207.2022.2163157]
23. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
24. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: a scoping review. Int J Med Inform 2019 Dec;132:103978. [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
25. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
26. D'Alfonso S. AI in mental health. Curr Opin Psychol 2020 Dec;36:112-117. [doi: 10.1016/j.copsyc.2020.04.005] [Medline: 32604065]
27. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: the "black swan" for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317. [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]

XSL•FO
RenderX

28. Elmer T, Mepham K, Stadtfeld C. Students under lockdown: comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. PLOS ONE 2020;15(7):e0236337. [doi: 10.1371/journal.pone.0236337] [Medline: 32702065]
29. Haque MDR, Rubya S. An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. JMIR mHealth uHealth 2023 May 22;11:e44838. [doi: 10.2196/44838] [Medline: 37213181]
30. Grové C. Co-developing a mental health and wellbeing chatbot with and for young people. Front Psychiatry 2020;11:606041. [doi: 10.3389/fpsyt.2020.606041] [Medline: 33597898]
31. Ma L, Huang C, Cui Z. A meta-analysis of cognitive-behavioral therapy for depressive symptoms among college students, alone and in comparison with other interventions. Journal of Behavioral and Cognitive Therapy 2020 Jun;30(2):127-139. [doi: 10.1016/j.jbct.2020.03.008]
32. Chen SX, Mak WWS. Seeking professional help: etiology beliefs about mental illness across cultures. J Couns Psychol 2008 Oct;55(4):442-450. [doi: 10.1037/a0012898] [Medline: 22017551]
33. Liu J, Ma H, He YL, et al. Mental health system in China: history, recent service reform and future challenges. World Psychiatry 2011 Oct;10(3):210-216. [doi: 10.1002/j.2051-5545.2011.tb00059.x] [Medline: 21991281]
34. Ryder AG, Chentsova-Dutton YE. Depression in cultural context: "Chinese somatization," revisited. Psychiatr Clin North Am 2012 Mar;35(1):15-36. [doi: 10.1016/j.psc.2011.11.006] [Medline: 22370488]
35. Chowdhary N, Jotheeswaran AT, Nadkarni A, et al. The methods and outcomes of cultural adaptations of psychological treatments for depressive disorders: a systematic review. Psychol Med 2014 Apr;44(6):1131-1146. [doi: 10.1017/S0033291713001785] [Medline: 23866176]
36. Hwang WC, Myers HF, Chiu E, et al. Culturally adapted cognitive-behavioral therapy for Chinese Americans with depression: a randomized controlled trial. Psychiatr Serv 2015 Oct;66(10):1035-1042. [doi: 10.1176/appi.ps.201400358] [Medline: 26129996]
37. Choi I, Zou J, Titov N, et al. Culturally attuned internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]
38. Li S, Xi Z, Barnett P, Saunders R, Shafran R, Pilling S. Efficacy of culturally adapted interventions for common mental disorders in people of Chinese descent: a systematic review and meta-analysis. Lancet Psychiatry 2023 Jun;10(6):426-440. [doi: 10.1016/S2215-0366(23)00118-9]
39. Rathod S, Gega L, Degnan A, et al. The current status of culturally adapted mental health interventions: a practice-focused review of meta-analyses. Neuropsychiatr Dis Treat 2018;14:165-178. [doi: 10.2147/NDT.S138430] [Medline: 29379289]
40. Sit HF, Hong IW, Burchert S, et al. A feasibility study of the WHO digital mental health Intervention step-by-step to address depression among Chinese young adults. Front Psychiatry 2021;12:812667. [doi: 10.3389/fpsyt.2021.812667] [Medline: 35069297]
41. Yeo G, Reich SM, Liaw NA, Chia EYM. The effect of digital mental health literacy interventions on mental health: systematic review and meta-analysis. J Med Internet Res 2024 Feb 29;26(1):e51268. [doi: 10.2196/51268] [Medline: 38421687]
42. Zhang X, Lewis S, Firth J, Chen X, Bucci S. Digital mental health in China: a systematic review. Psychol Med 2021 Nov;51(15):2552-2570. [doi: 10.1017/S0033291721003731] [Medline: 34581263]
43. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
44. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
45. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research... and three things we can do about them. Psychiatr Serv 2017 May 1;68(5):427-429. [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
46. Bernal G, Jiménez-Chafey MI, Domenech Rodríguez MM. Cultural adaptation of treatments: a resource for considering culture in evidence-based practice. Professional Psychology: Research and Practice 2009;40(4):361-368. [doi: 10.1037/a0016401]
47. Day S, Laver K, Jeon YH, Radford K, Low LF. Frameworks for cultural adaptation of psychosocial interventions: a systematic review with narrative synthesis. Dementia (London) 2023 Nov;22(8):1921-1949. [doi: 10.1177/14713012231192360] [Medline: 37515347]
48. Beck JS. Cognitive Behavior Therapy: Basics and Beyond: Guilford Publications; 2020. URL: https://books.google.com/books?hl=en&lr=&id=yb_nDwAAQBAJ&oi=fnd&pg=PP1&dq=beck+cbt+manual&ots=wjm_IGHYZ8&sig=EKqpTN79LQmg71rVnQAWqUfTPhs [accessed 2024-06-04]
49. Cuijpers P, Karyotaki E, Eckshtain D, et al. Psychotherapy for depression across different age groups: a systematic review and meta-analysis. JAMA Psychiatry 2020 Jul 1;77(7):694-702. [doi: 10.1001/jamapsychiatry.2020.0164] [Medline: 32186668]
50. Bocklisch T, Faulkner J, Pawlowski N, Nichol A. Rasa: open source language understanding and dialogue management. arXiv. Preprint posted online on Dec 14, 2017. [doi: 10.48550/arXiv.1712.05181]

XSL•FO
RenderX

51. Devlin J, Chang MW, Lee K, Toutanova K. B. BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein J, Doran C, Solorio T, editors. Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); Jun 2-7, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/N19-1423]
52. Brooke J. SUS - a quick and dirty usability scale. In: Usability Evaluation In Industry: CRC Press; 1996.
53. Determining what individual SUS scores mean: adding an adjective rating scale. J User Exp 2009:114-123 [FREE Full text]
54. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
55. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]
56. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
57. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
58. van Dijk WW, van der Werf MMB, van Dillen LF. The Psychological Inventory of Financial Scarcity (PIFS): a psychometric evaluation. Journal of Behavioral and Experimental Economics 2022 Dec;101:101939. [doi: 10.1016/j.socec.2022.101939]
59. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health—a systematic review and meta-analysis. npj Digit Med 2019 Dec 2;2(1):1-10. [doi: 10.1038/s41746-019-0188-8]
60. Eilert N, Enrique A, Wogan R, Mooney O, Timulak L, Richards D. The effectiveness of internet-delivered treatment for generalized anxiety disorder: an updated systematic review and meta-analysis. Depress Anxiety 2021 Feb;38(2):196-219. [doi: 10.1002/da.23115] [Medline: 33225589]
61. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10(1):113. [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
62. King CA, Eisenberg D, Pistorello J, et al. Electronic bridge to mental health for college students: a randomized controlled intervention trial. J Consult Clin Psychol 2022 Feb;90(2):172-183. [doi: 10.1037/ccp0000709] [Medline: 35099205]
63. Clement S, Schauman O, Graham T, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
64. Richardson T, Enrique A, Earley C, Adegoke A, Hiscock D, Richards D. The acceptability and initial effectiveness of “space from money worries”: an online cognitive behavioral therapy intervention to tackle the link between financial difficulties and poor mental health. Front Public Health 2022;10:739381. [doi: 10.3389/fpubh.2022.739381] [Medline: 35493363]
65. Barlow DH, Farchione TJ, Bullis JR, et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. JAMA Psychiatry 2017 Sep 1;74(9):875-884. [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
66. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146. [doi: 10.2196/jmir.3077] [Medline: 24905070]
67. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
68. Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 1;78(4):361-371. [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
69. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 6;20(6):e199. [doi: 10.2196/jmir.9275] [Medline: 29875089]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
CES-D: Center for Epidemiologic Studies Depression Scale
CONSORT: Consolidated Standards of Reporting Trials
GAD-7: Generalized Anxiety Disorder-7 scale
PIFS: Psychological Inventory of Financial Scarcity

XSL•FO
RenderX

Edited by L Buis; submitted 30.06.24; peer-reviewed by H Bu, OEV Bergeijk, S Zhu; revised version received 19.06.25; accepted 21.07.25; published 29.08.25.

Please cite as:
Wang Y, Li X, Zhang Q, Yeung D, Wu Y
Effect of a Cognitive Behavioral Therapy–Based AI Chatbot on Depression and Loneliness in Chinese University Students: Randomized Controlled Trial With Financial Stress Moderation
JMIR Mhealth Uhealth 2025;13:e63806
URL: https://mhealth.jmir.org/2025/1/e63806
doi:10.2196/63806

© Yahui Wang, Xuhong Li, Qiaochu Zhang, Dannii Yeung, Yihan Wu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness and Implementation Outcomes of an mHealth App Aimed at Promoting Physical Activity and Improving Psychological Distress in the Workplace Setting: Cluster-Level Nonrandomized Controlled Trial

Kazuhiro Watanabe[1], PhD; Mitsuhiro Sato[2], BPE; Shoichi Okusa[3], BE; Akizumi Tsutsumi[1], MD

[1]Department of Public Health, Kitasato University School of Medicine, 1-15-1 Kitazato, Minami-ku, Sagamihara, Japan

[2]Human Resources Unit, Fujitsu General Limited, Kawasaki, Japan

[3]TOSHIBA Health Insurance Society, Kawasaki, Japan

Corresponding Author:
Kazuhiro Watanabe, PhD
Department of Public Health, Kitasato University School of Medicine, 1-15-1 Kitazato, Minami-ku, Sagamihara, Japan

Abstract

Background: Encouraging physical activity improves mental health and is recommended in workplace mental health guidelines. Although mobile health (mHealth) interventions are promising for physical activity promotion, their impact on mental health outcomes is inconsistent. Furthermore, poor user retention rates of mHealth apps pose a major challenge.

Objective: This study aimed to examine the effectiveness and implementation outcomes of the smartphone app ASHARE in Japanese workplace settings, leveraging a deep learning model to monitor depression and anxiety through physical activity.

Methods: This hybrid effectiveness-implementation trial was a 3-month nonrandomized controlled trial conducted from October 2023 to September 2024. Work units and employees were recruited and allocated to the intervention or active control group based on preference. The intervention group installed the ASHARE app, whereas the control group participated in an existing multicomponent workplace program promoting physical activity. Changes in physical activity and psychological distress levels were compared between the groups. User retention rates, participation rates, acceptability, appropriateness, feasibility, satisfaction, and potential harm were also assessed.

Results: A total of 84 employees from 7 work units participated (67 from 5 units in the intervention group and 17 from 2 units in the control group). In total, 78 employees completed the 3-month follow-up survey (follow-up rate: 93%). Both groups showed increased physical activity, and the intervention group showed reduced psychological distress; however, the differences between groups were not statistically significant (P=.20; P=.36). In a sensitivity analysis of protocol-compliant employees (n=21), psychological distress levels were significantly reduced in the intervention group compared with the control group (coefficient=−3.68, SE 1.65; P=.03). The app's 3-month user retention rate was 20% (12/61), which was lower than the participation rate in each component of the control programs. Implementation outcomes evaluated by employees were less favorable in the intervention group than in the control group, whereas health promotion managers found them to be similar.

Conclusions: The ASHARE app did not show superior effectiveness compared with an existing multicomponent workplace program for promoting physical activity. An implementation gap may exist between health promotion managers and employees, possibly contributing to the app's low user retention rate. Future research should focus on examining the effectiveness of strategies to get engagement from managers and from segments of employees with favorable responses in the workplace at an early stage.

Trial Registration: UMIN-CTR UMIN000052374; https://tinyurl.com/4x4dxfv6

(JMIR Mhealth Uhealth 2025;13:e70473) doi:10.2196/70473

KEYWORDS

eHealth; behavioral change; mobile phone; smartphone; workplace; depression; anxiety

Introduction

Background

Mental health conditions among workers are widespread and adversely affect critical outcomes. In 2019, an estimated 301 million working-age adults experienced anxiety, while approximately 280 million had depression [1]. Furthermore, many workers report subclinical symptoms such as psychological distress [2-4]. These mental health concerns lead to increased absenteeism, deteriorating quality of life, and overall well-being [5]. The global economic burden of mental

XSL•FO
RenderX

health conditions is estimated at approximately US $1 trillion annually, primarily due to productivity losses [1,6,7]. Given their range of severity, early interventions and preventive measures are essential for managing mental health in workplace environments.

Promoting physical activity effectively treats and prevents depression and anxiety and plays a critical role in preventing mental health conditions [8]. A previous study has reported that promoting physical activity improves outcomes such as job stress, absenteeism, job performance, and turnover among workers [9]. It is among the most evidence-based interventions for preventing major mental disorders in workers and is recommended in the World Health Organization's mental health at work guidelines [1,10]. Due to the high prevalence of physical inactivity and mental health conditions among workers, encouraging physical activity has become a key strategy for improving workplace mental health.

Mobile health (mHealth) interventions delivered through smartphones and other devices effectively promote physical activity. These interventions typically use basic behavior change techniques, such as psychoeducation, goal setting, monitoring and providing feedback on physical activity, sharing and competing regarding activity levels among participants, or offering rewards [11]. mHealth interventions offer high accessibility and cost-effectiveness to diverse populations [12], and their effectiveness in promoting physical activity has been confirmed in various groups, including patients and the general adult population [13-15]. However, the quality of scientific evidence remains low, and the effects are inconsistent [16]. Furthermore, effectiveness is not always observed when depression, anxiety, or stress-related indicators are used as outcomes [17,18]. Although numerous mHealth interventions and apps aimed at promoting physical activity have been developed to date, studies in Asia or targeting healthy workers remain limited [13,19]. Furthermore, most of these interventions do not primarily focus on improving mental health, nor do they measure related outcomes [15].

The biggest challenge is poor implementation. For example, when programs are provided through smartphone apps, usage often declines due to poor design, obstacles, excessive notifications or ads, and insufficient efforts to prevent dropouts. In Japan, health care app retention rates are extremely low, ranging from 3.5% to 4% by the 30th day [20]. Low use rates also pose a major challenge in research [21]. Workers recognized that physical activity effectively improved mental health and expressed a desire for individualized health services. However, they highlighted the limited benefits of nonleisure physical activities and prioritized sleep and rest over physical activity [22]. Low organizational- and employer or manager-level engagement in programs are also important barriers to successful implementation in the Japanese workplace [23]. Therefore, while mHealth interventions may be scientifically effective, their impact is limited by poor user engagement. Given these challenges, further research is needed for workplace mHealth interventions to promote physical activity specifically designed to improve mental health outcomes, and tailored to the needs of the healthy working population to increase their long-term engagement.

Objectives

To bridge these research gaps, we developed a native smartphone app called ASHARE to encourage physical activity and alleviate psychological distress in the working population. Previous findings on the implemented deep learning model and a feasibility study have been published elsewhere [24,25]. Briefly, the app incorporates basic behavior change techniques to encourage physical activity, including self-monitoring, feedback, and user-driven data sharing. In addition, it uses a deep learning model to enable passive monitoring of depression and anxiety based on physical activity patterns [24]. The beta version of the ASHARE app did not effectively promote physical activity or alleviate psychological distress overall but reduced the number of participants with severe psychological distress [25]. This app may contribute to providing evidence for mHealth research on improving mental health through physical activity among healthy workers, an area that has been scarcely examined. In addition, the deep learning model with individualized feedback has the potential to enhance individual worker engagement and to encourage employers or managers to implement the app who face constraints in human resources and prioritization for health promotion. This study aimed to evaluate the effectiveness and implementation outcomes of the revised ASHARE app [26] compared with an existing multicomponent workplace program as an active control. Based on the substantial evidence supporting the effectiveness of mHealth apps that use basic behavior change techniques to encourage physical activity, along with the additional feature of the ASHARE app for passive monitoring of depression and anxiety, we hypothesized that the increase in physical activity and improvement in psychological distress in the intervention group would be significantly greater than those in the active control group. In addition, we explored the retention rate, acceptability, appropriateness, feasibility, and satisfaction with the app compared with the existing workplace program to assess its implementation in workplace settings.

Methods

Study Design

A 3-month superiority nonrandomized controlled trial was conducted to investigate the effectiveness and implementation outcomes of the revised full version of the ASHARE app. We adopted a hybrid effectiveness-implementation design. From October 2023 to September 2024, the authors recruited work units and company employees through personal networks and connections. When companies agreed to participate in the trial, health promotion managers were selected as representatives to communicate with the research teams and select participating work units. Employees from the selected work units were invited to participate and asked to complete an internet-based baseline survey provided by the research team. In addition, the research team assigned work units to either the intervention or active control group based on the preferences expressed by health promotion managers. We adopted a nonrandomized design because randomization would not have aligned with the work units' decision-making process. In addition, requiring randomization could have resulted in low participation rates, potentially compromising the successful execution of measuring

XSL•FO
RenderX

implementation outcomes. The intervention group installed and used the ASHARE app for 3 months, while the control group participated in the existing multicomponent workplace program to promote physical activity during the same period [27]. Immediately after completing the programs, the research team sent the employees the internet-based follow-up survey to measure the effectiveness and implementation outcomes (acceptability, appropriateness, feasibility, and satisfaction) of each program. Health promotion managers also assessed the implementation outcomes of each program. As an incentive, health promotion managers and employees received an Amazon gift certificate worth 3000 yen (US $20.5) upon completion of the 3-month follow-up survey. This manuscript follows the guidelines of nonrandomized evaluations of behavioral and public health interventions (TREND [Transparent Reporting of Evaluations with Nonrandomized Designs] statement; Checklist 1) and Standards for Reporting Implementation Studies (StaRI) statement [28,29] (Checklist 2).

Participants

This hybrid nonrandomized controlled trial included work units, health managers, and employees. No eligibility criteria were set for the work units. Eligible health promotion managers were employees in charge of human resources, labor affairs, general affairs, or health management, or employees in charge of promoting mental health in the workplace, according to the guidelines of the Japanese Ministry of Health, Labour and Welfare. Eligible employees within work units were aged 18 years or older, capable of completing questionnaires in Japanese, and owning a personal smartphone. Employees were excluded if they were absent during the enrollment period or had been absent within the previous 12 months.

Interventions

Intervention Group

Employees in the work units assigned to the intervention group were instructed to install the ASHARE app on their smartphones and use it for 3 months. ASHARE is a native app compatible with Android (version 5.0 and later) and iOS (version 12.0 and later) devices. Following an initial feasibility trial [25], the fully developed iOS and Android app versions were made available through Google Play and the App Store, respectively [26]. The app uses basic behavior change techniques to encourage physical activity, including self-monitoring, feedback, and user-to-user data sharing. It integrates with Google Fit (Android) and Apple Health (iOS) to track the duration of physical activity. Each time the user launches the app, demographic and physical activity data from the previous day are sent to a cloud server. The model, using long short-term memory, calculates the predicted psychological distress for the present day, and predicted scores are displayed on the app. After the initial trial [25], several modifications were made to backend processing, feedback messages, the display of predicted depression and anxiety scores, graph tracking score changes, and feedback to the prediction algorithm. Furthermore, related content, YouTube channels, RSS feeds, a social networking account (X), and promotional videos were added. During the intervention period, employees were not forced to use the app, they could open it anytime. In terms of ideal app usage, the researchers anticipated that participants would spend approximately 5 min each day launching the app to review their physical activity patterns and the results of their predicted mental health scores while simultaneously aiming to increase their physical activity. To improve adherence to the program, researchers sent weekly email reminders to use the app. The app also sends users smartphone notifications with new information, and their predicted depression and anxiety scores were provided daily at 6 am. In addition, we monitored participants' login frequency through the management dashboard during the study period once a week.

Control Group

Employees in the work units allocated to the control group were requested to participate in an existing multicomponent workplace intervention program featuring environmental changes for 3 months. The effectiveness of the program in promoting physical activity has been confirmed in Japanese workplace settings in a previous cluster randomized controlled trial [27]. The program comprises 13 elements across the different workplace components, including policymaking and declarations, posters detailing the program contents and physical activity recommendations, prompts for stair use near stairs and elevators, physical activity competitions between individuals within work units, and psychoeducation to increase self-regulation of physical activity. These elements, as evidenced by a literature review, build awareness and social norms around physical activity, enhance accessibility, and strengthen individual cognitive-behavioral skills [30-32]. Due to limited resources, only the elements considered relevant or feasible for each work unit were conducted after discussions with health promotion managers. To improve adherence to the program, researchers sent weekly email reminders to all participants in the work units. In addition, we monitored program participation through responses to the weekly survey, which tracked the number of steps for the physical activity competition.

Outcomes

Physical Activity

The primary outcome measure was total physical activity, which was assessed using the Japanese version of the Global Physical Activity Questionnaire (GPAQ) [33]. This scale is widely used to evaluate physical activity levels in epidemiological studies and has demonstrated acceptable reliability and convergent validity across multiple countries, including Japan [34]. The intensity of physical activity was measured in metabolic equivalents (METs). As a continuous indicator, the total amount of physical activity per week (MET-hours/week) was calculated according to the GPAQ analysis guidelines [35]. Furthermore, physical activity levels were classified into 3 ordinal categories: low, moderate, and high.

Psychological Distress

The secondary outcome, self-reported psychological distress, was measured using the 6-item Kessler Psychological Distress Scale (K6) [36]. This scale was also used as the supervised indicator when the deep learning model was implemented in the ASHARE app [24]. The K6 includes 6 items that evaluate the frequency of symptoms related to depression and anxiety,

XSL•FO
RenderX

rated on a 5-point Likert scale (0=none of the time to 4=all the time). The Japanese version of the K6 has been confirmed to be valid [37]. The total K6 score was calculated as a continuous variable. Levels of psychological distress were also categorized into 3 groups based on the proposed cut-off points: light level (<5), subthreshold level (≥5 and <13), and severe level (≥13) [37,38].

Implementation Outcomes

As an objective measurement of the program implementation, the user retention rate of the app (intervention) and the participation rate for each element of the program (control) were calculated. In the intervention group, app usage logs during the intervention period were retrieved from the app's cloud server. The user retention rate was determined weekly; employees who launched the app at least once per week were classified as continuing users.

Participants in the control group were asked whether they had read the policy on physical activity promotion or had watched the psychoeducational video provided by the research team. Participation rates in the physical activity competition were calculated from questionnaire logs registering the amount of physical activity during the period.

For the subjective assessment of the program implementation by the employees, the Implementation Outcome Scale of Digital Mental Health (iOSDMH) was used [39]. This scale consists of 19 items assessing 5 domains: acceptability, appropriateness, feasibility, satisfaction, and potential harm. Each item is rated on a 4-point Likert scale (1=disagree to 4=agree). The scale's reliability and validity have been verified in previous studies [39,40]. Subscale scores were calculated for each domain, and the overall score included the domains of acceptability, appropriateness, feasibility, and satisfaction. Furthermore, the iOSDMH for managers and policymakers was used to assess implementation outcomes by health promotion managers. This scale comprises 14 items covering the same domains as those for the user's version: acceptability, appropriateness, feasibility, satisfaction, and potential harm. Unlike the users' version, the version for managers and policymakers is rated on a 5-point Likert scale. The additional choice is "5=don't know." For score calculation, the average subscale score was calculated after excluding items rated as "don't know."

Demographic Variables

Age group (≤19, 20 - 29, 30 - 39, 40 - 49, 50 - 59, 60 - 69, ≥70 years), gender (men, women, not specified, and other), employment status (full-time, contract, part-time, and dispatched), working shifts (day shift, rotation, and night shift), occupation (manager; professional, engineer, or academic; clerk; sales; service, transportation; construction; production; agriculture, forestry, or fishery; security; and other), and working hours per week (1 - 34, 35 - 40, 41 - 50, 51 - 60, 61 - 65, 66 - 70, ≥71) were recorded as demographic variables. These were used to calculate a propensity score to control baseline imbalances between the intervention and control groups.

Sample Size Calculation

The required sample size was calculated at the individual (employee) level, although the work units were allocated to the 2 groups. The intraclass correlation coefficients for the primary and secondary outcome measures, physical activity, and psychological distress are known to be small in Japanese workplace settings; only 1% of variances are shared at the work-unit level [41,42]. The scientific evidence for using mHealth interventions to promote physical activity and reduce psychological distress is not well established, and the effect sizes are unavailable. In a previous meta-analysis, the standardized mean difference of mHealth interventions for increasing the number of steps and physical activity was estimated at 0.24 [43]. Based on this information, the required sample size was estimated as 274 employees per group (548 in total) using a significance level of .05 (2-sided) and statistical power (1-β) of 0.8 with G*Power (version 3.1.9.2, Heinrich Heine Universität, Düsseldorf) [44]. However, the required sample size could not be reached within the funded period, necessitating premature termination of recruitment and analysis.

Statistical Analysis

At the individual level, the demographic characteristics of the employees at baseline were compared between the intervention and control groups using *t* tests or chi-square tests. To investigate the effectiveness of the ASHARE app on physical activity and psychological distress, generalized linear models were used to estimate the interaction effects of time (baseline and 3-month follow-up) by group (intervention and control). Based on the high follow-up rates in this study, we included data from employees who completed the 3-month follow-up survey in the analysis. Although both physical activity (MET-hours/week) and psychological distress can be treated as continuous variables, their distributions were skewed with high variances. Therefore, we treated them as both continuous and binary variables. For the binary measures, the levels of physical activity and psychological distress were dichotomized into low and moderate or high for physical activity, and light and subthreshold or severe for psychological distress. To ensure comparability between the intervention and control groups, a propensity score was calculated and included as a covariate in the model. The propensity for intervention group allocation was estimated using a logistic regression model. This model included age group, gender, employment status, occupation, working hours, total physical activity, and baseline psychological distress scores as predictive variables. Modeling was implemented using "MIXED" for continuous outcomes and "GENLINMIXED" (binomial, logit link) for binary outcomes using SPSS (version 29.0, IBM Corp). The remaining analyses were performed using the same software.

For the sensitivity analysis, we conducted a per-protocol analysis focusing on employees who continued to use the ASHARE app (intervention) and who participated in all the elements of the multicomponent workplace intervention program conducted in their work units (control). To evaluate the implementation outcomes of each program, the user retention rate for the ASHARE app (intervention) and the participation rate for each program element (control) were compared. In addition, the mean

XSL•FO
RenderX

scores of the implementation outcomes (acceptability, appropriateness, feasibility, and satisfaction) assessed by the employees and health promotion managers were compared between the 2 groups. Furthermore, we compared the demographic characteristics, implementation outcome assessments, and qualitative evaluations between the app users who remained engaged and who dropped out.

Ethical Considerations

The study protocol was registered in the University Hospital Medical Information Network (UMIN) clinical trials registry (UMIN-CTR, ID: UMIN000052374, registered October 10, 2023) and approved by the Kitasato University Medical Ethics Organization (C22-137). Informed consent was obtained from all participants before completing the baseline survey. The data were anonymized when analyzed. As an incentive, health promotion managers and employees received an Amazon gift certificate worth 3,000 yen (US $20.5) upon completion of the 3-month follow-up survey.

Results

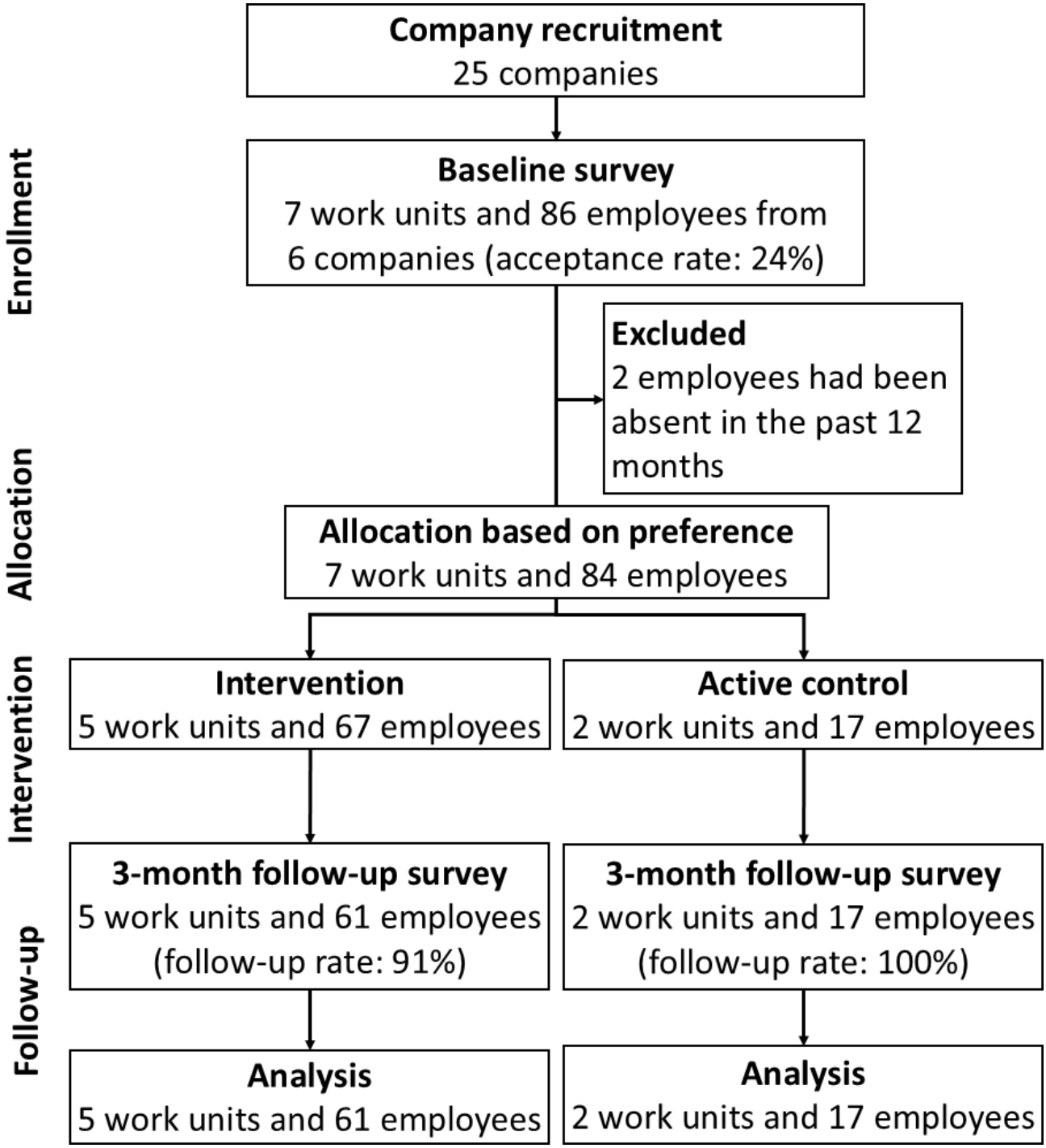
Participant Characteristics

Figure 1 illustrates the participant flow chart. A total of 7 work units and 86 employees from 6 companies participated (acceptance rate: 24%). After excluding 2 employees who had been absent due to illness in the previous 12 months, 5 work units and 67 employees preferred using the ASHARE app, while the other 2 work units and 17 employees preferred following the existing workplace program. The number of employees in the work units ranged from 7 to 22, with the mean number being 12 (SD 6.1). After completing the 3-month programs, 78 employees completed the follow-up survey (61 in the intervention group and 17 in the control group), and their data were included in the analysis (follow-up rate: 93%).

Table 1 presents the baseline demographic characteristics of the included employees. Most participants were full-time (65/78, 83%) and all were day shift (78/78, 100%) workers. Significant baseline imbalances in age group and employment status were found between the intervention and control group.

XSL•FO
RenderX

Figure 1. Participant flowchart.



XSL•FO
RenderX

Table . Characteristics of the participants at baseline (N=78).

Characteristics	Total (N=78), n (%)	Intervention group (n=61), n (%)	Control group (n=17), n (%)	*P* value
Age group (years)				.03
20 - 29	16 (21)	14 (23)	2 (12)	
30 - 39	28 (36)	22 (36)	6 (35)	
40 - 49	14 (18)	14 (23)	0 (0)	
50 - 59	14 (18)	8 (13)	6 (35)	
≥60	6 (8)	3 (5)	3 (18)	
Gender				.93
Men	42 (54)	33 (54)	9 (53)	
Women	36 (46)	28 (46)	8 (47)	
Employment status				.02
Full-time	65 (83)	54 (89)	11 (65)	
Other	13 (17)	7 (12)	6 (35)	
Shift type				—[a]
Day shift	78 (100)	61 (100)	17 (100)	
Occupation				.14
Manager	10 (13)	7 (12)	3 (18)	
Professional, engineer, or academic	28 (36)	26 (43)	2 (12)	
Clerk	37 (47)	26 (43)	11 (65)	
Other	3 (4)	2 (3)	1 (6)	
Working hours per week (hours)				.53
1 - 40	32 (41)	27 (44)	5 (29)	
41 - 50	31 (40)	22 (36)	9 (53)	
51 - 60	13 (17)	10 (16)	3 (18)	
≥61	2 (3)	2 (3)	0 (0)	

[a]Not applicable.

Effect of an Intervention Program on Outcome Measures

Table 2 shows the physical activity and psychological distress levels of the 2 groups at baseline and the 3-month follow-up. Both groups, especially the group following the existing workplace program, showed increased physical activity after completing 3 months of the programs. Employees with moderate and high levels of physical activity saw an increase from 59% (36/61) to 66% (40/61) and from 53% (9/17) to 82% (14/17) at the 3-month follow-up, respectively; however, these changes were not statistically significant in the continuous (*P*=.98) or binary (*P*=.20) analyses.

The mean psychological distress score decreased from 3.87 to 3.75 in the ASHARE app group and increased from 4.18 to 4.59 in the active control group. In addition, the proportion of patients with subthreshold or severe levels of distress increased in the control group (from 29% [5/17] to 47% [8/17]), whereas it was maintained in the intervention group (from 31% [19/61] to 31% [19/61]). However, the group differences were not significant in the generalized linear model adjusted for the propensity score (*P*=.75; *P*=.36).

XSL•FO
RenderX

Table . Changes in physical activity and psychological distress levels after the 3-month interventions (N=78).

Outcomes	Intervention group (n=61)		Control group (n=17)		Effectiveness		
	Baseline	3-month follow-up	Baseline	3-month follow-up	Interaction effect of time by group	SE[a]	*P* value
Physical activity (GPAQ)[b]							
Total amount (MET-hours/week)[c]							
Mean (SD)	20.63 (18.5)	21.58 (20.3)	34.86 (73.2)	36.21 (48.2)	—[d]	—	—
Mean (SE)[e]	23.60 (5.2)	24.54 (4.1)	24.23 (11.5)	25.58 (9.9)	−0.4	12.84	.98
Level of physical activity[f]							
Low, n (%)	25 (41)	21 (34)	8 (47)	3 (18)	−1.16	0.91	.2
Moderate or high, n (%)	36 (59)	40 (66)	9 (53)	14 (82)	—	—	—
Psychological distress (K6)[g]							
Total score							
Mean (SD)	3.87 (3.3)	3.75 (4.3)	4.18 (5.8)	4.59 (4.9)	—	—	—
Mean (SE)[e]	4.13 (0.6)	4.01 (0.6)	3.25 (1.3)	3.67 (1.4)	−0.53	1.62	.75
Level of distress							
Light (≤4)	42 (69)	42 (69)	12 (71)	9 (53)	−0.78	0.84	.36
Subthreshold or severe (≥5)	19 (31)	19 (31)	5 (29)	8 (47)	—	—	—

[a]SE: standard error.

[b]GPAQ: Global Physical Activity Questionnaire.

[c]MET: metabolic equivalent.

[d]Not available.

[e]Estimated mean scores adjusted by propensity score.

[f]High: ≥ 3 days of vigorous-intensity activity of ≥1500 MET-min/week or ≥7 days of any combination of walking or moderate- to vigorous–intensity physical activity of 3000 MET-min/week; moderate: ≥3 days of vigorous-intensity activity of ≥20 min/day or ≥5 days of moderate-intensity activity or walking of ≥30 min/day or ≥5 days of any combination of walking or moderate- to vigorous–intensity physical activity of ≥600 MET-min/week; low: not meeting the criteria for high or moderate.

[g]K6: Kessler Psychological Distress Scale.

Evaluation of User Retention Rate and Implementation Outcomes

Table 3 illustrates the user retention rate of the ASHARE app in the intervention group during the 3-month study period. The weekly user retention rate fell below 50% by the fourth week of the intervention and continued to decline steadily until the eighth week. In the last month, the retention rate stabilized at 20% - 30%. In the final week, the user retention rate was 20% (12/61), which was lower than the participation rate for each component of the control program: managers viewing the written policy (17/17, 100%), employees watching the psychoeducational YouTube (Google) video (9/17, 53%), and participation in the final week's physical activity competition (14/17, 82%). In the control group, the percentage of employees who participated in all the elements conducted in the work units was 53% (9/17). When we compared app users in the intervention group, those who remained engaged were more likely to be in their 50s and employed as professionals, engineers, or academics. In contrast, those who dropped out were more likely to be over 60 and employed as clerks (Table S1 in Multimedia Appendix 1).

Table 4 presents the mean scores for acceptability, appropriateness, feasibility, satisfaction, and harm in the 2 groups, as measured using the iOSDMH. The implementation outcomes, as evaluated by the employees, were less favorable in the intervention group than in the control group, except for harm. Conversely, the implementation scores, as evaluated by health promotion managers, show little difference between the 2 groups. When we compared app users in the intervention group, no significant differences were observed between those who remained engaged and those who dropped out (Table S2 in Multimedia Appendix 2). However, the retention group rated

XSL•FO
RenderX

the app's feasibility slightly more favorably. Qualitative feedback from the dropout group indicated several concerns, including slow app launch times, a lack of specificity in the provided advice, and a desire for enhanced gamification features.

Table . User retention rate of the ASHARE app over 3 months during the intervention period (n=61).

Week	Retention users, n (%)
1	46 (75)
2	42 (69)
3	33 (54)
4	25 (41)
5	20 (33)
6	18 (30)
7	16 (26)
8	14 (23)
9	16 (26)
10	14 (23)
11	12 (20)
12	11 (18)
13	12 (20)

Table . Assessment of the implementation outcomes of the interventions.

Outcomes	Intervention group, mean (SD)	Control group, mean (SD)	*P* value for *t* test	Beta version of ASHARE[a], mean (SD)
Employees (n=78), n	61	17	N/A[b]	23
Overall score (14 items; range 14 - 56)	36.21 (7.2)	48.00 (5.4)	<.001	36.65 (5.1)
Acceptability (3 items; range 3 - 12)	7.05 (1.9)	9.24 (1.5)	<.001	6.87 (1.5)
Appropriateness (4 items; range 4 - 16)	9.87 (2.4)	13.65 (2)	<.001	9.96 (2.3)
Feasibility (6 items; range 6 - 24)	17.00 (3.5)	21.59 (2.3)	<.001	17.74 (2.8)
Satisfaction (1 item; range 1 - 4)	2.30 (0.8)	3.53 (0.5)	<.001	2.09 (0.7)
Harm (5 items; range 5 - 20)	7.10 (2.8)	6.18 (1.5)	.19	6.43 (1.6)
Health promotion managers (n=7)	5	2	N/A	__[c]
Acceptability (4 items; range 1 - 4)	3.3 (0.5)	3.63 (0.5)	.5	—
Appropriateness (4 items; range 1 - 4)	3.35 (0.2)	3.42 (0.1)	.72	—
Feasibility (4 items; range 1 - 4)	3.8 (0.1)	3.75 (0.4)	.76	—
Satisfaction (1 item; range 1 - 4)	2.8 (0.8)	3.5 (0.7)	.35	—
Harm (1 item; range 1 - 4)	1.2 (0.4)	1 (0)	.58	—

[a]Scores were obtained in our previous feasibility trial [25].

[b]Not applicable.

[c]Not available.

XSL•FO
RenderX

Per-Protocol Analysis Among Program Completers

For the sensitivity analysis, the program effectiveness was investigated among 21 completers (12 in the intervention group and 9 in the control group). Table 5 presents the results of the generalized linear modeling. The analyses revealed more favorable changes in physical activity and psychological distress levels in the intervention group compared with the control group. Although these differences were not significant in terms of physical activity, psychological distress significantly decreased. The proportion of the subthreshold or severe distress levels in the intervention group was significantly reduced compared with the control group (coefficient=−3.68, SE 1.65; *P*=.03).

Table . Sensitivity analysis: changes in physical activity and psychological distress among program completers (n=21).

Outcomes	Intervention group (n=12)		Control group (n=9)		Effectiveness		
	Baseline	3-month follow-up	Baseline	3-month follow-up	Interaction effect of time by group	SE[a]	*P* value
Physical activity (GPAQ)[b]							
Total amount (MET-h/week)[c]							
Mean (SD)	28.31 (25.5)	28.92 (25.1)	49.7 (99)	16.1 (9.9)	—[d]	—	—
Mean (SE)[e]	26.9 (20)	27.51 (7.7)	51.58 (23.4)	17.97 (9.4)	34.22	31.1	.28
Level of physical activity[f]							
Low, n (%)	4 (33)	3 (25)	2 (22)	2 (22)	0.41	1.54	.80
Moderate or high, n (%)	8 (67)	9 (75)	7 (78)	7 (78)	—	—	—
Psychological distress (K6)[g]							
Total score							
Mean (SD)	5.58 (4.1)	3.17 (3.3)	2.44 (1.9)	4.89 (4.5)	—	—	—
Mean (SE)[e]	6.47 (1.2)	4.06 (1.3)	1.26 (1.4)	3.70 (1.5)	–4.86	2.23	.04
Level of distress							
Light (≤4)	6 (50)	9 (75)	8 (89)	4 (44)	–3.68	1.65	.03
Subthreshold or severe (≥5)	6 (50)	3 (25)	1 (11)	5 (56)	—	—	—

[a]SE: standard error.

[b]GPAQ: Global Physical Activity Questionnaire.

[c] MET: metabolic equivalent.

[d]Not available.

[e]Estimated mean scores adjusted by propensity score.

[f]High: ≥ 3 days of vigorous-intensity activity of ≥1500 MET-min/week or ≥7 days of any combination of walking or moderate- to vigorous–intensity physical activity of 3000 MET-min/week; moderate: ≥3 days of vigorous-intensity activity of ≥20 min/day or ≥5 days of moderate-intensity activity or walking of ≥30 min/day or ≥5 days of any combination of walking or moderate- to vigorous–intensity physical activity of ≥600 MET-min/week; low: not meeting the criteria for high or moderate.

[g]K6: Kessler Psychological Distress Scale.

Discussion

Principal Findings

Although the per-protocol sensitivity analysis revealed a significant improvement in psychological distress, the primary results did not support the hypothesis that the ASHARE app was more effective than the existing workplace program. Regarding implementation, the user retention rate and employee assessment outcomes were lower than those of active controls across the various measured indicators. This highlights that the app still has several shortcomings in its design and functionalities, including slow launch times, unoptimized content, and a lack of integrated gamification elements. No statistically significant differences between the app and the existing workplace program were observed in the primary and secondary outcomes. Active control, an existing workplace intervention program with environmental changes, successfully increased employee's physical activity levels. The developed app may not be more effective than traditional, face-to-face, multicomponent programs. The lack of a superior effect on increasing physical activity may be attributed to the overlap in behavior change techniques used in both programs. The psychoeducation and competitions in the control program were also designed to encourage self-monitoring and data sharing. Nevertheless, the changes in psychological distress, while not statistically significant, differed between the groups. This may be due to the aim of the app, which encourages the monitoring

XSL•FO
RenderX

of depression and anxiety while highlighting the association between physical activity and mental health through a deep learning model. Although this study did not provide sufficient evidence, it may suggest the potential of the deep learning model for outcome-specific effects in promoting physical activity.

The biggest challenge in implementing mHealth interventions in workplace settings was not addressed in this study. The user retention rate of the ASHARE app, 20% (12/61)at the 3-month follow-up, was not very low; it was higher than those of health care apps on the market (3.5% - 4%) and the 4-week follow-up rate in a previous intervention study (35.7%) [20,21]. Weekly reminders sent by the researchers in this study may have contributed to maintaining the user retention rate. However, this was insufficient to achieve favorable effects in the work units. Furthermore, compared with the active control group, employees rated the app more poorly across the dimensions of acceptability, appropriateness, feasibility, and satisfaction. Despite prerelease modifications, the ASHARE app's overall implementation score (36.21, SD 7.2) was similar to its beta version (36.65, SD 5.1) [25]. Conversely, assessments of the control program were mostly favorable. Discrepancies may have occurred between research and practical settings in the control group. Existing workplace intervention programs require substantial coordination from providers, including drafting written policies, distributing promotional materials (eg, flyers and posters), organizing psychoeducation sessions, and overseeing competitive activities. In this study, researchers managed these responsibilities instead of health promotion managers, effectively exempting the latter from typical program-related efforts. Under these conditions, the acceptability and feasibility of traditional face-to-face multicomponent programs in practical settings may be lower than those in this present study.

Some valuable insights were obtained from the comparison between the retention and dropout groups among app users, as well as from the qualitative feedback provided by those who dropped out. The retention group primarily consisted of younger, professional workers who exhibited higher levels of distress at baseline. These segments of workers could serve as the initial target for promoting the implementation of the app in the workplace. In contrast, older workers may require more intensive support to encourage app engagement. Improvements to the app's design are also necessary: the ASHARE app exchanges data with the server upon launch to predict mental health scores, which may contribute to slower startup times. In addition, since the program is written in Python, an interpreted language, its processing speed may be slower compared with compiled languages such as C++. Optimizing the program or converting it to a compiled language could help address user dissatisfaction related to slow launch times. Following a user-centered approach, enhancing the specificity and diversity of feedback messages, as well as incorporating gamification elements, such as badges, points, narratives, and characters could further improve user engagement [45]. An alternative explanation is that work-unit level participation could make them feel compelled, leading to fatigue in using the app.

A potential area for improving the app's effectiveness is addressing the evaluation discrepancy between app users (employees) and health promotion managers. While managers did not rate the app poorly regarding implementation outcomes, employees suggested room for improvement. This discrepancy indicates an implementation gap between managers and employees, potentially affecting the app's overall effectiveness. This suggestion is inconsistent with the previous findings from the qualitative study in Japan [23]. Future implementations may benefit from additional strategies, such as informational sessions on the app's objectives, benefits, and usage to enhance employees' understanding and encourage greater commitment to the program.

The per-protocol analysis indicated a significant reduction in psychological distress in the intervention group. Since the results were based on a post-hoc analysis and the baseline imbalance was more noticeable than in the overall sample, they must be carefully interpreted. This reduction is merely a regression to the mean. Despite these limitations, the app's efficacy in reducing psychological distress may be favorable if employees continuously use the app for 3 months. Furthermore, based on the characteristics of the program completers, the app may be particularly efficacious in users with lower activity and higher psychological distress. This aligns with findings from our feasibility trial, where the number of participants experiencing severe psychological distress decreased after using the app [25]. Future implementations may need to identify appropriate workplace intervention targets.

This study had a few limitations. First, we could not reach the required sample size within the funded period, leading to the termination of recruitment at our convenience. This limitation led to low statistical power in detecting the significant improvement in outcomes and disturbed the hypothesis-testing framework. In addition, the imbalance in allocation between the intervention and control groups may increase the risk of both type I and type II errors. The nonrandomized design has inherent biases that could not be adjusted even if propensity scores were used in statistical modeling. Other unmeasured confounders may include motivation for health promotion, job stress, and medical history. We adopted a nonrandomized design because randomization would not have agreed with the work units' decision-making process. Therefore, additional efforts are needed to ensure the success of cluster-randomized controlled trials. The outcome measures and implementation outcomes rely heavily on self-reported data, which are prone to recall bias and social desirability bias. Even if the program had been offered to the control group, given the study setting designed to investigate the effects of the newly developed mHealth app, participants in the intervention group would still be more likely to report favorable changes. Since most participants were full-time workers on day shifts, the generalizability of the findings is limited. In addition, compared with the overall population of Japanese workers, the participants in this study included a higher proportion of professionals. Other working populations, especially night-shift workers, were not covered in this study.

Conclusions

The ASHARE app did not show superior effectiveness compared with the existing workplace program. An implementation gap

XSL•FO
RenderX

may exist between health promotion managers and employees, possibly contributing to the app's low user retention rate. Future research should focus on improving the app design identified in this study, as well as examining the effectiveness of strategies to get engagement from managers and from segments of employees with favorable responses in the workplace at an early stage. Also for apps already implemented in society, strategies may be needed to bridge the engagement gap between decision-making managers and employees. Furthermore, efforts to ensure the success of cluster-randomized controlled trials will be necessary to generate more robust and reliable findings.

Acknowledgments

The authors thank Masahiro Morimoto and Hideki Miura for their contributions to the project's conceptualization and participant recruitment. This study was supported by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JP20K19671; JP24K20247), the Japan Agency for Medical Research and Development (JP21de0107006), and the 38th Research Grant from the Meiji Yasuda Life Foundation of Health and Welfare. The authors declare the use of generative artificial intelligence technology (ChatGPT) in editing the manuscript text for this study.

Data Availability

The raw datasets and program codes used for statistical analyses are not publicly accessible but are available for secondary use from the corresponding author upon reasonable request and with additional approval from the ethics committee.

Authors' Contributions

KW spearheaded the conceptualization of the project, developed the methodology, and oversaw data analysis. KW and AT drafted the initial version of the manuscript, while KW, AT, SO, and MS contributed to its review and revision. All authors reviewed and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the app users who retained and dropped out (n=61).
[PDF File, 122 KB - mhealth_v13i1e70473_app1.pdf]

Multimedia Appendix 2
Assessment of the implementation outcomes among the app users who retained and dropped out (n=61).
[PDF File, 105 KB - mhealth_v13i1e70473_app2.pdf]

Checklist 1
A completed TREND checklist.
[PDF File, 1195 KB - mhealth_v13i1e70473_app3.pdf]

Checklist 2
A completed StaRI checklist.
[PDF File, 253 KB - mhealth_v13i1e70473_app4.pdf]

References

1. Guidelines on mental health at work. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240053052 [accessed 2025-05-01]
2. Hilton MF, Whiteford HA, Sheridan JS, et al. The prevalence of psychological distress in employees and associated occupational risk factors. J Occup Environ Med 2008 Jul;50(7):746-757. [doi: 10.1097/JOM.0b013e31817e9171] [Medline: 18617830]
3. Fan ZJ, Bonauto DK, Foley MP, Anderson NJ, Yragui NL, Silverstein BA. Occupation and the prevalence of current depression and frequent mental distress, WA BRFSS 2006 and 2008. Am J Ind Med 2012 Oct;55(10):893-903. [doi: 10.1002/ajim.22094] [Medline: 22821712]
4. Fushimi M, Saito S, Shimizu T, Kudo Y, Seki M, Murata K. Prevalence of psychological distress, as measured by the Kessler 6 (K6), and related factors in Japanese employees. Community Ment Health J 2012 Jun;48(3):328-335. [doi: 10.1007/s10597-011-9416-7] [Medline: 21547569]

XSL•FO
RenderX

5. Duchaine CS, Aubé K, Gilbert-Ouimet M, et al. Psychosocial stressors at work and the risk of sickness absence due to a diagnosed mental disorder: a systematic review and meta-analysis. JAMA Psychiatry 2020 Aug 1;77(8):842-851. [doi: 10.1001/jamapsychiatry.2020.0322] [Medline: 32236498]
6. Chisholm D, Sweeny K, Sheehan P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424. [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
7. Kessler RC. The costs of depression. Psychiatr Clin North Am 2012 Mar;35(1):1-14. [doi: 10.1016/j.psc.2011.11.005] [Medline: 22370487]
8. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev 2015;9(3):366-378. [doi: 10.1080/17437199.2015.1022901] [Medline: 25739893]
9. Chu AHY, Koh D, Moy FM, Muller-Riemenschneider F. Do workplace physical activity interventions improve mental health outcomes? Occup Med 2014 Jun 1;64(4):235-245. [doi: 10.1093/occmed/kqu045]
10. Joyce S, Modini M, Christensen H, et al. Workplace interventions for common mental disorders: a systematic meta-review. Psychol Med 2016 Mar;46(4):683-697. [doi: 10.1017/S0033291715002408] [Medline: 26620157]
11. Hosseinpour M, Terlutter R. Your personal motivator is with you: a systematic review of mobile phone applications aiming at increasing physical activity. Sports Med 2019 Sep;49(9):1425-1447. [doi: 10.1007/s40279-019-01128-3] [Medline: 31144235]
12. Price M, Yuen EK, Goetter EM, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436. [doi: 10.1002/cpp.1855] [Medline: 23918764]
13. Mönninghoff A, Kramer JN, Hess AJ, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699. [doi: 10.2196/26699] [Medline: 33811021]
14. Tong HL, Alnasser A, Alshahrani NZ, et al. The use of mobile technologies to promote physical activity and reduce sedentary behaviors in the Middle East and North Africa region: systematic review and meta-analysis. J Med Internet Res 2024 Mar 19;26:e53651. [doi: 10.2196/53651] [Medline: 38502160]
15. Zhang M, Wang W, Li M, Sheng H, Zhai Y. Efficacy of mobile health applications to improve physical activity and sedentary behavior: a systematic review and meta-analysis for physically inactive individuals. Int J Environ Res Public Health 2022 Apr 18;19(8):4905. [doi: 10.3390/ijerph19084905] [Medline: 35457775]
16. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046. [doi: 10.2196/17046] [Medline: 32186518]
17. Murawski B, Plotnikoff RC, Rayward AT, et al. Efficacy of an m-Health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
18. Baumann H, Heuel L, Bischoff LL, Wollesen B. Efficacy of individualized sensory-based mHealth interventions to improve distress coping in healthcare professionals: a multi-arm parallel-group randomized controlled trial. Sensors (Basel) 2023 Feb 19;23(4):2322. [doi: 10.3390/s23042322] [Medline: 36850920]
19. Díaz-Silveira C, Santed-Germán M, Burgos-Julián FA, Ruiz-Íñiguez R, Alcover CM. Differential efficacy of physical exercise and mindfulness during lunch breaks as internal work recovery strategies: a daily study. European Journal of Work and Organizational Psychology 2023 Jul 4;32(4):549-561. [doi: 10.1080/1359432X.2023.2198706]
20. A report on the project related to optimizing the commercial transaction and service environment. Ministry of Economy, Trade and Industry, Japan. 2023. URL: https://www.meti.go.jp/meti_lib/report/2022FY/000193.pdf [accessed 2025-04-30]
21. Haque MS, Kangas M, Jämsä T. A persuasive mHealth behavioral change intervention for promoting physical activity in the workplace: feasibility randomized controlled trial. JMIR Form Res 2020 May 4;4(5):e15083. [doi: 10.2196/15083] [Medline: 32364506]
22. Watanabe K, Tsutsumi A. Workers' perceptions of mHealth services for physical activity and mental health: a qualitative study using a text-mining method. Environ Occup Health Pract 2023;5(1):2022-0022-OA. [doi: 10.1539/eohp.2022-0022-OA] [Medline: 40059926]
23. Saito J, Odawara M, Takahashi H, et al. Barriers and facilitative factors in the implementation of workplace health promotion activities in small and medium-sized enterprises: a qualitative study. Implement Sci Commun 2022 Mar 2;3(1):23. [doi: 10.1186/s43058-022-00268-4] [Medline: 35236511]
24. Watanabe K, Tsutsumi A. The passive monitoring of depression and anxiety among workers using digital biomarkers based on their physical activity and working conditions: 2-week longitudinal study. JMIR Form Res 2022 Nov 30;6(11):e40339. [doi: 10.2196/40339] [Medline: 36449342]
25. Watanabe K, Okusa S, Sato M, Miura H, Morimoto M, Tsutsumi A. mHealth intervention to promote physical activity among employees using a deep learning model for passive monitoring of depression and anxiety: single-arm feasibility trial. JMIR Form Res 2023 Nov 17;7:e51334. [doi: 10.2196/51334] [Medline: 37976094]
26. ASHARE. URL: https://www.ashare-hp.jp/ [accessed 2025-04-14]

XSL•FO
RenderX

27. Watanabe K, Kawakami N. Effects of a multi-component workplace intervention program with environmental changes on physical activity among Japanese white-collar employees: a cluster-randomized controlled trial. Int J Behav Med 2018 Dec;25(6):637-648. [doi: 10.1007/s12529-018-9747-7] [Medline: 30194533]
28. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004 Mar;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
29. Pinnock H, Barwick M, Carpenter CR, et al. Standards for Reporting Implementation Studies (StaRI) Statement. BMJ 2017 Mar 6;356:i6795. [doi: 10.1136/bmj.i6795] [Medline: 28264797]
30. Anderson LM, Quinn TA, Glanz K, et al. The effectiveness of worksite nutrition and physical activity interventions for controlling employee overweight and obesity: a systematic review. Am J Prev Med 2009 Oct;37(4):340-357. [doi: 10.1016/j.amepre.2009.07.003] [Medline: 19765507]
31. Lin YP, McCullagh MC, Kao TS, Larson JL. An integrative review: work environment factors associated with physical activity among white-collar workers. West J Nurs Res 2014 Feb;36(2):262-283. [doi: 10.1177/0193945913503417] [Medline: 24065693]
32. Beresford SAA, Bishop SK, Brunner NL, et al. Environmental assessment at worksites after a multilevel intervention to promote activity and changes in eating: the PACE project. J Occup Environ Med 2010 Jan;52 Suppl 1(Suppl 1):S22-S28. [doi: 10.1097/JOM.0b013e3181c7512c] [Medline: 20061883]
33. Physical activity surveillance. World Health Organization. URL: https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/physical-activity-surveillance [accessed 2025-04-30]
34. Bull FC, Maslin TS, Armstrong T. Global Physical Activity Questionnaire (GPAQ): nine country reliability and validity study. J Phys Act Health 2009 Nov;6(6):790-804. [doi: 10.1123/jpah.6.6.790] [Medline: 20101923]
35. Global Physical Activity Questionnaire (GPAQ): analysis guide. World Health Organization. URL: https://cdn.who.int/media/docs/default-source/ncds/ncd-surveillance/gpaq-analysis-guide.pdf?sfvrsn=1e83d571_2 [accessed 2025-04-30]
36. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003 Feb;60(2):184-189. [doi: 10.1001/archpsyc.60.2.184] [Medline: 12578436]
37. Furukawa TA, Kawakami N, Saitoh M, et al. The performance of the Japanese version of the K6 and K10 in the World Mental Health Survey Japan. Int J Methods Psychiatr Res 2008;17(3):152-158. [doi: 10.1002/mpr.257] [Medline: 18763695]
38. Sakurai K, Nishi A, Kondo K, Yanagida K, Kawakami N. Screening performance of K6/K10 and other screening instruments for mood and anxiety disorders in Japan. Psychiatry Clin Neurosci 2011 Aug;65(5):434-441. [doi: 10.1111/j.1440-1819.2011.02236.x] [Medline: 21851452]
39. Sasaki N, Obikane E, Vedanthan R, et al. Implementation Outcome Scales for Digital Mental Health (iOSDMH): scale development and cross-sectional study. JMIR Form Res 2021 Nov 23;5(11):e24332. [doi: 10.2196/24332] [Medline: 34817391]
40. Obikane E, Sasaki N, Imamura K, et al. Usefulness of Implementation Outcome Scales for Digital Mental Health (iOSDMH): experiences from six randomized controlled trials. Int J Environ Res Public Health 2022 Nov 27;19(23):15792. [doi: 10.3390/ijerph192315792] [Medline: 36497867]
41. Watanabe K, Otsuka Y, Shimazu A, Kawakami N. The moderating effect of health-improving workplace environment on promoting physical activity in white-collar employees: a multi-site longitudinal study using multi-level structural equation modeling. J Occup Environ Med 2016 Feb;58(2):178-184. [doi: 10.1097/JOM.0000000000000590] [Medline: 26849262]
42. Eguchi H, Tsutsumi A, Inoue A, Hikichi H, Kawachi I. Association of workplace social capital with psychological distress: results from a longitudinal multilevel analysis of the J-HOPE Study. BMJ Open 2018 Dec 22;8(12):e022569. [doi: 10.1136/bmjopen-2018-022569] [Medline: 30580264]
43. Feter N, Dos Santos TS, Caputo EL, da Silva MC. What is the role of smartphones on physical activity promotion? A systematic review and meta-analysis. Int J Public Health 2019 Jun;64(5):679-690. [doi: 10.1007/s00038-019-01210-7] [Medline: 30758514]
44. Faul F, Erdfelder E, Lang AG, Buchner AG. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
45. Brigham TJ. An introduction to gamification: adding game elements for engagement. Med Ref Serv Q 2015;34(4):471-480. [doi: 10.1080/02763869.2015.1082385] [Medline: 26496401]

Abbreviations

GPAQ: Global Physical Activity Questionnaire
iOSDMH: Implementation Outcome Scale of Digital Mental Health
K6: 6-item Kessler Psychological Distress Scale
MET: metabolic equivalent
StaRI: Standards for Reporting Implementation Studies
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs

XSL•FO
RenderX

UMIN: University Hospital Medical Information Network

Edited by M Giurgiu; submitted 28.12.24; peer-reviewed by T Yoshikawa, YH Lee; revised version received 20.03.25; accepted 21.03.25; published 06.05.25.

Please cite as:
Watanabe K, Sato M, Okusa S, Tsutsumi A
Effectiveness and Implementation Outcomes of an mHealth App Aimed at Promoting Physical Activity and Improving Psychological Distress in the Workplace Setting: Cluster-Level Nonrandomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e70473
URL: https://mhealth.jmir.org/2025/1/e70473
doi:10.2196/70473

© Kazuhiro Watanabe, Mitsuhiro Sato, Shoichi Okusa, Akizumi Tsutsumi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility and Effects of a Gait Assistance and Gait Resistance Training Program Using a Walking-Assist Wearable Robot for Community-Dwelling Older Adults: Single-Group, Pre-, and Posttest Study

Eunhee Cho[1], PhD; Sinwoo Hwang[2,3], MSN; Seok-Jae Heo[4], PhD; Bokman Lim[5], PhD; Jewoo Lee[5], MS; Younbaek Lee[5], MS

[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]College of Nursing, Graduate School, Yonsei University, 50-1 Yonsei-ro, Seodaemoon-gu, Seoul, Republic of Korea

[3]Army Cadet Military School, Goesan, Republic of Korea

[4]Biostatistics Collaboration Unit, Department of Biomedical Systems Informatics, College of Medicine, Yonsei University, Seoul, Republic of Korea

[5]WIRobotics Incorporation, Youngin, Republic of Korea

Corresponding Author:
Sinwoo Hwang, MSN
College of Nursing, Graduate School, Yonsei University, 50-1 Yonsei-ro, Seodaemoon-gu, Seoul, Republic of Korea

Abstract

Background: Two-thirds of people aged 65 years and older may require help with daily activities such as eating, bathing, and getting in and out of bed or a chair. Walking-assist wearable robots have shown significant improvements in physical function in controlled settings for patients.

Objective: In this study, we aimed to assess the feasibility and the effect of a gait assistance and gait resistance training program using a walking-assist wearable robot for community-dwelling older adults.

Methods: A total of 23 community-dwelling older adults aged 65 years and older (30 participants recruited, 7 dropped out) enrolled in a 12-session, 6-week gait assistance and gait resistance training program using a walking-assist wearable robot. A single-group, pre- and posttest design was employed to evaluate the feasibility based on program adherence and effectiveness. The primary and secondary outcomes for evaluating effectiveness were walking speed and functional performance, respectively.

Results: Regarding the feasibility, the average number of sessions attended was 11.7 out of 12, indicating a mean adherence rate of 97.8%. Linear mixed model analysis revealed significant improvements in walking speed and functional performance at the end of the program compared with baseline. Specifically, the walking speed measured using the 10-Meter Walk Test, which includes self-selected velocity and fastest safe velocity, improved by a mean of 0.15 (SD 0.13) m/s (P<.001) and 0.15 (SD 0.17) m/s (P<.001), respectively. Functional performance also improved, with faster performance in Timed Up-and-Go (mean −0.63, SD 0.92 s; P=.003) and Four Square Step Test (mean −1.71, SD 1.64 s; P≤.001). Leg muscle strength increased across all measured domains, including plantarflexion (mean +7.29, SD 4.92; P=.004), hip adduction (mean +3.03, SD 2.73; P≤.001), hip extension (mean +2.63, SD 2.50; P≤.001), knee extension (mean +2.33, SD 3.12; P≤.001), knee flexion (mean +2.19, SD 2.17; P≤.001), dorsiflexion (mean +2.10, SD 3.06; P≤.001), hip abduction (mean +1.59, SD 1.92; P=.002), and hip flexion (mean +0.90, SD 1.56; P≤.001).

Conclusions: This study stands out for applying gait assistance and resistance training across various terrains, unlike previous studies that only tested gait assistance in controlled environments. The results demonstrated significant improvements in walking speed and functional performance in older adults, suggesting the effectiveness of preventive health care services using a walking-assist wearable robot as an intervention that can contribute to improving independent functioning and frailty among community-dwelling older adults.

(JMIR Mhealth Uhealth 2025;13:e58142) doi:10.2196/58142

KEYWORDS

older adult; exercise; wearable robot; walking speed; physical functional performance; muscle strength

XSL•FO
RenderX

Introduction

Background

Global life expectancy has gradually increased from 66.8 years in 2000 to 73.4 years in 2019, although health-adjusted life expectancy (HALE) was only 63.7 years in 2019 (86.3% of life expectancy) [1]. In other words, HALE significantly lags behind life expectancy, with an increase in HALE (5.4 y) failing to keep up with the increase in life expectancy (6.6 y). This implies that people are experiencing prolonged periods of poor health in their older years. In particular, age-related declines in physical motor function can lead to limitations in activities of daily living (ADL) such as dressing, washing, getting out of bed, and using the restroom [2]. According to a recent report by the World Health Organization [3], approximately two-thirds of people aged 65 years and older may require help with daily activities such as eating, bathing, and getting in and out of bed or a chair. Among those aged 90 years and older, 96% may require assistance to maintain their daily functioning.

As awareness of the importance of physical activity for older adults increases, the US Department of Health and Human Services, which oversees the Office of Disease Prevention and Health Promotion [4], annually manages the physical activity rate of older adults with physical or cognitive health problems as a Healthy People 2030 indicator. In 2018, 41.3% of older adults were found to be physically active, and the target was set at 51%, which is managed as a national health policy indicator. Physical independence is primarily centered on gait. As older adults age, they tend to experience weakened lower extremity muscle strength and shorter stride lengths, resulting in slower walking speeds and difficulty maintaining balance and stability, which can lead to falls [5]. Furthermore, decreased walking speed is associated with functional loss and mortality, and is frequently used as not only a predictor of fall risk and disability but also an indicator of frailty in older adults [6]. Regular physical activity plays an important role in maintaining the health and independent functioning of older adults and has been linked to improvements in cognitive impairment and mental health. Previous studies [7-9] have shown that physical inactivity in older adults is a major contributor to the development of disuse syndrome, a vicious cycle that ultimately leads to disability in performing ADL if appropriate interventions are not staged.

Assist robots are recognized as a revolutionary technology that significantly enhances physical activity capabilities and provides rehabilitation or retraining opportunities for people with limited mobility due to injury. Wearable robots have been used to improve muscle strength in the general population; assist workers with mobility in industries such as construction, firefighting, or the military; and assist and rehabilitate people with gait weakness [10]. Unlike conventional robots, the latest commercialized robots can be worn and used to assist walking or enhance muscle strength based on the wearer's posture information and gait data. Wearing a robot can increase walking efficiency by more than 7% [11]. Furthermore, it can be used without restrictions when sitting in a chair or walking up and down stairs.

Previous studies evaluating the effects of training with walking-assist wearable robots in older adults have reported improvements in physical function, including increased walking speed and stride length [11-14] and increased muscle strength [13,15]. These findings suggest that physical activity promotion programs using walking-assist wearable robots can be extended to older adults in the community, although they have mostly been used for rehabilitation of patients with spinal cord injury [12], multiple sclerosis [16,17], and stroke [13]. A lack of research on preventive care for older adults remains. While some studies have targeted the general older adult population, they have involved single interventions in controlled laboratory settings, limiting their applicability to community settings [11,18]. Improving the HALE of community-dwelling older adults necessitates further studies that can evaluate the effectiveness of gait assistance training in a systematic physical activity promotion program. Progressive resistance training has been shown to positively impact muscle strength and functional limitations in older adults [19]. Thus, maximizing the effect of improving older adults' functional mobility and functional performance by applying the gait resistance function of wearable robots may not only assist gait but also strengthen muscle strength. Confirming the value of using walking-assist wearable robots to enhance walking speed and functional performance of community-dwelling older adults in the aging era may contribute to a preventive strategy that helps maintain independent physical activity while also improving HALE in the future.

In this study, we conducted a 6-week physical activity program incorporating gait assistance and gait resistance training for community-dwelling older adults using a walking-assist wearable robot, aiming to assess the feasibility and the effect on walking speed and functional performance, and to evaluate the validity of the intervention approach.

Purpose

The purpose of this study was to identify the feasibility and effects of a gait assistance and gait resistance training program using a walking-assist wearable robot on walking speed and functional performance for community-dwelling older adults. The primary outcome of this study was walking speed, which served as the main indicator of functional mobility. Secondary outcomes included comprehensive evaluation of the participants' functional performance, such as the Short Physical Performance Battery, Timed Up-and-Go, Four Square Step Test, Functional Reach Test, and measures of muscle strength.

Methods

Study Design and Setting

This study used a single-group, pretest-posttest design to determine the feasibility of program adherence and evaluate the effectiveness of a gait assistance and gait resistance training program using a walking-assist wearable robot on community-dwelling older adults. It was conducted in collaboration with a community service center and a senior living community to recruit community-dwelling older adults.

XSL•FO
RenderX

Participants

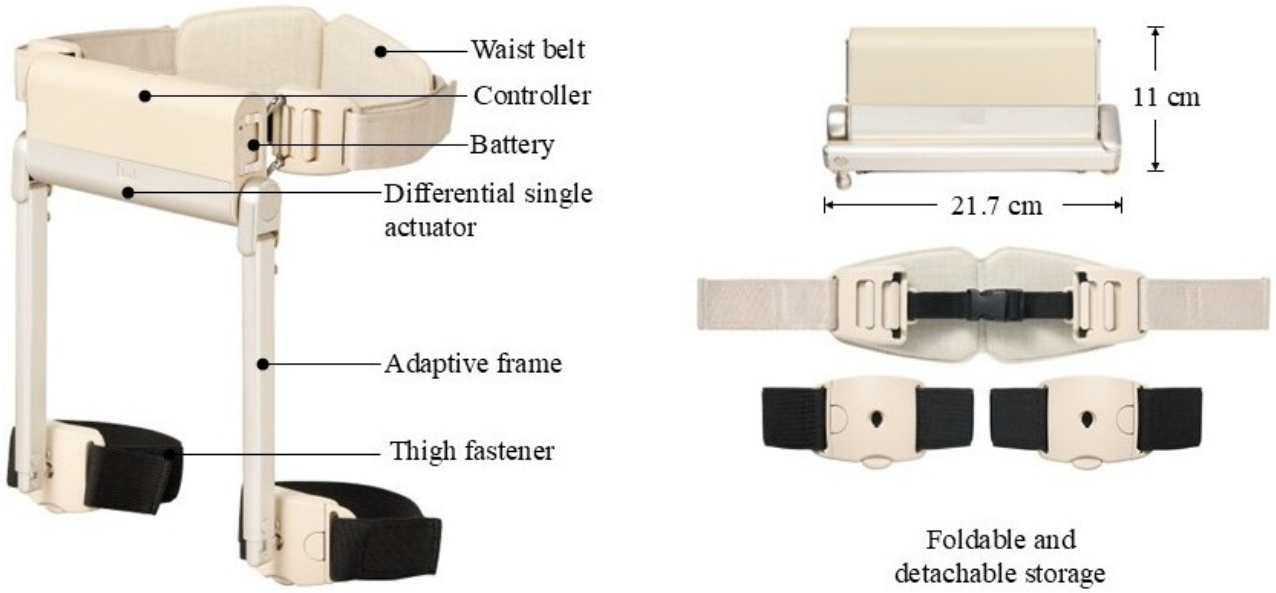
The participants were older adults living in the community, aged 65-90 years old, who were able to walk independently for more than 30 minutes, who understood the contents of the study, and who voluntarily agreed to participate. For the purpose of this study, it was essential for participants to be able to wear a wearable robot and undergo training. Thus, the specific exclusion criteria were as follows. First, individuals with difficulty walking due to vision defects, fractures, and so on. Second, individuals with body types that prevent them from wearing the robot, such as a height of 4.59 feet (140 cm) or less or 5.91 feet (185 cm) or more, or severe obesity with a BMI of 35 or more. Third, individuals with heart and circulatory conditions that may affect walking training. Fourth, individuals at a high risk of falling during walking training, owing to severe dizziness and having experienced a fall within the preceding 2 months.

The sample size was calculated using the G*Power 3.1 program for a paired *t* test with a significance level of *P*=.05, power of 0.80, and effect size of 0.59 [20], specifically based on the primary outcome of this study, which is walking speed. The minimum sample size required was 25, and a total of 30 participants underwent convenience sampling to accommodate an approximate 20% dropout rate. During the 6 weeks of the program, 7 participants withdrew from the study for reasons such as lagging behind in pace and feeling overwhelmed by the group exercise, leaving 23 participants in the final analysis.

Gait Assistance and Gait Resistance Training Program

The walking-assist wearable robot used in this study not only assists gait by recognizing the intention of the motion, thereby reducing the energy required for walking metabolism, but also strengthens muscles by adjusting the gait resistance intensity (Figures 1 and 2). The wearable robot (WIRobotics Inc) is ultralightweight at 1.6 kilograms, easy to wear, and can detect postures such as walking, climbing or descending stairs, and sitting. Its body-compatible articulation structure does not limit the ability to change positions, such as sitting or lying down [21]. Unlike conventional robots, the main body and actuation parts are not located at the hip joints and back to avoid restricting body movements and to facilitate movement even in cramped spaces. In addition, the device is portable, lightweight, easy to use, and cost-effective, as it uses a single actuator, unlike conventional dual-actuator motors. Korea certification and electromagnetic compatibility tests have been completed, confirming the device's safety and reliability.

Figure 1. The wearable robot and its components.



XSL•FO
RenderX

Figure 2. Wearing appearance and application of the wearable robot in an outdoor walking group exercise.

The gait assistance and gait resistance training program lasted 6 weeks and consisted of biweekly group exercise sessions. The gait assistance mode can be configured at 3 intensity levels, with level 1, level 2, and level 3 corresponding to peak assistance torques of approximately 4.0 Nm, 5.0 Nm, and 6.0 Nm, respectively. The intensity was set to the highest comfortable level for each participant, with level 2 being the most commonly used during the exercise sessions. During gait assistance, the device provides synchronized hip flexion and extension assistance, enhancing the wearer's natural gait by applying positive power transfer from the device to the wearer. Conversely, the gait resistance mode functions in opposition to the assistance mode, applying resistance to hip movements that counteract the wearer's gait. This mode facilitates negative power generation, where power is transferred from the wearer to the device, thereby increasing the difficulty of walking, like moving through water. The resistance levels are also adjustable, with peak torques of 1.5 Nm, 2.5 Nm, and 4.0 Nm for level 1, level 2, and level 3, respectively. The robot's interface is designed for ease of use, allowing mode transitions—such as assistance, resistance, or rest (0 torque)—with a single button press. Intensity adjustments within each mode can be managed by double-pressing the button, cycling through level 1 to level 3. This user-friendly control enables participants to conveniently modify assistance or resistance levels during training sessions, following guided instructions, and facilitates engagement in group exercise activities. The first 3 weeks focused on improving walking speed and correcting walking posture using only the wearable robot's gait assistance function, while the fourth week included 10 minutes of walking exercises with added gait resistance, increasing by 5 minutes per week to 15 minutes in the fifth week and 20 minutes in the sixth week. Participants could self-regulate the intensity of gait resistance and complete the program at a lower intensity if the exercise intensity was too much for them.

Each session of the program consisted of 10 minutes of warm-up exercises, 40 minutes of walking exercises while wearing the walking-assist wearable robot, and 10 minutes of cool-down exercises after removing the robot, totaling 1 hour. Before the training program, participants' blood pressure, blood glucose, and physical abnormalities (skin damage, joint pain, muscle pain, etc) were checked by a nurse and physiotherapist at the institution. If no abnormalities were detected, the researchers helped the participants wear the walking-assist wearable robots and proceeded with the training program, which involved walking on flat surfaces, inclines, and stairs. The exercise was conducted in settings such as parks and walking paths within senior residential areas, covering various terrains including flat paths, slopes, and staircases. The exercise intensity was based on the speed at which the participant could walk for 40 minutes, with instructions to maintain a natural walking pace and stride.

Feasibility (Program Adherence)

Feasibility was assessed based on program adherence, which was calculated as the number of sessions attended out of the total 12 exercise sessions for each participant.

Effectiveness Evaluation: Primary Outcome (Walking Speed)

The 10-Meter Walk Test (10MWT) comprises 2 components—walking at normal speed (10MWT for self-selected velocity) and walking at maximum speed (10MWT for fastest safe velocity). It involves measuring the time taken to walk 10 meters and calculating the time taken to walk the middle section of 10 meters, excluding the 2.5 meters of acceleration and deceleration at the beginning and end. These times are then converted to speed (m/s). A higher distance per second after conversion indicates a faster walking speed. Intra- and interrater reliability were 0.88 and 0.99, respectively [22].

Effectiveness Evaluation: Secondary Outcomes

Short Physical Performance Battery

The Short Physical Performance Battery (SPPB) is a tool developed in a multicenter study conducted by the National Institute of Aging in the United States [23]. It involves a brief physical performance test that assesses lower extremity function and includes 3 components—upright balance test, walking speed, and getting up from a chair. Each task is assigned a score of 0 to 4 [23]. The upright balance test consists of a side-by-side stance, a semitandem stance, and a tandem stance, each of which is tested in sequence and assessed by the examiner. Each stance is to be maintained for at least 10 seconds after being

XSL•FO
RenderX

demonstrated to the participant. The walking speed item asks the participant to walk 4 meters at their normal walking speed, and the time taken to walk is measured. The getting up from a chair test is evaluated by measuring the time it takes to stand up and sit down from a chair 5 times with arms crossed in front of the chest. A higher SPPB total score indicates better physical performance.

Timed Up-and-Go

The Timed Up-and-Go (TUG) is a get-up-and-walk test to assess dynamic balance and mobility. It measures the time it takes to get up from a chair, walk 6 meters round trip, and sit down. The shorter the time (sec) taken to complete the TUG test, the better the dynamic balance and mobility. It has a reported intra- and interrater reliability of 0.99 and 0.98, respectively [24].

Four Square Step Test

The Four Square Step Test (FSST) is a quadrant step test that uses 4 cylindrical rods 90 cm long and 1 cm in diameter to create a cross-shaped square on the floor. It measures the time taken for the participant to stand in square 1 and then move as quickly as possible clockwise to 2 (forward), 3, 4 (backward), and 1, and then counterclockwise to 4, 3, 2, and 1 [25]. The shorter the time taken (sec) to complete the FSST test, the higher the level of dynamic balance and mobility. It has a reported intra- and interrater reliability of 0.83 and 0.99, respectively [26].

Functional Reach Test

The Functional Reach Test (FRT) assesses balance ability. It measures the distance participants can reach with their arms maximally extended in front of them, with a reported interrater reliability of 0.98 [27]. A longer distance (cm) measured with the arm outstretched in the FRT test indicates better dynamic balance ability.

Muscle Strength

Muscle strength was measured at the hip, knee, and ankle using a digital dynamometer, micro FET2 (Hoggan Scientific; range 0 - 300 lb). Strength was measured in a seated and supine position using a bed, and 8 items related to strength were measured in hip flexion, hip extension, hip adduction, hip abduction, knee flexion, knee extension, dorsiflexion, and plantarflexion according to a standardized protocol. Higher values indicated greater strength of the muscle.

Data Collection

The program intervention spanned 6 weeks, from September 20 to October 27, 2023, at the community service center and from September 18 to October 25, 2023, at the senior living community. A preassessment of primary and secondary outcomes was conducted before the program, and a postassessment was conducted after. The pre- and postassessments were collected by the same researcher.

Ethical Considerations

This study was conducted after receiving approval from the Public Institutional Review Board of the National Center for Bioethics Policy (P01-202309-01-032). Interested participants were recruited from the community service center and the senior living community in Gyeonggi-do with previous approval of the managers and the cooperation of visiting nurses and physiotherapists in these institutions. Those who were willing to participate in the study were asked to contact the research director, who explained the purpose, method, and confidentiality in detail, first over the phone and then in person. Written consent (Multimedia Appendix 1) was obtained from all participants, an orientation session was conducted on how to use the wearable robot, and the participation schedule was coordinated. After participating in the program, light refreshments (such as water and bread valued at less than US $7) were provided to support hydration and energy replenishment. Participants who completed the training program were given a certificate and a small gift valued at approximately US $14 as appreciation for their participation in the study. Research data were stored in a place accessible only to the researcher, and electronic files were encrypted to protect the rights of research participants. Furthermore, all participant data were fully anonymized to prevent the identification of individuals. Consent to publish their image has been obtained from all individuals pictured in Figure 2.

Statistical Analysis

The general characteristics of the participants are presented as mean (SD) for continuous variables or number (percentage) for categorical variables. Comparisons of mean differences in the primary and the secondary outcome before and after the program were analyzed using paired *t* tests. The normality of the pre-post differences was assessed using histograms and Q-Q plots. While the SPPB scores exhibited some deviations from a normal distribution, their approximately symmetrical pattern supported the use of a parametric test (paired *t* test) for the analysis. We also used linear mixed models to adjust for baseline value, gender, age, and BMI while accounting for repeated measures. Furthermore, 2-sided *P* values less than .05 were considered statistically significant. All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc) and R software (version 4.2.0; R Foundation for Statistical Computing).

Results

General Characteristics of the Participants and Feasibility

The general characteristics of the 23 participants are presented in Table 1. Their mean age was 75.87 (SD 4.70) years. They included 18 (78%) women and 5 (22%) men. Their mean height was 156.00 (SD 10.69) cm, mean weight was 59.20 (SD 10.69) kg, and mean BMI was 24.20 (SD 3.32). Regarding the feasibility, the mean adherence rate was 97.8%, with participants attending an average of 11.7 out of 12 exercise sessions (range 10 - 12 sessions). The reasons for missed sessions were primarily scheduling conflicts with personal appointments (n=4), followed by confusion regarding the session time or location (n=2).

XSL•FO
RenderX

Table . General characteristics of study participants (N=23).

Variables and categories		Statistical value
Sex, n (%)		
	Male	5 (22)
	Female	18 (78)
Age (y), mean (SD)		75.87 (4.70)
Height (cm), mean (SD)		156.00 (10.69)
Weight (kg), mean (SD)		59.20 (10.69)
BMI (kg/m^2), mean (SD)		24.20 (3.32)

Effects of the Training on Community-Dwelling Older Adults

The results of the linear mixed model analysis adjusted for baseline values, gender, age, and BMI are presented in Table 2. The mean walking speed and functional performance of the participants were found to improve at the end of the program compared to the baseline values. Paired *t* tests showed statistically significant improvements in all posttests except SPPB and FRT.

XSL•FO
RenderX

Table . Pre- and posttest differences of the training program on community-dwelling older adults (N=23).

Variables	Pretest, mean (SD)	Posttest, mean (SD)	Difference, mean (SD)	Paired *t* test		Linear mixed model	
				t test (*df*)	*P* value	Standardized β coefficient (95% CI)	*P* value
Primary outcomes							
10MWT-SSV[a] (m/s)	1.48 (0.16)	1.33 (0.16)	0.15 (0.13)	−5.69 (22)	<.001	0.15 (0.10 to 0.20)	<.001
10MWT-FV[b] (m/s)	1.76 (0.21)	1.61 (0.20)	0.15 (0.17)	−4.29 (22)	<.001	0.15 (0.09 to 0.22)	<.001
Secondary outcomes							
SPPB[c] (score)	11.74 (0.61)	11.96 (0.20)	0.22 (0.67)	1.55 (22)	.14	0.22 (0.00 to 0.44)	.05
TUG[d] (sec)	8.30 (1.32)	7.67 (1.29)	−0.63 (0.92)	−3.30 (22)	.003	−0.63 (−1.02 to −0.25)	.002
FSST[e] (sec)	9.33 (2.21)	7.62 (1.18)	−1.71 (1.64)	−4.99 (22)	<.001	−1.71 (−2.29 to −1.13)	<.001
FRT[f] (cm)	30.37 (5.68)	32.09 (5.10)	1.72 (5.58)	1.48 (22)	.15	1.72 (−0.50 to 3.94)	.13
Muscle strength							
Hip flexion	12.59 (3.71)	13.49 (3.23)	0.90 (1.56)	2.79 (22)	.01	0.90 (0.31 to 1.50)	.004
Hip extension	12.40 (2.38)	15.03 (3.15)	2.63 (2.50)	5.05 (22)	<.001	2.63 (1.70 to 3.56)	<.001
Hip adduction	7.57 (1.93)	10.60 (3.70)	3.03 (2.73)	5.33 (22)	<.001	3.03 (2.02 to 4.04)	<.001
Hip abduction	8.19 (1.97)	9.78 (2.96)	1.59 (1.92)	3.98 (22)	<.001	1.59 (0.84 to 2.35)	<.001
Knee flexion	13.00 (3.36)	15.19 (2.25)	2.19 (2.17)	4.84 (22)	<.001	2.19 (1.40 to 2.98)	<.001
Knee extension	15.17 (4.01)	17.50 (3.49)	2.33 (3.12)	3.58 (22)	.002	2.33 (1.07 to 3.58)	<.001
Dorsiflexion	17.01 (3.20)	19.11 (3.83)	2.10 (3.06)	3.30 (22)	.003	2.10 (0.85 to 3.35)	.002
Plantarflexion	15.63 (2.05)	22.93 (4.18)	7.29 (4.92)	7.11 (22)	<.001	7.29 (5.41 to 9.17)	<.001

[a]10MWT-SSV: 10-Meter Walk Test for self-selected velocity.

[b]10MWT-FV: 10-Meter Walk Test for fastest safe velocity.

[c]SPPB: Short Physical Performance Battery.

[d]TUG: Timed Up-and-Go.

[e]FSST: Four Square Step Test.

[f]FRT: Functional Reach Test.

Specifically, the participants' 10MWT self-selected velocity improved by an average of 0.15 (SD 0.13) m/s, and their 10MWT fastest safe velocity improved by an average of 0.15 (SD 0.17) m/s. SPPB improved by an average of 0.22 (SD 0.67) points. TUG and FSST times decreased by –0.63 (SD 0.92) seconds and –1.71 (SD 1.64) seconds, respectively. FRT increased by 1.72 (SD 5.58) cm. All muscle strength measurements showed an increase, with the following values: plantarflexion (mean 7.29, SD 4.92), hip adduction (mean 3.03, SD 2.73), hip extension (mean 2.63, SD 2.50), knee extension (mean 2.33, SD 3.12), knee flexion (mean 2.19, SD 2.17), dorsiflexion (mean 2.10, SD 3.06), hip abduction (mean 1.59, SD 1.92), and hip flexion (mean 0.90, SD 1.56).

Discussion

Principal Findings

This study was conducted to assess the feasibility and the effect of a training program that uses a walking-assist wearable robot to enhance walking speed and functional performance in

XSL•FO
RenderX

community-dwelling older adults. The program incorporated gait resistance training as well as gait assistance on flat surfaces, slopes, and stairs in community settings, which distinguishes this study from previous studies that examined the effects of gait assistance on patients in a controlled setting. The results demonstrated significant effects in improving walking speed and functional performance in older adults. These findings confirm the potential for preventive health care services using a walking-assist wearable robot to contribute to the improvement of the health status and physical frailty of community-dwelling older adults.

The participants in the program showed a significant increase in walking speed of 0.15 m/s. This is consistent with previous studies that also showed an improvement in walking speed after wearable robot–assisted gait training [12-14]. Previous research [28] has established that a change in walking speed ranging from 0.10 to 0.14 m/s represents a minimal important change with a medium effect size. Therefore, the improvement observed in this study exceeds this threshold, indicating both statistical and clinical significance. Although this was achieved over a relatively short intervention period, it suggests that extended programs could lead to even greater improvements, further enhancing participants' mobility and quality of life. Previous studies have demonstrated that a walking-assist wearable robot can significantly reduce metabolic energy consumption during walking. For instance, using the device for walking assistance reduced energy expenditure by 15.7% on inclines and 10.6% on flat surfaces in young adults, which is equivalent to lowering the payload by approximately 9 kg [21,29]. In this study, the robot operates by applying torque directly to the hip joint, influencing the entire lower limb both directly and indirectly. In older adults, this hip-assist function, which aids in hip flexion and extension, can promote a more energy-efficient gait, enabling longer strides and sustained walking speed by reducing metabolic demand compared with unaided walking. In addition, as shown by Lenzi et al [30], hip exoskeleton assistance can decrease the activation of both proximal hip and distal ankle muscles. Given that aging is associated with decreased lower limb strength and reduced ankle propulsion, older adults tend to compensate by generating more power at the hip, resulting in about 17% higher metabolic energy consumption during walking compared with younger individuals [31,32]. Therefore, hip assistance is anticipated to enhance stride length and walking speed, helping maintain these improvements throughout extended periods. Furthermore, the findings are similar to the results of a study in which older adults were divided into 3 groups—independent walking, partially assisted by a robot, and fully assisted by a robot. The group that was fully assisted by a robot exhibited a significant increase in walking speed [11]. Decreased walking speed is a known risk factor for disability, frailty, falls, cognitive impairment, and mortality among community-dwelling older adults [6,33]. It is also a significant predictor of disability in ADL [34,35]. In view of these considerations, older adults' increased walking speed may contribute to maintaining their daily physical activity and independent functioning by reducing future rates of disability or impairment. Therefore, training programs using wearable robots can be considered a preventive health care service for community-based aging in place, aligning with a global policy trend in the era of aging.

The results also revealed statistically significant improvements in the TUG and FSST, with the time required for each test reducing by an average of 0.63 (SD 0.92) seconds (7.59%) and 1.71 (SD 1.64) seconds (18.33%), respectively. These results are similar to previous studies that have reported improvements in TUG [12-14,18,36] and FSST [18] after training with a wearable robot. In this study, the TUG showed a reduction of 0.63 seconds, exceeding the minimal important change threshold of 0.4 seconds. This suggests a clinically significant improvement, indicating enhanced mobility and a potential reduction in the participants' fall risk. In general, the TUG and FSST are reported to be associated with a higher risk of falling if they take longer than 15 seconds to complete [37]; community-dwelling older adults are expected to be able to perform these tests in less than 12 seconds [36,38]. Similar to walking speed, TUG and FSST in older adults have been reported to be key predictors of health deterioration, new ADL difficulties, and falls [35,39]. The results of this study highlight the importance of the use of wearable robots as a physical activity promotion interface to support independent community living in older adults.

The postintervention assessment also revealed significant improvements in leg muscle strength. These results are similar to those of previous studies on patients with impairments [15,17]. The participants' muscle strength potentially improved as the wearable robot supported their gait efforts to move their joints. In addition, the introduction of a gait resistance mode helped address the typical drawback of walking exercises, where it can be challenging to increase muscle strength load. By allowing participants to exert more force, particularly from week 4, and gradually increasing the resistance time by 5 minutes per week, the resistance mode contributed to strength gains. Furthermore, the combination of walking training on not only flat surfaces but also inclines and stairs may have contributed to muscle strength improvements. A previous study [19] also identified progressive resistance training programs as an effective strategy for increasing muscle strength and bone formation in older adults aged more than 75 years. Although this study used progressive training with gradually increasing gait resistance intensity, differences in the effectiveness of resistance training may exist among participants, as they were allowed to self-regulate the intensity to complete the exercise. Therefore, programs should ideally include a progressive training method customized to the individual's condition to improve muscle strength without causing muscle fatigue or injury.

Although the SPPB and FRT showed improvements in this study, they were not statistically significant. Regarding the SPPB, our study observed an improvement of 0.22 points. While this change did not reach the established clinical threshold of 0.54 points [28] typically required for a meaningful improvement, it is important to consider the high baseline performance of our participants, who had scores close to the maximum possible. The SPPB is a 12-point test, on which the participants scored an average of 11.74 (SD 0.61) in the pretest, indicating that most of them were close to perfect, which may

XSL•FO
RenderX

explain the lack of statistical significance due to the small variation and ceiling effect [40]. Ceiling effects in highly functional groups may limit the SPPB's sensitivity to detect subtle changes. Future studies should consider using more sensitive assessment tools to better capture subtle improvements in similar populations. The FRT is a test that reflects dynamic balance ability [41]; while a previous study [14] has reported significant effects, this study's results differed.

However, slight improvements in SPPB and FRT suggest that the program may have improved functional performance in older adults.

In addition, walking speed and functional performance, which were statistically significantly improved by paired *t* tests, remained statistically significant after adjusting for the baseline values, gender, age, and BMI, confirming the program's effectiveness. From the perspective of intrinsic capacity, a core element of the World Health Organization's framework on healthy aging [42], it is essential to maintain and enhance physical capacities even in individuals with high baseline function. The program's effectiveness in preserving mobility, strength, and overall functional performance underscores its potential to support healthy aging by fostering the sustained development of intrinsic capacity, which is crucial for long-term functional ability and independence. Furthermore, future research should target more frail populations, as they may experience even greater benefits from such interventions. This would allow for a clearer assessment of the program's potential to improve intrinsic capacity and functional ability in individuals with lower baseline performance, addressing a broader range of needs within the aging population.

Limitations

This study has several limitations. First, the absence of a control group and the lack of a postrepeat measurement limit the elucidation of the program's direct and residual effects, thus making it difficult to establish causality. Therefore, the long-term effects should be confirmed with a control group in future studies to address these limitations. Second, the evaluators who conducted both pre- and postassessments were not blinded. While this arrangement may introduce potential bias, we attempted to minimize subjective influence by using objective evaluation tools, such as walking speed and muscle strength measurements. Future studies should consider using independent, blinded evaluators to further reduce the risk of bias and ensure greater objectivity in outcome assessments. In addition, in cases where physical function is relatively high, as observed in the participants of this study, the SPPB may have limitations in detecting subtle changes. This can lead to difficulties in achieving statistical significance due to a ceiling effect, where participants score near the maximum limit, reducing the sensitivity of the measure. Therefore, future studies should consider using more sensitive and appropriate assessment tools that can accurately capture subtle variations, thereby improving the reliability of the findings. Finally, during the 6 weeks of the program, 7 participants dropped out because of the burden of group exercise, lowering the statistical power of the study, as only 23 participants in total were analyzed. In group training, a wide gap in pace with other people tends to be psychologically discouraging, making it difficult to continue exercising. It may be advisable to conduct group training in small groups of no more than 3‐4 people, considering individual physical abilities. Nonetheless, the significant walking speed and functional performance improvements revealed in this study support the effectiveness of the program using a walking-assist wearable robot for community-dwelling older adults.

Conclusions

In conclusion, implementing a preventive intervention of an appropriate physical activity program using a wearable robot proved effective in improving walking speed and functional performance in community-dwelling older adults. This study provides evidence of its feasibility and effectiveness as an intervention to promote independent and successful aging in older adults. In addition, our findings confirm the clinical potential of the gait assistance and gait resistance training program using a wearable robot as a preventive health care service for community-dwelling older adults in an aging society.

Acknowledgments

This work was supported by the Materials and Components Technology Development Program (RS-2024‐00420336, Development of 1.5kg ultralight wearable robot for preventive rehabilitation of seniors in daily life) funded By the Ministry of Trade Industry & Energy (MOTIE, Korea).

Authors' Contributions

EC, SH, BL, and YL contributed to the study concept and design. BL, JL, and YL were responsible for acquisition of the data. EC, SH, and SJH analyzed, interpreted the data, and drafted the manuscript. All authors participated in the revision of the manuscript and approved the final manuscript.

Conflicts of Interest

BL, YL, and JL are current employees of the company Wirobotics Incorporation. The other authors declare no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Multimedia Appendix 1

XSL•FO
RenderX

Informed consent form.
[DOCX File, 198 KB - mhealth_v13i1e58142_app1.docx]

References

1. Life expectancy and healthy life expectancy. World health organization (WHO). 2024. URL: https://apps.who.int/gho/data/node.main.688?lang=en [accessed 2024-01-18]
2. Seidel D, Brayne C, Jagger C. Limitations in physical functioning among older people as a predictor of subsequent disability in instrumental activities of daily living. Age Ageing 2011 Jul;40(4):463-469. [doi: 10.1093/ageing/afr054] [Medline: 21609999]
3. Framework for countries to achieve an integrated continuum of long-term care 2021. World Health Organization (WHO). URL: https://www.who.int/publications/i/item/9789240038844 [accessed 2024-01-18]
4. Older adults. Healthy People 2030. URL: https://health.gov/healthypeople/objectives-and-data/browse-objectives/older-adults [accessed 2024-02-26]
5. Afiah IN, Nakashima H, Loh PY, Muraki S. An exploratory investigation of changes in gait parameters with age in elderly Japanese women. Springerplus 2016;5(1):1069. [doi: 10.1186/s40064-016-2739-7] [Medline: 27462517]
6. Rodríguez-Molinero A, Herrero-Larrea A, Miñarro A, et al. The spatial parameters of gait and their association with falls, functional decline and death in older adults: a prospective study. Sci Rep 2019 Jun 19;9(1):8813. [doi: 10.1038/s41598-019-45113-2] [Medline: 31217471]
7. Femia EE, Zarit SH, Johansson B. The disablement process in very late life: a study of the oldest-old in Sweden. J Gerontol B Psychol Sci Soc Sci 2001 Jan;56(1):12-23. [doi: 10.1093/geronb/56.1.p12] [Medline: 11192333]
8. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
9. Visser M, Pluijm SMF, Stel VS, Bosscher RJ, Deeg DJH, Longitudinal Aging Study Amsterdam. Physical activity as a determinant of change in mobility performance: the longitudinal aging study Amsterdam. J Am Geriatr Soc 2002 Nov;50(11):1774-1781. [doi: 10.1046/j.1532-5415.2002.50504.x] [Medline: 12410894]
10. Bogue R. Robotic exoskeletons: a review of recent progress. Ind Rob 2015 Jan 19;42(1):5-10. [doi: 10.1108/IR-08-2014-0379]
11. Lee HJ, Lee S, Chang WH, et al. A wearable hip assist robot can improve gait function and cardiopulmonary metabolic efficiency in elderly adults. IEEE Trans Neural Syst Rehabil Eng 2017 Sep;25(9):1549-1557. [doi: 10.1109/TNSRE.2017.2664801] [Medline: 28186902]
12. Zhang L, Lin F, Sun L, Chen C. Comparison of efficacy of Lokomat and wearable exoskeleton-assisted gait training in people with spinal cord injury: a systematic review and network meta-analysis. Front Neurol 2022;13:772660. [doi: 10.3389/fneur.2022.772660] [Medline: 35493806]
13. Yoo HJ, Bae CR, Jeong H, Ko MH, Kang YK, Pyun SB. Clinical efficacy of overground powered exoskeleton for gait training in patients with subacute stroke: a randomized controlled pilot trial. Medicine (Baltimore) 2023 Jan 27;102(4):e32761. [doi: 10.1097/MD.0000000000032761] [Medline: 36705351]
14. Yoshimoto T, Shimizu I, Hiroi Y, Kawaki M, Sato D, Nagasawa M. Feasibility and efficacy of high-speed gait training with a voluntary driven exoskeleton robot for gait and balance dysfunction in patients with chronic stroke: nonrandomized pilot study with concurrent control. Int J Rehabil Res 2015 Dec;38(4):338-343. [doi: 10.1097/MRR.0000000000000132] [Medline: 26288120]
15. Park GM, Cho SH, Hong JT, Kim DH, Shin JC. Effects and safety of wearable exoskeleton for robot-assisted gait training: a retrospective preliminary study. J Pers Med 2023 Apr 18;13(4):676. [doi: 10.3390/jpm13040676] [Medline: 37109062]
16. Afzal T, Tseng SC, Lincoln JA, Kern M, Francisco GE, Chang SH. Exoskeleton-assisted gait training in persons with multiple sclerosis: a single-group pilot study. Arch Phys Med Rehabil 2020 Apr;101(4):599-606. [doi: 10.1016/j.apmr.2019.10.192] [Medline: 31821798]
17. Drużbicki M, Guzik A, Przysada G, et al. Effects of robotic exoskeleton-aided gait training in the strength, body balance, and walking speed in individuals with multiple sclerosis: a single-group preliminary study. Arch Phys Med Rehabil 2021 Feb;102(2):175-184. [doi: 10.1016/j.apmr.2020.10.122] [Medline: 33181115]
18. Kim J, Chun H, Lee SH, Lee HJ, Kim YH. Functional improvement and satisfaction with a wearable hip exoskeleton in community-living adults. Healthcare (Basel) 2023 Feb 22;11(5):643. [doi: 10.3390/healthcare11050643] [Medline: 36900648]
19. O'Bryan SJ, Giuliano C, Woessner MN, et al. Progressive resistance training for concomitant increases in muscle strength and bone mineral density in older adults: a systematic review and meta-analysis. Sports Med 2022 Aug;52(8):1939-1960. [doi: 10.1007/s40279-022-01675-2] [Medline: 35608815]
20. Jayaraman C, Embry KR, Mummidisetty CK, et al. Modular hip exoskeleton improves walking function and reduces sedentary time in community-dwelling older adults. J Neuroeng Rehabil 2022 Dec 30;19(1):144. [doi: 10.1186/s12984-022-01121-4] [Medline: 36585676]
21. Chang Y, Kang J, Jeong B, et al. Verification of industrial worker walking efficiency with wearable hip exoskeleton. Appl Sci (Basel) 2023;13(23):12609. [doi: 10.3390/app132312609]
22. Dobkin BH. Short-distance walking speed and timed walking distance: redundant measures for clinical trials? Neurology (ECronicon) 2006 Feb 28;66(4):584-586. [doi: 10.1212/01.wnl.0000198502.88147.dd] [Medline: 16505318]

XSL•FO
RenderX

23. Guralnik JM, Simonsick EM, Ferrucci L, et al. A Short Physical Performance Battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994 Mar;49(2):M85-M94. [doi: 10.1093/geronj/49.2.m85] [Medline: 8126356]
24. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
25. Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults. Arch Phys Med Rehabil 2002 Nov;83(11):1566-1571. [doi: 10.1053/apmr.2002.35469] [Medline: 12422327]
26. Goh EY, Chua SY, Hong SJ, Ng SS. Reliability and concurrent validity of Four Square Step Test scores in subjects with chronic stroke: a pilot study. Arch Phys Med Rehabil 2013 Jul;94(7):1306-1311. [doi: 10.1016/j.apmr.2013.01.027] [Medline: 23416218]
27. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. J Gerontol 1990 Nov;45(6):M192-M197. [doi: 10.1093/geronj/45.6.m192] [Medline: 2229941]
28. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006 May;54(5):743-749. [doi: 10.1111/j.1532-5415.2006.00701.x] [Medline: 16696738]
29. Bastien GJ, Willems PA, Schepens B, Heglund NC. Effect of load and speed on the energetic cost of human walking. Eur J Appl Physiol 2005 May;94(1-2):76-83. [doi: 10.1007/s00421-004-1286-z] [Medline: 15650888]
30. Lenzi T, Carrozza MC, Agrawal SK. Powered hip exoskeletons can reduce the user's hip and ankle muscle activations during walking. IEEE Trans Neural Syst Rehabil Eng 2013 Nov;21(6):938-948. [doi: 10.1109/TNSRE.2013.2248749] [Medline: 23529105]
31. DeVita P, Hortobagyi T. Age causes a redistribution of joint torques and powers during gait. J Appl Physiol (1985) 2000 May;88(5):1804-1811. [doi: 10.1152/jappl.2000.88.5.1804] [Medline: 10797145]
32. Das Gupta S, Bobbert MF, Kistemaker DA. The metabolic cost of walking in healthy young and older adults - a systematic review and meta analysis. Sci Rep 2019 Jul 10;9(1):9956. [doi: 10.1038/s41598-019-45602-4] [Medline: 31292471]
33. Binotto MA, Lenardt MH, Rodríguez-Martínez MDC. Physical frailty and gait speed in community elderly: a systematic review. Rev Esc Enferm USP 2018 Dec 13;52:e03392. [doi: 10.1590/S1980-220X2017028703392] [Medline: 30570081]
34. Vermeulen J, Neyens JCL, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. BMC Geriatr 2011 Jul 1;11:33. [doi: 10.1186/1471-2318-11-33] [Medline: 21722355]
35. Viccaro LJ, Perera S, Studenski SA. Is timed up and go better than gait speed in predicting health, function, and falls in older adults? J Am Geriatr Soc 2011 May;59(5):887-892. [doi: 10.1111/j.1532-5415.2011.03336.x] [Medline: 21410448]
36. Bischoff HA, Stähelin HB, Monsch AU, et al. Identifying a cut-off point for normal mobility: a comparison of the timed "up and go" test in community-dwelling and institutionalised elderly women. Age Ageing 2003 May;32(3):315-320. [doi: 10.1093/ageing/32.3.315] [Medline: 12720619]
37. Nordin E, Lindelöf N, Rosendahl E, Jensen J, Lundin-Olsson L. Prognostic validity of the Timed Up-and-Go test, a modified Get-Up-and-Go test, staff's global judgement and fall history in evaluating fall risk in residential care facilities. Age Ageing 2008 Jul;37(4):442-448. [doi: 10.1093/ageing/afn101] [Medline: 18515291]
38. Whitney SL, Marchetti GF, Morris LO, Sparto PJ. The reliability and validity of the Four Square Step Test for people with balance deficits secondary to a vestibular disorder. Arch Phys Med Rehabil 2007 Jan;88(1):99-104. [doi: 10.1016/j.apmr.2006.10.027] [Medline: 17207683]
39. Cleary K, Skornyakov E. Predicting falls in older adults using the four square step test. Physiother Theory Pract 2017 Oct;33(10):766-771. [doi: 10.1080/09593985.2017.1354951] [Medline: 28771062]
40. Kameniar K, Mackintosh S, Van Kessel G, Kumar S. The psychometric properties of the Short Physical Performance Battery to assess physical performance in older adults: a systematic review. J Geriatr Phys Ther 2024;47(1):43-54. [doi: 10.1519/JPT.0000000000000337] [Medline: 35442231]
41. Rosa MV, Perracini MR, Ricci NA. Usefulness, assessment and normative data of the Functional Reach Test in older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr 2019;81:149-170. [doi: 10.1016/j.archger.2018.11.015] [Medline: 30593986]
42. WHO global strategy and action plan on ageing and health (2016–2020). World Health Organization (WHO). 2016. URL: http://apps.who.int/gb/ebwha/pdf_files/WHA69/A69_17-en.pdf [accessed 2024-11-18]

Abbreviations

10MWT: 10-Meter Walk Test
ADL: activities of daily living
FRT: Functional Reach Test
FSST: Four Square Step Test
HALE: health-adjusted life expectancy
SPPB: Short Physical Performance Battery

XSL•FO
RenderX

TUG: Timed Up-and-Go

Edited by TDA Cardoso; submitted 12.03.24; peer-reviewed by K Miyata, M Colombo, S Muraki; revised version received 12.04.25; accepted 14.04.25; published 26.05.25.

Please cite as:
Cho E, Hwang S, Heo SJ, Lim B, Lee J, Lee Y
Feasibility and Effects of a Gait Assistance and Gait Resistance Training Program Using a Walking-Assist Wearable Robot for Community-Dwelling Older Adults: Single-Group, Pre-, and Posttest Study
JMIR Mhealth Uhealth 2025;13:e58142
URL: https://mhealth.jmir.org/2025/1/e58142
doi:10.2196/58142

© Eunhee Cho, Sinwoo Hwang, Seok-Jae Heo, Bokman Lim, Jewoo Lee, Younbaek Lee. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Behavioral Markers in Older Adults During COVID-19 Confinement: Secondary Analysis of In-Home Sensor Data

Knoo Lee[1,2], RN, PhD; Noah Marchal[2], MS, MFA; Erin L Robinson[3], PhD; Kimberly R Powell[1], RN, PhD

[1]Sinclair School of Nursing, University of Missouri, 915 Hitt St, S413, Columbia, MO, United States

[2]Institute for Data Science and Informatics, University of Missouri, Columbia, MO, United States

[3]College of Health Professions, University of Missouri, Columbia, MO, United States

Corresponding Author:
Knoo Lee, RN, PhD
Sinclair School of Nursing, University of Missouri, 915 Hitt St, S413, Columbia, MO, United States

Abstract

Background: Older adults were disproportionately affected by the COVID-19 pandemic, with a high number of deaths occurring in this age group. The impact of social isolation and home confinement continues to impact the mental and emotional health of older adults, despite the end of the COVID-19 pandemic. Unhealthy lifestyle behaviors, including physical and social inactivity, and poor sleep quality, have been reported. Recommendations for healthy lifestyle changes have primarily targeted the general population, highlighting the need for personalized recommendations for vulnerable older adults. Remote sensing technologies may offer an opportunity to understand behavior changes among older adults and provide personalized recommendations.

Objective: This study aims to describe the effects of home confinement and social isolation on community-dwelling older adults during the COVID-19 outbreak and investigate how integrated computing technologies, such as remote sensors installed in homes, can help inform recommendations for safe and healthy lifestyles.

Methods: As part of a larger study and ongoing research with community-dwelling older adults, remote sensors including bed transducers, 3D depth cameras, and passive infrared (PIR) motion sensors were installed in the homes of the study sample. We compared features derived from sensors for approximately one month before the COVID-19 outbreak (January 14, 2020-February 13, 2020) and one month after the onset of the pandemic (March 14, 2020-April 13, 2020). We used descriptive statistics and paired-sample *t* tests to compare the 2 time periods, pre-COVID-19 and early-COVID-19.

Results: Sensor data from 64 older adults were analyzed, the majority identifying as female (n=51, 80%), aged >76 years (n=58, 92%), and living alone (n=50, 78%). Results from paired-sample *t* tests demonstrated significant differences in sensor features between the pre–COVID-19 and early–COVID-19 time periods. We found statistically significant differences in bed restlessness (pre-COVID: mean 14.98, SD 5.10; early-COVID: mean 15.56, SD 5.25; t_{554}=−4.10; P<.001), time spent in bed (pre-COVID: mean 32,547.41, SD 9269.96; early-COVID: mean 33,494.73, SD 10,887.33; t_{554}=−2.81; P=.005), pulse (pre-COVID: mean 68.45, SD 3.30; early-COVID: mean 68.10, SD 3.36; t_{554}=3.66; P<.001), respiration (pre-COVID: mean 14.54, SD 1.32; early-COVID: mean 14.41, SD 1.31; t_{553}=3.72; P<.001), and stride length (pre-COVID: mean 29.10, SD 4.813; early-COVID: mean 28.76, SD 5.016; t_{595}=2.17; P=.03). Among the study sample, bed restlessness and time spent in bed increased between the 2 time periods, while pulse, respiration, and stride length decreased.

Conclusions: This study highlights that home confinement during the pandemic significantly impacted the behavior and health of older adults, leading to more sedentary lifestyles and poorer sleep quality. These changes may contribute to a decline in physical and mental health, increasing the risk of depression, lack of social contact, and diminished functional capacity. The findings underscore the need for older adults in future infectious disease outbreaks and suggest in-home sensor technology as a potential tool for monitoring their health and guiding decisions during periods of confinement.

(JMIR Mhealth Uhealth 2025;13:e56678) doi:10.2196/56678

KEYWORDS

older adults; aging-in-place; community dwelling; sensors; healthy lifestyle

Introduction

Nearly 75% of COVID-19–linked deaths (more than 630,000) between 2020 and 2022 occurred among people of 65 years and older [1]. This suggests that COVID-19 has taken the lives of nearly 1 in 100 US older adults in just over 2 years [2]. Although case demographics of COVID-19 in the United States fluctuated over time and between the US states since the beginning of the pandemic, older adults were affected disproportionately in terms of hospitalizations and deaths [3]. To limit the spread of

XSL•FO
RenderX

COVID-19, lockdowns with home confinement and social distancing measures were implemented by public health authorities around the world. Despite being recognized as effective measures to curb the spread of the COVID-19 outbreak, social distancing and home confinement caused profound disruption of normal lifestyles, exacerbating mental health concerns, and social isolation among older adults.

The home confinement and social distancing measures implemented to reduce the spread of COVID-19 during the COVID-19 outbreak resulted in significant negative effects on mental health. Studies have reported worsening mental health symptoms, including stress, anxiety, depression, insomnia, and intrusive thinking [4,5] among older adults as a result of pandemic-induced social isolation. In addition, these psychosocial symptoms have been found to be significantly associated with unhealthy lifestyle behaviors. Ammar et al [6], found that the COVID-19 home confinement resulted in increased physical and social inactivity, poor sleep quality, and unhealthy diet. Community-dwelling older adults who lived alone reported struggles with eating properly during the early stages of the pandemic [7]. It is also recognized that lifestyle behaviors, including physical inactivity, unhealthy diets, and poor sleep quality, significantly affect the severity of COVID-19 and can extend recovery time [8]. This illustrates the cyclical nature of home confinement, where measures taken to minimize COVID-19 infection risk inadvertently contribute to factors such as reduced physical activity, which in turn can lead to negative mental health outcomes and worsen overall COVID-19 impacts. These negative impacts during the COVID-19 lockdown were observed not only among older adults but also among elite athletes, underscoring the widespread adverse effects of home confinement across different populations [9]. Results from a worldwide cohort study revealed that older adults were more likely to experience weight gain, physical inactivity, and perceived frailty during the COVID-19 lockdown [10].

Alleviating the risk factors associated with negative mental health symptoms experienced by older adults is a high priority. However, available recommendations from organizations including the World Health Organization (WHO), American Heart Association (AHA), and American College of Sports Medicine (ACSM), mainly target the general population [11-13]. To move forward, it is imperative to consider the holistic behavioral patterns among older adults. This understanding forms the foundational groundwork for crafting personalized recommendations, which hinge on precise user profiling and the subsequent dissemination of these recommendations to the individuals themselves.

This study is uniquely positioned to examine the physiological and behavioral patterns of community-dwelling older adults during the onset of the COVID-19 pandemic. At this time, participants were already recruited into a larger, ongoing, observational study that installed embedded sensing technologies in their homes. The goal of that project was to develop a tailored interface for older adults and their family caregivers to monitor the older adults' health-generated sensor data. Since the sensor technologies were already in place before the onset of the COVID-19 pandemic, they offer us a real-time glimpse of community-dwelling older adults' patterns during the initial onset of a global pandemic. The purpose of this study was to describe the effects of home confinement and social isolation on community-dwelling older adults during the COVID-19 outbreak. By analyzing data derived from in-home sensor systems of older adults, we intend to inform recommendations for safe and healthy lifestyles during future infectious disease outbreaks.

Methods

Recruitment

Residents (n=143) were recruited from Americare Senior Living facilities located in mid-Missouri between June 2017 to July 2022. For this analysis, we selected residents (n=64) who had at least one of 3 types of sensor data (bed, gait, and motion) for the duration of the study time period (January 14, 2020-April 13, 2020).

Procedures

A suite of in-home monitoring sensors, developed at the Center to Stream Healthcare In-Place, and commercially licensed by Foresite Healthcare, LLC [14,15], were installed within the study participants' homes. Data were collected from 3 types of in-home sensors: a bed mat containing 4 hydraulic transducers that captured ballistocardiogram signals, a thermal depth sensor for capturing gait and detecting falls, and passive infrared (PIR) motion sensors [16-19]. The hydraulic bed mat was installed directly beneath the side of the bed that a participant sleeps on, minimizing the amount of error introduced by companions or pets. Errors in gait measurements are reduced by using a probability density function model to identify participant walks separate from those of visitors or caregivers. However, it is understood that the motion sensors capture resident-level activity throughout the home. The raw data were first preprocessed algorithmically to derive the individual health parameters using existing methods [17,20-24], with the exception of percent restless and walk density, which were calculated respectively as the percent of time in bed spent restless and the number of walks recorded per day. The raw ballistocardiogram waveforms were sampled and collected at 100 hertz (Hz), from which the pulse rate, respiration, and restlessness parameters were extracted using signal processing algorithms [17]. The resolution of the depth sensor allows for collecting gait parameters during purposeful walks at 15 frames per second [22], as well as detecting falls. The motion sensor firings were collected in 7-second intervals over the duration from the onset of detected motion [25,26]. To account for the differing sampling rates, we aggregated the derived health parameters as daily averages before analysis. Due to the heterogeneity of participant home environments not all possible sensor variables were collected from participants. Bed mats were not installed for participants who sleep in hospital beds with adjustable height or in a recliner chair. Gait parameters were not collected from the thermal depth sensor for participants who are wheelchair-bound. In addition, Zigbee motion sensors required frequent battery changes every 6 - 8 weeks, which reduced the availability for the period during which facility visitation was limited after the onset of the pandemic. In order to compare the periods before and after the shelter-in-place order took effect, we limited our participant

XSL•FO
RenderX

cohort to only those with data both before and immediately after. All 64 participants initially recruited were included in at least one of the 3 sensor analyses (bed, gait, or motion). However, not all participants had complete data for every sensor type. In specific, 43 participants had complete bed sensor data, 49 participants had complete gait sensor data, and 43 participants had complete motion sensor data. Participants were categorized based on the availability of usable data for each sensor type. Some participants contributed to more than one category, while others contributed to only one. For example, a participant with complete gait sensor data but no bed or motion sensor data was included in the gait analysis but not in the others. Importantly, no participants were excluded entirely, as all 64 provided usable data for at least one sensor type. This aligns with the inclusion criterion of having at least one form of sensor data during the study period. There were 14 individuals with roommates and 50 without roommates. Gait observations were filtered using Gaussian distributions of stride time, stride length, and walking speed, relative to participant height, to form clusters. These clusters were validated using a probability density function, effectively removing outliers caused by visitors' gait metrics [27].

Motion sensors capture room-level activity within the residence, including movement from visitors, family members, and facility staff. The data is calculated as the number of 7-second intervals during which motion is detected. Since our focus is on the overall patterns between the periods immediately before and after the onset of the pandemic, we consider these motion variables as indicators of isolation within the residence rather than as measurements of individual behavior.

Statistical Analysis

We compared 17 health parameters (Table 1) derived from sensors installed in apartments of older adults who resided in an independent living community pre- and early-COVID-19 outbreak. Descriptive statistics and paired-sample *t* tests for each health parameter were calculated between the one-month pre–COVID-19 and post–COVID-19 periods. The pre–COVID-19 period represented approximately one month before the Centers for Medicare and Medicaid Services (CMS) issuance of a mandatory stay at home order (January 14, 2020-February 13, 2020) and the early–COVID-19 period represented approximately one month after the stay-at-home order (March 14, 2020-April 13, 2020) [28]. The *t* tests were conducted in Python using Scipy.stats package's ttest_rel function with null observations omitted. We used group-level data for *t* tests. There was not a treatment, outcome effect measured, or comparison against control group involved in the analysis that would lead to confound or performance bias. Gait data is filtered to the resident with a probability density model to prevent noise introduced by walks captured by apartment visitors. We used only the available data without applying any imputation techniques for missing entries. Participants with complete data for both the pre- and post-COVID lockdown periods were included. If postlockdown data were missing, the participant was excluded from the *t* test analysis to maintain methodological consistency.

Table . Variables included in the analysis.

Variable	Description	Unit
Average pulse rate	Average pulse rate computed with energy algorithm	Per minute
Minimum pulse rate	Minimum pulse rate computed with energy algorithm	Per minute
Maximum pulse rate	Maximum pulse rate computed with energy algorithm	Per minute
Average respiration rate	Average respiration rate	Per minute
Maximum respiration rate	Maximum respiration rate	Per minute
Restlessness	Number of seconds of restlessness in bed	Count
Time in bed	Number of seconds in bed	Count
Percentage of restlessness	Ratio of seconds_restless to seconds_in_bed * 100	Percent
Height	Walking height	Inches
Walking speed	Walking speed	Centimeters per second
Stride time	Time taken for stride	Seconds
Stride length	Length of stride	Centimeters
Walk density	Count of walks	Count
Bathroom motion	Motion hits bathroom	Count
Bedroom motion	Motion hits bedroom	Count
Front door motion	Motion hits front door	Count
Living room motion	Motion hits living room	Count

XSL•FO
RenderX

Ethical Considerations

Participants consented through the study protocol approved by the University of Missouri Institutional Review Board for a National Library of Medicine award (R01LM012221) entitled "Linguistic Summarization of Sensor Data for Early Illness Recognition in Eldercare." The study was approved by the institutional review board of an R1 university. Written informed consent was obtained from all participants, and the original consent included provisions for the future use of anonymized data in secondary analyses. All data were deidentified before analysis, and strict measures were implemented to protect participant information, including secure storage and encryption. Data transmission adhered to Health Insurance Portability and Accountability Act requirements. Participants were not compensated for their involvement in this study.

Results

Participant Characteristics

There were 64 participants (See Table 2 for baseline characteristics), of whom 51 (80%) were female and 13 (20%) were male. Participants were grouped by age as follows: 1 (2%) was under 65 years old, 5 (8%) were between 66 and 75, 24 (38%) were between 76 and 85, 26 (41%) were between 86 and 95, and 8 (13%) were older than 95 years. A total of 14 participants (22%) reported using a wheelchair, while 55 (86%) reported using a walker. A total of 14 (22%) participants reported having a roommate, while 50 (78%) did not.

Table . Participant characteristics (N=64).

Baseline characteristics	Value, n (%)
Sex	
Female	51 (80)
Male	13 (20)
Age (years)	
<65	1 (2)
66 - 75	5 (8)
76 - 85	24 (38)
86 - 95	26 (41)
>95	8 (13)
Mobility aids	
Wheelchair	
Yes	14 (22)
No	50 (78)
Walker	
Yes	55 (86)
No	9 (14)
Has roommate	
Yes	14 (22)
No	50 (78)

Statistical Outcomes

Group-level paired *t* test revealed statistically significant differences in pulse, respiration, restlessness, stride, walk speed, and times spent in bed between pre– and early–COVID-19 time periods (See Table 3).

XSL•FO
RenderX

Table . Paired *t* test results.

Variables	Pre-COVID-19, mean (SD)	Early-COVID-19, mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
Bathroom	351.42 (305.82)	309.45 (275.82)	5.19 (1314)	<.001	0.144
Bedroom	499.57 (454.22)	497.95 (394.52)	0.17 (1314)	.87	0.004
Front door	121.34 (158.36)	134.07 (159.46)	−2.91 (1314)	.004	−0.08
Living room	431.11 (439.04)	371.4 (393.48)	5.14 (1314)	<.001	0.143
Pulse (mean)	68.45 (3.3)	68.1 (3.36)	3.66 (554)	<.001	0.107
Pulse (max)	88.61 (5.99)	88.29 (5.71)	1.28 (554)	.20	0.055
Pulse (min)	51.43 (3.51)	51.4 (3.72)	0.29 (554)	.78	0.01
Respiration(mean)	14.54 (1.32)	14.41 (1.31)	3.72 (553)	<.001	0.099
Respiration (max)	18.98 (0.75)	18.95 (0.81)	0.64 (554)	.52	0.031
In bed (s)	32,547.41 (9269.96)	33,494.73 (10,887.33)	−2.81 (554)	.005	−0.094
Restless (%)	14.98 (5.1)	15.56 (5.25)	−4.1 (554)	<.001	−0.111
Restless (s)	4790.82 (1876.29)	5223.73 (2376.45)	−6.15 (554)	<.001	−0.204
Height	63.75 (3.75)	63.37 (4.17)	4.38 (611)	<.001	0.095
Stride length	29.1 (4.81)	28.76 (5.02)	2.17 (595)	.03	0.071
Stride time	1.45 (0.25)	1.51 (0.3)	−4.94 (595)	<.001	−0.193
Walk speed	21.81 (5.27)	21.33 (5.79)	2.79 (611)	.006	0.086
Walk density	103.16 (87.75)	107.91 (111.56)	−1.24 (611)	.22	−0.048

Regarding monthly activity by room locations, there was a significant decrease in levels of detected motions in the bathroom and living room during early–COVID-19 (bathroom: mean 309.45, SD 275.82; t_{1314}=5.19; *P*<.001; living room: mean 371.40, SD 393.48; t_{1314}=5.14; *P*<.001) compared with pre–COVID-19 (bathroom: mean 351.42, SD 305.82; living room: mean 431.11, SD 439.04). The movement at the front door, on the other hand, significantly increased during early–COVID-19 (mean 134.07, SD 159.46; *P*=.004) compared with the pre–COVID-19 time period (mean 121.34, SD 158.36).

Regarding changes in walk-related measurements, stride time increased (pre-COVID: mean 1.45, SD 0.25; early-COVID: mean 1.51, SD 0.30; t_{595}=−4.94; *P*<.001), while stride length (pre-COVID: mean 29.104, SD 4.813; early-COVID: mean 28.756, SD 5.016; t_{595}=2.167; *P*=.03), walk speed (pre-COVID: mean 21.81, SD 5.27; early-COVID: mean 21.33, SD 5.79; t_{611}=2.79; *P*=.006), and height (pre-COVID: mean 63.75, SD 3.75; early-COVID: mean 63.37, *SD* 4.17, t_{611}=4.38; *P*<.001) decreased.

Other measurements with significant differences include restlessness (increase; pre-COVID: mean 14.98, SD 5.10; early-COVID: mean 15.56, SD 5.25; t_{554}=−4.10; *P*<.001), seconds in bed (increase; pre-COVID: mean 32,547.41, SD 9269.96; early-COVID: mean 33,494.73, SD 10,887.33, t_{554}=−2.81; *P*=.005), pulse (decrease; pre-COVID: mean 68.45, SD 3.30; early-COVID: mean 68.10, SD 3.36; t_{554}=3.66; *P*<.001), and respiratory rate (decrease; pre-COVID: mean 14.54, SD 1.32; early-COVID: mean 14.41, SD 1.31; t_{553}=3.72; *P*<.001).

Discussion

Principal Findings

Our findings provide evidence on the impact of home confinement and social distancing measures on the behavior and health of aging in place older adults. When we compared pre–COVID-19 to the early–COVID-19 time periods, participants in our study showed a significant decrease in motion detected in the bathroom and living room, while motion at the front door increased. There was a significant decrease in stride, walk speed, and height; conversely, stride time increased. Sensors detected an increase in in-bed restlessness and a decrease in both average pulse and average respiratory rate. These findings suggest the decline in activity during home confinement could negatively impact the physical and mental health of older adults.

Comparison With Previous Work

Regarding activities by location, the sedentary lifestyle that older adults exhibit in the living room may have long-term negative effects on their physical health. Previous research demonstrated a link between prolonged sitting and increased risk of obesity, cardiovascular disease, and type 2 diabetes [29-31]. These studies highlight the importance of promoting physical activity and reducing sedentary habits among older adults, further supporting the utility of remote sensors in discovering early signs of physical inactivity or sedentary lifestyles. Despite the lower priority assigned to the mentioned health risks (obesity, cardiovascular disease, type 2 diabetes) in comparison to safeguarding and quarantine due to the risks

XSL•FO
RenderX

and severity of COVID-19 infection during the pandemic, it is essential to highlight that addressing the issue of physical inactivity becomes even more crucial in these challenging circumstances necessitating home confinement. The decline in bathroom usage may impact a number of health issues, including urinary tract infections, constipation, dehydration, and kidney stones [32-34]. In addition, this pattern of decreased bathroom use may be due to residents not performing hygiene tasks such as getting dressed, combing hair, etc. These problems could lead to more serious complications for older adults, who might have additional health issues or mobility issues [7]. Early detection of changes in bathroom use could alert caregivers before the occurrence of a significant health event. Finally, the increase in activity at the front door may indicate a rise in both residents and building service providers, such as building managers and caregivers, frequenting that area to help support residents meeting a variety of needs (eg, retrieve meals, send mail, wellness, and vital checks). This could be attributed to the surge in the use of food delivery services following the stay-at-home orders during the COVID-19 pandemic [35,36].

Regarding the gait parameters, our findings of increased height and stride time, and decreased walk speeds could suggest a decline in physical activity and mobility, which can have negative effects on overall health and well-being. According to a study [37] examining the impact of COVID-19 on older adults using the Kihon checklist regarding physical activity, a significant decline in total physical activity time for all frailty categories was reported. The decline in physical activity, which aligns with our study's findings of decreased stride length and walk speeds, suggests that the confinement during the pandemic may have had a negative effect on the physical activity levels of older adults.

Our findings of increased restlessness and seconds spent in bed may be an indication of poor sleep quality, which can lead to negative mental health outcomes, including depression. As opportunities for socialization diminished, older adults spent more time in bed as they were adjusting to a more confined lifestyle. In a study [38], the researchers explored the relationship between social isolation, depression, and the well-being of older adults. Their findings revealed that older adults who experienced social isolation had a greater likelihood of developing depression and anxiety. In addition, A recent report indicates that physical exercise helped promote mental well-being while reducing the severity of acute respiratory infection symptoms and the number of symptom days (−2.24 d) during the follow-up period [39]. Findings from these studies further support that the social isolation resulting from the COVID-19 pandemic could be associated with symptoms of depression [35].

In examining the decline in average pulse and respiratory rates in our findings, research [40,41] shows that reduced physical activity, particularly in older adults with high-risk or cardiovascular disease, was a widespread consequence of lockdowns, leading to lower metabolic rates and decreased cardiovascular output, which could explain the drop in vitals such as heart and respiratory rates. However, the anticipated increase in stress-induced vitals, due to the pandemic's mental health toll, is not consistently reflected in physiological markers. While anxiety and depression rates among older adults did increase during COVID-19, research shows the factors such as older adults' emotional resilience and ability to manage stress may suggest that these psychological stressors do not consistently lead to elevated pulse or respiratory rates [42]. The physiological response to stress during quarantine may have been modulated by factors such as extended rest periods and reduced physical activity. Future research should investigate the complex interplay between stress, medication use, and physiological markers while also examining the long-term health implications of sedentary behavior in older adults. This insight is particularly relevant during prolonged periods of isolation or crisis, as it can inform policymakers in the development of remote health monitoring systems aimed at supporting the physical and mental well-being of elderly populations.

The variability of movement shown in the bedroom among participants, indicated by statistical insignificance in their movement patterns (P=.87), suggests that while home confinement significantly affects movement patterns mentioned above, individual responses can vary greatly. This variability highlights the need for personalized health care strategies, as home confinement impacts different people in different ways. For example, differences in movement patterns may be influenced by factors such as biological sex, age, or pre-existing health conditions. A recent study highlights how these demographics, such as biological sex and social determinants of health, can affect the psychological impact during home confinement [43]. Similarly, another systematic review on physical activities during home confinement points out the need for tailored recommendations for vulnerable populations, including older adults, further supporting the need for precision health care approaches [44].

Using in-home sensors for aging-in-place older adults can provide insights into their behavioral patterns. Moreover, a significant number of older adults prefer to remain in their home environment rather than move into institutional settings like nursing homes. The use of in-home sensor systems could accommodate this desire and maintain safety. Different types of contactless sensors can be integrated into the homes' infrastructure, for example, motion, temperature, pressure or bed sensors, and infrared imaging. These sensors enable the monitoring of activities of daily living of older adults such as personal hygiene, toileting, cooking, and managing medications. Gradual and acute changes can be detected by these sensors and shared with professional and informal caregivers allowing them the opportunity to provide assistance or early intervention. There has been long-standing interest surrounding in-home sensor systems for older adults as this technology could support their ability to live independently and delay institutionalization [45-48]. However, with the recent global COVID-19 pandemic, in-home sensor systems may have a secondary benefit. Sensor-based health parameter data could provide evidence for developing recommendations around when home confinement is necessary and when the adverse effects of staying home (eg, social isolation, decreased physical exercise, and poor mental health) outweigh the benefits. Such recommendations could provide support for evidence-based clinical decision-making.

XSL•FO
RenderX

This study has important implications for individuals providing care support to older adults aging in place. Study findings suggest an opportunity to use remote sensing technologies in the ongoing monitoring of community-dwelling older adults, particularly in times when quarantine and social distancing measures are in place. Previous research suggests that older adults living with embedded, sensor technologies found a sense of comfort and security in having the technologies installed, and a desire for those providing care support to view the data [49]. Strides have been made over the past 10 years to create a consumer-facing interface for older adults and their caregivers to access the sensor-generated data [50]. Individuals who provide care support to the aging population can review sensor-generated data to observe changes in activity levels and physiological functioning. Trends in the data over time can help inform early, at-home interventions. For instance, in this study the increase in restlessness in bed and time spent in bed could provide an opportunity for a conversation about sleep hygiene and mental health. Interventions to engage the older adult in activities that promote cognitive stimulation could be negotiated between the older adult and their caregiver. Social engagement activities can also be a point of intervention, even during quarantine and social distancing, such as video conferencing with loved ones and attending various group activities internet-based or remotely. Participation in such activities during the COVID-19 pandemic provided older adults with increased entertainment and mental stimulation [51].

Limitations

While our secondary analysis using remote sensors reveals significant findings about the impact of confinement on older adults, it is important to recognize some inherent limitations in our methodology. First, the dependence on observable metrics might have hindered fully capturing the range of subtleties associated with older adults' adherence. The nuances of individual and facility staff compliance, the variances in the use of social distancing techniques, or the degrees of adherence within particular demographic groups, for example, may not be fully captured by our assessment. Second, the unprecedented circumstances of the COVID-19 pandemic presented challenges in maintaining sensor networks, particularly for motion activity tracking. The need for manual battery replacements disrupted continuous data collection, potentially affecting the accuracy and reliability of insights derived from motion sensors. Third, while the behavioral markers used in the study are supported by previous research and established methodologies, their validation within the specific context of this cohort remains a limitation. Future studies should prioritize incorporating such validation measures to ensure the reliability of identified markers. Fourth, the sample consisted solely of residents from senior living facilities in mid-Missouri, which may limit the generalizability of the findings. This may limit the generalizability of the findings, as the sample may not adequately represent older adults residing in different geographic regions or those with diverse socioeconomic backgrounds with disparities of technology access in mind. As such, this limitation should be considered when drawing conclusions from the study's findings. Fifth, the limited exploration of the psychological impacts of isolation, with a focus on basic physical and mental health metrics, highlights the need for future studies to incorporate qualitative data, such as patient interviews or psychological assessments, to gain a more nuanced understanding of participants' mental health. Finally, while precautions are taken in ensuring accurate sensor measurements, the introduction of noise from pets and visits does occur and analysis was conducted on the daily aggregation of sensor measurements to help account for this noise. These limitations underscore the need for enhanced validation strategies, broader participant diversity, and more comprehensive data integration in future research.

A recent systematic review discerns the bidirectionality between loneliness and the functional trajectory of aging older adults [52]. This observation emphasizes the need for further investigative endeavors aimed at identifying the latent confounders affecting these phenomena. To systematically identify these determinants, the acquisition of a continuous, high-fidelity data stream is imperative. In this regard, the deployment of remote passive sensors emerges as a potent modality, while ensuring minimal disruption to aging in place older adults. In future research endeavors, the implementation of causal inference analyses on the dataset obtained through the aforementioned methods holds promise for uncovering the intricate causal foundations that underlie, the complex framework of aging in place among older adults [53].

Conclusions

This study presents important implications for the health and well-being of aging in place older adults, particularly during times of home confinement. The results suggest the potential value of increased support and interventions to help maintain physical and cognitive health, while also indicating the importance of tailoring these interventions to the unique needs and circumstances of each individual. By recognizing and addressing the unique challenges faced by older adults, we can design interventions that effectively support physical, cognitive, and emotional health.

Acknowledgments

This research was supported by R01 grants from the US National Institutes of Health (NIH; R01NR016423; principal investigators Marge Skubic, PhD) and National Library of Medicine (NLM; R01LM012221; principal investigator Mihail Popescu, PhD). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH or NLM.

XSL•FO
RenderX

Authors' Contributions

KL performed conceptualization, data curation, methodology, resources, software, supervision, validation, visualization, and wrote the original draft. NM performed data curation, formal analysis, methodology, software, and contributed to writing – review and editing. ELR performed conceptualization, methodology, resources, validation, and contributed to writing – review and editing. KRP performed investigation, methodology, resources, supervision, validation, and contributed to writing – review and editing.

Conflicts of Interest

None declared.

References

1. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2023-10-30]
2. COVID-19 mortality overview. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/covid19/mortality-overview.htm [accessed 2023-01-26]
3. Yanez ND, Weiss NS, Romand JA, et al. COVID-19 mortality risk for older men and women. BMC Public Health 2020 Nov 19;20(1):1742. [doi: 10.1186/s12889-020-09826-8] [Medline: 33213391]
4. Or G, Levi-Belz Y, Aisenberg D. Death anxiety and intrusive thinking during the COVID-19 pandemic. GeroPsych (Bern) 2021 Dec;34(4):201-212. [doi: 10.1024/1662-9647/a000268]
5. Rodrigues NG, Han CQY, Su Y, et al. Psychological impacts and online interventions of social isolation amongst older adults during COVID-19 pandemic: a scoping review. J Adv Nurs 2022 Mar;78(3):609-644. [doi: 10.1111/jan.15063] [Medline: 34625997]
6. Ammar A, Trabelsi K, Brach M, et al. Effects of home confinement on mental health and lifestyle behaviours during the COVID-19 outbreak: insights from the ECLB-COVID19 multicentre study. Biol Sport 2021 Mar;38(1):9-21. [doi: 10.5114/biolsport.2020.96857] [Medline: 33795912]
7. Mantantzis K, Drewelies J, Duezel S, et al. Dehydration predicts longitudinal decline in cognitive functioning and well-being among older adults. Psychol Aging 2020 Jun;35(4):517-528. [doi: 10.1037/pag0000471] [Medline: 32352804]
8. Yagmaee F, Taheri M, Irandoust K, et al. Lifestyle behaviors predict COVID-19 severity: a cross-sectional study. Int J Sport Stud Health 2023;6(2). [doi: 10.5812/intjssh-142413]
9. Taheri M, Esmaeili A, Irandoust K, et al. Mental health, eating habits and physical activity levels of elite Iranian athletes during the COVID-19 pandemic. Sci Sports 2023 May 29;38(5-6):527-533. [doi: 10.1016/j.scispo.2023.01.002] [Medline: 37362084]
10. de Leo D, Trabucchi M. COVID-19 and the fears of Italian senior citizens. Int J Environ Res Public Health 2020 May 20;17(10):3572. [doi: 10.3390/ijerph17103572] [Medline: 32443683]
11. Staying physically active during the COVID-19 pandemic. American College of Sports Medicine. URL: https://www.acsm.org/education-resources/trending-topics-resources/covid-19/active-at-home [accessed 2023-08-28]
12. Coronavirus precautions for patients and others facing higher risks. American Heart Association. 2020. URL: https://www.heart.org/en/coronavirus/coronavirus-covid-19-resources/coronavirus-precautions-for-patients-and-others-facing-higher-risks [accessed 2023-08-28]
13. WHO updates COVID-19 guidelines on masks, treatments, and patient care. World Health Organization. 2023. URL: https://www.who.int/news/item/13-01-2023-who-updates-covid-19-guidelines-on-masks-treatments-and-patient-care#:~:text=For%20those%20who%20test%20positive,an%20antigen%2Dbased%20rapid%20test [accessed 2023-08-28]
14. Innovative aging solutions for predictive and preventative care. Foresite. URL: https://www.foresitehealthcare.com [accessed 2023-07-01]
15. Statistical functions. SciPy. URL: https://docs.scipy.org/doc/scipy/reference/stats.html#module-scipy.stats [accessed 2023-07-01]
16. Heise D, Skubic M. Monitoring pulse and respiration with a non-invasive hydraulic bed sensor. Presented at: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology; Aug 31 to Sep 4, 2010; Buenos Aires, Argentina. [doi: 10.1109/IEMBS.2010.5627219]
17. Heise D, Rosales L, Skubic M, et al. Refinement and evaluation of a hydraulic bed sensor. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:4356-4360. [doi: 10.1109/IEMBS.2011.6091081] [Medline: 22255304]
18. Rosales L, Su BY, Skubic M, Ho KC. Heart rate monitoring using hydraulic bed sensor ballistocardiogram1. AIS 2017;9(2):193-207. [doi: 10.3233/AIS-170423]
19. Starr I, Rawson AJ, Schroeder HA, et al. Studies on the estimation of cardiac ouptut in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts; the ballistocardiogram. Am J Physiol-Legacy Content 1939 Jul 31;127(1):1-28. [doi: 10.1152/ajplegacy.1939.127.1.1]
20. Alexander GL, Wakefield BJ, Rantz M, et al. Passive sensor technology interface to assess elder activity in independent living. Nurs Res 2011;60(5):318-325. [doi: 10.1097/NNR.0b013e318225f3e1] [Medline: 21873920]

XSL•FO
RenderX

21. Stone EE, Skubic M. Unobtrusive, continuous, in-home gait measurement using the Microsoft Kinect. IEEE Trans Biomed Eng 2013 Oct;60(10):2925-2932. [doi: 10.1109/TBME.2013.2266341] [Medline: 23744661]
22. Stone E, Skubic M. Passive, in-home gait measurement using an inexpensive depth camera: initial results. Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2012; San Diego, United States. [doi: 10.4108/icst.pervasivehealth.2012.248731]
23. Su BY, Ho KC, Skubic M, et al. Pulse rate estimation using hydraulic bed sensor. : IEEE Presented at: 2012 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 28 to Sep 1, 2012; San Diego, CA p. 2587-2590. [doi: 10.1109/EMBC.2012.6346493]
24. Lydon K, Rosales L, et al. Robust heartbeat detection from in-home ballistocardiogram signals of older adults using a bed sensor. Presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Milan p. 7175-7179. [doi: 10.1109/EMBC.2015.7320047]
25. Wang S, Skubic M, Zhu Y. Activity density map visualization and dissimilarity comparison for eldercare monitoring. IEEE Trans Inf Technol Biomed 2012 Jul;16(4):607-614. [doi: 10.1109/TITB.2012.2196439] [Medline: 22547460]
26. Wang S, Skubic M, Zhu Y, et al. Using passive sensing to estimate relative energy expenditure for eldercare monitoring. Proc IEEE Int Conf Pervasive Comput Commun 2011 Mar 21:642-648. [doi: 10.1109/PERCOMW.2011.5766968] [Medline: 25266777]
27. Banerjee T, Keller JM, Skubic M. Resident identification using kinect depth image data and fuzzy clustering techniques. Presented at: 2012 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 28 to Sep 1, 2012; San Diego, CA p. 5102-5105. [doi: 10.1109/EMBC.2012.6347141]
28. CMS announces new measures to protect nursing home residents from COVID-19. Centers for Medicare & Medicaid Services. 2020 May 13. URL: https://www.cms.gov/newsroom/press-releases/cms-announces-new-measures-protect-nursing-home-residents-covid-19 [accessed 2023-11-02]
29. Larsen RN, Kingwell BA, Sethi P, et al. Breaking up prolonged sitting reduces resting blood pressure in overweight/obese adults. Nutr Metab Cardiovasc Dis 2014 Sep;24(9):976-982. [doi: 10.1016/j.numecd.2014.04.011] [Medline: 24875670]
30. Dunstan DW, Kingwell BA, Larsen R, et al. Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 2012 May;35(5):976-983. [doi: 10.2337/dc11-1931] [Medline: 22374636]
31. Hamilton MT, Hamilton DG, Zderic TW. Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease. Diabetes 2007 Nov;56(11):2655-2667. [doi: 10.2337/db07-0882] [Medline: 17827399]
32. Manz F. Hydration and disease. J Am Coll Nutr 2007 Oct;26(5 Suppl):535S-541S. [doi: 10.1080/07315724.2007.10719655] [Medline: 17921462]
33. Shepard CL, Wang G, Hopson BD, et al. Urinary tract stone development in patients with myelodysplasia subjected to augmentation cystoplasty. Rev Urol 2017;19(1):11-15. [doi: 10.3909/riu0741] [Medline: 28522925]
34. Hooper L, Bunn D, Jimoh FO, et al. Water-loss dehydration and aging. Mech Ageing Dev 2014;136-137:50-58. [doi: 10.1016/j.mad.2013.11.009] [Medline: 24333321]
35. Senior meal plans adjust, add deliveries as centers close amid COVID-19. MPR News. 2020. URL: https://www.mprnews.org/story/2020/03/20/senior-meal-plans-adjust-add-deliveries-as-centers-close-amid-covid19 [accessed 2023-03-04]
36. Steinman MA, Perry L, Perissinotto CM. Meeting the care needs of older adults isolated at home during the COVID-19 pandemic. JAMA Intern Med 2020 Jun 1;180(6):819-820. [doi: 10.1001/jamainternmed.2020.1661] [Medline: 32297903]
37. Yamada M, Kimura Y, Ishiyama D, et al. Effect of the COVID-19 epidemic on physical activity in community-dwelling older adults in Japan: a cross-sectional online survey. J Nutr Health Aging 2020 Sep;24(9):948-950. [doi: 10.1007/s12603-020-1501-6]
38. Robb CE, de Jager CA, Ahmadi-Abhari S, et al. Associations of social isolation with anxiety and depression during the early COVID-19 pandemic: a survey of older adults in London, UK. Front Psychiatry 2020;11:591120. [doi: 10.3389/fpsyt.2020.591120] [Medline: 33132942]
39. Grande AJ, Keogh J, Silva V, et al. Exercise versus no exercise for the occurrence, severity, and duration of acute respiratory infections. Cochrane Database Syst Rev 2020 Apr 4;4(4):CD010596. [doi: 10.1002/14651858.CD010596.pub3] [Medline: 32246780]
40. Wilke J, Rahlf AL, Füzéki E, et al. Physical activity during lockdowns associated with the COVID-19 pandemic: a systematic review and multilevel meta-analysis of 173 studies with 320,636 participants. Sports Med Open 2022 Oct 11;8(1):125. [doi: 10.1186/s40798-022-00515-x] [Medline: 36219269]
41. Elsworthy RJ, Jong ST, Hanson S, et al. Effects of the COVID-19 associated United Kingdom lockdown on physical activity in older adults at high risk of cardiovascular disease: a mixed methods perspective from the MedEx-UK multicenter trial. Front Public Health 2024;12:1371453. [doi: 10.3389/fpubh.2024.1371453] [Medline: 38784572]
42. Vahia IV, Jeste DV, Reynolds CF. Older adults and the mental health effects of COVID-19. JAMA 2020 Dec 8;324(22):2253-2254. [doi: 10.1001/jama.2020.21753] [Medline: 33216114]
43. Rodríguez-Fernández P, González-Santos J, Santamaría-Peláez M, et al. Psychological effects of home confinement and social distancing derived from COVID-19 in the general population-a systematic review. Int J Environ Res Public Health 2021 Jun 17;18(12):6528. [doi: 10.3390/ijerph18126528] [Medline: 34204403]

XSL•FO
RenderX

44. Bentlage E, Ammar A, How D, et al. Practical recommendations for maintaining active lifestyle during the COVID-19 pandemic: a systematic literature review. Int J Environ Res Public Health 2020 Aug 28;17(17):6265. [doi: 10.3390/ijerph17176265] [Medline: 32872154]
45. Baig MM, Afifi S, GholamHosseini H, et al. A systematic review of wearable sensors and IoT-based monitoring applications for older adults - a focus on ageing population and independent living. J Med Syst 2019 Jun 15;43(8):1-11. [doi: 10.1007/s10916-019-1365-7] [Medline: 31203472]
46. Peetoom KKB, Lexis MAS, Joore M, et al. Literature review on monitoring technologies and their outcomes in independently living elderly people. Disabil Rehabil Assist Technol 2015 Jul;10(4):271-294. [doi: 10.3109/17483107.2014.961179] [Medline: 25252024]
47. Rantz M, Phillips LJ, Galambos C, et al. Randomized trial of intelligent sensor system for early illness alerts in senior housing. J Am Med Dir Assoc 2017 Oct;18(10):860-870. [doi: 10.1016/j.jamda.2017.05.012]
48. Rantz M, Popejoy LL, Galambos C, et al. The continued success of registered nurse care coordination in a state evaluation of aging in place in senior housing. Nurs Outlook 2014;62(4):237-246. [doi: 10.1016/j.outlook.2014.02.005] [Medline: 24731918]
49. Galambos C, Rantz M, Craver A, et al. Living with intelligent sensors: older adult and family member perceptions. Comput Inform Nurs 2019 Dec;37(12):615-627. [doi: 10.1097/CIN.0000000000000555] [Medline: 31498250]
50. Robinson EL, Park G, Lane K, et al. Technology for healthy independent living: creating a tailored in-home sensor system for older adults and family caregivers. J Gerontol Nurs 2020 Jul 1;46(7):35-40. [doi: 10.3928/00989134-20200605-06] [Medline: 32597999]
51. Zhao W, Kelly RM, Rogerson MJ, et al. Understanding older adults' participation in online social activities: lessons from the COVID-19 pandemic. Presented at: Proceedings of the ACM on Human-Computer Interaction; Nov 11, 2022 p. 1-26. [doi: 10.1145/3564855]
52. Pollak C, Verghese J, Blumen H. Loneliness and functional decline in aging: a systematic review. Res Gerontol Nurs 2023;16(4):202-212. [doi: 10.3928/19404921-20230503-02] [Medline: 37159388]
53. Rawls E, Kummerfeld E, Zilverstand A. An integrated multimodal model of alcohol use disorder generated by data-driven causal discovery analysis. Commun Biol 2021 Mar 31;4(1):435. [doi: 10.1038/s42003-021-01955-z] [Medline: 33790384]

Abbreviations

ACSM: American College of Sports Medicine
AHA: American Heart Association
CMS: Centers for Medicare and Medicaid Services
PIR: passive infrared
WHO: World Health Organization

Edited by L Buis, TDA Cardoso; submitted 23.01.24; peer-reviewed by D Cook, M Taheri; revised version received 09.12.24; accepted 25.03.25; published 02.06.25.

Please cite as:
Lee K, Marchal N, L Robinson E, Powell KR
Behavioral Markers in Older Adults During COVID-19 Confinement: Secondary Analysis of In-Home Sensor Data
JMIR Mhealth Uhealth 2025;13:e56678
URL: https://mhealth.jmir.org/2025/1/e56678
doi:10.2196/56678

© Knoo Lee, Noah Marchal, Erin L Robinson, Kimberly R Powell. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparative Effectiveness of Wearable Devices and Built-In Step Counters in Reducing Metabolic Syndrome Risk in South Korea: Population-Based Cohort Study

Kyung-In Joung[1], PhD; Sook Hee An[2], PhD; Joon Seok Bang[3], PhD; Kwang Joon Kim[4], PhD

[1]School of AI Healthcare, College of Integrated Health Science, CHA University, Pocheon-si, Republic of Korea

[2]College of Pharmacy, Wonkwang University, Iksan, Republic of Korea

[3]College of Pharmacy, Sookmyung Women's University, Seoul, Republic of Korea

[4]College of Pharmacy, Chonnam National University, BLDG1, Room No. 407, 77 Yongbong-ro, Gwangju, Republic of Korea

Corresponding Author:
Kwang Joon Kim, PhD
College of Pharmacy, Chonnam National University, BLDG1, Room No. 407, 77 Yongbong-ro, Gwangju, Republic of Korea

Abstract

Background: Mobile health technologies show promise in addressing metabolic syndrome, but their comparative effectiveness in large-scale public health interventions remains unclear.

Objective: This study aims to compare the effectiveness of wearable devices (wearable activity trackers) and mobile app–based activity trackers (built-in step counters) in promoting walking practice, improving health behaviors, and reducing metabolic syndrome risk within a national mobile health care program operated by the Korea Health Promotion Institute.

Methods: This retrospective cohort study analyzed data from 46,579 participants in South Korea's national mobile health care program (2020 - 2022). Participants used wearable devices for 12 weeks, after which some switched to built-in step counters. The study collected data on demographics, health behaviors, and metabolic syndrome risk factors at baseline, 12 weeks, and 24 weeks. Outcomes included changes in walking practice, health behaviors, and metabolic syndrome risk factors. Metabolic syndrome risk was assessed based on 5 factors: blood pressure, fasting glucose, waist circumference, triglycerides, and high-density lipoprotein cholesterol. Health behaviors included low-sodium diet preference, nutrition label reading, regular breakfast consumption, aerobic physical activity, and regular walking. To address potential selection bias, propensity score matching was performed, balancing the 2 groups on baseline characteristics including age, gender, education level, occupation, insurance type, smoking status, and alcohol consumption.

Results: Both wearable activity tracker and built-in step counter groups exhibited significant improvements across all evaluated outcomes. The improvement rates for regular walking practice, health behavior changes, and metabolic syndrome risk reduction were high in both groups, with percentages ranging from 45.2% to 60.8%. After propensity score matching, both device types showed substantial improvements across all indicators. The built-in step counter group demonstrated greater reductions in metabolic syndrome risk compared to the wearable device group (odds ratio [OR] 1.20, 95% CI 1.05 - 1.36). No significant differences were found in overall health behavior improvements (OR 0.95, 95% CI 0.83 - 1.09) or walking practice (OR 0.84, 95% CI 0.70 - 1.01) between the 2 groups. Age-specific subgroup analyses revealed that the association between built-in step counters and metabolic syndrome risk reduction was more pronounced in young adults aged 19 - 39 years (OR 1.35, 95% CI 1.09 - 1.68). Among Android use subgroups, built-in step counters were associated with a higher reduction in health risk factors (OR 1.20, 95% CI 1.03 - 1.39).

Conclusions: Both wearable devices and built-in step counters effectively reduced metabolic syndrome risk in a large-scale public health intervention, with built-in step counters showing a slight advantage. The findings suggest that personalized device recommendations based on individual characteristics, such as age and specific health risk factors, may enhance the effectiveness of mobile health interventions. Future research should explore the mechanisms behind these differences and their long-term impacts on health outcomes.

(JMIR Mhealth Uhealth 2025;13:e64527) doi:10.2196/64527

KEYWORDS

wearable devices; built-in step counters; mobile health; public health intervention; physical activity; health behavior; metabolic syndrome; population-based; cohort study; South Korea; mobile health technologies; effectiveness; activity tracker; mobile app; retrospective; logistic regression; mHealth; digital health; mobile phone

XSL•FO
RenderX

Introduction

Metabolic syndrome, a cluster of interconnected factors, significantly increases the risk of cardiovascular disease, stroke, and type 2 diabetes, presenting a major challenge to global public health systems [1-3]. Its rising prevalence imposes substantial economic burdens on health care systems and society [4]. In response, digital health technologies, particularly mobile apps and wearable devices, have emerged as promising tools for health promotion and disease prevention. These technologies offer advantages such as real-time data collection, personalized feedback, and continuous monitoring, potentially enhancing user engagement and promoting sustainable behavior change to reduce metabolic syndrome risks [5,6].

Recent studies have demonstrated the potential of mobile health apps in supporting behavior change for diabetes prevention and control. A prospective cohort study in Singapore revealed that higher engagement with a mobile health app led to greater weight loss and HbA_{1c} reduction among adults with overweight or obesity and who had type 2 diabetes or prediabetes [7]. Similarly, a qualitative research highlighted the importance of integrating app use into routine care and developing guidelines for health care professionals to maximize the benefits of these technologies [8]. These findings underscore the growing importance of digital interventions in chronic disease management.

However, the efficacy of these digital interventions in public health settings remains a subject of debate. While some studies have reported significant benefits in health promotion and metabolic syndrome management [9,10], others have found limited or no effects [11,12]. This discrepancy underscores the need for further research to elucidate the factors that influence the effectiveness of digital health interventions in real-world settings.

Wearable devices are hypothesized to be more effective in promoting health behaviors due to their ability to provide continuous monitoring and immediate feedback, potentially enhancing user engagement and adherence to exercise regimens than mobile apps alone [10,13,14]. However, evidence supporting this hypothesis in large-scale public health interventions remains limited and inconsistent. Some studies suggest that wearables enhance physical activity and health monitoring more effectively than mobile apps due to their continuous tracking and user engagement features [15]. Conversely, other research indicates minimal differences in health outcomes between users of wearables and mobile apps, suggesting that the technology's effectiveness may heavily rely on individual user engagement and integration into comprehensive health management strategies [16,17].

In South Korea, a nationwide mobile health care program operated by the Korea Health Promotion Institute (KHPI) has been operational since 2018, providing an unprecedented opportunity to evaluate the effectiveness of digital health interventions at a population level. This program, implemented through over 220 health centers, offers free wearable activity trackers to eligible participants to mitigate metabolic syndrome risks. Interestingly, some participants have opted to switch to using mobile phone built-in step counters due to personal preference or device loss, creating a natural experiment to compare the efficacy of these 2 approaches.

This study aims to comprehensively evaluate the comparative effectiveness of wearable devices (wearable activity trackers) versus mobile app–based activity trackers (built-in step counters) within this public health service framework. Specifically, we seek to examine their differential impacts on three key outcomes: (1) the promotion of regular walking practices, (2) improvements in overall health behaviors, and (3) the reduction of metabolic syndrome risk factors. By leveraging a large-scale, real-world dataset, we aim to provide nuanced insights into how these technologies influence various aspects of health behavior and outcomes.

Methods

Data Source

The data source is the KHPI's health center mobile health care program data (2020 - 2022). This nationwide program uses Information and Communication Technology to improve preventive health care and self-health capabilities of local residents aged 19 years and older with at least one metabolic syndrome risk factor. The program targets individuals who are generally healthy or without diagnosed chronic diseases. Participants undergo comprehensive health assessments at baseline, 3 months, and 6 months, supported by continuous mobile health care interventions throughout the 6-month program duration.

On enrollment, participants receive a mobile app and an activity tracker (wristband or smartwatch) that monitors biometric data, syncing with the app. The app features automated exercise and dietary logging, delivers personalized health information, and sends exercise and food diaries to participants. From week 12, participants could opt to use their smartphone's built-in step counter instead of the provided tracker, integrating with fitness tracker apps.

Study Design and Participants

Participants receive medical check-ups, lifestyle assessments, and expert counseling at the beginning, week 12, and week 24. Exclusion criteria included: (1) nonparticipation for over 2 weeks, (2) onset of chronic diseases during the study, (3) incomplete final health questionnaire or pedometer loss by week 27, and (4) moving out of the service area. The Ministry of Health and Welfare oversees the program, with management by the KHPI (Figure 1) [18].

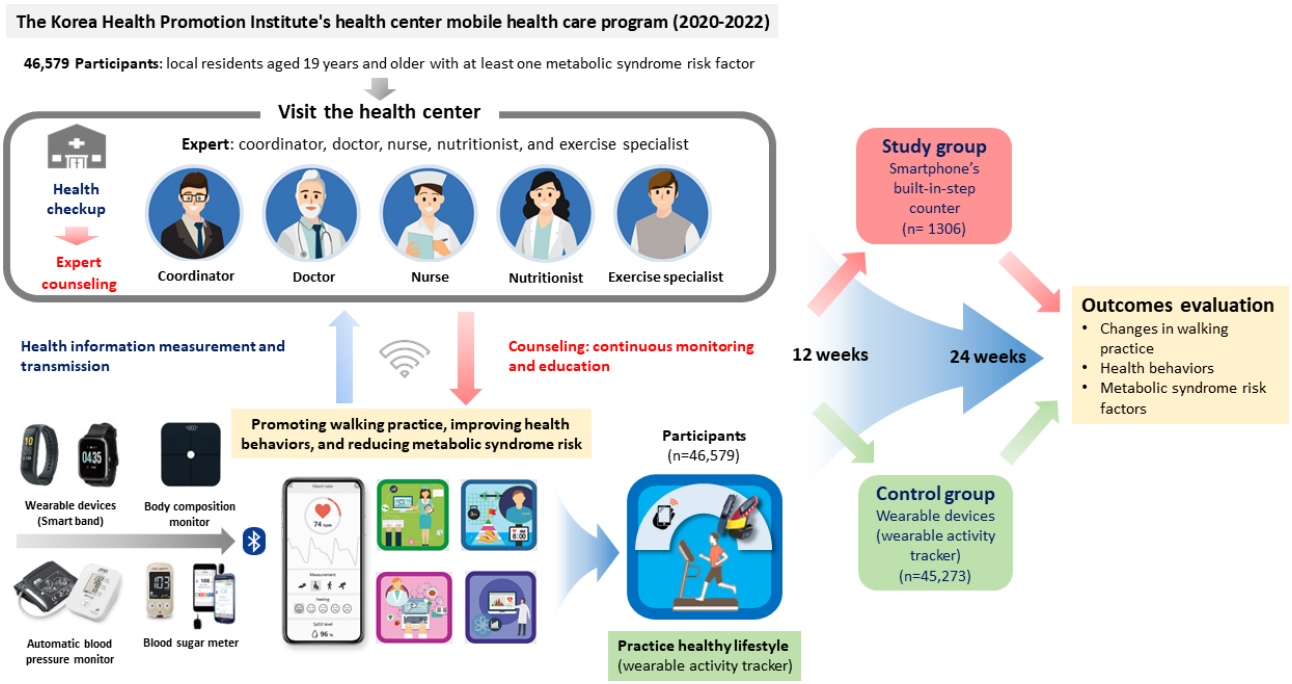
This retrospective cohort study examined changes in walking routines, health behavior patterns, and metabolic syndrome risk factors among participants who either continued using a wearable activity tracker or switched to a smartphone's built-in step counter after 12 weeks of the program.

The study group comprised individuals with at least 1 metabolic syndrome risk factor who switched from a wearable activity tracker to their smartphone's built-in step counter after 12 weeks. The control group consisted of individuals with at least 1 metabolic syndrome risk factor who continued using a

XSL•FO
RenderX

wearable activity tracker throughout the 24-week intervention period.

Figure 1. Health center mobile health care program and study design.



Outcome Measures

The practice of "regular walking" was defined based on a survey as engaging in walking for at least 10 minutes consecutively for 5 or more days in the past week. If individuals who did not practice "regular walking" before participating in the program started practicing it afterward, it was considered an improvement in "regular walking."

Health behaviors were assessed through a comprehensive survey that evaluated 5 key indicators: preference for low-salt diets, regular breakfast consumption, attentiveness to nutritional labels, engagement in regular walking, and participation in aerobic exercises. If any of these 5 health behaviors were not practiced before the program but were adopted afterward, it was considered an improvement in health behaviors.

Metabolic syndrome risk factors were composed of the following five criteria, adopting the diagnostic criteria used domestically based on the National Cholesterol Education Program Adult Treatment Panel III guidelines: (1) elevated blood pressure (systolic blood pressure equal to or above 130 mm Hg or diastolic blood pressure equal to or above 85 mm Hg); (2) elevated fasting blood glucose equal to or above 100 mg/dL; (3) abdominal obesity: male waist circumference equal to or above 90 cm, female waist circumference equal to or above 85 cm; (4) elevated triglycerides equal to or above 150 mg/dL; and (5) decreased high-density lipoprotein cholesterol: male less than 40 mg/dL, female less than 50 mg/dL. If 1 or more of these 5 metabolic syndrome risk factors decreased after participation in the program compared to before, it was considered a reduction in metabolic syndrome risk [19].

Potential Confounding Variables

The following potential confounding variables were included as covariates: gender, age group (in 10 y increments), BMI, education level (4 levels), occupation, family type, insurance type (as a proxy for economic status), smoking, alcohol consumption, and mobile operating system.

Statistical Analysis

Baseline demographic characteristics were compared between the built-in step counter users and wearable activity tracker users using chi-square tests. To enhance comparability between the study group and the control group, we performed propensity score matching as our primary analysis method [20]. Propensity scores were estimated using multivariable logistic regression for all variables mentioned earlier. Matching was performed using a 1:4 nearest neighbor method with the estimated propensity score, and the caliper was strictly set at 0.05. The standardized mean difference was used to compare baseline characteristics of the study group and control group, with imbalance defined as an absolute value greater than 0.1.

Using the propensity score-matched data, we assessed the associations between device type and (1) regular walking practice, (2) health behavior improvement, and (3) metabolic syndrome risk reduction. Odds ratios (OR) with 95% CIs were calculated to quantify these associations.

As a secondary analysis, we conducted multivariable logistic regression with backward elimination, adjusting for potential confounding variables, to further validate our findings. This analysis was performed on the full dataset before propensity score matching.

Subgroup analyses were performed based on age groups (10-30 years: young adults; 40-50 years: middle-aged adults; and 60 years and older: older people) and mobile operating systems (Android and iOS) using both the propensity score-matched data and the full dataset with multivariable logistic regression.

XSL•FO
RenderX

Ethical Considerations

This study was approved by the Bioethics Committee of the CHA University Institutional Review Board (approval E1705/001-003). This study used secondary data from the Public Health Center Mobile Health Care Project, a public health initiative in South Korea. The institutional review board deemed this study exempt from the requirement for additional informed consent, as it involved the secondary analysis of deidentified data from an existing public health program. All data were anonymized and deidentified by the public health centers before being provided for research purposes, with all personal identifiers removed to ensure participant privacy and confidentiality. The data management and analysis procedures strictly adhered to the Personal Information Protection Act of South Korea and institutional data security protocols. As this was a secondary analysis of data from a public health program, participants were not provided additional compensation for this specific research study. Their original participation in the Mobile Health Care Project was voluntary and part of a public health service initiative, without monetary compensation.

Results

Compared to continuous wearable users (n=45,273, 97.2%), the proportion of users who switched to built-in step counters was small (n=1306, 2.8%). The built-in step counter group had a higher proportion of younger adults (19 - 39 y), iOS users, and those with higher education levels. This group also included more managers, professionals, and office workers. In contrast, housewives were more likely to continue using wearable devices. Men showed a slightly higher proportion of built-in step counter use compared to women (n=486, 3.2% vs n=820, 2.6%). Among those with a Bachelor's degree or higher, built-in step counter use was slightly more common than wearable activity tracker use (n=1037, 79.4% vs n=33,659, 74.3%; Table 1).

XSL•FO
RenderX

Table . Basic characteristics of the participants by device type.

Variable and class		Wearable activity tracker (n=45,273, 97.2%)	Built-in step counter (n=1306, 2.8%)	Overall	*P* value
Sex, n (%)					<.001
	Male	14,800 (32.7)	486 (37.2)	15,286 (32.8)	
	Female	30,473 (67.3)	820 (62.8)	31,293 (67.2)	
Age group (years), n (%)					<.001
	10-19	79 (0.2)	3 (0.2)	82 (0.2)	
	20-29	3034 (6.7)	163 (12.5)	3197 (6.9)	
	30-39	9652 (21.3)	357 (27.3)	10,009 (21.5)	
	40-49	16,850 (37.2)	426 (32.6)	17,276 (37.1)	
	50-59	12,509 (27.6)	280 (21.4)	12,789 (27.5)	
	60-69	2937 (6.5)	70 (5.4)	3007 (6.5)	
	70-79	206 (0.5)	7 (0.5)	213 (0.5)	
	80-89	6 (0)	0 (0)	6 (0)	
mOS[a], n (%)					<.001
	Android	40,208 (88.8)	976 (74.7)	41,184 (86.7)	
	iOS	5062 (11.2)	330 (25.3)	5392 (13.3)	
Family type, n (%)					.030
	Single person	3727 (8.2)	134 (10.3)	3861 (8.3)	
	2 - 3 members	20,532 (45.4)	573 (43.9)	21,105 (45.3)	
	≥4 members	21,014 (46.4)	599 (45.9)	21,613 (46.4)	
Insurance type, n (%)					.363
	Health insurance	44,652 (98.6)	1287 (98.6)	45,939 (98.6)	
	Medical aid	359 (0.8)	8 (0.6)	367 (0.8)	
	Not enrolled	262 (0.6)	11 (0.8)	273 (0.6)	
Education, n (%)					<.001
	No education to middle school graduate	1037 (2.3)	19 (1.5)	1056 (2.3)	
	High school graduate	10,577 (23.4)	250 (19.1)	10,827 (23.2)	
	Bachelor's degree	29,340 (64.8)	900 (68.9)	30,240 (64.9)	
	Master's degree or higher	4319 (9.5)	137 (10.5)	4456 (9.6)	
Occupation, n (%)					<.001
	Manager, expert	9301 (20.5)	319 (24.4)	9620 (20.7)	
	Office worker, service worker	19,542 (43.2)	579 (44.3)	20,121 (43.2)	
	Sales people, agricultural, forestry, and fishery workers, technician, machine operator, elementary occupation, military personnel	5374 (11.9)	140 (10.7)	5514 (11.8)	
	Student	661 (1.5)	25 (1.9)	686 (1.5)	
	Homemaker	8944 (19.8)	193 (14.8)	9137 (19.6)	
	Unemployed	1451 (3.2)	50 (3.8)	1501 (3.2)	

XSL•FO
RenderX

Variable and class		Wearable activity tracker (n=45,273, 97.2%)	Built-in step counter (n=1306, 2.8%)	Overall	*P* value
Smoking, n (%)					.005
	No smoking	36,630 (80.9)	1031 (78.9)	37,661 (80.9)	
	Used to smoke but currently do not smoke	5037 (11.1)	146 (11.2)	5183 (11.1)	
	Smoke occasionally	815 (1.8)	37 (2.8)	852 (1.8)	
	Smoke every day	2791 (6.2)	92 (7)	2883 (6.2)	
Alcohol intake, n (%)					.61
	<0 - 1 times/month	20,632 (45.6)	558 (42.7)	21,190 (45.5)	
	2 - 4 times/month	16,768 (37)	486 (37.2)	17,254 (37)	
	2 - 3 times/week	6304 (13.9)	211 (16.2)	6515 (14)	
	≥4 times/week	1569 (3.5)	51 (3.9)	1620 (3.5)	
General or intensive management needed[c], n (%)					.98
	General	29,584 (65.4)	853 (65.3)	30,437 (65.3)	
	Intensive	15,689 (34.7)	453 (34.7)	16,142 (34.7)	
BMI[b] (kg/m^2), mean (SD)		25.4 (3.8)	25.7 (4)	25.4 (3.8)	.01
Waist circumference (cm), mean (SD)					
	Male	92.6 (8.9)	93.7 (9.8)	92.6 (9.0)	.02
	Female	85.2 (9.6)	84.5 (10.2)	85.1 (9.6)	.06
Blood pressure (mm Hg), mean (SD)					
	Systolic	126.3 (13.5)	125.4 (13.4)	126.3 (13.5)	.01
	Diastolic	80.4 (10.4)	79.8 (10.5)	80.4 (10.4)	.06
FBG[d] (mg/dL), mean (SD)		98.0 (13.2)	97.8 (11.9)	98.0 (13.1)	.60
Triglyceride (mg/dL), mean (SD)		150.0 (84.9)	150.0 (80.9)	150.0 (84.8)	.99
HDL[e] cholesterol (mg/dL), mean (SD)		55.7 (14.6)	54.9 (14.6)	55.7 (14.6)	.07

[a]mOS: mobile operating system.

[b]BMI: body mass index

[c]Individuals exhibiting 1-2 metabolic syndrome risk factors were allocated to a standard management group, whereas those presenting with 3-5 risk factors were assigned to an intensive management group.

[d]FBG: fasting blood glucose.

[e]HDL: high-density lipoprotein.

Overall, 54% (n=25,158) of participants showed a reduction in metabolic syndrome risk, with women demonstrating a higher rate than men (n=17,371, 55.5% vs n=7787, 50.9%). The reduced risk group had slightly higher proportions of young adults (10 - 39 years), built-in step counter users, and iOS users. Participants with healthier baseline lifestyle indicators (nonsmoking, lower alcohol consumption, and lower BMI) were more prevalent in the reduced-risk group. The proportion of individuals showing a reduction in metabolic syndrome risk was slightly higher among those who practiced fewer health behaviors at baseline (Table 2).

XSL•FO
RenderX

Table . Characteristics of the participants by the reduction in metabolic syndrome risk.

Characteristic		No metabolic syndrome risk reduction (n=21,421, 46%), n (%)	Metabolic syndrome risk reduction (n=25,158, 54%), n (%)	Overall (n=46,579, 100%), n (%)	*P* value
Sex					<.001
	Male	7499 (35)	7787 (31)	15,286 (32.8)	
	Female	13,922 (46)	17,371 (69.1)	31,293 (67.2)	
Age (years)					<.001
	10 - 29	1334 (6.2)	1945 (7.7)	3279 (7)	
	30-39	4612 (21.5)	5397 (21.5)	10,009 (21.5)	
	40-49	7979 (37.3)	9297 (37)	17,276 (37.1)	
	50-59	5912 (27.6)	6877 (27.3)	12,789 (27.5)	
	60-69	1470 (6.9)	1537 (6.1)	3007 (6.5)	
	70 - 89	114 (0.5)	105 (0.4)	219 (0.5)	
Family type					.47
	Single person	1749 (8.2)	2112 (8.4)	3861 (8.3)	
	2 - 3 members	9762 (45.6)	11,343 (45.1)	21,105 (45.3)	
	≥4 members	9910 (46.3)	11,703 (46.5)	21,613 (46.4)	
Education					.03
	No education to middle school graduate	508 (2.4)	548 (2.2)	1056 (2.3)	
	High school graduate	5062 (23.6)	5765 (22.9)	10,827 (23.2)	
	Bachelor's degree	13,762 (64.3)	16,478 (54.5)	30,240 (64.9)	
	Master's degree or higher	2089 (9.8)	2367 (9.4)	4456 (9.6)	
Insurance					.79
	Health insurance	21,125 (98.6)	24,814 (98.6)	45,939 (98.6)	
	Medical aid	174 (0.8)	193 (0.8)	367 (0.8)	
	Not enrolled	122 (0.6)	151 (0.6)	273 (0.6)	
Device type					.007
	Wearable	20,868 (97.4)	24,405 (97)	45,273 (97.2)	
	Switched to app	553 (2.6)	753 (3)	1306 (2.8)	
mOS[a]					<.001
	Android	19,097 (89.2)	22,087 (87.8)	41,184 (88.4)	
	iOS	2323 (10.9)	3069 (12.2)	5392 (11.6)	
Smoking					<.001
	No smoking	17,020 (79.5)	20,641 (82.1)	37,661 (80.9)	
	Used to smoke but currently do not smoke	2516 (11.8)	2667 (10.6)	5183 (11.1)	
	Smoke occasionally	426 (2)	426 (1.7)	852 (1.8)	
	Smoke every day	1459 (6.8)	1424 (5.7)	2883 (6.2)	
Alcohol intake					.001
	<0 - 1 times/month	9597 (44.8)	11,593 (46.1)	21,190 (45.5)	
	2 - 4 times/month	7970 (37.2)	9284 (36.9)	17,254 (37)	
	2 - 3 times/week	3073 (14.4)	3442 (13.7)	6515 (14)	

XSL•FO
RenderX

Characteristic		No metabolic syndrome risk reduction (n=21,421, 46%), n (%)	Metabolic syndrome risk reduction (n=25,158, 54%), n (%)	Overall (n=46,579, 100%), n (%)	*P* value
	≥4 times/week	781 (3.7)	839 (3.3)	1620 (3.5)	
BMI					<.001
	<18.5	221 (1)	381 (1.5)	602 (1.3)	
	18.5 - 23	4597 (21.5)	6657 (26.5)	11,254 (24.2)	
	23 - 25	4346 (20.3)	5887 (23.4)	10,233 (22)	
	25 - 30	9176 (42.8)	9849 (39.2)	19,025 (40.9)	
	30 - 35	2551 (11.9)	1998 (7.9)	4549 (9.8)	
	≥35	530 (2.5)	385 (1.5)	915 (2)	
Number of metabolic syndrome risk factors at baseline					<.001
	1	9439 (44.1)	6784 (27)	16,223 (34.8)	
	2	6705 (31.3)	7509 (29.9)	14,214 (30.5)	
	3	3641 (17)	6176 (24.6)	9817 (21.1)	
	4	1420 (6.6)	3602 (14.3)	5022 (10.8)	
	5	216 (1)	1087 (4.3)	1303 (2.8)	
Hypertension risk					<.001
	Yes	10,157 (47.4)	14,846 (59)	25,003 (53.7)	
	No	11,264 (52.6)	10,312 (41)	21,576 (46.3)	
Diabetes mellitus risk					<.001
	Yes	7079 (33.1)	11,379 (45.2)	18,458 (39.6)	
	No	14,342 (67)	13,779 (54.8)	28,121 (60.4)	
High waist circumference					.16
	Yes	12,298 (57.4)	14,607 (58.1)	26,905 (57.8)	
	No	9123 (42.6)	10,551 (42)	19,674 (42.2)	
High TG[b]					<.001
	Yes	6703 (31.3)	11,806 (46.9)	18,509 (39.7)	
	No	14,718 (68.7)	13,352 (53.1)	28,070 (60.3)	
Low HDL[c] cholesterol					
	Yes	4295 (20.1)	7535 (30)	11,830 (25.4)	
	No	17,126 (80)	17,623 (70.1)	34,749 (74.6)	
Number of health behaviors being practicing at baseline					<.001
	0	221 (1)	381 (1.5)	602 (1.3)	
	1	4597 (21.5)	6657 (26.5)	11,254 (24.2)	
	2	4346 (20.3)	5887 (23.4)	10,233 (22)	
	3	9176 (42.8)	9849 (39.2)	19,025 (40.9)	
	4	2551 (11.9)	1998 (7.9)	4549 (9.8)	
	5	530 (2.5)	385 (1.5)	915 (2)	

[a]mOS: mobile operation system.

[b]TG: triglyceride.

[c]HDL: high density lipoprotein.

After ensuring comparability between the wearable activity tracker and built-in step counter groups through propensity score matching, an analysis was conducted. Before matching, there were differences in various variables such as age, insurance

XSL•FO
RenderX

type, and education level. However, after matching, the absolute values of the standardized mean differences were all below 0.1, indicating no significant differences between the 2 groups (Multimedia Appendix 1).

Table 3 presents the results of the propensity score-matched analysis. For regular walking practice, there was no significant difference between the 2 groups (OR 0.84, 95% CI 0.70 - 1.01). Similarly, no significant differences were found in overall health behavior improvements (OR 0.95, 95% CI 0.83 - 1.09).

Notably, the built-in step counter group demonstrated greater reductions in metabolic syndrome risk compared to the wearable device group (OR 1.20, 95% CI 1.05 - 1.36). This significant association was particularly pronounced among young adults aged 10 - 30 years (OR 1.35, 95% CI 1.09 - 1.68). The effect was less pronounced and not statistically significant for the 40 - 50 years age group (OR 1.15, 95% CI 0.97 - 1.38) and the 60 - 80 years age group (OR 1.08, 95% CI 0.61 - 1.94).

When analyzed by the mobile operating system, Android users showed a significant association between built-in step counter use and metabolic syndrome risk reduction (OR 1.20, 95% CI 1.03 - 1.39). For iOS users, the association was in the same direction but not statistically significant (OR 1.14, 95% CI 0.88 - 1.50), possibly due to the smaller sample size in this subgroup (Table 3).

Table . Propensity score-matched analysis of device type association with walking practice, health behaviors, and metabolic syndrome risk reduction.

Study population and number of participants by device type		Propensity score matched model OR[a] (95% CI)
Not practicing regular walking at baseline[b]		
	Wearable activity tracker (n=2468, 79.9%)	reference
	Built-in step counter (n=621, 20.1%)	0.84 (0.70 - 1.01)
Not practicing any health behaviors at baseline[c]		
	Wearable activity tracker (n=4951, 80%)	reference
	Built-in step counter (n=1244, 20%)	0.95 (0.83 - 1.09)
One or more metabolic syndrome risk factors at baseline[d]		
Overall		
	Wearable activity tracker (n=5195, 80%)	reference
	Built-in step counter (n=1308, 20%)	1.20 (1.05 - 1.36)
10 - 39 years		
	Wearable activity tracker (n=2035, 83%)	reference
	Built-in step counter (n=516, 17%)	1.35 (1.09 - 1.68)
40 - 59 years		
	Wearable activity tracker (n=2808, 80%)	reference
	Built-in step counter (n=704, 20%)	1.15 (0.97 - 1.38)
60 - 89 years		
	Wearable activity tracker (n=288, 79.9%)	reference
	Built-in step counter (n=75, 20.1%)	1.08 (0.61 - 1.94)
Android		
	Wearable activity tracker (n=3893, 80%)	reference
	Built-in step counter (n=975, 20%)	1.20 (1.03 - 1.39)
iOS		
	Wearable (n=1250, 79.3%)	reference
	Built-in step counter (n=326, 21.7%)	1.14 (0.88 - 1.50)

[a]OR: odds ratio.

[b]Outcome: practicing walking.

[c]Outcome: improved in 1 or more health behaviors.

[d]Outcome: reduced in 1 or more metabolic syndrome risk factors.

XSL•FO
RenderX

Across all outcomes—regular walking practice, health behavior improvements, and metabolic syndrome risk reduction—both wearable activity tracker and built-in step counter groups showed high improvement rates, ranging from 45.2% to 60.8% (Table 4).

Table . Sensitivity analysis: multivariable analysis of device type association with walking practice, health behaviors, and metabolic syndrome risk reduction.

Study population and number of participants by device type	Improved, n (%)	Crude model, OR[a] (95% CI)	Covariates model I[b], OR (95% CI)	Covariates model II[c], OR (95% CI)
Not practicing regular walking at baseline[d]				
Wearable activity tracker (n=22,795, 97.3%)	11,705 (51.3)	reference	reference	reference
Built-in step counter (n=624, 2.7%)	282 (45.2)	0.78 (0.67 - 0.92)	0.82 (0.70 - 0.96)	0.83 (0.66 - 1.06)
Not practicing any health behaviors at baseline[e]				
Wearable activity tracker (n=43,666, 96.5%)	26,569 (60.8)	reference	reference	reference
Built-in step activity tracker (n=1253, 2.8%)	730 (58.3)	0.89 (0.80 - 1.00)	0.92 (0.82 - 1.03)	0.94 (0.84 - 1.04)
One or more metabolic syndrome risk factors at baseline[f]				
Wearable activity tracker (n=45,273, 97.2%)	24,405 (53.9)	reference	reference	reference
Built-in step counter (n=1306, 2.8%)	753 (57.7)	1.16 (1.04 - 1.30)	1.16 (1.04 - 1.29)	1.15 (1.03 - 1.30)

[a]OR: odds ratio.

[b]Covariates model Ⅰ: adjusted for age, sex.

[c]Covariates model Ⅱ: adjusted for age, sex, education, family type, insurance type, occupation, BMI, smoking, alcohol intake, mobile operation system type.

[d]Outcome: regular walking practice.

[e]Outcome: improved in 1 or more health behaviors.

[f]Outcome: reduced in 1 or more metabolic syndrome risk factors.

To further validate our findings, we conducted multivariable logistic regression analyses, presented in Table 4. These results were generally consistent with the propensity score-matched analysis. For regular walking practice, wearable activity trackers showed a slight advantage over built-in step counters in the crude model (OR 0.78, 95% CI 0.67 - 0.92). However, this association weakened after adjusting for multiple covariates (fully adjusted OR 0.83, 95% CI 0.66 - 1.06). No significant association was found between device type and overall health behavior improvement across all models (fully adjusted OR 0.94, 95% CI 0.84 - 1.04). Interestingly, the use of built-in step counters was associated with greater reductions in metabolic syndrome risk compared to wearable activity trackers. This relationship remained consistent and statistically significant across all 3 models (crude OR 1.16, 95% CI 1.04 - 1.30; fully adjusted OR 1.15, 95% CI 1.03 - 1.30; Table 4).

Table 5 presents the results of subgroup analyses using multivariable logistic regression. These analyses revealed nuanced differences across age groups and mobile operating systems. Wearable activity trackers were significantly associated with better improvements in regular walking practice only in the 40-59 year old group (OR 0.80, 95% CI 0.64 - 0.99). The device type did not significantly impact health behavior improvement across any age group. For metabolic syndrome risk, built-in step counters were significantly associated with greater risk reduction only among young adults (10-39 years; OR 1.21, 95% CI 1.01 - 1.45), aligning with the findings from the propensity score-matched analysis.

In analyses by the mobile operating system, Android users showed a significant association between built-in step counter use and (1) less regular walking practice (OR 0.81, 95% CI 0.67 - 0.97), and (2) metabolic syndrome risk reduction (OR 1.13, 95% CI 1.00 - 1.29). iOS users demonstrated no significant differences between device types for any outcome, possibly due to limited statistical power.

XSL•FO
RenderX

Table . Sensitivity analysis: multivariable analysis in subgroup for the association of device type with regular walking practice, improvement in health behaviors, and metabolic syndrome risk reduction.

Study population, subgroup, and number of participants by device type			Covariates model II, OR[a] (95% CI)
Not practicing walking at baseline[b]			
	10-39 years		
		Wearable activity tracker (n=6458, 96.1%)	reference
		Built-in step counter (n=261, 3.9%)	0.89 (0.69 - 1.14)
	40-59 years		
		Wearable activity tracker (n=15,144, 97.8%)	reference
		Built-in step counter (n=337, 2.2%)	0.80 (0.64 - 0.99)
	60-89 years		
		Wearable activity tracker (n=1193, 97.9%)	reference
		Built-in step counter 26 (2.1)	0.61 (0.27 - 1.37)
	Android		
		Wearable activity tracker (n=20,287, 97.7%)	reference
		Built-in step counter (n=477, 2.3%)	0.81 (0.67 - 0.97)
	iOS		
		Wearable activity tracker (n=2507, 94.5%)	reference
		Built-in step counter (n=147, 5.5%)	0.89 (0.63 - 1.25)
Not practicing any health behaviors at baseline[c]			
	10-39 years		
		Wearable activity tracker (n=12,270, 96.1%)	reference
		Built-in step counter (n=505, 4%)	0.86 (0.72 - 1.03)
	40-59 years		
		Wearable activity tracker (n=27,856, 97.7%)	reference
		Built-in step counter (n=668, 2.3%)	1.00 (0.85 - 1.17)
	60-89 years		
		Wearable activity tracker (n=2788, 97.5%)	reference
		Built-in step counter (n=73, 2.5%)	0.72 (0.45 - 1.17)
	Android		
		Wearable activity tracker (n=38,190, 97.6%)	reference
		Built-in step counter (n=932, 2.4%)	0.94 (0.83 - 1.08)
	iOS		
		Wearable activity tracker (n=4811, 93.9%)	reference
		Built-in step counter (n=314, 6.1%)	0.85 (0.67 - 1.07)
One or more Metabolic syndrome risk factors at baseline[d]			
	10-39 years		

XSL•FO
RenderX

Study population, subgroup, and number of participants by device type			Covariates model II, OR[a] (95% CI)
		Wearable activity tracker (n=12,765, 96.1%)	reference
		Built-in step counter (n=523, 3.9%)	1.21 (1.01 - 1.45)
	40-59 years		
		Wearable activity tracker (n=29,359, 97.7%)	reference
		Built-in step counter (n=706, 2.4%)	1.11 (0.96 - 1.29)
	60-89 years		
		Wearable activity tracker (n=3149, 97.2%)	reference
		Built-in step counter (n=77, 2.8%)	1.17 (0.74 - 1.86)
	Android		
		Wearable activity tracker (n=40,208, 97.6%)	reference
		Built-in step counter (n=976, 2.4%)	1.13 (1.00 - 1.29)
	iOS		
		Wearable activity tracker (n=5062, 93.9%)	reference
		Built-in step counter (n=330, 6.1%)	1.18 (0.94 - 1.49)

[a]OR: odds ratio.

[b]Outcome: regular walking practice.

[c]Outcome: improved in 1 or more health behaviors.

[d]Outcome: reduced in 1 or more metabolic syndrome risk factors.

Discussion

Principal Results

This study compared the effects of continuously using wearable devices versus switching to built-in step counters after 12 weeks on walking practice, health behaviors, and metabolic syndrome risk reduction. Both groups showed substantial improvements across all indicators, with 45% - 61% of participants demonstrating positive changes.

Using propensity score-matched analysis, we found that the group that switched to built-in step counters showed better results in metabolic syndrome risk reduction (OR 1.20, 95% CI 1.05 - 1.36). This effect was particularly pronounced among young adults aged 10 - 30 years (OR 1.35, 95% CI 1.09 - 1.68). There were no significant differences between groups in walking practice (OR 0.84, 95% CI 0.70 - 1.01) or overall health behaviors (OR 0.95, 95% CI 0.83 - 1.09).

Age-specific subgroup analyses revealed that the effects of device type varied across age groups, with the association between switching to built-in step counters and metabolic syndrome risk reduction being more pronounced in young adults (10-30 years).

Comparison With Previous Studies

Our research findings contribute to the evidence supporting the effectiveness of mobile technology in promoting physical activity and health behaviors. The substantial improvements observed across all indicators in both groups align with previous studies demonstrating the positive impact of smartphone apps and activity trackers on physical activity levels and health outcomes [17,21-24]. Recent research further supports this, showing that higher engagement with mobile health apps led to greater weight loss and HbA_{1c} reduction in adults with overweight or obesity and type 2 diabetes or prediabetes [7].

Our analysis showed no significant difference in walking practice between wearable device users and built-in step counter users (OR 0.84, 95% CI 0.70 - 1.01). This result contrasts with some previous studies demonstrating greater benefits of wearable devices. A randomized controlled trial (RCT) found that among inactive overweight or obese postmenopausal women, a Fitbit-based intervention significantly increased moderate to vigorous physical activity compared to standard pedometers (Fitbit: +62 min/week vs pedometer: –3 min/week; *P*=.01) [25]. Similarly, another RCT reported significantly higher daily average step counts in a group using both app and wearable devices compared to an app-only group (8165 vs 6034 steps; *P*=.02) [26]. The differences between our results and those of previous studies may stem from variations in study design, target population, and intervention duration. For instance, while Cadmus-Bertram et al [25] focused on a specific population (postmenopausal women) and measured moderate to vigorous physical activity as the primary outcome, this study included a broader population and evaluated walking practice and overall health behaviors.

XSL•FO
RenderX

Regarding health behavior improvements, our propensity score-matched analysis found no significant differences between the 2 groups (OR 0.95, 95% CI 0.83 - 1.09). This partially aligns with findings from a previous study, which reported no difference in energy intake changes between mobile app and wearable "Bite Counter" device groups [27]. However, the same study observed a significant increase in physical activity in the wearable device group (mean 2015.4, SD 684.6 metabolic equivalent (METs) min/week; *P*=.02), which contrasts with our findings. Recent research found that both synchronous (videoconference) and asynchronous (prerecorded videos) digital interventions demonstrated good acceptability among inactive adults, suggesting various digital tools can effectively promote physical activity [28]. The importance of integrating app use into routine care has been highlighted, which may explain the similar improvements in both groups in this study [8]. These findings underscore the complexity of digital health interventions and the need for tailored approaches.

In metabolic syndrome risk reduction, our propensity score-matched analysis revealed significant decreases in both groups, with a slight advantage in the group that switched to built-in step counters (OR 1.20, 95% CI 1.05 - 1.36). This contrasts with some previous studies, where mobile phone–based health coaching showed more effectiveness in improving metabolic syndrome–related indicators [29]. However, our results align with other studies demonstrating the potential of mobile software and wearable devices in reducing metabolic syndrome risk [9].

A notable aspect of our findings is the high rate of metabolic syndrome risk reduction (over 50%) in both groups, possibly due to the well-designed intervention characteristics of the national-level public health program. These results are similar to a 12-week RCT where the intervention group receiving physical activity feedback through wearable devices showed significant improvements in key clinical indicators of metabolic syndrome [22]. This supports the potential effectiveness of well-designed mobile health interventions in reducing metabolic syndrome risk.

However, mixed results are also present in the literature. Some studies report that mobile interventions have no significant impact on physical activity or health outcomes [12,30]. For instance, a study on overweight Taiwanese adults found that activity promotion systems provided short-term benefits for physical activity but did not significantly affect metabolic abnormalities [31]. Similarly, an RCT of multidimensional physical activity feedback intervention for primary care patients at risk of chronic diseases showed no effect on physical activity or health outcomes [32].

These conflicting results highlight the complexity of using technology for health risk management and physical activity promotion. The effectiveness of such interventions may depend on various factors, including intervention design, user engagement, and individual health status [9]. Our findings, along with these recent studies, highlight the potential of digital health interventions in chronic disease management. They suggest that the effectiveness of such interventions may not solely depend on the specific technology used (eg, wearable devices vs built-in step counters), but also on how well these technologies are integrated into comprehensive health programs and how they engage users in multiple aspects of health management. Our findings align with recent research on digital health interventions for chronic disease management. The mobile health care program incorporated gamification elements, defined as "the use of game design elements in nongame contexts" [33]. This was evident in features like national step count rankings and health behavior scores, which were designed to engage users and encourage healthier behaviors. These gamified components, along with performance-based rewards, likely enhanced user engagement and motivation. Our results support the effectiveness of gamification strategies in health interventions, consistent with findings from a systematic review on type 2 diabetes self-management, which reported improvements in key health outcomes, such as glycated hemoglobin levels and BMI [34]. This underscores the importance of considering these elements when implementing mobile health interventions in public health settings and suggests that while mobile health technologies show promise, their impact may vary across different contexts and populations.

Potential Mechanisms and Explanations

The differential effects observed between wearable devices and built-in step counters in this study can be explained by several potential mechanisms. The marginal advantage of wearable devices in promoting walking activity may stem from their continuous and real-time data collection capabilities. These devices provide immediate feedback, which has been shown to enhance user engagement and motivation [15]. This instant feedback loop can heighten awareness of physical activity levels and encourage users to be more active throughout the day.

The noninferiority or similar health impact of built-in step counters compared to wearable trackers suggests that simplicity and accessibility play crucial roles in promoting health behaviors. This aligns with findings emphasizing that simple pedometers could be more effective than complex wearable trackers or smartphone apps in public health interventions. A meta-regression analysis of 57 RCTs revealed that body-worn trackers or smartphone apps were less effective than pedometers in interventions lasting less than 4 months (–834 steps/day; 95% CI –1542 to –126]) [35]. This underscores the potential value of straightforward, easily accessible tools in promoting physical activity within public health initiatives.

In this study, younger individuals, those with higher education, and iOS users showed higher rates of switching to built-in step counters. This may reflect these groups' higher understanding and adherence, potentially relating to the slightly better health risk factor reduction observed in the built-in step counter group. However, the possibility of selection or residual bias should be considered.

Multivariable logistic regression analysis confirmed an association between built-in step counter use and greater metabolic syndrome risk reduction, though no difference was found in health behaviors between the 2 groups, and the strength of the association was weak. Subgroup analyses revealed differences only in specific groups, limiting result

XSL•FO
RenderX

generalizability and suggesting the need for personalized interventions.

Limitations

This study provides valuable insights into the real-world effectiveness of wearable devices and built-in step counters within a national public health program. Using a large-scale nationwide dataset enhances statistical power and generalizability. The application of diverse statistical methodologies, including propensity analysis, strengthens the reliability of the results.

However, several limitations should be noted. As an observational study, it has inherent limitations in establishing causality. The small number of participants who switched to built-in step counters may limit statistical power in some analyses. Reliance on self-reported data and the inability to assess long-term effects are significant limitations. Additionally, the study design, where all participants initially used wearable devices for 12 weeks before some switched to built-in step counters, limits the ability to compare effects from the intervention's start. This design feature means that the effects of built-in step counters cannot be completely isolated from the potential carryover effects of initial wearable device use. Finally, our analysis may be limited by unmeasured confounding factors not available in our dataset.

Implications and Future Research Directions

Our findings provide important implications for the design and implementation of public mobile health care initiatives. While wearable devices may offer some advantages in promoting physical activity, built-in step counters can be equally effective, particularly in reducing metabolic syndrome risk among young adults.

These results support a more personalized approach to device recommendations in public health programs, aligning with the growing trend toward personalized digital health interventions [36,37]. Programs could offer options between wearable devices and built-in step counters based on individual preferences, age, and other relevant factors, potentially improving intervention adherence and outcomes.

The high rate of metabolic syndrome risk reduction observed (over 50% in both groups) suggests that well-designed technology-supported public health interventions can significantly impact population health, especially when integrated into comprehensive national health programs.

Future research should explore the mechanisms behind age-related differences in device effectiveness. Longitudinal studies with longer follow-up periods could provide insights into the long-term effects of various device types on health outcomes, addressing gaps identified in previous research [38].

Conclusions

This study compared the effectiveness of wearable devices and built-in step counters in a large-scale national mobile health care program. Both device types were effective in reducing metabolic syndrome risk, with the group switching to built-in step counters showing slightly better results (OR 1.20, 95% CI 1.05 - 1.36). Wearable devices showed a marginal advantage in promoting walking practice, but there was no significant difference in overall health behavior improvement.

These findings suggest the importance of providing device selection options based on individual characteristics and preferences in public mobile health care programs. The high rate of metabolic syndrome risk reduction observed in both groups (over 50%) demonstrates the potential of well-designed mobile health interventions.

Acknowledgments

This study was supported by a grant (20230813B0AA-00) from the Korea Health Promotion Institute. This work was supported by the National Research Foundation of Korea grant funded by the Korea government (MSIT; 2022R1F1A1063404).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standard mean differences before and after propensity score matching.
[DOCX File, 24 KB - mhealth_v13i1e64527_app1.docx]

References

1. Mottillo S, Filion KB, Genest J, et al. The metabolic syndrome and cardiovascular risk a systematic review and meta-analysis. J Am Coll Cardiol 2010;56(14):1113-1132. [doi: 10.1016/j.jacc.2010.05.034] [Medline: 20863953]
2. Gami AS, Witt BJ, Howard DE, et al. Metabolic syndrome and risk of incident cardiovascular events and death: a systematic review and meta-analysis of longitudinal studies. J Am Coll Cardiol 2007;49(4):403-414. [doi: 10.1016/j.jacc.2006.09.032] [Medline: 17258085]
3. Alberti K, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the

XSL•FO
RenderX

Study of Obesity. Circulation 2009;120(16):1640-1645. [doi: 10.1161/CIRCULATIONAHA.109.192644] [Medline: 19805654]
4. Cornier MA, Dabelea D, Hernandez TL, et al. The metabolic syndrome. Endocr Rev 2008 Dec;29(7):777-822. [doi: 10.1210/er.2008-0024] [Medline: 18971485]
5. Huckvale K, Prieto JT, Tilney M, Benghozi PJ, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015;13:214. [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
6. Kim HJ, Lee KH, Lee JH, Youk H, Lee HY. The effect of a mobile and wearable device intervention on increased physical activity to prevent metabolic syndrome: observational study. JMIR Mhealth Uhealth 2022 Feb 24;10(2):e34059. [doi: 10.2196/34059] [Medline: 35200145]
7. Lim SL, Tay MHJ, Ong KW, et al. Association between mobile health app engagement and weight loss and glycemic control in adults with type 2 diabetes and prediabetes (D'LITE Study): prospective cohort study. JMIR Diabetes 2022;7(3):e35039. [doi: 10.2196/35039] [Medline: 36178718]
8. Bults M, van Leersum CM, Olthuis TJJ, Bekhuis REM, den Ouden MEM. Mobile health apps for the control and self-management of type 2 diabetes mellitus: qualitative study on users' acceptability and acceptance. JMIR Diabetes 2023;8:e41076. [doi: 10.2196/41076] [Medline: 36692927]
9. Keshet A, Reicher L, Bar N, Segal E. Wearable and digital devices to monitor and treat metabolic diseases. Nat Metab 2023;5(4):563-571. [doi: 10.1038/s42255-023-00778-y] [Medline: 37100995]
10. Natalucci V, Marmondi F, Biraghi M, Bonato M. The effectiveness of wearable devices in non-communicable diseases to manage physical activity and nutrition: where we are? Nutrients 2023;15(4):913. [doi: 10.3390/nu15040913] [Medline: 36839271]
11. Shi C. Mobile phone messaging for facilitating self-management of long-term illnesses: summaries of nursing care-related systematic reviews from the Cochrane library. JBI Evidence Implementation 2013 Dec;11(4):344-345. [doi: 10.1111/1744-1609.12041]
12. Vorrink SNW, Kort HSM, Troosters T, Zanen P, Lammers JWJ. Efficacy of an mHealth intervention to stimulate physical activity in COPD patients after pulmonary rehabilitation. Eur Respir J 2016;48(4):1019-1029. [doi: 10.1183/13993003.00083-2016] [Medline: 27587557]
13. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016;13(2):e1001953. [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
14. Lee SK, Kim MG, Oh Y, et al. Effects of walking promotion using smart mobile activity meter on changes in metabolic health. Phys Act Health 2023;7(1):153-165. [doi: 10.5334/paah.241]
15. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
16. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012;9:21. [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
17. Oba T, Takano K, Katahira K, Kimura K. Use patterns of smartphone apps and wearable devices supporting physical activity and exercise: large-scale cross-sectional survey. JMIR Mhealth Uhealth 2023;11:e49148. [doi: 10.2196/49148] [Medline: 37997790]
18. Ministry of health and welfare. Korea health promotion development institute. guide to community integrated health promotion projects: physical activity. 2023 URL: https://www.khepi.or.kr/kps/publish/view?menuId=MENU00808&page_no=B2017008&pageNum=2&sield=&srch_text=2023&srch_cate=28&srch_type=ALL&str_dft_cd_list=&str_dft_cd_type_list=&board_idx=10913 [accessed 2025-02-10]
19. Eckel RH, Cornier MA. Update on the NCEP ATP-III emerging cardiometabolic risk factors. BMC Med 2014;12:115. [doi: 10.1186/1741-7015-12-115] [Medline: 25154373]
20. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. J Econ Surv 2008;22(1):31-72. [doi: 10.1111/j.1467-6419.2007.00527.x]
21. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
22. Jang M, Park JH, Kim GM, et al. Health provider's feedback on physical activity surveillance using wearable device-smartphone application for adults with metabolic syndrome; a 12-week randomized control study. Diabetes Metab Syndr Obes 2023;16:1357-1366. [doi: 10.2147/DMSO.S406065] [Medline: 37193576]
23. Rosa I, Lages M, Grilo C, Barros R, Guarino MP. mHealth applications to monitor lifestyle behaviors and circadian rhythm in clinical settings: current perspective and future directions. Front Public Health 2022;10:862065. [doi: 10.3389/fpubh.2022.862065] [Medline: 35923965]
24. Dehghan Ghahfarokhi A, Vosadi E, Barzegar H, Saatchian V. The effect of wearable and smartphone applications on physical activity, quality of life, and cardiovascular health outcomes in overweight/obese adults: a systematic review and meta-analysis of randomized controlled trials. Biol Res Nurs 2022;24(4):503-518. [doi: 10.1177/10998004221099556] [Medline: 35535558]

XSL•FO
RenderX

25. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418. [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
26. Kim JW, Ryu B, Cho S, et al. Impact of personal health records and wearables on health outcomes and patient response: three-arm randomized controlled trial. JMIR Mhealth Uhealth 2019;7(1):e12070. [doi: 10.2196/12070] [Medline: 30609978]
27. Turner-McGrievy GM, Wilcox S, Boutté A, et al. The dietary intervention to enhance tracking with mobile devices (diet mobile) study: a 6-month randomized weight loss trial. Obesity (Silver Spring) 2017 Aug;25(8):1336-1342. [doi: 10.1002/oby.21889] [Medline: 28600833]
28. Dawson JK, Ede A, Phan M, Sequeira A, Teng HL, Donlin A. Feasibility and acceptability of a mobile health exercise intervention for inactive adults: 3-arm randomized controlled pilot trial. JMIR Form Res 2024;8:e52428. [doi: 10.2196/52428] [Medline: 39120078]
29. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HBA1C in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 5;17(10):e224. [doi: 10.2196/jmir.4871] [Medline: 26441467]
30. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-754. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
31. Yang YP, Wang CJ, Wang JJ, et al. The effects of an activity promotion system on active living in overweight subjects with metabolic abnormalities. Obes Res Clin Pract 2017;11(6):718-727. [doi: 10.1016/j.orcp.2017.06.002] [Medline: 28729003]
32. Peacock OJ, Western MJ, Batterham AM, et al. Effect of novel technology-enabled multidimensional physical activity feedback in primary care patients at risk of chronic disease—the MIPACT study: a randomised controlled trial. Int J Behav Nutr Phys Act 2020 Aug 8;17(1):99. [doi: 10.1186/s12966-020-00998-5] [Medline: 32771018]
33. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining gamification. Presented at: The 15th International Academic MindTrek Conference: Envisioning Future Media Environments; Sep 28-30, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
34. Ossenbrink L, Haase T, Timpel P, et al. Effectiveness of digital health interventions containing game components for the self-management of type 2 diabetes: systematic review. JMIR Serious Games 2023 Jun 1;11:e44132. [doi: 10.2196/44132] [Medline: 37261900]
35. Chaudhry UAR, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 9;17(1):129. [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
36. Schroé H, Carlier S, Van Dyck D, De Backere F, Crombez G. Towards more personalized digital health interventions: a clustering method of action and coping plans to promote physical activity. BMC Public Health 2022;22(1):2325. [doi: 10.1186/s12889-022-14455-4] [Medline: 36510181]
37. Chén OY, Roberts B. Personalized health care and public health in the digital age. Front Digit Health 2021;3:595704. [doi: 10.3389/fdgth.2021.595704] [Medline: 34713084]
38. Martos-Cabrera MB, Velando-Soriano A, Pradas-Hernández L, et al. Smartphones and apps to control glycosylated hemoglobin (HbA1c) level in diabetes: a systematic review and meta-analysis. J Clin Med 2020 Mar 4;9(3):693. [doi: 10.3390/jcm9030693] [Medline: 32143452]

Abbreviations

KHPI: Korea Health Promotion Institute
OR: odds ratio
RCT: randomized controlled trial

Edited by L Buis; submitted 19.07.24; peer-reviewed by A Eisingerich, N Jeon; revised version received 14.12.24; accepted 23.01.25; published 25.02.25.

Please cite as:
Joung KI, An SH, Bang JS, Kim KJ
Comparative Effectiveness of Wearable Devices and Built-In Step Counters in Reducing Metabolic Syndrome Risk in South Korea: Population-Based Cohort Study
JMIR Mhealth Uhealth 2025;13:e64527
URL: https://mhealth.jmir.org/2025/1/e64527
doi:10.2196/64527

XSL•FO
RenderX

© Kyung-In Joung, Sook Hee An, Joon Seok Bang, Kwang Joon Kim. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessment of Environmental, Sociocultural, and Physiological Influences on Women's Toileting Decisions and Behaviors Using "Where I Go": Pilot Study of a Mobile App

Abigail R Smith[1], PhD; Elizabeth R Mueller[2], MD; Cora E Lewis[3], MD; Alayne Markland[4,5], DO; Caroline Smerdon[6], MPH; Ariana L Smith[7], MD; Siobhan Sutcliffe[8,9], MHS, MSc, PhD; Jean F Wyman[10], NP; Lisa Kane Low[11,12], PhD; Janis M Miller[11,12], PhD; The Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium[13]

[1]Division of Biostatistics and Informatics, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, 680 North Lake Shore Drive, Suite 1400, Chicago, IL, United States

[10]School of Nursing, University of Minnesota, Minneapolis, MN, United States

[11]Department of Health Behavior and Biological Sciences, School of Nursing, University of Michigan, Ann Arbor, MI, United States

[12]Department of Obstetrics & Gynecology, School of Medicine, University of Michigan, Ann Arbor, MI, United States

[13]See Acknowledgments

[2]Departments of Urology & Obstetrics/Gynecology, Loyola University Medical Center, Loyola University Chicago, Chicago, IL, United States

[3]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, United States

[4]Division of Gerontology, Geriatrics, and Palliative Care, Department of Medicine, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[5]Geriatric Research, Education, and Clinical Center, Birmingham Veterans Affairs Health Care System, Birmingham, AL, United States

[6]Arbor Research Collaborative for Health, Ann Arbor, MI, United States

[7]Division of Urology, Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

[8]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, United States

[9]Department of Obstetrics and Gynecology, Washington University School of Medicine, St. Louis, MO, United States

Corresponding Author:
Abigail R Smith, PhD
Division of Biostatistics and Informatics, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, 680 North Lake Shore Drive, Suite 1400, Chicago, IL, United States

Abstract

Background: Little is known about women's decisions around toileting for urination and how those decisions influence moment-to-moment behaviors to manage bladder needs. The new smartphone app "Where I Go" captures such nuanced and granular data in real-world environments.

Objective: This study aims to describe participant engagement with "Where I Go", variation in novel parameters collected, and readiness for the data collection tool's use in population-based studies.

Methods: "Where I Go" has three components: (1) real-time data, (2) short look-back periods (3 - 4 h), and (3) event location (GPS recorded at each interaction). The sample size was 44 women. Recording of real-time toileting events and responding to look-back questions was measured over 2 days of data collection. The participant's self-entered location descriptions and the automatic GPS recordings were compared.

Results: A total of 44 women with an average age of 44 (range 21-85) years interacted with the app. Real-time reporting of at least 1 toileting event per day was high (38/44, 86%, on day 1 and 40/44, 91%, on day 2) with a median of 5 (IQR 3-7 on day 1 and IQR 3-8 on day 2) toileting events recorded each day. Toileting most commonly occurred at home (85/140, 61%, on day 1 and 129/171, 75%, on day 2) due to a need to go (114/140, 66%, on day 1 and 153/171, 74%, on day 2). The most common reasons for delaying toileting were "work duties" (33/140, 21%, on day 1 and 21/171, 11%, on day 2) and "errands or traveling" (19/140, 12%, on day 1 and 19/171, 10%, on day 2). Response to at least 1 look-back notification was similarly high (41/44, 93%, on day 1 and 42/44, 95%, on day 2), with number of responses higher on average on day 2 compared with day 1 (mean on day 1=3.2, 95% CI 3.0-3.5; mean on day 2=4.3, 95% CI 3.9-4.7; P<.001). Median additional toileting events reported on the look-back survey were 1 (IQR 1-2) and 2 (IQR 1-2) on days 1 and 2, respectively. Overall concordance between self-reported location recording and GPS was 76% (188/247). Participants reported lower urge ratings when at home versus away when reporting real-time toileting (median rating 61, IQR 41-84 vs 72, IQR 56-98), and daily fluid intake showed a small to medium positive correlation with toileting frequency (day 1 r=0.3, day 2 r=0.24). Toileting frequency reported in "Where I Go" showed a small

XSL•FO
RenderX

positive correlation with the frequency item from the International Consultation on Incontinence Questionnaire (r=0.31 with day 1 toileting frequency and r=0.21 with day 2 toileting frequency).

Conclusions: "Where I Go" has potential to increase the understanding of factors that affect women's toileting decisions and long-term bladder health. We anticipate its use as a data collection tool in population-based studies.

International Registered Report Identifier (IRRID): RR2-10.2196/54046

(JMIR Mhealth Uhealth 2025;13:e56533) doi:10.2196/56533

KEYWORDS

ecological momentary assessment; time location factors; voiding diary; voiding behaviors; population studies; mobile application; app; bladder health; data collection tool; decision support

Introduction

Little is known about how women make decisions in real-time around toileting for urination, and how those decisions result in moment-to-moment specific behaviors to manage bladder needs in real-world environments, the surrounding environment (eg, man-made structures and natural settings) and the sociocultural world (eg, work, play, school, family, and other social settings) in the context of physiological factors (eg, age, fluid intake, emotion, and discomfort). Women with lower urinary tract symptoms (LUTS) report more frequently engaging in toileting behaviors such as hovering to urinate, delayed toileting or holding, and straining during toileting [1-5]. However, whether these examples of toileting behaviors contribute to LUTS, represent adaptations to the presence of symptoms, or a complex interaction between environmental, sociocultural, and physiological factors is unclear.

Variance within and across women in influences leading to real-time decisions about toileting is difficult to measure. Validated questionnaires such as the "Toileting Behaviors: Women's Elimination Behaviors Scale" [6] are limited by reliance on recalled behaviors and do not assess potential influences on toileting decisions (eg, bathroom safety and cleanliness, personal or work activities) and reasons for toileting (eg, no time to go later, having a bowel movement, wanting a break, or changing pad or tampon) that occur at the time of initiating or delaying toileting. Although bladder (or frequency or volume) diaries are considered to have higher validity than surveys, as they can capture toileting frequency and volume, urgency ratings, type and amount of fluids consumed, and urinary incontinence episodes as they occur, they still do not capture specific reasons for toileting or delaying toileting. They are also limited by participant burden (which can affect adherence), recall bias, and administrative costs.

Due to the difficulty in measuring real-world influences by diary, survey, or interview, scientific studies lack evidence for many hypothesized relationships between sociocultural and physical environment factors and onset or worsening of LUTS or deterioration of bladder health. To address this gap, the Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium developed a smartphone app called "Where I Go." The Consortium's goal was to produce a sophisticated data collection tool to explore potential relationships between factors influencing toileting decisions and bladder health. Given the increasing ubiquity of smartphone use in the United States (as of 2021, 85% of people in the United States own a smartphone and 11% own another type of cell phone) [7], PLUS designed Where I Go to provide dense, accurate data about moment-to-moment toileting decisions and bladder management behaviors in the real world. This diary-like app was designed to limit recalled information and maximize end users' interests in the use of the data collection tool. A key design element was location tracking at the time of the toileting event. The objective was to produce a simple, easy, and fun way for women to record episodes of "thinking about my bladder," and "now I peed," along with influencing contextual factors and accurate location, that is, "where" the event took place.

The first report on Where I Go described the app's design, build process, and usability or acceptability testing, including extensive supplementary material: multiple video clips of the build process and composition of the finished app as shown by a complete set of screenshots. Initial testing demonstrated above-average usability as measured by the System Usability Scale (SUS) [8-11] and "good" functionality as measured by the Mobile Application Rating Scale [12]. Where I Go also demonstrated high appeal to users (eg, color choices, intuitiveness, and delight in the cartoon features) as documented via qualitative data obtained postuse (Miller et al [11]). The purpose of this second report is to (1) demonstrate that Where I Go can capture nuanced and granular data in real-time or via short look-back periods, and (2) provide an initial description of the variation in novel parameters collected within Where I Go to support readiness for use in large population-based studies.

Methods

Ethical Considerations

After institutional review board approval, we used a cross-sectional study design with sampling from regions surrounding 4 PLUS research centers (Washington University in St. Louis [201912088], Loyola University Medical Center [LU212615], University of Pennsylvania [protocol number 833718], and University of Alabama at Birmingham [IRB-300003879]) to maximize variability across key novel Where I Go variables. We purposely excluded the University of Michigan PLUS Center where the build and prior informal iterative usability testing was performed. Women were instructed to enter real-time data in the Where I Go app whenever they were thinking about their bladder or using the toilet over 2 days. In addition, ecological momentary assessment

XSL•FO
RenderX

(EMA) data were collected 3 to 5 times per day to capture contextual factors around toileting including stress, mood, and fluid intake in addition to potentially missed toileting events. EMA data were collected via equally spaced prompts across participants' usual waking hours as reported in the app set-up process at first log-in.

Sample Selection and Size

As published in our first report [11], we estimated that a minimal sample size of 40 women would be needed to evaluate an a priori cut-point score of 74 (representing the 70th percentile of SUS scores or a grade of B-) [13] on the SUS. Community-dwelling women were recruited through flyers, advertisements in local community centers, contacting participants from previous studies, social media, and word-of-mouth. Women were eligible if they (1) were 18 years and older; (2) were fluent in written and spoken English; (3) were able to stand or walk independently without assistance by another person (use of cane or walker allowed); (4) owned a smartphone (Android or iOS) on which calls could be made; (5) had downloaded at least 1 app in the past 6 months; (6) were willing to respond to Where I Go prompts or texts and input data about toileting behaviors for 48 consecutive hours, during which they were required to leave their home at least once, but not to change time zones; and (7) agreed to 2 in-person visits at the study site. Women were excluded if they had a physical or mental condition that prevented them from completing written questionnaires and interactions with Where I Go; lived in an institutional setting (ie, skilled nursing, long-term care, or rehabilitation center); were pregnant or had a neurogenic or congenital bladder condition; and were unable to independently use the toilet.

After consent, each participant downloaded and installed Where I Go onto her smartphone, and then entered basic demographic information and typical "wake" time and pinned her home and any other regularly frequented locations. Formal data collection began with a check-in notification the morning following the download, which was scheduled automatically based on the inputted "wake time." Starting with this "wake time," each woman was instructed to continue using Where I Go for 48 consecutive hours. Participants were provided with a modest honorarium for their involvement.

Measurement Instruments

Where I Go was built at the University of Michigan's Center for Health Communications Research (UM invention registration OTT Ref No: 2019‐292) using the Cordova dual platform (The Apache Software Foundation; simultaneous build for use on mobile phones of either iOS or Android operating systems).

As reported in detail elsewhere [11], Where I Go provides each woman with an opportunity for (1) real-time reporting and (2) EMA reporting of the last few hours. Real-time reporting allows the woman to report planning for toileting throughout her day, The app starts with an option to tap the opening screen of "I'm thinking about my bladder." Once tapped, these screens follow: "How strong is your urge to pee?" (slider scale 0‐100), "When are you planning to pee?" (response options "now" or "later"), and if later, "Why aren't you going to pee now?" (response options list "don't have to pee that bad yet," "no place to pee," "work duties," etc). The screen then changes to an option for tapping "Now I peed." Once tapped, these screens follow: "When did you pee" (response items right now, 10, 20, or 30+ min ago), "Where did you pee?" (with tailored locations options, the last one being "new location"), "How strong was your urge to pee?" (slider scale), "Why did you pee" (eg, "had to go," "no time to go later," "wanted a break," and "I leaked"), and "If you had to wait before you peed, why did you wait?" ("work duties," "errands or traveling," "long line for toilet," "no delay," "peed when I needed to," etc). Response option of "other - fill in" was also offered when applicable. An "other" response offered the option for typing a comment into the fill-in box to indicate the reason for delaying toileting. At each of these interactions with the app, geospatial data were gathered automatically.

In addition to real-time reporting, Where I Go has built-in EMA, which the woman sees as pushed "Check-In" notifications. Push notifications for additional assessments offer users the opportunity to update information in the app if the recording of a toileting event was missed in real time. The screen series starts with automatic feedback to the woman about her prior real time reports, "These are the times you said you peed. Are there more you want to report?" and she can input additional toilet events and approximate time they occurred.

Within the EMA section of the app, additional screens with questions are "During this time period, I felt…" with slider scale response options anchored at each end with calm or stressed (0=calm to 10=stressed) and similarly not busy to busy.

EMA check-ins offered reporting during the look-back period of "Any pain with peeing," and "Any leakage with peeing." These screens were followed by "How much liquid did you drink?" with response options of selecting drink portion-size pictorials to add, and finally a screen "Did any of these drinks have these ingredients?" with response options "caffeine," "artificial sweeteners," "alcohol," and "none of these." Push notifications offered women the opportunity to record information on variables that are easier or more accurately reported through short look-back periods over the interval of time since the last notification rather than in real time. Geospatial data were also gathered automatically at each one of these EMAs.

The International Consultation on Incontinence Questionnaire (ICIQ) Overactive Bladder Module [14], a validated measure assessing overactive bladder and its impact on quality of life, was administered to participants at the end of the study. Responses to novel measures collected by Where I Go were compared with responses to these validated survey items on frequency, urgency, and incontinence.

Response Rate to Prompts on Key Factors of Interest

Given the novel nature of the app and the exploratory nature of this report, there were no a priori rules for evaluating participant response rate to prompts. To assess the accuracy of the location-pinning function (home or away from home), we compared participants' GPS longitude and latitude pinned location of "home," set at baseline, to the longitude and latitude

XSL•FO
RenderX

automatically captured when a participant interacted with the app to record a real-time toileting event at home. For participants with a missing or inaccurate home location pinned at baseline, their home location was replaced with the longitude and latitude of their first toileting event when they indicated that they were at home.

Statistical Analysis

To illustrate if plausible values were captured across the wide variety of types of data collected by Where I Go, descriptive statistics were calculated for the daily number of toileting events, average urinary urge ratings, number of locations reported, locations of each toileting event, and reasons for delaying toileting. In addition, descriptive statistics were calculated on the EMA component of the app (check-ins), comparing a number of locations, stress ratings, busy ratings, leakage episodes, toileting locations, fluid intake volume and ingredients, and management strategies across day 1 and day 2.

Descriptive statistics were calculated to portray the frequency or lack of use of real-time and EMA survey elements. The ability to record toileting events in real time was assessed by comparing real-time recordings to the number of missed toileting events documented in the EMA section. We evaluated beverage intake response rates to prompts as at least 1 fluid intake recorded in a day, making the logical assumption that all women would have consumed at least a minimal amount of fluid on each day. Finally, the relationships between quantities measured within the app and the ICIQ items were assessed.

Descriptive statistics were calculated separately for day 1 and day 2 and reported as frequency and percentage for categorical variables and median and IQR for continuous variables. For comparisons of continuous variables (eg, daily toileting frequency and daily intake volume), Pearson correlations were calculated. For comparisons of a continuous variable across levels of categorical variables (eg, urge ratings at home vs away), box plots were used for qualitative visual assessment. All analyses were completed in the SAS Enterprise Guide (version 9.3; SAS Institute).

Results

Participant Characteristics

A total of 44 women completed the baseline survey with 48 hours of interaction with Where I Go. Participants had a mean age of 44 (range 21 - 85) years, with 25/44 (56%) of participants identifying as White, 30/44 (68%) working one or more jobs, and 32/44 (73%) having a college degree or higher. Most participants had 2 or 3 toilets in the home (32/44, 73%), and most used an Apple phone (33/44, 75%) to interact with the app. Additional details on demographic information are provided in Table 1.

XSL•FO
RenderX

Table . Summary of demographic information collected directly within the Where I Go app in a pilot study of community participants (N=44).

Characteristic	Values
Age (years), mean (range)	44.3 (21-85)
Age (years), n (%)	
18 - 25	8 (18.2)
26 - 45	16 (36.4)
46 - 65	15 (34.1)
≥66	5 (11.4)
Employment (check all that apply), n (%)	
Homemaker	5 (11.4)
Student	8 (18.2)
Not working or unable to work	8 (18.2)
Working one or more jobs	30 (68.2)
Income (US $), n (%)	
<25,000	9 (20.5)
25,000-49,999	9 (20.5)
50,000-99,999	15 (34.1)
≥100,000	9 (20.5)
Missing	2 (4.5)
Sought care from health care provider for bladder, n (%)	
Yes	9 (20.5)
Missing	1 (2.3)
Education, n (%)	
High school diploma, General Educational Development Test, or less	0 (0)
Some college	12 (27.3)
Associate or Bachelor degree	18 (40.9)
Master degree	11 (25)
Professional or Doctorate degree	3 (6.8)
Hispanic or Latino, n (%)	8 (18.2)
Primary language spoken at home (check all that apply), n (%)	
English	43 (97.7)
Spanish	3 (6.8)
Other	2 (4.5)
Race (check all that apply), n (%)	
American Indian or Alaska Native	1 (2.3)
Asian	4 (9.1)
Black or African-American	12 (27.3)
Middle Eastern or North African	1 (2.3)
Native Hawaiian or other Pacific Islander	0 (0)
Other race or ethnicity or origin	1 (2.3)
White or Caucasian	25 (56.8)
Participant did not answer	3 (6.8)
Number of toilets in participants' home?	
1	12 (27.3)

XSL•FO
RenderX

Characteristic	Values
2	23 (52.3)
3	9 (20.5)
Phone type	
Android	10 (22.7)
Apple	33 (75)
Missing	1 (2.3)

Real-Time Assessment of Toileting

On both days 1 and 2, the median number of toileting events recorded in real time was 5 (IQR 3 - 7 on day 1 and IQR 3 - 8 on day 2) and most toileting events were recorded at the time of the event (112/140, 80%, on day 1 and 136/171, 80%, on day 2, Table 2). The most common toileting locations were home (85/140, 61%, on day 1 and 129/171, 75%, on day 2) and workplace (34/140, 24%, on day 1 and 25/171, 15%, on day 2). Most participants reported toileting due to a need to go (114/140, 66%, on day 1 and 153/171, 74%, on day 2), with work duties and errands or traveling as common reasons for delaying toileting as well as selection of "other" both when reporting toileting now and later (Table 2).

Real-time urge ratings were similar across day 1 (median rating 64, IQR 52 - 81) and day 2 (median rating 62, IQR 46 - 85) when participants reported toileting now (Table 2). Ratings were lower but similarly consistent across days when participants reported toileting later (day 1 median 54.5, IQR 34-74; day 2 median 55, IQR 38-71).

XSL•FO
RenderX

Table . Real-time assessment descriptive statistics.

Construct	Day 1	Day 2
Number of voids[a], median (IQR)	5 (3-7)	5 (3-8)
Number of locations, median (IQR)	1 (1-2)	2 (1-2)
Urge rating[b], median (IQR)	60.5 (44.5-79.5)	60 (41-81)
When planning to pee		
Now, n (%)	140 (66)	171 (69)
Urge rating[b], median (IQR)	64 (52-81)	62 (46-85)
When did you pee, n (%)		
Right now	112 (80)	136 (80)
10 minutes ago	11 (8)	20 (12)
20 minutes ago	2 (1)	5 (3)
30+ minutes ago	15 (11)	10 (6)
Where did you pee, n (%)		
Home	85 (61)	129 (75)
Workplace	34 (24)	25 (15)
School	6 (4)	1 (1)
Restaurant	1 (1)	8 (5)
Gym	5 (4)	2 (1)
Place of worship	1 (1)	1 (1)
Retail store	2 (1)	1 (1)
Gas station	1 (1)	0 (0)
Prefer not to say	0 (0)	0 (0)
Other	5 (4)	4 (2)
Why did you pee[c], n (%)		
Had to go	114 (66)	153 (74)
No time to go later	10 (6)	12 (6)
My usual time to go	14 (8)	10 (5)
Went with others	0 (0)	0 (0)
Bowel movement	14 (8)	19 (9)
Wanted a break	4 (2)	2 (1)
Changed pad or tampon	5 (3)	5 (2)
I leaked	0 (0)	3 (1)
Other	13 (7)	4 (2)
Delay reason[c], n (%)		
No delay. Peed when I needed	77 (50)	108 (59)
Work duties	33 (21)	21 (11)
Errands or traveling	19 (12)	19 (10)
Caring for family	9 (6)	10 (5)
Dirty bathroom	2 (1)	1 (1)
Bathroom not private or safe	0 (0)	1 (1)
Long line for toilet	4 (3)	3 (2)
Other	10 (6)	21 (11)

XSL•FO
RenderX

Construct	Day 1	Day 2
Later, n (%)	72 (34)	78 (31)
Urge rating[b], median (IQR)	54.5 (34-74)	55.0 (38-71)
Delay reason[c], n (%)		
Don’t have to pee that bad yet	11 (23)	7 (16)
No place to pee	6 (13)	5 (11)
Work duties	11 (23)	5 (11)
Errands or traveling	7 (15)	6 (14)
Caring for family	1 (2)	3 (7)
Dirty bathroom	0 (0)	1 (2)
Bathroom not private or safe	0 (0)	0 (0)
Long line for toilet	1 (2)	0 (0)
Other	10 (21)	17 (39)

[a]Median number of voids including real-time assessment and EMA survey was 6 (IQR 4-7) on day 1 and 6 (IQR 5-8) on day 2.

[b]Urge rating: 0=no urge, 100=strong urge.

[c]Multiple reasons per void possible.

EMA Assessment of Toileting

A total of 6 out of 44 participants (14%) on day 1 reported toileting via the EMA survey only and did not use the real-time function. Similarly, on day 2, four out of 44 participants (9%) used the EMA survey only. In contrast, 3 participants out of 44 (7%) on day 1 and 2 participants out of 44 (5%) on day 2 did not respond to any EMA notifications. The average number of responses to the EMA survey was higher on day 2 than on day 1 (day 1=3.2, day 2=4.3; $P<.001$). On the EMA survey, participants reported a median of 1 (IQR 1 - 2) additional toileting event on day 1 and 2 (IQR 1 - 2) additional toileting events on day 2 (Table 3).

XSL•FO
RenderX

Table . Ecological momentary assessment (EMA) descriptive statistics.

Construct	Day 1	Day 2
Voiding-related		
Number of voids[a] added, median (IQR)	1 (1-3)	2.5 (1-4)
Number of locations, median (IQR)	1 (1-2)	2 (1-2)
Pain while peeing, n (%)	1 (1)	1 (1)
Leakage episodes, median (IQR)	0 (0-0)	0 (0-0)
Where did you pee, n (%)		
Home	45 (56)	73 (68)
Workplace	28 (35)	15 (14)
School	1 (1)	1 (1)
Restaurant	2 (3)	3 (3)
Gym	0 (0)	1 (1)
Place of worship	1 (1)	1 (1)
Retail store	1 (1)	2 (2)
Gas station	0 (0)	2 (2)
Prefer not to say	0 (0)	1 (1)
Other	2 (3)	8 (8)
Intake-related		
Fluid intakes added, median (IQR)	3 (2-4)	4 (3-5)
Approximate daily intake volume (oz), median (IQR)	32 (20-40)	41 (20-60)
Intake ingredient[b], n (%)		
Caffeine	38 (30)	45 (25)
Artificial sweeteners	8 (6)	15 (8)
Alcohol	1 (1)	5 (3)
None of these	78 (62)	115 (64)
Environment-related		
Stress rating[c], median (IQR)	1 (0-3)	2 (1-3)
Busy rating[d], median (IQR)	4 (1-7)	4 (1-6)
Highly focused, n (%)	103 (67)	130 (65)
Management strategies, n (%)		
Peed at certain times of the day because of schedule	32 (18)	21 (9)
Peed before certain activities	26 (15)	32 (14)
Knew I couldn’t go, so I wore a pad or diaper	0 (0)	1 (0.5)
Limited drinking	8 (5)	6 (3)
Had the urge but ignored it	11 (6)	17 (7)
No strategy; I just peed when I needed to	99 (55)	151 (66)
I used some other strategy	2 (1)	0 (0)

[a]Median number of voids including real-time assessment and EMA survey was 6 (IQR 4-7) on day 1 and 6 (IQR 5-8) on day 2.

[b]Drink-level, multiple ingredients per participants per day possible; multiple ingredients per intake possible.

[c]Stressing rating: 0=calm, 10=stress.

[d]Busy rating: 0=not busy, 10=busy.

XSL•FO
RenderX

Location Assessment

A total of 15 out of 44 participants (34%) had a missing or incorrect pinned home location at baseline. Two participants out of 44 (5%) did not have a pinned home location at baseline and 13 out of 44 (30%) had a pinned home location that was set incorrectly. Of all toileting events that had location automatically pinned by the app as "Home," 97% (103/106) of those events were self-reported as having occurred at home. However, for toileting events where participants' location was automatically pinned as "Away," only 60% (85/141) were self-reported by those participants as being away from home, potentially due to an incorrectly pinned home location.

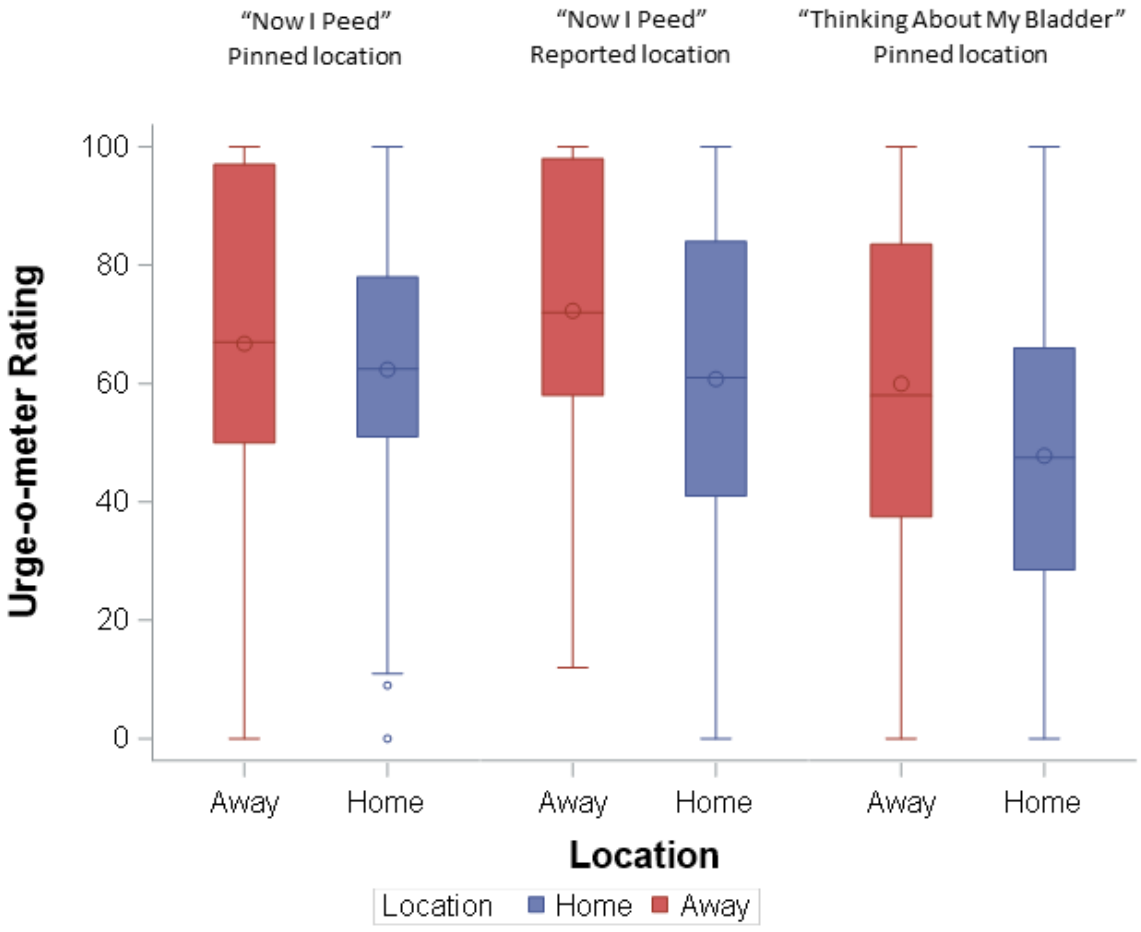
Look-Back Factors Potentially Influencing Toileting Decisions

Median number of beverage servings put into the EMA beverage section ranged from 3 to 4 per day, and the approximate median daily intake volume was 32 (IQR20-40) to 41 (IQR 20-60) ounces. While most participants did not report using strategies to manage their toileting events, those that did most reported toileting at certain times of the day due to their schedule (32/178, 18%, on day 1 and 21/228, 9%, on day 2) and before certain activities (26/178, 15%, on day 1 and 32/228, 14%, on day 2). Most participants reported low stress generally (median scores 1 - 2 on a scale from 0=calm to 10=stressed) during the look-back periods.

Within App Comparisons

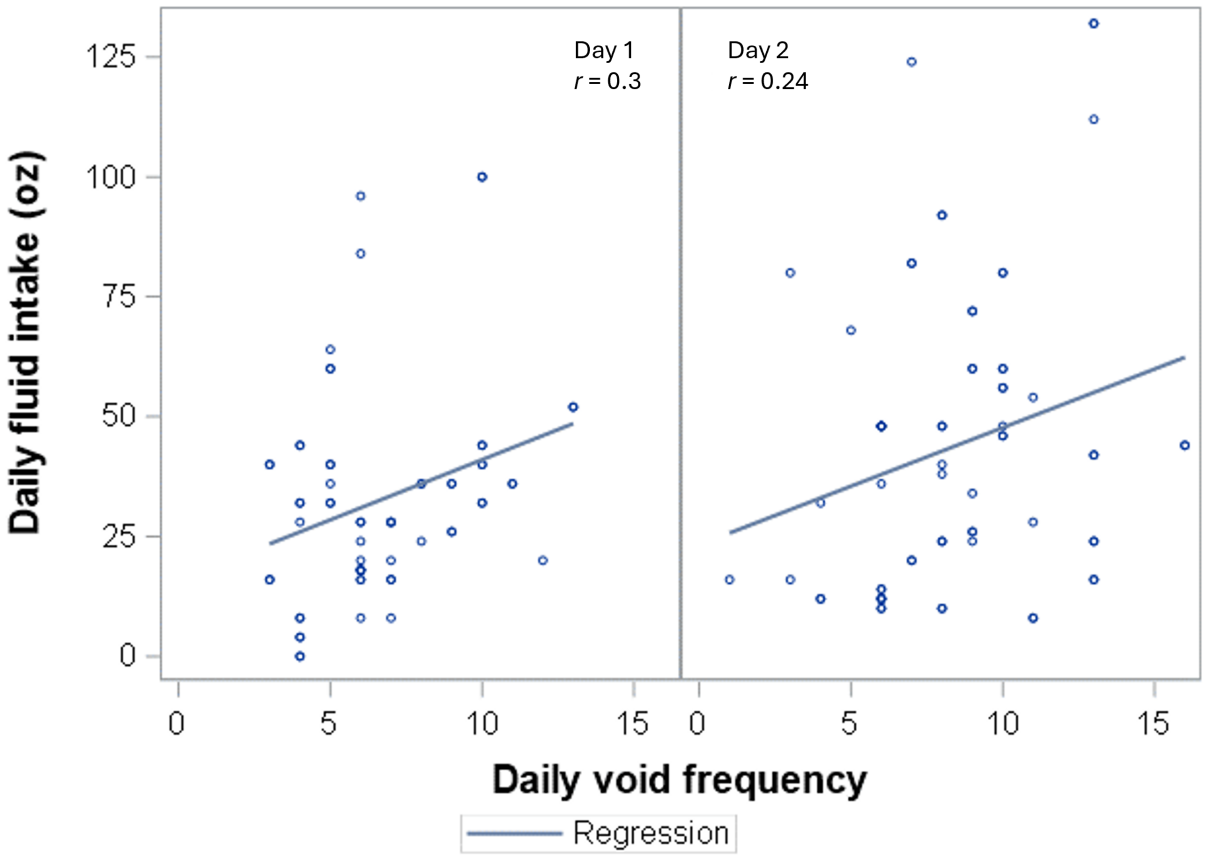
When comparing urge ratings between "Home" and "Away," urge ratings tended to be lower on average when participants were at home (Figure 1). This relationship was more pronounced for the "Now I Peed" survey using participants' reported location (median rating 72, IQR 56-98, when away vs 61, IQR 41-84, when at home) and the "Thinking About My Bladder" survey using participants' pinned location (median rating 58, IQR 41.5-74.5, when away vs 48, IQR 24-67, when at home). Daily fluid intake had a small to medium positive correlation with daily toileting frequency (day 1 r=0.3, day 2 r=0.24, Figure 2). In addition, participants typically reported higher urge ratings when they reported they would "Pee Now" versus "Pee Later," (median 64, IQR 52-81 vs 54.5, IQR 34-74, on day 1; median 62, IQR 46-85 vs 55, IQR 38-71, on day 2; Figure 3).

Figure 1. Comparisons between urge-o-meter ratings at home versus away from "Now I Peed" and "Thinking About my Bladder" surveys (The center line of the boxplot represents the median, while the center circle represents the mean. The outer edges of the boxplot represent the 25th-75th percentiles. Whiskers extend to 1.5 times the IQR. Outliers [values with distance>1.5*IQR] are represented by circles outside of the whiskers).



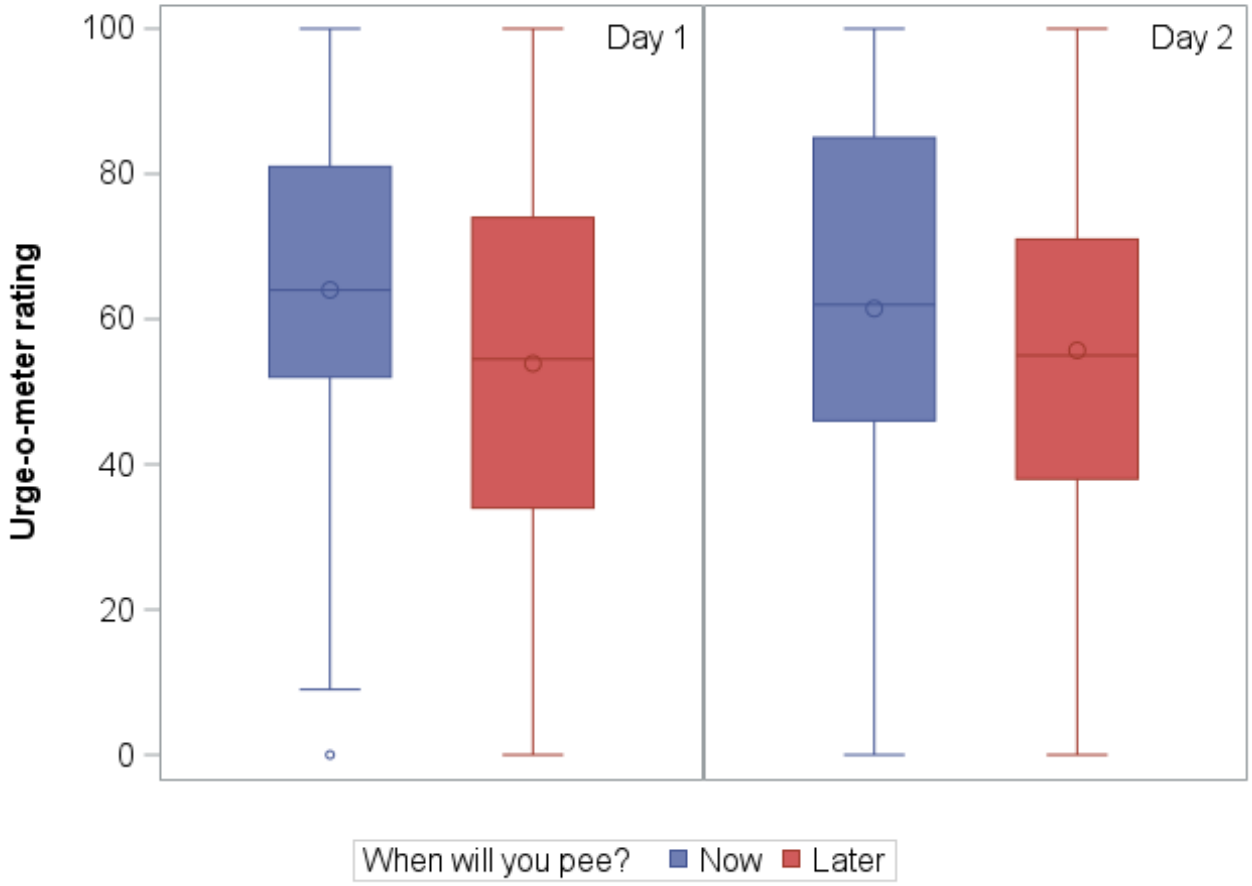
XSL•FO
RenderX

Figure 2. Comparison between daily fluid intake and daily void frequency.



XSL•FO
RenderX

Figure 3. Distribution of urge-o-meter ratings: pee now versus pee later (The center line of the boxplot represents the median, while the center circle represents the mean. The outer edges of the boxplot represent the 25th-75th percentiles. Whiskers represent data points outside the intra-quartile range but are not considered outliers. Outliers [values>1.5*IQR] are represented by circles outside of the whiskers).



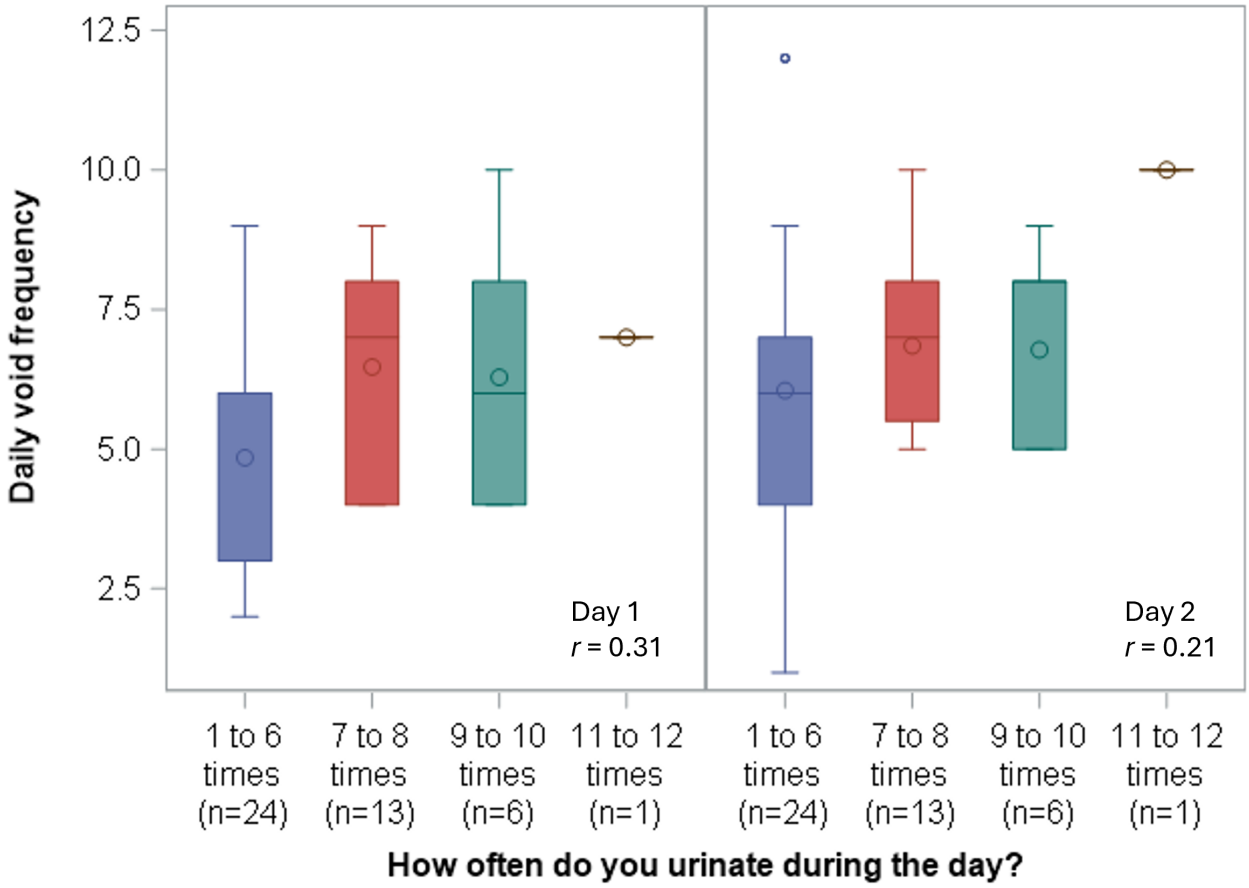
Relationships With ICIQ Responses

Participants' daily toileting frequency reported in Where I Go showed a small positive correlation with the daily urinary frequency item on the ICIQ (r=0.31 with day 1 toileting frequency and r=0.21 with day 2 toileting frequency, Figure 4).

Most participants reported urination frequencies of 1 to 6 times and 7 or 8 times on the ICIQ frequency item (37/44, 84%); therefore, the relationship between toileting events recorded in Where I Go and ICIQ responses cannot be assessed precisely for those reporting urination frequency greater than 8 times per day on the ICIQ.

XSL•FO
RenderX

Figure 4. Distribution of daily voids (app) by International Consultation on Incontinence Questionnaire frequency category. (The center line of the boxplot represents the median, while the center circle represents the mean. The outer edges of the boxplot represent the 25th-75th percentiles. Whiskers represent data points outside the intra-quartile range but are not considered outliers. Outliers [values>1.5*IQR] are represented by circles outside of the whiskers).



Discussion

In this paper, we report on the 2-day experience of 44 community women with the Where I Go app. Toileting behaviors, decisions around toileting, toileting location, and beverage consumption were recorded using real-time input by the user, EMA, and geospatial data [11]. This data collection tool's outputs, as described above, illustrate the wide range of data that Where I Go collects and the resulting plausible values, relationships, and preliminary variability of those parameters. The agreement with an extant survey instrument in the field is another positive indicator of its value. Strengths of this data collection tool include its continuous surveillance, user prompts, dynamic visualizations (eg, facial expressions to help users' rate their level of urinary urge before toileting), ecofriendly nature relative to paper diaries, and convenience.

Where I Go uses both real-time and EMA functions with high compliance, minimizing recall concerns typically associated with surveys or questionnaires. In addition, Where I Go has the capacity to capture contextual parameters and adaptive behaviors such as limiting fluid intake, toileting on a schedule, preemptive toileting, and use of absorbent products, all of which are missing from current tools. These findings support the utility of Where I Go for conducting large-scale population research. Users must own a smartphone and be able to download a mobile app that is designed to be used on either Android or iOS smartphones. The ability to measure location, as demonstrated by Where I Go, is a key advantage over paper diaries. The agreement between participant self-report and pinned GPS location was high overall (188/247, 76%). However, despite the seeming ubiquity of smartphones, this requirement may limit the inclusion of certain populations, particularly marginalized and minoritized populations, who may have more limited access to both smartphones as well as the data required to operate the app.

Where I Go offers opportunities to explore the impact of social and environmental determinants of toileting behavior from the individual to the societal level. Reasons for delaying voiding captured by Where I Go could reveal underlying factors impacting LUTS, from personal or work responsibilities, issues of privacy, cleanliness, and safety, to the lack of access to public toilets [15-18]. Delaying urination or holding urine to the point of discomfort or urgency may be a risk factor for bladder symptoms. Prior work has shown that women who limited their use of workplace toilets had higher prevalences of urinary urgency, overactive bladder symptoms, and urinary leakage, which could be more robustly investigated through large-scale use of Where I Go [19].

This study allowed us to identify areas of future improvement in Where I Go. The most notable future improvement expected is in the initial home location pinning. Discordance between app-recorded and self-reported location was most common when

XSL•FO
RenderX

participants reported toileting events at home, but the app GPS did not match their home location. Whether this was due to technological issues (eg, poor service at a participants' home location) or user error is unknown and should be investigated further. The ability of smartphone GPS systems to identify precise locations is improving over time, however, which should positively impact the utility of Where I Go. Other improvements include the exploration of reasons typed into the app when participants selected "other" with the goal of honing the list of given response options. Currently, Where I Go does have an automatic function that brings in the "other" option's typed-in response options when the participant is next queried. Anecdotally, women appreciated this feature.

Results from this study support the implementation of Where I Go as an important data collection tool for population-based studies investigating factors affecting bladder health; however, the small sample did not allow us to demonstrate its use in all subgroups, including individuals with lower education and higher income. Future work should include the use of this app among a more diverse and larger population, with adequate power for testing hypothesized relationships. In addition, data collection occurred over a 2-day period; however, the optimal data collection time frame is still to be determined. The app as currently built does offer the ability to extend data collection beyond 2 days. Finally, average daily fluid intake was lower than expected, potentially due to underreporting.

More research should also be directed towards comparing data obtained by Where I Go with validated patient-reported outcome measures (PROMs) on bladder health and LUTS, with an emphasis on environmental factors that may limit toileting [20]. When comparing traditional bladder diaries to PROMs, other studies have found that bladder diary data are comparable to PROMs for women with LUTS [20]. However, this small sample with low levels of LUTS demonstrated only weak associations with items from the ICIQ.

Where I Go measures toileting and associated symptoms across a fluidity of environments. The breadth of environments in a woman's day of managing her bladder function is not commonly measured in traditional paper or electronic diaries. Real-time data recording supplemented by EMA opportunities each day overcomes recall concerns inherent in most extant survey instruments and offers more nuanced data than standard surveys or bladder diaries.

We conclude that the "Where I Go" app is an instrument that can help us understand a fuller spectrum of potential factors that affect women's bladder health. The results reported support this novel data collection tool's use in a wider study; specifically in a general population of women with and without LUTS.

Acknowledgments

The authors wish to thank the "Where I Go" build team at the Center for Health Communications Research (CHCR), University of Michigan. This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases at the National Institutes of Health (NIH) by cooperative agreements (grants U24DK106786, U01 DK106853, U01 DK106858, U01 DK106898, U01 DK106893, U01 DK106827, U01 DK106908, U01 DK106892, and U01DK126045). Additional funding was from the National Institute on Aging, NIH Office on Research in Women's Health, the NIH Office of Behavioral and Social Science, and by the National Cancer Institutes of Health under Award Number P30CA046592 by the use of the following Rogel Cancer Center Health Communications Shared Resource: Center for Health Communications Research. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors attest there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this paper.

We acknowledge the contributions of the Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium Research Centers and Investigators: Multi-principal investigators Linda Brubaker and Colleen Fitzgerald; investigators Marian Acevedo-Alvarez, Cecilia T Hardacker, and Jeni Hebert-Beirne (Loyola University Chicago, Maywood, IL, U01DK106898). Multi-principal investigators James W Griffith, Kimberly Sue Kenton, and Melissa Simon; investigators Oluwateniola Brown, Julia Geynisman-Tan, and Margaret Mueller (Northwestern University, Chicago, IL, U01DK126045). Multi-principal investigators Alayne D Markland and Camille P Vaughan; investigators Tamera Coyne-Beasley, Kathryn L Burgio, Cora E Lewis, Gerald McGwin, and Beverly Rosa Williams (University of Alabama at Birmingham, Birmingham, AL, U01DK106858). Principal investigators Emily S Lukacz and D Yvette LaCoursiere; investigators Sheila Gahagan and Jesse Nodora (University of California San Diego, La Jolla, CA, U01DK106827). Principal investigator Lisa Kane Low; investigators Janis M Miller and Abigail R Smith (University of Michigan - Ann Arbor, MI, U01DK106893). Multi-principal investigators Gerald McGwin and Kyle D Rudser; investigators Sonya S Brady, Cynthia S Fok, Bernard L Harlow, Peter Scal, and Todd Rockwood (Scientific and Data Coordinating Center, University of Minnesota, Minneapolis, MN, U24DK106786). Multi-principal investigators Diane K Newman and Ariana L Smith; investigators Amanda Berry, Andrea Bilger, Terri H Lipman, Heather Klusaritz, Ann E Stapleton, Jean F Wyman, and Kristine Talley (University of Pennsylvania, PA, U01DK106892). Principal investigator Siobhan Sutcliffe; investigators Aimee S James, Jerry L Lowder, and Melanie R Meister (Washington University in St. Louis - Saint Louis, MO, U01DK106853). Principal investigator Leslie M Rickey; investigators Deepa R Camenga and Shayna D Cunningham (Yale University - New Haven, CT, U01DK106908). Steering committee chair Linda Brubaker ((UCSD, San Diego, January 2021-). NIH project scientist Jenna Norton (NIH Program Office: National Institute of Diabetes and Digestive and Kidney Diseases, Division of Kidney, Urologic, and Hematologic Diseases, Bethesda, MD).

XSL•FO
RenderX

Conflicts of Interest

ERM receives UpToDate author royalties and ALS is the Associate Editor of *Neurourology and Urodynamics* and on the Board of Directors for the Society for Urodynamics and Female Pelvic Medicine and Urogenital Reconstruction.

References

1. Willis-Gray MG, Wu JM, Sripad A, Newman D. Toileting behaviors in women presenting to a urogynecology clinic. UNJ 2017;37(5):251. [doi: 10.7257/1053-816X.2017.37.5.251]
2. Palmer B, Kropp B. Urologic evaluation and management of pediatric kidney transplant patients. Urol Clin North Am 2018 Nov;45(4):561-569. [doi: 10.1016/j.ucl.2018.06.004] [Medline: 30316311]
3. Newman DK, Burgio KL, Cain C, et al. Toileting behaviors and lower urinary tract symptoms: a cross-sectional study of diverse women in the United States. Int J Nurs Stud Adv 2021 Nov;3:100052. [doi: 10.1016/j.ijnsa.2021.100052] [Medline: 35498154]
4. Kowalik CR, Casteleijn FM, van Eijndhoven HWF, Zwolsman SE, Roovers JPWR. Results of an innovative bulking agent in patients with stress urinary incontinence who are not optimal candidates for mid-urethral sling surgery. Neurourol Urodyn 2018 Jan;37(1):339-345. [doi: 10.1002/nau.23299] [Medline: 28452427]
5. Daily AM, Kowalik CG, Delpe SD, Kaufman MR, Dmochowski RR, Reynolds WS. Women with overactive bladder exhibit more unhealthy toileting behaviors: a cross-sectional study. Urology 2019 Dec;134:97-102. [doi: 10.1016/j.urology.2019.08.038] [Medline: 31499079]
6. Wang K, Palmer MH. Development and validation of an instrument to assess women's toileting behavior related to urinary elimination: preliminary results. Nurs Res 2011;60(3):158-164. [doi: 10.1097/NNR.0b013e3182159cc7] [Medline: 21522032]
7. Mobile fact sheet. Pew Research Center. URL: Https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2025-01-28]
8. Brooke J. System usability scale (SUS): a quick-and-dirty method of system evaluation user information. : Digital Equipment Co Ltd; 1986.
9. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
10. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123.
11. Miller J, Wyman J, An L, et al. Design of a tool capable of assessing environmental sociocultural physical factors influencing women's decisions on when and where to toilet within real-world settings: protocol for the build and usability testing of a mobile app for use by community-dwelling women. JMIR Res Protoc 2024 Sep 18;13:e54046.
12. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015;3(1):e27. [doi: 10.2196/mhealth.3422]
13. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices: Measuring Usability LLC; 2011.
14. Jackson S, Donovan J, Brookes S, Eckford S, Swithinbank L, Abrams P. The Bristol female lower urinary tract symptoms questionnaire: development and psychometric testing. Br J Urol 1996 Jun;77(6):805-812. [doi: 10.1046/j.1464-410x.1996.00186.x] [Medline: 8705212]
15. Yuko E. Where did all the public bathrooms go? Bloomberg. URL: https://www.bloomberg.com/news/features/2021-11-05/why-american-cities-lost-their-public-bathrooms [accessed 2025-01-28]
16. Smoyer AB, Pittman A, Borzillo P. Humans peeing: justice-involved women's access to toilets in public spaces. PLoS One 2023;18(3):e0282917. [doi: 10.1371/journal.pone.0282917] [Medline: 36897925]
17. Maroko AR, Hopper K, Gruer C, Jaffe M, Zhen E, Sommer M. Public restrooms, periods, and people experiencing homelessness: an assessment of public toilets in high needs areas of Manhattan, New York. PLoS One 2021;16(6):e0252946. [doi: 10.1371/journal.pone.0252946] [Medline: 34161351]
18. Sommer M, Gruer C, Smith RC, Maroko A. Menstruation and homelessness: challenges faced living in shelters and on the street in New York City. H Place 2020 Nov;66:102431. [doi: 10.1016/j.healthplace.2020.102431] [Medline: 32987242]
19. Reynolds WS, Kowalik C, Delpe SD, Kaufman M, Fowke JH, Dmochowski R. Toileting behaviors and bladder symptoms in women who limit restroom use at work: a cross-sectional study. J Urol 2019 Nov;202(5):1008-1014. [doi: 10.1097/JU.0000000000000315] [Medline: 31059664]
20. Flynn KE, Wiseman JB, Helmuth ME, et al. Comparing clinical bladder diaries and recalled patient reports for measuring lower urinary tract symptoms in the symptoms of Lower Urinary Tract Dysfunction Research Network (LURN). Neurourol Urodyn 2022 Nov;41(8):1711-1721. [doi: 10.1002/nau.25030] [Medline: 36066068]

Abbreviations

CHCR: Center for Health Communications Research
EMA: ecological momentary assessment
ICIQ: International Consultation on Incontinence Questionnaire

XSL•FO
RenderX

LUTS: lower urinary tract symptoms
PLUS: Prevention of Lower Urinary Tract Symptoms
PROM: patient-reported outcome measure
SUS: System Usability Scale

Edited by L Buis; submitted 19.01.24; peer-reviewed by S Nyenhuis, S Bennis; revised version received 15.10.24; accepted 10.11.24; published 12.02.25.

Please cite as:
Smith AR, Mueller ER, Lewis CE, Markland A, Smerdon C, Smith AL, Sutcliffe S, Wyman JF, Low LK, Miller JM, The Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium
Assessment of Environmental, Sociocultural, and Physiological Influences on Women's Toileting Decisions and Behaviors Using "Where I Go": Pilot Study of a Mobile App
JMIR Mhealth Uhealth 2025;13:e56533
URL: https://mhealth.jmir.org/2025/1/e56533
doi:10.2196/56533

© Abigail R Smith, Elizabeth R Mueller, Cora E Lewis, Alayne Markland, Caroline Smerdon, Ariana L Smith, Siobhan Sutcliffe, Jean F Wyman, Lisa Kane Low, Janis M Miller, The Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

An Explanation Interface for Healthy Food Recommendations in a Real-Life Workplace Deployment: User-Centered Design Study

Robin De Croon[1], PhD; Daniela Segovia-Lizano[2], MSc; Paul Finglas[2], PhD; Vero Vanden Abeele[1], PhD; Katrien Verbert[1], PhD

[1]Department of Computer Science, KU Leuven, Celestijnenlaan 200A, Leuven, Belgium

[2]Food & Nutrition National Biosciences Research Infrastructure, Quadram Bioscience Institute, Norwich, United Kingdom

Corresponding Author:
Robin De Croon, PhD
Department of Computer Science, KU Leuven, Celestijnenlaan 200A, Leuven, Belgium

Abstract

Background: Despite widespread awareness of healthy eating principles, many individuals struggle to translate this knowledge into consistent, sustainable dietary change. Food recommender systems, increasingly used in various settings, offer the potential for personalized guidance and behavior change support. However, traditional approaches may prioritize user preferences or popularity metrics without sufficiently considering long-term nutritional goals. This can inadvertently reinforce unhealthy eating patterns. Emerging research suggests that incorporating explanations into recommender systems can increase transparency, promote informed decision-making, and potentially influence food choices. Yet, the effectiveness of explanations in promoting healthy choices within complex, real-world food environments remain largely unexplored.

Objective: This study aims to investigate the design, implementation, and preliminary evaluation of a food recommender system that integrates explanations in a real-world food catering application. We seek to understand how such a system can promote healthy choices while addressing the inherent tensions between user control, meal variety, and the need for nutritionally sound recommendations. Specifically, our objectives are to (1) identify and prioritize key design considerations for food recommenders that balance personalization, nutritional guidance, and user experience; and (2) conduct a proof-of-principle study in a real-life setting to assess the system's effect on user understanding, trust, and potentially on dietary choices.

Methods: An iterative, user-centered design process guided the development and refinement of the system across 4 phases: (Phase 0) an exploratory qualitative study (N=26) to understand stakeholder needs and initial system impressions, (Phases 1 and 2) rapid prototyping in real-life deployments (N=45 and N=16, respectively) to iteratively improve usability and features, and (Phase 3) a proof-of-principle study with employees (N=136) to evaluate a set of design goals. We collected a mix of data, including usage logs, pre- and post-study questionnaires, in-app feedback, and a pre- and post–Food Frequency Questionnaire to establish nutritional profiles.

Results: Although we experienced a high drop-out (77% after 7 weeks), motivated and remaining participants valued personalization features, particularly the ability to configure allergies and lifestyle preferences. Explanations increased understanding of recommendations and created a sense of control, even when preferences and healthy options did not fully align. However, a mismatch persisted between individual preferences and nutritionally optimal recommendations. This highlights the design challenge of balancing user control, meal variety, and the promotion of healthy eating.

Conclusions: Integrating explanations into personalized food recommender systems might be promising for supporting healthier food choices and creating a more informed understanding of dietary patterns. Our findings could highlight the importance of balancing user control with both the practical limitations of food service settings and the need for nutritionally sound recommendations. While fully resolving the tension between immediate preferences and long-term health goals is an ongoing challenge, explanations can play a crucial role in promoting more conscious decision-making.

(JMIR Mhealth Uhealth 2025;13:e51271) doi:10.2196/51271

KEYWORDS

food recommender systems; personalized nutrition; healthy eating; human-computer interaction; real-life deployment; food catering; meal recommendations; nutritional profile; transparency

XSL•FO
RenderX

Introduction

Despite growing awareness of healthy eating, the prevalence of obesity tripled between 1975 and 2016 [1]. This disconnect, driven largely by unhealthy diets and malnutrition [2], suggests that traditional information campaigns often fail to translate knowledge into sustained behavioral change [3]. Since workplace food choices occur in controlled retail environments like catering and restaurants, foodservice providers have a unique opportunity to guide healthier decisions. However, current offerings in these settings often prioritize convenience over health. Integrating personalized dietary guidance with foodservice providers could be a significant step in promoting healthier eating habits within [4] and beyond the workplace [5].

Recent research emphasizes the need for personalized approaches to promote healthy eating [6-8]. Food recommender systems offer the potential to both empower people to monitor and improve their food-intake through technology-assisted, personalized healthy recommendations [9,10]. These systems can personalize suggestions, engage users to change their consumption behavior [11], and provide actionable insights based on individual eating patterns [12,13]. While studies indicate greater adherence to personalized dietary advice compared to generic guidelines [14], many people still face challenges in achieving their health goals.

Food recommender systems are software applications that leverage user data and algorithms to provide personalized suggestions for meals, recipes, or food items. Studies typically focus on suggesting recipes [15,16], meal plans [17], menu items [18], and more recently personalized nutrition [19]. However, as indicated by Trattner and Elsweiler [20], food recommender systems are relatively under researched compared to other recommender domains. There are many reasons that make food recommendations challenging [20,21], not only in terms of encouraging healthy behavior, but also in predicting what people would like to eat. The challenges are multifaceted, culturally determined, and context dependent. Users may have complex, constrained needs, such as allergies or life-style preferences, such as the desire to eat only vegan or gluten-free food. In such cases, standard recommender techniques often fall short in these contexts [22,23]. Data sources in real-life settings are often limited, for example, food catering services usually have restricted menu options. Other challenges include retrieving and selecting user-nutrition profiles, labor intensive food tracking, missing ingredients, and atypical for recommender systems, the most popular meals are not always the healthiest options [24]. For example, El Majjodi et al [25] found that commonly used, preference-based and popularity-driven approach could lead to a decrease in the healthiness of chosen recipes. To address these challenges, explanations within food recommender systems are crucial. Such explanations would clarify why certain recommendations are made, considering the user's unique profile and health objectives.

Explanation interfaces have demonstrated positive impacts on user acceptance and trust in recommender systems [26,27]. A growing body of research highlights the benefits of explanations in enabling transparency and informed decision-making [28-30]. This focus on explanations is increasingly being applied in high-risk domains, such as health care [31] and job search [32], to support users in navigating complex recommendations. However, as recently highlighted by Musto et al [24], the use of personalized explanations in real-world food recommender systems remains underexplored. While some systems offer healthier alternatives [33,34] or incorporate user preferences [35], considerations around interface design and their impact on user choices are crucial [36]. In addition, few food recommender systems comprehensively integrate nutritional profiles [37], especially in real-life settings.

Although we experienced a high drop-out of 77% after 7 weeks, this paper makes an initial contribution by longitudinally deploying and evaluating an explanation interface within a food catering application. Research on user acceptance and trust of explanations in the food domain is limited [21], highlighting the need for studies that explore their potential to promote healthier eating habits. While early work [38] linked explanations to user preferences, or mainly evaluated system usability [4], our approach advances the field by tailoring explanations to individual nutritional profiles. Unlike studies requiring manual food tracking [39], our approach integrates explanations directly into the recommendation process.

This study aims to investigate the design, implementation, and evaluation of a food recommender system that integrates explanations in a real-world workplace setting. Our goal is to gain a deeper understanding of the user experience with explanations, specifically how they influence comprehension, trust, and ultimately, dietary decision-making. In addition, we seek to identify key design challenges and opportunities for balancing personalization with the promotion of healthy eating in the complex context of a real-life food service environment.

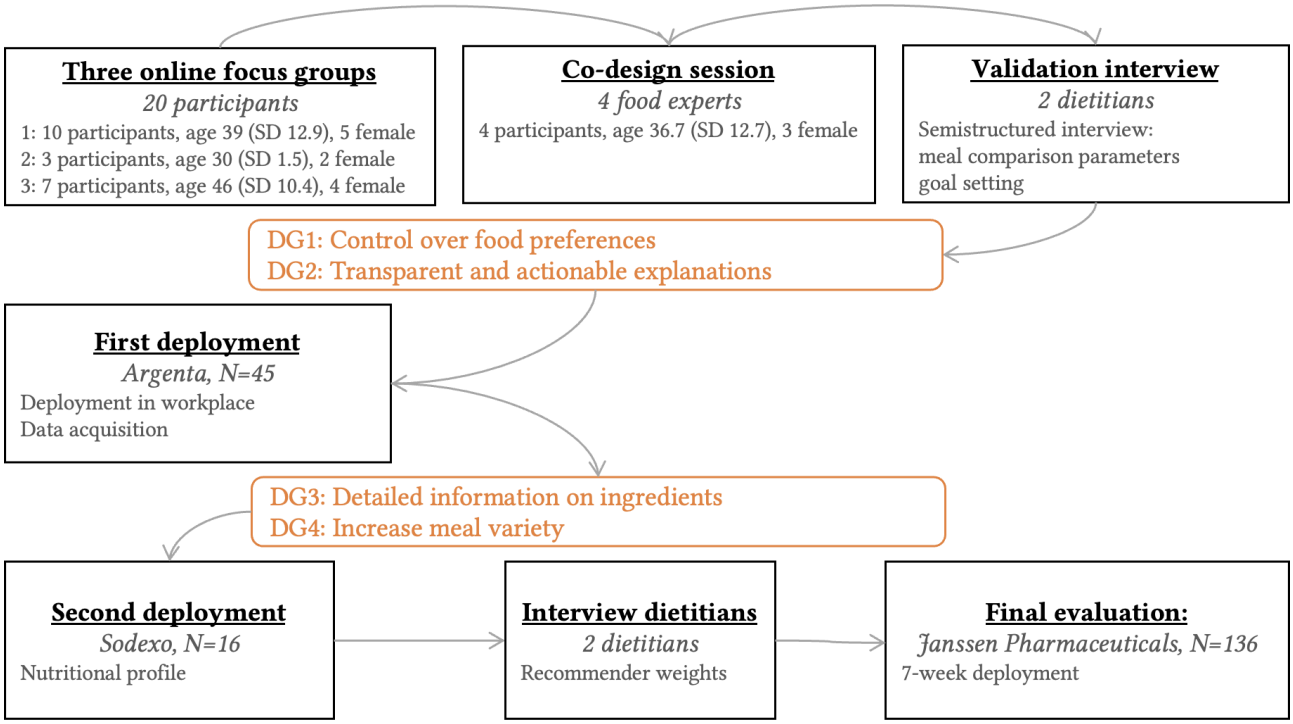
Methods

Study Design

This study used an iterative (see Figure 1), user-centered methodology to design, implement, and evaluate a food recommender system with an integrated explanation interface (see Figure 2). We gathered feedback through a multiphase approach:

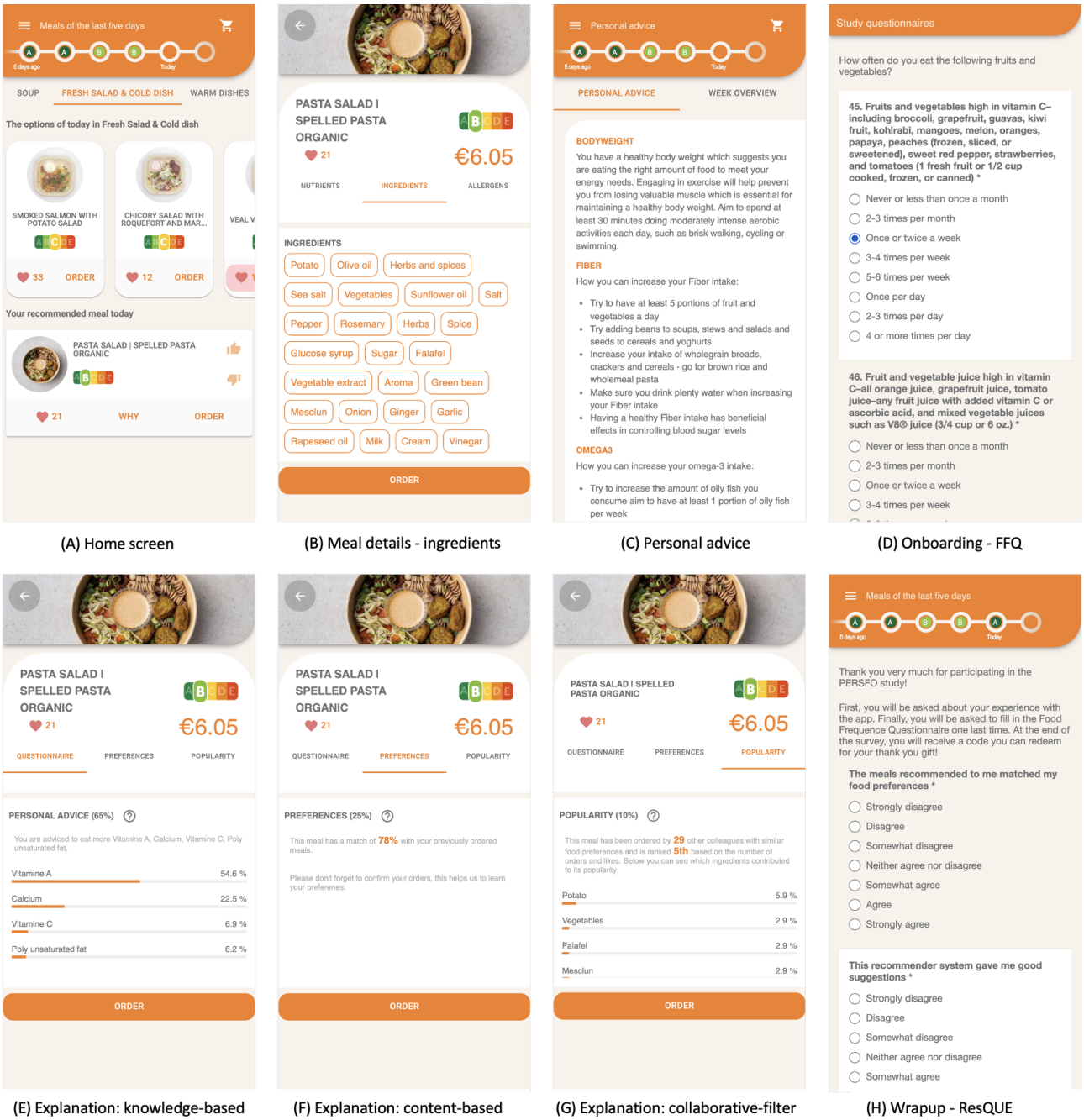
XSL•FO
RenderX

Figure 1. Timeline that shows the different steps. DG: design goals.



XSL•FO
RenderX

Figure 2. Screens and user interface components from the mobile application: (**A**) home screen shown after a user is logged in, (**B**) meal details shown when a user selects a meal, (**C**) personal advice based on the user's nutritional profile, part of the (**D**) Food Frequency Questionnaire (FFQ), (**E**) knowledge-based explanation, (**F**) content-based explanations, (**G**) collaborative-filter explanations, and part of the final (**H**) ResQue questionnaire.



Phase 0: Initial Exploration and Requirement Gathering

To understand potential barriers to integrating a transparent food recommender system within a food catering application, we conducted a multiphase stakeholder analysis. Two lead researchers, in close collaboration with food and technical experts, designed a set of 25 user interface designs. These were evaluated in 3 focus groups (N=20), a co-design session (N=4), and a semistructured interview with 2 dietitians. Participants rated and discussed the designs. The complete results are described in [40].

Phase 1 and 2: Data Acquisition and Rapid Prototyping

The goal of Phases 1 and 2 was to develop, deploy, and refine the system in real-life settings, gathering feedback on usability and functionality. The system was iteratively deployed at 2 companies chosen in collaboration with our food service partner (to ensure access to the necessary food management systems)—a bank (Phase 1, Argenta, N=45) and a facilities management company (Phase 2, Sodexo, N=16). These deployments with a diverse participant pool informed refinements of both the interface and the recommender engine before the proof-of-principle study.

To address the common "cold start" problem (ie, not enough initial data to create a user profile) in health recommenders [41] and personalize recommendations, we integrated a food frequency questionnaire (FFQ) [42]. FFQs offer a more comprehensive view of dietary habits than short-duration food records [43] and have demonstrated validity in previous research

XSL•FO
RenderX

[44]. We used a validated FFQ [45-47] and the European Union–funded Quisper platform [48] to generate personalized nutritional profiles [49,50]. The Quisper Food4Me service provided personalized feedback (Low, Normal, or High nutrient intake) based on pseudoanonymized FFQ data. Interaction logs and in-app user feedback were collected to inform system improvements. Ethical approval was granted by the university's ethics committee (approval number G2019-12-1911).

Phase 3: Proof-of-Principle Study

The goal of this third and final phase was to evaluate the system's impact on user understanding, trust, and dietary choices in a larger deployment. The refined system was deployed in a large pharmaceutical company restaurant (Janssen Pharmaceuticals, N=136). Pre- and post-study FFQ, the ResQue questionnaire, interaction logs, and open feedback functionality were used to collect comprehensive data about system usage, user perceptions, and potential dietary changes. Ethical approval was granted by the university's ethics committee (approval number G2021-3055-R2(MIN)).

Mobile Application and Explanation Interface

The home screen (Figure 2A) provides an overview of daily menu options, including courses (eg, soup, salads, etc), individual meals, and their nutriscore [51], a color-coded nutritional label that indicates the overall healthiness of a food item. Scores range from green (A, most nutritious) to red (E, least nutritious), helping users make informed choices at a glance. Users can tap a meal for detailed nutritional information, ingredients (Figure 2B), and allergens. Personalized dietary advice is accessible through the "Personal advice" screen (Figure 2C).

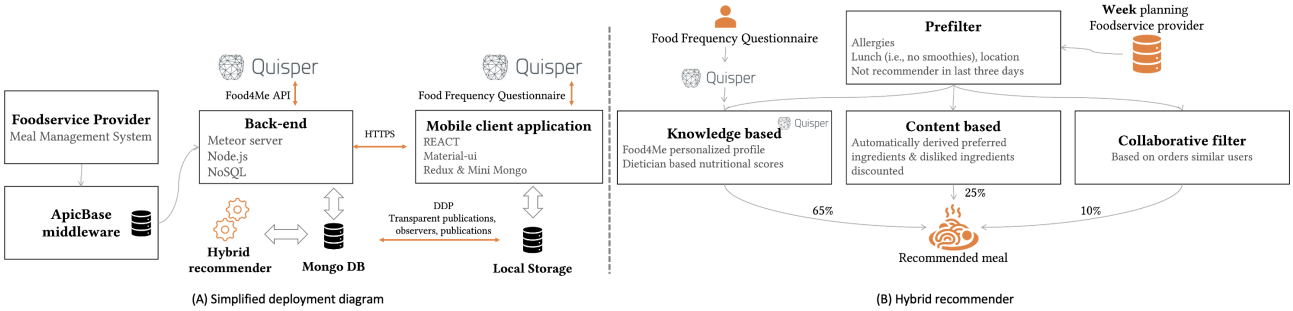
Users express preferences with a thumbs-up icon for liked recommendations (Figure 2A). If they disliked a recommendation, they were asked which ingredients contributed to their dislike. The why button reveals the following 3 explanations:

1. Nutritional alignment (Figure 2E): This explains how the recommended meal aligns with the user's nutritional profile and the weighting of each nutrient. A question mark icon offers additional context (eg, "Your recommendations are 65% influenced by your nutritional profile.").
2. Past choices (Figure 2F): This highlights how the recommendation relates to the user's previous orders and liked meals. A question mark icon clarifies the process (eg, "Based on your previously chosen meals, we try to find similar recipes.").
3. Popularity (Figure 2G): This reveals the meal's popularity among similar users and emphasizes that popularity is a minor recommendation factor (eg, "The most popular meals are not always the healthiest meals. Therefore, the popularity of a meal is only taken into account for 10%.").

Hybrid Recommender Engine

Our food recommender system (Figure 3A) integrates multiple strategies to personalize meal suggestions in real-world foodservice settings, with a focus on promoting healthy choices. Weekly menus designed by the main chefs of the company restaurants were accessed through standardized interfaces (RESTful API) and stored in a flexible database system (MongoDB). The underlying hybrid food recommender engine (Figure 3B) consists of 4 key building blocks:

Figure 3. Schematic overview of the (A) deployment diagram and (B) the different hybrid recommender building blocks.



Contextual Prefilter

The menu is initially filtered to ensure relevance and safety, considering:

- Location-specific offerings.
- User-reported allergies, promoting dietary safety.
- Recent recommendations (previous 3 days) to enhance meal variety and prevent repetition.

A Knowledge-Based Recommender

Leveraging the Quisper Food4Me integration [49], we generate personalized nutritional profiles that guide meal rankings. Users receive personalized feedback (Figure 2D and 2E), promoting dietary awareness. Based on the input from the stakeholder analysis, this emphasis on nutritional alignment accounts for 65% of the final recommendation score.

Content-Based Engine

A cosine-similarity algorithm assigned a personalized score to each meal based on the historical purchases of the user. Using the thumb-up and down icons, users were provided with additional controls to impact their recommendations. This content-based engine thus mainly assisted in differentiating between preferred options. The ranking score, therefore, only contributed 25% to the final ranking score.

Collaborative-Filter

After 2 weeks of data collection, a nearest-neighbor algorithm suggests meals based on the order histories of users with similar dietary patterns. Due to the intervention period, this

XSL•FO
RenderX

collaborative element contributes 10% to the final recommendation score.

Data Analysis and Recruitment

Qualitative data derived from the focus groups, co-design session, semistructured interviews, and in-app user feedback was analyzed using thematic analysis and the results are described in [40]. This analysis focused on identifying key themes related to user needs, preferences, design considerations, usability issues, and overall system perceptions. Interaction log data from all system deployments was analyzed using descriptive statistics. This analysis examined user engagement patterns (frequency of use, feature popularity, and interaction with meal recommendations) and provided insights into user behavior within the system. In Phase 3, data from the ResQue questionnaire [52] was analyzed using descriptive statistics to assess user perceptions of the explanations, their sense of control, and the system's impact on trust. Dietary intake data obtained through the FFQ at baseline and after the intervention was analyzed to examine potential changes in food choices and nutrient intake. We analyzed this data using the Belgian Food Composition Database (NUBEL) [53]. This analysis focused on comparing food group consumption and shifts in macronutrient proportions. User comments from the in-app feedback feature were analyzed to gain additional insights into user experiences, preferences, and suggestions for improvement.

Recruitment Strategies

Phases 1 and 2 prioritized usability testing and formative evaluation of the recommender system interface. Here, our primary goal was to gather feedback from a diverse range of users within each organization. This user diversity was essential for identifying usability issues, refining the interface, and optimizing the recommender engine's functionality. In this context, a group interested in exploring the system's features was more valuable than a perfectly representative sample of the overall employee population. For the later, proof-of-principle study (phase 3), we adopted a broader recruitment approach (emails likely supplemented by existing communication channels within the organizations) to ensure a larger, and more representative participant pool. While self-selection remains a limitation (more details in the Limitations section), our main objectives were to (1) identify key design considerations for balancing personalization, nutritional guidance, and user experience in food recommenders; and (2) assess the system's impact on understanding, trust, and dietary choices in a real-life setting.

Ethical Considerations

Participation was voluntary and fully anonymous in all phases to protect sensitive health data. The restaurant proprietor managed recruitment and communication with participants, while we handled only anonymized data, ie, participants were asked to use a pseudonym when creating an account on the study platform. Informed consent was obtained, emphasizing participant anonymity and secure data handling. Personally identifiable information was delinked from usage and dietary data as early as possible, and data was stored on secure servers with access limited to authorized researchers. The university ethics committee reviewed and approved all studies. Ethical approval for phase 1 and 2 was granted by the university's ethics committee (approval number G2019-12-1911). Ethical approval for phase 3 was granted by the university's ethics committee (approval number G2021-3055-R2(MIN)).

Results

Phase 0: Initial Exploration and Requirement Gathering

The complete results of the user centered design studies in Phase 0 are described in [40] and they provided insights into the needs of various stakeholders. This analysis yielded 2 key design goals (DG):

DG1: Control Over Food Preferences

End users and other stakeholders desire a food experience that balances authenticity, convenience, health, and environmental consciousness. The platform should accommodate varied preferences, including:

- Lifestyle and environmental: options such as vegan, halal, locally sourced food.
- Medical: clear allergen information.

DG2: Transparent and Actionable Explanations

Recommendations and explanations should enhance informed decision-making. Users expect control over food choices, requiring:

- Transparency: Clear reasoning behind recommendations.
- Actionable insights: Guidance on how recommendations align with individual preferences and health goals.

Phase 1: Data Acquisition

To test technical integration and gather data to train a recommender algorithm, we deployed an online, but fully working, mobile prototype (adapted due to COVID-19 restrictions) at Argenta Headquarters Belgium (December 7 - 18, 2020). Collaborating with the head chef, we designed a menu mirroring restaurant offerings. During the 2-week study, 45 participants placed 427, with an average of 9 orders per user and 9066 logged interactions. This yielded valuable technical insights and user feedback, leading to 2 additional design goals.

DG3: Fine-Grained Ingredient Data

Technical integration revealed limitations in the available ingredient data. Ingredient lists were often composite (eg, Bolognese sauce, without specifying its constituent ingredients) and precise portion sizes were unknown. This lack of granular data hinders effective ingredient-based recommendations [54]. In total. 10 participants echoed this concern, expressing difficulty making informed choices due to missing ingredient details: "I think it would be useful to list all ingredients in the description. For example, I see in the picture that there are pomegranate seeds with the salad, but that is nowhere to be read. Maybe it contains ingredients that someone is allergic to." Another participant highlighted the need for low-level ingredients: "Pasta with Alfredo sauce: how the hell do I know what Alfredo sauce is? The photo or the nutritional values give

XSL•FO
RenderX

me little or no info about that. I want to know what that is right (a kind of cheese sauce?) and what the predominant taste will be."

DG4: Increase Meal Variety

Although the menu used was prepared by the head chef as it would have been served in the restaurant, participants voiced concerns about limited healthy options, specifically a lack of vegetarian, gluten-free, and vegan choices. Comments like, "also today I find the healthy range too limited, far too little focus on vegetables;" or "it strikes me that there is little choice for people who really want to eat healthy, for example, there are virtually no vegetarian / gluten-free dishes or vegan options," highlight the need for greater meal variety to support diverse dietary needs.

Phase 2: Rapid Prototyping and Quisper Integration

We deployed an enhanced prototype for one week at Sodexo headquarters (December 3 - 10, 2021). A total of 16 participants completed the FFQ and received personalized feedback through the Food4Me Quisper service [49] that was tailored to local food. Users provided feedback on recommendations thumbs up and down icons or open feedback field. The main goal was to collect data for refining personalized recommendations based on nutritional profiles.

Our findings highlighted the need to balance personalized nutritional advice with the actual nutrient content of meals. We normalized meal nutritional values and developed a weighted scoring system with input from 2 professional dietitians (see Figure 1, Table 1). This aimed to optimize recommendations based on individual needs. For example, a person with low fiber intake would receive a higher score fiber-rich meals (weight=5). However, when fiber intake was already high, fiber-rich meals would not be further emphasized (weight=0). In contrast, when a person does not consume a lot of polyunsaturated fats, meals with a higher amount of polyunsaturated fat could receive a slightly higher ranking (weight=3). When a person consumes enough polyunsaturated fats, meals with a higher amount should be discounted (weight=−0.5). This approach, detailed in Table 1, ensures that recommendations align with individual user profiles without neglecting important dietary considerations.

Table . Nutrient weights for personalized recommendations. Higher weights for "low intake" indicate nutrients the user needs more of. Negative weights for "high intake" indicate nutrients the user should consume in smaller amounts.

Nutrient	Low intake, weight	High intake, weight
Fiber	5	0
Vitamin A	2	0
Vitamin B12	3	0
Vitamin C	5	0
Monounsaturated fat	1	0
Polyunsaturated fat	3	−0.5
Proteins	3	−0.3
Total fat	−1	−1
Calcium	5	0
Iron	2	0
Saturated fats	0	−1
Carbohydrate	3	−0.5

Phase 3: Proof-of-Principle Study

Our final proof-of-principle study ran from January 17, 2022 to March 4, 2022 at Janssen Pharmaceuticals. Each participant was asked to use the platform for 7 weeks and order their meals through the application, 176 employees initially expressed interest and 136 (77%) participants filled in the full FFQ and made at least 2 orders through the study application. Participants ranged in age from 23 to 61 (average age 41, SD 10 y), with 65% (88/136) female participants. Their BMI and weight followed a normal distribution. Despite COVID-19 restrictions limiting on-site work, participants made a mix of real and simulated food orders. We logged 17,597 interactions and 1031 orders placed by the app, demonstrating active usage. After completing the study, 41 participants who participated for 7 weeks and filled in the final questionnaires received a $42 (€40) healthy snack box and personalized dietary advice, similar to Figure 2C.

DG1: Control Over Food Preferences

User interaction logs demonstrate that users actively consulted and used the option to configure their own food preferences. Results show that participants consulted personalized recommendations and tried to keep their profile up to date: "I sometimes skip lunch due to time restricted feeding. There is no option to save this in the application." The most requested feature was to log meals from previous days or to add own options: For example, "you cannot enter fruit or yogurt into the system, which can also be useful. If necessary, you provide something that you can type yourself."

The food service provider supported 34 allergies (see Textbox 1) and 31 lifestyle preferences, yet some participants indicated they would have preferred even more options: "My allergy is

XSL•FO
RenderX

not in the list. Tomatoes and all preparations with tomato (such as with cocktail sauce, tomato powder in eg, marinades.)" Other potential features were options to integrate with existing diet programs, such as Weight Watchers, or the option to differentiate between a "cold" and "warm" dish: "I also like to eat warm, and I was always advised a cold dish [=salad]."

Textbox 1. Food preferences supported by the platform.

- Allergy: Barley, fish, lupine, mollusk, corn, wheat, sulfides, gluten, Brazil nuts, shellfish, mustard, sunflower seeds, egg, hazelnuts, celery, spelt, pistachio nuts, macadamia nuts, nut, lactose, poppy seeds, pecan nuts, rye, milk dairy, sesame, oats, walnuts, peanut, legume pulse, Khorasan wheat, cashews, soy, almonds, and seeds.
- Preference: Coriander, taurine, coloring, sulfur, phosphate, caffeine, nitrates, sorbates, alcohol, poultry meat, blackened, halal, beef, benzoates, kosher, acesulfame e962, vegan, aspartame e951, lamb, sweetener, preservative, pig meat, antioxidant, genetic modified, flavor enhancer, glutamate, quinine, vegetarian, carrot, cacao, and propionates.

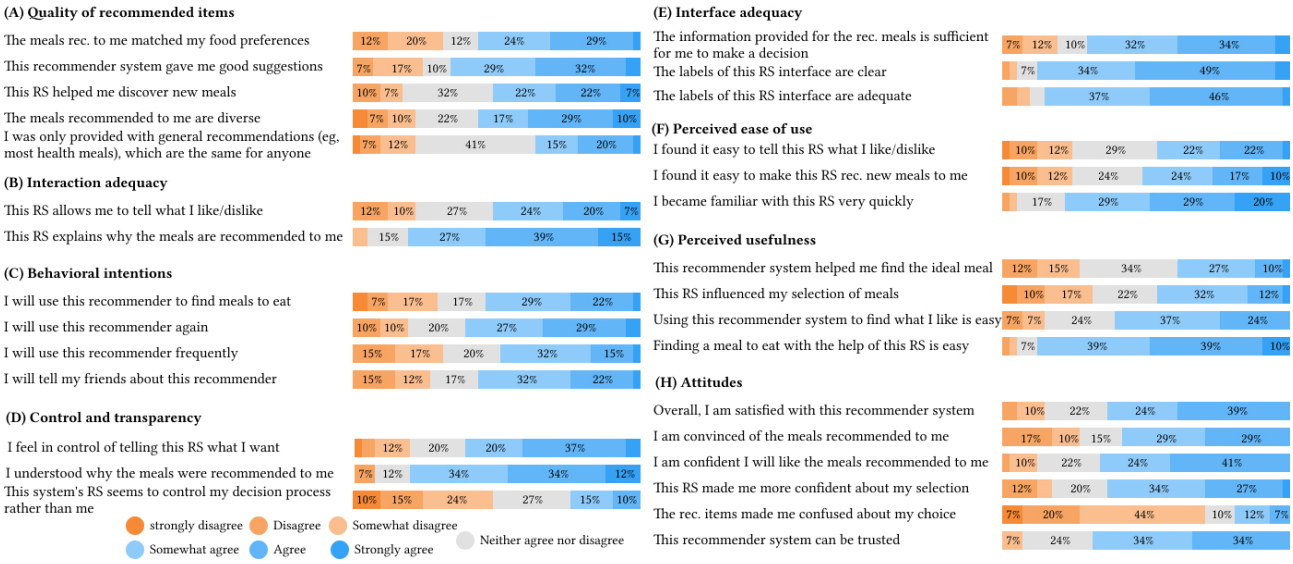
High-level dietary intake analysis of the 41 motivated participants who finished the 7-week study revealed positive initial trends in food choices. After the intervention, we observed an increased contribution of vegetables and fruits to total carbohydrate intake in several participants (3 and 1, respectively). This was accompanied by a decrease in reliance on starchy carbohydrates like potatoes, rice and pasta (6 at baseline, 3 after) and bread or savory biscuits (1 at baseline, 0 after). Protein intake patterns also shifted, with decreased reliance on dairy (3 at baseline, 2 after) and sweets or snacks (1 at baseline, 0 after), alongside increased vegetable intake (2 at baseline, 4 after). Differences in bread and savory biscuits (1 at baseline, 0 after); dairy (1 at baseline, 2 after); fats and spreads (7 at baseline, 10 after); meat and fish (17 at baseline, 13 after); and potatoes, rice, and pasta (0 at baseline, 1 after) were found when estimating food group contributions toward total fat intake.

DG2: Transparent and Actionable Explanations

ResQue questionnaire results (Figure 4, N=41) revealed a mismatch between what people wanted to eat and their healthy recommendations. Most respondents (33/41) found the explanations clear (36/41) and adequate (33/41) in explaining why meals were suggested. This understanding extended to the food recommender system and, 32/41 understood why meals were recommended to them, 32/41 indicated they quickly became familiar with the mobile user interface and found it easy to find a meal with the help of the app (36/41). Thanks to the explanations (Figure 2E), users received visual insight into their own dietary patterns and gained insights how the recommendations aligned with their nutritional needs.

However, the mismatch remains evident: only 22/41 felt the recommendations matched their food preferences. Despite this, explanations seem to have mitigated potential negative impacts.

Figure 4. Results from the final questionnaire: (**A**) quality of recommended items, (**B**) interaction adequacy, (**C**) behavioral intentions, (**D**) control and transparency, (**E**) interface adequacy, (**F**) perceived ease of use, (**G**) perceived usefulness, and (**H**) attitudes. RS: recommender system.



DG3: Fine-Grained Ingredient Data

Despite its importance to users, many food service providers lack detailed ingredient data. This study confirmed the need for ingredient breakdowns and precise portion information. Interaction logs revealed frequent use of the ingredient feature (961 views, 83% of orders), yet participants expressed a desire for more granularity. For example, one participant requested a breakdown of sugar types (glucose, fructose, etc) to support their diet. Interestingly, 69% (28/41) respondents found the provided information, including the summary information of the nutriscore [51], sufficient for making informed decision. Positive feedback like, "Definitely a nice system!" highlights the nutriscore's potential in aiding quick assessment of a meal's healthiness.

XSL•FO
RenderX

DG4: Meal Variety

Our real-life deployment of 7 weeks highlighted a challenge in balancing meal variety with personalized recommendations. The restaurant expanded its menu options during the study, but daily offerings remained limited (10 - 20 meals). Consequently, only half of participants perceived the recommendations as diverse. Most did not feel the system promoted generic options (14/41), and only 21/41 found it helpful for discovering new dishes. However, the extensive allergy and lifestyle filters further reduced recommendation variety for some users, leading to a perception of repetitiveness (eg, "It was unfortunate that I almost always got the same advice."). This reflects the design's focus on addressing individual nutritional needs. For example, when a participant scored low or high on a certain nutritional intake, they were recommended meals that contained more or less of that nutrient. After all, nutritional uptake is not "fixed" with 1 meal. This was also experienced by the participants: "I find it strange that I always get similar suggestions, there has never been a dish of the day or a salad of the week recommended [...]. This is very monotonous isn't it?" Despite this perceived monotony, 58% (24/41) of motivated participants liked the meals recommended to them and 65% (27/41) expressed confidence in future recommendations.

Discussion

Principal Findings

While the initial interest in the study was high (176 employees), the actual engagement was lower, with 136 participants (77%) completing the FFQ and actively using the application. It is also important to highlight that of these 136 participants, only 41 completed the ResQue questionnaire and the final FFQ after 7 weeks. This attrition rate is consistent with observations in other digital health interventions, which often face challenges in maintaining user engagement over time. This further reduction in participants was likely due to the extended length of the study and the requirement to consistently use the application for ordering meals. Therefore, it is important to note that the results from the ResQue questionnaire and the observed dietary changes are based on this subset of 41 participants. Despite the lower-than-expected uptake, the study provided valuable lessons learned into user preferences, design considerations, and the potential of personalized explanations to promote healthier food choices. Future research with a longer intervention period and post–follow-up is needed to fully assess sustained behavioral change and address the engagement challenges.

Lessons Learned

Our preliminary study suggest that motivated users might value user control features (DG1), such as allergy filters, feedback mechanisms, and personalized nutritional profiles. However, participants desired even greater customization, including additional allergy options, integration with dietary programs, and finer control over meal types (eg, warm or cold, only a sandwich during lunchtime). While increased control could potentially enhance meal variety (DG4), it must be balanced against practical limitations; for example, too strict food preferences limited the available meal options (eg, only 1 or 2 vegan options per day), and the potential for users to inadvertently undermine their own nutritional goals. Implementing meal customization options could potentially further increase meal variety, for example, "the choice of the type of sandwich would make up for a lot." But this has a lot of pragmatic implications, such as pricing and supply chain issues. Customization would also impact the nutritional values and require cumbersome actions from users to track a correct nutrition intake profile. Furthermore, increasing meal variety is more complex than simply introducing more variety in the menu options, dietary changes require an increased or decreased intake of the same user specific nutrients over a longer period of time [39].

The automatically generated nutritional profiles were crucial for the acceptance and trust (DG2). Motivated participants actively consulted profiles, tried to understand recommendations, and performed explicit actions to keep their nutritional profile up to date (eg, they even contacted the research team when they forgot to register a meal). However, at least 27% (11/41) felt the recommendations did not help them find healthier options, which could be due in part to their allergy or other lifestyle related filters, and 37% (15/41) felt recommendations were too general. We managed to support 80% (33/41) of remaining participants to understand their healthy food recommendations. However, consistent with previous research [29,47], we observed a mismatch between what people liked to eat and their healthy recommendations. While personalized nutritional profiles and explanations could have the potential to increase awareness of healthy eating, they cannot guarantee that all users will consistently opt for healthier choices.

Healthy Food Recommendations? Thanks, but No Thanks - Design Implications

Our study presents preliminary insights into the crucial dilemma between perceived user control and recommendation diversity, impacting both meal variety and potential dietary impact. Self-determination theory [55] underscores the importance of autonomy for lasting behavior change. The more individuals feel they understand their own eating habits, the more likely they are to adopt a healthy diet and achieve physical and psychological health [56]. However, if food recommenders mainly take user preferences into account, bad eating habits could be encouraged [21]. Furthermore, as our results demonstrate, excessive focus on preferences can reduce meal options, limiting variety and potentially hindering dietary change.

Prioritizing either nutrients or meal variety has implications for the system's effectiveness. Dietary changes can only be achieved after there is a significant increase (eg, calcium) or decrease (eg, saturated fats) of certain nutrients. However, emphasizing nutrient optimization could lead to recommendations users find unpalatable or repetitive, reducing perceived control [21,57]. Conversely, focusing solely on variety risks reinforcing existing preferences, even unhealthy ones [21].

By helping users understand the rationale behind recommendations and their connection to individual nutritional profiles, we could potentially increase perceived control and potentially motivate change. Therefore, a successful food

XSL•FO
RenderX

recommender design must carefully balance control features with personalized nutritional guidance [21]. Clear explanations can support understanding and could create a sense of autonomy, increasing acceptance and trust in the system.

Limitations and Future Work

There are important limitations to this preliminary work that need to be acknowledged.

First, participants were self-selected for greater interest in healthy eating, introducing bias. Despite this, our design implications remain valuable for supporting motivated users, who are the most likely adopters of food recommender applications. Our focus on human-computer interaction methods aims to address the gap between healthy eating motivation and action, making our findings relevant to this target audience.

Second, COVID-19 restrictions necessitated a mix of real and virtual orders. Virtual orders may reflect idealized choices rather than true consumption habits. However, we have examined if any significant trends exist between order types (simulated vs actual) related to both the healthiness of choices and user perceptions captured by the ResQue questionnaire. No difference could be observed within the group of 41 participants that filled in the final questionnaires. It is also worth noting that the abovementioned selection bias, could have positively influenced the similarity in trends between simulated and real orders. Participants who were genuinely interested in exploring healthier eating may have been more conscientious about both their real and virtual meal choices, minimizing the potential discrepancy. While this does not negate the limitations for translating virtual choices into real-world dietary change, it suggests a level of engagement that benefits the exploration of system design elements. Nonetheless, a longer intervention with post–follow-up is needed in future work to fully assess sustained behavioral change.

Finally, the field of transparent recommender systems is evolving [26], and comparative studies exploring alternative explanation methods, such as multilist interfaces [57], could offer insights into greater effectiveness. Future research possibilities include a more advanced content-based algorithm using detailed ingredient data (currently the exact ingredient contributions were not available for every meal); and the ability for users to log meals consumed outside the company restaurant for a more complete nutritional assessment.

Acknowledgments

The authors would like to thank all the participants. We would also like to thank all project partners. This work was undertaken within the PERSFO Project (ID 20291) which received funding from EIT Food, the innovation community on Food of the European Institute of Innovation and Technology (EIT), a body of the EU, under Horizon 2020, the EU Framework Programme for Research and Innovation.

Conflicts of Interest

None declared.

References

1. World Health Organization. Obesity and overweight. URL: https://www.who.int/newsroom/factsheets/detail/obesity-and-overweight [accessed 2023-07-18]
2. NCD Alliance. Unhealthy diets and malnutrition. URL: https://www.who.int/activities/preventingnoncommunicable-diseases [accessed 2023-07-18]
3. Spronk I, Kullen C, Burdon C, O'Connor H. Relationship between nutrition knowledge and dietary intake. Br J Nutr 2014 May 28;111(10):1713-1726. [doi: 10.1017/S0007114514000087] [Medline: 24621991]
4. Appleton KM, Bray J, Price S, et al. A mobile phone app for the provision of personalized food-based information in an eating-out situation: development and initial evaluation. JMIR Form Res 2019 Nov 4;3(4):e12966. [doi: 10.2196/12966] [Medline: 31682575]
5. Goossens J, Astley S, Mathijssen G, Simillion H, Finglas P, Lizano DS. Introducing personalised nutrition services in a behavioural motivation concept for connected food service environments. Agro Food Ind Hi Tech 2020;31(2):42-45.
6. de Hoogh IM, Reinders MJ, Doets EL, Hoevenaars FPM, Top JL. Design Issues in Personalized Nutrition Advice Systems. J Med Internet Res 2023 Mar 29;25:e37667. [doi: 10.2196/37667] [Medline: 36989039]
7. Lizano DS, Goossens J, Astley S, Berry R, Finglas P. Quisper and development of a personalised nutrition community. Agro Food Ind Hi Tech 2020;31(2):36-38.
8. Zenun Franco R, Fallaize R, Weech M, Hwang F, Lovegrove JA. Effectiveness of web-based personalized nutrition advice for adults using the eNutri web app: evidence from the EatWellUK randomized controlled trial. J Med Internet Res 2022 Apr 25;24(4):e29088. [doi: 10.2196/29088] [Medline: 35468093]
9. Elsweiler D, Harvey M, Ludwig B, Said A. Bringing the" healthy" into food recommenders. In: DMRS 2015:33-36.
10. Starke A, Trattner C, Bakken H, Johannessen M, Solberg V. The cholesterol factor: balancing accuracy and health in recipe recommendation through a nutrient-specific metric. Presented at: Proceedings of the 1st Workshop on Multi-Objective Recommender Systems (MORS 2021): co-located with 15th ACM Conference on Recommender Systems (RecSys 2021); Sep 25, 2021; Amsterdam, Netherlands p. 1-11 URL: http://ceur-ws.org/Vol-2959/paper8.pdf [accessed 2025-02-10]

XSL•FO
RenderX

11. Yürüten O. Recommender Systems for Healthy Behavior Change. Tech. rep. : EPFL; 2017.
12. Hidalgo JI, Maqueda E, Risco-Martín JL, Cuesta-Infante A, Colmenar JM, Nobel J. glUCModel: A monitoring and modeling system for chronic diseases applied to diabetes. J Biomed Inform 2014 Apr;48:183-192. [doi: 10.1016/j.jbi.2013.12.015] [Medline: 24407050]
13. Rabbi M, Aung MH, Zhang M, Choudhury T. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing UbiComp '15; Sep 7-11, 2015; Osaka Japan p. 707-718. [doi: 10.1145/2750858.2805840]
14. Ronteltap A, van Trijp H, Berezowska A, Goossens J. Nutrigenomics-based personalised nutritional advice: in search of a business model? Genes Nutr 2013 Mar;8(2):153-163. [doi: 10.1007/s12263-012-0308-4] [Medline: 22903899]
15. Ng YK, Jin M. Personalized recipe recommendations for toddlers based on nutrient intake and food preferences. Presented at: Proceedings of the 9th International Conference on Management of Digital EcoSystems MEDES '17; Nov 7-10, 2017; Bangkok, Thailand p. 243-250. [doi: 10.1145/3167020.3167057]
16. Pawar KR, Ghorpade T, Shedge R. Constraint based recipe recommendation using forward checking algorithm. Presented at: 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI); Sep 21-24, 2016; Jaipur, India p. 1474-1478. [doi: 10.1109/ICACCI.2016.7732256]
17. Elsweiler D, Harvey M. Towards automatic meal plan recommendations for balanced nutrition. Presented at: Proceedings of the 9th ACM Conference on Recommender Systems RecSys '15; Sep 16-20, 2015; Vienna, Austria p. 313-316. [doi: 10.1145/2792838.2799665]
18. Ntalaperas D, Bothos E, Perakis K, Magoutas B, Mentzas G. DISYS: an intelligent system for personalized nutritional recommendations in restaurants. Presented at: Proceedings of the 19th Panhellenic Conference on Informatics PCI '15; Oct 1-3, 2015; Athens, Greece p. 382-338. [doi: 10.1145/2801948.2801997]
19. Choi Y, Hoops SL, Thoma CJ, Johnson AJ. A Guide to Dietary Pattern-Microbiome Data Integration. J Nutr 2022 May 5;152(5):1187-1199. [doi: 10.1093/jn/nxac033] [Medline: 35348723]
20. Trattner C, Elsweiler D. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. Presented at: Proceedings of the 26th International Conference on World Wide Web WWW '17; Apr 3-7, 2017 p. 489-498. [doi: 10.1145/3038912.3052573]
21. Trang Tran TN, Atas M, Felfernig A, Stettinger M. An overview of recommender systems in the healthy food domain. J Intell Inf Syst 2018 Jun;50(3):501-526. [doi: 10.1007/s10844-017-0469-0]
22. Lawo D, Neifer T, Esau M, Stevens G. Buying the 'right' thing: designing food recommender systems with critical consumers. Presented at: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems CHI '21; May 8-13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445264]
23. Trattner C, Elsweiler D. Food recommender systems: important contributions, challenges and future research directions". arXiv. Preprint posted online on Nov 7, 2017 URL: https://doi.org/10.48550/arXiv.1711.02760
24. Musto C, Starke AD, Trattner C, Rapp A, Semeraro G. Exploring the effects of natural language justifications in food recommender systems. Presented at: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization; Jun 2021 to 2021; Utrecht Netherlands p. 147-157. [doi: 10.1145/3450613.3456827]
25. El Majjodi A, Starke AD, Trattner C. Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system. Presented at: UMAP '22Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization UMAP '22; Jul 4-7, 2022; Barcelona Spain p. 48-56. [doi: 10.1145/3503252.3531312]
26. He C, Parra D, Verbert K. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. Expert Syst Appl 2016 Sep;56:9-27. [doi: 10.1016/j.eswa.2016.02.013]
27. Ooge J, Kato S, Verbert K. Explaining recommendations in E-learning: effects on adolescents' trust. Presented at: 27th International Conference on Intelligent User Interfaces IUI '22; Mar 22-25, 2022; Helsinki Finland p. 93-105. [doi: 10.1145/3490099.3511140]
28. Rader E, Cotter K, Cho J. Explanations as mechanisms for supporting algorithmic transparency. Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems CHI '18; Apr 21-26, 2018; Montreal QC Canada p. 1-13. [doi: 10.1145/3173574.3173677]
29. Tintarev N, Masthoff J. Explaining recommendations: design and evaluation. In: Recommender Systems Handbook 2015.
30. Zhang Y, Chen X. Explainable Recommendation: A Survey and New Perspectives. Found Trends Inf Retr 2020;14(1):1-101. [doi: 10.1561/1500000066]
31. Gutiérrez F, Htun NN, Vanden Abeele V, De Croon R, Verbert K. Explaining call recommendations in nursing homes: a user-centered design approach for interacting with knowledge-based health decision support systems. Presented at: 27th International Conference on Intelligent User Interfaces IUI '22; Mar 22-25, 2022; Helsinki Finland. [doi: 10.1145/3490099.3511158]
32. Gutiérrez F, Charleer S, De Croon R, Htun NN, Goetschalckx G, Verbert K. Explaining and exploring job recommendations. Presented at: Proceedings of the 13th ACM Conference on Recommender Systems RecSys '19; Sep 16-20, 2019; Copenhagen Denmark. [doi: 10.1145/3298689.3347001]

XSL•FO
RenderX

33. Gutiérrez F, Htun NN, Charleer S, Verbert K. Designing augmented reality applications for personal health decision-making. : Hawaii, USA Presented at: Proceedings of the 52nd Hawaii International Conference on System Sciences; Jan 8-11, 2019 p. 1738-1747.
34. Werthner H, Zanker M, Golbeck J, Semeraro G. Nudging grocery shoppers to make healthier choices. Presented at: Proceedings of the 9th ACM Conference on Recommender Systems RecSys '15; Sep 16-20, 2015; Vienna Austria p. 289-292. [doi: 10.1145/2792838]
35. Ge M, Elahi M, Fernaández-Tobías I, Ricci F, Massimo D. Using tags and latent factors in a food recommender system. Presented at: Proceedings of the 5th International Conference on Digital Health 2015 DH '15; May 18-20, 2015; Florence Italy p. 105-112. [doi: 10.1145/2750511.2750528]
36. Starke AD, Asotic E, Trattner C, Van Loo EJ. Examining the User Evaluation of Multi-List Recommender Interfaces in the Context of Healthy Recipe Choices. ACM Trans Recomm Syst 2023 Dec 31;1(4):1-31. [doi: 10.1145/3581930]
37. Agapito G, Simeoni M, Calabrese B, et al. DIETOS: A dietary recommender system for chronic diseases monitoring and management. Comput Methods Programs Biomed 2018 Jan;153:93-104. [doi: 10.1016/j.cmpb.2017.10.014] [Medline: 29157465]
38. Elahi M, Ge M, Ricci F, Massimo D, Berkovsky S. Interactive food recommendation for groups. Presented at: Poster Proceedings of the 8th ACM Conference on Recommender Systems, RecSys 2014; Oct 16-20, 2014; Foster City, Silicon Valley, USA p. 1-2.
39. Hauptmann H, Leipold N, Madenach M, et al. Effects and challenges of using a nutrition assistance system: results of a long-term mixed-method study. User Model User-Adap Inter 2022 Nov;32(5):923-975. [doi: 10.1007/s11257-021-09301-y]
40. Gutiérrez F, Berry R, Astley S, Verbert K, Abeele VV. Motivational design techniques to support healthy eating habits at work. Presented at: 2020 IEEE International Conference on Healthcare Informatics (ICHI); Nov 30 to Dec 3, 2020 p. 1-3. [doi: 10.1109/ICHI48887.2020.9374350]
41. De Croon R, Van Houdt L, Htun NN, Štiglic G, Vanden Abeele V, Verbert K. Health Recommender Systems: Systematic Review. J Med Internet Res 2021 Jun 29;23(6):e18035. [doi: 10.2196/18035] [Medline: 34185014]
42. Forster H, Fallaize R, Gallagher C, et al. Online dietary intake estimation: the Food4Me food frequency questionnaire. J Med Internet Res 2014 Jun 9;16(6):e150. [doi: 10.2196/jmir.3105] [Medline: 24911957]
43. Paisley CM, Lloyd HM, Brown W, Mela DJ. Production and validation of a food frequency questionnaire to estimate dietary fat intake. J Human Nutrition Diet 1996 Jun;9(3):189-206. [doi: 10.1046/j.1365-277X.1996.00450.x]
44. Willett W. Nutritional Epidemiology: Oxford University Press; 2012, Vol. 40.
45. Kaiser B, Stelzl T, Finglas P, Gedrich K. The Assessment of a Personalized Nutrition Tool (eNutri) in Germany: Pilot Study on Usability Metrics and Users' Experiences. JMIR Form Res 2022 Aug 4;6(8):e34497. [doi: 10.2196/34497] [Medline: 35925664]
46. Valavanis IK, Mougiakakou SG, Grimaldi KA, Nikita KS. A multifactorial analysis of obesity as CVD risk factor: use of neural network based methods in A nutrigenetics context. BMC Bioinformatics 2010 Sep 8;11(1):1-10. [doi: 10.1186/1471-2105-11-453] [Medline: 20825661]
47. Valavanis IK, Mougiakakou SG, Marinos S, et al. Gene-nutrition interactions in the onset of obesity as cardiovascular disease risk factor based on a computational intelligence method. Presented at: 2008 8th IEEE International Conference on BioInformatics and BioEngineering; Oct 8-10, 2008; Athens, Greece.
48. Goossens J, Mellaerts R. Quisper - a neutral digital platform to enable the creation of scientifically validated personalised nutrition services. Agro Food Ind Hi Tech 2018;29(5):30-35.
49. Celis-Morales C, Livingstone KM, Marsaux CFM, et al. Effect of personalized nutrition on health-related behaviour change: evidence from the Food4me European randomized controlled trial. Int J Epidemiol 2016 Aug(2):578-588. [doi: 10.1093/ije/dyw186]
50. Celis-Morales C, Mathers J, Gibney M, et al. White paper on personalised nutrition - paving a way to better population health. 2015. [doi: 10.13140/RG.2.2.13147.16166]
51. Chantal J, Hercberg S, Organization WH, et al. Development of a new front-of-pack nutrition label in France: the five-colour Nutri-Score. Public Health Panor 2017;3(4):712-725. [doi: 10.1016/S2468-2667(18)30009-4]
52. Pu P, Chen L, Hu R. Evaluating recommender systems from the user's perspective: survey of the state of the art. User Model User-Adap Inter 2012 Oct;22(4-5):317-355. [doi: 10.1007/s11257-011-9115-7]
53. Seeuws C. Belgian Branded Food Products Database: Inform consumers on a healthy lifestyle in a public-private partnership. J Food Compost Anal 2017 Dec;64:39-42. [doi: 10.1016/j.jfca.2017.07.008]
54. Ludwig B, Meyer S, Dietz J, Donabauer G, Pfaffelhuber A, Elsweiler D. Recommending better food choices for clinically obese users. Presented at: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization; Jun 21-25, 2021; Utrecht Netherlands p. 79-82. [doi: 10.1145/3450614.3461677]
55. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior: Springer Science & Business Media; 2013.
56. Guertin C, Rocchi M, Pelletier LG, Émond C, Lalande G. The role of motivation and the regulation of eating on the physical and psychological health of patients with cardiovascular disease. J Health Psychol 2015 May;20(5):543-555. [doi: 10.1177/1359105315573471] [Medline: 25903242]

XSL•FO
RenderX

57. Starke A, Asotic E, Trattner C. “Serving each user”: supporting different eating goals through a multi-list recommender interface. Presented at: RecSys ’21: Fifteenth ACM Conference on Recommender Systems; Sep 27 to Oct 1, 2021; Amsterdam Netherlands p. 124-132. [doi: 10.1145/3460231.3474232]

Abbreviations

DG: design goals
EIT: European Institute of Innovation and Technology
FFQ: Food Frequency Questionnaire

Edited by L Buis; submitted 26.07.23; peer-reviewed by C Jacob, K Appleton, K Ferrini; revised version received 18.10.24; accepted 20.12.24; published 11.02.25.

Please cite as:
De Croon R, Segovia-Lizano D, Finglas P, Vanden Abeele V, Verbert K
An Explanation Interface for Healthy Food Recommendations in a Real-Life Workplace Deployment: User-Centered Design Study
JMIR Mhealth Uhealth 2025;13:e51271
URL: https://mhealth.jmir.org/2025/1/e51271
doi:10.2196/51271

© Robin De Croon, Daniela Segovia-Lizano, Paul Finglas, Vero Vanden Abeele, Katrien Verbert. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Effects of Self-Monitoring Using a Smartwatch and Smartphone App on Stress Awareness, Self-Efficacy, and Well-Being–Related Outcomes in Police Officers: Longitudinal Mixed Design Study

Herman Jaap de Vries[1], PhD; Roos Delahaij[1], PhD; Marianne van Zwieten[2], MSc; Helen Verhoef[3], MSc; Wim Kamphuis[1], PhD

[1]Department of Learning and Workforce Development, The Netherlands Organisation for Applied Scientific Research, Soesterberg, Netherlands

[2]Department of Work Health Technology, The Netherlands Organisation for Applied Scientific Research, Leiden, Netherlands

[3]Department of Sustainable Productivity and Employability, The Netherlands Organisation for Applied Scientific Research, Leiden, Netherlands

Corresponding Author:
Wim Kamphuis, PhD
Department of Learning and Workforce Development, The Netherlands Organisation for Applied Scientific Research, Soesterberg, Netherlands

Abstract

Background: Wearable sensor technologies, often referred to as "wearables," have seen a rapid rise in consumer interest in recent years. Initially often seen as "activity trackers," wearables have gradually expanded to also estimate sleep, stress, and physiological recovery. In occupational settings, there is a growing interest in applying this technology to promote health and well-being, especially in professions with highly demanding working conditions such as first responders. However, it is not clear to what extent self-monitoring with wearables can positively influence stress- and well-being–related outcomes in real-life conditions and how wearable-based interventions should be designed for high-risk professionals.

Objective: The aim of this study was to investigate (1) whether offering a 5-week wearable-based intervention improves stress- and well-being–related outcomes in police officers and (2) whether extending a basic "off-the-shelf" wearable-based intervention with ecological momentary assessment (EMA) questionnaires, weekly personalized feedback reports, and peer support groups improves its effectiveness.

Methods: A total of 95 police officers from 5 offices participated in the study. The data of 79 participants were included for analysis. During the first 5 weeks, participants used no self-monitoring technology (control period). During the following 5 weeks (intervention period), 41 participants used a Garmin Forerunner 255 smartwatch with a custom-built app (comparable to that of the consumer-available wearable), whereas the other 38 participants used the same system, but complemented by daily EMA questionnaires, weekly personalized feedback reports, and access to peer support groups. At baseline (T0) and after the control (T1) and intervention (T2) periods, questionnaires were administered to measure 15 outcomes relating to stress awareness, stress management self-efficacy, and outcomes related to stress and general well-being. Linear mixed models that accounted for repeated measures within subjects, the control and intervention periods, and between-group differences were used to address both research questions.

Results: The results of the first analysis showed that the intervention had a small (absolute Hedges g=0.25 - 0.46) but consistent effect on 8 of 15 of the stress- and well-being–related outcomes in comparison to the control group. The second analysis provided mixed results; the extended intervention was more effective than the basic intervention at improving recovery after work but less effective at improving self-efficacy in behavior change and sleep issues, and similarly effective in the remaining 12 outcomes.

Conclusions: Offering a 5-week wearable-based intervention to police officers can positively contribute to optimizing their stress-related, self-efficacy, and well-being–related outcomes. Complementing the basic "off-the-shelf" wearable-based intervention with additional EMA questionnaires, weekly personalized feedback reports, and peer support groups did not appear to improve the effectiveness of the intervention. Future work is needed to investigate how different aspects of these interventions can be tailored to specific characteristics and needs of employees to optimize these effects.

(JMIR Mhealth Uhealth 2025;13:e60708) doi:10.2196/60708

XSL•FO
RenderX

KEYWORDS

wearable electronic devices; ecological momentary assessment; psychological stress; psychological well-being; awareness; self-efficacy; occupational medicine; emergency responders; well-being; psychological; efficacy; stress; wearables; wearable device; smartwatch; smartphone app; app; sensor; sensor technology; police officers; questionnaire; stress awareness; stress management

Introduction

Background

First responders, such as police officers and firefighters, are regularly faced with highly demanding working conditions [1] such as threatening or emotionally demanding situations. These roles frequently involve irregular shifts and peak workloads, sometimes compounded by understaffing [2]. Such conditions can lead to increased stress levels [3], reduced mental well-being [1,4], sleep disturbances [5], and difficulties in maintaining a healthy lifestyle [6], thereby escalating the risks of long-term health issues [7] and absenteeism [8]. Although some of the inherent challenges of these professions are unavoidable, it could be beneficial for employees to gain insights into their own stress levels, recovery patterns, and lifestyle-related factors like sleep and physical activity [9]. Such personal monitoring could help them to improve their stress- and well-being–related awareness and self-efficacy to make behavioral adjustments in order to better manage demanding aspects of their professions and thus optimize their well-being.

The emergence of wearable sensor technology facilitates promising opportunities to monitor a broad set of health-related outcomes [10]. Besides the initial use cases of wearables in estimating daily physical activity levels [11] and sleep duration [12], the introduction of photoplethysmography sensors enabled the measurement of other metrics such as heart rate [13], heart rate variability (HRV) [14], oxygen saturation [15], and potentially blood pressure [16]. Wearable-based measurements such as sleep and resting HRV have already been linked to stress- and well-being–related outcomes in first responders [17,18], but are sometimes also used to estimate other relevant outcomes. Using measures such as HRV as inputs, wearable companies now also estimate mental stress levels [19,20] and physiological recovery [21] in an attempt to provide more easily interpretable feedback. Although conceptually useful, in consumer-available wearables this is almost exclusively done via black box algorithms and the created measures do not always correlate optimally with similar subjective outcomes [22]. Despite these limitations and need for more scientific evaluation of the performance of these metrics, the wearables that utilize them are already broadly commercially available and adopted and are being integrated into wearable-based interventions [23]. In these wearable-based interventions, individuals use real-time data from physiological measurements to improve their health and performance, also known as biofeedback [24]. Given the continuing adoption of these wearables, it is important to understand the impact of these devices on behavior change and stress- and well-being–related outcomes.

The Effectiveness of Wearable-Based Interventions

Wearable-based interventions can be supportive at increasing physical activity [25], reducing body weight [25-27], and improving sleep outcomes [28]. Regarding the application of wearables to reduce stress and support mental well-being, most research efforts are focused on the modeling of stress-related outcomes using wearables [29]. Although studies show that wearable-based stress detection is feasible, this is especially true in controlled (eg, laboratory) settings, whereas challenges remain for accurate measurement in real-life circumstances [30]. A few studies have been performed investigating the effectiveness of wearable-based interventions to reduce stress and improve well-being, and they found some promising results [31-33]. A recent review that surveyed the broader literature on this topic showed that these effect studies tend to assess acute stress responses in laboratory settings, often in young (student) populations [34]. The authors concluded that there is therefore a need for more studies that investigate the effects of longitudinal wearable-based interventions on stress- and well-being–related outcomes in a natural, real-life context.

To improve our understanding of the effects of wearable-based interventions, it is not only helpful to gain insight into their direct impact on outcomes such as stress and mental well-being, but also in the possible precursors to those outcomes. Effects on these precursors could be meaningful even in the absence of short-term effects on the primary outcomes and may substantiate why the found effects occur. With regard to these precursors, the Health Belief Model proposes that health-related awareness (eg, on the susceptibility for and severity of the potential health problem) and perceived self-efficacy to cope with the situation are necessary to come to action and change behavior, which in turn may improve health-related outcomes [35,36]. Metacognitive awareness (MCA) has been shown to be related to learning [37] in general and learning about coping with stress specifically [38]. MCA is defined as the ability to reflect upon, understand, and regulate one's learning [37]. This concept distinguishes between insight in one's responses and potential causes and ways to regulate these responses. MCA helps people to develop strategies to change their behavior in accordance with the situation.

Self-efficacy is considered a key mechanism through which behavioral change can be observed [39]. This has also been shown for the way people cope with stress [40]. People's beliefs about their ability to change behavior is determined, among other factors, by mastery experiences (succeeding at changing behavior) and vicarious learning (seeing other succeed). Daily feedback through monitoring can provide valuable insight into how to change behavior and mastery of these behaviors in attainable steps.

Since most wearables are designed to improve their users' awareness of their own health-related behavior and provide them with leads on what needs to change and how they are progressing, improvement in health outcomes may be preceded by increased awareness and self-efficacy. Some studies have

XSL•FO
RenderX

confirmed that wearable-based interventions can indeed positively influence health-related awareness [41] and self-efficacy in patient populations [42,43], but it is unknown to what extent these findings translate to occupational settings. Health-related behavior change interventions generally require a duration of months to years in order to ingrain the new behaviors [44], but when studying awareness and other precursors of behavior change, a shorter duration (eg, weeks) is sufficient.

Therefore, the first research question that this study addresses is the following:

Research question 1: Does offering a 5-week wearable-based intervention improve awareness, self-efficacy, and stress- and well-being–related outcomes in police officers in comparison to a control period?

The Potential Added Value of Enhanced Feedback and Implementation

For policymakers considering the deployment of wearable-based interventions, it is also helpful to know to what extent it is sufficient to use an "off-the-shelf" solution, where consumer-available wearables are directly deployed to the target group, or whether it is necessary, for instance, to provide (1) more extensive feedback and (2) (peer) support for the intervention to be effective. Based on self-efficacy theory, the latter could be of added value as it provides potential for vicarious learning and supports awareness of ways to attain mastery experience, whereas the former may contribute to the development of increased awareness.

First, it is possible to additionally collect subjective data via brief daily questionnaires, also known as ecological momentary assessment (EMA), and utilize the data to provide feedback on outcomes that cannot be measured using wearables. These self-report data can also be used to provide contextual and subjective information to the objective wearable-based measures [45]. Providing this more enriched feedback could potentially have beneficial effects on awareness as it links personal experience to objective measurement and provides information on the behavioral context that affects stress and well-being. This in turn can inform users on strategies to attain behavior change.

Second, organizing peer support groups can be a way to facilitate collaborative reflection and group learning [46]. Literature on the effectiveness of peer consultation groups in a preventive occupational health context is scarce, but there are indications that these can positively contribute to recovery from mental health conditions [47-49]. Although different from peer consultation through the guidance of more experienced professionals, peer supervision has also been shown to have positive effects on self-awareness and self-efficacy in occupational settings [50,51]. It is therefore possible that utilizing peer support groups may benefit the effectiveness of wearable-based interventions that aim to achieve health-related behavior change and optimization of stress- and well-being–related outcomes.

To contrast the effects of a condition that is enriched by both enhanced feedback and added peer support to a more simple "off-the-shelf" intervention that just utilizes a consumer-available wearable, the second research question is as follows:

Research question 2: Is offering a 5-week wearable-based intervention that is extended by EMA, personalized weekly feedback, and peer support meetings ("Wearable+") more effective at improving awareness, self-efficacy, and stress- and well-being–related outcomes in police officers than an "off-the-shelf" ("Wearable") intervention?

Methods

Ethical Considerations

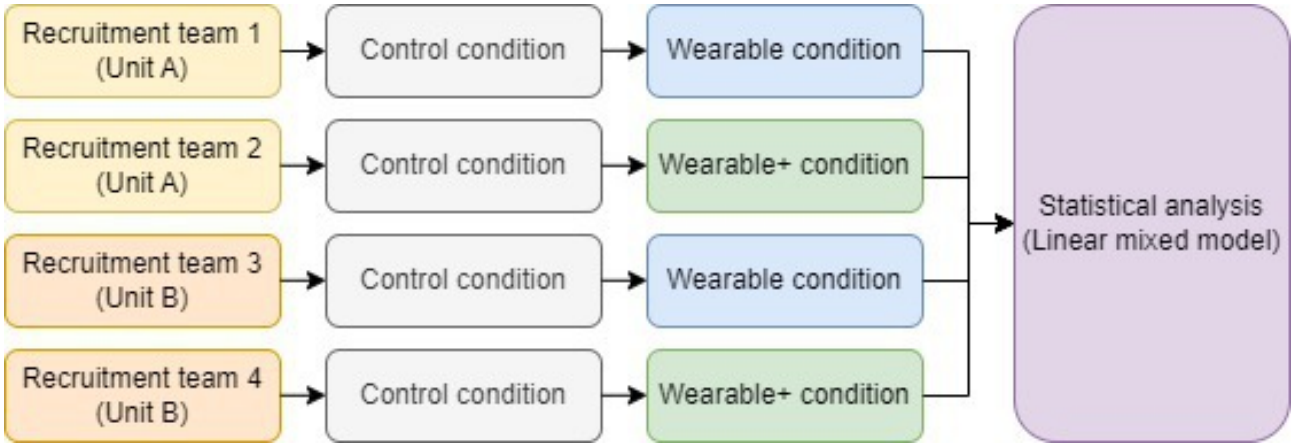
The study protocol was approved (case 2022 - 100) by the internal Research Ethics Committee of the Netherlands Organisation for Applied Scientific Research (TC-nWMO). Participation occurred voluntarily based on informed consent, and participants could opt out at any time without further consequences. Data were collected in a custom-built app that allowed for full control over the collected data without using consumer cloud services. During data collection, data were pseudonymously stored by using participant identifiers and anonymized after the data collection was completed by deleting the participant identifiers. Participants did not receive monetary compensation, but in each participating police team a lottery was held among the participants that adhered to the measurement protocol, with one wearable awarded per team.

Design

A mixed research design was applied, combining elements of a within-subject design with those of a between-subject design (Figure 1). In this study, 2 teams from 2 separate units of the Dutch police force were recruited. Of the teams in each unit, one team enrolled into the Wearable condition, whereas the other team enrolled into the Wearable+ condition (the between-subject comparison). Before undergoing the respective 5-week interventions, the participants in each team would first partake in a 5-week baseline control period (the within-subject comparison). This approach ensured that we would have a well-comparable control condition. No a priori stratification methods were applied, but χ^2 tests to assess possible between-group differences in gender, age group, and wearable ownership before their participation were performed afterwards.

XSL•FO
RenderX

Figure 1. A mixed design was applied in this study, combining elements of within- and between-subject designs.



Participants

The recruitment of participants was initiated in 2 teams from 2 units of the Dutch police. In each unit, police officers of one team were recruited for the Wearable condition, whereas the police officers of the other team were recruited for the Wearable+ condition. In each team, up to 25 participants were recruited (up to 100 participants in total). This clustered recruitment method ensured geographical distribution of participants while preventing contamination of the different interventions within teams. Due to a lack of response in one team (the Wearable condition team in the rural unit), participants were also recruited from an additional team within that same unit.

In each team, recruitment was done via a combination of posters, flyers, a mention of the study during a team meeting, and an email to all police officers. Interested police officers would use an online tool to check if they met the inclusion criteria, which meant that they were (1) aged 18 years or older, (2) executive police officers, and (3) willing to put aside their own wearable during the duration of the study if they owned one. The police officers then received an information brochure about the study and filled in an online informed consent form before they received their baseline questionnaire and started their participation in the 5-week control period. To stimulate participant adherence during the intervention condition, within each team, a lottery was held that included the participants that collected complete daily data for at least 80% of their participation period and filled in all 5-weekly outcome questionnaires, where the winning participant got to keep the wearable that was used during the study.

Intervention

Control Condition

After filling in their baseline questionnaire, all participants first enrolled in a 5-week control period. Participants were instructed to live their lives as usual but to not use their own wearable (eg, smartwatch) if they owned one.

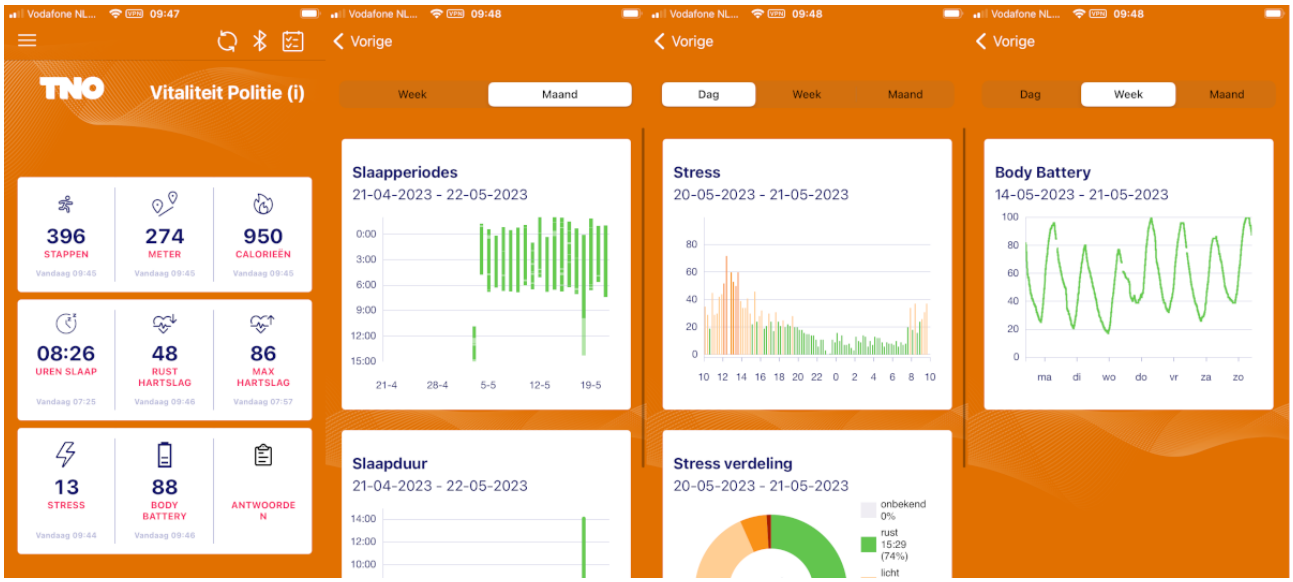
Wearable Condition

After the control period, the participants that enrolled into the Wearable condition received an “off-the-shelf” intervention that was based on just a consumer-available smartwatch. Due to the sensitive profession of the participants in combination with Dutch privacy laws, it was not possible to use a platform that stores data outside the European Union. To accommodate this need, a Garmin Forerunner 255 smartwatch was used in combination with a smartphone app that was approved for use with the Dutch police (built by Sensorium 69 BV), as it directed data directly from the device to self-controlled servers. The app displayed much of the same data as the Garmin Connect app that regular Garmin consumers use (Figure 2). The displayed wearable-based metrics were steps, walking distance, active calories, total sleeping time, resting and maximum heart rate, physiological stress, and a “body battery” metric that estimates the individual’s energy levels. The wearable also measured the daily HRV status and (after 3 weeks) provided feedback on the device itself regarding the extent in which the participant’s resting physiology was in a normal state; if it was not, the device would advise the individual to take it easy that day. Other functions such as sports tracking were also only available on the wearable device itself.

During the 5-week intervention period, Wearable condition participants used the wearable and app at their own discretion. The researchers were available via email and a messaging app to help with technical issues, but no other support was given.

XSL•FO
RenderX

Figure 2. Screenshots of the (Dutch-language) Sensorium smartphone app that used the Garmin Health software development kit to directly retrieve data from the Garmin Forerunner 255. From left-to-right, the screenshots display the home screen dashboard (the version of the Wearable+ condition, which includes answers to the ecological momentary assessment questionnaires on the bottom right corner), a monthly view of sleep periods, a daily view of physiological stress, and a weekly view of the "body battery" metric, which estimates the individual's energy levels.



Wearable+ Condition

The Wearable+ condition represented an enriched situation, where a consumer-available wearable was complemented by additional feedback and support. The composition of the Wearable+ condition was based on a prior (unpublished) study that was done with the Dutch police force, where police officers tested the "off-the-shelf" solution of the current Wearable condition in this study and described how the intervention could be improved to more optimally fit their needs and working context.

Participants enrolled in the 5-week Wearable+ condition used the same Garmin Forerunner 255 that the participants in the regular Wearable condition used. The accompanying custom-built smartphone app was also the same, but it included a feature that allowed the Wearable+ condition participants to also fill in a brief daily EMA questionnaire. The daily EMA questionnaire was available each day during the late afternoon, and participants were instructed to fill it in before they went to bed. The questionnaire contained items on the type of shift that they worked that day, as well as their perceived stress, need for recovery, affective state (valence and arousal), and an optional custom annotation to report incidents or issues that they thought were potentially relevant, where they could write a brief note of up to 35 characters.

The EMA questionnaire data were visible for the participants in their app and were also used for more comprehensive personalized feedback reports that the participants received on a weekly basis. In the clickable weekly feedback report that offered tabs with alternative views and metrics (based on an R Notebook), the data of the subjective EMA outcomes were combined with the objective wearable-based outcomes in order to provide enriched contextual information and compare trends. In addition, textual messages were added to support participants in reflecting on the data. The contextual information could help participants in the interpretation of the wearable-based outcomes, for instance, to better understand why certain scores were abnormally low or high and thus identify potentially actionable associations. When little EMA data was available, the participant was reminded to fill in their daily EMA questionnaire, as it would improve the richness of the feedback report. In the second-to-last week of the intervention period, the report also included a personalized reflection on the data (eg, highlighting relevant trends or associations) and, if possible, advice from the researchers (HJdV, HV, and MvZ) based on the data of the participant. This advice was seen as a proxy for a future algorithm-based function that provides overarching personalized advice based on the broader range of available data sources.

Finally, peer support groups were organized twice per week, where researchers (HV and MvZ) were also available, for instance, to help with the interpretation of the weekly reports. In the peer support groups, participants primarily discussed their personal experiences, for example, with the smartwatch, personal feedback reports, and potential lifestyle changes that they were considering or executing. The personal data of the feedback reports of the participants were not publicly shared by the researchers, but participant questions on the interpretation of certain outcomes were discussed in general terms in order to optimize participants' understanding of the feedback reports. Additionally, participants could contact the researchers via email, telephone, or an online video call on an individual basis for specific questions about their own device or data.

Data Collection

Overview

The intervention effects were assessed via an online questionnaire that was administered at baseline (T0), after the control period (T1), and after the intervention period (T2). For descriptive purposes, questions on the participants' gender, age category (under 30 years old, 30-45 years old, older than 45 years), and whether they already owned a wearable before

XSL•FO
RenderX

participating in the study were included. Wearable ownership was also used as a covariate during hypothesis testing to correct for potential differences in outcome variables among participants that did or did not own a wearable prior to their participation.

Metacognitive Awareness About Sleep, Stress, and Activity Assessment

To determine whether the intervention influenced the participants' stress- and well-being–related awareness, we used the Metacognitive Awareness about Sleep, Stress, and Activity Assessment (MASSAA). This scale was developed in a pilot study in preparation for this study for the Dutch police [52]. Items are scored on a 7-point Likert scale ranging from "completely disagree" to "completely agree." Items fall into 3 categories that are in line with the concept of MCA, namely awareness about responses, causes, and regulation. After performing post hoc factor analysis on the collected data, the responses to the used awareness and cause subscales were found to be very highly interrelated, after which these subscales were combined to minimize the number of performed analyses. Reliability of the subscales was good. Therefore, a total of 6 MASSAA subscales were assessed, namely stress awareness and cause (α=.87), stress regulation (α=.91), sleep awareness and cause (α=.86), sleep regulation (α=.93), physical activity awareness and cause (α=.90), and physical activity regulation (α=.92). The complete MASSAA questionnaire (both the original Dutch items and their English translations) can be found in Multimedia Appendix 1.

Self-Efficacy

Self-efficacy was measured via a series of propositions, which were scored on a 5-point Likert scale that ranges from "completely disagree" to "completely agree," in order to assess if the intervention also affected the individual's perceived capacity to execute behaviors necessary to produce specific outcomes [39] and therefore possibly indirectly affected stress-related outcomes [53]. Separate scales were used. First, a scale for behavior change relating to stress, recovery, sleep and activity was used. This scale was developed in a pilot study in preparation for this study. Reliability was adequate (α=.79). Furthermore, self-efficacy scales with respect to stress resilience (α=.82), coping (α=.83) and tasks (α=.86) were included [54]. For completeness, all used self-efficacy scales and items can be found in Multimedia Appendix 2.

Stress and Well-Being

Several stress- and well-being–related outcomes were also monitored. The participants' perceived need for recovery during and after work were queried using a series of propositions that were scored on a 5-point Likert scale that ranges from "completely disagree" to "completely agree" [55]. Reliability was good, with an α of .79 and .87 for recovery during and after work, respectively. A 3-item subscale for sleep issues regarding waking up too early, sleeping through, and having trouble falling asleep was also included and scored on a 5-point Likert scale, ranging from "never" to "always." Reliability was adequate (α=.74). The subscales for recovery during and after work, as well as for sleep issues, can be found in Multimedia Appendix 2. For the measurement of the emotional, psychological and social well-being of participants, the 14-item Dutch version of the Mental Health Continuum-Short Form (MHC-SF) was administered [56]. The MHC-SF asks how often the respondent experienced a range of well-being–related feelings during the last month and is scored on a 6-point Likert scale that ranges from "never" to "daily." Reliability was good (α=.92). Finally, the 10-item version of the Perceived Stress Scale, 10-item version (PSS-10) was taken [57]. The PSS-10 inquires about the frequency of the occurrence of stress-related feelings and thoughts during the last month, and it is evaluated on a 5-point Likert scale that ranges from "never" to "very often." Reliability was good (α=.78).

Data Analysis

Data management and statistical analyses were performed in RStudio (version 2023.06.2) [58] using R (version 4.2.2; R Foundation for Statistical Computing) [59].

Data Management

The data of participants that used their own wearable during the study or that did not fill in all outcome questionnaires (eg, dropouts or nonresponse) were filtered out. For the description of the participants' gender, age, and wearable ownership prior to the study, as well as the statistical testing for potential between-group baseline differences (via χ^2 tests) and the development of supporting data visualizations, the observed data were used. For hypothesis testing, the data were first standardized at the grand mean [60].

Hypothesis Testing

Hypothesis testing was done using version 1.1 - 33 of the lme4 R package, which allows for statistical testing with linear mixed models (LMMs) [61]. For each of the 15 awareness, self-efficacy, and well-being–related outcomes, modeling was performed using a 3-step process.

For analysis 1, a first model was formed based on solely the confounding variable wearable ownership prior to participating in the study as a fixed effect and the participant identifier as a random effect to account for repeated measures within-subject. Second, the model was extended by including the main variables of interest "time" (a binary label for the repeated measures before and after the control and intervention periods) and "period" (labels for the control and intervention periods). The third and final model was then extended by the interaction effect of time and period, which assesses if the changes over time differed between the control and intervention periods; it is therefore the key focus for answering research question 1. For each model, the marginal R^2 (mR^2) and conditional R^2 (cR^2) [62], as well as the difference (Δ) in each in comparison to the previous step, were described. For the final model, absolute values of Hedges *g* statistic were also calculated to estimate the observed effect size, along with its 95% CI. In the interpretation of Hedges *g*, effects below 0.20 were considered to be "very small," between 0.20 - 0.49 as "small," between 0.50 - 0.79 as "medium," and above 0.8 as "large" [63]. To accommodate the interpretation of the results, a data visualization with interaction plots for each of the 15 outcomes was created.

XSL•FO
RenderX

For analysis 2, this same approach was used, but also included the main effects of the "group" variable (Wearable or Wearable+) in step 2 and the 3-way effect of time, period, and group in step 3. The latter evaluates if the effectiveness of the intervention in comparison to the control condition differed between the Wearable and Wearable+ groups and therefore contributes to answering research question 2.

Results

Overview

A total of 95 participants were recruited (48 Wearable, 47 Wearable+), of which 16 participants (7 Wearable, 9 Wearable+) were excluded from statistical analysis. Of the excluded participants, 12 (3 Wearable, 9 Wearable+) used their own private wearable during the control period, 2 (Wearable) dropped out, and 2 (Wearable) did not fill in the final (T2) questionnaire. The data of the remaining 79 participants—who were predominantly males (n=57, 72%) in the age category of 30-45 years (n=46, 58%) and did not own a wearable before participating in the study (n=47, 59%)—were analyzed. Descriptive statistics of the analyzed participants of both intervention groups and the total sample are described in Table 1. No statistically significant baseline differences between the Wearable and Wearable+ groups in gender, age category, and wearable ownership prior to participation were identified based on χ^2 testing.

Table . Descriptive statistics of the analyzed participants in the Wearable group, Wearable+ group, and total sample.

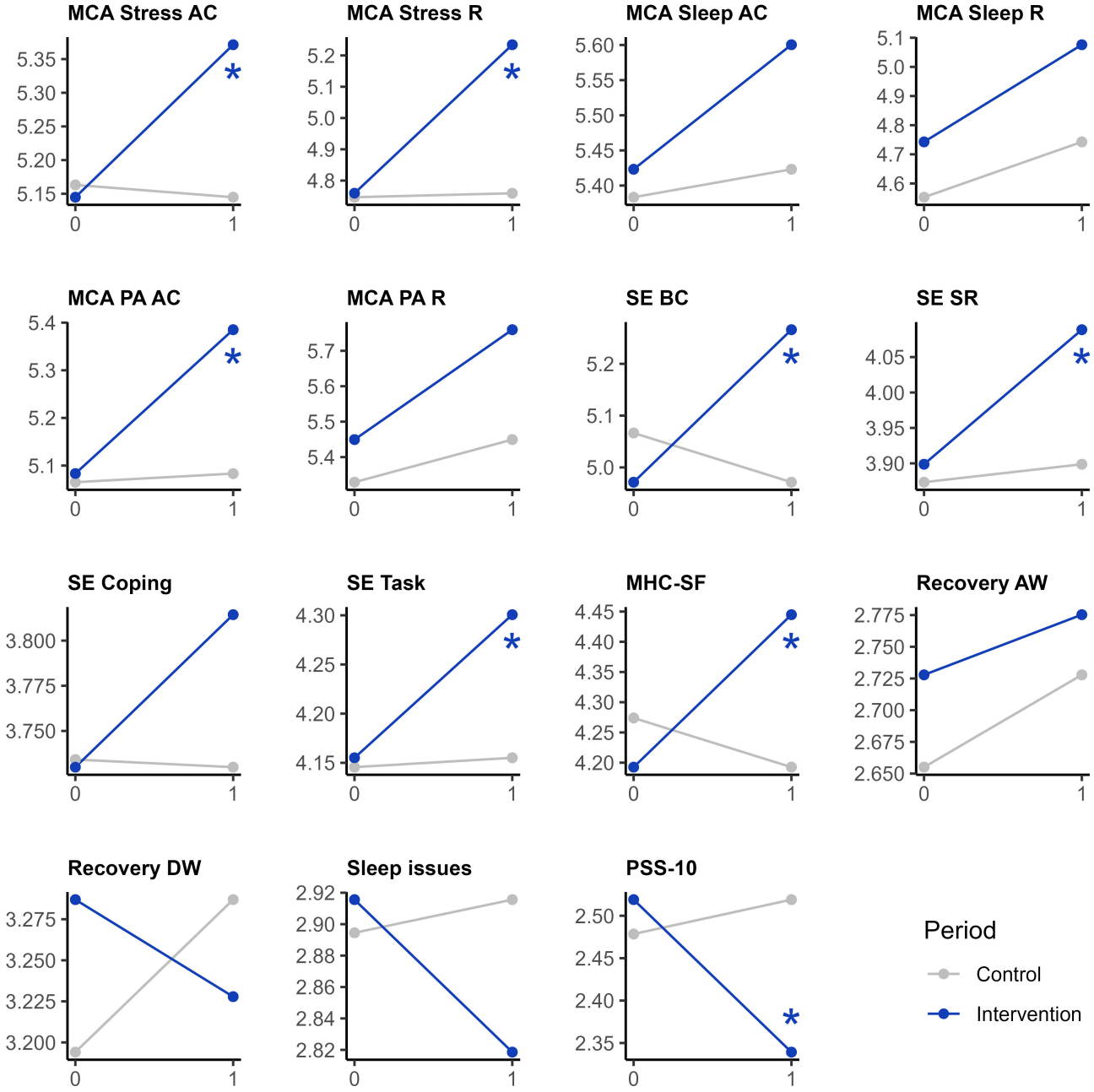
		Wearable, n (%)	Wearable+, n (%)	Total, n (%)
Participants		41 (52)	38 (48)	79 (100)
Gender				
	Male	28 (68)	29 (76)	57 (72)
	Female	13 (32)	9 (24)	22 (28)
Age (years)				
	<30	12 (29)	4 (11)	16 (20)
	30 - 45	21 (51)	25 (66)	46 (58)
	>45	8 (20)	9 (24)	17 (22)
Wearable owner				
	False	27 (66)	20 (53)	47 (59)
	True	14 (34)	18 (47)	32 (41)

Analysis 1: Effect of the Intervention Conditions Against the Control Condition

LMMs were created for each of the 15 awareness-, self-efficacy–, and well-being–related outcomes. Statistically significant 2-way interaction effects were found in 8 outcomes, namely MCA stress awareness and cause (P=.04, mR^2=.02, cR^2=.59, Hedges g=0.30, 95% CI 0.01-0.58), MCA stress regulation (P=.002, mR^2=.02, cR^2=.66, Hedges g=0.41, 95% CI 0.16-0.66), MCA physical activity awareness and cause (P=.03, mR^2=.07, cR^2=.64, Hedges g=0.30, 95% CI 0.04-0.57), self-efficacy behavior change (P=.002, mR^2=.06, cR^2=.73, Hedges g=0.37, 95% CI 0.13-0.60), self-efficacy stress resilience (P=.04, mR^2=.03, cR^2=.71, Hedges g=0.25, 95% CI 0.01-0.50), and self-efficacy task (P=.03, mR^2=.02, cR^2=.68, Hedges g=0.29, 95% CI 0.04-0.54), as well as the MHC-SF (P<.001, mR^2=.02, cR^2=.74, Hedges g=0.40, 95% CI 0.17-0.63) and PSS-10 (P<.001, mR^2=.02, cR^2=.71, Hedges g=0.46, 95% CI 0.22-0.70). The detailed results for the 3-step LMM process for each of the 15 outcomes is available in Multimedia Appendix 3. The interaction plots in Figure 3 summarize these findings by visualizing the average scores before ("0" on x-axis) and after ("1" on x-axis) the control (gray) and intervention (blue) conditions, and labeling the statistically significant interaction effects.

XSL•FO
RenderX

Figure 3. Interaction plots to support the interpretation of the results of analysis 1, which investigated the effectiveness of a wearable-based intervention in comparison to a control condition. For each of the 15 awareness-, self-efficacy–, and well-being–related outcomes, the mean scores before ("0" on x-axis) and after ("1" on x-axis) the control (gray) or intervention condition (blue) are visualized. Interaction plots with statistically significant (P<.05) 2-way interaction effects are annotated (*). AC: awareness and cause; AW: after work; BC: behavior change; DW: during work; MCA: metacognitive awareness; MHC-SF: Mental Health Continuum-Short Form; PA: physical activity; PSS-10: Perceived Stress Scale, 10-item version; R: regulation; SE: self-efficacy; SR: stress resilience.



For each of the 8 statistically significant outcomes, the wearable-based intervention had a favorable effect in comparison to the control condition. For 6 of the remaining 7 nonstatistically significant outcomes, the direction of the effect also favored the intervention, with the exception of recovery during work. Therefore, the observed effects consistently favored the wearable-based intervention over the control period, for which the statistically significant effects were estimated to be "small" (Hedges g=0.25-0.46) in size. In these outcomes, the fixed effects (the control or intervention period) explained between 2% and 7% of the variance, whereas the full model including the random effect (the participant identifier) explained between 59% and 74% of the variance in the outcomes.

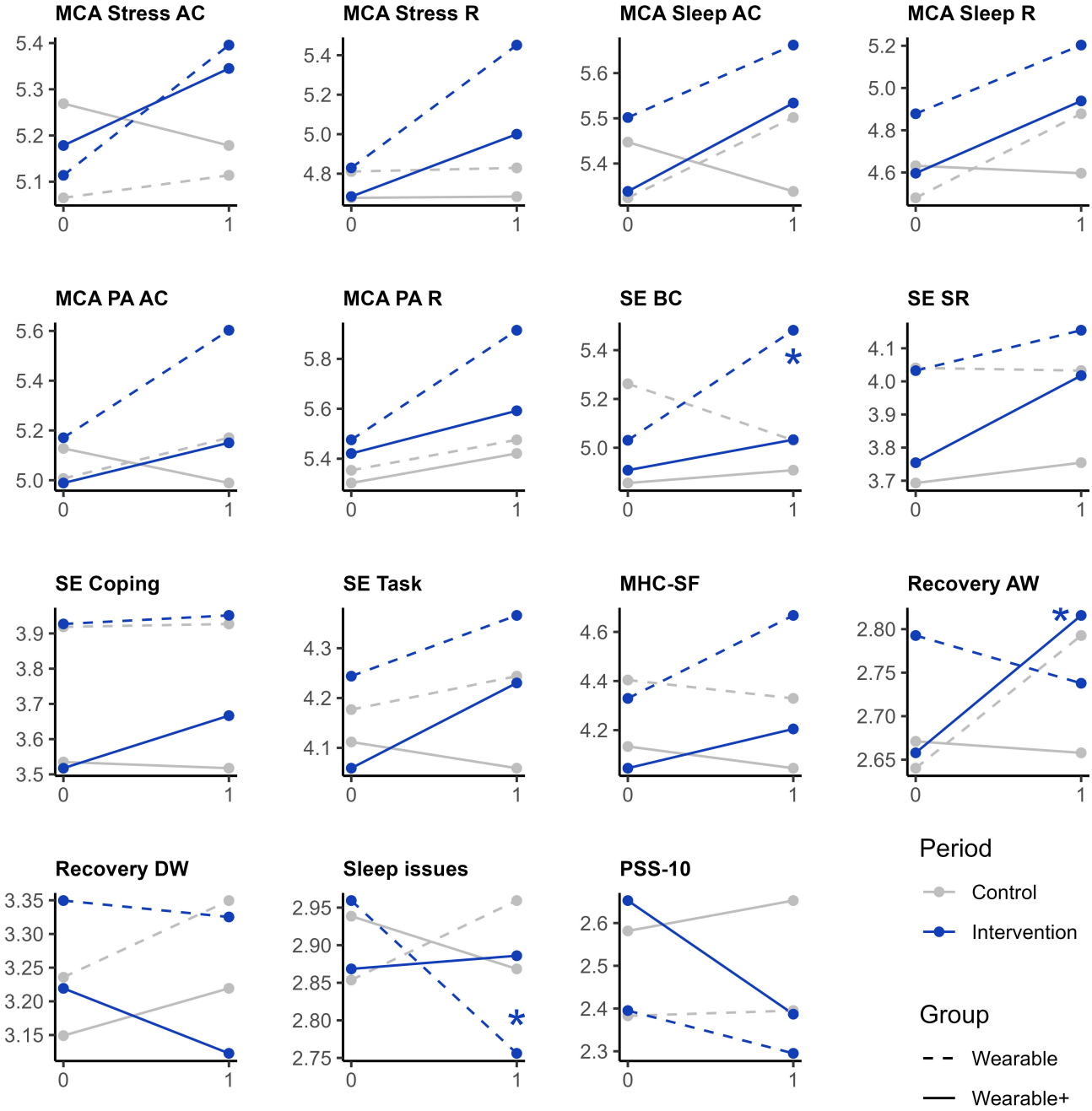
Analysis 2: Effects in the Wearable+ Group Compared to the Wearable Group

In the LMMs that were created for analysis 2, statistically significant 3-way interaction effects were found in 3 of 15 outcomes, namely self-efficacy behavior change (P=.02, mR^2=.10, cR^2=.75, Hedges g=0.58, 95% CI 0.11-1.05), recovery after work (P<.05, mR^2=.01, cR^2=.83, Hedges g=0.41, 95% CI 0.04-0.77), and sleep issues (P=.03, mR^2=.02, cR^2=.77, Hedges g=0.45, 95% CI 0.02-0.88). The detailed results for the 3-step LMM process are available in Multimedia Appendix 4. The interaction plots in Figure 4 summarize these findings by visualizing the average scores before ("0" on x-axis) and after

XSL•FO
RenderX

("1" on x-axis) the control (gray) and intervention (blue) conditions for both the Wearable (striped line) and Wearable+ (solid line) groups and labeling the statistically significant interaction effects.

Figure 4. Interaction plots to support the interpretation of the results of analysis 2, which investigated possible differences in the effectiveness in 2 wearable-based interventions. For each of the 15 awareness-, self-efficacy–, and well-being–related outcomes, the mean scores before ("0" on x-axis) and after ("1" on x-axis) the control (gray) or intervention condition (blue) are visualized, for both the Wearable (striped line) and Wearable+ (solid line) groups. Interaction plots with statistically significant (P<.05) 3-way interaction effects are annotated (*). AC: awareness and cause; AW: after work; BC: behavior change; DW: during work; MCA: metacognitive awareness; MHC-SF: Mental Health Continuum-Short Form; PA: physical activity; PSS-10: Perceived Stress Scale, 10-item version; R: regulation; SE: self-efficacy; SR: stress resilience.



Based on the interaction plots of the outcomes with statistically significant 3-way interaction effects (Figure 4), the enhanced Wearable+ intervention was more effective than the "off-the-shelf" Wearable intervention for improving recovery after work, but less effective for improving self-efficacy in behavior change and perceived sleep issues. However, the plots of these 3 outcomes all showed a noticeable change during the control period, all of which benefited the intervention group, which demonstrated statistically significant improvements in the respective analyses. Therefore, these changes during the control period may have influenced the results for these outcomes. Furthermore, the broad 95% CIs of the estimated effect sizes, compared to those in analysis 1, indicate decreased certainty in the magnitude of the reported effects. The results of analysis 2 should therefore be interpreted more cautiously than the results of analysis 1.

XSL•FO
RenderX

Discussion

Principal Results

The Effectiveness of Wearable-Based Interventions (Analysis 1)

The first aim of this study was to assess whether offering a 5-week wearable-based intervention would improve awareness, self-efficacy, and stress- and well-being–related outcomes in police officers in comparison to a control period. The results of analysis 1 showed that the wearable-based intervention consistently benefited the participants in 8 of the 15 outcomes, namely MCA of stress awareness and cause, MCA of stress regulation, MCA of physical activity awareness and cause, self-efficacy in behavior change, self-efficacy in stress resilience, self-efficacy in task execution, well-being, and stress. This means that the wearable-based intervention had a direct positive influence on the primary stress and well-being outcomes, as well as several aspects of MCA and self-efficacy, which were seen as precursors to relevant behavior change, based on the Health Behavior Model [35,36].

The effect sizes of the 8 statistically significant outcomes were estimated to be "small" (Hedges g=0.25-0.46). The variance that could be explained by the fixed intervention effects only was 2% - 7% (based on the mR^2), whereas the variance that could be explained when also including the random effects of participant identifiers ranged from 59% - 74% (based on the cR^2). This indicates that there are large interindividual differences among participants that can potentially play a role in how individuals respond to the intervention. Future studies with a larger sample size are needed to understand which individual factors influence the differential effectiveness of the intervention and the extent to which personalized approaches could enhance the effectiveness of the intervention, taking into account the variability among participants.

Based on the results of analysis 1, offering a 5-week wearable-based intervention has a small but positive impact on stress- and well-being–related outcomes in police officers. Further studies with a larger sample size are needed to explore if the intervention is particularly effective for certain subgroups and whether more personalized approaches could better accommodate interindividual differences, thereby potentially increasing its overall effectiveness.

The Potential Added Value of Enhanced Feedback and Implementation (Analysis 2)

This study also aimed to investigate if offering an enriched wearable-based intervention that is extended by additional EMA questionnaires, personalized weekly feedback, and peer support meetings would be more effective than offering a more basic "off-the-shelf" wearable-based intervention. Analysis 2 showed mixed results, where the enriched intervention was more effective than the basic intervention at improving recovery after work, less effective for self-efficacy in behavior change and sleep issues, and similarly effective for the other 12 outcomes.

Similar to the results of analysis 1, the estimated effect sizes of the differences between the Wearable and Wearable+ intervention groups were small (Hedges g=0.41-0.45) to medium (0.58), although the much larger confidence intervals in analysis 2 suggest that these results should be interpreted more cautiously. This, combined with the observation that changes in the control period might have exacerbated the between-group effects in the 3 statistically significant outcomes, warrants the conclusion that confirmation in future studies with a larger sample size is desired before policy decisions are made based on these insights. Due to the again observed large differences between the variance that were explained by the effects of the 2 interventions (1% - 10%) and by the full models (75% - 83%), exploration regarding whether certain individuals may benefit more from either intervention and whether additional personalization may increase their overall effectiveness would be interesting.

In view of these findings, the enriched wearable-based intervention was not more effective than the basic wearable-based intervention in its current form. Further research is needed to identify which individuals would particularly benefit from either intervention variation and explore if further personalization could increase their overall effectiveness.

Strengths and Limitations

This research contributes to the literature by assessing the effectiveness of wearable-based interventions for reducing stress and improving well-being in a real-world context. By utilizing consumer-available wearables and investigating the impact of the interventions over several weeks in a real-life setting, where the participants are dealing with true life demands, the current findings can be generalized toward occupational settings, especially highly demanding work settings. Additionally, this study used a nested design that optimized the comparability between the control and intervention periods (ie, for analysis 1). Due to the clustered recruitment strategy and geographical distribution across teams, the potential for spillover effects between the 2 investigated wearable-based interventions was considered to be low (ie, for analysis 2).

A limitation of the study that could have caused a potential underestimation of the true effect sizes via ceiling effects [64] is the inclusion of a relatively healthy population. To at least partially account for this limitation, an LMM approach that was optimized for repeated measures (and therefore accounted for potential baseline differences between participants) was used in favor of using change scores [65]. Another limitation was that the Wearable+ intervention simulated the application of enriched feedback by providing custom-made feedback reports. It is possible that more integrated and continuously available feedback could be more user-friendly and potentially more effective.

Comparison With Prior Work

The results of analysis 1, which showed that offering a wearable-based intervention improved stress- and well-being–related outcomes, are in line with those of prior studies that investigated the effectiveness of wearables in real-life settings. For instance, Ponzo et al [31] reported positive effects on anxiety, depression, and general well-being following a 4-week intervention that used a wearable device (BioBeam)

XSL•FO
RenderX

that measured sleep, physical activity, and heart rate, as well as an app that utilized EMA and relaxation exercises based on a randomized controlled trial in a student population. Another randomized controlled trial, by Smith et al [32], showed favorable effects from a 4-week intervention on stress and anxiety, but used a wearable (Spire Stone) that primarily focuses on optimizing breathing patterns. Other studies on the effectiveness of wearable-based interventions were often performed in laboratory settings, often in young (student) populations [34]. The results of this study therefore expand upon this prior work by exploring the effectiveness of wearable-based interventions in a working population with high job demands and assessing actual changes in lifestyle outcomes and potential precursors of those in a real-life context.

We are not aware of similar prior work comparing the effectiveness of enriched and basic wearable-based interventions as in analysis 2. The difference between our enriched and basic wearable-based intervention can be broken down into the addition of (1) EMA questionnaires, (2) complementary and personalized feedback, and (3) peer support groups. The potential of combining wearable data with EMA questionnaires and complementary feedback is an emerging research topic in this field [66-68], but information about the benefits of EMA and expanded feedback next to wearable data is scarce. Similarly, there are indications that peer support groups can have a positive effect on the recovery of patients with mental health conditions [47-49] and may benefit self-care in patients with medically unexplained symptoms [46], but the effects in a preventive occupational setting are unknown. The lack of prior research on these specific topics highlights the relevance of this study, though based on our results, no clear benefits of these 3 extensions to the basic wearable-based intervention were identified.

Conclusions

Given the evidence obtained from our analyses, offering a 5-week wearable-based intervention to police officers has small but consistent positive effects on their stress, well-being, and related MCA and self-efficacy. The results also showed that there are large interindividual differences in the outcomes that were analyzed, which may be a sign that there is potential for further assessment of possible differences in effectiveness in subgroups, as well as for improvements in the overall effectiveness via more personalized approaches. Complementing the basic “off-the-shelf” wearable-based intervention with additional EMA questionnaires, weekly personalized feedback reports, and peer support groups did not appear to improve the effectiveness of the intervention in its current form. Future research is needed to better understand how the properties of the enriched condition can be improved to increase the intervention’s effectiveness and which participants may or may not benefit from such extended feedback and support.

Acknowledgments

The authors thank the Dutch Police Force for the collaboration, especially the officers who participated or helped us during the recruitment and data collection process. We also thank Sensorium BV for facilitating the data collection via their platform. This study was funded by the Dutch Ministry of Economic Affairs and Climate and the Dutch Ministry of Justice.

Authors' Contributions

RD led the overarching program that this study was part of. WK led this study. All authors contributed to the study design. RD, HV, MvZ, and HJdV assisted in the recruitment of participants. RD, HV, and MvZ facilitated the peer support groups. HJdV created the weekly feedback reports, for which HV and MvZ assisted in the development of personalized advice. RD, WK, and HJdV analyzed the data. HJdV created the data visualizations and wrote the first draft of this article. RD and WK cowrote the final version via feedback and suggested revisions. All coauthors approved the final draft of this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The original Dutch versions and English translations of the items of the Metacognitive Awareness about Sleep, Stress, and Activity Assessment questionnaire. The items fall into 3 categories that are in line with the concept of metacognitive awareness, namely awareness about responses, causes, and regulation. All 32 items are scored on a 7-point Likert scale, ranging from “completely disagree” to “completely agree.”
[DOCX File, 43 KB - mhealth_v13i1e60708_app1.docx]

Multimedia Appendix 2
All other scales that were used besides the Metacognitive Awareness about Sleep, Stress, and Activity Assessment (Multimedia Appendix 1), the Mental Health Continuum Short-Form [56], and the 10-item version of Perceived Stress Scale [57] are described. The basis of and rationale for the use of these scales can be found in the Methods section.
[DOCX File, 40 KB - mhealth_v13i1e60708_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Results of the stepwise performed linear mixed models for analysis 1, where first the confounder "wearable ownership" is included, then the main effects of the variables of interest "time" (repeated measures before and after each period) and "period" (control or intervention), and finally the interaction of time × period, the key aspect for the current analyses.
[DOCX File, 62 KB - mhealth_v13i1e60708_app3.docx]

Multimedia Appendix 4
Results of the stepwise performed linear mixed models for analysis 2, where first the confounder "wearable ownership" is included, then the main effects of the variables of interest "time" (repeated measures before and after each period), "period" (control or intervention) and "group" (Wearable or Wearable+), and finally the interaction of time × period × group, which is the key aspect for these analyses.
[DOCX File, 69 KB - mhealth_v13i1e60708_app4.docx]

References

1. Bevan MP, Priest SJ, Plume RC, Wilson EE. Emergency first responders and professional wellbeing: a qualitative systematic review. Int J Environ Res Public Health 2022 Nov 8;19(22):14649. [doi: 10.3390/ijerph192214649] [Medline: 36429361]
2. Collins PA, Gibbs ACC. Stress in police officers: a study of the origins, prevalence and severity of stress-related symptoms within a county police force. Occup Med 2003 Jun 1;53(4):256-264. [doi: 10.1093/occmed/kqg061]
3. Cheung YK, Li JCM. Predictors, mediators and moderators of police work-related stress: a scoping review. Int J Environ Res Public Health 2023 Jan 27;20(3):2253. [doi: 10.3390/ijerph20032253] [Medline: 36767619]
4. Naushad VA, Bierens JJ, Nishan KP, et al. A systematic review of the impact of disaster on the mental health of medical responders. Prehosp Disaster Med 2019 Dec;34(6):632-643. [doi: 10.1017/S1049023X19004874] [Medline: 31625487]
5. Näslund-Koch C, Vedel-Krogh S, Bojesen SE, Skov L. Plasma vitamin D is not associated with moderate-to-severe psoriasis: results from Danish general population studies. J Invest Dermatol 2023 Oct;143(10):2068-2071. [doi: 10.1016/j.jid.2023.04.004] [Medline: 37121271]
6. Yao L, Xiong Y, Yuan F, Luo Y, Yan L, Li Y. Perceived stress and its impact on the health behavior of Chinese residents during the COVID-19 epidemic: an internet-based cross-sectional survey. Health Sci Rep 2022 Sep;5(5):e778. [doi: 10.1002/hsr2.778] [Medline: 36000081]
7. Magnavita N, Capitanelli I, Garbarino S, Pira E. Work-related stress as a cardiovascular risk factor in police officers: a systematic review of evidence. Int Arch Occup Environ Health 2018 May;91(4):377-389. [doi: 10.1007/s00420-018-1290-y] [Medline: 29344727]
8. Magnavita N, Garbarino S. Is absence related to work stress? A repeated cross‐sectional study on a special police force. Am J Ind Med 2013 Jul;56(7):765-775. [doi: 10.1002/ajim.22155] [Medline: 23334868]
9. Grothe J, Tucker S, Blake A, et al. Exploring first responders' use and perceptions on continuous health and environmental monitoring. Int J Environ Res Public Health 2023 Mar 8;20(6):4787. [doi: 10.3390/ijerph20064787] [Medline: 36981694]
10. Huhn S, Axt M, Gunga HC, et al. The impact of wearable technologies in health research: scoping review. JMIR mHealth uHealth 2022 Jan 25;10(1):e34384. [doi: 10.2196/34384] [Medline: 35076409]
11. Germini F, Noronha N, Borg Debono V, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices. Syst Rev Lit J Med Internet Res 2022 Jan 21;24(1):e30791. [doi: 10.2196/30791]
12. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273. [doi: 10.2196/16273] [Medline: 31778122]
13. Zhang Y, Weaver RG, Armstrong B, Burkart S, Zhang S, Beets MW. Validity of wrist-worn photoplethysmography devices to measure heart rate: a systematic review and meta-analysis. J Sports Sci 2020 Sep;38(17):2021-2034. [doi: 10.1080/02640414.2020.1767348] [Medline: 32552580]
14. Georgiou K, Larentzakis AV, Khamis NN, Alsuhaibani GI, Alaska YA, Giallafos EJ. Can wearable devices accurately measure heart rate variability? A systematic review. Folia Med (Plovdiv) 2018 Mar 1;60(1):7-20. [doi: 10.2478/folmed-2018-0012] [Medline: 29668452]
15. Almarshad MA, Islam MS, Al-Ahmadi S, BaHammam AS. Diagnostic features and potential applications of PPG signal in healthcare: a systematic review. Healthcare (Basel) 2022 Mar 16;10(3):547. [doi: 10.3390/healthcare10030547] [Medline: 35327025]
16. Elgendi M, Fletcher R, Liang Y, et al. The use of photoplethysmography for assessing hypertension. NPJ Digit Med 2019;2:60. [doi: 10.1038/s41746-019-0136-7] [Medline: 31388564]
17. de Vries H, Kamphuis W, van der Schans C, Sanderman R, Oldenhuis H. Trends in daily heart rate variability fluctuations are associated with longitudinal changes in stress and somatisation in police officers. Healthcare (Basel) 2022 Jan 12;10(1):144. [doi: 10.3390/healthcare10010144] [Medline: 35052307]

XSL•FO
RenderX

18. de Vries HJ, Pennings HJM, van der Schans CP, Sanderman R, Oldenhuis HKE, Kamphuis W. Wearable-measured sleep and resting heart rate variability as an outcome of and predictor for subjective stress measures: a multiple n-of-1 observational study. Sensors (Basel) 2022 Dec 28;23(1):332. [doi: 10.3390/s23010332] [Medline: 36616929]
19. Hickey BA, Chalmers T, Newton P, et al. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review. Sensors (Basel) 2021 May 16;21(10):3461. [doi: 10.3390/s21103461] [Medline: 34065620]
20. Morales A, Barbosa M, Morás L, et al. Occupational stress monitoring using biomarkers and smartwatches: a systematic review. Sensors (Basel) 2022 Sep 2;22(17):6633. [doi: 10.3390/s22176633] [Medline: 36081096]
21. Greenwalt CE, Angeles E, Vukovich MD, et al. Pre-sleep feeding, sleep quality, and markers of recovery in division I NCAA female soccer players. J Int Soc Sports Nutr 2023 Dec;20(1):2236055. [doi: 10.1080/15502783.2023.2236055] [Medline: 37470428]
22. Lundstrom EA, De Souza MJ, Koltun KJ, Strock NCA, Canil HN, Williams NI. Wearable technology metrics are associated with energy deficiency and psychological stress in elite swimmers. Int J Sports Sci Coach 2024 Aug;19(4):1578-1587. [doi: 10.1177/17479541231206424]
23. Jerath R, Syam M, Ahmed S. The future of stress management: integration of smartwatches and HRV technology. Sensors (Basel) 2023 Aug 22;23(17):7314. [doi: 10.3390/s23177314] [Medline: 37687769]
24. Schwartz MS, Andrasik F. Biofeedback: A Practitioner's Guide: Guilford Publications; 2017.
25. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
26. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. J Sports Med Phys Fitness 2018 Apr;58(4):534-548. [doi: 10.23736/S0022-4707.17.07437-0] [Medline: 28488834]
27. Goode AP, Hall KS, Batch BC, et al. The impact of interventions that integrate accelerometers on physical activity and weight loss: a systematic review. Ann Behav Med 2017 Feb;51(1):79-93. [doi: 10.1007/s12160-016-9829-1] [Medline: 27565168]
28. Lai MYC, Mong MSA, Cheng LJ, Lau Y. The effect of wearable‐delivered sleep interventions on sleep outcomes among adults: a systematic review and meta‐analysis of randomized controlled trials. Nurs Health Sci 2023 Mar;25(1):44-62. [doi: 10.1111/nhs.13011] [Medline: 36572659]
29. Namvari M, Lipoth J, Knight S, et al. Photoplethysmography enabled wearable devices and stress detection: a scoping review. J Pers Med 2022 Oct 31;12(11):1792. [doi: 10.3390/jpm12111792] [Medline: 36579537]
30. Gedam S, Paul S. A review on mental stress detection using wearable sensors and machine learning techniques. IEEE Access 2021;9:84045-84066. [doi: 10.1109/ACCESS.2021.3085502]
31. Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D. Efficacy of the digital therapeutic mobile app BioBase to reduce stress and improve mental well-being among university students: randomized controlled trial. JMIR mHealth uHealth 2020 Apr 6;8(4):e17767. [doi: 10.2196/17767] [Medline: 31926063]
32. Smith EN, Santoro E, Moraveji N, Susi M, Crum AJ. Integrating wearables in stress management interventions: promising evidence from a randomized trial. Int J Stress Manag 2020;27(2):172-182. [doi: 10.1037/str0000137]
33. Kocielnik R, Sidorova N. Personalized stress management: enabling stress monitoring with LifelogExplorer. Künstl Intell 2015 Jun;29(2):115-122. [doi: 10.1007/s13218-015-0348-1]
34. González Ramírez ML, García Vázquez JP, Rodríguez MD, Padilla-López LA, Galindo-Aldana GM, Cuevas-González D. Wearables for stress management: a scoping review. Healthcare (Basel) 2023 Aug 22;11(17):2369. [doi: 10.3390/healthcare11172369] [Medline: 37685403]
35. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
36. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
37. Schraw G, Dennison RS. Assessing metacognitive awareness. Contemp Educ Psychol 1994 Oct;19(4):460-475. [doi: 10.1006/ceps.1994.1033]
38. Delahaij R, van Dam K. Coping style development: the role of learning goal orientation and metacognitive awareness. Pers Individ Dif 2016 Apr;92:57-62. [doi: 10.1016/j.paid.2015.12.012]
39. Bandura A. Self-efficacy mechanism in human agency. Am Psychol 1982;37(2):122-147. [doi: 10.1037//0003-066X.37.2.122]
40. Delahaij R, van Dam K, Gaillard AWK, Soeters J. Predicting performance under acute stress: the role of individual characteristics. Int J Stress Manag 2011;18(1):49-66. [doi: 10.1037/a0020891]
41. Edward KL, Garvey L, Aziz Rahman M. Wearable activity trackers and health awareness: nursing implications. Int J Nurs Sci 2020 Apr 10;7(2):179-183. [doi: 10.1016/j.ijnss.2020.03.006] [Medline: 32685614]
42. Park Y, Kim SR, So HY, et al. Effect of mobile health intervention for self-management on self-efficacy, motor and non-motor symptoms, self-management, and quality of life in people with Parkinson's disease: randomized controlled trial. Geriatr Nurs (Lond) 2022 Jul;46:90-97. [doi: 10.1016/j.gerinurse.2022.05.003]

XSL•FO
RenderX

43. Wulfovich S, Fiordelli M, Rivas H, Concepcion W, Wac K. “I must try harder”: design implications for mobile apps and wearables contributing to self-efficacy of patients with chronic conditions. Front Psychol 2019;10:2388. [doi: 10.3389/fpsyg.2019.02388] [Medline: 31749733]
44. Singh N, Stewart RAH, Benatar JR. Intensity and duration of lifestyle interventions for long-term weight loss and association with mortality: a meta-analysis of randomised trials. BMJ Open 2019 Aug 18;9(8):e029966. [doi: 10.1136/bmjopen-2019-029966] [Medline: 31427335]
45. Gromatsky M, Sullivan SR, Spears AP, et al. Ecological momentary assessment (EMA) of mental health outcomes in veterans and servicemembers: a scoping review. Psychiatry Res 2020 Oct;292:113359. [doi: 10.1016/j.psychres.2020.113359] [Medline: 32777594]
46. Nebeker C, Weisberg B, Hekler E, Kurisu M. Using self-study and peer-to-peer support to change “sick” care to “health” care: the patient perspective. Front Digit Health 2020;2:2. [doi: 10.3389/fdgth.2020.00002] [Medline: 34713016]
47. Smit D, Miguel C, Vrijsen JN, Groeneweg B, Spijker J, Cuijpers P. The effectiveness of peer support for individuals with mental illness: systematic review and meta-analysis. Psychol Med 2023 Aug;53(11):5332-5341. [doi: 10.1017/S0033291722002422] [Medline: 36066104]
48. Lyons N, Cooper C, Lloyd-Evans B. A systematic review and meta-analysis of group peer support interventions for people experiencing mental health conditions. BMC Psychiatry 2021 Jun 23;21(1):315. [doi: 10.1186/s12888-021-03321-z] [Medline: 34162340]
49. White S, Foster R, Marks J, et al. The effectiveness of one-to-one peer support in mental health services: a systematic review and meta-analysis. BMC Psychiatry 2020 Nov 11;20(1):534. [doi: 10.1186/s12888-020-02923-3] [Medline: 33176729]
50. Tan SY, Chou CC. Supervision effects on self-efficacy, competency, and job involvement of school counsellors. J Psychol Couns Sch 2018 Jun;28(1):18-32. [doi: 10.1017/jgc.2017.19]
51. Lohani G, Sharma P. Effect of clinical supervision on self-awareness and self-efficacy of psychotherapists and counselors: a systematic review. Psychol Serv 2023 May;20(2):291-299. [doi: 10.1037/ser0000693] [Medline: 35913852]
52. Delahaij R, Kamphuis W. Measuring the effects of wearable use on metacognitive awareness and self efficacy for behavioral change [Poster presentation]. Presented at: European Association of Occupational Health Psychology; Jun 5-7, 2024; Granada, Spain.
53. Eikenhout LMJ, Delahaij R, Van Dam K, Kamphuis W, Hulshof IL, Van Ruysseveldt J. Chronic stressors and burnout in Dutch police officers: two studies into the complex role of coping self-efficacy. Front Psychol 2022;13:1054053. [doi: 10.3389/fpsyg.2022.1054053] [Medline: 36591030]
54. Kamphuis W, Delahaij R, Preenen P. Validatie weerbaarheidsmonitor politie–basis politie zorg. Validation of the resilience monitor for the police–basic police teams [Report in Dutch]. : Soesterberg, Netherlands: TNO; 2014 TNO 2014 M11035.
55. Demerouti E, Bakker AB, Sonnentag S, Fullagar CJ. Work‐related flow and energy at work and at home: a study on the role of daily recovery. J Organ Behav 2012 Feb;33(2):276-295 [FREE Full text] [doi: 10.1002/job.760]
56. Luijten CC, Kuppens S, van de Bongardt D, Nieboer AP. Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF) in Dutch adolescents. Health Qual Life Outcomes 2019 Oct 22;17(1):157. [doi: 10.1186/s12955-019-1221-y] [Medline: 31640806]
57. Lee EH. Review of the psychometric evidence of the Perceived Stress Scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127. [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
58. RStudio Team. RStudio: integrated development environment for R. RStudio, BPC. 2023. URL: https://www.rstudio.com [accessed 2025-01-13]
59. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2023. URL: https://www.R-project.org/ [accessed 2025-01-13]
60. Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. Psychol Methods 2007 Jun;12(2):121-138. [doi: 10.1037/1082-989X.12.2.121] [Medline: 17563168]
61. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015 Oct 7;67(1):1-48. [doi: 10.18637/jss.v067.i01]
62. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 2013;4(2):133-142. [doi: 10.1111/j.2041-210x.2012.00261.x]
63. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Academic Press; 2013.
64. Šimkovic M, Träuble B. Robustness of statistical methods when measure is affected by ceiling and/or floor effect. PLoS One 2019;14(8):e0220889. [doi: 10.1371/journal.pone.0220889] [Medline: 31425561]
65. Thomas DR, Zumbo BD. Difference scores from the point of view of reliability and repeated-measures ANOVA: in defense of difference scores for data analysis. Educ Psychol Meas 2012;72(1):37-43. [doi: 10.1177/0013164411409929]
66. Colombo D, Fernández-Álvarez J, Patané A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019 Apr 5;8(4):465. [doi: 10.3390/jcm8040465] [Medline: 30959828]
67. Kim J, Marcusson-Clavertz D, Yoshiuchi K, Smyth JM. Potential benefits of integrating ecological momentary assessment data into mHealth care systems. Biopsychosoc Med 2019;13:19. [doi: 10.1186/s13030-019-0160-5] [Medline: 31413726]

XSL•FO
RenderX

68. Seppälä J, De Vita I, Jämsä T, et al. Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: systematic review. JMIR Ment Health 2019 Feb 20;6(2):e9819. [doi: 10.2196/mental.9819] [Medline: 30785404]

Abbreviations

cR^2: conditional R^2
EMA: ecological momentary assessment
HRV: heart rate variability
LMM: linear mixed model
MASSAA: Metacognitive Awareness about Sleep, Stress, and Activity Assessment
MCA: metacognitive awareness
MHC-SF: Mental Health Continuum-Short Form
mR^2: marginal R^2
PSS-10: Perceived Stress Scale, 10-item version

Edited by L Buis; submitted 19.05.24; peer-reviewed by ML González Ramírez, T Chitra; revised version received 15.08.24; accepted 14.10.24; published 28.01.25.

Please cite as:
de Vries HJ, Delahaij R, van Zwieten M, Verhoef H, Kamphuis W
The Effects of Self-Monitoring Using a Smartwatch and Smartphone App on Stress Awareness, Self-Efficacy, and Well-Being–Related Outcomes in Police Officers: Longitudinal Mixed Design Study
JMIR Mhealth Uhealth 2025;13:e60708
URL: https://mhealth.jmir.org/2025/1/e60708
doi:10.2196/60708

© Herman Jaap de Vries, Roos Delahaij, Marianne van Zwieten, Helen Verhoef, Wim Kamphuis. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Mobile App–Based Gratitude Intervention's Effect on Mental Well-Being in University Students: Randomized Controlled Trial

Chloë Fuller[1], BA; Silvia Marin-Dragu[2,3], MSc; Ravishankar Subramani Iyer[4], BSc; Sandra Melanie Meier[2,4], PhD

[1]Department of Psychology, Saint Mary's University, Halifax, NS, Canada

[2]IWK Health Centre Department of Psychiatry & Specific Care Clinics, Department of Psychiatry, Dalhousie University, 5850/5980 University Ave, Halifax, NS, Canada

[3]Department of Psychology and Neuroscience, Dalhousie University, Halifax, NS, Canada

[4]Department of Computer Science, Dalhousie University, Halifax, NS, Canada

Corresponding Author:
Sandra Melanie Meier, PhD
IWK Health Centre Department of Psychiatry & Specific Care Clinics, Department of Psychiatry, Dalhousie University, 5850/5980 University Ave, Halifax, NS, Canada

Abstract

Background: Gratitude interventions are used to cultivate a sense of gratitude for life and others. There have been mixed results of the efficacy of gratitude interventions' effect on psychological well-being with a variety of populations and methodologies.

Objectives: The objective of our study was to test the effectiveness of a gratitude intervention smartphone app on university students' psychological well-being.

Methods: We used a randomized experimental design to test our objective. Participants were recruited undergraduate students from a web-based university study recruitment system. Participants completed 90 web-based survey questions on their emotional well-being and personality traits at the beginning and end of the 3-week research period. Their depression, anxiety, and stress levels were measured with the Depression, Anxiety, and Stress Scale (DASS-21). After the baseline survey, participants were randomly assigned to either the control or the intervention. Participants in the intervention group used both a fully automated mobile sensing app and a gratitude intervention mobile iOS smartphone app designed for youth users and based on previous gratitude interventions and exercises. The gratitude intervention app prompted users to complete daily gratitude exercises on the app including a gratitude journal, a gratitude photo book, an imagine exercise, a speech exercise, and meditation. Participants in the control group used only the mobile sensing app, which passively collected smartphone sensory data on mobility, screen time, sleep, and social interactions.

Results: A total of 120 participants met the inclusion criteria, and 27 were lost to follow-up for a total of 41 participants in the intervention group and 52 in the control group providing complete data. Based on clinical cutoffs from the baseline assessment, 56 out of 120 participants were identified as being in a subsample with at least moderate baseline symptomatology. Participants in the subsample with at least moderate baseline symptomatology reported significantly lower symptoms of depression, anxiety, and stress postintervention (Cohen d=−0.68; P=.04) but not in the full sample with low baseline symptomatology (Cohen d=0.16; P=.46). The number of times the app was accessed was not correlated with changes in either the subsample (r=0.01; P=.98) or the full sample (r=−0.04; P=.79).

Conclusions: University students experiencing moderate to severe distress can benefit from a gratitude intervention smartphone app to improve symptoms of depression, anxiety, and stress. The number of times the gratitude intervention app was used is not related to well-being outcomes. Clinicians could look at incorporating gratitude apps with other mental health treatments or for those waitlisted as a cost-effective and minimally guided option for university students experiencing psychological distress.

Trial Registration: Clinicaltrials.gov NCT06621745; http://clinicaltrials.gov/ct2/show/NCT06621745

(JMIR Mhealth Uhealth 2025;13:e53850) doi:10.2196/53850

KEYWORDS

gratitude intervention; smartphone app; gratitude exercises; psychological well-being; mobile phone

XSL•FO
RenderX

Introduction

Background

Gratitude is the awareness of and appreciation for the positive aspects of life and the goodness of others, resulting in positive affect [1,2]. Gratitude has been shown to be an important resource for promoting and protecting mental well-being. Cross-sectional research has consistently reported that more grateful individuals are less depressed, anxious, and stressed than less grateful individuals [3]. Longitudinal studies in first-year undergraduate students further provided evidence that gratitude leads to improved well-being, with students higher in gratitude becoming less stressed, less depressed, and having higher perceived social support over the first term [4]. Gratitude is theorized to improve mental well-being through the mechanisms of positive affect (the positive emotion of gratitude has inherently a protective effect on various mental health problems) [1], positive outlook (grateful individuals' cognitive schemas prompt them to interpret situations more positively) [5], coping strategies (grateful individuals make more positive coping appraisals) [6], and positive relationships (grateful individuals have broader skill repertoires helping them cultivate additional resources for support) [7]. In consequence, gratitude interventions have been developed to foster a sense of appreciation for life and others, yielding successful results and clinical relevance [1]. Gratitude interventions commonly consist of listing things to be grateful for, grateful contemplation, and behavioral expressions of gratitude [1]. Some exercises included in these interventions can be gratitude journals, gratitude letters, and practicing meditation and mindfulness. Gratitude journals prompt people to write about experiences that they are grateful for every day [8]. Gratitude letters are written expressions of gratitude to someone and their impact on the person's life [9]. Gratitude interventions are valuable as they are easy to understand and complete. Furthermore, they can be done in various settings such as at home or virtually by a wide range of age groups [10]. They are flexible in their application and can be completed alongside other forms of psychotherapy or alone. In addition, they are generally enjoyed by those who engage in them [10]. Therefore, the accessibility of gratitude interventions makes them easy to integrate for widespread use by various populations.

Effectiveness of Gratitude Interventions

Previous literature has suggested that gratitude interventions may help improve mental well-being by increasing gratitude, life satisfaction, and self-esteem while simultaneously decreasing negative affect, depression, anxiety, and stress. However, results have been mixed with effects varying by outcome, duration, follow-up length, format, and age. The first qualitative review investigating the efficacy of gratitude interventions included 12 studies and found that while a gratitude condition showed significant positive effects on mental well-being compared with a condition writing about daily hassles, no effects were observed in comparison with other control conditions [1]. Davis et al [10] re-evaluated the efficacy of gratitude interventions in 32 samples and found that gratitude interventions outperformed measurement-only control conditions for psychological well-being, but not for gratitude, suggesting overall weak efficacy of gratitude interventions compared with measurement-only controls. Renshaw and Olinger Steeves [11] examined gratitude interventions' effect on youth by analyzing 6 studies and found that gratitude exercises to promote youth's well-being and decrease distress were effective only for select outcomes and generally ineffective overall. Dickens [12] expanded on previous work with a series of meta-analyses on the effectiveness of gratitude interventions on unpooled outcome variables both postintervention and at a delayed follow-up. The results showed significant differences between control and intervention, providing evidence that gratitude interventions can positively impact people's well-being, affect, and gratitude, as well as decreased depressive symptoms, with small to medium effect sizes both immediately and long term [12]. A recent systematic review on controlled trials of positive psychology digital interventions for youth and young adults found that these interventions enhance gratitude, hope, compassion, positive coping, and body image, with medium to large effects [13]. However, the results of a more recent meta-analysis including 27 studies focusing more specifically on symptoms of depression and anxiety suggest that there is a limited effect of gratitude interventions on reducing symptoms of depression and anxiety at postintervention and at follow-up [3].

Web-Based and Mobile-Based Gratitude Interventions

Several recent studies have begun to explore the use of web-based gratitude interventions and gratitude apps. Web-based interventions can be delivered on a computer, smartphone, or other electronic devices, and are advantageous in that they are accessible, readily available, flexible in their use, and cost-effective; furthermore, they can be implemented into daily life, allow for monitoring engagement, and can encourage and improve consistent use [14-16]. A few studies have examined the effectiveness of gratitude apps on mental health in younger children [17], high school students [18], and adults [19]. Specifically, a study using a 6-week gratitude exercise app intervention found that intervention participants had significantly improved mental health outcomes compared with the control group, with effects continuing after the 6-week follow-up, which were not dependent on baseline well-being [19].

Focus on Youth

Overall, previous literature indicates that gratitude interventions can effectively improve participants' well-being. However, to our knowledge, no prior studies have conducted web-based gratitude interventions with a university-aged population. Furthermore, gratitude interventions have primarily focused on journaling and gratitude letters as forms of gratitude exercise. This study aims to address these gaps in the literature and expand the usage of gratitude interventions to a university student population, using diverse gratitude exercise modalities. University students experience higher stress levels than the rest of the population, making them an important population to provide mental health interventions [20]. Furthermore, youth mental health has been impacted by the COVID-19 pandemic, and university students have shown increased levels of depression, anxiety, and stress, as well as reported academic, health, and lifestyle concerns due to the pandemic [21]. Given

XSL•FO
RenderX

the strong link between university students' poor mental health [22] and the risk of low academic functioning and dropout [23], gratitude interventions might help mitigate not only mental health issues but also academic failure. Web-based interventions have proven to be generally effective, used, and liked by university students [24]. University students are less likely to seek help for psychological distress, but web-based interventions have been found to be a popular alternative to providing help and promoting well-being among students [25]. This is likely because web-based interventions are flexible, and university students can use them anywhere on their own time, which is advantageous with busy schedules or time constraints. Accordingly, university students, as an especially vulnerable population, might greatly benefit from a gratitude intervention app (GIA). If proven effective, the GIA could configure an easily accessible tool for university students to improve their mental well-being and academic performance.

Research Question

Our study provides a 3-week gratitude intervention delivered through a smartphone app with 5 gratitude exercises to university students aged 18‐30 years to examine the intervention's effects on depression, anxiety, and stress. This study will allow for a better understanding of a web-based GIA's effect on psychological functioning and inform the use of this intervention for young adults and university students to support their psychological well-being.

Hypotheses

We hypothesize the following:

1. Depression, anxiety, and stress symptoms will decrease from baseline to after the 3-week period in the group that practices the gratitude intervention tasks (intervention group) compared with the group that does not (control group).
2. The number of times the participants use the GIA will correlate with clinical symptomatology changes.

As many of the previous studies have focused on the general population with effectiveness depending on symptom severity [18], we explored our hypotheses in the full sample of university students and in a subsample of those who showed at least moderate symptomatology. Besides effectiveness, this study also investigated changes in potential mechanisms such as negative affectivity and social inhibition.

Methods

Participants

Participants were university students from Dalhousie University in Nova Scotia, Canada. They were recruited on the web through the Dalhousie SONA Experimental Participation System, a cloud-based research and participant management solution for universities. A priori power analyses were conducted using G*Power 5.0. A total sample size of 92 was estimated to be needed to achieve an 80% chance of detecting a small to medium effect size (d=0.25) at an α value of .05 for comparing the 2 groups at 2 time points. Anticipating approximately 30% attrition, we aimed to enroll a minimum of 120 participants.

The inclusion criteria were that participants must have been between ages of 18 and 30 years, spoke English, had access to a mobile phone with an iOS, had mobile phone and web-based literacy, and were not receiving treatment for any mental health condition. Participants also could not have any history of cardiovascular, endocrine, or kidney disease; tumor; hypertension; coarctation of the aorta; fibromuscular dysplasia; or sleep apnea to minimize possible confounding factors. A total of 137 participants met this inclusion criteria, and of those, 87.6% (120/137) enrolled and completed the baseline assessment.

Design

This study used a randomized experimental design to test the effectiveness of a gratitude intervention smartphone app. Participants were asked to answer 90 questions in a web-based survey on emotional well-being and personality traits at the beginning and end of the 3-week research period. The intervention group used the GIA, and the Predicting Risk and Outcomes of Social Interactions (PROSIT) app, which is a mobile sensing app (for more details, see the study by MacLeod et al [26]). All participants used the PROSIT app in the background of their mobile devices throughout the 3-week study period. In addition, those allocated to the intervention group practiced gratitude intervention tasks during the same 3-week period using the GIA app.

Procedure

The study took place between November 2021 and April 2023. Participants were recruited on the web through SONA Experimental Participation System at Dalhousie University. Individuals interested in registering for the study could sign up through SONA systems, which provided a link to the web-based consent form and study information on REDCap (Research Electronic Data Capture) [27]. REDCap is a free and secure application for web-based data collection approved by the Nova Scotia Health Authority and the Dalhousie University. Eligible participants who completed the consent form were sent a REDCap baseline assessment form on emotional well-being and personality traits through email. After completing the baseline measures, participants were randomly allocated to either the gratitude intervention group or the control group and were emailed instructions for downloading and using the iOS apps. Participants were randomized with a 1:1 allocation in blocks of 4 using a computer-generated algorithm. Randomization was conducted by a researcher who was not otherwise involved in the study. Data analysts had no interactions with participants. Participants in the gratitude intervention group downloaded and used both the PROSIT and GIA apps, while participants in the control group downloaded and used only the PROSIT app. Participants were instructed to complete the prompted gratitude exercise on the GIA app daily. After 3 weeks of consistent use of the apps, participants logged out of the apps and completed the outcome assessment form sent by email on REDCap. To improve adherence, participants using the GIA app received daily notifications via the app as well as a weekly reminder via email. Upon completion of the study, participants were further reimbursed through the assignment of US $21 Amazon gift card or 3 SONA points,

XSL•FO
RenderX

which can be used for 3% extra credit toward a chosen university course final grade. Participants were also encouraged to complete questionnaires regardless of app usage.

Gratitude Intervention App

The GIA app is an iOS smartphone app specifically designed for this study. The images of the GIA app can be found in Figures S1 and S2 in Multimedia Appendix 1. It is based on previous effective gratitude interventions and exercises [1,8,9], and its design and structure were specifically adapted for youth users. Users accessed it by logging in with an individualized username and password. The app consists of 5 different gratitude exercises: a gratitude journal, a photo book, an imagine exercise, a speech exercise, and meditation. Gratitude journals have been successfully used in previous gratitude intervention studies to encourage users to write a journal entry about what they are grateful for that day [8,26]. The photo book asks users to upload photographs of 5 things they are grateful for, such as people, things, or activities in their life, and encourages users to be mindful throughout the next week of other things they are grateful for to photograph and upload. Expressing gratitude through photography has previously been shown to improve well-being and decrease stress [26]. The imagine exercise asks users to take time to imagine 5 things they are most grateful for in their life and to explain the details as to why they are grateful for the listed thing and the aspects that make them happy. The speaking-out-loud exercise asks users to think about 5 things they are grateful for and talk about them for 90 seconds each, recording with the app. The meditation exercise prompts users to listen to a gratitude-based guided meditation recording and follow its instructions for being grateful in the moment and relaxing. Meditation has been used in previous gratitude interventions with significant effects on well-being and gratitude [26,28].

Every day during the 3-week study period, users received a notification from the app prompting them to complete the exercise for that day. The app is green and white in color, with an icon featured on the home and login pages of a flower with a heart rate design. Technical support was available through email for installation, usage, or login issues.

Measurements

Participants' demographic information was recorded at baseline via the secure REDCap survey application. Participants were asked to provide their age, biological sex, gender, sexual orientation, level of education, and race. Questionnaires measuring the outcome variables of interest (anxiety, depression, stress level, negative affectivity, and social inhibition) were assessed in the baseline and outcome assessment forms before and after the 3-week research period. They were automatically sent to participants via REDCap. PROSIT and GIA app usage (ie, registration date, frequency of accessing the app, and date of last visit) was logged.

Outcome

Participant's depression, anxiety, and stress levels were measured with the Depression, Anxiety, and Stress Scale (DASS-21) [29], which is a set of 3 self-report scales each consisting of 7 items to measure the individual's emotional states of depression ("I couldn't seem to experience any positive feeling at all"), anxiety ("I felt I was close to panic"), and stress ("I found it difficult to relax"). Participants rated how much each statement applied to them in the past week, from 0 ("Did not apply to me at all") to 3 ("Applied to me very much or most of the time"). High scores indicate higher levels of symptoms of anxiety, depression, and stress. There are subscale score cutoffs for normal, mild, moderate, severe, and extremely severe emotional states. Participants were considered to show at least moderate symptomatology if they scored higher than 7 on the depression subscale, 6 on the anxiety subscale, and 10 on the stress subscale. The DASS-21 has shown good psychometric properties, with strong internal consistency in both total scoring and scale set scoring [30,31], and has been found to have adequate construct validity [29,31]. In the current sample, the reliability of DASS-21 was good with Cronbach α values of 0.89, 0.82, and 0.76 for the subscales of depression, anxiety, and stress, respectively.

Mechanisms

Negative affectivity and social inhibition were measured with the Type D Personality Scale (DS14). The DS14 contains 14 items and consists of 2 subscales: negative affectivity ("I often make a fuss of unimportant things") and social inhibition ("I find it hard to start a conversation"). Participants answered the items from 0 ("False') to 4 ("True"), resulting in maximum scores of 28 for negative affectivity and social inhibition. The DS14 has shown good internal consistency for both negative affectivity and social inhibition [32]. In the current sample, the reliability of DS14 was good with Cronbach α values of 0.84 and 0.86 for the subscales of negative affectivity and social inhibition, respectively. All measures were completed on the web.

Data Analysis

All statistical analyses were completed using R programming language (R version 4.2.2) and RStudio (version 2023.03.0+ Build 386). Descriptive statistics were computed for demographics and outcome measures at baseline in the full sample of students and a subsample of those who showed at least moderate symptomatology. Two-tailed independent sample *t* tests and χ^2 tests were used to test for any differences between the gratitude intervention group and the control group at baseline. We further compared participants who dropped out with those who completed the study on demographics, and baseline outcome measures, using independent sample *t* tests and χ^2 tests.

A longitudinal analysis of covariance, using linear mixed modeling, was conducted to test the effects of treatment on outcome and mechanism measures (DASS-21 and DS14) to test the effectiveness of the intervention. Treatment effects were tested by entering 2-way interactions between time and group assignment. An α value of .05 was used to determine statistical significance. Cohen *d* effect sizes were derived based on recommendations for linear mixed models [32], where 0.20, 0.50, and 0.80 were interpreted as indicating small, medium, and large effects, respectively [33]. An intent-to-treat approach was used, consistent with CONSORT guidelines (ie, all participants who were randomized to receive enrollment

XSL•FO
RenderX

information for the intervention were included in analyses) [34]. Missing data were handled using maximum likelihood estimation [35]. To assess possible dose–response relationships in the gratitude intervention group, Spearman correlations were calculated between the number of times the app was used and changes in clinical symptomatology as measured by the DASS-21. To explore potential differences in technology engagement, a linear regression model compared the daily usage of the PROSIT app across groups.

Ethical Considerations

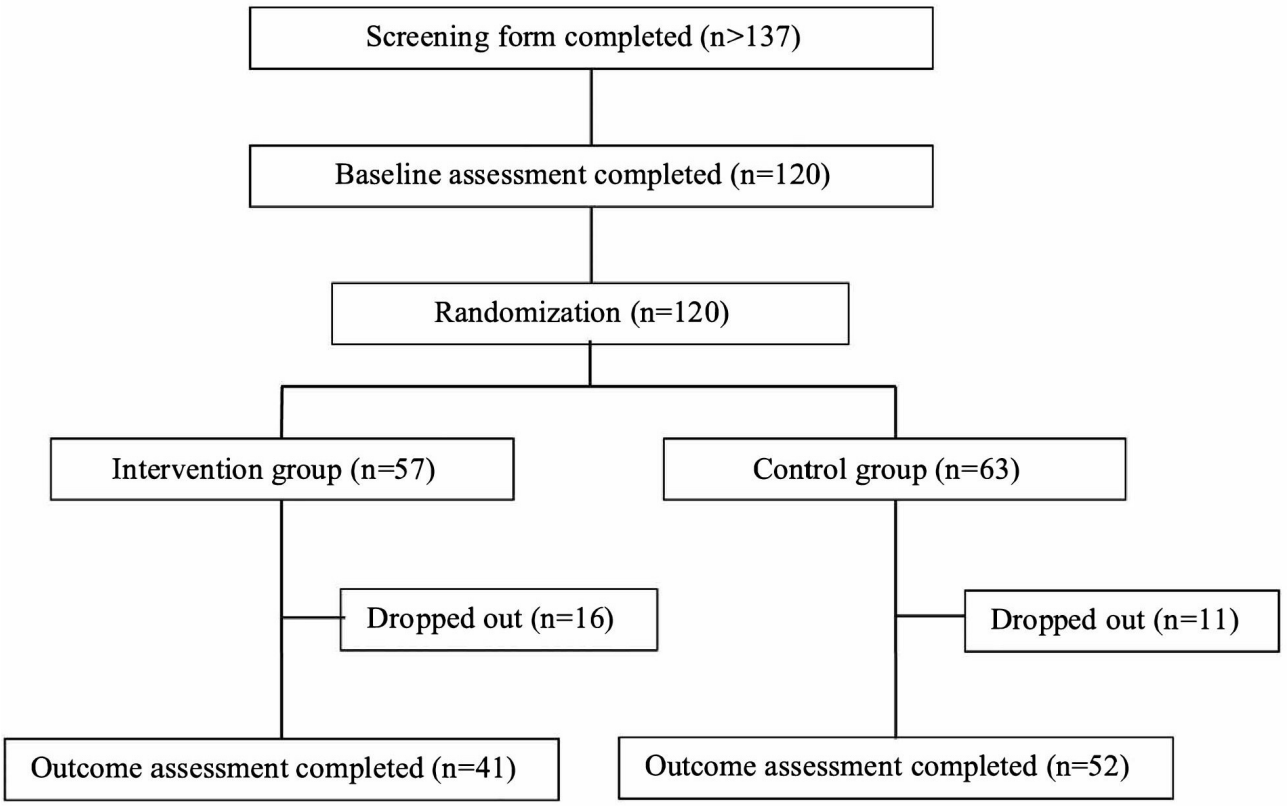
The study was approved by the research ethics board committee of Dalhousie University in Halifax, Nova Scotia, Canada (REB#2021-5460). Before participation, all study participants were required to sign a primary informed consent form, thereby confirming their engagement in the survey process, app usage, future secondary data analysis, and the ability to opt out of the study without reason. The original consent covered secondary data analysis without additional consent. All study data were deidentified. Participants who completed the study were reimbursed with a US $21 Amazon gift card.

Results

Recruitment and Participant Characteristics

Participant enrollment, allocation, and retention information are provided in line with the CONSORT (Consolidated Standards of Reporting Trails) flow diagram (Figure 1). Recruitment ran from November 11, 2021, to March 15, 2023. More than 137 individuals completed the eligibility screener for the study. Out of the 137 individuals who met the inclusion criteria, 87.6% (120/137) were enrolled and completed the baseline assessment. In the full sample, 57 participants were randomized to the gratitude intervention group and 63 to the control group of whom 31 and 25 displayed at least moderately significant symptomatology. Of those participants randomized to the gratitude intervention group, 77.2% (44/57) started and used the app in the full sample and 67.7% (21/31) in the subsample. Follow-ups were completed 3 weeks later. In the full sample, 28.1% (16/57) of gratitude intervention participants and 23.8% (15/63) of control participants were lost to follow-up. In the subsample 35.5% (11/31) of gratitude intervention participants and 24.0% (6/25) control participants were lost. These differences were not statistically significant (P=.59 in the full sample and P=.35 in the subsample).

Figure 1. CONSORT (Consolidated Standards of Reporting Trails) flow diagram of the study.



Participant sociodemographic characteristics and baseline outcome measures are presented for the intervention group and the control group for both the full sample and the subsample showing at least moderate symptomatology (Multimedia Appendix 2). Participants were on average 19.8 years of age and the large majority (102/120, 85%) identified as female. Across both samples, most participants (117/120, 98% in the full sample and 54/56, 96% in the subsample) had completed some form of postsecondary education and most (86/117, 73% and 39/56, 70%) identified as being of European descent. Mean levels of depression, anxiety, and stress symptoms as measured by the DASS-21 were below the established clinical cutoffs at baseline for the full sample but not for the subsample. After randomization, there was no difference between the gratitude intervention group and the control group for the full sample and

XSL•FO
RenderX

the subsample in any baseline sociodemographic characteristics and outcome measures.

Dropout and Adherence

No statistically significant differences were observed in baseline characteristics between participants who completed the follow-up assessment and those lost to attrition for the full sample and the subsample. Furthermore, participants in the gratitude intervention group were no more likely to drop out than participants in the control group.

Log data showed that most participants (44/57, 77%) in the full sample randomized to the gratitude intervention group accessed the app. Participants opened the app on average 18 times (SD 11.04) over 9 days (SD 6.90). Participants in the subsample showing at least moderate symptomatology used the app slightly less, which was not statistically significant (P=.34); specifically, 68% (21/31) of participants accessed the app on average 16 times (SD 11.27) over 9 days (SD 7.73).

Treatment Effects

The results of the mixed model analyses are shown in Table 1. There were no adverse events related to study participation. The results indicated a statistically significant medium treatment effect (ie, time × group interaction) for mental health symptoms (Cohen d=−0.68, 95% CI −1.34 to −0.03; P=.04) in the subsample showing at least moderate symptomatology but not in the full sample (d=0.16, 95% CI −0.26 to 0.58; P=.46). In the subsample, the gratitude intervention group scored significantly lower on depression, anxiety, and stress as measured by the DASS-21. The number of times the app was used was not correlated with changes in clinical symptomatology in the subsample with at least moderate symptomatology (r=0.01; P=.98) or in the full sample (r=−0.04; P=.79). The results indicated a statistically significant effect (ie, time × group interaction) for negative affectivity (d=−0.68, 95% CI −1.34 to −0.01; P=.046) but not for social inhibition (d=−0.61, 95% CI −1.27 to 0.05; P=.07) in the subsample showing at least moderate symptomatology. The gratitude intervention group scored significantly lower on negative affectivity and trendwise for social inhibition. No effects were observed in the full sample (negative affectivity: d=−0.18, 95% CI −0.60 to 0.24, P=.41; social inhibition: d=−0.18, 95% CI −0.60 to 0.24, P=.41). No differences in technology engagement as measured by the PROSIT app were observed in the gratitude intervention group compared with the control group in the subsample with at least moderate symptomatology (t_{42}=0.56; P=.58) or in the full sample (t_{100}=1.20; P=.23).

Table . Results of mixed model analyses.

	Value	SE	*t* value (*df*)	*P* value
Intercept	54.06	8.40	6.433 (54)	<.001
Baseline	8.76	8.39	1.045 (37)	.30
Treatment	−1.74	5.15	−0.338 (54)	.74
Baseline × treatment	−10.92	5.25	−2.081 (37)	.04

Discussion

Summary of Key Findings

Previous studies have examined the efficacy of gratitude interventions as a resource for psychological well-being with mixed results [1,10,12,14,18,36]. This study aimed to examine the effectiveness of a 3-week GIA on Canadian university students' symptoms of depression, anxiety, and stress. We hypothesized that (1) initial depression, anxiety, and stress symptoms would decrease after the 3-week gratitude intervention period compared with the control group, in both the full sample and the subsample with at least moderate symptomatology, and (2) the number of times the participants used the GIA would be correlated with changes in clinical symptomatology in both the full sample and the moderate symptomatology subsample.

Our first hypothesis was partially supported. The mobile gratitude intervention was found to effectively improve mental health symptoms in the subsample showing at least moderate symptomatology, with those in the intervention group scoring significantly lower for symptoms of depression, anxiety, and stress after the 3-week intervention period than those in the subsample control group. However, these results were not found in the sample consisting of participants of any mental health symptomatology (full sample). These results indicate that university students with moderate and severe distress benefited from the intervention and highlight the suitability of gratitude interventions for this population. In contrast, our second hypothesis was not supported in both the subsample and the full sample as there was no correlation observed between the number of times the GIA was used and the changes in depression, anxiety, and stress symptoms. Interestingly, no differences between the gratitude intervention group and the control group were observed in technology engagement as measured using the PROSIT app. Thus, differences in motivation and comfortability in using technology are unlikely to explain the observed changes in mental health symptoms in the intervention group.

Comparison With Other Research

The current gratitude intervention had a moderate effect size and exceeded previous effects found in meta-analyses of gratitude interventions [10,12,14]. The intervention had a similar effect size and ability to improve mental health outcomes as other recently conducted web-based gratitude interventions [18,37-39], although the current intervention examined here was probably less intense than those used in previous trials. This is of special interest as no additional guidance was provided in the current intervention, besides email-based technical support. Guidance can improve effectiveness [40,41] but it

XSL•FO
RenderX

simultaneously reduces the feasibility of scaling the intervention into population-wide implementation. The current findings could indicate that guidance may be less necessary in highly educated digital natives such as university students resulting in higher scalability. Furthermore, our study revealed a similar effect size although it used an active control group (the mobile sensing app) as the comparison, instead of an inactive waitlist control group like previous trials, which can frequently lead to an overestimation of intervention effects [42].

The current gratitude intervention's effect on mental health outcomes was dependent on participants' baseline depression, anxiety, and stress, which contrasts previous findings [18]. Differences in the sample baseline well-being could explain this dissimilarity, as Bono et al [18] conducted their gratitude intervention with older participants of whom less than 20% experienced moderate symptoms of depression or anxiety due to the pandemic. This is in contrast to this study, in which nearly 50% of university students displayed moderate symptomatology. As such, the current findings show that gratitude interventions might be helpful not only for health promotion and illness prevention purposes but also for those who are clinically distressed thereby advancing current knowledge. Furthermore, the number of times the participants used the GIA was unrelated to well-being outcomes after the 3-week study period, contrasting previous studies [18,37]. This could be explained by the shorter intervention time frame compared with previous studies.

Besides symptomatology, the intervention also showed an effect on negative affectivity. These results are in line with previous work suggesting positive emotion as a potential underlying mechanism to explain the effect of gratitude on mental health [1,7,43]. In addition, a nonsignificant effect of the intervention on social inhibition was observed. Social support has previously been suggested as a process of change in gratitude interventions' effects on mental well-being [1,2,7,37,43]. Yet other processes of change such as cognitive schemata and coping strategies were not assessed in this study [7,18,21,37,43] . Furthermore, this study describes only changes in potential mechanisms and thus does not establish them as mediating factors or their temporal precedence.

While this study did not include participants already receiving mental health treatment, future studies, interventions, and mental health treatments could look to incorporating gratitude interventions with other mental health treatments. The potential additive effect of gratitude interventions could be explored in combination with cognitive behavioral therapy or medication treatments. They could be offered as an interim approach for individuals on a waitlist as using the app may either reduce their symptoms or prevent them from getting worse. Gratitude interventions could also be used to maintain treatment success after discharge. Moreover, as the app targets gratitude rather than mental health problems, it may also help reduce the impact of stigma, which can prevent individuals from seeking support when distressed, and potentially increase engagement. Finally, because this intervention effectively improved mental health outcomes in university students with at least moderate baseline symptomatology, university health clinics could look to including mobile gratitude interventions as a cost-effective, minimally guided, and independent option for the promotion of student mental well-being, allowing more students to receive the help they are seeking. Specifically, students interested in self-driven mental well-being support may be encouraged to use gratitude interventions by counselors and other health care professionals. Gratitude could also be implemented in general university-led wellness programs.

Strengths and Limitations

This is the first study to examine the effectiveness of a GIA in improving the mental well-being of university students in North America. Compared with previous gratitude interventions, the present intervention was also minimally guided in its delivery to participants [37], making it flexibly and independently used. The app also had various gratitude exercise options, allowing for different modalities of gratitude intervention on 1 interface.

Another strength of the gratitude intervention was that the mobile intervention app automatically recorded when participants used the exercises, providing data on when the app was used and allowing for uniquely objective insight into intervention utilization. Finally, another strength of the study was the exclusion of individuals already receiving mental health treatment, allowing for internal validity of the effectiveness of the intervention.

This study had several limitations. First, the study had a small sample size with a relatively homogenous sample population. The participants were predominantly Canadian, and female, and all were students. Thus, the study findings might not be generalizable to males. However, females are more vulnerable to depression and anxiety than males and are more likely to seek mental health treatment [42-44], making this form of intervention potentially more valuable and relevant to the female population. Furthermore, as previously discussed, university students are a population susceptible to stress and mental health issues [21], and web-based mental health interventions are effective and liked by this population [23,45]. However, future studies would benefit from examining the effects of mobile GIAs on less educated and more diverse samples by adapting different recruitment strategies (eg, by advertising in sports clubs or public transportation). Second, study dropout in the gratitude intervention group was relatively high and somewhat higher than in the control group. While this is a common finding in web-based intervention studies in university student populations (intervention group completion rates: 43% - 66%), this might limit the interpretation of results [46-48]. Reported trends could be partially related to specific participant characteristics (eg, having a stronger motivation for gratitude interventions) by the participants in the sample that adhered to the intervention. In addition, the app format was not compared with other forms of delivery and the effects of certain app features or persuasive elements were not tested. Further experiments are needed to unravel the specific effects of delivery features [49]. Third, there was no long-term follow-up on well-being outcomes after the intervention and follow-up occurred immediately only postintervention. Therefore, we are unable to conclude the long-term effects of mobile gratitude interventions in this population. Fourth, we investigated the gratitude intervention's effect on the potential mechanisms of

XSL•FO
RenderX

negative affectivity and social inhibition. Yet, this study did not have the power to examine them as mediators. In addition, more proximal mediators need to be considered for their effect on these more distal mechanisms. Future research should assess and test even more proximal mediators in serial multiple mediation models to further understand how the gratitude intervention works.

Conclusions

This study is one of the first gratitude interventions delivered through an app to be used by a university student population. It demonstrates the effectiveness of a GIA in improving university students' mental well-being by decreasing depression, anxiety, and stress symptoms after a 3-week intervention period. This study also found that app usage was not related to mental health outcomes, and intervention effectiveness was dependent on the participant's baseline symptomatology. This app is valuable in its ability to be used virtually and independently, its potential to be used in conjunction with other mental health treatments, and its cost-effectiveness. Future studies should examine this intervention's effectiveness in other populations and explore the unique impacts of different gratitude exercise modalities.

Acknowledgments

The authors would like to thank Marcus Cormier and Abiaz Hossain (both of Dalhousie University) for providing a critical review of this manuscript. This work was partly funded by the Canadian Institute of Health Research (CIHR) Canadian Research Chairs (CRC) stipend (award 1024586) and the Canada Foundation for Innovation (CFI) John R. Evans Leaders Fund (JELF) (grant 38835).

Data Availability

The data used in this study are confidential.

Authors' Contributions

CF contributed to conceptualization, methodology, statistical analysis, writing of the original draft, and writing—review and editing; SMD contributed to the development of mobile app and writing—review and editing; RSI contributed to the development of mobile app, statistical analysis, and writing—review and editing; and SMM contributed to conceptualization, methodology, statistical analysis, writing—review and editing—project administration, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The gratitude intervention app.
[DOCX File, 426 KB - mhealth_v13i1e53850_app1.docx]

Multimedia Appendix 2
Sociodemographic characteristics and outcome measures at baseline.
[DOCX File, 17 KB - mhealth_v13i1e53850_app2.docx]

Checklist 1
CONSORT-EHEALTH (V 1.6.1).
[PDF File, 434 KB - mhealth_v13i1e53850_app3.pdf]

References

1. Wood AM, Froh JJ, Geraghty AWA. Gratitude and well-being: a review and theoretical integration. Clin Psychol Rev 2010 Nov;30(7):890-905. [doi: 10.1016/j.cpr.2010.03.005] [Medline: 20451313]
2. Sansone RA, Sansone LA. Gratitude and well being: the benefits of appreciation. Psychiatry (Edgmont) 2010 Nov;7(11):18-22. [Medline: 21191529]
3. Cregg DR, Cheavens JS. Gratitude interventions: effective self-help? A meta-analysis of the impact on symptoms of depression and anxiety. J Happiness Stud 2021 Jan;22(1):413-445. [doi: 10.1007/s10902-020-00236-6]
4. Wood AM, Maltby J, Gillett R, Linley PA, Joseph S. The role of gratitude in the development of social support, stress, and depression: two longitudinal studies. J Res Pers 2008 Aug;42(4):854-871. [doi: 10.1016/j.jrp.2007.11.003]
5. Watkins PC, Woodward K, Stone T, Kolts RL. Gratitude and happiness: development of a measure of gratitude, and relationships with subjective well-being. Soc Behav Pers 2003 Jan 1;31(5):431-451. [doi: 10.2224/sbp.2003.31.5.431]

XSL•FO
RenderX

6. Algoe SB, Zhaoyang R. Positive psychology in context: effects of expressing gratitude in ongoing relationships depend on perceptions of enactor responsiveness. J Posit Psychol 2016;11(4):399-415. [doi: 10.1080/17439760.2015.1117131] [Medline: 27800009]
7. Lambert NM, Fincham FD, Stillman TF. Gratitude and depressive symptoms: the role of positive reframing and positive emotion. Cogn Emot 2012;26(4):615-633. [doi: 10.1080/02699931.2011.595393] [Medline: 21923564]
8. Bohlmeijer E, Westerhof G. The model for sustainable mental health: future directions for integrating positive psychology into mental health care. Front Psychol 2021;12:747999. [doi: 10.3389/fpsyg.2021.747999] [Medline: 34744925]
9. Emmons RA, McCullough ME. Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. J Pers Soc Psychol 2003;84(2):377-389. [doi: 10.1037//0022-3514.84.2.377]
10. Davis DE, Choe E, Meyers J, et al. Thankful for the little things: a meta-analysis of gratitude interventions. J Couns Psychol 2016 Jan;63(1):20-31. [doi: 10.1037/cou0000107] [Medline: 26575348]
11. Renshaw TL, Olinger Steeves RM. What good is gratitude in youth and schools? A systemic review and meta-analysis of correlates and intervention outcomes. Psychol Sch 2016 Mar;53(3):286-305. [doi: 10.1002/pits.21903]
12. Dickens LR. Using gratitude to promote positive change: a series of meta-analyses investigating the effectiveness of gratitude interventions. Basic Appl Soc Psych 2017 Jul 4;39(4):193-208. [doi: 10.1080/01973533.2017.1323638]
13. Saboor S, Medina A, Marciano L. Application of positive psychology in digital interventions for children, adolescents, and young adults: systematic review and meta-analysis of controlled trials. JMIR Ment Health 2024 Aug 14;11:e56045. [doi: 10.2196/56045] [Medline: 39141906]
14. Conley CS, Durlak JA, Shapiro JB, Kirsch AC, Zahniser E. A meta-analysis of the impact of universal and indicated preventive technology-delivered interventions for higher education students. Prev Sci 2016 Aug;17(6):659-678. [doi: 10.1007/s11121-016-0662-3] [Medline: 27225631]
15. Bidargaddi N, Almirall D, Murphy S, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123. [doi: 10.2196/10123] [Medline: 30497999]
16. Mohr DC, Siddique J, Ho J, Duffecy J, Jin L, Fokuo JK. Interest in behavioral and psychological treatments delivered face-to-face, by telephone, and by internet. Ann Behav Med 2010 Aug;40(1):89-98. [doi: 10.1007/s12160-010-9203-7] [Medline: 20652466]
17. Lau N, O'Daffer A, Colt S, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798. [doi: 10.2196/17798] [Medline: 32357125]
18. Bono G, Mangan S, Fauteux M, Sender J. A new approach to gratitude interventions in high schools that supports student wellbeing. J Posit Psychol 2020 Sep 2;15(5):657-665. [doi: 10.1080/17439760.2020.1789712]
19. Kloos N, Austin J, van 't Klooster JW, Drossaert C, Bohlmeijer E. Appreciating the good things in life during the Covid-19 pandemic: a randomized controlled trial and evaluation of a gratitude app. J Happiness Stud 2022;23(8):4001-4025. [doi: 10.1007/s10902-022-00586-3] [Medline: 36245700]
20. Ribeiro Í, Pereira R, Freire IV, de Oliveira BG, Casotti CA, Boery EN. Stress and quality of life among university students: a systematic literature review. Health Prof Educ 2018 Jun;4(2):70-77. [doi: 10.1016/j.hpe.2017.03.002]
21. Wang X, Hegde S, Son C, Keller B, Smith A, Sasangohar F. Investigating mental health of US college students during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2020 Sep 17;22(9):e22817. [doi: 10.2196/22817] [Medline: 32897868]
22. Mental health. ScienceDirect. URL: https://www.sciencedirect.com/topics/psychology/mental-health [accessed 2025-01-06]
23. Becker TD, Torous JB. Recent developments in digital mental health interventions for college and university students. Curr Treat Options Psych 2019 Sep;6(3):210-220. [doi: 10.1007/s40501-019-00178-8]
24. Bruffaerts R, Mortier P, Auerbach RP, et al. Lifetime and 12‐month treatment for mental disorders and suicidal thoughts and behaviors among first year college students. Int J Methods Psych Res 2019 Jun;28(2):e1764. [doi: 10.1002/mpr.1764]
25. Ryan ML, Shochet IM, Stallman HM. Universal online interventions might engage psychologically distressed university students who are unlikely to seek formal help. Adv Ment Health 2010 Aug;9(1):73-83. [doi: 10.5172/jamh.9.1.73]
26. MacLeod L, Suruliraj B, Gall D, et al. A mobile sensing app to monitor youth mental health: observational pilot study. JMIR Mhealth Uhealth 2021 Oct 26;9(10):e20638. [doi: 10.2196/20638] [Medline: 34698650]
27. O'Leary K, Dockray S. The effects of two novel gratitude and mindfulness interventions on well-being. J Altern Complement Med 2015 Apr;21(4):243-245. [doi: 10.1089/acm.2014.0119] [Medline: 25826108]
28. Esquivel HF. Decreasing stress by expressing gratitude through photography [Dissertation]. : Notre Dame de Namur University; 2023 Sep 1 URL: https://www.proquest.com/openview/0b2d49c72e5d2e7b279c17efd96ea7cc/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2025-01-06]
29. Denollet J. DS14: standard assessment of negative affectivity, social inhibition, and Type D personality. Psychosom Med 2005;67(1):89-97. [doi: 10.1097/01.psy.0000149256.81953.49] [Medline: 15673629]
30. Duthely LM, Nunn SG, Avella JT. A novel heart‐centered, gratitude‐meditation intervention to increase well‐being among adolescents. Educ Res Int 2017 Jan;2017(1):1-12. [doi: 10.1155/2017/4891892]

XSL•FO
RenderX

31. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010]
32. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
33. Osman A, Wong JL, Bagge CL, Freedenthal S, Gutierrez PM, Lozano G. The Depression Anxiety Stress Scales-21 (DASS-21): further examination of dimensions, scale reliability, and correlates. J Clin Psychol 2012 Dec;68(12):1322-1338. [doi: 10.1002/jclp.21908] [Medline: 22930477]
34. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
35. Ng F, Trauer T, Dodd S, Callaly T, Campbell S, Berk M. The validity of the 21-item version of the Depression Anxiety Stress Scales as a routine clinical outcome measure. Acta Neuropsychiatr 2007 Oct;19(5):304-310. [doi: 10.1111/j.1601-5215.2007.00217.x] [Medline: 26952943]
36. Feingold A. A regression framework for effect size assessments in longitudinal modeling of group differences. Rev Gen Psychol 2013 Mar;17(1):111-121. [doi: 10.1037/a0030048] [Medline: 23956615]
37. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: L. Erlbaum Associates; 1988.
38. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
39. Kwok OM, Underhill AT, Berry JW, Luo W, Elliott TR, Yoon M. Analyzing longitudinal data with multilevel models: an example with individuals living with lower extremity intra-articular fractures. Rehabil Psychol 2008 Aug;53(3):370-386. [doi: 10.1037/a0012765] [Medline: 19649151]
40. Adair KC, Rodriguez-Homs LG, Masoud S, Mosca PJ, Sexton JB. Gratitude at work: prospective cohort study of a web-based, single-exposure well-being intervention for health care workers. J Med Internet Res 2020 May 14;22(5):e15562. [doi: 10.2196/15562] [Medline: 32406864]
41. Kirca A, M. Malouff J, Meynadier J. The effect of expressed gratitude interventions on psychological wellbeing: a meta-analysis of randomised controlled studies. Int J Appl Posit Psychol 2023;PMID(1):63-86. [doi: 10.1007/s41042-023-00086-6]
42. Tudge JRH, Freitas LBL, O'Brien LT. The virtue of gratitude: a developmental and cultural approach. Hum Dev 2015;58(4-5):281-300. [doi: 10.1159/000444308]
43. Heckendorf H, Lehr D, Ebert DD, Freund H. Efficacy of an internet and app-based gratitude intervention in reducing repetitive negative thinking and mechanisms of change in the intervention's effect on anxiety and depression: results from a randomized controlled trial. Behav Res Ther 2019 Aug;119:103415. [doi: 10.1016/j.brat.2019.103415] [Medline: 31202003]
44. Srivastava M, Ghosh A. Online gratitude interventions during COVID-19 pandemic on youth: randomized trial. J Evid Based Soc Work 2023 Mar 4;20(2):272-298. [doi: 10.1080/26408066.2022.2157691]
45. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208. [doi: 10.2196/17208] [Medline: 32196462]
46. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions—a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
47. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
48. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146. [doi: 10.2196/jmir.3077] [Medline: 24905070]
49. Alonso J, Angermeyer MC, Bernert S, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand 2004 Jun;109(s420):21-27. [doi: 10.1111/j.1600-0047.2004.00327.x]

Abbreviations

DASS-21: Depression, Anxiety, and Stress Scale
DS14: Type D Personality Scale
GIA: gratitude intervention app
PROSIT: Predicting Risk and Outcomes of Social Interactions
REDCap: Research Electronic Data Capture

XSL•FO
RenderX

Edited by L Buis; submitted 23.10.23; peer-reviewed by M Pritchard, YH Lee; revised version received 02.10.24; accepted 19.11.24; published 14.01.25.

Please cite as:
Fuller C, Marin-Dragu S, Iyer RS, Meier SM
A Mobile App–Based Gratitude Intervention's Effect on Mental Well-Being in University Students: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e53850
URL: https://mhealth.jmir.org/2025/1/e53850
doi:10.2196/53850

© Chloë Fuller, Silvia Marin-Dragu, Ravishankar Subramani Iyer, Sandra Melanie Meier. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smoking Cessation Smartphone App for Nondaily Smoking With Telephone Onboarding: Proof-of-Concept Randomized Controlled Trial

Bettina B Hoeppner[1,2], MS, PhD; Kaitlyn R Siegel[1], BA; Allison E Futter[1], BS; Diadora Finley-Abboud[1], BS; Alivia C Williamson[1], BA; Christopher W Kahler[3], PhD; Elyse R Park[2,4], MPH, PhD; Susanne S Hoeppner[1,2], MApStat, PhD

[1]Health through Flourishing (HtF) program, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

[3]Center for Alcohol and Addiction Studies, Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, United States

[4]Health Promotion and Resiliency Intervention Research Center, Mongan Institute/Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Bettina B Hoeppner, MS, PhD
Health through Flourishing (HtF) program
Department of Psychiatry
Massachusetts General Hospital
125 Nashua St, 4th Fl.
Boston, MA, 02114
United States
Phone: 1 617 643 198
Email: bhoeppner@mgh.harvard.edu

Abstract

Background: Nondaily smoking is a widespread and increasingly prevalent pattern of use. To date, no effective treatment approach for nondaily smoking has been identified.

Objective: This study aimed to conduct an unblinded randomized controlled trial to evaluate proof-of-concept markers of the Smiling instead of Smoking (SiS) app, a smartphone app for smoking cessation, designed specifically for people who smoke less than daily, within the framework of positive psychology.

Methods: Overall, 226 adults who smoke less than daily were recruited on the web and asked to undertake a quit attempt while using assigned smoking cessation support materials. Participants were randomly assigned to 1 of 3 materials: the SiS smartphone app, the National Cancer Institute's smartphone app QuitGuide (QG), or the National Cancer Institute's smoking cessation brochure, "Clearing the Air" (CtA). All participants engaged in a 15-minute scripted onboarding phone call and were introduced to their support materials to use for the next 7 weeks. Follow-up self-assessment web surveys were sent 2, 6, 12, and 24 weeks after participants' initially chosen quit date (ie, 1 week after onboarding). The primary outcome for this study was self-efficacy to remain abstinent from smoking at treatment end. Secondary outcomes assessed treatment acceptability, treatment feasibility (eg, number of days of app use, time spent using app, and use of smoking cessation strategies), and secondary proof-of-concept efficacy outcomes (eg, positive affect, craving, and attitudes toward smoking). Smoking outcomes (ie, 30-day point prevalence abstinence and smoking reduction) were also assessed.

Results: Results indicated a significant effect of treatment on the primary outcome, where SiS participants (n=80) reported higher self-efficacy to abstain from smoking at the end of treatment than the 2 control groups (QG: n=75; *P*=.02; Cohen *d*=0.40 and CtA: n=71; *P*=.007; Cohen *d*=0.50). This effect was also significant on both self-efficacy subscales (ie, internal cues and external cues) with effect sizes ranging from Cohen *d*=0.34 to 0.50 across the pairwise comparisons. The SiS app group also reported lower craving (QG: *P*=.005; Cohen *d*=–0.57 and CtA: *P*=.005; Cohen *d*=–0.57) and higher positive affect than QG (QG: *P*=.01; Cohen *d*=0.44 and CtA: *P*=.05; Cohen *d*=0.38); attitudes toward smoking were largely similar across groups. Treatment acceptability was comparable across groups (*P* values for all groups >.05; Cohen *d* range 0.06-0.23). Treatment feasibility measures

XSL•FO
RenderX

indicated that participants used the SiS app on 33 out of 49 days, for 35 to 40 minutes per week, resulting in greater use of smoking cessation strategies than QG (QG: *P*=.04; Cohen *d*=0.38 and CtA: *P*=.16; Cohen *d*=0.24).

Conclusions: These findings provide strong evidence for the conceptual underpinnings of the SiS app, and thereby provide compelling justification for conducting a large-scale randomized controlled trial that can test the effectiveness of the SiS app on smoking cessation.

Trial Registration: ClinicalTrials.gov NCT04672239; https://clinicaltrials.gov/study/NCT04672239

International Registered Report Identifier (IRRID): RR2-10.2196/40867

(JMIR Mhealth Uhealth 2025;13:e53971) doi:10.2196/53971

KEYWORDS

mobile health; mHealth; smoking cessation; nondaily smoking; smartphone; smoking; positive psychology; mobile phone

Introduction

Background

Nondaily smoking is an increasingly prevalent pattern of smoking, which despite perceptions to the contrary, results in substantial health detriments. Currently, 25% of all US adults who smoke do so less frequently than daily [1]. This prevalence has increased by 27% in the last decade [1]. Formerly believed to be a transient pattern of smoking [2,3], research has established that nondaily smoking is a persistent pattern [4-8]. Nondaily smoking is more prevalent among Black and Latinx populations [9-11] and is increasingly prevalent among people with serious mental health issues [12]. People who smoke less than daily and have never smoked daily are younger than those who smoke daily or have previously smoked daily [13]. More than 10 years ago, nondaily smoking was highlighted as an important public health issue [14]. This call to action has produced compelling evidence on the substantial negative impact of nondaily smoking on health [15], observed for smoking as few as 6 to 10 cigarettes per month [13].

Despite the prevalence of nondaily smoking, the US Clinical Practice Guidelines for smoking cessation [16] offer no guidance on how to support people who smoke less than daily in smoking cessation due to a lack of evidence for efficacious approaches. People who smoke less than daily do not view themselves as "smokers" who need "treatment" and thus are challenging to engage in traditional smoking cessation treatments [17-20]. They do, however, report high motivation to quit smoking, more so than people who smoke daily [3,17,18], which manifests in more recent and planned cessation efforts [18,21-23]. To date, only 2 trials have tested interventions for nondaily smoking cessation. Both focused on pharmacological treatments (ie, nicotine replacement therapy) and both failed to show efficacy in achieving smoking abstinence [24,25]. This lack of efficacy is in line with perceptions by people who smoke less than daily that withdrawal is not a barrier to their smoking cessation [26,27]. Behavioral intervention approaches may be more effective. Smartphone apps are a highly sought after source of behavioral support, as demonstrated by >33 million downloads that smoking cessation apps have generated [28]. Moreover, this technology-facilitated approach invokes less treatment resistance among people who smoke less than daily than more traditional smoking cessation treatments [20].

Using the Smiling instead of Smoking Smartphone App to Support People Who Smoke Less Than Daily in Quitting Smoking

Building on research that found that people who smoke less than daily prioritized having positive self-identity and wellness [26], our team built a smartphone app that focused on fostering the experience of positive emotions, the Smiling instead of Smoking (SiS) app [29,30]. As detailed in the protocol paper matching this outcome report [31], the treatment approach of the SiS app was inspired by the development of positive psychotherapy for smoking cessation [32,33]. Typically, during smoking cessation, positive affect decreases temporally, following a U-shaped trajectory consistent with a withdrawal effect [34]. However, research has shown that having high positive affect is beneficial to quitting in that it is related to increased self-efficacy to abstain from smoking [35], decreased desire to smoke [36,37], and greater readiness to process self-relevant health information [38]. While positive psychotherapy for smoking cessation was originally developed for daily smokers, research with people who smoke less than daily has highlighted the impact of positive affect on reducing craving [36]. As many other constructs related to smoking cessation are less salient to people who smoke less than daily [26], we chose "fostering positive affect" as the therapeutic goal of the SiS app.

To develop the SiS app, we used an iterative, staged process, consisting of 3 studies. This paper reports on the outcomes of the third study (NCT04672239). The 2 prior studies demonstrated the app's ability to engage people who smoke less than daily when onboarded in-person and when onboarded remotely [30,39]. User experiences by participants in both studies were used to further adapt and develop the app, as described previously [31,39]. In both studies, within-person changes were observed in line with our conceptual model [30,39]. Moreover, in a feature-level analysis of data obtained during the second study (NCT03951766), the number of days participants used the app significantly predicted 30-day point prevalence abstinence (PPA) at the end of treatment and 6-month follow-up [40]. Further analyses indicated that this effect was primarily due to interacting with the positive psychology components of the SiS app, rather than the app's tools dedicated to smoking cessation guidance.

XSL•FO
RenderX

Objectives

In this paper, we present the outcomes of the first randomized trial testing the SiS app, using a parallel, unblinded, randomized design. As described in the protocol paper [31], this study is a proof-of-concept randomized controlled trial (RCT), where we targeted increased self-efficacy by treatment end as a proximal, conceptual proof-of-concept indicator of treatment efficacy. We chose self-efficacy as our primary, proximal outcome variable because of its prominence in theoretical models of health behavior change [41-44] and because of prior demonstrable effects. Namely, a large-scale efficacy trial of an SMS text messaging intervention for smoking cessation identified self-efficacy as the primary mediator of treatment on conferring benefit for smoking cessation [45].

In this trial, we tested the SiS app against 2 control conditions to allow us to choose the more rigorous control condition for a subsequent, large-scale efficacy RCT. Both control conditions use smoking cessation materials developed and disseminated by the National Cancer Institute (NCI): the NCI's smartphone app QuitGuide (QG) and the NCI's smoking cessation brochure "Clearing the Air" (CtA), respectively. Secondary goals of this proof-of-concept RCT were to compare treatment acceptability and feasibility across groups, describe smoking cessation outcomes, and test for differences on secondary proof-of-concept efficacy outcomes. In line with our conceptual model, we chose positive affect, craving, and attitudes toward smoking as these secondary conceptual markers of efficacy.

Methods

Participants

Participants were adults who smoke less than daily and were interested in using support materials to help them quit smoking (recruitment period: February 25, 2021, to June 29, 2022). Study recruitment information was displayed on Craigslist, Facebook, Reddit, Smokefree.gov, ClinicalTrials.gov, a study recruitment website at Massachusetts General Hospital, and a study specific website, and was shown to people interacting with study recruitment websites (ie, Clinical Connection and Wayturn). To be eligible, participants had to be above the age of 18 years, own an Android or iPhone smartphone, smoke cigarettes at least weekly but no more than 25 out of the past 30 days, have a lifetime history of having smoked >100 cigarettes, be willing to make a quit attempt as part of the study, and currently reside in the United States.

For this proof-of-concept RCT, we selected a sample size that would allow us to detect group differences on our primary outcome variable (ie, self-efficacy) [31]. A large RCT testing a mobile health intervention for smoking cessation found group differences of Cohen d=0.66 on self-efficacy 1 month after quitting [45]. We conservatively chose to power the trial to detect Cohen d=0.50, as our primary end point was further out (ie, 6 weeks after quitting).

Procedure

The study was conducted entirely remotely. Participants were recruited on the web nationwide within the United States, using the tagline "are you smoking nondaily and want to quit?." Interested participants were phone screened, where they learned more about the study (ie, that the study involves randomization, that some but not all groups would use a smartphone app, and that there would be web-based survey spanning 6 months). Following this phone screen, interested people were emailed a screening survey. This survey contained check items (eg, "Please indicate 'strongly agree'") to assess participants' ability to successfully interact with the web-based surveying platform, REDCap (Research Electronic Data Capture; Vanderbilt University) [46]. Screening participants who completed this survey were invited to an enrollment phone call and asked to provide contact information for family or friends who would be available to the study staff to help contact participants, in case the participants changed their contact information during the course of the study. Participants were advised that the enrollment phone call needed to be scheduled to occur 1 week before their quit date, but that they could choose their quit day, and the enrollment call would be scheduled to accommodate their chosen quit day. If their chosen quit date occurred before the study closed for enrollment, their preference was accommodated. On average, 19 (SD 15) days elapsed between participants completing the screening survey and their chosen quit day.

During the enrollment phone call, the study fact sheet was reviewed, smoking status was reconfirmed via self-report, and participants provided verbal consent to enroll in the study. Following consent, while on the call, study staff emailed participants a link to download their assigned app (for both apps, links to the GooglePlay and iStore listings were sent; both apps were freely available to the public at no cost) or the PDF for CtA, as chosen by randomization using a 1:1:1 allocation; randomization occurred via randomization sheet, with staff looking up a participant's group assignment within the 24 hours before the enrollment visit to prepare for the onboarding. Staff then engaged participants in a scripted 15-minute onboarding dialogue based on their assigned treatment condition. Staff did not specify which treatment condition was the intervention of treatment versus the control conditions, but rather presented each assigned smoking cessation material as potentially useful to help support smoking cessation. The onboarding script length was matched between randomized groups. This scripted dialogue systematically led participants through their assigned smoking cessation materials, with study staff asking participants to read out loud the text displayed by their assigned app or brochure and asking participants to share answers to questions their smoking materials asked them to think about.

At the end of onboarding, study staff instructed participants to use their assigned app or brochure for a period of 7 weeks (1 week before quitting and 6 weeks after quitting) and to complete follow-up surveys on the web 2, 6, 12, and 24 weeks after their initially chosen quit date. Participants were told that they could change their quit date as needed, but that this initially chosen quit date would be the anchor date for surveying.

In total, we screened 1268 individuals over phone. Of these, 41.4% (525/1268) of the individuals were found ineligible during the phone screening (primarily due to not smoking less than daily; 497/525, 94.7%), and 15.3% (194/1268) of the individuals decided against the study (most commonly due to

XSL•FO
RenderX

having lost interest in the study; 56/194, 28.9%) or not liking the study as described (22/194, 11.3%), but also due to study logistics, such as needing to provide their social security number for payment by check (36/194, 18.6%) or having to wait for checks to be mailed (26/194, 13.4%). The remaining 43.3% (549/1268) of the individuals were emailed the screening survey. Of these, 175 (31.9%) chose not to complete the survey, 86 (15.7%) failed the check items embedded in the survey, and 2 (0.4%) completed the survey after the study had closed enrollment. The remaining 52.1% (286/549) of the individuals were invited to the enrollment phone call. Of these, 28 (9.8%) participants did not show up for the enrollment phone call, 20 (7%) participants decided against the study at this point, and 1 (0.3%) participant could not be reached to schedule the enrollment. The remaining 82.9% (237/289) of the individuals started the enrollment phone call. During this phone call, 2.8% (8/237) of the individuals were found ineligible (ie, n=5 did not smoke less than daily, n=2 did not want to quit smoking, and n=1 did not reside in the United States). The remaining 96.6% (229/237) of the individuals were enrolled and proceeded to onboarding. During onboarding, after having learned about their group assignment, 3 participants (one in each treatment group) decided against the study, resulting in a final sample size of 226 participants who were enrolled and successfully onboarded to their smoking cessation materials.

Survey responses were obtained from 95.1% (215/226), 89.4% (202/226), 82.3% (186/226), and 81.9% (185/226) of the participants at 2, 6, 12, and 24 weeks after the initially chosen quit day, respectively; these included participants who only completed partial surveys (12/226, 5.3%; 8/226, 3.5%; 7/226, 3.1%; and 9/226, 4%, respectively) and participants who completed the survey but incorrectly responded to 2+ check items (6/226, 2.7%; 6/226, 2.7%; 7/226, 3.1%; and 2/226, 0.9%, respectively). Obtaining survey responses did not differ between groups (all $P>.07$ at all assessments).

Treatment Condition: SiS Smartphone App

Participants in the treatment condition were asked to use the SiS app every day for 7 weeks. As described in more detail elsewhere [31], the SiS app provides smoking cessation tools within the framework of positive psychology. Thus, app users were asked to engage in activities that foster positive affect and in activities that focus on smoking cessation (Figure 1). To enhance prescriptive clarity [40], a tool called “Today’s Tasks” listed all tasks to be completed that day; clicking on the task brought the app user directly to that task in the app.

XSL•FO
RenderX

Figure 1. Summary of the content provided in the Smiling instead of Smoking (SiS) app.

The positive psychology content was derived from positive psychology findings, as summarized previously [47]. App users were asked to complete a positive psychology habit-building exercise every day to develop a habit of noticing and savoring the positive experiences app users encountered in everyday life. For moments of low positive affect, the SiS app offered a variety of happiness boost activities. The rationale for engaging in positive affect fostering activities was reinforced throughout the app: the overall rationale was provided in a section called "why work on happiness"; "why" buttons were linked to all exercises, which provided information about the rationale for each specific exercise; and push notifications, called "Owl Wisdoms," were sent every 2-3 days to provide positive psychology science findings relevant to the tasks the app assigned at that time.

The smoking cessation content built directly on the materials provided by the NCI's Smokefree.gov resources, which are in line with United States Clinical Practice Guidelines [16]. The

XSL•FO
RenderX

smoking cessation content consisted of tools to engage app users in tracking their cigarette use, understanding their triggers for smoking, learning about benefits of quitting smoking, and reflecting on personal reasons to quit smoking. Guidance to engage with these tools was provided via time-anchored push notifications, called "Behavioral Challenges," which asked app users to complete a specific task within the app on that specific day.

During the course of this trial, the SiS app, as hosted on hospital servers, experienced three downtimes, lasting from 1 hour to 2 days, where the app did not record its use, but users could still interact with the app in its intended way. The downtimes were noted by the research team at Massachusetts General Hospital; the app programmers (ie, PreviewLabs Inc) then implemented the needed changes in the programming to overcome outages. No changes were made to the overall functionality of the app during the duration of this study.

Control Condition 1: NCI QG Smartphone App

Participants in control condition 1 were asked to use the QG app every day for 7 weeks. As described in more detail elsewhere [31], the QG app provided many tools analogous to the smoking cessation content provided in the SiS app, just designed differently. Specifically, functionality exists to set the quit date, track mood and craving, read about best practices for quitting smoking, log cigarettes, see graphs about smoking patterns based on one's cigarette log, set reminders to stay smoke free (time- and location-based), and receive tips in response to reporting craving or negative affect. QG is frequently used as a comparison app in smartphone app smoking cessation studies [28,48-53] because it controls for time, attention, and modality of delivery of smoking cessation support.

Control Condition 2: NCI CtA Brochure

Participants in control condition 2 were asked to use the CtA brochure every day for 7 weeks. As described in more detail elsewhere [31], this 36-page brochure [54] provided analogous smoking cessation information as the QG app. It is composed of both text and worksheets to engage the reader in reflecting on reasons to quit, learning about the benefits of quitting, logging smoking and its triggers, and formulating strategies for dealing with triggers. This booklet has been used as a "treatment as usual" comparison condition in past RCTs evaluating phone-based and other mHealth smoking cessation technologies [55-58].

Measures

Two sources of data were collected: self-report answers to web-based surveys and passively recorded app use data. The primary end point for this study was treatment end, which occurred 6 weeks after the originally chosen quit date.

Primary Outcome

The primary outcome for this proof-of-concept RCT was self-efficacy to abstain from smoking, as measured using the Self-Efficacy Questionnaire, assessed at baseline, and 2, 6, and 12 weeks after the initially chosen quit day. This 12-item scale asks participants to rate their confidence on a 0 to 100 slider scale in their ability to abstain from smoking when facing internal stimuli (eg, feeling depressed) and external stimuli (eg, being with people who smoke). Scale scores are reported as mean scores (range 0-100), where higher scores indicate greater self-efficacy to abstain from smoking. The primary outcome variable was the total Self-Efficacy Questionnaire score. Subscale scores (ie, internal stimuli and external stimuli) were also calculated to provide more nuanced insight and are presented on the same mean score scale. The internal consistency (Cronbach α) of the total score ranged from 0.89 at baseline to 0.95 at the 6-month follow-up, for the internal subscale from 0.86 at baseline to 0.95 at the 6-month follow-up, and for the external subscale from 0.80 at week 2 to 0.91 at the 6-month follow-up in this sample.

Secondary Outcomes: Treatment Acceptability

Treatment acceptability was assessed at treatment end using the Client Satisfaction Questionnaire (CSQ) [59] to assess satisfaction with the received smoking cessation support, the System Usability Scale (SUS) [60] to assess acceptability of the technology used to provide smoking cessation support, and 2 single item measures to assess overall ratings of treatment likeability and satisfaction. The CSQ scores are reported as total scores that can range from 0 to 27, where higher scores indicate higher client satisfaction; the internal consistency of CSQ scores was 0.95 in this sample. The SUS scale scores are transformed to a 0 to 100 range, where higher scores mean greater system usability. Interpreted on a letter grading system, SUS scores in the range from 78.9 to 80.7 correspond to an A-, 80.8 to 84.0 to an A, and 84.1 to 100 to an A+ [61]. The internal consistency of SUS scale scores was 0.88 in this study. Both single item measures were rated on 5-point Likert scales ranging from 1="I strongly disliked using the app" to 5="I strongly liked using the app" for likeability and from 1="very unsatisfied" to 5="very satisfied" for satisfaction.

Secondary Outcomes: Treatment Feasibility

Treatment acceptability was assessed at multiple time points. For participants randomized to use apps, acceptability was measured by their actual engagement with their assigned apps (ie, number of days participants used their assigned apps) and their subjective appraisal of how much (in minutes per week) they used their assigned apps at midtreatment (week –1 to week 2 after the quit date) and at the end of treatment (week 3 to 6 after the quit date). All participants, regardless of whether they were assigned a smartphone app, were asked to reflect at treatment end on how much time they spent applying content learned through the smoking cessation support materials (in minutes per week), to report on recommended smoking cessation strategies they used, and to rate the perceived impact of their assigned smoking cessation treatment. Smoking cessation strategies used were evaluated with an 8-item survey rated on 5-point Likert scales ranging from 1="strongly disagree" to 5="strongly agree," and aggregate scores were presented as mean scores, where higher scores indicate a higher use of multiple cessation strategies. The internal consistency of this Use of Smoking Cessation Strategies scale was 0.80 in this sample. Because the SiS app uses a positive psychology approach to smoking cessation, we also included items that addressed positive psychology approaches to smoking cessation

XSL•FO
RenderX

in this list (eg, "I focused on feeling as happy as possible while I was quitting smoking" and "I focused on the good things that happened each day"), as used in prior research on positive psychotherapy for smoking [33]. These items were summarized into a similar mean score of Use of Positive Psychology Strategies (same range and interpretation as the Use of Smoking Cessation Strategies) and had an internal consistency of 0.88 in this sample. The perceived impact of assigned smoking cessation treatments was evaluated with a 17-item survey, as used in prior studies on the SiS app [30,39], (eg, [The assigned treatment] "...made me think that it was worthwhile for me to quit.", "...gave me the feeling I could get trusted advice at any time.") rated on a 5-point Likert scale ranging from 1="strongly disagree" to 5="strongly agree." Aggregate scores were presented as mean scores, where higher scores indicated a higher perceived impact of the assigned treatment on participants' smoking cessation support. The internal consistency of the Perceived Impact scale was 0.95 in this sample.

Secondary Outcomes: Exploratory Treatment Effectiveness Outcomes

While this trial was not powered to detect differences in smoking cessation, we did assess smoking cessation outcomes for descriptive purposes. Specifically, we assessed 30-day PPA, as self-reported smoking status 6, 12, and 24 weeks after the initially chosen quit day. Participants who did not report on smoking status were presumed to be smoking. We also assessed smoking reduction (ie, the difference in cigarettes smoked in the past week, assessed at baseline vs the end of treatment).

Secondary Outcomes: Secondary Proof-of-Concept Markers

In addition to the primary outcome, we assessed additional constructs in line with our conceptual model. Specifically, we used the positive affect subscale of the Positive and Negative Affect Schedule [62] to assess positive affect within the past week and the Brief Questionnaire of Smoking Urges (Brief-QSU) [63] to assess in the moment craving. The 10-item positive affect subscale of the Positive and Negative Affect Schedule is reported as a total score that can range from 10 to 50, with higher scores representing higher levels of positive affect; the internal consistency of this scale ranged from 0.91 at week 2 to 0.94 at the 6-month follow-up. The 10-item Brief-QSU scale is reported as a total score that can range from 7 to 70, with higher scores indicating greater cigarette craving. In this sample, the internal consistency of the Brief-QSU ranged from 0.92 at baseline to 0.96 at the 6-month follow-up. To assess attitudes toward smoking, we used the 3 subscales of the Attitudes Toward Smoking [64] scale, and the Decisional Balance Inventory for Smoking short form (DCB-SF) [65]. The Attitudes Toward Smoking subscales have different numbers of items, so we present each subscale score as a mean score with a possible range from 1="strongly disagree" to 5="strongly agree," so that higher scores indicate stronger agreement with statements about the Adverse Effects of Smoking (10 items), the Psychoactive Benefits of Smoking (4 items), and the Pleasure of Smoking (4 items) for the respective subscales. The internal consistencies of all 3 subscales ranged from 0.82 for the Psychoactive Benefits of Smoking at baseline to 0.94 for the Adverse Effects of Smoking at the 3-month follow-up. The DCB-SF is a 6-item scale that is rated on 0 to 100 slider scales. Three DCB-SF items evaluate the positive smoking expectancies (pros) and 3 items evaluate the negative smoking expectancies (cons); item scores are averaged within each subscale for a score range of 0 to 100, where higher scores indicate greater agreement with the pros or cons of smoking, respectively. The internal consistency of the subscales ranged from 0.74 to 0.85 for pros and 0.68 to 0.78 for cons across the assessments.

Participant Descriptors: Demographics, Smoking Characteristics, and Clinical Characteristics

At baseline, participants reported on demographic information, smoking characteristics, and clinical characteristics. Most items were stand-alone multiple-choice items. We used validated scales to assess nicotine dependence, depression severity, anxiety severity, and capacity to experience pleasure. Specifically, we used the Fagerström Test for Cigarette Dependence [66], the Center of Epidemiologic Studies Depression Scale [67], the Generalized Anxiety Disorder Screener [68], and the Snaith-Hamilton Pleasure Scale [69], respectively. For cigarette dependence, Fagerström Test for Cigarette Dependence total scores can range from 0 to 10, where higher scores indicate greater dependence. To summarize clinical characteristics, we used a Center of Epidemiologic Studies Depression Scale cutoff score of ≥10 to indicate the risk of depression [67], a Generalized Anxiety Disorder Screener cutoff score of ≥10 to indicate moderate anxiety [68], and a Snaith-Hamilton Pleasure Scale cutoff score of >2 to indicate anhedonia [69], an inability to feel pleasure.

Analytic Strategy

Data Preparation

Before analysis, we reviewed all check items participants completed as part of their web-based surveys. Data were set to missing for a specific scale if the participant incorrectly responded to the check item embedded in that scale. Data were set to missing for the entire time point, if the participant responded incorrectly to ≥2 check items. Scale scores were then calculated; if an item was left blank, the scale score was calculated using the average of the remaining items (after accounting for reverse coding, as applicable), so long as ≥80% of the items were completed.

Analysis of Primary Outcome

To test if randomized group assignment was significantly related to treatment outcome, we used a generalized linear mixed model, where repeated observations per person were modeled using an unstructured covariance matrix. Predictors included in the model were GROUP (ie, SiS vs QG vs CtA), TIME (ie, baseline, week 2, week 6, week 12, and week 24), and the GROUP*TIME interaction effect. A contrast statement was used to derive the test statistic for the GROUP comparison at week 6, our primary endpoint, given the overall longitudinal model. The contrast statement tested the null hypothesis that SiS=QG=CtA. If this null hypothesis was rejected (P<.05) for the primary outcome, we followed this up with pair-wise follow-up tests to compare the SiS app treatment to each of the specific control groups (ie, SiS vs QG, SiS vs CtA); for secondary outcomes, we provided

XSL•FO
RenderX

these secondary pairwise comparisons for all outcomes, regardless of statistical significance of the overall effect. Between-group effect sizes were calculated using Cohen *d*, where effect sizes are interpreted such that 0.2 is small, 0.5 is medium, and 0.8 is large [70]. Details about the significance tests and pairwise comparisons are available in Tables S1 and S2 in Multimedia Appendix 1.

Analysis of Secondary Outcomes

The same analytic strategy was used for secondary outcomes, but adjusted for fewer assessments (ie, no TIME predictor, if outcome was only assessed once), and the distribution of the variable of interest (ie, logit model used for the binary outcome 30-day point prevalence smoking abstinence). Some outcomes (including change in past week cigarettes, time spent applying content, app use, and app use) were heavily skewed to the right; for these outcomes, we used the nonparametric Wilcoxon rank sum test to analyze group differences (Table S3 in Multimedia Appendix 1). Effect sizes for these outcomes were reported as $r=Z/\text{sqrt}(N)$, where Z is the test statistic from the Wilcoxon rank sum test and N is the total number of observations (ie, participants), and which are interpreted such that 0.1 is small, 0.3 is medium, and 0.5 is large [70]. Effect sizes for the logit models are presented as odds ratios with 95% CIs. Because these secondary outcomes cover distinctly different domains and are interpreted marginally to give a fuller picture of the participant experience, we did not correct for multiple testing [71]. We calculated effect sizes to provide insight into the relative strength of these effects. All analyses were performed using SAS System for Windows (version 9.4; IBM Corp).

Ethical Considerations

The study was approved by the Mass General Brigham institutional review board (2020P003466) and registered on ClinicalTrials.gov (NCT04672239). All participants provided informed consent; specifically, they provided verbal consent after reviewing the study fact sheet and engaging in a true or false knowledge test of the content of the study fact sheet in conversation with study staff. Surveys were collected via Health Insurance Portability and Accountability Act compliant technology (ie, REDCap), which was only accessible to staff trained in the conduct of research. Smartphone data were stored on password protected Mass General Brigham servers and linked to survey data via study-generated app ID number. Participants received US $25 for completed surveys or US $10 for incomplete surveys or surveys with failed check items. They received US $50 for the week 6 survey (end of treatment), which was longer than the other surveys. Participants provided their social security number to enable remuneration by check.

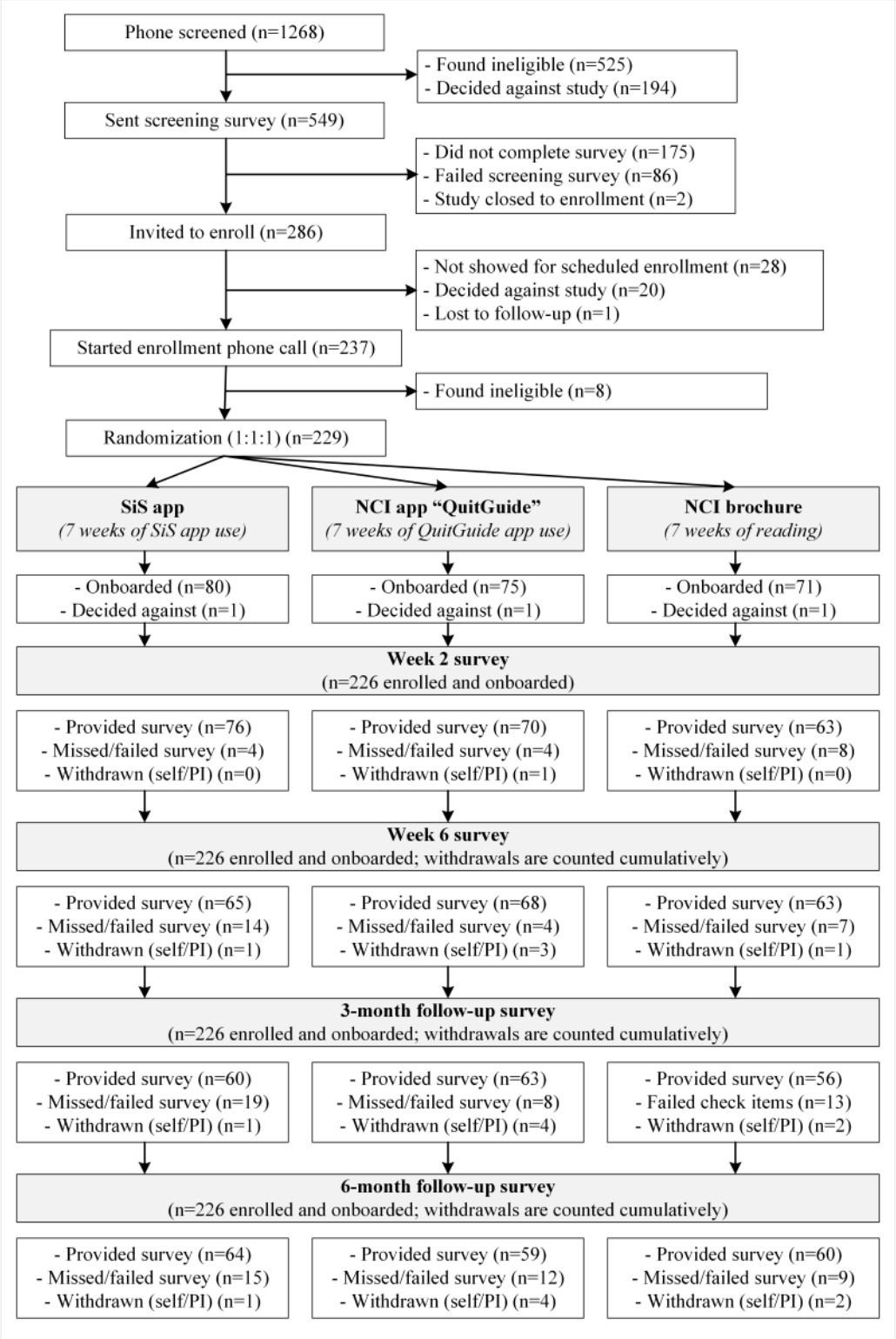
Results

Participant Characteristics

Study participants (Figure 2; CONSORT-eHEALTH checklist is provided in Multimedia Appendix 2) lived in urban, suburban, and rural areas, located in 42 out of 50 US states (Northeast region: 51/226, 22.6%; Midwest region: 31/226, 13.7%; South region: 97/226, 42.9%; West region: 47/226, 20.8%). They were largely (151/226, 66.8%) people who had smoked daily previously, who smoked an average of 3.4 (SD 2.7) cigarettes per smoking day on 14.9 (SD 4.9) days out of the past 30 days. Many had made a previous quit attempt; many had tried e-cigarettes; few were using them at the time of the study (Table 1).

XSL•FO
RenderX

Figure 2. Flow of participants throughout the study from phone screening to the end of 6-month follow-up. NCI: National Cancer Institute; PI: principal investigator; SiS: Smiling instead of Smoking.



XSL•FO
RenderX

Table 1. Baseline characteristics of study participants.

Characteristic	SiS3[a] (n=80)	QuitGuide (n=75)	Clearing the Air (n=71)	Total (N=226)
Demographics				
Age (y), mean (SD)	41.3 (11.1)	41.6 (13.1)	40.0 (12.5)	41.0 (12.2)
Gender, n (%)				
Men	28 (35)	37 (49.3)	30 (42.3)	95 (42)
Women	51 (63.7)	36 (48)	41 (57.7)	128 (56.6)
Nonbinary	1 (1.3)	2 (2.7)	0 (0)	3 (1.3)
Race, n (%)				
American Indian or Alaska Native	0 (0)	0 (0)	1 (1.4)	1 (0.4)
Asian or Pacific Islander	1 (1.3)	1 (1.3)	4 (5.6)	6 (2.7)
Black	23 (28.7)	19 (25.3)	13 (18.3)	55 (24.3)
White	50 (62.5)	46 (61.3)	48 (67.6)	144 (63.7)
Other or unknown	2 (2.5)	3 (4)	1 (1.4)	6 (2.7)
More than one race	4 (5)	6 (8)	4 (5.6)	14 (6.2)
Hispanic, n (%)	15 (18.7)	16 (21.3)	6 (8.5)	37 (16.4)
Education, n (%)				
High school or less	7 (8.7)	12 (16)	5 (7)	34 (15)
Some college	31 (38.7)	38 (50.7)	25 (35.2)	94 (41.6)
Bachelor of Arts, Bachelor of Science, or higher	42 (52.5)	25 (33.3)	30 (42.3)	97 (42.9)
Employment, n (%)				
Full-time	29 (36.3)	30 (40)	27 (38)	86 (38)
Part-time	14 (17.5)	11 (14.7)	18 (25.3)	43 (19)
Unemployed	32 (40)	33 (44)	23 (32.4)	88 (38.9)
Household income (US $), n (%)				
≤39,999	41 (51.3)	42 (56)	30 (42.3)	113 (50)
40,000-69,999	18 (22.5)	18 (24)	22 (31)	58 (25.7)
70,000-99,999	16 (20)	8 (10.7)	10 (14.1)	34 (15)
≥100,000	4 (5)	7 (9.3)	8 (11.3)	19 (8.4)
Geographic location, n (%)				
Urban	39 (48.7)	34 (45.3)	27 (38)	100 (44.2)
Suburban	29 (36.3)	24 (32)	35 (49.3)	88 (38.9)
Rural	12 (15)	17 (22.7)	9 (12.7)	38 (16.8)
Smoking characteristics				
Number of days smoked in past 30 days, mean (SD)	15.3 (5.4)	14.7 (4.8)	14.6 (4.5)	14.9 (4.9)
Number of cigarettes smoked per smoking day, mean (SD)	3.5 (2.9)	3.0 (2.2)	3.8 (3.0)	3.4 (2.7)
Cigarette dependence (FTCD[b]), mean (SD)	1.7 (1.9)	1.6 (1.8)	1.7 (1.6)	1.7 (1.8)
Ever smoked daily? (yes), n (%)	54 (67.5)	53 (70.7)	44 (62)	151 (66.8)
Ever quit before? (yes), n (%)	49 (61.3)	56 (74.7)	52 (73.2)	157 (69.5)
Ever used E-cigarette? (yes), n (%)	46 (57.5)	41 (54.7)	46 (64.8)	133 (58.8)
Currently using E-cigarette? (yes), n (%)	14 (17.5)	10 (13.3)	9 (12.7)	33 (14.6)
Clinical characteristics				

XSL•FO
RenderX

Characteristic	SiS3[a] (n=80)	QuitGuide (n=75)	Clearing the Air (n=71)	Total (N=226)
Depressive symptoms (CES-D-10[c]) (at risk), n (%)	30 (37.5)	37 (49.3)	41 (57.7)	108 (47.8)
Moderate to severe anxiety (GAD-7[d]) (yes), n (%)	19 (23.7)	21 (28)	19 (26.8)	59 (26.1)
Anhedonia (SHAPS[e]) (yes), n (%)	11 (13.7)	18 (24)	17 (23.9)	46 (20.3)
Ever diagnosed mental health condition (yes), n (%)	24 (30)	23 (30.7)	22 (31)	69 (30.5)
Alcohol use, past 30 days (yes), n (%)				
Any alcohol use	45 (56.3)	37 (49.3)	40 (56.3)	122 (54)
Exceeded drinking guidelines	23 (28.7)	18 (24)	19 (26.8)	60 (26.5)
Any binge drinking episode	25 (31.3)	15 (20)	18 (25.3)	58 (25.7)

[a]SiS3: version 3 of the Smiling instead of Smoking app.

[b]FTCD: Fagerström Test for Cigarette Dependence (range: 0-10, where higher scores indicate greater nicotine dependence).

[c]CES-D-10: Center of Epidemiologic Studies Depression Scale, 10-item version (yes=total score of ≥10, indicating risk of depression).

[d]GAD-7: Generalized Anxiety Disorder Screener (yes=total score of ≥10, indicating moderate or more severe anxiety).

[e]SHAPS: Snaith-Hamilton Pleasure Scale (yes=total score >2, indicating mild or more severe anhedonia, an inability to feel pleasure).

Scores on validated scales indicated presence of significant depressive symptoms for 47.8% (108/226) and moderate anxiety symptoms for 26.1% (59/226) of the participants. One-third of the participants (69/226, 30.5%) reported having been diagnosed with a mental health condition in their lifetime. Capacity to experience pleasure was in the normal range for most participants (180/226, 79.6%). A quarter of participants exceeded drinking guidelines [69] during the 30 days before the study.

Primary Outcome: Self-Efficacy

Analyses indicated a significant difference between randomized groups at treatment end ($F_{2,198}$=4.32; *P*=.01), where pair-wise follow-up tests indicated higher self-efficacy for participants randomized to SiS compared to QG ($t_{198.8}$=2.31; *P*=.02; Cohen *d*=0.40) and CtA ($t_{199.7}$=2.72; *P*=.007; Cohen *d*=0.50; Table 2; Tables S1 and S2 in Multimedia Appendix 1). The effect was consistent with the effect observed in subscale specific analyses, with effect sizes ranging from Cohen *d*=0.34 to 0.50 across the pairwise comparisons.

XSL•FO
RenderX

Table 2. Study outcomes at end of treatment (6 weeks after initially chosen quit date).

Outcome	SiS3[a] app	QG[b] app	CtA[c]	Effect size and *P* value, SiS3 versus					
				QG			CtA		
				Cohen *d*[d]	Cohen r[e]	*P* value	Cohen *d*	Cohen *r*	*P* value
Primary outcomes									
Self-efficacy (SEQ-12[f]): overall, mean (SD)	78.3 (17.3)	71.3 (17.7)	69.0 (19.8)	0.40	—[g]	.02	0.50	—	.007
Self-efficacy (SEQ-12): internal, mean (SD)	76.5 (20.8)	69.6 (19.4)	66.9 (21.4)	0.34	—	.03	0.45	—	.01
Self-efficacy (SEQ-12): external, mean (SD)	80.2 (16.4)	72.7 (19.4)	70.8 (21.3)	0.42	—	.03	0.50	—	.009
Secondary outcomes									
Treatment acceptability									
Client satisfaction (CSQ-8[h]), mean (SD)	26.9 (5.3)	26.6 (4.5)	25.7 (5.8)	0.06	—	.75	0.23	—	.17
System usability (SUS[i]), mean (SD)	81.9 (17.1)	79.5 (16.0)	—	0.15	—	.41	—	—	—
App likability rating, mean (SD)	4.2 (1.0)	4.1 (0.9)	—	0.12	—	.50	—	—	—
App satisfaction rating, mean (SD)	4.3 (1.0)	4.1 (0.9)	—	0.14	—	.42	—	—	—
Treatment feasibility									
Use of smoking cessation strategies, mean (SD)	4.1 (0.6)	3.9 (0.6)	4.0 (0.7)	0.38	—	.04	0.24	—	.16
Use of positive psychology strategies, mean (SD)	4.3 (0.6)	4.0 (0.5)	4.0 (0.6)	0.55	—	.003	0.42	—	.02
Perceived impact on quitting, mean (SD)	4.1 (0.7)	4.0 (0.6)	3.8 (0.7)	0.23	—	.20	0.40	—	.01
Applying content (min/wk), median (IQR)[j]	30 (20-70)	25 (10-60)	30 (10-60)	—	0.17	.13	—	0.12	.34
App use[k] (number of days used), median (IQR)[j]	33 (11-40)	26 (11-38)	—	—	0.07	.42	—	—	—
App use, weeks 1-3 (min/wk), median (IQR)[j]	40 (28-68)	28 (16-50)	—	—	0.23	.007	—	—	—
App use, weeks 4-7 (min/wk), median (IQR)[j]	35 (21-70)	23 (14-60)	—	—	0.17	.06	—	—	—
Exploratory outcomes (ie, secondary proof-of-concept efficacy outcomes)									
Positive affect									
Positive affect (PANAS[l]), mean (SD)	37.2 (8.2)	33.4 (9.1)	34.0 (8.4)	0.44	—	.01	0.38	—	.05
Desire to smoke									
Craving (Brief-QSU[m]), mean (SD)	14.9 (6.5)	19.7 (10.1)	19.7 (10.0)	–0.57	—	.005	–0.57	—	.005
Attitudes toward smoking									
Adverse effects (ATS[n]), mean (SD)	4.4 (0.6)	4.4 (0.6)	4.4 (0.6)	0.03	—	.74	0.05	—	.65
Psychoactive benefits (ATS), mean (SD)	2.6 (1.1)	2.8 (1.0)	2.9 (1.0)	–0.23	—	.17	–0.33	—	.1

XSL•FO
RenderX

Outcome	SiS3[a] app	QG[b] app	CtA[c]	Effect size and *P* value, SiS3 versus					
				QG			CtA		
				Cohen *d*[d]	Cohen r[e]	*P* value	Cohen *d*	Cohen *r*	*P* value
Pleasure (ATS), mean (SD)	2.3 (0.9)	2.4 (0.9)	2.7 (1.0)	–0.11	—	.44	–0.45	—	.04
Positive expectancies (DCB-SF[o]), mean (SD)	28.6 (26.5)	36.0 (24.2)	39.1 (25.6)	–0.29	—	.12	–0.40	—	.08
Negative expectancies (DCB-SF), mean (SD)	69.0 (27.6)	71.3 (23.0)	67.1 (24.6)	–0.09	—	.63	0.07	—	.54
Treatment effectiveness									
Change in past week cigarettes (cigarettes per week), median (IQR)[j]	–5 (–12 to –1)	–6 (–11 to –3)	–7 (–13 to –4)	—	0.00	1.00	—	0.02	.97

[a]SiS3: version 3 of the Smiling instead of Smoking app.

[b]QG: National Cancer Institute (NCI) QuitGuide app.

[c]CtA: National Cancer Institute "Clearing the Air" brochure.

[d]Cohen *d* effect size, where effects are interpreted as 0.2=small, 0.5=medium, and 0.8=large.

[e]Cohen *r* effect size, where effects are interpreted as 0.1=small, 0.3=medium, and 0.5=large.

[f]SEQ-12: Self-Efficacy Questionnaire (range 0-100, where higher scores indicate greater self-efficacy to abstain from smoking).

[g]Not applicable.

[h]CSQ-8: Client Satisfaction Questionnaire (range 0-27, where higher scores indicate higher client satisfaction).

[i]SUS: System Usability Scale (range 0-100, where higher scores mean greater app usability).

[j]Outcomes that were not normally distributed and were analyzed with Wilcoxon rank sum tests (rather than generalized linear mixed model used for all other outcomes).

[k]App use is reported for all who were onboarded (SiS3 app n=80, QuitGuide app n=75).

[l]PANAS: Positive and Negative Affect Schedule (range 10-50, where higher scores indicate higher levels of positive affect).

[m]Brief-QSU: Brief Questionnaire of Smoking Urges (range 7-70, with higher scores indicating greater cigarette craving).

[n]ATS: Attitudes Toward Smoking scale (subscale mean score range 1-5, where higher scores indicate stronger agreement with the subscale statements).

[o]DCB-SF: Decisional Balance Inventory for Smoking short form (subscale range 0-100, where higher scores indicate greater agreement with the pros and cons of smoking, respectively).

Secondary Outcomes: Treatment Acceptability and Feasibility

Treatment acceptability was comparable across the 3 randomized groups across all indices (Table 2; Tables S1 and S2 in Multimedia Appendix 1), indicating credibility of both control groups. Technological acceptability of SiS, as measured by the SUS, was rated as an "A" on average [61]; satisfaction with the smoking cessation treatment was similarly high.

Treatment feasibility was higher for SiS compared to QG and CtA across several indices (Table 2; Tables S1 and S2 in Multimedia Appendix 1), but in different ways. In comparison to QG, SiS participants more frequently used recommended smoking cessation strategies during their treatment (Cohen *d*=0.38). Subjectively, they felt that they had spent more time with the SiS than QG app (*r*=0.23) during the first 3 weeks of treatment. This effect was not significant for the second half of treatment, or significant on the objectively measured number of days with which participants engaged with their assigned apps. In comparison to CtA, SiS participants perceived a stronger impact of their treatment on their quitting experience (Cohen *d*=0.40). In comparison to both QG and CtA, SiS participants more frequently used positive psychology strategies to help them quit smoking than QG (Cohen *d*=0.55) or CtA (Cohen *d*=0.42) participants.

Exploratory Treatment Effectiveness Outcomes: Impact on Smoking

Self-reported 30-day PPA tended to be higher among SiS than QG or CtA participants, but not significantly so (Table 3). This trend was observable (but not significant) at both treatment end (SiS3 vs QG: *P*=.06; SiS3 vs CtA: *P*=.23) and 6-month follow-up (SiS3 vs QG: *P*=.64; SiS3 vs CtA: *P*=.55). Reductions in cigarette smoking from baseline to treatment end were comparable across treatment groups (|*r*|<0.03; Table 2).

XSL•FO
RenderX

Table 3. Point prevalence smoking abstinence at end of treatment and at end of follow-up.

	SiS3[a] app (n=80), n (%)	QG[b] app (n=75), n (%)	CtA[c] (n=71), n (%)	OR[d] (95% CI) SiS3 vs	
				QG	CtA
30-day PPA[e,f] 6 weeks after quitting	32 (40)	23 (31)	18 (25)	1.5 (0.8-2.9)	2.0 (0.98-3.9)
30-day PPA 6 months after quitting	38 (48)	32 (43)	31 (44)	1.2 (0.6-2.3)	1.2 (0.6-2.2)

[a]SiS3: version 3 of the Smiling instead of Smoking app.

[b]QG: National Cancer Institute (NCI) QuitGuide app.

[c]CtA: NCI "Clearing the Air" brochure.

[d]OR: odds ratio.

[e]PPA: point prevalence abstinence.

[f]PPA is reported for all n=226 who started treatment, with missing surveys interpreted as "smoking."

Exploratory Outcomes: Secondary Proof-of-Concept Efficacy Outcomes

Significant group differences were found for some of the secondary proof-of-concept outcomes (Table 2; Tables S1-S3 in Multimedia Appendix 1). The strongest effect was observed for craving, where SiS participants reported significantly lower craving at treatment end than QG or CtA participants, an effect that was consistent in size across the 2 control groups (both Cohen *d*=–0.57). This effect was also stronger than the effect on self-efficacy, the primary outcome of this study. There was also a significant effect on positive affect, where SiS participants had higher positive affect at treatment end, but this effect was only significant when comparing SiS and QG participants (Cohen *d*=0.44), but not in comparing SiS and CtA participants (Cohen *d*=0.38). In terms of attitudes toward smoking, participants reported largely similar attitudes at the end of treatment (all but one|d|<0.34). The only exception was that SiS participants reported significantly lower pleasure in smoking than CtA participants at the end of treatment (Cohen *d*=–0.45). This effect was not significant compared to QG participants (Cohen *d*=–0.11).

Discussion

Principal Findings

We tested version 3 of the SiS app in a proof-of-concept RCT using a remote-only design and brief onboarding procedures, matched for time and content across treatment groups. We found a significant treatment effect on our primary outcome target, such that participants receiving smoking cessation support via the SiS app were more confident at treatment end in their ability to remain abstinent than participants in 2 relevant control conditions (ie, the smoking cessation app QG and the smoking cessation pamphlet CtA). Importantly, engaging with the SiS app appeared to help people who smoke less than daily face both internal and external cues to smoking. Prior research has highlighted their susceptibility to smoke in response to cues in their environment, which is stronger in people who smoke less than daily than in those who smoke daily [72]. Thus, the capacity of treatment with the SiS app to impact the ability to abstain from smoking in these circumstances is particularly promising.

Further supporting proof-of-concept evidence for the efficacy of the SiS app were the observed strong impacts of the SiS app on secondary proof-of-concept markers, which were additional therapeutic targets. These findings suggested that engaging with the SiS app also helped people who smoke less than daily experience lower levels of craving and higher levels of positive affect. These effects are consistent with the conceptual model underpinning this treatment approach, which hypothesized that regular engagement with positive psychology exercises would increase positive affect. This increased level of positive affect would then lead to a decreased desire to smoke [31]. Previously, a large laboratory cue-reactivity study with people who smoke less than daily [36], had demonstrated the potency of positive affect cues to reduce craving. Our results are in line with these findings and extend them beyond the laboratory environment. Taken together, the conceptual evidence for the efficacy of the SiS app is strong.

The observed lack of group differences between the SiS app and the frequently used and widely distributed smoking cessation tools QG and CtA on treatment acceptability indices further supports the potential impact of the novel SiS intervention. In this RCT, we assessed both content and technological satisfaction with the treatment approach. Across the 3 randomized groups, people who smoke less than daily were, by and large, equally satisfied with the smoking cessation treatment they received. In comparison to the QG app, participants were also equally satisfied with the level of technology they had to master in SiS to engage with the treatment. Given that participants in the control conditions were asked to engage with the standard of care smoking cessation materials developed and delivered by the NCI, it is quite encouraging that the SiS app was able to achieve similar levels of satisfaction. The fact that engagement with the SiS app resulted in better outcomes in terms of self-efficacy, positive affect, and craving, despite a lack of differences in likeability of the treatment, speaks to the effectiveness of the content experienced in using the SiS app.

The feasibility indices tracked as part of this RCT showed that participants meaningfully interacted with the app. Participants used the app on 33 out of 49 days of the treatment. From prior work, we know that the majority of time SiS users spent with the app is focused on completing the daily happiness exercises.

XSL•FO
RenderX

For version 2 of the SiS app, 96% of the days participants used the SiS app, they completed the daily happiness exercise [39]. For version 1, app use was largely driven by completing happiness exercises, with 73% of participants completing happiness exercises on any given day [30]. Thus, it is likely that participants in this study as well spent a majority of their app time on happiness exercises. In self-report surveys, participants reported spending 35 to 40 minutes each week using the app. This length of app use time is comparable to the length of a clinician-delivered session, but it is achieved in small intervals over many days at the convenience of the participant. Completing the type of happiness exercises included in the SiS app only takes approximately 4 minutes, with demonstrable, immediate impacts on positive affect [47], and requires no scheduling or wait times. Results further suggest that people using the SiS app were more able to convert exposure to information to practice, with SiS users using more smoking cessation strategies in their daily lives. This effect was particularly strong for use of positive psychotherapy smoking cessation strategies [33] emphasized in the SiS app (Cohen d=0.55 SiS3 vs QG, Cohen d=0.42 SiS3 vs CtA), which are not typically part of smoking cessation treatment. A full-scale trial that is powered to detect differences in 30-day smoking cessation PPA is needed to test whether greater engagement in positive psychology strategies and greater self-efficacy translate to meaningful difference in smoking abstinence rates.

Regarding smoking abstinence rates, abstinence rates in our study were unusually high across all treatment groups. These higher abstinence rates are unlikely to be due to our focus on people who smoke less than daily, as prior RCTs showed low abstinence rates for people who smoke less than daily (5%-11%) [24,25]. These high abstinence rates may be due to our reliance on self-report; prior RCTs involving people who smoke less than daily used biochemical verification. However, it is important to note that the only fully powered RCT to date on the efficacy of a smartphone app for smoking cessation also relied solely on self-report and reported much lower abstinence rates (ie, 20% at the 6-month follow-up) compared to what we observed in this study [28]. Moreover, the self-reported abstinence rates in this study are in line with our previous findings in participants using prior versions of the SiS app (ie, 53% in study 1 on the SiS app, 56% in study 2) [30,39]. This leads us to believe that abstinence rates were higher in this study compared to other RCTs evaluating smoking cessation smartphone apps [28,49] because our procedures were more demanding. Unlike the largest RCT on a smoking cessation app to date [28], our study required participants to engage in an onboarding session with study staff. A smaller RCT, however, provided extensive technical support to study participants [49]. Possibly the nature of the support was more technical in nature and less focused on engaging with smoking cessation tools. In addition, our study was relatively demanding in terms of the steps participants had to take to get into the study, including the potential to not be able to proceed to enrollment if participants did not carefully read surveys. It is possible that these factors could explain the unusually high abstinence rates observed in this study. If so, it is worth investigating exactly what contributed to these high abstinence rates, as perhaps these mechanisms could be leveraged to increase therapeutic benefit.

Perhaps even more puzzling is the rising abstinence rates over time, with more participants reporting abstinence at 6-month follow-up than 6-week follow-up. This increase in abstinence occurred in all 3 groups. Our best guess for this phenomenon is that our study procedures resulted in selecting participants who took their quit attempt very seriously, thereby enabling deep learning of smoking cessation strategies during this quit attempt, which may have aided them in staying quit or using these skills for a new quit attempt with greater success. They also had full access to the assigned materials throughout the duration of the study, and thus could have used all materials well beyond the “prescribed” treatment period.

Strengths and Limitations

This study’s strengths include using 2 active, standard of care control conditions, controlling for onboarding time and attention between conditions, recruiting a national sample with racial-ethnic diversity similar to the racial-ethnic diversity of the US population of people who smoke less than daily [73], recruiting a sample comparable to people who smoke less than daily reported on in other studies, based on smoking behavior [74] and quit attempts [18], and the use of repeated measures per participant over time to increase measure sensitivity to individual differences and reliability.

This trial also had some limitations. First, this proof-of-concept trial was powered to detect differences in psychosocial constructs relevant to smoking cessation and not powered to detect differences in smoking cessation rates. Thus, the implications of this trial are limited to how well the treatment achieved differences in self-efficacy, positive affect, craving, and the use of smoking cessation strategies, and it remains to be tested how well these psychosocial constructs translate into successful smoking cessation. Second, it should be kept in mind that only a subset of those people initially interested in the study ultimately participated in the study. Thus, our study procedures may have inadvertently screened-out people less committed to making a quit attempt, thereby restricting implications of this study to people committed to making a quit attempt. Third, our onboarding procedures were more hands-on than a real-world scenario. In the real world, a person seeking smoking cessation support via smartphone app would simply browse through options in their phone’s app store, choose an app, download it, and start using it. In our study, participants did not need to choose an app. Instead, they were given clear guidance on how to use a specific app in a one-on-one phone call with clear instructions and comprehension checks. This type of onboarding is not standard. Interestingly, however, providing such guidance could be done on a larger scale through video onboarding, the help from digital navigators, or a combination of both [75,76]; this would still be shorter and require less training than telephone Quitline support. Future studies on this issue may be warranted. Such studies should examine the use of smartphone apps by people who smoke less than daily. Very little research exists regarding the type of smoking cessation support that people who smoke less than daily may accept. Existing evidence [20] suggests smartphone apps may be a particularly potent way in which to engage people who smoke less than daily in smoking cessation support, but more research is needed to demonstrate the real-world use of these technologies by this population.

XSL•FO
RenderX

It is also notable that elevated depressive symptoms varied across the three treatment groups, ranging from 38% in the SiS group to 58% in the CtA group, despite randomization. Overall, this prevalence of depressive symptoms is in line with findings on help-seeking people who smoke, where 39%-40% of callers to state quitlines reported symptoms in line with mild depression or major depression [77,78]. It is also important to note that the prevalence of nondaily smoking patterns among people who smoke is rising among those with serious mental health issues [12]. It is unclear how this difference between groups may have impacted study results. Past research has highlighted that smoking cessation may be harder for people with depressive symptoms [79,80]. At the same time, our previous findings on the same positive psychology exercises used in the SiS app have suggested that people with low positive affect benefit the most from positive psychology exercises [47]. Future research should include presence of depressive symptoms as a stratification factor to remove this ambiguous confounder from the interpretation of future results. Finally, we restricted study participation to residents of the United States. Thus, it is unclear how the findings of this study translate to findings that may be obtained when testing this treatment approach in other countries with different smoking cessation attitudes, infrastructure, and resources.

Conclusions

The SiS app demonstrated superior performance on increasing self-efficacy to abstain from smoking and other targeted psychological constructs compared to 2 active and relevant control conditions. These superior treatment effects cannot be attributed to greater likeability of the SiS treatment approach because satisfaction ratings were largely similar across randomized groups, and thus speak to the effectiveness of the content experienced by individuals who smoke less than daily when using the SiS app. Feasibility indices showed that people who smoke less than daily interacted meaningfully with the SiS app, suggesting that they successfully translated the knowledge gained from the app into practical use in their daily lives. Together, these findings provide strong evidence for the conceptual underpinnings of the SiS app and provide a good rationale for conducting a large-scale RCT that can test the effectiveness of the SiS app on smoking cessation.

Acknowledgments

This research was supported by a grant from the American Cancer Society (grant RSG CPPB—130323) to BBH.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BBH (lead), SSH (supporting), ERP (supporting), CWK (supporting), KRS (supporting)
Data curation: SSH (lead), BBH (supporting)
Formal analysis: SSH (lead), BBH (supporting)
Funding acquisition: BBH (lead), SSH (supporting), ERP (supporting), CWK (supporting)
Investigation: BBH (equal), SSH (supporting), KRS (equal), AEF (supporting), DF-A (supporting), ACW (supporting)
Methodology: BBH (lead), SSH (supporting), CWK (supporting), ERP (supporting), KRS (supporting)
Project administration: BBH (lead), KRS (supporting), SSH (supporting), AEF (supporting), DF-A (supporting), ACW (supporting)
Resources: BBH (lead), SSH (supporting), ERP (supporting), CWK (supporting)
Software: BBH (lead), SSH (supporting), ERP (supporting), CWK (supporting), KRS (supporting)
Supervision— BBH (lead), SSH (supporting), ERP (supporting), CWK (supporting), KRS (supporting)
Validation: SSH (lead), BBH (supporting), ERP (supporting), CWK (supporting), KRS (supporting)
Visualization: BBH (lead), SSH (supporting), KRS (supporting)
Writing—original draft: BBH (lead), SSH (supporting), KRS (supporting)
Writing—review & editing: BBH (lead), SSH (equal), ERP (equal), CWK (equal), KRS (equal), AEF (equal), DF-A (equal), ACW (equal)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables of study outcomes.
[DOCX File , 43 KB - mhealth_v13i1e53971_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3353 KB - mhealth_v13i1e53971_app2.pdf]

XSL•FO
RenderX

References

1. Creamer MR, Wang TW, Babb S, Cullen KA, Day H, Willis G, et al. Tobacco product use and cessation indicators among adults - United States, 2018. MMWR Morb Mortal Wkly Rep 2019 Nov 15;68(45):1013-1019 [FREE Full text] [doi: 10.15585/mmwr.mm6845a2] [Medline: 31725711]
2. Wortley PM, Husten CG, Trosclair A, Chrismon J, Pederson LL. Nondaily smokers: a descriptive analysis. Nicotine Tob Res 2003 Oct;5(5):755-759. [doi: 10.1080/1462220031000158753] [Medline: 14577992]
3. Hassmiller KM, Warner KE, Mendez D, Levy DT, Romano E. Nondaily smokers: who are they? Am J Public Health 2003 Aug;93(8):1321-1327. [doi: 10.2105/ajph.93.8.1321] [Medline: 12893622]
4. Hennrikus DJ, Jeffery RW, Lando HA. Occasional smoking in a Minnesota working population. Am J Public Health 1996 Sep;86(9):1260-1266. [doi: 10.2105/ajph.86.9.1260] [Medline: 8806378]
5. Hoeppner BB, Bidwell LC, Colby SM, Barnett NP. Smoking patterns and their relationship to drinking among first-year college students. Nicotine Tob Res 2014 Jun 10;16(6):743-752 [FREE Full text] [doi: 10.1093/ntr/ntt205] [Medline: 24415586]
6. Gilpin E, Cavin SW, Pierce JP. Adult smokers who do not smoke daily. Addiction 1997 Apr;92(4):473-480. [Medline: 9177069]
7. Schane RE, Glantz SA, Ling PM. Nondaily and social smoking: an increasingly prevalent pattern. Arch Intern Med 2009 Oct 26;169(19):1742-1744 [FREE Full text] [doi: 10.1001/archinternmed.2009.315] [Medline: 19858429]
8. Schane RE, Glantz SA, Ling PM. Social smoking implications for public health, clinical practice, and intervention research. Am J Prev Med 2009 Aug;37(2):124-131 [FREE Full text] [doi: 10.1016/j.amepre.2009.03.020] [Medline: 19589449]
9. Reyes-Guzman CM, Pfeiffer RM, Lubin J, Freedman ND, Cleary SD, Levine PH, et al. Determinants of light and intermittent smoking in the United States: results from three pooled national health surveys. Cancer Epidemiol Biomarkers Prev 2017 Feb 19;26(2):228-239 [FREE Full text] [doi: 10.1158/1055-9965.EPI-16-0028] [Medline: 27760782]
10. Arauz RF, Mayer M, Reyes-Guzman C, Ryan BM. Racial disparities in cigarette smoking behaviors and differences stratified by metropolitan area of residence. Int J Environ Res Public Health 2022 Mar 02;19(5):2910 [FREE Full text] [doi: 10.3390/ijerph19052910] [Medline: 35270603]
11. Weinberger AH, Giovenco DP, Zhu J, Lee J, Kashan RS, Goodwin RD. Racial/ethnic differences in daily, nondaily, and menthol cigarette use and smoking quit ratios in the United States: 2002 to 2016. Prev Med 2019 Aug;125:32-39. [doi: 10.1016/j.ypmed.2019.04.009] [Medline: 31004620]
12. Weinberger AH, Streck JM, Pacek LR, Goodwin RD. Nondaily cigarette smoking is increasing among people with common mental health and substance use problems in the United States: data from representative samples of US adults, 2005-2014. J Clin Psychiatry 2018 Aug 14;79(5):57 [FREE Full text] [doi: 10.4088/JCP.17m11945] [Medline: 30153404]
13. Inoue-Choi M, Christensen CH, Rostron BL, Cosgrove CM, Reyes-Guzman C, Apelberg B, et al. Dose-response association of low-intensity and nondaily smoking with mortality in the United States. JAMA Netw Open 2020 Jun 01;3(6):e206436 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.6436] [Medline: 32492162]
14. U.S. Department of Health and Human Services, National Cancer Insitute. What's in a Name? Examination of Light and Intermittent Smokers: Helping to Set a Tobacco Control Research Agenda. Washington, DC: National Institutes of Health; 2008.
15. Inoue-Choi M, McNeel TS, Hartge P, Caporaso NE, Graubard BI, Freedman ND. Non-daily cigarette smokers: mortality risks in the U.S. Am J Prev Med 2019 Jan;56(1):27-37 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.025] [Medline: 30454906]
16. Clinical Practice Guideline Treating Tobacco Use and Dependence 2008 Update Panel, Liaisons,Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. public health service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
17. Rutten LJ, Augustson EM, Doran KA, Moser RP, Hesse BW. Health information seeking and media exposure among smokers: a comparison of light and intermittent tobacco users with heavy users. Nicotine Tob Res 2009 Feb;11(2):190-196 [FREE Full text] [doi: 10.1093/ntr/ntn019] [Medline: 19264865]
18. Tong EK, Ong MK, Vittinghoff E, Pérez-Stable EJ. Nondaily smokers should be asked and advised to quit. Am J Prev Med 2006 Jan;30(1):23-30. [doi: 10.1016/j.amepre.2005.08.048] [Medline: 16414420]
19. Berg CJ, Sutfin EL, Mendel J, Ahluwalia JS. Use of and interest in smoking cessation strategies among daily and nondaily college student smokers. J Am Coll Health 2012;60(3):194-202 [FREE Full text] [doi: 10.1080/07448481.2011.586388] [Medline: 22420696]
20. Schauer GL, Malarcher AM, Berg CJ. Differences in smoking and cessation characteristics among adult nondaily smokers in the United States: findings from the 2009-2010 National Adult Tobacco Survey. Nicotine Tob Res 2014 Jan;16(1):58-68. [doi: 10.1093/ntr/ntt113] [Medline: 23925825]
21. Levy DE, Biener L, Rigotti NA. The natural history of light smokers: a population-based cohort study. Nicotine Tob Res 2009 Feb;11(2):156-163 [FREE Full text] [doi: 10.1093/ntr/ntp011] [Medline: 19264862]
22. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013 Aug 16;15(8):e171 [FREE Full text] [doi: 10.2196/jmir.2636] [Medline: 23956024]

XSL•FO
RenderX

23. Cooper TV, Taylor T, Murray A, DeBon MW, Vander Weg MW, Klesges RC, et al. Differences between intermittent and light daily smokers in a population of U.S. military recruits. Nicotine Tob Res 2010 May;12(5):465-473 [FREE Full text] [doi: 10.1093/ntr/ntq025] [Medline: 20203108]
24. Shiffman S, Scholl SM, Mao J, Ferguson SG, Hedeker D, Primack B, et al. Using nicotine gum to assist nondaily smokers in quitting: a randomized clinical trial. Nicotine Tob Res 2020 Mar 16;22(3):390-397 [FREE Full text] [doi: 10.1093/ntr/ntz090] [Medline: 31125988]
25. Nollen NL, Cox LS, Mayo MS, Ellerbeck EF, Ahluwalia JS. Counseling alone or in combination with nicotine replacement therapy for treatment of black non-daily smokers: a randomized trial. Addiction 2020 Aug 06;115(8):1547-1560. [doi: 10.1111/add.14948] [Medline: 31899564]
26. Hoeppner BB, Hoeppner SS, Carlon HA, Abry A, Darville A, Rohsenow DJ. Preparing for the quit day: comparing beliefs of nondaily versus daily young adult smokers as they prepare for a quit attempt. Nicotine Tob Res 2021 May 24;23(6):1038-1046 [FREE Full text] [doi: 10.1093/ntr/ntaa166] [Medline: 32882037]
27. Scheuermann TS, Thomas JL, Hayes RB, Harwood EM, Shanafelt AP, Wolff ER, et al. A qualitative exploration of smoking influences and quit attempts among nondaily smokers. Health Behav Policy Rev 2014 May 15;1(3):172-182. [doi: 10.14485/hbpr.1.3.1]
28. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
29. Hoeppner BB, Hoeppner SS, Kelly L, Schick M, Kelly JF. Smiling instead of smoking: development of a positive psychology smoking cessation smartphone app for non-daily smokers. Int J Behav Med 2017 Oct;24(5):683-693. [doi: 10.1007/s12529-017-9640-9] [Medline: 28197846]
30. Hoeppner BB, Hoeppner SS, Carlon HA, Perez GK, Helmuth E, Kahler CW, et al. Leveraging positive psychology to support smoking cessation in nondaily smokers using a smartphone app: feasibility and acceptability study. JMIR Mhealth Uhealth 2019 Jul 03;7(7):e13436 [FREE Full text] [doi: 10.2196/13436] [Medline: 31271147]
31. Hoeppner BB, Siegel KR, Dickerman SR, Todi AA, Kahler CW, Park ER, et al. Testing the outcomes of a smoking cessation smartphone app for nondaily smokers: protocol for a proof-of-concept randomized controlled trial. JMIR Res Protoc 2023 Feb 14;12:e40867 [FREE Full text] [doi: 10.2196/40867] [Medline: 36787172]
32. Kahler CW, Spillane NS, Day A, Clerkin EM, Parks A, Leventhal AM, et al. Positive psychotherapy for smoking cessation: treatment development, feasibility and preliminary results. J Posit Psychol 2014 Jan 01;9(1):19-29 [FREE Full text] [doi: 10.1080/17439760.2013.826716] [Medline: 24683417]
33. Kahler CW, Spillane NS, Day AM, Cioe PA, Parks A, Leventhal AM, et al. Positive psychotherapy for smoking cessation: a pilot randomized controlled trial. Nicotine Tob Res 2015 Nov 02;17(11):1385-1392 [FREE Full text] [doi: 10.1093/ntr/ntv011] [Medline: 25646352]
34. Klemperer EM, Hughes JR, Peasley-Miklus CE, Callas PW, Cook JW, Streck JM, et al. Possible new symptoms of tobacco withdrawal III: reduced positive affect-a review and meta-analysis. Nicotine Tob Res 2021 Jan 22;23(2):259-266 [FREE Full text] [doi: 10.1093/ntr/ntaa044] [Medline: 32188995]
35. Hoeppner BB, Kahler CW, Gwaltney CJ. Relationship between momentary affect states and self-efficacy in adolescent smokers. Health Psychol 2014 Dec;33(12):1507-1517 [FREE Full text] [doi: 10.1037/hea0000075] [Medline: 25020151]
36. Shiffman S, Dunbar MS, Kirchner TR, Li X, Tindle HA, Anderson SJ, et al. Cue reactivity in non-daily smokers: effects on craving and on smoking behavior. Psychopharmacology (Berl) 2013 Mar;226(2):321-333 [FREE Full text] [doi: 10.1007/s00213-012-2909-4] [Medline: 23142992]
37. Rabois D, Haaga DA. The influence of cognitive coping and mood on smokers' self-efficacy and temptation. Addict Behav 2003 Apr;28(3):561-573. [doi: 10.1016/s0306-4603(01)00249-0] [Medline: 12628627]
38. Das E, Vonkeman C, Hartmann T. Mood as a resource in dealing with health recommendations: how mood affects information processing and acceptance of quit-smoking messages. Psychol Health 2012;27(1):116-127 [FREE Full text] [doi: 10.1080/08870446.2011.569888] [Medline: 21678163]
39. Hoeppner BB, Siegel KR, Carlon HA, Kahler CW, Park ER, Hoeppner SS. A smoking cessation app for nondaily smokers (version 2 of the smiling instead of smoking app): acceptability and feasibility study. JMIR Form Res 2021 Nov 17;5(11):e29760 [FREE Full text] [doi: 10.2196/29760] [Medline: 34787577]
40. Hoepper BB, Siegel KR, Carlon HA, Kahler CW, Park ER, Taylor ST, et al. Feature-level analysis of a smoking cessation smartphone app based on a positive psychology approach: prospective observational study. JMIR Form Res 2022 Jul 28;6(7):e38234 [FREE Full text] [doi: 10.2196/38234] [Medline: 35900835]
41. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991 Dec;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
42. Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. J Exp Soc Psychol 1986 Sep;22(5):453-474. [doi: 10.1016/0022-1031(86)90045-4]
43. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983;51(3):390-395. [doi: 10.1037//0022-006X.51.3.390]

XSL•FO
RenderX

44. Marlatt GA, Donovan DM. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. 2nd edition. New York, NY: The Guilford Press; 2005.
45. Hoeppner BB, Hoeppner SS, Abroms LC. How do text-messaging smoking cessation interventions confer benefit? A multiple mediation analysis of Text2Quit. Addiction 2017 Apr 12;112(4):673-682 [FREE Full text] [doi: 10.1111/add.13685] [Medline: 27943511]
46. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
47. Hoeppner BB, Schick MR, Carlon H, Hoeppner SS. Do self-administered positive psychology exercises work in persons in recovery from problematic substance use? An online randomized survey. J Subst Abuse Treat 2019 Apr;99:16-23. [doi: 10.1016/j.jsat.2019.01.006] [Medline: 30797389]
48. Vilardaga R, Rizo J, Ries RK, Kientz JA, Ziedonis DM, Hernandez K, et al. Formative, multimethod case studies of learn to quit, an acceptance and commitment therapy smoking cessation app designed for people with serious mental illness. Transl Behav Med 2019 Nov 25;9(6):1076-1086 [FREE Full text] [doi: 10.1093/tbm/iby097] [Medline: 30445507]
49. Vilardaga R, Rizo J, Palenski PE, Mannelli P, Oliver JA, Mcclernon FJ. Pilot randomized controlled trial of a novel smoking cessation app designed for individuals with co-occurring tobacco use disorder and serious mental illness. Nicotine Tob Res 2020 Aug 24;22(9):1533-1542 [FREE Full text] [doi: 10.1093/ntr/ntz202] [Medline: 31667501]
50. Janes AC, Datko M, Roy A, Barton B, Druker S, Neal C, et al. Quitting starts in the brain: a randomized controlled trial of app-based mindfulness shows decreases in neural responses to smoking cues that predict reductions in smoking. Neuropsychopharmacology 2019 Aug;44(9):1631-1638 [FREE Full text] [doi: 10.1038/s41386-019-0403-y] [Medline: 31039580]
51. Hébert ET, Ra CK, Alexander AC, Helt A, Moisiuc R, Kendzor DE, et al. A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial. J Med Internet Res 2020 Mar 09;22(3):e16907 [FREE Full text] [doi: 10.2196/16907] [Medline: 32149716]
52. Bricker JB, Watson NL, Heffner JL, Sullivan B, Mull K, Kwon D, et al. A smartphone app designed to help cancer patients stop smoking: results from a pilot randomized trial on feasibility, acceptability, and effectiveness. JMIR Form Res 2020 Jan 17;4(1):e16652 [FREE Full text] [doi: 10.2196/16652] [Medline: 31951215]
53. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
54. Clearing the air. National Cancer Insitute, U.S. Department of Health and Human Services. 2008. URL: https://www.cancer.gov/publications/patient-education/clearing-the-air-pdf [accessed 2022-06-08]
55. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
56. Abroms L, Windsor R, Simons-Morton B. Getting young adults to quit smoking: a formative evaluation of the X-pack program. Nicotine Tob Res 2008 Jan;10(1):27-33 [FREE Full text] [doi: 10.1080/14622200701767852] [Medline: 18188742]
57. Jose PE, Lim BT, Bryant FB. Does savoring increase happiness? A daily diary study. J Posit Psychol 2012 May;7(3):176-187. [doi: 10.1080/17439760.2012.671345]
58. Prokhorov AV, Yost T, Mullin-Jones M, de Moor C, Ford KH, Marani S, et al. "Look at your health": outcomes associated with a computer-assisted smoking cessation counseling intervention for community college students. Addict Behav 2008 Jun;33(6):757-771 [FREE Full text] [doi: 10.1016/j.addbeh.2007.12.005] [Medline: 18280668]
59. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
60. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
61. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, CO: Measuring Usability LLC; 2011.
62. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
63. Cox LS, Tiffany ST, Christen AG. Evaluation of the brief questionnaire of smoking urges (QSU-brief) in laboratory and clinical settings. Nicotine Tob Res 2001 Feb;3(1):7-16. [doi: 10.1080/14622200020032051] [Medline: 11260806]
64. Etter JF, Humair JP, Bergman MM, Perneger TV. Development and validation of the Attitudes Towards Smoking scale (ATS-18). Addiction 2000 Apr;95(4):613-625. [doi: 10.1046/j.1360-0443.2000.95461312.x] [Medline: 10829336]
65. Ward RM, Velicer WF, Rossi JS, Fava JL, Prochaska JO. Factorial invariance and internal consistency for the decisional balance inventory--short form. Addict Behav 2004 Jul;29(5):953-958. [doi: 10.1016/j.addbeh.2004.02.042] [Medline: 15219341]

XSL•FO
RenderX

66. Heatherton TF, Kozlowski LT, Frecker RC, Rickert W, Robinson J. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. Br J Addict 1989 Jul;84(7):791-799. [doi: 10.1111/j.1360-0443.1989.tb03059.x] [Medline: 2758152]
67. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]
68. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
69. Snaith RP, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P. A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. Br J Psychiatry 1995 Jul;167(1):99-103. [doi: 10.1192/bjp.167.1.99] [Medline: 7551619]
70. Cohen J. Quantitative methods in psychology: a power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
71. Cook RJ, Farewell VT. Multiplicity considerations in the design and analysis of clinical trials. J R Stat Soc Ser A Stat Soc 1996;159(1):93. [doi: 10.2307/2983471]
72. Shiffman S, Dunbar MS, Li X, Scholl SM, Tindle HA, Anderson SJ, et al. Smoking patterns and stimulus control in intermittent and daily smokers. PLoS One 2014;9(3):e89911 [FREE Full text] [doi: 10.1371/journal.pone.0089911] [Medline: 24599056]
73. Public use data release. National Health Interview Survey (NHIS). URL: https://www.cdc.gov/nchs/nhis/2020nhis.htm [accessed 2024-04-29]
74. Shiffman S, Tindle H, Li X, Scholl S, Dunbar M, Mitchell-Miland C. Characteristics and smoking patterns of intermittent smokers. Exp Clin Psychopharmacol 2012 Aug;20(4):264-277 [FREE Full text] [doi: 10.1037/a0027546] [Medline: 22390208]
75. Calton BA, Nouri S, Davila C, Kotwal A, Zapata C, Bischoff KE. Strategies to make telemedicine a friend, not a foe, in the provision of accessible and equitable cancer care. Cancers (Basel) 2023 Oct 24;15(21):5121 [FREE Full text] [doi: 10.3390/cancers15215121] [Medline: 37958296]
76. Smith KA, Blease C, Faurholt-Jepsen M, Firth J, Van Daele T, Moreno C, et al. Digital mental health: challenges and next steps. BMJ Ment Health 2023 Feb;26(1):e300670 [FREE Full text] [doi: 10.1136/bmjment-2023-300670] [Medline: 37197797]
77. Hebert KK, Cummins SE, Hernández S, Tedeschi GJ, Zhu SH. Current major depression among smokers using a state quitline. Am J Prev Med 2011 Jan;40(1):47-53 [FREE Full text] [doi: 10.1016/j.amepre.2010.09.030] [Medline: 21146767]
78. Tedeschi GJ, Cummins SE, Anderson CM, Anthenelli RM, Zhuang Y, Zhu SH. Smokers with self-reported mental health conditions: a case for screening in the context of tobacco cessation services. PLoS One 2016;11(7):e0159127 [FREE Full text] [doi: 10.1371/journal.pone.0159127] [Medline: 27391334]
79. Niaura R, Britt DM, Shadel WG, Goldstein M, Abrams D, Brown R. Symptoms of depression and survival experience among three samples of smokers trying to quit. Psychol Addict Behav 2001 Mar;15(1):13-17. [doi: 10.1037/0893-164x.15.1.13] [Medline: 11255933]
80. Cinciripini PM, Wetter DW, Fouladi RT, Blalock JA, Carter BL, Cinciripini LG, et al. The effects of depressed mood on smoking cessation: mediation by postcessation self-efficacy. J Consult Clin Psychol 2003 Apr;71(2):292-301. [doi: 10.1037/0022-006x.71.2.292] [Medline: 12699023]

Abbreviations

Brief-QSU: Brief Questionnaire of Smoking Urges
CSQ: Client Satisfaction Questionnaire
CtA: Clearing the Air
DCB-SF: Decisional Balance Inventory for Smoking short form
NCI: National Cancer Institute
PPA: point prevalence abstinence
QG: QuitGuide
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SiS: Smiling instead of Smoking
SUS: System Usability Scale

XSL•FO
RenderX

Edited by L Buis; submitted 26.10.23; peer-reviewed by S Shiffman, D Chinwong, S Chinwong, S Alexander; comments to author 30.04.24; revised version received 25.07.24; accepted 14.11.24; published 15.01.25.

Please cite as:
Hoeppner BB, Siegel KR, Futter AE, Finley-Abboud D, Williamson AC, Kahler CW, Park ER, Hoeppner SS
Smoking Cessation Smartphone App for Nondaily Smoking With Telephone Onboarding: Proof-of-Concept Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e53971
URL: https://mhealth.jmir.org/2025/1/e53971
doi:10.2196/53971
PMID:

©Bettina B Hoeppner, Kaitlyn R Siegel, Allison E Futter, Diadora Finley-Abboud, Alivia C Williamson, Christopher W Kahler, Elyse R Park, Susanne S Hoeppner. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Proximal Effects of a Just-in-Time Adaptive Intervention for Smoking Cessation With Wearable Sensors: Microrandomized Trial

Christine Vinci[1], PhD; Steve K Sutton[1], PhD; Min-Jeong Yang[1], PhD; Sarah R Jones[1], BS; Santosh Kumar[2], PhD; David W Wetter[3], PhD

[1]Moffitt Cancer Center, Tampa, FL, United States

[2]University of Memphis, Memphis, TN, United States

[3]University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Christine Vinci, PhD
Moffitt Cancer Center
12902 Magnolia Dr.
Tampa, FL, 33612
United States
Phone: 1 813 745 5421
Fax: 1 813 449 6871
Email: Christine.vinci@moffitt.org

Abstract

Background: Tobacco use remains the leading preventable cause of morbidity and mortality in the United States. Novel interventions are needed to improve smoking cessation rates. Mindfulness-based interventions (MBIs) for cessation address tobacco use by increasing awareness of the automatic nature of smoking and related behaviors (eg, reactivity to triggers for smoking) from a nonjudgmental stance. Delivering MBIs for smoking cessation via innovative technologies allows for flexibility in the timing of intervention delivery, which has the potential to improve the efficacy of cessation interventions. Research shows MBIs target key mechanisms in the smoking cessation process and can be used to minimize drivers of smoking lapse.

Objective: This single-arm study investigated the impact of mindfulness-based strategies and motivational messages on proximal outcomes, collected via ecological momentary assessment (EMA), relevant to tobacco abstinence via a microrandomized trial. This approach allows for the evaluation of intervention content on proximal outcomes (eg, reduced negative affect) that are thought to impact positive distal outcomes (eg, smoking abstinence).

Methods: All participants were motivated to quit smoking, and the intervention they received included nicotine replacement therapy, brief individual counseling, and a 2-week Just-in-Time Adaptive Intervention (JITAI) with wearable sensors. Throughout the JITAI period, a single strategy was randomly pushed (vs not) multiple times per day through the smartphone application. An EMA next assessed negative affect, positive affect, mindfulness, abstinence self-efficacy, motivation to quit, craving, and smoking motives. The primary analyses evaluated differences in EMA outcomes (proximal) for when a strategy was pushed versus not pushed. Additional analyses evaluated changes in similar outcomes collected from surveys at the baseline and end-of-treatment visits.

Results: Participants (N=38) were 63% (24/38) female, 18% (7/38) Hispanic or Latino, and 29% (11/38) African American. They had an average age of 49 years and smoked an average of 15 (SD 7.9) cigarettes per day. Results indicated that receiving the JITAI significantly reduced proximal negative affect in the second (and final) week of the intervention. Self-reports provided at baseline and end of treatment showed significant decreases in perceived stress, automaticity of smoking and craving, and a significant increase in abstinence self-efficacy. Increases in abstinence self-efficacy significantly predicted abstinence.

Conclusions: To our knowledge, this is the first study to test the proximal impact of a mindfulness-based JITAI on key variables associated with smoking cessation. Our primary finding was that negative affect was lower following the completion of a strategy (vs when no strategy was delivered) in the final week of the JITAI. Among a larger sample size, future research should extend the length of the intervention to further evaluate the impact of the JITAI, as well as include a comparison condition to further evaluate how each component of the intervention uniquely impacts outcomes.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT03404596; https://clinicaltrials.gov/study/NCT03404596

(JMIR Mhealth Uhealth 2025;13:e55379) doi:10.2196/55379

KEYWORDS

smoking cessation; mindfulness; ecological momentary assessment; micro-randomized trial; Just-in-Time Adaptive Intervention; JITAI; EMA; ecological momentary; smoking; smokers; quitting; cessation; meditation; mind body; sensors; motivational; tobacco; nicotine; NRT; counseling; wearables; abstinence; stress; craving; adaptive intervention; mobile phone

Introduction

Tobacco use remains the leading preventable cause of morbidity and mortality in the United States. Although the prevalence of smoking has decreased dramatically over the past several decades, 12.5% of the US adult population continues to smoke [1]. Novel smoking cessation interventions are needed, and one approach is to target key precipitants of lapse and relapse that may ultimately lead to long-term abstinence. Mindfulness-based interventions (MBIs) for smoking cessation address tobacco use by increasing awareness of the automatic nature of smoking and related behaviors (eg, reactivity to triggers for smoking) from a nonjudgmental stance [2-6]. Formal and informal mindfulness practices are taught and practiced, and individuals are encouraged to implement these practices while quitting smoking. MBIs have been shown to both reduce negative affect and decouple the link between negative affect and craving or smoking behavior [7-13].

Although MBIs have demonstrated efficacy for smoking cessation via traditional in-person delivery [2-4,6], requiring individuals to attend treatment solely in-person can be burdensome and may result in high attrition. As such, various mHealth MBIs delivered via smartphone have been developed, both as module-based apps and apps that adapt to the unique circumstances of the individual [14-16]. For instance, Craving to Quit, a module-based mindfulness smartphone app, showed promise, although there were no significant differences between craving to quit and the comparison condition (inactive attention control) on biochemically-verified abstinence at 6 months (9.8% vs 12.1%) [15]. In our recent publication of a Just-in-Time MBI for smoking cessation delivered via smartphone, results showed high acceptability and moderate feasibility; biochemically-confirmed abstinence was promising with 34% abstinent at the end of treatment and 21% one month later [16]. The primary focus of this study is to examine the proximal impact of this MBI on intermediary variables (eg, negative affect) of tobacco abstinence via a microrandomized trial (MRT), using the same sample of participants as in [16].

MRTs can allow for the evaluation of a Just-in-Time Adaptive Intervention (JITAI) on proximal outcomes (eg, negative affect) that are thought to mediate distal outcomes (eg, smoking abstinence) [17]. In most cases, MRTs aid in the refinement of intervention components. This includes assessment and evaluation of proximal outcomes over time and across situations. Since smoking lapses are very common during a quit attempt [18-22] and most occur within 1 week of quitting [18,21,22], the JITAI tested in this study was designed to address quitting smoking in the first 2 weeks of a quit attempt. This JITAI was created to present an MBI when triggered (1) by momentary negative affect (NA; both high and low) given it is a key precipitant of lapse [23-27] and (2) by smoking behavior itself. Both NA and smoking were designed to be detected in real-time on a smartphone via the artificial intelligence algorithms cStress and puffMarker, respectively, using data captured by the AutoSense wearable wireless sensor suite [28-30] which was worn throughout the study.

As described in more detail below and elsewhere [16], technology-related issues resulted in challenges regarding the detection of high NA and active smoking, despite extensive testing of the app before data collection. Our experience is consistent with recent reports showing the ability to detect rare events using mHealth apps is a challenge [31-33]. Once we determined that high NA and active smoking were not detected as planned, we focused analyses on all other detected moments (low NA and no smoking periods; additional details are provided below).

Individuals who smoked and were treatment-seeking received the JITAI and brief counseling, which were implemented in the first 2 weeks, along with 6 weeks of nicotine replacement therapy (NRT). Mindfulness-based content delivered via the JITAI was the primary treatment element. Data for the proximal effects of the MRT was acquired via ecological momentary assessment (EMA). That is when an eligible moment of NA or smoking was detected from sensor data, that moment was randomized to either deliver a mindfulness or motivational strategy or not. Motivational messages were included due to high desirability in previous mHealth cessation studies [34-36]. An EMA was pushed immediately after the participant completed the strategy or 5 minutes after the strategy was pushed, whichever came first.

This study focuses on the proximal effects of the JITAI on targeted outcomes collected via EMA among adherent participants, previously defined as those who wore the sensors for more than 70% of the time during the 2-week period and completed a majority of the strategies [16]. Our rationale for examining adherent participants is 2-fold. First, this approach ensured that our analysis had EMA data from participants that covered most if not all, days across the 2-week period and for different moments triggered by the JITAI. Second, given this study presents the first test of the JITAI, focusing our analysis among those who received a sufficient dose of the JITAI allows us to determine if a “signal” of the intervention on proximal outcomes is present. That is, if no signal was detected even among those who adhered to the intervention, then no further study of the intervention would be indicated. On the other hand, some signals would suggest that further study of the JITAI is warranted, along with addressing adherence in future studies. We selected the following outcomes based on previous research

XSL•FO
RenderX

indicating their importance in the prediction of lapse [8,10,21,24,37-41] and their potential to be impacted by mindfulness [10,42-45]. We hypothesized that following the delivery of a strategy (vs none) participants would report higher positive affect (PA [43,46-48]), mindfulness [15,49-51], abstinence self-efficacy [44,45,52,53], and motivation to quit [54,55], and report lower NA [3,44,48,56-59], craving [9,11,44,51,60,61], and smoking expectancies [49,62,63].

Additional analyses focused on changes from baseline through end of treatment on conceptually similar constructs (affect, mindfulness, abstinence self-efficacy, and smoking motives) using self-report survey data collected at the in-person baseline and end of treatment visits to capture change over the entire intervention. We present results from the entire sample, as well as the subset of participants defined as adherent. In sum, this paper is the first to report the results of a mindfulness-based JITAI on key variables central to smoking cessation.

Methods

Overview

We provide here a high-level overview of the study procedure given the multiple study visits and intervention components, with details presented below. After consent, participants were scheduled to attend 5 in-person visits, which consisted of delivering intervention content and study questionnaires [64]. Visit 1 was scheduled 4 days before quit day, visit 2 was on the quit day, visit 3 was 3 days postquit day, visit 4 was 10 days postquit day, and visit 5 was 28 days postquit day. Visits 1-4 included the first 2-week period when participants wore the study equipment, received the JITAI, completed EMAs, and received brief counseling. Visit 4 is referred to as the "end of treatment" throughout the manuscript. Visit 5 only included study questionnaires. Participants received a 6-week supply of NRT to be used at visit 2 (quit day). Refer to Table 1 for the overall study design.

Table 1. Study design.

	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5
Day	–4	Day 0 (Quit Day)	+3	+10 (End of treatment)	+28
Intervention					
Brief counseling	✓	✓	✓	✓	N/A[a]
JITAI[b]	✓	✓	✓	✓	N/A
NRT[c]		✓	✓	✓	N/A
Assessments					
Questionnaires	✓	✓	✓	✓	✓
EMAs[d]	✓	✓	✓	✓	N/A

[a]Not applicable.

[b]JITAI: Just-in-Time Adaptive Intervention.

[c]NRT: nicotine replacement therapy.

[d]EMA: ecological momentary assessment.

Participants

Inclusion criteria were (1) age of 18 years or older; (2) having smoked≥ 3 cigarettes per day for the past year; a carbon monoxide (CO) reading of ≥6 ppm; (3) motivated to quit smoking in the next 30 days; (4) a valid home address and functioning phone number; and (5) being able to read, write, and speak English. Exclusion criteria were (1) contraindication for wearing the nicotine patch, (2) psychosis, (3) having an implanted cardiac rhythm device, (4) use of other tobacco products (e-cigarette use okay), (5) pregnancy or lactation, (6) being unable to wear the sensors or provide good readings of physiological data, (7) actively trying to quit smoking, (8) having another member of the household enrolled in the study, and (9) no previous use of a smartphone.

Measures

Overview

Demographics (eg, sex and age) were collected at baseline, and questionnaire data were collected at the 5 in-person visits [64].

EMA data was collected via the study smartphone app during the 2-week JITAI. Tobacco use was self-reported by participants via the Timeline Followback [65,66] and abstinence was biochemically verified by a CO reading of <6 ppm at each visit. The primary tobacco abstinence outcome was a biochemically-confirmed 7-day point prevalence of abstinence at the end of treatment and follow-up, as reported elsewhere [16].

The EMA

EMAs were sent to participants about 50% of the time (to reduce burden) after a randomized push or no push of the JITAI. All EMAs presented the same questions, which took 3-5 minutes to complete. EMA variables examined in this analysis included: NA, PA, smoking-related variables (craving, abstinence self-efficacy, motivation to quit, and smoking expectancies), and state mindfulness. All variables were rated on a 5-point Likert scale (1=strongly disagree through 5=strongly agree).

XSL•FO
RenderX

Negative Affect and Positive Affect

Each NA and PA item was prefaced with "Right now, I feel…" Responses to 5 items were averaged for NA: anxious, ashamed, guilty, irritable, and restless. Four responses were averaged for PA: active, determined, enthusiastic, and proud.

Smoking-Related Variables

Individual items assessed craving ("I have an urge to smoke"), abstinence self-efficacy ("I am confident in my ability to be abstinent from cigarettes"), motivation to quit ("I am very motivated to be abstinent from cigarettes"), and smoking expectancies ("Smoking would improve my mood, be pleasurable, or help me cope with this situation").

State Mindfulness

Each state mindfulness item was prefaced with "When the phone beeped…" The 6 items assessed 3 constructs: attention ("I was doing things automatically, without being aware of what I was doing" and "I found myself doing things without paying attention"), nonjudgment ("I was telling myself that I shouldn't be thinking the way I'm thinking" and "I was thinking that some of my emotions are bad or inappropriate and I shouldn't feel them"), and decentering ("I was experiencing my thoughts more as events in my mind than a direct reflection of the way things really are" and "I was experiencing my changing thoughts and feelings as separate from myself").

Questionnaires

The Positive Affect and Negative Affect Scale (PANAS) is a self-report measure that assesses PANAS-PA and PANAS-NA [67]. The Perceived Stress Scale (PSS) measures the degree to which participants find their lives to be stressful [68,69]. The Five Facet Mindfulness Questionnaire (FFMQ) assesses 5 components of mindfulness including observing, describing, awareness, nonjudgment, and nonreactivity [70]. The Self-Efficacy Scale – Smoking (SES) determines an individual's level of confidence for not smoking in positive or social situations, negative affect situations, and out of habit [71]. Finally, 2 subscales from the Brief Wisconsin Inventory of Smoking Dependence Motives were analyzed to assess smoking motives related to automaticity and craving [72]. The automaticity subscale was chosen due to its conceptual overlap with mindfulness (ie, automaticity should theoretically reduce overtime via mindfulness) [73,74] and craving was chosen due to its association with lapse [24,75,76].

Procedures

Participants were recruited from the community via advertisements and contacted the study team to complete an initial phone screen, where they were provided with general information about the study. Those eligible and interested were invited to an in-person orientation session where more detail about the study was provided and eligibility screening was finalized (ie, CO, pregnancy test, assessment of psychosis symptoms via the Mini-International Neuropsychiatric Interview [77]). Those still eligible and interested were scheduled for 5 in-person visits.

At visit 1 (4 days prequit), participants completed the informed consent, baseline questionnaires, and CO measurement. They were taught how to wear and use the equipment. The study smartphone was provided with instructions on its use and for strategy and EMA completion. Participants then completed brief counseling (refer to below). Visits 2 (quit day), 3 (3 days postquit day), and 4 (10 days postquit day) included the administration of visit questionnaires, CO measurement, an equipment check, and brief counseling. Use of NRT began at visit 2 and continued for 6 weeks; dosing was based on participants' reported cigarettes smoked per day. At visit 4 (end of treatment), all equipment was returned. At visit 5 (28 days postquit day), participants completed questionnaires and CO measurements, and additional smoking cessation resources were provided.

Ethical Considerations

The study protocol was approved by the institution's institutional review board (MCC19002). Informed consent was obtained from all participants and included the option to opt out of the study at any time. The data collected and results presented in this manuscript were covered within the scope of the informed consent document. Data was deidentified. Participants were compensated at each visit for completing measures (US $10 at orientation, US $30 at visits 1-3; US $50 at visits 4 and 5). Bonus compensation was earned by completion of the EMAs. If participants wore the equipment for at least 60% of the time since the last phone prompt, they received US $1.25 for each EMA completed. If equipment had not been worn at least 60% of the time, they received US $.50 for completing each EMA.

AutoSense

AutoSense [28-30] is a wearable sensor suite that consists of a chest strap with 2 electrodes and 2 wrist sensors worn by participants during the study. A smartphone with the study app collected all data from AutoSense sensors wirelessly. NA was detected from respiration and electrocardiogram data, and the cStress algorithm was used to categorize a given moment as low versus high NA [29]. Smoking status was detected from a combination of respiration and hand-to-mouth movement from wrist sensors. The PuffMarker algorithm determined whether a given moment was smoking versus no smoking [30]. A more detailed description of the cStress and Puffmarker algorithms can be found elsewhere [16,29,30,64].

MRT Design

Specific moments during the day were designed to be randomized to either push intervention content to a participant, or not, and the decision point for randomization was based on smoking status, NA, and availability (which were continuously monitored via AutoSense and the study smartphone). Each day was separated into six 2-hour blocks for the delivery of mindfulness or motivational strategies (ie, the JITAI). For low NA and no smoking, the decision to randomize was based on a randomly chosen time at the beginning of the 2-hour block. If the participant was available and experiencing low NA (or was not smoking) based on data collected from the equipment, randomization occurred. If unavailable (eg, driving), a new time was chosen within the block, and randomization was again attempted at that time. This continued until the end of the block. Within a 2-hour block, randomization was limited to once for

XSL•FO
RenderX

low NA and once for no smoking to limit the burden. Participants could accept, delay, or ignore the strategy, and after completion, participants rated the strategy. An EMA then appeared about 50% of the time following randomization of a strategy and participants had the option to complete, delay, or ignore.

Intervention Content

The JITAI consisted of mindfulness or motivational strategies sent to participants each day of the 2-week period when wearing the equipment. Mindfulness strategies (n=76) were designed to prompt individuals to engage in mindfulness skills at that moment and fell into one of 5 topic areas: breath, thoughts, sensations, acceptance or nonjudgment, and craving. The first 4 topic areas were chosen as they represent the broad areas of focus that mindfulness-based meditations and exercises often fall under [78], while also providing a practical framework to develop statements that individuals can understand and implement with ease in the present moment. Craving was added as a category to provide mindfulness strategies that were tailored to the experience of quitting smoking. Examples included "Turn your attention to your breathing. Notice where you feel your breathing most in your body," "Our thoughts change each moment. Notice how quickly your thoughts may shift around over the next several moments," and "Notice how the clouds in the sky are constantly moving and never stay in the same place. Cravings are similar – they come and go." Motivational messages (n=40) were designed to support motivation to quit smoking, as previous research has identified these types of messages as appealing and helpful for those making a quit attempt [34-36]. Examples included "Every cigarette you don't smoke is money saved. Great job!" and "You got this. Keep moving towards your goal of being tobacco free." Participants could also access the mindfulness strategies on demand (ie, when not prompted) through a button within the app. Another on-demand button addressed 13 frequently asked questions about mindfulness (eg, Is mindfulness the same thing as meditation?).

Brief individually-delivered counseling was consistent with the Treating Tobacco Dependence Guidelines (eg, removing smoking paraphernalia, and social support) [79] and lasted about 20-30 minutes at each visit. Visit 1 provided participants with the National Cancer Institute's Clearing the Air booklet and a single-page handout on mindfulness. During visits 1-3, participants also listened to a 10-minute audio recording of mindfulness meditation (either breath meditation or urge surfing), followed by a brief discussion about the meditation with the study counselor.

Analytic Plan

Descriptive statistics were used to summarize participant characteristics, and to review EMA item responses to identify outliers and patterns of responding that would suggest invalid data (eg, straight-lining). Participants who wore the sensors for at least 8 days and completed at least 60% of the delivered strategies were previously identified as adherent [16]. The primary analyses evaluated the JITAI impact on proximal outcomes among this subset who engaged with the JITAI as intended on self-reported NA, PA, smoking-related variables (craving, abstinence self-efficacy, motivation to quit, and smoking expectancies), and state mindfulness using generalized estimating equations (GEE). The primary model included strategy push (vs no-push), condition (low NA vs not smoking), day of intervention (1-14), and hour of the day (0-23). Three intervention periods were analyzed separately: (1) prequit period (4 days) evaluated differences while still smoking, (2) postquit period (10 days) evaluated differences while attempting to quit, and (3) the final 7 days evaluated differences after 1 week of experience with the intervention and practicing mindfulness.

Unexpected technology-related issues resulted in the detection of very few moments of high NA (the algorithm was too stringent) and smoking (there was a mismatch of orientation parameters in the wrist sensors) among participants [16]. Specifically, 6 EMAs were completed following the detection of high NA and 5 following smoking (less than 2% of all EMAs completed). Thus, we did not include these few moments in the analysis because of the technology-related concerns already noted, as well as the inability of these moments to reasonably represent the intended condition.

Additional analyses focused on the impact of the entire intervention by applying paired sample *t* tests to evaluate changes in PANAS-NA, PANAS-PA, perceived stress, mindfulness, abstinence self-efficacy, and smoking motives (automaticity and craving) at baseline and end of treatment. These analyses were performed for the entire sample (n=38) as well as the subset of participants considered adherent (n=16). Changes in these measures were also evaluated as predictors of abstinence at the end of treatment using simple logistic regression.

Results

Overview

Of the 43 who consented, four dropped out of the study after visit 1, and one was unable to wear the sensors due to being ill. Of the remaining 38 participants, 24 wore the sensors for at least 8 of the 14 days. Of those 24 participants, 16 participants completed at least 60% of the strategies sent. Table 2 presents participant characteristics for those who completed visit 4 (n=38), as well as adherent participants (n=16). There were no significant group differences between those who were adherent and those who were not (n=22).

There were 704 EMAs completed by the 16 adherent participants. This included 178 EMAs during prequit, 526 during postquit, and 374 during the final 7 days. There were no response patterns observed that necessitated the removal of data before analysis (eg, straight-lining).

XSL•FO
RenderX

Table 2. Participant characteristics.

Variable	Adherent (n=16)	All (n=38)	*P* value[a]
Age (years), mean (SD)	49.6 (13.5)	49.4 (13.5)	.96
Gender: women, n (%)	10 (62.5)	24 (63.2)	.94
Ethnicity: Hispanic or Latinx	3 (18.8%)	7 (18.4%)	.98
Race, n (%)			.79
Black or African American	5 (31.3)	11 (28.9)	
White	11 (68.8)	27 (71.1)	
Relationship status, n (%)			.94
Married or living together as married	6 (37.5)	13 (34.2)	
Single	6 (37.5)	15 (39.5)	
Other	4 (25)	10 (26.3)	
Education: high school diploma, General Education Development, or less, n (%)	5 (31.3)	11 (28.9)	.72
Annual household income <US $30,000, n (%)	9 (56.3)	22 (57.9)	.38
Heaviness of Smoking Index[b], mean (SD)	3.0 (1.4)	2.8 (1.3)	.55
Cigarettes per day, mean (SD)	16.6 (8.5)	15.4 (7.9)	.43

[a]*P* value for statistical test comparing the 16 participants categorized as adherent versus the 22 participants considered not adherent. The *t* test was used for continuous variables and the chi-square test was used for categorical variables.

[b]Heaviness of Smoking Index consisted of 2 items: time to first cigarette and cigarettes per day.

JITAI Proximal Outcomes

Table 3 presents the results from the GEE model of the effect of pushing a strategy on each proximal outcome for each of the three JITAI periods. Multimedia Appendix 1 presents results for all model components. NA was significantly lower after a strategy was pushed (estimated mean [EM] 1.98) versus not pushed (EM 2.07) during the final week of the intervention. No significant differences were observed during the prequit or postquit periods between when the intervention was pushed versus not. PA did not exhibit any statistically significant differences.

Table 3. Results by intervention period among adherent participants (n=16).

Outcome	Intervention period, B[a] (*P* value)		
	Prequit	Postquit	Final week
Negative affect	0.061 (.34)	–0.041 (.20)	–0.090 (.01)[b]
Positive affect	–0.095 (.09)	0.010 (.80)	0.032 (.47)
Motivation	0.063 (.50)	–0.015 (.71)	0.052 (.33)
Abstinence self-efficacy	0.005 (.97)	–0.001 (.98)	0.088 (.07)
Expectancies	–0.011 (.90)	–0.062 (.28)	–0.123 (.051)
Craving	0.055 (.60)	0.017 (.82)	–0.022 (.83)
Attention	–0.083 (.41)	0.014 (.82)	–0.042 (.60)
Non-judgement	0.032 (.51)	0.032 (.35)	0.067 (.10)
Decentering	–0.097 (.32)	0.063 (.10)	0.096 (.08)

[a]B: unstandardized beta.

[b]Statistically significant differences (α=.014).

Refer to Table S1 in Multimedia Appendix 1 for a presentation of all model components.

There were no statistically significant differences for any of the 4 smoking-related variables. However, during the final week, differences trended in the expected direction for expectancies (EMs are 2.41 and 2.53, respectively) and for abstinence self-efficacy (EMs are 3.88 and 3.79), although not significant.

There were no statistically significant differences for any of the 3 mindfulness-related variables. However, during the final week, differences for push versus no-push for decentering trended in

XSL•FO
RenderX

the expected direction, albeit not significant (EMs are 2.13 and 2.04).

Baseline to End of Treatment Changes

Table 4 presents descriptive statistics and results of paired *t* tests. Among adherent participants, a significant increase in PANAS-PA (*P*=.03) and a significant decrease in perceived stress were observed (*P*=.02). The nonjudgment subscale of the FFMQ increased (*P*=.03). All subscales of the Abstinence Self-Efficacy Scale significantly increased: Negative/Affective, Positive/Social, Habit/Addictive, and total score (all *P*s<.001). Regarding smoking motives, both automaticity (*P*<.01) and craving (*P*=.02) significantly decreased.

Among the entire sample, a significant decrease in perceived stress was observed (*P*<.01). All subscales of the Abstinence Self-Efficacy significantly increased: negative or affective, positive or social, habit or addictive, and total score (all *P*s<.001). Both automaticity (*P*<.01) and craving subscales (*P*<.01) significantly decreased. No other significant changes were observed.

In total, 13 of 38 (34%) participants were biochemically confirmed abstinent at the end of treatment [16]. Increases in abstinence self-efficacy (total score) significantly predicted abstinence odd ratio (odds ratio [OR] 2.57, 95% CI 1.21-5.44; *P*=.01). Of all other measures, only PANAS-NA showed a trend in the expected direction as related to abstinence (OR 0.87, 95% CI 0.75-1.02; *P*=.08).

Table 4. Change in affect, mindfulness, and smoking-related variables from baseline to end of treatment.

	Adherent participants (n=16)					Full sample (n=38)				
	Baseline, mean (SD)	End of treatment, mean (SD)	Test statistic			Baseline, mean (SD)	End of treatment, mean (SD)	Test statistic		
			t test (*df*)	*P* values	Cohen *d*			*t* test (*df*)	*P* values	Cohen *d*
Positive and Negative Affect Scale										
Negative affect	16.44 (4.76)	15.75 (4.65)	0.49 (15)	.63	0.12	19.47 (8.74)	18.84 (8.52)	0.74 (37)	.46	0.12
Positive affect	30.38 (8.02)	35.25 (7.93)	–2.33 (15)	.03	0.58	32.55 (9.56)	34.82 (8.74)	–1.81 (37)	.08	0.29
Perceived stress scale	7.38 (2.58)	5.44 (3.08)	2.55 (15)	.02	0.64	7.37 (3.78)	6.11 (3.73)	2.93 (37)	.01	0.48
Five Facet Mindfulness Questionnaire										
Observation	25.73 (8.03)	26.67 (5.73)	–0.60 (14)	.56	0.16	27.32 (6.99)	27.14 (5.76)	0.23 (36)	.82	0.04
Describing	29.93 (5.87)	29.40 (6.19)	0.34 (14)	.74	0.09	29.73 (6.53)	29.89 (6.20)	–0.18 (36)	.86	0.03
Attention or awareness	30.20 (8.55)	29.07 (7.41)	0.60 (14)	.56	0.15	30.11 (7.39)	29.76 (7.01)	0.37 (36)	.72	0.06
Nonjudgment	30.60 (6.85)	33.20 (5.66)	–2.35 (14)	.03	0.61	28.38 (7.85)	30.16 (8.45)	–2.03 (36)	.05	0.33
Nonreactivity	20.33 (6.01)	19.87 (5.51)	0.30 (14)	.77	0.08	21.35 (6.40)	20.84 (5.67)	0.53 (36)	.60	0.09
Self-Efficacy Scale – Smoking										
Negative or effective	1.73 (0.85)	3.19 (0.86)	–5.50 (15)	<.001	1.38	1.70 (0.76)	2.96 (1.20)	–7.14 (37)	<.001	1.16
Positive or social	1.56 (0.76)	3.33 (1.02)	–5.21 (15)	<.001	1.30	1.81 (0.84)	3.30 (1.20)	–7.10 (37)	<.001	1.15
Habit or addictive	1.96 (0.97)	3.67 (0.92)	–5.33 (15)	<.001	1.33	2.14 (1.01)	3.60 (1.21)	–7.14 (37)	<.001	1.16
Total	1.75 (0.78)	3.40 (0.80)	–5.82 (15)	<.001	1.46	1.88 (0.78)	3.28 (1.11)	–7.62 (37)	<.001	1.24
Brief Wisconsin Inventory of Smoking Dependence Motives										
Automaticity	4.23 (1.63)	2.73 (2.05)	3.32 (15)	.01	0.83	4.32 (1.86)	3.16 (2.14)	3.71 (37)	.01	0.60
Craving	4.70 (1.38)	3.34 (1.72)	2.70 (15)	.02	0.67	4.65 (1.32)	3.42 (1.85)	4.08 (37)	<.001	0.66

XSL•FO
RenderX

Discussion

Overview

This study evaluated the momentary impact of mindfulness-based JITAI on proximal outcomes, as well as the broader impact of the entire intervention on conceptually similar constructs assessed via self-report surveys at baseline and end of treatment. The JITAI strategies significantly reduced proximal NA in the second (ie, final) week. The intervention also produced hypothesized changes from baseline to end of treatment with decreases in perceived stress, automaticity of smoking, and craving, and increases in abstinence self-efficacy. Additional increases in PANAS-PA and nonjudgment were observed among adherent participants.

Principal Findings

To our knowledge, this is the first study to test the proximal impact of mindfulness-based JITAI on key variables associated with smoking cessation. Our primary finding was that NA reduced following the completion of a strategy (vs when no strategy was delivered) in the final week of the JITAI. Both mindfulness and motivational strategies were sent, and although the majority were mindfulness, these results suggest that such a combination has a positive effect on NA in the context of quitting smoking. As reported previously [16], participants reported that the just-in-time aspect was one of their favorite parts of the intervention, as well as the ability of the strategies to help them self-regulate, stay present, and manage cravings. Future research should evaluate whether sending a strategy at different moments in time (eg, high vs low NA) results in different proximal effects.

Other proximal outcomes (ie, expectancies that smoking would improve mood, abstinence self-efficacy, decentering) exhibited differences that trended in the expected direction, albeit nonsignificant. Together with the significant reduction in NA, these findings suggest that there may be a cumulative impact of receiving these strategies that results in observable differences in the final week. In other words, increased experience and practice with the strategies result in a more substantial effect later in the intervention period. Future testing of this JITAI should examine this hypothesis by extending the delivery of the JITAI over additional weeks and among a larger sample, as well as including a comparison condition to determine if these findings are unique to this JITAI. For instance, Garrison et al [15] found that craving to quit resulted in increased mindfulness, although this occurred for both the mindfulness and control condition.

This JITAI was delivered in the context of other treatment components, including brief in-person counseling and NRT. When examining how the entire intervention impacted conceptually similar outcomes over time, we observed reductions in perceived stress and smoking motives of automaticity and craving, along with increases in abstinence self-efficacy; additionally, among adherent participants, significant increases in PANAS-PA and the mindfulness construct of nonjudging were observed. Finally, increases in abstinence self-efficacy were significantly predictive of abstinence. These results are consistent with the broader literature on mindfulness interventions for smoking cessation [10,42-45]. For those receiving the JITAI as intended (ie, adherent participants), not only were the additional benefits of increased PA and nonjudgment found, but the effect sizes on these outcomes were larger for this subset of participants. This finding may not be particularly surprising, as we would expect that those who were more adherent would also obtain greater benefits. Given the numerous components of this intervention and the potential to vary the intensity of each, future research may consider an optimization trial (eg, factorial experiment) to examine the unique benefits of each component as related to treatment efficacy and cost-effectiveness.

Results from this study point to the importance of addressing engagement in the context of mHealth interventions. For instance, our previous paper found that adherent participants reported significantly more experience with wearable sensors at baseline, as compared to nonadherent participants [16]. This suggests future research should directly address adherence by (1) allowing participants a few days to gain experience with the equipment before starting the study and/or (2) using less burdensome equipment. Fortunately, rapid advances in technology are making equipment more feasible to wear (eg, wristbands only). Extant research indicates that adding features such as gamification and progress tracking increases adherence and engagement [35,80] (which were not present in this version of the app). Future research should incorporate such features to hopefully increase adherence. Relatedly, although biochemically confirmed abstinence was not statistically different [16], these numbers trended toward higher abstinence among those who were more engaged with the intervention, further underscoring the importance of addressing engagement. Finally, advances in artificial intelligence could address engagement by leveraging patterns of when participants do engage with the app (eg, sending more vs less prompts), current activities (eg, driving), and general schedule (eg, at work so unavailable).

A brief discussion of nonsignificant findings is warranted despite the smaller-than-intended sample size and low statistical power. First, the JITAI intervention did not impact proximal outcomes during prequit nor during the first several days post quitting. One potential explanation is that participants needed more time to practice the intervention strategies. Second, even in the final week of the JITAI, changes were not observed in some variables (eg, aspects of mindfulness, PA, craving). This may be due to limited power to detect changes, that the intervention content (mindfulness and motivational strategies) did not have an immediate impact on these variables, or even that negative affect (relative to the other variables we tested) may be more sensitive to change via brief mindfulness strategies. Future research is needed to investigate these possibilities. Third, we observed changes from baseline to end of treatment on certain variables (eg, positive affect) that we did not see immediately impacted by the strategies. The reverse is also true – NA was immediately impacted following the completion of a strategy within the JITAI, whereas PANAS-NA assessed at baseline and end of treatment did not change. There are various explanations for these conflicting findings including that the variables were measured in different ways (ie, the survey measure items were different than the EMA measures); that one set of measures

XSL•FO
RenderX

captured the impact of the entire intervention across all components over a long period of time whereas the other set captured the immediate impact of a single intervention strategy; and the expectation was that the JITAI would have a small effect on proximal outcomes, although extending the JITAI for another week may have strengthened these effect sizes. Fourth, we did not observe significant changes on all subscales of the FFMQ, which may be due to the brief period of the intervention, the intervention itself, or other measurement-related issues (eg, demand characteristics, specific constructs measured by the FFMQ [81]).

Limitations first include that there was no control condition, as the primary goal was to evaluate the proximal impact of receiving (vs not) a strategy. Thus, it is unclear whether the results may simply reflect the natural history of these outcomes over the first 2 weeks of any cessation intervention or even just the impact of brief counseling and NRT, particularly for changes observed from baseline through the end of treatment. Without a comparison condition sending non-mindfulness or motivational strategies, it is also not possible to conclude that the NA reductions were due to mindfulness or motivational strategies (vs some other type of strategy). Results should be interpreted with caution until future studies can be conducted with a control condition. Second, given the small sample size, additional research is needed to confirm our findings with a larger sample. Third, although our JITAI was deliberately designed to be implemented within the first 2 weeks of quitting smoking to address the period when most smoking lapses occur [18,22,27], we are limited in our ability to draw conclusions on the impact of the intervention beyond this short timeframe. Fourth, given technology issues, we are unable to know with certainty that the low NA and no smoking moments did not include moments of high NA and smoking. As such, we focused our conclusions solely on whether a strategy was pushed or not, although future research using appropriate sensor detection could further investigate such important distinctions.

Conclusion

In sum, this paper reports on the first investigation of a mindfulness-based JITAI on proximal outcomes related to smoking cessation. Key takeaways include the fact that proximal NA was affected by strategies in the final week of the JITAI with other variables trending in the expected direction. One recommendation is to extend the intervention so that participants have more time to learn and practice the strategies. Whereas there were observed changes from baseline to end of treatment for many variables, there needs to be a comparison condition to better understand these changes and evaluate how different components of the intervention (JITAI, in-person counseling, and NRT) impact outcomes. Finally, given that participants who were adherent to wearing the equipment and engaging with the intervention content exhibited larger differences in outcomes, it would be beneficial to address engagement with this type of intervention in future studies.

Acknowledgments

This research was supported by grants from the National Center on Minority Health and Health Disparities (R00MD010468), National Cancer Institute (K99CA271040), the National Institutes of Health (P41EB028242), NIBIB (U54EB020404), through funds provided by the trans-NIH Big Data-to-Knowledge (BD2K) initiative, by the National Science Foundation (NSF) under awards IIS-1722646, ACI-1640813, and CNS-1823221, and the Biostatistics and Bioinformatics Shared Resources at the H. Lee Moffitt Cancer Center and Research Institute (P30-CA076292).

Conflicts of Interest

SK has financial interests in CuesHub, and PBC, including stock ownership and management role.

Multimedia Appendix 1
Results for all model components for each outcome by each intervention period.
[DOCX File , 23 KB - mhealth_v13i1e55379_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1081 KB - mhealth_v13i1e55379_app2.pdf]

References

1. Cornelius ME, Wang TW, Jamal A, Loretan CG, Neff LJ. Tobacco product use among adults - United States, 2019. MMWR Morb Mortal Wkly Rep 2020;69(46):1736-1742 [FREE Full text] [doi: 10.15585/mmwr.mm6946a4] [Medline: 33211681]
2. Brewer JA, Mallik S, Babuscio TA, Nich C, Johnson HE, Deleone CM, et al. Mindfulness training for smoking cessation: results from a randomized controlled trial. Drug Alcohol Depend 2011;119(1-2):72-80 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.05.027] [Medline: 21723049]
3. Davis JM, Fleming MF, Bonus KA, Baker TB. A pilot study on mindfulness based stress reduction for smokers. BMC Complement Altern Med 2007;7(1):2 [FREE Full text] [doi: 10.1186/1472-6882-7-2] [Medline: 17254362]

XSL•FO
RenderX

4. Davis JM, Manley AR, Goldberg SB, Stankevitz KA, Smith SS. Mindfulness training for smokers via web-based video instruction with phone support: a prospective observational study. BMC Complement Altern Med 2015;15(1):95 [FREE Full text] [doi: 10.1186/s12906-015-0618-3] [Medline: 25886752]
5. Vidrine JI, Spears CA, Heppner WL, Reitzel LR, Marcus MT, Cinciripini PM, et al. Efficacy of mindfulness-based addiction treatment (MBAT) for smoking cessation and lapse recovery: a randomized clinical trial. J Consult Clin Psychol 2016;84(9):824-838 [FREE Full text] [doi: 10.1037/ccp0000117] [Medline: 27213492]
6. Vinci C, Sutton SK, Yang M, Baban S, Sauls R, Witkiewitz K, et al. Pilot randomized controlled trial of mindfulness-based relapse prevention vs cognitive behavioral therapy for smoking and alcohol use. Drug Alcohol Depend 2023;244:109768 [FREE Full text] [doi: 10.1016/j.drugalcdep.2023.109768] [Medline: 36645978]
7. Adams CE, Cano MA, Heppner WL, Stewart DW, Correa-Fernández V, Vidrine JI, et al. Testing a moderated mediation model of mindfulness, psychosocial stress, and alcohol use among African American smokers. Mindfulness (N Y) 2015;6(2):315-325 [FREE Full text] [doi: 10.1007/s12671-013-0263-1] [Medline: 25848408]
8. Adams CE, Chen M, Guo L, Lam CY, Stewart DW, Correa-Fernández V, et al. Mindfulness predicts lower affective volatility among African Americans during smoking cessation. Psychol Addict Behav 2014;28(2):580-585 [FREE Full text] [doi: 10.1037/a0036512] [Medline: 24955676]
9. Bowen S, Marlatt A. Surfing the urge: brief mindfulness-based intervention for college student smokers. Psychol Addict Behav 2009;23(4):666-671. [doi: 10.1037/a0017127] [Medline: 20025372]
10. Cambron C, Hopkins P, Burningham C, Lam C, Cinciripini P, Wetter DW. Socioeconomic status, mindfulness, and momentary associations between stress and smoking lapse during a quit attempt. Drug Alcohol Depend 2020;209:107840 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.107840] [Medline: 32058242]
11. Elwafi HM, Witkiewitz K, Mallik S, Thornhill TA, Brewer JA. Mindfulness training for smoking cessation: moderation of the relationship between craving and cigarette use. Drug Alcohol Depend 2013;130(1-3):222-229 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.11.015] [Medline: 23265088]
12. Witkiewitz K, Bowen S. Depression, craving, and substance use following a randomized trial of mindfulness-based relapse prevention. J Consult Clin Psychol 2010;78(3):362-374 [FREE Full text] [doi: 10.1037/a0019172] [Medline: 20515211]
13. Witkiewitz KS, Bowen DM. Depression, craving, and substance use following a randomized trial of mindfulness-based relapse prevention. J Consult Clin Psychol 2010;78(3):362-374 [FREE Full text] [doi: 10.1037/a0019172] [Medline: 20515211]
14. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
15. Garrison K, Pal P, O'Malley SS, Pittman BP, Gueorguieva R, Rojiani R, et al. Craving to Quit: A randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2020;22(3):324-331 [FREE Full text] [doi: 10.1093/ntr/nty126] [Medline: 29917096]
16. Yang M, Sutton SK, Hernandez LM, Jones SR, Wetter DW, Kumar S, et al. A just-in-time adaptive intervention (JITAI) for smoking cessation: feasibility and acceptability findings. Addictive Behaviors 2023;136:107467. [doi: 10.1016/j.addbeh.2022.107467] [Medline: 36037610]
17. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
18. Hughes JR, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 2004;99(1):29-38. [doi: 10.1111/j.1360-0443.2004.00540.x] [Medline: 14678060]
19. Marlatt GA, Curry S, Gordon JR. A longitudinal analysis of unaided smoking cessation. J Consult Clin Psychol 1988;56(5):715-720. [doi: 10.1037//0022-006x.56.5.715] [Medline: 3192787]
20. Piasecki TM, Fiore MC, McCarthy DE, Baker TB. Have we lost our way? The need for dynamic formulations of smoking relapse proneness. Addiction 2002;97(9):1093-1108. [doi: 10.1046/j.1360-0443.2002.00216.x] [Medline: 12199822]
21. Shiffman S, Balabanis MH, Paty JA, Engberg J, Gwaltney CJ, Liu KS, et al. Dynamic effects of self-efficacy on smoking lapse and relapse. Health Psychol 2000;19(4):315-323. [doi: 10.1037//0278-6133.19.4.315] [Medline: 10907649]
22. Zhu SH, Stretch V, Balabanis M, Rosbrook B, Sadler G, Pierce JP. Telephone counseling for smoking cessation: effects of single-session and multiple-session interventions. J Consult Clin Psychol 1996;64(1):202-211. [doi: 10.1037//0022-006x.64.1.202] [Medline: 8907100]
23. Brandon TH. Negative affect as motivation to smoke. Curr Dir Psychol Sci 1994;3(2):33-37. [doi: 10.1111/1467-8721.ep10769919]
24. Cambron C, Haslam AK, Baucom BRW, Lam C, Vinci C, Cinciripini P, et al. Momentary precipitants connecting stress and smoking lapse during a quit attempt. Health Psychol 2019;38(12):1049-1058 [FREE Full text] [doi: 10.1037/hea0000797] [Medline: 31556660]
25. McKee SA, Sinha R, Weinberger AH, Sofuoglu M, Harrison EL, Lavery M, et al. Stress decreases the ability to resist smoking and potentiates smoking intensity and reward. J Psychopharmacol 2011;25(4):490-502 [FREE Full text] [doi: 10.1177/0269881110376694] [Medline: 20817750]

XSL•FO
RenderX

26. Oberleitner LMS, Moore KE, Verplaetse T, Roberts W, McKee SA. Developing a laboratory model of smoking lapse targeting stress and brief nicotine deprivation. Exp Clin Psychopharmacol 2018;26(3):244-250 [FREE Full text] [doi: 10.1037/pha0000187] [Medline: 29863382]
27. Shiffman S, Waters AJ. Negative affect and smoking lapses: a prospective analysis. J Consult Clin Psychol 2004;72(2):192-201. [doi: 10.1037/0022-006X.72.2.192] [Medline: 15065954]
28. Ertin E, Stohs N, Kumar S, Raij, A, al'Absi M, Shah S. AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. 2011 Presented at: in Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems; 2011 November 4; Seattle Washington. [doi: 10.1145/2070942.2070970]
29. Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: towards a gold standard for continuous stress assessment in the mobile environment. Proc ACM Int Conf Ubiquitous Comput 2015;2015:493-504 [FREE Full text] [doi: 10.1145/2750858.2807526] [Medline: 26543926]
30. Saleheen N, Ali AA, Hossain SM, Sarker H, Chatterjee S, Marlin B, et al. puffMarker: A multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. Proc ACM Int Conf Ubiquitous Comput 2015;2015:999-1010 [FREE Full text] [Medline: 26543927]
31. Carpenter SM, Menictas M, Nahum-Shani I, Wetter DW, Murphy SA. Developments in mobile health just-in-time adaptive interventions for addiction science. Curr Addict Rep 2020;7(3):280-290 [FREE Full text] [doi: 10.1007/s40429-020-00322-y] [Medline: 33747711]
32. Epstein DH, Tyburski M, Kowalczyk WJ, Burgess-Hull AJ, Phillips KA, Curtis BL, et al. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. NPJ Digit Med 2020;3(1):26 [FREE Full text] [doi: 10.1038/s41746-020-0234-6] [Medline: 32195362]
33. Nakajima M, Lemieux AM, Fiecas M, Chatterjee S, Sarker H, Saleheen N, et al. Using novel mobile sensors to assess stress and smoking lapse. Int J Psychophysiol 2020;158:411-418 [FREE Full text] [doi: 10.1016/j.ijpsycho.2020.11.005] [Medline: 33189770]
34. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. J Med Internet Res 2016;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]
35. Hartzler AL, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth design space: a mixed-methods analysis of smokers' perspectives. JMIR Mhealth Uhealth 2016;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
36. Yingst JM, Veldheer S, Hrabovsky S, Hammett E, Nicholson J, Berg A, et al. Pilot randomized trial of an automated smoking cessation intervention via mobile phone text messages as an adjunct to varenicline in primary care. J Health Commun 2018;23(4):370-378 [FREE Full text] [doi: 10.1080/10810730.2018.1453890] [Medline: 29578832]
37. Businelle MS, Kendzor DE, Reitzel LR, Costello TJ, Cofta-Woerpel L, Li Y, et al. Mechanisms linking socioeconomic status to smoking cessation: a structural equation modeling approach. Health Psychol 2010;29(3):262-273 [FREE Full text] [doi: 10.1037/a0019285] [Medline: 20496980]
38. Gwaltney CJ, Shiffman S, Balabanis MH, Paty JA. Dynamic self-efficacy and outcome expectancies: prediction of smoking lapse and relapse. J Abnorm Psychol 2005;114(4):661-675. [doi: 10.1037/0021-843X.114.4.661] [Medline: 16351387]
39. Minami H, Yeh VM, Bold KW, Chapman GB, McCarthy DE. Relations among affect, abstinence motivation and confidence, and daily smoking lapse risk. Psychol Addict Behav 2014;28(2):376-388 [FREE Full text] [doi: 10.1037/a0034445] [Medline: 24955665]
40. Spears CA, Li L, Wu C, Vinci C, Heppner WL, Hoover DS, et al. Mechanisms linking mindfulness and early smoking abstinence: an ecological momentary assessment study. Psychol Addict Behav 2019;33(3):197-207 [FREE Full text] [doi: 10.1037/adb0000451] [Medline: 30829517]
41. Vinci C, Li L, Wu C, Lam CY, Guo L, Correa-Fernández V, et al. The association of positive emotion and first smoking lapse: an ecological momentary assessment study. Health Psychol 2017;36(11):1038-1046 [FREE Full text] [doi: 10.1037/hea0000535] [Medline: 28726478]
42. Brewer JA, Elwafi HM, Davis JH. Craving to quit: psychological models and neurobiological mechanisms of mindfulness training as treatment for addictions. Psychol Addict Behav 2013;27(2):366-379 [FREE Full text] [doi: 10.1037/a0028490] [Medline: 22642859]
43. Garland EL, Geschwind N, Peeters F, Wichers M. Mindfulness training promotes upward spirals of positive affect and cognition: multilevel and autoregressive latent trajectory modeling analyses. Front Psychol 2015;6:15 [FREE Full text] [doi: 10.3389/fpsyg.2015.00015] [Medline: 25698988]
44. Spears CA, Hedeker D, Li L, Wu C, Anderson NK, Houchins SC, et al. Supplemental material for mechanisms underlying mindfulness-based addiction treatment versus cognitive behavioral therapy and usual care for smoking cessation. Journal of Consulting and Clinical Psychology 2017;85(11):1029-1040. [doi: 10.1037/ccp0000229.supp]
45. Vidrine JI, Businelle MS, Cinciripini P, Li Y, Marcus MT, Waters AJ, et al. Associations of mindfulness with nicotine dependence, withdrawal, and agency. Subst Abus 2009;30(4):318-327 [FREE Full text] [doi: 10.1080/08897070903252973] [Medline: 19904667]

XSL•FO
RenderX

46. Davidson RJ, Kabat-Zinn J, Schumacher J, Rosenkranz M, Muller D, Santorelli SF, et al. Alterations in brain and immune function produced by mindfulness meditation. Psychosom Med 2003;65(4):564-570. [doi: 10.1097/01.psy.0000077505.67574.e3] [Medline: 12883106]
47. Geschwind N, Peeters F, Drukker M, van Os J, Wichers M. Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: a randomized controlled trial. J Consult Clin Psychol 2011;79(5):618-628. [doi: 10.1037/a0024595] [Medline: 21767001]
48. Tang Y, Ma Y, Wang J, Fan Y, Feng S, Lu Q, et al. Short-term meditation training improves attention and self-regulation. Proc Natl Acad Sci U S A 2007;104(43):17152-17156 [FREE Full text] [doi: 10.1073/pnas.0707678104] [Medline: 17940025]
49. Heppner WL, Spears CA, Correa-Fernández V, Castro Y, Li Y, Guo B, et al. Dispositional mindfulness predicts enhanced smoking cessation and smoking lapse recovery. Ann Behav Med 2016;50(3):337-347 [FREE Full text] [doi: 10.1007/s12160-015-9759-3] [Medline: 26743533]
50. Luberto CM, McLeish AC. The effects of a brief mindfulness exercise on state mindfulness and affective outcomes among adult daily smokers. Addict Behav 2018;77:73-80. [doi: 10.1016/j.addbeh.2017.09.013] [Medline: 28965070]
51. Weiss de Souza IC, Kozasa E, Bowen S, Richter KP, Sartes LMA, Colugnati FAB, et al. Effectiveness of mindfulness-based relapse prevention program as an adjunct to the standard treatment for smoking: a pragmatic design pilot study. Nicotine Tob Res 2020;22(9):1605-1613. [doi: 10.1093/ntr/ntaa057] [Medline: 32222767]
52. Britton WB, Bootzin RR, Cousins JC, Hasler BP, Peck T, Shapiro SL. The contribution of mindfulness practice to a multicomponent behavioral sleep intervention following substance abuse treatment in adolescents: a treatment-development study. Subst Abus 2010;31(2):86-97. [doi: 10.1080/08897071003641297] [Medline: 20408060]
53. Chang VY, Palesh O, Caldwell R, Glasgow N, Abramson M, Luskin F, et al. The effects of a mindfulness‐based stress reduction program on stress, mindfulness self‐efficacy, and positive states of mind. Stress and Health 2004;20(3):141-147. [doi: 10.1002/smi.1011]
54. Jackson S, Brown J, Norris E, Livingstone-Banks J, Hayes E, Lindson N. Mindfulness for smoking cessation. Cochrane Database Syst Rev 2022;4(4):CD013696 [FREE Full text] [doi: 10.1002/14651858.CD013696.pub2] [Medline: 35420700]
55. Witkiewitz K, Marlatt GA. Relapse prevention for alcohol and drug problems: that was Zen, this is Tao. Am Psychol 2004;59(4):224-235. [doi: 10.1037/0003-066X.59.4.224] [Medline: 15149263]
56. Arch JJ, Craske MG. Mechanisms of mindfulness: emotion regulation following a focused breathing induction. Behav Res Ther 2006;44(12):1849-1858. [doi: 10.1016/j.brat.2005.12.007] [Medline: 16460668]
57. Gotink RA, Chu P, Busschbach JJV, Benson H, Fricchione GL, Hunink MGM. Standardised mindfulness-based interventions in healthcare: an overview of systematic reviews and meta-analyses of RCTs. PLoS One 2015;10(4):e0124344 [FREE Full text] [doi: 10.1371/journal.pone.0124344] [Medline: 25881019]
58. Ortner CNM, Kilner SJ, Zelazo PD. Mindfulness meditation and reduced emotional interference on a cognitive task. Motiv Emot 2007;31(4):271-283. [doi: 10.1007/s11031-007-9076-7]
59. Vinci C, Peltier MR, Shah S, Kinsaul J, Waldo K, McVay MA, et al. Effects of a brief mindfulness intervention on negative affect and urge to drink among college student drinkers. Behav Res Ther 2014;59:82-93. [doi: 10.1016/j.brat.2014.05.012] [Medline: 24972492]
60. Adams CE, Benitez L, Kinsaul J, Apperson McVay M, Barbry A, Thibodeaux A, et al. Effects of brief mindfulness instructions on reactions to body image stimuli among female smokers: an experimental study. Nicotine Tob Res 2013;15(2):376-384 [FREE Full text] [doi: 10.1093/ntr/nts133] [Medline: 22987786]
61. Westbrook C, Creswell JD, Tabibnia G, Julson E, Kober H, Tindle HA. Mindful attention reduces neural and self-reported cue-induced craving in smokers. Soc Cogn Affect Neurosci 2013;8(1):73-84 [FREE Full text] [doi: 10.1093/scan/nsr076] [Medline: 22114078]
62. Gonzalez A, Vujanovic AA, Johnson KA, Leyro TM, Zvolensky MJ. The role of mindful attention in regard to the relation between negative affect reduction outcome expectancies and emotional vulnerability among adult cigarette smokers. Cognit Ther Res 2009;33(6):645-656 [FREE Full text] [doi: 10.1007/s10608-009-9246-x] [Medline: 32773908]
63. Vinci C, Spears CA, Peltier MR, Copeland AL. Facets of mindfulness mediate the relationship between depressive symptoms and smoking behavior. Mindfulness (N Y) 2016;7(6):1408-1415 [FREE Full text] [doi: 10.1007/s12671-016-0582-0] [Medline: 28083076]
64. Hernandez LM, Wetter DW, Kumar S, Sutton SK, Vinci C. Smoking cessation using wearable sensors: protocol for a microrandomized trial. JMIR Res Protoc 2021;10(2):e22877 [FREE Full text] [doi: 10.2196/22877] [Medline: 33625366]
65. Sobell L, Sobell M. Alcohol Timeline Followback (TLFB). Handbook of psychiatric measures B2. Handbook of psychiatric measures. Washington DC: American Psychiatric Association; 2000.
66. Sobell LC, Sobell MB. Timeline follow-back: A technique for assessing selfreported alcohol consumption, in Measuring Alcohol Consumption: Psychosocial and Biochemical Methods. Totowa, NJ, US: Humana Press; 1992:41-72.
67. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
68. Cohen S, Lichtenstein E. Perceived stress, quitting smoking, and smoking relapse. Health Psychol 1990;9(4):466-478. [doi: 10.1037//0278-6133.9.4.466] [Medline: 2373070]

XSL•FO
RenderX

69. Warttig SL, Forshaw MJ, South J, White AK. New, normative, English-sample data for the short form perceived stress scale (PSS-4). J Health Psychol 2013;18(12):1617-1628. [doi: 10.1177/1359105313508346] [Medline: 24155195]
70. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
71. Velicer WF, Diclemente CC, Rossi JS, Prochaska JO. Relapse situations and self-efficacy: an integrative model. Addict Behav 1990;15(3):271-283. [doi: 10.1016/0306-4603(90)90070-e] [Medline: 2378287]
72. Smith SS, Piper ME, Bolt DM, Fiore MC, Wetter DW, Cinciripini PM, et al. Development of the brief wisconsin inventory of smoking dependence motives. Nicotine Tob Res 2010;12(5):489-499 [FREE Full text] [doi: 10.1093/ntr/ntq032] [Medline: 20231242]
73. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, et al. Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. Assessment 2008;15(3):329-342. [doi: 10.1177/1073191107313003] [Medline: 18310597]
74. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
75. Businelle MS, Ma P, Kendzor DE, Frank SG, Wetter DW, Vidrine DJ. Using intensive longitudinal data collected via mobile phone to detect imminent lapse in smokers undergoing a scheduled quit attempt. J Med Internet Res 2016;18(10):e275 [FREE Full text] [doi: 10.2196/jmir.6307] [Medline: 27751985]
76. Lam CY, Businelle MS, Aigner CJ, McClure JB, Cofta-Woerpel L, Cinciripini PM, et al. Individual and combined effects of multiple high-risk triggers on postcessation smoking urge and lapse. Nicotine Tob Res 2014;16(5):569-575 [FREE Full text] [doi: 10.1093/ntr/ntt190] [Medline: 24323569]
77. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-34. [Medline: 9881538]
78. Bishop SR, Mark Lau, M, Shauna Shapiro, S, Linda Carlson L, Nicole D. Anderson, ND, James Carmody, J, et al. Mindfulness: a proposed operational definition. Clinical Psychology: Science and Practice 2004;11(3):230-241. [doi: 10.1093/clipsy/bph077]
79. Fiore M. Treating tobacco use and dependence: 2008 update: clinical practice guideline. United States: Diane Publishing; 2009.
80. McClure JB, Hartzler AL, Catz SL. Design Considerations for Smoking Cessation Apps: feedback from nicotine dependence treatment providers and smokers. JMIR Mhealth Uhealth 2016;4(1):e17 [FREE Full text] [doi: 10.2196/mhealth.5181] [Medline: 26872940]
81. Van Dam NT, van Vugt MK, Vago DR, Schmalzl L, Saron CD, Olendzki A, et al. Mind the Hype: a critical evaluation and prescriptive agenda for research on mindfulness and meditation. Perspect Psychol Sci 2018;13(1):36-61 [FREE Full text] [doi: 10.1177/1745691617709589] [Medline: 29016274]

Abbreviations

CO: carbon monoxide
EM: estimated mean
EMA: ecological momentary assessment
FFMQ: Five Facet Mindfulness Questionnaire
GEE: generalized estimating equations
JITAI: Just-in-Time Adaptive Intervention
MBI: mindfulness-based intervention
MRT: microrandomized trial
NA: negative affect
NRT: nicotine replacement therapy
OR: odds ratio
PA: positive affect
PANAS: Positive Affect and Negative Affect Scale
PSS: Perceived Stress Scale
SES: Self-Efficacy Scale – Smoking

XSL•FO
RenderX

Edited by L Buis; submitted 11.12.23; peer-reviewed by N Leonard, H Tateno; comments to author 24.04.24; revised version received 18.06.24; accepted 13.02.25; published 19.03.25.

Please cite as:
Vinci C, Sutton SK, Yang MJ, Jones SR, Kumar S, Wetter DW
Proximal Effects of a Just-in-Time Adaptive Intervention for Smoking Cessation With Wearable Sensors: Microrandomized Trial
JMIR Mhealth Uhealth 2025;13:e55379
URL: https://mhealth.jmir.org/2025/1/e55379
doi:10.2196/55379
PMID:

©Christine Vinci, Steve K Sutton, Min-Jeong Yang, Sarah R Jones, Santosh Kumar, David W Wetter. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of a Mobile Intervention for Procrastination Augmented With a Semigenerative Chatbot for University Students: Pilot Randomized Controlled Trial

Seonmi Lee[1], BS; Jaehyun Jeong[2,3], MS; Myungsung Kim[1], BS; Sangil Lee[2,4], PhD; Sung-Phil Kim[2], PhD; Dooyoung Jung[1,2], MD, PhD

[1]Graduate School of Health Science and Technology, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

[2]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

[3]LVIS Korea, Daegu, Republic of Korea

[4]School of Liberal Arts, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

Corresponding Author:
Dooyoung Jung, MD, PhD
Graduate School of Health Science and Technology
Ulsan National Institute of Science and Technology
UNIST-gil 50
Ulsan, 44919
Republic of Korea
Phone: 82 522174010
Email: dooyoung@unist.ac.kr

Abstract

Background: Procrastination negatively affects university students' academics and mental health. Traditional time management apps lack therapeutic strategies like cognitive behavioral therapy to address procrastination's psychological aspects. Therefore, we developed and integrated a semigenerative chatbot named Moa into a to-do app.

Objective: We intended to determine the benefits of the Moa-integrated to-do app over the app without Moa by verifying behavioral and cognitive changes, analyzing the influence of engagement patterns on the changes, and exploring the user experience.

Methods: The developed chatbot Moa guided users over 30 days in terms of self-observation, strategy establishment, and reflection. The architecture comprised response-generating and procrastination factor–detection algorithms. A pilot randomized controlled trial was conducted with 85 participants (n=37, 44% female; n=48, 56% male) from a university in South Korea. The control group used a to-do app without Moa, whereas the treatment group used a fully automated Moa-integrated app. The Irrational Procrastination Scale, Pure Procrastination Scale, Time Management Behavior Scale, and the Perceived Stress Scale were examined using linear mixed models with repeated measurements obtained before (T0) and after (T1) 1-month use and after 2-month use (T2) to assess the changes in irrational procrastination, pure procrastination, time management and behavior, academic self-regulation, and stress. Intervention engagement, divided into "high," "middle" and "low" clusters, was quantified using app access and use of the to-do list and grouped using k-means clustering. In addition, changes in the psychological scale scores between the control and treatment groups were analyzed within each cluster. User experience was quantified based on the usability, feasibility, and acceptability of and satisfaction with the app, whereas thematic analysis explored the users' subjective responses to app use.

Results: In total, 75 participants completed the study. The interaction of time × procrastination was significant during the required use period (P=.01). The post hoc test indicated a significant improvement from T0 to T1 in the Time Management Behavior Scale and Perceived Stress Scale scores only in the treatment group (P<.001 and P=.009). The changes in Pure Procrastination Scale score after the required use period were significant in all clusters except for the low cluster of the control group. The high cluster in the treatment group exhibited a significant change in the Irrational Procrastination Scale after Bonferroni correction (P=.046). Usability was determined to be good in the treatment group (mean score 72.8, SD 16.0), and acceptability was higher than in the control group (P=.03). Evaluation of user experience indicated that only the participants in the treatment group achieved self-reflection and experienced an alliance with the app.

XSL•FO
RenderX

Conclusions: The chatbot-integrated app demonstrated greater efficacy in influencing user behavior providing psychological support. It will serve as a valuable tool for managing procrastination and stress together.

Trial Registration: Clinical Research Information Service (CRIS) KCT0009056; https://tinyurl.com/yc84tedk

(JMIR Mhealth Uhealth 2025;13:e53133) doi:10.2196/53133

KEYWORDS

procrastination; chatbot; generative model; semigenerative model; time management; cognitive behavioral therapy; psychological assessment; intervention engagement; emotional support; user experience; mobile intervention; artificial intelligence; AI

Introduction

Background

In psychology, procrastination refers to delaying an intended course of action despite expecting that this will lead to a worse outcome [1]. Rothblum et al [2] reported that fear of failure and task aversiveness are the primary causes of procrastination, indicating that students who feel overwhelmed by tasks delay the action further. Prolonged procrastination leads to poor academic performance, which impacts the careers of university students, elevates stress, and forms a negative cycle that postpones work [3,4].

Time management apps such as to-do apps are convenient for controlling this delaying behavior for better performance. Time management involves "behaviors that aim at achieving an effective use of time while performing certain goal-directed activities" [5]. Methods for implementing time management strategies such as planning and monitoring have been adapted to mobile apps to enable users to manage their tasks efficiently and effectively [6]. For instance, certain apps assist in subdividing tasks into smaller goals or performing self-assessment by maintaining daily journals [7-9].

However, existing time management apps face challenges in fundamentally resolving the issue of procrastination. They often do not manage various aspects of procrastination, including self-esteem, motivation, and mood regulation [10]. Procrastination is not solely a matter of time management but is closely associated with psychological problems such as low life satisfaction and poor mental health [11,12]. Recognizing procrastination as a psychological issue, psychologists have adapted cognitive behavioral therapy (CBT) programs to help students identify irrational thought processes, which in turn can motivate them to change their behavior [13]. Numerous studies have validated the effectiveness of CBT in treating procrastination [14-16].

However, engaging in offline counseling is often challenging for students due to difficulties in accessing mental health care services compounded by time constraints and a lack of familiarity with the therapeutic approaches used to address procrastination [17]. Consequently, they may prefer using time management apps for self-management such as Google Keep [18] or TickTick [19] owing to their convenience, time-saving features, and effectiveness in task management. Typically, these apps do not contain sufficient therapeutic content to raise awareness of irrational behavioral patterns, provide motivation, or assist in behavior change.

In this study, we developed and integrated a chatbot into a time management app to provide sufficient therapeutic content while ensuring convenience. Chatbots are widely used to support task performance in various fields because they efficiently deliver information by interacting with users in natural languages [20]. In addition, chatbots enable emotional interaction in conversations and form alliances that prolong the interest of users in the app [21,22]. Verbal communication, which is the primary feature of chatbots, is the basis of psychological therapy [23]. Both clients and counselors can share their thoughts and feelings verbally, whereas the counselors can lead conversations with adequate questions and responses. These strengths have increased the efforts to use chatbots for psychological assistance in the mental health care field [21,22].

Most health care chatbots are rule based and retrieval based, meaning that they do not use generative models such as GPT-3.5 and LaMDA [24,25]. Rule-based chatbots offer predetermined options for users, ensuring cost-effective information delivery yet limiting conversational flexibility [20]. Retrieval-based chatbots determine responses based on an artificial intelligence model to classify intent, enabling natural language interactions for goal-oriented conversations, such as in CBT [26,27]. Despite their advantages, these chatbots are confined to preexisting responses, hindering various reactions, resulting in limited empathetic conversations, and restricting the formation of alliances from interactions. Generative models address these limitations by providing relevant responses and empowering users to control conversation contents [28,29]. However, relying solely on generative models for therapeutic purposes is challenging because they cannot consistently maintain context and may produce inaccurate outputs, particularly in sensitive discussions related to severe mental health problems [30]. Therefore, a predetermined framework is necessary in the psychological field to avoid distraction from the conversation topic and minimize detrimental effects caused by erroneous responses when using the generative model.

Objectives

In this study, we developed a semigenerative chatbot named Moa to facilitate natural conversations within a predetermined scenario. This was integrated into a to-do app to mediate procrastination using a psychological approach. The architecture of Moa included a generative model to facilitate flexible responses and was tailored to user responses. Moa was interlocked with a to-do app that primarily included time management skills, planning, and monitoring and was endowed with the functions of listing tasks to be executed and visualizing achievement through a calendar [5]. After integrating Moa into the to-do app, we tested whether the developed chatbot could

XSL•FO
RenderX

provide additional benefits over an app that only supports time management without a chatbot. To this end, we conducted a randomized controlled trial (RCT) setting the intervention for the control group as the to-do app without Moa and that for the treatment group as the Moa-integrated to-do app. In addition to assessing the impact of Moa based on differences in survey scores between the control and treatment groups, we examined the changes in scores by clustering the intervention engagement patterns. This analysis acknowledged that the effectiveness of an intervention can significantly vary depending on user engagement [31]. In addition, both quantitative and qualitative analyses of the user experience were conducted to identify the specific elements influenced by Moa. In summary, our study aimed to determine the additional benefits of using the to-do app with Moa in comparison with using the app without Moa. The primary objectives of this study can be summarized as follows: (1) to validate changes in behavior related to procrastination and psychological aspects such as emotions and stress through self-reported survey scores, (2) to analyze the influence of intervention engagement on the improvement in users' behavior and cognition, and (3) to quantitatively and qualitatively explore the user experience.

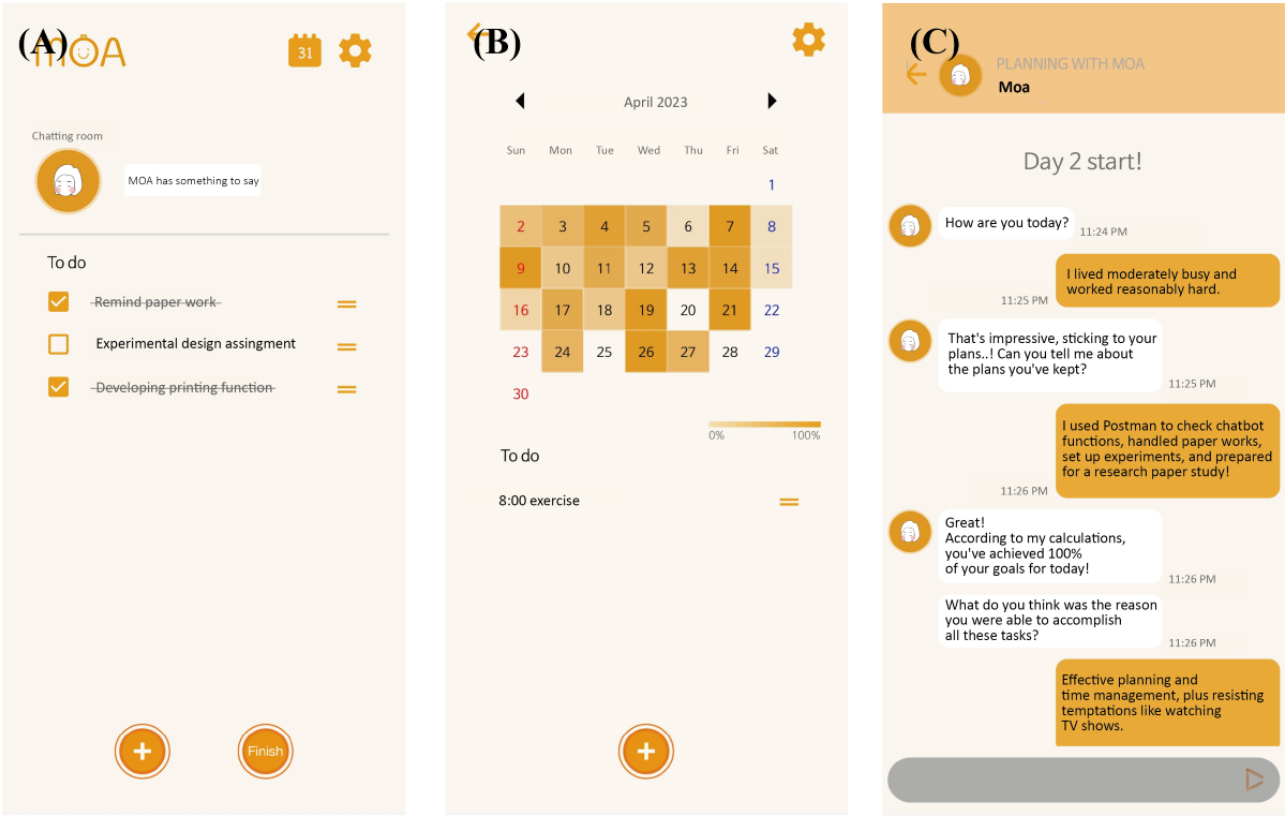
Methods

Development of the Intervention

Overview

The intervention for the control group involved performing to-do app functions, and the Moa chatbot was added only for the treatment group. A typical to-do app pushes alarms to write plans in the morning if users do not note any lists for the day. Users tap the Finish button after checking the tasks completed on the list. Subsequently, the success rate, which denotes the ratio of the number of checked lists to the total number of lists, is automatically calculated, and the saturation of the calendar differs from the success rate. For the treatment group, the chatbot Moa initiates a conversation based on the success rate. Figures 1A and 1B depict the overall user interface (UI) of the to-do app.

Figure 1. Screenshots of the time management app we used for the treatment group: (A) main screen with to-do list; (B) calendar screen visualizing success rate; (C) chatting room screen for conversation with the chatbot Moa. The content that was originally in Korean has been translated into English.



Moa is a semigenerative chatbot interlocked with the to-do app to facilitate conversations tailored according to the users' delaying behavior. Moa aims to foster users' awareness of behavioral patterns and mediate procrastination. To realize the intervention for the treatment group, the function of Moa was interlocked as depicted in Figure 1C.

Moa includes generative models in a predetermined scenario based on the CBT process to encourage active user engagement in conversations while maintaining consistency to guide behavior change; therefore, it is referred to as a semigenerative chatbot. Typically, CBT chatbots for mental health are rule-based systems driven by knowledge-based communication [30]. Generative models of mental health–related chatbots are infrequent because of the potential for erroneous responses, which could lead to major issues such as inadvertent promotion of self-harm or suicide [20]. To minimize the damage from incorrect responses, we restricted the topic of Moa to problems related to procrastination rather than severe mental illnesses. In addition, we incorporated predetermined questionnaires and established a scenario to maintain context within a conversation,

XSL•FO
RenderX

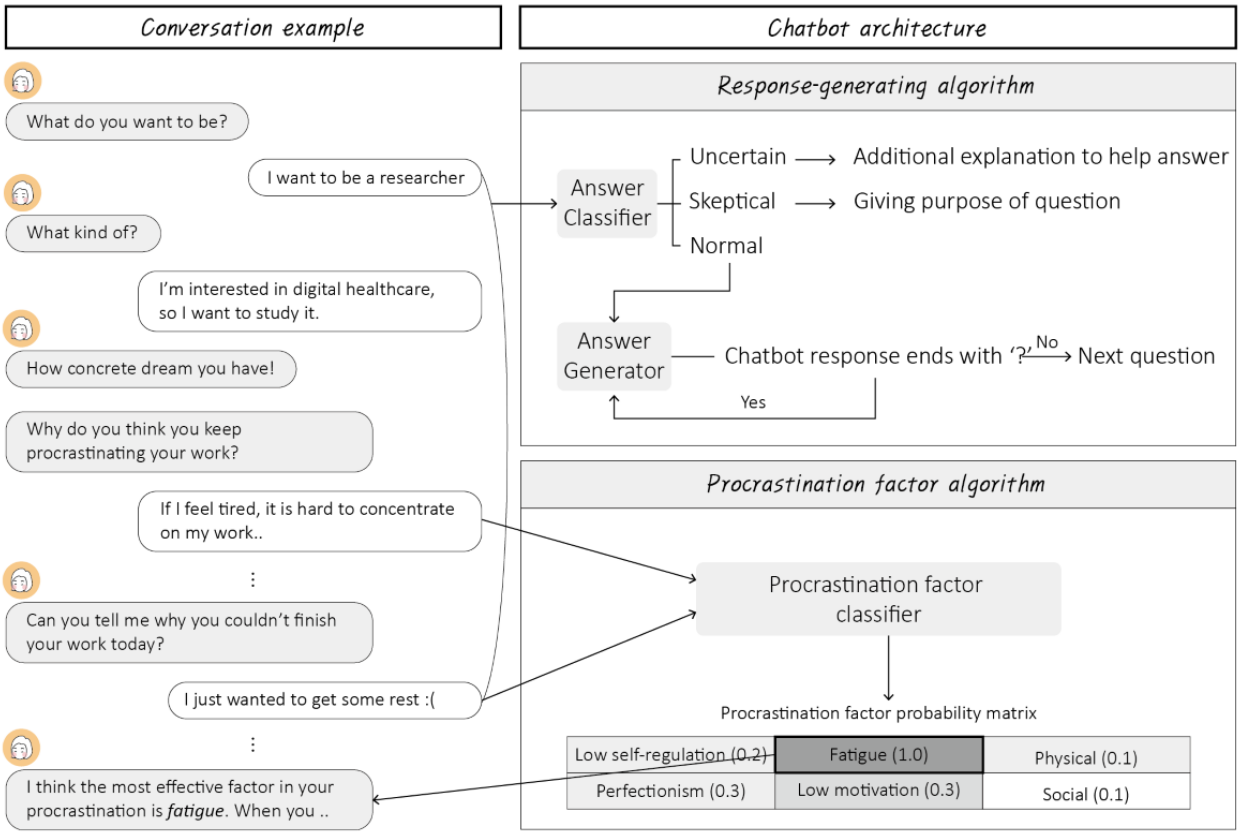
ensuring that, even if the chatbot responds awkwardly, the conversation stays on track and avoids harmful directions. These questionnaires were a part of the therapeutic process and provided knowledge pertinent to time management. Furthermore, we collected our own data, and the authors reviewed them to remove harmful utterances from the training dataset used for generative pretrained transformer development.

The overall Moa chatbot scenario comprises the process of procrastination counseling in academic settings that observes maladaptive delaying behavior, searching for its causes and modifying behavior and cognition [32]. In addition, procrastination is highly associated with a lack of self-determined motivation, and most people with high procrastination levels struggle to establish reasonable goals to initiate the tasks [33]. To encourage people to set achievable goals, we adapted the early part of the life-crafting process, which involved identifying life values and goals to motivate users to implement actions and plan accordingly [34]. This scenario was thoroughly reviewed by the coauthors, including a psychiatrist and a clinical psychologist.

The Moa chatbot sessions spanned 1 month and featured varied weekly topics. The sessions included the following steps: (1) observing oneself, (2) understanding strengths and weaknesses while working, (3) establishing personalized strategies by reflecting on one's behavioral patterns, (4) learning effective strategies, and (5) reflecting on personal change. These sessions encouraged users to develop an awareness of their behavioral patterns and pursue changes. The detailed scenario for each session is provided in Multimedia Appendix 1.

To operate Moa with tailored responses and advice, we designed a chatbot architecture comprising 2 primary algorithms: a response-generating and a procrastination factor algorithm (Figure 2). A survey dataset was collected from researchers, and an open-source dataset was used to train the models embedded in the algorithms. The survey data were obtained via questionnaires designed with open-ended questions. The questionnaires were administered between 2021 and 2022 to 447 undergraduate students at a university in South Korea. Survey data were used after elaboration to train each model. We also used an open-source dataset, namely, the wellness conversation script dataset publicly available from the AI Hub, an open-data web platform [35]. This dataset comprises sentence pairs with problem statements and the corresponding chatbot responses. Sentences labeled as issues that college students commonly face, such as emotional regulation difficulties, worries, self-blame, and interpersonal relationships, were selected; a total of 765 sentence pairs were included.

Figure 2. Moa chatbot architecture.



Response-Generating Algorithm

The response-generating algorithm describes how Moa responds to user utterances, as depicted in Figure 2. When the user responds to the question posed by Moa, the answer classifier classifies the user's response into 1 of 3 categories: "uncertain," "skeptical," or "normal." A response classified as uncertain or skeptical indicates a situation in which the user experiences difficulty in responding. In such cases, the algorithm retrieves sentences from the corpus to help the user reconsider the

XSL•FO
RenderX

question. If the user's response is classified as normal, the answer generator generates a sentence using the response. When an interrogative sentence is generated, the algorithm waits for the subsequent response from the user. Conversely, in the case of a declarative sentence, the algorithm proceeds to the subsequent step in the scenario.

The answer classifier was trained using the Korean Bidirectional Encoder Representations From Transformers [36] with 3356 survey data sentences labeled as uncertain, skeptical, or normal. The test accuracy of the model was 0.86 when the training and test datasets were used in the ratio of 7:3. The answer generator used the Korean generative pretrained transformer 2 [37] released by SK Telecom and fine-tuned it using the data suitable for the scenario. A total of 18,453 sentences from the survey data and the open-source dataset were used for fine-tuning.

Procrastination Factor Algorithm

The procrastination factor algorithm was designed to detect the leading causes of procrastination based on user utterances to provide tailored advice. Initially, we conducted a literature review [1,38-46] using the keywords "procrastination," "risk factors," and "students" to define the principal aspects of procrastination. The reasons for procrastination from each reference were extracted and categorized within common themes, and 6 factors were identified: physical and emotional stress, social relationships, environmental distractions, perfectionism, low self-regulation, and lack of motivation. The results of the literature review can be found in Multimedia Appendix 2. These factors served as labels for the training datasets.

We trained a multilayered convolutional neural network using 719 sentences from the survey data to develop a procrastination factor classifier. These sentences were divided in the ratio of 7:3 for training and test sets to validate the model. We achieved an accuracy of 51%, with a top-2 accuracy of 81%. The classifier was adopted to answer several questions related to work inefficiency, and the resulting probability matrices were combined to improve the ability to detect procrastination factors. The chatbot provided tailored advice relevant to the factor with the highest score in the final matrix. If the user disagreed with the results inferred from the model, one of the provided advices for the second-highest factor was randomly selected and recommended, and the top 1 and 2 values from the matrix were switched and saved to reflect the genuine information. To generate content for advice, we conducted brief discussions with 4 certified clinical psychologists. These counselors presented advice on the role of Moa in situations in which users could not complete their work owing to each procrastination factor. Refined responses were saved in a database for generating tailored advice.

Study Design

To validate the effect of the Moa-integrated to-do app, we conducted a pilot study using a split-plot RCT design among university students.

Ethical Considerations

Facilities Ethics Board approved the experimental procedures of the Ulsan National Institute of Science and Technology (UNISTIRB-22-72-A). Every participant provided their consent via email after they received the explanation of the overall experiment process through a web-based meeting. This study did not collect personally identifiable information from participants, and each individual's data were deidentified. Participants were compensated based on the duration of their involvement in the experiment, with those who completed the 2-month experiment and submitted the questionnaire receiving approximately US $70.

Participants

In March 2023, a total of 151 undergraduate students from a Korean university were initially selected to participate in this experiment. We recruited participants through social media and campus posters using the following statement: "Recruitment for user test of an application for effective time management." Participants were not informed of the specific characteristics of the interventions, which allowed us to blind participants to whether they were in the treatment or control group. Interested individuals voluntarily completed web-based application forms containing recruitment surveys related to demographic information and the Patient Health Questionnaire–9 (PHQ-9), Generalized Anxiety Disorder–7 (GAD-7), Pure Procrastination Scale (PPS), and Irrational Procrastination Scale (IPS) to determine their eligibility for participating in the experiment.

We focused on individuals who had procrastination without other severe mental health conditions such as depression and anxiety. Therefore, individuals scoring <29 on the PPS, the average score in a previous study that assessed PPS scores from Korean university students, were not allowed to participate in the experiment [47]. The presence of other mental health conditions was examined based on the severity indicated by the PHQ-9 and GAD-7, and individuals with scores of >14 on both these scales, indicating a cutoff for a severe level, were excluded to avoid including those with potentially severe mood disorders that required clinical treatment initially [48,49].

Participants who failed to use the to-do list app with the chatbot at least 3 times during a 1-week practice session were also excluded to ensure consistent engagement throughout the experiment.

After the practice session, participants were asked to complete a baseline survey (T0) just before receiving treatment. Participants who exhibited moderate depression or anxiety as determined using a Beck Depression Inventory (BDI) score of >23 or a Beck Anxiety Inventory (BAI) score of >15 during the baseline survey (T0) were also excluded. This additional screening was necessary because there was a gap between the recruitment survey and T0, and we wanted to ensure that participants with clinical levels of depression or anxiety were excluded to minimize the influence of external mental health issues on the experiment's outcomes [50,51].

On the basis of these eligibility criteria, 85 participants were deemed suitable for evaluating the effectiveness of Moa. The participants were randomly divided into 2 groups: a treatment

XSL•FO
RenderX

group comprising 55% (47/85) of the individuals, who used the Moa-integrated app, and a control group comprising 45% (38/85) of the individuals, who used the app without Moa. We used Python (Python Software Foundation) to create random allocation lists to assign participants to the treatment and control groups in a 4:3 ratio until the distribution of PHQ-9, GAD-7, PPS, and IPS scores became similar. The list generation process was repeated until the distributions were equal. Researchers were blinded to group assignment until the randomization was performed. The number of participants was powered to detect a weak effect size of 0.25 using G*Power (version 3.1.9.6; Heinrich-Heine-Universität Düsseldorf) for repeated-measure ANOVA. Assuming that 15% of the participants would leave the study by the end of the experiment given our participants' high educational background and the incentives provided to complete the sessions [52-54], the required sample size was 82.

Procedure

Eligible participants received a concise explanation regarding the installation of the app, use guidelines for the intervention, and the criteria for continued participation. The participants were initially asked to complete recruitment surveys that included demographic information, psychological questions such as the PHQ-9 and GAD-7, and details about their previous experiences with other time management apps and chatbots.

Participants with severe depression or anxiety, as indicated by the scores on the PHQ-9 and GAD-7, were excluded from the study. The participants were not allowed to use other to-do apps or offline journals to note their everyday tasks during the experiment. The APK file for installing the app was sent via email, and each participant signed up for the app at no cost. Participants were then randomly assigned to the treatment or control group based on the scores of their demography.

Following randomization, the participants were instructed to use the Moa intervention at least 3 or 4 times a week for a month, which was the required use period. In the treatment group, participants used the Moa-integrated to-do app, whereas the participants in the control group used the app without Moa. A push alarm reminded the participants to make lists and complete their daily tasks every morning and evening regardless of their group assignment. The only difference between the groups was the presence of a Moa message accompanying the push alarm in the evening. Participants who failed to satisfy the use requirements for a particular week were disqualified and received a reward based on the duration of their use. After the required use period, participants were allowed to use the app freely without communicating with Moa for 1 month; this was referred to as the free use period. Minor bugs in the app were fixed during the required use sessions, but no changes to the content were made.

We conducted 3 web-based assessments during the experiment: before the start of the required use period, which was immediately after the practice session (T0); after the required use period (T1); and at the end of the free use period (T2). A web-based survey was conducted to assess the overall user experience, including subjective feelings and usability, feasibility, and acceptability of and satisfaction while using the app during T1. Data on app use were automatically stored in a secure database (MongoDB; MongoDB Inc). All data obtained through the experiment were protected through anonymization.

Measurements

Verification of Behavior and Cognitive Changes

The psychological scales were administered during T0, T1, and T2 to analyze any changes in the behavior of the participants during and after app use. The procrastination level of individuals was examined using the PPS (Cronbach α=0.93) and IPS (Cronbach α=0.85) [47]. The PPS typically measures general procrastination with a low likelihood of clinical problems. In contrast, the IPS measures irrational behaviors associated with mental health problems such as stress and worries. The PPS and IPS comprise 12-item and 9-item questionnaires, respectively, measured on a 5-point Likert scale. The Time Management Behavior Scale (TMBS), developed by considering the fundamental theories and strategies for time management, was used to evaluate the time management abilities of the participants (Cronbach α=0.88) [55]. Here, the factors of time management were factorized into "goal-setting and planning" (Cronbach α=0.82), "procrastination" (Cronbach α=0.86), and "practice strategies" (Cronbach α=0.86), with 9-item, 8-item, and 6-item questionnaires for each factor on a 5-point Likert scale, respectively. We used the Academic Self-Regulation Scale (ASRS) developed by the Brain and Motivation Research Institute in South Korea to identify the difference in regulative behavior with respect to academic work (Cronbach α=0.89) [56]. This scale comprises 8 items rated on a 7-point scale. Furthermore, stress, depression, and anxiety were measured to assess changes in mental status. The Perceived Stress Scale (PSS) comprises 10 items measured on a 5-point Likert scale based on 2 factors: negative perception (Cronbach α=0.77) and positive perception (Cronbach α=0.74) [57]. The PSS score was calculated by reverse-coding only the items related to positive perception among the total items, and subsequently summing the scores of all items. We used the BDI and BAI, which comprise 21 items each measured on a 4-point scale, for depression and anxiety, respectively [50,51]. The BDI and BAI were used to screen people for anxiety and depression to eliminate nuisance factors only except for measuring the intervention effect.

Exploration of Behavior and Cognitive Changes via Engagement

Engagement measurement tools are essential for evaluating user interaction with an app and their subsequent behavior and psychological changes [58]. Given the variety of methods for measuring engagement in digital interventions, it is essential to clearly define what specific aspect of engagement will be measured in our research [59,60]. In this context, we quantified the engagement of participants primarily based on app and chatbot engagement because the chatbot was used only by the treatment group, whereas both the control and treatment groups used the to-do app.

To obtain the detailed app engagement score, data on app use were logged each time the participants accessed the app and made entries. App engagement was measured daily using three subscores: (1) frequency of app access, (2) number of to-do

XSL•FO
RenderX

lists, and (3) elapsed time of the to-do list. Each subscore was adjusted to remain in the range of 0 to 1 to ensure comparability. The daily app engagement score was obtained by summing the 3 subscores using the specifics and formulas detailed in Table S1 in Multimedia Appendix 3. These daily scores were aggregated into weekly averages, enabling the discernment and comparison of patterns across groups every week. Subsequently, we explored weekly app engagement patterns over time using k-means clustering analysis based on the correlation distance. The optimal number of clusters was determined using the elbow method. Each cluster was then characterized according to its unique engagement pattern over time. Furthermore, we evaluated the changes in behavioral and psychological scores within each of these clusters for each group.

To verify the effects of the chatbot intervention, we calculated the chatbot engagement score exclusively for the treatment group. This score incorporated several elements, including the length and duration of user responses and emotional scores derived from conversations with Moa. Emotional scores ranging from 0 (indicating negative emotions) to 1 (indicating positive emotions) were extracted using the Korean Bidirectional Encoder Representations from Transformers sentiment classifier; simple responses such as “yes” or “no” were excluded from emotional scoring. Considering the variability in the quality and content of user responses required in scenario-based chatbot conversations, temporal changes in chatbot engagement were not considered in our analyses. By comparing the chatbot engagement scores across different app engagement clusters, we aimed to elucidate the potential influence of chatbot interventions on behavioral modification.

Exploration of User Experience

User experience was quantified via the user test surveys. The usability of the developed apps was measured using the 10-item System Usability Scale [61] on a 5-point Likert scale. The total score was calculated by multiplying 2.5 by the sum of the item scores. The standards for the total scores were “OK if >55,” “good if >75,” and “excellent if >87.5” [62]. The acceptability of the app, management of procrastination, and feasibility were assessed considering the Usage Rating Profile–Intervention (URP-I) [63]. The URP-I was designed to measure the factors that individuals adopt and subsequently use in the intervention over time. We used the 8-item subscale associated with acceptability and the 6-item subscale for feasibility in the URP-I, adapting the questionnaires to the purpose of our app using a 6-point Likert scale. The general satisfaction with the intervention in the treatment and control groups was assessed using an 8-item client satisfaction questionnaire measured on a 4-point Likert scale [64].

User experience in the treatment and control groups was assessed based on the overall feeling of the users when using the app, collected via a subjective questionnaire (“Please write the overall feeling freely after using the app”). Participants were given additional explanations of the questions to guide their responses regarding aspects such as the advantages and disadvantages, similarities to prior expectations, and any behavioral changes they experienced.

Analytic Strategy

To analyze the self-reported surveys, linear mixed models were fitted for repeated measures of each feature investigated in the questionnaires (ie, procrastination, time management, academic self-regulation, and stress). We critically considered 3 effects: the effect of the chatbot intervention by comparing the treatment and control groups (*group*), the effect of time (T0, T1, and T2; *time*), and the interaction of *group* and *time*. Both *group* and *time* were nominalized and set as fixed effects. The interaction effect of *group* and *time* was set as an effect of the model to explain how chatbots impact behavior or cognitive changes. We set the score of each scale at T0 as a covariate to control for the initial difference between the 2 groups. Participants considered as random effects in the model were nested into the *group* effect. A linear mixed model was fitted using the RStudio (version 4.3; Posit PBC) *nlme* package [65]. We conducted post hoc pairwise comparisons considering different numbers of samples from each group using the Scheffé method with the *emmeans* library in RStudio to conservatively determine the primary effect [66,67].

In the clusters derived from the temporal patterns of app engagement, the objective was to compare the changes in questionnaire scores obtained from T0 to T1 across each group. The sample sizes were reduced because of cluster division, necessitating nonparametric testing. Consequently, we used a 2-sided Wilcoxon signed-rank test in our analyses. In addition, separate tests were conducted for the 2 groups (treatment and control) and 3 clusters (high, mid, and low) within each questionnaire, resulting in a total of 6 comparisons. To control for the risk of type-I error due to multiple comparisons, we applied a Bonferroni correction. To investigate the differences in chatbot engagement scores among the clusters, we conducted the Kruskal-Wallis ANOVA test followed by the Scheffé post hoc test for multiple cluster comparisons.

To account for the potential influence of participants who did not consistently engage with the intervention, we performed an intention-to-treat analysis initially. In this case, we filled out the missing values for withdrawals by using the scores that were previously measured during the prior assessment. In addition, we conducted a completer-only analysis excluding participants who withdrew during the required use period (T1) to assess the effect of the intervention while controlling more precisely for nuisance variables such as use frequency. A post hoc test was subsequently conducted on the completer data.

To quantitatively analyze the user experience, we calculated the total score of each test survey to examine user experience through simple statistics. The difference between the treatment and control groups was validated via a proportion *z* test using positive survey results. For the review data, we conducted a thematic analysis to elucidate the findings derived from subjective responses using QDA Miner Lite (Provalis Research). The smallest level, features of the intervention, was identified from the users’ reviews, and the responses with similar features were combined to form codes. They were categorized into 2 groups: the intervention effect that individuals can experience cognitively and the functions that individuals mentioned positively or negatively. The functions of the intervention were

XSL•FO
RenderX

further categorized into graphical UI–based to-do app and chatbot-based to-do app. The coding of the data was discussed with 3 coauthors to ensure accuracy, and the codes were refined by carefully reviewing the content.

Finally, we calculated the percentage of sentences that were coded with codes from the intervention effect and positive function categories together to explore relationships between the functions and the effect that users experience. If a code from the intervention effect category co-occurred with one from the positive function category in ≥40% of the sentences, they were assumed to be related.

All statistical analyses were conducted using RStudio (version 4.3) [68].

Results

Participant Characteristics

We recruited participants starting in February 2023 and completed our experiment in May 2023. Initially, 85 participants were eligible for inclusion in the study. Participants were randomly divided into 2 groups based on their balanced characteristics. We validated the balanced distribution of the demographic, psychological, or behavioral characteristics using a 2-tailed *t* test. Table 1 summarizes the detailed information about the participants. Figure 3 shows the withdrawal of participants based on the inclusion and exclusion criteria.

Table 1. Participants' demographics.

Characteristics	Treatment (n=47)	Control (n=38)	*P* value[a]
Demographic characteristics			
Age (y), mean (SD)	20.6 (2.0)	20.2 (1.9)	.29
Sex (female), n (%)	19 (40)	18 (47)	—[b]
Baseline psychological and behavioral characteristics, mean (SD)			
PHQ-9[c] score	3.6 (3.2)	4.4 (3.5)	.26
GAD-7[d] score	2.6 (2.8)	2.8 (2.7)	.75
PPS[e] score	42.9 (6.7)	40.6 (6.8)	.13
IPS[f] score	32.2 (5.4)	31.4 (4.9)	.48

[a]*P* values are the result of 2-sided *t* tests.

[b]*t* test results are not available for sex as it is a binary variable.

[c]PHQ-9: Patient Health Questionnaire–9.

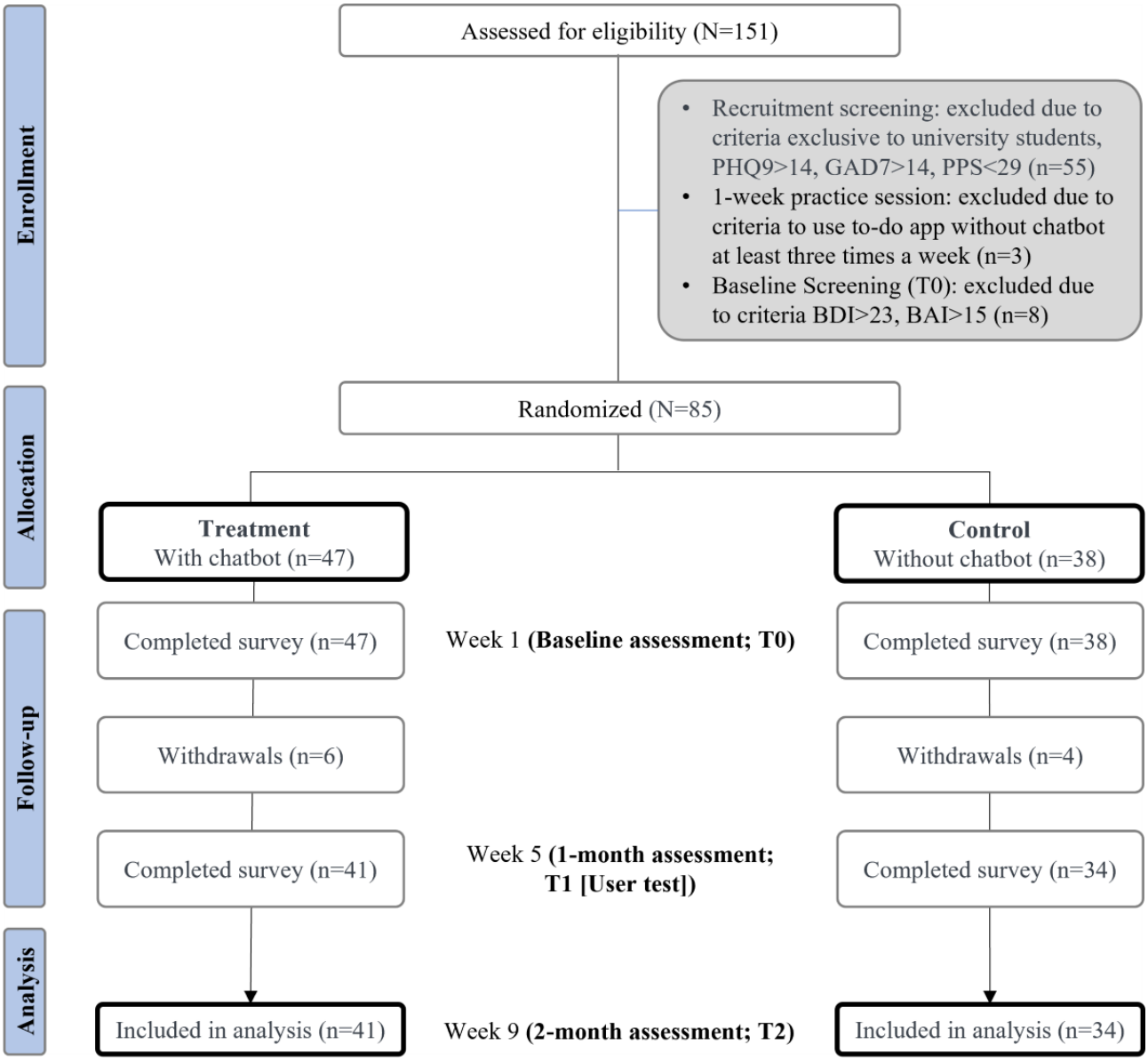
[d]GAD-7: Generalized Anxiety Disorder–7.

[e]PPS: Pure Procrastination Scale.

[f]IPS: Irrational Procrastination Scale.

XSL•FO
RenderX

Figure 3. Study flowchart representing randomization and criteria to exclude participants throughout the study. BAI: Beck Anxiety Inventory; BDI: Beck Depression Inventory; GAD-7: Generalized Anxiety Disorder–7; PHQ-9: Patient Health Questionnaire–9; PPS: Pure Procrastination Scale.



Verification of Behavior and Cognitive Changes

Table 2 summarizes the results of the *F* test conducted via the linear mixed model using the data from T0 to T1 and the differences between these time points using the intention-to-treat method. The results indicate that the absolute value of the difference in the treatment group was generally higher than that in the control group. The interaction of *time* × *group* was significant in the procrastination subscale of the TMBS (*P*=.03). The *time* effect on all questionnaires except for "Goal-setting/planning," which is a subscale of the TMBS and ASRS, was significant (*P*<.001), whereas the *group* effect was not.

XSL•FO
RenderX

Table 2. Difference between T0 and T1 with F statistic with a linear mixed model setting scores at T0 as a covariate (intention-to-treat analysis).

	T1–T0, mean (SD)		*F* test (*df*)			*P* value		
	Treatment (n=47)	Control (n=38)	Time	Group	Interaction	Time	Group	Interaction
Pure Procrastination Scale	−10.29 (7.74)	−7.68 (6.93)	103.6 (1, 83)	4.05 (1, 82)	1.08 (1, 83)	<.001	.048	.30
Irrational Procrastination Scale	−4.54 (5.59)	−2.53 (6.13)	40.44 (1, 83)	2.04 (1, 82)	2.7 (1, 83)	<.001	.16	.10
Time Management Behavior Scale	7.56 (10.6)	4.09 (9.69)	30.46 (1, 83)	0.09 (1, 82)	1.98 (1, 83)	<.001	.77	.16
Goal setting/planning	0.49 (4.71)	−0.68 (4.62)	1.77 (1, 83)	1.01 (1, 82)	1.01 (1, 83)	.19	.32	.32
Procrastination	4.61 (4.47)	2.24 (4.06)	4.74 (1, 83)	4.77 (1, 82)	4.77 (1, 83)	.03	.03	.03
Practice strategies	2.46 (4.59)	2.53 (4.39)	2.33 (1, 83)	0.01 (1, 82)	0.01 (1, 83)	.13	.92	.92
Academic Self-Regulation Scale	−0.12 (6.12)	−0.21 (6.24)	0.13 (1, 83)	7.69 (1, 82)	0.04 (1, 83)	.72	.007	.84
Perceived Stress Scale	−2.46 (4.64)	−1.38 (4.63)	19.37 (1, 83)	0.03 (1, 82)	1.05 (1, 83)	<.001	.87	.30

The results from a linear mixed model from T0 to T1 to T2, which included a month after the end of the required app use period, revealed that the *time* effect was significant at every scale (P<.001).

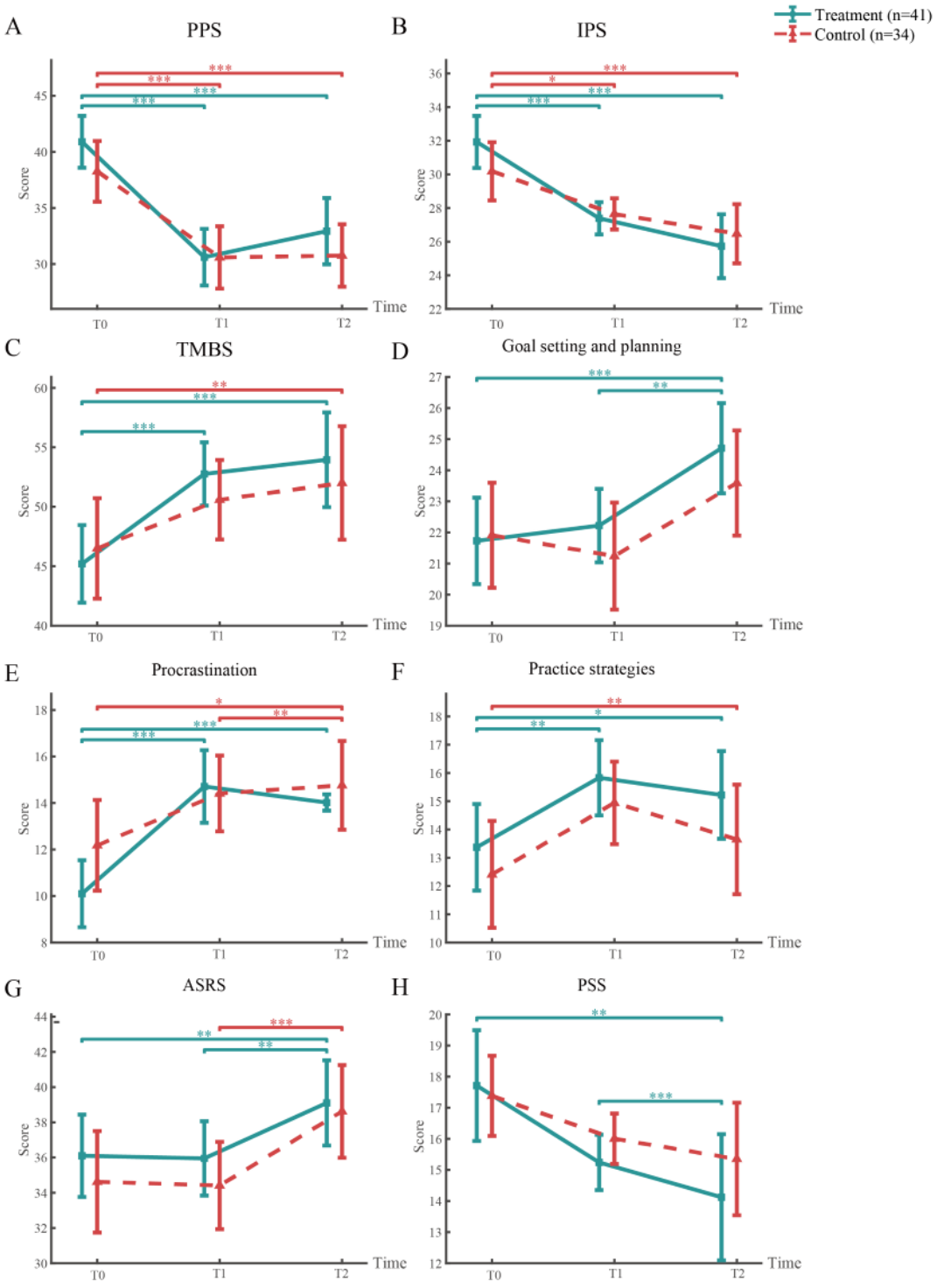
We also analyzed the completers excluding participants who withdrew from the study due to insufficient app use. The results are presented in Table S3 in Multimedia Appendix 3. With the exclusion of 12% (10/85) of the participants, the overall P value decreased slightly. Notably, the interaction of the procrastination subscale of the TMBS showed significant results (P=.01). In addition, irrational procrastination measured using the IPS was close to the significance level (P=.07).

An overview of each questionnaire examined at the time points before and after the required app use period (T0 and T1) and after a month (T2) is provided in Multimedia Appendix 3 with 95% CIs. Most participants exhibited improved scores for all variables except the ASRS.

We conducted a post hoc test using the Scheffé method to identify behavioral and psychological change within each group among completers. Figure 4 illustrates the scale variations, indicating the significance level based on the results of the Scheffé test. As indicated in the figure, the PPS scores decreased in both the treatment and control groups regardless of the use of Moa. Conversely, irrational procrastination measured using the IPS exhibited a strong significant tendency to decline in the treatment group between T0 and T1 (mean −4.54, SD 5.59, 95% CI −6.30 to −2.77; P<.001) and between T0 and T2 (mean −6.20, SD 5.78, 95% CI −8.02 to −4.37; P<.001), whereas the control group showed a smaller significant variation between T0 and T1 (mean −2.53, SD 6.13, 95% CI −4.67 to −0.39; P=.02).

XSL•FO
RenderX

Figure 4. Changes from T0 to T1 to T2 in the questionnaire scores with significance from the Scheffé test among completers (mean and 95% CI). *P<.05, **P<.01, ***P<.001. ASRS: Academic Self-Regulation Scale; IPS: Irrational Procrastination Scale; PPS: Pure Procrastination Scale; PSS: Perceived Stress Scale; TMBS: Time Management Behavior Scale.



The TMBS score, which measures time management, increased significantly from T0 to T1 (mean 7.56, SD 10.6, 95% CI 4.22-10.91; P<.001) and maintained the intervention effect until T2 in the treatment group; however, it was not statistically significant in the control group (mean 4.09, SD 9.69, 95% CI 0.71-7.47; P=.05). The control group did not exhibit significant changes between T0 and T2 in the "Goal-setting/planning" subscale (mean 1.68, SD 4.54, 95% CI 0.09-3.26; P=.07). In contrast, the treatment group exhibited a significant increase in scores from T0 to T2 (mean 2.98, SD 4.93, 95% CI 1.42-4.53; P<.001).

The PSS scores decreased over time regardless of the group, but only the treatment group showed a significant change. Remarkably, the treatment group exhibited a significant

XSL•FO
RenderX

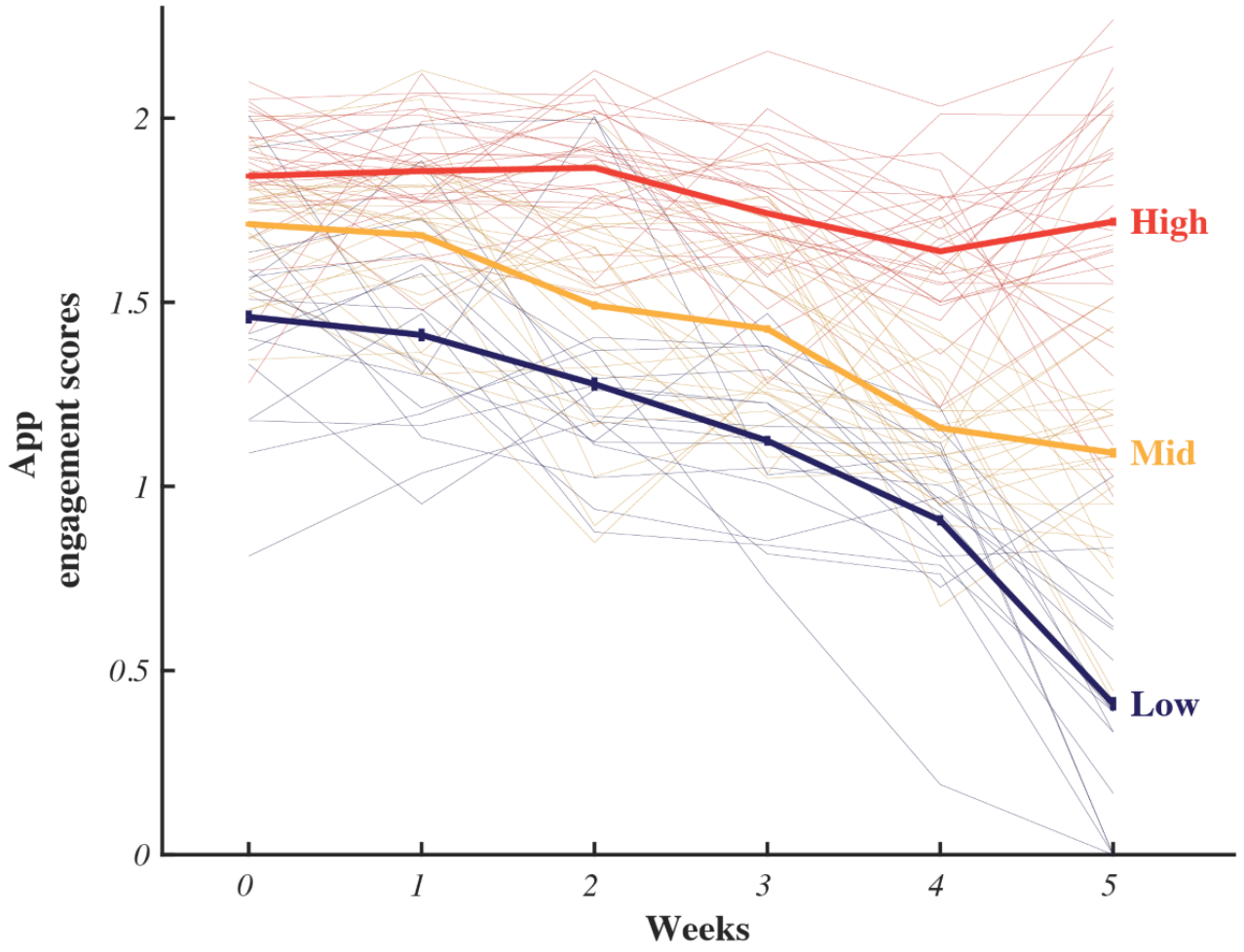
difference in PSS scores between T0 and T2 (mean −3.59, SD 6, 95% CI −5.48 to −1.69; *P*<.001).

The effect of *time* between T0 and T1 was insignificant on the ASRS; however, the difference between T0 and T2 (treatment: *P*=.003; control: *P*<.001) and between T1 and T2 (treatment: *P*=.002; control: *P*<.001) was significant.

Exploration of Behavior and Cognitive Changes via Engagement

We compared the patterns of temporal change in app engagement scores between the treatment and control groups (Table S4 in Multimedia Appendix 3). Although the results of the *F* test showed that both groups exhibited a decrease in app engagement scores over time ($F_{5,\ 355}$=43.44; *P*<.001), no interaction effect existed between *time* and *group* ($F_{5,\ 365}$=0.36; *P*=.94). Therefore, we extracted clusters of app engagement changes over time based on correlation distance–based k-means clustering. The optimal number of clusters, determined using the elbow method, was 3 (Figure S3 in Multimedia Appendix 3). We categorized users into 3 clusters: “high,” “middle,” and “low,” representing their level of sustained engagement with the app during the intervention (Figure 5). In the high cluster, users maintained their engagement scores relatively consistent throughout the practice (week 0) and required use (weeks 1-5) periods in comparison with the middle and low clusters. Although the app engagement scores of both the middle and low clusters decreased over time, the scores of the middle cluster remained higher than those of the low cluster, and the scores of the high cluster remained significantly higher than those of the middle cluster at all time points (Tables S5 and S6 in Multimedia Appendix 3).

Figure 5. Temporal changes in app engagement scores within each cluster (mean and 95% CI).



We observed changes in the behavioral assessment results corresponding to variations in app engagement. Table 3 lists these changes by presenting the differences in questionnaire scores between T0 and T1 per cluster within each group. Notably, the findings mirror those in Table 2, indicating an overall score improvement across the groups. The Bonferroni-corrected *P* values for each of the 6 comparisons (treatment: high, mid, and low engagement; control: high, mid, and low engagement) are also reported in Table 3, reflecting the significance of these changes after adjusting for multiple comparisons.

XSL•FO
RenderX

Table 3. Descriptive statistics and Wilcoxon signed-rank test for behavioral and psychological assessment—comparing T1 and T0 within groups in each cluster.

	Changes during the required use period (T1–T0), mean (SD)					
	Treatment			Control		
	High engagement (n=13)	Mid engagement (n=15)	Low engagement (n=13)	High engagement (n=11)	Mid engagement (n=12)	Low engagement (n=11)
PPS[a]	−10.7[b] (6.5)	−9.4[c] (7.7)	−10.8[c] (9.9)	−8[c] (5.7)	−7.5[c] (7.4)	−7.4 (9.0)
IPS[d]	−4[c] (4.8)	−5.1 (5.1)	−4.7 (7.1)	−3 (7.2)	−0.8 (5.2)	−5.1 (5.4)
TMBS[e]	5.9 (8.3)	7.8 (11.2)	9.7 (13.4)	8.9 (10.8)	1.9 (7.8)	−0.4 (8.1)
Goal setting/planning	−0.1 (4.7)	0.2 (3.9)	1.6 (5.7)	1.8 (4.5)	−2.1 (4.6)	−2.6 (2.7)
Procrastination	4.9[b] (3.7)	4.8[c] (4.8)	3.8 (5.5)	3.5 (3.7)	2.1 (3.7)	0.3 (5.1)
Practice strategies	1.1 (3.9)	2.8 (5.0)	4.3 (4.8)	3.6 (4.8)	1.9 (3.3)	0.3 (5.6)
ASRS[f]	−1.3 (5.7)	0.4 (7.5)	1.1 (5.1)	2.1 (6.3)	−1.6 (6.1)	−1.6 (6.0)
PSS[g]	−2.7 (4.1)	−3.2 (4.7)	−1.3 (5.5)	−1.7 (4.1)	−0.9 (5.0)	−1.7 (5.3)

[a]PPS: Pure Procrastination Scale.

[b]The statistics are significant at $P<.001$ after Bonferroni correction.

[c]The statistics for the 2-sided Wilcoxon signed-rank test comparing the scores at T1 and T0 are significant at $P<.01$ after Bonferroni correction.

[d]IPS: Irrational Procrastination Scale.

[e]TMBS: Time Management Behavior Scale.

[f]ASRS: Academic Self-Regulation Scale.

[g]PSS: Perceived Stress Scale.

The PPS scores revealed a significant difference in all clusters within the treatment group and all except for the low cluster within the control group. Even after applying the Bonferroni correction, this difference remained significant in every cluster of the treatment group (high: $P=.004$; middle: $P=.01$; low: $P=.002$) and in the high ($P=.01$) and middle ($P=.03$) clusters of the control group. The IPS scores indicated a significant difference in the high and middle clusters of the treatment group only, which remained significant in the high cluster ($P=.046$) after correction.

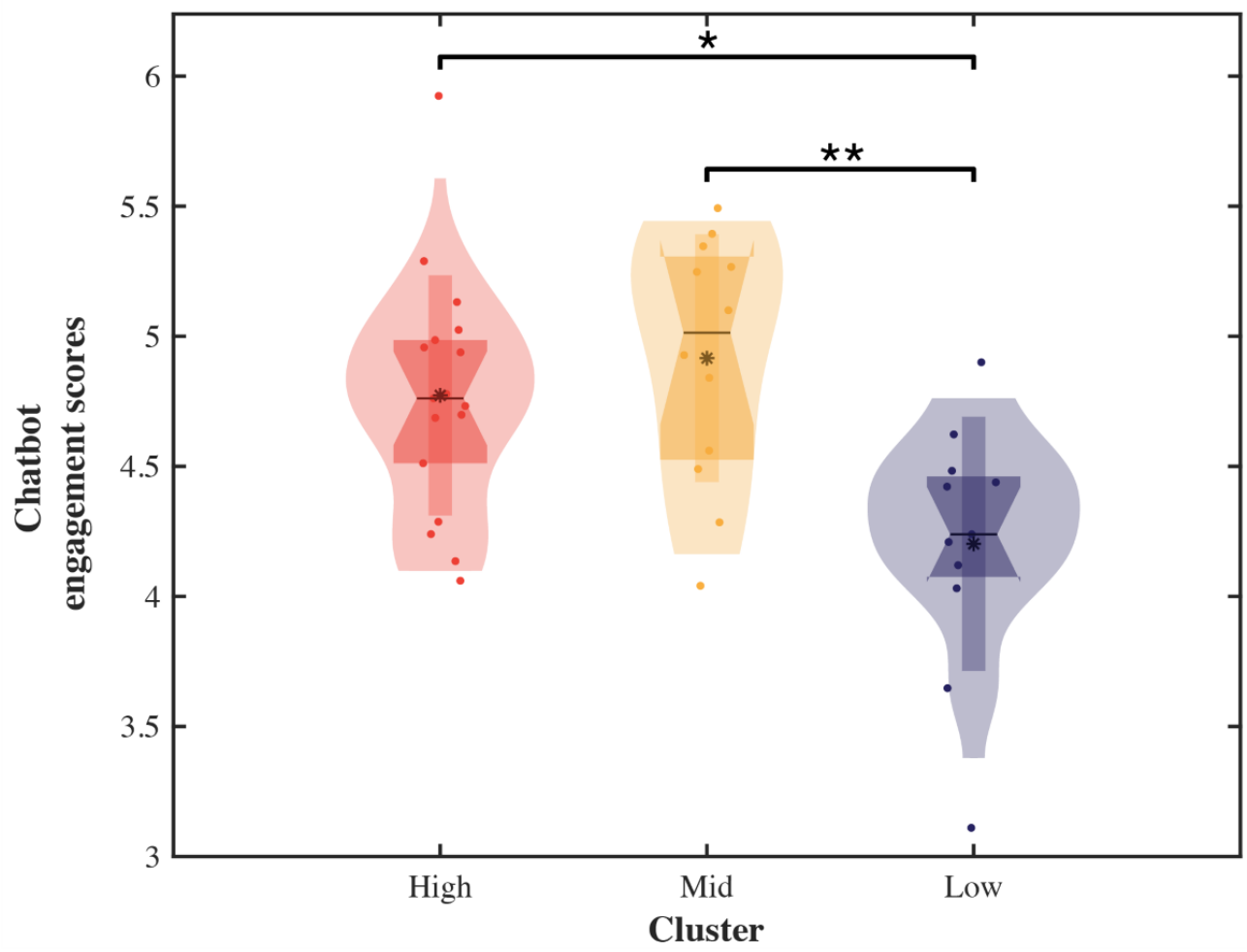
We examined the Procrastination subscale scores on the TMBS and found significant differences in all clusters within the treatment group; however, only the high cluster was significant in the control group. After Bonferroni correction, the high ($P=.004$) and middle ($P=.02$) clusters of the treatment group remained significant.

The analysis of app engagement scores revealed that, unlike the control group, the treatment group exhibited a continued intervention effect even in the middle cluster, with decreased app engagement over time.

The analysis of app engagement indicated that even users with relatively low app engagement experienced behavior changes under the treatment conditions, unlike what was observed in the control group. Therefore, we aimed to elucidate these effects through an engagement analysis of the Moa chatbot provided in the treatment group based on app engagement score cluster. The chatbot engagement scores exhibited significant differences between the clusters ($H2=10.79$; $P=.005$; Kruskal-Wallis ANOVA test). Following a Dunn post hoc analysis, the high and low ($P=.02$) and middle and low ($P=.006$) clusters demonstrated significant differences in chatbot engagement scores between them. Conversely, the disparities between the high and middle clusters were not statistically significant (Figure 6). The significantly higher chatbot engagement scores in the middle cluster compared with those of the low cluster support the possibility that significant behavior changes occurred in the users in the middle cluster in the treatment group, unlike in those from the control group.

XSL•FO
RenderX

Figure 6. Comparison of chatbot engagement scores between clusters.



Exploration of User Experience

We conducted surveys to test the usability, feasibility, and acceptability of and satisfaction with the intervention in each group (details are presented in Multimedia Appendix 3). All participants who completed the psychological questionnaires also took part in the user experience survey (n=75). Our analysis determined that >70% of the participants (58/75, 77%) scored above the midpoint of the Likert scale in every dimension of the user test. The average usability score assessed using the System Usability Scale was 72.8 (SD 16.0) in the treatment group and 75.15 (SD 10.1) in the control group, indicating that the intervention was “good” to use regardless of the presence of Moa [61]. No significant differences were observed between the control and treatment groups in terms of usability, feasibility, and satisfaction. However, participants in the treatment group exhibited significantly better acceptability than those in the control group (P=.03).

Table 4 summarizes the results of the thematic analysis of the qualitative responses regarding the app use experience of the participants in each group. The contents were categorized as intervention effects on behavioral and cognitive improvements and positive and negative features of the intervention.

XSL•FO
RenderX

Table 4. Results of the thematic analysis.

Theme and codes	Cases, n (%)	
	Treatment (n=41)	Control (n=34)
Intervention effect on the user		
Superficial effect		
Take action for today	11 (27)	7 (21)
Preparation to do	8 (20)	9 (26)
Review what the user did	5 (12)	2 (6)
Inner effect		
Diligence and sincerity	16 (39)	6 (18)
Motivation	12 (29)	6 (18)
Self-reflection	5 (12)	1 (3)
Alliance	9 (22)	0 (0)
Positive features of...		
To-do app		
Function for to-do list up	8 (20)	18 (53)
Simple	5 (12)	2 (6)
Alarm	2 (5)	6 (18)
Calendar	3 (7)	3 (9)
Moa augmentation		
Advice or feedback	11 (27)	—[a]
Consolation or cheering utterance	7 (17)	—
Conversation for reflection	5 (12)	—
Caring	4 (10)	—
Friendly tone	2 (5)	—
Negative features of...		
To-do app		
Fixed time to finish work	4 (10)	7 (21)
Lack of function	5 (12)	7 (21)
System error	3 (7)	5 (15)
Inconvenient UI[b]	4 (10)	2 (6)
Moa augmentation		
Illogical answers	16 (39)	—
General feedback or not interesting	4 (10)	—

[a]The control group did not interact with the chatbot, resulting in missing values for this variable.

[b]UI: user interface.

The effects of the intervention on the users were divided into superficial and inner effects. The superficial effect theme included 3 codes: *Take an action for today*, *Preparation to do*, and *Review what the user did*. The answers that facilitated task performance, such as “I was able to perform well without forgetting things” and “Writing things down in advance, I have reduced the number of missed assignments,” were coded as *Take an action for today*. We observed that participants could plan and envision their tasks (*Preparation to do*) based on responses such as “I could check what to do tomorrow” and “I liked being able to think and organize my tasks every morning.” Several participants reviewed their tasks as follows: “It was great having time to review the plans I had set and what I’ve done that day,” which was coded as *Review what the user did.* The percentage of the cases in the treatment and control groups was not remarkably different except for *Review what the user did*, which was mentioned by 12% (5/41) of the participants in the treatment group and only 6% (2/34) of the participants in the control group. These effects were considered superficial changes because they were directly associated with the to-do

XSL•FO
RenderX

functions, including establishing lists of tasks to execute, checking the tasks completed by users, and visually displaying these to the users.

The inner effect did not directly vary according to the functions of the app and was developed internally using traits and mindsets. We identified *Diligence and sincerity*, *Motivation*, *Self-reflection*, and *Alliance* as inner effects. Participants who experienced an improvement in their procrastination habits or enhanced sincerity answering "I have significantly reduced the habit of procrastinating on tasks" and "I lead a more fulfilling life than before" were coded as *Diligence and sincerity*. Participants coded as *Motivation* were motivated to plan or execute their tasks, and stated the following: "(Moa) gave me more determination to stick to my plans" and "Looking at the lists, I was able to receive the spirit of challenge that encourages to put things into practice." Several participants reflected on their habits and goals, indicated by the following statements: "I could notice my procrastination pattern" and "I started to contemplate how I should live my day," which were classified as *Self-reflection*. In comparison with the control group, more participants in the treatment group experienced internal changes. Only the participants in the treatment group perceived the app as a mode of communication rather than as a mere tool, enabling them to gain internal support that represented an *Alliance* between the users and Moa; this was indicated by statements such as the following: "Sometimes, Moa became a source of comfort" and "Feeling like having a friend who constantly takes care of my schedule."

There were functions that were evaluated positively by the users. Participants evaluated the *Function for to-do list* as helpful, saying the following:

> *By obligingly jotting down today's tasks, I could lead a more fulfilling life.*

The simple UI of the app was positively evaluated (eg, "UI is neat, and app usage is intuitive" [*Simple*]). Although the treatment group's app contained more features and was more complicated to use, there were more responses coded as *Simple* than in the control group. The alarm function coded as *Alarm* was considered positive, indicated by statements such as the following:

> *Getting daily reminders helped me develop a good habit.*

Participants who evaluated *Calendar* as suitable visual feedback mentioned the following:

> *It was also great to see my achievements at a glance through the calendar.*

Among the positive aspects after Moa augmentation, over one-fourth of the participants (11/41, 27%) mentioned that *Advice/feedback* from Moa was a positive feature, saying the following:

> *During conversations with Moa, I learned various (time management) tips.*

Similarly, 12% (5/41) of the participants, who responded, "I enjoyed having the opportunity to think about things I don't usually consider and reflect on overlooked aspects while chatting with Moa," received support from *Conversation for reflection*. In addition, *Consolation/cheering utterance* was considered a positive aspect, indicated by statements such as the following:

> *I appreciated having someone in Moa who could cheer me on and support my plans during our conversations.*

Caring was also one of the positive aspects of the chatbot, indicated by statements such as the following:

> *...I received confirmation through the chatbot, the success rate of my plans increased.*

Responses such as the following—"The cute and friendly tone made me feel comfortable, allowing me to have conversations with Moa with ease"—were coded as *Friendly tone*.

The app had certain negative aspects as well. Participants were uncomfortable with the *fixed time to finish the work*, which was set to 11 PM in the experiment. They commented the following:

> *Sometimes, I have experienced forgetting to check and press the "Finish" button when a planned task extends beyond midnight.*

Several participants criticized that the app had a *lack of functions* during use:

> *I often find it inconvenient that I have tasks that frequently extend well beyond the 11 p.m. deadline.*

Participants who complained about the errors were coded as *System error*. In contrast to the opinion that the UI was simple and intuitive, several participants evaluated it by saying the following—"it seems that the convenience of the to-do list up feature has somewhat declined, possibly due to a focus on the Moa"—indicating an *Inconvenient UI* of the app.

Furthermore, 39% (16/41) of the participants conversing with Moa experienced *illogical answers*, indicated by statements such as the following: "Sometimes, artificial intelligence can speak awkwardly" or "I felt a bit frustrated while conversing with Moa." Additional opinions mentioned the following: "It seems Moa's instructions are too generic, so it doesn't provide much help" or "The responses were more limited than expected, and the chatbot's lack of understanding made the conversation less enjoyable."

The codes from *intervention effect* that were present along with *positive features* were analyzed to identify the relationship between functions of the app and associated effects.

The codes for superficial effects did not have a significant relationship with other positive features in the treatment group. In the case of the control group, 86% (6/7) and 67% (6/9) of the codes associated with *Take an action for today* and *Preparation to do* appeared with *Function for to-do list up*, respectively.

Over 64% of the participants who experienced the inner effects of *Motivation* and *Diligence and sincerity* in the control group (7/11) mentioned *Function for to-do list up* as a positive feature of the app.

In contrast, the treatment group participants tended to touch the features of the Moa chatbot more than the elements of the to-do

XSL•FO
RenderX

app with inner effects. For instance, 44% (7/16) of the people who experienced more *Diligence and sincerity* mentioned the *Advice/feedback* of the chatbot as a positive feature. One participant stated the following:

> *I found it fascinating that the app taught me how to plan and stick to my schedule, and my procrastination habit has significantly reduced in practice.*

Furthermore, 80% (4/5) and 60% (3/5) of the participants underwent *Self-reflection* based on *Conversation for reflection* and *Advice/feedback*, respectively. One participant stated the following:

> *The best part was being able to receive feedback and reflect while making plans through conversations with Moa.*

The *Consolation/cheering utterance* of Moa appeared with *Motivation* and *Alliance* with an occurrence of 42% (5/12) and 67% (6/9), respectively. Remarkably, 44% (4/9) of the participants who mentioned *Alliance* as the effect of the app considered *Management* as a positive feature of the chatbot, indicated by statements such as the following:

> *Unlike other apps, Moa monitors my behavior and supports me, which motivated me to participate more actively.*

Discussion

Principal Findings

In this study, we developed a CBT-based semigenerative chatbot named Moa and integrated it with a to-do list app for time management. To validate its effectiveness and usability, we designed a graphical UI–based to-do list app with the features of planning, checking tasks, and monitoring. The control group used the app without Moa, whereas the treatment group used the app with the chatbot, receiving CBT sessions for 5 weeks.

We demonstrated the necessity of integrating the chatbot with the to-do app to provide further assistance in facilitating behavior changes, planning, and monitoring. In comparison with the control group, the treatment group exhibited higher variations in time management and procrastination scale scores. Notably, individuals who were not used to maintaining to-do lists yet sincerely communicated with the chatbot reported a significant reduction in procrastination. Conversely, little or no change was observed in the control group unless the participants continued using the to-do list consistently. This result confirms that a to-do app with Moa can instigate behavior changes, particularly for people not experienced in managing tasks. In the treatment group, a higher proportion of individuals subjectively perceived reduced procrastination habits compared to the control group. In addition, according to the results of the thematic analysis, the chatbot provided self-reflection opportunities and emotional support, which are not offered by apps without a chatbot.

Procrastination, rooted in the "avoidance of the implementation of an intention" [69], has multiple underlying causes, including poor time management, pressure-seeking behavior, anxiety, and perfectionism [70]. Although improved time management can mediate procrastination, it does not serve as a comprehensive treatment for it. The chatbot, built on the offline academic procrastination CBT process, offered tailored advice and feedback based on the various causes of procrastination. This principle enabled the chatbot to intervene effectively by following the different procrastination factors for each individual.

The effectiveness of the chatbot originated from the possibility of Moa to provide emotional assistance while forming an alliance with users based on their reviews. The variation in scores for irrational procrastination and stress in the treatment group was significantly greater than that in the control group compared with the other scales from the ANOVA result. Although conversing with Moa cannot equal interacting with humans, users experienced emotional interventions by feeling cared for and supported. This indicated that engaging in conversations with the developed chatbot enabled emotional interventions to influence stress relief and emotional regulation, which are critical for treating irrational delaying behaviors [71].

Despite the addition of the chatbot, the app was evaluated as simple and feasible to use. The only difference was in acceptability, which was higher in the treatment group than in the control group.

Strengths and Limitations

Considering that most time management apps focus on productivity and organizing tasks rather than on the psychological concept of procrastination [7-9,72], this study leveraged a psychological approach by combining CBT with a chatbot to directly address procrastination. The developed chatbot not only addressed procrastination as a behavioral issue but also considered various causes such as stress and perfectionism, thereby enabling tailored feedback. Moreover, it could effectively handle procrastination behaviors grounded in the offline counseling process and reviews provided by counselors and psychologists.

Our findings confirmed that the chatbot not only accelerates behavior changes but also significantly impacts the psychological aspects. This was achieved by fostering introspection and providing valuable support and advice to users, which enabled them to cultivate a sense of connection with the chatbot. These results indicate that chatbots can address procrastination issues from a broader perspective than traditional time management apps.

Although generative chatbots enable natural conversations, they are less suitable for goal-oriented conversations because the overall conversation content can vary based on user utterances [30]. Furthermore, users may experience critical confusion if chatbots respond inappropriately. However, the developed semigenerative chatbot ensured a consistent dialogue within a predefined framework, allowing it to retain the direction of the conversation even when several responses were not entirely reasonable. In other words, the semigenerative chatbot enabled users to engage in relatively natural conversations while maintaining a consistent dialogue flow.

In this study, the control group was allocated to use a to-do app without the chatbot rather than assigning the participants to a

XSL•FO
RenderX

waitlist. This facilitated the determination of distinct characteristics of the chatbot compared with a basic time management app. One perspective suggests that using a to-do app on its own can enable individuals to develop time management skills, establish habits, and initiate behavior changes. Therefore, we aimed to investigate the additional effects of using a chatbot when combined with existing apps. This was in contrast to typical validation studies on new interventions because the proposed approach assessed the benefits of incorporating a chatbot into an existing time management app from various perspectives. Rather than relying solely on self-report questionnaires to compare the treatment and control groups, we conducted a comprehensive analysis of the impact of the chatbot by clustering intervention engagement and overall user experience. These additional analyses provided a detailed understanding of the advantages of chatbot integration. Our study provides practical insights that can be applied to real-world scenarios, facilitating the adoption of chatbot technology in future studies and apps.

We observed the change from before to immediately after the intervention and validated that the effectiveness of the chatbot was prolonged even after the users completed conversing with the chatbot.

However, several limitations were observed in the analysis. Although the intervention effectively changed the behavior of users and received an excellent evaluation in the user test, participants reported troublesome unusual responses because of the fine-tuning of the model using limited data. Continuous errors may affect the satisfaction, trustworthiness, and communicative effectiveness of the app. Therefore, the model performance must be improved to reduce errors. In addition, the finite number of chatbot sessions poses a challenge for sustained use, suggesting the necessity for additional research to enable prolonged engagement. Moreover, participants were obligatorily advised to use the app at least 3 times per week for experimental purposes. However, the feasibility of consistent use without such obligations remains uncertain. Therefore, further research is required in terms of app use in unconstrained environments.

Finally, our target population consisted of university students without severe illnesses, limiting the generalizability of the findings to a broader demographic. To enhance generalizability, future studies should extend their scope to include diverse populations across various age groups and psychopathological conditions.

Conclusions

We developed Moa, a semigenerative chatbot based on the CBT process, to address procrastination and support the behavior changes intended to be derived from time management apps. This study demonstrated that integrating a chatbot into a basic time management app can effectively change user behavior and reduce procrastination based on an RCT. The chatbot addresses procrastination as more than a simple time management issue, recognizing it as a complex interplay of various psychological and behavioral factors. This trial suggests that chatbot intervention can provide additional psychological assistance, such as fostering introspection, forming an alliance with Moa, and providing support and advice, which are not offered by a basic time management app. However, as this study represents the initial stages of chatbot development and implementation, further improvement is necessary in chatbot performance and app functionality.

Acknowledgments

This research was supported by the National Research Foundation of Korea funded by the Korean government (Ministry of Science and Information and Communication Technology; NRF-2020R1F1A1072003) and the Information and Communications Technology and Future Planning for Convergent Research in the Development Program for Research and Development (R&D) Convergence Over Science and Technology Liberal Arts (NRF-2022M3C1B6080866). This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute funded by the Ministry of Health and Welfare, Republic of Korea (HI22C0646). This study was supported by the Ulsan National Institute of Science and Technology–Korea research brand program (1.230016.01) funded by the Ulsan National Institute of Science and Technology. In addition, this research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute funded by the Ministry of Health and Welfare, Republic of Korea (grant HI19C1234).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to participant confidentiality agreements but are available from the corresponding author on reasonable request.

Authors' Contributions

SL contributed to conceptualization, development of the intervention, methodology, ANOVA and thematic analysis, and writing of the original draft. JJ contributed to conceptualization, design of the experiment, analysis of engagement, and writing of the original draft. MK contributed to thematic analysis and revision of the draft. SL contributed to conceptualization, advice for analysis, and revision of the draft. SPK contributed to advice for analysis and revision of the draft. DJ contributed to conceptualization, methodology, revision of the paper, and funding acquisition.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Moa scenario content.
[PDF File (Adobe PDF File), 109 KB - mhealth_v13i1e53133_app1.pdf]

Multimedia Appendix 2
The 6 factors of procrastination—results of the literature review.
[PDF File (Adobe PDF File), 274 KB - mhealth_v13i1e53133_app2.pdf]

Multimedia Appendix 3
Supplementary information, tables, and figures for the experimental results.
[PDF File (Adobe PDF File), 1619 KB - mhealth_v13i1e53133_app3.pdf]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3125 KB - mhealth_v13i1e53133_app4.pdf]

References

1. Steel P. The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. Psychol Bull 2007 Jan;133(1):65-94. [doi: 10.1037/0033-2909.133.1.65] [Medline: 17201571]
2. Rothblum ED, Solomon LJ, Murakami J. Affective, cognitive, and behavioral differences between high and low procrastinators. J Couns Psychol 1986;33(4):387-394. [doi: 10.1037//0022-0167.33.4.387]
3. Hussain I, Sultan S. Analysis of procrastination among university students. Procedia Soc Behav Sci 2010;5:1897-1904. [doi: 10.1016/j.sbspro.2010.07.385]
4. Afzal S, Jami H. Prevalence of academic procrastination and reasons for academic procrastination in university students. J Behav Sci 2018;28(1):51-69 [FREE Full text]
5. Claessens BJ, Van Eerde W, Rutte CG, Roe RA. A review of the time management literature. Pers Rev 2007;36(2):255-276. [doi: 10.1108/00483480710726136]
6. Baker R, Evans B, Li Q, Cung B. Does inducing students to schedule lecture watching in online classes improve their academic performance? An experimental analysis of a time management intervention. Res High Educ 2018 Jul 28;60(4):521-552. [doi: 10.1007/S11162-018-9521-3]
7. Liyanage A, Siriwardana S, Reyal S, Mithsara M. Design-thinking and UCD combination for designing effective time management assistant mobile app. In: Proceedings of the 2021 Conference on International Conference on Human-Computer Interaction. 2021 Presented at: RoCHI '21; September 16-17, 2021; Virtual Event p. 111-118 URL: https://rochi.utcluj.ro/articole/9/RoCHI2021-Ashana.pdf [doi: 10.37789/rochi.2021.1.1.17]
8. Guimarães C, Lazaro DJ. Design of a task-management APP to avoid procrastination. In: Proceedings of the 7th International Conference on Design, User Experience, and Usability: Users, Contexts and Case Studies. 2018 Presented at: DUXU '18; July 15-20, 2018; Las Vegas, NV p. 16-27 URL: https://link.springer.com/chapter/10.1007/978-3-319-91806-8_2 [doi: 10.1007/978-3-319-91806-8_2]
9. The habit of making a pleasant day. Myroutine. URL: https://myroutine.today/ko [accessed 2024-04-29]
10. Danaei R, Hashemi T. Role of perfectionism and perceived stress on behavioral procrastination: mediating role of self-regulation. Int J Philos Soc Psychol Sci 2019;5(2):13-22 [FREE Full text]
11. Özer BU, Saçkes M. International conference on education and educational psychology (ICEEPSY 2010) effects of academic procrastination on college students' life satisfaction. Procedia Soc Behav Sci 2011;12:512-519. [doi: 10.1016/j.sbspro.2011.02.063]
12. Stead R, Shanahan MJ, Neufeld RW. "I'll go to therapy, eventually": procrastination, stress and mental health. Pers Individ Dif 2010 Aug;49(3):175-180. [doi: 10.1016/j.paid.2010.03.028]
13. Hollon SD, Beck AT. Cognitive and cognitive-behavioral therapies. In: Castonguay LG, Lutz W, editors. Bergin and Garfield's Handbook of Psychotherapy and Behavior Change. Hoboken, NJ: John Wiley & Sons; 2013:393-442.
14. Scent CL, Boes SR. Acceptance and commitment training: a brief intervention to reduce procrastination among college students. J College Stud Psychother 2014 Apr 07;28(2):144-156. [doi: 10.1080/87568225.2014.883887]
15. Marrero RJ, Carballeira M, Martín S, Mejias M, Hernández JA. Effectiveness of a positive psychology intervention combined with cognitive behavioral therapy in university students. An Psicol 2016;32(3):728-740 [FREE Full text]

XSL•FO
RenderX

16. Ugwuanyi CS, Gana CS, Ugwuanyi CC, Ezenwa DN, Eya NM, Ene CU, et al. Efficacy of cognitive behaviour therapy on academic procrastination behaviours among students enrolled in physics, chemistry and mathematics education (PCME). J Rat Emo Cognitive Behav Ther 2020 Apr 22;38(4):522-539. [doi: 10.1007/S10942-020-00350-7]
17. Czyz EK, Horwitz AG, Eisenberg D, Kramer A, King CA. Self-reported barriers to professional help seeking among college students at elevated risk for suicide. J Am Coll Health 2013 Oct;61(7):398-406 [FREE Full text] [doi: 10.1080/07448481.2013.820731] [Medline: 24010494]
18. Dutton L. Google Keep. Gale Academic. URL: https://link.gale.com/apps/doc/A431446130/AONE?u=anon~eac4c7f2&sid=googleScholar&xid=ac96de08 [accessed 2025-03-18]
19. Masaat MF, Almeyjidu DA, Saman AK, Kamal MK. Smartphone mobile application time-based features in managing daily activities. Malays J Inf Commun Technol 2023:1-10 [FREE Full text]
20. Adamopoulou E, Moussiades L. Chatbots: history, technology, and applications. Mach Learn Appl 2020 Dec;2:100006. [doi: 10.1016/j.mlwa.2020.100006]
21. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
22. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul 21;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
23. Rimondini M. Communication in Cognitive Behavioral Therapy. Cham, Switzerland: Springer; 2010:2010.
24. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
25. Talebi S, Tong E, Mofrad MR. Beyond the hype: assessing the performance, trustworthiness, and clinical suitability of GPT3.5. arXiv. Preprint posted online June 28, 2023 [FREE Full text]
26. Pereira J, Díaz Ó. Struggling to keep tabs on capstone projects: a chatbot to tackle student procrastination. ACM Trans Comput Educ 2021 Oct 18;22(1):1-22. [doi: 10.1145/3469127]
27. Wang L, Mujib MI, Williams J, Demiris G, Huh-Yoo J. An evaluation of generative pre-training model-based therapy chatbot for caregivers. arXiv. Preprint posted online July 28, 2021 [FREE Full text]
28. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. IEEE Trans Emerg Topics Comput 2021 Jul 1;9(3):1170-1182. [doi: 10.1109/tetc.2020.2974478]
29. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
30. Korngiebel DM, Mooney SD. Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. NPJ Digit Med 2021 Jun 03;4(1):93 [FREE Full text] [doi: 10.1038/s41746-021-00464-x] [Medline: 34083689]
31. Chang S, Gray L, Torous J. Smartphone app engagement and clinical outcomes in a hybrid clinic. Psychiatry Res 2023 Jan;319:115015. [doi: 10.1016/j.psychres.2022.115015] [Medline: 36549096]
32. Schouwenburg HC, Lay CH, Pychyl TA, Ferrari JR. Counseling the Procrastinator in Academic Settings. Washington, DC: American Psychological Association; 2004.
33. Lee E. The relationship of motivation and flow experience to academic procrastination in university students. J Genet Psychol 2005 Mar;166(1):5-15. [doi: 10.3200/gntp.166.1.5-15]
34. Schippers MC, Ziegler N. Life crafting as a way to find purpose and meaning in life. Front Psychol 2019 Dec 13;10:2778 [FREE Full text] [doi: 10.3389/fpsyg.2019.02778] [Medline: 31920827]
35. Lee B. Wellness conversation script dataset. AI Hub. URL: https://aihub.or.kr/aihubdata/data/view.do?currMenu=120&topMenu=100&aihubDataSe=extrldata&dataSetSn=267 [accessed 2024-04-29]
36. Jeon H. KoBERT. GitHub. URL: https://github.com/SKTBrain/KoBERT [accessed 2025-03-06]
37. Jeon H. SKT-AI / KoGPT2. GitHub. URL: https://github.com/SKT-AI/KoGPT2 [accessed 2024-04-29]
38. Flett GL, Hewitt PL. Perfectionism: Theory, Research and Treatment. Washington, DC: American Psychological Association; 2002.
39. Harris NN, Sutton RI. Task procrastination in organizations: a framework for research. Hum Relat 1983 Nov 01;36(11):987-995. [doi: 10.1177/001872678303601102]
40. Hinsch C, Sheldon KM. The impact of frequent social Internet consumption: increased procrastination and lower life satisfaction. J Consum Behav 2013 Nov 21;12(6):496-505. [doi: 10.1002/cb.1453]
41. Kuhl J. A functional-design approach to motivation and self-regulation: the dynamics of personality systems interactions. In: Boekaerts M, Pintrich PR, Zeidner M, editors. Handbook of Self-Regulation. New York, NY: Academic Press; 2000:111-169.
42. Lee DG, Kelly KR, Edwards JK. A closer look at the relationships among trait procrastination, neuroticism, and conscientiousness. Pers Individ Dif 2006 Jan;40(1):27-37. [doi: 10.1016/j.paid.2005.05.010]

XSL•FO
RenderX

43. Nordby K, Klingsieck KB, Svartdal F. Do procrastination-friendly environments make students delay unnecessarily? Soc Psychol Educ 2017 Jul 14;20(3):491-512. [doi: 10.1007/s11218-017-9386-x]
44. Rice KG, Richardson CM, Clark D. Perfectionism, procrastination, and psychological distress. J Couns Psychol 2012 Apr;59(2):288-302. [doi: 10.1037/a0026643] [Medline: 22352949]
45. Schunk DH. Self-efficacy, motivation, and performance. J Appl Sport Psychol 2008 Jan 14;7(2):112-137. [doi: 10.1080/10413209508406961]
46. Svartdal F, Dahl TI, Gamst-Klaussen T, Koppenborg M, Klingsieck KB. How study environments foster academic procrastination: overview and recommendations. Front Psychol 2020 Nov 2;11:540910 [FREE Full text] [doi: 10.3389/fpsyg.2020.540910] [Medline: 33224046]
47. Kim H, Kim H, Lee WK, Han S, Carlbring P, Rozental A. assessing procrastination in Korean: a study of the translation and validation of the pure procrastination scale and a reexamination of the irrational procrastination scale in a student and community sample. Cogent Psychol 2020 Aug 17;7(1):1809844. [doi: 10.1080/23311908.2020.1809844]
48. Löwe B, Spitzer RL, Williams JB, Mussell M, Schellberg D, Kroenke K. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. Gen Hosp Psychiatry 2008 May;30(3):191-199. [doi: 10.1016/j.genhosppsych.2008.01.001] [Medline: 18433651]
49. Lee S, Lim J, Lee S, Heo Y, Jung D. Group-tailored feedback on online mental health screening for university students: using cluster analysis. BMC Prim Care 2022 Jan 25;23(1):19 [FREE Full text] [doi: 10.1186/s12875-021-01622-6] [Medline: 35172741]
50. Hyung-Mo S, Jung-Bum K, Young-Nam P, Dai-Seok B, Sun-Hee L, Hyun-Nie A. A study on the reliability and the validity of Korean version of the Beck Depression Inventory-II (BDI-II). J Korean Soc Biol Ther Psychiatry 2008;14(2):201-212 [FREE Full text]
51. Kim SG, Park J, Kim HT, Pan Z, Lee Y, McIntyre RS. The relationship between smartphone addiction and symptoms of depression, anxiety, and attention-deficit/hyperactivity in South Korean adolescents. Ann Gen Psychiatry 2019 Mar 9;18(1):1 [FREE Full text] [doi: 10.1186/s12991-019-0224-8] [Medline: 30899316]
52. Boucher EM, Raiker JS. Engagement and retention in digital mental health interventions: a narrative review. BMC Digit Health 2024 Aug 08;2(1):52. [doi: 10.1186/s44247-024-00105-9]
53. Schlosser DA, Campellone TR, Truong B, Anguera JA, Vergani S, Vinogradov S, et al. The feasibility, acceptability, and outcomes of PRIME-D: a novel mobile intervention treatment for depression. Depress Anxiety 2017 Jun 18;34(6):546-554 [FREE Full text] [doi: 10.1002/da.22624] [Medline: 28419621]
54. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 2017 Jan 05;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
55. Donghyeock P, Lee MK, Heecheon S. Development and validation of the Time Management Behavior Scale. Korean J Counsel Psychotherapy 2006;18(4):801-816 [FREE Full text]
56. Bong M, Kim SI, Reeve J, Im H, Lee W, Jiang Y, et al. SMILES (Student Motivation in the Learning Environment Scales). SMILE Academic Self-Regulation. 2012. URL: https://bmri.korea.ac.kr/user/kcp670pm/list.asp?listType=3 [accessed 2023-10-01]
57. Park J, Seo Y. Validation of the Perceived Stress Scale (PSS) on samples of Korean University students. Korean J Psychol 2010;29(3):611-629 [FREE Full text]
58. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
59. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
60. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]
61. Brooke J. SUS - a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996:4-7.
62. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
63. Briesch AM, Chafouleas SM, Neugebauer SR, Riley-Tillman TC. Assessing influences on intervention implementation: revision of the usage rating profile-intervention. J Sch Psychol 2013 Feb;51(1):81-96. [doi: 10.1016/j.jsp.2012.08.006] [Medline: 23375174]
64. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979 Jan;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
65. Pinheiro J, Bates D, DebRoy S, Sarkar D, Heisterkamp S, Van Willigen B, et al. Package 'nlme'. Linear and nonlinear mixed effects models, version. URL: https://cran.r-project.org/web/packages/nlme/nlme.pdf [accessed 2024-04-29]
66. Lenth RV. Emmeans: estimated marginal means, aka least-squares means. Cran R. URL: https://CRAN.R-project.org/package=emmeans [accessed 2024-04-29]

XSL•FO
RenderX

67. Hilton A, Armstrong R. Statnote 6: post-hoc ANOVA tests. Microbiologist 2006;7(3):34-36 [FREE Full text]
68. Team R. RStudio: integrated development for R. RStudio. URL: http://www.rstudio.com/ [accessed 2024-04-29]
69. Van Eerde W. Procrastination: self‐regulation in initiating aversive goals. Applied Psychol 2001 Dec 25;49(3):372-389. [doi: 10.1111/1464-0597.00021]
70. Grunschel C, Patrzek J, Fries S. Exploring different types of academic delayers: a latent profile analysis. Learn Individ Differ 2013 Feb;23:225-233. [doi: 10.1016/j.lindif.2012.09.014]
71. Claessens BJ, Van Eerde W, Rutte CG, Roe RA. Things to do today . . . : a daily diary study on task completion at work. Applied Psychol 2010 Feb 16;59(2):273-295. [doi: 10.1111/j.1464-0597.2009.00390.x]
72. Van Eerde W. Time management and procrastination. In: Mumford MD, Frese M, editors. The Psychology of Planning in Organizations. New York, NY: Routledge; 2015:312-333.

Abbreviations

ASRS: Academic Self-Regulation Scale
BAI: Beck Anxiety Inventory
BDI: Beck Depression Inventory
CBT: cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder–7
IPS: Irrational Procrastination Scale
PHQ-9: Patient Health Questionnaire–9
PPS: Pure Procrastination Scale
PSS: Perceived Stress Scale
RCT: randomized controlled trial
TMBS: Time Management Behavior Scale
UI: user interface
URP-I: Usage Rating Profile–Intervention

Edited by L Buis; submitted 03.10.23; peer-reviewed by S Campos, M Augustin; comments to author 04.09.24; revised version received 29.10.24; accepted 26.02.25; published 10.04.25.

Please cite as:
Lee S, Jeong J, Kim M, Lee S, Kim SP, Jung D
Development of a Mobile Intervention for Procrastination Augmented With a Semigenerative Chatbot for University Students: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e53133
URL: https://mhealth.jmir.org/2025/1/e53133
doi:10.2196/53133
PMID:40208664

©Seonmi Lee, Jaehyun Jeong, Myungsung Kim, Sangil Lee, Sung-Phil Kim, Dooyoung Jung. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring the Potential of a Digital Intervention to Enhance Couple Relationships (the Paired App): Mixed Methods Evaluation

Catherine Aicken[1], PhD; Jacqui Gabb[2], PhD; Salvatore Di Martino[3], PhD; Tom Witney[4], PhD; Mathijs Lucassen[5], PhD

[1]School of Education, Sport and Health Sciences, University of Brighton, Brighton, United Kingdom

[2]Faculty of Arts & Social Sciences, The Open University, Milton Keynes, United Kingdom

[3]School of Social Sciences, University of Bradford, Bradford, United Kingdom

[4]Institute of Epidemiology and Health Care, University College London, London, United Kingdom

[5]School of Health and Medical Sciences, City St George's, University of London, London, United Kingdom

Corresponding Author:
Jacqui Gabb, PhD
Faculty of Arts & Social Sciences
The Open University
Walton Hall
Milton Keynes, MK7 6BJ
United Kingdom
Email: jacqui.gabb@open.ac.uk

Abstract

Background: Despite the effects of poor relationship quality on individuals', couples', and families' well-being, help seeking often does not occur until problems arise. Digital interventions may lower barriers to engagement with preventive relationship care. The *Paired* app, launched in October 2020, aims to strengthen and enhance couple relationships. It provides daily questions, quizzes, tips, and detailed content and facilitates in-app sharing of question and quiz responses and tagged content between partners.

Objective: To explore the potential of mobile health to benefit couple relationships and how it may do this, we examined (1) *Paired's* impact on relationship quality and (2) its mechanisms of action.

Methods: This mixed methods evaluation invited *Paired* subscribers to complete (1) brief longitudinal surveys over 3 months (n=440), (2) a 30-item web-based survey (n=745), and (3) in-depth interviews (n=20). For objective 1, survey results were triangulated to determine associations between relationship quality measures and the duration and frequency of *Paired* use, and qualitative data were integrated to provide explanatory depth. For objective 2, mechanisms of action were explored using a dominant qualitative approach.

Results: Relationship quality improved with increasing duration and frequency of *Paired* use. Web-based survey data indicate that the Multidimensional Quality of Relationship Scale score (representing relationship quality on a 0-10 scale) was 35.5% higher (95% CI 31.1%-43.7%; *P*=.002), at 7.03, among people who had used *Paired* for >3 months compared to 5.19 among new users (≤1 wk use of *Paired*), a trend supported by the longitudinal data. Of those who had used *Paired* for >1 month, 64.3% (330/513) agreed that their relationship felt stronger since using the app (95% CI 60.2%-68.4%), with no or minimal demographic differences. Regarding the app's mechanisms of action, interview accounts demonstrated how it prompted and habituated meaningful communication between partners, both within and outside the app. Couples made regular times in their day to discuss the topics *Paired* raised. Daily questions were sometimes lighthearted and sometimes concerned topics that couples might find challenging to discuss (eg, money management). Interviewees valued the combination of fun and seriousness. It was easier to discuss challenging topics when they were raised by the "neutral" app, rather than during stressful circumstances or when broached by 1 partner. Engagement seemed to be enhanced by users' experience of relationship benefits and by the app's design.

Conclusions: This study demonstrates proof of concept, showing that *Paired* may have the potential to improve relationship quality over a relatively short time frame. Positive relationship practices became embedded within couples' daily routines, suggesting that relationship quality improvements might be sustained. Digital interventions can play an important role in the relationship care ecosystem. The mixed methods design enabled triangulation and integration, strengthening our findings. However, app users were self-selecting, and methodological choices impact our findings' generalizability.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e55433) doi:10.2196/55433

KEYWORDS

digital intervention; couple relationships; romantic relationships; relationship quality; app; digital technology; couples; internet-based intervention; evaluation research; mobile phone

Introduction

Background

Due to its prevalence and impacts, poor relationship health has been identified as a public health issue and deterioration of couple relationships as an epidemic [1]. Poor relationship quality (and similar constructs; eg, marital strain) negatively affects individual partners' well-being [2,3], mental health [4,5], and physical health [6,7]; moreover, children are negatively impacted by poor parental relationship quality [8], conflict [9,10], and divorce [11,12]. Although an overwhelming majority of people worldwide ever marry [13], relationship distress and breakdowns are common. Approximately 50% of marriages in the United States and >40% of marriages in the United Kingdom end in divorce [14,15]; in addition, in the United Kingdom, 18% of ongoing couple relationships are estimated to be distressed [16].

Despite the importance of couple relationships, help seeking for relationship problems is often delayed or inadequate. Among people who are currently in a relationship, half of US adults, as well as a higher proportion of UK adults, do not seek advice for relationship issues from any source (not even web-based searches), and among those who do, the most common sources are friends and family [17]. Rigorously evaluated interventions such as couple therapy [18,19] and couple relationship education [20] (terms we use to include marital therapy and marital relationship education, respectively) are underused compared to less evaluated sources of advice (eg, self-help books; or talking with family, friends, or religious leaders) [21]. Possible reasons include accessibility, cost, social stigma, and the need for both partners' simultaneous participation [22,23]. Couple relationship education is preventive, seeking to promote healthy relationships; yet, it often has a religious basis, conventionally focusing on marriage [24]. There is an opportunity for relationship care interventions that are independent of faith-based organizations and inclusive of diverse relationships.

Potential for Digital Interventions in Relationship Care

Digital interventions may increase the accessibility of relationship care, their use fits with contemporary couple behaviors, and they may be effective in supporting behavior change. Taking these points in turn, compared to face-to-face interventions, mobile health (mHealth) interventions can be delivered discreetly, which is appropriate for sensitive or stigmatized matters, and they may further increase access by overcoming barriers such as inconvenience and cost [25]. Communications technology is embedded in many areas of contemporary coupledom: meeting romantic and sexual partners on the web is common [26-29], and information and communications technologies are used to complement couples' in-person communication, sustain intimacy while apart, and end relationships [30]. Digital interventions can facilitate self-monitoring, which raises awareness and thus facilitates self-management of users' behavior and emotions [31], and through cues, routines, and rewards, apps may be effective in habit formation within relationships [32].

Although digital interventions can be effective for supporting behavior change across diverse health areas [33-36], disengagement is generally high [37,38], and app quality is often poor. Many of the >250 new mHealth apps that are available daily [39] are not based on robust evidence or theory and may contain inaccurate or harmful information [40,41]. Empirical evidence on digital interventions for couples lags behind the emergence of new apps [42]. A recent review [42] found that most digital couple interventions were treatment focused (eg, for relationally distressed couples or for prevention of intimate partner violence) or lacked clarity about whether they were treatment focused or for primary prevention (ie, positive relationship care), making it unclear who should use them. The digital couple intervention that was most clearly focused on primary prevention consisted of a simple 1-time relationship assessment, while others were modular courses for couples to complete together, often linked to professional or coaching support [42]. Although human support may enhance engagement, these "blended" and structured interventions may negate some of the accessibility, convenience, and scalability advantages that mHealth can offer. We offer an evaluation of a fully digital and flexible intervention.

The Paired App

Paired is a commercially available mobile app, launched in October 2020 and designed to help couples enhance their relationships. Its intended users are couples at any stage in their relationship, including same- and opposite-sex couples. It is not intended for couples experiencing relationship distress, who may require intensive intervention, and there is no human input targeted at the individual or couple (ie, it is not a blended or guided intervention). It is not based on any single theory of how relationships are sustained or improved but was developed with input from relationship science experts and informed by findings from the *Enduring Love?* study [43]. This study built upon practices theorizing—notably family practices [44] and practices of intimacy [45]—to examine the ways that daily interactions generate relationship quality, developing the concept of everyday relationship work [43]. The findings highlighted the importance of daily gestures and relationship work in the maintenance of long-term couple relationships [43]. During our study, *Paired* was available for free, in English, and in multiple countries, with most users in the United States and the United Kingdom. A paid-for version provides access to additional content and features. New content is continually added, and there is no defined course or set sequence of activities to complete. The app's main features and functionality—described in the next subsection—did not change during data collection (and were common to free and paid-for versions). Therefore,

XSL•FO
RenderX

this study evaluates an early version of *Paired* within the first 3 months of its launch.

Paired provides daily questions and weekly quizzes, intended to prompt couples to have frequent, open conversations on diverse topics. If individuals link their app account with their partner's, question and quiz responses become mutually available when both partners have responded. Users can reply to their partner within the app and receive tips and links to preprepared topical content, including from therapists and academics specializing on couple relationships (which are searchable and accessible at any time). *Paired* can also be used independently: a user can access content and discuss questions with a partner who does not have the app.

This Study

Approach and Theoretical Basis

Digital health interventions are complex interventions, operating within complex systems [46]. *Paired* has several dimensions of complexity; for example, users "receive" the intervention differently because it has multiple components that can be engaged with in different ways and because users engage with it for as long and as frequently as they choose [47,48] (vs completing a course). It may lead to improvement in relationship quality through complex causal pathways, and its use and effectiveness may be shaped by the social and relationship contexts and settings in which it is used [49]. It is also complex because it requires unaided use of technology [50] (ie, users' devices and the app), relying on digital literacy skills that may differ between users.

To evaluate *Paired*, we drew on established guidance for evaluating complex interventions [50]. Evaluations can address a range of related questions, such as whether, in what contexts, and for whom an intervention works [51]; how it works; and how it may be further developed. Mixed methods and interdisciplinary approaches are recommended for evaluating complex interventions [52], specifically digital interventions [53-55]. The integration of qualitative and quantitative approaches can aid an understanding of user behavior and issues affecting intervention success [56] and strengthens the conclusions that can be drawn [57]. Qualitative research aids the development of causal explanations by describing the processes that produce an outcome [58].

Relationship quality, our main outcome, is measured in various ways [59]. We believe it is multidimensional [60] (ie, it entails factors such as communication quality and how couples deal with conflict) and is best assessed as such (information received from Di Martino et al [email, January 16, 2025]). In this study, we use a broad definition of relationship quality: "how positive or negative individuals feel about their relationship" [61], which acknowledges that it is subjective yet measurable at the individual level.

We developed a provisional theory of change to guide our evaluation, representing how we expected that *Paired* might "work" based on findings from relationships research and behavioral science. This included the following 3 strands:

- The functionality, design, and content of the app may prompt or facilitate within-couple communication about the relationship (links between improved couple communication and improved relationship quality are well substantiated [62]).
- Daily notifications and in-app interactions may prompt or facilitate daily conversations and "relationship work" by the couple, which may help users learn and advance relationship maintenance skills, including daily gestures [43], benefiting the relationship.
- "Dose-response" (where "dose" is duration and frequency or intensity of use): greater use of *Paired* may lead to greater improvements in relationship quality.

We sought to develop this provisional theory through our research. We consider the role of *Paired* within relationships from a digital sociology perspective in the Discussion section.

Aim and Objectives

This study aimed to explore the potential for an mHealth intervention to benefit relationships and how it might do so, using *Paired* as an exemplar. The objectives were (1) to assess the app's impact on relationship quality by examining associations between relationship quality and duration, frequency and intensity of use, and by comparing the perceived impact of *Paired* among its users (as indicative of the direction of causation); and (2) to develop a refined, empirically informed understanding of how *Paired* may deliver improved relationship quality. To put it simply: (1) "Does it appear to work?" (2) "How does it work?"

Methods

Study Design and Participants

The app was evaluated using mixed methods among users of *Paired*. Quantitative and qualitative datasets were analyzed separately, and the findings were integrated (parallel design). Objective 1 was addressed using primarily quantitative approaches, with qualitative data providing explanatory depth and evidencing the likely direction of causation. Objective 2 was addressed primarily qualitatively, supplemented by quantitative findings, because this objective is exploratory. In concordance with our aim (to explore the intervention's potential), we make no claim of generalizability to all *Paired* users, and our sampling methods reflect this.

Recruitment, Sampling, and Procedures

Data collection materials, including the wording of survey items, are provided in Multimedia Appendix 1.

Brief in-App Survey

At 3 time points with 1-month intervals, starting October 30, 2020, the *Paired* weekly quiz was replaced by a 5-item in-app survey, identified to users as university-led research. This brief survey asked for agreement or disagreement (on a 5-point Likert scale) with 4 statements about aspects of relationship quality (ie, communication, emotional connection, conflict, and sex and intimacy). One further statement concerned participants' perception of the impact of *Paired* on their relationship communication. Researchers obtained a deidentified extract of

XSL•FO
RenderX

data from *Paired* users who had completed at least 1 brief survey. This included data from a "relationship check-up" that users can complete when they first download the app, containing 4 statements similar or identical to those in the brief survey, which we used as baseline measures (the data extract also included demographic details for a small minority of users, which were not used). From 3717 unique participants, the sample was restricted to 440 (11.84%) individuals who completed the relationship check-up in October 2020 and all 3 brief surveys thereafter (4 data points), with at least 7 days between each data point to allow for a realistic prospect of change (in explanation, the data showed that some users delayed completion of the brief surveys; if they completed, say, November's survey just before completing December's survey, we would not expect to detect much change). This gave complete longitudinal data spanning approximately 3 months (October-December 2020), referred to as "brief survey data" for brevity.

Web-Based Survey

A 30-item survey, administered in December 2020 and hosted securely outside the app, collected quantitative and limited free-text data. *Paired* users were invited to complete it via 3 in-app messages containing the survey link. Those willing to be contacted regarding a research interview provided email addresses.

The survey collected demographic data but not race or ethnicity, religiosity, income, or socioeconomic status (despite well-documented associations with marital satisfaction and quality in the United States [63-65]). These constructs are differently defined and delineated in different cultures, posing challenges for our relatively small-scale international study.

From the web-based survey data, we obtained self-reported duration and frequency of *Paired* use, as well as relationship quality, measured by (1) the Multidimensional Quality of Relationship Scale (MQoRS) and (2) direct questions about relationship quality change since using the app. The MQoRS, developed for this study, combines and weights responses to 17 statements on 5 aspects of relationship quality and is expressed on a scale ranging from 0 to 10. Suited to smartphone-based self-completion (ie, with relatively few survey items comprising simple statements and with responses using 5-point Likert scales that fit on the smartphone's screen, thus requiring minimal scrolling), it overcomes the shortcomings of unidimensional relationship quality measures and is described in detail elsewhere (Multimedia Appendix 2).

In-Depth Interviews

Individual interviews were conducted on the web with purposively sampled web-based survey participants (January-April 2021). The interviewer (TW) identified himself as a university researcher, independent from *Paired*. The primary sampling criteria were gender and country, with initial targets of 5 women and 5 men each from the United States and the United Kingdom. The secondary sampling criteria, by which we sought diversity across the sample, were sexuality, relationship duration, cohabitation, and the presence or absence of children in the household. The interviews followed a topic guide and lasted 40 to 75 minutes. Similar descriptions of *Paired* use and its impacts were heard repeatedly toward the end of data collection, suggesting that saturation had been reached.

Analyses and Integration

Objective 1: Impact of Paired on Relationship Quality

Quantitative data were analyzed using SPSS software (version 26.0; IBM Corp) for descriptive statistics and R software (R Foundation for Statistical Computing) for inferential statistical analyses and data visualization. Statistical significance was considered as $P<.05$.

Brief longitudinal survey data were analyzed using multilevel multinomial logistic regression models with cumulative logit link functions. Predicted values were used to generate odds ratios (ORs) and 95% CIs for change in each dimension of relationship quality over time compared to baseline (the initial "relationship check-up").

Drawing on data from the web-based survey, MQoRS scores were compared between different categories of reported duration and frequency of using *Paired*, using 1-way ANOVAs with Welch approximation due to unbalanced sample sizes and post hoc Tukey tests. Responses to direct questions about change in relationship quality since using *Paired* were compared using chi-square tests to make binary comparisons between people reporting using *Paired* for ≥1 month versus <1 month and people reporting using *Paired* on 6 to 7 days per week versus ≤5 days per week (among those reporting <1 mo use). Of note, because we lacked the data to predict our sample and subsample sizes in advance, decisions on the cutoffs for these categories were not made a priori but were data driven and were also informed by our provisional theory of change.

Evidence from the 2 surveys on the association between relationship quality and the duration and frequency of *Paired* use was triangulated. The direction of causation of this association was explored qualitatively, informing the development of the theory of change (objective 2).

We compared the perceived effectiveness of *Paired* between demographic groups, using the proportions agreeing with statements in the web-based survey (detailed in the Results section and Multimedia Appendix 3). We used chi-square tests for association and univariable logistic regression to obtain ORs for demographic differences (using the characteristics we used for interview sampling and additionally age, reflecting widespread age-related assumptions regarding technology engagement). Analysis was restricted to the 514 participants who reported ≥1 month's use of *Paired* (because it is perhaps less plausible that change in relationship quality among new users is attributable to *Paired*). Qualitative evidence to help explain quantitative findings was provided by interview responses to direct questions and spontaneous remarks about the app's inclusivity.

Objective 2: Theory of Change

TW and CA, both experienced qualitative researchers, conducted the qualitative analysis. We used reflexive thematic analysis [66] to describe how *Paired* seems to work, using a largely inductive process but also informed by the provisional theory

XSL•FO
RenderX

of change. "Dosage" and its association with relationship quality were explored quantitatively (refer to the Objective 1: Impact of Paired on Relationship Quality subsection) and qualitatively.

Ethical Considerations

Ethics approval was provided by The Open University Human Research Ethics Committee (brief survey including linked data: HREC/3759/Gabb; web-based survey and interviews: HREC/3797/Gabb). After reading the study information, web-based survey participants indicated their informed consent on the web before survey completion and reconfirmed this afterwards. They were offered the chance to win one £100 e-voucher (or equivalent value, ie, US $135) as a thank you. Individual interview participants provided oral informed consent. The interviewees were sent an e-voucher worth £20 or US $27 as a thank you.

Results

Sample Descriptions

Of the 3717 brief survey participants, 440 (11.84%) provided complete data suitable for longitudinal analysis.

A total of 745 participants completed the web-based survey, with diversity by relationship type and duration, sexual orientation, age, relationship status, and parenthood. Comparison with aggregate data from *Paired* suggests that this sample was broadly representative of *Paired* users (Multimedia Appendix 4).

Within the sample of 20 interviews, quotas were filled, and diversity was achieved by all secondary sampling criteria (Table 1). With each quote, we provide age and sampling characteristics.

Table 1. Qualitative interview sample (N=20).

Sampling criteria and categories	Interviewees, n (%)
Primary sampling criteria	
UK woman	5 (25)
UK man	5 (25)
US woman	5 (25)
US man	5 (25)
Secondary sampling criteria	
Sexuality	
Heterosexual	14 (70)
LGBTQ+[a]	6 (30)
Relationship duration (y)	
≤1	6 (30)
1-5	8 (40)
>5	6 (30)
Living with partner? (all married or civil partnered couples were cohabiting)	
Yes (cohabiting)	14 (70)
No (living apart together)	6 (30)
Children aged <18 y living in household?	
Yes	6 (30)
No	14 (70)

[a]LGBTQ+: lesbian, gay, bisexual, transgender/trans, queer, and other sexuality and gender minoritized individuals.

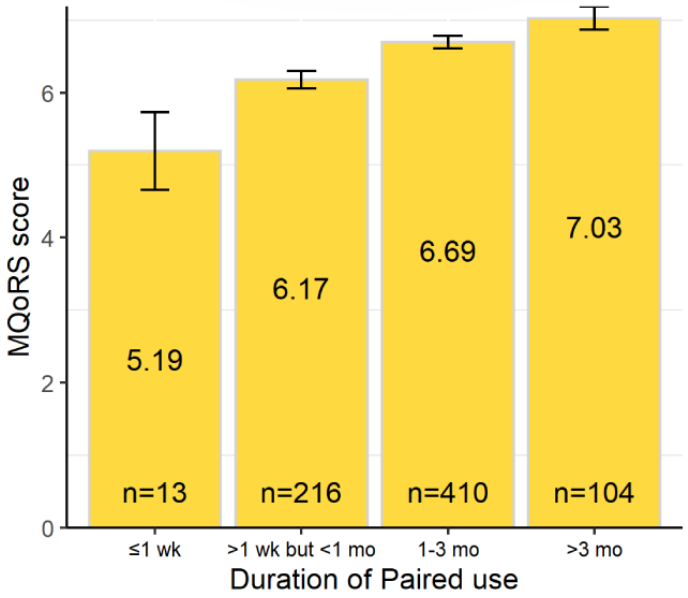
Objective 1: Paired Use and Relationship Quality

Overview

Web-based survey results show that relationship quality, measured by the MQoRS, was 35.5% higher among those with >3 months' *Paired* use compared to new users (5.19 among people who reported having used the app for ≤1 wk vs 7.03 among those who reported >3 mo use; 95% CI 31.1%-43.7%; *P*=.002; Figure 1). Supporting this, 59.5% (440/740) agreed or strongly agreed that their relationship felt stronger since using *Paired* (95% CI 56%-63%), and people who had been using the app for ≥1 month were more likely to report this than newer users (330/513, 64.3%, 95% CI 60.2%-68.4% vs 110/227, 48.5%, 95% CI 42%-55%; *P*<.001).

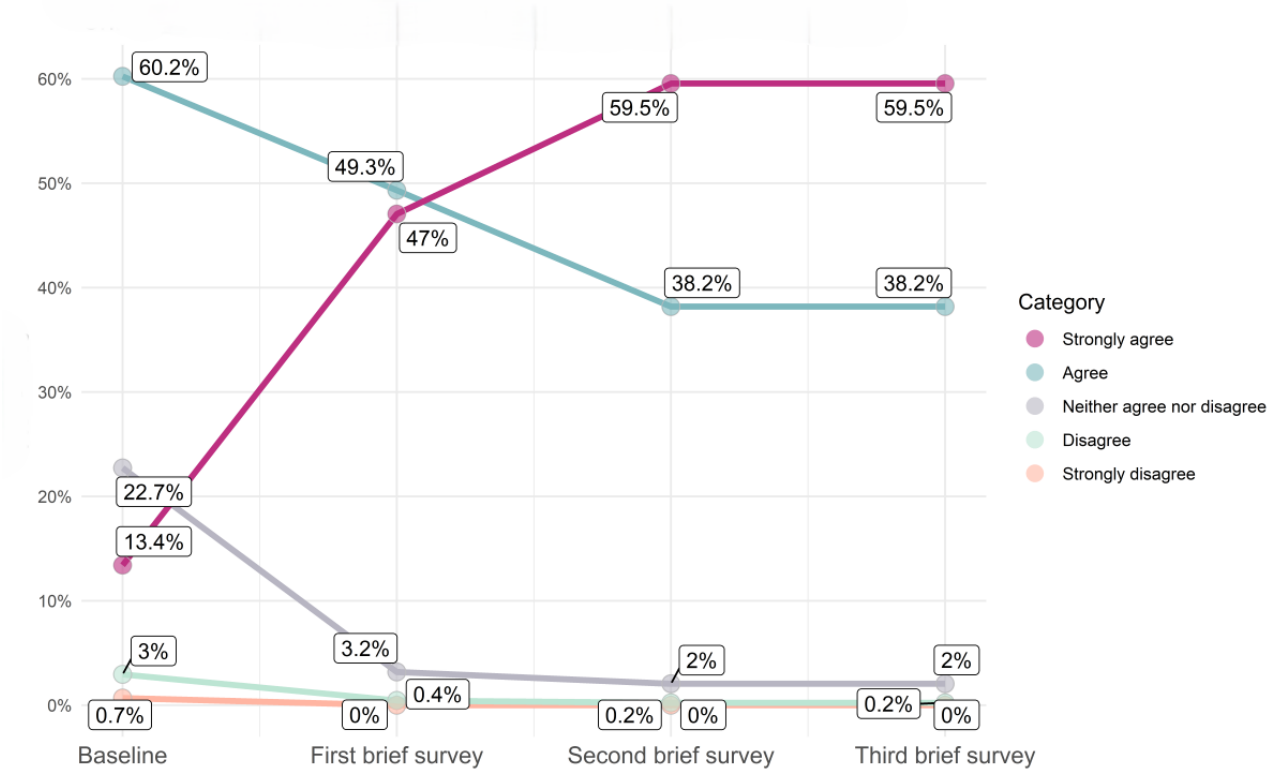
XSL•FO
RenderX

Figure 1. Multidimensional Quality of Relationship Scale (MQoRS) scores by duration of Paired use (showing SDs).



The longitudinal brief survey data support this dose-response finding, showing positive changes over time in a cohort of *Paired* users in 4 distinct aspects of relationship quality (Figures 2-5; for each data point in Figures 2-5 [x-axis], the percentages sum to 100%; an increase in the extent of agreement is particularly marked in Figures 2-4, where there is an increase in the percentages responding "strongly agree"; the decreasing percentages responding "agree" represent a shift toward this stronger level of agreement [and not a decline in overall agreement]). The web-based survey analyses also support this (Multimedia Appendix 3).

Figure 2. Change over time in responses to the statements "I am very satisfied with how we communicate with each other" and "We communicate openly with each other."



XSL•FO
RenderX

Figure 3. Change over time in response to the statement "We are able to discuss and resolve conflict."

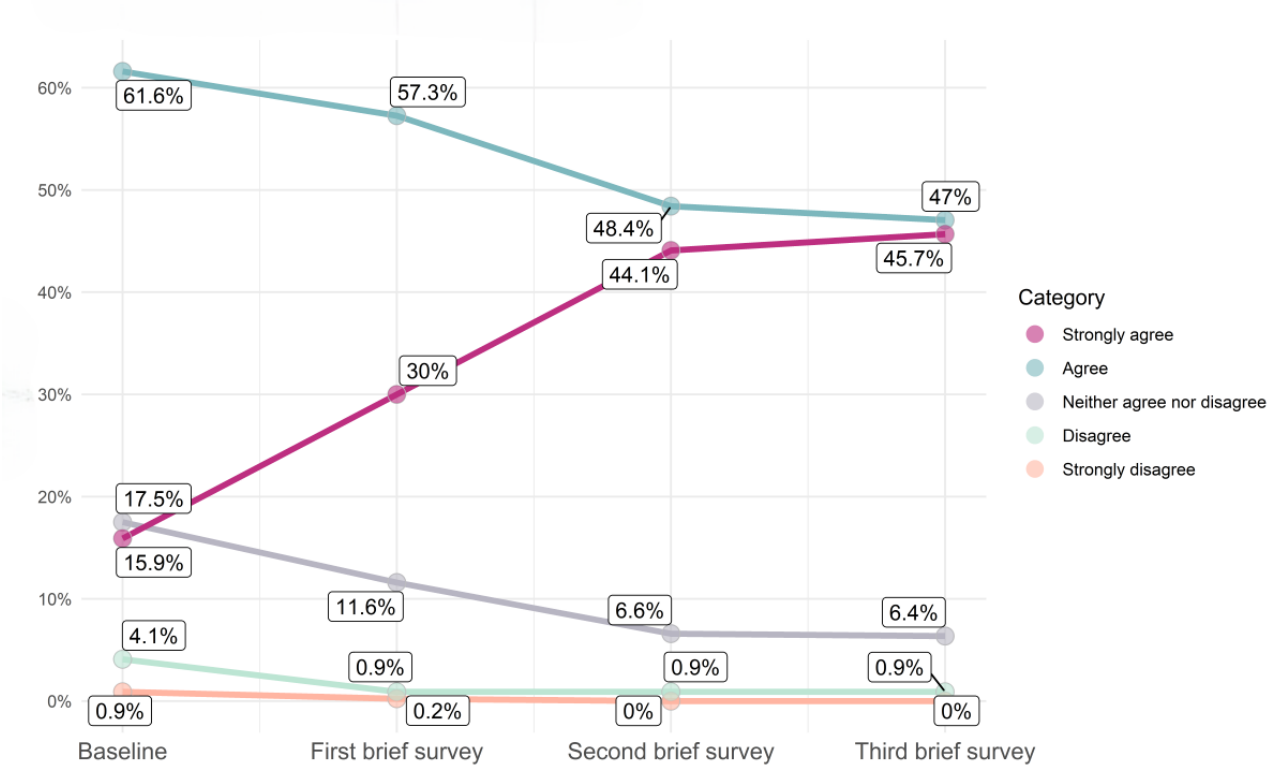
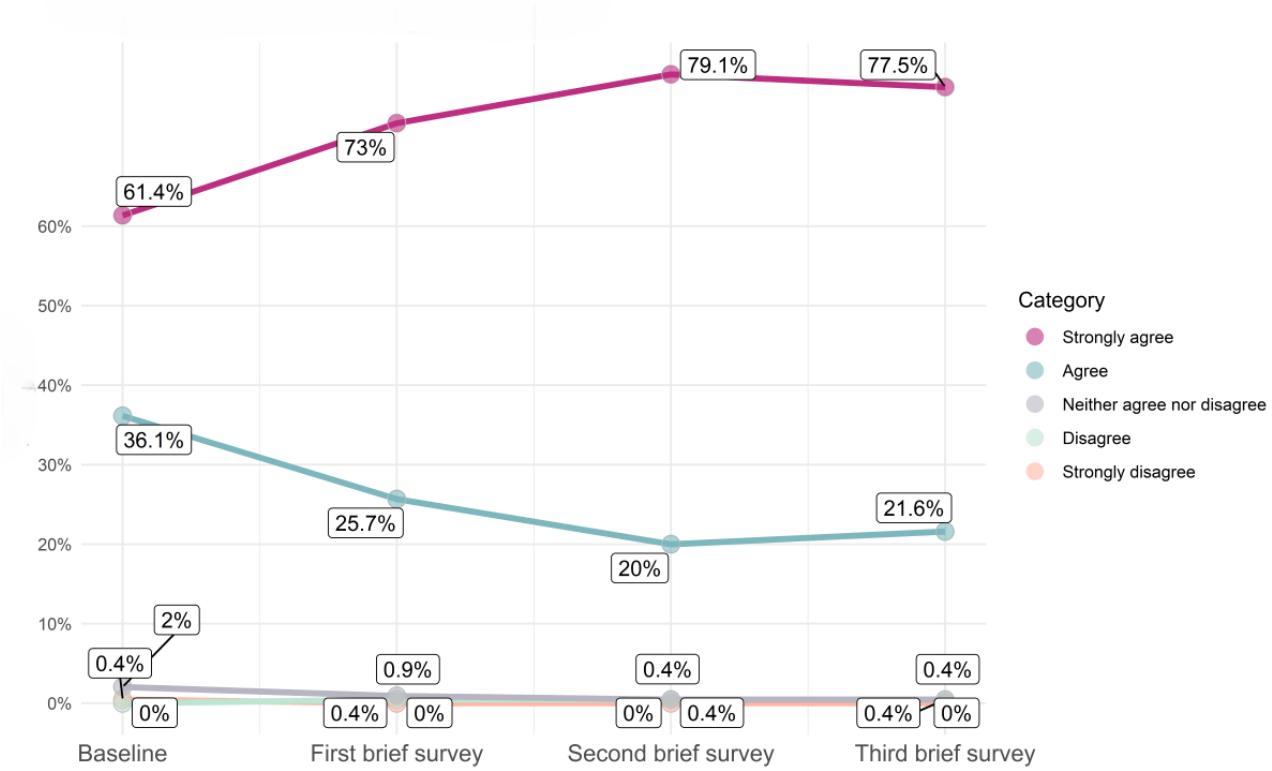
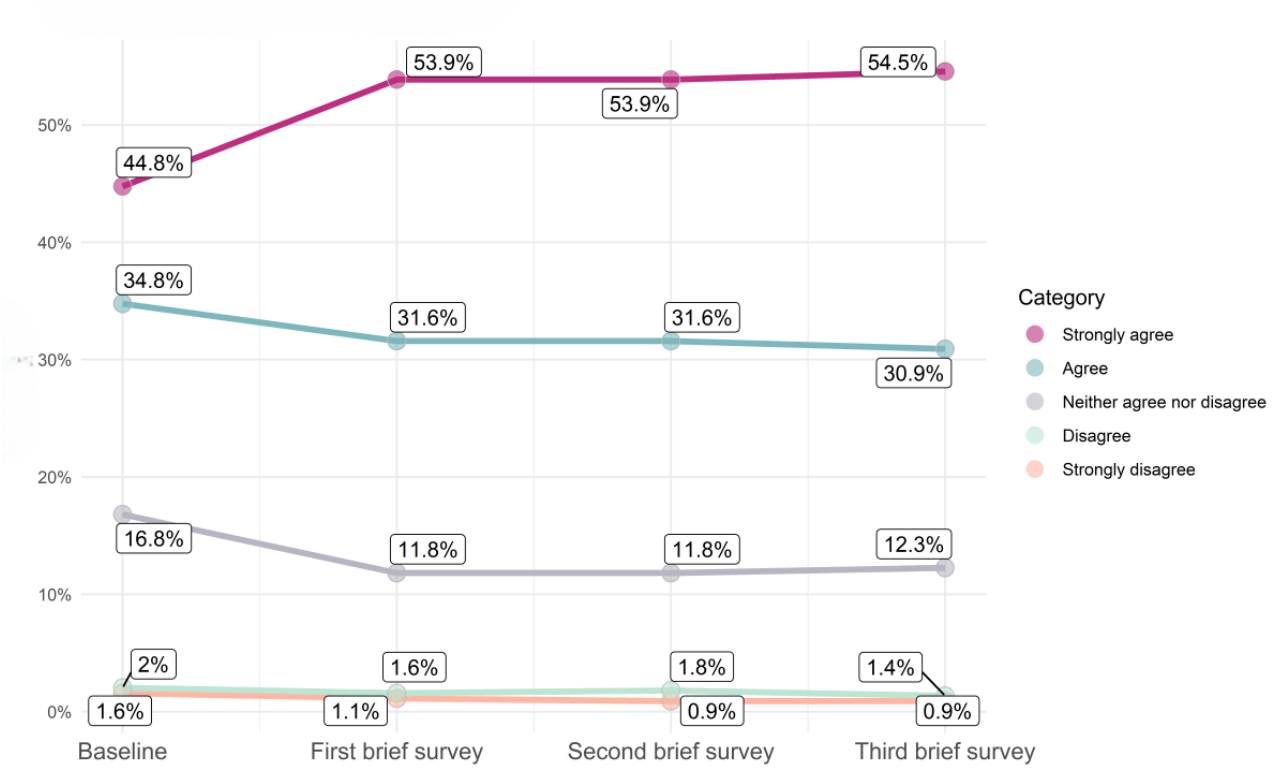


Figure 4. Change over time in responses to the statements "I feel connected with my partner emotionally" and "We enjoy a positive emotional connection."



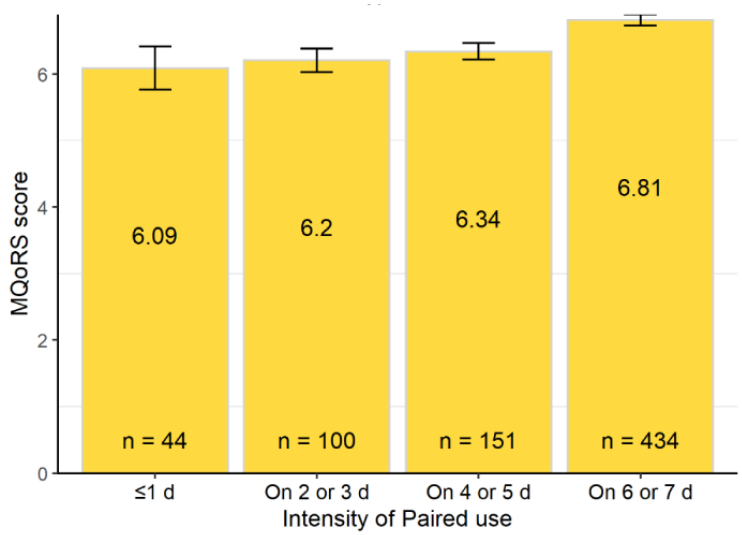
XSL•FO
RenderX

Figure 5. Change over time in response to the statement "We are comfortable with discussing our sex life (with each other)."



Reporting more frequent *Paired* use was associated with higher MQoRS scores: people using *Paired* on 6 to 7 days per week had 11.8% higher relationship quality than those using *Paired* on ≤1 day per week (MQoRS scores: 6.81 vs 6.09), and this difference was statistically significant (95% CI 0.19-1.42; *P*=.04; Figure 6; details in Multimedia Appendix 2). Supporting this finding, among people who had used *Paired* for at least 1 month, those who reported using it on 6 to 7 days in a typical week were more likely to agree or strongly agree that their relationship felt stronger since using *Paired* compared to those who used it less often (223/316, 70.6%, 95% CI 65.5%-75.6% vs 106/196, 54.1%, 95% CI 47.1%-61.1%; *P*<.001). However, we found no statistically significant association between the total amount of time spent using *Paired* in a typical week (intensity of use) and MQoRS scores.

Figure 6. Multidimensional Quality of Relationship Scale (MQoRS) score by number of days using Paired in a typical week (showing SDs).



Comparison of Perceived Effectiveness by Demographic Characteristics

Among those who reported having used Paired for ≥1 month, there was no statistically significant difference in the proportion of web-based survey participants who agreed or strongly agreed that "Paired is improving how we communicate as a couple" (overall 414/514, 80.5%), "Our relationship feels stronger since we've been using Paired" (overall 330/513, 64.3%), or "The longer I use Paired, the better my relationship gets" (overall

XSL•FO
RenderX

252/514, 49%) between men and women or by age group, sexuality, country, relationship status, or relationship duration (refer to Multimedia Appendix 3 for details of the demographic categories used). Participants without children (aged <18 y) in their household were more likely to agree that their relationship felt stronger since using *Paired* compared to those living with children (OR 1.49, 95% CI 1.00-2.21), but this was of borderline statistical significance (*P*=.048), and no differences were observed in agreement with the other statements.

Regarding the qualitative data, the inclusivity of *Paired* was described positively; for example, it was described as suitable for a range of ages ("kind of age agnostic" [UK02, woman, aged 58 y, heterosexual, relationship duration: ≤1 y, living apart, no children]) and as potentially benefiting couples in different ways over the course of their relationship ("as a prompt of finding out more about your partner early on, or [for] mature couples, the stuff you take for granted" [UK07, man, aged 36 y, heterosexual, relationship duration: >5 y, cohabiting, children]). While some commented that some of the app's content did not apply to their relationship (eg, sharing domestic chores among noncohabiting couples), this was not necessarily perceived negatively. Interviewees valued questions and content that were broadly relevant to diverse couples. As an interviewee in a same-sex relationship explained, he had never come across a question which was not inclusive:

> *I've never furrowed my brow at a question, [and never thought] oh my god how heteronormative, how cis, clearly it's a bunch of cisgendered white men programming this app.* [US01, man, aged 34 y, LGBTQ+ (lesbian, gay, bisexual, transgender/trans, queer, and other sexuality and gender minoritized individuals), relationship duration: ≤1 y, cohabiting, no children]

> *We've found it more useful than faith-based apps. And I think sometimes you can be a Christian couple that has conversations. You don't have to be a Christian couple that has Christian couple conversations.* [UK10, man, aged 27 y, LGBTQ+, relationship duration: 1-5 y, living apart, no children]

Another interviewee considered *Paired* "just so useful for anyone" but wondered whether some people would be reticent to start using it:

> *...there's plenty of those "typical guys" who are going to be like, "I don't need no stupid app to get to know this and this" [...] But I feel like once someone gets into it and starts using it and understands how beneficial it is that it can pretty much benefit anyone.* [US04, man, aged 38 y, heterosexual, relationship duration: >5 y, cohabiting, no children]

Objective 2: Refined Theory of Change

Overview

A total of 4 themes were identified from the qualitative analysis, 3 (75%) of which drew upon the provisional theory of change (refer to the Approach and Theoretical Basis subsection in the Introduction section). Subthemes and the theme *Engagement* (which describes how *Paired* use is sustained) were derived from the data. Theme and subtheme descriptions (subheadings) in this subsection summarize the refined theory of change.

Communication: Practicing Communication to Strengthen Emotional Connection and Improve Relationship Communication Skills

The App Is a Neutral Prompt for Relationship Conversations

Interviewees found it uncomfortable to raise difficult topics and felt as though they were nagging their partner by doing so. The following interviewee describes the consequences of this:

> *...a much more tense, stressful situation to talk about, than "this wasn't on my mind but the app brought it up so why don't we talk about it?"* [US06, woman, aged 44 y, LGBTQ+, relationship duration: 1-5 y, cohabiting, children]

Without the app, couples might avoid discussing a challenging topic until circumstances meant that it had to be addressed, whereupon they might struggle to communicate without arguing:

> *[Discussions about bills] used to be sort of heated arguments. [But] Paired allowed us to kind of discuss such things more easily and freely because we were discussing them in the context of a question that it says, discuss.* [UK05, man, aged 45 y, heterosexual, relationship duration: >5 y, cohabiting, children]

By interspersing serious and lighter questions, relationship communication became easier. As an interviewee described, this brought "balance" to relationship communication between her and her partner:

> *A lot of the questions that the app has are ones we can talk about easily without one of us getting upset, and it makes it easier to talk when you don't dread it.* [US07, woman, aged 23 y, heterosexual, relationship duration: 1-5 y, cohabiting, no children]

Questions and Quizzes Help Couples Get to Know Each Other Better, Stay Connected, and Refresh or Deepen Intimacy

Interviewees described how the app's daily questions and quizzes helped them to gain new insights about their partner and their relationship, which informed changes to their own behavior:

> *...it's definitely given me more insight on his position, how he feels [...] I didn't realize he was really not sure about how he expresses his feelings, so now I can take some time and try to really draw that out before we discuss.* [US07, woman, aged 23 y, heterosexual, relationship duration: 1-5 y, cohabiting, no children]

Question and quiz responses could remind couples of what they shared:

> *It's funny there was one question about a favorite memory of the year [...] [We wrote] the exact same memory. Yeah, I think it has been reassuring. It has reinforced our already strong connection.* [US06,

XSL•FO
RenderX

woman, aged 44 y, LGBTQ+, relationship duration: 1-5 y, cohabiting, children]

This enabled interviewees to gain "new insights" about themselves and each other and to "expand and communicate better" through questions coming from the app, a "third party" (UK04, woman, aged 23 y, LGBTQ+, relationship duration: >5 y, living apart, no children). It helped new couples get to know each other, and long-term couples to become "better connected." For example, an interviewee noticed his wife's responses to some of the app's questions changing over time, which, he felt, helped sustain the connection between them:

> *[I'm] no longer keeping this stale, unevolving picture of who she is and I think that is the key to being connected to someone, to remain connected.* [US04, man, aged 38 y, heterosexual, relationship duration: >5 y, cohabiting, no children]

As a way of maintaining emotional connection, *Paired* was welcomed by couples in contrasting contexts. This included physical separation (eg, a long-distance relationship, COVID-19 quarantine, or self-isolation) and being in lockdown together. An interviewee described how while he and his wife were working from home and homeschooling their children together, his wife was "just overstimulated with communication that she already has with the kids and work and stuff like that." He explained how using *Paired* helped improve the frequency and quality of their relationship communication, making it "easy to just keep that tank topped up a bit" when they were "very much in survival mode" (UK07, man, aged 36 y, heterosexual, relationship duration: >5 y, cohabiting, children).

Relationship Communication Skills Are Learnt and Developed Through Regular Practice, Sometimes Supplemented by Guidance

By regularly responding to and discussing questions and quizzes, accessing content and tips, and putting their learning into practice, couples developed their relationship skills. They observed how these skills improved. One interviewee stated that "the communication is so much better and it's stopped a lot of shouting at each other," and she went on to explain how when arguments begin, they are more able to stop and give each other space before apologizing; app notifications could distract the couple from an argument, prompting them to engage with *Paired*, which "just changes the whole atmosphere again, which is just amazing" (UK09, woman, aged 32 y, heterosexual, relationship duration: 1-5 y, cohabiting, children).

Within and between couples, there was a wide range of use of the additional content and resources that *Paired* provides: some read widely, others rarely accessed these parts of the app (instead mostly using questions and quizzes). Some shared what they found with their partner:

> *If maybe we're like not super aligned on [daily questions], I know sometimes they give you tips so I'll read it and I'll be like hey read this [...] having that knowledge in the background is more helpful if it comes to a conflict later.* [US10, woman, aged 25 y, heterosexual, relationship duration: ≤1 y, living apart, no children]

Engagement: Getting Hooked on the App, Getting Hooked on Each Other

Being Interested and Curious About New Content—and About Each Other

Regular engagement is prompted via the app (eg, automated in-app messages and daily questions) and by partners reminding each other to complete questions or quizzes or sharing articles. Interest and curiosity provoked their engagement:

> *I'll get notifications that she has answered it and I want to see what she's said so I go onto my one.* [UK06, man, aged 33 y, LGBTQ+, relationship duration: 1-5 y, cohabiting, no children]
>
> *Because it is a pleasant interaction [...] I'm curious what today's question is. You know, if my partner's gone into it before me [...] it's like, ooh what did they answer? Or, I wonder what they will answer, let me put mine and that will nudge them to get their answer in.* [UK07, man, aged 36 y, heterosexual, relationship duration: >5 y, cohabiting, children]

Fun and Entertainment

Lighthearted questions helped make *Paired* enjoyable and facilitated engagement in relationship work:

> *I think [the questions are] fun and I think they're informative and I think the way they're done is pretty low pressure. Like I'm never scared, like "oh I have to put the right answer, it's scary." I think they're pretty low effort but still can lead to good conversations.* [US10, woman, aged 25 y, heterosexual, relationship duration: ≤1 y, living apart, no children]

As one interviewee explained, she and her partner enjoyed the quizzes the most:

> *...the questions sometimes are very informative. But when we have the quizzes it feels really easy to compare our views on things. So I asked my partner yesterday how do you actually feel about this app [...] he definitely said "oh quizzes, yeah I love quizzes, it's just so fun."* [UK01, woman, aged 18 y, heterosexual, relationship duration: ≤1 y, cohabiting, no children]

Another interviewee mentioned a Christmas-themed quiz, accessed at a time of year which was often "stressful" for the couple:

> *...it was nice to have something to talk about that wasn't stressful, and that was little. Like "what is a holiday tradition that you want to do with your kids?" [...] It's not tiny because it's important, but it's also not like "how are we going to get [partner's parent] to respect me with our child?" [...] it's a little more chill.* [US07, woman, aged 23 y, heterosexual, relationship duration: 1-5 y, living apart, no children]

While some interviewees sought out quizzes within *Paired* to complete with their partner for fun, others found the "fun" questions and quizzes too trivial:

XSL•FO
RenderX

> *Sometimes it's not serious enough but maybe we're just too serious.* [UK10, man, aged 27 y, LGBTQ+, relationship duration: 1-5 y, living apart, no children]

As such, there was a tension between this subtheme and the following subtheme.

Experiencing Meaningful Benefits and Wanting More

Experiencing short- and longer-term benefits was a key motivator for sustained *Paired* use. These benefits ranged from enjoying a moment of connection to looking back over past months and recognizing that relationship communication had improved (refer to the theme *Communication*). This created a virtuous cycle, where benefits attributed to the app led couples to continue using it.

Maintaining a "Streak" Can Be a Motivator and Signifier of Commitment

Paired informs users of their "streak" (number of continuous days of answering daily questions). For some, the streak was unimportant, while for others it was a motivator to engage daily:

> *I've almost never found myself not in the mood to answer one of these questions [...] because I get a lot out of it, conversely there's time where my wife destroys her streak only because of the fact that she's just like, either she was too busy or whatever, "I just couldn't be bothered doing it today." Where I'm just like, this is fun!* [UK07, man, aged 36 y, heterosexual, relationship duration: >5 y, cohabiting, children]

Partners can see how long each other's streak is, which motivated some but was off-putting for others: a partner's lower streak could be perceived as signifying lower commitment to the relationship:

> *God forbid I miss my streak and have to start over, that's my personality. But when I look at the app and I see that I'm on a streak of 27 days and he's on 2, it gets to me [...] feeling like he's not as committed to working on our relationship. I think maybe [the streak's] good for some people and it's reassuring and it's motivating, but for us it just causes more problems.* [US06, woman, aged 44 y, LGBTQ+, relationship duration: 1-5 y, cohabiting, children]

Dailiness: Paired Use and Regular Relationship Communication Can Become Embedded in Couples' Daily Lives

App Use Can Become Habitual

Through regular engagement, *Paired* use could become a pleasant habit, which was therefore sustained:

> *...it's just become a habit now [...] automatically just do it at a certain time of the day. Knowing [...] this is what we'll be discussing shortly. I think we both kind of look forward to having stuff to talk about that is not work, not kids, not finances.* [UK05, man, aged 45 y, heterosexual, relationship duration: >5 y, cohabiting, children]

Regular Relationship Communication Can Become Habitual

Regular relationship communication prompted by the app can become part of couples' intimate lives because it "prompts conversations [...] forces you to talk about something proper every day" (UK01, woman, aged 18 y, heterosexual, relationship duration: ≤1 y, living apart, no children). Interviewees described how they tended to respond to the questions and discuss them at particular times in their daily routines:

> *I usually come out into the living room because I wake up earlier most days [...] And when he answers his Paired question is how I know he's awake.* [US01, man, aged 34 y, LGBTQ+, relationship duration: ≤1 y, cohabiting, no children]

This became regular time devoted to the relationship:

> *Initially we would talk about Paired in the morning but probably I'm already in work mode. Or we'd talk about Paired at lunchtime and eventually we realized it was probably best to talk about it just before bed. When we were together, relaxed, just before we start bingeing on our box sets.* [UK05, man, aged 45 y, heterosexual, relationship duration: >5 y, cohabiting, children]

Discussions occurred in person, by telephone, SMS text message, or on the web (eg, Snapchat, Instagram, or WhatsApp). Users came to expect and look forward to daily relationship communication:

> *[Paired] creates these little windows and pockets of time for us to talk about something that has a benefit or is something silly or makes us laugh or something that just reinforces that it's difficult, but it's not really that difficult.* [UK08, man, aged 44 y, heterosexual, relationship duration: >5 y, cohabiting, children]

"Dosage": Regular Use of Paired Delivers Incremental Benefits

The theme *Communication* described how *Paired* can cumulatively improve relationship quality through prompting meaningful conversations and providing topics, enabling couples to regularly practice communication and enhance their emotional connection. In the themes *Engagement* and *Dailiness*, we described how the use of *Paired* can be sustained and become embedded in couples' daily lives. Supporting the quantitative findings, interviewees noticed gradual, incremental gains with regular, frequent use of the app:

> *...I'm struggling to find something that's major and has left an imprint. But the bottom line is all those interactions have had a positive impact.* [UK07, man, aged 36 y, heterosexual, relationship duration: >5 y, cohabiting, children]

The survey findings showing that the amount of time (per week) spent on the app was relatively unimportant may be explained by the qualitative findings that *Paired* facilitates relationship maintenance behaviors that may occur offline:

> *...it's a little time spent [on the app] for a lot of love gained.* [UK10, man, aged 27 y, LGBTQ+, relationship duration: 1-5 y, living apart, no children]

XSL•FO
RenderX

Discussion

Principal Findings

The regular use of an mHealth intervention, specifically *Paired*, seems to improve relationship quality over a relatively short time frame. The MQoRS score was 35.5% higher among people reporting >3 months' use of *Paired* compared to new users (≤1 wk), with longitudinal data showing similar positive trends. Regular, daily use benefited relationships the most and did not require intensive use of the app. The interview data suggest that this may be due to how the app prompts enjoyable and meaningful conversations, and these interactions help couples develop and practice relationship maintenance skills and feel more emotionally connected. Using *Paired*, and the habits it engenders, can become embedded in couples' daily lives as they look forward to their partner's response to the daily questions and make time for regular relationship communication. Noticing the app's impact helps to sustain its regular use in a positive feedback loop (virtuous cycle).

Strengths and Limitations

Our study was relatively small scale (eg, the web-based survey sample included just 13 new users of the app [≤1 wk use], which is something we could not have anticipated). Data collection occurred over 3 months when there were no major changes to the app and among people who had chosen to use it. This self-selected sample may be more open to positive relationship care and more digitally literate than the general population, which may limit the transferability of our findings; however, our international sample is a strength. Regular users may be overrepresented among participants because they were more likely to see the research invitations and more likely to provide complete longitudinal brief survey data. These factors and our sampling methods preclude generalization to all *Paired* users but are unproblematic to the study's aim of exploring the *potential* of a stand-alone mHealth intervention in relationship care.

Our study design is appropriate for this early stage in the app's evaluation. As well as providing proof of concept, we developed a theory about how *Paired* works in context. This work could inform a future randomized controlled trial to quantify the app's effectiveness compared to an alternative (or no) intervention [50]. Future evaluative work will need to take into account the app's widespread availability and assess its effectiveness during its iterative, ongoing development [53]. Proceeding straight to a trial would have been unnecessary (and potentially wasteful) without the indicative evidence of potential effectiveness, or proof of concept [52,67], that we have now provided.

Our use of mixed methods design is appropriate to our aim of developing an in-depth, contextualized understanding of whether and how a complex digital intervention can "work" in practice. The integration of data from complementary sources increases the validity of the main finding that *Paired* could improve relationship quality. In explanation, analysis of the cross-sectional web-based survey demonstrates that relationship quality (measured by the MQoRS and by responses to direct questions about the app's impact) is positively associated with the reported duration of *Paired* use. This finding could represent a positive association between relationship quality and duration of use or higher disengagement from the app among people in poor-quality relationships. The latter explanation can be discounted as unlikely because longitudinal (brief survey) data show improvements in various aspects of relationship quality in a cohort of users over 3 months. The interviewees' descriptions of how *Paired* helped them demonstrate a perceived causal association: use of the app contributes to improved relationship quality. It is possible that analyses based on subgroups of the web-based survey data may be underpowered, although the *P* values (and CIs) obtained suggest that this is unlikely. The broad agreement between the findings obtained from multiple data sources in this mixed methods study provides further reassurance.

We used a multidimensional measure of relationship quality, the MQoRS, which was developed based on theory and evidence derived from the *Enduring Love?* study [43,68,69] using robust statistical analyses to assess its validity and reliability. It overcomes the shortcomings of existing unidimensional scales (acknowledging that relationship quality is a complex construct) and is suited to mHealth research (information received from Di Martino et al [email, January 16, 2025]). The in-app survey was brief to increase the likelihood of repeat completions and thus obtain longitudinal data, which precluded use of the 17-item MQoRS; yet, encouragingly, the results support the same main finding.

The surveys used convenience sampling; the brief survey lacked demographic data; and, during our study, the app's data on its users' demographics were very incomplete. Therefore, we could not check the survey samples' representativeness to users overall, but we made such comparisons as were possible. These were favorable: the web-based survey participants seemed broadly representative of the active users of *Paired* (Multimedia Appendix 4). Although our sample is international, we could not explore use or effectiveness by race, ethnicity, religion, or socioeconomic status because we did not collect these data, for reasons explained in the Methods section. Both survey datasets may include nonindependent data if both partners in a couple participated; however, we cannot account for this, which is a limitation (matching data within anonymous surveys poses feasibility, acceptability, ethical, and privacy issues because asking for participants' own and their partners' names renders the data identifiable). The web-based survey data on the duration and frequency of engagement with *Paired* are self-reported and therefore could be subject to recall bias. App-collected metrics would be preferable but were not obtained (this would have raised ethical and data governance issues). However, the longitudinal brief survey data effectively provide the approximate duration of *Paired* use because these surveys were repeated over time in a cohort of app users. Triangulation and integration of data from multiple sources strengthens the conclusions we are able to draw regarding the "dosage" of use of the app.

Meaning and Implications

Digital Relationship Care

The findings provide proof of concept of mHealth for relationship care for supporting modest improvements in

XSL•FO
RenderX

relationship quality, over a relatively short time frame, in a self-selected population of users of 1 app. In the context of limited research on fully digital couple interventions for relationship care [45] (ie, positive, preventive interventions that are not blended or therapist guided), our rigorous evaluation makes an important contribution to the evidence base. The findings fit with existing evidence and theory about the importance of regular daily "relationship work" in sustaining couple relationships [43] and demonstrate how this can be supported in practice. The interview accounts of how benefits can accrue from regular, positive interactions and moments of connection evoke the concept of "relationship banking" [70] and demonstrate how mHealth can prompt these interactions. We would not expect large changes in relationship quality, given the many contextual influences on relationships, the time frame, and the fact that the study population included users whose relationship quality was already good (among whom scope for improvement is limited).

Inclusivity and Accessibility

Similar perceptions of the app's effectiveness among people of different ages, sexualities, and relationship types and durations suggest that it is inclusive by these characteristics. Therefore, its appeal is likely to be broad, although ethnic and cultural inclusivity has not been explored. Evidence of the effectiveness of *Paired* is promising in the context of underuse of existing preventive interventions. The convenience and flexibility of mHealth suits the ebbs and flows of couple relationships because partners have control over when and how much they engage and apparently benefit without needing to use the app intensively. Barriers to uptake are low for digitally literate web-based populations, and digital interventions are scalable (eg, the work of creating new questions and content for *Paired* is the same, irrespective of the number of users), contrasting with face-to-face or "blended" courses or therapy, which require conjoint commitment and a professional's input with each couple or group.

Theory and Practice

Where mHealth interventions incorporate tailoring to individual users' characteristics, this can increase the personal relevance of content and messaging, which in turn can increase engagement with the intervention and improve learning [71]. Blended interventions may achieve similar effects through personalized messaging from a coach to each individual user. We have shown that an intervention without these features can be personally relevant because partners effectively create content for each other. They "receive" an intervention that has a unique human touch (their partner's) that is not only personalized but intimately personal. Peer-to-peer digital communication ordinarily requires resource-intensive moderation, but this does not apply to *Paired* because communication is between partners only.

Theories of behavior change tend to be individualistic, although some acknowledge the social dimensions of learning [31,72]. The refined theory of change that we developed is novel in encompassing the dyadic nature of behavior change within relationships, and in emphasizing how human-technology interactions and between-partner interactions support this change and sustain engagement. *Paired* engagement is sustained directly (eg, in-app messages) and indirectly (partners remind each other to answer questions; they experience relationship improvements and are motivated to continue to engage).

A combination of fun and meaningfulness seemed to help maintain engagement, and the app prompted regular daily "relationship work." These findings may inform self-help and professionally delivered courses and therapy, both digital and face-to-face, including future iterations of *Paired*.

Digital Intimacy

The use of apps to score and rate users' sexual lives has been described by Lupton [73] as reinforcing reductive and normative perspectives on what is "good" within an area of intimate life. Levels of agreement in *Paired* quizzes are undoubtedly reductive assessments of how "good" a relationship is, but in contrast to the apps reviewed by Lupton [73], these comparisons are not designed for ranking or sharing outside of the couple. Instead, they stimulate within-couple discussion and interpretation. Daily questions elicit free-text responses, eluding quantification, but "streaks" are comparable between partners. The work conducted by Lupton [55,74] explains how mHealth apps become more-than-human individual-app assemblages, with their own agency. We suggest, based on our interviews, that *Paired* creates an individual-partner-app assemblage, functioning within and as part of couples' intimate lives [75]; for example, daily question completion indicates to one partner that the other is thinking of the relationship, both documenting and becoming part of couples' relationship work; differences in streak length may signify lower commitment to the relationship. App-based indicators of conjoint accountability may cause difficulties for couples where problematic relationship dynamics already exist [42]. In a study of young people's communication regarding their Snapchat streaks, Hristova et al [76] found that not maintaining a streak could have significance for peer relationships similar to what we found for couples (unlike *Paired* streaks, which are individualized, a Snapchat streak is a continuous period of at least daily "snaps" exchanged between a pair of users).

Transferability and Relevance

We evaluated 1 app, within the first 3 months of its availability, as an exemplar to explore whether mHealth can benefit relationships. Choosing *Paired* was fortuitous, given the rapid turnover of apps, most of which fail (an estimated 99.5% [77]). *Paired* became the leading relationships app worldwide in terms of revenue and downloads in 2021 and 2022 [78]. It expanded from 10,000 active users during our study to >1 million monthly active users in November 2024 (information received from *Paired* [personal communication, December 10, 2024]). Therefore, the findings inform an app with considerable reach. *Paired* became more sophisticated since data collection, but our findings remain relevant because core features remain, including the daily questions that were so important. The app's impact may be affected by how new technologies' user populations tend to be differently constituted over time [79]. Our findings should not be generalized uncritically to apps with different functionality, features, and content. We make no claim that our findings are generalizable to all couples or all users of

XSL•FO
RenderX

Paired. We have not explored what happens to relationship quality beyond 3 months' use of *Paired* or whether changes are sustained if couples stop using it.

Pandemic Context

This research took place during the COVID-19 pandemic, and couples may have turned to *Paired* during relationship strain; however, the data are not suggestive of severe relationship problems among participants. Media predictions of large increases in relationship breakdown and divorce during the pandemic did not occur [15,80]. Instead, impacts on relationships were heterogeneous—and positive for some couples [17,81]—although long-term impacts remain unknown [82,83]. However, couples' access to formal and informal relationship support was reduced or interrupted. Some people invested in personal growth and relationships during lockdowns and quarantines [84], often supported by digital resources. Couples may use *Paired* differently now, perhaps preferring to engage with relationship care through other means or not at all. Nevertheless, if the digital engagement with services, social lives, and entertainment that accelerated during COVID-19 is here to stay, interest in digital relationships care may persist—as *Paired* subscription patterns suggest.

Future Directions

Paired may complement existing relationship self-help and support interventions, for example, for couples who are unready or waiting for therapy, as blended care alongside conventionally delivered relationship therapy, or during "offboarding" (after the cessation of therapy). Future work could explore reasons for disengagement with *Paired*, which could be interpreted positively (it has done its job), negatively (it is not working), or neutrally (it would be surprising if any intervention suited everyone). Digital solutions are not a panacea due to inequalities in technology access and digital skills [85]. Further evaluative research could take a theory-based and systems perspective [67] regarding how digital interventions fit within the relationship care and support ecosystem and extending our findings to delineate optimal contexts of use, that is, for which couples, in which circumstances, and how *Paired* may best enhance couple relationships. Data from couple dyads and longitudinal qualitative data would enable an exploration of the influences, effects, and contexts of different patterns of use (eg, daily, weekly, or episodic). The meaning of *Paired* use within relationships has been explored in greater depth in a separate paper [75].

Conclusions

Paired has the potential to improve relationship quality over a relatively short time frame. It does this by being an engaging mHealth intervention that prompts regular daily relationship communication, supporting the development of relationship skills and increasing feelings of emotional connection.

Acknowledgments

This work was funded by The Open University, with in-kind contributions from the University of Brighton and Better Half/Paired. It builds on the *Enduring Love?* study (Economic and Social Research Council; ESRC RES-062-23-3056). The authors are grateful to their participants, funders, and match funders. Data collection was carried out independently from Better Half/Paired and the funders, with the exception of the brief in-app survey (hosted on *Paired*). Better Half/Paired was not involved in the analysis, interpretation, writing, or decision to submit this manuscript.

Conflicts of Interest

JG completed a secondment as Chief Relationships Officer at Paired. She retained her academic and professional independence throughout this public engagement role, which was unremunerated aside from her university salary. Before the study reported here, Better Half/Paired provided a modest amount of funding to The Open University for a small-scale survey. CA worked for The Open University on a consultancy basis (10 d) on this project, under the supervision of JG and ML. All other authors declare no conflicts of interest. Furthermore, during and since this study, no authors have received funding indirectly or directly from Paired, and no authors stand to benefit or suffer financially from the success or otherwise of the Paired app.

Multimedia Appendix 1
Data collection materials.
[DOCX File , 1175 KB - mhealth_v13i1e55433_app1.docx]

Multimedia Appendix 2
Multidimensional Quality of Relationship Scale development, content, and analyses.
[DOCX File , 27 KB - mhealth_v13i1e55433_app2.docx]

Multimedia Appendix 3
Additional statistical analyses.
[DOCX File , 31 KB - mhealth_v13i1e55433_app3.docx]

Multimedia Appendix 4

XSL•FO
RenderX

Comparison of web-based survey participants and Paired users.
[DOCX File , 24 KB - mhealth_v13i1e55433_app4.docx]

References

1. Cordova JV, Fleming CJ, Morrill MI, Hawrilenko M, Sollenberger JW, Harp AG, et al. The Marriage Checkup: a randomized controlled trial of annual relationship health checkups. J Consult Clin Psychol 2014 Aug;82(4):592-604 [FREE Full text] [doi: 10.1037/a0037097] [Medline: 24932565]
2. Proulx CM, Helms MH, Buehler C. Marital quality and personal well‐being: a meta‐analysis. J Marriage Fam 2007 Jul 05;69(3):576-593 [FREE Full text] [doi: 10.1111/j.1741-3737.2007.00393.x]
3. Carr D, Cornman J, Freedman V. Marital quality and negative experienced well-being: an assessment of actor and partner effects among older married persons. J Gerontol B Psychol Sci Soc Sci 2016 Jan;71(1):177-187 [FREE Full text] [doi: 10.1093/geronb/gbv073] [Medline: 26329115]
4. Leach LS, Butterworth P, Olesen SC, Mackinnon A. Relationship quality and levels of depression and anxiety in a large population-based survey. Soc Psychiatry Psychiatr Epidemiol 2013 Mar;48(3):417-425 [FREE Full text] [doi: 10.1007/s00127-012-0559-9] [Medline: 22875222]
5. Whisman MA. Marital distress and DSM-IV psychiatric disorders in a population-based national survey. J Abnorm Psychol 2007 Aug;116(3):638-643 [FREE Full text] [doi: 10.1037/0021-843X.116.3.638] [Medline: 17696721]
6. Robles TF, Slatcher RB, Trombello JM, McGinn MM. Marital quality and health: a meta-analytic review. Psychol Bull 2014 Jan;140(1):140-187 [FREE Full text] [doi: 10.1037/a0031859] [Medline: 23527470]
7. Kiecolt-Glaser JK, Wilson SJ. Lovesick: how couples' relationships influence health. Annu Rev Clin Psychol 2017 May 08;13:421-443 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-045111] [Medline: 28301763]
8. Garriga A, Kiernan K. Parents' relationship quality, mother-child relations and children's behaviour problems: evidence from the UK Millennium Cohort Study. University of York. URL: https://www.york.ac.uk/media/spsw/documents/research-and-publications/Garriga-KiernanWP2014.pdf [accessed 2024-04-29]
9. Jekielek SM. Parental conflict, marital disruption and children's emotional well-being. Social Forces 1998 Mar 01;76(3):905-936 [FREE Full text] [doi: 10.1093/sf/76.3.905]
10. Vandewater EA, Lansford JE. Influences of family structure and parental conflict on children's well-being. Fam Relat 1998 Oct;47(4):323-339 [FREE Full text] [doi: 10.2307/585263]
11. Amato PR. Children of divorce in the 1990s: an update of the Amato and Keith (1991) meta-analysis. J Fam Psychol 2001 Sep;15(3):355-370 [FREE Full text] [doi: 10.1037//0893-3200.15.3.355] [Medline: 11584788]
12. Amato PR, Keith B. Parental divorce and the well-being of children: a meta-analysis. Psychological Bulletin 1991;110(1):26-46 [FREE Full text] [doi: 10.1037//0033-2909.110.1.26]
13. World marriage data 2019 (POP/DB/Marr/Rev2019). United Nations. URL: https://www.un.org/development/desa/pd/data/world-marriage-data [accessed 2024-04-29]
14. Divorce rate by state. World Population Review. URL: https://worldpopulationreview.com/state-rankings/divorce-rate-by-state [accessed 2024-04-29]
15. Divorces in England and Wales: 2020. Office for National Statistics (ONS). URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/divorce/bulletins/divorcesinenglandandwales/previousReleases [accessed 2024-04-29]
16. Sserwanja I, Marjoribanks D, Relationship DM. Relationship distress monitor: estimating levels of adult couple distress across the UK. NACCC. 2016. URL: https://naccc.org.uk/wp-content/uploads/2020/06/Relationship_Distress_Monitor.pdf [accessed 2024-04-29]
17. Gabb J, Aicken C. Summary report: COVID-19 relationships survey. Milton Keynes: Open University. URL: http://www.open.ac.uk/researchprojects/enduringlove/sites/www.open.ac.uk.researchprojects.enduringlove/files/files/Summary%20OU-Paired%20Report.pdf [accessed 2024-04-29]
18. Lebow JL, Chambers AL, Christensen A, Johnson SM. Research on the treatment of couple distress. J Marital Fam Ther 2012 Jan;38(1):145-168 [FREE Full text] [doi: 10.1111/j.1752-0606.2011.00249.x] [Medline: 22283385]
19. Shadish WR, Baldwin SA. Meta-analysis of MFT interventions. J Marital Fam Ther 2003 Oct;29(4):547-570 [FREE Full text] [doi: 10.1111/j.1752-0606.2003.tb01694.x] [Medline: 14593694]
20. Meier R. Relationship education programmes for adults - an overview of research. What Works in Relationship Support: An Evidence Review.: Tavistock Centre For Couple Relationships; 2015. URL: https://www.healthymarriageinfo.org/wp-content/uploads/2017/12/20150608EvidenceReview2015pdf-min.pdf [accessed 2024-04-29]
21. Stewart JW, Bradford K, Higginbotham BJ, Skogrand L. Relationship help-seeking: a review of the efficacy and reach. Marriage Fam Rev 2016 Apr 25;52(8):781-803 [FREE Full text] [doi: 10.1080/01494929.2016.1157559]
22. Parnell KJ, Scheel MJ, Davis CK, Black WW. An investigation of couples' help-seeking: a multiple case study. Contemp Fam Ther 2017 Sep 5;40(1):110-117 [FREE Full text] [doi: 10.1007/S10591-017-9427-9]
23. Williamson HC, Karney BR, Bradbury TN. Barriers and facilitators of relationship help-seeking among low-income couples. J Fam Psychol 2019 Mar;33(2):234-239 [FREE Full text] [doi: 10.1037/fam0000485] [Medline: 30489129]

XSL•FO
RenderX

24. Markman HJ, Hawkins AJ, Stanley SM, Halford WK, Rhoades G. Helping couples achieve relationship success: a decade of progress in couple relationship education research and practice, 2010-2019. J Marital Fam Ther 2022 Jan;48(1):251-282 [FREE Full text] [doi: 10.1111/jmft.12565] [Medline: 34783038]
25. Bailey J, Mann S, Wayal S, Abraham C, Murray E. Digital media interventions for sexual health promotion-opportunities and challenges: a great way to reach people, particularly those at increased risk of sexual ill health. BMJ 2015 Mar 03;350(mar03 8):h1099 [FREE Full text] [doi: 10.1136/bmj.h1099] [Medline: 25736806]
26. Cabecinha M, Mercer CH, Gravningen K, Aicken C, Jones KG, Tanton C, et al. Finding sexual partners online: prevalence and associations with sexual behaviour, STI diagnoses and other sexual health outcomes in the British population. Sex Transm Infect 2017 Dec;93(8):572-582 [FREE Full text] [doi: 10.1136/sextrans-2016-052994] [Medline: 28396557]
27. Watchirs Smith L, Guy R, Degenhardt L, Yeung A, Rissel C, Richters J, et al. Meeting sexual partners through internet sites and smartphone apps in Australia: national representative study. J Med Internet Res 2018 Dec 18;20(12):e10683 [FREE Full text] [doi: 10.2196/10683] [Medline: 30563809]
28. Deogan C, Jacobsson E, Mannheimer L, Björkenstam C. Meeting sexual partners online and associations with sexual risk behaviors in the Swedish population. J Sex Med 2020 Nov;17(11):2141-2147 [FREE Full text] [doi: 10.1016/j.jsxm.2020.08.001] [Medline: 32873533]
29. Anderson M, Vogels EA, Turner E. The virtues and downsides of online dating. Pew Research Center. URL: https://www.pewresearch.org/internet/2020/02/06/the-virtues-and-downsides-of-online-dating/ [accessed 2024-04-29]
30. Eichenberg C, Huss J, Küsel C. From online dating to online divorce: an overview of couple and family relationships shaped through digital media. Contemp Fam Ther 2017 Nov 2;39(4):249-260 [FREE Full text] [doi: 10.1007/S10591-017-9434-X]
31. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649 [FREE Full text] [doi: 10.1080/08870449808407422]
32. Danaher J, Nyholm S, Earp B. The quantified relationship. Am J Bioeth 2018 Feb;18(2):3-19 [FREE Full text] [doi: 10.1080/15265161.2017.1409823] [Medline: 29393796]
33. Burns K, Keating P, Free C. A systematic review of randomised control trials of sexual health interventions delivered by mobile technologies. BMC Public Health 2016 Aug 12;16(1):778 [FREE Full text] [doi: 10.1186/s12889-016-3408-z] [Medline: 27514851]
34. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb 14;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
35. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342 [FREE Full text] [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
36. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
37. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
38. Watson I. Unmasking uninstalls: three data points to think about. Adjust GmbH. 2018. URL: https://www.adjust.com/blog/unmasking-uninstalls/ [accessed 2024-04-29]
39. Innovation, evidence, regulation and adoption. IQVIA Institute. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021 [accessed 2024-04-29]
40. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
41. Gibbs J, Gkatzidou V, Tickle L, Manning SR, Tilakkumar T, Hone K, et al. 'Can you recommend any good STI apps?' A review of content, accuracy and comprehensiveness of current mobile medical applications for STIs and related genital infections. Sex Transm Infect 2017 Jun;93(4):234-235 [FREE Full text] [doi: 10.1136/sextrans-2016-052690] [Medline: 27884965]
42. Knopp K, Schnitzer JS, Khalifian C, Grubbs K, Morland LA, Depp C. Digital interventions for couples: state of the field and future directions. Couple Fam Psychol Res Pract 2023 Dec;12(4):201-217. [doi: 10.1037/cfp0000213]
43. Gabb J, Fink J. Couple Relationships in the 21st Century: Research, Policy, Practice. Cham, Switzerland: Springer; 2015.
44. Morgan DH. Family Connections: An Introduction to Family Studies. Cambridge, MA: Polity Press; 1996.
45. Jamieson L. Intimacy: Personal Relationships in Modern Societies. Cambridge, MA: Polity Press; 1998.
46. McNamee P, Murray E, Kelly MP, Bojke L, Chilcott J, Fischer A, et al. Designing and undertaking a health economics study of digital health interventions. Am J Prev Med 2016 Nov;51(5):852-860 [FREE Full text] [doi: 10.1016/j.amepre.2016.05.007] [Medline: 27745685]
47. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. BMJ 2007 Mar 01;334(7591):455-459 [FREE Full text] [doi: 10.1136/bmj.39108.379965.be]

XSL•FO
RenderX

48. Oakley A, Strange V, Bonell C, Allen E, Stephenson J, RIPPLE Study Team. Process evaluation in randomised controlled trials of complex interventions. BMJ 2006 Feb 18;332(7538):413-416 [FREE Full text] [doi: 10.1136/bmj.332.7538.413] [Medline: 16484270]
49. Hawe P, Shiell A, Riley T. Theorising interventions as events in systems. Am J Community Psychol 2009 Jun;43(3-4):267-276 [FREE Full text] [doi: 10.1007/s10464-009-9229-9] [Medline: 19390961]
50. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
51. Pawson R, Tilley N. Realistic Evaluation. Thousand Oaks, CA: Sage Publications; 2019.
52. Drabble SJ, O'Cathain A. Moving from randomized controlled trials to mixed methods intervention evaluations. In: Hesse-Biber SN, Johnson RB, editors. The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry. Oxford, UK: Oxford Academic Press; 2015:406-425.
53. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
54. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
55. Lupton D. Toward a more-than-human analysis of digital health: inspirations from feminist new materialism. Qual Health Res 2019 Dec;29(14):1998-2009 [FREE Full text] [doi: 10.1177/1049732319833368] [Medline: 30964392]
56. Lilford RJ, Foster J, Pringle M. Evaluating eHealth: how to make evaluation more methodologically robust. PLoS Med 2009 Nov;6(11):e1000186 [FREE Full text] [doi: 10.1371/journal.pmed.1000186] [Medline: 19956674]
57. Bazeley P. Integrating Analyses in Mixed Methods Research. Thousand Oaks, CA: Sage Publications; 2018.
58. Burke Johnson R, Russo F, Schoonenboom J. Causation in mixed methods research: the meeting of philosophy, science, and practice. J Mix Methods Res 2017 Jul 25;13(2):143-162 [FREE Full text] [doi: 10.1177/1558689817719610]
59. Gabb J, Aicken C, Di Martino S, Witney T, Lucassen M. More–than–relationship quality: a feminist new materialist analysis of relationship quality and the potential of digital couple interventions. J Fam Theory Rev 2023 May 03;15(4):685-705 [FREE Full text] [doi: 10.1111/jftr.12509]
60. Fletcher GJ, Simpson JA, Thomas G. The measurement of perceived relationship quality components: a confirmatory factor analytic approach. Pers Soc Psychol Bull 2000 Mar 01;26(3):340-354. [doi: 10.1177/0146167200265007]
61. Morry MM, Reich T, Kito M. How do I see you relative to myself? Relationship quality as a predictor of self- and partner-enhancement within cross-sex friendships, dating relationships, and marriages. J Soc Psychol 2010;150(4):369-392 [FREE Full text] [doi: 10.1080/00224540903365471] [Medline: 20718222]
62. Kanter JB, Lavner JA, Lannin DG, Hilgard J, Monk JK. Does couple communication predict later relationship quality and dissolution? A meta-analysis. J Marriage Fam 2021 Oct 30;84(2):533-551 [FREE Full text] [doi: 10.1111/jomf.12804]
63. Bulanda JR, Brown SL. Race-ethnic differences in marital quality and divorce. Soc Sci Res 2007 Sep;36(3):945-967 [FREE Full text] [doi: 10.1016/j.ssresearch.2006.04.001]
64. Conger RD, Conger KJ, Martin MJ. Socioeconomic status, family processes, and individual development. J Marriage Fam 2010 Jun;72(3):685-704 [FREE Full text] [doi: 10.1111/j.1741-3737.2010.00725.x] [Medline: 20676350]
65. Ellison CG, Burdette AM, Bradford Wilcox WB. The couple that prays together: race and ethnicity, religion, and relationship quality among working‐age adults. J Marriage Fam 2010 Jul 09;72(4):963-975 [FREE Full text] [doi: 10.1111/j.1741-3737.2010.00742.x]
66. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
67. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
68. Chonody JM, Gabb J, Killian M, Dunk-West P. Measuring relationship quality in an international study: exploratory and confirmatory factor validity. Res Soc Work Pract 2018 Nov;28(8):920-930 [FREE Full text] [doi: 10.1177/1049731516631120] [Medline: 30369776]
69. Chonody J, Gabb J. Understanding the role of relationship maintenance in enduring couple partnerships in later adulthood. Marriage Fam Rev 2018 Apr 27;55(3):216-238 [FREE Full text] [doi: 10.1080/01494929.2018.1458010]
70. Gottman J, Gottman J. The natural principles of love. J Fam Theory Rev 2017 Mar 02;9(1):7-26 [FREE Full text] [doi: 10.1111/jftr.12182]
71. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693 [FREE Full text] [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
72. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991:179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-t]

XSL•FO
RenderX

73. Lupton D. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. Cult Health Sex 2015;17(4):440-453 [FREE Full text] [doi: 10.1080/13691058.2014.920528] [Medline: 24917459]
74. Lupton D. The thing-power of the human-app health assemblage: thinking with vital materialism. Soc Theory Health 2019 Feb 22;17(2):125-139 [FREE Full text] [doi: 10.1057/s41285-019-00096-y]
75. Witney T, Gabb J, Aicken C, Di Martino S, Lucassen M. Configuring the digital relationship landscape: a feminist new materialist analysis of a couple relationship app. Fam Relatsh Soc 2021;13(2):181-197 [FREE Full text] [doi: 10.1332/20467435Y2024D000000015]
76. Hristova D, Jovicic S, Göbl B, de Freitas S, Slunecko T. "Why did we lose our snapchat streak?". Social media gamification and metacommunication. Comput Human Behav Rep 2022 Mar;5:100172 [FREE Full text] [doi: 10.1016/j.chbr.2022.100172]
77. Steiner R. Why 99.5% of consumer apps fail (and how to keep yours alive). Fyresite. URL: https://www.fyresite.com/how-many-apps-fail/ [accessed 2024-04-29]
78. Reports. Sensor Tower. 2022. URL: https://sensortower.com/ [accessed 2024-04-29]
79. Rogers EM. Diffusion of Innovations. 5th edition. New York, NY: Free Press; 2003.
80. Marriages and divorces. Centers for Disease Control (CDC). URL: https://www.cdc.gov/nchs/nvss/marriage-divorce.htm [accessed 2024-04-29]
81. Mitchell KR, Shimonovich M, Bosó Pérez R, Dema E, Clifton S, Riddell J, et al. Initial Impacts of COVID-19 on sex life and relationship quality in steady relationships in Britain: findings from a large, quasi-representative survey (Natsal-COVID). J Sex Res 2023 Jan;60(1):1-12 [FREE Full text] [doi: 10.1080/00224499.2022.2035663] [Medline: 35286182]
82. Fancourt D, Bu F, Mak HW, Steptoe A. COVID-19 social study: results release 16. University College London.: UCL URL: https://www.covidsocialstudy.org/_files/ugd/3d9db5_dc64263647624fd3842e6521c186aa69.pdf [accessed 2024-04-29]
83. Pietromonaco PR, Overall NC. Applying relationship science to evaluate how the COVID-19 pandemic may impact couples' relationships. Am Psychol 2021 Apr;76(3):438-450 [FREE Full text] [doi: 10.1037/amp0000714] [Medline: 32700937]
84. Kim JJ, Munroe M, Feng Z, Morris S, Al-Refae M, Antonacci R, et al. Personal growth and well-being in the time of COVID: an exploratory mixed-methods analysis. Front Psychol 2021;12:648060 [FREE Full text] [doi: 10.3389/fpsyg.2021.648060] [Medline: 33841285]
85. McClain C, Vogels EA, Perrin A, Rainie L. The internet and the pandemic. Pew Research Center. URL: https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/ [accessed 2024-04-29]

Abbreviations

LGBTQ+: lesbian, gay, bisexual, transgender/trans, queer, and other sexuality and gender minoritized individuals
mHealth: mobile health
MQoRS: Multidimensional Quality of Relationship Scale
OR: odds ratio

Edited by M Sobolev; submitted 13.12.23; peer-reviewed by M Brinberg, S Locke; comments to author 16.07.24; revised version received 10.12.24; accepted 23.01.25; published 14.04.25.

Please cite as:
Aicken C, Gabb J, Di Martino S, Witney T, Lucassen M
Exploring the Potential of a Digital Intervention to Enhance Couple Relationships (the Paired App): Mixed Methods Evaluation
JMIR Mhealth Uhealth 2025;13:e55433
URL: https://mhealth.jmir.org/2025/1/e55433
doi:10.2196/55433
PMID:

©Catherine Aicken, Jacqui Gabb, Salvatore Di Martino, Tom Witney, Mathijs Lucassen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessment of an App-Based Sleep Program to Improve Sleep Outcomes in a Clinical Insomnia Population: Randomized Controlled Trial

Walter Staiano[1,2], PhD; Christine Callahan[3], PhD; Michelle Davis[3], PhD; Leah Tanner[3], PhD; Chelsea Coe[3], PhD; Sarah Kunkle[3], PhD; Ulrich Kirk[2,4], PhD

[1]Department of Physical Education and Sport, Universitat de València, Valencia, Spain

[2]Department of Psychology, University of Southern Denmark, Odense, Denmark

[3]Headspace, Santa Monica, CA, United States

[4]Fralin Biomedical Research Institute, Roanoke, VA, United States

Corresponding Author:
Ulrich Kirk, PhD
Department of Psychology
University of Southern Denmark
Campusvej 55
Odense, 5230
Denmark
Phone: 45 65502695
Email: ukirk@health.sdu.dk

Abstract

Background: Insomnia is the most commonly reported sleep disturbance and significantly impacts mental health and quality of life. Traditional treatments for insomnia include pharmacological interventions or cognitive behavioral therapy for insomnia (CBT-I), but these options may not be accessible to everyone who needs treatment.

Objective: This study aims to assess the effectiveness of the app-based Headspace Sleep Program in adults with clinical insomnia on sleep disturbance and mental health outcomes, compared with a waitlist control group.

Methods: This randomized controlled trial included 132 adults with clinical insomnia who were assigned to either the Headspace Sleep Program (an 18-session self-guided, in-app program utilizing CBT-I techniques augmented by mindfulness) or a waitlist control group. Sleep disturbance outcomes were assessed by changes in insomnia symptoms (measured using the Insomnia Severity Index) and sleep efficiency (measured via sleep diary and actigraphy). Mental health outcomes included perceived stress (measured by the 10-item Perceived Stress Scale), depressive symptoms (measured by the 8-item Patient Health Questionnaire), sleep quality (measured by the Pittsburgh Sleep Quality Index), anxiety symptoms (measured by the 7-item Generalized Anxiety Disorder Scale), and mindfulness (measured by the Mindful Attention Awareness Scale). Changes from baseline to postintervention and follow-up were assessed for each outcome.

Results: Participants had a mean (SD) age of 37.2 (10.6) years, with 69 out of 132 (52.3%) identifying as female. Those randomized to the Headspace Sleep Program group experienced significantly greater improvements in insomnia symptoms from baseline to postintervention and follow-up compared with participants in the waitlist control group ($P<.001$, $\eta^2p=0.107$). Improvements from baseline to postintervention and follow-up were also observed in the Headspace Sleep Program group for sleep efficiency, as measured by both sleep diary ($P=.01$, $\eta^2p=.03$) and actigraphy outcomes ($P=.01$, $\eta^2p=.03$). Participants in the Headspace Sleep Program group showed greater improvements in insomnia remission rates (8/66, 12%, at postintervention and 9/66, 14%, at follow-up) and treatment response (11/66, 17%, at postintervention and 15/66, 23%, at follow-up) compared with the control group (remission rate 2/66, 3%, at postintervention and 0/66, 0%, at follow-up; treatment response 3/66, 5%, at postintervention and 1/66, 2%, at follow-up). The results suggest significant improvements in depressive symptoms ($P=.01$, $\eta^2p=.04$), anxiety symptoms ($P=.02$, $\eta^2p=.02$), and mindfulness ($P=.01$, $\eta^2p=.03$) in the Headspace Sleep Program group.

Conclusions: The Headspace Sleep Program is an effective intervention for improving sleep disturbances in adults with clinical insomnia.

Trial Registration: ClinicalTrials.gov NCT05872672; https://clinicaltrials.gov/ct2/show/NCT05872672

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e68665) doi:10.2196/68665

KEYWORDS

cognitive behavioral therapy for insomnia; mindfulness; randomized controlled trial; RCT; therapy; insomnia; behavioral; app based; app

Introduction

Sleep plays a vital role in human functioning and is essential for quality of life [1]. Sleep disorders affect the sleep quality of millions and may impact physical and mental health, increasing the risk of obesity, hypertension, heart attack, and stroke [2-8], as well as elevating stress, depression, and anxiety symptoms [9]. Insomnia, defined as difficulty initiating or maintaining sleep, is the most common sleep disorder, with 20% of adults reporting occasional insomnia symptoms and 10%-30% experiencing chronic insomnia [10,11]. In addition to its physical and mental health impacts, insomnia has significant societal and economic consequences, including reduced productivity, absenteeism, higher accident rates, increased health care costs, and an overall decline in quality of life [12]. Research suggests that insomnia imposes an annual economic burden of US $95.5 to US $107.5 billion [13].

Sleep disturbances, including insomnia, can be treated with pharmacological interventions; however, these treatments may have adverse side effects [14]. Consequently, behavioral and psychological interventions, such as cognitive behavioral therapy for insomnia (CBT-I) [15] and mindfulness [16,17], have gained popularity for treating sleep disturbances and insomnia, with a robust body of literature supporting their efficacy [15-18]. Specifically, CBT-I is the first-line treatment recommended by the American College of Physicians [19] and the European Insomnia Guidelines [20] for improving insomnia symptoms, with interventions demonstrating moderate to large and durable effects on sleep quality and sleep efficiency [21,22]. CBT-I, like traditional therapy, is most often delivered in person by a licensed mental health professional, a mode of treatment that may not be accessible to everyone in need. To improve accessibility, recent studies have explored delivering CBT-I interventions digitally, with results suggesting similar outcomes to in-person treatment [23-25]. Similarly, app-based mindfulness interventions, also offered digitally, have been shown to improve sleep quality and insomnia symptoms [16,26].

To address insomnia treatment gaps, particularly by improving accessibility (Multimedia Appendix 1), Headspace developed the Headspace Sleep Program—an 18-session, app-based, self-guided program that integrates CBT-I techniques with mindfulness. This program aims to provide individuals with insomnia an accessible way to change their relationship with sleep and adopt strategies to build and maintain sleep hygiene. While Headspace and other digital mental health companies [27-29] have developed sleep-specific content, to our knowledge, this is the first self-guided, app-based program to combine CBT-I techniques with mindfulness as a core component in a structured, 18-session sequential format. Previous research has explored mindfulness as an adjunct or alternative to CBT-I [30], with initial findings suggesting that combining the 2 interventions may enhance overall sleep outcomes [31]. However, this research still followed the traditional CBT-I structure (ie, 6 weekly sessions lasting 90-120 minutes) and was not tested in a randomized controlled trial (RCT) [31]. While mindfulness is often incorporated as part of the cognitive component of CBT-I [32], the Headspace Sleep Program positions mindfulness as a foundational approach, introducing it first as a tool for building awareness of sleep habits and patterns. This approach differs from traditional CBT-I frameworks, where mindfulness is typically introduced as a complementary cognitive strategy. Additionally, the Headspace Sleep Program integrates mindfulness-based perspectives and practices throughout the program to encourage ongoing reflection, self-compassion, and nonjudgmental awareness, further distinguishing its emphasis on mindfulness as a core rather than a supplementary component. By incorporating mindfulness as a core component, the Headspace Sleep Program aims to equip individuals with tools to address not only sleep issues but also broader concerns such as stress, depression, and anxiety symptoms, which are assessed in this study.

By integrating both CBT-I and mindfulness, the Headspace Sleep Program was designed to be more effective than mindfulness alone and more accessible than traditional CBT-I for individuals with clinical insomnia. Several features were incorporated to enhance accessibility: (1) the program is delivered via the Headspace app, offering a subscription at a lower cost than traditional CBT-I; (2) all sessions are self-guided, allowing users to access content at their convenience without relying on provider availability; (3) the intervention consists of 18 daily sessions, a significantly shorter time frame than traditional CBT-I, which is typically delivered weekly over several weeks; and (4) offering the program through the Headspace app, which may help reduce treatment stigma and increase awareness among individuals who may not realize that insomnia treatment is available to them (Multimedia Appendix 1).

This study aimed to evaluate the effectiveness of the Headspace Sleep Program in adults with clinical insomnia by assessing its impact on sleep disturbance and mental health outcomes, including perceived stress, depression, anxiety, and mindfulness. We hypothesized that participants in the Headspace Sleep Program group would show significant improvements in insomnia and mental health outcomes compared with the waitlist control group. Additionally, in an exploratory analysis, we sought to determine whether participants with mild baseline insomnia symptoms responded differently to the intervention compared with those with moderate baseline insomnia symptoms.

XSL•FO
RenderX

Methods

Study Design

This RCT, using a pre-post and 3-week follow-up design, evaluated the effectiveness of the Headspace Sleep Program on insomnia outcomes in 132 adults with clinical insomnia. Sample size determination was based on 2 parameters. First, the minimally important difference in insomnia severity (ie, treatment response), defined as an 8-point reduction on the Insomnia Severity Index (ISI) [10,16], with the sample size calculated using an SD of 6 points based on previous studies [33,34]. Second, a priori power calculations using G*Power (Heinrich-Heine-Universität Düsseldorf) indicate that, with a sample size of 108, our study had 80% power to detect significant (P<.05) between-within interaction effects (effect size f=0.12), corresponding to a small-to-medium effect size by ANOVA (up to 3-way) [35]. In both parameters, we accounted for a 20% attrition rate. Participants completed a 7-day baseline period in which they maintained their normal sleep patterns while wearing an actigraphy watch during sleep and completing a sleep diary the following morning. After the baseline period, participants completed the baseline sleep disturbance and mental health questionnaires. The randomization procedure was determined after study recruitment but before the study launch. Specifically, participants were randomly allocated to 1 of 2 groups: the Sleep Program or the waitlist control. The research team was not formally blinded to group allocation. Participants were not informed of their group assignment until after completing the baseline assessment. Those in the Headspace Sleep Program were instructed to follow the 18-session program daily, while those in the waitlist control group were asked to maintain their regular routines and were granted access to the Sleep Program after the study.

After the intervention, participants were reassessed using the sleep disturbance and mental health questionnaires before completing a 7-day postintervention period. During this period, as in the baseline phase, they maintained their normal sleep patterns while wearing an actigraphy watch during sleep and completing a sleep diary the following morning. Participants were instructed not to access the app-based Sleep Program from the time of the postintervention assessment for the remainder of the study. Deidentified in-app user data for each participant were extracted by the study team to assess adherence to this requirement. Three weeks later, participants completed a final 7-day follow-up period with actigraphy, a sleep diary, and the questionnaires.

Ethical Considerations

Study procedures were approved by the Institutional Review Board at Virginia Tech (approval number 21-295). Participants signed a consent form before enrolling in the study and were informed that they could withdraw at any time. All data were deidentified before analysis.

Participants

Participants were recruited through flyers and advertisements at a local university (Virginia Tech, Virginia, USA). Inclusion criteria were assessed during the screening phase and included meeting the insomnia research diagnostic criteria (ISI score ≥11) [36], being 18-65 years old, and having access to a smartphone. Participants using sleep medications were included if their self-reported dosage had been stable for at least 6 weeks (medication dosage and type were balanced between study groups). Exclusion criteria, also self-reported and assessed during the screening phase, included (1) risk factors that may be comorbid with severe insomnia (eg, depression and self-harm suspected to interfere with the study protocol or requiring immediate treatment), assessed using the question, “The thought of harming myself has occurred to me,” and the 8-item Patient Health Questionnaire (PHQ-8) [37]. Participants scoring >0 (range 0-3) on the self-harm screening question and >20 on the PHQ-8 screening questionnaire were not included in the study. For safety reasons, individuals meeting these criteria were referred to mental health resources (National Alliance on Mental Illness helpline). Additional exclusion criteria included the following: (2) enrollment in a CBT-I program within the past 6 months; (3) regular mindfulness meditation practice (≥10 minutes, ≥2 times per week); and (4) inadequate English proficiency. Participants were compensated for their participation according to the following structure: US $50 for completing the baseline assessment, US $50 for the postintervention assessment, US $50 for the follow-up measures, and a US $100 completion bonus for completing >75% of the content. While compensation was intended to encourage participants to engage with >75% of the app content, it may have also limited the generalizability of the findings by contributing to higher adherence rates in this study.

Headspace Sleep Program (Intervention)

The Headspace Sleep Program is an app-based, self-guided program that incorporates CBT-I techniques augmented by mindfulness to help individuals adopt a nonjudgmental, self-compassionate approach to improving sleep habits. The 18-session program, designed for daily completion, includes mindfulness principles, cognitive techniques, relaxation techniques, and CBT-I behavioral techniques (eg, sleep hygiene, stimulus control, and wind-down routines). Notably, to enhance approachability and safety, the program excludes traditional CBT-I sleep restriction protocols. Instead, it provides information on the concept of sleep restriction and suggests it as an option for those who wish to try it. Sessions 1-11 are considered the core elements of the program and are assumed to provide the minimum effective dose, as all key techniques (eg, mindfulness, CBT-I, relaxation) are introduced by this point. Sessions 12-18 focus on integrating these techniques into daily practice. The content is delivered through short videos (5-10 minutes per day) in the Headspace app, and users are encouraged to engage in small daily actions (eg, trying a new wind-down routine). To enhance engagement, the program includes animations and videos featuring mindfulness teachers, sleep experts, and individuals with sleep difficulties. A supplemental workbook helps users track sleep, reflect on daily experiences, and create personalized sleep plans. Participants received instructions for app usage and followed the program without an introductory session to best reflect real-world engagement. They had the option to set up the app during the initial laboratory visit and were encouraged to contact the study

XSL•FO
RenderX

team if they encountered technical issues related to app setup or usage. None of the participants reported technical difficulties when setting up the app in the laboratory, at home, or during the study. Multimedia Appendix 2 outlines and describes all sessions in the Headspace Sleep Program.

Waitlist Control

The waitlist control group was instructed to maintain their regular routines and was given access to the Headspace Sleep Program after study completion.

Insomnia Outcomes

The ISI was used to assess self-reported insomnia symptoms at screening, baseline, postintervention, and the 3-week follow-up (score range 0-28; higher scores indicate more insomnia symptoms; subgroups: score range for moderate insomnia 15-21, score range for severe insomnia 22-28; Cronbach α=0.87) [38,39]. Additionally, the ISI has been validated as a clinical end point for a minimally important treatment response (ISI total score reduction >7) and remission (ISI total score <8) [38,39]. Total scores and clinical end points (ie, treatment response and remission) were assessed.

Subjective sleep quality was measured using a sleep diary, which calculated the following:

- Sleep efficiency (SE, %): the percentage of the sleep period spent sleeping (duration of sleep/duration of time in bed);
- Sleep onset latency (SOL; minutes): the duration of time from when the lights are turned off until falling asleep;
- Wake after sleep onset (WASO; minutes): the minutes awake after initially falling asleep; and
- Total sleep time (TST; minutes): the amount of sleep obtained at night [40].

The sleep diary was completed every morning over a period of 7 days at 3 time points (baseline, postintervention, and 3-week follow-up). For each time point, sleep diary data were averaged across the 7-day period, with a minimum of 4 days required for inclusion.

Objective sleep quality was measured using wrist actigraphy, a valid measure commonly used to assess sleep patterns [41]. Sleep was measured with the ActiGraph wGT3X-BT accelerometer at a frequency of 50 Hz (ActiGraph Corporation). Participants were instructed to wear the watch on their nondominant wrist from 2 hours before bedtime until the next morning for 7 consecutive days, at the same 3 time points as the sleep diary. Identical sleep parameters were computed for actigraphy as for the sleep diary (SE, SOL, WASO, and TST). For all 3 time points, data were averaged across the 7-day period, with a minimum of 4 days required for inclusion. Sleep patterns were assessed using a previously validated software algorithm based on the Cole-Kripke scoring method [42] and processed with ActiLife software (version 6.13.5; ActiGraph Corporation).

Mental Health Outcomes

All mental health questionnaires were collected at baseline, postintervention, and the 3-week follow-up. Perceived stress was assessed using the 10-item Perceived Stress Scale (PSS-10) (total score range 0-40, with higher scores indicating higher perceived stress; Cronbach α=0.85) [43,44]. The 8-item PHQ was used to measure depressive symptoms (total score range 0-24; higher scores indicate more depressive symptoms; Cronbach α=0.86) [37]. The Pittsburgh Sleep Quality Index (PSQI) was used to assess sleep quality (total score range 0-21, where "0" indicates no sleep difficulty and "21" indicates severe sleep difficulties [45]; Cronbach α=0.83) [45]. The 7-item Generalized Anxiety Disorder (GAD-7) was completed to measure anxiety symptoms (total score range 0-21, with higher scores indicating more anxiety symptoms; Cronbach α=0.89) [46]. Mindfulness was assessed using the Mindful Attention Awareness Scale (MAAS) (mean of the 15 items, range 1-6, with higher scores indicating higher levels of trait mindfulness; Cronbach α=0.80) [47]. The MAAS was chosen because it has been used in previous studies assessing the effects of mindfulness on sleep quality [48,49]. At postintervention, potential adverse effects were assessed by asking participants to report the number of potential adverse effects experienced during the intervention period. Specifically, during the postintervention in-person visit, we asked participants, "Did you experience any adverse effects as a participant in the app-based intervention?" Participants were given the option to write down any comments or responses to this question. No adverse effects were reported. Usage of sleep medication in both the intervention group and the waitlist control group was reported at baseline and follow-up assessments.

Adherence to the Headspace Sleep Program was obtained from in-app user data for each participant. It was calculated by summing the total number of completed sessions in the Headspace Sleep Program during the intervention period.

Statistical Analysis

All analyses were conducted using the intention-to-treat (ITT) and study-completer approaches. Study-completer results, shown in Multimedia Appendix 3, are comparable to the ITT analyses. For the ITT analysis, we used multiple imputations to handle missing outcomes (15/132, 11.4%, values were missing at postintervention and follow-up). Multiple imputations (10 imputations) were used to achieve a relative efficiency of 99%, ensuring in-range values.

Assumptions of statistical tests for normal distribution and sphericity of data were checked using the Shapiro-Wilk test, histograms, Q-Q plots, and box plots. A series of mixed 2 (groups: intervention and control) × 3 (time: preintervention, postintervention, and follow-up) ANOVAs was performed on the ISI, sleep diary variables, actigraphy variables, PSS-10, PSQI, PHQ-8, GAD-7, and MAAS. To analyze insomnia subgroups, a series of mixed 2 (groups: intervention and control) × 3 (time: preintervention, postintervention, and follow-up) × 2 (subgroups: mild insomnia and severe insomnia) ANOVAs was performed on the ISI, PSS-10, PSQI, PHQ-8, GAD-7, and MAAS. To examine clinical end points, logistic regression analyses were conducted to test differences in treatment response and remission using the cut scores described above on the ISI across postintervention and follow-up. Logistic regression analyses on remission and response status were performed to compare the 2 groups (intervention and control) at postintervention and 3-week follow-up (outcome variables) in

XSL•FO
RenderX

relation to preintervention values (covariate), without stepwise regression.

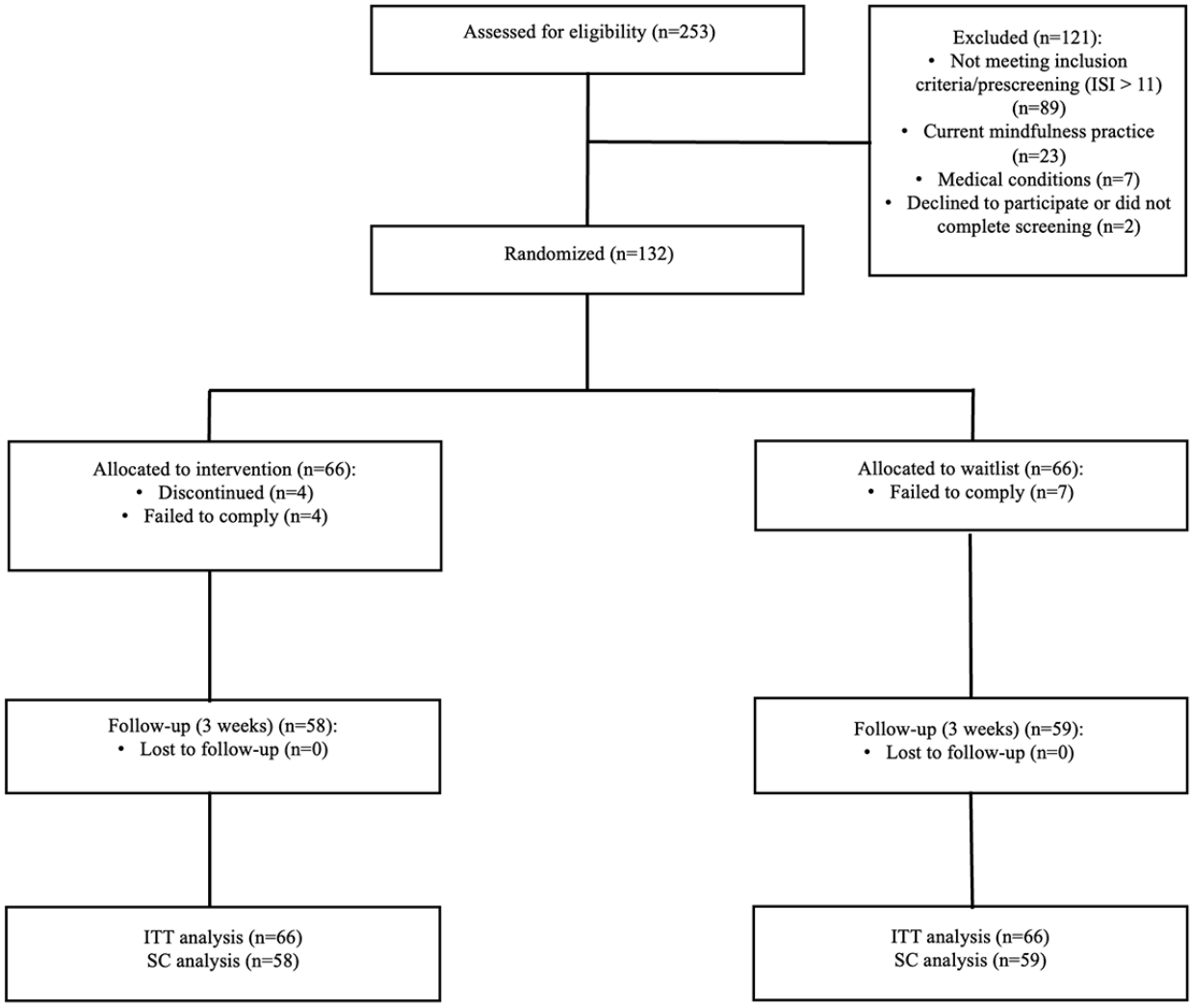
All significant interactions were followed up with relevant corrected pairwise comparisons using the Bonferroni method for simple main effects within each group. Where no significant interactions were found, the main effects of time and group were reported. Significance was set at .05 (2-tailed) for all analyses, and ANOVA effect sizes were calculated as partial eta squared (η^2p), with 0.02, 0.13, and 0.26 indicating small, medium, and large effects, respectively. Pearson correlations were used to assess whether a dosage-response relationship was present. Data analysis was conducted using SPSS (version 27; IBM Corp).

Results

Participant Enrollment and Intervention Adherence

Overall, 253 potential participants were assessed for eligibility, with 132 meeting the inclusion criteria, completing the informed consent process, and enrolling in the study (Figure 1; also see Multimedia Appendix 4). The 132 participants included in the analysis had a mean (SD) age of 37.2 (10.6) years, and 69 (52.3%) females identified as female (Table 1). No significant differences (all *P*s>.32; discussed later) were found between the intervention and control groups at baseline for demographic characteristics or study outcomes. On average, participants in the intervention group completed 13.0 (SD 5.4) sessions, corresponding to 13 of the 18 (72%) intervention contents. Of the 66 participants, 65 (98%) completed at least two sessions of the app-based Sleep Program. Additionally, 24 (36%) participants completed all 18 sessions of the intervention and 51 (77%) completed at least half of the intervention.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram and recruitment flowchart. ISI: Insomnia Severity Index; ITT: intention-to-treat analysis; SC: study completer analysis.



XSL•FO
RenderX

Table 1. Demographics and clinical characteristics by study group of the intention-to-treat sample.

Characteristics	Sleep intervention (n=66)	Waitlist control (n=66)
Gender, n (%)		
Male	32 (48)	31 (47)
Female	34 (52)	35 (53)
Age, mean (SD)	37.3 (11.5)	37.0 (9.6)
Marital status, n (%)		
Married	26 (39)	24 (36)
In a relationship	14 (21)	14 (21)
Single	18 (27)	22 (33)
Divorced	8 (12)	6 (9)
Employment, n (%)		
Full time	35 (53)	32 (48)
Part time	17 (26)	14 (21)
Retired	2 (3)	4 (6)
Unemployed	12 (18)	16 (24)
Household income (US $), n (%)		
<20,000	9 (14)	9 (14)
20,000-35,000	12 (18)	14 (21)
35,001-50,000	14 (21)	13 (20)
50,001-75,000	11 (17)	11 (17)
75,001-100,000	9 (14)	10 (15)
>100,000	11 (17)	9 (14)
Ethnicity, n (%)		
White	34 (52)	32 (48)
Asian	7 (11)	11 (17)
Black	10 (15)	12 (18)
Hispanic	9 (14)	7 (11)
Other	6 (9)	4 (6)
Sleep medication, n (%)		
Yes (baseline)	17 (26)	17 (26)
No (baseline)	49 (74)	49 (74)
Yes (follow-up)	18 (27)	18 (27)
No (follow-up)	48 (73)	48 (73)

At baseline, medication status in the intervention group showed that 17 of the 66 (26%) participants used sleep medication regularly, while 49 of the 66 (74%) participants did not. In the control group, 18 of the 66 (27%) participants used sleep medication regularly at baseline, whereas 48 of the 66 (73%) did not. At follow-up, there was no change in sleep medication usage in either group. There were no differences between the 2 groups at baseline or follow-up, and no changes within either group (Table 1). Medication status was used as a control variable to assess its impact on the outcomes. No significant interactions or main effects of medication status were observed for any variable analyzed (P=.61).

Insomnia Outcomes

ISI results indicated a group × time interaction (P<.001, η^2p=0.107), with the Headspace Sleep Program group showing a significant decrease in the ISI scores from pre- to postintervention (from a mean score of 14.8 to a mean score of 11.95, 19.26% decrease; P<.001), from preintervention to follow-up (from a mean score of 14.8 to a mean score of 10.70, 27.70% decrease; P<.001), and from postintervention to follow-up (from a mean score of 11.95 to a mean score of 10.70, 10.46% decrease; P=.02). No significant changes were observed in the control group (P=.35; Table 2).

XSL•FO
RenderX

Table 2. Behavioral data by study group and time point of the intention-to-treat sample.

Clinical value	Sleep intervention (n=66)			Waitlist control (n=66)		
	Pre, mean (SD)	Post, mean (SD)	Follow-up, mean (SD)	Pre, mean (SD)	Post, mean (SD)	Follow-up, mean (SD)
Insomnia Severity Index	14.80 (3.13)	11.95 (3.93)[a,b]	10.70 (3.37)[b,c,d,e]	14.81 (2.93)	14.88 (4.60)	14.80 (3.39)
Pittsburgh Sleep Quality Index	10.31 (2.48)	9.41 (2.37)[a,b]	8.12 (2.43)[b,c,d,e]	10.43 (2.58)	10.64 (2.25)	10.18 (1.83)
10-item Perceived Stress Scale	16.89 (3.91)	18.17 (3.59)	17.73 (3.63)	17.09 (3.65)	17.84 (3.06)	17.97 (2.49)
8-item Patient Health Questionnaire	11.42 (3.69)	10.03 (4.52)[a,b]	9.01 (3.87)[b,c,d,e]	11.56 (3.37)	11.41 (2.78)	10.72 (2.97)
7-item Generalized Anxiety Disorder	10.89 (4.40)	8.76 (3.84)[a,b]	8.21 (4.25)[b,c,d,f]	10.80 (3.89)	11.27 (3.57)	11.08 (3.52)
Mindful Attention Awareness Scale	50.28 (15.38)	52.83 (15.11)[a,e]	54.80 (14.40)[b,c,d,e]	50.69 (13.89)	50.35 (12.28)	50.83 (11.28)

[a]Significant difference between pre and post time points.

[b]*P*<.001.

[c]Significant difference between pre and follow-up time points.

[d]Significant difference between post and follow-up time points.

[e]*P*<.01.

[f]*P*<.05.

Remission rates (ISI total score <8) [38,39] in the intervention group increased significantly (*P*=.03) over time (8/66, 12%, at postintervention and 9/66, 14%, at follow-up). By contrast, remission rates in the control group remained unchanged (*P*=.29; 2/66, 3%, at postintervention and 0/66, 0%, at follow-up; Multimedia Appendix 5). Similarly, treatment response (ISI total score reduction >7) [38,39] showed a steady increase from pre- to postintervention (11/66, 17%) to follow-up (15/66, 23%) in the intervention group. By contrast, treatment response did not significantly change in the control group (*P*=.27; 3/66, 5%, at postintervention and 1/66, 2%, at follow-up; Multimedia Appendix 6).

Sleep diary results suggest that the Headspace Sleep Program group significantly increased SE from pre- to postintervention (*P*=.002) and from preintervention to follow-up (*P*=.01), but not from postintervention to follow-up (*P*=.69; Table 3). A significant group × time interaction was identified (*P*=.01, η^2p=0.05). No significant changes in SE were observed in the control group (*P*=.19). No group × time interactions were found for SOL (*P*=.15, η^2p=.01), WASO (*P*=.18, η^2p=.01), and TST (*P*=.11, η^2p=.01). For TST, a significant main effect of group (*P*=.04, η^2p=.03) indicated that the groups were significantly different, regardless of time points (pre-, postintervention, or follow-up; Table 3). No main effects of time were observed for SOL (*P*=.15, η^2p=.01) or WASO (*P*=.38, η^2p=.01). A significant main effect of group for SOL (*P*=.03, η^2p=.04) showed that both groups increased in SOL from pre- to postintervention and follow-up (Table 3).

Table 3. Subjective measures of sleep (ie, sleep diary) by study group and time point of the intention-to-treat sample.

Clinical value	Sleep intervention (n=66)			Waitlist control (n=66)		
	Pre, mean (SD)	Post, mean (SD)	Follow-up, mean (SD)	Pre, mean (SD)	Post, mean (SD)	Follow-up, mean (SD)
Sleep efficiency (%)	75.21 (16.51)	82.38 (9.98)[a,b]	82.01 (13.10)[c,d]	77.10 (12.84)	75.04 (14.24)	78.29 (9.58)
Sleep onset latency (minutes)	39.66 (21.97)	29.45 (23.31)	26.97 (29.71)	38.33 (28.23)	42.08 (39.76)	35.92 (25.14)
Wake after sleep onset (minutes)	85.59 (83.53)	60.40 (53.97)	69.43 (77.78)	75.74 (69.17)	82.22 (65.93)	70.90 (62.75)
Total sleep time (minutes)	383.93 (104.59)	402.65 (78.64)	410.53 (104.67)	390.98 (101.91)	378.67 (95.34)	369.41 (86.64)

[a]Significant difference between pre and post time points.

[b]*P*<.05.

[c]*P*<.01.

[d]Significant difference between pre and follow-up time points.

Objective sleep quality actigraphy results mirrored those of the subjective sleep diary. A group × time interaction was identified for SE (*P*<.001, η^2p=.085), SOL (*P*<.001, η^2p=.06), and WASO (*P*=.01, η^2p=.03), but not for TST (*P*=.06, η^2p=.02). The

XSL•FO
RenderX

Headspace Sleep Program group showed significant improvements in SE, SOL, and WASO (P=.01), with no significant improvements observed in the control group (P=.25; Table 4). For TST, a significant main effect of group (P=.02, η^2p=.04) indicated a significant difference between the groups, regardless of time points (Table 4).

Table 4. Objective measures of sleep (ie, actigraphy) by study group and time point of the intention-to-treat sample.

Clinical value	Sleep intervention (n=66)			Waitlist control (n=66)		
	Pre, mean (SD)	Post, mean (SD)	Follow-up, mean (SD)	Pre, mean (SD)	Post, mean (SD)	Follow-up, mean (SD)
Sleep efficiency (%)	74.40 (7.17)	82.20 (8.26)[a,b]	81.85 (9.93)[b,c]	76.49 (9.61)	74.61 (13.04)	75.08 (11.12)
Sleep onset latency (minutes)	50.34 (22.23)	31.36 (20.58)[a,b]	28.15 (21.98)[b,c]	39.89 (23.39)	43.57 (38.30)	42.83 (24.98)
Wake after sleep onset (minutes)	78.71 (45.94)	56.92 (45.08)[a,d]	62.15 (53.62)[c,d]	67.60 (48.58)	82.65 (64.63)	73.14 (46.95)
Total sleep time (minutes)	377.71 (58.49)	395.52 (52.14)	391.80 (60.90)	383.30 (69.51)	374.54 (89.25)	354.60 (70.32)

[a]Significant difference between pre and post time points.

[b]P<.001.

[c]Significant difference between pre and follow-up time points.

[d]P<.05.

Subgroup analysis compared mild versus moderate baseline insomnia symptoms (mild: ISI range 8-14, 63 patients; moderate: ISI range 15-21, 69 patients). ANOVA results showed a significant 3-way interaction (P=.02, η^2p=.03). Follow-up testing revealed that participants with moderate baseline insomnia symptoms in the Headspace Sleep Program group experienced significant improvements in the ISI from pre- to postintervention (P<.001), preintervention to follow-up (P<.001), and postintervention to follow-up (P=.03; Multimedia Appendix 7). Participants with mild baseline insomnia symptoms in the Headspace Sleep Program group did not show a significant change from preintervention to postintervention (P=.79). However, they did show significant improvements from preintervention to follow-up (P=.04) and from postintervention to follow-up (P=.04).

Mental Health Outcomes

Group × time interactions were identified for the PHQ-8 (P=.01, η^2p=.03), PSQI (P<.001, η^2p=.10), GAD-7 (P<.001, η^2p=.11), and MAAS (P=.01, η^2p=.04). Results for the PHQ-8, PSQI, GAD-7, and MAAS showed that the Headspace Sleep Program group significantly improved at all time points (P=.02), whereas the control group did not show significant changes at any time point (P=.19). PSS-10 results did not show any group × time interaction (P=.63, η^2p=.005) or main effect of group (P=.42, η^2p=.001). However, a significant main effect of time (P=.01, η^2p=.03) indicated that both groups showed an increase in PSS-10 scores over time, from pre- to postintervention to follow-up (Table 2).

Subgroup ANOVA results showed no significant 3- or 2-way interactions for the PSS-10 (P=.41, η^2p=.003). However, a main effect of time was observed (P=.005, η^2p=.03). Similarly, no significant 3- or 2-way interactions were found for the PHQ-8 (P=.33, η^2p=.004). A main effect of time (P<.001, η^2p=.05) and group (P=.007, η^2p=.04) was observed. No significant 3- or 2-way interactions were found for the PSQI (P=.43, η^2p=.002). A main effect of time was observed (P=.02, η^2p=.03). No significant 3- or 2-way interactions were found for GAD-7 (P=.27, η^2p=.001). A main effect of time (P<.001, η^2p=.05) and group (P=.01, η^2p=.04) was observed.

ANOVA results showed a significant 3-way interaction for the MAAS (P=.02, η^2p=.03), with follow-up testing revealing that the moderate insomnia group in the experimental condition showed significant improvement in the MAAS scores from pre- to postintervention (P=.003) and from preintervention to follow-up (P<.001), but not from postintervention to follow-up (P=.20). By contrast, the mild insomnia group in the experimental condition did not show a significant change from preintervention to postintervention (P=.60) or from preintervention to follow-up (P=.08). However, a significant difference was observed from postintervention to follow-up (P=.03). No significant differences were found for either moderate (P=.31) or mild insomnia (P=.35) in the waitlist control group at any time point. Significant dosage-response correlations (P=.02) were found with weak to moderate strength (ranging from 0.28 to 0.34) only for SE at postintervention (for the sleep diary variable) and WASO at postintervention (for the actigraphy variable). No other variables showed significant differences (P=.23).

Discussion

Principal Findings

This study aimed to evaluate the efficacy of the Headspace Sleep Program in improving sleep disturbances and mental health outcomes in adults with clinical insomnia. The results provide support for the study hypotheses, showing that insomnia symptoms significantly decreased and both subjective and objective sleep quality, as well as mental health outcomes, significantly improved for participants in the Headspace Sleep Program group compared with those in the waitlist control group. Importantly, the program's benefits were maintained 3 weeks postintervention. Clinically, participants in the Headspace Sleep Program group exhibited higher remission rates and better

XSL•FO
RenderX

treatment responses compared with those in the control group, although the effect sizes were small.

The findings of this study contribute to existing research supporting the efficacy of CBT-I techniques and mindfulness as nonpharmacological treatments for insomnia symptoms [16,23-26]. Importantly, this clinical trial extends the literature by demonstrating that an app-based, self-guided sleep program incorporating CBT-I techniques augmented with mindfulness is an effective intervention for reducing insomnia symptoms. The results of this trial suggest that the Headspace Sleep Program could serve as a viable alternative for improving sleep and mental health in individuals unable to access in-person or alternative CBT-I treatments. While CBT-I remains the first-line treatment for insomnia, as endorsed by the American College of Physicians [19] and the European Insomnia Guidelines [20], it may not be accessible to everyone in need of treatment. The results of this trial suggest that the Headspace Sleep Program could serve as a viable alternative for improving sleep and mental health in individuals unable to access in-person or alternative CBT-I treatments. As such, digital CBT-I interventions have been developed, suggesting outcomes similar to in-person treatment [23-25]. However, digital CBT-I can still be inaccessible to those who need it due to its cost, reliance on a provider's schedule, and the fact that treatment can take weeks to show results [18]. The Headspace Sleep Program was designed to be lower in cost, more timely, and more flexible compared with traditional CBT-I. Results suggest that the program improves insomnia symptoms and mental health while also being more accessible than traditional treatments (Multimedia Appendix 1). Specifically, the Headspace Sleep Program is more accessible than both in-person and digital CBT-I because (1) the program is delivered via the Headspace app, which offers a subscription at a lower cost than both digital and in-person CBT-I; (2) all sessions are self-guided, allowing patients to access content at their convenience without relying on providers' availability; (3) the intervention consists of 18 daily sessions, a much shorter time frame than traditional CBT-I, which is typically delivered weekly over several weeks; and (4) offering the program through the Headspace app may reduce treatment stigma and make it available to those who may not be aware that insomnia treatment is accessible to them (Multimedia Appendix 1). For these reasons, we note that the program was not designed to be a digital CBT-I intervention, but rather to use CBT-I techniques augmented by mindfulness, creating a program that is more effective than mindfulness-focused sleep content while being more accessible than traditional CBT-I. As such, research on digital CBT-I suggests greater remission rates compared with the current program [32,50]. Specifically, the SHUTi CBT-I intervention reported a remission rate of 56.7% (vs the current study's 9/66, 14%) [16]. However, the SHUTi CBT-I involved 9 weeks of clinical sessions, whereas the current program consisted of 18 consecutive self-guided sessions. Our study's findings suggest results that are similar to or stronger than those from interventions using mindfulness alone to improve insomnia [16]. Therefore, the Headspace Sleep Program's design, which combines CBT-I techniques with mindfulness, may be more beneficial than mindfulness alone, and its greater accessibility could provide an effective intervention for individuals who are unable to receive traditional CBT-I to improve sleep quality.

This study also evaluated the efficacy of the Headspace Sleep Program on mental health outcomes, including perceived stress, depression, anxiety, and mindfulness. The results suggest that the intervention led to improvements in depressive and anxiety symptoms, as well as in mindfulness. On average, participants entered the study with mild depressive symptoms and moderate anxiety symptoms. While this study cannot fully disentangle the confounding relationship between participants' baseline insomnia, depressive, and anxiety symptoms, or determine whether the program's impact on insomnia symptoms directly caused the decrease in depressive and anxiety symptoms, the findings suggest that mental health outcomes significantly improved for those who engaged in the program. In addition, the PSQI showed a significant improvement in sleep quality in the intervention group, supporting the ISI, sleep diary, and actigraphy results. However, the results did not show improvements in perceived stress, despite the improvements in depressive and anxiety symptoms and the comorbidity between these disorders [51]. A possible explanation for the nonsignificant perceived stress result may be situational circumstances driving overall stress. The relationship between sleep and stress is bidirectional and complex. Our app-based intervention focused primarily on sleep, which might not have adequately addressed the multifaceted nature of stress, often involving psychosocial factors [52]. However, our app-based intervention did incorporate mindfulness, and previous studies have demonstrated significant reductions in self-reported stress using app-based mindfulness interventions in participants with elevated stress [53]. Future studies are needed to further disentangle the impact of app-based sleep interventions on stress. Trait mindfulness increased in the intervention group, which may be attributed to the mindfulness components embedded within the Headspace Sleep Program. This finding supports previous research suggesting that mindfulness is effective in patients with chronic insomnia [16,17].

As is typical with digital programs, there was a dropout from the intervention (15/132, 11.4%, values were missing at post and follow-up); however, ITT analyses still identified significant improvements. Sensitivity and missing data analyses did not alter the study conclusions when compared with study-completer analyses (Multimedia Appendix 3).

There were no changes in sleep medication usage in either the intervention group or the waitlist control group from baseline to follow-up. This nonsignificant finding may be explained by the relatively short period between baseline and follow-up (approximately 6 weeks). Regarding adherence rates in this study, participants, on average, completed 13 of the 18 (72%) intervention contents. Previous studies using digital sleep interventions have found comparable results. Specifically, one study reported that participants completed, on average, 50% of the intervention [54], another found a completion rate of 56.7% [55], and a third study reported 93.3% adherence [56]. Thus, our study demonstrates adherence rates similar to those observed in other app-based sleep intervention trials.

XSL•FO
RenderX

Strengths, Limitations, and Future Research

To our knowledge, this is the first RCT to investigate the impact of an app-based, self-guided program that uses CBT-I techniques augmented by mindfulness to improve sleep disturbance in adults with clinical insomnia. The adherence data suggest that the intervention is feasible, while both subjective and objective sleep and mental health outcomes indicate that the intervention is effective. However, it should be noted that the financial compensation provided to participants may have inflated adherence rates. Other limitations of the study include the potential impact of response bias on the self-report survey results. Specifically, because participants were compensated, we cannot rule out the possibility that their responses to participant-reported outcomes may have been biased. However, as the study also used objective measures, it reduced the likelihood of bias. Additionally, the requirement for participants to have smartphone access may have impacted the overall sample, making the results less generalizable. Ethical guidelines, such as those from the Council for International Organizations of Medical Sciences, recommend that an established effective intervention be offered to study participants in waitlist control groups, which was not adhered to in this study. However, all participants in the waitlist control group were given the option to access the Headspace Sleep Program after study completion. Furthermore, although group randomization was performed before the study launch, the research team was not formally blinded, and thus, allocation concealment may have introduced biased estimates of intervention effects, as shown in previous research [57]. Finally, this study used a waitlist control group; future research should include comparisons to sleep-specific mindfulness content or other traditional insomnia treatments. Future research could expand on this study by evaluating the Headspace Sleep Program in a real-world setting and investigating the integration of the Headspace Sleep Program with a Headspace mental health coach. Research shows that pairing digital mental health interventions with human support may improve adherence and overall outcomes [58].

Conclusions

This RCT found that the Headspace Sleep Program significantly improved insomnia symptoms, sleep quality, sleep disturbance, depressive symptoms, anxiety symptoms, and mindfulness. Insomnia remission rates and treatment response increased over time in the Headspace Sleep Program group, but not in the waitlist control group. These results suggest that the program is an effective intervention for improving sleep disturbance in adults with clinical insomnia. However, further research is needed to better understand the impact of the Headspace Sleep Program compared with traditional treatments and in real-world settings.

Authors' Contributions

WS and UK were involved in the study design, data collection, data analysis, interpretation of data, and writing of the paper. C Callahan, MD, LT, C Coe, and SK were involved in the study design, interpretation of data, and writing of the paper.

Conflicts of Interest

C Callahan, MD, LT, C Coe, and SK are currently employed by Headspace.

Multimedia Appendix 1
Differences between the Headspace Sleep Program, sleep mindfulness content, and cognitive behavioral therapy for insomnia.
[DOCX File , 15 KB - mhealth_v13i1e68665_app1.docx]

Multimedia Appendix 2
Headspace Sleep Program session outline and descriptions.
[DOCX File , 16 KB - mhealth_v13i1e68665_app2.docx]

Multimedia Appendix 3
Supplemental results.
[DOCX File , 2868 KB - mhealth_v13i1e68665_app3.docx]

Multimedia Appendix 4
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 436 KB - mhealth_v13i1e68665_app4.pdf]

Multimedia Appendix 5
Percentage of participants who met the criteria for treatment remission (ISI score<8 at both postintervention and follow-up). ISI: Insomnia Severity Index. Figure S1
[PNG File , 44 KB - mhealth_v13i1e68665_app5.png]

Multimedia Appendix 6

XSL•FO
RenderX

Percentage of participants who met the criteria for a minimally important treatment response (ISI total score reduction from baseline >7 points at postintervention and follow-up). ISI: Insomnia Severity Index.
[PNG File , 48 KB - mhealth_v13i1e68665_app6.png]

Multimedia Appendix 7
Subgroup analysis comparing baseline mild versus moderate insomnia symptoms. $ indicates 3-way interaction.
[PNG File , 91 KB - mhealth_v13i1e68665_app7.png]

References

1. Walker W, Walton JC, DeVries AC, Nelson RJ. Circadian rhythm disruption and mental health. Transl Psychiatry 2020 Jan 23;10(1):28 [FREE Full text] [doi: 10.1038/s41398-020-0694-0] [Medline: 32066704]
2. Arble D, Bass J, Behn C, Butler MP, Challet E, Czeisler C, et al. Impact of sleep and circadian disruption on energy balance and diabetes: a summary of workshop discussions. Sleep 2015 Dec 01;38(12):1849-1860 [FREE Full text] [doi: 10.5665/sleep.5226] [Medline: 26564131]
3. Altman NG, Izci-Balserak B, Schopfer E, Jackson N, Rattanaumpawan P, Gehrman PR, et al. Sleep duration versus sleep insufficiency as predictors of cardiometabolic health outcomes. Sleep Medicine 2012 Dec;13(10):1261-1270. [doi: 10.1016/j.sleep.2012.08.005]
4. Buxton OM, Marcelli E. Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States. Social Science & Medicine 2010 Sep;71(5):1027-1036. [doi: 10.1016/j.socscimed.2010.05.041]
5. Grandner MA, Schopfer EA, Sands‐Lincoln M, Jackson N, Malhotra A. Relationship between sleep duration and body mass index depends on age. Obesity 2015 Nov 02;23(12):2491-2498. [doi: 10.1002/oby.21247]
6. Grandner MA, Chakravorty S, Perlis ML, Oliver L, Gurubhagavatula I. Habitual sleep duration associated with self-reported and objectively determined cardiometabolic risk factors. Sleep Medicine 2014 Jan;15(1):42-50. [doi: 10.1016/j.sleep.2013.09.012]
7. Knutson KL. Does inadequate sleep play a role in vulnerability to obesity? Am J Hum Biol 2012 Jan 24;24(3):361-371. [doi: 10.1002/ajhb.22219]
8. Knutson KL, Spiegel K, Penev P, Van Cauter E. The metabolic consequences of sleep deprivation. Sleep Medicine Reviews 2007 Jun;11(3):163-178. [doi: 10.1016/j.smrv.2007.01.002]
9. Mental health and sleep. Sleep Foundation. URL: https://www.sleepfoundation.org/mental-health#references-80280 [accessed 2025-09-01]
10. Morin CM, Jarrin DC. Epidemiology of insomnia: prevalence, course, risk factors, and public health burden. Sleep Med Clin 2022 Jun;17(2):173-191. [doi: 10.1016/j.jsmc.2022.03.003] [Medline: 35659072]
11. Bhaskar S, Hemavathy D, Prasad S. Prevalence of chronic insomnia in adult patients and its correlation with medical comorbidities. J Family Med Prim Care 2016;5(4):780-784 [FREE Full text] [doi: 10.4103/2249-4863.201153] [Medline: 28348990]
12. Léger D, Bayon V. Societal costs of insomnia. Sleep Med Rev 2010 Dec;14(6):379-389. [doi: 10.1016/j.smrv.2010.01.003] [Medline: 20359916]
13. Stoller MK. Economic effects of insomnia. Clin Ther 1994;16(5):873-97; discussion 854. [Medline: 7859246]
14. Richey MS, Krystal DA. Pharmacological advances in the treatment of insomnia. CPD 2011 May 01;17(15):1471-1475. [doi: 10.2174/138161211796197052]
15. Trauer JM, Qian MY, Doyle JS, Rajaratnam SM, Cunnington D. Cognitive behavioral therapy for chronic insomnia. Ann Intern Med 2015 Aug 04;163(3):191-204. [doi: 10.7326/m14-2841]
16. Ong JC, Manber R, Segal Z, Xia Y, Shapiro S, Wyatt JK. A randomized controlled trial of mindfulness meditation for chronic insomnia. Sleep 2014 Sep 01;37(9):1553-1563 [FREE Full text] [doi: 10.5665/sleep.4010] [Medline: 25142566]
17. Garland SN, Carlson LE, Stephens AJ, Antle MC, Samuels C, Campbell TS. Mindfulness-based stress reduction compared with cognitive behavioral therapy for the treatment of insomnia comorbid with cancer: a randomized, partially blinded, noninferiority trial. JCO 2014 Feb 10;32(5):449-457. [doi: 10.1200/JCO.2012.47.7265]
18. Mitchell MD, Gehrman P, Perlis M, Umscheid CA. Comparative effectiveness of cognitive behavioral therapy for insomnia: a systematic review. BMC Fam Pract 2012 May 25;13(1):1471 [FREE Full text] [doi: 10.1186/1471-2296-13-40] [Medline: 22631616]
19. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2016 Jul 19;165(2):125-133. [doi: 10.7326/m15-2175]
20. Riemann D, Espie C, Altena E, Arnardottir ES, Baglioni C, Bassetti CLA, et al. The European Insomnia Guideline: an update on the diagnosis and treatment of insomnia 2023. J Sleep Res 2023 Dec;32(6):e14035. [doi: 10.1111/jsr.14035] [Medline: 38016484]

XSL•FO
RenderX

21. van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. Sleep Med Rev 2018 Apr;38:3-16. [doi: 10.1016/j.smrv.2017.02.001] [Medline: 28392168]
22. Chen T, Chang S, Hsieh H, Huang C, Chuang J, Wang H. Effects of mindfulness-based stress reduction on sleep quality and mental health for insomnia patients: a meta-analysis. J Psychosom Res 2020 Aug;135:110144. [doi: 10.1016/j.jpsychores.2020.110144] [Medline: 32590218]
23. Luik AI, Kyle SD, Espie CA. Digital cognitive behavioral therapy (dCBT) for insomnia: a state-of-the-science review. Curr Sleep Med Rep 2017 May 8;3(2):48-56 [FREE Full text] [doi: 10.1007/s40675-017-0065-4] [Medline: 28553574]
24. Seyffert M, Lagisetty P, Landgraf J, Chopra V, Pfeiffer PN, Conte ML, et al. Internet-delivered cognitive behavioral therapy to treat insomnia: a systematic review and meta-analysis. PLoS One 2016 Feb 11;11(2):e0149139 [FREE Full text] [doi: 10.1371/journal.pone.0149139] [Medline: 26867139]
25. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
26. Gong H, Ni C, Liu Y, Zhang Y, Su W, Lian Y, et al. Mindfulness meditation for insomnia: a meta-analysis of randomized controlled trials. J Psychosom Res 2016 Oct;89:1-6. [doi: 10.1016/j.jpsychores.2016.07.016] [Medline: 27663102]
27. Low T, Conduit R, Varma P, Meaklim H, Jackson ML. Treating subclinical and clinical symptoms of insomnia with a mindfulness-based smartphone application: a pilot study. Internet Interv 2020 Sep;21:100335 [FREE Full text] [doi: 10.1016/j.invent.2020.100335] [Medline: 32939342]
28. Huberty JL, Green J, Puzia ME, Larkey L, Laird B, Vranceanu A, et al. Testing a mindfulness meditation mobile app for the treatment of sleep-related symptoms in adults with sleep disturbance: a randomized controlled trial. PLoS One 2021 Jan 7;16(1):e0244717 [FREE Full text] [doi: 10.1371/journal.pone.0244717] [Medline: 33411779]
29. Lin W, Li N, Yang L, Zhang Y. The efficacy of digital cognitive behavioral therapy for insomnia and depression: a systematic review and meta-analysis of randomized controlled trials. PeerJ 2023;11:e16137 [FREE Full text] [doi: 10.7717/peerj.16137] [Medline: 37927792]
30. Ong JC, Kalmbach DA. Mindfulness as an adjunct or alternative to CBT-I. Sleep Med Clin 2023 Mar;18(1):59-71. [doi: 10.1016/j.jsmc.2022.09.002] [Medline: 36764787]
31. Ong JC, Shapiro SL, Manber R. Combining mindfulness meditation with cognitive-behavior therapy for insomnia: a treatment-development study. Behav Ther 2008 Jun;39(2):171-182 [FREE Full text] [doi: 10.1016/j.beth.2007.07.002] [Medline: 18502250]
32. Furukawa Y, Sakata M, Yamamoto R, Nakajima S, Kikuchi S, Inoue M, et al. Components and delivery formats of cognitive behavioral therapy for chronic insomnia in adults: a systematic review and component network meta-analysis. JAMA Psychiatry 2024 Apr 01;81(4):357-365 [FREE Full text] [doi: 10.1001/jamapsychiatry.2023.5060] [Medline: 38231522]
33. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
34. Savard J, Simard S, Ivers H, Morin CM. Randomized study on the efficacy of cognitive-behavioral therapy for insomnia secondary to breast cancer, part I: Sleep and psychological effects. J Clin Oncol 2005 Sep 01;23(25):6083-6096. [doi: 10.1200/JCO.2005.09.548] [Medline: 16135475]
35. Kang H. Sample size determination and power analysis using the G*Power software. J Educ Eval Health Prof 2021 Jul 30;18:17 [FREE Full text] [doi: 10.3352/jeehp.2021.18.17] [Medline: 34325496]
36. Bastien C, Vallières A, Morin C. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
37. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
38. Morin C. Insomnia: Psychological Assessment and Management. New York City, NY: Guilford Press; 1993.
39. Morin CM, Vallières A, Guay B, Ivers H, Savard J, Mérette C, et al. Cognitive behavioral therapy, singly and combined with medication, for persistent insomnia: a randomized controlled trial. JAMA 2009 May 20;301(19):2005-2015 [FREE Full text] [doi: 10.1001/jama.2009.682] [Medline: 19454639]
40. Morin CM, Espie CA. Insomnia: A Clinical Guide to Assessment and Treatment. New York City, NY: Springer; Jan 01, 2004.
41. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003 May 01;26(3):342-392. [doi: 10.1093/sleep/26.3.342] [Medline: 12749557]
42. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992 Oct;15(5):461-469. [doi: 10.1093/sleep/15.5.461] [Medline: 1455130]
43. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. Journal of Health and Social Behavior 1983 Dec;24(4):385. [doi: 10.2307/2136404]
44. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]

XSL•FO
RenderX

45. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
46. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
47. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003 Apr;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
48. Ye J, Jia X, Zhang J, Guo K. Effect of physical exercise on sleep quality of college students: chain intermediary effect of mindfulness and ruminative thinking. Front Psychol 2022;13:987537-987537 [FREE Full text] [doi: 10.3389/fpsyg.2022.987537] [Medline: 36262438]
49. Zuo X, Tang Y, Chen Y, Zhou Z. The efficacy of mindfulness-based interventions on mental health among university students: a systematic review and meta-analysis. Front Public Health 2023 Nov 30;11:1259250-1259250 [FREE Full text] [doi: 10.3389/fpubh.2023.1259250] [Medline: 38146476]
50. Simon L, Steinmetz L, Feige B, Benz F, Spiegelhalder K, Baumeister H. Comparative efficacy of onsite, digital, and other settings for cognitive behavioral therapy for insomnia: a systematic review and network meta-analysis. Sci Rep 2023 Feb 02;13(1):1929-1929 [FREE Full text] [doi: 10.1038/s41598-023-28853-0] [Medline: 36732610]
51. Manzar MD, Salahuddin M, Pandi-Perumal SR, Bahammam AS. Insomnia may mediate the relationship between stress and anxiety: a cross-sectional study in university students. Nat Sci Sleep 2021;13:31-38 [FREE Full text] [doi: 10.2147/NSS.S278988] [Medline: 33447116]
52. Nieuwenhuijsen K, Bruinvels D, Frings-Dresen M. Psychosocial work environment and stress-related disorders: a systematic review. Occup Med (Lond) 2010 Jun;60(4):277-286. [doi: 10.1093/occmed/kqq081] [Medline: 20511268]
53. Kirk U, Staiano W, Hu E, Ngnoumen C, Kunkle S, Shih E, et al. App-based mindfulness for attenuation of subjective and physiological stress reactivity in a population with elevated stress: randomized controlled trial. JMIR Mhealth Uhealth 2023 Oct 13;11:e47371 [FREE Full text] [doi: 10.2196/47371] [Medline: 37831493]
54. Vollert B, Müller L, Jacobi C, Trockel M, Beintner I. Effectiveness of an app-based short intervention to improve sleep: randomized controlled trial. JMIR Ment Health 2023 Mar 21;10:e39052 [FREE Full text] [doi: 10.2196/39052] [Medline: 36943337]
55. Lancee J, van den Bout J, van Straten A, Spoormaker VI. Internet-delivered or mailed self-help treatment for insomnia?: a randomized waiting-list controlled trial. Behav Res Ther 2012 Jan;50(1):22-29. [doi: 10.1016/j.brat.2011.09.012] [Medline: 22055281]
56. Lorenz N, Heim E, Roetger A, Birrer E, Maercker A. Randomized controlled trial to test the efficacy of an unguided online intervention with automated feedback for the treatment of insomnia. Behav Cogn Psychother 2019 May;47(3):287-302 [FREE Full text] [doi: 10.1017/S1352465818000486] [Medline: 30185239]
57. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008 Mar 15;336(7644):601-605 [FREE Full text] [doi: 10.1136/bmj.39465.451748.AD] [Medline: 18316340]
58. Werntz A, Amado S, Jasman M, Ervin A, Rhodes JE. Providing human support for the use of digital mental health interventions: systematic meta-review. J Med Internet Res 2023 Feb 06;25:e42864 [FREE Full text] [doi: 10.2196/42864] [Medline: 36745497]

Abbreviations

CBT-I: cognitive behavioral therapy for insomnia
GAD-7: 7-item Generalized Anxiety Disorder
ISI: Insomnia Severity Index
ITT: intention-to-treat
MAAS: Mindful Attention Awareness Scale
PHQ-8: 8-item Patient Health Questionnaire
PSQI: Pittsburgh Sleep Quality Index
PSS-10: 10-item Perceived Stress Scale
RCT: randomized controlled trial
SOL: sleep onset latency
TST: total sleep time
WASO: wake after sleep onset

XSL•FO
RenderX

Edited by L Buis; submitted 11.11.24; peer-reviewed by M Merlini, CJ Lin, D Lehr; comments to author 11.12.24; revised version received 22.12.24; accepted 18.03.25; published 23.04.25.

Please cite as:
Staiano W, Callahan C, Davis M, Tanner L, Coe C, Kunkle S, Kirk U
Assessment of an App-Based Sleep Program to Improve Sleep Outcomes in a Clinical Insomnia Population: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e68665
URL: https://mhealth.jmir.org/2025/1/e68665
doi:10.2196/68665
PMID:

©Walter Staiano, Christine Callahan, Michelle Davis, Leah Tanner, Chelsea Coe, Sarah Kunkle, Ulrich Kirk. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Promoting Hand Hygiene During the COVID-19 Pandemic: Randomized Controlled Trial of the Optimized Soapp+ App

Dario Baretta[1], PhD; Carole Lynn Rüttimann[1], MSc; Melanie Alexandra Amrein[1], PhD; Jennifer Inauen[1], PhD

Institute of Psychology, University of Bern, Bern, Switzerland

Corresponding Author:
Dario Baretta, PhD
Institute of Psychology
University of Bern
Fabrikstarsse 8
Bern, 3012
Switzerland
Phone: 41 31 684 81 11
Email: dario.baretta@unibe.ch

Abstract

Background: The adoption of protective behaviors represents a crucial measure to counter the spread of infectious diseases. The development of effective behavior change techniques therefore emerged as a public health priority during the COVID-19 pandemic, but randomized controlled trials (RCTs) testing such interventions during the pandemic were scarce. We conducted a Multiphase Optimization Strategy to develop, optimize, and evaluate a smartphone app, Soapp+, to promote hand hygiene during the COVID-19 pandemic.

Objective: This RCT aims to evaluate the efficacy of the Soapp+ app (intervention group) targeting motivation and habit compared to a simplified version of the app mainly delivering hand hygiene information (active control group). We hypothesize that, compared to the control group, the intervention group will show greater improvements in hand hygiene behavior and behavioral determinants post intervention and at a 6-month follow-up.

Methods: We conducted an RCT from March 2022 to April 2023, recruiting 193 adults living in Switzerland online. Following baseline assessment, the intervention lasted 32 days, followed by a postintervention assessment and a 6-month follow-up. The primary outcome was the change in hand hygiene behavior from pre- to postintervention and preintervention to follow-up. Hand hygiene was assessed with electronic diaries. The intervention group received content incorporating various behavior change techniques designed to address key motivational and volitional determinants of hand hygiene behavior (eg, skills, knowledge, intention, attitudes toward hand hygiene, risk perception, outcome expectancies, self-efficacy, action planning, coping planning, action control, habit). In contrast, the active control group was exposed to behavior change techniques targeting only a subset of these determinants (ie, skills, knowledge, and intention). The delivery of the intervention content was fully automated. Group differences were tested using an intention-to-treat approach with the nonparametric Wilcoxon rank sum test.

Results: Of the 193 randomized participants, 146 completed the first hand hygiene diary preintervention and were included in the main analysis. The mean age was 41 (SD 17) years, and 69.2% (n=101) were women. The main analysis revealed significant superiority of the intervention compared to controls in the change in hand hygiene pre-post intervention (W=2034; $P<.04$; effect size $r=0.17$) and between preintervention and follow-up (W=2005; $P<.03$; effect size $r=0.18$). Regarding behavioral determinants, the change in coping planning pre-post intervention (W=3840; $P=.03$, effect size $r=0.16$) was significantly greater in the intervention group using Soapp+ compared to controls.

Conclusions: Soapp+ was developed through a rigorous experimental method during the ongoing COVID-19 pandemic. The RCT provided evidence for the efficacy of Soapp+ to promote hand hygiene in the context of a pandemic.

Trial Registration: ClinicalTrials.gov NCT04830761; https://clinicaltrials.gov/study/NCT04830761

(JMIR Mhealth Uhealth 2025;13:e57191) doi:10.2196/57191

KEYWORDS

COVID-19; hand hygiene; behavior change technique; Multiphase Optimization Strategy; randomized controlled trial; smartphone apps; mobile phones

XSL•FO
RenderX

Introduction

Background

The COVID-19 pandemic represented an exceptional threat to human health worldwide. In that context, the collective adoption of protective behaviors (eg, hand hygiene, physical distancing, mask wearing) became an essential measure to prevent the transmission of the virus, especially during the pandemic outbreak when there was no vaccine available [1,2]. Consequently, the development of effective behavior change techniques aiming at supporting the uptake and adoption of such behaviors emerged as a public health priority [3]. However, during the course of the outbreak, there was limited or no contextualized knowledge about how to effectively promote protective behaviors in the general population [4,5].

During the early stages of the pandemic, we addressed the call for applying principles of behavior change to reduce the transmission of COVID-19 [2,3] and devised a Multiphase Optimization Strategy (MOST) [6] to develop, optimize, and evaluate a smartphone-based behavior change technique to promote hand hygiene during the COVID-19 pandemic [7,8]. We opted for a digital intervention to avoid in-person interaction, which is crucial for preventing virus transmission. Additionally, digital interventions can be tailored to individual needs and have the potential to reach a large number of users. The MOST framework was deemed highly suitable for rigorously contextualizing the intervention to the pandemic through optimization. We targeted correct hand hygiene (ie, hand washing or sanitizing) at key times because it represents an effective strategy in preventing the transmission of respiratory illnesses, including COVID-19 [9,10], and was therefore included in public health guidelines [11].

Following the MOST framework [12], the project included 3 phases: preparation, optimization, and evaluation. In the preparation phase [7], we developed a set of intervention modules and defined the optimization criteria. The development of intervention modules was guided by the Theoretical Domain Framework [13]. First, we identified behavioral determinants of hand hygiene through a literature review, and through 2 focus group discussions with the target population [7]. Then, we mapped behavior change techniques [14] to the behavioral determinants. Specifically, three intervention modules were developed targeting (1) motivation, (2) habit formation, and (3) social influence. The identified modules were paired and sequenced, resulting in 9 different intervention conditions. Each intervention condition was hosted in the developed smartphone app Soapp. During the subsequent optimization phase, we devised a mixed method double-blind parallel randomized trial between March 2021 and August 2021, which lasted 34 days. The trial aimed at identifying the intervention condition that met our optimization criteria: superiority regarding change in hand hygiene pre-post intervention, and regarding user satisfaction, usability, and engagement [8]. The results of the trial showed that hand hygiene improved while using the Soapp app. However, no superiority was found for any of the versions of the app. The qualitative results indicated that participants preferred the motivation and habit formation modules over the social influence ones. A further emerging theme pointed out the need for better distribution of intervention content over time. Thus, in the optimized version of the Soapp app, named Soapp+, we selected the motivation and habit formation modules and delivered them in a parallel fashion rather than sequentially.

Soapp+ comprises behavior change techniques aimed at targeting key motivational and volitional behavioral determinants including intention, attitudes toward hand hygiene, risk perception, outcome expectancies, self-efficacy, action planning, coping planning, action control, and habit. These determinants represent fundamental psychological constructs in prominent behavior change theories [15], addressing both reflective and automatic processes. Reflective processes are deliberate and conscious, involving setting intended goals, planning, and monitoring progress [16,17]. In contrast, automatic processes, such as habit, operate unconsciously, enabling individuals to perform behaviors without deliberate thought once learned through repetition. These automatic processes help sustain behavior change by reducing the cognitive load associated with decision-making and promoting consistency in behavior performance [18,19].

Aims

The current paper focuses on the evaluation phase of the MOST in which we tested the efficacy of Soapp+ (intervention group) by comparison to an active control group in a randomized controlled trial (RCT). As presented in the study protocol [7], we tested the following preregistered hypotheses

- H1: The intervention group shows a greater increase in correct hand hygiene behavior at key times at the postmeasure (H1a) and at the 6-month follow-up (H1b) than the control group.
- H2: The intervention group shows a significant increase in the targeted behavioral determinants compared with the control group at the postintervention measure (H2a) and at the 6-month follow-up (H2b). Specifically, the target behavioral determinants were intention, attitude, risk perception, outcome expectancies, self-efficacy, action planning, coping planning, action control, and habit strength.

As the secondary aim, we examined between-group differences in flu-like infection symptoms and occurrences of COVID-19 (secondary outcomes). The corresponding results are reported in Multimedia Appendix 1.

Methods

Study Design

The study design for the evaluation phase was a double-blind RCT (1:1 ratio) comparing the optimized app Soapp+ with an active control group using a simplified version of the app that focused on providing information on hand hygiene. The trial was carried out from March 9, 2022, to April 18, 2023, a few days before the World Health Organization declared the end of the global health emergency [20].

XSL•FO
RenderX

Participants

The target population was German-speaking adults from the Swiss population interested in using an app to improve hand hygiene. Inclusion criteria were (1) being at least 18 years of age, (2) owning a smartphone with mobile access to the internet, (3) being proficient in the German language, and (4) having signed an informed consent form to participate in the study. Additionally, smartphone proficiency was an implicit eligibility criterion, as participants needed to perform various tasks on their smartphones prior to the study's commencement (eg, downloading the app and registering). As per protocol [7], the initial target sample size was 205 participants. The sample size was calculated a priori for an independent sample 1-tailed *t* test (β=.80; α=.05; Cohen *d*=0.35). The target sample size was raised to 245 participants to account for a 20% dropout rate.

Measures

Primary Outcome

The frequency of correct hand hygiene at key times was assessed using ecological momentary assessment with an electronic diary embedded in the study apps (ie, Soapp+ and active control apps). This approach was used to avoid retrospective bias in reporting hand hygiene [21]. On study days 2, 8, 16, 24, and 32, participants were prompted 5 times per day to indicate whether each of the 13 key times to perform hand hygiene defined by the Swiss Federal Office of Public Health occurred (eg, arriving home, after using the toilet; Table S2 in Multimedia Appendix 1). For each key situation that occurred, participants were asked how often they correctly washed or disinfected their hands. The 5-point response scale ranged from 0=never to 4=always. For each diary day, the hand hygiene score was calculated by (1) computing the mean frequency of correct hand hygiene across all indicated key times in a single diary and then (2) averaging the hand hygiene frequency across the five diaries.

Secondary Outcomes

Targeted Behavioral Determinants

The following behavioral determinants were assessed at T1, T2, and the follow-up. These variables were selected based on the results of the optimization phase and represented the determinants targeted by the Soapp+ intervention.

Behavioral Intention

One self-reported item was adapted from a study by Allan, Sniehotta, and Johnston [22] asked "To what extent do you intend to correctly perform your hand hygiene behavior at key times?" The response options ranged from 1=not at all to 6=very strongly.

Action Planning

This was assessed using the mean score of 3 self-reported items [23], such as "I have made a detailed plan about when I am going to wash and disinfect my hands." The response options ranged from 1=not at all to 6=very strongly. The scale was reliable (Cronbach α=0.90 at T1).

Coping Planning

This was assessed using the mean score of 4 self-reported items, such as "I have made a detailed plan about how I can perform correct hand hygiene behavior at key times when soap and water are not available." The response options ranged from 1=not at all to 6=very strongly (Cronbach α=0.84 at T1).

Habit Strength

This was assessed with the "Self-Report Behavioural Automaticity Index" [24] with 4 items, such as "Correct hand hygiene behavior at key times is something that I do automatically." The response options ranged from 1=strongly disagree to 7=strongly agree (Cronbach α=0.93 at T1).

Action Control

This was assessed using the mean score of 3 self-reported items, such as "During the last two weeks I have constantly monitored myself whether I washed or disinfected my hands according to my plans." The response options were adapted from a study by Sniehotta, Scholz, and Schwarzer [25] and ranged from 1=not at all to 6=very strongly (Cronbach α=0.91 at T1).

Self-Efficacy

This was assessed using the mean score of 8 self-reported items, such as "How confident are you that you correctly wash/disinfect your hands at key times, when you are on the way." The response options ranged from 1=not at all to 6=very strongly. Self-efficacy items were adapted from a study by Schwarzer [23] (Cronbach α=0.91 at T1).

Attitude

This was assessed using the mean score of 6 self-reported bipolar items [26] introduced by the wording "Correct hand hygiene behavior at key times is…" and followed by pairs of bipolar adjectives (eg, useless vs useful; Cronbach α=0.81 at T1).

Outcome Expectancies

This was assessed using the mean score of 8 self-reported items [27], such as "If I engage in correct hand hygiene, then then the risk drops that I contract pathogens." The response options ranged from 1=this doesn't apply at all to 6=this applies very precisely (Cronbach α=0.72 at T1).

Risk Perception

This was assessed using the mean score of 5 self-reported items [27], such as "to what extent do you think that you could contract corona, if you don't engage in correct hand hygiene in key situations?" The response options ranged from 1=this doesn't apply at all to 6=this applies very precisely (Cronbach α=0.90 at T1).

Flu-Like Infection Symptoms and Occurrences of COVID-19

Participants reported on these secondary outcomes by answering the items "In the last two weeks, did you have flu-like infection symptoms?" and "In the last two weeks, did you have a positive corona test?"

Finally, social desirability was assessed at T1 using the German Short Scale for the two-dimensional measurement of social desirability [28].

XSL•FO
RenderX

Procedure

Participants were recruited via social media (eg, Facebook, Instagram), mailing lists, and leaflets with the support of a market research company. We aimed at recruiting a broad range of people from the German-speaking adult Swiss general population in terms of sex, age, and socioeconomic status. The content of the recruitment materials targeted individuals interested in forming new hand-hygiene habits and learning how to correctly wash or disinfect their hands to counter the spread of the virus. Interested people were redirected to the study page on RedCap (Vanderbilt University) [29] where they could go through the study information, fill out an eligibility and consent survey, and sign the e-consent form. Afterward, participants received a registration code via email and were guided to download fthe Soapp+ app from the Apple App Store or Google Play Store and register it. The day after the registration, participants received the baseline questionnaire (T1) and were then randomized to an intervention or active control group. A simple randomization was implemented in Qualtrics, which preserved the allocation concealment. Additionally, the researchers involved in the study were blinded to intervention assignment because the participant identifier was pseudoanonymized before randomization (Figure S1 in Multimedia Appendix 1). During the intervention phase, participants completed the hand hygiene diary on days 2, 8, 16, 24, and 32. At the end of the intervention phase (day 34), participants received a second questionnaire (T2). At 6 months (182 days) after T1, participants received the last follow-up questionnaire. Questionnaires and diaries were integrated into Qualtrics services using Soapp+'s application programming interface, and the participants' data were stored on Qualtrics. The recruitment for the evaluation trial began on March 9, 2022, and ended on October 20, 2022. Follow-up data were collected between September 5, 2022, and April 18, 2023.

Intervention

The content of the intervention was delivered to participants via their personal smartphone through the study app Soapp+. The intervention group received the optimized Soapp+ app while the active control group was exposed to a simplified version of Soapp+. Table 1 shows the content delivered. The exact wording of the intervention messages and the intervention timeline can be found in Table S3 and Figure S2 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Intervention content overview.

Condition and module	TDF[a] domain	Behavioral determinant	Behavior change technique
Intervention and active control groups			
Basic	Goals	Intention	1.1 Goal setting (behavior)
Basic	Skills	Skills	4.1 Instruction on how to perform behavior
Basic	Knowledge	Knowledge	5.1 Information about health consequences
Active control group			
Quizzes	Knowledge	Knowledge	5.1 Information about health consequences
Quizzes	Skills	Skills	4.1 Instruction on how to perform behavior
Intervention group			
Motivation	Goals	Intention	1.1 Goal setting (behavior)
Motivation	Beliefs about consequences	Risk perception	5.1 Information about health consequences
Motivation	Beliefs about consequences	Attitude	5.2 Salience of consequences
Motivation	Beliefs about consequences	Outcome expectancies	9.2 Pros and cons
Motivation	Beliefs about consequences	Intention	5.2 Salience of consequences
Motivation	Beliefs about capabilities	Self-efficacy	1.2 Problem-solving 15.1 Verbal persuasion about capabilities 15.3 Focus on past success
Motivation	Beliefs about capabilities	Coping planning	1.2 Problem solving
Motivation	Reinforcement	Intention	10.9 Self-reward
Habit	Knowledge	Knowledge	4.2. Information about antecedents
Habit	Memory, attention, and decision processes	Action control	2.3 Self-monitoring of behavior
Habit	Goals	Action planning	1.4 Action planning 7.1. Prompts or cues
Habit	Skills and goals	Habit strength	8.1 Behavioral practice or rehearsal 8.3 Habit formation
Habit	Behavioral regulation	Habit strength	7.1 Prompts or cues (physical cue)

[a]TDF: theoretical domains framework [29].

Active Control Group

The active control group received information on how to perform correct hand hygiene (both in written and video formats), the key times to perform hand hygiene, and a statement of the goal of the app (basic module). This content was accessible throughout the intervention phase. Based on the results of the optimization phase, we also added a list of fun facts that were delivered to the participants via push notifications throughout the intervention phase. Finally, quizzes on hand hygiene aimed at ensuring that the active control group interacted with the app throughout the intervention phase.

Intervention Group

Like the control group, the intervention group received the basic module. They additionally received the content targeting motivation and habit. Based on the results of the optimization phase (ie, request for better-distributed content or tasks), this content was delivered in parallel instead of sequentially as done during the optimization trial.

Data Analysis

Handling of Missing Data

To handle missing data, the intention-to-treat (ITT) principle was adopted [30]. According to the ITT principle, all the randomized participants are included in the analysis, regardless of what happens after randomization (eg, noncompliance, protocol deviations). Therefore, the ITT approach avoids overoptimistic estimates of the efficacy of an intervention resulting from the removal of noncompliers. Missing data and dropouts were addressed using the last observation carried forward approach. The last observation carried forward approach represents a parsimonious yet effective method to maintain appropriate type 1 error protection [31]. This approach has been used in previous behavior change research [32], including the optimization trial of the *Soapp* app [8]. This methodological choice was reinforced by the results of a secondary analysis conducted on the optimization trial [33]. The analysis revealed a noteworthy intraclass correlation of 0.60 in hand hygiene, signifying that a substantial portion of the observed behavior

XSL•FO
RenderX

can be attributed to individual differences between persons. This implies that the behavior tends to be more consistent or less variable within individuals over time. Finally, we also conducted a set of sensitivity analyses without missing value imputation; results are available in Multimedia Appendix 1.

Hypothesis Testing

In order to test the hypotheses, the assessment points considered for hand hygiene behavior were the first diary on day 2 (T1), the diary day on day 32 (T2, postintervention), and the last diary day on day 181 (follow-up). The difference in hand hygiene between T2 and T1 and follow-up and T1 was compared between the intervention and the active control groups using the nonparametric Wilcoxon rank sum test. A positive value in hand hygiene difference over assessments indicated an increase in hand hygiene. Likewise, positive values in the targeted behavioral determinants over assessments indicated an increasing trend. The Wilcoxon rank sum test was chosen over the preregistered 1-tailed Student *t* test due to the nonnormal distribution of the data, as assessed by the Shapiro-Wilk normality test, representing a deviation from the registered protocol. Results of the analysis with the Student *t* test are available in Table S4 in Multimedia Appendix 1.

Analytical Software

The Soapp+ app was developed by the study authors using technological resources provided by the Faculty of Human Sciences at the University of Bern, with support from an internal software engineer. The statistical software R (version 4.3.0; R Foundation for Statistical Computing) was used to process the data and run the analyses. The data and R code used for the main analyses are available on the Open Science Framework repository platform [34].

Ethical Considerations

The trial received ethical approval from the Swiss Ethics Committee of the canton of Bern (ID 2021-00164). The reporting of the trial is in line with the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Multimedia Appendix 2) [35]. Informed consent was obtained from all participants, who were also informed of their right to opt out at any stage of the study. To protect privacy and confidentiality, data were deidentified through pseudonymization. As compensation, participants had the opportunity to win one of three iPhone 12s as an incentive for their participation. Additionally, after completing T2, participants were offered a small gift (ie, a bar of hand soap and a thank you card) in order to prevent dropout, which was sent to their homes.

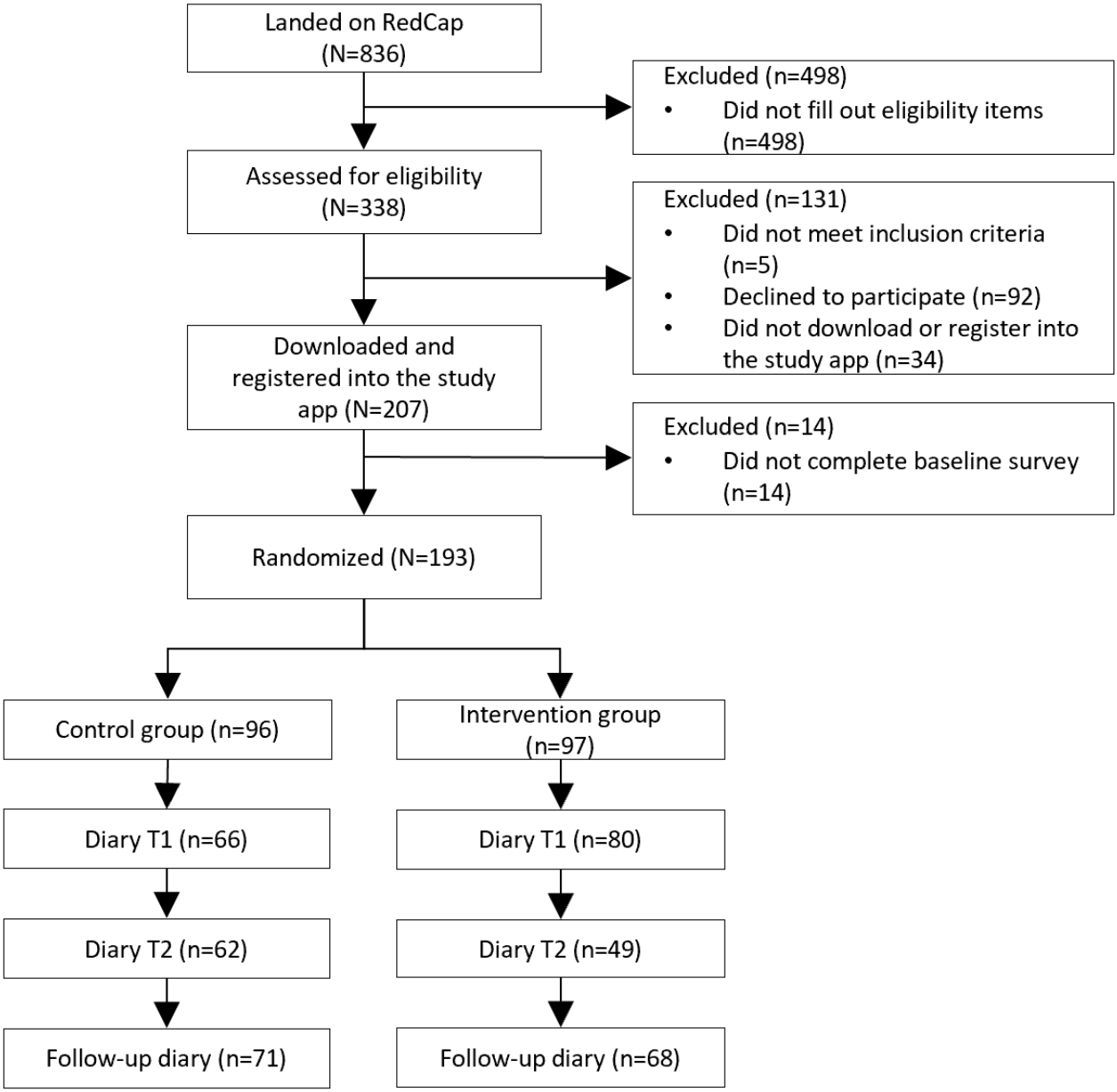
Results

Overview

We stopped the trial 13 months after the start of the study due to the end of the project timeline. A total of 193 participants were recruited and randomized into one of the 2 experimental conditions. Among these, 12 participants did not complete any of the 6 hand hygiene diary days, while the other 35 participants did not fill out the first diary at T1. Further, 3 participants completed the first diary but did not encounter any of the key situations to perform hand hygiene during that day. These participants (n=50) were excluded from the analysis because the main outcome (ie, hand hygiene) at T1 was missing. Out of the 193 participants who were randomized, 146 (75.6%) filled out the hand hygiene diary at T1, 91 of these (47.2% of the randomized participants) completed the hand hygiene diary at T2, and 109 (56.5% of the randomized participants) completed the hand hygiene diary at the 6-month follow-up. Figure 1 shows the participants' flow through randomization, T1 diary assessment, T3 diary assessment, and follow-up for each intervention group.

XSL•FO
RenderX

Figure 1. Participant recruitment flow.



Baseline Characteristics

Sociodemographics and hand hygiene behavior at baseline are reported in Table 2 (see Table S1 in Multimedia Appendix 1 for a comparison with the sociodemographic characteristics of the Swiss population). The figures refer to the 146 participants who completed the first diary at T1. Participants' mean age was 41 (SD 17) years, 69.2% (n=101) were women, 69.9% (n=102) had high school qualifications, 45.9% (n=67) were employed, and 24% (n=35) were living alone. Descriptive statistics in hand hygiene behavior (mean 3.05, SD 0.78; median 3.25, IQR 2.63-3.60; skewness=–1.24) suggested that hand hygiene behavior was already high at baseline and characterized by a moderate left-tailed distribution. No significant differences in social desirability were found between the intervention and the control group (M_1=4.72; M_2=4.79; $F_{1,189}$=0.44; P=.51).

XSL•FO
RenderX

Table 2. Sociodemographic characteristics at baseline (T1; N=146).

Variable	Overall	Experimental group		*P* value[a]
		Control (n=66)	Intervention n=80)	
Age (years), mean (SD)	41 (17)	42 (17)	40 (17)	.56
Sex, n (%)				.49
Diverse	3 (2.1)	1 (1.5)	2 (2.5)	
Male	42 (28.8)	16 (24)	26 (33)	
Female	101 (69.2)	49 (74)	52 (65)	
Marital status, n (%)				.01
Divorced	12 (8.2)	8 (12)	4 (5)	
Single	88 (60.3)	32 (48)	56 (70)	
Married	44 (30.1)	26 (39)	18 (23)	
Widowed	2 (1.4)	0 (0)	2 (2.5)	
Have children, n (%)				.047
Yes	43 (29.5)	25 (38)	18 (23)	
No	103 (70.5)	41 (62)	62 (78)	
Education, n (%)				.15
High school diploma	102 (69.9)	42 (64)	60 (75)	
Completed primary school	2 (1.4)	2 (3)	0 (0)	
Completed secondary school	42 (28.8)	22 (33)	20 (25)	
Higher education, n (%)				.20
Completed vocational training (apprenticeship)	35 (24)	21 (32)	14 (18)	
Completed (applied) university degree	59 (40.4)	26 (39)	33 (41)	
Other educational or vocational qualification	14 (9.6)	5 (7.6)	9 (11)	
No educational or vocational training	38 (26)	14 (21)	24 (30)	
Profession, n (%)				.59
Unemployed	4 (2.7)	1 (1.5)	3 (3.8)	
Employed	67 (45.9)	35 (53)	32 (40)	
Homemaker	8 (5.5)	4 (6.1)	4 (5)	
In training or retraining	2 (1.4)	1 (1.5)	1 (1.3)	
Retired	18 (12.3)	8 (12)	10 (13)	
Student	47 (32.2)	17 (26)	30 (38)	
Monthly income[b], n (%)				.25
Less than 2000 Fr	11 (7.5)	3 (4.5)	8 (10)	
Between 2000 and 4000 Fr	17 (11.6)	12 (18)	5 (6.3)	
Between 4001 and 6000 Fr	29 (19.9)	13 (20)	16 (20)	
Between 6001 and 8000 Fr	23 (15.8)	7 (11)	16 (20)	
Between 8001 and 10,000 Fr	23 (15.8)	11 (17)	12 (15)	
More than 10,000 Fr	26 (17.8)	12 (18)	14 (18)	
I don’t know	17 (11.6)	8 (12)	9 (11)	
Living situation, n (%)				.42
Living by themselves	35 (24)	15 (23)	20 (25)	
Living with family	67 (45.9)	35 (53)	32 (40)	

XSL•FO
RenderX

Variable	Overall	Experimental group		*P* value[a]
		Control (n=66)	Intervention n=80)	
Living in a shared apartment	33 (22.6)	12 (18)	21 (26)	
Other	11 (7.5)	4 (6.1)	7 (8.8)	

[a]Welch 2-sided *t* test or Fisher exact test.

[b]Swiss francs. When the study started in March 2022, the average conversion rate between Swiss francs and US dollars was 0.93, meaning that US $1 was equal to 0.93 Swiss francs

Dropout Analysis

Dropout analysis was performed to investigate baseline differences between participants who reached the end of the intervention phase and those who dropped out at any point during the intervention. We analyzed all 193 randomized participants and the ones who did not complete the T2 panel assessment were categorized as dropouts (n=120). The results suggested that participants who dropped out were significantly younger than retainers (M_1=39; M_2=46; $F_{1, 191}$=7.06; *P*=.009). Dropouts and retainers did not differ regarding sex (male vs female; χ^2_1=1.72; *P*=.19), hand hygiene ($F_{1,190}$=0.12; *P*=.73), intention to increase hand hygiene behavior ($F_{1,191}$=2.36; *P*=.13), or intervention group allocation (χ^2_1=2.04; *P*=.15).

Hand Hygiene Behavior

Hand hygiene behavior at T1, T2, and follow-up is summarized in Table 3. The whole sample exhibited a nonsignificant increase in hand hygiene behavior from T1 (median 3.25, IQR 2.63-3.60) to T2 (median 3.42, IQR 2.62-3.88; *W*=8904; *P*=.06, *r*=0.11). This was followed by a significant decrease between T2 and the 6-month follow-up (median 3.25, IQR 2.50-3.65; *W*=11,655; *P*=.04; *r*=0.12). When looking at changes in hand hygiene within each experimental group, the data shows that hand hygiene between T1 and T2 remained almost stable for the active control group (*W*=2050; *P*=.77; *r*=0.03), while the intervention group showed a significant increase in hand hygiene after the intervention phase (*W*=2386; *P*=.02; *r*=0.19). At follow-up, both the active control group (*W*=2313; *P*=.35; *r*=0.08) and the intervention group (*W*=2924; *P*=.68; *r*=0.03) exhibited a nonsignificant decrease in hand hygiene compared to T1.

The between-group differences in hand hygiene behavior between T1 and T2 and T1 and follow-up are reported in Table 4. Nonparametric analysis with the Wilcoxon rank sum test showed that the change in hand hygiene behavior between T1 and T2 (*W*=2034; *P*<.04; effect size *r*=0.17) and between T1 and follow-up (*W*=2005; *P*<.03; effect size *r*=0.18) significantly differed between the groups, indicating better hand hygiene in the intervention group.

XSL•FO
RenderX

Table 3. Hand hygiene and targeted behavioral determinants median values at each assessment point and Shapiro-Wilk normality test.

Variable and assessment point	Participants, n	Overall, median (IQR)	Active control group, median (IQR)	Intervention group, median (IQR)	Shapiro-Wilk normality test	
					Statistic	*P* value
Target behavior						
Hand hygiene						
T1	143	3.25 (2.63-3.60)	3.23 (2.56-3.57)	3.25 (2.71-3.61)	0.897	<.001
T2	143	3.42 (2.62-3.88)	3.29 (2.55-3.71)	3.58 (3.00-3.94)	0.852	<.001
Follow-up	143	3.25 (2.50-3.65)	3.14 (2.33-3.55)	3.29 (2.57-3.77)	0.894	<.001
Behavioral determinants						
Intention						
T1	193	5.00 (5.00-6.00)	5.00 (5.00-5.25)	5.00 (5.00-6.00)	0.823	<.001
T2	193	5.00 (5.00-6.00)	5.00 (5.00-6.00)	5.00 (5.00-6.00)	0.796	<.001
Follow-up	193	.00 (4.00-6.00)	5.00 (4.00-5.00)	5.00 (4.00-6.00)	0.841	<.001
Action planning						
T1	193	4.33 (4.00-5.00)	4.33 (3.67-5.00)	4.67 (4.00-5.00)	0.922	<.001
T2	193	4.67 (4.00-5.00)	4.67 (4.00-5.00)	5.00 (4.00-5.00)	0.890	<.001
Follow-up	193	4.67 (3.67-5.00)	4.33 (3.67-5.00)	4.67 (3.67-5.00)	0.916	<.001
Coping planning						
T1	193	3.75 (2.75-4.25)	3.50 (2.75-4.25)	3.75 (2.75-4.50)	0.974	.001
T2	193	4.00 (3.25-4.75)	3.75 (2.94-4.50)	4.25 (3.50-5.00)	0.971	<.001
Follow-up	193	4.00 (2.75-4.75)	3.75 (2.69-4.50)	4.00 (3.00-5.00)	0.966	<.001
Habit strength						
T1	187	5.25 (4.00-6.25)	5.00 (3.50-6.00)	5.75 (4.19-6.25)	0.933	<.001
T2	187	5.25 (4.25-6.00)	5.25 (4.00-6.00)	5.25 (4.50-6.25)	0.940	<.001
Follow-up	187	5.25 (4.00-6.00)	5.00 (3.75-6.00)	5.25 (4.63-6.00)	0.932	<.001
Action control						
T1	152	4.00 (3.33-5.00)	4.00 (3.58-5.00)	4.00 (3.33-5.00)	0.947	<.001
T2	152	4.67 (4.00-5.00)	4.33 (4.00-5.00)	4.67 (4.00-5.33)	0.938	<.001
Follow-up	152	4.33 (3.33-5.00)	4.33 (3.33-5.00)	4.33 (3.58-5.00)	0.955	<.001
Self-efficacy						
T1	193	4.25 (3.63-4.88)	4.13 (3.50-4.75)	4.38 (3.63-5.00)	0.986	.049
T2	193	4.38 (3.75-5.00)	4.25 (3.75-4.88)	4.38 (3.88-5.00)	0.979	.005
Follow-up	193	4.25 (3.63-4.88)	4.13 (3.59-4.88)	4.25 (3.63-4.88)	0.984	.03
Attitudes						
T1	188	5.17 (4.67-5.62)	5.17 (4.67-5.50)	5.17 (4.67-5.67)	0.932	<.001
T2	188	5.17 (4.67-5.50)	5.17 (4.50-5.50)	5.08 (4.67-5.67)	0.885	<.001
Follow-up	188	5.00 (4.33-5.50)	5.00 (4.46-5.38)	5.08 (4.33-5.67)	0.880	<.001
Outcome expectancies						
T1	193	4.63 (4.25-5.13)	4.63 (4.25-5.00)	4.75 (4.25-5.13)	0.979	.005
T2	193	4.75 (4.25-5.13)	4.75 (4.25-5.13)	4.75 (4.25-5.13)	0.987	.08
Follow-up	193	4.63 (4.13-5.13)	4.63 (4.13-5.13)	4.63 (4.25-5.13)	0.987	.07
Risk perception						
T1	193	4.60 (4.00-5.00)	4.60 (4.00-5.00)	4.60 (3.80-5.00)	0.942	<.001

XSL•FO
RenderX

Variable and assessment point	Participants, n	Overall, median (IQR)	Active control group, median (IQR)	Intervention group, median (IQR)	Shapiro-Wilk normality test	
					Statistic	*P* value
T2	193	4.60 (4.00-5.00)	4.60 (4.00-5.00)	4.60 (4.00-5.00)	0.953	<.001
Follow-up	193	4.60 (3.60-5.00)	4.60 (3.60-5.00)	4.60 (3.80-5.00)	0.950	<.001

Table 4. Differences in changes in hand hygiene and behavioral determinants between the intervention group and active control group with the Wilcoxon rank sum test.

Variables and change	Participants, n	Overall, median (IQR)	Active control group, median (IQR)	Intervention group, median (IQR)	*W* statistic	*P* value	Effect size *r*[a]
Target behavior							
Hand hygiene							
T1-T2	143	0.00 (–0.11 to 0.38)	0.00 (–0.23 to 0.35)	0.00 (0.00 to 0.41)	2034	.04	0.17
T1-Follow-up	143	0.00 (–0.41 to 0.23)	–0.04 (–0.52 to 0.05)	0.00 (–0.23 to 0.32)	2005	.03	0.18
Behavioral determinants							
Intention							
T1-T2	193	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	4560	.75	0.02
T1-Follow-up	193	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (–1.00 to 0.00)	4845	.57	0.04
Action planning							
T1-T2	193	0.00 (0.00 to 0.33)	0.00 (0.00 to 0.33)	0.00 (0.00 to 0.33)	4395	.47	0.05
T1-Follow-up	193	0.00 (–0.33 to 0.33)	0.00 (0.00 to 0.33)	0.00 (–0.33 to 0.33)	4917	.49	0.05
Coping planning							
T1-T2	193	0.00 (0.00 to 0.75)	0.00 (–0.25 to 0.50)	0.00 (0.00 to 1.00)	3840	.03	0.16
T1-Follow-up	193	0.00 (–0.25 to 0.75)	0.00 (–0.31 to 0.56)	0.00 (0.00 to 0.75)	4388	.48	0.05
Habit strength							
T1-T2	187	0.00 (0.00 to 0.25)	0.00 (–0.13 to 0.25)	0.00 (0.00 to 0.25)	4585	.54	0.04
T1-Follow-up	187	0.00 (–0.50 to 0.38)	0.00 (–0.63 to 0.38)	0.00 (–0.31 to 0.31)	4469	.79	0.02
Action control							
T1-T2	152	0.00 (0.00 to 0.67)	0.00 (0.00 to 0.67)	0.00 (0.00 to 0.42)	2996	.68	0.03
T1-Follow-up	152	0.00 (–0.33 to 0.33)	0.00 (–0.33 to 0.33)	0.00 (–0.33 to 0.08)	3075	.48	0.06
Self-efficacy							
T1-T2	193	0.00 (0.00 to 0.25)	0.00 (0.00 to 0.25)	0.00 (0.00 to 0.25)	4824	.65	0.03
T1-Follow-up	193	0.00 (–0.38 to 0.25)	0.00 (–0.25 to 0.38)	0.00 (–0.38 to 0.25)	4881	.56	0.04
Attitudes							
T1-T2	188	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.17)	0.00 (0.00 to 0.00)	4650	.50	0.05
T1-Follow-up	188	0.00 (–0.33 to 0.17)	0.00 (–0.25 to 0.17)	0.00 (–0.33 to 0.00)	4747	.37	0.07
Outcome expectancies							
T1-T2	193	0.00 (0.00 to 0.25)	0.00 (0.00 to 0.25)	0.00 (0.00 to 0.13)	5059	.28	0.08
T1-Follow-up	193	0.00 (–0.25 to 0.13)	0.00 (–0.25 to 0.25)	0.00 (–0.25 to 0.13)	5125	.22	0.09
Risk perception							
T1-T2	193	0.00 (–0.20 to 0.00)	0.00 (–0.20 to 0.00)	0.00 (0.00 to 0.00)	4300	.33	0.07
T1-Follow-up	193	0.00 (–0.40 to 0.20)	0.00 (–0.60 to 0.20)	0.00 (–0.20 to 0.40)	3971	.07	0.13

[a]The following guidelines are used to interpret the effect size (*r*): a large effect is defined as $r \geq 0.50$, a medium effect as approximately $r=0.30$, and a small effect as approximately $r=0.10$ [36].

XSL•FO
RenderX

Targeted Behavioral Determinants

Behavioral determinants values at T1, T2, and follow-up are summarized in Table 3, while results from the paired Wilcoxon tests between T1 and T2 and T1 and follow-up are reported in Table S5 in Multimedia Appendix 1. Shapiro-Wilk normality test showed that almost all the behavioral determinants were not normally distributed at all assessment points, with the exception of outcome expectancies at T2 and follow-up (Table 3). The variables that increased the most in the whole sample between T1 and T2 were coping planning (W=1474; P<.001) and action control (W=611; P<.001), followed by action planning (W=1338; P=.006) and self-efficacy (W=2115; P=.01). The change in behavioral determinants between T1 and follow-up was less pronounced and sometimes negative (ie, intention, self-efficacy, attitudes, outcome expectancies, and risk perception). The between-group differences in the targeted behavioral determinants between T1 and T2 and T1 and follow-up are reported in Table 4. Results showed that the change in coping planning between T1 and T2 (W=3840;P=.03; effect size r=0.16) significantly differed between the intervention groups.

Sensitivity Analysis

All the hypotheses were also tested without applying any missing value imputation algorithm. Results are available in Tables S6 and S7 in Multimedia Appendix 1 and confirm most of the findings. Additionally, Wilcoxon tests without missing values imputation showed between-group differences in coping planning between T1 and T2, and in risk perception between T1 and follow-up in favor of the intervention group.

Discussion

Efficacy of the Soapp+ App

As part of a MOST to develop, optimize, and test a smartphone-based hand hygiene intervention during the COVID-19 pandemic, this RCT aimed at evaluating the efficacy of the Soapp+ app intervention against an active control group. Results showed that hand hygiene can be increased short and long term using the Soapp+ app intervention compared to active controls toward the end of the COVID-19 pandemic. Specifically, we observed small but stable effect sizes (r=0.17 between pre-post intervention and r=0.18 between preintervention and follow-up) for an increase in hand hygiene among the intervention group compared to the control group. The results of this study contribute to the notion that simultaneously targeting both reflective and automatic processes is an effective strategy for promoting and sustaining behavior change, even in challenging contexts such as a pandemic.

Discussing the clinical relevance of the effect size is challenging due to the limited evidence available regarding the necessary target effect size for a hand hygiene intervention to prevent the acquisition or transmission of COVID-19 [37]. However, we can note that the effect sizes we observed were smaller than the average effect size typically reported in previous interventions, primarily RCTs, aimed at promoting hand hygiene to mitigate the transmission of respiratory viruses (Cohen d=0.62) [4]. One plausible reason for this difference is that the study's target population consists of individuals who are already inclined toward hand hygiene behavior. This is evident from the fact that the frequency of hand hygiene in key situations was already relatively high in our sample at baseline, with median values of 3.25 and 3.23 out of 4 for the intervention and active control group, respectively, where a value of 3 indicates that hand hygiene was performed "often." In spite of such a ceiling effect, participants in the intervention group still managed to increase the frequency of correct hand hygiene compared to those in the active control group.

Additionally, even though between-group differences in hand hygiene were small, we saw a trend that they resulted in fewer self-reported flu-like symptoms at postintervention and follow-up for the intervention group (Table S8 in Multimedia Appendix 1). Although this difference was not statistically significant due to the small sample size and claims about the clinical significance of the intervention cannot be made, it does indicate a tendency toward an intervention effect on flu-like symptoms. Finally, it should be noted that digital interventions like Soapp+ offer the advantage of scalability and the potential to reach a broader segment of the population. This is especially relevant in a pandemic context, where the diffusion of the virus's transmission grows exponentially, and even small changes in individual behavior, enacted on a larger scale, have the potential to significantly contribute to countering the pandemic's spread at the community level.

Targeted Behavioral Determinants

Regarding our second hypothesis, results evidenced significant group differences pre-post intervention in coping planning in favor of the intervention group (Table 4). The fact that coping planning was the only behavioral determinant associated with group differences pre-post intervention aligns well with the intervention content targeting coping planning skills (ie, problem-solving) delivered toward the end of the intervention and close to the postintervention assessment (Multimedia Appendix 1). A possible explanation is that coping planning was a determinant with a low baseline value. Therefore, the intervention content might have played a significant role on this variable as there were more margins for improvement compared to the other targeted determinants.

Strengths and Limitations

To the best of our knowledge (based on a search on ClinicalTrials.gov with the following key terms: condition or disease "COVID 19," age "Adult (18-64)," study type "Interventional," outcome measure "Hand hygiene," this study is one of the few registered RCTs testing the effectiveness of a behavior change technique targeting hand hygiene in the general population during the COVID-19 pandemic. Additionally, this RCT study is the third of a series where we have used the MOST approach for intervention development. This feature of our project stands as a significant strength, given that the MOST places strong emphasis on a rigorous, evidence-based approach to developing, refining, and evaluating interventions. Moreover, the mixed but overall encouraging evidence for the efficacy of Soapp+ points to the potential value of this app as an effective tool to address the next pandemic. This aspect is of primary relevance, as before the pandemic, there was only limited or no

XSL•FO
RenderX

contextualized knowledge about how to effectively promote hand hygiene during an outbreak [3,4]. In case of future epidemics, Soapp+ is a contextualized intervention that is readily available to be tested at a larger scale.

This study is not without limitations. First, the achieved sample size (N=143) was smaller than planned. Therefore, the probability of detecting a true effect of the intervention was below standard (ie, .80) and some small effect sizes might have been missed. The decision to stop the recruitment even though the target sample size was not reached was due to the project timeline [7], but also considering that the pandemic situation was developing toward better scenarios to the point where the World Health Organization declared the end of the global health emergency [20]. Based on the found effect size, future efficacy trials should be powered to detect small effects.

A second limitation is related to the dropouts between randomization and the first diary assessment. Indeed, the first diary was scheduled for the day after the randomization, and some participants who had been randomized (n=47) did not complete it. Other participants dropped out through the course of the study without filling out the assessment at T2 and follow-up. A possible explanation for the attrition might be due to the longitudinal study design with multiple diaries and other daily tasks that might have generated some fatigue. To address potential differences between dropouts and retainers, we conducted a dropout analysis, indicating a higher likelihood of younger participants dropping out. Thus, when extending these results to young adults, caution is advised.

Another limitation is the high prevalence of women in the trial (69%), resulting in a sex ratio that does not mirror the target population. This ratio aligns with the sample characteristics of the optimization trial [8] and other COVID-19 hand hygiene research [38]. This imbalance can be explained in light of empirical evidence suggesting that women were more worried about the COVID-19 pandemic and keener to adopt hand hygiene behavior [39]. Finally, the primary outcome of the study was self-reported, which can be biased by social desirability and recall issues [21]. The latter were minimized using the electronic hand hygiene diary. Nevertheless, social desirability might still have played a role.

Implications, Challenges, and Considerations

A key practical implication is that the Soapp+ app, a behavior change technique contextualized to a pandemic, is now available and can be readily offered to the general public at the emergence of a future infectious disease outbreak (eg, via public health institutions). However, since the app's efficacy has not yet been robustly demonstrated due to sample size, we recommend conducting a larger-scale RCT in a future outbreak. It is also important to note that testing behavior change techniques like Soapp+ during a pandemic remains challenging due to the constantly evolving context, including shifts in the pandemic trajectory, public health policies, government regulations, and medical advancements such as vaccine availability. These multifaceted and evolving factors likely contributed to the observed temporal variations in the adoption of hand hygiene behaviors, with implications for the evaluation of behavior change techniques [33,38]. For this reason, we believe that further testing of Soapp+ could extend beyond pandemic contexts and be conducted in more feasible settings, provided that the behavior change techniques specifically referring to the COVID-19 pandemic (eg, the ones targeting risk perception and outcome expectancies; Table S3 in Multimedia Appendix 1) are adapted to the specific characteristics of the new context. Upon confirmation of the current positive effect of Soapp+ on hand hygiene, this further testing beyond the pandemic context would also pave the way for a broader application of Soapp+, particularly in settings where there is a demand for hand hygiene interventions (eg, health care settings, during regular flu seasons, and to prevent the transmission of other viruses).

A methodological lesson learned regards the suitability of the MOST as a reference framework for intervention development in a pandemic context. The sequential decision-making process unique to the MOST may be too rigid for dealing with a continuously changing context. Indeed, the contextual conditions under which the intervention is developed may differ from those in which the intervention is later optimized and evaluated. Therefore, we foresee a validity risk when using the results from one phase to inform the implementation of the upcoming phase. Additionally, the pandemic context has amplified the already known challenges in completing the full MOST cycle—preparation, optimization, and evaluation—within the project timeline [40], as evidenced by the slow recruitment pace and the high dropout rates we saw in this study.

One strategy to enhance the feasibility of RCTs in future pandemics is to join forces and contribute to large-scale multicenter trials [41]. This approach was successfully adopted by observational studies investigating protective behaviors during the COVID-19 pandemic [38,42] but was less commonly used in behavior change techniques for protective behaviors, which saw fewer collaborative efforts [43]. Another strategy is represented by adaptive clinical trials [44], whose adaptive design aims to decrease the time needed to complete a trial and increase the likelihood of detecting intervention effects. Specifically, they allow for interim analysis, which is used to enhance the efficiency of the trials by adapting the original study design. Examples include response-adaptive randomization (ie, decreasing the allocation ratio to less promising arms), sample size reassessment (ie, redefining the target sample size), or seamless trials (ie, discontinuing the most unpromising arms of the trial).

Conclusions

This study described the evaluation phase of Soapp+, a smartphone app to promote hand hygiene in the context of the COVID-19 pandemic. We leveraged digital technology and the MOST approach to address the call raised by public health experts for developing evidence-based behavior change techniques that are built to be used in a pandemic context [3]. Our findings provided tentative support for the efficacy of Soapp+, and for targeting reflective and automatic behavior change processes more broadly, in promoting hand hygiene during an ongoing pandemic. Beyond our individual results, the trial has provided important learnings for conducting rigorous trials of contextualized behavior change techniques

XSL•FO
RenderX

during a pandemic, which can be further developed in the next outbreak.

Acknowledgments

The authors want to thank the technology platform of the Faculty of Human Sciences, especially Stefan Kodzhabashev and Gian Giacomo Ruschetti for supporting the development of the Soapp+ app. The authors also want to thank Annalea Venzin and Nadine Bürgi for supporting the recruitment of participants. The funding for this study was provided by Ursula-Wirz Stiftung.

Authors' Contributions

DB conceptualized the study, provided resources, curated the data, contributed to the methodology, conducted the formal analysis, and prepared the original draft. CLR assisted with data curation and resource provision. MAA contributed to conceptualization, methodology, and resource provision. JI secured funding, contributed to conceptualization, methodology, and formal analysis, supervised the research, and contributed to the original draft. All authors participated in writing, reviewing, and editing the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental materials.
[DOCX File , 121 KB - mhealth_v13i1e57191_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1077 KB - mhealth_v13i1e57191_app2.pdf]

References

1. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. Nat Hum Behav 2021;5(7):947-953. [doi: 10.1038/s41562-021-01122-8] [Medline: 33972767]
2. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav 2020;4(5):451-459. [doi: 10.1038/s41562-020-0887-9] [Medline: 32377018]
3. Michie S, West R. Behavioural, environmental, social, and systems interventions against COVID-19. BMJ 2020;370:m2982. [doi: 10.1136/bmj.m2982] [Medline: 32723732]
4. Perski O, Szinay D, Corker E, Shahab L, West R, Michie S. Interventions to increase personal protective behaviours to limit the spread of respiratory viruses: a rapid evidence review and meta-analysis. Br J Health Psychol 2022;27(1):215-264. [doi: 10.1111/bjhp.12542] [Medline: 34173697]
5. White S, Thorseth AH, Dreibelbis R, Curtis V. The determinants of handwashing behaviour in domestic settings: an integrative systematic review. Int J Hyg Environ Health 2020;227:113512 [FREE Full text] [doi: 10.1016/j.ijheh.2020.113512] [Medline: 32220763]
6. Collins LM, Murphy SA, Strecher V. The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007;32:S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
7. Amrein MA, Ruschetti GG, Baeder C, Bamert M, Inauen J. Mobile intervention to promote correct hand hygiene at key times to prevent COVID-19 in the Swiss adult general population: study protocol of a Multiphase Optimisation Strategy. BMJ Open 2022;12(3):e055971 [FREE Full text] [doi: 10.1136/bmjopen-2021-055971] [Medline: 35351716]
8. Baretta D, Amrein MA, Bäder C, Ruschetti GG, Rüttimann C, Del Rio Carral M, et al. Promoting hand hygiene during the COVID-19 pandemic: parallel randomized trial for the optimization of the Soapp app. JMIR Mhealth Uhealth 2023;11:e43241 [FREE Full text] [doi: 10.2196/43241] [Medline: 36599056]
9. Aiello AE, Coulborn RM, Perez V, Larson EL. Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis. Am J Public Health 2008;98(8):1372-1381. [doi: 10.2105/AJPH.2007.124610] [Medline: 18556606]
10. Beale S, Johnson AM, Zambon M, Watch Group F, Hayward AC, Fragaszy EB. Hand hygiene practices and the risk of human coronavirus infections in a UK community cohort. Wellcome Open Res 2020;5:98 [FREE Full text] [doi: 10.12688/wellcomeopenres.15796.2] [Medline: 34250260]
11. Lee JKC, Bullen C, Ben Amor Y, Bush SR, Colombo F, Gaviria A, et al. Institutional and behaviour-change interventions to support COVID-19 public health measures: a review by the Lancet Commission Task Force on public health measures to suppress the pandemic. Int Health 2021;13(5):399-409 [FREE Full text] [doi: 10.1093/inthealth/ihab022] [Medline: 33974687]

XSL•FO
RenderX

12. Collins LM. Conceptual introduction to the multiphase optimization strategy (MOST). In: Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland: Springer International Publishing; 2018:1-34.
13. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
14. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
15. Hamilton K, Cameron LD, Hagger MS, Hankonen N, Lintunen T. The Handbook of Behavior Change. Cambridge, UK: Cambridge University Press; 2020.
16. Carver CS, Scheier MF. On the Self-Regulation of Behavior. New York, NY: Cambridge University Press; 1998:439.
17. Schwarzer R, Hamilton K. Changing behavior using the health action process approach. In: Hamilton K, Cameron LD, Hagger MS, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change. Cambridge, UK: Cambridge University Press; 2020:89-103.
18. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
19. Orbell S, Verplanken B. Changing behavior using habit theory. In: Hamilton K, Cameron LD, Hagger MS, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change. Cambridge, UK: Cambridge University Press; 2020:178-192.
20. Wise J. Covid-19: WHO declares end of global health emergency. BMJ 2023;381:1041. [doi: 10.1136/bmj.p1041] [Medline: 37160309]
21. Contzen N, De Pasquale S, Mosler HJ. Over-reporting in handwashing self-reports: potential explanatory factors and alternative measurements. PLoS One 2015;10(8):e0136445 [FREE Full text] [doi: 10.1371/journal.pone.0136445] [Medline: 26301781]
22. Allan JL, Sniehotta FF, Johnston M. The best laid plans: planning skill determines the effectiveness of action plans and implementation intentions. Ann Behav Med 2013;46(1):114-120. [doi: 10.1007/s12160-013-9483-9] [Medline: 23456214]
23. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
24. Gardner B, Abraham C, Lally P, de Bruijn GJ. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int J Behav Nutr Phys Act 2012;9:102 [FREE Full text] [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
25. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005;20(2):143-160. [doi: 10.1080/08870440512331317670]
26. Ajzen I. Constructing a Theory of Planned Behaviour Questionnaire. UMass Amherst. 2006. URL: http://www.people.umass.edu/aizen/pdf/tpb.measurement.pdf [accessed 2025-03-29]
27. Schwarzer R, Schuz B, Ziegelmann JP, Lippke S, Luszczynska A, Scholz U. Adoption and maintenance of four health behaviors: theory-guided longitudinal studies on dental flossing, seat belt use, dietary behavior, and physical activity. Ann Behav Med 2007;33(2):156-166. [doi: 10.1007/BF02879897] [Medline: 17447868]
28. Winkler N, Kroh M, Spieß M. Entwicklung Einer Deutschen Kurzskala zur Zweidimensionalen Messung von Sozialer Erwünschtheit [dissertation]. Bielefeld University. 2006. URL: https://pub.uni-bielefeld.de/record/2918149 [accessed 2024-10-17]
29. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
30. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
31. Overall JE, Tonidandel S, Starbuck RR. Last-observation-carried-forward (LOCF) and tests for difference in mean rates of change in controlled repeated measurements designs with dropouts. Soc Sci Res 2009;38(2):492-503. [doi: 10.1016/j.ssresearch.2009.01.004]
32. Garnett C, Oldham M, Angus C, Beard E, Burton R, Field M, et al. Evaluating the effectiveness of the smartphone app, drink less, compared with the NHS alcohol advice webpage, for the reduction of alcohol consumption among hazardous and harmful adult drinkers in the UK at 6-month follow-up: protocol for a randomised controlled trial. Addiction 2021;116(2):412-425 [FREE Full text] [doi: 10.1111/add.15287] [Medline: 33067856]
33. Baretta D, Inauen J. Implications of the COVID-19 trajectory for the evaluation of hand hygiene interventions: secondary analysis of the soapp trial. Transl Behav Med 2024;14(1):66-71 [FREE Full text] [doi: 10.1093/tbm/ibad075] [Medline: 38001051]
34. SOAPP - hand hygiene promotion in the context of the COVID-19 pandemic. OSF. URL: https://osf.io/pzhqx/ [accessed 2024-11-20]

XSL•FO
RenderX

35. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
36. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012;141(1):2-18. [doi: 10.1037/a0024338] [Medline: 21823805]
37. Gozdzielewska L, Kilpatrick C, Reilly J, Stewart S, Butcher J, Kalule A, et al. The effectiveness of hand hygiene interventions for preventing community transmission or acquisition of novel coronavirus or influenza infections: a systematic review. BMC Public Health 2022;22(1):1283 [FREE Full text] [doi: 10.1186/s12889-022-13667-y] [Medline: 35780111]
38. Szczuka Z, Abraham C, Baban A, Brooks S, Cipolletta S, Danso E, et al. The trajectory of COVID-19 pandemic and handwashing adherence: findings from 14 countries. BMC Public Health 2021;21(1):1791 [FREE Full text] [doi: 10.1186/s12889-021-11822-5] [Medline: 34610808]
39. Bronfman N, Repetto P, Cordón P, Castañeda J, Cisternas P. Gender differences on psychosocial factors affecting COVID-19 preventive behaviors. Sustainability 2021;13(11):6148. [doi: 10.3390/su13116148]
40. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, Smith SS, et al. The Multiphase Optimization Strategy for engineering effective tobacco use interventions. Ann Behav Med 2011;41(2):208-226 [FREE Full text] [doi: 10.1007/s12160-010-9253-x] [Medline: 21132416]
41. Janiaud P, Hemkens LG, Ioannidis JPA. Challenges and lessons learned from COVID-19 trials: Should we be doing clinical trials differently? Can J Cardiol 2021;37(9):1353-1364 [FREE Full text] [doi: 10.1016/j.cjca.2021.05.009] [Medline: 34077789]
42. Koller JE, Villinger K, Lages NC, Absetz P, Bamert M, Branquinho C, et al. Individual and collective protective responses during the early phase of the COVID-19 pandemic in 10 different countries: results from the EUCLID online survey. Int J Infect Dis 2022;122:356-364 [FREE Full text] [doi: 10.1016/j.ijid.2022.06.012] [Medline: 35709963]
43. Smith SR, Hagger MS, Keech JJ, Moyers SA, Hamilton K. Improving hand hygiene behavior using a novel theory-based intervention during the COVID-19 pandemic. Ann Behav Med 2022;56(11):1157-1173 [FREE Full text] [doi: 10.1093/abm/kaac041] [Medline: 36099420]
44. Park JJ, Thorlund K, Mills EJ. Critical concepts in adaptive clinical trials. Clin Epidemiol 2018;10:343-351 [FREE Full text] [doi: 10.2147/CLEP.S156708] [Medline: 29606891]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
ITT: intention-to-treat
MOST: Multiphase Optimization Strategy
RCT: randomized controlled trial

Edited by L Buis; submitted 08.02.24; peer-reviewed by J Garlasco, H Giese; comments to author 07.10.24; revised version received 25.10.24; accepted 11.03.25; published 24.04.25.

Please cite as:
Baretta D, Rüttimann CL, Amrein MA, Inauen J
Promoting Hand Hygiene During the COVID-19 Pandemic: Randomized Controlled Trial of the Optimized Soapp+ App
JMIR Mhealth Uhealth 2025;13:e57191
URL: https://mhealth.jmir.org/2025/1/e57191
doi:10.2196/57191
PMID:

©Dario Baretta, Carole Lynn Rüttimann, Melanie Alexandra Amrein, Jennifer Inauen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of Technology-Enabled Medical Nutrition Therapy on Weight Loss in Adults With Overweight and Obesity: Retrospective Observational Study

Emily A Hu[1], PhD; Tommy Kelley[1], BA; Ajay Haryani[2], MD, MBA

[1]Nourish, New York, NY, United States

[2]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Emily A Hu, PhD
Nourish
Floor 15
853 Broadway
New York, NY, 10003
United States
Phone: 1 929 460 8273
Email: emily.hu@usenourish.com

Abstract

Background: Obesity represents a major public health crisis in the United States, imposing substantial health risks and economic costs. Medical nutrition therapy (MNT) is an evidence-based treatment where a registered dietitian provides personalized nutrition and lifestyle guidance to patients. MNT has been demonstrated to be effective for weight loss and managing chronic diseases in patients with obesity. With the rise of telehealth, MNT has gained popularity as an accessible alternative to traditional in-person care. While a nationwide program integrating MNT with a companion mobile app offers a comprehensive weight management solution, data supporting its clinical effectiveness is limited.

Objective: This study aimed to evaluate the effectiveness of an MNT program with a companion mobile app on weight loss among adults with overweight and obesity.

Methods: This retrospective cohort study included users of Nourish, an MNT program with a companion mobile app, who attended at least 1 appointment between August 2023 and October 2024 and had a baseline BMI≥30 kg/m² or a BMI between 27-30 kg/m² with diabetes or prediabetes. The primary outcome was the proportion of participants who achieved at least 5% weight loss; secondary outcomes included mean weight change, mean percent weight change, and the proportion of participants who achieved at least 3% weight loss. Statistical significance of weight change was determined using 2-tailed *t* tests. Subgroup analyses were performed by sex, BMI, follow-up time between weights, number of appointments completed, and levels of engagement according to appointment frequency and app usage.

Results: In total, 3951 participants were included in the analysis. The mean age was 38 (SD 10) years, and 78% (3082/3951) of participants were female. Weight loss was reported as a program goal by 70% (2748/3951) of participants, while 31% (1204/3951) and 24% (939/3951) reported diabetes or prediabetes and a cardiovascular condition, respectively. Over a median follow-up of 2.2 months, 17% (689/3951) of participants achieved at least 5% weight loss. The mean weight change was –4.5 (SD 8.9) pounds, corresponding to a mean percent weight change of –2% (SD 3.9; $P<.001$). Males and participants aged 60 years or older were more likely to experience at least 5% weight loss. Longer follow-up time between weights and a higher number of completed appointments (≥5 appointments) were significantly associated with a significantly higher likelihood of achieving at least 5% weight loss ($P<.001$ for both). In addition, participants who were most engaged, based on appointment frequency and app usage, were more likely to achieve at least 5% weight loss compared with those who were less engaged ($P<.001$).

Conclusions: Engagement with an MNT program and companion mobile app is associated with significant weight loss for adults with overweight and obesity and may serve as an effective, scalable weight management solution.

(JMIR Mhealth Uhealth 2025;13:e70228) doi:10.2196/70228

XSL•FO
RenderX

KEYWORDS

medical nutrition therapy; mHealth; obesity; overweight; registered dietitian; smartphone application; telehealth; virtual; weight loss

Introduction

Obesity remains a critical public health challenge nationwide, affecting over 100 million adults across the United States according to recent estimates [1]. In the United States, the prevalence of obesity has continued to rise, with over 40% of adults classified as obese in 2023 [2] and rates expected to exceed 60% in 2050, posing a significant public health threat given its impact on the existing chronic disease epidemic [3]. This trend is associated with substantial health and economic burden, as obesity-related conditions, such as diabetes and other cardiovascular disease risk factors, are projected to worsen cardiovascular health and cost up to US $1.49 trillion annually, along with additional losses from reduced productivity [4].

A cornerstone of obesity prevention and treatment is an optimization of behavioral and lifestyle modifications, including adherence to regular physical activity and a well-balanced, heart-healthy diet [5]. Effective nutritional counseling, as part of a medical nutrition program, has repeatedly proven to result in weight loss, thereby lowering an individual's risk of obesity-related comorbidities, including diabetes and cardiovascular disease [6-10].

Medical nutrition therapy (MNT) is a treatment modality where a registered dietitian (RD) meets one-on-one with a patient to provide individualized lifestyle recommendations. MNT can be used to address a wide range of health conditions and may occur in various practice settings, such as outpatient and inpatient facilities. Following an initial intake, an RD meets with a patient regularly to individualize their treatment plan and help patients achieve their health goals. Data have consistently demonstrated the impact of MNT on reducing body weight, BMI, and waist circumference and have been shown to improve weight-related risk factors, including hypertension, diabetes, hyperlipidemia, and quality of life [6,8]. For weight management specifically, previous research has demonstrated that participants receiving MNT were twice as likely to achieve 5% weight loss compared with controls [6]. Despite its effectiveness, the utilization of MNT in practice is low relative to the prevalence of obesity for multiple reasons, including access, affordability, and awareness of this resource [11,12].

In recent years, mobile health interventions targeting weight management have gained popularity. These apps often feature calorie and meal logging, meal planning, and grocery ordering. Studies show that such apps can lead to significant short-term and long-term weight reductions [13-15]. However, many mobile apps lack tailored feedback, which may limit engagement and long-term use, underscoring the need for complementary interventions.

During the COVID-19 pandemic, the adoption of telehealth significantly broadened health care accessibility [16]. In response, the American Heart Association (AHA) issued a policy statement emphasizing the importance of expanding access to and optimizing the use of MNT, citing its numerous benefits for managing chronic diseases [17]. MNT delivered via telehealth, which emerged as an opportunity to bridge the gap highlighted by the AHA, had the potential to not only improve access and convenience of MNT but to also help achieve clinical goals. Emerging research has validated the potential of MNT delivered via telehealth to deliver on the AHA mission by improving access, and in fact, has resulted in better engagement and adherence resulting from the increased convenience of telehealth [18]. The emergence of telehealth is especially important for rural populations, who face unique challenges, such as limited access to health care providers and greater distances to fresh food sources [19,20]. These barriers are associated with worse health outcomes, particularly for those with chronic conditions. Telehealth offers a promising solution to address these disparities by providing on-demand access to health care services.

Starting in 2021, a national program, that combines MNT delivered via telehealth with a companion mobile app, was launched to provide care for individuals with overweight and obesity seeking to lose weight and improve their health. The app includes features such as meal logging, meal planning, and outcomes tracking. By combining 2 proven interventions for weight loss, the program sought to deepen patient engagement with their MNT treatment, and in turn, see meaningful clinical outcomes. This study aimed to evaluate the effectiveness of this combined solution on weight loss among adults with overweight and obesity across the United States.

Methods

MNT Program With a Companion Mobile App

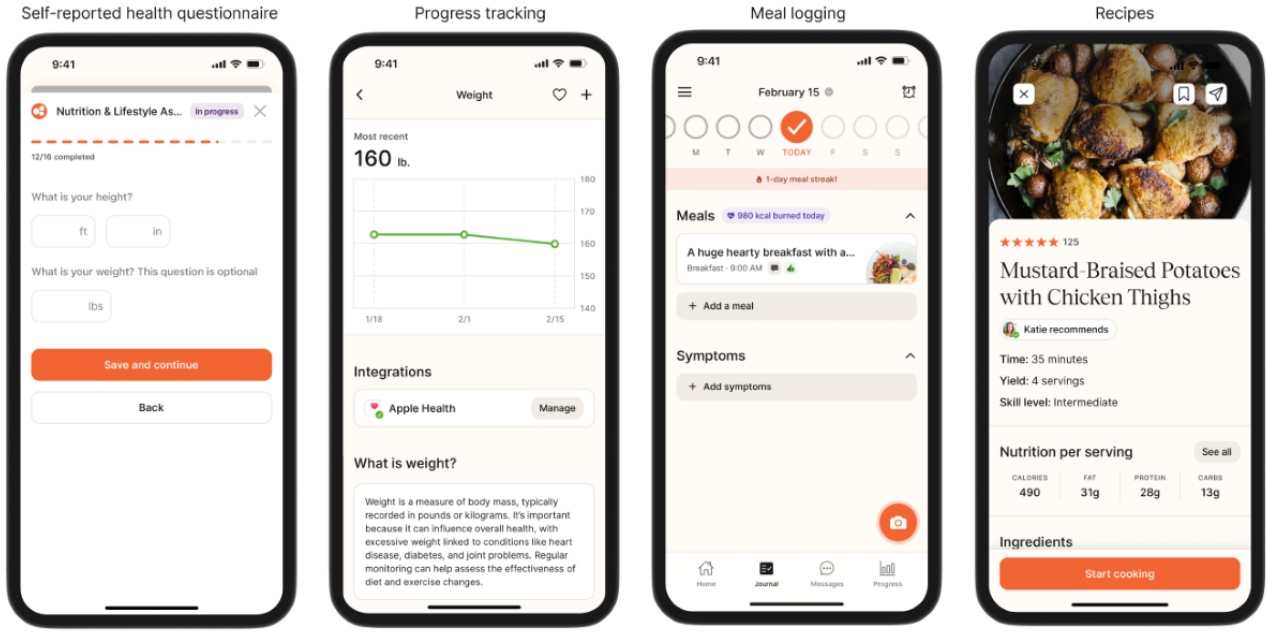
Nourish is a national telehealth provider of MNT delivered by a licensed RD, paired with a companion mobile app to offer asynchronous support and tools. Nourish has a network of over 1500 RDs covering all major specialties, including overweight and obesity, diabetes, and heart disease, and can serve patients across all 50 states. Patients find Nourish independently or through a referral from a clinician. The cost of Nourish is covered by insurance for the majority of patients as Nourish is in a network with the nation's largest commercial insurance plans. Upon enrollment, patients provide demographic information, medical history, program goals, and personal preferences, which are used to match patients with recommended RDs that best fit their clinical needs and goals. Patients then schedule their first appointment with an RD they would like to work with. Patients have the option to switch providers at any time during their program.

Patients meet virtually in one-on-one sessions with their RD for an initial nutritional intake assessment through the Nourish web portal or app. The RD then develops a personalized treatment plan focused on nutrition and lifestyle recommendations to help patients achieve their health goals. Furthermore, 1-hour sessions are scheduled every 1 to 3 weeks, depending on the treatment plan. In addition to telehealth

XSL•FO
RenderX

appointments, patients also have access to a proprietary mobile app (Android [Google], iOS [Apple Inc]) and web portal with features designed to complement the clinical intervention and increase engagement. Features of the app include a self-reported health questionnaire that assesses lifestyle and nutritional habits, progress tracking of weight and other physiological measurements and laboratory data, meal logging, recipe and meal planning content, nutrition education resources, and on-demand chat with an RD (Figure 1). Patients receive automated reminders to track their progress and log meals through push notifications within the app and can also set up customized reminders. Patients also receive notifications to schedule their next appointment.

Figure 1. Screenshots of Nourish mobile app.



Study Population

This retrospective cohort study evaluated weight change and engagement among newly enrolled users who completed their first appointment with an RD between August 2023 and October 2024. Individuals were eligible for inclusion if they were at least 18 years old and met one of the following BMI criteria: baseline BMI≥30 kg/m^2 or baseline BMI between 27 and 30 kg/m^2 with self-reported diabetes or prediabetes. In addition, we required individuals to have at least 2 nonidentical weight measurements at least 30 days apart with their first weight within 2 weeks of their first appointment and their last at least 4 weeks after their first appointment and within 90 days of their last appointment. Exclusion criteria included reported usage of weight loss medications, history of bariatric surgery, or history of thyroid disorders. Given the retrospective and observational nature of this study, a prospective sample size and power analysis were not performed.

Assessments

Data used for this analysis were self-reported by participants in the app or during the appointment, where RDs entered data on the participants' behalf. In addition, platform-collected app engagement data were used to derive the number of appointments, app touchpoints, and time between recorded weights. Participants were prompted to enter demographic and clinical information when they enrolled in the program. These self-reported assessments included sex, age, weight, height, and goals for the program (eg, weight loss, managing comorbidities such as diabetes or prediabetes, and cardiovascular conditions such as high blood pressure and high cholesterol). Participants could update their weight at any time through the app, and RDs were prompted to record weight during appointments if participants did not update them in the app. All fields were optional except for date of birth and goal for enrolling in the program.

Based on self-reported height and weight data, BMI was calculated. Weight change was calculated as the difference between baseline weight and follow-up weight. Percent weight change was calculated as weight change divided by baseline weight. Follow-up time between weights was defined as days between the date that baseline weight was recorded and the date that follow-up weight was recorded. In addition, follow-up time between appointments for participants who attended at least 2 appointments was calculated. The primary outcome was the proportion of participants who achieved at least 5% weight loss. Secondary outcomes included mean weight change (in pounds), mean percent weight change, and the proportion of participants who achieved at least 3% weight loss. Additional outcomes included the proportion of participants who achieved any weight loss, at least 7% weight loss, and at least 10% weight loss.

To examine the relationship between engagement and weight reduction, engagement metrics based on app-generated data were also analyzed. Engagement was assessed based on the number of completed appointments and the number of total touchpoints with the app. The number of touchpoints with the app was calculated based on total engagements with key app features, including the number of meals logged, the number of messages sent to their RD, and the number of vital or laboratory

XSL•FO
RenderX

values (eg, weight, height, blood pressure, and cholesterol) entered.

Statistical Analyses

Baseline characteristics were summarized for the total population and further stratified by whether participants achieved at least 5% weight loss. Descriptive statistics, including means, SDs, medians, IQRs, and ranges for continuous variables and percentages for categorical variables were reported. Group comparisons were conducted using ANOVA for continuous variables and chi-square tests for categorical variables. Furthermore, 2-tailed *t* tests were conducted to determine whether weight changes significantly differed from zero. Participants with missing data for age and sex were excluded from the analyses for those specific metrics.

Mean (SD) weight change, mean (SD) percent weight change, and the proportion of participants who achieved at least 5% weight loss were calculated by baseline characteristics. The mean (SD) weight change and mean (SD) percent weight change were calculated for the total population and stratified by whether participants achieved at least 5% weight loss, with results presented by sex and BMI category. ANOVA and chi-square tests were used to test for significant differences within each subgroup.

To address potential bias from loss to follow-up, the proportion of participants achieving at least 3% and 5% weight loss were analyzed according to follow-up time between weights (<2 months, 2 to up to 3 months, 3 to up to 6 months, and ≥6 months) and the number of appointments completed (1-2, 3-4, 5-7, and ≥8 appointments). Furthermore, the proportion of participants who achieved any and at least 3%, 5%, 7%, and 10% weight loss was calculated among the total population and by whether participants completed at least 5 appointments (the mean number of appointments in the study population). To evaluate the potential interaction of appointments and app touchpoints on weight loss, participants were stratified into low- or high-engagement groups using mean values as cutoffs to compare relative engagement within the study population. In total, four subgroups were defined: (1) low appointments (<5) and low app usage (<100 touchpoints), 2) low appointments (<5) and high app usage (≥100 touchpoints), (3) high appointments (≥5) and low app usage (<100 touchpoints), and (4) high appointments (≥5) and high app usage (≥100 touchpoints).

A *P* value less than .05 was considered to be statistically significant for all tests. Stata (version 18; StataCorp) was used for all analyses.

Ethical Considerations

This study adhered to ethical standards for human participants research. At sign-up, all participants were notified that their data may be used for research purposes, and data were deidentified before analysis to ensure privacy. No compensation was provided to participants. The study was declared exempt from institutional review board oversight by Advarra institutional review board (protocol #00083062) given the retrospective design of the study and less than minimal risk to participants.

Results

Participant Characteristics

In total, 3951 participants were included in this analysis. Baseline characteristics are presented in Table 1. Participants had a mean age of 38 (SD 10) years, mean BMI of 37 (SD 6) kg/m^2, and were primarily female (3082/3951, 78%). Furthermore, 52% (2066/3951) of participants had a BMI greater than 35 kg/m^2 and a majority (2748/3951, 70%) reported that weight loss was a goal of the program. Diabetes or prediabetes and cardiovascular conditions were reported among 31% (1204/3951) and 24% (939/3951) of participants, respectively. The mean and median follow-up time between weight entries was 80 (SD 49) days and 67 (IQR 47-95) days, respectively. The mean and median time between appointments was 70 (SD 61) days and 52 (IQR 26-95) days. In total, participants attended a total of 18,966 appointments (mean 4.8, SD 4.0) and had 393,131 touchpoints with the app (mean 100, SD 101; data not shown).

XSL•FO
RenderX

Table 1. Baseline characteristics.

Characteristics	Total population (N=3951)	Achieved ≥5% weight loss (n=689)	Achieved <5% weight loss (n=3262)	*P* value
Sex[a], n (%)				
Male	759 (19)	151 (22)	608 (19)	.05
Female	3082 (78)	514 (75)	2568 (79)	—[b]
Age [a] (y)				
Mean (SD)	38 (10)	38 (11)	38 (10)	.02
Median (IQR)	36 (30-44)	36 (31-44)	36 (30-43)	—
Range	18-78	19-78	18-78	—
Age group (y), n (%)				
18-29	809 (20)	138 (20)	671 (21)	.03
30-39	1653 (42)	280 (41)	1373 (42)	—
40-59	1237 (31)	209 (30)	1028 (32)	—
≥60	141 (4)	38 (6)	103 (3)	—
Follow-up time between weights (d)				
Mean (SD)	80 (49)	97 (53)	76 (47)	<.001
Median (IQR)	67 (47-95)	85 (62-116)	63 (45-90)	—
Range	30-424	30-403	30-424	—
Follow-up time between appointments[c] (d)				
Mean (SD)	70 (61)	81 (67)	67 (60)	.002
Median (IQR)	52 (26-95)	64 (29-112)	49 (24-91)	—
Range	1-412	3-361	1-412	—
Weight (pounds)				
Mean (SD)	228 (44)	229 (44)	227 (44)	.4
Median (IQR)	220 (195-252)	220 (196-255)	220 (195-252)	—
Range	131-398	142-396	131-398	—
BMI (kg/m^2)				
Mean (SD)	37 (6)	37 (6)	37 (6)	.4
Median (IQR)	35 (32-40)	36 (32-40)	35 (32-40)	—
Range	27-75	27-75	27-67	—
Baseline BMI, n (%)				
27 to <30 kg/m^2	181 (5)	28 (4)	153 (5)	.80
30 to <35 kg/m^2	1704 (43)	287 (42)	1417 (43)	—
35 to <40 kg/m^2	1048 (27)	193 (28)	855 (26)	—
40 to <50 kg/m^2	834 (21)	149 (22)	685 (21)	—
≥50 kg/m^2	184 (5)	32 (5)	152 (5)	—
Diabetes or prediabetes, n (%)	1204 (31)	227 (33)	977 (30)	.10
Cardiovascular condition, n (%)	939 (24)	168 (24)	771 (24)	.70
Weight loss goal, n (%)	2748 (70)	487 (71)	2261 (69)	.50

[a]Overall, 3% (n=151) of the total population did not report sex nor age and were not included in summaries of sex and age.

[b]Not applicable.

[c]Among participants who had at least 2 appointments (n=3269).

XSL•FO
RenderX

Weight Change

In total, 17% (689/3951) of participants achieved at least 5% weight loss (Table 2). When comparing results by sex, males were more likely to achieve at least 5% weight loss compared with females (151/759, 20% vs 514/3082, 17%; *P*=.05, 95% CI 1.1-4.9). Participants who were 60 years or older were more likely to achieve at least 5% weight loss compared with participants who were younger (38/141, 27% vs 627/3699, 17%; *P*=.03, 95% CI 5.2-14.2). While those who had a higher baseline BMI lost more weight compared with those with a lower BMI (*P*<.001, 95% CI –6.0 to –3.0), there was no statistically significant association between BMI and achieving 5% weight loss. Notably, participants who entered the program with a goal of losing weight experienced significantly more weight loss than those without a weight loss goal (*P*<.001, 95% CI –1.8 to –0.6). However, the weight loss goal was not statistically and significantly associated with achieving at least 5% weight loss. In total, 74% (2929/3951) of participants lost any weight, 34% (1350/3951) of participants achieved at least 3% weight loss, 9% (336/3951) achieved at least 7% weight loss, and 3% (123/3951) achieved at least 10% weight loss (Table S1 in Multimedia Appendix 1).

Table 2. Proportion of participants who achieved at least 5% weight loss, mean weight change, and percent weight change by baseline characteristics

Population	Participants, N	Achieved ≥5% weight loss			Weight change (pounds)			Percent weight change (%)		
		n (%)	*P* value	95% CI	Mean (SD)	*P* value	95% CI	Mean (SD)	*P* value	95% CI
Total population	3951	689 (17)	—[a]	—	–4.5 (8.9)	—	—	–2 (3.9)	—	—
Sex[b]										
Male	759	151 (20)	.05	1.1 to 4.9	–6.1 (9.4)	<.001	–2.6 to –1.4	–2.4 (3.8)	.002	–0.9 to –0.3
Female	3082	514 (17)	—	—	–4.1 (8.7)	—	—	–1.8 (3.9)	—	—
Age group[b] (y)										
18-29	805	138 (17)	.03	5.2 to 14.2	–4.0 (9.2)	.20	–2.5 to –1.1	–1.7 (4.2)	.05	–1.5 to –0.5
30-39	1653	280 (17)	—	—	–4.6 (8.4)	—	—	–2 (3.7)	—	—
40-59	1240	209 (17)	—	—	–4.7 (9.3)	—	—	–2 (4)	—	—
≥60	142	38 (27)	—	—	–5.8 (7.9)	—	—	–2.7 (3.6)	—	—
Baseline BMI, %										
27-<30 kg/m^2	181	28 (16)	.80	0.5 to 4.3	–3.5 (7.8)	<.001	–6.0 to –3.0	–1.9 (4.5)	.02	–1.1 to –0.1
30-<35 kg/m^2	1704	287 (17)	—	—	–3.5 (8.2)	—	—	–1.7 (4.1)	—	—
35-<40 kg/m^2	1048	193 (18)	—	—	–4.7 (9)	—	—	–2 (3.9)	—	—
40-<50 kg/m^2	834	149 (18)	—	—	–5.8 (9.6)	—	—	–2.2 (3.6)	—	—
≥50 kg/m^2	184	32 (17)	—	—	–8.0 (10.6)	—	—	–2.5 (3.2)	—	—
Diabetes or prediabetes										
Yes	1204	227 (19)	.10	0.9 to 3.2	–5.2 (9)	.002	–1.7 to –0.3	–2.2 (4.0)	.02	–0.6 to –0.1
No	2747	462 (17)	—	—	–4.2 (8.8)	—	—	–1.9 (3.9)	—	—
Cardiovascular condition										
Yes	939	168 (18)	.70	–0.6 to 1.9	–5.0 (8.3)	.10	–1.2 to 0.0	–2.1 (3.6)	.10	–0.3 to 0.3
No	3012	521 (17)	—	—	–4.4 (9.1)	—	—	–2.1 (3.6)	—	—
Weight loss goal										
Yes	2748	487 (18)	.50	–0.4 to 2.0	–4.9 (8.6)	<.001	–1.8 to –0.6	–2.1 (3.7)	<.001	–0.8 to –0.2
No	1203	202 (17)	—	—	–3.7 (9.6)	—	—	–1.6 (4.4)	—	—

[a]Not applicable.

[b]Overall, 3% (n=151) of the total population did not report sex nor age and were not included in summaries for sex and age.

XSL•FO
RenderX

The mean weight change was –4.5 (SD 8.9) pounds (*P*<.001) and the mean percent weight change was –2% (SD 3.9; *P*<.001; Table 3). Among those who achieved at least 5% weight loss, the mean weight change was –17.6 (SD 6.3) pounds (*P*<.001) and the mean percent weight change was –7.7% (SD 2.4; *P*<.001), while those who did not achieve at least 5% weight loss had a mean weight change of –1.8 (SD 6.6) pounds and a mean percent weight change of –0.7% (SD 3.0; *P*<.001). While males lost more weight compared with females (mean weight change –6.1, SD 9.4 pounds for males vs mean weight change –4.1, SD 8.7 pounds for females), the mean percent weight change for those who achieved at least 5% weight loss was similar across males (mean –7.7%, SD 2.2) and females (mean –7.7%, SD 2.5). In addition, participants with a higher baseline BMI experienced greater weight loss compared with those with a lower BMI. Mean weight change was –3.4 (SD 8.2) pounds for participants with BMI<35 kg/m², –4.7 (SD 9.0) pounds for those with BMI 35-40 kg/m², and –6.2 (SD 9.8) pounds for those with BMI≥40 kg/m², demonstrating a progressive increase in weight loss with higher BMI categories.

Table 3. Mean weight change and percent weight change among the total population and stratified by 5% weight loss among all participants, by sex, and BMI category.

Category	Total population	*P* value[a]	Achieved ≥5% weight loss	*P* value[a]	Achieved <5% weight loss	*P* value[a]
All participants						
n	3951	—[b]	689	—	3262	—
Weight change (pounds), mean (SD)	–4.5 (8.9)	<.001	–17.6 (6.3)	<.001	–1.8 (6.6)	<.001
Percent weight change (%), mean (SD)	–2.0 (3.9)	<.001	–7.7 (2.4)	<.001	–0.7 (3.0)	<.001
Males						
n	759	—	151	—	608	—
Weight change (pounds), mean (SD)	–6.1 (9.4)	<.001	–19.5 (6.0)	<.001	–2.7 (6.9)	<.001
Percent weight change (%), mean (SD)	–2.4 (3.8)	<.001	–7.7 (2.2)	<.001	–1.1 (2.8)	<.001
Females						
n	3082	—	514	—	2568	—
Weight change (pounds), mean (SD)	–4.1 (8.7)	<.001	–17.0 (6.1)	<.001	–1.6 (6.6)	<.001
Percent weight change (%), mean (SD)	–1.8 (3.9)	<.001	–7.7 (2.5)	<.001	–0.7 (3.0)	<.001
BMI<30 kg/m^2						
n	1885	—	315	—	1570	—
Weight change (pounds), mean (SD)	–3.4 (8.2)	<.001	–15.6 (5.4)	<.001	–1.1 (6.3)	<.001
Percent weight change (%), mean (SD)	–1.7 (4.4)	<.001	–7.9 (2.6)	<.001	–0.5 (3.2)	<.001
BMI 30-35 kg/m^2						
n	1048	—	193	—	855	—
Weight change (pounds), mean (SD)	–4.7 (9.0)	<.001	–17.5 (5.3)	<.001	–1.8 (6.8)	<.001
Percent weight change (%), mean (SD)	–2.0 (3.9)	<.001	–7.5 (2.2)	<.001	–0.8 (3.0)	<.001
BMI ≥35 kg/m ^{2}						
n	1018	—	181	—	837	—
Weight change (pounds), mean (SD)	–6.2 (9.8)	<.001	–21.1 (7.0)	<.001	–3.0 (7.0)	<.001
Percent weight change (%), mean (SD)	–2.2 (3.5)	<.001	–7.6 (2.3)	<.001	–1.1 (2.5)	<.001

[a]*t* test to test whether weight change was different from zero.

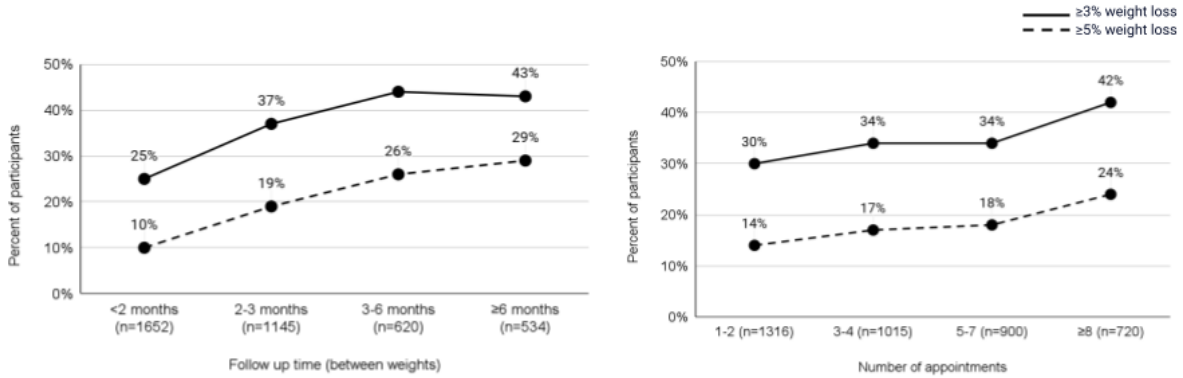
[b]Not applicable.

Engagement With Program

The proportion of participants who achieved at least 3% and 5% weight loss, stratified by time between weights, is presented in Figure 2A. Participants who engaged with the program for longer periods of time were more likely to achieve at least 3% (*P*<.001, 95% CI 15.0-20.8) and 5% weight loss (*P*<.001, 95% CI 16.2-21.9). Figure 2B illustrates the proportion of participants achieving at least 3% and 5% weight loss, stratified by the number of completed appointments (1-2, 3-4, 5-7, and ≥8). Participants who attended more appointments had a significantly greater likelihood of achieving at least 3% weight loss (*P*<.001, 95% CI 9.4-14.7) and 5% weight loss (*P*<.001, 95% CI 7.5-12.3).

XSL•FO
RenderX

Figure 2. Percent of participants who achieved at least 3% and 5% weight loss according to (A) follow-up time and (B) number of appointments.



Those who completed at least 5 appointments (n=1620) had significantly higher rates of achieving weight loss milestones compared with those who attended fewer than 5 appointments (n=2331): 38% (613/1620) vs 32% (737/2331) for at least 3% weight loss (*P*<.001, 95% CI 3.6-8.4), 20% (331/1620) vs 15% (358/2331) for at least 5% weight loss (*P*<.001, 95% CI 2.8-7.2), 11% (179/1620) vs 7% (157/2331) for at least 7% weight loss (*P*<.001, 95% CI 2.1-5.9), and 5% (73/1620) vs 2% (50/2331) for at least 10% weight loss (*P*<.001, 95% CI 1.4-4.6; Table S1 in Multimedia Appendix 1).

Engagement With MNT and App

A significant interaction was observed between appointments and app engagement in predicting the likelihood of achieving at least 5% weight loss (*P*<.001, 95% CI 4.1-9.9; Table 4). Participants with high engagement in both appointments (≥5 appointments) and app usage (≥100 total touchpoints) had the greatest likelihood of achieving at least 5% weight loss. The likelihood of this occurrence followed in descending order by participants with a high number of appointments and low app usage, a low number of appointments and high app usage, and low appointment completion and low app usage.

Table 4. Percent of participants who achieved at least 5% weight loss by engagement data.

App usage[a]	Appointments[b]	
	Low (<5 appointments)	High (≥5 appointments)
Low (<100 touchpoints)	15% (280/1833)	19% (165/875)
High (≥100 touchpoints)	16% (78/498)	22% (167/745)

[a]App usage includes meal logging, messages to registered dietician, or health questionnaire entries.

[b]Percents (n/N) are displayed.

Discussion

Principal Findings

Among adults with overweight and obesity across the United States, enrollment in an MNT program with a companion mobile app was associated with clinically meaningful weight loss. Over a median of 67 (IQR 47-95) days, 17% (689/3951) of participants lost at least 5% of initial weight and the majority (2929/3951, 74%) of participants lost any weight. Males and participants older than 60 years old were more likely to achieve at least 5% weight loss compared with females and participants younger than 60 years old, respectively. The greatest weight loss was achieved by participants who were most engaged with both MNT appointments and the mobile app, suggesting a significant benefit of integrating both resources into a nutritional therapy solution.

MNT has been proven to be an effective strategy to treat obesity by achieving significant weight loss. A systematic review and meta-analysis of 14 randomized controlled trials comparing individualized nutrition care for weight management with usual care found a pooled mean difference of –2.3 pounds with 95% CI –3.1 to –1.5 (*P*<.001) over a median of 6 months [8]. In this study, participants who engaged with an MNT program experienced a mean weight change of –4.5 (SD 8.9) pounds over a median of 67 (IQR 47-95) days. Of note, several of the studies in the systematic review were focused on primary outcomes of blood pressure or diabetes, rather than weight loss. This suggests that when MNT is implemented with the primary goal of achieving weight loss, potentially more meaningful weight loss can be achieved.

In this study, while as few as 1 appointment resulted in meaningful weight loss, a higher number of appointments and longer time engaged in the program were directly correlated with a higher likelihood of clinically significant weight loss. In fact, the greatest outcomes were observed for participants who completed at least eight 1-hour appointments. Therefore, in addition to prioritizing clinically effective MNT, an effective program must also emphasize long-term engagement and adherence to optimize clinically meaningful weight loss for individuals. In fact, this was substantiated by the Academy of Nutrition and Dietetics who reported that the most beneficial outcomes for blood pressure and fasting glucose resulted after at least 5 sessions with an RD over a period of at least 12 months [21].

XSL•FO
RenderX

Although MNT helps individuals successfully lose weight, uptake has been low over the years, in part due to accessibility, affordability, and convenience [11,12]. At the same time, there has been a rise in digital health interventions for weight loss through mobile and web-based applications [22,23]. Several analyses evaluating the effectiveness of these commercial apps have been published and report meaningful weight loss among users in a real-world setting.

A retrospective cohort study of 35,921 participants with overweight and obesity who used a mobile app for weight loss, which included daily meal logging, activity monitoring, and personalized feedback, found that 78% (27983/35921) of participants lost weight over a median of 8.8 months [14]. While this study had a smaller sample size, a comparable percentage of participants lost weight over a shorter time frame. One additional benefit of this study compared with the study by Chin et al [14] is the inclusion of self-reported comorbidities and program goals, which allows for further understanding of how a weight loss program impacts patients based on underlying comorbidities and program goals.

Similarly, among 8977 adults with obesity, a digital nutrition platform with personalized recipes and grocery tools resulted in a mean weight loss of 3.7 (1.5%; SD 17.4) pounds over 10 months [13]. In one-fifth the follow-up time, participants engaged with the current MNT program with a companion mobile app achieved greater mean weight loss and mean percent weight loss. Furthermore, this study includes detailed engagement data, whereas the previous study did not report engagement metrics; as a result, it remains unclear whether weight loss in the previous study was associated with active app use or passive engagement with the platform. The greater weight loss achieved in a shorter amount of time with this solution may be attributed to the integration of individualized MNT provided by RDs with a companion app. These combined modalities may provide synergistic benefits compared with stand-alone app or MNT interventions.

Recognizing the importance of a multifaceted approach to weight loss and patients' desires to have a mobile app or web-based platform to engage with during their weight loss journey, this study aimed to better understand how app engagement further impacts weight loss success associated with MNT. The features of the mobile app component of this MNT program were designed to drive engagement and facilitate consistent behavior change to ultimately result in better adherence to diet recommendations and greater weight loss. This has been validated in the analysis by Chin et al [14], where meal logging (odds ratio [OR] 10.69, 95% CI 6.20-19.53; $P<.001$) and weight logging (OR 0.59, 95% CI 0.39-0.89; $P<.001$) were both associated with a greater likelihood of successful weight loss.

Alternatively, the trial evaluated the impact of a single 90-minute visit with an RD combined with an app for weight loss among individuals with overweight and obesity. The study concluded that the RD visit, along with app features such as meal logging and physical activity tracking via a Fitbit (Google) device, led to meaningful weight loss. However, they reported no additional benefit of receiving personalized, real-time feedback through in-app messages [24], which is in contrast to another study highlighting the benefit of an RD chat feature [25]. Therefore, while the utility of each individual app-based feature requires additional investigation, multiple studies, including this analysis, highlight that patients who additionally engaged with an app achieved the greatest weight loss.

Limitations

There are several limitations of this study to consider. First, this study is observational and does not include a control group, which limits any causal inferences of the intervention on reported outcomes. However, this study provides valuable insights into the practical implementation and applicability of this solution across a nationwide patient population.

Second, although RDs were instructed to validate weights during each appointment, this was not a requirement and further exacerbated the limitation of self-reported data. Self-reported weights are subject to recall bias; however, previous research has suggested a moderate to strong correlation between self-reported weight data collected and actual values [26].

Third, the follow-up time was relatively short compared with other retrospective cohort studies. While longer follow-up is imperative to demonstrate the long-term utility of this program and its effect on weight loss, meaningful reductions in weight coupled with high patient engagement over several months in the program highlight the potential of this solution over an extended period.

Fourth, physical activity, which is a significant factor in achieving and maintaining weight loss, was not systematically assessed in this analysis, which prevents appropriately accounting for its impact on weight loss outcomes.

Fifth, although the program covered a diverse group of patients across the 50 states, the lack of comprehensive demographic data capture beyond age and sex limits the generalizability of these findings and prevents evaluation of the program's effectiveness in those individuals disproportionately affected by social determinants of health.

Finally, the incomplete data capture of these variables noted above restricted adjustment for potential cofounders, thereby limiting the feasibility and interpretability of regression-based analyses.

Conclusions

The obesity epidemic in the United States highlights the need for effective and comprehensive approaches to weight management. This study provides evidence that engagement with an MNT program and a companion mobile app is associated with clinically meaningful weight loss among individuals with overweight and obesity. It also demonstrates that combining MNT with a companion mobile app may amplify the program's effect on weight loss. These findings highlight the potential for MNT programs with mobile apps to bridge the gap in access to effective nutrition care, offering a scalable and effective solution for weight management in individuals with overweight and obesity across the United States.

XSL•FO
RenderX

Acknowledgments

The authors would like to acknowledge David Feldman and Julia Rayis for providing feedback and assistance with the manuscript.

No generative artificial intelligence (AI) tools were used in the creation of this manuscript. This study was funded by Nourish.

Data Availability

The data analyzed in this study are proprietary and cannot be publicly shared.

Authors' Contributions

EAH contributed to conceptualization, formal analysis, methodology, writing—original draft, and writing—review and editing. TK contributed to data curation, resources, supervision, and writing—review and editing. AH contributed to supervision and writing—review and editing. All authors reviewed and approved the final manuscript.

Conflicts of Interest

EAH and TK are current employees of Nourish. AH is a paid advisor.

Multimedia Appendix 1
Percent of patients who achieved any and at least 3%, 5%, 7%, and 10% weight loss.
[DOCX File , 14 KB - mhealth_v13i1e70228_app1.docx]

References

1. CDC. Adult Obesity Facts. 2024. URL: https://www.cdc.gov/obesity/adult-obesity-facts/index.html [accessed 2025-04-10]
2. Emmerich SD, Fryar CD, Stierman B, Ogden CL. Obesity and severe obesity prevalence in adults: United States, August 2021-August 2023. NCHS Data Brief 2024(508). [doi: 10.15620/cdc/159281] [Medline: 39808758]
3. Joynt Maddox KE, Elkind MSV, Aparicio HJ, Commodore-Mensah Y, de Ferranti SD, Dowd WN, American Heart Association. Forecasting the burden of cardiovascular disease and stroke in the United States through 2050-prevalence of risk factors and disease: a presidential advisory from the American Heart Association. Circulation 2024;150(4):e65-e88 [FREE Full text] [doi: 10.1161/CIR.0000000000001256] [Medline: 38832505]
4. Kazi DS, Elkind MSV, Deutsch A, Dowd WN, Heidenreich P, Khavjou O, American Heart Association. Forecasting the economic burden of cardiovascular disease and stroke in the United States through 2050: a presidential advisory from the American Heart Association. Circulation 2024;150(4):e89-e101 [FREE Full text] [doi: 10.1161/CIR.0000000000001258] [Medline: 38832515]
5. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019;140(11):e563-e595. [doi: 10.1161/CIR.0000000000000677] [Medline: 30879339]
6. Morgan-Bathke M, Baxter SD, Halliday TM, Lynch A, Malik N, Raynor HA, et al. Weight management interventions provided by a dietitian for adults with overweight or obesity: an evidence analysis center systematic review and meta-analysis. J Acad Nutr Diet 2023;123(11):1621-1661.e25. [doi: 10.1016/j.jand.2022.03.014] [Medline: 35788061]
7. Dudzik JM, Senkus KE, Evert AB, Raynor HA, Rozga M, Handu D, et al. The effectiveness of medical nutrition therapy provided by a dietitian in adults with prediabetes: a systematic review and meta-analysis. Am J Clin Nutr 2023;118(5):892-910 [FREE Full text] [doi: 10.1016/j.ajcnut.2023.08.022] [Medline: 37689140]
8. Williams LT, Barnes K, Ball L, Ross LJ, Sladdin I, Mitchell LJ. How effective are dietitians in weight management? A systematic review and meta-analysis of randomized controlled trials. Healthcare (Basel) 2019;7(1):20 [FREE Full text] [doi: 10.3390/healthcare7010020] [Medline: 30717197]
9. Evert AB, Dennison M, Gardner CD, Garvey WT, Lau KHK, MacLeod J, et al. Nutrition therapy for adults with diabetes or prediabetes: a consensus report. Diabetes Care 2019;42(5):731-754 [FREE Full text] [doi: 10.2337/dci19-0014] [Medline: 31000505]
10. Ross LJ, Barnes KA, Ball LE, Mitchell LJ, Sladdin I, Lee P, et al. Effectiveness of dietetic consultation for lowering blood lipid levels in the management of cardiovascular disease risk: a systematic review and meta-analysis of randomised controlled trials. Nutr Diet 2019;76(2):199-210. [doi: 10.1111/1747-0080.12509] [Medline: 30714668]
11. Skea ZC, Aceves-Martins M, Robertson C, De Bruin M, Avenell A, REBALANCE team. Acceptability and feasibility of weight management programmes for adults with severe obesity: a qualitative systematic review. BMJ Open 2019;9(9):e029473. [doi: 10.1136/bmjopen-2019-029473] [Medline: 31511284]

XSL•FO
RenderX

12. Thorndike AN, Gardner CD, Kendrick KB, Seligman HK, Yaroch AL, Gomes AV, American Heart Association Advocacy Coordinating Committee. Strengthening US food policies and programs to promote equity in nutrition security: a policy statement from the American Heart Association. Circulation 2022;145(24):e1077-e1093. [doi: 10.1161/CIR.0000000000001072] [Medline: 35535604]
13. Hu EA, Nguyen V, Langheier J, Shurney D. Weight reduction through a digital nutrition and food purchasing platform among users with obesity: longitudinal study. J Med Internet Res 2020;22(9):e19634 [FREE Full text] [doi: 10.2196/19634] [Medline: 32792332]
14. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
15. Hu EA, Pasupuleti M, Nguyen V, Langheier J, Shurney D. Sustaining weight loss among adults with obesity using a digital meal planning and food purchasing platform for 12, 24, and 36 months: a longitudinal study. Nutr J 2021;20(1):8. [doi: 10.1186/s12937-021-00666-9] [Medline: 33478516]
16. Calton B, Abedini N, Fratkin M. Telemedicine in the time of coronavirus. J Pain Symptom Manage 2020;60(1):e12-e14. [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]
17. American Heart Association. Strategies to Address Socioeconomic and Racial and Ethnic Disparities in Chronic Diseases by Incorporating Food and Nutrition Programs into the Primary Healthcare Setting. 2022. URL: https://www.heart.org/-/media/Files/About-Us/Policy-Research/Policy-Positions/Access-to-Healthy-Food/Medical-Nutrition-Therapy-Policy-Statement-2022.pdf [accessed 2024-11-05]
18. Alolayan RA, Aldisi DA, Hussain DS, Alafif N, Abulmeaty MMA. The efficacy of telehealth versus in-person management delivery in adult patients with obesity. Healthcare (Basel) 2024;12(21):2190 [FREE Full text] [doi: 10.3390/healthcare12212190] [Medline: 39517400]
19. Lundeen EA, Park S, Pan L, O'Toole T, Matthews K, Blanck HM. Obesity prevalence among adults living in metropolitan and nonmetropolitan counties - United States, 2016. MMWR Morb Mortal Wkly Rep 2018;67(23):653-658 [FREE Full text] [doi: 10.15585/mmwr.mm6723a1] [Medline: 29902166]
20. Morris AA, Masoudi FA, Abdullah AR, Banerjee A, Brewer LC, Commodore-Mensah Y, et al. 2024 ACC/AHA Key data elements and definitions for social determinants of health in cardiology: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical data standards. Circ Cardiovasc Qual Outcomes 2024;17(10):e000133. [doi: 10.1161/HCQ.0000000000000133] [Medline: 39186549]
21. Morgan-Bathke M, Raynor HA, Baxter SD, Halliday TM, Lynch A, Malik N, et al. Medical nutrition therapy interventions provided by dietitians for adult overweight and obesity management: an academy of nutrition and dietetics evidence-based practice guideline. J Acad Nutr Diet 2023;123(3):520-545.e10. [doi: 10.1016/j.jand.2022.11.014] [Medline: 36462613]
22. Han M, Rhee SY. Effectiveness of mobile health applications for 5% body weight reduction in obese and overweight adults. J Obes Metab Syndr 2021;30(4):354-364 [FREE Full text] [doi: 10.7570/jomes21014] [Medline: 34853189]
23. Ufholz K, Werner J. The efficacy of mobile applications for weight loss. Curr Cardiovasc Risk Rep 2023;17(4):83-90 [FREE Full text] [doi: 10.1007/s12170-023-00717-2] [Medline: 36974130]
24. Burke LE, Sereika SM, Parmanto B, Bizhanova Z, Kariuki JK, Cheng J, et al. Effect of tailored, daily feedback with lifestyle self-monitoring on weight loss: the SMARTER randomized clinical trial. Obesity (Silver Spring) 2022;30(1):75-84. [doi: 10.1002/oby.23321] [Medline: 34898011]
25. Brunton C, Arensberg MB, Drawert S, Badaracco C, Everett W, McCauley SM. Perspectives of registered dietitian nutritionists on adoption of telehealth for nutrition care during the COVID-19 pandemic. Healthcare (Basel) 2021;9(2):235. [doi: 10.3390/healthcare9020235] [Medline: 33672179]
26. Pursey K, Burrows TL, Stanwell P, Collins CE. How accurate is web-based self-reported height, weight, and body mass index in young adults? J Med Internet Res 2014;16(1):e4. [doi: 10.2196/jmir.2909] [Medline: 24398335]

Abbreviations

AHA: American Heart Association
MNT: medical nutrition therapy
OR: odds ratio
RD: registered dietitian

XSL•FO
RenderX

Edited by L Buis; submitted 18.12.24; peer-reviewed by P Chidipudi, C Aloba, A Jamal; comments to author 20.01.25; revised version received 03.02.25; accepted 07.04.25; published 06.05.25.

Please cite as:
Hu EA, Kelley T, Haryani A
The Impact of Technology-Enabled Medical Nutrition Therapy on Weight Loss in Adults With Overweight and Obesity: Retrospective Observational Study
JMIR Mhealth Uhealth 2025;13:e70228
URL: https://mhealth.jmir.org/2025/1/e70228
doi:10.2196/70228
PMID:40327381

©Emily A Hu, Tommy Kelley, Ajay Haryani. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 06.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Outdoor Exercise Facility–Based Integrative Mobile Health Intervention to Support Physical Activity, Mental Well-Being, and Exercise Self-Efficacy Among Older Adults With Prefrailty and Frailty in Hong Kong: Pilot Feasibility Randomized Controlled Trial Study

Janet Lok Chun Lee[1,2,3], PhD; Arnold Y L Wong[1,2,4,5], PhD; Peter H F Ng[1,6], PhD; S N Fu[1,4], PhD; Kenneth N K Fong[1,5], PhD; Andy S K Cheng[1,7], PhD; Karen Nga Kwan Lee[8], MPA; Rui Sun[1], PhD; Hao Yi Zhang[1], MSc; Rong Xiao[1], MSc

[1]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[2]Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[3]Mental Health Research Centre, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[4]Research Institute for Sports Science and Technology, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[5]Research Centre for Assistive Technology, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[6]Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[7]School of Health Sciences, Western Sydney University, Sydney, Australia

[8]Jockey Club Design Institute for Social Innovation, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

Corresponding Author:
Janet Lok Chun Lee, PhD
Department of Rehabilitation Sciences
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
China (Hong Kong)
Phone: 852 27664354
Email: lcjlee@polyu.edu.hk

Abstract

Background: Engaging in an adequate amount of physical activity (PA) serves as a protective factor against frailty. While previous PA interventions have been effective in improving physical functioning outcomes, they have not consistently succeeded in sustaining PA behavioral changes.

Objective: The primary aim of this pilot randomized controlled trial (RCT) is to explore the feasibility and acceptability of an integrative mobile health (mHealth) intervention among community-dwelling older adults with prefrailty and frailty. The secondary aim was to investigate the potential effects of the intervention on sustaining PA levels and improving mental well-being and exercise self-efficacy in this population.

Methods: A 2-armed pilot feasibility randomized controlled trial was conducted. A total of 38 inactive, community-dwelling older adults (aged>55 years) with prefrailty and frailty were randomized to either the intervention group (n=19), which received 4 weekly educational workshops at a university and a mobile app to support their use of outdoor exercise facilities in their neighborhood, or the control group (n=19), which received 4 weekly health education workshops with exercise experiential sessions tailored for older adults with frailty. To assess the acceptability of the intervention, individual semistructured interviews were conducted with, and a self-developed questionnaire was administered to, 14 participants from the intervention group.

Results: The mean age of the participants was 71.8 (SD 9.34) years, and 24 out of 34 (71%) were female. As many as 34 participants out of 38 (89%) completed the study (18/19 in the control group and 16/19 in the intervention group). Workshop attendance rates were very high in both groups (intervention group, 63/68, 93%, and control group, 72/76, 95%). Self-reported adherence to the unsupervised outdoor practical sessions and engagement with the app was over 65% (36/51, 71%, and 35/51, 69%, in the intervention group. Two adverse events were reported in the intervention group, and none in the control group. As

XSL•FO
RenderX

hypothesized, secondary outcome analyses showed that both groups increased their PA levels immediately after the intervention; however, only the intervention group maintained this increase at the 3-month follow-up. Additionally, favorable changes in mental well-being and exercise self-efficacy were observed in the intervention group. Feasibility and acceptability data also highlighted areas for improvement that should be addressed before a larger trial.

Conclusions: This study provides initial proof-of-concept evidence for the integrative mHealth intervention. However, modifications are needed to enhance user adherence to both the mobile app and the outdoor practice component before proceeding to a larger trial.

Trial Registration: ClinicalTrials.gov NCT06326710; https://clinicaltrials.gov/ct2/show/NCT06326710

(JMIR Mhealth Uhealth 2025;13:e69259) doi:10.2196/69259

KEYWORDS

integrative intervention; mHealth; outdoor exercise facilities; physical activity; built environment; frailty; community-dwelling older adults

Introduction

Background

Population aging is a global phenomenon and an irreversible trend. Statistics show that by 2050, the world population of individuals aged 65 years or older will surpass 1.6 billion [1]. It is also estimated that 20% of the aged population in Asia [2], and 16.6% of older adults in Hong Kong, are frail [3]. Frailty affects various aspects of older adults' lives, including gait, mobility, balance, muscle strength, motor processing, cognition, nutrition, endurance, and physical activity (PA) [4]. Importantly, high levels of frailty increase the risks of adverse health outcomes such as falls, hospitalization, and mortality [5].

Frailty is a modifiable, dynamic process characterized by frequent transitions between different frailty states over time. Research has shown that interventions and health strategies can be effectively used to prevent, delay, or even reverse frailty among community-dwelling older adults [6]. PA interventions are considered one of the main strategies to counteract frailty-related physical impairments in older adults [7]. Engaging in an adequate level of PA has been identified as a protective factor against frailty in this population [8]. Participating in regular PA improves the quality of life of older adults with frailty by enhancing balance and mobility, and by reducing the risks of falls, hospitalization, institutionalization, and mortality [9]. Therefore, it is crucial to develop effective and sustainable PA interventions for community-dwelling older adults with frailty.

According to review studies, most PA interventions designed for older adults with frailty have demonstrated effectiveness in improving functional performance and reducing the incidence of falls. However, they have not consistently succeeded in facilitating sustainable behavioral changes related to PA [10,11]. For instance, a 14-week exercise intervention that used wearable activity trackers and incorporated behavioral change techniques for older adults with frailty succeeded in improving physical performance 3 months after the intervention but did not enhance PA levels [12].

In recent years, outdoor exercise facilities (OEFs) have become increasingly prevalent worldwide as a means of promoting PA among citizens. Interventions have also been developed to educate older adults on using OEFs to improve health outcomes such as muscle strength, balance, mobility, PA levels, and weight control [13,14]. OEF-based interventions have demonstrated effectiveness in supporting sustained behavioral changes related to PA [13,15]. More recently, an OEF-based intervention integrated with mobile health (mHealth) was found to be effective in promoting and sustaining resistance training–specific PA changes in adults [16].

Since 2002, the Hong Kong government has been installing senior-friendly OEFs throughout the territory. Unlike in other regions, where OEFs have been designed and installed for the general adult population, those in Hong Kong have been specifically tailored to meet the needs of older adults. The exercise equipment is low impact and nonresistance, aiming to provide a safe and accessible environment for seniors [17]. Statistics show that senior-friendly OEFs have now been installed at more than 440 locations across Hong Kong [18]. Because of the nonresistance nature of the equipment, exercising with these OEFs has been objectively shown to be of low-to-moderate intensity [19]. Additionally, an in-depth qualitative investigation revealed that older adults with frailty in Hong Kong preferred using OEFs, as they perceived the exercise intensity to be "nonvigorous" [20].

Despite the availability of numerous senior-friendly OEFs in Hong Kong, older adults—especially those with frailty—may hesitate to use or may misuse these facilities due to a lack of proper training. To address this limitation, an integrative mHealth intervention can be implemented to support the proper use of OEFs by community-dwelling older adults with frailty in Hong Kong and to foster sustainable behavioral changes related to PA. Research has shown that integrative mHealth interventions can facilitate the development of sustainable PA-related behavioral changes in the general population [21]. Recent studies also indicate that older adults hold a positive perception of mHealth, provided they receive social support and instructional guidance [22]. The proposed integrative mHealth intervention incorporates educational workshops and supports unsupervised outdoor self-practice sessions in participants' neighborhoods via a mobile app, demonstrating its potential to enhance and sustain PA levels, as well as to improve mental well-being and exercise self-efficacy among inactive and community-dwelling older adults with frailty. The primary aim of this pilot randomized controlled trial (RCT) was to explore the feasibility and acceptability of the integrative

XSL•FO
RenderX

mHealth intervention among community-dwelling older adults with prefrailty and frailty. The secondary aim was to examine the potential effects of the intervention in sustaining PA levels and improving mental well-being and exercise self-efficacy in this population. It was hypothesized that both groups would show improvement in PA immediately after the intervention, but only the intervention group would maintain increased PA levels over time. Additionally, it was hypothesized that the intervention group would demonstrate greater improvements in mental well-being and exercise self-efficacy compared with the control group.

Methods

Trial Design

This study was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) statement extension for pilot trials [23]. Please refer to Multimedia Appendix 1 for the CONSORT checklist. This study was a prospective, 2-armed, single-blinded (participant-blinded), parallel-group pilot RCT with an allocation ratio of 1:1. A qualitative interview was conducted with participants in the intervention group following the pilot RCT. The trial was conducted between May and September 2024 at a university in Hong Kong and was registered on ClinicalTrials.gov (March 22, 2024).

Participants

Participants were recruited in April 2024 using convenience sampling. Recruitment was conducted through public talks delivered online, at community health centers, and at health promotional booths. Enrollment screening was carried out by the research assistant according to the inclusion and exclusion criteria (Textbox 1).

Textbox 1. Inclusion and exclusion criteria for participants.

1. Inclusion criteria

- Community-dwelling older adults aged 55 years or over.
- Self-report of frailty or prefrailty as determined by a score of ≥1 and ≤5 on the Chinese version of the FRAIL (Fatigue, Resistance, Ambulation, Illness, and Loss of Weight) scale [24].
- A score of 0 out of 7 on the Physical Activity Readiness Questionnaire [25].
- Use a smartphone on a daily basis.

2. Exclusion criteria

- Those who have already achieved an adequate level of daily physical activity, defined as engaging in at least 150 minutes of moderate-to-vigorous physical activity per week by a single-item question with a dichotomous response (ie, yes/no). They were also asked to provide their exercise habits to validate their response to this single-item question.

Sample Size

This was a pilot feasibility trial; therefore, it was not designed to test the effect of the intervention, which requires a fully powered study. Instead, this pilot trial aimed to estimate the potential effects in a small sample, providing a reference for calculating the sample size of the main trial.

The determination of the sample size for this study was based on 2 key factors: Hertzog's sample size recommendation [26] and insights from recent mHealth studies using a pilot RCT design. Hertzog suggested that pilot studies should include between 10 and 40 participants per group [26]. While this range is broad, the research team also considered recent mHealth studies focusing on PA behavioral change to guide the decision on sample size. Recent mHealth studies with a similar focus on PA behavior reported that recruiting and randomly allocating between 33 and 42 participants to the intervention and control groups was sufficient to assess preliminary differences between groups [27,28]. Considering these factors, this study aimed to recruit a total of 30-40 participants for random allocation. It is important to note that this study was not powered to detect significant changes in secondary outcomes; instead, the mean difference between groups observed at the 3-month follow-up, along with the SD, will be used to estimate the sample size for a future large-scale definitive RCT.

Intervention

The integrative mHealth intervention group comprised 3 components: (1) a mobile app [29], (2) educational workshops, and (3) unsupervised outdoor self-practice sessions. Participants in the intervention group attended 4 weekly 1-hour educational workshops led by an interventionist with expertise in exercise and health. These workshops covered various topics, including how to use the app to identify OEFs in their neighborhood and how to correctly use each OEF. Additional workshop content addressed risk management during outdoor exercise, the association between frailty and exercise, selecting appropriate exercise intensity, strengthening and balance exercises, postural awareness, pain management during exercise, and principles for modifying exercises when using OEFs. Over the 4-week intervention period, participants were encouraged to use the OEFs with the assistance of the app to learn the correct usage of the facilities. During the workshops, they had the opportunity to share their experiences using the app and discuss any challenges encountered during the unsupervised outdoor practice sessions. Their adherence to the outdoor practice sessions was recorded by the research assistant at each workshop.

The Outdoor Rehab-Fit App

The Outdoor Rehab-Fit App was developed by PA researchers, clinician-scientists, and computer scientists at the Department of Rehabilitation Sciences, The Hong Kong Polytechnic

XSL•FO
RenderX

University. The app offers (1) an OEF library, (2) OEF locations, and (3) risk management guidance for outdoor exercise, among other features. For each piece of outdoor exercise equipment in the library, the app provides the health benefits of using the equipment, tips from a physiotherapist and an occupational therapist, and advice for caregivers—all presented in text format. This additional health information, tailored to each piece of equipment by exercise and health professionals, aims to enhance users' perceived benefits and outcome expectancies of using the equipment for PA. Additionally, an audio guide provides step-by-step voice-over instructions to help users learn about the OEFs. Short videos for each piece of equipment are also available to remind users of the correct postural alignment while exercising. These multimedia formats improve users' self-efficacy in performing the exercises.

Active Control

Participants in the active control group attended four 60-minute health education workshops led by an interventionist with expertise in exercise and physical health management. These workshops focused on the benefits of various exercises for older adults, particularly those who are frail. In addition to health knowledge, participants were given opportunities to try different modified exercises suitable for older adults with frailty, such as chair yoga, seated Baduanjin, Otago exercises, and seated fitness exercises. The structure and content of these workshops mirrored common community-based programs promoting PA in older adults. This comparator was selected to compare the integrative mHealth intervention against existing practices. Additionally, by providing structured sessions with the same exercise theme to both groups, the comparator minimized potential confounding factors such as participant expectations. Finally, the active control ensured ethical fairness by offering all participants support for PA.

Outcomes

Primary Outcomes

Overview

The primary outcomes were the feasibility, safety, and acceptability of the intervention.

Feasibility

The feasibility of the integrative mHealth intervention was assessed based on the recruitment rate, intervention completion rate, and attrition rate. The recruitment rate refers to the percentage of individuals who signed up for the intervention out of all those who were approached. The intervention completion rate reflects the extent to which participants completed all 3 components (app, workshops, and outdoor self-practice) in the intervention group, or attended the education talks in the control group, during the 4-week intervention period. Intervention completion was recorded using a self-developed form and calculated as the percentage of intervention components completed (Multimedia Appendix 2).

Safety

Any adverse events occurring in either group during the intervention and follow-up periods were recorded. These events included incidents such as falls and muscle or joint pain, as well as serious occurrences such as difficulty breathing, new or persistent chest pain, or sudden changes in consciousness that require immediate medical attention.

Acceptability

The acceptability of the app and the integrative mHealth intervention were evaluated using a self-developed questionnaire (Multimedia Appendix 3) and a semistructured interview. The interview guide was developed based on a similar mHealth intervention study [30] (Multimedia Appendix 4). The interviews gathered participants' overall views of the intervention, perceived barriers and facilitators, and recommendations, and were conducted either face-to-face or online.

Secondary Outcomes

Overview

The secondary outcomes included measures of PA behaviors (utilization of OEFs and overall PA), exercise self-efficacy, and mental well-being. These measures were selected to assess the intervention's effectiveness across multiple domains.

PA Behaviors

- Participants' utilization of OEFs was assessed using a self-developed questionnaire that collected information on the frequency, duration, and types of OEFs typically used.
- The Rapid Assessment of Physical Activity Scale (RAPA) [31] was used to measure self-reported PA. RAPA is commonly used in community-based interventional studies [32,33]. It consists of 9 yes-or-no items and evaluates PA levels ranging from sedentary to regular vigorous activity, as well as exercise types. The questionnaire has 2 sections: $RAPA_1$ and $RAPA_2$. $RAPA_1$ includes 7 questions related to the level of leisure-time aerobic activity, illustrated with pictures depicting light, moderate, and vigorous activities. $RAPA_2$ consists of 2 questions focusing on strength and flexibility training. RAPA has 81% sensitivity, 69% specificity, and a 77% positive predictive value. It is significantly (P<.001) correlated with the frequently used Community Healthy Activities Model Program for Seniors Physical Activity Questionnaire (CHAMPS) for older adults. The PA level assessed by $RAPA_1$ can be treated as a continuous variable to observe changes over time [33]. This study used only the data from $RAPA_1$ to examine changes in aerobic activity over time. RAPA was chosen because it requires less time to administer and imposes a lower cognitive load compared with CHAMPS. Given that the participants are older adults with frailty, RAPA was selected to minimize their burden and reduce respondent fatigue.
- Objective measurement of moderate-to-vigorous intensity physical activity (MVPA) was collected using an accelerometer (ActiGraph GT3X, ActiGraph LLC). This activity monitor has been validated and shown to have good criterion validity against calorimetry in distinguishing MVPA from non-MVPA during treadmill walking at various speeds [34]. The monitor was worn on the participant's nondominant wrist for 7 consecutive days during waking hours, with a minimum of 10 hours of wear

XSL•FO
RenderX

time each day [35,36]. A 7-day monitoring period is recommended to capture typical daily PA patterns [35,37]. Literature indicates that wrist-worn activity monitors outperform hip-worn monitors when participants walk at speeds of 3-6 km/h [34]. Participants received instructions on how to properly wear the monitor on their nondominant wrist before the assessment. Participants were allowed to remove the ActiGraph during specific activities such as bathing, swimming, and sleeping. They also maintained a daily wear-time log to record the monitor's on and off times throughout the monitoring period. Participants were instructed not to engage in any additional PA beyond their usual routine while wearing the monitor. A research assistant contacted participants on the first or second day after device distribution to ensure proper use [38]. The ActiGraph devices were initialized to collect raw data at a sampling rate of 60 Hz. After screening for data validity, only complete and valid data sets were processed using ActiLife software in low-frequency extension mode [39-41]. Valid data were defined as wear time exceeding 10 hours per day for at least 5 days, or totaling more than 50 hours across 4 days [35]. Activity counts per minute from the monitor were converted into minutes of MVPA using age-specific cutoff points established in the literature, specifically 4117.7 counts or more per minute [34]. These cutoffs were derived from a previous study [34] validating ActiGraph use in identifying PA intensity levels among the Hong Kong older adult population.

- The Chinese version of the Exercise Self-Efficacy Scale [42] was used to assess participants' confidence in maintaining regular exercise. This scale comprises 9 items, each rated on an 11-point Likert scale from 0 (not confident) to 10 (very confident). The mean score of these items, ranging from 0 to 10, represents the overall exercise self-efficacy, with higher scores indicating greater confidence in regularly engaging in exercise.
- The mental well-being of participants was assessed using the Chinese version [43] of the 7-item Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) [44]. This scale addresses both hedonic and eudaimonic aspects of well-being, with items rated on a 5-point Likert scale from 1 (none of the time) to 5 (all the time). The mean score, calculated by averaging responses across the 7 items, reflects the overall mental well-being, with higher scores indicating greater well-being.

Randomization and Blinding

Randomization was performed after the completion of the screening. Participants were randomly allocated to either the intervention or active control group on a 1:1 basis using block randomization by a research assistant. The randomization process did not consider whether participants were couples. Block sizes of 4 or 6 were randomly generated using random number generator software [45].

Blinding

Participants were blinded to their group allocation and informed only that they had been assigned to 1 of 2 groups to evaluate the feasibility and effects of the exercise programs. All assessments were conducted by trained assessors with health sciences backgrounds.

Statistical Analysis

Quantitative Data

Descriptive analyses were conducted to summarize demographic characteristics, frailty levels, chronic disease diagnoses, hospitalizations, history of falls in the past 12 months, and use of outdoor walking aids. Means and SDs were used to report continuous parametric data, while medians and IQRs summarized continuous nonparametric data. Categorical data were presented as frequencies and percentages. Recruitment rate, intervention completion rate, attrition rate, and attendance rate were reported as percentages. Adverse events were recorded and described narratively.

Although this study was not specifically powered to detect statistically significant differences between groups for the secondary outcomes, analyses were conducted to explore these differences. This approach aimed to inform the selection of outcome measures for a future fully powered clinical trial and to facilitate preliminary exploration of potential intervention effects. Data analyses followed the intention-to-treat principle. Descriptive statistics were used to examine statistical trends. Generalized estimating equations were used to analyze temporal changes and between-group differences in self-reported and objectively measured PA levels, exercise self-efficacy scores, and mental well-being scores. To account for multiplicity in pairwise comparisons of secondary outcomes, least squares difference correction was applied. Effect sizes were estimated using Cohen *d*, adjusted for generalized estimating equations analysis, based on the following formula: $\beta_{interaction}/SD_{residual}$.

Qualitative Data

Interviews were audio-recorded and transcribed verbatim. Thematic analysis was conducted following the 6-phase framework outlined by Braun and Clarke [46]. Data were inductively coded and then deductively organized into 4 key domains: (1) overall perception of the intervention, (2) facilitators, (3) barriers, and (4) suggestions for improvement. NVivo version 14 (QSR International Pty Ltd) was used to support data management and analysis. To enhance the credibility of the findings, analyst triangulation was applied. The first author conducted the initial coding and analysis, while all coauthors reviewed and refined the themes. Any discrepancies were resolved through discussion and consensus.

Ethics Considerations

Ethical approval has been obtained from The Hong Kong Polytechnic University Institutional Review Board (PolyU IRB; reference number HSEARS20230306001-3). All participants will be required to sign a consent form before participation. Where a participant lacks the capacity to provide consent, a nominated representative will sign the consent form, along with obtaining assent from the participant.

Any information obtained in this study will remain strictly confidential and be used for research purposes only. Codes, not names, are used on all reports and publications related to this study to protect confidentiality. All collected questionnaires

XSL•FO
RenderX

will be kept in locked cabinets. The electronic data set will be stored in encrypted computer storage. Data containing personal identifiers will be kept for a maximum of 3 years after publication of the first paper upon the end of the study.

Dissemination Policy

Manuscripts resulting from this trial will be published in academic journals or abstracts of papers will be presented in academic conferences. Authorship eligibility includes (1) substantial contributions to the study design or the process of data acquisition, analysis, or interpretation, (2) drafting or revising the manuscript, and (3) approving the final manuscript. There is no intention to use professional writers.

Results

Demographic and Other Characteristics of the Participants

The demographic and health characteristics of the participants are presented in Table 1. No significant differences ($P<.05$ in all cases; see Table 1 for the complete P values) were observed between groups at baseline, except for the incidence of falls in the past 6 months ($P=.04$). Additionally, due to attrition occurring immediately after group allocation, there was an imbalance in gender distribution between the groups. To enhance the robustness of the results, both gender and incidence of falls were included as covariates in the statistical analyses.

XSL•FO
RenderX

Table 1. Demographic characteristics of the participants.

Characteristics	Total, n (%)	Intervention (n=16), n (%)	Control (n=18), n (%)	*P* value
Age				.09
Mean (SD)	71.8 (9.34)	69.0 (8.18)	74.28 (9.82)	
Median	70	66.5	72.5	
Gender				.20
Male	10 (29)	3 (19)	7 (39)	
Female	24 (71)	13 (81)	11 (61)	
Education				.38
No formal education	2 (6)	0 (0)	2 (11)	
Primary level	5 (15)	1 (6)	4 (22)	
Secondary level	15 (44)	10 (63)	5 (28)	
Tertiary level	8 (24)	2 (13)	6 (33)	
Postgraduate level	4 (12)	3 (19)	1 (6)	
Living arrangement				.93
Living alone	11 (32)	5 (31)	6 (33)	
Living with family	23 (68)	11 (69)	12 (67)	
Weight (kg)				.75
Mean (SD)	58.4 (9.36)	59.0 (8.19)	57.9 (10.5)	
Median	59.5	59.5	59.5	
Height (cm)				.35
Mean (SD)	155.4 (7.14)	157.6 (7.56)	155.2 (6.77)	
Median	155.5	155.0	155.5	
BMI (kg/m^2)				.82
Mean (SD)	23.8 (2.48)	23.7 (2.04)	23.9 (2.87)	
Median	23.1	23.0	23.6	
Frailty level (FRAIL[a] scale)				.75
Score=1	25 (74)	12 (75)	13 (72)	
Score=2	4 (12)	3 (19)	1 (6)	
Score=3	4 (12)	1 (6)	3 (17)	
Score=4	1 (3)	0 (0)	1 (6)	
Chronic disease				.82
Yes	24 (71)	11 (69)	13 (72)	
No	10 (29)	5 (31)	5 (28)	
Use walking aids outdoors				.48
Wheelchair	2 (6)	1 (6)	1 (6)	
Walking stick	5 (15)	1 (6)	4 (22)	
None	28 (82)	14 (88)	13 (72)	
Hospitalized in the past 12 months				1.00
Yes	0 (0.0)	0 (0.0)	0 (0)	
No	34 (100)	16 (100)	18 (100)	
Fallen in the past 6 months				.04
Yes	4 (12)	0 (0)	4 (22)	

XSL•FO
RenderX

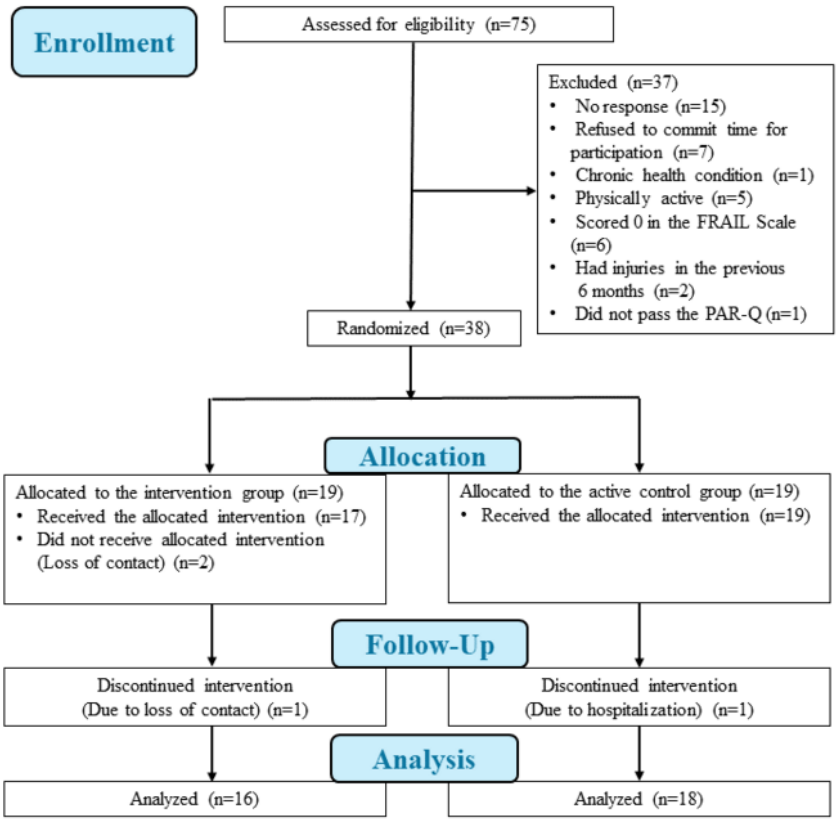
Characteristics	Total, n (%)	Intervention (n=16), n (%)	Control (n=18), n (%)	*P* value
No	30 (88)	16 (100)	14 (78)	

[a]FRAIL: Fatigue, Resistance, Ambulation, Illness, and Loss of Weight.

Feasibility of the Intervention

Of the 75 individuals, 38 (51%) were recruited. Of the 75 individuals who met the eligibility criteria, 38 were randomized into the study groups. In the intervention group, 17 of the 19 participants received the integrative mHealth intervention, while 2 participants were lost to follow-up before the intervention began. All 19 participants allocated to the active control group received the control sessions. The attrition rate was less than 20% (3/19, 16%) in the intervention group and under 10% (1/19, 5%) in the active control group. Among the participants who dropped out, the reasons included loss of contact (n=3) and hospitalization (n=1; see Figure 1). The overall intervention completion rate in the integrative mHealth intervention group was nearly 80% (134/170, 78.8%). The average completion rates for individual components were as follows: workshop attendance, 63 out of 68 (93%); self-reported adherence to unsupervised outdoor practice sessions, 36 out of 51 (71%); and self-reported mobile app engagement, 35 out of 51 (69%). The active control group, which included only the workshop component, had a completion rate of over 90% (72/76, 95%). Two adverse events were reported in the intervention group, while none was reported in the control group. In the intervention group, 1 participant experienced an upper body slipping incident while using faulty exercise equipment. Another participant reported knee and lower back pain after using the stepper and twister; the discomfort was alleviated by modifying the exercise duration and reducing the range and speed of movement. No other adverse events related to the interventions were reported by participants.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart.



Acceptability of the App and the Integrative mHealth Intervention

A total of 14 participants from the intervention group, including 11 women and 3 men aged 61-77 years, with frailty levels ranging from 1 to 3, took part in the semistructured interviews. Additionally, 13 of these participants completed the self-developed questionnaire on the perceived usefulness of the app. The questionnaire results showed that all respondents agreed or strongly agreed that the app improved their exercise performance, and at least 12 (86%) of them agreed or strongly agreed that they intended to continue using the app to support their outdoor workouts. All respondents found the app's "Introduction by an Exercise Specialist," "Tips from a Physiotherapist," "Tips from an Occupational Therapist," and "My Favorites" features useful or extremely useful (Tables 2 and 3). Themes and subthemes related to participants' overall perceptions of the intervention, as well as the barriers,

XSL•FO
RenderX

facilitators, and suggestions for future improvements, were identified through qualitative analysis (Table 4). Participants appreciated the structured design of the intervention, which included the app, workshops, and follow-ups. They reported increased motivation and confidence in using the free-of-charge OEFs to self-manage their physical health. The intervention also enhanced their awareness of the health benefits associated with OEF use. Barriers and facilitators related to the various components of the integrative mHealth intervention—the app, outdoor practice, and workshops—were identified. For the app component, participants reported barriers such as small image sizes, the inability to geolocate certain outdoor exercise equipment, and the unavailability of some equipment listed in the app. Barriers to the outdoor practice component included equipment being out of order, hygiene concerns, lack of storage for personal belongings, disruptive behavior from other users, unfavorable weather conditions, and social factors. By contrast, the app's content and the workshop's delivery mode were viewed as key facilitators of engagement. To enhance the intervention, participants suggested several improvements to the app: enabling image enlargement; integrating a geolocation feature for exercise equipment, including demonstrations of both correct and incorrect exercise techniques; and adding an exercise reminder function. Regarding the outdoor practice component, participants suggested that incorporating a supervised outdoor practical session would be beneficial to support their exercise adherence and technique.

Table 2. Perceived usefulness of the app (N=13).

Usefulness	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
Part 1: User's exercise performance					
Enhanced my knowledge of outdoor exercise facilities.	0 (0)	0 (0)	0 (0)	6 (46)	7 (54)
Allowed me to execute exercise movements accurately when using outdoor exercise facilities.	0 (0)	0 (0)	0 (0)	4 (31)	9 (69)
Allowed me to execute exercise movements safely when using outdoor exercise facilities.	0 (0)	0 (0)	0 (0)	7 (54)	6 (46)
Enhanced my exercise efficiency, enabling me to effectively achieve health benefits, such as improved balance, muscle strength, and overall subjective well-being.	0 (0)	0 (0)	0 (0)	7 (54)	6 (46)
Assisted me in reaching my physical activity goals.	0 (0)	0 (0)	0 (0)	8 (62)	5 (38)
Part 2: User's intention to use the app to support outdoor workouts					
When I use the outdoor exercise facilities, I intend to incorporate the Outdoor Rehab-Fit App into my physical activity routine.	0 (0)	0 (0)	0 (0)	6 (46)	6 (46)
I like to use the Outdoor Rehab-Fit App.	0 (0)	0 (0)	0 (0)	8 (62)	5 (38)
I will use the Outdoor Rehab-Fit App both during my outdoor workouts and whenever I want to learn more about the outdoor exercise facilities.	0 (0)	1 (8)	1 (8)	7 (54)	4 (31)

XSL•FO
RenderX

Table 3. Perceived usefulness of the app content (N=13).

Perceived usefulness	Extremely useless, n (%)	Useless, n (%)	Neutral, n (%)	Useful, n (%)	Extremely useful, n (%)
Part 3: Mobile app content					
Introduction by an exercise specialist	0 (0)	0 (0)	0 (0)	8 (62)	5 (38)
Tips from a physiotherapist	0 (0)	0 (0)	0 (0)	5 (38)	8 (62)
Tips from an occupational therapist	0 (0)	0 (0)	0 (0)	5 (38)	8 (62)
Tips for a caregiver	0 (0)	1 (8)	2 (15)	7 (54)	3 (23)
Audio guide	0 (0)	0 (0)	2 (15)	7 (54)	4 (31)
Demonstration video	0 (0)	0 (0)	0 (0)	8 (62)	4 (31)
Detailed instructions	0 (0)	1 (8)	0 (0)	9 (69)	2 (15)
Educational video	0 (0)	0 (0)	0 (0)	5 (38)	7 (54)
Safety precautions and risk management	0 (0)	0 (0)	0 (0)	6 (46)	6 (46)
Physical fitness examination questionnaire	0 (0)	1 (8)	0 (0)	8 (62)	3 (23)
Facility location	0 (0)	1 (8)	4 (31)	7 (54)	1 (8)
My favorites	0 (0)	0 (0)	0 (0)	6 (46)	7 (54)
The “By Categories” function	0 (0)	0 (0)	0 (0)	5 (38)	7 (54)

XSL•FO
RenderX

Table 4. Acceptability of the intervention.

Themes and subthemes	Illustrative quotes
The overall perception of the intervention	
Appreciate the structure of the intervention, which consists of an app, workshops, and follow-ups	"You need to have all three components, and [each component] complements each other. Once you're familiar with the app, you can go out and play [with the exercise equipment], right? But I think you still need to attend workshops. Workshops could explain things in more detail. Sometimes, if the app explains things too thoroughly, you might not want to read it." [P37]
	"The organization [of the intervention] is compact. It is good. We can ask questions on-site. The physiotherapist and the instructor can show us the exercise movements, which help to correct our previous misconceptions and serve as good encouragement. This gives us the confidence to exercise whenever we have free time or at a scheduled time." [P29]
Increased motivation and confidence in using the free-of-charge OEFs to self-manage physical health	"I really have a lot of issues all over my body, various aches and pains. Although there may be no way to improve them further, I'm still willing to try. Like before, as they said, I'd walk in the park, and if I had time, I'd do a couple of things. But after looking at the app at home, I realized that by taking steps, you can learn the correct way to improve. I really recommend this [the intervention], especially the medical and risk management aspects. For me, it's really good. From my personal experience, my legs used to be painful when I pedaled the cycling machine, but now, by positioning my toes correctly and not using excessive force, my legs hurt less. My shoulder, too—I don't know if it was a frozen shoulder or what, but after a year of occupational therapy, it was the same. Using the correct stretching technique, my upper arms feel much more relaxed now. This is one of the things that truly met my expectations." [P25]
	"Basically [now] I would attempt to try [the outdoor exercise equipment], explore them, and look at their introductions in the app. For example, if my shoulder feels painful, using these two [outdoor exercise equipment] may help me. For the legs, these [outdoor exercise equipment] are useful for the knees and muscle strength. [If] I don't need those right now, I can skip them and focus on the ones that help me." [P7]
	"Ah, because I knew these [outdoor exercise equipment] were available in the parks, but I did not know there was an app that explained them so clearly. Knowing this [information] gives me more confidence when exercising, and I know it helps me, especially with muscle training and cardiovascular functions. So I think it's okay." [P37]
	"Regarding posture and which exercises could help with certain pain issues in your body, you can focus on those [outdoor exercise equipment] that target your specific needs." [P44]
Improved awareness of the health benefits of using outdoor exercise facilities	"Before joining this intervention, I just passed by the government facilities in parks without knowing much about them and thought they were too simple and childish. I wasn't interested in trying them out. But after learning about this intervention and its background, I realized that these facilities are actually thoughtfully designed for the older adults. I didn't know about the government's intentions [before], but after attending talks delivered by professionals, I felt we shouldn't waste the effort put into installing these facilities. [If we are not using these facilities], these facilities will become rusty. It seems like such a waste. So, I decided to learn more and engage with them. Even though the movements seem simple, they actually provide good support. On another level, I realized I was too proud, thinking these simple movements were easy. [Now I realize] if you use the correct posture and are willing to spend time understanding how the outdoor exercise equipment can improve your capabilities, it's really beneficial." [P45]
	"I think it's quite good because you can understand more about the [outdoor sport] facilities available near you." [P6]
Barriers	
App component	
Size of the videos	"Some videos [size] are quite small and can't be enlarged, so they're not very clear. The equipment shown [in the video] is very small. Even...our phones are small, unless you're using an iPad. If those [videos] could be enlarged, you could see more details." [P25]
Inability to geolocate the exercise equipment	"The biggest problem is that not all types of outdoor exercise equipment [shown in the app] are available everywhere. I mean, the app categorizes things like strength training and aerobic exercises. But if you want to locate the specific outdoor exercise equipment, this app doesn't have that information. For example, if I want to do strength training, which park has these facilities? Where are the facilities for various types of training? It doesn't have this information." [P8]
Unavailability of some exercise equipment	"I think the barrier is...I remember once I went to an outdoor area by the seaside, but I can't remember the exact location. The explanations on the facilities were blurred, and coincidentally, this [outdoor exercise equipment] wasn't covered in your app." [P16]
Outdoor practice component	
Equipment being faulty	"And those facilities, especially the ones with pull ropes [upper limb stretcher], are often faulty." [P8]
Hygiene concerns	"Because those places are touched by many people, you need to bring disinfectant wipes to clean them first. When I talked to my friends, they said, 'Are you kidding me? So many people have used them, the hygiene situation [is relatively bad]...' So you always have to remember to bring disinfectant wipes." [P3]

XSL•FO
RenderX

Themes and subthemes	Illustrative quotes
Storage problem for belongings	"When I visit the [outdoor exercise facilities], I usually keep my backpack on while I am exercising. It might look a bit awkward, but I don't want to put it on the ground. Not just in parks, but even in other places, I don't like putting my things on the ground. I don't know why, but I'd rather keep it on me than placing it down." [P3]
Other users' behavior	"Those facilities are installed outdoors, so they're often affected by the weather and the environment. For example, if it's too sunny, it can have an impact [reducing people's willingness to use them]." [P7]
Weather	"Those facilities are installed outdoors, so they're often affected by the weather and the environment. For example, if it's too sunny, it can have an impact [on reducing people's willingness in using them]." [P7]
Social factors	"If a group of friends or someone like that goes to the park, then I'd be more interested in going there to do that activity. But for now, it's just me alone, so I don't have as much motivation." [P7]
Facilitators	
App component	
App content facilitates revision and consolidation of new knowledge	"And I think the advantage of having it [the app] is that you can check it anytime, anywhere. For example, if the weather is nice today, I can look at it today or tomorrow. If I go to the park to exercise, I can take it out to look at, or I can look at it before going down. So, there's no time or space limitation. But if you only attend a class, then I have to rely on my own memory, and I might not remember everything." [P16]
Professional content	"I want to use physiotherapy, occupational therapy, or do exercises myself to improve my physical health, so I try to look for more information on these things. For example, I don't know if your exercise program will help me, but after listening, I might think, 'Oh, this is useful and can help me'." [P7]
Workshop component	
Delivery mode	"When it's an in-person class, like when the coach teaches us how to warm up or when the interventionist shows us [correct] postures, I think this face-to-face instruction is really good." [P37]
Suggestions	
App component	
Enable image to be enlarged	"For example, can the image be enlarged? I feel like when I'm exercising, everything looks far away and really small." [P1]
Enable locating certain equipment	"For example, if I want to do strength training, it would be great if you could provide information on where to find these outdoor fitness facilities." [P8] "Actually, I've suggested geo-tagging, so you know where the locations are. The reason is quite simple: I initially had a goal, which is also simple. For example, I want to improve certain joints; I'm not someone who doesn't move at all..." [P39]
Incorporate both incorrect and correct exercise demonstration	"It's like we need to come here [to attend workshop], but in the future, the workshop will not be available. So, having correct and incorrect demonstrations in the video would allow more in-depth understanding [of how to exercise correctly]. Of course, it would be great if we could all sit here together [to learn the correct posture], right? But I might not have that opportunity. Or in the video, you could include some incorrect demonstrations along with the correct ones, which would help reinforce this [learning outcomes]." [P25]
Reminders	"Yes, yes, yes. It's like when someone asks you, 'Oh, have you been to so-and-so recently?' It reminds me. Sometimes people's mindset is quite interesting. There are certain times when people will ask me if I'm doing something, and I have to commit to doing it. I can't say I have done it when I actually have not done. So, it becomes a way for me to motivate myself to take action." [P45]
Outdoor practice component	
Provide supervised outdoor practical sessions	"I think that [performing] any type of exercise requires attention to your posture, application of force, and the steps involved. I believe it's important to have proper guidance to truly achieve the effectiveness of the exercise." [P8] "For example, if you all have time, you could organize a group to go outdoors and do something together. Like, 'Hey, on June 2nd, we have an outdoor activity from 2 to 3 PM. Is everyone interested?' That way, we can gather at Victoria Park and do the activities together in person." [P45]

Results of Secondary Outcomes

Statistically significant improvements from baseline to the 3-month follow-up were observed in 3 of the 5 outcomes in the integrative mHealth intervention group, specifically, usage of OEFs (P<0.001), self-reported PA (P=0.002), and mental well-being (P=0.002; Figures 2-4), while no significant changes were found in any of the 4 outcomes for the active control group.

As hypothesized, the intervention group demonstrated a significant increase in OEF usage, reflected in both weekly frequency (P<.001, W=0.628) and duration of PA (P<.001, W=0.635), compared with the control group. Additionally, the intervention group showed a significant improvement in self-reported PA level as measured by the $RAPA_1$ aerobic score (Wald $\chi^2_{2.96}$=21.02, P<.001, effect size 1.67). Although objectively measured weekly minutes of MVPA increased from

XSL•FO
RenderX

baseline to the 3-month follow-up (147.8, 166.75, and 174.29 minutes per day at T0, T1, and T2, respectively), this change was not statistically significant ($\chi^2_{2.87}$=2.29, *P*=.32, effect size 0.15) when compared with the control group. Furthermore, a statistically significant group-by-time effect was observed in favor of the integrative mHealth intervention on mental well-being (Wald $\chi^2_{2.81}$=7.24, *P*=.03, effect size 0.29), while changes in exercise self-efficacy from baseline to the 3-month follow-up were not statistically significant for T0, T1, and T2 (5.67, 6.03, and 5.45, respectively; Wald $\chi^2_{2.81}$=4.42, *P*=.11, effect size 0.63; Multimedia Appendix 5). Post hoc analyses indicated that the intervention group had significantly higher exercise self-efficacy immediately after the intervention (mean difference 1.42; 95% CI 0.14-2.71; *P*=.03; effect size 0.80) and significantly better self-reported PA levels at the 3-month follow-up (mean difference 1.31; 95% CI 0.52-2.11; *P*<.01; effect size 1.15) compared with the control group (Table 5).

Figure 2. Usage of outdoor exercise facilities (OEFs).

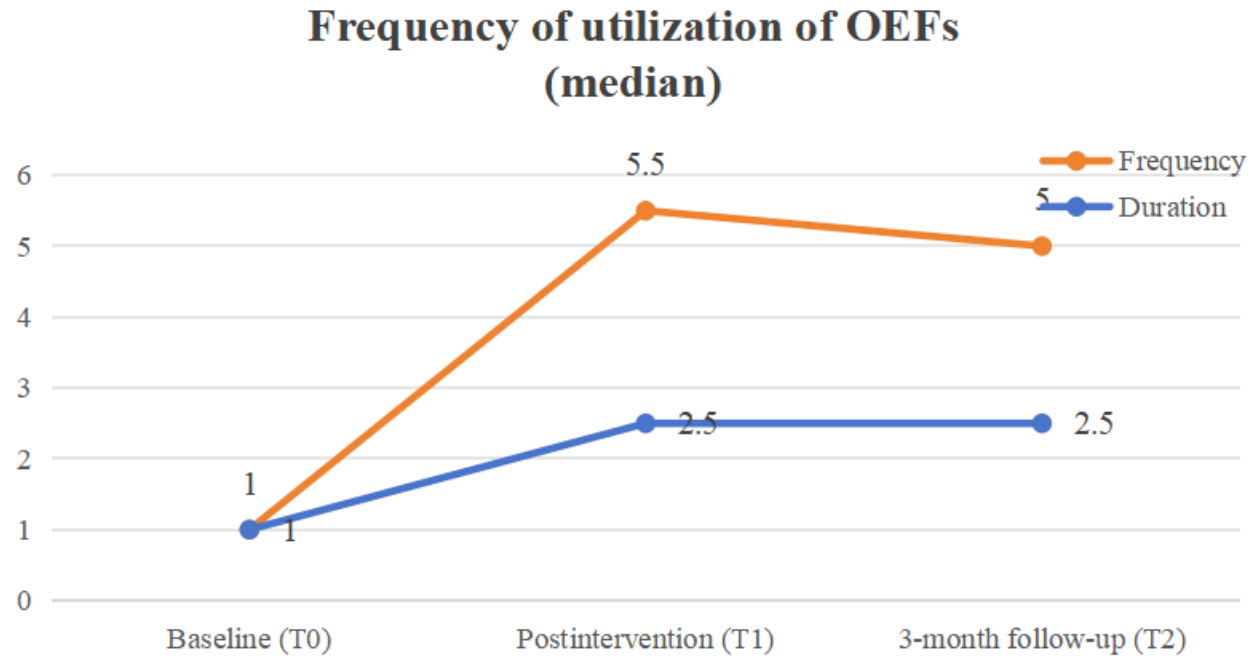
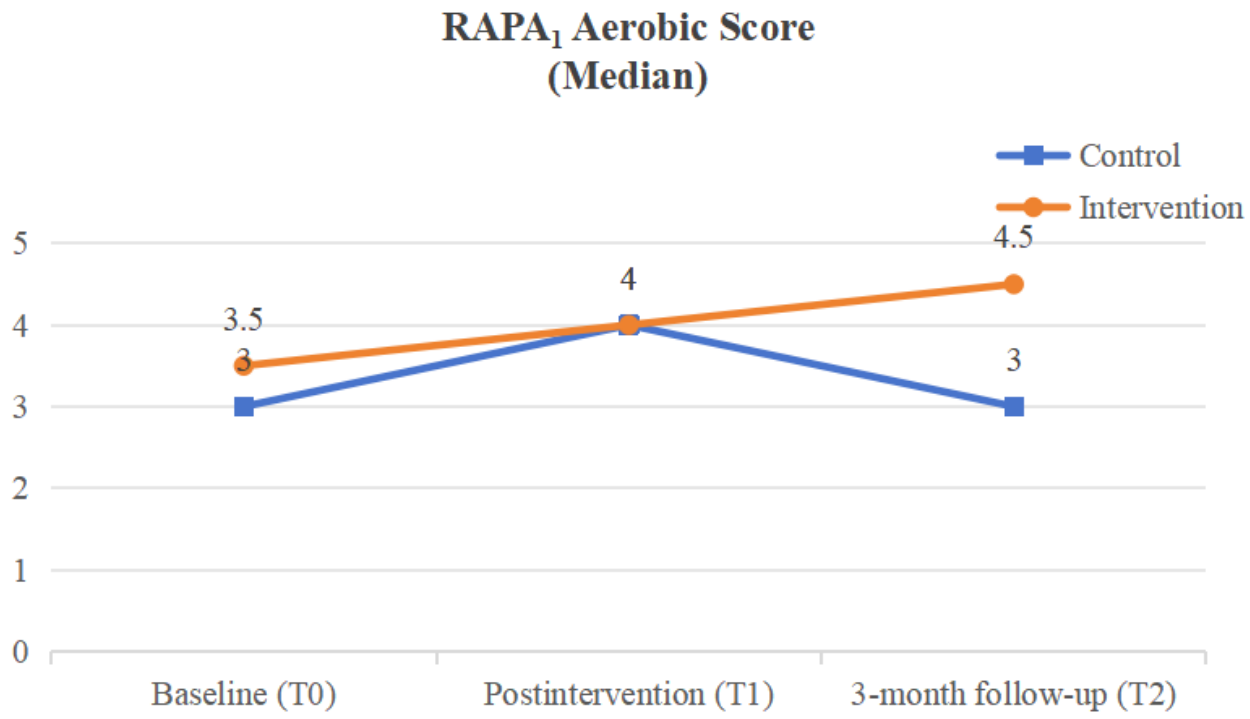


Figure 3. Self-reported physical activity level. RAPA: Rapid Assessment of Physical Activity.



XSL•FO
RenderX

Figure 4. Mental well-being.

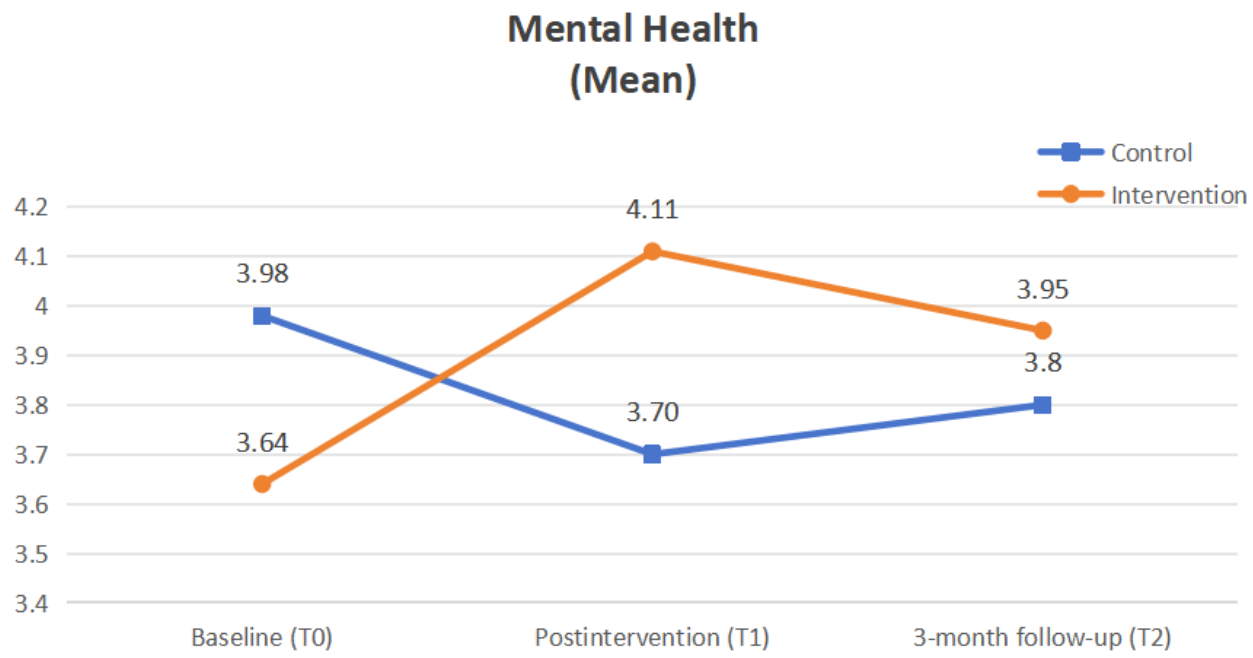


Table 5. Effects of the intervention at postintervention (T1) and the 3-month follow-up (T2).

Outcomes measurement	Postintervention (T1)			3-month follow up (T2)		
	Difference (95% CI)	*P* value	Effect size	Difference (95% CI)	*P* value	Effect size
PA[a] level						
Self-reported PA (RAPA[b]): $RAPA_1$ aerobic score	0.21 (–0.65 to 1.06)	.17	0.17	1.31 (0.52 to 2.11)[c]	<.01[c]	1.15[c]
Objectively measured PA: MVPA[d] time (minutes/week)	36.11 (–52.78 to 125.00)	.20	0.30	53.15 (–41.25 to 147.56)	.10	0.41
Exercise self-efficacy						
SEE-C[e] scores	1.42 (0.14 to 2.71)[c]	.03[c]	0.80[c]	0.37 (–0.71 to 1.46)	.50	0.25
Mental well-being						
SWEMWBS[f] scores	0.27 (–0.29 to 0.82)	.35	0.36	0.07 (–0.37 to 0.52)	.74	0.11

[a]PA: physical activity.
[b]RAPA: Rapid Assessment of Physical Activity Scale.
[c]Significant in favor of the intervention group.
[d]MVPA: moderate-to-vigorous physical activity.
[e]SEE-C: Exercise Self-Efficacy Scale—Chinese.
[f]SWEMWBS: Short Warwick-Edinburgh Mental Well-Being Scale.

Discussion

Principal Findings

This study aimed to assess the feasibility of conducting a full-powered RCT to evaluate the integrative mHealth intervention and to identify any necessary modifications before its implementation. The main findings indicated a high recruitment rate, a marginally acceptable intervention completion rate, and an acceptable attrition rate. Overall, the intervention was well accepted, as reflected in the semistructured interviews. Participants in the intervention group identified barriers and facilitators related to the integrative mHealth intervention and offered suggestions for its improvement. Exploratory analyses of the intervention's potential effects supported the study hypotheses; however, the feasibility outcomes suggested that the protocol requires modification before conducting a future trial.

XSL•FO
RenderX

Feasibility of the Intervention

Recruitment, Attrition, and Intervention Completion Rates

Our study achieved a recruitment rate of slightly over 50% (38/75, 51%), defined as the percentage of individuals who enrolled in the intervention out of those approached. This rate exceeds those reported in previous lifestyle intervention studies targeting older adults with frailty, which averaged around 20% [47,48], indicating the effectiveness of our recruitment strategies. The attrition rates were 3 out of 19 (16%) in the intervention group and 1 out of 19 (5%) in the control group. Although the attrition rates in both groups were lower than those reported in previous PA interventions for older adults with frailty (>20%) [49,50], the higher attrition rate in our intervention group compared with the control group may suggest the presence of "selective attrition." Notably, 2 participants dropped out immediately after being allocated to the integrative mHealth intervention group, as the researchers lost contact with them following the communication of randomization results. This may indicate that these participants lacked the motivation to engage with mHealth as a means to improve their health behavior [51]. Regarding the intervention completion rate, the overall completion rate—comprising workshop attendance, self-reported app engagement, and self-reported outdoor practice—was 134 out of 170 (78.8%). Specifically, the outdoor practice completion rate (36/51, 71%), app engagement (35/51, 69%), and workshop attendance (63/68, 93%) were well over 60%. When comparing the workshop components, both the intervention and control groups demonstrated similarly high completion rates (active control: 72/76, 95%; integrative mHealth: 63/68, 93%). These high workshop attendance rates in both groups suggest that participating in workshops to gain health knowledge or exercise experience is highly acceptable among community-dwelling older adults. Although direct comparisons of intervention completion rates across different mHealth studies are challenging—due to variations in intervention components (eg, some include only an app, while others add workshops and self-practice sessions) and differing definitions of app engagement [52]—it is noteworthy that the self-reported app engagement rate (35/51, 69%) in our study was lower than the 88.4% app log-in rate reported in a similar mHealth intervention for older adults with frailty in Japan [48]. Although participants perceived the app as useful, as reflected in the self-developed questionnaire responses, the marginally acceptable adherence suggested usability issues. Qualitative feedback highlighted that app usability could be enhanced by improving its visuals and functionality. Before a future trial, an upgraded app version should incorporate features such as image enlargement, exercise reminders, assistance with locating specific OEFs, and detailed guidance on correct and incorrect exercise postures. These enhancements are expected to improve user engagement and the overall usability of the app. To increase app engagement among older adults with prefrailty and frailty, the overall user experience should be optimized before implementing a full-powered trial. Regarding the outdoor practice component, the completion rate of self-reported outdoor exercise (134/170, 78.8%) was comparable to the logged workout rate (72%) reported in a previous mHealth intervention involving outdoor gym workouts for inactive adults [16]. Qualitative data from semistructured interviews suggest that incorporating at least one supervised outdoor practical session in future interventions could enhance the overall participant experience. To improve adherence to the outdoor practical sessions, future trials should identify specific outdoor gyms located near the community centers and offer supervised outdoor practical sessions alongside the educational workshops held within these centers.

Safety

The adverse events reported in our study included a slipping incident caused by faulty outdoor exercise equipment and musculoskeletal pain experienced after using certain exercise devices. The participant who reported musculoskeletal pain had a history of occasional, nonspecific low back and knee pain before participating in the intervention. After advising the participant to reduce the range of motion and exercise at a slower tempo to maintain proper posture and body alignment, the pain resolved. Previous studies on exercise apps [53] and exercise interventions [54] have reported musculoskeletal pain as a common issue among older adult participants. Additionally, research indicates that 66.5% of OEF users experience injuries related to faulty equipment or improper usage [55]. These findings offer valuable insights for improving safety in future full-scale trials. Interventionists with expertise in exercise and health should advise participants with frailty to begin exercising with appropriate dosage, type, and intensity. The principle of gradual exercise progression should be emphasized to help prevent musculoskeletal injuries. Participants should also be reminded to inspect exercise equipment before each use.

Acceptability

The current integrative mHealth intervention demonstrated a level of acceptability comparable to that reported in a previous mHealth intervention for older adults with frailty [48] and an OEF-based mHealth intervention for inactive adults [16]. Consistent with the findings of Jansson and colleagues [16], our participants perceived the app content as sufficient to support and motivate unsupervised PA in park settings. Additionally, participants appreciated the intervention's comprehensive structure, which included 3 components: the app, workshops, and unsupervised outdoor practice. This finding aligns with recent research indicating that interventions combining workshop and digital components yield more promising outcomes in older adults than those relying on a single component [56,57]. While the app was generally found to be acceptable—reflected in responses to the self-developed questionnaire and semistructured interviews—the results also provided important insights for further improving both the app and the overall integrative mHealth intervention before future trials. Unlike the study by Jansson and colleagues [16], which suggested that the Ecofit app could effectively support unsupervised PA in park settings, participants in our study—who were community-dwelling older adults with prefrailty and frailty—recommended incorporating at least one supervised session in parks equipped with OEFs.

XSL•FO
RenderX

Preliminary Effects of the Intervention

Overview

This study explored the preliminary effects of an integrative mHealth intervention on older adults with prefrailty and frailty, focusing on the use of OEFs, PA levels, exercise self-efficacy, and mental well-being. Given the exploratory and underpowered nature of the analysis, results should be interpreted with caution. To provide insights into the magnitude and direction of effects, results are presented alongside 95% CIs [58,59].

Self-Reported and Objectively Measured PA Levels

Sustaining PA behavioral change among older adults with prefrailty and frailty is challenging, as demonstrated in previous research [12]. Therefore, the sustained increase in PA levels—measured both subjectively and objectively—at the 3-month follow-up in our study is an encouraging finding. This result aligns with a prior outdoor gym mHealth study, which showed that a mobile app providing education on OEFs effectively maintained resistance training–specific PA behavior changes among inactive adults at 3 and 9 months [21]. While the integrative mHealth intervention demonstrated potential treatment effects, it is important to note that only a subsample of participants agreed to wear accelerometers to objectively measure PA. At baseline, 16 out of 18 (89%) in the control group and 15 out of 16 (94%) in the intervention group consented to wear the device, but only 24 out of 34 (71%) of these participants wore it across all time points. This level of compliance is consistent with previous PA studies involving older adults [60]. To enhance accelerometer adherence in future trials, providing financial incentives and ensuring logistical convenience will be essential.

Mental Well-Being and Exercise Self-Efficacy

The observed improvements in mental well-being at the 3-month follow-up and exercise self-efficacy immediately postintervention align with findings from previous OEF-based PA and mHealth studies [14,61]. However, these results require confirmation in a larger trial. Notably, 5 (15%) participants showed signs of cognitive impairment that prevented them from completing assessments related to mental well-being and exercise self-efficacy. Therefore, future trials should consider excluding individuals with cognitive impairment to ensure data accuracy.

Novelty of the Integrative mHealth Intervention of This Study

Although integrative health interventions supporting the use of public OEFs to promote PA are not new, this study presents the first integrative mHealth intervention specifically designed for community-dwelling older adults with prefrail or frail health status. The intervention aimed to improve PA, exercise self-efficacy, and mental well-being in this population through the combined use of an app, educational workshops, and unsupervised outdoor engagement with senior-friendly OEFs available in public open spaces. The novelty of this integrative mHealth intervention lies in its unique combination of PA and rehabilitation knowledge. The “Outdoor Rehab-Fit” app is the first to offer rehabilitation insights and guidance from physiotherapists and occupational therapists for each piece of outdoor exercise equipment available in public open spaces in Hong Kong. This added rehabilitation perspective helps older adults better understand how using specific outdoor equipment can guide movements and activities that improve their physical condition. For example, using the “calf stretch” equipment (an inclined ramp) can help alleviate foot and knee joint pain by improving calf flexibility, while exercising with the “bench stepper” equipment can strengthen the quadriceps and gluteus maximus muscles, ultimately enhancing walking and stair-climbing abilities. Unlike previous apps that primarily focus on fitness enhancement [21], the current app aims to boost users’ expectations regarding how OEFs can address functional limitations while also increasing PA levels. This intervention acknowledges that health concerns and functional limitations are significant barriers to exercise among this population [62].

Additionally, the integrative delivery structure is another novel aspect of the current intervention. Unlike previous outdoor gym mHealth interventions designed for adults, which included only a single introductory workshop [21], this intervention emphasizes the importance of a comprehensive approach by incorporating 4 educational workshop sessions alongside the app. These workshops recognize that older adults generally have lower digital literacy compared with the broader adult population, thereby providing more support to facilitate effective app use and engagement. Older adults require real-time, interactive support not only to learn how to use the app but also to gain knowledge relevant to their engagement with OEFs. The workshop sessions complement the app by delivering educational content and offering opportunities for real-time interaction, enabling older adults with frailty to better understand essential exercise concepts tailored to their needs. The high adherence to the workshops observed in this study underscores the importance of this integrated delivery approach.

Implications to Future Large-Scale Study

The pilot trial provides preliminary evidence that the proposed integrative mHealth intervention may promote sustainable improvements in PA, mental well-being, and exercise self-efficacy among older adults with prefrailty and frailty. The demonstrated effectiveness, along with the widespread availability of OEFs in the region, supports the scalability of this intervention. However, the feasibility and acceptability findings from this study highlight areas of the intervention that require refinement before conducting a future full-scale trial. Future large-scale trials should be conducted in community settings in collaboration with local community organizations and utilize a cluster RCT design. To improve adherence to the outdoor practice component, the intervention should incorporate at least one supervised outdoor session at a nearby outdoor gym, in addition to educational workshops held at the community center. Enhancing app adherence will require usability improvements, including providing exercise demonstrations in video format with the option to enlarge for better visibility. Additionally, the app should incorporate a reminder function to prompt users about their exercise routines. The educational content within the app also requires updating to enhance its relevance and usefulness. To improve safety, future interventions should include guidance and reminders for participants on how

XSL•FO
RenderX

to inspect exercise equipment before each use. Workshops should emphasize the concept of gradual exercise progression tailored to each piece of equipment. Regarding PA measurement, our study observed a decline in accelerometer compliance over repeated time points. Therefore, future trials should consider using self-reported PA measurement tools with strong psychometric properties to ensure higher compliance in longitudinal assessments. Lastly, because the study was conducted within a university setting, it naturally attracted participants with higher education levels and greater digital literacy. Future community-based trials are likely to recruit individuals with lower digital literacy. To support these participants in the integrative mHealth group, it will be essential to mobilize adequate social assistance from community volunteers or university students [63].

Limitations

This study has several limitations. First, the integrative mHealth intervention consisted of 3 components: the mobile app, OEFs, and workshops. Consequently, isolating the effects of each component was challenging. However, the inclusion of an active control group, which matched contact hours and exercise knowledge, helped minimize confounding factors and allowed for a clearer estimation of the intervention's potential treatment effect. Second, the study required participants to own a smartphone, excluding those without one. However, smartphone penetration in Hong Kong is high, with 97.1% of the population aged 10 years and above owning a smartphone [64]. Additionally, the mandatory use of the Leave Home Safe app during the 2020 COVID-19 pandemic further increased smartphone ownership among older adults [65]. Third, participants had access to varying numbers and types of outdoor exercise facilities (OEFs) in their neighborhoods, which may have influenced their engagement with the intervention. It is impossible to standardize the built physical environment, and this variation among neighborhoods may have influenced the study outcomes. Nevertheless, this pilot feasibility RCT estimated potential treatment effects within a real-world context. Finally, it is important to emphasize that the results of this pilot RCT provide only preliminary evidence regarding the feasibility of the integrative mHealth intervention for community-dwelling older adults with prefrailty and frailty.

Strengths

To the best of our knowledge, this is the first study to examine the feasibility of an integrative mHealth intervention combining an app, workshops, and unsupervised outdoor practice specifically for community-dwelling older adults with prefrailty and frailty. The study gathered valuable feasibility and acceptability data to inform important modifications ahead of a full-powered trial. Additionally, a health education workshop with exercise experiential sessions—commonly offered in community centers for older adults—served as the comparator group, allowing for a meaningful comparison. Finally, the study protocol was registered before data collection, ensuring transparency in all aspects of reporting.

Conclusion

Preliminary evidence supports the foundation of this study; however, strategies to improve intervention adherence are needed, and modifications should be made before future trials.

Acknowledgments

We thank Kong Chung Hei William, Siu Wai Ting, Tang Zengjin, and Xiao Kejing for their support in the promotion and data collection of the project; and Sang Mingjian and Gu Yanwu for their assistance with the software modification of the app. We also thank AsiaEdit for their professional English language editing support. This work is supported by the Better Utilization Fund, Department of Rehabilitation Sciences, The Hong Kong Polytechnic University (project code P0045122). This funding source had no role in the design of the study and will not have any role during its execution, data analysis, interpretation, or decision to submit the results.

Data Availability

The data that support the findings of this study are available upon reasonable request from the authors.

Authors' Contributions

JLC conceptualized the study, obtained protocol registration, supervised the execution of the project, analyzed and interpreted the data, and wrote the manuscript. AYL provided important advice during the conceptualization stage and contributed to the educational content of the app. RS provided advice during app development. PHF contributed to the software development of the app. HYZ supported data analysis. RX assisted manuscript writing. JLC, AYL, PHF, SNF, KNK, ASK, and KNK obtained the funding. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 197 KB - mhealth_v13i1e69259_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Intervention completion assessment.
[PDF File (Adobe PDF File), 103 KB - mhealth_v13i1e69259_app2.pdf]

Multimedia Appendix 3
Self-developed questionnaire on the perceived usefulness of the mobile app.
[PDF File (Adobe PDF File), 119 KB - mhealth_v13i1e69259_app3.pdf]

Multimedia Appendix 4
Interview guide.
[PDF File (Adobe PDF File), 166 KB - mhealth_v13i1e69259_app4.pdf]

Multimedia Appendix 5
Secondary outcomes of both the intervention and control groups at baseline, postintervention, and the 3-month follow-up.
[DOCX File , 19 KB - mhealth_v13i1e69259_app5.docx]

References

1. Wilmoth J, Bas D, Mukherjee S, Hanif N. World Social Report 2023: Leaving No One Behind in an Ageing World. New York, NY: DESA Publications; 2023.
2. To T, Doan T, Ho W, Liao W. Prevalence of frailty among community-dwelling older adults in Asian countries: a systematic review and meta-analysis. Healthcare (Basel) 2022 May 12;10(5):1-14 [FREE Full text] [doi: 10.3390/healthcare10050895] [Medline: 35628034]
3. Yu R, Wu W, Leung J, Hu SC, Woo J. Frailty and its contributory factors in older adults: a comparison of two Asian regions (Hong Kong and Taiwan). Int J Environ Res Public Health 2017 Sep 21;14(10):1-16 [FREE Full text] [doi: 10.3390/ijerph14101096] [Medline: 28934150]
4. Gobbens RJ, Luijkx KG, Wijnen-Sponselee MT, Schols JM. Toward a conceptual definition of frail community dwelling older people. Nurs Outlook 2010;58(2):76-86. [doi: 10.1016/j.outlook.2009.09.005] [Medline: 20362776]
5. Kojima G. Frailty as a predictor of disabilities among community-dwelling older people: a systematic review and meta-analysis. Disabil Rehabil 2017 Sep;39(19):1897-1908. [doi: 10.1080/09638288.2016.1212282] [Medline: 27558741]
6. Gill TM, Gahbauer EA, Allore HG, Han L. Transitions between frailty states among community-living older persons. Arch Intern Med 2006 Feb 27;166(4):418-423. [doi: 10.1001/archinte.166.4.418] [Medline: 16505261]
7. Angulo J, El Assar M, Álvarez-Bustos A, Rodríguez-Mañas L. Physical activity and exercise: strategies to manage frailty. Redox Biol 2020 Aug;35:1-21 [FREE Full text] [doi: 10.1016/j.redox.2020.101513] [Medline: 32234291]
8. Rogers NT, Marshall A, Roberts CH, Demakakos P, Steptoe A, Scholes S. Physical activity and trajectories of frailty among older adults: evidence from the English Longitudinal Study of Ageing. PLoS One 2017;12(2):e0170878 [FREE Full text] [doi: 10.1371/journal.pone.0170878] [Medline: 28152084]
9. Peterson MJ, Giuliani C, Morey MC, Pieper CF, Evenson KR, Mercer V, Health, AgingBody Composition Study Research Group. Physical activity as a preventative factor for frailty: the health, aging, and body composition study. J Gerontol A Biol Sci Med Sci 2009 Jan;64(1):61-68 [FREE Full text] [doi: 10.1093/gerona/gln001] [Medline: 19164276]
10. de Labra C, Guimaraes-Pinheiro C, Maseda A, Lorenzo T, Millán-Calenti JC. Effects of physical exercise interventions in frail older adults: a systematic review of randomized controlled trials. BMC Geriatr 2015 Dec 02;15:1-16 [FREE Full text] [doi: 10.1186/s12877-015-0155-4] [Medline: 26626157]
11. Haider S, Grabovac I, Dorner TE. Effects of physical activity interventions in frail and prefrail community-dwelling people on frailty status, muscle strength, physical performance and muscle mass-a narrative review. Wien Klin Wochenschr 2019 Jun;131(11-12):244-254 [FREE Full text] [doi: 10.1007/s00508-019-1484-7] [Medline: 30941525]
12. Liu JYW, Kwan RYC, Yin Y, Lee PH, Siu JY, Bai X. Enhancing the physical activity levels of frail older adults with a wearable activity tracker-based exercise intervention: a pilot cluster randomized controlled trial. IJERPH 2021 Sep 30;18(19):1-23. [doi: 10.3390/ijerph181910344]
13. Lee JLC, Ho RTH. Engaging community-dwelling older adults as co-developers in a public outdoor exercise facilities-based physical activity education intervention: a mixed-method participatory study in Hong Kong. Health Soc Care Community 2022 Sep;30(5):e1862-e1874. [doi: 10.1111/hsc.13616] [Medline: 34693590]
14. Ng YL, Hill KD, Levinger P, Jacques A, Burton E. A seniors exercise park program for older adults with mild balance dysfunction - a feasibility study. Disabil Rehabil 2023 Sep;45(19):3143-3154. [doi: 10.1080/09638288.2022.2112984] [Medline: 36031870]
15. Levinger P, Panisset M, Dunn J, Haines T, Dow B, Batchelor F, et al. Exercise interveNtion outdoor proJect in the cOmmunitY for older people - results from the ENJOY Seniors Exercise Park project translation research in the community. BMC Geriatr 2020 Nov 04;20(1):1-13 [FREE Full text] [doi: 10.1186/s12877-020-01824-0] [Medline: 33148182]

XSL•FO
RenderX

16. Jansson A, Duncan M, Bauman A, Smith J, Lubans D, Attia J, et al. A mobile health resistance training intervention using outdoor gym equipment: process evaluation of the Ecofit effectiveness randomized controlled trial. J Phys Act Health 2024 Apr 01;21(4):405-412. [doi: 10.1123/jpah.2023-0228] [Medline: 38335945]
17. Hong Kong Government Information Centre. LCSD to set up exercise stations at leisure venues for the elderly. Hong Kong Government. 2002. URL: https://www.info.gov.hk/gia/general/200201/24/0124155.htm [accessed 2024-11-25]
18. Use of public open space managed by the Leisure and Cultural Serivces Department. Legislative Council Panel on Home Affairs. 2016. URL: https://www.legco.gov.hk/yr15-16/english/panels/ha/papers/ha20160614cb2-1680-3-e.pdf [accessed 2024-11-25]
19. Chow H, Ho C. Does the use of outdoor fitness equipment by older adults qualify as moderate to vigorous physical activity? PLoS One 2018;13(4):1-13 [FREE Full text] [doi: 10.1371/journal.pone.0196507] [Medline: 29709035]
20. Lee J, Ho R. Exercise spaces in parks for older adults: a qualitative investigation. J Aging Phys Act 2021 Apr 01;29(2):233-241. [doi: 10.1123/japa.2019-0397] [Medline: 32919382]
21. Plotnikoff RC, Jansson AK, Duncan MJ, Smith JJ, Bauman A, Attia J, et al. mHealth to support outdoor gym resistance training: the ecofit effectiveness RCT. Am J Prev Med 2023 Jun;64(6):853-864 [FREE Full text] [doi: 10.1016/j.amepre.2023.01.031] [Medline: 36804197]
22. Trinh M, Harris MT, Azevedo RFL, Rogers WA. Understanding older adults' perceptions of mHealth apps. Gerontechnology 2023 Mar 17;22(1):1-7 [FREE Full text] [doi: 10.4017/gt.2023.22.1.841.10]
23. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. The Lancet 2001 Apr;357(9263):1191-1194. [doi: 10.1016/s0140-6736(00)04337-3]
24. Dong L, Qiao X, Tian X, Liu N, Jin Y, Si H, et al. Cross-cultural adaptation and validation of the FRAIL Scale in Chinese community-dwelling older adults. J Am Med Dir Assoc 2018 Jan;19(1):12-17. [doi: 10.1016/j.jamda.2017.06.011] [Medline: 28757330]
25. Cardinal BJ, Cardinal MK. Screening efficiency of the revised Physical Activity Readiness Questionnaire in older adults. Journal of Aging and Physical Activity 2022 Jul 01;3(5):299-308. [doi: 10.1123/japa.3.3.299] [Medline: 36626690]
26. Hertzog MA. Considerations in determining sample size for pilot studies. Res Nurs Health 2008 Apr;31(2):180-191. [doi: 10.1002/nur.20247] [Medline: 18183564]
27. Van Blarigan EL, Chan H, Van Loon K, Kenfield SA, Chan JM, Mitchell E, et al. Self-monitoring and reminder text messages to increase physical activity in colorectal cancer survivors (Smart Pace): a pilot randomized controlled trial. BMC Cancer 2019 Mar 11;19(1):218 [FREE Full text] [doi: 10.1186/s12885-019-5427-5] [Medline: 30866859]
28. Kwan RY, Lee D, Lee PH, Tse M, Cheung DS, Thiamwong L, et al. Effects of an mHealth brisk walking intervention on increasing physical activity in older people with cognitive frailty: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):1-16 [FREE Full text] [doi: 10.2196/16596] [Medline: 32735218]
29. Poly COMP Lab. Outdoor Rehab-Fit. GooglePlay. URL: https://play.google.com/store/apps/details?id=com.hayleyzzz.outdoorfit [accessed 2024-11-24]
30. Nicholas JC, Ntoumanis N, Smith BJ, Quested E, Stamatakis E, Thøgersen-Ntoumani C. Development and feasibility of a mobile phone application designed to support physically inactive employees to increase walking. BMC Med Inform Decis Mak 2021 Jan 21;21(1):23. [doi: 10.1186/s12911-021-01391-3] [Medline: 33478495]
31. Topolski T, LoGerfo J, Patrick D, Williams B, Walwick J, Patrick M. The rapid assessment of physical activity (RAPA) among older adults. Prev Chronic Dis 2006 Oct;3(4):A118. [Medline: 16978493]
32. Richardson J, Tang A, Guyatt G, Thabane L, Xie F, Sahlas D, et al. FIT for FUNCTION: study protocol for a randomized controlled trial. Trials 2018 Jan 15;19(1):39-15 [FREE Full text] [doi: 10.1186/s13063-017-2416-3] [Medline: 29335013]
33. Montgomery CA, Henning KJ, Kantarzhi SR, Kideckel TB, Yang CFM, O'Brien KK. Experiences participating in a community-based exercise programme from the perspective of people living with HIV: a qualitative study. BMJ Open 2017 Apr 04;7(4):e015861-e015810 [FREE Full text] [doi: 10.1136/bmjopen-2017-015861] [Medline: 28377397]
34. Kwan RYC, Liu JYW, Lee D, Tse CYA, Lee PH. A validation study of the use of smartphones and wrist-worn ActiGraphs to measure physical activity at different levels of intensity and step rates in older people. Gait Posture 2020 Oct;82:306-312. [doi: 10.1016/j.gaitpost.2020.09.022] [Medline: 33007688]
35. Choi L, Liu Z, Matthews C, Buchowski M. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
36. Newton JR, Beyl R, Hebert C, Harris M, Carter L, Gahan W, et al. A physical activity intervention in older African Americans: the PAACE pilot randomized controlled trial. Med Sci Sports Exerc 2022 Oct 01;54(10):1625-1634 [FREE Full text] [doi: 10.1249/MSS.0000000000002956] [Medline: 35522253]
37. Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. J Sports Sci 2015;33(10):991-1000. [doi: 10.1080/02640414.2014.981848] [Medline: 25517396]
38. Airlie J, Forster A, Birch KM. An investigation into the optimal wear time criteria necessary to reliably estimate physical activity and sedentary behaviour from ActiGraph wGT3X+ accelerometer data in older care home residents. BMC Geriatr 2022 Feb 17;22(1):136 [FREE Full text] [doi: 10.1186/s12877-021-02725-6] [Medline: 35177023]

XSL•FO
RenderX

39. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int J Behav Nutr Phys Act 2011 Jun 16;8:62 [FREE Full text] [doi: 10.1186/1479-5868-8-62] [Medline: 21679426]
40. Tudor-Locke C, Camhi S, Troiano R. A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006. Prev Chronic Dis 2012 Jun;9:E113 [FREE Full text] [doi: 10.5888/pcd9.110332] [Medline: 22698174]
41. Vanhelst J, Mikulovic J, Bui-Xuan G, Dieu O, Blondeau T, Fardy P, et al. Comparison of two ActiGraph accelerometer generations in the assessment of physical activity in free living conditions. BMC Res Notes 2012 Apr 25;5:187 [FREE Full text] [doi: 10.1186/1756-0500-5-187] [Medline: 22534207]
42. Lee L, Perng S, Ho C, Hsu H, Lau S, Arthur A. A preliminary reliability and validity study of the Chinese version of the self-efficacy for exercise scale for older adults. Int J Nurs Stud 2009 Feb;46(2):230-238. [doi: 10.1016/j.ijnurstu.2008.09.003] [Medline: 18950769]
43. Sun Y, Luk TT, Wang MP, Shen C, Ho SY, Viswanath K, et al. The reliability and validity of the Chinese Short Warwick-Edinburgh Mental Well-being Scale in the general population of Hong Kong. Qual Life Res 2019 Oct;28(10):2813-2820 [FREE Full text] [doi: 10.1007/s11136-019-02218-5] [Medline: 31144205]
44. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
45. CalculatorSoup. URL: https://www.calculatorsoup.com/calculators/statistics/random-number-generator.php [accessed 2024-11-25]
46. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Rydwik E, Frändin K, Akner G. Effects of a physical training and nutritional intervention program in frail elderly people regarding habitual physical activity level and activities of daily living--a randomized controlled pilot study. Arch Gerontol Geriatr 2010;51(3):283-289. [doi: 10.1016/j.archger.2009.12.001] [Medline: 20044155]
48. Ohta T, Osuka Y, Shida T, Daimaru K, Kojima N, Maruo K, et al. Feasibility, acceptability, and potential efficacy of a mobile health application for community-dwelling older adults with frailty and pre-frailty: a pilot study. Nutrients 2024 Apr 16;16(8):1-12 [FREE Full text] [doi: 10.3390/nu16081181] [Medline: 38674872]
49. Giné-Garriga M, Guerra M, Unnithan VB. The effect of functional circuit training on self-reported fear of falling and health status in a group of physically frail older individuals: a randomized controlled trial. Aging Clin Exp Res 2013 Jun;25(3):329-336. [doi: 10.1007/s40520-013-0048-3] [Medline: 23740589]
50. Pollock RD, Martin FC, Newham DJ. Whole-body vibration in addition to strength and balance exercise for falls-related functional mobility of frail older adults: a single-blind randomized controlled trial. Clin Rehabil 2012 Oct;26(10):915-923. [doi: 10.1177/0269215511435688] [Medline: 22324058]
51. Schreurs K, Quan-Haase A, Martin K. Problematizing the digital literacy paradox in the context of older adults' ICT use: aging, media discourse, and self-determination. Canadian Journal of Communication 2017 May 23;42(2):359-377. [doi: 10.22230/cjc.2017v42n2a3130]
52. Yang Y, Boulton E, Todd C. Measurement of adherence to mHealth physical activity interventions and exploration of the factors that affect the adherence: scoping review and proposed framework. J Med Internet Res 2022 Jun 08;24(6):e30817 [FREE Full text] [doi: 10.2196/30817] [Medline: 35675111]
53. Daly RM, Gianoudis J, Hall T, Mundell NL, Maddison R. Feasibility, usability, and enjoyment of a home-based exercise program delivered via an exercise app for musculoskeletal health in community-dwelling older adults: short-term prospective pilot study. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e21094 [FREE Full text] [doi: 10.2196/21094] [Medline: 33439147]
54. Liu C, Latham N. Adverse events reported in progressive resistance strength training trials in older adults: 2 sides of a coin. Arch Phys Med Rehabil 2010 Sep;91(9):1471-1473. [doi: 10.1016/j.apmr.2010.06.001] [Medline: 20801270]
55. Xie J. An analysis of the injury accidents of fitness route in non-central city, Sichuan Province. Sports Cult Guide 2012;10:82-84.
56. Mehra S, van den Helder J, Visser B, Engelbert RHH, Weijs PJM, Kröse BJA. Evaluation of a blended physical activity intervention for older adults: mixed methods study. J Med Internet Res 2020 Jul 23;22(7):e16380 [FREE Full text] [doi: 10.2196/16380] [Medline: 32459652]
57. van der Veen DJ, Jellema S, van der Wees PJ, Graff MJL, de Swart BJM, Steultjens EMJ. Enhancing the role of the social network in activity (re)engagement post-stroke: a focus group study with rehabilitation professionals. BMC Prim Care 2022 Nov 17;23(1):285 [FREE Full text] [doi: 10.1186/s12875-022-01897-3] [Medline: 36397003]
58. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
59. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]

XSL•FO
RenderX

60. Shvedko AV, Thompson JL, Greig CA, Whittaker AC. Physical Activity Intervention for Loneliness (PAIL) in community-dwelling older adults: a randomised feasibility study. Pilot Feasibility Stud 2020;6:73 [FREE Full text] [doi: 10.1186/s40814-020-00587-0] [Medline: 32489675]
61. Levinger P, Goh AMY, Dunn J, Katite J, Paudel R, Onofrio A, et al. Exercise interveNtion outdoor proJect in the cOmmunitY - results from the ENJOY program for independence in dementia: a feasibility pilot randomised controlled trial. BMC Geriatr 2023 Jul 12;23(1):426 [FREE Full text] [doi: 10.1186/s12877-023-04132-5] [Medline: 37438710]
62. Freiberger E, Kemmler W, Siegrist M, Sieber C. Frailty and exercise interventions: evidence and barriers for exercise programs. Z Gerontol Geriatr 2016 Oct;49(7):606-611. [doi: 10.1007/s00391-016-1134-x] [Medline: 27655437]
63. Kwan RYC, Ng F, Lai M, Wong D, Chan S. The effects of Digital Buddy programme on older adults' mental well-being: study protocol for a multi-centre, cluster randomized controlled trial. Trials 2023 Feb 07;24(1):95 [FREE Full text] [doi: 10.1186/s13063-023-07130-5] [Medline: 36750879]
64. Slotta D. Penetration rate of smartphones among people over ten years old in Hong Kong from June to September 2022. Statista. 2023. URL: https://www.statista.com/statistics/1093416/hong-kong-penetration-rate-of-smartphone/ [accessed 2024-11-25]
65. Yang C, Lai DWL, Sun Y, Ma C, Chau AKC. Mobile application use and loneliness among older adults in the digital age: insights from a survey in Hong Kong during the COVID-19 pandemic. Int J Environ Res Public Health 2022 Jun 23;19(13):1-22 [FREE Full text] [doi: 10.3390/ijerph19137656] [Medline: 35805316]

Abbreviations

CHAMPS: Community Healthy Activities Model Program for Seniors Physical Activity Questionnaire
CONSORT: Consolidated Standards of Reporting Trials
FRAIL: Fatigue, Resistance, Ambulation, Illness, and Loss of Weight
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
OEF: outdoor exercise facility
PA: physical activity
RAPA: Rapid Assessment of Physical Activity
RCT: randomized controlled trial
SWEMWBS: Short Warwick-Edinburgh Mental Well-Being Scale

Edited by G Badicu; submitted 25.11.24; peer-reviewed by A Samuel, I Liang; comments to author 17.04.25; revised version received 27.04.25; accepted 14.05.25; published 05.06.25.

Please cite as:
Lee JLC, Wong AYL, Ng PHF, Fu SN, Fong KNK, Cheng ASK, Lee KNK, Sun R, Zhang HY, Xiao R
Outdoor Exercise Facility–Based Integrative Mobile Health Intervention to Support Physical Activity, Mental Well-Being, and Exercise Self-Efficacy Among Older Adults With Prefrailty and Frailty in Hong Kong: Pilot Feasibility Randomized Controlled Trial Study
JMIR Mhealth Uhealth 2025;13:e69259
URL: https://mhealth.jmir.org/2025/1/e69259
doi:10.2196/69259
PMID:40471669

©Janet Lok Chun Lee, Arnold Y L Wong, Peter H F Ng, S N Fu, Kenneth N K Fong, Andy S K Cheng, Karen Nga Kwan Lee, Rui Sun, Hao Yi Zhang, Rong Xiao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Association of Previous Day Carbohydrate Consumption With Fasted, Exhaled Carbon Dioxide in Lumen Users: Retrospective Real-World Study

Shlomo Yeshurun[1], PhD; Tomer Cramer[1], PhD; Daniel Souroujon[1,2], MPH, MD; Merav Mor[1], PhD

[1]Metaflow Ltd, Tel Aviv, Israel

[2]School of Public Health, Tel Aviv University, Tel Aviv, Israel

Corresponding Author:
Shlomo Yeshurun, PhD
Metaflow Ltd
HaArba'a Street 30
Tel Aviv
Israel
Phone: 972 37684062
Email: shlomi@lumen.me

Abstract

Background: The increasing prevalence of obesity and related metabolic disorders has highlighted the need for innovative tools to monitor and manage metabolic health. The Lumen device offers a novel approach to assess the metabolic state through exhaled carbon dioxide (CO_2) measurements, providing real-time feedback via a mobile app. This app-driven experience allows users to track their metabolic state and receive personalized nutrition and lifestyle recommendations, potentially supporting long-term metabolic health improvements.

Objective: This study aimed to investigate the association between the previous day's carbohydrate consumption with fasted, exhaled CO_2 levels in female and male Lumen users while also examining the influence of fasting duration, BMI, and age.

Methods: We conducted a retrospective, observational study using deidentified data from 48,058 Lumen users, comprising 707,372 fasted sessions. Separate linear mixed models were fitted for female (n=520,269 sessions) and male (n=187,103 sessions) users due to observed sex differences in metabolism. User ID was included as a random effect to account for repeated measures. The models analyzed the relationship between fasted %CO_2 levels and reported carbohydrate intake, fasting duration, BMI, and age.

Results: Higher reported carbohydrate intake from the previous day was significantly associated with increased morning %CO_2 levels in both male and female users (β=.032; P<.001; Cohen d=0.0691 in women and β=.024; P<.001; Cohen d=0.0534 in men), while a longer fasting duration was linked to decreased %CO_2 levels in both sexes (β=−.017; P<.001 for both and Cohen d=−0.0374 in women and Cohen d=−0.0382 in men). A higher BMI was associated with elevated %CO_2 levels in both sexes (β=.018; P<.001; Cohen d=0.0390 in women and β=.017; P<.001; Cohen d=0.0384 in men). Age had a statistically significant but modest effect in women (β=.008; P<.001; Cohen d=0.0180), whereas the effect size was minimal and did not meet the stricter significance threshold in men (β=.001; P=.02; Cohen d=0.0016). Cohen d values indicated that reported carbohydrate intake had the strongest effect size, while fasting duration and BMI had relatively smaller effects in both models.

Conclusions: The Lumen device is able to detect changes in fasted %CO_2 levels based on the previous day's reported carbohydrate intake and fasting duration with sex-specific metabolic responses. These findings highlight the potential of Lumen as a personalized metabolic health monitoring tool, providing insights into the influence of dietary intake and fasting on metabolic state. Future research should investigate the hormonal and physiological mechanisms contributing to the observed sex differences and assess the long-term impact of app-guided metabolic feedback on user behaviors and metabolic health outcomes.

(JMIR Mhealth Uhealth 2025;13:e64604) doi:10.2196/64604

KEYWORDS

Lumen; fasting; carbohydrate; mHealth; BMI; metabolism

XSL•FO
RenderX

Introduction

Background

There has been a dramatic increase in the prevalence of obesity in recent years, posing significant health challenges globally [1,2]. Obesity is closely linked to an elevated risk of chronic diseases, including type 2 diabetes and cardiovascular diseases, which are leading causes of mortality worldwide [3,4]. Despite efforts to promote lifestyle changes and reduce caloric intake, sustained weight management remains challenging for many individuals, highlighting the need for innovative and effective interventions [5,6].

It has been proposed that modifying the macronutrient composition of the diet could induce weight loss [7,8]. For instance, diets low in carbohydrates or high in protein have shown promise in promoting satiety and supporting weight loss [9-11]. By modulating macronutrient intake, individuals may experience improved metabolic flexibility, which is the body's ability to efficiently switch between burning carbohydrates and fats for energy, enabling better regulation of blood glucose levels and lipid metabolism [12,13]. Metabolic flexibility can be assessed by measuring the respiratory exchange ratio (RER) from metabolic cart measurements, where a lower RER during fasted states indicates greater reliance on fat oxidation over carbohydrate oxidation [13]. Short-term fasting has been shown to decrease RER values, reflecting the shift toward lipid metabolism and improved metabolic flexibility [14,15]. Metabolic flexibility is also linked to obesity, as individuals with a higher BMI are less flexible and more likely to experience metabolic disorders [16-19]. Therefore, enhancing metabolic flexibility through dietary adjustments may offer a viable strategy for mitigating the adverse effects of obesity.

Time-restricted eating is another recently popular diet, in which meals are limited to a specific period during the day. Several studies have shown that it is beneficial in weight loss, as well as other health benefits such as improved insulin sensitivity, reduced inflammation, and higher metabolic flexibility [20-22]. In addition, it may help to lower the risk of chronic diseases such as heart disease and diabetes [23]. Time-restricted eating has also been shown to decrease RER values during the fasted state, indicating a shift toward greater fat oxidation [24]. This reduction in RER during fasting periods suggests an enhanced capacity to use fat as a fuel source, which may contribute to the beneficial metabolic effects of intermittent fasting regimens [25].

Mobile health technologies, such as the Lumen breath analyzer, present innovative opportunities for real-time monitoring of metabolic responses to dietary changes [26,27]. Moreover, we recently showed that increased carbohydrate consumption in meals was associated with increased postprandial $\%CO_2$ response as measured by Lumen [28]. Similarly, applying continuous glucose monitoring (CGM) system with a mobile app was associated with improved glycemic control [29]. Such technologies empower users to make informed decisions about their diet and lifestyle, thereby potentially enhancing long-term health outcomes [30-33].

Objective

This study aimed to investigate whether fasted $\%CO_2$ measurements are associated with the previous day's carbohydrate intake, fasting duration as logged by Lumen users, and users' age, sex, and BMI.

Methods

Study Design

This is a retrospective observational study based on deidentified data collected from users of the Lumen device and app.

Participants

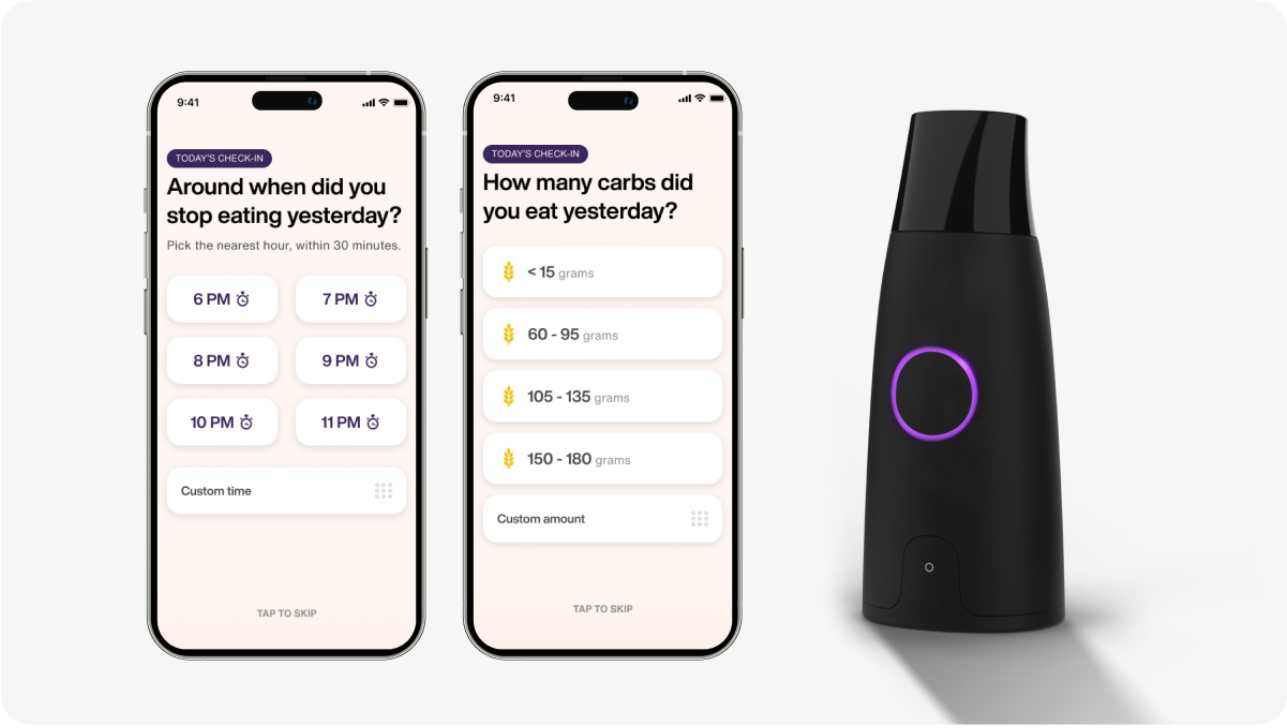
Participants' data were collected retrospectively between September 16, 2023, and March 14, 2024. All users in the analysis were aged ≥18 years. As only 150 users in the database were found to be underweight (BMI ≤18.5), they were removed from the analysis.

Data Sources

During the onboarding process, users had to specify their sex (assigned at birth), date of birth, height, and weight. The Lumen app allows users to perform a variety of optional measurements with the Lumen device (MetaFlow Ltd), such as morning fasted, as well as pre- and postmeal measurements and a bedtime measurement before sleeping. The app guides the user through each phase of the breathing maneuver, and exhaled CO_2 levels are obtained, as previously described [26]. The app provides nutritional recommendations based on the user's personal preferences and their morning $\%CO_2$ measurements regarding how much macronutrients should be consumed during the day. In addition, with each morning measurement, the app asks users about the number of portions of carbohydrates they consumed the day before (1 portion=15 g) and when was the last time they consumed any food. An example of the morning questions in the app is shown in Figure 1. For this study, only fasted measurements where users reported both their carbohydrate intake and the time they stopped eating the previous day before were included. To avoid discrepancies, we removed the top and bottom 2.5% of data points for both the amount of carbohydrate portions consumed the previous day, and the duration of fasting hours, effectively eliminating outliers from both extremes. Moreover, in instances where more than half of a user's measurements were removed for these reasons, that user's entire dataset was consequently excluded from the analysis.

XSL•FO
RenderX

Figure 1. Lumen mobile app screenshots (A and B) detailing the user input asked for following a fasted $\%CO_2$ measurement with the Lumen device. (C) An illustration of the Lumen device.



Statistical Analysis

Data were analyzed using R Studio (version 2023.12.1; Posit PBC), and all statistical analyses were conducted with the R (version 4.3.3; R Core Team) programming language, using custom scripts and the *lme4* package [34].

Linear mixed models were fitted separately for women and men to explore how the previous day's carbohydrate consumption, fasting duration, and personal characteristics (age and BMI) influence fasted $\%CO_2$ levels. Separating the sexes for analysis was based on well-documented metabolic differences and our previous work, which identified distinct $\%CO_2$ patterns in female Lumen users related to menstrual cycles and menopause [35]. Users' unique IDs were included as a random effect to account for individual variability. Model fit was evaluated using Akaike Information Criterion, Bayesian Information Criterion, and R^2 values, with detailed results provided in Table S1 in Multimedia Appendix 1. Full random effects results and diagnostic checks are also presented in Table S1 and Figure S1 in Multimedia Appendix 1. Effect sizes were quantified using Cohen *d* to assess the relative strength of each predictor variable in explaining $\%CO_2$ variation.

Among all performed analysis, a 2-sided $P \leq .01$ was considered statistically significant, following Maier and Lakens [36] and Wulff and Taylor [37] to mitigate inflated type I errors in large datasets and ensure robust findings.

Ethical Considerations

This study was determined to be exempt from institutional review board (IRB) under category 2, as detailed in 45 CFR 46.104(d) and the standard operating procedure of the Biomedical Research Alliance of New York (BRANY), by the BRANY Social, Behavioral, and Educational Research IRB on May 9, 2023 (BRANY IRB File 23-119-1476). Exemption was granted because the study involved the secondary use of deidentified data that were originally collected for purposes other than this research, ensuring that participants could not be identified. All user data were fully anonymized to minimize the risk of violating participants' privacy, and no identifiable information was accessible to the researchers. As such, informed consent was not required, and no compensation was given to the participants.

Results

Participants

Overall, 48,058 users have completed 707,372 fasted sessions reporting their carbohydrate intake from the previous day. These sessions were included in the final analysis, with a median of 8 (IQR 6-10) sessions per user.

Descriptive Characteristics

Among the users, 71.24% (34,238/48,058) were women. Female participants were slightly older than their male counterparts, with a mean age of 49.9 (SD 10) years compared to 48.1 (SD 10.2) years, and had a slightly lower BMI, averaging 29.2 (SD 6) kg/m^2 compared to 30.4 (SD 5.3) kg/m^2. A summary of the participants' characteristics is presented in Table 1.

Across 707,372 morning measurements, users reported consuming an average of 7.9 (SD 3) portions of carbohydrates the previous day. Women consumed an average of 7.5 (SD 2.7) portions, while men consumed 9.1 (SD 3.5) portions. The average fasting duration was 11.3 (SD 1.8) hours, with women fasting for 11.3 (SD 1.7) hours and men for 11.1 (SD 1.8) hours.

XSL•FO
RenderX

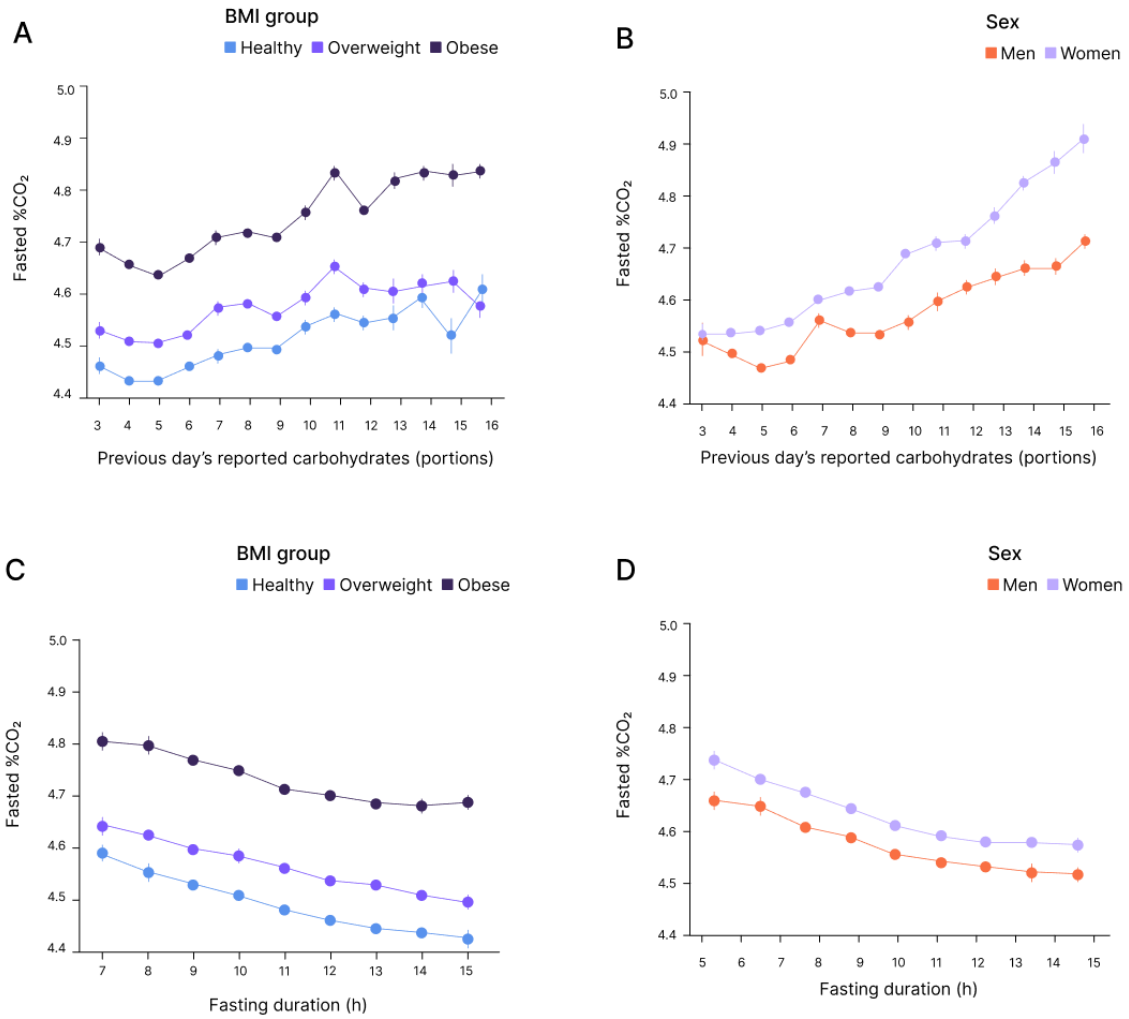
Table 1. Sample characteristics of users (n=48,058).

Characteristics	Female	Male	Overall
Sample, n (%)	34,238 (71.24)	13,820 (28.76)	48,058 (100)
Age (y), mean (SD)	49.9 (10)	48.1 (10.2)	50.0 (9.7)
BMI (kg/m^2), mean (SD)	29.2 (6)	30.4 (5.3)	28.9 (5.6)

Evaluation Outcomes

Figure 2 presents the relationship between fasted %CO_2 levels and previous day's reported carbohydrates (A and B) and fasting duration (C and D), categorized by BMI (A and C) and sex (B and D). As expected, increased reported carbohydrate consumption in the previous day resulted in increased %CO_2, and longer fasting duration resulted in decreased %CO_2. In addition, users with obesity showed higher %CO_2 than users with overweight, which in turn had higher %CO_2 than those with healthy BMI. Moreover, female users displayed higher %CO_2 than male users, with a mean morning %CO_2 of 4.60% (SD 0.47) in women and 4.56% (SD 0.46%) in men.

Figure 2. Fasted %CO_2 measurements in response to different parameters (mean and 95% CI): (A) previous day's reported carbohydrate consumption, categorized by BMI groups (healthy, overweight, and obese); (B) previous day's carbohydrate consumption, categorized by sex (male and female); (C) fasting duration, categorized by BMI groups; and (D) fasting duration, categorized by sex.



The relationship between fasted %CO_2 levels and previous day's carbohydrates, fasting duration, and users' personal information as recorded in the Lumen app was examined using separate linear mixed models for women (Table 2) and men (Table 3). User ID was modeled as a random effect to account for interindividual variability (see full random effects results in Table S1 in Multimedia Appendix 1). In both sexes, higher BMI and reported carbohydrate intake were associated with increased

XSL•FO
RenderX

$\%CO_2$ levels (reported carbohydrate intake: β=.032; *P*<.001; Cohen *d*=0.0691 in women and β=.024; *P*<.001; Cohen *d*=0.0534 in men and BMI: β=.018; *P*<.001; Cohen *d*=0.0390 in women and β=.017; *P*<.001; Cohen *d*=0.0384 in men), while prolonged fasting was linked to decreased $\%CO_2$ levels (β=−.017; *P*<.001 for both and Cohen *d*=−0.0374 in women and Cohen *d*=−0.0382 in men), corroborating the observations in Figure 2. Cohen *d* values suggest that carbohydrate intake had the strongest effect size, while fasting duration and BMI had relatively smaller effects in both models. Age had a small but statistically significant effect only in women (β=.008; *P*<.001; Cohen *d*=0.0180).

Both models demonstrated good fit, with diagnostic checks supporting their validity. Further details, including Akaike Information Criterion, Bayesian Information Criterion, R^2 values, and full diagnostic results, are provided in Table S1 and Figure S1 in Multimedia Appendix 1.

Table 2. Determinants of $\%CO_2$ in a fasted state in women (n=520,269).

Variables	β (95% CI)	*z* statistics	Cohen *d*	*P* value
Intercept	3.671 (3.642 to 3.700)	253.68	8.0015	<.001
Previous day's carbohydrates (portions)	0.032 (0.031 to 0.032)	133.00	0.0691	<.001
Fasting duration (h)	−0.017 (−0.0018 to −0.017)	−55.28	−0.0374	<.001
BMI (kg/m^2)	0.018 (0.017 to 0.019)	53.11	0.0390	<.001
Age (y)	0.008 (0.008 to 0.009)	40.54	0.0180	<.001

Table 3. Determinants of $\%CO_2$ in a fasted state in men (N=187,103).

Variables	β (95% CI)	z statistics	Cohen *d*	*P* value
Intercept	4.010 (3.963 to 4.057)	165.581	8.8970	<.001
Previous day's carbohydrates (portions)	0.024 (0.023 to 0.025)	85.741	0.0534	<.001
Fasting duration (h)	−0.017 (−0.0018 to −0.016)	−35.173	−0.0382	<.001
BMI (kg/m^2)	0.017 (0.016 to 0.019)	29.257	0.0384	<.001
Age (y)	0.001 (0.000 to 0.002)	2.317	0.0016	.021

Discussion

Principal Findings

This study provides evidence that the Lumen device can detect changes in morning $\%CO_2$ levels in relation to reported carbohydrate intake from the previous day, with similar metabolic responses observed in male and female Lumen users, apart from a notable age effect in women (Tables 2 and 3). The response is also influenced by fasting duration and BMI to a lesser extent (Figure 2).

Higher reported carbohydrate intake was associated with increased morning $\%CO_2$ in both women and men (Figure 2) aligning with previous studies showing similar results for RER values measured by metabolic carts [13,16]. In addition, prolonged fasting duration was linked to decreased $\%CO_2$, highlighting the influence of fasting on metabolic processes, which is consistent with findings on time-restricted feeding [24].

The observed shifts in morning $\%CO_2$ values occurred within a physiologically plausible range; the mean morning $\%CO_2$ was 4.60% (SD 0.47%) in women and 4.56% (SD 0.46%) in men. Although the effect sizes were small, even modest changes on the order of 0.03% to 0.05% may reflect meaningful shifts in metabolic substrate use, particularly in large-scale, real-world datasets, in accordance with our previous retrospective analysis of female users [35]. In addition, in a controlled study, fasted $\%CO_2$ levels increased by approximately 0.11% when comparing high- versus low-carbohydrate diets over a 7-day period, demonstrating that even modest shifts can reflect meaningful metabolic adaptations [27]. These results support the physiological relevance of such variations in exhaled CO_2, particularly in relation to carbohydrate intake and metabolic flexibility.

As shown in our previous research, higher BMI was associated with increased $\%CO_2$ levels [28], suggesting ameliorated metabolic flexibility in individuals with obesity. While most associations were consistent across both sexes, age demonstrated a stronger effect in female Lumen users compared to men, as older age was correlated with elevated $\%CO_2$ levels in women. This may be due to differences in menstrual cycle phases and menopause status, as previously observed in Lumen users [35], where eumenorrheic women exhibited lower $\%CO_2$ levels compared to postmenopausal women, suggesting hormonal influence on exhaled $\%CO_2$.

Comparison With Prior Work

The findings of our study align with and expand upon previous research in the field of metabolic health and nutrition. Our observation that higher reported carbohydrate intake is associated with increased morning $\%CO_2$ is consistent with findings from studies using traditional RER measurements. For instance, Galgani et al [13] demonstrated that carbohydrate overfeeding led to increased RER values, indicating a shift

XSL•FO
RenderX

toward carbohydrate oxidation. In addition, this retrospective analysis showed similar results to a previous prospective study using the Lumen device. Roberts et al [27] observed significant changes in exhaled $\%CO_2$ in response to short-term low and high carbohydrate diets, where high carbohydrate diet resulted in higher fasted $\%CO_2$ levels. Our study adds to this by demonstrating that these effects are observable in a larger, real-world setting, with variations in carbohydrate intake from the previous day significantly influencing morning $\%CO_2$ measurements.

The inverse relationship we observed between fasting duration and morning $\%CO_2$ levels aligns with established literature on the metabolic effects of fasting. Ravussin et al [24] demonstrated that time-restricted feeding increased fat oxidation and decreased RER values in addition to improved metabolic flexibility in overweight adults. Similarly, Vinales et al [15] showed that fasting led to decreased RER values from the metabolic chamber, while high carbohydrate intake resulted in increased RER. As longer overnight fasting was previously associated with reduced BMI [38], this study suggests that Lumen could be a valuable tool for monitoring the metabolic effects of various fasting regimens which might assist in weight loss.

Our study identified significant associations between $\%CO_2$ levels and various demographic factors, including BMI, age, and sex. The positive correlation between BMI and $\%CO_2$ levels is consistent with previous research on metabolic flexibility. For instance, Begaye et al [14] found that individuals with higher BMI exhibited impaired metabolic flexibility, which could explain the higher $\%CO_2$ levels observed in our study. In addition, our previous retrospective analysis demonstrated the relationship between low metabolic flexibility in users with obesity [28].

The observed associations between carbohydrate intake, fasting duration, and exhaled CO_2 may be explained by the role of insulin in substrate oxidation, as carbohydrate intake stimulates insulin secretion, promoting carbohydrate use while suppressing fat oxidation [13,19]. Similarly, the reduction in $\%CO_2$ with prolonged fasting reflects a metabolic shift from carbohydrate to fat oxidation, indicative of improved metabolic flexibility [14,16]. Hormonal responses to fasting, including increased glucagon and cortisol and reduced insulin, further modulate substrate use [24,39]. In women, progesterone fluctuations during the menstrual cycle may also contribute to metabolic flexibility, as higher progesterone levels have been linked to increased lipid oxidation and lower RER values, which could partly explain the observed sex differences in age-related $\%CO_2$ patterns [35,40,41]. Future research should further investigate these hormonal influences to better understand individual metabolic responses.

The observed sex-based differences in our study are particularly interesting in light of our recent work by Cramer et al [35], which explored the relationship between the menstrual cycle and menopause to exhaled CO_2 in Lumen users. This study showed cyclical $\%CO_2$ patterns in eumenorrheic women, with lower values compared to menopausal women, suggesting a hormonal influence on metabolic fuel use. Consistently, our findings indicate that age was significantly associated with $\%CO_2$ in women but had a negligible effect in men. Given that most Lumen users analyzed in this study are women, this could explain the stronger association of age with fasted $\%CO_2$ levels in the female cohort. Future research should aim to elucidate the underlying hormonal and physiological mechanisms driving these sex-specific responses and their potential implications for personalized nutrition and health strategies.

Longitudinal investigations into the relationship between exhaled CO_2 levels and weight loss outcomes could provide crucial insights into the potential of Lumen device for weight management. These studies could explore whether fluctuations in morning $\%CO_2$ levels serve as correlates of long-term weight loss success. In addition, the applicability of the device should be assessed in diverse populations, including individuals with metabolic disorders such as type 2 diabetes or polycystic ovary syndrome, as well as athletes and older adults, to determine whether metabolic responses differ across various health and lifestyle conditions. Future research should also investigate the integration of Lumen with other health monitoring technologies, such as CGMs or wearable activity trackers, to provide a more comprehensive view of metabolic health. Furthermore, future research should delve into the impact of additional parameters on exhaled CO_2 levels. For instance, examining both the acute and chronic effects of physical activity on $\%CO_2$ measurements could yield valuable data as it was previously shown to influence RER levels [42]. Given the established links between sleep and metabolic health [43], investigating how sleep duration and quality influence $\%CO_2$ levels would also be beneficial. Similarly, exploring the role of stress levels is important, as cortisol is known to affect metabolic processes and was found to be elevated during prolonged fasting [39]. Ultimately, this research could pave the way for more personalized and effective strategies in metabolic health management.

Our study contributes significantly to the growing field of mobile health (mHealth) and ubiquitous health technologies. The use of the Lumen device as a portable metabolic measurement tool aligns with the broader trend of leveraging consumer-grade devices for health monitoring and management. Similar to how CGMs have revolutionized diabetes management [29], our findings suggest that breath-based metabolic monitoring could play a crucial role in personalized nutrition and weight management strategies. The large-scale, real-world nature of our dataset is particularly valuable in the context of mHealth research, as it provides insights into how these technologies perform in everyday settings, outside of controlled laboratory environments. This approach is consistent with recent mHealth studies that have demonstrated the effectiveness of mobile apps in promoting health and managing diseases [30,31]. Furthermore, our study exemplifies the potential of mHealth devices to generate large datasets that can be used to identify patterns and personalize interventions, a key advantage noted in reviews of digital health interventions for weight loss [32]. By demonstrating the ability of Lumen device to detect metabolic changes in response to dietary habits and individual characteristics, our research contributes to the broader goal of

XSL•FO
RenderX

empowering users to take greater responsibility for their health through accessible, user-friendly technologies [33].

Limitations

Despite the large dataset of real-world evidence that was used in this analysis, several limitations need to be mentioned. First, as this is a retrospective and observational study, causal relationships between variables cannot be definitively established. While the results provide valuable insights into associations between carbohydrate intake, fasting duration, and metabolic responses, future research using prospective or experimental designs is needed to strengthen causal inferences. Second, as this analysis is based on real-world data, the reported carbohydrate and the reported fasting duration may be miscalculated or incomplete (recall bias). Future studies should consider integrating objective dietary tracking technologies, such as food-tracking tools, to cross-validate self-reported data and reduce potential recall bias. Moreover, as only reported carbohydrate intake was considered in this analysis, it is difficult to determine if other factors in the meal might influence the $\%CO_2$ as well, such as fats and protein as well as the type of carbohydrate and the fiber composition. In addition, the macronutrient intake of Lumen users might not be representative of most typical diets, as the Lumen app guides them toward a specific diet, primarily low in carbohydrate, and with most of them on a weight loss journey. As this study has specific user characteristics, caution should be exercised in interpreting these conclusions for a broader audience. Future research should aim to validate findings in more diverse populations, including those with different dietary habits and health goals. Our model showed adequate but imperfect homoscedasticity (Figure S1 in Multimedia Appendix 1). Future studies could use generalized additive or robust regression techniques to address this limitation. Finally, fasted $\%CO_2$ levels might be affected by other confounding variables, including preexisting conditions, medication use, dietary restrictions, menstrual cycle, and lifestyle factors such as stress, sleep, and physical activity, which were not controlled for in this study. Future studies should control these variables through comprehensive data collection and statistical adjustments, or by designing prospective studies that account for these potential confounders.

Conclusions

In summary, this study demonstrates that the Lumen device can detect the impact of the previous day's reported carbohydrate consumption on fasted $\%CO_2$ measurements in both female and male users. In addition, the device's sensitivity to variations in fasting duration and BMI suggests its utility in providing personalized dietary recommendations. By accounting for individual differences in metabolic responses, Lumen can offer valuable insights into metabolic health and assist users in optimizing their carbohydrate intake and fasting routines.

Acknowledgments

This work was funded by MetaFlow Ltd. The authors would like to express their thanks to all app users who contributed their data to this study and the Lumen team for their support.

Authors' Contributions

SY, TC, and MM led the design and conceptualization of the study. MM was the principal supervisor of the study. SY led the curation of data and conducted data interpretation and formal analysis. SY wrote the first draft of the manuscript. TC and DS contributed to the data analysis. All authors critically revised the manuscript and gave their final approval.

Conflicts of Interest

MM is the chief scientific officer and cofounder of MetaFlow Ltd. DS is a consultant at MetaFlow. SY and TC are paid employees at MetaFlow. To promote transparency and mitigate potential conflicts of interest, the study was reviewed and approved by an independent institutional review board. Data analysis followed established methodologies, and statistical interpretations were based on predefined analytical approaches in accordance with existing literature. While the study was funded by MetaFlow Ltd, the decision to publish and the interpretation of findings were guided by scientific integrity, independent of the funder's influence. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Supplementary materials including model fit statistics for the linear mixed models (Table S1) and residual diagnostic plots assessing homoscedasticity, normality, and distribution of residuals (Figure S1).
[DOC File , 32 KB - mhealth_v13i1e64604_app1.doc]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. Lancet 2016 Apr 02;387(10026):1377-1396 [FREE Full text] [doi: 10.1016/S0140-6736(16)30054-X] [Medline: 27115820]
2. Powell-Wiley TM, Poirier P, Burke LE, Després JP, Gordon-Larsen P, Lavie CJ, American Heart Association Council on Lifestyle and Cardiometabolic Health; Council on CardiovascularStroke Nursing; Council on Clinical Cardiology; Council on EpidemiologyPrevention;Stroke Council. Obesity and cardiovascular disease: a scientific statement from the American

XSL•FO
RenderX

Heart Association. Circulation 2021 May 25;143(21):e984-1010 [FREE Full text] [doi: 10.1161/CIR.0000000000000973] [Medline: 33882682]
3. Joseph JJ, Deedwania P, Acharya T, Aguilar D, Bhatt DL, Chyun DA, American Heart Association Diabetes Committee of the Council on LifestyleCardiometabolic Health; Council on Arteriosclerosis, ThrombosisVascular Biology; Council on Clinical Cardiology;Council on Hypertension. Comprehensive management of cardiovascular risk factors for adults with type 2 diabetes: a scientific statement from the American Heart Association. Circulation 2022 Mar;145(9):e722-e759 [FREE Full text] [doi: 10.1161/CIR.0000000000001040] [Medline: 35000404]
4. Censin JC, Peters SA, Bovijn J, Ferreira T, Pulit SL, Mägi R, et al. Causal relationships between obesity and the leading causes of death in women and men. PLoS Genet 2019 Oct;15(10):e1008405 [FREE Full text] [doi: 10.1371/journal.pgen.1008405] [Medline: 31647808]
5. Wadden TA, Tronieri JS, Butryn ML. Lifestyle modification approaches for the treatment of obesity in adults. Am Psychol 2020;75(2):235-251 [FREE Full text] [doi: 10.1037/amp0000517] [Medline: 32052997]
6. MacLean PS, Wing RR, Davidson T, Epstein L, Goodpaster B, Hall KD, et al. NIH working group report: innovative research to improve maintenance of weight loss. Obesity (Silver Spring) 2015 Jan;23(1):7-15 [FREE Full text] [doi: 10.1002/oby.20967] [Medline: 25469998]
7. Ge L, Sadeghirad B, Ball GD, da Costa BR, Hitchcock CL, Svendrovski A, et al. Comparison of dietary macronutrient patterns of 14 popular named dietary programmes for weight and cardiovascular risk factor reduction in adults: systematic review and network meta-analysis of randomised trials. BMJ 2020 Apr 01;369:m696 [FREE Full text] [doi: 10.1136/bmj.m696] [Medline: 32238384]
8. Austin GL, Ogden LG, Hill JO. Trends in carbohydrate, fat, and protein intakes and association with energy intake in normal-weight, overweight, and obese individuals: 1971-2006. Am J Clin Nutr 2011 Apr;93(4):836-843 [FREE Full text] [doi: 10.3945/ajcn.110.000141] [Medline: 21310830]
9. Halford JC, Harrold JA. Satiety-enhancing products for appetite control: science and regulation of functional foods for weight management. Proc Nutr Soc 2012 May;71(2):350-362 [FREE Full text] [doi: 10.1017/S0029665112000134] [Medline: 22401600]
10. Kim JY. Optimal diet strategies for weight loss and weight loss maintenance. J Obes Metab Syndr 2021 Mar 30;30(1):20-31 [FREE Full text] [doi: 10.7570/jomes20065] [Medline: 33107442]
11. Wycherley TP, Moran LJ, Clifton PM, Noakes M, Brinkworth GD. Effects of energy-restricted high-protein, low-fat compared with standard-protein, low-fat diets: a meta-analysis of randomized controlled trials. Am J Clin Nutr 2012 Dec;96(6):1281-1298 [FREE Full text] [doi: 10.3945/ajcn.112.044321] [Medline: 23097268]
12. Russell WR, Baka A, Björck I, Delzenne N, Gao D, Griffiths HR, et al. Impact of diet composition on blood glucose regulation. Crit Rev Food Sci Nutr 2016;56(4):541-590 [FREE Full text] [doi: 10.1080/10408398.2013.792772] [Medline: 24219323]
13. Galgani JE, Moro C, Ravussin E. Metabolic flexibility and insulin resistance. Am J Physiol Endocrinol Metab 2008 Nov;295(5):E1009-E1017 [FREE Full text] [doi: 10.1152/ajpendo.90558.2008] [Medline: 18765680]
14. Begaye B, Vinales KL, Hollstein T, Ando T, Walter M, Bogardus C, et al. Impaired metabolic flexibility to high-fat overfeeding predicts future weight gain in healthy adults. Diabetes 2020 Feb;69(2):181-192 [FREE Full text] [doi: 10.2337/db19-0719] [Medline: 31712321]
15. Vinales KL, Schlögl M, Piaggi P, Hohenadel M, Graham A, Bonfiglio S, et al. The consistency in macronutrient oxidation and the role for epinephrine in the response to fasting and overfeeding. J Clin Endocrinol Metab 2017 Jan 01;102(1):279-289 [FREE Full text] [doi: 10.1210/jc.2016-3006] [Medline: 27820654]
16. Goodpaster BH, Sparks LM. Metabolic flexibility in health and disease. Cell Metab 2017 May 02;25(5):1027-1036 [FREE Full text] [doi: 10.1016/j.cmet.2017.04.015] [Medline: 28467922]
17. Galgani JE, Fernández-Verdejo R. Pathophysiological role of metabolic flexibility on metabolic health. Obes Rev 2021 Feb;22(2):e13131. [doi: 10.1111/obr.13131] [Medline: 32815226]
18. Palmer BF, Clegg DJ. Metabolic flexibility and its impact on health outcomes. Mayo Clin Proc 2022 Apr;97(4):761-776 [FREE Full text] [doi: 10.1016/j.mayocp.2022.01.012] [Medline: 35287953]
19. Smith RL, Soeters MR, Wüst RC, Houtkooper RH. Metabolic flexibility as an adaptation to energy resources and requirements in health and disease. Endocr Rev 2018 Aug 01;39(4):489-517 [FREE Full text] [doi: 10.1210/er.2017-00211] [Medline: 29697773]
20. Harvie M, Howell A. Potential benefits and harms of intermittent energy restriction and intermittent fasting amongst obese, overweight and normal weight subjects-a narrative review of human and animal evidence. Behav Sci (Basel) 2017 Jan 19;7(1):4 [FREE Full text] [doi: 10.3390/bs7010004] [Medline: 28106818]
21. Gabel K, Hoddy KK, Haggerty N, Song J, Kroeger CM, Trepanowski JF, et al. Effects of 8-hour time restricted feeding on body weight and metabolic disease risk factors in obese adults: a pilot study. Nutr Healthy Aging 2018 Jun 15;4(4):345-353 [FREE Full text] [doi: 10.3233/NHA-170036] [Medline: 29951594]
22. Patterson RE, Sears DD. Metabolic effects of intermittent fasting. Annu Rev Nutr 2017 Aug 21;37:371-393 [FREE Full text] [doi: 10.1146/annurev-nutr-071816-064634] [Medline: 28715993]

XSL•FO
RenderX

23. Patikorn C, Roubal K, Veettil SK, Chandran V, Pham T, Lee YY, et al. Intermittent fasting and obesity-related health outcomes: an umbrella review of meta-analyses of randomized clinical trials. JAMA Netw Open 2021 Dec 01;4(12):e2139558 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.39558] [Medline: 34919135]
24. Ravussin E, Beyl RA, Poggiogalle E, Hsia DS, Peterson CM. Early time-restricted feeding reduces appetite and increases fat oxidation but does not affect energy expenditure in humans. Obesity (Silver Spring) 2019 Aug;27(8):1244-1254 [FREE Full text] [doi: 10.1002/oby.22518] [Medline: 31339000]
25. Antoni R, Johnston KL, Collins AL, Robertson MD. Intermittent v. continuous energy restriction: differential effects on postprandial glucose and lipid metabolism following matched weight loss in overweight/obese participants. Br J Nutr 2018 Mar;119(5):507-516. [doi: 10.1017/S0007114517003890] [Medline: 29508693]
26. Lorenz KA, Yeshurun S, Aziz R, Ortiz-Delatorre J, Bagley JR, Mor M, et al. A handheld metabolic device (Lumen) to measure fuel utilization in healthy young adults: device validation study. Interact J Med Res 2021 May 17;10(2):e25371 [FREE Full text] [doi: 10.2196/25371] [Medline: 33870899]
27. Roberts J, Dugdale-Duwell D, Lillis J, Pinto JM, Willmott A, Yeshurun S, et al. The efficacy of a home-use metabolic device (Lumen) in response to a short-term low and high carbohydrate diet in healthy volunteers. J Int Soc Sports Nutr 2023 Dec;20(1):2185537 [FREE Full text] [doi: 10.1080/15502783.2023.2185537] [Medline: 36862060]
28. Yeshurun S, Cramer T, Souroujon D, Mor M. The association of macronutrient consumption and BMI to exhaled carbon dioxide in lumen users: retrospective real-world study. JMIR Mhealth Uhealth 2024 Apr 01;12:e56083 [FREE Full text] [doi: 10.2196/56083] [Medline: 38439744]
29. Kumbara AB, Iyer AK, Green CR, Jepson LH, Leone K, Layne JE, et al. Impact of a combined continuous glucose monitoring-digital health solution on glucose metrics and self-management behavior for adults with type 2 diabetes: real-world, observational study. JMIR Diabetes 2023 Sep 11;8:e47638 [FREE Full text] [doi: 10.2196/47638] [Medline: 37590491]
30. Kang HS, Exworthy M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth 2022 Jul 13;10(7):e35684 [FREE Full text] [doi: 10.2196/35684] [Medline: 35830222]
31. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
32. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
33. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MD, WARIFA Consortium. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 04;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]
34. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):33. [doi: 10.18637/jss.v067.i01]
35. Cramer T, Yeshurun S, Mor M. Changes in exhaled carbon dioxide during the menstrual cycle and menopause. Digit Biomark 2024;8(1):102-110 [FREE Full text] [doi: 10.1159/000539126] [Medline: 39015514]
36. Maier M, Lakens D. Justify your alpha: a primer on two practical approaches. Adv Methods Pract Psychol Sci 2022 May 30;5(2):251524592210803. [doi: 10.1177/25152459221080396]
37. Wulff JN, Taylor L. How and why alpha should depend on sample size: a Bayesian-frequentist compromise for significance testing. Strateg Organ 2024 Jan 06;22(3):550-581. [doi: 10.1177/14761270231214429]
38. Xiao Q, Bauer C, Layne T, Playdon M. The association between overnight fasting and body mass index in older adults: the interaction between duration and timing. Int J Obes (Lond) 2021 Mar;45(3):555-564 [FREE Full text] [doi: 10.1038/s41366-020-00715-z] [Medline: 33214704]
39. Nakamura Y, Walker BR, Ikuta T. Systematic review and meta-analysis reveals acutely elevated plasma cortisol following fasting but not less severe calorie restriction. Stress 2016;19(2):151-157 [FREE Full text] [doi: 10.3109/10253890.2015.1121984] [Medline: 26586092]
40. Burger HG, Dudley EC, Robertson DM, Dennerstein L. Hormonal changes in the menopause transition. Recent Prog Horm Res 2002;57:257-275. [doi: 10.1210/rp.57.1.257] [Medline: 12017547]
41. Rogan MM, Black KE. Dietary energy intake across the menstrual cycle: a narrative review. Nutr Rev 2023 Jun 09;81(7):869-886 [FREE Full text] [doi: 10.1093/nutrit/nuac094] [Medline: 36367830]
42. San-Millán I, Brooks GA. Assessment of metabolic flexibility by means of measuring blood lactate, fat, and carbohydrate oxidation responses to exercise in professional endurance athletes and less-fit individuals. Sports Med 2018 Feb;48(2):467-479 [FREE Full text] [doi: 10.1007/s40279-017-0751-x] [Medline: 28623613]
43. Reutrakul S, Van Cauter E. Sleep influences on obesity, insulin resistance, and risk of type 2 diabetes. Metabolism 2018 Jul;84:56-66. [doi: 10.1016/j.metabol.2018.02.010] [Medline: 29510179]

XSL•FO
RenderX

Abbreviations

BRANY: Biomedical Research Alliance of New York
CGM: continuous glucose monitoring
IRB: institutional review board
RER: respiratory exchange ratio

Edited by L Buis; submitted 23.07.24; peer-reviewed by Y Miao, P Shamanna; comments to author 04.02.25; revised version received 01.04.25; accepted 05.08.25; published 02.09.25.

Please cite as:
Yeshurun S, Cramer T, Souroujon D, Mor M
The Association of Previous Day Carbohydrate Consumption With Fasted, Exhaled Carbon Dioxide in Lumen Users: Retrospective Real-World Study
JMIR Mhealth Uhealth 2025;13:e64604
URL: https://mhealth.jmir.org/2025/1/e64604
doi:10.2196/64604
PMID:40772962

©Shlomo Yeshurun, Tomer Cramer, Daniel Souroujon, Merav Mor. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 02.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of a Self-Guided Transdiagnostic Intervention for Adults With Anxiety and Depression: Randomized Controlled Trial

Walter Staiano[1,2], PhD; Christine E Callahan[3], PhD; Michelle Davis[3], PhD; Leah Tanner[3], MSPH; Sarah Kunkle[3], MPH; Jenna Glover[3], PhD; Jon Kole[3,4], MD; Neeru Bakshi[3], MD; Marco Romagnoli[1], PhD; Ulrich Kirk[2,5], PhD

[1]Department of Physical Education and Sport, University of Valencia, Valencia, Spain

[2]Department of Psychology, University of Southern Denmark, Odense, Denmark

[3]Headspace Inc, Santa Monica, CA, United States

[4]Department of Psychiatry and Human Behavior, Brown University, Providence, RI, United States

[5]Fralin Biomedical Research Institute, Roanoke, VA, United States

Corresponding Author:
Ulrich Kirk, PhD
Department of Psychology
University of Southern Denmark
Campusvej 55
Odense, 5230
Denmark
Phone: 45 65502695
Email: ukirk@health.sdu.dk

Abstract

Background: Anxiety and depressive disorders affect hundreds of millions globally, with substantial barriers limiting access to evidence-based treatments like cognitive behavioral therapy. Digital mental health interventions offer potential solutions to improve access to support. However, evidence of efficacy through randomized controlled trials is essential for clinical adoption.

Objective: This study evaluated the efficacy of the Headspace Anxiety and Depression Program, a digitally delivered 21-session intervention grounded in the Unified Protocol that integrates cognitive behavioral and mindfulness-based strategies to target common drivers of emotional disorders, compared to a waitlist control group. Primary aims were to assess reductions in anxiety (Generalized Anxiety Disorder-7 [GAD-7]) and depression (Patient Health Questionnaire-8 [PHQ-8]) symptoms. Secondary aims examined improvements in sleep quality (Pittsburgh Sleep Quality Index), perceived stress (Perceived Stress Scale-10), mindfulness (Mindful Attention Awareness Scale), and overall well-being (Warwick-Edinburgh Mental Well-Being Scale), as well as clinical remission and treatment response rates.

Methods: A fully remote, 2-arm parallel-group randomized controlled trial was conducted with 168 participants (aged ≥18 years) who had clinically significant anxiety (GAD-7≥10) or depression symptoms (PHQ-8≥10). Participants were randomized to either the Headspace Anxiety and Depression Program (n=84) or waitlist control (n=84). The intervention consisted of 21 daily sessions (5-10 minutes each) delivered via the Headspace app. Primary (GAD-7 and PHQ-8) and secondary outcomes (Pittsburgh Sleep Quality Index, Perceived Stress Scale-10, Mindful Attention Awareness Scale, and Warwick-Edinburgh Mental Well-Being Scale) were assessed at baseline, postintervention, and a 3-week follow-up using intention-to-treat analysis with mixed-model ANOVAs.

Results: Study retention was high, with the majority (154/168, 91.7%) of participants completing the 3-week follow-up visit, and no serious adverse events were reported. Intervention adherence was high, with 82.1% (69/84) in the intervention group completing all 21 sessions. Significant group×time interactions were observed for both primary outcomes (P<.001). The Headspace group showed substantial reductions in anxiety symptoms (GAD-7: 34.5% reduction from baseline: mean 11.9, SD 2.8 to follow-up: mean 7.8, SD 2.3; η^2_p=0.350) and depression symptoms (PHQ-8: 33.9% reduction from baseline: mean 12.1, SD 2.8 to follow-up: mean 8.0, SD 2.0; η^2_p=0.370), while the control group did not show significant reductions. Combined anxiety and depression remission rates were significantly higher in the Headspace group (18/84, 21.4%) versus controls (7/84, 8.3%; P=.03), as were treatment response rates (23/84, 27.3% vs 2/84, 2.3%; P=.03). Participants in the Headspace group also demonstrated significant improvements in secondary outcomes, including sleep quality (30.2% improvement), perceived stress (13.2% reduction), mindfulness (10.3% increase), and mental well-being (10.7% increase).

XSL•FO
RenderX

Conclusions: The Headspace Anxiety and Depression Program demonstrated significant efficacy in reducing anxiety and depression symptoms comparable to established treatments. Benefits were sustained at the 3-week follow-up. These findings support the potential of evidence-based transdiagnostic digital mental health interventions to address critical gaps in mental health care delivery.

Trial Registration: ClinicalTrials.gov NCT06434883; https://clinicaltrials.gov/study/NCT06434883

(JMIR Mhealth Uhealth 2025;13:e79759) doi:10.2196/79759

KEYWORDS

cognitive behavioral therapy; anxiety; depression; mindfulness; meditation; mobile health; smartphone intervention; Headspace; Generalized Anxiety Disorder-7; GAD-7; Patient Health Questionnaire-8; PHQ-8

Introduction

Anxiety and depressive disorders represent 2 of the most prevalent and debilitating mental health conditions globally, affecting hundreds of millions of individuals worldwide and imposing substantial personal, social, and economic burdens [1]. In the United States, approximately 31% of adults report experiencing an anxiety disorder at some point in their lives, with 20% reporting anxiety symptoms within the past year [2]. Additionally, 13% of adolescents and adults experience depression, with 88% reporting functional impairment in work, home, and social activities [3]. For the purposes of this study, we focus on generalized anxiety symptoms as measured by the Generalized Anxiety Disorder-7 (GAD-7), which assesses worry, nervousness, irritability, and difficulty relaxing—core features of generalized anxiety disorder that often co-occur with depression and significantly impair daily functioning. The comorbidity between these conditions is particularly concerning, with approximately 50%-60% of those with anxiety symptoms also experiencing depression symptoms [4], which may further worsen mental health and overall quality of life.

Cognitive behavioral therapy (CBT) has been extensively validated as an effective treatment for both anxiety and depression disorders [5], with both in-person and teletherapy-delivered CBT showing consistent outcomes [6]. However, both in-person and teletherapy CBT face substantial barriers including insurance coverage limitations, treatment costs, uncertainty about where to seek treatment, lack of access to qualified therapists, and scheduling inconvenience due to workday constraints [7]. While teletherapy has made CBT more accessible, these barriers continue to limit the accessibility of evidence-based treatments for addressing the growing mental health needs of the population. Recognizing these barriers, empirically validated components of psychological interventions that were once delivered solely via CBT have increasingly been translated into digital mental health interventions (DMHIs) [8,9]. These DMHIs are often delivered via app or online, completed asynchronously, and require less time and monetary commitment than CBT. The emergence of the DMHIs represents a paradigm shift in treatment accessibility and delivery, offering potential solutions to overcome traditional barriers to care. Smartphone apps are one form of digital treatment delivery receiving substantial attention, with mental health apps for depression and anxiety being among the most widely downloaded categories of health apps [10]. The ubiquity of smartphone technology presents unprecedented opportunities to deliver evidence-based interventions directly to individuals in need, transcending traditional barriers of geography, cost, scheduling constraints, and stigma associated with help-seeking [11].

Recent meta-analytic evidence has provided increasingly robust support for the efficacy of DMHIs [12], where they showed that apps have overall small but significant effects on symptoms of depression (n=33,567; g=0.28) and generalized anxiety (n=22,394; g=0.26) compared to control conditions [13-15]. While evidence exists on the impact of DMHIs, there is still a critical shortage in the understanding of smartphone-delivered mental health interventions, as only 3% of apps that claimed to offer a therapeutic treatment for anxiety and depression had published peer-reviewed research to back up their assertions [16]. Additionally, engagement and adherence represent persistent challenges in DMHIs. Studies have demonstrated that engagement with DMHIs varies widely, with naturalistic use often showing considerable attrition even among those who initially download apps [17]. A lack of engagement with mental health apps has been reported as a key limitation for realizing the potential of apps to broadly disseminate mental health treatments.

This study addresses several critical gaps in the existing literature by conducting a randomized controlled trial (RCT) examining the efficacy of the Headspace Anxiety and Depression Program. The DMHI is grounded in the Unified Protocol (UP) [18], a transdiagnostic treatment that integrates cognitive behavioral and mindfulness-based strategies to target common drivers of emotional disorders, such as avoidance and inflexible thinking, seen across conditions like anxiety and depression. The Headspace Anxiety and Depression Program is a self-guided program consisting of 21 daily 5- to 10-minute sessions. This study evaluated the efficacy of the Headspace Anxiety and Depression Program by addressing both primary symptoms of depression and anxiety as well as secondary outcomes including sleep quality, perceived stress, mindfulness, and overall well-being. Furthermore, the study explored treatment response and remission rates, in addition to symptom reduction measures, to allow for the evaluation of clinical outcomes. The Headspace Anxiety and Depression Program is grounded in a transdiagnostic framework based on the UP, which recognizes that anxiety and depression share common underlying mechanisms, particularly emotion dysregulation and maladaptive coping strategies. The program operates through several interconnected mechanisms within this UP framework. First, mindfulness-based components enhance emotional

XSL•FO
RenderX

awareness and reduce automatic reactivity to distressing thoughts and feelings. Second, cognitive flexibility exercises help participants recognize and modify maladaptive thought patterns common in both anxiety and depression. Third, behavioral experiments and exposure components directly address avoidance behaviors that maintain both conditions. The integration of these approaches within the UP transdiagnostic framework targets shared vulnerability factors such as neuroticism and emotion dysregulation that underlie both anxiety and depressive disorders. While anxiety and depression have distinct symptomatic presentations, the transdiagnostic approach assumes that they share core underlying mechanisms, particularly related to emotion regulation difficulties, cognitive biases, and behavioral avoidance. The intervention addresses these common mechanisms while allowing for individual adaptation to specific symptom presentations. We hypothesized that participants using the Headspace Anxiety and Depression Program would demonstrate significantly greater reductions in depression and anxiety symptoms compared to a waitlist control group, with effect sizes comparable to those reported in recent meta-analyses (Cohen *d*≥0.26-0.28). Secondary hypotheses predicted that the intervention would also yield improvements in sleep quality, perceived stress levels, mindfulness, and overall well-being.

Methods

Study Design

This study was an app-based, 2-arm parallel-group RCT comparing the Headspace Anxiety and Depression Program (intervention) to a waitlist control group. The study followed CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines for reporting parallel group randomized trials [19]. A completed CONSORT checklist is provided in Multimedia Appendix 1.

Ethical Considerations

Study procedures were approved by the institutional review board of Virginia Tech (#21-295), and the study was preregistered (ClinicalTrials.gov NCT06434883). Participants were asked to sign an electronic consent form prior to enrolling in the study and were informed that they could opt out of the study at any time. All data were deidentified prior to analysis. Participants were compensated for taking part in the study using the following compensation structure: US $50 for completing the baseline assessment, US $50 for completing the postintervention assessment, and US $50 for completing the follow-up measures.

Participants

Overall, 514 potential participants were assessed for eligibility, with 168 meeting the inclusion criteria and enrolling in the study. Inclusion criteria were assessed through a structured online eligibility screening process. All potential participants completed the Patient Health Questionnaire-8 (PHQ-8) and GAD-7 as part of this screening to determine eligibility. Inclusion required (1) clinically significant symptoms of depression or anxiety as determined by validated cutoff scores (PHQ-8≥10 or GAD-7≥10) [20,21], (2) no prescription medication for anxiety or depressive symptoms or a stable dose for ≥4 weeks, (3) 18 years or older, (4) located in the United States, and (5) access to a compatible smartphone (Android or iOS). Exclusion criteria, also assessed by self-report during the eligibility screening phase, were (1) severe depression (PHQ-8>20) or risk of self-harm (positive response to the self-harm screening question, “The thought of harming myself has occurred to me.”); (2) not maintaining a stable dose of psychotropic medication for ≥4 weeks prior to enrollment; (3) psychiatric hospitalizations within the past 6 months; (4) self-reported diagnosis of any of the following conditions: schizophrenia, psychosis, bipolar disorder, seizure disorder, substance use disorder, recent trauma to the head or brain damage, severe cognitive impairment, serious physical health concerns necessitating surgery, or pregnancy; and (5) having completed CBT or other psychotherapy that included self-monitoring and cognitive or behavioral exercises delivered by a licensed therapist in the past 6 months. Individuals reporting self-harm risk were referred to mental health resources (National Alliance on Mental Illness helpline) per safety protocol.

Study Procedures

Participants were recruited from across the United States via targeted social media advertisements on Facebook and Instagram. The advertisements featured brief descriptions of the study purpose (evaluating a digital mental health program for anxiety and depression), eligibility criteria, and compensation details. Potential participants accessed a secure web-based platform using REDCap (Research Electronic Data Capture; Vanderbilt University) hosted at Virginia Tech. The platform facilitated all study procedures including eligibility screening, electronic informed consent, randomization, and data collection. Apart from app use data, all data were collected and managed through the REDCap platform. This infrastructure enabled fully remote participation via a web-enabled device. App use data was automatically captured and stored securely within Headspace’s servers.

Following the initial overview of the study, potential participants completed an eligibility screening to assess inclusion and exclusion criteria. After eligibility screening, the research team assessed the input from potential participants and notified eligible participants with instructions on how to proceed. Subsequently, eligible participants reviewed the consent document through the research platform, with an option to download it as a PDF before providing electronic consent.

Enrolled participants completed all baseline outcome assessments. Specifically, participants completed the primary outcome measures (PHQ-8 and GAD-7) and secondary outcome measures Perceived Stress Scale-10 (PSS-10) for measuring self-reported stress [22], Pittsburgh Sleep Quality Index (PSQI) for measuring sleep quality [23], Mindful Attention Awareness Scale (MAAS) for measuring mindfulness [24], and Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) for measuring mental well-being [25].

Following baseline data collection, participants were randomized to either the intervention (Headspace Anxiety and Depression Program) or the waitlist control group. The randomization sequence was created by the MATLAB randi() function to

XSL•FO
RenderX

provide an array of "1" or "2." The randomization sequence was uploaded to the REDCap study platform to allow for allocation concealment and automated randomization of enrolled participants. The sequence was concealed until analysis; however, 1 member of the research team was unblinded to track adherence and provide technical assistance. The research team emailed participants an auto-generated email based on their group allocation. Participants randomized to the intervention group received instructions on how to download and access the Headspace app including instructions about the Anxiety and Depression program. Participants were provided with written instructions to only access the Anxiety and Depression Program in the Headspace app during the study. Participants were provided with details about how to contact the research team if they experienced technical challenges related to app setup and app use. None of the participants in the study reported technical challenges when setting up the app or reported experiencing technical challenges during the study.

Participants responded to outcome measures at the end of the intervention period (postintervention assessment). Participants in the intervention group were asked not to access the Headspace Anxiety and Depression Program from the time of postintervention assessment until completion of the 3-week follow-up assessment to allow for evaluation of intervention durability without ongoing exposure. This protocol decision was made to maintain scientific rigor in assessing sustained effects, though we acknowledge this created an ethical tension between research methodology and participant benefit. This temporary limitation was included in the study protocol, approved by the institutional review board, and participants were informed of this during the consent process. Participants were advised that they could access other mental health resources if needed during this period. Follow-up outcome measures were collected 3 weeks following the postintervention assessment.

Safety

Safety was assessed through monitoring of adverse events throughout the study period. Participant responses indicating clinical deterioration or elevated risk were automatically flagged in the REDCap platform for review by the research team.

Headspace Anxiety and Depression Program (Intervention)

The Headspace Anxiety and Depression Program integrates CBT and mindfulness techniques similar to the approach used in the UP, a transdiagnostic approach to reducing anxiety and depression symptoms (Multimedia Appendix 2). The 21-session program (5-10 minutes per session, intended for daily completion) includes education and skill-building in 4 core areas: mindful emotional awareness (cultivating nonjudgmental recognition of thoughts, emotions, and behaviors that contribute to anxiety and depression), behavioral experiments (testing new actions in response to strong emotions), cognitive flexibility (shifting unhelpful thinking patterns), and emotional exposure (reducing avoidance of distressing emotions and situations; Multimedia Appendix 3). The theoretical framework underlying this intervention is based on the principle that anxiety and depression share common underlying mechanisms, particularly emotion dysregulation, cognitive inflexibility, and behavioral avoidance. The program systematically addresses these through a structured progression: early sessions build mindful awareness of emotional patterns, middle sessions introduce cognitive and behavioral tools for managing difficult emotions, and final sessions consolidate these skills for long-term application (Multimedia Appendix 4). Sessions 1-5 are focused on building awareness through mindfulness skills, values exploration, and self-monitoring. Sessions 6-14 introduce cognitive, behavioral, and emotional regulation strategies, emphasizing active experimentation and skill development; and sessions 15-21 reinforce these techniques through continued practice, with a focus on problem-solving, behavior change, and relapse prevention. Sessions 1-14 comprise the program's core content and are considered the minimum effective dose, as all key techniques are introduced during this phase. In contrast to other CBT programs, which often run 30-60 minutes per session [26], the content is delivered through brief daily videos (5-10 minutes) in the Headspace app, each paired with a simple, actionable suggestion (or "daily action") to help users apply the day's learning (eg, trying a new wind-down routine). To increase relatability and engagement, the program features animations and real-world video segments with mindfulness teachers, clinical experts, and individuals sharing personal experiences with anxiety and depression. This approach was intended to help users see their experiences reflected in the content and better understand how techniques can be adapted to apply them to everyday life. Participants received automated in-app reminders, but no additional human support beyond the technical assistance contact information provided at enrollment.

Waitlist Control

Participants in the waitlist control group were instructed to maintain their regular daily routines and activities throughout the study period. They received no intervention content during this time but completed all assessment measures at identical time points to the intervention group. Control participants were provided with contact information for mental health resources and crisis support services. To minimize potential differential attrition, waitlist participants were informed during consent that they would receive full access to the Headspace Anxiety and Depression Program immediately following the study completion.

Anxiety and Depression Symptoms (Primary Outcomes)

All questionnaires were collected at baseline, postintervention, and 3-week follow-up. Depression and anxiety symptom severity were assessed as primary outcomes using the PHQ-8 and the GAD-7, respectively. The PHQ-8 is an 8-item measure assessing depression symptoms (total score range 0-24; higher scores indicate greater symptom severity; Cronbach α=0.83) [19]. A score ≥10 indicates moderate depression. The GAD-7 is a 7-item measure assessing anxiety symptoms (total score range 0-21; higher scores indicate greater symptom severity; Cronbach α=0.89) [21]. A score ≥10 indicates moderate anxiety. Treatment response was defined as a ≥50% reduction in symptoms from baseline to the 3-week follow-up time point on PHQ-8 or

XSL•FO
RenderX

GAD-7. Remission was defined as an absolute score <5 at the 3-week follow-up time point on PHQ-8 or GAD-7.

Secondary Outcome Measures

Perceived stress was assessed using the PSS-10 (total score range 0-40; higher scores indicate higher perceived stress; Cronbach α=0.85) [22]. Sleep quality was assessed using the PSQI (total score range 0-21; higher scores indicate higher levels of sleep difficulties; Cronbach α=0.83) [23]. Mindfulness was assessed using the MAAS (total score range 15-90; higher scores indicate higher levels of trait mindfulness; Cronbach α=0.89) [24]. Mental well-being was assessed using the WEMWBS (total score range 14-70; higher scores indicate greater mental well-being; Cronbach α=0.91) [25]. Adherence to the Headspace Anxiety and Depression Program was obtained from in-app user data for each participant. Adherence or app use was calculated by summing the total number of completed sessions. Current psychotropic medication use was assessed at baseline and follow-up and included as a potential moderator in exploratory analyses.

Power Analysis

The sample size calculation was conducted using G*Power [27]. The study aimed to recruit 168 participants, with 84 randomized to each arm and with a 20% attrition rate. The sample size was determined based on an estimated medium effect (0.4-0.5) on postintervention scores between the intervention and waitlist arms, with 90% power, a 1:1 allocation ratio, and with α set at .05. The effect size is based on recent studies of smartphone-delivered psychological treatment for anxiety versus waitlist, which demonstrated a small to moderate between-group postintervention effect on anxiety symptoms when compared to any active or waitlist or inactive control condition [19,28].

Statistical Analysis

All analyses were conducted using the intention-to-treat approach. For the intention-to-treat analysis, we used multiple imputations to handle missing outcomes (11/168, 6.6%; values were missing at postintervention and follow-up). Multiple imputations (10 imputations) were used to achieve a relative efficiency of 99%, ensuring in-range values.

Assumptions of statistical tests for normal distribution and sphericity of data were checked using the Shapiro-Wilk test, histograms, quantile-quantile plots, and box plots. A series of mixed 2 (groups: intervention and waitlist control)×3 (time: preintervention, postintervention, and follow-up) ANOVAs was performed on GAD-7, PHQ-8, PSQI, PSS-10, MAAS, and WEMWBS. Covariates for elevated GAD-7 and PHQ-8 were included for potential confiding effects. To examine clinical end points, logistic regression analyses were conducted to test differences in treatment response and remission using the cutoff scores described earlier on GAD-7 and PHQ-8 across baseline and follow-up. Logistic regression analyses on remission and response status were performed to compare the 2 groups (intervention and control) at 3-week follow-up (outcome variables) in relation to preintervention values (covariates), without stepwise regression.

All significant interactions were followed up with relevant corrected pairwise comparisons using the Bonferroni method for simple main effects within each group. Where no significant interactions were found, the main effects of time and group were reported. Significance was set at .05 (2-tailed) for all analyses, and ANOVA effect sizes were calculated as partial eta squared (η^2_p), with 0.02, 0.13, and 0.26 indicating small, medium, and large effects, respectively. Data analysis was conducted using SPSS (version 27; IBM Corp). Statistical analyses were performed by a member of the research team who was not involved in the intervention delivery. The analyst was blinded to group allocation until completion of primary outcome analyses.

Results

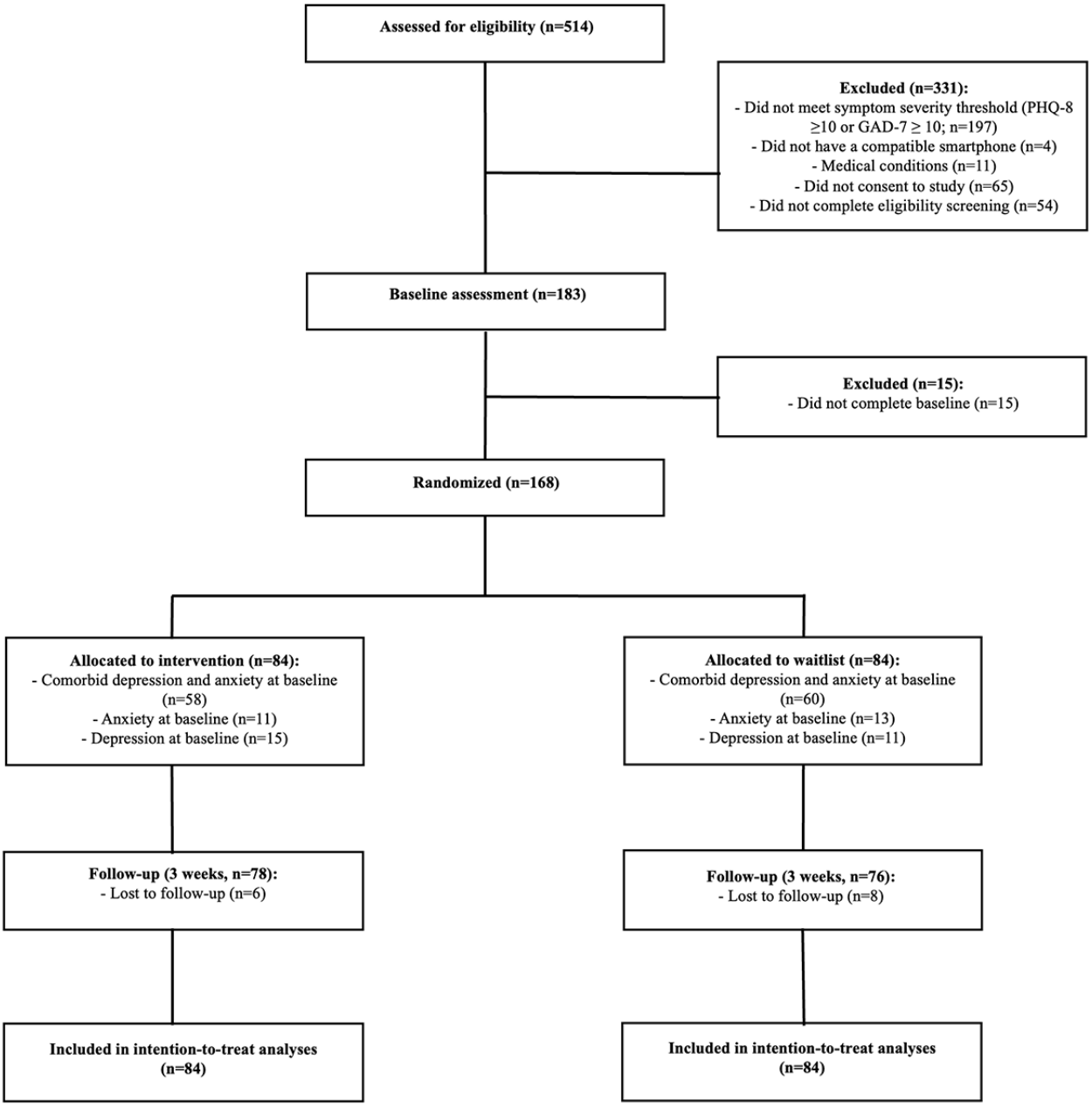
Overview

A total of 514 potential participants were assessed for eligibility, with 168 participants meeting the inclusion criteria and enrolled in the study (Figure 1). Participants were randomized equally between the intervention group (n=84) and the waitlist control group (n=84). Overall retention was high, with only 14 of 168 (8.3%) participants dropping out before the follow-up assessment (6 from the intervention group and 8 from the waitlist group).

Baseline demographic characteristics were well-balanced between groups (Table 1). The sample consisted predominantly of female participants (intervention: 57/84, 67.9% and control: 60/84, 71.4%), with a mean age of 34.8 (SD 9.6) years in the intervention group and 35.1 (SD 10.7) years in the waitlist control group. The sample was ethnically diverse, with approximately half identifying as White (intervention: 44/84, 52.4% and control: 43/84, 51.2%) and substantial representation of Black (intervention: 18/84, 21.4% and control: 43/84, 23.8%), Asian (intervention: 10/84, 11.9% and control: 8/84, 9.5%), and Hispanic (both groups: 8/84, 9.5%) participants. Most participants were employed (intervention: 60/84, 71.4% and control: 59/84, 70.3%) and had completed at least some college education (intervention: 73/84, 89.3% and control: 74/84, 88.1%). Psychotropic medication use was common at baseline, with 66.7% (56/84) of intervention participants and 63.1% (53/84) of control participants reporting current use. In both groups, the majority reported elevated baseline comorbid anxiety and depression symptoms (intervention: 58/84, 69% and control: 60/84, 71.4%). Psychotropic medication use remained stable from baseline to follow-up in both the intervention and waitlist control groups.

XSL•FO
RenderX

Figure 1. CONSORT diagram and recruitment flowchart. GAD-7: Generalized Anxiety Disorder-7; PHQ-8: Patient Health Questionnaire-8.



XSL•FO
RenderX

Table 1. Baseline demographics by study group of the complete sample.

	Headspace group (n=84)	Waitlist group (n=84)
Sex, n (%)		
Male	27 (32.1)	24 (28.6)
Female	57 (67.9)	60 (71.4)
Age (years), mean (SD)	34.8 (9.6)	35.1 (10.7)
Race or ethnicity, n (%)		
Asian	10 (11.9)	8 (9.5)
Black	18 (21.4)	20 (23.8)
Hispanic or Latino	8 (9.5)	8 (9.5)
White	44 (52.4)	43 (51.2)
Other	4 (4.8)	5 (6)
Employment, n (%)		
Employed	60 (71.4)	59 (70.3)
Unemployed	14 (16.7)	16 (19)
Retired or disability or other	10 (11.9)	9 (10.7)
Education, n (%)		
High school or less	9 (10.7)	10 (11.9)
College or associate degree	32 (38.1)	27 (32.1)
Bachelor's degree	28 (33.3)	33 (39.3)
Graduate degree or higher	15 (17.9)	14 (16.7)
Marital status, n (%)		
Married	33 (39.3)	36 (42.9)
In a relationship	18 (21.4)	17 (20.2)
Single	23 (27.4)	22 (26.2)
Divorced	10 (11.9)	9 (10.7)
Clinical characteristics, n (%)		
Both anxiety and depression (baseline PHQ-8[a]≥10 and GAD-7[b]≥10)	58 (69)	60 (71.4)
Depression only (baseline PHQ-8≥10, GAD-7<10)	15 (17.9)	11 (13.1)
Anxiety only (baseline GAD-7≥10, PHQ-8<10)	11 (13.1)	13 (15.5)
Current psychotropic medication		
Baseline, n (%)	56 (66.7)	53 (63.1)
Follow-up[c], n/N (%)	53/78 (67.9)	50/76 (65.8)

[a]PHQ-8: Patient Health Questionnaire-8.

[b]GAD-7: Generalized Anxiety Disorder-7.

[c]Follow-up percentages calculated from participants who completed the study (intervention: n=78 and control: n=76 after 14 dropouts: 6 intervention and 8 control).

Anxiety and Depression Symptoms

Both primary outcome measures demonstrated significant interactions between group and time, indicating differential treatment effects (Table 2). Anxiety symptoms (GAD-7) showed a significant group×time interaction (P<.001; η^2_p=0.350). Follow-up analyses revealed that participants in the intervention group experienced significant reductions in anxiety scores from baseline to postintervention (P<.001), baseline to follow-up (P<.001), and continued improvement from postintervention to follow-up (P=.009). Mean GAD-7 scores in the intervention group decreased from 11.9 (SD 2.8) at baseline to 8.2 (SD 2.6) after the intervention and 7.8 (SD 2.3) at follow-up, representing a 34.5% reduction from baseline (11.9 to 7.8). In contrast, the waitlist control group showed no significant changes across time points, with scores remaining stable around 12.2 at baseline, 11.8 after the intervention, and 11.6 at follow-up.

XSL•FO
RenderX

Significant main effects were observed for both time (P<.001; η^2_p=0.422) and group (P<.001; η^2_p=0.222).

Depression symptoms (PHQ-8) similarly demonstrated a significant group×time interaction (P<.001; η^2_p=0.370). The intervention group showed significant reductions from baseline to postintervention (P<.001) and baseline to follow-up (P<.001), though the change from postintervention to follow-up approached but did not reach significance (P=.05). Mean PHQ-8 scores decreased from 12.1 (SD 2.8) at baseline to 8.3 (SD 2.2) after the intervention and 8.0 (SD 2.0) at follow-up in the intervention group, representing a 33.9% reduction from baseline (12.1 to 8.0). The waitlist control group remained stable across all time points (11.8 to 11.4 to 11.3). Significant main effects were found for time (P<.001; η^2_p=0.472) and group (P<.001; η^2_p=0.163).

Table 2. Primary and secondary outcomes by group and time point for the complete sample.

	Baseline		Postintervention		Follow-up		Model results	
	Headspace group, mean (SD)	Waitlist control, mean (SD)	Headspace group, mean (SD)	Waitlist control, mean (SD)	Headspace group, mean (SD)	Waitlist control, mean (SD)	*P* value	Partial eta squared (η^2_p)
Anxiety (GAD-7[a])	11.9 (2.8)	12.2 (2.6)	8.2 (2.6)	11.8 (2.9)	7.8 (2.3)	11.6 (2.8)	<.001[b]	0.350
Depression (PHQ-8[c])	12.1 (2.8)	11.8 (2.7)	8.3 (2.2)	11.4 (2.6)	8.0 (2.0)	11.3 (2.9)	<.001[b]	0.370
Sleep quality (PSQI[d])	11.6 (2.0)	11.6 (1.7)	9.1 (2.2)	11.1 (2.4)	8.1 (2.6)	10.8 (2.3)	<.001[b]	0.172
Perceived stress (PSS-10[e])	27.2 (3.2)	26.9 (3.1)	25.0 (3.3)	26.9 (2.8)	23.6 (3.4)	26.2 (3.3)	<.001[b]	0.130
Mindfulness (MAAS[f])	2.9 (0.4)	2.9 (0.3)	3.1 (0.4)	2.9 (0.3)	3.2 (0.5)	3.0 (0.4)	<.001[b]	0.046
Well-being (WEMWBS[g])	36.4 (8.9)	36.3 (8.1)	37.7 (8.0)	37.1 (7.2)	40.3 (8.1)	37.7 (6.5)	.01[b]	0.030

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]Significant difference between groups (P<.05).

[c]PHQ-8: Patient Health Questionnaire-8.

[d]PSQI: Pittsburgh Sleep Quality Index.

[e]PSS-10: Perceived Stress Scale-10.

[f]MAAS: Mindful Attention Awareness Scale.

[g]WEMWBS: Warwick-Edinburgh Mental Well-Being Scale.

Remission Rates and Treatment Response

Exploratory analyses examining clinically meaningful outcomes revealed significant differences between groups. For remission (defined as scores <5) [20,21,29], GAD-7 remission rates increased significantly in the intervention group over time (P=.03), with 11 of 84 (13.1%) participants achieving remission at follow-up compared to 3 of 84 (3.5%) in the waitlist control group (P=.12). PHQ-8 remission rates showed improvement in the intervention group (7/84, 8.3%) versus control group (4/84, 4.7%), though this difference was not statistically significant (intervention: P=.22 and control: P=.41). When combining remission rates across both measures, the intervention group showed significantly higher rates (18/84, 21.4%) compared to controls (7/84, 8.3%; intervention: P=.03 and control: P=.41).

Treatment response (defined as ≥50% symptom reduction) [20,21,29] also favored the intervention group. GAD-7 response rates increased significantly in the intervention group (P=.03), with 9 of 84 (10.7%) participants achieving response at follow-up compared to 1 of 84 (1.1%) in waitlist controls (P=.57). PHQ-8 response rates were higher in the intervention group (14/84, 16.6%) versus waitlist controls (1/84, 1.1%; intervention: P=.03 and control: P=.37). Combined treatment response rates across both measures were significantly higher in the intervention group (23/84, 27.3%) compared to waitlist controls (2/84, 2.3%; intervention: P=.03 and waitlist control: P=.39).

Secondary Outcomes

All secondary outcome measures showed significant improvements in the intervention group compared to the waitlist control group (Table 2). Sleep quality (PSQI) demonstrated a significant group×time interaction (P<.001; η^2_p=0.172). The intervention group showed significant improvements from baseline to postintervention (P<.001) and baseline to follow-up (P<.001) and continued improvement from postintervention to follow-up (P<.001). Mean scores improved from 11.6 (SD 2.0) at baseline to 9.1 (SD 2.2) after the intervention and 8.1 (SD 2.6) at follow-up, representing a 30.2% improvement from baseline. The waitlist control group showed minimal change, though a small improvement from baseline to follow-up was observed (P=.04). Main effects were significant for time (P<.001; η^2_p=0.332) and group (P<.001; η^2_p=0.149).

Perceived stress (PSS-10) showed a significant group×time interaction (P<.001; η^2_p=0.130). The intervention group demonstrated significant reductions across all time points: baseline to postintervention (P<.001), baseline to follow-up

XSL•FO
RenderX

(*P*<.001), and postintervention to follow-up (*P*<.001). Mean scores decreased from 27.2 (SD 3.2) at baseline to 25.0 (SD 3.3) after the intervention and 23.6 (SD 3.4) at follow-up, representing a 13.2% reduction from baseline. No significant changes were observed in the waitlist control group. Significant main effects emerged for time (*P*<.001; η^2_p=0.239) and group (*P*=.002; η^2_p=0.052).

Mindfulness (MAAS) exhibited a significant group×time interaction (*P*<.001; η^2_p=0.046). The intervention group showed significant increases in mindfulness scores from baseline to postintervention (*P*<.001), baseline to follow-up (*P*<.001), and postintervention to follow-up (*P*<.001), improving mean scores from 2.9 (SD 0.4) at baseline to 3.1 (SD 0.4) after the intervention to 3.2 (SD 0.5) at follow-up, representing a 10.3% increase from baseline. The control group remained stable. A significant main effect of time was found (*P*<.001; η^2_p=0.176) but not for group (*P*=.05; η^2_p=0.023).

Mental well-being (WEMWBS) demonstrated a significant group×time interaction (*P*=.01; η^2_p=0.030). The intervention group showed significant improvements from baseline to postintervention (*P*=.02), baseline to follow-up (*P*<.001), and postintervention to follow-up (*P*<.001). Mean scores increased from 36.4 (SD 8.9) at baseline to 37.7 (SD 8.0) after the intervention and 40.3 (SD 8.1) at follow-up, representing a 10.7% increase from baseline. No significant changes occurred in the control group. A significant main effect of time was observed (*P*<.001; η^2_p=0.114), but not for group.

Adherence

Overall retention was high, with 154 of 168 (91.7%) participants completing all 3 study time points. Additionally, adherence to the Headspace Anxiety and Depression Program was high, with participants completing an average of 18.7 (SD 5.5) sessions of the 21 total sessions. The majority of participants (69/84, 82.1%) completed all 21 sessions, and 89.2% (75/84) completed the core elements of the program (ie, 14 sessions where all techniques are introduced). Only 9 (10.7%) participants did not complete the first module (ie, 5 sessions). No significant correlation was found between the number of completed sessions and treatment response across all outcome variables (all *P*s=.14).

Medication Status

Psychotropic medication use remained stable throughout the study. At baseline, 66.7% (56/84) of intervention participants and 63.1% (53/84) of waitlist control participants reported current medication use. At follow-up, these rates were 67.9% (53/78) and 65.8% (50/76), respectively, with no significant changes in either group. When medication status was included as a covariate in analyses, no significant interactions or main effects were observed for any outcome variable (all *P*s>.15; η^2_p<0.010).

Safety

No serious adverse events were reported during the study. Safety monitoring was conducted through automated flagging in the REDCap platform for clinical deterioration or elevated risk. No technical problems or privacy breaches were reported by participants or detected by the research team. No participants reported difficulty accessing or using the Headspace app interface.

Discussion

Principal Findings

This RCT demonstrates that the Headspace Anxiety and Depression Program significantly reduced anxiety and depression symptoms and improved overall well-being in adults with clinically significant anxiety or depression symptoms. Participants in the Headspace group showed significant and sustained improvements across all primary and secondary outcomes, with effect sizes ranging from small to large. The program's benefits were maintained at a 3-week follow-up, and remission and treatment response rates were significantly higher in the Headspace group versus the waitlist control. While this study did not evaluate outcomes in comparison to an existing intervention, the observed effect sizes for anxiety (η^2_p=0.35) and depression (η^2_p=0.37) symptom reduction represent large effects and are comparable to established interventions [5,30]. Meta-analytic evidence indicates that teletherapy-delivered CBT achieves similar outcomes to in-person treatment, with effect sizes typically ranging from 0.4 to 0.8 for anxiety and depression symptoms [30].

The 34.5% reduction in anxiety symptoms and 33.9% reduction in depression symptoms observed in our study align with recent meta-analytic findings on DMHIs. Linardon et al [13] reported overall small but significant effects for apps on depression (*g*=0.28) and anxiety (*g*=0.26) symptoms, based on analysis of 176 RCTs examining content-based smartphone interventions incorporating therapeutic elements such as CBT, mindfulness, or behavioral activation. Our intervention demonstrated larger effects, potentially due to the evidence-based, transdiagnostic approach combining CBT and mindfulness techniques focusing on improving emotional regulation. Additionally, the structured, sequential 21-session format ensures comprehensive skill development and consistent engagement. Importantly, 21.4% (18/84) of participants achieved combined remission across both anxiety and depression measures compared to only 8.3% (7/84) in the control group. These remission rates, while modest compared to intensive clinical interventions, represent meaningful recovery for individuals who might otherwise lack access to evidence-based treatment. The demonstrated improvement for participants with both elevated depression and anxiety symptoms also signals the utility of applying a transdiagnostic approach that targets common mechanisms for emotion dysregulation across anxiety and depression. Given the population included in this study (individuals with elevated but not severe anxiety and depression symptoms) and associated improvement, our findings suggest that this program could be a scalable and cost-effective alternative for individuals preferring self-guided treatment or as an adjunct to treatment with a licensed professional (eg, coach, therapist, and psychiatrist).

Beyond primary anxiety and depression outcomes, participants demonstrated significant improvements across all secondary

XSL•FO
RenderX

measures, including sleep quality, perceived stress, mindfulness, and overall mental well-being. The improvement in sleep quality is particularly relevant, given the bidirectional relationship between sleep disturbances and mood disorders. Specifically, previous research demonstrates that 50%-80% of individuals with anxiety and depression experience comorbid sleep problems [31]. The increase in mindfulness scores suggests that participants successfully acquired and applied mindfulness skills, which may contribute to long-term resilience and symptom management [32]. These findings are consistent with extensive research on mindfulness-based interventions for anxiety and depression, which typically show moderate effect sizes (anxiety: d=0.59 and depression: d=0.37) when delivered as stand-alone interventions [33]. The Headspace Anxiety and Depression Program's integration of mindfulness within a structured CBT framework may have provided participants with both contemplative practice (mindfulness) and practical cognitive tools (CBT) to recognize, understand, and respond to difficult emotions and thoughts. We also note that all mindfulness techniques incorporated into the program were introduced in the first 14 sessions, which 89.2% (75/84) of participants completed. This high engagement may have impacted overall mindfulness outcomes. The improvements in mental well-being are particularly noteworthy, as these measures are typically challenging to influence and represent broader functional improvements beyond symptom reduction.

The high retention (154/168, 91.7%) and adherence rates (69/84, 82.1% completed all sessions) observed in this study are also notable. These rates substantially exceed those typically reported in DMHI research, where adherence rates often range from 50% to 70% [17,28], with Headspace's similarly designed Sleep Program RCT achieving 72% adherence [34]. Several factors likely contributed to these high engagement rates. The Anxiety and Depression Program's design emphasized accessibility through brief daily sessions (5-10 minutes), representing a significantly lower time commitment than traditional CBT (typically 50-90 minutes weekly) or even digital CBT programs (often 30-60 minutes per session). Additionally, the sequential structure of the Anxiety and Depression Program, which builds skills progressively across 3 distinct modules, may have enhanced motivation and completion rates. This structured approach contrasts with digital interventions that offer less guided, more self-directed content access. Qualitative data from the follow-up surveys support these reasons for high engagement with participants, indicating that the program being app-based made it very easy to access, the length of the sessions made it doable to fit into their daily schedule, and the program structure making it easy to follow. However, it is important to acknowledge that the financial compensation structure (US $150 total) may have artificially inflated adherence rates, potentially limiting the generalizability of these findings to real-world implementation where such incentives are not provided.

This study's sample reported high baseline psychotropic medication use (intervention: 56/84, 66.7% and control: 53/84, 63.1%), which remained stable throughout the study. This prevalence exceeds general population estimates of antidepressant use (13.2% in US adults during 2015-2018) [32]. The stability of medication use suggests that the intervention's benefits were independent of pharmacological changes, suggesting that the intervention can enhance outcomes for individuals already receiving pharmacological care.

Strengths, Limitations, and Future Research

To our knowledge, this is the first RCT investigating the impact of a self-guided DMHI that uses a transdiagnostic approach with an intentional focus on integrating mindfulness and CBT techniques in a structured, sequential format specifically designed to improve anxiety and depression symptoms. Unlike many digital CBT programs that incorporate mindfulness as a secondary component, the Headspace program was designed with mindfulness as a core therapeutic element. The short daily time commitment, broad commercial availability via an existing app, and the adherence data suggest that the intervention can be practically integrated in a real-world setting. Despite these strengths, the study did have several limitations. First, this study used a waitlist control design, rather than an active control that might provide differential benefits. Second, the relatively short 3-week follow-up period limits conclusions about the long-term durability of treatment effects. Future studies should incorporate longer-term follow-up assessments to evaluate the sustainability of clinical benefits. Third, the moderate baseline symptom severity in our sample may limit the intervention's applicability to individuals with more severe symptoms. While we excluded participants with severe depression (PHQ-8>20) for safety reasons, future research should examine the program's effectiveness across the full spectrum of symptom severity. Fourth, while our recruitment approach through social media advertisements may have introduced self-selection bias toward individuals already interested in digital mental health solutions, this limitation is somewhat mitigated by our use of validated screening instruments to confirm clinically significant symptom levels. Fifth, participants were recruited based on validated screening questionnaire scores using self-report symptoms rather than professional clinical diagnoses. While the PHQ-8 and GAD-7 are well-validated screening tools with established clinical cutoffs, they do not replace formal diagnostic evaluation. Future research areas include expanding upon the current RCT to evaluate the effectiveness of the Headspace Anxiety and Depression Program in a real-world setting and examining the use of the program alongside Headspace's meditation content library and human-delivered care (mental health coaching, therapy, and psychiatry). Future research should also explore how evidence-based DMHIs like the Headspace program might work synergistically with other innovative approaches to well-being, including digital gaming interventions, virtual reality applications, and other media-based therapeutic tools [33]. The integration of multiple digital modalities may offer personalized pathways to mental health improvement.

Conclusions

The Headspace Anxiety and Depression Program represents an efficacious, accessible, and scalable approach to mental health treatment for individuals with clinically significant anxiety and depression symptoms. The transdiagnostic intervention consisting of 21 daily brief 5- to 10-minute sessions demonstrated substantial improvements across multiple domains

XSL•FO
RenderX

of mental health and well-being, with benefits maintained at follow-up. These findings support the potential for evidence-based DMHIs to address critical gaps in mental health care delivery, particularly for populations who face barriers accessing traditional treatment modalities or have a preference for self-guided care.

Acknowledgments

This study was funded by Headspace.

Data Availability

The deidentified datasets generated and analyzed during the study are available from the corresponding author on reasonable request.

Authors' Contributions

WS and MR were involved in data analysis, interpretation of data, and writing of the paper. UK was involved in the study design, data collection, data analysis, interpretation of data, and writing of the paper. CEC, MD, LT, SK, JG, JK, and NB were involved in the study design, interpretation of data, and writing of the paper.

Conflicts of Interest

CEC, MD, LT, SK, JG, JK, and NB are currently employed by Headspace.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 460 KB - mhealth_v13i1e79759_app1.pdf]

Multimedia Appendix 2
Differences between the Headspace Anxiety and Depression Program, other digital cognitive behavioral therapy programs, and traditional cognitive behavioral therapy.
[DOCX File , 19 KB - mhealth_v13i1e79759_app2.docx]

Multimedia Appendix 3
Headspace Anxiety and Depression Program session outline and descriptions.
[DOCX File , 19 KB - mhealth_v13i1e79759_app3.docx]

Multimedia Appendix 4
Headspace Anxiety and Depression Program within the Headspace app.
[PNG File , 46 KB - mhealth_v13i1e79759_app4.png]

References

1. Santomauro DF, Herrera AMM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021;398(10312):1700-1712. [doi: 10.1016/s0140-6736(21)02143-7]
2. Any anxiety disorder. National Institute of Mental Health (NIMH). 2023. URL: https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder [accessed 2025-09-27]
3. Brody DJ, Hughes JP. Depression prevalence in adolescents and adults: United States, August 2021–August 2023. NCHS Data Brief. 2025. URL: https://www.cdc.gov/nchs/products/databriefs/db527.htm [accessed 2025-09-27]
4. Hirschfeld RMA. The comorbidity of major depression and anxiety disorders: recognition and management in primary care. Prim Care Companion J Clin Psychiatry 2001;3(6):244-254 [FREE Full text] [doi: 10.4088/pcc.v03n0609] [Medline: 15014592]
5. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
6. Lin A, Espay AJ. Remote delivery of cognitive behavioral therapy to patients with functional neurological disorders: promise and challenges. Epilepsy Behav Rep 2021;16:100469. [doi: 10.1016/j.ebr.2021.100469] [Medline: 34409282]
7. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2013;44(6):1303-1317. [doi: 10.1017/s0033291713001943]

XSL•FO
RenderX

8. Brinsley J, O'Connor EJ, Singh B, McKeon G, Curtis R, Ferguson T, et al. Effectiveness of digital lifestyle interventions on depression, anxiety, stress, and well-being: systematic review and meta-analysis. J Med Internet Res 2025;27:e56975. [doi: 10.2196/56975] [Medline: 40112295]
9. Lau CKY, Saad A, Camara B, Rahman D, Bolea-Alamanac B. Acceptability of digital mental health interventions for depression and anxiety: systematic review. J Med Internet Res 2024;26:e52609. [doi: 10.2196/52609] [Medline: 39466300]
10. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
11. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-11-13]
12. Valentine L, Hinton JDX, Bajaj K, Boyd L, O'Sullivan S, Sorenson RP, et al. A meta-analysis of persuasive design, engagement, and efficacy in 92 RCTs of mental health apps. NPJ Digit Med 2025;8(1):229 [FREE Full text] [doi: 10.1038/s41746-025-01567-5] [Medline: 40301581]
13. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. World Psychiatry 2024;23(1):139-149 [FREE Full text] [doi: 10.1002/wps.21183] [Medline: 38214614]
14. Kirk U, Staiano W, Hu E, Ngnoumen C, Kunkle S, Shih E, et al. App-Based Mindfulness for Attenuation of Subjective and Physiological Stress Reactivity in a Population With Elevated Stress: Randomized Controlled Trial. JMIR Mhealth Uhealth 2023 Oct 13;11(10312):e47371-e41712 [FREE Full text] [doi: 10.2196/47371] [Medline: 37831493]
15. Piil F, Axelsen J, Staiano W, Kirk U. Mindfulness Passes the Stress Test: Attenuation of Behavioral Markers of Mind Wandering During Acute Stress. J Cogn Enhanc 2020 Oct 07;5(2):155-163 [FREE Full text] [doi: 10.1007/s41465-020-00193-1] [Medline: 37831493]
16. Marshall JM, Dunstan DA, Bartik W. The digital psychiatrist: in search of evidence-based apps for anxiety and depression. Front Psychiatry 2019;10:831 [FREE Full text] [doi: 10.3389/fpsyt.2019.00831] [Medline: 31803083]
17. Renn BN, Walker TJ, Edds B, Roots M, Raue PJ. Naturalistic use of a digital mental health intervention for depression and anxiety: a randomized clinical trial. J Affect Disord 2025 Jan 01;368:429-438. [doi: 10.1016/j.jad.2024.09.104] [Medline: 39299591]
18. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, et al. The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. JAMA Psychiatry 2017;74(9):875-884. [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
19. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332. [doi: 10.1136/bmj.c332] [Medline: 20332509]
20. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
21. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
22. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [doi: 10.2307/2136404]
23. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
24. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
25. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
26. How long does psychotherapy take? Institute for Quality and Efficiency in Health Care (IQWiG). 2013. URL: https://www.ncbi.nlm.nih.gov/books/NBK618444/#n0079.howlongdoespsychothe [accessed 2025-09-27]
27. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
28. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, et al. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. JAMA Psychiatry 2020;77(9):906-914. [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
29. Coley RY, Boggs JM, Beck A, Hartzler AL, Simon GE. Defining success in measurement-based care for depression: a comparison of common metrics. Psychiatr Serv 2020;71(4):312-318. [doi: 10.1176/appi.ps.201900295] [Medline: 31847739]
30. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]

XSL•FO
RenderX

31. Harvey AG. Sleep and circadian rhythms in bipolar disorder: seeking synchrony, harmony, and regulation. Am J Psychiatry 2008;165(7):820-829. [doi: 10.1176/appi.ajp.2008.08010098] [Medline: 18519522]
32. Brody DJ, Gu Q. Antidepressant use among adults: United States, 2015-2018. NCHS Data Brief 2020(377):1-8. [Medline: 33054926]
33. Arigayota A, Duffek B, Hou C, Eisingerich AB. Effects of The Legend of Zelda: Breath of the Wild and Studio Ghibli films on young people's sense of exploration, calm, mastery and skill, purpose and meaning, and overall happiness in life: exploratory randomized controlled study. JMIR Serious Games 2025;13:e76522. [doi: 10.2196/76522] [Medline: 40750097]
34. Staiano W, Callahan C, Davis M, Tanner L, Coe C, Kunkle S, et al. Assessment of an app-based sleep program to improve sleep outcomes in a clinical insomnia population: randomized controlled trial. JMIR Mhealth Uhealth 2025;13:e68665. [doi: 10.2196/68665] [Medline: 40267472]

Abbreviations

CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
DMHI: digital mental health intervention
GAD-7: Generalized Anxiety Disorder-7
MAAS: Mindful Attention Awareness Scale
PHQ-8: Patient Health Questionnaire-8
PSQI: Pittsburgh Sleep Quality Index
PSS-10: Perceived Stress Scale-10
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
UP: Unified Protocol
WEMWBS: Warwick-Edinburgh Mental Well-Being Scale

Edited by L Buis; submitted 28.06.25; peer-reviewed by R Weiskittle, A Eisingerich; comments to author 11.09.25; revised version received 16.09.25; accepted 25.09.25; published 23.10.25.

Please cite as:
Staiano W, Callahan CE, Davis M, Tanner L, Kunkle S, Glover J, Kole J, Bakshi N, Romagnoli M, Kirk U
Efficacy of a Self-Guided Transdiagnostic Intervention for Adults With Anxiety and Depression: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e79759
URL: https://mhealth.jmir.org/2025/1/e79759
doi:10.2196/79759
PMID:

©Walter Staiano, Christine E Callahan, Michelle Davis, Leah Tanner, Sarah Kunkle, Jenna Glover, Jon Kole, Neeru Bakshi, Marco Romagnoli, Ulrich Kirk. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Requirements and Concerns of Individuals Remitted From Depression for an Early Relapse Detection mHealth App: Focus Group Study

Tina Coenen[1], MA; Matthias Maerevoet[2], MA; Stephanie Chen[3], MA; Mathias De Brouwer[3], PhD; Sofie Van Hoecke[3], PhD; Ernst HW Koster[2], PhD; Mariek MP Vanden Abeele[1], PhD; Klaas Bombeke[1], PhD

[1]imec-mict-UGent, Department of Communication Sciences, Ghent University, Ghent, Belgium

[2]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

[3]IDLab, Ghent University – imec, Ghent, Belgium

Corresponding Author:
Tina Coenen, MA
imec-mict-UGent
Department of Communication Sciences
Ghent University
Miriam Makebaplein 1
Ghent, 9000
Belgium
Phone: 32 9 264 97 07
Email: tina.coenen@ugent.be

Abstract

Background: Major depressive disorder is often a recurrent condition, with a high risk of relapse for individuals remitted from depression. Early detection of relapse is critical to improve clinical outcomes. Mobile health (mHealth) technologies offer new opportunities for real-time monitoring and prevention of relapse, if the user requirements of the target population are effectively implemented.

Objective: This study investigated the requirements and concerns of individuals remitted from depression for an mHealth app aimed at monitoring depressive symptoms and detecting early signs of relapse through integrating both active ecological momentary assessment data and passive data from the user's smartphone and smartwatch.

Methods: Three focus group discussions were conducted with 17 participants remitted from depression. Before the focus group, participants had gained some experience with an in-house designed ecological momentary assessment monitoring app, prompting questions regarding their mood multiple times throughout the day. During the focus groups, feedback and insights were gathered on participants' expectations, requirements, concerns, and attitudes toward a depression monitoring app. A thematic analysis was performed to identify recurring themes and subthemes, shedding light on the desired user experience and functionalities.

Results: We identified 5 main themes. Participants highlighted (1) a need for customization settings, particularly in terms of data collection and sharing, and frequency of self-assessments. They also valued (2) positivity in the app's design through positive reinforcement and journaling features. Additionally, participants emphasized (3) interventions to be the main motivator for adoption and long-term usage. More specifically, they wanted the app to foster self-awareness, self-reflection, and insights, and to offer support during deteriorations in mental health. Furthermore, participants deemed (4) transparency in data use and machine learning predictions to be essential for building trust. Participants required these functionalities to bear (5) the user burdens of self-monitoring. Key concerns were for passive monitoring to cause a privacy burden and for active monitoring to raise an emotional burden.

Conclusions: Considering the vulnerability of potential users, the design of an mHealth app for early depression relapse detection should be guided by user preferences and approached with caution. Requirements for customization, positivity, interventions, and transparency must be addressed, while minimizing both the emotional and privacy burden. Future iterations should implement these findings to improve and test the app's acceptability, adoption, and usability for clinical use.

(JMIR Mhealth Uhealth 2025;13:e67141) doi:10.2196/67141

XSL•FO
RenderX

KEYWORDS

mHealth; mobile health; apps; mental health; depression; relapse prevention; self-monitoring; ecological momentary assessment; passive monitoring; smartphone; mobile phone; wearable devices; qualitative research; thematic analysis

Introduction

Background

Major depressive disorder is a debilitating condition that is associated with tremendous personal suffering and substantial societal costs [1,2]. It is the most prevalent of all mental disorders [3]. Moreover, its prevalence is increasing over time [4], and it is often a recurrent or even chronic disorder [5]. Despite the availability of pharmacological and behavioral treatments that can be quite effective in the acute phase, the risk of relapse remains high for individuals remitted from depression [6-8], with depressive remission being defined as the asymptomatic state after a depressive episode [9].

A contributing factor to the high relapse rates is the presence of both a treatment gap [10] and lag [11]. That is, many patients—even those who have previously sought treatment for depression [12]—only seek help when depressive symptoms become severe [13] or persist for a longer period [14]. Provided the reality of waiting lists, this often means that they only receive mental health care when a full-blown depressive episode has unfolded, and intensive treatment is required. Hence, delayed treatment is harmful to patients [15-17]. Recovery rates decline while the risk of persistence increases if patients do not start treatment within the first few months after the manifestation of an episode [18]. Should patients become aware of the re-emergence of depressive symptoms sooner and, therefore, seek help at an earlier stage, symptoms are less likely to escalate. Early detection and intervention are therefore paramount. Scholars and practitioners have broadened their focus accordingly to preventing the onset of depression, especially among individuals remitted from depression [19-23]. Of key importance to such prevention is predicting if and when there will be a novel depressive episode, in which case patients should be notified and encouraged to seek help early on.

In this realm, recent innovations in mobile health (mHealth), that is, the application of mobile technologies for health care services [24], are promising for personalized mental health monitoring. The ubiquity of affordable smartphones and—to a lesser extent—smartwatches makes mHealth increasingly accessible to patients. mHealth technology also provides opportunities for empowerment by fostering self-management. mHealth apps might have less stigma associated with them, lowering the barrier for adoption. Furthermore, most review papers conclude that mHealth interventions are promising in terms of cost-effectiveness as well, especially for depression management [25-29]. The current landscape of depression management apps largely consists of functionalities providing therapeutic support, intervention, treatment, and psychoeducation, followed by tracking of mood, behavior, or thoughts and symptom management [30-32].

In contrast to the myriad of therapeutic apps currently available in the marketplace, this study focuses on an mHealth app for prevention through monitoring depressive symptoms. More specifically, it is about a future app that captures and feeds both active and passive data sources into a personalized supervised machine learning model, being trained for just-in-time detection of early warning signs that are associated with depression relapse. Active data is gathered through ecological momentary assessment (EMA) questionnaires that participants can answer on their smartphones. EMA is a monitoring strategy to assess users' feelings, behavior, and context in real time in natural settings, thus maximizing ecological validity while avoiding retrospective recall [33,34]. Passive data is sourced from the user's smartphone and smartwatch. Smartphone data contain phone usage patterns (eg, screen time per app), revealing a user's digital footprint. Additionally, smartphones include a multitude of sensors (eg, accelerometers, GPS location sensors, and light and proximity sensors), providing detailed measurements about users' daily life and their environment. Similarly, smartwatches capture physiological, behavioral, and contextual data (eg, heart rate and step count) in a noninvasive and continuous way.

Passive data tracking via these mobile devices is increasingly relevant for digital phenotyping in mental health care. Digital phenotyping is defined as the "moment-by-moment quantification of the individual-level human phenotype in-situ using data from smartphones and other personal digital devices" [35]. The promise of digital phenotyping through mobile devices is in its unprecedented opportunity to objectively and continuously measure an individual's real-world functioning. It offers novel ways of understanding and measuring disease manifestations, previously reliant on subjective or self-reported data [36]. More specifically, it could be a fruitful approach to identifying symptoms and early signals of depression relapse [37,38]. Furthermore, recent technological advancements in predictive algorithms, machine learning techniques, and artificial intelligence can facilitate making sense of the unprecedented amounts of granular passive data [39] and pave the way toward personalized prevention for depression [40,41]. For example, the relationship between disturbed sleep and depression is well-established [42-44]. Hence, digital phenotyping by tracking changes in sleep patterns via a smartphone or smartwatch may indicate an impending depressive episode [45].

Smartphones are especially well-suited for digital phenotyping due to their worldwide adoption and users' frequent engagement with them. Combining smartphone data with the wealth of information from smartwatch data [36] and real-time self-reported data through EMA can maximize the utility of digital phenotyping for prevention [46]. Despite the clinical value of this information, respecting patient privacy should remain a priority when collecting their personal data [35,37,47]. It is also important to note that accessibility to a smartwatch and a smartphone with a wireless plan is not a given for everyone. The digital divide is particularly visible among ethnic minorities, those who are older, less educated, have a lower household income, and those who have a disability [48]. However, in the near future, it can be reasonably expected that most of the world will have access to mobile devices. In the

XSL•FO
RenderX

meantime, the combination of smartphone, smartwatch, and EMA data is worth exploring. Indeed, answers to EMA questionnaires enrich a passive dataset with labels, with which supervised machine learning models can be trained. Building upon the previous example, passively monitored sleep patterns can be labeled with a daily morning EMA question on subjective sleep quality.

Notwithstanding the potential of mHealth apps in preventing depression [49-51] or at least in preventing the transition from minor to severe depressive symptoms [52], engagement is low. Most of mHealth app users try an app only once [53,54]. mHealth trials endure high dropout rates, giving clues for real-life adoption problems [55]. There is also room for improvement when it comes to engagement in the long term [39]. For the population with a history of depression in particular, the lack of motivation and behavioral avoidance inherent to the mood disorder [56] might pose a major challenge to engagement and retention [57]. Furthermore, research shows that the implementation of mHealth tools in routine mental health care often does not meet expectations [58]. This is because certain prerequisites for achieving successful implementation of mHealth in practice are not met. mHealth tools need to be accepted by patients, appropriate in addressing their mental health disorder, and meet their expectations and preferences. This has thus far not always been the case. Our study closes this gap by unraveling the specific requirements and concerns of individuals remitted from depression in particular, whose opinions have been under-researched. The choice for this population goes hand in hand with the shift in mental health care from reactive to preventive approaches [59-61]. Considering the potentially far-reaching personal, social, and ethical implications this technology can have [62], user-centric design is warranted. As such, we make sure the app is acceptable, aligned with their preferences, and usable, to ensure future adoption [63].

Objectives

Unlike previous research, which focused on preferences of individuals with current depression for interventional apps based on active data collection, this study examines the requirements and concerns of those in remission toward an early relapse detection app based on both active and passive data collection. We conducted focus groups to gauge the attitudes of individuals remitted from depression toward self-tracking. A major question was to what extent these individuals are open to the idea of monitoring early warning signs of relapse. We assessed the role technology can play in this regard. We investigated adoption intention and participants' preferences regarding the app and its functionalities. We examined the differences between active and passive monitoring in terms of privacy and other user burdens. Thematic analysis revealed 5 main themes (ie, customization, positivity, interventions, transparency, and user burden) representing the different requirements and concerns participants voiced regarding a prevention tool, the design of which should carefully consider an equilibrium between both.

Methods

Reporting

To ensure transparent, explicit, and comprehensive reporting of qualitative research, we report this focus group study according to the relevant guidelines in the Standards for Reporting Qualitative Research [64] and in the Consolidated Criteria for Reporting Qualitative Research [65].

Participants

We made use of 2 channels for recruitment. Most participants were recruited through a flyer on social media. The inclusion criteria were (1) ≥18 years, (2) Dutch speaking, (3) in possession of an Android smartphone, and (4) having experienced a depressive episode in the past, from which they had recovered at the time of this study. Prior experiences with depression were self-reported in this part of the sample, and no further clinical screening was carried out by the researchers. Other participants were recruited via email through a previous study on web-based cognitive control training for depression relapse prevention [66]. These participants had already been screened for depressive symptoms in the past. Participants were recruited until data saturation was reached, that is, until no new information was heard during focus groups. In total, 17 participants were recruited to host 3 focus groups.

Materials

The app used in this study is currently in development by a consortium of research partners, as part of the DEDICAT ("Depression's Digital Forecasting Tool") project. At the time of the focus groups, the app was in the prototyping phase and was thus not yet publicly available. The app captures both active (EMA) and passive (smartphone and smartwatch data) data sources. While the active data is self-reported and subjective, the passive data is automatic and objective. They both complement each other, with active EMA data acting as labels for the passive data. In the future, these data sources will be fed into a personalized supervised machine learning model aimed at early detection of depression relapse.

Study Design

Two days before the focus group, participants were requested to download a version of our app that only triggered the EMA questionnaires (for screenshots, see Multimedia Appendix 1). Fifteen EMA questions were selected from an EMA item repository [67] and reflect psychological constructs that are relevant to depression relapse. Examples of EMA questions include "right now, I'm feeling excited" or "since the last measurement, I have been worrying" (Multimedia Appendix 2). Participants were prompted by a notification 6 times throughout the day at random times. This sampling frequency is common in mental health research [68-75], given that mood states can change multiple times a day. Furthermore, frequent measurements are necessary to compensate for missing data, to obtain sufficient labeling for passive data, and to allow for meaningful statistical analyses. If the questionnaire was not completed within 15 minutes, a reminder notification was sent once. After another 15 minutes, the questionnaire was no longer available. The EMA data was not stored and was deleted

XSL•FO
RenderX

immediately. No passive data was tracked from participants. The purpose of this preparation was merely to provide participants with an accurate experience with an active monitoring app.

Focus groups were chosen over one-to-one interviews as they facilitate dynamic interactions among participants, resulting in a broader range of ideas and insights [76-80]. Moreover, focus groups can uncover requirements and concerns that might not surface in individual interviews [76,81-83]. These advantages also hold when the topics discussed are personal and sensitive [84]. Especially in the context of the design of an app that is not yet available to the public or technology that participants might never have used before, focus group discussions can be an effective qualitative data collection method to encourage creative and abstract thinking.

Three face-to-face focus group discussions took place in the research facility in Ghent, Belgium, from February to March 2024. They were conducted by TC (MA, female, PhD candidate) as moderator, with the assistance of MM (MA, male, PhD candidate and clinical psychologist). Other than having made practical arrangements for the focus groups, there was no further contact between researchers and participants before this study. None of the participants knew TC and MM. The 3 groups counted 2, 7, and 8 participants, respectively. The number of participants in the first group was limited due to an unexpectedly high number of no-shows. The size of the group, however, did not affect the quality of the data. Reasons for cancellations across all focus groups included forgetfulness, changes in work schedules, and overrun meetings. One person canceled as they felt uncomfortable participating because they had been in a guest lecture by MM.

The focus group was audio-recorded and lasted for approximately one and a half hours. It started with an explanation of this study's rationale and the purpose of the app. By means of a semistructured interview guide, feedback and insights were then gathered by probing into participants' expectations, needs, requirements, concerns, and attitudes toward a depression monitoring app. The interview guide was refined after a pilot focus group hosted by TC and MM with colleagues. Example questions included "What functionalities would you like to use?" "What kind of information or feedback would you like to see in this app?" and "How would you improve this app so you would be intended to use it?"

Ethical Considerations

This study design and informed consent form were approved by the Ghent University Department of Political and Social Sciences ethics committee (2024-05) before this study commenced. Participants read and signed an informed consent form at the beginning of the focus group. They were informed through this form that their privacy would be protected by pseudonymizing the transcripts. Afterward, they received a financial compensation of €25 (US $27.17) via bank transfer.

Data Analysis

The focus group discussions were transcribed verbatim and pseudonymized, after which a thematic analysis was carried out by TC according to the step-by-step guide by Braun and Clarke [85]. Transcripts were coded in NVivo (Lumivero) using an inductive approach. In this approach, the researcher looks at the data from a data-driven or bottom-up perspective and identifies codes that are strongly linked to the data themselves. Codes were aggregated, split up, or discarded, resulting in a code tree that eventually translated into themes and subthemes. They were identified at the semantic level by iteratively deriving and refining the codes against the data. Semantic themes represent the more explicit or surface meanings of the data, rather than the researcher's interpretation beyond what the participant has said.

The criteria of what counts as a theme were a combination of the importance of the theme in terms of answering the research question and the prevalence of the theme, that is, the number of focus groups it appeared in and the number of participants that brought it up. In the latter sense, themes are repeated patterns of meaning [85]. The quality of a theme also depends on its ability to overarch contradictions between participants' preferences (see, for example, the theme of customization). The resulting themes and subthemes were discussed with MM, who comoderated the focus groups, and all other authors. Finally, the active role of the researchers' judgment, their assumptions, and theoretical backgrounds in determining themes should be acknowledged.

Results

Participants' Characteristics

The sample of 17 participants consisted of 3 men and 14 women, aged 21 to 49 (mean 27.7, SD 6.8) years. All participants were White.

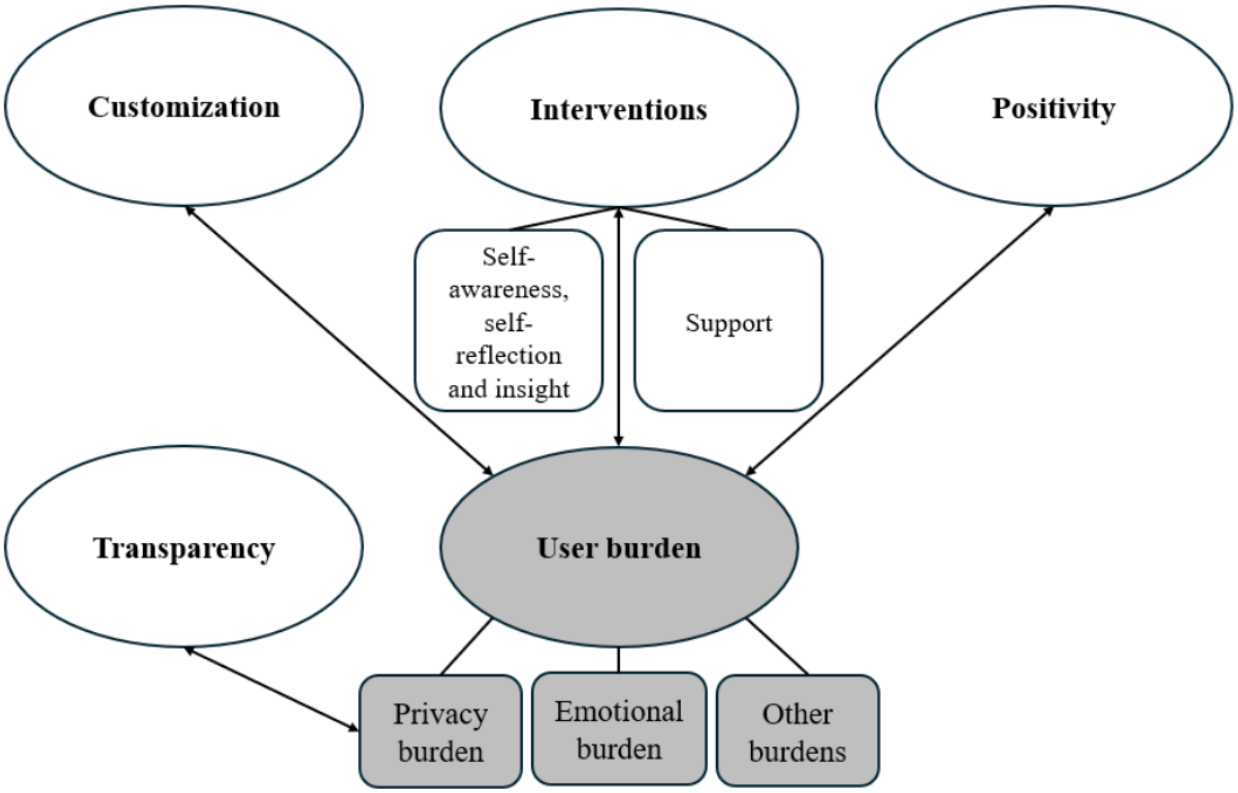
Themes

Overview

The 5 main themes resulting from the thematic analysis and the relationships between them are visualized in a thematic map (Figure 1). The ovals depict the main themes, while the rounded rectangles represent the subthemes. The themes or subthemes with a white background represent what participants require in an mHealth app, whereas the themes or subthemes with a grey background depict their concerns. Many of the topics discussed within themes are interconnected with other themes. The double-headed arrows visualize which requirements balance out which concerns. Each theme is outlined below and is illustrated with a few quotes from participants (Multimedia Appendix 3).

XSL•FO
RenderX

Figure 1. Thematic map of the 5 main themes: customization, interventions, positivity, transparency, and user burden.



Customization

The customization theme consistently appeared across all focus groups and was echoed by numerous participants. The need for customization was apparent in multiple aspects of design: (1) data and privacy issues, (2) the EMA questionnaires, and (3) the feedback functionalities.

Participants wanted to customize which data parameters are tracked by the app. Such a customization feature would allow users to disable certain parameters they deem sensitive. Furthermore, they wished to configure whether the data would be shared with their mental health care providers or not, and if so, which mental health care provider would see which data parameters at what time. Participants preferred to decide for themselves whether and when to contact their mental health care provider in case of early detection of a relapse in depression.

Another topic of customization was related to the EMA questionnaires and the possibilities for feedback (eg, on tracked mood, sleep, or physical activity). Participants wanted to configure the frequency of the questionnaires. Next to the frequency, some liked to configure the timings of the questionnaires as well, whereas others embraced the random character in which the notifications appeared. These configurations would allow users to embed the usage of the app into their daily life—for example, shifts and breaks at work—and would allow them to decide how often and when they are confronted with questions asking about their emotional state. Concerning the content of those questions, participants expressed the need to provide context by linking mood and daily life activities. This functionality would allow for more personalized feedback, which in turn would provide the user with more valuable insights. However, a considerable number of participants preferred not to see any feedback at all to prevent adverse emotional effects—demonstrating the need for customization of feedback functionalities.

Positivity

Participants made it clear that they wanted positivity to be included in an mHealth app and that design efforts should be directed toward establishing (1) a positive user experience, and (2) positive reinforcement and affirmation. An interesting observation here is that participants alluded to the importance of the app being more than just a relapse monitor; they saw opportunities for prevention. Participants mentioned, for example, that the app should also include (3) journaling functionalities.

With respect to a positive user experience, participants sought the ability to customize what the app looks like. The idea of choosing between a few options of themes was put forward in every focus group. Themes consist of a combination of colors, visuals, fonts, and shapes that are consistently applied throughout the app. Participants expressed explicit interest in themes with more vivid color palettes, next to the default light and dark themes, as colors were said to represent positivity. What participants meant by a positive user experience is that the app does not look clinical, dull, or depressing.

To prevent this, some participants proposed implementing positive reinforcement via gamification into the app. Gamification refers to enhancing an app with game-like elements to keep users engaged and motivated. Whereas some participants enthusiastically suggested creative ideas for

XSL•FO
RenderX

gamification features, other participants thought of gamification as something inappropriate or immature that would make the app less suitable for them. Participants did, however, agree on the omission of 1 particular gamification feature: the streak. A streak is a consecutive series of actions performed without interruption. They are used to incentivize users to maintain consistent usage. Multiple anecdotes were given on how the loss of a streak in other apps caused demotivation, disengagement, or even triggered them to stop using the app altogether.

What appealed to participants in terms of gamification was the potential for positive reinforcement through game rewards, without the risk of punishment. As 1 participant (female, 31 years, ID22) put it, "we are already far too hard on ourselves," implying that the app should reflect mildness and forgivingness. A self-monitoring app that is flexible and forgiving after noncompliance would cultivate self-compassion [86]. An example of a reward is a motivational, positive, or appreciative message after having completed an EMA questionnaire. Finishing tasks would be more rewarding when receiving such messages. Participants proposed to have the principles of positive reinforcement and affirmation implemented in the feedback functionality as well. More importantly, they wanted such encouraging messages regardless of whether the feedback indicates that the user's mental health remains stable, improves, or deteriorates.

Other functionalities proposed by participants were related to journaling, such as a gratitude journal and the possibility to write down positive events of the day or week. What appealed to them was the ability and encouragement to reflect on positivity. This would make for a more positive user experience. In turn, a weekly feedback overview could bring back gratitude and positive events to the user and could therefore serve as a reminder of the positivity during the week. The features mentioned above were desired by participants because of their potential to facilitate a more positive outlook on life.

Interventions

In every focus group, participants came up with ideas for functionalities that we categorized as possible interventions that an mHealth app could offer. The content of this theme is twofold. In return for personal data and consistent efforts to comply with EMA questionnaires, participants would want both (1) self-awareness, self-reflection, and insight, and (2) support during times when their mental health deteriorates.

Self-Awareness, Self-Reflection, and Insight

Participants were interested in gaining self-awareness, self-reflection, and insights from their tracked data. They preferred feedback to be displayed in the form of weekly overviews of the monitored mental health parameters. It would be even more interesting to them if they were able to link their monitored mood with activities in the EMA questionnaires. This idea of adding context to mood experiences was echoed in every focus group. Participants believed that bringing back this relationship to the user in the feedback overviews could lead to the detection of patterns and help them get to know themselves. They expected the feedback to provide insight into their triggers, coping mechanisms, self-care activities, and the activities that bring them energy. Even more so, they believed it could lead to action points or motivation for behavior change, upon learning what they felt when they were doing certain activities or when they were around certain people. When asked what appeals most to participants or what would be a motivator for adoption and long-term use of the app, the main answer boiled down to the valuable insights that the app could offer. Participants stated that for weekly feedback overviews to promote self-awareness, self-reflection, and insight, the language used should not be stigmatizing. Furthermore, they required the language to be subordinate to the user's feelings and to be carefully formulated to avoid triggering self-fulfilling prophecies. Finally, they wanted the app to be honest about the accuracy of its results and predictions.

Support

In addition to gaining self-awareness, self-reflection, and insight, participants expressed a need for the app to provide support during difficult times. Participants proposed a few functionalities to support them in case the machine learning model detects a decline in mental health. As the app would collect a myriad of data, participants would prefer to see this with an explanation and some examples. In their view, realizing and understanding the situation are the first steps in coping with it. Participants argued that a difficult time is often accompanied by a drop in self-reliance. Self-reliance refers to trusting and relying on one's own capabilities, judgments, and efforts without excessive reliance on others. As self-reliance is less evident during more vulnerable periods, they saw opportunities for personalized support beyond merely follow-up. An example they gave is that the app could display personalized tips. By providing the user with their own coping mechanisms or self-care activities, the app would primarily focus on self-reliance by supporting the users' resilience. According to participants, this advice would be empowering and would be more accessible than a visit to their mental health care provider. However, in case the problems persist, participants would like the app to advise them to contact their mental health care provider. Should they, however, feel an immediate need for a conversation in anticipation of a consultation with a professional, a help page in the app could provide an accessible way for them to contact a mental health care service. The participants argued that during visits with their mental health care provider, the feedback overviews could also act as conversation material. A few participants mentioned that the link between mood and activities is especially relevant here because it would help to overcome potential recall biases. For instance, they might struggle to accurately recall their activities from the previous week when they reported low mood in the EMA questionnaire. Related to this, as 1 participant pointed out, the EMA questionnaires could replace journals that mental health professionals ask them to keep.

Participants argued that they can recognize a relapse in themselves. However, some saw the potential of an mHealth app recognizing it in a timelier manner [57]. This early warning sign could lead them to consult their mental health care provider sooner than they would without the support of an app. Another way the interventions could prove useful in their lives—as indicated by 1 participant—is by bridging the waiting time until

XSL•FO
RenderX

the visit with their mental health care provider. In the meantime, there would be something to hold on to.

Transparency

Participants requested transparency regarding (1) the data monitoring and (2) the explainability of the machine learning model. They required information on what data parameters would be tracked, where they would be stored, and who would be able to access the data. A better understanding was needed of why certain data parameters would be monitored and how they link to mental health. The more sensitive the data parameter, the greater their need to understand this link. If the benefit were to be explained to them, participants said they would have a less reluctant attitude. Participants also demanded transparency regarding the timeframe during which data would be monitored before they could expect any meaningful feedback.

Participants claimed that machine learning in general is nontransparent and therefore requested more explanation on how the machine learning model would work. This was said to help them in deciding whether to trust or distrust the algorithm [87]. As opposed to a lengthy and tedious mandatory read at installation of the app, participants would prefer to have explanations displayed little by little, where relevant and upon request. Progressive disclosure—where additional explanation is provided on demand—could therefore meet participants' need for transparency in intelligent systems [88]. In case the system would detect a decline in mental health, participants would prefer to know why, in the form of an explanation, preferably with some examples unveiling which parameters are most crucial in their case. When it comes to the language used in the explanations, they would prefer the accuracy of the predictions to be communicated honestly and transparently. This way, the participants would have access to sufficient information to decide for themselves whether to trust the system and whether they can rely on the technology.

User Burden

The 4 main themes mentioned above can be seen as antidotes to this theme. The user burden theme contains three subthemes: (1) privacy burden, (2) emotional burden, and (3) other burdens. We classified the user burdens into subthemes according to an existing user burden scale [89].

Privacy Burden

A repeated observation across the focus groups was that participants raised privacy concerns. The privacy burden is defined as a system's risk of revealing information about a user that they would prefer not to share [89]. As discussed in the customization theme, participants would prefer to configure which parameters would be tracked by the mHealth app. They also expected an option to put in personal preferences regarding whether to allow sharing data with mental health care providers. Among the most sensitive data to share, according to participants, would be location. They indicated resistance to allowing an app to track their phone usage per app category (eg, social media, gaming, communication, etc) as well, but to a lesser extent than their location. Participants considered location very personal, as it reveals a lot about one's daily life. Furthermore, not everyone was convinced of the link between passive data (ie, smartphone and smartwatch data) and mental health monitoring. Participants viewed passive data as more burdensome in terms of privacy to have tracked compared to active data (ie, EMA questionnaires on mood), which they deemed more relevant for monitoring mental health. The initiative of a mental health care provider would, however, increase trust and credibility in the system, enhancing participants' intention to use it [90].

Participants would expect benefits in return for tracking and sharing data to make the privacy burden fairer. In other words, they would seek a balance between the data they would "give" and what would be in it for them. Examples of benefits expected by participants were positive affirmations, customization settings, a highly personalized detection model, self-awareness, self-reflection, and insight based on personalized feedback, personalized intervention by the app in case of difficult times, and personalized help from a mental health care provider based on the tracked data.

Emotional Burden

Another subtheme of user burden is the emotional burden, which is defined as a system's risk of making the user feel bad or unnecessarily worried [89]. Participants raised concerns about the potential detrimental side effects of (1) the EMA questionnaires and (2) the feedback overviews. Based on the experience that participants had gained with EMA before the focus group, they talked about the psychological influences it had on them. Even though the questionnaire was balanced in terms of measuring positive (eg, "right now, I'm feeling satisfied") and negative affect (eg, "right now, I'm feeling down"), a lot of participants stated that the introspection for filling out the questionnaire made them feel worse than before. They personally focused primarily on the negative emotions that were asked, even after the interviewers reminded them of the fact that positive emotions were gauged as well. Indeed, individuals remitted from a depressive disorder have a stronger attentional bias toward negative adjectives (eg, worthless), compared to a never depressed comparison group [91].

Participants in the focus groups warned about the downside effect of the confrontation with these questions multiple times a day on their mental well-being. The EMA questions forced them to think about how they were feeling, while they might not otherwise have thought about their mood. This had caused some participants to start ruminating about why they were feeling a certain way. Finally, some participants indicated that the EMA questionnaires fed preoccupation with having (had) depression. Filling out the questions acted as an unwanted reminder of a challenging time in their life. They compared it to visiting their psychologist, and how they wound down these visits or stopped them altogether when their mental health started to improve again. Similarly, those participants stated that they would install and use the app in times when they feel depressed. Consequently, they would delete the app again whenever they felt better.

Participants also worried about the emotional burden of feedback that the app could display to the user. Participants were inclined to envision feedback in the more negative scenarios, for example, displaying a downward trend in mood over the last

XSL•FO
RenderX

week. Confrontation with this kind of feedback could also trigger rumination and would cause participants to be worried about a downward spiral due to a self-fulfilling prophecy. The latter would occur when the user would internalize this feedback. A participant explained that feedback stating a deterioration in sleep quality could cause trouble sleeping, which might otherwise not have occurred without seeing this feedback. To prevent any potential self-fulfilling prophecies, some participants indicated that they would prefer to have the option not to see any feedback at all. Seeing their own feedback would trigger anxiety, especially if they are alone at that moment. They proposed that only their mental health care provider would be able to see the feedback and would interpret and explain it to them during a consultation. However, not everyone shared this opinion, as others would worry that their mental health care provider would confront them with their data and feedback. This discrepancy in participants' requirements once again proves the need for customization in the design of mHealth apps.

Other Burdens

The last subtheme of user burden is "other burdens." It is an umbrella term for all other practical difficulties that participants brought forward. The following types of burdens were mentioned in the focus groups: (1) time burden, (2) social burden, (3) mental burden, and (4) access burden [89]. The time burden refers to the frequent use of the system due to the EMA data collection. Many participants indicated that filling out a questionnaire daily or multiple times a day is too burdensome. It was often not compatible with their work, education, or daily life in general [57]. The social burden applies to the difficulty or inappropriateness of responding to the EMA notifications in social settings. Those notifications also distracted them from what they are doing, resulting in a mental burden. Lastly, the access burden refers to the incompatibility of filling out EMA questionnaires and wearing a smartwatch in certain jobs. For example, nurses are not allowed to wear a smartwatch [57], or some company policies prohibit employees from carrying a smartphone with them during their shift. Additionally, participants working in jobs that involve a fixed timetable (eg, a teacher) or jobs that involve consultations with people (eg, a psychologist) were not able to fill out a questionnaire at any time of the day. Many of these issues could be mitigated by customization settings that would allow the user to choose the frequency and timings of notifications.

Discussion

Principal Findings

This paper presents a focus group study with individuals remitted from depression on requirements and concerns related to the design of an mHealth app for early detection of depression relapse. Participants provided feedback on an mHealth app that would capture both active EMA data and passive sensing data from the user's smartphone and smartwatch. The aim of this app in development is to feed all data into a personalized supervised machine learning model, being trained for just-in-time detection of depression relapse. In general, 5 main themes were identified by means of a thematic analysis: customization, positivity, interventions, transparency, and user burden. Whereas the first 4 themes reveal participants' requirements of the mHealth app's design and functionalities, the fifth and last theme covers their concerns regarding the use of such an app.

Our findings contribute to the field of mHealth and depression management in several ways. First, our unique sample consisted of individuals remitted from depression, a group at high risk of relapse but often overlooked in research and app development. Second, feedback was gathered toward a preventative tool integrating both active and passive data. Technical feasibility and scientific validation are essential, though insufficient for successful implementation in clinical practice. Considering the poor engagement with this technology thus far, mHealth apps have to be aligned with user needs as well to ensure acceptance, adoption, and usability. Therefore, our findings offer in-depth insight into the requirements and concerns of potential end users, which are discussed below.

Requirements

First, participants emphasized the need for customization in several areas of design: data and privacy settings, the EMA questionnaires, and the feedback functionalities. Configuring what data parameters would be tracked, who they would be shared with, and when would allow participants to maintain control over their data. This would mitigate the privacy burden for them. The assumption of ownership over one's own data and the choice of sharing it with clinicians is in line with previous studies [70,92,93]. Participants expected functionalities to customize the timing and frequency of the EMA notifications to alleviate the time, social, mental, and access burden by embedding the usage of the app seamlessly into their daily routines. This confirms the findings of many previous studies stressing users' need for customization of the timing [92] and frequency of self-assessments [67,70,71,86,94]. The same functionalities would also address the emotional burden by allowing participants to decide how often and when they would be confronted with questions asking about their emotional state that could trigger rumination. An important remark here is that from a psychometric point of view, customization concerning timescales can lead to significantly different conclusions [95-98]. Furthermore, from the perspective of training a supervised machine learning model, sufficient EMA data (ie, a labeled dataset) is needed to generate accurate predictions. Another example given by participants of how customization could address the emotional burden was a feature to opt out of seeing one's own feedback overviews. This was assumed to prevent adverse emotional side effects. Allowing users to select or deselect specific visualizations in a dynamic feedback interface might provide an even better solution [71]. The importance of customization settings cannot be overstated, as insufficient personalization is a perceived barrier to using mHealth apps for depression [99] and one of the main reasons for disliking the use of apps for mental health in general [90].

The second preference of participants is for the app to display and encourage positivity. As depression is associated with biased attention to negative information [100-102], participants might require positivity in an app to compensate for this information processing bias. According to them, a positive user experience

XSL•FO
RenderX

can be established through colorful visuals, thereby avoiding clinical or monotonous aesthetics [71,103]. Additionally, participants suggested directing design efforts toward positive reinforcement and affirmation in the form of rewards. This can be a motivational, positive, or appreciative message after having completed an EMA questionnaire. Previous studies suggested rewarding users for their efforts with advice and compliments [92] or encouragement messages [86]. Other functionalities proposed by participants—such as gratitude journaling and the possibility to write down positive events of the day or week—are inspired by positive psychology. This field of psychology is about positive subjective experiences: well-being, contentment, and satisfaction in the past; flow and happiness in the present; and hope and optimism for the future [104]. Positive psychology interventions such as writing down good things in life or cultivating gratitude have the ability to—at least temporarily—decrease depressive symptoms [105]. Researchers provided ample empirical support in favor of positive psychology practices alleviating depressive symptoms and enhancing well-being [106-109]. Furthermore, positive psychology strategies can act as a prevention mechanism by helping to build resilience and coping capabilities that reduce relapses in the treatment of depression [110].

The third finding is that participants requested interventions to be the focal point in the app's design. This is interesting because participants were in remission, and the researchers clearly stated that the app would aim at mere monitoring and follow-up. The importance of interventions cannot be overlooked, as it was said to be the main motivator for long-term use of the app. In exchange for accepting a certain degree of user burden, participants would expect benefits in addition to early relapse detection. Those benefits can come in the form of self-awareness, self-reflection, and insight, or support during difficult times. Linking mood with context was deemed crucial for obtaining meaningful feedback. Contextual data can be gathered through a smartwatch [111] or through EMA diaries. It is important for the latter to include (social) contextual factors next to symptoms [67,70,92,112]. Combining mood with context is key to reflecting on emotional patterns and detecting triggers to mood changes [72,113]. Participants of previous studies agreed that the act of self-monitoring in itself and the feedback based on this data can result in self-awareness, self-reflection, insight, and behavior change [57,70-73,75,86,87,92,112,113]. Increased self-awareness is expected to help in self-assessing whether they need to reach out for help from others or not [86].

During periods of mental health decline, participants would want the app to provide them with support. In line with previous studies, they would prefer the app to display personalized tips and coping strategies [86,94]. In another study, patients put forward the suggestion to have the system alert their clinician in case of a significant change in symptom levels [92]. They liked the idea of the alerts triggering personalized therapeutic advice. However, the study also highlighted clinicians' points of view, stating that they were hesitant about the possibility of receiving alerts. Alerts would make them worry about patient safety. The clinicians were also concerned about an increased burden of responsibility, given their time constraints. A more realistic way in which the app can offer support in difficult times is by referring the user to professional help. Participants also saw value in having the feedback overviews serve as a starting point for dialogue with mental health care providers. Another way in which feedback can help, according to them, is in overcoming the recall bias, especially if significant time has passed since a prior appointment [71,86]. EMA-derived feedback can provide a more reliable overview of how patients have felt since the previous session compared to directly asking the patient. This is because the latter is often influenced by current mood [92]. Overall, feedback overviews were expected to improve communication with clinicians [57,70] and make time spent with them more efficient [72].

The fourth and last requirement from participants was transparency concerning the data monitoring and the explainability of the machine learning model. Participants viewed transparency as an antidote to the privacy burden. However, the literature is in disagreement regarding this relationship. For example, information use transparency features—as requested by participants in the focus groups—inform users about how and for what purposes the acquired information is used [114]. While many authors have speculated that said transparency features alleviate privacy concerns, this assumption has not been empirically supported [115,116]. Therefore, a dual effect is assumed. Whereas transparency as a signal of trustworthiness and fairness [117] can help to drive information disclosure [118], it can also make privacy hazards explicit and users reluctant to share sensitive information [119].

Transparency regarding the machine learning model can be addressed with explainable machine learning. By providing a transparent explanation for its results, reasoning, and predictions, this form of machine learning is made understandable for humans. Participants' requirement for transparency is aligned with the literature. A qualitative study found that patients with bipolar disorder would prefer to see simple explanations of the data analysis behind the experience sampling feedback [70]. In another interesting study that measured stress by combining smartwatch and EMA data into a predictive algorithm, participants expressed curiosity to understand how stress can be determined [87]. More specifically, they were interested in the relationship between certain data parameters and stress. Especially those participants who articulated distrust in the stress algorithm felt the need to identify the quality and reliability of the algorithm. In contrast to the dearth of studies that did not find an empirical relationship between transparency and privacy, participants in this study viewed transparency as essential for building trust and thereby reducing the expected privacy burden.

Concerns

Participants required customization, positivity, interventions, and transparency to bear the anticipated user burdens of mental health monitoring. The most prevalent user burdens highlighted by participants in this study were the privacy and emotional burden. Privacy concerns were related to the tracking and sharing of sensitive data. Other studies confirm that privacy concerns are among the biggest barriers to the use of apps for depression [57,99], apps for mental illness [90], and health apps

XSL•FO
RenderX

in general [120]. Given that the data would be highly personal, participants require it to be safely stored to protect their privacy [67]. Among the most personal data to track and share according to participants, would be location, which is known to be sensitive to share [57,93,121]. By extension, passive data tracking instilled more privacy concerns compared to active data. This finding fits seamlessly with earlier work on willingness to participate and actual participation in passive mobile data collection [122,123]. Even though active data collection requires more effort than passive tracking, users were more willing to agree to it, as it rendered them a greater sense of control over the content [124]. By consequence, the possibility for direct control over the passive data collection process—such as an option to temporarily switch off tracking—can lighten the privacy burden [125]. What might reduce the privacy burden even more would be a functionality to specifically pick which data parameters to share and which not to ahead of time (ie, à la carte consent). This functionality is exactly what was proposed by participants in this study.

Interestingly, however, privacy issues do not necessarily translate into poor adoption, as the literature describes a discrepancy between user attitudes and actual behavior. This is called the privacy paradox [126]. Despite reporting high privacy concerns, web users disclose private information, nonetheless. Among the possible explanations of the paradox are consumers lacking sufficient information to make informed privacy decisions, and in the case of sufficient information, they are likely to trade off privacy risks for possibly gained benefits [127-129]. The latter is also known as the privacy calculus theory [130,131]. The theory holds in different contexts, including the context of mobile apps [132] and the Internet of Things environment [133]. As perceived benefits, users often expect a better personalized service [134,135]. This is where the personalization-privacy paradox comes into play, stating that users pursue personalized services, yet the collection of personal information to provide such services raises privacy concerns [114,136]. The personalization-privacy paradox is highly relevant in the mHealth context, where trust largely mediates the effects of perceived personalization and privacy concerns on adoption intention [137]. Participants in our focus groups indeed expected benefits in exchange for sharing personal information. Personalized services were on top of the list, such as a highly personalized detection model, personalized feedback providing self-awareness, self-reflection, and insight, personalized intervention during difficult times, and personalized help from a mental health care provider based on the tracked data.

In addition to privacy concerns, participants also feared the emotional burden of self-tracking. While frequent self-assessment and feedback are promising in terms of self-awareness, insights, and motivation for positive lifestyle choices, many authors also warned against emotionally adverse side effects. EMA assessments can be regarded as a continuous reminder of symptoms rather than what goes well, possibly resulting in preoccupation with mental illness [72] or symptom worsening [92]. Frequent confrontation with mood could impair the user's adaptive coping strategies, such as seeking distraction [70]. Interestingly, emotion tracking seems more burdensome for people prone to depressive symptoms [74,138], especially because this population is known to have a greater focus on negative emotions [91]. It is, however, worth noting that participants underwent intense EMA assessments (ie, 6 times per day). While EMA assessments in themselves are predictive of depression relapse [69,139-141], they are also needed to obtain labeling for supervised machine learning. Frequent EMA assessments are therefore inevitable in the beginning stages of development of the depression relapse monitoring tool. Considering that the prediction model will be personalized, frequent EMA assessments will also be needed in the beginning for every new user to establish a baseline and labeling. For long-term usage, however, we expect a greater contribution of passive monitoring.

Participants were also apprehensive about the idea of seeing their own data and the psychological consequences of receiving said feedback. For example, some participants worried about negative feedback triggering self-fulfilling prophecies. The accuracy of feedback is also important, as suggestive or misleading feedback can lead to self-reported symptoms worsening [142]. Similar to the findings of other studies, some participants in this study therefore proposed to reduce the emotional burden by having the EMA-derived feedback explained and interpreted by their mental health professional in person [70-73,92]. The main reason for preferring not to look at their own data would be to prevent anxiety [57] and rumination [87]. Rumination is characterized by repetitive negative thoughts and is related to a deterioration in mood over time [143]. Rumination is both a symptom of depression [144] and a risk factor for the onset, maintenance, and recurrence of depressive episodes [6,145]. Given the vulnerability of potential users, mental well-being–related assessments and feedback should therefore be well considered and carefully designed to avoid harming or at least to minimize mood reactivity [87]. The emotional burden would make participants refrain from self-reporting during periods in remission, when they do not want to be occupied with their mental (ill) health [70]. This runs counter to the idea of a relapse detection app, which can only work when users start monitoring in times of remission. Contrary to the scientific evidence that follow-up is valuable in relapse prevention [146-148], many participants stated that they would only consider self-monitoring in the presence of more severe depressive symptoms.

Recommendations

Our focus groups have identified several themes that can guide technological development. The requirements direct the attention of mHealth app developers toward the features that participants most value, whereas their concerns expose pain points for adoption and engagement. The first recommendation is to build in a wide range of customization settings for users to have a sense of agency, ownership, and autonomy regarding their mental health monitoring. Whereas the possibility to configure data tracking and sharing has the potential to counter the privacy burden, the emotional burden is tackled through customization settings regarding the frequency and timing of EMA assessments and control over how feedback is displayed. Developers can also address the latter burden by further refining passive tracking, instead of the emotionally straining active data

XSL•FO
RenderX

collection. Based on the second theme, we also recommend that app developers create a positive user experience. This can be done via relatively straightforward design efforts toward colorful visuals and gratitude journaling features. We strongly discourage integrating punitive gamification features such as streaks. Instead, developers can incorporate positively reinforcing and affirmative messages to obtain appropriate user behavior. Third, it is also recommended to add functionalities that users can gain insight and support from. That includes weekly feedback overviews and presenting the user with personalized tips conditional on their mental state. The fourth recommendation for mHealth developers is to disclose transparent information about all facets of the data processing. Progressive disclosure, for example, fulfills users' need for transparency without overwhelming them. Transparency was said to mitigate the privacy burden, for instance, by clarifying the relevance of passive data collection. Developers should also address privacy concerns by prioritizing data protection and security.

Our results also led to some recommendations for clinicians. First, as clinicians can play an important role in the follow-up of people in remission, we recommend that they continue participating in research concerning the development of this technology. Their participation could pave the way for clinical integration. They could educate patients on how the app works and how predictions are made, particularly to make them understand the relevance of passive data collection. Clinicians' explanations and recommendations would increase patients' trust and credibility in the system, enhancing their intention to use it. During early use or emotionally vulnerable periods, clinicians can assist patients in interpreting feedback from the app to mitigate the risk of misinterpretation, rumination, anxiety, or self-fulfilling prophecies. Clinicians are also encouraged to explore this technology's opportunities and risks. A potential opportunity identified in this study is that EMA can act as conversation material to improve recall during talk therapy. A risk, as mentioned during the focus groups, could be that patients feel surveilled by their mental health provider. Based upon these real-world experiences, the tool could be iteratively improved.

Limitations and Future Work

A limitation of this study is that the requirements and concerns of our sample may not be generalizable to the entire population. This is because our convenience sample lacked some diversity in terms of demographics. More specifically, most participants were relatively young, female, and White. Under-recruitment of ethnic minorities is a well-known issue in mental health research [149]. A predominantly female sample is the result of a prevalent self-selection bias in this kind of research [57,69-72,86,87,99]. This is not necessarily problematic, as females experience a higher risk of onset of depression, longer duration of episodes, and a tendency for higher risk of recurrence [150]. Women also find it easier to confide in others about psychological problems and, therefore, are more likely to seek help by attending health care services and receive diagnoses for mental health problems [151]. Females may thus form a target population likely to adopt an app that monitors depression relapse. The sample also consisted mainly of people in their twenties, which can be explained by the first recruitment method through a social media channel linked to the university. This age group is highly relevant for our research objectives because the first onset of depression is usually in adolescence, with a substantial risk of recurrence afterward [152,153]. Additionally, younger individuals are among the most prevalent users of mobile technologies and are more likely to use mHealth apps [120]. Younger individuals and females may therefore generate a wide range of ideas toward a future mHealth app for depression detection. Granted that our results are not validated in a larger and more diverse sample, we elaborated in depth on the requirements and concerns of people who have experienced depression. Despite the relevant profiles included in our convenience sample, the generalizability of the findings to more diverse populations should be tested in future work.

Another limitation is that for a large part of the sample, experiences with depression were self-reported. Research, however, indicates high accuracy in self-diagnosis for commonly occurring mental disorders, such as depression [154]. Additionally, we acknowledge that examining the requirements and concerns of only 1 group of end users (ie, individuals remitted from depression) is a limitation of this study. Future research should consider other stakeholders as well [155]. In this regard, the requirements and concerns of clinicians are of utmost importance in implementing this technology into routine practice [58]. Nonetheless, interviewing clinicians was beyond the scope of this study.

A further limitation relates to the dynamics within focus groups. While focus groups provide researchers with many advantages, there are certain drawbacks to them as well. These include the potential for dominant voices to influence the discussion [156] and the possibility for groupthink [157]. Groupthink occurs when people strive for consensus within a group, thereby setting aside personal opinions or agreeing with those of the rest of the group. The moderators aimed at preventing groupthink by providing different perspectives to the group, asking new questions, or framing questions differently [158]. The moderators also prevented outspoken participants from dominating the discussion by encouraging silent individuals to speak within the group by directly asking them for their opinion.

Finally, participants provided feedback based on a concept for an app, rather than an app they had actual experience with. Participants did interact with a prototype of the app. However, this only covered EMA functionality thus far. The remainder of their feedback was based on a hypothetical scenario for app usage (ie, active and passive mood monitoring and early detection of a potential depression relapse). In this sense, our prototype acted as a "tabula rasa" for participants to generate unbiased and diverse ideas for. Nonetheless, we also acknowledge that our findings are limited to their imagination, creativity, expectations, and perceptions. A self-evident area for future research is to evaluate the acceptability and usability of a prototype containing the depression relapse detection model that takes into account the requirements and concerns resulting from this study. Moving forward, it will also be necessary to assess whether the implementation of our findings is associated with a positive impact on adoption and engagement in a real-world mHealth context. It could, for example, be possible that customization settings induce a cognitive or mental burden, or that journaling features heighten the privacy burden. Further

XSL•FO
RenderX

user testing can reveal whether expected benefits or risks will materialize when participants experience this new tool in real life. Additionally, A/B testing can empirically test the assumptions that came forward during the focus groups. For example, it can contribute to the debate in the literature on whether transparency features alleviate the privacy burden or not.

Conclusions

This study underscores the importance of designing mHealth apps for depression relapse detection that carefully balance requirements and concerns [57]. More specifically, their design should include customization, positivity, interventions, and transparency while minimizing the privacy and emotional user burden. According to participants, active data collection was most burdensome emotionally, while passive tracking would mainly contribute to the privacy burden. Whereas externalizing tracking by moving from user-driven to more technology-driven, automated data collection has the potential to reduce the emotional burden, it would likely aggravate the privacy burden. Passive mood monitoring will only become clinically useful if patients are open to learning about its potential, tolerate it, and perceive it to be helpful. Participants required certain functionalities to be willing to bear both burdens of self-monitoring. Only then can mHealth apps start to be widely accepted, adopted, and used in clinical practice. Considering the vulnerability of the potential end users, caution is warranted in the design of mHealth apps. Therefore, collaborative efforts should persist between researchers, mHealth app developers, mental health care providers, and individuals remitted from depression. Our findings reiterate other authors' call for designer responsibility in the mHealth landscape, especially because researchers have yet to unravel the complicated relationship between the usage of emerging technologies and mental well-being.

Acknowledgments

We are grateful to Yannick Vander Zwalmen, PhD and David Demeester, MA for their help in recruitment. We are also thankful to the participants for sharing their thoughts, ideas, and insights. This paper was established without the use of generative artificial intelligence technology. This research was supported and funded by a concerted project (Geconcerteerde onderzoeksacties) from the Special Research Fund (Bijzonder Onderzoeksfonds) of Ghent University, called DEDICAT (Depression's Digital Forecasting Tool; BOF23-GOA-006).

Data Availability

The pseudonymized focus group transcripts are not publicly available, given the personal nature of the stories that were shared. They are, however, available from the corresponding author upon reasonable request.

Authors' Contributions

TC worked on the conceptualization, methodology, and visualization of this study; formally analyzed, investigated, and administered this study; and wrote the original draft and reviewed and edited the writing. MM equally contributed to the investigation. KB, MMPVA, MM, and EHWK assisted with the conceptualization and methodology. KB supervised this study. KB, MM, SC, MDB, SVH, EHWK, and MMPVA helped with reviewing and editing the writing. EHWK, KB, MMPVA, and SVH acquired funding for this project. MDB managed the resources and software for this study. SC, SVH, and MM aided MDB with resources. SC and SVH supported MDB with the software.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the EMA functionality in the DEDICAT app. DEDICAT: Depression's Digital Forecasting Tool; EMA: ecological momentary assessment.
[DOCX File , 120 KB - mhealth_v13i1e67141_app1.docx]

Multimedia Appendix 2
Details of the EMA questionnaire.
[DOCX File , 21 KB - mhealth_v13i1e67141_app2.docx]

Multimedia Appendix 3
Quotes from participants per theme.
[DOCX File , 35 KB - mhealth_v13i1e67141_app3.docx]

References

XSL•FO
RenderX

1. Kessler RC. The costs of depression. Psychiatr Clin North Am 2012;35(1):1-14 [FREE Full text] [doi: 10.1016/j.psc.2011.11.005] [Medline: 22370487]
2. Ekman M, Granström O, Omérov S, Jacob J, Landén M. The societal cost of depression: evidence from 10,000 Swedish patients in psychiatric care. J Affect Disord 2013;150(3):790-797. [doi: 10.1016/j.jad.2013.03.003] [Medline: 23611536]
3. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
4. Moreno-Agostino D, Wu Y, Daskalopoulou C, Hasan MT, Huisman M, Prina M. Global trends in the prevalence and incidence of depression:a systematic review and meta-analysis. J Affect Disord 2021;281:235-243. [doi: 10.1016/j.jad.2020.12.035] [Medline: 33338841]
5. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
6. Buckman J, Underwood A, Clarke K, Saunders R, Hollon S, Fearon P, et al. Risk factors for relapse and recurrence of depression in adults and how they operate: a four-phase systematic review and meta-synthesis. Clin Psychol Rev 2018;64:13-38 [FREE Full text] [doi: 10.1016/j.cpr.2018.07.005] [Medline: 30075313]
7. Bockting CLH, Spinhoven P, Koeter MWJ, Wouters LF, Schene AH, Depression Evaluation Longitudinal Therapy Assessment Study Group. Prediction of recurrence in recurrent depression and the influence of consecutive episodes on vulnerability for depression: a 2-year prospective study. J Clin Psychiatry 2006;67(5):747-755. [Medline: 16841624]
8. Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman ATF. Prevalence and predictors of recurrence of major depressive disorder in the adult population. Acta Psychiatr Scand 2010;122(3):184-191. [doi: 10.1111/j.1600-0447.2009.01519.x] [Medline: 20003092]
9. de Zwart PL, Jeronimus BF, de Jonge P. Empirical evidence for definitions of episode, remission, recovery, relapse and recurrence in depression: a systematic review. Epidemiol Psychiatr Sci 2019;28(5):544-562 [FREE Full text] [doi: 10.1017/S2045796018000227] [Medline: 29769159]
10. Twynam-Perkins J, Pollock A, Brilikhova P. Treatment gap in depression. J Epidemiol Community Health 2011;65(Suppl 2):A36.2-A3A36. [doi: 10.1136/jech.2011.143586.81]
11. Doblyte S, Jiménez-Mejías E. Understanding help-seeking behavior in depression: a qualitative synthesis of patients' experiences. Qual Health Res 2017;27(1):100-113. [doi: 10.1177/1049732316681282] [Medline: 27956660]
12. Sherwood C, Salkovskis PM, Rimes KA. Help-seeking for depression: the role of beliefs, attitudes and mood. Behav Cogn Psychother 2007;35(5):541-554. [doi: 10.1017/s1352465807003815]
13. Doll CM, Michel C, Rosen M, Osman N, Schimmelmann BG, Schultze-Lutter F. Predictors of help-seeking behaviour in people with mental health problems: a 3-year prospective community study. BMC Psychiatry 2021;21(1):432 [FREE Full text] [doi: 10.1186/s12888-021-03435-4] [Medline: 34479537]
14. Boerema AM, Kleiboer A, Beekman ATF, van Zoonen K, Dijkshoorn H, Cuijpers P. Determinants of help-seeking behavior in depression: a cross-sectional study. BMC Psychiatry 2016;16:78 [FREE Full text] [doi: 10.1186/s12888-016-0790-0] [Medline: 27009062]
15. Cheung R, O'Donnell S, Madi N, Goldner EM. Factors associated with delayed diagnosis of mood and/or anxiety disorders. Health Promot Chronic Dis Prev Can 2017;37(5):137-148 [FREE Full text] [doi: 10.24095/hpcdp.37.5.02] [Medline: 28493658]
16. Kisely S, Scott A, Denney J, Simon G. Duration of untreated symptoms in common mental disorders: association with outcomes: international study. Br J Psychiatry 2006;189:79-80. [doi: 10.1192/bjp.bp.105.019869] [Medline: 16816310]
17. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. J Affect Disord 2014;152-154:45-51. [doi: 10.1016/j.jad.2013.10.002] [Medline: 24183486]
18. Spijker J, de Graaf R, Bijl RV, Beekman ATF, Ormel J, Nolen WA. Duration of major depressive episodes in the general population: results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry 2002;181:208-213. [doi: 10.1192/bjp.181.3.208] [Medline: 12204924]
19. Biesheuvel-Leliefeld KEM, Kok GD, Bockting CLH, Cuijpers P, Hollon SD, van Marwijk HWJ, et al. Effectiveness of psychological interventions in preventing recurrence of depressive disorder: meta-analysis and meta-regression. J Affect Disord 2015;174:400-410. [doi: 10.1016/j.jad.2014.12.016] [Medline: 25553400]
20. Bockting CLH, Smid NH, Koeter MWJ, Spinhoven P, Beck AT, Schene AH. Enduring effects of preventive cognitive therapy in adults remitted from recurrent depression: a 10 year follow-up of a randomized controlled trial. J Affect Disord 2015;185:188-194. [doi: 10.1016/j.jad.2015.06.048] [Medline: 26188380]
21. Holländare F, Anthony SA, Randestad M, Tillfors M, Carlbring P, Andersson G, et al. Two-year outcome of internet-based relapse prevention for partially remitted depression. Behav Res Ther 2013;51(11):719-722. [doi: 10.1016/j.brat.2013.08.002] [Medline: 24021360]

XSL•FO
RenderX

22. Krijnen-de Bruin E, Scholten W, Muntingh A, Maarsingh O, van Meijel B, van Straten A, et al. Psychological interventions to prevent relapse in anxiety and depression: a systematic review and meta-analysis. PLOS One 2022;17(8):e0272200 [FREE Full text] [doi: 10.1371/journal.pone.0272200] [Medline: 35960783]
23. Kuyken W, Warren FC, Taylor RS, Whalley B, Crane C, Bondolfi G, et al. Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials. JAMA Psychiatry 2016;73(6):565-574 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.0076] [Medline: 27119968]
24. Iyawa GE, Ondiek CO, Osakwe JO. mHealth: a low-cost approach for effective disease diagnosis, prediction, monitoring and management – effective disease diagnosis. In: Smart Medical Data Sensing and Iot Systems Design in Healthcare. United States: IGI Global; 2020:1-21.
25. Rohrbach PJ, Dingemans AE, Evers C, Van Furth EF, Spinhoven P, Aardoom JJ, et al. Cost-effectiveness of internet interventions compared with treatment as usual for people with mental disorders: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2023;25:e38204 [FREE Full text] [doi: 10.2196/38204] [Medline: 36602854]
26. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
27. Kählke F, Buntrock C, Smit F, Ebert DD. Systematic review of economic evaluations for internet- and mobile-based interventions for mental health problems. npj Digit Med 2022;5(1):175 [FREE Full text] [doi: 10.1038/s41746-022-00702-w] [Medline: 36424463]
28. Massoudi B, Holvast F, Bockting CLH, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: a systematic review and meta-analysis. J Affect Disord 2019;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]
29. Franquez RT, Del Grossi Moura M, McClung DCF, Barberato-Filho S, Cruz Lopes L, Silva MT, et al. E-Health technologies for treatment of depression, anxiety and emotional distress in person with diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin Pract 2023;203:110854 [FREE Full text] [doi: 10.1016/j.diabres.2023.110854] [Medline: 37499879]
30. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AAR, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR mHealth uHealth 2015;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
31. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of top-rated mobile apps for depression: systematic search and evaluation. JMIR Ment Health 2020;7(1):e15321 [FREE Full text] [doi: 10.2196/15321] [Medline: 32012079]
32. Leong QY, Sridhar S, Blasiak A, Tadeo X, Yeo G, Remus A, et al. Characteristics of mobile health platforms for depression and anxiety: content analysis through a systematic review of the literature and systematic search of two app stores. J Med Internet Res 2022;24(2):e27388 [FREE Full text] [doi: 10.2196/27388] [Medline: 35119370]
33. Stone A, Shiffman S. Ecological momentary assessment (ema) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
34. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
35. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
36. De Boer C, Ghomrawi H, Zeineddin S, Linton S, Kwon S, Abdullah F. A call to expand the scope of digital phenotyping. J Med Internet Res 2023;25:e39546 [FREE Full text] [doi: 10.2196/39546] [Medline: 36917148]
37. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
38. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018;17(3):276-277 [FREE Full text] [doi: 10.1002/wps.20550] [Medline: 30192103]
39. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
40. Torous J, Choudhury T, Barnett I, Keshavan M, Kane J. Smartphone relapse prediction in serious mental illness: a pathway towards personalized preventive care. World Psychiatry 2020;19(3):308-309 [FREE Full text] [doi: 10.1002/wps.20805] [Medline: 32931109]
41. Aleem S, Huda N, Amin R, Khalid S, Alshamrani S, Alshehri A. Machine learning algorithms for depression: diagnosis, insights, and research directions. Electronics 2022 Mar 31;11(7):1111 [FREE Full text] [doi: 10.3390/electronics11071111] [Medline: 23844154]
42. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, et al. Major depressive disorder. Nat Rev Dis Primers 2016;2:16065. [doi: 10.1038/nrdp.2016.65] [Medline: 27629598]
43. Tsuno N, Besset A, Ritchie K. Sleep and depression. J Clin Psychiatry 2005;66(10):1254-1269. [doi: 10.4088/jcp.v66n1008] [Medline: 16259539]

XSL•FO
RenderX

44. Mendelson W, Gillin JC, Wyatt RJ. Human Sleep and Its Disorders. New York: Plenum Press; 1977.
45. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
46. Mestdagh M, Verdonck S, Piot M, Niemeijer K, Kilani G, Tuerlinckx F, et al. m-Path: an easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. Front Digit Health 2023;5:1182175 [FREE Full text] [doi: 10.3389/fdgth.2023.1182175] [Medline: 37920867]
47. Onnela J. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021;46(1):45-54 [FREE Full text] [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
48. Roberts ET, Mehrotra A. Assessment of disparities in digital access among medicare beneficiaries and implications for telemedicine. JAMA Intern Med 2020;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
49. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
50. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
51. Hennemann S, Farnsteiner S, Sander L. Internet- and mobile-based aftercare and relapse prevention in mental disorders: a systematic review and recommendations for future research. Internet Interv 2018;14:1-17 [FREE Full text] [doi: 10.1016/j.invent.2018.09.001] [Medline: 30510909]
52. Moritz S, Grudzień DP, Gawęda, Aleksandrowicz A, Balzan R, Shaffy A, et al. A randomized controlled trial on COGITO, a free self-help smartphone app to enhance mental well-being. J Psychiatr Res 2024;174:254-257 [FREE Full text] [doi: 10.1016/j.jpsychires.2024.04.021] [Medline: 38670060]
53. Su L, Anderson PL. User behavior of a publicly available, free-to-use, self-guided mhealth app for depression: observational study in a global sample. JMIR Form Res 2022;6(10):e35538 [FREE Full text] [doi: 10.2196/35538] [Medline: 36282559]
54. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. J Med Internet Res 2014;16(4):e109 [FREE Full text] [doi: 10.2196/jmir.3084] [Medline: 24735567]
55. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
56. Beck AT, Alford BA. Depression: Causes and Treatment. Pennsylvania, PA: University of Pennsylvania Presss, Inc; 2009.
57. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and facilitators of engagement with mhealth technology for remote measurement and management of depression: qualitative analysis. JMIR mHealth uHealth 2019;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]
58. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of eMental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
59. Budd M, Iqbal A, Harding C, Rees E, Bhutani G. Mental health promotion and prevention in primary care: what should we be doing vs. what are we actually doing? Ment Health Prev 2021 Mar;21(19):200195-201616. [doi: 10.1016/j.mhp.2020.200195] [Medline: 22489612]
60. Thomas S, Jenkins R, Burch T, Calamos Nasir L, Fisher B, Giotaki G, et al. Promoting mental health and preventing mental illness in general practice. London J Prim Care (Abingdon) 2016;8(1):3-9 [FREE Full text] [doi: 10.1080/17571472.2015.1135659] [Medline: 28250821]
61. Singh V, Kumar A, Gupta S. Mental health prevention and promotion-a narrative review. Front Psychiatry 2022;13:898009 [FREE Full text] [doi: 10.3389/fpsyt.2022.898009] [Medline: 35958637]
62. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: a systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Trans Comput-Hum Interact 2020;27(5):1-53. [doi: 10.1145/3398069]
63. Patton J. Understanding user centricity. IEEE Softw 2007;24(6):9-11. [doi: 10.1109/ms.2007.181]
64. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
65. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
66. Vander Zwalmen Y, Demeester D, Hoorelbeke K, Verhaeghe N, Baeken C, Koster EHW. The more, the merrier? Establishing a dose-response relationship for the effects of cognitive control training on depressive symptomatology. J Consult Clin Psychol 2025;93(3):161-175. [doi: 10.1037/ccp0000945] [Medline: 40014506]

XSL•FO
RenderX

67. Bos FM, von Klipstein L, Emerencia AC, Veermans E, Verhage T, Snippe E, et al. A web-based application for personalized ecological momentary assessment in psychiatric care: user-centered development of the PETRA application. JMIR Ment Health 2022;9(8):e36430 [FREE Full text] [doi: 10.2196/36430] [Medline: 35943762]
68. Janssens KAM, Bos EH, Rosmalen JGM, Wichers MC, Riese H. A qualitative approach to guide choices for designing a diary study. BMC Med Res Methodol 2018;18(1):140 [FREE Full text] [doi: 10.1186/s12874-018-0579-6] [Medline: 30445926]
69. Snippe E, Smit AC, Kuppens P, Burger H, Ceulemans E. Recurrence of depression can be foreseen by monitoring mental states with statistical process control. J Psychopathol Clin Sci 2023;132(2):145-155. [doi: 10.1037/abn0000812] [Medline: 36808958]
70. Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: a qualitative study of patient and clinician experiences. Int J Bipolar Disord 2020;8(1):38 [FREE Full text] [doi: 10.1186/s40345-020-00201-5] [Medline: 33258015]
71. Weermeijer JDM, Wampers M, de Thurah L, Bonnier R, Piot M, Kuppens P, et al. Usability of the experience sampling method in specialized mental health care: pilot evaluation study. JMIR Form Res 2023;7:e48821 [FREE Full text] [doi: 10.2196/48821] [Medline: 37988137]
72. Saunders KEA, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes JR, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. Eur Psychiatry 2017;41:115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
73. Rimpler A, Siepe BS, Rieble CL, Proppert RKK, Fried EI. Introducing FRED: software for generating feedback reports for ecological momentary assessment data. Adm Policy Ment Health 2024;51(4):490-500 [FREE Full text] [doi: 10.1007/s10488-023-01324-4] [Medline: 38200261]
74. van Genugten CR, Schuurmans J, Lamers F, Riese H, Penninx BW, Schoevers RA, et al. Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. J Clin Med 2020;9(2):322 [FREE Full text] [doi: 10.3390/jcm9020322] [Medline: 31979340]
75. Kramer I, Simons CJP, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. World Psychiatry 2014;13(1):68-77 [FREE Full text] [doi: 10.1002/wps.20090] [Medline: 24497255]
76. Kidd PS, Parshall MB. Getting the focus and the group: enhancing analytical rigor in focus group research. Qual Health Res 2000;10(3):293-308. [doi: 10.1177/104973200129118453] [Medline: 10947477]
77. Kitzinger J. The methodology of focus groups: the importance of interaction between research participants. Sociol Health Illness 1994;16(1):103-121. [doi: 10.1111/1467-9566.ep11347023]
78. Krueger R, Casey M. Focus Groups: a Practical Guide for Applied Research. Thousand Oaks, CA: Sage; 2015.
79. Powell RA, Single HM. Focus groups. Int J Qual Health Care 1996;8(5):499-504. [doi: 10.1093/intqhc/8.5.499] [Medline: 9117204]
80. Robinson N. The use of focus group methodology—with selected examples from sexual health research. J Adv Nurs 1999;29(4):905-913. [doi: 10.1046/j.1365-2648.1999.00966.x] [Medline: 10215982]
81. Agar M, MacDonald J. Focus groups and ethnography. Hum Organ 1995;54(1):78-86. [doi: 10.17730/humo.54.1.x102372362631282]
82. Albrecht T, Johnson G, Walther J. Understanding communications processes in focus groups. In: Successful Focus Groups: Advancing the State of the Art. Newbury Park, CA: Sage; 1993:51-64.
83. Kaplowitz MD, Hoehn JP. Do focus groups and individual interviews reveal the same information for natural resource valuation? Ecol Econ 2001;36(2):237-247. [doi: 10.1016/s0921-8009(00)00226-3]
84. Guest G, Namey E, Taylor J, Eley N, McKenna K. Comparing focus groups and individual interviews: findings from a randomized study. Int J Soc Res Methodol 2017;20(6):693-708. [doi: 10.1080/13645579.2017.1281601]
85. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
86. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. J Am Med Inform Assoc 2016;23(3):477-484. [doi: 10.1093/jamia/ocv165] [Medline: 26911822]
87. Smitt M, Montakhabi M, Morton J, van LC, Bombeke K, Jacobs A. Users' perceptions of a digital stress self-monitoring application: research insights to design a practical innovation. 2022 Presented at: Universal Access in Human-Computer Interaction. Novel Design Approaches and Technologies: 16th International Conference, UAHCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022; June 26–July 1, 2022; Virtual Event p. 325-341. [doi: 10.1007/978-3-031-05028-2_22]
88. Springer A, Whittaker S. Progressive disclosure: empirically motivated approaches to designing effective transparency. 2019 Presented at: IUI '19: Proceedings of the 24th International Conference on Intelligent User Interfaces; March 17-20, 2019; Marina del Ray California p. 107-120. [doi: 10.1145/3301275.3302322]

XSL•FO
RenderX

89. Suh H, Shahriaree N, Hekler E, Kientz J. Developing and validating the user burden scale: a tool for assessing user burden in computing systems. 2016 Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose California USA p. 3988-3999. [doi: 10.1145/2858036.2858448]
90. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
91. Elgersma HJ, Koster EHW, van Tuijl LA, Hoekzema A, Penninx BWJH, Bockting CLH, et al. Attentional bias for negative, positive, and threat words in current and remitted depression. PLOS One 2018;13(10):e0205154 [FREE Full text] [doi: 10.1371/journal.pone.0205154] [Medline: 30379840]
92. Bos FM, Snippe E, Bruggeman R, Wichers M, van der Krieke L. Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care. Psychiatr Serv 2019;70(11):983-991. [doi: 10.1176/appi.ps.201900050] [Medline: 31434558]
93. Hartmann R, Sander C, Lorenz N, Böttger D, Hegerl U. Utilization of patient-generated data collected through mobile devices: insights from a survey on attitudes toward mobile self-monitoring and self-management apps for depression. JMIR Ment Health 2019;6(4):e11671 [FREE Full text] [doi: 10.2196/11671] [Medline: 30942693]
94. Alqahtani F, Winn A, Orji R. Co-designing a mobile app to improve mental health and well-being: focus group study. JMIR Form Res 2021;5(2):e18172 [FREE Full text] [doi: 10.2196/18172] [Medline: 33635281]
95. Langener AM, Stulp G, Jacobson NC, Costanzo A, Jagesar RR, Kas MJ, et al. It's all about timing: exploring different temporal resolutions for analyzing digital-phenotyping data. Adv Methods Pract Psychol Sci 2024;7(1):1-22. [doi: 10.1177/25152459231202677]
96. Bos FM, Schreuder MJ, George SV, Doornbos B, Bruggeman R, van der Krieke L, et al. Anticipating manic and depressive transitions in patients with bipolar disorder using early warning signals. Int J Bipolar Disord 2022;10(1):12 [FREE Full text] [doi: 10.1186/s40345-022-00258-4] [Medline: 35397076]
97. Cai L, Boukhechba M, Wu C, Chow P, Teachman B, Barnes L. State affect recognition using smartphone sensing data. 2018 Presented at: CHASE '18: Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies; September 26-28, 2018; DC, Washington.
98. Heijmans M, Habets J, Kuijf M, Kubben P, Herff C. Evaluation of parkinson's disease at home: predicting tremor from wearable sensors. 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany p. 584-587.
99. Stiles-Shields C, Montague E, Lattie EG, Kwasny MJ, Mohr DC. What might get in the way: Barriers to the use of apps for depression. Digit Health 2017;3:2055207617713827 [FREE Full text] [doi: 10.1177/2055207617713827] [Medline: 29942605]
100. Peckham AD, McHugh RK, Otto MW. A meta-analysis of the magnitude of biased attention in depression. Depress Anxiety 2010;27(12):1135-1142. [doi: 10.1002/da.20755] [Medline: 21049527]
101. Beck AT, Clark DA. Anxiety and depression: an information processing perspective. Anxiety Res 1988;1(1):23-36. [doi: 10.1080/10615808808248218]
102. De Raedt R, Koster EHW. Understanding vulnerability for depression from a cognitive neuroscience perspective: a reappraisal of attentional factors and a new conceptual framework. Cogn Affect Behav Neurosci 2010;10(1):50-70. [doi: 10.3758/CABN.10.1.50] [Medline: 20233955]
103. Rosenfeld EA, Lyman C, Roberts JE. Development of an mHealth app-based intervention for depressive rumination (RuminAid): mixed methods focus group evaluation. JMIR Form Res 2022;6(12):e40045 [FREE Full text] [doi: 10.2196/40045] [Medline: 36512400]
104. Seligman MEP, Csikszentmihalyi M. Positive psychology: an introduction. Am Psychol 2000;55(1):5-14. [doi: 10.1037//0003-066x.55.1.5] [Medline: 11392865]
105. Seligman MEP, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
106. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
107. Pietrowsky R, Mikutta J. Effects of positive psychology interventions in depressive patients—A randomized control study. Psychol 2012;03(12):1067-1073. [doi: 10.4236/psych.2012.312158]
108. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009;65(5):467-487 [FREE Full text] [doi: 10.1002/jclp.20593] [Medline: 19301241]
109. Goodmon LB, Middleditch AM, Childs B, Pietrasiuk SE. Positive psychology course and its relationship to well-being, depression, and stress. Teach Psychol 2016;43(3):232-237. [doi: 10.1177/0098628316649482]
110. Santos V, Paes F, Pereira V, Arias-Carrión O, Silva AC, Carta MG, et al. The role of positive emotion and contributions of positive psychology in depression treatment: systematic review. Clin Pract Epidemiol Ment Health 2013;9:221-237 [FREE Full text] [doi: 10.2174/1745017901309010221] [Medline: 24358052]

XSL•FO
RenderX

111. Stojchevska M, Steenwinckel B, Van Der Donckt J, De Brouwer M, Goris A, De Turck F, et al. Assessing the added value of context during stress detection from wearable data. BMC Med Inform Decis Mak 2022;22(1):268 [FREE Full text] [doi: 10.1186/s12911-022-02010-5] [Medline: 36243691]
112. Wichers M, Simons CJP, Kramer IMA, Hartmann JA, Lothmann C, Myin-Germeys I, et al. Momentary assessment technology as a tool to help patients with depression help themselves. Acta Psychiatr Scand 2011;124(4):262-272. [doi: 10.1111/j.1600-0447.2011.01749.x] [Medline: 21838742]
113. Walsh AEL, Naughton G, Sharpe T, Zajkowska Z, Malys M, van Heerden A, et al. A collaborative realist review of remote measurement technologies for depression in young people. Nat Hum Behav 2024;8(3):480-492 [FREE Full text] [doi: 10.1038/s41562-023-01793-5] [Medline: 38225410]
114. Awad NF, Krishnan MS. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. MIS Q 2006;30(1):13-28. [doi: 10.2307/25148715]
115. Karwatzki S, Dytynko O, Trenz M, Veit D. Beyond the personalization–privacy paradox: privacy valuation, transparency features, and service personalization. J Manage Inf Syst 2017;34(2):369-400. [doi: 10.1080/07421222.2017.1334467]
116. Betzing JH, Tietz M, vom Brocke J, Becker J. The impact of transparency on mobile privacy decision making. Electron Mark 2020;30(3):607-625. [doi: 10.1007/s12525-019-00332-3]
117. Zimmer JC, Arsal R, Al-Marzouq M, Moore D, Grover V. Knowing your customers: using a reciprocal relationship to enhance voluntary information disclosure. Dec Support Syst 2010;48(2):395-406. [doi: 10.1016/j.dss.2009.10.003]
118. Tsai JY, Egelman S, Cranor L, Acquisti A. The effect of online privacy information on purchasing behavior: an experimental study. Inf Syst Res 2011;22(2):254-268. [doi: 10.1287/isre.1090.0260]
119. John LK, Acquisti A, Loewenstein G. Strangers on a plane: context-dependent willingness to divulge sensitive information. J Consum Res 2011;37(5):858-873. [doi: 10.1086/656423]
120. Krebs P, Duncan DT. Health app use among us mobile phone owners: a national survey. JMIR mHealth uHealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
121. Zimmer M, Kumar P, Vitak J, Liao Y, Chamberlain Kritikos K. 'There's nothing really they can do with this information': unpacking how users manage privacy boundaries for personal fitness information. Inf Commun Soc 2018;23(7):1020-1037. [doi: 10.1080/1369118x.2018.1543442]
122. Elevelt A, Lugtig P, Toepoel V. Doing a time use survey on smartphones only: what factors predict nonresponse at different stages of the survey process? Surv Res Methods 2019;13(2):195-213. [doi: 10.18148/srm/2019.v13i2.7385]
123. Wenz A, Jäckle A, Couper M. Willingness to use mobile technologies for data collection in a probability household panel. Surv Res Methods 2017;13(1):1-22. [doi: 10.18148/srm/2019.v1i1.7298]
124. Revilla M, Couper M, Ochoa C. Willingness of online panelists to perform additional tasks. Methods Data Anal 2019;13(2):223-252. [doi: 10.12758/mda.2018.01]
125. Keusch F, Struminskaya B, Antoun C, Couper MP, Kreuter F. Willingness to participate in passive mobile data collection. Public Opin Q 2019;83(Suppl 1):210-235 [FREE Full text] [doi: 10.1093/poq/nfz007] [Medline: 31337924]
126. Norberg PA, Horne DR, Horne DA. The privacy paradox: personal information disclosure intentions versus behaviors. J Consum Aff 2007;41(1):100-126. [doi: 10.1111/j.1745-6606.2006.00070.x]
127. Barth S, de Jong MDT. The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – a systematic literature review. Telemat Inform 2017;34(7):1038-1058. [doi: 10.1016/j.tele.2017.04.013]
128. Gerber N, Gerber P, Volkamer M. Explaining the privacy paradox: a systematic review of literature investigating privacy attitude and behavior. Comput Secur 2018;77:226-261. [doi: 10.1016/j.cose.2018.04.002]
129. Acquisti A, Grossklags J. Privacy and rationality in individual decision making. IEEE Secur Priv 2005;3(1):26-33. [doi: 10.1109/MSP.2005.22]
130. Culnan MJ, Armstrong PK. Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation. Organ Sci 1999;10(1):104-115. [doi: 10.1287/orsc.10.1.104]
131. Laufer RS, Wolfe M. Privacy as a concept and a social issue: a multidimensional developmental theory. J Soc Issues 1977;33(3):22-42. [doi: 10.1111/j.1540-4560.1977.tb01880.x]
132. Wottrich VM, van Reijmersdal EA, Smit EG. The privacy trade-off for mobile app downloads: the roles of app value, intrusiveness, and privacy concerns. Decis Support Syst 2018;106:44-52. [doi: 10.1016/j.dss.2017.12.003]
133. Kim D, Park K, Park Y, Ahn J. Willingness to provide personal information: perspective of privacy calculus in IoT services. Comput Hum Behav 2019;92:273-281. [doi: 10.1016/j.chb.2018.11.022]
134. Wang T, Duong TD, Chen CC. Intention to disclose personal information via mobile applications: a privacy calculus perspective. Int J Inf Manage 2016;36(4):531-542. [doi: 10.1016/j.ijinfomgt.2016.03.003]
135. Lee A. Investigating the personalization–privacy paradox in Internet of Things (IoT) based on dual-factor theory: moderating effects of type of IoT service and user value. sustainability 2021;13(19):10679. [doi: 10.3390/su131910679]
136. Chellappa RK, Sin RG. Personalization versus privacy: an empirical examination of the online consumer's dilemma. Inf Technol Manage 2005;6(2):181-202. [doi: 10.1007/s10799-005-5879-y]
137. Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups. Electron Commer Res Appl 2016;16:55-65. [doi: 10.1016/j.elerap.2015.11.001]

XSL•FO
RenderX

138. Conner TS, Reid KA. Effects of intensive mobile happiness reporting in daily life. Soc Psychol Pers Sci 2011;3(3):315-323. [doi: 10.1177/1948550611419677]
139. Smit AC, Snippe E. Real-time monitoring of increases in restlessness to assess idiographic risk of recurrence of depressive symptoms. Psychol Med 2023;53(11):5060-5069 [FREE Full text] [doi: 10.1017/S0033291722002069] [Medline: 35833374]
140. Schreuder MJ, Schat E, Smit AC, Snippe E, Ceulemans E. Monitoring emotional intensity and variability to forecast depression recurrence in real time in remitted adults. J Consult Clin Psychol 2024;92(8):505-516. [doi: 10.1037/ccp0000871] [Medline: 38512172]
141. van Rijsbergen GD, Burger H, Hollon SD, Elgersma HJ, Kok GD, Dekker J, et al. How do you feel? Detection of recurrent major depressive disorder using a single-item screening tool. Psychiatry Res 2014;220(1-2):287-293. [doi: 10.1016/j.psychres.2014.06.052] [Medline: 25070177]
142. van Helvoort D, Otgaar H, Merckelbach H. Worsening of self-reported symptoms through suggestive feedback. Clin Psychol Sci 2020;8(2):359-365. [doi: 10.1177/2167702619869184]
143. Ehring T, Watkins ER. Repetitive negative thinking as a transdiagnostic process. Int J Cogn Ther 2008;1(3):192-205. [doi: 10.1680/ijct.2008.1.3.192]
144. Bartoskova M, Sevcikova M, Durisko Z, Maslej MM, Barbic SP, Preiss M, et al. The form and function of depressive rumination. Evol Hum Behav 2018;39(3):277-289. [doi: 10.1016/j.evolhumbehav.2018.01.005]
145. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. Perspect Psychol Sci 2008;3(5):400-424 [FREE Full text] [doi: 10.1111/j.1745-6924.2008.00088.x] [Medline: 26158958]
146. Dobson KS, Hollon SD, Dimidjian S, Schmaling KB, Kohlenberg RJ, Gallop RJ, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. J Consult Clin Psychol 2008;76(3):468-477 [FREE Full text] [doi: 10.1037/0022-006X.76.3.468] [Medline: 18540740]
147. Clarke K, Mayo-Wilson E, Kenny J, Pilling S. Can non-pharmacological interventions prevent relapse in adults who have recovered from depression? A systematic review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2015;39:58-70. [doi: 10.1016/j.cpr.2015.04.002] [Medline: 25939032]
148. Katon W, Rutter C, Ludman EJ, Von Korff M, Lin E, Simon G, et al. A randomized trial of relapse prevention of depression in primary care. Arch Gen Psychiatry 2001;58(3):241-247. [doi: 10.1001/archpsyc.58.3.241] [Medline: 11231831]
149. Brown G, Marshall M, Bower P, Woodham A, Waheed W. Barriers to recruiting ethnic minorities to mental health research: a systematic review. Int J Methods Psychiatr Res 2014;23(1):36-48 [FREE Full text] [doi: 10.1002/mpr.1434] [Medline: 24474683]
150. Eaton WW, Shao H, Nestadt G, Lee HB, Bienvenu OJ, Zandi P. Population-based study of first onset and chronicity in major depressive disorder. Arch Gen Psychiatry 2008;65(5):513-520 [FREE Full text] [doi: 10.1001/archpsyc.65.5.513] [Medline: 18458203]
151. Corney RH. Sex differences in general practice attendance and help seeking for minor illness. J Psychosom Res 1990;34(5):525-534. [doi: 10.1016/0022-3999(90)90027-2] [Medline: 2231486]
152. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022;27(1):281-295 [FREE Full text] [doi: 10.1038/s41380-021-01161-7] [Medline: 34079068]
153. Lewinsohn PM, Clarke GN, Seeley JR, Rohde P. Major depression in community adolescents: age at onset, episode duration, and time to recurrence. J Am Acad Child Adolesc Psychiatry 1994;33(6):809-818. [doi: 10.1097/00004583-199407000-00006] [Medline: 7598758]
154. Rutter LA, Howard J, Lakhan P, Valdez D, Bollen J, Lorenzo-Luaces L. "I haven't been diagnosed, but I should be"-insight into self-diagnoses of common mental health disorders: cross-sectional study. JMIR Form Res 2023;7:e39206 [FREE Full text] [doi: 10.2196/39206] [Medline: 36637885]
155. Craven MP, Lang AR, Martin JL. Developing mHealth apps with researchers: multi-stakeholder design considerations. In: Design, User Experience, and Usability. User Experience Design for Everyday Life Applications and Services. DUXU 2014. Lecture Notes in Computer Science. Switzerland: Springer, Cham; 2014.
156. Smithson J. Using and analysing focus groups: limitations and possibilities. Int J Soc Res Methodol 2000;3(2):103-119. [doi: 10.1080/136455700405172]
157. Janis I. Victims of Groupthink. Boston: Houghton Mifflin; 1972.
158. MacDougall C, Baum F. The devil's advocate: a strategy to avoid groupthink and stimulate discussion in focus groups. Qual Health Res 1997;7(4):532-541. [doi: 10.1177/104973239700700407]

Abbreviations

DEDICAT: Depression's Digital Forecasting Tool
EMA: ecological momentary assessment
mHealth: mobile health

XSL•FO
RenderX

Edited by L Buis; submitted 04.10.24; peer-reviewed by C Lidynia, K Mouloudj; comments to author 13.03.25; revised version received 17.04.25; accepted 02.10.25; published 23.10.25.

Please cite as:
Coenen T, Maerevoet M, Chen S, De Brouwer M, Van Hoecke S, Koster EHW, Vanden Abeele MMP, Bombeke K
Requirements and Concerns of Individuals Remitted From Depression for an Early Relapse Detection mHealth App: Focus Group Study
JMIR Mhealth Uhealth 2025;13:e67141
URL: https://mhealth.jmir.org/2025/1/e67141
doi:10.2196/67141
PMID:

©Tina Coenen, Matthias Maerevoet, Stephanie Chen, Mathias De Brouwer, Sofie Van Hoecke, Ernst HW Koster, Mariek MP Vanden Abeele, Klaas Bombeke. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Pilot Study of a Mobile, Virtual Reality–Based Digital Therapeutic for Smoking Cessation: Randomized Controlled Trial

Yeong-Seon Jo[1*], PhD; Arom Pyeon[1*], MA; Min-Kyung Hu[2], BA; SungMin Kim[1], MA; In-Young Choi[1], PhD; Dai-Jin Kim[1], MD, PhD; Ji-Won Chun[2,3], PhD

[1]Department of Medical Science, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, Seoul, Republic of Korea
[2]Department of Medical Informatics, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, Seoul, Republic of Korea
[3]Catholic Medical Center Institute for Basic Medical Science, The Catholic University of Korea, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Ji-Won Chun, PhD
Department of Medical Informatics
Seoul St. Mary's Hospital
The Catholic University of Korea College of Medicine
222 Banpo-daero, Seocho-gu
Seoul, 06591
Republic of Korea
Phone: 82 2 3147 9727
Email: jwchun@catholic.ac.kr

Abstract

Background: Smoking cessation remains a global challenge, with traditional treatments showing limited long-term success due to low adherence and accessibility issues. Digital therapeutics, such as mobile apps and virtual reality (VR)–based interventions, could offer innovative solutions for smoking cessation treatment. NICO-THERA is a digital therapeutic program integrating cognitive behavioral therapy (CBT) and motivational enhancement therapy (MET) to address nicotine dependence.

Objective: This study aims to evaluate the safety and efficacy of NICO-THERA, a digital therapeutic intervention combining VR and mobile apps, for supporting smoking cessation among individuals with nicotine dependence. The primary focus was on smoking abstinence, nicotine dependence reduction, and motivation to quit smoking over a 12-week intervention period.

Methods: This open-label, exploratory, randomized controlled trial involved 30 participants randomly assigned to the digital therapeutic group (DTG; n=15; mean age 43.1 years; mean number of daily cigarettes: 9.5) or basic treatment group (BTG; n=15; mean age 48.7 years; mean number of daily cigarettes: 9.2). The DTG received the NICO-THERA program, involving VR sessions (relaxation training, craving coping, and refusal skills) and mobile app–based CBT and MET modules for daily therapeutic exercises. The BTG received basic care including video education and printed materials. The primary outcomes were the 7-day point prevalence abstinence (PPA) and 30-day PPA at 8 weeks and 12 weeks. Secondary outcomes included nicotine dependence (Fagerström Test for Nicotine Dependence [FTND]) and motivation to change (Stages of Change Readiness and Treatment Eagerness Scale–Smoking [SOCRATES-S]).

Results: There were no significant differences in 7-day and 30-day PPA between the DTG and BTG. However, in within-group analyses, the DTG showed significant improvements in both 7-day and 30-day PPA at week 8 (z=–2.00, P=.046), along with consistent reductions in smoking days and cigarette consumption across all time points. The BTG only significantly decreased cigarette consumption at week 8. Additionally, the DTG exhibited a significant increase in taking steps of motivation at week 12 (U=19.00, P=.048) compared with the BTG. No adverse device effects were reported. Adherence to smoking cessation diaries and medication logs was higher in the DTG (mean 83 of 84 days, 99%; 9 participants) than in the BTG (mean 74 of 84 days, 88%; 12 participants), based on the 12-week average adherence among those who completed the study.

Conclusions: The NICO-THERA digital therapeutic program demonstrated preliminary effectiveness at reducing nicotine dependence. Additionally, participants in the DTG group exhibited a progressive improvement in their motivation to quit smoking, as reflected by a significant reduction in ambivalence at week 8 and increase in proactive cessation efforts at week 12. These results suggest that the structured integration of MET and CBT within the NICO-THERA program effectively reinforced psychological readiness for smoking cessation, fostering a sustained commitment to behavioral change.

XSL•FO
RenderX

Trial Registration: Clinical Research Information Service, Republic of Korea KCT0009801; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=28285&status=5&seq_group=28285&search_page=M

(JMIR Mhealth Uhealth 2025;13:e66411) doi:10.2196/66411

KEYWORDS

digital therapeutics; virtual reality; VR; mobile health; mHealth; smoking cessation; nicotine dependence; cognitive behavioral therapy; CBT; motivational enhancement therapy; MET; randomized controlled trial; RCT

Introduction

The global tobacco crisis continues to be one of the most severe public health challenges, with more than 8 million fatalities annually worldwide [1], of which 6.2 million deaths are directly linked to tobacco consumption and approximately 1.3 million are owing to exposure to second-hand smoke among nonsmokers [2]. According to the 2023 Organisation for Economic Co-operation and Development (OECD) Health Statistics, the average smoking rate across OECD countries is 15.9%, with South Korea showing a comparable rate of 15.4% [3]. Tobacco use is a leading contributor to a multitude of health issues, including lung cancer, heart disease, and chronic respiratory conditions, resulting in significant social and economic repercussions [4]. Although South Korea has implemented comprehensive smoking cessation policies and invested substantially in public health initiatives [5], as the variety of tobacco products increases, the willingness of smokers to quit smoking has been under threat [6]. These trends indicate that smoking cessation policies must evolve to address the changing landscape of tobacco use and cessation efforts.

Despite the known harm of smoking, many smokers find quitting to be a challenging endeavor. Characterized as a chronic relapsing condition [7], smoking cessation frequently leads to a range of withdrawal symptoms such as cravings, irritability, depression, and restlessness [8]. Notably, the success rate of individuals attempting to quit smoking unaided without professional support remains low, with estimates ranging from 3% to 5% [9]. Medications are designed to mitigate the discomfort associated with nicotine withdrawal symptoms, thereby significantly enhancing the likelihood of smoking cessation [10,11]. The most effective smoking cessation strategy appears to be a combination of behavioral therapy and pharmacotherapy [10-13], a method widely supported by national smoking cessation programs across various countries [14-19].

To overcome the limitations of accessibility identified in traditional smoking cessation treatments and provide cost-effective and time-efficient interventions, the development of digital therapeutics (DTx) for smoking cessation has been extensively pursued. Mobile health (mHealth) technologies for smoking cessation began as basic text message–based interventions. Although effective [20-23], these early digital tools had limited functionality [21,24]. Over time, these have evolved into complex smartphone apps and other forms of DTx that use technology to offer more accessible, flexible, and personalized support [25-28]. DTx are characterized as interventions based on evidence provided by certified software programs to prevent, manage, or treat medical conditions, separating them from conventional wellness products [29]. Following the development of wellness apps for smoking cessation, some DTx specifically designed for smoking cessation have incorporated behavioral strategies such as cognitive behavioral therapy (CBT) to support cessation efforts, while others have focused on alternative behavioral modification techniques [30-34]. The efficacy of DTx devices has also been confirmed in studies with small sample sizes [35].

In addition to smartphone-based interventions, encouraging advances have been made in the use of virtual reality (VR) in digital medicine. VR has been implemented in consultations and hospital settings under the supervision of health professionals [36]. Incorporating VR into treatments can alleviate patient symptoms in conditions such as claustrophobia [37,38], post-traumatic stress disorder [39], and smoking cessation [40,41] and may improve adherence to weight loss behaviors [42]. Expanding the role of VR in smoking cessation, recent studies have explored its potential to not only assist with reducing cravings and withdrawal symptoms but also modify behavioral patterns and increase long-term abstinence rates. For instance, the cue exposure therapy approach, which uses VR to simulate real-life smoking triggers in a controlled environment, has demonstrated promising results for helping individuals resist cravings by practicing coping strategies in virtual scenarios [40]. Similarly, immersive VR environments have been used for mindfulness-based interventions to reduce stress and anxiety associated with quitting smoking, thereby supporting overall mental well-being during cessation efforts [41].

Building on recent advances in digital-based smoking cessation treatments, such as mobile apps and VR, we developed NICO-THERA, a unified DTx that integrates both a mobile app and VR to promote smoking cessation. Although some prior studies have combined VR with therapeutic components—such as cue exposure therapy or mindfulness-based approaches [40,41]—these were typically delivered alongside CBT or as independent modules, rather than through an integrated, week-by-week structure. In contrast, NICO-THERA incorporates CBT and motivational enhancement therapy (MET) techniques into a structured digital program in which VR sessions are systematically aligned with weekly cognitive-behavioral content. This study examined the safety and efficacy of NICO-THERA as a comprehensive, multimodal intervention, rather than relying on standalone VR or mobile app treatments.

XSL•FO
RenderX

Methods

Study Design

This study was conducted as a 2-arm, open-label, exploratory, randomized controlled clinical trial (RCT). The primary objectives were to evaluate the preliminary efficacy and safety of NICO-THERA for smoking cessation. Outcomes were assessed directly by the researchers without independent evaluators. Although the absence of blinding may introduce potential biases, the focus of this study was on generating initial data rather than establishing definitive efficacy conclusions.

Participants were randomly assigned to either the digital therapeutic group (DTG) or the basic treatment group (BTG) using a pregenerated 1:1 random allocation sequence. Each participant was assigned to a group according to their screening number in the order of arrival. The allocation sequence was predetermined for screening numbers 1 to 30, ensuring that neither the participants nor the researchers knew in advance which group the participants would be allocated to, thereby maintaining the randomness of the allocation process. According to guidelines for pilot trials, a minimum sample size of 12 participants per treatment arm was recommended to ensure meaningful preliminary analysis [43,44]. Additionally, small-sample RCT studies, including those related to smoking cessation, have been conducted with a minimum of 9 and a maximum of 19 participants per group. [45-48]. In this study, 15 participants were enrolled in each group (DTG and BTG) to account for potential attrition, resulting in a total enrollment of 30 individuals. No significant changes in the trial methodology were made after the trial commenced.

We used the Simulator Sickness Questionnaire (SSQ), administering it before and after each VR session to monitor potential adverse effects and evaluate safety in DTG. To indicate preliminary efficacy, we measured 7-day and 30-day point prevalence abstinence (PPA) rates as primary outcomes, as well as changes in nicotine dependence and motivation to quit smoking as secondary outcomes. Additionally, pharmacological therapy and participation in outpatient counseling were monitored through clinical records.

Participants and Recruitment

The study population consisted of patients visiting Seoul St. Mary's Hospital for smoking cessation and those who responded to a clinical study recruitment advertisement. Written informed consent was obtained from all participants before eligibility screening. Medical staff evaluated the participants based on the inclusion and exclusion criteria through individual interviews. Data were collected in a clinical setting at Seoul St. Mary's Hospital, where the participants attended outpatient visits. Screening and follow-up assessments were performed in clinical settings. Participants who met the eligibility criteria were randomly assigned to either the DTG or BTG. Regardless of the assigned group, all participants received usual care for smoking cessation, which included counseling from medical staff at every visit as well as personalized pharmacological therapy with varenicline or bupropion. Of the 21 participants, 14 (including 7 in the DTG) began pharmacotherapy at baseline, while the remaining 7 (including 2 in the DTG) initiated pharmacotherapy later in the study. The in-treatment period commenced with randomization and lasted until week 12, during which follow-up visits were conducted at 4-week intervals (weeks 4, 8, and 12).

The inclusion criteria were as follows: (1) adults who were at least 19 years old and younger than 75 years; (2) individuals diagnosed with nicotine addiction according to the clinical criteria for mental and behavioral disorders owing to use of tobacco (F17) or toxic effect of tobacco and nicotine (T652) as per the Korean Classification of Diseases and in accordance with the clinical criteria for nicotine use disorders specified in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [49]; (3) individuals who smoked an average of 5 or more cigarettes per day in the past 6 months; (4) participants who were able to communicate with the researcher and consented to the procedures required by the study protocol, understanding the objectives of the research and signing the consent form; (5) smartphone users who agreed to install and use an app for data collection and management; and (6) smartphone users who agreed to receive text messages and respond to survey items via their smartphones, consenting to data collection and management.

The exclusion criteria included (1) individuals who did not meet the clinical criteria for mental and behavioral disorders owing to use of tobacco (F17) or toxic effect of tobacco and nicotine (T652) according to the Korean Classification of Diseases, based on individual interviews with a specialist or researcher; (2) individuals with cognitive impairments affecting decision-making capacity; (3) individuals with logistical or personal constraints preventing full participation in the study, including scheduled visits and intervention activities (eg, long-term travel plans, conflicting work schedules, or severe mobility limitations); (4) those who had started or were undergoing another CBT within the last 3 months; (5) individuals who either did not own a smartphone registered in their name or lacked sufficient ability to use a smartphone for study-related tasks, such as receiving text messages, completing self-report logs, and engaging with app-based interventions; (6) individuals without a smartphone or significantly limited ability to use smartphones including receiving text messages; (7) those at severe risk of depression, suicidal ideation, or suicide attempts; and (8) other reasons deemed by the investigator as inappropriate for participation in the clinical trial. The participants completed eligibility screening and provided informed consent.

Intervention Condition

DTG

NICO-THERA, the DTx used in this study, is software as a medical device designed to treat and manage nicotine dependence. NICO-THERA has been classified as a Class 2 (low-risk) medical device for cognitive therapy software by the Ministry of Food and Drug Safety in South Korea (E066060.02). Figure 1 illustrates the interface elements of the NICO-THERA, including screenshots of its main features. The app's content was reviewed by clinical experts in smoking cessation and CBT to ensure the accuracy and effectiveness of the therapeutic

XSL•FO
RenderX

modules. The NICO-THERA app was developed through several iterations, including usability testing with target users, to refine the interface and content based on feedback. The intervention for the DTG was a structured 12-week program consisting of 2 key components: VR therapy content provided via a head-mounted display and a smartphone app. This nicotine DTx content is based on CBT and MET, which are prominent evidence-based therapies aimed at improving addictive disorders. The NICO-THERA app and VR content used in this study were all from Version 1.0, ensuring consistent therapeutic content with no updates nor modifications throughout the trial.

The VR content was delivered using the HTC Vive Cosmos headset, which includes integrated headphones for immersive audio output. VR sessions were conducted exclusively in the hospital at designated PC stations, ensuring a controlled and consistent environment. Participants underwent 3 VR sessions at weeks 0, 4, and 8 during their clinic visits, with each session lasting approximately 10 minutes. Each VR session focused on a specific module, progressing sequentially throughout the intervention period: Image Relaxation Training at week 0, Craving Coping Training at week 4, and Refusal Training at week 8. Image relaxation training, rooted in behavior modification theory, helps participants cope with stress-related triggers [50,51] such as depression and anxiety by practicing deep breathing and relaxation techniques within a VR environment that replicates real-life settings. Participants can choose a preferred virtual setting among the forest, at a campfire, or on the beach to conduct the imagery relaxation training (Multimedia Appendix 1). Craving coping training (Multimedia Appendix 2) uses role-playing in a realistic VR setting to apply learned relaxation and behavioral strategies to manage situations that trigger smoking cravings [52,53]. Finally, refusal training (Multimedia Appendix 3) focuses on strengthening the participants' ability to clearly refuse smoking offers by practicing assertive communication and refusal skills [54] in VR scenarios that mirror real-life interactions, thereby bolstering their smoking cessation efforts. Participants interacted with the VR environment using handheld controllers to navigate scenarios and respond to prompts, such as selecting preferred settings (eg, forest, campfire, or beach) or engaging in role-playing scenarios. In addition, although VR sessions were separate from the smartphone app, data from participant responses in the VR session were synchronized with the smartphone app for program monitoring.

The participants accessed the NICO-THERA app on their smartphones. The app was provided free of charge, and the participants were given detailed instructions on how to download and install it. The app required an internet connection for certain features, and participants were encouraged to maintain access to the internet throughout the study.

The NICO-THERA app integrates CBT and MET to deliver structured therapeutic interventions tailored to nicotine addiction. According to the stages of change model, intentional behavior change, such as quitting smoking, typically progresses through 5 stages: Precontemplation, Contemplation, Preparation, Action, and Maintenance [55-57]. NICO-THERA is a 12-week program that aligns with the 5 stages of change. It is also designed to support smoking cessation and maintenance by integrating evidence-based educational, cognitive, behavioral, and emotional techniques derived from MET and CBT [30,58]. This app is designed to facilitate users' engagement in their treatment processes by providing therapeutic sessions that help identify triggers, manage cravings, and develop effective coping strategies. Each of the 12 weekly topics was released sequentially, with one topic becoming available each week to align with the structured program. Participants were allowed to review past topics multiple times but could not access future topics ahead of schedule. This comprehensive approach encompasses operant conditioning to modify behaviors associated with smoking, cognitive restructuring to address dysfunctional thoughts related to nicotine use, and relaxation training to manage stress and emotional triggers. Specifically, the initial 1 to 4 weeks of the program emphasized MET approaches, which are effective for promoting behavioral changes for smoking cessation. From week 5 onward, the program gradually introduced cognitive, behavioral, and emotional techniques focused on CBT, helping individuals acquire strategies that are practically useful for maintaining smoking cessation. Each lesson took approximately 10 minutes per day to complete. Upon completion of the content each week, a quiz was provided to evaluate the effectiveness of the weekly program and reinforce the users' knowledge. The weekly topics and contents are detailed in Table 1.

In addition to these core therapeutic strategies, the app is equipped with several practical tools to support patients and health care providers. It includes a smoking cessation diary and medication adherence log, which are instrumental for tracking daily smoking behaviors and medication intake. Participants were instructed to record their smoking cessation diary and medication adherence log daily and to complete the weekly lessons provided through the app. Each lesson took approximately 10 minutes per day to complete. These features allow health care providers to monitor the patient's progress and compliance more effectively during clinical visits, providing crucial data that can be used to adjust treatment plans and interventions accordingly (Multimedia Appendix 4). The participants received weekly push notifications from the app reminding them to complete their therapy sessions and log their smoking behaviors.

XSL•FO
RenderX

Figure 1. Overview of the NICO-THERA digital therapeutic program, including app screenshots of the main dashboard and cognitive behavioral training (CBT)-based contents and virtual reality (VR)-based sessions for relaxation training, craving coping, and refusal training.

XSL•FO
RenderX

Table 1. Cognitive behavioral therapy (CBT) and motivational enhancement therapy (MET) techniques included in the NICO-THERA app.

Time point	Weekly program topics	Detailed content
Registration	Understanding Application Usage	Familiarization with the application's functionalities; completion of registration and provide basic information
Week 1	Treatment Plan and Functional Analysis	Structured approach for addiction treatment and identification of users' nicotine usage patterns
Week 2	Motivation Enhancement Training	Education on motivational theories, change motivations, and stages of change
Week 3	Creating an Environment for Success	Understanding conditioning and association and eliminating factors related to smoking
Week 4	Identifying Triggers and Cravings	Pinpointing specific smoking triggers and understanding the underlying mechanisms of cravings
Week 5	Relaxation Training	Education and implementation of relaxation training with biofeedback signals
Week 6	Behavioral Craving Coping Strategies	Understanding cravings and learning strategies to cope with them
Week 7	Identifying Cognitive Distortions	Recognizing and categorizing thought patterns that contribute to smoking behavior
Week 8	Responding to Cognitive Distortions	Developing strategies to challenge and reframe distorted thoughts
Week 9	Refusal Training	Learning assertive communication skills useful in smoking cessation
Week 10	Coping with Depression and Anxiety	Learning about and dealing with depression and anxiety
Week 11	Coping with Anger	Learning about and dealing with anger
Week 12	Relapse Prevention and Review	Recognizing the possibility of relapse and responding to warning signs

BTG

In the control group, participants received both video and printed educational materials as part of their treatment for nicotine addiction. Specifically, video education comprised CBT-focused content tailored to manage nicotine addiction. We provided 3 videos during the study: the first video at baseline (week 0), the second at week 4, and the third at week 8. Each video introduced distinct strategies for managing nicotine addiction, aligned with the participants' progress in the study. Additionally, during their 12-week outpatient visits, they were provided with standardized educational booklets on CBT and MET prepared by the Ministry of Health and Welfare of the Republic of Korea. The first booklet (weeks 0-3) focused on the harmful effects of smoking and the mechanisms of nicotine dependence, the second (weeks 4-7) addressed coping strategies for withdrawal symptoms and cravings and offered relapse prevention techniques, and the third (weeks 8-11) emphasized strategies for sustaining abstinence and managing stress and negative emotions. Printed materials, including educational booklets, smoking cessation diaries, and medication logs, were provided at baseline (week 0), week 4, and week 8. The smoking cessation diaries and medication logs were collected at the subsequent visits (weeks 4, 8, and 12) to monitor participants' adherence and progress. These resources were designed to improve the understanding and handling of nicotine-use disorders.

Assessments and Outcomes

Prior to the commencement of the intervention, participants completed a comprehensive baseline questionnaire designed to collect demographic information and detailed smoking history. Demographic information included age, sex, education level, past and current medical histories, and information on concurrent medications. The questionnaire also inquired about participants' smoking history, including the type of tobacco products, number of cigarettes consumed per day, age at initiation, previous attempts at quitting, and methods previously used in those attempts.

For the primary efficacy indicators of smoking cessation maintenance, assessments were conducted at 2 time points: during treatment at weeks 8 and 12. At each of these time points, both the 7-day and 30-day PPA were assessed. Self-reported smoking abstinence was determined by asking patients whether they had smoked one or more cigarettes in the last 7 and 30 days. All assessments were conducted through verbal self-reports in clinical interviews with medical staff. Additionally, to enhance the validity of these self-reports, a qualitative saliva cotinine test and an exhaled carbon monoxide test, a recognized biological marker of nicotine exposure, were concurrently administered. If the results of these biological tests contradicted the self-reported abstinence, the participants were considered to have failed to maintain smoking cessation.

Secondary efficacy indicators (supplementary indicators) included the Fagerström Test for Nicotine Dependence (FTND) [59], a widely used instrument for evaluating the severity of nicotine addiction. The FTND consists of 6 questions that assess various aspects of nicotine dependence, including the time to the first cigarette after waking up, difficulty of refraining from smoking in nonsmoking areas, and number of cigarettes smoked per day. Each question is scored on a scale, with higher total scores indicating greater nicotine dependence. Additionally, the Stages of Change Readiness and Treatment Eagerness Scale-Smoking (SOCRATES-S) was used to measure smokers'

XSL•FO
RenderX

motivational readiness [60]. It assesses 3 key dimensions of motivation: recognition, ambivalence, and the steps taken. The recognition reflects a smoker's acknowledgment of the problem and the understanding that smoking is detrimental to health. Ambivalence measures the smoker's mixed feelings about quitting, capturing the internal conflict between the desire to quit and fear of change. Taking steps evaluates the proactive efforts that the smoker has already made toward cessation. Each item is scored on a 5-point Likert scale, ranging from "strongly disagree" (1 point) to "strongly agree" (5 points). Subscale scores are calculated by summing the scores of the individual items within each dimension, providing a range of possible scores for each subscale. For example, higher recognition scores indicate a stronger acknowledgment of the problem, while higher ambivalence scores reflect greater internal conflict. Higher scores in the taking steps dimension signify more active efforts toward cessation. The SOCRATES-S score provides insight into where smokers are on the continuum of change, ranging from precontemplation to maintenance, thereby guiding individualized treatment approaches.

Safety assessments for the VR intervention were conducted using the SSQ [61], which evaluates potential adverse effects such as nausea or disorientation, ensuring participant safety in the VR components of the study. The SSQ was administered twice, before and after each VR session, to monitor any changes in symptoms. The SSQ includes 16 symptoms, each rated on a 4-point Likert scale (0=none, 1=slight, 2=moderate, 3=severe). The scores are aggregated into 3 subscales (nausea, oculomotor, and disorientation) and a total severity score using weighted formulas. A cutoff score of 20 points was used in our study; if this threshold was exceeded, VR sessions would have been discontinued, and an adverse event report would have been filed in accordance with our protocol [62,63]. This protocol was established to prioritize participant safety throughout the study.

Data Analysis

In this study, intention-to-treat (ITT) analysis was primarily used to evaluate the primary efficacy outcomes. For the ITT analysis, a total of 30 participants were included, with 15 participants in the DTG group and 15 participants in the BTG group. To handle missing data, the last observation carried forward method was used. However, as the majority of participant dropouts occurred prior to the first follow-up visit (week 4) after enrollment, a per-protocol (PP) analysis was additionally conducted for the secondary efficacy outcomes to assess the treatment effects among those who adhered to the intervention. The PP analysis included only participants who completed the 12-week study protocol, excluding those who dropped out. Specifically, data from participants with complete records at all time points were analyzed, resulting in a total of 21 participants: 9 from the DTG and 12 from the BTG. All statistical analyses were performed using SPSS version 28 (IBM Corp). Descriptive statistics were used to examine the demographic characteristics and smoking or quit-related attributes of participants. Subsequently, at baseline and during treatment weeks 4, 8, and 12, the mean values of efficacy indicators, nicotine use assessments, and motivation to change indicators were compared between the DTG and the BTG to evaluate changes over time. This exploratory clinical study had a small sample size, with only 15 participants assigned to each group. Normality tests indicated that the data did not follow a normal distribution (P<.05), thereby invalidating the assumption of normality. Consequently, nonparametric tests that are appropriate for small sample sizes and do not require the assumption of normality were used [64]. Specifically, statistical significance was assessed using the Wilcoxon signed rank test and the Mann-Whitney U test, which analyzes rank-based data [65,66].

Ethical Considerations

This study was conducted according to the ethical guidelines outlined in the Declaration of Helsinki. Ethical approval for the study protocol was obtained from the Ministry of Food and Drug Safety (number 1271) and the Institutional Review Board of Seoul St. Mary's Hospital (KC21DNSS0706). All participants provided written informed consent after being thoroughly informed about the study's objectives, procedures, risks, and benefits. Participant data were anonymized and securely stored to ensure confidentiality. Identifiable information was removed from the dataset, and only deidentified data were used for analysis. Data access was restricted to authorized researchers. Participants were compensated ₩40,000 (US $28.69) per visit for their participation, amounting to a total of ₩160,000 (US $114.76) for completing the 4 study visits. This compensation was intended to cover transportation and time-related costs and was not contingent upon their smoking cessation outcomes.

Results

Participant Engagement and Safety Outcomes

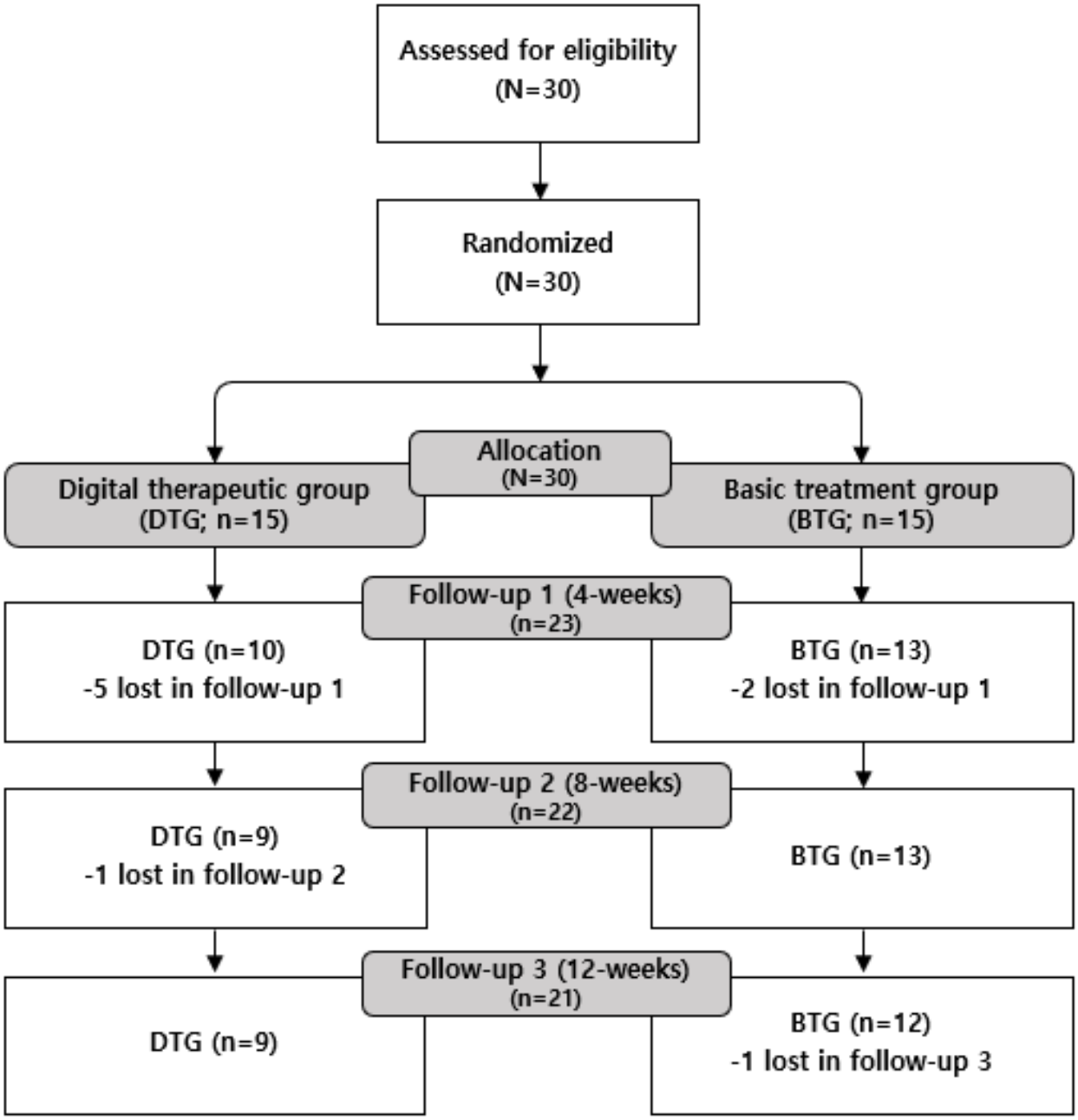
A CONSORT flow diagram is depicted in Figure 2. In this study, 30 participants were randomly assigned to either the DTG (15 participants) or the BTG (15 participants). Both groups completed the baseline visit at week 0 (visit 2). By the fourth week (visit 3), 7 participants had dropped out (5 from the DTG and 2 from the BTG), leaving 23 participants in the study follow-up period (10 in the DTG and 13 in the BTG). By week 8 (visit 4), 1 additional participant from the DTG had dropped out, resulting in 22 participants remaining in the study (9 in the DTG and 13 in the BTG). By the 12th week (visit 5), 1 more participant from the BTG had dropped out, and a total of 21 participants (9 in the DTG and 12 in the BTG) successfully completed the study follow-up period. The overall attrition rate was 30% (9/30), with 40% (6/15) in the DTG and 20% (3/15) in the BTG. The adherence rates for maintaining a smoking cessation diary and medication log were 99% (mean 83 of 84 days) in the DTG (12-week average diary and log adherence rate of 9 participants) and 88% (mean 74 of 84 days) in the BTG (12-week average diary and log adherence rate of 12 participants). These adherence rates were calculated by dividing the number of diary and log entries each participant recorded over the 12-week (84-day) intervention period by the total expected number of entries then averaging these individual adherence rates across participants in each group. This indicates higher treatment adherence in the DTG group. Upon reviewing the DTG's app usage monitoring, the average completion rate of the app-based treatment program was 91% (11 weekly modules of the 12 weeks of modules; 9 participants), while the

XSL•FO
RenderX

average achievement rate of the app treatment program was 98% (12-week average quiz achievement rate by week for 9 participants) based on the proportion of completed quizzes relative to the number of accessed modules).

In this exploratory clinical trial, the SSQ was used to assess the symptoms of motion sickness induced by a virtual environment before and after the use of a VR device. The results indicated that neither the pre- nor postuse scores exceeded the cutoff of 20 and no significant differences were observed between the 2 scores. Additionally, no adverse events related to the use of the VR or smartphone app were reported, confirming the safety of the DTx software NICO-THERA.

Figure 2. Study participant CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



Participant Characteristics at Baseline

The demographic characteristics of the participants are presented in Table 2. The mean ages were 43.07 (SD 12.15) years in the DTG and 48.67 (SD 14.53) years in the BTG. Of the 15 participants in the DTG, there were 11 (73%) men and 4 (27%) women, whereas of the 15 participants in the BTG, there were 12 (80%) men and 3 (20%) women. Regarding educational attainment, the highest proportion of participants (9/15, 60%) had completed a bachelor's degree in both the DTG and BTG.

The average length of time participants had been smoking was 23.32 (SD 11.22) years in the DTG and 29.21 (SD 12.90) years in the BTG. Daily cigarette consumption was similar, with DTG participants smoking an average of 9.53 (SD 8.16) cigarettes per day and BTG participants smoking an average of 9.20 (SD 7.42) cigarettes per day. Regarding smoking cessation

XSL•FO
RenderX

medication, 10 of 15 participants in the DTG (67%) and 8 of 15 participants in the BTG (53%) were using bupropion at baseline. Regarding the type of tobacco used, among the 15 participants of the DTG, 8 participants (53%) used manufactured cigarettes, 2 participants (13%) used heat-not-burn tobacco, 1 participant (7%) used liquid e-cigarettes, and 4 participants (27%) used multiple tobacco products. Among the 15 participants in the BTG, 13 participants (87%) had a higher proportion of manufactured cigarette use compared than the DTG. No participants used heat-not-burn tobacco, 1 participant (1/15, 7%) used liquid e-cigarettes, and 1 participant (1/15, 7%) used multiple tobacco products. The mean FTND scores were 2.80 (SD 2.48) in the DTG and 4.87 (SD 2.62) in the BTG. Exhaled carbon monoxide levels averaged 6.13 (SD 6.21) ppm in the DTG and 6.60 (SD 4.63) ppm in the BTG. Salivary cotinine test results were positive for 11 participants (73%) and negative for 4 participants (27%) of 15 participants in the DTG, whereas 14 participants (93%) tested positive and 1 participant (7%) tested negative out of 15 participants in the DTG.

Table 2. Demographics and baseline characteristics of the participants (N=30).

Characteristics	DTG[a] (n=15)	BTG[b] (n=15)
Age (years), mean (SD)	43.07 (12.15)	48.67 (14.53)
Sex, n (%)		
Male	11 (73)	12 (80)
Female	4 (26)	3 (20)
Education level, n (%)		
Elementary school graduate	0 (0)	1 (7)
High school graduate	4 (27)	2 (13)
Two-year college degree graduate	0 (0)	1 (7)
Bachelor's degree	9 (60)	9 (60)
Graduate degree	2 (13)	2 (13)
Length of smoking (years), mean (SD)	23.32 (11.22)	29.21 (12.90)
Number of cigarettes per day, mean (SD)	9.53 (8.16)	9.20 (7.42)
Medication for smoking cessation, n (%)		
Bupropion	10 (67)	8 (53)
None	5 (33)	7 (47)
Type of tobacco used, n (%)		
Manufactured cigarettes	8 (53)	13 (87)
Heat-not-burn tobacco	2 (13)	0 (0)
Liquid e-cigarettes	1 (7)	1 (7)
Multiple tobacco product use	4 (27)	1 (7)
FTND[c], mean (SD)	2.80 (2.48)	4.87 (2.62)
Exhaled carbon monoxide levels, mean (SD)	6.13 (6.21)	6.60 (4.63)
Salivary cotinine qualitative test results, n (%)		
Positive	11 (73)	14 (93)
Negative	4 (27)	1 (7)
Dropouts, n (%)	6 (40)	3 (20)

[a]DTG: digital therapeutic group.

[b]BTG: basic treatment group.

[c]FTND: Fagerström Test for Nicotine Dependence.

Primary Efficacy Indicators

Initially, the ITT analysis was performed to assess the primary efficacy outcome. At the end of the 12-week study period, 6 of 15 participants (40%) in both the DTG and BTG reported achieving 7-day PPA. Moreover, there were no significant differences between the 2 groups at any time point (week 0: U=97.50, P=.37; week 4: U=90.00, P=.27; week 8: U=112.50, P≥.99; week 12: U=112.50, P≥.99). The mean number of smoking days over a 7-day period at the 12-week mark was 3.27 (SD 3.31) days for the DTG and 3.67 (SD 3.33) days for the BTG, with no statistically significant difference observed.

XSL•FO
RenderX

In addition, the average number of cigarettes smoked over 7 days at the 12-week mark was 36.77 (SD 56.73; range 0-168) cigarettes for the DTG and 28.53 (SD 41.91; range 0-140) cigarettes for the BTG; however, this difference was not significant.

After analyzing the difference in 7-day point abstinence between baseline and each time point after the DTx program in the DTG, there was a statistically significant difference only at the 8-week point (z=–2.00, negative ranks=0, positive ranks=4, P=.046) confirming the efficacy of the DTx program (Table 3). At the 12-week point, the difference before and after the intervention within the DTG was not statistically significant but showed an estimated value relatively close to the significance level (z=–1.73, negative ranks=0, positive ranks=3, P=.08). On the other hand, in the BTG, none of the differences between the baseline and each time point after basic treatment were statistically significant at any time point.

More specifically, changes from baseline in the number of smoking days and the quantity of cigarettes smoked over a 7-day period were analyzed within each group at each time point (Figure 3, Tables 4 and 5). Regarding the number of smoking days over a 7-day period, the DTG demonstrated statistically significant reductions at all post-treatment time points compared with baseline (Figure 3, Table 4), whereas no significant changes were observed in the BTG (Figure 3, Table 4).

Table 3. Intention-to-treat analysis of the difference in 7-day and 30-day point prevalence abstinence before and after treatment, by assigned group.

Evaluation time points, by group		Negative ranks[a], n (mean ranks)	Positive ranks[b], n (mean ranks)	Ties, n	z-score	P value
7-day point prevalence abstinence						
BTG[c] (n=1 5)						
	Week4-week0	0 (0.00)	3 (2.00)	12	–1.73	.08
	Week8-week0	1 (2.50)	3 (2.50)	11	–1.00	.32
	Week12-week0	1 (2.00)	2 (2.00)	12	–0.58	.56
DTG[d] (n= 15)						
	Week4-week0	0 (0.00)	2 (1.50)	13	–1.41	.16
	Week8-week0	0 (0.00)	4 (2.50)	11	–2.00	.046
	Week12-week0	0 (0.00)	3 (2.00)	12	–1.73	.08
30-day point prevalence abstinence [e]						
BTG (n=1 5)						
	Week4-week0	1 (1.50)	1 (1.50)	13	0.00	≥.99
	Week8-week0	1 (2.50)	3 (2.50)	11	–1.00	.32
	Week12-week0	1 (1.50)	1 (1.50)	13	0.00	≥.99
DTG (n= 15)						
	Week4-week0	0 (0.00)	1 (1.00)	14	–1.00	.32
	Week8-week0	0 (0.00)	4 (2.50)	11	–2.00	.046
	Week12-week0	0 (0.00)	3 (2.00)	12	–1.73	.08

[a]Smoking (0) > smoking abstinence (1).

[b]Smoking (0) < smoking abstinence (1).

[c]BTG: basic treatment group.

[d]DTG: digital therapeutic group.

[e]The 30-day point prevalence abstinence (PPA) baseline (week 0) was replaced with the 7-day PPA.

XSL•FO
RenderX

Figure 3. Changes in (A) and (B) over time in the digital therapeutic group (DTG) and basic treatment group (BTG).

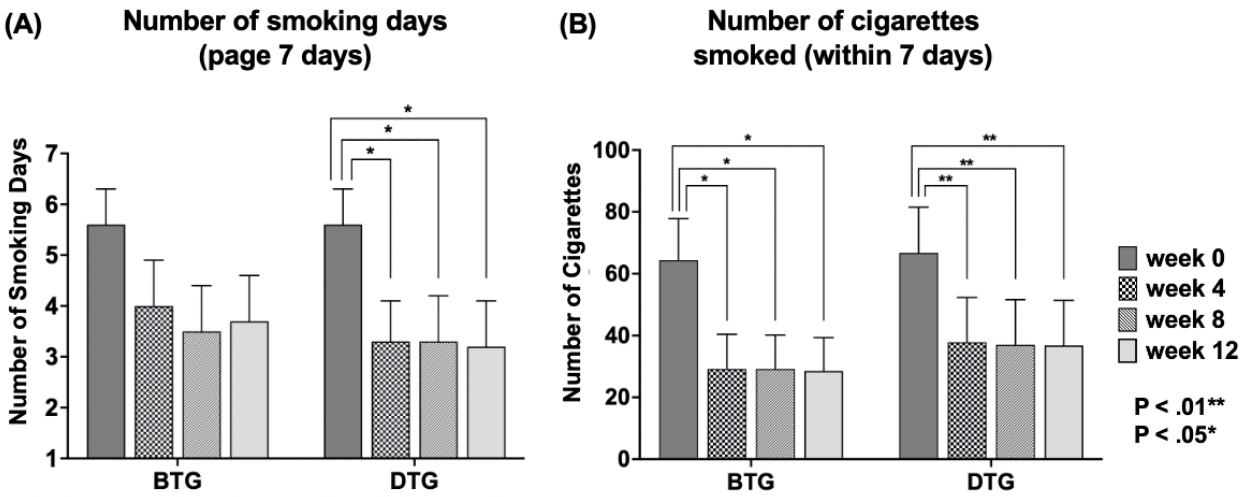


Table 4. Intention-to-treat analysis of the difference in the number of smoking days within 7 days before and after treatment, by assigned group.

Evaluation time points, by group	Negative ranks[a], n (mean ranks)	Positive ranks[b], n (mean ranks)	Ties, n	*z*-score	*P* value
BTG[c] (n=1 5)					
Week4-week0	0 (0.00)	4 (2.50)	11	–1.89	.06
Week8-week0	1 (1.50)	5 (3.90)	9	–1.91	.06
Week12-week0	1 (1.00)	4 (3.50)	10	–1.76	.08
DTG[d] (n= 15)					
Week4-week0	0 (0.00)	6 (3.50)	9	–2.27	.02
Week8-week0	0 (0.00)	7 (4.00)	8	–2.41	.02
Week12-week0	0 (0.00)	7 (4.00)	8	–2.40	.02

[a]Baseline number of smoking days > number of smoking days in weeks 4, 8, and 12.

[b]Baseline number of smoking days < number of smoking days in weeks 4, 8, and 12.

[c]BTG: basic treatment group.

[d]DTG: digital therapeutic group.

Table 5. Intention-to-treat analysis of the difference in the number of cigarettes smoked within 7 days before and after treatment, by assigned group.

Evaluation time points, by group	Negative ranks[a], n (mean ranks)	Positive ranks[b], n (mean ranks)	Ties, n	*z*-score	*P* value
BTG[c] (n=1 5)					
Week4-week0	0 (0.00)	8 (4.50)	7	–2.52	.01
Week8-week0	1 (1.00)	8 (5.50)	6	–2.55	.01
Week12-week0	2 (2.50)	8 (6.25)	5	–2.30	.02
DTG[d] (n= 15)					
Week4-week0	0 (0.00)	8 (4.50)	7	–2.52	.01
Week8-week0	0 (0.00)	8 (4.50)	7	–2.53	.01
Week12-week0	0 (0.00)	8 (4.50)	7	–2.52	.01

[a]Baseline number of smoking days > number of smoking days in weeks 4, 8, and 12.

[b]Baseline number of smoking days < number of smoking days in weeks 4, 8, and 12.

[c]BTG: basic treatment group.

[d]DTG: digital therapeutic group.

XSL•FO
RenderX

Both the DTG and BTG demonstrated statistically significant reductions in the number of cigarettes smoked within 7 days at all time points (Figure 3, Table 5; all $P<.05$). At the 12-week mark, the 30-day PPA rate was 33% (5/15) in the DTG and 27% (4/15) in the BTG, with no significant group differences at any of the time points (week 4: $U=105.00$, $P=.67$; week 8: $U=112.50$, $P\geq.99$; week 12: $U=120.00$, $P=.70$). Additionally, since the number of smoking days and the quantity of cigarettes smoked over a 30-day period were not assessed at baseline, the results were examined using a PP analysis. The mean number of smoking days over a 30-day period at the 12-week mark was 1.89 (SD 2.85) days for the DTG and 11.67 (SD 13.19) days for the BTG, with no statistically significant difference observed. In addition, the average number of cigarettes smoked over 30 days at the 12-week mark was 3.89 (SD 6.45; range 0-16) cigarettes for the DTG and 72.63 (SD 97.94; range 0-300) cigarettes for the BTG, even though this difference in average number of cigarettes smoked between the 2 groups was not statistically significant.

Similar to the 7-day PPA, the 30-day PPA in the DTG also showed statistically significant differences at evaluation points after 8 weeks of the DTx program compared with baseline (Table 3; $z=-2.00$, negative ranks=0, positive ranks=4, $P=.046$). Additionally, at the 12-week point, the difference before and after the intervention in the 30-day PPA in the DTG was not statistically significant but showed an estimated value relatively close to the significance level ($z=-1.73$, negative ranks=0, positive ranks=3, $P=.08$). This indicates that the DTx program was effective at achieving not only 7-day smoking abstinence but also 30-day smoking abstinence. On the other hand, in the BTG, none of the differences between the baseline and each time point after basic treatment were statistically significant.

Consistency Between Self-Reported Results and Biological Test Results

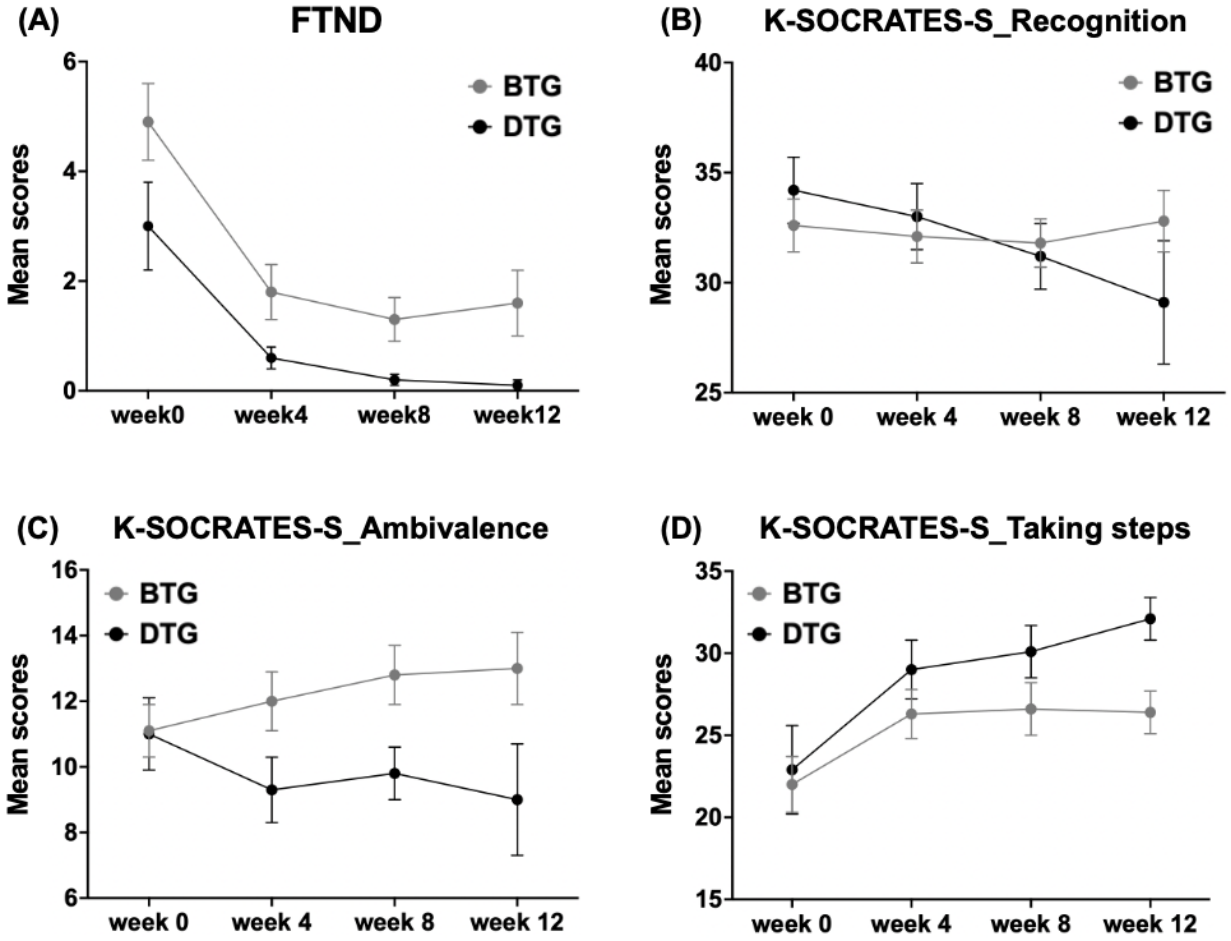
Of all 21 participants, the discordance rate between the qualitative saliva cotinine test results and the patients' verbal self-reports was 13% (n=4) at week 4, 3% (n=1) at week 8, and 3% (n=1) at week 12. When examining the discordance rates, no significant differences were observed between the groups across the weeks. At the 4-week assessment, false negatives included participants who reported smoking over the past 30 days but reported abstinence in the past 7 days. Specifically, their exhaled carbon monoxide levels were all found to be 0 ppm or 1 ppm, indicating a nonsmoker level (≤7 ppm). On average, these false-negative participants had smoked for 6 (range 1-12) days in the past 30 days, with an average daily consumption of 5 (range 1.5-15) cigarettes. In cases of false negatives, participants reported exposure to secondhand smoke or were likely owing to a small amount of smoking at least 7 days prior. Their exhaled carbon monoxide levels were 5 ppm and 7 ppm, respectively, corresponding to nonsmoker levels. Overall, the concordance rate between self-reports and the saliva cotinine qualitative test was approximately 86% to 97%, indicating a high level of reliability for the self-reported data. Detailed concordance rates at each assessment time point are presented in Multimedia Appendix 5.

Secondary Efficacy Indicators

To evaluate the treatment effect, a PP analysis was conducted for the secondary efficacy outcomes. The FTND analysis indicated no statistically significant differences between the groups prior to treatment, and a similar pattern was observed following the program, with the results remaining comparable to those observed before treatment (Figure 4, Table 6).

XSL•FO
RenderX

Figure 4. Changes in nicotine dependence and motivation for change over time in the digital therapeutic group (DTG) and basic treatment group (BTG) as shown in (A)–(D).



The results of comparing the subfactor scores of the K-SOCRATES-S to examine the motivation to change smoking cessation between the DTG and BTG are shown in Figure 4 and Table 6. There was no significant difference in motivation recognition between the DTG and BTG owing to the random assignment (adjusted *P* values: week 0: *P*=.71; week 4: *P*=.77; week 8: *P*=.77; week 12: *P*=.71). Although no statistically significant difference was observed in motivation-ambivalence, a trend approaching significance was observed from week 4 (*U*=28.50; DTG=9, BTG=12; adjusted *P*=.09) to week 12 (*U*=28.50; DTG=9, BTG=12; adjusted *P*=.09), with differences nearing meaningful levels compared with baseline. Specifically, BTG participants consistently maintained ambivalence levels slightly higher than their baseline, whereas DTG participants tended to exhibit slightly lower levels of ambivalence compared with their baseline. Furthermore, there was a significant group difference at 12 weeks (*U*=19.00; DTG n=9, BTG n=12; adjusted *P*=.048) to the motivation-taking steps.

XSL•FO
RenderX

Table 6. Per-protocol comparative analysis results of supplementary indicators between the digital therapeutic group (DTG) and basic treatment group (BTG).

Assessment by week and group	Mean rank	Sum of ranks	*U*	*P* value	Adjusted *P* value[a]
FTND[b]					
Week 0			30.50	.09	.09
BTG (n=12)	12.96	155.50			
DTG (n=9)	8.39	75.50			
Week 4			29.00	.05	.07
BTG (n=12)	13.08	157.00			
DTG (n=9)	8.22	74.00			
Week 8			26.00	.03	.07
BTG (n=12)	13.33	160.00			
DTG (n=9)	7.89	71.00			
Week 12			31.00	.05	.07
BTG (n=12)	12.92	155.00			
DTG (n=9)	8.44	76.00			
K-SOCRATES-S [c] -recognition					
Week 0			41.00	.35	.71
BTG (n=12)	9.92	119.00			
DTG (n=9)	12.44	112.00			
Week 4			50.00	.77	.77
BTG (n=12)	10.67	128.00			
DTG (n=9)	11.44	103.00			
Week 8			49.50	.75	.77
BTG (n=12)	11.38	136.50			
DTG (n=9)	10.50	94.50			
Week 12			37.50	.24	.71
BTG (n=12)	12.38	148.50			
DTG (n=9)	9.17	82.50			
K-SOCRATES-S [c] -a mbivalence					
Week 0			45.00	.51	.51
BTG (n=12)	11.75	141.00			
DTG (n=9)	10.00	90.00			
Week 4			28.50	.07	.09
BTG (n=12)	13.13	157.50			
DTG (n=9)	8.17	73.50			
Week 8			23.50	.03	.09
BTG (n=12)	13.54	162.50			
DTG (n=9)	7.61	68.50			
Week 12			28.50	.07	.09
BTG (n=12)	13.13	157.50			
DTG (n=9)	8.17	73.50			
K-SOCRATES-S-t aking steps					

XSL•FO
RenderX

Assessment by week and group	Mean rank	Sum of ranks	*U*	*P* value	Adjusted *P* value[a]
Week 0			48.50	.70	.70
BTG (n=12)	10.54	126.50			
DTG (n=9)	11.61	104.50			
Week 4			38.00	.25	.34
BTG (n=12)	9.67	116.00			
DTG (n=9)	12.78	115.00			
Week 8			35.50	.19	.34
BTG (n=12)	15.70	235.50			
DTG (n=9)	15.30	229.50			
Week 12			19.00	.01	.048
BTG (n=12)	14.47	217.00			
DTG (n=9)	16.53	248.00			

[a]Adjusted using the Benjamini-Hochberg method.

[b]FTND: Fagerström Test for Nicotine Dependence.

[c]K-SOCRATES-S: Korean Stages of Change Readiness and Treatment Eagerness Scale-Smoking.

Discussion

Principal Findings

This exploratory clinical trial aimed to evaluate the safety and preliminary efficacy of NICO-THERA, a DTx software program developed to alleviate nicotine dependence symptoms in patients diagnosed with nicotine use disorder.

Nicotine dependence is a chronic condition with a high relapse rate necessitating long-term and consistent therapeutic management. Traditional treatments for nicotine addiction typically involve medication and face-to-face CBT sessions. However, adherence to prescribed medications is often suboptimal, which reduces the efficacy of pharmacotherapy. Additionally, CBT sessions require considerable time and financial investment from the patients, which can be a significant barrier. In practice, the national smoking cessation support services in South Korea are typically available only during working hours, which limits physical and temporal accessibility. Although available, online services are typically limited to one-time consultations via telephone or messaging. Consequently, many smokers struggle to maintain their motivation and commitment to quit outside the clinical settings in their daily lives. The number of smokers utilizing national smoking cessation services in South Korea is steadily declining, with only approximately 30% completing the program [67,68]. These challenges highlight the need for alternative strategies to address the limitations of traditional smoking cessation treatments in the digital age.

In this context, DTx, such as NICO-THERA, could play a pivotal role as an adjunct to existing smoking cessation treatments. Specifically, these interventions can be integrated into daily life by providing continuous support, along with national smoking cessation programs as a complementary method. As a digital therapy based on MET and CBT, NICO-THERA has the potential to enhance treatment accessibility for nicotine addiction and advance patient-centered participatory medicine. It can also assist health care providers with monitoring patients' condition daily and incorporating this information into treatment plans. In this study, both the DTG and BTG participants followed identical outpatient visit schedules and pharmacotherapy regimens. However, differences in the interventions implemented in each group resulted in varying adherence rates, which may have contributed to observed differences in smoking cessation outcomes between the 2 groups. In the DTG, relatively favorable treatment adherence was observed among participants who completed the program, and no adverse device effects were reported. These findings suggest that NICO-THERA can be regarded as a safe medical device with the potential to contribute to smoking cessation. Several previous studies reported on the effectiveness of pharmacotherapy or behavioral interventions for smoking cessation, emphasizing the combination of both as clinical guidelines for successfully providing smoking cessation treatment [10-13,69]. In fact, a meta-analysis reported that combining pharmacotherapy with intensive behavioral support significantly increases the likelihood of successful smoking cessation compared with providing simple advice or support [13,58]. This suggests that NICO-THERA, as a digital therapy, has the potential to enhance the effectiveness of smoking cessation treatments by supplementing pharmacotherapy and counseling.

NICO-THERA is a 12-week DTx that systematically integrates a mobile app and VR to promote smoking cessation based on the stages of change model [55]. In the DTx program, the participants engaged in VR content for relaxation training, craving coping training, and refusal training in clinical settings. Additionally, they undertook the smoking cessation program, including cognitive restructuring, refusal skills, and coping with negative emotions through the app. Regarding primary efficacy, the 7-day and 30-day PPA of the DTG significantly improved at week 8. Additionally, significant reductions in both the

XSL•FO
RenderX

number of smoking days and the number of cigarettes smoked were observed at all time points compared with baseline in the DTG. In the BTG, despite a significant reduction in cigarette consumption at weeks 4, 8, and 12, no statistically significant improvements were found in the number of smoking days and in abstinence rates. Although the group differences at each 4-week interval were not statistically significant due to the small sample size, the improvement in abstinence observed in the DTG group may be attributable to the DTx program. In this study, we identified a series of positive behavioral changes among DTG participants—including increased motivation to quit smoking—that aligned with their learning of craving management skills and strategies for overcoming cognitive distortions from week 4 to week 12. Also, from week 4 to week 12, no significant differences in nicotine dependence were observed between the DTG and BTG participants, with levels remaining comparable before and after treatment. However, the DTG exhibited a consistently significant reduction in both the number of smoking days and the number of cigarettes smoked, compared with the BTG. Therefore, by acquiring and repeatedly practicing cognitive, behavioral, and emotional coping strategies conducive to smoking cessation, participants were able to establish and consolidate resources to effectively apply coping skills. The DTx intervention may have contributed to maintaining the stability of both physical and psychological dependence on nicotine, thereby preventing its deterioration and ultimately supporting a reduction in nicotine consumption.

The progressive decline in ambivalence from week 4 and the significant increase in taking steps at week 12 suggest that the DTG participants exhibited a progressive change in their motivation to quit smoking. The initial 1 to 4 weeks of the DTx program included a series of content focused on MET approaches, such as addiction education, addressing ambivalent emotions, creating supportive environments, and understanding confidence in quitting smoking. From week 4, they demonstrated an improved ability to resolve ambivalence and strengthen their commitment to change, thereby enhancing their motivation. By week 12, DTG participants had solidified their resolve to sustain smoking cessation, further demonstrating the effectiveness of the program's approach. Notably, at week 8, the DTG exhibited a significant improvement in the primary efficacy outcome, abstinence, compared with baseline, with this effect being relatively sustained through week 12. In the 8 to 12 weeks of the program, the emphasis shifted to consolidating and maintaining the behavioral skills acquired during earlier phases, with a focus on training participants to manage not only cravings but also emotional regulation strategies. These strategies aim to address factors such as depression, anxiety, and anger, which may influence smoking behavior. As a result, the impact of the intervention may have become most evident by week 8, with the remainder of the program serving to reinforce and sustain these changes. This pattern is consistent with those found in previous studies that demonstrate a progression of motivation and behavioral change from the precontemplation stage to the contemplation stage then to the preparation stage and finally the action stage [55-57]. In contrast, the BTG maintained a similar level of recognition, whereas the scores for ambivalence and taking steps gradually increased. This pattern suggests that the BTG remained at the precontemplation stage and contemplation stage without fully transitioning to the action stage, likely owing to unresolved ambivalence and ongoing conflict between the cognitive and behavioral dimensions [57]. Smoking cessation is not merely a matter of short-term efforts; it requires ongoing management to ensure long-term success. NICO-THERA, as a software that combines app and VR components, provides cognitive, behavioral, and emotional strategies designed to support smoking cessation across various stages of behavioral change. Notably, VR allows participants to engage in motivational, cognitive, and behavioral training in clinical settings, helping them develop strategies to cope with smoking triggers. The smartphone app supports self-directed therapy and habit formation in daily life through its content and daily functions. As a digital therapy based on MET and CBT, NICO-THERA has the potential to enhance the effectiveness of smoking cessation treatments by expanding combined DTx interventions and improving accessibility.

In summary, the results demonstrated that adults who participated in NICO-THERA had a significantly higher success rate for smoking abstinence. Additionally, compared with those who received basic treatment, participants in the DTx program exhibited higher motivation to quit smoking. Participants in the DTx program exhibited adaptive changes conducive to smoking cessation consistent with the weekly intervention strategies of the MET and CBT. Thus, the efficacy of the DTx program was validated, suggesting its greater potential to facilitate successful smoking cessation than basic treatment.

Limitations

This study has some limitations. As a pilot study, the sample size was kept small, limiting the statistical power to detect significant differences in some analyses.

The small sample size (30 participants) necessitated the use of nonparametric tests, which affected the generalizability of the results. In addition, there were limitations in analyzing the impact of medication and counseling adherence on the efficacy of DTx. Future studies with larger sample sizes are required to confirm the observed trends and to establish statistical significance through parametric tests. Additionally, the open-label nature of the study, in which both participants and researchers were aware of group assignments, may have introduced bias in outcome reporting.

The DTG had a higher dropout rate than the BTG. Specifically, 5 participants in the DTG dropped out in week 4; notably, 3 of them expressed a willingness to continue but were unable to visit the hospital due to scheduling issues. NICO-THERA not only required in-person VR sessions but also involved relatively greater time demands for device setup and execution compared with the passive activity of video viewing. Therefore, limited device and temporal accessibility may have contributed to the higher dropout rate. In the future, developing portable or home-based VR devices could help mitigate this limitation by enabling patients to receive treatment without being restricted by the temporal constraints of hospital visits.

Future directions for VR in smoking cessation may involve more personalized VR experiences that adapt to users' physiological and psychological states in real time, potentially

XSL•FO
RenderX

through the integration of biofeedback mechanisms. As VR technology continues to evolve, such personalized interventions could allow users to engage in dynamically responsive therapeutic environments that match their stress levels, craving intensity, or motivational readiness.

In this study, NICO-THERA enabled a degree of personalization by allowing participants to choose among VR scenarios tailored for relaxation, craving coping, and refusal training. This scenario selection was based on the user's preferences or situational relevance, offering flexibility within the intervention. However, the program did not yet incorporate adaptive customization based on the participants' stage of change or real-time physiological input. We acknowledge this as a limitation and propose it as a valuable direction for future development.

Building on this pilot study, future confirmatory trials should aim to address the limitations identified herein, particularly those related to accessibility and personalization. By refining the VR delivery method and enhancing the adaptability of mobile and VR content, the NICO-THERA intervention may further improve its clinical efficacy and user engagement.

Conclusion

We identified the preliminary safety and efficacy of NICO-THERA, a DTx combining MET and CBT that has the potential to improve smoking cessation outcomes. By incorporating CBT and MET content into a sequenced digital format and linking each weekly module with a corresponding VR session, NICO-THERA offers a distinctive treatment flow that contrasts with mobile-only interventions primarily focused on information delivery or tracking tools. Participants in the DTG experienced significant reductions in nicotine dependence and maintained a stronger practical motivation for quitting smoking. These results suggest that integrating immersive behavioral training with structured cognitive strategies may enhance engagement and support core behavioral change in smoking cessation. It has been suggested that NICO-THERA may serve as a valuable complement to traditional pharmacotherapy and counseling in smoking cessation programs. Although the statistical significance of the group differences in smoking cessation rates could not be confirmed due to the limitations of the sample size, the smoking cessation rate of the DTG group increased significantly at week 8, as compared with baseline. Future research should confirm these findings in larger clinical trials and explore the long-term efficacy of integrating VR-based interventions into smoking cessation programs. Additionally, studies should assess their scalability across diverse clinical and community settings.

Acknowledgments

This work was supported by a grant from the Korea Evaluation Institute of Industrial Technology (KEIT) and the Ministry of Trade, Industry and Energy, Republic of Korea (Development of CBT-based Digital Therapeutics for Alcohol and Nicotine Addiction Disorder, Study Number 20009227). This research was supported by a grant from the Mental Health related Social Problem-Solving Project, Ministry of Health & Welfare, Republic of Korea (RS-2024-00403474). This research was supported by the Bio&Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT; RS-2023-00223559).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Background settings for image relaxation training through virtual reality (VR), presented sequentially: forest, campfire, beach.
[PNG File , 345 KB - mhealth_v13i1e66411_app1.png]

Multimedia Appendix 2
Virtual reality (VR)-based training session for craving coping: A concept image from the storyboard of the VR content illustrating a situation at a social gathering where friends are stepping out to smoke.
[PNG File , 886 KB - mhealth_v13i1e66411_app2.png]

Multimedia Appendix 3
Virtual reality (VR)-based training session for refusal training: A concept image from the storyboard of the VR content depicting a situation where a friend offers a cigarette to someone who is in the process of quitting smoking.
[PNG File , 960 KB - mhealth_v13i1e66411_app3.png]

Multimedia Appendix 4
App screenshots of the digital therapeutic intervention NICO-THERA, presented sequentially from left to right: Smoking cessation therapy contents, Smoking cessation diary and medication log, and a graph displaying the number of cigarettes smoked and changes in motivation levels.
[PNG File , 170 KB - mhealth_v13i1e66411_app4.png]

XSL•FO
RenderX

Multimedia Appendix 5
Concordance rate between self-reported and saliva cotinine qualitative test results.
[DOCX File , 15 KB - mhealth_v13i1e66411_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1956 KB - mhealth_v13i1e66411_app6.pdf]

References

1. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1223-1249 [FREE Full text] [doi: 10.1016/S0140-6736(20)30752-2] [Medline: 33069327]
2. Global Burden of Disease 2021: Findings from the GBD 2021 Study. IHME. URL: https://www.healthdata.org/research-analysis/library/global-burden-disease-2021-findings-gbd-2021-study [accessed 2025-09-15]
3. Health at a Glance 2023: OECD Indicators. OECD. 2023 Nov 07. URL: https://www.oecd.org/en/publications/health-at-a-glance-2023_7a7afb35-en.html [accessed 2025-09-15]
4. WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke. World Health Organization. 2023 Jul 31. URL: https://www.who.int/publications/i/item/9789240077164 [accessed 2025-09-15]
5. Oh S. The current status and challenges of national smoking cessation support program in Korea. Korean J Fam Med 2019 Nov;40(6):351-352. [doi: 10.4082/kjfm.40.6e]
6. Lee S, Kimm H, Yun JE, Jee SH. Public health challenges of electronic cigarettes in South Korea. J Prev Med Public Health 2011 Nov;44(6):235-241 [FREE Full text] [doi: 10.3961/jpmph.2011.44.6.235] [Medline: 22143173]
7. Steinberg MB, Schmelzer AC, Lin PN, Garcia G. Smoking as a chronic disease. Curr Cardio Risk Rep 2010 Aug 29;4(6):413-420. [doi: 10.1007/s12170-010-0125-5]
8. Shiffman S, West R, Gilbert D, SRNT Work Group on the Assessment of Craving and Withdrawal in Clinical Trials. Recommendation for the assessment of tobacco craving and withdrawal in smoking cessation trials. Nicotine Tob Res 2004 Aug;6(4):599-614. [doi: 10.1080/14622200410001734067] [Medline: 15370156]
9. Hughes JR, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 2004 Jan;99(1):29-38. [doi: 10.1111/j.1360-0443.2004.00540.x] [Medline: 14678060]
10. Barua RS, Rigotti NA, Benowitz NL, Cummings KM, Jazayeri M, Morris PB, et al. 2018 ACC Expert Consensus Decision Pathway on Tobacco Cessation Treatment: a report of the American College of Cardiology Task Force on Clinical Expert Consensus Documents. J Am Coll Cardiol 2018 Dec 25;72(25):3332-3365 [FREE Full text] [doi: 10.1016/j.jacc.2018.10.027] [Medline: 30527452]
11. Stead L, Koilpillai P, Fanshawe T, Lancaster T. Combined pharmacotherapy and behavioural interventions for smoking cessation. Cochrane Database Syst Rev 2016 Mar 24;3(3):1-96 [FREE Full text] [doi: 10.1002/14651858.CD008286.pub3] [Medline: 27009521]
12. Lemmens V, Oenema A, Knut IK, Brug J. Effectiveness of smoking cessation interventions among adults: a systematic review of reviews. Eur J Cancer Prev 2008 Nov;17(6):535-544 [FREE Full text] [doi: 10.1097/CEJ.0b013e3282f75e48] [Medline: 18941375]
13. Onwuzo CN, Olukorode J, Sange W, Orimoloye DA, Udojike C, Omoragbon L, et al. A review of smoking cessation interventions: efficacy, strategies for implementation, and future directions. Cureus 2024 Jan;16(1):e52102 [FREE Full text] [doi: 10.7759/cureus.52102] [Medline: 38344627]
14. Müller F, Wehbe L. Smoking and smoking cessation in Latin America: a review of the current situation and available treatments. Int J Chron Obstruct Pulmon Dis 2008;3(2):285-293 [FREE Full text] [doi: 10.2147/copd.s2654] [Medline: 18686737]
15. Tobacco: Preventing Uptake, Promoting Quitting and Treating Dependence. National Institute for Health and Care Excellence (NICE). 2023 Jan 16. URL: https://www.nice.org.uk/guidance/ng209 [accessed 2025-09-15]
16. New Zealand Guidelines for Helping People to Stop Smoking 2021. Ministry of Health. 2021. URL: https://www.health.govt.nz/system/files/2014-06/the-new-zealand-guidelines-for-helping-people-to-stop-smoking-2021.pdf [accessed 2025-09-15]
17. Risk estimation and the prevention of cardiovascular disease. Scottish Intercollegiate Guidelines Network. 2017 Jun. URL: https://www.sign.ac.uk/assets/sign149.pdf [accessed 2025-09-15]
18. US Preventive Services Task Force, Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, et al. Interventions for tobacco smoking cessation in adults, including pregnant persons: US Preventive Services Task Force recommendation statement. JAMA 2021 Jan 19;325(3):265-279. [doi: 10.1001/jama.2020.25019] [Medline: 33464343]
19. Supporting Smoking Cessation: A Guide for Health Professionals. The Royal Australian College of General Practitioners. 2021 Sep 29. URL: https://www.racgp.org.au/clinical-resources/clinical-guidelines/key-racgp-guidelines/view-all-racgp-guidelines/supporting-smoking-cessation [accessed 2025-09-15]

XSL•FO
RenderX

20. Jamison J, Naughton F, Gilbert H, Sutton S. Delivering smoking cessation support by mobile phone text message: What information do smokers want? A focus group study. J Appl Biobehavioral Res 2013 Mar 08;18(1):1-23 [FREE Full text] [doi: 10.1111/jabr.12004]
21. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
22. Keoleian V, Polcin D, Galloway GP. Text messaging for addiction: a review. J Psychoactive Drugs 2015 May 07;47(2):158-176 [FREE Full text] [doi: 10.1080/02791072.2015.1009200] [Medline: 25950596]
23. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS Text Message Interventions for Smoking Cessation: A Meta-Analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
24. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
25. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
26. Seidman DF, Westmaas JL, Goldband S, Rabius V, Katkin ES, Pike KJ, et al. Randomized controlled trial of an interactive internet smoking cessation program with long-term follow-up. Ann Behav Med 2010 Feb 23;39(1):48-60. [doi: 10.1007/s12160-010-9167-7] [Medline: 20177844]
27. Ben-Zeev D. Mobile technologies in the study, assessment, and treatment of schizophrenia. Schizophr Bull 2012 May;38(3):384-385 [FREE Full text] [doi: 10.1093/schbul/sbr179] [Medline: 22234487]
28. Browne J, Halverson TF, Vilardaga R. Engagement with a digital therapeutic for smoking cessation designed for persons with psychiatric illness fully mediates smoking outcomes in a pilot randomized controlled trial. Transl Behav Med 2021 Sep 15;11(9):1717-1725 [FREE Full text] [doi: 10.1093/tbm/ibab100] [Medline: 34347865]
29. Hong J, Wasden C, Han D. Introduction of digital therapeutics. Comput Methods Programs Biomed 2021 Sep;209:106319 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106319] [Medline: 34364181]
30. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
31. Masaki K, Tateno H, Kameyama N, Morino E, Watanabe R, Sekine K, et al. Impact of a novel smartphone app (CureApp Smoking Cessation) on nicotine dependence: prospective single-arm interventional pilot study. JMIR Mhealth Uhealth 2019 Feb 19;7(2):e12694 [FREE Full text] [doi: 10.2196/12694] [Medline: 30777848]
32. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. Am J Drug Alcohol Abuse 2015 Jan;41(1):68-73 [FREE Full text] [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]
33. Tudor-Sfetea C, Rabee R, Najim M, Amin N, Chadha M, Jain M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. JMIR Mhealth Uhealth 2018 Apr 18;6(4):e98 [FREE Full text] [doi: 10.2196/mhealth.9405] [Medline: 29669708]
34. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective behavioral changes through a digital mHealth app: exploring the impact of hedonic well-being, psychological empowerment and inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024 [FREE Full text] [doi: 10.2196/10024] [Medline: 29907557]
35. Newton A, Bagnell A, Rosychuk R, Duguay J, Wozney L, Huguet A, et al. A mobile phone-based app for use during cognitive behavioral therapy for adolescents with anxiety (MindClimb): user-centered design and usability study. JMIR Mhealth Uhealth 2020 Dec 08;8(12):e18439 [FREE Full text] [doi: 10.2196/18439] [Medline: 33289671]
36. Dascal J, Reid M, IsHak WW, Spiegel B, Recacho J, Rosen B, et al. Virtual reality and medical inpatients: a systematic review of randomized, controlled trials. Innov Clin Neurosci 2017;14(1-2):14-21 [FREE Full text] [Medline: 28386517]
37. Botella C, Villa H, Baños R, Perpiñá C, García-Palacios A. The treatment of claustrophobia with virtual reality: changes in other phobic behaviors not specifically treated. Cyberpsychol Behav 1999;2(2):135-141. [doi: 10.1089/cpb.1999.2.135] [Medline: 19178249]
38. Botella C, Baños R, Villa H, Perpiñá C, García-Palacios A. Virtual reality in the treatment of claustrophobic fear: a controlled, multiple-baseline design. Behavior Therapy 2000 Feb;31(3):583-595 [FREE Full text] [doi: 10.1016/s0005-7894(00)80032-5]
39. Heo S, Park J. Effects of virtual reality-based graded exposure therapy on PTSD symptoms: a systematic review and meta-analysis. Int J Environ Res Public Health 2022 Nov 29;19(23):1 [FREE Full text] [doi: 10.3390/ijerph192315911] [Medline: 36497989]
40. Pericot-Valverde I, Secades-Villa R, Gutiérrez-Maldonado J. A randomized clinical trial of cue exposure treatment through virtual reality for smoking cessation. J Subst Abuse Treat 2019 Jan;96:26-32 [FREE Full text] [doi: 10.1016/j.jsat.2018.10.003] [Medline: 30466545]

XSL•FO
RenderX

41. Goldenhersch E, Thrul J, Ungaretti J, Rosencovich N, Waitman C, Ceberio MR. Virtual reality smartphone-based intervention for smoking cessation: pilot randomized controlled trial on initial clinical efficacy and adherence. J Med Internet Res 2020 Jul 29;22(7):e17571 [FREE Full text] [doi: 10.2196/17571] [Medline: 32723722]
42. Coons MJ, Roehrig M, Spring B. The potential of virtual reality technologies to improve adherence to weight loss behaviors. J Diabetes Sci Technol 2011 Mar 01;5(2):340-344 [FREE Full text] [doi: 10.1177/193229681100500221] [Medline: 21527103]
43. Browne RH. On the use of a pilot sample for sample size determination. Stat Med 1995 Sep 15;14(17):1933-1940. [doi: 10.1002/sim.4780141709] [Medline: 8532986]
44. Julious S. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceutical Statistics 2005 Nov 24;4(4):287-291 [FREE Full text] [doi: 10.1002/pst.185]
45. Athilingam P, Jenkins B, Johansson M, Labrador M. A mobile health intervention to improve self-care in patients with heart failure: pilot randomized control trial. JMIR Cardio 2017 Aug 11;1(2):e3 [FREE Full text] [doi: 10.2196/cardio.7848] [Medline: 31758759]
46. Hu L, Shi Y, Wylie-Rosett J, Sevick MA, Xu X, Lieu R, et al. Feasibility of a family-oriented mHealth intervention for Chinese Americans with type 2 diabetes: a pilot randomized control trial. PLoS One 2024 Mar 11;19(3):e0299799 [FREE Full text] [doi: 10.1371/journal.pone.0299799] [Medline: 38466714]
47. Trojak B, Meille V, Achab S, Lalanne L, Poquet H, Ponavoy E, et al. Transcranial magnetic stimulation combined with nicotine replacement therapy for smoking cessation: a randomized controlled trial. Brain Stimul 2015 Nov;8(6):1168-1174. [doi: 10.1016/j.brs.2015.06.004] [Medline: 26590478]
48. Kitsiou S, Gerber BS, Buchholz SW, Kansal MM, Sun J, Pressler SJ. Patient-centered mHealth intervention to improve self-care in patients with chronic heart failure: phase 1 randomized controlled trial. J Med Internet Res 2025 Jan 15;27:e55586 [FREE Full text] [doi: 10.2196/55586] [Medline: 39813671]
49. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). In: American Psychiatric Association. Washington, DC: American Psychiatric Association Publishing; 2013.
50. Achterberg J. Imagery in Healing. Boulder, CO: Shambhala Publishing; 2002.
51. Davis M, Eshelman ER, McKay M. Progressive Muscle Relaxation: The Relaxation and Stress Reduction Workbook Chapter Singles. Oakland, CA: New Harbinger Publications; 2008.
52. Conklin CA, Tiffany ST. Applying extinction research and theory to cue-exposure addiction treatments. Addiction 2002 Feb;97(2):155-167. [doi: 10.1046/j.1360-0443.2002.00014.x] [Medline: 11860387]
53. Culbertson CS, Shulenberger S, De La Garza R, Newton TF, Brody AL. Virtual reality cue exposure therapy for the treatment of tobacco dependence. J Cyber Ther Rehabil 2012;5(1):57-64 [FREE Full text] [Medline: 25342999]
54. Carroll KM. Therapy Manuals for Drug Addiction: Manual One: Cognitive-Behavioral Approach: Treating Cocaine Addiction. Internet Archive. 1998. URL: https://archive.org/details/manual1cbt [accessed 2025-09-15]
55. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. Journal of Consulting and Clinical Psychology 1983;51(3):390-395. [doi: 10.1037/0022-006X.51.3.390]
56. Prochaska JO, Johnson S, Lee P. The transtheoretical model of behavior change. In: Ockene JK, Riekert KA, Shumaker SA, editors. The Handbook of Health Behavior Change, Third Edition. New York, NY: Springer Publishing Company; 2009:59-83.
57. Hsu C, Liao H, Huang L. Exploring smoking cessation behaviors of outpatients in outpatient clinics: application of the transtheoretical model. Medicine (Baltimore) 2020 Jul 02;99(27):e20971 [FREE Full text] [doi: 10.1097/MD.0000000000020971] [Medline: 32629709]
58. Sussman S, Sun P, Dent CW. A meta-analysis of teen cigarette smoking cessation. Health Psychol 2006 Sep;25(5):549-557. [doi: 10.1037/0278-6133.25.5.549] [Medline: 17014271]
59. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127 [FREE Full text] [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
60. Park JW, Park KH, Joe KH, Choi SH, Lee IJ, Hwang JH, et al. Standardization study of the Korean version of the Stages of Change Readiness and Treatment Eagerness Scale for Smoking Cessation (K-SOCRATES-S) and its predictive validity. Psychiatry Investig 2012 Sep;9(3):223-228 [FREE Full text] [doi: 10.4306/pi.2012.9.3.223] [Medline: 22993520]
61. Biernacki MP, Kennedy RS, Dziuda. Simulator sickness and its measurement with Simulator Sickness Questionnaire (SSQ). Med Pr 2016;67(4):545-555 [FREE Full text] [doi: 10.13075/mp.5893.00512] [Medline: 27623835]
62. Stanney KM, Kennedy RS, Drexler JM. Cybersickness is not simulator sickness. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 1997 Oct 01;41(2):1138-1142. [doi: 10.1177/107118139704100292]
63. Bimberg P, Weissker T, Kulik A. On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research. 2020 Presented at: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); March 22-26, 2020; Atlanta, GA. [doi: 10.1109/vrw50115.2020.00098]
64. Conover WJ. Practical Nonparametric Statistics. Hoboken, NJ: John Wiley & Sons Inc; 1999.
65. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bulletin 1945 Dec;1(6):80. [doi: 10.2307/3001968]

XSL•FO
RenderX

66. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist 1947 Mar;18(1):50-60. [doi: 10.1214/aoms/1177730491]
67. Smoking Cessation Project Invested 263.1 Billion KRW Over 5 Years, But Completion Rate Remains at 30%. Medipharm Health. 2023 Oct 09. URL: https://www.medipharmhealth.co.kr/mobile/article.html?no=94415 [accessed 2025-09-15]
68. Limited Effectiveness of Smoking Cessation Programs in South Korea. 2Firsts News. 2023 Oct 08. URL: https://www.2firsts.com/news/limited-effectiveness-of-smoking-cessation-programs-in-south-korea [accessed 2025-09-15]
69. Hartmann-Boyce J, Hong B, Livingstone-Banks J, Wheat H, Fanshawe TR. Additional behavioural support as an adjunct to pharmacotherapy for smoking cessation. Cochrane Database Syst Rev 2019 Jun 05;6(6):CD009670 [FREE Full text] [doi: 10.1002/14651858.CD009670.pub4] [Medline: 31166007]

Abbreviations

BTG: basic treatment group
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
DTG: digital therapeutic group
DTx: digital therapeutic
FTND: Fagerström Test for Nicotine Dependence
ITT: intention to treat
K-SOCRATES-S: Korean Stages of Change Readiness and Treatment Eagerness Scale-Smoking
MET: motivational enhancement therapy
mHealth: mobile health
OECD: Organisation for Economic Co-operation and Development
PP: per-protocol
PPA: point prevalence abstinence
RCT: randomized controlled trial
SSQ: Simulator Sickness Questionnaire
VR: virtual reality

Edited by L Buis; submitted 27.09.24; peer-reviewed by A Chieng, Z Guo; comments to author 19.12.24; revised version received 13.02.25; accepted 02.09.25; published 12.11.25.

Please cite as:
Jo YS, Pyeon A, Hu MK, Kim S, Choi IY, Kim DJ, Chun JW
Pilot Study of a Mobile, Virtual Reality–Based Digital Therapeutic for Smoking Cessation: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e66411
URL: https://mhealth.jmir.org/2025/1/e66411
doi:10.2196/66411
PMID:

©Yeong-Seon Jo, Arom Pyeon, Min-Kyung Hu, SungMin Kim, In-Young Choi, Dai-Jin Kim, Ji-Won Chun. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Unraveling the Factors Associated With Digital Health Intervention Uptake: Cross-Sectional Study

Ilona Ruotsalainen[1], PhD; Mikko Valtanen[2,3], MSc; Riikka Kärsämä[2], MA; Adil Umer[4], MSc; Hilkka Liedes[4], DSC; Suvi Parikka[2], PhD; Annamari Lundqvist[2], PhD; Kirsikka Aittola[5], PhD; Suvi Koivunen[5], PhD; Jussi Pihlajamäki[5,6], PhD; Anna-Leena Vuorinen[7], PhD; Jaana Lindström[2,5], PhD

[1]VTT Technical Research Centre of Finland Ltd, Kuopio, Finland

[2]Population Health Unit, Finnish Institute for Health and Welfare, Helsinki, Finland

[3]Department of Mathematics and Statistics, University of Turku, Turku, Finland

[4]VTT Technical Research Centre of Finland Ltd., Espoo, Finland

[5]Institute of Public Health and Clinical Nutrition, School of Medicine, University of Eastern Finland, Kuopio, Finland

[6]Department of Medicine, Endocrinology and Clinical Nutrition, Kuopio University Hospital, Kuopio, Finland

[7]Unit of Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland

Corresponding Author:
Ilona Ruotsalainen, PhD
VTT Technical Research Centre of Finland Ltd
Microkatu 1, Building M
Kuopio, 70211
Finland
Phone: 358 469209193
Email: ilona.ruotsalainen@vtt.fi

Abstract

Background: Chronic noncommunicable diseases (NCDs) remain a leading health challenge worldwide, and reducing modifiable lifestyle risk factors is a key prevention strategy. Digital health interventions (DHIs) offer scalable, cost-effective tools to support healthy behaviors, but concerns persist about their equitable reach and uptake across population groups.

Objective: This study aimed to examine how socioeconomic factors, health status, lifestyle behaviors, and attitudes and experiences related to the use of electronic services (e-services) are associated with the uptake of a DHI.

Methods: In this cross-sectional study, we invited (through mail or SMS) a subgroup of 6978 participants aged 20-74 years from the population-based Healthy Finland survey to take part in a DHI. The DHI, delivered via the web-based BitHabit app, aimed to support the adoption of healthy lifestyle habits. Uptake was defined as successful registration, agreeing to the terms of use, and accepting the invitation to participate. Predictor variables were drawn from national registry and self-reported survey data and included socioeconomic status, health indicators, lifestyle behaviors, and attitudes and experiences related to the use of e-services. Adjusted logistic regression models were used to identify significant predictors of DHI uptake.

Results: Of the final sample of 6975 participants, 1287 (18.5%) started using the DHI. Uptake was significantly higher among women (adjusted odds ratio [aOR] 1.69, 95% CI 1.49-1.93), middle-aged individuals (aOR 1.47, 95% CI 1.21-1.79), and those with higher income (aORs 1.76-1.97, 95% CIs 1.37-2.59) and more years of education (aOR 1.10, 95% CI 1.08-1.12). Healthier lifestyle indicators, including better diet quality (aOR 1.07, 95% CI 1.04-1.10), less frequent smoking or nonsmoking (aORs 1.59-2.29, 95% CIs 1.08-3.12), sleep (aOR 0.58, 95% CI 0.37-0.86), higher functional capacity (aOR 1.06, 95% CI 1.02-1.11), and good overall current health (aOR 1.46, 95% CI 1.15-1.89), were associated with increased likelihood of DHI uptake. The strongest predictors were related to the use of e-services: Individuals who used e-services (aORs 2.48-6.08, 95% CIs 1.19-11.92) reported higher competence to use e-services (aORs 2.00-4.10, 95% CIs 1.44-5.92), had low concerns about data security (aORs 1.37-1.76, 95% CIs 1.03-2.33), believed in the benefits of digital services (aOR 1.04, 95% CI 1.02-1.05), and had better internet connections had higher odds of uptake.

Conclusions: Our findings show that DHI uptake is associated with socioeconomic status, health and lifestyle factors, and, especially, individuals' experience and attitudes toward e-services. Individuals with lower education levels, lower income, and poorer health and lifestyle habits are less likely to adopt DHIs, raising concerns about potential digital health inequities. These

XSL•FO
RenderX

results underscore the need for targeted strategies to reduce barriers and ensure more equitable reach and engagement in future DHI implementations.

(JMIR Mhealth Uhealth 2025;13:e63896) doi:10.2196/63896

KEYWORDS

digital health intervention; DHI; uptake; lifestyle; adoption; digital literacy; health equity; cross-sectional study; literacy; cross-sectional; socioeconomic; habits; e-services; questionnaire survey; regression model; Finland; healthy lifestyle; logistic regression model

Introduction

Chronic noncommunicable diseases (NCDs) represent a significant health challenge worldwide, imposing an increasing economic burden on health care services and resulting in premature deaths [1,2]. To reduce the burden of NCDs, reducing modifiable lifestyle risk factors has been highlighted as one of the most important actions worldwide [3]. Adopting a healthy lifestyle can potentially reduce the risk of developing common chronic diseases by up to 80% [4]. There is a pressing need to develop strategies that promote the widespread adoption of healthy lifestyle habits, aiming to achieve impactful results at a large scale.

Digital health interventions (DHIs) offer a viable way to support a healthy lifestyle and hold promise as a low-cost means of delivering interventions to large population groups. However, DHIs can be effective only if they reach the target population, especially those who need them most. Ensuring that DHIs reach their intended users requires awareness of the potential barriers to uptake, which has been broadly defined as the initial step of registering for, downloading, or otherwise beginning to use a DHI [5,6]. This is particularly important in the context of NCDs, which disproportionately affect populations with lower socioeconomic status and limited access to health care services [7,8]. DHIs have the potential to improve access to preventive care and support individuals in managing their health, but to truly make a difference, it is essential to remove barriers that could otherwise deepen existing health inequalities [9]. Currently, there is a lack of comprehensive understanding regarding the factors influencing DHI uptake. Identifying these factors will facilitate the customization of future interventions and recruitment strategies, thereby enhancing adoption rates and promoting digital equity.

Digital health promotion is an emerging field, and therefore, research on DHIs remains scarce [10]. Much of the existing literature has concentrated on health apps, which often lack a defined duration, evidence-based modules, and explicit behavior change techniques that are typically built into DHIs [11,12]. These studies have shown that females [6,13,14], young or middle-aged individuals [6,13,14], and those with higher education levels [6,13,15] and higher income [6,15,16] tend to be more frequent users of health apps. Additionally, conflicting findings exist regarding the relationship between overall health [15,17,18], lifestyle habits [13,15,17,18], and health app usage. Recent research has also suggested that additional factors, such as perceived usefulness and privacy concerns, might play a significant role in influencing the adoption of health apps and digital health services [19-21]. Perceived usefulness has been associated with greater acceptance and intention to use digital health services across age groups [19,20]. In contrast, concerns regarding privacy and data security, particularly among older adults, can reduce usage intentions and act as barriers to adoption [19-22].

Studying health app users, although informative, does not offer comprehensive insights into potential barriers to DHI uptake. First, the lack of comparative analysis between adopters (individuals who start using DHIs) and nonadopters raises questions about whether observed characteristics reflect a higher uptake probability or the characteristics of the populations to which the apps were advertised. Second, the usage of health apps itself may influence individuals' lifestyle habits and health [23], introducing a potential bias and thereby complicating the interpretation of these findings in relation to DHI adoption. Lastly, usage is also influenced by other factors, such as user experience, which may be affected by the app quality [24].

Here, our aim was to study factors predicting DHI uptake to identify potential groups that may be less likely to start using DHIs. We examined whether socioeconomic factors, self-reported health and lifestyle habits, or habitual use of electronic services (e-services) predicts DHI uptake. Understanding these factors will help identify population groups that are underrepresented in DHIs and find alternative approaches for them. These findings will also aid in integrating potential barriers into the formulation of recruitment strategies and the development of DHIs themselves.

Methods

Participants and Study Design

This cross-sectional study was a substudy of the Healthy Finland population survey, the methodology of which is described in detail elsewhere [25,26]. In this survey, a questionnaire on health, well-being, and service use was sent to randomly selected individuals over the age of 20 years, representing the entire adult population of Finland. For our study, Finnish-speaking individuals aged 20-74 years who had completed the Healthy Finland questionnaire between September 2022 and the end of 2022 were considered eligible, excluding those randomly selected for the health examination component to prevent potential influence of the DHI on examination results. Among this eligible population (N=10,207), an SMS invitation was sent to all individuals with a known phone number (n=4978, 48.8%). Additionally, 2000 (19.6%) individuals from the remaining eligible population were sampled based on 5-year age groups to receive a letter invitation. Three participants asked their data to be removed, resulting in a final sample size of 6975 (99.9%)

XSL•FO
RenderX

individuals. The sample size was chosen to detect small effects in logistic regression, with a statistical power of 0.8 and a type I error of 0.05, assuming that approximately 10%-15% of invitees would adopt the DHI. Based on this, a minimum total sample size of 2538 participants was required.

Participants were offered the opportunity to use a DHI app (BitHabit) for 3 months [27]. The invitation letter included a web address, and the SMS message contained a direct link to the project's web page (hosted by the Finnish Institute for Health and Welfare), where participants could find information regarding the study, a link to the BitHabit app, and brief instructions on how to get started with the app.

Ethical Considerations

Ethical approval for the study was obtained from the research ethics committee of the Finnish Institute for Health and Welfare (THL/5335/6.02.01/2022). The approval covered all the procedures, data, and analysis presented in this study. Participation in the Healthy Finland survey was voluntary, and participants were informed that completing and returning the questionnaire constituted consent to participate. In addition, digital informed consent was obtained following BitHabit registration. Participants were informed of their right to decline or withdraw participation at any time. All data were pseudonymized before analysis, and the study was conducted in accordance with the General Data Protection Regulation (GDPR) and applicable Finnish data protection legislation and the analysis conducted in a secure analysis environment. No compensation was provided to participants.

The Digital Health Intervention App (BitHabit)

The BitHabit app was originally developed to support the formation of healthy lifestyle habits in adults at increased risk of type 2 diabetes [27,28]. The app includes a broad selection of small health-promoting actions ("habits") in 14 different categories related to physical activity, diet, sleep, stress management, positive mood, smoking, and alcohol consumption. It is a web-based app that can be used on a mobile phone, tablet, or computer. To log into the app, participants had to provide a phone number and a user ID that was specified either in the letter or in the SMS message. Uptake of the app was defined as the process of registering for the BitHabit app with a phone number and user ID, agreeing to the terms of use, and, subsequently accepting the invitation to participate. No additional confirmation of refusal was collected from those who did not register.

Predictor Variables

We examined predictors from the following categories: socioeconomics, health, lifestyle habits, and the use of e-services. All predictor variables except age and sex (assigned at birth), which were obtained from the Finnish National Population Register, were derived from the Healthy Finland questionnaire and were thus self-reported. The survey was conducted approximately 2-4 months prior to sending invitations to take part in the DHI substudy, and the BitHabit app was offered only after the survey responses had been received. Following the invitation, participants had approximately 4 weeks to begin using the app. All questions and the answer options are presented in Multimedia Appendix 1.

Socioeconomics

Socioeconomic variables included age, sex, education (years of education, including primary and comprehensive school), household income (€ per year before taxes), and employment status. Household income was coded as a five-category ordinal variable: <€15,000 (<US $17,437; calculated at an exchange rate of €1=US $1.16 approximately), €15,001-35,000 (US $17,438-$40,686), €35,001-55,000 (US $40,687-$63,935), €55,001-75,000 (US $63,936-$87,184), and >€75,000 (>US $87,184). Age was categorized into four groups: 20-34, 35-49, 50-64, and 65-74 years.

Health

The BMI was calculated based on self-reported height and weight. Other health-related predictor variables included functional capacity, ability to work, current health status, prevalence of (any) long-term illness, health care service use (number of visits during the past 12 months), perceived ability to learn, and perceived limitations due to health problems. Functional capacity was calculated as a summary score for five questions on the abilities to perform the following activities: run for about 100 m, walk for about 500 m without stopping to rest, see ordinary newspaper print, hear what is said in a conversation between several people, and walk up one flight of stairs without stopping to rest. Each subquestion had four choices (yes, with no problem [3 points]; yes, with some difficulty [2 points]; yes, but with great difficulty [1 point]; no, I cannot [0 points]). A higher summary score indicated better functional capacity.

Lifestyle Habits

Lifestyle-related information included physical activity, sleep, smoking, and diet quality. Participants were divided into two groups based on whether they achieved the Finnish physical activity recommendations [29]. Sleep adequacy categories were created based on the question, "Do you get enough sleep?" The diet quality score was calculated following the method presented by Lindström et al [30], and the details of the scoring are in the *Diet Quality Scoring* section in Multimedia Appendix 1.

Use of e-Services

The following predictors relating to the use of e-services were considered: independence of e-service use, competence to use e-services, accessibility of e-services, concerns about data security, poor internet connections, and perceived benefits of e-services. The "perceived benefits of e-services" score was a summary measure calculated based on participants' level of agreement with six claims: e-services (1) help me assess the need for services, (2) support me in finding and choosing the most suitable services, (3) make it easier for me to use the services regardless of where I am and when, (4) make it easier for me to collaborate with professionals, (5) help me take an active role in looking after my own health and welfare, and (6) help me take care of the health, welfare, and functional capacity of family or friends. Each claim had five categories (completely agree, somewhat agree, neither agree nor disagree, somewhat

XSL•FO
RenderX

disagree, strongly disagree). A higher score indicated that the participant perceived e-services as more beneficial.

Statistical Analysis

Logistic regression models were used to assess the association between the predictor variables and the uptake of the BitHabit DHI. The results were presented as adjusted odds ratios (aORs) with 95% CIs. Models examining variables related to health, lifestyle, and use of e-services were adjusted by contact method (SMS or letter), age, sex, education, and income. Regarding socioeconomic factors, models for age, sex, and contact method were adjusted by each other; models for income, BMI, and employment status by education, contact method, age, and sex; and the model for education by contact method, age, and sex. These covariates were selected a priori as background variables likely to be confounders of the predictor-outcome relationship based on the previous literature. The only exception was contact method (SMS vs letter), which was included to adjust for potential confounding introduced by the recruitment strategy: letter invitations were sampled by age to ensure representation across age groups, potentially correlating with both participant characteristics and DHI uptake. Participants with missing data (missing answers) were excluded from the analyses, and a complete-case analysis was performed. The amount of missing data for each predictor variable is presented in the *Results* section.

The overall significance of the associations between the categorical predictors and the outcome was assessed with likelihood ratio tests (LRTs), for which a false discovery rate (FDR)–corrected *P* value of <.05 was determined to indicate a significant association. R software version 4.3.1 (R Foundation for Statistical Computing [31]) was used to perform all statistical analyses. The reporting of this study followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting observational studies [32].

Results

Baseline Information

The final sample of 6975 participants, 1287 (18.5%) started using the BitHabit app (“adopters”). The uptake proportion was 16.2% (806/4975) among those who were invited via SMS and 24.1% (481/2000) among those with a letter invitation. Those invited via SMS were less likely to start using the app than those invited via mail (aOR 0.65, 95% CI 0.55-0.75). Background information for the entire study sample and separately for app adopters and nonadopters is presented in Tables 1-4.

XSL•FO
RenderX

Table 1. Socioeconomic characteristics of the study sample.[a]

Characteristics	Total sample (N=6975)	Adopters (n=1287)	Nonadopters (n=5688)
Age (years), n (%)			
20-34	1229 (17.6)	250 (19.4)	979 (17.2)
35-49	1254 (18.0).	303 (23.5)	951 (16.7)
50-64	2066 (29.6)	361 (28.0)	1705 (30.0)
65-74	2426 (34.8)	373 (29.0)	2053 (36.1)
Sex, n (%)			
Female	3975 (57.0)	868 (67.4)	3107 (54.6)
Male	3000 (43.0)	419 (32.6)	2581 (45.4)
Education (years)			
Mean (SD)	14.0 (3.6)	15.2 (3.4)	13.8 (3.6)
Missing, n (%)	95 (1.4)	10 (0.8)	85 (1.5)
Annual household income (€; US $[b]), n (%)			
<15,000; <17,437	761 (10.9)	94 (7.3)	667 (11.7)
15,001-35,000; 17,438-40,686	1913 (27.4)	266 (20.7)	1647 (29.0)
35,001-55,000; 40,687-63,935	1916 (27.5)	380 (29.5)	1536 (27.0)
55,001-75,000; 63,936-87,184	1171 (16.8)	264 (20.5)	907 (15.9)
>75,000; >87,184	1119 (16.0)	272 (21.1)	847 (14.9)
Missing, n (%)	95 (1.4)	11 (0.9)	84 (1.5)
Employment status, n (%)[c]			
Employed full-time	1798 (44.2)	389 (51.1)	1409 (42.6)
Employed part-time	211 (5.2)	41 (5.4)	170 (5.1)
Retired	1463 (35.0)	228 (30.0)	1235 (37.4)
On a disability pension or rehabilitation benefit	161 (4.0)	24 (3.2)	137 (4.1)
Part-time retirement	27 (0.7)	1 (0.1)	26 (0.8)
Unemployed	169(4.2)	36 (4.7)	133 (4.0)
Family leave or stay-at-home parent	67(1.6)	19 (2.5)	48 (1.5)
Other	123(3.0)	21 (2.8)	102 (3.1)
Missing, n (%)	48 (1.2)	2 (0.3)	46 (1.4)

[a]Data are reported as the mean (SD) for continuous variables and as n (%) for categorical variables.

[b]An exchange rate of €1=US $1.16 approximately was applied.

[c]The number of participants was as follows: total, N=4067; adopters, n=761; and nonadopters, n=3306.

XSL•FO
RenderX

Table 2. Health-related characteristics of the study sample.[a]

Characteristics	Total sample (N=6975)	Adopters (n=1287)	Nonadopters (n=5688)
BMI (kg/m^2)			
Mean (SD)	27.4 (5.3)	27.5 (5.5)	27.3 (5.2)
Missing, n (%)	75 (1.1)	6 (0.5)	69 (1.2)
Functional capacity			
Mean (SD)	13.8 (2.0)	14.1 (1.5)	13.7 (2.1)
Missing, n (%)	144 (2.1)	15 (1.2)	129 (2.3)
Ability to work			
Completely unable to work	350 (5.0)	49 (3.8)	301 (5.3)
Partially unable to work	1688 (24.2)	247 (19.2)	1441 (25.3)
Completely fit for work	4878 (69.9)	986 (76.6)	3892 (68.4)
Missing, n (%)	59 (0.8)	5 (0.4)	54 (0.9)
Current health status			
Poor or fairly poor	676 (9.7)	83 (6.4)	593 (10.4)
Average	1740 (24.9)	278 (21.6)	1462 (25.7)
Good or fairly good	4520 (64.8)	925 (71.9)	3595 (63.2)
Missing, n (%)	39 (0.6)	1 (0.1)	38 (0.7)
Long-term illnesses			
Yes	3937 (56.4)	713 (55.4)	3224 (56.7)
No	2932 (42.0)	562 (43.7)	2370 (41.7)
Missing, n (%)	106 (1.5)	12 (0.9)	94 (1.7)
Health care visits			
Mean (SD)	5.7 (7.4)	5.8 (6.6)	5.6 (7.6)
Missing, n (%)	655 (9.4)	125 (9.7)	530 (9.3)
Memory and learning			
Poorly or very poorly	277 (4.0)	34 (2.6)	243 (4.3)
Adequately	1779 (25.5)	249 (19.3)	1530 (26.9)
Well or very well	4864 (69.7)	1001 (77.8)	3863 (67.9)
Missing, n (%)	55 (0.8)	3 (0.2)	52 (0.9)
Limitations due to health problems			
Severely limited	384 (5.5)	51 (4.0)	333 (5.9)
Limited but not severely	2424 (34.8)	413 (32.1)	2011 (35.4)
Not limited at all	4032 (57.8)	803 (62.4)	3229 (56.8)
Missing, n (%)	135 (1.9)	20 (1.6)	115 (2.0)

[a]Data are reported as the mean (SD) for continuous variables and as n (%) for categorical variables.

XSL•FO
RenderX

Table 3. Lifestyle-related characteristics of the study sample.[a]

Characteristics	Total sample (N=6975)	Adopters (n=1287)	Nonadopters (n=5688)
Diet quality			
Mean (SD)	8.1 (2.5)	8.5 (2.5)	7.9 (2.5)
Missing, n (%)	193 (2.8)	31 (2.4)	162 (2.8)
Physical activity			
Did not achieve recommended level	4122 (59.1)	741 (57.6)	3381 (59.4)
Achieved recommended level	2530 (36.3)	515 (40.0)	2015 (35.4)
Missing, n (%)	323 (4.6)	31 (2.4)	292 (5.2)
Enough sleep			
Yes, almost always or often	5324 (76.3)	991 (77.0)	4333 (76.2)
Rarely or hardly ever	1315 (18.9)	261 (20.3)	1054 (18.5)
Not sure	286 (4.1)	29 (2.3)	257 (4.5)
Missing, n (%)	50 (0.7)	6 (0.4)	44 (0.8)
Smoking			
Daily	675 (9.7)	56 (4.4)	619 (10.9)
Occasionally	419 (6.0)	64 (5.0)	355 (6.2)
Not at all	3137 (45.0)	586 (45.5)	2551 (44.8)
Have never smoked	2548 (36.5)	548 (42.6)	2000 (35.2)
Missing, n (%)	196 (2.8)	33 (2.5)	163 (2.9)

[a]Data are reported as the mean (SD) for continuous variables and as n (%) for categorical variables.

XSL•FO
RenderX

Table 4. E-service[a] use–related characteristics of the study sample.[b]

Characteristics	Total sample (N=6975)	Adopters (n=1287)	Nonadopters (n=5688)
Independence of e-service use			
Do not use	447 (6.4)	11 (0.9)	436 (7.7)
Use with help or someone else uses on my behalf	306 (4.4)	20 (1.5)	286 (5.0)
Use independently	6177 (88.6)	1252 (97.3)	4925 (86.6)
Missing, n (%)	45 (0.6)	4 (0.3)	41 (0.7)
Competence to use online services			
No competence	208 (3.0)	6 (0.5)	202 (3.6)
Low competence	550 (7.9)	37 (2.9)	513 (9.0)
Moderate competence	2076 (29.7)	268 (20.8)	1808 (31.8)
High competence	1976 (28.3)	446 (34.6)	1530 (26.8)
Very high competence	2103 (30.2)	529 (41.1)	1574 (27.7)
Missing, n (%)	62 (0.9)	1 (0.1)	61 (1.1)
E-services not accessible to me			
Completely or somewhat agree	831 (11.9)	118 (9.2)	713 (12.5)
Neither agree nor disagree	1583 (22.7)	274 (21.3)	1309 (23.0)
Strongly or somewhat disagree	4022 (57.7)	838 (64.1)	3184 (56.0)
Missing, n (%)	539 (7.7)	57 (4.4)	482 (8.5)
Concerned about data security			
Completely agree	724 (10.3)	80 (6.2)	644 (11.3)
Somewhat agree	1441 (20.7)	255 (19.8)	1186 (20.9)
Neither agree nor disagree	1242 (17.8)	203 (15.8)	1039 (18.3)
Somewhat disagree	1392 (20.0)	310 (24.1)	1082 (19.0)
Strongly disagree	1882 (27.0)	415 (32.2)	1467 (25.8)
Missing, n (%)	294 (4.2)	24 (1.9)	270 (4.7)
Poor internet connections			
Completely or somewhat agree	727 (10.4)	111 (8.6)	616 (10.8)
Neither agree nor disagree	1050 (15.1)	151 (11.7)	899 (15.8)
Strongly or somewhat disagree	4820 (69.1)	996 (77.4)	3824 (67.2)
Missing, n (%)	378 (5.4)	29 (2.3)	349 (6.1)
Benefits of digital services			
Mean (SD)	16.4 (5.1)	17.4 (4.7)	16.1 (5.2)
Missing, n (%)	465 (6.7)	54 (4.2)	411 (7.2)

[a]e-service: electronic service.

[b]Data are reported as the mean (SD) for continuous variables and as n (%) for categorical variables.

Predictor Variables Associated With DHI Uptake

Factors associated with the uptake of the BitHabit app were studied in four domains: socioeconomics, health, lifestyle habits, and use of e-services. The aORs from all the models, along with associated *P* values and 95% CIs, are listed here, while details of the adjustments for each variable are listed in *Methods* section.

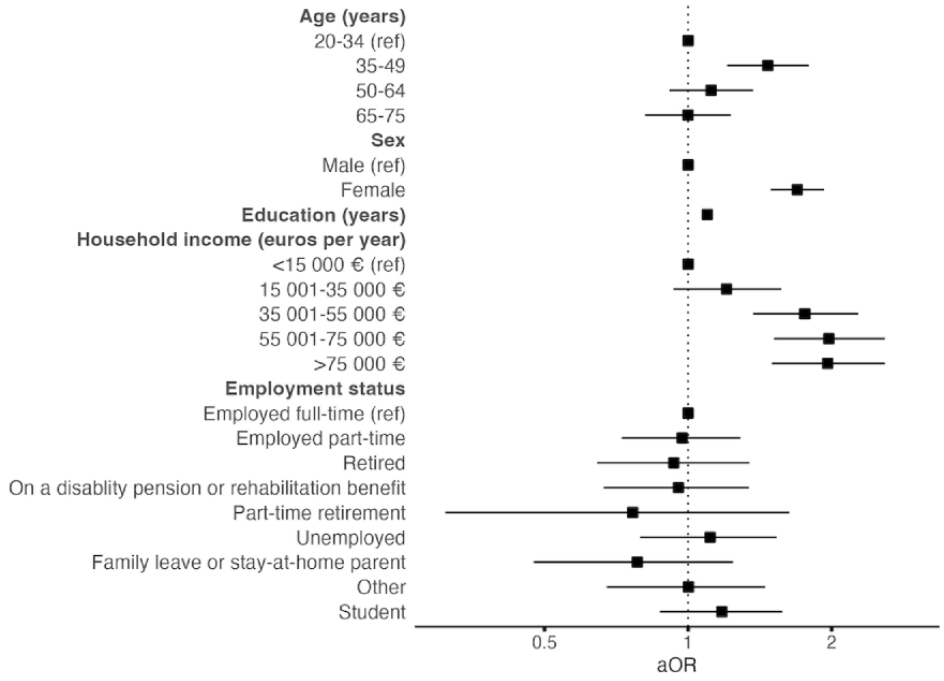
Socioeconomics

Age, sex, education, and household income demonstrated a significant association with uptake of the BitHabit DHI (Figure 1 and Table 5). Odds of adopting the DHI were higher for individuals aged 35-49 years compared to the reference group of 20-34 years (aOR 1.47, 95% CI 1.21-1.79). The two older age groups (50-64 and 65-74 years) showed no significant difference from the reference group. Higher educational attainment was associated with greater odds of DHI uptake (aOR 1.10, 95% CI 1.08-1.12). Similarly, higher income levels

XSL•FO
RenderX

were associated with increased uptake, though the two highest categories had comparable estimates (€35,001-55,000: aOR 1.76, 95% CI 1.37-2.27; €55,001-75,000: aOR 1.97, 95% CI 1.51-2.59; >€75,000: aOR 1.96, 95% CI 1.50-2.59). In addition, women were more likely than men to adopt the DHI (aOR 1.69, 95% CI 1.49-1.93), whereas employment status was not significantly associated with DHI uptake.

Figure 1. Forest plot showing aORs and 95% CIs for socioeconomic factors. Each aOR is adjusted for relevant covariates, as detailed in the *Statistical Analysis* section. aOR: adjusted odds ratio; ref: reference.



XSL•FO
RenderX

Table 5. Logistic regression and LRT[a] results for socioeconomic predictor variables.

Predictor variables	LRT		Logistic regression	
	χ^2 (*df*)	FDR[b]-adjusted *P* value	aOR[c] (95% CI)	*P* value (group comparison)
Age (years)	**22.37 (3)**	**<.001**	—[d]	—
20-34 (reference)	—	—	—	—
35-49	—	—	1.47 (1.21-1.79)	<.001
50-64	—	—	1.12 (0.91-1.37)	.28
65-75	—	—	1.00 (0.81-1.22)	.99
Sex	**66.75 (1)**	**<.001**	—	—
Male (reference)	—	—	—	—
Female	—	—	1.69 (1.49-1.93)	<.001
Education (years)	98.8 (1)	<.001	1.10 (1.08-1.12)	
Annual household income (€; US $[e])	**50.19 (4)**	**<.001**	—	—
<15,000; <17,437 (reference)	—	—	—	—
15,001-35,000; 17,438-40,686	—	—	1.20 (0.93-1.57)	.16
35,001-55,000; 40,687-63,935	—	—	1.76 (1.37-2.27)	<.001
55,001-75,000; 63,936-87,184	—	—	1.97 (1.51-2.59)	<.001
>75,000; >87,184	—	—	1.96 (1.50-2.59)	<.001
Employment status	**3.6 (8)**	**.89**	—	—
Employed full-time (reference)	—	—	—	—
Employed part-time	—	—	0.97 (0.73-1.29)	.85
Retired	—	—	0.93 (0.64-1.34)	.71
On a disability pension or rehabilitation benefit	—	—	0.95 (0.67-1.34)	.79
Part-time retirement	—	—	0.77 (0.31-1.63)	.52
Unemployed	—	—	1.11 (0.79-1.53)	.53
Family leave or stay-at-home parent	—	—	0.78 (0.47-1.24)	.32
Other	—	—	1.00 (0.68-1.45)	.99
Student	—	—	1.18 (0.87-1.58)	.28

[a]LRT: likelihood ratio test.

[b]FDR: false discovery rate.

[c]aOR: adjusted odds ratio.

[d]Not applicable.

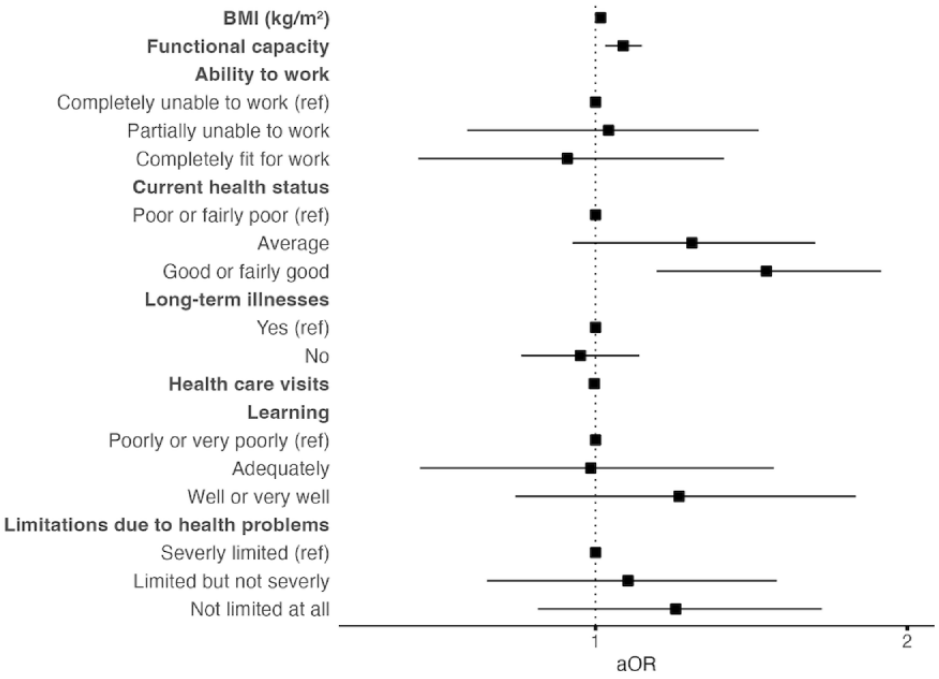
[e]An exchange rate of €1=US $1.16 approximately was applied.

Health

Among the health-related predictors, functional capacity and current health status showed significant associations with DHI uptake (Figure 2 and Table 6). The odds of adopting the DHI were higher for individuals with better functional status (aOR 1.06, 95% CI 1.02-1.11) and better current health (good or fairly good compared to poor: aOR 1.46, 95% CI 1.15-1.89). In contrast, none of the other health indicators, including the BMI, ability to work, ability to learn, the presence of long-term illnesses, the number of health care visits during the past 12 months, and limitations due to health problems, were significantly associated with uptake of the DHI.

XSL•FO
RenderX

Figure 2. Forest plot showing aORs and 95% CIs for health-related predictors. Each aOR is adjusted for relevant covariates, as detailed in the *Statistical Analysis* section. aOR: adjusted odds ratio; ref: reference.



XSL•FO
RenderX

Table 6. Logistic regression and LRT[a] results for health-related predictor variables.

Predictor variables	LRT		Logistic regression	
	χ^2 (*df*)	FDR[b]-adjusted *P* value	aOR[c] (95% CI)	*P* value (group comparison)
BMI (kg/m^2)	3.6 (1)	.08	1.01 (1.00-1.02)	—[d]
Functional capacity	9.27 (1)	.005	1.06 (1.02-1.11)	—
Ability to work	**1.16 (2)**	**.61**	—	—
Completely unable to work (reference)	—	—	—	—
Partially unable to work	—	—	1.03 (0.75-1.44)	.86
Completely fit for work	—	—	0.94 (0.67-1.33)	.72
Current health status	**11.99 (2)**	**.005**	—	—
Poor or fairly poor (reference)	—	—	—	—
Average	—	—	1.24 (0.95-1.63)	.12
Good or fairly good	—	—	1.46 (1.15-1.89)	.003
Long-term illnesses	**0.25 (1)**	**.64**	—	—
Yes (reference)	—	—	—	—
No	—	—	0.97 (0.85-1.10)	.62
Health care visits	**0.41 (1)**	**.60**	**1.00 (0.99-1.01)**	—
Ability to learn	**6.24 (2)**	**.06**	—	—
Poorly or very poorly (reference)	—	—	—	—
Adequately	—	—	0.99 (0.68-1.49)	.96
Well or very well	—	—	1.20 (0.84-1.78)	.33
Limitations due to health problems	**3.06 (2)**	**.26**	—	—
Severely limited (reference)	—	—	—	—
Limited but not severely	—	—	1.07 (0.79-1.50)	.66
Not limited at all	—	—	1.19 (0.88-1.65)	.27

[a]LRT: likelihood ratio test.

[b]FDR: false discovery rate.

[c]aOR: adjusted odds ratio.

[d]Not applicable.

Lifestyle Habits

Diet quality, smoking, and sleeping were associated with the uptake of the DHI (Figure 3 and Table 7). Participants with better diet quality (aOR 1.07, 95% CI 1.04-1.10) had higher odds of DHI uptake. Similarly, smoking behavior showed a clear gradient: Compared to daily smokers, those who smoked occasionally (aOR 1.59, 95% CI 1.08-2.36), did not smoke at the time of the survey (aOR 2.26, 95% CI 1.69-3.06), or had never smoked (aOR 2.29, 95% CI 1.72-3.12) had higher odds of adopting the DHI. Sleeping enough was also a significant predictor; however, no meaningful difference was estimated between those sleeping mostly enough and the reference category of those sleeping mostly not enough. Instead, participants who were unsure of whether they slept enough had significantly lower odds of DHI uptake compared to the reference category (aOR 0.58, 95% CI 0.37-0.86). In contrast, achieving physical activity recommendations was not significantly associated with uptake of the DHI.

XSL•FO
RenderX

Figure 3. Forest plot showing aORs and 95% CIs for lifestyle habits. Each aOR is adjusted for relevant covariates, as detailed in the *Statistical Analysis* section. aOR: adjusted odds ratio; ref: reference.

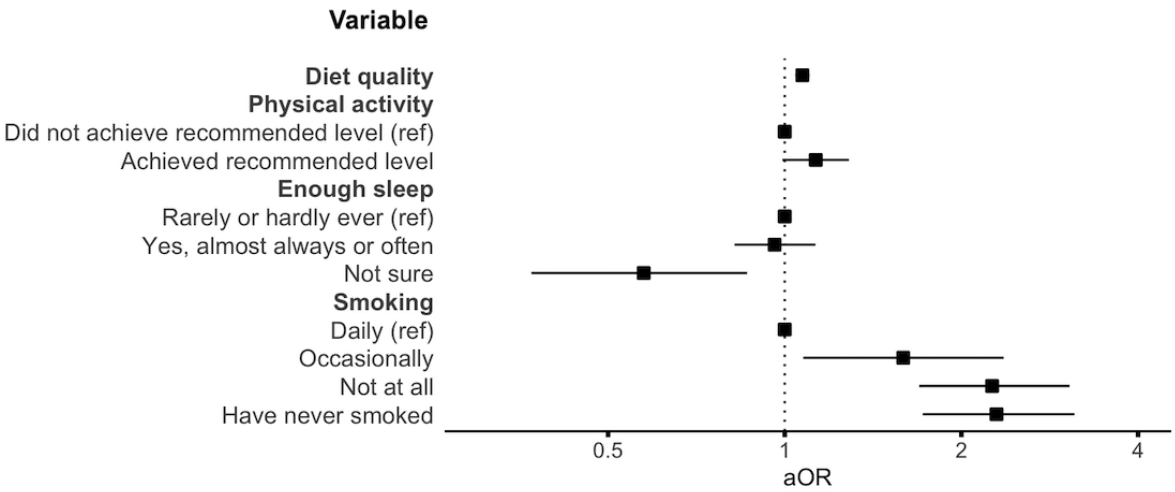


Table 7. Logistic regression and LRT[a] results for lifestyle-related predictor variables.

Predictor variables	LRT		Logistic regression	
	χ^2 (*df*)	FDR[b]-adjusted *P* value	aOR[c] (95% CI)	*P* value (group comparison)
Diet quality	23.67 (1)	<.001	1.07 (1.04-1.10)	—[d]
Physical activity	**3.34 (1)**	**.09**	—	—
Did not achieve recommended level (reference)	—	—	—	—
Achieved recommended level	—	—	1.13 (0.99-1.29)	.07
Enough sleep	**7.54 (2)**	**.04**	—	—
Rarely or hardly ever (reference)	—	—	—	—
Yes, almost always or often	—	—	0.96 (0.82-1.13)	.62
Not sure	—	—	0.58 (0.37-0.86)	.01
Smoking	**41.76 (3)**	**<.001**	—	—
Daily (reference)	—	—	—	—
Occasionally	—	—	1.59 (1.08-2.36)	.02
Not at all	—	—	2.26 (1.69-3.06)	<.001
Have never smoked	—	—	2.29 (1.72-3.12)	<.001

[a]LRT: likelihood ratio test.

[b]FDR: false discovery rate.

[c]aOR: adjusted odds ratio.

[d]Not applicable.

Use of e-Services

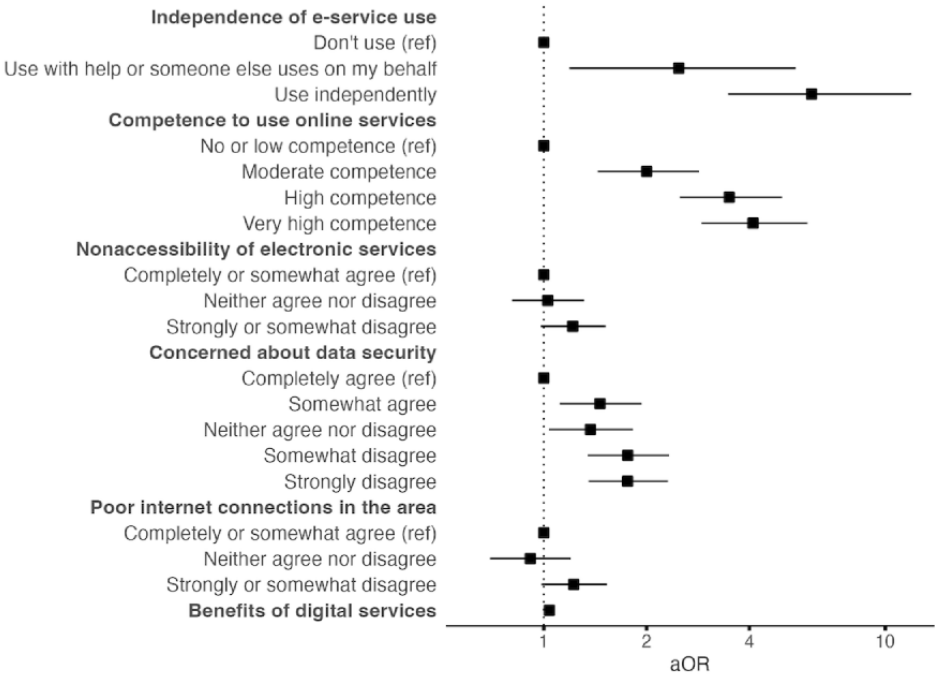
All examined factors related to the use of e-services, except the perceived accessibility of e-services, showed a significant association with DHI uptake (Figure 4 and Table 8). Individuals who reported using e-services had notably higher odds of adopting the DHI. Compared to those who did not use e-services at all, participants who used them with assistance had more than twice the odds of DHI uptake (aOR 2.48, 95% CI 1.19-5.47), while those who used them independently had over six times the odds of DHI uptake (aOR 6.08, 95% CI 3.46-11.92). Competence in using online services also showed a strong relationship: Those with moderate competence had an aOR of 2.00 (95% CI 1.44-2.85), those with high competence had an aOR of 3.49 (95% CI 2.50-4.98), and those with very high competence had an aOR of 4.10 (95% CI 2.90-5.92).

Attitudes toward data security also played a role. Participants who expressed fewer concerns about data security (somewhat agree that I am concerned about data security vs completely

XSL•FO
RenderX

agree: aOR 1.46, 95% CI 1.11-1.93; neither agree nor disagree vs completely agree: aOR 1.37, 95% CI 1.03-1.82; somewhat disagree vs completely agree: aOR 1.76, 95% CI 1.35-2.33; strongly disagree vs completely agree: aOR 1.76, 95% CI 1.35-2.31) had higher odds of DHI uptake. Additionally, a stronger belief in the benefits of digital services (aOR 1.04, 95% CI 1.02-1.05) was associated with increased DHI uptake. Finally, individuals who reported good connections had the highest estimated odds of adopting the DHI.

Figure 4. Forest plot showing aORs and 95% CIs for e-service use–related factors. Each aOR is adjusted for relevant covariates, as detailed in the *Statistical Analysis* section. aOR: adjusted odds ratio; e-service: electronic service; ref: reference.



XSL•FO
RenderX

Table 8. Logistic regression and LRT[a] results for e-service[b] use–related predictor variables.

Predictor variables	LRT		Logistic regression	
	χ^2 (*df*)	FDR[c]-adjusted *P* value	aOR[d] (95% CI)	*P* value (group comparison)
Independence of e-service use	**70.22 (2)**	**<.001**	—[e]	—
Do not use (reference)	—	—	—	—
Use with help or someone else uses on my behalf	—	—	2.48 (1.19-5.47)	.02
Use independently	—	—	6.08 (3.46-11.92)	<.001
Competence to use online services	**97.61 (3)**	**<.001**	—	—
No or low competence (reference)	—	—	—	—
Moderate competence	—	—	2.00 (1.44-2.85)	<.001
High competence	—	—	3.49 (2.50-4.98)	<.001
Very high competence	—	—	4.10 (2.90-5.92)	<.001
E-services not accessible to me	**6.41 (2)**	**.06**	—	—
Completely or somewhat agree (reference)	—	—	—	—
Neither agree nor disagree	—	—	1.03 (0.81-1.31)	.83
Strongly or somewhat disagree	—	—	1.22 (0.98-1.52)	.08
Concerned about data security	**24.75 (4)**	**<.001**	—	—
Completely agree (reference)	—	—	—	—
Somewhat agree	—	—	1.46 (1.11-1.93)	.007
Neither agree nor disagree	—	—	1.37 (1.03-1.82)	.03
Somewhat disagree	—	—	1.76 (1.35-2.33)	<.001
Strongly disagree	—	—	1.76 (1.35-2.31)	<.001
Poor internet connections	**11.11 (2)**	**.007**	—	—
Completely or somewhat agree (reference)	—	—	—	—
Neither agree nor disagree	—	—	0.91 (0.70-1.20)	.51
Strongly or somewhat disagree	—	—	1.22 (0.98-1.53)	.07
Benefits of digital services	27.38 (1)	<.001	1.04 (1.02-1.05)	—

[a]LRT: likelihood ratio test.

[b]e-service: electronic service.

[c]FDR: false discovery rate.

[d]aOR: adjusted odds ratio.

[e]Not applicable.

Discussion

Principal Findings

This study examined the relationship between background characteristics in four different domains (socioeconomics, health, lifestyle habits, and use of e-services) and the uptake of a lifestyle DHI (BitHabit). Overall, 18.2% (n=1287) of participants who received the invitation started using the BitHabit app. Women, middle-aged individuals, and those with higher income and education levels were more likely to take up the DHI. Moreover, better lifestyle choices as well as better functional capacity and current health were positively associated with higher odds of DHI uptake. Our research also underscored the impact of digital literacy and positive attitudes towards online services, revealing that individuals with greater familiarity and more positive views about online services were more likely to register for the DHI. On the contrary, men and individuals with lower socioeconomic status, limited digital skills, and poorer health and lifestyle habits were less likely to engage with the DHI. These findings imply that when offering DHIs, it may be challenging to engage those individuals who would benefit the most from a lifestyle DHI, emphasizing the importance of tailoring recruitment and intervention methods accordingly.

XSL•FO
RenderX

Comparison With Prior Work

Although there is a scarcity of studies directly comparing DHI adopters and nonadopters, several studies exploring health app usage report that women and individuals in the young or middle-aged categories use more health apps [6,13,33,34]. These findings are also supported by recent results from a large population-level study comparing the characteristics of individuals who registered to use a mobile health physical activity intervention (N=696,907) with the entire population of Singapore; the results showed that women and younger participants are more inclined to register for a health app [14]. Our finding that women and those aged 35-49 years are more likely to take up a DHI is mainly consistent with previous studies examining health app usage. Interestingly, our results do not support the idea that young adults are more likely to take up DHIs. These conflicting findings could be due to differences in the study design, as most of the previous studies examining usage have focused on consumer-targeted health apps, which may be more appealing to younger adults.

Regarding income and education, previous studies have suggested that individuals with higher income levels or socioeconomic status and those with higher levels of education tend to use more health apps [6,13,15,16]. Our results showing that those with higher income levels and higher education are more likely to take up a DHI align with the previous findings on health app usage. This suggests that socioeconomic factors may continue to play a role in shaping the adoption of DHIs, highlighting the importance of considering these factors in future implementation strategies. However, although our study sheds light on who is more likely to adopt a DHI, it provides limited insight into the reasons why some individuals choose not to adopt DHIs. A deeper understanding of nonadopters is essential for designing inclusive digital interventions. Future research should explore the motivations and contextual factors that influence nonparticipation in DHIs.

DHIs and digital lifestyle interventions aim to promote healthy lifestyles and improve health outcomes. However, in general, it is not well known whether an individual's lifestyle or health influences DHI uptake. Our results suggest that healthier lifestyle choices, such as better diet quality and nonsmoking behavior, along with better self-reported functional capacity and current health status, are associated with higher odds of DHI uptake. In this study, perceived lack of sleep was not associated with uptake of the BitHabit DHI; however, those individuals who were not sure of whether they got enough sleep had lower odds of taking up the DHI. This response might indicate a lack of personal interest in the individual's lifestyle habits. However, it is challenging to draw strong conclusions based on this answer. Prior studies imply a potential link between health app usage and lifestyle choices, but results regarding diet quality, smoking, and physical activity still remain inconclusive in earlier investigations [13,15,17,18,35]. Conflicting results have also been found on the link between health app use and the BMI [14,15,18], as well as between chronic diseases and health app possession [15,17,18,36]. These earlier contradictory findings may be due to variations in the study design. Interestingly, those with better self-reported health status have been reported to be more likely to download health apps [36]. Overall, the results of this study regarding lifestyle habits and health imply that if a DHI aims to target those with poorer health and a nonhealthy lifestyle, we may not get the intended target population involved when using a "one-size-fits-all" recruitment strategy.

Digital literacy, defined as individuals' ability to use digital technologies and understand their associated risks, emerged as a key factor in the uptake of the BitHabit DHI [37]. Our findings indicate that digital competence and views about online services are associated with the uptake of DHIs. In particular, individuals' competence and their ability to use e-services independently are closely linked to their likelihood of taking up DHIs. Those reporting very high competence in using e-services had four times the odds to register for the DHI compared to those with little to no competence. Additionally, individuals who independently used e-services had six times the odds to take up the DHI compared to those who did not use e-services at all. These results highlight the importance of addressing the needs of individuals with limited e-service competence to enable them to access DHIs or of exploring alternative approaches. Furthermore, concerns about data security were associated with DHI uptake. Previous research on digital health services [19,38] and mobile health apps [20,21] has also highlighted the role of data security concerns as potential barriers to the usage of these services. Addressing security-related issues, such as ensuring compliance with regulations and enhancing transparency regarding data collection and handling, may help alleviate some of these concerns.

Strengths and Limitations

Our approach has some strengths. Most importantly, our study population was drawn from participants of a recent population-based survey that included a wide age and socioeconomic distribution. The available knowledge of the background characteristics of the entire invited population enabled a direct comparison between adopters and nonadopters, instead of only studying adopters, which has been a common approach in earlier research.

There are also some potential limitations. It is worth noting that individuals who chose not to participate in the Healthy Finland survey were excluded from the sample used in the study. In addition, participants were asked to first fill the Healthy Finland survey online, which could introduce bias among respondents who were later approached and offered the DHI. Nevertheless, this bias may have been mitigated by the option to request a paper format for the survey. However, this potential bias may have led to a slight overrepresentation of digitally literate individuals, which could influence the generalizability of the findings to less digitally engaged populations. In addition, there was a 2-4–month delay between baseline data collection and app uptake, which may have allowed changes in predictors. This study was conducted in a single country (Finland), and participation was limited to Finnish-speaking individuals due to practical constraints, which may limit generalizability to other countries and language minorities within Finland. Specifically, the significance of the quality of internet connections in the area may be more pronounced in regions with a higher proportion of individuals experiencing poor connections. Finally, we did

XSL•FO
RenderX

not collect explicit refusal data from nonparticipants. As such, we cannot distinguish between active refusal, missed invitations, or disinterest, which may affect the interpretation of nonuptake.

Conclusion

DHIs have significant potential to deliver services to large population groups and thereby decrease the burden of NCDs at the population level. Despite this potential, there are notable concerns about unequal access to and use of DHIs and how this will affect health equity [39]. Our findings indicate that those with lower socioeconomic status, poorer digital literacy, and poorer health and lifestyle habits are less likely to take up DHIs. Therefore, effectively reaching individuals likely to benefit most from lifestyle DHIs may be complex. Consequently, it is important consider the targeted population group when designing recruitment strategies and intervention modalities. Otherwise, the digitalization of lifestyle interventions might lead to widening health equity gaps between population groups.

Acknowledgments

JL, A-LV, and IR designed the study. IR, AU, AL, and SP were involved in the data collection, and AU maintained the BitHabit software. The analysis plan was designed by MV, JL, and IR, and analyses were conducted by MV. IR, MV, and JL interpreted the data, and IR visualized the results. The original draft was written by IR, MV, RK, and JL. All authors reviewed and edited the manuscript. JL and JP acquired the funding for this project. This study was funded by the Academy of Finland (grant numbers 332464, 332465, and 332466).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions on predictor variables in the Healthy Finland survey and diet quality scoring.
[PDF File (Adobe PDF File), 125 KB - mhealth_v13i1e63896_app1.pdf]

Multimedia Appendix 2
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for cross-sectional studies.
[DOCX File , 33 KB - mhealth_v13i1e63896_app2.docx]

References

1. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
2. World Health Organization global status report on noncommunicable diseases 2014. World Health Organization. 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf/ [accessed 2023-08-30]
3. Discussion paper for the regional expert consultaitons – development of an implementation roadmap 2023–2030 for the global action plan for the prevention and control of NCDs 2013–2030. World Health Organization. 2021. URL: https://shorturl.at/t6aUY [accessed 2025-11-14]
4. Ford ES, Bergmann MM, Kröger J, Schienkiewitz A, Weikert C, Boeing H. Healthy living is the best revenge: findings from the European prospective investigation into cancer and nutrition–Potsdam study. Arch Intern Med 2009 Aug 10;169(15):1355-1362 [FREE Full text] [doi: 10.1001/archinternmed.2009.237] [Medline: 19667296]
5. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
6. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
7. Di Cesare M, Khang Y, Asaria P, Blakely T, Cowan MJ, Farzadfar F, Lancet NCD Action Group. Inequalities in non-communicable diseases and effective responses. Lancet 2013 Feb 16;381(9866):585-597 [FREE Full text] [doi: 10.1016/S0140-6736(12)61851-0] [Medline: 23410608]
8. Sommer I, Griebler U, Mahlknecht P, Thaler K, Bouskill K, Gartlehner G, et al. Socioeconomic inequalities in non-communicable diseases and their risk factors: an overview of systematic reviews. BMC Public Health 2015 Sep 18;15(1):914 [FREE Full text] [doi: 10.1186/s12889-015-2227-y] [Medline: 26385563]
9. Veinot T, Mitchell H, Ancker J. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 01;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
10. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021 Nov 17;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]

XSL•FO
RenderX

11. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
12. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
13. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
14. Yao J, Lim N, Tan J, Matthias Müller A, Martinus van Dam R, Chen C, et al. Evaluation of a population‐wide mobile health physical activity program in 696 907 adults in Singapore. J Am Heart Assoc 2022 Jun 21;11(12):e022508 [FREE Full text] [doi: 10.1161/JAHA.121.022508] [Medline: 35699174]
15. Shen C, Wang MP, Chu JT, Wan A, Viswanath K, Chan SSC, et al. Health app possession among smartphone or tablet owners in Hong Kong: population-based survey. JMIR Mhealth Uhealth 2017 Jun 05;5(6):e77 [FREE Full text] [doi: 10.2196/mhealth.7628] [Medline: 28583905]
16. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
17. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
18. Stühmann LM, Paprott R, Heidemann C, Baumert J, Hansen S, Zahn D, et al. Health app use and its correlates among individuals with and without type 2 diabetes: nationwide population-based survey. JMIR Diabetes 2020 May 20;5(2):e14396 [FREE Full text] [doi: 10.2196/14396] [Medline: 32432555]
19. Jokisch MR, Schmidt LI, Doh M. Acceptance of digital health services among older adults: findings on perceived usefulness, self-efficacy, privacy concerns, ICT knowledge, and support seeking. Front Public Health 2022;10:1073756 [FREE Full text] [doi: 10.3389/fpubh.2022.1073756] [Medline: 36582385]
20. Wang C, Qi H. Influencing factors of acceptance and use behavior of mobile health application users: systematic review. Healthcare (Basel) 2021 Mar 22;9(3):357 [FREE Full text] [doi: 10.3390/healthcare9030357] [Medline: 33809828]
21. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
22. Klaver NS, van de Klundert J, van den Broek RJGM, Askari M. Relationship between perceived risks of using mHealth applications and the intention to use them among older adults in the Netherlands: cross-sectional study. JMIR Mhealth Uhealth 2021 Aug 30;9(8):e26845 [FREE Full text] [doi: 10.2196/26845] [Medline: 34459745]
23. Pujia C, Ferro Y, Mazza E, Maurotti S, Montalcini T, Pujia A. The role of mobile apps in obesity management: systematic review and meta-analysis. J Med Internet Res 2025 May 06;27:e66887 [FREE Full text] [doi: 10.2196/66887] [Medline: 40327853]
24. Park Y, Tak YW, Kim I, Lee HJ, Lee JB, Lee JW, et al. User experience and extended technology acceptance model in commercial health care app usage among patients with cancer: mixed methods study. J Med Internet Res 2024 Dec 18;26:e55176 [FREE Full text] [doi: 10.2196/55176] [Medline: 39693615]
25. Lundqvist A, Tolonen H. Healthy Finland survey as the latest uptake for 50 years of population health monitoring in Finland. Eur J Public Health 2023 Oct 24;33(Supplement_2):ckad160.720 [FREE Full text] [doi: 10.1093/eurpub/ckad160.720]
26. Sääksjärvi K, Elonheimo H, Ikonen J, Lehtoranta L, Parikka S, Härkänen T, et al. Cohort profile: Healthy Finland survey. Int J Epidemiol 2024 Dec 23;54(1):dyae166 [FREE Full text] [doi: 10.1093/ije/dyae166] [Medline: 39715488]
27. Harjumaa M, Absetz P, Ermes M, Mattila E, Männikkö R, Tilles-Tirkkonen T, et al. Internet-based lifestyle intervention to prevent type 2 diabetes through healthy habits: design and 6-month usage results of randomized controlled trial. JMIR Diabetes 2020 Aug 11;5(3):e15219 [FREE Full text] [doi: 10.2196/15219] [Medline: 32779571]
28. Lavikainen P, Mattila E, Absetz P, Harjumaa M, Lindström J, Järvelä-Reijonen E, et al. Digitally supported lifestyle intervention to prevent type 2 diabetes through healthy habits: secondary analysis of long-term user engagement trajectories in a randomized controlled trial. J Med Internet Res 2022 Feb 24;24(2):e31530 [FREE Full text] [doi: 10.2196/31530] [Medline: 35200147]
29. Weekly physical activity recommendations. UKK Institute. 2023. URL: https://ukkinstituutti.fi/en/products-services/physical-activity-recommendations/weekly-physical-activity-recommendation-for-over-65-year-olds/ [accessed 2023-12-07]
30. Lindström J, Aittola K, Pölönen A, Hemiö K, Ahonen K, Karhunen L, et al. Formation and validation of the Healthy Diet Index (HDI) for evaluation of diet quality in healthcare. Int J Environ Res Public Health 2021 Feb 28;18(5):2362 [FREE Full text] [doi: 10.3390/ijerph18052362] [Medline: 33670967]
31. A language and environment for statistical computing. R Foundation for Statistical Computing. 2023. URL: https://www.R-project.org/ [accessed 2025-11-14]

XSL•FO
RenderX

32. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
33. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. JMIR Mhealth Uhealth 2016 Mar 24;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
34. Paradis S, Roussel J, Bosson J, Kern J. Use of smartphone health apps among patients aged 18 to 69 years in primary care: population-based cross-sectional survey. JMIR Form Res 2022 Jun 16;6(6):e34882 [FREE Full text] [doi: 10.2196/34882] [Medline: 35708744]
35. Coughlin JW, Martin LM, Zhao D, Goheer A, Woolf TB, Holzhauer K, et al. Electronic health record–based recruitment and retention and mobile health app usage: multisite cohort study. J Med Internet Res 2022 Jun 10;24(6):e34191 [FREE Full text] [doi: 10.2196/34191] [Medline: 35687400]
36. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
37. Arias López MDP, Ong BA, Borrat Frigola X, Fernández AL, Hicklent RS, Obeles AJT, et al. Digital literacy as a new determinant of health: a scoping review. PLOS Digit Health 2023 Oct;2(10):e0000279 [FREE Full text] [doi: 10.1371/journal.pdig.0000279] [Medline: 37824584]
38. Foley K, Freeman T, Ward P, Lawler A, Osborne R, Fisher M. Exploring access to, use of and benefits from population-oriented digital health services in Australia. Health Promot Int 2021 Aug 30;36(4):1105-1115 [FREE Full text] [doi: 10.1093/heapro/daaa145] [Medline: 33367568]
39. Iyamu I, Gómez-Ramírez O, Xu AX, Chang H, Watt S, Mckee G, et al. Challenges in the development of digital public health interventions and mapped solutions: findings from a scoping review. Digit Health 2022;8:20552076221102255 [FREE Full text] [doi: 10.1177/20552076221102255] [Medline: 35656283]

Abbreviations

aOR: adjusted odds ratio
DHI: digital health intervention
e-service: electronic service
FDR: false discovery rate
LRT: likelihood ratio test
NCD: noncommunicable disease

Edited by L Buis; submitted 02.07.24; peer-reviewed by YF Luo, W Evans, V Mohan; comments to author 17.02.25; revised version received 30.05.25; accepted 23.10.25; published 09.12.25.

Please cite as:
Ruotsalainen I, Valtanen M, Kärsämä R, Umer A, Liedes H, Parikka S, Lundqvist A, Aittola K, Koivunen S, Pihlajamäki J, Vuorinen AL, Lindström J
Unraveling the Factors Associated With Digital Health Intervention Uptake: Cross-Sectional Study
JMIR Mhealth Uhealth 2025;13:e63896
URL: https://mhealth.jmir.org/2025/1/e63896
doi:10.2196/63896
PMID:

©Ilona Ruotsalainen, Mikko Valtanen, Riikka Kärsämä, Adil Umer, Hilkka Liedes, Suvi Parikka, Annamari Lundqvist, Kirsikka Aittola, Suvi Koivunen, Jussi Pihlajamäki, Anna-Leena Vuorinen, Jaana Lindström. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 09.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Internet-Based Cognitive Behavioral Therapy for Depressive Symptoms During Pregnancy by Using Real-World Data: Retrospective Cohort Study

Asuka Takae[1], BSN, MPH; Natsu Sasaki[1], MD, PhD; Kotaro Imamura[1,2], PhD; Daisuke Nishi[1], MD, PhD

[1]Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

[2]Department of Digital Mental Health, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

Corresponding Author:
Daisuke Nishi, MD, PhD
Department of Mental Health
Graduate School of Medicine
The University of Tokyo
7-3-1 Hongo
Bunkyo-ku, Tokyo, 113-0033
Japan
Phone: 81 3 5841 3364
Fax: 81 3 5841 3392
Email: d-nishi@m.u-tokyo.ac.jp

Abstract

Background: Approximately 1 out of 5 pregnant women develops depression. Internet-based cognitive behavioral therapy (iCBT) is an effective way to treat not only depression but also mild depressive symptoms or subthreshold depression. While numerous iCBT programs have been developed and tested through randomized controlled trials for various mental health conditions and specific populations, research on their effectiveness and application in the real world remains limited.

Objective: This study aimed to examine the effectiveness of a previously developed iCBT program implemented in an existing app for improving depressive symptoms among pregnant women in a real-world setting.

Methods: The previously developed iCBT program for preventing perinatal depression was already implemented in an existing app called Luna Luna Baby by MTI Ltd. The app aims to provide information to pregnant women about pregnancy and babies, and potential users can download it from the Japanese version of the Apple App Store or Google Play Store without any fee. The program does not require any additional fees. The log data stored on the app identified iCBT program users and nonusers, allowing us to conduct this retrospective cohort study. Data from September 2022 to September 2024 were extracted from the app after anonymous processing. The primary outcome was the score on the self-reported Edinburgh Postnatal Depression Scale (EPDS), which participants answer by themselves on the app. The exposure group was defined as completers of all 6 modules of the iCBT program. The nonexposure group was defined as users who did not use any module of the program and matched the baseline characteristics of the exposure group. The change in EPDS score before and after using the program was compared using effect sizes, and repeated 2-way ANOVA was conducted to test the difference between the exposure and nonexposure groups.

Results: Data from 119 women who completed the iCBT program and 448 pair-matched controls were selected. The average EPDS scores at baseline were 7.24 (SD 5.30) in the exposure group and 7.25 (SD 5.18) in the nonexposure group. After using the iCBT program, the group mean EPDS scores changed by −0.69 (SD 4.92) and +0.99 (SD 5.56) over time in the exposure and nonexposure groups, respectively (Cohen *d*=0.31, 95% CI 0.11-0.51). The repeated 2-way ANOVA showed statistical significance in the interaction terms between the groups and the measurement time points (*P*=.04).

Conclusions: The previously developed iCBT program showed a significant effect with a modest effect size on decreasing depressive symptoms among pregnant women in a real-world setting. Future research should attempt to minimize dropouts and increase participation in the program.

(JMIR Mhealth Uhealth 2025;13:e73512) doi:10.2196/73512

XSL•FO
RenderX

KEYWORDS

antenatal depressive symptom; prevention; internet-based cognitive behavioral therapy; real-world data; digital mental health; mobile health

Introduction

Depression is common among pregnant women, with a global prevalence of antenatal depression estimated at 20.7% [1]. In Japan, prevalence rates are 14% during midpregnancy and 16.3% in late pregnancy [1,2], suicide is the most frequent cause of perinatal death, and depression is the major risk factor for suicide [3]. Depressive symptoms during pregnancy affect not only pregnant women themselves but also fetuses and babies and can lead to adverse outcomes for them, such as impaired physical growth, preterm birth, low birth weight, decreased reactivity, changes in temperament, and behavioral disorders [4-6]. Therefore, it is essential to work on preventing depression among pregnant women.

Cognitive behavioral therapy is a way to prevent depression during pregnancy [7]. Internet-based cognitive behavioral therapy (iCBT) has recently increased accessibility to this therapy modality and convenience. Although iCBT was initially applied for treating depression, a recent study reported that iCBT was also effective for those who have mild depressive symptoms or subthreshold depression [8,9]. An iCBT program for preventing depression among pregnant women on a smartphone app has already been developed and investigated for its effectiveness in a randomized controlled trial (RCT) in Japan [10,11]. This RCT did not find an overall effect of the program in preventing perinatal depression; however, post hoc analyses suggest potential effectiveness among individuals with mild psychological distress. The program is now available as a free service on a popular Japanese app for anyone to use [12].

RCTs are conducted in a strictly controlled setting, which may be inconsistent with the real world. Significant differences have been reported between study populations from RCTs and study populations from real-world data, even using the same sampling criteria [13]. Engagement with mobile mental health apps is also lower in the real-world setting than in the clinical trial setting [14,15]. Therefore, it is necessary to confirm whether the developed iCBT program is effective not only in study settings but also in real-world settings to ensure its external validity [16-18]. Even though several studies have evaluated the actual use of digital interventions and identified their effectiveness for the treatment of depression and anxiety [19-23], we found no such efforts for iCBT programs for prevention or among pregnant women.

The objective of this observational retrospective cohort study was to investigate the effectiveness of iCBT use among pregnant women in improving depressive symptoms in a real-world setting.

Methods

Study Design and Data Setting

This observational study was conducted based on a previous RCT [10]. The users evaluated the iCBT program highly in terms of acceptability, appropriateness, fidelity, and satisfaction [24]. After the study, this developed iCBT program was implemented in the real world. The iCBT program was released in an existing app for perinatal women called Luna Luna Baby by MTI Ltd as a free standard feature in September 2022 [12]. Luna Luna Baby is a smartphone app that aims to provide pregnant women with information about pregnancy and babies. The app can be downloaded from the Japanese version of the Apple App Store or Google Play Store without any fee, and it does not require any additional fees to use the program on the app. A self-assessment service based on the Edinburgh Postnatal Depression Scale (EPDS) questionnaire was also implemented on the app at the same time, allowing pregnant women who use this app to check their current mental status whenever they want. The app routinely logs the results of users' self-checked EPDS scores. These data include not only the EPDS score and its measurement date but also basic information about the user at the time of EPDS measurement, such as age, weeks of pregnancy, and the number of times the user read the iCBT modules. This study used the sample from September 2022 to September 2024 extracted from these EPDS service log data after anonymous processing. There is no linkage across other databases. Given that user access typically decreases with increasing depth in a hierarchical app structure, we collected information on the number of visits to pages at each hierarchical level. We also collected information regarding any modifications made to the structure over the observational period. This allowed us to analyze user navigation patterns during this period. These data were all provided by MTI Ltd, the app developer.

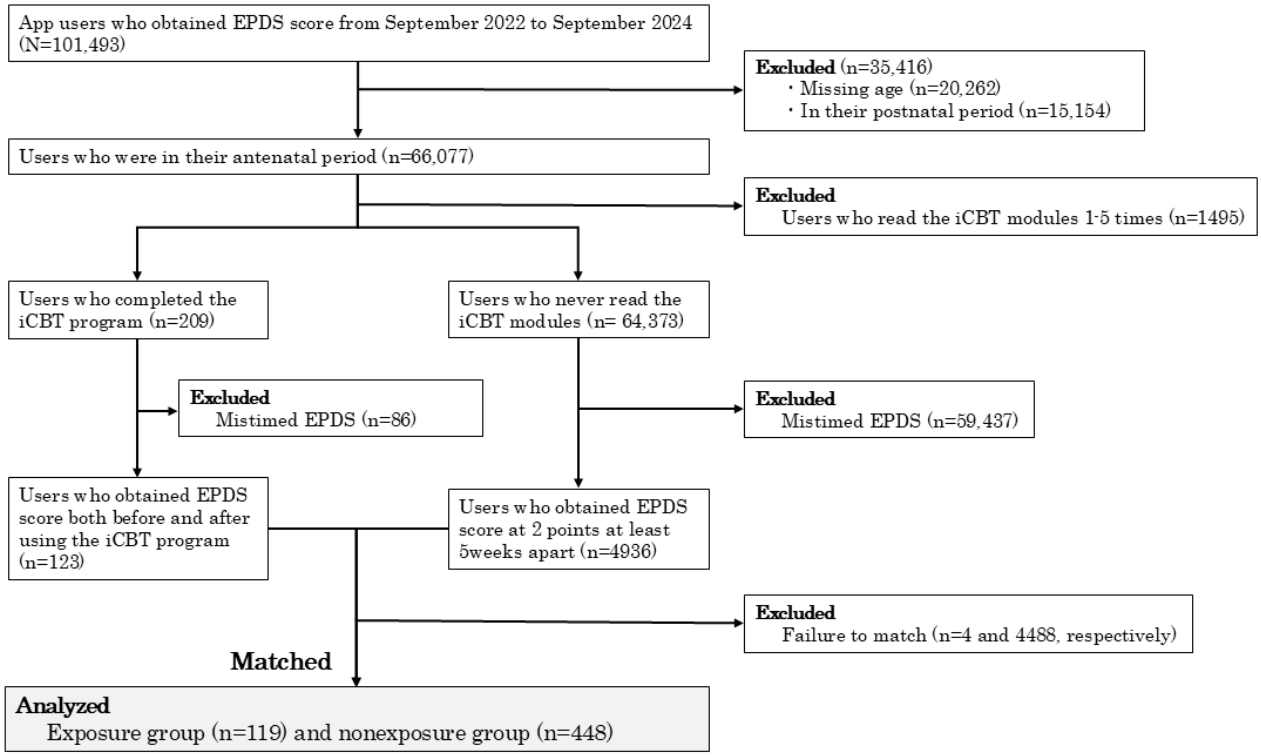
Participants and Measurements

Participant Selection

Pregnant women who had downloaded and used the Luna Luna Baby app from September 2022 to September 2024 were possible participants (N=101,493). Among them, those who were in their antenatal period were included (66,077/101,493, 65.1%). Those who were in their postnatal period or those with missing age data were excluded. The exposure group was defined as those who completed the iCBT program (209/66,077, 0.31%). The nonexposure group was defined as those who never read the iCBT modules (64,373/66,077, 97.42%). To compare changes in depressive symptoms, participants were limited to those who had EPDS scores from 2 time points, before and after exposure to the iCBT program (123/209, 58.9%). For the nonexposure group, participants were limited to those who had EPDS scores from 2 time points at least 5 weeks apart based on the estimated iCBT program duration (4936/64,373, 7.66%). The details of the selection are shown in Figure 1.

XSL•FO
RenderX

Figure 1. Participant selection flowchart. EPDS: Edinburgh Postnatal Depression Scale; iCBT: internet-based cognitive behavioral therapy.



Exposure: iCBT Program Use

Exposure Group

The iCBT program consists of a total of 6 modules, each including psychoeducation, case formulation based on a cognitive behavioral model, behavioral activation, self-compassion, mindfulness, and problem-solving. All modules were based on the topics that many women tend to be concerned about during pregnancy. The details of the module contents have been explained elsewhere [11]. The iCBT program within the app becomes accessible only after 16 weeks of pregnancy, after which users may access one module per week—each requiring approximately 5 minutes for completion—at a time convenient to the user. Although the program can be completed within 5 weeks at minimum, it can be used at any time as long as the users have the app downloaded. The program's progress is managed by clicking the "Read" button at the bottom of each module page, and the next module is opened 1 week after clicking the button on the previous module page. The number of modules read is recorded when the EPDS is measured, which enables us to identify the timing of program use from EPDS log data.

Nonexposure Group

A previous study has demonstrated that baseline factors such as maternal age and stress levels influence the trajectories of depressive symptoms from pregnancy to the postpartum period [25]. In particular, the severity of depressive symptoms at baseline has been shown to affect the efficacy of interventions [10,26]. Therefore, to ensure comparability between the exposure and nonexposure groups, participants in the nonexposure group were selected using matching based on the following baseline variables with the exposure group: age, weeks of pregnancy, and total EPDS score. The caliper widths were set at –2 to +2 for age, –1 to +1 for weeks of pregnancy, and –1 to +1 for total EPDS score following previous research [27-29].

The total EPDS score was added in this study as a matching variable because it was suggested that baseline severity of depressive symptoms affects intervention effects [10,26]. We randomly selected up to 4 controls per iCBT program completer without replacement, such that any individual selected as a control could not be chosen again for another case within the study [30,31].

Outcome: Antenatal Depressive Symptoms

To evaluate depressive symptoms during pregnancy, we used the Japanese version of the EPDS [32,33]. This scale consists of 10 questions about perinatal-specific depressive symptoms measured on a 4-point scale ranging from 0 to 3, with a total score ranging from 0 to 30. A higher score indicates more severe depressive symptoms. A cutoff of 12 or 13 is considered appropriate for antenatal depression [34]. In addition, a cutoff of 8 or 9 is often used clinically in Japan to screen for perinatal depression [32], so this cutoff was also considered. Although participants could answer the questionnaire only at given time points (eg, 16 to 20 weeks of pregnancy and 32 weeks of pregnancy) in the previous RCT [10], after the launching the service in public, app users can answer the EPDS whenever they wanted to check their mental status. The app displays advice to the users depending on the score obtained. When the score is 0 to 8, the advice is as follows: "There seems to be no problem with your mental status. Let's continue with the self-check of your mental status." When the score is 9 to 12, the advice is as follows: "You may not be feeling well. We recommend that you take extra care of yourself by getting

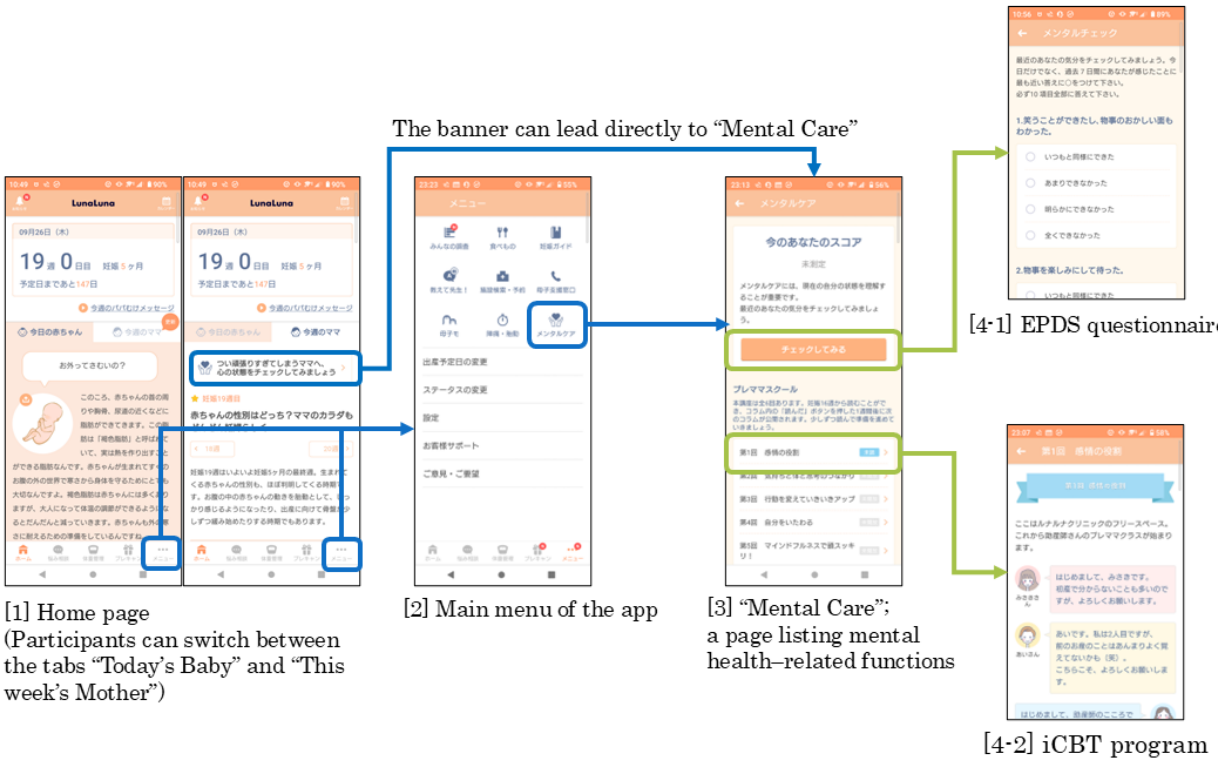
XSL•FO
RenderX

enough sleep, taking time for yourself, and doing activities that will refresh you. Please also use our Pre-mother School modules." When the score is 13 or higher, the advice is as follows: "It is possible that your mental status is not at its best. We recommend that you take better care of yourself than usual and consult with a trusted family member, friend, supervisor, or colleague. If you continue to be in the wrong mood, please consider consulting your obstetrician-gynecologist or public consultation service." These messages appear automatically, and there is no tracking of what actions participants take afterward. The app also encourages users to answer the EPDS using push notifications. All users regardless of exposure to iCBT receive a notification to answer the EPDS when they access the "Mental Care" section with the following message: "Do a self-check once every two weeks." For users who have previously completed an EPDS assessment, a follow-up notification is sent 2 weeks after their last measurement. These notifications are not linked to iCBT program use status and are delivered uniformly across users.

Process Evaluation

We reported the reach and completion rate of the iCBT program among users who downloaded the app. After launching the service to the public, both the iCBT program and the EPDS questionnaire were stored on a higher-level page named "Mental Care." The "Mental Care" page lists mental health–related functions, 2-levels-lower from the top page (Figure 2). In February 2023, the specifications were updated to include a banner that enables users to jump from the home page to the "Mental Care" page with a single click. The banner shows the following message: "For mothers who try too hard, let's check your mental state." Under these specifications, "reach" was defined as the proportion of those who visited the "Mental Care" page, iCBT program page, and EPDS page at least once each. The completion rate was defined as the proportion of those who completed all 6 modules.

Figure 2. Screenshots of the Luna Luna Baby app. EPDS: Edinburgh Postnatal Depression Scale; iCBT: internet-based cognitive behavioral therapy.



Covariates

Age and weeks of pregnancy at the time of the baseline EPDS score were measured and used as basic information. For service use information, the observational period was defined as the period between the baseline and follow-up time points when EPDS scores were measured because the log data did not include date information about iCBT program use. The number of additional EPDS measurements at time points within the observational period other than baseline and follow-up was also calculated.

Statistical Analysis

First, the baseline characteristics of the exposure and nonexposure groups were compared using standardized differences. Service use information was also compared between groups using 2-tailed *t* tests. We calculated changes in EPDS scores after program use, 95% CIs, and Cohen *d* effect sizes to assess the association between the iCBT program and depressive symptoms. Repeated-measure ANOVA was also conducted to identify the statistical significance of interaction terms between the groups and the measurement time points. Regarding the process evaluation, reach and completion rate were calculated using information on the number of users visiting the "Mental Care" page, iCBT program page, and EPDS questionnaire page

XSL•FO
RenderX

per month. All analyses were conducted using R (version 4.4.1; R Foundation for Statistical Computing), and $P<.05$ was set as statistically significant.

Ethical Considerations

This study was approved by the Research Ethics Committee of the Graduate School of Medicine and the Faculty of Medicine of the University of Tokyo (approval 2019150NI). Informed consent to collect data from the app was obtained from users who had an EPDS score after informing them that the data would be used for research purposes. The data from the app were anonymously processed. There was no compensation provided to participants for this study as we tracked natural use of the app.

Results

Participant Characteristics

Of the 101,493 women (173,712 records) who answered the EPDS on the app from September 2022 to September 2024, a total of 209 (0.2%; 1675 records) completed the 6 iCBT modules, and 64,373 (63.42%; 101,251 records) never read any of the iCBT modules. After excluding those without EPDS measurements before and after using the iCBT program or those who were excluded from the matching process either because they did not fall within the specified caliper range for one or more matching variables or because there were more than 4 eligible control candidates for a single exposed participant, 119 women who completed the iCBT program and 448 pair-matched controls were identified.

All variables showed standardized differences of <0.1, which suggests that baseline characteristics were balanced between the exposure and nonexposure groups after matching (Table 1). The mean age of the participants was 33.31 (SD 4.12) years and 33.04 (SD 3.85) years in the exposure and nonexposure groups, respectively. The mean weeks of pregnancy at baseline were 18.00 (SD 5.01) and 17.63 (SD 4.85) in the exposure and nonexposure groups, respectively. The mean EPDS scores were 7.24 (SD 5.30) and 7.25 (SD 5.18) in the exposure and nonexposure groups, respectively. Demographic data of the full sample (N=101,493 women; 173,712 records) are shown in Multimedia Appendix 1, with a mean EPDS score of 8.33 (SD 5.71) to 8.53 (SD 5.88). The prevalence of antenatal depressive symptoms above the EPDS cutoff of 12 or 13 was 15.9% (19/119) and 17.4% (78/448) of the participants in the exposure and nonexposure groups, respectively. The mean observational period was 11.67 (SD 5.80) weeks in the exposure group and 13.61 (SD 6.13) weeks in the nonexposure group ($P=.002$). The mean number of additional EPDS measurements during the observational period was 3.17 (SD 1.70) in the exposure group and 0.51 (SD 0.95) in the nonexposure group ($P<.001$).

Table 1. Baseline participant characteristics after matching.

Variable	Exposure group (n=119)	Nonexposure group (n=448)	Standardized difference
Age (y), mean (SD)	33.31 (4.12)	33.04 (3.85)	0.068
Pregnancy duration (wk), mean (SD)	18.00 (5.01)	17.63 (4.85)	0.075
EPDS[a] score (0-30), mean (SD)	7.24 (5.30)	7.25 (5.18)	0.002
EPDS score of ≥9, n (%)	40 (33.6)	147 (32.8)	0.017
EPDS score of ≥13, n (%)	19 (16.0)	78 (17.4)	0.039

[a]EPDS: Edinburgh Postnatal Depression Scale.

Antenatal Depressive Symptoms Before and After Using the iCBT Program

Table 2 and Figure 3 show the EPDS scores and the differences before and after using the program. The average EPDS score at baseline was 7.24 (SD 5.30) in the exposure group and 7.25 (SD 5.18) in the nonexposure group. Compared to before using the program, the EPDS scores changed by −0.69 and +0.99 in the exposure and nonexposure groups, respectively (Cohen $d=0.31$, 95% CI 0.11-0.51). The repeated-measure ANOVA showed statistical significance in interaction terms between the groups, and the measurement time points ($P=.04$).

Table 2. Changes in the depressive symptoms after the internet-based cognitive behavioral therapy intervention.

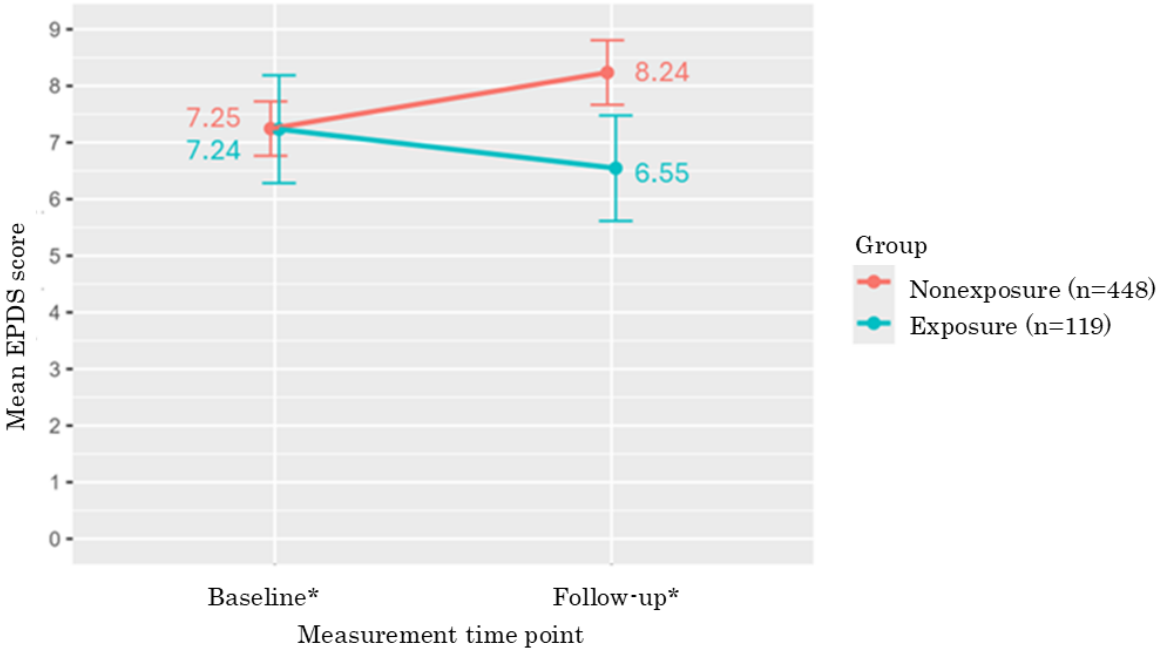
Variable	Exposure group (n=119), mean (SD; 95% CI)	Nonexposure group (n=448), mean (SD; 95% CI)	Effect size (Cohen *d*)
EPDS[a] score at baseline (0-30)	7.24 (5.30; 6.27 to 8.20)	7.25 (5.18; 6.76 to 7.73)	—[b]
EPDS score at follow-up (0-30)	6.55 (5.19; 5.60 to 7.49)	8.24 (6.16; 7.66 to 8.81)	—
Change in EPDS score	−0.69 (4.92; –1.58 to 0.20)	0.99 (5.56; 0.47 to 1.51)	0.31

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]Not applicable.

XSL•FO
RenderX

Figure 3. Changes in the EPDS score after iCBT program use. EPDS: Edinburgh Postnatal Depression Scale; iCBT: internet-based cognitive behavioral therapy. *The mean observational period between baseline and follow-up was 11.67 (SD 5.80) weeks in the exposure group and 13.61 (SD 6.13) weeks in the nonexposure group.



Process Evaluation

In terms of reach, 5.4% of app users who visited the "Mental Care" page. Of these, 26.5% visited the iCBT program page, and 67.5% visited the EPDS page. Regarding completion rate, 2.6% of those who visited the iCBT program page at least once completed all 6 modules. When the population was limited to those who visited the program page and clicked the "Read" button at least once, 32.3% completed the 6 modules.

Discussion

Principal Findings

This study examined the effectiveness of the iCBT program for pregnant women in preventing depressive symptoms during pregnancy in a real-world setting. This study found a statistically significant trend in the reduction in depressive symptoms after using the iCBT program in the exposure group compared with the nonexposure group. The effect size of the program for reducing depressive symptoms was 0.31 (95% CI 0.11-0.51). This is similar to a previous meta-analysis that reported an effect size of 0.28 for smartphone app interventions aimed at treating depressive symptoms [18], suggesting that this iCBT program also had a mild effect on preventing depressive symptoms among pregnant women.

This study showed that the average EPDS score at baseline was 7.24 (SD 5.30) in the exposure group and 7.25 (SD 5.18) in the nonexposure group, which indicated that the participants in this study had mild depressive symptoms under the EPDS cutoff score of 9. The finding that this iCBT program was effective among those with mild depressive symptoms was consistent with those of a previous meta-analysis and RCT [8,9], which showed that unguided iCBT was slightly less, but also similarly significantly effective for those who have mild depressive symptoms or subthreshold depression compared to guided iCBT. This unguided iCBT program for pregnant women may also play a role in preventing depression symptoms from exceeding the cutoff point. The baseline EPDS scores in this study were higher than the baseline EPDS scores reported in a previous RCT that evaluated the effectiveness of this iCBT program [10]. In that study, individuals with mild psychological distress at baseline had EPDS scores of 5.66 in the intervention group and 5.45 in the control group. The differences might be due to the participants' characteristics. The log data of this study included only those who voluntarily obtained EPDS scores at least once. A previous study showed that women with antenatal depression symptoms at baseline were more likely to use the screening features of pregnancy-related apps and spend more time using apps [35]. It is possible that the potential participants in this study had mild mental health issues and greater motivation to learn about mental health. When selecting final participants from the full sample to investigate the change in depressive symptoms at 2 time points, it is possible that some outliers who had more severe distress were dropped. It should be noted that this study was unable to observe the effects on those with particularly severe symptoms among those who obtained EPDS scores. Future studies are needed to detect the effectiveness of this program for those who have more severe distress.

The number of additional EPDS measurements during the observational period was higher in the exposure group than in the nonexposure group. It has been pointed out that repeated scale measurements themselves may have a transformative effect on cognition and behavior [36]. It has also been reported that repeated measurements of self-reported questionnaires have little impact that might raise serious concerns [37]. A study examining the effects of stress screening in the workplace and the efforts to improve the work environment found that stress screening alone did not improve workers' psychological distress.

XSL•FO
RenderX

Significant improvements were observed only when both stress screening and efforts to improve the work environment were implemented together [38]. Considering these studies, it may be reasonable to assume that the EPDS measurement itself in this study had little effect on depressive symptoms.

The average proportion of visitors to the "Mental Care" page among app users was 5.4%. This was similar to the results of a previous study, which reported a 4% open rate by daily active users of popular mental health apps in the real world [15]. Even though the iCBT program and the EPDS questionnaire were stored on the same "Mental Care" page, the visit rates were more than double, at 26.5% and 67.5%, respectively, than visits to the "Mental Care" page. This may suggest that many users who visited the "Mental Care" page were aiming to self-check their mental health using the EPDS questionnaire, not to use the iCBT program. Furthermore, the completion rate of the iCBT program was 2.6%. This result is consistent with previous reports that a digital mental health intervention could only attract participants to the first module and many people dropped out before reaching the final module [15,19,39]. However, when the population was limited to those who clicked the "Read" button at the bottom of the module pages at least once, which indicates at least some engagement, the completion rate increased to 32.3%. This was close to the completion rate of 37.2% in the previous RCT [10]. This suggests that the program may be feasible for a subset of app users in real-world settings. It is also possible that the requirement to click the "Read" button after reading the module contents unintentionally hindered program progress. Further study will be needed to reveal a way to increase participation in the program.

Limitations

This study has several limitations. First, the population selected for effectiveness evaluation did not represent all the app users as only 5.4% of app users visited the "Mental Care" page and 26.5% of them visited the iCBT program page. Furthermore, only 2.6% completed all 6 modules among those who visited the iCBT program page. To evaluate the effectiveness of the iCBT program with more app users, including stratified analysis by the number of iCBT modules completed, it may be necessary to improve the program's delivery environment. Although the number of participants was small, we were still able to evaluate the effectiveness of the iCBT program on those who completed it in a real-world environment. The final participants of this study were also limited compared to the app users who obtained EPDS score because of the study design, which compared the change in depressive symptoms between 2 time points among similar populations. While estimating the average treatment effect in the broader population is an important future direction, establishing the effectiveness of the program in engaged users is a necessary and informative first step. Second, there was a difference in the baseline and follow-up time points between individuals because participants were able to complete the EPDS at any time after accessing the app, resulting in variability in when these measurements were taken. The observational period was 2 weeks longer in the nonexposure group. As depressive symptoms are reported to remain stable between the middle and late stages of pregnancy [2], it is likely that the 2-week extension in the follow-up period for the control group did not substantially affect the outcomes. However, considering the relatively short time frame of this study, the possibility that this extension influenced the results to some extent cannot be entirely excluded. Third, there is a possibility of unmeasured confounding. The log data used in this study only included EPDS scores and related information; no socioeconomic factors, such as economic status, educational level, and employment status, could be considered. Previous studies have identified higher educational level and employment status as predictors of seeking mental health support through the internet [40]. It has also been suggested that, when digital mental health interventions are offered at no cost, there was not a reduction in willingness to use these interventions among individuals from socioeconomically disadvantaged groups, potentially contributing to a reduction in health disparities [41]. In this study, the iCBT program was available within the app completely free of charge, which may minimize the economic barrier. However, the engagement with mental health features or its effect may still be influenced by differences in health awareness associated with educational and employment backgrounds. Further research is warranted to clarify the impact of socioeconomic factors on effectiveness. Fourth, although the findings were statistically significant, their clinical relevance remains open to further consideration. Finally, despite this iCBT program targeting perinatal depression prevention, only the effect on antenatal depressive symptoms could be evaluated because few app users continued to use the EPDS measurement service after delivery.

Conclusions

An iCBT program targeting the prevention of perinatal depression was significantly effective in decreasing depressive symptoms among pregnant women who downloaded the app and engaged with it in a real-world setting. Future studies should explore efforts to reduce dropouts and increase participation in the program to help more pregnant women reduce their depressive symptoms.

Acknowledgments

The Japan Society for the Promotion of Science funded this work under a Grant-in-Aid for Scientific Research (A; 19H01073 and 23H00443 to DN). The study's sponsor had no role in the study design, data collection, data analysis, data interpretation, or report writing. The authors had access to the data and the final responsibility for submitting the paper. The authors thank all the pregnant women who used the program. They are grateful to MTI Ltd for providing the data from the app.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are not publicly available because data sharing was not approved by the institutional review board, but are available from the corresponding author on reasonable request.

Authors' Contributions

DN was in charge of this study, supervising the process and providing his expert opinion. AT and DN designed the study and analyzed the data. AT wrote the first draft of the manuscript. NS and KI ensured that questions related to the accuracy or integrity of all parts of the work were appropriately investigated and resolved. All authors approved the final version of the manuscript.

Conflicts of Interest

DN received personal fees outside the submitted work from Startia, Inc, and MD.net and an honorarium from Takeda Pharmaceutical Company and Otsuka Medical Devices Co, Ltd. NS reports personal fees from Medilio, Inc, outside the submitted work. KI is employed at the Department of Digital Mental Health, an endowment department supported with an unrestricted grant from 15 enterprises [42] outside the submitted work. The other author declares no other conflicts of interest.

Multimedia Appendix 1
Demographic data of the full sample.
[DOCX File , 18 KB - mhealth_v13i1e73512_app1.docx]

References

1. Yin X, Sun N, Jiang N, Xu X, Gan Y, Zhang J, et al. Prevalence and associated factors of antenatal depression: systematic reviews and meta-analyses. Clin Psychol Rev 2021 Feb;83:101932. [doi: 10.1016/j.cpr.2020.101932] [Medline: 33176244]
2. Tokumitsu K, Sugawara N, Maruo K, Suzuki T, Shimoda K, Yasui-Furukori N. Prevalence of perinatal depression among Japanese women: a meta-analysis. Ann Gen Psychiatry 2020;19:41 [FREE Full text] [doi: 10.1186/s12991-020-00290-7] [Medline: 32607122]
3. Takeda S, Takeda J, Murakami K, Kubo T, Hamada H, Murakami M, et al. Annual Report of the Perinatology Committee, Japan Society of Obstetrics and Gynecology, 2015: proposal of urgent measures to reduce maternal deaths. J Obstet Gynaecol Res 2017 Jan;43(1):5-7. [doi: 10.1111/jog.13184] [Medline: 28074548]
4. Glover V. Maternal depression, anxiety and stress during pregnancy and child outcome; what needs to be done. Best Pract Res Clin Obstet Gynaecol 2014 Jan;28(1):25-35. [doi: 10.1016/j.bpobgyn.2013.08.017] [Medline: 24090740]
5. Field T. Prenatal depression effects on early development: a review. Infant Behav Dev 2011 Feb;34(1):1-14. [doi: 10.1016/j.infbeh.2010.09.008] [Medline: 20970195]
6. Zhang L, Li P, Ge Q, Sun Z, Cai J, Xiao C, et al. Maternal prenatal depressive symptoms and fetal growth during the critical rapid growth stage. JAMA Netw Open 2023 Dec 01;6(12):e2346018 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.46018] [Medline: 38048129]
7. Yasuma N, Narita Z, Sasaki N, Obikane E, Sekiya J, Inagawa T, et al. Antenatal psychological intervention for universal prevention of antenatal and postnatal depression: a systematic review and meta-analysis. J Affect Disord 2020 Aug 01;273:231-239. [doi: 10.1016/j.jad.2020.04.063] [Medline: 32421608]
8. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
9. Taguchi K, Miyoshi M, Seki Y, Baba S, Shimizu E. Effectiveness of cognitive behavioral therapy provided through a web application for subthreshold depression, subthreshold insomnia, and subthreshold panic: open-labeled 6-arm randomized clinical trial pilot study. JMIR Form Res 2025 Feb 03;9:e63139 [FREE Full text] [doi: 10.2196/63139] [Medline: 39899369]
10. Nishi D, Imamura K, Watanabe K, Obikane E, Sasaki N, Yasuma N, et al. The preventive effect of internet-based cognitive behavioral therapy for prevention of depression during pregnancy and in the postpartum period (iPDP): a large scale randomized controlled trial. Psychiatry Clin Neurosci 2022 Nov;76(11):570-578 [FREE Full text] [doi: 10.1111/pcn.13458] [Medline: 36114656]
11. Nishi D, Imamura K, Watanabe K, Obikane E, Sasaki N, Yasuma N, et al. Internet-based cognitive-behavioural therapy for prevention of depression during pregnancy and in the post partum (iPDP): a protocol for a large-scale randomised controlled trial. BMJ Open 2020 May 17;10(5):e036482 [FREE Full text] [doi: 10.1136/bmjopen-2019-036482] [Medline: 32423941]
12. Luna Luna Baby adds mental care features to prevent pregnancy and postpartum depression!. MTI. 2022 Sep 15. URL: https://www.mti.co.jp/?p=31896 [accessed 2024-08-06]
13. Averitt AJ, Weng C, Ryan P, Perotte A. Translating evidence into practice: eligibility criteria fail to eliminate clinically significant differences between real-world and study populations. NPJ Digit Med 2020 May 11;3:67 [FREE Full text] [doi: 10.1038/s41746-020-0277-8] [Medline: 32411828]

XSL•FO
RenderX

14. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
15. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
16. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials 2015 Nov 03;16:495 [FREE Full text] [doi: 10.1186/s13063-015-1023-4] [Medline: 26530985]
17. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
18. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. World Psychiatry 2024 Feb;23(1):139-149 [FREE Full text] [doi: 10.1002/wps.21183] [Medline: 38214614]
19. Guiney H, Mahoney A, Elders A, David C, Poulton R. Internet-based cognitive behavioural therapy in the real world: naturalistic use and effectiveness of an evidence-based platform in New Zealand. Aust N Z J Psychiatry 2024 Mar;58(3):238-249 [FREE Full text] [doi: 10.1177/00048674231183641] [Medline: 37366145]
20. Fundoiano-Hershcovitz Y, Breuer Asher I, Ritholz MD, Feniger E, Manejwala O, Goldstein P. Specifying the efficacy of digital therapeutic tools for depression and anxiety: retrospective, 2-cohort, real-world analysis. J Med Internet Res 2023 Sep 22;25:e47350 [FREE Full text] [doi: 10.2196/47350] [Medline: 37738076]
21. Dzubur E, Yu J, Hoffman J, Painter S, James R, Shah B. The effect of a digital mental health program on anxiety and depression symptoms: retrospective analysis of clinical severity. JMIR Form Res 2023 Oct 03;7:e36596 [FREE Full text] [doi: 10.2196/36596] [Medline: 37788069]
22. Tsirmpas C, Nikolakopoulou M, Kaplow S, Andrikopoulos D, Fatouros P, Kontoangelos K, et al. A digital mental health support program for depression and anxiety in populations with attention-deficit/hyperactivity disorder: feasibility and usability study. JMIR Form Res 2023 Oct 11;7:e48362 [FREE Full text] [doi: 10.2196/48362] [Medline: 37819688]
23. Callahan C, Kimber J, Hu E, Tanner L, Kunkle S. The real-world impact of app-based mindfulness on headspace members with moderate and severe perceived stress: observational study. JMIR Mhealth Uhealth 2024 Mar 04;12:e52968 [FREE Full text] [doi: 10.2196/52968] [Medline: 38488513]
24. Obikane E, Sasaki N, Imamura K, Nozawa K, Vedanthan R, Cuijpers P, et al. Usefulness of Implementation Outcome Scales for Digital Mental Health (iOSDMH): experiences from six randomized controlled trials. Int J Environ Res Public Health 2022 Nov 27;19(23):15792 [FREE Full text] [doi: 10.3390/ijerph192315792] [Medline: 36497867]
25. Ahmed A, Feng C, Bowen A, Muhajarine N. Latent trajectory groups of perinatal depressive and anxiety symptoms from pregnancy to early postpartum and their antenatal risk factors. Arch Womens Ment Health 2018 Dec;21(6):689-698. [doi: 10.1007/s00737-018-0845-y] [Medline: 29654413]
26. Terhorst Y, Kaiser T, Brakemeier EL, Moshe I, Philippi P, Cuijpers P, et al. Heterogeneity of treatment effects in internet- and mobile-based interventions for depression: a systematic review and meta-analysis. JAMA Netw Open 2024 Jul 01;7(7):e2423241 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.23241] [Medline: 39023887]
27. Feldman DE, Vinet É, Bérard A, Duffy C, Hazel B, Meshefedjian G, et al. Heart disease, hypertension, gestational diabetes mellitus, and preeclampsia/eclampsia in mothers with juvenile arthritis: a nested case-control study. Arthritis Care Res (Hoboken) 2017 Feb;69(2):306-309. [doi: 10.1002/acr.22925] [Medline: 27111101]
28. Stephansson O, Johansson K, Söderling J, Näslund I, Neovius M. Delivery outcomes in term births after bariatric surgery: population-based matched cohort study. PLoS Med 2018 Sep;15(9):e1002656 [FREE Full text] [doi: 10.1371/journal.pmed.1002656] [Medline: 30256796]
29. Munger KL, Hongell K, Cortese M, Åivo J, Soilu-Hänninen M, Surcel HM, et al. Epstein-barr virus and multiple sclerosis risk in the finnish maternity cohort. Ann Neurol 2019 Sep;86(3):436-442 [FREE Full text] [doi: 10.1002/ana.25532] [Medline: 31226219]
30. Iwagami M, Shinozaki T. Introduction to matching in case-control and cohort studies. Ann Clin Epidemiol 2022;4(2):33-40 [FREE Full text] [doi: 10.37737/ace.22005] [Medline: 38504854]
31. Fabre C, Pauly V, Baumstarck K, Etchecopar-Etchart D, Orleans V, Llorca PM, et al. Pregnancy, delivery and neonatal complications in women with schizophrenia: a national population-based cohort study. Lancet Reg Health Eur 2021 Nov;10:100209 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100209] [Medline: 34806069]
32. Okano T. Validation and reliability of a Japanese version of the EPDS. Arch Psychiatr Diagn Clin Eval 1996;7:525 [FREE Full text]
33. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]

XSL•FO
RenderX

34. Usuda K, Nishi D, Okazaki E, Makino M, Sano Y. Optimal cut-off score of the Edinburgh Postnatal Depression Scale for major depressive episode during pregnancy in Japan. Psychiatry Clin Neurosci 2017 Dec;71(12):836-842 [FREE Full text] [doi: 10.1111/pcn.12562] [Medline: 28767198]
35. Mo Y, Gong W, Wang J, Sheng X, Xu DR. The association between the use of antenatal care smartphone apps in pregnant women and antenatal depression: cross-sectional study. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e11508 [FREE Full text] [doi: 10.2196/11508] [Medline: 30497996]
36. French DP, Sutton S. Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it? Br J Health Psychol 2010 Sep;15(Pt 3):453-468. [doi: 10.1348/135910710X492341] [Medline: 20205982]
37. Arslan RC, Reitz AK, Driebe JC, Gerlach TM, Penke L. Routinely randomize potential sources of measurement reactivity to estimate and adjust for biases in subjective reports. Psychol Methods 2021 Apr;26(2):175-185. [doi: 10.1037/met0000294] [Medline: 32584065]
38. Imamura K, Asai Y, Watanabe K, Tsutsumi A, Shimazu A, Inoue A, et al. Effect of the National Stress Check Program on mental health among workers in Japan: a 1-year retrospective cohort study. J Occup Health 2018 Jul 25;60(4):298-306 [FREE Full text] [doi: 10.1539/joh.2017-0314-OA] [Medline: 29669966]
39. Schueller SM, Torous J. Scaling evidence-based treatments through digital mental health. Am Psychol 2020 Nov;75(8):1093-1104 [FREE Full text] [doi: 10.1037/amp0000654] [Medline: 33252947]
40. Villagonzalo KA, Arnold C, Farhall J, Rossell SL, Foley F, Thomas N. Predictors of overall and mental health-related internet use in adults with psychosis. Psychiatry Res 2019 Aug;278:12-18. [doi: 10.1016/j.psychres.2019.05.034] [Medline: 31132571]
41. Lorenzo-Luaces L, Wasil A, Kacmarek CN, DeRubeis R. Race and socioeconomic status as predictors of willingness to use digital mental health interventions or one-on-one psychotherapy: national survey study. JMIR Form Res 2024 Apr 11;8:e49780 [FREE Full text] [doi: 10.2196/49780] [Medline: 38602769]
42. Participating companies and cooperative courses. Graduate School of Medicine, The University of Tokyo. URL: https://dmh.m.u-tokyo.ac.jp/c [accessed 2025-05-10]

Abbreviations

EPDS: Edinburgh Postnatal Depression Scale
iCBT: internet-based cognitive behavioral therapy
RCT: randomized controlled trial

Edited by L Buis; submitted 10.03.25; peer-reviewed by S Saarni, R Pearson; comments to author 10.06.25; revised version received 24.11.25; accepted 24.11.25; published 11.12.25.

Please cite as:
Takae A, Sasaki N, Imamura K, Nishi D
Effectiveness of Internet-Based Cognitive Behavioral Therapy for Depressive Symptoms During Pregnancy by Using Real-World Data: Retrospective Cohort Study
JMIR Mhealth Uhealth 2025;13:e73512
URL: https://mhealth.jmir.org/2025/1/e73512
doi:10.2196/73512
PMID:

©Asuka Takae, Natsu Sasaki, Kotaro Imamura, Daisuke Nishi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Text Messaging Versus Postal Reminders to Improve Participation in a Colorectal Cancer Screening Program: Randomized Controlled Trial

Nuria Vives[1,2], MPH; Gemma Binefa[1,2,3], PhD; Noemie Travier[1,2], PhD; Albert Farre[4], PhD; Jon Aritz Panera[1,2], BMATH; Berta Casas[1,2], BS; Carmen Vidal[1,2,3], MPH; Gemma Ibáñez-Sanz[5,6], PhD; Montse Garcia[1,2,3], PhD; M-TICS research group[7]

[1]Catalan Institut of Oncology, Hospitalet del Llobregat, Spain

[2]Early Detection of Cancer Research Group, Epidemiology, Public Health, Cancer Prevention and Palliative Cures Program, Bellvitge Biomedical Research Institute, Hospitalet del Llobregat, Spain

[3]Consortium for Biomedical Research in Epidemiology and Public Health (CIBEResp), Madrid, Spain

[4]School of Health Sciences, University of Dundee, Dundee, United Kingdom

[5]Gastroenterology Department, Bellvitge University Hospital, Hospitalet del Llobregat, Spain

[6]Colorectal Cancer Group, Molecular Mechanisms and Experimental Therapeutics in Oncology Program, Bellvitge Biomedical Research Institute, Hospitalet del Llobregat, Spain

[7]see Acknowledgments

Corresponding Author:
Montse Garcia, PhD
Catalan Institut of Oncology
Gran Via de L'Hospitalet 199-203
Hospitalet del Llobregat, 08908
Spain
Phone: 34 93 260 72 05
Email: mgarcia@iconcologia.net

Abstract

Background: Mobile phone SMS text message reminders have shown moderate effects in improving participation rates in ongoing colorectal cancer screening programs.

Objective: This study aimed to assess the effectiveness of SMS text messages as a replacement for routine postal reminders in a fecal immunochemical test–based colorectal cancer screening program in Catalonia, Spain.

Methods: We conducted a randomized controlled trial among individuals aged 50 to 69 years who were invited to screening but had not completed their fecal immunochemical test within 6 weeks. The intervention group (n=12,167) received an SMS text message reminder, while the control group (n=12,221) followed the standard procedure of receiving a reminder letter. The primary outcome was participation within 18 weeks of the invitation. The trial was stopped early, and a recovery strategy was implemented for nonparticipants in the intervention group. We performed a final analysis to evaluate the impact of the recovery strategy on the main outcome of the trial. Participation was assessed using a logistic regression model adjusting for potential confounders (sex, age, and deprivation score index) globally and by screening behavior.

Results: The trial was discontinued early in September 2022 due to the results of the interim analysis. The interim analysis included 5570 individuals who had completed 18 weeks of follow-up (intention-to-treat). The SMS text message group had a participation rate of 17.2% (477/2781), whereas the control group had a participation rate of 21.9% (610/2789; odds ratio 0.71, 95% CI 0.62-0.82; *P*<.001). As a recovery strategy, 7591 (72.7%) out of 10,442 nonparticipants in the SMS text message group had an open screening episode and received a second reminder by letter, reaching a participation rate of 23% (1748/7591). The final analysis (N=24,388) showed a participation rate of 29.3% (3561/12,167) in the intervention group, which received 2 reminders, while the participation rate was 26.5% (3235/12,221) in the control group (odds ratio 1.16, 95% CI 1.09-1.23; *P*<.001).

Conclusions: Replacing SMS text messages with reminder letters did not increase the participation rate but also led to a decline in participation among nonparticipants 6 weeks after the invitation. However, sending a second reminder by letter significantly increased participation rates among nonparticipants within 6 weeks in the SMS text message group compared with those who

XSL•FO
RenderX

received 1 postal reminder (control group). Additional research is essential to determine the best timing and frequency of reminders to boost participation without being intrusive in their choice of participation.

Trial Registration: ClinicalTrials.gov NCT04343950; https://www.clinicaltrials.gov/study/NCT04343950

(JMIR Mhealth Uhealth 2025;13:e64243) doi:10.2196/64243

KEYWORDS

text message; mobile health; colorectal cancer; screening; participation; reminders; text messaging; colorectal cancer screening; fecal immunochemical test

Introduction

Colorectal cancer (CRC) is a significant global health concern, with millions of new cases diagnosed and thousands of lives lost each year [1,2]. Early detection through screening is vital for reducing the morbidity and mortality associated with CRC [3,4]. However, despite the proven benefits of early detection, participation rates in CRC screening programs remain suboptimal in most parts of the world [5]. A recent systematic review that analyzed the effectiveness of invitation schemes in fecal occult blood test (FOBT)–based CRC screening found that reminders had a positive impact, increasing participation by 8.5% to 15.8% points depending on the method used [6]. The current European guidelines for CRC screening recommend using reminders through evidence-based communication channels such as letters and telephone reminders to improve cancer screening rates; however, some of these approaches can be costly [4]. Therefore, in order to broaden the reach of screening efforts, there is a need to explore and establish innovative and cost-effective strategies aimed at improving screening uptake among the target population.

In the contemporary era of digital communication, mobile phones have become indispensable tools that can influence various aspects of health care delivery. SMS text messaging has emerged as a versatile and convenient channel for health care professionals to connect with patients. Its utility lies in its accessibility, affordability, and effectiveness as a platform for conveying health-related information, reminders, and motivation [7-9]. SMS text messaging has also shown promising results in addressing some challenges presented by CRC screening programs to overcome certain traditional barriers. A study conducted in Israel found that sending SMS reminders increased participation by 1.8% points compared with no reminder [10]. However, a study in the United Kingdom showed that adding an SMS reminder to the standard letter reminder resulted in an increase in participation only among first-time invitees [11]. In addition, a systematic review of SMS text message interventions for cancer screening found that they had a small effect on CRC screening rates, ranging from 0.6% to 3.3% [12]. There is, therefore, a need for further research to explore the potential of this approach more comprehensively.

This randomized controlled trial (RCT) assessed the effectiveness of SMS text messages versus postal reminders in a fecal immunochemical test (FIT)–based CRC screening program and is part of the larger Mobile Phone Messaging as a Tool to Improve Cancer Screening (M-TICS) study, which aims to evaluate the impact of different SMS text messaging interventions on participation in the ongoing colorectal and breast cancer screening programs in the metropolitan area of Barcelona, Catalonia, Spain [13].

Methods

Study Design

We conducted an RCT that compared an SMS text message reminder with the standard reminder procedure (a letter sent by postal mail) in a FIT-based CRC screening program. Recruitment was conducted from May 9, 2022, to September 9, 2022. The trial was stopped early, and we launched a recovery strategy for nonparticipants in the intervention group.

Ethical Considerations

The study received ethical approval from the Ethics Committee of the Bellvitge University Hospital (reference PR042/20), which waived the requirement to obtain the participant's signature as part of the consent process, as the intervention was a minor variation on the invitation practice. The study was performed under Good Clinical Practice and the Declaration of Helsinki. The data utilized in this study were de-identified prior to analysis to ensure the protection of individuals' privacy. The participants in this study did not receive any form of compensation.

Setting

The Catalan Institute of Oncology screening hub, as part of the CRC screening program in Catalonia, Spain, manages a biennial FIT-based program for CRC in northern and southern metropolitan areas of Barcelona and covers a target population of 502,348 men and women aged 50-69 years (January 1, 2022). The hub identifies individuals due for screening from the Central Register of Insured Persons of the Catalan Health Service. Eligible individuals receive an invitation letter to pick up a FIT kit at any pharmacy participating in the CRC program. A reminder invitation letter is sent to those nonparticipants in the sixth week. A total of 10 weeks after the reminder letter, the screening episode is automatically closed as a nonparticipant if no response is received. Community pharmacies collect and send completed FIT kits to their allocated laboratory to be processed. Individuals with positive FIT results are offered a diagnostic colonoscopy. If an invitation letter is returned to our hub due to an incorrect mailing address, we verify its current and correct status. If the address is wrong, we send a new letter to the updated address.

Participants and Randomization

Eligible individuals were men and women who received an invitation letter but did not participate within 6 weeks. Simple

XSL•FO
RenderX

randomization was performed to allocate the participants. We designed an application using JavaScript's built-in *Math.random* function to select and randomize eligible individuals in a 1:1 ratio stratified by previous screening behavior to either the intervention or the control group. From May 9, 2022, onward, eligible individuals were randomized to the intervention weekly until the target sample size was achieved. Individuals without a registered mobile phone were excluded. No blinding was considered at any step. However, the end point of this study did not require subjective judgment.

Intervention Description

Individuals randomly assigned to the control group received the standard reminder procedure: a reminder letter 6 weeks after the invitation. Individuals randomly assigned to the intervention group received an SMS text message reminder 6 weeks after the invitation instead of the reminder letter. SMS text messages were bidirectional (enabling 2-way messaging) and fully automated delivery through a platform. The screening hub staff members managed the incoming individual responses. The initial of the first name and the entire last name were part of the SMS text message. A link to request a hard copy of the original invitation letter was provided, as this includes a unique code needed to pick up the FIT kit in their local pharmacy. A reminder letter was sent to individuals whose SMS text messages failed to be delivered (Multimedia Appendix 1).

Outcomes and Baseline Variables

The primary outcome was the screening participation defined as individuals who completed an adequate FIT kit (positive or negative) within 18 weeks of the invitation. When the laboratory is unable to analyze a test (spoilt kit or technical failure), the participant is asked to repeat the FIT. If they refuse, the final test result is coded as indeterminate. Baseline variables were sex, age at the invitation, previous screening behavior (first-time invitees, previous nonparticipants, and previous screenees), and tertiles of deprivation score (DS) index based on the individual's Catalan primary health care referral area [14].

Sample Size

The sample size was estimated to detect differential effects on previous screening behavior (first-time invitees, previous nonparticipants, and previous screenees). We hypothesized that SMS text message reminders would boost the participation of first-time invitees. A small, nonsignificant effect was expected for previous nonparticipants or previous screenees.

We considered a proportion difference in participation of 3% points between the intervention and control groups for each previous screening stratum.

Participation rates after 6 weeks from the invitation letter of 9%, 17%, and 52% were considered for first-time invitees, previous nonparticipants, and previous screenees (based on our retrospective data). We anticipated that 10% of the phone numbers would be wrongly recorded. Finally, accepting an α risk of .05 and a power of 0.90 in a 2-tailed test, the total sample size was estimated at 25,572 individuals.

Interim Analysis

We performed an interim analysis when about 25% of the individuals recruited had completed the 18-week follow-up period. The criterion we used to stop the trial early was based on the Haybittle-Peto boundary, which sets a threshold of $P<.001$ when 1 interim analysis is performed while maintaining the α level and statistical power determined in the sample size calculation [15].

Recovery Strategy

A recovery strategy was designed to be implemented if results from interim analyses met the criteria for stopping the intervention. The recovery strategy involved a postal reminder for all nonrespondents within the intervention group if they were still eligible to participate (SMS text message reminder received within the last 10 weeks). This strategy aimed to reduce nonparticipation in the intervention group and, therefore, mitigate the impact of the nonresponse of the intervention group and prevent ethical issues.

Statistical Analysis

All primary analyses were conducted as an intention-to-treat analysis, including all eligible individuals, regardless of whether they received the complete or partial intended intervention. We compared baseline characteristics between the study groups using the 2-tailed Student *t* test for continuous data and chi-square tests for categorical data to identify imbalances during randomization. Similarly, we made comparisons of baseline characteristics in trial groups based on previous screening behavior. The effectiveness of the SMS text message reminders in participation was assessed using a logistic regression model adjusting for potential confounders (sex, age, and DS index) globally and by screening behavior. Results were reported as odds ratio (OR) and 95% CIs. A 2-sided *P* value of less than .05 was considered statistically significant. We also performed a post hoc subgroup analysis according to sex, age, and DS index. For the post hoc comparisons, we applied the Bonferroni correction by setting a significance level of $P<.03$ when 2 tests were conducted (for each sex and each age group) and a threshold of $P<.02$ when 3 tests were performed (one for each DS index). All the analyses were performed using Stata (version 18.0; StataCorp LP).

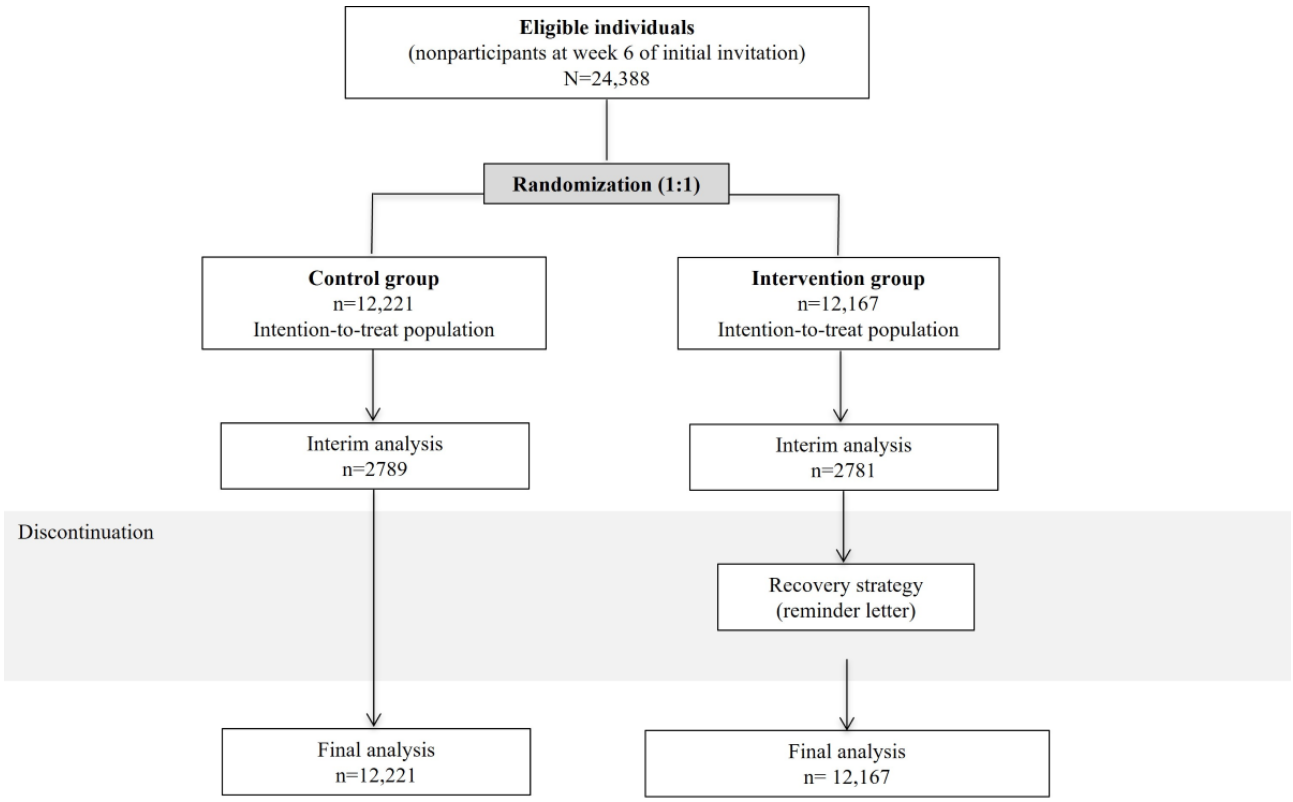
Results

Participants

Between May 9 and September 9, 2022, a total of 24,388 individuals were enrolled in this study. A total of 12,167 individuals were randomly assigned to the SMS text message group and 12,221 to the control group (Figure 1). SMS text messages failed to be delivered to 457 (3.8%) out of 12,167 individuals in the intervention group, and letters failed to be delivered to 625 (5.1%) out of 12,221 individuals in the control group. Furthermore, a total of 113 individuals who were lost follow-up due to changes in their individual addresses were classified as nonparticipants. The intention-to-treat approach included all individuals selected and randomized, regardless of their final status.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of reminder intervention to participate in a colorectal cancer screening program.



Interim Analysis

The interim analysis included a sample of 5570 individuals who had completed the 18 weeks of follow-up, thus representing 23% of the total 24,388 individuals recruited. Among these 5570 individuals, 2825 (50.7%) were women, the mean age was 57.1 (SD 5.78) years, 2123 (38.1%) were from the high DS area, and 2613 (46.9%) had previously undergone a screening test. Baseline characteristics were similar in both groups (Multimedia Appendix 2). The participation rate of nonparticipants within 6 weeks of invitation was significantly higher in the letter group (610/2789, 21.9%) compared with the SMS text message group (477/2781, 17.2%) at 18 weeks, with an OR of 0.71 (95% CI 0.62-0.82). The stratified analysis by previous screening behavior also showed differences in participation between the two groups, favoring the letter group (Table 1).

Table 1. Interim analysis of the participation by previous screening behavior among nonparticipants within 6 weeks of invitation (intention-to-treat).

Previous screening behavior	Letter, n/N (%)	SMS text message, n/N (%)	OR[a] (95% CI)	*P* value
Previous screenees	475/1315 (36.1)	400/1298 (30.8)	0.79 (0.67-0.92)	.004
First-time invitees	135/1474 (9.2)	77/1483 (5.2)	0.55 (0.4-0.76)	<.001
Previous nonparticipants	32/651 (4.9)	11/585 (1.9)	0.38 (0.19-0.76)	.004
Global	610/2789 (21.9)	477/2781 (17.2)	0.71 (0.62-0.82)	<.001

[a]OR: odds ratio.

Recovery Strategy

At the time of the trial interruption, we identified 10,431 individuals who did not participate in the SMS text message group. Of these individuals, 7591 (72.7%) were still eligible to participate as they received a reminder (SMS text message) fewer than 10 weeks ago. Therefore, we sent a second reminder by letter to encourage them to participate and help mitigate the negative impact of the SMS text message (Figure 2). The time of sending reminder letters varied between 11 and 103 days after receiving the SMS text message. The screening participation rate among those who received a second reminder by letter was 23% (1748/7591). However, participation was higher among those to whom we sent the reminder letter fewer than 30 days before (528/1652, 31%) than when the reminder letter was sent more than 60 days after the initial SMS text message reminder (594/3221, 18.4%; Table 2)

XSL•FO
RenderX

Figure 2. Diagram of the recovery strategy launched in the intervention group among individuals with an open screening episode. If no response is received 6 weeks after the reminder letter, the screening episode is automatically closed as a nonparticipant.

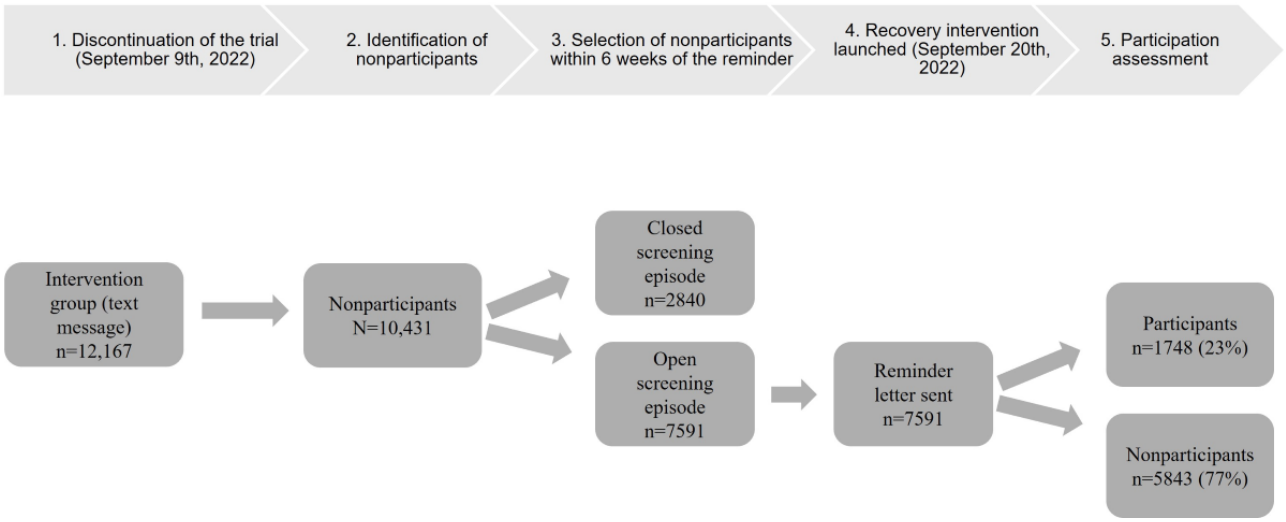


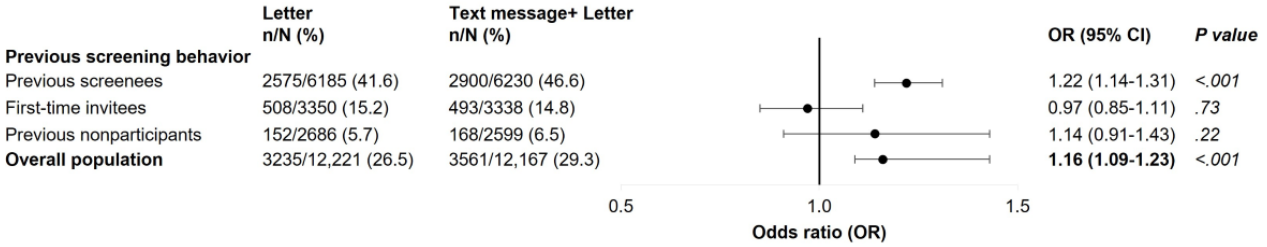
Table 2. Participation in the intervention group by the time the recovery strategy was launched following the SMS text message reminder.

	Participants, n/N (%)	*P* value
Fewer than or equal to 30 days	528/1652 (32)	<.001
Between 31 and 60 days	626/2718 (23)	<.001
More than 60 days	594/3221 (18.4)	<.001
Global	1748/7591 (23)	__[a]

[a]Not applicable.

The overall population baseline characteristics distribution was similar in both trials' groups (Multimedia Appendix 2). When stratified by previous screening behavior, overall population baseline characteristics also remained similar (Multimedia Appendix 3). According to the data presented in Figure 3, the sending of a reminder letter as a recovery strategy for nonparticipants from the intervention group resulted in a significantly higher final screening participation rate of 29.3% (3561/12,167) as compared with the letter group (3235/12,221, 26.5%), with an OR of 1.16 (95% CI 1.09-1.23). The stratified analysis by previous screening behavior showed a higher participation rate in the SMS text message group after the recovery strategy was launched compared with the letter group among previous screenees (OR 1.22, 95% CI 1.14-1.31). No significant differences were observed in participation rates of first-time invitees and previous nonparticipants after sending the reminder letter in the intervention group compared with the letter group (Figure 3).

Figure 3. Impact of the recovery strategy on participation globally and by screening behavior (intention-to-treat). The analysis was adjusted for sex, age, and deprivation score, considering the letter group as a reference.



Post hoc subgroup analyses showed that the recovery strategy of sending a reminder letter to previous nonparticipants in the SMS text message group significantly increased overall final participation, regardless of sex and age group, compared with the control group. However, when compared with the control group, significant increases in participation in the intervention group were only observed in the first and third tertile of the DS index (Multimedia Appendix 4).

Discussion

Principal Findings

This 2-arm RCT examined an SMS text message reminder intervention versus the usual reminder letter to nonparticipants in a FIT-based CRC screening program within 6 weeks of the invitation. The interim analysis showed that the participation rate was 4.5% points lower when an SMS text message reminder

XSL•FO
RenderX

was sent than the usual reminder letter. This decline in participation in the intervention group was consistent when stratified by previous screening behavior. As a result, the intervention was discontinued, and a recovery strategy was implemented, which involved sending a second reminder by letter to nonparticipants in the intervention group. These combined reminders (SMS text message followed by a letter) led to an increase of 2.8% points in the final participation in the intervention group after the recovery strategy compared with the usual reminder letter group.

Previously published studies have shown variable results regarding the effectiveness of SMS text message reminders for increasing uptake in CRC screening, depending on the invitation scheme used. Our program's global participation rate is around 45%. Out of the total participants, 70% participated within 6 weeks, while the remaining 30% required a reminder to participate. Therefore, using reminders is crucial to ensure acceptable levels of participation. As far as we know, no studies compared the impact of replacing a letter reminder with an SMS text message reminder. However, a study in the United Kingdom found that adding a second reminder through SMS text message did not improve overall participation compared with a reminder letter. Still, it did show a small benefit for newly screened participants [11]. In contrast, our study showed that replacing a letter with SMS text message reminders had a negative effect on both global participation and on each stratum of previous screening behavior. First-time invitees are crucial population subgroups for boosting participation in cancer screening programs because they are more likely to adhere to regular screening if engaged positively. On the other hand, previous nonparticipants have missed earlier opportunities, and efforts to engage them might require more intensive interventions, such as addressing specific concerns or barriers. Previous screenees have already participated, indicating they understand the importance of screening. Efforts to boost their participation might be less critical since they are likely to continue attending screenings without additional encouragement.

The obtained results were unexpected, as SMS text messaging is currently a routine part of cancer screening programs in the metropolitan Area of Barcelona. For example, SMS text messaging has been used as an appointment reminder for screening mammography since 2019, with an improvement in attendance rates [16]. Also, in other studies of the M-TICS project, we found that an SMS text message reminder to complete the FIT kit increased participation among individuals who had picked up the kit from the pharmacy but did not return it within 14 days [17]. There are different possible explanations for this unexpected result. First, letters (invitations or reminders) from our CRC screening program contain specific codes required for picking up FIT kits from the pharmacy. These codes serve as a unique identification for the recipients and are necessary for acquiring the kits. This requirement may have created a logistical barrier to access screening in the intervention group because the SMS text message alone did not allow them to pick up the test kit [18]. As a result, they had to present the invitation letter to the pharmacy, and if they were unable to locate it, they had to contact the screening hub to request a new invitation letter to be sent to them. Second, pharmacists who collaborated with the CRC screening program were informed by email that SMS text message reminders would be sent as part of a study to improve participation in the program. However, it is possible that not all the pharmacy clerks received the information, or its content was not sufficiently clear instructions about what they should explain to individuals who show up at their pharmacy with an SMS text message to pick up a FIT kit. This lack of dissemination may be partly responsible for the failure of the SMS text message approach. Providing precise and easy-to-follow instructions to all involved parties is essential to ensure better participation rates in similar programs.

In our study, we randomly assigned nonparticipants to receive a single reminder through letter or SMS text message at 6 weeks. When we stopped the intervention and sent a second reminder by letter to those nonparticipants in the SMS text message arm, we finally reached higher participation in the intervention than those in the letter arm, where individuals only received one reminder to participate in our CRC screening program. These findings are of particular importance as they highlight that the use of more than 1 reminder, such as SMS text messaging and postal delivery, can be more effective than using only one. In this regard, 2 previous studies found that the sending of 2 reminders combined with a prenotification message significantly increased participation rates compared with a single reminder [19,20]. We should investigate the potential impact of sending more than one SMS text message reminder in our program, carefully establishing the timing and frequency of reminders to maximize effectiveness without causing annoyance or being intrusive in their choice of participation [21].

Conclusions

Although our initial attempt to replace a letter with an SMS text message reminder did not achieve the desired result, we eventually found that sending more than 1 reminder combining both SMS text messaging and postal delivery methods may increase participation rates. Integrating SMS text messages into cancer screening should ensure that the screening process is streamlined and minimize access barriers. When implementing an intervention, all stakeholders should be informed and receive clear and concise instructions. In doing so, we will facilitate that more people participate in cancer screening programs and receive the benefits of early detection and timely treatment.

Acknowledgments

We gratefully thank Virtudes Guardiola, Maria Jose Ruiz, Natividad Valera, and Àngels Ventura of the Colorectal Cancer Screening Program of the Catalan Institute of Oncology for their administrative support. We thank the CERCA Program/Generalitat de Catalunya for institutional support. We sincerely thank all members of the M-TICS Group for their dedication and expertise in conducting this research: Guillermo Bagaria, Gemma Binefa, Josep A Espinàs, Albert Farré, Montse Garcia (Principal

XSL•FO
RenderX

Investigator), Gemma Ibáñez-Sanz, Núria Milà, M José Pérez-Lacasta, Noemie Travier, Carmen Vidal, and Núria Vives (PhD Candidate). This study was partially cofounded by Instituto de Salud Carlos III (ISCIII)-Subdirección General de Evaluación and by FEDER funds/European Regional Development Fund (ERDF)—a way to build Europe (PI19/00226). The funders have not participated in the study design, collection, management, analysis, interpretation of data, writing of the report, or decision to submit the report for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Content of the text message and reminder letter sent to nonparticipants.
[DOCX File , 1545 KB - mhealth_v13i1e64243_app1.docx]

Multimedia Appendix 2
Baseline characteristics of enrolled individuals by trial arm (intention-to-treat population).
[DOCX File , 19 KB - mhealth_v13i1e64243_app2.docx]

Multimedia Appendix 3
Baseline characteristics of enrolled individuals by previous screening behavior and trial arm (intention-to-treat population).
[DOCX File , 20 KB - mhealth_v13i1e64243_app3.docx]

Multimedia Appendix 4
Subgroup analysis of the impact of the recovery strategy on participation (intention-to-treat).
[DOCX File , 21 KB - mhealth_v13i1e64243_app4.docx]

Multimedia Appendix 5
CONSORT (Consolidated Standards of Reporting Trials) 2020 checklist.
[PDF File (Adobe PDF File), 82 KB - mhealth_v13i1e64243_app5.pdf]

References

1. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut 2023;72(2):338-344. [doi: 10.1136/gutjnl-2022-327736] [Medline: 36604116]
2. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: an overview. Int J Cancer 2021. [doi: 10.1002/ijc.33588] [Medline: 33818764]
3. Lin JS, Perdue LA, Henrikson NB, Bean SI, Blasi PR. Screening for colorectal cancer: an evidence update for the U.S. preventive services task force. Rockville (MD): Agency for Healthcare Research and Quality (US) 2021. [Medline: 34097369]
4. IARC. Colorectal Cancer Screening. IARC Handbooks of Cancer Prevention. Lyon, France: International Agency for Research on Cancer; 2019.
5. Klabunde C, Blom J, Bulliard J, Garcia M, Hagoel L, Mai V, et al. Participation rates for organized colorectal cancer screening programmes: an international comparison. J Med Screen 2015;22(3):119-126. [doi: 10.1177/0969141315584694] [Medline: 25967088]
6. Gruner LF, Amitay EL, Heisser T, Guo F, Niedermaier T, Gies A, et al. The effects of different invitation schemes on the use of fecal occult blood tests for colorectal cancer screening: systematic review of randomized controlled trials. Cancers (Basel) 2021;13(7):1520 [FREE Full text] [doi: 10.3390/cancers13071520] [Medline: 33806234]
7. Car J, Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2012(7):CD007458 [FREE Full text] [doi: 10.1002/14651858.CD007458.pub2] [Medline: 22786507]
8. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
9. Schwebel FJ, Larimer ME. Using text message reminders in health care services: a narrative literature review. Internet Interv 2018;13:82-104 [FREE Full text] [doi: 10.1016/j.invent.2018.06.002] [Medline: 30206523]
10. Hagoel L, Neter E, Stein N, Rennert G. Harnessing the question-behavior effect to enhance colorectal cancer screening in an mHealth experiment. Am J Public Health 2016;106(11):1998-2004. [doi: 10.2105/AJPH.2016.303364] [Medline: 27631750]

XSL•FO
RenderX

11. Hirst Y, Skrobanski H, Kerrison RS, Kobayashi LC, Counsell N, Djedovic N, et al. Text-message reminders in colorectal cancer screening (TRICCS): a randomised controlled trial. Br J Cancer 2017;116(11):1408-1414 [FREE Full text] [doi: 10.1038/bjc.2017.117] [Medline: 28441381]
12. Uy C, Lopez J, Trinh-Shevrin C, Kwon SC, Sherman SE, Liang PS. Text messaging interventions on cancer screening rates: a systematic review. J Med Internet Res 2017;19(8):e296 [FREE Full text] [doi: 10.2196/jmir.7893] [Medline: 28838885]
13. Vives N, Farre A, Ibáñez-Sanz G, Vidal C, Binefa G, Milà N, et al. Text messaging as a tool to improve cancer screening programs (M-TICS Study): a randomized controlled trial protocol. PLoS One 2021;16(1):e0245806 [FREE Full text] [doi: 10.1371/journal.pone.0245806] [Medline: 33481914]
14. Colls C, Mias M, García-Altés A. A deprivation index to reform the financing model of primary care in Catalonia (Spain). Gac Sanit 2020;34(1):44-50 [FREE Full text] [doi: 10.1016/j.gaceta.2018.07.015] [Medline: 30595339]
15. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. Lancet 2005;365(9471):1657-1661. [doi: 10.1016/S0140-6736(05)66516-6] [Medline: 15885299]
16. Vidal C, Garcia M, Benito L, Milà N, Binefa G, Moreno V. Use of text-message reminders to improve participation in a population-based breast cancer screening program. J Med Syst 2014;38(9):118. [doi: 10.1007/s10916-014-0118-x] [Medline: 25073694]
17. Vives N, Travier N, Farre A, Binefa G, Vidal C, Pérez Lacasta MJ, et al. Effectiveness and acceptability of targeted text message reminders in colorectal cancer screening: randomized controlled trial (M-TICS Study). JMIR Public Health Surveill 2024;10:e57959 [FREE Full text] [doi: 10.2196/57959] [Medline: 39083331]
18. Honein-AbouHaidar GN, Kastner M, Vuong V, Perrier L, Daly C, Rabeneck L, et al. Systematic review and meta-study synthesis of qualitative studies evaluating facilitators and barriers to participation in colorectal cancer screening. Cancer Epidemiol Biomarkers Prev 2016;25(6):907-917. [doi: 10.1158/1055-9965.EPI-15-0990] [Medline: 27197277]
19. Larsen MB, Hedelund M, Flander L, Andersen B. The impact of pre-notifications and reminders on participation in colorectal cancer screening - a randomised controlled trial. Prev Med 2022;164:107229. [doi: 10.1016/j.ypmed.2022.107229] [Medline: 36057390]
20. Sallis A, Sherlock J, Bonus A, Saei A, Gold N, Vlaev I, et al. Pre-notification and reminder SMS text messages with behaviourally informed invitation letters to improve uptake of NHS Health Checks: a factorial randomised controlled trial. BMC Public Health 2019;19(1):1162 [FREE Full text] [doi: 10.1186/s12889-019-7476-8] [Medline: 31438908]
21. Acharya A, Judah G, Ashrafian H, Sounderajah V, Johnstone-Waddell N, Harris M, et al. Investigating the national implementation of SMS and mobile messaging in population screening (The SIPS study). EBioMedicine 2023;93:104685 [FREE Full text] [doi: 10.1016/j.ebiom.2023.104685] [Medline: 37384997]

Abbreviations

CRC: colorectal cancer
DS: deprivation score
FIT: fecal immunochemical test
FOBT: fecal occult blood test
M-TICS: Mobile Phone Messaging as a Tool to Improve Cancer Screening
OR: odds ratio
RCT: randomized controlled trial

Edited by L Buis; submitted 12.07.24; peer-reviewed by C Wilson, L Aljerf; comments to author 14.08.24; revised version received 08.10.24; accepted 05.11.24; published 01.01.25.

Please cite as:
Vives N, Binefa G, Travier N, Farre A, Panera JA, Casas B, Vidal C, Ibáñez-Sanz G, Garcia M, M-TICS research group
Text Messaging Versus Postal Reminders to Improve Participation in a Colorectal Cancer Screening Program: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e64243
URL: https://mhealth.jmir.org/2025/1/e64243
doi:10.2196/64243
PMID:

©Nuria Vives, Gemma Binefa, Noemie Travier, Albert Farre, Jon Aritz Panera, Berta Casas, Carmen Vidal, Gemma Ibáñez-Sanz, Montse Garcia, M-TICS research group. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 01.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of a Pilot mHealth Intervention to Engage Primary Care Clients at an Urban Clinic Serving Marginalized Populations: Mixed-Methods Cohort Study

Lauren Harrison[1], MSc; Christina Fulton[1]; Antonio Marante Changir[1], MSc; Adedayo Ajidahun[1]; Cassidy Tam[1]; Wendy Zhang[1]; Tian Shen[1]; Rolando Barrios[1]; Julio Montaner[1], MD; Kate Salters[1], PhD; Richard T Lester[1], MD; Surita Parashar[1]; David M Moore[1,2], MDCM

[1]British Columbia Centre for Excellence in HIV/AIDS, #608-1081 Burrard Street, Vancouver, BC, Canada

[2]Facutly of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
David M Moore, MDCM
British Columbia Centre for Excellence in HIV/AIDS, #608-1081 Burrard Street, Vancouver, BC, Canada

Abstract

Background: Many individuals in urban low-income settings face barriers to engaging in primary care and experience systemic challenges such as homelessness and discrimination in the health care system. This study was conducted in the Downtown Eastside of Vancouver, Canada, a low-income neighborhood with intersecting structural vulnerabilities and disproportionate rates of substance use disorders. Advancements in mobile health expand options for facilitating communication between primary care providers and clients.

Objective: We conducted a pilot project that provided primary care clients a mobile phone and access to WelTel, a mobile health tool that uses a 2-way texting approach and sends weekly automated check-in messages. Our study measured phone retention, defined as retaining a study-supplied device and being reachable on the study-supplied device over a 6-month period and explored the acceptability and feasibility of WelTel among a cohort of clients with complex health challenges.

Methods: Stratified random sampling was used to recruit participants from a larger cohort study of primary care clinic clients in the Downtown Eastside. The sample was stratified to ensure equal participation based on gender and Indigenous and non-Indigenous participants. In this mixed methods research, participants completed 3 surveys over a 6-month period from November 2022 to May 2023. The surveys assessed phone retention and functionality as well as phone use, including use of the WelTel platform. Clients who had access to a functional mobile phone after the follow-up period were invited to complete an in-person interview. The semistructured interviews explored clients' experiences with WelTel, primary care, and engagement with technology.

Results: We enrolled 49 participants (median age 48 y; 53% women and 49% Indigenous) and interviewed 16 participants. A total of 44 clients completed the 6-month survey, and of those clients, 26 (59%) had a functional phone. However, only 14 (29%) clients retained the mobile device supplied during the study and completed the 6-month survey. Phone retention or access to a nonstudy phone was lower among those who had used opioids (18% for both) in the past 3 months (P<.01), as well as those who reported not having access to a cellphone at enrollment (25% and 15%, respectively) compared to those who did have phone access at enrollment (P=.05). Both the surveys and semistructured interviews indicated that WelTel was generally well received; 25 (96%) of those with a functional phone reported that they liked receiving the weekly messages. During qualitative interviews, the WelTel intervention was reported to strengthen client-provider relationships and create pathways for receiving care.

Conclusions: Overall, the pilot study found this intervention feasible and acceptable to clients; however, barriers to phone retention were an ongoing challenge. Expanded enrollment in the WelTel service will allow us to examine whether it also facilitates engagement in primary care among marginalized urban populations.

(JMIR Mhealth Uhealth 2025;13:e68561) doi:10.2196/68561

KEYWORDS

text messaging; primary health care; marginalized populations; substance use; mental health; mHealth; mobile phone

XSL•FO
RenderX

Introduction

Client engagement in primary health care can improve health outcomes and lower health care costs [1]. All Canadian residents receive health care through Canada's publicly funded health care system, which covers physician and hospital services [2]. However, disparities in access to regular primary care have been growing for marginalized groups in Canada [3]. Primary care visits and continuity of care have decreased from 1999 to 2018, whereas emergency room visits have increased, and this dichotomy has been the strongest in lower-income neighborhoods [4]. People who use drugs (PWUDs) have been found to have lower primary health care engagement, despite complex chronic health needs and high acute emergent care access [5,6]. Better engagement in primary health care with a regular physician among PWUD has been shown to reduce the use of emergency services [7]. Strong provider-client relationships established by building rapport and trust can facilitate patient engagement and retention for those facing systemic barriers to health care engagement [8,9].

Virtual health care, including telehealth and mobile health (mHealth), expanded during the SARS-CoV-2 pandemic, during which Canadian provinces and territories placed lockdowns on residents to mitigate infectious disease spread in 2020 [10]. The lockdowns necessitated virtual care options as movement was restricted, and therefore funding was increased for these modalities; while virtual care peaked during these lockdowns, it appears to have decreased in 2021 [10]. Before the SARS-CoV-2 pandemic, mHealth interventions were already being recognized as innovative tools to support disease self-management, health behavior change, and communication with health care providers [11]. One-way messaging symptoms, where participants receive but do not respond to messages, can nudge users toward behaviors such as prescription refills, and the use of chatbots is being explored as artificial intelligence in health care expands [12].

WelTel is an mHealth tool that provides a secure virtual health platform and allows for 2-way texting between health care providers and clients [13]. The platform was found to place a low burden on health care practitioners with a modest relative cost [14]. WelTel has been used for HIV care in Canada and East Africa and has been found to improve medication adherence, connectedness to care, and health outcomes for people living with HIV [15-17]. In addition to HIV care, previous pilot projects evaluating WelTel in British Columbia and Alberta, Canada, have shown promising results; WelTel was well received by health care providers and patients at a tuberculosis clinic [18], improved asthma control outcomes for patients with physician-diagnosed asthma [19], and was found to support high-quality care among rheumatoid arthritis patients [20]. Among Indigenous youth in British Columbia who use substances, there was interest in using mHealth; however, mobile phone ownership was recognized as a potential barrier to the use of WelTel for clients facing multiple structural vulnerabilities [21].

The Downtown Eastside (DTES) is a neighborhood in Vancouver, Canada, that is disproportionately impacted by poverty, unstable and inadequate housing, and food insecurity [22]. As of 2016, the life expectancy in the DTES was 62.2 years for females and 60.2 years for males compared to 86.6 and 82.5 years, respectively, within the general population of Metro Vancouver [23]. Additionally, the median income in 2016 was CAD $23,352 (US $17,963) in the DTES, whereas the median income in the City of Vancouver was $65,423 [24]. High rates of conditions such as HIV, hepatitis C virus, and mental health and substance use disorders are found among many intersecting social inequities [22]. While a strong sense of community can be found among residents in the DTES, stigma and discrimination are common and can act as a barrier to accessing primary health care [25,26]. Indigenous peoples face health care discrimination and racism and are more likely to reside in the DTES, with approximately 10% of DTES residents identifying as Indigenous compared to 2% in the City of Vancouver [27]. In the DTES, PWUDs have reported multiple barriers to accessing primary health care [5]. As such, this population may benefit greatly from enhanced interventions to improve access to primary health care.

Previous qualitative research from our team indicated that mHealth would be desirable in this community and may help to mitigate barriers to accessing primary care [26]. We conducted a pilot project that supplied clients of one primary care clinic in the DTES with a mobile phone and data, talk, and text mobile phone plan as well as access to the WelTel service over a 6-month period. We used quantitative methods to measure phone retention, defined as retaining the study-supplied device and being reachable on the study-supplied device, phone usage, and client experiences of the WelTel platform. Qualitative methods were used to garner a deeper understanding of user experiences, tech engagement, and health care experiences. This mixed methods approach allowed us to generate insight into the feasibility and acceptability of the program.

Methods

Setting

In October 2019, the Hope to Health Clinic was opened to provide interdisciplinary primary health care to residents of the DTES with unmet health needs who were previously unattached to primary care [28]. The Hope to Health Clinic was designed to engage and retain clients facing multiple barriers to accessing primary health care and to improve health outcomes for this population [28]. Clients at the Hope to Health Clinic receive low-barrier psychosocial and primary care services to address health, mental health, and addiction challenges [29]. The innovative work at the Hope to Health Clinic is connected to onsite research teams that allow for collaboration between clinicians, researchers, and community members.

Study Sample

Beginning June 2021, all clients of the Hope to Health Clinic were offered enrollment in the Hope to Health Engagement and Retention Evaluation (HERE) study, a longitudinal cohort study consisting of a baseline survey and follow-up surveys every 6 months for 3 years. A convenience sample of clients who attended the clinic and were interested in the study was enrolled; posters at the Hope to Health Clinic informed clients of their

XSL•FO
RenderX

eligibility for the study, medical practitioners facilitated connections between clinic clients and the research team, and research team members were available for face-to-face conversations. At enrollment, participants in the HERE study were given the option to consent to being contacted by the research team for the purpose of participating in a substudy on mobile phones and the WelTel intervention.

Of those who provided this consent, a stratified random sample of participants was selected using random allocation in Microsoft Excel. Participants were stratified based on self-identified Indigenous or non-Indigenous status and self-identified gender to ensure equal opportunities for participation across these strata. All transgender participants in the HERE study were offered enrollment in the WelTel pilot. Additionally, all participants who identified as nonbinary or other gender identities were offered enrollment, but none enrolled. Study staff made outreach attempts to selected participants to offer enrollment, and the random selection procedure was repeated until enrollment was completed with a target of 50 participants. All participants were supplied with a mobile phone device and registered in the WelTel platform. The mobile phones had touchscreens and additional software functions, including internet and the option to download numerous games and applications. All phones were previously used with Android operating systems and had been rebuilt and certified by the provider. A member of the research team provided instructions on how to use the mobile phone and WelTel; ensured the participant had password protected their device; and answered any questions that arose about the phone, the data plan, or WelTel. During this session, participants also consented to complete interviewer-administered surveys every 2 months for a 6-month follow-up period (3 surveys total). Clients who no longer had the study-supplied device but provided their own mobile phone during the follow-up period were able to register their new device with the phone plan and WelTel. Participants were enrolled from November 2022 to May 2023 with follow-up continuing until November 2023.

Intervention

Participants received a weekly automated check-in message from WelTel that asked “how are you?” On the basis of participants’ responses, WelTel categorized the messages as “okay,” “not okay,” or “unrecognized.” Messages labeled “unrecognized” were not automatically designated by the WelTel system as okay or not okay. Participants were instructed to minimize personal information shared with WelTel when responding to the initial check-in messages. Medical office assistants monitored incoming messages and responded according to the need expressed in the participant’s message, including messages that were unrecognized by the system. Participants who did not reply to the original check-in message did not receive further follow-up from clinic staff for that week.

This study used a mixed methods design to evaluate provisioning of the phone and the acceptability of the WelTel messaging system. Quantitative surveys were used to characterize phone retention and functionality, general phone use, WelTel use, and clinic communication. Qualitative interviews elucidated a deeper understanding of WelTel accessibility and user experiences as well as tech engagement and primary care experiences.

Data Collection and Analysis

Surveys

Following enrollment in the pilot project, participants completed bimonthly surveys for 6 months with a Peer Research Associate (PRA). PRAs are trained researchers with lived experience of the marginalization experienced by residents of the DTES. Surveys were completed over the phone or in person at the Hope to Health Clinic. The surveys took approximately 10 minutes.

HERE study baseline surveys were used to describe participant characteristics and phone access and to assess interest in using technology to engage with health care. Phone retention was summarized using the pilot project follow-up survey data, and χ^2 and Fisher exact tests were used to assess HERE study baseline variables associated with phone retention. Descriptive statistics from the follow-up surveys and the WelTel platform show WelTel usage and client satisfaction with WelTel.

Semistructured Interviews

At the end of the 6-month follow-up period, all participants who had a functional mobile phone were offered enrollment in the qualitative substudy up to a maximum of 20 participants. Interviews were completed at the Hope to Health Clinic and were facilitated in pairs by a PRA and the Research Coordinator. Participants provided consent and completed the semistructured interview in 1 session, which lasted approximately 30 to 45 minutes. All interviews were audio recorded and transcribed verbatim. After completing each interview, the Research Coordinator and PRA debriefed about the interview and discussed early observations of emergent codes.

We used Braun and Clarke’s [30] thematic analysis to provide a flexible framework for identifying themes and took a codebook approach [31]. Two research coordinators trained in qualitative analysis began by open-coding 3 transcripts and identifying emergent codes. PRAs provided insight and feedback as these emergent codes were organized into the codebook. Using the codebook, all interviews were coded independently in NVivo (Lumivero, Denver, CO) by both research coordinators, and the research coordinator who co-facilitated all interviews acted as the primary coder. The primary and secondary coders frequently collaborated to refine the codebook and found a high degree of consensus between their codes.

Ethical Considerations

Ethics approval was obtained from the University of British Columbia/Providence Health Care Research Ethics Board (#H20-03256). All clients provided informed consent for the quantitative surveys at enrollment in the study and had the option of indicating their interest in the qualitative interviews. Clients who agreed to take part in the qualitative interviews provided informed consent through a separate process before the interview took place. Clients were made aware that they could withdraw from the study at any time. The data were deidentified and securely stored. Participants were provided a CAD $15 ($US 10.80) honorarium per survey and a CAD $30 ($US 21.60) honorarium for the qualitative interview. While

XSL•FO
RenderX

financial incentives may influence participants' willingness to participate in research, we felt that it was appropriate to reimburse participants for their time, as is standard research practice. Additionally, the participants were provided mobile phones to participate in this study, which were considered their personal property and not expected to be returned. The talk/text/data plans were provided to participants for up to 2 years, provided they retained their phone provided by the study.

Results

Quantitative Surveys

Baseline Characteristics

Of the 49 participants who enrolled in the pilot study, 26 (53%) identified as women and 23 (47%) identified as men inclusive of cisgender and transgender participants. The median age was 48 years (IQR 37-56). Nearly half (n=24, 49%) of participants identified as Indigenous, 20 (41%) identified as White, and 5 (10%) identified as other ethnicities. At the time of the HERE study baseline survey, 44 (90%) participants reported an annual income below $20,000, and 12 (25%) participants were currently homeless; an additional 9 (18%) participants had been homeless in the previous 3 months. A total of 19 (39%) participants were taking medications to treat opioid use disorder (ie, methadone), 24 (49%) participants used opioids, 21 (43%) participants used stimulants, and 16 (33%) participants reported injecting substances in the previous 3 months.

The HERE study baseline data indicated 27 (55%) of the participants enrolled in the phone study had regular access to a cell phone, and 11 (22%) had used a cell phone to text their health care provider before enrollment in the pilot. Thirty-eight (78%) participants indicated they would like to text with their health care provider; of these participants, 33 (87%) were interested in receiving appointment reminders, 31 (82%) in receiving health information, and 31 (82%) in texting to set up appointments. Additionally, 38 (78%) participants indicated they would like to be able to use a mobile phone to remind the clinic about their prescription refills, and 35 (71%) participants reported they would like to use a mobile phone to have a clinic visit using video.

Follow-Up Surveys and WelTel Data

Of the 49 participants enrolled in the pilot study, 46 (94%) completed at least one of 3 follow-up surveys, and 44 (90%) completed the third follow-up survey at the end of the 6-month period. Of these 44 participants, 14 (32%) had retained their study-supplied mobile phone, and an additional 12 (27%) participants had access to another functional phone (Table 1). Of the 30 participants who did not retain their phone, 28 (64%) reported that their study supplied mobile phone had been stolen. No participants reported that their phone had been sold or damaged. While we found similar proportions of participants with no access to a phone at the end of the study based on age (44% of participants aged ≥50 y vs 39% for those aged <50 y), a higher proportion of those aged ≥50 years had retained their phone (44% vs 25%), but a lower proportion had access to a nonstudy phone at the end of the pilot (13% vs 36%; P=.04 for all comparisons). Phone retention or access to a nonstudy phone was lower among those who had used opioids (18% for both) in the past 3 months compared to those who did not (45% and 36% respectively; P<.01). As well, those who reported not having access to a cellphone at the HERE study baseline survey also had lower proportions of phone retention or access to a nonstudy phone (25% and 15%, respectively) compared to those who did have phone access at enrollment (38% for both; P=.05). Stimulant or alcohol use in the past 3 months, homelessness, ethnicity, and gender were not associated with phone retention and access to a functional phone (Table 1).

XSL•FO
RenderX

Table . Factors associated with phone retention and access after 6 months of follow-up.

Variable	Total	Retained the study supplied device (n=14), n (%)	Access to a functional phone (not the study supplied device; n=12), n (%)	No access to a functional phone (n=18), n (%)	*P* value
Age (y)					
19 - 49	28	7 (25)	10 (36)	11 (39)	
50 - 65+	16	7 (44)	2 (13)	7 (44)	.04
Gender					
Male	20	7 (35)	4 (20)	9 (45)	
Female	24	7 (29)	8 (33)	9 (38)	.64
Ethnicity					
Indigenous	21	9 (43)	3 (14)	9 (43)	
Caucasian/White	18	4 (22)	6 (33)	8 (44)	
Other ethnicity	5	1(20)	3 (60)	1 (20)	.22
Homelessness					
Currently homeless	11	4 (36)	0 (0)	7 (64)	
Homeless, past 3 mo (not current)	9	2 (22)	3 (33)	4 (44)	
Homeless, more than 3 mo ago	17	5 (29)	8 (47)	4 (24)	
Never been homeless	7	3 (43)	1 (14)	3 (43)	.13
Regular access to a cell phone at baseline					
No	20	5 (25)	3 (15)	12 (60)	
Yes	24	9 (38)	9 (38)	6 (25)	.05
Opioid use in the past 3 mo					
No	22	10 (45)	8 (36)	4 (18)	
Yes	22	4 (18)	4 (18)	14 (64)	.01
Stimulant use in the past 3 mo					
No	24	10 (42)	7 (29)	7 (29)	
Yes	20	4 (20)	5 (25)	11 (55)	.18
Alcohol use in the past 3 mo					
No	28	11 (25)	6 (21)	11 (39)	
Yes	16	3 (19)	6 (38)	7 (18)	0.31

At the third bimonthly survey, 25 (96%) of the 26 participants with a functional phone reported that they liked receiving the weekly WelTel text messages. Across the 3 follow-up surveys, participants reported 16 unique incidents of responding to a WelTel message that they needed help; 7 (44%) of these incidents were pertaining to a medical problem, and 7 (44%) participants needed help with appointment reminders. Of these incidents, participants reported 12 (75%) of their problems were fully addressed. WelTel data showed that 643 (37%) of responses from clients indicated that they were okay, 75 (4%) indicated that the client was not okay or the response type was unrecognized, and 1027 (59%) of check-in messages received no response (Table 2). The median response time for clients to reply to the WelTel message was 221 minutes (Q1-Q3: 23 - 1415 min).

XSL•FO
RenderX

Table . Frequency and type of WelTel response. The median of the total number of texts sent by clients over the study period was 30 (Q1 to Q3: 0-45).

Type of response	Vaues, n (%)
Not okay	17 (1)
Okay	643 (37)
Unrecognized (not designated as not okay or okay)	58 (3)
No response	1027 (59)

Semistructured Interviews

A total of 16 participants who had access to a functional phone at the third bimonthly survey completed the qualitative interviews. Of the 16 participants who completed in-depth interviews, 7 (44%) were men, and 9 (56%) were women; 6 (38%) identified their ethnicity as Caucasian/White; 8 (50%) identified as Indigenous; and 2 (13%) identified as other ethnicities. The median age was 48.5 years (Q1-Q3: 47-56 y). At the time of participant enrollment in the WelTel pilot study, 4 (29%) were currently homeless, 4 (29%) were not homeless but had been homeless in the previous 3 months, 4 (29%) had previously been homeless more than 3 months before, and 2 (14%) had never been homeless. Another 2 had missing information on homeless status.

Many participants who completed a qualitative interview viewed the WelTel platform as a tool to support health. Two key themes emerged: client-provider connection and program improvement. Communication and care, as well as self-assessment, were identified as important aspects of the client-provider connection.

Client-Provider Connection

Communication and Care

The importance of the relationship between clients and their providers was emphasized by participants throughout this study, and feeling cared for was frequently mentioned as a benefit of the WelTel program. While one participant described feeling that:

> *someone who cares is on the other side.* [Man, Indigenous]

Another participant voiced that:

> *It makes me feel like they're concerned and that they care even a little bit even though I think it's a robot.* [Woman, Indigenous]

Having weekly texts from the clinic was well received and fostered positive relationships with the clinic staff; the automation of the initial check-in text did not interfere with this relationship building between clients and providers.

Many participants expressed that WelTel facilitated "open and honest" communication with their primary care providers and removed feelings of judgment. The improved communication appeared to both contribute to and result from strengthened client-provider relationships. Increased comfort disclosing medical issues through WelTel provided participants with a pathway to receive care that they may not have otherwise accessed. This was explained by a participant when she compared her health care before using WelTel to her experience with the platform:

> *I wouldn't call. I would just deal with whatever was going on, whether I had a small stroke or whatever. I would just deal with it. Now with the WelTel I feel like I can communicate with them without feeling embarrassed.* [Woman, Indigenous]

By attenuating negative feelings around accessing medical care, participants were empowered to engage in their healthcare. This process was further described by the participant:

> *I have an abscess and it's got me scared and they sent the [WelTel] text asking me how I was and I told them I was in a lot of pain, but otherwise I was doing good and they texted back. And that led to talking to a nurse and now I'm going to get medical assessment done because of that text.* [Woman, Indigenous]

The participant directly attributed the care she received to WelTel, highlighting the potential for WelTel to improve patient care and health outcomes:

> *it's helping me with my mental health, it's helping me with other problems that I have and it's all because that one little text, how are you doing.* [Woman, Indigenous]

It should be noted that this was not the experience of all participants, as 1 participant explained:

> *they try to help fix my problem or whatever by putting me in touch with somebody else, just that person can only do so much. No one can believe a magic wand.* [Man, Indigenous]

Ultimately, while WelTel may improve clients' connection to primary care, limitations in the health care system are still present. Furthermore, although generally perceived as beneficial, a minority of participants felt that WelTel did not impact their communication with the clinic, and some participants expressed a preference for in-person communication with clinic staff:

> *I'm not a big fan of sitting or text messaging or talking to somebody on the phone. I'd rather face-to-face.* [Man, Indigenous]

In recognizing the benefits of communicating through text message, it should be noted that texting was not all participants' preferred method of communication.

XSL•FO
RenderX

Self-Assessment

For some participants, WelTel acted as a prompt to check in with themselves and assess their current health status. This was described by a participant who said:

> *I can get stuck in my head, and not really realise where I'm at. Sometimes when I've gotten that message, I've been like, how are you actually doing, are you doing okay or are you struggling with stuff right now.* [Man, White]

One participant explained that this self-assessment motivated personal action to support their overall well-being when they shared:

> *the phone call says hey, how are you. Hey, I need a shower.* [Woman, Indigenous]

A similar viewpoint was shared by another participant who felt their increased awareness of their health and wellness was inextricably linked with their primary care:

> *I think this reminds me about my health and wellness because of where it came from. So, it has automatically got a subliminal message, I guess, it just makes me more aware. That's about all because I don't really think about Hope to Health with the regular telephone, so it's easier to slip my mind or forget about.* [Man, Indigenous]

While the "how are you" message does not explicitly mention the Hope to Health Clinic, receiving the automated message created awareness of the participant's well-being and acted as a reminder to engage with the clinic.

Program Improvement

The WelTel program had a high degree of acceptability among participants in the study. The automated check-in messages were sent weekly on Mondays; all participants reported satisfaction with the frequency of text messages, and multiple participants stated that Mondays were an ideal day to receive the text messages. As reported by the participants, at times, the conversations that took place through WelTel contained sensitive medical information. However, all participants in the study had no privacy concerns about the WelTel platform and discussing medical issues on their mobile phones. While some participants indicated that they were receiving appointment reminders through WelTel, others indicated that they would like to see the program expanded to include appointment reminders.

There were several recommendations made regarding the messages that were exchanged between participants and the clinic staff after the initial automated check-in message. The medical office assistants at the Hope to Health Clinic monitored the incoming text messages and often engaged in text conversation with the clients. Clients would not know which clinic staff were replying to their messages, and 1 participant suggested it would be helpful if the clinic staff included their name. When clients indicated they were okay in response to the automated check-in, some reported that they did not receive any further messages, while others stated they would receive a reply.

Some participants felt follow-up messages were unnecessary when they had reported that they were okay, with 1 client stating:

> *it really doesn't make a difference. The only time it would maybe make a difference would be if I'm not okay.* [Man, Indigenous]

Although overall well received, there did not appear to be a high demand for clinic staff to reply to text messages from clients when they indicated that they were okay.

As shown in the quantitative surveys, 59% of Weltel check-in messages received no response, and reasons given for nonresponse in the qualitative interviews included busy schedules, being preoccupied, or sleeping when messages were received. When participants did not respond to the weekly check-in messages, they did not receive communication until the following Monday. For 1 participant, this exposed a gap in the service provided:

> *There have been a couple times where it has gone over a week without my responding, and no one has questioned it or anything. And I was using it as my safety net, I guess, that I'm 50 so anything could happen, I could slip in a tub, who knows whatever. So, after a few days I would have expected, hey just checking in again.* [Man, Indigenous]

While participants were instructed at enrollment that the WelTel program was not an emergency service, providing follow-up messages to nonresponse may provide extra support while remaining within the scope of the program. Furthermore, this participant's experience highlights how creating clear expectations at enrollment can aid clients in understanding how WelTel may be able to support their health.

Similarly, at enrollment, participants were informed that messages would receive a response within 2 business days, excluding weekends and holidays. When asked how long it takes for the clinic to respond to their messages, 1 participant reported that:

> *the longest it has taken would be the next day, and the fastest would have probably been within probably a few hours. I may have gotten one or two that were half an hour or about.* [Man, White]

This was perceived as a lag in response time by the participant when he described his experience with the WelTel check-in message:

> *There's not a lot attached to it, like, how are you. You're either fine, not okay, or you're doing pretty good. I found there's quite a lag in the response time, so if I'm really not doing okay, it would be nice to hear back from someone sooner. And there's not always a lot attached to it, if you ... I think a couple of times I've been like, no, I'm not having a very good day. Someone has asked, is there anything we can do to help, but other times it hasn't happened, it's just like, I'm sorry to hear that, hope things get better.* [Man, White]

XSL•FO
RenderX

In addition to expressing dissatisfaction with the response time, this participant did not find the type of responses they received was consistently adequate. Recognizing limitations introduced by the workload of clinic staff, reducing the time to response, and providing additional training to clinic staff could address the concerns presented by this participant.

Discussion

Principal Findings

Our study provides evidence that the WelTel intervention is acceptable and has the potential to facilitate engagement in primary care for a cohort of clients facing systemic barriers in an inner-city urban low-income neighborhood where substance use, mental health disorders, and homelessness are highly prevalent. Participants were largely positive about the program, and the majority of participants had favorable views about receiving the weekly text messages. As found among a cohort of people living with HIV in Vancouver [32], participants expressed that the WelTel intervention made them feel cared for, which stands in contrast to findings that residents of the DTES face discrimination and dehumanization by primary care providers [26]. In the context of the DTES where clients face stigma and judgment, the finding that the WelTel program provided a nonjudgmental space to share health care concerns with open communication meets a basic need that had been lacking [26]. Participants were not only able to express their concerns more freely with their providers but were also made more aware of their health needs, which in turn acted as a pathway to receive care via strengthened relationship with the clinic. Additionally, the self-assessment prompted by this increased awareness may aid in detecting health issues earlier and before the progression of the health issue [15].

At times, participants' expectations did not match the scope and/or capacity of the program, providing guidance on both how the program can be strengthened and ways in which instruction and support at the time of registration for WelTel can be improved. As opposed to 1-way mHealth texting programs, or programs that use chatbots, a benefit of WelTel is the 2-way communication approach that can facilitate discussion between clients and providers [33]; this potential for conversation was used and understood differently by clients in this study. In a similar study using WelTel, clients receive a follow-up call from a case manager if they did not respond to their check-in message by the end of the week [21], which may address nonresponse as well as align more closely with the expectation that WelTel could be used as a safety net. Text-based appointment reminders have been shown to reduce missed medical appointments [34,35], and many participants showed interest in receiving text-based appointment reminders, but only some expressed that they were currently receiving appointment reminders through WelTel. Participants reported varying experiences regarding the type of follow-up messages they received when they replied to the automated WelTel check-in, including whether they received any messages when they reported they were okay. Improving consistency regarding the type of messages clients can expect may help manage expectations and maintain quality. Additional training and workflow management could support health care providers in integrating WelTel into their daily tasks, as well as ensuring all key messages about the functionality of WelTel are communicated to clients at the time of registration.

Phone access and retention were a challenge throughout this study, mirroring similar work in the DTES that identified lost/damaged/stolen phones as the primary barrier to using WelTel [32]. A significant proportion (45%) of our participants did not have access to a phone at the HERE study baseline survey, and those who had used opioids in the past 3 months were less likely to have retained the study-supplied device or have access to a phone after 6 months. Phone retention was dynamic, and the number of functional phones participants had fluctuated throughout the study period, with some participants replacing the study-supplied phone after it was lost or stolen. The qualitative interviews primarily explored participants' experiences with WelTel and were limited to participants who had a functional phone after the 6-month follow-up period. While WelTel was shown in this study to provide benefit for those able to use the platform, a major gap exists for those who may be facing additional vulnerabilities that create challenges for phone retention. The scalability of this pilot project may be limited by clients' access to a mobile device, as providing a phone did not sufficiently address barriers to phone access. Notably, PRAs made considerable effort to connect with study participants for follow-up interviews but had limited success (n=33 at the 2-month follow-up), highlighting limitations in the potential for outreach using mobile phones.

This intervention aims to increase engagement for people who experience barriers to remaining attached to primary care; however, these same barriers may have impeded our teams' ability to connect with participants selected for the pilot project. Therefore, those with the highest need may not have been included in this cohort. Participants may have felt motivated to provide positive feedback about the program because they had received a phone and talk/text/data plan. However, many participants felt comfortable sharing that they were not happy with the quality of the phones, which indicates that responder bias was mitigated in this study, likely through the inclusion of skilled PRAs who shared lived/living experience with participants. This relatively small cohort provided valuable, but limited insight into the impact of an mHealth program on engagement and retention in primary care for clients with complex, chronic health needs in a Canadian context. Further research will offer all clients at the Hope to Health Clinic enrollment in WelTel to explore the scalability of providing WelTel services as part of ongoing primary care.

Conclusions

This study showed that the WelTel intervention was generally well received by Hope to Health Clinic clients. The intervention facilitated communication and connection between clients and health care providers and fostered positive relationships for some clients. The WelTel check-in messages have the potential to act as a prompt for self-reflection and provide pathways to receive care that otherwise would not have been used. Phone retention and access were an ongoing challenge in this study. These findings support the expansion of mHealth programs to

XSL•FO
RenderX

improve engagement and retention in primary care as an innovative service delivery tool in underserved populations; however, there remains a need to explore additional solutions for individuals who are not consistently reachable via a mobile phone. In the next phase of our research, we will offer WelTel enrollment to all Hope to Health Clinic clients who have access to a mobile phone, to see if, indeed, it can assist in better communication between clients and clinic staff and ultimately result in improved retention in primary care for this population.

Acknowledgments

The authors respectfully acknowledge that their work takes place on the traditional, ancestral, and unceded territories of the Coast Salish Peoples, including the Musqueam, Squamish, and Tsleil-Waututh Nations. They also would like to thank everyone who contributed their time and expertise to the Hope to Health Engagement and Retention Evaluation study, including the study participants, Peer Research Associates, co-investigators and co-collaborators, and the Community Engagement Committee at the Hope to Health Clinic who provided invaluable feedback.

Funding

The Hope to Health Engagement and Retention Evaluation study was funded by the Canadian Institutes for Health Research (grant PJT-173270).

Data Availability

The British Columbia Center for Excellence in HIV/AIDS (BC-CfE) is prohibited from making individual-level data available publicly due to provisions in our service contracts, institutional policy, and ethical requirements. The datasets generated and analyzed during this study are not publicly available to protect the privacy of study participants. As this pilot project had a small number of participants, the risk of reidentification is high.

Authors' Contributions

DM, RB, SP, and KS designed the HERE study, including the WelTel Pilot Study. CS, LH supervised the data collected. AM, TW, and CF collected the survey data. CT prepared the survey data, and WZ and TS conducted the analysis. LH and AA conducted the qualitative interviews and conducted the analysis. LH wrote the first draft of the manuscript and incorporated feedback from all other authors. All authors provided input and approved the final manuscript.

Conflicts of Interest

RL has a financial and professional interest in WelTel Incorporated, a company for which he is Chief Scientific Officer. WelTel Incorporated was subcontracted to provide the SMS software platform and maintenance for the project. However, RL did not provide input to the data analysis due to this potential conflict of interest. JSGM has received limited unrestricted funding, paid to his institution, from Gilead Sciences, Merck, and ViiV Healthcare. All other authors declare that they have no competing interests.

References

1. van Weel C, Kidd MR. Why strengthening primary health care is essential to achieving universal health coverage. CMAJ 2018 Apr 16;190(15):E463-E466. [doi: 10.1503/cmaj.170784] [Medline: 29661815]
2. Government of Canada. About Canada's health care system. 2025 Jun 25. URL: https://www.canada.ca/en/health-canada/services/canada-health-care-system.html [accessed 2025-07-16]
3. Lavergne MR, Bodner A, Allin S, et al. Disparities in access to primary care are growing wider in Canada. Healthc Manage Forum 2023 Sep;36(5):272-279. [doi: 10.1177/08404704231183599] [Medline: 37340726]
4. Lavergne MR, Bodner A, Peterson S, et al. Do changes in primary care service use over time differ by neighbourhood income? Population-based longitudinal study in British Columbia, Canada. Int J Equity Health 2022 Jun 7;21(1):80. [doi: 10.1186/s12939-022-01679-4] [Medline: 35672744]
5. Brar R, Milloy MJ, DeBeck K, et al. Inability to access primary care clinics among people who inject drugs in a Canadian health care setting. Can Fam Physician 2021 Dec;67(12):e348-e354. [doi: 10.46747/cfp.6712e348] [Medline: 34906953]
6. Kendall CE, Boucher LM, Donelle J, et al. Engagement in primary health care among marginalized people who use drugs in Ottawa, Canada. BMC Health Serv Res 2020 Sep 7;20(1):837. [doi: 10.1186/s12913-020-05670-z] [Medline: 32894114]
7. Kendall CE, Boucher LM, Mark AE, et al. A cohort study examining emergency department visits and hospital admissions among people who use drugs in Ottawa, Canada. Harm Reduct J 2017 May 12;14(1):16. [doi: 10.1186/s12954-017-0143-4] [Medline: 28494791]
8. Mallinson RK, Rajabiun S, Coleman S. The provider role in client engagement in HIV care. AIDS Patient Care STDS 2007;21 Suppl 1:S77-S84. [doi: 10.1089/apc.2007.9984] [Medline: 17563293]

XSL•FO
RenderX

9. Komaromy M, Madden EF, Zurawski A, et al. Contingent engagement: what we learn from patients with complex health problems and low socioeconomic status. Patient Educ Couns 2018 Mar;101(3):524-531. [doi: 10.1016/j.pec.2017.08.019] [Medline: 28890084]
10. Patterson PB, Roddick J, Pollack CA, Dutton DJ. Virtual care and the influence of a pandemic: necessary policy shifts to drive digital innovation in healthcare. Healthc Manage Forum 2022 Sep;35(5):272-278. [doi: 10.1177/08404704221110084] [Medline: 35775144]
11. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415. [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
12. Luong P, Glorioso TJ, Grunwald GK, et al. Text message medication adherence reminders automated and delivered at scale across two institutions: testing the nudge system: pilot study. Circ Cardiovasc Qual Outcomes 2021 May;14(5):e007015. [doi: 10.1161/CIRCOUTCOMES.120.007015] [Medline: 33993727]
13. WelTel Health, a secure saasvirtual care platform. WelTel Health. URL: https://www.weltelhealth.com/platform [accessed 2024-07-14]
14. Campbell AR, Kinvig K, Côté HC, et al. Health care provider utilization and cost of an mHealth intervention in vulnerable people living with HIV in Vancouver, Canada: prospective study. JMIR Mhealth Uhealth 2018 Jul 9;6(7):e152. [doi: 10.2196/mhealth.9493] [Medline: 29986845]
15. Chiang N, Guo M, Amico KR, Atkins L, Lester RT. Interactive two-way mHealth interventions for improving medication adherence: an evaluation using the behaviour change wheel framework. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e87. [doi: 10.2196/mhealth.9187] [Medline: 29650504]
16. Lester RT, Ritvo P, Mills EJ, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
17. Memetovic J, van der M. A pilot study of text-message support for HIV-positive patients at the oak tree clinic in Vancouver: WelTel HAART BC1. 2013. URL: https://smartsexresource.com/for-health-providers/sti-updates/2013/04/pilot-study-text-message-support-hiv-positive-patients-oak-tree-clinic/ [accessed 2024-07-14]
18. van der Kop ML, Memetovic J, Smillie K, et al. Use of the WelTel mobile health intervention at a tuberculosis clinic in British Columbia: a pilot study. JournalMTM 2013;2(3):7-14. [doi: 10.7309/jmtm.2.3.3]
19. Poureslami I, Shum J, Lester RT, Tavakoli H, Dorscheid DR, FitzGerald JM. A pilot randomized controlled trial on the impact of text messaging check-ins and a web-based asthma action plan versus a written action plan on asthma exacerbations. J Asthma 2018 Oct 16;56(8):1-13. [doi: 10.1080/02770903.2018.1500583] [Medline: 30003851]
20. Zafar SN, Hazlewood G, Dhiman K, et al. "How are you?" Perspectives from patients and health care providers of text messaging to support rheumatoid arthritis care: a thematic analysis. ACR Open Rheumatol 2024 May;6(5):276-286. [doi: 10.1002/acr2.11652] [Medline: 38376004]
21. Jongbloed K, Pearce ME, Thomas V, et al. The Cedar Project - mobile phone use and acceptability of mobile health among young Indigenous people who have used drugs in British Columbia, Canada: mixed methods exploratory study. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e16783. [doi: 10.2196/16783] [Medline: 32716311]
22. Linden IA, Mar MY, Werker GR, Jang K, Krausz M. Research on a vulnerable neighborhood-the vancouver downtown eastside from 2001 to 2011. J Urban Health 2013 Jun;90(3):559-573. [doi: 10.1007/s11524-012-9771-x] [Medline: 23161093]
23. Yu J, Dwyer-Lindgren L, Bennett J, et al. A spatiotemporal analysis of inequalities in life expectancy and 20 causes of mortality in sub-neighbourhoods of Metro Vancouver, British Columbia, Canada, 1990-2016. Health Place 2021 Nov;72:102692. [doi: 10.1016/j.healthplace.2021.102692] [Medline: 34736154]
24. City of Vancouver. Jobs and economy snapshot for the Downtown Eastside. 2019. URL: https://vancouver.ca/files/cov/jobs-and-economy-snapshot-dtes.pdf [accessed 2024-07-14]
25. Ivsins A, Benoit C, Kobayashi K, Boyd S. From risky places to safe spaces: re-assembling spaces and places in Vancouver's Downtown Eastside. Health Place 2019 Sep;59:102164. [doi: 10.1016/j.healthplace.2019.102164] [Medline: 31382220]
26. Pakhomova TE, Nicholson V, Fischer M, et al. Exploring primary healthcare experiences and interest in mobile technology engagement amongst an urban population experiencing barriers to care. Qual Health Res 2023 Jul;33(8-9):765-777. [doi: 10.1177/10497323231167829] [Medline: 37225177]
27. City of Vancouver. Downtown Eastside plan three year summary of implementation (2017-2019). 2020. URL: https://vancouver.ca/files/cov/dtes-plan-implementation-summary-2017-2019.pdf [accessed 2024-07-14]
28. BC Center for Excellence in HIV/AIDS. Improving patient engagement and retention in healthcare at Hope to Health. URL: https://bccfe.ca/news/forecast/improving-patient-engagement-and-retention-healthcare-hope-health [accessed 2024-07-14]
29. Primary care clinic. BC Centre for Excellence in HIV/AIDS. URL: https://bccfe.ca/primary-care-clinic [accessed 2024-10-25]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2022 Jun;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]

XSL•FO
RenderX

32. Smillie K, Van Borek N, Abaki J, et al. A qualitative study investigating the use of a mobile phone short message service designed to improve HIV adherence and retention in care in Canada (WelTel BC1). J Assoc Nurses AIDS Care 2014;25(6):614-625. [doi: 10.1016/j.jana.2014.01.008] [Medline: 24768442]
33. Lester RT. Ask, don't tell - mobile phones to improve HIV care. N Engl J Med 2013 Nov 7;369(19):1867-1868. [doi: 10.1056/NEJMc1310509] [Medline: 24195570]
34. Anthony N, Molokwu J, Alozie O, Magallanes D. Implementation of a text message to improve adherence to clinic and social service appointments. J Int Assoc Provid AIDS Care 2019;18:2325958219870166. [doi: 10.1177/2325958219870166] [Medline: 31456465]
35. Schwebel FJ, Larimer ME. Using text message reminders in health care services: a narrative literature review. Internet Interv 2018 Sep;13:82-104. [doi: 10.1016/j.invent.2018.06.002] [Medline: 30206523]

Edited by L Buis; submitted 08.11.24; peer-reviewed by C Baxter, KJC De Angeles; revised version received 29.07.25; accepted 22.10.25; published 27.11.25.

Please cite as:
Harrison L, Fulton C, Marante Changir A, Ajidahun A, Tam C, Zhang W, Shen T, Barrios R, Montaner J, Salters K, Lester RT, Parashar S, Moore DM
Evaluation of a Pilot mHealth Intervention to Engage Primary Care Clients at an Urban Clinic Serving Marginalized Populations: Mixed-Methods Cohort Study
JMIR Mhealth Uhealth 2025;13:e68561
URL: https://mhealth.jmir.org/2025/1/e68561
doi:10.2196/68561

© Lauren Harrison, Christina Fulton, Antonio Marante Changir, Adedayo Ajidahun, Cassidy Tam, Wendy Zhang, Tian Shen, Rolando Barrios, Julio Montaner, Kate Salters, Richard T Lester, Surita Parashar, David M Moore. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Implementation and Evaluation of a Virtual Transitional Care Intervention Using Automated Text Messaging and Virtual Visits After Emergency Department Discharges: Retrospective Cohort Study

Grace Lee[1,2], BA; Courtenay Bruce[2,3], MA, JD; Tariq Nisar[2], MPH; Brendan Holderread[1,2], MD; Sarah Pletcher[1,2,4], MD, MHCDS; Ngoc Anh Nguyen[1,2,5], MD

[1]Center for Connected Care, Innovation & Implementation - Research, Houston Methodist Hospital, 6445 Main Street, Houston, TX, United States

[2]Houston Methodist Academic Institute, Houston Methodist Hospital, Houston, TX, United States

[3]System Quality & Patient Safety, Houston Methodist Hospital, Houston, TX, United States

[4]Department of Surgery, Houston Methodist Hospital, Houston, TX, United States

[5]Department of Medicine, Houston Methodist Hospital, Houston, TX, United States

Corresponding Author:
Ngoc Anh Nguyen, MD
Center for Connected Care, Innovation & Implementation - Research, Houston Methodist Hospital, 6445 Main Street, Houston, TX, United States

Abstract

Background: Emergency department (ED) overcrowding and avoidable revisits challenge health systems, with approximately 20% of patients returning within 30 days. ED-based transitional care interventions, including automated SMS text messaging, offer scalable, cost-effective means to improve follow-up, though evidence remains limited.

Objective: This study evaluated a transitional care intervention combining SMS text messaging and virtual transitional care visits to reduce ED revisits and improve outpatient follow-up.

Methods: This retrospective observational cohort study included patients discharged from 4 EDs within a single US health system between September 2023 and September 2024. Patients were categorized into two groups based on intervention engagements: (1) completed (requested, scheduled, and completed a visit) and (2) noncompleted (requested, scheduled, and did not complete). The primary outcome was spontaneous, unplanned ED revisits within 90 days; secondary outcomes included outpatient follow-up and time to first outpatient evaluation. Between-group differences were assessed using descriptive statistics and multivariable regression models (with $P<.05$ considered statistically significant).

Results: Of 68,115 discharged patients, 42.72% (29,100/68,115) received an automated SMS text messaging for the virtual transitional care program, and 2.93% (853/29,100) accessed the scheduling link. Of these, 56.5% (482/853) requested a visit, 49.8% (240/482) scheduled, and 70% (168/240) completed the visit (completed group). Among 72 noncompleted patients, 57% (n=41) did not show, 32% (n=23) canceled, and 11% (n=8) scheduled 2 appointments but completed neither. Nearly half (35/72, 49%) of the noncompleted group had a subsequent ambulatory follow-up. Demographics, comorbidities, and acuity were similar. The noncompleted group was nearly twice as likely to return to the ED within 90 days (21/72, 29% vs 28/150 18.7%; $\chi^2_1=4.20$, $P=.04$; odds ratio 2.11, 95% CI 1.02 - 4.33), while the completed group was more likely to complete outpatient follow-up (35/72, 49% vs 51/168, 30.4%; $\chi^2_1=6.60$, $P=.01$; odds ratio 2.15, 95% CI 1.03 - 4.77). Time to first outpatient visit did not differ significantly between groups (mean 15.7, SD 19.0 d vs mean 19.8, SD 20.7 d; $\Delta\beta=-1.93$, 95% CI –10.09 to 6.42; $P=.65$).

Conclusions: A combined SMS text messaging and virtual transitional care program lowered 90-day ED revisits and increased outpatient follow-up, but engagement was low (2.9%). Future work should focus on optimizing care delivery and developing strategies to expand reach across the broader ED discharge population.

(JMIR Mhealth Uhealth 2025;13:e77973) doi:10.2196/77973

KEYWORDS

transitional care intervention; SMS text messaging; virtual care; telemedicine; emergency department; hospital readmission; postdischarge follow-up

XSL•FO
RenderX

Introduction

Emergency department (ED) overcrowding and avoidable return visits represent significant challenges for health care systems, with approximately 20% of patients returning to the ED within 30 days of discharge [1]. ED return visits can signal gaps in postdischarge care coordination, unresolved clinical concerns, or inadequate follow-up, all of which contribute to worse patient outcomes and increased strain on health care resources. Notably, about 14% of ED return visits result in hospital admission, underscoring the role of the ED as a critical entry point for both initial and subsequent acute care use [2]. These findings suggest that timely post-ED follow-up could help prevent return visits and downstream hospitalizations. This dual impact—reducing ED overcrowding while preventing readmissions—makes ED-discharged patients a strategic target for transitional care interventions (TCIs).

A wide range of TCIs have been implemented to reduce avoidable acute care reuse with varying outcomes. Timely face-to-face primary care follow-ups have been shown to reduce readmissions, but only 30% to 49% of patients secure an appointment within 14 to 28 days after discharge [3,4]. This gap is often even wider among patients discharged from the ED, who may lack a primary care provider, health insurance, or familiarity with outpatient care navigation, factors that further complicate successful follow-up. Telephone outreach is an area of potential promise that has previously demonstrated mixed results. A retrospective cohort study analyzing all patients discharged from 15 Californian hospitals over 2 years found that telephone follow-up had a marginal impact on 30-day readmissions in high-risk medicine patients and no effect in low-risk patients [5]. In contrast, Constantino et al [6] demonstrated that telephone intervention reduced ED revisits and 30-day hospital readmissions in Medicare Advantage beneficiaries enrolled in a large national health care company. Other TCIs, such as postdischarge home visits, multidisciplinary transition care clinics, and pharmacist-led medication management, have been either ineffective overall, effective only in specific subgroups, or inconclusive due to limited enrollment [7-9].

These inconsistent outcomes, likely driven by the heterogeneity of patient populations across prior studies, combined with the labor-intensive and logistically complex nature of these interventions, raise concerns about their scalability and long-term sustainability. These challenges highlight the need for a TCI following acute care or ED visits that is effective; cost-efficient; and (1) broadly implementable across diverse patient populations while allowing flexibility for individual patient needs and preferences, (2) capable of providing timely follow-up without delays, and (3) scalable across different health care settings and clinical contexts. SMS text messaging and virtual transitional care visits have significant potential to address these gaps. SMS text messaging is cost-effective, widely accessible to all patients with mobile devices, and convenient due to its asynchronous nature [10]. SMS text messaging outreach further enables clinicians to target defined patient populations or permit text recipients to self-select into TCI. This flexibility facilitates the identification of patients most likely to benefit from additional postdischarge care navigation, thereby optimizing resource use while maintaining broad implementability.

In this study, we describe the conceptualization, development, implementation, and outcomes of a novel virtual TCI that integrates automated unidirectional SMS text messaging with virtual transitional care visits. While the visits were staffed by the system's existing virtual urgent care (VUC) team, this was not a standard on-demand VUC encounter. Recognizing that the ED frequently represents the first step in the hospital readmission cycle, this intervention specifically targeted patients discharged directly from the ED through a virtual transitional care program designed to reconnect them with timely care, reinforce discharge instructions, and route unresolved concerns to appropriate next steps.

Methods

Study Design and Setting

This retrospective observational cohort study analyzed patients discharged from four EDs within our health system between September 2023 and September 2024. This study was reviewed and classified as a quality improvement initiative and "Not Human Research" by the health system's institutional review board.

The study was conducted within a nonprofit health system in Houston, Texas, consisting of 1 academic quaternary hospital (1020 operating beds), 6 community hospitals (average of 277 beds each), and 1 long-term acute care facility (98 beds) [11]. The academic hospital also receives patients from 3 surrounding freestanding emergency care centers (ECCs).

VUC Program Background

The system launched its VUC program in 2019 as part of an institutional innovation initiative to expand access to on-demand care through virtual platforms. For standard urgent care needs, patients can directly schedule VUC appointments online through the health system's patient portal for new acute complaints. The VUC service operates 24/7/365 and is staffed by family nurse practitioners under the supervision of an emergency medicine physician. Patients can connect with providers via interactive, 2-way audio-video from their homes, offering convenient, on-demand care. In addition to treating the usual urgent care conditions, the program also facilitates referrals for diagnostics, establishing care with a primary care physician, and specialty care referrals. During the COVID-19 pandemic, VUC was also used to assess and refer patients to the health system's monoclonal antibody infusion program. Establishing accessible referral pathways was critical to managing high patient volumes, and the VUC program served as a reliable source alongside freestanding ECCs and hospital-based EDs [12].

By April 28, 2019, the program had completed over 4500 VUC visits, with an average patient satisfaction score of 4.9 out of 5 [13]. Building on this early success, the VUC infrastructure and staffing model were adapted in 2023 to support a distinct use case: virtual transitional care visits for patients recently discharged from the ED. This adaptation was designed to leverage existing virtual care capabilities to address gaps in

XSL•FO
RenderX

post-ED discharge care and represents the intervention evaluated in this study. Unlike standard VUC encounters, which address new symptoms, the virtual transitional care visits described in this study were initiated through our ED discharge workflows and specifically focused on postdischarge care coordination and assessment of ongoing concerns related to the initial ED visit.

Participants

Eligibility Criteria

Inclusion Criteria

Eligibility required a documented phone number in the Epic electronic medical record (EMR) system (Epic Systems Corp) and aged ≥2 years. SMS text messaging was covered under the hospital-wide consent for postdischarge contact obtained at ED registration. Not all ED-discharged patients during the study period (September 2023 to September 2024) were assigned to receive automated SMS text messaging due to phased implementation and evolving eligibility criteria. During the pilot phase (September 2023 to February 2024), assignment to the virtual transitional care program was limited to patients diagnosed with chronic obstructive pulmonary disease, pneumonia, upper respiratory infection, or COVID-19. This respiratory-focused pilot approach served dual purposes: enabling targeted VUC nurse practitioner training in respiratory care management and chest imaging interpretation while establishing standardized virtual transitional care protocols for visit workflows, documentation, and care coordination before expanding to all ED discharge diagnoses. Beginning in March 2024, virtual transitional care pathway assignment expanded to include all discharged patients, regardless of diagnosis.

Exclusion Criteria

Patients who were directed to return to the ED by a VUC nurse practitioner were excluded from the outcome analysis. These cases represented provider-directed referrals for emergent evaluation, distinct from the spontaneous, patient-initiated ED revisits that this study sought to assess.

Group Classification

Patients were categorized into two comparison groups based on their level of engagement with the virtual TCI:

1. Completed virtual transitional care visit group: patients who received an automated SMS text messaging, requested and scheduled a virtual transitional care visit through a patient service representative (PSR), and *completed* the visit. These patients are referred to as the “completed group” or “completed patients” throughout this study.
2. Noncompleted virtual transitional care visit group: Patients who received the SMS text messaging, requested and scheduled a virtual transitional care visit with a PSR, but *did not complete* the visit (due to a cancelation or no-show). These patients are referred to as the “noncompleted group” or “noncompleted patients” throughout this paper.

Intervention: Virtual Transitional Care Program

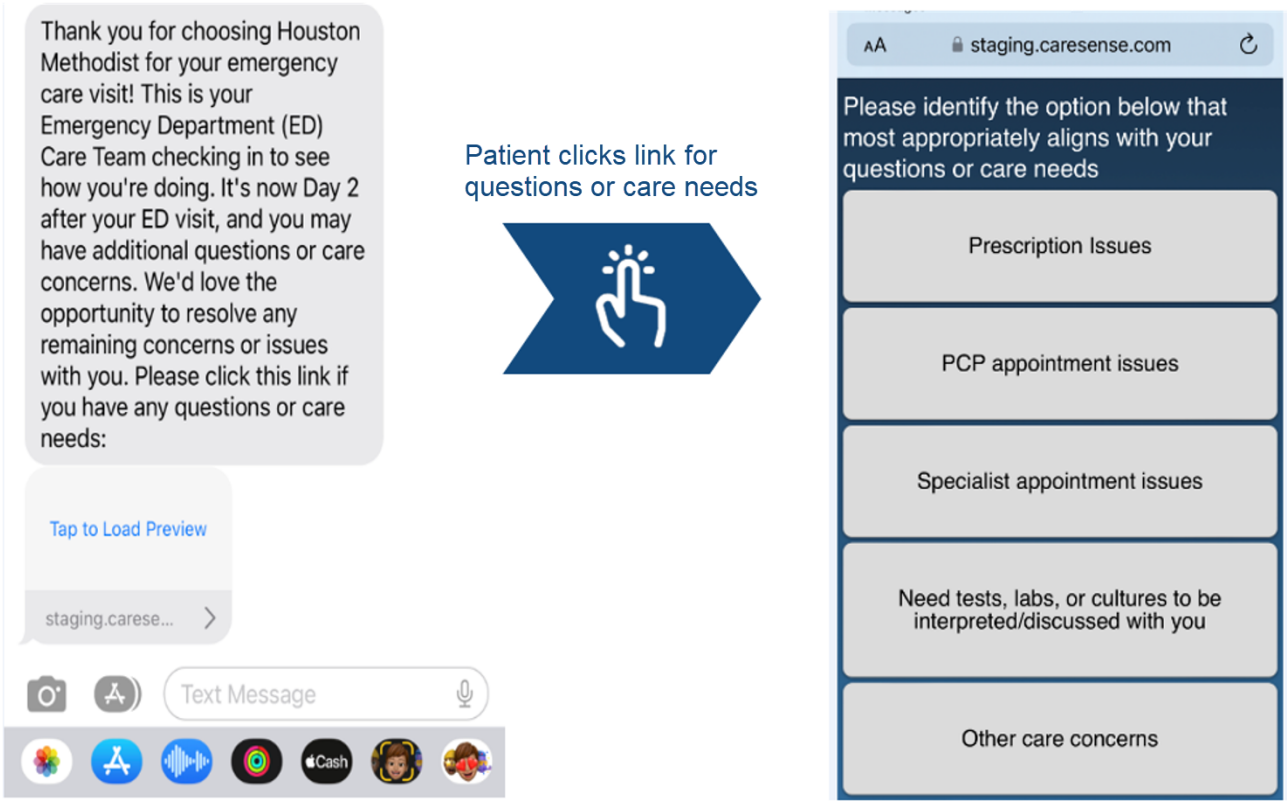
The virtual transitional care program was developed with the following objectives: (1) direct patients with nonemergent concerns to alternate care options to reduce ED overcrowding and boarding times, (2) facilitate timely transitions to outpatient care through the existing VUC infrastructure, (3) improve continuity of care and prevent patients from becoming “lost to follow-up” after ED discharge, (4) enhance patient satisfaction through additional resources and referrals, and (5) provide timely clinical reassessment by trained clinicians.

Program Workflow and Technology Platform

The program leveraged *CareSense* (MedTrak, LLC), a third-party digital engagement platform, configured to align with institutional security and communication protocols. Within 48 hours of discharge, eligible patients receive an automated SMS text message with a link to a brief questionnaire (Figure 1).

XSL•FO
RenderX

Figure 1. Day 2 postdischarge SMS text messaging and care needs survey interface platform. Clinical concerns route to virtual transitional care scheduling, while administrative concerns route to referral management. PCP: primary care physician.



Patient responses trigger different care pathways based on concern type. Clinical concerns (prescription issues, need for tests, laboratory results discussion, or "other") route patients to PSRs who schedule virtual transitional care visits through Epic or MyChart infrastructure. Administrative concerns (appointment scheduling issues) redirect patients to the referral management center for outpatient appointment scheduling, bypassing the virtual transitional care pathway.

Virtual Transitional Care Visit Process

For patients routed to virtual transitional care, their medical record numbers are queued in the CareSense database for PSR review, who then contact patients by phone to confirm ongoing concerns and schedule virtual transitional care visits. Although visits are staffed by the health system's VUC clinicians, these are structured follow-ups specifically designed for post-ED discharge care coordination rather than standard urgent care encounters addressing new acute symptoms. VUC nurse practitioners assess patients clinically to determine whether care plans require continuation, modification, or escalation to an in-person evaluation.

Outcomes and Variables

Primary Outcome

The primary outcome was the occurrence of spontaneous ED revisits within 90 days for the same or a similar chief complaint as the initial ED visit. Spontaneous revisits were defined as patient-initiated encounters, excluding those directed by health care providers.

Secondary Outcomes

Secondary outcomes included the rate of outpatient follow-up and the number of days to the first outpatient evaluation following ED discharge.

Data Collection

Patient data were collected through manual chart review using the Epic EMR. Variables collected include patient demographics (age, gender, race, and ethnicity), Emergency Severity Index (ESI) at the time of the ED visit, and 6 common comorbidities (coronary artery disease, chronic obstructive pulmonary disease, congestive heart failure, hypertension, diabetes mellitus, and transplant status). Additional data included ED visit details (date and care received), follow-up appointments (virtual transitional care visits, primary care, and specialist visits), and, if applicable, dates and chief complaints for ED revisits.

Statistical Analysis

Descriptive statistics were calculated for baseline characteristics. A generalized linear model was used to analyze the continuous outcome (time to first outpatient evaluation in days), and a logistic regression model was built for the binary outcome (occurrence of spontaneous ED revisits within 90 days). Multivariable models were adjusted for patient demographics (age, gender, race, and ethnicity) and clinical characteristics (ESI and comorbidities). Statistical significance was established at $P<.05$. All analyses were performed in R (version 4.4.1; R Foundation for Statistical Computing).

Ethical Considerations

This study was reviewed and classified as a quality improvement initiative and "Not Human Research" by the health system's

XSL•FO
RenderX

institutional review board, exempting it from full human subjects research approval under institutional guidelines for quality improvement activities. No informed consent was required as this study involved retrospective analysis of existing medical records and quality improvement data. Patients received standard clinical care and routine follow-up outreach as part of established care protocols. All patient data were deidentified for analysis purposes. Data collection was conducted through the institution's secure Epic EMR system with appropriate access controls and institutional data security protocols in place to safeguard patient information. No compensation was provided to participants as this was a retrospective analysis of routine clinical care data without direct patient participation.

Results

Study Population and Engagement Cascade

Over the study period, the 4 EDs recorded 111,648 total encounters, including 38,155 admissions, 68,115 discharges, 741 transfers, and 4637 other dispositions. Of the 68,115 discharges, 51.76% (35,256/68,115) were assigned to the virtual transitional care pathway. This percentage reflects the program's phased rollout, in which eligibility was initially restricted to selected freestanding ECCs and 4 diagnoses before expansion to the main academic ED and all diagnoses. Of the 35,256 pathway assignments, 82.54% (29,100/35,256) received an SMS text message. The remaining cases were excluded because patients either had no phone number on file or had a landline number ineligible for text delivery.

Among text recipients, 2.93% (853/29,100) clicked the scheduling link. Among those who clicked the link, 56.5% (482/853) requested a virtual transitional care visit, and 49.8% (240/482) of requesters successfully scheduled an appointment. Ultimately, 168 patients (70% of those scheduled) completed the visit, while 72 (30%) did not. The stepwise retention of patients through this engagement pathway is presented in Table 1. Figure 2 provides a visual summary of the engagement pathway and illustrates how patients were categorized into the completed and noncompleted virtual transitional care visit groups.

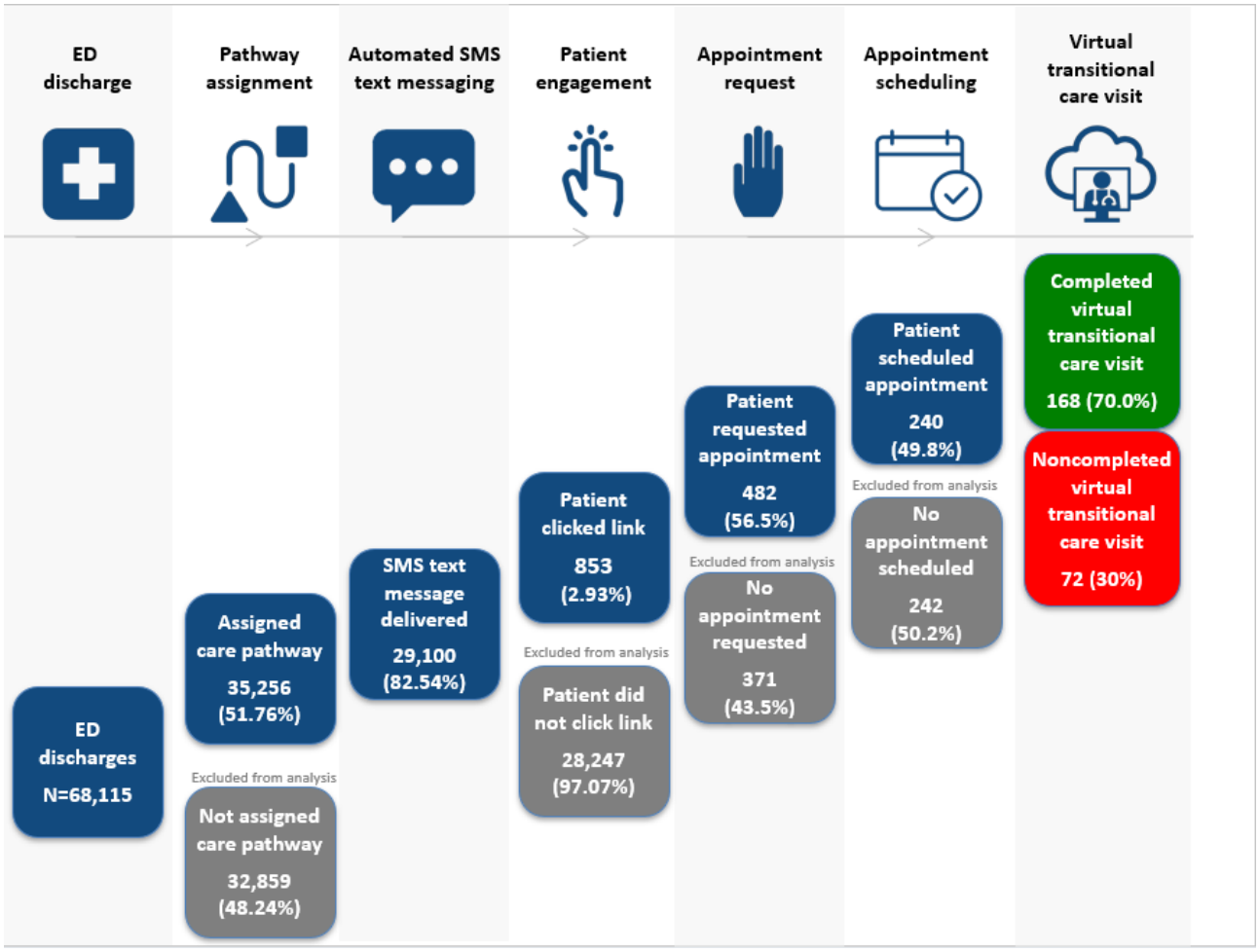
Table . Patient engagement cascade following emergency department (ED) discharge (N=68,115).

Event in engagement cascade	Total, n	Prior step, n/N (%)	Total ED discharges (%)
Total ED discharges	68,115	N/A[a]	100
Assigned to virtual transitional care pathway	35,256	51.76 (35,256/68,115)	51.76
Received automated text message	29,100	82.54 (29,100/35,256)	42.72
Clicked scheduling link	853	2.93 (853/29,100)	1.25
Requested a virtual transitional care visit	482	56.51 (482/853)	0.71
Scheduled an appointment	240	49.79 (240/482)	0.35
Completed (completed virtual transitional care visit group)	168	70.00 (168/240)	0.25
Did not complete visit (noncompleted virtual transitional care visit group)	72	30.00 (72/240)	0.11

[a]N/A: not applicable.

XSL•FO
RenderX

Figure 2. Emergency department (ED) discharge pathway and group classification by virtual transitional care visit completion status.



Engagement Funnel Analysis

The virtual transitional care visit program demonstrated substantial attrition at each engagement step. Of the 68,115 total discharged patients, only 0.25% (168/68,115) ultimately completed a virtual transitional care visit. The largest drop-off occurred at the initial SMS text messaging stage, where 97.07% (28,247/29,100) of recipients did not click the scheduling link. Among those who did engage, conversion rates improved: 56.5% (482/853) of link-clickers requested visits, 49.8% (240/482) of requesters successfully scheduled appointments, and 70% (168/240) of scheduled patients completed their visits. This pattern indicates that while initial engagement was challenging, patients who demonstrated early interest showed progressively higher commitment to completing the care pathway.

Baseline Characteristics

As shown in Table 2, patients in both completed and noncompleted groups were predominantly women (71.4% and 72.2%, respectively). The mean age was similar between groups (54.7, SD 17.7 y completed group versus 52.7, SD 21.6 y noncompleted group; P=.55). There were no significant differences between the groups in terms of race, ethnicity, or clinical comorbidities. The mean ESI was identical in both groups at 3.3 (SD 0.6), indicating comparable clinical acuity.

XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics.

Characteristic	Completed (n=168)	Noncompleted (n=72)	*P* value
Age (y), mean (SD)	54.7 (17.7)	52.7 (21.6)	.55
Gender, n (%)			
Women	120 (71.4)	52 (72.2)	>.99
Men	48 (28.6)	20 (27.8)	>.99
Race, n (%)			.19
African American	55 (32.7)	33 (45.8)	
Asian	8 (4.8)	4 (5.6)	
Other	6 (3.6)	2 (2.8)	
White	99 (58.9)	33 (45.8)	
Ethnicity, n (%)			
Non-Hispanic	141 (84.9)	63 (88.7)	.54
Hispanic	25 (15.1)	8 (11.3)	.54
Comorbidities, n (%)			
CAD[a]	147 (87.5)	58 (80.6)	.17
CHF[b]	159 (94.6)	64 (88.9)	.17
COPD[c]	164 (97.6)	70 (97.2)	>.99
Hypertension	74 (44)	34 (47.2)	.67
DM[d]	132 (78.6)	58 (80.6)	.86
Transplant	166 (98.8)	72 (100)	>.99
Emergency Severity Index, mean (SD)	3.3 (0.6)	3.3 (0.6)	.49

[a]CAD: coronary artery disease.

[b]CHF: congestive heart failure.

[c]COPD: chronic obstructive pulmonary disease.

[d]DM: diabetes mellitus.

At the time of their ED visit, most patients in both groups had at least 1 provider of record. Among patients in the completed group, 26.2% (44/168) were linked to a primary care physician within the health system and 28% (47/168) to an outside primary care physician. In contrast, 23.6% (17/72) of the noncompleted group were linked to a primary care physician within the health system and 20.8% (15/72) to an outside primary care physician. From both groups, 13 (7.7%) patients in the completed group and 11 (15.2%) in the noncompleted group had no provider listed. Although not statistically analyzed, these differences may reflect broader patterns in health care access and engagement.

Clinical Interventions Delivered During Virtual Transitional Care Visits

Out of 168 patients in the completed group, 106 (63.1%) received additional care from VUC nurse practitioners during their visit. The most common interventions included issuing a new prescription (39/113, 34.5%) and directing patients back to the ED (28/113, 24.8%). Other actions included medication adjustments (14/113, 12.4%), referrals to specialists (15/113, 13.3%) or primary care providers (9/113, 7.9%), and instructions to return to urgent care (4/113, 3.5%). A small number of patients received other support such as documentation or clarification of care instructions (3/113, 2.7%). Because some patients received more than one intervention, the total number of interventions (N=113) exceeds the number of patients who received them (106/168, 63.1%). The remaining 62 (36.9%) patients were evaluated but did not require any further clinical action; their concerns were either resolved, self-limited, or deemed appropriate for routine outpatient follow-up without nurse practitioner–initiated intervention. All patients, regardless of whether a clinical intervention was provided, were encouraged to follow up with their outpatient provider.

Notably, 24.8% (28/113) of patients who received a clinical intervention—16.7% (28/168) of all patients in the completed group—were directed back to the ED by nurse practitioners, representing a substantial proportion that required escalation to higher-level care. Of the 28 (16.7%) patients directed to return to the ED, 18 (64%) followed through with this recommendation, indicating good patient adherence to clinical guidance provided during virtual visits.

XSL•FO
RenderX

Outcomes

We assessed three outcomes: spontaneous ED revisits within 90 days of discharge, outpatient follow-up rates, and time to first outpatient evaluation.

Patients (29%, 21/72) in the noncompleted group were almost twice as likely to revisit the ED within 90 days compared to the completed group (18.7%, 28/150). This difference remained significant in a multivariate logistic regression model adjusting for demographics, comorbidities, and ESI. As shown in Table 3, patients who did not complete a virtual transitional care visit were more than twice as likely to return to the ED within 90 days (odds ratio 2.11; 95% CI 1.02 - 4.33; *P*=.04). Among the completed group patients who were advised by a nurse practitioner to return to the ED, 18 out of 28 (64%) followed through.

Table . Multivariate logistic regression: predictors of spontaneous emergency department (ED) revisits within 90 days.

Variables	Odds ratio (95% CI)	*P* value
(Intercept)	0.31 (0.08-1.22)	.09
Group		
Noncompleted (ref=completed)	2.11 (1.02-4.33)	.04
Age (y)	0.99 (0.97-1.02)	.48
Gender		
Men (ref=women)	0.75 (0.34-1.7)	.50
Race		
Asian (ref=White)	0 (0-inf)[a]	.99
Black (ref=White)	0.58 (0.26-1.32)	.19
Others (ref=White)	1.09 (0.11-11.08)	.94
Ethnicity		
Hispanics (ref=non-Hispanics)	0.67 (0.21-2.12)	.49
Comorbidities (yes)		
Congestive heart failure	0.68 (0.15-3.09)	.62
Coronary artery disease	0.97 (0.36-2.63)	.95
Chronic obstructive pulmonary disease	3.43 (0.56-20.9)	.18
Diabetes mellitus	1.32 (0.53-3.29)	.55
Hypertension	1.85 (0.75-4.54)	.18
Transplant	4.94 (0.24-100.17)	.30

[a]There is no difference between White versus Asian patients with the outcome (ED visit), that is, they are extremely similar or have extreme similar patterns.

On average, patients in the completed group completed their virtual transitional care visit within 3.83 (SD 1.22) days. They were also twice as likely to complete outpatient follow-up after ED visit compared to the noncompleted group (48.6% vs 30%, respectively; odds ratio 2.17; 95% CI 1.23 - 3.83; *P*=.01).

Although patients in the completed group were more likely to follow up, there were no significant differences in the timing of the first outpatient evaluation between the groups. As shown in Table 4**,** the mean time to first outpatient visit was 15.7 (SD 19.0) days in the noncompleted group versus 19.8 (SD 20.7) days in the completed group (Δβ=−1.93, 95% CI −10.09 to 6.42; *P*=.65).

XSL•FO
RenderX

Table . Multivariate linear regression: timing of first outpatient evaluation.

Variables	Estimated days (95% CI)	*P* value
(Intercept)	9.69 (−6.68 to 26.05)	.25
Group		
Noncompleted (ref=completed)	−1.93 (−10.29 to 6.42)	.65
Age (y)	0.23 (−0.03 to 0.49)	.09
Gender		
Men (ref=women)	2.80 (−5.74 to 11.35)	.52
Race		
Asian (ref=White)	2.97 (−14.60 to 20.54)	.74
Black (ref=White)	−2.66 (−11.27 to 5.94)	.55
Others (ref=White)	12.22 (−9.99 to 34.43)	.28
Ethnicity		
Hispanics (ref=non-Hispanics)	−3.17 (−15.92 to 9.58)	.63
Comorbidities (yes)		
Congestive heart failure	−11.18 (−25.34 to 2.98)	.12
Coronary artery disease	−5.17 (−14.71 to 4.38)	.29
Chronic obstructive pulmonary disease	5.00 (−14.39 to 24.39)	.61
Diabetes mellitus	0.12 (−9.20 to 9.44)	.98
Hypertension	−3.84 (−13.00 to 5.32)	.41
Transplant	2.99 (−28.07 to 34.04)	.85

Discussion

Principal Findings

Among 68,115 ED-discharged patients, we implemented a TCI using automated SMS text messaging linked to a virtual transitional care program. While only 0.25% (168/68,115) of discharged patients ultimately completed a virtual transitional care visit, those who did demonstrated significantly improved outcomes: 53% lower odds of 90-day ED revisits and doubled rates of outpatient follow-up completion compared to patients who scheduled but did not attend virtual transitional care visits. These findings suggest the intervention effectively identified and served patients with ongoing postdischarge needs while maintaining high operational efficiency through patient self-selection.

To our knowledge, this study is the first to describe a virtual TCI that integrates automated text messaging with virtual transitional care visits, representing a novel application of existing virtual care infrastructure for postdischarge care coordination.

Impact on ED Revisits

Our finding was a lower rate of spontaneous ED revisits among patients who completed a virtual transitional care visit compared to the noncompleted group. By excluding provider-directed revisits, we isolated the effect of postdischarge care through virtual transitional care visits. These results contrast with findings from Bressman et al [14], whose similar SMS text messaging TCI did not reduce ED revisit rates within 30 days. One probable factor for this variance is that their population was limited to patients who had already established care at primary care practices within their health system. We build on their work by expanding our TCI's target population to all patients discharged from the ED, including those without a primary care provider and those not established in our health system. This allowed our program to reach a broader, more diverse patient population with varying health care access, which may explain the differential impact on revisit rates.

The substantial proportion of patients (28/168, 16.7% of the completed group) who were directed back to the ED during virtual transitional care visits deserves particular attention. This rate is notably higher than typical urgent care referral patterns and likely reflects the postdischarge context, in which patients may have evolving or incompletely resolved conditions. Rather than representing a program failure, this pattern suggests successful identification of patients requiring escalated care who might otherwise have delayed seeking appropriate treatment. 18 out of 28 (64%) patients adhered to instructions to return to the ED, underscoring the clinical value of structured postdischarge follow-up. The case examples in Multimedia Appendix 1 illustrate how virtual reassessment enabled timely recognition of clinical deterioration, potentially preventing more serious complications. However, this high referral rate also highlights the complexity of post-ED discharge care and implies that a significant subset of discharged patients may benefit from more intensive monitoring or alternative care models.

XSL•FO
RenderX

Outpatient Follow-Up Engagement

Our major finding was that completed participants were significantly more likely to complete an outpatient follow-up visit compared to those in the noncompleted group. Similar trends have been reported by Bhandari et al [15] and Costantino et al [6]. However, this finding should be interpreted cautiously given potential selection bias, as patients who completed virtual transitional care visits likely represent a more engaged population.

Engagement Patterns and Gender Distribution

While 72.0% (121/168) of the patients who used virtual transitional care visits were women, a distribution consistent with higher telehealth uptake among women during the VUC program pilot study [13] and prior studies [16-18], this pattern is especially notable given the ED context. It suggests that gender-based consumer preferences for timely, convenient care persist even in postacute scenarios, potentially shaping who engages with digital transitional care models.

Despite this observed preference pattern, overall engagement with the virtual transitional care program was lower than expected. Only 2.93% (853/29,100) of SMS text messaging recipients clicked the virtual transitional care link, and just over half of those (482/853, 56.5%) requested a visit. This low engagement rate contrasts with the higher engagement observed in prior studies using bidirectional texting, which allowed patients to respond to the SMS text messaging outreach [10,14]. One possible contributor to the low uptake was the wording of the SMS text messaging: patients without “remaining concerns or issues” may have been unlikely to click the scheduling link, and they did not have the opportunity to ask clarifying questions as would have been possible in a bidirectional model. While the automated unidirectional texting approach helped minimize staffing burden and operational complexity, it limited real-time responsiveness. Future iterations of the program may benefit from a more seamless, patient-driven experience, such as direct scheduling or on-demand access, as well as exploration of bidirectional texting models.

Workflow Considerations and Operational Efficiency

Additional patients were lost when PSRs reached out via phone to confirm the appointment request and schedule a virtual transitional care visit. This outcome was somewhat anticipated, as telephone outreach necessitates the patient’s availability at the time the PSR places the call. Based on operational workflows, patients typically waited 1 to 2 days between clicking the link and completing their scheduled virtual transitional care visit. This added step and subsequent delays may have introduced friction or led to disengagement, particularly for those with lower-acuity concerns or limited availability.

However, this step was included in the program workflow to address concerns about potentially unmanageable, high ED-discharge patient volume flowing into the virtual transitional care program (68,115 discharges during the study period). By having PSRs confirm patients’ remaining concerns and schedule virtual transitional care visits, it was also possible to distribute the patient volume throughout each day, making patient visits more manageable for VUC clinicians. PSRs also supported care navigation by redirecting patients with concerns outside of the scope of virtual transitional care, such as requests for controlled substances, to more appropriate care settings. Once scheduled, 70% (168/240) of patients successfully completed their virtual transitional care visit. The high adherence rate indicates this multistep process effectively triages patients with ongoing needs and routes them to appropriate care.

Program Reach and Resource Efficiency

The engagement challenges revealed by the cascade analysis demonstrate both limitations and strengths of this approach. With only 1 in 400 discharged patients ultimately completing a virtual transitional care visit, the program’s reach was more limited than anticipated. The 97% nonresponse rate to text messages represents the most significant barrier and likely reflects multiple factors: SMS text messaging fatigue in health care settings, competing priorities during postdischarge recovery, resolution of concerns before follow-up, and potential technological or literacy barriers.

From an operational perspective, the low engagement rate meant that the 29,100 (82.54%) SMS text messages sent ultimately generated only 168 completed visits (0.58% conversion), making this intervention highly efficient in terms of provider time and resources. Once patients demonstrated initial engagement by clicking the link, conversion rates improved substantially. This suggests the program successfully identified and served patients with the highest need for postdischarge support while avoiding overwhelming virtual transitional care capacity.

Outpatient Care Timing

While the program appears to improve the likelihood of follow-up, there was no significant difference in the timing of the first outpatient evaluation after the ED visit between completed and noncompleted groups. This finding may reflect the retrospective nature of our study, which prevented differentiation between appointments scheduled before versus after the ED visit, and our inclusion of all specialty providers rather than focusing on visits related to the presenting complaint.

Limitations

This study has several important limitations that should be considered when interpreting the findings.

Study Design and Comparison Groups

First, we did not compare the completed group to all ED-discharged patients but instead focused on patients who had both requested and scheduled a virtual transitional care visit. While this cohort design allowed for a more targeted comparison of clinical outcomes, as both groups expressed lingering concerns after their ED discharge, it limits the generalizability of our findings to the broader ED discharge population. The overall sample size (240 patients) also reduced statistical power to detect differences between groups.

The small sample sizes within racial and ethnic subgroups further limited our ability to detect meaningful differences between the completed and noncompleted groups. While we observed variations in racial distribution (White: 99/168, 58.9% vs 33/72, 45.8%; African American: 55/168, 32.7% vs 33/72,

XSL•FO
RenderX

45.8%), the restricted sample size constrained statistical power for subgroup analyses. This represents a missed opportunity for robust statistical inference regarding potential disparities in program access, engagement, or outcomes. Future studies with larger samples are needed to adequately assess whether racial or ethnic factors influence engagement with virtual transitional care visits.

Another important limitation is our inability to compare outcomes between the completed group and the larger population of patients who received automated SMS text messaging but did not access the scheduling link (28,247 patients, representing 97.07% of text recipients). Even the noncompleted group demonstrated engagement by clicking the link and requesting virtual transitional care visits, representing a higher level of activation than the nonclicking population. We did not collect or analyze demographic or clinical characteristics of patients who received texts but did not engage with the scheduling link, which could have provided valuable insights into differences between engaged and nonengaged populations. Furthermore, we did not conduct qualitative interviews to explore reasons for nonengagement. These limitations restrict our understanding of the full spectrum of patient engagement and reduce our ability to evaluate the program's impact across all text recipients.

Selection Bias and Patient Engagement

Second, our findings are significantly influenced by selection bias. While both the completed and noncompleted groups demonstrated initial motivation by requesting and scheduling virtual transitional care visits, patients who actually completed visits may still represent a more activated or health-engaged subgroup compared to those who scheduled but did not attend. This could partially explain the higher observed rate of outpatient follow-up among completed patients. However, the more significant limitation is our inability to compare either engaged group to the 28,247 (97.07%) text recipients who never clicked the scheduling link, representing the vast majority of the ED discharge population.

The low overall engagement with the intervention (only 853/29,100, 2.93% of text recipients clicked the link) raises important questions about program reach. We did not track SMS text messaging read receipts, so it remains unclear whether the 97% nonresponse rate reflects unread messages, conscious decisions not to engage, or technological barriers. This limitation prevents us from distinguishing between patients who were unaware of the opportunity versus those who actively chose not to participate.

It is possible that some patients failed to complete their scheduled virtual transitional care visit because they were already experiencing worsening symptoms or barriers to care, factors that may have directly contributed to their subsequent ED return. As such, the association between visit completion and improved outcomes should be interpreted with caution, given potential confounding by severity, access, or health literacy.

Data Collection and Measurement Limitations

Third, our study's retrospective design introduced limitations in data collection and measurement. It was not possible to differentiate through chart review whether the outpatient appointment was scheduled before or after the ED visit. In addition, we included providers across all specialties for the first outpatient appointment, making it unclear whether patients were seen for the same chief complaint that brought them to the ED or for unrelated issues. For example, if a patient presented to the ED with chest pain and saw an otolaryngologist 3 days later, this would still be recorded as an outpatient follow-up at 3 days after discharge. This limitation may have contributed to the lack of observed difference in follow-up timing between the 2 groups.

External acute care visits, outpatient visits, and other postdischarge outreach could have been missed if they were not available through Epic Care Everywhere for retrospective chart review; however, this would have occurred randomly in both groups and is unlikely to have systematically biased our results.

We did not systematically document PSR contact rates during phone outreach, though anecdotal reports suggest approximately 50% of patients who clicked the scheduling link did not answer subsequent phone calls. This represents an unmeasured but potentially significant source of attrition that limits our understanding of barriers to care access among initially engaged patients and may underestimate true demand for virtual transitional care services.

Accessibility and Generalizability Concerns

This TCI design may have also excluded patients facing language and technological barriers. Those who cannot read English would have been unable to understand the SMS text messaging and would have required additional assistance to participate. Patients with technological difficulties, such as unreliable internet access or unfamiliarity with telehealth, may have been less inclined to click on the link. Alternatively, it is possible that patients who did not click on the virtual transitional visit link had no remaining issues after their ED discharge. In addition, this study was conducted within a single health system, which may limit the generalizability of our findings to other health care settings with different patient populations, technological infrastructure, or staffing models. While this TCI framework demonstrates potential for scalability across diverse health care settings, successful implementation would require customization to accommodate local patient populations, technological capabilities, language preferences, and existing care delivery models. Health systems considering adoption should adapt messaging content, scheduling workflows, and clinical protocols to align with their specific patient demographics and operational constraints.

Future Work

Building on the promising findings of this study, efforts are underway to expand the virtual transitional care program across all hospitals and affiliated freestanding ECCs within our health system. Prior to broader implementation, optimizing program workflows should be considered. The current multistep process—requiring patients to wait for a PSR to call and

XSL•FO
RenderX

schedule a visit after they engage with the SMS text messaging—may have introduced friction and contributed to lower overall participation. While most virtual transitional care visits were scheduled for the day after patient outreach, this 1- to 2-day delay may not align with the expectations of patients seeking more immediate care access. Streamlining this process through integrated self-scheduling or offering on-demand virtual transitional care visits could reduce attrition. In addition, incorporating a bidirectional texting model may enhance responsiveness, foster more dynamic communication, and increase overall engagement, an approach that has shown promise in prior postdischarge care studies. Integrating patient language preferences from Epic into the messaging workflow and tailoring communication accordingly could further extend the program's reach to linguistically diverse populations.

Prior ED Use Patterns

An important area for future investigation involves examining baseline ED use patterns among patients who complete virtual transitional care visits versus those who do not. Understanding whether patients who engage with postdischarge virtual care have different historical patterns of acute care use could contribute key knowledge about patient activation, health-seeking behaviors, and care preferences. Such analysis could help identify whether the program primarily attracts patients who are already engaged with the health care system or successfully reaches patients with traditionally poor care continuity. This information would be crucial for targeting interventions and understanding the mechanisms by which virtual transitional care visits influence subsequent health care use.

Qualitative Research Opportunities

Future studies should incorporate qualitative methods to better understand patient motivations for engagement or nonengagement with postdischarge SMS text messaging and virtual transitional care programs. Interviews with patients who do not click scheduling links could reveal whether nonresponse reflects resolved concerns, technological barriers, competing priorities, or other factors. Similarly, qualitative work with patients who schedule but do not complete virtual transitional care visits could identify modifiable barriers to care completion. This information would be invaluable for refining program design and improving patient engagement strategies.

Validation and Long-Term Impact Studies

While this study offers promising early insights, further research is needed to validate the impact of the virtual transitional care program on acute care reuse using a larger sample size. Advanced analytical techniques such as latent class analysis may help identify patient subgroups with differential responses to virtual TCIs, enabling more targeted and effective program design. The long-term impact of this virtual TCI remains uncertain, particularly with regard to hospital readmission rates, continuity of care from new referrals, and patient satisfaction. If the program demonstrates positive outcomes in these areas, it could offer valuable administrative insights by potentially reducing readmission-related financial penalties and improving patient retention within the health system. Notably, the program was implemented without additional resources or staffing, indicating that meaningful patient engagement and clinical follow-up can be achieved using existing infrastructure. This is a key consideration for long-term sustainability and scalability. Addressing these factors will help refine virtual TCI models and guide broader implementation strategies in postacute transitional care.

Conclusions

The virtual TCI, combining automated SMS text messaging with virtual follow-up visits, demonstrated significant clinical benefits among engaged patients: those who scheduled and completed virtual transitional care visits were less likely to return to the ED within 90 days compared to those who scheduled but did not complete visits. However, given that only 2.9% (853/29,100) of eligible patients engaged with the virtual TCI initially, the broader population-level impact remains limited. While the program demonstrated clinical benefits among engaged patients, the substantial engagement challenges highlight the need for both optimizing the virtual TCI delivery for the small subset who do engage and also developing approaches to expand engagement more broadly across the ED discharge population. In addition, the program's multistep scheduling workflow, requiring phone-based appointment confirmation, may have introduced friction that limited overall participation. Future work should prioritize both streamlined virtual transitional care processes for engaged patients and innovative outreach strategies to reach the 97% who did not initially respond to the SMS text messaging component of the intervention. Streamlined, patient-driven scheduling and bidirectional communication models may further enhance engagement and scalability. Further research is needed to validate the program's impact across larger and more diverse populations.

Acknowledgments

The authors thank the virtual urgent care nurse practitioners, the patient service representatives, the administrative team, Ashleigh Kamencik-Wright, the correspondence teams at CareSense and Epic, Erica Richardson, and Aydin Dursunov for their invaluable contributions to the implementation of the virtual transitional care program, and Sarah Sossong for her contributions to manuscript editing, formatting, and figure development. The authors received no external funding for this work.

XSL•FO
RenderX

Data Availability

The data used in this study were extracted from the institution's electronic medical record system. A deidentified dataset supporting the conclusions of this article is available in Multimedia Appendix 2.

Authors' Contributions

Conceptualization: NAN, GL, CB
Data Curation: GL, BH, NAN
Formal Analysis: GL, NAN, TN
Investigation: GL, BH
Methodology: NAN, GL, CB
Visualization: GL, NAN
Writing – original draft: GL, NAN, CB, BH
Writing – review & editing: NAN, GL, CB, BH, SP
Statistical analysis: TN, GL, NAN
Supervision: NN, SP

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinical case examples: emergency department (ED) return and admission following virtual transitional care visit.
[DOCX File, 15 KB - mhealth_v13i1e77973_app1.docx]

Multimedia Appendix 2
Virtual transitional care program deidentified dataset for statistical analysis.
[XLSX File, 66 KB - mhealth_v13i1e77973_app2.xlsx]

References

1. Duseja R, Bardach NS, Lin GA, et al. Revisit rates and associated costs after an emergency department encounter: a multistate analysis. Ann Intern Med 2015 Jun 2;162(11):750-756. [doi: 10.7326/M14-1616] [Medline: 26030633]
2. Sabbatini AK, Kocher KE, Basu A, Hsia RY. In-hospital outcomes and costs among patients hospitalized during a return visit to the emergency department. JAMA 2016 Feb 16;315(7):663-671. [doi: 10.1001/jama.2016.0649] [Medline: 26881369]
3. Misky GJ, Wald HL, Coleman EA. Post-hospitalization transitions: examining the effects of timing of primary care provider follow-up. J Hosp Med 2010 Sep;5(7):392-397. [doi: 10.1002/jhm.666] [Medline: 20578046]
4. Anderson TS, Herzig SJ, Marcantonio ER, Yeh RW, Souza J, Landon BE. Medicare transitional care management program and changes in timely postdischarge follow-up. JAMA Health Forum 2024 Apr 5;5(4):e240417. [doi: 10.1001/jamahealthforum.2024.0417] [Medline: 38607641]
5. Nguyen HQ, Baecker A, Ho T, et al. Association between post-hospital clinic and telephone follow-up provider visits with 30-day readmission risk in an integrated health system. BMC Health Serv Res 2021 Aug 17;21(1):826. [doi: 10.1186/s12913-021-06848-9] [Medline: 34404408]
6. Costantino ME, Frey B, Hall B, Painter P. The influence of a postdischarge intervention on reducing hospital readmissions in a Medicare population. Popul Health Manag 2013 Oct;16(5):310-316. [doi: 10.1089/pop.2012.0084] [Medline: 23537154]
7. Bloodworth LS, Malinowski SS, Lirette ST, Ross LA. Pharmacist linkage in care transitions: from academic medical center to community. J Am Pharm Assoc 2019;59(6):896-904. [doi: 10.1016/j.japh.2019.08.011] [Medline: 31590926]
8. Rasmussen LF, Barat I, Riis AH, Gregersen M, Grode L. Effects of a transitional care intervention on readmission among older medical inpatients: a quasi-experimental study. Eur Geriatr Med 2023 Feb;14(1):131-144. [doi: 10.1007/s41999-022-00730-5] [Medline: 36564644]
9. Liss DT, Ackermann RT, Cooper A, et al. Effects of a transitional care practice for a vulnerable population: a pragmatic, randomized comparative effectiveness trial. J Gen Intern Med 2019 Sep;34(9):1758-1765. [doi: 10.1007/s11606-019-05078-4] [Medline: 31144279]
10. Bruce C, Pinn-Kirkland T, Meyers A, et al. Investigating patient engagement associations between a postdischarge texting programme and patient experience, readmission and revisit rates outcomes. BMJ Open 2024 Mar 13;14(3):e079775. [doi: 10.1136/bmjopen-2023-079775] [Medline: 38485169]
11. Facts and statistics. Houston Methodist. 2024. URL: https://www.houstonmethodist.org/newsroom/facts-statistics [accessed 2024-06-18]
12. Berry JR, Liebl MG, Todd PH, Brownewell V. Rapid operationalization of COVID-19 monoclonal antibody infusion clinics. NEJM Catal Innov Care Deliv 2021;2(2). [doi: 10.1056/CAT.21.0040]

XSL•FO
RenderX

13. Case study: Virtual health care. Houston Methodist. URL: https://www.houstonmethodist.org/center-for-innovation/our-work/virtual-health [accessed 2024-06-18]
14. Bressman E, Long JA, Burke RE, et al. Automated text message–based program and use of acute health care resources after hospital discharge: a randomized clinical trial. JAMA Netw Open 2024 Apr 1;7(4):e243701. [doi: 10.1001/jamanetworkopen.2024.3701] [Medline: 38564221]
15. Bhandari N, Epane J, Reeves J, Cochran C, Shen J. Post-discharge transitional care program and patient compliance with follow-up activities. J Patient Exp 2022;9:23743735221086756. [doi: 10.1177/23743735221086756] [Medline: 35647268]
16. Chang E, Davis TL, Berkman ND. Differences in telemedicine, emergency department, and hospital utilization among nonelderly adults with limited English proficiency post-COVID-19 pandemic: a cross-sectional analysis. J Gen Intern Med 2023 Dec;38(16):3490-3498. [doi: 10.1007/s11606-023-08353-7] [Medline: 37592119]
17. Chang E, Penfold RB, Berkman ND. Patient characteristics and telemedicine use in the US, 2022. JAMA Netw Open 2024 Mar 4;7(3):e243354. [doi: 10.1001/jamanetworkopen.2024.3354] [Medline: 38517438]
18. Lucas JW, Villarroel MA. Telemedicine use among adults: United States, 2021. NCHS Data Brief 2022 Oct(445):1-8. [Medline: 36255940]

Abbreviations

ECC: emergency care center
ED: emergency department
EMR: electronic medical record
ESI: Emergency Severity Index
PSR: patient service representative
TCI: transitional care intervention
VUC: virtual urgent care

Edited by L Buis; submitted 28.05.25; peer-reviewed by FN Jafri, RK Green; revised version received 09.09.25; accepted 10.10.25; published 29.10.25.

Please cite as:
Lee G, Bruce C, Nisar T, Holderread B, Pletcher S, Nguyen NA
Implementation and Evaluation of a Virtual Transitional Care Intervention Using Automated Text Messaging and Virtual Visits After Emergency Department Discharges: Retrospective Cohort Study
JMIR Mhealth Uhealth 2025;13:e77973
URL: https://mhealth.jmir.org/2025/1/e77973
doi:10.2196/77973

© Grace Lee, Courtenay Bruce, Tariq Nisar, Brendan Holderread, Sarah Pletcher, Ngoc Anh Nguyen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of an mHealth Intervention With Short Text Messages to Promote Treatment Adherence Among HIV-Positive Mexican Adults: Randomized Controlled Trial

Luis Eduardo Del Moral Trinidad[1*], MPH; Jaime Federico Andrade Villanueva[2*], MCR; Pedro Martínez Ayala[2*], IDP; Rodolfo Ismael Cabrera Silva[3*], MSc; Melva Guadalupe Herrera Godina[4*], PhD; Luz Alicia González-Hernández[2*], IMD, PhD

[1]Public Health Sciences, Universidad de Guadalajara, Guadalajara, Mexico

[2]HIV Unit, Hospital Civil de Guadalajara, Hospital 278, Guadalajara, Mexico

[3]Department of Medical Sciences, University of Guadalajara, Guadalajara, Mexico

[4]Biomedical Sciences Research Institute, Universidad de Guadalajara, Guadalajara, Mexico

[*]all authors contributed equally

Corresponding Author:
Luz Alicia González-Hernández, IMD, PhD
HIV Unit, Hospital Civil de Guadalajara, Hospital 278, Guadalajara, Mexico

Abstract

Background: HIV continues to be a public health concern in Mexico and Latin America due to an increase in new infections, despite a decrease being observed globally. Treatment adherence is a pillar for achieving viral suppression. It prevents the spread of the disease at a community level and improves the quality and survival of people living with HIV. Thus, it is important to implement strategies to achieve sustained treatment adherence.

Objective: The objective of this study is to evaluate the effectiveness of a mobile health (mHealth) intervention based on SMS text messages to increase antiretroviral therapy (ART) adherence for HIV-positive adults.

Methods: A randomized controlled trial was performed at the Hospital Civil de Guadalajara – Fray Antonio Alcalde on HIV-positive adults who had initiated ART. The mHealth intervention included the use of SMS text messages as a reminder system for upcoming medical examinations and ART resupply to increase adherence. This intervention was provided to 40 participants for a 6-month period. A control group (n=40) received medical attention by the standard protocol used in the hospital. Intervention effectiveness was assessed by quantifying CD4+ T cells and viral load, as well as a self-report of adherence by the patient.

Results: The intervention group had greater adherence to ART than the control group (96% vs 92%; *P*<.001). In addition, the intervention group had better clinical characteristics, including a lower viral load (141 copies/mL vs 2413 copies/mL; *P*<.001) and a trend toward higher CD4+ T cells counts (399 cells/µL vs 290 cells/µL; *P*=.15).

Conclusions: These results show that an mHealth intervention significantly improves ART adherence. Implementing mHealth programs could enhance the commitment of HIV-positive adults to their treatment.

Trial Registration: ClinicalTrials.gov NCT05187741; https://clinicaltrials.gov/study/NCT05187741

(JMIR Mhealth Uhealth 2025;13:e57540) doi:10.2196/57540

KEYWORDS

HIV; treatment adherence; mobile health; mHealth; mHealth intervention; randomized clinical trial; text messages; eHealth intervention; sexual health; randomized controlled trial; RCT

Introduction

In 2022, 39 million people were living with HIV, of which 37.5 million were over the age of 15 years. Most people living with HIV are female (53%), and 630,000 people died from AIDS-related illness in 2022 [1]. International efforts have been implemented to achieve the 95-95-95 goal by 2030. This goal is that for all people living with HIV, 95% be aware of their HIV status, 95% of all diagnosed people be treated with antiretroviral therapy (ART), and 95% of treated people have an undetectable viral load [2]. HIV continues to be a public health concern in Mexico and Latin America due to an increase in new cases, despite a decrease being observed globally. The high-risk populations in Mexico reflects global trends and include men who have sex with men, injection drug users, sex workers, and transgender individuals. By the end of 2020, there

XSL•FO
RenderX

were 3389 new HIV cases reported. Males accounted for 80.2% of cases and females accounted for 19.8%, with a sex ratio of 4:1 [1-4].

Adherence is not only an important pillar for achieving viral suppression, but it also prevents the spread of the virus at a community level. Moreover, viral suppression is a contributing factor for improving the quality of life for people living with HIV [3]. The World Health Organization defines treatment adherence as "the extent to which a person's behavior, taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider" [4,5]. Treatment adherence is complex and is influenced by the personal perceptions that people living with HIV have toward the disease and the relationship they have with health care providers and institutions [6,7]. Multiple factors can impact adherence to treatment, such as the patient's understanding of the disease, side effects of the treatment, perceived treatment efficacy, trust in health care providers, financial and social access to ART, mental health status, and availability of family or community support [8,9].

Improving treatment adherence is beneficial to the overall health of people living with HIV by reducing disease progression and the development of complications from infection. In addition, treatment adherence is favorable for health systems as it reduces the number of patients with infections, diminishes the need for specialized medical attention, reduces the need for alternative treatments due to treatment failure, and prevents patient disability [10]. Recent research suggests that individuals newly diagnosed with a chronic condition often cultivate their behaviors and perceptions regarding treatment during the initial 6-month period. This time frame is critical for them to establish treatment adherence, which can have far-reaching implications for their long-term health and well-being [11].

Literature has reported a wide range of interventions to improve adherence, including digital interventions (also known as digital health). Digital health is defined as the cost-effective and secure use of information and communication technologies to support health and health-related fields. An integral part of digital health is using wireless technologies like phones, tablets, and smartphones for public health, which is known as mobile health (mHealth) [12]. eHealth and mHealth strategies have been proposed for many aspects of health services, including medication adherence, diagnosis, training, recruitment, follow-up, disease prevention, and guidance about a disease or treatment [13]. In a systematic review analyzing 27 studies regarding the use of mHealth for people living with HIV in low- and middle-income countries, it was shown that 56% of the studies had a significant positive impact on treatment adherence. The most common intervention was the use of reminders through SMS text messages [14].

The duration of interventions can vary depending on the population and objectives of the study, as described in a systematic review by Taylor et al [15] that examined a range of SMS-based interventions across different settings. Additionally, it has been documented that prolonged interventions could potentially lead to participant fatigue and disengagement, which is an important consideration when designing such interventions [16]. One of the main reasons why SMS interventions have been used is due to their extensive coverage in areas where wireless internet is not available. Additionally, SMS text messages can be received by users even if they do not have a data plan. This intervention allows patients to stay connected despite the coverage limitations they may face. Thus, this study aimed to evaluate the efficacy of an mHealth intervention, based on SMS text messages, in increasing ART adherence for adults living with HIV in contrast to standard medical conventions.

Methods

Eligibility Criteria

We performed a randomized clinical trial in which we assessed the efficacy of using SMS text messages to improve treatment adherence. People living with HIV were recruited from the HIV clinic of the tertiary care Hospital Civil de Guadalajara – Fray Antonio Alcalde from February to November 2022. The study included both men and women aged 18 years or older who were diagnosed with HIV and had been on ART for less than 6 months. Additionally, participants needed to have a mobile device capable of receiving SMS text messages and sign the informed consent to participate. Excluded from the study were individuals currently enrolled in another clinical trial, pregnant women, minors, and individuals with a mobile phone capable of receiving SMS messages but lacking coverage or service in their residential area.

Sample Size Determination

The sample size was determined using the OpenEpi software version 3.01, using an anticipated mean difference of 3 percentage points in the rates of adherence between the study groups, based on existing literature [17]. The study was designed with a 95% CI, 80% statistical power, and a 1:1 sample size ratio between the groups being compared. A power analysis was done and determined that a sample size of 29 participants per group was sufficient to detect a mean difference of 3 points in adherence with 80% power and a significance level of .05. Increasing the sample size to 40 participants per group would raise the power to 91%, providing a greater likelihood of detecting a true effect if it exists.

Randomization

To begin randomization, the protocol nurse generated a table of random numbers in Excel version 16.92 (Microsoft), which was used to assign participants to their respective groups based on whether the randomly generated number was even or odd, at a 1:1 ratio. If a participant who was invited declined the intervention, the allocation sequence proceeded with the next randomly assigned number. This approach ensured that each patient had an equal chance of being allocated to either group, enhancing the randomization process and minimizing selection biases.

Variables of the Study

Sociodemographic variables were obtained at the time of recruitment using a form that was filled out with the aid of the research team. Viral load and number of CD4+ T cells were

XSL•FO
RenderX

measured at the time of enrollment and after the 6-month follow-up. Self-adherence was measured at the 6-month follow-up. To measure adherence, the 4-day adherence index was used, which was calculated using the following formula [18]:

$$\text{4-day adherence index} = 1 - (\text{forgotten pills}/\text{prescribed pills}) \times 100$$

For the measurement of variables, participants were considered employed if they reported engaging in paid work, whether formal or informal. Tobacco, alcohol, and illicit drug consumption were determined if participants reported consumption more than twice a week in the last month. The duration of time living with HIV and the duration on ART were calculated using the study's completion date.

mHealth Strategy

The intervention consisted of the following procedures. Participants received 2 messages per week on Mondays and Thursdays in the morning during the first 3 months, then 1 message per week for the last 3 months. These messages were prevalidated by a research group and categorized into four types: motivational messages for self-care, reminders for ART collection, reminders for medical appointments, and laboratory reminders. The content encouraged patients to maintain adherence to their medication, provided reminders about upcoming medical visits, and reinforced the importance of regular lab tests, such as CD4 and viral load monitoring. To avoid repetition, the messages were rotated throughout the intervention. The full set of validated messages and their detailed content can be found in the corresponding publication [19].

Additionally, participants received reminder messages for appointments or prescription refills 24 hours before the scheduled date during the 6-month intervention period. Moreover, participants had the option to communicate with the researchers to address any queries, which were resolved through a consensus among the research team with support from relevant departments. To ensure ongoing communication, the telephone line accepted collect messages and collect calls, enabling participants to stay connected throughout the study. Both groups received the standard medical treatment provided by the hospital, which consisted of medical appointments tailored to each patient's needs, but the appointments were scheduled no more than once every 3 months. The treatment also included free ART with a maximum supply of 90 pills. Additionally, laboratory studies and complementary tests were performed in accordance with national treatment guidelines.

Statistical Analysis

Proportions were analyzed by χ^2 tests and a Student *t* test was used for variables. Relative risk based on treatment adherence (>95% or >90%) was measured and the Fisher exact test was used for statistical evaluation. A multivariate analysis was done to calculate relative risks of variables associated with adherence. All statistical analyses were performed on SPSS version 25 (IBM Corp) and an α level of .05 was used to determine significance.

Ethical Considerations

The study protocol was approved by the Ethics Committee in Research of the Hospital Civil de Guadalajara – Fray Antonio Alcalde (registry number 095/18) and was registered as a clinical trial (NCT05187741). In line with the Council for International Organizations of Medical Sciences guidelines, all participants provided written inform consent, which was supplemented by a family witnesses in vulnerable cases to ensure fully informed decision-making. The consent form included all relevant information from the protocol and was supplemented with direct clarification of any concerns to ensure patients fully understood. Participants were informed that they could withdraw from the study at any time without any impact on their access to medical care or other services. Privacy was safeguarded by anonymizing participant data with numerical codes, with only the principal research team having access. In the event of unintentional disclosure, a potential risk identified in our study, psychological support channels were made available. While no monetary compensation was provided, participants benefited from direct access to health care professionals via text or call support, and the treatment also included free ART with a maximum supply of 90 pills.

Results

From 217 potential candidates, 80 patients fulfilled the inclusion criteria and agreed to participate in the study. The 80 participants were randomly assigned to the intervention or control group (40 in each). In the intervention group, 4 participants were lost to follow-up: 3 due to clinic migrations and 1 for unspecified reasons. In the control group, 2 participants were lost to follow-up due to clinic migrations, and an additional 2 were lost for other reasons. All participants analyzed belonged to their original assigned group (Figure 1). Intervention and control groups had a median age of 31 (IQR 27.8 - 36.0) years and 37.5 (IQR 30.0 - 49.5) years, respectively. Both groups were mostly men (intervention: 38/40, 96%; control: 37/40, 93%). At the time of enrollment, the intervention group had more CD4+ T cells (227 cells/µL vs 169 cells/µL) and a higher viral load (65,050 copies/mL vs 57,800 copies/mL) compared to the control. In addition, the intervention group had a greater number of days living with HIV and more days on ART when compared to the control group (Table 1). Tobacco consumption was greater in the control group (15/40, 38% vs 23/40, 58%). Most of the participants were single, employed, and reported alcohol consumption (Table 1).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the intervention.

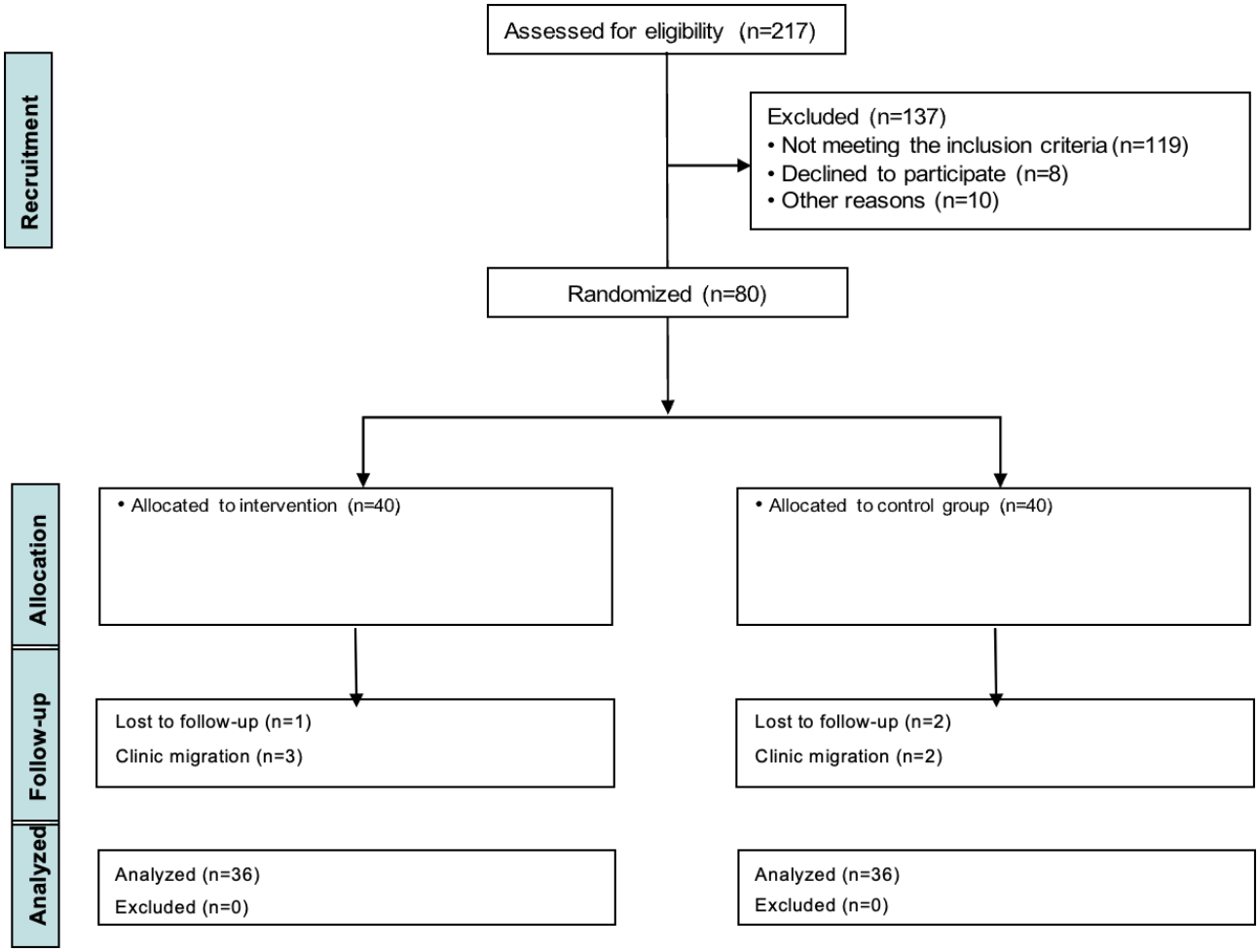


Table . Participants' clinical and sociodemographic characteristics.

	Intervention group (n=40)	Control group (n=40)	*P* value
Clinical variables			
Age (years), median (IQR)	31.0 (27.8 - 36.0)	37.5 (30.0 - 49.5)	.001
Sex, n (%)			—[a]
Male	38 (96)	37 (92)	
Female	2 (4)	3 (8)	
Days after HIV diagnosis, median (IQR)	283.5 (58.3 - 519.3)	303 (101.3 - 1128)	.50
Days on ART[b], median (IQR)	122 (15.75 - 192.3)	51 (19.75 - 171.0)	.73
CD4+ T cell count (cells/µL), median (IQR)	227 (88 - 435)	169 (25 - 291)	.09
Viral load (copies/mL), median (IQR)	65,050 (17,282-343,750)	57,800 (21,900-168,750)	.46
Sociodemographic variables, n (%)			
Employed	25 (62)	24 (60)	.60
Tobacco consumption	15 (38)	23 (58)	.10
Alcohol consumption	25 (62)	30 (75)	.13
Illegal drug consumption	15 (38)	14 (35)	.81

[a]Analysis was not performed.

[b]ART: antiretroviral therapy.

XSL•FO
RenderX

After 6 months, the viral load and CD4+ T cell count were reassessed. The intervention group trended toward higher CD4+ T cell counts compared to the control group (399 cells/µL vs 290 cells/µL; *P*=.15). In addition, the intervention group had a significantly reduced viral load compared to the control group (141 copies/mL vs 2413 copies/mL; *P*<.001). Regarding treatment adherence, the intervention group self-reported a higher level of adherence than the control group (96% vs 92%; *P*<.001; Table 2).

Table . Clinical characteristics and adherence at the intervention period endpoint.

	Intervention group (n=36)	Control group (n=36)	*P* value
CD4+ T cell count (cells/µL), median (IQR)	399 (238.8 - 634.0)	290 (195.3 - 557.5)	.15
Viral load (copies/mL), mean (SD)	141 (60)	2413 (1606)	<.001
Adherence rate (%), median (IQR)	96 (94.3 - 98.0)	92 (88.0 - 95.0)	<.001

We further analyzed adherence and separated the members of each group based on whether they achieved adherence above 90% or 95%. In both cases, the intervention group had a significantly greater chance of reaching the optimal levels of adherence and were approximately 3 times more likely to reach 95% adherence than the control (Table 3).

Table . Comparison of treatment adherence between the intervention and control groups.

Group	Adherence, n (%)		Relative risk (95% CI)	*P* value
	Yes	No		
Adherence >90%				
Intervention	36 (100)	0 (0)	1.39 (1.31 - 1.70)	<.001
Control	26 (72)	10 (28)	1.0 (reference)	—
Adherence >95%				
Intervention	26 (72)	10 (28)	2.6 (1.48 - 4.57)	<.001
Control	10 (28)	26 (72)	1.0 (reference)	—

To identify if other variables impacted the probability of reaching 95% adherence, we performed a multivariate analysis. Aside from the SMS intervention, no other variables significantly contributed to reaching 95% adherence (Table 4).

Table . Multivariate analysis of variables associated with 95% treatment adherence.

Variable	Relative risk (95% CI)	*P* value
Age >45 years	0.73 (0.15 - 3.60)	.69
Employed	0.39 (0.10 - 1.41)	.15
Tobacco consumption	1.56 (0.436 - 5.60)	.49
Alcohol consumption	2.31 (0.59 - 8.98)	.22
Use of illegal drugs	0.81 (0.25 - 2.62)	.73
Detectable viral load	4.26 (0.67 - 26.83)	.12
CD4+ T cell count >200 cells/µL	4.09 (0.38 - 43.05)	.24
Study intervention	11.66 (2.88 - 47.05)	.001

Discussion

Principal Findings

This study aimed to evaluate the effectiveness of an mHealth intervention on adherence for adults starting ART. As previously mentioned, ART adherence is important for people living with HIV, allowing them to reach viral suppression and increasing their overall quality of life. This highlights the importance of exploring interventions to support patients recently diagnosed with HIV to reach optimal adherence. This study enrolled patients with an age ranging from 30 to 40, which corresponds with the national and worldwide age with the highest HIV prevalence. Also, our participants were mostly men, which aligns with Mexico's national trend, as men who have sex with men and transgender people are the main groups affected by HIV in Mexico [20].

Both the intervention and control groups shared previously reported factors that can impact adherence like alcohol consumption and being single [5,6,10,21]. Although further analysis of the factors associated with adherence was beyond the scope of this study, our descriptive results provide a preliminary view for further investigations aiming to study the

XSL•FO
RenderX

importance of these factors on a patient's adherence to treatment. Even though the participants were randomly assigned to each group, our intervention group was younger than our control group. We believed that this was related to the recruitment process, as older patients were more likely to reject the invitation of the study due to the concern of using digital conventions over in-person conventions. Giebel et al [22] mention that interventions must be adapted to different age groups due to the diversity of digital health users. These findings need to be explored and contemplated in further studies.

Regarding our clinical findings, the control group had a lower CD4+ T cell count (168 cells/µL), suggesting a delay in HIV detection. The fact that one of our study groups had less than 200 cells/µL highlights the relevance of finding methods to improve early detection in high-risk populations and strategies to prevent low adherence to ART or losses to follow-up [23]. One of the potential implications of these findings is the application of mHealth interventions to provide a faster linkage-to-care after HIV diagnosis, as a previous study has shown that these interventions can reduce the gap between health care clients and institutions to increase retention in care and impact mortality [24].

In this study, we show that mHealth strategies improve adherence to treatment by 4% (P<.001) and increase the probability of reaching a 95% of treatment adherence by 3 times. However, we wanted to evaluate the impact of age and other variables on the probability of reaching 95% treatment adherence. For this, we performed a multivariate analysis, in which the mHealth intervention statistically correlated with ART adherence regardless of other independent variables like age, marital status, and employment. These results suggest that mHealth strategies can be beneficial in multiple scenarios.

As previously demonstrated, simply sending SMS text messages had minimal to no impact on enhancing treatment adherence [25,26]. Therefore, we implemented an mHealth strategy based on the use of interactive SMS text messages, in which the patients had the opportunity to interact by text or phone call with a health professional. Interactive SMS text messages are effective and participants are more likely to engage actively and adhere to their treatment. [27-29]. Another key aspect of the intervention was the frequency of the text messages. The goal was to help establish a habit in the patients while still providing access to assistance if needed. Also, it has been reported that a high frequency of messages over an extended period can become repetitive and lose their beneficial impact [30-32].

Although other studies have used similar strategies, this is the first study in the Mexican population to use an mHealth approach that decreased message frequency to enhance adherence. In addition, we implemented messages that were validated by a Mexican group of health professionals, researchers, and people living with HIV on ART [19]. Few studies looking into the efficacy of mHealth SMS interventions use validated messages for their population [33-35]. Like other studies, we identified an improvement in treatment adherence by using SMS text messages [14,26,36,37]. However, it is important to consider that adherence evaluation was mainly assessed by a self-report that was provided by the patient, which had an intrinsic bias. As a method to address this bias, we measured the HIV viral load along with adherence reported by the 4-day adherence index, which has been reported to have an adequate sensitivity and specificity in Mexican populations [18]. This allowed us to provide a more accurate estimation of adherence levels, as self-reported measures are known to overestimate true adherence levels.

Although we identified an increased adherence using an mHealth strategy, further studies on a larger population are required to confirm our results. It is also important to recognize the variability among different populations and regions [38]. This study was performed in the State of Jalisco, and we found a 93% treatment adherence rate in our control group. This result was similar to another study performed in the state of Mexico that had a 93% adherence rate [39] but double that of a study done in the state of Tamaulipas that had a 46% adherence rate [40]. These state-specific differences highlight how treatment adherence is contextual and is determined largely by the conditions surrounding the patients [41].

This study had some limitations. The evaluation period was relatively brief, lasting only 6 months, which raised uncertainty about the long-term persistence of the intervention's effect. Additionally, the reliance on self-reported adherence data, without the use of more objective measures such as pill counts at each appointment, introduced the potential for bias. However, the investigators ensured that all participants underwent viral load testing after the study, providing an objective assessment of treatment outcomes.

Conclusions

In summary, this study demonstrates that integrating mHealth strategies as a supplementary approach to existing interventions can enhance treatment adherence among people living with HIV. Additionally, our findings highlight significant results that may inform the implementation of larger and more intricate interventions aimed at bridging the communication gap between health care clients and medical service providers.

This kind of mHealth-based intervention could serve as a precursor for larger-scale interventions aimed at increasing patient adherence and retention for chronic illnesses, which is a significant public health concern that we must address. By leveraging digital health technologies, we can develop more comprehensive strategies to support patients in managing their conditions and improve overall treatment outcomes. Such approaches have the potential to enhance communication between health care providers and clients, thereby promoting better treatment adherence and increasing retention rates across diverse populations with chronic diseases.

XSL•FO
RenderX

Acknowledgments

We would like to thank our institution for allowing us to carry out the protocol and giving us the support for its implementation. Most importantly, we would like to thank the patients whose participation was vital for this research.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Checklist 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (version 1.6.1).
[PDF File, 1235 KB - mhealth_v13i1e57540_app1.pdf]

References

1. UNAIDS. National commitments and policy instrument 2022. : Joint United Nations Programme on HIV/AIDS; 2022 URL: https://www.aidsdatahub.org/resource/danger-unaids-global-aids-update-2022 [accessed 2024-12-23]
2. Global commitments, local action. UNAIDS. 2021. URL: https://www.unaids.org/en/resources/documents/2021/global-commitments-local-action [accessed 2024-12-17]
3. Espinosa A, Sánchez M, Lamadrid M. An updated understanding of adherence to antiretroviral treatment. Rev Cubana Enferm 2018;34(1):220-232 [FREE Full text]
4. Adherence to Longterm Therapies: Evidence for Action: World Health Organization; 2003.
5. Nachega JB, Marconi VC, van Zyl GU, et al. HIV treatment adherence, drug resistance, virologic failure: evolving concepts. Infect Disord Drug Targets 2011 Apr;11(2):167-174. [doi: 10.2174/187152611795589663] [Medline: 21406048]
6. Peñarrieta MI, Kendall T, Martinez N, et al. Adherencia al tratamiento antirretroviral en personas con VIH en Tamaulipas, México. Rev Peru Med Exp Salud Publica 2009 [FREE Full text]
7. Ferraz Pamplona L, Romano Rennó Costa G, e Silva BM, De Araújo Pinto G, Souza de Faria R, Oliveira Cortez PJ. Factors associated to the control of viral load in HIV positive patients. arch med 2020;21(1). [doi: 10.30554/archmed.21.1.3823.2021]
8. Miller TA. Health literacy and adherence to medical treatment in chronic and acute illness: a meta-analysis. Pat Educ Couns 2016 Jul;99(7):1079-1086. [doi: 10.1016/j.pec.2016.01.020] [Medline: 26899632]
9. Sánchez-Rivero I, Madoz-Gúrpide A, Parro-Torres C, Hernández-Huerta D, Ochoa Mangado E. Influence of substance use and cognitive impairment on adherence to antiretroviral therapy in HIV+ patients. Adicciones 2020 Jan 1;32(1):7-18. [doi: 10.20882/adicciones.1025] [Medline: 30627722]
10. Schaecher KL. The importance of treatment adherence in HIV. Am J Manag Care 2013 Sep;19(12 Suppl):s231-s237. [Medline: 24495293]
11. Thomson P, Rushworth GF, Andreis F, Angus NJ, Mohan AR, Leslie SJ. Longitudinal study of the relationship between patients' medication adherence and quality of life outcomes and illness perceptions and beliefs about cardiac rehabilitation. BMC Cardiovasc Disord 2020 Feb 11;20(1):71. [doi: 10.1186/s12872-020-01378-4] [Medline: 32046646]
12. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 5;16 Suppl 5(Suppl 5):290. [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
13. Bervell B, Al-Samarraie H. A comparative review of mobile health and electronic health utilization in sub-Saharan African countries. Soc Sci Med 2019 Jul;232:1-16. [doi: 10.1016/j.socscimed.2019.04.024] [Medline: 31035241]
14. Demena BA, Artavia-Mora L, Ouedraogo D, Thiombiano BA, Wagner N. A systematic review of mobile phone interventions (SMS/IVR/calls) to improve adherence and retention to antiretroviral treatment in low-and middle-income countries. AIDS Patient Care STDS 2020 Feb;34(2):59-71. [doi: 10.1089/apc.2019.0181] [Medline: 32049555]
15. Taylor D, Lunny C, Lolić P, et al. Effectiveness of text messaging interventions on prevention, detection, treatment, and knowledge outcomes for sexually transmitted infections (STIs)/HIV: a systematic review and meta-analysis. Syst Rev 2019 Jan 8;8(1):12. [doi: 10.1186/s13643-018-0921-4] [Medline: 30621784]
16. Mulawa MI, LeGrand S, Hightow-Weidman LB. eHealth to enhance treatment adherence among youth living with HIV. Curr HIV/AIDS Rep 2018 Aug;15(4):336-349. [doi: 10.1007/s11904-018-0407-y] [Medline: 29959649]
17. Côté J, Godin G, Ramirez-Garcia P, et al. Virtual intervention to support self-management of antiretroviral therapy among people living with HIV. J Med Internet Res 2015 Jan 6;17(1):e6. [doi: 10.2196/jmir.3264] [Medline: 25563775]
18. Balandrán DA, Gutiérrez JP, Romero M. Evaluación de la adherencia antirretroviral en México: adherencia de cuatro días vs. índice de adherencia. Rev Invest Clin 2013;65(5):384-391. [Medline: 24687337]

XSL•FO
RenderX

19. Pérez-Sánchez IN, Candela Iglesias M, Rodriguez-Estrada E, Reyes-Terán G, Caballero-Suárez NP. Design, validation and testing of short text messages for an HIV mobile-health intervention to improve antiretroviral treatment adherence in Mexico. AIDS Care 2018;30(sup1):37-43. [doi: 10.1080/09540121.2018.1524115] [Medline: 30235937]
20. Centro Nacional para la Prevención y Control del VIH y el sida. Boletín Día mundial del sida 2022. Gobierno de México. 2022. URL: https://www.gob.mx/cms/uploads/attachment/file/778212/BOLETIN_DAI_DIA_MUNDIAL_DEL_SIDA_.pdf [accessed 2024-12-17]
21. Soares RDC, Brito AD, Lima K, Lapa TM. Adherence to antiretroviral therapy among people living with HIV/AIDS in northeastern Brazil: a cross-sectional study. Sao Paulo Med J 2019;137(6):479-485. [doi: 10.1590/1516-3180.2019.0212170919] [Medline: 32159632]
22. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808. [doi: 10.2196/43808] [Medline: 37171838]
23. Johnson DF, Sorvillo FJ, Wohl AR, et al. Frequent failed early HIV detection in a high prevalence area: implications for prevention. AIDS Patient Care STDS 2003 Jun;17(6):277-282. [doi: 10.1089/108729103322108157] [Medline: 12880491]
24. Hodges J, Zhdanova S, Koshkina O, et al. Implementation of a mobile health strategy to improve linkage to and engagement with HIV care for people living with HIV, tuberculosis, and substance use in Irkutsk, Siberia. AIDS Patient Care STDS 2021 Mar;35(3):84-91. [doi: 10.1089/apc.2020.0233] [Medline: 33538649]
25. Linnemayr S, Huang H, Luoto J, et al. Text messaging for improving antiretroviral therapy adherence: no effects after 1 year in a randomized controlled trial among adolescents and young adults. Am J Public Health 2017 Dec;107(12):1944-1950. [doi: 10.2105/AJPH.2017.304089] [Medline: 29048966]
26. Georgette N, Siedner MJ, Petty CR, Zanoni BC, Carpenter S, Haberer JE. Impact of a clinical program using weekly short message service (SMS) on antiretroviral therapy adherence support in South Africa: a retrospective cohort study. BMC Med Inform Decis Mak 2017 Feb 20;17(1):18. [doi: 10.1186/s12911-017-0413-9] [Medline: 28219372]
27. King E, Kinvig K, Steif J, et al. Mobile text messaging to improve medication adherence and viral load in a vulnerable Canadian population living with human immunodeficiency virus: a repeated measures study. J Med Internet Res 2017 Jun 1;19(6):e190. [doi: 10.2196/jmir.6631] [Medline: 28572079]
28. Jopling R, Nyamayaro P, Andersen LS, Kagee A, Haberer JE, Abas MA. A cascade of interventions to promote adherence to antiretroviral therapy in African countries. Curr HIV/AIDS Rep 2020 Oct;17(5):529-546. [doi: 10.1007/s11904-020-00511-4] [Medline: 32776179]
29. Hendricks L, Eshun-Wilson I, Rohwer A. A mega-aggregation framework synthesis of the barriers and facilitators to linkage, adherence to ART and retention in care among people living with HIV. Syst Rev 2021 Feb 11;10(1):54. [doi: 10.1186/s13643-021-01582-z] [Medline: 33568216]
30. Bartlett YK, Farmer A, Newhouse N, Miles L, Kenning C, French DP. Effects of using a text message intervention on psychological constructs and the association between changes to psychological constructs and medication adherence in people with type 2 diabetes: results from a randomized controlled feasibility study. JMIR Form Res 2022 Apr 29;6(4):e30058. [doi: 10.2196/30058] [Medline: 35486430]
31. Lin B, Liu J, He W, Pan H, Ma Y, Zhong X. Effect of a reminder system on pre-exposure prophylaxis adherence in men who have sex with men: prospective cohort study based on WeChat intervention. J Med Internet Res 2022 Aug 11;24(8):e37936. [doi: 10.2196/37936] [Medline: 35969436]
32. Griffee K, Martin R, Chory A, Vreeman R. A systematic review of digital interventions to improve ART adherence among youth living with HIV in sub-Saharan Africa. AIDS Res Treat 2022;2022:9886306. [doi: 10.1155/2022/9886306] [Medline: 36199816]
33. Kibu OD, Siysi VV, Albert Legrand SE, Asangbeng Tanue E, Nsagha DS. Treatment adherence among HIV and TB patients using single and double way mobile phone text messages: a randomized controlled trial. J Trop Med 2022;2022:2980141. [doi: 10.1155/2022/2980141] [Medline: 35996467]
34. Kawichai S, Songtaweesin WN, Wongharn P, et al. A mobile phone app to support adherence to daily HIV pre-exposure prophylaxis engagement among young men who have sex with men and transgender women aged 15 to 19 years in Thailand: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Apr 21;10(4):e25561. [doi: 10.2196/25561] [Medline: 35451976]
35. Garofalo R, Kuhns LM, Hotton A, Johnson A, Muldoon A, Rice D. A randomized controlled trial of personalized text message reminders to promote medication adherence among HIV-positive adolescents and young adults. AIDS Behav 2016 May;20(5):1049-1059. [doi: 10.1007/s10461-015-1192-x] [Medline: 26362167]
36. Haberer JE, Musiimenta A, Atukunda EC, et al. Short message service (SMS) reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May 15;30(8):1295-1300. [doi: 10.1097/QAD.0000000000001021] [Medline: 26760452]
37. Ingersoll KS, Dillingham RA, Hettema JE, et al. Pilot RCT of bidirectional text messaging for ART adherence among nonurban substance users with HIV. Health Psychol 2015 Dec;34S:1305-1315. [doi: 10.1037/hea0000295] [Medline: 26651472]
38. Camacho G, Kalichman S, Katner H. Anticipated HIV-related stigma and HIV treatment adherence: the indirect effect of medication concerns. AIDS Behav 2020 Jan;24(1):185-191. [doi: 10.1007/s10461-019-02644-z] [Medline: 31432297]

XSL•FO
RenderX

39. Torres Torija CS, Cruz GV, Noguez Noguez S, Jaimes Hernándezndez AC, Martín Chapa MI. Indicadores de adherencia al tratamiento antiretroviral de una población mexicana VIH+. : Alternativas en psicología; 2017 URL: http://www.alternativas.me/attachments/article/126/Indicadores de adherencia al tratamiento antiretroviral de una población mexicana VIH.pdf [accessed 2024-12-18]
40. Pérez Bastán JE. Adherence to antiretroviral treatment in people with HIV/AIDS in primary health care. Rev Haban Cienc Med 2020 Oct;19(5):1-15 [FREE Full text]
41. Giuliani M, Donà MG, La Malfa A, et al. Ensuring retention in care for people living with HIV during the COVID-19 pandemic in Rome, Italy. Sex Transm Infect 2021 Jun;97(4):317. [doi: 10.1136/sextrans-2020-054650] [Medline: 33046581]

Abbreviations

ART: antiretroviral therapy
mHealth: mobile health

Edited by L Buis; submitted 19.02.24; peer-reviewed by C Chiaranai, J Mitchell; revised version received 23.09.24; accepted 02.12.24; published 28.01.25.

Please cite as:
Del Moral Trinidad LE, Andrade Villanueva JF, Martínez Ayala P, Cabrera Silva RI, Herrera Godina MG, González-Hernández LA Effectiveness of an mHealth Intervention With Short Text Messages to Promote Treatment Adherence Among HIV-Positive Mexican Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e57540
URL: https://mhealth.jmir.org/2025/1/e57540
doi:10.2196/57540

© Luis Eduardo Del Moral Trinidad, Jaime Federico Andrade Villanueva, Pedro Martínez Ayala, Rodolfo Ismael Cabrera Silva, Melva Guadalupe Herrera Godina, Luz Alicia González-Hernández. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Factors Influencing Engagement in a Digital Substance Use Prevention Program: Qualitative Study Using the Capability, Opportunity, Motivation Model and Theoretical Domains Framework

Nikolai Kiselev, PhD; Rebecca Schaffner, MSc; Zsofia Csepregi, BSc; Andreas Wenger, MSc; Raquel Paz Castro, PhD; Severin Haug, PhD

Swiss Research Institute for Public Health and Addiction (ISGF), University of Zurich, Konradstrasse 32, Zurich, Switzerland

Corresponding Author:
Severin Haug, PhD
Swiss Research Institute for Public Health and Addiction (ISGF), University of Zurich, Konradstrasse 32, Zurich, Switzerland

Abstract

Background: Digital interventions are promising for the prevention of substance use in young people. However, engagement with these interventions is often insufficient, and their full potential remains unrealized. Given the established link between engagement in digital interventions and their effectiveness, understanding user factors that influence involvement in eHealth and mobile health (mHealth) interventions is essential.

Objective: This study aimed to identify factors influencing user engagement in a mobile phone–based life skills training program for substance use prevention among adolescents and to collect suggestions for program optimization.

Methods: A qualitative study was conducted with 171 participants of the mHealth prevention SmartCoach (Pathmate Technologies) program. The program provided individualized text messages to foster life skills for 4 months and proved to be effective in preventing the onset of cigarette and cannabis use. Semistructured phone interviews were conducted with program participants to explore factors associated with program engagement and to gather suggestions for program optimization. Interviews were recorded, transcribed, and analyzed using thematic analysis with both inductive and deductive coding. The capability, opportunity, and motivation model of behavior change (COM-B) model and Theoretical Domains Framework (TDF) were used to assess behavioral influences.

Results: Key factors positively influencing program engagement included the timing of text messages, social influences and support, engaging and helpful content, and rewards (points and prizes). Conversely, barriers to engagement were identified as forgetfulness, short response time limits, limited time resources, lack of interest, concerns related to personal disclosure, and difficulty identifying with the challenge task type (posting). Suggestions for optimization included implementing reminders, providing better guidance for using tips, allowing personalization of message timing and content, extending time limits for tasks, and reducing the concerns related to personal disclosure.

Conclusions: The study confirms the critical role of timing, content relevance, and social support in enhancing engagement with digital interventions. Specific recommendations for optimizing the SmartCoach program were derived, highlighting the importance of reminders, personalization, and addressing concerns related to personal disclosure.

(JMIR Mhealth Uhealth 2025;13:e64861) doi:10.2196/64861

KEYWORDS

SmartCoach; program engagement; digital intervention; substance use prevention; COM-B model; theoretical domains framework; engagement optimization; adolescent; teenager; substance use; drug use; engagement; qualitative study; training program; prevention; mobile health; mHealth; text; cigarette; cannabis; social support; social influences

Introduction

Background

The consumption of tobacco, alcohol, and cannabis poses serious health risks and can lead to addiction. According to the World Health Organization (WHO), millions of individuals die each year from the consequences of consuming these substances [1]. Despite the known risks, usage rates remain high. Across Europe, there are significant variations in consumption, with almost 10% of the population consuming alcohol daily, over a quarter being smokers, and around 1.3% using cannabis [2-5].

The prevalence of substance use increases sharply between the ages of 11 and 15 years [6]. In Switzerland, the lifetime

XSL•FO
RenderX

prevalence of alcohol use has been documented to increase from 22% among 11-year-old boys to 70% among 15-year-old boys and from 11% in 11-year-old girls to 69% in 15-year-old girls [7]. The proportion of students who reported having smoked cigarettes at least once in their life similarly increased from 6% to 35% in 11-year-old boys versus 15-year-old boys and from 2% to 30% in 11-year-old girls versus 15-year-old girls, respectively. On average, in the Canton of Zurich, alcohol consumption begins at around 15 years of age, tobacco at 14.6 years, and cannabis at 15.6 years [8]. Therefore, early and midadolescence appear to be a critical period for introducing substance use prevention interventions. School settings are particularly suitable, as they facilitate delivery and access, especially within compulsory education systems.

Life skills–based intervention programs to prevent substance use delivered in the school have been effective at preventing the onset of specific substances [9-11] and at decreasing problematic substance use [12]. However, their implementation and dissemination in schools present serious challenges [13], as teachers and other professionals need the time, motivation, knowledge, and skills to deliver the program [14].

Digital interventions have great potential to overcome these obstacles. Excluding development costs, mobile health (mHealth) interventions have an extensive reach at low cost and offer the ability to deliver uniquely personalized content automatically, which can be accessed at any time and anywhere [15].

However, a major challenge in delivering digital preventive measures is the lack of control over adolescents' engagement with the intervention. Several reviews on digital interventions to promote mental health [16,17] or to prevent substance use [15] in young people point to relatively low levels of user engagement. Regular use of mHealth programs is often the exception rather than the rule, which is subsumed under the term "engagement crisis" [18].

There is a wide body of evidence showing that higher engagement (ie, a higher number of user interactions within a program) in a prevention program leads to better outcomes for participants [19-22]. The results of the mobile phone–based SmartCoach (Pathmate Technologies) program to prevent substance use in adolescents also revealed this relationship. For instance, participants with a higher engagement rate in the intervention showed a lower likelihood of tobacco consumption in the follow-up survey. Furthermore, participants who actively participated in almost all program components reported the highest increase in well-being [23]. In addition to user engagement during the program, several studies emphasize the positive effects of active youth involvement in the development of prevention messages and program content [24,25]. These findings support the idea that engagement should not only be seen as a behavioral outcome but also as a process of cocreation, which can contribute to stronger identification with the program and its goals.

The challenge to increase participant engagement in digital preventive intervention programs is increasingly discussed from both scientific and practical perspectives; however, there is a lack of generalizable approaches that show how engagement in existing programs can be systematically improved [18].

Promising starting points for the identification of factors related to program engagement and behavior change are provided by the capability, opportunity, and motivation model of behavior change (COM-B) model [23,26] and the Theoretical Domains Framework (TDF) [27].

According to the COM-B model, behavior (eg, program engagement) arises from the interaction between the individual's psychological and physical capacity to engage in the activity (eg, literacy skills), the physical and social opportunity (eg, peer support), and automatic and reflective motivation (eg, rewards). The model offers a high-level yet structured perspective to identify which components may need to be addressed to achieve behavior change. It is particularly suitable for designing and analyzing health interventions, as it helps to distinguish whether a lack of engagement, for instance, stems from skill deficits (capability), contextual barriers (opportunity), or limited intention or drive (motivation).

The TDF is a synthesis of 33 theories and 128 psychological constructs [27]. It includes 14 domains (eg, knowledge and environmental resources) that can be mapped under the 3 components of the COM-B model [27,28]. The framework was developed to support the systematic identification of behavioral influences and to guide the development of theory-informed interventions. In qualitative research, the TDF is frequently used to structure deductive coding, as it allows researchers to break down complex behavioral phenomena into well-defined, theory-based domains.

The combination of these 2 models has been widely applied in research. For example, Szinay et al [29] used the COM-B and TDF models to investigate experiences with and reasons for using a broad range of health and well-being apps through interviews. Rosário et al [30] used the COM-B and TDF framework to identify factors influencing the implementation of screening and brief interventions for alcohol consumption in primary care.

Research Aim

This study aimed to identify factors derived from the COM-B and TDF framework that might increase engagement in the SmartCoach program, an mHealth intervention to prevent substance use by fostering life skills in adolescents. Specifically, this qualitative study aimed to address the following research questions: (1) What factors positively or negatively influence engagement with the SmartCoach program? and (2) What suggestions do participants offer to optimize the program?

Methods

Overview

A qualitative study was conducted based on free-listing interviews. In the course of the first research question, an exploratory and explanatory approach was pursued, following the framework proposed by Guest et al [31]. The second research question was exploratory. Thematic analysis was used to analyze the interviews, using NVivo 14 software (Lumivero). To assess

XSL•FO
RenderX

the generalizability of the results, a nonresponder analysis was also conducted. This paper was prepared following the Journal Article Reporting Standards for Qualitative Research (JARS), as proposed by the American Psychological Association [32], to ensure transparent and rigorous reporting.

Intervention Program and Setting

The fully automated, internet- and SMS text messaging–based intervention program SmartCoach is based on social cognitive theory and addresses self-management skills, social skills, and substance use resistance skills [33,34]. Participants receive 1-4 SMS text messages per week during the 4-month intervention period. Active program engagement was encouraged through interactive features such as quiz questions, messages, and picture contests, and the integration of a friendly competition with prizes, where users collect credits with each interaction. The efficacy of the program was tested in a cluster randomized controlled trial with 1759 students from 89 Swiss secondary and upper secondary school classes [34,35]. Of these, 1473 (83.7%) students participated in both the program and the respective study; their mean age was 15.4 years. On average, program participants responded to about half of the prompted activities, with adolescents who did not drink in a problematic manner and those with higher educational levels showing the highest usage rates [23].

Longer-term results at the 18-month follow-up showed a reduction in smoking prevalence in the intervention group compared to controls. However, no significant effect was observed for at-risk drinking. No significant moderators of the primary outcomes were observed; that is, the program's effectiveness was independent of age, sex, migration background, and school level.

While participant access and engagement in the previous efficacy study [34,35] may have been influenced by the assessment-only control condition, this study was conducted under more naturalistic conditions, without a control group, and focused on identifying factors associated with program uptake and engagement [36].

Sample

For the abovementioned study focusing on program uptake and engagement, we recruited a total of 476 students from 28 secondary and upper secondary school classes in the Swiss cantons of Zurich and Aargau between September and December 2022. Recruitment was conducted within a school workshop about the topic of stress, with the option for further participation in the SmartCoach program offered at the end of the workshop. Of these, 315 (66.2%) students were willing to participate in the SmartCoach program and received weekly text messages for 4 months [36]. This subsample constitutes the population for the results presented in this qualitative study.

Interviewers

All interviews were conducted by 2 female psychology master's students from the University of Zurich (in their final year). The interviewers had been involved in interview preparation since the recruitment phase for the intervention. Besides previous interviewing experience in other research projects, they were additionally trained in practicing free-listing interviews with this study's target population by the project's senior researcher. In preparation for the data collection, 2 half-day workshops were held with the interviewers, during which the methodological approach and procedures were discussed in detail. In addition, the interviewers conducted practice interviews both within the research team and with 3 adolescents, each of whom was not part of the SmartCoach program, to refine their interviewing skills and ensure clarity and appropriateness of the questions.

Interview Guide

To collect the data, an interview guide was developed in collaboration with researchers from the SmartCoach project. The questions were free-listing, allowing respondents to provide open-ended answers without being constrained by the question [37]. The interview guide comprised 22 questions, mostly about general program participation, including reasons for high and low involvement (first part) and optimization suggestions based on experiential feedback (second part). The last part consisted of a think-aloud task, during which personal stressors and possible coping strategies were queried (for possible future development of the program). This part of the interview guide is neither included in the present work nor published elsewhere. It is important to note that participants (50:50) were asked questions alternating between the first and second or the first and last parts of the interview guide.

A total of 5 pilot interviews were conducted to validate the interview guide. Based on these, the interview guide was discussed, and several questions were revised. During the pilot interviews, it was noticed that responses, particularly in the first part, might be influenced by social desirability. In response, a section explaining the interview's goals was added to the introduction, emphasizing that the aim was not to evaluate SmartCoach positively but to answer the questions as honestly as possible.

Recruitment

All interviews were conducted between January 2023 and May 2023 by phone in German, encompassing both regional dialects and standard German. The prevalent call timeframe was on weekdays, typically from approximately 4 to 7 PM, chosen to optimize accessibility for students during those hours.

Procedure

After the interventional part of the SmartCoach program was finished, participants received an SMS text message reminding them about the postprogram interview in the subsequent days. In addition, this message informed the students that they would receive CHF 10 (approximately US $10) after completing the interview. The participants' list was split randomly between the interviewers. After reaching out, the participants were informed about the interview aims. Inclusion criteria included willingness to participate in the interview and consent to the audio recording. Subsequently, the interviews started.

Participants who could not be reached during the first call attempt were called for 2 more times on other days. If these attempts were unsuccessful, a text message was sent requesting

XSL•FO
RenderX

appointment suggestions for the interview. Further contact attempts were discontinued if there was no response to this message either. In total, a maximum of 4 contact attempts were made: 3 calls and 1 text message. All audio recordings of the interviews were subsequently transcribed.

Analysis

The interviews were analyzed using the thematic analysis method [38]. NVivo 14 software was used for data analysis. Following the approach outlined by Gale et al [39] and Braun et al [40], the subsequent steps detail the process undertaken during the application of thematic analysis. Subsequently, codes were created using both an inductive and deductive approach. Inductive codes emerged directly from the data, while deductive codes followed the structure of the COM-B and TDF, as also presented in Table 1 in the "Results" section (Braun et al [40] and Castleberry and Nolen [41]). The research team approved the coding framework after several rounds of iterative adaptation. After establishing the final coding framework, the interrater reliability was calculated based on 5 transcripts (k=0.88 indicating excellent agreement). Thereafter, transcripts were split between 2 coders. Finally, 2 senior researchers additionally cross-checked the final coding results. Furthermore, it is important to mention that it was possible to allocate a statement to different codes simultaneously.

Table . Number of sources citing promoting and inhibiting factors.

COM-B[a] model components and TDF[b] domains	Source		
	Promoting factor, n	Inhibiting factor, n	Sum, n
Capability			
Physical			
Skills	2	0	2
Psychological			
Skills	9	6	15
Knowledge	3	9	12
Memory, attention, and decision process	7	44	51
Behavior regulation	1	21	22
Opportunity			
Physical			
Environmental context and resources	44	74	118
Social			
Social influences	38	3	41
Motivation			
Reflective			
Social identity	3	0	3
Beliefs about capabilities	3	1	4
Beliefs about consequences	44	2	46
Optimism	17	7	24
Goals	9	2	11
Intentions	7	13	20
Automatic			
Emotions	60	32	92
Optimism	0	2	2
Social identity	0	20	20
Reinforcement	20	3	23

[a]COM-B: capability, opportunity, and motivation model of behavior change.

[b]TDF: Theoretical Domains Framework.

XSL•FO
RenderX

Conducting a nonresponder analysis was essential to assess the generalizability of the results. This analysis examined the differences between the students who participated in the interview and those who did not. The following variables were of interest: age, gender, number of interactions in the SmartCoach program, type of school (secondary vs upper secondary school), and migration background (defined as the student themselves, their father, or their mother being from a non-German-speaking country vs none). For the analysis of the variables age and the number of interactions in the SmartCoach program, independent sample *t* tests were conducted for each. For the analysis of the variables age, school type (secondary vs upper secondary school), and migration background, the Pearson chi-square test for associations was applied. The analyses were conducted in IBM SPSS Statistics version 29.

Ethical Considerations

Ethical approval for this study was granted on April 17, 2022, by the Ethics Committee of the Faculty of Philosophy at the University of Zurich (22.2.15). In Switzerland, parental consent is not required for participation in prevention programs such as SmartCoach for adolescents aged 14 years or older. All participants provided active digital informed consent by registering for the program. Study participants completed the screening procedure and baseline assessment anonymously. No names were collected; instead, participants chose nicknames. Password protection and Secure Sockets Layer encoding were used to ensure the privacy and safety of data transfer. Study participants received an incentive of CHF 10 (approximately US $12) for taking part in the follow-up survey.

Results

Participant Characteristics

Out of the 315 students (who participated in SmartCoach) contacted, only 184 students were reached, and interviews could be conducted with 171 (54.3%) students. Another 13 (4.1%) students were reached but did not consent to the interview, all citing a lack of interest. A majority of the interviewed students were female (107/171, 62.6%). There were no students who identified themselves aside from traditional binary genders. Their age ranged from 13 to 19 years, with an average age of 15.4 (SD 1.1) years. In total, 99 (57.9%) students attended secondary school, while 72 (42.1%) students were in upper secondary school. For 84 (49.1%) students, at least 1 person in the household was from a non-German-speaking country. At the same time, the other half of the participants (85/171) were considered native-born since no person in the household was from a non–German-speaking country. Two students did not provide information about their potential migration background. The shortest interview lasted 5:46 minutes, and the longest lasted 17:44 minutes. The average interview duration was 11:33 minutes.

Nonresponder Analysis

Participants did not differ from nonresponders regarding age (t_{312}=0.281; *P*=.78), school type (χ^2_1=.206; *P*=.65; n=315) and migration background (χ^2_1=2.17; *P*=.14; n=315). However, significant differences regarding the number of interactions with the program (t_{312}=−10.63; *P*<.001) and gender (χ^2_1=5.6; *P*=.02; n=315) were found.

Factors Influencing Engagement

Every element of the TDF and COM-B model was identified during the interviews, but the frequencies varied strongly among the factors, as presented in Table 1. A total of 7 factors were mentioned fewer than 18 times by each interviewee. They are briefly summarized in the last subchapter, whereas other factors are discussed in more detail below, depending on their role and possible positive, negative, or rather ambivalent (not mostly positive or negative) influencing factors.

Factors Influencing Engagement Positively

Based on the participants' answers, 3 notably mentioned factors ("beliefs about consequences," "social influences," and "reinforcement") were identified as positively influencing engagement in the SmartCoach program. The TDF domain "beliefs about consequences" (within the reflective part of motivation in the COM-B model) was mentioned 44 times, making it the second most frequently mentioned positively influencing factor. In most cases, participants reported intrinsic expectations such as:

> *I think that can help me.* [ID 197]
>
> *[...] that you can also learn from it.* [ID 137]
>
> *[...] I thought if I participate, it might help me, and that's why I just tried it.* [ID 32]

This was often related to the topic of stress, particularly in the school context, as the following quote shows:

> *For example, in times of stress, I participated in everything […] and SmartCoach helps me a lot.* [ID 117]

The TDF domain "social influences" (as part of the social subcategory within the COM-B model) was mentioned almost as frequently as "beliefs about consequences," with 38 sources. However, only 2 sources of social influence were discussed. On the one hand, students stated that their peers motivated them a lot, and they participated because others were also involved, as the following 2 quotes show:

> *Because of my colleague, we were always together outside, and when she participated, I thought, I'll join in too.* [ID 315]
>
> *[…] my colleagues participated, and if my colleagues participate, then I participate too.* [ID 122]

On the other hand, the information and recruiting session in the classes was mentioned as an influencing factor that positively impacted engagement. Students found it informative and motivating:

> *So, this woman came to our school and explained everything, and from that moment, it really motivated me to participate, and that's why I answered all the questions.* [ID 213]
>
> *Yes [the reason was] the person who came into the classroom, the energy from that person.* [ID 281]

XSL•FO
RenderX

Another factor strengthening participation (in the SmartCoach program) was incentives (students had the opportunity to receive CHF 10 (approximately US $12) after participating in the program and to win prizes through a periodic lottery among program participants). The 20 statements regarding them, such as "simply because of the prize" [ID 205] or because "money motivates"[ID 66], were allocated to the TDF domain "reinforcement" within the automatic subcategory of motivation in the COM-B model. However, the reinforcement was not limited to material incentives, as the following 2 quotes show:

> *I liked the credit system quite a bit, that drove me, I really wanted to reach the 100 credits.* [ID 195]

> *And, on the other hand, it was an incentive that you got credits through participation.* [ID 62]

In contrast, these 3 domains were mentioned only a couple of times as inhibiting factors (a maximum of 3 times).

Factors Influencing Engagement Negatively

Next to the strengthening factors, 3 notable factors ("memory, attention, and decision process," "behavioral regulation," and "social identity") were identified as negatively influencing engagement in the SmartCoach program. The TDF domain "memory, attention, and decision processes" within the psychological component of the capability of the COM-B model was the second most frequently (n=44) cited domain that may have had inhibiting effects on engagement. Statements such as "it slipped my mind" [ID 142] and "Sometimes I forgot to respond" [ID 296] were predominant. Most students acknowledged receiving a text message and, in some cases, even read it. However, they often forgot to participate afterward, as expressed by ID 122:

> *I saw it, and then I had to leave immediately and didn't see it again, and then it got lost or forgotten.* [ID 122]

Overall, the key issue here was that students frequently forgot to respond.

Another TDF domain—"behavioral regulation"—was mentioned frequently (n=21) as an inhibitor and was allocated to the same psychological component of the capability of the COM-B model. Two main themes were prominent in this domain. First, some students mentioned that they did not even look at text messages because they do not use SMS text messaging or have turned off notifications, so the message was not perceived at all:

> *[...] because I have turned off the notification for text messages on my phone, so I don't receive notifications for text messages.* [ID 131]

Second, there was often a discussion about the time limit imposed on some interactive content by SmartCoach (the opportunity to respond was limited to 2 hours in some cases (eg, the picture contests). Many were too late to respond:

> *[...] except for the pictures and videos. Because I always saw them too late. Because in the evening, I don't always check my phone. It was until eight o'clock, and then I didn't check it anymore. [...] I just always saw the contests too late.* [ID 108]

A further group of statements (n=20) allocated to the TDF domain "social identity" within the automatic category of motivation of the COM-B model was strongly related to the issue of posting within the contests. Many interviewees could not identify themselves with uploading personal-related pictures or statements, as was asked in the program contests, as shown in the following quotes:

> *Posting is not really my thing.* [ID 129]

> *I'm not the type to post.* [ID 161]

> *Because I don't post much on social media myself, I don't have any social media except WhatsApp, and that's generally not my area, so I didn't really participate.* [ID 49]

In contrast, these 3 domains were mentioned very rarely as strengthening factors (a maximum of up to 7 times).

Factors Influencing Engagement in an Ambivalent Way

In addition to TDF domains that were reported as positive or negative influencing factors, 2 domains were reported as contradictory. On the one hand, 44 statements allocated to the TDF domain "environmental context and resources" within the physical subcategory of the COM-B model's opportunity might be interpreted as facilitators for program participation. On the other hand, according to 74 sources, this domain was reported to have the strongest inhibiting influence on engagement among all interviews. In this context, the majority of statements addressed time constraints, which were observed to be the most controversial issues, as presented in Textbox 1.

Textbox 1. Controversial quotes related to time issues were stated by interviewees.

Time as a strengthening factor:

- "So the messages always came at five o'clock on Tuesday or so. We were always on our way home from school and always had time to respond." [ID 108]
- "Most of the time it was easy because the message just came, and I didn't have anything else to do." [ID 225]

Time as an inhibiting factor:

- "It was just always at an inconvenient time." [ID 142]
- "Because the text messages always came at a certain time when you might not have needed it." [ID 44]

XSL•FO
RenderX

Other statements closely related to the time matter were the issue of overload with other duties, as the following quote shows:

> *I have enough stress because of football and school, and when I'm tired in the evening from school or because I have stress with the coach, then I consider the text message rather unimportant and annoying. I think, 'Oh, not again the text message,' and then I swipe it away. Not meant to be rude, but in that moment, I just needed time for myself.* [ID 263]

Furthermore, the statements indicated that for many, the lack of internet was also a reason for their lower participation. Some needed the internet to respond to the text message for free, and others could not open the links to interactive parts of SmartCoach without the internet:

> *I received the text message, and I wasn't at home, then I couldn't open the link because I didn't have Internet. Then I came home and forgot about it, and then it was too late, which I found unfortunate.* [ID 291]

Another group of statements that received ambivalent feedback was related to the TDF domain "emotions" from the automatic subcategory within the component "motivation" of the COM-B model. In contrast to the previous domain, 60 people believed it was a factor positively impacting program engagement. Fun, interest in the program, and perception of the content as exciting were particularly prominent. Many mentioned that it was "interesting to see what comes as a response" [ID 53] and found the program "really good" [ID 250]. A respondent said:

> *Honestly, I found it still exciting, answering the questions was fun, thinking about what you are interested in, and reading the tips.* [ID 165]

And another respondent expressed:

> *Because I found it super cool, I always looked forward to it; it was also my favorite task in the whole study. I thought it was cool to show what helps me and to see what others have said, what target groups were also participating in the study.* [ID 223]

However, for 32 interviewees, this domain inhibited engagement. Furthermore, concerning the frequency of mentioning, this domain was named the second most frequent inhibiting factor. In the absolute majority of answers, the reasons were lack of interest and lack of desire to participate, as the following 2 statements present:

> *So if the topics didn't interest me 100%, then I skipped them.* [ID 152]

> *I just received the messages and didn't look at them closely anymore. Actually, there wasn't a real reason; I had seen it over time and didn't feel like participating anymore.* [ID 93]

Factors Influencing Engagement Mentioned Rarely

From the rest of the domains that were rarely reported (under 18 times), 2 domains were named in sum (promoting + inhibiting influence) by over 10% of the sources: the TDF domains "Optimism" (n=24) and "Intentions" (n=20)—both within the automatic part of the component "Motivation" in the COM-B model. However, the statements for both were somewhat ambiguous. They might not be allocated firmly to promoting or inhibiting factors, making it impossible to consider them a silent influencing factor for a specific minority.

Suggestions Regarding Program Adjustments

The responses regarding possible adjustment of the program (these questions formed the second part of the interview and were asked only to half of the respondents [Methods section]) were categorized into 2 notable subcodes with the themes of "Text Messages" (n=43) and "Contests" (n=29). Most of the suggestions for optimizing text messages were related to reminder messages (n=19) and the timing of the messages (n=30). However, a closer analysis of the suggestions associated with the time of the messages showed that the opinion regarding better messaging time split almost equally among 3 proposals: SMS text message on the weekend (n=15), in the late evening (n=14), or the morning (n=10). Besides a couple of statements about the number of contests (more [n=5] vs less [n=2]), participants who talked about the issue of contests recommended increasing the time limit to react to the contest (n=19).

The same picture showed the replies to the question about the communication medium. One-half of the participants rated SMS text messaging positively (n=73), calling SMS text messaging "*uncomplicated*"[ID 195], *"great"* [ID 19], *"clear"* [ID 197], and *"practical"* [ID 225]. However, another half was convinced that better communication methods, instead of SMS text messaging, should be used (n=69). Among them, WhatsApp (Meta) was mentioned most frequently (n=20). In addition, it should be noted that almost half of this subgroup added that even though better communication methods exist, SMS text messaging still seemed acceptable (n=36).

Overall, participants evaluated the length of the program as well as the number and frequency of the messages mostly positively (n=60 and n=72), with a considerable number of individuals wishing for a shorter program (n=29), compared with a few people wanting a more extended program (n=11). Only a few individuals wished for more (n=14) or fewer (n=6) SMS text messages during the program.

In addition to the two main subcodes presented above ("Text Messages" and "Contests"), two topics within the sum of over 30 residual statements regarding program adjustment (n=32) were mentioned more than twice and therefore warrant further discussion. Within this residual subcategory, personalization (n=5) was most related to the issue that the messages seemed to be too "robotic," as the following example shows:

> *The response sometimes seemed almost programmed, with the 'Too bad' and all, sometimes you almost felt like you were talking to a robot, […] maybe making it more personal when there is capacity for it.* [ID 223]

Another topic mentioned 5 times (n=5) within the residual category was related to the tips or suggestions that sounded simple but were challenging to implement. A few participants wished for more detailed guidance on how to implement the

XSL•FO
RenderX

tips. The remaining suggestions regarding program adjustments were strongly heterogeneous and sometimes opposed to one another.

Furthermore, it needs to be noted that a significant part of the participants reported positive statements (n=57), such as "very exciting" [ID 165], "very helpful" [ID 107], and "understandable" [ID 285] regarding the program, and at the same time, also addressed the issue of potential adjustment. The following statement summarizes the positive answers as follows:

> *I found it good, you filled out at the beginning which topics are important to you, and there were always questions/tips in that area, so I found it good.* [ID 255]

Feedback on the Program's Content

Since the program consisted of 3 major blocks (self-management skills, social skills, and substance use resistance skills), interviewees were asked about feedback on these 3 superior topics. Regarding self-management skills, most responses were positively allocated to the subcodes "helpful" (n=29) and "interesting" (n=32). At the same time, 20 individuals considered the topic less relevant to them. Many expressed that they were already familiar with the tips and strategies or did not personally experience a personal necessity, making them less relevant, as the following quote shows:

> *So, I didn't really need the tips for school stress; I'm doing well in school. They were good, but I didn't need them.* [ID 193]

Only 5 people stated that the tips related to social skills were not helpful. The majority considered the tips concerning social skills as "interesting" (n=34) or helpful (n=17). At the same time, 26 interviewees believed that the statements were not relevant to them, as the following quotation highlights:

> *[…] I don't really have problems dealing with others, so I couldn't really use them […].* [ID 171]

Answers related to substance use resistance skills mirrored the same pattern as the previous 2 content areas: 37 persons considered the topic "interesting," and another 23 described it as "important to discuss." In contrast, 23 interviewees thought the topic was irrelevant to them.

Discussion

Principal Findings

This study aimed to identify factors derived from the COM-B model and the TDF influencing engagement in an mHealth program for adolescent substance use prevention. The findings provide valuable insights into promoting and inhibiting factors affecting engagement, accessed by means of structured qualitative interviews, offering practical implications for enhancing the effectiveness of digital interventions. In the following section, the results are first discussed individually in relation to their specific outcomes and then synthesized to highlight overarching patterns, including contradictory findings and their implications for future implementation.

Promoting Factors for Engagement

Qualitative analyses revealed that the following 5 TDF domains most commonly positively influenced program participation: "environmental context and resources," "social influences," "emotions," "reinforcement," and "beliefs about consequences." Specifically, the following factors were perceived as promoting engagement: the right timing of text messages, the involvement of other classmates, collaborative content processing, and the informational session during recruitment. In addition, engaging intervention content, rewards, and incentives in the form of points and prizes, and the inherent curiosity of participants about the program content further enhanced engagement.

These findings align with previous research emphasizing the importance of timely and relevant content in digital interventions [42,43]. The role of social support and interactive content in promoting engagement is well-documented [43-45]. The effectiveness of rewards and incentives in motivating participation has also been supported by studies on digital health interventions [44].

Inhibiting Factors Regarding Engagement

In terms of inhibiting factors regarding engagement, the following 5 TDF domains were identified as particularly relevant: memory, attention, decision process, behavior regulation, environmental context and resources, emotion, and social identity. Specifically, the following contents were found to inhibit engagement: forgetting to participate, inactivity on the communication channel and the time limit, lack of time resources and internet access, concerns related to personal disclosure, lack of interest and desire, and a lack of identification with a topic.

These barriers are consistent with challenges noted in previous studies. Forgetfulness and timing issues have been highlighted as common obstacles in digital interventions [43,44]. The need for personalization and flexibility in intervention delivery is crucial to address diverse user needs and preferences [42,46-48]. Concerns related to personal disclosure, particularly regarding not wishing to share personal or even any information, remain a significant barrier to digital health engagement [42,43].

Domains That Can Have Both Inhibiting and Facilitating Influences

Interestingly, some factors, such as the "Environmental context and resources" and "Emotion" domains, had both promoting and inhibiting influences. Positive emotions, such as finding the content fun and interesting, facilitated engagement, while negative emotions, such as a lack of interest or feeling overwhelmed, inhibited participation. Similarly, the timing and context of text messages could either support or hinder engagement depending on individual circumstances.

Implications for the Further Development of the SmartCoach Program

To maintain the successful elements of the SmartCoach program, it is essential to preserve the passionate and engaging introduction sessions conducted in schools, as these have proven to significantly boost initial interest and participation. In addition, the use of SMS text messages or comparable

XSL•FO
RenderX

messaging systems as a communication medium should be retained, given its widespread acceptance and effectiveness in reaching adolescents. These components have demonstrated their value in fostering engagement and should remain integral parts of the program.

At the same time, based on the findings of the study, several optimization suggestions concerning the SmartCoach intervention program were derived. First, to increase engagement regarding the usage of program content (reading the messages, answering them, watching additional topic-related materials [eg, videos]), the introduction of a text message reminder for participation is suggested. This can help reduce the likelihood of forgetting to participate, even if the text message is received at an inconvenient time. This solution would address the "memory, attention, and decision process" aspect frequently mentioned in the context of inhibited acting. The use of reminders to enhance participation is also found in the literature [46,48]. However, it is important to ensure that sending reminders via text message does not exceed the tolerated number of messages. Intrusive notifications can trigger negative emotions or reactance, which, in turn, can harm engagement [44].

Closely connected to this aspect is also the reported fact related to the "Environmental context and resources" that participation was often forgotten because the text message arrived at an inconvenient time for many participants. In such cases, responding was postponed and then forgotten. Therefore, changing the text message's timing may be meaningful in reaching the users when they are receptive and available to engage with the intervention [49]. A significant number of respondents expressed a preference for the weekend as a better time. However, opinions vary regarding the time of day. In contrast, the timing of the text message was also found to be among the facilitating factors, indicating that many participants were satisfied with the timing. Here, it was a clear aspect that might be both inhibiting and promoting activity depending on the individual situation. A solution could be stronger personalization: allowing participants to choose or define an individual time slot for receiving text messages during the initial survey. Personalization as a strategy to increase engagement is well-represented in the literature [42,46-48].

An issue that also frequently led to low participation was the time limit for completing picture or message contests. Participants' suggestions in this regard were to increase the time limit, with the majority desiring an extension from the original 2 hours to several days. It is important to note that while it may be sensible to increase the time limit to this extent, there is also a risk that this could lead to an increased likelihood of forgetting to participate. Therefore, in the context of extending the time limit, the consideration of reminders should also be taken into account. In combination with a later reminder, an extension for a few hours might also be considered, giving participants a longer time slot, such as an entire evening, and allowing them to react later in case of other activities during the evening. This optimization suggestion is specific to SmartCoach and has not been confirmed by the existing literature.

In addition to the options for addressing the inhibiting factors discussed above, 3 further domains—"behavior regulation," "social identity," and "emotions"—were frequently mentioned as barriers but appear to be more challenging to address directly through program design alone. These domains are closely tied to individual preferences, values, and psychological traits, making one-size-fits-all solutions less feasible.

For example, behavior regulation, such as turning off all notifications, is typically a matter of personal autonomy. A possible response could be to inform participants at the start of the program about the importance of notifications and recommend enabling them during participation. Still, this measure may have a limited effect, as SMS text message notifications cannot easily be customized for individual programs such as SmartCoach, and participants might deliberately choose to restrict digital interruptions as part of their general phone usage behavior.

Similarly, social identity-related barriers, such as discomfort with posting or participating in peer group challenges, often reflect deeper self-concepts or social positioning. Removing these program components entirely could have unintended negative consequences for other users by weakening positively perceived elements such as "beliefs about consequences," which emerged as a key engagement facilitator in this study. As a possible optimization, offering participants the choice to engage in such content at the beginning of the program—similar to the possibility of choosing the preferred time for receiving messages as described above—could help accommodate individual preferences. In addition, providing alternative task types for those who are reluctant to take part in group interactions or offering clearer information about who can see uploaded content may further reduce discomfort and increase perceived safety. Maintaining a flexible structure that allows optional participation in such tasks may help to protect both individual comfort and the overall program effect.

In addition, the domain of emotions—although mentioned in both promoting and inhibiting ways—was described more frequently as a positive contributor to engagement. Therefore, attempting to influence this factor through design changes should be cautiously approached. Interventions that try to evoke specific emotions or suppress negative ones risk being perceived as manipulative and could reduce user trust or trigger reactance [50,51]. Thus, maintaining an emotionally balanced, user-sensitive communication tone may be the most appropriate strategy. The domains "reinforcement," "beliefs about consequences," and "social influences" from the TDF were identified in this study as positively associated with participant engagement. From this perspective, it is essential to maintain and further strengthen program elements that foster these perceptions. For instance, financial incentives such as the small monetary reward at the end of the program served as a reinforcement mechanism; statements such as "*everyone participated, so I did too*" reflected the impact of peer dynamics; and beliefs that the program would be useful or bring something new were key motivators for engagement.

However, an overly strong emphasis on these aspects might have unintended consequences. For example, excessive reliance

XSL•FO
RenderX

on social conformity ("everyone else is doing it") could undermine intrinsic motivation and lead to disengagement among those who value autonomy. This may also relate to the previously discussed aspect of social identity, particularly in cases where individuals prefer not to share personal content or engage in group-based activities. Similarly, if reinforcement mechanisms such as financial incentives are overused, participants may perceive the program as externally controlling rather than personally meaningful. A careful balance is therefore needed to preserve the motivational power of these domains without inducing dependency, loss of authenticity, or reactance.

Since, as already mentioned, not all topics addressed during the SmartCoach program were relevant to everyone, there were suggestions for additional topics that participants might find more interesting. One possible solution might be applying the topics' selection for the individual challenges to the regular messages or adapting the messages more closely to the preselected topics for the individual challenges. However, one must consider whether participants might misappropriate the possibility of strong personalization of topics; ultimately, SmartCoach remains a program for addiction prevention and should not serve as entertainment.

Finally, a few participants also wished for a more detailed explanation of how to implement the tips and strategies. Providing a clear user guide for using an intervention is also discussed in previous literature as a facilitating factor for engagement [44]. However, the situations participants face might be much more complex than what a single tip or text can address directly. Therefore, more comprehensive support mechanisms, such as an "Ask-The-Expert" or similar, would be more valuable.

It is also important to continue involving students in the content creation and gathering feedback on the program. Literature confirms as well that the participatory development of comparable interventions with specific target groups (eg, disadvantaged groups) often results in higher engagement regarding the usage of the "products" [24,25,52]. Furthermore, consideration could also be given to surveying teachers. On one hand, the support of teachers is essential for recruiting school classes, as they do not receive compensation for the lessons they provide for recruitment. In addition, their perception of the class during the ongoing SmartCoach program could be valuable.

Strengths and Limitations

The main strength of this study is that comprehensive qualitative interviews were conducted using a structured and theoretically sound framework (thereby enhancing the validity of results)—the COM-B model and TDF framework—in a relatively large sample of adolescents to derive factors associated with engagement in a digital substance use prevention program among adolescents. This participatory approach ensures that the optimization suggestions are practical and aligned with user needs. In addition to the interviews, nonresponder analyses were conducted to assess the generalizability of the results (regarding the sample) and understand differences between participants and nonparticipants.

The use of qualitative methods, particularly the use of NVivo for thematic analysis, is another significant strength [31]. This approach allows for a deep, nuanced understanding of participant experiences and perspectives, which is often not achievable through quantitative methods alone. The process is time-intensive and requires meticulous attention to detail, reflecting a high level of commitment and expertise. In general, qualitative research of this nature is relatively rare in the setting of the present research area, but is crucial for complementing quantitative findings identified in the primary study [36]. It adds depth to the understanding of engagement factors and provides rich, context-specific insights that can drive more effective program improvements.

At the same time, some limitations of this study should also be considered. First, it was challenging to reach students by phone to conduct the qualitative interviews. After the third attempt at contact, many still could not be reached, and only a few responded to the subsequent text message requesting a proposed appointment. As a result, 4 of 10 (41%) students could not be reached and thus were not interviewed. This limits the representativeness and generalizability of the results. In this regard, the results of the nonresponder analysis indicated a significant association between participation in the interview and gender as well as the number of interactions in the SmartCoach program. Students who did not participate in the interview were more likely to be male and showed lower program engagement. This suggests a bias in the results reported toward female students with higher program engagement.

Second, the study sample consists of adolescents from specific regions in Switzerland, which may limit the generalizability of the findings to other populations or cultural contexts. Variations in substance use norms, digital health literacy, and educational systems across different regions and countries could impact the applicability of the results.

Third, socially desirable responses could not be entirely ruled out, despite a section that addressed this issue during the validation of the questionnaire. Fourth, qualitative research is characterized by subjectivity. On one hand, interviews are a form of self-disclosure and can therefore be distorted by phenomena such as memory distortion, personal perspectives, or the aforementioned social desirability. On the other hand, research questions, theoretical approaches, methods, and interpretations of results are also influenced to some extent by the subjectivity of the researchers [53]. This leads to a limitation of objectivity.

Finally, the timing of the study may have influenced engagement levels, especially if conducted during periods of heightened stress or significant events (eg, exam periods, pre-Christmas, and holidays). These temporal factors were not controlled for, which could affect the consistency of engagement data.

Conclusion

This study identified key factors influencing engagement in a digital life skills training program for addiction prevention in adolescents. By addressing the identified barriers (eg, "Environmental context and resources" by adjusting the timing of messages and "Memory, attention, and decision process" by

XSL•FO
RenderX

implementing reminders) and enhancing promoting factors such as "Beliefs about positive consequences," digital interventions such as SmartCoach can achieve higher engagement rates and, ultimately, greater effectiveness in preventing substance use among adolescents. The insights gained contribute to the broader understanding of engagement in digital health interventions and offer a concrete approach for the continued development of SmartCoach and similar programs. The combination of these findings derived from qualitative research, in combination with results from quantitative predictor analysis [36], might also be of interest to optimize and tailor intervention programs for specific subgroups, for example, those defined by specific social stratifying factors or different levels of program engagement.

Acknowledgments

Funding for this project was provided by the Swiss National Science Foundation (10001C_204412/1). Since the authors are not native English speakers, we used ChatGPT and Grammarly to improve the language, grammar, and clarity of the manuscript. No content was generated by artificial intelligence. All scientific content and interpretations were developed by the authors.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SH contributed to the conceptualization, methodology, supervision, funding acquisition, resources, validation, and writing—review and editing. NK was responsible for methodology, supervision, project administration, investigation, formal analysis, validation, and both the original draft and review and editing of the manuscript. RS conducted the investigation and formal analysis and contributed to the original draft. ZS was involved in the investigation, formal analysis, and drafting of the manuscript. AW supported data curation, software development, visualization, and writing—original draft. RPC contributed to the conceptualization, methodology, validation, and writing—original draft.

Conflicts of Interest

None declared.

References

1. Peacock A, Leung J, Larney S, et al. Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. Addiction 2018 Oct;113(10):1905-1926. [doi: 10.1111/add.14234] [Medline: 29749059]
2. Manthey J. Cannabis use in Europe: current trends and public health concerns. Int J Drug Policy 2019 Jun;68:93-96. [doi: 10.1016/j.drugpo.2019.03.006] [Medline: 31030057]
3. Davies P, Walsh D. Alcohol Problems and Alcohol Control in Europe: Taylor & Francis; 2023. URL: https://tinyurl.com/2vnzwvr5 [accessed 2025-07-30]
4. Dai X, Gakidou E, Lopez AD. Evolution of the global smoking epidemic over the past half century: strengthening the evidence base for policy action. Tob Control 2022 Mar;31(2):129-137. [doi: 10.1136/tobaccocontrol-2021-056535] [Medline: 35241576]
5. European Monitoring Centre for Drugs and Drug Addiction. European drug report 2023: trends and developments. : EUDA; 2023 URL: https://www.euda.europa.eu/publications/european-drug-report/2023_en [accessed 2025-07-30]
6. Inchley J, Currie DB, Budisavljevic S, Torsheim T, Jastad A, Cosma A, et al. Spotlight on adolescent health and well-being. findings from the 2017/2018 health behaviour in school-aged children (HBSC) survey in europe and canada. international report. volume 1. key findings. : World Health Organization; 2020 URL: https://iris.who.int/bitstream/handle/10665/332091/9789289055000-eng.pdf?sequence=1 [accessed 2025-08-07]
7. Jordan MD, Balsiger N, Schmidhauser V. Consommation de substances psychoactives chez les 11 à 15 ans en suisse–situation en 2022 et évolution dans le temps. : Addiction Suisse; 2023 URL: https://www.addictionsuisse.ch/wp-content/uploads/2023/03/HBSC_2022_Rapport_substances_DEF_26032023.pdf [accessed 2025-08-07]
8. Quednow BB, Steinhoff A, Bechtiger L, Ribeaud D, Eisner M, Shanahan L. High prevalence and early onsets: legal and illegal substance use in an urban cohort of young adults in Switzerland. Eur Addict Res 2022;28(3):186-198. [doi: 10.1159/000520178] [Medline: 34864731]
9. Weichold K, Blumenthal A. Long-term effects of the life skills program IPSY on substance use: results of a 4.5-year longitudinal study. Prev Sci 2016 Jan;17(1):13-23. [doi: 10.1007/s11121-015-0576-5] [Medline: 26202801]
10. Velasco V, Griffin KW, Botvin GJ, Corrado Celata and Gruppo LST Lombardia. Preventing adolescent substance use through an evidence-based program: effects of the Italian adaptation of life skills training. Prev Sci 2017 May;18(4):394-405. [doi: 10.1007/s11121-017-0776-2] [Medline: 28353126]
11. Thomas RE, McLellan J, Perera R. School-based programmes for preventing smoking. Cochrane Database Syst Rev 2013 Apr 30;2013(4):CD001293. [doi: 10.1002/14651858.CD001293.pub3] [Medline: 23633306]

XSL•FO
RenderX

12. Foxcroft DR, Tsertsvadze A. Universal school-based prevention programs for alcohol misuse in young people. Cochrane Database Syst Rev 2011 May 11;11(5):CD009113. [doi: 10.1002/14651858.CD009113] [Medline: 21563171]
13. Ellickson PL. You've shown the program model is effective. Now what? New Dir Youth Dev 2014;2014(141):95-105. [doi: 10.1002/yd.20089] [Medline: 24753281]
14. Domitrovich CE, Durlak JA, Staley KC, Weissberg RP. Social-emotional competence: an essential factor for promoting positive adjustment and reducing risk in school children. Child Dev 2017 Mar;88(2):408-416. [doi: 10.1111/cdev.12739] [Medline: 28213889]
15. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
16. Garrido S, Millington C, Cheers D, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019;10(10):759. [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
17. MacDougall S, Jerrott S, Clark S, Campbell LA, Murphy A, Wozney L. Text message interventions in adolescent mental health and addiction services: scoping review. JMIR Ment Health 2021 Jan 8;8(1):e16508. [doi: 10.2196/16508] [Medline: 33416504]
18. Perski O, Hébert ET, Naughton F, Hekler EB, Brown J, Businelle MS. Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. Addiction 2022 May;117(5):1220-1241 [FREE Full text] [doi: 10.1111/add.15687] [Medline: 34514668]
19. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852. [doi: 10.1037/amp0000983] [Medline: 35298199]
20. Couper MP, Alexander GL, Zhang N, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52. [doi: 10.2196/jmir.1430] [Medline: 21087922]
21. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3(3):764079. [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
22. Graham AK, Kwasny MJ, Lattie EG, et al. Targeting subjective engagement in experimental therapeutics for digital mental health interventions. Internet Interv 2021 Sep;25:100403. [doi: 10.1016/j.invent.2021.100403] [Medline: 34401363]
23. Paz Castro R, Haug S, Debelak R, Jakob R, Kowatsch T, Schaub MP. Engagement with a mobile phone-based life skills intervention for adolescents and its association with participant characteristics and outcomes: tree-based analysis. J Med Internet Res 2022 Jan 19;24(1):e28638. [doi: 10.2196/28638] [Medline: 35044309]
24. Greene K, Choi HJ, Glenn SD, Ray AE, Hecht ML. The role of engagement in effective, digital prevention interventions: the function of engagement in the REAL media substance use prevention curriculum. Prev Sci 2021 Feb;22(2):247-258. [doi: 10.1007/s11121-020-01181-9] [Medline: 33140287]
25. Greene K. The theory of active involvement: processes underlying interventions that engage adolescents in message planning and/or production. Health Commun 2013;28(7):644-656. [doi: 10.1080/10410236.2012.762824] [Medline: 23980581]
26. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
27. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77. [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
28. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572. [doi: 10.2196/17572] [Medline: 32348255]
29. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of health and well-being apps and curated app portals: think-aloud and interview study. JMIR Mhealth Uhealth 2021 Apr 27;9(4):e27173. [doi: 10.2196/27173] [Medline: 33904827]
30. Rosário F, Santos MI, Angus K, Pas L, Ribeiro C, Fitzgerald N. Factors influencing the implementation of screening and brief interventions for alcohol use in primary care practices: a systematic review using the COM-B system and Theoretical Domains Framework. Implement Sci 2021 Jan 7;16(1):6. [doi: 10.1186/s13012-020-01073-0] [Medline: 33413487]
31. Guest G, MacQueen K, Namey E. Applied Thematic Analysis: SAGE Publications; 2012.
32. Levitt HM, Bamberg M, Creswell JW, Frost DM, Josselson R, Suárez-Orozco C. Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. Am Psychol 2018 Jan;73(1):26-46. [doi: 10.1037/amp0000151] [Medline: 29345485]
33. Haug S, Paz Castro R, Wenger A, Schaub MP. Efficacy of a mobile phone-based life-skills training program for substance use prevention among adolescents: study protocol of a cluster-randomised controlled trial. BMC Public Health 2018 Sep 10;18(1):1102. [doi: 10.1186/s12889-018-5969-5] [Medline: 30200928]

XSL•FO
RenderX

34. Haug S, Paz Castro R, Wenger A, Schaub MP. A mobile phone-based life-skills training program for substance use prevention among adolescents: cluster-randomized controlled trial. JMIR Mhealth Uhealth 2021 Jul 13;9(7):e26951. [doi: 10.2196/26951] [Medline: 34255703]
35. Paz Castro R, Haug S, Wenger A, Schaub MP. Longer-term efficacy of a digital life-skills training for substance use prevention. Am J Prev Med 2022 Dec;63(6):944-953. [doi: 10.1016/j.amepre.2022.06.017] [Medline: 35985899]
36. Haug S, Boumparis N, Wenger A, Schaub MP, Kiselev N. Predictors of youth accessibility for a mobile phone-based life skills training program for addiction prevention. Int J Environ Res Public Health 2023 Jul 17;20(14):6379. [doi: 10.3390/ijerph20146379] [Medline: 37510611]
37. Keddem S, Barg FK, Frasso R. Practical guidance for studies using freelisting interviews. Prev Chronic Dis 2021 Jan 14;18(18):E04. [doi: 10.5888/pcd17.200355] [Medline: 33444525]
38. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
39. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117. [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
40. Braun V, Clarke V, Hayfield N, Terry G. Thematic analysis. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences: Springer; 2019:843-860 URL: https://link.springer.com/rwe/10.1007/978-981-10-5251-4_103#citeas [accessed 2025-07-30]
41. Castleberry A, Nolen A. Thematic analysis of qualitative research data: is it as easy as it sounds? Curr Pharm Teach Learn 2018 Jun;10(6):807-815. [doi: 10.1016/j.cptl.2018.03.019] [Medline: 30025784]
42. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120. [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
43. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
44. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Perceptions of factors influencing engagement with health and well-being apps in the United Kingdom: qualitative interview study. JMIR Mhealth Uhealth 2021 Dec 16;9(12):e29098. [doi: 10.2196/29098] [Medline: 34927597]
45. Sharpe EE, Karasouli E, Meyer C. Examining factors of engagement with digital interventions for weight management: rapid review. JMIR Res Protoc 2017 Oct 23;6(10):e205. [doi: 10.2196/resprot.6059] [Medline: 29061557]
46. Milward J, Drummond C, Fincham-Campbell S, Deluca P. What makes online substance-use interventions engaging? A systematic review and narrative synthesis. Digit HEALTH 2018;4:2055207617743354. [doi: 10.1177/2055207617743354] [Medline: 29942622]
47. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
48. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000. [doi: 10.2196/30000] [Medline: 34931995]
49. Künzler F, Mishra V, Kramer JN, Kotz D, Fleisch E, Kowatsch T. Exploring the state-of-receptivity for mHealth interventions. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Dec;3(4):140. [doi: 10.1145/3369805] [Medline: 36159209]
50. Dillard JP, Shen L. On the nature of reactance and its role in persuasive health communication. Commun Monogr 2005 Jun;72(2):144-168. [doi: 10.1080/03637750500111815]
51. Brown SL. Emotive health advertising and message resistance. Aust Psychol 2001 Nov 1;36(3):193-199. [doi: 10.1080/00050060108259655]
52. Radu I, Scheermesser M, Spiess MR, Schulze C, Händler-Schuster D, Pehlke-Milde J. Digital health for migrants, ethnic and cultural minorities and the role of participatory development: a scoping review. Int J Environ Res Public Health 2023 Oct 23;20(20):6962. [doi: 10.3390/ijerph20206962] [Medline: 37887700]
53. Kalu ME. How does "subjective I" influence a qualitative research question, theoretical approach and methodologies? Glo Jnl Pure Appl Sci 2019;25(1):97. [doi: 10.4314/gjpas.v25i1.13]

Abbreviations

COM-B: Capability, Opportunity, Motivation-Behavior
JARS: Journal Article Reporting Standards for Qualitative Research
mHealth: mobile health
TDF: Theoretical Domains Framework
WHO: World Health Organization

XSL•FO
RenderX

Edited by L Buis; submitted 30.07.24; peer-reviewed by M Hecht, S Bull; revised version received 25.05.25; accepted 28.05.25; published 14.08.25.

Please cite as:
Kiselev N, Schaffner R, Csepregi Z, Wenger A, Paz Castro R, Haug S
Factors Influencing Engagement in a Digital Substance Use Prevention Program: Qualitative Study Using the Capability, Opportunity, Motivation Model and Theoretical Domains Framework
JMIR Mhealth Uhealth 2025;13:e64861
URL: https://mhealth.jmir.org/2025/1/e64861
doi:10.2196/64861

© Nikolai Kiselev, Rebecca Schaffner, Zsofia Csepregi, Andreas Wenger, Raquel Paz Castro, Severin Haug. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

WhatsApp-Based Coaching Program to Support Smoking and Vaping Cessation Among Young People: Pre-Post Study on Acceptance and Preliminary Efficacy

Severin Haug, PhD; Lisa Caputo, MSc; Andreas Wenger, MSc; Nikolai Kiselev, PhD; Olivia Studhalter, MSc; Michael P Schaub, PhD

Swiss Research Institute for Public Health and Addiction, University of Zurich, Konradstrasse 32, Zurich, Switzerland

Corresponding Author:
Severin Haug, PhD
Swiss Research Institute for Public Health and Addiction, University of Zurich, Konradstrasse 32, Zurich, Switzerland

Abstract

Background: The use of tobacco cigarettes and electronic nicotine products is widespread among young people in Switzerland. At the same time, the instant messaging platform WhatsApp (Meta Platforms, Inc) is the most frequently used smartphone app in this population group. The provision of individually tailored, evidence-based coaching messages via WhatsApp seems promising to support smoking cessation in adolescents and young adults.

Objective: This study aims to test the feasibility, acceptance, and preliminary efficacy of a newly developed, semiautomated WhatsApp-based intervention program to support smoking and vaping cessation and reduction in adolescents and young adults.

Methods: Recruitment took place in Switzerland in 2023 and 2024 via various channels, both online and offline. For a period of 11 weeks, regular users of cigarettes or electronic cigarettes, aged between 16 and 30 years, received individually tailored messages on how to deal with cravings or stressful situations and how to stop or reduce smoking. A separate WhatsApp channel provided the opportunity to ask individual questions to a counselor. A one-group pre-post design was used to obtain preliminary information on the acceptability and potential efficacy of the program.

Results: A total of 167 young people (mean age 23.2, SD 4.0 years; n=95, 56.9% women and n=72, 43.1% men) who regularly smoked tobacco cigarettes (n=81, 48.5%), vaped electronic nicotine products (n=17, 10.2%), or used both (n=69, 41.3%) were recruited for participation in the program. Of these, 100 (59.9%) intended to stop smoking or vaping while 67 (40.1%) aimed at reducing their use. The participants actively engaged in an average of 5.5 (SD 3.5) of the 11 program weeks, the average number of interactions with the program was 26.8 (SD 26.1), and the average duration from the start of the program to the last interaction was 45.0 (SD 31.1) days. The follow-up survey at the end of the 11-week coaching program was completed by 108 (64.7%) participants. The generalized estimating equation (GEE) analyses revealed significant reductions in the mean number of days in the last 30 days on which tobacco cigarettes were used from 20.6 (SD 11.8) at baseline to 14.0 (SD 12.0) at post assessment (incidence rate ratio [IRR] 0.68, P<.001) and for electronic nicotine products from 11.1 1 (SD 13.1) days at baseline to 7.7 (SD 11.3) days at follow-up (IRR 0.71, P=.005). Overall, 6/108 (5.6%) participants in the follow-up survey stated that they neither consumed tobacco cigarettes nor electronic nicotine products in the last 30 days.

Conclusions: The WhatsApp-based program appears to be a feasible, moderately accepted, and promising intervention for reducing the consumption of tobacco cigarettes and electronic nicotine products among young people. A larger-scale randomized controlled trial would be reasonable in order to make more substantiated statements about the program's efficacy.

(JMIR Mhealth Uhealth 2025;13:e65301) doi:10.2196/65301

KEYWORDS

smoking cessation; vaping; nicotine; mobile phone; WhatsApp; young people; adolescents

Introduction

Nicotine and tobacco use is a major cause of the disease burden in most countries of the world [1]. Every year, it causes 7.7 million deaths worldwide and more than 9000 deaths in Switzerland, with the majority of tobacco-related deaths due to noncommunicable chronic diseases, such as cardiovascular diseases, respiratory diseases, and tumors [2,3]. Despite the well-documented health risks, many young people, that is, adolescents and young adults, start smoking or experiment with smoking. In Switzerland, 27% of men and 21% of women are regularly smoking tobacco, with the highest prevalence rate of 30% among young adults between the ages of 20 and 24 years [4]. Young people should therefore be at the center of preventive and early interventions, particularly because quitting smoking

XSL•FO
RenderX

before the age of 30 years can prevent more than 97% of the excess mortality caused by continued smoking [5].

Although the consumption of tobacco cigarettes has declined in many European countries in recent years, there has been an increase in the consumption of electronic nicotine products, such as electronic cigarettes (e-cigarettes), also known as vapes, particularly among adolescents and young adults between the ages of 15 and 24 years [6]. There is substantial evidence that e-cigarettes can aid in smoking cessation for adults [7] and that e-cigarettes may be less harmful than cigarettes [8]; however, there is currently insufficient scientific evidence to fully understand the short- and long-term health effects of these electronic nicotine products [7]. They are not harm-free, particularly for adolescents, as vaping is associated with similar nicotine metabolite levels compared with smoking tobacco [9]. Furthermore, there is substantial evidence that electronic nicotine products can cause addiction and that young nonsmokers who use these products are more likely than nonusers to initiate tobacco smoking [10,11].

There is extensive literature demonstrating the effectiveness of traditional behavioral counseling and pharmacological interventions for smoking cessation in adults [12-14]; however, there is only little evidence regarding their effectiveness in adolescents and young adults [15]. Furthermore, their widespread implementation faces barriers, and they are only used by a minority of young people with a high motivation to quit [16,17].

Mobile phone–based interventions have the potential to reach young smokers and vapers who are not or only partially reached by traditional interventions, owing to their 24/7 availability and anonymity [18]. A review on mobile phone–based smoking cessation interventions among young people demonstrated the efficacy of text messaging–based interventions [17], while the evidence regarding the effectiveness of smartphone app–based interventions is currently inconclusive [17,19]. In recent years, social media–based interventions have been increasingly developed and shown to be a promising way of delivering smoking cessation content to young people [20,21]. Social media–based smoking cessation interventions that use platforms like Facebook (Meta Platforms, Inc), TikTok (ByteDance), or WeChat (Tencent) typically provide peer support, individually tailored content, and community engagement, while traditional text message–based interventions deliver more structured, automated, and individually tailored messages. A randomized controlled trial with young adults from the United States investigated the effectiveness of a Facebook-based smoking cessation program and revealed significant short-term effects on abstinence [22]. In a controlled trial with adult Chinese smokers, the effectiveness of a WeChat-based smoking cessation intervention was demonstrated [23].

To date, no studies have been conducted on the acceptance or effectiveness of automated WhatsApp-based smoking cessation programs. Nevertheless, WhatsApp has characteristics that make it a promising platform for delivering smoking and vaping cessation interventions. In Switzerland, as in many other European countries, WhatsApp is the most widely used instant messenger and the most used smartphone app among adolescents and young adults [24]. Similar to SMS text messaging, WhatsApp is a regularly used communication channel that requires neither familiarization with the functionality of a new platform nor the opening of a standalone app. In addition, WhatsApp offers significant advantages over text messaging, including the ability to directly incorporate media objects and the option to send messages free of charge with an existing internet connection.

The objectives of this study were (1) to test the feasibility of a WhatsApp-based smoking and vaping cessation intervention, (2) to investigate its acceptance among adolescent and young adult nicotine users, and (3) to provide a test of its preliminary efficacy.

Methods

Intervention Program

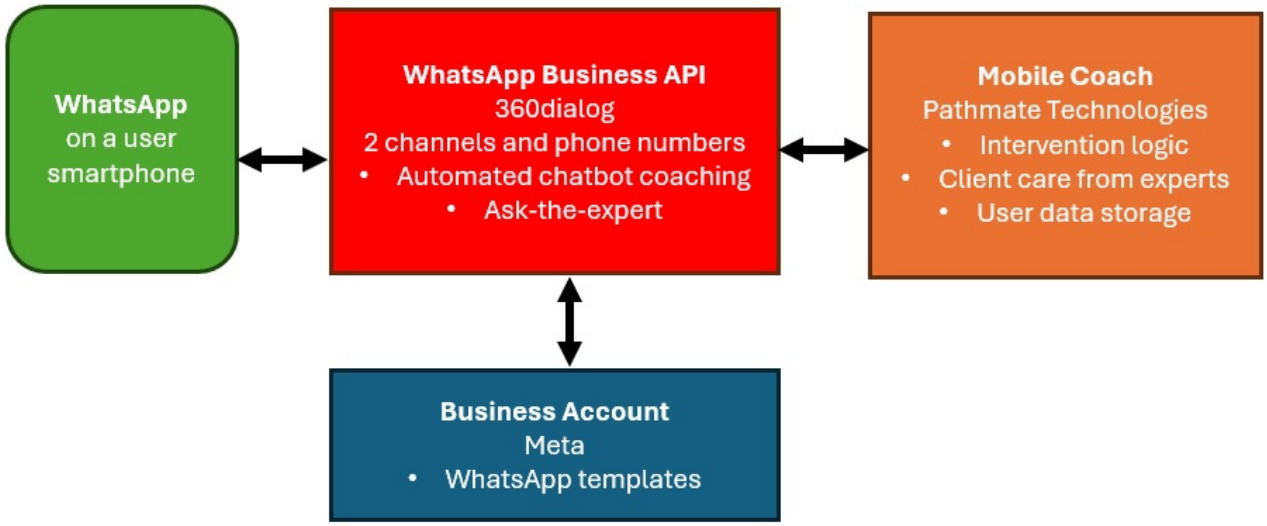
Technological Background

The rauchfreiCoach (English translation: SmokefreeCoach) intervention program was developed in 2022 and 2023 by the Swiss Research Institute for Public Health and Addiction (Institut für Sucht- und Gesundheitsforschung; ISGF) and PathMate Technologies as a German-language program for use in the German-speaking part of Switzerland. SmokefreeCoach is the world's first automated WhatsApp-based program to promote smoking and vaping cessation among adolescents and young adults.

The SmokeFreeCoach consists of 4 essential components; refer to Figure 1. The first component is a participant using a smartphone with WhatsApp installed. The second component is a business solution provider who provides a WhatsApp Business API (application programming interface). Facebook launched this API in 2018 but only for certified partners, which was 360dialog in our case. The third component is a business account with Meta by which one can create message templates. Those templates are needed to initiate a dialog with a user. The fourth component is a technological partner who realizes the actual intervention. In our case, this was PathMate Technologies, which ran the intervention logic, provided a web interface to chat with the participants, and stored the scientific data.

XSL•FO
RenderX

Figure 1. Software architecture of SmokefreeCoach. API: application programming interface.



Intervention Content

The program's target group was young people between the ages of 16 and 30 years who consumed tobacco cigarettes or electronic nicotine products at least once a week and used WhatsApp on their smartphones. The program was based on principles from Prochaska and DiClemente's transtheoretical model [25] and Bandura's social cognitive theory [26]. In addition, SmokefreeCoach applied various techniques that, according to recent literature reviews [18,27], were promising for increasing the use of digital interventions and promoting the desired behavioral changes. These include the individualization of content, individual support from a professional, regular coaching messages, ease of use and technical stability, consideration of personally relevant goals, regular monitoring of goal achievement, providing information on the health risks associated with smoking, weighing up the pros and cons of quitting smoking, and highlighting alternative behaviors. The program was available in 3 versions, whose contents differed depending on the individual's specified behavioral goal. Version 1 was designed for participants who neither wanted to quit nor wanted to smoke or vape less, version 2 for participants who wanted to reduce their consumption, and version 3 for participants who wanted to stop smoking or vaping. Participants had the possibility to modify their behavioral goal at any time during program runtime. For participants who regularly smoked tobacco cigarettes and vaped electronic nicotine products, the specified goal applied to both products.

Table 1 displays the intervention program elements according to the individual consumption goal. Due to the optional on-demand tips and the higher-intensity quit program, the intervention dose was very heterogeneous, ranging from 1 message per week for participants without a reduction or cessation goal and no use of on-demand tips to 3 messages per day for participants with a cessation goal and an activated higher-intensity quit program.

XSL•FO
RenderX

Table . Intervention program elements according to the selected consumption goal.

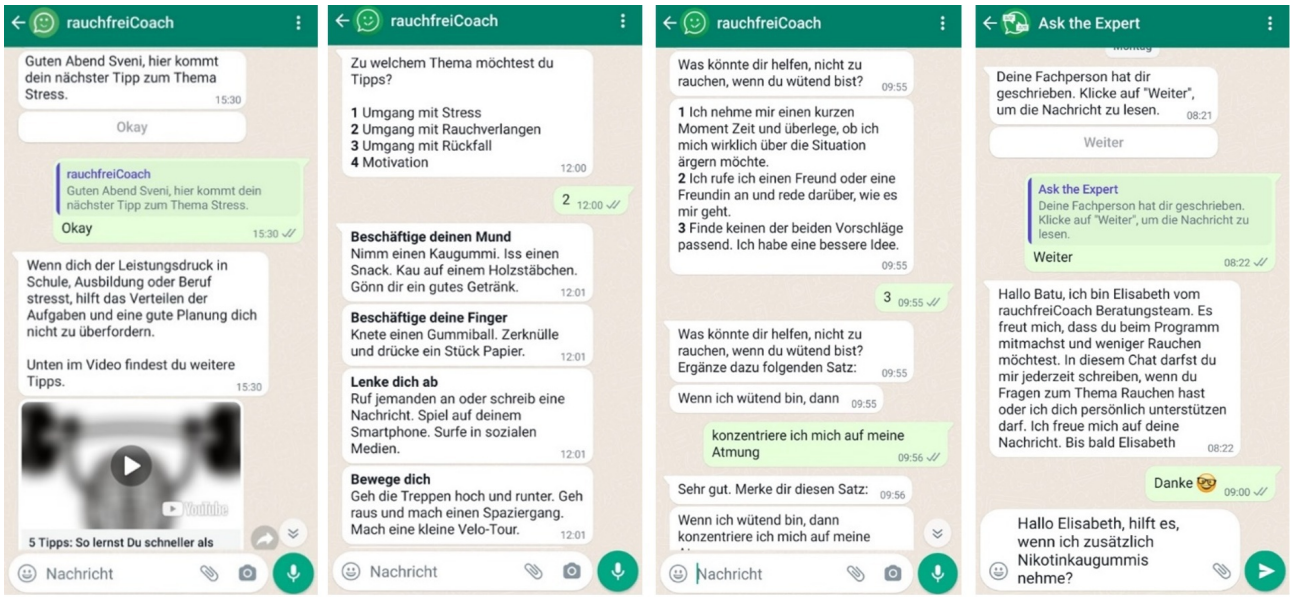
Intervention program element	Neither reduction nor cessation goal	Reduction goal	Cessation goal
Stress management tips tailored to individually indicated stressors	✓	✓	✓
Motivational messages tailored to individually indicated benefits of smoking or vaping cessation or reduction	✓	✓	✓
On-demand tips on motivation, stress, and craving	✓	✓	✓
Ask-the-Expert	✓	✓	✓
Strategies to deal with individually indicated barriers and obstacles during smoking or vaping reduction		✓	
Monitoring of goal achievement, that is, smoking or vaping cessation in self-selected trigger situations		✓	
If-then behavior plans and strategies to refrain from smoking or vaping in self-selected trigger situations		✓	
Strategies to deal with individually indicated barriers and obstacles during smoking or vaping cessation			✓
Monitoring of goal achievement, that is, the number of consumption days in the previous week			✓
If-then behavior plans and strategies to refrain from smoking or vaping in self-selected trigger situations			✓
Higher-intensity quit the program around the self-selected quit date			✓

Program Flow

All participants, regardless of their consumption goal, received 2 - 4 weekly WhatsApp messages that provided tips on how to deal with stress as well as motivational tips for a period of 11 weeks. Furthermore, all participants could access motivational tips and tips on dealing with stress, cravings, and relapse at any time. Through a separate WhatsApp channel, all participants also had the opportunity to ask individual questions to a smoking cessation expert anonymously and free of charge (Ask-the-Expert). This expert was a trained smoking cessation counselor from the Zurich Lung Association, that is, an external counselor who was not part of the research team. These individual questions were answered during working hours within a maximum of 2 working days. Figure 2 shows a selection of screenshots of the program (in German), which illustrate different types of conversation within the program.

XSL•FO
RenderX

Figure 2. Sample screenshots of the program (from left to right): (A) individual tip to cope with performance pressure, (B) on-demand tips to cope with craving for cigarettes, (C) if-then behavior plan on how not to smoke when feeling angry, and (D) Ask-the-Expert channel to pose individual questions to a smoking cessation expert.



In addition to the program elements already mentioned, participants with a reduction goal received weekly monitoring of goal achievement, that is, smoking or vaping cessation in certain situations, as well as WhatsApp messages that helped them with strategies to stop smoking or vaping in these selected situations. Participants who intended to quit received weekly monitoring of their goal achievement by assessing the consumption days in the previous week as well as messages helping them with strategies to stop smoking or vaping. Participants who were determined to give up smoking or vaping could also sign in to a higher intensity quit program. This optional component of the program started 7 days before an individually selected quit date and accompanied the participants with an additional 2 messages per day for cessation preparation and relapse prevention until 14 days after the self-selected quit date.

Study Design and Procedure

A longitudinal pre-post study design was used to test the feasibility, acceptance, and preliminary efficacy of the program. In order to identify small to medium effect sizes, we planned to recruit at least 100 people for this pre-post pilot study. Study participants were recruited between May 26, 2023, and March 12, 2024. The program and the associated website were advertised via various channels in the German-speaking part of Switzerland. These included unpaid posts on an Instagram (Meta Platforms, Inc) channel, an announcement in the Swiss free newspaper 20 Minuten, announcements in the mailing list of the University of Zurich, information for vocational school teachers and vocational trainers about the program, and a market stall at the Zurich University of Applied Sciences. In addition, the program was promoted through personal contacts with colleagues, family members, friends, and acquaintances of the Zurich Center for the Prevention of Substance Abuse (Zürcher Fachstelle zur Prävention des Suchtmittelmissbrauchs; ZFPS) and the ISGF.

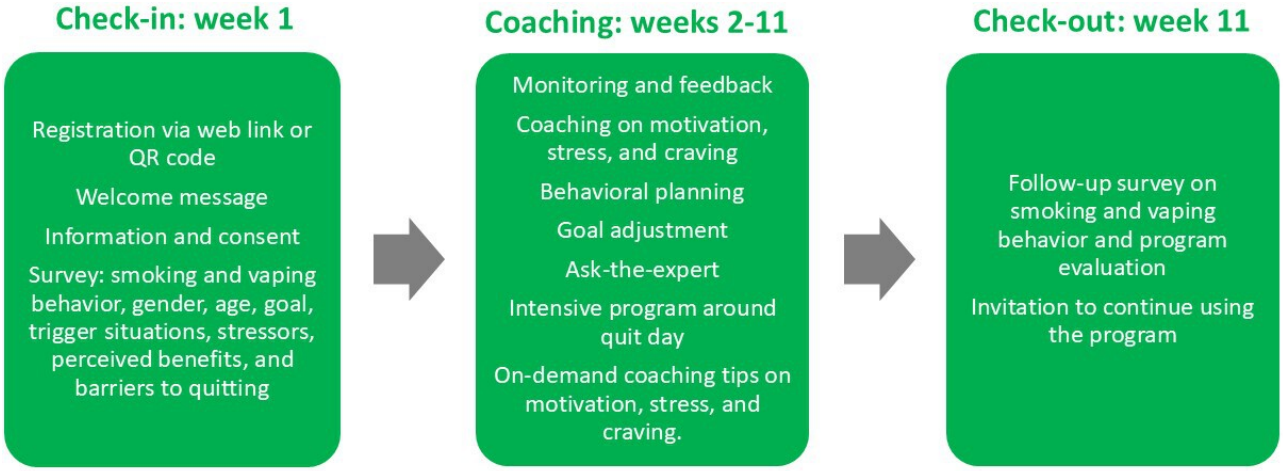
People interested in the program were asked to register using the sign-in button or QR code. Subsequently, the inclusion criteria for study participation (ages between 16 and 30 years and at least weekly consumption of tobacco or electronic nicotine products in the previous 4 weeks) were assessed and checked within a brief WhatsApp dialog. If the inclusion criteria were met, detailed information on the program, its aims, and the associated study was provided. After agreeing to participate in the program and the related study within the WhatsApp dialog via tickbox, participants were asked to select a username and to complete the baseline survey (refer to “Measures and Outcome Criteria” below).

The 10-week coaching always began on the first Monday after registration. During the period between registration and the actual start of coaching, which could last from 1 to 7 days, participants already had the opportunity to interact with the program. Therefore, the total duration of the program was 10 weeks plus 1 - 7 days.

At the end of week 11, participants were invited via the WhatsApp channel to complete an online follow-up survey (refer to “Measures and Outcome Criteria”). If they did not do so, they were contacted again via an SMS text message and finally by phone call. Following the follow-up survey, participants had the opportunity to continue using the program. Figure 3 illustrates and summarizes the course of the program and the study.

XSL•FO
RenderX

Figure 3. Procedure of the program and the study.



Measures and Outcome Criteria

The baseline survey included the following variables: (1) gender, (2) canton of residence, (3) recruitment channel, (4) the consumption goal (quitting, reducing, or neither quitting nor reducing), (5) emotional states that often tempt participants to smoke or vape, (6) situations in which participants smoke or vape frequently, (7) situations in which participants would like to stop smoking or vaping in the future, (8) current sources of stress and strain, (9) potential benefits of quitting or reducing nicotine use, and (10) participants' concerns about quitting or reducing. Items (5) to (10) were multiple-response items, derived from the current literature on nicotine and tobacco use in young people and consolidated within 2 focus groups with 8 and 7 adolescents and young adults, who regularly smoked or vaped nicotine products. Beyond consolidating the response items, the focus groups aimed at optimizing the program flow, the logo, and the integrated images and illustrations.

Variables that were assessed in both the baseline and follow-up survey and used as outcome criteria included (1) the number of days in the past 30 days smoking tobacco cigarettes, (2) the number of tobacco cigarettes smoked per smoking day, and (3) the number of days in the past 30 days using electronic nicotine products (eg, e-cigarettes, electronic shisha). By using (1) and (2), we calculated the average number of tobacco cigarettes consumed per day in the last 30 days. Participants who indicated that they did not smoke tobacco or electronic nicotine products in the past 30 days in the follow-up survey were considered as nonsmokers.

To assess program use and acceptance, we also assessed in the follow-up survey whether participants (1) had usually read the program messages thoroughly, (2) had usually only looked at the program messages briefly, or (3) had not read the program messages. Participants also had to indicate whether the content and tips of the program were (1) understandable, (2) helpful, (3) adapted to their personal situation, (4) motivating to smoke less or quit smoking, and (5) helpful to smoke less or quit smoking using the response options 1=thumbs up, 2=thumbs up and down, and 3=thumbs down. In addition, participants were asked to rate specific program elements (the form of coaching via WhatsApp, the tips received, the opportunity to ask questions to an expert, and the intensive quit program) using the response options, 1=thumbs up, 2=thumbs up and down, 3=thumbs down, and 4=can't judge, as well as to give the program an overall rating on a scale of 1-5 stars. Finally, participants were asked to indicate which functions or content of the program they found particularly interesting or helpful and to make suggestions for improving the program. The detailed baseline and follow-up survey questions are available in Multimedia Appendix 1.

In addition to the variables recorded in the baseline and follow-up surveys, the following indicators of program engagement were also extracted from the program log files for each participant: (1) whether or not the participant posed individual questions in the Ask-the-Expert channel, (2) the number of active weeks, that is, weeks during which at least 1 coaching message was retrieved or requested by a participant, (3) the total number of retrieved and requested coaching messages of a participant, and (4) the number of days from the start of the program to the last interaction with the program. Interaction with the program was defined as any response by the participant to the messages automatically sent by the program or any spontaneous request by the participant to the program.

Data Analyses

An attrition analysis was conducted to test for significant differences between participants who completed the follow-up survey and those who did not. Mann-Whitney *U* tests were used for non-normally distributed metric variables and chi-square tests for categorical variables.

We used generalized estimating equation (GEE) analyses to investigate the longitudinal course of the outcome criteria over the study period of 10 weeks. GEE is a repeated-measures regression model that takes into account the correlation of the repeated measures within a person [28]. It is a powerful and versatile procedure for analyzing longitudinal data under minimal assumptions about time dependence and allowed us to use all available longitudinal data, irrespective of single missing values at follow-up. We created separate GEE models for our outcomes: (1) number of days in the past 30 days smoking tobacco cigarettes, (2) number of days in the past 30 days using electronic nicotine products, and (3) average number of tobacco cigarettes consumed per day in the last 30 days. Each model

XSL•FO
RenderX

contained the time variable as an independent variable (baseline vs follow-up survey), one of the three variables associated with the consumption of nicotine-containing products as a dependent variable, and all variables that proved to be significant in the attrition analysis as covariates. In this way, it was possible to control for attrition bias. As all participants were smokers at baseline assessment, it was not possible to develop a converging GEE model for the 30-day point prevalence abstinence rate, and we only provided descriptive statistics for this outcome.

The open-text data on helpful program elements and suggestions for improving the program were analyzed and presented in a quantitative-descriptive manner. Based on qualitative content analysis [29], a PhD-level scientist determined the level of abstraction of the categories in the first step. In a second step, the responses were assigned to one or more of the available categories. An alpha level of .05 (2-tailed) was chosen for all statistical tests in this study. The analyses were performed using the Stata version 12 (StataCorp LLC) software package, or SPSS version 29 (IBM Corp).

Ethical Considerations

The study was approved by the Ethics Committee of the Faculty of Philosophy of the University of Zurich (approval date: December 20, 2021; approval number 21.12.15). Digital informed consent was obtained from the participants within the onboarding WhatsApp dialog. No additional parental consent was necessary for 16- and 17-year-olds. Study participants completed the screening procedure and baseline assessment anonymously. No names were collected, and participants instead chose nicknames. The business solution provider, which provided the API for WhatsApp communication to end users, ensured General Data Protection Regulation (GDPR) compliance. Data were transmitted with end-to-end encryption and were stored on a server in Switzerland. Participants received an incentive of 10 Swiss Francs (around US $12) for taking part in the follow-up survey.

Results

Study Participants

A total of 167 young people (mean age 23.2, SD 4.0 years; n=95; 56.9% women, n=72, 43.1% men) who regularly smoked tobacco cigarettes (n=81, 48.5%), vaped electronic nicotine products (n=17, 10.2%), or both (n=69, 41.3%) could be recruited for participation in the program and the associated study. Of these, 59.9% (n=100) intended to quit, while 40.1% (n=67) aimed at reducing their use of cigarettes or electronic nicotine products. None of the participants stated that they wanted neither to stop nor to reduce. Table 2 presents the baseline characteristics of the program participants.

XSL•FO
RenderX

Table . : Baseline characteristics of program participants (N=167).

Characteristic	Value
Recruitment channel (multiple answers), n (%)	
Social media	10 (6)
Advertising in a commuter newspaper	12 (7.2)
Friends or acquaintances	35 (21)
School or University	93 (55.7)
Work or vocational trainer	20 (12)
Gender, n (%)	
Men	72 (43.1)
Women	95 (56.9)
Diverse	0 (0)
Age (years), mean (SD)	23.2 (4.0)
Canton of residence, n (%)	
Zurich	99 (59.3)
Aargau	17 (10.2)
St. Gallen	11 (6.6)
Grisons	7 (4.2)
Other canton	26 (15.6)
Other country	7 (4.2)
Products consumed, n (%)	
Solely tobacco cigarettes	81 (48.5)
Solely electronic nicotine products	17 (10.2)
Both tobacco cigarettes and electronic nicotine products	69 (41.3)
Behavioral goal, n (%)	
Stop smoking or vaping	100 (59.9)
Reduce smoking or vaping	67 (40.1)
Neither stop nor reduce	0 (0)
Number of days in the last 30 days on which tobacco cigarettes were smoked, mean (SD)	20.8 (11.6)
Average number of tobacco cigarettes smoked per day, mean (SD)	7.4 (7.0)
Number of days in the last 30 days on which electronic nicotine products were consumed, mean (SD)	10.0 (12.7)

After week 11, 108/167 (64.7%) program participants completed the follow-up survey. The attrition analyses revealed that follow-up responders did not differ from nonresponders with respect to any of the baseline characteristics. However, nonresponders were active during fewer program weeks (mean 3.6, SD 2.6 vs 6.5, SD 3.6; z=–5.3; P<.001), had fewer interactions with the program (mean 15.7, SD 20.1 vs 32.9, SD 27.0; z=–4.8; P<.001) and showed a lower number of days from program start to last interaction with the program (mean 27.6, SD 27.7 vs 54.5, SD 28.8; z=4.9; P<.001). These engagement variables were included as covariates in the GEE models investigating the longitudinal course of the outcome criteria.

Acceptance of the Intervention Program

Program Attrition

During the program, which lasted for 11 weeks, none of the 167 participants actively unsubscribed from the program or withdrew from study participation.

Program Use

Among the 167 program participants, the average number of active weeks was 5.5 (SD 3.5). A total of 16 (9.6%) participants were inactive in any program weeks, while 26 (15.6%) participants were active in all program weeks. In addition, 80 (47.9%) participants were active in more than half, that is, at least 6, of the 11 program weeks.

XSL•FO
RenderX

The average number of interactions with the program was 26.8 (SD 26.1), and the mean duration from the start of the program to the last interaction was 45.0 (SD 31.1) days.

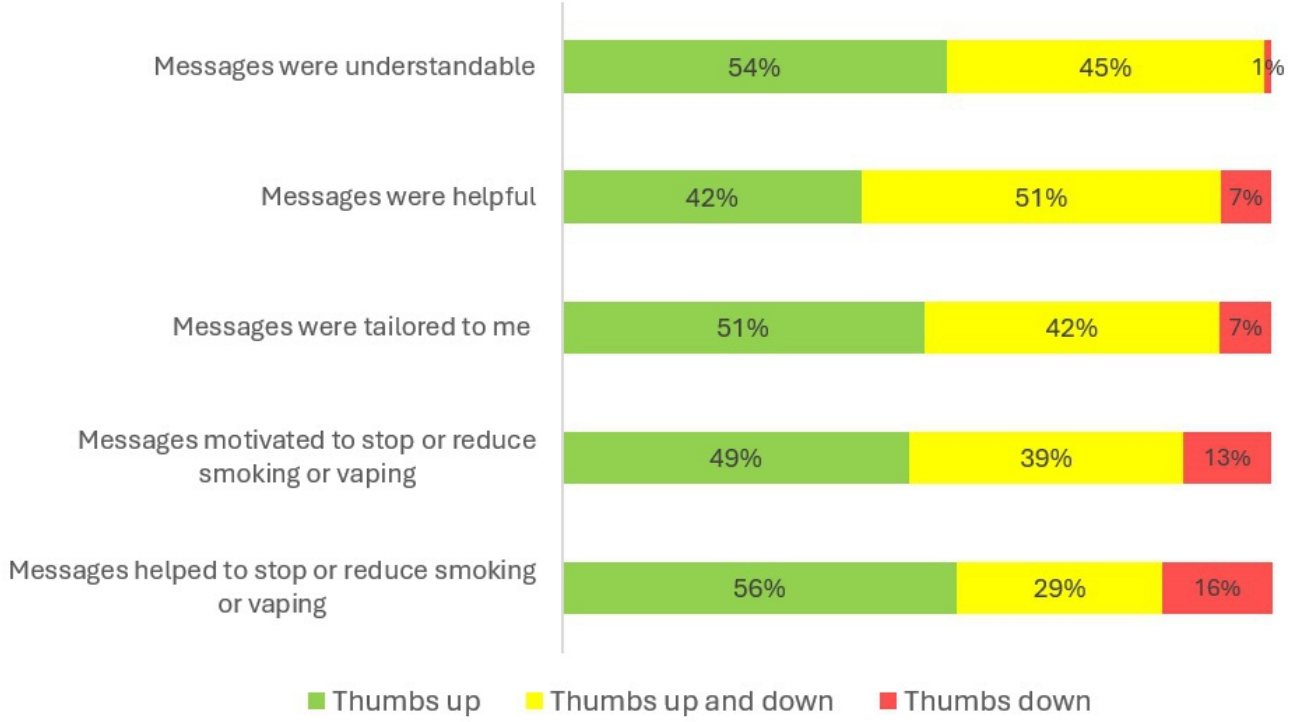
In the follow-up survey, 38/108 (35.2%) participants stated that they had usually read the program messages thoroughly, 58 (53.7%) participants stated that they usually looked at them briefly, and 12 (11.1%) participants stated that they usually had not read the program messages.

In total, 8/167 (4.8%) program participants used the Ask-the-Expert channel to pose individual questions to the smoking cessation expert on the consumption of nicotine-containing products.

Program Evaluation

Data on the evaluation of the program messages were available for 96/108 follow-up survey respondents. This is due to the fact that participants who reported that they usually had not read the program messages were not asked to rate them. Of the 96 participants, 52 (54%) rated the comprehensibility of the content positively, 40 (42%) stated that the program messages were helpful, 49 (51%) stated that the tips and information contained in the program messages were tailored to their personal situation, 47 (49%) stated that the program messages motivated them to smoke or vape less or to quit, and 54 (56%) stated that the program messages helped them to smoke or vape less or to quit. The results on the evaluation of the program messages are also displayed in Figure 4.

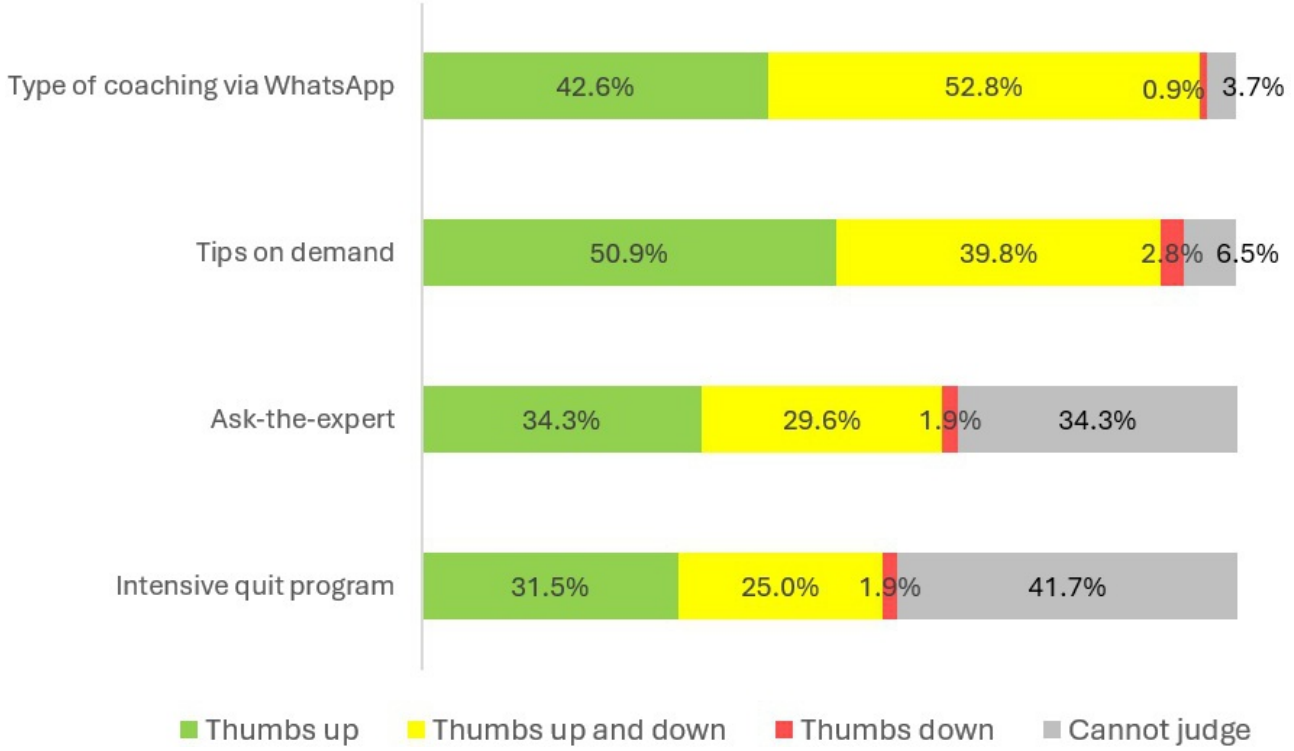
Figure 4. Evaluation of the coaching messages of the SmokeFreeCoach program (n=96).



Concerning the evaluation of specific program elements, we had data from all participants who completed the follow-up survey (n=108). The form of coaching via WhatsApp was rated positively by 46 (42.6%) participants, and 55 (50.9%) participants liked the available tips on demand. Overall, 36 (34.3%) participants rated the option to ask questions to an expert positively, and 34 (31.5%) participants rated the intensive quit program positively. The percentages of respondents who rated these program elements negatively were very low, ranging from 0.9% to 2.8%. The results concerning the evaluation of specific program elements are also shown in Figure 5.

XSL•FO
RenderX

Figure 5. Evaluation of specific program elements of the SmokeFreeCoach (n=108).



The 108 postsurvey participants gave the program an overall rating on a scale of 1-5 stars. The mean program rating was 3.5 (SD 0.8) stars.

We received valid open-ended responses from 65 participants to the question "What content or functions were particularly interesting or helpful for you?" in the follow-up survey. A total of 5 participants stated that all of the program's content and functions were interesting and helpful to them. The responses of 17 participants indicated that the tips sent by the program and the tips that participants could access themselves, such as tips on dealing with stress and distraction tips, were particularly interesting and useful. The fact that the SmokeFreeCoach is a WhatsApp-based program and easily accessible was appreciated by 9 participants, while 4 participants praised the ability to ask questions to experts. A total of 5 participants found the planning of individual strategies helpful. Altogether, 8 participants found the content of the program as well as the way the content was delivered (eg, videos, links, and statistics) helpful and interesting. Finally, 21 participants expressed their appreciation for the regular messages and reminders, which enabled them to constantly engage with the topic.

For the question "How could we improve the program?" in the follow-up survey, we received open-ended responses from 45 participants, of which 3 indicated having technical difficulties during the course of the intervention. Of the people making specific suggestions, 11 wanted the program to be better tailored to their individual life situation. Deepening or expanding the topics covered (eg, by suggesting more strategies for distraction during moments of craving, or by providing more information about the consequences of consuming nicotine or tobacco and also about other products, such as smokeless tobacco products) was suggested by 19 respondents. Four articipants expressed the wish that they could choose for themselves how many messages one gets and that the statements should be shortened in length. The need to avoid repetition of messages and content was also mentioned by 4 respondents. In addition, 3 participants specifically asked for a monitoring system with which they could track their daily cigarette use and earn credits. Finally, 1 participant reported having to click on links within the program to receive additional information as a barrier. Another one would have liked the expert to be available more often.

Preliminary Program Efficacy

Days With Tobacco Cigarette Consumption

The GEE analyses revealed a statistically significant decrease in the number of days in the last 30 days on which tobacco cigarettes were smoked between baseline and the follow-up survey (incidence rate ratio [IRR] 0.68, $P<.001$). Considering only the data from participants who completed the follow-up survey, the average number of days in the last 30 days on which tobacco cigarettes were smoked decreased by 6.6 days from 20.6 (mean, SD 11.8) at baseline to 14.0 (mean, SD 12.0) at follow-up (Table 3).

XSL•FO
RenderX

Table . Pre-post comparison of indicators of nicotine use. The first 3 columns show the pre-post comparisons of the mean values, whereby only the data of the participants in the follow-up survey (n=108) were taken into account. The last 2 columns summarize the main results of the generalized estimating equation (GEE) analyses.

Indicator of nicotine use	Baseline mean (SD)	Follow-up mean (SD)	Mean value difference	IRR[a] (95% CI)	*P* value
Number of days of tobacco cigarette consumption in the last 30 days	20.6 (11.8)	14.0 (12.0)	–6.6	0.68 (0.59-0.77)	<.001
Average number of tobacco cigarettes consumed per day in the last 30 days	7.2 (6.8)	4.1 (6.4)	–3.1	0.57 (0.45-0.73)	<.001
Number of days using electronic nicotine products in the last 30 days	11.1 (13.1)	7.7 (11.3)	–3.4	0.71 (0.55-0.90)	.005

[a]IRR: incidence rate ratio.

Quantity of Tobacco Cigarettes Smoked Per Day

Concerning the average number of tobacco cigarettes consumed per day in the last 30 days, the GEE analyses revealed a statistically significant decrease between the baseline and the follow-up survey (IRR 0.57, 95% CI 0.45-0.73; *P*<.001). Considering only the data from participants who completed the follow-up survey, the average number of tobacco cigarettes consumed per day in the last 30 days decreased by 3.1 cigarettes per day from 7.2 (mean, SD 6.8) at baseline to 4.1 (mean, SD 6.4) at follow-up (Table 3).

Days With Electronic Cigarette Consumption

Concerning the number of days using electronic nicotine products in the last 30 days, the GEE analyses also revealed a statistically significant decrease between the baseline and the follow-up survey (IRR 0.71, 95% CI 0.55-0.90; *P*<.005). Considering only the data from participants who completed the follow-up survey, the average number of days on which electronic nicotine products were used decreased by 3.4 days from 11.1 (mean, SD 13.1) days at baseline to 7.7 (mean, SD 11.3) days at follow-up (Table 3).

30-Day Point Prevalence Abstinence

A total of 6/108 (5.6%) participants in the follow-up survey stated that they had not consumed either conventional cigarettes or electronic nicotine products in the last 30 days and could be considered as abstinent for the last 30 days at follow-up.

Discussion

Principal Findings

The study revealed three main findings: (1) the automated provision of individually tailored messages to reduce or stop smoking or vaping via WhatsApp was feasible, (2) the evaluation and use of the program in the target group of young people between the ages of 16 and 30 years was acceptable, and (3) the program may support cessation and reduction of tobacco cigarettes and electronic nicotine products.

The overall use and evaluation of the intervention program was acceptable. None of the 167 participants actively unsubscribed from the program or withdrew from study participation. On average, participants actively interacted in 5.5 (SD 3.5) of 11 program weeks with an average of 26.8 (SD 26.1) interactions over a period of 45.0 (SD 31.1) days. A comparison of these results with findings on program engagement in other digital programs developed specifically for young people shows that participation in the SmokefreeCoach program is comparable or even higher. For example, young people participating in a text messaging–based smoking cessation program in Switzerland responded to an average of 6.5 (SD 3.7) of 11 SMS text messaging questions asked [30]. Within a multiple behavior prevention program provided via smartphone app and aimed at Swiss adolescents, the average number of dialogs completed with a conversational agent was 2.1 (SD 3.5) out of 16 [31]. In an app-based alcohol prevention program for young people in Switzerland, an average of 3.8 (SD 3.6) out of 10 dialogs with a conversational agent were completed, and the average duration of program use was 38.3 (SD 52.1) days [32]. Adolescents who registered for the Australian health promotion app Health4life were active for an average of 9.7 (SD 18.8) days [33].

The vast majority of respondents expressed positive or at least neutral evaluations of the program, the program messages, and the program elements. However, the answers to the question “How can we improve the program?” indicate that there is room for improvement. Among the respondents’ suggestions for improving the program, better tailoring of the program content to the respective person, and expanding the program to include information on untreated topics, such as substitution of tobacco cigarettes with other nicotine-containing products and on nicotine products that are not included, such as smokeless tobacco products, stood out in particular.

Although the results on the preliminary efficacy of the intervention program are derived from a pre-post study, they suggest that the SmokeFreeCoach could be a promising tool for supporting smoking and vaping cessation among adolescents and young adults. Throughout the study, a significant reduction

XSL•FO
RenderX

was observed in 3 indicators of cigarette and nicotine consumption, and 6 participants stopped smoking.

As there are hardly any low-threshold evidence-based programs to support young people in quitting smoking or vaping [15], it seems sensible to pursue this intervention approach further on the basis of the results obtained. Although text messaging–based interventions are among the most promising digital smoking and vaping cessation interventions for young people [15], it is questionable whether such good results can be achieved in the future as, for example, in the study from Graham et al [34] or Ybarra et al [35], in view of the increasingly reduced use and limited possibilities of this channel. Communication via instant messenger apps like WhatsApp or Viber (Rakuten) has largely replaced SMS text messaging and offers more options for integrating media content and also for combining personal and automated modes of communication.

Compared to cessation apps, for example, [36], which have to be installed and opened separately, messenger apps offer the advantage that they are already available and are used regularly. Therefore, their access is likely to be lower-threshold and program use higher, which is indicated by the results of this study. On the other hand, the now numerous features of WhatsApp, such as “channels,” “communities,” or the recently integrated “Meta AI,” also pose the risk that messages are more likely to be ignored within WhatsApp intervention programs than in app-based applications developed solely for this purpose.

Limitations

This study has several limitations:

1. The longitudinal one-group pre-post design does not allow sound conclusions on the effectiveness of the program.
2. The reported outcomes on tobacco smoking and nicotine use are based on self-reports and were not verified using biochemical parameters. However, numerous studies have shown that self-reports on nicotine consumption are mostly valid; especially when delivered online and anonymously, they are not subject to strong bias [37].
3. The results of the attrition analysis reveal significant differences in the use of the program between those who participated in the follow-up survey and those who did not. Therefore, a bias toward those who found the program effective and acceptable is likely.

Outlook

In view of the promising results of this study, it seems crucial to conduct a randomized controlled trial in order to draw sound conclusions on the effectiveness of the program. Based on the suggestions for improvement made by the participants in this study and the increasing popularity of smokeless nicotine products, such as snus or nicotine pouches among young people [38], it seems reasonable to expand the program to include coaching content also addressing this variant of nicotine use.

Conclusions

This is the first study that investigated the acceptability and preliminary efficacy of an automated WhatsApp-based program to support smoking and vaping cessation. Its results suggest that the automated provision of individually tailored coaching messages via WhatsApp is feasible and the SmokefreeCoach program was modestly accepted by the target group of young people. The preliminary results on program efficacy suggest that the program could provide a low-cost and scalable measure to decrease the prevalence rate of nicotine use. Overall, the results of this study, the promising efficacy of SMS text messaging and WeChat-based smoking cessation programs [17,23,37], and the popularity of instant messaging apps [24] suggest that WhatsApp-based interventions are a promising means of helping young people to quit smoking and vaping.

Acknowledgments

This study was funded by Health Promotion and Prevention in the Canton of Zurich as part of the Cantonal Tobacco Prevention program. The tobacco industry did not have any influence on the funding or the conduct of the study. We would like to thank Dominique Lorandt and her team from the Zurich Specialist Center for the Prevention of Substance Abuse (Zürcher Fachstelle zur Prävention des Suchtmittelmissbrauchs; ZFPS) and Sabine Jenny from the Epidemiology Biostatistics and Prevention Institute (EBPI) for supporting the recruitment. Furthermore, we would like to thank Andreas Filler and his team from PathMate Technologies for the technical development of the program.

Conflicts of Interest

Single authors (SH, AW, OS) were also involved in the development of the intervention program.

Multimedia Appendix 1
Baseline- and follow-up assessments.
[DOCX File, 72 KB - mhealth_v13i1e65301_app1.docx]

References

1. GBD 2019 Tobacco Collaborators. Study, spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet 2021;397(10292):2337-2360. [doi: 10.1016/S0140-6736(21)01169-7]
2. Jakob J, Cornuz J, Diethelm P. Prevalence of tobacco smoking in Switzerland: do reported numbers underestimate reality? Swiss Med Wkly 2017;147:w14437. [doi: 10.4414/smw.2017.14437] [Medline: 28634966]

XSL•FO
RenderX

3. West R. Tobacco smoking: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):1018-1036. [doi: 10.1080/08870446.2017.1325890] [Medline: 28553727]
4. Tobacco consumption (age: 15+). Swiss Monitoring System of Addiction and Noncommunicable Diseases. 2025. URL: https://ind.obsan.admin.ch/en/indicator/monam/tobacco-consumption-age-15 [accessed 2025-04-22]
5. Pirie K, Peto R, Reeves GK, Green J, Beral V, Million Women Study Collaborators. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. Lancet 2013 Jan 12;381(9861):133-141. [doi: 10.1016/S0140-6736(12)61720-6] [Medline: 23107252]
6. Consumption of e-cigarettes (age: 15+) [Web page in German]. Swiss Monitoring System of Addiction and Noncommunicable Diseases. URL: https://ind.obsan.admin.ch/indicator/monam/konsum-von-e-zigaretten-alter-15 [accessed 2025-04-22]
7. Lindson N, Butler AR, McRobbie H, et al. Electronic cigarettes for smoking cessation. Cochrane Database Syst Rev 2024 Jan 8;1(1):CD010216. [doi: 10.1002/14651858.CD010216.pub8] [Medline: 38189560]
8. Hartmann-Boyce J, Butler AR, Theodoulou A, et al. Biomarkers of potential harm in people switching from smoking tobacco to exclusive e-cigarette use, dual use or abstinence: secondary analysis of Cochrane systematic review of trials of e-cigarettes for smoking cessation. Addiction 2023 Mar;118(3):539-545. [doi: 10.1111/add.16063] [Medline: 36208090]
9. Hammond D, Reid JL, Goniewicz ML, et al. Nicotine exposure from smoking tobacco and vaping among adolescents. JAMA Netw Open 2025 Mar 3;8(3):e2462544. [doi: 10.1001/jamanetworkopen.2024.62544] [Medline: 40072439]
10. Anderer S. Rise in vaping keeps tobacco use high. JAMA 2025 May 13;333(18):1570. [doi: 10.1001/jama.2025.4024]
11. Banks E, Yazidjoglou A, Brown S, et al. Electronic cigarettes and health outcomes: umbrella and systematic review of the global evidence. Med J Aust 2023 Apr 3;218(6):267-275. [doi: 10.5694/mja2.51890] [Medline: 36939271]
12. Hartmann-Boyce J, Chepkin SC, Ye W, Bullen C, Lancaster T. Nicotine replacement therapy versus control for smoking cessation. Cochrane Database Syst Rev 2018 May 31;5(5):CD000146. [doi: 10.1002/14651858.CD000146.pub5] [Medline: 29852054]
13. Hartmann-Boyce J, Livingstone-Banks J, Ordóñez-Mena JM, et al. Behavioural interventions for smoking cessation: an overview and network meta-analysis. Cochrane Database Syst Rev 2021 Jan 4;1(1):CD013229. [doi: 10.1002/14651858.CD013229.pub2] [Medline: 33411338]
14. Wu AD, Lindson N, Hartmann-Boyce J, et al. Smoking cessation for secondary prevention of cardiovascular disease. Cochrane Database Syst Rev 2022 Aug 8;8(8):CD014936. [doi: 10.1002/14651858.CD014936.pub2] [Medline: 35938889]
15. Park E, Zhou Y, Chen C, Chacko T, Mahoney M, Chang YP. Systematic review: interventions to quit tobacco products for young adults. BMC Public Health 2023 Jun 26;23(1):1233. [doi: 10.1186/s12889-023-15900-8] [Medline: 37365562]
16. Thrul J, Stemmler M, Goecke M, Bühler A. Are you in or out? Recruitment of adolescent smokers into a behavioral smoking cessation intervention. Addict Behav 2015 Jun;45:150-155. [doi: 10.1016/j.addbeh.2015.01.030] [Medline: 25678303]
17. Zhou X, Wei X, Cheng A, et al. Mobile phone-based interventions for smoking cessation among young people: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Sep 12;11:e48253. [doi: 10.2196/48253] [Medline: 37706482]
18. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
19. Guo YQ, Chen Y, Dabbs AD, Wu Y. The effectiveness of smartphone app-based interventions for assisting smoking cessation: systematic review and meta-analysis. J Med Internet Res 2023 Apr 20;25:e43242. [doi: 10.2196/43242] [Medline: 37079352]
20. Luo T, Li MS, Williams D, et al. Using social media for smoking cessation interventions: a systematic review. Perspect Public Health 2021 Jan;141(1):50-63. [doi: 10.1177/1757913920906845] [Medline: 32077368]
21. Sweileh WM. Technology-based interventions for tobacco smoking prevention and treatment: a 20-year bibliometric analysis (2003-2022). Subst Abuse Treat Prev Policy 2024 Feb 6;19(1):13. [doi: 10.1186/s13011-024-00595-w] [Medline: 38321493]
22. Ramo DE, Thrul J, Delucchi KL, et al. A randomized controlled evaluation of the tobacco status project, a Facebook intervention for young adults. Addiction 2018 May 24;113:1683-1695. [doi: 10.1111/add.14245] [Medline: 29797621]
23. Tang J, Yang J, Liu Y, et al. Efficacy of WeChat-based online smoking cessation intervention ('WeChat WeQuit') in China: a randomised controlled trial. EClinicalMedicine 2023 Jun;60:102009. [doi: 10.1016/j.eclinm.2023.102009] [Medline: 37251625]
24. Külling C, et al. JAMES: jugend, aktivitäten, medien - erhebung schweiz [Report in German]. : Zürcher Hochschule für angewandte Wissenschaften; 2022 URL: https://www.zhaw.ch/storage/psychologie/upload/forschung/medienpsychologie/james/2018/Bericht_JAMES_2022_de.pdf [accessed 2025-09-24]
25. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
26. McAlister AL, Perry CL, Parcel GS. How individuals, environments, and health behaviour interact: social cognitive theory. In: Practice KG, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research: Jossey-Bass; 2008:169-188.

XSL•FO
RenderX

27. Mair JL, Salamanca-Sanabria A, Augsburger M, et al. Effective behavior change techniques in digital health interventions for the prevention or management of noncommunicable diseases: an umbrella review. Ann Behav Med 2023 Sep 13;57(10):817-835. [doi: 10.1093/abm/kaad041] [Medline: 37625030]
28. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988 Dec;44(4):1049-1060. [Medline: 3233245]
29. Mayring P. Qualitative Content Analysis: Fundamentals and Techniques: Weinheim: Beltz; 2010.
30. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013 Aug 16;15(8):e171. [doi: 10.2196/jmir.2636] [Medline: 23956024]
31. Haug S, Boumparis N, Wenger A, Schaub MP, Paz Castro R. Efficacy of a mobile app-based coaching program for addiction prevention among apprentices: a cluster-randomized controlled. Int J Environ Res Public Health 2022 Nov 25;19(23):15730. [doi: 10.3390/ijerph192315730] [Medline: 36497804]
32. Haug S, Boumparis N, Wenger A, Paz Castro R, Schaub MP. Mobile app-based coaching for alcohol prevention among adolescents: pre-post study on the acceptance and effectiveness of the program “MobileCoach Alcohol”. Int J Environ Res Public Health 2023 Feb 13;20(4):3263. [doi: 10.3390/ijerph20043263] [Medline: 36833958]
33. Thornton L, Brown HM, Osman B, et al. Factors associated with adolescents’ engagement with a Healthy Lifestyles app. Procedia Comput Sci 2022;206:56-67. [doi: 10.1016/j.procs.2022.09.085]
34. Graham AL, Amato MS, Cha S, Jacobs MA, Bottcher MM, Papandonatos GD. Effectiveness of a vaping cessation text message program among young adult e-Cigarette users: a randomized clinical trial. JAMA Intern Med 2021 Jul 1;181(7):923-930. [doi: 10.1001/jamainternmed.2021.1793] [Medline: 33999133]
35. Ybarra ML, Holtrop JS, Prescott TL, Rahbar MH, Strong D. Pilot RCT results of Stop My Smoking USA: a text messaging–based smoking cessation program for young adults. Nicotine Tob Res 2013 Aug;15(8):1388-1399. [doi: 10.1093/ntr/nts339] [Medline: 23348969]
36. Baskerville NB, Struik LL, Guindon GE, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893. [doi: 10.2196/10893] [Medline: 30355563]
37. Fang YE, Zhang Z, Wang R, et al. Effectiveness of eHealth smoking cessation interventions: systematic review and meta-analysis. J Med Internet Res 2023 Jul 28;25:e45111. [doi: 10.2196/45111] [Medline: 37505802]
38. Patel M, Kierstead EC, Kreslake J, Schillo BA. Patterns of oral nicotine pouch use among U.S. adolescents and young adults. Prev Med Rep 2023 Aug;34:102239. [doi: 10.1016/j.pmedr.2023.102239] [Medline: 37228836]

Abbreviations

API: application programming interface
e-cigarette: electronic cigarette
GDPR: General Data Protection Regulation
GEE: generalized estimating equation
IRR: incidence rate ratio
ISGF: Institut für Sucht- und Gesundheitsforschung
ZFPS: Zürcher Fachstelle zur Prävention des Suchtmittelmissbrauchs

Edited by L Buis; submitted 14.08.24; peer-reviewed by EA Vogel, L Egan; revised version received 15.09.25; accepted 15.09.25; published 02.10.25.

Please cite as:
Haug S, Caputo L, Wenger A, Kiselev N, Studhalter O, Schaub MP
WhatsApp-Based Coaching Program to Support Smoking and Vaping Cessation Among Young People: Pre-Post Study on Acceptance and Preliminary Efficacy
JMIR Mhealth Uhealth 2025;13:e65301
URL: https://mhealth.jmir.org/2025/1/e65301
doi:10.2196/65301

© Severin Haug, Lisa Caputo, Andreas Wenger, Nikolai Kiselev, Olivia Studhalter, Michael P Schaub. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is

XSL•FO
RenderX

properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

mHealth-Based Diabetes Prevention Program for Chinese Mothers With Abdominal Obesity: Randomized Controlled Trial

Qinyuan Huang[1,2*], MSN; Qinyi Zhong[1,3*], MSN; Yanjing Zeng[1], BSN; Yimeng Li[1], BSN; James Wiley[4], PhD; Man Ping Wang[5], PhD; Jyu-Lin Chen[6], PhD; Jia Guo[1], PhD

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

[2]Department of nursing, The 921st Hospital of Chinese People's Liberation Army (Second Affiliated Hospital of Hunan Normal University), Changsha, China

[3]Manchester Centre for Health Psychology, University of Manchester, Manchester, United Kingdom

[4]University of California, San Francisco, San Francisco, CA, United States

[5]School of Nursing, University of Hong Kong, Hong Kong, China (Hong Kong)

[6]School of Nursing, University of California, San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Jia Guo, PhD
Xiangya School of Nursing
Central South University
No.172 Tongzipo Road, Yuelu District
Changsha, 410013
China
Phone: 86 13875947418
Email: guojia621@163.com

Abstract

Background: Among people with abdominal obesity, women are more likely to develop diabetes than men. Mobile health (mHealth)–based technologies provide the flexibility and resource-saving opportunities to improve lifestyles in an individualized way. However, mHealth-based diabetes prevention programs tailored for busy mothers with abdominal obesity have not been reported yet.

Objective: The aim of this study is to evaluate the feasibility and acceptability of an mHealth-based diabetes prevention program and its preliminary efficacy in reducing weight-related variables, behavioral variables, psychological variables, and diabetes risk among Chinese mothers with abdominal obesity over 6 months.

Methods: A randomized controlled trial was conducted at health management centers in 2 tertiary hospitals in Changsha, China. The mHealth group (n=40) received 12 weekly web-based lifestyle modification modules for diabetes prevention, 6 biweekly individualized health education messages based on their goal settings, and a Fitbit tracker. The control group (n=40) received 12 weekly web-based general health education modules, 6 biweekly general health education messages, and a Fitbit tracker. Data were collected at baseline, 3 months, and 6 months on the feasibility and acceptability outcomes, weight-related variables (waist circumference and BMI), diabetes risk scores, glycemic levels, behavioral variables (daily step count, active minutes, fruit and vegetable intake, calorie consumption, and sleep duration), and psychological variables (self-efficacy and social support for physical activity and diet, perceived stress, and quality of life). Generalized estimating equations were used for data analysis.

Results: Approximately 85% (68/80) of the participants completed 6 months of follow-up assessments. Regarding the feasibility and acceptance of the program in the mHealth group, the average number of modules reviewed was 7.9 out of 12, and the satisfaction score was 4.37 out of 5. Significant improvements at 6 months between the intervention and control groups were found in waist circumference (β=–2.24, 95% CI –4.12 to –0.36; *P*=.02), modifiable diabetes risk scores (β=–2.5, 95% CI –4.57 to –0.44; *P*=.02), daily steps (β=1.67, 95% CI 0.06-3.29; *P*=.04), self-efficacy for physical activity (β=1.93, 95% CI 0.44-3.43; *P*=.01), social support for physical activity (β=2.27, 95% CI 0.80-3.74; *P*=.002), and physical health satisfaction (β=0.82, 95% CI 0.08-1.55; *P*=.03). No differences were found in BMI, total diabetes risk score, daily active minutes, daily intake of fruits and vegetables, sleep duration, daily calorie consumption, self-efficacy, and social support for diet (*P*>.05).

XSL•FO
RenderX

Conclusions: This study addresses the potential role of tailored lifestyle interventions based on mHealth technology by offering tailored web-based health modules and health information in managing diabetes risk among mothers with abdominal obesity. The mHealth diabetes prevention program provides a flexible, customized, and resource-saving model for busy mothers. Future research could further explore the efficacy improvement on dietary behaviors to better serve the health care needs of this population.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400090554; https://www.chictr.org.cn/showproj.html?proj=226411

(JMIR Mhealth Uhealth 2025;13:e47837) doi:10.2196/47837

KEYWORDS

type 2 diabetes; mHealth; obesity; prevention; mothers

Introduction

Diabetes is the leading cause of death [1], accounting for 12.2% of all deaths in 2021 [2]. Approximately 537 million people have reported diabetes worldwide. There are complicated and various risk factors for developing diabetes or prediabetes, including aging, family history, unhealthy lifestyle, gut metagenome, and being overweight or obese [3,4]. By far, the strongest risk factor is being obese [5]. Moreover, compared to peripheral obesity, abdominal obesity is more closely associated with insulin resistance and metabolic disorders because the metabolic activity of visceral fat located in the abdomen is much higher than that of subcutaneous fat [6]. The waist circumference (WC) reflects the degree of abdominal fat accumulation in individuals [7]. Therefore, the reduction of WC plays a role in preventing and delaying the onset and progression of diabetes.

China has the largest diabetic population in the world [8]. According to a recent national study in 2020, a substantial number of new-onset type 2 diabetes mellitus cases in Chinese adults can be attributed to abdominal obesity, accounting for 50.4% in women and 30.3% in men [9]. Mothers rarely had free time to participate in face-to-face health promotion programs [10], albeit abdominal obesity is highly prevalent among middle-aged mothers with a normal BMI [11,12]. How to prevent diabetes among abdominally obese people has become an increasingly critical question in China, especially among mothers.

A healthy lifestyle, including low-calorie diets, high levels of physical activity, and self-management to lose weight or WC, benefits high-risk groups with diabetes. It is an effective approach to diabetes prevention [13]. For women with abdominal obesity, 2 lifestyle modification programs in the form of face-to-face diet and exercise coaching were effective in reducing their risk of developing diabetes [14,15]. However, interventions via face-to-face approaches require a large number of highly qualified health professionals to achieve the desired results [10]. The shortage of primary health care professionals and the strain on medical resources in low-income countries make it difficult to implement such interventions on a large scale for people at risk for diabetes [16]. Therefore, there is a great need to explore more flexible and resource-saving lifestyle modification interventions.

Mobile health (mHealth)–based and tailored health promotion programs via the WeChat platform have been widely used in diabetes and obesity management and have achieved satisfactory results [17,18]. However, these health promotion programs for lifestyle change are mostly used for people who have been diagnosed with diabetes [19]. There is little evidence of using mHealth technology for preventing diabetes [20]. For people with high risk but not diagnosed with diabetes, less adherence to lifestyle change compared to people with diabetes has been reported because they usually have lower awareness and lesser motivation to make changes [21]. In other words, whether lifestyle modification interventions via mHealth technology could work among high-risk people for diabetes, especially among busy mothers who have abdominal obesity, is still an open question.

The goal of this study was to promote changes in modifiable diabetes risk factors, thereby improving metabolic levels and reducing the risk of diabetes. The primary objective was to evaluate the feasibility and the acceptability of an mHealth-based diabetes prevention program and its preliminary efficacy in reducing weight-related variables, that is, WC and BMI, and diabetes risk (type 2 diabetes risk score) among Chinese mothers with abdominal obesity in 6 months. The secondary objective was to assess the preliminary efficacy of the intervention on glycated hemoglobin (HbA_{1c}) levels, behavioral variables (daily steps and daily moderate to vigorous activity time), and psychological variables (self-efficacy and social support for physical activity and diet, perceived stress, and quality of life) at 6 months.

Methods

Ethics Approval

This study received approval from the ethics review committee of the Xiangya School of Nursing, Central South University (batch 2019002). This study was retrospectively registered with the Chinese Clinical Trial Registry (ChiCTR2400090554) on October 24, 2024.

Study Design and Participants

This study utilizes a randomized controlled trial design. The inclusion criteria for the participants were as follows: (1) women as defined by biological sex, aged 18 years and older; (2) having access to a smartphone (Android and iOS phones compatible with Fitbit) and being willing to wear a Fitbit sports tracker for the entire study period; (3) willing to share their behavioral data collected via Fitbit with researchers in this study; (4) WC>80 cm and BMI>24.5; (5) having at least 1 child in the age range of 1-12 years; (6) not reached menopause; (7) able to read Chinese and speak Mandarin; and (8) planning to live locally for at least 8 months. The exclusion criteria were as follows:

XSL•FO
RenderX

(1) being pregnant, (2) having given birth within 12 months prior to the enrollment date, (3) having an acute or life-threatening illness (eg, kidney failure), (4) having a condition that requires dietary and activity control (eg, diabetes, hypertension, hyperthyroidism), (5) having irregular periods for 6 months, and (6) having plans to become pregnant within 1 year.

Sample Size

Monte Carlo simulations were conducted, aiming to estimate the power for effect sizes to be tested with multilevel regression models. Using G*power 3.1 software, sample size estimation was conducted based on the effect size of HbA_{1c}. The effect size of HbA_{1c} was calculated according to relevant literature as 0.83 [22], using a 2-tailed test with a significance level set at .05. To ensure a power of 90%, the total sample size required was determined to be 64. Accounting for a 20% dropout rate, the total sample size needed for this study was 80 individuals, with 40 participants in each group.

Recruitment

Participants were recruited from Health Management Centers for Adults and Children of 2 tertiary hospitals in Changsha, a provincial capital city of Hunan Province in China via posters and web-based recruitment. Two research assistants put up posters at the health management centers. The web-based recruitment was promoted in the WeChat Moments through Eqxiu. Eqxiu is an all-in-one creative design and marketing platform that uses design tools such as posters, long pages, and videos to produce marketing content for users. Interested mothers could contact the research assistants at the research sites. Research assistants measured the WC, height, and weight onsite to ensure they met inclusion and exclusion criteria.

If interested participants met the eligibility criteria, the researchers would provide them with a description of the purpose, content, process, risks, benefits, and the right to withdraw. Written informed consent was obtained from the participants in this study. All research-related information, including basic demographic data of the research participants and the original data recorded in this study, strictly adhered to the principles of research confidentiality. Individuals participating in this study received a designated compensation for the time and effort they expended in support of this research following the completion of each data collection session.

Randomization

Randomization was performed using a random sequence generated by SPSS software (v. 25.0; SPSS Inc) in this study. The mHealth group was designated as number 1 beforehand and the control group as number 2. According to the time order of entry into the study, the researchers used SPSS software to generate a corresponding random number for each study participant and then coded half of the random numbers as 1 and the other half as 2. The study participants were entered into 2 groups according to this code. After randomization, the intervention and control groups included 40 participants, respectively. Neither the study participants nor the medical staff involved in recruitment were informed of the randomization allocation results. All research assistants involved in this study possessed extensive clinical and research experience. Prior to the commencement of the program, they underwent specialized training provided by the program lead. Notably, the research assistants responsible for the interventions in the mHealth group and the control group were different, with each receiving separate training to avoid potential contamination.

mHealth Group Intervention

The mHealth-based lifestyle intervention was developed based on the social cognitive theory. The intervention consisted of 3 components: (1) recording participants' exercise, sleep, and diet data via Fitbit during the whole study period (6 months); (2) accessing the 12 weekly lifestyle modification modules delivered on WeChat; and (3) receiving personalized biweekly messages after the modules, informed by Fitbit data, which were sent via WeChat for 3 months.

Fitbit Tracker (6 months)

Participants received personalized, one-on-one instruction from research assistants on using the Fitbit tracker and its mobile app prior to wearing the device. This instruction covered setting behavioral goals, learning to monitor and record behavior-related data systematically, interpreting displayed behavior values, and synchronizing/uploading data. Research assistants sent reminders via WeChat to participants who did not synchronize their data for over a week to improve compliance. Throughout the intervention, participants wore Fitbit trackers to record steps, moderate to vigorous activity time, and sleep. They could also manually enter dietary intake, automatically converted into calories. Fitbit, available as a mobile app for Android and iOS, facilitated automatic data syncing via Bluetooth. Individual sync data were accessible to researchers on the Fitbit website. Fitbit Alta HR trackers were obtained directly from the official Fitbit website using project funds. No additional payments were made to Fitbit, and participants incurred no usage fees. Chargers were provided with the trackers, and participants received both for free after the intervention.

WeChat is a Chinese instant messaging, social media, and mobile payment app with over 1 billion monthly active users. The rationale for disseminating educational modules to participants via WeChat stems from its ubiquitous presence as an everyday app for nearly all Chinese individuals owning a smartphone in China. Leveraging this platform ensures that each participant promptly receives our educational modules. The widespread use of WeChat in the daily lives of the Chinese population facilitates efficient and timely communication of educational content, maximizing the accessibility and reach of our intervention.

Weekly Educational Modules Posted on WeChat (12 Weeks)

The researcher sent out 1 educational module per week via WeChat for 12 weeks. The educational modules consist of the following 5 categories: (1) introduction to diabetes and obesity (modules 1-3), (2) healthy diet (modules 4-5), (3) various exercise patterns (modules 6-9), (4) stress (modules 10-11), and (5) weight maintenance strategies (module 12). The selection of topics and content for the 12 educational modules was based on the etiology and factors contributing to diabetes [23],

XSL•FO
RenderX

evidence from the American Diabetes Association, and consensus decisions reached by our expert panel. Figure 1 shows the details of the weekly educational modules. Each exercise module consists of 3 videos of varying difficulty levels: low, moderate, and high. Participants can select the appropriate level based on their individual fitness capabilities. All modules can be viewed repeatedly. Research assistants remind participants to review modules promptly with messages and phone calls. A point system, rewarding 10 points per module, motivated ongoing participation, with a small gift at 120 points.

Figure 1. WeChat weekly educational modules for the mobile health group.

Tailored WeChat Messages Based on Individual Behavioral Patterns and Preferences (6 Messages in Total)

Research assistants assessed exercise and dietary patterns for tailored messages based on the recordings. After completing the 12 weekly web-based educational modules on WeChat, the research team analyzed the Fitbit data of participants in the mHealth group in order to send tailored messages aligned to the participant's behavioral patterns based on their exercise and diet data for the last 2 weeks. According to the participants' situations of goal completion, individual behavioral patterns were categorized into 3 levels: below target, meeting target, and exceeding target. Exercise steps up to 0%-70% of the target steps were identified as below target, 71%-99% were identified as meeting target, and 100% and above were identified as exceeding target. Dietary targets were set according to the dietary guidelines: total daily intake of vegetables and fruits was set at 5 standard servings (approximately 500 g, 100 g raw weight per standard serving), with 1-2 servings identified as below target, 3-4 servings identified as meeting target, and 5 servings and above identified as exceeding target. Tailored biweekly WeChat messages consisted of 3 components: assessment of current progress, transition words, and recommendations. For instance, "Congratulations, Ms Wang! You've reached your exercise goal, moving closer to a healthier lifestyle. Find more dietary advice on module 4 'Healthy Eating.' Keep up the good work!"

Control Group Interventions

The control group interventions consisted of 3 main components: Fitbit trackers identical to those used by the mHealth group, 12 educational modules addressing general health information delivered on WeChat, and health-improvement messages. The 12 modules, which did not focus on lifestyle modification, covered a variety of topics: (1) obesity and diabetes, (2) parenting education, (3) domestic violence, (4) sourcing food, (5) Chinese Adult Physical Activity Guidelines, (6) cardiovascular disease, (7) health checkups, (8) anxiety, (9) sexually transmitted diseases, (10) contraception, (11) hepatitis B, and (12) AIDS. Figure 2 shows the samples of the weekly educational modules for the control group. The supervision pattern for module completion in the control group mirrored that of the mHealth group. Additionally, nontailored informational tips were delivered to the control group, aligning with the educational content of the modules. For instance, participants received messages such as "AIDS, a class B legal infectious disease in China, lacks effective cure drugs. Click to review the tenth module on WeChat."

XSL•FO
RenderX

Figure 2. WeChat weekly educational modules for the control group.

Data Collection

At baseline, 3 months, and 6 months, research assistants who were unaware of group allocation contacted the participants via phone or WeChat for data collection, including questionnaires and physical examinations. The research assistants responsible for contacting participants were distinct from those who implemented the interventions and were blinded to group assignments. Participants completed questionnaires in a designated hospital room, with researchers available for assistance. Physical examinations, conducted by trained nurses, included height (measured at baseline), weight, and WC, with data averaged from 3 measurements. Fasting blood samples for HbA_{1c} were collected by hospital nurses at baseline and 6 months. Participants received travel and meal allowances as appreciation for their participation. To avoid contamination, data collection for both groups occurred on separate dates at the study sites.

Outcomes Measures

Feasibility and Acceptability

The feasibility was evaluated by calculating the number of module views by participants, which is automatically recorded by the WeChat platform. The acceptability outcome measure is a satisfaction question concerning each module. Satisfaction questions asked participants whether they thought they had learnt something after watching the content of a module, with scores ranging from 1 to 5—a higher score indicating a higher level of satisfaction.

Demographics

General sociodemographic information was collected at baseline, consisting of 26 questions on the study population's demographic and sociological information, including maternity history, lifestyle habits, and family history of diabetes and other chronic diseases.

Measurement of Weight-Related Variables

WC is the circumference of the horizontal line between the upper edge of the iliac crest and the lowest point of the rib cage—generally the horizontal circumference of the center of the umbilicus. A trained nurse instructed the participant to breathe regularly and read the value of the soft ruler (accuracy of 0.1 cm) at the end of exhalation. The value was recorded as the average of the measurements. BMI is calculated by dividing body weight (kg) by the square of height (m). Height is measured using the same calibrated height gauge (0.5 cm accuracy) and weight scale (0.1 kg accuracy).

Measurement of Diabetes Risk Score

The Chinese Diabetes Risk Questionnaire (CHINARISK) as the diabetes risk assessment tool consists of 14 modifiable and unmodifiable items covering age, gender, ethnicity, residence, education level, WC, BMI, daily exercise, daily intake of fruits and vegetables, history of hypertension, treatment history of taking medication for hypertension, history of giving birth to a large child, blood glucose status, and family history of diabetes. Among these, the 4 modifiable risk factors were WC, BMI, daily exercise, and daily intake of fruits and vegetables. The scores of the 4 entries were summed to calculate the CHINARISK modifiable scores. The modifiable score implies that by modifying lifestyle and achieving weight loss, it is possible to lower the scores in CHINARISK, thereby contributing to a reduction in the risk of diabetes. The possible total score range of CHINARISK is 0-82 (including the modifiable and unmodifiable scores). A total score above 30 is considered a high risk of developing diabetes in the future and requires prompt intervention and further diagnosis. The retest reliability of CHINARISK was 0.988, with a positive predictive value of 57%, a negative predictive value of 78%, and sensitivity of 73% [24].

Measurement of Metabolic Variables

HbA_{1c} reflects average blood glucose levels over 2-3 months, with a normal range of 4%-5.6% [25]. Fasting blood specimens were sent to Changsha Goldfield Medical Testing Center within

XSL•FO
RenderX

2 hours for standardized testing, thereby reducing methodological errors across sites. Standardized HbA_{1c} reports were issued by the Goldfield Medical Testing Center.

Measurement of Behavioral Variables

All behavioral variables in this study, including step count, moderate to vigorous activity time, sleep duration, daily fruit and vegetable intake, and total daily calorie intake/burn, were recorded in real time by the Fitbit tracker. It accurately measures step counts through a built-in gyroscope and acceleration sensors and tracks moderate to vigorous activity time by using the accelerometer. Sleep duration is determined by the frequency and amplitude of tossing and turning. The Fitbit tracker estimates daily calorie intake/consumption from exercise time, intensity, sleep duration, BMI, etc. It has shown good reliability and validity in various large lifestyle intervention studies, with intraclass correlation coefficients ranging from 0.71 to 1.00 [26]. In this study, daily steps and daily moderate to vigorous activity time were averaged over the last week at baseline, 3 months, and 6 months.

Measurement of Diet and Exercise-Related Psychological Variables

The Dietary and Physical Self-Efficacy Scale and Social Support Scale are used to assess participants' self-efficacy and the level of social support provided by family and friends in relation to dietary and exercise behaviors [27,28]. The Self-Efficacy Scale and the Social Support Scale has 4 and 6 entries, respectively. These 4 scales are scored on a 4-point scale, with higher scores indicating higher levels of individual self-efficacy and social support for eating and exercise. The Cronbach coefficients for the Chinese version of the Dietary Self-Efficacy Scale and the Physical Activity Self-Efficacy Scale ranged from 0.93 to 0.95 and 0.87 to 0.97, respectively. The Cronbach coefficients of the Chinese version of the Dietary Social Support Scale and the Physical Activity Social Support Scale ranged from 0.86 to 0.89 and 0.78 to 0.87, respectively. Perceived Stress Scale-14 items (PSS-14) is used to assess the amount of stress perceived by the participant over the past month [29]. The scores for each item are summed to give a total score on the PSS-14, which ranges from 0 to 56. The higher the score, the higher the perceived stress. The Cronbach coefficient of the Chinese version of the PSS-14 scale was 0.78.

Measurement of Quality of Life

The World Health Organization Quality of Life Scale-Brief Version is used to quantify participants' self-assessment of their own quality of life and general health status [30]. It contains a total of 26 items and addresses 4 quality of life domains: physical health (7 items), psychological health (6 items), social relationships (3 items), and environment (8 items). The items are scored on a 5-point Likert scale. Higher scores in a domain indicate higher self-rated quality of life in that domain. Domain scores range from 4 to 20, with a Cronbach coefficient of 0.88 for the Chinese version of World Health Organization Quality of Life Scale-Brief Version.

Statistical Methods

Raw data were entered in pairs using EpiData 3.1 software [31] to verify accuracy. Data were analyzed using SPSS 25.0 software [32] with a test level of α taken as .05. Statistical description methods were proposed to describe the demographic and sociological data and all scale data for both study groups. One-way analysis of variance (for measurement information), least significance difference tests (2-way comparisons between factors), and χ^2 tests (for counting information) were performed to compare the differences in demographic information and primary and secondary variables between the 2 groups of enrolled and study participants who did not complete the follow-up assessments. Intention-to-treat analysis was used, with all participants (N=80) analyzed based on their original group assignments, including those excluded or lost. To compare outcome indicators (including psychological, behavioral, weight-related, and diabetes risk–related variables) between the 2 groups over 6 months, we established a 3 (time: baseline, 3 months, and 6 months) * 2 (group: intervention and control) generalized estimating equation (GEE) model by using general sociodemographic information and disease risk–related data as covariates to control for the confounding factors. Additionally, we developed a 2 (time: baseline and 6 months) * 2 (group: intervention and control) GEE model to evaluate the time trend of HbA_{1c} levels and compare the differences in HbA_{1c} between the mHealth group and the control group.

The GEE model offers key advantages over traditional repeated measures analysis. It accounts for within-subject correlations in longitudinal data, provides reliable estimates even when the correlation structure is incorrectly specified, and focuses on population-average effects. GEE can also handle missing data by assuming it occurs completely at random, allowing for the inclusion of incomplete datasets. Additionally, by incorporating covariates, GEE controls for confounding factors, making it a robust choice for analyzing repeated measurements over time.

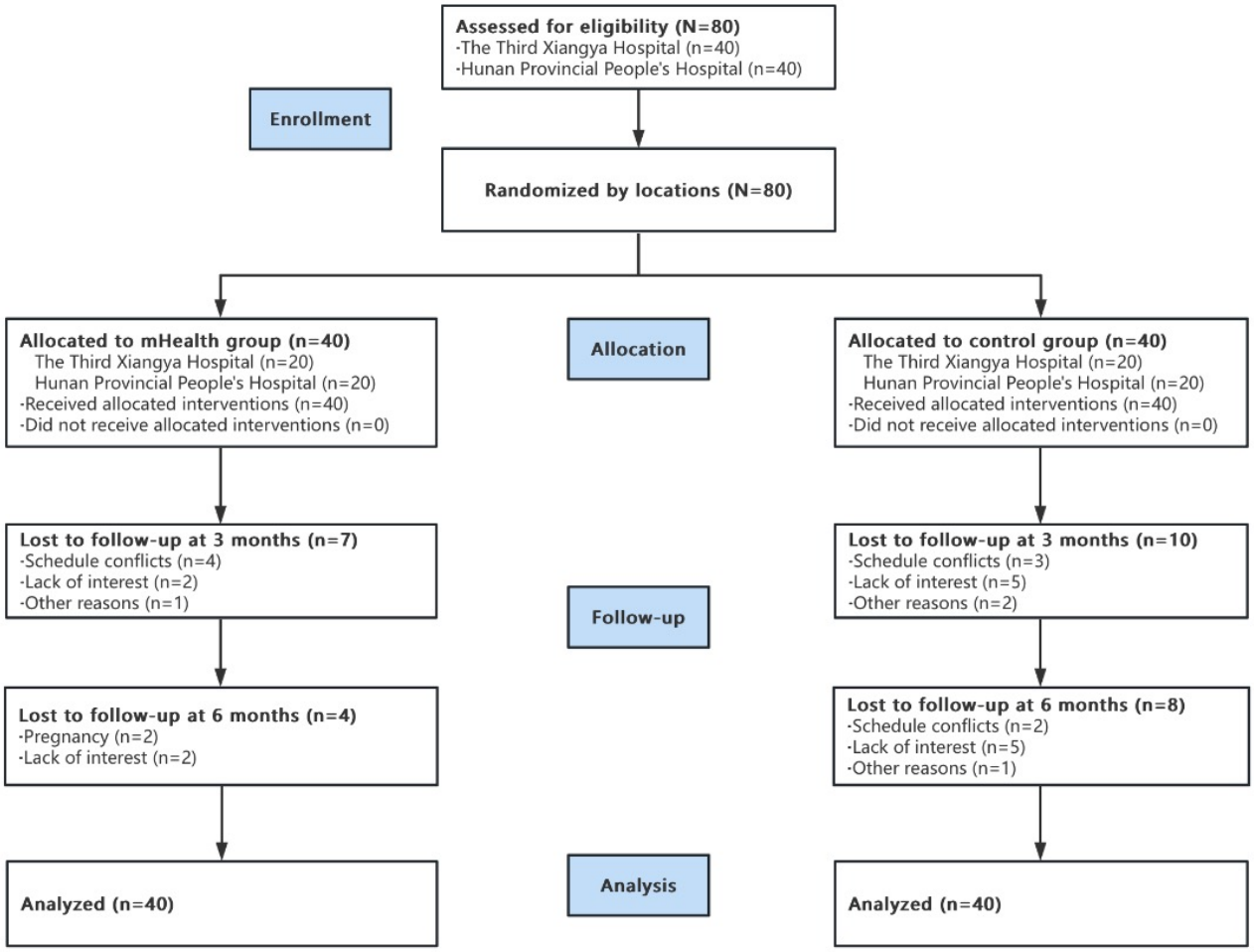
Results

Baseline Sociodemographic and Clinical Characteristics

A total of 80 participants were enrolled and randomly assigned to the mHealth group (n=40) and the control group (n=40). Figure 3 lists the details of inclusion and loss to follow-up. The mean age of the participants was 34.86 (SD 4.36; range 26-45) years. About 79% (63/80) of the participants had a university or higher level of education. The mean WC and BMI of the sample were 92.61 (SD 9.58) cm and 28.12 (SD 3.93) kg/m^2, respectively. The mean type 2 diabetes risk score of the sample was 21.50 (SD 9.66). There were no statistically significant differences in baseline sociodemographic and clinical characteristics between the mHealth and control groups. There were no significant differences in the demographic and clinical characteristics between participants who completed the follow-ups and those who missed the follow-ups (P>.05; Table 1).

XSL•FO
RenderX

Figure 3. Study design and participant flow.



XSL•FO
RenderX

Table 1. Comparison of the baseline sociodemographic and clinical characteristics between the control and mobile health groups.

Variables	Total (N=80)	Mobile health group (n=40)	Control group (n=40)	χ^2/*t* test (*df*)	*P* value
Age (years), mean (SD)	34.86 (4.36)	34.56 (4.01)	35.19 (4.77)	0.63 (77)[a]	.53
Education, n (%)				1.9 (1)[b]	.17
Senior high school and below	17 (21)	6 (15)	11 (28)		
College and above	63 (79)	34 (85)	29 (72)		
Marital status, n (%)				0.2 (1)[b]	.64
Married	75 (94)	37 (92)	38 (95)		
Divorced	5 (6)	3 (8)	2 (5)		
Occupation, n (%)				0.49 (1)[b]	.49
Employed	51 (64)	27 (67)	24 (60)		
Unemployed	29 (36)	13 (33)	16 (40)		
Youngest child age (years), mean (SD)	5.99 (2.98)	5.78 (3.01)	6.22 (2.98)	0.65 (75)[a]	.52
BMI (kg/m^2), mean (SD)	28.12 (3.93)	28.19 (4.12)	28.05 (3.76)	–0.15 (76)[a]	.88
Waist circumference (cm), mean (SD)	92.61 (9.58)	93.09 (10.45)	92.07 (8.63)	–0.47 (76)[a]	.64
Hemoglobin A_{1c}, mean (SD)	5.74 (0.74)	5.83 (0.93)	5.65 (0.45)	–1.06 (74)[a]	.29
Five servings of fruits and vegetables daily intake, n (%)				1.4 (1)[b]	.25
Yes	29 (36)	12 (30)	17 (43)		
No	51 (64)	28 (70)	23 (57)		
30 minutes of daily exercise, n (%)				0.4 (1)[b]	.50
Yes	35 (44)	19 (48)	16 (40)		
No	45 (56)	21 (52)	24 (60)		
Type 2 diabetes risk score, mean (SD)	21.50 (9.66)	21.46 (9.05)	21.54 (10.42)	0.04 (76)[a]	.97

[a]*t* test values (2-sided).

[b]Chi-square values.

Feasibility and Acceptability

The total number of WeChat module views for the mHealth and control groups were 697 and 189, respectively, and the top 3 view sessions in the mHealth group were modules 1, 6, and 3 with 89, 80, and 77 views, respectively. All modules can be viewed repeatedly but will also be counted repeatedly. The mean score of the satisfaction questionnaire about weekly educational modules posted on WeChat was 4.37 out of 5 for the mHealth group and 3.71 for the control group.

Changes in the Primary and Secondary Outcomes Over Time Between mHealth and Control Groups

With a reduction of 3.39 cm and 1.50 points in the mHealth group versus 2.37 cm and 0.56 points in the control group at the 6-month follow-up (Figures 4 and 5; Table 2), respectively, we evaluated the effectiveness of the intervention and accounted for missing data through an intention-to-treat analysis.

The intention-to-treat analysis of the adjusted GEE model revealed a significant group-by-time interaction effect on WC and modifiable type 2 diabetes risk scores (Table 3). We observed a gradual decrease in the primary outcome measures due to the group-by-time interaction effect. No significant effect on BMI was found (Table 3). For secondary outcomes, there was a significant group-by-time interaction effect for the mHealth group on daily steps, self-efficacy, social support for physical activity, and the physiological domain of quality of life (Table 3). Specifically, the daily steps and daily moderate to vigorous activity time in the mHealth group were significantly higher than those in the control group at baseline, 3 months, and 6 months. Although the daily moderate to vigorous activity time in the mHealth group was consistently higher than that in the control group, there was no significant group-by-time interaction effect for this variable (Table 2 and Table 3).

XSL•FO
RenderX

Figure 4. Changes in the waist circumference between the mobile health and control groups over 6 months. mHealth: mobile health.

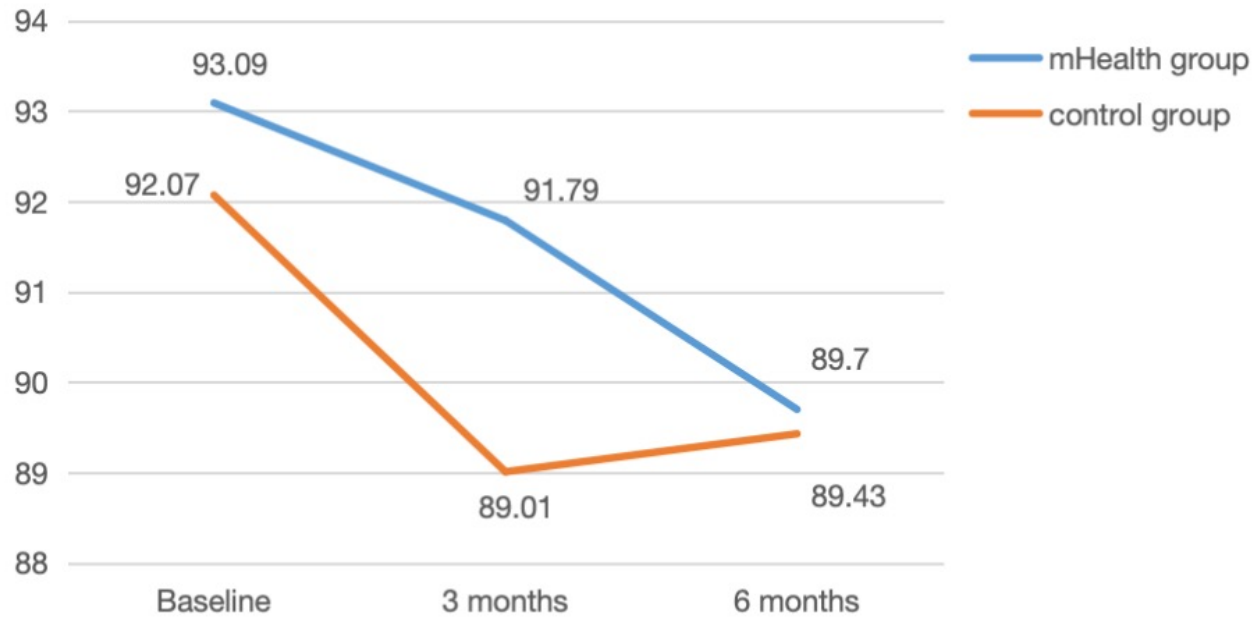
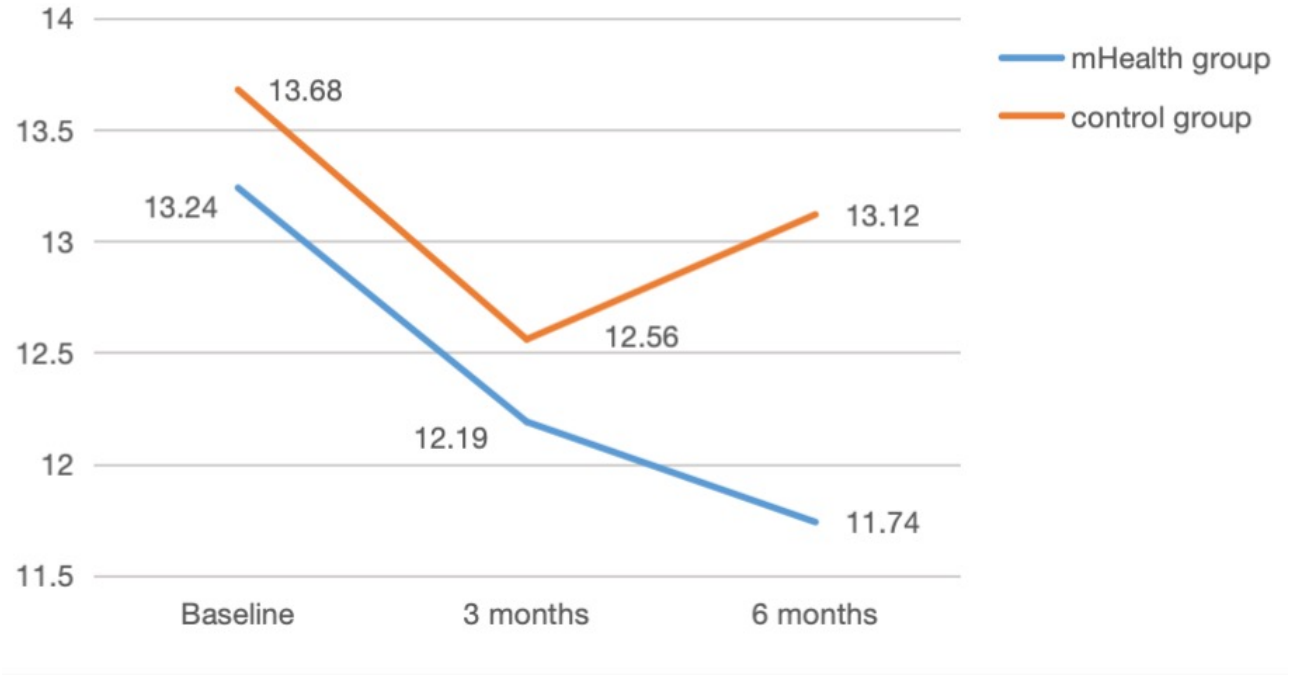


Figure 5. Changes in the modifiable type 2 diabetes mellitus risk scores between the mobile health and control groups over 6 months. mHealth: mobile health.



XSL•FO
RenderX

Table 2. Comparison of the primary and secondary outcomes between the mobile health group and control group (N=80).

Variables	Mobile health group			Control group		
	Baseline (n=40)	3 months (n=33)	6 months (n=36)	Baseline (n=40)	3 months (n=30)	6 months (n=32)
Primary outcome, mean (SD)						
Waist circumference	93.09 (10.46)	91.79 (11.22)	89.70 (10.13)	92.07 (8.63)	89.01 (9.13)	89.70 (9.45)
BMI	28.19 (4.12)	27.82 (6.79)	28.42 (4.49)	28.06 (3.77)	28.26 (3.70)	28.23 (3.51)
Type 2 diabetes risk scores	21.46 (9.05)	21.09 (10.35)	21.26 (11.58)	21.54 (10.43)	20.88 (9.94)	21.88 (11.31)
Modifiable type 2 diabetes risk scores	13.24 (4.43)	12.19 (5.11)	11.74 (5.75)	13.68 (5.13)	12.56 (5.86)	13.12 (5.64)
Secondary outcomes, mean (SD)						
Hemoglobin A_{1c}	5.84 (0.93)	N/A[a]	5.78 (1.07)	5.65 (0.46)	N/A	5.85 (1.61)
Daily steps	10,662.11 (4542.38)	10,793.70 (4825.32)	8967.02 (6431.57)	8441.45 (3666.66)	8657.78 (4915.79)	4934.83 (3692.10)
Daily moderate to vigorous activity time	40.06 (40.69)	42.09 (46.51)	31.82 (46.15)	27.49 (19.41)	15.25 (19.54)	5.63 (15.60)
Self-efficacy for diet	6.73 (2.86)	6.91 (2.81)	6.32 (2.77)	6.32 (4.11)	6.70 (3.14)	6.88 (3.44)
Self-efficacy for physical activity	6.98 (3.42)	6.31 (2.81)	6.47 (2.79)	7.43 (3.73)	6.06 (3.24)	6.27 (3.25)
Social support for diet	5.00 (1.52)	5.66 (1.98)	4.97 (1.38)	5.24 (2.15)	5.31 (1.64)	4.76 (1.71)
Social support for physical activity	11.51 (3.63)	11.91 (3.73)	11.50 (3.47)	12.68 (4.68)	11.53 (3.89)	11.61 (4.11)
Perceived stress	23.68 (8.78)	25.31 (8.472)	23.53 (9.33)	22.57 (6.62)	22.66 (7.89)	23.91 (6.82)
Physiological QoL[b]	14.04 (2.04)	13.20 (4.08)	14.06 (2.26)	14.10 (1.93)	13.95 (2.16)	14.06 (1.96)
Psychological QoL	13.96 (2.35)	12.71 (4.23)	13.78 (2.69)	13.77 (1.98)	14.27 (1.94)	13.82 (1.80)
Environment QoL	13.69 (2.21)	12.64 (4.02)	13.37 (2.00)	13.87 (1.94)	14.27 (1.84)	13.83 (1.80)
Social relationships QoL	14.13 (3.01)	12.94 (4.07)	13.92 (2.25)	14.41 (2.06)	14.54 (2.18)	14.30 (2.57)

[a]N/A: not applicable. Hemoglobin A_{1c} was only measured twice in the program.

[b]Qol: quality of life.

XSL•FO
RenderX

Table 3. Generalized estimating equation analysis of the 6-month efficacy of the intervention on the primary outcomes and the secondary outcomes[a].

Outcome	β estimate (SE; 95% CI)	*P* value
Primary outcomes, mean (SD)		
Waist circumference		
Group	4.90 (2.83; –0.65 to 10.46)	.08
Time	2.10 (1.61; –1.07 to 5.27)	.20
Group × time	–2.24 (0.95; –4.12 to –0.36)	.02
BMI		
Group	1.23 (1.07; –0.88 to 3.34)	.26
Time	0.89 (0.50; –0.09 to 1.87)	.08
Group × time	–0.39 (0.26; –0.92 to 0.12)	.14
Type 2 diabetes risk scores		
Group	6.82 (4.07; –1.16 to 14.80)	.09
Time	7.89 (4.91; –1.73 to 17.51)	.11
Group × time	–4.18 (2.57; –9.22 to 0.86)	.10
Modifiable type 2 diabetes risk scores		
Group	2.74 (1.98; –1.14 to 6.63)	.17
Time	4.35 (2.04; 0.34 to 8.36)	.03
Group × time	–2.50 (1.05; –4.57 to –0.44)	.02
Secondary outcomes, mean (SD)		
Hemoglobin A_{1c}		
Group	0.32 (0.26; –0.18 to 0.84)	.21
Time	0.42 (0.46; –0.48 to 1.32)	.36
Group × time	–0.23 (0.23; –0.69 to 0.21)	.31
Daily steps		
Group	0.56 (1.58; –2.54 to 3.67)	.72
Time	–4.82 (1.21; –7.19 to –2.45)	<.001
Group × time	1.67 (0.82; 0.06 to 3.29)	.04
Daily moderate to vigorous activity time		
Group	0.39 (5.63; –10.64 to 11.44)	.94
Time	–9.11 (3.53; –16.03 to –2.18)	.01
Group × time	3.28 (2.25; –1.13 to 7.70)	.15
Self-efficacy for diet		
Group	2.59 (1.99; –1.31 to 6.50)	.19
Time	3.33 (2.01; –0.62 to 7.28)	.10
Group × time	–1.52 (1.07; –3.63 to 0.58)	.16
Self-efficacy for physical activity		
Group	–2.77 (1.65; –6.01 to 0.45)	.09
Time	–4.05 (1.27; –6.55 to –1.55)	.001
Group × time	1.93 (0.76; 0.44 to 3.43)	.01
Social support for diet		
Group	–0.56 (0.96; –2.45 to 1.31)	.55
Time	–1.12 (1.03; –3.16 to 0.90)	.28
Group × time	0.51 (0.55; –0.56 to 1.59)	.35

XSL•FO
RenderX

Outcome	β estimate (SE; 95% CI)	*P* value
Social support for physical activity		
Group	–0.36 (1.45; –6.46 to –0.75)	.01
Time	–4.38 (1.32; –6.98 to –1.79)	.001
Group × time	2.27 (0.74; 0.80 to 3.74)	.002
Perceived stress		
Group	2.02 (3.30; –4.45 to 8.51)	.54
Time	2.11 (3.37; –4.50 to 8.73)	.53
Group × time	–1.32 (1.88; –5.02 to 2.38)	.49
Physiological quality of life		
Group	–1.04 (0.73; –2.48 to 0.39)	.15
Time	–1.59 (0.61; –2.80 to –0.38)	.01
Group × time	0.82 (0.37; 0.08 to 1.55)	.03
Psychological quality of life		
Group	0.43 (0.81; –1.16 to 2.03)	.60
Time	0.89 (0.69; –0.46 to 2.25)	.20
Group × time	–0.45 (0.41; –1.26 to 0.36)	.28
Environment quality of life		
Group	–0.02 (0.72; –1.44 to 1.39)	.97
Time	0.40 (0.30; –0.20 to 1.00)	.20
Group × time	–0.35 (0.24; –0.83 to 0.11)	.14
Social relationships quality of life		
Group	–0.40 (0.81; –1.99 to 1.18)	.61
Time	0.07 (0.46; –0.83 to 0.98)	.88
Group × time	–0.04 (0.32; –0.67 to 0.58)	.89

[a]When constructing the generalized estimating equation model, time was categorized into baseline, 3 months, and 6 months for all outcome indicators, except hemoglobin A_{1c}. When hemoglobin A_{1c} was the dependent variable, time was categorized into baseline and 6 months. The participants were divided into 2 groups: intervention and control.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to explore the use of mHealth technology for diabetes prevention in mothers with abdominal obesity in China. We demonstrated that a tailored mHealth-based diabetes prevention intervention is both feasible and resource-efficient, yielding significant positive effects. Specifically, our findings provide preliminary evidence that such interventions can reduce WC, increase daily step count, enhance exercise-related self-efficacy and social support, improve quality of life, and lower modifiable diabetes risk. The high level of compliance among participants suggests potential for implementation in primary health care centers.

The success of this program can be attributed to several key factors. First, the intervention was grounded in the social cognition theory [33], which posits that psychosocial influences on health behavior are shaped by 3 factors: goals, outcome expectancies, and self-efficacy. In this study, participants set realistic, personalized goals for diet and exercise, greatly enhancing the likelihood of achieving desired outcomes. In addition, Fitbit's data visualization feature encouraged self-motivation and boosted participants' self-efficacy [34]. Second, our study population consisted of mothers whose children were younger than 12 years, a stage when increased maternal involvement is often required in various aspects of life, including academics, physical health, and emotional well-being. At the same time, they were at a life stage marked by active professional pursuits. As a result, these mothers may have inadvertently neglected their own health due to the competing demands of family and work. However, the mHealth-based lifestyle management intervention overcame time and space constraints, enhancing adherence, particularly for busy mothers [35,36].

The mHealth-based, tailored diabetes prevention intervention effectively reduced WC in mothers with abdominal obesity, which was consistent with findings from a recent meta-analysis [37]. The intervention included components aimed at reducing body fat and WC, such as 12 web-based modules (4 on exercise and 3 on diet). The 15-minute exercise videos were designed for both individual use and joint participation with children.

XSL•FO
RenderX

These short sessions made it easier for busy mothers to stay active daily. Additionally, wearing exercise trackers and receiving personalized health text messages further supported WC reduction. This approach highlights that small, consistent daily efforts play a key role in achieving long-term health improvement.

In this study, while the overall diabetes risk score did not show a significant reduction, it is noteworthy that a significant difference in diabetes risk scores emerged between the 2 groups when the 10 nonmodifiable risk factors from the CHINARISK diabetes risk scale were excluded. This left only 4 modifiable risk factors: WC, BMI, physical activity, and dietary habits. These findings suggest that tailored diabetes risk management interventions using mHealth technology can effectively target controllable risk factors such as WC, BMI, diet, and physical activity, ultimately helping to lower an individual's future risk of developing diabetes.

Exercise self-efficacy, as an essential component of health self-efficacy, reflects an individual's belief in their ability to resist temptations and maintain a healthy exercise routine [38]. This intervention led to significant improvements in exercise-related psychological variables, including exercise self-efficacy and exercise social support, which are likely to encourage women to engage in more physical activity. This aligns with findings from a systematic review of lifestyle interventions [39]. Our 12 web-based educational modules incorporated family-based exercise training, offering participants new formats that mobilize social resources and enhance support for exercise. However, there is limited national and international research on exercise-related social support for women with children at risk of diabetes. Therefore, the findings regarding social support in this study should be interpreted with caution.

This study shows that tailored mHealth-based diabetes risk management improved the quality of life in the physiological domain for abdominally obese mothers. Participants reported greater satisfaction with their sleep, increased energy levels for daily activities, and enhanced ability to perform everyday tasks. These findings align with weight management studies that link weight reduction to improved sleep quality [40]. Our intervention included exercise training, which encouraged behaviors such as increasing daily step counts. According to the pendulum effect [41], adequate physical activity during the day promotes deeper sleep at night, reduces sleep latency, and enhances overall sleep quality and satisfaction. This, in turn, leads to improved energy levels for daily activities, ultimately enhancing participants' quality of life in the physiological domain.

The frequent module reviews and the high satisfaction ratings reflect the feasibility of our research. The traditional face-to-face diabetes risk management program is costly and time-consuming, posing challenges in low-income countries with a shortage of primary health care professionals and limited medical resources [16]. WeChat, a popular multifunctional mobile app, is China's leading instant messaging tool and offers a wide range of connectivity features [42]. Utilizing WeChat as a web-based platform for health education has proven effective in promoting weight loss among Chinese obese women and even preventing breast cancer [43]. Our diabetes risk management program on WeChat leverages its flexible learning formats to engage participants and ensure high adherence to the intervention. Additionally, it requires less reliance on health care professionals, which helps conserve both human and economic resources—a significant advantage in the current context of low-income countries.

Limitations

First, participants were recruited exclusively from Changsha, limiting the generalizability of the findings to mothers with abdominal obesity across China. Second, the study lasted only 6 months, which may be insufficient to fully understand the long-term development of diabetes, given its gradual onset. Third, there could be discrepancies between the participants' actual eating behaviors and the diet data they manually entered in the Fitbit app. Fourth, the intervention and data collection mainly took place in autumn, a time when people tend to eat more [44], potentially contributing to the lack of significant changes in dietary variables in this study.

Implications

Future trials should place importance on long-term follow-ups to confirm the sustained effectiveness of this risk management intervention. Using factorial design methodology can help identify which specific elements of the intervention directly influence changes in patient behavior. This study, delivered through the WeChat platform, uses flexible learning formats to actively engage participants and ensure high adherence to the intervention. Moreover, by reducing reliance on health care professionals, this approach provides cost-effective benefits, which are particularly valuable in resource-limited health care settings.

Conclusion

This mHealth-based and tailored diabetes prevention model lays both the theoretical groundwork and practical strategies crucial for scaling up diabetes prevention and risk management initiatives in public health nursing practice. Although our study sheds light on the initial effects of mHealth-based interventions, it addresses the current scarcity of medical resources in China, offering a potential blueprint for tackling various chronic conditions.

Acknowledgments

The authors thank all the health care providers, research assistants, and participants in this work. This study was supported by the 2022 Central South University Graduate Student Independent Exploration and Innovation Program (grant 1053320216633) and the Central South University Research Programme of Advanced Interdisciplinary Studies (grant 2023QYJC041).

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

QH and QZ were involved in literature search, study design, data collection, statistical analysis, interpretation, and writing and revision of the manuscript. YZ was involved in writing the manuscript. YL was involved in data organization. JW, JLC, MPW, and JG were involved in the conception and supervision of this study. All authors revised the manuscript and approved the final report. We did not use generative artificial intelligence in any part of the manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online Telehealth) checklist.
[PDF File (Adobe PDF File), 3610 KB - mhealth_v13i1e47837_app1.pdf]

References

1. Papatheodorou K, Banach M, Bekiari E, Rizzo M, Edmonds M. Complications of diabetes 2017. J Diabetes Res 2018;2018:3086167 [FREE Full text] [doi: 10.1155/2018/3086167] [Medline: 29713648]
2. IDF Diabetes Atlas, 10th edn. Brussels, Belgium: International Diabetes Federation; 2021.
3. Kyrou I, Tsigos C, Mavrogianni C, Cardon G, Van Stappen V, Latomme J, Feel4Diabetes-study Group. Sociodemographic and lifestyle-related risk factors for identifying vulnerable groups for type 2 diabetes: a narrative review with emphasis on data from Europe. BMC Endocr Disord 2020 Mar 12;20(Suppl 1):134 [FREE Full text] [doi: 10.1186/s12902-019-0463-3] [Medline: 32164656]
4. Wu Y, Ding Y, Tanaka Y, Zhang W. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. Int J Med Sci 2014;11(11):1185-1200 [FREE Full text] [doi: 10.7150/ijms.10001] [Medline: 25249787]
5. Roglic G. WHO global report on diabetes: A summary. Int J Non-Commun Dis 2016;1(1):3. [doi: 10.4103/2468-8827.184853]
6. Chau Y, Bandiera R, Serrels A, Martínez-Estrada OM, Qing W, Lee M, et al. Visceral and subcutaneous fat have different origins and evidence supports a mesothelial source. Nat Cell Biol 2014 Apr;16(4):367-375 [FREE Full text] [doi: 10.1038/ncb2922] [Medline: 24609269]
7. Ross R, Neeland IJ, Yamashita S, Shai I, Seidell J, Magni P, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. Nat Rev Endocrinol 2020 Mar;16(3):177-189 [FREE Full text] [doi: 10.1038/s41574-019-0310-7] [Medline: 32020062]
8. Wang L, Peng W, Zhao Z, Zhang M, Shi Z, Song Z, et al. Prevalence and treatment of diabetes in China, 2013-2018. JAMA 2021 Dec 28;326(24):2498-2506 [FREE Full text] [doi: 10.1001/jama.2021.22208] [Medline: 34962526]
9. Feng G, Li H, Shen Q, Li Z, Ji X, Xiang Y. Population attributable risk of excess weight, abdominal obesity and physical inactivity for type 2 diabetes in Chinese men and women. Ann Transl Med 2021 Feb;9(4):326 [FREE Full text] [doi: 10.21037/atm-20-6121] [Medline: 33708953]
10. Haw JS, Galaviz KI, Straus AN, Kowalski AJ, Magee MJ, Weber MB, et al. Long-term sustainability of diabetes prevention approaches: a systematic review and meta-analysis of randomized clinical trials. JAMA Intern Med 2017 Dec 01;177(12):1808-1817 [FREE Full text] [doi: 10.1001/jamainternmed.2017.6040] [Medline: 29114778]
11. Kautzky-Willer A, Harreiter J, Pacini G. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. Endocr Rev 2016 Jun;37(3):278-316 [FREE Full text] [doi: 10.1210/er.2015-1137] [Medline: 27159875]
12. Du T, Sun X, Yin P, Huo R, Ni C, Yu X. Increasing trends in central obesity among Chinese adults with normal body mass index, 1993-2009. BMC Public Health 2013 Apr 10;13:327 [FREE Full text] [doi: 10.1186/1471-2458-13-327] [Medline: 23575244]
13. Paulweber B, Valensi P, Lindström J, Lalic N, Greaves C, McKee M, et al. A European evidence-based guideline for the prevention of type 2 diabetes. Horm Metab Res 2010 Apr;42(1):S3-36 [FREE Full text] [doi: 10.1055/s-0029-1240928] [Medline: 20391306]
14. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V, Indian Diabetes Prevention Programme (IDPP). The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). Diabetologia 2006 Feb;49(2):289-297. [doi: 10.1007/s00125-005-0097-z] [Medline: 16391903]
15. Kosaka K, Noda M, Kuzuya T. Prevention of type 2 diabetes by lifestyle intervention: a Japanese trial in IGT males. Diabetes Res Clin Pract 2005 Feb;67(2):152-162. [doi: 10.1016/j.diabres.2004.06.010] [Medline: 15649575]

XSL•FO
RenderX

16. Huang W, Long H, Li J, Tao S, Zheng P, Tang S, et al. Delivery of public health services by community health workers (CHWs) in primary health care settings in China: a systematic review (1996-2016). Glob Health Res Policy 2018;3:18 [FREE Full text] [doi: 10.1186/s41256-018-0072-0] [Medline: 29992191]
17. So CF, Chung JW. Telehealth for diabetes self-management in primary healthcare: A systematic review and meta-analysis. J Telemed Telecare 2017 May 02;24(5):356-364. [doi: 10.1177/1357633x17700552]
18. Chen X, Zhou X, Li H, Li J, Jiang H. The value of WeChat application in chronic diseases management in China. Comput Methods Programs Biomed 2020 Nov;196:105710 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105710] [Medline: 32858284]
19. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
20. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: A systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
21. Mainous AG, Tanner RJ, Scuderi CB, Porter M, Carek PJ. Prediabetes screening and treatment in diabetes prevention: the impact of physician attitudes. J Am Board Fam Med 2016 Nov 12;29(6):663-671 [FREE Full text] [doi: 10.3122/jabfm.2016.06.160138] [Medline: 28076248]
22. Jaana M, Paré G. Home telemonitoring of patients with diabetes: a systematic assessment of observed effects. J Eval Clin Pract 2007 Apr;13(2):242-253. [doi: 10.1111/j.1365-2753.2006.00686.x] [Medline: 17378871]
23. Ismail L, Materwala H, Al Kaabi J. Association of risk factors with type 2 diabetes: A systematic review. Comput Struct Biotechnol J 2021;19:1759-1785 [FREE Full text] [doi: 10.1016/j.csbj.2021.03.003] [Medline: 33897980]
24. Guo J, Shi Z, Chen J, Dixon JK, Wiley J, Parry M. Translation and validation of the Canadian diabetes risk assessment questionnaire in China. Public Health Nurs 2018 Jan;35(1):18-28. [doi: 10.1111/phn.12380] [Medline: 29285800]
25. Nomura S, Sakamoto H, Rauniyar SK, Shimada K, Yamamoto H, Kohsaka S, et al. Analysis of the relationship between the HbA1c screening results and the development and worsening of diabetes among adults aged over 40 years: a 4-year follow-up study of 140,000 people in Japan - the Shizuoka study. BMC Public Health 2021 Oct 18;21(1):1880 [FREE Full text] [doi: 10.1186/s12889-021-11933-z] [Medline: 34663286]
26. Coughlin SS, Thind H, Liu B, Wilson LCC. Towards research-tested smartphone applications for preventing breast cancer. Mhealth 2016 Jun;2:26 [FREE Full text] [doi: 10.21037/mhealth.2016.06.02] [Medline: 27390745]
27. Sallis JF, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for healthrelated diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]
28. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987 Nov;16(6):825-836. [doi: 10.1016/0091-7435(87)90022-3] [Medline: 3432232]
29. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. Journal of Health and Social Behavior 1983 Dec;24(4):385. [doi: 10.2307/2136404]
30. No authors listed. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
31. Kafatos G, Verlander N, Gelb D, Arnold E, Charlett A. A comparison of public domain software for outbreak investigations: Epi Info and EpiData. Euro Surveill 2007 Jan 11;12(1):E070111.4. [doi: 10.2807/esw.12.02.03111-en] [Medline: 17370933]
32. Cronk BC. How to Use SPSS®: A Step-by-Step Guide to Analysis and Interpretation. New York: Routledge; Oct 11, 2019.
33. Bandura A. Social foundations of thought and action: a social-cognitive view. Academy of Management Review 1986;12:1. [doi: 10.5465/amr.1987.4306538]
34. Asimakopoulos S, Asimakopoulos G, Spillers F. Motivation and user engagement in fitness tracking: heuristics for mobile healthcare wearables. Informatics 2017 Jan 22;4(1):5. [doi: 10.3390/informatics4010005]
35. Cavero-Redondo I, Martinez-Vizcaino V, Fernandez-Rodriguez R, Saz-Lara A, Pascual-Morena C, Álvarez-Bueno C. Effect of behavioral weight management interventions using lifestyle mHealth self-monitoring on weight loss: a systematic review and meta-analysis. Nutrients 2020 Jul 03;12(7):1977 [FREE Full text] [doi: 10.3390/nu12071977] [Medline: 32635174]
36. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
37. Lee HS, Lee J. Effects of combined exercise and low carbohydrate ketogenic diet interventions on waist circumference and triglycerides in overweight and obese individuals: a systematic review and meta-analysis. Int J Environ Res Public Health 2021 Jan 19;18(2):828 [FREE Full text] [doi: 10.3390/ijerph18020828] [Medline: 33478022]
38. Fletcher J, Banasik J. Exercise self-efficacy. Clin Excell Nurse Pract 2001 May;5(3):134-143. [doi: 10.1054/xc.2001.24203] [Medline: 11381353]
39. Hamidi S, Gholamnezhad Z, Kasraie N, Sahebkar A. The effects of self-efficacy and physical activity improving methods on the quality of life in patients with diabetes: a systematic review. J Diabetes Res 2022;2022:2884933 [FREE Full text] [doi: 10.1155/2022/2884933] [Medline: 35936392]

XSL•FO
RenderX

40. Kline CE, Chasens ER, Bizhanova Z, Sereika SM, Buysse DJ, Imes CC, et al. The association between sleep health and weight change during a 12-month behavioral weight loss intervention. Int J Obes (Lond) 2021 Mar;45(3):639-649 [FREE Full text] [doi: 10.1038/s41366-020-00728-8] [Medline: 33414489]
41. Bliwise DL. The pit (of sleeplessness) and the pendulum (of regulation). Sleep Med 2010 Jan;11(1):7-8. [doi: 10.1016/j.sleep.2009.08.002] [Medline: 19945339]
42. Montag C, Becker B, Gan C. The multipurpose application WeChat: a review on recent research. Front Psychol 2018;9:2247 [FREE Full text] [doi: 10.3389/fpsyg.2018.02247] [Medline: 30618894]
43. Chen J, Guo J, Zhong Q, Jiang Y, Zhang H, Mao P, et al. Smartphone-based cancer and obesity prevention education program for Chinese women (SCOPE): a pilot RCT. Int J Environ Res Public Health 2023 May 09;20(10):5768 [FREE Full text] [doi: 10.3390/ijerph20105768] [Medline: 37239495]
44. Ma Y, Olendzki BC, Li W, Hafner AR, Chiriboga D, Hebert JR, et al. Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. Eur J Clin Nutr 2006 Apr;60(4):519-528 [FREE Full text] [doi: 10.1038/sj.ejcn.1602346] [Medline: 16340952]

Abbreviations

CHINARISK: Chinese Diabetes Risk Questionnaire
GEE: generalized estimating equation
HbA_{1c}: glycated hemoglobin A_{1c}
mHealth: mobile health
PSS-14: Perceived Stress Scale-14 items
WC: waist circumference

Edited by L Buis; submitted 08.04.23; peer-reviewed by Y Jeem, F Tian, E Aigbokhan; comments to author 17.11.23; revised version received 11.02.24; accepted 19.11.24; published 24.01.25.

Please cite as:
Huang Q, Zhong Q, Zeng Y, Li Y, Wiley J, Wang MP, Chen JL, Guo J
mHealth-Based Diabetes Prevention Program for Chinese Mothers With Abdominal Obesity: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e47837
URL: https://mhealth.jmir.org/2025/1/e47837
doi:10.2196/47837
PMID:

©Qinyuan Huang, Qinyi Zhong, Yanjing Zeng, Yimeng Li, James Wiley, Man Ping Wang, Jyu-Lin Chen, Jia Guo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Tailored mHealth Intervention for Improving Antenatal Care Seeking and Its Determinants Among Pregnant Adolescent Girls and Young Women in South Africa: Pilot Randomized Controlled Trial

Ronel Sewpaul[1,2], PhD; Ken Resnicow[3], PhD; Rik Crutzen[1], PhD; Natisha Dukhi[2], PhD; Priscilla Reddy[4], PhD

[1]Department of Health Promotion, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands

[2]Public Health, Societies and Belonging, Human Sciences Research Council, Pretoria, South Africa

[3]School of Public Health, University of Minnesota, Minneapolis, MN, United States

[4]Business School, Imperial College London, London, United Kingdom

Corresponding Author:
Ronel Sewpaul, PhD
Public Health, Societies and Belonging
Human Sciences Research Council
134 Pretorius Street
Pretoria, 0002
South Africa
Phone: 27 4668000
Email: rsewpaul@hsrc.ac.za

Abstract

Background: Adolescent pregnancy is of public health concern due to high rates of pregnancy-related complications and lower antenatal attendance among adolescent girls and young women. Mobile health (mHealth) interventions have the potential to improve pregnancy health behaviors and thereby birth outcomes.

Objective: This pilot randomized controlled trial with pre-post design evaluated user acceptability and preliminary efficacy of an mHealth intervention to improve antenatal appointment attendance and its determinants among pregnant adolescent girls and young women in South Africa.

Methods: The "Teen MomConnect" intervention entailed both fixed and 2-way tailored SMS text messages about antenatal appointment keeping and pregnancy health behaviors. The intervention content and functionality were adapted from MomConnect, a national mHealth program that sends fixed SMS text messages to pregnant women in South Africa. Pregnant adolescent girls and young women aged 13-20 years were recruited from health facilities and community networks in Cape Town during May-December 2018. Simple 1:1 randomization was used to allocate participants into the control group that received the standard MomConnect maternal health messages or the experimental group that received the Teen MomConnect intervention. A subset of experimental group participants received an in-person motivational interviewing session. Questionnaires were administered at baseline and after the end of the participants' pregnancies. Appointment attendance data were obtained from clinic records. ANOVA, ANCOVA, and logistic regression models assessed the differences in appointments attended, awareness of HIV status, and the psychosocial determinants of antenatal attendance between the control and experimental groups.

Results: Overall, 412 adolescent girls and young women were enrolled, of which 254 (62%) completed the posttest survey (64% control, 59% intervention). Patient record data were obtained for 222 of the 412 (54%; in both control and intervention) participants. A total of 84% (63/75) and 72% (54/75) rated the intervention messages highly regarding their content value and their motivational nature for behavior change, respectively. Participants responded to an average of 20% of the 2-way messages they received. Mean appointment attendance did not differ significantly between the experimental (4.86, SD 1.76) and control (4.79, SD 1.74; *P*=.79) groups. Appointment attendance was higher among intervention participants who responded to ≥50% of messages ("high-responders"; 5.08, SD 1.66) than intervention participants who responded to fewer messages (4.82, SD 1.79) and control participants (4.79, SD 1.74; *P*=.86). The mean increase in knowledge scores was significantly higher among experimental group high-responders (2.1, SD 3.17) than the control group (0.7, SD 2.73; β=1.50; *P*=.045).

XSL•FO
RenderX

Conclusions: Engagement with the intervention's 2-way messaging was low, which could have impacted the outcomes. However, the intervention content was deemed acceptable. Appointment attendance did not vary significantly between the intervention and control groups. More intensive intervention may be needed to impact appointment adherence.

Trial Registration: Pan African Clinical Trial Registry (PACTR) PACTR201912734889796; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=9565

International Registered Report Identifier (IRRID): RR2-10.2196/43654

(JMIR Mhealth Uhealth 2025;13:e59144) doi:10.2196/59144

KEYWORDS

antenatal care; adolescent girls and young women; adolescent pregnancy; mobile health; mHealth; tailoring; motivational interviewing; South Africa; mobile phone; adolescent; teens; youth; young women; pregnancy; pregnant adolescent; low income country; middle income country; antenatal; pilot randomized controlled trial; design; user acceptability; psychosocial

Introduction

South Africa has a high prevalence of pregnancy among adolescent girls and young women. The South African Demographic and Health Survey in 2016 found that 27.8% of 19-year-olds had begun childbearing [1]. This is of public health concern due to the higher risk of pregnancy complications and maternal and infant mortality for adolescent girls and young women compared to older adult pregnant women [2-4]. Antenatal care (ANC) and healthy behaviors during pregnancy are vital for healthy pregnancy outcomes. Early and routine attendance of ANC facilitates prevention, detection, management of risk factors and pregnancy complications, and the delivery of pregnancy health education [5]. However, adolescent girls and young women in South Africa generally have lower ANC attendance, and they book in later along their pregnancy for their first antenatal visit than older pregnant women [6,7]. Almost a quarter of pregnant adolescents do not attend at least four antenatal visits [8].

There is therefore a need for behavioral interventions to improve antenatal attendance and health behaviors during pregnancy among adolescent girls and young women. Mobile health (mHealth) behavior change interventions are on the rise and have the power to reach large numbers of people [9,10]. Previous mHealth interventions using SMS text messages have increased ANC uptake among pregnant women in low- and middle-income countries [11]. The South African National Department of Health's MomConnect program [12] is one of the largest mHealth interventions in the world, having cumulatively reached over 1.7 million pregnant women between 2014 and 2017 [13]. Pregnant women are routinely registered on MomConnect when they attend ANC at public health facilities throughout South Africa. They then receive SMS text messages with maternal health information during various stages of their pregnancy, up to 1 year post birth. The messages were founded upon maternal health SMS text messages in the Mobile Alliance for Maternal Action in South Africa, which were previously shown to improve antenatal attendance in adult women [14]. MomConnect has successfully registered 63% of women attending their first antenatal appointment in South Africa [13].

The MomConnect messages are designed more for adult pregnant women, and opt-out rates of MomConnect are higher for women younger than 25 years of age [15]. We recently developed a mHealth intervention targeted specifically for pregnant adolescent girls and young women that aimed to improve antenatal attendance and its determinants, as well as health behaviors during pregnancy. The intervention, titled Teen MomConnect, included age-appropriate, tailored motivational behavior-change and reinforcement messages with iterative two-way message feedback on behaviors. Teen MomConnect was embedded in an SMS text messaging–based system similar to MomConnect. The Teen MomConnect intervention was also designed to include a motivational interviewing (MI) counseling session delivered by a trained research assistant. The development and description of the intervention and the protocol for a pilot two-arm randomized controlled trial (RCT) to evaluate the intervention can be found elsewhere [16]. The aim of this paper is to evaluate the user acceptability of the Teen MomConnect intervention and to assess its preliminary efficacy on the primary outcome of antenatal attendance. Secondary outcomes reported are awareness of HIV status, tobacco and alcohol use, and the psychosocial determinants of antenatal attendance, namely knowledge, risk perceptions, attitudes, and self-efficacy regarding antenatal attendance.

Methods

Study Design

We conducted a pilot 2-arm RCT with a pre-post design (registered at the Pan African Clinical Trial Registry with trial number PACTR201912734889796). The study follows the CONSORT (Consolidated Standards of Reporting Trials) statement on randomized pilot and feasibility trials (Multimedia Appendix 1) [17]. Participants were randomized into an experimental and a usual care control condition, where the experimental condition received the new Teen MomConnect intervention while the control condition received the standard MomConnect mobile SMS text messaging. The study was conducted in Cape Town, South Africa, from April 2018 to December 2019. Participants completed a posttest questionnaire after the end of their pregnancy, which included similar items to the baseline questionnaire and additional items on user acceptability. Patient records of the participants were retrieved from the health facilities after the end of their pregnancy.

XSL•FO
RenderX

Participants and Procedure

ANC services are available free of charge at primary health care facilities within South Africa's public health care system, which serves over 83% of the population [18]. Pregnant adolescent girls and young women aged 13-20 years of all gestational ages and who had access to a mobile phone were recruited to participate from a list of priority clinics and surrounding communities in Cape Town, South Africa. Discussions with the Western Cape Provincial Department of Health led to the identification of 16 facilities that provide ANC, where the number of pregnancies among adolescent girls and young women in the surrounding communities was highest, and the socioeconomic circumstances in those areas were challenging. Participants were approached and invited to participate while attending ANC at these facilities. Furthermore, participant recruitment was conducted in the surrounding communities by word of mouth. Participants were excluded from enrollment in the study if any one or more of their estimated date of delivery (EDD), language, name, age, and mobile number were missing or if they were already registered on the original version of MomConnect.

The initial inclusion criteria were participants up to 24 weeks pregnant. However, substantial numbers of pregnant adolescent girls and young women were more than 24 weeks pregnant at the time of being approached during recruitment. The inclusion criteria were therefore amended to include pregnant adolescent girls and young women of all gestational ages. Participant recruitment was conducted from May to December 2018. For further details on the recruitment procedure, see the study protocol [16].

A sample of 200 (100 participants per group) was targeted. It was determined that this sample size would allow us to detect a difference of 0.52 visits between the treatment groups at posttest measurement, that is, a mean of 4.1 visits in the intervention group and 3.58 visits in the control group with a pooled SD of 1.3 visits. This translated to a Cohen *d* effect size of at least 0.40 (0.52/1.3). Sample calculations applied 80% power and a type I error of 5%. We anticipated large dropout rates as observed in many adolescent studies. Given the fact that a large number of prospective participants would be expected to have missing contact details and pregnancy characteristics or to be already registered on MomConnect and would therefore be excluded from registration onto the mobile intervention, it was decided to recruit and screen three times as many participants for the baseline survey.

Recruited participants completed a consent form (for participants aged 18 years or older) or assent form (for participants younger than 18 years of age) before being enrolled into the study. In addition, parent consent forms were completed for participants aged younger than 18 years. Consenting individuals completed a self-administered baseline questionnaire using an electronic tablet or mobile phone provided by the research assistants.

Information on demographic characteristics, name, mobile phone number, age, EDD, estimated last menstrual date (LMD), and current enrollment on MomConnect was extracted from the baseline questionnaire. This information was used to determine the individuals who could be enrolled in the study. Participants were asked for the number of their own mobile phone or, alternatively, the number of a mobile phone they shared or which they could use to access the messages sent during the study. User registration to receive the MomConnect and Teen MomConnect messages required information on the user's name, mobile phone number, language, and EDD. The EDD was necessary for the participants to receive messages tailored to the pregnancy week. In cases where the EDD was missing, an estimate was determined using the LMD. When both dates were missing, the participant was contacted telephonically numerous times to find either their EDD or LMD. Participants indicated their preferred language for their messages. The language choices were English, Afrikaans, and isiXhosa, which are the three primary languages spoken in the Western Cape province.

We used simple 1:1 randomization to allocate enrolled participants to the experimental group that received the Teen MomConnect intervention and the control group that received the standard MomConnect messages. Generation of the random allocation sequence and allocation of participants into intervention or control groups were performed by a statistician on the research team. The study was blinded for the participants because they were not told which group they were allocated to. We registered participants into the Teen MomConnect and MomConnect messaging programs in weekly or biweekly recruitment batches. Toward the end of study enrollment, we needed to randomize participants unequally to the control group to correct for a minor mistake in the randomization schema. There was a delay in turning off the unequal randomization, which resulted in a slightly uneven number of participants by treatment group at the end of study enrollment.

Participants were contacted telephonically during September-December 2018 to schedule a face-to-face MI session. Completing MI sessions between enrollment and the end of pregnancy proved to be challenging as some participants did not answer their phones after multiple calls, some failed to attend their scheduled sessions, many could not make time to attend a session due to school or other commitments, and there were difficulties with grouping participants into sessions at times and days when they were available. We therefore decided to conduct MI sessions until a third of the participants had received an MI session. The subset of MI participants was thus quota-sample driven. This would facilitate a subgroup analysis of participants within the intervention group who completed and did not complete MI, although this was not by random allocation.

After each participant's EDD had passed, they were programmed to receive the standard MomConnect postpartum messages about caring for the newborn and postpartum maternal health. The research assistants contacted participants from a week after their EDD had passed to schedule administration of the posttest questionnaire. The posttest questionnaire administration followed the same procedure as that of the baseline. To encourage participation, incentives were provided of $3 for completing each of the baseline and posttest questionnaires and $6 for attending an MI session. At the time of writing the paper, the conversion rate was 16.7 South African Rand to US $1.

XSL•FO
RenderX

We also visited the facilities where the participants were receiving ANC to retrieve information from their patient folders. The folders contained a maternity case booklet and related information sheets that contained data from the booking appointment and subsequent antenatal appointments, as well as information recorded about the birth outcomes. The dates of the appointments attended were summed to calculate the total number of appointments attended. Maternal health care workers usually record this information in the patient records during the antenatal appointments and during and after labor.

Intervention

The development of the intervention content was informed by formative qualitative research with pregnant adolescents about their experiences with MomConnect and with antenatal attendance, reviews of maternal health education resources, the literature on factors associated with health care seeking in pregnant adolescent girls and young women, a review of the existing MomConnect messages, and various stakeholder consultative discussions. More details about the development of the intervention can be found elsewhere [16]. Both intervention and control messaging programs used an SMS text messaging platform where participants were registered using Unstructured Supplementary Service Data [15]. An SMS text messaging platform was chosen for Teen MomConnect to facilitate comparison with MomConnect, which, at the time of intervention development in 2016, used an SMS text messaging platform. Discussions with the Department of Health revealed that SMS text messaging was preferred over a mobile app at the time because SMS text messaging did not require access to smartphones, and users would not incur data or other costs from sending and receiving SMS text messages.

The Teen MomConnect intervention content comprised an SMS text message library of 66 messages that covered 13 behavioral domains, as well as an MI counseling session. The domains were determined as being relevant to pregnancy among adolescent girls and young women in the South African context and were: appointment keeping, attitudes toward pregnancy, information about clinic attendance, general health information, alcohol, illegal drugs, smoking, nutrition, self-care and hygiene, social stigma regarding adolescent pregnancy, social support, sexually transmitted diseases, and tuberculosis. The number of messages received per participant was dependent on their pregnancy week, join week, and their responses to the tailored questions. The length of messages was restricted to 160 characters, consistent with the maximum length of an SMS text message.

The message library included 48 fixed messages and 18 two-way tailored SMS text message sets. The fixed messages did not elicit a response from the participant, while the two-way messages asked the participant to respond to a closed-ended question. The participant could respond to the question by sending back an SMS text message with the letter corresponding to their response option. The participant then received a feedback message that was tailored to their response. A customized response was also sent when the participant did not respond to the message. Participants who had missed messages from earlier pregnancy weeks were sent relevant "catch-up" messages of the ones they had missed. The language was motivational and was designed to reinforce behaviors. The development of messages was grounded in self-determination theory [19] and was based on techniques for tailoring health promotion messages used in previous studies [20-22].

There were layered, two-way tailored message sets on appointment keeping that were repeated every 4 weeks. Therefore, if a participant joined the intervention at 14 weeks pregnant, for example, they would receive six sets of appointment-keeping messages over 22 weeks. The appointment-keeping message set included asking about whether the participant knew when their next appointment was, tailored appointment reminders every 4 weeks for those who did and did not know their appointment dates, checkups of whether the appointments were actually attended, asking about the reasons for not attending their last appointment, and tailored feedback addressing various reasons for not attending each appointment. In addition, there were five fixed messages on information about appointment attendance, and the messages from other domains, like substance use, general health information, nutrition, self-care, social stigma, sexually transmitted diseases, and tuberculosis, also included content that emphasized seeking further advice, treatment, or testing from a clinic.

Messages asking about smoking, alcohol, and illegal drug use were sent at enrollment, and participants who reported past-month use of these substances were sent tailored messages every 4 weeks to follow up on the status of their substance use and motivate reduced use for those using. Examples of messages and their logic flows and timings are detailed in the study protocol [16].

A single face-to-face MI counseling session was administered by a trained research assistant during a participant's pregnancy. MI is a style of counseling where the participant reflects on reasons for and against changing their behavior, how their behavior aligns with their goals and values, and opportunities for changing their behavior [23,24]. Research assistants underwent a 9-hour training workshop in MI from an expert MI counselor, which covered the skills of open-ended questions, positive affirmations, reflective listening, and summarizing. During the training, the research assistants role-played an MI session to practice their skills, and they were assessed and critiqued by the expert trainer. The main content areas covered in the participants' MI sessions included appointment keeping, substance use, HIV and tuberculosis testing, medication adherence, nutrition and overall health, and pregnancy behaviors. Each session was tailored to the participant's specific behaviors and questions so that a session could focus more on some behaviors than others. Depending on their availability, participants could schedule a one-on-one session or a group session with up to five participants.

Measures

Demographic Variables

The baseline questionnaire assessed age, race (Black African, Coloured, Other), currently attending an educational institution such as school or a tertiary college (yes, no), dwelling type (informal, formal), self-perceived cost of living (food and

XSL•FO
RenderX

clothes shortage, shortage of other important things, have important basics), and having previously been pregnant (yes, no). The EDD and LMD from the baseline questionnaire were used to calculate the gestational age at baseline.

Baseline Psychosocial Determinants of Antenatal Attendance

The baseline questionnaire included items on the following determinants: knowledge about antenatal attendance and pregnancy, risk perceptions, social support for attending ANC; peer, family, partner, and individual participant attitudes toward antenatal attendance; and self-efficacy, intention, and action-planning to attend antenatal appointments. Knowledge statements were informed by maternity booklets. The attitude, self-efficacy, peer norm, and intention items were informed by the reasoned action approach [25], and the action planning, risk perception, and social support items were informed by the I-Change Model [26]. The knowledge items were measured where 1=correct and 0=incorrect, and the other determinant items were measured on 4- or 5-point Likert scales with responses ranging from strongly disagree to strongly agree. Previous publications provide further details on the items measured per determinant [16,27].

Exploratory principal component factor analysis was performed on the items to derive scales for each determinant. The sum score of the items was used as the scale. Multimedia Appendix 2 shows the determinant scales and their reliability information. The McDonald's omega total is often presented as a more accurate approximation of a scale's reliability than the Cronbach α [28]. However, both indices are used as indicators of scale structure [29].

Outcomes

The number of appointments attended by the end of the pregnancy was extracted from the patient records and was the primary outcome to assess preliminary efficacy. In addition to appointments attended as a continuous variable, the number of antenatal appointments attended was also recoded into a binary variable for <4 versus ≥4 to distinguish the participants who attended the World Health Organization's (WHO's) basic ANC recommendation of at least four appointments [30]. While the WHO increased its recommended number of ANC appointments to eight contacts in recent years [31], international reporting on ANC coverage still focuses on the 4-visit indicator [8,32], and the new guidelines were relatively new to South Africa at the time of this study. We therefore reported on the percentage attending ≥4 appointments.

Secondary outcomes were self-reported awareness of HIV status (yes or no) postpartum, self-reported postpartum alcohol use in the past month among those who reported alcohol use at baseline; self-reported postpartum tobacco smoking in the past month among those who reported tobacco smoking at baseline; changes in three knowledge scale scores between baseline and posttest; and risk perceptions, positive participant attitudes, negative participant attitudes, and self-efficacy regarding antenatal attendance at posttest after the end of the pregnancy. The three knowledge scales assessed the risks of not attending ANC; preparation for childbirth, substance use, and sexually transmitted infections; and misinformation about routine antenatal and pregnancy care, respectively. The same knowledge items were asked in both baseline and posttest questionnaires. However, the other psychosocial determinant items were asked in the past tense form in the posttest questionnaire. For example, "attending clinic appointments is important for me" at baseline was rephrased to "attending clinic appointments was important for me." This meant that the items were not directly comparable between the two time points, and the differences in their scale scores between baseline and posttest could not be computed. The reliability information on the outcome scales (knowledge, risk perceptions, attitudes, and self-efficacy) is presented in Multimedia Appendix 2.

The trial preregistration included the number of HIV tests as a secondary outcome, and it did not include awareness of HIV status. It appeared later that the number of HIV tests administered at a facility was not straightforwardly captured in the patient records. Self-reported awareness of HIV status was also deemed a more appropriate measure of awareness of one's health than the number of HIV tests conducted per participant. Furthermore, the trial preregistration did not list knowledge, risk perceptions, attitudes, and self-efficacy as secondary outcomes.

The feasibility outcomes were satisfaction and user acceptability of the Teen MomConnect messages and the MI counseling session, as well as the number of intervention messages responded to and self-reported message usability. A total of 21 items and 14 items assessed the SMS text messages and the MI session, respectively. The items evaluating the SMS text messages asked about understandability, appropriateness, and motivation to change behaviors, while the items evaluating the MI session asked about perceptions and feelings after attending the session and motivation to change behaviors. All the items were measured on a Likert scale from 0=strongly disagree to 4=strongly agree with each statement. Exploratory principal component factor analyses were used to determine scales for each of the SMS text message and MI evaluations (Multimedia Appendix 3). The items per scale were used to create a sum score, where higher values indicated more positive evaluations of each component. The ranges of possible scores were categorized into tertiles corresponding to low, moderate, and high scores.

Ethical Considerations

Ethical approval for the study was obtained from the Research Ethics Committee of the Human Sciences Research Council (REC 2/17/08/16), which is aligned with the Declaration of Helsinki. In addition, permissions were granted from the Provincial Government of the Western Cape and the National Department of Health to access the public health facilities. Written informed consent was obtained from participants aged 18 years or older. Written assent and written informed parental consent were obtained for participants younger than 18 years of age. The consent forms informed participants and parents of voluntary participation, the anonymity and confidentiality of their responses, and the right to withdraw from the study at any time.

XSL•FO
RenderX

Statistical Analysis

Descriptive statistics were used to describe the characteristics of both the full sample enrolled at baseline and the analytic sample for whom data on appointments attended were available after removal of dropouts. The differences in characteristics between the groups were assessed using 2-tailed *t* tests for continuous variables and chi-square tests for discrete variables. Dropouts included participants who had a miscarriage or abortion, who opted out of the messaging platforms within the first few days, and for whom there were technical issues with registering or receiving message content. These dropouts were excluded from the analysis for efficacy and user acceptability. Descriptive statistics were also used to characterize the control and experimental conditions for whom appointment data was available. The number of antenatal appointments attended; the differences in knowledge scores between baseline and posttest; and the risk perception, positive attitude, negative attitude, and self-efficacy scores at posttest were compared between the experimental and the control group using one-way ANOVA and ANCOVA. The covariates adjusted for were age, gestational age, race, attendance at an educational institution, and self-perceived cost of living (a proxy for socioeconomic status). These covariates were informed by the literature on sociodemographic factors associated with antenatal attendance [3,33,34]. In addition, sociodemographic variables whose distributions differed significantly between the full sample and the subsample for whom appointment attendance data were available were also included as covariates. For the ANCOVA models for risk perceptions, positive attitudes, negative attitudes, and self-efficacy at posttest, their respective baseline scores were also included as covariates in the models. The underlying regression coefficients from the ANOVA and ANCOVA models were reported with 95% CIs. Simple logistic regression and multiple logistic regression adjusting for covariates were used to compare the following outcome measures between the experimental and control group: the percentage of participants who attended at least four appointments, the percentage who were aware of their HIV status at the end of their pregnancy, and the percentages of postpartum tobacco and alcohol use during the past month among those who reported using those substances at baseline. As part of the sensitivity analyses, the ANOVA and ANCOVA analyses were performed on the primary outcome of the number of appointments attended after multiple imputation via chained equations was performed to impute missing data on the number of appointments attended.

The treatment condition was recoded into three categories: control group, experimental group high responders (responded to ≥50% of the Teen MomConnect messages), and experimental group low responders (responded to <50% of the Teen MomConnect messages), to assess the possible dose-response effect of the level of engagement with the Teen MomConnect messaging content on the outcome variables. The treatment condition was also recoded into three categories: control group, experimental group who did not receive MI counseling, and experimental group who received MI counseling, to assess the effect of the intervention with and without MI on the outcome variables. These were exploratory analyses to supplement the primary intention-to-treat analyses. The same ANOVA, ANCOVA, and logistic regression analyses for the outcome variables were then performed using each of the condition-by-level-of-engagement and the condition-by-MI variables to investigate whether participants who engaged more with the intervention messages or who received MI had higher outcome values. The level of engagement and MI was not investigated for the outcomes of postpartum tobacco smoking and alcohol use because the number of participants who reported baseline use of these substances was too small.

Descriptive statistics were used to analyze the user acceptability of the Teen MomConnect messages and MI sessions, the number of Teen MomConnect messages responded to, and self-reported message usability for the participants enrolled in the intervention.

Results

Participation and Attrition

Figure 1 illustrates the participant flow. Recruitment occurred during May-December 2018, and posttest surveys were administered postpartum during October 2018-July 2019. A total of 626 adolescent girls and young women aged 13-20 years were assessed for eligibility and completed the baseline survey. Of these adolescent girls and young women, 412 participants were enrolled in the study, where 219 were randomized into MomConnect and 193 into Teen MomConnect. After the end of the pregnancy, 254 (62%) completed the posttest questionnaire (140/219, 64% MomConnect and 114/193, 59% Teen MomConnect). Patient records were successfully reviewed for 222 (54%) of the 412 enrolled participants (118/219, 54% MomConnect and 104/193, 54% Teen MomConnect). After removing records of participants who reported a miscarriage or abortion, who opted out of the Teen MomConnect service within the first week, and participants with technical issues later discovered with registering onto the messaging platforms, posttest data were analyzed for 241 participants (132/219, 60% MomConnect and 109/193, 56% Teen MomConnect). Furthermore, after removing records with incomplete data on appointments attended in the patient records, data on appointment attendance were analyzed for 194 participants (104/219, 47% MomConnect and 90/193, 47% Teen MomConnect).

XSL•FO
RenderX

Figure 1. Participant flow and attrition. *Numbers do not add to 14 records because 1 participant had missing data on appointments attended and had a miscarriage. EDD: estimated date of delivery.

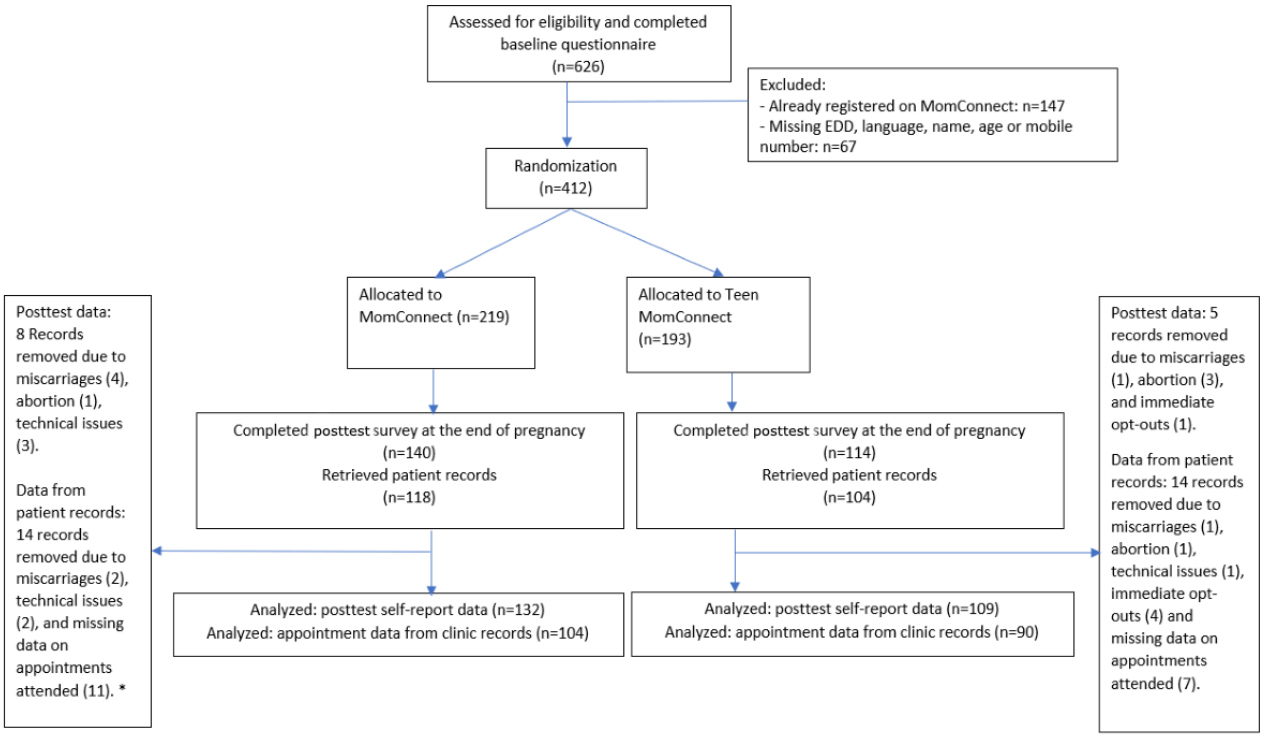


Table 1 shows the baseline characteristics of the full sample (n=412) and the sample for whom there were data on appointments (n=194). The sample with appointment attendance data differed significantly from the full sample for the following baseline variables: age (mean 17.8, SD 1.6 vs 18.0, SD 1.6 years), gestational age (mean 22.7, SD 7.3 vs 22.0, SD 7.4 weeks), race group (40/194, 20.6% vs 123/410, 30% for Black African race), attendance at an educational institution (80/194, 41.2% vs 149/412, 36.2%) and knowledge of the risks of not attending ANC or not practicing healthy behaviors (mean score 3.8, SD 2.5 vs 3.4, SD 2.3).

XSL•FO
RenderX

Table 1. Characteristics of the full baseline sample and the analytic sample with data on appointment attendance.

	Full baseline sample (n=412)	Analytic sample with data on appointments (n=194)	*P* value
Age (years), mean (SD)	18.0 (1.6)	17.8 (1.6)	.01[a]
Gestational age (weeks), mean (SD)	22.0 (7.4)	22.7 (7.3)	.086
Race group[b], n (%)			.001[a]
Black African	123 (30)	40 (20.6)	
Coloured	285 (69.5)	153 (78.9)	
Other	2 (0.5)	1 (0.5)	
Attending an educational institution, n (%)			.04[a]
Do not attend an educational institution	263 (63.8)	114 (58.8)	
Attend an educational institution	149 (36.2)	80 (41.2)	
Dwelling type, n (%)			.54
Informal	100 (24.6)	44 (23.2)	
Formal	307 (75.4)	146 (76.8)	
Self-perceived cost of living, n (%)			.82
Food and clothes shortage	92 (22.7)	41 (21.5)	
Shortage of other important things	150 (37)	73 (38.2)	
Have important basics	164 (40.4)	77 (40.3)	
Had been pregnant before, n (%)			.31
Yes	45 (10.9)	18 (9.3)	
No	367 (89.1)	176 (90.7)	
Past month tobacco smoking, n (%)	162 (39.4)	80 (41.5)	.43
Past month alcohol use, n (%)	119 (29)	56 (29)	.98
Knowledge score 1[c] (range: 0-7), mean (SD)	3.8 (2.5)	3.4 (2.3)	.001[a]
Knowledge score 2[d] (range: 0-4), mean (SD)	3.4 (1.0)	3.3 (1.1)	.08
Knowledge score 3[e] (range: 0-3), mean (SD)	1.2 (1.0)	1.2 (1.0)	.69
Risk perceptions (range: 5-25), mean (SD)	16.3 (2.0)	16.4 (2.1)	.41
Positive participant attitudes toward ANC[f] attendance (range: 7-35), mean (SD)	29.7 (5.2)	29.5 (5.3)	.41
Negative participant attitude toward ANC attendance (range: 5-25), mean (SD)	9.0 (3.8)	8.7 (3.6)	.10
Social support score (range: 3-12), mean (SD)	10.2 (1.8)	10.4 (1.6)	.17
Positive peer attitudes toward ANC attendance (range: 2-8), mean (SD)	6.4 (1.1)	6.4 (1.1)	.69
Negative peer attitudes toward ANC attendance (range: 3-12), mean (SD)	8.9 (1.7)	8.9 (1.8)	.78
Positive family attitudes toward ANC attendance (range: 4-20), mean (SD)	17.1 (2.7)	17.1 (2.6)	.86
Negative family attitudes toward ANC attendance (range: 3-15), mean (SD)	6.0 (2.2)	5.8 (2.1)	.09
Positive partner attitudes toward ANC attendance (range: 3-12), mean (SD)	10.0 (1.7)	10.0 (1.6)	.79
Self-efficacy score regarding ANC attendance (range: 8-32), mean (SD)	25.8 (5.0)	25.4 (4.8)	.22
Intention to attend ANC (range: 5-20), mean (SD)	16.8 (2.7)	16.7 (2.6)	.61
Action planning to attend ANC (range: 4-16), mean (SD)	12.1 (1.6)	12.0 (1.5)	.19

[a]*P*<.05.

XSL•FO
RenderX

[b]Race was self-reported by participants. Racial categories provided in the questionnaire were according to Statistics South Africa's standard population groups: Black African, Coloured, White, and Indian. The White and Indian categories were collapsed for analysis due to small sample sizes. Race was reported not with the intention of reifying sociocultural constructs, but rather to study ongoing health disparities across groups.

[c]Knowledge score 1: knowledge of risks of not attending antenatal care or not practicing healthy behaviors.

[d]Knowledge score 2: Knowledge about preparation for childbirth, substance use, and sexually transmitted infections.

[e]Knowledge score 3: Misconceptions about antenatal care and pregnancy behaviors.

[f]ANC: antenatal care.

Appointment Attendance and Related Determinants

Table 2 shows the estimates of appointment attendance, HIV awareness, and the psychosocial determinants of antenatal attendance after the end of pregnancy between the control and experimental groups. It further shows the regression coefficients from the underlying ANOVA and ANCOVA models for continuous outcomes and the odds ratios from the logistic regression models for the binary outcomes.

Table 2. Estimates of appointment attendance, risk behaviors, and the psychosocial determinants of antenatal attendance after the end of pregnancy between the control and experimental groups.

Primary outcome	Control	Intervention	Coefficient, OR[a] (95% CI)	Adjusted coefficient, AOR[b] (95% CI)[c]
Number of appointments attended, mean (SD)	4.8 (1.7)	4.9 (1.8)	0.07 (–0.43 to 0.57)	0.02 (–0.50 to 0.55)
Attended at least 4 appointments, n (%)	78 (75)	69 (77)	1.10 (0.57 to 2.12)	1.00 (0.49 to 2.06)
Risk perception sum score at posttest, mean (SD)	16.0 (2.4)	15.6 (1.9)	–0.43 (–1.02 to 0.15)	–0.49 (–1.11 to 0.12)
Positive participant attitudes sum score toward ANC[d] attendance at posttest, mean (SD)	30.0 (4.8)	29.1 (5.4)	–0.89 (–2.22 to –0.44)	–0.38 (–1.73 to 0.97)
Negative participant attitudes sum score toward ANC attendance at posttest, mean (SD)	8.4 (2.8)	9.5 (4.0)	1.01 (0.13 to 1.90)	0.48 (–0.41 to 1.37)
Self-efficacy to attend ANC sum score at posttest, mean (SD)	26.5 (4.4)	25.2 (4.4)	–1.31 (–2.54 to –0.07)	–1.30 (–2.57 to –0.03)
Change in knowledge score 1[e] between baseline and posttest survey, mean (SD)	0.7 (2.7)	0.8 (3.2)	0.08 (–0.72 to 0.89)	0.07 (–0.75 to 0.90)
Change in knowledge score 2[f] between baseline and posttest survey, mean (SD)	0 (1.3)	0.1 (1.3)	0.14 (–0.21 to 0.48)	0.14 (–0.22 to 0.50)
Change in knowledge score 3[g] between baseline and posttest survey, mean (SD)	0.3 (1.2)	0.5 (1.3)	0.24 (–0.09 to 0.56)	0.23 (–0.10 to 0.56)
Self-reported awareness of HIV status at posttest, n (%)	120 (93.8)	99 (94.3)	1.10 (0.37 to 3.28)	1.33 (0.41 to 4.31)
Tobacco smoking at posttest among participants who reported tobacco smoking at baseline, n (%)	32 (65.3)	26 (65.0)	1.01 (0.42 to 2.43)	0.68 (0.25 to 1.86)
Alcohol use at posttest among participants who reported alcohol use at baseline, n (%)	8 (21.1)	11 (35.5)	0.48 (0.17 to 1.42)	0.32 (0.09 to 1.10)

[a]OR: odds ratio.

[b]AOR: adjusted odds ratio.

[c]Adjusted for age, gestational age, perceived cost of living, race group, and attendance at an educational institution. The ANCOVA models for risk perception, positive and negative participant attitudes, and self-efficacy outcomes were also adjusted for their respective baseline scores.

[d]ANC: antenatal care.

[e]Knowledge score 1: knowledge of risks of not attending antenatal care or not practicing healthy behaviors.

[f]Knowledge score 2: knowledge about preparation for childbirth, substance use, and sexually transmitted infections.

[g]Knowledge score 3: misconceptions about antenatal care and pregnancy behaviors.

The number of appointments attended did not differ significantly between the Teen MomConnect intervention group (mean 4.9, SD 1.8) and the MomConnect control group (mean 4.8, SD 1.7) in both the unadjusted (β=0.07; *P*=.79) and adjusted models

XSL•FO
RenderX

(β=0.02; *P*=.93). The percentage of participants who attended at least four appointments did not differ significantly between the intervention (69/90, 77%) and control groups (78/104, 75%), nor did risk perceptions, positive participant attitudes, negative participant attitudes, increase in the knowledge scores, and awareness of HIV status. Among participants who reported baseline use of tobacco (n=89) and alcohol (n=69), past-month use of each of these substances did not differ significantly between the intervention and control groups. Self-efficacy scores to attend appointments at posttest were significantly higher in the control group than in the intervention group after adjusting for the baseline score and the other covariates.

Of the participants in the intervention group, 14.3% (n=26) responded to ≥50% of the Teen MomConnect messages and were termed as the "high-responders." The average number of appointments attended was 5.1 (SD 1.7) among participants in the intervention group who were high responders, 4.8 (SD 1.8) among intervention group participants who responded to <50% of the Teen MomConnect messages ("low-responders"), and 4.8 (SD 1.7) among control group participants (Table 3). However, these differences were not statistically significant in the unadjusted or adjusted models. The average increase in knowledge about the risks of not attending ANC was significantly higher among participants in the intervention group who were high-responders (2.1, SD 3.2) than those in the control group (0.7, SD 2.7; β=1.50; *P*=.045). There were no significant associations between the condition-by-level-of-engagement variable and the other outcome variables after adjusting for covariates.

XSL•FO
RenderX

Table 3. Estimates of appointment attendance, risk behaviors, and the psychosocial determinants of antenatal attendance after the end of pregnancy between the control group, intervention group low-responders, and intervention group high-responders.

	Control	Intervention—low message response	Intervention—high message response	Unadjusted model		Adjusted model[a]	
				Coefficient, OR[b] for intervention—low-message response (95% CI)	Coefficient, OR for intervention—high-message response (95% CI)	Coefficient, AOR[c] for intervention—low-message response (95% CI)	Coefficient, AOR for Intervention—high-message response (95% CI)
Number of appointments attended, mean (SD)	4.8 (1.7)	4.8 (1.8)	5.1 (1.7)	0.03 (–0.49 to 0.55)	0.29 (–0.73 to 1.31)	–0.04 (–0.59 to 0.51)	0.38 (–0.65 to 1.41)
Attended at least 4 appointments, n (%)	75 (78)	59 (77)	10 (77)	1.09 (0.55 to 2.18)	1.11 (0.28 to 4.35)	0.98 (0.46 to 2.07)	1.17 (0.28 to 4.91)
Risk perception sum score at posttest, mean (SD)	16.0 (2.4)	15.5 (2.0)	15.8 (1.4)	–0.48 (–1.09 to 0.14)	–0.22 (–1.35 to 0.91)	–0.53 (–1.19 to 0.12)	–0.31 (–1.47 to 0.85)
Positive participant attitudes sum score toward ANC[d] attendance at posttest, mean (SD)	30.0 (4.8)	29.0 (5.8)	29.4 (3.1)	–0.96 (–2.38 to 0.45)	–0.53 (–3.04 to 1.98)	–0.44 (–1.88 to 1.00)	–0.11 (–2.56 to 2.34)
Negative participant attitudes sum score toward ANC attendance at posttest, mean (SD)	8.4 (2.8)	9.4 (4.2)	9.7 (3.1)	0.95 (0.01 to 1.89)	1.29 (–0.35 to 2.94)	0.33 (–0.62 to 1.28)	1.05 (–0.54 to 2.64)
Self-efficacy to attend ANC sum score at posttest, mean (SD)	26.5 (4.4)	25.4 (4.7)	24.6 (2.1)	–1.18 (–2.48 to 0.12)	–1.96 (–4.41 to 0.48)	–1.27 (–2.60 to 0.07)	–1.49 (–4.03 to 1.05)
Change in knowledge score 1[e] between baseline and posttest survey, mean (SD)	0.7 (2.7)	0.5 (3.2)	2.1 (3.2)	–0.2 (–1.04 to 0.64)	1.38 (–0.10 to 2.86)	–0.27 (–1.13 to 0.60)	1.5 (0.03 to 2.96)
Change in knowledge score 2[f] between baseline and posttest survey, mean (SD)	0.0 (1.3)	0.1 (1.4)	0.3 (0.7)	0.1 (–0.27 to 0.47)	0.3 (–0.33 to 0.93)	0.09 (–0.29 to 0.48)	0.33 (–0.30 to 0.95)
Change in knowledge score 3[g] between baseline and posttest survey, mean (SD)	0.3 (1.2)	0.5 (1.3)	0.7 (1.3)	0.19 (–0.15 to 0.54)	0.43 (–0.17 to 1.03)	0.17 (–0.18 to 0.53)	0.45 (–0.15 to 1.04)

XSL•FO
RenderX

	Control	Intervention—low message response	Intervention—high message response	Unadjusted model		Adjusted model[a]	
				Coefficient, OR[b] for intervention—low-message response (95% CI)	Coefficient, OR for intervention—high-message response (95% CI)	Coefficient, AOR[c] for intervention—low-message response (95% CI)	Coefficient, AOR for Intervention—high-message response (95% CI)
Self-reported awareness of HIV status at posttest, n (%)	120 (93.8)	78 (93)	21 (100)	0.87 (0.29 to 2.59)	—[h]	1.02 (0.31 to 3.36)	—

[a]Adjusted for age, gestational age, perceived cost of living, race group, and attendance at an educational institution. The ANCOVA models for risk perception, participant attitudes, and self-efficacy outcomes were also adjusted for their respective baseline scores.

[b]OR: odds ratio.

[c]AOR: adjusted odds ratio.

[d]ANC: antenatal care.

[e]Knowledge score 1: knowledge of risks of not attending antenatal care or not practicing healthy behaviors.

[f]Knowledge score 2: knowledge about preparation for childbirth, substance use, and sexually transmitted infections.

[g]Knowledge score 3: misconceptions about antenatal care and pregnancy behaviors.

[h]Not applicable.

Sensitivity analyses, after multiple imputation of the number of appointments attended, showed no significant differences between the number of appointments attended between the Teen MomConnect intervention group and the control group (Multimedia Appendix 4). Furthermore, the number of appointments attended did not differ significantly by the condition-by-level-of-engagement.

A third (64/193, 33.2%) of the participants in the intervention received an MI counseling session. The average number of appointments was higher in the intervention group who received MI (5.4, SD 1.7) than the control group (4.8, SD 1.7) and the intervention group who did not receive MI (4.6, SD 1.8; Table 4). Risk perceptions were significantly lower among intervention participants who did not receive MI (15.3, SD 2.1) than control group participants (16.0, SD 2.4; β=–0.80; *P*=.04). Self-efficacy to attend appointments was significantly lower in the intervention group who received MI (24.7, SD 3.5) than in the control group (26.5, SD 4.4) after adjusting for the baseline self-efficacy score and the other covariates (β=–1.84; *P*=.03).

XSL•FO
RenderX

Table 4. Estimates of appointment attendance, risk behaviors, and the psychosocial determinants of antenatal attendance after the end of pregnancy between the control group, intervention group without MI[a], and intervention group with MI.

	Control	Intervention without MI	Intervention with MI	Unadjusted model		Adjusted model[b]	
				Coefficient, OR[c] for intervention without MI (95% CI)	Coefficient, OR for intervention with MI (95% CI)	Coefficient, AOR[d] for intervention without MI (95% CI)	Coefficient, AOR for intervention with MI (95% CI)
Number of appointments attended, mean (SD)	4.8 (1.7)	4.6 (1.8)	5.4 (1.7)	–0.23 (–0.79 to 0.34)	0.58 (–0.11 to 1.26)	–0.17 (–0.75 to 0.41)	0.42 (–0.32 to 1.17)
Attended at least 4 appointments, n (%)	78 (75)	41 (72)	28 (85)	0.85 (0.41 to 1.77)	1.87 (0.65 to 5.33)	0.9 (0.41 to 1.95)	1.36 (0.44 to 4.25)
Risk perception sum score at posttest, mean, (SD)	16.0 (2.4)	15.3 (2.1)	15.9 (1.6)	–0.71 (–1.44 to 0.01)	–0.13 (–0.87 to 0.61)	–0.8 (–1.56 to –0.05)	–0.15 (–0.94 to 0.63)
Positive participant attitudes sum score toward ANC[e] attendance at posttest, mean (SD)	30.0 (4.8)	28.8 (6.7)	29.5 (3.4)	–1.23 (–2.87 to 0.42)	–0.53 (–2.21 to 1.15)	–0.91 (–2.55 to 0.72)	0.21 (–1.48 to 1.91)
Negative participant attitudes sum score toward ANC attendance at posttest, mean (SD)	8.4 (2.8)	9.2 (4.4)	9.8 (3.6)	0.73 (–0.37 to 1.82)	1.32 (0.20 to 2.43)	0.24 (–0.84 to 1.32)	0.74 (–0.39 to 1.87)
Self-efficacy to attend ANC sum score at posttest, mean (SD)	26.5 (4.4)	25.6 (5.0)	24.7 (3.5)	–0.89 (–2.38 to 0.60)	–1.82 (–3.42 to –0.21)	–0.86 (–2.39 to 0.67)	–1.84 (–3.47 to –0.21)
Change in knowledge score 1[f] between baseline and posttest survey, mean (SD)	0.7 (2.7)	0.1 (3.3)	1.4 (3.1)	–0.60 (–1.58 to 0.39)	0.76 (–0.22 to 1.75)	–0.45 (–1.45 to 0.55)	0.63 (–0.39 to 1.65)
Change in knowledge score 2[g] between baseline and posttest survey, mean (SD)	0.0 (1.3)	0.0 (1.5)	0.2 (1.0)	0.04 (–0.39 to 0.47)	0.24 (–0.20 to 0.67)	0.05 (–0.39 to 0.49)	0.24 (–0.21 to 0.69)
Change in knowledge score 3[h] between baseline and posttest survey, mean (SD)	0.3 (1.2)	0.6 (1.3)	0.5 (1.3)	0.25 (–0.15 to 0.65)	0.22 (–0.18 to 0.63)	0.32 (–0.09 to 0.72)	0.13 (–0.28 to 0.55)
Self-reported awareness of HIV status at posttest, n (%)	120 (93.8)	49 (93)	50 (96)	0.82 (0.24 to 2.84)	1.67 (0.34 to 8.13)	0.87 (0.23 to 3.30)	2.35 (0.43 to 12.70)

[a]MI: motivational interviewing.

[b]Adjusted for age, gestational age, perceived cost of living, race group, and attendance at an educational institution. The ANCOVA models for risk perception, participant attitudes, and self-efficacy outcomes were also adjusted for their respective baseline scores.

[c]OR: odds ratio.

[d]AOR: adjusted odds ratio.

[e]ANC: antenatal care.

[f]Knowledge score 1: knowledge of risks of not attending antenatal care or not practicing healthy behaviors.

[g]Knowledge score 2: knowledge about preparation for childbirth, substance use, and sexually transmitted infections.

[h]Knowledge score 3: misconceptions about antenatal care and pregnancy behaviors.

User Acceptability of Teen MomConnect

Analysis of the messages sent and responded to on the Teen MomConnect SMS text messaging platform showed that, on average, participants in the Teen MomConnect intervention responded to 20% of the messages that asked them questions (Table 5). Over three-quarters (143/182, 78.6%) responded to less than a third of the messages that contained questions. The mean message response rate did not differ significantly by age group, gestational age, race, self-perceived cost of living, or educational attendance (P>.05). When asked if they were able to respond to messages, over 60% (52/86) reported that they

XSL•FO
RenderX

were able to respond always or most of the time. More than half (43/81, 53%) reported that they were able to respond to messages within 6 hours of receiving them, while 16% (13/81) were only able to respond to them after a day or longer. Among those who reported taking a long time to respond to messages, the most common reasons for not responding to messages were not checking their phones all the time (40/89, 49%) and their phones not being with them all the time due to phone sharing (31/89, 38%). Overall, 59% (52/88) reported that they would not prefer the intervention to be delivered by an app and therefore preferred SMS text messaging. In total, 84% (63/75) and 72% (54/75) rated the Teen MomConnect SMS text messaging content highly in terms of the value of the message content and the motivational nature of the messages for behavior change, respectively. The majority (64/81, 79%) expressed low negative opinions of the message content. A total of 64 (33.2%) Teen MomConnect participants received an MI counseling session. Of those who participated in the posttest survey and responded to the questions evaluating the session (n=44), 84% (37/44) and 77% (34/44) rated the MI session highly in terms of positive opinions about the MI and the motivational nature of the MI for behavior change for exercise and substance use, respectively. Among all participants (control and intervention), 22.7% (41/181) reported that they were using a different mobile number from the one used at baseline. There were no recorded harms or unintended effects among participants during the study period.

XSL•FO
RenderX

Table 5. Evaluation of the Teen MomConnect intervention content among participants enrolled in Teen MomConnect who participated in the posttest survey.

	Value
Were you able to reply to all the messages? n (%)	
Always	24 (27.9)
Most of the time	28 (32.6)
Sometimes	25 (29.1)
Never	9 (10.5)
Missing	23 (21.1)
How soon after seeing the messages did you reply to them? n (%)	
Within 6 hours	43 (53.1)
The same day	25 (30.9)
The next day or later	13 (16.1)
Missing	28 (25.7)
If you usually replied to messages long after they were sent, what were the reasons that took you so long to reply?[a] n (%)	
Phone is not with me all the time because I share it with someone else	31 (37.8)
Don’t check my phone all the time, and I see the messages much later	40 (48.8)
I see the messages, but I do not have time to reply to them	10 (12.2)
I thought that replying to the messages would use my airtime	8 (9.8)
Percentage (%) of messages responded to[b]	
Mean (SD)	20.0 (25.6)
<33%, n (%)	143 (78.6)
33%-66%, n (%)	24 (13.2)
>=66%, n (%)	15 (8.2)
Would you have preferred the messages be delivered via an app rather than SMS text messages? n (%)	
Yes	36 (40.9)
No	52 (59.1)
Is your mobile number still the same as at baseline?[c] n (%)	
Yes	140 (77.4)
No	41 (22.7)
Evaluations of the SMS text messages	
Value of the message content	
Sum score (range: 0-36; mean, SD)	27.3 (5.9)
Low (0-12), n (%)	2 (2.7)
Moderate (13-24), n (%)	10 (13.3)
High (25-36), n (%)	63 (84.0)
Missing, n (%)	34 (31.2)
Motivational nature of messages for behavior change	
Sum score (range: 0-20; mean, SD)	14.5 (3.9)
Low (0-6), n (%)	2 (2.7)
Moderate (7-13), n (%)	19 (25.3)
High (14-20), n (%)	54 (72.0)
Missing, n (%)	34 (31.2)
Negative opinions of the messages	

XSL•FO
RenderX

	Value
Sum score (range: 0-12; mean, SD)	3.0 (2.1)
Low (0-4), n (%)	64 (79.0)
Moderate (5-8), n (%)	15 (18.5)
High (9-12), n (%)	2 (2.5)
Missing, n (%)	28 (25.7)
Evaluations of the MI[d] session, n (%)	
Attended an MI session	64 (33.2)
Positive opinions about MI	
Sum score (range: 0-20; mean, SD)	15.5 (3.2)
Low (0-6), n (%)	1 (2.3)
Moderate (7-13), n (%)	6 (13.6)
High (14-20), n (%)	37 (84.1)
Missing, n (%)	9 (17.0)
Motivational nature of MI for behavior changes in substance use and exercise	
Sum score (range: 0-16; mean, SD)	12.3 (3.1)
Low (0-5), n (%)	1 (2.3)
Moderate (6-11), n (%)	9 (20.5)
High (12-16), n (%)	34 (77.3)
Missing, n (%)	9 (17.0)

[a]Multiple responses possible.

[b]Percentage of messages responded to per participant=(total number of messages responded to/total number of messages received that asked a question)×100. The percentages ranged from 0% to 100% and a mean and SD were calculated for the percentage of messages responded to.

[c]Both Teen MomConnect and MomConnect participants.

[d]MI: motivational interviewing.

Discussion

Principal Findings

This pilot RCT investigated the user acceptability and preliminary efficacy of Teen MomConnect. The number of antenatal appointments attended did not differ significantly between Teen MomConnect and MomConnect. While higher self-efficacy regarding attending appointments at posttest was observed in the MomConnect control group, this did not translate into higher rates of appointments attended. Higher mean increases in knowledge about the risks of not attending ANC and not practicing health behaviors during pregnancy were found among Teen MomConnect participants who interacted more frequently with the 2-way messages than the control group. In addition, the inclusion of MI in the intervention resulted in a 12% higher mean number of appointments attended (a difference of 0.57 visits) compared to the control group. This was not statistically significant, which was likely attributed to the small sample size for those who received MI. The majority of participants enrolled in the intervention regarded the messages highly in terms of content and their motivational nature for behavior change. Despite this, engagement with the 2-way messages was generally low, with participants replying to an average of 20% of the messages that asked them questions.

In comparison with our findings, a 3-arm trial among pregnant women living with HIV in Kenya also found that 2-way SMS text messages did not improve appointment attendance when compared to one-way messaging and standard care [35]. The findings of no preliminary efficacy on appointment attendance in this study can be contextualized with respect to the motivational nature, tailoring, and dose of the messages, and the fidelity of the intervention components, which relied on the frequency of engagement with the messages. First, behaviors with high resistance to change and that require larger motivation and competence to change usually need more intense or interpersonal interventions [36]. Moreover, the MomConnect fixed messages were previously tested and have already been shown to improve antenatal uptake when compared to standard care. Therefore, the Teen MomConnect intervention messages might not have adequately targeted the salient constructs affecting attendance behavior to result in higher antenatal attendance rates than MomConnect; they might not have been motivating enough, and their content may not have been sufficiently targeted to adolescent girls and young women. Second, the novelty of Teen MomConnect's two-way SMS text messages was not used as frequently as intended. This impacted intervention fidelity because the level of tailored message feedback is dependent on participants' responses to the two-way messages. The intervention could perhaps be more effective if it could identify and tailor appointment messages to the

XSL•FO
RenderX

adolescent girls and young women with high risk for low antenatal attendance, as determined by behavioral determinants. Finally, while the intervention contained more messages on appointment attendance, it also covered a range of other behavioral domains. It is unclear the dose and intensity of messages that would be required per domain to effect behavior changes. In this study, appointment reminders were based on asking participants when their next appointment was. Precise appointment date reminders that are linked to medical records have shown more promise in improving appointment attendance behaviors [35].

mHealth interventions generally experience low engagement rates [37,38], particularly among young people who become habituated to the mobile interventions and become distracted by other activities on their mobile phones [39]. It is easy to be passively involved in mHealth interventions, and there is no direct follow-up on why messages are not responded to. Intervention engagement can be defined as the perceptions and experiences with the intervention or as intervention use [40], and higher engagement is associated with intervention effectiveness [41]. In this study, participants reported high satisfaction with the message content, suggesting favorable perceptions and experiences with the intervention and high acceptability. In terms of message use, however, the number of messages responded to was very low. The small number of participants who replied frequently to the two-way messages impacted the power to detect a dose-response relationship for high versus low message engagement. Low responsiveness to bidirectional SMS text messages has also been observed in the WelTel text-based intervention for pregnant women in Kenya [42]. Younger women aged 18-24 years were twice as likely not to reply to messages than older women, but also had high satisfaction with the intervention content. Reasons for low engagement with the messages were forgetting to reply, loss of mobile phones, and message fatigue over the duration of the interventions, and when messages seemed repetitive [42]. The iterative nature of some Teen MomConnect messages could therefore have induced message fatigue and nonresponse. Regular access to a mobile phone is crucial for adherence to the intervention and engagement with the 2-way messages. A fifth of participants changed their mobile phone number during the study, which is common in South Africa, where the majority of phones are prepaid [15]. This would have affected the messages that they were exposed to and hence responded to. Furthermore, participants cited not always checking their phones and sharing phones as common reasons for not replying to messages timeously. In addition, the intervention's program logic allowed a few hours for participants to reply to some messages in cases where multiple messages were to be sent on the same day. This could have affected whether participants responded to the messages.

This study can be used to inform engagement rates in future text-based interventions for adolescent girls and young women. Features associated with increased adolescent engagement in mHealth studies, including apps, are increased tailoring, theory-informed messages, provision of self-monitoring and feedback, message reminders, game-based elements with rewards, and provision of peer or health care provider support [39]. The Teen MomConnect intervention included these elements to varying degrees, within the feature limitations of an SMS text messaging–based platform. Approaches that other adolescent text-based interventions used to increase engagement are quizzes or polls where the user responds by answering "myth" or "fact" and then receives the correct answer, and messages with pop culture references or slang phrases [43,44]. Teen MomConnect messages could have incorporated these features in order to increase engagement, while keeping in mind the restricted length of an SMS text message that balances both engaging and educational content, and the avoidance of sending too many messages to participants. We could have also called participants regularly to remind them to respond to messages, but this might have had limited success because previous attempts to contact them for scheduling MI sessions and posttest surveys presented difficulties with unanswered calls and phone calls directed to voicemail.

Participants who received MI reviewed it favorably. However, the feasibility of MI was low in terms of difficulties in scheduling MI sessions with participants due to unanswered calls, conflicting commitments, and participants not attending their sessions. This resulted in only one MI session being able to be delivered. Some mHealth studies confirm no significant difference in health behavior outcomes between text-based interventions and text supplemented with counseling or support groups [45,46]. The 12% increase in appointment attendance with MI found in this study was not statistically significant but warrants further investigation using larger samples. It is also possible that participants with fewer barriers to attending ANC visits were more likely to be willing to attend the MI session. Furthermore, the fidelity of the MI sessions was not assessed in this study. While the research assistants were trained and tested in MI skills, their proficiency in all aspects of MI while delivering their sessions was not measured. Suboptimal proficiency among counselors following MI training has been confirmed in other studies in South Africa [47,48]. More MI sessions may also be needed to elicit change, as effective MI interventions in pregnant women typically use 2-6 sessions [49].

Higher knowledge levels about the risks of not attending ANC and practicing healthy behaviors during pregnancy were found for participants who interacted more frequently with the Teen MomConnect intervention messages than those who did not and those in the control group. Other mHealth interventions with bidirectional messaging for adolescent girls and young women in low- and middle-income countries resulted in improvements in knowledge levels about sexual and reproductive health [50,51]. Mwenda et al [46] found that bidirectional text messages improved maternal knowledge about infant care during pregnancy. Rokicki and Fink [50] also found higher knowledge in girls who interacted more with the intervention content. Hence, the Teen MomConnect intervention messages provided participants with reproductive health information that is traditionally offered by antenatal services in a convenient, confidential, cost-effective, and adolescent-responsive way. This, in turn, enhanced their prenatal knowledge, which would equip them to make better health decisions during and after pregnancy.

XSL•FO
RenderX

A strength of the study is that the intervention development was informed by qualitative and quantitative research with pregnant adolescent girls and young women and by behavioral theories. The intervention applied tailored two-way messages and was designed to incorporate MI counseling. The intervention used SMS text messaging to facilitate comparison with the pre-existing MomConnect intervention, and SMS text messaging was the preferred method by the Department of Health at the time of the study. SMS text messages have the advantage of reaching all mobile phone users at no user cost when compared to mobile apps. Consequently, almost 60% did not prefer the messages to be delivered by an app, which can be attributed to high mobile data costs and problems with speed and coverage in South Africa [52]. In terms of limitations, first, there was low engagement with the messages, and despite participation incentives, high attrition at follow-up. Similarly, high attrition rates were found in other adolescent longitudinal studies [53,54] as well as in a study testing the MomConnect messages in pregnant women [14]. The low response rate in the follow-up survey could be because the in-person follow-up surveys were conducted postpartum at a time when the participants were preoccupied with looking after their newborns [54] and would therefore have little time and interest to be interviewed. Second, follow-up data from patient records were not collected for many participants due to difficulties with accessing patient records at their facilities. Maternal case records were not always found due to the filing systems at facilities, and some patient records were relocated to another facility or tertiary hospital where the participants gave birth. The low completion of the posttest survey and the numbers for whom appointment data were available limited the final samples used for analysis. Extensions of this work would benefit from improved strategies to improve participant engagement and interaction with the two-way messages, track and reregister participants who change mobile phone numbers, lessen attrition, and find ways in which MI can be implemented more effectively in this population.

SMS text message was relevant at the time of intervention development in 2016-2017, when smartphone access and use of the internet, social media, and apps were less prevalent in low socioeconomic groups in South Africa, among which adolescent pregnancy was highest. Smartphone access, frequent use of SMS text messaging and other apps, and internet use have now become almost ubiquitous in South Africa. Therefore, instead of text messages, content can be delivered in a wide variety of modes, such as social media sources using videos and images. The theory-based tailored two-way content used in Teen MomConnect lays a foundation for future mHealth interventions for pregnant adolescent girls and young women that use more advanced modes of technology, such as interventions that are app- or web-based or that are embedded within instant messaging services like WhatsApp. In recent years, MomConnect expanded to sending messages on WhatsApp to users who prefer this, with additional features like a chatbot answering frequently asked questions and directing questions to a helpdesk [55]. Opportunities for increased engagement and more in-depth tailoring to pregnant adolescent girls and young women, however, are provided in mobile apps. There are a few mHealth apps designed for young people in South Africa [56,57]. Apps allow for features such as gamification, greater customization and personalization, entertainment education, more efficient self-monitoring and feedback, peer forums, reminders, and notifications, which are all shown to be associated with improved appeal and relevance to youth in mHealth interventions [39].

Conclusions

The Teen MomConnect intervention was developed to send tailored 2-way text messages on antenatal attendance and pregnancy health behaviors to pregnant adolescent girls and young women. A specific intervention was developed for adolescent girls and young women because they are inherently different from older adult women in terms of their agency, decision-making, health determinants, and social and cognitive functioning. In this pilot RCT, the Teen MomConnect intervention did not have enhanced preliminary efficacy in improving antenatal attendance when compared to the national MomConnect messages, which could be attributed to low engagement with the two-way messages. However, the intervention messages had high acceptability. Participants with high engagement had higher knowledge about ANC and pregnancy risks. Future work can investigate strategies for improving engagement and retention in mHealth studies among pregnant adolescent girls and young women. Furthermore, a more nuanced approach is needed to identify and tailor messages to adolescent girls and young women with low motivation to attend antenatal appointments. This pilot study is a stepping stone for further work on sexual and reproductive mHealth interventions in adolescent girls and young women that build on the need for tailoring in young people and that also use modes of communication that go beyond SMS text messages.

Acknowledgments

The authors thank the participants and their parents for taking part in the study. They thank the team of research assistants for their dedicated work. The authors acknowledge the public health facilities that participated in this study for their support and assistance with recruiting participants and accessing patient records. This study was funded by a grant from the UK Medical Research Council (MR/N021355/1). The funding body played no role in the study design, data collection, analysis and interpretation of the data, or writing of the manuscript.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT 2010 checklist of information to include when reporting a pilot or feasibility trial.
[DOC File , 228 KB - mhealth_v13i1e59144_app1.doc]

Multimedia Appendix 2
Determinant Scales and reliability information.
[DOCX File , 18 KB - mhealth_v13i1e59144_app2.docx]

Multimedia Appendix 3
Characteristics of the analytic sample with data on appointment attendance by treatment condition.
[DOCX File , 18 KB - mhealth_v13i1e59144_app3.docx]

Multimedia Appendix 4
Sensitivity analyses.
[DOCX File , 21 KB - mhealth_v13i1e59144_app4.docx]

References

1. National Department of Health. South African medical research council, and ICF, South Africa demographic and health survey 2016. Statistics South Africa. 2019. URL: https://dhsprogram.com/pubs/pdf/FR337/FR337.pdf [accessed 2025-08-29]
2. Leftwich HK, Alves MVO. Adolescent pregnancy. Pediatr Clin North Am 2017;64(2):381-388. [doi: 10.1016/j.pcl.2016.11.007] [Medline: 28292453]
3. Adolescent pregnancy—unmet needs and undone deeds. In: Issues in Adolescent Health and Development. Geneva: World Health Organization; 2007.
4. Nove A, Matthews Z, Neal S, Camacho AV. Maternal mortality in adolescents compared with women of other ages: evidence from 144 countries. Lancet Global Health 2014;2(3):e155-e164 [FREE Full text] [doi: 10.1016/S2214-109X(13)70179-7] [Medline: 25102848]
5. WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience. Geneva: World Health Organization; 2017.
6. Gumede S, Black V, Naidoo N, Chersich MF. Attendance at antenatal clinics in inner-city Johannesburg, South Africa and its associations with birth outcomes: analysis of data from birth registers at three facilities. BMC Public Health 2017;17(Suppl 3):443 [FREE Full text] [doi: 10.1186/s12889-017-4347-z] [Medline: 28832284]
7. Ebonwu J, Mumbauer A, Uys M, Wainberg ML, Medina-Marino A. Determinants of late antenatal care presentation in rural and peri-urban communities in South Africa: a cross-sectional study. PLoS One 2018;13(3):e0191903 [FREE Full text] [doi: 10.1371/journal.pone.0191903] [Medline: 29518082]
8. Antenatal care coverage—at least four visits (%). World Health Organization. 2024. URL: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/antenatal-care-coverage-at-least-four-visits [accessed 2025-03-21]
9. Chan KL, Chen M. Effects of social media and mobile health apps on pregnancy care: meta-analysis. JMIR Mhealth Uhealth 2019;7(1):e11836 [FREE Full text] [doi: 10.2196/11836] [Medline: 30698533]
10. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
11. Wagnew F, Dessie G, Alebel A, Mulugeta H, Belay YA, Abajobir AA. Does short message service improve focused antenatal care visit and skilled birth attendance? A systematic review and meta-analysis of randomized clinical trials. Reprod Health 2018;15(1):191 [FREE Full text] [doi: 10.1186/s12978-018-0635-z] [Medline: 30466453]
12. Bateman C. Using basic technology—and corporate social responsibility—to save lives. S Afr Med J 2014;104(12):839-840. [doi: 10.7196/samj.8992] [Medline: 26042259]
13. Barron P, Peter J, LeFevre AE, Sebidi J, Bekker M, Allen R, et al. Mobile health messaging service and helpdesk for South African mothers (MomConnect): history, successes and challenges. BMJ Global Health 2018;3(Suppl 2):e000559 [FREE Full text] [doi: 10.1136/bmjgh-2017-000559] [Medline: 29713503]
14. Coleman J, Black V, Thorson AE, Eriksen J. Evaluating the effect of maternal mHealth text messages on uptake of maternal and child health care services in South Africa: a multicentre cohort intervention study. Reprod Health 2020;17(1):160 [FREE Full text] [doi: 10.1186/s12978-020-01017-3] [Medline: 33081795]
15. LeFevre AE, Dane P, Copley CJ, Pienaar C, Parsons AN, Engelhard M, et al. Unpacking the performance of a mobile health information messaging program for mothers (MomConnect) in South Africa:. BMJ Global Health 2018;3(Suppl 2):e000583 [FREE Full text] [doi: 10.1136/bmjgh-2017-000583] [Medline: 29713510]

XSL•FO
RenderX

16. Sewpaul R, Resnicow K, Crutzen R, Dukhi N, Ellahebokus A, Reddy P. A tailored mHealth intervention for improving antenatal care seeking and health behavioral determinants during pregnancy among adolescent girls and young women in South Africa: development and protocol for a pilot randomized controlled trial. JMIR Res Protoc 2023;12:e43654 [FREE Full text] [doi: 10.2196/43654] [Medline: 37703092]
17. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
18. General household survey 2021. Stats SA. 2022. URL: https://www.statssa.gov.za/publications/P0318/P03182021.pdf [accessed 2025-08-29]
19. Ryan RMH, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. Eur Psychol 2008;10:2 [FREE Full text]
20. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
21. Resnicow K, Davis RE, Zhang G, Konkel J, Strecher VJ, Shaikh AR, et al. Tailoring a fruit and vegetable intervention on novel motivational constructs: results of a randomized study. Ann Behav Med 2008;35(2):159-169. [doi: 10.1007/s12160-008-9028-9] [Medline: 18401673]
22. Robbins LB, Pfeiffer KA, Vermeesch A, Resnicow K, You Z, An L, et al. "Girls on the Move" intervention protocol for increasing physical activity among low-active underserved urban girls: a group randomized trial. BMC Public Health 2013;13:474 [FREE Full text] [doi: 10.1186/1471-2458-13-474] [Medline: 23672272]
23. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change. New York: Guilford press; 2013.
24. Resnicow K, Baskin M, Rahotep S, Periasamy S, Rollnick S. Handbook of Motivational Counseling: Motivating People for Change. Sussex, UK: Wiley and Sons; 2011:581-595.
25. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach. New York: Psychology Press; 2010.
26. de Vries H. An integrated approach for understanding health behavior; the I-change model as an example. PBSIJ 2017;2(2):55585. [doi: 10.19080/pbsij.2017.02.555585]
27. Sewpaul R, Crutzen R, Reddy P. Psychosocial determinants of the intention and self-efficacy to attend antenatal appointments among pregnant adolescents and young women in Cape Town, South Africa: a cross-sectional study. BMC Public Health 2022;22(1):1809 [FREE Full text] [doi: 10.1186/s12889-022-14138-0] [Medline: 36151528]
28. Crutzen R, Peters GY. Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. Health Psychol Rev 2017;11(3):242-247 [FREE Full text] [doi: 10.1080/17437199.2015.1124240] [Medline: 26602990]
29. Peters GJY. The alpha and the omega of scale reliability and validity: why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. Euro Health Psychol 2014;16(2):56-69. [doi: 10.31234/osf.io/h47fv]
30. Villar J, Ba'aqeel H, Piaggio G, Lumbiganon P, Belizán JM, Farnot U, et al. WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care. Lancet 2001;357(9268):1551-1564. [doi: 10.1016/s0140-6736(00)04722-x] [Medline: 11377642]
31. World Health Organization. WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience. Geneva: World Health Organization; 2016.
32. United Nations Children's Fund. The state of the world's children 2021. In: On My Mind: Promoting, Protecting and Caring for Children’s Mental Health. New York: UNICEF; 2021.
33. Khuzaiyah S, Mumin KHA, McKenna L, Hashim SH. Health-seeking behaviours of pregnant adolescents: a scoping review. Br J Midwifery 2023;31(7):408-417. [doi: 10.12968/bjom.2023.31.7.408]
34. Hackett K, Lenters L, Vandermorris A, LaFleur C, Newton S, Ndeki S, et al. How can engagement of adolescents in antenatal care be enhanced? Learning from the perspectives of young mothers in Ghana and Tanzania. BMC Pregnancy Childbirth 2019;19(1):184 [FREE Full text] [doi: 10.1186/s12884-019-2326-3] [Medline: 31122199]
35. Kinuthia J, Ronen K, Unger JA, Jiang W, Matemo D, Perrier T, et al. SMS messaging to improve retention and viral suppression in prevention of mother-to-child HIV transmission (PMTCT) programs in Kenya: a 3-arm randomized clinical trial. PLoS Med 2021;18(5):e1003650 [FREE Full text] [doi: 10.1371/journal.pmed.1003650] [Medline: 34029338]
36. Resnicow K, Teixeira PJ, Williams GC. Efficient allocation of public health and behavior change resources: the "Difficulty by Motivation" matrix. Am J Public Health 2017;107(1):55-57. [doi: 10.2105/AJPH.2016.303526] [Medline: 27925833]
37. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
38. Hesser H. Estimating causal effects of internet interventions in the context of nonadherence. Internet Interventions 2020;21:100346 [FREE Full text] [doi: 10.1016/j.invent.2020.100346] [Medline: 32983907]
39. Hightow-Weidman LB, Horvath KJ, Scott H, Hill-Rorie J, Bauermeister JA. Engaging youth in mHealth: what works and how can we be sure? Mhealth 2021;7:23 [FREE Full text] [doi: 10.21037/mhealth-20-48] [Medline: 33898592]

XSL•FO
RenderX

40. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
41. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
42. Nordberg B, Kaguiri E, de Angeles KJC, Gabriel EE, van der Kop ML, Mwangi W, et al. The use, adherence, and evaluation of interactive text-messaging among women admitted to prevention of mother-to-child transmission of HIV care in Kenya (WelTel PMTCT). BMC Pregnancy Childbirth 2024;24(1):25 [FREE Full text] [doi: 10.1186/s12884-023-06194-0] [Medline: 38172881]
43. Devine S, Leeds C, Shlay JC, Leytem A, Beum R, Bull S. Methods to assess youth engagement in a text messaging supplement to an effective teen pregnancy program. J Biomed Inform 2015;56:379-386 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.003] [Medline: 26173038]
44. Wrobel J, Silvasstar J, Peterson R, Sumbundu K, Kelley A, Stephens D, et al. Text messaging intervention for mental wellness in American Indian and alaska native teens and young adults (BRAVE Study): analysis of user engagement patterns. JMIR Form Res 2022;6(2):e32138 [FREE Full text] [doi: 10.2196/32138] [Medline: 35212633]
45. Head KJ, Noar SM, Iannarino NT, Harrington NG. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
46. Mwenda V, Makena I, Ogweno V, Obonyo J, Were V. The effectiveness of interactive text messaging and structured psychosocial support groups on developmental milestones of children from adolescent pregnancies in Kenya: quasi-experimental study. JMIR Pediatr Parent 2023;6:e37359 [FREE Full text] [doi: 10.2196/37359] [Medline: 37126373]
47. Mash R, Baldassini G, Mkhatshwa H, Sayeed I, Ndapeua S. Reflections on the training of counsellors in motivational interviewing for programmes for the prevention of mother to child transmission of HIV in sub-Saharan Africa. S Afr Fam Pract 2014;50(2):53-59. [doi: 10.1080/20786204.2008.10873697]
48. Dewing S, Mathews C, Cloete A, Schaay N, Shah M, Simbayi L, et al. From research to practice: lay adherence counsellors' fidelity to an evidence-based intervention for promoting adherence to antiretroviral treatment in the Western cape, South Africa. AIDS Behav 2013;17(9):2935-2345. [doi: 10.1007/s10461-013-0509-x] [Medline: 23666183]
49. van der Windt M, van Zundert S, Schoenmakers S, Jansen P, van Rossem L, Steegers-Theunissen R. Effective psychological therapies to improve lifestyle behaviors in (pre)pregnant women: a systematic review. Prev Med Rep 2021;24:101631 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101631] [Medline: 34976684]
50. Rokicki S, Fink G. Assessing the reach and effectiveness of mHealth: evidence from a reproductive health program for adolescent girls in Ghana. BMC Public Health 2017;17(1):969 [FREE Full text] [doi: 10.1186/s12889-017-4939-7] [Medline: 29262823]
51. Akande OW, Muzigaba M, Igumbor EU, Elimian K, Bolarinwa OA, Musa OI, et al. The effectiveness of an m-health intervention on the sexual and reproductive health of in-school adolescents: a cluster randomized controlled trial in Nigeria. Reprod Health 2024;21(1):6 [FREE Full text] [doi: 10.1186/s12978-023-01735-4] [Medline: 38218840]
52. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Global Health Action 2017;10(sup3):1327686 [FREE Full text] [doi: 10.1080/16549716.2017.1327686] [Medline: 28838302]
53. Laurenzi CA, du Toit S, Ameyan W, Melendez-Torres GJ, Kara T, Brand A, et al. Psychosocial interventions for improving engagement in care and health and behavioural outcomes for adolescents and young people living with HIV: a systematic review and meta-analysis. J Int AIDS Soc 2021;24(8):e25741 [FREE Full text] [doi: 10.1002/jia2.25741] [Medline: 34338417]
54. Pinto-Foltz MD, Logsdon MC, Derrick A. Engaging adolescent mothers in a longitudinal mental health intervention study: challenges and lessons learned. Issues Ment Health Nurs 2011;32(4):214-219 [FREE Full text] [doi: 10.3109/01612840.2010.544841] [Medline: 21355755]
55. South African programme MomConnect. Jembi Health Systems. 2023. URL: https://www.jembi.org/Project/momconnect [accessed 2025-08-29]
56. mHealth for adolescents and young people living with HIV: best practices, tools, and case studies. Paediatric-Adolescent Treatment Africa (PATA). 2021. URL: https://teampata.org/wp-content/uploads/2021/08/UB11-mHealth-Report_FINAL.pdf [accessed 2025-08-29]
57. Mulawa MI, Mtukushe B, Knippler ET, Matiwane M, Al-Mujtaba M, Muessig KE, et al. Supporting adolescents with HIV in South Africa through an adherence-supporting app: mixed methods beta-testing study. JMIR Form Res 2023;7:e47575 [FREE Full text] [doi: 10.2196/47575] [Medline: 37261883]

Abbreviations

ANC: antenatal care
CONSORT: Consolidated Standards of Reporting Trials

XSL•FO
RenderX

EDD: expected date of delivery
LMD: last menstrual date
mHealth: mobile health
MI: motivational interviewing
RCT: randomized controlled trial
WHO: World Health Organization

Edited by M Waring; submitted 03.04.24; peer-reviewed by B Nwankwo, DT Hendrixson; comments to author 24.07.24; revised version received 19.11.24; accepted 16.05.25; published 03.10.25.

Please cite as:
Sewpaul R, Resnicow K, Crutzen R, Dukhi N, Reddy P
A Tailored mHealth Intervention for Improving Antenatal Care Seeking and Its Determinants Among Pregnant Adolescent Girls and Young Women in South Africa: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e59144
URL: https://mhealth.jmir.org/2025/1/e59144
doi:10.2196/59144
PMID:

©Ronel Sewpaul, Ken Resnicow, Rik Crutzen, Natisha Dukhi, Priscilla Reddy. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 03.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Cerina—Cognitive Behavioral Therapy–Based Mobile App for Managing Generalized Anxiety Disorder Symptoms Among University Students: Results From a Pilot Feasibility Randomized Controlled Trial

Ozlem Eylem-van Bergeijk[1,2], PhD; Tony Robinson[3], MSc; Matthew Manktelow[3], PhD; Michail Olympios[4], MSc; Siobhan Poulter[3], PhD; Prasannajeet Mane[2], MSc; Maria Panagioti[1], PhD; Joan Condell[3], PhD; Gerard Leavey[5], PhD

[1]Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, United Kingdom

[2]Cerina Health, London, United Kingdom

[3]School of Computing, Engineering and Intelligent Systems Human Centred Computing, Intelligent Systems Research Centre, Ulster University, Londonderry, United Kingdom

[4]Faculty of Medicine, School of Life Sciences, Utrecht University, Utrecht, The Netherlands

[5]The Bamford Centre for Mental Health and Wellbeing, School of Psychology, Ulster University, Londonderry, United Kingdom

Corresponding Author:
Ozlem Eylem-van Bergeijk, PhD
Division of Population Health, Health Services Research and Primary Care
University of Manchester
Oxford Road
Manchester, UK M13 9PL
United Kingdom
Phone: 44 440161306
Email: ozlem.eylem-vanbergeijk@manchester.ac.uk

Abstract

Background: Generalized anxiety disorder (GAD) is common among university students due to academic pressure and financial uncertainty, among other challenges. Despite the need, the receipt of available psychological services is often low.

Objective: This study investigates the feasibility of a digital unguided cognitive behavioral therapy (CBT)–based mobile app, Cerina, and examines the likely effects of this intervention in reducing GAD symptoms compared to the waitlist control group.

Methods: Eligible students (n=158) with mild to moderate GAD symptoms were self-assessed through web-based questionnaires and were randomly allocated to the intervention group (n=79) or to the waitlist control group (n=79) following their informed consent. The intervention group had direct access to Cerina and followed CBT-based interactive sessions for 6 weeks. The waitlist control group participants had access to optional on-campus well-being services, and they were given access to Cerina 6 weeks after their randomization. Participants completed assessments on anxiety, depression, worry, and usability at three time points. Additionally, upon completing the intervention, they were invited to a web-based interview to understand the implementation of the intervention in more depth.

Results: On average, 13% (10/79) intervention group participants dropped out, 61% (36/69) completed the core clinical content (2 sessions), and 12% (7/69) completed the desired number of sessions (6 or 7 sessions). Analyses of the completers (2 or more sessions) revealed significant group differences in GAD (mean 8.4, SD 3.7; t_{42}=–2.25; P=.03; d=–0.7) and worry symptoms (mean 42.3, SD 10.8; t_{42}=–2.50; P=.02; d=–0.8), as well as functional impairment (mean 16.7, SD 2.44; t_{42}=–2.12; P=.04; d=–0.6) in favor of the intervention group at posttest with medium to large effect sizes. The intention-to-treat analyses confirmed significant group differences in GAD (mean 8.47, SD 2.7; t_{156}=–2.23; P=.03; d=–0.4), and there were marginally nonsignificant group differences in worry symptoms (mean 41.5, SD 8.40; t_{156}=–1.94; P=.05; d=–0.3) in favor of the intervention group at posttest with medium effect sizes. These results suggest that the intervention had a meaningful impact on reducing GAD symptoms and a modest impact on reducing worry symptoms among participants.

Conclusions: The Cerina app showed promising results in reducing GAD symptoms among students. This result supports findings from other randomized controlled trials showing that digital CBT-based interventions are effective and feasible for a

XSL•FO
RenderX

wide range of age groups and populations experiencing GAD symptoms. The low number of participants completing the recommended number of sessions suggests a usability issue. To address this, the intervention could be refined through an iterative design process informed by user feedback, and the long-term impact of specific engagement features in improving usability and retention could be assessed through extended evaluations.

Trial Registration: ClinicalTrials.gov NCT06146530; https://clinicaltrials.gov/study/NCT06146530

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2023-083554

(JMIR Mhealth Uhealth 2025;13:e70691) doi:10.2196/70691

KEYWORDS

generalized anxiety disorder; mHealth; mobile health; CBT; students; feasibility RCT; randomized controlled trial; cognitive behavioral therapy

Introduction

Background

Generalized anxiety disorder (GAD) is a common mental disorder (CMD) and is the most common anxiety disorder, affecting almost 30%-40% of the global population [1]. It is estimated that the prevalence of GAD varies widely across countries, with lifetime prevalence highest in high-income countries (5% [0.1%]), lower in middle-income countries (2.8% [0.1%]), and lowest in low-income countries (1.6% [0.1%]) [1,2]. Lifetime comorbidity is also high (81.9% [0.7%]), particularly with mood (63% [0.9%]) and other anxiety (51.7% [0.9%]) disorders [1-3]. Treatment is sought by approximately half of the affected individuals (49.2% [1.2%]) in high-income countries [1-3]. This suggests that GAD often goes unrecognized and untreated, even in high-income countries where access to mental health services is more readily available [4].

University students represent a vulnerable group at increased risk for anxiety disorders, particularly GAD. [5-12]. Data from the World Health Organization (WHO) World Mental Health International College Student project, conducted across 19 universities in eight countries (Australia, Belgium, Germany, Mexico, Northern Ireland, South Africa, Spain, and the United States), indicate that GAD was the most prevalent anxiety disorder among university students (n=14,371), with a lifetime prevalence of 18.6% and a 12-month prevalence of 16.7% [12]. Although longitudinal studies indicate mixed results about the increased risk for anxiety disorders among university students in the long run [13-16], there is growing evidence pointing to the subgroups of students who might be more at risk for anxiety disorders such as those who report financial uncertainty, those with a history of other CMDs [13-15], and students from ethnic minorities and lesbian, gay, bisexual, transgender, and queer groups [17].

According to the diathesis-stress model, it could be argued that common stressors faced by university students, such as academic pressure and financial uncertainty, may interact with pre-existing vulnerabilities, including a history of CMDs, developmental factors such as identity exploration and engagement in risky behaviors, and stressors specific to the university environment. Together, these interactions may heighten susceptibility to GAD in this population [16]. Additionally, Minority Stress Theory suggests that students from lesbian, gay, bisexual, transgender, and queer groups and ethnic minority backgrounds may face additional stressors, including discrimination and reduced access to social capital, which may further elevate their risk of developing GAD [18].

Psychotherapies are recommended for treating GAD [19,20]. Recent network meta-analysis examining the short- and long-term effects of different psychotherapies in reducing GAD symptoms suggested that cognitive behavioral therapy (CBT) remained significantly effective in reducing GAD symptoms at 3 to 12 months after completion of the intervention (standardized mean difference –0.60, 95% CI –0.99 to –0.21) [19]. However, relaxation therapy and third-wave CBTs were associated with short-term effectiveness [19]. Although pharmacotherapies are also recommended, CBT has larger overall treatment effects and fewer side effects than pharmacotherapy in the long run (6-12 mo follow-up; g=0.34) [21]. Although CBT represents the first-line therapy for GAD [20], social (eg, perceived stigma), organizational (eg, waiting times and service availability), and financial barriers (eg, in health care contexts where one must pay for their treatments) can restrict its access [22,23].

In recent years, digital CBT, including CBT-based smartphone apps and web-based interventions, has been proposed as an adjunct to or alternative to face-to-face CBT to overcome the above-mentioned barriers [24]. Despite the assumed benefits of digital CBT, such as reduced costs, anonymity, and accessibility [25], the existing randomized controlled trials (RCTs) point to mixed results about the effectiveness of these interventions in reducing GAD symptoms among student populations [26-31]. These mixed results may be attributed to the heterogeneity of CBT techniques used across treatment protocols, as well as methodological limitations such as small sample sizes, high dropout rates, and a lack of long-term follow-up [26]. Furthermore, high dropout rates suggest that the engagement with digital CBT might still be a challenge, especially among student populations [24-31]. For instance, an open feasibility trial [32] indicated that only a relatively small portion of a sample of university students, 26.5% (31/117), downloaded one or more of the CBT-based apps tested in the study, and approximately 24% (28/117) participants used these apps only once [32]. The literature indicates several person- and intervention-related reasons for dropout from digital CBT-based interventions. The most common being lengthy sessions conflicting with students' busy academic schedules, needs for different kinds of help, and absence of personal contact [33,34]. There is, therefore, a need to improve adherence to

XSL•FO
RenderX

digital CBT-based interventions and test new approaches to treat GAD symptoms among students [33].

One way of mitigating low engagement is through encouragement strategies such as automatic notifications, follow-up via phone or email, or automated feedback on homework exercises [35]. However, these methods are not prescriptive, and sustained engagement could be achieved if the intended population coleads the development and implementation of the digital CBT-based interventions [36,37]. More recently, technology-driven engagement features such as customized notifications, reminders, avatars (ie, web-based therapists), and chatbots have shown promise in changing users' attitudes and behaviors [38-40]. These features differ from standard app functionalities by leveraging artificial intelligence to initiate interactions with users about specific content or exercises. Given the heterogeneity in mental health and help-seeking profiles among student populations [41-43], such technology-driven engagement strategies might mitigate the issue of low engagement [43,44].

Aims

Overall, GAD is prevalent among university students, yet access to effective, scalable interventions remains limited due to barriers such as stigma, service availability, and time constraints. To help address this unmet need, Cerina Health, a UK-based mental health start-up, has developed Cerina, a mobile app grounded in CBT principles for the treatment of GAD symptoms. The app incorporates technology-driven engagement features within its user interface (UI), including customized notifications, reminders, and a chatbot that initiates conversations with users around specific therapeutic content or exercises, following co-design principles [36,37]. This study, therefore, has two primary objectives: (1) to investigate the feasibility of Cerina among a university student population, focusing on adherence, usability, and its potential for scaling to a full trial; and (2) to test the preliminary effects of Cerina in reducing GAD symptoms compared to a waitlist control group in university students presenting with self-reported GAD symptoms.

Methods

Study Design and Settings

The study is a single-blind pilot feasibility RCT with two conditions: an intervention group and a waitlist control group. This design addresses the primary research question: Is the Cerina app feasible to implement in a university setting in terms of usability, adherence, and potential for future scale-up? It also allows for the examination of the preliminary effects of Cerina compared to a waitlist control condition. A waitlist control group was selected to ensure that participants, many of whom may not otherwise have access to psychological support, would still have the opportunity to complete the intervention following the study [36]. Participants in the intervention group received immediate access to the Cerina app for 6 weeks, while those in the waitlist control group had access to optional on-campus well-being services during that period. After completing the posttest assessments at 6 weeks, the waitlist group was then offered access to Cerina.

All participants were asked to fill in web-based self-report measures of GAD, depression, worry, and work and social adjustment at baseline (t0), at 3 weeks (t1), and at 6 weeks (t2). Additionally, intervention group participants completed the system usability questionnaire at t2, while the waitlist group participants completed it after they accessed Cerina at t2. The 3-week (t1) assessment was included to capture early cognitive and behavioral changes, which are often expected to emerge within the first few sessions of CBT. The 6-week (t2) point aligns with the full delivery of core CBT components, and was used to assess short-term outcomes and initial consolidation of change. Semistructured interviews were conducted with all participants to explore which elements of the intervention facilitated or hindered participants' progress during the study and how the intervention can be optimized to increase its acceptability, relevance, and user-friendliness in real-life settings. Intervention group participants were invited to the semistructured interviews right after they completed their follow-up assessments at t2. Waitlist control group participants were invited to the interviews after they accessed Cerina and completed their 6-week follow-up assessments.

The study was conducted in partnership between Cerina Health and Ulster University as part of the North West Europe INTERREG IT4Anxiety project. The primary objective of the IT4Anxiety project is to support the development and implementation of innovative, start-up-led solutions aimed at reducing anxiety. This collaboration supported Cerina Health's research capacity to design and implement the study in a university setting. To avoid potential bias or conflicts of interest, the Chief Executive Officer of Cerina Health (PM) was involved only during the conceptualization phase of the study. The study was managed by OEB in collaboration with a research team at Ulster University. Study participant recruitment took place at Ulster University campuses from April 2023 until April 2024. Ulster University has approximately 30,000 students enrolled in undergraduate, postgraduate, and e-learning programs across its four main campuses: Belfast, Coleraine, Londonderry Derry, and Jordanstown in Northern Ireland.

Ethical Considerations

Ethical approval for the study has been granted by the Ulster University Research Ethics Committee (FCPSY-22-084). The study adhered to the rights of the participants. All participants provided written informed consent on enrollment, and each participant had the right to make an informed, voluntary decision to participate in this study. They had the right to withdraw from the trial at any stage or access their data without any consequences. Participants also had the right to access available services offered by the university Student Well-being team, their general practitioner (GP), and additional resources (eg, Inspire Wellbeing, Samaritans, Lifeline). All data were stored according to the Ulster University General Data Protection Regulation. To adhere to privacy and confidentiality, each participant was given a participation number, and their data were stored in a deidentified format in the Ulster University infrastructure by the research team. User name and password were used, and the password was encrypted. The participant's name and other identifiers were replaced by a unique code. To reduce the risk of disclosure, personal identifiers were stored

XSL•FO
RenderX

separately from the research data at the Bamford Centre, Ulster University storage by the Principal Investigator and were only accessible to the research team. There was a key document that linked their participation number to their real identity. This was kept at the Bamford Centre, Ulster University storage, and only the research team and the Student Well-being team were able to access this key document as part of the safety protocol [36]. In addition, monetary compensation was not offered to the participants.

Participants

Participants were adult students attending Ulster University who had reported anxiety symptoms based on the Generalized Anxiety Disorder Scale-7 (GAD-7) questionnaire. Eligible participants were (1) 18 years of age and older, (2) reported mild to severe self-reported GAD symptoms (ie, those who scored between 5 and 19 on the GAD-7 questionnaire were accepted as eligible), (3) enrolled as a student at Ulster University (ie, having a valid student ID number), (4) fluent in English, (5) had smartphone (ie, Android device or iPhone) and (6) internet connection, and (7) they gave informed consent. The exclusion criteria were (1) reporting minimal anxiety symptoms (ie, defined by a score of 5 and below on the GAD-7), (2) scoring 19 and above on the GAD-7, (3) reporting suicidal thoughts (ie, based on their scores on the Patient Health Questionnaire-9 [PHQ-9]), (4) having recently started taking psychotropic medication (ie, within the last 6 wk), and (5) not giving informed consent. Receiving psychological treatment was not part of the exclusion criteria.

Recruitment and Randomization

Information about the study was available on Cerina Health's website and Ulster University Student Well-being team's website, where potential participants could register by filling out a web-based registration form. Those who registered received an email containing an information sheet, consent form, and a link to the baseline questionnaire. Participants who were eligible (based on the baseline questionnaire) and provided consent received a follow-up email informing them of their group allocation. Participants in the intervention group received further instructions on how to download the Cerina app, and technical assistance was provided if required. Those allocated to the waitlist control group received an email confirming their group allocation and were informed about the optional support services available through the Student Well-being team while waiting to access Cerina. All the participants, regardless of group allocation, received weekly email reminders to follow their sessions and complete their assessments.

Several channels were used to recruit participants: (1) invitation emails were circulated within the student email lists (global emails authorized by the university's Pro Vice Chancellor for Academic Quality and Student Experience office), and the study-related communications (ie, banners, social media posts, and posters) were posted on Cerina Health's LinkedIn and Instagram accounts, the Ulster University Student Well-being Facebook page, and the Ulster University Instagram account; (2) on-campus drop-ins (ie, brief face-to-face interactions between researchers and students where leaflets or posters are handed out to students and students asked questions about the study); and (3) Ulster University Students Union were approached in person and through emails to circulate information about the study within their networks of students.

The randomization scheme was derived using a random allocation software by an independent person who was not involved in the study. Randomization took place in blocks of 6 and in a 1:1 ratio. Allocation was concealed from the researchers involved in the study. It was not possible to mask participants due to the waitlist control group and the nature of the intervention. All participants were informed of their assigned condition. The study protocol is described in more detail elsewhere [36]. The ClinicalTrials.gov ID assigned to this study was NCT06146530.

Intervention and the Waitlist Control Condition

Cerina is an unguided CBT-based mobile app for GAD symptoms. It does not provide any professional support related to the therapeutic content of the intervention [36,45]. If participants experience any difficulty with a particular therapeutic technique on the app or a technical problem, they are advised to contact the Cerina team at support@cerina.co. If the users experience a distressing situation unrelated to Cerina, they are advised to contact their GP or a health care professional. In addition, there is a self-care page available on the app, including links to external resources for further help, such as the National Health Service helpline, mental health charities, support groups, and Samaritans [36].

Following National Institute for Clinical Excellence guidelines for treating GAD and Panic Disorder [20], Cerina is positioned as a STEP 2 intervention: this includes low-intensity psychological interventions, individual guided self-help, and psychoeducation. The sessions are based on the treatment protocol of the Roth and Pilling UCL Competency Framework [46]. Clinical features include evaluation of worry (using the Penn State Worry Questionnaire [PSWQ]), worry awareness training, understanding and coping with intolerance of uncertainty, re-evaluating beliefs about the usefulness of worrying, worry management techniques, problem-solving strategies, cognitive exposure, and relapse prevention [20,47]. Cerina consists of 7 sessions of CBT for the treatment of GAD. Each session includes information, tasks, and exercises to help the participants understand GAD, the CBT approach, and how it applies to them. The participants complete the sessions in a progressive direction, and they are advised to complete the first six sessions on a weekly basis, that is, one per week. However, they can repeat a session before going on to the next session. The last session is designed as a review of progress session rather than a full therapy session and does not include additional exercises or homework assignments. These seven therapy sessions are (1) learning about GAD, (2) dealing with worry, (3) managing worry and anxiety, (4) beliefs about worry, (5) managing uncertainty, (6) testing beliefs, and (7) review and therapy blueprint [36].

Several design features were identified during the co-design process [36] and were incorporated into the UI to promote user engagement. These include (1) vibrant and simple illustrations to present the case studies in each session; (2) customized notification settings; (3) a separate self-care page including

XSL•FO
RenderX

podcasts, anxiety management techniques, and available resources for further help if needed; (4) regular check-ins to rate mood; (5) a safety mechanism embedded within the participants' conversations with the chatbot, for instance, if a participant talks about death and suicidal thoughts, the chatbot provides contact details of the available crisis lines and services for further help; and (6) flexibility regarding the navigation, for example, participants can resume where they are left off upon logging into the app, and they can revisit previous sessions or review their therapy journey at any time [36].

The waitlist control condition includes optional on-campus well-being services offered by the Student Well-being team. The Student Well-being team on the Belfast, Coleraine, and Derry Londonderry campuses provides free and confidential support and guidance to students with a broad range of issues, concerns, and challenges, helping them to successfully engage in their studies. Student Well-being assistants provide an initial assessment to determine a student's primary need, and then well-being advisers, student mental health advisers, and accessibility advisers are available to provide a variety of solution-focused interventions based on that need. Each campus team is led by a Student Well-being manager to support the advisers in the management of risk and response to clinical incidents. Additionally, therapeutic counseling support is also available free to students through the external counseling provider via a dedicated 24/7 counseling helpline.

Study Outcomes

Primary Outcome

Feasibility was defined in terms of adherence to the intervention, its usability, and potential to deliver a full RCT in the future [47].

Adherence was defined based on Staudt adherence model [37,48], and it includes behavioral (ie, performance on tasks that are necessary to implement the intervention in daily life, such as homework completion) and attitudinal (ie, emotional commitment to the intervention) adherence [48]. In this study, behavioral adherence was assessed based on the number of (%) login attempts and the number of (%) participants who followed the sessions [30]. Attitudinal adherence was measured based on the number of participants (%) who completed the core therapeutic content (ie, entries in their worry diary in session 1 and reflections in session 2), some therapeutic content (ie, psychoeducational part), or no therapeutic content at all (ie, few taps and invalid entries) [36]. Additionally, we looked at the desired number of sessions followed (n=6) by the intervention and waitlist group participants.

Usability was measured with the System Usability Scale (SUS) [49]. The SUS consists of 10 items rated on a 5-point Likert scale, with total scores ranging from 0 to 100. It has demonstrated good reliability (Cronbach α=0.91) [49]. The scale also showed good internal consistency in this study (α=0.81). Scores of 70 and above are considered indicative of high usability, suggesting that the intervention is easy to use and navigate, and that users feel confident implementing it in real-life contexts [49]. Scores between 50 and 70 indicate acceptable usability, though they may point to minor issues or concerns in everyday use. Scores below 50 raise major concerns about usability and warrant further investigation [49]. SUS results should be interpreted alongside data from semistructured interviews to inform potential refinements to the intervention [49].

The potential to deliver a full-scale RCT in the future was assessed based on the response rates to the self-report questionnaires, the recruitment procedures, and participants' willingness to be randomized to a waitlist control group [36].

Secondary Outcomes

GAD-7, PSWQ-PW, PHQ-9, and the Work and Social Adjustment Scale (WSAS) were selected secondary outcome measures as these instruments are routinely used to assess clinical symptoms across diverse settings [50-54].

GAD symptoms were assessed through the GAD-7 [50], a 7-item self-report scale designed to identify the severity of GAD. Scores range from 0 to 21, with a cutoff score of 5 used to distinguish between clinical and nonclinical populations. The scale has good psychometric properties [50]. In this study, the scale showed acceptable internal consistency (α=0.7).

Worry symptoms were assessed using the PSWQ-PW [50], a 15-item inventory assessing both the weekly status of pathological worry and treatment-related changes of worry during the treatment [50]. Each item is scored on a 7-point rating scale, ranging from 0=never to 6=almost always. The total score ranges from 0 to 90, with a high score indicating higher levels of worry. PSWQ-PW shows good reliability and convergent validity [50]. In this study, the scale showed good internal consistency (α=0.8).

Depression symptoms were assessed using the PHQ-9 instrument [51]. Scores range from 0 to 27, with a score of 10 and above considered to be a clinically significant level of depression. The PHQ-9 has good reliability and validity [51]. In this study, the scale showed good internal consistency (α=0.8).

Functional impairment was assessed using WSAS [52]. Scores range from 0 to 40. The scale assesses the impact on work, home, social and private activities, and personal and family relationships. A score of 20 and above is considered to indicate severe functional impairment, scores between 10 and 20 suggest severe but functional impairment, and scores of 10 and less are considered subclinical. The scale has good reliability and validity [52]. In this study, the scale showed good internal consistency (α=0.8).

Statistical Analysis

The process evaluation examined usability (ie, pathways and barriers to using Cerina in daily life), the clinical utility of the app, the context in which the implementation of the clinical content takes place, and the processes involved in delivering the intervention [36]. These aspects were investigated through semistructured web-based interviews or web-based questionnaires with participants from both the intervention and waitlist control groups who completed the intervention. The results of the process evaluation are reported in detail elsewhere [55].

XSL•FO
RenderX

The RCT was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials; Multimedia Appendix 1) and CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) [56] checklists (Multimedia Appendix 2). First, 2-tailed *t* tests and chi-square tests were performed using SPSS (IBM Corp) to compare the baseline characteristics between participants allocated to the intervention group and those allocated to the waitlist control group. Second, independent 2-tailed *t* tests were used to compare changes in secondary outcomes across time points between the two groups. These analyses were conducted in RStudio (version 3.6.1; Posit Software, PBC) using the completers dataset [57]. A 2-tailed *t* test was chosen instead of a linear mixed model with random intercepts, as it is equivalent in terms of yielding the same test statistic for evaluating the treatment effect [58]. Third, after reviewing the data collected, 41.83% of the dataset was missing; this missing data was deemed to be completely at random [59]. To address the missing data, a single imputation-linear regression model [60] was trained, using scikit-learn in Python (version 3.11.0; Python Software Foundation) to estimate and balance the missing data [61]. The model was trained per item, based on the pre-existing data of weeks t0, t1, and t2, with a "bias gradient" represented by the items' values at t0. Therefore, a feature prediction was made for each (row) subject considering their t0 outcomes, the pre-existing t1 and t2 results, and the "bias gradient" created. The rationale for this imputation technique was three-fold: (1) it factors a multivariate approach, considering the relationship between the items [62,63], and the relationship and underlying patterns between the same item in a longitudinal nature [60]; (2) it illustrates a linearity between the three time points, showcasing a decrease of symptomatology based on the existed, preimputation data; and (3) it creates an acceptable normal distribution [64] in comparison with the pre-existed data with a "trade-off" of shifting the residuals, extreme values toward the mean (decreasing variance of residuals) [65]. Following this data analysis and transformation, the intention-to-treat (ITT) analyses were performed to test the likely effects of the intervention compared to the waitlist control group, as described for the completers dataset. Cohen *d* for the effect of the intervention was estimated by calculating the difference between estimated means (corrected for baseline) divided by the raw pooled SD. Effect sizes of 0.8 were accepted as large, effect sizes of 0.5 were moderate, and effect sizes of 0.2 were small [66]. A two-sided $P<.05$ indicated statistical significance.

Results

Participants

This subsection provides information on the recruitment channels, participant flow throughout the study, and baseline characteristics and their distribution across the intervention and the waitlist control group.

Overall, a total of 514 students registered for the study through the registration link provided in the global emails (n=488, 95%), social media channels including Instagram (n=3, 0.6%), Facebook (n=4, 0.7%), LinkedIn (n=1, 0.2%) and on the leaflets and posters (n=4, 0.7%) shared during the on-campus drop-ins. Additionally, a few participants registered for the study through word of mouth from the Student Well-being team (n=3, 0.6%), friend (n=2, 0.4%), and tutor (n=2, 0.4%).

Figure 1 shows the participant flow through the trial. Out of 514 registered people, 65% (n=356) were excluded mainly due to not completing the baseline questionnaire (n=240, 47%), not giving consent (n=63, 12%), presenting suicidal thoughts (n=26, 5%), presenting severe GAD symptoms (n=10, 2%), and presenting minimal GAD symptoms (n=17, 3%).

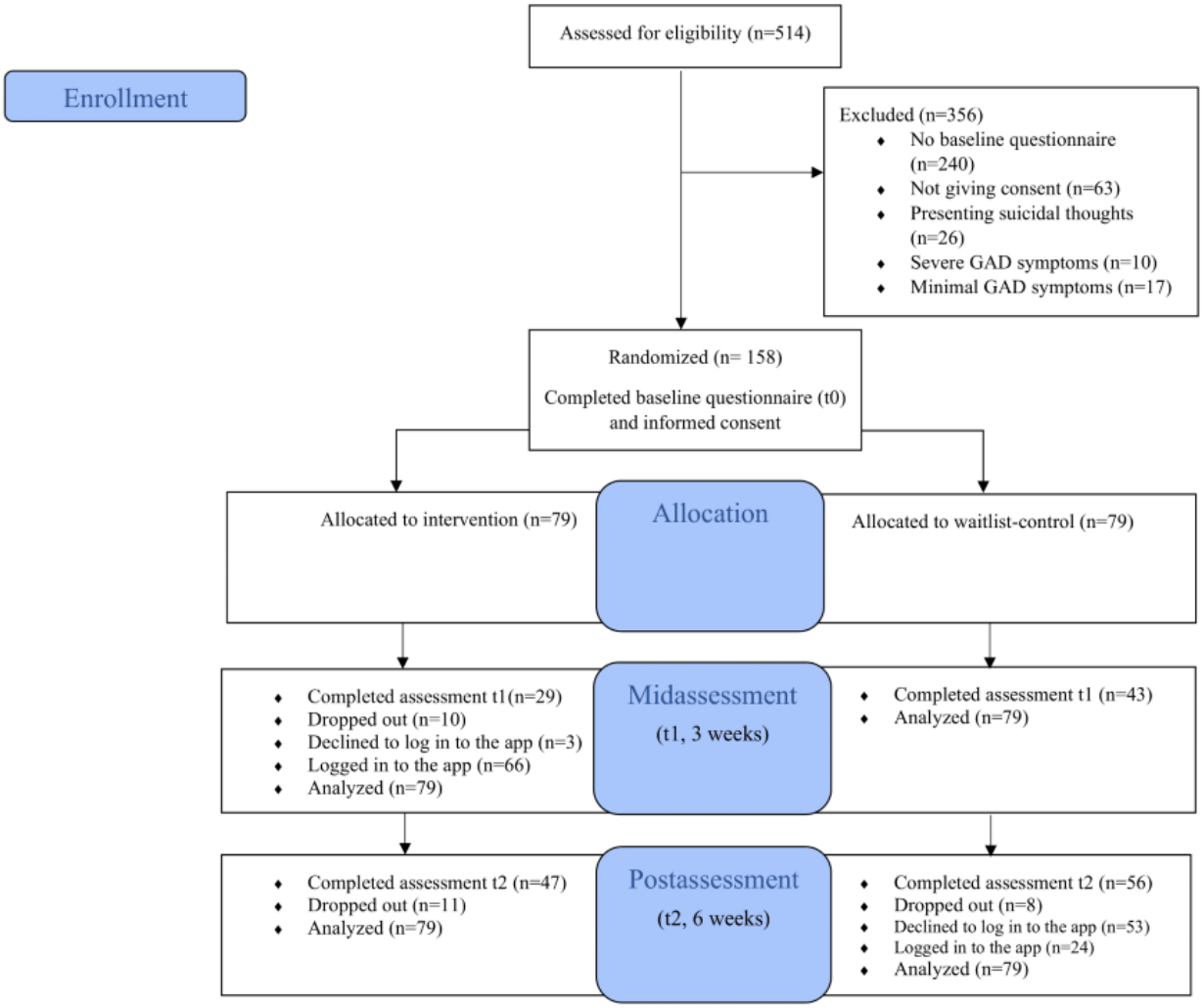
The remaining 158 people were eligible, gave informed consent, and were randomized (n=79 intervention; n=79 waitlist control group). In the intervention group, a total of 37% (n=29) of participants completed their 3-week assessments, and a total of 59% (n=47) of participants completed their 6-week postassessments. In the waitlist control group, a total of 54% (n=43) of participants completed their 3-week assessments, and a total of 71% (n=56) of participants completed their 6-week assessments. Additionally, all participants who used Cerina and completed their 6-week assessments (n=56, 54%) were invited to the web-based postassessment feedback interviews. Of those who agreed to give postassessment feedback (n=20, 36%), the majority of them opted out of interviews and completed web-based questionnaires (n=16, 80%), while only 20% (n=4) agreed to be interviewed.

A significantly higher number of participants in the intervention group (n=41, 71%) than in the waitlist control group (n=17, 29%) did not complete their 3- and 6-week assessments ($P<.05$). There were no significant differences in baseline characteristics between participants who completed the assessments and those who did not.

Multimedia Appendix 3 displays the baseline characteristics for all the participants included in the study. Overall, there were no imbalances between the intervention and the waitlist group participants. The majority were female (n=129, 81.6%), undergraduate students (n=78, 49.4%), and aged between 18 and 30 years (n=104, 65.8%). While a considerable number of participants chose friends and family as their main source of informal support (n=42, 26.5%), the majority of them were not receiving professional support at the time of their baseline assessment (n=83, 52.5%). Of those who were receiving professional support, 13.3% (n=21) chose GP as the main source of professional support. More than half of the participants (n=105, 66.5%) considered professional support as neither helpful nor unhelpful. While the majority of participants were not using psychotropic medication for a mental health problem (n=105, 66.5%), a considerable number of them (n=53, 34%) were using psychotropic medication at the time of the baseline assessment. Of these, 24.7% (n=39) of participants had been on a stable dose for more than 6 months, 6.3% (n=10) for the past 2 months, and 2.5% (n=4) had recently started medication.

XSL•FO
RenderX

Figure 1. Participant flowchart. GAD: generalized anxiety disorder.



On average, participants reported moderate levels of generalized anxiety symptoms (mean 11.64, SD 3.84), clinically significant levels of depression (mean 12.71, SD 5.16), a substantial level of worry (mean 51.08, SD 8.75), and severe but functional levels of impairment (mean 19.41, SD 8.53). Finally, there were no statistically significant differences between the intervention and waitlist control groups on any baseline characteristics.

Safety

This subsection summarizes information on the implementation of the safety protocol throughout the study.

As part of the safety protocol, students (n=26) who reported suicidal thoughts at the baseline assessment (measured by PHQ-9) were signposted to the Student Well-being team and excluded from the study. These student participants received a phone call from the Student Well-being team to schedule a follow-up phone call with an adviser for further assessment. Out of those who were signposted, 27% (n=7) did not respond, 38% (n=10) did not require support, 27% (n=7) started engaging with the Student Well-being team, and the remaining were receiving other external support (n=2, 8%). In addition, all participants were closely monitored throughout the study via their responses to regular in-app check-in questions and their 3- and 6-week assessments. None of the participants' responses indicated worsening of symptoms, and none of the participants contacted the research team to be signposted to the Student Well-being team for further help.

Feasibility

This subsection presents findings related to participants' adherence to the intervention and the usability of the Cerina app.

Behavioral Adherence

Of the 79 participants allocated to the intervention group, 13% (n=10) dropped out, and 87% (n=69) were given access to Cerina. Among those, 97% (n=66) logged in to Cerina, and more than three-quarters (n=59, 89.3%) followed the sessions. In the waitlist control group, less than half (n=26, 33%) were given access to Cerina, while the remaining participants (n=53, 67%) did not follow the sessions. Of the 26 waitlist group participants who were given access to Cerina, the majority (n=24, 92.3%) logged in to the app, and 96% (n=23) followed the sessions.

Attitudinal Adherence

On average, 61% (n=36) of intervention group participants engaged with the core therapeutic content (ie, have entries in their worry diary in session 1 and reflect on their worry diary

XSL•FO
RenderX

in session 2), while 25% (n=15) engaged with the psychoeducational content in session 1, and 14% (n=8) did not engage with any content (ie, few taps and invalid entries). Additionally, only 12% (n=7) of participants completed 6 or 7 sessions in the intervention group. Similarly, of all the waitlist group participants who followed the sessions, 56% (n=13) engaged with the core therapeutic content, 35% (n=8) engaged with psychoeducational content in session 1, and 9% (n=2) did not engage with any content. Only 4% (n=1) of participants completed 6 or 7 sessions in the waitlist control group.

Usability

Participants in the intervention group reported an average score of 71.08 (SD 12.04) at 6-week assessments (t2) on the SUS scale. This score indicates highly acceptable usability [49] and suggests that overall, the intervention was considered easy to use, various UI functions were well-integrated, and users felt confident while using it.

Intervention Effects

This subsection presents the results of the completers and the ITT analyses.

Completers Analyses

Overall, the completers analyses indicated changes in secondary outcomes: GAD-7, PHQ-9, PSWQ-PW, and WSAS scores from baseline (t0) to 6-week (t2) assessments.

The results of the independent 2-tailed *t* test indicated that there was significantly greater reduction in GAD-7 scores among participants in the intervention group (mean 8.4, SD 3.7; 95% CI 6.66-9.86) compared to the waitlist control group (mean 11, SD 3.9; 95% CI 8.39-12.52; t_{42}=–2.25; *P*=.03) at t2 with moderate to large effect size (*d*=–0.7), suggesting that the intervention had a meaningful impact on reducing GAD-7 scores among completers. The point bilateral correlation (*r*=–0.328) supports this finding (Multimedia Appendix 4).

There was significantly greater reduction in PSWQ-PW scores among participants in the intervention group (mean 42.3, SD 10.8; 95% CI 36.3-46.31) compared to the waitlist control group (mean 49.8, SD 9.06; 95% CI 43.86-52.81; t_{42}=–2.50; *P*=.02) at t2 with large effect sizes (*d*=–0.8), suggesting that the intervention had a strong impact on reducing PSWQ-PW scores among completers. The point-biserial correlation (*r*=–0.360) indicates that the intervention may be highly effective in alleviating worry symptoms among completers (Multimedia Appendix 5).

Similarly, the independent 2-tailed *t* test results comparting the intervention (mean 16.7, SD 2.44; 95% CI 11.99-21.66) and the waitlist control group (mean 23.2, SD 8.35; 95% CI 16.38-25.12) on WSAS scores revealed significant group differences (t_{42}=–2.12; *P*=.04) with moderate effect size (*d*=–0.6) at t2 in favor of the intervention, suggesting that the intervention had a meaningful impact on reducing WSAS scores among completers. The point bilateral correlation (*r*=–0.311) supports the notable impact of the intervention on improving completers' functional impairment (Multimedia Appendix 6).

Finally, the independent 2-tailed *t* test results comparing the intervention (mean 10.1, SD 6.1; 95% CI 7.23-12.51) and the waitlist control group (mean 13.3, SD 7.1; 95% CI 9.06-15.53) on PHQ-9 scores at t2 did not reveal significant group differences (*t*=–1.5678; *P*=.12) with moderate effect size (*d*=–0.5) suggesting some potential impact of the intervention on depression symptoms among completers, and the point-biserial correlation (*r*=–0.2351) suggest a trend points toward the intervention having a beneficial effect among completers (Multimedia Appendix 7).

ITT Analyses

The overall changes in secondary outcomes from t0 to t2 were also observed in the ITT analyses (Table 1).

There was significantly greater reduction in GAD-7 scores among participants in the intervention group (mean 8.47, SD 2.7; 95% CI 7.86-9.07) compared to the waitlist control group (mean 9.6, SD 3.8; 95% CI 8.79-10.47; t_{156}=–2.23; *P*=.03) at t2 with moderate effect size (*d*=–0.4), suggesting that the intervention had a meaningful impact on reducing GAD-7 scores. The point-biserial correlation (*r*=–0.176) supports the conclusion of a moderate effect (Multimedia Appendix 8).

There were marginally nonsignificant group differences between the intervention (mean 41.5, SD 8.40; 95% CI 36.65-43.41) and waitlist control group (mean 44.2, SD 8.9; 95% CI 42.21-46.22) in PSWQ-PW scores (t_{156}=–1.94; *P*=.05) at t2 with moderate effect size (*d*=–0.3), suggesting that the intervention had a modest impact on reducing PSWQ-PW scores. The point-biserial correlation (*r*=–0.154) supports the conclusion of a moderate effect (Multimedia Appendix 9).

The independent 2-tailed *t* test results comparing the intervention and the waitlist control group on PHQ-9 and WSAS scores at t2 did not reveal significant group differences with small effect sizes (Multimedia Appendices 10 and 11).

XSL•FO
RenderX

Table 1. Mean changes from baseline to posttest (n=158).

Secondary outcomes	Intervention (n=79), mean (SD)		Waitlist control (n=79), mean (SD)		Cohen *d*
	Baseline (t0)	Posttest (t2)	Baseline (t0)	Posttest (t2)	
GAD-7[a]	12.25 (3.70)	8.47 (2.71)[b]	11.03 (3.89)	9.63 (3.76)	–0.355
PHQ-9[c]	13.73 (5.16)	10.00 (3.99)	11.68 (4.99)	10.78 (5.40)	–0.165
PSWQ[d]	50.96 (8.65)	41.53 (8.40)	51.19 (8.90)	44.21 (8.96)	–0.309
WSAS[e]	20.94 (7.85)	17.21 (7.75)	17.89 (8.61)	18.25 (7.80)	–0.133

[a]GAD-7: Generalized Anxiety Disorder Scale-7.

[b]*P*<.05.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]PSWQ: Penn State Worry Questionnaire.

[e]WSAS: Work and Social Adjustment Scale.

Correlations Between the Support Measures and Secondary Outcomes

We investigated the correlations between the support measures (psychological support, professional support, and helpfulness of the professional support) at baseline (t0) and the improvements in the secondary outcomes to explore whether any of the support measures could have predicted the improvements in secondary outcomes at 6 weeks (t2; Multimedia Appendix 12).

The correlations between psychological support at t0 and the secondary outcomes were all negative, with the strongest negative correlation observed with WSAS scores (r=–0.186). This suggests that higher psychological support at t0 was associated with lower scores on secondary outcomes, indicating a reduction in anxiety, worry, depression, and functional impairment.

Overall, the correlations between professional support at t0 and the secondary outcomes were very weak. There is a slight positive correlation with PHQ-9 (r=0.058) and WSAS (r=0.126) scores. This implies that professional support at t0 had a minimal impact on secondary outcomes, indicating almost no relationship.

The correlations involving helpfulness of the professional support at t0 were also very weak, with the highest correlation being with WSAS (r=0.038) scores. This indicates that the perceived helpfulness of professional support at t0 had a negligible relationship with secondary outcomes.

Discussion

Principal Results

This pilot feasibility RCT included 158 Ulster University students, who at baseline reported moderate levels of generalized anxiety, depression, worry symptoms, and severe but functional levels of impairment. Students' interest in registering for the study, along with the safety of the trial procedures, which did not restrict access to the existing Student Well-being services, provided supportive evidence for the feasibility of the current trial. These procedures, including risk monitoring and participant autonomy in accessing care, align with ethical standards for low-risk digital intervention trials and offer a model for future studies in similar populations.

Although many participants engaged with the Cerina app, a notable proportion did not complete the desired number of sessions and skipped their 3-week assessments (t1). Nevertheless, the ITT analyses revealed significant group differences in GAD symptoms and marginally nonsignificant group differences in worry symptoms in favor of the intervention group with medium effect sizes at 6-week assessments (t2). These medium effect sizes are consistent with benchmarks reported in digital CBT-based interventions, which typically range between 0.3 and 0.6 for anxiety symptoms [19,21,23,26-31]. This indicates that even with limited engagement, the Cerina app was able to produce clinically meaningful improvements in GAD symptoms. However, it is plausible that adherence challenges may have diluted the observed effects, particularly in the ITT analyses.

Future iterations of the app could consider enhancing personalization, integrating gamification, or incorporating brief synchronous contact to optimize engagement and maximize therapeutic benefits. The digital format, along with the successful student-led recruitment, indicates that this intervention could be scaled across university contexts, especially where institutional counseling services are overburdened. However, to broaden accessibility, future adaptations may need to consider linguistic, cultural, and digital literacy differences to ensure inclusivity in more diverse educational or community-based settings.

Participants

Overall, the total number of students registered for the study (n=514) indicates university students' interest in line with the existing literature on the reach and access of the mental health apps among this population [33,67].

Despite the anticipated potential of social network channels such as Facebook and Instagram to support the recruitment for digital interventions [68,69], the majority of the students in this study self-registered through the university's global email system. It may be that different communication channels need

XSL•FO
RenderX

to be adapted to the target population's needs, enabling age- and culture-specific implementation strategies, as well as content adaptations to enhance relevance and effectiveness [35]. Furthermore, in the context of student populations, present results suggest that using recruitment channels that are of interest to students and using time-efficient strategies providing anonymity could be important during the recruitment process [33,67]. This highlights the effectiveness of digital, anonymous, and institution-based recruitment strategies in university settings. Compared to previous studies with limited uptake [32,47], our approach facilitated self-initiated participation with minimal burden on institutional services—an important strength for scalable interventions.

The baseline characteristics of the included participants in this study are consistent with findings from existing studies on digital CBT-based interventions [26-32] and the World Mental Health International College Student study [12,42,70]. Notably, in this study, the considerable number of students on medication because of mental health problems (n=53, 34%) provides further support for the prevalence of CMDs, their severity and complexity, particularly among first-year students transitioning from adolescence to adulthood [12]. Supporting this, data from a retrospective cohort study looking at the use of psychotropic medication among student populations between 2007 and 2019 in America reported a 2-fold increase in use of medication in nearly all classes of medication, including mood stabilizers and antianxiety medication [71]. In this study, 24.7% (n=39) of participants reported being on a stable dose of medication for the past 6 months. This is comparable to the number of students (n=65,899, 20.5%) who reported being prescribed medication because of mental health problems over the past 12 months [71].

The combination of high rates of psychotropic medication use and low uptake of psychological treatment among student populations in general [51] and in Northern Ireland specifically [12] underscores the potential of digital CBT-based interventions to improve access to first-line psychological help for GAD and other CMD symptoms among university students [72-77]. These findings are particularly relevant for universities with limited counseling capacity, as digital CBT offers a scalable adjunct or alternative pathway for mental health support.

Safety

In alignment with ethical standards, the safety procedures in this study were integrated with Ulster University's Student Well-being services, allowing participants to access the available psychological services throughout the study. Notably, none of the included participants reported a worsening of symptoms or requested to be signposted to the Student Well-being team for further help. These findings are consistent with existing RCTs indicating that digital interventions are safe to use among a diverse user population [74-77]. Furthermore, these findings have implications for future trials emphasizing the importance of such embedded strategies for the successful implementation of digital interventions in diverse settings [78,79].

Feasibility

Adherence to the Cerina app in this study was comparable to the existing literature on digital CBT-based interventions and mental health apps [67,69-81]. On average, 75% (59/79) of intervention group participants followed the sessions, while 25% (20/79) did not. This finding is consistent with a recent meta-analysis indicating that the attrition rate of mental health apps was 24.1% in the short-term and 35.5% in the longer-term follow-up [24].

The study also provided evidence of attitudinal adherence (emotional investment) to the Cerina app among the participants. In the intervention group, 61% (n=36) of participants followed the core psychoeducational and anxiety-related therapeutic content, while 25% (n=15) accessed the psychoeducational content only, and just a few (n=8, 14%) did not engage with any content. In comparison, in an open feasibility trial with university students in America, just a small portion of the sample (31/117, 26.5%) downloaded a mental health app, and 24% (28/117) of participants used the app only once [32]. In the same study, depression and anxiety-related content of the mental health app were only accessed by 11.1% (13/117) of participants, and users mostly accessed the psychoeducation lessons (27/117, 23.1%) within the app [32]. Thus, it could be argued that engagement features may need to be refined or combined with minimal human support to sustain behavioral adherence, particularly in unsupervised settings.

Several factors may account for the differences observed between this study and the previous feasibility trial by Lattie et al [32]. One explanation could be the differences in the baseline characteristics of the students in both studies. In our sample, the included participants had moderate depression, generalized anxiety, worry, and severe functional impairment, whereas the clinical symptom severity at the baseline is not known for the sample of students in Lattie et al's [32] study. It is plausible that the higher level of clinical need in our sample increased the perceived relevance of the intervention, thereby enhancing user engagement [79-81]. Alternatively, differences in the nature of the interventions may also explain the variation in adherence.

It is noted that in this study, several engagement features were incorporated into the UI of the Cerina app to optimize the user engagement [36]. It is possible these features improved initial engagement but did not sustain use over time. Future versions of the app might incorporate adaptive elements, social reinforcement features, or real-time feedback to maintain interest. As such, improving sustained user engagement, rather than recruitment, should be the primary focus of refinement. It is also possible that the engagement features helped make the psychoeducational and anxiety-related content more interactive in line with the supportive evidence in the existing literature [38,41]. It may also be beneficial to explore how motivational or cognitive framing in early sessions might better retain users, particularly for worry-specific modules, which showed marginal effects in ITT but stronger effects among completers.

Although the attitudinal adherence among waitlist group participants was similar to the intervention group participants, the behavioral adherence was much lower. It is well established that participation in waitlist groups often heightens expectations about the intervention, which may in turn motivate participants to remain in the study to access the intervention [82]. This pattern was also observed in this study, where significantly

XSL•FO
RenderX

more participants in the intervention group (n=41, 71%) failed to complete assessments at t1 and t2 compared to the waitlist control group (n=17, 29%). However, it could be that participants in the waitlist control group got better over time, and they did not feel the need to follow the Cerina app by the time they were given access. Such human-related factors, such as losing motivation, feeling better over time, and not feeling the need to access the intervention, are reported as common reasons for disengagement among waitlist group participants in the existing literature [82].

Despite the attitudinal adherence to the core clinical content of the Cerina app in this study, only a few participants in both the intervention (n=7, 12%) and the waitlist control group (n=1, 4%) completed the desired number of sessions, accessing the whole therapeutic content. This raises interpretive caution, particularly in the completer analyses, which may overestimate the impact of the intervention due to selective dropout. In contrast, the ITT analyses, although more conservative, offer a more comprehensive view of intervention effectiveness across the full sample. Future trials should aim to balance these two approaches and consider hybrid analytic models (eg, multiple imputation or pattern-mixture models) to address bias.

Strategies such as adaptive messaging or brief check-ins from human guides may also reduce dropout and strengthen trial integrity. The low engagement with the whole therapeutic content may point to potential usability issues in the real-world implementation of the Cerina app, despite overall acceptable usability scores. Such challenges are common to digital CBT-based interventions and mental health apps, often leading to poor adoption, implementation barriers, and poor long-term maintenance [79,80]. Known common human-related barriers include disappointment with the lack of human contact, the expectation that the intervention would mimic face-to-face psychotherapy, busy schedules, differing support needs, forgetfulness, and privacy concerns [33,34]. Intervention-related factors include lack of guidance regarding the customized content of the intervention and technical problems (eg, problems with navigation, bugs or glitches, and lack of flexibility with the UI) [33,34]. In this study, several engagement features were used, such as customized notifications, more flexibility regarding navigation, regular check-ins to rate mood, and a chatbot initiating more interaction with the user [36]. However, it could be argued that the evidence base for the effectiveness of these features on long-term adherence and adoption of the digital CBT-based interventions remains limited [38-40,83].

Although the recruitment channels used in this were effective in generating initial interest, the content engagement features alone might not be sufficient to optimize user engagement in the long run. Recent advancements in large language models (LLMs) have provided a promising framework for enhancing engagement features in digital interventions [84]. However, several challenges remain that limit their potential for improving user engagement, such as risk of biases, literacy gaps, and inequalities in data presentation [84]. Additionally, although chatbots and web-based therapist features in particular promise to initiate more interaction with users, the conversation flow can still be perceived as static and redundant by the users [83,84]. It could be that these key challenges contributed to the poor long-term adherence in this study. The ongoing process evaluation is expected to shed more light on these issues, potentially restricting the implementation of the Cerina app in real-life settings, especially after the 2nd week on the app [55]. Future trials could test combinations of interactive components, early engagement tactics, and optional human support to reduce attrition. Importantly, mental health apps using these advanced engagement features should be grounded in a sociocultural-technical framework [84]. Such a framework should provide a global repository of health information to train LLMs to ensure inclusivity, minimize biases, and inequalities in data presentation [84]. Additionally, the sociotechnical limitations of LLMs, such as the structure of a conversation flow with users in mental health settings, should be coproduced with users with lived experience, mental health professionals, and computer scientists [83,84].

Intervention Effects

Overall, there was an indication that the intervention group achieved better secondary outcomes compared to the waitlist control group. Both completers and ITT analyses showed significant improvements in GAD symptoms from t0 to t2. Therefore, it is pertinent to compare with other similar RCTs investigating the effectiveness of digital CBT-based interventions among clinical [74-77] and student populations [26-31]. The findings of this study align with the recent network meta-analysis supporting both acute and long-term effectiveness of CBT in reducing GAD symptoms among clinical populations [19]. However, the present results contradict the previous meta-analysis, indicating no evidence of a difference in the effects of the digital interventions compared to controls for college students with anxiety symptoms [85]. Similarly, the present findings on GAD symptoms contrast with a previous trial on a transdiagnostic CBT-based digital intervention for university students that found no evidence for the effects of the intervention on anxiety symptoms [26]. Possible explanations of this discrepancy include differences in the comparison group, baseline symptom severity, type of CBT intervention, length, and type of support [26].

Interestingly, although the completers analyses indicated significant improvements in worry symptoms in favor of the intervention with a large effect size, the ITT analyses showed marginally nonsignificant group differences in worry symptoms favoring the intervention. This may be due to early disengagement before the worry-specific modules were completed, suggesting that worry-focused content could be introduced earlier in the program to maximize impact across the full sample. In this study, only a small number of participants completed the desired number of sessions. It is well known that "worry" represents the cognitive part of anxiety that is characterized by unhelpful thoughts such as "I am worthless" and "I am a failure" [46,86]. Since the Cerina app follows a CBT-based protocol, the first 2 sessions focus on psychoeducation on GAD, conceptualizing one's anxiety symptoms, and setting up goals. Users start working on their worries after the 2nd session [46]. Therefore, though completing the first 2 sessions might have been sufficient to reduce the GAD symptoms in this study, it could be that receiving the intended dose of the intervention is crucial for the treatment of

XSL•FO
RenderX

the cognitive aspect of anxiety. Alternatively, the completers were more motivated to use the Cerina app, and this could have inflated their results [87]. A meta-analysis on existing RCTs supports the former argument and suggests that longer treatment (in weeks) was associated with better worry outcomes [73]. To enhance outcomes related to worry, future versions of the intervention could be adapted to include an earlier focus on cognitive aspects of anxiety. Additionally, engagement features could be leveraged to customize the UI to those presenting more severe worry symptoms. They could, for instance, receive notifications and reminders directing them to the sessions and CBT techniques focusing on cognitive aspects of anxiety more often. The effectiveness of these strategies in improving worry outcomes should be tested in future studies.

Despite the well-known close ties between anxiety, depression, and functional impairment (ie, subjective interpretation of the clinical symptoms) [88], the positive trend observed in treating depression did not persist in the ITT analyses. Similarly, there was no favorable effect of the intervention in treating the functional effects of GAD symptoms in the ITT analyses.

With regard to depression symptoms, the present findings contradict the previous meta-analysis, suggesting a large overall effect (g=–0.70) of digital CBT-based interventions for GAD on depression symptoms [73]. One plausible explanation for these contradictory results could be related to differences in the types of CBT-based interventions between the current and former studies. This study used a targeted CBT-based intervention for treating a single condition (GAD symptoms) rather than a transdiagnostic (unified) CBT-based intervention, which addresses overlapping characteristics of GAD and depression, such as uncontrollable worry about multiple future events and rumination (ie, repeatedly thinking about past or current concerns) [26]. Therefore, it could be argued that completing more sessions focusing on understanding and managing worry among the completers potentially had an effect on reducing their depression symptoms in this study.

Regarding functional impairment, the discrepancy between the ITT and completers analyses could be explained by greater engagement with the intervention among completers. Consistent with the present findings, a previous meta-analysis reported a large overall effect of digital CBT-based interventions on functional outcomes (g=−0.66) [73]. This suggests that user engagement and completing the intended number of sessions (ie, receiving the intended dose of the intervention) could play an important role in improving the functional effects of anxiety [86].

It is also likely that the improvements in secondary outcomes can be attributed to the group status (intervention group) rather than the access to usual care, which cannot be ruled out in this study. Even though receiving psychological support at baseline showed a slight trend toward improving secondary outcomes, all support measures (ie, professional support, psychological support from the informal network, and helpfulness of the professional support) generally showed weak correlations with the improvements in secondary outcomes. This could indicate that accessing usual care (ie, psychological support from informal networks or professional support) alone may not be a strong predictor of improvements in secondary outcomes in this study.

Clinical Implications

This pilot feasibility trial highlights several important clinical implications. Findings suggest that digital CBT interventions such as Cerina can be feasibly and safely integrated into university well-being services, offering accessible and scalable support for student mental health. These tools are particularly valuable for students who may be reluctant to seek face-to-face help due to stigma, limited time, or other barriers.

Universities could use this emerging evidence to inform the development, procurement, or commissioning of digital mental health resources. Specific features likely to enhance uptake and sustained engagement include user-friendly onboarding, options for personalization, and interactive content that maintains interest over time.

Digital CBT interventions can also be embedded within stepped care models in both university and primary care settings. As low-intensity, first-line options, platforms like Cerina may help manage the growing demand for mental health support by providing early intervention and reducing pressure on mental health services and general practices. In primary care, these tools could assist in more effective triage, support self-management, and facilitate continuity of care for people with emerging or mild symptoms, or those referred but awaiting treatment by mental health services.

Importantly, presenting digital CBT as a universal well-being resource rather than a treatment may help reduce stigma and increase engagement. As both education and health systems look to expand capacity and address mental health needs more efficiently, interventions like Cerina can offer a flexible, cost-effective, and user-centered solution.

Strengths and Limitations

There are several limitations to this study. As discussed above, the attrition rate was one of the main limitations, as this can result in an imbalance occurring by chance and can introduce bias [87]. However, our analyses confirmed that the data missing due to attrition were missing completely at random, and variables in this study were unaffected by the attrition between time points. Notably, completer and ITT analyses showed differing outcomes. While ITT offers a more conservative estimate of effectiveness, completers analyses may be biased due to dropout. [87]. It could be that many participants dropped out due to the inefficacy of the intervention, or they could have felt better and did not feel the need to use the Cerina app anymore. Therefore, the completers analyses might fail to capture the limitations or potential downsides of the intervention [87]. Understanding patterns of dropout can inform future designs aimed at minimizing attrition, such as adaptive intervention strategies or just-in-time support.

Another limitation relates to the sample size. Since the sample size is attenuated in a completers analysis, the statistical power is reduced. Additionally, the benefits of randomization may be compromised due to changes in group composition [87]. In contrast, the ITT sample in the present study comprises all

XSL•FO
RenderX

participants who were randomized, reducing the likelihood of selection bias. Additionally, to minimize any bias in the ITT sample, we tested several imputation models and selected a linear regression model, which confirmed the data distribution while preserving an acceptable level of variance within the imputed data.

Additionally, this study was designed as a feasibility pilot RCT; therefore, it was not powered to detect significant intervention effects [36]. Despite this limitation, we exceeded the intended sample size, and both the completers and the ITT analyses indicated significant group differences in GAD symptoms in favor of the intervention group with a medium effect size. This finding shows the meaningful impact of the intervention and suggests that the effect size of the intervention is likely to be bigger in a future full-scale RCT.

Another important limitation of this study concerns the operationalization of adherence (referred to as engagement in the Staudt model) [37,48]. The way we defined behavioral and attitudinal adherence might not be representative of the multidimensionality of the construct. For example, the way users navigate sessions or engagement sequences was not measured in this study, and they may yield additional insights [34,47]. Furthermore, unlike the approach taken by Hanano et al [48], we did not operationalize attitudinal adherence as "the extensiveness of users' open-ended reflective responses on weekly check-ins" [47]. Instead, we focused on participants' engagement with the content and relied on the flow of the CBT protocol to define different levels of attitudinal adherence in this study [36]. This approach is more efficient and clinically relevant [66]. Another approach is to compare the qualitative and quantitative analyses of user engagement [55]. The process evaluation will enable us to combine different aspects and facilitate a more accurate understanding of adherence in this study [55].

A further limitation is that the sample of this study is not representative of the university student population in Northern Ireland, which limits the generalizability of the findings [36]. However, the demographic and clinical characteristics of the current sample mimic the previous findings on these characteristics among student populations in Northern Ireland [42,70] and globally [88].

Finally, it is well established that participation in a waitlist control group can increase motivation to remain in the trial and may elevate participants' expectations regarding the intervention's effectiveness [89,90]. This cannot be ruled out in this study. The waitlist control group was chosen as a comparison group to offer the possibility of completing the intervention to a vulnerable population who might not have access to any form of psychological help otherwise.

There are several strengths of this study that warrant acknowledgment. First, a recent audit of the health apps available in the Apple store highlighted that less than half (n=220, 44%) claimed to be evidence-based, and more than 70% of these evidence-based apps relied on low-quality evidence [91]. Moreover, significant methodological limitations were identified in the small subset of developers who published studies [91]. This pilot feasibility study addresses this gap and contributes high-quality evidence on the safety, feasibility, and effectiveness of mental health apps in general.

Second, the low-threshold CBT-based mobile app in this study was tailored to the student population, developed using co-design principles [36]. This approach informed the prioritization of specific engagement features and enabled the evaluation of their impact on user engagement. The study contributes to the growing body of evidence suggesting that further research, particularly with longer-term follow-up, is needed to better understand the role of engagement features in treatment outcomes [38,44].

Another notable strength is the collaborative partnership underpinning the study involving Ulster University, the Ulster University Student Well-being team, and Cerina Health (a mental health start-up company). This collaboration contributed to both the reach and acceptability of the Cerina app among students, supporting its alignment with the goals of university-based digital mental health interventions.

Conclusions

In conclusion, the design of this trial provides the groundwork for future development of full-scale RCTs embedded in Student Well-being services. Though this study was designed as a feasibility pilot trial, it exceeded the intended sample size (n=90) and was successfully conducted in a university setting. Furthermore, the results highlight the likely effects of the Cerina app in reducing anxiety symptoms, offering a basis for further research and potential implementation of the intervention in broader contexts to alleviate anxiety. Importantly, the usability issues identified in this study underscore the need for iterative refinement of digital CBT-based interventions. Ongoing feedback and the evaluation of long-term effects of engagement features will be essential to enhancing user experience and sustaining engagement in future trials.

Acknowledgments

The authors would like to acknowledge Momotaj Islam (Cerina Health Chief Psychologist) for writing the cognitive behavioral therapy–based treatment protocol for the Cerina app. Additionally, Charlotte Monnickendam (Cerina Health Head of Partnerships) assisted the recruitment process, Omkar Kadam (Cerina Health Head of Innovation) and Priyadharshini Rajaram (Cerina Health Technical Product Manager) managed the randomization procedure, and provided technical help whenever it was needed by the participants. The authors also wish to thank Prof Nicos Savva and Dr Kurtis Ashcroft for their advice and input on the data analysis. They would like to thank the Student Well-being Team, Ulster University, and Ulster University Students' Union for their collaboration and support throughout the study. Finally, the authors would like to extend their gratitude to all the participants for their time and valuable contributions to this study. This work was conducted as part of the North West Europe INTERREG

XSL•FO
RenderX

IT4Anxiety project in partnership between Cerina Therapeutics and Ulster University. The project is supported by the European Commission. The sponsor was not involved in the writing of the manuscript or the decision to submit it for publication.

Authors' Contributions

OEB: writing—original draft, review and editing, conceptualization, methodology, project administration.
TR: writing—review and editing, conceptualization, methodology.
MM: writing—review and editing, conceptualization, methodology.
SP: writing—review and editing, project administration, methodology.
MO: writing—review and editing, methodology.
PM: writing—review and editing, conceptualization.
JC: writing—review and editing, conceptualization.
MP: writing—review and editing, methodology.
GL: writing—review and editing, conceptualization.

Conflicts of Interest

PM is the Chief Executive Officer of Cerina Health and was not involved in data collection, analysis, or interpretation. All other authors declare no conflicts of interest.

Multimedia Appendix 1
CONSORT 2010 checklist.
[PDF File (Adobe PDF File), 86 KB - mhealth_v13i1e70691_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 619 KB - mhealth_v13i1e70691_app2.pdf]

Multimedia Appendix 3
Baseline characteristics.
[DOCX File , 22 KB - mhealth_v13i1e70691_app3.docx]

Multimedia Appendix 4
(Left) the change in GAD7 score distributions across t0-t2, (middle) the datapoint swam of GAD7 total scores for t0-t2, (right) the change in mean GAD7 score across t0-t2 among completers.
[PNG File , 109 KB - mhealth_v13i1e70691_app4.png]

Multimedia Appendix 5
(Left) the change in PSWQ score distributions across t0-t2, (middle) the datapoint swam of GAD7 total scores for t0-t2, (right) the change in mean PSWQ score across t0-t2 among completers.
[PNG File , 114 KB - mhealth_v13i1e70691_app5.png]

Multimedia Appendix 6
(Left) the change in PHQ9 score distributions across t0-t2, (middle) the datapoint swam of GAD7 total scores for t0-t2, (right) the change in mean PHQ9 score across t0-t2 among completers.
[PNG File , 103 KB - mhealth_v13i1e70691_app6.png]

Multimedia Appendix 7
(Left) the change in WSAS score distributions across t0-t2, (middle) the datapoint swam of GAD7 total scores for t0-t2, (right) the change in mean WSAS score across t0-t2 among completers.
[PNG File , 113 KB - mhealth_v13i1e70691_app7.png]

Multimedia Appendix 8
(Left) the change in GAD7 score distributions across t0-t2, (middle) the datapoint swam of GAD7 total scores for t0-t2, (right) the change in mean GAD7 score across t0-t2 among the ITT sample.
[PNG File , 112 KB - mhealth_v13i1e70691_app8.png]

Multimedia Appendix 9

XSL•FO
RenderX

(Left) the change in PSWQ score distributions across t0-t2, (middle) the datapoint swam of GAD7 total scores for t0-t2, (right) the change in mean PSWQ score across t0-t2 among the ITT sample.
[PNG File , 128 KB - mhealth_v13i1e70691_app9.png]

Multimedia Appendix 10
(Left) the change in WSAS score distributions across t0-t2, (middle) the datapoint swam of GAD7 total scores for t0-t2, (right) the change in mean WSAS score across t0-t2 among the ITT sample.
[PNG File , 126 KB - mhealth_v13i1e70691_app10.png]

Multimedia Appendix 11
(Left) the change in PHQ9 score distributions across t0-t2, (middle) the datapoint swam of GAD7 total scores for t0-t2, (right) the change in mean PHQ9 score across t0-t2 among the ITT sample.
[PNG File , 117 KB - mhealth_v13i1e70691_app11.png]

Multimedia Appendix 12
Correlation Matrix.
[PNG File , 43 KB - mhealth_v13i1e70691_app12.png]

References

1. Anxiety disorders. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders [accessed 2023-07-03]
2. Ruscio AM, Hallion LS, Lim CCW, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, et al. Cross-sectional comparison of the epidemiology of DSM-5 generalized anxiety disorder across the globe. JAMA Psychiatry 2017;74(5):465-475 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0056] [Medline: 28297020]
3. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
4. Szuhany KL, Simon NM. Anxiety disorders: a review. JAMA 2022;328(24):2431-2445. [doi: 10.1001/jama.2022.22744] [Medline: 36573969]
5. Sampson K, Priestley M, Dodd AL, Broglia E, Wykes T, Robotham D, et al. Key questions: research priorities for student mental health. BJPsych Open 2022;8(3):e90 [FREE Full text] [doi: 10.1192/bjo.2022.61] [Medline: 35535504]
6. Schulenberg JE, Sameroff AJ, Cicchetti D. The transition to adulthood as a critical juncture in the course of psychopathology and mental health. Dev Psychopathol 2004;16(4):799-806. [doi: 10.1017/s0954579404040015] [Medline: 15704815]
7. de Girolamo G, Dagani J, Purcell R, Cocchi A, McGorry PD. Age of onset of mental disorders and use of mental health services: needs, opportunities and obstacles. Epidemiol Psychiatr Sci 2012;21(1):47-57. [doi: 10.1017/s2045796011000746] [Medline: 22670412]
8. Arnett JJ. College students as emerging adults: the developmental implications of the college context. Emerg Adulthood 2015;4(3):219-222. [doi: 10.1177/2167696815587422]
9. Acharya L, Jin L, Collins W. College life is stressful today—emerging stressors and depressive symptoms in college students. J Am Coll Health 2018;66(7):655-664. [doi: 10.1080/07448481.2018.1451869] [Medline: 29565759]
10. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, et al. WHO world mental health surveys international college student project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]
11. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
12. Ebert DD, Mortier P, Kaehlke F, Bruffaerts R, Baumeister H, Auerbach RP, et al. Barriers of mental health treatment utilization among first-year college students: first cross-national results from the WHO world mental health international college student initiative. Int J Methods Psychiatr Res 2019;28(2):e1782 [FREE Full text] [doi: 10.1002/mpr.1782] [Medline: 31069905]
13. Voltmer E, Köslich-Strumann S, Walther A, Kasem M, Obst K, Kötter T. The impact of the COVID-19 pandemic on stress, mental health and coping behavior in German University students—a longitudinal study before and after the onset of the pandemic. BMC Public Health 2021;21(1):1385 [FREE Full text] [doi: 10.1186/s12889-021-11295-6] [Medline: 34256717]
14. Saraswathi I, Saikarthik J, Kumar KS, Srinivasan KM, Ardhanaari M, Gunapriya R. Impact of COVID-19 outbreak on the mental health status of undergraduate medical students in a COVID-19 treating medical college: a prospective longitudinal study. PeerJ 2020;8:e10164 [FREE Full text] [doi: 10.7717/peerj.10164] [Medline: 33088628]

XSL•FO
RenderX

15. Reuter PR, Forster BL, Kruger BJ. A longitudinal study of the impact of COVID-19 restrictions on students' health behavior, mental health and emotional well-being. PeerJ 2021;9:e12528 [FREE Full text] [doi: 10.7717/peerj.12528] [Medline: 34993018]
16. Calkins AW, Otto MW, Cohen LS, Soares CN, Vitonis AF, Hearon BA, et al. Psychosocial predictors of the onset of anxiety disorders in women: results from a prospective 3-year longitudinal study. J Anxiety Disord 2009;23(8):1165-1169 [FREE Full text] [doi: 10.1016/j.janxdis.2009.07.022] [Medline: 19699609]
17. Pei J, Amanvermez Y, Vigo D, Puyat J, Kessler RC, Mortier P, et al. Sociodemographic correlates of mental health treatment seeking among college students: a systematic review and meta-analysis. Psychiatr Serv 2024;75(6):556-569. [doi: 10.1176/appi.ps.20230414] [Medline: 38291886]
18. Frost DM, Meyer IH. Minority stress theory: application, critique, and continued relevance. Curr Opin Psychol 2023;51:101579 [FREE Full text] [doi: 10.1016/j.copsyc.2023.101579] [Medline: 37270877]
19. Papola D, Miguel C, Mazzaglia M, Franco P, Tedeschi F, Romero SA, et al. Psychotherapies for generalized anxiety disorder in adults: a systematic review and network meta-analysis of randomized clinical trials. JAMA Psychiatry 2024;81(3):250-259 [FREE Full text] [doi: 10.1001/jamapsychiatry.2023.3971] [Medline: 37851421]
20. Generalised anxiety disorder and panic disorder in adults: management. National Institute for Health and Care Excellence (NICE). 2011. URL: https://www.nice.org.uk/guidance/cg113 [accessed 2020-06-15]
21. Cuijpers P, Miguel C, Harrer M, Plessen CY, Ciharova M, Ebert D, et al. Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. World Psychiatry 2023;22(1):105-115 [FREE Full text] [doi: 10.1002/wps.21069] [Medline: 36640411]
22. Comer JS, Barlow DH. The occasional case against broad dissemination and implementation: retaining a role for specialty care in the delivery of psychological treatments. Am Psychol 2014;69(1):1-18 [FREE Full text] [doi: 10.1037/a0033582] [Medline: 23915401]
23. Cuijpers P, Eylem O, Karyotaki E, Zhou X, Sijbrandij M. Psychotherapy for depression and anxiety in low- and middle-income countries. In: Global Mental Health and Psychotherapy: Adapting Psychotherapy for Low- and Middle-Income Countries. London: Academic Press; 2019:173-192.
24. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. World Psychiatry 2024;23(1):139-149 [FREE Full text] [doi: 10.1002/wps.21183] [Medline: 38214614]
25. Ebert DD, Harrer M, Apolinário-Hagen J, Baumeister H. Digital interventions for mental disorders: key features, efficacy, and potential for artificial intelligence applications. Adv Exp Med Biol 2019;1192:583-627. [doi: 10.1007/978-981-32-9721-0_29] [Medline: 31705515]
26. Karyotaki E, Klein AM, Ciharova M, Bolinski F, Krijnen L, de Koning L, et al. Guided internet-based transdiagnostic individually tailored cognitive behavioral therapy for symptoms of depression and/or anxiety in college students: a randomized controlled trial. Behav Res Ther 2022;150:104028 [FREE Full text] [doi: 10.1016/j.brat.2021.104028] [Medline: 35066365]
27. Weisel KK, Zarski AC, Berger T, Krieger T, Schaub MP, Moser CT, et al. Efficacy and cost-effectiveness of guided and unguided internet- and mobile-based indicated transdiagnostic prevention of depression and anxiety (ICare Prevent): a three-armed randomized controlled trial in four European countries. Internet Interventions 2019;16:52-64 [FREE Full text] [doi: 10.1016/j.invent.2018.04.002] [Medline: 30775265]
28. Weisel KK, Zarski AC, Berger T, Krieger T, Moser CT, Schaub MP, et al. User experience and effects of an individually tailored transdiagnostic internet-based and mobile-supported intervention for anxiety disorders: mixed-methods study. J Med Internet Res 2020;22(9):e16450 [FREE Full text] [doi: 10.2196/16450] [Medline: 32936085]
29. Rahmadiana M, Karyotaki E, Schulte M, Ebert DD, Passchier J, Cuijpers P, et al. Transdiagnostic internet intervention for Indonesian university students with depression and anxiety: evaluation of feasibility and acceptability. JMIR Ment Health 2021;8(3):e20036 [FREE Full text] [doi: 10.2196/20036] [Medline: 33666553]
30. Lee RA, Jung ME. Evaluation of an mHealth app (DeStressify) on university students' mental health: pilot trial. JMIR Ment Health 2018;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
31. Becker TD, Torous JB. Recent developments in digital mental health interventions for college and university students. Curr Treat Options Psych 2019;6(3):210-220. [doi: 10.1007/s40501-019-00178-8]
32. Lattie EG, Cohen KA, Hersch E, Williams KD, Kruzan KP, MacIver C, et al. Uptake and effectiveness of a self-guided mobile app platform for college student mental health. Internet Interventions 2022;27:100493 [FREE Full text] [doi: 10.1016/j.invent.2021.100493] [Medline: 35036331]
33. Ciharova M, Cuijpers P, Amanvermez Y, Riper H, Klein AM, Bolinski F, et al. Use of tailoring features and reasons for dropout in a guided internet-based transdiagnostic individually-tailored cognitive behavioral therapy for symptoms of depression and/or anxiety in college students. Internet Interventions 2023;34:100646 [FREE Full text] [doi: 10.1016/j.invent.2023.100646] [Medline: 38099094]
34. Gericke F, Ebert DD, Breet E, Auerbach RP, Bantjes J. A qualitative study of university students' experience of internet‐based CBT for depression. Couns Psychother Res 2021;21(4):792-804. [doi: 10.1002/capr.12465]

XSL•FO
RenderX

35. Elkes J, Cro S, Batchelor R, O'Connor S, Yu L, Bell L, et al. User engagement in clinical trials of digital mental health interventions: a systematic review. BMC Med Res Methodol 2024;24(1):184 [FREE Full text] [doi: 10.1186/s12874-024-02308-0] [Medline: 39182064]
36. Eylem-van Bergeijk O, Poulter S, Ashcroft K, Robinson T, Mane P, Islam M, et al. Cerina: cognitive-behavioural therapy-based mobile application for managing GAD symptoms among Ulster University students in Northern Ireland—a protocol for a pilot feasibility randomised controlled trial. BMJ Open 2024;14(6):e083554 [FREE Full text] [doi: 10.1136/bmjopen-2023-083554] [Medline: 38950994]
37. Staudt M. Treatment engagement with caregivers of at-risk children: gaps in research and conceptualization. J Child Fam Stud 2006;16(2):183-196. [doi: 10.1007/s10826-006-9077-2]
38. Gan DZQ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user engagement with digital mental health interventions: a systematic review. Digital Health 2022;8:20552076221098268 [FREE Full text] [doi: 10.1177/20552076221098268] [Medline: 35677785]
39. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
40. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022;9(1):e30766 [FREE Full text] [doi: 10.2196/30766] [Medline: 35099398]
41. Choi SK, Bruehlman-Senecal E, Green A, Lavra J, Bauermeister J. Patterns of engagement in digital mental health intervention for LGBTQ+ youth: a latent profile analysis. Front Digital Health 2023;5:1254929 [FREE Full text] [doi: 10.3389/fdgth.2023.1254929] [Medline: 38075522]
42. McLafferty M, Lapsley CR, Ennis E, Armour C, Murphy S, Bunting BP, et al. Mental health, behavioural problems and treatment seeking among students commencing university in Northern Ireland. PLoS One 2017;12(12):e0188785 [FREE Full text] [doi: 10.1371/journal.pone.0188785] [Medline: 29236727]
43. Broglia E, Ryan G, Williams C, Fudge M, Knowles L, Turner A, et al. Profiling student mental health and counselling effectiveness: lessons from four UK services using complete data and different outcome measures. Br J Guid Couns 2021;51(2):204-222. [doi: 10.1080/03069885.2020.1860191]
44. Borghouts J, Eikey EV, Mark G, De Leon C, Schueller SM, Schneider M, et al. Understanding mental health app use among community college students: web-based survey study. J Med Internet Res 2021;23(9):e27745 [FREE Full text] [doi: 10.2196/27745] [Medline: 34519668]
45. Andersson E, Enander J, Andrén P, Hedman E, Ljótsson B, Hursti T, et al. Internet-based cognitive behaviour therapy for obsessive-compulsive disorder: a randomized controlled trial. Psychol Med 2012;42(10):2193-2203 [FREE Full text] [doi: 10.1017/S0033291712000244] [Medline: 22348650]
46. Roth, Pilling. The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders: best practise guidance, PTSD, Ehlers and Clark model. IAPT. 2007. URL: http://www.ucl.ac.uk/clinical-psychology/CORE/CBT_Competences/CBT_Competence_List.pdf [accessed 2007-09-10]
47. Eylem O, van Straten A, de Wit L, Rathod S, Bhui K, Kerkhof AJFM. Reducing suicidal ideation among Turkish migrants in the Netherlands and in the UK: the feasibility of a randomised controlled trial of a guided online intervention. Pilot Feasibility Stud 2021 Jan 25;7(1):30 [FREE Full text] [doi: 10.1186/s40814-021-00772-9] [Medline: 33494831]
48. Hanano M, Rith-Najarian L, Boyd M, Chavira D. Measuring adherence within a self-guided online intervention for depression and anxiety: secondary analyses of a randomized controlled trial. JMIR Ment Health 2022;9(3):e30754 [FREE Full text] [doi: 10.2196/30754] [Medline: 35343901]
49. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
50. Spitzer R, Kroenke K, Williams J, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
51. Stöber J, Bittencourt J. Weekly assessment of worry: an adaptation of the Penn State worry questionnaire for monitoring changes during treatment. Behav Res Ther 1998;36(6):645-656 [FREE Full text] [doi: 10.1016/s0005-7967(98)00031-x] [Medline: 9648338]
52. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
53. Mundt JC, Marks IM, Shear MK, Greist JM. The work and social adjustment scale: a simple measure of impairment in functioning. Br J Psychiatry 2002;180(05):461-464. [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]
54. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of patient health questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019;365:l1781 [FREE Full text] [doi: 10.1136/bmj.l1781] [Medline: 30979729]
55. Eylem-van Bergeijk O, Islam M, Panagioti M, Mane P, Poulter S, Condell S, et al. Acceptability, use and implementation of the Cerina app: qualitative findings from the Cerina-pilot feasibility randomised controlled trial. Manuscript in preparation 2026.

XSL•FO
RenderX

56. Eysenbach G, Stoner S, Drozd F. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
57. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2023.
58. Laube M. The paired t-test and linear mixed models. Medium. 2021. URL: https://medium.com/@marco_laube/the-paired-t-test-and-linear-mixed-models-185a084d7813 [accessed 2025-08-30]
59. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006;59(10):1087-1091. [doi: 10.1016/j.jclinepi.2006.01.014] [Medline: 16980149]
60. Raudenbush SW. Comparing personal trajectories and drawing causal inferences from longitudinal data. Annu Rev Psychol 2001;52:501-525. [doi: 10.1146/annurev.psych.52.1.501] [Medline: 11148315]
61. Python: programming language. Python Software Foundation. 2023. URL: https://www.python.org/ [accessed 2025-08-30]
62. Zhang Z. Missing data imputation: focusing on single imputation. Ann Transl Med 2016;4(1):9 [FREE Full text] [doi: 10.3978/j.issn.2305-5839.2015.12.38] [Medline: 26855945]
63. Bennett DA. How can I deal with missing data in my study? Aust N Z J Public Health 2001;25(5):464-469. [Medline: 11688629]
64. Demir S. Comparison of normality tests in terms of sample sizes under different skewness and kurtosis coefficients. Int J Assess Tools Educ 2022;9(2):397-409. [doi: 10.21449/ijate.1101295]
65. Osman MS, Abu-Mahfouz AM, Page PR. A survey on data imputation techniques: water distribution system as a use case. IEEE Access 2018;6:63279-63291. [doi: 10.1109/access.2018.2877269]
66. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ, USA: Lawrence Erlbaum Associates; 1988.
67. Melcher J, Camacho E, Lagan S, Torous J. College student engagement with mental health apps: analysis of barriers to sustained use. J Am Coll Health 2022;70(6):1819-1825. [doi: 10.1080/07448481.2020.1825225] [Medline: 33048626]
68. Brooks H, Devereux-Fitzgerald A, Richmond L, Bee P, Lovell K, Caton N, et al. Assessing the effectiveness of social network interventions for adults with a diagnosis of mental health problems: a systematic review and narrative synthesis of impact. Soc Psychiatry Psychiatr Epidemiol 2022;57(5):907-925 [FREE Full text] [doi: 10.1007/s00127-022-02242-w] [Medline: 35138427]
69. Zarrabian S, Nooripour R, Babaeian-koopaie M, Hasani-abharian P. Psychometric properties of the persian version of the instagram addiction test (IAT) in the Iranian population. Int J Behav Sci 2024;17(4):186-194. [doi: 10.30491/ijbs.2024.404557.1981]
70. Lagdon S, Ross J, Robinson M, Contractor AA, Charak R, Armour C. Assessing the mediating role of social support in childhood maltreatment and psychopathology among college students in Northern Ireland. J Interpers Violence 2021;36(3-4):NP2112-2136NP. [doi: 10.1177/0886260518755489] [Medline: 29448910]
71. Morris MR, Hoeflich CC, Nutley S, Ellingrod VL, Riba MB, Striley CW. Use of psychiatric medication by college students: a decade of data. Pharmacotherapy 2021;41(4):350-358 [FREE Full text] [doi: 10.1002/phar.2513] [Medline: 33583071]
72. Ennis E, McLafferty M, Murray E, Lapsley C, Bjourson T, Armour C, et al. Readiness to change and barriers to treatment seeking in college students with a mental disorder. J Affect Disord 2019;252:428-434. [doi: 10.1016/j.jad.2019.04.062] [Medline: 31003112]
73. Eilert N, Enrique A, Wogan R, Mooney O, Timulak L, Richards D. The effectiveness of internet-delivered treatment for generalized anxiety disorder: an updated systematic review and meta-analysis. Depress Anxiety 2021;38(2):196-219 [FREE Full text] [doi: 10.1002/da.23115] [Medline: 33225589]
74. Carl JR, Miller CB, Henry AL, Davis ML, Stott R, Smits JAJ, et al. Efficacy of digital cognitive behavioral therapy for moderate-to-severe symptoms of generalized anxiety disorder: a randomized controlled trial. Depress Anxiety 2020;37(12):1168-1178. [doi: 10.1002/da.23079] [Medline: 32725848]
75. Miller CB, Gu J, Henry AL, Davis ML, Espie CA, Stott R, et al. Feasibility and efficacy of a digital CBT intervention for symptoms of generalized anxiety disorder: a randomized multiple-baseline study. J Behav Ther Exp Psychiatry 2021;70:101609. [doi: 10.1016/j.jbtep.2020.101609] [Medline: 32950939]
76. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
77. Paxling B, Almlöv J, Dahlin M, Carlbring P, Breitholtz E, Eriksson T, et al. Guided internet-delivered cognitive behavior therapy for generalized anxiety disorder: a randomized controlled trial. Cogn Behav Ther 2011;40(3):159-173. [doi: 10.1080/16506073.2011.576699] [Medline: 21770848]
78. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
79. Milne-Ives M, Homer SR, Andrade J, Meinert E. Potential associations between behavior change techniques and engagement with mobile health apps: a systematic review. Front Psychol 2023;14:1227443 [FREE Full text] [doi: 10.3389/fpsyg.2023.1227443] [Medline: 37794916]

XSL•FO
RenderX

80. Ramos G, Hernandez-Ramos R, Taylor M, Schueller SM. State of the science: Using digital mental health interventions to extend the impact of psychological services. Behav Ther 2024;55(6):1364-1379. [doi: 10.1016/j.beth.2024.04.004] [Medline: 39443071]
81. Chan AHY, Honey MLL. User perceptions of mobile digital apps for mental health: acceptability and usability—an integrative review. J Psychiatr Ment Health Nurs 2022;29(1):147-168. [doi: 10.1111/jpm.12744] [Medline: 33604946]
82. Tønning ML, Kessing LV, Bardram JE, Faurholt-Jepsen M. Methodological challenges in randomized controlled trials on smartphone-based treatment in psychiatry: systematic review. J Med Internet Res 2019;21(10):e15362 [FREE Full text] [doi: 10.2196/15362] [Medline: 31663859]
83. Zilcha-Mano S. A glance into the future of artificial intelligence-enhanced scalable personalized training: a response to Kopelovich, Brian, et al. (2025) and Kopelovich, Slevin, et al. (2025). Psychotherapy (Chic) 2025;62(1):22-27. [doi: 10.1037/pst0000547] [Medline: 40095966]
84. Malgaroli M, Schultebraucks K, Myrick KJ, Loch AA, Ospina-Pinillos L, Choudhury T, et al. Large language models for the mental health community: framework for translating code to care. Lancet Digital Health 2025;7(4):e282-e285. [doi: 10.1016/s2589-7500(24)00255-3]
85. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
86. Erickson TM, Scarsella G, Clark DA, Beck AT. The anxiety and worry workbook: the cognitive behavioral solution New York: the Guilford Press. Cogn Behav Pract 2013;20(1):119-120. [doi: 10.1016/j.cbpra.2012.06.004]
87. Ahn E, Kang H. Intention-to-treat versus as-treated versus per-protocol approaches to analysis. Korean J Anesthesiol 2023;76(6):531-539 [FREE Full text] [doi: 10.4097/kja.23278] [Medline: 38031328]
88. Cao H, Wu Y, Yin H, Sun Y, Yuan H, Tao M. Global trends in the incidence of anxiety disorders from 1990 to 2019: joinpoint and age-period-cohort analysis study. JMIR Public Health Surveill 2024;10:1-15 [FREE Full text] [doi: 10.2196/49609] [Medline: 38285497]
89. Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research: study design: randomised controlled trials. BJOG 2018;125(13):1716-1716 [FREE Full text] [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
90. Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. Acta Psychiatr Scand 2014;130(3):181-192. [doi: 10.1111/acps.12275] [Medline: 24697518]
91. Morley J, Laitila J, Ross JS, Schamroth J, Zhang J, Floridi L. An app a day will (probably not) keep the doctor away: an evidence audit of health and medical apps available on the apple app store. Minds Mach 2025;35(1):11. [doi: 10.1007/s11023-025-09710-7]

Abbreviations

CBT: cognitive behavioral therapy
CMD: common mental disorder
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GAD: generalized anxiety disorder
GAD-7: Generalized Anxiety Disorder Scale-7
GP: general practitioner
ITT: intention to treat
LLM: large language model
PHQ-9: Patient Health Questionnaire-9
PSWQ: Penn State Worry Questionnaire
RCT: randomized controlled trial
SUS: System Usability Scale
UI: user interface
WHO: World Health Organization
WSAS: Work Social Adjustment Scale

XSL•FO
RenderX

Edited by L Buis; submitted 30.12.24; peer-reviewed by R Nooripour, S Locke; comments to author 19.02.25; revised version received 07.05.25; accepted 04.08.25; published 09.10.25.

Please cite as:
Eylem-van Bergeijk O, Robinson T, Manktelow M, Olympios M, Poulter S, Mane P, Panagioti M, Condell J, Leavey G
Cerina—Cognitive Behavioral Therapy–Based Mobile App for Managing Generalized Anxiety Disorder Symptoms Among University Students: Results From a Pilot Feasibility Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e70691
URL: https://mhealth.jmir.org/2025/1/e70691
doi:10.2196/70691
PMID:41066166

©Ozlem Eylem-van Bergeijk, Tony Robinson, Matthew Manktelow, Michail Olympios, Siobhan Poulter, Prasannajeet Mane, Maria Panagioti, Joan Condell, Gerard Leavey. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 09.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Participant Evaluation of Blockchain-Enhanced Women's Health Research Apps: Mixed Methods Experimental Study

Madelena Y Ng[1,2], DrPH; Jodi Halpern[1], MD, PhD; Olivia Shane[1], MPH; Tina Teng[3], BA, BS; Michael Nguy n[4], MBA; Casey Ryan Alt[4], MA, MFA; Anaïs Barthe Leite[4], BA; Sean Moss-Pultz[4], BS; Courtney R Lyles[1,5,6*], PhD; Coye Cheshire[7*], PhD

[1]School of Public Health, University of California, Berkeley, Berkeley, CA, United States

[2]Department of Medicine (Biomedical Informatics), Stanford University, Stanford, CA, United States

[3]Haas School of Business, University of California, Berkeley, Berkeley, CA, United States

[4]Bitmark, Inc, Taipei, Taiwan

[5]Department of Public Health Sciences, University of California, Davis, Davis, CA, United States

[6]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[7]School of Information, University of California, Berkeley, Berkeley, CA, United States

[*]these authors contributed equally

Corresponding Author:
Madelena Y Ng, DrPH
School of Public Health
University of California, Berkeley
2121 Berkeley Way
Berkeley, CA, 94704
United States
Phone: 1 5106426000
Email: madelena.ng@berkeley.edu

Abstract

Background: Blockchain technology has capabilities that can transform how sensitive personal health data are safeguarded, shared, and accessed in digital health research. Women's health data are considered especially sensitive, given the privacy and safety risks associated with their unauthorized disclosure. These risks may affect research participation. Using a privacy-by-design approach, we developed 2 app-based women's health research study prototypes for user evaluation and assessed how blockchain may impact participation.

Objective: This study aims to seek the perspectives of women to understand whether applications of blockchain technology in app-based digital research would affect their decision to participate and contribute sensitive personal health data.

Methods: A convergent, mixed methods, experimental design was used to evaluate participant perceptions and attitudes toward using 2 app-based women's health research study prototypes with blockchain features. Prototype A was based on the status quo ResearchKit framework and had extensive electronic informed consent, while prototype B minimized study onboarding requirements and had no informed consent; the mechanisms of how the contributed data flowed and were made pseudonymous were the same. User evaluations were carried out in February and March 2021 and consisted of a think-aloud protocol, a perception survey, and a semistructured interview. Findings were mapped to the technology acceptance model to guide interpretation.

Results: We recruited 16 representative female participants from 175 respondents. User evaluations revealed that while participants considered prototype B easier to use on intuitive navigation (theme 1) of specified tasks and comprehension (theme 2) of research procedures, prototype A trended toward being perceived more favorably than prototype B across most perception survey constructs, with an overall lower level of privacy concern (mean [SD]: 2.22 [1.10] vs 2.95 [1.29]) and perceived privacy risk (2.92 [1.46] vs 3.64 [1.73]) and higher level of perceived privacy (5.21 [1.26] vs 4.79 [1.47]), trust (5.46 [1.19] vs 4.76 [1.27]), and usability (67.81 [21.77] vs 64.84 [23.69]). Prototype B was perceived more favorably than prototype A with perceived control (4.92 [1.32] vs 4.89 [1.29]) and perceived ownership (5.18 [0.59] vs 5.01 [0.96]). These constructs, except for perceived ownership, were significantly correlated with behavioral intention to use the app (P<.05). Participants perceived the usefulness of these prototypes in relation to the value of research study to women's health field (theme 3), the value of research study to self (theme 4), and the value of blockchain features for participation (theme 5).

XSL•FO
RenderX

Conclusions: This study provides nuanced insights into how blockchain applications in app-based research remain secondary in value to participants' expectations of health research, and hence their intention to participate and contribute data. However, with impending data privacy and security concerns, it remains prudent to understand how to best integrate blockchain technology in digital health research infrastructure.

(JMIR Mhealth Uhealth 2025;13:e65747) doi:10.2196/65747

KEYWORDS

blockchain technology; privacy; trust; data control; data ownership; digital health study; user-centered design; user experience; mHealth; mobile health; women's health; reproductive health; data sharing; research participation; bioethics

Introduction

Blockchain, or distributed ledger technology, has capabilities that can alleviate persistent concerns in digital health research [1-3]. A blockchain is a decentralized database that maintains a growing list of records or transactions using cryptographically linked blocks [1,2]. The blockchain requires consensus across all peers or "nodes" of the network before a block can be added, thus enabling the validation of transactions in a decentralized manner [1,2]. Once recorded on a blockchain, transactions are immutable (cannot be changed or deleted), existing in a verifiable and auditable manner that fosters transparency and trust [1,3]. While blockchain is best known as the underlying technology of cryptocurrencies like Bitcoin, the application of its properties in the health research sector can transform the ways we share, access, and use data [1,4].

For digital health research, vast quantities of health data captured through wearables, mobile health (mHealth) apps, smartphones, and Internet of Things devices coupled with clinical and "omics" data hold valuable insights into individuals' health and behavior. As these large datasets or "big data" grow in scope, they also become more susceptible and vulnerable to privacy risks (eg, data breaches, privacy violations, and unauthorized access) that can cause undue harm to participants, organizations, and society [5]. Health research is also increasingly conducted outside the purview of academic health care institutions, which must comply with stringent regulations (eg, Health Insurance Portability and Accountability Act and human research subject protections) to safeguard patient privacy and ensure ethical research by independent entities unencumbered by these requirements. For example, many marketplace mHealth apps that use user data for health research or algorithm training purposes lack formal informed consent processes or are unclear about how to opt in or opt out of research, further challenging the ethical norms of human subject research [6].

These risks are salient in women's health research, a field that is understudied with large existing health disparity gaps [7] and historic injustices that left lingering mistrust in biomedical interventions [8]. A boom of "femtech" (female- or women-centered technologies) tools are enabling women to regularly report and track data related to menstruation, pregnancy, fertility, pelvic or sexual health, and cancers, among other relevant topics [9-11]. While these data offer immense opportunities to advance women's health, the data are also inherently sensitive, intimate, and taboo in certain sociocultural contexts [9]. Exposure can cause women potential harm such as shame, discrimination, and even violence [9]. Women's willingness to share personal health data may be further dampened by the 2022 US landmark *Dobbs v Jackson Women's Health Organization* case that overturned the privacy precedents established by *Roe v Wade* [12,13]. Privacy concerns and discomfort with data sharing were also reported as reasons for the mHealth app abandonment [14]. Thus, the presence of robust data security safeguards is critical to the outlook of participation, trust, and quality of data generated for health research.

Blockchain can reinforce the security of digital health study infrastructure and offer participants greater protection over their data. Data stored on the blockchain can be analyzed but remain private, as the owner is pseudonymous, and a private key is needed to access their data [1,3]. Participants can grant researchers access to select personal health data by initiating transactions signed by their private key, thereby exercising control and ownership over their data in a digital environment [1-4]. This process would require participants to be active contributors of their data, with each transaction on the blockchain serving as a record of their consent. Marketplace health tools can use this approach to be resilient toward evolving data-sharing regulations; however, it remains to be seen how users perceive its value and impact in driving women's health research.

This study sought the perspectives of women, centering on their perceptions and attitudes, to understand how applications of blockchain technology in app-based digital research affect their decision to participate and contribute data. In collaboration with Bitmark Inc, a domain-agnostic blockchain applications company, we took a "privacy-by-design" [15] approach in creating 2 app-based women's health research study prototypes—prototypes A and B. Prior to the launch of a "live" or in-production app, rigorous user testing on prototypes helps uncover core real-world matters that may affect adoption. We, therefore, carried out a convergent, mixed methods, experimental study of the 2 prototype variants with 16 female participants between February and March 2021. This study aims to show the complexity of human and sociotechnical considerations that arise with blockchain integration into digital health research infrastructure.

Methods

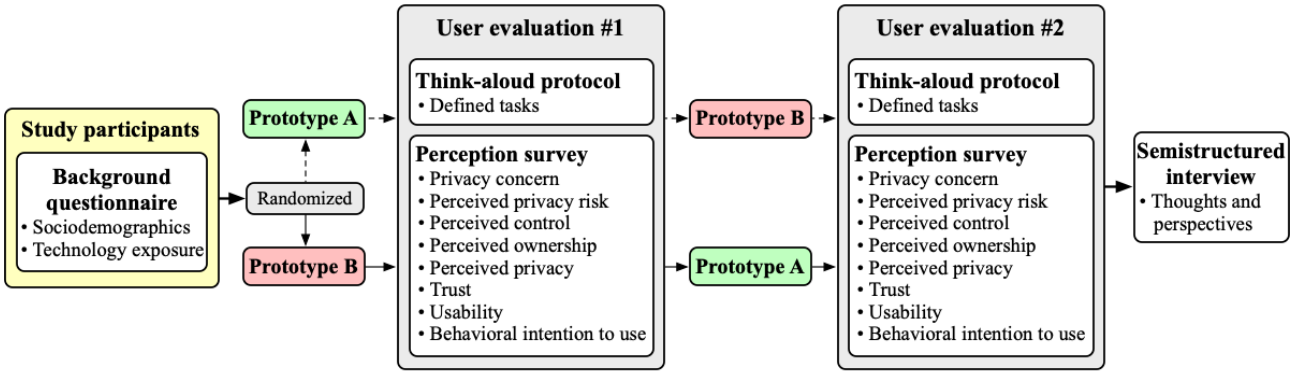
Study Design

A convergent, mixed methods, experimental design [16] was used to ascertain the data for our study (Figure 1). The experimental portion consisted of a crossover trial design where female participants were randomly assigned to either test (1)

XSL•FO
RenderX

prototype A first, followed by prototype B, or (2) prototype B first, followed by prototype A. This served to address any potential spillover effects from testing one prototype first over the other. The mixed methods portion consisted of (1) usability tests for each prototype with a think-aloud protocol (during testing) and perceptions survey (after testing) and (2) a semistructured interview to gauge a more nuanced understanding of participant thoughts and perspectives.

Figure 1. Summary of the convergent, mixed methods, experimental design.



Industry Partnership

A partnership with Bitmark Inc produced 2 unique prototypes guided by real-world blockchain applications for our comprehensive evaluation. The Bitmark system is built on a public blockchain, which is a digitized, decentralized, and public ledger that records all occurring data transfers. Bitmark issues digital property titles, known as "bitmarks," for a person's digital data and records them on the blockchain. In the context of digital health research, this system enables participants to contribute their "bitmarked" data to researchers pseudonymously. Bitmark Inc provided technical and design support for the prototypes.

Prototype Development and Description

Development cycles were carried out between study researchers and Bitmark engineers to design the 2 baseline prototypes: prototype A and prototype B. These prototypes each present a hypothetical digital research infrastructure where blockchain capabilities were differentially operationalized and conveyed to participants. Prototype A proposed a more familiar process of carrying out app-based digital research where an extensive electronic informed consent (eIC) process exists, and participant health data were actively transferred from the participant to the researcher; the ResearchKit framework from Apple served as the basis for the design. Prototype B envisioned a more radical process of app-based research participation where informed consent was not required, and participant health data were stored in a "data vault" for open aggregate analyses by the research community. The features of both prototypes, including their similarities and differences, are listed in Multimedia Appendix 1. The prototypes were produced as a series of high-fidelity wireframes and presented using Figma (Figma Inc), a design and prototyping software tool. Relevant wireframe designs and brief descriptions are provided in Multimedia Appendix 2.

Participant Recruitment and Sample

Women, aged 18 years or older, were recruited through the University of California, Berkeley (UC Berkeley) Experimental Social Science Laboratory (Xlab) for user evaluations. A recruitment email with a background questionnaire, administered through the survey platform Qualtrics (Qualtrics International Inc), was sent to Xlab's participant pool to obtain sociodemographic, smartphone, and health app use information from interested respondents (Multimedia Appendix 3). Email recruitment yielded a total of 175 respondents who completed the questionnaire.

We assessed the eligibility of the 175 respondents and purposively sampled across 4 groups to obtain a representative sample of 16 participants for user evaluations (Multimedia Appendix 4). Our 4 groups consisted of (1) White, non-Hispanic women of higher socioeconomic status (SES), (2) White, non-Hispanic women of lower SES, (3) women of color with higher SES, and (4) women of color with lower SES. Women of color were defined as those respondents who did not identify as White and non-Hispanic (eg, non-Hispanic Black, non-Hispanic Asian, Hispanic, or Latino). SES was defined by either above (higher SES) or below (lower SES) 100% area median income (AMI; adjusted to San Francisco Bay Area standards) or presence (higher SES) or lack of (lower SES) college-level education. We took an intersectional approach in determining the 4 groups, which accounted for differences in the association between certain sociodemographic variables (eg, race or ethnicity, income, education) and the likelihood of participating in health research [17]. Within each of these groups, we also considered the respondents' level of digital health research and technology exposure. A higher level of digital health literacy was defined as those who had either previously joined a digital research study or ever used a women's health app.

Study Setting

All user evaluation sessions were conducted in English and remotely through Zoom (Zoom Video Communications) from February to March 2021. Sessions were video and audio recorded with participant consent. Following our study design (Figure 1), participants were randomized to a testing arm to evaluate the prototypes. We first asked participants to complete 3 tasks in 1 prototype while thinking aloud (think-aloud protocol), followed by completing a perception survey about the prototype; these procedures were repeated for the next

XSL•FO
RenderX

prototype. Participants accessed each prototype through a Figma link and shared their screen in Zoom during the session recording. Participants completed a perception survey for prototypes A and B after each think-aloud protocol. The total mean (SD) time to complete both think-aloud protocols was 39 (10) minutes; recordings were paused during the completion of the perception surveys. After completing user testing for both prototypes, we conducted a semistructured interview with participants to capture their experiences and thoughts. The total mean (SD) time of the interview was 13 (4) minutes. Participants took approximately 60-90 minutes to complete the entire evaluation session. We employed Rev (Rev.com Inc), a reputable commercial human transcription service, to transcribe session audio recordings.

Data Collection Instruments

We describe the data collection instruments used in our mixed methods, experimental design (Figure 1) in detail below. All data collection instruments can be found in Multimedia Appendix 3.

Background Questionnaire

A background questionnaire gathered sociodemographic and general technology exposure information (ie, usage of smartphone, app-based health research study, or women's health-related app) from interested study respondents. Respondents accessed the questionnaire through a Qualtrics link in the recruitment email.

Think-Aloud Protocol

The usability of the prototypes was assessed using a think-aloud protocol with a defined interaction approach. Defined tasks are specified for the participant to complete. Participants were asked to "think-aloud" or verbalize their thought processes while they completed (or tried to complete) each task. Participants were asked to carry out the following three tasks: (1) create an account or register in the app, (2) join a women's health research study, and (3) contribute health data to the study. These tasks were selected because they form the basis of app-based digital health research participation, and their usability can influence early attitudes toward intent to participate.

Perceptions Survey

Our perception survey measured eight constructs: (1) privacy concern, (2) perceived privacy risk, (3) perceived control, (4) perceived ownership, (5) perceived privacy, (6) trust, (7) usability, and (8) behavioral intention to use the app. Measurement items for each construct were gathered and adapted from validated surveys or scales in the literature [18-23]. Items were selected to capture participants' perceptions in relation to their intention to use either prototype. Items are assessed on a 5- or 7-point Likert scale ranging from 1=strongly disagree to 5 or 7=strongly agree, depending on the construct. Summative scales or a calculated score (for usability only) were produced to represent each construct. Participants accessed the perceptions survey through a Qualtrics link after completing the think-aloud protocol for each prototype.

Semistructured Interview

A semistructured interview helped gather more in-depth thoughts or perspectives from participants that may not have been captured through the task-based activities or surveys from the testing session. The interview guide was informed by salient questions surrounding our study topic and was purely exploratory.

Data Analysis

We describe the analysis approach for the data of each instrument.

Background Questionnaire

Descriptive statistics were used to summarize the sociodemographic and general technology exposure characteristics of the women who completed the entire user evaluation of prototypes A and B.

Think-Aloud Protocol

The study lead (MYN) reviewed the cognitive interviews, audiovisual recordings during user testing, and corresponding transcript data to ascertain information about task performance. For each task, it was determined whether the participant (1) completed the task with ease, (2) completed the task with help, or (3) did not complete the task. Task completion was defined as the correct completion of the task and participant acknowledgment. Ease of completion was defined by whether the participant experienced a low versus high level of difficulty or confusion in completing the task; this was determined based on a combination of time elapsed, clicks (trial-and-error), and expressed emotions (familiarity, excitement, frustration, and doubt) until completion acknowledgment. Those who completed the task with ease were generally more confident in stating their completion, while those who completed the task with help were less certain or had lingering concerns. Task incompletion was defined by incorrect completion of the task (eg, incorrect stop points) or an explicit statement that (1) they did not believe they completed the task or (2) they would not have proceeded to complete the task under usual circumstances.

Perceptions Survey

We carried out reliability analyses for each construct's set of measurement items to determine their "internal consistency" or how well they go together [15]. Reliability was determined using Cronbach α; a value of more than 0.70 means high reliability, 0.35 to 0.70 means acceptable reliability, and less than 0.35 means low reliability. Summative scales were created for each construct, with the exception of usability. For the usability construct, we generated a System Usability Scale (SUS) score between 0-100 following the scale's validated process; an SUS score of 68 is considered average. Pearson correlation analyses were performed to determine the magnitude and direction of the linear relationship between all the perception survey constructs and behavioral intention to use the app (outcome variable) for prototypes A and B; P values of <.05 were considered statistically significant. An absolute Pearson correlation coefficient of more than 0.70 means strong to very strong correlation and 0.40 to 0.69 means moderate to good correlation [24]. Due to the small sample size, we used the

XSL•FO
RenderX

Wilcoxon rank-sum (Mann-Whitney *U*) test, a nonparametric test of mean difference, to determine whether there were significant differences between each construct's mean or score in prototype A versus prototype B. The significance level for all tests was set at the .05 level (5%). All statistical analyses were conducted using STATA (version 16.0; StataCorp).

Semistructured Interview

Three coders (MYN, OS, and TT) used a thematic analysis approach to process transcripts in MaxQDA (VERBI Software), a qualitative data analysis software program. First, the study lead (MYN) generated initial codes from the raw semistructured interview data using a mixed deductive-inductive approach. Deductive codes were generated based on broad content areas (eg, perceptions about blockchain, research participation, and health data contribution) that corresponded to the interview questions. Inductive codes emerged from close reading of the interview transcript data. Second, the 3 coders (MYN, OS, and TT) used an initial codebook to independently code a subset of transcripts. The coding team discussed discrepancies until a consensus was reached. The initial codebook was iteratively refined throughout this process. The study lead (MYN) reapplied the updated coding frame to all coded transcripts to attain alignment. Finally, identified emergent concepts and connections were examined to identify themes relevant to the use of an app-based women's health research study. Key themes were mapped to the constructs of the technology acceptance model (TAM; ie, perceived ease of use and perceived usefulness) to explore how they relate to participants' attitudes and behavioral intention to use an app-based research study with integrated blockchain features [25,26].

Ethical Considerations

Human subject ethical review approval for this study and all procedures were obtained from the UC Berkeley Institutional Review Board (2021-02-14021). All participants provided verbal consent upon reviewing the study information and before taking part in study procedures. Participant data are safeguarded and stored on encrypted devices with minimal identifiers. Respondents to the background questionnaire were entered into a US $15 raffle. Those who participated in the full evaluation session were paid US $30 to US $35 for their time. We followed guidelines from the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) for the perception survey and the Consolidated Criteria for Reporting Qualitative Studies (COREQ) [27,28] for the semistructured interview.

Results

Participant Characteristics

A total of 175 respondents completed the background questionnaire in the recruitment email. Using purposive sampling, we reached out to respondents until we attained a balanced representation of 4 participants each for the aforementioned 4 groups (Multimedia Appendix 4). These 16 women participated in the user evaluation of prototypes A and B, which consisted of a think-aloud protocol (task performance and cognitive interview) and perception survey for each prototype, followed by a concluding semistructured interview.

Participant characteristics are displayed in Table 1. For race or ethnicity, of the 16 women, 8 (50%) identified as non-Hispanic White and 8 (50%) identified with other racial and ethnic groups. Participants' average age was 29 years, with 8 (50%) women in the 18-29 years age group, 6 (38%) in the 30-39 years age group, and 2 (13%) in the 40-49 years age group. We did not have any respondents in the 50-59 years or 60 years or older age groups. For education level, 6 (38%) women did not have a college-level degree while 10 (63%) had a college-level degree or above. For AMI, 9 (56%) women were below 100% AMI, while 7 (44%) were above 100% AMI. In terms of having previous exposure to the study topic, 8 (50%) women have either ever used a women's health app or ever joined a digital research study, while 8 (50%) have done neither. Most participants (15/16, 94%) have owned a smartphone for over 5 years and indicated they mostly (14/16, 88%) use their smartphones for less than 5 hours each day.

XSL•FO
RenderX

Table 1. Characteristics of women participants who evaluated prototype A and prototype B (N=16).

Characteristic	Participants, n (%)
Race or ethnicity	
Asian, non-Hispanic	5 (31)
Black, non-Hispanic	1 (6)
Hispanic or Latino	2 (13)
White, non-Hispanic	8 (50)
Age group (years)	
18-29	8 (50)
30-39	6 (38)
40-49	2 (13)
Education	
Some college, no degree	4 (25)
Associate's degree	2 (13)
Bachelor's degree	6 (38)
Master's degree	2 (13)
Doctoral or professional degree	2 (13)
AMI[a] (%)	
Less than 50	4 (25)
50-99	5 (31)
100-149	2 (13)
150-200	1 (6)
More than 200	4 (25)
Familiarity with study topic	
Ever use a women's health app	7 (44)
Ever join a digital research study	1 (6)
Neither	8 (50)
Smartphone ownership (years)	
3-5	1 (6)
More than 5	15 (94)
Smartphone usage per day (hours)	
1-2	6 (38)
3-4	8 (50)
More than 5	2 (13)

[a]AMI: area median income.

Think-Aloud Protocol

The 16 participants' task performance of the 3 tasks for each prototype is summarized in Table 2. We identified a key theme from the participants' cognitive interviews relevant to the perceived ease of use construct.

XSL•FO
RenderX

Table 2. Task performance identified through usability testing (N=16).

Task performance	Prototype A, n (%)			Prototype B, n (%)		
	Task 1[a]	Task 2[b]	Task 3[c]	Task 1[a]	Task 2[b]	Task 3[c]
Completed						
With ease	5 (31)	0 (0)	16 (100)	4 (25)	11 (69)	15 (94)
With help	4 (25)	16 (100)	0 (100)	10 (63)	5 (31)	0 (0)
Not completed	7 (44)	0 (0)	0 (0)	2 (13)	0 (0)	1 (6)

[a]Task 1: Create an account or register in the app.

[b]Task 2: Join a women's health research study.

[c]Task 3: Contribute health data to the study.

Theme 1: Intuitive Navigation

Overall, the task performance and cognitive interviews provided important context about intuitive navigation in facilitating participants' perceived ease of use with the health research study app. Intuitive navigation was assessed in our study based on whether participants were able to (1) definitively determine whether an account was created (task 1), (2) locate the research study within the app in a straightforward manner (task 2), and (3) easily contribute health data to the research study (task 3).

Task 1 was easier to complete in prototype B than in prototype A. Although more participants completed task 1 overall in prototype B (14/16, 88%) than in prototype A (9/16, 56%), many acknowledged completion with a level of help. The findings from task 1 suggest that the concept of a pseudonymous account is still quite foreign to users and the prototypes can better expound on its significance.

Task 2 was easier to complete in prototype B than in prototype A. Although all participants completed task 2 in both prototypes, more completed the task with help in prototype A (16/16, 100%) than in prototype B (5/16, 31%). The findings from task 2 stress the importance of a clear and intuitive interface to effectively direct participants from account creation to the research study.

Task 3 was largely completed with ease by participants in both prototypes, with 16 (100%) in prototype A and 15 (94%) in prototype B. The findings from task 3 show that participants are familiar with the concept of data contribution to research in a remote setting.

Perceptions Survey

The quantitative assessments of the 16 participants' perceptions toward each prototype are presented in Table 3. Reliability analyses revealed that each construct's set of measurement items had Cronbach α values over 0.70, which indicates high reliability.

Table 3. Mean values and intercorrelations between perception survey constructs and behavioral intention to use the app (N=16).

Perception constructs	Prototype A			Prototype B		
	Cronbach α	Mean (SD)	*r*	Cronbach α	Mean (SD)	*r*
Behavioral intention to use app[a]	0.79	3.52 (0.68)	1.00	0.89	3.33 (1.04)	1.00
Privacy concern[a]	0.96	2.22 (1.10)	–0.61[d]	0.96	2.95 (1.29)	–0.59[d]
Perceived privacy risk[b]	0.90	2.92 (1.46)	–0.61[d]	0.96	3.64 (1.73)	–0.52[d]
Perceived control[b]	0.85	4.89 (1.29)	0.73[d]	0.88	4.92 (1.32)	0.61[d]
Perceived ownership[b]	0.88	5.01 (0.96)	0.26	0.77	5.18 (0.59)	0.17
Perceived privacy[b]	0.97	5.21 (1.26)	0.65[d]	0.96	4.79 (1.47)	0.66[d]
Trust[b]	0.95	5.46 (1.19)	0.62[d]	0.96	4.76 (1.27)	0.68[d]
Usability[c]	0.90	67.81 (21.77)	0.45	0.95	64.84 (23.69)	0.87[d]

[a]Assessed on a 5-point Likert scale ranging from 1="strongly disagree" to 5="strongly agree."

[b]Assessed on a 7-point Likert scale ranging from 1="strongly disagree" to 7="strongly agree."

[c]Calculated System Usability Scale score between 0-100.

[d]Denotes *P*<.05 and statistical significance in correlation with behavioral intention to use the app.

Pearson correlation analyses of the resultant summative scales and scores revealed insights into the 8 constructs, with a focus on the others' relationship to behavioral intention to use the app (outcome variable). For both prototypes A and B, participants' level of privacy concern (prototype A: *r*=–0.61; *P*=.01 and prototype B: *r*=–0.59; *P*=.02) and perceived privacy risk (prototype A: *r*=–0.61; *P*=.01 and prototype B: *r*=–0.52; *P*=.04) were shown to have a good negative correlation with the behavioral intention to use the app. In contrast, participants' level of perceived control (prototype A: *r*=0.73; *P*=.001 and

XSL•FO
RenderX

prototype B: r=0.61; P=.01), perceived privacy (prototype A: r=0.65; P=.01 and prototype B: r=0.66; P=.01), and trust (prototype A: r=0.62; P=.01 and prototype B: r=0.68; P=.004) with both prototypes were shown to have a good positive correlation with the behavioral intention to use the app. Usability was shown to only have a strong positive correlation with the behavioral intention to use the app in prototype B (r=0.87; P<.001). Perceived ownership of one's health data in either prototype was not shown to have a significant correlation with the behavioral intention to use the app.

Overall, across all the perception constructs, a comparison of their mean scales or scores between prototype A and prototype B revealed no significant differences. However, prototype A trended toward being perceived more favorably than prototype B across most perception survey constructs, with an overall lower level of mean privacy concern (2.22, SD 1.10 vs 2.95, SD 1.29) and mean perceived privacy risk (2.92, SD 1.46 vs 3.64, SD 1.73) and a higher level of mean perceived privacy (5.21, SD 1.26 vs 4.79, SD 1.47), mean trust (5.46, SD 1.19 vs 4.76, SD 1.27), and mean usability (67.81, SD 21.77 vs 64.84, SD 23.69). Prototype B was perceived more favorably than prototype A with mean perceived control (4.92, SD 1.32 vs 4.89, SD 1.29) and mean perceived ownership (5.18, SD 0.59 vs 5.01, SD 0.96).

Semistructured Interview

From 16 participants' concluding interviews, we identified 4 more key themes relevant to perceived ease of use and perceived usefulness that drive participants' overall attitude toward consistent use of the apps. All themes informed prototype improvement recommendations in Multimedia Appendix 5.

Theme 2: Comprehension

Participant comprehension of the research study and other technical safeguards (eg, blockchain applications) within the app facilitated their perceived ease of use. The presence of a formal eIC process in prototype A was helpful in informing and reassuring participants about the research study, hence fostering trust in health data contribution.

> *Prototype A just seemed more trustworthy design-wise and also because it really emphasized consenting… it gave more information on how the data was used.* [User 3]

However, the absence of an eIC process or other extensive explanatory content about blockchain technology (in prototype B) did not completely hinder one's decision to participate and share health data.

> *As far as what information was being shared, and how secure I feel about it, I thought both apps did a good job… Prototype B was weaker on explaining security protocols, but it didn't leave me feeling uncertain about the app or unwilling to share [data].* [User 1]

> *I'd probably contribute the same. Just because I trust the overall aim of what the study is trying to go for. I didn't feel there was a difference in security. I just think the description of one was more convoluted than the other.*" [User 5]

Theme 3: Value of Research Study to Women's Health Field

The perceived usefulness of the research study app is discussed in the following 3 themes. Participants were generally comfortable contributing to a majority of the requested women's health data categories, including those typically considered to be more stigmatized. Most apprehensive thoughts centered around categories that required added acquisition costs (eg, ovulation test results, basal body temperature), additional clarification (eg, cervical mucus quality), or sensitivity tolerance (eg, sexual activity) to contribute data.

> *I'm just not a person that is too uncomfortable with a lot of these things because I just consider them all to be pretty human experiences that… a lot of women experience.* [User 13]

Overall, participants' perceived value of their data contributions to the advancement of women's health in research and medicine outweighed any discomforts or inconveniences with continued participation.

> *I feel like this information, given to the right institution or to the right study, has the potential to be instrumental and helpful for further generations of women for medical knowledge… I would love to see our medical knowledge expand in this area to be able to help more women, and for medicine… to be more equal.* [User 7]

Theme 4: Value of Research Study to Self

Participants expressed their continued participation depended on the extent of direct or indirect benefits to themselves, including positive feelings associated with seeing how their contributed data advanced study objectives and goals, their participation was appreciated and valued, and the presence of financial compensation. Regular updates about study progress and research findings were considered helpful for participants to assess whether continued participation was worth their time.

> *It would be nice to see how [my] information is being used... Not just knowing that I'm giving my information away with no real work to show for it. It's not… set it to forget it, but you set it, and you're not seeing results. It's like, where's my information going? What are they really using it for? So to see some updates of how that information is being used makes me want to keep using it. Makes me feel like... I'm contributing to these new findings... So [I would like to] see that update of: "Hey, with the recent health data being donated, we've realized that women from the ages 21 to 25 have an uptick in this, or there's this pattern of this…"* [User 14]

A more granular understanding of how the research study is relevant to them and the return of insights about their contributed data is also a driver for their continued participation.

> *There's also always: 'What am I getting out of it?' I've been more interested in signing up for studies*

XSL•FO
RenderX

> *that do compile your information and give it to you at the end, or tell you something about what they've learned about using your data.* [User 8]

Both prototypes did not intend to financially compensate participants for their time or data, aligning with the design of other app-based research studies. However, financial incentives were considered compelling for continued participation.

> *If it's 10 minutes every day for 3 months, plus sharing this data and being thoughtful about my feedback, I'd want to know what the financial compensation is, if any… To participate in the study, I'd want it to be worth my time.* [User 2]

Theme 5: Value of Blockchain Features for Participation

Participants generally believed blockchain to be trustworthy and can enhance the privacy and security of contributed health data, but given their limited understanding of the technology, they could not definitively say whether its application in the study would affect their decision to participate.

> *On one hand, I don't know if I'd want to really do a long-term study in general, but I guess knowing that the specific type of encryption technology, from my vague knowledge, I've heard is a relatively new and relatively well-respected technology, might make me more interested in contributing more data.* [User 6]

> *My doctor's office… [has] an online portal where they keep all of my data and super intimate health information… I'm sure there's special encryption that they use, but I don't know the specifics of it. I just trust them. So I don't know if necessarily having this blockchain technology is making me more likely [to participate]... but that's probably because I don't have a background of knowing anything about blockchain technology.* [User 8]

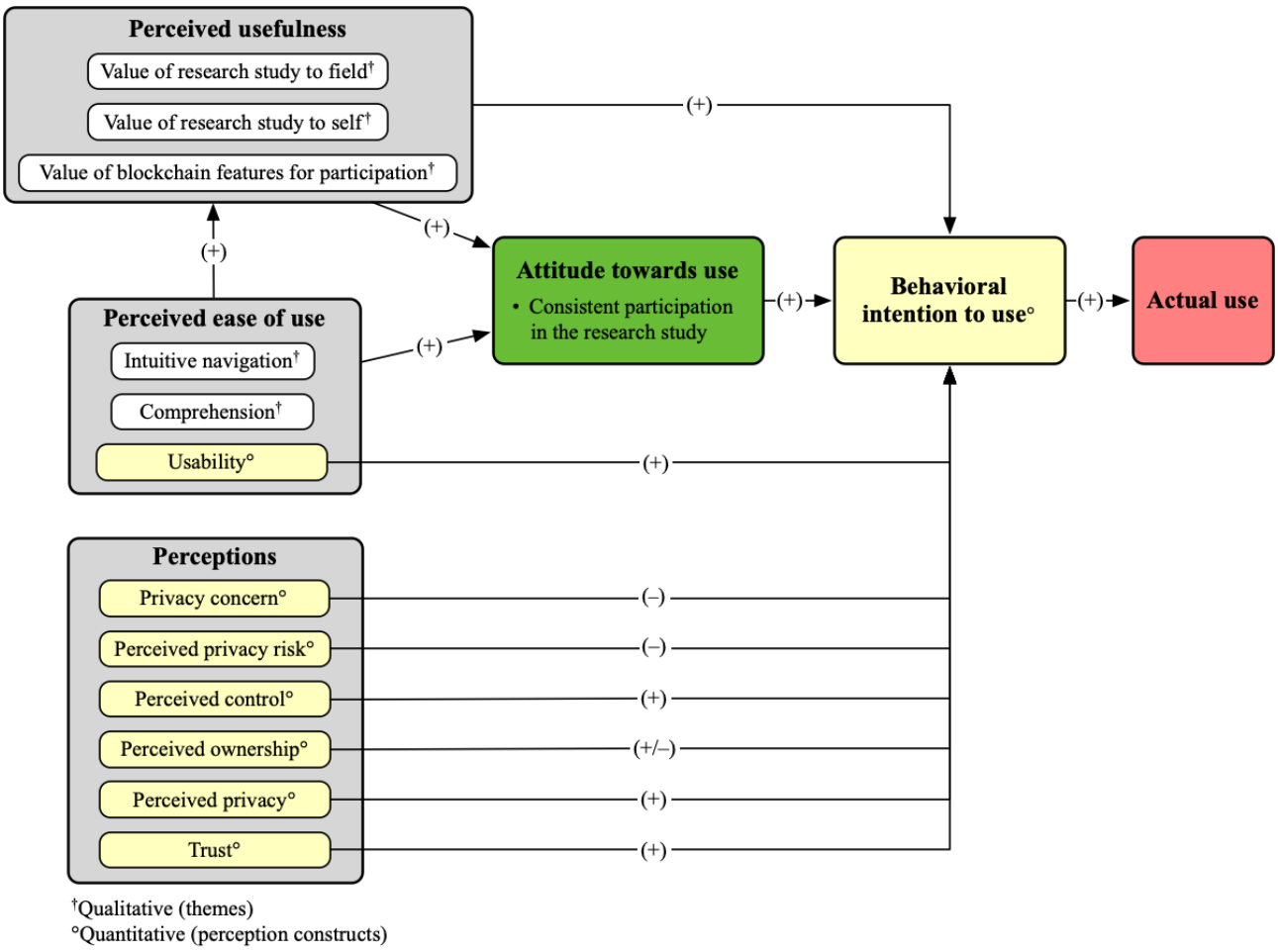
Also, given the novelty of blockchain technology in health care and its limited application, one participant expressed that it would take time for trust to be established and drive decision-making.

> *Typically, a new technology wouldn't allure me until it had become more established, and maybe more generally accepted and widely used. I would be wary of it, just being the first time I've ever heard of it or used it. I don't know that I would trust it right away.* [User 1]

Integrated Conceptual Framework

We mapped qualitative and quantitative findings from the user evaluations to TAM constructs. Figure 2 provides an overview of the integrated conceptual framework.

Figure 2. Integrated conceptual framework with qualitative and quantitative findings mapped to technology acceptance model constructs.



XSL•FO
RenderX

Discussion

Principal Findings

This study set out to understand potential women participants' perceptions and attitudes toward the use of blockchain-enhanced women's health research study apps. The prototypes proved to be effective representations for participants to ground their assessments. User evaluations reveal the limited influence of blockchain technology in contrast to a thoughtful research operations pipeline while fostering trust in digital health research participation. Our integrated conceptual framework (Figure 2) shows the nuanced factors that affect participation in this context.

Our study revealed 4 key findings on how digital health research teams and industry partners may approach the integration and use of emerging technology. First, while optimizing user interface design can directly improve the intuitive navigation of a prototype, it does not sufficiently address participant comprehension, which ultimately affects perceived usefulness. Recommendations in Multimedia Appendix 5 involve reworking and reorganizing the app content to better communicate key concepts about the study's purpose and significance. These improvements reiterate the importance of a joint focus on optimizing technical design layouts alongside content usability, which aims to improve users' uptake and processing of key information [29]. While this can enhance digital literacy in app-based research, the addition of blockchain features complicates these efforts [30-32]. User-centered design and user testing the comprehension of complex information are essential to uncover such challenges sooner.

Participants also revealed how differing privacy-by-design approaches can affect research participation. Even though prototype A had extensive explanatory content, participants' perceptions were collectively better compared to prototype B, which lacked a formal informed consent process and emphasis on blockchain capabilities. Prototype A notably trended toward a higher level of perceived privacy and trust and a lower level of privacy concern and perceived privacy risk compared to prototype B. Although we were unable to detect statistically significant differences between means across all constructs of both prototypes, we were able to show the practical significance of these constructs (that most are highly correlative) on participants' behavioral intention to use the app.

Second, we found that what was most valuable to women participants in app-based research was consistent with qualms already expressed in conventional research. Wilkins et al [33] found through a national survey that research participants preferentially valued receiving research results and updates on how their data were used, which is also associated with a higher likelihood of trusting researchers. Perceived usefulness, particularly the value of the research to women's health and oneself was an important driver of participation in the proposed research apps. Participants also periodically made appraisals of that value (eg, progress, outcomes, impact) to determine their continued participation. This demonstrates that one's decision to participate in research is a dynamic process that requires constant cultivation, especially if a study calls for long-term participation.

To meet these fundamental participant needs and expectations, design, and research teams need to prioritize efforts in building channels of communication with participants for the "return of value" and information [33]. Rather than acquiesce to the usual unidirectional data flow from participants to researchers, the return of study findings or data analytics to participants can make great strides in promoting their prolonged engagement and participation. There are also ethical benefits of providing data back to participants from the standpoint of respecting their autonomy and fairness of research benefits. If this option is not feasible, providing compensation for participants' time and contributions is another way to motivate their participation. Otherwise, the study runs the risk of losing consistent contributions of quality data for women's health research and wasting already expended efforts and resources.

Third, data privacy and security concerns were not at the forefront of women participants' minds. Given the stigma in reproductive and sexual health, we had initial concerns about the sensitivity and extent of data elements requested from participants [9,34]. The list of data elements was originally drawn from Apple Health's "Reproductive Health" category, which has since been rebranded to "Cycle Tracking" with a considerably longer list of data elements. Surprisingly, participants were open and willing, without an overly expressed sense of stigma, to share these data with researchers. While this may be partially attributed to the qualities of our recruited population, such as their interest and willingness to participate in research, it may also point to a larger cultural or generational shift toward women's health issues. An important caveat is that these attitudes may have evolved considering the post-*Dobbs* era and threats to the privacy and safety of women [12,13].

This willingness to contribute data is predicated on an inherent trust in systems, particularly with affiliated academic institutions and the protections they have in place. Despite potentially having control over data sharing and establishing ownership over one's data via blockchain capabilities, participants were generally satisfied with fewer technical safeguards, such as statements about privacy and security protections. Although participants may be content with deferring the safeguard of their data and interests to the study and its partnering entities, this does not absolve research stakeholders from further reinforcing their security protections. Participants' trust in academic institutions is fragile and research stakeholders have a responsibility to minimize threats to its credibility. Trust, once lost, is very hard to rebuild. Research stakeholders need to remain vigilant in upholding strong data privacy and security practices regardless of participants' changing attitudes toward data contribution.

Finally, even though our evaluation suggests a wavering level of perceived usefulness and comprehension with blockchain technology among women participants, it does not undercut the technology's actual capabilities and value to society. While blockchain-enhanced features in the prototypes did not impact participant perceptions and attitudes toward use as much as we had expected, the technology encompasses societal values (eg,

XSL•FO
RenderX

privacy, democratization, transparency, trustworthiness, reliability) that will endure in their relevance to digital health research [35]. Nonetheless, the technical benefits of blockchain will not be sufficient in affecting digital health research without detailed work on comprehension, explainability, and sustainable implementation.

Limitations

The study had several limitations. First, we recruited women from the Xlab human subject pool, which consisted of students, staff, and alumni from UC Berkeley. This exposed our study to potential selection bias, where participants may not be representative of the general population. The sociodemographic characteristics of the UC Berkeley community, and the San Francisco Bay Area at large, are different (eg, higher household income and educational attainment) compared to the rest of the United States [36-38]. Also, those who volunteer to participate in digital health research have been found to have different characteristics compared to the general population [39]. To mitigate bias, we used purposive sampling to create a more representative sample across sociodemographic characteristics and levels of study topic literacy or familiarity.

Second, our study was partially limited by its small sample size of 16 participants, which reduced the statistical power of our perception survey analyses. Although we were unable to find a statistical effect between construct means of prototypes A and B, the sample size yielded insights into their practical significance. A sample size of 16 participants, however, is sufficient for usability testing, where approximately 5 users are considered adequate to reveal major heuristic violations. For the interviews, we reached a level of thematic saturation with the given sample size.

Finally, the prototypes were not technically functional, which limited participants' assessment of their full capability. We also did not fully design and operate prototype features that were deemed not critical to the research study workflow. While we cannot determine the actual use of a live research study app, our use of mixed methods helped triangulate and identify positive features and processes that can enhance research participation.

Conclusions

A privacy-by-design approach in digital health research proves to be promising for research participation. This study revealed that blockchain-enhanced capabilities in health research apps were ultimately secondary in value to participant fundamental expectations (eg, communication with researchers and return of findings) with health research. Hence, the use of novel technological solutions may not be sufficient to significantly affect participation. However, with growing data privacy and security concerns, it remains prudent for organizations to explore effective and sustainable integration of blockchain capabilities.

Acknowledgments

The authors acknowledge Bitmark Inc (MN, CA, AB, SMP, and others) and the Experimental Social Science Laboratory (Xlab) at the University of California, Berkeley for their support throughout this study.

Authors' Contributions

MYN, CRL, and CC conceptualized the study. CC, CRL, and JH provided analytical guidance and oversight. MYN, OS, and TT carried out the analyses. MYN wrote the initial draft, and all authors edited and critically revised the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Design features of prototype A vs prototype B.
[PDF File (Adobe PDF File), 41 KB - mhealth_v13i1e65747_app1.pdf]

Multimedia Appendix 2
Screenshots of prototype A vs prototype B (selection).
[PDF File (Adobe PDF File), 387 KB - mhealth_v13i1e65747_app2.pdf]

Multimedia Appendix 3
Data collection instruments.
[PDF File (Adobe PDF File), 75 KB - mhealth_v13i1e65747_app3.pdf]

Multimedia Appendix 4
Participant user IDs and grouping details (N=16).
[PDF File (Adobe PDF File), 31 KB - mhealth_v13i1e65747_app4.pdf]

Multimedia Appendix 5

XSL•FO
RenderX

Overview of themes organized to technology acceptance model constructs and recommendations for prototype improvement. [PDF File (Adobe PDF File), 28 KB - mhealth_v13i1e65747_app5.pdf]

References

1. Elangovan D, Long CS, Bakrin FS, Tan CS, Goh KW, Yeoh SF, et al. The use of blockchain technology in the health care sector: systematic review. JMIR Med Inform 2022;10(1):e17278 [FREE Full text] [doi: 10.2196/17278] [Medline: 35049516]
2. Lee S, Kim J, Kwon Y, Kim T, Cho S. Privacy preservation in patient information exchange systems based on blockchain: system design study. J Med Internet Res 2022;24(3):e29108 [FREE Full text] [doi: 10.2196/29108] [Medline: 35315778]
3. Sengupta A, Subramanian H. User control of personal mHealth data using a mobile blockchain app: design science perspective. JMIR Mhealth Uhealth 2022;10(1):e32104 [FREE Full text] [doi: 10.2196/32104] [Medline: 35049504]
4. Xie Y, Zhang J, Wang H, Liu P, Liu S, Huo T, et al. Applications of blockchain in the medical field: narrative review. J Med Internet Res 2021;23(10):e28613 [FREE Full text] [doi: 10.2196/28613] [Medline: 34533470]
5. Schroeder T, Haug M, Gewald H. Data privacy concerns using mHealth apps and smart speakers: comparative interview study among mature adults. JMIR Form Res 2022;6(6):e28025 [FREE Full text] [doi: 10.2196/28025] [Medline: 35699993]
6. Farag N, Noë A, Patrinos D, Zawati MH. Mapping the apps: ethical and legal issues with crowdsourced smartphone data using mHealth applications. Asian Bioeth Rev 2024;16(3):437-470. [doi: 10.1007/s41649-024-00296-3] [Medline: 39022376]
7. National Academies of Sciences, Engineering, and Medicine. Overview of Research Gaps for Selected Conditions in Women's Health Research at the National Institutes of Health: Proceedings of a Workshop—in Brief. Washington, DC: National Academies Press; 2024.
8. Ho IK, Sheldon TA, Botelho E. Medical mistrust among women with intersecting marginalized identities: a scoping review. Ethn Health 2022;27(8):1733-1751. [doi: 10.1080/13557858.2021.1990220] [Medline: 34647832]
9. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022;10(5):e33735 [FREE Full text] [doi: 10.2196/33735] [Medline: 35522465]
10. Karim JL, Wan R, Tabet RS, Chiu DS, Talhouk A. Person-generated health data in women's health: scoping review. J Med Internet Res 2024;26:e53327 [FREE Full text] [doi: 10.2196/53327] [Medline: 38754098]
11. Krishnamurti T, Birru Talabi M, Callegari LS, Kazmerski TM, Borrero S. A framework for femtech: guiding principles for developing digital reproductive health tools in the United States. J Med Internet Res 2022;24(4):e36338 [FREE Full text] [doi: 10.2196/36338] [Medline: 35482371]
12. Swanson K, Ravi A, Saleh S, Weia B, Pleasants E, Arvisais-Anhalt S. Effect of recent abortion legislation on Twitter user engagement, sentiment, and expressions of trust in clinicians and privacy of health information: content analysis. J Med Internet Res 2023;25:e46655 [FREE Full text] [doi: 10.2196/46655] [Medline: 37171873]
13. Hendricks-Sturrup R, Lu CY. An assessment of perspectives and concerns among research participants of childbearing age regarding the health-relatedness of data, online data privacy, and donating data to researchers: survey study. J Med Internet Res 2023;25:e41937 [FREE Full text] [doi: 10.2196/41937] [Medline: 36897637]
14. Kidman PG, Curtis RG, Watson A, Maher CA. When and why adults abandon lifestyle behavior and mental health mobile apps: scoping review. J Med Internet Res 2024;26:e56897 [FREE Full text] [doi: 10.2196/56897] [Medline: 39693620]
15. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. J Med Internet Res 2017;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
16. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research, 3rd Edition. Thousand Oaks, CA: Sage Publications; 2018.
17. Abrams JA, Tabaac A, Jung S, Else-Quest NM. Considerations for employing intersectionality in qualitative health research. Soc Sci Med 2020;258:113138 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113138] [Medline: 32574889]
18. Liu C, Marchewka JT, Lu J, Yu CS. Beyond concern: a privacy–trust–behavioral intention model of electronic commerce. Inf Manag 2004;42(1):127-142. [doi: 10.1016/j.im.2004.01.002]
19. Xu H, Dinev T, Smith J, Hart P. Information privacy concerns: linking individual perceptions with institutional privacy assurances. J Assoc Inf Syst 2011;12(12):798-824. [doi: 10.17705/1jais.00281]
20. Dinev T, Xu H, Smith JH, Hart P. Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts. Eur J Inf Syst 2017;22(3):295-316. [doi: 10.1057/ejis.2012.23]
21. Malhotra NK, Kim SS, Agarwal J. Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model. Inf Syst Res 2004;15(4):336-355. [doi: 10.1287/isre.1040.0032]
22. van Dyne L, Pierce JL. Psychological ownership and feelings of possession: three field studies predicting employee attitudes and organizational citizenship behavior. J Organ Behavior 2004;25(4):439-459. [doi: 10.1002/job.249]
23. Brooke J. SUS: A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London, UK: Taylor & Francis; 1996.
24. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018;126(5):1763-1768. [doi: 10.1213/ANE.0000000000002864] [Medline: 29481436]

XSL•FO
RenderX

25. Koul S, Eydgahi A. A systematic review of technology adoption frameworks and their applications. J Technol Manag Innov 2017;12(4):106-113. [doi: 10.4067/s0718-27242017000400011]
26. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319. [doi: 10.2307/249008]
27. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
28. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
29. Ginossar T, Shah SFA, West AJ, Bentley JM, Caburnay CA, Kreuter MW, et al. Content, usability, and utilization of plain language in breast cancer mobile phone apps: a systematic analysis. JMIR Mhealth Uhealth 2017;5(3):e20 [FREE Full text] [doi: 10.2196/mhealth.7073] [Medline: 28288954]
30. Kemp E, Trigg J, Beatty L, Christensen C, Dhillon HM, Maeder A, et al. Health literacy, digital health literacy and the implementation of digital health technologies in cancer care: the need for a strategic approach. Health Promot J Austr 2021;32:104-114. [doi: 10.1002/hpja.387] [Medline: 32681656]
31. Nouri SS, Avila-Garcia P, Cemballi AG, Sarkar U, Aguilera A, Lyles CR. Assessing mobile phone digital literacy and engagement in user-centered design in a diverse, safety-net population: mixed methods study. JMIR Mhealth Uhealth 2019;7(8):e14250 [FREE Full text] [doi: 10.2196/14250] [Medline: 31469083]
32. Blacklow SO, Lisker S, Ng MY, Sarkar U, Lyles C. Usability, inclusivity, and content evaluation of COVID-19 contact tracing apps in the United States. J Am Med Inform Assoc 2021;28(9):1982-1989 [FREE Full text] [doi: 10.1093/jamia/ocab093] [Medline: 34022053]
33. Wilkins CH, Mapes BM, Jerome RN, Villalta-Gil V, Pulley JM, Harris PA. Understanding what information is valued by research participants, and why. Health Aff 2019;38(3):399-407 [FREE Full text] [doi: 10.1377/hlthaff.2018.05046] [Medline: 30830824]
34. Worthington AK, Burke EE, Shirazi TN, Leahy C. US women's perceptions and acceptance of new reproductive health technologies. Women's Health Rep 2020;1(1):402-412 [FREE Full text] [doi: 10.1089/whr.2020.0063] [Medline: 33786505]
35. Nguyen LTQ, Hoang TG, Do LH, Ngo XT, Nguyen PH, Nguyen GD, et al. The role of blockchain technology-based social crowdfunding in advancing social value creation. Technol Forecast Soc Change 2021;170:120898. [doi: 10.1016/j.techfore.2021.120898]
36. McNee G, Pojani D. NIMBYism as a barrier to housing and social mix in San Francisco. J Hous Built Environ 2022;37(1):553-573 [FREE Full text] [doi: 10.1007/s10901-021-09857-6] [Medline: 34054378]
37. San Francisco County, California—Census Bureau Profile. United States Census Bureau. URL: https://data.census.gov/profile/San_Francisco_County,_California?g=050XX00US06075 [accessed 2025-01-10]
38. Alameda County, California—Census Bureau Profile. United States Census Bureau. URL: https://data.census.gov/profile/Alameda_County,_California?g=050XX00US06001 [accessed 2025-01-10]
39. Ng MY, Olgin JE, Marcus GM, Lyles CR, Pletcher MJ. Email-based recruitment into the Health eHeart Study: cohort analysis of invited eligible patients. J Med Internet Res 2023;25:e51238. [doi: 10.2196/51238] [Medline: 38133910]

Abbreviations

AMI: area median income
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COREQ: Consolidated Criteria for Reporting Qualitative Studies
eIC: electronic informed consent
mHealth: mobile health
SES: socioeconomic status
SUS: System Usability Scale
TAM: technology acceptance model
UC Berkeley: University of California, Berkeley
Xlab: Experimental Social Science Laboratory

XSL•FO
RenderX

Edited by L Buis; submitted 26.08.24; peer-reviewed by H Subramanian, TBDS Costa, M Gaynor; comments to author 18.12.24; revised version received 05.02.25; accepted 18.02.25; published 25.03.25.

Please cite as:
Ng MY, Halpern J, Shane O, Teng T, Nguy n M, Alt CR, Leite AB, Moss-Pultz S, Lyles CR, Cheshire C
Participant Evaluation of Blockchain-Enhanced Women's Health Research Apps: Mixed Methods Experimental Study
JMIR Mhealth Uhealth 2025;13:e65747
URL: https://mhealth.jmir.org/2025/1/e65747
doi:10.2196/65747
PMID:40131317

©Madelena Y Ng, Jodi Halpern, Olivia Shane, Tina Teng, Michael Nguy n, Casey Ryan Alt, Anaïs Barthe Leite, Sean Moss-Pultz, Courtney R Lyles, Coye Cheshire. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Investigating the Quality of Mobile Apps for Drug-Drug Interaction Management Using the Mobile App Rating Scale and K-Means Clustering: Systematic Search of App Stores

Ayush Bhattacharya*, MS; Jose Fernando Florez-Arango*, PhD, MD

Department of Population Health Sciences, Weill Cornell Medicine, 425 E 61st, Room 323, New York, NY, United States

*all authors contributed equally

Corresponding Author:
Jose Fernando Florez-Arango, PhD, MD
Department of Population Health Sciences, Weill Cornell Medicine, 425 E 61st, Room 323, New York, NY, United States

Abstract

Background: Drug-drug interactions (DDIs) pose a significant risk to patient safety and increase health care costs. Mobile apps offer potential solutions for managing DDIs, yet their quality and effectiveness from the user's perspective remain unclear.

Objective: The aim is to evaluate the quality of publicly available mobile apps for DDI management in the US using the Mobile App Rating Scale (MARS) and to identify patterns that reflect user satisfaction and preferences.

Methods: A structured review was conducted to identify mobile apps for DDI management, resulting in 19 eligible apps. Two health care–affiliated evaluators independently assessed each app using the mobile app rating scale (MARS). Dimensionality scores were calculated, and correlation analysis was conducted to examine relationships among dimensions. K-means clustering was applied to group apps based on their MARS scores. Scatter plots visualized app distributions across clusters. To validate the clustering model and assess alignment with user satisfaction, mean weighted user ratings were compared with mean MARS scores per cluster. Correlation analysis was also performed between individual MARS dimensions and user ratings within each cluster.

Results: The mean MARS score was 3.54 out of 5, with the Information dimension scoring the highest (mean 3.68, SD 0.51) and Engagement the lowest (mean 3.42, SD 0.80). The Kruskal-Wallis test revealed no significant differences in median scores across the four dimensions (χ^2_3=2.109, *P*=.55). All MARS dimensions were positively correlated (*r*=0.65 to 0.92), indicating interrelated quality characteristics. K-means clustering identified three app groups with varying quality profiles: Cluster 1 (n=7, mean MARS=2.86), Cluster 2 (n=7, mean=3.57), and Cluster 3 (n=5, mean=4.44). Cluster 1 apps showed strongest correlations between user satisfaction and functionality (*r*=0.74) and engagement (*r*=0.53). Cluster 2 users prioritized information (*r*=0.41) and aesthetics (*r*=0.58), and Cluster 3 exhibited balanced influence from information (*r*=0.62), aesthetics (*r*=0.58), and functionality (*r*=0.39). Scatter plots indicated that engagement, functionality, and aesthetics were key drivers of user perception, while information, though consistently strong, played a lesser role in differentiating the apps. The weighted user ratings aligned with MARS scores, supporting the validity of the clustering model.

Conclusions: This study assesses the quality of mobile apps for DDI management by integrating MARS with K-means Clustering. This approach enabled a structured classification of apps based on the MARS scores, identifying distinct clusters that reflect overall app quality profiles across key usability dimensions. The study revealed that the influence of MARS dimensions on app ratings varies by cluster, highlighting that the significance of these dimensions shifts according to the specific needs and preferences of different user groups.

(JMIR Mhealth Uhealth 2025;13:e65927) doi:10.2196/65927

KEYWORDS

drug interactions; Mobile App Rating Scale (MARS); mobile health (mHealth); K-Means clustering; correlation analysis; app quality assessment; digital health; mobile app; United States

Introduction

Drug-drug interactions (DDIs) pose significant risks to patient safety and contribute substantially to adverse drug reactions, which can result in serious health complications and increased health care costs. DDIs occur when the pharmacological or clinical response to a drug is altered by the presence of another drug, potentially leading to harmful effects [1]. In the United States, DDIs account for approximately 26% of adverse drug reaction–related hospital admissions [2], highlighting the critical need for effective DDI management strategies.

The advent of mobile health (mHealth) applications has introduced new possibilities for managing DDIs by providing

XSL•FO
RenderX

decision support tools directly to consumers and health care professionals. These apps can facilitate the detection, avoidance, and reporting of potential DDIs, thereby enhancing medication safety and improving patient outcomes. The increasing availability and use of smartphones have made these tools accessible to a broad audience, including older adults who are often at higher risk for medication errors due to polypharmacy [3].

Despite the growing availability of mHealth apps for DDI management, their quality and usability vary significantly. Previous studies have assessed app quality using the mobile app rating scale (MARS) and found that while most apps scored well on information and functionality, they often lacked engaging features and consistent usability. For instance, Kim et al [4] and Shen et al [5] both reported that the information and functionality dimensions were rated the highest, while engagement consistently received the lowest scores. However, these studies relied primarily on descriptive MARS scores and did not explore how different quality dimensions influence user satisfaction or reveal patterns in app quality.

Building on these insights, our study aims to evaluate the quality of publicly available mobile apps for DDI management in the United States using the MARS and to identify patterns in user preferences [6-8]. Unlike previous studies that relied on descriptive MARS summaries, we employ k-means clustering, an unsupervised machine learning technique that has not been previously applied for MARS analysis before. This innovative approach enabled the identification of clusters of apps with similar quality profiles and helped uncover patterns in how MARS dimensions relate to user preferences. By linking expert-based MARS scores with real-world app store ratings, this study offers a more user-centered perspective and provides practical insights to guide future app development and evaluation.

Methods

App Selection Process and Distribution

The app selection process conducted in this study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol [9] as closely as possible. Although this was not a formal systematic review, our methodology was aligned with PRISMA principles, and the checklist was followed where applicable to ensure transparent and robust reporting of methods and results. Some deviations from PRISMA were necessary due to the distinct characteristics of mHealth app store databases, which were dissimilar from conventional scholarly reference databases. App store search algorithms are not transparent, and results may vary over time, limiting reproducibility.

The app identification process followed a structured approach to search for relevant apps using specific keywords (Table 1) related to drug interactions in the Apple App Store and Google Play Store. The search was conducted in March 2024 on a designated set of devices to ensure uniformity of results. All searches were performed on devices physically located in the United States using US-registered Apple ID and Google Play accounts to ensure inclusion of apps available in the US market only. The identified apps were screened for eligibility in 2 stages. First, the app descriptions were reviewed to determine if they claimed to check for DDIs and were published in English. Second, apps were excluded if they were not intended for general consumers, did not allow for multiple or combination interaction checks, were designed for pets or animals, or were specific to a particular disease or drug class. The selected app names were compiled, and the apps were downloaded and installed to verify their eligibility. A total of 19 apps met the inclusion criteria and were selected for subsequent analysis. User ratings and the number of reviews for these apps were recorded from the respective app stores. All data processing and statistical analyses were conducted using the latest version of RStudio.

The quality of the mHealth apps was assessed using MARS, a 23-item expert-based rating scale designed to evaluate various aspects of app quality [10]. The MARS scale provides a 5-point rating system to assess quality dimensions, such as end-user engagement, features, aesthetics, information, and subjective quality. In this study, we excluded the subjective quality dimension of the MARS scale. This decision was made to maintain the objectivity and consistency of our evaluation, as the subjective quality dimension is inherently more prone to bias, even among evaluators with similar professional backgrounds. Two evaluators independently assessed each app. Both were graduates of Weill Cornell Medicine with health care-related backgrounds and substantial experience in digital health. One evaluator is a licensed physician with clinical experience and advanced training in medical informatics. Another evaluator holds a master's degree in Biostatistics and has over five years of professional experience working with health data in research and industry settings. Their combined expertise ensured both clinical insight and technical rigor in evaluating app content, usability, and functionality. The interrater solid reliability (IRR) reflecting the degree of agreement [11] between the two evaluators was estimated using Cohen's kappa (κ) statistic. The IRR was calculated per dimension and for all apps, using a two-way random model for agreement level.

XSL•FO
RenderX

Table . Search Terms with REsulTS.

Search Term	Raw Results	Results After Exclusion Criteria
“Drug-Drug Interaction”	70	12
“Drug Checker”	56	5
“Medication Tracker”	145	1
“Drug Interaction”	85	6
“Drugs Interaction”	69	12
“Drug Interactions”	83	10
“Drugs Interactions”	83	10
“Drug-Interaction”	81	11
“Pill Interaction”	1	0
“Pills Interaction”	0	0
“Pill Interactions”	0	0
“Pills Interactions”	0	0
“Pill-Interaction”	1	0
“Medication Interaction”	18	2
“Medications Interaction”	0	0
“Medication Interactions”	105	4
“Medications Interactions”	2	1
“Medication-Interaction”	18	2

To understand the relationship between the different MARS dimensions, we first assessed the distribution of the data. We used quantile-quantile (Q-Q) normal plots to evaluate this. The Q-Q plots helped us to visually determine how well the data for each dimension fit a standard normal distribution. The Q-Q plots helped us to determine deviations from normality; therefore, we performed the Kruskal-Wallis test to compare the dimension groups and draw further inferences. The Kruskal-Wallis test is a non-parametric method used to determine whether there are differences between two or more groups by examining differences in their median values. We then performed correlation analysis to understand the strength and direction of the linear relationships between each pair of dimensions. The correlation coefficient ranges from −1 to 1, where 1 indicates a perfect positive correlation, 0 indicates no correlation, and −1 indicates a perfect negative correlation.

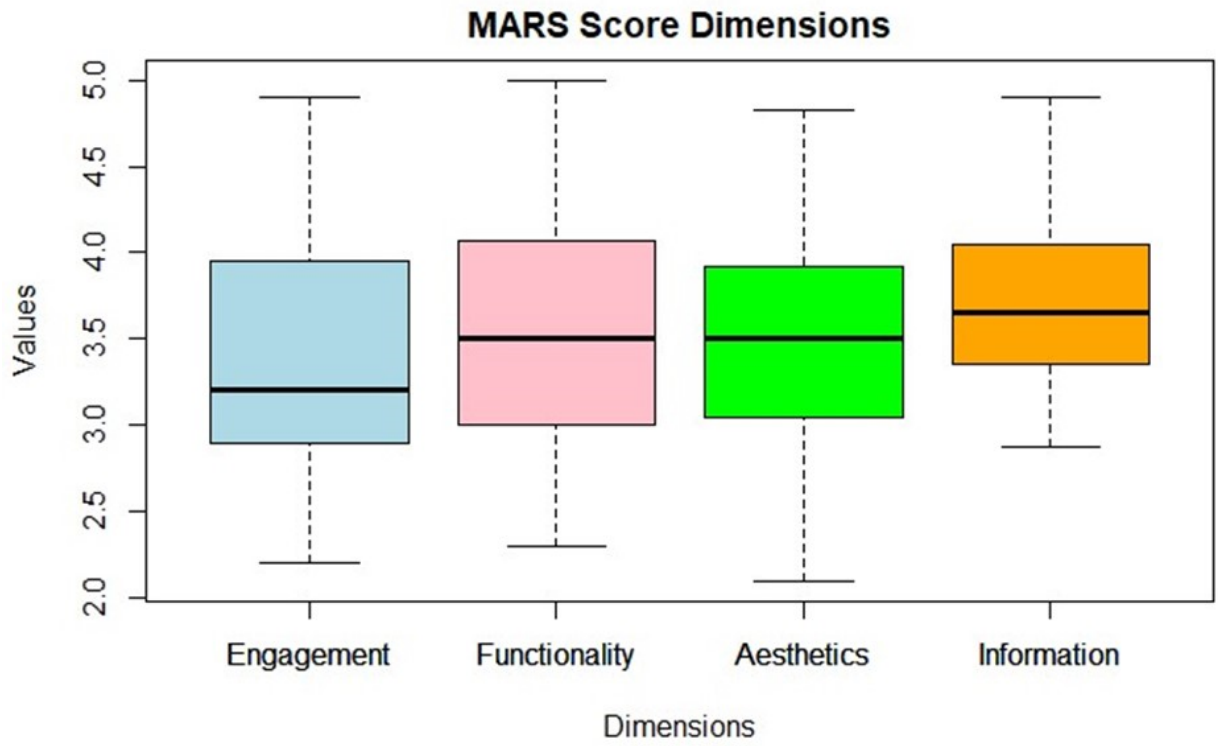
We then categorized the apps into distinct groups based on their MARS dimension scores using k-means clustering [12]. This unsupervised machine learning technique groups similar data points into distinct clusters, helping us identify patterns and relationships between the dimensions and group apps with similar characteristics. The first step in k-means clustering is to determine the optimal number of clusters. We used the elbow plot approach to identify this number [13]. The elbow plot shows the number of clusters against the within-cluster sum of squared errors (SSE). The plot shows a decreasing SSE as the number of clusters increases. The “elbow” point on the plot indicates where the rate of decrease in SSE slows down and becomes less pronounced. This point represents the optimal number of clusters, suggesting that adding more clusters does not significantly improve the clustering solution. While commonly used, this approach involves some degree of visual interpretation and subjectivity, and the inherent number of clusters may influence the resulting groupings. After identifying the optimal number of clusters, we then perform the k-means clustering (Figure 1).

After k-means, to further explore and visualize the clusters, we used scatter plots. The scatter plots allowed us to visually examine the distribution of MARS scores across different dimensions and understand the clustering patterns more clearly. By plotting the dimensions against each other, we identified how the clusters are separated and discerned relationships and trends within and between the clusters, providing a more intuitive understanding of the data.

Following this, we integrated user ratings and the number of reviews from both the Google Play Store and Apple App Store to validate our k-means clustering results with user feedback. To provide a more accurate estimate of app ratings, we calculated a weighted average that considers both the ratings and the number of reviews on each platform (Table 2). We rounded the number of reviews to the nearest multiple of 5. To evaluate the effectiveness of our clustering model, we compared the mean of weighted average ratings for all apps within each cluster with the mean MARS score for that cluster. This approach allowed us to validate our clustering methodology, ensuring it produces meaningful results that align with actual user experiences and preferences.

XSL•FO
RenderX

Figure 1. Mars Dimensions Score Box PloT. Mars: Mobile App Rating Scale.



XSL•FO
RenderX

Table . Ratings and Review COunts for APps on Google Play Store and Apple App Store

App Name	Google Play Store		Apple App Store		Weighted Average Rating
	Rating	Number of Ratings	Rating	Number of Ratings	
Apothera	0	0	4.3	15	4.30
Avicenna - Drug Interactions	0	0	0	0	0.00
Drug Compatibility Checker Tuh	1	25	0	0	1.00
Drug Interaction Checker +	4.7	145	4.1	60	4.52
Drugs.com Medication Guide	4.6	32,600	4.8	8600	4.64
Elsevier Clinical Pharmacology	3.3	400	2.6	5	3.29
Epocrates	4.1	26,500	4.5	6700	4.18
Everydose Medication Reminder	4.3	1250	4.8	3500	4.67
Goodrx	4.8	298,000	4.8	652,000	4.80
Lexidrug	3.3	3300	3.1	150	3.29
Medisafe Medication Manager	4.6	240,000	4.7	90,000	4.63
Medscape	4	635,00	3.6	1250	3.99
Micromedex Drug Interactions	4.1	500	3.8	100	4.05
Myrxprofile	5	85	3.6	25	4.68
Pill Identifier & Drug Search	3.9	1000	4.2	100	3.93
Pillbox	3.3	630	3.8	85	3.36
Pocket Pharmacist	4.5	11,770	4.7	3535	4.55
The Washington Manual	4.7	155	4.7	180	4.70
Uptodate	4	10,800	3.9	450	4.00

We then used quantile-quantile (Q-Q) plots to evaluate the distribution patterns of the Weighted Average Ratings and the MARS Mean Scores. The Q-Q plots revealed deviations from normality, and therefore, we used the paired Wilcoxon signed-rank test to compare these two metrics within each cluster. The Wilcoxon signed-rank test is a nonparametric test used to compare the median of two related samples when the data does not follow a normal distribution [14,15]. We chose the paired version of this test because the measures being compared—weighted average ratings and MARS scores—are evaluations of the same group of apps [16].

Additionally, to further explore the relationship between the MARS dimensions and the weighted average ratings within each cluster, we conducted a correlation analysis. This analysis aimed to identify the specific MARS dimensions that align closely with user ratings, as well as those that do not, thus providing a deeper understanding of how MARS dimensions relate to user perceptions.

Ethical Considerations

This study did not involve human participants, patient data, or identifiable information. The analysis was limited to publicly available mobile applications and app store metadata (descriptions, features, ratings). Institutional review board (IRB) approval was not required, and informed consent was not applicable.

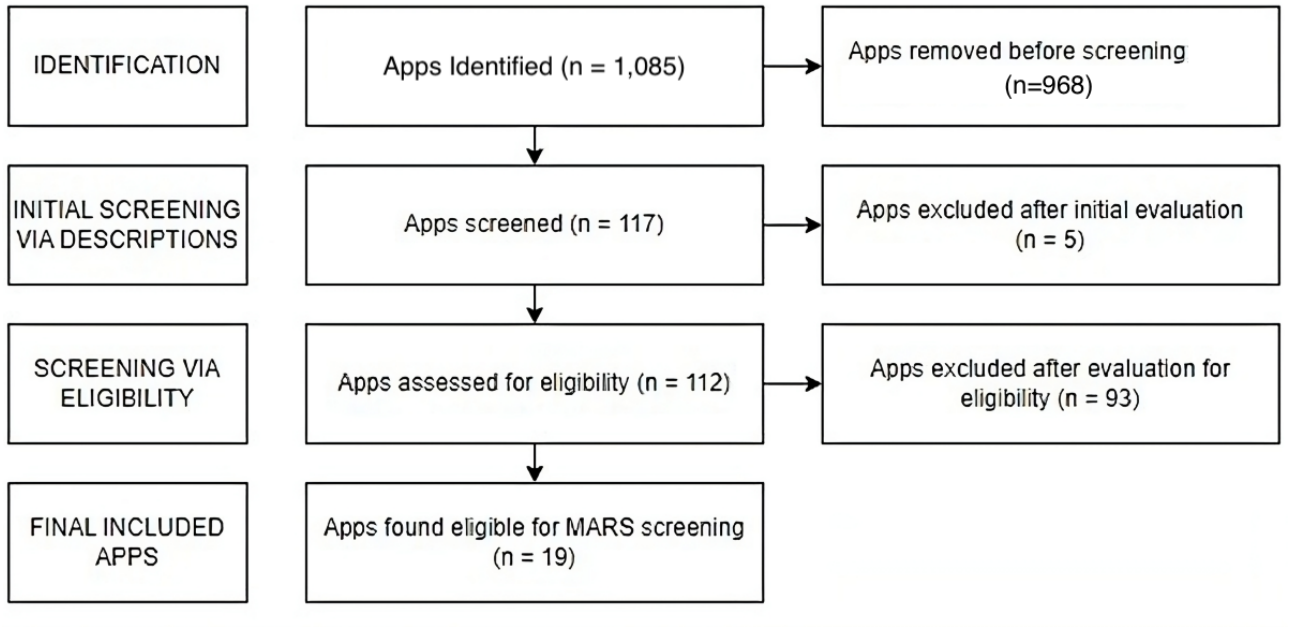
Results

App Identification and Selection

The app store search identified a total of 1085 apps from the Google Play Store and Apple App Store. After removing duplicates, 117 unique apps remained for screening. Based on their descriptions and functionality, 19 apps met all inclusion and exclusion criteria and were selected for full evaluation (Figure 2); a completed PRISMA checklist is provided in Checklist 1.

XSL•FO
RenderX

Figure 2. App Selection PRocesS. Mars: Mobile App Rating Scale.



Overall Scores

The mean MARS score of the 19 apps was 3.54. Some notable apps with high ratings were Medisafe Medication Manager, GoodRx, and Drugs.com Medication Guide (Table 3). On the other hand, Elsevier Clinical Pharmacology and Drug Compatibility Checker TUH received relatively low ratings across all categories. The IRR between two raters as assessed by the kappa statistic was 0.46.

The mean scores of the four dimensions of MARS were examined to investigate the magnitude of the differences in quality in each dimension (Table 4). The information dimension resulted in the highest mean and median scores (3.68 and 3.65 (IQR 3.35-4.05), respectively), whereas the engagement dimension showed the lowest mean score (3.42) and lowest median score (3.21 (IQR 2.90). These findings align with observations in similar studies [4,5]. The aesthetics dimension also had the most variability (SD 0.81).

XSL•FO
RenderX

Table . Apps and Their Mars[a] ScoreS.

App Name	Engagement	Functionality	Aesthetics	Information	Mean Mars Score
Apothera	3.10	4.00	3.83	3.40	3.58
Avicenna - Drug Interactions	3.00	3.63	3.00	3.00	3.16
Drug Compatibility Checker Tuh	2.45	2.90	2.67	2.87	2.72
Drug Interaction Checker +	3.80	3.30	3.75	3.60	3.61
Drugs.com Medication Guide	4.10	4.75	4.20	4.30	4.34
Elsevier Clinical Pharmacology	2.20	2.30	2.15	2.95	2.40
Epocrates	4.50	4.38	4.00	3.80	4.17
Everydose Medication Reminder	3.80	3.75	3.40	4.10	3.76
Goodrx	4.85	5.00	4.75	4.14	4.69
Lexidrug	2.90	3.10	3.30	3.12	3.11
Medisafe Medication Manager	4.90	5.00	4.83	4.75	4.87
Medscape	3.30	3.45	3.50	3.70	3.49
Micromedex Drug Interactions	2.60	2.72	2.10	3.65	2.77
Myrxprofile	3.10	2.90	3.12	3.30	3.11
Pill Identifier & Drug Search	2.60	2.70	2.10	3.60	2.75
Pillbox	3.21	3.45	3.53	3.87	3.52
Pocket Pharmacist	4.20	4.13	4.17	4.00	4.13
The Washington Manual	2.90	3.80	3.10	4.24	3.51
Uptodate	3.40	3.50	3.70	3.60	3.55

[a]Mars: Mobile App Rating Scale.

Table . Mars[a] Dimension SCoreS.

Dimensions	Mean Ratings	Median Ratings (Iqr)	Standard Deviations
Engagement	3.42	3.21 (2.90 -3.95)	0.80
Functionality	3.62	3.50 (3.0 - 4.06)	0.78
Aesthetics	3.43	3.50 (3.05 - 3.91)	0.81
Information	3.68	3.65 (3.35 - 4.05)	0.51

[a]Mars: Mobile App Rating Scale.

The Q-Q plots (Figure 3) reveal that engagement, functionality, and aesthetics dimensions deviate from a standard normal distribution, particularly at the tails, indicating the presence of more extreme values than expected under normality. In contrast, the information dimension adheres more closely to a normal distribution, with some minor deviations.

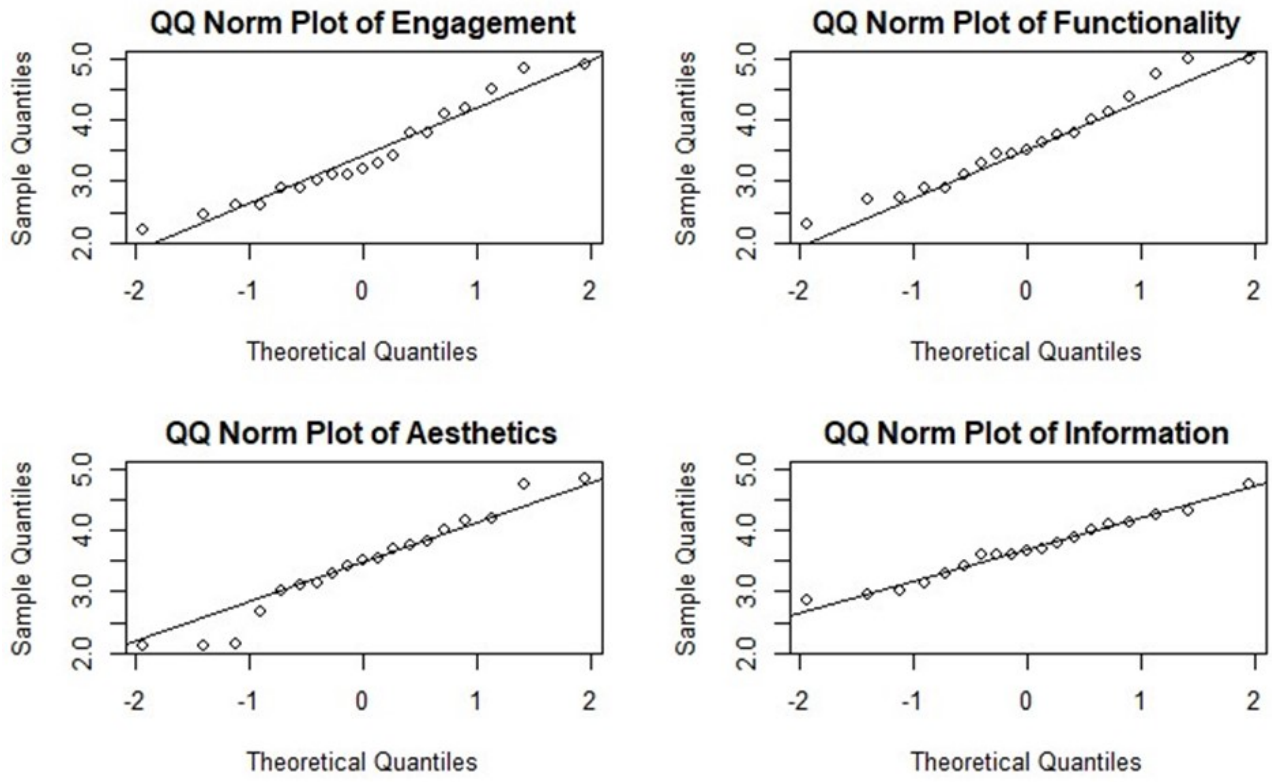
Since the dimension distributions are not uniformly normal, we perform the Kruskal-Wallis test. The null hypothesis is that the medians of all groups are equal, indicating no difference in the central tendency of these dimensions. The alternative hypothesis is that at least one dimension's median is different from the others. The significance level is taken as .05. The test resulted in a χ^{2}_{3} value of 2.109 and a *P* value of .55.

Since the *P* value is greater than .05, we have sufficient evidence to accept the null hypothesis that there is no significant difference between the median values of the dimension groups for engagement, functionality, aesthetics, and information. Therefore, there is no need to treat any dimension differently

XSL•FO
RenderX

in terms of interventions, improvements, or prioritization based on median scores alone.

Figure 3. Quantile-Quantile Normal Plots of Mars Dimensions. Mars: Mobile App Rating Scale.

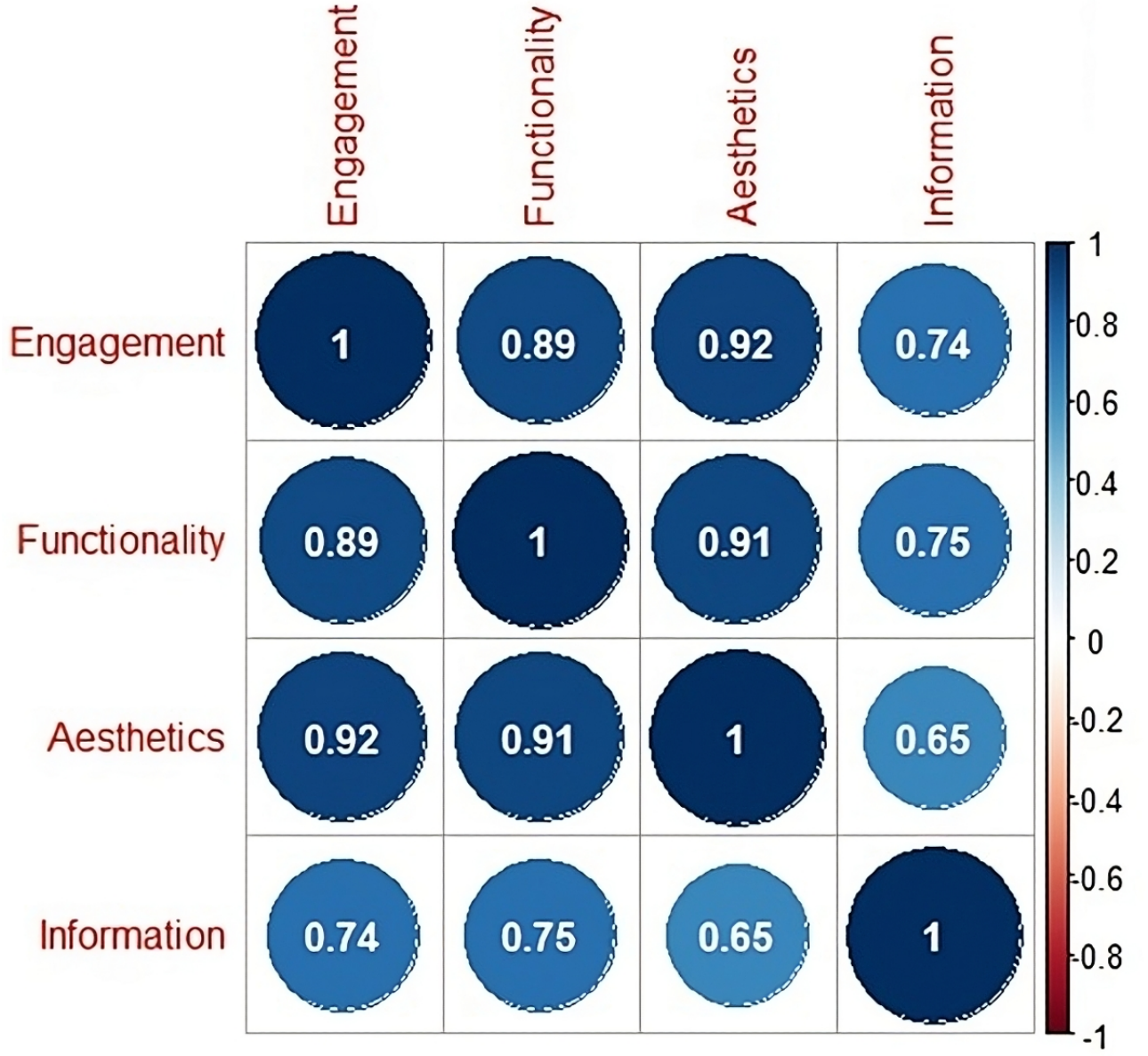


Correlation Between Dimensions

From the correlation plot (Figure 4), we see that all the correlation coefficients are positive, indicating a positive association between the dimensions. The strongest correlation is between aesthetics and engagement, with a coefficient of 0.92, followed closely by functionality and aesthetics with a coefficient of 0.91. The weakest correlation is between aesthetics and information, with a coefficient of 0.65.

XSL•FO
RenderX

Figure 4. Correlation Plot of Mars DImensionS. Mars: Mobile App Rating Scale.



K-Means Clustering of Dimensions

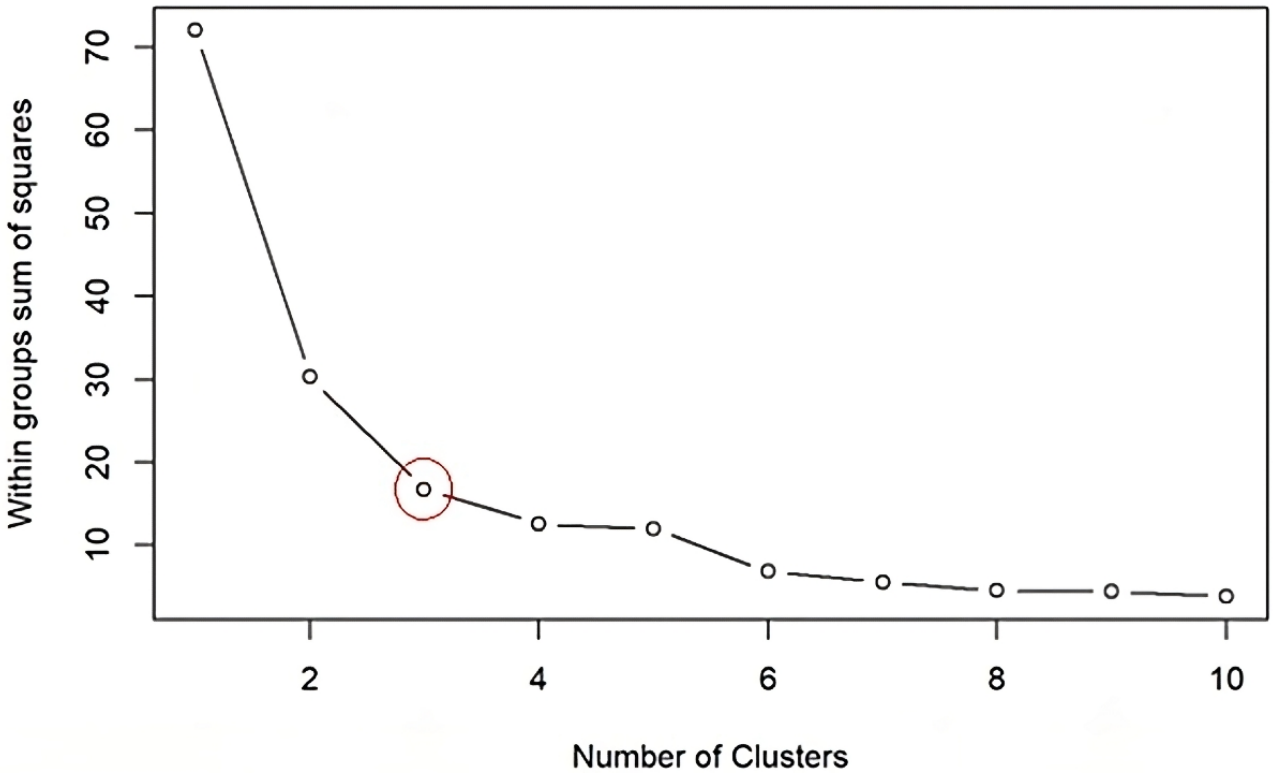
Given the above findings, we apply K-means clustering to categorize the apps into distinct groups based on their MARS dimension scores.

From the elbow plot (Figure 5), we see that 3 clusters are optimal for our study. We then run the k-means clustering, and each app is assigned to a cluster based on characteristics derived from the dimensions.

We see that cluster 3 apps consistently have the highest mean scores across all dimensions: engagement (4.51, SD 0.36), functionality (4.65, SD 0.39), aesthetics (4.39, SD 0.37), and information (4.20, SD 0.36; Table 5). This indicates that apps in cluster 3 generally perform exceptionally well across all dimensions. In contrast, cluster 2 apps show moderate mean scores for engagement (3.36, SD 0.34), functionality (3.61, SD 0.25), aesthetics (3.54, SD 0.25), and information (3.79, SD 0.30). Cluster 1 apps exhibit the lowest mean scores for engagement (2.69, SD 0.32), functionality (2.89, SD 0.41), aesthetics (2.63, SD 0.52), and information (3.21, SD 0.31), indicating lower overall performance.

XSL•FO
RenderX

Figure 5. Elbow Plot to Identify the Optimal Number of ClusterS.



XSL•FO
RenderX

Table . Apps in Cluster and the Mean and STandard DEviation (Sd) Scores by ClusteR.

Cluster	App Name	Engagement		Functionality		Aesthetics		Information	
		Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
1	Avicenna - Drug Interactions, Myrxprofile, Pill Identifier & Drug Search, Micromedex Drug Interactions, Lexidrug, Elsevier Clinical Pharmacology, Drug Compatibility Checker Tuh	2.69	0.32	2.89	0.41	2.63	0.52	3.21	0.31
2	The Washington Manual, Everydose Medication Reminder, Drug Interaction Checker +, Apothera, Uptodate, Medscape, Pillbox	3.36	0.34	3.61	0.25	3.54	0.25	3.79	0.30
3	Goodrx, Drugs.com Medication Guide, Medisafe Medication Manager, Pocket Pharmacist, Epocrates	4.51	0.36	4.65	0.39	4.39	0.37	4.20	0.36

From the box plots (Figure 6), we see that cluster 3 apps show consistently high median scores and minimal variability, indicating well-rounded apps. Cluster 2 has moderate medians with some variability, particularly in functionality and aesthetics, indicating a solid performance with certain areas that may need improvement. Cluster 1 exhibits the lowest medians and greatest variability, especially in aesthetics and engagement. This highlights a diverse performance within cluster 1, where some apps may have specific strengths, but overall, they are of lower quality compared to those in clusters 2 and 3.

To further explore and visualize these clusters, we use scatter plots to visually examine the distribution of app scores across different dimensions and understand the clustering patterns more clearly (Figure 7).

Cluster 3 (blue) consistently occupies the higher end of the scale in most dimensions, indicating that apps in this cluster are perceived to be of higher quality overall. Cluster 2 (green) apps show moderate ratings, occupying the middle range across dimensions. In contrast, cluster 1 (red) apps are positioned towards the lower end of the scale.

XSL•FO
RenderX

Figure 6. Box Plot of Mars Dimension Scores by ClusterS. Mars: Mobile App Rating Scale.

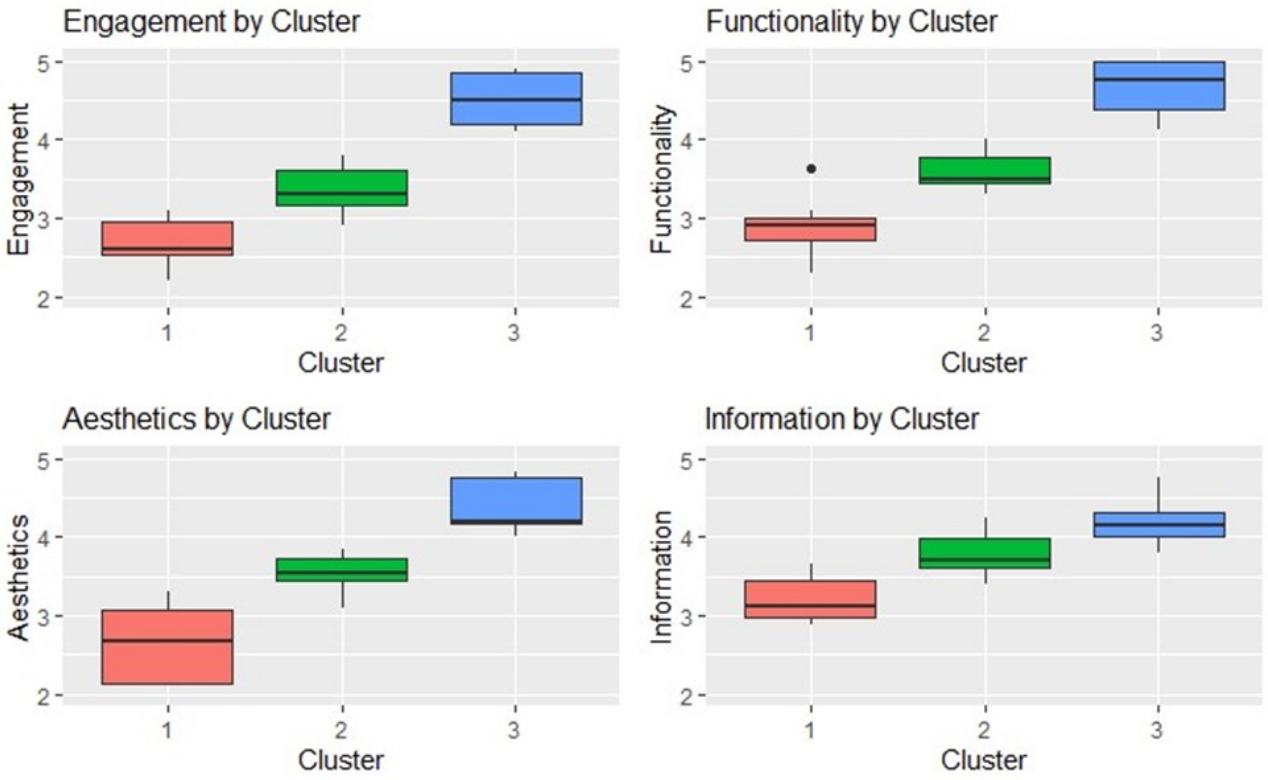
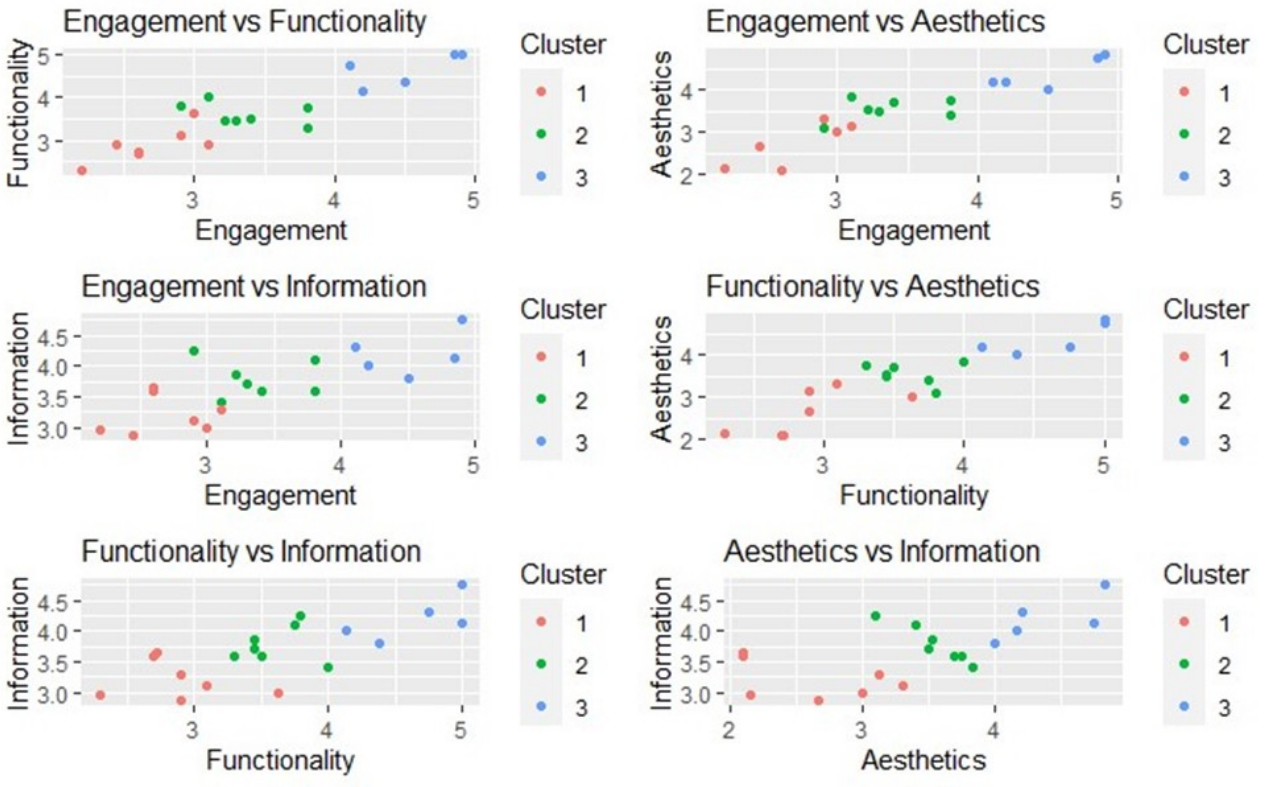


Figure 7. Scatter Plots of DImensionS.



Model Evaluation Based on User Ratings

To validate our K-means clustering model, we use the weighted app ratings. We first calculate the mean of weighted average ratings of all apps in the cluster and compare it with the mean MARS score of the cluster (Table 6).

Cluster 1 has the lowest mean weighted average rating of 3.30 and a relatively high standard deviation of 1.31, indicating lower user satisfaction with significant variability. This is mirrored by the mean MARS score, which is also the lowest at 2.86, suggesting that these apps are perceived as lower quality across the MARS dimensions, with a somewhat consistent perception

XSL•FO
RenderX

reflected by the standard deviation of 0.28. In cluster 2, we see moderate user satisfaction, with a mean weighted average rating of 4.22 and a lower standard deviation of 0.64, indicating more consistent user experiences. This is closely aligned with the mean MARS score of 3.57, which, along with a very low standard deviation of 0.09, reflects a consistent perception of moderate app quality. Cluster 3 stands out with the highest mean weighted average rating of 4.84 and the lowest variability (SD=0.36), indicating that these apps are consistently well received by users. The high mean MARS score of 4.44, despite a slightly higher standard deviation of 0.33, reinforces the perception that these apps are of high quality across all MARS dimensions. To further substantiate these observations, we then proceeded to determine the statistical significance of the differences between the clusters.

The Q-Q plots for the weighted average ratings and mean MARS scores (Figure 8) show significant deviations from the reference line, indicating that the distribution of these ratings deviates from normality. Hence, we perform the paired Wilcoxon signed-rank test to compare the weighted average ratings and MARS scores. The null hypothesis is that there is no difference in the median values between the weighted average ratings and the MARS scores for the same apps. The alternative hypothesis is that there is a difference in the median values, and the significance level is taken as 0.05.

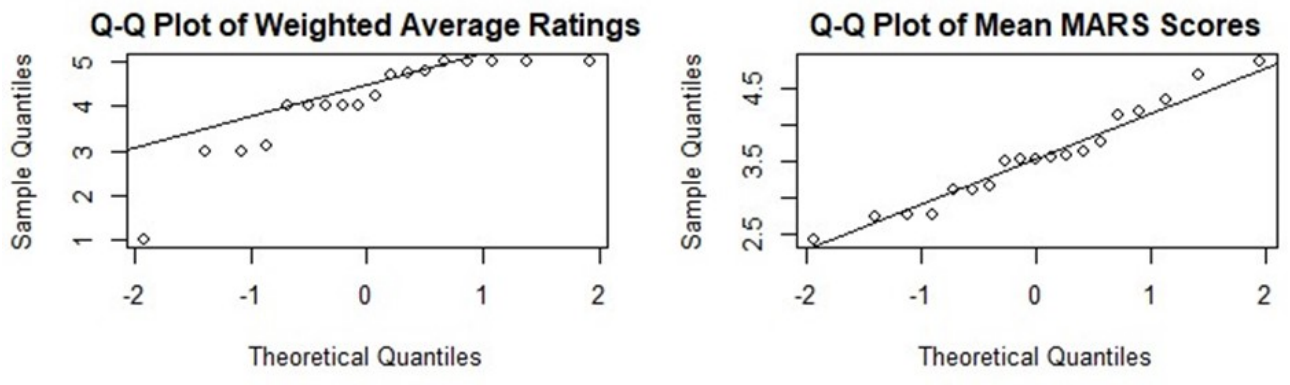
Table . Comparison of Weighted Average Ratings and Mars[a] Score by ClusteR.

Cluster	Mean of Weighted Average Ratings	Median of Weighted Average Ratings	Sd of Weighted Average Ratings	Mean Mars Score	Median Mars Score (Iqr)	Sd of Mars Score	Test Statistic (V) of The Wilcoxon Signed-Rank Test	*P* VAlue
1	3.30	3.50	1.31	2.86	2.77 (2.74-3.11)	0.28	14	.56
2	4.22	4.00	0.64	3.57	3.55 (3.51-3.60)	0.09	27	.03*[b]
3	4.84	5.00	0.36	4.44	4.34 (4.17-4.69)	0.33	15	.06

[a]Mars: Mobile App Rating Scale.

[b]*P*<.05

Figure 8. Q-Q Plots of Weighted Average Ratings and Mean Mars Scores. Mars: Mobile App Rating Scale.



The Wilcoxon signed-rank tests (Table 6) revealed distinct variations in the relationship between weighted average ratings and MARS scores across the clusters. In cluster 1, the *P* value of .56 suggests that there is no statistically significant difference between the median weighted average ratings and the median MARS scores, indicating that these two metrics are fairly aligned within this cluster. For cluster 2, however, the *P* value of .03 is below the significance threshold of .05, indicating a statistically significant difference between the two medians. This suggests that the weighted average ratings and MARS scores in cluster 2 may reflect different aspects of app quality or user perceptions. This could imply that the factors contributing to higher or lower user ratings on app stores may differ from those emphasized by the MARS criteria. In cluster 3, the *P* value of .06 is slightly above the .05 threshold, suggesting a marginal difference between the medians that needs more exploring.

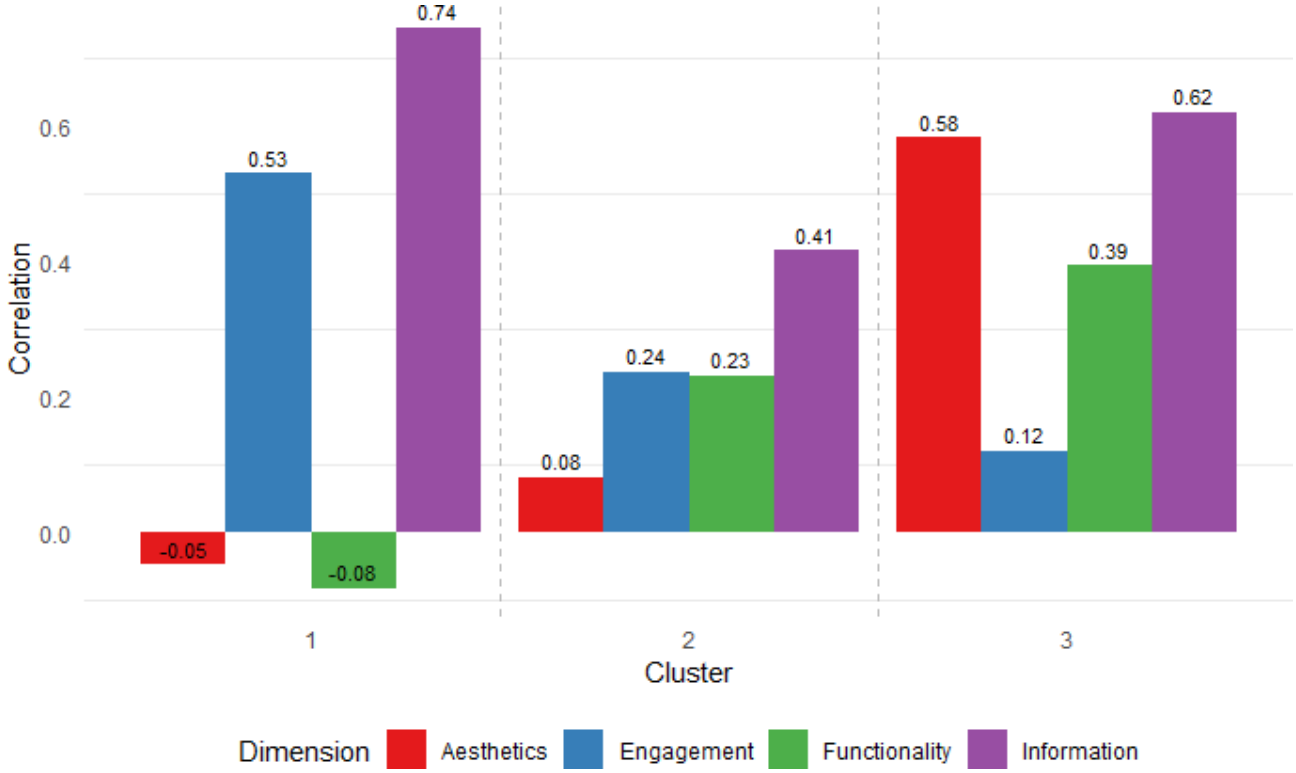
To further understand the reasons behind these observations, we conducted a correlation analysis between the MARS dimensions and weighted average ratings within each cluster (Figure 9). This analysis helped identify which specific MARS components align closely with user ratings and which do not, providing deeper insights into the facets of app quality that resonate with users compared to those emphasized by the MARS criteria.

We observed that for cluster 1 apps, information and engagement are key drivers of user satisfaction, while aesthetics and functionality have less influence. In cluster 2, information quality stood out as the most significant factor, with

XSL•FO
RenderX

functionality and engagement having a more moderate impact. Cluster 3 showed that aesthetics and information are the primary factors influencing user ratings, with engagement playing a minimal role. These insights suggest that user priorities differ across clusters, with some clusters valuing functionality and engagement more, while others placing greater importance on aesthetics and information.

Figure 9. Correlations Between The Weighted Average Rating and Mars Dimension Scores by CLusteR. Mars: Mobile App Rating Scale.



Discussion

Principal Findings and Comparison With Previous Works

This study evaluated the quality of 19 mobile apps for DDI management using MARS and applied K-means clustering to identify patterns in app quality and user satisfaction. The results showed that while the information dimension received the highest scores, engagement was consistently rated the lowest. Strong correlations were observed across MARS dimensions, suggesting that app quality is multidimensional and interrelated. Cluster analysis revealed 3 distinct app quality profiles, with variations in how specific dimensions aligned with user ratings, highlighting differences in user preferences and design priorities across app types.

Among the MARS dimensions, the information dimension received the highest scores, indicating that these apps generally provide comprehensive and reliable content. This aligns with the core function of DDI apps, which are designed to deliver accurate drug interaction information quickly and efficiently. This finding is consistent with prior studies on drug-related mHealth apps, which also identified information as the most consistently strong dimension [4,5].

In contrast, the engagement dimension scored the lowest and exhibited considerable variability, suggesting that many DDI apps struggle to retain user interest. This is not unexpected, as these apps are typically used as reference tools for quick, task-oriented interactions rather than prolonged engagement. As such, lower engagement scores may reflect their intended purpose rather than a design flaw. This trend was also observed in prior studies by Kim et al [4] and Shen et al [5], where engagement received the lowest or most variable scores among MARS dimensions. It also aligns with broader mHealth research, which shows that reference apps often score lower in engagement due to their transactional design focus [17,18].

The functionality dimension scored relatively high, which indicates that these apps generally offer effective and user-friendly features. This is crucial for DDI apps, as users expect reliable and straightforward tools to check drug interactions efficiently. The high functionality scores were seen to be in alignment with previous studies [4,5] and suggest that most of these apps succeed in delivering on their core promise: providing users with easily accessible information.

Finally, the aesthetics dimension showed moderate scores with noticeable variability. This suggests that while some apps are visually appealing and offer a pleasant user experience, others may fall short in terms of design and visual coherence. Aesthetics can influence user perceptions, even for utilitarian apps like those for drug interactions. Although it might not be the primary concern for users, a well-designed interface can enhance usability and overall satisfaction.

The Kruskal-Wallis test showed no significant difference between the median values of the dimensions. This suggests that while there are notable differences in variability and

XSL•FO
RenderX

individual scores, the overall central tendency of these dimensions does not differ significantly, indicating that no single dimension consistently outperforms or underperforms the others.

The correlation analysis showed strong positive relationships among MARS dimensions, indicating that apps performing well in one area tend to perform well across others. This interrelatedness suggests that improvements in any one domain may contribute to overall app quality and higher user satisfaction. For example, an app that is engaging and visually appealing (high engagement and aesthetics scores) is more likely to also have good functionality and provide useful information (high functionality and information scores), leading to a higher overall rating. Conversely, an app that performs poorly in one or more dimensions is more likely to have a lower overall rating. Therefore, app developers and designers should strive to create apps that perform well in all dimensions, as this is likely to lead to higher user satisfaction and better ratings.

The K-means clustering analysis revealed 3 distinct app groupings based on MARS dimension scores. Cluster 3 apps demonstrated high quality and consistency across all dimensions; cluster 2 showed moderate performance with variability; and cluster 1 included underperforming apps with the greatest variation. These groupings were supported by box plots showing clear differences in overall score distributions. Scatter plot analysis showed distinct clustering patterns for engagement, functionality, and aesthetics—dimensions that most strongly differentiated app quality. Cluster 3 apps consistently scored high across these, while cluster 1 scored lowest. However, plots involving information showed more overlap, particularly between clusters 2 and 3, indicating that information quality alone was less effective in distinguishing apps. This suggests that engagement, aesthetics, and functionality play a more significant role in shaping user perceptions.

We validated our clustering model by comparing mean MARS scores with weighted user ratings. Cluster 3 had the highest satisfaction, followed by cluster 2, with cluster 1 scoring lowest—demonstrating strong alignment between expert evaluation and user feedback. This correlation suggests that our K-means clustering model effectively distinguishes apps based on their quality and user satisfaction.

To further substantiate our observations, we assessed the statistical significance of the differences between clusters. Wilcoxon tests revealed no significant difference between MARS and user ratings in clusters 1 and 3, indicating alignment between expert and user perceptions. In contrast, cluster 2 showed a statistically significant divergence, suggesting that user satisfaction in this group may depend on different factors than those captured by MARS.

To explore these differences further, we conducted a correlation analysis between the MARS dimensions and weighted average ratings. This helped to identify which MARS dimensions were most closely aligned with user ratings across clusters, revealing distinct priorities for app quality.

In cluster 1, information and engagement emerged as the primary drivers of user satisfaction, with strong positive correlations to user ratings (information=0.74, engagement=0.53). Although functionality had a higher average MARS score (mean 2.89), it showed a negligible negative correlation (−0.08) with user ratings, suggesting that users in this cluster did not perceive added functionality as a key value driver. Aesthetics showed similarly low influence. These patterns may reflect preferences of clinically oriented users, such as health care professionals or informed patients, who prioritize reliable, content-driven tools over design or feature-rich interfaces.

In cluster 2, the significant divergence between metrics, as revealed by the Wilcoxon test, is supported by the correlation analysis. Information quality (mean 3.79, SD 0.30) correlates strongly with user ratings (correlation=0.41), highlighting its importance to users. Although aesthetics received a relatively high MARS score (mean=3.54), it showed the weakest correlation with user ratings (r=0.08), suggesting that visual design had little influence on user satisfaction in this cluster. In contrast, engagement (mean=3.36, r=0.24) and functionality (mean=3.61, r=0.23) both showed modest associations, indicating these dimensions were somewhat more aligned with user expectations. These results suggest that while MARS evaluates certain design and usability features highly, users in this group appear to place greater value on information quality, contributing to the observed divergence between expert and user evaluations.

Cluster 3 shows a borderline difference between MARS scores and user ratings. The correlation analysis reveals that information (mean 4.20, SD 0.36) and aesthetics (mean 4.39, SD 0.37) are the primary factors driving user satisfaction, with strong correlations of 0.62 and 0.58, respectively. Functionality (mean 4.65, SD 0.39) and engagement (mean 4.51, SD 0.36) have moderate to low correlations (functionality=0.39, engagement=0.12). This suggests that while overall alignment exists, users in this cluster place higher value on design and content than interactivity.

Overall, these findings demonstrate that while MARS scores and user ratings often align—particularly in cluster 1—important exceptions exist, most notably in cluster 2, where user feedback may reflect aspects of app quality not fully captured by MARS. By integrating dimension-level MARS scores with user rating correlations, this study offers a more nuanced understanding of the components of app quality that matter most to different user groups. The results highlight that user preferences are not uniform; some users prioritize functionality and efficiency, while others value informational depth or visual appeal. Unlike prior studies by Kim et al [4] and Shen et al [5], which focused primarily on aggregate MARS scores, our use of K-means clustering captures variation in how different quality dimensions resonate with users. This layered approach provides more actionable guidance for tailoring app design to diverse user needs.

Conclusion

The study assessed 19 mobile apps designed to check for DDIs using the MARS rating scale, with an overall mean MARS score of 3.54. Among the MARS dimensions, information scored the highest, while engagement had the lowest scores and showed the most variability. These results are consistent with the nature

XSL•FO
RenderX

of DDI apps, which are primarily used as reference tools where users prioritize quick access to accurate information over prolonged interaction. The positive correlations between all MARS dimensions highlight their interrelatedness, emphasizing the need for well-rounded apps to achieve the most favorable user reviews.

K-means clustering grouped the apps into three distinct clusters: cluster 3, consisting of 5 apps, demonstrated the highest ratings across all dimensions, indicating well-rounded app quality. Cluster 2, with 7 apps, exhibited moderate ratings, reflecting a balanced but less consistent performance. Cluster 1, also with 7 apps, had the lowest ratings, indicating poorer overall performance compared to the other clusters. Scatter plot analysis further revealed that while maintaining strong informational content is important, engagement, functionality, and aesthetics are the key drivers of user perceptions and app quality. These dimensions are critical for distinguishing an app and enhancing user satisfaction.

The comparison of mean weighted app ratings with mean MARS scores confirmed that the K-means clustering model effectively differentiated app quality, with higher MARS scores correlating with greater user satisfaction across clusters. Analyzing the clusters revealed strong alignment between user satisfaction and MARS scores in cluster 1, significant divergence in cluster 2, and slight variation in cluster 3, highlighting differences in how app quality is evaluated. The correlation analysis showed that cluster 1 aligned with MARS scores through engagement and information, while cluster 2 diverged, with users prioritizing information quality. Cluster 3 mostly aligned with MARS scores, favoring well-designed, informative apps, offering valuable insights for developers to better meet user expectations.

In summary, this study offers valuable insights into medication and DDI-checking apps by highlighting the key dimensions that drive user satisfaction. The research introduces an innovative approach by applying K-means clustering to MARS ratings, segregating apps into distinct clusters, and analyzing how different dimensions influence user satisfaction within each cluster. The findings reveal that MARS ratings are cluster-dependent, with dimensions like functionality, engagement, information, and aesthetics playing varying roles depending on the user base. The study also indicates that app ratings received are not uniformly correlated with MARS scores and will differ by cluster based on the specific preferences and needs of different user groups. By uncovering these relationships, the study provides developers with actionable insights to tailor app improvements more effectively according to their targeted user base.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, it focused exclusively on English-language apps available in the United States, which may limit the global applicability of the findings. Second, mobile apps are dynamic and frequently updated, so our evaluations represent a snapshot in time and may not reflect current app versions. Third, the lack of transparency in app store algorithms affects search reproducibility, as results can vary by device, location, or personalization. Fourth, although rigorous inclusion criteria were applied, the final sample of 19 apps may not capture the full spectrum of DDI-related apps, limiting generalizability. Additionally, the search was limited to the Apple App Store and Google Play Store, which may have excluded relevant apps available through other sources such as web-based tools or alternative platforms, introducing potential sample bias. Fifth, user ratings are dynamic and may be influenced by external factors such as app updates or public sentiment. Sixth, while the MARS tool provides a structured framework, it is based on subjective assessment and may introduce evaluator bias. Two expert raters were used in the evaluation process, which may not fully capture the diversity of professional perspectives. Finally, the study did not include assessments from typical end-users, which may limit the extent to which patient-centered usability perspectives are represented.

Future Directions

Future studies could expand the analysis by incorporating lay user perspectives through the use of the uMARS tool, which would offer deeper insights into real-world usability and patient-centered experiences. Additionally, integrating specific app features—such as interactive tools, visual aids, and personalization settings—into the clustering model alongside MARS dimensions could enhance the understanding of which functionalities most influence user satisfaction. Given the rapid evolution of mHealth technologies, periodic reassessment of app quality is also recommended to ensure continued relevance and accuracy of evaluations.

Acknowledgments

We would like to express our sincere gratitude to Dr. Prachi Khandare and Rhia Singh from Cornell University for their invaluable assistance in rating the apps based on the MARS scale and providing insightful feedback. No external writing assistance was used in the preparation of this manuscript.

No generative AI or large language models were used in the writing, editing, or analysis of this manuscript.

This research received no external funding.

Data Availability

The data analyzed in study are available from the corresponding author upon request. The evaluation metrics (MARS scores) and search strategy are documented within the manuscript.

XSL•FO
RenderX

Authors' Contributions

AB and JFFA jointly contributed to the conceptualization and methodology of the study. AB conducted the formal analysis and data curation and prepared the original draft of the manuscript. Both authors contributed to the review and editing of the manuscript. JFFA provided overall supervision and guidance throughout the project.

Conflicts of Interest

None declared.

Checklist 1
PRISMA Checklist.
[PDF File, 139 KB - mhealth_v13i1e65927_app1.pdf]

References

1. Scheife RT, Hines LE, Boyce RD, et al. Consensus recommendations for systematic evaluation of drug-drug interaction evidence for clinical decision support. Drug Saf 2015 Feb;38(2):197-206. [doi: 10.1007/s40264-014-0262-8] [Medline: 25556085]
2. Yee JL, Hasson NK, Schreiber DH. Drug-related emergency department visits in an elderly veteran population. Ann Pharmacother 2005 Dec;39(12):1990-1995. [doi: 10.1345/aph.1E541] [Medline: 16288080]
3. Peng Y, Wang H, Fang Q, et al. Effectiveness of mobile applications on medication adherence in adults with chronic diseases: a systematic review and meta-analysis. J Manag Care Spec Pharm 2020 Apr;26(4):550-561. [doi: 10.18553/jmcp.2020.26.4.550] [Medline: 32223596]
4. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74. [doi: 10.2196/mhealth.8613] [Medline: 29592848]
5. Shen C, Jiang B, Yang Q, et al. Mobile apps for drug-drug interaction checks in Chinese app stores: systematic review and content analysis. JMIR Mhealth Uhealth 2021 Jun 15;9(6):e26262. [doi: 10.2196/26262] [Medline: 33962910]
6. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422]
7. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, et al. Assessment of the quality of mobile applications (apps) for management of low back pain using the mobile app rating scale (MARS). Int J Environ Res Public Health 2020 Dec 9;17(24):9209. [doi: 10.3390/ijerph17249209] [Medline: 33317134]
8. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718. [doi: 10.2196/10718] [Medline: 30361196]
9. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 1;4(1):1. [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
10. Seo S, Cho SI, Yoon W, Lee CM. Classification of smoking cessation apps: quality review and content analysis. JMIR Mhealth Uhealth 2022 Feb 17;10(2):e17268. [doi: 10.2196/17268] [Medline: 35175213]
11. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34. [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
12. Siezenga AM, Mertens E, Gelder JV. FutureU - clustering users of a smartphone app-intervention based on usage and engagement. CrimRxiv 2023. [doi: 10.21428/cb6ab371.34389086]
13. Kodinariya TM, Makwana PR. Review on determining number of cluster in K-means clustering. IJARCSMS 2013;1(6):90-95 [FREE Full text]
14. Anaene Oyeka IC, Ebuh GU. Modified Wilcoxon signed-rank test. In: Bernard RJG, Lee MLT, editors. OJS 2012;02(2):172-176. [doi: 10.4236/ojs.2012.22019]
15. Agiro A, Bhattacharya A, Khan Z, Tung A, Brahmbhatt Y. PUK13 characteristics of patients receiving sodium zirconium cyclosilicate during hospital-based treatment of hyperkalemia. Value Health 2021 Jun;24:S236. [doi: 10.1016/j.jval.2021.04.1181]
16. Rosner B, Glynn RJ, Lee MLT. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics 2006 Mar;62(1):185-192. [doi: 10.1111/j.1541-0420.2005.00389.x] [Medline: 16542245]
17. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
18. Arthurs N, Tully L, O'Malley G, Browne S. Usability and engagement testing of mhealth apps in paediatric obesity: a narrative review of current literature. Int J Environ Res Public Health 2022 Jan 27;19(3):1453. [doi: 10.3390/ijerph19031453] [Medline: 35162470]

XSL•FO
RenderX

Abbreviations

DDI: drug-drug interaction
IRR: inter-rater solid reliability
MARS: mobile app rating scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
SSE: sum of squared errors

Edited by L Buis; submitted 29.08.24; peer-reviewed by A Schmieder, R Pryss; revised version received 23.04.25; accepted 14.05.25; published 13.10.25.

Please cite as:
Bhattacharya A, Florez-Arango JF
Investigating the Quality of Mobile Apps for Drug-Drug Interaction Management Using the Mobile App Rating Scale and K-Means Clustering: Systematic Search of App Stores
JMIR Mhealth Uhealth 2025;13:e65927
URL: https://mhealth.jmir.org/2025/1/e65927
doi:10.2196/65927

© Ayush Bhattacharya, Jose Fernando Florez-Arango. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of Mobile Intermittent Fasting Applications in Chinese App Stores: Quality Evaluations and Content Analysis

Laihao Fang[1,2,3], ME; Cheng Huang[1,2], MA; Bing Lin[4], PhD; Kuanlin Lei[1], BMgt; Jiazhen Zhou[1], BMgt; Xiaoni Zhong[4], MA; Yanbing Liu[1], PhD; Jiaxiu Liu[1,2], PhD

[1]College of Artifical Intelligence Medicine, Chongqing Medical University, No.1 Medical College Road, Chongqing, China

[2]Medical Data Science Academy, Chongqing Medical University, Chongqing, China

[3]School of Software, Shandong University, Jinan, China

[4]School of Public Health, Chongqing Medical University, Chongqing, China

Corresponding Author:
Jiaxiu Liu, PhD
College of Artifical Intelligence Medicine, Chongqing Medical University, No.1 Medical College Road, Chongqing, China

Abstract

Background: Obesity and related disorders are rising globally, especially in China, where they are linked to chronic diseases like diabetes and cardiovascular issues. As intermittent fasting (IF) gains popularity for weight management, the use of IF apps has increased, yet their quality varies significantly. A systematic evaluation of these apps is essential to assess their effectiveness and reliability.

Objective: This study aimed to conduct a comprehensive evaluation of IF apps available in the Chinese mobile app market. We concentrated on evaluating their features, quality, and overall user experience to help users avoid low-quality options and direct app developers to enhance their offers.

Methods: A systematic search was performed across 5 major app stores in China, including the Apple App Store, Huawei AppGallery, Oppo Software Store, Vivo App Store, and Xiaomi Market. “Fasting”, “Intermittent Fasting”, “Time-Restricted Feeding”, “Time-Restricted Fasting”, “Time-Restricted Eating” and “Meal Skipping” were used as keywords to identify relevant apps, which were then screened based on inclusion and exclusion criteria. The evaluation was conducted using the user version of the Mobile Application Rating Scale (uMARS). The uMARS assessment examined 4 key subscales: engagement, functionality, aesthetics, and information. Each app was independently evaluated by 2 raters who underwent uniform training to ensure consistency in scoring.

Results: A total of 35 apps were assessed for the study. These apps mostly contain features such as fasting timer (100.0%), recording weight (97.14%), fasting reminder (85.71%), and recording water intake (85.71%). All of the apps have an obvious privacy protection. Most of the apps (79%) have tools for quantifying users’ health status. The results showed that the overall average uMARS score across the apps was 4.35 (SD 0.51). The subscale scores were as follows: engagement 4.42 (SD 0.47), functionality 4.65 (SD 0.31), aesthetics 4.19 (SD 0.64), and information 4.15 (SD 0.58). The functionality subscale had the highest mean score, while the aesthetic subscale showed the greatest range of scores, from 2.17 to 5.00. The overall uMARS score was significantly positively correlated with the subscale scores (r=0.786 - 0.953, P<.001). The user ratings in the app stores did not significantly correlate with the uMARS overall scores (r=−0.290, P=.091). Strong inter-rater reliability was confirmed by intraclass correlation coefficients (ICC=0.809 - 0.909 across subscales).

Conclusions: All the apps reveal high overall quality but gaps in professional engagement and social features. Limited clinical input may undermine the evidence-based accuracy and long-term applicability of some apps. Developers are encouraged to collaborate with health care professionals to enhance content reliability and incorporate social features to boost user engagement, while ensuring robust privacy protections and reasonable use of artificial intelligence.

(JMIR Mhealth Uhealth 2025;13:e66339) doi:10.2196/66339

KEYWORDS

mHealth; mobile health; mobile applications; digital interventions; app evaluation; intermittent fasting; weight management; content analysis

XSL•FO
RenderX

Introduction

Recent statistics indicate that overweight/obesity continues to rise globally, with more than 2 billion people, or 30% of the world's population, already overweight [1], prompting increased public health concerns. Nowadays, China has the highest number of people who are overweight and obese [2]. Obesity is linked to numerous chronic health conditions, including cardiovascular disease, cancer, diabetes, osteoarthritis, and chronic kidney disease [3-6], making weight management a crucial public health priority. Dietary interventions are stepping into the spotlight as people seek effective ways to lose weight and improve overall health [7].

With the rise of digital health technologies, mobile health (mHealth) apps have become an increasingly popular tool for managing various health conditions and lifestyle modifications [8]. Among these, intermittent fasting apps have gained significant attention due to the growing interest in intermittent fasting as a method for weight management and metabolic health improvement. Intermittent fasting, characterized by alternating periods of eating and fasting [9], is purported to offer various health benefits, such as improvements in dyslipidemia, blood pressure, and longer-term effects on cardiometabolic health [10,11]. A meta-analysis examined a randomized controlled trial of 899 Chinese participants, found that intermittent fasting significantly reduced body weight and BMI and was effective in improving insulin resistance and lipid indices [12], which not only supports the potential of intermittent fasting as a weight management strategy in the Chinese population but also demonstrates its scientific validity as a metabolic health intervention. These findings confirm the metabolic modulatory effects of intermittent fasting at the level of biological mechanisms. In this context, global studies are deepening the understanding of intermittent fasting from multiple dimensions, including the exploration of its mechanisms, the validation of an application for the assessment of dietary compliance [13], retention rates, fasting patterns, and weight loss effects of an intermittent fasting app [14]. Moreover, a qualitative analysis of the descriptions of weight loss apps on the Spanish market showed that features like "quick results" and "personalized plans" dominate marketing, but often lack empirical support, suggesting gaps in the quality of the apps and the need for a systematic evaluation [15]. Current studies are limited to systematic evaluations of fitness and weight management apps or focus on single intermittent fasting applications, and there have been no systematic evaluations of multiple intermittent fasting applications [13,16-18]. The main features offered by intermittent fasting apps in the Chinese app market range from tracking fasting periods and providing reminders to providing educational content about intermittent fasting and nutrition.

Given the significant role that intermittent fasting apps can play in influencing users' health behaviors and aiding the development and improvement of mHealth applications, it is crucial to evaluate their quality. Therefore, we used the user version of the Mobile Application Rating Scale (uMARS) because it is a widely recognized and validated tool specifically designed for assessing the quality of mobile health apps [19]. uMARS is an optimized and user-friendly version of the original Mobile Application Rating Scale (MARS). Its reliability and validity have been verified, demonstrating a high Cronbach α coefficient in different linguistic contexts [20-24] and various scenarios [25] such as weight management and health tracking applications.

This study seeks to deliver a comprehensive functional analysis and systematic evaluation of intermittent fasting applications available in the Chinese mobile application market. In 2025, designated as the "Year of Weight Management" for Chinese citizens, China is implementing a 3-year action plan for weight management as a part of "Healthy China 2030" national strategy [26]. Adopting a user-centered approach, this study evaluates these applications through subscales reflecting user needs, preferences, and viewpoints. This research on intermittent fasting apps in China App Stores has significant scientific value and practical guidance, helping users identify low-quality options, determine effectiveness for weight loss and health management, and enhance user overall experience. Additionally, evaluating app quality yields valuable data for academic research, thereby highlighting strengths, limitations, and potential for enhancement.

Methods

Ethical Considerations

According to Article 32 of the Ethical Review Measures for Human Life Science and Medical Research Involving Human Participants (2023), jointly issued by the National Health Commission, the Ministry of Education, the Ministry of Science and Technology, and the State Administration of Traditional Chinese Medicine, research in human life science and medical studies that utilizes legally obtained public data or data generated through observation without interfering with public behavior may be exempt from ethical review [27]. Upon verification, the research described in the manuscript conforms to the circumstances outlined in the aforementioned regulations for exemption from ethical review. This study involves no human participants, and all data used were collected from legally and publicly accessible sources without any interaction or intervention.

Search Strategy

This study provides a functional analysis and systematic evaluation of intermittent fasting apps in the mobile phone markets of China on April 1, 2025. Huawei AppGallery, Oppo Software Store, VIVO App Store, and Xiaomi Market are the most widely used Android app markets in China [28]. Thus, we searched the Apple App Store (for iOS apps), Huawei AppGallery, Oppo Software Store, VIVO App Store, and Xiaomi Market (for Android apps).

We conducted keyword searches using "Fasting", "Intermittent Fasting", "Time-Restricted Feeding", "Time-Restricted Fasting", "Time-Restricted Eating" and "Meal Skipping" across each app store in Chinese terms, respectively, without logging into any user accounts. If an app was found in multiple app markets, the version with the highest user score was selected for evaluation.

XSL•FO
RenderX

Eligibility

After screening out duplicate apps, we further screened applications using certain exclusion-inclusion criteria. All apps that met the exclusion-inclusion criteria were downloaded to the test device. Apps that did not function properly due to technical reasons after being downloaded were excluded from the evaluation.

We will include apps related to intermittent fasting, with fasting timer features and those that have been continuously updated since January 1, 2023, and will exclude apps if they (1) have a user score of less than 3 and (2) are not in Chinese.

Data Extraction

For each eligible app, data were extracted using a predefined extraction form. The data collected included general information and features of apps. The platform, user ratings from the app markets, the number of downloads, privacy protection measures, and any evidence-based or professional background supporting the app were noted as general information. All of the retrieved data were then arranged for additional analysis into structured tables. Table 1 shows the data that will be extracted from the apps.

Table . Data that will be extracted from the apps.

Assessment measure	Definition
General information	
Platform	Store platform from which the app originates (eg, Apple App Store, Huawei AppGallery)
User ratings from the app markets	Average rating given by users in the app store (eg, star rating)
Number of downloads	The total number of times the app has been downloaded from the app store
Privacy protection	Applying measures to protect user data privacy
Evidence-based and professional background	Whether the content or methodology of the app is based on scientific evidence or developed by professionals in a relevant field
Features	
App features	Specific features provided by the app (eg,notifications, timer)
Quality	
Engagement	The extent to which the app captures the user's interest and encourages continued use (the uMARS subscale)
Functionality	The effectiveness and efficiency with which the app performs its intended function (the uMARS subscale)
Aesthetic	Visual appeal and design quality of the app, such as color scheme and layout organization (the uMARS subscale)
Information	Accuracy and usefulness of the information provided by the app (the uMARS subscale)

Quality Appraisal of Apps

The uMARS will be used to evaluate the quality of the included apps. It is a widely recognized tool specifically designed for assessing the quality of mobile health apps [19] has been shown to be useful for assessing various aspects of mHealth apps in areas such as chronic diseases [29-31], mental disorders [32], and nutrition [33-35], and has been translated into several languages [20-24]. The uMARS scale consists of 26 items divided into 5 different subscales: engagement, functionality, aesthetics, information, and subjective. The focus of this study was to provide an objective assessment of the quality of the apps and therefore did not include the evaluation of subjective scale. Each item is rated on a 5-point Likert scale. The overall score is the mean of the averages of the 4 objective subscales. The reliability of the two raters' scores will also be assessed.

Before formal scoring, the two raters underwent a uniform training supervised by a senior researcher with more than 5 years of expertise in digital health evaluation and preassessment of the applications and discussed inconsistencies in the results to ensure a common understanding of the criteria for uMARS. During the evaluation process, each app was evaluated independently by both raters, with at least 15 minutes of usage before scoring.

Statistical Analysis

Mean values and SDs were used to describe quantitative variables in the uMARS in the qualitative evaluation. Frequencies and percentages were used to describe classification variables such as the features of the apps. Pearson correlation was calculated to compare the uMARS score with each subscale score, the uMARS score with user ratings, and the user ratings with each subscale score. The intraclass correlation coefficient (ICC) was calculated to measure the consistency of ratings across the subscales (engagement, functionality, aesthetics, and information) and the overall score. The ICC is a commonly used statistical method for assessing the reliability between raters [36]. All statistical analyses were performed using IBM SPSS Statistics 26. All figures were performed using PyCharm 2024.1.

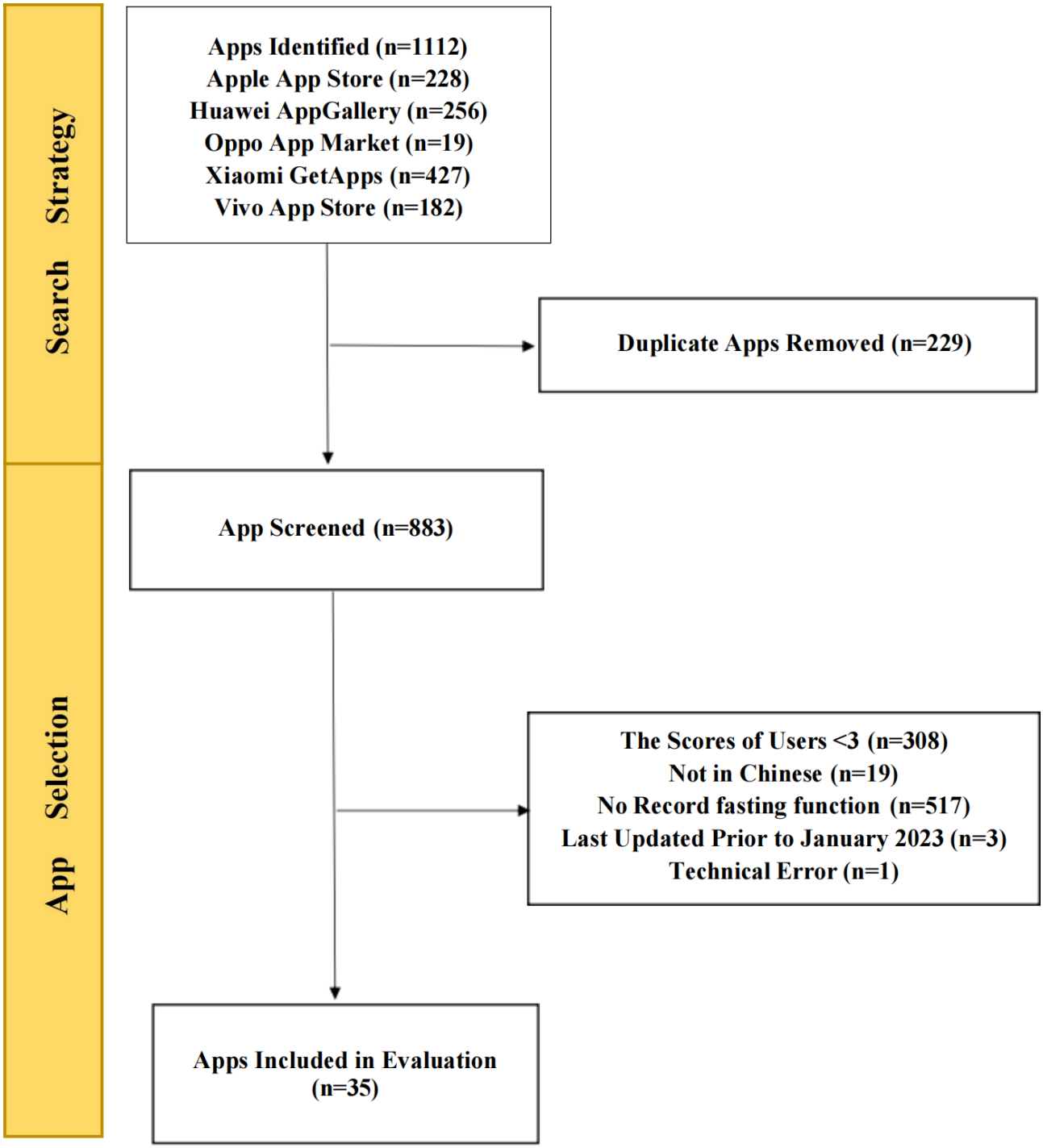
XSL•FO
RenderX

Results

App Selection

A total of 1112 apps were identified from the Apple App Store (n=228), Huawei AppGallery (n=256), Oppo App Market (n=19), Xiaomi GetApp (n=427), and Vivo App Store (n=182). Combining the search results from the 5 app stores, we excluded 229 duplicate apps. A total of 848 apps were excluded on the basis of the exclusion-inclusion criteria. The remaining 35 apps were downloaded for evaluation. Figure 1 shows a flowchart of the intermittent fasting app selection process.

Figure 1. Flowchart for the systematic search and selection of apps.



General Information

Of the 35 apps included in the study, 12 (34%) were from the Apple App Store, 10 (29%) were from Xiaomi GetApp, and 10 (29%) were from the Vivo App Store. Only 2 (6%) apps were from the Huawei AppGallery, and another 1 (3%) app was from the Oppo App Market.

In terms of privacy protection, all 35 (100%) apps have obvious privacy protection measures in place.

Regarding the evidence-based and professional background, 11 (31%) apps have clinical physicians, nutritionists, and dietary experts involved in them. A total of 23 (65%) apps are supported by peer-reviewed academic research, ensuring the reliability of

XSL•FO
RenderX

the content provided, and a total of 28 (80%) apps include tools for quantifying users' health status. Table 2 shows the data that were extracted from the apps by analyzing apps store descriptions, in-app information, and developer statements.

Table . Data that were extracted from the apps.

Assessment measure	N (%)
Platform	
Apple App Store	12 (34)
Huawei AppGallery	2 (6)
Oppo App Market	1 (3)
Xiaomi GetApp	10 (29)
Vivo App Store	10 (29)
Privacy protection	
An obvious privacy protection	35 (100)
No privacy protection	0 (0)
Evidence-based and professional background	
Involvement of clinical physicians, nutritionists, and dietary experts	11 (31)
Peer-reviewed academic research supporting the content's reliability	23 (65)
Tools for quantifying users' health status	28 (80)

Features of the Included Apps

Of the 35 intermittent fasting apps downloaded, 100% (35/35) contain a fasting timer function, 85.71% (30/35) contain a fasting reminder function, and 60% (21/35) contain a provision of fasting tip functions. In addition to the abovementioned core functions, the features of these apps include 4 major categories: recording, calculation, recommendation, and socialization. Among the functions in the recording category are recording weight (34/35, 97.14%), water intake (30/35, 85.71%), fasting experiences (26/35, 74.29%), diet (24/35, 68.57%), exercise (22/35, 62.86%), physical dimensions (17/35, 48.57%), mood (13/35, 37.14%), diary (11/35, 31.43%), menstruation (7/35, 20%), steps (6/35, 17.14%), sleep (3/35, 8.57%), and egestion (2/35, 5.71%). For calculation functions, there are BMI calculation (28/35, 80%) and BMR calculation (25/35, 71.43%). The recommendation function has recipe recommendations (26/35, 74.26%) and sports courses recommendations (4/35, 11.43%), and the social function has a weight loss forum (3/35, 8.57%) and friend connections (4/35, 11.43%). Notably, several apps incorporate artificial intelligence (AI)-driven modules for automated diet and exercise and personalized meal planning. Figure 2 shows the main functionalities and their percentages. For more detailed information on the apps' features, please refer to Multimedia Appendix 1.

XSL•FO
RenderX

Figure 2. Heatmap showing the main functionalities of the intermittent fasting apps (n=35).

XSL•FO
RenderX

App Quality Score

Table 3 presents the final scores, mean values, and SDs for the following: (1) 4 subscales of the uMARS (engagement mean, functionality mean, aesthetic mean, and information mean), (2) app quality mean score (mean of 4 subscales), and (3) user ratings from the app markets.

XSL•FO
RenderX

Table . The user version of the Mobile Application Rating Scale mean scores for intermittent fasting apps[a].

App	Engagement	Functionality	Aesthetic	Information	uMARS[b]	User ratings[c]
Boohee Health	4.90	4.88	4.67	4.63	4.77	4.70
Monster Intermittent Fasting[d]	4.70	4.88	4.33	4.63	4.63	4.70
Lanren Health	4.40	4.63	4.17	4.38	4.39	5.00
Lemon Fasting[d]	4.80	4.88	4.83	4.38	4.72	4.50
Grapefruit Intermittent Fasting[d]	5.00	5.00	5.00	4.88	4.97	4.20
Bigu Intermittent Fasting[d]	4.90	5.00	4.67	4.75	4.83	4.70
Fasting	4.80	5.00	4.50	4.50	4.70	3.80
Gugu Fasting[d]	5.00	4.75	4.83	4.75	4.83	4.90
FastingKeepFit	4.50	4.63	4.00	4.13	4.31	5.00
Health Fasting[d]	4.80	4.75	4.33	4.00	4.47	4.40
Fasting Bigu[d]	4.30	4.50	3.50	4.13	4.11	4.70
168 Intermittent Fasting App	4.80	5.00	4.50	4.50	4.70	4.80
Fasting now[d]	4.50	4.75	3.83	3.63	4.18	4.83
Weightloss Fasting	3.60	4.00	3.33	3.38	3.58	4.80
Tomato Health	4.90	4.88	5.00	5.00	4.94	4.60
Fasting Go[d]	4.90	4.75	4.50	4.50	4.66	4.62
Miaomiao Foods	4.80	4.38	4.83	4.75	4.69	4.81
Xifengyinlu Fasting[d]	3.90	4.75	4.00	3.38	4.01	4.93
Bohe Fasting[d]	4.30	4.88	3.83	3.88	4.22	5.00
Fasting Tracker	3.50	3.63	2.67	2.75	3.14	5.00
Nectarine Fasting[d]	4.60	5.00	4.83	4.50	4.73	4.55
Bianshenxiu[d]	4.70	4.75	4.00	4.50	4.49	5.00
Fastin: Intermittent Fasting	4.60	5.00	4.17	4.38	4.54	4.80
Fasting Tracker	3.20	4.13	2.17	2.38	2.97	4.50
Fasting Weightloss	3.00	3.75	2.17	2.88	2.95	5.00
Daily Fasting[d]	3.80	5.00	3.67	4.00	4.12	4.00
LemonFast	5.00	4.75	5.00	4.63	4.84	3.70
Diet Plan[d]	4.70	4.38	4.00	4.63	4.43	5.00
BodyOK	4.00	4.88	4.33	3.63	4.21	4.90
Fasting Timer[d]	3.80	4.63	4.50	4.25	4.29	4.80
Fasta Fasting Tracker	4.10	4.50	4.50	4.00	4.28	4.80
FastEasy	4.80	4.38	5.00	4.50	4.67	4.30
Shiguang Fasting[d]	4.30	4.63	4.33	4.00	4.31	4.10

XSL•FO
RenderX

App	Engagement	Functionality	Aesthetic	Information	uMARS[b]	User ratings[c]
Window - Fasting Tracker	4.60	4.63	4.17	4.00	4.35	4.00
Tomato Fasting[d]	4.20	4.38	4.33	4.25	4.29	4.70
Means	4.42	4.65	4.19	4.15	4.35	4.63
SD	0.47	0.31	0.64	0.58	0.46	0.36

[a]All items were rated on a 5-point scale from 1=inadequate to 5=excellent.

[b]The user version of the Mobile Application Rating Scale.

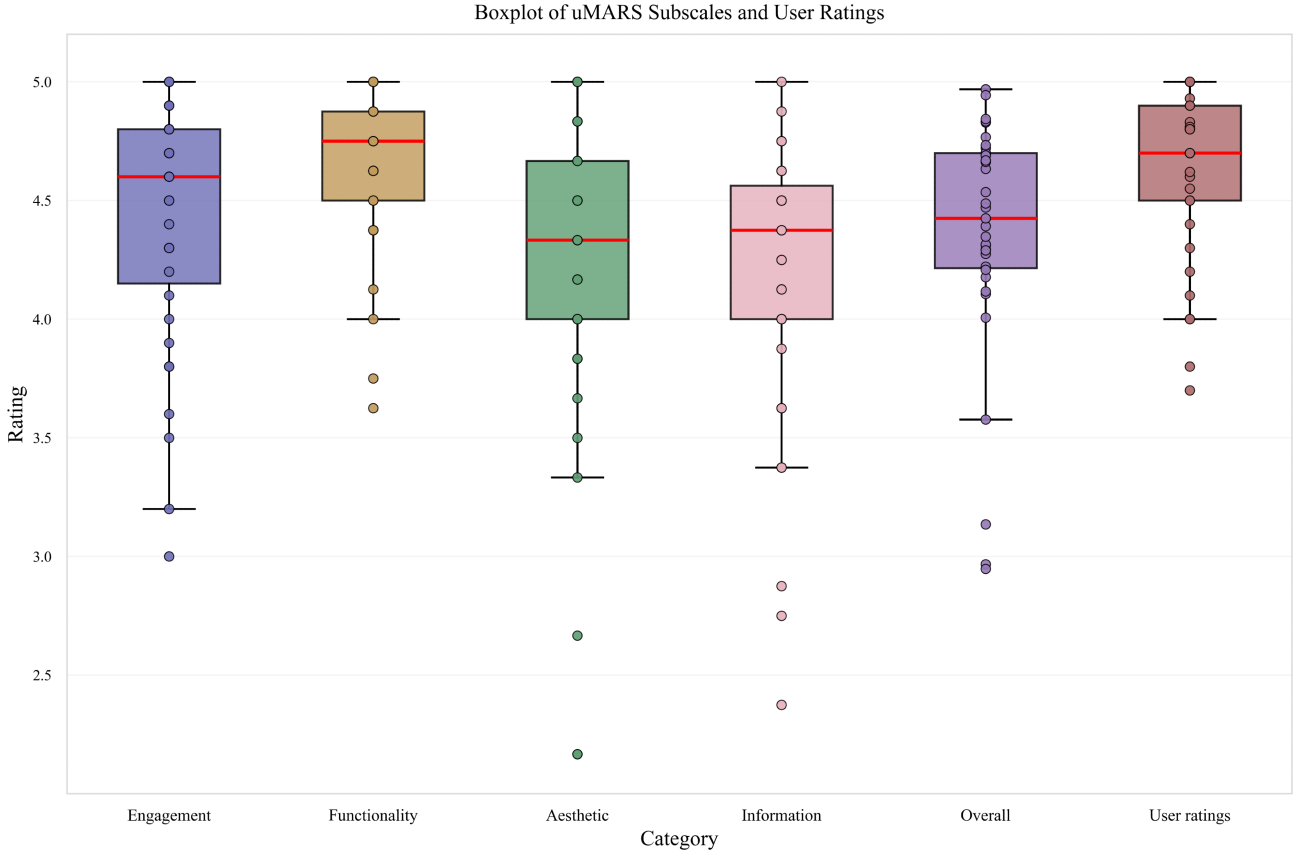
[c]The user ratings from the app markets.

[d]The app does not have an official English name, which is a literal translation in Chinese.

The overall average uMARS score of all apps was 4.35 (SD 0.46), with scores ranging from 2.95 (Fasting Weightloss) to 4.97 (Grapefruit Intermittent Fasting). The reliability of the uMARS is calculated as Cronbach α=.959. The uMARS score of 15 of 35 (42.9%) apps was ≥4.50. There are 16 of 35 (45.7%) apps with a uMARS score between 4.50 and 4.00. Furthermore, 2 of 35 apps (5.7%) had uMARS scores ranging from 3.00 to 3.99. And uMARS scores for 2 of 35 apps (5.7%) fell within the range of 2.90-2.99. There were no apps with a score of <2.90.

The mean scores of each subscale were as follows: engagement quality score=4.42 (SD 0.47), functionality quality score=4.65 (SD 0.31), aesthetic quality score=4.19 (SD 0.64), and information quality score=4.15 (SD 0.58). The aesthetic quality scores showed the greatest span, ranging from a minimum of 2.17 to a maximum of 5.00. The distribution of scores for overall quality and the 4 subscale dimensions is shown in Figure 3.

Figure 3. Graphical representation of the distribution of the uMARS overall, subscale score, and the user ratings. The median, the interquartile distance, and the range were given (n=35). uMARS: the user version of the Mobile Application Rating Scale.



Correlation and Reliability

Table 4 and Figure 4 show the correlation between the uMARS subscales and the overall score and user ratings. The overall uMARS score was significantly positively correlated with the subscale scores (r=0.786 - 0.953, P<.001). However, user ratings of the app market did not correlate with either the overall uMARS score (r=−0.290, P=.091) or the subscale scores (r=-0.305 - -0.207, P=.075-.233).

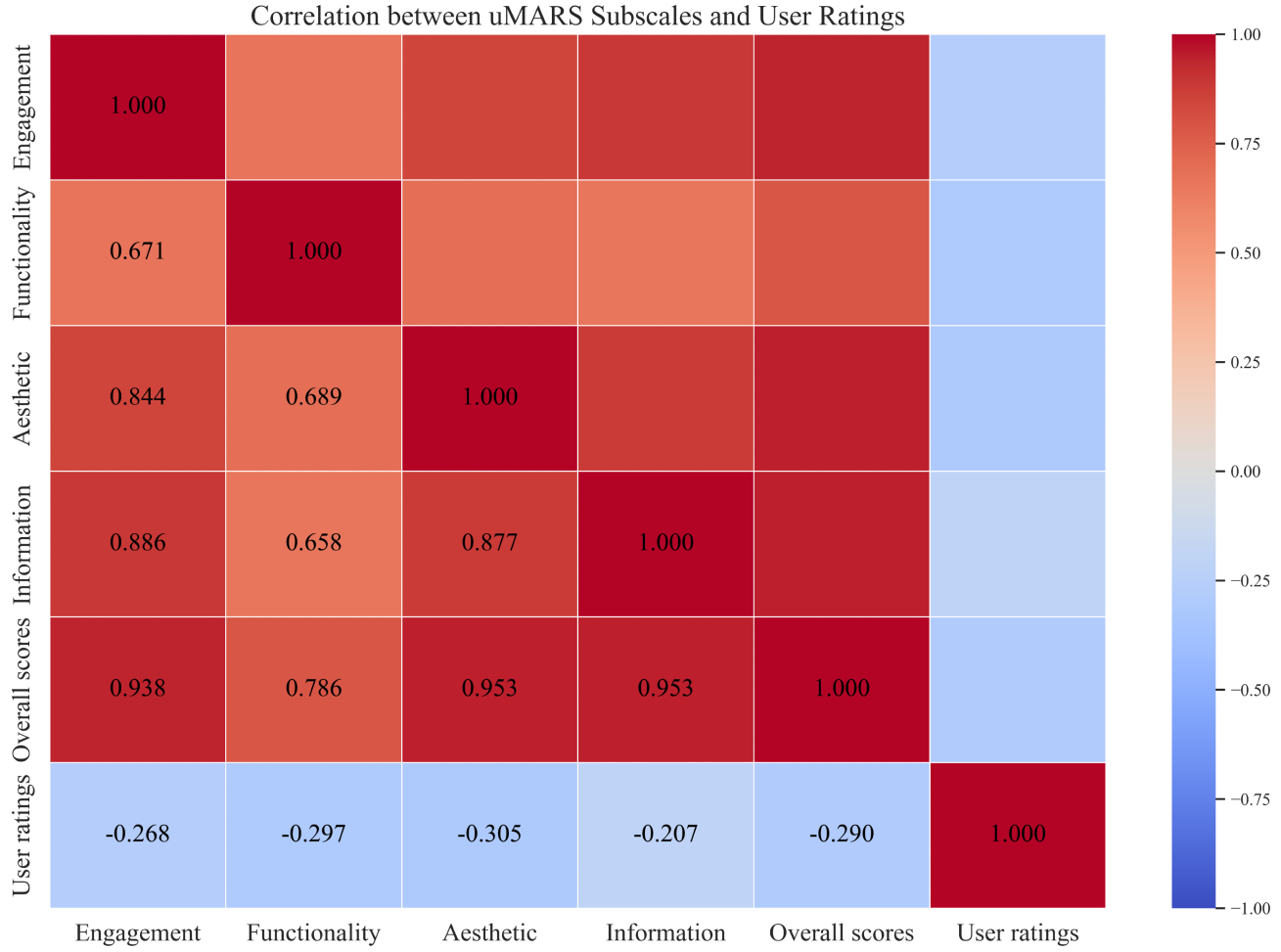
XSL•FO
RenderX

Table . Correlation between the user version of Mobile App Rating Scale subscale and the overall score and user ratings.

Characteristic	Engagement	Functionality	Aesthetic	Information	Overall scores
Engagement					
P value	—[a]				
Functionality					
P value	0.671 (<.001)	—			
Aesthetic					
P value	0.844 (<.001)	0.689 (<.001)	—		
Information					
P value	0.886 (<.001)	0.658 (<.001)	0.877 (<.001)	—	
Overall scores					
P value	0.938 (<.001)	0.786 (<.001)	0.953 (<.001)	0.952 (<.001)	—
User ratings					
P value	−0.268 (.119)	−0.297 (.084)	−0.305 (.075)	−0.207(.233)	−0.290 (.091)

[a]Not applicable

Figure 4. Heatmap showing the correlation between the user version of the Mobile App Rating Scale subscale and the overall score and user ratings.



The overall uMARS score showed high reviewer reliability (ICC 0.909, 95% CI 0.788 - 0.957). Also, all subscales showed good consistency: engagement ICC 0.855 (95% CI 0.715 - 0.927), functionality ICC 0.880 (95% CI 0.758 - 0.940), aesthetics ICC 0.877 (95% CI 0.758 - 0.938), and information ICC 0.809 (95% CI 0.537 - 0.912). Table 5 shows the ICC between the Mobile App Rating Scale subscale and the overall score, as measured by the 2 raters.

XSL•FO
RenderX

Table . Intraclass Correlation Coefficient between the Mobile App Rating Scale subscale and the overall score, as measured by the 2 raters.

	Intraclass correlation[a]	95% CI		F test with true value 0			
		Lower bound	Upper bound	Value	df1	df2	*P* value
Engagement	0.855	0.715	0.927	6.932	34	34	<.001
Functionality	0.880	0.758	0.940	8.969	34	34	<.001
Aesthetic	0.877	0.758	0.938	8.370	34	34	<.001
Information	0.809	0.537	0.912	6.533	34	34	<.001
Overall	0.909	0.788	0.957	13.098	34	34	<.001

[a]Type A intraclass correlation coefficients using an absolute agreement definition.

Discussion

Principal Findings

This study aimed to provide a comprehensive functional analysis and systematic evaluation of intermittent fasting apps available in the Chinese market. We analyzed a total of 35 intermittent fasting apps, examining the various features, privacy protection, evidence-based support, and quality of each app.

Our results showed that of all 35 apps, almost all of them have a fasting timer (35/35, 100%) and the ability to record weight (34/35, 97.14%), and most offered water intake recording (30/35, 85.71%), BMI calculation (28/35, 80%), recipe recommendations (26/35, 74.29%), fasting experience recording (26/35, 74.29%), BMR calculation (25/35, 71.43%), and more. The presence of these features suggests that most apps are able to enable users to achieve their fasting plans by providing guidance and features required for fasting during the user's weight loss phase, such as self-tracking, controlling dietary intake, and improving adherence to self-monitoring eating behaviors, which are important to support the user's health goals [9,37]. Existing research suggests that social media and teammates' influence facilitate the weight management process to a certain extent [38-41]. But the discrepancy is that very few apps (4/35, 11.43%) are able to provide social features. Notably, studies have shown that social features have a significant motivational effect on users with normal or low BMI but may trigger negative emotions in users with a higher BMI [39]. Thus, the low prevalence of social features may reflect developers' considerations regarding user privacy protection and individual differences. Meanwhile, we noticed that some apps introduced AI tools to record diet, exercise, and customize recipes, etc. Some studies have shown that the introduction of AI modules indeed enhances user engagement and opens up new opportunities for apps, but the risks of AI for user privacy and algorithmic errors remain unpredictable to us [42,43].

Privacy protection measures were evident in all apps, reflecting the developers' sensitivity to health data. A study has shown that users are more likely to use apps that clearly state privacy policies and data handling practices [44]. This is especially important for intermittent fasting apps as they collect sensitive data such as weight, eating habits, and fasting times. In addition to these, users are more afraid of receiving more personalized advertisements than the misuse of personal health data [45]. This suggests that the boundaries regarding the extraction of personal information also need to be rethought by developers.

Regarding evidence-based support, most of the apps provide tools to quantify the state of physical health and are backed by peer-reviewed academic research, ensuring the reliability of the content provided [46]. But there is evidence in the literature that app developers lack a comprehensive understanding of study design in translating scientific evidence into practical functionality, ignoring possible side effects or limitations. It is more appropriate for apps to focus on the longevity and completeness of the scientific intervention with applicability to different populations; otherwise, it may lead to misleading user behavior [47-50]. However, only a few apps claimed that their content had been developed with the involvement of clinicians, nutritionists, and dietary experts. This highlights a common flaw in the development process of intermittent fasting apps, which is the lack of the direct involvement of professionals with a systematic validation process, in line with previous literature findings, and such a shortcoming may affect the clinical accuracy and practical application of the information provided [51]. Weight loss applications should emphasize the importance of collaborative development with clinicians, nutritionists, and dietary experts and the need for their involvement early in the development process to ensure that the content is consistent with clinical practice guidelines and evidence-based medical standards [52-55].

The overall average uMARS score for all apps was 4.35 (SD 0.46), indicating high quality, which is consistent with international research performance [56]. The 3 apps with the highest uMARS quality score all received perfect scores for aesthetics, and all of them were designed with some kind of fruit as a specific theme, which could appeal to the user. Unlike traditional pages, several apps designed their interfaces with a cute scene along with cute characters to attract users. In the 3 apps with the lowest scores, both the aesthetic and information scores showed lower scores, which may also suggest a lack of evidence-based content for the apps [57,58]. The above indicates that it is crucial to balance aesthetics and information quality when designing apps [59]. In particular, the functionality subscale score of 4.65 (SD 0.31) is in line with the results of some previous literature, which found that the effectiveness of mobile apps in weight management is closely related to the quality of their design and that mobile app interventions can significantly enhance users' physical activity levels and weight loss [60,61].

The various subscales in this study showed consistency with the total scale, and there is observational research that further

XSL•FO
RenderX

suggests that frequency of app use and engagement are important factors in weight management success, which coincides with the positive correlation between engagement and overall score found in this study. These consistencies suggest that high-quality functional design and user engagement are key to making intermittent fasting apps work [62].

There was no significant correlation between user ratings in the app market and uMARS scores, which is consistent with existing research that suggests that user ratings may not always reflect the actual quality of the apps assessed by the standardized tool [25,63]. They pointed out that scores from app markets are usually based on the subjective experience of users and may be influenced by user preferences, advertisements, and other factors. A meta-analytic study showed that influencer marketing significantly affects consumer attitudes and behavioral engagement [64], and research has also shown that apps improve ratings by improving the user experience through fun or customization [65]. Also, developers may use bots or puppet accounts to post fake reviews [66]. These behaviors may result in high user ratings but not actual quality, and many user reviews are of low quality and hardly reflect the actual quality of the app [67]. Whereas the uMARS is an objective, professional assessment tool, this may lead to inconsistencies between user ratings and actual app quality. Similarly, a randomized controlled trial showed that app-based multimodal interventions were effective for weight management, but user ratings were not associated with the actual effects [68]. Meanwhile, our study found that the selected high user-rated health apps exhibit 2 key characteristics: systematic assessment scores on the uMARS are generally lower than consumer-side user ratings, and there is a notable lack of involvement of clinical medical experts in the product development process. This phenomenon echoes the findings of the Spanish language market study, which similarly found that weight loss apps, while achieving high user satisfaction, had a disconnect between their functional design and their scientific evidence base. Notably, both sets of studies revealed a common problem of existing apps in the market failing to effectively integrate specialized medical resources [69].

Limitations and Perspectives

There are several limitations to consider. First, the keywords used in the app store searches may not have captured all relevant intermittent fasting apps as some may use different terms or be categorized differently, resulting in their exclusion from this study. Second, the app market is constantly evolving, with new apps being introduced and existing apps being updated or removed. The cross-sectional nature of this study means it provides a snapshot in time that may soon become outdated as the app landscape changes. Third, this study focused on the functionality and quality of the apps but did not measure long-term user engagement, adherence, or compliance with the apps, which are key factors in the success of a health intervention. Additionally, this study did not assess the actual effectiveness of the apps in promoting weight loss or improving metabolic health. The existence of a privacy policy does not necessarily guarantee that data protection measures are effectively enforced. Furthermore, the study evaluated only the free version of some of the apps due to lack of funding, and paying to provide additional functionality may have yielded different results. Last but not least, the subjective scale component of the uMARS was not included in this study when assessing app quality. Since subjective scales usually reflect users' personal experiences and feelings, which may vary greatly depending on individual backgrounds, cultural differences, or other external factors. Meanwhile, the focus of the study is to provide an objective assessment of the functionality and quality of the applications, and the subjective feelings component may not directly affect the core objective of the study. Therefore, to ensure the objectivity of the data and the consistency of the assessment results, this study focuses on the objective dimensions of engagement, functionality, information, and aesthetic design of the apps.

Subsequent research will explore the long-term impact of using intermittent fasting apps on health outcomes, in addition to investigating user experience and engagement to understand the needs and preferences of different user groups to provide more targeted and effective information for development. Meanwhile, this research team is further expanding the evaluation team in upcoming studies to increase the data value of the study.

Conclusions

This study identified 35 intermittent fasting apps in Chinese App Stores, analyzed the distribution of their main features, and assessed their content and quality. Most of the apps fulfill the basic requirements of intermittent fasting and are of high quality overall, but there are still gaps in terms of professional engagement and social features as well. Based on these findings, we recommend that developers consider engaging more health care professionals early in the app development process to ensure the content is more evidence-based and accurate. More evidence-based content will help enhance the informational quality of the apps, addressing the gaps identified in some lower-scoring apps. Furthermore, adding social features, such as community forums or friend connections, could increase user engagement and provide additional support, which has been shown to aid in weight management. By focusing on these areas, developers can create more comprehensive and user-friendly intermittent fasting apps that promote better adherence and outcomes for users. However, enhanced user social engagement should be accompanied by good user privacy protection to avoid counterproductive results. AI-driven tools have the potential to enhance user personalization, but they need to be used in reasonable moderation.

Acknowledgments

The authors would like to sincerely appreciate the support of Chongqing Medical University for this study. The authors would also like to thank seniors Lyu, Liu, and Zhou for their guidance in writing.

XSL•FO
RenderX

This study was supported by the Chongqing Science and Technology Commission (No. CSTB2025NSCQ-GPX1200), the Chongqing Social Science Planning Cultivation Project (No. 2025PY23), and the Intelligent Medicine Young Talents Key Research Project of Chongqing Medical University (No. ZHYXQNRC202203).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Main functionalities of the fasting apps (n=35).
[DOCX File, 13 KB - mhealth_v13i1e66339_app1.docx]

References

1. Caballero B. Humans against obesity: who will win? Adv Nutr 2019 Jan 1;10(suppl_1):S4-S9. [doi: 10.1093/advances/nmy055] [Medline: 30721956]
2. Wang Y, Zhao L, Gao L, Pan A, Xue H. Health policy and public health implications of obesity in China. The Lancet Diabetes & Endocrinology 2021 Jul;9(7):446-461. [doi: 10.1016/S2213-8587(21)00118-2]
3. Emerging Risk Factors Collaboration, Wormser D, Kaptoge S, et al. Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. Lancet 2011 Mar 26;377(9771):1085-1095. [doi: 10.1016/S0140-6736(11)60105-0] [Medline: 21397319]
4. Prospective Studies Collaboration, Whitlock G, Lewington S, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. Lancet 2009 Mar 28;373(9669):1083-1096. [doi: 10.1016/S0140-6736(09)60318-4] [Medline: 19299006]
5. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. The Lancet 2008 Feb;371(9612):569-578. [doi: 10.1016/S0140-6736(08)60269-X]
6. Ni MC, Rodgers A, Pan WH, Gu DF, Woodward M. Body mass index and cardiovascular disease in the Asia-Pacific Region: an overview of 33 cohorts involving 310 000 participants. Int J Epidemiol 2004;33(4):751-758. [doi: 10.1093/ije/dyh163]
7. Chao AM, Quigley KM, Wadden TA. Dietary interventions for obesity: clinical and mechanistic findings. J Clin Invest 2021 Jan 4;131(1):e140065. [doi: 10.1172/JCI140065] [Medline: 33393504]
8. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
9. Varady KA, Cienfuegos S, Ezpeleta M, Gabel K. Clinical application of intermittent fasting for weight loss: progress and future directions. Nat Rev Endocrinol 2022 May;18(5):309-321. [doi: 10.1038/s41574-022-00638-x] [Medline: 35194176]
10. Vasim I, Majeed CN, DeBoer MD. Intermittent fasting and metabolic health. Nutrients 2022 Jan 31;14(3):631. [doi: 10.3390/nu14030631] [Medline: 35276989]
11. Khalafi M, Maleki AH, Ehsanifar M, Symonds ME, Rosenkranz SK. Longer‐term effects of intermittent fasting on body composition and cardiometabolic health in adults with overweight and obesity: a systematic review and meta‐analysis. Obes Rev 2025 Feb;26(2):e13855. [doi: 10.1111/obr.13855]
12. Cheung K, Chan V, Chan S, et al. Effect of intermittent fasting on cardiometabolic health in the Chinese population: a meta-analysis of randomized controlled trials. Nutrients 2024 Jan 25;16(3):357. [doi: 10.3390/nu16030357] [Medline: 38337642]
13. Baum Martinez I, Peters B, Schwarz J, et al. Validation of a smartphone application for the assessment of dietary compliance in an intermittent fasting trial. Nutrients 2022 Sep 7;14(18):3697. [doi: 10.3390/nu14183697] [Medline: 36145073]
14. Torres L, Lee JL, Park S, et al. Retention, fasting patterns, and weight loss with an intermittent fasting app: large-scale, 52-week observational study. JMIR Mhealth Uhealth 2022 Oct 4;10(10):e35896. [doi: 10.2196/35896] [Medline: 36194463]
15. Martin-Vicario L, Bustos Díaz J, Nicolas-Sans R. Weight loss app descriptors in app stores: a qualitative analysis of the Spanish market. Human Behavior and Emerging Technologies 2023 Nov 9;2023(1):1-7. [doi: 10.1155/2023/4104229]
16. Selvaraj SN, Sriram A. The quality of indian obesity-related mHealth apps: PRECEDE-PROCEED model-based content analysis. JMIR Mhealth Uhealth 2022 May 11;10(5):e15719. [doi: 10.2196/15719] [Medline: 35544318]
17. Milne-Ives M, Lam C, van Velthoven M, Meinert E. Mobile fitness and weight management apps: protocol for a quality evaluation. JMIR Res Protoc 2020 Sep 24;9(9):e17685. [doi: 10.2196/17685] [Medline: 32969830]
18. Alshathri DM, Alhumaimeedy AS, Al-Hudhud G, Alsaleh A, Al-Musharaf S, Aljuraiban GS. Weight management apps in Saudi Arabia: evaluation of features and quality. JMIR Mhealth Uhealth 2020 Oct 26;8(10):e19844. [doi: 10.2196/19844] [Medline: 33104013]
19. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]

XSL•FO
RenderX

20. Vasconcelos JM, Viana YG, Duarte NP, Polese JC. Translation and cross-cultural adaptation into Portuguese-Brazil of NuHISS, Enlight, and uMARS. J Bodyw Mov Ther 2024 Apr;38:437-448. [doi: 10.1016/j.jbmt.2024.01.009] [Medline: 38763590]
21. Chasiotis G, Stoyanov SR, Karatzas A, Gravas S. Greek validation of the user version of the mobile application rating scale (uMARS). J Int Med Res 2023 Mar;51(3):655745155. [doi: 10.1177/03000605231161213]
22. Martin-Payo R, Carrasco-Santos S, Cuesta M, Stoyan S, Gonzalez-Mendez X, Fernandez-Alvarez MDM. Spanish adaptation and validation of the user version of the Mobile Application Rating Scale (uMARS). J Am Med Inform Assoc 2021 Nov 25;28(12):2681-2686. [doi: 10.1093/jamia/ocab216]
23. Morselli S, Sebastianelli A, Domnich A, et al. Translation and validation of the Italian version of the user version of the Mobile Application Rating Scale (uMARS). J Prev Med Hyg 2021 Mar;62(1):E243-E248. [doi: 10.15167/2421-4248/jpmh2021.62.1.1894] [Medline: 34322643]
24. Shinohara Y, Yamamoto K, Ito M, et al. Development and validation of the Japanese version of the uMARS (user version of the mobile app rating system). Int J Med Inform 2022 Sep;165:104809. [doi: 10.1016/j.ijmedinf.2022.104809]
25. Ko S, Lee J, An D, Woo H. Menstrual tracking mobile app review by consumers and health care providers: quality evaluations study. JMIR Mhealth Uhealth 2023 Mar 1;11:e40921. [doi: 10.2196/40921] [Medline: 36857125]
26. Notice on incorporating the healthy weight management action and three other actions into the healthy china action. National Health Commission of the People's Republic of China. URL: https://www.nhc.gov.cn/guihuaxxs/c100133/202504/b5d296bc536e42aa95baead98e792c67.shtml [accessed 2025-04-15]
27. Chonqing medical university. URL: https://kyc.cqmu.edu.cn/ [accessed 2025-10-07]
28. AppInChina. The AppInChina app store index. URL: https://www.appinchina.co/market/app-stores/ [accessed 2023-11-11]
29. Welch WA, Solk P, Auster-Gussman L, et al. User-centered development of a smartphone application (Fit2Thrive) to promote physical activity in breast cancer survivors. Transl Behav Med 2022 Feb 16;12(2):203-213. [doi: 10.1093/tbm/ibab112]
30. Lambrecht A, Vuillerme N, Raab C, et al. Quality of a supporting mobile app for rheumatic patients: patient-based assessment using the user version of the Mobile Application Scale (uMARS). Front Med 2021;8:715345. [doi: 10.3389/fmed.2021.715345]
31. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. 'Prostate cancer risk calculator' mobile applications (apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
32. Strodl E, Shakespeare-Finch J, Alichniewicz KK, et al. Clinicians' perceptions of PTSD Coach Australia. Internet Interv 2020 Sep;21:100333. [doi: 10.1016/j.invent.2020.100333] [Medline: 32939341]
33. Wu Y, Wang X, Gao F, Liao J, Zeng J, Fan L. Mobile nutrition and health management platform for perioperative recovery: an interdisciplinary research achievement using WeChat Applet. Front Med 2023;10. [doi: 10.3389/fmed.2023.1201866]
34. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: user-centered study with employees from a Lebanese University. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e9836. [doi: 10.2196/mhealth.9836] [Medline: 30672742]
35. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the China app store: assessment of functionality and quality. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e13261. [doi: 10.2196/13261] [Medline: 31364606]
36. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
37. Azar KMJ, Lesser LI, Laing BY, et al. Mobile applications for weight management. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005]
38. Petkovic J, Duench S, Trawin J, et al. Behavioural interventions delivered through interactive social media for health behaviour change, health outcomes, and health equity in the adult population. Cochrane Database Syst Rev 2021 May 31;5(5):CD012932. [doi: 10.1002/14651858.CD012932.pub2] [Medline: 34057201]
39. Laranjo L, Quiroz JC, Tong HL, Arevalo Bazalar M, Coiera E. A mobile social networking app for weight management and physical activity promotion: results from an experimental mixed methods study. J Med Internet Res 2020 Dec 8;22(12):e19991. [doi: 10.2196/19991] [Medline: 33289670]
40. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294. [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]
41. Leahey TM, Kumar R, Weinberg BM, Wing RR. Teammates and social influence affect weight loss outcomes in a team ‐ based weight loss competition. Obesity (Silver Spring) 2012 Jul;20(7):1413-1418. [doi: 10.1038/oby.2012.18]
42. Lee D, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges. IJERPH 2021;18(1):271. [doi: 10.3390/ijerph18010271]
43. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7]
44. Chauhan GK, Vavken P, Jacob C. Mobile apps and wearable devices for cardiovascular health: narrative review. JMIR Mhealth Uhealth 2025 Apr 4;13:e65782. [doi: 10.2196/65782] [Medline: 40184552]

XSL•FO
RenderX

45. Schroeder T, Haug M, Gewald H. Data privacy concerns using mHealth apps and smart speakers: comparative interview study among mature adults. JMIR Form Res 2022 Jun 13;6(6):e28025. [doi: 10.2196/28025] [Medline: 35699993]
46. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile health apps for self-management of rheumatic and musculoskeletal diseases: systematic literature review. JMIR Mhealth Uhealth 2019 Nov 26;7(11):e14730. [doi: 10.2196/14730] [Medline: 31769758]
47. Hahn SL, Hazzard VM, Loth KA, Larson N, Klein L, Neumark-Sztainer D. Using apps to self-monitor diet and physical activity is linked to greater use of disordered eating behaviors among emerging adults. Prev Med 2022 Feb;155:106967. [doi: 10.1016/j.ypmed.2022.106967] [Medline: 35065981]
48. Holzmann SL, Holzapfel C. A scientific overview of smartphone applications and electronic devices for weight management in adults. J Pers Med 2019 Jun 7;9(2):31. [doi: 10.3390/jpm9020031] [Medline: 31181705]
49. Ghelani DP, Moran LJ, Johnson C, Mousa A, Naderpoor N. Mobile apps for weight management: a review of the latest evidence to inform practice. Front Endocrinol (Lausanne) 2020;11:412. [doi: 10.3389/fendo.2020.00412] [Medline: 32670197]
50. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903. [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
51. Najm A, Nikiphorou E, Kostine M, et al. EULAR points to consider for the development, evaluation and implementation of mobile health applications aiding self-management in people living with rheumatic and musculoskeletal diseases. RMD Open 2019;5(2):e001014. [doi: 10.1136/rmdopen-2019-001014] [Medline: 31565245]
52. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. Designing health apps to support dietetic professional practice and their patients: qualitative results from an international survey. JMIR Mhealth Uhealth 2017 Mar 31;5(3):e40. [doi: 10.2196/mhealth.6945] [Medline: 28363882]
53. Choi J, Chung C, Woo H. Diet-related mobile apps to promote healthy eating and proper nutrition: a content analysis and quality assessment. IJERPH 2021;18(7):3496. [doi: 10.3390/ijerph18073496]
54. Bakker CJ, Wyatt TH, Breth MC, et al. Nurses' roles in mHealth app development: scoping review. JMIR Nurs 2023 Oct 17;6:e46058. [doi: 10.2196/46058] [Medline: 37847533]
55. Tang T, Lim ME, Mansfield E, McLachlan A, Quan SD. Clinician user involvement in the real world: designing an electronic tool to improve interprofessional communication and collaboration in a hospital setting. Int J Med Inform 2018 Feb;110:90-97. [doi: 10.1016/j.ijmedinf.2017.11.011] [Medline: 29331258]
56. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903. [doi: 10.1016/j.amepre.2018.12.005]
57. Cowan LT, Van Wagenen SA, Brown BA, et al. Apps of steel: are exercise apps providing consumers with realistic expectations? Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126]
58. Breton ER, Fuemmeler BF, Abroms LC. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? Transl Behav Med 2011 Dec;1(4):523-529. [doi: 10.1007/s13142-011-0076-5] [Medline: 24073074]
59. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Dec;13(1):35. [doi: 10.1186/s12966-016-0359-9]
60. Ufholz K, Werner J. The efficacy of mobile applications for weight loss. Curr Cardiovasc Risk Rep 2023;17(4):83-90. [doi: 10.1007/s12170-023-00717-2] [Medline: 36974130]
61. Islam MM, Poly TN, Walther BA, Jack Li YC. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039. [doi: 10.2196/17039] [Medline: 32706724]
62. Huntriss R, Salimgaraev R, Nikogosov D, Powell J, Varady KA. The effectiveness of mobile app usage in facilitating weight loss: an observational study. Obes Sci Pract 2024 Jun;10(3):e757. [doi: 10.1002/osp4.757] [Medline: 38745944]
63. Wu X, Xu L, Li P, Tang T, Huang C. Multipurpose mobile apps for mental health in Chinese app stores: content analysis and quality evaluation. JMIR Mhealth Uhealth 2022 Jan 4;10(1):e34054. [doi: 10.2196/34054] [Medline: 34982717]
64. Pan M, Blut M, Ghiassaleh A, Lee ZWY. Influencer marketing effectiveness: a meta-analytic review. J of the Acad Mark Sci 2025 Jan;53(1):52-78. [doi: 10.1007/s11747-024-01052-7]
65. Kim E, Lin JS, Sung Y. To app or not to app: engaging consumers via branded mobile apps. Journal of Interactive Advertising 2013 Jan;13(1):53-65. [doi: 10.1080/15252019.2013.782780]
66. Genc-Nayebi N, Abran A. A systematic literature review: opinion mining studies from mobile app store user reviews. Journal of Systems and Software 2017 Mar;125:207-219. [doi: 10.1016/j.jss.2016.11.027]
67. Sadiq S, Umer M, Ullah S, Mirjalili S, Rupapara V, Nappi M. Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning. Expert Syst Appl 2021 Nov;181:115111. [doi: 10.1016/j.eswa.2021.115111]
68. Gemesi K, Winkler S, Schmidt-Tesch S, Schederecker F, Hauner H, Holzapfel C. Efficacy of an app-based multimodal lifestyle intervention on body weight in persons with obesity: results from a randomized controlled trial. Int J Obes (Lond) 2024 Jan;48(1):118-126. [doi: 10.1038/s41366-023-01415-0] [Medline: 38017117]

XSL•FO
RenderX

69. Martin-Vicario L, Bustos Díaz J, Martínez-Sánchez ME, Nicolas-Sans R. Mobile applications for weight-loss in the Spanish-speaking market: usability and engagement. Obesity Medicine 2023 Aug;41:100499. [doi: 10.1016/j.obmed.2023.100499]

Abbreviations

AI: artificial intelligence
BMR: basal metabolic rate
ICC: intraclass correlation coefficient
the uMARS: the user version of the Mobile Application Rating Scale

Edited by L Buis; submitted 13.09.24; peer-reviewed by R Nicolas-Sans, R Pryss; revised version received 15.04.25; accepted 12.09.25; published 07.10.25.

Please cite as:
Fang L, Huang C, Lin B, Lei K, Zhou J, Zhong X, Liu Y, Liu J
Evaluation of Mobile Intermittent Fasting Applications in Chinese App Stores: Quality Evaluations and Content Analysis
JMIR Mhealth Uhealth 2025;13:e66339
URL: https://mhealth.jmir.org/2025/1/e66339
doi:10.2196/66339

© Laihao Fang, Cheng Huang, Bing Lin, Kuanlin Lei, Jiazhen Zhou, Xiaoni Zhong, Yanbing Liu, Jiaxiu Liu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 7.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Apps for HIV and Sexually Transmitted Infection Prevention in Canada, Mexico, and the United States: Environmental Scan

Higinio Fernández-Sánchez[1], PhD, RN; Javier Salazar-Alberto[2], MSc; Jhan Carlos Manuel Fernández-Delgado[3], BSN; Annalynn Galvin[4], PhD; Michael J Mugavero[5], PhD; Carlos E Rodriguez-Diaz[6], PhD; Diane Santa Maria[4], PhD

[1]Cizik School of Nursing, The University of Texas Health Science Center Houston, 6901 Bertner Avenue, Houston, TX, United States

[2]National Institute of Public Health, Estado de Mexico, Mexico

[3]National University of Cajamarca, Jaén, Peru

[4]The University of Texas Health Science Center Houston, Houston, TX, United States

[5]University of Alabama at Birmingham, Birmingham, AL, United States

[6]Boston University School of Public Health, Boston, MA, United States

Corresponding Author:
Higinio Fernández-Sánchez, PhD, RN
Cizik School of Nursing, The University of Texas Health Science Center Houston, 6901 Bertner Avenue, Houston, TX, United States

Abstract

Background: Canada, Mexico, and the United States are primary transit destinations for migrants in the Western Hemisphere. Migrants face barriers to accessing health services, including HIV and AIDS and sexually transmitted infection (STI) prevention. Mobile apps may enhance public health access for these populations.

Objective: This study aims to systematically identify and evaluate mobile apps supporting HIV and STI prevention in Canada, Mexico, and the United States.

Methods: An environmental scan of 357 mobile apps from the Google Play and Apple App stores was conducted on June 18, 2024, following the rigorous 6-step framework proposed by Fernández-Sánchez to ensure a systematic and comprehensive evaluation of apps for HIV and STI prevention. Predefined inclusion and exclusion criteria were applied, resulting in 6 eligible apps. Each app was assessed using the 29-item Mobile App Rating Scale (MARS), scored on a 5-point Likert scale (1=inadequate, 5=excellent), and categorized as high (3), medium (2), or low (1) based on mean scores. Internal consistency was excellent (Cronbach α=0.90), and interrater reliability demonstrated near-perfect agreement (Cohen κ=0.862). Data analyses were performed using SPSS (version 27; IBM Corp).

Results: All 6 apps were available in Canada, Mexico, and the United States, with 33.3% (2/6) from Google Play, 16.7% (1/6) from Apple, and 50% (3/6) from both platforms. MARS evaluation revealed high quality ratings for engagement (83.0%), functionality (88.9%), aesthetics (83.3%), and information quality (100%), as well as high subjective quality (83.3%) and app-specific quality (88.9%). Life4Me+ was the highest-rated app (4.6), while HIV-TEST received the lowest rating (3.4). Most apps (5/6, 83.3%) were only available in English, and 16.7% (1/6) supported multiple languages, which may limit accessibility for non–English-speaking migrant populations. In addition, 83.3% (5/6) were updated in 2024, 33.3% (2/6) were linked to nongovernmental organization, 16.7% (1/6) to a university, and 50% (3/6) had no clear affiliation. Regarding their focus, 50% (3/6) addressed STI prevention, diagnosis, and treatment, 16.7% (1/6) combined HIV and STI prevention, and 33.3% (2/6) provided pre-exposure prophylaxis–related resources.

Conclusions: These 6 apps stand out for their high functionality, engagement, and accessibility, establishing themselves as effective tools for HIV and STI prevention education among migrant populations. This study highlights the critical role of digital resources in addressing public health challenges faced by vulnerable and minority groups. Integrating these apps into health promotion strategies is essential to improve health literacy and encourage preventive behaviors. Moreover, ensuring the quality, credibility, linguistic diversity, and continuous updating of these digital interventions is crucial to achieving a real and sustained impact on public health. Policies should promote clear standards that guarantee accessibility, transparency, and accuracy, thereby facilitating access to health care services in complex migratory contexts.

(JMIR Mhealth Uhealth 2025;13:e72009) doi:10.2196/72009

KEYWORDS

mobile health; mobile app; sexual health; HIV; MARS; Mobile App Rating Scale; STI; sexually transmitted infections

XSL•FO
RenderX

Introduction

Background

Migration to North America is one of the most active migration routes globally [1]. Latin America plays a central role in this dynamic, as Mexico is the largest source of migrants to the United States [2,3]. Other Latin American nations also contribute to migration flows toward Canada, the United States, and Mexico, as these regions serve as key transit and destination points for diverse migrant populations. While the United States remains the primary destination owing to perceived economic and social opportunities [4], increasingly restrictive immigration policies have led to heightened detentions and deportations. Canada has emerged as an alternative destination, but accessibility challenges and strict migration controls complicate pathways, compelling migrants to adapt their routes based on shifting opportunities and policy restrictions [5].

This evolving migration landscape introduces critical public health challenges, as migrants face substantial barriers to accessing health care and are at increased risk for infectious diseases, including HIV and sexually transmitted infections (STIs) [6]. Without adequate prevention and treatment, HIV progresses to AIDS, significantly impacting individual health, quality of life, and the capacity of health care systems in host countries [7]. For migrants, vulnerabilities such as undocumented status, limited health care access, and cultural stigmas exacerbate their susceptibility to HIV and STIs and their negative outcomes [8-10].

The socioeconomic conditions of many migrants, characterized by poverty, instability, and precarious living arrangements, further heighten their exposure to risk factors, including sexual violence, human trafficking, and substance and alcohol use [11-13]. These vulnerabilities not only undermine individual health but also contribute to the transmission of HIV, with significant public health implications for both migrant communities and their host countries [14,15]. The compounded risks underscore the urgent need for effective HIV and STI prevention strategies that can reach these transient populations.

Innovations such as pre-exposure prophylaxis (PrEP), nonoccupational postexposure prophylaxis, postexposure prophylaxis, doxycycline postexposure prophylaxis, and self-testing kits have revolutionized HIV and STI prevention. However, migrants often encounter barriers that limit their access to these tools, including linguistic and cultural challenges [16]. Mobile health (mHealth) apps have emerged as promising platforms to address these barriers, providing accessible, discreet, and culturally relevant health information [17]. Evidence suggests that mHealth apps dedicated to prevention can enhance access to information, especially for minoritized populations such as migrants and others, racially and ethnically diverse groups, as found in recent randomized controlled trials of HIV prevention apps [18-21]. These apps facilitate equitable health care access by offering content in native languages and by integrating functionalities such as testing reminders and treatment adherence tracking [22].

Despite the potential of these apps, ensuring the quality and reliability of the information they provide remains challenging, particularly among migrant populations [23]. Current app stores, such as Google Play and Apple App Store, host numerous health-related apps. Yet only some are designed to meet the specific needs of racially and ethnically diverse populations. For instance, a review of 55 HIV and STI prevention apps found that only 9 catered to racialized or ethnically diverse groups [24]. This limited scope highlights the pressing need to evaluate app quality, participant acceptability, and language appropriateness to ensure these tools effectively support prevention efforts for marginalized groups.

Objectives

To address this gap, this study used an environmental scan (E-scan) methodology to systematically identify and evaluate mobile apps supporting HIV and STI prevention in Canada, Mexico, and the United States. The E-scan methodology enables a comprehensive analysis of app features, accessibility, and information quality, offering a structured and innovative approach to app evaluation [25]. Given the lack of regulation in health app content [26], this study fills a critical void by evaluating the quality of digital resources tailored to migrant populations. Such research can help establish a foundation for leveraging mHealth tools to enhance HIV and STI prevention across North America, bridging the gap between technological innovation and public health needs. The objective of this study was to systematically identify and evaluate mobile apps that support HIV and STI prevention in Canada, Mexico, and the United States.

Methods

Study Design

We conducted an environmental scan (E-scan) following the 6 steps proposed by Fernández-Sánchez et al [25], to systematically identify, screen, and analyze mobile apps available for HIV and STI prevention as of June 18, 2024. This method was combined with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines to document the identification, screening, eligibility, and inclusion process through a standardized flow diagram.

Step I: Research Questions

Our research focused on the following overarching question: What is the quality (ie, engagement, functionality, aesthetics, and information quality) of mobile apps designed for HIV and STI prevention in Canada, Mexico, and the United States? These criteria were selected based on the Mobile App Rating Scale (MARS) [27]. In addition, the study addressed several sub-questions to provide a comprehensive evaluation of the mobile apps: (1) What is the quality of these apps according to user criteria (ie, app subjective quality)? (2) Do these apps allow for user input, feedback, and other prompts (ie, app interactivity)? and (3) How accurate is the HIV and STI prevention information on the app (ie, app-specific quality)? Other considerations included the number of languages each app supports, the countries where they are available, the target age range, and the year of their most recent update. Combined,

XSL•FO
RenderX

these aspects provided a broader understanding of the apps' usability, accessibility, and relevance to diverse user needs.

Step II: Selection Criteria

To answer these research questions, we included mobile apps that were (1) related to HIV and STI prevention and (2) available in Canada, the United States, and Mexico. To broaden the scope of the search, we did not exclude apps-based release date, cost, number of downloads, or the digital medium. We excluded apps developed for health care professionals and not for a public audience.

Step III: Search Strategy and Data Sources

We thoroughly searched the Google Play Store and Apple App Store by accessing the platforms in all 3 countries. The search process was developed following the PRISMA 2020 guidelines to ensure methodological rigor, transparency, and reproducibility to ensure organized selection and a clear summary of findings [28]. Overall, 3 team members created accounts in the Google Play Store and Apple App Store in Canada, Mexico, and the United States. Apps were identified from May 23 to June 18, 2024, using mobile devices, computers, and iPads with Android and iOS operating systems. A manual search was also conducted on Google to identify potential complementary apps. Keywords in Spanish, English, and French were based on recommendations from similar articles and the authors' experience. These keywords included: "HIV," "AIDS," "PrEP," "human immunodeficiency virus," "acquired immunodeficiency syndrome," "HIV prevention," "AIDS prevention," "HIV education," "HIV/AIDS," "HIV treatment," "HIV app," "HIV health," "STIs," "sexually transmitted infection," "STI prevention," "STI app," "sexual health," "STI education," and "STI treatment." Boolean operators (AND, OR) were used to combine keywords and maximize search coverage effectively. The first 100 results of each keyword combination were reviewed, considering that limited time and few human resources were available.

Step IV: Data Extraction

Data were extracted via Microsoft Excel. The research team designed an Excel table aligned with the purpose of this E-scan to collect relevant information about the apps, including the app name, developer contact information, digital store type, rating, number of downloads, dates of release and last update, version, cost, focus and brief description, theoretical background or strategies, affiliation, target age group, and technical aspects. Data extraction was independently performed by 2 reviewers from the team (JS-A and JCMF-D). Discrepancies were resolved by consensus, and a third reviewer, the principal investigator, validated the data.

Step V: Quality Assessment

The selected apps underwent a quality assessment using the MARS, developed by Stoyanov et al [29]. The MARS is a multidimensional, reliable, and customizable app quality assessment tool, suitable for health researchers. A team of 3 researchers with quantitative experience consulted articles on the use of the MARS, which are fundamental for avoiding deviations in the quantitative research process [30]. The test-completion procedures were carried out with the principal investigator to gain an initial understanding of the MARS evaluation instrument. During the anchoring process, the use of basic concepts for completing the instrument was standardized to ensure clarity and effectiveness in data collection [31]. Subsequently, the apps were divided for peer evaluation, and once 2 evaluations were obtained, averages were calculated for use. The scale demonstrated high internal consistency, with a Cronbach α coefficient of 0.90. In addition, the subscales (app quality rating, app subjective quality, and app-specific quality) showed strong internal consistency, with a mean coefficient of 0.85 [29].

The MARS uses 3 categories to evaluate several aspects of mobile app quality. The first category, "app quality rating," is assessed across 4 dimensions [27]: engagement (5 items: fun, interest, individual adaptability, interactivity, and target group), functionality (4 items: performance, usability, navigation, and gestural design), aesthetics (3 items: layout, graphics, and visual appeal), and information quality (7 items: accuracy of app description, goals, quality of information, quantity of information, quality of visual information, credibility, and evidence base). The second category, app subjective quality, comprises 4 questions (ie, "What is your overall star rating of the app?"). The third category, app-specific quality, comprises 6 questions (ie, "Is app content correct, well-written, and relevant to the goal/topic of the app?"). All items were rated on a 5-point scale, where 1 indicates "inadequate," 2 indicates "poor," 3 indicates "acceptable," 4 indicates "good," and 5 indicates "excellent." Thus, the scale contained 29 items that enabled a comprehensive and accurate assessment of mobile app quality.

For scoring, points were summed for each subscale, and tertile categories (low, medium, and high) were constructed in Microsoft Excel, considering the minimum and maximum scores possible and their ranges. This allowed for precise categorization of quality levels. For the first category, app quality rating, in the engagement subscale, 3 levels were defined: low (5 to 11 points), medium (12 to 18 points), and high (19 to 25 points). For functionality, the levels were: low (4 to 9 points), medium (10 to 14 points), and high (15 to 20 points). In terms of aesthetics, levels were classified as low (3 to 7 points), medium (8 to 11 points), and high (12 to 15 points). Finally, for information, the levels were: low (7 to 16 points), medium (17 to 25 points), and high (26 to 35 points). The second category, app subjective quality, had the following levels: low (4 to 9 points), medium (10 to 14 points), and high (15 to 20 points). Finally, in the third category, app-specific quality, levels were defined as low (6 to 14 points), medium (15 to 22 points), and high (23 to 30 points).

Step VI: Analysis and Synthesis

Based on the mean obtained by the app evaluators, values per level were assigned for each app as evaluation categories. These values were coded on a scale of 1 to 3, where high corresponds to 3 points, medium to 2 points, and low to 1 point. Percentages were established based on the criteria and evaluated apps. Two reviewers (JS-A and JCMF-D) independently conducted the assessment, resolving discrepancies by consensus. Results were processed using SPSS v27, and Cohen κ was used to measure

XSL•FO
RenderX

inter-rater agreement [32]. The calculated Cohen κ coefficient was 0.862, indicating near-perfect agreement between the evaluators. Descriptive statistics, frequencies, and percentages were used for data analysis. The MARS results are presented through stacked bar graphs and box plots. Directed content analysis was applied to narrative information, identifying patterns related to HIV and STI prevention methods. This methodology ensures a comprehensive and accurate evaluation of mobile apps for HIV and STI prevention, providing a solid foundation for future research and public health practice.

Results

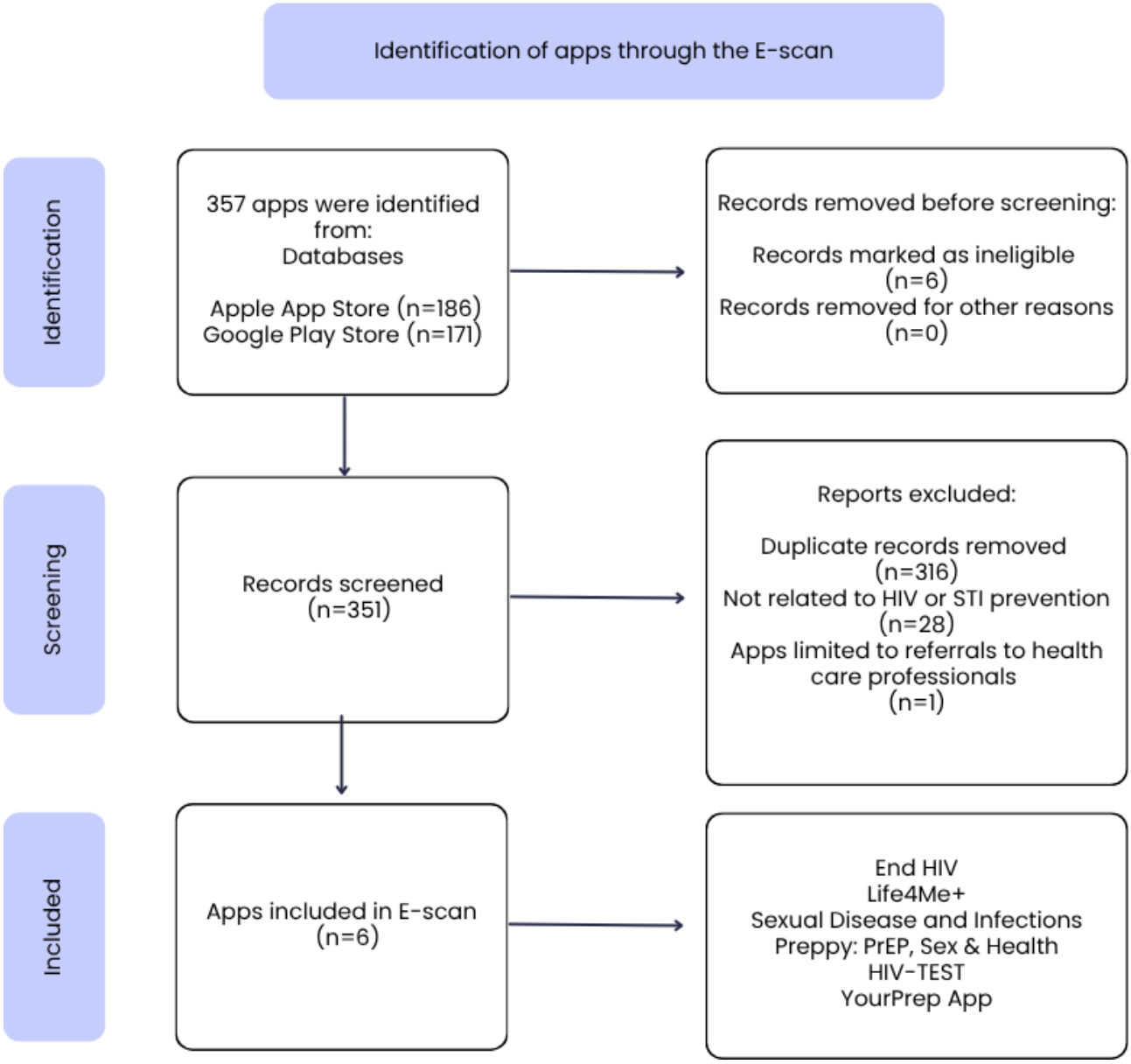
General Characteristics of the Final Apps

Figure 1 shows the PRISMA flowchart. The search identified a total of 357 apps, of which 186 were from the Apple App Store and 171 from the Google Play Store. The distribution by country was as follows: an average of 78 were available in Canada, 148 in the United States, and 131 in Mexico (overall mean 119, SD 36.52). The detailed organization of the apps facilitated selection by country, with independent review by JSA, JCMFD, and the principal investigator. As a result, 351 screened apps were eliminated, leaving 6 apps included in this review: "End HIV," "Life4Me+," "Sexual Disease and Infections," "Preppy: PrEP, Sex & Health," "HIV-TEST," and "Your Prep App."

All 6 identified mobile apps were available on either the Apple App Store (1/6, 16.7%), Google Play Store (2/6, 33.3%), or both platforms (3/6, 50%). Overall, 5 of the 6 apps (5/6, 83.3%) were available exclusively in English, while 1 app (1/6, 16.7%) supported 25 languages, including Spanish (Life4Me+). Most of the apps (5/6, 83.3%) had been updated in 2024, with the remaining 1 app (1/6, 16.7%) last updated in 2023. Two apps (2/6, 33.3%) were linked to nongovernmental organizations, 1 app (1/6, 16.7%) was linked to a university, and 3 apps (3/6, 50%) had no clearly identified affiliations. Most apps (5/6, 83.3%) were designed for individuals 17 years and older, while 1 app (1/6, 16.7%) was reported to be suitable for users 12 years and older (End HIV). The apps shared common objectives, including promoting PrEP, facilitating HIV testing, providing HIV risk assessment tools, and supporting HIV/STI treatment. Their specific focuses on HIV and STI prevention and care were categorized as follows: (1) prevention, diagnosis, and treatment of STIs (3/6, 50%); (2) combined HIV and STI prevention (1/6, 16.7%); and (3) PrEP-related resources (2/6, 33.3%) (Table 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for included apps. E-scan: environmental scan; STI: sexually transmitted infection.



XSL•FO
RenderX

Table . Characteristics of HIV and sexually transmitted infection (STI) prevention mobile apps in Canada, Mexico, and the United States.

App number and name	Platform (Android and iOS), languages, privacy, customization, and connectivity.	Developer	Goals	Approach	Year of update	Tested app version	Mean User Star Rating[a]	MARS[b] app quality rating score, mean (SD)[a]	MARS app subjective quality score, mean (SD)[a]	MARS app-specific quality, mean (SD)[a]
End HIV	Both, English, does not share data, and it is personalized, free.	University of Mississippi Medical Center	Inform about HIV prevention, testing, PrEP[c], and health care	HIV prevention, diagnosis, and treatment	2024	2.1.0	—[d]	4.55 (0.11)	3.9 (0.40)	4.5 (0.17)
Life4Me+	Both, English, Spanish, and 16 languages, does not share data, and is personalized, free.	Life4me.plus fight to AIDS, Hepatitis C, and Tuberculosis	Prevents HIV, or people already living with it, as well as other STIs	HIV prevention, STI prevention	2024	2.0.40	—	4.68 (0.13)	4.5 (0.10)	4.4 (0.22)
Sexual Disease and Infections	Android, English, does not share data, and is personalized, free.	Encoded Knowledge Ltd. Silversky Technology	Provide information about STDs[e], including symptoms, treatments, and prevention	STI prevention, diagnosis, and treatment	2024	1.25	5	4.6 (0.22)	3.5 (0.25)	3.3 (0.34)
Preppy: PrEP, Sex & Health	Both, English, does not share data, and is personalized free.	Prepsafe Global	Helps users manage PrEP, their sexual health, and well-being	Pre-exposure prevention	2024	1.2.3	4.6	4.57 (0.13)	4.1 (0.17)	4.5 (0.16)
HIV-TEST	Android, English, does not share data, and is personalized free.	Merezha	Provide a risk measurement tool that estimates the probability of HIV infection	HIV prevention	2024	2.3.1	—	4.18 (0.14)	3.1 (0.38)	4.3 (0.21)
YourPrEP App	Android, English, does not share data, and is personalized, free.	GGD Amsterdam	Inform about the use of PrEP	Pre-exposure prevention	2023	1.0.3.51	—	4.47 (0.10)	3.1 (0.32)	4.3 (0.17)

[a]Overall mean values are 4.8 (SD 0.28) for Mean User Star Rating, 4.5 (SD 0.17) for Mobile App Rating Scale (MARS) app quality score, 3.7 (SD 0.56) for MARS app subjective quality score, and 4.4 (SD 0.08) for app-specific quality.

[b]MARS: Mobile App Rating Scale.

[c]PrEP: pre-exposure prophylaxis.

[d]Not available.

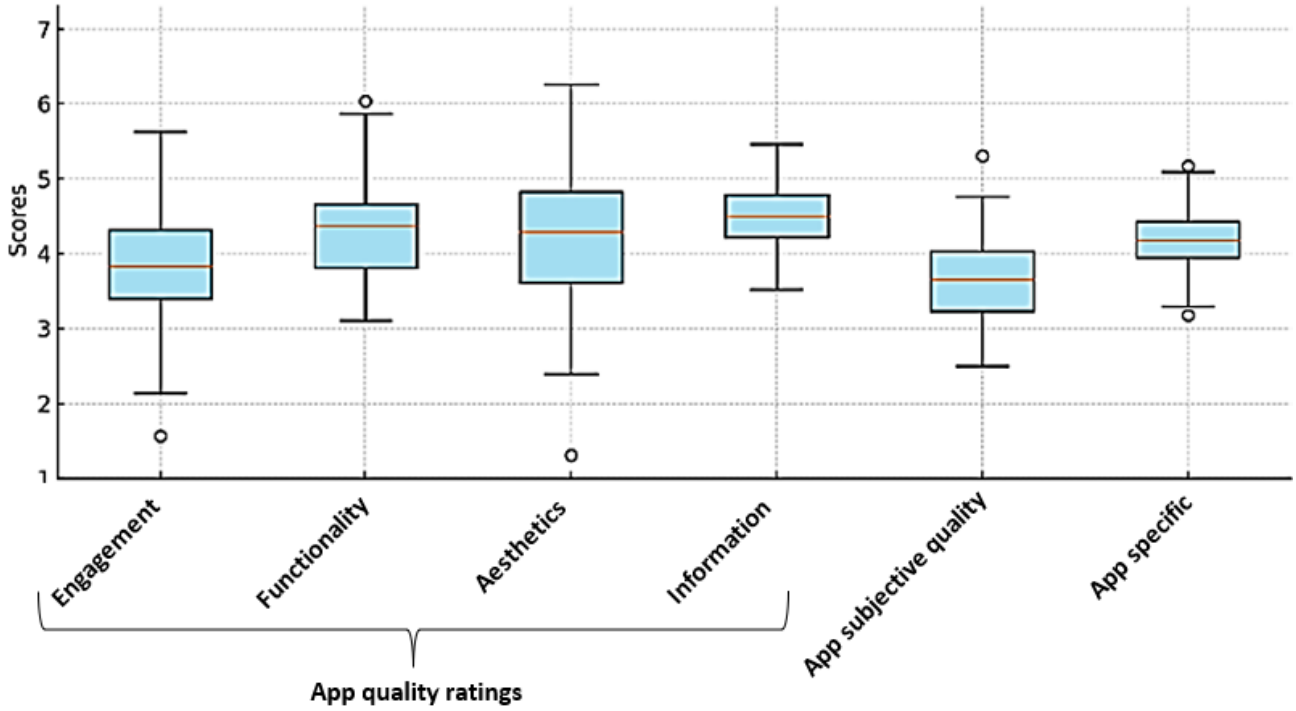
[e]STD: sexually transmitted disease.

XSL•FO
RenderX

Overall App Quality Assessment With MARS

The overall app quality assessment using the MARS framework revealed positive results across all 4 dimensions. Engagement had a mean score of 3.95 of 5 (SD 0.91), with apps achieving high ratings of 83.0%. Functionality demonstrated a mean of 4.32 (SD 0.63), with apps scoring highly (88.9%). Aesthetics achieved a mean score of 4.20, with apps receiving high ratings of 83.3%. The information dimension had the highest mean score of 4.48 (SD 0.45), with the apps scoring highly (100%). Among the 6 highest-rated apps, Preppy: PrEP, Sex & Health scored highest in engagement (4.8/5), End HIV excelled in functionality (4.9/5), Sexual Disease and Infections led in aesthetics (4.9/5), and the YourPrEP app received a perfect score in the information dimension (5/5) (Figure 2).

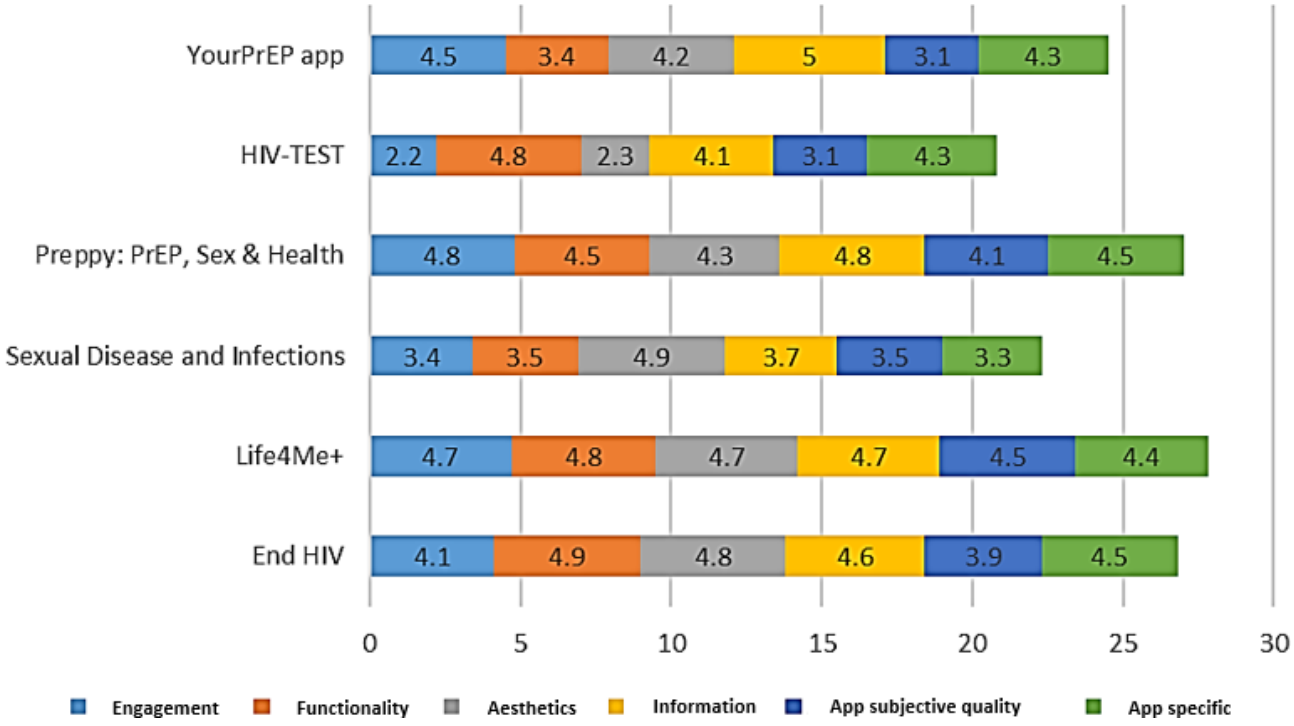
Figure 2. Mobile App Rating Scale (MARS) scores by dimensions.



The subjective quality of the apps also yielded favorable results, with a mean score of 3.70 (SD 0.52) and the apps receiving high ratings of 83.3%. The highest subjective quality score was recorded for Life4Me+ (4.5/5), while the lowest ratings were observed for HIV-TEST and the YourPrEP app, both scoring 3.1/5. Similarly, the app-specific quality assessment revealed a mean score of 4.22 (SD 0.42), with apps achieving high ratings of 88.9%. The highest scores in this category were shared by Preppy: PrEP, Sex & Health, and End HIV, both averaging 4.5/5, while Sexual Disease and Infections received the lowest score (3.3/5). Overall, the results highlight the strong performance of these apps across multiple quality dimensions, underscoring their potential use in addressing HIV prevention and care (Figure 3).

XSL•FO
RenderX

Figure 3. Cumulative score of each app's Mobile App Rating Scale (MARS) subdomains.



Barriers and Facilitators

The following are the main barriers and facilitators identified and also the strengths and weaknesses found for each app were addressed (seeMultimedia Appendix 1 and Garcia et al [33]). The relevant aspects of barriers and facilitators were grouped into three sections that are (1) technological, (2) social and psychological, and (3) governance; further details about these barriers are listed below.

Technological

Technical

Facilitating features included aspects such as the type of operating system and the speed of downloads. The diversity of mobile devices, browsers, and assistive technologies was noted, following platform accessibility guidelines with which the apps are compatible. The apps generally used good visual or auditory design.

Content-Related

Facilitating features included the year of creation and update, availability of languages, free access in most cases, and continuous online availability. They used simple language, clear information, and appropriate hierarchy in information distribution, since without these aspects, legibility, comprehension, and usability could be compromised. In this study, all the apps addressed topics related to HIV and STIs to a greater or lesser degree. On the other hand, barriers included some apps requiring user registration for access, though this was less common. A lack of or inconsistency in app metadata (downloads and ratings) was also observed, which may lead to data loss and variation in the classification of information during analysis. Among the studied apps that mentioned HIV and STI detection and treatment, all emphasized that their use does not substitute for physical testing, treatment, or medical consultations, recognizing that use is the responsibility of the user [34]. In addition, there was very little mention of migrant populations, as the apps primarily targeted men who have sex with men [35].

Social and Psychological

Age

This was generally a facilitator, as the apps were understandable for almost all audiences. Most were identified as suitable for users older than 10 years. *Economic:* Mobile phones can serve as facilitators since they are relatively accessible and portable when small. However, they can become barriers if users cannot afford a smartphone, or if the device is larger (such as an iPad, tablet, or laptop), making transport more difficult.

Cultural Sensitivity (Stigma)

Facilitating aspects included awareness and respect for diversity, culture, values, beliefs, and personal preferences. Most apps did not request personal information beyond an email account, alias, and confirmation of age. Generally, if personal information was requested, a privacy notice was provided. In this study, people living with HIV were not involved in the app evaluations.

Cognitive Characteristics (Ease of Use, Usefulness, Satisfaction, Digital Literacy, Online Participation, and Trust in the Internet)

Facilitating aspects included ease of use. Usefulness and satisfaction were supported by positive user comments, which could serve as suggestions for improvement. Digital literacy was observed through the rapid learning curve when using the apps. Online participation was reflected in the number of stars and downloads, indicating that apps had been marked as highlighted or favorites. However, a main barrier was trust in internet access: without internet connectivity, users cannot download or use the app [36].

XSL•FO
RenderX

Governance (Legal compliance)

As a facilitating feature, most apps had country-specific but legal permissions. However, this aspect varied, and consultation with legal experts was deemed essential to avoid lawsuits, fines, or damages.

Discussion

Principal Findings

Digital interventions offer an inclusive, adaptable, and confidential approach to prevention, contrasting with traditional in-person interventions that are often inaccessible, fear-based, and ineffective for populations made vulnerable [37]. This study focused on mobile apps related to HIV and STI prevention across Canada, Mexico, and the United States. Using the E-Scan methodology, recognized for its reliability in selecting mobile apps [25], the study used the MARS tool to evaluate 6 apps. These apps were assessed based on quality ratings (engagement, functionality, information, and aesthetics), app subjective quality, and app-specific criteria.

All the apps included in this study shared common goals related to HIV and STI prevention. This availability of the service, but at the same time, the limited access for the migrant population underscores a gap in addressing public health issues such as HIV and STI prevention, particularly among vulnerable populations [38,39]. High information and functionality scores reflect intuitive and practical interactions that encourage adoption [29]. However, differences in app quality, such as low interactivity, can impede user engagement and learning [40]. This study revealed high scores for information, with slightly lower engagement scores, highlighting areas for improvement to enhance user satisfaction. Appealing aesthetics further help sustain user attention [41], who noted that app quality significantly influences user trust and decision-making. Despite the proliferation of health apps, concerns remain about the reliability of information for nonprofessional users [42]. Users often face an overwhelming number of options without clear quality indicators owing to a scarcity of scientifically validated evidence underpinning health app content [43]. Evidence-based apps supporting HIV and STI prevention among minoritized populations, including migrants, show promise [44-46] but must overcome structural barriers and implementation costs to increase their reach. Ensuring updated, relevant, and reliable content is critical to maintaining credibility, while subjective and specific quality criteria ensure personalized and trustworthy user experiences.

Understanding user engagement and app scalability is crucial for addressing health access barriers. While most apps in this study were accessible in English, fewer supported other languages. Language availability is essential for reducing barriers to HIV care and prevention [47,48]. Language-specific studies revealed a lack of interventions tailored for Spanish-speaking Latino individuals, adolescents, and men who have sex with men [49]. Accurate translation of HIV and STI information was inconsistently noted in app descriptions, emphasizing the need for linguistic rigor. These findings highlight the importance of leveraging MARS ratings to guide future app development and improvements in accessibility, reach, acceptability, and effectiveness.

Implications for Practice, Research, and Policy

The findings from the E-Scan and MARS evaluation have critical implications for practice, research, and policy. Practitioners and public health organizations should integrate mobile apps into comprehensive health promotion strategies, particularly for populations made vulnerable, such as migrants. High engagement, functionality, and information quality make these apps valuable tools for enhancing health literacy and promoting preventive behaviors, though addressing aesthetic and subjective quality concerns is necessary to improve user experience and sustain engagement.

Regarding usefulness, given the increasing use of mHealth apps alongside the growth of social networks, smartphones, and advancing technology, these tools should be leveraged and translated into activities focused on HIV and STI prevention, detection, and medical care among migrant and undocumented populations. For example, among young men, several factors are associated with increased risk of infection, such as risky sexual behaviors through dating apps, a history of sexually transmitted infections, sexualized drug use, and pornography consumption, among others [50-52]. These actions can be implemented across multiple settings, since this population is immersed in diverse environments: educational centers (schools and universities), workplaces, government centers (migrant detention centers, immigration embassies, and legal agencies), community centers (shelters, civil society organizations, and faith-based organizations), cultural spaces, and commercial venues, among the most relevant [53]. Efforts should include the distribution and dissemination of flyers, brochures, posters, and wall newspapers, as well as incorporation into pedagogical activities, nursing talks, and health promotion sessions, with explicit mention of the existence of these apps. Strategies from mass media (eg, television, radio, social networks such as Facebook [Meta], Twitter [Meta], and Instagram [Meta]) can also be leveraged.

Undocumented migrant populations are also served in health care institutions, where strategies can be developed across the 3 levels of health care. At the first level of care, with a community-based focus, activities may include communication, health promotion, prevention, and early detection of HIV and STIs, including dissemination of information about the existence of apps [54]. At the second and third levels of care, which are more focused on treatment and referral of newly diagnosed individuals to other services, these activities can be incorporated into health programs and projects (eg, best practices or evidence-based practices). Information about HIV and STI-related apps can also be included on hospital websites and in prevention campaigns delivered by nurses, physicians, social workers, psychologists, and health educators. Finally, at tertiary-level or specialty hospitals that function as research centers, the topic of HIV and apps can be formally integrated into clinical and educational initiatives.

This article underscores the need for ongoing evaluation of digital health tools using frameworks such as MARS to assess their real-world effectiveness. Future studies should refine

XSL•FO
RenderX

evaluation tools and methodologies such as the E-Scan to incorporate diverse user experiences and cultural contexts, ensuring broader accessibility and relevance. Ensuring apps adhere to evidence-based practices and remain updated with reliable information is crucial for advancing digital health interventions. Policymakers should prioritize developing guidelines and standards to ensure health apps meet high-quality criteria, focusing on transparency, accuracy, and accessibility [55,56]. Given the increasing barriers to accessing health care, apps with accurate and reliable information can support migrant populations navigating changing political landscapes in search of health care services.

Changes in the political climate in the United States and globally have increased fear among immigrants, leading to missed appointments and relocations [57]. State and local policies across sectors such as labor, health, and education influence HIV vulnerability among Latino migrants by shaping access to institutions [58]. For example, state-level legislation that requires hospitals to collect immigration status may lead to less health care use, as such measures can deter undocumented or mixed-status families from seeking care due to fears of disclosure, discrimination, or legal repercussions. This reduced engagement with health care services can delay HIV testing, prevention, and treatment, thereby increasing vulnerability within migrant communities [59]. Moreover, immigration policies are impacted by partisan polarization and right-wing populism, aspects in which progress is not achieved, such as full access to health care for migrants, the availability of international medical graduates, which are essential for primary care in underserved areas, among other aspects [60]. Language policies addressing social inclusion for migrants are critical to reducing inequality and ensuring access to public services [61].

Strengths and Limitations

This study systematically evaluated mHealth apps across 3 countries, key destinations for migrants, leveraging the bilingual expertise of its researchers in Spanish and English. The use of 3 reviewers per app and Cohen κ calculations ensured accuracy and consistency. However, limitations include the rapid evolution of health apps, potentially excluding newer options, and challenges in evaluating instruments in languages not native to the researchers. The study also did not incorporate direct user feedback or usability testing, which could provide critical insights into real-world adoption, engagement, and satisfaction. The generalizability of the MARS ratings to undocumented migrants may be limited, as this group may face distinct access barriers and usage patterns that were not directly assessed. In addition, the use of tertiles to interpret MARS scores may be less reliable with a small sample size, potentially limiting the robustness of these categorizations. Another limitation is the lack of comparison between app content and authoritative online resources, such as Centers for Disease Control and Prevention websites, which could help assess the relative quality and completeness of the information provided. Moreover, while this study was unable to assess the efficacy or perception of behavior change resulting from the apps, doing so would be a good next step in their evaluation. Despite these limitations, the study provides a comprehensive view of the availability, accessibility, and quality of HIV and STI prevention apps, emphasizing the need for linguistic accessibility and optimized user experiences.

Conclusions

The mobile apps analyzed in this study, focusing on HIV and STI prevention, are available in Canada, the United States, and Mexico through major app stores. These apps primarily offer free access and support multiple languages. The MARS evaluation revealed high performance in engagement, functionality, and information quality, with room for improvement in aesthetics and subjective ratings. These findings underscore the critical role of digital resources in addressing public health challenges faced by minoritized populations, such as migrants. Expanding methodologies like the E-Scan and tools like the MARS could further enhance the field. Addressing quality concerns is essential to ensuring that digital interventions provide credible, satisfying, and effective solutions that drive tangible health impacts. High standards in engagement, functionality, and information are not only technical requirements but fundamental necessities for promoting app credibility and effectiveness in public health promotion.

Acknowledgments

We would like to thank MSN Héctor Miguel Melchor Ortiz for his valuable contribution in the search for information and Mrs Marketa Wade for copyediting. No generative artificial intelligence tools were used in the writing, editing, data analysis, or figure creation for this article. All content was produced by the authors.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: HF-S, JS-A, JCMF-D, AG, MJM, CER-D, DSM.
Formal analysis: HF-S, JS-A, JCMF-D, AG, MJM, CER-D, DSM.
Funding acquisition: not applicable
Investigation: HF-S, JS-A, JCMF-D, AG, MJM, CER-D, DSM.
Methodology: HF-S, JS-A, JCMF-D, AG, MJM, CER-D, DSM.
Project administration: HF-S.

XSL•FO
RenderX

Supervision: HF-S.
Writing – original draft: HF-S.
Writing – review and editing: HF-S, JS-A, JCMF-D, AG, MJM, CER-D, DSM.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table of strengths and weaknesses of each application.
[DOCX File, 16 KB - mhealth_v13i1e72009_app1.docx]

References

1. Giourguli-Saucedo SE, Lazcano-Ponce E. Migración y salud: el reflejo de la inequidad social global y la necesidad de investigación en salud pública [Article in Spanish]. Salud Publica Mex 2024;66(2, mar-abr):115-116. [doi: 10.21149/15530] [Medline: 39977033]
2. Bojórquez I, Cerecero-García D, Orraca-Romano PP, Fernández-Niño J, Rojas-Botero M, Infante C. Atención en salud a migrantes en tránsito: estimación del costo para el sistema de salud en México [Article in Spanish]. Salud Publica Mex 2024;66(6 (nov-dec):816-823. [doi: 10.21149/15736] [Medline: 39977155]
3. Zamora Salazar C, Casillas R. Detención de migrantes indocumentados en Estados Unidos: ¿quién es quién en aprehensiones? [Article in Spanish]. Migr Inter 2024 Jul 29;15:0. [doi: 10.33679/rmi.v1i1.2798]
4. Migraciones sur-norte desde sudamérica: rutas, vulnerabilidades y contextos del tránsito de migrantes extrarregionales [Report in Spanish]. : Oficina Regional de la OIM para América del Sur, OIM para América del Sur (International Organization for Migration); 2022 URL: https://lac.iom.int/sites/g/files/tmzbdl626/files/documents/oim_migraciones-sur-norte-desde-suramerica.pdf [accessed 2024-07-08]
5. Reitz JG, Hernández Jasso M. Growth in high-skilled Mexican migration northward: American and Canadian destinations [Article in Spanish]. Migr Inter 2023;14:0. [doi: 10.33679/rmi.v1i1.2593]
6. Guerra-Ordoñez JA, Benavides-Torres RA, Zapata-Garibay R, Ruiz-Cerino JM, Ávila-Alpirez H, Salazar-Barajas ME. Percepción de riesgo para VIH y sexo seguro en migrantes de la frontera norte de México [Article in Spanish]. Revista Internacional de Andrología 2022 Apr;20(2):80-85. [doi: 10.1016/j.androl.2020.10.010] [Medline: 35078727]
7. Parco Gavilánez MC. Calidad de vida de los pacientes VIH positivo [Article in Spanish]. Ciencia Latina 2023;7(4):1732-1746. [doi: 10.37811/cl_rcm.v7i4.7006]
8. Swinkels HM, Gulick PG, Nguyen AD. HIV and AIDS. In: StatPearls [Internet]: StatPearls Publishing; 2024. [Medline: 30521281]
9. Aumento de la migración en las américas en 2023: retos para garantizar la salud de las personas migrantes y respuesta de la organización panamericana de la salud [Web page in Spanish]. OPS/OMS. 2023. URL: https://www.paho.org/es/noticias/18-12-2023-aumento-migracion-americas-2023-retos-para-garantizar-salud-personas-migrantes [accessed 2025-10-31]
10. Kaur H, Saad A, Magwood O, et al. Understanding the health and housing experiences of refugees and other migrant populations experiencing homelessness or vulnerable housing: a systematic review using GRADE-CERQual. CMAJ Open 2021;9(2):E681-E692. [doi: 10.9778/cmajo.20200109] [Medline: 34145051]
11. Zhang L, Au W, Ewesesan R, Yakubovich AR, Brownridge DA, Urquia ML. Intimate partner violence among international and interprovincial migrants: a population-based analysis of Canadian linked immigration and justice data. Violence Against Women 2024 Nov;30(14):3455-3473. [doi: 10.1177/10778012231178001] [Medline: 37272037]
12. Sabri B, Njie-Carr VPS, Messing JT, et al. The weWomen and ourCircle randomized controlled trial protocol: a web-based intervention for immigrant, refugee and indigenous women with intimate partner violence experiences. Contemp Clin Trials 2019 Jan;76:79-84. [doi: 10.1016/j.cct.2018.11.013] [Medline: 30517888]
13. Santoso D, Asfia S, Mello MB, et al. HIV prevalence ratio of international migrants compared to their native-born counterparts: a systematic review and meta-analysis. EClinicalMedicine 2022 Nov;53:101661. [doi: 10.1016/j.eclinm.2022.101661] [Medline: 36147629]
14. Sanchez MA, Lemp GF, Magis-Rodríguez C, Bravo-García E, Carter S, Ruiz JD. The epidemiology of HIV among Mexican migrants and recent immigrants in California and Mexico. J Acquir Immune Defic Syndr 2004 Nov 1;37 Suppl 4:S204-S214. [doi: 10.1097/01.qai.0000141253.54217.24] [Medline: 15722863]
15. Li K, Thaweesee N, Kimmel A, Dorward E, Dam A. Barriers and facilitators to utilizing HIV prevention and treatment services among migrant youth globally: a scoping review. PLOS Glob Public Health 2024;4(2):e0002851. [doi: 10.1371/journal.pgph.0002851] [Medline: 38354206]
16. Benoit JRA, Louie-Poon S, Kauser S, Meherali S. Promoting adolescent sexual and reproductive health in North America using free mobile apps: environmental scan. JMIR Pediatr Parent 2022 Oct 4;5(4):e33826. [doi: 10.2196/33826] [Medline: 36194450]

XSL•FO
RenderX

17. Muñoz D, Gomez J, Andrade KV, Institución Universitaria Antonio José Camacho. Salud móvil (mHealth) en profundidad: un estudio bibliométrico [Article in Spanish]. Espacios 2025 Jul;46(3):130-152. [doi: 10.48082/espacios-a25v46n03p11]
18. Cunha-Oliveira AD, Gómez Cantarino S, Santos D, et al. Use of digital technologies to promote sexual health in young adults: a narrative review. Rev Bras Enferm 2025;78Suppl 1(Suppl 1):e20230434. [doi: 10.1590/0034-7167-2023-0434] [Medline: 40561348]
19. Biello KB, Marrow E, Mimiaga MJ, Sullivan P, Hightow-Weidman L, Mayer KH. A mobile-based app (MyChoices) to increase uptake of hiv testing and pre-exposure prophylaxis by young men who have sex with men: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2019 Jan 7;8(1):e10694. [doi: 10.2196/10694] [Medline: 30617042]
20. Jones J, Dominguez K, Stephenson R, et al. A theoretically based mobile app to increase pre-exposure prophylaxis uptake among men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Feb 21;9(2):e16231. [doi: 10.2196/16231] [Medline: 32130178]
21. Biello KB, Hill-Rorie J, Valente PK, et al. Development and evaluation of a mobile app designed to increase hiv testing and pre-exposure prophylaxis use among young men who have sex with men in the United States: open pilot trial. J Med Internet Res 2021 Mar 24;23(3):e25107. [doi: 10.2196/25107] [Medline: 33759792]
22. Reis L, Pereira M, Silveira C. The social challenge of migrant integration: the role of mobile apps. In: Handbook of Research on Solving Societal Challenges Through Sustainability-Oriented Innovation: IGI Gobal; 2023:307-327. [doi: 10.4018/978-1-6684-6123-5.ch016]
23. Velázquez Hidalgo AS, Fernández Pérez Y, Pacheco Jeréz YS, Zulueta Véliz Y. Métodos y criterios para valorar la calidad en uso de las aplicaciones Móviles para atención al ciudadano cubano [Article in Spanish]. Rev Cubana Cienc Inform 2023 Sep;17(3) [FREE Full text]
24. Bilal A, Mirza HT, Hussain I, Ahmad A. Investigating influence of Google-Play application titles on success. Big Data Research 2024 May;36:100443. [doi: 10.1016/j.bdr.2024.100443]
25. Fernández-Sánchez H, Guerrero-Castañeda RF, Jones J, King K. Exploraciones ambientales en ciencias de la salud [Article in Spanish]. Horiz Enferm 2024;35(2):703-715. [doi: 10.7764/Horiz_Enferm.35.2.703-715]
26. Harris B, Brooker J. Environmental scanning: a look to the future [Article in Spanish]. New Drctns Evaluation 2025 Mar;2025(185-186):33-41. [doi: 10.1002/ev.20633]
27. Bekar Adiguzel M, Cengiz MA. Model selection in multivariate adaptive regressions splines (MARS) using alternative information criteria. Heliyon 2023 Sep;9(9):e19964. [doi: 10.1016/j.heliyon.2023.e19964] [Medline: 37809827]
28. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. J Clin Epidemiol 2021 Jun;134:178-189. [doi: 10.1016/j.jclinepi.2021.03.001] [Medline: 33789819]
29. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
30. Delgado-Morales C, Duarte-Hueros A. Una Revisión sistemática de instrumentos que evalúan la calidad de aplicaciones móviles de salud [Article in Spanish]. Pixel-Bit (67):35-58. [doi: 10.12795/pixelbit.97867]
31. Márquez JL. ¿Somos predeciblemente racionales o predeciblemente irracionales? Un estudio sobre el “efecto anclaje” [Article in Spanish]. ddee 2023(3):51-63. [doi: 10.26439/ddee2023.n003.6052]
32. Duarte-Anselmi G, Leiva-Pinto E, Vanegas-López J, Thomas-Lange J. Experiences and perceptions on sexuality, risk and STI/HIV prevention campaigns by university students. designing a digital intervention. Cien Saude Colet 2022 Mar;27(3):909-920. [doi: 10.1590/1413-81232022273.05372021] [Medline: 35293468]
33. Garcia Saiso S, Marti MC, Malek Pascha V, et al. Implementation of telemedicine in the Americas: barriers and facilitators [Article in Spanish]. Rev Panam Salud Publica 2021;45:e131. [doi: 10.26633/RPSP.2021.131] [Medline: 34703460]
34. Schaab BL, Remor E. Development, feasibility testing and perceived benefits of a new app to help with adherence to antiretroviral therapy in people living with HIV in Brazil. Pilot Feasibility Stud 2023 Jul 26;9(1):130. [doi: 10.1186/s40814-023-01370-7] [Medline: 37496084]
35. Pravosud V, Ballard AM, Holloway IW, Young AM. Latent class analysis of online platforms for partner-seeking and sexual behaviors among men who have sex with men from Central Kentucky. AIDS Behav 2024 Mar;28(3):1015-1028. [doi: 10.1007/s10461-023-04210-0] [Medline: 37824036]
36. Mola DJ, Josefina F, Sonia D, Cecilia R. Barreras y facilitadores en el uso de las herramientas digitales de gobierno: resultados preliminares [Article in Spanish]. Presented at: XIX Reunión Nacional y VIII Encuentro Internacional de La Asociación Argentina de Ciencias Del Comportamiento; Aug 23-25, 2023 URL: https://ri.conicet.gov.ar/handle/11336/244461 [accessed 2025-10-31]
37. Blamey R, Sciaraffia A, Piñera C, et al. Situación epidemiológica de VIH a nivel global y nacional: puesta al día [Article in Spanish]. Rev chil infectol 2024;41(2):248-258. [doi: 10.4067/s0716-10182024000200248]
38. Rojas D. Acceso y uso de datos móviles en poblaciones migrantes [Web page in Spanish]. Centro LATAM Digital. 2023. URL: https://centrolatam.digital/publicaciones/acceso-y-uso-de-datos-moviles-por-poblaciones-migrantes [accessed 2024-11-08]
39. Sullivan PS, Hightow-Weidman L. Mobile apps for HIV prevention: how do they contribute to our epidemic response for adolescents and young adults? Mhealth 2021;7:36. [doi: 10.21037/mhealth-20-71] [Medline: 33898605]

XSL•FO
RenderX

40. Acosta Espinoza JL, León Yacelga LR, Sanafria Michilena WG. Las aplicaciones móviles y su impacto en la sociedad [Article in Spanish]. Rev Univ Soc 2022;14(2):237-243 [FREE Full text]
41. Preciado-Ortiz CL. Quality and use of mobile applications for transportation service: influence on satisfaction. MYN 2021 Jul 1(44):21-42. [doi: 10.32870/myn.v0i44.7646]
42. Tala Á, Vásquez E, Rojas E, Marín R. Apps y Medicina: una visión global y la situación chilena [Article in Spanish]. Rev méd Chile 2022;150(2):206-215. [doi: 10.4067/S0034-98872022000200206] [Medline: 36156647]
43. Collado-Borrell R, Escudero-Vilaplana V, Narrillos-Moraza Á, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Sáez M. Patient-reported outcomes and mobile applications. A review of their impact on patients' health outcomes. Farm Hosp 2022 Apr 27;46(3):173-181. [doi: 10.7399/fh.11830] [Medline: 36183211]
44. Curioso WH, Blas MM, Kurth AE, Klausner JD. Tecnologías de información y comunicación para la prevención y control de la infección por el VIH y otras ITS [Article in Spanish]. Rev Peru Med Exp Salud Publica 2007;24(3):262-271. [Medline: 26339254]
45. Santa Maria D, Padhye N, Businelle M, et al. Efficacy of a just-in-time adaptive intervention to promote HIV risk reduction behaviors among young adults experiencing homelessness: pilot randomized controlled trial. J Med Internet Res 2021 Jul 6;23(7):e26704. [doi: 10.2196/26704] [Medline: 34255679]
46. Nahum-Shani I, Naar S. Digital adaptive behavioral interventions to improve HIV prevention and care: innovations in intervention approach and experimental design. Curr HIV/AIDS Rep 2023 Dec;20(6):502-512. [doi: 10.1007/s11904-023-00671-z] [Medline: 37924458]
47. Mounkoro I. Revolucionando el aprendizaje de idiomas: el impacto transformador de las aplicaciones móviles [Article in Spanish]. Apertura 2024;16(2):22-35. [doi: 10.32870/Ap.v16n2.2528]
48. Marzan-Rodriguez M, Rodriguez-Diaz CE, Mustanski B. Recommendations for the development of HIV prevention interventions among Latino young sexual minority groups. Sex Res Social Policy 2021 Dec;18(4):992-1001. [doi: 10.1007/s13178-020-00494-2] [Medline: 38124992]
49. Border health and migration. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/migration-border-health/work-in-mexico/index.html [accessed 2024-11-08]
50. Naranjo-Márquez M, Bocchino A, Gilart E, Cotobal-Calvo EM, Procentese F, Palazón-Fernández JL. Risk determinants of sexual behaviors: dating apps, history of sexually transmitted infections, substance use, and pornography consumption in health science students. Nurs Rep 2025 Feb 28;15(3):83. [doi: 10.3390/nursrep15030083] [Medline: 40137657]
51. Marqués-Sánchez P, Bermejo-Martínez D, Quiroga Sánchez E, Calvo-Ayuso N, Liébana-Presa C, Benítez-Andrades JA. Men who have sex with men: an approach to social network analysis. Public Health Nurs 2023 Jan;40(1):73-79. [doi: 10.1111/phn.13138] [Medline: 36217564]
52. Lisboa C, Stuardo V, Folch C. Sexualized drug use among gay men and other men who have sex with men in Latin America: a description of the phenomenon based on the results of LAMIS-2018. PLoS ONE 2023;18(10):e0287683. [doi: 10.1371/journal.pone.0287683] [Medline: 37856552]
53. Rafful C, Orozco R, Peralta D, et al. Feasibility, acceptability, and perceived usefulness of a community-evidence-based harm reduction intervention for sexualized stimulant use among Mexican gay, bisexual, and other men who have sex with men. Harm Reduct J 2024 May 16;21(1):95. [doi: 10.1186/s12954-024-01020-y] [Medline: 38755623]
54. González Martínez M, Castaño Suero MJ, Guerrero Muñoz M, Francisco Rossetti A, Sequeira Aymar E, Roca Saumell C. Initial assessment of immigrant patients in primary care. Aten Primaria 2024 Jul;56(7):102896. [doi: 10.1016/j.aprim.2024.102896] [Medline: 38417201]
55. Savas ST, Knipper M, Duclos D, Sharma E, Ugarte-Gurrutxaga MI, Blanchet K. Migrant-sensitive healthcare in Europe: advancing health equity through accessibility, acceptability, quality, and trust. Lancet Reg Health Eur 2024 Jun;41:100805. [doi: 10.1016/j.lanepe.2023.100805] [Medline: 39444446]
56. Tummala-Narra P. El miedo de los inmigrantes [Article in Spanish]. Aperturas Psicoanaliticas 2022(70):e5 [FREE Full text]
57. Pérez CS. Sistemas de salud y políticas de atención al migrante [Article in Spanish]. FMC - Formación Médica Continuada en Atención Primaria 2025 Aug;32(7):57-62. [doi: 10.1016/j.fmc.2025.07.005]
58. Executive order GA-46 reporting for collection of information on hospital costs related to immigration status. Texas Health and Human Services. 2024. URL: https://content.govdelivery.com/accounts/TXHHSC/bulletins/3b0acec [accessed 2024-08-11]
59. Conozca sus derechos a obtener atención médica y seguro médico [Web page in Spanish]. National Immigration Law Center. URL: https://www.nilc.org/resources/derechos-seguro-y-cuidado-de-salud/ [accessed 2024-11-08]
60. González Delgado D. La política migratoria de Estados Unidos en el siglo XXI: evolución y desafíos [Article in Spanish]. Novedades en Población 2024:125-156 [FREE Full text]
61. Gazzola M, Wickström BA, Fettes M. Towards an index of linguistic justice. Politics, Philosophy & Economics 2023 Aug;22(3):243-270. [doi: 10.1177/1470594X231158657]

XSL•FO
RenderX

Abbreviations

E-scan: environmental scan
MARS: Mobile App Rating Scale
PrEP: pre-exposure prophylaxis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
STI: sexually transmitted infection

Edited by L Buis; submitted 31.01.25; peer-reviewed by F Carrouel, K Fiscella; revised version received 06.09.25; accepted 20.10.25; published 14.11.25.

Please cite as:
Fernández-Sánchez H, Salazar-Alberto J, Fernández-Delgado JCM, Galvin A, Mugavero MJ, Rodriguez-Diaz CE, Santa Maria D Mobile Apps for HIV and Sexually Transmitted Infection Prevention in Canada, Mexico, and the United States: Environmental Scan JMIR Mhealth Uhealth 2025;13:e72009
URL: https://mhealth.jmir.org/2025/1/e72009
doi:10.2196/72009

© Higinio Fernández-Sánchez, Javier Salazar-Alberto, Jhan Carlos Manuel Fernández-Delgado, Annalynn Galvin, Michael J Mugavero, Carlos E Rodriguez-Diaz, Diane Santa Maria. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone Apps for Cardiovascular and Mental Health Care: Digital Cross-Sectional Analysis

Manjot Singh[1,2], BSc; Julian Herpertz[3], MD; Noy Alon[3], BSc; Sarah Perret[3], BSc; John Torous[3], MD, MBI; Daniel Kramer[1], MD, MPH

[1]Richard A. and Susan F. Smith Center for Outcomes Research, Beth Israel Deaconess Medical Center, 375 Longwood Ave, Boston, MA, United States
[2]University of Connecticut School of Medicine, UConn Health, Farmington, CT, United States
[3]Division of Digital Psychiatry, Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, United States

Corresponding Author:
Daniel Kramer, MD, MPH
Richard A. and Susan F. Smith Center for Outcomes Research, Beth Israel Deaconess Medical Center, 375 Longwood Ave, Boston, MA, United States

Abstract

Background: The rapidly expanding digital health landscape offers innovative opportunities for improving health care delivery and patient outcomes; however, regulatory and clinical frameworks for evaluating their key features, effectiveness, and outcomes are lacking. Cardiovascular and mental health apps represent 2 prominent categories within this space. While mental health apps have been extensively studied, limited research exists on the quality and effectiveness of cardiovascular care apps. Despite their potential, both categories of apps face criticism for a lack of clinical evidence, insufficient privacy safeguards, and underuse of smartphone-specific features alluding to larger shortcomings in the field.

Objective: This study extends the use of the MINDApps framework to compare the quality of cardiovascular and mental health apps framework to compare the quality of cardiovascular and mental health apps with regard to data security, data collection, and evidence-based support to identify strengths, limitations, and broader shortcomings across these domains in the digital health landscape.

Methods: We conducted a systematic review of the Apple App Store and Google Play Store, querying for cardiovascular care apps. Apps were included if they were updated within the past 90 days, available in English, and did not require a health care provider's referral. Cardiovascular care apps were matched to mental health apps by platform compatibility and cost. Apps were evaluated using the M-Health Index & Navigation Database (MIND; MINDApps), a comprehensive tool based on the American Psychiatric Association's app evaluation model. The framework includes 105 objective questions across 6 categories of quality, including privacy, clinical foundation, and engagement. Statistical differences between the 2 groups were assessed using two-proportion Z-tests.

Results: In total, 48 cardiovascular care apps and 48 matched mental health apps were analyzed. The majority of apps in both categories included a privacy policy; yet, the majority in both samples shared user data with third-party companies. Evidence for effectiveness was limited, with only 2 (4%) cardiovascular care apps and 5 (10%) mental health apps meeting this criterion. Cardiovascular care apps were significantly more likely to be used in external devices such as smartphone-based electrocardiograms and blood pressure monitors.

Conclusions: Both categories lack robust clinical foundations and face substantial privacy challenges. Cardiovascular apps have the potential to revolutionize patient monitoring; yet, their limited evidence base and privacy concerns highlight opportunities for improvement. Findings demonstrate the broader applicability of the MINDApps framework in evaluating apps across medical fields and stress the significant shortcomings in the app marketplace for cardiovascular and mental health. Future work should prioritize evidence-based app development, privacy safeguards, and the integration of innovative smartphone functionalities to ensure that health apps are safe and effective for patient use.

(JMIR Mhealth Uhealth 2025;13:e63642) doi:10.2196/63642

KEYWORDS

smartphones; cardiovascular disease; cross sectional; application; app; smartphone app; cardiovascular; cardio; mental health; effectiveness; patient-reported symptoms; mental disorder; symptoms; effective

XSL•FO
RenderX

Introduction

The rapidly growing digital health landscape enhances accessibility and broadens the dissemination of health-related products that were once limited to traditional medical settings [1]. According to recent reports by Quintiles and IMS Health Incorporated, more than 350,000 digital health apps are currently available on the Apple App Store and Google Play Store [2]. These include apps that connect to external devices, record and transmit patient health data, or use diagnostic sensors integrated into smartphones. With the growing number of apps on the market and an increasing number of users downloading them, there is an urgent need to evaluate their quality based on criteria such as evidence base, usability, data privacy, and security [3].

The US Food and Drug Administration (FDA) plays a key role in regulating the marketing of medical products and monitoring of safety signals. However, policies to ensure a reasonable assurance of safety and effectiveness' for digital health apps remain underdeveloped [4]. Many apps marketed as tools for health improvement are categorized as health and wellness products, which exempts them from FDA regulation [5]. To address these gaps, the FDA recently concluded its precertification pilot program, aimed at streamlining the evaluation of new apps within resource constraints. Despite its innovative approach, the program yielded mixed results, hindered by the lack of formal legislative authority to enforce and implement new evaluation criteria [6]. Recognizing the need for standardized quality assessment, numerous frameworks for evaluating health apps have been proposed. Defined as criteria-based tools for assessing the quality of mobile health apps, a recent review identified 44 distinct frameworks, each varying in evaluation methodologies and quality criteria [7].

One of the most widely used frameworks is the American Psychiatric Association (APA) app evaluation model, which uses a hierarchical structure to assess apps based on key quality criteria: access and background, privacy and safety, clinical foundation, usability, and data integration toward therapeutic goals [8]. To operationalize this model, our team developed the M-Health Index & Navigation Database (MIND) platform—a user-friendly database that provides more than 100 yes-or-no answers to critical evaluation questions [9]. These include whether an app costs money to download, has a privacy policy, is supported by studies, offers real-time responses to enhance engagement, or integrates with electronic health records [10].

MINDApps has already been used in numerous studies evaluating the quality of mental health apps, establishing its quality standards as an excellent control group for comparison [10-12]. The quality of cardiovascular care apps has yet to be evaluated by MINDApps. Mobile health apps offer 2 key potential benefits in cardiology: they empower and engage individuals to make healthier lifestyle changes, and they enhance patient monitoring through features like electrocardiography and blood pressure tracking via wristbands and watches connected to an app [13]. Wearable devices integrated with apps are widely regarded as the future of cardiology, marking a clear shift toward more digitized and patient-centered care in the field [14-16]. With an increasing number of people relying on these apps and their functionality, it is crucial to evaluate their effectiveness, safety, and privacy standards. This is particularly important given recent literature indicating that the evidence base for these apps and the integration of wearable data into clinical cardiology are still in their early stages [14].

We conducted a systematic review of the Google Play Store and Apple App Store to identify smartphone apps used in cardiovascular care. Using the MINDApps framework, we evaluated these apps to assess their clinical foundation, privacy standards, security measures, platform compatibility, and cost. The results were then compared to the quality criteria of a comparable sample of mental health apps. We hypothesized that only a minority of both cardiovascular care and mental health apps would be supported by evidence, that the majority of apps would share personal health data with third-party companies, and that cardiovascular care apps would be used in external devices more frequently than mental health apps. We aim to identify and compare the limitations in both mental health and cardiovascular care apps by extending the application of the MINDApps framework to highlight areas within the broader digital health landscape that require improvement.

Methods

App Identification and Selection Criteria

The Apple App Store and Google Play Store were queried the for cardiovascular apps from January 19, 2023, to March 16, 2023, using the search terms "Blood pressure," "Heart," and "Cardiovascular," yielding a total of 159 apps across both sites. Terms were selected to mimic patient and provider search terms when identifying a potential assistive app. Apps were eligible for inclusion if they were updated within the past 90 days and in English. We excluded apps that required referral from a health care provider, had not been updated in the past 90 days, or were otherwise not functional. These limits were applied to ensure user accessibility, app functionality, and recency. Information was collected via app home pages and by trialing each app.

Cardiology apps were then rated using the MINDApps framework. Cardiology apps were matched to a subset of mental health apps from the MIND database by considering platform compatibility, such as whether the apps were available only on Android, only on iOS, on both Android and iOS, or across Android, iOS, and web platforms. To closely mimic the functionality and intention of cardiology apps, the mental health apps were chosen from the MINDApps database, which included journaling/mood trackers as they reflect the functionality of cardiology apps better than apps designed to provide a psychotherapy intervention. Additionally, apps were matched based on their cost structure (totally free vs some form of payment) and platform available (Apple App Store, Google Play Store, or both). No citation chaining, registry searching, or developer contacting was conducted, as this study focused solely on publicly available app marketplace content.

Evaluation Framework and Statistical Analysis

MINDApps is the largest publicly available database of mental health apps [17]. The index consists of 105 objective questions (yes/no) based on the APA's app evaluation model and one

XSL•FO
RenderX

free-text section for reviewers to add any additional information [9]. Apps are evaluated across six categories: (1) app origin and accessibility, (2) privacy and security, (3) clinical foundation, (4) features and engagement, (5) inputs and outputs, and (6) interoperability (see Multimedia Appendix 1 for the full list of questions). Reviews are conducted by evaluators who have undergone interrater reliability training. App ratings are updated every 6 months recognizing the fast dynamic of the health app marketplace. Using the MINDApps framework, reviewers having undergone the same interrater reliability training evaluated the identified cardiovascular care apps by applying its 105 questions, which assess key app features such as privacy policies, clinical foundation, and engagement. For example, one question asks whether the app includes a privacy policy, allowing us to determine the proportion of cardiovascular care apps that meet this basic privacy standard. We then compared the results with those of mental health apps to identify differences in quality criteria. Statistical differences between the 2 groups were assessed using two-proportion Z-tests.

Results

App Overview

We identified 48 cardiovascular care apps on the Apple App store and Google Play store, which met the eligibility criteria (Table 1). These apps were compared to 48 mental health apps derived from MINDApps. Refer to Multimedia Appendix 2 for the full list of included apps. The app selection process is shown in Figure 1.

Table . Comparing the quality criteria of cardiovascular care and mental health apps.

Category	Cardiology apps, n (%)	Mental health apps, n (%)	*P* value
Total number of apps	48 (100)	48 (100)	—[a]
Data security			
No privacy policy	6 (12.50)	5 (10.42)	>.99
Sharing PHI[b] with third party	30 (62.50)	28 (58.33)	.68
Cost			
Completely free option	12 (25.00)	11 (22.92)	.81
For profit	48 (100)	44 (91.67)	.04[c]
Platform			
Android	13 (27.08)	8 (16.67)	.22
iOS	22 (45.83)	17 (35.42)	.30
Android + iOS	13 (27.08)	21 (43.75)	.09
Data collection			
Survey/diary	39 (81.25)	46 (95.83)	.03[c]
Microphone	2 (4.17)	18 (37.50)	<.001[d]
Step count	5 (10.42)	3 (6.25)	.46
Camera	21 (43.75)	19 (39.58)	.68
Requires external device	14 (29.17)	6 (12.50)	.04[c]
Reports summary of data	44 (91.67)	44 (91.67)	>.99
Email or export data	39 (81.25)	32 (66.67)	.10
Clinical foundation	2 (4.17)	5 (10.42)	.24
Features			
Biodata	30 (62.50)	0 (0)	<.001[d]
Biodata with feedback	12 (25.00)	1 (2.08)	.001[e]
Symptom tracking	34 (70.83)	47 (97.92)	<.001[d]

[a]Not applicable.

[b]PHI: patient health information.

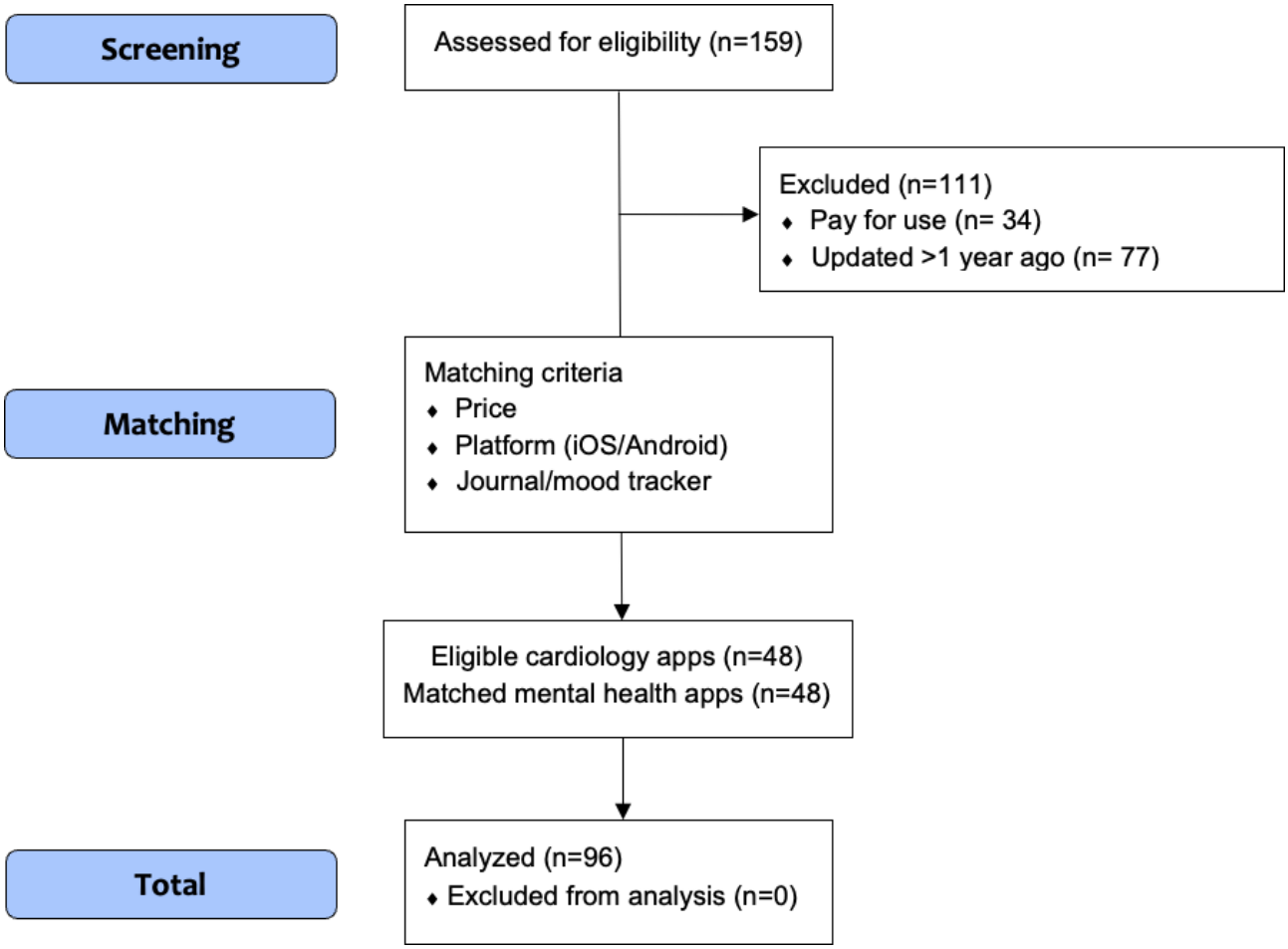
[c]*P*<.05.

[d]*P*<.001.

[e]*P*<.01.

XSL•FO
RenderX

Figure 1. App selection Process.

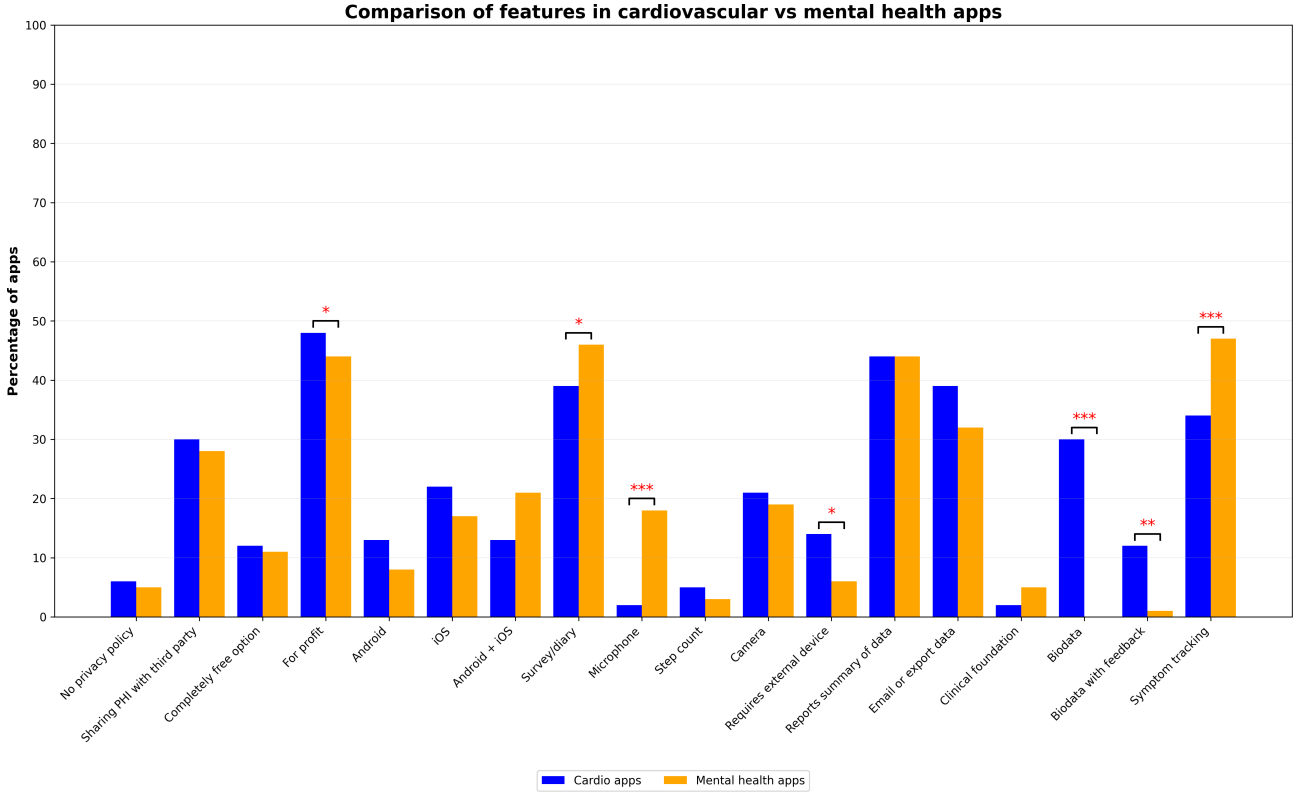


Data Security and Privacy

The majority of apps in both the cardiovascular care and mental health samples had a privacy policy. However, many of these apps disclosed user data to third-party companies (Table 1). A significantly higher proportion of cardiovascular care apps was developed by for-profit companies compared to the mental health app sample (P=.04; Figure 2): 44 of all 48 (92%) mental health apps were developed by a for-profit company, while all of the 48 identified cardiovascular care apps originated from a for-profit company.

XSL•FO
RenderX

Figure 2. Comparison of features in cardiovascular vs mental health apps. PHI: patient health information. *P<.05, **P<.01, ***P<.001.



Features and Functionality

Apps in both samples used similar features to assess user data. However, it was significantly more likely for mental health apps to collect speech data using the smartphone's microphone than it was for the cardiovascular care apps (P<.001). Cardiovascular care apps on the other hand had a higher likelihood of being connected to an external device (P=.04). In terms of features, a significantly higher proportion of cardiovascular care apps used biodata (P<.001), whereas mental health apps were significantly more likely to include symptom tracking capabilities (P<.001).

Clinical Foundation and Evidence

A minority of apps in both samples provided evidence for their effectiveness, with only 2 of all 48 (4%) cardiovascular care apps and 5 of all 48 (10%) mental health apps meeting this criterion.

Discussion

Summary of Key Findings

Both cardiovascular and mental health apps suffer from limited clinical validation and substantial data sharing practices. While cardiovascular apps excel in integration with external devices and bio capture, they lack evidence in support and transparency. Both classes of apps lack supporting clinical evidence while including a large price tag, highlighting the importance of need for regulation and counseling for integrated patient care use.

Limitations in Technology Integration and Data Privacy in Health Apps

We found that the current apps targeted to cardiovascular conditions and mental health both fail to leverage new technology for data collection. The majority of cardiology and mental health apps collect data through diaries and surveys, which rely on self-report and/or users being able to obtain their vitals independently from the app. Though monitoring and tracking symptoms in real time can be beneficial to patients, these "digital notebooks" do not robustly use functionalities that are unique to smartphones (ie, microphones, cameras, and sensors) [18-20]. In behavioral health, research teams continue to innovate how the sensors in smartphones can be used to monitor patients' symptoms, behaviors, and environment in real time; yet, there is little translation from the academic literature to marketplace options [21-23]. Integrating the data streams and functionalities that are unique to smartphones can help marketplace options present more innovative interventions for health conditions [18,21-23].

A major concern with health apps, which is reflected in our hypotheses, is the sharing of patient health information with third-party companies, which would typically be protected in standard health care settings [24-26]. Our findings highlight this issue, with the majority of both cardiovascular and mental health apps sharing data with third-party companies. Recent investigations on mental health apps reported similar results [10]. In our study, we demonstrate that this issue is not exclusive to the field of mental health, as cardiovascular care apps also face significant challenges in ensuring data ownership. The lack of evidence for supporting effectiveness is another frequently reported issue [27,28]. With only about one-tenth of mental health apps and fewer than one-twentieth of cardiovascular care

XSL•FO
RenderX

apps in our sample providing evidence for their effectiveness, it becomes clear why the health app marketplace is often likened to a "wild west" of medical software [29,30]. Cardiovascular care apps are more likely to integrate with external devices than are mental health apps, which is unsurprising given that many are equipped with smartphone-based electrocardiograms or blood pressure monitoring functions [31,32]. These external devices have the potential to revolutionize patient monitoring in cardiology. However, without app developers providing evidence of their feasibility and efficacy, we remain cautious about the current status of health app adoption in this field.

Future Directions and Quality Assurance in Health Apps

According to data from the Medical Expenditure Panel Survey, heart disease in the United States incurred an estimated cost of $229 billion between 2017 and 2018, encompassing health care services, medications, and lost productivity [33]. Digital health apps may offer opportunities to mitigate this public health challenge by providing more streamlined access to care. The current app ecosystem, however, exhibits significant shortcomings. Despite the concerns, we are convinced that health apps will continue to grow in prominence in society and health care. As medicine becomes increasingly digitized, ensuring its quality assurance is indispensable. App libraries such as MINDApps are able to alleviate concerns and increase trust in health apps. Patients have expressed a desire for app libraries to provide personalized recommendations tailored to their needs [34]. MINDApps already provides personalized recommendations in the field of mental health, and in this study, we demonstrate that its framework can be effectively applied to other medical fields. We present the first use case of MINDApps being adapted to cardiovascular health apps, demonstrating that this framework can be effectively extended beyond mental health to categorize available apps and assist patients and consumers in making informed decisions when selecting health apps.

Limitations

This study has several limitations: we included only apps that provided content in English, which may limit generalizability. Additionally, by including only apps updated within the last 90 days, there is a potential for selection bias and, thus, this sample may not represent the full range of available apps, potentially excluding newer or less popular options that offer better or different features. Furthermore, as a cross-sectional study, the included apps may have been updated or modified since the time of analysis and publication. While this study provides insights into the current state of app quality and features, it cannot determine causality or the long-term effectiveness of whether the use of these apps leads to sustained improvements in health outcomes over time.

Conclusion

Our findings reveal that the current app marketplace for cardiovascular and mental health on both the Apple App Store and Google Play Store is built on a limited clinical foundation and presents significant privacy and security concerns. We argue that there is need for better-quality assurance of health apps. MINDApps provides valuable guidance in identifying effective and secure apps tailored to the individual preferences of both clinicians and users. In this study, we demonstrate that this capability extends beyond the field of mental health, showing that the framework can be successfully applied to other medical domains as well.

Conflicts of Interest

JT is a Scientific Advisor of Precision Mental Wellness. JT also serves as editor-in-chief of *JMIR Mental Health*. DBK provides consulting to Whoop Inc and the Peterson Health Technology Institute. There are no other relevant conflicts of interest to declare.

Multimedia Appendix 1
List of questions.
[DOCX File, 32 KB - mhealth_v13i1e63642_app1.docx]

Multimedia Appendix 2
List of included apps.
[DOCX File, 32 KB - mhealth_v13i1e63642_app2.docx]

References

1. Kario K, Nomura A, Kato A, et al. Digital therapeutics for essential hypertension using a smartphone application: A randomized, open-label, multicenter pilot study. J Clin Hypertens (Greenwich) 2021 May;23(5):923-934. [doi: 10.1111/jch.14191] [Medline: 33484628]
2. Kasperbauer TJ, Wright DE. Expanded FDA regulation of health and wellness apps. Bioethics 2020 Mar;34(3):235-241. [doi: 10.1111/bioe.12674] [Medline: 31633831]
3. Wells C, Spry C. An Overview of Smartphone Apps: CADTH Horizon Scan: Canadian Agency for Drugs and Technologies in Health; 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK595384/ [accessed 2025-10-14]
4. Boulton KA, Hilton M, Sutton E, Guastella AJ. Apps and digital resources for child neurodevelopment, mental health, and well-being: Review, evaluation, and reflection on current resources. J Med Internet Res 2025 Jan 1;27:e58693. [doi: 10.2196/58693] [Medline: 39742455]

XSL•FO
RenderX

5. Camacho E, Cohen A, Torous J. Assessment of mental health services available through smartphone apps. JAMA Netw Open 2022 Dec 1;5(12):e2248784. [doi: 10.1001/jamanetworkopen.2022.48784] [Medline: 36576737]
6. Kidman PG, Curtis RG, Watson A, Maher CA. When and why adults abandon lifestyle behavior and mental health mobile apps: Scoping review. J Med Internet Res 2024 Dec 18;26:e56897. [doi: 10.2196/56897] [Medline: 39693620]
7. Patel MS, Small DS, Harrison JD, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: The STEP UP randomized clinical trial. JAMA Intern Med 2019 Dec 1;179(12):1624-1632. [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]
8. Stoeckl SE, Torres-Hernandez E, Camacho E, Torous J. Assessing the dynamics of the mental health Apple and Android App marketplaces. J Technol Behav Sci 2023 Jan 25;8(4):1-8. [doi: 10.1007/s41347-023-00300-x] [Medline: 36712910]
9. Mittal A, Elkaldi Y, Shih S, Nathu R, Segal M. Mobile electrocardiograms in the detection of subclinical atrial fibrillation in high-risk outpatient populations: Protocol for an observational study. JMIR Res Protoc 2024 May 27;13:e52647. [doi: 10.2196/52647] [Medline: 38801762]
10. Stern AD, Brönneke J, Debatin JF, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health 2022 Mar;4(3):e200-e206. [doi: 10.1016/S2589-7500(21)00292-2] [Medline: 35216754]
11. Velmovitsky PE, Kirolos M, Alencar P, Leatherdale S, Cowan D, Morita PP. Leveraging mHealth technologies for public health. JMIR Public Health Surveill 2024 Sep 12;10:e49719. [doi: 10.2196/49719] [Medline: 39265164]
12. Ramos G, Ponting C, Labao JP, Sobowale K. Considerations of diversity, equity, and inclusion in mental health apps: A scoping review of evaluation frameworks. Behav Res Ther 2021 Dec;147:103990. [doi: 10.1016/j.brat.2021.103990] [Medline: 34715396]
13. Nichol AA, Sankar PL, Halley MC, Federico CA, Cho MK. Developer perspectives on potential harms of machine learning predictive analytics in health care: Qualitative analysis. J Med Internet Res 2023 Nov 16;25:e47609. [doi: 10.2196/47609] [Medline: 37971798]
14. Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing mental health apps marketplaces with objective metrics from 29,190 data points from 278 apps. Acta Psychiatr Scand 2021 Aug;144(2):201-210. [doi: 10.1111/acps.13306] [Medline: 33835483]
15. Motta-Yanac E, Riley V, Ellis NJ, Mankoo A, Gidlow CJ. The digital prescription: A systematic review and meta-analysis of smartphone apps for blood pressure control. Int J Med Inform 2025 Mar;195:105755. [doi: 10.1016/j.ijmedinf.2024.105755] [Medline: 39657401]
16. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digit Health 2019 Jun;1(2):e52-e54. [doi: 10.1016/S2589-7500(19)30013-5] [Medline: 33323229]
17. Agency for healthcare research and quality. Medical expenditure panel survey (MEPS): Household component summary tables: Medical conditions, United States. URL: https://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp [accessed 2025-10-14]
18. Torous J, Stern AD, Bourgeois FT. Regulatory considerations to keep pace with innovation in digital health products. NPJ Digit Med 2022 Aug 19;5(1):121. [doi: 10.1038/s41746-022-00668-9] [Medline: 35986056]
19. Hughes A, Shandhi MMH, Master H, Dunn J, Brittain E. Wearable devices in cardiovascular medicine. Circ Res 2023 Mar 3;132(5):652-670. [doi: 10.1161/CIRCRESAHA.122.322389] [Medline: 36862812]
20. APP Advisor. American Psychiatric Association. 2024. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model [accessed 2025-10-14]
21. Bhaltadak V, Ghewade B, Yelne S. A comprehensive review on advancements in wearable technologies: Revolutionizing cardiovascular medicine. Cureus 2024 May;16(5):e61312. [doi: 10.7759/cureus.61312] [Medline: 38947726]
22. Torous J, Benson NM, Myrick K, Eysenbach G. Focusing on digital research priorities for advancing the access and quality of mental health. JMIR Ment Health 2023 Apr 24;10:e47898. [doi: 10.2196/47898] [Medline: 37093624]
23. Tumuhimbise W, Theuring S, Kaggwa F, et al. Enhancing the implementation and integration of mHealth interventions in resource-limited settings: a scoping review. Implement Sci 2024 Oct 14;19(1):72. [doi: 10.1186/s13012-024-01400-9] [Medline: 39402567]
24. Williams GJ, Al-Baraikan A, Rademakers FE, et al. Wearable technology and the cardiovascular system: the future of patient assessment. Lancet Digit Health 2023 Jul;5(7):e467-e476. [doi: 10.1016/S2589-7500(23)00087-0] [Medline: 37391266]
25. Alon N, Stern AD, Torous J. Assessing the Food and Drug Administration's risk-based framework for software precertification with top health apps in the United States: Quality improvement study. JMIR Mhealth Uhealth 2020 Oct 26;8(10):e20482. [doi: 10.2196/20482] [Medline: 32927429]
26. Yang Q, Van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: Meta-analytic review. JMIR Mhealth Uhealth 2019 Apr 3;7(4):e11244. [doi: 10.2196/11244] [Medline: 30942695]
27. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226. [doi: 10.1037/prj0000130] [Medline: 25844912]

XSL•FO
RenderX

28. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of health and well-being apps and curated app portals: Think-aloud and interview study. JMIR Mhealth Uhealth 2021 Apr 27;9(4):e27173. [doi: 10.2196/27173] [Medline: 33904827]
29. Gagnon L. Time to rein in the “Wild West” of medical apps. CMAJ 2014 May 13;186(8):E247. [doi: 10.1503/cmaj.109-4772] [Medline: 24733763]
30. Hare AJ, Chokshi N, Adusumalli S. Novel digital technologies for blood pressure monitoring and hypertension management. Curr Cardiovasc Risk Rep 2021;15(8):11. [doi: 10.1007/s12170-021-00672-w] [Medline: 34127936]
31. Zych MM, Bond R, Mulvenna M, Martinez Carracedo J, Bai L, Leigh S. Quality assessment of digital health apps: Umbrella review. J Med Internet Res 2024 Oct 10;26:e58616. [doi: 10.2196/58616] [Medline: 39388698]
32. Minen MT, George A, Camacho E, et al. Assessment of smartphone apps for common neurologic conditions (headache, insomnia, and pain): Cross-sectional study. JMIR Mhealth Uhealth 2022 Jun 21;10(6):e36761. [doi: 10.2196/36761] [Medline: 35727625]
33. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med 2019 Jul 17;17(1):137. [doi: 10.1186/s12916-019-1377-7] [Medline: 31311535]
34. Katz ME, Mszar R, Grimshaw AA, et al. Digital health interventions for hypertension management in US populations experiencing health disparities: A systematic review and meta-analysis. JAMA Netw Open 2024 Feb 5;7(2):e2356070. [doi: 10.1001/jamanetworkopen.2023.56070] [Medline: 38353950]

Abbreviations

APA: American Psychiatric Association
FDA: Food and Drug Administration
MIND: M-Health Index & Navigation Database

Edited by L Buis; submitted 25.06.24; peer-reviewed by F Carrouel, L Schmidt; revised version received 13.07.25; accepted 25.09.25; published 13.11.25.

Please cite as:
Singh M, Herpertz J, Alon N, Perret S, Torous J, Kramer D
Smartphone Apps for Cardiovascular and Mental Health Care: Digital Cross-Sectional Analysis
JMIR Mhealth Uhealth 2025;13:e63642
URL: https://mhealth.jmir.org/2025/1/e63642
doi:10.2196/63642

© Manjot Singh, Julian Herpertz, Noy Alon, Sarah Perret, John Torous, Daniel Kramer. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Apps and Wearable Devices for Cardiovascular Health: Narrative Review

Gauri Kumari Chauhan[1], BSc; Patrick Vavken[2,3,4*], MD, MSc, MBA, PD; Christine Jacob[1*], PhD

[1]FHNW - University of Applied Sciences Northwestern Switzerland, Bahnhofstrasse 6, Windisch, Switzerland

[2]Vavken Health Lab, Zürich, Switzerland

[3]ETH, Zürich, Switzerland

[4]University of St. Gallen, St. Gallen, Switzerland

[*]these authors contributed equally

Corresponding Author:
Christine Jacob, PhD
FHNW - University of Applied Sciences Northwestern Switzerland, Bahnhofstrasse 6, Windisch, Switzerland

Abstract

Background: Cardiovascular diseases (CVDs) continue to be the leading cause of global morbidity and mortality. Aiming to reduce the risk of CVD development and better manage them, an increasing number of individuals are adopting mobile health (mHealth) apps and wearable devices (wearables). These technologies provide critical insights into heart health and fitness, supporting users to monitor their lifestyle behaviors and adhere to preventative medication.

Objective: In this review, we aimed to investigate the current state of mHealth apps and wearables designed for cardiovascular health, with a specific focus on the DACH region (Germany, Austria, and Switzerland). We assessed the benefits these technologies provide to clinicians and patients, particularly in addressing unmet needs like sex-specific symptoms, while also examining their potential integration into the broader health care ecosystem.

Methods: To identify heart health apps, a keyword search was performed on both the Swiss Apple App Store and Google Play Stores. A separate search was performed on Google to identify heart health wearables. The identified apps and wearables were evaluated using the foundational and contextual criteria of the sociotechnical framework for assessing patient-facing eHealth tools.

Results: After filtering out apps and wearables that did not meet our inclusion criteria, 20 apps and 22 wearables were included in the review. While all the apps were available in the DACH region, only 30% (6/20) were specifically designed for these countries. Only 25% (5/20) of the apps included sex-specific information; 40% (8/20) provided information from evidence-based research, 35% (7/20) provided general health information without academic and clinical references, and 25% (5/20) did not include any evidence-based or general health information. While 20% (4/20) of the included apps had clinical integration features such as clinician dashboards, only 10% (2/20) had the potential to effectively enhance clinician workflows. Privacy policies were present in 95% (19/20) of the apps, with 75% (15/20) adhering to General Data Protection Regulation (GDPR) regulations; 1 app had no data protection policy. Only 20% (4/20) of the apps were medically certified. For wearables, only 9% (2/22) were tailored to the DACH region, and 40% (9/22) addressed women's health. While around 60% (13/22) offered features to support clinical integration, only 9% (2/22) had the potential to improve clinical workflows. More than half (12/22) of the wearables were medically certified, and 77% (17/22) referenced scientific or peer-reviewed research. All wearables included a privacy policy.

Conclusions: While many mHealth tools for cardiovascular health are available, only a few provide meaningful value to both patients and clinicians or have the potential to integrate effectively into the health care system. Women's sex-specific needs are often overlooked, and the benefits for clinicians are limited. In addition, mHealth apps largely lack robust evidence, whereas wearables showed comparatively stronger support through evidence-based and medical certification.

(JMIR Mhealth Uhealth 2025;13:e65782) doi:10.2196/65782

KEYWORDS

eHealth; mobile health; mHealth; digital health; technology assessment; technology adoption; technology implementation; cardiovascular diseases; cardiovascular health; Germany; Austria; Switzerland; wearables; apps; smartphones; Swiss Apple App; Google Play Store; reviews; morbidity; mortality; well-being; care management; health technologies

XSL•FO
RenderX

Introduction

Background

Cardiovascular diseases (CVDs) are the leading global cause of death, responsible for approximately 17.9 million fatalities annually and affecting over 500 million people worldwide [1-3]. The widespread prevalence of CVD highlights the urgent need for effective diagnosis, treatment, management, and preventive measures [3]. A notable gap in cardiovascular health is the tendency of current approaches to CVD prevention, diagnosis, and treatment to overlook the physiological differences between sexes [4]. This has contributed to a lack of awareness regarding women's cardiovascular risks, resulting in delayed diagnoses and suboptimal care [5].

The advancement of digital technologies presents a significant opportunity to monitor and manage lifestyle factors and health risks, enabling the early detection and prevention of cardiovascular diseases; through innovative tools like telemonitoring, remote patient monitoring, mobile health (mHealth) apps, and wearable health devices, individuals can engage in proactive health management, providing a transformative approach to cardiovascular care [1]. The World Health Organization's Global Observatory of Electronic Health (eHealth) considers mHealth a subcategory of eHealth and defines it as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, Personal Digital Assistants (PDAs), and other wireless devices" [6]. The integration and widespread adoption of these technologies present a promising shift toward more personalized and proactive health care solutions, with the potential to enhance patient outcomes while simultaneously reducing health care costs [7].

Coorey et al [8] found that mHealth apps significantly reduce hospital readmission rates, improve blood pressure, support healthy dietary habits, and enhance cardiovascular disease management. Smartphones equipped with photoplethysmography technology can detect atrial fibrillation and assess cardiovascular health by measuring blood volume changes using infrared light, this technique provides insights into heart rate and variability, offering a cost-effective and noninvasive method for evaluating cardiovascular fitness [9].

Wearable devices, typically worn on the wrist, arms, chest, or hips, can also be effective tools for managing CVD risks [10]. These devices track various health metrics including heart rate, heart rate variability (HRV), blood oxygen levels, sleep patterns, and physical activity using photoplethysmography or electrocardiogram (ECG) technology to detect atrial fibrillation (AFib) [10,11]. For example, a study by Guo et al [12] involving over 187,000 users identified 265,139 potential AFib cases among 424 users using the smartwatch. Follow-up testing confirmed AFib in 227 out of 262 users, demonstrating that wearables can effectively alert users to potential AFib, prompting timely medical evaluation and early diagnosis. A systematic review of interventions using smartwatches revealed favorable outcomes across various health aspects including improvements in lifestyle changes, medication adherence, reduction in unplanned hospital readmissions, enhanced AFib diagnosis, and better adherence to self-monitoring practices [13].

In the DACH region (Germany, Austria, and Switzerland), cardiovascular diseases also rank as the top cause of death [14]. This region accounts for about one third of Europe's medical technology market, presenting a key opportunity for health care innovation and research [15]. With its growing focus on health tech, the DACH region is positioned as an emerging leader in advancing medical solutions, making it a compelling market for cardiovascular health interventions [14,15].

Objectives

In this review, we set out to investigate the landscape of mHealth apps and wearables designed for cardiovascular health globally, with a special focus on tools available in the DACH region. The objective was to identify potential gaps by assessing the value these digital health tools bring to both patients and clinicians. Given the well-documented gap in cardiovascular care related to women's health, we also aimed to explore whether the analyzed apps and wearables provide tailored support for sub-groups of users, such as addressing sex-specific symptoms. In addition, we aimed to assess the potential of these technologies to integrate into the broader health care ecosystem, with the goal of identifying areas that require improvement and opportunities for future innovation.

Methods

Search Strategy

The search for heart health apps involved a screening of the Swiss Apple's App Store and Google's Play Store. The Apple Store search was conducted using an iPhone, while Google Play Store apps were identified through its web-based platform. In addition, a supplementary Google search was performed using a combination of English and German keywords to ensure a comprehensive discovery of heart health apps. Keywords were: "heart," "health," "heart health," "cardiac," "pulse monitor," "heart rate," "heart monitor," "best heart apps," "fitness app," "Herz-Apps," "Herzmonitor," and "Herzfrequenz." The search process spanned from April 8, 2024, to June 20, 2024, during which time each app was downloaded, reviewed, and evaluated.

The screening for wearables took place via Google search aimed at identifying the most relevant heart health wearables available both in the DACH region and globally. The keywords "heart wearables," "heart device tracking," "health wearables," "healthcare wearable devices," "smart wearables," "fitness wearables," "heart health tracking," "medical trackers," "DACH herzmonitor," and "Herzmonitor" were used. The search for wearables occurred between April 8, 2024, and July 11, 2024, during which time each vendor's website and accompanying app if available were thoroughly reviewed, and evaluated.

Inclusion and Exclusion Criteria

The included apps were required to meet one of two criteria: (1) measure or log health and well-being, typically categorized as health and fitness apps, (2) measure specific metrics such as heart rate and heart rate variability. Apps that did not meet at least one of these criteria were excluded.

XSL•FO
RenderX

Similarly, the selection of wearables was based on two criteria: (1) focus on health and wellness features, (2) possess the capability to measure health vitals such as heart rate, heart rate variability, or ECG. Only wearables meeting both criteria were included. The inclusion criteria of the mobile apps and wearables included in this review are outlined in Textbox 1.

Textbox 1. Inclusion criteria of the identified mobile apps and wearable devices

The included mobile apps were required to meet at least 1 of the following criteria:

- Health and well-being tracking: apps that measured or logged health and well-being were typically categorized as health and fitness apps.
- Specific metric measurement: apps that measured specific metrics, such as heart rate and heart rate variability.

The included wearable devices were required to meet both of the following criteria:

- Health and wellness features: wearables offering features related to health and wellness.
- Health vitals measurement: wearables had to be capable of measuring key health vitals, such as heart rate, heart rate variability, or ECG.

Analysis of Included Apps and Wearables

Each app that met the criteria was downloaded, analyzed, and evaluated. In addition, each website associated with the identified wearables, as well as the accompanying apps where available, were reviewed and downloaded. The features detailed on the wearable companies' websites were then used to assess their core functionalities and offerings. The first author (GC) devoted 2 - 3 hours to testing and evaluating each included app and wearable. The assessments were then reviewed by the last author (CJ), and any discrepancies or disagreements were collaboratively discussed and resolved between them.

An app or wearable was deemed DACH-specific based on 2 key criteria. Firstly, it had to be explicitly designed for the German-speaking region and the DACH countries. While simply providing an app or wearable in German did not automatically qualify it as DACH-specific, those developed by creators from Germany, Austria, or Switzerland, with a focus on their respective health care systems, were deemed as such. Secondly, the app or wearable was considered DACH-specific if the company behind it was based in one of the DACH countries.

The analysis was guided by a sociotechnical framework for evaluating patient-facing eHealth tools, which highlights the significance of context in assessing health care technologies [16]. This framework considers critical factors affecting the adoption of mHealth solutions by both patients and clinicians and evaluates how these technologies fit within the broader health care ecosystem [16-20]. To determine the value for patients, the evaluation focused on the technical features, safety, and functionality of each app and wearable from the patient's perspective. For clinicians, the analysis extended to include a review of the potential impact on clinical workflow, clinical utility, and the presence of a cost-benefit analysis to substantiate cost-efficiency claims. In addition, the evaluation criteria encompassed data protection, safety and regulatory compliance, interoperability and data sharing, revenue models, and certifications, to assess each tool's integration within the health care ecosystem. The assessment criteria for the mobile apps and wearables included in this review are outlined in Textbox 2.

XSL•FO
RenderX

Textbox 2. Assessment criteria for the included mobile apps and wearable devices.

General features

- Do they offer features relevant for CVD such as heart rate measurements?
- Are they DACH (Germany, Austria, and Switzerland) specific?
- Do they offer tailored support to underserved subgroups of users such as sex-specific symptoms?

Value for patients

- What are the main benefits for patients?
- Did they consider the whole patient journey? (eg, context, overall treatment path)
- Do they offer proper training and support material for patients?
- Do they provide valid clinical evidence to support their claims? (eg, published results of clinical trials, user research, real world evidence)

Value for clinicians

- What are the main benefits for clinicians? And for clinic or hospital management?
- Do they address the potential impact on clinical workflow?
- Do they offer proper training and support material for clinicians?
- Do they provide valid evidence to support their cost-efficiency claims? (eg, a cost-benefit analysis)

Fit into the ecosystem

- What is their business model? (eg, are they transparent about how they get paid and by whom?)
- Do they explicitly provide information about their data management and privacy policies?
- Do they address data sharing and interoperability?
- Does the tool require any additional infrastructure to function?
- Are they properly certified according to their risk tier? (eg, FDA [US Food and Drug Administration] approved, CE [Conformité Européenne] marked, certified as software as a medical device)

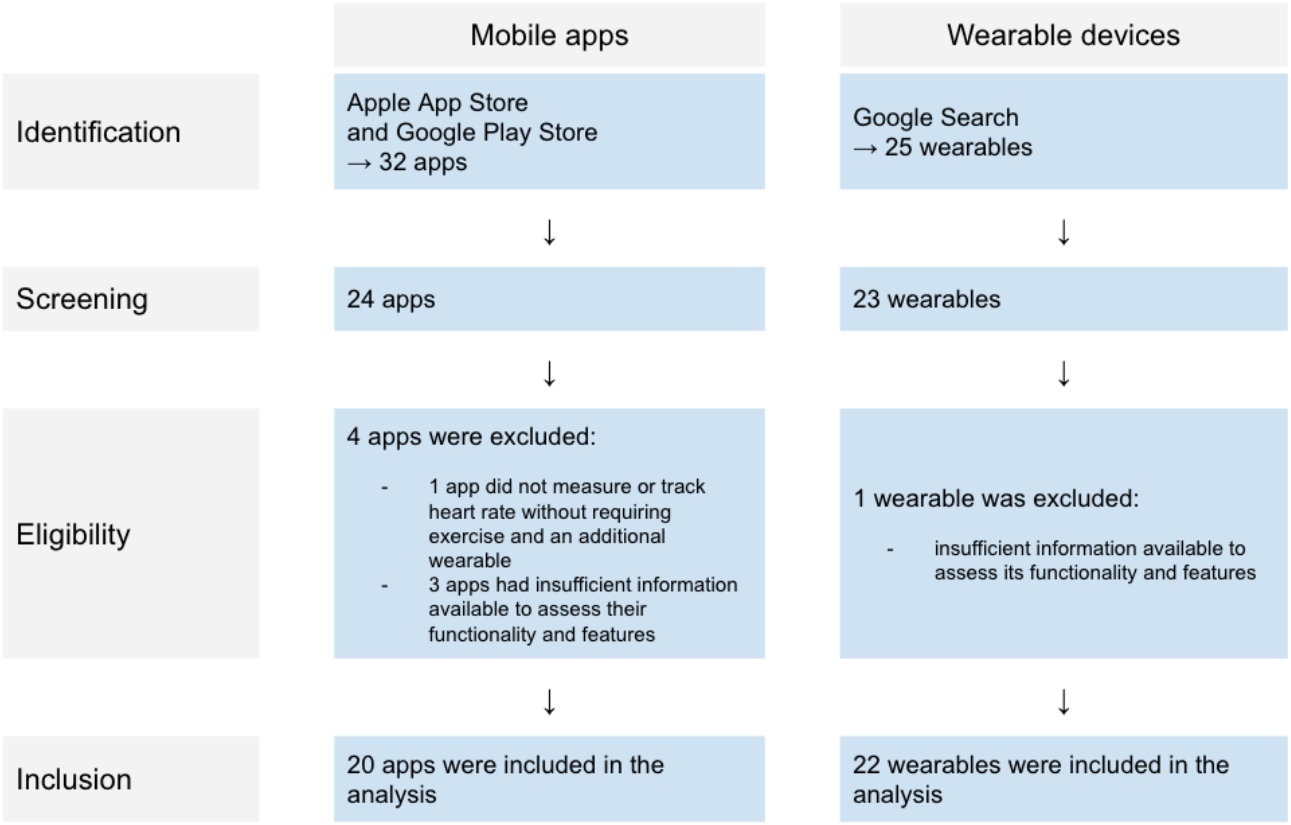
Results

Search Results

Only 20 out of 24 identified apps were included in the analysis. The 4 excluded apps were removed for the following reasons: one app did not measure or track heart rate without requiring exercise and an additional wearable; the remaining 3 apps had insufficient information available to assess their functionality and features. Out of 23 identified wearables, 1 wearable was excluded due to insufficient information available to assess its functionality and features, despite being intended for patient use. The selection diagram for apps and wearables is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Apps and wearables selection diagram.



Mobile Apps for Managing Cardiovascular Health

Ultimately, 20 apps were included according to the inclusion criteria. Only 30% (6/20) of the apps are tailored specifically for the DACH region, with just one, Preventicus Heartbeats GmbH, providing comprehensive heart health monitoring by measuring vital metrics, including heart rate, and integrating clinical research. This app is Conformité Européenne (CE)–certified and focuses on delivering valuable insights into heart health, particularly for women, which is not commonly addressed by other apps.

Our analysis highlighted that apps can serve either as heart rate measurement tools using photoplethysmography technology or as logbooks or tracking systems. Around 60% (12/20) of apps use the phone's photoplethysmography capabilities, offering a convenient way to measure heart rate without additional devices. Conversely, apps like Heart Analyzer (Helix Apps Ltd) and Cardiogram (Cardiogram Inc), which do not use photoplethysmography, rely on external devices for data collection. Furthermore, some apps with photoplethysmography functionality, such as Cardi Mate (Gismart), require users to watch advertisements before taking a measurement.

Regarding sex-specific considerations, only 25% (5/20) of the apps address women's specific health needs. Preventicus Heartbeats GmbH is notable for explicitly including women in its analysis, particularly menopausal women, and assessing the impact of hormonal changes on heart health. Other apps like Welltory (Welltory Inc) and Azumio's InPulse heart rate monitor include sex-specific considerations but do not provide as detailed an analysis.

Out of the reviewed apps, only 40% (8/20) are supported by scientific research to validate their claims. Meanwhile, 35% (7/20) provide general heart health information but lack direct research support. And 25% (5/20) include neither scientific evidence nor general health data, raising concerns about their credibility.

In terms of value for health care providers, 60% (12/20) of the apps allow users to share heart health reports with clinicians, which enhances their practical value in clinical settings. However, only 20% (4/20) offer platforms for clinicians to manage patient data or provide telehealth services, and just 10% (2/20) positively impact clinician workflows by facilitating appointment bookings, prescription management, or offering interoperability with patient health records.

Most apps (60%, 12/20) are interoperable with devices like Bluetooth-enabled blood pressure monitors or wearables. However, the ability to export and share data from the app itself is available in 75% (15/20) of the apps. The predominant revenue model is a free download with in-app purchases, used by 70% (14/20) of the apps. Only one app, ProHerz (ProCarement GmbH), is offering insurance coverage in Germany with a doctor's prescription.

Privacy is comprehensively addressed, with 95% (19/20) of apps including a privacy policy and 75% (15/20) ensuring General Data Protection Regulation (GDPR) compliance. Only 20% lack explicit GDPR mention but still provide privacy policies. Finally, 20% (4/20) of the apps are medically certified, with 3, Preventicus Heartbeats GmbH, CardioSignal, and FibriCheck, being clinically tested and capable of detecting

XSL•FO
RenderX

AFib. Multimedia Appendix 1 provides a detailed analysis of the included apps based on predefined assessment criteria.

Table 1 provides a consolidated overview of the characteristics of the analyzed apps. Figures 2 and 3 offer a detailed overview of the assessment of each included app, categorized into two distinct groups: “Heart Health Apps” and “Certified Apps.” The Heart Health Apps are distinguished by a red color coding, indicating that these apps either use photoplethysmography technology to measure heart rate directly through the smartphone or support data logging from other devices such as Bluetooth sensors or blood pressure monitors. These apps mostly function as trackers or logbooks for recording heart health metrics. Conversely, certified apps are highlighted in green, signifying that these apps possess formal certifications and evidence-based information. This includes various certifications such as the CE mark or FDA (US Food and Drug Administration) approval.

Table . Aggregated analysis of the included apps.

App characteristics	Value (N=20), n (%)
DACH[a] specific	6 (30)
Value for patients	
Measure heart rate	12 (60)
Medical diagnosis	3 (15)
Training and support	12 (60)
Women-specific measures	5 (25)
Value for clinician	
Receives heart report	12 (60)
Clinician integration	4 (20)
Positive impact on workflow	2 (10)
Training and support	3 (15)
Fit into the ecosystem	
Interoperable	12 (60)
Data sharing	15 (75)
Revenue model type	
Free download and no in-app purchase	5 (25)
Free download and with in-app purchase	14 (70)
Insurance covered	1 (5)
Evidence-based	
Research highlighted	8 (40)
General knowledge, no references	7 (35)
No research or scientific evidence	5 (25)
Privacy policy	
GDPR-[b] or FADP[c]-compliant	15 (75)
Privacy policy but not GDPR-compliant	4 (20)
No privacy policy	1 (5)
Certification	
Certified as a medical app	4 (20)
Not certified	16 (80)

[a]DACH: Germany, Austria, and Switzerland.

[b]GDPR: General Data Protection Regulation.

[c]FADP: Swiss Federal Act on Data Protection.

XSL•FO
RenderX

Figure 2. Analysis of each of the included apps – part I. BPM: beats per minute; DACH: Germany, Austria, and Switzerland; GDPR: General Data Protection Regulation; HR: heart rate; PPG: photoplethysmography.

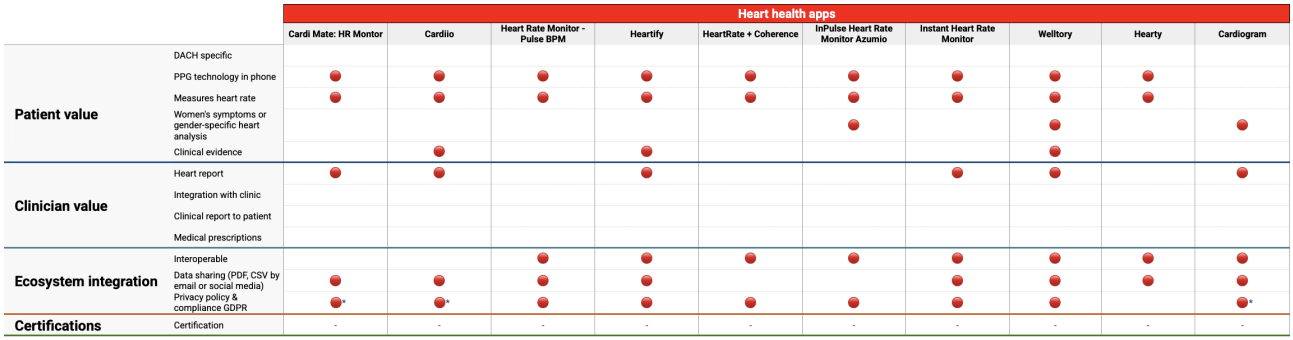
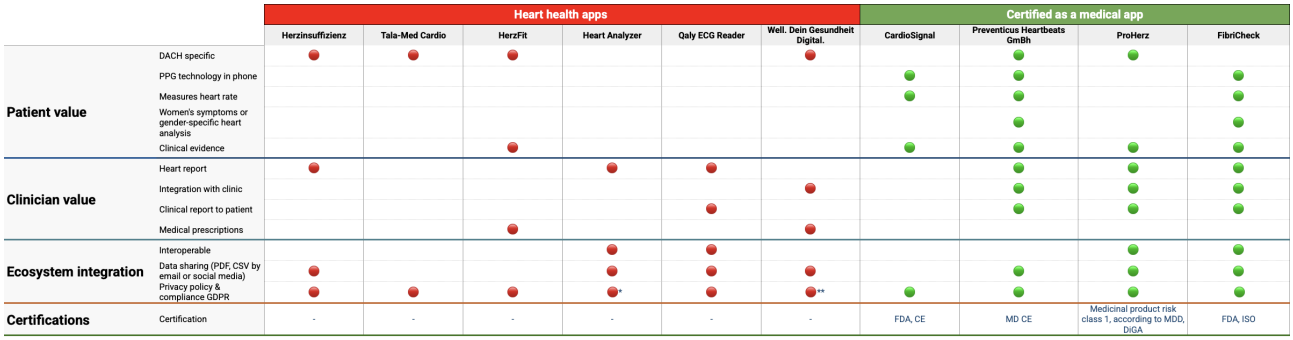


Figure 3. Analysis of each of the included apps – part II. CE: Conformité Européenne; DACH: Germany, Austria, and Switzerland; DiGA: digital health applications; ECG: electrocardiogram; FDA: US Food and Drug Administration; GDPR: General Data Protection Regulation; HR: heart rate; ISO: International Organization for Standardization; PPG: photoplethysmography.



Wearable Devices for Managing Cardiovascular Health

The analysis of wearables reveals that only 9% (2/22) are specific to the DACH region, with Aktiia (Aktiia SA) and SmartCardia (SmartCardia SA) representing Swiss companies focused on cardiovascular health. Aktiia emphasizes blood pressure measurement, while SmartCardia specializes in detecting AFib.

Among the wearables, all measure heart rate, but only 55% (12/22) provide reliable measures through medically certified devices. Photoplethysmography technology is used by 9 wearables, whereas 13 use ECG for detecting heart irregularities. Most wearables (82%, 18/22) offer comprehensive user instructions. However, only 41% (9/22) explicitly include features for women, such as menstrual cycle tracking or sex-specific heart rate analysis.

The Oura Ring (Oura Health Ltd) and Whoop (Whoop Inc) are particularly notable for their benefits to women. The Oura Ring tracks menstrual cycles and body temperature fluctuations, providing insights into heart health changes throughout the cycle. Whoop offers personalized pregnancy insights and tracks various health metrics through a daily journal.

From a clinical standpoint, wearables can be categorized into 3 distinct groups: fitness-centric, quality-certified, and medically certified. Fitness-centric wearables are primarily designed for general health and fitness purposes, emphasizing exercise and recovery rather than detailed heart health metrics. In contrast, quality-certified and medically certified wearables provide more comprehensive heart health reports, offering valuable data for clinical evaluation. However, a notable gap exists in the cost-benefit analysis provided by these wearables. Most lack robust evidence supporting their efficiency claims, with Philips Mobile Cardiac being the exception. Philips Mobile Cardiac stands out by presenting research that highlights its cost-effectiveness compared to alternative monitoring methods.

All of the included wearables had a privacy policy in place. Among them, 77% (17/22) were compliant with GDPR or Swiss Federal Act on Data Protection (FADP) standards, while 23% (5/22) included a privacy policy but did not explicitly mention GDPR compliance. While integration with clinical workflows varied: Fitness-centric wearables do not support clinician integration, some quality certified wearables offer APIs for data sharing, and medically certified wearables often have platforms for real-time remote monitoring. All wearables support data sharing with smartphones and allow the export of data, though medically certified devices often provide automatic synchronization.

Revenue models vary, with 50% (11/22) of wearables requiring a one-time purchase and 32% (7/22) using a subscription model. Over 75% (17/22) are backed by research or clinical trials, while the rest provide minimal or no scientific evidence. Privacy policies are universally provided, with most wearables ensuring GDPR or FADP compliance. About two-thirds of the wearables

XSL•FO
RenderX

have certifications, with 32% (7/22) being CE or ISO (International Organization for Standardization)–certified and 55% (12/24) holding medical certifications such as CE marks or FDA approvals. The distinction between certified and medically certified wearables often lies in the explicit proof of certification and medical approval. Multimedia Appendix 2 provides a detailed analysis of the included wearables based on the predefined assessment criteria.

Table 2 provides a consolidated overview of the characteristics of the analyzed wearables. Figures 4 and 5 offer a detailed overview of the assessment of each included wearable, categorized into 3 distinct groups: “Fitness-centric wearables,” “Quality certified wearables,” and “Medically certified wearables”. Fitness-centric wearables are represented in blue, indicating their primary focus on general fitness and training. Wearables with quality certification, marked in orange, offer detailed health measurements, including heart rate, blood oxygen levels, stress, and recovery, while also meeting quality certification standards such as FCC (Federal Communications Commission), DoC (Declaration of Conformity), or ISO. Wearables certified as medical devices are distinguished in purple, denoting their status as medical devices with regulatory clearance such as FDA clearance, or CE mark, enabling their use in clinical settings and providing a higher level of accuracy.

XSL•FO
RenderX

Table . Aggregated analysis of the included wearables.

Wearables characteristics	Value (N=22), n (%)
DACH[a] specific	2 (9)
Value for patients	
Measure heart rate	22 (100)
Medical diagnosis	12 (55)
Training and support	18 (82)
Women-specific measures	9 (41)
Value for clinician	
Receives heart report	18 (82)
Clinician integration	13 (59)
Positive impact on workflow	2 (9)
Training and support	9 (41)
Fit into the ecosystem	
Interoperable	22 (100)
Data sharing	22 (100)
Revenue model type	
One-off purchase of the wearable	11 (50)
Subscription-based app with wearable	7 (32)
Not stated	4 (18)
Evidence-based	
Research highlighted	17 (77)
General knowledge, no references	4 (18)
No research or scientific evidence	1 (5)
Privacy policy	
GDPR-[b] or FADP[c]-compliant	17 (77)
Privacy policy but not GDPR-compliant	5 (23)
No privacy policy	0 (0)
Certification	
Holds a quality certificate	7 (32)
Certified as a medical device	12 (55)
Not certified	3 (14)

[a]DACH: Germany, Austria, and Switzerland.

[b]GDPR: General Data Protection Regulation.

[c]FADP: Swiss Federal Act on Data Protection.

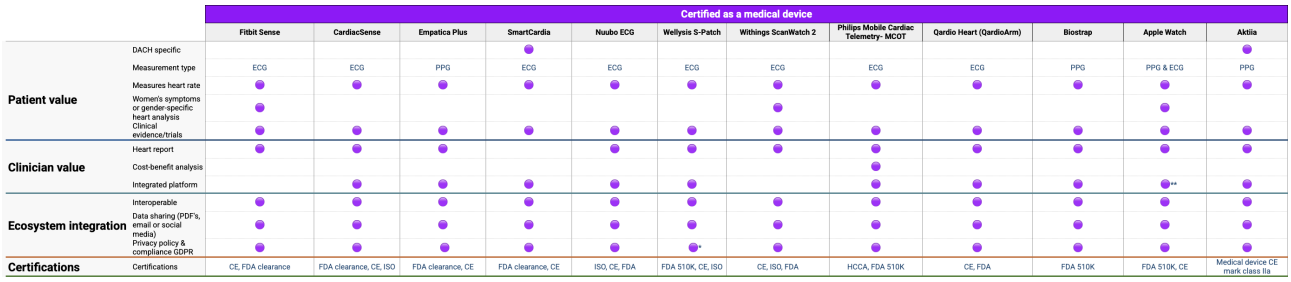
XSL•FO
RenderX

Figure 4. Analysis of each of the included wearables – part I. DACH: Germany, Austria, and Switzerland; DoC: Declaration of Conformity; ECG: electrocardiogram; FCC: Federal Communications Commission; GDPR: General Data Protection Regulation; ISO: International Organization for Standardization; PPG: photoplethysmography.

		Fitness-centric wearables			Holds a quality certificate						
		Amazfit GTR 4	Amazfit Helio Ring	GloryFit	Hexoskin Smart Shirt	Oura Ring	Fourth Frontier	Polar	Samsung Galaxy S6	Garmin Venu 3	Whoop
Patient value	DACH specific										
	Measurement type	PPG	PPG	N/A	ECG	PPG	ECG	ECG	ECG	PPG	PPG
	Measures heart rate	●	●	●	●	●	●	●	●	●	●
	Women's symptoms &/or cycles tracking or gender-specific heart analysis	●	●			●			●	●	●
	Clinical evidence/trials				●	●		●		●	●
Clinician value	Heart report	●	●		●	●	●	●	●	●	●
	Cost-benefit analysis										
	Integrated platform							●		●	
Ecosystem integration	Interoperable	●	●	●	●	●	●	●	●	●	●
	Data sharing (PDF's, email or social media)	●	●	●	●	●	●	●	●	●	●
	Privacy policy & compliance GDPR	●*	●*	●	●	●*	●	●	●**	●	●
Certifications	Certifications	-	-	-	FCC	DoC	ISO, DoC	DoC, ISO	FCC	ISO	DoC

Legend:
* Does not explicitly mention that GDPR is applied to the privacy policy
**Policy available and compliance with DPF (Data Privacy Framework - EU, UK, & CH)
N/A: data not available

Figure 5. Analysis of each of the included wearables – part II. CE: Conformité Européenne; DACH: Germany, Austria, and Switzerland; ECG: electrocardiogram; FDA: US Food and Drug Administration; GDPR: General Data Protection Regulation; HCCA: Health Care Compliance Association; ISO: International Organization for Standardization; PPG: photoplethysmography.



		Certified as a medical device										
		Fitbit Sense	CardiacSense	Empatica Plus	SmartCardia	Nuubo ECG	Wellysis S-Patch	Withings ScanWatch 2	Philips Mobile Cardiac Telemetry- MCOT	Qardio Heart (QardioArm)	Biostrap	Apple Watch	Aktiia
Patient value	DACH specific				●								●
	Measurement type	ECG	ECG	PPG	ECG	ECG	ECG	ECG	ECG	ECG	PPG	PPG & ECG	PPG
	Measures heart rate	●	●	●	●	●	●	●	●	●	●	●	●
	Women's symptoms or gender-specific heart analysis	●						●				●	
	Clinical evidence/trials	●	●	●	●	●	●	●	●	●	●	●	●
Clinician value	Heart report	●	●	●		●	●	●	●	●	●	●	●
	Cost-benefit analysis								●				
	Integrated platform		●	●	●	●	●		●	●	●	●**	●
Ecosystem integration	Interoperable	●	●	●	●	●	●	●	●	●	●	●	●
	Data sharing (PDF's, email or social media)	●	●	●	●	●	●	●	●	●	●	●	●
	Privacy policy & compliance GDPR	●	●	●	●	●	●*	●	●	●	●	●	●
Certifications	Certifications	CE, FDA clearance	FDA clearance, CE, ISO	FDA clearance, CE	FDA clearance, CE	ISO, CE, FDA	FDA 510K, CE, ISO	CE, ISO, FDA	HCCA, FDA 510K	CE, FDA	FDA 510K	FDA 510K, CE	Medical device CE mark class IIa

Legend:
* Does not explicitly mention that GDPR is applied to the privacy policy
**Only available in Canada, the UK, and the US

Discussion

Primary Findings

Our analysis showed that many apps do not provide a solid evidence base to substantiate their claims, and many apps lack medical certification and scientific validation, which may diminish trust and adoption by both users and clinicians. Conversely, wearables showed greater credibility, with many devices supported by evidence-based claims and validated through medical certifications. Privacy concerns also emerged as some apps fail to provide clear GDPR-compliant privacy policies, raising risks around data security and user trust. Usability remains a challenge, with some apps requiring additional devices for tracking, which can discourage engagement. Clinical integration faces significant barriers, including few clinician-specific features such as dashboards, and little information about interoperability and data sharing possibilities, limiting the effective integration of these tools in clinical practice. Furthermore, reimbursement remains a major hurdle, as most of the included apps and wearables rely on user payments with only 1 app that is insurance-covered.

Value for Patients

Our analysis highlights a significant presence of mHealth apps for managing heart health available in the DACH region. However, only 30% (6/20) of the apps are specifically tailored for this region, and just 20% (4/20) are medically certified. Only 40% (8/20) of the apps incorporate scientific research to validate their claims, while 35% (7/20) of the apps provide general heart health information without specific research backing. It is worth noting that 25% (5/20) of the apps lack any scientific or general health data, underscoring a notable gap in clinical research and evidence supporting these apps. This can prove problematic as the absence of clinical trials and scientific validation can undermine both user trust and clinician endorsement. The credibility of mHealth apps relies heavily on accessible, cited scientific research [21]. Without such evidence, users may question the app's reliability, and clinicians may hesitate to recommend these tools, impacting their adoption in health care [21]. Wearables, on the other hand, performed notably better in this area, with over 75% (17/22) supported by research or clinical trials, providing a stronger evidence base compared to the included apps.

From a usability perspective, our review indicates that 40% (8/20) of the analyzed apps do not enable users to measure their heart rate directly using their smartphones and instead require users to pair the app with an additional device, such as a wearable or a blood pressure monitor, to log and track their heart health. This additional requirement can be cumbersome and may discourage users from engaging with the app, as highlighted by research showing that manual tracking methods often reduce user readiness for self-monitoring [22]. Some apps in our analysis use the index finger for AFib measurements, others require users to lie down and place the phone on their sternum, which may not be practical in all situations, particularly if a rapid heart rate change occurs. Wearables offer a distinct

XSL•FO
RenderX

advantage by allowing heart vitals to be measured directly and automatically synced with a mobile app. Previous research showed that users prefer devices like Fitbit trackers for their seamless integration and automatic data synchronization [22]. This automatic syncing not only benefits users but also provides clinicians with valuable insights into the patient's activities, such as exercise intensity and duration, which are essential for accurate monitoring and assessment [22].

Value for Clinicians

Over half of the reviewed apps and wearables offer some level of clinician integration, including features like dashboards for remote patient monitoring and data reporting. However, many fall short of significantly improving clinicians' workflows. Key functionalities, such as appointment scheduling and e-prescription capabilities, are often missing, and important aspects like interoperability with electronic health record are not adequately addressed in the information provided on their websites and supporting documentation. This disconnect complicates processes for both patients and health care providers. Many clinicians are cautious about these tools due to their poor interoperability with existing health care systems and electronic health records, which limits their impact on workflow efficiency [19,20]. Furthermore, research indicates that patients are more likely to trust tools that are recommended by a physician [18,23]; this underscores the necessity of involving clinicians in the app's use to build trust and enhance its effectiveness.

The lack of scientific evidence, particularly among the identified apps (with wearables performing slightly better in this regard), may pose a challenge for clinician adoption. Previous research has shown that clinicians are often hesitant to recommend mHealth apps due to concerns about insufficient evidence and supporting research. A survey found that 62% of physiotherapists would not recommend apps lacking sufficient evidence and quality due to potential risks to patient health [24]. Similarly, 81.1% of general practitioners believe that mHealth apps should undergo clinical testing and receive certification from independent experts to ensure their effectiveness and reliability [25].

Our analysis suggests that medically certified apps and wearables are better-suited to support clinical integration, as the evidence base required for certification may enhance their credibility and acceptance among medical professionals. However, their effectiveness can still be hindered by several barriers. Achieving seamless integration into clinicians' workflows is vital to enable efficient remote monitoring and promote effective 2-way communication between patients and health care providers. Furthermore, challenges such as lack of compensation, increased workload, and insufficient digital literacy are significant obstacles for clinician adoption [19,20].

Fit Into the Ecosystem

According to the analysis, 75% of the reviewed apps explicitly state their GDPR compliance, while 20% do not mention it at all, and 1 app, the Hearty app (Pathfinder DMCC), lacks a privacy policy altogether. All of the included wearables featured a privacy policy. Of these, 77% (17/22) adhered to GDPR or FADP standards, ensuring compliance with data protection regulations. However, 23% (5/22) had a privacy policy in place but made no explicit reference to GDPR compliance, leaving room for ambiguity regarding their adherence to these standards. Previous research demonstrated the crucial role of privacy policies and data protection in building user trust, indicating that apps with privacy policies adhering to GDPR and employing high-security measures are viewed as more trustworthy [21].

GDPR regulations require that privacy policies be prominently displayed on an app's landing page and remain accessible throughout use [26]. However, some heart health apps analyzed in this review fall short of these transparency standards, making it harder for users to access their privacy policies. Noncompliance with GDPR can erode user trust, as mHealth apps often handle sensitive personal data shared with clinicians, family members, or third parties [21]. This kind of noncompliance poses risks like data breaches and unauthorized access to personal information, which can adversely affect user perceptions [27]. Apps with unclear or insecure privacy policies may face adoption barriers [28].

Our analysis highlights a significant gap in demonstrating cost reimbursement for eHealth tools. Among the included apps, only one is covered by health insurance, while the majority rely on in-app purchases for funding. Wearables, on the other hand, are generally available either as one-time purchases or through subscription-based models. Despite evidence suggesting that mHealth wearables can reduce stroke frequency and lower costs per stroke for high-risk patients, the lack of a clear reimbursement model remains a major barrier [29]. Previous research indicates that while initial costs for using wearables to detect AFib may be higher, they ultimately lead to reduced stroke incidence and lower overall costs due to effective anticoagulation therapy [29].

Another significant gap lies in the cost-benefit analyses provided by the included apps and wearables, as most fail to offer robust evidence to substantiate their cost-efficiency claims. An exception is Philips Mobile Cardiac, which distinguishes itself by presenting research that demonstrates its cost-effectiveness when compared to other monitoring methods. This adds to the challenges faced by health care providers in integrating mHealth tools primarily due to reimbursement issues [30]. The complex reimbursement pathways and lack of incentives for remote care significantly hinder the adoption of these technologies, particularly when their cost-efficiency claims are not supported by robust evidence [31].

Limitations and Future Research

In our review, we faced several limitations that should be acknowledged. The rapid evolution of technology means that the apps and wearables included in our assessment may have undergone updates or feature changes since the research was conducted, potentially altering their functionalities. In addition, some apps and wearables were inaccessible to us due to stringent verification requirements, such as restrictions limiting access to paying users only. While we evaluated whether the apps and wearables included statements about data security, we were unable to verify whether these tools effectively implemented secure data-handling practices. Moreover, despite our efforts

XSL•FO
RenderX

to use comprehensive search terms to capture as many relevant apps and wearables as possible, there remains a possibility that some significant tools were inadvertently excluded.

It is also important to note that our quality assessment was conducted as a high-level narrative review, focusing primarily on the features we could test after downloading the apps and information provided by developers through their websites and related documentation. This approach did not extend to verifying the success of real-world implementation, evaluating the quality of specific claimed functionalities (eg, integration into clinical workflows or interoperability), or assessing long-term performance in practical health care settings. Future studies could complement our initial findings by conducting comprehensive evaluations under real-world conditions. Such research would enable a deeper understanding of these tools' effectiveness, including their performance over extended periods and their impact on specific health care workflows, offering a more detailed comparison of their overall quality and value.

Conclusion

Our findings showed that mHealth apps and wearables hold considerable promise for improving the interaction between patients and clinicians in the management of cardiovascular diseases. We identified several strengths in these technologies, including accurate measurements, robust privacy policies, and useful clinical features. However, our analysis showed that the mHealth apps included in the study largely lack a solid foundation of scientific evidence. In contrast, wearables demonstrated a higher degree of credibility, with many supported by evidence-based claims and validated through medical certification.

Significant gaps remain, particularly in addressing women's cardiovascular health. A notable percentage of apps and wearables do not adequately meet the specific needs of women, who experience cardiovascular diseases differently than men. This underrepresentation in clinical research results in a lack of tailored diagnostic and treatment options, leading to less effective care. Our findings emphasize the need for health apps and wearables that are more inclusive of sex-specific cardiovascular needs throughout different life stages, such as pregnancy and menopause.

In addition, our results show that there is a critical need for better integration of these technologies into clinical practices through features that allow better integration into clinical workflows and interoperability with hospital IT systems. Our assessment indicates that most of the current mHealth tools do not offer enough evidence to improve clinician workflows sufficiently, limiting their ability to provide seamless and efficient care. Further research and case studies are required to demonstrate the financial benefits of these technologies as they mostly lack any evidence to support their cost-effectiveness claims. These gaps present a vital opportunity for innovation in mHealth, fostering stronger connections between patients and clinicians while ultimately improving heart health outcomes for all.

Acknowledgments

This study was part of a research project conducted by the first author (GKC), under the supervision of the last author (CJ). The research was commissioned by Vavken Health Labs, a startup founded by the second author (PV), which is dedicated to transforming health care through the development and implementation of innovative technologies.

Conflicts of Interest

CJ is an Associate Editor of *JMIR Human Factors* at the time of this publication. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Analysis of the included apps based on the predefined assessment criteria.
[XLSX File, 262 KB - mhealth_v13i1e65782_app1.xlsx]

Multimedia Appendix 2
Analysis of the included wearables based on the predefined assessment criteria.
[XLSX File, 114 KB - mhealth_v13i1e65782_app2.xlsx]

References

1. Tromp J, Jindal D, Redfern J, et al. World Heart Federation roadmap for digital health in cardiology. Glob Heart 2022;17(1):61. [doi: 10.5334/gh.1141] [Medline: 36051317]
2. Leading causes of death. National Center for Health Statistics, CDC. 2024. URL: https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm [accessed 2025-03-26]
3. Cardiovascular diseases. World Health Organization. 2024. URL: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 [accessed 2025-03-27]
4. Bairey Merz CN, Andersen H, Sprague E, et al. Knowledge, attitudes, and beliefs regarding cardiovascular disease in women: the women's heart alliance. J Am Coll Cardiol 2017 Jul 11;70(2):123-132. [doi: 10.1016/j.jacc.2017.05.024] [Medline: 28648386]

XSL•FO
RenderX

5. Al Hamid A, Beckett R, Wilson M, et al. Gender bias in diagnosis, prevention, and treatment of cardiovascular diseases: a systematic review. Cureus 2024 Feb;16(2):e54264. [doi: 10.7759/cureus.54264] [Medline: 38500942]
6. Global observatory for eHealth mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2025-03-28]
7. Singh PK, Kolekar MH, Tanwar S, et al. Emerging Technologies for Computing, Communication and Smart Cities: Springer; 2022. [doi: 10.1007/978-981-19-0284-0]
8. Coorey GM, Neubeck L, Mulley J, et al. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018 Mar;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
9. Castaneda D, Esparza A, Ghamari M, et al. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195-202. [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]
10. Scheid JL, Reed JL, West SL. Commentary: is wearable fitness technology a medically approved device? Yes and no. Int J Environ Res Public Health 2023 Jun 27;20(13):6230. [doi: 10.3390/ijerph20136230] [Medline: 37444078]
11. Cowie MR, Lam CSP. Remote monitoring and digital health tools in CVD management. Nat Rev Cardiol 2021 Jul;18(7):457-458. [doi: 10.1038/s41569-021-00548-x] [Medline: 33824486]
12. Guo Y, Wang H, Zhang H, et al. Mobile photoplethysmographic technology to detect atrial fibrillation. J Am Coll Cardiol 2019 Nov 12;74(19):2365-2375. [doi: 10.1016/j.jacc.2019.08.019] [Medline: 31487545]
13. Triantafyllidis A, Kondylakis H, Katehakis D, et al. Smartwatch interventions in healthcare: a systematic review of the literature. Int J Med Inform 2024 Oct;190:105560. [doi: 10.1016/j.ijmedinf.2024.105560] [Medline: 39033723]
14. Haas S, Camm JA, Harald D, et al. GARFIELD-AF: risk profiles, treatment patterns and 2-year outcomes in patients with atrial fibrillation in Germany, Austria and Switzerland (DACH) compared to 32 countries in other regions worldwide. Clin Res Cardiol 2023 Jun;112(6):759-771. [doi: 10.1007/s00392-022-02079-y] [Medline: 36094573]
15. Digital health in Europe: a record-breaking year for founders and investors in DACH. Speedinvest. 2022. URL: https://www.speedinvest.com/blog/digital-health-in-europe-record-breaking-2021-for-founders-and-investors-in-dach [accessed 2025-03-28]
16. Jacob C, Lindeque J, Müller R, et al. A sociotechnical framework to assess patient-facing eHealth tools: results of a modified Delphi process. NPJ Digit Med 2023 Dec 15;6(1):232. [doi: 10.1038/s41746-023-00982-w] [Medline: 38102323]
17. Jacob C, Lindeque J, Klein A, et al. Assessing the quality and impact of eHealth tools: systematic literature review and narrative synthesis. JMIR Hum Factors 2023 Mar 23;10:e45143. [doi: 10.2196/45143] [Medline: 36843321]
18. Jacob C, Sezgin E, Sanchez-Vazquez A, et al. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR mHealth uHealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
19. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR mHealth uHealth 2020 Feb 20;8(2):e15935. [doi: 10.2196/15935] [Medline: 32130167]
20. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR mHealth uHealth 2020 Jul 6;8(7):e18072. [doi: 10.2196/18072] [Medline: 32442132]
21. van Haasteren A, Gille F, Fadda M, et al. Development of the mHealth app trustworthiness checklist. Digit Health 2019;5:2055207619886463. [doi: 10.1177/2055207619886463] [Medline: 31803490]
22. Bezerra Giordan L, Ronto R, Chau J, et al. Use of mobile apps in heart failure self-management: qualitative study exploring the patient and primary care clinician perspective. JMIR Cardio 2022 Apr 20;6(1):e33992. [doi: 10.2196/33992] [Medline: 35442205]
23. Baldauf M, Fröehlich P, Endl R. Trust me, I'm a doctor – user perceptions of AI-driven apps for mobile health diagnosis. Presented at: MUM 2020; Nov 22-25, 2020; Essen, Germany. [doi: 10.1145/3428361.3428362]
24. Rowe M, Sauls B. The use of smartphone apps in clinical practice: a survey of South African physiotherapists. S Afr J Physiother 2020;76(1):1327. [doi: 10.4102/sajp.v76i1.1327] [Medline: 32391442]
25. Della Vecchia C, Leroy T, Bauquier C, et al. Willingness of French general practitioners to prescribe mHealth apps and devices: quantitative study. JMIR mHealth uHealth 2022 Feb 11;10(2):e28372. [doi: 10.2196/28372] [Medline: 35147508]
26. Muchagata J, Ferreira A. Translating GDPR into the mHealth practice. Presented at: 2018 International Carnahan Conference on Security Technology (ICCST); Oct 22-25, 2018; Montreal, QC, Canada. [doi: 10.1109/CCST.2018.8585546]
27. Lv Z, Qiao L. Analysis of healthcare big data. Future Gener Comp Sy 2020 Aug;109:103-110. [doi: 10.1016/j.future.2020.03.039]
28. Zhou L, Bao J, Watzlaf V, et al. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR mHealth uHealth 2019 Apr 16;7(4):e11223. [doi: 10.2196/11223] [Medline: 30990458]
29. Giebel GD. Use of mHealth devices to screen for atrial fibrillation: cost-effectiveness analysis. JMIR mHealth uHealth 2020 Oct 6;8(10):e20496. [doi: 10.2196/20496] [Medline: 33021489]
30. Lin CCC, Dievler A, Robbins C, et al. Telehealth in health centers: key adoption factors, barriers, and opportunities. Health Aff (Millwood) 2018 Dec;37(12):1967-1974. [doi: 10.1377/hlthaff.2018.05125] [Medline: 30633683]

XSL•FO
RenderX

31. Rajan B, Sainathan A, Agnihothri S, et al. The promise of mHealth for chronic disease management under different payment systems. Manuf Serv Oper Manag 2022 Nov;24(6):3158-3176. [doi: 10.1287/msom.2022.1143]

Abbreviations

AFib: atrial fibrillation
CE: Conformité Européenne
CVD: cardiovascular disease
DACH: Germany, Austria, and Switzerland
DoC: Declaration of Conformity
FADP: Swiss Federal Act on Data Protection
FCC: Federal Communications Commission
FDA: US Food and Drug Administration
GDPR: General Data Protection Regulation
HRV: heart rate variability
ISO: International Organization for Standardization
mHealth: mobile health

Edited by L Buis; submitted 26.08.24; peer-reviewed by E Patil, L Allan; revised version received 23.02.25; accepted 10.03.25; published 04.04.25.

Please cite as:
Chauhan GK, Vavken P, Jacob C
Mobile Apps and Wearable Devices for Cardiovascular Health: Narrative Review
JMIR Mhealth Uhealth 2025;13:e65782
URL: https://mhealth.jmir.org/2025/1/e65782
doi:10.2196/65782

© Gauri Kumari Chauhan, Patrick Vavken, Christine Jacob. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Trustworthiness of Web-Based Pharmacy Apps in Pakistan Based on the Mobile App Rating Scale: Content Analysis and Quality Evaluation

Anum Sattar[1,2*], PharmD, MPhil; Hina Rehman[3*], PharmD, MPhil, PhD; Safila Naveed[4*], PharmB, MPhil, PhD; Sumaira Khadim[5*], PharmD, MPhil, PhD; Nargis Khan[6*], MBBS, PhD; Ahmad Furqan Kazi[7], Pharm.D, Mphil; Wajid Syed[8], PhD; Mahmood Basil A Al-Rawi[9], PhD; Shazia Jamshed[10*], PharmB, MPhil, PhD

[1]Department of Pharmacy Practice, Jinnah University for Women, 5C, Block 5 Nazimabad, Karachi, Pakistan

[10]Department of Pharmacy Practice, School of Pharmacy, International Medical University, Kuala Lumpur, Malaysia

[2]Department of Pharmacy Practice, Faculty of Pharmacy and Pharmaceutical Sciences, Ziauddin University, Karachi, Pakistan

[3]Department of Pharmacy Practice, Institute of Pharmaceutical Sciences, Jinnah Sindh Medical University, Karachi, Pakistan

[4]Department of Pharmaceutical Chemistry, University of Karachi, Karachi, Pakistan

[5]Department of Pharmacy Practice, Faculty of Pharmacy, Iqra University, Karachi, Pakistan

[6]Department of General Medicine, Dow University of Health Sciences, Karachi, Pakistan

[7]Department of Pharmacy Practice, College of Pharmacy, Purdue University in Indianapolis, Indianapolis, United States

[8]Department of Clinical Pharmacy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

[9]Department of Optometry College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

[*]these authors contributed equally

Corresponding Author:
Anum Sattar, PharmD, MPhil
Department of Pharmacy Practice, Jinnah University for Women, 5C, Block 5 Nazimabad, Karachi, Pakistan

Abstract

Background: Web-based pharmacy apps facilitate the electronic exchange of health-related supplies. They are digital platforms that run on websites and smartphones. Pakistan is experiencing significant progress in smartphone integration and digital services, leading to the expansion of the online pharmacy business. However, concerns remain over the legitimacy and precision of these apps.

Objective: The aim of this study was to undertake a thorough assessment of digital pharmacy apps accessible in Pakistan. Specifically, our focus was on apps accessible via the Google Play Store and the iOS App Store. To fulfill this objective, an evaluation of these apps was performed using the Mobile App Rating Scale (MARS).

Methods: A research investigation was conducted to analyze the online pharmacy apps in Pakistan. Initially, 50 apps were identified, but 10 were excluded for not meeting pre-established criteria, 10 were excluded for being in languages other than English, and 7 could not be downloaded. All paid and non-English apps were also excluded. A total of 23 apps were selected for the study, acquired via the Google Play Store and iOS App Store. The evaluation was conducted by 2 researchers who maintained independence from one another by using the MARS.

Results: Initially, 50 apps were identified, of which 27 were excluded for not meeting the predetermined criteria. A total of 23 apps were selected for the study, acquired via the Google Play Store and iOS App Store. Strong positive correlations between higher user engagement and better app functionality and information quality were observed. The average rating of the 23 apps ranged between 2.64 and 4.00 on a scale up to 5. The aesthetics dimension had the highest mean score of 3.6, while the information dimension had the lowest mean score of 3.2. For credibility and reliability, different tests (intraclass correlation, Cohen κ, Krippendorff α, and Cronbach α) on each dimension of the MARS were performed by using SPSS Statistics 27. The intraclass correlation of all MARS dimensions ranged from 0.702 - 0.913 (95% CI 0.521 - 0.943), the Cohen κ of all MARS dimensions ranged from 0.388 - 0.907 (95% CI 0.151 - 0.994), the Krippendorff α of all MARS dimensions ranged from 0.705 - 0.979 (95% CI 0.657 - 0.923), and Cronbach α had a lower score of 0.821 in the information dimension and a higher score of .911 in the subjective quality dimension of the MARS.

Conclusion: This study evaluated online pharmacy apps in Pakistan by using the MARS. It is the first study on online pharmacy apps in Pakistan. The findings of the evaluation have provided insights into the reliability and efficacy of these apps.

(JMIR Mhealth Uhealth 2025;13:e59884) doi:10.2196/59884

XSL•FO
RenderX

KEYWORDS

online pharmacy; online pharmacy apps; apps; mobile app; smartphone app; trustworthiness; pharmacy; digital platforms; questionnaire; engagement; functionality; Mobile App Rating Scale; MARS; user; efficacy; Pakistan

Introduction

Background

Over the past decade, the internet has evolved into a primary source of knowledge for many individuals. The demand for purchasing things online, including medications, has experienced a parallel surge [1]. In the last few years, a discernible transformation has occurred within the pharmacy industry, whereby a considerable number of conventional pharmacies have adopted e-commerce platforms. These internet-based pharmacy firms engage in commercial activities through web-based platforms, enabling the immediate distribution of drugs and health care equipment to clients using postal services.

The field of online pharmacies has seen significant market value growth in the last few years. In the United States, online pharmacies have raised about $29.35 billion in 2014 and are projected to reach US $128 billion by 2023, reflecting a yearly increase of 17.7% [2]. The main factors influencing the customer preference for online pharmacies are convenience, affordability, and availability of drugs that may not normally be accessible or may be subject to cancellation [3,4]. The spread of mobile and wireless gadgets on a worldwide scale has generated an exceptional and incomparable potential for the provision of global medical care. According to data provided by the International Telecommunication Union, the global number of digital phone connections reached over 6 billion in 2012, representing an 86% market share globally. It is noteworthy that more than 70% of these customers were located in economically deprived nations such as Pakistan [5]. However, a significant number of customers residing in developing nations have shown reluctance toward accessing pharmacies on the internet. The reluctance to use such networks might be linked to multiple factors, which include a lack of awareness of their functionality, concerns regarding their reliability and authenticity, and limited accessibility. The main catalyst for the widespread use of e-pharmacies by most clients is expediency [6]. A survey reveals that over 80% of persons living in disadvantaged areas have use of mobile telephone connections. The geographical reach of these networks encompasses approximately 90% of the globe's population [7]. Internet pharmacies are rapidly growing in Pakistan. Organizations want to build trust and establish a favorable bond with customers. Pakistan's internet-based pharmacy business is expected to develop at a compound annual growth rate of 16.8% between 2020 and 2025 [8]. The growth might be ascribed to a significant shift in customer preferences, as they progressively favor the use of such services. An extensive examination is necessary to investigate the current condition and incidence of online shopping for medicines in Pakistan, including established sites like Dawaai.pk, DVAGO, and Tabiyat.pk, among others. The predominantly digital nature of web-based pharmacies, except for the combination model of DVAGO, has significantly influenced client preferences [9]. The industry ensures medicine legitimacy and follows regulations [10]. The use of health care apps and electronic gadgets is crucial for both persons who are using them for self-management and clinicians who are using them to enhance health care, as they have the ability to access metrics that allow them to evaluate the effectiveness and security of these innovations [11]. Potential risks to people's safety and well-being may arise if health-related apps and gadgets are not subjected to thorough assessments of their suitability and credibility. The lack of reliability in corroborating information inside medical apps has been emphasized in multiple studies. The increasing concern over the possible risks linked to the use of medical apps has led to an upsurge in curiosity in the surveillance of the effectiveness and dependability of these apps. Although there are no regulations in Pakistan regarding the usage of mobile health (mHealth) apps, we used international guidelines as a reference. In 2013, the Food and Drug Administration issued a handbook with guidelines for evaluating the quality of mobile health apps. Another guideline on digital health was established by the World Health Organization in order to enhance patient outcomes and provide direction for the application of initiatives related to digital health [12,13]. In the last decade, a research team led by Stoyanov developed the Mobile App Rating Scale (MARS) [14]. The MARS constitutes a well-recognized, dependable evaluation instrument designed exclusively for the assessment of the capabilities of (mHealth) apps [14,15]. The method has repeatedly shown a high level of dependability and a broad range of applications after being used for comprehensive evaluation of smartphone app content. The MARS is composed of 23 discrete elements that are organized into several areas, including participation, utility, design, accuracy of data, and subjective assessment [15-26].

Study Objective

The aim of our research was to evaluate digital pharmacy apps in Pakistan that were available on the Google Play Store and the iOS App Store. To achieve the aforementioned goal, an evaluation of these apps was conducted using the MARS.

Methods

Ethical Considerations

The investigation was done with scrupulous attention to ethical guidelines. The investigators acquired ethical permission from the Institutional Ethics Review Board committee of Jinnah University for Women. Their authorization was granted under the reference number JUW/IERB/PHARM-ERA-004/2023. This measure was used to enhance the dependability and credibility of the study while upholding the highest ethical standards.

Details of Assets and Investigation Techniques

A pair of assessors (pharmacists) was selected to evaluate the online pharmacy apps. The inclusion criteria required both assessors to have expertise in pharmacy and training in the MARS. The exclusion criterion was to eliminate individuals who were not trained in the MARS. Both assessors searched

XSL•FO
RenderX

for web-based pharmacy apps on the official marketplaces of the 2 prominent operating systems, namely Android (Play Store) and iOS (App Store). To guarantee the comprehensiveness of the findings, a widely searched phrase, namely "online pharmacy apps in Pakistan," was used in both stores.

App Criteria for Acceptability

The study's sample selection criteria were used to determine which apps met the specific inclusion criteria and exclude those apps that did not meet the inclusion criteria. Inclusion criteria included apps solely originating from Pakistan, accessible from the Google Play Store and iOS App Store, and which involved online pharmacies. Financial or premium app-related biases were not taken into account, and only English-language apps were considered. Exclusion criteria encompassed the exclusion of paid apps, apps in languages other than English, apps not available for download on the Google Play Store and iOS App Store, and any duplicate apps.

The Attributes and Qualities Associated With the Apps

The complete dataset of the apps included the input of crucial information. The collection of data encompassed several attributes of the apps, such as their title, platform (Android or iOS), costs, class (about health care, wellness, and physical activity), date of the latest update, the newest language, number of reviewers, score, creator, and number of downloads.

Selection Process of Web-Based Apps

The 2 neutral assessors assessed the apps by name and downloaded them from the iOS App Store and Google Play Store. The chosen apps demonstrated probable suitability and were entered into a database. The apps that satisfied the established criteria were kept, but those that failed to meet the requirements were eliminated. In instances when uncertainty arose over the suitability of an app, a third assessor was engaged.

Gathering and Assessment of Information Integrity

The 2 assessors received instructions regarding the use of the MARS via individually viewing tutorial videos accessible on YouTube. Subsequently, the users proceeded to independently download, use, and assess the quality of apps. To gather the data, the researchers used a collection form that encompassed various details such as the app's creator, device, version, release year of the latest release, costs, the number of downloaded files, feedback from consumers, the existence of a statement of confidentiality, privacy-related technical factors, medical device data, and the elements evaluated through the MARS.

Investigational Tool

The assessment of the usability of the incorporated apps was carried out by using the MARS, a comprehensive tool including 23 criteria that are categorized into 5 distinct sections.

The engagement component of the assessment evaluates many factors including enjoyment, interest among users, personalization, interaction (including alerts, messages, signals, and comments), and suitability for the target demographic.

The functionality component encompasses 4 key aspects related to operational efficiency, including (1) utilization, (2) directions, (3) conceptual flow, and (4) gestural design. The part on aesthetics assesses the elements of graphic layout, visual attractiveness, chromatic palette, and aesthetic coherence. The next part examines the existence of quality data such as textual content, comments, evaluations, and citations derived from trustworthy sources.

The subjective quality component consists of 4 items. This section assesses the consumer's level of enthusiasm for the app. Every single aspect was assessed using a rating system ranging from 1 (representing inadequacy) to 5 (representing excellence). The composite app grade was determined by computing the average score across parts A, B, C, and D. The app's performance results are a subjective measure, which was derived separately by calculating the mean value of subsection E. Furthermore, subsection F included 6 app-specific questions that evaluated the perceived influence of the app on the user's understanding, mindset, intention to make a change, and the probability of achieving successful modification of behavior concerning the targeted wellness behavior. To improve the objectivity of the MARS in evaluating app excellence, the inclusion of the subjective aspect of the quality part was omitted from the computation of the general average app performance grade. Moreover, the robust association shown in prior research between the MARS sum score and the individual's star rating suggests its efficacy in reflecting the overall perceived excellence [14].

Data Compilation and Assessment

Each reviewer conducted an independent evaluation of all the elements under the criteria of MARS for each of the apps. The average rating for each item within each app was computed using the numbers supplied by both evaluators. The mean scores for every category and part of the MARS were computed by taking the average values obtained from all the apps. As a result, mean scores were calculated for each category, and the same method was used to get the overall score and its corresponding standard deviation for each app.

Statistical Evaluation

The mean score for each dimension of the MARS was determined by calculating the mean values as recommended by the raters. In addition, the mean and standard deviation were obtained and Pearson correlation analysis was conducted for all dimensions of the MARS. The interrater reliability estimate for the MARS tool was conducted using 3 statistical indicators (Cohen κ coefficient, Krippendorff α, and intraclass correlation coefficient [ICC]). The κ coefficient was evaluated by assigning quadratic weights to different values. The ICC was calculated using a 2-way random model to assess the degree of agreement. Weighted κ, Krippendorff α, and ICC were calculated for each dimension.

Results

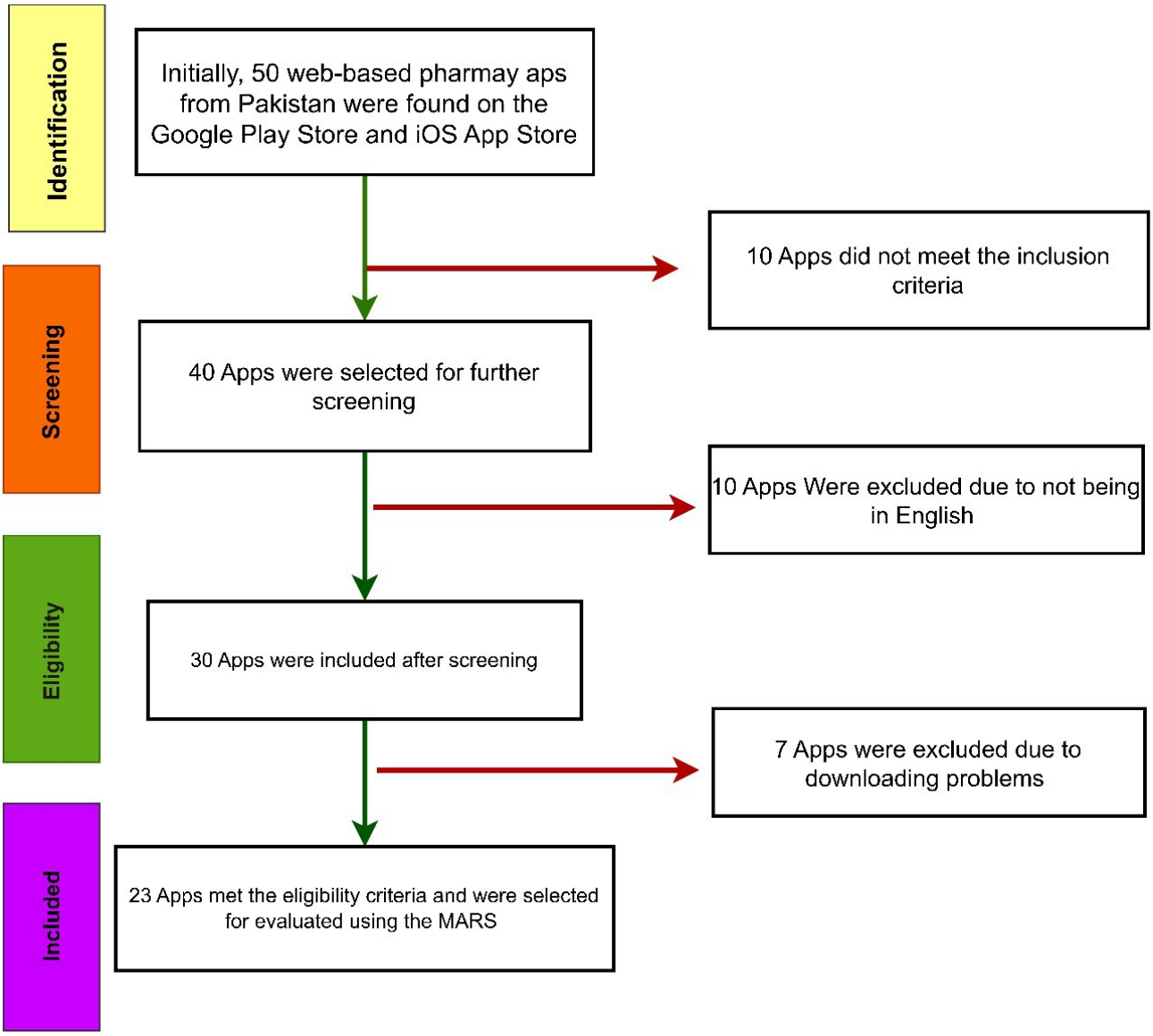
Overview

A preliminary search was conducted on the Google Play Store and iOS App Store to identify 50 online pharmacy apps accessible in Pakistan, as demonstrated in Figure 1. A total of 10 apps were eliminated from the study due to their failure to meet the predetermined inclusion criteria. Additionally, 10 apps

XSL•FO
RenderX

were found to be in languages other than English, so they were not included in the analysis. Furthermore, 7 apps could not be downloaded. A total of 23 apps were chosen for inclusion in the study, as shown in Figure 1. The apps have been downloaded from the Google Play Store and the iOS App Store, and they were evaluated using the MARS.

Figure 1. Flowchart for the selection of apps for this study.



App Characteristics

Web-based pharmacy apps in Pakistan offer a wide range of health care and pharmaceutical products and services. The tables in this section provide detailed information on technical features, creator details, customer reviews, pricing, version specifications, affiliations, and other relevant details of apps. These tables can aid users in selecting suitable apps and provide valuable insights into Pakistani health care and pharmaceutical competition. Table 1 presents a complete overview of the technological attributes associated with various pharmaceutical apps. Every app has been assigned a distinct serial number that uniquely identifies it and this table provides a comprehensive overview of the apps' special technological capabilities. These qualities encompass the ability to distribute material across multiple platforms and offer secure login capabilities, among other functionalities. The table is presented as a valuable resource for both customers and developers, facilitating their comprehension of the various functionalities and fundamental potentials of these apps. Table 2 presents a summary of the main domains of focus (ie, strategies for utilization and the affiliations with which they are connected). The analysis suggests that the major objective of these kinds of apps is to enhance the welfare of individuals, while concurrently implementing a business-oriented approach that involves severing connections between companies. This information facilitates users' understanding of the underlying goals and target market of these apps. Table 3 provides a comprehensive analysis of the manufacturers of pharmacy apps, encompassing essential details such as the number of evaluations, feedback, the date of the most recent update, and the pricing model of the app (whether it is free or paid). The table presented herein can be used by users to evaluate the reliability of these programs by taking into account both the reputation of the creators and the feedback offered by consumers. Table 4 presents a comprehensive examination of several versions of pharmaceutical software including their

XSL•FO
RenderX

download statistics, release dates, and the operating systems on which they may be accessed. The provided data hold significant value for users who are in search of compatibility with their devices and the most up-to-date versions of apps. Table 5 presents concise and informative descriptions of pharmacy apps, furnishing users with a concise summary of the features and distinctive value proposition of each app. This feature is especially advantageous for individuals who are in search of a suitable app that meets their health care and pharmaceutical needs.

Table . Technological components of mobile apps for online pharmacies in Pakistan.

Serial number	App name	Technical aspects of the app (all that apply)
1	DVAGO	Allow sharing of different platforms, send reminders
2	Dawaai - Medicine & Healthcare	Allow sharing of different platforms, login protection
3	Servaid plus	Allow sharing of different platforms
4	Ehad Pharmacy	Allow sharing of different platforms, login protection
5	Goli Monthly Pharmacy	Allow sharing of different platforms, login protection
6	Dawawala	Allow sharing of different platforms
7	Sakoon Pharmacy and Healthcare	Allow sharing of different platforms
8	Emeds Pharmacy	Allow sharing of different platforms
9	Najeeb pharmacy	Allow sharing of different platforms
10	Tabiyat.pk	Allow sharing of different platforms
11	DawaAppTak	Allow sharing of different platforms, login protection
12	e-Clinix	Allow sharing of different platforms
13	Medicalstore.com.pk	Allow sharing of different platforms
14	Dava Pakistan	Allow sharing of different platforms
15	Qarshi	Allow sharing of different platforms
16	Medipanda	Allow sharing of different platforms
17	DH Pharmacy	Allow sharing of different platforms
18	Global Care	Allow sharing of different platform
19	Dawa Asaan	Allow sharing of different platforms
20	Bin Hashim	Allow sharing of different platforms
21	Dawa Online	Allow sharing of different platforms, login protection
22	Caplet	Allow sharing of different platforms
23	Care pharmacy	Allow sharing of different platforms, login protection

XSL•FO
RenderX

Table . Qualities and targets of online pharmacy apps in Pakistan.

Serial number	App names	Focus (what the app targets)	Theoretical background and strategies	Affiliations
1	DVAGO	Well-being	Business	Commercial
2	Dawaai - Medicine & Healthcare	Well-being	Business	Commercial
3	Servaid plus	Well-being	Business	Commercial
4	Ehad	Well-being	Business	Commercial
5	Goli monthly pharmacy	Well-being	Business	Commercial
6	Dawawala	Well-being	Business	Commercial
7	Sakoon pharmacy and health care	Well-being	Business	Commercial
8	Emeds	Well-being	Business	Commercial
9	Najeeb pharmacy	Well-being	Business	Commercial
10	Tabiyat.pk	Well-being	Business	Commercial
11	DawaAppTak	Well-being	Business	Commercial
12	e-Clinix	Well-being	Business	Commercial
13	Medicalstore.com.pk	Well-being	Business	Commercial
14	Dava Pakistan	Well-being	Business	Commercial
15	Qarshi	Well-being	Business	Commercial
16	Medipanda	Well-being	Business	Commercial
17	DH Pharmacy	Well-being	Business	Commercial
18	Global Care	Well-being	Business	Commercial
19	Dawa Asaan	Well-being	Business	Commercial
20	Bin Hashim	Well-being	Business	Commercial
21	Dawa Online	Well-being	Business	Commercial
22	Caplet	Well-being	Business	Commercial
23	Care pharmacy	Well-being	Business	Commercial

XSL•FO
RenderX

Table . Information about the creator and specific characteristics of online pharmacy apps in Pakistan.

Serial number	App name	Developers	Number of reviews	Rating	Last update	Cost
1	DVAGO	Novacare (Pvt) Ltd	725	3.5	July 3, 2023	Free
2	Dawaai - Medicine & Healthcare	Dawaai Pvt Ltd	12,600	3.6	April 21, 2023	Free
3	Servaid plus	Servaid Pharmacy	461	2.4	July 19, 2023	Free
4	Ehad	Ehad Virtual Healthcare	72	3.9	December 12, 2022	Free
5	Goli monthly pharmacy	Atlash Tech LLC	85	4.5	January 31, 2023	Free
6	Dawawala	Dawawala	1000	4.9	April 29, 2022	Free
7	Sakoon pharmacy and health care	Healthcare Mart Pvt Ltd	33	4.5	August 25, 2023	Free
8	Emeds	Emeds	10	3.7	March 15, 2021	Free
9	Najeeb pharmacy	Xperia Tech	630	2.8	September 27, 2022	Free
10	Tabiyat.pk	Medznmore	4000	1.9	April 19, 2023	Free
11	DawaApptak	Muller & Phipps (Pvt.) Ltd	6	5	September 8, 2023	Free
12	e-Clinix	Bright-line solutions	195	3.2	April 5, 2022	Free
13	Medical-store.com.pk	Medical-store.com.pk	746	2.3	N/A[a]	Free
14	Dava Pakistan	B techno Media	16	3.9	November 25, 2021	Free
15	Qarshi	Obrotu	112	3.8	February 28, 2021	Free
16	Medipanda	Medipanda Pvt Ltd	42	3.8	November 24, 2020	Free
17	DH Pharmacy	Digital Health Services	0	N/A	August 30, 2023	Free
18	Global Care	Devcon Digital	10	4.0	July 20, 2017	Free
19	Dawa Asaan	Dawa Asaan	29	4.5	July 31, 2023	Free
20	Bin Hashim	Blink Co.Technologies	151	4.6	June 10, 2023	Free
21	Dawa Online	Dawa online	56	2	April 12, 2023	Free
22	Caplet	Salman abid	0	N/A	June 27, 2023	Free
23	Care pharmacy	Sibyl technologies	0	N/A	May 4, 2023	Free

[a]N/A: not applicable.

XSL•FO
RenderX

Table . Information about the apps, including details regarding their version updates, download statistics, platform compatibility, and associated age group.

Serial number	App name	Version	Downloads	Release date	Platform	Age group
1	DVAGO	2.1.5	≥100,000	January 31, 2022	iOS	General
2	Dawaai - Medicine & Healthcare	4.2.1	≥1,000,000	December 30, 2017	Google Play Store	General
3	Servaid Plus	1.2.2	≥50,000	December 4, 2020	iOS	General
4	Ehad	1.2.4	≥10,000	October 1, 2021	Google Play Store	General
5	Goli Monthly Pharmacy	1.0.1.0	≥10,000	October 31, 2022	iOS	General
6	Dawawala	1.0.9	≥1000	April 14, 2022	Google Play Store	General
7	Sakoon Pharmacy and Healthcare	1.0.2	≥1000	November 8, 2022	iOS	General
8	Emeds	1.0	≥1000	March 15, 2021	Google Play Store	General
9	Najeeb Pharmacy	31.1	≥10,000	August 15, 2020	iOS	General
10	Tabiyat.pk	4.1.1	≥100,000	April 16, 2021	Google Play Store	General
11	DawaAppTak	8.0.0	≥500	August 13, 2022	Google Play Store	General
12	e-Clinix	4.4	≥10,000	July 10, 2021	iOS	General
13	Medical-store.com.pk	N/A[a]	100,000	N/A	Google Play Store	General
14	Dava Pakistan	0.0.1	≥1000	March 31, 2021	Google Play Store	General
15	Qarshi	2.0.6	≥10,000	April 22, 2020	Google Play Store	General
16	Medipanda	1.18	1000	November 9, 2019	Google Play Store	General
17	DH Pharmacy	1.01	≥10	April 21, 2022	Google Play Store	General
18	Global Care	1.1.4	≥100	July 11, 2017	Google Play Store	General
19	Dawa Asaan	3.1.0	≥1000	February 10, 2021	iOS	General
20	Bin Hashim	1.1.2	≥10,000	April 7, 2022	iOS	General
21	Dawa Online	1.0.4	≥10,000	January 31, 2020	Google Play Store	General
22	Caplet	2.01	≥50	April 7, 2022	iOS	General
23	Care pharmacy	1.01	≥1000	May 4, 2023	Google Play Store	General

[a]N/A: not applicable.

XSL•FO
RenderX

Table . Summary of online pharmacy apps in Pakistan.

Serial number	App name	Description provided by the app
1	DVAGO	DVAGO is Pakistan‘s most trusted pharmacy chain for complete health care
2	Dawaai - Medicine & Healthcare	Dawaai Pakistan, also known as Dawaai.pk, is Pakistan’s premier online pharmacy, offering a comprehensive 360-degree medical solution for all your health care requirements. Dawaai holds the distinction of being one of Pakistan’s longest-established and most reliable digital health platforms, providing access to affordable medications.
3	Servaid plus	Your medicine is at your doorstep, with Servaid, no more waiting in lines anymore.
4	Ehad	A better world with quality health care.
5	Goli monthly pharmacy	Goli is an online pharmacy platform.
6	Dawawala	B2B marketplace for pharmaceutical distributors and wholesalers.
7	Sakoon pharmacy and health care	A pharmacy of branded medicines in your pocket.
8	Emeds	Pakistan’s number 1 online pharmacy for online medicines, lab tests, and health tips.
9	Najeeb pharmacy	We provide free home delivery service in Islamabad within an hour.
10	Tabiyat.pk	Tabiyat.pk is an online pharmacy app where your medicines and health care products get delivered anywhere in Pakistan.
11	DawaApptak	DawaApptak customer app for our valuable customers.
12	e-Clinix	e-Clinix, the leading retail pharmacy brand of Pakistan, is now at your door.
13	Medicalstore.com.pk	The biggest online pharmacy app and medical store offers medicine delivery services in Pakistan.
14	Dava Pakistan	An online directory that is in the process of transitioning into an online pharmacy. Our goal is to offer people access to information about alternative medicines, helping them find cost-effective options rather than paying exorbitant prices.
15	Qarshi	Authentic health products with the fastest delivery.
16	Medipand	Get your medicines delivered right to your doorstep with convenient cash payment upon delivery. We also provide herbal deliveries, surgical machines, patient-use devices, wheelchairs, baby essentials, and food supplements, all available with free delivery and the option to pay cash on delivery. As Pakistan’s first trusted online pharmacy, we are your go-to source for purchasing medicines online in Lahore, Punjab, Pakistan.
17	DH Pharmacy	Pakistan‘s number 1 pharmacy for medicines, beauty, and surgical products.

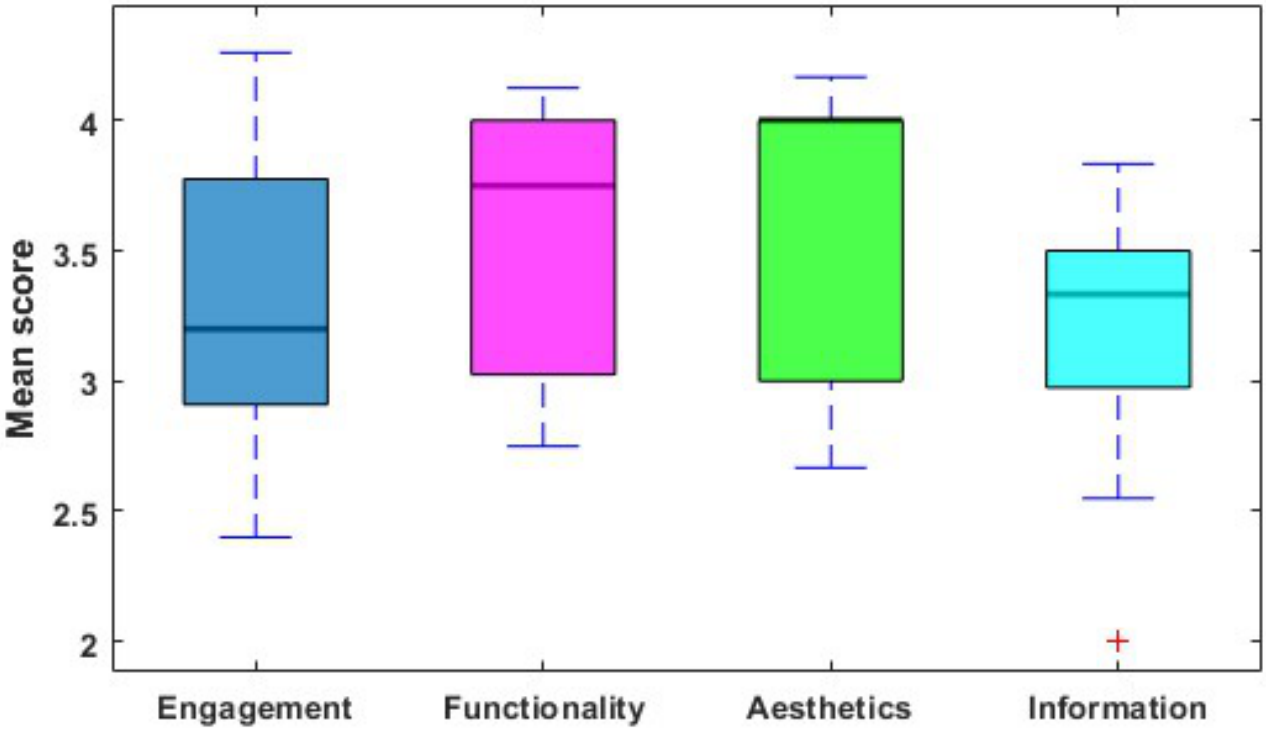
XSL•FO
RenderX

Serial number	App name	Description provided by the app
18	Global Care	Global Care stands as Pakistan's pioneering online pharmacy, offering the option to purchase affordable medicines through the convenience of uploading your doctor's prescription or placing an order independently. We are a licensed pharmacy providing home delivery services, with delivery times of 1 to 2 hours in Rawalpindi - Islamabad and 2-3 days across Pakistan.
19	Dawa Asaan	Dawa Asaan is Pakistan's first smart pharmacy.
20	Bin Hashim	Order your medications and various items online, and have them conveniently delivered to your doorstep.
21	Dawa Online	Dawa Online Medical Store is a prominent and recognized medical establishment situated in the city of Raipur. Their services encompass the timely delivery of medicines, over-the-counter products, cosmetics, and generic medications to cater to the needs of all their patients.
22	Caplet	The Caplet Pharmacy app facilitates a swift sign-up procedure and efficiently locates the medicines prescribed to you by your doctor. Additionally, Caplet Pharmacy extends its services to provide over-the-counter products to its customers.
23	Care pharmacy	Online marketplace for acquiring prescription medications and health-related items.

MARS Assessment

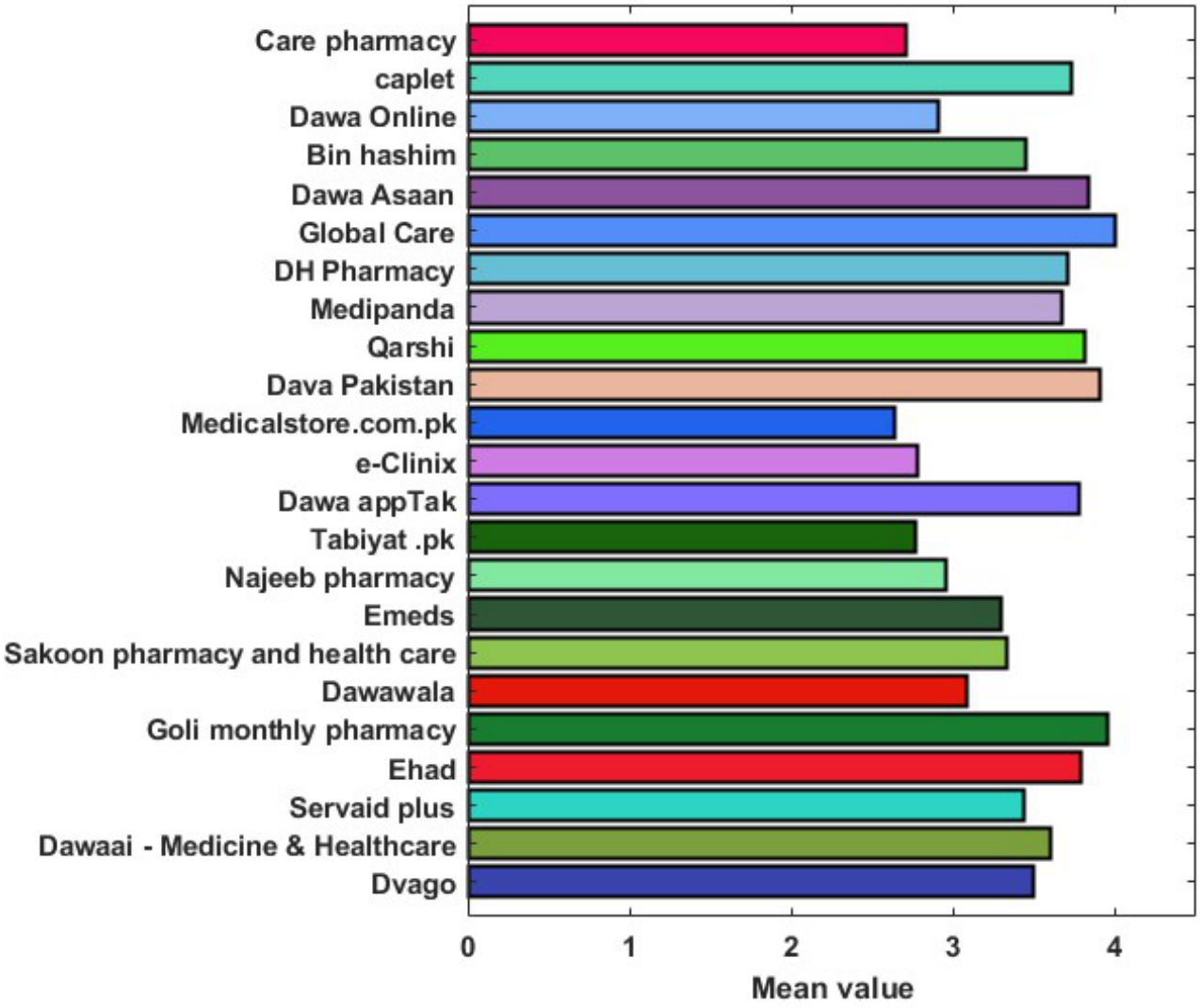
The mean scores of the 23 apps ranged from 2.64 to 4.00. The mean scores of each dimension of the MARS are demonstrated in Figure 2. The individual scores of each app are presented in Figure 3.

Figure 2. Mean scores of each dimension of the Mobile App Rating Scale.



XSL•FO
RenderX

Figure 3. Mean scores of individual apps as evaluated by the Mobile App Rating Scale.



The ICC coefficient was computed for the 5 aspects of the MARS (Table 6). The ICC coefficient for the variable "Engagement" was determined to be 0.764, with a 95% CI ranging from 0.521 to 0.893. This finding indicates a considerable degree of consensus among raters about the concept of participation. Similarly, the construct labeled "Functionality" demonstrated a notable ICC coefficient of 0.824 (95% CI 0.630-0.922), which signifies a robust consensus among raters on their evaluations of the construct's functionality. The variable "Aesthetics" demonstrated the highest ICC coefficient, measuring 0.949 (95% CI 0.885-0.978), indicating a substantial level of consensus about aesthetic judgments. The dimension of "Information" demonstrated a moderate ICC cofficient of 0.702 (95% CI 0.601-0.875), and "Subjective quality" exhibited strong agreement, as indicated by an ICC coefficient of 0.871 (95% CI 0.720-0.943). The Krippendorff α coefficient was computed to evaluate the reliability of the MARS over the identical set of 5 dimensions. The Krippendorff α coefficients for the dimensions of "Engagement," "Functionality," "Aesthetics," "Information," and "Subjective quality" were 0.759 (95% CI 0.661-0.857), 0.825 (95% CI 0.727-0.923), 0.979 (95% CI 0.881-1.077), 0.705 (95% CI 0.689-0.885), and 0.755 (95% CI 0.657-0.853), respectively. The Krippendorff α values, accompanied by their corresponding CIs, serve to affirm the reliability and coherence of raters' interpretations across the assessed factors, so enhancing the overall comprehension of the data's integrity. Furthermore, the investigators used weighted κ statistics to evaluate the characteristics of the MARS. The weighted κ value for the parameter "Engagement" was calculated to be 0.764 (95% CI 0.350-0.722), indicating a reasonable degree of consistency that has been adjusted for chance. The assessment of "Functionality" resulted in a Cohen κ score of 0.658 (95% CI 0.510-0.805), showing a statistically significant degree of consensus among evaluators about the meaning of functionality. The variable "Aesthetics" demonstrated a statistically significant κ value of 0.907 (95% CI 0.819-0.994), showing a strong agreement in the evaluation of aesthetic attributes. The κ statistic for the value of the parameter "Information" was calculated to be 0.388 (95% CI 0.151-0.626). The study's findings indicate a rather low degree of consensus between assessors about the dimension of Information. The subjective quality assessment resulted in a κ score of 0.670 (95% CI 0.488-0.851), showing a statistically significant level of agreement. The reliability and internal consistency of every element of the MARS questionnaire were assessed using Cronbach α. The rating of "Engagement" had a Cronbach α value of 0.867, indicating a high degree of internal consistency. The concept of "Functionality" demonstrated a

XSL•FO
RenderX

significant degree of internal consistency, as indicated via a Cronbach α value of 0.904. The concept of "Aesthetics" had a notably high Cronbach α rating of 0.905, suggesting a commendable level of internal consistency. The Cronbach α value of 0.821 indicates that the "Information" parameter has a high level of internal coherence. The construct of "Subjective quality" demonstrated a Cronbach α coefficient of 0.911, indicating a substantial degree of internal consistency.

Table . Evaluating the consistency and reliability of online pharmacy apps in Pakistan using the intraclass correlation coefficient, Krippendorff α, κ, and Cronbach α.

	Values
Mobile App Rating Scale dimensions, intraclass correlation coefficient (95% CI)	
Engagement	0.764 (0.521-0.893)
Functionality	0.824 (0.630-0.922)
Aesthetics	0.913 (0.885-0.978)
Information	0.702 (0.601-0.875)
Subjective quality	0.871 (0.720-0.943)
Mobile App Rating Scale dimensions, Krippendorff α (95% CI)	
Engagement	0.759 (0.661-0.857)
Functionality	0.825 (0.727-0.923)
Aesthetics	0.979 (0.881-0.077)
Information	0.705 (0.689-0.885)
Subjective quality	0.755 (0.657-0.853)
Mobile App Rating Scale dimensions, weighted κ (95% CI)	
Engagement	0.536 (0.350-0.722)
Functionality	0.658 (0.510-0.805)
Aesthetics	0.907 (0.819-0.994)
Information	0.388 (0.151-0.626)
Subjective quality	0.670 (0.488-0.851)
Mobile App Rating Scale dimensions, Cronbach α	
Engagement	0.867
Functionality	0.904
Aesthetics	0.905
Information	0.821
Subjective quality	0.911

Relationship of MARS Dimension With Total MARS Scores and the Mean (SD) of Each Dimension

Table 7 presents a detailed overview of the relationship values, means, and standard deviations of various elements of the MARS. The dimension denoted as "Engagement" exhibited a notable favorable correlation of 0.949. The average score for this aspect was 3.324, accompanied by a rather small standard deviation of 0.560. The results of this study indicate a substantial degree of concurrence and uniformity among participants in their assessments of engagement. The parameter "Functionality" demonstrated a significant positive relationship of 0.923. The data exhibited a marginally elevated mean rating of 3.520, accompanied by an acceptable standard deviation of 0.491. These findings indicate consistent perceptions of the functionality of the subject matter. The aspect referred to as "Aesthetic" demonstrated a considerable positive association of 0.798. Additionally, it had a mean value of 3.626 and a standard deviation of 0.504. The aforementioned figures indicate a substantial agreement among participants on their assessments of beauty. The aspect labeled as "Information" exhibited a noteworthy positive relationship of 0.763. The average grade for this aspect was 3.044, suggesting a consistent perspective. Moreover, the observed standard deviation of 0.359 indicates a rather small range of variation in individuals' judgments regarding the information aspect. The study's findings provide valuable insights into the prevailing viewpoints of the participants and the level of agreement seen across multiple aspects. These observations enhance our understanding of the information quality derived from the MARS.

XSL•FO
RenderX

Table . Mean and standard deviation of the correlation among various aspects of the Mobile Rating App Scale.

Mobile App Rating Scale dimensions	Correlation value	Mean (SD)
Engagement	0.949	3.324 (0.560)
Functionality	0.923	3.520 (0.491)
Aesthetic	0.798	3.626 (0.504)
Information	0.763	3.044 (0.359)

Discussion

Principal Findings

Online pharmacies not only offer medication selling, purchasing, and delivery to patients but they also provide medication details, which include brand, dosage form, indication, side effects, warning, and price. Thus, the critical first step in the treatment process is to evaluate the standards of online pharmacies. This investigation represents a significant achievement as it constitutes the inaugural evaluation of the quality of online pharmacy apps in Pakistan through the utilization of a mobile app grading scale developed by Stoyanov et al [14]. In this comprehensive study, a meticulous analysis was undertaken of the pharmacy apps available in Pakistan; as Amor-García et al [12] did with genitourinary tumor apps, we chose and downloaded our apps using both the Google Play Store and the iOS App Store. The sample size of 23 apps was comparable to the sample size in research carried out by Kim et al [26]. These apps met the established criteria. As previously reported in other research, we discovered that the Google Play Store had a wider variety of apps than the iOS App Store. Nine apps were downloaded from the iOS App Store, while 14 came from the Google Play Store [21,27].

The MARS is an innovative and effective tool that is used for evaluating mobile health apps. The mean score of all apps ranged from 2.64 to 3.92, which was similar to scores reported by previous studies [17,28-30]. The subjective quality component of the MARS was omitted from the collective mean score [24,31]. Furthermore, we determined the average scores for several dimensions of “Engagement” (mean 3.32, SD 0.560), “Functionality” (mean 3.52, SD 0.49), “Aesthetics” (mean 3.62, SD 0.50), and “Information quality” (mean 3.04, SD 0.35). Several methodologies have been developed to evaluate the efficacy of mHealth apps. Nevertheless, the absence of a universally accepted method and the limited participation of scientific professionals have made it impractical to regularly put them into practice [32]. The MARS approach, as documented by Stoyanov et al [14], is regarded as a user-friendly, straightforward, and coherent instrument that is highly regarded for its reliability due to its endorsement by skilled technicians and health care experts. This scale enables a comprehensive analysis that encompasses several disciplines and is intended for the assessment of health care apps. It consists of 5 domains that encompass the key factors necessary for a thorough evaluation. In their study, Stoyanov et al [14] found that the MARS assessment had excellent internal consistency (Cronbach α=0.9) and interrater reliability (2-way mixed ICC coefficient=0.79, 95% CI 0.75 - 0.83) when used to assess 50 mental health apps.

Our findings were similar to both the original MARS and other translated versions, according to reliability and validity studies [33-36]. As per globally recognized quality standards, all sections’ internal reliability and Cronbach α as well as the overall scores were found to be good [37]. Subsequent translation studies and the original MARS study also noted this high level of internal consistency. In contrast, our research revealed a good degree of interrater reliability, with an ICC ranging from 0.702 to 0.913. To ensure construct validity, we opted for a multitrait scale evaluation with the item-subscale association over a factor analysis, as this approach has proven effective in prior research [26]. Our findings met the predetermined success level and demonstrated good parallel and divergent reliability, except for the “Information” subscale because many parameters were not measurable and less information on content related to medications. Nonetheless, we discovered a low ICC for the “Information” subscale. Prior research has similarly documented the diminished outcomes associated with the information variable [17].

The mobile apps provided did not consistently cite the authors or websites that provided information, and there was no assurance of the accuracy and timeliness of the material.

The diminished quality of data has been linked to a variety of dangers for mobile app users, primarily due to the potential for misinformation leading to inaccurate self-diagnoses and unfavorable health-related decisions in prevention, wellness, and therapy [38,39]. The majority of apps were simple to download and log-in friendly. Of the 23 apps, we only encountered login issues with 2 or 3. In contrast to other aspects of the MARS, the information dimension was reduced because, of the 23 apps, some apps provided information about the specifics of medications, such as dosage forms, indications, side effects, precautions, categories, and alternative brands, and many apps only presented brand names with the cost and category of drugs. The medication information is useful for patients to have in addition to the drugs they choose or for other purposes relevant to medicines or managing their treatment.

As a result of transportation and public gathering limitations brought on by the COVID-19 epidemic that struck the nation in February 2020, the selling of medicines online expanded significantly in the years that followed [40]. Many people started to order medications from online pharmacies even if the quality of the products was still questionable according to medical professionals. Pharmacists argued that online pharmacies put their businesses at risk and allow the free selling of potentially

XSL•FO
RenderX

harmful drugs [41-46]. The actual number of online pharmacies in Pakistan is unknown, but according to authorities, most of them are located in the country's 2 biggest cities, Karachi and Lahore. They claim that, in specific locations, clients will get prescription drugs and other medical supplies within 4 hours of placing their order, or, in other cities, within 48 hours via a shipping service. In addition, many online pharmacies provide free shipping to client residences and discounts of as much as 20%. Online pharmacies help people get medicines in just a few clicks. Online pharmacies are not only a source of medicine delivery. Online pharmacy apps provide other services including prescription refills, information about the availability of drugs, and information about drugs. Pharmacies are the main pillar for this treatment method [47,48]. However, there have been no studies performed on this topic. Our study provides information about these pharmacy services, including engagement, functions, layout, and basic information.

Limitations

The scope of our research was mostly limited to a single geographic region, namely Pakistan, and specifically focused on apps accessible in the English language. Additionally, the limitations of our study stem from the narrow focus on free apps. We did not conduct a thorough literature analysis and our technique was primarily focused on app stores. As a result, it might be possible that we did not fully explore the entire collection of available data. The evaluation was done by just 2 assessors. There were more assessors in previous studies that used the MARS.

Future Recommendations

To bolster the reliability of online pharmacy apps in Pakistan, it is advisable to enforce more stringent laws and guidelines, regularly assess the content of these apps through audits, fortify data security measures, and provide standardized and evidence-based medical information. Subsequent investigations may focus on the exploration of client observations and opinions about these apps, with a particular emphasis on the dynamics of user-platform interaction and the establishment of trustworthiness.

Conclusion

The MARS is a trustworthy and adaptable tool for assessing the reliability of mobile apps. The evaluation of online pharmacy apps of Pakistan using the MARS provided helpful data on the reliability and efficacy of these apps. The results showed the trustworthiness and dependability of web-based pharmacy apps. Many of the evaluated online pharmacy apps exhibited high reliability and credibility in all dimensions of the MARS.

Acknowledgments

The authors of this study extend their appreciation to the Researchers Supporting Project (project number RSP2025R378), King Saud University, Riyadh, Saudi Arabia, which provided the funding for this work.

Conflicts of Interest

None declared.

References

1. Gray NJ. The evolution of online pharmacies. Self Care Journal 2011 Doi 10.2217/14750708.2.3.76.;2:76-86.
2. Global industry analysis, size, share, growth, trends and forecast, 2013-2019. Transparency Market Research. 2014. URL: https://www.transparencymarketresearch.com [accessed 2025-04-08]
3. Quon BS, Firszt R, Eisenberg MJ. A comparison of brand-name drug prices between Canadian-based internet pharmacies and major U.S. drug chain pharmacies. Ann Intern Med 2005 Sep 20;143(6):397-403. [doi: 10.7326/0003-4819-143-6-200509200-00004] [Medline: 16172438]
4. Liang BA, Mackey TK. Prevalence and global health implications of social media in direct-to-consumer drug advertising. J Med Internet Res 2011 Aug 31;13(3):e64. [doi: 10.2196/jmir.1775] [Medline: 21880574]
5. Burney A, Abbas Z, Mahmood N, Arifeen QU. Prospects for mobile health in Pakistan and other developing countries. AIT 2013;03(2):27-32. [doi: 10.4236/ait.2013.32A004]
6. Dutta D, Bhattacharjee B. Consumer preference and buying pattern of medicines through e-pharmacy during COVID-19 pandemic in Silchar, Assam. Current Trends in Pharmaceutical Research 2021;8(1):192-211.
7. Callan P, et al. mHealth Education: Harnessing the Mobile Revolution to Bridge the Health Education and Training Gap in Developing Countries: iHeed Institute; 2011. [doi: 10.4236/ait.2013.32A004]
8. Javed M, Shaiq M. The effects of online pharmacy on consumer behaviour. PJAR 2023;3(1):384-402.
9. Agarwal S, Bhardwaj G. A study of consumer buying behaviour towards e-pharmacies in Delhi NCR. IJFE 2020;4(4):255. [doi: 10.1504/IJFE.2020.115022]
10. Gupta MS. Consumer buying behavior towards e-pharmacy. Dogo Rangsang Research Journal 2020;10(3):183-190.
11. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210. [doi: 10.2196/jmir.3133] [Medline: 25223398]
12. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, et al. Assessing apps for patients with genitourinary tumors using the mobile application rating scale (MARS): systematic search in app stores and content analysis. JMIR mHealth uHealth 2020 Jul 23;8(7):e17609. [doi: 10.2196/17609] [Medline: 32706737]

XSL•FO
RenderX

13. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening: Web Supplement 2: Summary of Findings and GRADE Tables: World Health Organization; 2019. URL: https://iris.who.int/handle/10665/324998 [accessed 2025-04-08]
14. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
15. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR mHealth uHealth 2015 Aug 19;3(3):e82. [doi: 10.2196/mhealth.4328] [Medline: 26290327]
16. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
17. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR mHealth uHealth 2018 Oct 25;6(10):e10718. [doi: 10.2196/10718] [Medline: 30361196]
18. Kalhori SRN, Hemmat M, Noori T, Heydarian S, Katigari MR. Quality evaluation of English mobile applications for gestational diabetes: app review using Mobile Application Rating Scale (MARS). Curr Diabetes Rev 2021;17(2):161-168. [doi: 10.2174/1573399816666200703181438] [Medline: 32619173]
19. Davalbhakta S, Advani S, Kumar S, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164. [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
20. Romero RL, Kates F, Hart M, Ojeda A, Meirom I, Hardy S. Quality of deaf and hard-of-hearing mobile apps: evaluation using the Mobile App Rating Scale (MARS) with additional criteria from a content expert. JMIR mHealth uHealth 2019 Oct 30;7(10):e14198. [doi: 10.2196/14198] [Medline: 31670695]
21. Knitza J, Tascilar K, Messner EM, et al. German mobile apps in rheumatology: review and analysis using the Mobile Application Rating Scale (MARS). JMIR mHealth uHealth 2019 Aug 5;7(8):e14991. [doi: 10.2196/14991] [Medline: 31381501]
22. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13(1):35. [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
23. Steeb T, Wessely A, French LE, Heppt MV, Berking C. Skin cancer smartphone applications for German-speaking patients: review and content analysis using the Mobile App Rating Scale. Acta Derm Venereol 2019 Oct 1;99(11):1043-1044. [doi: 10.2340/00015555-3240] [Medline: 31233185]
24. Sharif MO, Alkadhimi A. Patient focused oral hygiene apps: an assessment of quality (using MARS) and knowledge content. Br Dent J 2019 Sep;227(5):383-386. [doi: 10.1038/s41415-019-0665-0] [Medline: 31520041]
25. Choi YK, Demiris G, Lin SY, et al. Smartphone applications to support sleep self-management: review and evaluation. J Clin Sleep Med 2018 Oct 15;14(10):1783-1790. [doi: 10.5664/jcsm.7396] [Medline: 30353814]
26. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the Mobile App Rating Scale (MARS). JMIR mHealth uHealth 2018 Mar 28;6(3):e74. [doi: 10.2196/mhealth.8613] [Medline: 29592848]
27. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334. [doi: 10.2196/20334] [Medline: 32614777]
28. Böhme C, von Osthoff MB, Frey K, Hübner J. Qualitative evaluation of mobile cancer apps with particular attention to the target group, content, and advertising. J Cancer Res Clin Oncol 2018 Jan;144(1):173-181. [doi: 10.1007/s00432-017-2533-0] [Medline: 29052039]
29. Richardson B, Dol J, Rutledge K, et al. Evaluation of mobile apps targeted to parents of infants in the neonatal intensive care unit: systematic app review. JMIR mHealth uHealth 2019 Apr 15;7(4):e11620. [doi: 10.2196/11620] [Medline: 30985282]
30. Siddique AB, Krebs M, Alvarez S, et al. Mobile apps for the care management of chronic kidney and end-stage renal diseases: systematic search in app stores and evaluation. JMIR mHealth uHealth 2019 Sep 4;7(9):e12604. [doi: 10.2196/12604] [Medline: 31486408]
31. Roy A, Nayak PP, Shenoy PR, Somayaji K. Availability and content analysis of smartphone applications on conservative dentistry and endodontics using Mobile Application Rating Scale (MARS). Curr Oral Health Rep 2022;9(4):215-225. [doi: 10.1007/s40496-022-00325-w]
32. Meyer AJ, Armstrong-Hough M, Babirye D, et al. Implementing mHealth interventions in a resource-constrained setting: case study from Uganda. JMIR mHealth uHealth 2020 Jul 13;8(7):e19552. [doi: 10.2196/19552] [Medline: 32673262]
33. Martin Payo R, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov SR, Llaneza Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. Int J Med Inform 2019 Sep;129:95-99. [doi: 10.1016/j.ijmedinf.2019.06.005] [Medline: 31445295]

XSL•FO
RenderX

34. Domnich A, Arata L, Amicizia D, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016 Jul 7;16(1):83. [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
35. Messner EM, Terhorst Y, Barke A, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR mHealth uHealth 2019;8(3):e14479. [doi: 10.2196/14479]
36. Saliasi I, Martinon P, Darlington E, et al. Promoting health via mHealth applications using a French version of the Mobile App Rating Scale: adaptation and validation study. JMIR mHealth uHealth 2021 Aug 31;9(8):e30480. [doi: 10.2196/30480] [Medline: 34463623]
37. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
38. Albrecht UV. Chancen und risiken von gesundheits-apps. In: Albers M, Katsivelas I, editors. Recht & Netz [Book in German]: Nomos Verlagsgesellschaft; 2018. [doi: 10.5771/9783845293288-417]
39. Gigerenzer G, Schlegel-Matthies K, Wagner GG. : SVRV, Sachverständigenrat Für Verbraucherfragen; 2016. URL: https://www.svr-verbraucherfragen.de/wp-content/uploads/2024/10/Digitale-Welt-und-Gesundheit.pdf [accessed 2025-04-08]
40. Jairoun AA, Al-Hemyari SS, Abdulla NM, et al. Correction to: Online medication purchasing during the Covid-19 pandemic: a pilot study from the United Arab Emirates. J Pharm Policy Pract 2021 May 11;14(1):44. [doi: 10.1186/s40545-021-00324-9] [Medline: 33975632]
41. Orizio G, Schulz P, Domenighini S, et al. Cyberdrugs: a cross-sectional study of online pharmacies characteristics. Eur J Public Health 2009 Aug;19(4):375-377. [doi: 10.1093/eurpub/ckn146] [Medline: 19151103]
42. Fox S, Duggan M. Information triage. : Pew Research Center; 2013. URL: https://www.pewresearch.org/internet/2013/01/15/information-triage [accessed 2025-04-08]
43. Bukhari N, Rasheed H, Nayyer B, Babar ZUD. Pharmacists at the frontline beating the COVID-19 pandemic. J Pharm Policy Pract 2020;13(1):8. [doi: 10.1186/s40545-020-00210-w] [Medline: 32328285]
44. Bonyan R, Al-Karasneh AF, El-Dahiyat F, Jairoun AA. Identification of the awareness level by the public of Arab countries toward COVID-19: cross-sectional study following an outbreak. J Pharm Policy Pract 2020;13:43. [doi: 10.1186/s40545-020-00247-x] [Medline: 32724659]
45. Algunmeeyn A, El-Dahiyat F, Altakhineh MM, Azab M, Babar ZUD. Understanding the factors influencing healthcare providers' burnout during the outbreak of COVID-19 in Jordanian hospitals. J Pharm Policy Pract 2020;13:53. [doi: 10.1186/s40545-020-00262-y] [Medline: 32974035]
46. Zeenny RM, Ramia E, Akiki Y, Hallit S, Salameh P. Assessing knowledge, attitude, practice, and preparedness of hospital pharmacists in Lebanon towards COVID-19 pandemic: a cross-sectional study. J Pharm Policy Pract 2020;13:54. [doi: 10.1186/s40545-020-00266-8] [Medline: 32959004]
47. Alwon BM, Solomon G, Hussain F, Wright DJ. A detailed analysis of online pharmacy characteristics to inform safe usage by patients. Int J Clin Pharm 2015 Feb;37(1):148-158. [doi: 10.1007/s11096-014-0056-1] [Medline: 25564180]
48. Chaturvedi AK, Singh UK, Kumar A. Online pharmacy: an e-strategy for medication. Int J Pharm Front Res 2011 http://www.ijpfr.com;1(1):146-158.

Abbreviations

ICC: intraclass correlation
MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by L Buis; submitted 25.04.24; peer-reviewed by F Carrouel, H Ali, S Shakeel; revised version received 18.11.24; accepted 10.02.25; published 17.04.25.

Please cite as:
Sattar A, Rehman H, Naveed S, Khadim S, Khan N, Kazi AF, Syed W, Al-Rawi MBA, Jamshed S
Trustworthiness of Web-Based Pharmacy Apps in Pakistan Based on the Mobile App Rating Scale: Content Analysis and Quality Evaluation
JMIR Mhealth Uhealth 2025;13:e59884
URL: https://mhealth.jmir.org/2025/1/e59884
doi:10.2196/59884

© Anum Sattar, Hina Rehman, Safila Naveed, Sumaira Khadim, Nargis Khan, Ahmad Furqan Kazi, Wajid Syed, Mahmood Basil A Al-Rawi, Shazia Jamshed. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

mHealth Apps Available in Italy to Support Health Care Professionals in Antimicrobial Stewardship Implementation: Systematic Search in App Stores and Content Analysis

Giuseppa Russo[1], MD; Annachiara Petrazzuolo[1], MD; Marino Trivisani[1], MD; Giuseppe Virone[1], MD; Elena Mazzolini[2], MD; Davide Pecori[3], MD; Assunta Sartor[4], MD; Sergio Giuseppe Intini[5], MD; Stefano Celotto[6], MD; Rossana Roncato[1], MD; Roberto Cocconi[7], MD; Luca Arnoldo[1,7], MD; Laura Brunelli[1,7], MD, PhD

[1]Department of Medicine, University of Udine, Via Colugna 50, Udine, Italy

[2]Department of Epidemiology, Experimental Zooprophylactic Institute of Venezia, Udine and Legnaro, Legnaro, Italy

[3]Infectious Disease Unit, Friuli Centrale Healthcare University Trust, Udine, Italy

[4]Microbiology Unit, Friuli Centrale Healthcare University Trust, Udine, Italy

[5]General Surgery Clinic and Liver Transplant Center, Friuli Centrale Healthcare University Trust, Udine, Italy

[6]Primary Care Department, Friuli Centrale Healthcare University Trust, Udine, Italy

[7]Accreditation, Quality, and Clinical Risk Unit, Friuli Centrale Healthcare University Trust, Udine, Italy

Corresponding Author:
Marino Trivisani, MD
Department of Medicine, University of Udine, Via Colugna 50, Udine, Italy

Abstract

Background: Antimicrobial resistance (AMR) is a major challenge of the 21st century. Digital technologies are now an increasingly effective means of supporting optimal health care delivery and public health.

Objective: The aim of this study was to explore the apps available to support health care professionals in the fight against AMR.

Methods: A total of 4 independent researchers conducted a systematic search of the App Store and Google Play Store using the following keywords: “antimicrobial resistance,” “antibiotic resistance,” “antimicrobial stewardship,” “antibiotic stewardship,” “antibiotic guide,” “antibiotic therapy,” and “antimicrobial therapy.” The same keywords were also searched in Italian. The apps whose contents were in languages other than Italian or English, or apps which were games, or had multimedia or paid content and advertising, or apps for only specific pathologies were not considered. A set of basic information was collected for all apps found. After downloading the apps, they were evaluated using an 86-item checklist containing expert-validated criteria aggregated in the domains of pathogens and etiological agents, diagnosis and therapy support, AMR, dashboard function, antimicrobial stewardship (AMS), notes and recordings, network, and technical characteristics of the app.

Results: First, 115 apps were identified: 31 apps for Android and 84 apps for iOS. By applying the exclusion criteria, 31 apps were excluded (16 for Android and 15 for iOS) for the following reasons: not available in Italian or English (6 apps), not freely available (14 apps), required registration (5 apps), and games (6 apps). The remaining 84 eligible apps (15 for Android and 69 for iOS) were downloaded, installed, and further analyzed using the same criteria, excluding 57 apps (48 for iOS and 9 for Android) for the following reasons: required further registration (16 apps), language other than Italian or English (17 apps), pathology specific (5 apps), paid content (8 apps), specific to veterinarians (4 apps), recreational apps (2 apps), referred to only scientific articles (1 app), no longer available (1 app), and not health care objectives (3 apps). The remaining 27 apps (6 for Android and 21 for iOS) were selected for in-depth analysis. Of the 27 apps that met the inclusion criteria, most apps did not fulfill the desirable aspects and only 2 of them achieved a fulfillment score of 36%. The highest scores were achieved for support for diagnosis and therapy (37%) and technical characteristics of the app (23%). Lower scores were achieved for AMS (8%), pathogens and etiological agents (4%), notes and records (3%), network (2%), AMR (1%), and dashboard function (1%).

Conclusions: None of the apps examined successfully provided the desired features and functions. To better engage of prescribers in the fight against AMR, the development of an app that meets the requirements is needed.

(JMIR Mhealth Uhealth 2025;13:e51122) doi:10.2196/51122

KEYWORDS

antimicrobial resistance; antimicrobial stewardship; mHealth; app; prescribing behaviors; one health; antimicrobial; Italy; application; support; therapy support; diagnosis; engagement; mobile health

XSL•FO
RenderX

Introduction

Antimicrobial resistance (AMR) is a serious problem [1] and remains one of the greatest global threats to public health at the beginning of the 21st century [2]. There is no single solution that can solve the problem alone, but a so-called theragnostic model based on infection prevention, antibiotic stewardship, and diagnostic stewardship programs is required [3]. It is estimated that bacterial infections will cause around 10 million deaths per year worldwide in 2050, far exceeding deaths from cancer (8.2 million), diabetes (1.5 million), or traffic accidents (1.2 million), with projected costs exceeding US $100 trillion [4]. In 2019, up to 4.95 million people died from diseases in which bacterial AMR played a role. Of these, 1.27 million deaths were the direct result of AMR—meaning that drug-resistant infections killed more people than HIV/AIDS (864,000 deaths) or malaria (643,000 deaths) [5]. In the United States alone, more than 2 million people become infected with multidrug-resistant organisms each year, resulting in 23,000 deaths per year, and the fact that the cost of antimicrobial resistance in the United States is estimated at US $20 billion per year underscores the need for action [6,7]. In addition, in the European Union (EU) and in the European Economic Area (EEA), there are more than 670,000 infections caused by antibiotic-resistant bacteria every year, and about 33,000 people die, as a direct result of these infections [8]. In Italy, antibiotic resistance remains one of the highest in Europe, and in our country up to 10% of patients contract multidrug-resistant bacterial infections every day, resulting in thousands of deaths. In Italy, about 284,100 patients are affected by health care–associated infections every year, resulting in about 4500 - 7000 deaths [9].

It is known that the inappropriate use of antimicrobials leads to unnecessary exposure of pathogens to antimicrobials, which over time leads to the development of AMR [10]. It is estimated that a mixed package of measures including responsible antibiotic use programs, improved hygiene, mass media campaigns, and the use of rapid diagnostic tests could prevent around 27,000 deaths per year in the EU and EEA [11]. However, the problem of antimicrobial resistance does not only affect the human sector, but also the environmental and veterinary sectors, where antibiotics are also widely used, and the impact of antimicrobial resistance is equally significant. Therefore, combating AMR cannot be done independently of a “One Health” approach that promotes coordinated action in the different sectors concerned [12]. As the World Health Organization (WHO) points out, the problem of AMR could also have a significant impact on a number of Sustainable Development Goals (SDGs). For example, the 2030 Agenda includes two specific indicators for AMR under SDG 3 (good health and well-being), including number 3.2—percentage of bloodstream infections due to selected antimicrobial-resistant organisms and number 3.3—proportion of health facilities that have a core stock of relevant essential medicines that are sustainably available and affordable [13].

In addressing such an impressive morbidity and mortality burden, digital mobile health (mHealth) technologies could play a crucial role in this worrying scenario. In fact, AMR counseling and support could be effectively aided by the use of information technology, big data, and artificial intelligence (AI) in health care. Despite debates over terminology, mHealth encompasses various types of solutions that could support medical and public health practices using mobile devices, such as cell phones, patient monitoring devices, personal digital assistants, and other wireless devices [14]. Such digital technologies are becoming an important resource for healthcare service delivery and public health. Among them, mobile wireless technologies are particularly important due to their ease of use, wide reach, and broad adoption [15]. To support health care professionals in antimicrobial stewardship, workflow integration with systems that use electronic health records (EHRs), including specific software or predictive AI, could be a good solution, but in areas where technology is not as advanced in the use of HHR or where access to EHR is limited or unavailable, the use of mobile phones, especially well-developed apps that support health care workers in antimicrobial stewardship (AMS), could be an effective key. Today, a large proportion of innovative digital health solutions come from start-ups that are building their services and technologies on Apple and Google mobile app platforms, using powerful analytics based on AI and machine learning, and developing new patient-oriented business models [16]. Among these solutions, smartphone apps are among the top 10 digital health technologies expected to impact the workforce between 2020 and 2040 [17]. Given this complex and challenging scenario, the aim of our study was to evaluate the mHealth apps available in Italy to support health care professionals and in particular prescribers, in the fight against AMR.

Methods

Step 1: Systematic Search

A systematic search was conducted by 4 independent researchers from the University of Udine, Italy, for the selection, data extraction, and functional evaluation of antimicrobial resistance apps available on the Apple App Store and Google Play Store. The Italian and English terms for antimicrobial resistance, antibiotic resistance, antimicrobial stewardship, antibiotic stewardship, antibiotic guide, antibiotic therapy and antimicrobial therapy were used as keywords in both stores. All apps that were found with at least 1 keyword mentioned above were included in the study. The choice to start the systematic search using keywords was necessary because there are too many healthcare apps and it was essential to reduce the action area. The following exclusion criteria applied to the search: content in languages other than Italian or English, games, leisure apps, photo and video apps, paid apps, paid content, advertising, apps that require registration, apps that only contain references to scientific articles, and apps that are not focused on health care. Paid apps and paid content were excluded as the focus of the research was to find apps that support health care professionals in the implementation of AMS and are, therefore, accessible to all health care professionals regardless of their financial means or affiliation with a public or private institution. Pathology specific apps were excluded as the aim of the study was to find a generic app that covers all diseases and systems. A set of basic information was collected for all apps found. This included the name of the app, operating system (iOS or

XSL•FO
RenderX

Android), availability of additional paid content, size (MB), number of downloads, position in the download ranking, user rating, copyright, privacy policy, possible presence of European medical device (CE) labelling, number of languages, store category, recipients, date of first release, last update, and responsibility of support. A total of 4 researchers conducted the search independently between June 15, 2022, and October 18, 2022, (2 researchers worked with iOS and 2 researchers worked with Android). All the contradictory information collected was discussed until a consensus was reached. Then the results were merged and app deduplication was performed.

Step 2: App Evaluation

A first draft of a checklist was created, listing the app features that were considered ideal for the app users. The main topics (domains) to be evaluated were related to the content of the app, notes and records, social support, and technical features. Individual questions were formulated for each domain based on the scientific literature [2,7,12] and the authors' previous experience in the field of mHealth apps for postnatal care [18] and for antenatal and postnatal care [19]. Possible answers were yes (y), no (n), or partially (p). The first draft of the checklist was commented on by a group of experts involved in AMR at different levels, including infectious disease specialists, microbiologists, clinicians, public health and infection risk management specialists, pharmacologists, and veterinarians. The valuable multidisciplinary collaboration between these professionals [20] helped to refine the assessment tool for the final draft of the 86-item checklist, which covers 8 domains, namely: pathogens and etiological agents, diagnostic and therapeutic support, AMR, dashboard function, AMS, notes and records, network, and technical features. Table 1 shows the final version of the checklist.

XSL•FO
RenderX

Table . Checklist for apps to support antimicrobial stewardship (AMS).

Domain	Item
Pathogens or etiological agents	1. Does the app contain information about the main microbiological characteristics of microorganisms?
	2. Does the app contain information about the characteristics of aerobic and anaerobic microorganisms?
	3. Does the app contain information about the cell wall of microorganisms?
	4. Does the app contain information about the main characteristics of fungal pathogens?
	5. Does the app contain information about the availability of vaccines that can prevent infections/diseases caused by infectious agents?
	6. Does the app include information about the methods and timeline for reporting infectious diseases?
	7. Does the app provide information on how antimicrobial drugs work?
	8. Does the app provide guidance on infection prevention and control (personal protective equipments, procedures)?
Diagnostic and therapeutic support	9. Does the app contain information about infections and their etiology?
	10. Does the app include information to help distinguish between a community-acquired infection (CAI) and a health care–acquired infection (HAI)?
	11. Does the app include references to diagnostic tests that are useful in formulating the diagnosis?
	12. Does the app support the user in selecting a drug for empiric therapy according to the clinical condition?
	13. Does the app indicate the route of administration of the medications indicated for therapy?
	14. Does the app provide specific information on drug pharmacokinetics and pharmacodynamics?
	15. Does the app indicate the expected duration of therapy, depending on the type of infection?
	16. Does the app support the user in adjusting medications based on the patient's kidney function?
	17. Does the app support the user in adjusting the dosage/medications to be prescribed in the case of a pediatric patient?
	18. Does the app provide information about possible side effects and allergic reactions to these medications?
	19. Does the app provide useful information for prescribing alternative medications if a patient is allergic to first-line medications?
	20. Does the app indicate the safety class of the medication for use during pregnancy?
	21. Does the app provide information about antifungal therapies?
	22. Does the app support the user in correctly reading an antibiogram?
	23. Does the app support the user in identifying the drug for targeted therapy against a specific microorganism?
	24. Does the app support the user in switching from empiric therapy to targeted therapy?
	25. Does the app enable information about the dosage to be prescribed?
	26. Does the app consider clinical variables related to the location of the infection and patient characteristics when choosing therapy?
	27. Does the app indicate the range of prescription of the drug (range A or range C or hospital regimen)?

XSL•FO
RenderX

Domain	Item
Antimicrobial resistance (AMR)	28. Does the app provide information on the developmental mechanism and types of antibiotic resistance (eg, plasmid, chromosomal)?
	29. Does the app provide data on the prevalence of antimicrobial resistance worldwide?
	30. Does the app provide data on the prevalence of antimicrobial resistance phenomena at the European level?
	31. Does the app provide data on the prevalence of antimicrobial resistance phenomena at the Italian level?
	32. Does the app provide data on the prevalence of antimicrobial resistance phenomena at the regional level?
	33. Does the app report the need to isolate the patient?
Dashboard function	34. Does the app provide up-to-date global/worldwide data on antimicrobial resistance profiles of isolates?
	35. Does the app report updated European-level data on antimicrobial resistance profiles of isolates?
	36. Does the app report updated data on the Italian-level on isolate antimicrobial resistance?
	37. Does the app report regionally updated data on the antimicrobial resistance profiles of isolates?
Antimicrobial stewardship (AMS)	38. Does the app provide information on the latest available guidelines?
	39. Do the guidelines or indications included in the app differentiate between prophylaxis and therapy?
	40. Does the app describe the use of antibiotics in prophylaxis (eg, antibiotic therapy in neutropenic leukemic patients, endocarditis prophylaxis)?
	41. Does the app include information about the circulation of the same pathogen between human-animal-environment?
	42. Does the app inform the user about the role of the veterinary component in AMR/AMS from a One Health perspective?
	43. Does the app inform the user about the role of the environmental component in AMR/AMS from a One Health perspective?
Notes and records	44. Does the app allow the user to enter data for an antimicrobial resistance database (eg, lack of response to therapy)?
	45. Does the app allow users to store notes or personal information?
	46. Does the app allow users uploading of clinical patient data?
	47. Does the app allow users to store individua therapy regimens (eg, clinical trials, innovative therapy associations)?
	48. Does the app provide warnings about potential unsafe uses of a particular therapy associated with specific clinical conditions?
	49. Does the app provide warnings about potential drug-drug or drug-food interactions?
Network	50. Does the app provide mechanisms for users to interact with other users?
	51. Does the app provide references and contacts of a service that prescribers can contact with clinical questions?
	52. Does the app provide different levels of interaction between users (eg, between peers, different professional profiles)?
	53. Does the app provide different levels of interaction that allow consultation with experts, pharmacologists, microbiologists, etc?
	54. Does the app provide other social mechanisms for users to share experiences?
	55. Is the app connected to the Italian post-marketing surveillance network (vigifarmaco) for reporting adverse effects?
	56. Is the app connected to the infectious disease reporting system?

XSL•FO
RenderX

Domain	Item
Technical features	57. Does the app ask users for authentication?
	58. Does the app have a privacy policy?
	59. Are all app contents freely available to the users (without any payment)?
	60. Are there specific inclusion criteria for full app usage (eg, authorization from a company/region/professional order)?
	61. Does the app require a workplace as an inclusion criterion for its full use?
	62. Does the app allow for preselection of treatment setting (primary care/long-term care facilities/nursing home/hospital)?
	63. Does the app require to "sign" an informed consent for app usage?
	64. Does the app include a glossary of the most used terms/abbreviations?
	65. Does the app identify the scientific responsibility of the provided contents?
	66. Is there a possibility to back-up/restore data within the app?
	67. Is there a possibility to download data collected through the app?
	68. Does the app have multilanguage support?
	69. Does the app interact with the hospital or local medical management software?
	70. Does the app geolocate the user to provide more detailed information?
	71. Does the app provide a real-time map of the AMR in relation to the country of reference (even if it is not the usual location of the practice)?
	72. Does the app allow users to update their account preferences?
	73. Does the app adapt to screen orientation (both portrait and landscape)?
	74. Does the app learn user's preferences over time?
	75. Does the app implement intuitive navigation patterns?
	76. Does the app implement predictable navigation patterns?
	77. Has the app content been validated by a local institutional source?
	78. Has the app content been validated by a regional institutional source?
	79. Has the app content been validated by a national institutional source?
	80. Has the app content been validated by a global institutional source?
	81. Is the app a certified medical device according to Italian law?
	82. Does the app provide content in text mode?
	83. Does the app provide content in audio mode?
	84. Does the app provide content in video mode?
	85. Can the app be used offline in its entirety?
	86. Does the app offer a technical support center to contact?

Ethical Considerations

This study did not include human subjects research (no human subjects experimentation or intervention was conducted) and so did not require institutional review board approval.

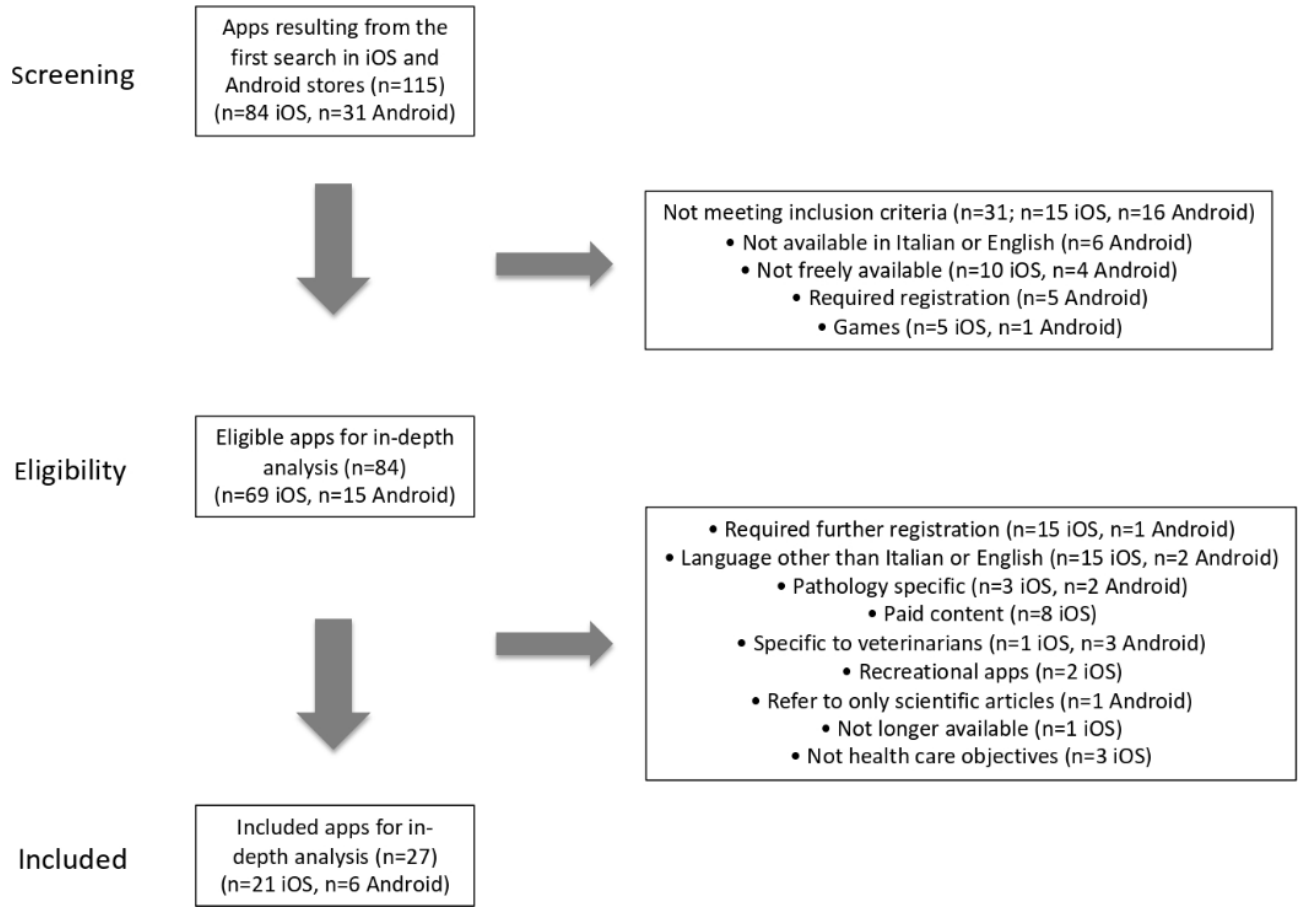
Results

Step 1: Systematic Search

First, 115 apps were identified: 31 apps for Android and 84 apps for iOS. By applying the exclusion criteria, 31 apps were excluded: 16 for Android and 15 for iOS. The remaining 84 eligible apps (15 for Android and 69 for iOS) were downloaded, installed, and further analyzed using the same criteria, excluding 57 apps: 48 for iOS and 9 for Android. The remaining 27 apps (6 for Android and 21 for iOS) were selected for in-depth analysis. Based on additional inclusion and exclusion criteria, as outlined in Figure 1, the flowchart illustrates the app selection process and outcomes.

XSL•FO
RenderX

Figure 1. Results of the app selection process.



Step 2: App Evaluation

After the entire selection process, 27 apps were evaluated using the checklist questions, of which 6 were available for Android (22%), 21 (78%) were available for iOS, and 1 was available in both stores. In total, 12 of the 27 apps were associated or affiliated with health care institutions or other international organizations. As shown in Table 2, considering all the potential items that could be fulfilled by the apps for each domain, of the 27 apps selected, diagnosis and therapy support (90/513, 37%) and app technical characteristics (187/810, 23%) were the most frequently fulfilled domains, followed by AMS (13/162, 8%), pathogens and etiological agents (8/216, 4%), notes and records (5/162, 3%), network (4/189, 2%), AMR (2/162, 1%), and dashboard function (1/108, 1%).

In the domain of pathogens and etiological agents, only 5 apps (19%) provided information on the mechanism of action of drugs, 2 apps (7%) provided information on the main characteristics of microorganisms, and 1 app provided information on the characteristics of aerobic and anaerobic microorganisms. In the domain of diagnostic and therapeutic support, 21 (78%) apps assisted the user in identifying the drug for targeted therapy against a specific microorganism, 18 (67%) apps provided information on the dosage to be prescribed, 17 apps (63%) indicated the route of administration for the drugs indicated for therapy, and 15 apps (56%) assisted the user in selecting the drug for empirical therapy according to the patient's clinical condition, and 15 apps (56%) considered the clinical variables related to the site of infection and patient characteristics in their therapy suggestions. In contrast, most apps (n=23, 85%) did not support the user in reading the antibiogram correctly. In the domain of AMR, hardly any app fulfilled the required characteristics, with the exception of one app that provided information on the evolution and types of AMR and one app that provided prevalence data on AMR in Italy, while no app provided prevalence data at European and Italian level.

In terms of dashboard functions, only 1 app responded to one of the items by providing regional data on AMR profiles of isolates; national and international data were not provided to the user. In relation to AMS, 8 apps (30%) provided information on the latest available guidelines, but little other information was provided on this topic. In particular, no app provided information on the possible circulation of the same pathogen between humans, animals, and the environment and did not inform the user about the role of the veterinary component in AMR and AMS from a One Health perspective. In terms of the network domain, no app was connected to the Italian postmarketing surveillance network for reporting possible adverse effects or to the infectious disease reporting system. No app offered opportunities to interact with other users or experts.

Regarding technical features, all apps (100%) had intuitive and predictable navigation patterns and offered content in text mode, the content of most apps (26/27, 96%) was available to users without payment, and 81% of the apps (22/27) could be used completely offline. However, none of the apps provided a

XSL•FO
RenderX

real-time map of the AMR in relation to the reference country or located the user to provide more detailed information. In addition, none of the apps interacted with the hospital or local medical management software, nor did they offer backup, restore, or download functions. In addition, no app offered multilingual support or was certified as a medical device under Italian law, and only 11 apps (41%) claimed scientific responsibility for the content provided. Multimedia Appendix 1 shows the detailed results of the app analysis.

Multimedia Appendix 2 shows the detailed results of the in-depth analysis of each of the 27 apps.

Considering the full list of items, no app met all the desirable criteria and the highest level of fulfilment of the desirable items identified by the checklist was 36%, which was only achieved by 2 apps. All other apps achieved fulfilment rates between 9% and 26%.

Table . Level of fulfillment for each domain.

	Information, functionality, and features provided by the app		
Domain[a]	Yes, n (%)	No, n (%)	Partially, n (%)
Pathogens and etiological agents (n=216)	8 (4)	208 (96)	0 (0)
Diagnostic and therapeutic support (n=513)	190 (37)	320 (62)	3 (1)
Antimicrobial resistance (n=162)	2 (1)	160 (99)	0 (0)
Dashboard function (n=108)	1 (1)	107 (99)	0 (0)
Antimicrobial stewardship (n=162)	13 (8)	149 (92)	0 (0)
Notes and records (n=162)	5 (3)	156 (96)	1(1)
Network (n=189)	4 (2)	185 (98)	0 (0)
Technical features (n=810)	187 (23)	623 (77)	0 (0)

[a]In brackets are reported the total number of items evaluated for accomplishment for each domain on the total of 27 apps.

Discussion

The aim of this study was to search and evaluate the mHealth apps available in Italy to support health care professionals and, in particular, prescribers in the fight against AMR.

Principal Findings

Most of the apps analyzed did not have most of the desirable features, with only two of them achieving a 36% fulfilment level. Most of these apps lacked the provision of content on pathogen characteristics, diagnostic and therapeutic support, AMR, and AMS. The dashboard function was generally lacking and did not allow users to view data on local, national, and international AMR prevalence. The mHealth solutions did not support a network system, which meant that organizational and technical support between professionals could not be provided. One of the most likely reasons why these apps did not meet the criteria selected by the expert group could be that they were designed and developed for other purposes and were not aimed at supporting health care professionals in AMS. Not having a dashboard function is missing an opportunity both at the national and international level, because having local data available could be useful to prescribers.

We are aware that we identified many features and that probably no app could have responded positively to all the items. Nevertheless, to reach our purpose to fighting AMR, we inserted all the elements we considered useful for the good functioning of an app in order to verify the distance between apps available for free and our ideal standards. Starting from this study, the next step we are planning is to create an app that fulfills most of our items.

We chose to provide data about the evaluation of each of the 27 apps in Multimedia Appendix 1. This makes it possible for readers to analyze apps also using subgroups of items. Ensuring an audit and feedback mechanism for prescribers by bringing together relevant information can indeed promote changes in practice, as is done in many clinical contexts [21] through quality improvement programs [22]. Therefore, the audit and feedback mechanism for prescribers and the opportunity for benchmarking should be supported to enhance their role in practice [23]. In addition, the recent pandemic has revolutionized the previous approach of health care professionals and researchers seeking access to organizational and institutional sources of information. This process toward greater transparency, accountability, and democratization of health data began with the introduction of the free availability of Italian COVID-19 surveillance data in 2020 [24]. The use of technology is not only changing the way people and professionals communicate with each other, but is also providing innovative ways to monitor the health and well-being of patients and the planet, and giving health care professionals better access to information for a more holistic view of health issues. Digital technologies are emerging as an important resource for health care and public health [15], but much remains to be done to translate evidence-based, validated information into clinical practice through medically certified devices. Probably the most recent, scientifically pervasive, and currently ground-breaking topic that is receiving massive and increasing investment in the public sector, but even more so in the private sector, is AI. In

XSL•FO
RenderX

health care, AI is already being used to some extent for chatbots (software applications or web interfaces, designed to have textual or spoken conversations) and triage systems to assess the need for doctor visits. It is also sometimes used in diagnostics, where it achieves accuracy comparable to that of experts, in predictive modelling, pharmaceutical research, and epidemiological studies [25]. As for its use in antimicrobial devices, ChatGPT (a comprehensive language model developed by OpenAI) appears to have access to sufficient training data (although it does not have access to specific medical databases), but has shortcomings in consistency, reasoning, and situational awareness that could jeopardize patient health. missing important aspects in scenarios of increasing complexity [26]. Despite these advances, problems remain regarding the robustness of studies and algorithmic bias. Furthermore, there are currently no specific regulations for AI-assisted care. If AI systems prove to be sufficiently reliable, they could provide significant support to health care providers. Although the use of AI in clinical practice is still limited, initial studies show promising results, suggesting that external validation and clinical trials are likely to increase in the near future [27]. Most importantly, local, regional, and national institutions and professionals need real-time data at the local, regional, and national level to make the best clinical and organizational decisions regarding antimicrobial stewardship. Therefore, the availability of features to train algorithms with real AMR and AMS data without the risk of data breaches is critical to the pursuit of this goal.

The Italian National Plan to Combat Antimicrobial Resistance (Piano Nazionale di Contrasto dell'Antimicrobico-Resistenza, PNCAR) recommends a One Health approach to contain AMR [12]. The apps analyzed in this study lack information on the circulation of zoonotic AMR pathogens between humans and animals and on the role of the veterinary component and the environment in the AMR and AMS problem, thus missing the opportunity to consider the broader One Health perspective. To effectively contribute to global health security by supporting prevention, early detection and warning, response and recovery for emerging and endemic communicable diseases, the One Health approach requires cross-sectoral integration and collaboration to efficiently and effectively bring together information, capacity and expertise to protect people, animals and our environment. On this basis, the tripartite collaboration between the Food and Agriculture Organization (FAO), the World Organisation for Animal Health (OIE) and the WHO has recently reaffirmed its commitment to cross-sectoral collaboration on disease risks at the human-animal-environment interface [28], which is also supported by many other global and local stakeholders [29]. Given the high economic risks of disease epidemics and their far-reaching societal and economic impacts, cross-sectoral One Health approaches are critical to improving prevention, detection, response, and recovery capacity [30]. In line with this vision, any action to address climate change can in turn benefit the global burden of disease, migration, conflict over natural resources and political instability—all of which are closely linked to the economic, environmental, and social determinants of health. In Europe, for example, climate change has been observed to lead to changes in rates of communicable and noncommunicable diseases due to temperature changes, increase the risk of outbreaks and exacerbations of clinical diseases (especially cardiovascular diseases), influence extreme meteorological events (eg, rainfall, floods, and hurricanes), cause sea level rise, and alter the distribution of plant species and arthropod vectors, as well as the quality of air and food. Nonetheless, climate change also has an impact on food security, as an increase in global temperature is associated with an increase in global demand for food [31].

Measures to improve the appropriateness of antimicrobial prescribing are a key component of AMS programs [10,32], as it is widely recognized that inappropriate use of antimicrobials increases the risk of treatment failures and adverse events in individual patients and accelerates the selection and transmission of antimicrobial-resistant pathogens in health care facilities and populations [33]. Nonetheless, the awareness and knowledge of prescribing physicians on these issues still needs to be improved and strengthened. To achieve this goal, much needs to be done, from preparation to ongoing education of health care professionals, to reinforce the message and support the behaviors. To promote a culture of optimal antibiotic use, the WHO published guidance in 2022 based on the AWaRe classification, which was introduced to support measures to improve antibiotic management. Antibiotics are categorized into 3 groups (Aware, Watch, and Reserve) according to their clinical importance and the risk that their use promotes the development of resistance. The guidelines contain evidence-based information on the selection, formulation, dosing, and duration of administration of antibiotics and defines clinical situations in which antibiotics are not recommended based on the best scientific evidence [34]. The WHO's "Firstline" app, which is only partially available to unregistered users, is also moving in this direction. This app provides various national and some local guidelines on different types of infections and some of their antibiotic treatment for adults and children, and its content is also available online [35]. In Italy, for example, "Firstline" is temporarily available for the pilot phase of the Enforcing Surveillance of Antimicrobial Resistance and Antibiotic Use to Drive Stewardship (ENSURE project) in Padua or for the users of the Integrated University Hospital of Verona after authentication. However, as this is a first attempt to address the problem, more advanced solutions are being sought to match guidelines with data and prescribing behavior to reverse the general trend. Antibiogo (Médecins Sans Frontières Foundation) is a mobile app, which was developed to help laboratory workers without expertise in low-resource settings measure and evaluate antibiotic susceptibility tests so that clinicians can prescribe the right antibiotics to their patients [36].

The European Centre for Disease Prevention and Control (ECDC) has defined core competencies for professionals in the field of infection control and hospital hygiene. These include minimum requirements common to all professionals in this field, ranging from leadership and management of infection prevention and control programs, microbiology and surveillance, infection prevention and control in clinical practice, education and training, quality, patient safety and occupational health [37]. Another issue to consider in relation to this topic is the sale of

XSL•FO
RenderX

over-the-counter (OTC) antibiotics. This is seen as a significant problem, particularly in regions of the world (ie, low- and middle-income countries) where prevalence is quite high. In these areas, it is not enough just to influence prescribers to tackle the problem of antimicrobial resistance; other measures must also be taken to control the use of antimicrobials. For this reason, the WHO has launched an electronic consultation in 2023 involving key stakeholders such as politicians, pharmacist associations, and civil society to identify possible actions for a global campaign to phase out over-the-counter antibiotics [38]. Despite these considerations and the awareness that there is still much to be done in this area, we believe that this study was able to provide an initial insight into the information that already exists and what could be done to further support health care professionals in combating AMR. For example, providing an institutionally validated app with content tailored to the individual needs of the health care professional and setting would help to tailor prescribing decisions to the context and tailor support to the prescriber's expertise. For example, the app could differentiate between primary care providers (general practitioners, family medicine, pediatrics, and internal medicine), acute care providers (inpatient or hospitalists), and infectious disease specialists to provide the most needed information.

Limitations

This study has several limitations. First, the exclusion criteria of this study limited the systematic search for the apps available in the Apple App Store and Google Play Store, specifically the requirement that the app must be available for free and without registration to be included in the in-depth analysis. These criteria aimed to evaluate the apps available to all health care professionals, regardless of their geographical or institutional affiliation. During the screening phase and the eligibility phase excluded apps were overall 22 of 115 (19%). Although aware that this share is relevant, we opted to exclude them anyway, because of the consideration that was more important that the contents were freely available to health care professionals, without the obligation of payment, in order to ensure equity for all economic levels. However, by choosing to include both Italian- and English-language apps, apps were selected that primarily target the educational level and professional profile of prescribing physicians. Second, we cannot know whether presenting the same checklist to other larger groups of experts would confirm all the items on the checklist we used to test the apps. Therefore, further future research along these lines may be useful to confirm, model, refine, and integrate domains and items. Third, the availability of the apps in stores at the time of the systematic search and a limited app validation panel may have limited this study. The search strategy, which involved the use of keywords, was probably unable to find some apps containing AMR modules. However, our goal was to find an app totally and not partially focused on AMR. Furthermore, this study does not value the length of therapy based on the organism and site of infection, surely an AMS domain. It could be a parameter to include in future works.

Finally, the actual effectiveness of the apps was not evaluated as this was not our focus. Nonetheless, this initial evaluation could form the basis for future work exploring this aspect in more depth.

Conclusions

This systematic search for apps available to Italian health care professionals to support antimicrobial prescribing in line with AMS principles shows that there is still much room for improvement in the development of apps and more generally in mHealth solutions to combat antimicrobial resistance. There is currently no user-friendly and useful app that meets all the requirements identified by the expert group; such a tool would support the engagement of prescribers in the fight against AMR.

Acknowledgments

Artificial intelligence was not used in the writing of this manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

LB, LA, RC, and EM designed the research. GR, AP, MT, and GV collected data. GR, AP, MT, GV, DP, AS, SGI, SC, RR, and LB discussed investigation methodology and contributed to result interpretation. GR, AP, GV, and MT performed data analysis. LB supervised the conduct of the study. GR and AP wrote the original draft. MT revised the paper and all authors agreed with the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Total number and percentage of apps meeting the items.
[DOCX File, 35 KB - mhealth_v13i1e51122_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Total number and percentage of accomplishments for each questionnaire item and globally for each of the 27 apps.
[DOCX File, 56 KB - mhealth_v13i1e51122_app2.docx]

References

1. MacDougall C, Polk RE. Antimicrobial stewardship programs in health care systems. Clin Microbiol Rev 2005 Oct;18(4):638-656. [doi: 10.1128/CMR.18.4.638-656.2005] [Medline: 16223951]
2. Global action plan on antimicrobial resistance. : World Health Organization; 2015 URL: https://iris.who.int/bitstream/handle/10665/193736/9789241509763_eng.pdf?sequence=1 [accessed 2025-04-14]
3. Dik JWH, Poelman R, Friedrich AW, et al. An integrated stewardship model: antimicrobial, infection prevention and diagnostic (AID). Future Microbiol 2016;11(1):93-102. [doi: 10.2217/fmb.15.99] [Medline: 26323589]
4. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations / the review on antimicrobial resistance chaired by Jim O'Neill. : Government of the United Kingdom; 2016 URL: https://apo.org.au/sites/default/files/resource-files/2016-05/apo-nid63983.pdf [accessed 2025-04-14]
5. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 2022 Feb 12;399(10325):629-655. [doi: 10.1016/S0140-6736(21)02724-0] [Medline: 35065702]
6. Core elements of hospital antibiotic stewardship programs. CDC. 2019. URL: https://www.cdc.gov/antibiotic-use/core-elements/hospital.html [accessed 2025-04-14]
7. Antibiotic resistance threats in the United States. CDC. 2013. URL: https://stacks.cdc.gov/view/cdc/20705/cdc_20705_DS1.pdf [accessed 2025-04-14]
8. Cassini A, Högberg LD, Plachouras D, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect Dis 2019 Jan;19(1):56-66. [doi: 10.1016/S1473-3099(18)30605-4] [Medline: 30409683]
9. Antibiotico-resistenza: un rischio globale che richiede strategie condivise [Web page in Italian]. AIFA. URL: https://www.aifa.gov.it/en/-/antibiotico-resistenza-un-rischio-globale-che-richiede-strategie-condivise [accessed 2025-04-14]
10. Ostrowsky B, Banerjee R, Bonomo RA, et al. Infectious diseases physicians: leading the way in antimicrobial stewardship. Clin Infect Dis 2018 Mar 19;66(7):995-1003. [doi: 10.1093/cid/cix1093] [Medline: 29444247]
11. Tackling the burden in the European Union. OECD. URL: https://www.oecd.org/health/health-systems/AMR-Tackling-the-Burden-in-the-EU-OECD-ECDC-Briefing-Note-2019.pdf [accessed 2025-04-14]
12. Piano nazionale di contrasto dell' antimicrobico-resistenza 2017-2020 [Report in Italian]. : Ministero della Salute; 2017 Oct 24 URL: Https://cdn.who.int/media/docs/default-source/antimicrobial-resistance/amr-spc-npm/nap-library/italy-national-plan-against-antimicrobial-resistance-2017-2020-(italian).pdf?sfvrsn=374e556e_1&download=true [accessed 2025-04-14]
13. Antimicrobial Resistance and the United Nations Sustainable Development Cooperation Framework: Guidance for United Nations Country Teams: WHO; 2021.
14. mHealth: new horizons for health through mobile technologies. : World Health Organization; 2011 URL: https://iris.who.int/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2025-04-14]
15. Measuring the information society report. : International Telecommunication Union; 2015 URL: https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2015.aspx [accessed 2025-04-14]
16. Shivkova M, Jahns R. Partnerships in digital health – breaking points and success factors. Research2Guidance. 2019. URL: https://research2guidance.com/product/partnerships-in-digital-health-breaking-points-and-success-factors/ [accessed 2025-04-18]
17. The Topol Review. NHS. 2019. URL: https://topol.hee.nhs.uk/the-topol-review/ [accessed 2025-04-14]
18. Sardi L, Idri A, Redman LM, Alami H, Bezad R, Fernández-Alemán JL. Mobile health applications for postnatal care: review and analysis of functionalities and technical features. Comput Methods Programs Biomed 2020 Feb;184:105114. [doi: 10.1016/j.cmpb.2019.105114] [Medline: 31655305]
19. Brunelli L, De Vita C, Cenedese F, et al. Gaps and future challenges of Italian apps for pregnancy and postnatal care: systematic search on app stores. J Med Internet Res 2021 Aug 10;23(8):e29151. [doi: 10.2196/29151] [Medline: 34383668]
20. D'Amour D, Ferrada-Videla M, San Martin Rodriguez L, Beaulieu MD. The conceptual basis for interprofessional collaboration: core concepts and theoretical frameworks. J Interprof Care 2005 May;19 Suppl 1(sup1):116-131. [doi: 10.1080/13561820500082529] [Medline: 16096150]
21. Colquhoun HL, Carroll K, Eva KW, et al. Advancing the literature on designing audit and feedback interventions: identifying theory-informed hypotheses. Implement Sci 2017 Sep 29;12(1):117. [doi: 10.1186/s13012-017-0646-0] [Medline: 28962632]
22. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012 Jun 13;2012(6):CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]
23. Di Blasio N, Acampora A, Angelici L, Pagano E, Ciccone G. Audit & feedback: come funziona. Recenti Prog Med 2022 Dec;113(12):733-738. [doi: 10.1701/3914.38975] [Medline: 36420849]

XSL•FO
RenderX

24. Ministerial circular N.1997 of 22 January 2020. : Ministero della Salute URL: https://www.trovanorme.salute.gov.it/norme/renderNormsanPdf?anno=2020&codLeg=72796&parte=1%20&serie=null [accessed 2025-04-14]
25. Panetto M. In salute. Intelligenza artificiale in medicina: buone prospettive ma studi ancora poco solidi [Web page in Italian]. : Il Bo Live; 2023 Mar 24. URL: https://ilbolive.unipd.it/it/news/salute-intelligenza-artificiale-medicina-buone [accessed 2025-04-14]
26. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? Lancet Infect Dis 2023 Apr;23(4):405-406. [doi: 10.1016/S1473-3099(23)00113-5] [Medline: 36822213]
27. De Corte T, Van Hoecke S, De Waele J. Artificial intelligence in infection management in the ICU. Crit Care 2022 Mar 22;26(1):79. [doi: 10.1186/s13054-022-03916-2] [Medline: 35337363]
28. Memorandum of Understanding between the United Nations Food and Agriculture Organization and the World Organisation for Animal Health and World Health Organization regarding cooperation to combat health risks at the human-animal-ecosystems interface in the context of the “One Health" approach and including antimicrobial resistance. FAO Knowledge Repository. 2018 May 30. URL: https://openknowledge.fao.org/server/api/core/bitstreams/0b6a5a41-4383-4840-acf0-ef374e07a4b3/content [accessed 2018-08-23]
29. Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, World Health Organization. The tripartite’s commitment: providing multi-sectoral, collaborative leadership in addressing health challenges. World Organisation for Animal Health. 2017 Oct. URL: https://www.woah.org/app/uploads/2018/05/tripartite-2017.pdf [accessed 2018-08-23]
30. Fan VY, Jamison DT, Summers LH. Pandemic risk: how large are the expected losses? Bull World Health Organ 2018 Feb 1;96(2):129-134. [doi: 10.2471/BLT.17.199588] [Medline: 29403116]
31. Climate change and health: fact sheet on Sustainable Development Goals (SDGs) : health targets. World Health Organization. 2018. URL: https://apps.who.int/iris/handle/10665/340820 [accessed 2025-04-14]
32. Doyle D, McDonald G, Pratt C, et al. Impact of a mobile decision support tool on antimicrobial stewardship indicators in St. John’s, Canada. PLoS ONE 2021;16(6):e0252407. [doi: 10.1371/journal.pone.0252407] [Medline: 34133424]
33. Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. BMC Infect Dis 2014 Jan 9;14:13. [doi: 10.1186/1471-2334-14-13] [Medline: 24405683]
34. Manuale Antibiotici AWaRe (Access, Watch, Reserve) Edizione Italiana Del “The WHO AWaRe Antibiotic Book [Book in Italian]: Agenzia Italiana del Farmaco (AIFA); 2023. URL: https://www.aifa.gov.it/documents/20142/1811463/Manuale_antibiotici_AWaRe.pdf [accessed 2025-04-14]
35. World Health Organization. Firstline. URL: https://app.firstline.org/en/clients/276-world-health-organization [accessed 2025-04-14]
36. Antibiogo. URL: http://www.antibiogo.org [accessed 2025-04-14]
37. Core competencies for infection prevention and control professionals. : World Health Organization; 2020 URL: Https://iris.who.int/bitstream/handle/10665/335821/9789240011656-eng.pdf?sequence=1 [accessed 2025-04-13]
38. Global campaign to phase out over-the-counter sales of antibiotics. : World Health Organization URL: https://cdn.who.int/media/docs/default-source/antimicrobial-resistance/amr-gcp-asa/report_antibiotics-campaign_web.pdf [accessed 2025-04-14]

Abbreviations

AI: artificial intelligence
AMR: antimicrobial resistance
AMS: antimicrobial stewardship
ECDC: European Centre for Disease Prevention and Control
EEA: European Economic Area
EHR: electronic health record
ENSURE: Enforcing Surveillance of Antimicrobial Resistance and Antibiotic Use to Drive Stewardship
EU: European Union
FAO: Food and Agriculture Organization
mHealth: mobile health
OTC: over-the-counter
PNCAR: Piano Nazionale di Contrasto dell’Antimicrobico-Resistenza
SDG: Sustainable Development Goal
WHO: World Health Organization

XSL•FO
RenderX

Edited by L Buis; submitted 21.07.23; peer-reviewed by KYY Kok, M Lee, R Marshall; revised version received 30.11.24; accepted 11.02.25; published 29.04.25.

Please cite as:
Russo G, Petrazzuolo A, Trivisani M, Virone G, Mazzolini E, Pecori D, Sartor A, Intini SG, Celotto S, Roncato R, Cocconi R, Arnoldo L, Brunelli L
mHealth Apps Available in Italy to Support Health Care Professionals in Antimicrobial Stewardship Implementation: Systematic Search in App Stores and Content Analysis
JMIR Mhealth Uhealth 2025;13:e51122
URL: https://mhealth.jmir.org/2025/1/e51122
doi:10.2196/51122

© Giuseppa Russo, Annachiara Petrazzuolo, Marino Trivisani, Giuseppe Virone, Elena Mazzolini, Davide Pecori, Assunta Sartor, Sergio Giuseppe Intini, Stefano Celotto, Rossana Roncato, Roberto Cocconi, Luca Arnoldo, Laura Brunelli. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review and Comparative Evaluation of Mobile Apps for Cardiovascular Risk Estimation: Usability Evaluation Using mHealth App Usability Questionnaire

Adrijana Svenšek[1*], RN, MSc; Lucija Gosak[1], RN, MSc; Mateja Lorber[1], RN, BSc, MSc, PhD; Gregor Štiglic[1,2,3], PhD; Nino Fijačko[1,4*], RN, MSc

[1]Faculty of Health Sciences, University of Maribor, Žitna ulica 15, Maribor, Slovenia

[2]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[3]Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[4]University Medical Centre, Maribor, Slovenia

[*]these authors contributed equally

Corresponding Author:
Adrijana Svenšek, RN, MSc
Faculty of Health Sciences, University of Maribor, Žitna ulica 15, Maribor, Slovenia

Abstract

Background: Cardiovascular diseases (CVD) are the leading cause of death and disability worldwide, and their prevention is a major public health priority. Detecting health issues early and assessing risk levels can significantly improve the chances of reducing mortality. Mobile apps can help estimate and manage CVD risks by providing users with personalized feedback, education, and motivation. Incorporating visual analysis into apps is an effective method for educating society. However, the usability evaluation and inclusion of visualization of these apps are often unclear and variable.

Objective: The primary objective of this study is to review and compare the usability of existing apps designed to estimate CVD risk using the mHealth App Usability Questionnaire (MAUQ). This is not a traditional usability study involving user interaction design, but rather an assessment of how effectively these applications meet usability standards as defined by the MAUQ.

Methods: First, we used predefined criteria to review 16 out of 2238 apps to estimate CVD risk in the Google Play Store and the Apple App Store. Based on the apps' characteristics (ie, developed for health care professionals or patient use) and their functions (single or multiple CVD risk calculators), we conducted a descriptive analysis. Then we also compared the usability of existing apps using the MAUQ and calculated the agreement among 3 expert raters.

Results: Most apps used the Framingham Risk Score (8/16, 50%) and Atherosclerotic Cardiovascular Disease Risk (7/16, 44%) prognostic models to estimate CVD risk. The app with the highest overall MAUQ score was the MDCalc Medical Calculator (mean 6.76, SD 0.25), and the lowest overall MAUQ score was obtained for the CardioRisk Calculator (mean 3.96, SD 0.21). The app with the highest overall MAUQ score in the "ease-of-use" domain was the MDCalc Medical Calculator (mean 7, SD 0); in the domain "interface and satisfaction," it was the MDCalc Medical Calculator (mean 6.67, SD 0.33); and in the domain "usefulness," it was the ASCVD Risk Estimator Plus (mean 6.80, SD 0.32).

Conclusions: We found that the Framingham Risk Score is the most widely used prognostic model in apps for estimating CVD risk. The "ease-of-use" domain received the highest ratings. While more than half of the apps were suitable for both health care professionals and patients, only a few offered sophisticated visualizations for assessing CVD risk. Less than a quarter of the apps included visualizations, and those that did were single calculators. Our analysis of apps showed that they are an appropriate tool for estimating CVD risk.

(JMIR Mhealth Uhealth 2025;13:e56466) doi:10.2196/56466

KEYWORDS

cardiovascular diseases; MAUQ; prognostic models; mobile applications; visualization; PRISMA

Introduction

The major cause of death worldwide is cardiovascular diseases (CVD). The prevalence of CVDs has increased globally due to factors such as obesity, poor nutrition, hypertension, and the incidence of type 2 diabetes [1,2]. CVDs still affect more than half a billion people worldwide, causing 20.5 million deaths in 2021, which is almost a third of all deaths worldwide [3]. The

XSL•FO
RenderX

approach to preventing CVD morbidity is to reduce and control risk factors through lifestyle changes and medication. Early detection and risk stratification further augment the chances of reducing mortality [4,5]. Many prognostic models have been developed in recent decades that assess the risk of CVDs [6,7], for example, the Framingham Heart Score for the American population [8,9], SCORE2 (Systematic Coronary Risk Evaluation 2) for European populations [10], QRISK (Cardiovascular Risk Score) for the British population [11-13], and artificial intelligence (AI)-based prediction model's [14] and others.

The rise of IT, such as mobile health (mHealth) tools, can make health care more accessible because of the integration of many prognostic models. It has been shown they can be used to reduce the risk of CVDs [6,15-17], improve glycemic control [18,19], control elevated blood pressure [20,21], contribute to the reduction in hospital admissions as part of cardiac rehabilitation [17,22], and also eHealth information exchange could give patients, who are from rural areas or have mobility problems, access to their health data – including the ability to capture and aggregate multiple sources of clinical data for performance measurement [23,24].

Growing research evidence supports the effectiveness of mobile apps for the prevention of CVDs [25-34]. Raising public awareness of various health issues can lead to the implementation of preventive measures [35]. The guidelines [36] state that, in general, visual aids (eg, graphs or images presenting the risk estimation) improve the understanding of risk and are more understandable than providing information in the form of numbers intended to improve results [35-38]. There is currently no official app or computer-based valid screening test that a health professional can use to show patients their health status through images [39]. Visual analytics enables efficient analysis and understanding of large datasets in real time [40]. Visual analysis, which is incorporated in apps, is an effective technique for educating society. The results of measures can be presented to help adults increase motivation and successfully change their behavior or lifestyle [28,34]. So, the results of measures can be presented to help adults increase motivation and successfully change their behavior or lifestyle [28,34].

The objective of this study is to review and compare the usability of existing estimations of CVD risk apps using the mHealth App Usability Questionnaire (MAUQ). Rather than focusing on traditional usability aspects related to user interaction design, this study assesses how effectively these apps adhere to usability standards as defined by the MAUQ.

Methods

This review used quantitative research methodology to review and compare the usability of existing estimations of CVD risk apps.

Systematic Review of Apps for Estimation of CVD Risk

In March 2023, we reviewed apps for estimating CVD risk available in the Google Play Store [41] using the Samsung Galaxy Tab S6 Lite and Samsung Galaxy S8+ with Android operating systems. In addition, we reviewed apps in the Apple App Store [42] using the iPhone 13 Pro Max and iPad Pro (3rd generation) with iOS operating systems. The review of the apps was conducted by two researchers separately from each other. To ensure consistency, any discrepancies between the two reviewers were resolved through discussion and consultation with a third reviewer. Search strings including "CVD", "CVD risk calculator", "cardiovascular diseases", and "cardiovascular risk" were used for searching the apps suitable for evaluation. According to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)(Checklist 1) [43] flow diagram, we searched for apps and excluded or included them based on their title, icon, and description in the app store.

The coding process categorized key aspects from all apps, names, and website links from each web store were manually transferred into Excel spreadsheets. The next step was to open and analyze each app on distribution platforms by looking into their titles, icons, screenshots or videos, and descriptions. Apps were included in the following steps if the title of the app is referenced to prognostic models for estimating CVD risk; the icon of the app represents some information about a possible way of estimating CVD risk; if the screenshots or video of the app gives an inside look at questions or results of prognostic models for estimation of the CVDs risk; if the description in the software platform provides information or references to a prognostic model for estimating CVD risk. Each relevant app was downloaded on a mobile phone or tablet in the final step. We have included free apps in the English language for various software platforms and their sections (Medical, Health and Wellness, Health and Fitness, etc), which estimate CVD risk and can be used either for personal (eg, estimating individual CVD risk in home environments) or for professional (eg, assessing patient CVD risk in clinical settings) purposes. Whether the app was developed for health care professionals or patients was determined according to the manufacturer's intended use. Apps were excluded in case of duplicates (from the search string and distribution platforms), non-English language, technical issues, paid apps, and apps not related to estimating CVD risk and having different purposes (games, quizzes, journals, etc).

We also categorized apps according to whether they were single calculators containing one prognostic model for estimating CVD risk or multi-calculators containing more than one prognostic model for estimating CVD risk, as done by Fijačko et al [44]. Results were summarized with means and SDs. We also calculated the intraclass correlation coefficient ($ICC_{2,k}$, intraclass correlation coefficient, 2-way random, average measures, and absolute agreement) [45-47].

Evaluating the Usability of Apps for Estimation of CVD Risk

To measure the usability of existing apps designed to estimate CVD risk, we used the MAUQ [48]. MAUQ responses can be aggregated into (1) "Ease of Use" (8 items), (2) "Interface and Satisfaction" (6 items), and (3) "Usefulness" domains (7 items). All responses to the evaluating questions use a Likert scale, rated on a scale of 1-7 and higher scores mean better usability.

XSL•FO
RenderX

The apps were evaluated by 3 experts with experience in health care informatics. The intraclass correlation coefficient ($ICC_{2,k}$, intraclass correlation coefficient, 2-way random, average measures, and absolute agreement) [45-47] was calculated to represent the agreement of the app ratings.

Statistical Analysis

Statistical analyses were conducted in October and November, 2023. Continuous variables were analyzed according to their Gaussian distribution and reported as mean with SD or 95% CI, whichever was appropriate. The data were analyzed using the SPSS statistical package version 29.0.0 (IBM Corp).

Results

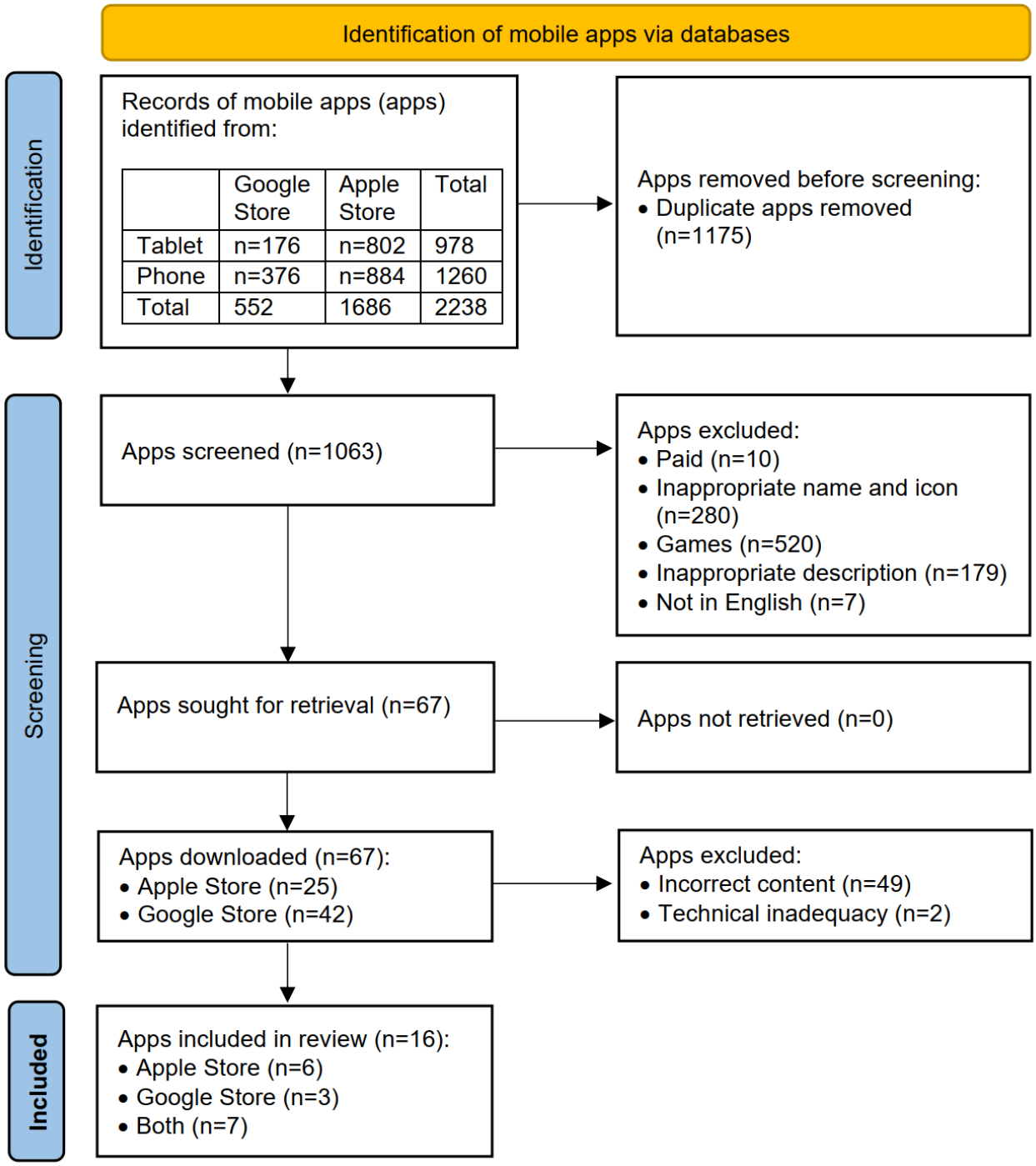
App Selection

We identified a total of 2238 apps across both platforms. In the Google Play Store, we found 176 apps using the tablet and 376 apps using the phone. In the Apple App Store, we identified 802 apps on the tablet and 884 apps on the phone. In the next step, duplicates were removed (1175/2238, 53%), leaving 1063 apps for review. Then we excluded apps for several reasons: they were not available for free (10/1063, 1%), had inappropriate icons or names (280/1063, 26%), were game based (520/1063, 47%), had inappropriate descriptions (179/1063, 17%), or were not available in English (7/1063, 1%). This process narrowed the selection to 67 apps for further review. In the second phase, we installed and examined these 67 apps in detail, ultimately excluding those with incorrect content (49/67, 76%) and technical inadequacy (2/67, 3%). Incorrect content was identified based on criteria such as lack of evidence-based information, incompleteness, or lack of personalization features tailored to individual user profiles. Apps that did not provide a prognostic model for the estimation of risk or did not include critical risk factors (eg, blood pressure, cholesterol levels) were also considered poor content. Technical shortcomings were identified by several key factors, including app stability (frequent crashes or failure to load), usability problems (difficult navigation, nonintuitive interface), and compatibility problems (apps that did not work properly across different devices or operating systems). Apps with outdated or nonfunctional features were also excluded, as were those that caused security problems (eg, lack of data encryption for sensitive user information). As these factors have a significant impact on the reliability and user experience of apps, they were crucial in our elimination process.

A total of 16 apps remained for analysis (Figure 1 [43]).

XSL•FO
RenderX

Figure 1. PRISMA flow diagram of the selection of included apps.



More characteristics of apps related to usability are presented in Multimedia Appendix 1.

Out of the 16 apps reviewed, 7 (44%) were compatible with both iOS and Android operating systems. Specifically, 4/16 (25%) apps were compatible with iOS, and 5/16 (31%) were compatible with Android. Compatibility refers to the ability of the apps to function on the respective operating systems without technical issues. Most of the apps (12/16, 75%) were designed for use by health care professionals in clinical settings as well as by individuals in home environments. A few apps (3/16, 19%) specifically indicated that they were intended for clinical use by health care professionals, including CV Risk Estimation, MediCalc (ScyMed, Inc), and CardioExpert I (Farid Belialov). In addition, one app (1/16, 6%), the Indigenous CVD Risk Calculator (An Tran Duy), was explicitly designed for use by individuals in home environments. The estimated results for CVD risk in apps were visualized in three formats across the apps: text, numerical, and graphical. Most apps (12/16, 75%) visualized results in both textual and numerical form. A smaller portion (4/16, 25%) presented CVD risk estimates in all 3 formats—text, numerical, and graphical—specifically in the following apps: ESC CVD Risk Calculation (European Society of Cardiology), Epi-RxlSK (University of Alberta), Cardiovascular Risk Calculator (www Machealth Pty Ltd), and Heartcare Lite (Heartcare Lite; Multimedia Appendix 2). All apps with graphical visualization were designed for use by

XSL•FO
RenderX

health care professionals and patients (Table 1 and Multimedia Appendix 2).

Table . Descriptive characteristics of apps.

Name of apps	MOS[a]	Single (S[b]) or multi-calculators (M[c])	Determination of use by developers (H[d] and I[e])	Visualization form of the cardiovascular disease (CVD) risk results
ESC CVD Risk Calculation	AND[f] and iOS[g]	M	H/I	Text, numerical, graphical
CardioCal	AND and iOS	S	H/I	Text, numerical
CardioRisk Calculator	AND and iOS	S	H/I	Text, numerical
ASCVD[j] Risk Estimator Plus	AND and iOS	S	H/I	Text, numerical
MediCalc	AND and iOS	M	H	Text, numerical
MDCalc Medical Calculator	AND and iOS	M	H/I	Text, numerical
Calculate by QxMD	AND and iOS	M	H/I	Text, numerical
CV Risk Estimation	iOS	M	H	Text, numerical
Indigenous CVD Risk Calculator	iOS	S	I	Text, numerical
Cardiovasculator Risk Calculator	iOS	S	H/I	Text, numerical, graphical
Epi-RxlSK	iOS	S	H/I	Text, numerical, graphical
Heartcare Lite	AND	S	H/I	Text, numerical, graphical
Framingham Score Heart Age	AND	S	H/I	Text, numerical
WHO[h]and ISH[i] Cardiovascular risk prediction charts	AND	S	H/I	Text, numerical
CardioExpert I	AND	S	H	Text, numerical
ASCVD[j] Risk	AND	S	H/I	Text, numerical

[a]MOS: mobile operating system.

[b]S: single calculator.

[c]M: multi-calculator.

[d]H: health professionals.

[e]I: individuals.

[f]AND: operation system for Android smartphones.

[g]iOS: Operation system for Apple smartphones.

[h]WHO: World Health Organization.

[i]ISH: International Society of Hypertension.

[j]ASCVD: atherosclerotic cardiovascular disease.

The reviewed apps feature 8 distinct prognostic models for estimating CVD risk, with the most common in all apps being the Framingham Risk Score (8/16, 50%), followed by the Atherosclerotic Cardiovascular Disease Risk model (7/16, 44%). Of the apps, 11/16 (69%) are single calculators using one prognostic model, while 5/16 (31%) are multi-calculators that incorporate multiple models. The multi-calculator app with the highest number of distinct prognostic models for estimating CVD risk was ESC CVD Risk Calculation, featuring 4 models (Table 2).

XSL•FO
RenderX

Table . Risk score method in apps.

Mobile apps	Single calculators (11/16, 69%)											Multi-calculators (5/16; 31%)					
Apps for estimating CVD risk	Cardio-Cal	CardioRisk Calculator	AS-CVD Risk Estimator Plus	Indigenous CVD Risk Calculator	Cardiovasculator Risk Calculator	Epi-RxI-SK	Heartcare Lite	Framingham Score Heart Age	WHO[b] and ISH[c] Cardiovascular risk prediction charts	CardioExpert I	AS-CVD Risk	ESC CVD Risk Calculation	MediCalc	MD-Calc Medical Calculator	Calculate by QxMD	CV Risk Estimation	Total PM[aa] in apps
Prognostic models (PM[a]) (n=8)																	
Framingham Risk Score		✓			✓	✓		✓					✓	✓	✓	✓	8
Atherosclerotic Cardiovascular Disease Risk			✓								✓	✓	✓	✓	✓	✓	7
Reynolds Risk Score													✓	✓	✓		3
WHO[b] and ISH[c] risk prediction charts				✓			✓		✓								3
SCORE[d] (Europe)										✓		✓					2
CVD[e] Risk	✓																1
SMART risk score[f]												✓					1
ADVANCE[g] risk score												✓					1
Total PM[a]:	1	1	1	1	1	1	1	1	1	1	1	4	3	3	3	2	26

[a]PM: prognostic models.

[b]WHO: World Health Organization.

[c]ISH: International Society of Hypertension.

[d]SCORE: Systematic Coronary Risk Evaluation.

XSL•FO
RenderX

[e]CVD: Cardiovascular disease.

[f]SMART risk score: which estimates 10-year risk of myocardial infarction (MI), stroke, or cardiovascular death.

[g]ADVANCE: the Action in Diabetes and Vascular disease: preterax and diamicron-MR controlled evaluation.

Empirical Evaluation of Estimation of CVD Risk Apps

The overall average MAUQ score for all the apps was 5.6/7 (95% CI 4.33 - 5.78). The app with the highest overall MAUQ score was MDCalc Medical Calculator 6.76 (0.25), followed by Heartcare Lite 6.58 (0.12), ASCVD Risk Estimator Plus 6.57 (0.18), ESC CVD Risk Calculation 6.11 (0.35), and Framingham Score Heart Age 6.03 (0.24). The app with the lowest MAUQ score was CardioRisk Calculator, 3.96 (0.21).

The "ease of use" domain had the highest overall average score across all domains (mean 6.20, SD 0.51). The apps with the highest overall MAUQ scores for ease of use were MDCalc Medical Calculator (mean 7, SD 0), followed by ESC CVD Risk Calculation (mean 6.92, SD 0.29), CardioCal (mean 6.73, SD 0.63), and Heartcare Lite (mean 6.60, SD 0.58). The lowest score in this domain was recorded for the CardioRisk Calculator (mean 5.27, SD 1.32).

The "interface and satisfaction" domain had the second highest overall average score (mean 5.25, SD 0.93). MDCalc Medical Calculator achieved the highest MAUQ score in this domain (mean 6.67, SD 0.33), followed by Heartcare Lite (mean 6.48, SD 0.79) and ASCVD Risk Estimator Plus (mean 6.24, SD 0.44). The lowest score in the "interface and satisfaction" domain was recorded by CardioRisk Calculator (mean 3.19, SD 1.13).

The "usefulness" domain received the lowest overall average score (mean 5.17, SD 1.13). The highest scores in this domain were achieved by ASCVD Risk Estimator Plus (mean 6.8, SD 0.32), followed by Heartcare Lite (mean 6.67, SD 0.58), MDCalc Medical Calculator (mean 6.60, SD 0.48), and Framingham Score Heart Age (mean 6, SD 0.32). The lowest score was recorded by MediCalc (mean 3.33, SD 0.47) (Table 3).

Table . mHealth App Usability Questionnaire mean scores.

Apps for estimating cardiovascular disease (CVD) risk	Ease of use, mean (SD)	Interface and satisfaction, mean (SD)	Usefulness, mean (SD)	Overall, mean (SD)[a]
ESC CVD Risk Calculation	6.92 (0.29)	5.95 (0.94)	5.47 (0.84)	6.11 (0.35)
CardioCal	6.73 (0.63)	5.48 (0.50)	4.07 (0.96)	5.43 (0.24)
CardioRisk Calculator	5.27 (1.32)	3.19 (1.13)	3.43 (0.89)	3.96 (0.21)
ASCVD[b] Risk Estimator Plus	6.67 (0.67)	6.24 (0.44)	6.80 (0.32)	6.57 (0.18)
MediCalc	5.73 (0.75)	4.29 (0.81)	3.33 (0.47)	4.45 (0.18)
MDCalc Medical Calculator	7.00 (0.00)	6.67 (0.33)	6.60 (0.48)	6.76 (0.25)
Calculate by QxMD	6.47 (0.26)	5.62 (0.37)	4.93 (1.11)	5.67 (0.46)
CV Risk Estimation	5.60 (0.95)	4.29 (0.58)	3.60 (0.41)	4.50 (0.42)
Indigenous CVD Risk Calculator	5.80 (0.63)	4.24 (0.49)	4.33 (0.52)	4.79 (0.08)
Cardiovasculator Risk Calculator	6.07 (0.86)	5.24 (0.50)	5.33 (0.54)	5.55 (0.20)
Epi-RxlSK	6.13 (0.48)	5.52 (0.57)	5.8 (1.62)	5.82 (0.64)
Heartcare Lite	6.6 (0.58)	6.48 (0.79)	6.67 (0.58)	6.58 (0.12)
Framingham Score Heart Age	6.33 (0.52)	5.76 (0.80)	6 (0.32)	6.03 (0.24)
WHO[c] and ISH[d] Cardiovascular risk prediction charts	6.2 (0.58)	4.57 (0.80)	5.20 (0.42)	5.32 (0.19)
CardioExpert I	5.93 (0.45)	5.19 (0.35)	5.53 (0.71)	5.55 (0.18)
ASCVD[b] Risk	5.73 (0.44)	5.29 (0.52)	5.6 (0.40)	5.54 (0.06)
Average score	6.2 (0.51)	5.25 (0.93)	5.17 (1.13)	5.60 (0.27)

[a]For the overall MAUQ score, we calculated reviewers inter-rater reliability using ICC, which showed excellent reliability ($ICC_{2,k}$ 0.87; 95% CI 0.80 - 0.91).

[b]ASCVD: atherosclerotic cardiovascular disease.

[c]WHO: World Health Organization.

[d]ISH: International Society of Hypertension.

XSL•FO
RenderX

Discussion

Principal Findings

Our review presented that apps for prognostic CVD risk estimation were available for both Android and iOS smartphones, designed for use by both health care professionals and patients. The most popular apps were single-calculator tools, each using only one CVD prognostic model. In contrast, the MAUQ highest-rated app, MDCalc Medical Calculator, offered access to multiple prognostic models, with the Framingham Risk Score being the most widely represented CVD prognostic model.

All apps were average quality, each achieving a high overall MAUQ score. In particular, the 'ease-of-use' domain received the highest scores. More than half of the apps can be used by health care professionals or patients, but only a few offer more sophisticated visualizations for assessing estimation of CVD risk. Less than a quarter of apps included visualization. The apps that did include visualization were single calculators. In this article, we found considerable variation in the usability of apps for estimation of CVD risk, highlighting areas where developers can focus on improving the user experience. These findings are key to guide future improvements of mHealth, in terms of accessibility, user satisfaction, and overall functionality.

Empirical Evaluation of Estimation of CVD Risk Apps

Our study assessed the effectiveness of these apps meeting usability standards as defined by the MAUQ of various apps designed to estimate CVD risk. Our findings revealed that most of the apps were predominantly characterized by positive feedback across multiple MAUQ questionnaire domains. In particular, the "ease-of-use" domain indicates a high level of practical acceptance and usability of the apps. Other studies have shown that the use of technology can promote shared decision-making by enabling health care professionals to manage chronic conditions [26,49-51]. The MAUQ evaluation tool has been used by several authors in different areas of health care. From ophthalmology (Australia) [52] to fitness (Malay) [53], breast cancer (German) [54], chatbot for reaching a patient (China) [55] etc. We also believe apps are an effective way to communicate the complex concept of clinical trials to patients and should also be incorporated in education curricula [37,56,57]. Based on the results, it can be concluded that the apps with the highest MAUQ scores are suitable for use by both health care professionals and patients. Rowland, et al [58] claim that despite the current limitations of diagnostic apps, there is a huge potential, and evidence is starting to emerge to demonstrate clinically significant improvements in morbidity and mortality outcomes in specific scenarios.

Forms of CVD Risk Score Presented in Apps

Recently, there has been a growing interest in the use of visualization in digital humanities, which extends the way we interact with traditional information visualization methods and guides analytical processes. The main goal is to incorporate user feedback to improve automated analysis [59-63]. In our study, a few apps, such as ESC CVD Risk Calculation, Epi-RxlSK, Cardiovascular Risk Calculator, and Heartcare Lite, featured a visual risk display that helped to understand the risks of CVD to patients in a comprehensible manner. The Zolezzi et al [64] article also reported that participants found the EPI-RXISK app (University of Alberta) visually appealing, with a professional layout and use of simple technology. Visualization offers new capabilities to analyze health care systems and support better decision-making and patient motivation, but on the other hand, the presence of visualization does not automatically guarantee good usability, and an app without advanced visualization can still provide a good user experience [57,65,66].

From our review of various apps, it is evident that most patient-oriented apps incorporate visualization. Typically, these are single-function calculators and are predominantly available for free. Conversely, apps aimed at CVD risk estimation can assist patients in modifying risk factors and lifestyle habits, improving medication adherence, quality of life, and psychosocial well-being, and are associated with better outcomes in managing CVD risk factors [67,68] and lead to increased adherence to primary prevention strategies and reduced health care costs [69,70]. For health care professionals, the priority is to obtain results quickly and have access to multiple prognostic models and calculators within a single app [71-74]. Consequently, apps designed for estimating CVD risk not only deliver crucial information but also provide educational resources and guidance to health care professionals. They help with prevention, raise awareness, and encourage control of risk factors [71,72]. In addition, apps for estimation of CVD risk employing visualization to display data can increase patient motivation [73,74].

Suggestions for Future Research

Electronic health record (EHR) systems are a digital representation of a patient's paper-based medical documentation. They have emerged in recent years as a promising avenue for advancing clinical research [75-77]. In a recent study, the authors examined differences in improvements in hypertension guideline implementation using standard EHRs and EHRs that incorporated the use of apps with visual analytics dashboards for the estimation of CVD risk. They found that incorporating a specific app's dashboard into EHRs has the potential to reduce the time and improve the implementation of hypertension guidelines [78]. The integration of apps with prognostic models for estimation of CVD risk directly into EHRs promises to be a valuable advancement for health care professionals [79-81]. This integration would allow the immediate estimation of CVD risk at the point of patient data entry, simplifying the process compared to using separate apps [82,83]. This efficiency could increase the effectiveness and quality of care management of CVD prevention strategies [84]. However, thorough evaluation and clinical testing are essential before apps can be seamlessly integrated into EHR systems and used in hospitals. In hospitals where the standard EHRs lack the capability to automatically calculate patient health risks, transitioning to EHRs with integrated apps with prognostic models would represent a substantial improvement in health care efficiency and accuracy. Even more, in the research by the author Serbanati [85] and other studies, they integrated AI into the EHRs and created an intelligent agent that acts as an avatar of the individual's health.

XSL•FO
RenderX

This agent collaborates with health care professionals, provides information on all aspects of an individual's health, proactively offers solutions, and helps them diagnose and decide on the right treatment [79,86].

Apps for estimation of CVD risk should be designed based on prognostic models for effective use in health care. Other studies conclude that the Framingham Risk Score is the most widely used prognostic model for the estimation of CVD risk. We also found out that the most used prognostic model in analyzed apps was the Framingham Risk Score, and no AI prognostic models were used in any of the apps [49-51,87]. The authors of the review article reference findings indicating that 50% of their studies have already incorporated prognostic models for estimating CVD risk, developed using AI [88]. Estimation of CVD risk using AI in general and CVDs prediction specifically is becoming increasingly common, but it is crucial to ask critical questions before using these prognostic models [89].

Limitations

Our study also has a few limitations. The first limitation is that we excluded paid apps, so we may have missed some quality apps that are maybe connected to EHRs or had prognostic models made by AI. The second limitation is that we focused only on usability and the inclusion of visualization in apps; we did not focus on other contextual analyses of the data, such as app usage in countries with the same Gross Domestic Product (GDP), apps for different races, disability levels, and so on. The latter could be a target for further research. The third limitation of the study is the apps industry, which is changing so rapidly that this type of review of apps is limited in time reliability. The fourth limitation is the number of MAUQ reviewers in this study and the representativeness of the target users' evaluators. The future of integrating apps for estimation of CVD risk and EHR in the field of prevention of CVDs is very promising. Data collected through EHRs, apps, wearables, and remote monitoring systems will enable health care organizations to identify trends, risk factors, and patterns in cardiovascular outcomes [90].

Conclusions

We presented the current state of apps for the estimation of CVD risk. We found that the most common prognostic model used in these apps was the Framingham Risk Score and that most of the apps were easy to use, indicating a high level of user satisfaction and acceptance. We discussed the benefits and challenges of using apps for the estimation of CVD risk in hospital settings by health care professionals or in home environments by patients. We suggested that apps for estimation of CVD risk be integrated into EHRs or other systems supported by health authorities to simplify the task of community risk estimation and improve the implementation of CVD prevention. However, we also acknowledged the need for further evaluation and testing of apps for estimation of CVD risk and prognostic models in clinical practice before they can be widely adopted and used in the hospital setting. We also argued that apps for estimation of CVD risk could provide patients valuable information, education, and guidance for prevention, as well as help modify risk factors and lifestyle habits, improve medication adherence, quality of life and psychosocial well-being, and reduce health care costs.

Acknowledgments

We would like to express our sincere gratitude to all the contributors and partners of the Innovative Didactic Technologies for Human and Environmental Health (InTECHiE; INOTEH-ZDRAV) project.

Adrijana Svenšek, Gregor Štiglic, Mateja Lorber and Nino Fijačko are founded by the Republic of Slovenia, Ministry of Higher Education, Science and Innovation, and the European Union – NextGenerationEU (Grant number: C3330-22-953012).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information of apps.
[DOCX File, 21 KB - mhealth_v13i1e56466_app1.docx]

Multimedia Appendix 2
Apps with visualization.
[PNG File, 357 KB - mhealth_v13i1e56466_app2.png]

Checklist 1
Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.
[DOCX File, 26 KB - mhealth_v13i1e56466_app3.docx]

References

1. Reynolds R, Dennis S, Hasan I, et al. A systematic review of chronic disease management interventions in primary care. BMC Fam Pract 2018 Jan 9;19(1):11. [doi: 10.1186/s12875-017-0692-3] [Medline: 29316889]

XSL•FO
RenderX

2. Timmis A, Vardas P, Townsend N, et al. European society of cardiology: cardiovascular disease statistics 2021. Eur Heart J 2022 Feb 22;43(8):716-799. [doi: 10.1093/eurheartj/ehab892] [Medline: 35016208]
3. Di Cesare M, Honor Bixby T, Lisa Hadeed C, et al. Global progress against cardiovascular disease (CVD) is flatlining. though rates of CVD deaths globally have fallen in the last three decades, this trend has begun to stall and, without concerted efforts, is at risk of reversing. In: Fox E, editor. World Heart Report 2023: Confronting the World's Number One Killer: World Heart Federation; 2023:1-52 URL: https://world-heart-federation.org/resource/world-heart-report-2023/ [accessed 2024-12-12] [doi: 10.5334/gh.1288]
4. Raihan M, Mondal S, More A, et al. Smartphone based heart attack risk prediction system with statistical analysis and data mining approaches. Adv sci technol eng syst j 2017;2(3):1815-1822 [FREE Full text] [doi: 10.25046/aj0203221]
5. Noncommunicable diseases. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2025-04-21]
6. Damen J, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016 May 16;353:i2416. [doi: 10.1136/bmj.i2416] [Medline: 27184143]
7. Wang X, Bakulski KM, Fansler S, et al. Improving the prediction of death from cardiovascular causes with multiple risk markers. medRxiv. Preprint posted online on Jan 23, 2023. [doi: 10.1101/2023.01.21.23284863] [Medline: 36747865]
8. Mahmood SS, Levy D, Vasan RS, et al. The Framingham heart study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014 Mar 15;383(9921):999-1008. [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]
9. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. Circulation 1998 May 12;97(18):1837-1847. [doi: 10.1161/01.cir.97.18.1837] [Medline: 9603539]
10. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J 2021 Jul 1;42(25):2439-2454. [doi: 10.1093/eurheartj/ehab309] [Medline: 34120177]
11. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ 2017 May 23;357:j2099. [doi: 10.1136/bmj.j2099] [Medline: 28536104]
12. Parsons RE, Liu X, Collister JA, et al. Independent external validation of the QRISK3 cardiovascular disease risk prediction model using UK Biobank. Heart 2023 Oct 26;109(22):1690-1697. [doi: 10.1136/heartjnl-2022-321231] [Medline: 37423742]
13. Badawy M, Naing L, Johar S, et al. Evaluation of cardiovascular diseases risk calculators for CVDs prevention and management: scoping review. BMC Public Health 2022 Sep 14;22(1):1742. [doi: 10.1186/s12889-022-13944-w] [Medline: 36104666]
14. Schorr EN, Gepner AD, Dolansky MA, et al. Harnessing mobile health technology for secondary cardiovascular disease prevention in older adults: a scientific statement from the American heart association. Circ Cardiovasc Qual Outcomes 2021 May;14(5):e000103. [doi: 10.1161/HCQ.0000000000000103] [Medline: 33793309]
15. Iyengar N, Vivekanandan T, Ahmed ST. Big data analytic approach to predict risk assessment for cardiovascular diseases using framingham risk score. IJCLI 2023;2:32-38 [FREE Full text]
16. Krzowski B, Peller M, Boszko M, et al. Mobile app and digital system for patients after myocardial infarction (afterAMI): study protocol for a randomized controlled trial. Trials 2022 Jun 21;23(1):522. [doi: 10.1186/s13063-022-06463-x] [Medline: 35729626]
17. Widmer RJ, Allison TG, Lerman LO, et al. Digital health intervention as an adjunct to cardiac rehabilitation reduces cardiovascular risk factors and rehospitalizations. J Cardiovasc Transl Res 2015 Jul;8(5):283-292. [doi: 10.1007/s12265-015-9629-1] [Medline: 25946990]
18. Hoda F, Arshad M, Khan MA, et al. Impact of a mHealth intervention in type 2 diabetes mellitus patients: randomized clinical trial. SN Compr Clin Med 2023;5(1). [doi: 10.1007/s42399-023-01564-3]
19. Zimmermann G, Venkatesan A, Rawlings K, et al. Improved glycemic control with a digital health intervention in adults with type 2 diabetes: retrospective study. JMIR Diabetes 2021 Jun 2;6(2):e28033. [doi: 10.2196/28033] [Medline: 34075880]
20. Skolarus LE, Dinh M, Kidwell KM, et al. Reach out emergency department: a randomized factorial trial to determine the optimal mobile health components to reduce blood pressure. Circ Cardiovasc Qual Outcomes 2023 May;16(5):352-362. [doi: 10.1161/CIRCOUTCOMES.122.009606] [Medline: 37192282]
21. Yilmaz E, Uzuner A, Bajgora M, et al. Effect of eTansiyon smartphone application on hypertension control. Prim Health Care Res Dev 2022 Oct 19;23:e64. [doi: 10.1017/S146342362100058X] [Medline: 36259143]
22. Braver J, Marwick TH, Oldenburg B, Issaka A, et al. Digital health programs to reduce readmissions in coronary artery disease. JACC: Advances 2023 Oct;2(8):100591. [doi: 10.1016/j.jacadv.2023.100591]
23. Aminizadeh S, Heidari A, Toumaj S, et al. The applications of machine learning techniques in medical data processing based on distributed computing and the internet of things. Comput Methods Programs Biomed 2023 Nov;241:107745. [doi: 10.1016/j.cmpb.2023.107745] [Medline: 37579550]
24. Paul M, Maglaras L, Ferrag MA, et al. Digitization of healthcare sector: a study on privacy and security concerns. ICT Express 2023 Aug;9(4):571-588. [doi: 10.1016/j.icte.2023.02.007]

XSL•FO
RenderX

25. Adler AJ, Martin N, Mariani J, et al. Mobile phone text messaging to improve medication adherence in secondary prevention of cardiovascular disease. Cochrane Database Syst Rev 2017 Apr 29;4(4):CD011851. [doi: 10.1002/14651858.CD011851.pub2] [Medline: 28455948]
26. Chatterjee A, Prinz A, Gerdes M, et al. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021 Nov 17;23(11):e26931. [doi: 10.2196/26931] [Medline: 34787575]
27. Lee C, Lee K, Lee D. Mobile healthcare applications and gamification for sustained health maintenance. Sustainability 2017;9(5):772. [doi: 10.3390/su9050772]
28. Lobelo F, Kelli HM, Tejedor SC, et al. The wild wild West: a framework to integrate mHealth software applications and wearables to support physical activity assessment, counseling and interventions for cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016;58(6):584-594. [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
29. Maddison R, Rawstorn JC, Shariful Islam SM, et al. mHealth Interventions for Exercise and Risk Factor Modification in Cardiovascular Disease. Exerc Sport Sci Rev 2019 Apr;47(2):86-90. [doi: 10.1249/JES.0000000000000185] [Medline: 30883472]
30. Martin SS, Feldman DI, Blumenthal RS, et al. mActive: a randomized clinical trial of an automated mHealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 9;4(11):1-9. [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
31. Näslund U, Ng N, Lundgren A, et al. Visualization of asymptomatic atherosclerotic disease for optimum cardiovascular prevention (VIPVIZA): a pragmatic, open-label, randomised controlled trial. Lancet 2019 Jan 12;393(10167):133-142. [doi: 10.1016/S0140-6736(18)32818-6] [Medline: 30522919]
32. Park LG, Howie-Esquivel J, Chung ML, et al. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268. [doi: 10.1016/j.pec.2013.10.027] [Medline: 24321403]
33. Park CW, Seo SW, Kang N, et al. Artificial intelligence in health care: current applications and issues. J Korean Med Sci 2020 Nov 2;35(42):e379. [doi: 10.3346/jkms.2020.35.e379] [Medline: 33140591]
34. Stiglic G, Kocbek P, Fijacko N, et al. Interpretability of machine learning‐based prediction models in healthcare. WIREs Data Min & Knowl 2020 Sep;10(5):1-12 [FREE Full text] [doi: 10.1002/widm.1379]
35. Mansoor H, Gerych W, Alajaji A, et al. INPHOVIS: Interactive visual analytics for smartphone-based digital phenotyping. Visual Informatics 2023 Jun;7(2):13-29. [doi: 10.1016/j.visinf.2023.01.002]
36. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2021 Sep 7;42(34):3227-3337. [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
37. Chatzimparmpas A, Martins RM, Kerren A. VisRuler: visual analytics for extracting decision rules from bagged and boosted decision trees. Inf Vis 2023 Apr;22(2):115-139. [doi: 10.1177/14738716221142005]
38. Moradi H, Al-Hourani A, Concilia G, et al. Recent developments in modeling, imaging, and monitoring of cardiovascular diseases using machine learning. Biophys Rev 2023 Feb;15(1):19-33. [doi: 10.1007/s12551-022-01040-7] [Medline: 36909958]
39. Zhang C, Lakens D, IJsselsteijn WA. Theory integration for lifestyle behavior change in the digital age: an adaptive decision-making framework. J Med Internet Res 2021 Apr 9;23(4):e17127. [doi: 10.2196/17127] [Medline: 33835036]
40. Raghupathi W, Raghupathi V, Saharia A. Analyzing health data breaches: a visual analytics approach. AppliedMath 2023;3(1):175-199. [doi: 10.3390/appliedmath3010011]
41. Google play store. Google Play Store. 2023. URL: https://play.google.com/store/games?hl=en_IN [accessed 2025-04-22]
42. Apple Inc. Apple Store. 2023. URL: https://www.apple.com/in/app-store/ [accessed 2025-04-22]
43. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
44. Fijačko N, Masterson Creber R, Gosak L, et al. A review of mortality risk prediction models in smartphone applications. J Med Syst 2021 Nov 4;45(12):107. [doi: 10.1007/s10916-021-01776-x] [Medline: 34735603]
45. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic 1981 Sep;86(2):127-137. [Medline: 7315877]
46. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
47. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice: Prentice Hall; 2009. URL: https://www.scirp.org/reference/referencespapers?referenceid=1952134 [accessed 2024-12-22]
48. Zhou L, Bao J, Setiawan IMA, et al. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
49. Abohelwa M, Kopel J, Shurmur S, et al. The Framingham study on cardiovascular disease risk and stress-defenses: a historical review. JVD 2023;2(1):122-164. [doi: 10.3390/jvd2010010]
50. Arabiyat S, Tadros O, Al-Daghastani T, et al. Applying the Framingham risk score for cardiovascular diseases in Jordan: a cross sectional study. Trop J Pharm Res 2022;21(12):2659-2667. [doi: 10.4314/tjpr.v21i12.22]

XSL•FO
RenderX

51. Hasani-Ranjbar S, Razmandeh R, Ghodssi-Ghassemabadi R, et al. Comparison of Framingham cardiovascular risk criteria and ASCVD score in Iranian obese patients. Iran J Public Health 2023 Feb;52(2):420-426. [doi: 10.18502/ijph.v52i2.11895] [Medline: 37089142]
52. Chumkasian W, Fernandez R, Win KT, et al. Adaptation of the MAUQ and usability evaluation of a mobile phone-based system to promote eye donation. Int J Med Inform 2021 Jul;151:104462. [doi: 10.1016/j.ijmedinf.2021.104462] [Medline: 33933903]
53. Mustafa N, Safii NS, Jaffar A, et al. Malay version of the mHealth app usability questionnaire (M-MAUQ): translation, adaptation, and validation study. JMIR Mhealth Uhealth 2021 Feb 4;9(2):e24457. [doi: 10.2196/24457] [Medline: 33538704]
54. Moorthy P, Weinert L, Harms BC, et al. German version of the mHealth app usability questionnaire in a cohort of patients with cancer: translation and validation study. JMIR Hum Factors 2023 Nov 1;10:e51090. [doi: 10.2196/51090] [Medline: 37910144]
55. Shan Y, Ji M, Xie W, et al. Chinese Version of the Mobile Health App Usability Questionnaire: Translation, Adaptation, and Validation Study. JMIR Form Res 2022 Jul 6;6(7):e37933. [doi: 10.2196/37933] [Medline: 35793132]
56. Iyengar N, Vivekanandan T, Thouheed Ahmed S. Big data analytic approach to predict cardiovascular diseases risk assessment using Framingham risk score. Int J Comput Intell 2020;2(1):32-38.
57. Ding MY, Wang WT. Analysis of factors influencing we-intention in healthcare applications based on the AISAS model. Int J Hum-Comput Interact 2024 May 18;40(10):2560-2577. [doi: 10.1080/10447318.2022.2163566]
58. Rowland SP, Fitzgerald JE, Holme T, et al. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
59. Ooge J, Stiglic G, Verbert K. Explaining artificial intelligence with visual analytics in healthcare. WIREs Data Min & Knowl 2022 Jan;12(1):1-19. [doi: 10.1002/widm.1427]
60. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. Presented at: 1996 IEEE Symposium on Visual Languages; Sep 3-6, 1996; Boulder, CO, USA p. 336-343. [doi: 10.1109/VL.1996.545307]
61. Afzal S, Ghani S, Hittawe MM, et al. Visualization and visual analytics approaches for image and video datasets: a survey. ACM Trans Interact Intell Syst 2023 Mar 31;13(1):1-41. [doi: 10.1145/3576935]
62. Backonja U, Haynes SC, Kim KK. Data visualizations to support health practitioners' provision of personalized care for patients with cancer and multiple chronic conditions: user-centered design study. JMIR Hum Factors 2018 Oct 16;5(4):e11826. [doi: 10.2196/11826] [Medline: 30327290]
63. Whitmore K, Zhou Z, Chapman N, et al. Impact of patient visualization of cardiovascular images on modification of cardiovascular risk factors: systematic review and meta-analysis. JACC Cardiovasc Imaging 2023 Aug;16(8):1069-1081. [doi: 10.1016/j.jcmg.2023.03.007] [Medline: 37227327]
64. Zolezzi M, Elhakim A, Hejazi T, et al. Translating and piloting a cardiovascular risk assessment and management online tool using mobile technology. Saudi Pharm J 2023 Apr;31(4):492-498. [doi: 10.1016/j.jsps.2023.02.001] [Medline: 37063447]
65. Chatterjee P, Cymberknop LJ, Armentano RL. IoT-based decision support system for intelligent healthcare — applied to cardiovascular diseases. Presented at: 2017 7th International Conference on Communication Systems and Network Technologies (CSNT); Nov 11-13, 2017; Nagpur, India p. 362-366. [doi: 10.1109/CSNT.2017.8418567]
66. Mishra N, Desai NP, Wadhwani A, et al. Visual analysis of cardiac arrest prediction using machine learning algorithms: a health education awareness initiative. In: Garcia MB, Lopez Cabrera MV, Almeida RPP, editors. Handbook of Research on Instructional Technologies in Health Education and Allied Disciplines: IGI Global; 2023:331-363. [doi: 10.4018/978-1-6684-7164-7.ch015]
67. Spaulding EM, Marvel FA, Piasecki RJ, et al. User engagement with smartphone apps and cardiovascular disease risk factor outcomes: systematic review. JMIR Cardio 2021 Feb 3;5(1):e18834. [doi: 10.2196/18834] [Medline: 33533730]
68. Tekkeşin A, Hayıroğlu M, Çinier G, et al. Lifestyle intervention using mobile technology and smart devices in patients with high cardiovascular risk: a pragmatic randomised clinical trial. Atherosclerosis 2021 Feb;319:21-27. [doi: 10.1016/j.atherosclerosis.2020.12.020] [Medline: 33465658]
69. Iuga AO, McGuire MJ. Adherence and health care costs. Risk Manag Healthc Policy 2014;7:35-44. [doi: 10.2147/RMHP.S19801] [Medline: 24591853]
70. Kaminsky LA, German C, Imboden M, et al. The importance of healthy lifestyle behaviors in the prevention of cardiovascular disease. Prog Cardiovasc Dis 2022;70:8-15. [doi: 10.1016/j.pcad.2021.12.001] [Medline: 34922952]
71. Baek H, Suh JW, Kang SH, et al. Enhancing user experience through user study: design of an mHealth tool for self-management and care engagement of cardiovascular disease patients. JMIR Cardio 2018 Feb 9;2(1):e3. [doi: 10.2196/cardio.9000] [Medline: 31758783]
72. Svenšek A, Lorber M, Gosak L, et al. The role of visualization in estimating cardiovascular disease risk: scoping review. JMIR Public Health Surveill 2024 Oct 14;10:e60128. [doi: 10.2196/60128] [Medline: 39401079]
73. Nelson AJ, Pagidipati NJ, Bosworth HB. Improving medication adherence in cardiovascular disease. Nat Rev Cardiol 2024 Jun;21(6):417-429. [doi: 10.1038/s41569-023-00972-1] [Medline: 38172243]
74. Moungui HC, Nana-Djeunga HC, Anyiang CF, et al. Dissemination of mobile health applications: a systematic review (preprint). Preprint posted online on Jun 26, 2023. [doi: 10.2196/preprints.50293]

XSL•FO
RenderX

75. Forde-Johnston C, Butcher D, Aveyard H. An integrative review exploring the impact of electronic health records (EHR) on the quality of nurse-patient interactions and communication. J Adv Nurs 2023 Jan;79(1):48-67. [doi: 10.1111/jan.15484] [Medline: 36345050]
76. Nath KA. In the limelight: March 2023. Mayo Clin Proc 2023 Mar;98(3):357-359. [doi: 10.1016/j.mayocp.2023.01.017]
77. Weiskopf NG, Dorr DA, Jackson C, Lehmann HP, Thompson CA. Healthcare utilization is a collider: an introduction to collider bias in EHR data reuse. J Am Med Inform Assoc 2023 Apr 19;30(5):971-977. [doi: 10.1093/jamia/ocad013] [Medline: 36752649]
78. Fadel RA, Ross J, Asmar T, et al. Visual analytics dashboard promises to improve hypertension guideline implementation. Am J Hypertens 2021 Oct 27;34(10):1078-1082. [doi: 10.1093/ajh/hpab081] [Medline: 34043744]
79. Holmgren AJ, Thombley R, Sinsky CA, et al. Changes in physician electronic health record use with the expansion of telemedicine. JAMA Intern Med 2023 Dec 1;183(12):1357-1365. [doi: 10.1001/jamainternmed.2023.5738] [Medline: 37902737]
80. Ratwani RM. Electronic health records and improved patient care: opportunities for applied Psychology. Curr Dir Psychol Sci 2017 Aug;26(4):359-365. [doi: 10.1177/0963721417700691] [Medline: 28808359]
81. Goldstein BA, Navar AM, Pencina MJ, et al. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208. [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
82. Eapen ZJ, Turakhia MP, McConnell MV, et al. Defining a mobile health roadmap for cardiovascular health and disease. J Am Heart Assoc 2016 Jul 12;5(7):e003119. [doi: 10.1161/JAHA.115.003119] [Medline: 27405809]
83. Neubeck L, Coorey G, Peiris D, et al. Development of an integrated e-health tool for people with, or at high risk of, cardiovascular disease: the consumer navigation of electronic cardiovascular tools (CONNECT) web application. Int J Med Inform 2016 Dec;96:24-37. [doi: 10.1016/j.ijmedinf.2016.01.009] [Medline: 26847070]
84. Plantier M, Havet N, Durand T, et al. Does adoption of electronic health records improve the quality of care management in France? Results from the French e-SI (PREPS-SIPS) study. Int J Med Inform 2017 Jun;102:156-165. [doi: 10.1016/j.ijmedinf.2017.04.002] [Medline: 28495344]
85. Serbanati LD. Health digital state and smart EHR systems. Inform Med Unlocked 2020;21:100494. [doi: 10.1016/j.imu.2020.100494]
86. Militello C, Iacona SL, Dan Serbanati L, et al. A virtual health record-based EHR system. Presented at: 2015 E-Health and Bioengineering Conference (EHB); Nov 19-21, 2015; lasi, Romania p. 1-6. [doi: 10.1109/EHB.2015.7391425]
87. Raj MP, Krishnamurthy V, Basu E, et al. Comparison of Framingham risk scores (FRS), joint British society (JBS3), and American college of cardiology/American heart association (ACC/AHA) cardiovascular risk scores among adults with first myocardial infarction. Cureus 2023 Jan;15(1):e33221. [doi: 10.7759/cureus.33221] [Medline: 36733551]
88. Mohsen F, Al-Saadi B, Abdi N, et al. Artificial intelligence-based methods for precision cardiovascular medicine. J Pers Med 2023 Aug 16;13(8):1268. [doi: 10.3390/jpm13081268] [Medline: 37623518]
89. van Smeden M, Heinze G, Van Calster B, et al. Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. Eur Heart J 2022 Aug 14;43(31):2921-2930. [doi: 10.1093/eurheartj/ehac238] [Medline: 35639667]
90. Ullah M, Hamayun S, Wahab A, et al. Smart technologies used as smart tools in the management of cardiovascular disease and their future perspective. Curr Probl Cardiol 2023 Nov;48(11):101922. [doi: 10.1016/j.cpcardiol.2023.101922] [Medline: 37437703]

Abbreviations

ADVANCE: Action in Diabetes and Vascular disease: preterAx and diamicroN-MR ControllEd evaluation
AI: artificial intelligence
CVD: cardiovascular disease
CVI: Content Validity Index
EHR: electronic health record
GDP: gross domestic product
I-CVI: Item-CVI
ICC_{2k}: intraclass correlation coefficient
iOS: operation system for Apple smartphones
ISH: International Society of Hypertension
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
MOS: Mobile Operating System
PM: Prognostic models
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QRISK: Cardiovascular Risk Score
S-CVI: Scale-level CVI

XSL•FO
RenderX

SCORE: Systematic Coronary Risk Evaluation
SCORE2: Systematic Coronary Risk Evaluation 2
SMART RISK score: which estimates 10 years risk of myocardial infarction (MI), stroke, or cardiovascular death
WHO: World Health Organization

Edited by L Buis, TDA Cardoso; submitted 17.01.24; peer-reviewed by M Chlabicz, O El-Gayar; revised version received 21.11.24; accepted 25.03.25; published 08.05.25.

Please cite as:
Svenšek A, Gosak L, Lorber M, Štiglic G, Fijačko N
Review and Comparative Evaluation of Mobile Apps for Cardiovascular Risk Estimation: Usability Evaluation Using mHealth App Usability Questionnaire
JMIR Mhealth Uhealth 2025;13:e56466
URL: https://mhealth.jmir.org/2025/1/e56466
doi:10.2196/56466

© Adrijana Svenšek, Lucija Gosak, Mateja Lorber, Gregor Stiglic, Nino Fijačko. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Artificial Intelligence–Based Mobile Phone Apps for Child Mental Health: Comprehensive Review and Content Analysis

Fan Yang[1], LGSW, PhD; Jianan Wei[2], MA; Xuejun Zhao[3], MA; Ruopeng An[4], PhD

[1]School of Social Work, University of Illinois Urbana Champaign, Urbana, IL, United States

[2]School of Public Administration, Dongbei University of Finance and Economics, Dalian, China

[3]School of Computer Science and Engineering, Macau University of Science and Technology, Macau, China

[4]Constance and Martin Silver Center on Data Science and Social Equity, Silver School of Social Work, New York University, 726 Broadway, New York, NY, United States

Corresponding Author:
Ruopeng An, PhD
Constance and Martin Silver Center on Data Science and Social Equity, Silver School of Social Work, New York University, 726 Broadway, New York, NY, United States

Abstract

Background: Mobile phone apps powered by artificial intelligence (AI) have emerged as powerful tools to address mental health challenges faced by children.

Objective: This study aimed to comprehensively review AI-driven apps for child mental health, focusing on their availability, quality, readability, characteristics, and functions.

Methods: This study systematically analyzed AI-based mobile apps for child mental health. Quality was evaluated using the Mobile Application Rating Scale, which assessed various dimensions of app quality, including subjective quality, engagement, functionality, aesthetics, and information. An automatic readability index calculator was implemented to assess readability by using the count of words, syllables, and sentences to generate a score indicative of the reading difficulty level. Content analysis was conducted to examine the apps' availability, characteristics, and functionality.

Results: Out of 369 apps initially identified, 27 met the eligibility criteria for inclusion. The quality of the apps was assessed using Mobile Application Rating Scale, with an average score of 3.45 out of 5 (SD 0.5), indicating a need for quality improvement. The readability analysis revealed suboptimal scores, with an average grade level of 6.62 (SD 2.2) for in-app content and 9.93 (SD 2.6) for app store descriptions. These results, combined with a monotonous user interface, suggest that many apps lack a child-friendly design, potentially hindering their usability and engagement for young users. Content analysis categorized the apps into 3 functional groups—chatbot-based apps (15 apps), journal logging apps (9 apps), and psychotherapeutic treatment apps (3 apps). While 20 out of 27 apps (74%) used clinically validated technologies, rigorous clinical tests of the apps were often missing, with only 2 apps undergoing clinical trials. Of the 27 apps analyzed, only 7 (26%) were free to use, while the majority, 20 apps, required a subscription or one-time payment. Among the paid apps, the average cost was US $20.16 per month, which may pose a financial barrier and limit accessibility for some users, particularly those from lower-income households.

Conclusions: AI-based mental health apps hold significant potential to address the unique challenges of child mental health but face critical limitations in design, accessibility, and validation. To fully realize their benefits, future research and development should focus on integrating child-centric design principles, ensuring affordability, and prioritizing rigorous clinical testing. These efforts are essential to harness the power of AI technologies in creating equitable, effective, and engaging solutions for improving child mental health outcomes.

(JMIR Mhealth Uhealth 2025;13:e58597) doi:10.2196/58597

KEYWORDS

mental health; children; artificial intelligence; mobile health; AI-driven mobile applications; mobile phone

Introduction

Mental health, as defined by the World Health Organization, is a state of well-being where individuals effectively manage stress, recognize their abilities, work productively, and contribute to their community. It includes conditions such as mental disorders, psychosocial disabilities, and states involving significant distress or functional impairment [1]. Concerns regarding mental health, particularly among children—defined as individuals younger than 18 years of age [2]—have garnered increasing attention in recent years. Around the world, 1 in 7 children experiences mental health conditions, yet many more remain largely unrecognized and untreated [1]. Children with mental health

XSL•FO
RenderX

challenges face various risks, including social exclusion, discrimination, stigma, educational difficulties, risk-taking behaviors, physical health symptoms, and human rights violations [3]. To safeguard their lifelong health and well-being, it is crucial to shield children from adversity, promote psychological wellness, and ensure their accessibility to mental health care [4,5].

Many children grapple with mental health issues, and an overwhelming fraction encounter barriers to accessing effective, affordable, and high-quality mental health services [6]. Several factors contribute to this limited access, including lack of health insurance, inadequate financial support, cultural constraints, and extended waiting times due to a shortage of professional services [4]. Beyond these microlevel challenges, public health crises, notably the COVID-19 pandemic, have exacerbated disparities between the escalating mental health needs of children and the availability of quality care. During the pandemic, national surveys across various countries revealed a heightened mental health strain on children, exacerbated by lockdowns that limited access to services and reduced help-seeking out of fear of contracting virus [1,7,8].

The swift advancement of technology, particularly in communication, has paved the way for novel mental health support mechanisms. The widespread availability of mobile phones and associated apps makes it possible for individuals to easily tap into essential resources, irrespective of where they are [9]. This democratization of access has been particularly impactful for mental health services, which are now available at the fingertips of those in need, often at a fraction of the traditional cost [10]. The integration of artificial intelligence (AI) in these apps has further revolutionized the landscape. As noted by Boucher et al [11], the infusion of AI technology not only enhances the precision of these platforms but also imbues them with a more empathetic and human touch. This allows for a tailored intervention, optimizing the support offered based on an objective analysis of user data.

Growing up in the digital era, children are inherently familiar with digital devices such as smartphones and tablets [12,13]. These tools offer an intuitive and user-friendly interface, eliminating complexities or abstract concepts and reducing potential barriers for younger users [13]. Previous research highlights a strong preference among children and adolescents for AI-integrated tools [13]. Studies suggest that AI-informed mobile health apps hold significant potential in addressing mental health issues in children [14]. These apps are generally well-received, with children showing positive attitudes toward their integration. Experts view such tools as complementary to traditional in-person interventions, offering low-threshold access via smartphones and extending their reach to children in high-risk situations or underserved communities [14]. This trend underscores the readiness of these age groups to embrace and engage with emerging technologies.

While AI-powered mental health apps are increasingly prevalent in the market, concerns have been raised about their effectiveness and quality [15,16]. These concerns are heightened by the rapid pace of AI advancement and evolving mental health care guidelines. It is important to note that most of the current scientific literature concentrates on adult mental health, often overlooking the unique needs of children. Recent comprehensive reviews by Li et al [17] and Hua et al [18] have made significant contributions to the field by providing valuable insights into AI-based conversational agents and large language models, respectively. Our study builds upon this foundation by encompassing a diverse array of AI-driven mobile phone apps, both those exclusively designed for children and those that include children as part of their broader target audience. This includes not only conversational agents, but also AI techniques used for emotion and journal tracking, as well as psychotherapeutic treatment within mental health apps. Furthermore, while these previous works have addressed AI applications in a general population context, our research uniquely targets the mental health of children, a demographic with specific needs and challenges that are often underrepresented in the literature. By systematically reviewing AI-based mental health apps accessible to children, our study aimed to highlight the current gaps and challenges in the development of child-specific resources within this emerging field. We evaluated 5 core areas—availability, quality, readability, characteristics, and functionality. Our insights aim to lay the groundwork for future research and guide the creation and development of AI-driven mental health apps that better cater to the needs of children, ultimately fostering the long-term psychological well-being of our future generations.

Methods

This study aimed to systematically review AI-driven mobile apps designed for children's mental health. The review process involved multiple steps, including a thorough literature search, screening, and evaluating selected apps.

App Search and Screening Strategy

An extensive search was conducted in the 2 most widely used marketplaces for mobile phone apps: Android's Google Play Store and Apple's iOS App Store. Search terms included "AI mental health apps for child/children" and "AI mental health apps for teenagers." These apps were subsequently screened based on the following criteria: (1) available in English; (2) can be downloaded for current use from Google Play or the Apple App Store; (3) the primary function is improving mental health; (4) the app must explicitly state that it is AI-based, with the AI component being central to its functionality in addressing mental health issues; and (5) the minimum age of the user shall be less than 18 years old.

The minimum age for app users in app stores is determined by a combination of factors, including developer-provided information, content ratings, and compliance with legal regulations. Developers are required to disclose the intended audience and the highest age-rated content available in their apps, which is then aligned with standardized rating systems like those developed by the International Age Rating Coalition. These systems ensure that apps are categorized appropriately based on their content, features, and potential risks. Legal frameworks, such as the Children's Online Privacy Protection Act in the United States and the General Data Protection Regulation in the European Union, further influence these age

XSL•FO
RenderX

ratings by setting strict guidelines for data collection from children, thereby shaping the minimum age requirements [19]. Using the app stores' minimum age ratings as a criterion for this study ensured that the reviewed apps include both those explicitly designed for children and those that, while not exclusively targeting children, include children as part of their broader user population. This dual inclusion provides a comprehensive analysis of AI-based mental health support tools available to children aged less than 18 years.

Apps that did not meet these criteria were filtered out. Furthermore, 2 research investigators (JW and XZ) independently conducted this screening process and reached a consensus on the final selection.

App Evaluation

Two investigators (JW and XZ) independently evaluated the apps. The Mobile Application Rating Scale (MARS) was primarily used to assess the quality of the apps, while a readability calculator was used to evaluate the readability. Content analysis was conducted to determine the descriptive characteristics and functionality. Descriptive characteristics codes included (1) price (free vs for-purchase); (2) confidentiality, concerning whether the app communicated and managed policies related to user privacy and data security; and (3) AI techniques adopted, referring to specific AI techniques used by the apps. Mobile app functionality components were also evaluated, including (1) clinically-proven psychotherapeutic treatment, detailing whether the app's treatments or the app itself have been clinically validated; (2) target mental health problems about specific mental health issues the app was designed to address; and (3) functions. Any disagreements about an app were resolved through discussion until a consensus was reached.

Readability of Mobile Apps

The readability of the mobile apps was assessed by 2 independent researchers (JW and XZ) using an automatic readability index calculator. Considering the prevalence of children's apps being downloaded by parents but primarily used by the children themselves [20], our analysis strategically separated the readability assessment into 2 distinct parts—examining the descriptions listed in the app store and the in-app text. For the in-app text, all visible text on the app pages during typical use was included in the analysis. If a page contained minimal text insufficient for readability evaluation, further interaction with the app was conducted to generate additional text output, such as engaging with chat features. Unrelated content, such as advertisements, was excluded to focus on text that directly impacts user interaction and comprehension. The Flesch-Kincaid Grade Level, an integral component within the readability calculator, was selected as our standard for readability, given its widespread recognition within the field [21-23]. The readability score was calculated using the formula below. The index corresponds to the US elementary school grade levels associated with the evaluated text, offering a universally comprehensible measure of text readability. A higher value on the readability index denotes a higher grade level, implying that more advanced reading proficiency is required.

Reading ease = 206.835 − (1.015 × number of words/number of sentences) − (84.6 × number of syllables/number of words)

MARS Assessment

The quality of the apps was evaluated using MARS, a widely recognized standardized quality measurement tool for health-related mobile apps [23-26]. MARS was characterized by 2 principal dimensions—the subjective quality scale, which imparted valuable insights into user experiences and formed a crucial part of the comprehensive evaluation index, and the objective scale, which provided a rigorous, structured evaluation of the apps' various technical and functional attributes. The objective scale comprised 19 items rated on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent). The items within the objective scale were further divided into 4 thematic categories: engagement, functionality, aesthetics, and information. Each category assessed diverse elements, from user interactivity and design appeal to the accuracy and credibility of the information provided. The final MARS score was the mean of the 4 subscores in these dimensions. This structured approach accomplished a comprehensive and detailed evaluation of the app's quality.

Content Analysis

In addition to evaluating the apps by the readability index calculator and the MARS scale, we also used content analysis to assess the included apps. Content analysis encompassed 3 main areas—availability, the descriptive characteristics of the apps, and functionality. Availability evaluated the presence of the apps across various app stores. Descriptive characteristics codes included (1) price (free vs for-purchase); (2) confidentiality, concerning whether the app communicated and managed policies related to user privacy and data security; (3) AI techniques adopted, referring to specific AI techniques used by the apps; and (4) children-specificity. Mobile app functionality components were also recorded, including (1) clinically-proven psychotherapeutic treatment, detailing whether the app's treatments or the app itself have been clinically validated; (2) target mental health problems about specific mental health issues the apps were designed to address; and (3) functions.

Two coders (JW and XZ) underwent training to attain reliability across several rounds of sample mental health apps designed for children. They were required to achieve an interrater reliability criterion of K=0.80 on 10 independently coded apps before commencing coding on the included apps in this study [27]. The 27 apps were downloaded from their respective app stores, and each was assessed and coded by these 2 coders. An overall interrater reliability of 0.94 was achieved, indicating strong reliability [28].

Results

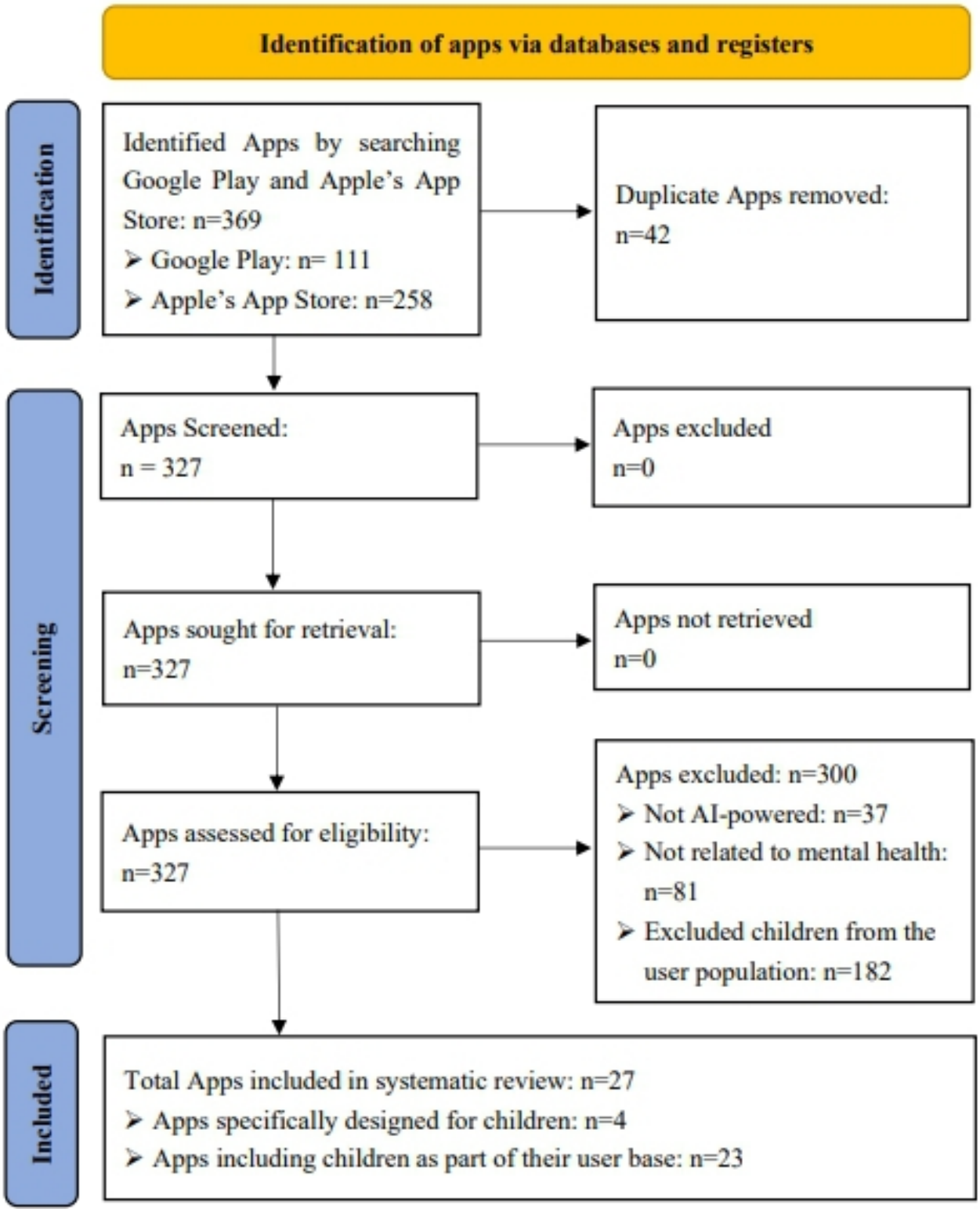
Overview

As of December 2024, this search yielded 369 apps. After removing 42 duplicated apps, 327 apps remained. The 327 apps were evaluated based on the inclusion criteria. This selection process resulted in 27 apps being included in this review study, with 4 apps specifically designed for children and 23 apps included children as part of their broad user base. The detailed

XSL•FO
RenderX

selection process was demonstrated in Figure 1; the PRISMA checklist has been provided (Checklist 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) of app search and screening strategy. AI: artificial intelligence.



Readability

Our study used the Flesch-Kincaid Grade Level index in a readability calculator to evaluate the readability of the app descriptions in the app store and in-app contents. The test revealed that the content readability within the apps spanned from grades 3 to 11, with an average readability of grade 6.62 (SD 2.2). This suggested a certain threshold of readability for younger children. The readability of the app description in the app store had a mean grade level of 9.93 (SD 2.6), which was notably higher than the content's readability. This could be due to the anticipation of parental involvement during the app download and installation processes, as parents are generally

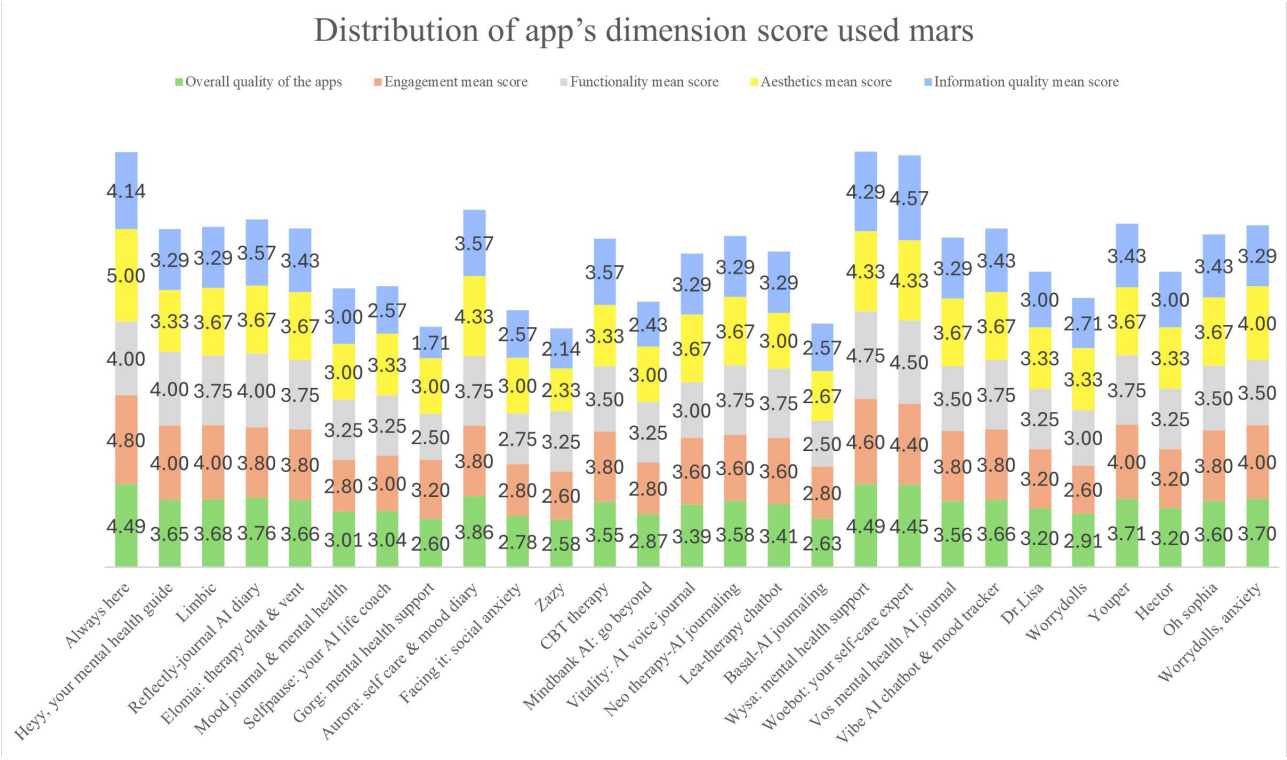
XSL•FO
RenderX

expected to have higher reading proficiency and are thus more likely to comprehend such introductions.

Quality Evaluation

Our findings revealed a suboptimal overall quality among existing apps, with MARS scores varying from 2.58 to 4.49, with an average of 3.45 (SD 0.5) on a scale of 1 to 5. Figure 2 employed a stacked bar graph to delineate the subscores and the composite MARS scores across these 4 dimensions. Of particular concern was the information quality, which had the lowest mean score of 3.19 (SD 0.6) and a minimum of 1.71. In contrast, engagement scored the highest, with a mean of 3.56 (SD 0.6) and a score range of 2.6 - 4.8.

Figure 2. Total MARS scores and subscores across 4 dimensions. MARS: Mobile Application Rating Scale.



The Engagement subscale emerged as the highest-scoring category in our study. This subscale was assessed through 5 perspectives—customization, interactivity, entertainment, interest, and target group. Delving deeper into the results, 44% (12/27) of the apps incorporated a baseline log-in assessment, emphasizing user customization, while 56% (15/27) did not include this feature. The apps displayed varied engagement strategies; 56% (15/27) primarily used dialogue interactions, and 33% (9/27) focused on diary log inputs. However, many apps needed more rigorous tracking and oversight mechanisms, suggesting a gap in ensuring long-term user interactivity. In addition, the apps exhibited diverse levels of entertainment and interest, with scores ranging from 1 to 5. These variances, especially in entertainment and interest, could affect the target group's appropriate level.

In total, 4 primary characteristics, performance, ease of use, navigation, and gestural design, were used to assess functionality. In terms of performance, 7 (26%) apps out of 27 demonstrated low accuracy or speed in their features and components, such as buttons and menus, adversely affecting the user's experience. Ease of use was apparent in 24 apps out of 27 (89%), as they facilitated effortless use for users with distinct menu labels and instructions. When evaluating navigation, these apps generally exhibited a coherent and uninterrupted transition between screens, with essential links in place. The scale score of this component ranged from 3 to 5, with an average score of 3.41 (SD 0.5), indicating that the apps' navigation is relatively well-integrated and logically structured. In the domain of gestural design, the apps consistently exhibited intuitive user interactions, such as tapping, swiping, pinching, and scrolling, across various components and screens. The evaluation scores for this aspect ranged from 2 to 5, with an average score of 3.48 (SD 0.6).

In evaluating the aesthetic appeal of the apps, particularly given their primary target audience of children, our findings revealed some limitations in their overall design. Based on the aesthetic subscale's criteria – which included layout, graphics, and visual appeal – only 11 (41%) of the reviewed apps showcased vibrant and engaging designs that effectively captivated the intended population. In stark contrast, 56% (15/27) of the apps exhibited monotonous designs that, while functional, lacked the visual allure necessary to maintain the interest of a younger audience. This aspect was particularly evident in the layout and graphics, where the arrangement and quality of buttons, icons, menus, and content were often not visually appealing. Especially in the visual appeal part, the score range was 2 - 5, and most ratings were on the lower end, indicating that the overall appeal of the app design was not strong. Furthermore, 15% (4/27) of these apps predominantly featured dark color schemes, which could potentially create negative psychological impressions and detract from their attractiveness to children. This mismatch between the apps' aesthetic elements and their intended users'

XSL•FO
RenderX

preferences highlighted a significant area for improvement in app design.

In assessing the apps from an information perspective, our findings demonstrated high accuracy in describing features, aligning well with the first characteristic of the information subscale. Nearly all apps accurately reflected their related features in app stores, avoiding exaggeration or misinformation. Regarding the goals, the majority of the apps were explicit in stating specific, measurable, and achievable objectives. Regarding the quality of information, 68% (18/27) of the apps, primarily developed by research laboratories and scientific institutions, provided correct, well-written content relevant to their goals. This underlined their credibility as sources of information. Regarding quantity, most apps offer comprehensive yet concise information within the scope of their intended use. In addition, the visual information in these apps generally presented concepts clearly and logically through various means like charts, graphs, images, and videos.

Content Analysis

Content analysis systematically evaluated app availability, descriptive characteristics, and functionality (Table 1). In terms of app availability, all apps were sourced from the Apple App Store and Google Play Store. There were 22 apps available in both app stores. Furthermore, 4 apps were exclusively available in the Apple App Store, and 1 was solely offered in Google Play.

Table . Content analysis for availability, descriptive characteristics, and functionality.

Analysis criteria	Subcategory	Value, n (%)
Availability		
Availability	Google Play	1 (4)
Availability	Apple’s App Store	4 (15)
Availability	Both Google and Apple	22 (82)
Descriptive characteristics		
Cost	Free	7 (26)
Cost	Purchased	20 (74)
Confidentiality	Yes	27 (100)
Confidentiality	No	0 (0)
AI techniques	Natural language processing	20 (74)
AI techniques	AI-based recommendation system	2 (7)
AI techniques	Automatic speech recognition	1 (4)
AI techniques	Data analysis and tracking	7 (26)
Children specificity	Yes	4 (15)
Children specificity	No	23 (85)
Functionality		
Clinical-proven	Yes	25 (93)
Clinical-proven	No	2 (7)
Psychotherapeutic treatment	None	7 (26)
Psychotherapeutic treatment	Cognitive behavioral therapy	19 (70)
Psychotherapeutic treatment	Dialectical behavior therapy	4 (15)
Psychotherapeutic treatment	Acceptance and commitment therapy	1 (4)
Psychotherapeutic treatment	Exposure therapy	1 (4)
Target problems	General	21 (78)
Target problems	Anxiety, depression, and stress	6 (22)
Functions	Dialogue interaction	15 (56)
Functions	Emotion and journal tracking	9 (33)
Functions	Psychotherapeutic treatment	3 (11)

We have identified critical, descriptive characteristics of these apps in terms of cost, confidentiality, AI techniques, and children’s specificity. There were 20 apps requiring payment to access additional content, with prices ranging from US $2.99 to US $120 per month (mean US $20.16), while only 7 apps offer basic functionality for free. All apps ensured privacy

XSL•FO
RenderX

protection by including privacy policies in the App Store and Google Play. AI technologies include natural language processing (NLP), AI recommendation systems, automatic speech recognition, and data analysis and tracking. NLP was used in 74% (20/27) of the apps, mainly for chatbot interactions, while automatic speech recognition was adopted in 1 voice input app. Furthermore, 9 emotion and journal tracking apps used data analysis to monitor and track users' mental health. Although all the apps have age limits that make them accessible to children, most (23/27, 85%) were not explicitly designed for a young audience, often lacking content and interfaces tailored to be child-friendly. Only a small proportion (4/27, 15% apps) were explicitly tailored for children, featuring age-appropriate content and user interfaces.

Functionality components were evaluated from the perspectives of clinically proven psychotherapeutic treatment, targeting mental health problems, functions, and challenges. Despite the utilization of clinically validated psychotherapeutic treatments like cognitive behavioral therapy and dialectical behavior therapy in 74% (20/27) of the apps, the overall effectiveness of the apps largely remained untested, with 2 apps, Woebot developed by Woebot Health and Youper developed by Youper Inc, having undergone clinical trials. Woebot underwent a clinical trial involving 36,070 participants, which revealed that users of the app experienced a significant reduction in depression and anxiety symptoms after 3 - 5 days of use [29]. Similarly, Youper has shown clinical evidence of benefit in a study involving its paying users. This study found that Youper was effective in reducing symptoms of anxiety and depression, with significant improvements observed [30]. Although this provides promising support for both Woebot and Youper, further rigorous testing, particularly through additional randomized controlled trials, is needed to establish their clinical effectiveness more definitively. In terms of targeted mental health problems, 22 apps focused on general mental health problems, while 5 apps were explicitly designed to treat anxiety and depression. Included apps can be divided into 3 major categories based on functions—dialogue interaction, journal log tracking, and psychotherapeutic interventions, which accounted for 56% (15/27), 33% (9/27), and 11% (3/27), respectively.

The utilization of these apps revealed several challenges, which can be categorized into 3 primary areas. Initially, a significant concern was the reliance on static large language models, which do not update their parameters after deployment and therefore may not align with rapid advancements in the field. This led to 12 apps needing help with complex or nuanced user inputs, resulting in ineffective or delayed processing. The responses of 7 apps tended to be repetitive and needed help to deliver meaningful conversation after multiple trials of conversation or when topics extended beyond standardized themes. Second, 9 apps exhibited a critical deficiency in user stickiness and feedback mechanisms. User stickiness was evaluated by examining factors including the app's ability to re-engage users after periods of inactivity and the presence of features that incentivize continuous interaction, such as personalized content, reminders, and progress tracking. The lack of robust tracking and feedback systems may precipitate user attrition, diminishing the apps' potential advantages. The third and equally significant challenge was observed in 4 apps, particularly concerning their approach to providing clear guidance. Furthermore, 1 app showed inconsistent dialogue transitions, where choices made by users were not appropriately reflected in the following responses, leading to confusion.

In addition, the game elements within one app needed explicit instructions, creating usability challenges for young users. Explanations or instructions were absent from the results displayed on the card surface, making it difficult for children to comprehend. Collectively, these challenges resulted in a less intuitive and supportive user experience, highlighting the need for more dynamic and user-centric design in these apps.

Discussion

Summary of Findings

In an era characterized by the rapid progression of AI, an increasing interest in leveraging technological advancements to address population-level mental health care needs has emerged [27]. In the context of growing societal attention to child mental health challenges [4], our research assessed the quality, readability, characteristics, and limitations of AI-driven mental health apps oriented to children. We identified 3 main functional categories—chatbots (15 apps), journal logging (9 apps), and psychotherapeutic treatment (3 apps), with 74% (20/27) employing NLP technology. The average MARS score of 3.45 out of 5 (SD 0.5) indicated a need for quality improvement. Readability was not sufficiently child-friendly. While most apps (25/27, 93%) used clinically validated technologies, rigorous clinical testing was lacking, with only 2 apps having undergone clinical trials. In addition, concerns about response accuracy and the high costs associated with these apps—74% (20/27) required payment, averaging US $20.16 per month—could limit their accessibility and effectiveness for children.

Underrepresentation of Child-Specific Design

Empirical research has underscored a positive inclination among children and adolescents toward AI and its integration with interventions, suggesting a solid willingness within these age groups to explore these technologies [13]. However, the evident lack of child-centric design may undermine these potential advantages, as identified in our investigation.

In the context of readability, these apps primarily required advanced reading skills. Previous studies have shown that parents are often the gatekeepers of the app download and installation process. At the same time, children are the users [31]; we evaluated the readability of app descriptions in the app store and content separately. Our findings correspond with the broader readability challenges of mobile apps designed for vulnerable populations [23]. The readability of user interface content—consistently 2 - 4 grade levels lower than the app descriptions—still presented a considerable readability threshold for younger users, raising the accessibility barrier for app usage and potentially hindering independent use by children [4].

The content incorporated in these apps was not explicitly tailored to children. For example, our findings revealed that certain apps primarily focused on alleviating occupational stress and

XSL•FO
RenderX

addressing professional interpersonal conflict, thereby falling short of addressing the unique mental health requirements of children. This mismatch can be attributed to 2 key factors. First, due to the limited availability of apps exclusively designed for children's mental health, our study included apps that, while not specifically targeted at children, were accessible to them within the broader age range intended by the developers. As a result, some apps that children could access contained content not relevant to their specific needs. Second, this mismatch also reflects the developers' limited understanding of children's mental health needs, leading to the inclusion of content that does not align with children's requirements. App developers were predominantly technology firms and research laboratories without professional child mental health experts to inform app design. Hence, based on the recommendations made by previous studies [13], we advocated for a multidisciplinary team, comprised of AI and app technicians alongside child mental health professionals, to guide the future development of AI-driven apps designed to address children's mental health concerns. This approach should involve collaboration between AI and app developers, child mental health professionals, and endorsement by peak mental health organizations. The involvement of these key stakeholders ensures that the apps are clinically effective, relevant to the target population, and endorsed by authoritative bodies, thereby increasing their credibility and acceptance in clinical practice. Furthermore, the cocreation of these apps with children and clinicians can enhance their usability and ensure that they meet the specific needs of the target population.

The intervention delivery method within these apps also demonstrated a need for more child-friendly techniques. Diary-logging apps and chatbot dialogues were the dominating intervention strategies employed by these apps. However, diary logs and chatbot interactions needed substantial text input and the ability for users to clearly articulate their mental health concerns, an approach that has been evidenced as being more challenging for child users [4,27]. The expressive capabilities of children may need to be sufficiently developed to leverage these apps, thereby impacting their effectiveness. Incorporating child-friendly tools such as art therapy [4], an area made achievable by recent developments in text-to-image models like stable diffusion models and DALL·E, could offer potential solutions.

In addition to readability, app content, and intervention delivery method, the existing apps' visual aesthetics and user interface design also demonstrated a need for child-centric considerations. The color palette of many current apps could have been more varied, simplistic, and potentially unengaging, lacking vibrant imagery that could capture children's attention [32]. On the contrary, some apps adopted a dark aesthetic, which could appear somber and oppressive, potentially intensifying feelings of discomfort or even burden, particularly in users already experiencing low mood or mental health issues [33]. Hence, it was advisable for app designs to integrate richer, brighter colors and more engaging images. Including popular animation characters could enhance children's interest and engagement [4].

Barriers to Optimal Mental Health Outcome Concerns

To effectively embed the findings of this study into clinical practice, it is crucial to integrate AI-driven mental health apps into existing care frameworks. This integration can enhance the accessibility and quality of mental health care for children, particularly in underserved populations. However, for these tools to be successfully adopted in clinical settings, their effectiveness in producing measurable behavior change and improving mental health outcomes must first be validated. While our review highlights the potential of AI-driven mobile phone apps in addressing child mental health, it is important to note that the effectiveness of these interventions in producing measurable behavior change remains largely untested. Therefore, future research should focus on evaluating the impact of these apps on users' mental health outcomes and behavioral modifications. Establishing this evidence is essential to ensure that these apps can be confidently and effectively incorporated into routine care, ultimately delivering the intended benefits to the target population.

The effectiveness of an app was fundamentally linked to the therapeutic technology it employed [27]. Most reviewed mobile apps used clinically validated technologies, such as cognitive behavioral therapy and dialectical behavior therapy [34]. The effectiveness of an app was fundamentally linked to the diagnostic and therapeutic technology it employed [31]. However, these apps' effectiveness in enhancing child mental health largely remained clinically unproven, a conclusion drawn from our review of 27 such apps. Even though these apps primarily serve adjunctive support tools and cannot fully replace professional psychological diagnosis and treatment [11], we strongly recommend that these apps undergo rigorous clinical validation to ensure their effectiveness, underpinned by solid evidence.

User engagement and stickiness are critical factors that can influence the effectiveness of mental health interventions. The journey toward mental health improvement is cumulative, requiring consistent commitment over time [11]. This underscores the importance of regular app usage and the fostering of robust user engagement for optimal intervention functionality [35]. However, a significant shortcoming in the design of most current apps is the absence of effective user tracking mechanisms. User tracking mechanisms were examined by examining the presence of features such as usage analytics, personalized reminders, and progress tracking, which are essential for promoting regular interaction. These features are known to encourage regular use and engagement, which are important for the success of any intervention [36-38]. Several apps have implemented systems requiring notification activation to promote more regular user interaction [11], contributing to a more effective mental health intervention experience.

Furthermore, the interplay between AI-powered mental health apps and user privacy presented unique strengths and challenges that can indirectly affect mental health outcomes by altering the users' decisions on whether to engage with these apps. The stigma attached to mental health issues in various cultural and societal contexts heightens privacy concerns [31], especially among underrepresented populations with limited access to

XSL•FO
RenderX

formal mental health care [39]. AI-powered mobile phone apps can positively enhance the sense of privacy credited to no human involvement [40]. However, chatbot dialogues and journal logging may involve significant exposure of personal and traumatic experiences, raising critical privacy concerns about handling such sensitive data [10]. This potential exposure may deter many privacy-sensitive individuals from using these apps. To alleviate these concerns, many apps have adopted privacy safeguards such as presenting privacy policies during initial installation, anonymizing data, and postsession deletion of user histories. Transparency in implementing and communicating these privacy-related strategies remained crucial.

The association between low socioeconomic status and adverse child mental health outcomes has been widely described [8,41-43]. On the flip side, cost often presents a significant obstacle in accessing mental health services, with affordability issues preventing nearly half of the individuals requiring mental health treatment in the United States from receiving it [44]. In this context, the presumed high accessibility of AI-powered mental health apps, both in academia and public perception, was considered a significant advantage, fostering wider user reach and coverage [32]. However, most apps we evaluated incorporate fee-based features, raising the barrier to use. This increased cost compromises the perceived accessibility of the AI and mobile app approach, potentially curtailing its effectiveness.

Challenges and Considerations in AI Techniques for Child-Oriented Mental Health Apps

The current landscape of AI in child-oriented mental health apps reveals a significant reliance on NLP to assess and intervene in users' mental health states. While NLP is a powerful tool for processing and analyzing language, this predominant focus on text-based interactions may not fully address the needs of children, who often struggle with limited reading, verbal, and writing skills, particularly when expressing complex mental health concerns. The lack of diverse interaction methods could hinder children's ability to engage with and benefit from these mental health interventions. In light of this, we suggest future research and apps incorporating child-friendly AI techniques, such as generative AI for image creation using natural language prompts, to complement NLP-based approaches. Image-based tools can help children express their emotions and thoughts more effectively, broadening the accessibility and impact of mental health apps for this demographic. By offering multiple modes of interaction, these apps could better accommodate the diverse needs and abilities of children, ultimately making them more effective tools for promoting mental health among younger users [3].

Consistent with previous research [10,11,40,45], our findings revealed that the quality of the AI techniques currently in use raised considerable concerns. The performance of NLP, in particular, demonstrated notable inconsistencies and disparities among different mobile phone apps. For instance, some apps needed help processing simple questions, often responding with repetitive and uninspiring answers after prolonged periods. This issue was especially evident when voice input—an interaction mode that naturally resonates with and is easily used by children [5]—was used. These apps' inability to handle such a fundamental form of interaction effectively raised significant questions about their usability and overall quality [46]. Given these limitations, users might be deterred from relying on these tools for consistent support, limiting their potential benefits.

In light of these quality issues, we suggest that the application of implementation science principles is vital for the practical deployment of AI-driven mental health apps. One approach is for developers to prioritize incorporating state-of-the-art language models. One of the key challenges in this context is preventing AI hallucinations—where AI generates inaccurate or misleading information. To mitigate this risk, we propose linking AI systems within these apps to established clinical guidelines using advanced information retrieval methods. This approach ensures that AI provides evidence-based recommendations, thereby reducing the likelihood of errors and ensuring that the advice given is both accurate and appropriate for children. Another approach involves the use of rule-based models, as seen in apps like Woebot, which prioritize safety by avoiding generative AI for user-facing text. Even with rule-based models, there is room for improvement by updating and refining the model parameters. This approach can enhance the user experience while maintaining the safety and reliability of the app. By integrating these strategies, AI techniques in AI-driven mental health apps for children can be better positioned to meet clinical standards and effectively support children's mental health needs.

Limitations

It is important to acknowledge the limitations of this study. The first limitation lies in the time-sensitive nature of the app search process. AI-based mental health apps are being developed and launched at an unprecedented rate, reflecting the rapid advancements in AI technology. To mitigate this limitation, we conducted 2 app searches in October 2023 and December 2023. However, this rapid pace of development means that some newly launched apps may not have been captured in our analysis. Future studies could address this limitation by using longitudinal designs with repeated app searches over extended periods to better capture the evolving landscape of AI-based mental health apps. The second limitation concerns the scope of the apps included. Currently, there are very few AI-based mental health apps specifically designed for children available in the market. As a result, we included all AI-based mental health mobile phone apps, including both those explicitly designed for children and those that, while not exclusively targeting children, still consider them part of their intended user base. As AI-based mental health interventions specifically designed for children continue to grow, future research should focus on systematically reviewing apps developed exclusively for children to provide a more targeted analysis. Third, we only included apps available in English, which may have excluded apps offered in other languages. This limitation highlights the need for future studies to explore apps in diverse languages to provide a more inclusive understanding of the global availability and accessibility of AI-based mental health apps for children.

XSL•FO
RenderX

Conclusion and Implications

AI-powered mobile apps were positioned at the frontier, addressing children's mental health issues, uniquely combining accessibility and acceptance among the younger demographics. Our study, however, emphasized that the future of these AI-based tools was intrinsically tied to their capacity for significant evolution and enhancement. This included enhanced child-friendly design, clinically impactful outcomes, and superior quality of AI techniques, all of which should be considered nonnegotiable facets of their development. At this juncture, our choices will undeniably carve out the path for AI-powered mental health support for children. It is our collective responsibility as a profession to ensure these digital solutions extend beyond technological sophistication, embodying empathy, effectiveness, and genuine child-centricity. By doing so, we can unlock the full potential of AI, creating a future where every child, regardless of their location or circumstances, has access to robust, reliable, and adequate mental health support.

Conflicts of Interest

None declared.

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 31 KB - mhealth_v13i1e58597_app1.docx]

References

1. World Health Organizaiton. Improving the mental and brain health of children and adolescents. URL: https://www.who.int/activities/improving-the-mental-and-brain-health-of-children-and-adolescents [accessed 2023-09-03]
2. Lansdown G, Vaghri Z. Article 1: definition of a child. In: Vaghri Z, Zermatten J, Lansdown G, Ruggiero R, editors. Monitoring State Compliance with the UN Convention on the Rights of the Child Children's Well-Being: Indicators and Research: Cham: Springer; 2022, Vol. 25. [doi: 10.1007/978-3-030-84647-3_1]
3. World Health Organizaiton. Mental health of adolescents. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2023-09-03]
4. Zhou X, Edirippulige S, Bai X, Bambling M. Are online mental health interventions for youth effective? A systematic review. J Telemed Telecare 2021 Dec;27(10):638-666. [doi: 10.1177/1357633X211047285] [Medline: 34726992]
5. Guo Y, Hopson LM, Yang F. Socio-ecological factors associated with adolescents' psychological well-being: a multilevel analysis. International Journal of School Social Work 2018;3(1). [doi: 10.4148/2161-4148.1032]
6. Patrick SW, Henkhaus LE, Zickafoose JS, et al. Well-being of parents and children during the COVID-19 pandemic: a national survey. Pediatrics 2020 Oct;146(4):e2020016824. [doi: 10.1542/peds.2020-016824] [Medline: 32709738]
7. Jones Harden B, Slopen N. Inequitable experiences and outcomes in young children: addressing racial and social-economic disparities in physical and mental health. Annu Rev Dev Psychol 2022 Dec 9;4(1):133-159 [FREE Full text] [doi: 10.1146/annurev-devpsych-121020-031515]
8. Gadermann AC, Thomson KC, Richardson CG, et al. Examining the impacts of the COVID-19 pandemic on family mental health in Canada: findings from a national cross-sectional study. BMJ Open 2021 Jan 12;11(1):e042871. [doi: 10.1136/bmjopen-2020-042871] [Medline: 33436472]
9. Welch V, Wy TJ, Ligezka A, et al. Use of mobile and wearable artificial intelligence in child and adolescent psychiatry: scoping review. J Med Internet Res 2022 Mar 14;24(3):e33560. [doi: 10.2196/33560] [Medline: 35285812]
10. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
11. Boucher EM, Harake NR, Ward HE, et al. Artificially intelligent chatbots in digital mental health interventions: a review. Expert Rev Med Devices 2021 Dec;18(sup1):37-49. [doi: 10.1080/17434440.2021.2013200] [Medline: 34872429]
12. Linder LA, Newman AR, Stegenga K, et al. Feasibility and acceptability of a game-based symptom-reporting app for children with cancer: perspectives of children and parents. Support Care Cancer 2021 Jan;29(1):301-310. [doi: 10.1007/s00520-020-05495-w] [Medline: 32358779]
13. Tan K, Chen XS. Keeping up with technology: socioemotional and equity challenges with children and schools. Child Sch 2023 Jun 22;45(3):127-130. [doi: 10.1093/cs/cdad014]
14. Götzl C, Hiller S, Rauschenberg C, et al. Artificial intelligence-informed mobile mental health apps for young people: a mixed-methods approach on users' and stakeholders' perspectives. Child Adolesc Psychiatry Ment Health 2022 Nov 17;16(1):86. [doi: 10.1186/s13034-022-00522-6] [Medline: 36397097]
15. Zidaru T, Morrow EM, Stockley R. Ensuring patient and public involvement in the transition to AI-assisted mental health care: a systematic scoping review and agenda for design justice. Health Expect 2021 Aug;24(4):1072-1124. [doi: 10.1111/hex.13299] [Medline: 34118185]

XSL•FO
RenderX

16. Hamdoun S, Monteleone R, Bookman T, Michael K. AI-based and digital mental health apps: balancing need and risk. IEEE Technol Soc Mag 2023;42(1):25-36. [doi: 10.1109/MTS.2023.3241309]
17. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digit Med 2023 Dec 19;6(1):236. [doi: 10.1038/s41746-023-00979-5] [Medline: 38114588]
18. Hua Y, Liu F, Yang K, et al. Large language models in mental health care: a systematic scoping review (preprint). JMIR Mental Health. Preprint posted online on 2024. [doi: 10.2196/preprints.64088]
19. App Store Connect Help. Manage app information: Set an app age rating. URL: https://developer.apple.com/help/app-store-connect/manage-app-information/set-an-app-age-rating/ [accessed 2024-12-18]
20. Pearson HA, Montazami A, Dubé AK. Why this app: can a video‐based intervention help parents identify quality educational apps? Brit J Educational Tech 2023 May;54(3):712-733 [FREE Full text] [doi: 10.1111/bjet.13284]
21. Fitzsimmons PR, Michael BD, Hulley JL, Scott GO. A readability assessment of online Parkinson's disease information. J R Coll Physicians Edinb 2010 Dec;40(4):292-296. [doi: 10.4997/JRCPE.2010.401] [Medline: 21132132]
22. Munsour EE, Awaisu A, Hassali MAA, Darwish S, Abdoun E. Readability and comprehensibility of patient information leaflets for antidiabetic medications in Qatar. J Pharm Technol 2017 Aug;33(4):128-136. [doi: 10.1177/8755122517706978] [Medline: 34860991]
23. Guo Y, Yang F, Hu F, Li W, Ruggiano N, Lee HY. Existing mobile phone apps for self-care management of people with Alzheimer disease and related dementias: systematic analysis. JMIR Aging 2020 Jan 24;3(1):e15290. [doi: 10.2196/15290] [Medline: 32012045]
24. Hensley CP, Witte MM, Cai J, et al. Assessment of mobile health applications for management of knee and/or hip osteoarthritis using the mobile application rating scale. J Clin Rheumatol 2023 Aug 1;29(5):245-253. [doi: 10.1097/RHU.0000000000001896] [Medline: 36256541]
25. Musgrave LM, Kizirian NV, Homer CSE, Gordon A. Mobile phone apps in Australia for improving pregnancy outcomes: systematic search on app stores. JMIR Mhealth Uhealth 2020 Nov 16;8(11):e22340. [doi: 10.2196/22340] [Medline: 33196454]
26. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
27. Bry LJ, Chou T, Miguel E, Comer JS. Consumer smartphone apps marketed for child and adolescent anxiety: a systematic review and content analysis. Behav Ther 2018 Mar;49(2):249-261. [doi: 10.1016/j.beth.2017.07.008] [Medline: 29530263]
28. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76(5):378-382. [doi: 10.1037/h0031619]
29. Darcy A, Daniels J, Salinger D, Wicks P, Robinson A. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. JMIR Form Res 2021 May 11;5(5):e27868. [doi: 10.2196/27868] [Medline: 33973854]
30. Mehta A, Niles AN, Vargas JH, Marafon T, Couto DD, Gross JJ. Acceptability and effectiveness of artificial intelligence therapy for anxiety and depression (Youper): longitudinal observational study. J Med Internet Res 2021 Jun 22;23(6):e26771. [doi: 10.2196/26771] [Medline: 34155984]
31. Barua PD, Vicnesh J, Gururajan R, et al. Artificial intelligence enabled personalised assistive tools to enhance education of children with neurodevelopmental disorders-a review. Int J Environ Res Public Health 2022 Jan 21;19(3):1192. [doi: 10.3390/ijerph19031192] [Medline: 35162220]
32. Lipschitz JM, Connolly SL, Miller CJ, Hogan TP, Simon SR, Burdick KE. Patient interest in mental health mobile app interventions: demographic and symptom-level differences. J Affect Disord 2020 Feb 15;263:216-220. [doi: 10.1016/j.jad.2019.11.083] [Medline: 31818779]
33. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176. [doi: 10.2196/jmir.7332] [Medline: 28546138]
34. Denecke K, Schmid N, Nüssli S. Implementation of cognitive behavioral therapy in e-mental health apps: literature review. J Med Internet Res 2022 Mar 10;24(3):e27791. [doi: 10.2196/27791] [Medline: 35266875]
35. Schachner T, Keller R, V Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep 14;22(9):e20701. [doi: 10.2196/20701] [Medline: 32924957]
36. Kankanhalli A, Xia Q, Zhao X, National University of Singapore. Understanding personalization for health behavior change applications: a review and future directions. THCI 2021;13:316-349. [doi: 10.17705/1thci.00152]
37. Alshurafa N, Jain J, Alharbi R, Iakovlev G, Spring B, Pfammatter A. Is more always better?: Discovering incentivized mHealth intervention engagement related to health behavior trends. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec;2(4):1-26. [doi: 10.1145/3287031] [Medline: 32318650]
38. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and e-Health: a scoping review. Telemed J E Health 2022 Jun;28(6):768-780. [doi: 10.1089/tmj.2021.0176] [Medline: 34637651]

XSL•FO
RenderX

39. Timmons AC, Duong JB, Simo Fiallo N, et al. A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. Perspect Psychol Sci 2023 Sep;18(5):1062-1096. [doi: 10.1177/17456916221134490] [Medline: 36490369]
40. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019 May 9;21(5):e13216. [doi: 10.2196/13216] [Medline: 31094356]
41. Ravens-Sieberer U, Kaman A, Otto C, et al. Mental health and psychological burden of children and adolescents during the first wave of the COVID-19 pandemic-results of the COPSY study. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Dec;64(12):1512-1521. [doi: 10.1007/s00103-021-03291-3] [Medline: 33649901]
42. Lampert T, Hoebel J, Kuntz B, et al. Gesundheitliche ungleichheiten bei kindern und jugendlichen in deutschland–zeitliche entwicklung und trends der kiggs-studie. 2019. URL: https://www.rki.de/DE/Aktuelles/Publikationen/Journal-of-Health-Monitoring/GBEDownloadsJ/Focus/JoHM_01_2019_Gesundheitliche_Ungleichheiten.pdf?__blob=publicationFile&v=2 [accessed 2025-05-14]
43. Reiss F, Meyrose AK, Otto C, Lampert T, Klasen F, Ravens-Sieberer U. Socioeconomic status, stressful life situations and mental health problems in children and adolescents: results of the German BELLA cohort-study. PLoS ONE 2019;14(3):e0213700. [doi: 10.1371/journal.pone.0213700] [Medline: 30865713]
44. Skorburg JA, Yam J. Is there an app for that?: Ethical issues in the digital mental health response to COVID-19. AJOB Neurosci 2022;13(3):177-190. [doi: 10.1080/21507740.2021.1918284] [Medline: 33989127]
45. Velez JA, Keene JR, Corwin M, Elko S, Potter RF. A visual interactive narrative intervention (VINI) for aphasia education: can digital applications administer augmented input to educate stroke survivors with aphasia? Patient Educ Couns 2021 Oct;104(10):2536-2543. [doi: 10.1016/j.pec.2021.03.018] [Medline: 33810913]
46. Pham KT, Nabizadeh A, Selek S. Artificial intelligence and chatbots in psychiatry. Psychiatr Q 2022 Mar;93(1):249-253. [doi: 10.1007/s11126-022-09973-8] [Medline: 35212940]

Abbreviations

AI: artificial intelligence
MARS: Mobile Application Rating Scale

Edited by TDA Cardoso; submitted 28.03.24; peer-reviewed by B Thies, L Musgrave; revised version received 11.03.25; accepted 02.04.25; published 06.06.25.

Please cite as:
Yang F, Wei J, Zhao X, An R
Artificial Intelligence–Based Mobile Phone Apps for Child Mental Health: Comprehensive Review and Content Analysis
JMIR Mhealth Uhealth 2025;13:e58597
URL: https://mhealth.jmir.org/2025/1/e58597
doi:10.2196/58597

© Fan Yang, Jianan Wei, Xuejun Zhao, Ruopeng An. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Users' Needs for Mental Health Apps: Quality Evaluation Using the User Version of the Mobile Application Rating Scale

Siyeon Ko[1], MPH; Hyekyung Woo[1,2], PhD

[1]Department of Health Administration, College of Nursing & Health, Kongju National University, 56 Gongjudaehak-ro, Gongju-Si, Chungcheongnam-do, Republic of Korea

[2]Institute of Health and Environment, Kongju National University, Gongju-si, Chungcheongnam-do, Republic of Korea

Corresponding Author:
Hyekyung Woo, PhD
Department of Health Administration, College of Nursing & Health, Kongju National University, 56 Gongjudaehak-ro, Gongju-Si, Chungcheongnam-do, Republic of Korea

Abstract

Background: Mental health is an essential element of life. However, existing mental health services face challenges in utilization due to issues such as societal prejudices and a shortage of counselors. Mobile health is gaining attention as an alternative approach to improving mental health by addressing the shortcomings of traditional services. As a result, various mental health apps are being developed, but there is a lack of evaluation research on whether these apps meet users' needs.

Objective: This study aims to evaluate the content and quality of mental health apps from the user's perspective and identify the content features that influence evaluation scores. We also aim to guide future updates and improvements in mental health apps to deliver high-quality solutions to users.

Methods: We searched the Google Play Store and iOS App Store using Korean keywords "mental health," "mental health care," "depression," and "stress." Apps meeting the following criteria were selected for the study: relevance to the topic, written in Korean, more than 700 reviews (Android) or more than 200 reviews (iOS), updated within the past 365 days, available for free, nonduplicate, and currently operational. After identifying and defining the primary contents of the apps, 7 users evaluated their quality using the user version of the Mobile Application Rating Scale (uMARS). Correlation analysis was performed to examine the relationships among app content, uMARS scores, star ratings, and the number of reviews. Multiple regression analysis was conducted to identify the factors influencing uMARS scores and each evaluation item.

Results: The analysis included a total of 41 mental health apps. Content analysis revealed that reminders (n=29, 71%), recording and statistics features (n=29, 71%), and diaries (n=24, 59%) were the most common app components. The top-rated apps, as determined by uMARS evaluations, consistently provided information about counselors and counseling agencies, and included counseling services. uMARS scores were significantly correlated with the presence of health care provider information (r=0.53; P<.001) and counseling/question and answer services (r=0.55; P<.001). Multiple regression analysis indicated that providing more relevant information was associated with higher uMARS scores (β=.361; P=.02).

Conclusions: The quality of mental health apps was evaluated from the user's perspective using a validated scale. To deliver a high-quality mental health app, it is essential to incorporate app technologies such as generative artificial intelligence during development and to continuously monitor app quality from the user's perspective.

(JMIR Mhealth Uhealth 2025;13:e64622) doi:10.2196/64622

KEYWORDS

app user perspective; user evaluation; mental health care; digital health; mobile app; mHealth; mental health; app; evaluation study; mobile health; quality; user; user perspective; Google Play Store; correlation analysis; regression analysis; technology; generative AI; app quality; smartphone

Introduction

Mental health is an essential aspect of life, representing a positive and peaceful state that enables individuals to effectively manage daily stress [1]. Mental health problems can greatly reduce the quality of life and are often recognized as major contributors to weakened immunity and an increased risk of disease. Consequently, there is an urgent need for ongoing mental health management [2,3]. Interest in mental health care has grown markedly due to the rise in mental health issues, such as psychological stress, depression, and anxiety, exacerbated by prolonged social distancing and self-isolation during the COVID-19 pandemic [4]. Notably, Korea has the highest suicide rate among member countries of the Organization for Economic

XSL•FO
RenderX

Cooperation and Development, underscoring the importance of mental health management [5].

Current mental health care and promotion services primarily focus on 1-on-1, face-to-face counseling. Consistent emotional exploration and management are key to recovering from mental health issues [6]. However, these services are often insufficient because of factors such as a shortage of counselors, inadequate facilities, financial barriers, and societal prejudices [7].

Recently, ChatGPT, a generative artificial intelligence (AI) model that generates contextually relevant responses based on user input, has gained significant popularity for its ability to reduce human intervention while enhancing efficiency. When asked about mental health care, it provides tailored solutions related to treatment, self-management, interpersonal relationships, and goal setting. This adaptive responsiveness is especially valuable for AI chatbots and other apps that require personalized services [8].

Within the field of digital health, interventions such as a "web-based tool for health promotion and stress management" and "online educational program focusing on depression and anxiety" have shown promise in addressing various mental health problems [9-11]. The proliferation of mobile health (mHealth) apps that incorporate digital health content offers an alternative approach to improving mental health, whether through face-to-face or remote interactions, while also addressing the limitations of traditional services.

mHealth apps for mental health promotion target a range of issues, including anxiety, depression, and stress. They provide information to support self-diagnosis, treatment options, and access to counseling services. The inclusion of features such as health data collection and personalized feedback helps foster user engagement and supports mental health improvement [12]. These apps have been well-received, as they compensate for the limitations of the existing services by offering a discreet, low-risk alternative that helps mitigate barriers to psychiatric counseling [13].

The growing interest in mHealth for mental health has led various institutions and developers to create apps featuring psychological support programs, mental health assessments, counseling, and management tools. However, limited awareness of these apps and the proliferation of similar apps with nearly identical functions pose challenges for users in selecting customized solutions [4,14].

Accordingly, research on the quality evaluation of mental health apps is actively underway in many countries. Although such evaluations exist, they often emphasize expert perspectives, highlighting the need for more user-centered research to assess satisfaction and usability [15-17]. In Korea, research targeting specific user groups and behaviors, such as hospital employee–focused case studies and app development for anger management, is ongoing, yet a broader understanding of whether these apps meet users' needs remains limited [18-20].

This study assessed the content and quality of mental health apps available in South Korea from the user perspective, focusing on apps from the Google Play Store and iOS App Store. We aim to guide future updates and improvements in mental health apps to deliver high-quality solutions to users.

Methods

App Selection

The selection process of the apps was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [21]. Between September 30 and October 5, 2022, we conducted searches on both the Google Play Store and iOS App Store using Korean keywords, including "mental health," "mental health care," "depression," and "stress." These keywords were chosen based on previous studies related to mental health and mental health apps. Following this initial search, we established specific criteria for app inclusion, considering prior research findings [22-24]. As free apps are generally more accessible to users and tend to be the most widely used, free downloads were included as a selection criterion [15]. The inclusion criteria for apps selected for the final analysis required that all of the following conditions (C1-C7) be satisfied:

- C1. Relevance to the topic
- C2. App language is Korean
- C3. App has received over 700 reviews on Android and over 200 reviews on iOS (calculated in proportion to each platform's market share in Korea as of Q2 2022 [25])
- C4. The app was updated within the last 365 days
- C5. The app is free to use (including those with in-app purchases)
- C6. The app is nonduplicate
- C7. The app is currently operational

The star ratings for each app were extracted from the respective app stores as of the search date. On both platforms, star ratings are based on a 5-point scale. The review counts represent the cumulative total number of user reviews recorded for each app up to the search date.

Analysis of App Contents

We downloaded all selected research apps and categorized their main contents in Multimedia Appendix 1, classifying them into categories, such as information provision, self-diagnosis, and monitoring functions. A frequency analysis was performed to determine the number of apps offering these functions.

Evaluation of App Quality

For each of the final selected apps, we assessed quality using the Korean-translated version of the user version of the Mobile Application Rating Scale (uMARS) [26]. uMARS is a 5-point scale developed from the original Mobile Application Rating Scale to enable users to evaluate mHealth apps [27,28]. The scale demonstrates excellent internal consistency and has been widely validated, confirming its reliability and validity. The scale has been successfully applied to assess various mHealth apps, including those related to weight management, rheumatic conditions, menstrual cycles, and cancer risk calculators [22,29-31]. The overall uMARS score was calculated by averaging the scores across all 5 categories: engagement, aesthetics, functionality, information, and subjective quality.

XSL•FO
RenderX

Between October 8 and October 14, 2022, a total of 7 users of real mental health apps participated in the uMARS evaluation. The number of evaluators was determined based on user-centered evaluation standards, which suggest that 5 or more users are typically sufficient to identify most usability issues. This sample size ensures diversity in user experiences and allows for reliable results [32,33]. Apps were randomly assigned to evaluators to minimize potential bias due to subjectivity. Each evaluator was instructed to use the assigned app for at least 10 minutes daily, and evaluations were conducted in a blind test format.

Statistical Analysis

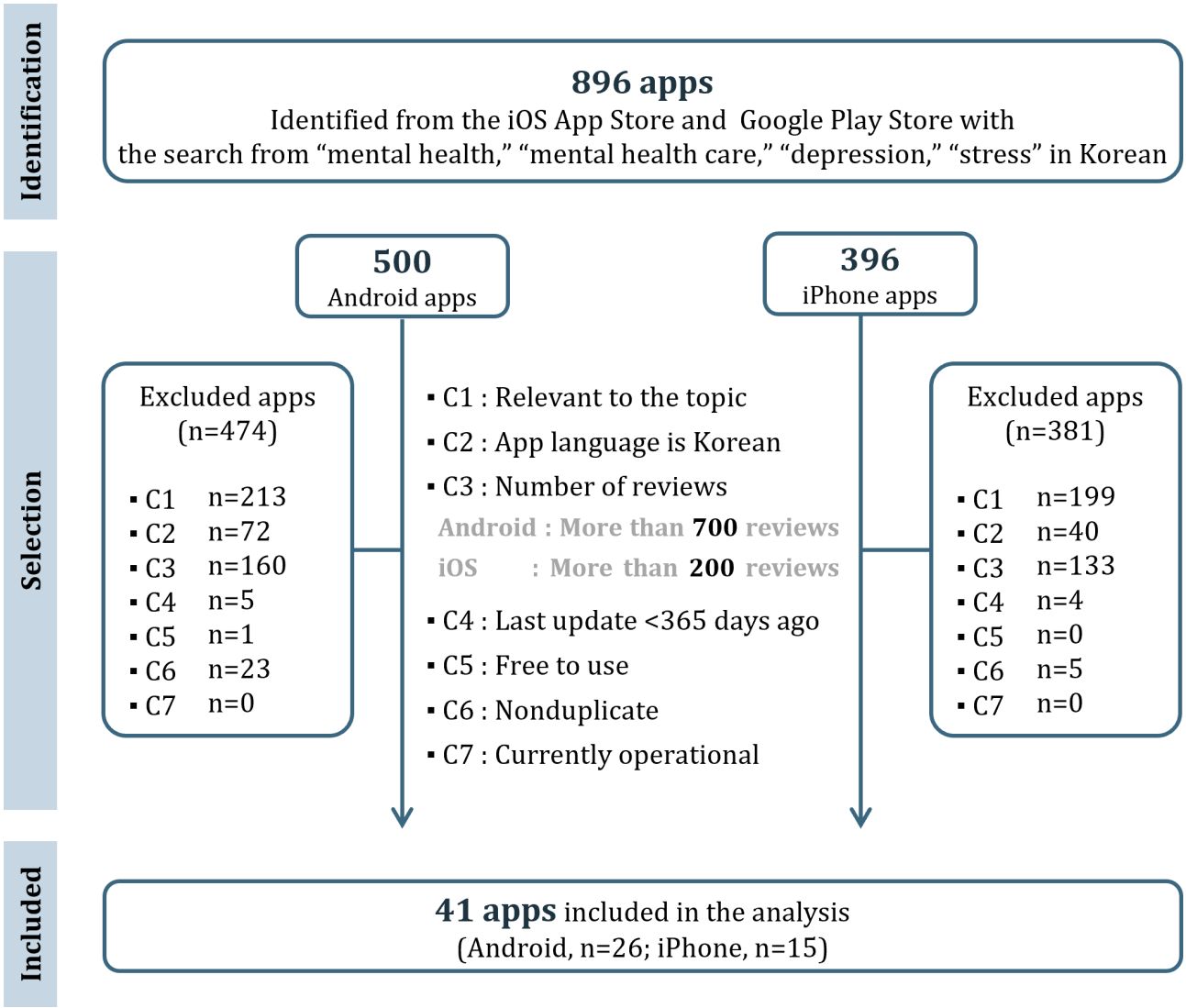
After normalizing the data, the correlations among app content, uMARS scores, star ratings, and number of reviews were analyzed using Pearson correlation analysis. To explore the relationship between objective and subjective indicators, 4 separate correlation analyses were conducted by categorizing the aforesaid data into all apps, the top 20 apps, Android apps, and iOS apps. Multiple regression analysis was performed to identify the content factors influencing uMARS scores and evaluation items. All analyses were conducted using R software (version 4.1.2; R Core Team).

Results

App Selection

A total of 896 apps were identified using the search keywords. Apps that did not meet the inclusion criteria (C1-C7) were excluded, resulting in 41 apps being selected for the final analysis. This included 26 Android apps and 15 iOS apps (Figure 1).

Figure 1. Flowchart depicting the app review and selection process (n=41). C1-7: criteria 1 - 7.



The app's operating system (OS) is linked to the App Store, allowing app updates whenever the OS is updated [34]. Therefore, depending on the timing of the OS update, the version and function of each OS may differ, and the same app on different OS versions was considered a different app.

Analysis of App Contents

Of the 41 apps, 29 (71%) included reminders, as well as recording and statistics features. Among the monitoring functions, 24 (59%) offered diary features, and 21 (51%) provided emotion tracking features. In terms of privacy, 26 (63%) included a lock feature, which was more common than

XSL•FO
RenderX

in existing mHealth apps. However, only 1 (2%) provided an icon change option (Table 1).

Table . App contents analysis results.

Contents	Android (n=26), n (%)	iPhone (n=15), n (%)	Total (N=41), n (%)
Information			
Health information	9 (35)	3 (20)	12 (29)
Health care provider information	7 (27)	5 (33)	12 (29)
Cognitive behavioral therapy	8 (31)	2 (13)	10 (24)
Review	4 (15)	3 (20)	7 (17)
Function			
Reminder	18 (69)	11 (73)	29 (71)
Connected wearable devices	3 (12)	2 (13)	5 (12)
Fingerprint sensor and camera	3 (12)	1 (7)	4 (10)
Entertainment	3 (12)	1 (7)	4 (10)
Social networking site log-in interlock	15 (58)	10 (67)	25 (61)
Direct log-in	8 (31)	7 (47)	15 (37)
Service			
Mindfulness and meditation	11 (42)	7 (47)	18 (44)
Counseling/question and answer	8 (31)	6 (40)	14 (34)
Recording and statistics	20 (77)	9 (60)	29 (71)
Chatting based on algorithm	6 (23)	5 (33)	11 (27)
Self-diagnosis			
Self-diagnosis through questionnaire	10 (38)	5 (33)	15 (37)
Mood checking via artificial intelligence	2 (8)	2 (13)	4 (10)
Monitoring			
Tracker anxiety	4 (15)	2 (13)	6 (15)
Tracker depression	7 (27)	3 (20)	10 (24)
Tracker stress	7 (27)	3 (20)	10 (24)
Tracker emotions	12 (46)	9 (60)	21 (51)
Daily diary	14 (54)	10 (67)	24 (59)
Behavior change technology			
Goal and planning	5 (19)	2 (13)	7 (17)
Community and social networking sites share	14 (54)	7 (47)	21 (51)
Privacy			
Icon change	0 (0)	1 (7)	1 (2)
Lock	15 (58)	11 (73)	26 (63)

Evaluation of App Quality

An evaluation of the quality using uMARS resulted in an overall mean score of 3.53 (SD 0.34). Among the apps, “Mind Cafe Lite” for Android achieved the highest score of 4.08 (SD 0.30), while “Depression Test” received the lowest score of 2.68 (SD 0.57). In the uMARS evaluation, the top 5 apps received their highest scores in the information category. They shared a common feature: providing information and counseling services from counselors and counseling agencies. However, most of these services, including counseling and meditation, required a fee.

Conversely, the bottom 5 apps received the lowest scores on subjective evaluation items. These apps offered limited or no

XSL•FO
RenderX

counseling-related services commonly provided by the top 5 apps. Moreover, most of them offered only a single function, such as stress measurement or depression testing, which limited their utility for consistent mental health management and promotion (Table 2). uMARS scores of all 41 apps are presented in Multimedia Appendix 2.

Table . The 5 highest and lowest scoring apps based on the user version of the Mobile Application Rating Scale (n=41).

App	Operating system	Score, mean (SD)					
		Engagement	Functionality	Aesthetics	Information	Subjective quality	Overall score
Top 5[a]							
	Android	3.89 (0.46)	4.43 (0.35)	4.14 (0.39)	4.43 (0.39)	3.54 (0.45)	4.08 (0.30)
	Android	4.37 (0.20)	4.11 (0.29)	4.29 (0.49)	4.29 (0.47)	3.29 (0.43)	4.07 (0.31)
	iOS	4.14 (0.33)	4.11 (0.35)	4.19 (0.35)	4.36 (0.40)	3.29 (0.60)	4.02 (0.31)
	Android	3.74 (0.35)	4.36 (0.40)	4.57 (0.61)	4.46 (0.43)	2.89 (0.73)	4.01 (0.43)
	Android	3.97 (0.39)	4.21 (0.47)	4.05 (0.42)	4.32 (0.37)	3.29 (0.54)	3.97 (0.33)
Bottom 5[b]							
	Android	2.66 (0.45)	3.54 (0.62)	2.57 (0.87)	2.96 (0.52)	1.68 (0.61)	2.68 (0.57)
	Android	2.66 (0.30)	3.57 (0.32)	2.86 (0.16)	3.32 (0.49)	1.96 (0.34)	2.87 (0.20)
	Android	2.86 (0.21)	3.32 (0.49)	3.29 (0.93)	3.18 (0.69)	2.21 (0.49)	2.97 (0.48)
	iOS	3.09 (0.50)	3.14 (0.80)	3.48 (0.56)	3.14 (0.80)	2.39 (0.72)	2.98 (0.49)
	Android	2.26 (0.38)	4.18 (0.42)	3.14 (0.39)	3.64 (0.77)	2.18 (0.68)	3.08 (0.37)

[a]Top 5 apps: 1. Mind Cafe Lite, 2. Trost, 3. Trost, 4. Elephants – Sleep, Meditation, 5. Mind Cafe.

[b]Bottom 5 apps: 1. Depression Test, 2. Instantaneous Heartbeat Monitor, Health, Home Training, Stress Measurement, 3. Dochi - My Own Emotional Diary, 4. Owaves, 5. eMoods.

The top 5 apps commonly provided information about health care providers, while 3 of the apps included content related to mental health and cognitive behavioral therapy. One app offered counseling and meditation services, but these were available only through paid features such as in-app purchases. By contrast, among the bottom 5 apps, none provided information about health care providers, and only 1 offered mental health information. Most of the bottom 5 apps had single functions, such as stress measurement, depression testing, or emotional record-keeping, making them less suitable for ongoing mental health management and promotion (Multimedia Appendix 3).

Statistical Analysis

Table 3 shows the correlations among the content of mental health apps, uMARS scores, star ratings, and number of reviews. Health care provider information (r=0.53; P<.001) and counseling/question and answer (Q&A; r=0.55; P<.001) had the strongest correlations with the uMARS score. Mindfulness and meditation (r=0.40; P=.01) and chatting algorithms (r=0.34; P=.03) showed weak correlations with it. Star rating was not correlated with app content. Recording and statistics (r=0.41; P=.009) and community and social networking site sharing (r=0.33; P=.04) showed weak correlations with the number of reviews.

XSL•FO
RenderX

Table . Correlation analysis of uMARS[a] scores, star rating, number of reviews, and app content. Pearson *r* values are presented.

App contents	uMARS score	Star rating	Number of reviews
Information			
Health information	0.20	–0.05	0.30
Health care provider information	0.53[b]	0.03	–0.11
Cognitive behavioral therapy	0.29	0.16	0.04
Review	0.37[c]	0.06	–0.06
Function			
Reminder	0.11	0.26	–0.05
Connected wearable devices	–0.13	–0.18	0.14
Fingerprint sensor and camera	–0.05	–0.21	0.23
Entertainment	0.17	0.06	0.08
Social networking site log-in interlock	0.36[c]	0.00	–0.05
Direct log-in	0.19	–0.16	0.11
Service			
Mindfulness and meditation	0.40[c]	0.20	0.08
Counseling/question and answer	0.55[b]	0.10	–0.20
Recording and statistics	0.00	0.10	0.41[d]
Chatting based on algorithm	0.34[c]	0.05	–0.12
Self-diagnosis			
Self-diagnosis through questionnaire	0.15	–0.11	–0.05
Mood checking via artificial intelligence	0.14	–0.03	0.10
Monitoring			
Tracker anxiety	0.00	0.12	0.12
Tracker depression	–0.01	–0.05	0.17
Tracker stress	–0.01	–0.05	0.17
Tracker emotions	0.17	0.20	–0.01
Daily diary	0.33[c]	0.26	0.13
Behavior change technology			
Goal and planning	0.20	0.09	0.29
Community and social networking site share	0.25	0.20	0.33[c]
Privacy			
Icon change	–0.26	0.05	–0.24
Lock	0.30	0.22	–0.07

[a]uMARS: user version of the Mobile Application Rating Scale.

[b]*P*<.001.

[c]*P*<.05.

[d]*P*<.01.

Figure 2 depicts the correlation between users' objective uMARS score, subjective star ratings, and number of reviews. Among all apps (r=0.51; P<.001) and Android apps (r=0.61; P<.001), there was a significant correlation between the uMARS scores and star ratings. Meanwhile, no significant correlation

XSL•FO
RenderX

was found between uMARS score and star ratings among the top 20 apps (r=–0.08; P=.74) and iOS apps (r=–0.04; P=.89).

Figure 2. Correlation analysis for the user version of the Mobile Application Rating Scale (uMARS), star ratings, and number of reviews.

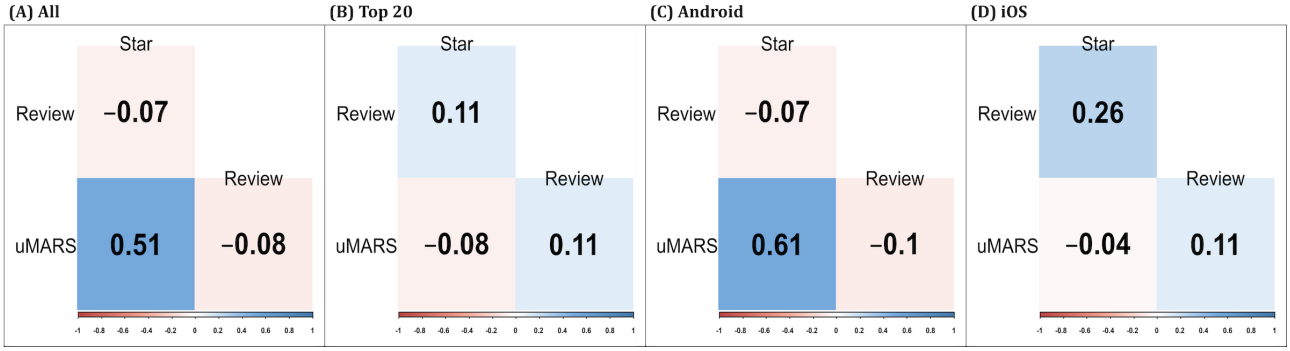


Table 4 presents the outcomes of the multiple regression analysis used to identify the factors influencing uMARS scores. A higher presence of information content in an app (β=.56; P<.001) and a higher star rating (β=.36; P=.02) were associated with higher uMARS scores. Furthermore, we examined the factors influencing each evaluation item within the uMARS assessment. The presence of information content significantly influenced the overall uMARS overall score. Moreover, apps with higher star ratings had correspondingly higher scores in the aesthetics category (β=.12; P<.001) and the subjective evaluation category (β=.34; P=.04; Table 5).

Table . Multiple regression of user version of the Mobile Application Rating Scale scores.[a]

Variables	Regression		β coefficient	*t* test (*df*)
	B	SE		
Intercept	2.29	0.44	0	5.225 (30)[b]
Information	0.63	0.20	.56	3.134 (30)[c]
Functions	–0.19	0.25	–.13	–0.761 (30)
Service	0.19	0.18	.17	1.041 (30)
Self-diagnosis	–0.06	0.19	–.06	–0.323 (30)
Monitoring	–0.01	0.15	–.01	–0.058 (30)
Behavior change technology	–0.05	0.17	–.06	–0.306 (30)
Privacy	0.15	0.19	.11	0.785 (30)
Star rating	0.24	0.09	.36	2.408 (30)[d]
Number of reviews	0	0	.06	0.380 (30)
Last update	0	0	–.05	–0.333 (30)

[a]Adjusted R^2=0.40, $F_{10,30}$=3.68 (P<.001).

[b]P<.001.

[c]P<.01.

[d]P<.05.

XSL•FO
RenderX

Table . Multiple regression for the user version of the Mobile Application Rating Scale items.

Variables	Regression		β coefficient	*t* test (*df*)	Adjusted R^2/*F* test (*df*); *P* value
	B	SE	β		
Engagement					0.37/3.36 (10, 30); <.001
Intercept	2.72	0.58	0	5.23 (30)[a]	
Information	0.60	0.27	.42	3.13 (30)[b]	
Functions	0.07	0.33	.04	–0.76 (30)	
Service	0.26	0.24	.18	1.041 (30)	
Self-diagnosis	–0.12	0.26	–.08	–0.32 (30)	
Monitoring	–0.10	0.20	–.08	–0.06 (30)	
Behavior change technology	–0.04	0.22	–.03	–0.31 (30)	
Privacy	0.51	0.25	.28	0.79 (30)	
Star rating	0.09	0.13	.11	2.41 (30)	
Number of reviews	0	0	–.07	0.38 (30)	
Last update	0	0	–.21	–0.33 (30)	
Functionality					0.24/2.28 (10, 30); <.001
Intercept	3.20	0	.46	6.92 (30)[a]	
Information	0.14	0.14	.21	0.69 (30)	
Functions	–0.82	–0.62	.26	–3.12 (30)[c]	
Service	0.28	0.26	.19	1.44 (30)	
Self-diagnosis	0.16	0.16	.20	0.79 (30)	
Monitoring	0.05	0.05	.16	0.31 (30)	
Behavior change technology	0.05	0.06	.18	0.30 (30)	
Privacy	0.12	0.09	.20	0.60 (30)	
Star rating	0.18	0.29	.10	1.72 (30)	
Number of reviews	0	0.12	0	0.66 (30)	
Last update	0	0.20	0	1.26 (30)	
Information					0.47/4.61 (10, 30); <.001
Intercept	2.69	0.00	.50	5.34 (30)[a]	
Information	1.12	0.82	.23	4.89 (30)[a]	
Functions	–0.22	–0.13	.29	–0.78 (30)	
Service	0.07	0.05	.21	0.32 (30)	
Self-diagnosis	–0.16	–0.12	.22	–0.73 (30)	
Monitoring	0.18	0.15	.17	1.02 (30)	
Behavior change technology	–0.16	–0.14	.19	–0.82 (30)	
Privacy	–0.09	–0.05	.22	–0.43 (30)	
Star rating	0.18	0.22	.11	1.54 (30)	
Number of reviews	0	0.07	0	0.48 (30)	

XSL•FO
RenderX

Variables	Regression		β coefficient	*t* test (*df*)	Adjusted R^2/*F* test (*df*); *P* value
	B	SE	β		
Last update	0	–0.04	0	–0.31 (30)	
Aesthetics					0.36/3.28 (10, 30); <.001
Intercept	1.81	0	.54	3.35 (30)[a]	
Information	0.55	0.41	.25	2.21 (30)[b]	
Functions	0.13	0.07	.31	0.41 (30)	
Service	0.04	0.03	.22	0.20 (30)	
Self-diagnosis	–0.24	–0.19	.24	–1.02 (30)	
Monitoring	0.03	0.02	.19	0.15 (30)	
Behavior change technology	–0.11	–0.09	.21	0.51 (30)	
Privacy	–0.09	–0.05	.24	–0.38 (30)	
Star rating	0.44	0.55	.12	3.56 (30)[c]	
Number of reviews	0	0.03	0	0.18 (30)	
Last update	0	–0.10	0	–0.71 (30)	
Subjective quality					0.34/3.05 (10, 30); <.001
Intercept	0.99	0	.65	1.54 (30)	
Information	0.67	0.43	.29	2.26 (30)[b]	
Functions	0.05	0.03	.37	0.14 (30)	
Service	0.29	0.18	.27	1.08 (30)	
Self-diagnosis	–0.02	–0.01	.29	–0.06 (30)	
Monitoring	–0.22	–0.16	.22	–0.99 (30)	
Behavior change technology	0.03	0.02	.25	0.13 (30)	
Privacy	0.36	0.18	.28	1.28 (30)	
Star rating	0.31	0.34	.15	2.14 (30)[b]	
Number of reviews	0	0.06	0	0.35 (30)	
Last update	0	0.02	0	0.11 (30)	

[a]*P*<.001.

[b]*P*<.05.

[c]*P*<.01.

Discussion

Principal Findings and Comparison With Prior Work

The development of various mental health apps is on the rise, and the use of mHealth apps is becoming increasingly vital in the mental health field. Nevertheless, few studies have assessed user satisfaction and app quality in the context of mental health apps. In this study, we used uMARS to evaluate mental health apps from the user perspective and to identify factors that influence user evaluations. Users preferred apps that provided information related to mental health promotion, including details about counselors and counseling institutions. However, only a limited number of apps in our final selection provided this type of information. Mental health apps are effective in enhancing patient knowledge and symptom management, primarily due to their multimedia and audiovisual content, which can deliver mental health information and education [35,36]. Users generally perceive online health information as reliable, fostering trust in their treatment regimens and encouraging active participation in disease management [37]. Moreover, the inclusion of mental health information within these apps can contribute to improved mental health literacy among users [38]. To effectively support users in promoting their mental health through apps, future app

XSL•FO
RenderX

development and updates should prioritize the inclusion of mental health–related content.

Counseling and meditation services, which constituted the primary content of the top-rated apps in user evaluations, were mostly provided for a fee. Although these apps can be downloaded for free, they often require payment for in-app purchases or access beyond a limited free trial period. The cost factor remains a prominent concern among mental health app users, and apps offering paid services may be less likely to be used [15]. Users generally prefer free apps that support independent management and monitoring of their mental health [39]. In Korea, public institutions are developing various mental health management apps; 58 of the 896 (6.47%) apps in our initial search were developed by public institutions. However, these apps were not included in our final analysis due to their extended update cycles, limited utilization, and small number of user reviews. Furthermore, mental health–related apps developed as part of research projects often focus on study objectives rather than user needs [35]. Evidence-based mental health promotion projects in the public sector are also relatively rare. To enhance app quality and encourage the use of public apps for mental health promotion, it is advisable to benchmark the services offered by top-ranking paid apps and align them with research-driven initiatives.

Chatting based on algorithms demonstrated a weak but significant correlation with uMARS scores. Users are increasingly interested in utilizing AI chatbots that provide convenient access to medical information without relying solely on health care providers [40]. Emerging technologies, such as generative AI, have the potential to significantly impact the health care sector by delivering tailored health advice, support, and integrated systems for detecting and managing health issues [8,41]. In the field of mental health, generative AI can serve as a valuable tool to support the treatment process by offering functions such as 24-hour online support, counseling record maintenance, and follow-up management [42]. It also has the potential to bridge gaps in the diagnosis and treatment of mental disorders, expanding access to psychiatric care [43]. Currently, AI chatbot functions in mental health apps are somewhat limited, often restricted to depression self-diagnosis tests (eg, the 9-item Patient Health Questionnaire). Additionally, technical issues such as limited response capabilities and unnatural conversation flow have been noted. The lack of personalized feedback for continuous engagement also hinders user satisfaction and long-term participation [44-47]. If AI counseling services that provide evidence-based, customized solutions by integrating existing mental health apps with generative AI functions are developed, they are expected to deliver higher-quality mental health services to users. To achieve this, it is essential to explore future app development strategies that incorporate generative AI in diverse ways. These efforts should focus on developing advanced algorithms that enhance natural conversation flow and provide sophisticated, personalized support. Furthermore, continuous clinical validation is necessary to ensure the reliability, accuracy, and safety of AI technologies.

The correlation between the uMARS score, an objective quality evaluation tool for mHealth apps, and user star ratings differed depending on the platform. Considering previous studies reporting digital disparities and age differences between Android and iOS users [48,49], these factors may have contributed to variations in app evaluation criteria, user preferences, and expectation levels. However, some studies have found no significant differences in personality traits or characteristics between users of the 2 platforms [50]. Combining these conflicting findings, it can be inferred that differences in user evaluations across platforms may be influenced not only by user characteristics but also by app usage experiences and platform-specific features. Future research should develop evaluation strategies that systematically analyze these differences and account for platform-specific characteristics.

Mental health apps collect and share sensitive information, including user symptoms and counseling records. When handling such sensitive personal data, it is crucial to build trust with users by ensuring the protection of their personal information [51]. Many users consider features such as app locking to be essential when using mental health apps [39]; however, few of the apps included in our final study offered the option to change the app icon for privacy purposes, and some even lacked lock functionality. Moreover, concerns about unclear personal information protection, the absence of relevant regulations, and a lack of personalized attention from apps have been steadily raised in mHealth research. As a result, some countries are now beginning to implement personal data protection regulations [52-54]. While regulations such as the Health Insurance Portability and Accountability Act and the General Data Protection Regulation were developed to protect personal information, they are not universally applied, leaving privacy and security challenges unresolved [55]. In response to these challenges, researchers are actively exploring sustainable systems to safeguard personal information using blockchain technology [56]. Blockchain-based health care systems offer personal information protection and provide users the option to request the deletion of their data [57]. In the mental health field, preserving the security and confidentiality of users' mental health data is considered paramount [58]. Consequently, app developers must prioritize including privacy features suited to the app's characteristics and aligned with existing personal information protection regulations.

None of the apps in our final analysis targeted specific population groups. Although apps for children and adolescents are in development, they remain difficult to find, and their specific benefits for particular target groups are still unclear [59]. Mental health issues vary across life stages, implying that desired app content may also differ by age group [60]. Previous research highlights the importance of tailoring interventions to life stages due to variations in mental health problems [61,62]. However, such a life stage–specific approach has yet to be fully integrated into app development. Future efforts should focus on developing mental health apps that address the unique needs of each life stage, considering factors such as sex and age. This requires active research into the usability and effectiveness of apps across different population groups to better address the diverse needs of users.

The World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC) offer targeted guidelines for mental health management, categorized by different

XSL•FO
RenderX

population groups such as parents, adults, children, individuals with mental health conditions, public health workers, and health care professionals [63,64]. WHO also provides translated stress-management guides for use in different countries, along with illustrated materials that are easy for users to follow [65]. The CDC has created a dedicated page that consolidates guidelines from various institutions and includes links to mental health experts that can be accessed during emergencies. This allows users to access free or low-cost mental health services through a structured system based on standardized criteria, enabling them to choose an expert, appointment time, and office environment. By contrast, such detailed interventions are not yet standard practice for mental health management in Korea. Consequently, mental health services are often underutilized. To promote mental health services in the future, concerted efforts are required to raise awareness and ensure easy access to services through diverse interventions, such as national guidelines and expert-led initiatives offered by the WHO and CDC.

Limitations

This study had some limitations. First, the apps were selected based on searches conducted within app stores during a specific period. As app stores are continuously updated, the content analyzed in this study may differ from that of future app versions. Second, the data sources were not electronic academic databases but app stores, and the accuracy of the search process may have varied depending on each store's recommendation algorithms.

Conclusions

In this study, users evaluated the quality of mental health apps using a validated scale. Users preferred apps that included information related to mental health. To develop high-quality mental health apps, it is essential to incorporate technologies such as generative AI into the app development process. This study is meaningful in that it identified the types of content preferred by users through a quality evaluation and suggested directions for developing apps that meet current needs. We expect our results to serve as guidelines for creating evidence-based data and developing health promotion apps in the public sector. Further research focusing on the long-term usability, effectiveness, and user engagement of mental health apps—through extended evaluations from a user perspective—could contribute to improving their overall quality and impact.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2024-00350688) and this paper was written based on SK's thesis [66].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the app contents.
[DOCX File, 21 KB - mhealth_v13i1e64622_app1.docx]

Multimedia Appendix 2
The user version of the Mobile Application Rating Scale score of all 41 apps.
[XLSX File, 13 KB - mhealth_v13i1e64622_app2.xlsx]

Multimedia Appendix 3
Content comparison between the top and bottom 5 apps (ranked according to the user evaluation scores).
[DOCX File, 44 KB - mhealth_v13i1e64622_app3.docx]

References

1. World Health Organization (WHO). World Mental Health Report: Transforming Mental Health for All: WHO; 2022. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2025-06-30]
2. Drissi N, Ouhbi S, Janati Idrissi MA, Fernandez-Luque L, Ghogho M. Connected mental health: systematic mapping study. J Med Internet Res 2020 Aug 28;22(8):e19950. [doi: 10.2196/19950] [Medline: 32857055]
3. Westerhof GJ, Keyes CLM. Mental illness and mental health: the two continua model across the lifespan. J Adult Dev 2010 Jun;17(2):110-119. [doi: 10.1007/s10804-009-9082-y] [Medline: 20502508]
4. Lee SE, Shin H, Hur JW. mHealth for mental health in the COVID-19 era. kjcp 2020 Nov 30;39(4):325-354. [doi: 10.15842/kjcp.2020.39.4.007]
5. Suicide rates. OECD. 2018. URL: https://www.oecd.org/en/data/indicators/suicide-rates.html [accessed 2025-06-23]
6. An S, Lee H. Use of mobile mental health application for mental health promotion: based on the information-motivation-behavioral skills model. Korean Journal of Journalism & Communication Studies 2018 Dec 31;62(6):167-194. [doi: 10.20879/kjjcs.2018.62.6.006]

XSL•FO
RenderX

7. Han SY, Kim HS, Kim J. Crowdsourcing-based psychotherapy application UX design. In: Proceedings of HCI Korea 201 2017:1976-0671.
8. Kalla D, Smith N. Study and analysis of ChatGPT and its impact on different fields of study. International Journal of Innovative Science and Research Technology 2023 Mar 24;8(3):827-833. [doi: 10.5281/zenodo.7767675]
9. Griffiths KM, Bennett K, Walker J, Goldsmid S, Bennett A. Effectiveness of MH-Guru, a brief online mental health program for the workplace: a randomised controlled trial. Internet Interv 2016 Nov;6:29-39. [doi: 10.1016/j.invent.2016.09.004] [Medline: 30135812]
10. Irvine AB, Philips L, Seeley J, Wyant S, Duncan S, Moore RW. Get moving: a website that increases physical activity of sedentary employees. Am J Health Promot 2011;25(3):199-206. [doi: 10.4278/ajhp.04121736] [Medline: 21192750]
11. Gan DZQ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user engagement with digital mental health interventions: a systematic review. Digit Health 2022;8:20552076221098268. [doi: 10.1177/20552076221098268] [Medline: 35677785]
12. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046. [doi: 10.2196/17046] [Medline: 32186518]
13. Brian RM, Ben-Zeev D. Mobile health (mHealth) for mental health in Asia: objectives, strategies, and limitations. Asian J Psychiatr 2014 Aug;10:96-100. [doi: 10.1016/j.ajp.2014.04.006] [Medline: 25042960]
14. van Velsen L, Beaujean DJMA, van Gemert-Pijnen JEWC. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013 Feb 11;13(1):1-5. [doi: 10.1186/1472-6947-13-23] [Medline: 23399513]
15. Hudson G, Negbenose E, Neary M, et al. Comparing professional and consumer ratings of mental health apps: mixed methods study. JMIR Form Res 2022 Sep 23;6(9):e39813. [doi: 10.2196/39813] [Medline: 36149733]
16. Martín-Martín J, Muro-Culebras A, Roldán-Jiménez C, et al. Evaluation of Android and Apple Store depression applications based on Mobile Application Rating Scale. Int J Environ Res Public Health 2021 Nov 27;18(23):12505. [doi: 10.3390/ijerph182312505] [Medline: 34886232]
17. Carrouel F, du Sartz de Vigneulles B, Bourgeois D, et al. Mental health mobile apps in the French App Store: assessment study of functionality and quality. JMIR Mhealth Uhealth 2022 Oct 12;10(10):e41282. [doi: 10.2196/41282] [Medline: 36223178]
18. Kang GS, Roh SS. Development of mental health self-care app for university student. JKEIA 2019 Jan 31;13(1):25-34. [doi: 10.21184/jkeia.2019.1.13.1.25]
19. Park JJ, Choi GS, Kim JL, Park IK, Kang JJ, Son BK. Development of mobile healthcare app for mental health management - focused on anger management. JIIBC 2014 Dec 31;14(6):13-18. [doi: 10.7236/JIIBC.2014.14.6.13]
20. Kim SM, Yang JY, Lee JH. Case study of mobile app-based mental health management: focused on hospital employees. JNCIST 2022 Sep 30;6(9):1634-1642. [doi: 10.33097/JNCTA.2022.06.09.1634]
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Ko S, Lee J, An D, Woo H. Menstrual tracking mobile app review by consumers and health care providers: quality evaluations study. JMIR Mhealth Uhealth 2023 Mar 1;11:e40921. [doi: 10.2196/40921] [Medline: 36857125]
23. Davalbhakta S, Advani S, Kumar S, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164. [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
24. Wu X, Xu L, Li P, Tang T, Huang C. Multipurpose mobile apps for mental health in chinese app stores: content analysis and quality evaluation. JMIR Mhealth Uhealth 2022 Jan 4;10(1):e34054. [doi: 10.2196/34054] [Medline: 34982717]
25. Korea smartphone market share: by quarter. Counterpoint Technology Market Research. 2022. URL: https://korea.counterpointresearch.com/ [accessed 2024-05-28]
26. Byun AJ. Development and Evaluation of Smartphone App “Healthy Kids” for Children’s Weight Control with Their: Seoul National University; 2018. URL: https://www.dcollection.net/handler/snu/000000151003 [accessed 2025-06-23]
27. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
28. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015;3(1):e27. [doi: 10.2196/mhealth.3422]
29. Lambrecht A, Vuillerme N, Raab C, et al. Quality of a supporting mobile app for rheumatic patients: patient-based assessment using the user version of the Mobile Application Scale (uMARS). Front Med (Lausanne) 2021;8:715345. [doi: 10.3389/fmed.2021.715345] [Medline: 34368202]
30. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: user-centered study with employees from a Lebanese university. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e9836. [doi: 10.2196/mhealth.9836] [Medline: 30672742]

XSL•FO
RenderX

31. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. “Prostate Cancer Risk Calculator” mobile applications (apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
32. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. Presented at: The SIGCHI Conference; Apr 24-29, 1993; Amsterdam, The Netherlands. [doi: 10.1145/169059.169166]
33. Kang YM, Lee SY. User evaluation of the mobile application, CAMBRA-kids, for caries management in preschoolers. Korean Society of Dental Hygiene 2019;19(5):859-871. [doi: 10.13065/jksdh.20190073]
34. Novac OC, Novac M, Gordan C, Berczes T, Bujdoso G. Comparative study of google android, apple ios and microsoft windows phone mobile operating systems. Presented at: 2017 14th International Conference on Engineering of Modern Electric Systems (EMES); Jun 1-2, 2017; Oradea, Romania. [doi: 10.1109/EMES.2017.7980403]
35. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]
36. Timmers T, Janssen L, Kool RB, Kremer JA. Educating patients by providing timely information using smartphone and tablet apps: systematic review. J Med Internet Res 2020 Apr 13;22(4):e17342. [doi: 10.2196/17342] [Medline: 32281936]
37. Eastin MS. Credibility assessments of online health information: the effects of source expertise and knowledge of content. J Comput Mediat Commun 2001 Jul 1;6(4). [doi: 10.1111/j.1083-6101.2001.tb00126.x]
38. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. Am Psychol 2012 Apr;67(3):231-243. [doi: 10.1037/a0025957] [Medline: 22040221]
39. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64. [doi: 10.2196/jmir.1475] [Medline: 21169174]
40. Johnson D, Goodman R, Patrinely J, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the chat-GPT model. Res Sq. Preprint posted online on Feb 28, 2023 URL: https://www.researchsquare.com/article/rs-2566942/v1 [accessed 2025-06-12] [doi: 10.21203/rs.3.rs-2566942/v1] [Medline: 36909565]
41. Biswas SS. Role of ChatGPT in public health. Ann Biomed Eng 2023 May;51(5):868-869. [doi: 10.1007/s10439-023-03172-7] [Medline: 36920578]
42. He Y, Liang K, Han B, Chi X. A digital ally: the potential roles of ChatGPT in mental health services. Asian J Psychiatr 2023 Oct;88(103726):103726. [doi: 10.1016/j.ajp.2023.103726] [Medline: 37567084]
43. Singh OP. Artificial intelligence in the era of ChatGPT - opportunities and challenges in mental health care. Indian J Psychiatry 2023 Mar;65(3):297-298. [doi: 10.4103/indianjpsychiatry.indianjpsychiatry_112_23] [Medline: 37204980]
44. Chung K, Cho HY, Park JY. A chatbot for perinatal women’s and partners’ obstetric and mental health care: development and usability evaluation study. JMIR Med Inform 2021 Mar 3;9(3):e18607. [doi: 10.2196/18607] [Medline: 33656442]
45. Kleinau E, Lamba T, Jaskiewicz W, et al. Effectiveness of a chatbot in improving the mental wellbeing of health workers in Malawi during the COVID-19 pandemic: a randomized, controlled trial. PLoS One 2024;19(5):e0303370. [doi: 10.1371/journal.pone.0303370] [Medline: 38805444]
46. Daley K, Hungerbuehler I, Cavanagh K, Claro HG, Swinton PA, Kapps M. Preliminary evaluation of the engagement and effectiveness of a mental health chatbot. Front Digit Health 2020;2(576361):576361. [doi: 10.3389/fdgth.2020.576361] [Medline: 34713049]
47. Anmella G, Sanabra M, Primé-Tous M, et al. Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and health care professionals: development, feasibility, and potential effectiveness studies. J Med Internet Res 2023 Apr 3;25:e43293. [doi: 10.2196/43293] [Medline: 36719325]
48. Pryss R, Schobel J, Hoppenstedt B, et al. Ecological momentary assessment based differences between Android and iOS users of the TrackYourHearing mHealth crowdsensing platform. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3951-3955. [doi: 10.1109/EMBC.2019.8857854] [Medline: 31946737]
49. Ubhi HK, Kotz D, Michie S, van Schayck OCP, West R. A comparison of the characteristics of iOS and Android users of a smoking cessation app. Transl Behav Med 2017 Jun;7(2):166-171. [doi: 10.1007/s13142-016-0455-z] [Medline: 28168609]
50. Götz FM, Stieger S, Reips UD. Users of the main smartphone operating systems (iOS, Android) differ only little in personality. PLoS One 2017;12(5):e0176921. [doi: 10.1371/journal.pone.0176921] [Medline: 28467473]
51. Hollis C, Morriss R, Martin J, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
52. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women’s mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022 May 6;10(5):e33735. [doi: 10.2196/33735] [Medline: 35522465]
53. Koh KS, Lim HS, Lim J, Wei Y, Minn PW, Wong J. Development and implementation of a national mobile health application: a case study from Brunei. J Glob Health 2022 Dec 21;12(3083):03083. [doi: 10.7189/jogh.12.03083] [Medline: 36538470]
54. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8. [doi: 10.2196/mhealth.3672] [Medline: 25599627]

XSL•FO
RenderX

55. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 5;2(4):e192542. [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
56. Upadhyay S, Kumar M, Kumar A, et al. Feature extraction approach for speaker verification to support healthcare system using blockchain security for data privacy. Comput Math Methods Med 2022;2022:8717263. [doi: 10.1155/2022/8717263] [Medline: 35924113]
57. Sengupta A, Subramanian H. User control of personal mHealth data using a mobile blockchain app: design science perspective. JMIR Mhealth Uhealth 2022 Jan 20;10(1):e32104. [doi: 10.2196/32104] [Medline: 35049504]
58. Bondre A, Pathare S, Naslund JA. Protecting mental health data privacy in India: the case of data linkage with Aadhaar. Glob Health Sci Pract 2021 Sep 30;9(3):467-480. [doi: 10.9745/GHSP-D-20-00346] [Medline: 34593574]
59. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
60. Kang MK, Kim MS, Gang M, Oh K, Kwon JS, Lee SH. Factors affecting the mental health related quality of life in adults across the lifespan. Korean J Rehabil Nurs 2012 Dec 31;15(2):73-82. [doi: 10.7587/kjrehn.2012.73]
61. Park YR, Son YJ. Relationship of satisfaction with appearance, self-esteem, depression, and stress to health related quality of life in women across the lifespan. Korean Acad Nurs 2009;16(3):353-360 [FREE Full text]
62. Jeon J, Choi J. Mental health level and mental health support status by life cycle. Health and Welfare Policy Forum 2017(1):75-85. [doi: 10.23062/2017.01.7]
63. #HealthyAtHome--mental health. World Health Organization (WHO). URL: https://www.who.int/campaigns/connecting-the-world-to-combat-coronavirus/healthyathome/healthyathome---mental-health?utm_source=HowRightNow&utm_medium=HRNwebsite [accessed 2025-06-30]
64. Mental health. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/mentalhealth/index.htm [accessed 2024-05-28]
65. Mental Health, Brain Health and Substance Use (MSD). Doing What Matters in Times of Stress: An Illustrated Guide: World Health Organization (WHO); 2020. URL: https://www.who.int/publications/i/item/9789240003927 [accessed 2025-06-12]
66. Ko S. Content analysis and quality evaluation of mhealth app for mental health management [Master's thesis]. : Kongju National University; 2024 URL: https://www.riss.kr/link?id=T16939784 [accessed 2025-06-23]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
mHealth: mobile health
OS: operating system
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Q&A: question and answer
uMARS: user version of the Mobile Application Rating Scale
WHO: World Health Organization

Edited by L Buis; submitted 22.07.24; peer-reviewed by F Carrouel, HY Lee, S Lee, YH Lee; revised version received 24.12.24; accepted 06.05.25; published 04.07.25.

Please cite as:
Ko S, Woo H
Users' Needs for Mental Health Apps: Quality Evaluation Using the User Version of the Mobile Application Rating Scale
JMIR Mhealth Uhealth 2025;13:e64622
URL: https://mhealth.jmir.org/2025/1/e64622
doi:10.2196/64622

© Siyeon Ko, Hyekyung Woo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cognitive Training Mobile Apps for Older Adults With Cognitive Impairment: App Store Search and Quality Evaluation

Leyi Wu[1], MSN; Jiajuan Pan[1], BN; Chuwen Dou[1], MSN; An Gu[1], MSN; An Huang[1], MSN; Hong Tao[2], PhD; Xiaoyan Wang[3], BN; Chen Zhang[3], BA; Lina Wang[1], PhD

[1]School of Medicine, Huzhou Key Laboratory of Precise Prevention and Control of Major Chronic Diseases, Huzhou University, Erhuan East Road 759, Longquan Street, Wuxing District, Huzhou, China

[2]Center for Whole-Person Research, AdventHealth Whole-Person Research, Orlando, FL, United States

[3]Department of General Medicine, Community Health Service Center of Renhuangshan, Huzhou, China

Corresponding Author:
Lina Wang, PhD
School of Medicine, Huzhou Key Laboratory of Precise Prevention and Control of Major Chronic Diseases, Huzhou University, Erhuan East Road 759, Longquan Street, Wuxing District, Huzhou, China

Abstract

Background: As the population ages, cognitive impairment is becoming increasingly prevalent. Mobile apps offer a scalable platform for delivering cognitive training interventions. However, their variable quality and lack of rigorous evaluation underscore the need for further research to guide optimization and ensure their effective application in improving cognitive health outcomes.

Objective: This study aimed to evaluate the characteristics and quality of cognitive training apps designed for older adults with cognitive impairment.

Methods: A comprehensive search of the Google Play Store and Apple App Store was conducted using predefined terms and inclusion criteria, with the search completed on July 13, 2024. Eligible apps were assessed for quality by two independent reviewers using the Mobile App Rating Scale (MARS), with interrater reliability evaluated via quadratic weighted kappa (κ). The Kruskal-Wallis test analyzed differences in MARS scores across subgroups for each dimension, and Spearman correlation was applied to examine the relationship between user star ratings and overall mean scores.

Results: A total of 4822 potential apps were identified, of which 24 met eligibility criteria. Among these, 13 (54%) were available on both platforms, 5 (21%) were exclusive to the Google Play Store, and 6 (25%) to the Apple App Store. Notably, 5 (20.8%) apps offered user-tailored training modules and 8 (33%) involved medical professionals in development. Interrater agreement was high (k=0.88; 95% CI, 0.80 - 0.95). Global quality scores based on the MARS dimensions ranged from 2.38 to 4.13, with a mean (SD) of 3.57 (0.43) across 24 apps, indicating generally acceptable quality. The functionality dimension received the highest score, while engagement scored the lowest. Brain HQ and Peak had scores above 4 and were rated as good, whereas Memory Trainer, Cognitive Skill Training, and Ginkgo Memory & Brain Training scored below 3 and were rated as insufficient. Spearman correlation showed no significant association between mean score and app rating.

Conclusions: Current cognitive training apps for older adults with cognitive impairment demonstrate moderate quality with considerable variability. Improvements are needed in the engagement and information dimensions. Future development should prioritize enhancing user engagement, incorporating personalized features, and involving health care professionals and experts to align with evidence-based guidelines.

(JMIR Mhealth Uhealth 2025;13:e69637) doi:10.2196/69637

KEYWORDS

cognitive training; older adults; mobile apps; mobile app rating scale; eHealth

Introduction

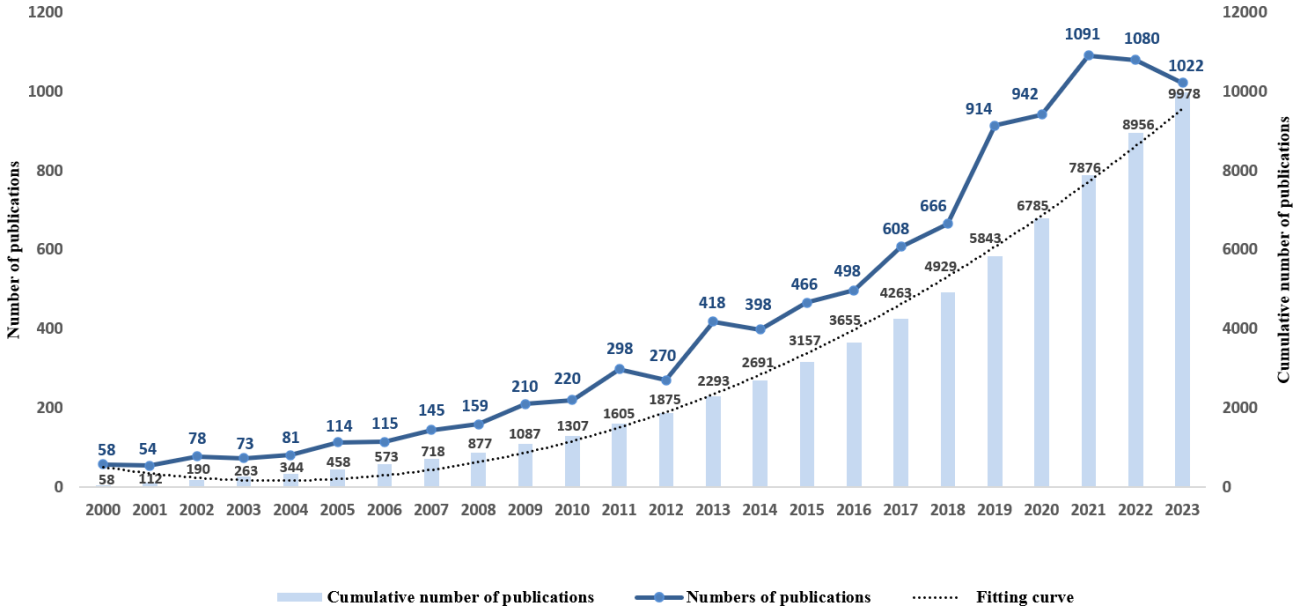
The global aging population presents a significant public health challenge due to the increasing prevalence of dementia. Aging is the primary risk factor for dementia, which affects an estimated 50 million individuals worldwide, with projections reaching 152 million by 2050 [1]. Age-related processes, including neurofibrillary tangles, amyloid-beta plaque deposition, and cerebrovascular changes, contribute to the pathogenesis of Alzheimer disease and other dementias [2]. Therefore, efforts to address cognitive decline in older adults are critical to mitigating the growing global burden of dementia.

Cognitive interventions have emerged as a promising strategy for maintaining cognitive function in older adults, particularly those with cognitive impairment. The popularity and development of digital cognitive training technology as one of

XSL•FO
RenderX

the key applications of mobile health (mHealth) app has further facilitated the transfer of cognitive training from traditional clinical or laboratory settings to everyday life. Among these, cognitive training has shown potential in mitigating cognitive decline by engaging in specific exercises targeting various cognitive domains, such as memory, attention, processing speed, and executive function, to stimulate and enhance the connectivity and efficiency of brain neural networks [3]. A systematic review by Gates et al [4] found that cognitive training resulted in significant improvements in memory and attention in older adults with mild cognitive impairment, with long-term benefits persisting beyond the intervention period. Lampit et al [5] demonstrated that cognitive training in cognitively healthy older adults, particularly computerized interventions, resulted in moderate improvements in global cognition and specific domains such as processing speed and memory. Notably, a systematic search of the scientific literature was conducted in the Web of Science using a topic search with the terms "cognitive training" and "cognitive impairment." The results show a steady increase in publications over the past 2 decades, as illustrated in Figure 1. Annual output rose from 58 papers in 2000 to over 1,000 per year after 2020, peaking at 1,091 in 2021. Despite slight declines in 2022 and 2023, the cumulative total reached 9,978 by 2023, reflecting sustained and growing academic interest in cognitive training for cognitive impairment.

Figure 1. Trends in annual and cumulative number of publications on cognitive training (2000 - 2023).



In recent years, digital interventions have attracted growing attention as promising strategies to support cognitive health among older adults with cognitive impairment. A range of modalities, including mHealth apps, digital art therapy, and digital storytelling, have been explored with the aim of enhancing cognitive engagement and promoting rehabilitation [6-9]. Systematic and narrative reviews underscore both the potential benefits of these approaches and notable challenges related to accessibility, user engagement, and clinical validation [6,10]. Among these interventions, mHealth technologies have demonstrated particular promise by improving access to care, enabling real-time monitoring, and facilitating personalized health interventions [11,12]. Cognitive training apps delivered through mobile platforms offer greater accessibility, convenience, and cost-effectiveness compared to other cognitive training methods such as paper-and-pencil tasks, computer-based programs, or virtual reality–based interventions [13-15]. A growing number of cognitive training apps are now available, providing diverse exercises aimed at enhancing memory, attention, and executive function. Studies have shown that mHealth-based cognitive training apps can maintain or improve cognitive performance, especially when incorporating features such as adaptive difficulty levels, gamification, and feedback mechanisms to promote sustained engagement [16,17]. However, despite these advances, concerns remain regarding the methodological rigor, scientific validity, and evidence base underpinning many commercially available cognitive training apps. Recent reviews have highlighted substantial variability in app quality and the limited methodological rigor and standardization in existing evaluations of cognitive training apps [18,19]. Furthermore, it remains unclear to what extent current cognitive training apps have integrated recent innovations, such as individualized digital therapies and user-centered design principles. Therefore, a systematic and comprehensive quality evaluation of cognitive training apps is critically needed to inform future development and ensure these digital interventions truly meet the complex needs of specific population.

In the field of cognitive impairment, quality assessments of caregiver-related apps have been conducted [20]. Meanwhile, recent reviews have evaluated the content and quality of cognitive training apps available to the general population [18,19]; however, these evaluations often did not specifically target individuals with cognitive impairment. In addition, a recent scoping review by Silva et al [21] emphasized the substantial heterogeneity in intervention approaches among existing cognitive training apps and the lack of a standardized framework for app quality assessment. In order to optimize mobile apps, effectively meet user needs, and promote cognitive health, it is essential to establish a reliable metric for assessing

XSL•FO
RenderX

app quality. Various tools have been developed to evaluate the quality of mHealth apps, such as the User Version of the Mobile App Rating Scale (uMARS) developed by Stoyanov et al [22], the Evaluation Tool for Mobile and Web-Based eHealth Interventions (ENLIGHT) developed by Baumel et al [23], and the System Usability Scale (SUS) developed by Brooke [24]. However, each of these instruments presents specific limitations. uMARS is primarily designed to capture feedback from general users, ENLIGHT focuses on behavioral change and lacks emphasis on broader usability and design elements, while SUS is limited to evaluating system usability alone. In contrast, the Mobile App Rating Scale (MARS), developed by Leanne's research team, offers a simple, objective, and reliable tool for evaluating the overall quality of mobile apps. It assesses five core dimensions: engagement, functionality, aesthetics, information quality, and subjective quality, with clearly defined descriptors for each rating anchor to ensure consistency and accuracy [25]. MARS has been widely applied in assessing various health-related apps, including targeting genitourinary tumors, inflammatory bowel disease, psoriasis, and heart failure, with demonstrated reliability and applicability in related research contexts [26-29].

Given these gaps in the literature, the aim of this study was to identify and evaluate publicly available cognitive training apps designed for older adults with cognitive impairment. Specifically, this study aims to (1) identify and summarize existing cognitive training apps and their core features and (2) assess the quality of these apps using the MARS. It was hypothesized that existing cognitive training apps would exhibit variable quality and might not fully meet the specific needs of older adults with cognitive impairment. The findings from this study are expected to inform researchers, developers, and clinicians, contributing to the development of more effective, user-centered digital interventions for cognitive health promotion in this vulnerable population.

Methods

Study Design

Although the search strategy differed slightly from traditional methods, the study adhered to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses standards) Checklist present in Checklist 1 [30], and was consistent with recent reviews of mobile apps that applied similar methods [26,31]. A neuropsychologist from a large medical center and a certified software engineer oversaw and reviewed the research process and findings.

App Search Strategy

A comprehensive search of the Google Play Store and Apple App Store was conducted up to July 13, 2024, using a 2-step approach, developed in collaboration with a librarian. In step 1, both app stores were searched with 67 relevant key terms (eg, "cognitive impairment," "cognitive decline," "brain," "cognitive functions," "cognitive therapy," "cognitive training," etc.). In step 2, string keywords were applied, combining multiple forms of cognitive impairment (eg, "cognitive dysfunction," "cognitive declines," "mental deterioration," etc.) with terms such as "cognitive therapy," "cognitive training," "brain training," etc. This approach accounted for differences in search algorithms: Apple's optimized for single keywords and Google Play's for string keywords. The details were presented in Multimedia Appendix 1. Relevant apps were identified based on the following inclusion criteria: (1) English language, (2) relevance to the subject matter, (3) free to download, (4) available for individual use, and (5) normal functionality.

App Content Analysis

The primary content of the 24 representative apps was assessed by the two reviewers (LW and JP). These apps were required to focus primarily on cognitive training, targeting the enhancement and improvement of memory, attention, language, comprehension, and other cognitive functions.

App Quality Evaluation

The quality of the included apps was evaluated using the MARS, a widely used tool developed by Stoyanov et al [25] in 2015 for evaluating mHealth apps. MARS has demonstrated high internal consistency (Cronbach α coefficient=0.90) and interrater reliability [25]. The scale consists of 23 items covering objective quality dimensions, including engagement, functionality, aesthetics and information, and subjective dimension. Specifically, engagement assesses whether the app is fun, interesting, customizable, interactive, and targeted; functionality assesses whether the app is easy to learn, easy to use, and logical; aesthetics assesses the graphic design, overall visual appeal, color scheme, and stylistic consistency of the app; information assesses whether the app contains high-quality information from credible sources; and the subjective quality dimension reflects user satisfaction, application adoption, and continuity of use.

Each item was scored on a 1-5 scale, with a "not applicable" option available. Following the MARS guidelines, the scoring procedure involved: (1) average score for each dimension for a single app is calculated by summing the scores of all items under that dimension and dividing by the total number of items in the dimension; (2) average score of each dimension across all apps was calculated by summing the dimension's average scores for all apps and dividing by the number of apps; (3) overall quality score for a single app is determined by averaging the average scores of the four objective dimensions (engagement, functionality, aesthetics, and information), excluding the subjective quality dimension to maintain an unbiased evaluation; and (4) average overall quality score across all apps is calculated by summing the overall quality scores of all apps and dividing by the total number of apps. An average MARS score of ≥3 points (out of 5) is considered to be of "acceptable" quality [25].

Data Collection

Two independent reviewers (LW and JP), both with medical backgrounds and specializing in cognitive impairment in older adults, conducted the app evaluations. Prior to the assessment, both reviewers received formal training in the use of the MARS to ensure consistency and accuracy. Each app was used for at least 10 minutes during evaluation to allow sufficient interaction with its functionalities. Apps were randomly assigned and

XSL•FO
RenderX

independently evaluated without communication between reviewers to minimize potential bias. Discrepancies greater than one point on any subscale were resolved through discussion; if consensus could not be reached, a third reviewer (CD) was consulted to adjudicate the final score. Interrater reliability between the two reviewers was assessed using the intraclass correlation coefficient (ICC), and the agreement was good (ICC 0.965 [95% CI 0.92 - 0.985]).

Data Analysis

The two reviewers (LW and JP) reconfirmed the dates and entered the scores into a cloud-based Microsoft Excel spreadsheet, followed by the computation of descriptive statistics for each rating. The data analysis procedure consisted of the following steps: (1) calculating the mean score for each quality dimension for each app and calculating its variance, (2) adding the mean scores for the engagement, functionality, aesthetics, and information dimensions and dividing them by the average of the number of dimensions as the MARS final score and calculating its variance, (3) calculating the overall mean score for each dimension for all apps and calculating its variance, and (4) calculating the overall mean score by summing and dividing by the total number of apps to calculate the overall average score of all apps and calculate its variance. In addition, the apps were categorized into three groups based on the overall mean score: less than 3 (Group 1), 3-4 (Group 2), and greater than 4 (Group 3). The Kruskal-Wallis test was used to compare median scores across the three groups for each of the four MARS dimensions: engagement, functionality, aesthetics, and information, identifying significant variations among the subgroups. The Spearman correlation coefficient was used to assess the correlation between the star rating score and the overall mean score.

Ethical Considerations

This study was exempt from institutional review board approval, as no risk to human participants was involved, and was registered at PROSPERO (International Prospective Register of Systematic Reviews, number CRD42024602240). This study involved a systematic search and quality assessment of publicly available mobile applications, conducted using the validated Mobile App Rating Scale (MARS). The study did not include the recruitment of human participants, collection of personal or health-related information, or any form of intervention involving individuals. All information was obtained from open-access app store listings and publicly available app descriptions. According to established ethical guidelines, including those outlined by institutional review boards (IRBs) in China and internationally (the Declaration of Helsinki, Council for International Organizations of Medical Sciences [CIOMS] guidelines), research that does not involve human participants or identifiable personal data is generally exempt from ethical review. Our study aligns with these criteria. Consistent with this, numerous high-impact, peer-reviewed studies involving app evaluations have also reported similar exemptions from ethical review, for example in references [32-36]. Based on this precedent and our institution's policy, we confirm that no ethical approval was required for this study.

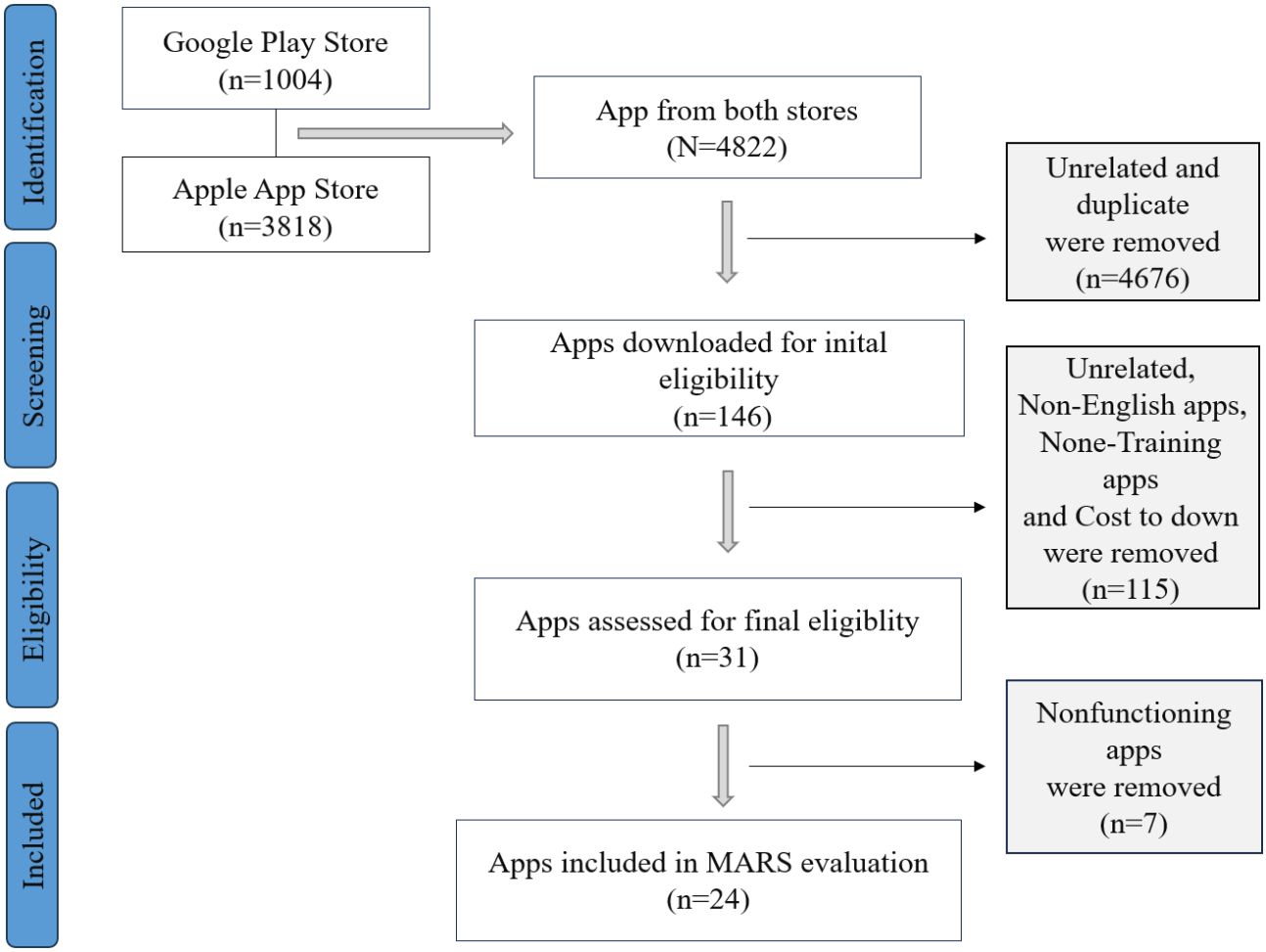
Results

App Selection

A total of 4822 apps were retrieved from the Google Play Store and Apple App Store (1004 from Google Play Store and 3818 from Apple App Store). After preliminary screening, 146 apps were assessed for eligibility with duplicate and obviously irrelevant apps were removed. Following a detailed review, 31 apps were included, but 7 apps were found to be nonfunctioning, leaving 24 apps that met the inclusion criteria and were eventually included for evaluation as shown in Figure 2.

XSL•FO
RenderX

Figure 2. The process of apps selection. MARS: Mobile App Rating Scale.



Characteristics and Purposes of Included Apps

Of these 24 apps, 13 (54%) are available on both platforms, 5 (21%) are exclusive to the Google Play Store, and 6 (25%) are accessible on the Apple App Store. Table 1 outlines the content characteristics of each app. The primary focus of the included apps is cognitive training aimed at improving memory, attention, language comprehension, and other cognitive functions. In total, 5 apps (20.8%) allow users to select training modules tailored to their individual needs. Overall, 8 apps (33.3%) indicate the involvement of medical and related professionals in their development process, as stated in the app profiles or within the apps.

XSL•FO
RenderX

Table . Description of the apps included in the study.

App name	Platform	Developer	Age group (years)	Involvement of health care	Focus (targeted areas of the app)
Memory Trainer	iOS	Svyatoslav Ivashchenko	4 years or older	Unknown	Memory
Brave	iOS	University of Hong Kong	17 years or older	Unknown	MCI[a] education and related physical training
Dementia Learning Game	iOS	Bom Soft	4 years or older	Unknown	Concentration, observation,and memory
Mindmate	iOS	Mindmate Ltd	12 years or older	Unknown	Problem-solving, attention, and lifestyle support
Cognitive Skill Training	iOS	Kazuaki Matayoshi	4 years or older	Unknown	Arithmetic, memory, and attention
Reminding	iOS	Sync International Pty Ltd	12 years or older	Unknown	Game-based cognitive training
Mindpal	Android	Elektron Labs Inc	All ages	Unknown	Language, memory, attention, and problem-solving
Elevate	Android	Elevate Labs	All ages	Yes	Attention, language, memory, processing speed, and math
Neurobics	Android	Peoresnada.com	All ages	Unknown	Mental, attention, calculation, and memory, analysis
Train Your Brain Memory Games	Android	Senior Games	All ages	Yes	Memory-focused cognitive stimulation
Abrain	Android	Abrain Labs	All ages	Unknown	Memory, attention, reaction, and math
Cognishape	iOS and Android	Cognishape	12 years or older	Yes	Attention, memory, creativity, and problem-solving
Q4 Active - Brain Health	iOS and Android	Genius Gyms LLC	4 years or older	Yes	Cognitive enhancement and decline prevention
Brain HQ	iOS and Android	Posit Science	4 years or older	Yes	Cognition, function, and brain plasticity
Alzlife	iOS and Android	Alzheimer's Light	17 years or older	Unknown	Sensory stimulation and cognitive games
Ginkgo Memory & Brain Training	iOS and Android	Ginkgo Academy	4 years or older	Unknown	Cognition and memory retention
Recover Brain	iOS and Android	Imagiration LLC	4 years or older	Yes	Memory, language, comprehension and executive
Lumosity	iOS and Android	Lumos Labs, Inc.	4 years or older	Unknown	Attention, flexibility, and problem-solving
Constant Therapy: Brain Rehab	iOS and Android	Constant Therapy, Inc	12 years or older	Yes	Language, memory, attention, reading, math, and comprehension
Memorie	iOS and Android	Neeuro Pte Ltd	4 years or older	Unknown	Attention, memory, decision-marking, and spatial perception

XSL•FO
RenderX

App name	Platform	Developer	Age group (years)	Involvement of health care	Focus (targeted areas of the app)
Focus - Train Your Brain	iOS and Android	Tellmewow	4 years or older	Unknown	Memory, coordination, and attention
Impulse	iOS and Android	Gmrd Apps Limited	4 years or older	Unknown	Memory, attention, and concentration
Peak	iOS and Android	Synaptic Labs	4 years or older	Yes	Brain assessment, memory, attention, and language
Neuronation	iOS and Android	Synaptikon Gmbh	4 years or older	Unknown	Memory and concentration

[a]MCI: mild cognitive impairment

MARS Evaluation of Included Apps

The 24 included apps were evaluated by two reviewers, with substantial interrater agreement (Quadratic Weighted κ=0.88, 95% CI 0.80 - 0.95), indicating a high level of agreement between the 2 reviewers. Based on the MARS evaluation, apps quality was assessed across four dimensions: engagement, functionality, aesthetics, and information, shown in Table 2. The overall mean score was 3.57 (SD 0.43), indicating that the apps were generally acceptable. Furthermore, 2 apps, Brain HQ and Peak, received scores above 4 and were rated as good, while 3 apps, Memory Trainer, Cognitive Skill Training, and Ginkgo Memory & Brain Training, scored below 3 and were considered insufficient. Among the 4 dimensions, functionality had the highest mean score, while engagement scored the lowest. In addition, the Spearman correlation showed no significant relationship between the overall mean score and star rating (R=−0.331, P=.18).

XSL•FO
RenderX

Table . The mean (SD) scores of each Mobile App Rating Scale dimension for included apps.

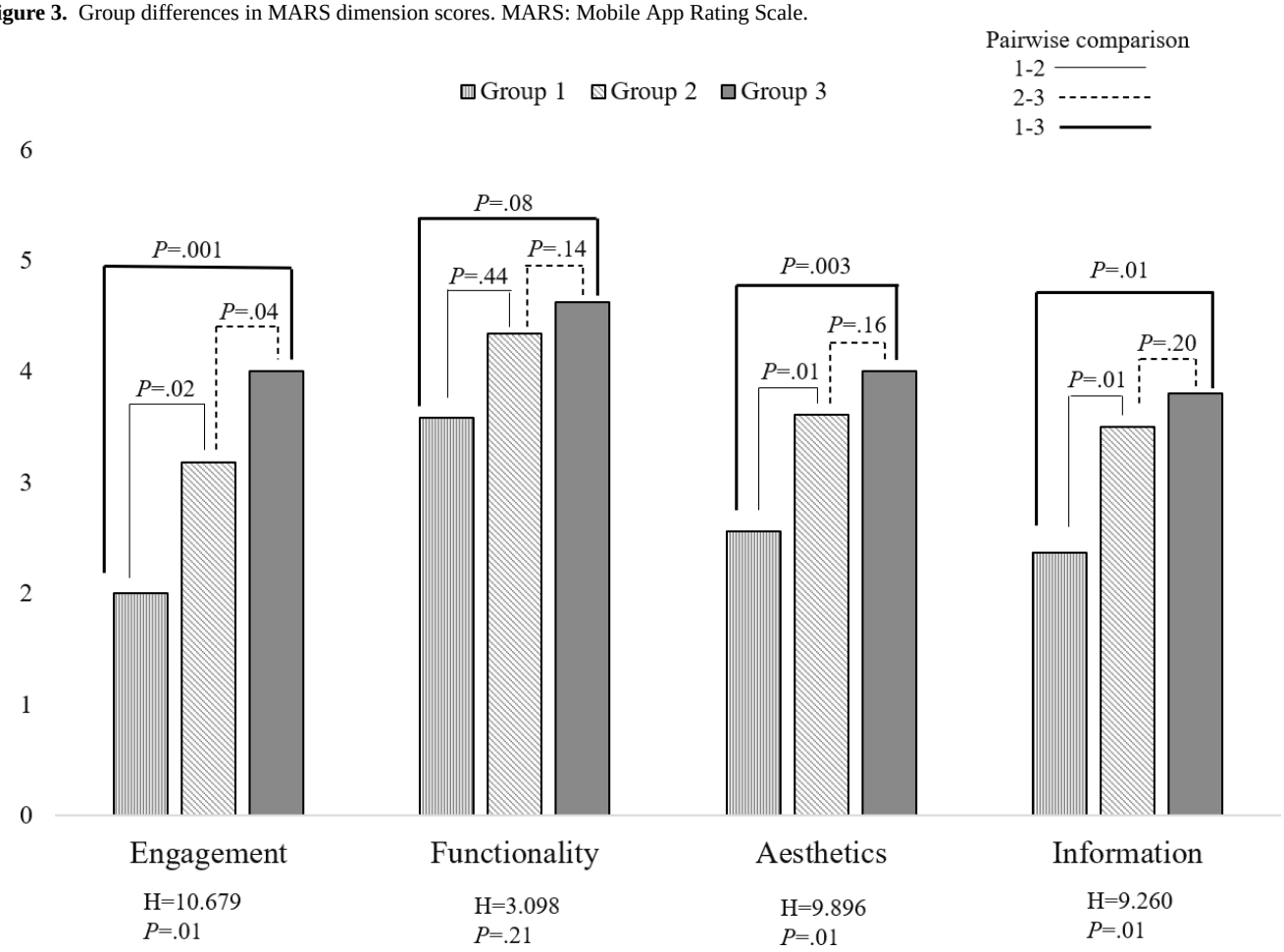
App name	Engagement	Functionality	Aesthetics	Information	Overall score	Star rating score
Memory Trainer	1.4 (0.49)	5 (0)	2 (0)	1.75 (0.43)	2.54 (1.44)	[a] —
Brave	3 (1.1)	4.75 (0.43)	4 (0.82)	4.2 (0.4)	3.99 (0.63)	—
Dementia Learning Game	3.2 (1.17)	4.25 (0.43)	3.67 (0.47)	3.5 (0.5)	3.65 (0.38)	—
Mindmate	3.2 (0.98)	4.5 (0.5)	3.67 (0.47)	3.8 (0.4)	3.79 (0.47)	5
Cognitive Skill Training	2.6 (1.02)	3.5 (0.5)	3 (0.82)	2.75 (0.43)	2.96 (0.34)	4.5
Reminding	2.6 (0.49)	4 (0)	3.33 (0.47)	3.5 (0.5)	3.36 (0.5)	—
Mindpal	3.6 (0.49)	4.5 (0.5)	3.67 (0.47)	3.6 (0.49)	3.84 (0.38)	4.5
Elevate	3.2 (0.4)	4.5 (0.5)	3.33 (0.47)	3.33 (0.47)	3.59 (0.53)	4.6
Neurobics	3.6 (0.49)	4.5 (0.5)	3.33 (0.47)	3.6 (0.49)	3.76 (0.44)	4
Train Your Brain Memory Games	3 (0.89)	4.25 (0.83)	4 (0)	3.25 (0.43)	3.63 (0.52)	4.5
Abrain	3.4 (0.49)	4.25 (0.43)	3.33 (0.47)	3.4 (0.49)	3.60 (0.38)	4.5
Cognishape	2.4 (0.49)	4.25 (0.43)	3 (0.82)	3.25 (0.43)	3.23 (0.67)	4.1
Q4 Active-Brain Health	2.6 (0.49)	4 (0.71)	3.33 (0.47)	3 (0)	3.23 (0.51)	—
Brain HQ	4 (0.63)	4.5 (0.5)	4 (0)	4 (0.58)	4.13 (0.22)	3.7
Alzlife	3.8 (1.17)	4 (0)	3.67 (0.47)	3.67 (0.47)	3.78 (0.14)	4.2
Ginkgo Memory& Brain Training	2 (0)	2.25 (0.83)	2.67 (0.47)	2.6 (0.49)	2.38 (0.27)	4.6
Recover Brain	3.6 (0.49)	4.5 (0.5)	4 (0)	3.8 (0.4)	3.98 (0.33)	4.2
Lumosity	3.2 (0.98)	4.25 (0.43)	4 (0)	3.43 (0.49)	3.72 (0.42)	4.2
Constant Therapy: Brain Rehab	3.4 (0.49)	4.5 (0.5)	4 (0)	3.6 (0.49)	3.88 (0.42)	4.15
Memorie	3 (1.26)	4.25 (0.43)	3.33 (0.47)	3.6 (0.49)	3.55 (0.46)	—
Focus - Train Your Brain	3.6 (0.49)	4.5 (0.5)	3.67 (0.47)	3.4 (0.49)	3.79 (0.42)	4.8
Impulse	3.4 (1.36)	4.25 (0.43)	4 (0)	3.4 (0.8)	3.76 (0.37)	4.5
Peak	4 (0.89)	4.75 (0.43)	4 (0)	3.6 (0.49)	4.09 (0.42)	4.3
Neuronation	2.6 (0.49)	4.5 (0.5)	3.33 (0.47)	3.2 (0.4)	3.41 (0.69)	4.55
Total (SD)	3.1 (0.61)	4.27 (0.51)	3.51 (0.49)	3.38 (0.48)	3.57 (0.43)	—

[a]Not applicable.

The included apps were categorized into three groups based on their overall mean scores: less than 3 (Group 1), between 3 and 4 (Group 2), and greater than 4 (Group 3). Detailed descriptions of these groupings are provided in Multimedia Appendix 2. The H value refers to the Kruskal–Wallis test statistic that indicates the degree of difference in rank sums among groups; higher values suggest greater group differences. Figure 3 shows the differences in MARS dimension scores between groups. No significant difference was observed in the functionality dimension (P=.21); however, differences in all other dimensions were statistically significant, with the engagement dimension showing the largest difference (P=.01). Pairwise comparisons revealed the following: (1) engagement: significant differences were observed between all groups, with score differences of 1.18 (Group 1 vs Group 2, P=.02), 0.82 (Group 2 vs Group 3, P=.04), and 2.00 (Group 1 vs Group 3, P=.001); (2) aesthetics: significant differences were found between Group 1 versus Group 2 (score difference=1.05, P=.01) and Group 1 versus Group 3 (score difference=1.44, P=.003), but not between Group 2 versus Group 3; (3) information: significant differences were noted between Group 1 versus Group 2 (score difference=1.13, P=.01) and Group 1 versus Group 3 (score difference=1.43, P=.01), with no difference between Group 2 versus Group 3.

XSL•FO
RenderX

Figure 3. Group differences in MARS dimension scores. MARS: Mobile App Rating Scale.



Bivariate Spearman correlation was calculated to test the relationship between the overall mean score and the star rating score. The results show that the overall mean score and the star rating score are not statistically significantly correlated (R=−0.331; P=.18).

Discussion

Principal Findings

Compared to previous studies that reviewed cognitive training apps primarily for the general population [18,19], this study is the first to systematically evaluate apps specifically designed for individuals with cognitive impairment, a population with distinct cognitive and usability needs. By applying the MARS to assess 24 apps across multiple quality dimensions, this study provides a comprehensive, standardized evaluation of app strengths and limitations. Overall, the apps demonstrated an acceptable level of quality. The functionality dimension received the highest scores, indicating strong usability and practicality, whereas the engagement dimension scored the lowest, reflecting limited user interactivity. Significant variability was observed across apps in engagement, aesthetics, and information quality. Furthermore, the findings highlight critical deficiencies in user-centered design and the limited involvement of health care professionals in app development. These insights identify key areas for improvement and offer important guidance for developers and health care providers seeking to optimize cognitive health interventions for individuals with cognitive impairment.

The quality of the 24 cognitive training apps was assessed using the MARS, with an average score of 3.57 (SD 0.43). Most apps (79.2%) were rated as acceptable, while only a small proportion (8.3%) received a good rating. Among the 4 MARS dimensions, functionality scored the highest, reflecting intuitive navigation, reliable performance, and device integration. In contrast, engagement received the lowest scores, followed by information quality, indicating a lack of features that sustain user interest and provide credible, evidence-based content. This quality pattern has also been reported in evaluations of other mHealth apps. Carrouel et al [37] evaluated mental health apps and found that although most demonstrated sound technical performance, many lacked meaningful user engagement and evidence-based content. Similarly, Martinon et al [38] reported that nutrition-related apps were generally usable but underperformed in terms of personalization and information quality. These similarities suggest a broader trend in mHealth app development, where technical functionality often exceeds performance in user-centered and content-related dimensions. As earlier research has pointed out, low engagement scores in health-related apps are often linked to a lack of personalized and adaptive content, which is particularly important for diverse populations such as older adults with cognitive impairments [39]. Furthermore, low engagement is often linked to poor user retention. In one study on diabetes management apps, low engagement scores were associated with higher dropout rates [40]. Therefore,

XSL•FO
RenderX

incorporating personalized features and adaptive content, such as gamification, user incentives, and social interaction tools, is essential to enhancing engagement and ensuring sustained use of cognitive training apps.

This study further revealed that group comparisons of the 4 MARS dimensions showed no significant differences in functionality; however, significant differences were observed in engagement, aesthetics, and information across app groups. The consistent functionality suggests that most cognitive training apps meet baseline standards for technical performance, reliability, and ease of use. In contrast, the observed differences in engagement, aesthetics, and information highlight key areas where apps diverge in their ability to meet user needs. Variations in engagement likely reflect differing levels of investment in user-focused features [41]. Differences in aesthetics and information may indicate disparities in design quality and content accuracy, with higher-scoring apps often incorporating professional input and evidence-based content [42,43]. Addressing these gaps by enhancing engagement features, improving design, and involving professionals during development could enhance app effectiveness and user satisfaction. Notably, limited involvement of health care professionals during app development for older adults with cognitive impairment, as observed in this study, is consistent with previous findings [27,44]. Only 8 of the included apps (33%) explicitly mentioned health care professional involvement. This reflects a broader challenge in the mHealth field [45], where insufficient engagement from both health care professionals and patients often results in apps that fail to provide reliable, evidence-based information or meet user needs effectively. Evidence indicates that the quality and trustworthiness of mHealth apps are strongly linked to the involvement of health care professionals, who ensure clinical accuracy and adherence to evidence-based practices [46]. Similarly, excluding patients from the development process frequently leads to features misaligned with real-world needs, reducing engagement and limiting the apps' effectiveness in supporting health management [47].

The findings of this study have important implications for both app developers and health care providers. Several key deficiencies were identified across many of the reviewed apps, including user engagement, information quality, aesthetic design, and the limited involvement of health care professionals, which underscores the need for closer collaboration with health care professionals to ensure content accuracy, adherence to medical guidelines, and alignment with real-world health care needs [46,48]. Similarly, incorporating patient feedback during the development process can help tailor cognitive training apps to better meet the cognitive, physical, and sensory needs of older adults with cognitive impairment, thereby improving accessibility and engagement. Furthermore, the results of this study also underscore the importance of user interface design, balancing aesthetic appeal with intuitive functionality can significantly enhance usability, particularly in cognitively impaired populations. On the technical side, integrating technical features such as adaptive difficulty levels and real-time feedback could enhance personalization and maintain user motivation over time. Overall, these insights underscore the value of a cross-disciplinary approach in which developers, clinicians, patients, designers, and technical teams collaborate to produce mHealth solutions that are user-centered, clinically credible, and sustainable in real-world use.

This study identified key limitations in existing cognitive training apps for individuals with cognitive impairments, particularly in visual presentation, such as dense text blocks and a lack of instructional images and videos, which are particularly challenging for this population. As cognitive impairments often lead to age-related visual degradation, including decreased visual acuity, reduced field of vision, and diminished color recognition, these issues exacerbate difficulties in using digital technologies, especially for cognitive training [49]. Previous studies indicated that cognitive impairments also introduce unique visual challenges, such as slower visual processing speed and difficulties in spatial perception, which hinder the effective use of apps [50,51]. To improve the usability and effectiveness of cognitive training apps for older adults with cognitive impairment, it is essential to incorporate user-centered design strategies that accommodate both cognitive and sensory limitations. Interface design should emphasize simplicity and clarity through the use of high-contrast visuals, large fonts and icons, linear navigation structures, and minimal interaction steps to reduce operational complexity and cognitive load. Readability can be further enhanced by optimizing text presentation with adjustable font sizes, low content density, and voice-assisted reading or playback features to support information processing and minimize visual fatigue. Furthermore, incorporating multimodal support, such as auditory prompts, visual cues, haptic feedback, and the inclusion of icons, images, and instructional videos can enhance task comprehension and user engagement. In particular, the integration of auditory and visual stimuli has been shown to improve performance and reduce cognitive burden [52]. To further support visual processing and maintain attention, optimizing visual presentation (such as enhancing contrast and minimizing screen glare) is also important, as higher contrast improves readability while glare reduction contributes to visual comfort and uninterrupted use [53,54]. In addition, personalized cognitive training modules that adapt to users' cognitive profiles, including their impairment levels, response speed, learning preferences, and progress over time, may help maintain motivation and optimize training outcomes. Finally, enhancing feedback mechanisms is also critical. Real-time audio cues, digital guidance, and multisensory reinforcement, including auditory and tactile feedback, can promote sustained attention and support memory retention [55]. These strategies will facilitate the improvement of usability and effectiveness in cognitive training apps for older adults with cognitive impairments, enhancing engagement and supporting better cognitive performance.

Limitations

Several limitations should be considered in this study. First, data were sourced primarily from major app stores (Google Play Store and Apple App Store), which may have excluded apps available on regional or specialized platforms, limiting the comprehensiveness of the findings. Second, although the MARS provided a standardized evaluation framework, inherent

XSL•FO
RenderX

subjectivity in the scoring process may have affected consistency, particularly in the evaluation of visual design. Third, the study focused exclusively on apps identified using specific cognitive training-related keywords, resulting in a limited sample size and potentially reducing the generalizability of the findings. Fourth, the dynamic nature of app stores posed challenges to the study's timeliness, as some apps may have been removed or updated after data collection. Finally, while MARS is a robust tool, it does not comprehensively address data privacy concerns or sustained app usage outcomes. Future research should address these limitations by expanding data sources, incorporating user experience and longitudinal data, and exploring additional domains of app quality.

Conclusions

This study evaluated the quality of cognitive training mobile apps for older adults with cognitive impairment using the MARS. The findings highlight the promise of these apps as accessible tools for cognitive health management while revealing notable deficiencies in quality, usability, and evidence-based functionality. As mHealth technologies evolve, rigorous and standardized evaluation frameworks and longitudinal studies are needed to validate their efficacy and ensure safety in real-world settings. Such efforts are critical to fully realize the potential of these technologies in supporting cognitive health in aging populations.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 72174061, No.71704053); China Scholarship Council foundation (NO. 202308330251), and Health Science and Technology Project of Zhejiang Provincial Health Commission (2022KY370, 2024KY1662).

Authors' Contributions

LW contributed to conceptualization, methodology, and writing-original draft. JP and CD were involved in conceptualization, methodology, and data curation. AG and AH were responsible for supervision and validation. HT and XW managed writing-review and editing. CZ handed conceptualization and methodology. LW was involved in conceptualization, methodology, validation, writing-review and editing, and funding acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search Strategy Process.
[DOCX File, 21 KB - mhealth_v13i1e69637_app1.docx]

Multimedia Appendix 2
Grouping.
[XLSX File, 12 KB - mhealth_v13i1e69637_app2.xlsx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 277 KB - mhealth_v13i1e69637_app3.docx]

References

1. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 8;396(10248):413-446. [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
2. Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA research framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement 2018 Apr;14(4):535-562. [doi: 10.1016/j.jalz.2018.02.018] [Medline: 29653606]
3. Chapman SB, Aslan S, Spence JS, et al. Neural mechanisms of brain plasticity with complex cognitive training in healthy seniors. Cereb Cortex 2015 Feb;25(2):396-405. [doi: 10.1093/cercor/bht234] [Medline: 23985135]
4. Gates NJ, Vernooij RW, Di Nisio M, et al. Computerised cognitive training for preventing dementia in people with mild cognitive impairment. Cochrane Database Syst Rev 2019 Mar 13;3(3):CD012279. [doi: 10.1002/14651858.CD012279.pub2] [Medline: 30864747]
5. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLOS Med 2014 Nov;11(11):e1001756. [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
6. Bateman DR, Srinivas B, Emmett TW, et al. Categorizing health outcomes and efficacy of mHealth apps for persons with cognitive impairment: a systematic review. J Med Internet Res 2017 Aug 30;19(8):e301. [doi: 10.2196/jmir.7814] [Medline: 28855146]

XSL•FO
RenderX

7. Maggio MG, Luca A, Calabrò RS, et al. Can mobile health apps with smartphones and tablets be the new frontier of cognitive rehabilitation in older individuals? A narrative review of a growing field. Neurol Sci 2024 Jan;45(1):37-45. [doi: 10.1007/s10072-023-07045-8] [Medline: 37702829]
8. Shojaei F, Bergvist ES, Shih PC. Exploring the impact of digital art therapy on people with dementia: a framework and research-based discussion. In: Shaikh A, Alghamdi A, Tan Q, El Emary IMM, editors. Advances in Emerging Information and Communication Technology ICIEICT 2023 Signals and Communication Technology: Springer; 2024:133-143. [doi: 10.1007/978-3-031-53237-5_10]
9. Zhu D, Al Mahmud A, Liu W, et al. Digital storytelling for people with cognitive impairment using available mobile apps: systematic search in app stores and content analysis. JMIR Aging 2024 Oct 24;7:e64525. [doi: 10.2196/64525] [Medline: 39446478]
10. Shojaei F, Osorio Torres J, Shih PC. Exploring the integration of technology in art therapy: insights from interviews with art therapists. Art Ther (Alex) 2024:1-7. [doi: 10.1080/07421656.2024.2383826]
11. World Health Organization. Seventy-first World Health Assembly: Provisional agenda item 124 (A71/20). 2018. URL: https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_20-en.pdf [accessed 2025-06-03]
12. Silva BMC, Rodrigues J, de la Torre Díez I, et al. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272. [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
13. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. Am J Geriatr Psychiatry 2015 Apr;23(4):335-359. [doi: 10.1016/j.jagp.2014.04.009] [Medline: 24998488]
14. Chan ATC, Ip RTF, Tran JYS, et al. Computerized cognitive training for memory functions in mild cognitive impairment or dementia: a systematic review and meta-analysis. NPJ Digit Med 2024 Jan 3;7(1):1. [doi: 10.1038/s41746-023-00987-5] [Medline: 38172429]
15. Chan JYC, Kwong JSW, Wong A, et al. Comparison of computerized and paper-and-pencil memory tests in detection of mild cognitive impairment and dementia: a systematic review and meta-analysis of diagnostic studies. J Am Med Dir Assoc 2018 Sep;19(9):748-756. [doi: 10.1016/j.jamda.2018.05.010] [Medline: 29921507]
16. Vermeir JF, White MJ, Johnson D, et al. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644. [doi: 10.2196/18644] [Medline: 32773374]
17. Koivisto J, Malik A. Gamification for older adults: a systematic literature review. Gerontologist 2021 Sep 13;61(7):e360-e372. [doi: 10.1093/geront/gnaa047] [Medline: 32530026]
18. Bang M, Jang CW, Kim HS, et al. Mobile applications for cognitive training: content analysis and quality review. Internet Interv 2023 Sep;33:100632. [doi: 10.1016/j.invent.2023.100632] [Medline: 37312799]
19. Cha SM. Mobile application applied for cognitive rehabilitation: a systematic review. Life (Basel) 2024 Jul 18;14(7):891. [doi: 10.3390/life14070891] [Medline: 39063644]
20. Werner NE, Brown JC, Loganathar P, et al. Quality of mobile apps for care partners of people with Alzheimer disease and related dementias: mobile app rating scale evaluation. JMIR Mhealth Uhealth 2022 Mar 29;10(3):e33863. [doi: 10.2196/33863] [Medline: 35348467]
21. Silva AF, Silva RM, Murawska-Ciałowicz E, et al. Cognitive training with older adults using smartphone and web-based applications: a scoping review. J Prev Alzheimers Dis 2024;11(3):693-700. [doi: 10.14283/jpad.2024.17] [Medline: 38706285]
22. Stoyanov SR, Hides L, Kavanagh DJ, et al. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
23. Baumel A, Faber K, Mathur N, et al. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82. [doi: 10.2196/jmir.7270] [Medline: 28325712]
24. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor and Francis; 1996:189-194. [doi: 10.1201/9781498710411]
25. Stoyanov SR, Hides L, Kavanagh DJ, et al. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
26. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, et al. Assessing apps for patients with genitourinary tumors using the Mobile Application Rating Scale (MARS): systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e17609. [doi: 10.2196/17609] [Medline: 32706737]
27. Gerner M, Vuillerme N, Aubourg T, et al. Review and analysis of German mobile apps for inflammatory bowel disease management using the Mobile Application Rating Scale: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2022 May 3;10(5):e31102. [doi: 10.2196/31102] [Medline: 35503246]
28. Lull C, von Ahnen JA, Gross G, et al. German mobile apps for patients with psoriasis: systematic search and evaluation. JMIR Mhealth Uhealth 2022 May 26;10(5):e34017. [doi: 10.2196/34017] [Medline: 35617014]
29. Diaz-Skeete YM, McQuaid D, Akinosun AS, et al. Analysis of apps with a medication list functionality for older adults with heart failure using the Mobile App Rating Scale and the IMS Institute for Healthcare Informatics Functionality score: evaluation study. JMIR Mhealth Uhealth 2021 Nov 2;9(11):e30674. [doi: 10.2196/30674] [Medline: 34726613]

XSL•FO
RenderX

30. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
31. Narrillos-Moraza Á, Gómez-Martínez-Sagrera P, Amor-García M, et al. Mobile apps for hematological conditions: review and content analysis using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2022 Feb 16;10(2):e32826. [doi: 10.2196/32826] [Medline: 35171109]
32. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
33. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
34. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718. [doi: 10.2196/10718] [Medline: 30361196]
35. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of top-rated mobile apps for depression: systematic search and evaluation. JMIR Ment Health 2020 Jan 24;7(1):e15321. [doi: 10.2196/15321] [Medline: 32012079]
36. Duarte-Díaz A, Perestelo-Pérez L, Gelabert E, et al. Efficacy, safety, and evaluation criteria of mHealth interventions for depression: systematic review. JMIR Ment Health 2023 Sep 27;10:e46877. [doi: 10.2196/46877] [Medline: 37756042]
37. Carrouel F, du Sartz de Vigneulles B, Bourgeois D, et al. Mental health mobile apps in the French App Store: assessment study of functionality and quality. JMIR Mhealth Uhealth 2022 Oct 12;10(10):e41282. [doi: 10.2196/41282] [Medline: 36223178]
38. Martinon P, Saliasi I, Bourgeois D, et al. Nutrition-related mobile apps in the French app stores: assessment of functionality and quality. JMIR Mhealth Uhealth 2022 Mar 14;10(3):e35879. [doi: 10.2196/35879] [Medline: 35285817]
39. Grua EM, Sanctis M, Malavolta I, et al. Social sustainability in the e-health domain via personalized and self-adaptive mobile apps. In: Calero C, editor. Software Sustainability: Springer Nature; 2021:301-328. [doi: 10.1007/978-3-030-69970-3_13]
40. Priesterroth L, Grammes J, Holtz K, et al. Gamification and behavior change techniques in diabetes self-management apps. J Diabetes Sci Technol 2019 Sep;13(5):954-958. [doi: 10.1177/1932296818822998] [Medline: 30762420]
41. Hibbard JH, Stockard J, Mahoney ER, et al. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026. [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
42. Kuck N, Dietel FA, Nohr L, et al. A smartphone app for the prevention and early intervention of body dysmorphic disorder: development and evaluation of the content, usability, and aesthetics. Internet Interv 2022 Apr;28:100521. [doi: 10.1016/j.invent.2022.100521] [Medline: 35281703]
43. Ke Z, Qian W, Wang N, et al. Improve the satisfaction of medical staff on the use of home nursing mobile APP by using a hybrid multi-standard decision model. BMC Nurs 2024 May 9;23(1):302. [doi: 10.1186/s12912-024-01918-9] [Medline: 38724959]
44. Hundert AS, Huguet A, McGrath PJ, et al. Commercially available mobile phone headache diary apps: a systematic review. JMIR Mhealth Uhealth 2014 Aug 19;2(3):e36. [doi: 10.2196/mhealth.3452] [Medline: 25138438]
45. Chudyk AM, Ragheb S, Kent D, et al. Patient engagement in the design of a mobile health app that supports enhanced recovery protocols for cardiac surgery: development study. JMIR Perioper Med 2021 Nov 30;4(2):e26597. [doi: 10.2196/26597] [Medline: 34851299]
46. Hamilton AD, Brady RRW. Medical professional involvement in smartphone “apps” in dermatology. Br J Dermatol 2012 Jul;167(1):220-221. [doi: 10.1111/j.1365-2133.2012.10844.x] [Medline: 22283748]
47. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022 Sep 9;9(3):e35693. [doi: 10.2196/35693] [Medline: 36083628]
48. Visvanathan A, Hamilton A, Brady RRW. Smartphone apps in microbiology--is better regulation required? Clin Microbiol Infect 2012 Jul;18(7):E218-E220. [doi: 10.1111/j.1469-0691.2012.03892.x] [Medline: 22563840]
49. Colombo G, Minta K, Grübel J, et al. Detecting cognitive impairment through an age-friendly serious game: The development and usability of the Spatial Performance Assessment for Cognitive Evaluation (SPACE). Comput Human Behav 2024 Nov;160:108349. [doi: 10.1016/j.chb.2024.108349]
50. Vergani L, Marton G, Pizzoli SFM, et al. Training cognitive functions using mobile apps in breast cancer patients: systematic review. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e10855. [doi: 10.2196/10855] [Medline: 30888326]
51. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
52. Iarlori S, Monteriú A, Perpetuini D, et al. Editorial: Affective computing and mental workload assessment to enhance human-machine interaction. Front Neuroergon 2024;5:1412744. [doi: 10.3389/fnrgo.2024.1412744] [Medline: 38867943]
53. Bright K, Egger (is-design GmbH) V. Using visual contrast for effective, inclusive environments. IDJ 2008 Dec 8;16(3):178-189 [FREE Full text] [doi: 10.1075/idj.16.3.02bri]

XSL•FO
RenderX

54. Lee CC, Chen CH, Chien WC, Wu FG. Improving pedestrians' navigation safety at night by enhancing legibility of foreground and background information on the display. Int J Ind Ergon 2023 Mar;94:103383. [doi: 10.1016/j.ergon.2022.103383]
55. Geurts JAP, van Vugt TAG, Arts JJC. Use of contemporary biomaterials in chronic osteomyelitis treatment: clinical lessons learned and literature review. J Orthop Res 2021 Feb;39(2):258-264. [doi: 10.1002/jor.24896] [Medline: 33098587]

Abbreviations

ENLIGHT: Evaluation Tool for Mobile and Web-Based eHealth Interventions
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO : International Prospective Register of Systematic Reviews
SUS: System Usability Scale
uMARS: User Version of the Mobile App Rating Scale

Edited by NM Ali; submitted 04.12.24; peer-reviewed by A Bhattacharya, F Shojaei, F Carrouel; revised version received 05.05.25; accepted 12.05.25; published 04.07.25.

Please cite as:
Wu L, Pan J, Dou C, Gu A, Huang A, Tao H, Wang X, Zhang C, Wang L
Cognitive Training Mobile Apps for Older Adults With Cognitive Impairment: App Store Search and Quality Evaluation
JMIR Mhealth Uhealth 2025;13:e69637
URL: https://mhealth.jmir.org/2025/1/e69637
doi:10.2196/69637

© Leyi Wu, Jiajuan Pan, Chuwen Dou, An Gu, An Huang, Hong Tao, Xiaoyan Wang, Chen Zhang, Lina Wang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Grouping Digital Health Apps Based on Their Quality and User Ratings Using K-Medoids Clustering: Cross-Sectional Study

Maciej Marek Zych[1], PhD; Raymond Bond[1], BSc, PhD; Maurice Mulvenna[1], BSc, MPhil, PhD; Lu Bai[2], BEng, PhD; Jorge Martinez-Carracedo[1], BSc, MSc, PhD; Simon Leigh[3], BSc, MSc, PhD

[1]School of Computing, Ulster University, 2-24 York Street, Belfast, United Kingdom

[2]School of Electronics, Electrical Engineering and Computer Science, Queen's University, Belfast, United Kingdom

[3]Organisation for the Review of Care and Health Apps (ORCHA), Daresbury, United Kingdom

Corresponding Author:
Maciej Marek Zych, PhD
School of Computing, Ulster University, 2-24 York Street, Belfast, United Kingdom

Abstract

Background: Digital health apps allow for proactive rather than reactive health care and have the potential to take the pressure off health care providers. With over 350,000 digital health apps available on the app stores today, those apps need to be of sufficient quality to be safe to use. Discovering the typology of digital health apps regarding professional and clinical assurance (PCA), user experience (UX), data privacy (DP), and user ratings may help in determining the areas where digital health apps can improve.

Objective: This study has two objectives: (1) discover the types (clusters) of digital health apps with regards to their quality (scores) across 3 domains (their PCA, UX, and DP) and user ratings and (2) determine whether the National Institute for Health and Care Excellence (NICE) Evidence Standard Framework's (ESF's) tier, target users of the digital health apps, categories, or features have any association with this typology.

Methods: Data were obtained from 1402 digital health app assessments conducted using the Organisation for the Review of Care and Health Apps Baseline Review (OBR), evaluating PCA, UX, and DP. K-medoids clustering identified app typologies, with the optimal number of clusters determined using the elbow method. The Shapiro-Wilk test assessed normality of user ratings and OBR scores. Nonparametric Wilcoxon rank sum tests compared cluster differences in these metrics. Post hoc analysis examined the distribution of NICE ESF tiers, target users, categories, and features across clusters, using Fisher exact test with Bonferroni correction. Effect sizes were calculated using Cohen *w*.

Results: A total of four distinct app clusters emerged: (1) apps with poor user ratings (220/1402, 15.7%), (2) apps with poor PCA and DP scores (252/1402, 18%), (3) apps with poor PCA scores (415/1402, 29.6%), and (4) higher quality apps with high user ratings and OBR scores (515/1402, 36.7%). While some statistically significant associations were found between clusters and NICE ESF tiers (2/3), target users (0/14), categories (4/33), and features (6/19), all had small effect sizes (Cohen $w<0.3$). The strongest associations were for the "Service Signposting" feature (Cohen $w=0.24$) and NICE ESF tier B (Cohen $w=0.19$).

Conclusions: The largest cluster comprised high-quality apps with strong user ratings and OBR scores (515/1402, 36.7%). A significant proportion (415/1402, 29.6%) performed poorly in PCA despite performing well in other domains. Notably, user ratings did not consistently align with PCA scores; some apps scored highly with users but poorly in PCA and DP. The 4-cluster typology underscores areas needing improvement, particularly PCA. Findings suggest limited association between the examined app characteristics and quality clusters, indicating a need for further investigation into what factors truly influence app quality.

(JMIR Mhealth Uhealth 2025;13:e57279) doi:10.2196/57279

KEYWORDS

digital health; ORCHA assessment; quality assessment; cluster analysis; digital health app; app; user; clinical assurance; user experience; data privacy; mHealth; Organisation for the Review of Care and Health Apps; mobile health

Introduction

Digital health apps allow for proactive rather than reactive health care and have the potential to take the pressure off health care providers. Furthermore, the integration of digital health apps into clinical practice was shown to reduce the clinician's workload in some cases [1,2], as well as increase workload in others [3]. According to an umbrella review from 2023 [4], digital health apps for noncommunicable diseases are effective in improving health outcomes. For mental health apps, it was reported that advances in digital health are changing mental health care in multiple ways, such as making mental health care more accessible [5]. The public appears to have a positive attitude regarding the use of digital health apps. According to

XSL•FO
RenderX

the Organisation for the Review of Care and Health Apps (ORCHA) website [6], "68% of people surveyed agreed, or strongly agreed, that to take pressure off our health care system, it is vital we all look at new ways to manage our health, including using high-quality health apps" [7]. A meta-ethnographic review from 2019 [8], with the aim to explore the public perception of digital health apps, found that users viewed digital health apps as useful complementary tools. However, there are still barriers regarding their use and quality that need to be resolved.

A study from 2020 [9] indicates that the evidence whereby digital health apps improve patient outcomes is scarce. Continued use of digital health apps (ie, user retention) after their installation is also a concern. A systematic review from 2020 [10] found a high dropout rate (47.8% when adjusting for publication bias) among digital health apps for depressive symptoms. A study from 2019 [11], focused on the "continued use" of mobile health apps, found that for users to continue using health apps, users must have a positive view of the app's user experience (UX) and be persistent at achieving their health goals. The effectiveness of digital health apps in the management of various chronic diseases remains unclear [12]. There are also barriers regarding the apps' integration into clinical practice, such as patient or provider support [13].

A mixed methods study from 2019 [14] found that users are concerned with the security and privacy of digital health apps. Furthermore, the current literature suggests that digital health apps could be improved regarding their quality [5,15-21]. The limitation of these studies is that they were often conducted using a small sample of data and the analyzed apps were for a specific category or health condition, making their findings category or condition specific.

In our study, collaborating with the ORCHA, we included digital health apps from 33 different categories allowing us to get a "big picture" of their performance regarding quality. ORCHA used their assessment tool called ORCHA Baseline Review (OBR) [22] to quality assess over 1400 digital health apps. The tool assesses digital health apps regarding their UX, data privacy (DP), and professional and clinical assurance (PCA). The scores in the 3 assessment areas are then combined into an overall ORCHA score out of 100. The aim of this study is to uncover similarities and differences in traits among digital health apps regarding characteristics related to their quality (as indicated by OBR's PCA, UX, and DP scores) and user ratings. We accomplish this aim with two objectives: (1) to discover the types (clusters) of digital health apps with regards to their quality (scores) across three domains (their PCA, UX, and DP) and user ratings and (2) to determine whether the National Institute for Health and Care Excellence (NICE) Evidence Standard Framework's (ESF) tier, target users of the digital health apps, categories, or features have any association with this typology.

Uncovering the similarities and differences in the traits of digital health apps via k-medoids cluster analysis can indicate areas where digital health apps can improve regarding their quality assessment and inform the state of digital health apps today. It also allows us to see how different categories of digital health apps are impacted by the same drawbacks. Similar work, using different methods, was done to explore the traits of mental health apps [21]. In this cross-sectional study, researchers examined 578 mental health apps regarding features that they offer and found that the most common features are psychoeducation, goal tracking, and mindfulness. The study also found that DP is not associated with user ratings and is weakly associated with the number of downloads. A study from 2019 [23] analyzed the most popular mental health apps with the aim to understand how their attributes relate to user ratings, app quality, and classification by the World Health Organization (WHO) health app classification framework [24]. However, due to the heterogeneity of the apps, they were unable to define a core set of features that would accurately assess app quality. In our study, as a post hoc analysis, we examined each cluster regarding each digital health app's NICE ESF tier [25], target users, categories, and features. This is done to examine whether there are any other relationships between a cluster or type of app and its NICE ESF tier, category, target users, or features of a digital health app. This study extends our "work in progress" paper from 2023 [26], applying more rigor and additional analyses. In our previous work, we conducted an analysis of digital health apps' quality across different health care categories [27].

Methods

The Dataset

ORCHA has conducted digital health apps' assessment as part of their standard workflow resulting in the original dataset of 2127 digital health apps. The OBR version 6 evolved from earlier versions of the OBR during the height of the COVID-19 pandemic. Originally, version 6 was created as a more stringent version of the OBR so that ORCHA could recommend the most compliant digital health apps to members of the UK population with confidence. ORCHA tested version 6 on a selection of highly compliant digital health apps (as determined by previous versions of the OBR). This set of 30 digital health apps served as the pilot group, with the subsequent 2097 apps being assessed with ORCHA's typical assessment approach of categorizing apps into categories, ordering by number of downloads, and assessing the most downloaded apps in each category, followed by the second, and so forth. The apps have been selected from the original 2127 app assessments by removing apps without user ratings, second assessments of the same app, apps that were not digital health apps, and taking the mean results of app assessments with both Android and iOS versions of the app. Resulting in the dataset of 1402 digital health apps that were used in this study (see Results section).

The dataset consists of digital health app assessments that were published between January 18, 2021, and January 6, 2022. All digital health apps were characterized using 14 target user groups (eg, adult, carer, etc), 33 categories (eg, allergy, blood, etc) and 19 features (eg, behavioral change techniques, condition management, etc), as well as to one of the three NICE ESF tiers [25]: (1) Tier A (n=9), "Digital health technologies intended to save costs or release staff time, no direct patient, health or care outcomes;" (2) Tier B (n=1018), "Digital health technologies for helping citizens and patients to manage their own health and wellness;" and Tier C (n=375), "Digital health technologies for

XSL•FO
RenderX

treating and diagnosing medical conditions, or guiding care choices. Includes digital health technologies with direct health outcomes, and those that are likely to be regulated medical devices." Details on how the digital health apps have been selected can be found in the Results section.

Assessment With ORCHA Baseline Review

A digital health app with satisfactory PCA, UX, and DP could lead to greater digital health app adoption, confident app recommendations by clinicians, and prevent risks that could arise from using the app. OBR assesses these 3 aspects of digital health apps. This is because assessing PCA ensures that the app is evidence-based and reduces the risks of an app having harmful effects on users' health. PCA assessment includes questions such as: "Is the developer or publisher registered with the Care Quality Commission (CQC)?" and "Is the app a medical device?" Assessing UX ensures that the app is useful, usable, and desirable by the user. UX assessment includes questions such as: "Is there a statement about user feedback during design/ development?" and "Is there any evidence of user involvement in testing?." Assessing DP ensures that users' rights and data are protected and not exploited in any way (eg, financially). DP assessment includes questions such as: "Is there a Privacy Policy clearly available via the App/ Web App/ Website?" and "Is the policy made available when the user is signing up to the service?" Digital health apps in the dataset are assigned an ORCHA score (range 0 - 100). ORCHA score is based on the 3 assessment areas (UX, PCA, and DP), each of which has their own score out of 100, where higher value indicates better compliance.

Each app was assessed by 2 ORCHA reviewers using OBR. In the case of a dispute, a third reviewer would be involved to resolve it. The reviewers assessed digital health apps using OBR assessment to get the following outcomes: PCA (indicated by PCA score), DP (indicated by DP score), UX (indicated by UX score), and an overall ORCHA score. In the dataset, 5-star user ratings at the app store were recorded at the time of assessment. An ORCHA threshold score of 65 out of 100 is a National Health Service (NHS) accepted cut-off point that indicates compliance with best practice standards for digital health apps, meaning that the digital health app may be used or recommended by NHS staff. ORCHA threshold is the point at which excess risks are avoided, that is, an app cannot possibly score above 65 while having no privacy policy, having no relevant evidence, or being a medical device which is not certified. The score of 65 was established with NHS partners in 2020 and has remained since.

An ORCHA score of 65 is an initial score for all the digital health apps being assessed in all assessment areas (UX, PCA, and DP). Meaning that the initial score at the beginning of the assessment is 65 for each assessment area and overall ORCHA score. Based on answers to assessment questions, this score is altered via value and risk points. Value points increase the score, and risk points reduce the score. This process changes the initial score of 65 for each assessment area and is then combined to give an overall ORCHA score. For example, for the apps that store personal and sensitive information, value points are assigned to such an app if they make their privacy policy immediately available when the user first uses the app. Risk points are assigned if a privacy policy is not clearly available via the app. The amount of value and risk points assigned per question varies based on the NICE ESF tier that was assigned to an app. If no value or risk points were assigned during the assessment, then the ORCHA score remains 65 [22]. Furthermore, to receive full points for appropriate evidence for its NICE ESF tier, a tier B digital health app (depending on its exact functionality) may only require a user benefits statement (eg, based on pilot results) and validation of the provided information by experts or references, while a tier C digital health app will likely require a full-scale observational study or randomized controlled trial to meet the same evidence threshold. These differences in evidence requirements were introduced by the NICE ESF and adopted with slight amendments by the ORCHA assessment. This was done to ensure that standards are realistic and achievable for digital health app companies, without placing an undue burden on developers of low-risk apps, while at the same time setting expectations sufficiently high for high-risk apps.

In summary, digital health apps with an ORCHA score of <65 are not recommended to be used by the public. Having a score <65 on an assessment area (UX, PCA, or DP) indicates that improvements should be made in that area by developers. However, it may still be recommended for use if other areas (especially PCA, due to its reliance on evidence) are >65.

Statistical Analysis

R studio and the R programming language was used to conduct the analysis and produce the graphs. Spearman correlation among the OBR scores (ORCHA, PCA, UX, and DP) and user ratings was calculated to examine the relationship among variables. Median (IQR) was calculated for the OBR scores and user ratings for reference. The Shapiro-Wilk test was used to check if the OBR scores or user ratings were normally distributed. K-medoids clustering was used to group together (classify) health apps based on their user ratings and PCA, UX, and DP scores. Due to the scores not being normally distributed, K-medoids clustering was used to allow for a more fitting interpretation of clustering results, not skewed by outliers in the cluster. K-medoids clustering is a partitioning technique based on medoids. A medoid is the most centrally located point in a cluster, with the minimum sum of distances to other points. Clustering was used as it simplifies the grouping together (classification) of apps. We have chosen clustering to discover the types of apps because this approach is objectively data-driven and does not result in bias and error from manual grouping (classification). A widely used elbow method was used to determine the optimal number of clusters for the K-medoids cluster analysis. Following the results of the Shapiro-Wilk tests (indicating that user ratings and PCA, UX, and DP scores are not normally distributed), the unpaired 2-samples Wilcoxon rank sum test (also known as Mann-Whitney *U* test) was used to compare the corresponding user ratings and the scores among clusters, to check for statistical significance. The Wilcoxon test was used due to the data not being normally distributed. A *P* value less than .05 was considered statistically significant. Bonferroni corrected alpha

XSL•FO
RenderX

value was used when multiple hypothesis testing was conducted on the same data.

The following steps have been taken when conducting the analysis:

- Step 1, outcomes under investigation: digital health apps' user ratings, PCA, UX, and DP scores will be investigated from the assessment data in the dataset. Those values were obtained by assessing digital health apps with OBR.
- Step 2, outcomes assessment: K-medoids clustering will be used on the user ratings, PCA, UX, and DP scores from the dataset.
- Step 3, outcomes interpretation: each cluster will have a medoid value where user ratings range from 1 to 5, where higher numbers indicate better rating. PCA, UX, and DP scores range from 0 to 100 where higher numbers indicate better compliance with assessment area.
- Step 4, variables generated: clusters will be investigated and labeled based on the medoid value of user ratings, PCA, UX, and DP scores in each of the clusters.

Post hoc analysis will be conducted by counting the prevalence of each target user group, categories, and features in each cluster. Fisher exact test will be used to determine whether the difference in proportion was statistically significant among the clusters. When the Fisher exact test has a statistically significant Bonferroni corrected alpha, the effect size will be calculated using Cohen *w*. A widely used "rule of thumb" interpretation of Cohen *w* for the effect size is as follows: small (Cohen *w*=.10), medium (Cohen *w*=.30), and large (Cohen *w*=.50) [28]. However, this interpretation was criticized, and a new interpretation was proposed by Funder and Ozer [29]: very small (Cohen *w*=.05), small (Cohen *w*=.10), medium (Cohen *w*=.20), large (Cohen *w*=.30), and very large (Cohen *w*≥.40). Both interpretations will be considered in this study.

Ethical Considerations

This secondary data analysis study gained ethical approval by Ulster University (ethics filter committee, Faculty of Computing, Engineering, and the Built Environment; project CEBE_RE-22-002). The data used in this study has been anonymized. The developers under consideration provided implicit consent for use of their data for research purposes. All reviews, unless explicitly asked to be removed by the developer, are covered as suitable for research in ORCHA's privacy policy [30].

Results

Principal Results

Figure 1 shows how the 1402 digital health apps have been selected from the original 2127 digital health apps assessed by ORCHA. To avoid inclusion of the same app twice in the analysis, the mean of the scores (ORCHA, PCA, UX, and DP) and user ratings was taken from the 2 versions (iOS and Android) and included in the analysis as one digital health app.

Table 1 depicts Spearman correlations and *P* values among the OBR scores (ORCHA, PCA, UX, and DP) and user ratings. All the OBR scores obtained a *P* value of <.001 when correlated with other OBR scores. User ratings did not obtain statistically significant correlations with any of the OBR scores. Shapiro-Wilk test results were <.001 for all the OBR scores and user ratings. Indicating that the variables are not normally distributed. Table 1 shows that achieving a high score in one dimension of PCA, UX, and DP is not a good predictor for the other two dimensions. However, PCA score is a good predictor of an overall ORCHA score, as can be seen by high Spearman correlation of .940.

K-medoids clustering was performed on the sample size (n) of 1402 digital health apps regarding their quality scores (UX, PCA, and DP) on the ORCHA assessment tool and user ratings. K-medoids had been used due to being less sensitive algorithm to outliers in data not assuming all variable to have the same variance. Figure 2 shows the results of the elbow method that was used to determine optimal number of clusters for the analysis. Considering the Elbow method, 4 clusters have been determined as the most optimal; see Multimedia Appendix 1 for analysis with 3, 5, and 6 clusters. Figure 3 depicts a k-medoids cluster plot of 4 clusters that were used to assign digital health apps based on their quality scores (UX, PCA, and DP) and their app store user ratings.

Each of the 4 clusters used for the analysis was assigned a label that best describes them. Table 2 depicts cluster number, label, and description. The labels indicate where improvements can be made, except for cluster 4 labeled "Higher quality apps with higher user ratings," as this cluster represents ideal (regarding quality) digital health apps. Table 3 depicts the median (IQR) of each variable used in the k-medoids clustering and cluster medoids for each of the four clusters, as well as ORCHA score median (IQR) and the cluster size.

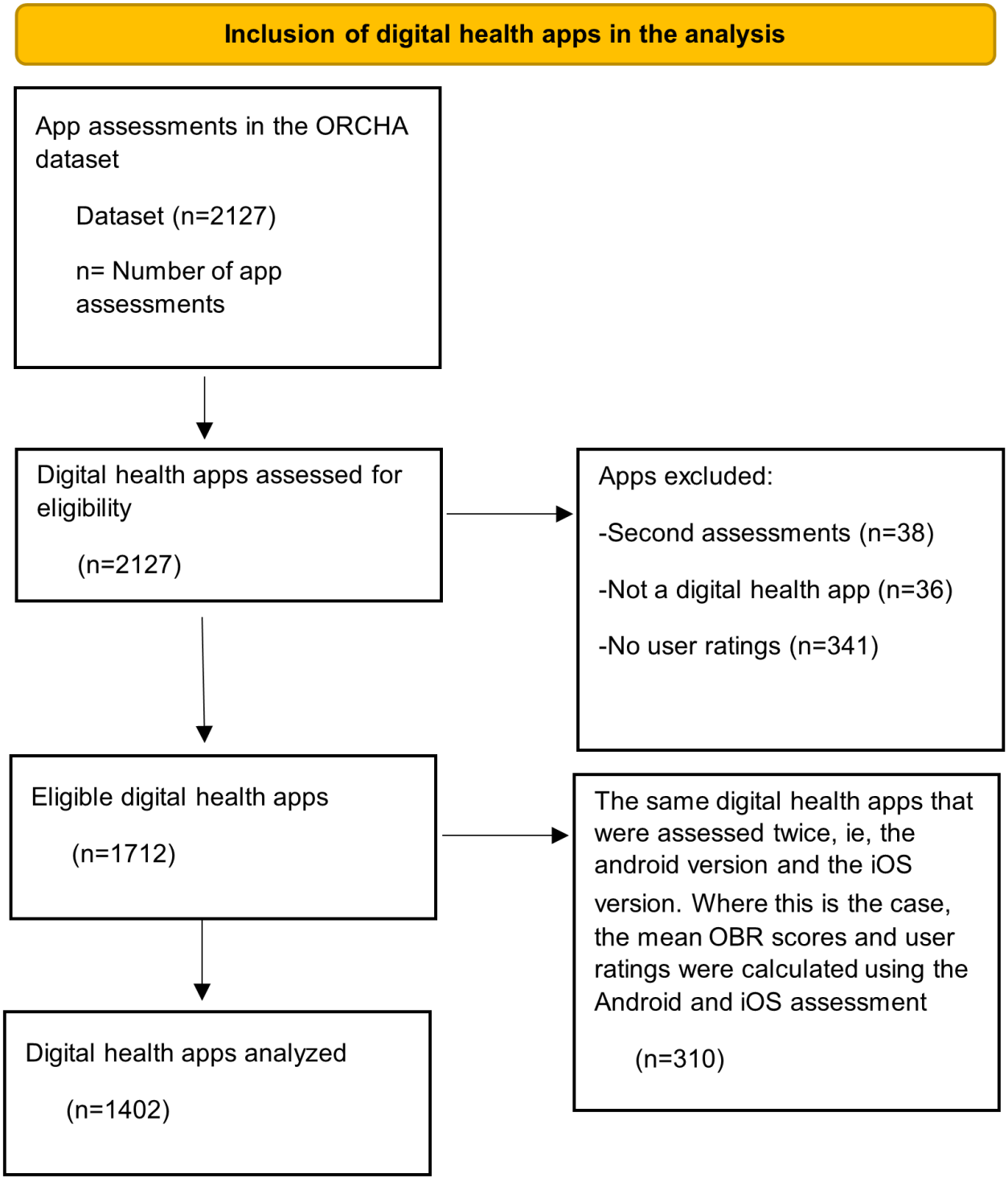
Figure 4 shows the value of each variable used in clustering per cluster in boxplots. The labels for each of the variables used in clustering were determined based on values in Table 3 and on depictions in Figure 4 boxplots. Figure 4A depicts user rating boxplots that show that user ratings are high for all clusters except for the cluster number 1 "Apps with poor user ratings." Figure 4B depicts UX score boxplots that show that UX is high for all clusters. In Figure 4C, PCA score boxplots show that PCA is high for cluster number 4 "Higher quality apps with higher user ratings," variable for cluster number 1 "Apps with poor user ratings," and low for 2 clusters "Apps with poor PCA/DP" and "Apps with poor PCA." Finally, Figure 4D shows DP score boxplots that show that DP is high for 3 clusters and lower for cluster number 2 "Apps with poor PCA/DP." Multimedia Appendix 2 shows Wilcoxon rank sum test for each of the variables against each cluster.

K-medoids clustering conducted with 3, 5, and 6 clusters can be found in Multimedia Appendix 1. Clustering results with 3 clusters indicated that further clustering may be possible, whereas clustering with 5 and 6 clusters indicated that the number can be reduced. For clustering with 5 clusters, cluster number 5 has intermediate results for PCA and DP scores, indicating that digital health apps in this cluster may be split into clusters with higher and lower PCA and DP scores. For clustering with 6 clusters, 2 of the same labels can be used to describe 4 (out of 6) different clusters, further indicating that

XSL•FO
RenderX

the number of clusters should be lower. Hence, for the rest of the analyses, results with 4 clusters have been used.

Figure 1. Digital health apps selection diagram. ORCHA: Organisation for the Review of Care and Health Apps; OBR: ORCHA Baseline Review.



XSL•FO
RenderX

Table . ORCHA Baseline Review (OBR) scores and user ratings Spearman correlations and *P* values.

Variables		Correlation	*P* value
ORCHA[a] score			
	User ratings	.002	..94
	PCA[b] score	.94	<.001
	UX[c] score	.53	<.001
	DP[d] score	.57	<.001
PCA score			
	User ratings	−.02	.57
	UX score	.41	<.001
	DP score	.33	<.001
UX score			
	User ratings	.04	.14
	DP score	.23	<.001
DP score			
	User ratings	−.02	.37

[a]ORCHA: Organisation for the Review of Care and Health Apps.

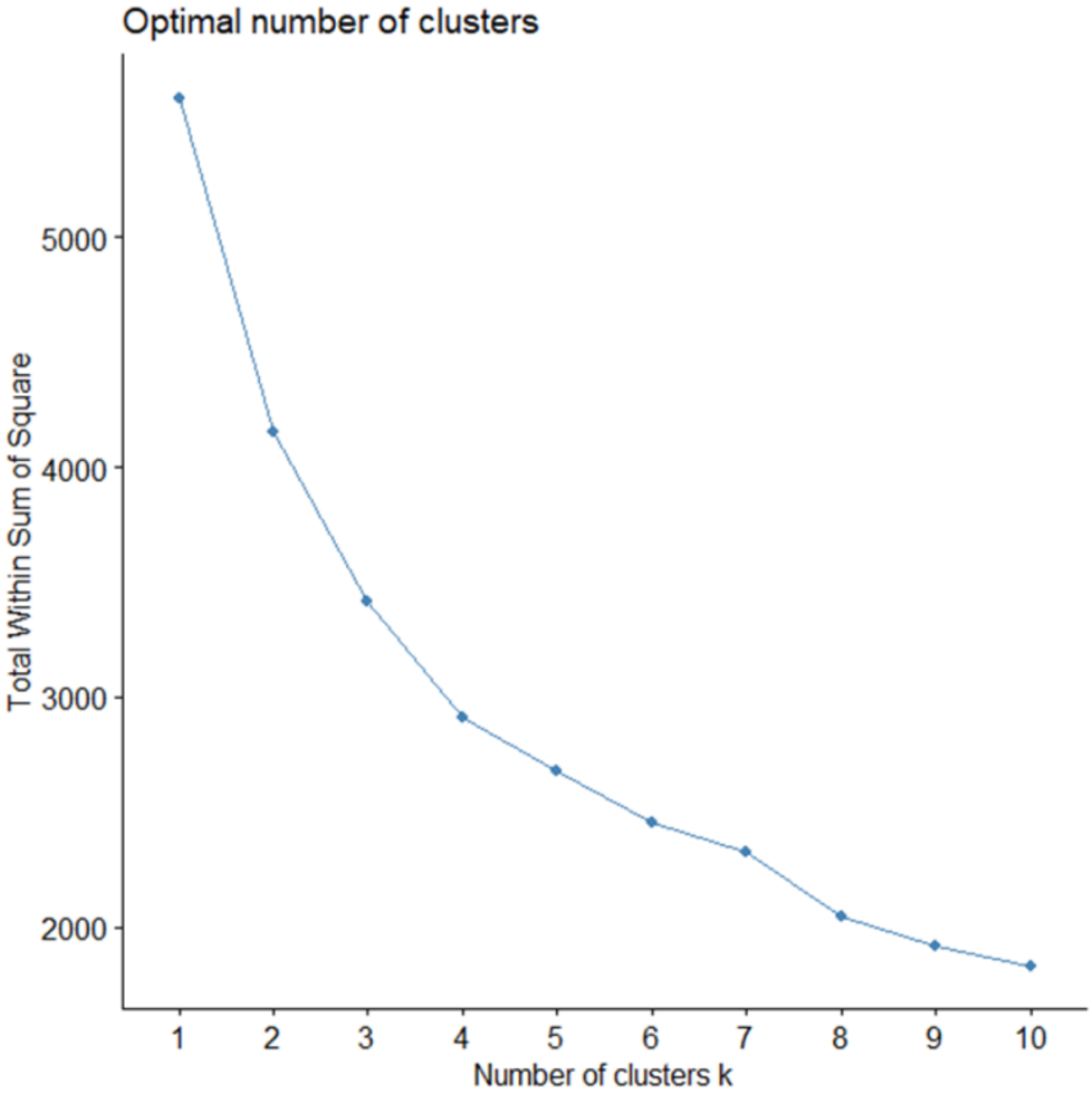
[b]PCA: professional and clinical assurance.

[c]UX: user experience.

[d]DP: data privacy.

XSL•FO
RenderX

Figure 2. Optimal cluster selection—elbow method.



XSL•FO
RenderX

Figure 3. (A) K-medoids cluster plot and (B) 3D plot of user experience (UX), professional and clinical assurance (PCA), and data privacy (DP) separated by clusters.

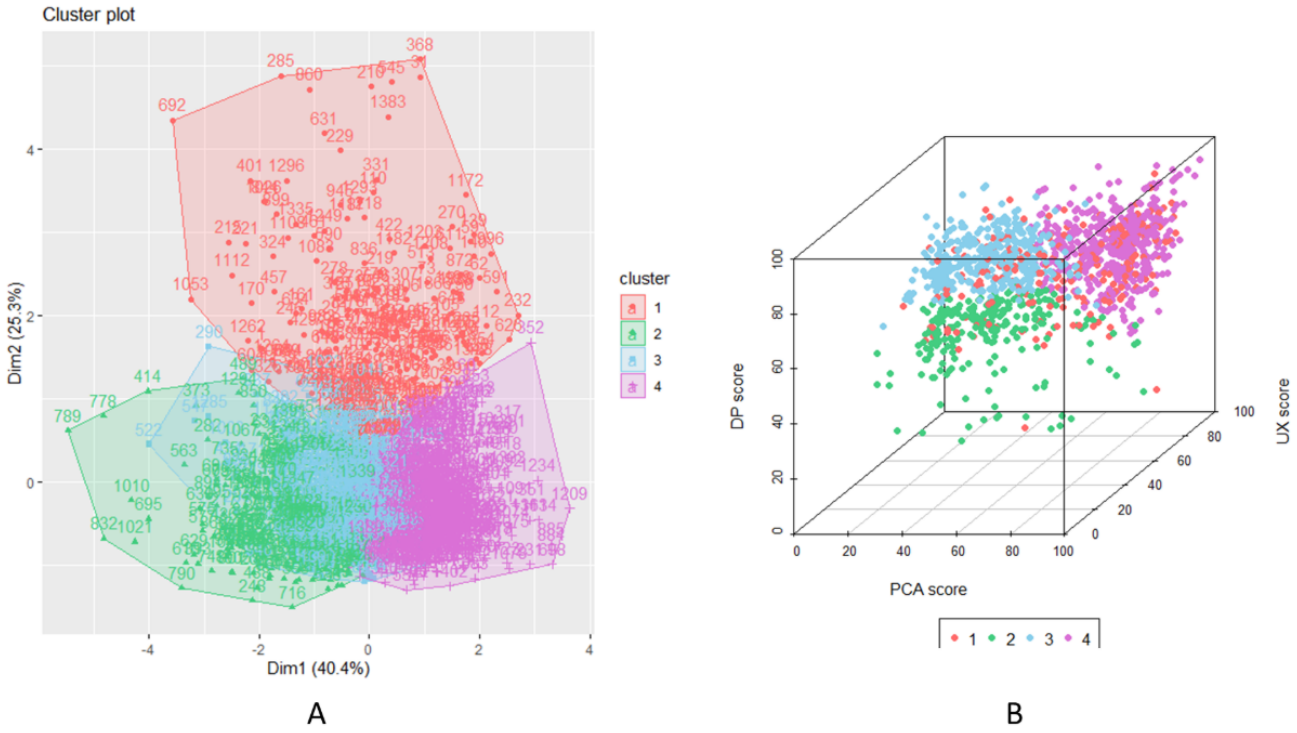


Table . Cluster number, label, and description.

Cluster number	Cluster label	Description	Cluster size, n (%)
1	Apps with poor user ratings	These are the apps that have low user ratings, but have intermediate DP[a], PCA[b] scores, and a high UX[c] score.	220 (15.7)
2	Apps with poor PCA and DP	These are the apps with low PCA and DP scores but have high UX scores and user ratings.	252 (18)
3	Apps with poor PCA	These are the apps with a low PCA score but high UX and DP scores and high user ratings.	415 (29.6)
4	Higher quality apps with higher user ratings	These are apps with high user ratings and high scores across all 3 domains.	515 (36.7)

[a]DP: data privacy.

[b]PCA: professional and clinical assurance.

[c]UX: user experience.

XSL•FO
RenderX

Table . Cluster centers of 1402 digital health apps, with cluster size percentage (%) out of 1402.

Variables	Median (IQR)	Cluster medoids			
		1	2	3	4
		Apps with poor user ratings	Apps with poor PCA[a] and DP[b] scores	Apps with poor PCA scores	Higher quality apps with higher user ratings
User ratings	4.49 (0.707)	3.23	4.58	4.47	4.60
PCA score	49 (45.3)	57.0	35.1	31.9	73.0
UX[c] score	75.2 (9.42)	72.9	71.6	71.6	80.3
DP score	65.3 (18.3)	61.3	42.4	67.5	66.4
ORCHA[d] score, median (IQR)	—[e]	63 (21)	46 (12)	54 (9)	74 (10)
Cluster size, n (%)	—	220 (15.7)	252 (18)	415 (29.6)	515 (36.7)

[a]PCA: professional and clinical assurance.

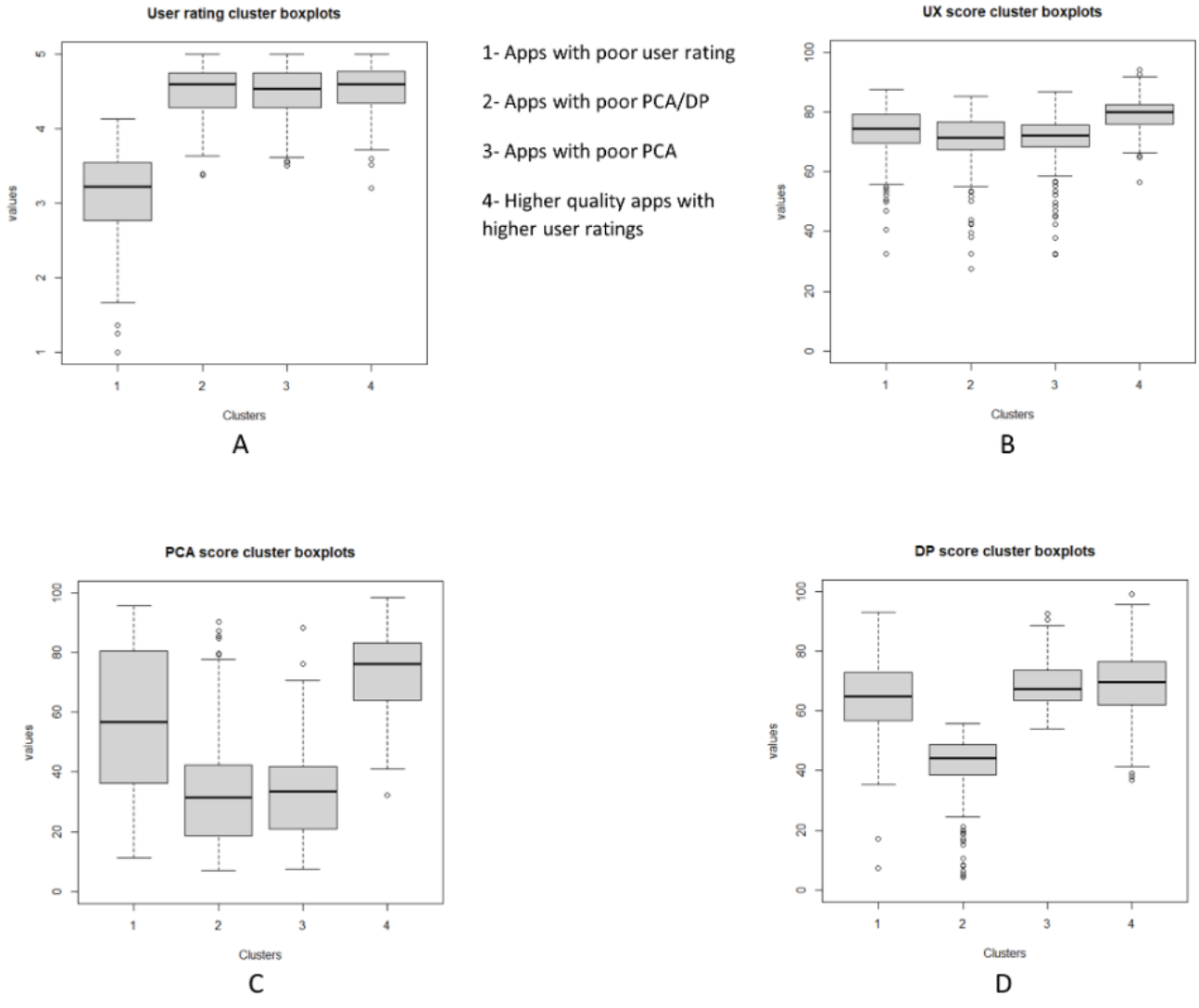
[b]DP: data privacy.

[c]UX: user experience.

[d]ORCHA: Organisation for the Review of Care and Health Applications.

[e]Not applicable.

Figure 4. Boxplot of user ratings and Organisation for the Review of Care and Health Applications. Baseline Review (OBR) scores per cluster. (A) user ratings cluster boxplots, (B) user experience (UX) cluster boxplots, (C) professional and clinical assurance (PCA) cluster boxplots, and (D) data privacy (DP) cluster boxplots.



XSL•FO
RenderX

Post Hoc Analysis

Fisher exact test with Bonferroni corrected alpha value was conducted between smallest and largest cluster percentage for NICE ESF tiers, target users, categories, and features. Statistical significance with Bonferroni corrected alpha value was found in 2/3 NICE ESF tiers, 0/14 target users, 4/33 categories, and 6/19 features. Results are presented in Multimedia Appendix 2 and Multimedia Appendix 3.

Discussion

Principal Findings

K-medoids clustering analysis was conducted using 4 clusters. The elbow method was used to select the optimal number of clusters. Different clusters were also tried; see Multimedia Appendix 1 for analysis with 3, 5, and 6 clusters. The clustering was performed on 4 variables: user ratings, PCA score, UX score, and DP score (see Table 3 and Figure 4). The four clusters were labeled: (1) apps with poor user ratings (n=220), (2) apps with poor PCA and DP scores (n=252), (3) apps with poor PCA scores (n=415), and (4) higher quality apps with higher user ratings (n=515). See Table 2 for label description. Considering the OBR scores, user ratings, and cluster size, the following conclusions were drawn:

The most common digital health app (in this dataset) is those with high user ratings and high OBR quality scores, as indicated by the cluster labeled “higher quality apps with higher user ratings,” n=515 (36.7%).

There are many digital health apps that lack PCA but excel in user ratings, UX, and DP scores, as indicated by the cluster “Apps with poor PCA scores,” n=415 (29.6%). Meaning these apps may be portrayed as good quality apps, despite not being evidence-based.

User ratings are not necessarily indicative of OBR quality assessment scoring. A digital health app can receive high user ratings and low OBR scores and vice versa, as indicated by Table 1. The finding that user ratings are not linked to OBR scores was confirmed with different analysis using the subset of the same original dataset [31].

As revealed by Hyzy et al [31], widely available proxies that users may perceive to signify the quality of health apps, namely user ratings and downloads, were found to be inaccurate predictors for estimating quality. That study indicated the need for the wider use of quality assurance methodologies that can accurately determine the quality, safety, and compliance of health apps, and this research clearly confirms that finding indicating that more should be done to enable users to recognize high-quality health apps, including digital health literacy training and the provision of nationally endorsed “libraries.”

The discovered typology could be used by assessors to classify future apps. The typology can also be used to track the size of these clusters over time. Looking for changes and trends—a kind of digital health app surveillance.

To further understand the clusters, a post hoc analysis was conducted, examining the spread of NICE ESF tiers, target users, categories, and features across the 4 clusters. Multimedia Appendix 3 shows the assigned NICE ESF tiers for each cluster. Multimedia Appendix 4 shows assigned target users, categories, and features for each of the clusters. This was done to check whether there is any relationship among NICE ESF tiers, target users, categories and features, and the clusters.

Fisher exact test with Bonferroni corrected alpha value was used between the largest and the smallest percentage of prevalence of NICE ESF tiers, target users, categories, and features relative to the cluster size. Statistical significance with Bonferroni corrected alpha value was achieved by 2/3 NICE ESF tiers, 0/14 target users, 4/33 categories, and 6/19 features. Results are presented in Multimedia Appendix 2 and Multimedia Appendix 3. For example, for the category “Ophthalmology,” the cluster labeled “Higher quality apps with higher user ratings” had the smallest percentage of prevalence of .583% and the cluster labeled “Apps with poor PCA/DP” had the largest percentage of prevalence of 6.75%, relative to cluster size. The Fisher exact test *P* value for the 2 clusters was <.001, with Bonferroni corrected alpha value of .002. These results are indicating statistically significant differences between the clusters with a Cohen *w* of .182. This means that there is a statistically significant association between some NICE ESF tiers, target users, categories, and features, and the clusters. However, the effect size was small, Cohen *w*<.2, for most NICE ESF tiers, target users, categories, and features that achieved statistical significance. Most noteworthy was the NICE ESF tier B and C with Cohen *w* of .193 and .190, respectively. As well as the category “Ophthalmology” and the app feature “Service Signposting” with Cohen *w* of .182 and .241, respectively, indicating small-medium effect size when considering both Cohen *w*s [28] and Funder and Ozer [29] interpretation. Hence, the results seem to indicate that NICE ESF tiers, target user, categories, and features, by and large, do not have an effect on the typology of digital health apps.

Previous studies have been done examining quality and impact of digital health apps, as well as finding areas for improvement. A review from 2022 [32] examined the impact of digital health apps. The review discussed that there is little evidence to show that digital health apps impact health outcomes either positively or negatively. However, a study from 2021 [33] discussed the impact of telemedicine and the use of digital health apps for healthcare. The study found that the use of telemedicine and digital health apps can streamline the workflow of hospitals and clinics; for example, scheduling follow-up visits may allow doctors and patients to be more effective and optimize patient outcomes. Hence, digital health apps can be a good complementary tool to standard health care. A review from 2023 [34] discussed barriers to the use of digital health apps. The review revealed 10 major barriers and problems associated with the use of digital health apps. Those were “validity,” “usability,” “technology,” “use and adherence,” “data privacy and data security,” “patient-physician relationship,” “knowledge and skills,” “individuality,” “implementation,” and “costs.” Furthermore, the results showed that more research is needed to study the problems and barriers of digital health apps. A review from 2021 [35] examined the effectiveness of digital health apps to promote health and manage disease. The review found a steady increase in the rigorous assessment of digital

XSL•FO
RenderX

health apps. Although it also found that there is a need for improved methodological and assessment approaches. Our study of categorizing digital health apps into a 4-cluster typology could lead to improving the quality of digital health apps, as it identifies areas of improvement and how prevalent they are.

Further analyses could be conducted on OBR scores and NICE ESF tiers, target user, categories, and app features. To examine how the OBR score changed when partitioned by NICE ESF tiers, target users, and categories. Or whether having specific features is linked to higher OBR scores. Furthermore, studies can be done to attempt to find the underlying cause for the 4-cluster typology. Possible analyses include whether the inclusion of a clinician during app development improves the app's quality. Or does having a higher budget during app development lead to higher quality of apps?

Further analysis could be done to reproduce this study using different assessment frameworks. The study conducted in this paper was done using the ORCHA assessment tool called OBR. However, there are alternative frameworks that could be used to assess apps, such as the Mobile App Rating Scale (MARS) [36]. OBR and MARS also use different rating techniques, where OBR uses mostly polar yes or no questions assigning risk and value points, whereas MARS uses 5-point scale questions. Another alternative is the Enlight [37] framework, which uses 5-point scale questions and checklists.

It may be helpful if app stores, or app developers, presented PCA of digital health apps in a standardized way akin to a privacy policy. Perhaps users can view the underpinning evidence of the app or content in the app itself and see answers to basic questions such as "how many healthcare professionals reviewed or were involved in the design of the app?" This would allow users to make more informed decisions on whether to use an app or not.

Limitations

This analysis was conducted using 4 clusters, and using a different number of clusters may have led to slightly different results. However, the elbow method has been used to select the number of clusters, and using different number of clusters in Multimedia Appendix 1 indicated that 4 clusters is most suitable. Using a different clustering solution such as k-means, self-organizing maps (SOMs), etc, may have also led to different results. Initially, we considered using k-means clustering but changed to k-medoids due to the presence of outliers in the data. When conducting post hoc analysis on target users, categories, and features, only the lowest and highest relative percentage (to cluster size) of prevalence have been used in Fisher exact test. When partitioning the clusters by NICE ESF tiers, target users, categories, and features, sample sizes have been low for some, making results less reliable. A dataset with a larger sample size could lead to more accurate results. Furthermore, the dataset may not be representative of all health apps since it is based on apps that were submitted to ORCHA for assessment. Meaning the dataset does not include apps that were not submitted to be assessed. Hence, the dataset is not a random sample of health apps and there may be a sample bias. As a result, a random sample of health apps might result in different results and clusters.

This analysis was done on the assessment conducted using the OBR assessment tool, using a different tool (eg, MARS [36] or Enlight [37]) could lead to different results. Using a quality assessment tool designed to assess a specific category of a digital health app could lead to more accurate results (category-specific typology) that could be used to improve their quality. Analysis of the OBR assessment questions for each of the clusters could have explained or indicated what led to the 4-cluster typology.

Conclusion

This study, based on the OBR assessments of 1402 digital health apps, shows that digital health apps can be assigned into a typology of 4 clusters (based on k-medoids clustering). The clusters were labeled: (1) apps with poor user ratings (220/1402, 15.7%), (2) apps with poor PCA and DP scores (252/1402, 18.0%), (3) apps with poor PCA scores (415/1402, 29.6%), and (4) higher quality apps with higher user ratings (515/1402, 36.7%). The principal findings of the analysis were: (1) the most common digital health apps are those with high user ratings and with high OBR scores indicating that these apps are of high quality and are well received by users (515/1402, 36.7%); (2) there are many digital health apps that lack PCA yet excel in user ratings, UX, and DP scores (415/1402, 29.6%), which may be dangerous as it means that many digital health apps are not evidence-based, but due to high user ratings, as well as good UX and DP policy, they may be portrayed as good quality apps; and (3) user ratings are not necessarily indicative of quality (according to OBR assessment scores), as digital health apps can receive high user ratings and low OBR scores and vice versa. Furthermore, for some digital health apps, NICE ESF tier and, to a lesser extent, category and features had a statistically significant effect on the assigned cluster. However, effect size with Cohen w was <.3 for all, indicating small association. This study indicates that many digital health apps could be improved regarding either their PCA or DP (667/1402, 47.6%). Also, 15.7% (220/1402) of digital health apps were of good quality but received poorer user ratings than other clusters; this might be due to apps not meeting user demands but providing accurate research-informed content. Further research is needed to understand the underlying cause for the 4-cluster typology of digital health apps.

Knowledge of the quality shortcomings in digital health apps and how prevalent they are as shown by the 4 clusters and their cluster size can inform the direction needed for future research. This study showed that the examined NICE ESF tiers, target users, categories, and features of digital health apps are not strongly associated with the 4-cluster typology of digital health apps.

XSL•FO
RenderX

Acknowledgments

This study was done as part of a PhD Co-operative Awards in Science and Technology (CAST) award, with funding from the Department for the Economy (DfE) in Northern Ireland and the Organisation for the Review of Care and Health Applications (ORCHA) in the United Kingdom.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Digital health apps clusters.
[DOCX File, 18 KB - mhealth_v13i1e57279_app1.docx]

Multimedia Appendix 2
Digital health apps cluster values comparison (4-clusters typology).
[DOCX File, 16 KB - mhealth_v13i1e57279_app2.docx]

Multimedia Appendix 3
Clusters per National Institute for Health and Care Excellence Evidence Standard Framework (NICE ESF) tier.
[DOCX File, 15 KB - mhealth_v13i1e57279_app3.docx]

Multimedia Appendix 4
Digital health apps' target user, categories, and features per cluster.
[DOCX File, 32 KB - mhealth_v13i1e57279_app4.docx]

References

1. Shah SS, Gvozdanovic A, Knight M, Gagnon J. Mobile app-based remote patient monitoring in acute medical conditions: prospective feasibility study exploring digital health solutions on clinical workload during the COVID crisis. JMIR Form Res 2021 Jan 15;5(1):e23190. [doi: 10.2196/23190] [Medline: 33400675]
2. Shah SS, Safa A, Johal K, Obika D, Valentine S. A prospective observational real world feasibility study assessing the role of app-based remote patient monitoring in reducing primary care clinician workload during the COVID pandemic. BMC Fam Pract 2021 Dec 16;22(1):248. [doi: 10.1186/s12875-021-01594-7] [Medline: 34911442]
3. Bezerra Giordan L, Ronto R, Chau J, Chow C, Laranjo L. Use of mobile apps in heart failure self-management: qualitative study exploring the patient and primary care clinician perspective. JMIR Cardio 2022 Apr 20;6(1):e33992. [doi: 10.2196/33992] [Medline: 35442205]
4. Mair JL, Salamanca-Sanabria A, Augsburger M, et al. Effective behavior change techniques in digital health interventions for the prevention or management of noncommunicable diseases: an umbrella review. Ann Behav Med 2023 Sep 13;57(10):817-835. [doi: 10.1093/abm/kaad041] [Medline: 37625030]
5. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
6. ORCHA. URL: https://orchahealth.com/ [accessed 2022-02-02]
7. Digital health attitudes and behaviour report 2023. : ORCHA; 2023 URL: https://info.orchahealth.com/digital-health-attitudes-behaviour-2023-report [accessed 2023-10-17]
8. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817] [Medline: 31293246]
9. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
10. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
11. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983. [doi: 10.2196/12983] [Medline: 31469081]
12. Cucciniello M, Petracca F, Ciani O, Tarricone R. Development features and study characteristics of mobile health apps in the management of chronic conditions: a systematic review of randomised trials. NPJ Digit Med 2021 Oct 5;4(1):144. [doi: 10.1038/s41746-021-00517-1] [Medline: 34611287]
13. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14. [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]

XSL•FO
RenderX

14. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223. [doi: 10.2196/11223] [Medline: 30990458]
15. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046. [doi: 10.2196/17046] [Medline: 32186518]
16. Stern AD, Brönneke J, Debatin JF, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health 2022 Mar;4(3):e200-e206. [doi: 10.1016/S2589-7500(21)00292-2] [Medline: 35216754]
17. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Feb;1(2):87-100. [doi: 10.1038/s44159-021-00003-1] [Medline: 38515434]
18. Galetsi P, Katsaliaki K, Kumar S. Assessing technology innovation of mobile health apps for medical care providers. IEEE Trans Eng Manage 2023;70(8):2809-2826. [doi: 10.1109/TEM.2022.3142619]
19. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118. [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
20. Larsen ME, Huckvale K, Nicholas J, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18. [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
21. Vaitkienė G, Kuzborska Z, Žukauskienė M. Digital health solutions for chronic illnesses: a systematic review of mobile health apps and quality analysis with mobile app rating scale. JAL 2022;2(3):193-205. [doi: 10.3390/jal2030016]
22. Sophie H. Review documentation - review development & resources. External Share for Confluence. URL: https://confluence.external-share.com/content/b6055aac-83e4-4947-be0e-ebb8c39559ef [accessed 2022-03-13]
23. Wisniewski H, Liu G, Henson P, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9. [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
24. Classification of Digital Interventions, Services and Applications in Health: World Health Organization; 2023:1-66.
25. Evidence standards framework for digital health technologies. NICE. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2022-06-02]
26. Hyzy M, Bond R, Mulvenna M, Bai L, Leigh S. The typology of digital health apps according to their quality scores and user ratings: k-means clustering. Presented at: ACM International Conference Proceeding Series 2023; Sep 19-22, 2023; Swansea, United Kingdom. [doi: 10.1145/3605655.3605786]
27. Hyzy M, Bond R, Mulvenna M, et al. Quality of digital health interventions across different health care domains: secondary data analysis study. JMIR Mhealth Uhealth 2023 Nov 23;11:e47043. [doi: 10.2196/47043] [Medline: 37995121]
28. Cohen J. Statistical Power Analysis for the Behavioral Sciences Second, 2nd edition: Routledge; 1988.
29. Funder DC, Ozer DJ. Evaluating effect size in psychological research: sense and nonsense. Advances in Methods and Practices in Psychological Science 2019 Jun;2(2):156-168. [doi: 10.1177/2515245919847202]
30. Privacy policy. ORCHA. URL: https://appfinder.orcha.co.uk/privacy-policy/ [accessed 2024-08-11]
31. Hyzy M, Bond R, Mulvenna M, et al. Don't judge a book or health app by its cover: user ratings and downloads are not linked to quality. PLoS ONE 2024;19(3):e0298977. [doi: 10.1371/journal.pone.0298977] [Medline: 38437233]
32. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 5;43(1):117-134. [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
33. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117. [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
34. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808. [doi: 10.2196/43808] [Medline: 37171838]
35. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563. [doi: 10.2196/21563] [Medline: 33427672]
36. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
37. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82. [doi: 10.2196/jmir.7270] [Medline: 28325712]

Abbreviations

DP: data privacy
ESF: Evidence Standard Framework
MARS: Mobile App Rating Scale
MAUQ: mHealth App Usability Questionnaire

XSL•FO
RenderX

NHS: National Health Service
NICE: National Institute for Health and Care Excellence
OBR: Organisation for the Review of Care and Health Applications Baseline Review
ORCHA: Organisation for the Review of Care and Health Applications
PCA: professional and clinical assurance
UX: user experience
WHO: World Health Organization

Edited by L Buis, TDA Cardoso; submitted 10.02.24; peer-reviewed by J Liang, R Pryss; revised version received 25.05.25; accepted 26.05.25; published 23.07.25.

Please cite as:
Marek Zych M, Bond R, Mulvenna M, Bai L, Martinez-Carracedo J, Leigh S
Grouping Digital Health Apps Based on Their Quality and User Ratings Using K-Medoids Clustering: Cross-Sectional Study
JMIR Mhealth Uhealth 2025;13:e57279
URL: https://mhealth.jmir.org/2025/1/e57279
doi:10.2196/57279

© Maciej Marek Zych, Raymond Bond, Maurice Mulvenna, Lu Bai, Jorge Martinez-Carracedo, Simon Leigh. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identifying and Evaluating Mobile and Web Apps for Patients to Manage Hidradenitis Suppurativa: Systematic Search in App Stores and Content Analysis

Caroline Glatzel, MD; Tassilo Dege, MSc, MD; Bernadette Glatzel; Matthias Goebeler, MD; Dagmar Presser, MD; Astrid Schmieder, MD

Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, Josef-Schneider-Strasse 2, Würzburg, Germany

Corresponding Author:
Caroline Glatzel, MD
Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, Josef-Schneider-Strasse 2, Würzburg, Germany

Abstract

Background: Hidradenitis suppurativa (HS) is a chronic inflammatory skin disease characterized by painful nodules, abscesses, and fistulas in intertriginous sites. It significantly impacts patients' quality of life. Early diagnosis and timely treatment are essential for disease control. Recurrent flares, suboptimal therapies, and prolonged misdiagnosis place a significant burden on both patients and health care systems.

Objective: We aimed to identify mobile health apps (MHAs) for patients with HS and evaluate their quality through assessments by both patients and physicians.

Methods: Two reviewers searched for mobile and web apps for HS, including those only available in German or English. Apps with advertising or non–patient-centered content and apps related to trials or conferences were excluded. Two apps met the criteria and were evaluated by 20 physicians and 27 patients using the Mobile App Rating Scale (MARS), user version of the MARS (uMARS), German Mobile App Usability Questionnaire, and technology affinity tools (Affinity for Technology Interaction Scale and Mobile Device Proficiency Questionnaire).

Results: We identified 2 apps for managing HS that met the inclusion criteria—the HSR-Patients app and the EHSF-Hidradenitis Suppurativa app—from an initial pool of 29 proposed apps that included many nonmedical, non–HS-specific, and non–patient-centered apps. Patients rated the quality of the HSR-Patients app significantly higher than physicians (MARS: mean 3.01, SD 0.60 vs. uMARS: mean 3.53, SD 0.69; *P*=.009). In contrast, ratings for the EHSF-Hidradenitis Suppurativa app did not differ significantly (physicians: mean 2.81, SD 0.55; patients: mean 2.72, SD 0.79; *P*=.69). Usability, assessed with the German Mobile App Usability Questionnaire, showed no significant difference between physicians and patients for either app. For the HSR-Patients app, physicians and patients rated usability at 4.37 (SD 0.86) and 4.72 (SD 1.21; *P*=.27), respectively. For the EHSF-Hidradenitis Suppurativa app, physicians and patients rated usability at 3.88 (SD 0.77) and 3.38 (SD 1.35; *P*=.11), respectively. Patients showed a significantly higher general affinity for technology than physicians, as measured by the Affinity for Technology Interaction Scale (physicians: mean 3.62, SD 0.61; patients: mean 4.38, SD 1.30; *P*=.01). However, there was no significant difference in affinity for technology specifically when using mobile devices, as assessed by the Mobile Device Proficiency Questionnaire (physicians: mean 4.83, SD 0.25; patients: mean 4.69, SD 0.72; *P*=.41).

Conclusions: This evaluation highlights the limited availability of high-quality, HS management–specific MHAs and underscores the need for more targeted digital tools. Differences in evaluations between patients and physicians were evident, with patients focusing on usability and practical guidance, while physicians prioritized content and usability. Neither the HSR-Patients app or the EHSF-Hidradenitis Suppurativa app demonstrated sufficient potential for long-term use, indicating the need for participatory development that includes all stakeholders.

(JMIR Mhealth Uhealth 2025;13:e69030) doi:10.2196/69030

KEYWORDS

hidradenitis suppurativa; acne inversa; mobile health apps; MHA; mobile apps; Mobile Application Rating Scale; MARS; uMARS; German mHealth App Usability Questionnaire; G-MAUQ; Mobile Device Proficiency Questionnaire; MDPQ-16; Affinity for Technology Interaction Scale; ATI Scale; skin

XSL•FO
RenderX

Introduction

Background

Hidradenitis suppurativa (HS; also known as acne inversa) is a chronic, recurrent skin disease characterized by painful inflammatory nodules, abscesses, and tunnels in areas such as the axilla, groin, and genitalia. HS leads to a significant physical and psychological burden for those affected [1]. The global prevalence of HS is reported to be between 0.4% and 1%, with many cases remaining unreported [2,3]. Early and stage-appropriate treatment is essential to prevent irreversible tissue damage [4]. Chronic skin inflammation is frequently associated with relevant comorbidity including metabolic syndrome, chronic inflammatory bowel disease, axial spondyloarthropathies, and depression, which require interdisciplinary management. There is also a clear association with active smoking and obesity [3].

HS often manifests after puberty, significantly impacting patients' lives due to chronic pain, limited mobility, persistent secretions, and unpleasant odors [5].

The diagnosis of HS is often considerably delayed, with a mean time to diagnosis of approximately 10 years from symptom onset reported in the literature [6]. While delays are sometimes attributed to patients, due to factors such as embarrassment or long waiting times for specialist consultations, it is important to recognize that diagnostic challenges in clinical practice may also play a role. In primary care, early HS lesions can resemble common skin infections and may be misinterpreted as such [3]. Smaller fistulas and abscesses are not always identified as manifestations of HS and may be diagnosed and treated as isolated dermatological findings [7,8]. Thus, patients often undergo multiple surgeries for recurrent abscesses, which greatly reduces their quality of life [9].

In the search for answers and support, many patients turn to the internet. Once a diagnosis has been made, HS-specific mobile health apps (MHAs) can serve as valuable tools for both patients and physicians. These apps may help improve medical care by supporting patients' education, explaining complex aspects of HS, enabling symptom tracking, and facilitating communication. This can empower patients to better manage their condition and help physicians save time during consultations. To fully leverage this potential, the integration of MHAs into routine care for HS should be further explored and promoted.

MHAs or digital health apps (DiGAs) already offer a promising opportunity to support self-management [10] of diseases and ideally have a positive impact on care. Since 2019, specially certified, patient-centered MHAs can be prescribed in Germany as DiGAs, which are covered by public health insurances. These DiGAs support disease self-management, improve care coordination, facilitate access to health care, enhance health literacy, and provide guideline-based recommendations. They also allow users to track disease progression (eg, through photo documentation and questionnaires), monitor therapy progress, and access helpful resources during acute episodes, ultimately improving disease control and quality of life. Germany, with its large insured population, is an attractive market for health care digitalization. However, there are currently no empirically validated DiGAs specifically developed for patients with HS, highlighting the need for a thorough evaluation of available MHAs. Given the large number of apps on the market, critical evaluation is essential [11]. The quality and effectiveness of apps can vary widely, and few offer the features that patients and clinicians actually need. Collecting user experiences and feedback is therefore vital for identifying the best digital tools for patients with HS [12,13].

Aim of This Study

The aims of this study were to analyze the quality and usability of currently available apps, to gain insight into users' technology affinity and requirements, and to draw conclusions for the development of future HS apps as DiGAs. We sought to answer these questions using validated questionnaires, including the Mobile Application Rating Scale (MARS), German mHealth App Usability Questionnaire (G-MAUQ), Affinity for Technology Interaction (ATI) Scale, Mobile Device Proficiency Questionnaire (MDPQ-16), and an open-ended questionnaire developed specifically for this study.

We hypothesize that there are currently no high-quality HS apps available in app stores, but the development of such apps will be essential in the future—not least due to the high technology affinity of patients with HS. We believe that patients with HS, who are typically young and tech-savvy, would greatly benefit from a well-designed, customized app that tracks their therapy, documents involved anatomical locations for better treatment control, provides support during acute stress situations, and strengthens the connection with their health care providers.

Methods

Ethical Considerations

This analysis was conducted in accordance with the Declaration of Helsinki and ethics approval was obtained from the local ethics committee of the University of Würzburg (2023091901). All participants provided written informed consent. Privacy and confidentiality were maintained in accordance with applicable data protection regulations.All personal data were anonymized and securely stored. No compensation was provided for participation.

App Selection

Two reviewers systematically and independently searched the Apple App Store, the Google Play Store, and the internet for mobile apps designed for patients with HS from November to December 2023. The following search terms were used: *hidradenitis suppurativa*, *acne inversa*, *abscesses*, *furuncles*, *fistulas*, *skin condition*, and *wound*. The inclusion criteria required the apps to be available in one of the aforementioned app stores and in either German or English. The exclusion criteria included apps that contained advertising; were not patient oriented, which means they were intended for health care professionals rather than patients; or were related to clinical trials or conferences. In total, 29 apps were identified, of which 27 apps were excluded for the following reasons: not patient centered (n=2; not focusing on patients' needs), not available in English or German (n=1), not specific to HS (n=24), and

XSL•FO
RenderX

containing advertisements (n=4). After applying the inclusion and exclusion criteria, 2 apps met the requirements: the HSR-Patients app and the EHSF-Hidradenitis Suppurativa app. The HSR-Patients app was available in both the Apple App Store and the Google Play Store and in the German language. The EHSF-Hidradenitis Suppurativa app was only available in the Apple App Store and only in the English language. No mobile apps for patients with HS were found on the internet. Patients and physicians tested the app version corresponding to the operating system available on their personal smartphones (iOS or Android). If they did not own an iOS device, the EHSF-Hidradenitis Suppurativa app, available only for iOS, was tested using a study-provided iPhone to ensure both apps could be evaluated by all participants.

Information about the identified apps was systematically collected, including the app name, platform availability, target audience, language, app store rating, number of reviews, chargeability, presence of advertising, and developer details.

Evaluation of App Quality

The MARS is one of the most commonly used scores to evaluate the quality of mobile apps [14]. The MARS score is calculated based on 19 questions in 4 different categories (engagement, functionality, aesthetics, and information). Each item is rated on a 5-point Likert scale. Higher scores indicate better quality. The overall MARS score is calculated as the mean of the 4 objective subscales. The German translation of the MARS, the validated MARS-G, was used in this study [15]. For readability, the term "MARS" is used throughout the manuscript to refer to the validated German version (MARS-G) applied in this study.

The uMARS is a simplified end-user version of the MARS that reliably measures the quality of MHAs [16]. The German translation of the uMARS used in this study has not yet been formally published; however, its items are identical to those in the validated MARS-G.

Evaluation of App Usability

The G-MAUQ [17]—the German, validated version of the MAUQ [11]—is a tool to determine the usability of an MHA. It measures the ease of use, the user's satisfaction with the design of the app (interface and satisfaction), and the usefulness of an MHA for the user. The questionnaire has 18 questions that are to be answered on a 7-point Likert scale. Higher scores indicate better usability. Unlike the MARS score, it was developed specifically for evaluating medical apps [18].

The 2 apps suitable for patients with HS were evaluated by a total of 20 physicians using the MARS-G for quality and the G-MAUQ for usability [19]. The same apps were rated simultaneously by 27 patients with HS. They used the uMARS and evaluated the usability of the apps using the G-MAUQ. Before evaluating the apps using the questionnaires, the testers were asked to test the apps for at least 10 minutes each.

Evaluation of Technical Affinity

In addition, physicians' and patients' affinity for technology was assessed using the ATI Scale and the MDPQ-16. The ATI Scale, a 9-item scale, is a widely used, well-established tool for quantifying people's general affinity for technology [20]. For the ATI Scale, higher scores indicate a greater affinity for technology. In addition, we used the MDPQ-16, a short version of the MDPQ, and also asked about affinity for technology, specifically when using mobile devices [21]. For the MDPQ-16, higher scores indicate greater proficiency in using mobile devices.

Evaluation of Patients' Need for a Patients' App

To explore the potential role of digital support tools in the management of HS, both patients and physicians were asked to complete a questionnaire specifically developed for this study. It included open-ended questions addressing current experiences, unmet needs, and expectations regarding MHAs tailored to HS. The aim was to assess overall interest in digital support and to identify key features that could enhance disease monitoring, communication, and self-management.

Comparative Analysis of Patients' and Physicians' Data

Following the evaluation of MARS, uMARS, G-MAUQ, ATI Scale, and MDPQ-16, we performed a comparative analysis including physicians and patients. Calculations were performed using SPSS (version 23; IBM Deutschland GmbH). After testing for normal distribution, either the Mann-Whitney *U* or 2-tailed *t* test was used, with *P* values <.05 considered statistically significant.

Results

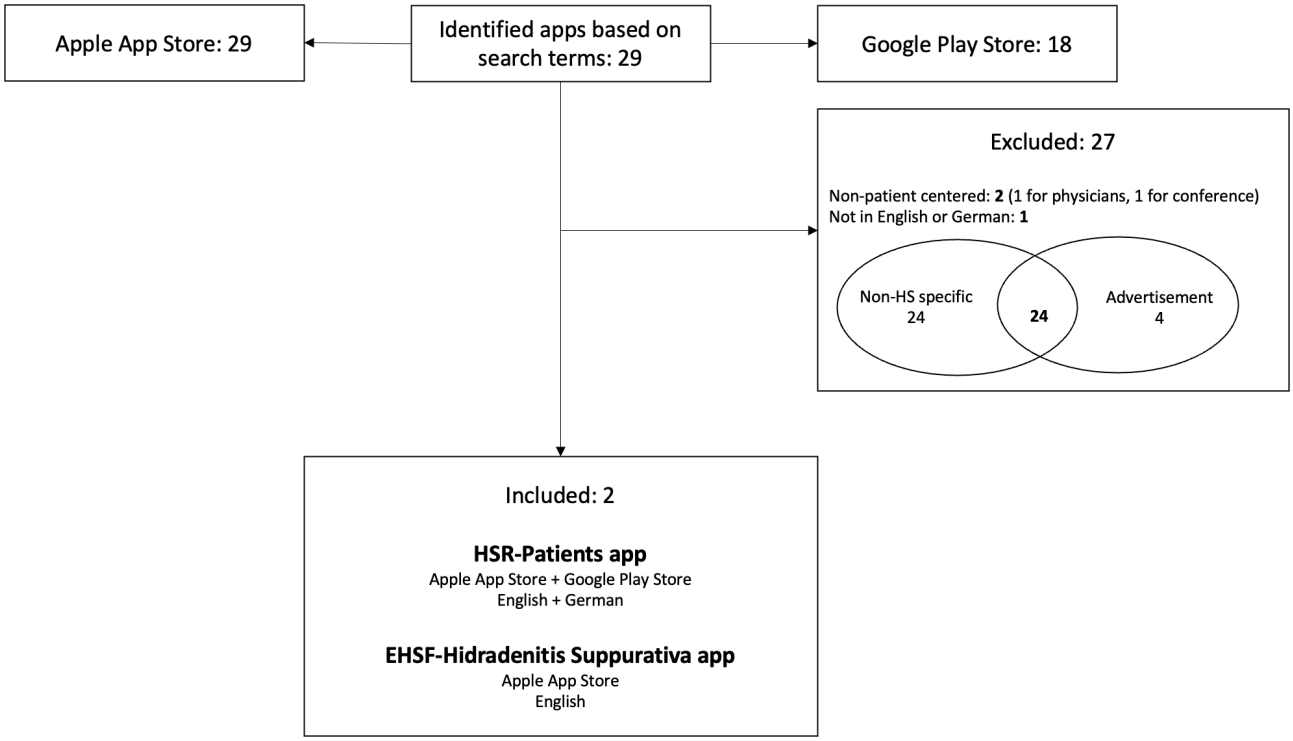
App Selection

From a total of 29 proposed apps, which included many nonmedical, non–HS-specific, and non–patient-centered apps, the 2 currently available patient-centered HS apps were identified: the HSR-Patients app and the EHSF-Hidradenitis Suppurativa app (Figure 1). The HSR-Patients app was developed by swiss4ward Europe S.L. with support from the Swiss Hidradenitis Suppurativa Foundation. It is available in English, French, German, and Spanish and serves as an information resource for patients with HS. Last updated in 2020, it offers various sections where patients can access information about HS, including "What is Hidradenitis Suppurativa," "Definition," "Lifestyle Adjustments," "Dietary Changes for HS," "Important Points for Effective Wound Management," and "Intimate Lesions."

The EHSF-Hidradenitis Suppurativa app is a Conformité Européenne–certified, English-language app developed in 2018 by the European Hidradenitis Suppurativa Foundation e.V. in collaboration with everywhereIM BV. It supports HS classification based on the "refined Hurley classification" [22] and presents treatment options. In the "Classification" tab, patients can indicate the number of body parts affected by HS, the number of abscesses or inflammatory nodules, and the presence of tunnels. Based on these data, the patient is shown the Hurley stage and appropriate therapy suggestions. The patient can then export a summary of this information as a PDF document. There are 2 references under the "Literature" tab. Only one tab can be opened at a time. The patient can access information on "Refined Hurley Staging," while the other tab IHS-4 cannot be opened at the same time.

XSL•FO
RenderX

Figure 1. Flowchart illustrating the process of app selection. HS: hidradenitis suppurativa.



Raters' Characteristics

A total of 20 physicians and 27 patients affected by HS for at least 1 year evaluated the 2 apps. The average age of the physicians was 31.5 (SD 3.12) years, while the average age of the patients was 41.8 (SD 9.5) years. There were 7 men and 13 women among the physicians and 15 men and 12 women among the patients with HS.

Evaluation of App Quality

A total of 2 apps, the HSR-Patients and the EHSF-Hidradenitis Suppurativa apps, met the inclusion criteria and were included in this analysis. The results of the MARS and uMARS scores for the respective apps are shown in Tables 1 and 2.

The quality of the HSR-Patients app was assessed by physicians using the MARS and by patients using the uMARS. Higher scores indicate a better result. Physicians rated the app's quality significantly lower than patients, with a mean score of 3.01 (SD 0.60) out of a possible 5 points, compared to the patients' average rating of 3.53 (SD 0.69), indicating a statistically significant difference (P=.009).

In all uMARS subcategories—engagement, functionality, aesthetics, information, and subjective quality—patients consistently rated the HSR-Patients app higher than the EHSF-Hidradenitis Suppurativa app. Notably, there were significant rating differences between physicians and patients for the HSR-Patients app in the aesthetics (P=.008) and information (P=.04) categories. In both categories, patients rated the app significantly higher than physicians.

Table . MARS[a] score of the HSR-Patients app and EHSF-Hidradenitis Suppurativa app assessed by 20 physicians.

		HSR-Patients app	EHSF-Hidradenitis Suppurativa app
MARS score[b], mean (SD)		3.01 (0.60)	2.81 (0.55)
	A: engagement score	2.18 (0.54)	2.20 (0.66)
	B: functionality score	4.21 (0.58)	3.68 (0.77)
	C: aesthetics score	2.97 (0.72)	2.65 (0.63)
	D: information score	2.99 (0.67)	3.05 (0.77)
	PT: psychotherapy score	2.85 (0.51)	2.82 (0.76)
	E: subjective quality	2.13 (0.55)	1.98 (0.72)

[a]MARS: Mobile Application Rating Scale.

[b]Maximum MARS score: 5 points.

XSL•FO
RenderX

Table . uMARS[a] score of the HSR-Patients app and EHSF-Hidradenitis Suppurativa app assessed by 27 patients with hidradenitis suppurativa.

		HSR-Patients app	EHSF-Hidradenitis Suppurativa app
uMARS score[b], mean (SD)		3.53 (0.69)	2.72 (0.79)
	A: engagement score	2.50 (0.81)	2.16 (0.95)
	B: functionality score	4.37 (0.59)	3.26 (1.02)
	C: aesthetics score	3.58 (0.76)	2.63 (0.83)
	D: information score	3.35 (0.79)	2.84 (1.02)
	E: subjective quality	2.31 (0.88)	1.77 (0.83)

[a]uMARS: user version of the Mobile Application Rating Scale.

[b]Maximum uMARS score: 5 points.

Evaluation of Usability

Usability of the HSR-Patients app was rated similarly by physicians and patients. Higher scores are correlated with better results. The G-MAUQ for the usability of the HSR-Patients app was on average 4.37 (SD 0.86) out of 7 points for physicians and 4.72 (SD 1.21) out of 7 points for patients. There was no significant difference between physicians and patients (P=.27; Tables 3 and 4).

The EHSF-Hidradenitis Suppurativa app received lower overall usability ratings than the HSR-Patients app. Although patients rated the usability of the English-language EHSF-Hidradenitis Suppurativa app to be lower than physicians, this difference was not statistically significant (P=.11). Physicians gave the EHSF-Hidradenitis Suppurativa app a mean usability score of 3.88 (SD 0.77), while patients rated it at 3.38 (SD 1.35). However, a significant difference was found between physicians and patients in the ease of use category. Patients rated the app's ease of use significantly lower, with a mean score of 4.10 (SD 1.74), than the physicians' higher rating of 4.91 (SD=0.75), reflecting a statistically significant difference (P=.04).

Table . Evaluation of G-MAUQ[a] and G-MAUQ subscales by physicians.

		HSR-Patients app	EHSF-Hidradenitis Suppurativa app
G-MAUQ score[b], mean (SD)		4.37 (0.86)	3.88 (0.77)
	Ease of use	5.77 (0.69)	4.91 (0.75)
	Interface and satisfaction	3.82 (1.19)	3.59 (1.20)
	Usefulness	3.71 (1.09)	3.31 (1.00)

[a]G-MAUQ: German mHealth App Usability Questionnaire.

[b]Maximum G-MAUQ score: 7 points.

Table . Evaluation of G-MAUQ[a] and G-MAUQ subscales by patients.

		HSR-Patients app	EHSF-Hidradenitis Suppurativa app
G-MAUQ score[b], mean (SD)		4.72 (1.21)	3.38 (1.35)
	Ease of use	6.07 (1.00)	4.10 (1.74)
	Interface and satisfaction	4.65 (1.59)	3.40 (1.55)
	Usefulness	3.66 (1.60)	2.84 (1.54)

[a]G-MAUQ: German mHealth App Usability Questionnaire.

[b]Maximum G-MAUQ score: 7 points.

Evaluation of Technical Affinity

The ATI Scale score and the MDPQ-16 score, which assess the user's affinity for technology, were collected from both patients and physicians. The higher the ATI Scale and MDPQ-16 scores, the higher the affinity for technology.

Physicians rated their overall affinity for technology with an ATI Scale score of 3.62 (SD 0.61) out of 6 points and with an MDPQ-16 score of 4.83 (SD 0.25) out of 5 points. Patients rated their affinity for technology with an ATI Scale score of 4.38 (SD 1.30) out of 6 points and with an MDPQ-16 score of 4.69 (SD 0.72) out of 5 points.

A statistically significant difference was found between physicians and patients in their overall affinity for technology, as measured by the ATI Scale (physicians: mean 3.62, SD 0.61; patients: mean 4.38, SD 1.30; P=.01), with patients rating themselves as more technology oriented. However, no significant difference was observed between physicians and

XSL•FO
RenderX

patients in their affinity for technology when using mobile devices, as measured by the MDPQ-16 (physicians: mean 4.83, SD 0.25; patients: mean 4.69, SD 0.72; *P*=.41).

Patients' Expectations and Comments

Most patients with HS expressed interest in using an app to support the management of their condition. Specifically, 85% (23/27) of patients indicated that they could envision using such an app if it provided a "value-added, targeted service." If these criteria were met, 93% (25/27) of patients stated they would be willing to use the app regularly as a constant companion.

The most frequently mentioned features included symptom diary (21/27, 78%) for documenting therapy outcomes, identifying trigger factors, and monitoring quality of life and photo documentation (19/27, 70%) for tracking the condition's progression.

All patients expected the app to provide "useful information for daily life," with a particular emphasis on it being "regularly updated" and delivering "the latest insights on HS" in an easily understandable format. During acute flare-ups, patients wanted the app to offer "personalized tips and instructions for symptom relief" (such as pain management) in "clear language."

The "constant availability" of the app was highly valued, as it would allow users to access information and support at any time, offering guidance on therapy and disease prevention, complementing health care provider visits. In addition, the app should facilitate "easy, straightforward contact with a physician" during acute situations.

About half of the patients (14/27, 52%) expressed interest in receiving an introduction to the app and assistance with using the technology. Furthermore, 41% (11/27) of patients desired support from a digital companion to guide them through the app's features.

Discussion

Principal Findings

MHAs are increasingly being used as digital therapeutics in the medical field. As their use grows, so does the need for a robust and standardized evaluation of these apps [11].

A systematic inventory and assessment of currently available medical apps, using established and validated evaluation tools, is essential for obtaining feedback from both users (ie, patients) and prescribers (ie, health care professionals) [23]. This feedback is crucial for the continuous improvement of MHAs. The better an MHA aligns with user needs, the higher the likelihood of sustained engagement and long-term use [24].

We present the first systematic analysis of MHAs available for patients with HS in app stores and on the web. For all instruments used (eg, MARS, MARS-G, uMARS, ATI Scale, and G-MAUQ), higher scores consistently indicate better results.

The range of MHAs for managing HS is still very limited, with only 2 out of the 29 proposed apps meeting the inclusion and exclusion criteria. This reflects a broader gap in the availability of HS-specific digital tools and highlights the need for further development and evaluation of mobile health solutions tailored to the needs of this patient population.

The quality of the included apps was assessed by physicians and patients using the validated and widely used MARS score [16,18], and the overall score (mean MARS and uMARS score) ranged from a minimum of 2.72 to a maximum of 3.53 out of 5 possible points.

At first glance, the HSR-Patients app appeared to be the superior option of the two. Both physicians and patients rated it higher than the EHSF-Hidradenitis Suppurativa app across nearly all categories. However, physicians rated the app significantly lower than patients (*P*=.009).

Patients and physicians particularly appreciated the ease of use (ie, functionality) of the HSR-Patients app. Patients also rated the app highly for its appealing design (ie, aesthetics) and content (ie, information). However, physicians rated the information content (*P*=.04) and aesthetics (*P*=.008) of the HSR-Patients app significantly lower than patients.

One possible explanation for this discrepancy could be that some practical tips for daily life in the app did not align with current medical practice, a factor more apparent to physicians with medical knowledge. Patients who are lacking this expertise are less able to verify the accuracy of the content.

Previous literature has shown that patients tend to trust the content of medical apps more than physicians [25]. As a result, the app's design may have been of greater importance to patients than to health care providers.

The HSR-Patients app's attractiveness (engagement and subjective quality)—that is, the motivation to use the app repeatedly over time or to recommend it to others—was rated as low by both physicians and patients. Once the initial content had been read, there was no additional benefit in the second review. Only looking up individual facts could motivate the patient to use the app again. However, long-term use is important to achieve good results in maintaining the patient engaged with the MHA [25].

In contrast, the EHSF-Hidradenitis Suppurativa app offered slightly more interactivity. As the severity of HS progressed, users could adjust therapy options by entering them into the app's interactive scoring system.

Regular monitoring and, when necessary, adjusting therapy based on disease severity is strongly recommended in the S2k guideline to ensure treatment adherence [5].

This feature in the EHSF-Hidradenitis Suppurativa app may explain the higher ratings from physicians in the "information" subcategory.

Unlike patients, physicians were able to quickly categorize and apply the information provided by the EHSF-Hidradenitis Suppurativa app during the 10-minute testing phase. For physicians, the app's design was of secondary importance.

For patients, an appealing design, good functionality, and practical information and tips for everyday use seem to be the fundamental requirements for the successful integration of an MHA into routine medical practice.

XSL•FO
RenderX

In addition, documenting the disease course in the app could help physicians make more informed decisions about systemic therapy. By tracking the progression of the disease over time, physicians might categorize patients as moderate to severe HS, even if they would not qualify based on a single scheduled assessment (eg, by IHS4).

The results from the G-MAUQ (usability) supported this finding. Overall, both physicians and patients rated the HSR-Patients app as more user-friendly (mean G-MAUQ). The design (ie, interface), patient satisfaction with the app's functions, and the perceived usefulness of the app were rated between 2.84 and 3.59 out of a possible 7 points by both groups. Both physicians and patients gave the highest ratings for ease of use. However, patient satisfaction was lower (P=.04), likely due to the challenge of using an app in a foreign language and the novelty of classifying HS, which presented a greater barrier for patients than for physicians.

Overall, patients with HS rated themselves as more tech-savvy than physicians, according to the ATI Scale, a broader technology familiarity questionnaire (P=.01).

Physicians and patients rated themselves as having the same affinity for using MHAs. This suggests that a well-designed MHA would certainly be well received by both patients and health care providers.

A comparison of our results with MHA assessments in other chronic dermatological conditions has revealed interesting findings.

For example, in a recent study on wound care mobile apps, both patients and physicians rated the quality of an MHA as moderate, with a mean MARS score of 3.88 and a mean uMARS score of 3.89, suggesting relatively aligned expectations. In contrast, in our HS study, patient ratings of the HSR app (uMARS: mean 3.53, SD 0.69) were significantly higher than those of physicians (MARS: mean 3.01, SD 0.60), indicating a greater divergence in perceived value and user priorities [23].

A more detailed investigation of the subcategory ratings reveals that the WUND App was rated less favorably by physicians in the "information" subcategory and by patients in the "functionality" subcategory, which is consistent with our findings that physicians tend to prioritize content quality while patients place greater emphasis on usability and practical relevance [23].

In the wound care study, patients rated their technical affinity (ATI Scale) slightly lower than physicians, although the difference was not statistically significant. In contrast, patients with HS in our cohort reported significantly higher ATI Scale scores than their physicians. This difference may reflect demographic factors, such as the younger average age of patients with HS, and highlights the need to tailor MHAs to the digital affinity of different patient groups.

In a systematic evaluation of MHAs for psoriasis, the app Psoriasis Helferin was rated significantly lower by patients (uMARS: mean 3.48) than by physicians (MARS: mean 4.18), with lower scores across all subcategories. The authors emphasized that this discrepancy underscores the importance of involving patients in both the development and evaluation of MHAs, as perceptions of quality and usefulness may differ considerably between end users and health care professionals [26].

The lack of co-design and user-centered development in most currently available MHAs contributes to inconsistencies and hinders the effective development and continuous improvement of MHAs. This is particularly evident in the HS apps we analyzed, which demonstrated limited potential for long-term usability and user acceptance. To address these shortcomings, future HS-specific app development should be guided by established regulatory frameworks that emphasize clinical relevance, user engagement, data protection, and technical interoperability. A structured co-design process—bringing together patients, dermatologists, app developers, and regulatory experts from the outset—could ensure the creation of digital tools that are both clinically effective and tailored to user needs and the requirements of modern health care systems.

Limitations

The small number of included apps (n=2) represents a major limitation of this study. However, this does not primarily reflect a methodological weakness but rather reveals a substantial gap in the availability of HS-specific mobile health tools. This market deficiency underlines the need for increased attention to digital innovation in the care of patients with HS.

In most cases, the apps included in the study were tested by participants for 10 minutes each, allowing them to form an initial impression. However, it was not possible to draw conclusions regarding the long-term benefits of these apps. Further research is needed to assess the long-term effectiveness of MHAs for individuals with HS. Nevertheless, the 2 apps designed exclusively for patients with HS appear to be less suitable for this purpose.

Conclusions

MHAs are viewed differently by patients and physicians. In this analysis, patients emphasized the importance of ease of use, an appealing interface, and practical information that helps manage their condition in everyday life. However, patients often face challenges in verifying the accuracy of the information provided by the app. Physicians, while also appreciating user-friendliness and a well-designed interface, prioritize the quality and accuracy of the content. Neither of the currently available apps appears to fully engage physicians or patients for long-term use. However, the EHSF-Hidradenitis Suppurativa app shows promise in supporting patients throughout their disease progression by personalizing treatment recommendations based on symptoms.

For an MHA to be more effective, it should combine an intuitive interface with interactive features that assess disease activity and allow for real-time treatment adjustments. In addition, the app should support a more individualized treatment approach, empowering physicians to offer tailored care rather than relying solely on scheduled check-ups. This would make the app not only a tool for disease management but also an extension of the patient-physician partnership.

XSL•FO
RenderX

Future development of HS-specific MHAs should be guided by participatory co-design approaches that actively involve patients and health care professionals to ensure usability, clinical relevance, and long-term acceptance.

Acknowledgments

The authors would like to express their sincere gratitude to all the patients and physicians who contributed to this study. The participation and insights of the patients and physicians were invaluable in advancing the authors' understanding of mobile health apps for managing hidradenitis suppurativa.

CG was supported by the TWINSIGHT Clinician Scientist Program (TWINSIGHT-09) at the Medical Faculty, University of Würzburg, that is funded by the Else-Kröner-Fresenius Foundation.

Data Availability

All data generated or analyzed during this study are included in this published article.

Conflicts of Interest

None declared.

References

1. Ralf Paus L, Kurzen H, Kurokawa I, et al. What causes hidradenitis suppurativa? Exp Dermatol 2008 May;17(5):455-456. [doi: 10.1111/j.1600-0625.2008.00712_1.x]
2. Jfri A, Nassim D, O'Brien E, Gulliver W, Nikolakis G, Zouboulis CC. Prevalence of hidradenitis suppurativa: a systematic review and meta-regression analysis. JAMA Dermatol 2021;157(8):924-931. [doi: 10.1001/jamadermatol.2021.1677] [Medline: 34037678]
3. Zouboulis CC, Del Marmol V, Mrowietz U, Prens EP, Tzellos T, Jemec GBE. Hidradenitis suppurativa/acne inversa: criteria for diagnosis, severity assessment, classification and disease evaluation. Dermatology (Basel) 2015;231(2):184-190. [doi: 10.1159/000431175] [Medline: 26139027]
4. Martorell A, Giovanardi G, Gomez-Palencia P, Sanz-Motilva V. Defining fistular patterns in hidradenitis suppurativa: impact on the management. Dermatol Surg 2019;45(10):1237-1244. [doi: 10.1097/DSS.0000000000001916] [Medline: 30893168]
5. Zouboulis CC, Bechara FG, Fritz K, et al. S2k-Leitlinie zur therapie der hidradenitis suppurativa/acne inversa (ICD-10-Code: L73.2). Aktuelle Derm 2024;50(1/02):30-83. [doi: 10.1055/a-2225-7983]
6. Kokolakis G, Wolk K, Schneider-Burrus S, et al. Delayed diagnosis of hidradenitis suppurativa and its effect on patients and healthcare system. Dermatology 2020;236(5):421-430. [doi: 10.1159/000508787] [Medline: 32610312]
7. Kirsten N, Frings V, Nikolakis GD, et al. Epidemiology, patient quality of life, and treatment costs of hidradenitis suppurativa/acne inversa. Hautarzt 2021;72(8):651-657. [doi: 10.1007/s00105-021-04851-z] [Medline: 34223939]
8. Kirsten N, Petersen J, Hagenström K, Augustin M. Epidemiology of hidradenitis suppurativa in Germany - an observational cohort study based on a multisource approach. J Eur Acad Dermatol Venereol 2020;34(1):174-179. [doi: 10.1111/jdv.15940] [Medline: 31494987]
9. Matusiak Ł, Bieniek A, Szepietowski JC. Hidradenitis suppurativa markedly decreases quality of life and professional activity. J Am Acad Dermatol 2010;62(4):706-708. [doi: 10.1016/j.jaad.2009.09.021] [Medline: 20227585]
10. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164. [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
11. Kopka M, Slagman A, Schorr C, et al. German mHealth App Usability Questionnaire (G-MAUQ) and short version (G-MAUQ-S): translation and validation study. Smart Health 2024 Dec;34(4):100517. [doi: 10.1016/j.smhl.2024.100517]
12. Thangada ND, Garg N, Pandey A, Kumar N. The emerging role of mobile-health applications in the management of hypertension. Curr Cardiol Rep 2018 Jul 26;20(9):78. [doi: 10.1007/s11886-018-1022-7] [Medline: 30046971]
13. Patel B, Thind A. Usability of mobile health apps for postoperative care: systematic review. JMIR Perioper Med 2020;3(2):e19099. [doi: 10.2196/19099] [Medline: 33393925]
14. Terhorst Y, Philippi P, Sander LB, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480. [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
15. Messner EM, Terhorst Y, Barke A, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020;8(3):e14479. [doi: 10.2196/14479] [Medline: 32217504]
16. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]

XSL•FO
RenderX

17. Moorthy P, Weinert L, Harms BC, Anders C, Siegel F. German version of the mHealth App Usability Questionnaire in a cohort of patients with cancer: translation and validation study. JMIR Hum Factors 2023;10:e51090. [doi: 10.2196/51090] [Medline: 37910144]
18. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022;22(1):22. [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]
19. Rathner EM, Probst T, Terhorst Y, Barke A, Baumeister H. MARS - Mobile Anwendungen Rating Skala. YouTube. URL: https://www.youtube.com/watch?v=5vwMiCWC0Sc [accessed 2025-07-27]
20. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the Affinity for Technology Interaction (ATI) scale. Int J Hum Comput Interact 2019;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
21. Roque NA, Boot WR. A new tool for assessing mobile device proficiency in older adults: the Mobile Device Proficiency Questionnaire. J Appl Gerontol 2018 Feb;37(2):131-156. [doi: 10.1177/0733464816642582] [Medline: 27255686]
22. Horváth B, Janse IC, Blok JL, et al. Hurley staging refined: a proposal by the Dutch hidradenitis suppurativa expert group. Acta Derm Venereol 2017;97(3):412-413. [doi: 10.2340/00015555-2513] [Medline: 27535129]
23. Dege T, Glatzel B, Borst V, et al. Patient-centered chronic wound care mobile apps: systematic identification, analysis, and assessment. JMIR Mhealth Uhealth 2024;12:e51592. [doi: 10.2196/51592] [Medline: 38533818]
24. Kushendriawan MA, Santoso HB, Putra POH, Schrepp M. Evaluating user experience of a mobile health application "Halodoc" using User Experience Questionnaire and usability testing. J Sistem Inf (J Inf Sys) 2021 Apr 29;17(1):58-71. [doi: 10.21609/jsi.v17i1.1063]
25. Madeira RN, Germano H, Macedo P, Correia N. Personalising the user experience of a mobile health application towards patient engagement. Procedia Comput Sci 2018;141:428-433. [doi: 10.1016/j.procs.2018.10.173]
26. Lull C, von Ahnen JA, Gross G, et al. German mobile apps for patients with psoriasis: systematic search and evaluation. JMIR Mhealth Uhealth 2022;10(5):e34017. [doi: 10.2196/34017] [Medline: 35617014]

Abbreviations

ATI: Affinity for Technology Interaction
DiGA: digital health app
G-MAUQ: German mHealth App Usability Questionnaire
HS: hidradenitis suppurativa
MARS: Mobile Application Rating Scale
MARS-G: Mobile Application Rating Scale–German
MDPQ-16: Mobile Device Proficiency Questionnaire
MHA: mobile health app
uMARS: user version of the Mobile Application Rating Scale

Edited by L Buis; submitted 25.11.24; peer-reviewed by F Carrouel, MGC Iwunwa, M Ronicke; revised version received 13.05.25; accepted 22.06.25; published 29.08.25.

Please cite as:
Glatzel C, Dege T, Glatzel B, Goebeler M, Presser D, Schmieder A
Identifying and Evaluating Mobile and Web Apps for Patients to Manage Hidradenitis Suppurativa: Systematic Search in App Stores and Content Analysis
JMIR Mhealth Uhealth 2025;13:e69030
URL: https://mhealth.jmir.org/2025/1/e69030
doi:10.2196/69030

© Caroline Glatzel, Tassilo Dege, Bernadette Glatzel, Matthias Goebeler, Dagmar Presser, Astrid Schmieder. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Application of Behavior Change Techniques and Rated Quality of Smoking Cessation Apps in China: Content Analysis

Qiumian Hong*, BSc; Shuochi Wei*, MSc; Hazizi Duoliken, BSc; Lefan Jin, BSc; Ning Zhang, PhD

School of Public Health and the Second Affiliated Hospital, School of Medicine, Zhejiang University, No. 866 Yuhangtang Road, Hangzhou, China
*these authors contributed equally

Corresponding Author:
Ning Zhang, PhD
School of Public Health and the Second Affiliated Hospital, School of Medicine, Zhejiang University, No. 866 Yuhangtang Road, Hangzhou, China

Abstract

Background: Smoking cessation apps are increasingly being used to help smokers quit smoking. In China, whether behavioral science–based techniques are being incorporated into smoking cessation apps remains unknown.

Objectives: This study aims to describe the usage of behavior change techniques (BCTs) among smoking cessation apps available in China and to evaluate the relationship between BCT utilization and the quality of available smoking cessation apps.

Methods: We searched eligible smoking cessation apps twice on September 12 and October 4, 2022. We coded them with BCTs and assessed their quality by the Mobile App Rating Scale (MARS) and rating score in the App Store. We described the quality of each app (ie, engagement, function, esthetic, and information) and the BCTs used within it, as well as the amount and proportion of all BCTs used. Correlation analysis and linear regression analysis were used to assess the association between the number of BCTs used and the quality of apps.

Results: Nine apps were included in the final analyses. The average number of BCTs being used was 11.44 (SD 2.57), ranging from 5 to 29. Only 1 app used more than 20 BCTs. The most frequently used BCTs were providing feedback on current smoking behavior (9/9, 100%), prompting review of goals (8/9, 88.89%), prompting self-monitoring of one's smoking behavior (7/9, 77.78%), and assessing current and past smoking behavior (7/9, 77.78%). The most commonly used BCTS specifically focus on behavior, including BM (B refers to behavior change, M focuses on addressing motivation; 4.44/11, 40.36%) and BS (B refers to behavior change, S refers to maximizing self-regulatory capacity or skills; 3.78/11, 34.36%). The average score of MARS for the apps was 3.88 (SD 0.38), ranging from 3.29 to 4.46, which was positively correlated with the number of BCTs used (r=0.79; P=.01). Specifically, more usage of BCTs was associated with higher engagement score (β=.74; P=.02; R^2=0.52) and higher information score (β=.76; P=.02; R^2=0.52).

Conclusions: The quality of smoking cessation apps assessed by MARS was correlated with the number of BCTs used. However, overall, the usage of BCTs was insufficient and imbalanced, and the apps demonstrated low quality of engagement and information dimensions. Coordinated efforts from policy makers, technology companies, health behavior professionals, and health care providers should be made to reduce tobacco consumption and to develop high-quality, widely accessible, and effective smoking cessation apps to help smokers quit smoking.

(JMIR Mhealth Uhealth 2025;13:e56296) doi:10.2196/56296

KEYWORDS

smoking cessation; behavior change techniques; mobile application; content analysis; China

Introduction

Tobacco smoking is one of the major public health challenges worldwide [1]. According to the World Health Organization (WHO) Western Pacific [2], the prevalence of smoking in the Chinese adult population was 28% in 2015. In addition, tobacco smoking is a major risk factor for many types of cancer, cardiovascular, and respiratory diseases [3-6], placing a heavy burden on both smokers and health care systems. Smoking cessation is an effective approach to preventing diseases related to smoking and reducing the risk of death due to smoking [7]. Traditional ways to increase smoking cessation mainly include providing advice, taking medicines such as bupropion and varenicline, carrying out cessation counseling, and organizing smoking cessation support groups [8,9].

With the development of mobile health, smoking cessation apps are increasingly being used to help smokers quit [10,11]. The convenience and accessibility of smoking cessation apps make it easier for smokers to monitor and change their smoking behaviors. In China, the widespread adoption of mobile

XSL•FO
RenderX

technology, with about 1.1 billion mobile web users in 2024, has led to the rapid growth of mobile apps. Public awareness of health has increased, leading to greater adoption of health-related apps for services such as telemedicine and chronic disease management (300 billion in 2022 to 365 billion in 2024) [12]. There were several smoking cessation apps available in China [13]. Based on the previous studies, the function and the quality of these apps varied [13-15]. It is important to evaluate these apps with a scientific approach to help smokers choose the most useful and appropriate app and improve the design and development of smoking cessation apps.

The behavior change techniques (BCTs) developed by Michie and colleagues [16] provide a detailed taxonomy of the methods to change behaviors. During the past decade, BCTs have been increasingly used among mobile apps for promoting health behavior change. To explore the components that make a behavior change app effective and assess the reliability of the function, researchers coded behavior change strategies used in apps focused on health behaviors with BCTs, such as physical activity [17], dietary [18], alcohol consumption [19], and emotion management [20].

Furthermore, researchers investigated the impact of BCTs usage on app quality. A study aimed to recommend apps used for health management during pregnancy showed a positive and significant relationship between the number of BCTs used and the quality assessed with the Mobile App Rating Scale (MARS) [21]. Another research in France also demonstrated that the mean MARS score was positively correlated with the amount of BCTs used, but it also reported that fewer popular smoking cessation apps were developed based on BCTs [14]. There have been studies investigating the basic features (such as release date, frequency of downloads, user ratings, and so on) of smoking cessation apps in China [13,22]. However, no previous research has examined the use of BCTs and its relationship to app quality among smoking cessation apps in China. The study aims to fill this gap by describing the number and type of BCTs being used among smoking cessation apps in China and evaluating the relationship between BCT usage and app quality rated by MARS.

Methods

Study Design

We conducted a content analysis of the smoking cessation apps in China. First, one author searched Huawei App Market (Android) and App Store (iOS) twice by keywords (in Chinese, “戒烟,” “吸烟,” and “抽烟”) to initially identify relevant smoking cessation apps. Then, the final app list was selected by 2 researchers (SW and LJ) based on the inclusion and exclusion criteria. Next, 2 researchers independently coded usage of BCTs in the smoking cessation apps based on a taxonomy of smoking cessation BCTs [23] and rated their quality according to the MARS. Two researchers (SW and LJ) had a moderate to high levels of agreement (κ=0.67) on coding the apps. For inconsistent codings, we discussed and consulted the opinion of the corresponding author to reach a consensus for the final coding.

Search Strategy and Data Extraction

According to the latest report from AppInChina, Huawei App Market was the most used Android App Store in China, with a market share of 44.31%. Considering the market share of iPhone in the Chinese smartphone market is 15% - 20%, we searched both on the Huawei App Market (Android) and App Store (iOS). One researcher (SW) searched twice on the 2 app stores on September 12 and October 4, 2022, to make sure that all apps were included. The inclusion criteria were as follows: (1) the app’s aim was to help users quit smoking, and (2) it was available for free download. A total of 69 apps were found. Meanwhile, the researcher collected the rating score (1 - 5 points), number of downloads for Android, and number of ratings for all iOS apps. Then, 2 researchers (SW and LJ) downloaded and tested these apps. Finally, 9 apps were included based on the inclusion criteria. The exclusion criteria were apps that (1) did not support Chinese or did not have a Chinese version, (2) were not designed for smoking cessation only, (3) were irrelevant to smoking cessation, (4) were unable to use because of parameter problems or version incompatibility, and (5) had <10,000 downloads for Android or <50 raters for iOS. The reason for the different criteria for Android and iOS is that the Huawei App Market (Android) and App Store (iOS) have different popularity indicators.

BCTs Coding

We adopted the 43-item taxonomy of BCT for smoking cessation developed by Michie et al [16], which was widely used in studies on smoking cessation interventions [23]. The taxonomy identified 43 BCTs and these techniques could be classified into six functions: (1) directly addressing motivation (BM; B refers to behavior change, M focuses on addressing motivation), (2) maximizing self-regulatory capacity or skills (BS; B refers to behavior change, S refers to self-regulatory capacity or skills), (3) promoting adjuvant activities (A), (4) supporting other BCTs (RD; R relates to general aspects of the interaction, D refers to development of intervention), (5) supporting other BCTs (RI; R relates to general aspects of the interaction, I refers to information gathering), and (6) supporting other BCTs (RC; R relates to general aspects of the interaction, C refers to general communication).

MARS Rating

We adopted the MARS to assess the quality of the apps [24]. This 5-point scale was developed to assess the mHealth app quality with the following subscales: Engagement, Functionality, Esthetics, Information, and a subjective subscale. In reference to previous studies with good consistency [25,26], an overall quality score was calculated by taking the mean of items from the Engagement, Functionality, Esthetics, and Information subscales. Two researchers rated the included smoking apps using this scale together and had a high interclass correlation coefficient (0.85).

Statistical Analysis

The interrater reliability was assessed by interclass correlation coefficient (for MARS) and kappa (for BCT coding). Descriptive analysis was used to describe the characteristics of the apps, including the overall, categorized, and individual BCT usage,

XSL•FO
RenderX

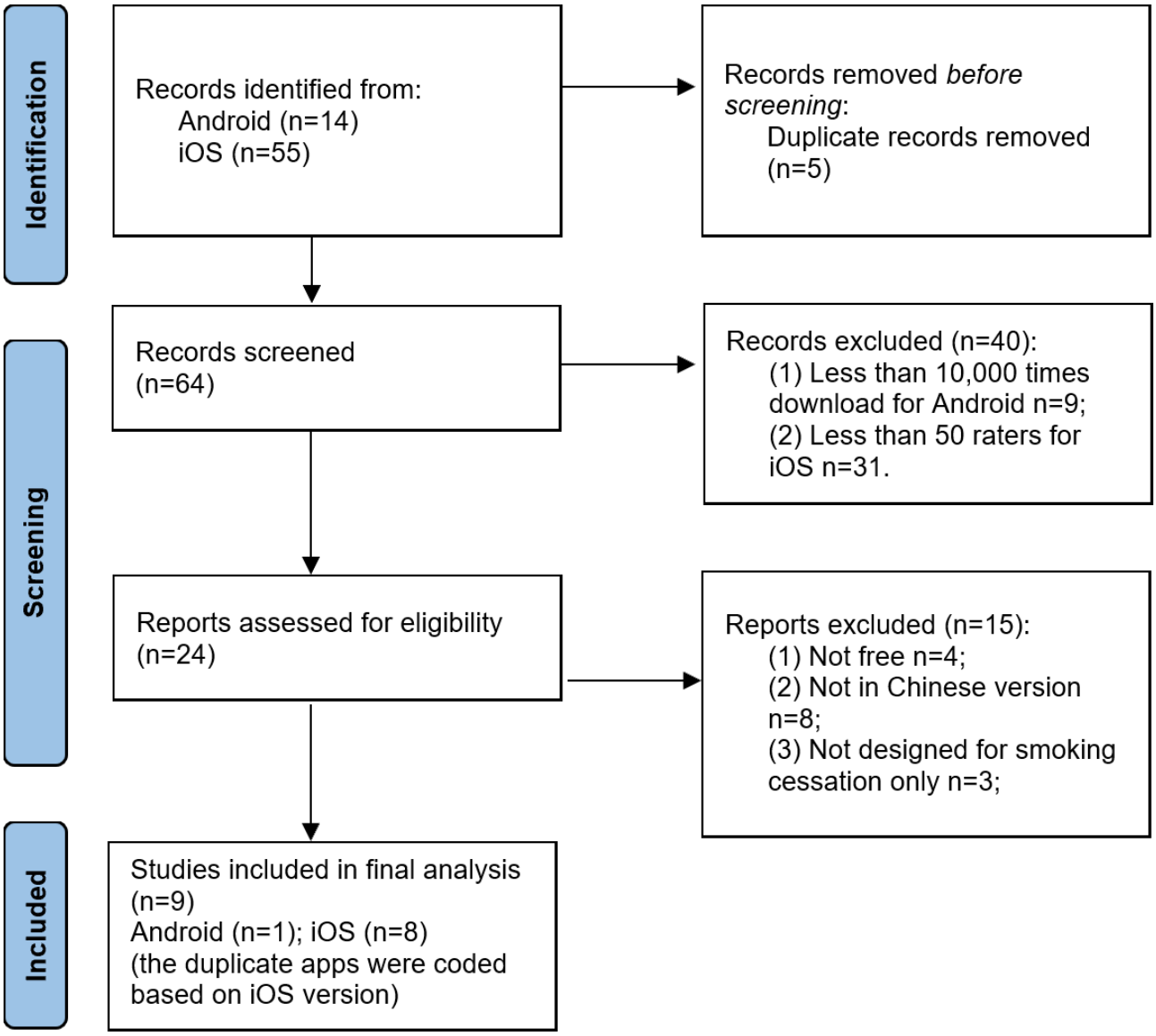
as well as the app's quality (MARS and user-rating scores). Correlation analysis was adopted to assess the relationship between the number of BCTs and the quality of apps (including MARS score and user rating in the App Store). Linear regression analysis was used to investigate whether the number of BCTs used was a significant predictor of the total score and the scores of each dimension of MARS. All the analyses were conducted using SPSS 19.0 (IBM Corporation), with a significance level of α=.05.

Results

Overview

A total of 69 apps were found after 2 rounds of searching. Of them, 20.29% (14/69) were developed for the Android system, 79.71% (55/69) of apps were developed for the iOS system, and 7.24% (5/69) of apps were excluded as they were duplicates in both systems. Of the remaining 64 apps, 55 apps were excluded based on the exclusion criteria (Figure 1), and finally, a total of 9 apps were coded. Detailed information for each app is shown in Table 1.

Figure 1. App screening flowchart.



XSL•FO
RenderX

Table . Detailed information of each app (n=9).

	App name[a]	Number of BCTs[b] used	MARS[c] score					User rating	Number of raters	File size (MB)
			Engagement	Functionality	Esthetics	Information	Total			
1	JYJT	29	4.50	4.75	4.50	4.08	4.46	4.50	2016	24.4
2	WY	16	3.70	4.88	4.50	3.67	4.19	4.80	111	13.5
3	XZJY	14	3.20	4.88	4.17	3.42	3.91	4.90	336	77.2
4	JYYS	12	3.30	4.88	3.67	3.92	3.94	4.80	593	39.3
5	JY	9	3.80	5.00	4.33	3.67	4.20	4.90	309	3.2
6	JRCY	7	3.90	4.25	4.17	3.17	3.87	4.70	120,000	88.3
7	JRJY	6	2.50	4.00	3.50	3.17	3.29	4.80	546	2.5
8	CYYS	5	2.80	4.75	3.83	3.46	3.71	4.70	525	101.7
9	CYJL	5	2.20	4.88	3.00	3.42	3.37	4.60	30,000	16.3

[a]The app name is the pinyin initials of the Chinese name.

[b]BCT: behavior change technique.

[c]MARS: Mobile App Rating Scale.

Usage of BCTs in Smoking Cessation Apps

As shown in Table 2, the average number of BCTs used by the 9 apps was 11.44 (SD 2.57), ranging from 5 to 29. Among the apps, 1 app that used the most BCTs used 29 BCTs, accounting for 67.44% of all BCTs, while 2 apps that used the least amount used only 5 BCTs, accounting for 11.63% of the BCTs for smoking cessation. The items in the BM (focusing on behavior change and addressing motivation) category were the most used BCTs in the apps (4.44/11, 40.36%), followed by the items in the BS (focusing on behavior change and self-regulatory capacity or skills) category (3.78/11, 34.36%). The least used items were those in the interaction category, focusing on the RD aspect (0.22/2, 11%).

Table . Average usage of behavior change techniques in smoking cessation apps in China.

Item	Label	Mean (SD)	Percentage (mean/total)
Average	The average number of BCT[a]	11.44 (2.57)	26.60 (11.44/43)
BM	Specific focus on behavior (B) and addressing motivation (M)	4.44 (2.60)	40.36 (4.44/11)
BS	Specific focus on behavior (B) and maximizing self-regulatory capacity or skills (S)	3.78 (2.82)	34.36 (3.78/11)
A	Promote adjuvant activities (A)	1.00 (1.11)	20 (1/5)
RD	General aspects of the interaction (R) focusing on delivery of the intervention (D)	0.22 (0.44)	11 (0.22/2)
RI	General aspects of the interaction (R) focusing on information gathering (I)	0.78 (0.44)	19.5 (0.78/4)
RC	General aspects of the interaction (R) focusing on general communication (C)	1.22 (1.79)	12.2 (1.22/10)

[a]BCT: behavior change technique.

The use of each BCT item is shown in Table 3. The most-used BCTs were BM3 (provide feedback on current behavior), BS5 (prompt review of goals), BS6 (prompt self-recording), RI1 (assess current and past smoking behavior), BM2 (boost motivation and self-efficacy), and BM6 (prompt commitment from the client there and then).

XSL•FO
RenderX

Table . Behavior change techniques (BCTs) used in each app.

BCT item	Label	App name[a]									Total (n=9)	Percentage
		JYJT (n=29)	WY (n=16)	XZJY (n=14)	JYYS (n=12)	JY (n=9)	JRCY (n=7)	JRJY (n=6)	CYYS (n=5)	CYJL (n=5)		
BM3	Provide feedback on current behavior	1	1	1	1	1	1	1	1	1	9	100
BS5	Prompt review of goals	1	1	1	1	1	0	1	1	1	8	88.89
BS6	Prompt self-recording	1	1	1	0	1	1	1	0	1	7	77.78
RI1	Assess current and past smoking behavior	1	1	1	1	1	1	0	0	1	7	77.78
BM2	Boost motivation and self-efficacy	1	1	1	1	1	0	1	0	0	6	66.67
BM6	Prompt commitment from the client there and then	1	1	1	1	1	0	0	1	0	6	66.67
BS4	Facilitate goal setting	1	1	0	1	1	0	1	0	0	5	55.56
BS9	Set graded tasks	1	1	0	1	1	0	1	0	0	5	55.56
BM1	Provide information on consequences of smoking and smoking cessation	1	1	0	1	0	0	0	0	1	4	44.44
BM7	Provide rewards contingent on effort or progress	1	1	1	1	0	0	0	0	0	4	44.44
BS3	Facilitate action planning/develop treatment plan	1	1	0	1	1	0	0	0	0	4	44.44

XSL•FO
RenderX

BCT item	Label	App name[a]									Total (n=9)	Percentage
		JYJT (n=29)	WY (n=16)	XZJY (n=14)	JYYS (n=12)	JY (n=9)	JRCY (n=7)	JRJY (n=6)	CYYS (n=5)	CYJL (n=5)		
A2	Advise on/facilitate use of social support	1	1	1	0	0	1	0	0	0	4	44.44
A5	Give options for additional and later support	1	1	1	0	0	0	0	1	0	4	44.44
BM4	Provide rewards contingent on successfully stopping smoking	1	1	1	0	0	0	0	0	0	3	33.33
BM5	Provide normative information about others' behavior and experiences	1	0	1	0	0	1	0	0	0	3	33.33
RC1	Build general rapport	1	1	0	0	0	1	0	0	0	3	33.33
RC8	Elicit client views	1	0	1	0	0	1	0	0	0	3	33.33
BM8	Strengthen ex-smoker identity	1	0	0	0	0	0	0	1	0	2	22.22
BM9	Identify reasons for wanting and not wanting to stop smoking	1	0	0	1	0	0	0	0	0	2	22.22
RD2	Emphasize choice	1	1	0	0	0	0	0	0	0	2	22.22
RC5	Offer/direct toward appropriate written materials	1	0	1	0	0	0	0	0	0	2	22.22

XSL•FO
RenderX

BCT item	Label	App name[a]									Total (n=9)	Percentage
		JYJT (n=29)	WY (n=16)	XZJY (n=14)	JYYS (n=12)	JY (n=9)	JRCY (n=7)	JRJY (n=6)	CYYS (n=5)	CYJL (n=5)		
BM10	Explain the importance of abrupt cessation	0	0	0	1	0	0	0	0	0	1	11.11
BS1	Facilitate barrier identification and problem solving	1	0	0	0	0	0	0	0	0	1	11.11
BS2	Facilitate relapse prevention and coping	1	0	0	0	0	0	0	0	0	1	11.11
BS7	Advise on changing routine	1	0	0	0	0	0	0	0	0	1	11.11
BS8	Advise on environmental restructuring	1	0	0	0	0	0	0	0	0	1	11.11
BS11	Advise on avoiding social cues for smoking	1	0	0	0	0	0	0	0	0	1	11.11
A1	Advise on stop-smoking medication	1	0	0	0	0	0	0	0	0	1	11.11
RC6	Provide information on withdrawal symptoms	1	0	0	0	0	0	0	0	0	1	11.11
RC9	Summarize information/confirm on client decisions	0	0	1	0	0	0	0	0	0	1	11.11
RC10	Provide reassurance	1	0	0	0	0	0	0	0	0	1	11.11

[a]The app name is the pinyin initials of the Chinese name.

XSL•FO
RenderX

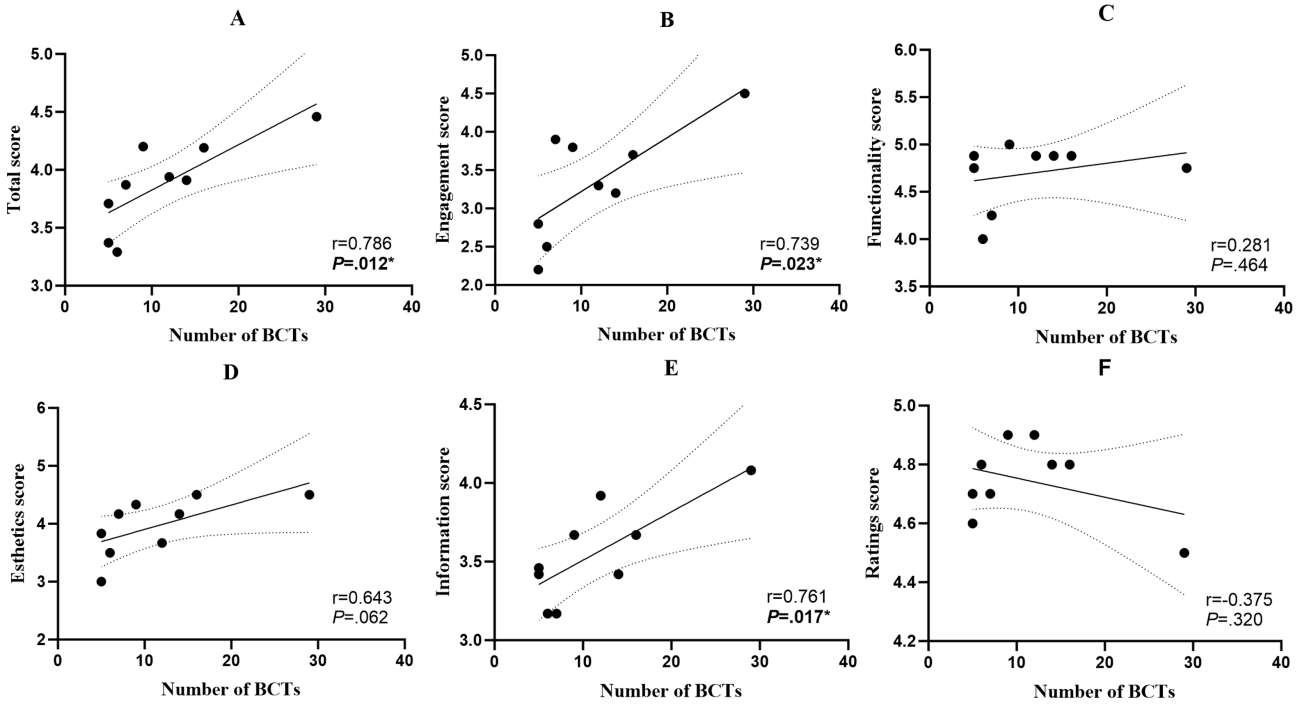
General Quality of the Apps

The quality of apps was assessed by MARS and the rating score in the App Store. The average rating score of the 9 apps in the App Store was 4.74 (SD 0.13) with a range from 4.50 to 4.90. The average score of MARS was 3.88 (0.38) with a range from 3.29 to 4.46. The average scores of Engagement, Functionality, Esthetics, and Information were 3.32 (SD 0.73) (ranging from 2.20 to 4.50), 4.69 (SD 0.33) (ranging from 4.00 to 5.00), 3.96 (SD 0.51) (ranging from 3.00 to 4.50), and 3.55 (SD 0.31) (ranging from 3.17 to 4.08), respectively.

Relationship Between Usage of BCTs and App Quality

The results of the correlation analysis are shown in Figure 2. The number of BCTs was positively correlated with the total score of MARS (Figure 2A; r=0.786; P<.05) and the subscores of Engagement (Figure 2B; r=0.739; P<.05) and Information (Figure 2E; r=0.761; P<.05). There was no significant relationship between the number of BCTs being used and Functionality (Figure 2C; r=0.281; P=.46) or Esthetics (Figure 2D; r=0.643; P=.06) dimension of the apps. Besides, there was no significant relationship between the number of BCTs being used and the rating score (Figure 2F; r=0.375; P=.32).

Figure 2. Correlation analysis between the number of BCTs and (A) the total score, (B) Engagement score, (C) Function score, (D) Esthetics score, (E) Information score, and (F) Rating score of apps. * P<.05. BCT: behavior change technique.



For further verification, we conducted univariate linear regression analysis with the number of BCTs used as the independent variable and the MARS total score, the Engagement score, and the Information score as dependent variables. The results showed that the number of BCTs used in each app significantly predicted its MARS total score (β=.79; P=.01; adjusted R^2=0.56; $F_{1,\ 7}$=11.31, P=.01), its Engagement score (β=.74, P=.02; adjusted R^2=0.48; $F_{1,\ 7}$=8.40, P=.02), and its Information score (β=.76; P=.02; adjusted R^2=0.52; $F_{1,\ 7}$=9.76, P=.02). The amount of BCT used explained 56%, 48%, and 52% of the variation in the MARS total score, the Engagement score, and the Information score, respectively.

Discussion

Principal Findings

To our best knowledge, this was the first study to evaluate the appliance of BCTs and the quality of smoking cessation apps in China. The main findings were as follows: (1) the average number of BCTs used among smoking cessation apps was 11.44, with a range from 5 to 29; (2) the average score of MARS was 3.88, with a range from 3.29 to 4.46; and (3) the number of BCTs was positively correlated with the quality of apps assessed by MARS but not significantly associated with rating scores in App Stores.

The average number of BCTs being used in Chinese smoking cessation apps was 11.44 (SD 2.57), which was lower than that used among smoking cessation apps in other countries. A study reported that the mean number of BCTs was 22 in the French smoking cessation apps [14], which adopted the same BCT taxonomy in its coding process. In the detailed analysis, the included apps contained the BCTs that were mostly in the category of behavior and addressing motivation (BM) and behavior and maximizing self-regulatory capacity or skills (BS), which was in line with the characteristics of mobile apps. App is for users to download and operate independently; therefore, the focus of app design is to stimulate users' motivation and enhance their self-management ability or skills. However, the most used BCT items served only to remind smokers to quit smoking and record smoking behaviors and lacked the assessment of smoking cessation status, explanation of smoking cessation behaviors, or communication and interaction with users, which can be integrated into app functions in the future. Furthermore, it is possible to evaluate the effect of these

XSL•FO
RenderX

functions on improving the quality of apps and even increasing the smoking cessation rate of users.

As our results showed, the general quality of smoking cessation apps was not as high as assessed by MARS. The engagement score was the lowest, which indicated that smoking cessation apps in China may lack promotion or appeal. When the researchers rated the apps, they also found that most apps were not interesting or attractive enough to use, which could affect users' engagement and stickiness. In future developments of smoking cessation apps, integrating personalized content and social features such as web-based communities and chat groups can help users feel that the app better meets their social needs and enhances their emotional attachment with the app. Gamification is also an effective way to increase engagement with apps [27,28]. Meanwhile, the information component scored second lowest among all the rated dimensions, indicating a lack of professional information and adequate data on smoking cessation (the benefits of quitting, potential obstacles and coping strategies, and information about withdrawal symptoms). It is worth noting that none of the apps we coded were evidence-based or validated by evidence or research. Researchers reported similar results when they coded the apps in France [14]. Notably, the relatively high functionality score may be due to the fact that most of the included apps have basic smoking cessation functions and are concise and smooth to use, meeting the scoring rules of this dimension.

We found that there was a positive relationship between the usage of BCTs and the quality of the apps as assessed by MARS. As hypothesized, the more BCTs being used, the higher the quality of smoking cessation apps. Specifically, the number of BCTs being used was significantly positively correlated with the engagement and information score of the apps. According to the MARS scoring rules, the use of BCTs may play a role in improving the pertinence, interactivity, and attractiveness of apps, and may provide guidance for the design of smoking cessation apps. However, we did not find a significant relationship between the usage of BCTs and rating scores in the App Store. For example, JY and XZJY were the 2 highest rating (4.9/5.0) apps but had lower usage of BCTs, and conversely, JYJT got the lowest rating (4.5/5.0) but used the most BCTs (29) and was rated with the highest MARS score (4.46). This finding suggested that ratings in stores were not sufficient indicators of app quality for smokers to choose quitting apps. Thus, to assist smokers select better apps, the App Store rating system should be more detailed, requiring users to rate different aspects of the app, such as functionality, information, and user engagement. In addition, users should compare multiple apps to find the one that best suits their needs and provide specific, timely feedback, such as leaving comments on the app.

Compared with previous studies, the number of smoking cessation apps in China was relatively small. There was a total of 69 apps available in the Huawei App Market (Android) and App Store (iOS), but only 9 apps met the inclusion criteria. A study from the United States searched 273 smoking cessation apps and analyzed 225 of them [29]. Another study from Australia searched 315 apps and analyzed 112 of them [26]. A team from the United Kingdom analyzed 140 apps [30], and a study from France searched 107 apps and analyzed 14 of them [14]. A recent study from Korea finally included 104 apps in analysis [15]. The difference in the number of smoking cessation apps revealed the big gap in the concept and market for smoking cessation or tobacco control between China and other countries. In fact, there is a growing number of studies on developing effective and scalable smoking cessation apps [31]. The behavior change theory, cognitive behavior therapy, and mindfulness training were the most frequently used theories in app development [32-34]. An app was developed under the guidance of BCTs for pregnant smokers [34]. Another app for smoking cessation was developed under the Behavioral Change Taxonomy version 1 [35]. Some apps were designed according to the mindfulness-based intervention [32], and others were designed based on cognitive behavioral therapy [33]. Several Chinese smoking cessation apps were developed using cognitive-behavioral theory [36] and protection motivation theory [37], and studies have demonstrated the feasibility and acceptability of the smartphone app intervention for smoking cessation [22]. However, we did not find them in the App Store, which indicated the need to translate the scientific evidence into public health practice.

In China, the development and promotion of health apps have seen significant progress. The government has actively supported the "Healthy China 2030" initiative [38], providing policy and financial backing for health app innovation. Furthermore, integration with telemedicine services, electronic health records, and artificial intelligence model has improved access to professional advice and streamlined patient care [12]. However, the frequency of use for health-related apps is much lower than that for communication, video, and food delivery apps. Moreover, although the usage of mobile health apps is increasing, there are demographic differences. Younger people and urban residents are more likely to use these apps than older adults and rural populations. Meanwhile, data privacy concerns, inconsistent app quality, a lack of collaboration with professional health care institutions, a lack of standardized regulations, and low user retention still exist and need urgent resolution. According to a report by WHO Western Pacific [2], there is the largest amount of tobacco consumption and number of smokers in China. However, the findings suggested that there were relatively few high-quality smoking cessation apps to help smokers quit. Coordinated efforts from policy makers, technology companies, and health care providers should be made to reduce tobacco consumption and to develop high-quality, widely accessible, and effective smoking cessation apps to help smokers quit in China.

Limitations

There were some limitations in this study. First, the smoking cessation app market is changing and evolving, so this study can reflect only the status of smoking cessation apps so far. With the update and improvement of the app, we believe that the development of smoking cessation apps will be greatly improved. Second, we coded only the presence of the BCTs but did not evaluate or compare the design and presentation forms of the BCTs in each app, nor did we assess the actual usage of each BCTs by users of the smoking cessation apps. These factors were also important influences on the quality of the apps. In the future, researchers could analyze the actual use of BCTs and

XSL•FO
RenderX

the differences between initial design and subsequent usage among smokers. Third, since fewer apps were eventually included in the analysis under strict inclusion and exclusion criteria, regression analysis could not be conducted to explore the effect of using a certain strategy on the app quality. Fourth, using the scores in the App Store as an evaluation metric to explore the relationship between app quality and BCTs usage may yield less accurate results. Ratings may come from individuals who have not fully experienced the app, leading to inaccuracies. Commercial influences can skew ratings, causing them to misrepresent the app's true quality. Furthermore, a single flaw or standout feature can disproportionately affect the overall rating, making it an imperfect indicator of the app's overall quality. Finally, the taxonomy of the BCTs we adopted was not specially designed for the mobile health app evaluation but instead used a taxonomy for smoking cessation. In this case, some of the items do not apply to apps, such as "measuring CO" and "Adopt appropriate local procedures to enable clients to obtain free medication." Future research could develop BCTs that are more appropriate for the design and evaluation of smoking cessation apps.

Conclusions

We searched and screened smoking cessation apps in China and evaluated the BCTs application and quality of the apps. We found that the BCTs were insufficiently used and the quality of the smoking cessation apps in China was generally low. This study provided a snapshot of the smoking cessation apps in China. There is an urgent need to design and develop more effective, attractive, and scalable smoking cessation apps to help smokers quit smoking.

Acknowledgments

This work was supported and funded by National Natural Science Foundation of China (82473741), One Hundred Talents Program Research Initiatives Fund of Zhejiang University, and Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province (2020E10004)

Authors' Contributions

QH participated in the methodology, formal analysis, validation, writing—reviewing and editing, and visualization of this study. SW participated in the conceptualization, methodology, software, formal analysis, investigation, resources, data curation, and writing—original draft of this study. HD participated in the methodology, formal analysis, investigation, resources, data curation, and writing of this study. LJ participated in the formal analysis, resources, and data curation of this study. NZ participated in the conceptualization, writing—review and editing, supervision, project administration, and funding acquisition of this study.

Conflicts of Interest

None declared.

References

1. WHO global report on trends in prevalence of tobacco use 2000-2025. World Health Organization. 2021 Nov 16. URL: https://www.who.int/publications/i/item/9789240039322 [accessed 2025-03-07]
2. The bill China cannot afford: health, economic and social costs of China's tobacco epidemic. World Health Organization. 2012 Mar 31. URL: https://www.who.int/publications/i/item/9789290617907 [accessed 2025-03-07]
3. Aredo JV, Luo SJ, Gardner RM, et al. Tobacco smoking and risk of second primary lung cancer. J Thorac Oncol 2021 Jun;16(6):968-979. [doi: 10.1016/j.jtho.2021.02.024] [Medline: 33722709]
4. Kondo T, Nakano Y, Adachi S, Murohara T. Effects of tobacco smoking on cardiovascular disease. Circ J 2019 Sep 25;83(10):1980-1985. [doi: 10.1253/circj.CJ-19-0323] [Medline: 31462607]
5. Yang JJ, Yu D, Wen W, et al. Tobacco smoking and mortality in Asia: a pooled meta-analysis. JAMA Netw Open 2019 Mar 1;2(3):e191474. [doi: 10.1001/jamanetworkopen.2019.1474] [Medline: 30924901]
6. Yoshida K, Gowers KHC, Lee-Six H, et al. Tobacco smoking and somatic mutations in human bronchial epithelium. Nature New Biol 2020 Feb;578(7794):266-272. [doi: 10.1038/s41586-020-1961-1] [Medline: 31996850]
7. Cao M, Chen W. Epidemiology of lung cancer in China. Thorac Cancer 2019 Jan;10(1):3-7. [doi: 10.1111/1759-7714.12916] [Medline: 30485694]
8. García-Gómez L, Hernández-Pérez A, Noé-Díaz V, Riesco-Miranda JA, Jiménez-Ruiz C. Smoking cessation treatments: current psychological and pharmacological options. Rev Invest Clin 2019;71(1):7-16. [doi: 10.24875/RIC.18002629] [Medline: 30810545]
9. Rigotti NA, Kruse GR, Livingstone-Banks J, Hartmann-Boyce J. Treatment of tobacco smoking: a review. JAMA 2022 Feb 8;327(6):566-577. [doi: 10.1001/jama.2022.0395] [Medline: 35133411]
10. Formagini TDB, Ervilha RR, Machado NM, Andrade BD, Gomide HP, Ronzani TM. A review of smartphone apps for smoking cessation available in Portuguese. Cad Saude Publica 2017 Mar 9;33(2):e00178215. [doi: 10.1590/0102-311X00178215] [Medline: 28300972]
11. Regmi K, Kassim N, Ahmad N, Tuah NA. Effectiveness of mobile apps for smoking cessation: a review. Tob Prev Cessat 2017;3:12. [doi: 10.18332/tpc/70088] [Medline: 32432186]

XSL•FO
RenderX

12. Statistical report on the development of the 54th China internet network. China Internet Network Information Center. 2024 Aug 29. URL: https://www.cnnic.cn/NMediaFile/2024/0911/MAIN1726017626560DHICKVFSM6.pdf [accessed 2024-08-29]
13. Cheng F, Xu J, Su C, Fu X, Bricker J. Content analysis of smartphone apps for smoking cessation in China: empirical study. JMIR Mhealth Uhealth 2017 Jul 11;5(7):e93. [doi: 10.2196/mhealth.7462] [Medline: 28698170]
14. Bustamante LA, Gill Ménard C, Julien S, Romo L. Behavior change techniques in popular mobile apps for smoking cessation in France: content analysis. JMIR Mhealth Uhealth 2021 May 13;9(5):e26082. [doi: 10.2196/26082] [Medline: 33983130]
15. Seo S, Cho SI, Yoon W, Lee CM. Classification of smoking cessation apps: quality review and content analysis. JMIR Mhealth Uhealth 2022 Feb 17;10(2):e17268. [doi: 10.2196/17268] [Medline: 35175213]
16. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
17. Wang Y, Wang Y, Greene B, Sun L. An analysis and evaluation of quality and behavioral change techniques among physical activity apps in China. Int J Med Inform 2020 Jan;133:104029. [doi: 10.1016/j.ijmedinf.2019.104029] [Medline: 31759245]
18. Mahdi S, Michalik-Denny EK, Buckland NJ. An assessment of behavior change techniques in two versions of a dietary mobile application: the Change4Life Food Scanner. Front Public Health 2022;10:803152. [doi: 10.3389/fpubh.2022.803152] [Medline: 35284376]
19. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118. [doi: 10.2196/jmir.4060] [Medline: 25977135]
20. Chan P, Furlonger BE, Leif ES, D'Souza LA, Phillips KJ, Di Mattia M. An evaluation of the behaviour change content and quality of smartphone apps designed for individuals experiencing anxiety: an illustrative example for school psychologists. J Educ Develop Psychol 2022 Jul 3;39(2):209-218. [doi: 10.1080/20590776.2022.2066994]
21. Muñoz-Mancisidor A, Martin-Payo R, Gonzalez-Mendez X, Fernández-Álvarez MDM. Content, behavior change techniques, and quality of pregnancy apps in Spain: systematic search on App Stores. JMIR Mhealth Uhealth 2021 Nov 17;9(11):e27995. [doi: 10.2196/27995] [Medline: 34787587]
22. Liao Y, Tang J. Feasibility and acceptability of a cognitive behavioral therapy-based smartphone app for smoking cessation in China: a single-group cohort study. Front Psychiatry 2021;12:759896. [doi: 10.3389/fpsyt.2021.759896] [Medline: 35309757]
23. West R, Walia A, Hyder N, Shahab L, Michie S. Behavior change techniques used by the English stop smoking services and their associations with short-term quit outcomes. Nicotine Tob Res 2010 Jul;12(7):742-747. [doi: 10.1093/ntr/ntq074] [Medline: 20478957]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
25. Noser AE, Lancaster BD, Hommel KA, et al. Use of behavior change techniques and quality of commercially available inflammatory bowel disease apps. Dig Dis Sci 2023 Jul;68(7):2908-2920. [doi: 10.1007/s10620-023-07884-7] [Medline: 36933116]
26. Thornton L, Quinn C, Birrell L, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
27. Von Bargen T, Zientz C, Haux R. Gamification for mHealth—a review of playful mobile healthcare. Stud Health Technol Inform 2014;202:225-228. [doi: 10.3233/978-1-61499-423-7-225] [Medline: 25000057]
28. Xu L, Shi H, Shen M, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 3;10(2):e27794. [doi: 10.2196/27794] [Medline: 35113034]
29. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736. [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
30. Rajani NB, Weth D, Mastellos N, Filippidis FT. Adherence of popular smoking cessation mobile applications to evidence-based guidelines. BMC Public Health 2019 Jun 13;19(1):743. [doi: 10.1186/s12889-019-7084-7] [Medline: 31196062]
31. Barroso-Hurtado M, Suárez-Castro D, Martínez-Vispo C, Becoña E, López-Durán A. Smoking cessation apps: a systematic review of format, outcomes, and features. Int J Environ Res Public Health 2021 Nov 6;18(21):11664. [doi: 10.3390/ijerph182111664] [Medline: 34770178]
32. Black DS, Kirkpatrick M. Test of daily app-based mindfulness meditation preceding a planned smoking quit attempt date on abstinence: protocol for a randomized controlled trial recruiting across the 58 counties of California. Contemp Clin Trials 2022 Sep;120:106855. [doi: 10.1016/j.cct.2022.106855] [Medline: 35863695]
33. García-Pazo P, Sesé A, Llabrés J, Fornés-Vives J. NoFumo+: a clinical trial of an mHealth for smoking cessation with hospitalized patients. Int J Environ Res Public Health 2021 Oct 6;18(19):10476. [doi: 10.3390/ijerph181910476] [Medline: 34639776]
34. Tombor I, Shahab L, Brown J, Crane D, Michie S, West R. Development of SmokeFree Baby: a smoking cessation smartphone app for pregnant smokers. Transl Behav Med 2016 Dec;6(4):533-545. [doi: 10.1007/s13142-016-0438-0] [Medline: 27699682]

XSL•FO
RenderX

35. van Agteren JEM, Lawn S, Bonevski B, Smith BJ. Kick.it: The development of an evidence-based smoking cessation smartphone app. Transl Behav Med 2018 Mar 1;8(2):243-267. [doi: 10.1093/tbm/ibx031] [Medline: 29447386]
36. Liao Y, Tang J. Efficacy of cognitive behavioural therapy-based smartphone app for smoking cessation in China: a study protocol of a randomised controlled trial. BMJ Open 2021 Jan 13;11(1):e041985. [doi: 10.1136/bmjopen-2020-041985] [Medline: 33441359]
37. Lin H, Lin Y, Zheng Y, Liu Z, Chang C. Design, development and randomised controlled trial of a smartphone application, "QinTB", for smoking cessation in tuberculosis patients: study protocol. BMJ Open 2019 Dec 2;9(12):e031204. [doi: 10.1136/bmjopen-2019-031204] [Medline: 31796480]
38. "Healthy China 2030" initiative. The State Council of the People's Republic of China. 2016 Oct 25. URL: https://www.gov.cn/xinwen/2016-10/25/content_5124174.htm [accessed 2014-09-21]

Abbreviations

BCT: behavior change technique
BM: specific focus on behavior and addressing motivation
BS: specific focus on behavior and maximizing self-regulatory capacity or skills
MARS: Mobile App Rating Scale
RC: general aspects of the interaction focusing on general communication
RD: general aspects of the interaction focusing on delivery of the intervention
RI: general aspects of the interaction focusing on information gathering
WHO: World Health Organization

Edited by L Buis; submitted 12.01.24; peer-reviewed by A Sriram, MDMF Álvarez, R Pryss; revised version received 02.10.24; accepted 26.11.24; published 24.03.25.

Please cite as:
Hong Q, Wei S, Duoliken H, Jin L, Zhang N
Application of Behavior Change Techniques and Rated Quality of Smoking Cessation Apps in China: Content Analysis
JMIR Mhealth Uhealth 2025;13:e56296
URL: https://mhealth.jmir.org/2025/1/e56296
doi:10.2196/56296

©Qiumian Hong, Shuochi Wei, Hazizi Duoliken, Lefan Jin, Ning Zhang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Health Resilience and Well-Being Interventions for Military Members, Veterans, and Public Safety Personnel: Environmental Scan and Quality Review

Rashell R Allen[1,2], MEd; Myrah A Malik[2,3], MSc; Carley Aquin[2,4], BA; Lucijana Herceg[2,4], MEd; Suzette Brémault-Phillips[2,3], PhD; Phillip R Sevigny[1,2,4], RPsych, PhD

[1]School and Clinical Child Psychology, Faculty of Education, University of Alberta, Edmonton, AB, Canada

[2]Heroes in Mind, Advocacy and Research Consortium, Edmonton, AB, Canada

[3]Department of Occupational Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[4]Counselling Psychology, Faculty of Education, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Rashell R Allen, MEd
School and Clinical Child Psychology
Faculty of Education
University of Alberta
1121-87 Avenue
Edmonton, AB, T6G2G5
Canada
Phone: 1 7804920415
Email: wozniak@ualberta.ca

Abstract

Background: Accessible mental health care, delivered via mobile apps or web-based services, may be essential for military members, public safety personnel (PSP), and veterans, as they report numerous barriers to seeking in-person care and are at an increased risk for a number of psychological disorders.

Objective: We aimed to identify, describe, and evaluate apps, resource banks (RBs), and web-based programs (WBPs), referred to as digital mental health interventions (DMHIs), recommended for military members, PSP, and veterans. A multidimensional and multisystemic view of resilience and well-being were maintained throughout this environmental scan.

Methods: Information was gathered from a comprehensive review of peer-reviewed literature, a Google search, and a targeted search of websites relevant to the study populations. DMHIs aimed at supporting resilience or well-being were included in the review, including those published in peer-reviewed articles, and those offered to these populations without research or literature backing their use.

Results: In total, 69 DMHIs were identified in this study, including 42 apps, 19 RBs, and 8 WBPs, and were described based on 3 questions related to purpose, strategies, and evidence from the adapted Mobile App Rating Scale and the Mobile App Rating Scale. Each WBP and RB was then reviewed via the adapted Mobile App Rating Scale and each app via the Alberta Rating Index for Apps (ARIA). Overall, 24 (35%) of the DMHIs were recommended for military members, 20 (29%) for PSP, and 41 (59%) for veterans. The most common aim across apps, RBs, and WBPs was to increase happiness and well-being, and the most common strategies were advice, tips, and skills training. In total, 2 apps recommended for military members—PTSD Coach and Virtual Hope Box—received a high rating on the ARIA subscales and have also been trialed in pilot randomized control trial (RCT) and RCT evaluations, respectively, with positive initial results. Similarly, 2 apps recommended for PSP—PeerConnect and R2MR—have been trialed in non-RCT studies, with partially positive outcomes or little to no contradictory evidence and received a high rating on the ARIA. Finally, 2 apps recommended for veteran populations—PTSD Coach and VetChange—received high ratings on the ARIA and have been trialed via pilot-RCT and RCT studies, respectively, with positive outcomes.

Conclusions: In conclusion, there is a need for efficacy and effectiveness trials for DMHIs for military members, PSP, and veterans to ensure that they are effectively meeting the population's needs. While there appears to be many promising DMHIs, further research is needed before these interventions continue to be promoted as effective and widely distributed.

(JMIR Mhealth Uhealth 2025;13:e64098) doi:10.2196/64098

XSL•FO
RenderX

KEYWORDS

public safety personnel; veteran; military member; web-based program; resources; resilience; mobile app; quality review; well-being; military; environmental; review

Introduction

Background

Military members (eg, Canadian Armed Forces, sailors, soldiers, aviators, and special forces members) and public safety personnel (PSP), or a person who fulfills duties related to public safety [1], show extraordinary commitment to the service and care of Canadians. These individuals, including those involved in emergency response, disaster relief, and national security, are our most critical defenses. Their positions require that they remain operationally ready to respond at a moment's notice to local, provincial, national, and international needs. Similarly, veterans, or those who have served with the military and have transitioned to civilian life, must remain supported after their service. The purpose of this paper is to complete an environmental scan to summarize key information and review the quality of resilience and well-being app- and web-based programming for military members, PSP, and veterans. To do so, population characteristics, the barriers they experience accessing mental health care, and the current landscape of mental health mobile apps, resource banks (RBs), and web-based programs (WBPs), hereafter referred to as digital mental health interventions (DMHIs), must be considered.

Population Characteristics

Occupational duties conducted by military members, PSP, and veterans have several implications on their health and well-being. This can include exposure to potentially psychologically traumatic events (PPTEs), such as sudden violent death, sudden accidental death, serious transportation accident, and physical assault [2,3]. Carleton et al [3] reported that, for PSP, exposure to PPTEs was associated with positive screens for a variety of psychological disorders, including posttraumatic stress disorder (PTSD), depression, anxiety, and substance use.

In addition, military members are at increased risk for exposure to PPTEs [4]. Notably, Iraq and Afghanistan war veterans are likely to report lifetime PTSD, substance use disorders, and nonsuicidal self-injurious behaviors, as well as current alcohol use disorder, substance use disorder, and suicidal ideation [5]. In a population-based sample of US veterans, 86.12% (2719/3157) reported exposure to at least one PPTE, with a mean of 3.4 (SD 2.8) different PPTEs in their lifetime [6]. The number of PPTEs experienced have been shown to exacerbate PTSD symptoms, which is related to greater cognitive difficulties, increased loneliness, and lower functional social support [7]. Understanding occupational and operational PPTEs and their impact is critical when working with military members, PSP, and veterans.

Barriers to Mental Health Care

Access to care is essential for populations that experience an increased risk of exposure to PPTEs and mental and physical health difficulties. There are several barriers to accessing and seeking care at both the individual and systems levels. From a strictly physical perspective, Canada is a vast country and access to services may vary depending upon where one lives. For example, while there may be a multitude of service options available in large urban settings, the same is not necessarily true for those living in rural or remote areas. On a psychological level, stigma is one of the most common barriers for military members, PSP, and veterans seeking care. PSP often attribute perceived or experienced stigma as a barrier not only because of the psychological disorder itself but because of care-seeking and the cause of the injury, complicating the notion of when care-seeking is acceptable [8]. Some PSP view individuals who are seeking support as "claiming" their mental injuries and "milking the system" for unwarranted personal gain, burdening their colleagues, organizations, and taxpayers by seeking care or speaking up about their mental health difficulties [8].

Stigma is also a perceived barrier to care for veteran and military member populations [4,9,10]. A population-based sample of US veterans found that only 12% reported engagement in mental health care [10]. Of 4069 veterans sampled, 924 (22.7%) met criteria for 1 or more psychological disorders, with 73.1% (675/924) of these individuals reporting no current engagement in treatment [10]. These data are similar for military members, with estimates of 23% to 40% of individuals who screened positive for a mental disorder seeking care [4]. Those who met criteria for a mental disorder were found to be twice as likely to report concern for possible stigmatization as a barrier to seeking care [4]. Some key stigma-related barriers to seeking care identified by military member respondents who met criteria for a mental disorder included: the belief that they will be seen as weak, belief that their unit leadership might treat them differently, and belief it would cause members of their unit to have less confidence in them [4]. Hoge et al [4] found that 55.1% (354/642) of military members in their sample who met the criteria for a mental disorder believed they would have difficulty getting time off work for treatment. Those with negative beliefs about mental health care are also less likely to seek care for their mental health needs [9]. Given that military members, PSP, and veterans are at increased risk of PTSD, mood, anxiety, and substance use disorders [3,6,11], it is paramount to consider factors or alternative treatment modalities that increase their access to support and services.

Well-Being and Resilience

A holistic well-being approach to working with these populations is essential considering the importance of well-being on a person's ability to adapt to stress and PPTEs [12]. Research has shown that the domains of well-being interact and affect other areas of well-being, as well as impact one's resilience response [12,13]. Our current review maintains a multidimensional perspective of well-being, which considers the degree to which individual, dyadic, community, and organizations' needs are satisfied [14,15]. For example, enhanced community well-being has been shown to lead to an increased likelihood of a resilience response [16,17] and has

XSL•FO
RenderX

potential positive effects across other domains of well-being and systems surrounding the individual [18,19].

Access to holistic care is essential to promoting resilience. Resilience is understood by the authors to be a system's process of adaptation (eg, individuals, groups, and organizations) following adversity or risk exposure. This multisystemic perspective views resilience as a dynamic process influenced by socioecological system interactions, available resources, and individual qualities, which function as protective factors and increase the likelihood of coping with and overcoming adversity [20-22].

Background on DMHIs

Asynchronous interventions, available on the web or via mobile phone, may offer increased access to care and potentially decrease military members', PSP, and veterans' experience of stigma. In response to the COVID-19 pandemic, there was a notable increase in the development of technologies in response to the needs of the public [23]. Digitally delivered mental health interventions via the internet or mobile app are relevant for purposes of this paper. Most internet-based interventions are psychological treatments that are disorder-specific and are typically grounded in cognitive behavioral therapy (CBT) [24]. Potential benefits of asynchronous self-guided internet-based interventions include that they are easily accessible, anonymous, and may reach populations that may not otherwise seek treatment [24]. A systematic review and meta-analysis (k=11, 12 comparisons) by Díaz-García et al [25] reported nonsignificant changes in direct and proximal resilience measures for internet-based interventions (effect size for resilience was g=0.12, *P*=.32). Kuester et al [24] conversely reported that internet-based interventions using a cognitive-behavioral therapy framework (k=7) were effective compared to passive control conditions in reducing PTSD symptoms (g=0.72, *P*<.001). When compared to active control conditions, there was no significant difference for changes in PTSD symptoms [24]. Further research is yet needed to further evaluate such programs.

Mobile apps appear to be a less effective form of treatment modality. In a review of meta-analyses, pooled effect sizes across 4 meta-analyses reported small-to-moderate results (g=0.28-0.38), with the small or nonsignificant effect sizes for intervention versus active controls (g=0.17-0.21) [23]. A meta-analysis found that smartphone interventions for depression (k=6, g=0.33, *P*=.005), anxiety (k=6, g=0.30, *P*=.15) suicidal ideation (k=4, g=–0.14, *P*=.25), self-injurious behaviors (k=3, g=–0.04, *P*=.75), smoking (k=3, g=0.39, *P*.001), drinking (k=3, g=–0.03, *P*=.77), sleep problems (k=2, nonpooled effects g=0.72 and 0.84, *P*<.05), and PTSD (k=2, g=–0.05 and 0.15, *P*>.05) resulted in small or nonsignificant changes compared to controls [26]. Caution was also noted for the moderate to high heterogeneity between trials for anxiety, depression, and substance use (smoking and drinking) [26]. Donker et al [27] also highlighted mixed results in terms of 5 mobile apps targeting depression, anxiety, and substance use. Due to the quality of these studies and the risk of bias, Donker et al [27] cautioned against the results reported by these studies. Overall, the within-group effect sizes ranged from –0.45 to 2.28 following the intervention, and 0.45-2.11 after follow-up. Some mobile apps were compared to control groups (ie, a CBT computer program, attentional control, and attentional control plus data summaries and meeting with general practitioner), and the between group effects ranged from –0.14 to 0.25 following the intervention, and –0.28 to 0.58 after follow-up [27]. Finally, a review of mental health mobile apps found that there is little evidence to suggest that mobile-based interventions are helpful, and that some are harmful [28].

Several limitations have been reported regarding app use to address mental health concerns. Many mobile apps have not been subjected to research validation and have privacy and confidentiality concerns [29], with evidence suggesting that the effectiveness and efficacy of mobile apps are questionable [23,26-29]. Skorburg and Yam [23] highlighted further concerns that the apps may exacerbate health inequalities. Veterans living in rural locations particularly report that apps are hard to navigate, and their use is impacted by financial and connectivity limitations [30]. Attitudes of app use in veteran populations also appear extreme, being either strongly positive or strongly negative [30]. Lack of awareness of the apps and low rate of veteran app use was also noted to be common [31].

Objectives

This environmental scan aimed to provide a review and quality assessment of DMHIs, including apps (ie, mobile apps), RBs (ie, websites with resources and information), and WBPs (ie, interactive programs available on the web), recommended for military members, PSP, and veterans. Apps and WBPs are similar in that they are interactive resources that may include modules, questionnaires, audio or video information, and intervention-specific activities. RBs may not be interactive in nature, but they were included in this review as they provide the target populations with valuable information related to resilience and well-being. In addition, the review was meant to be as inclusive as possible in identifying resources available to these populations and assessing the quality of such resources. Therefore, all available resources for the target populations, including apps, RBs, and WBPs, were examined.

Environmental scanning is the acquisition and use of information about trends and relationships in the environment to determine information needs and use [32]. The objectives of this project were to (1) conduct an environmental scan of well-being and resilience DMHIs available in Canada for military members, PSP, and veterans and (2) review the quality of the available programs. Information was gathered through an iterative search of peer-reviewed literature (ie, a scoping review search), a Google search, and a targeted search of websites relevant to the study populations. As DMHIs are widely available and accessible, regardless of the evidence to support their use, it is essential to review all DMHIs to evaluate their quality, including usability, privacy, functionality, and information quality. DMHIs vary in terms of their interactive or static nature of their delivery format. Therefore, we aimed to assess both the DMHIs that are more interactive in nature (ie, apps and programs) and those that are more static (ie, resources). Therefore, this environmental scan will add to the academic literature by reviewing DMHIs

XSL•FO
RenderX

with and without peer-reviewed literature backing their use and evaluating those that are available in Canada.

Methods

Overview

This environmental scan used 3 steps, including app identification, description, and evaluation. Step 1 involved several methodologies to first identify relevant DMHIs: (1) an iterative search of all available peer-reviewed literature, conducted in February 2024; (2) a Google search, conducted in June 2023; and (3) a targeted web search of 12 websites, completed in August 2023 (eg, Veterans Affairs Canada website). Once identified, step 2 included assessing each DMHI via questions on the adapted Mobile App Rating Scale (A-MARS) [33] and the Mobile App Rating Scale (MARS) [34] to evaluate DMHI purpose, strategies, and evidence base. In step 3, each app was evaluated with the Alberta Rating Index for Apps (ARIA) [35], and each RB and WBP was assessed with the A-MARS [33]. While there is no registered protocol for the current project, the information provided in the methods may be used to replicate the current search. A description of the methods associated with each step is provided in subsequent sections.

Step 1: Identification of DMHIs

Search Strategy and Information Sources

The literature search was used to gather key information to determine the depth and breadth of peer-reviewed literature related to DMHIs well-being and resilience resources for military members, PSP, and veterans, with key terms related to (1) population (eg, military members), (2) resilience and well-being related constructs (eg, hardiness), and (3) web- or mobile-based programs (eg, mobile apps; an example of a full search string is provided in Multimedia Appendix 1). The final search was conducted using a Boolean format of the following databases: Academic Search Complete, CINAHL, APA PsychInfo, Embase, SocINDEX, and MEDLINE. The Google search was conducted using key terms based on the same 3 concepts, but because Google limits the characters for each search, multiple searches were conducted (full search string is provided in Multimedia Appendix 2). Finally, the targeted website search included websites developed for each population (eg, Veterans Affairs), and each website was thoroughly searched for any well-being or resilience resources recommended for their members (the list of targeted websites is provided in Multimedia Appendix 3).

Eligibility Criteria

Apps, RBs, and WBPs included in this study encompassed resources that were aimed at supporting resilience and well-being in military members, PSP, and veterans (Textbox 1). While the focus of our study was to review resources relevant to military members, PSP, and veterans, DMHIs developed for and tested with the general population were incorporated if they were also recommended for (but not necessarily trialed with) one of these populations. In total, 4 researchers (RRA, MAM, CA, and LH) were involved in the eligibility assessment and selection process. Only DMHIs that were free (or had free components), available in Canada, and available on Apple, Google Play, and on the web without enrollment access met the criteria. At all levels of screening, a minimum of 2 researchers reviewed each app and website independently. All apps, programs, and websites with discrepant ratings were reviewed independently by a third reviewer. If necessary, the resource was discussed in a research team meeting before final eligibility decisions were made. Each DMHI was grouped by type of resource (ie, app, RB, or WBP) for the remaining steps: description and quality assessment.

Textbox 1. Eligibility criteria for digital mental health interventions (DMHIs).

Inclusion criteria

- Self-directed DMHIs meant to improve well-being or resilience
- The DMHI was recommended for or developed to support military members, public safety personnel, or veteran populations
- DMHI was available on the Apple or the Google Play Store or free to access on the web (or some features were accessible at no cost)

Exclusion criteria

- The DMHI included a guided table 1(synchronous) support component for the intervention (eg, in-person or virtual check-ins [or sessions], therapist interaction, virtual reality, or in-person meetings or discussions)
- DMHI was not available in the English language or not available in Canada
- DMHI required enrollment into the program or access code

Step 2: Description of DMHIs

Once eligibility was determined, each DMHI was reviewed and data were extracted and recorded by a minimum of 2 researchers to ensure relevant information was included in the review (eg, the description of the resource and the population the resource was developed, trialed, tested, or recommended for). With environmental scans being largely descriptive in nature and aimed at capturing the current state of the literature, the authors collected information about the type of research available (ie, no research and level of evidence), as well as the focus and strategy of each DMHI. These were then used to describe each DMHI and to describe the intervention focus, the strategies commonly used, and the overall state of the evidence.

The classification section of the MARS [34] was used to describe the focus and strategies used by each DMHI. The focus question evaluated what the intervention targeted and included

XSL•FO
RenderX

13 items in which the researchers selected all that applied, including whether the DMHI targeted (1) to increase happiness or well-being, (2) mindfulness or meditation or relaxation, (3) to reduce negative emotions, (4) depression, (5) anxiety or stress, (6) anger, (7) behavior change, (8) alcohol or substance use, (9) goal setting, (10) entertainment, (11) relationships, (12) physical health, and (13) other (specify). Next, the researchers selected all that applied in terms of the theoretical background and strategies used by each app, RB, and WBP, including (1) assessment, (2) feedback, (3) information or education, (4) monitoring or tracking, (5) goal setting, (6) advice or tips or strategies or skills training, (7) CBT—behavioral, (8) CBT—cognitive, (9) acceptance and commitment therapy, (10) mindfulness or meditation, (11) relaxation, (12) gratitude, (13) strengths-based, and (14) other (specify).

A question on the A-MARS [33] related to evidence base was used to describe the depth and breadth of the evidence for all DMHIs. Each DMHI was summarized based on this question to determine whether the app or electronic tool (e-tool) had been trialed or tested and verified by evidence in published scientific literature. Each app, RB, and WBP was classified into 1 of 6 groups based on their evidence in the literature:

- It has not been trialed or tested.
- The evidence suggests the app or e-tool does not work.
- App or e-tool has been trialed (eg, acceptability, usability, and satisfaction ratings) and has partially positive outcomes in studies that are not randomized controlled trials (RCTs), or there is little or no contradictory evidence.
- App or e-tool has been trialed (eg, acceptability, usability, and satisfaction ratings) and has positive outcomes in studies that are not RCTs, and there is no contradictory evidence.
- App or e-tool has been trialed and outcome tested in 1 to 2 RCTs indicating positive results.
- App or e-tool has been trialed and outcome tested in >3 high quality RCTs indicating positive results.

Step 3: Evaluation of DMHIs

Each DMHI was evaluated based on 2 quality rating scales: the A-MARS and the ARIA. A description of each is provided in the subsequent sections.

Adapted MARS: RBs and WBPs

The A-MARS [33] is a rating scale adapted from the MARS [34] and was used to review RBs and WBPs. The A-MARS was developed to evaluate health-related e-tools, with a specific expansion of the engagement subscale. The A-MARS is a 29-item scale rated on a scale from 1 (*inadequate*) to 5 (*excellent*), with the following subscales: engagement (5 items), functionality (4 items), aesthetic (3 items), information (6 items), subjective quality (4 items) and health-related quality (6 items). The subscale items were summed and averaged for a subscale score, then the engagement, functionality, esthetics, and information subscales were summed and averaged for a mean quality score, and finally, all subscales were summed and averaged for an overall mean total score. The A-MARS total mean score is a reflection of the overall quality of the e-tool, whereas the subscale scores and mean quality score describe the specific strengths and weaknesses of the e-tool.

For all subscales of the A-MARS, intraclass correlation coefficient (ICC) estimates and their 95% CIs were calculated using SPSS statistics (version 29; IBM Corp) based on a mean rating (k=2), absolute-agreement, 2-way mixed-effects model. For RBs, the reliability of raters fell in the excellent range (ie, ICC>0.90), with the exception of the aesthetic domain (ICC=0.73; 95% CI 0.33-0.90) [36]. For WBPs, the reliability of raters fell in the excellent range (ICC>0.90) for all domains [36].

The ARIA: Apps

The ARIA [35] served as a measure to evaluate the quality of eligible apps. For this study, the care provider version was used, with 2 sections: part A and part B. Part A was completed before downloading the app and was based on the information on the app store page. Part A was used to assess goal fit, trustworthiness, privacy, and affordability. Next, part B was completed after 2 researchers used the app independently for at least 10 minutes to assess quality related to security, trustworthiness, ease of use, functionality, fit for population, usefulness, and satisfaction [37]. The original ARIA scale was on a scale from 0 to 4, but for the purpose of this review, and to remain consistent with the A-MARS, the scale was changed so each item in part A and part B were rated on a scale of 1 (*strongly disagree*) to 5 (*strongly agree*). Subjective quality of each app was also assessed, and these items included “I would recommend using this app to the user” and “the number of stars that best represents your overall rating for the quality of this app is,” on a scale from 1 (*strongly disagree; 1 star, worst app*) to 5 (*strongly agree; 5 stars, best app I have every used*).

After completing part A and B, the scores for each item were added up for a total score out of 30 for part A, and 60 for part B. A higher score in part A indicates that the app fits the purpose of the project, is trustworthy, has adequate privacy, and is affordable [37]. A higher score in part B indicates better quality of content and usability of the app [37]. ICC estimates and their 95% CIs were calculated using SPSS statistical package version 29 based on a mean rating (k=2), absolute-agreement, 2-way mixed-effects model. The ICC for part A (ICC=0.86; 95% CI 0.75-0.92) and part B (ICC=0.51; 95% CI 0.07-0.75) indicate good and moderate reliability across raters, respectively [36].

Results

Step 1: Identification of DMHIs

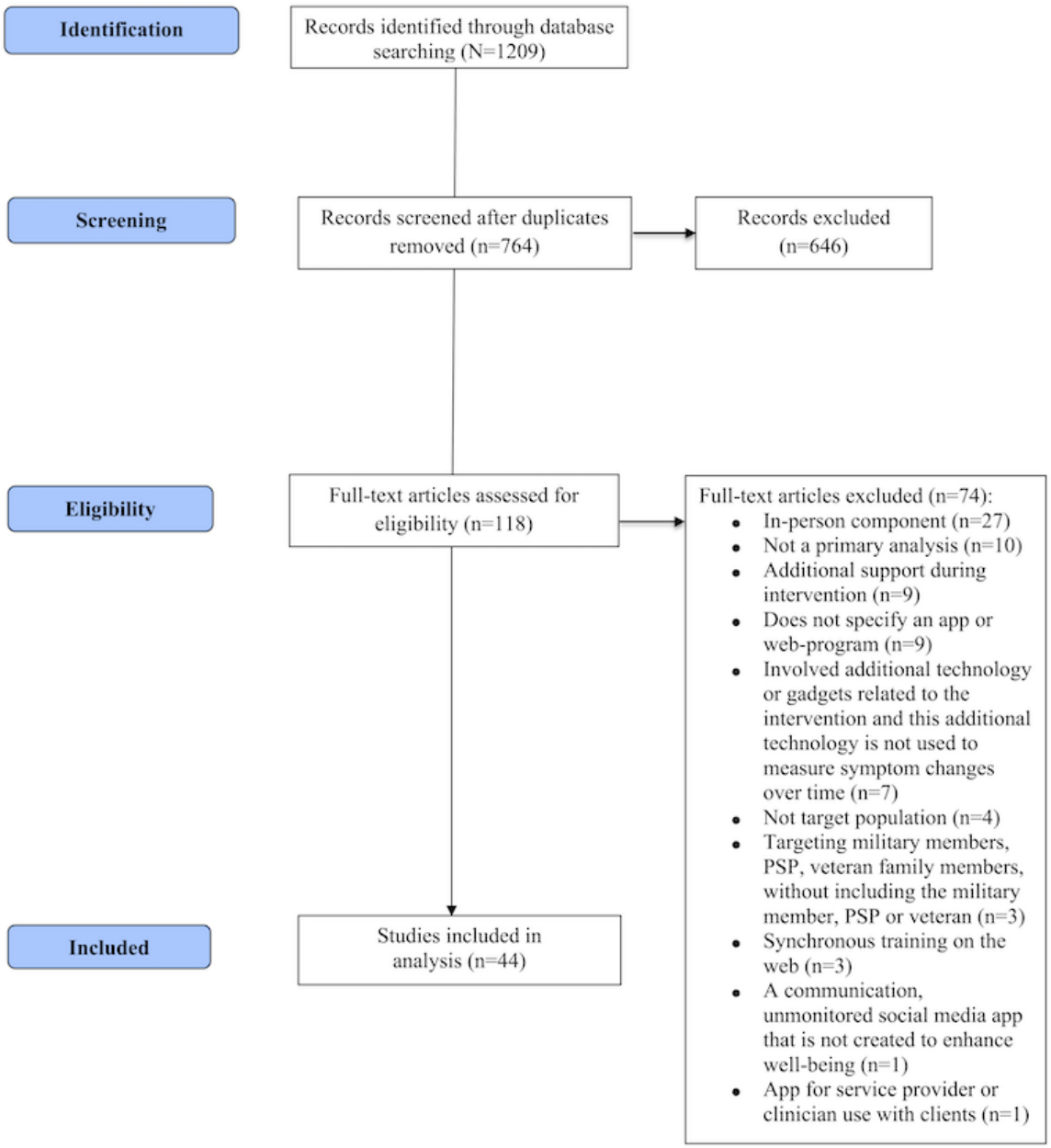
Literature Review Article Identification

The initial review of the literature yielded a total of 1209 articles, with 764 duplicates, resulting in 646 articles. In total 3 researchers (RRA, MAM, and CA) completed initial reviews of titles and abstracts identified during the literature review, with a range of agreement (Cohen k) from moderate (k=0.48) to substantial (k=0.79). The authors included all articles with disagreement in the full-text review to allow a more comprehensive evaluation of the discrepancies. A total of 118 articles met the criteria for full-text review. The agreement for

XSL•FO
RenderX

full-text reviews was fair (k=0.20) to substantial (k=0.79). In the end, 44 articles were included for the DMHI screening. Figure 1 illustrates the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) literature review flowchart.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flowchart. PSP: public safety personnel.

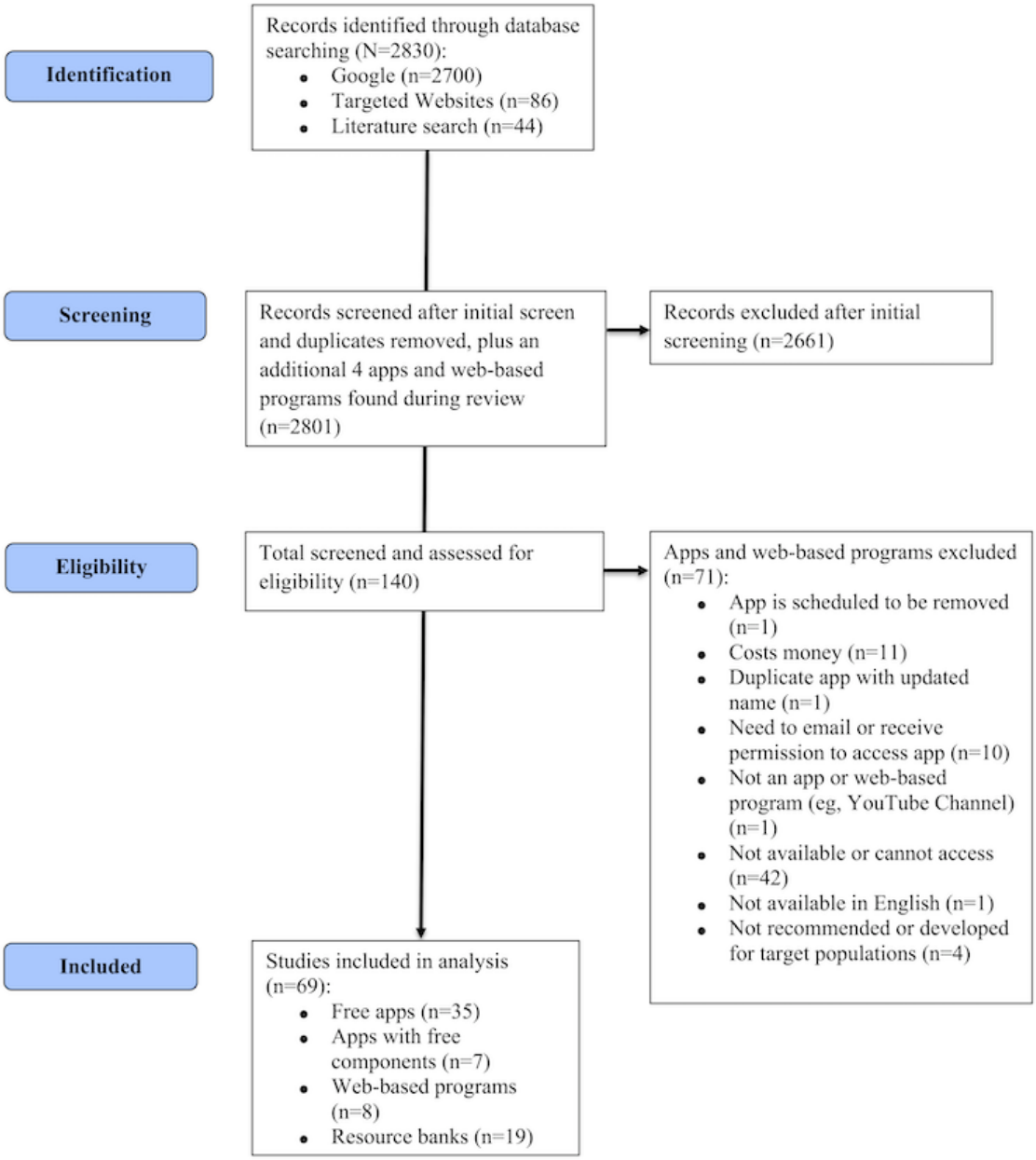


DMHI Identification

To review DMHIs available to the target populations in Canada, a comprehensive Google search and targeted web search were also conducted. In the end, the literature search yielded 44 digital wellness resources, the Google search yielded a total of 2700 Google records, and the targeted web search yielded 12 websites with 86 digital wellness resources. Each DMHI and resource was reviewed by a minimum of 2 reviewers, and discrepancies were rectified by a third reviewer or group discussion (authors involved in review included RRA, MAM, CA, and LH). Details of the review and screening process for accessible DMHIs are provided in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (Figure 2). After the initial screening of the literature search, Google search, and targeted web search, 140 apps, RBs, and WBPs were identified for further review. In total, 2 authors (RRA and MAM) conducted a further review to determine if the app or website was accessible and available (ie, available in Canada, free [or free components], and no sign-up required). A total of 69 relevant digital wellness programs were identified and included in this study, including 42 apps (7 apps with free components), 8 WBPs, and 19 web-based RBs.

XSL•FO
RenderX

Figure 2. Environmental scan flowchart.



Step 2: Description of DMHIs

The following results are a descriptive summary of all information gathered via the A-MARS and MARS. Key information was also recorded for each included program (eg, population recommended for) and is summarized in subsequent sections.

Characteristics of DMHIs

Overview

Most of the programs (41/69, 59%) focused on targeting veteran populations, followed by 35% (24/69) of programs recommended for military members, and 29% (20/69) of programs recommended for PSP. A small portion of programs (2/69, 3%) were recommended for military member and veteran families (Table 1). The aim of this project was to assess DMHIs available to military members, PSP, and veteran populations in Canada. Although a majority of DMHIs were developed outside of Canada, with 10% (7/69) DMHIs originating in Canada, the information, resources, tools, and activities provided by the DMHIs appeared to transcend national borders. In addition, it appeared that all DMHIs were developed without targeting a specific sex or gender, or it was unclear whether the intervention was developed for a specific gender.

XSL•FO
RenderX

Table 1. Digital mental health interventions key information.

Digital mental health resource name	Type of resource	Location the digital mental health intervention was created	Suggested population
7-Minute Chi[a,b,c]	App	Belgium	Veterans
AboutFace[c]	Resource bank	United States	Veterans
ACT Coach[d]	App	United States	Veterans
AIMS for Anger Management[c]	App	United States	Military members and veterans
AIMS for Anger Management[e]	Web-based program	United States	Military members and veterans
Beyond MST[b,c]	App	United States	Military members and veterans
Breathe2Relax[a,e]	App	United States	Military members
Calm[b,e]	App	United States	Public safety personnel
CBT-Insomnia Coach[e]	App	United States	Veterans
Chill Drills[c]	App	United States	Military members and veterans
Chris Germer Meditations[c]	Resource bank	United States	Veterans
Comfort Talk Pro[e]	App	United States	Veterans
Couples Coach[e]	App	United States	Veterans
COVID Coach[e]	App	United States	Military member, PSP, and veterans
CPT Coach[e]	App	United States	Veterans
CrewCare[c]	App	United States	Public safety personnel (fire)
Daily Yoga[b,e]	App	China	Military members
Drinks:Ration[a,f]	App	United Kingdom	Military members and veterans
Driven Resilience[b,c]	App	Australia	Public safety personnel and military member
Equipt[c]	App	Canada	Public safety personnel (police)
eQuoo[a,d]	App	United Kingdom	Veterans
Exalted Warrior Foundation[c]	Resource bank	United States	Veterans
First Responders First[c]	Resource bank	Canada	Public safety personnel
FOCUS on the Go![a,c]	App	United States	Military member families and veteran families
Freedom Qigong[c]	Web-based program	United Kingdom	Veterans
Head to Health[c]	Resource bank	Australia	Military members
HeadFIT[d]	Web-based program	United Kingdom	Military members
Insight Timer[a,b,d]	App	United States	Public safety personnel
Insomnia Coach[d]	App	United States	Veterans
Lighthouse Health and Wellness[c]	App	United States	Public safety personnel
Manage Stress: VA National Center for Health Promotion and Disease Prevention[c]	Resource bank	United States	Veterans
Meditation Oasis Podcasts[c]	Resource bank	United States	Veterans
Meditation Rx[a,e]	App	United States	Veterans

XSL•FO
RenderX

Digital mental health resource name	Type of resource	Location the digital mental health intervention was created	Suggested population
Mind Resilience[c]	Resource bank	United Kingdom	Public safety personnel
Mindarma[a,c]	App	Australia	Public safety personnel
Mindfulness Coach[d]	App	United States	Public safety personnel and veterans
Mindshift[d,g]	App	Canada	Public safety personnel
Misadventures in Money management[c]	Web-based program	United States	Military members
MOVE! Coach[c]	App	United States	Military members and veterans
National Sleep Foundation[c]	Resource bank	United States	Veterans
NHS Every Mind Matters[c]	Resource bank	United Kingdom	Military members
OSI Connect[e,g]	App	Canada	Veterans
Pain and Opioid Safety[c]	Resource bank	United States	Military members
Pain eHealth for Activity, Skills, and Education[f]	Resource bank	United States	Veterans
PeerConnect[e]	App	United States and Canada	Public safety personnel
PE Coach 2[e]	App	United States	Military members
Provider Resilience[c]	Web-based program	United States	Public safety personnel
PTSD Coach[e]	App	United States	Veterans
PTSD Coach Canada[c]	App	Canada	Military members and veterans
PTSD Family Coach[e]	App	United States	Military members (and families) and veterans (and families)
R2MR[a,e]	App	Canada	Military members, Public safety personnel, and veterans
Responder Strong[c]	Resource bank	United States	Public safety personnel
Shield of Resilience Training[c]	Web-based program	United States	Public safety personnel
Simply Yoga[b,e]	App	United States	Military members
STAIR Coach[e]	App	United States	Veterans
Stand Down: Think Before You Drink[a,b,e]	App	United States	Veterans
Stay Quit Coach[e]	App	United States	Veterans
Substance Abuse and Mental Health Services Administration Disaster App[c]	App	United States	Public safety personnel
SwapMyMood[c]	App	United States	Military members and veterans
Tactical Breather[e]	Web-based program	United States	Military members
Tao Connect[c]	Web-based program	United States	Public safety personnel
Ten Percent Happier[c]	Resource bank	United States	Public safety personnel
VA Make the Connection[c]	Resource bank	United States	Veterans
VA National Center for PTSD[c]	Resource bank	United States	Veterans
VA Public Health[c]	Resource bank	United States	Veterans
VetChange[f]	App	United States	Veterans
Veterans Yoga Project[c]	Resource bank	United States	Veterans

XSL•FO
RenderX

Digital mental health resource name	Type of resource	Location the digital mental health intervention was created	Suggested population
Virtual Hope Box[f]	App	United States	Military members
Yoga Journal[c]	Resource bank	United States	Veterans

[a]Available on Apple App Store only.

[b]App with free component available, but add-ons that cost money

[c]App or electronic tool (e-tool) has not been trialed or tested.

[d]App or e-tool has been trialed (eg, acceptability, usability, and satisfaction ratings) and has positive outcomes in studies that are not randomized controlled trials (RCTs), and there is no contradictory evidence.

[e]App or e-tool has been trialed (eg, acceptability, usability, and satisfaction ratings) and has partially positive outcomes in studies that are not RCTs, or there is little or no contradictory evidence.

[f]App or e-tool has been trialed and outcome tested in 1 to 2 RCTs indicating positive results.

[g]Available on Google Play only.

A summary description of the 8 WBPs, 19 web-based RBs, and 42 apps (7 apps with free components) can be found in the subsequent section.

Description of WBPs

Of the 8 WBPs that met the criteria, their purpose varied (the purpose and theoretical orientation for each WBP is provided in Multimedia Appendix 4). Notably, most WBP had multiple areas of focus for both the purpose and theoretical orientation. Therefore, the following totals and percentages represent this overlap and the general representation of each domain across all WBPs. In total, 75% (6/8) of the WBPs focused on increasing happiness and well-being. Next, 50% (4/8) focused on mindfulness, meditation, or relaxation, 38% (3/8) aimed to reduce anxiety and stress, 38% (3/8) focused on reducing negative emotions, and 25% (2/8) aimed to facilitate behavioral change. Finally, a few of the WBPs aimed to support anger management (1/8, 12%), financial management (1/8, 12%), goal setting (1/8, 12%), physical health (1/8, 12%), relationships (1/8, 12%), and resilience (1/8, 12%).

In terms of theoretical background or strategies used, 66% (5/8) of the WBPs focused on advice, tips, strategies, or skills training, 50% (4/8) on information or education, 50% (4/8) on mindfulness or meditation, 38% (3/8) on CBT (behavioral) techniques, 38% (3/8) on CBT (cognitive) techniques, and 38% (3/8) on relaxation strategies. Other theoretical background or strategies highlighted by WBPs included goal setting (2/8, 25%), gratitude modules (2/8, 25%), monitoring or tracking (2/8, 25%), or strengths-based techniques (1/8, 12%). The A-MARS evidence base review demonstrated a lack of evidence for WBPs for military members, PSP, and veterans, with 75% (6/8) of them not being trialed or tested, and only 25% (2/8; AIMS for Anger Management and HeadFIT) of them having been trialed by studies that are not RCTs, with partial positive outcomes, or little contradictory evidence.

Description of RBs

Similar to WBPs, each RB had multiple areas of focus for both the purpose and theoretical orientation. The following totals and percentages represent this overlap and the representation of domains across RBs. Of the 19 RBs included in the study, 42% (8/19) aimed to increase happiness or well-being; 42% (8/19) to facilitate mindfulness, meditation, or relaxation; 37% (7/19) to reduce negative emotions; and 37% (7/19) to support physical health (the purpose and theoretical orientation for each RB is provided in Multimedia Appendix 5). Less commonly, RBs aimed to support anxiety or stress (3/19, 16%), behavior change (2/19, 10%), relationships (2/19, 10%), alcohol or substance use (1/19, 5%), resilience (1/19, 5%), or sleep (1/19, 5%).

In terms of theoretical background or strategies, 58% of RBs (11/19) used advice or tips or strategies or skills training, 58% (11/19) used information or education, and 42% (8/19) used mindfulness or meditation techniques. The remaining and less common theoretical background or strategies included relaxation techniques (5/19, 26%), monitoring or tracking (4/19, 21%), goal setting (3/19, 16%), assessment (2/19, 10%), gratitude (2/19, 10%), CBT (behavioral) skills (1/19, 5%), or CBT (cognitive) skills (1/19, 5%). The A-MARS review of the evidence base for the RBs revealed that 18 (95%) have not been trialed or tested. One RB (Pain eHealth for Activity, Skills, and Education) has been trialed and outcome tested in 1 to 2 RCTs and reported positive results, but as this program requires access to be granted, our review was based on the “resources” tab of their website and not the program itself.

Apps

Each app also had multiple areas of focus for both the purpose and theoretical orientation. The following section represents this overlap, and the domains emphasized across all apps. Of the 42 reviewed apps, 42% (18/42) focused on increasing happiness and well-being (the purpose and theoretical orientation for each app is provided in Multimedia Appendix 6). In addition, 36% (15/42) aimed to support mindfulness, meditation, or relaxation, 31% (13/42) focused on reducing negative emotions, 21% (9/42) aimed to support behavior change, 21% (9/42) focused on reducing anxiety or stress, 17% (7/42) targeted relationship support, and 17% (7/42) aimed to improve physical health. Fewer programs targeted alcohol or substance use (4/42, 9%), PTSD (5/42, 12%), anger (4/42, 9%), sleep (2/42, 5%), depression (4/42, 9%), and goal setting (1/42, 2%).

In terms of the 42 apps’ theoretical background or strategies used, 62% (26/42) used advice, tips, strategies, or skills training, 60% (25/42) used information or education, 43% (18/42) used monitoring or tracking, 26% (11/42) used relaxation strategies, 24% (10/42) used mindfulness or meditation, and 24% (10/42)

XSL•FO
RenderX

used goal setting. Less common theoretical background or strategies used included the following: acceptance and commitment therapy (1/42, 2%), assessment (6/42, 1%), CBT (cognitive) skills (7/42, 2%), CBT (behavioral) skills (7/42, 2%), cognitive processing therapy (CPT; 1/42, 2%), feedback (3/42, 7%), gratitude (1/42, 2%), prolonged exposure (PE; 1/42, 2%), resilience (1/42, 2%), skills training (1/42, 2%), and family resilience training (1/42, 2%). On the basis of the A-MARS evidence base review, the efficacy and effectiveness of apps are still in its infancy. In total, 33% (14/42) apps have not been trialed or tested. A total of 45% (19/42) of apps have been trialed by studies that are not RCTs with partially positive outcomes, or little or no contradictory evidence. Around 14% (6/42) of apps have been trialed by studies that are not RCTs, with positive outcomes and no contradictory evidence. Finally, 7% (3/42) of apps have been trialed and outcome tested in 1 to 2 RCTs with positive results reported.

Step 3: Evaluation of DMHIs

Step 3: Procedures for Quality Review

Each RB and WBP was then evaluated via the A-MARS, and each app via the ARIA. A summary of key themes for each survey is described in the subsequent sections.

A-MARS: Quality Review of WBPs and RBs

Overview

All mean totals reported (ie, subscale mean, quality mean, and total mean scores) are based on the average ratings provided by 2 researchers (RRA and MAM) with the highest possible mean score being 5.0 (the mean scores for all domains for each e-tool are provided in Multimedia Appendix 7).

A-MARS Engagement

A higher average rating on the engagement subscale indicates greater levels of engagement, interest, customization, interactivity or interoperability, and target group fit. The average engagement mean rating was 2.93 (SD 0.52), ranging from 2.10 to 4.47. In total, 4 e-tools received an average score for engagement 1 SD above the mean—2 RBs (VA National Center for PTSD and VA Public Health) and 2 WBPs (AIMS for Anger Management and Misadventures in Money Management).

A-MARS Functionality

A higher score on the functionality subscale indicates high performance, ease of use, navigation, and design. This subscale received the highest overall ratings across all subscales with an average score 4.03 (SD 0.52), ranging from 3.00 to 4.88. The highest scoring e-tools (1 SD above the mean) for functionality were 3 WBPs (AIMS for Anger Management, Shield of Resilience Training, and Tactical Breather).

A-MARS Aesthetic

The aesthetic subscale represents the layout, graphics, and visual appeal of the e-tool, with a higher mean score representing a higher aesthetic rating. The average aesthetic mean rating was 3.62 (SD 0.40), ranging from 3.00 to 4.33. For aesthetics, 3 e-tools received an average score 1 SD above the mean—1 RB (About Face), and 2 WBPs (AIMS for Anger Management and HeadFIT).

A-MARS Information

The information subscale evaluates whether the e-tool contains high quality information (eg, text, feedback, measures, and references) from a credible source. A higher score for the information subscale indicates greater quality of information, greater quantity of information, precise goals, clear visual information, high credibility of sources, and strong evidence base. The average information subscale score was 3.54 (SD 0.73) and ranged from 1.67 to 4.80. The highest scoring e-tools (1 SD above the mean) for information were 3 RBs (Manage Stress: VA National Center for Health Promotion and Disease Prevention, National Sleep Foundation, and VA National Center for PTSD).

A-MARS Quality

The A-MARS quality subscale consists of the average rating across the engagement, functionality, aesthetic, and information subscales. Therefore, a higher score indicates higher overall quality without raters' subjective quality and health-related information ratings. Average quality scores ranged from 2.51 to 4.19, with an average rating of 3.53 (SD 0.44). In terms of e-tool quality, 3 e-tools received an average rating 1 SD above the mean, 2 were WBPs (AIMS for Anger Management; Misadventures in Money Management), and one was an RB (Manage Stress: VA National Center for Health Promotion and Disease Prevention).

A-MARS Subjective Quality

Subjective quality subscale score was determined by assessing whether the researchers would recommend the tool, how many times they thought they would use the tool, whether they would pay for the tool, and their overall rating of the tool. A higher score represents a higher subjective quality rating. This subscale received the overall lowest ratings, with an average of 2.53 (SD 0.60), ranging from 1.75 to 4.25. There were 2 WBPs that scored 1 SD above the mean for the subjective quality domain (AIMS for Anger Management and Misadventures in Money Management).

A-MARS Health-Related Quality

Finally, the health-related quality subscale was assessed by rating the e-tools on subject matter related to whether there were additional resources provided, other strategies recommended for the user, multiple solutions offered for the presenting issue, multiple symptoms addressed, opportunities for real time tracking, and obvious access to health-related help. A higher score indicates greater health-related quality. This subscale had an average rating of 3.24 (SD 0.98) and ranged from 1.58 to 4.67. Within this domain, 1 WBP (AIMS for Anger Management) and 4 RBs (Manage Stress: VA National Center for Health Promotion and Disease Prevention, NHS Every Mind Matters, VA Make the Connection, and VA National Center for PTSD) received an average score 1 SD above the mean.

A-MARS Total

The A-MARS total mean score is a reflection of the overall quality of the RB and WBP (ie, e-tool), with an average score across all e-tools of 3.32 (SD 0.49), ranging from 2.38 to 4.11. There were 5 e-tools that scored 1 SD above the mean, 2 WBPs (AIMS for Anger Management and Misadventures in Money

XSL•FO
RenderX

Management), and 3 RBs (Manage Stress: VA National Center for Health Promotion and Disease Prevention, VA Make the Connection, and VA National Center for PTSD).

ARIA: Quality Review of Apps

ARIA Part A

Part A was completed before downloading the app and assessed based on the app store page to evaluate goal fit, trustworthiness, privacy, and affordability. The average rating between the 2 researchers was 24.39 (SD 1.85) and ranged from a total score of 21 to 29 (the mean scores for each section of the ARIA for each app are provided Multimedia Appendix 8). In total, 7 apps received an average score 1 SD above the mean (Couples Coach, CPT Coach, Meditation Rx, OSI Connect, PeerConnect, PE Coach 2, and SwapMyMood).

In terms of individual item analysis, the researchers felt that the app fits the user's (ie, military member, PSP, and veteran) purpose, with the average rating of 3.9 (SD 0.34), ranging from 3 (*neutral*) to 5 (*strongly agree*). The largest range was 2.0 (disagree) to 5.0 (strongly agree) on an item assessing whether the user can trust that relevant experts in the field developed the app, with an average rating of 3.6. Most apps on the app store page did not include a statement about the risks associated with using the app, with an average rating of 2.5 (SD 0.75), a range of 2 to 4.50, and 27 (64%) receiving a score of 2 (disagree). In terms of declaring a conflict of interest, only 1 app (*Mindarma*) provided a conflict of interest statement. Most apps (36/42, 86%) had a privacy policy that explained what information is collected by the app, who will have access to this information, and how the information will be used.

ARIA Part B

Part B was completed after using each app for a minimum of 10 minutes, using all links, and checking all sounds, videos, and tools associated with the app. A higher score on part B indicates a higher quality app related to trustworthiness, security, ease of use, functionality, usefulness, and satisfaction. Overall, the average rating was 43.38 (SD 3.92) and ranged from 36 to 50. In total, 6 apps scored 1 SD above the mean (Couples Coach, PTSD Coach, PTSD Family Coach, R2MR, VetChange, and Virtual Hope Box).

ARIA Subjective Quality

The subjective quality of each app was assessed via 2 items. First, 2 researchers rated whether they would recommend the app, with an average rating of 3.73 (SD 0.52), ranging from 2.5 to 5. There were 5 apps that received an average score 1 SD above the mean (AIMS for Anger Management, FOCUS on the Go!, Insomnia Coach, PTSD Family Coach, and R2MR).

The researchers also rated the number of stars that they felt represented the overall quality of the app. The average rating for this item was 3.08 (SD 0.69) and ranged from 1.5 to 4.5. There were 8 apps that scored 1 SD above the mean (AIMS for Anger Management, CBT-Insomnia Coach, Couples Coach, COVID Coach, FOCUS on the Go!, Mindfulness Coach, R2MR, and Virtual Hope Box).

Discussion

Principal Findings

This environmental scan was conducted to review, describe, and qualitatively evaluate DMHIs recommended for military member, PSP, and veteran populations. This information is vital as high quality, accessible care for military members, PSP, and veterans is sorely needed [38]. This environmental scan reviewed 42 mobile apps, 19 RBs, and 8 WBPs, totaling to 69 DMHIs.

According to our review, 55% (38/69) of the apps, RBs, and WBPs have not been trialed or tested, while 30% (21/69) of the apps and WBPs have been trialed by studies that are not RCTs (eg, acceptability, usability, and satisfaction ratings) with partially positive outcomes or little to no contradictory evidence. In total, 10% (7/69) of the apps and WBPs have been trialed by studies that are not RCTs (eg, acceptability, usability, and satisfaction ratings) and have positive outcomes with no contradictory evidence. Finally, 6% (4/69), 3 apps (Drinks:Ration, Virtual Hope Box, and VetChange), and 1 RB (Pain eHealth for Activity, Skills, and Education), have been trialed and outcome tested in 1 to 2 RCTs and recorded positive results. Although Pain eHealth for Activity, Skills, and Education is a WBP that has been trialed and tested, but as this program requires access to be granted to the user, our review was based on the "resources" tab of their website, and not the program itself.

DMHI Summary

Summary Based on Aim, Strategy, Quality, and Evidence

Across apps, RBs, and WBPs, the most common aim was to increase happiness and well-being, and the most common strategies used were advice, tips, strategies, and skills training. In terms of the highest rated DMHIs, AIMS for Anger Management received the highest quality mean score and the highest total mean score (A-MARS), and PTSD Family Coach was rated the highest for overall quality (ARIA part B). SwapMyMood received the highest rating for overall fit, trustworthiness, privacy, and affordability (ARIA part A). In total, 2 e-tools received the highest score on 2 A-MARS subscales: Manage Stress: VA National Center for Health Promotion and Disease Prevention (information and health-related quality) and Misadventures in Money Management (engagement and subjective quality). No app, RB, or WBP has been trialed and outcome tested in more than 3 high quality RCTs with positive results.

Summary Based on Population

Summary Section Organization

The following section will be used to discuss the quality and usability of the highly rated DMHIs (ie, DMHIs that received a rating 1 SD above the mean) on the A-MARS [33] and the ARIA [35] and organized by the 3 populations of interest. Each of these highly rated DMHIs will be organized and synthesized based on the A-MARS levels of evidence [33], and the Canadian Psychological Association Task Force's [39] recommendations

XSL•FO
RenderX

for evidence-based practice (EBP) will be used as a framework to interpret these results.

Military Members

In total, 2 apps received a high rating on ARIA subscales with promising initial results, including PTSD Coach, which has been trialed for military member or veteran populations in pilot RCT and non-RCT satisfaction and useability studies, and Virtual Hope Box, which has been evaluated in proof-of-concept and RCT studies. PTSD Coach (also for veteran populations) is an app developed to support individuals with PTSD through psychoeducation and self-management tools and received a high rating on ARIA part B, which indicates high levels of trustworthiness, ease of use, usefulness, and user satisfaction. Virtual Hope Box, designed to support stress management and emotion regulation, has promising evidence and also scored highly on the ARIA part B. These apps are recommended for military members as they have been trialed through initial studies and received high ratings related to the overall quality of the apps; however, multiple efficacy trials have not been conducted to date.

One app, Drinks:Ration (also for veteran populations), was also trialed with 1 or 2 RCT studies, with positive initial results. Although this is the case, this app did not receive a high rating (1 SD above the mean) on any of the ARIA subscales, suggesting that this app is likely efficacious at supporting individuals in managing and monitoring alcohol consumption, but, in terms of user satisfaction and app quality (as assessed by the ARIA), this app comparatively did not score in the upper percentile. In addition, the ARIA assesses constructs related to app security, privacy, and trustworthiness, which may not impact the overall app efficacy. Therefore, this app is recommended to military members as a helpful app for monitoring and potentially reducing their drinking, and future research is suggested to improve app security, privacy, functionality, and trustworthiness.

Next, there were a number of DMHIs recommended for military members that have trailed initially (eg, acceptability studies), have partially positive outcomes (in non-RCT studies), and have little or no contrary evidence. AIMS for Anger Management (also for veteran populations) is a WBP developed to provide individuals with tools to track and manage anger, with information about anger, and with opportunities for support. AIMS for Anger Management received high scores across many domains, suggesting that this WBP is highly engaging, interesting, aesthetically pleasing, functional, and high quality. Next, HeadFIT, a WBP developed to support individuals with stress management skills, received a high rating on the aesthetic subscale indicating that it is aesthetically pleasing; however, it did not rank highly in other areas related to quality or information. PE Coach 2, an app with the goal of providing psychoeducation about PTSD-related symptoms and coping tools, scored highly on the ARIA part A, suggesting that it is trustworthy, has adequate privacy, and is affordable. PE Coach was assessed via RCT; however, the intervention condition included multiple apps and therefore it is difficult to discern the efficacy of PE Coach on its own. Next, PTSD Family Coach (also for veteran populations and their families) was developed to support family members of those living with PTSD, and has been assessed with a pilot RCT, feasibility, and acceptability study. This app received a high rating on ARIA part B indicating that the app is functional, useful, and satisfactory. Finally, R2MR (also for PSP and veteran populations) was rated highly for ARIA part B. This app was developed, as an adjunct to in-person R2MR training, to support short-term performance and long-term mental health outcomes. These apps are recommended as potentially helpful. Importantly, a stance of EBP must continue to be maintained, and further research is needed on each of these DMHIs to suggest them as effective tools for military members [39].

Finally, of the DMHIs rated highly on either the ARIA or A-MARS subscales, there were 4 that have not been trialed and tested. First, Misadventures in Money Management received high ratings related to engagement and overall quality. Second, NHS Every Mind Matters was rated positively for health-related information. Third, Tactical Breather received a high rating related to its ease of use. Importantly, Tactical Breather has been assessed via RCT, however, the intervention condition included the use of multiple other apps; therefore, it is difficult to discern and determine the efficacy of this app. Finally, SwapMyMood (also for veteran populations) was rated highly regarding its fit for the population, trustworthiness, privacy, and affordability. These DMHIs appear engaging, secure, and informative, and were rated highly in terms of user experience. While these are all important considerations, these DMHIs cannot be definitively recommended to military members until more evidence, in particular effectiveness and efficacy trials, support their use with this population [39].

Public Safety Personnel

In total there were 3 DMHIs that received high ratings on the ARIA or A-MARS subscales that are recommended for and trialed with PSP populations. PeerConnect was developed to provide mental health service to support wellness and belonging. This app has a high rating on the ARIA part A suggesting that it provides information on app security and privacy and appears to be a good fit for PSP. PeerConnect has promising preliminary evidence such that it has been trialed with PSP populations in non-RCT studies (eg, satisfaction and acceptability), with partially positive outcomes or little or no contradictory evidence. R2MR (also for military member and veteran populations), meant to support short-term health and long-term mental health, received high ratings on the ARIA part B (see Military Members section) and has similar positive preliminary evidence to PeerConnect. Finally, Shield of Resilience Training has not been trialed or tested for PSP but received a high rating on functionality suggesting that it is easy to use and navigate. Overall, there are few DMHIs recommended for or trialed with PSP populations that also received a high rating. There are no DMHIs for PSP that received a high rating and have a strong evidence base. As per the Canadian Psychological Association Task Force's recommendations, more research is needed before specific DMHIs can be recommended for PSP [39], and this is a notable gap as PSP do not have many high quality DMHIs developed for them.

XSL•FO
RenderX

Notably, there were 3 apps (Insight Timer, Mindfulness Coach, and Mindshift) that were noted for higher levels of evidence (ie, have been trialed in non-RCT trials and have positive outcomes with no contradictory results) but did not receive high ratings on the ARIA. These apps were made for the public, and therefore, were only recommended as potential resources for PSP populations. In addition, these apps have not been studied specifically with PSP populations and instead have been recommended for these populations and trialed with the general population. These apps with promising evidence may have had lower ratings because the measures used to assess these apps generally evaluate app satisfaction, usability, and security or privacy, and do not assess app effectiveness.

Veterans

In total, 16 DMHIs received high ratings on either the A-MARS or ARIA for veteran populations. Around 2 apps received high ratings related to ARIA part B, indicating that they are secure, trustworthy, usable, and functional, and have also been trialed via pilot RCT, RCT, and non-RCT (eg, usability, satisfaction, effectiveness trials) studies, with positive initial outcomes. These apps include, first, PTSD Coach (also for military members, assessed via pilot RCT), which supports individuals in managing PTSD symptoms, and second, VetChange, an app developed to reduce drinking and to support PTSD symptoms (evaluated by one RCT with military members indicating positive results). These apps are recommended for veteran populations based on preliminary research with positive results and based on high ratings related to user experience. Again, although they have positive research outcomes reported in preliminary RCT studies, in line with EBP, additional studies to assess internal and external validity are necessary to effectively address the empirical evidence supporting these programs [39].

In total, 7 DMHIs scored highly on subscales related to the ARIA and A-MARS and have also been trialed via non-RCT studies (eg, acceptability) with partially positive outcomes or little to no contradictory evidence: (1) AIMS for Anger Management (also for military members) was developed to provide users with tools to track and manage their anger, with education about anger, and with opportunities to find support. This WBP received high ratings on A-MARS engagement, functionality, aesthetic, quality, subjective quality, and health-related quality, suggesting that it is of high quality; (2) Couples Coach, an app designed to support partners in improving their relationship and exploring new ways for connection, scored high on the ARIA part A and part B. This suggests that it is secure, private, affordable, usable, and functional; (3) CPT Coach is an app developed to support individuals participating in CPT by providing worksheets, readings, and symptom monitoring tools. This app received a high score on ARIA part A, suggesting it is affordable, secure, and private; (4) Meditation Rx, an app designed to support stress reduction through guided meditations, also received a high score on ARIA part A (eg, security and privacy); (5) OSI Connect, an app with the aim of teaching self-management tools to assist with common mental health concerns, also received a high score on ARIA part A (eg, security and privacy); (6) PTSD Family Coach (also for military members) aims to support family members of those with PTSD (evaluated by pilot RCT, feasibility, and acceptability trial). This app received a high score on ARIA part B, suggesting it is secure, functional, and user friendly; (7) R2MR (also for military members and PSPs), an app developed to support (adjunct to in-person training) short-term performance and long-term mental health received a high score on ARIA part B, suggesting that it is also secure, functional, and user friendly. Overall, these DMHIs for veteran populations have preliminary evidence supporting their use, with partially positive outcomes or little to no contradictory evidence. On the basis of our review of the user experience, these DMHIs received the highest rating and may be used by veteran populations. However, in line with EBP, these DMHIs require further research, including effectiveness and efficacy trials [39].

In total, 7 DMHIs recommended for veteran populations do not appear to have evidence supporting their use and have not been trialed or tested: (1) AboutFace received high ratings by the researchers on aesthetics; (2) Manage Stress: VA National Center for Health Promotion and Disease Prevention received high ratings across A-MARS subscales (information, quality, health-related quality, and total score); (3) National Sleep Foundation scored highly on information; (4) SwapMyMood (also for military members) received a high score related to the ARIA part A, which indicates that it is affordable, secure, and private; (5) VA Make the Connection scored highly on health-related quality and total score; (6) VA National Center for PTSD scored highly across many A-MARS subscales, including engagement, information, health-related quality, and total score; and (7) VA Public Health scored highly on the engagement subscale. Although these DMHIs received high quality ratings based on their usability, experience, and information, in line with EBP [39], these must be trialed and tested before they are recommended for veteran populations.

Limitations and Strengths of the Environment Scan

Some methodological limitations exist in this environmental scan. Only DMHIs that were in English and available in Canada were included, which may limit the scope of this review. A further potential limitation is that there was no clear definition for RBs and WBPs in literature. Instead, the research team engaged in discussions to differentiate RBs and WBPs from other web-based programs (eg, web-based courses, chat rooms, Facebook groups, and YouTube channels). The research team decided a web-based program was an *interactive web-based platform* that does not involve an invitation or log-in code, and an RB was a web-based platform that was not interactive but provided *key information* or *links to the user*. In the future, it would be beneficial to have a clear definition and to differentiate between the aforementioned web-based resources.

In addition, many of the DMHIs have not been formally evaluated to determine their fit with the target populations, or to determine the efficacy or effectiveness at supporting users' well-being or resilience. Therefore, these are subjective ratings and assessments of fit and quality and do not represent the effectiveness or efficacy. The research team that completed this review has worked closely with military members, PSP, and veteran populations; however, they are not members of these populations. This limitation may have impacted the subjective

XSL•FO
RenderX

quality rating and fit for population rating, as it was the researchers' perspectives and not target population members' perspectives.

There are also several strengths related to this environmental scan. An extensive search of 6 databases, gray literature (ie, Google search), and 12 targeted websites were completed, strengthening the scope of this review. We also established clear inclusion and exclusion criteria that were refined throughout. This project also applied a whole-person and multidimensional well-being approach to evaluate the programs. This framework prioritizes multiple dimensions of wellness to emphasize one's well-being within and across domains, allowing researchers and clinicians to focus on protective networks rather than deficit-based frameworks. Finally, scales developed for the purpose of reviewing app and e-tool quality were used to organize and critically analyze the included DMHIs.

Future Directions

Overall, our findings are consistent with Miralles et al [29], who suggested that, for mobile apps, research is lacking, that they are not empirically validated interventions, and that there are privacy and confidentiality concerns. There has been a surge of DMHIs because of the COVID-19 pandemic, and the current evidence base for these interventions reflects this novelty. It is imperative that researchers aim to continue to push for EBP, even for self-managed DMHIs. Broadly, the evidence base for DMHIs for military members, PSP, and veterans is weak, with over half of them not having been trialed or tested, and no program being assessed with 3 or more RCTs with positive results. Efficacy and effectiveness trials are key to progressing EBP of DMHIs, and therefore, must be prioritized [39].

The current literature is yet in the early stages, with many studies focusing on acceptability, usability, or initial pilot studies. An important next step in establishing the evidence for DMHI will be conducting carefully controlled RCTs to demonstrate the efficacy of their use. In addition, developers must continue to consider the population characteristics and treatment preferences of those they are creating the program for [39]. It is key that future DMHI developers and researchers maintain a multidimensional well-being approach to support military members, PSP, and veterans across well-being domains, and to support their resilience [16,17].

Conclusions

This environmental scan contributes to the literature by collecting and reviewing available DMHIs intended to promote resilience and well-being in military members, PSP, and veterans. The purpose of this study was to determine the quality of DMHIs. Using an environmental scan methodology, we were able to capture the breadth and depth of literature and available resources (without research literature) to review and categorize the available DMHIs recommended or developed for these populations. Although the results of this environmental scan highlight key apps, WBPs, and RBs that are higher in quality, there is little research that supports their effectiveness and efficacy. Further research is needed before DMHIs can be confidently promoted and widely distributed.

Acknowledgments

This research has been funded by the government of Alberta, Mental Health and Addiction COVID-19 Community Funding Grant to support the authors in their time committed to the review.

Authors' Contributions

RRA maintained a leadership role in developing this project and collaborated with the remaining authors in the conceptualization of the manuscript. RRA took leadership and MAM assisted in writing the literature review and the first draft of the manuscript. RRA, MAM, CA, and LH were involved in the data search, review, and eligibility assessment. RRA and MAM completed the survey reviews of the digital mental health resources. SB-P and PRS both took leadership roles in reviewing and editing the manuscript. RRA amalgamated edit suggestions to write the final draft. All authors significantly contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and examples.
[DOCX File , 17 KB - mhealth_v13i1e64098_app1.docx]

Multimedia Appendix 2
Google search terms.
[DOCX File , 22 KB - mhealth_v13i1e64098_app2.docx]

Multimedia Appendix 3
Targeted website search.
[DOCX File , 18 KB - mhealth_v13i1e64098_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Web-based program purpose and theoretical background.
[DOCX File , 21 KB - mhealth_v13i1e64098_app4.docx]

Multimedia Appendix 5
Resource bank purpose and theoretical background.
[DOCX File , 23 KB - mhealth_v13i1e64098_app5.docx]

Multimedia Appendix 6
App purpose and theoretical background.
[DOCX File , 31 KB - mhealth_v13i1e64098_app6.docx]

Multimedia Appendix 7
Mean adapted Mobile App Rating Scale ratings.
[DOCX File , 27 KB - mhealth_v13i1e64098_app7.docx]

Multimedia Appendix 8
Mean Alberta Rating Index for Apps ratings.
[DOCX File , 26 KB - mhealth_v13i1e64098_app8.docx]

References

1. Standing Committee on Public Safety and National Security. Healthy minds, safe communities: supporting our public safety officers through a national strategy for operational stress injuries. House of Commons Canada. 2016. URL: https://www.ourcommons.ca/Content/Committee/421/SECU/Reports/RP8457704/securp05/securp05-e.pdf [accessed 2025-03-16]
2. Heber A, Schick A, Bose E. Addressing the mental health impacts of COVID-19 on Canada's frontline workers. J Community Saf Well Being 2023 Feb 23;8(Suppl_1):S9-11. [doi: 10.35502/jcswb.300]
3. Carleton RN, Afifi TO, Taillieu T, Turner S, Krakauer R, Anderson GS, et al. Exposures to potentially traumatic events among public safety personnel in Canada. Can J Behav Sci 2019 Jan;51(1):37-52. [doi: 10.1037/cbs0000115]
4. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL. Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N Engl J Med 2004 Jul 01;351(1):13-22. [doi: 10.1056/NEJMoa040603] [Medline: 15229303]
5. Na PJ, Schnurr PP, Pietrzak RH. Mental health of U.S. combat veterans by war era: results from the national health and resilience in veterans study. J Psychiatr Res 2023 Feb;158:36-40. [doi: 10.1016/j.jpsychires.2022.12.019] [Medline: 36565542]
6. Wisco BE, Marx BP, Wolf EJ, Miller MW, Southwick SM, Pietrzak RH. Posttraumatic stress disorder in the US veteran population: results from the national health and resilience in veterans study. J Clin Psychiatry 2014 Dec 24;75(12):1338-1346. [doi: 10.4088/jcp.14m09328]
7. Mota N, Tsai J, Kirwin PD, Harpaz-Rotem I, Krystal JH, Southwick SM, et al. Late-life exacerbation of PTSD symptoms in US veterans: results from the national health and resilience in veterans study. J Clin Psychiatry 2016 Mar 23;77(03):348-354. [doi: 10.4088/jcp.15m10101]
8. Ricciardelli R, Carleton RN, Mooney T, Cramm H. "Playing the system": structural factors potentiating mental health stigma, challenging awareness, and creating barriers to care for Canadian public safety personnel. Health (London) 2020 May 16;24(3):259-278. [doi: 10.1177/1363459318800167] [Medline: 32283964]
9. Blais RK, Tsai J, Southwick SM, Pietrzak RH. Barriers and facilitators related to mental health care use among older veterans in the United States. Psychiatr Serv 2015 May 01;66(5):500-506. [doi: 10.1176/appi.ps.201300469] [Medline: 25639990]
10. Kline AC, Panza KE, Nichter B, Tsai J, Harpaz-Rotem I, Norman SB, et al. Mental health care use among U.S. military veterans: results from the 2019-2020 national health and resilience in veterans study. Psychiatr Serv 2022 Jun;73(6):628-635. [doi: 10.1176/appi.ps.202100112] [Medline: 34775790]
11. Berger W, Coutinho ES, Figueira I, Marques-Portella C, Luz MP, Neylan TC, et al. Rescuers at risk: a systematic review and meta-regression analysis of the worldwide current prevalence and correlates of PTSD in rescue workers. Soc Psychiatry Psychiatr Epidemiol 2012 Jun;47(6):1001-1011 [FREE Full text] [doi: 10.1007/s00127-011-0408-2] [Medline: 21681455]
12. de Vries LP, Baselmans BM, Luykx JJ, de Zeeuw EL, Minică CC, de Geus EJ, et al. Genetic evidence for a large overlap and potential bidirectional causal effects between resilience and well-being. Neurobiol Stress 2021 May;14:100315 [FREE Full text] [doi: 10.1016/j.ynstr.2021.100315] [Medline: 33816719]
13. Wild J, Smith KV, Thompson E, Béar F, Lommen MJ, Ehlers A. A prospective study of pre-trauma risk factors for post-traumatic stress disorder and depression. Psychol Med 2016 Sep;46(12):2571-2582 [FREE Full text] [doi: 10.1017/S0033291716000532] [Medline: 27348599]

XSL•FO
RenderX

14. Keyes CL. The mental health continuum: from languishing to flourishing in life. J Health Soc Behav 2002 Jun;43(2):207-222. [doi: 10.2307/3090197]
15. Prilleltensky I. Wellness as fairness. Am J Community Psychol 2012 Mar 04;49(1-2):1-21. [doi: 10.1007/s10464-011-9448-8] [Medline: 21643926]
16. Prilleltensky I, Dietz S, Prilleltensky O, Myers ND, Rubenstein CL, Jin Y, et al. Assessing multidimensional well-being: development and validation of the I Coppe Scale. J Community Psychol 2015 Jan 27;43(2):199-226. [doi: 10.1002/jcop.21674]
17. Pietrzak RH, Cook JM. Psychological resilience in older U.S. veterans: results from the national health and resilience in veterans study. Depress Anxiety 2013 May 06;30(5):432-443. [doi: 10.1002/da.22083] [Medline: 23468170]
18. Masten AS. Risk and resilience in development. In: Zelazo PD, editor. The Oxford Handbook of Developmental Psychology, Vol. 2. Self and Other. Oxford, UK: Oxford University Press; 2013.
19. Ungar MC. Modeling multisystemic resilience: connecting biological, psychological, social, and ecological adaptation in contexts of adversity. In: Multisystemic Resilience: Adaptation and Transformation in Contexts of Change. Oxford, UK: Oxford University Press; 2021.
20. Ungar M. The social ecology of resilience: addressing contextual and cultural ambiguity of a nascent construct. Am J Orthopsychiatry 2011 Jan;81(1):1-17. [doi: 10.1111/j.1939-0025.2010.01067.x] [Medline: 21219271]
21. Ungar M. Resilience, trauma, context, and culture. Trauma Violence Abuse 2013 May 03;14(3):255-266. [doi: 10.1177/1524838013487805]
22. Luthar SS, Cicchetti D, Becker B. The construct of resilience: a critical evaluation and guidelines for future work. Child Dev 2003 Jan 28;71(3):543-562. [doi: 10.1111/1467-8624.00164]
23. Skorburg JA, Yam J. Is there an app for that?: ethical issues in the digital mental health response to COVID-19. AJOB Neurosci 2022 May 14;13(3):177-190. [doi: 10.1080/21507740.2021.1918284] [Medline: 33989127]
24. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
25. Díaz-García A, Franke M, Herrero R, Ebert DD, Botella C. Theoretical adequacy, methodological quality and efficacy of online interventions targeting resilience: a systematic review and meta-analysis. Eur J Public Health 2021 Jul 07;31(31 Suppl 1):i11-i18 [FREE Full text] [doi: 10.1093/eurpub/ckaa255] [Medline: 34240159]
26. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 2;2(1):118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
27. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
28. Lui JH, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Prof Psychol Res Pract 2017 Jun;48(3):199-210. [doi: 10.1037/pro0000122]
29. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone apps for the treatment of mental disorders: systematic review. JMIR Mhealth Uhealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
30. Connolly SL, Miller CJ, Koenig CJ, Zamora KA, Wright PB, Stanley RL, et al. Veterans' attitudes toward smartphone app use for mental health care: qualitative study of rurality and age differences. JMIR Mhealth Uhealth 2018 Aug 22;6(8):e10748 [FREE Full text] [doi: 10.2196/10748] [Medline: 30135050]
31. Reger GM, Harned M, Stevens ES, Porter S, Nguyen J, Norr AM. Mobile applications may be the future of veteran mental health support but do veterans know yet? A survey of app knowledge and use. Psychol Serv 2022 Aug 03;19(3):480-487. [doi: 10.1037/ser0000562] [Medline: 34081527]
32. Choo CW. The art of scanning the environment. In: Reframing Environmental Scanning: A Reader on the Art of Scanning the Environment. Victoria, Australia: Australian Foresight Institute Swinburne University; 2003.
33. Roberts AE, Davenport TA, Wong T, Moon HW, Hickie IB, LaMonica HM. Evaluating the quality and safety of health-related apps and e-tools: adapting the Mobile App Rating Scale and developing a quality assurance protocol. Internet Interv 2021 Apr;24:100379 [FREE Full text] [doi: 10.1016/j.invent.2021.100379] [Medline: 33777705]
34. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
35. Azad Khaneghah P. Alberta Rating Index for Apps (ARIA): an index to rate the quality of mobile health applications. University of Alberta. 2020. URL: https://era.library.ualberta.ca/items/3c62ca2d-d7a1-4808-b1c0-4a60b05eecaa [accessed 2024-03-06]
36. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
37. Azad-Khaneghah P, Roduta Roberts M, Stroulia E, Ferguson-Pell M, Liu L. Alberta Rating Index for Apps (ARIA): helping older adults find acceptable mobile health apps. Gerontechnology 2020 Oct 03;19:1. [doi: 10.4017/gt.2020.19.s.69884]

XSL•FO
RenderX

38. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]
39. Dozois DJ, Mikail SF, Alden LE, Bieling PJ, Bourgon G, Clark DA, et al. The CPA Presidential Task Force on evidence-based practice of psychological treatments. Can Psychol 2014 Aug;55(3):153-160. [doi: 10.1037/a0035767]

Abbreviations

A-MARS: adapted Mobile App Rating Scale
ARIA: Alberta Rating Index for Apps
CBT: cognitive behavioral therapy
CPT: cognitive processing therapy
DMHI: digital mental health intervention
EBP: evidence-based practice
e-tool: electronic tool
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
PE: prolonged exposure
PPTE: potentially psychologically traumatic event
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
PSP: public safety personnel
PTSD: posttraumatic stress disorder
RB: resource bank
RCT: randomized controlled trial
WBP: web-based program

Edited by L Buis; submitted 09.07.24; peer-reviewed by M Notarianni, BA Pratt; comments to author 09.11.24; revised version received 04.01.25; accepted 12.02.25; published 01.04.25.

Please cite as:
Allen RR, Malik MA, Aquin C, Herceg L, Brémault-Phillips S, Sevigny PR
Digital Health Resilience and Well-Being Interventions for Military Members, Veterans, and Public Safety Personnel: Environmental Scan and Quality Review
JMIR Mhealth Uhealth 2025;13:e64098
URL: https://mhealth.jmir.org/2025/1/e64098
doi:10.2196/64098
PMID:40168068

©Rashell R Allen, Myrah A Malik, Carley Aquin, Lucijana Herceg, Suzette Brémault-Phillips, Phillip R Sevigny. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 01.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating Characteristics and Quality of Mental Health Apps Available in App Stores for Indian Users: Systematic App Search and Review

Seema Mehrotra[1], PhD; Ravikesh Tripathi[1], PhD; Pramita Sengupta[1], PhD; Abhishek Karishiddimath[1], MSc; Angelina Francis[1], MA; Pratiksha Sharma[1], MSc; Paulomi Sudhir[1], PhD; T K Srikanth[2], PhD; Girish Rao[3], MD; Rajesh Sagar[4], MD

[1]Department of Clinical Psychology, National Institute of Mental Health and Neurosciences, Bengaluru, India

[2]E-Health Research Centre, International Institute of Information Technology Bangalore, Bengaluru, India

[3]Department of Epidemiology, Centre for Public Health, National Institute of Mental Health and Neurosciences, Bengaluru, India

[4]Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, India

Corresponding Author:
Seema Mehrotra, PhD
Department of Clinical Psychology
National Institute of Mental Health and Neurosciences
Hosur Road
Bengaluru, 560029
India
Phone: 91 9448503853
Email: drmehrotra_seema@yahoo.com

Abstract

Background: The mental health app sector in India is expanding rapidly, driven by increasing smartphone usage, growing internet penetration, the popularity of digital initiatives, and heightened recognition of mental health challenges in public discourse. This growth is also influenced by both supply- and demand-side barriers to seeking professional help and the rise of mental health tech startups. While digital mental health solutions provide scalable ways to address unmet needs, concerns persist regarding app quality, privacy, and safety due to rapid market expansion, regulatory challenges, and limited empirical research. We conducted a comprehensive and systematic review of smartphone-based mental health apps accessible to Indian users through app stores.

Objective: This study aims to describe apps in terms of characteristics such as the nature of their functions, involvement of mental health professionals in development, reference to an empirical basis, and inclusion of nudges to seek professional help, as well as to evaluate app quality.

Methods: This systematic review of mental health apps was conducted using the TECH (Target user, Evaluation focus, Connectedness, and Health domain) approach, along with the PASSR (Protocol for App Store Systematic Reviews) checklist. Fifteen search terms covering mental health conditions and therapies were applied to both Google Play and Apple App Store. Identified apps were screened according to predefined inclusion and exclusion criteria and subsequently downloaded for detailed review. Data were extracted based on prespecified parameters. Additionally, app quality was evaluated using the Mobile Application Rating Scale (MARS).

Results: The initial search identified 5827 apps, of which 350 were reviewed in detail after removing duplicates and applying eligibility criteria. Common search terms such as “depression” and “anxiety” yielded nearly a quarter of relevant apps (128/495, 25.9% to 133/497, 26.8%); 62 (17.7%) of the 350 reviewed apps originated from Asia, and 131 (37.4%) focused on a single mental health condition. Multifunction apps (eg, those combining assessment and intervention) constituted the largest category (230/350, 65.7%). Privacy concerns were notable; for example, 54 (15.4%) apps did not mention a data-sharing policy. Most apps were developed by commercial organizations, and 228 (65.1%) did not report involvement of mental health professionals, while 45 (12.9%) mentioned it only cursorily. Only 38 (10.9%) apps referenced empirical research, and more than half did not indicate an empirical basis for their content. Pointers to seek professional help were present in 139 (39.7%) apps, mostly in the form of disclaimers, whereas nudges or motivational prompts to seek help appeared in slightly less than a quarter. Only 105 (30%) apps attempted to dispel mental health myths. Functionality and aesthetics ratings on the MARS were relatively high, but 50 (14.3%) apps scored 3 or lower on the information subscale.

XSL•FO
RenderX

Conclusions: This study is among the first systematic evaluations of mental health apps accessible to Indian users on Google Play and Apple App Store. The findings provide insights to guide future research, app development, and policy making in the digital mental health space.

Trial Registration: International Platform of Registered Systematic Review and Meta-analysis Protocols (INPLASY) INPLASY2024100035; https://inplasy.com/inplasy-2024-10-0035/

International Registered Report Identifier (IRRID): RR2-10.2196/71071

(JMIR Mhealth Uhealth 2025;13:e79238) doi:10.2196/79238

KEYWORDS

mental health apps; mHealth; review of apps; smartphone apps; MHApps for Indian users; India; mobile phones

Introduction

Background

Digital mental health solutions offer a scalable approach to addressing the unmet mental health needs of the population. When effectively integrated into community-based services and health care delivery systems, they can facilitate early intervention, reduce the burden on mainstream health care, and promote mental well-being through preventive strategies and sustained support for recovery [1,2].

Mental Health App Sector in India

The mental health app sector in India is witnessing significant growth. India accounts for over 6% of global mental health tech start-ups, with its app market valued at US $112.75 million in 2023 and projected to grow at an annual rate of 18.2% until 2030 [3,4]. This growth is shaped by multiple factors, including a significant treatment gap due to the limited number of trained professionals, unevenly distributed services, and psychosocial barriers that discourage help-seeking (eg, concerns about social stigma, the perception that mental health struggles are temporary, inadequate support from significant others for professional help-seeking, and gaps in awareness about when to seek professional help) [5-8]. At the same time, the rapid increase in smartphone users, improved internet access, growing familiarity with digital platforms, and greater recognition of mental health challenges in public discourse are driving the expansion of the digital mental health market in India [9-14]. In this context, the appeal of digital mental health tools lies in their ease of access, affordability, and the privacy and anonymity they provide.

Mental Health Apps: Opportunities and Challenges

Mental health apps can serve a wide range of functions, including enhancing mental health literacy, fostering self-awareness of one's concerns, reducing barriers to seeking help, and providing basic crisis support. They can offer self-help tools for practice, enable access to online support from professionals or supportive communities, facilitate symptom monitoring and reporting to assist health care providers, and create opportunities for blended interventions that integrate digital self-help with brief therapy sessions [15,16].

While mental health apps present significant opportunities, it is equally crucial to recognize and address key challenges and potential risks to maximize their impact for the public good. Globally, about 10,000 mental health apps are available to the public [17]. The rapid and volatile growth of these apps in virtual stores, combined with insufficient regulatory oversight, makes it extremely difficult for both end users and professionals to identify the most suitable options at any given time [18,19]. Moreover, many well-researched apps remain unavailable to the public, while the majority of readily accessible apps lack direct evidence of effectiveness [20,21]. This underscores the importance of ensuring that all apps include evidence-informed content, which becomes especially vital when users rely on them without professional guidance or oversight [22,23]. Furthermore, privacy and security concerns surrounding mental health apps remain a significant issue, as some handle sensitive data without adequate legal and ethical safeguards, such as clear data-sharing policies, strong encryption, and responsible use of permissions sought [24-26].

Review of Mental Health Apps in App Stores

Evaluation of mental health apps available to the public in app stores has garnered significant research attention in recent years. These reviews have focused on a wide range of aspects, such as specific mental health conditions (eg, depression, social anxiety [27,28]), app types (eg, artificial intelligence chatbots [29]), age-specific apps (eg, youth-focused apps [30]), and apps available in particular regions [31]. They have also examined issues including privacy and security [25], content quality, user engagement and functionality [31,32], content analysis of consumer reviews [33], and compatibility with evidence-based practices [27,34]. Only a handful of reviews have examined mental health apps from India, addressing themes such as app store descriptions, downloads, pricing, user ratings, and privacy policies [26], as well as apps focused on depression [35], tobacco cessation [36], mental health apps in general [37], and suicide prevention [38]. Most of these studies focused on specific mental health concerns and screened approximately 200-700 apps available to Indian users in app stores, ultimately conducting detailed reviews of about 20-40 apps. To the best of our knowledge, no comprehensive review has yet been conducted on mental health apps available in virtual stores for Indian users.

Objectives

This study aimed to conduct a comprehensive and systematic review of smartphone-based mental health apps available in app stores accessible to Indian users. The specific objectives were to (1) describe the apps in terms of their characteristics—such as function, type of intervention, target mental health conditions, involvement of mental health

XSL•FO
RenderX

professionals in development, reference to an empirical evidence base, and inclusion of content that encourages professional help-seeking (nudges); and (2) evaluate the quality of these apps.

Methods

Ethical Considerations

As the study was a systematic review of apps available in virtual stores, an exemption was sought and obtained from the Institute Ethics Committee, National Institute of Mental Health and Neuro Sciences (NIMHANS) (No. NIMHANS/EC/(BEH.SC.DIV.)MEETNG/2024, dated February 6, 2024).

App Review Approach

A framework called "TECH" (Target user, Evaluation focus, Connectedness, and Health) has been specifically developed to formulate research questions and establish eligibility criteria for systematic reviews of health apps. Traditional frameworks such as PICO (Population, Intervention, Comparison, and Outcome) and SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research) are better suited for systematic reviews assessing intervention effectiveness or qualitative studies, respectively [39]. The TECH framework consists of 4 key elements: (1) target user, referring to the specific population of interest; (2) evaluation focus, which may include aspects such as app characteristics, quality, usability, techniques, or components; (3) connectedness, indicating whether the app integrates with other apps or services; and (4) health domain, representing the specific health condition or concern being addressed. This approach helps shape research questions and guide eligibility criteria for app reviews. In this study, the "target user" referred to Indian adults interested in using mobile apps to understand or manage mental health concerns. The "evaluation focus" was broad, encompassing descriptive details about app characteristics, content, and quality. Regarding "connectedness," the review included both stand-alone apps and those connected to services, but excluded apps that integrated with wearables or other apps. The "health domain" was also broad, covering apps that addressed a range of mental health conditions and concerns. The eligibility criteria outlined later in the study were informed by these elements of the TECH framework.

As this review focused on mental health apps rather than research studies on such apps, the PASSR (Protocol for App Store Systematic Reviews) checklist (see Multimedia Appendix 1) was utilized [21]. PASSR is a combination and adaptation of items from the AMSTAR (A Measurement Tool to Assess Systematic Reviews) and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklists, and can be applied to the systematic search of app stores for any category of apps [40,41].

Eligibility Criteria

The following inclusion criteria were applied to select apps for review:

- Apps available in the Google Play or Apple App Store for Indian adults.
- Store descriptions indicating that the app offers guidance on mental health problems or therapy, information, self-help, or support, and is therefore relevant to the review.
- Apps available in English.
- Fully free apps.
- Free trial version of fully paid apps.
- Free-to-use content within apps that include in-app purchases.

Apps available only in Indian languages, without an English version, were documented but not included in the detailed review. Fully free apps were reviewed in their entirety. Fully paid apps were reviewed based on the features accessible during the free trial period. For apps with in-app purchases, only the freely available content was reviewed as part of this study.

The following exclusion criteria were applied:

- Mental health apps that were fully paid and did not offer a free trial period.
- Apps available solely in non-English languages.
- Apps intended exclusively for use by health professionals (eg, for training or education).
- Apps requiring additional wearable devices or sensors.
- Apps focused solely on wellness enhancement without addressing mental health concerns (eg, improving time management, productivity, or relationships).
- Apps available only to research participants of a given study.

In addition, apps that were no longer available for download at the time of detailed review were excluded.

Rationale for the Eligibility Criteria

The search criteria were established to effectively address the research objectives while ensuring feasibility for this comprehensive review. The search was conducted in Google Play and Apple App Store, which together account for approximately 99% of the Indian market share [42]. Most mental health apps available to Indian users are in English. Apps in Indian languages that also had an English version were planned to be included in the review. Given that India is a multilingual country [43], apps available exclusively in non-English languages were documented but excluded from detailed review, as their assessment would have required language proficiency. Additionally, the vast and diverse range of apps related to general well-being warrants a separate review. To maintain focus and ensure feasibility, this review was limited to apps addressing 1 or more mental health conditions. Apps requiring additional wearable devices were excluded due to accessibility and affordability concerns. Likewise, apps designed specifically for health professionals were not included, as the review aimed to evaluate apps intended for public use.

Search Strategy

A keyword-driven search was used to identify apps for review. Fifteen keywords were applied in the virtual stores: mental health, depression, anxiety, PTSD, OCD, addiction, schizophrenia, bipolar, BPAD, CBT, ACT, DBT, cognitive

XSL•FO
RenderX

behavior therapy, acceptance and commitment therapy, and dialectical behavior therapy. The selection of these keywords was guided by the purpose of this study—to cover a broad range of prevalent mental health concerns, commonly used psychotherapy interventions, and their acronyms. In addition, a broad scan of recent review studies on mental health apps and an initial scan of the app stores informed the finalization of keywords.

A systematic keyword-based search was conducted between October and November 2024 on both the Google Play and Apple App Store. To avoid algorithmic personalization, each term was searched on freshly reset devices with new Gmail or iOS accounts. Search histories were cleared between each keyword search, and reviewers scrolled through all results to the end of the screen, ensuring they went beyond in-feed advertisements. The full search sessions were screen-recorded for verification and saved on secure drives. All apps identified in each search were manually entered into an Excel (Microsoft Corporation) database using the recordings. Four devices (2 Android and 2 iOS) were used. The review was conducted by 4 primary reviewers (AF, P Sharma, AK, and P Sengupta), with support from expert mentors (SM and RT).

Initial Screening

Duplicate apps were removed from the Excel sheet. The 4 primary reviewers independently screened the apps for eligibility by examining app store descriptions. If an app was available on both Google Play and Apple App Store for Android and iOS, respectively, only the Google Play version was reviewed, and the app was counted once. This screening task was divided among the 4 members of the review team. Any uncertainty regarding eligibility was resolved through joint discussions. If the reviewers were unable to reach a unanimous decision, consultations were held with the 2 expert mentors (SM and RT) to arrive at a final determination. Apps that met the eligibility criteria were then downloaded for detailed review. During the detailed review, any app deemed ineligible was excluded.

App Description and Quality Evaluation

Each eligible app was explored in detail, and its characteristics were documented across 21 parameters, including privacy, payment information, type of developer, target mental health condition, and mention of empirical research, among others. Table 1 provides a brief description of these parameters and the approach used for coding them.

App quality was evaluated using the Mobile Application Rating Scale (MARS) [44]. This scale consists of 19 items across 4 domains: engagement, functionality, aesthetics, and information quality. Each item is rated on a 5-point Likert scale: (1) inadequate, (2) poor, (3) acceptable, (4) good, and (5) excellent. The total score and the 4 objective domains demonstrate high internal consistency. An additional 4-item subjective quality scale is available; however, it is not included in the total score and was not used in this review.

XSL•FO
RenderX

Table 1. Parameters and coding explanations used to evaluate mental health apps.

No.	Parameter	Coding explanation
1	Privacy policy	• Recorded as "accessible to users" if it is easily located and readable, such as being available on the app's home page, footer, or settings menu. • Classified as "clearly explained" if the privacy documentation provides explicit information regarding user responsibilities, app scope, legal compliance, data security, confidentiality, and terms of use in a straightforward and nonambiguous manner. • "Mention of data-sharing policy with third parties" is recorded if the policy explicitly states whether and how user data may be shared with external entities, including relevant details about who the third parties are and the purposes for which data are shared. • "Mention of data retention duration" is recorded if the policy specifies how long user data will be stored, including time frames or the criteria for deletion or anonymization.
2	Provision for deletion of user account/data	• Recorded as "yes" if either the privacy policy or in-app settings specify a mechanism enabling users to delete their account or remove personal information from their device.
3	Payment requirements	• Categorized as (1) completely free to use, (2) fully paid, or (3) involving in-app purchases.
4	Nature of developer	• Classified according to institutional type: (1) academic institution, (2) commercial organization, (3) government agency, or (4) not-for-profit organization. When information provided is insufficient, it is recorded as (5) insufficient information available.
5	Continent of origin	• Designated as North America, South America, Asia, Europe, or Oceania, based on the developer's location.
6	Release year	• Coded as either "released before 2020" or "released after 2020."
7	Last updated	• Indicated as "updated within the past 6 months" or "updated 1 year ago or earlier," according to the app store information.
8	App downloads	• Recorded as reported in app stores (eg, ranges or absolute numbers, as available).
9	App reviews/ratings	• Extracted from app store data and coded according to the numerical average user rating displayed on the app store. Ratings were categorized as "5," "4-4.9," "3-3.9," and "2-2.9" for the apps.
10	Target age group	• Categorized as children (<12 years), adolescents (13-17 years), adults (18 and above), or as the general population if it is open to all.
11	Language	• Coded as (1) English only, or (2) English and 1 or more Indian languages.
12	Technical features	• Assessed for the presence of the following: sharing capability (eg, social media integration), app community features, password protection, log-in requirement, reminder functionality, and dependence on internet access.
13	Target mental health condition(s)	• Labeled as (1) focuses on a single mental health condition, (2) addresses multiple mental health conditions, or (3) not specified.
14	Type of intervention	• Coded as (1) cognitive behavioral therapy, (2) third-wave cognitive behavioral therapy, (3) mindfulness or meditation, (4) behavior change approaches, or (5) other therapeutic approach.
15	Involvement of mental health professionals in the development	• Labeled as "clearly specified" when names and qualifications/designations of professionals are provided; as "generally mentioned" when involvement is referenced without sufficient detail.
16	Inclusion of myth-busting content	• Recorded as "yes" if content explicitly seeks to clarify or correct common myths or stigmatizing beliefs regarding mental health, disorders, or treatment.
17	Mention of crisis support strategies	• Recorded as "yes" if the app provides access to helplines, emergency websites, or outlines basic crisis management strategies.
18	Pointers to seek professional help	• Recorded as "yes" if the app or its descriptive content explicitly states it is not a replacement for professional help and encourages users to consult a mental health professional as appropriate.
19	Nudges to seek professional help	• Recorded as "yes" if the app delivers prompts or suggestions that encourage users to seek professional support (eg, in response to high scores on screening tools or chatbot interactions).

XSL•FO
RenderX

No.	Parameter	Coding explanation
20	Empirical basis of intervention/assessment	• Recorded as "yes" if the app's intervention or assessment content is explicitly described as empirically supported in the app or store description.
21	Empirical research conducted on the app	• Recorded as "yes" if any published empirical studies evaluating the app's usability, effectiveness, or related outcomes are cited within the app's content or app store description.

Review Process

All primary reviewers held at least a master's degree in psychology and had a minimum of 3 years of research experience. For this review, they underwent structured training in the use of the MARS, a validated tool for evaluating the quality of mobile health apps. The training included viewing the official instructional videos, and the authors of MARS were also contacted to enhance conceptual clarity and ensure a consistent interpretation of the rating criteria.

Phase 1: Individual Review and Expert Discussion for Calibration

Following MARS training, the 4 primary reviewers independently rated 3 randomly selected apps from the pool of eligible apps, after removing duplicates across search terms and app stores. These independent ratings were then discussed in detail, with discrepancies deliberated and resolved in consultation with expert mentors (clinical psychology faculties SM and RT, each with over a decade of professional experience and prior involvement in app evaluations). This phase facilitated a shared understanding and helped identify specific domains where raters needed to refine and align their judgments.

Phase 2: Joint Review of Three Apps

In this phase, the 4 reviewers jointly evaluated 3 additional apps, discussing each item and rating in real time. This collaborative exercise reinforced shared understanding and facilitated cross-learning among the reviewers. An internal guide was created to provide a common framework for applying a nuanced approach when rating apps in cases of ambiguity or confusion.

Phase 3: Independent Review With Joint Discussion

Each of the 4 reviewers was then assigned a new set of 6 apps to review independently. Once rated, these apps were brought back for group discussion, during which all discrepancies were carefully analyzed. These discussions played a crucial role in refining rating standards and were instrumental in identifying specific domains—such as tracking-based apps (eg, mood tracking, sobriety tracking) and assessment/screening test-based apps—where scoring discrepancies arose due to differences in interpretation.

Following this, the remaining pool of apps was divided among the reviewers for independent evaluation, using the calibration established in the earlier rounds. For apps that raised minor uncertainties—due to content ambiguity, unclear functionalities, or cross-domain features—other reviewers were consulted. Apps involving a high degree of uncertainty were recorded in a shared "doubtful apps" sheet, primarily concerning eligibility for inclusion. These apps were flagged for expert-level deliberation in the next phase.

Phase 4: Expert Review

A total of 7 (2%) apps reviewed in detail were selected by the primary review team for evaluation by a team of 2 experts (SM and RT). These apps were purposively chosen to represent different functions, including assessment, monitoring and tracking, information provision, and interventions.

Comparison of expert ratings with those provided by the primary reviewers revealed discrepancies of up to 2 points on the engagement and information subscales of MARS. These discrepancies occurred only in 2 types of apps: monitoring and tracking, and assessment apps. Following recalibration of the rating norms for these app types, the primary reviewers re-evaluated the apps in these categories.

Additionally, all apps flagged as "doubtful" were discussed with SM and RT during this round, and decisions were made regarding their inclusion in the final pool of reviewed apps.

Results

Overview of the Review Process

The review process is shown in Figure 1.

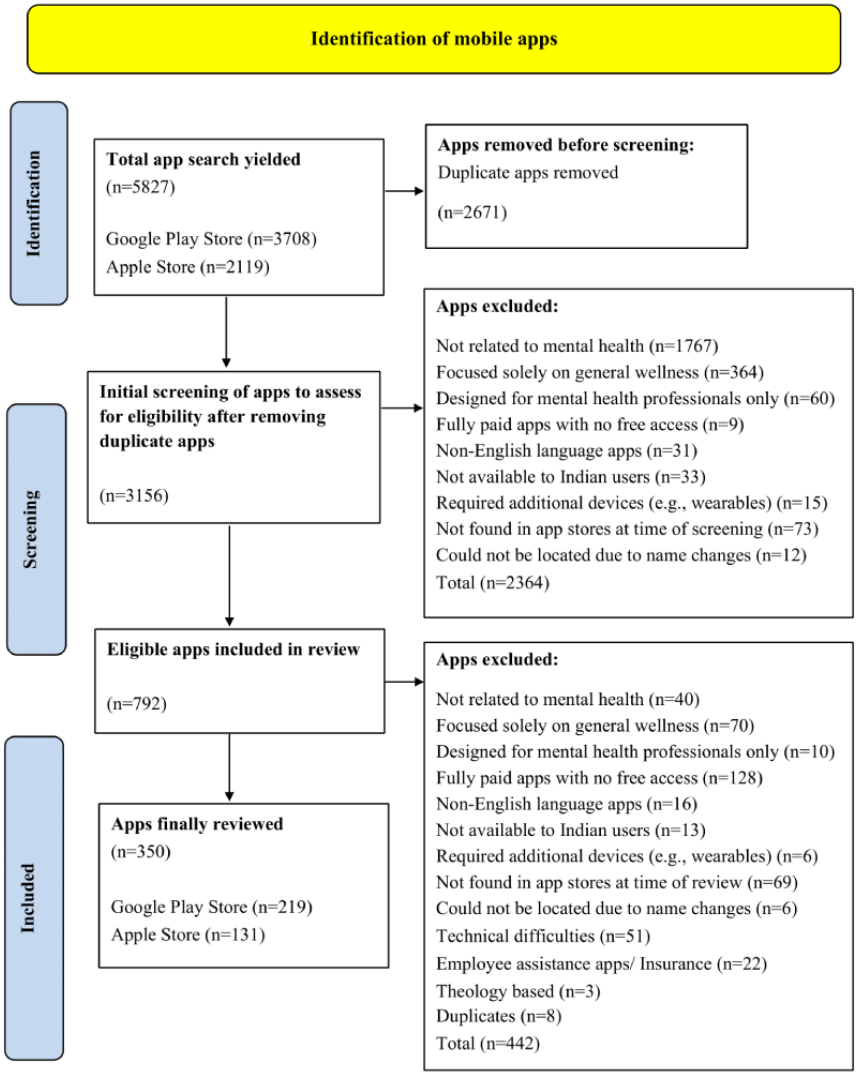
Table 2 shows that a total of 5827 apps were identified through the keyword-based search. The top 5 keywords yielding the highest number of apps were "mental health" (n=500), "depression" (n=497), "anxiety" (n=495), "OCD" (n=492), and "addiction" (n=484), collectively contributing to more than one-third (n=2468, 42.35%) of the total apps retrieved in the initial search. As indicated in Table 2, for most search terms, only about 20%-30% (for example, the keyword "bipolar" initially yielded 388 apps of which only 78 apps, ie, 20% were retained in the final review, whereas the keyword "cognitive behavioral therapy" initially yielded 303 apps of which 91 apps, ie, 30% were retained in the final review) of the initially retrieved apps were retained after duplicate removal and eligibility screening, highlighting the high proportion of nonrelevant or duplicate content in app store searches.

Of the apps initially identified, a subset of 350 apps met the predefined inclusion criteria and was deemed eligible for final review. Among these, the largest numbers were contributed by apps retrieved using the keywords "depression," "anxiety," "mental health," "CBT," and "PTSD." Of the 350 apps reviewed, 164 were available only on Google Play, 130 only on Apple App Store, and 56 were available on both platforms. In other words, 220 (62.9%) of the 350 reviewed apps were available on Google Play, and 186 (53.1%) were available on Apple App Store. As noted earlier, for apps available on both platforms, only the Google Play version was reviewed to maintain consistency in evaluation.

XSL•FO
RenderX

Tables 3 and 4 highlight the characteristics of the apps across various review parameters.

Figure 1. Flowchart of the review process.



XSL•FO
RenderX

Table 2. Number of apps identified using key search terms.

Apps yielded in the initial search on key search terms[a]	Apps eligible for final review, n (%)[b]
Mental health (n=500)	126 (25.2)
Depression (n=497)	133 (26.8)
OCD[c] (n=492)	59 (12.0)
Addiction (n=484)	50 (10.3)
Schizophrenia (n=265)	67 (25.3)
Anxiety (n=495)	128 (25.9)
PTSD[d] (n=431)	101 (23.4)
BPAD[e] (n=234)	11 (4.7)
Bipolar (n=388)	78 (20.1)
Cognitive behavioral therapy (n=303)	91 (30.0)
Dialectical behavioral therapy (n=250)	68 (27.2)
Acceptance commitment therapy (n=257)	58 (22.6)
CBT[f] (n=456)	122 (26.8)
DBT[g] (n=324)	40 (12.3)
ACT[h] (n=451)	38 (8.4)

[a]The initial search yielded a total of 5827 apps.

[b]Apps eligible for final review were 350. An app could be retrieved under multiple search terms. Hence, the total of frequencies corresponding to apps eligible for final review across all search terms exceeds 350.

[c]OCD: obsessive compulsive disorder.

[d]PTSD: posttraumatic stress disorder.

[e]BPAD: bipolar affective disorder.

[f]CBT: cognitive behavioral therapy.

[g]DBT: dialectical behavior therapy.

[h]ACT: acceptance and commitment therapy.

XSL•FO
RenderX

Table 3. Characteristics of reviewed apps on basic parameters (N=350).

Review parameters	Values, n (%)
Privacy policy	
Privacy policy accessible to users	310 (88.6)
Privacy terms clearly explained	301 (86.0)
Mention of data-sharing policy with third party	296 (84.6)
Mention of data retention duration	207 (59.1)
Mention of the provision for deletion of the user account/information from the users' device	
Yes	181 (51.7)
Not applicable	1 (0.3)
Payment requirements	
Completely free to use	119 (34.0)
Fully paid	5 (1.4)
Involves in-app purchases	226 (64.6)
Nature of the developer	
Academic institutions	8 (2.3)
Commercial organizations	302 (86.3)
Government agency	13 (3.7)
Not-for-profit organizations	15 (4.3)
Insufficient information available	12 (3.4)
Continent of origin	
North America	93 (26.6)
South America	5 (1.4)
Asia	62 (17.7)
Europe	89 (25.4)
Oceania	5 (1.4)
Insufficient information available	96 (27.4)
Release year	
Released after 2020	141 (40.3)
Released before 2020	69 (19.7)
Information not available[a]	140 (40.0)
Last updated (at the time of data collection)	
Updated 1 year ago or earlier	127 (36.3)
Updated within the past 6 months	125 (35.7)
Information not available[a]	98 (28.0)
App downloads	
Up to 10,000	62 (17.7)
10,000+	37 (10.6)
50,000+	16 (4.6)
100,000 to 10 million	88 (25.1)
10+ million	7 (2.0)
Information not available[a]	140 (40.0)
App reviews/ratings	

XSL•FO
RenderX

Review parameters	Values, n (%)
2-2.9	2 (0.6)
3-3.9	26 (7.4)
4-4.9	170 (48.6)
5	17 (4.9)
Information not available[a]	135 (38.6)
Technical aspects[b]	
Allow sharing (Facebook, Twitter, etc)	101 (28.9)
Has an app community	68 (19.4)
Allows password protection	63 (18.0)
Requires log-in	163 (46.6)
Sends reminders	199 (56.9)
Needs web access to function	214 (61.1)

[a]Information not available: due to insufficient information mentioned for the apps in the app stores.

[b]Frequency of technical aspect does not add up to 350, as 1 app can have more than 1 technical aspect.

Table 4. Characteristics of reviewed apps on mental health parameters (N=350).

Review parameters	Values, n (%)
Mental health condition	
Single mental health condition focused	131 (37.4)
Multiple mental health condition(s)	159 (45.4)
Not mentioned	60 (17.1)
Type of intervention[a]	
Cognitive behavioral therapy	152 (43.4)
Third-wave cognitive behavioral therapies	52 (14.9)
Mindfulness/meditation	123 (35.1)
Behaviour change–focused approaches	49 (14.0)
Other therapeutic approaches	27 (7.7)
Apps attempt to dispel myths	105 (30.0)
Apps mentioning basic crisis support strategies	94 (26.9)
Mental health professionals' involvement in the development of the app	
Clearly specified	77 (22.0)
Generally mentioned	45 (12.9)
Not mentioned	228 (65.1)
Apps providing pointers to seek professional help when needed	139 (39.7)
Apps providing nudges to seek professional help	80 (22.9)
Apps mentioning empirical basis of intervention/assessment	161 (46.0)
Empirical research on the app mentioned within its in-app content or app store description	38 (10.9)

[a]Frequency of intervention type does not add up to 350, as 1 app can have more than 1 intervention type.

Privacy Policy, Pricing, and Origin of the Apps

A total of 301 (86%) apps had a clearly stated privacy policy. Nearly half of the apps provided users with the option to delete their information if desired. Approximately two-thirds of the apps included in-app purchases, while slightly more than one-third were completely free to use. Of the 350 apps, 36 (10.3%) were developed by academic institutions, government agencies, or not-for-profit organizations. Half of the apps originated from North America and Europe, whereas one-sixth were developed in Asia.

XSL•FO
RenderX

Recency, Downloads, and Target Population of the Apps

Of the 350 reviewed apps, 141 (40.3%) were released after 2020; 10 (2.9%) were available in an Indian language in addition to English. Download information was not available for more than one-third of the reviewed apps. Among the remaining apps with available download data, approximately one-quarter had between 100,000 and 10 million downloads, while slightly more than one-quarter had fewer than 50,000 downloads. About half of the apps had a user rating of 4 or higher on app stores, although the number of reviews varied. Based on app store descriptions, 2 apps targeted children under 12 years, while 32 apps targeted adolescents (13-17 years). A total of 120 apps specified a target population of 18 years or older. Some apps targeted multiple age groups. However, 299 apps (85.4%) did not mention a specific target age group in their description.

Type of Interventions

More than half of the apps provided self-help strategies informed by CBT, dialectical behavior therapy (DBT), acceptance and commitment therapy (ACT), and compassion-focused therapy. As DBT, ACT, and compassion-focused therapy are commonly classified as third-wave CBTs, these interventions were grouped accordingly [45,46]. Mindfulness- and meditation-based self-help apps were coded separately, as they did not explicitly indicate that they were informed by CBT principles. Behavior change–focused approaches comprised 49 (14%) of the 350 apps, including 12-step interventions, motivational enhancement therapy, relapse prevention for addiction, exposure and response prevention, other behavioral therapies, and tracking. A wide range of other therapeutic approaches was mentioned in 27 of the reviewed apps, including hypnotherapy, eye movement desensitization and reprocessing, neurofeedback, trauma-informed approaches, Jungian therapy, existential therapy, schema therapy, psychodynamic therapy, humanistic approaches, safety planning, community/peer support, positive psychology, clinical sexology, Gestalt therapy, and somatic therapy. More than two-thirds of the apps did not include content specifically aimed at dispelling myths about mental health. Only about one-fourth of the apps provided crisis support strategies.

Involvement of Mental Health Professionals and Encouragement for Professional Help-Seeking

Of the 350 reviewed apps, 228 (65.1%) did not indicate whether mental health professionals were involved in their development. Pointers to seek professional help were coded when apps included disclaimers stating that they are neither diagnostic tools nor substitutes for professional assistance. Nudges to seek professional help were coded when an app went beyond disclaimers to actively encourage users to seek professional support based on assessment results or user inputs. More than half of the apps included neither pointers nor nudges to seek professional help.

Table 5 presents the classification of the apps included in the final review. Apps were categorized as “interventions” if they focused on providing therapeutic strategies intended to help users. Apps that focused exclusively on tools for recording and monitoring parameters such as thoughts, moods, or behaviors over time, to gain insights, were classified under “monitoring and tracking.” The “assessment” category included apps that exclusively provide screening tools or tests. Apps that offered only information or resources on mental health or mental health conditions were classified as “informative/educative.” The “other” category comprised artificial intelligence–assisted tools, journaling/diary apps, community support apps, game-based tools, and apps for connecting users to therapists.

Apps were classified as “multifunction” if they combined 2 or more single functions. Among the 230 multifunction apps, the most common functions were monitoring/tracking (186/230, 80.9%), interventions (177/230, 77%), and information/education (130/230, 56.5%). Other functions available in multifunction apps were journaling, assessments, community support, artificial intelligence–based support, and connecting with or finding a therapist.

Table 5. Classification of reviewed apps (N=350).

Apps category	Values, n (%)
Single function	120 (34.3)
Interventions (based on therapeutic approaches)	44 (12.6)
Assessment	24 (6.9)
Monitoring and tracking	33 (9.4)
Informative/educative	12 (3.4)
Other	7 (2.0)
Multifunction	230 (65.7)

Table 6 presents the mean and SD of the final reviewed app categories on MARS subscales and overall average ratings. The overall MARS average rating for all 350 apps was 4.09 (SD 0.44), with a range of 2.52-4.95. Intervention apps received the highest overall MARS average rating, while informative/educative apps scored the lowest. Assessment apps had the lowest mean rating for engagement, whereas multifunction apps scored the highest. All app categories had a mean rating above 4 on the Functionality subscale. The informative/educative category received the lowest average rating on the MARS aesthetics subscale.

XSL•FO
RenderX

Table 6. MARSa subscale ratings of reviewed app categories.

App category and MARS subscales	Mean (SD)
Multifunction (n=230)	
Engagement	3.92 (0.58)
Functionality	4.44 (0.55)
Aesthetics	4.28 (0.53)
Information	3.91 (0.49)
MARS average rating	4.14 (0.43)
Assessment (n=24)	
Engagement	2.89 (0.47)
Functionality	4.69 (0.49)
Aesthetics	3.92 (0.51)
Information	3.60 (0.64)
MARS average rating	3.77 (0.41)
Intervention (n=44)	
Engagement	3.88 (0.74)
Functionality	4.72 (0.47)
Aesthetics	4.30 (0.52)
Information	3.95 (0.54)
MARS average rating	4.21 (0.44)
Informative/educative (n=12)	
Engagement	3.10 (0.62)
Functionality	4.36 (0.67)
Aesthetics	3.49 (0.62)
Information	3.65 (0.26)
MARS average rating	3.65 (0.34)
Monitoring and tracking (n=33)	
Engagement	3.47 (0.48)
Functionality	4.62 (0.41)
Aesthetics	4.04 (0.39)
Information	3.74 (0.37)
MARS average rating	3.97 (0.28)
Other (n=7)	
Engagement	3.68 (0.72)
Functionality	4.44 (0.26)
Aesthetics	4.42 (0.42)
Information	3.50 (0.39)
MARS average rating	4.01 (0.33)

[a]MARS: Mobile Application Rating Scale.

The mean ratings on the information subscale of MARS were below 4 for all app categories; 50 (14.3%) reviewed apps received ratings of 3 or lower on the information quality item of the MARS information subscale. Among these, 7 apps received ratings below 3, and their content was examined in detail. Several concerning patterns were observed in these apps regarding the accuracy, safety, and scientific basis of their content. One notable issue was the presence of overstated or misleading claims, portraying the app as a comprehensive therapeutic tool or offering “highly accurate” diagnostic assessments without clear evidence of validation or involvement of qualified mental health professionals. In some instances,

XSL•FO
RenderX

there was a notable disconnect between how the app was described (eg, promising depression support or CBT-based interventions) and the features actually provided, which were often limited to basic journaling or habit-tracking functions. Such inconsistencies can be problematic for users seeking targeted mental health support. Additionally, some apps included components that are not widely recognized as interventions without providing users with explanations of their relevance or intended benefits. In a few cases, apps generated alarming feedback (eg, "high suicide risk") without context or guidance, raising ethical concerns about supporting such users. The absence of crisis support or access to professional help within these apps further compounds the risks. These observations underscore the importance of aligning app content with its stated purpose, ensuring transparency regarding the scientific basis of interventions, and including support mechanisms for users who may be in distress.

Further inspection of MARS ratings revealed that approximately two-thirds of multifunction apps achieved an average rating of 4 or higher, whereas only one-fourth of assessment apps reached this threshold. Three-quarters of intervention apps received ratings of 4 or higher. Less than one-fifth of apps in the informative/educative category achieved ratings of 4 or above, while about half of the monitoring and tracking apps scored 4 or higher on the MARS average rating.

Discussion

Principal Findings

This systematic review provided valuable insights into publicly accessible mental health apps available in virtual stores for Indian users. It identified several key findings related to the emergent app categories, the involvement of mental health professionals in app development, the presence of an evidence base, content quality, the extent to which apps encourage professional help-seeking, and overall app quality as evaluated using the MARS framework. Each of these aspects is discussed in the context of existing literature.

Reviewed Apps: Emerging Categories

The reviewed apps varied in the functions they offered, their focus on single versus multiple mental health conditions, and the types of interventions provided.

The mental health apps offered a variety of functions, including symptom assessment, information/education provision, symptom or progress monitoring, and interventions. Assessment and monitoring apps were the most common among single-function apps, consistent with previous literature [26,31,35]. Multifunction apps, incorporating 2 or more types of functions (eg, assessment and intervention), constituted the largest category (230/350, 65.7%). This is a positive trend, as such apps may address multiple user needs within a single platform. Further examination of app content revealed that 179 of the 350 (51.1%) apps incorporated multiple intervention strategies, while 126 (36%) were designed around a single component or strategy (eg, breathing exercises). The latter may be useful for individuals with highly specific or well-defined needs, whereas multistrategy apps offer a broader range of options for users to select and utilize according to their needs and relevance.

While 131 (37.4%) apps focused on a single mental health condition (eg, depression), 60 (17.1%) did not specify a focus. The majority targeted multiple mental health conditions, consistent with the overlap observed among common mental health problems in terms of symptoms, comorbidity, and intervention strategies [47-49].

CBT and mindfulness were the most common interventions, followed by DBT, ACT, and motivational enhancement therapy. CBT and related approaches have received considerable attention from researchers and developers in digital mental health, likely due to their widespread popularity and structured framework, which facilitates adaptation into digital self-help tools [50]. The distinction between mindfulness meditation practice and mindfulness-based therapeutic approaches for mental health problems was not always clear, which may be confusing for users. Basic crisis management strategies were included in 94 of the 350 (26.9%) apps, while helpline numbers or emergency service contacts were provided in 109 (31.1%) of the 350 reviewed apps.

Only 105 (30%) of the reviewed apps made a clear effort to dispel common myths about mental health problems. These observations are concerning and are particularly significant in settings where mental health literacy may be relatively low and when apps are accessed by inexperienced or naive users.

Reflections on, Expert Involvement, Evidence Base, and Content Quality

The majority of the reviewed apps were developed by for-profit organizations. Involvement of mental health professionals in app development was not mentioned in 228 (65.1%) apps, mentioned only cursively in 45 (12.9%), and clearly specified in just 77 (22%) of the 350 apps. Only 38 (10.9%) reviewed apps mentioned empirical research related to the app, with the majority (n=24) focusing on usability, acceptability, and user satisfaction. Direct empirical research on the apps themselves was rare, and more than half of the apps did not provide users with information about the empirical basis for their content. Generating high-quality evidence through rigorous trials is an important goal, but it is a cumulative process that does not always result in the immediate public availability of such apps. Against the backdrop of low mental health literacy [6], the absence of empirically validated or empirically informed content—often due to limited involvement of mental health professionals in app development [1]—further increases challenges and risks for users, who may struggle to assess the validity of app content [25,33].

The validity and quality of app content have been identified as concerns in several studies. For example, one study on depression apps found that only 40% involved a mental health expert in their development [26], while another study on social anxiety apps reported that the content often lacked sufficient information to determine whether it was based on expert validation [28]. Another review of depression apps, examining adherence to evidence-based guidelines, found that 21% of the reviewed apps suggested scientifically unproven treatments

XSL•FO
RenderX

[34]. Similarly, a review of suicide prevention apps reported that only one-third included supplementary information on scientifically validated content for users [38]. Another study noted that the number of evidence-based components in apps was not significantly associated with average consumer ratings [32]. Consistent with these findings, existing reviews also indicate the presence of unscientific or misleading information in mental health apps, although this issue has not been comprehensively or thoroughly examined. As the majority of mental health apps available to consumers have not been validated through empirical studies, it is especially important that their content is evidence-informed and developed in collaboration with mental health experts [22].

Despite the increasing number of studies on mobile apps for mental health, previous reviews have highlighted a lack of empirical research on apps publicly available in virtual stores [1,34]. This suggests a gap between research and the accessibility of research-based apps on one hand, and the rapid proliferation of apps with minimal involvement of mental health professionals on the other. The current regulatory landscape for digital mental health needs further development to safeguard the interests of end users [18]. Additionally, there are relatively few implementation studies on mental health apps that could inform real-world adoption, engagement, and effectiveness across different settings [1]. The barriers limiting the public availability of evidence-based or evidence-informed mental health apps must be identified and addressed.

Encouragement for Professional Help-Seeking

A notable concern was the lack of adequate information, guidance, and prompts to help users of self-help apps recognize when professional support is necessary. This issue is particularly critical in unguided self-help apps, as users may not be sufficiently aware that the app may not be suitable for everyone or that certain situations require transitioning from self-help to professional care. Pointers to seek professional help were present in 139 of the 350 (39.7%) reviewed apps, mostly in the form of disclaimers stating that the apps were not diagnostic tools or substitutes for consultation with a health professional. Nudges or motivational prompts to seek professional help, based on assessment results or user inputs, were noted in slightly less than a quarter of the apps. Given the mental barriers to help-seeking—such as normalization of distress, stigma, and preference for self-reliance—alongside low mental health literacy, particularly in the Indian context [6-8], it is ethically crucial for mental health apps to clearly delineate their scope, including for whom they may be most useful. Beyond providing information or guidance, these apps need to actively nudge users to seek professional help when appropriate, rather than relying solely on self-help [34,35].

Evaluation of App Quality Using MARS

Existing reviews indicate that most apps score high on functionality (eg, performance, ease of use) and that MARS ratings for functionality and aesthetics tend to correlate significantly with consumer ratings of the respective apps [32,51]. The MARS average rating of all apps reviewed in this study shows a similar trend to that observed in previous studies [51-53].

In this study, all app categories—except informative/educative and assessment apps—had average ratings above 4 on both the functionality and aesthetics subscales. Engagement ratings were highest (close to 4) for multifunction apps and lowest for assessment and informative/educative apps. Average ratings across all app categories were below 4 on the information subscale. A closer examination of apps with low information quality (ratings <3) highlighted several areas of concern, as discussed earlier.

Examining the apps with relatively higher MARS ratings, it was noted that two-thirds or more of the apps in the multifunction and intervention categories had average MARS ratings of 4 or higher. This review considered both MARS ratings and app descriptions based on several other parameters deemed relevant for mental health apps. While the MARS framework provided a consistent and standardized approach to evaluation, certain nuances important for mental health apps were not fully captured within its subscales. Ratings for each subscale were based on individual item scores, which showed some variability. Moreover, the average rating across subscales—where all subscales were weighted equally—tended to be relatively high, particularly because many apps received high scores on functionality and aesthetics. This issue has been noted in previous studies using MARS [54]. These findings are consistent with existing studies on mental health apps assessed using MARS [51,52]. Future studies could benefit from considering additional parameters specific to mental health to achieve a more comprehensive understanding of apps available in virtual stores.

Additional Observations

A slightly larger proportion of reviewed apps (220/350, 62.9%) were available on Google Play, which may partly reflect the market share of Android phones [42] and the relative ease of hosting mental health apps on this platform, making it appealing to developers [55]. It should be noted that both factors are rapidly evolving, with a growing market for Apple phones in India [56] and changes in the requirements and validation steps for hosting health apps on the Google Play [57]. Nearly 40% of the reviewed apps were released after 2020 (141/350, 40.3), potentially reflecting the rapid advancement of digital health and increased awareness of mental well-being, particularly during and after the COVID-19 pandemic. Common search terms such as “depression” and “anxiety” yielded nearly a quarter of relevant apps (128/495, 25.9% to 133/497, 26.8%) after duplicates were removed and ineligible apps filtered out according to the review criteria. Previous studies have similarly documented low rates of finding relevant apps for specific purposes using keyword-driven searches [28,58,59]. These figures also offer a rough indication of the challenges users may face when trying to find relevant apps using specific keywords.

Of the 350 reviewed apps, 40 (11.4%) did not have a privacy policy accessible to users, and in another 49 (14%), the policy was not clearly explained. Details on data sharing and the option to delete accounts were missing in 54 (15.4%) and 168 (48%) apps, respectively, while the duration of data retention was not specified in 143 (40.9%) apps. These observations raise concerns regarding privacy and security, consistent with findings from

XSL•FO
RenderX

other reviews [35,60]. In an Indian review of 50 depression apps, more than one-fourth did not have a privacy policy accessible through the store description, and some apps requested potentially risky permissions (eg, modify or delete USB storage) [26]. Another review of depression apps found that although most included a privacy policy, only one-third presented it to users before account creation [27]. End user reviews have also highlighted privacy concerns [33].

Our review found that several apps required in-app purchases, although evaluating their costs was beyond the scope of this analysis. Notably, only 119 of the 350 (34%) apps were entirely free, which could limit accessibility for users in low- and middle-income countries. Most apps originated from North America and Europe, while in about a quarter of cases, this information was unclear. Of these 350, 62 (17.7%) apps were developed in Asia, and 10 apps were available in an Indian language alongside English. This underscores the need for investment in the development and adaptation of evidence-informed and evidence-based apps that provide socioculturally appropriate content for Indian users.

Strengths and Limitations

This is one of the first systematic reviews to comprehensively examine mental health apps accessible on Google Play and Apple App Store for Indian users. The review had several strengths, including the use of a broad range of search terms related to mental health conditions and therapies, which ensured comprehensive coverage. Additionally, it incorporated recommended guidelines for systematic app reviews, employed a detailed set of evaluation parameters, and used a standardized rating tool, all of which enhance its potential utility for future research.

Nevertheless, the review had some limitations. First, we were able to evaluate only free apps, apps with free trials, and the freely accessible portions of paid apps. This may have provided an incomplete basis for rating some apps. Second, the review was intentionally limited to apps addressing mental health concerns and therapies, and did not include search terms related to general well-being. This was intentionally done because wellness apps constitute a large and heterogeneous category. However, this approach may have excluded some apps that provide useful information for managing mental health concerns but do not explicitly highlight this in their names or app store descriptions. Third, the detailed and comprehensive nature of the review, combined with the constantly evolving content of apps on virtual stores, imposed constraints on the review strategy. Apps were assigned to 4 different members of the review team for evaluation, rather than being reviewed concurrently by multiple raters. To enhance consistency across evaluators, several measures were implemented: joint training sessions for reviewers; discussions following independent ratings of a subset of apps to establish consensus; joint ratings and deliberations on selected apps; adherence to an internal guideline for MARS ratings; and a cross-validation exercise involving senior faculty members on a selected subsample of apps.

Future Directions and Conclusions

Overall, the findings of this systematic review of mental health apps available in virtual stores for Indian users provide valuable insights to guide future research, development, and policy making in the digital mental health sector. The limited empirical evidence supporting publicly available apps, coupled with indications of insufficient involvement of mental health professionals in their development, underscores critical gaps that require attention. While further studies are needed to strengthen the evidence base on app effectiveness and related factors, future research should also aim to identify barriers that limit the meaningful involvement of mental health experts in app design and development. A related concern, noted previously, is that many well-researched apps remain unavailable in app stores [19,50]. Efforts to identify and address bottlenecks in deploying rigorously researched apps on virtual platforms, and to promote their adoption across diverse settings, could enhance both the accessibility and impact of mental health apps.

This review highlights a marked underrepresentation of mobile mental health apps from the Asian continent, with just 62 (17.7%) developed in the region and only a handful originating from India. This gap underscores the urgent need for research that evaluates the sociocultural relevance of app content and rigorously tests the effectiveness of publicly available tools within Indian contexts. Addressing this challenge will require concerted efforts, including adequate funding support from local government agencies and other stakeholders, as well as active involvement of mental health professionals in both clinical and research domains. Such engagement is crucial for the development of indigenous apps that incorporate evidence-informed content tailored to local needs.

As the mental health app sector in India continues to expand rapidly, it is increasingly important to strengthen regulatory guidelines that protect user privacy, address safety concerns, and uphold quality standards—measures that are essential for fostering trust in digital mental health solutions. Additionally, exploring public-private partnership models could support the deployment and long-term sustainability of these apps. Equally important are initiatives to improve digital mental health literacy among potential users. With an increasing number of mental health apps of varying quality available in app stores, educating individuals about the scope and limitations of these tools—and guiding them in identifying apps that best meet their needs—can empower users, particularly those unfamiliar with mental health apps, to make informed decisions and maximize the benefits of digital mental health interventions.

Acknowledgments

The study was conducted at the Indian Council of Medical Research (ICMR)-Centre for Advanced Research in Digital Interventions for Mental Healthcare at the National Institute of Mental Health and Neuro Sciences (NIMHANS), Bengaluru. We express our

XSL•FO
RenderX

sincere gratitude to the ICMR for their continued support in advancing digital mental health research in India. We also thank the NIMHANS, Bengaluru, for institutional support throughout the study. We acknowledge and appreciate the efforts of the developers who made their mental health apps freely available to the public, enabling us to conduct this systematic review. Generative artificial intelligence tools were not used in the preparation of this manuscript. This study was funded by the ICMR as part of funding support for the Centre for Advanced Research in Digital Interventions for Mental Healthcare, NIMHANS.

Data Availability

The data of this study will be made available upon request to the corresponding author (SM).

Authors' Contributions

SM contributed to conceptualization, design, analysis, supervision, writing, and revising. RT was involved in conceptualization, design, development of selection criteria, analysis, supervision, and editing. P Sengupta and AK contributed to data analysis, validation, and synthesis revisions. AF handled data collection, analysis, synthesis, and writing portions of the first draft. P Sharma conducted the literature review, data collection, analysis, synthesis, and contributed to writing portions of the first draft. P Sudhir was involved in analysis, validation, supervision, writing, and revision. TKS reviewed and wrote the technological aspects and contributed to revision. GNR contributed to analysis, supervision, writing, and revision. RS provided supervision and consultations on revision and finalization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PASSR checklist.
[PDF File (Adobe PDF File), 139 KB - mhealth_v13i1e79238_app1.pdf]

References

1. Mehrotra S, Tripathi R. Recent developments in the use of smartphone interventions for mental health. Curr Opin Psychiatry 2018 Sep;31(5):379-388. [doi: 10.1097/YCO.0000000000000439] [Medline: 29985179]
2. Malhotra S. Mental health apps: a new field in community mental health care. Indian J Soc Psychiatry 2023 Jun 27;39(2):a. [doi: 10.4103/ijsp.ijsp_145_23]
3. Grand VR. Grand View Research. URL: https://www.grandviewresearch.com/industry-analysis/mental-health-apps-market-report [accessed 2025-06-13]
4. India mental health apps market analysis (2023-2030). Insights10. 2024 Jul 11. URL: https://www.insights10.com/report/india-mental-health-apps-market-analysis/ [accessed 2025-06-13]
5. Gautham MS, Gururaj G, Varghese M, Benegal V, Rao GN, Kokane A, NMHS Collaborators Group<xref ref-type="fn" rid="132fn1-0020764020907941" ptype="f0020764020907941" citart="citart1">*</xref>. The National Mental Health Survey of India (2016): prevalence, socio-demographic correlates and treatment gap of mental morbidity. Int J Soc Psychiatry 2020 Jun;66(4):361-372. [doi: 10.1177/0020764020907941] [Medline: 32126902]
6. Patel V, Xiao S, Chen H, Hanna F, Jotheeswaran AT, Luo D, et al. The magnitude of and health system responses to the mental health treatment gap in adults in India and China. Lancet 2016 Dec 17;388(10063):3074-3084. [doi: 10.1016/S0140-6736(16)00160-4] [Medline: 27209149]
7. Sanghvi PB, Mehrotra S. Help-seeking for mental health concerns: review of Indian research and emergent insights. JHR 2021 Mar 10;36(3):428-441. [doi: 10.1108/jhr-02-2020-0040]
8. Sanghvi P, Mehrotra S, Das SK, Sharma MK. Barriers to seek professional help for mental health scale: development and psychometric validation. Indian J Psychol Med 2024 Jun 22:02537176241256480. [doi: 10.1177/02537176241256480] [Medline: 39564249]
9. Mahapatra P, Seshadri S. Mental health in India: evolving strategies, initiatives, and prospects. Lancet Reg Health Southeast Asia 2024 Jan;20:100300 [FREE Full text] [doi: 10.1016/j.lansea.2023.100300] [Medline: 38236493]
10. Maroju RG, Choudhari SG, Shaikh MK, Borkar SK, Mendhe H. Role of telemedicine and digital technology in public health in India: a narrative review. Cureus 2023 Mar;15(3):e35986 [FREE Full text] [doi: 10.7759/cureus.35986] [Medline: 37050980]
11. Statista. Number of smartphone users in India from 2010 to 2023, with projections up to 2040. Statista. URL: https://www.statista.com/statistics/467163/forecast-of-smartphone-users-in-india/ [accessed 2025-06-13]
12. India mental health market: industry trends, forecasts, and outlook (2024-2032). Astute Analytica. 2024 Aug. URL: https://www.astuteanalytica.com/industry-report/india-mental-health-market [accessed 2025-06-13]
13. Stewart M. The rise of mental health apps in India: are they effective? Northern Feeling. 2025 Jun 7. URL: https://northernfeeling.com/the-rise-of-mental-health-apps-in-india-are-they-effective/ [accessed 2025-06-13]

XSL•FO
RenderX

14. Remedios F. The 5 mental health apps that deserve space on your phone. Vogue India. URL: https://www.vogue.in/beauty/content/the-5-mental-health-apps-that-deserve-space-on-your-phone? [accessed 2025-06-13]
15. Almuqrin A, Hammoud R, Terbagou I, Tognin S, Mechelli A. Smartphone apps for mental health: systematic review of the literature and five recommendations for clinical translation. BMJ Open 2025 Feb 11;15(2):e093932 [FREE Full text] [doi: 10.1136/bmjopen-2024-093932] [Medline: 39933815]
16. Sudarshan S, Mehrotra S, Thirthalli J. Effectiveness of blended internet-based self-help and face-to-face intervention for depression: a pilot study from India. Indian J Psychol Med 2025 Mar;47(2):142-153 [FREE Full text] [doi: 10.1177/02537176241238289] [Medline: 39564242]
17. Division of Digital Psychiatry, Beth Israel Deaconess Medical Center. Agency for Health care Research and Quality. 2023. URL: https://mindapps.org/Apps [accessed 2025-06-13]
18. Singh Sethi MI, Kumar RC, Manjunatha N, Naveen Kumar C, Math SB. Mental health apps in India: regulatory landscape and future directions. BJPsych Int 2025 Feb 10;22(1):2-5. [doi: 10.1192/bji.2024.20] [Medline: 40290478]
19. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. JMIR Mhealth Uhealth 2016 Aug 09;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
20. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
21. Marshall JM, Dunstan DA, Bartik W. Apps with maps-anxiety and depression mobile apps with evidence-based frameworks: systematic search of major app stores. JMIR Ment Health 2020 Jun 24;7(6):e16525 [FREE Full text] [doi: 10.2196/16525] [Medline: 32579127]
22. Montijo S, Grohol J. The 9 best evidence-based mental health apps in 2022. Psych Central. 2022 Feb 21. URL: https://psychcentral.com/health/evidence-based-mental-health-apps [accessed 2025-06-14]
23. Mehrotra S, Kumar S, Sudhir P, Rao GN, Thirthalli J, Gandotra A. Unguided mental health self-help apps: reflections on challenges through a clinician's lens. Indian J Psychol Med 2017;39(5):707-711. [doi: 10.4103/IJPSYM.IJPSYM_151_17] [Medline: 29200577]
24. Al Dweik R, Ajaj R, Kotb R, Halabi DE, Sadier NS, Sarsour H, et al. Opportunities and challenges in leveraging digital technology for mental health system strengthening: a systematic review to inform interventions in the United Arab Emirates. BMC Public Health 2024 Sep 27;24(1):2592 [FREE Full text] [doi: 10.1186/s12889-024-19980-y] [Medline: 39334131]
25. Camacho E, Cohen A, Torous J. Assessment of mental health services available through smartphone apps. JAMA Netw Open 2022 Dec 01;5(12):e2248784 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.48784] [Medline: 36576737]
26. Singh S, Sharma P, Ghimire P, Shrestha R, Gnanavel S. Assessment of app store description and privacy policy to explore ethical and safety concerns associated with the use of mental health apps for depression. Indian J Psychol Med 2023 Mar 14;45(2):173-178 [FREE Full text] [doi: 10.1177/02537176221142046] [Medline: 36925500]
27. Martinengo L, Stona A, Griva K, Dazzan P, Pariante CM, von Wangenheim F, et al. Self-guided cognitive behavioral therapy apps for depression: systematic assessment of features, functionality, and congruence with evidence. J Med Internet Res 2021 Jul 30;23(7):e27619 [FREE Full text] [doi: 10.2196/27619] [Medline: 34328431]
28. Hammond TE, Lampe L, Campbell A, Perisic S, Brakoulias V. Psychoeducational social anxiety mobile apps: systematic search in app stores, content analysis, and evaluation. JMIR Mhealth Uhealth 2021 Sep 21;9(9):e26603 [FREE Full text] [doi: 10.2196/26603] [Medline: 34546179]
29. Haque MDR, Rubya S. An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. JMIR Mhealth Uhealth 2023 May 22;11:e44838. [doi: 10.2196/44838] [Medline: 37213181]
30. Widnall E, Grant CE, Wang T, Cross L, Velupillai S, Roberts A, et al. User perspectives of mood-monitoring apps available to young people: qualitative content analysis. JMIR Mhealth Uhealth 2020 Oct 10;8(10):e18140 [FREE Full text] [doi: 10.2196/18140] [Medline: 33037875]
31. Carrouel F, du Sartz de Vigneulles B, Bourgeois D, Kabuth B, Baltenneck N, Nusbaum F, et al. Mental health mobile apps in the French app store: assessment study of functionality and quality. JMIR Mhealth Uhealth 2022 Oct 12;10(10):e41282 [FREE Full text] [doi: 10.2196/41282] [Medline: 36223178]
32. Lau N, O'Daffer A, Yi-Frazier JP, Rosenberg AR. Popular evidence-based commercial mental health apps: analysis of engagement, functionality, aesthetics, and information quality. JMIR Mhealth Uhealth 2021 Jul 14;9(7):e29689 [FREE Full text] [doi: 10.2196/29689] [Medline: 34259639]
33. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
34. Martinengo L, Stona A, Tudor Car L, Lee J, Griva K, Car J. Education on depression in mental health apps: systematic assessment of characteristics and adherence to evidence-based guidelines. J Med Internet Res 2022 Mar 09;24(3):e28942 [FREE Full text] [doi: 10.2196/28942] [Medline: 35262489]
35. Kumar S, Mehrotra S. Free mobile apps on depression for Indian users: a brief overview and critique. Asian J Psychiatr 2017 Aug;28:124-130. [doi: 10.1016/j.ajp.2017.03.031] [Medline: 28784364]

XSL•FO
RenderX

36. Bhat B, Nayak PP, Vasthare R, Singhal DK. Mobile applications (apps) for tobacco cessation: behaviour change potential and heuristic analysis using the Mobile Application Rating Scale (MARS). F1000Res 2023;12:1413. [doi: 10.12688/f1000research.142038.2] [Medline: 40433635]
37. Tanvi Gadgil, Shailaja Jadhav, Anjali Kumari, Mrunali Dasari, Chanchal Bhangdia. Mental health mobile application with diagnosis, sentiment analysis and chatbot. IJSRCSEIT 2022 Jun 30;8(3):94-100 [FREE Full text] [doi: 10.32628/cseit228327]
38. Sudarshan S, Mehrotra S. Suicide prevention mobile apps for Indian users: an overview. Cureus 2021 Jul;13(7):e16770. [doi: 10.7759/cureus.16770] [Medline: 34476140]
39. Gasteiger N, Dowding D, Norman G, McGarrigle L, Eost-Telling C, Jones D, et al. Conducting a systematic review and evaluation of commercially available mobile applications (apps) on a health-related topic: the TECH approach and a step-by-step methodological guide. BMJ Open 2023 Jun 12;13(6):e073283. [doi: 10.1136/bmjopen-2023-073283] [Medline: 37308269]
40. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
41. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
42. Android and iOS own 95.1% and 3.93% market share. ET Brand Equity (Economic Times). 2022 May 19. URL: https://brandequity.economictimes.indiatimes.com/news/research/android-and-ios-own-95-1-and-3-93-market-share-report/91657405 [accessed 2025-06-13]
43. Census of India 2011 - Language Atlas - India. Census of India (Government of India). 2022 Apr. URL: https://censusindia.gov.in/nada/index.php/catalog/42561 [accessed 2025-06-13]
44. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
45. Hayes SC, Hofmann SG. "Third-wave" cognitive and behavioral therapies and the emergence of a process-based approach to intervention in psychiatry. World Psychiatry 2021 Oct 09;20(3):363-375 [FREE Full text] [doi: 10.1002/wps.20884] [Medline: 34505370]
46. Cunha O, Pereira B, Sousa M, Rodrigues AC. Cognitive behavioural “third wave” therapies in the treatment of justice-involved individuals: a systematic review. Aggression and Violent Behavior 2024 May;76:101923. [doi: 10.1016/j.avb.2024.101923]
47. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content. Behav Res Ther 2019 Dec;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]
48. Eysenck MW, Fajkowska M. Anxiety and depression: toward overlapping and distinctive features. Cogn Emot 2018 Nov 13;32(7):1391-1400. [doi: 10.1080/02699931.2017.1330255] [Medline: 28608767]
49. Jiménez-Orenga N, Miguel C, González-Robles A, Fernández-Álvarez J, Grimaldos J, Bretón-López J, et al. Transdiagnostic psychological interventions for emotional disorders: a comprehensive meta-analysis. J Affect Disord 2025 Nov 01;388:119537. [doi: 10.1016/j.jad.2025.119537] [Medline: 40449740]
50. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016;11(5):e0154248. [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
51. Ki E, Jang J, Kang D. Evaluating mobile mental health apps using the Mobile Application Rating Scale (MARS) in the U.S. Advances in Mental Health 2024 Aug 15;23(2):222-238. [doi: 10.1080/18387357.2024.2388680]
52. Martín-Martín J, Muro-Culebras A, Roldán-Jiménez C, Escriche-Escuder A, De-Torres I, González-Sánchez M, et al. Evaluation of Android and Apple Store depression applications based on Mobile Application Rating Scale. Int J Environ Res Public Health 2021 Nov 27;18(23):1-9 [FREE Full text] [doi: 10.3390/ijerph182312505] [Medline: 34886232]
53. Domhardt M, Messner E, Eder A, Engler S, Sander LB, Baumeister H, et al. Mobile-based interventions for common mental disorders in youth: a systematic evaluation of pediatric health apps. Child Adolesc Psychiatry Ment Health 2021 Sep 13;15(1):49 [FREE Full text] [doi: 10.1186/s13034-021-00401-6] [Medline: 34517896]
54. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020 Nov 2;15(11):e0241480. [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
55. Lee S. Apple App Store vs Google Play Store: how different are their app store approval processes? Glance Group. 2024 May 3. URL: https://thisisglance.com/blog/apple-store-vs-google-play-store [accessed 2025-06-14]
56. India to hit 10 million Apple iPhone users this year. UMA Technology. 2025 Jan 11. URL: https://umatechnology.org/india-to-hit-10-million-apple-iphone-users-this-year-report/ [accessed 2025-06-15]
57. App testing requirements for new personal developer accounts. Google Play Store. URL: https://support.google.com/googleplay/android-developer/answer/14151465?hl=en [accessed 2025-06-14]

XSL•FO
RenderX

58. Alyami M, Giri B, Alyami H, Sundram F. Social anxiety apps: a systematic review and assessment of app descriptors across mobile store platforms. Evid Based Ment Health 2017 Aug 30;20(3):65-70 [FREE Full text] [doi: 10.1136/eb-2017-102664] [Medline: 28666986]
59. Psihogios AM, Stiles-Shields C, Neary M. The needle in the haystack: identifying credible mobile health apps for pediatric populations during a pandemic and beyond. J Pediatr Psychol 2020 Nov 01;45(10):1106-1113. [doi: 10.1093/jpepsy/jsaa094] [Medline: 33068424]
60. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]

Abbreviations

ACT: acceptance and commitment therapy
AMSTAR: A Measurement Tool to Assess Systematic Reviews
CBT: cognitive behavioral therapy
DBT: dialectical behavior therapy
ICMR: Indian Council of Medical Research
INPLASY: International Platform of Registered Systematic Review and Meta-analysis Protocols
MARS: Mobile Application Rating Scale
PASSR: Protocol for App Store Systematic Reviews
PICO: Population, Intervention, Comparison, and Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
SPIDER: Sample, Phenomenon of Interest, Design, Evaluation, Research
TECH: Target user, Evaluation focus, Connectedness, and Health

Edited by L Buis; submitted 19.06.25; peer-reviewed by V Malhotra, A Eisingerich; comments to author 18.07.25; revised version received 23.07.25; accepted 28.08.25; published 26.09.25.

Please cite as:
Mehrotra S, Tripathi R, Sengupta P, Karishiddimath A, Francis A, Sharma P, Sudhir P, Srikanth TK, Rao G, Sagar R
Evaluating Characteristics and Quality of Mental Health Apps Available in App Stores for Indian Users: Systematic App Search and Review
JMIR Mhealth Uhealth 2025;13:e79238
URL: https://mhealth.jmir.org/2025/1/e79238
doi:10.2196/79238
PMID:

©Seema Mehrotra, Ravikesh Tripathi, Pramita Sengupta, Abhishek Karishiddimath, Angelina Francis, Pratiksha Sharma, Paulomi Sudhir, T K Srikanth, Girish Rao, Rajesh Sagar. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Selection of Behavior Change Techniques for Asthma Medication Adherence Apps: Evidence-Based Design Study

Alison J Wright[1], PhD; Jeremy Holland[2], PhD; Iain Simpson[3], PhD; Samantha Walker[4], PhD; Naomi Bennett-Steele[4], PhD; John Weinman[1], PhD, DSC

[1]Institute of Pharmaceutical Science, King's College London, London, United Kingdom

[2]Healthcare Programmes, Centre for Business Innovation, 18 Sedley Taylor Road, Cambridge, United Kingdom

[3]Pharmaceutical Drug Delivery, Phillips Medisize, Cambridge, United Kingdom

[4]Research and Innovation, Asthma + Lung UK, London, United Kingdom

Corresponding Author:
Jeremy Holland, PhD
Healthcare Programmes, Centre for Business Innovation, 18 Sedley Taylor Road, Cambridge, United Kingdom

Abstract

Background: Poor medication adherence is a widespread issue that causes adverse patient outcomes and is expensive for all aspects of the health care system. Developing cost-effective and scalable interventions to promote medication adherence is a key goal. Mobile apps hold promise as a mode of delivery for adherence interventions, but app design rarely takes into account the behavioral influences on nonadherence with sufficient rigor. As a result, apps may not realize their full potential in enhancing adherence. Medication nonadherence is common among adults prescribed preventer inhalers for asthma and has a variety of influences, creating a need to identify what components behavior change technique (BCT) apps should include to effectively tackle each influence.

Objective: This study aimed to identify the most acceptable and practicable BCTs to include in a medication adherence app targeting factors that influence preventer inhaler adherence in adults with asthma.

Methods: Key influences on preventer inhaler adherence in adults with asthma were identified based on reviews of peer-reviewed and gray literature and domain expert knowledge. These influences were then mapped to a published set of 26 mechanisms of action (MoAs) of behavior change interventions. Next, candidate BCTs to change each MoA were identified using the Theory and Techniques tool, a web-based resource that reflects almost 100 expert behavioral scientists' consensus about which BCTs are most likely to change particular MoAs. Finally, candidate BCTs were filtered by considering their potential acceptability and practicability.

Results: A total of 31 influences on preventer inhaler adherence were identified and coded to 15/26 of the influences on behavior listed by the Theory and Techniques tool. The initial mapping of influences on behavior to candidate BCTs to change those influences identified 41 candidate BCTs. After considering the potential acceptability and practicability of the candidate BCTs, the number of BCTs suggested for inclusion was reduced to 24.

Conclusions: Using an evidence-based approach, this study identified 24 BCTs that may be particularly useful to include in apps promoting adherence to preventer inhalers in order to target particular influences on adherence. The list can be used by app developers to improve the quality of adherence behavior change support that their apps provide or by health care decision-makers to identify which apps contain elements addressing a range of adherence difficulties.

(JMIR Mhealth Uhealth 2025;13:e49348) doi:10.2196/49348

KEYWORDS

medication adherence; patient adherence; behavior change; asthma; apps for medication adherence; medication; app; behavior; patient outcome; app design; design; acceptability; practicability; decision-making; decision

Introduction

Background

The World Health Organization estimates that 30% - 50% of patients do not take their long-term medication as prescribed [1]. Medication nonadherence in long-term conditions is associated with reduced quality of life, poorer clinical outcomes, increased risk of premature death, and increased use of health care services. Estimates suggest that nonadherence contributes to nearly 200,000 premature deaths per year and costs European governments €125 billion (1 €=US $1.15) annually in excess health care services [2]. Taking medication as prescribed is a behavior, and so it is helpful to use a behavioral science perspective to understand and promote it. While there are a

XSL•FO
RenderX

number of approaches to behavior change intervention design, many share the approach of identifying modifiable influences on the target behavior and then identifying appropriate intervention components to change these modifiable influences.

Considerable effort has been devoted to the development and evaluation of interventions to enhance medication adherence [3]. Complex, carefully designed interventions have been shown to improve both adherence and health outcomes [4,5]. However, the difficulty is making these interventions sufficiently cost-effective and scalable. Moreover, adherence interventions delivered by health care professionals may not be offered at the times or in the locations most convenient to people with the health problem in question. In contrast, the pervasiveness of smartphones in high-income countries makes the potential of apps for providing convenient, lower-cost, and scalable yet personalized medication adherence interventions very appealing in this context. A recent meta-analysis of randomized controlled trials found that app-based medication adherence interventions enhanced self-reported adherence [6], but firm conclusions were hard to make due to the small number of included studies, the heterogeneity of study populations, and variation in how adherence was measured in different trials.

Medication adherence is a behavior, and so interventions to target adherence will benefit from building on behavioral science evidence. Based on the literature, behavioral scientists developed a taxonomy of 93 behavior change techniques (BCTs), defined as the smallest self-contained components of an intervention aimed at changing behavior [7]. Examples of BCTs include "action planning" (making a detailed plan for the performance of the behavior, which includes at least one of context, frequency, duration, or intensity) and "social comparison" (drawing attention to other people's behavior and comparing their behavior with the person's own behavior). A review of 166 apps targeting medication adherence [8] discovered that only 12 of over 90 possible BCTs were found in at least one of the apps. The most commonly used BCTs were "action planning" and "prompts or cues," both of which target developing a medication-taking routine and reducing the chances of forgetting to take a dose of medication. However, medication adherence is influenced by many factors, of which having a routine and remembering to take the medication are only 2 [9,10]. Apps are likely to be more effective when they target the full range of key influences on medication adherence behavior. A related issue is that many apps are not tailored to the specific causes of medication nonadherence for individual users. For example, an app providing simple reminders to take medication is unlikely to benefit a patient who is concerned about potential side effects. Similarly, providing information about medication benefits is unlikely to make a difference to the adherence of a patient who is already motivated to take their medicine but is nonadherent due to a lack of a routine for doing so.

Where apps do attempt to tackle specific influences on medication adherence, their developers face several challenges. First, they need to integrate a method of assessing the influences that are most relevant to each app user's adherence. Second, app designers may have difficulty applying evidence as to which of these BCTs are most likely to change these influences, due to the expertise required to assess and interpret this evidence. A further issue in selecting intervention components is that some BCTs may not be acceptable to app users or practical for individuals to use in their daily lives. Therefore, we need a way to identify what BCTs a medication adherence app could usefully contain based on evidence about the causes of nonadherence and considering the likely acceptability and practicality of deploying these BCTs in an app. Identifying these BCTs gives app developers and designers a set of potential intervention components to work with.

This study originated from a working group of the Centre for Business Innovation Medical Adherence and Digital Health consortium, a European interdisciplinary and primarily industrial grouping, which meets regularly to consider key issues around the role of digital health in improving medication adherence. For the present piece of work, the Centre for Business Innovation collaborated with Asthma UK (now Asthma+ Lung UK) to map out the behavioral factors influencing adherence to preventer medication for adults with asthma and apply a rigorous approach to identify the BCTs most relevant to promoting medication adherence for people with asthma. Asthma is a common long-term condition, affecting an estimated 260 million people worldwide in 2019 [11]. Each year, there are 455,000 asthma-related deaths, two-thirds of which are preventable [12]. For the vast majority of people with asthma, the regular use of "preventer" inhalers is recommended to avoid or reduce experiencing asthma attacks. However, nonadherence is common, with observed adherence rates between 22% and 63% [13]. Nonadherence is associated with poorer symptom control, reduced quality of life, and an increased risk of emergency health care use and even mortality [13]. Delivering medication adherence support via an app may be particularly helpful for adults with asthma, many of whom are of working age and so may find attending clinic appointments for adherence support inconvenient.

Nonadherence to preventer inhalers has a variety of causes, such as not perceiving asthma inhalers as necessary, being concerned about side effects, or having difficulties accessing medication, which vary for different people with asthma [14,15]. Therefore, there is a need to identify the optimal BCTs to tackle specific influences on preventer inhaler nonadherence via an app-based intervention. Medication adherence apps have the potential to be tailored to tackle the barriers most relevant to a particular person with asthma. However, a small systematic review of app-based interventions targeting medication adherence in people with asthma found only 9 BCTs in the 4 evaluated apps [16]. Another recent study examined the extent to which 23 asthma apps contained BCTs alongside rating their quality in terms of 5 user experience domains [17]. The number of BCTs in each app ranged from 1 to 11, and those rated as higher quality contained more BCTs. However, these BCTs were not linked to specific influences or causes of nonadherence in any logical way. A second recent study [18] examined the functionality and quality of asthma mHealth apps, but focused only on the broad categories of BCTs that are included in the Mobile App Rating Scale. In the 53 apps they reviewed, the most commonly used BCTs were in the broad categories of "information or education" (62% of apps) and "self-monitoring

XSL•FO
RenderX

or tracking" (72% of apps). However, there was no assessment of whether the included BCTs were appropriately linked to targeted influences on asthma self-management. The process of identifying BCTs to tackle different influences on nonadherence to preventer inhalers serves as a good exemplar of the process of identifying BCTs to feature in apps promoting adherence to medication in other conditions.

Objectives

The purpose of this article is to describe and report the outcomes of a process to identify which BCTs could be included in a medication adherence app to promote adherence to preventer inhalers in people with asthma.

The article also serves as a case study of using behavioral science evidence to move from understanding the influences on nonadherence to deriving a set of BCTs that can be used in a medication adherence app.

Methods

Overview

To identify which BCTs might be usefully included in an adherence app, we used a 4-stage process:

1. Identify modifiable influences on adherence.
2. Map the influences to a published list of mechanisms of action (MoA) of behavior change interventions.
3. Map MoAs to the BCTs, which expert consensus indicated were likely to change those MoAs.
4. Filter BCTs in terms of their likely acceptability and practicability.

We describe each step in more detail in later sections.

Identifying Influences on Medication Adherence

For any particular health condition and therapy combination, we must first identify the influences on medication adherence. Asthma + Lung UK worked with Public Health England to compile an evidence-based list of barriers and facilitators of preventer inhaler adherence. Please see the multimedia appendix for the details of this process. Two types of sources were used to provide data on the barriers and enablers to the target behavior. The first, identified by Asthma + Lung UK and classified as gray literature, were unpublished reports of findings from qualitative studies on asthma self-management and barriers to self-management behaviors that were commissioned or conducted by Asthma + Lung UK as part of their work on behavioral insights in asthma (these reports are listed in Table S1 in the Multimedia Appendix 1). The second source was a rapid review of barriers and enablers to preventer inhaler adherence reported in the published literature. The rapid review synthesized published, peer-reviewed reviews of either quantitative or qualitative studies that reported barriers or facilitators to asthma self-management behaviors, including medication adherence (please see the Multimedia Appendix 1 for a list of the included reviews). Rapid review methods [19,20] were chosen in order to provide a solid overview of the evidence on barriers and facilitators to adherence in a more timely and less resource-intensive fashion than a full systematic review would have required.

Data pertinent to the influences (barriers and facilitators) to adherence were extracted from the results sections of the gray literature reports and published reviews. Supporting statements for each identified theme were extracted, including quotes from the original participants, if available, or the author's interpretation/narration of the findings. Text reporting barriers and facilitators was coded using inductive thematic coding by one reviewer. The findings were then checked by another reviewer, and any disagreements were resolved through discussion.

Mapping Influences to Behavior Change Intervention MOAs

To be able to use expert consensus evidence about which BCTs were best to change particular influences on behavior, we mapped each influence identified in Step 1 to the list of 26 intervention mechanisms of action used in a recent expert consensus project [21-24]. For example, "beliefs about side effects" was mapped to the MoA called "beliefs about consequences" which is defined as "Beliefs about the consequences of a behavior (ie, perceptions about what will be achieved and/or lost by undertaking a behavior, as well as the probability that a behavior will lead to a specific outcome,)" because side effects are a consequence of taking medication. Mapping was conducted by 1 author (AJW). A second author (JW) then independently reviewed the mapping. Any disagreements between the authors were resolved through discussion.

Map MoAs to BCTs That Expert Consensus Indicates Will Change Each MoA

Ideally, the choice of a BCT to change a particular MoA would be based on evidence showing that the BCT changed the MoA in experimental studies targeting the relevant behavior. However, such evidence does not yet exist for every BCT-MoA-behavior combination that intervention designers might be interested in. Given this, intervention designers may instead have to rely on expert consensus. A recent project [24] triangulated expert consensus on the links between 26 MoAs and 56 BCTs commonly used in behavior change interventions and created a web-based tool that allows users to view which BCTs experts thought were likely to influence particular MoAs. We used this web-based tool to identify what BCT experts thought were linked to each of the MoAs identified in the previous step. For many MoAs, more than one BCT was linked to that MoA. For example, the MoA of "beliefs about consequences" was linked to ten BCTs, such as "provide information about health consequences," "anticipated regret," "pros and cons," "comparative imagining of future outcomes," and "material incentive for behavior" to name but a few.

Filter BCTs in Terms of Their Likely Acceptability and Practicability

To reduce the number of candidate BCTs that intervention developers might need to consider, we applied the "acceptability" and "practicability" elements of the APEASE (Acceptability, Practicability, Effectiveness and cost-effectiveness, Affordability, Side effects/safety, and Equity) criteria suggested by West et al and Michie et al [25,26], as a

XSL•FO
RenderX

way to select elements of interventions. “Acceptability” concerns how far an intervention component is likely to be acceptable to key stakeholders, including target participants. “Practicability” concerns whether the intervention component can be implemented at scale within the intended context, material, and human resources. “Effectiveness” (and cost-effectiveness) concerns how likely the intervention is to achieve the intended outcomes and be good value for money. “Side effects/safety” concerns whether the intervention poses any safety concerns, while “Equity” refers to whether the intervention might reduce or increase social inequalities.

To apply the APEASE criteria, one has to specify the context in which the behavior change intervention will take place. In this case, the relevant context was that the app should work as a standalone intervention for adults with asthma prescribed preventer inhalers. Although it might be recommended to a patient by a health care professional, the app should not require further health care professional involvement because part of the impetus for creating an app was to have an adherence intervention that did not need considerable input from time-pressed health care professionals. We also specified that the app should not require the patient to have any further technology (eg, a laptop or a way of monitoring lung functioning) beyond their smartphone. Finally, we assumed that the app was to be rolled out in a high-income country where asthma is largely managed in primary care.

Each candidate BCT was appraised by 2 behavioral science experts with expertise in the design of behavior change interventions targeting adherence (AJW and JW). Any disagreements were resolved through discussion between the 2 experts. The experts first considered the practicability of delivering each BCT via an app in the context specified above. For BCTs judged practicable, the experts then considered their potential acceptability to app users. For example, the BCT of “provide information about health consequences” was seen as practicable to deliver within an app and likely to be acceptable to app users. In contrast, the BCT of “anticipated regret” was viewed as potentially practicable to deliver within an app but unlikely to be acceptable to app users, who might view it as “guilt tripping” them about taking their medication, potentially leading to disengagement with the intervention.

Ethical Considerations

This study did not involve human participants, primary data collection, observations of public behavior, or secondary analysis of previously collected individual-level data. All information analyzed was obtained from existing literature reviews and the published results of expert consensus exercises. The judgments and interpretations presented were made solely by the study authors as part of the analytic methodology and do not constitute human subjects research. As such, institutional review board approval and informed consent were not required. No individuals were recruited or compensated.

Results

Step 1: Identifying Influences on Medication Adherence

The review conducted by Asthma + Lung UK and Public Health England identified 31 influences on preventer inhaler adherence (Table 1).

XSL•FO
RenderX

Table . Identified influences on preventer inhaler adherence mapped to mechanisms of action.

Influence on adherence	Example	Mapping of influence to the mechanisms of action defined by Michie et al [24]
Limited awareness that there is a need to take inhalers even if symptoms are controlled	"I think if it's quite severe, you have to control it, it's a real part of your life in terms of medication, everything, lifestyle. Whereas if it's mild, it rarely flares up, but when it does, then you really manage it to the limit."	Knowledge
Confusion about how much to take and when	"But obviously it says two puffs on there, but how many is too many?"	Knowledge
Confusion over which inhaler to take when if multiple	"There are so many different pumps, by the time you've gone through every single colour, which one's got steroids, which one hasn't, you must be thinking 'I'm going to die anyway.'"	Knowledge
Change environment to accommodate inhalers (eg, change of bag size)	"If you want to put a spacer as well as inhalers, etc, in it, you would need to go for a large one. I have a pretty toiletry bag. It keeps everything together and clean."	Environmental context and resources
Formation of habits as a result of taking regularly	"Having a fixed daily routine can promote adherence"	Behavioral cuing
Always have inhaler to hand or deliberately having multiple inhalers in different places	"So, I've done functional things, like I'll leave my asthma pump at my work drawer, places where you can go and leave things"	Environmental context and resources
Use of prompts and reminders	"Using environmental cues to trigger habitual taking of medication, such as reminder devices … can promote adherence"	Behavioral cuing
Symptoms serve as a reminder	"Experiencing symptoms serves as a reminder to use inhalers"	Behavioral cuing
Preference to use natural/herbal remedies over medication	"Patients may prefer to use folk remedies, behavioural strategies, and religious approaches before progressing to prescription medications"	Attitude toward the behavior
Viewing asthma outcomes as having more to do with faith/chance than treatment effectiveness	"Barriers…included the wrong beliefs [that]… asthma is more subject to faith and chance than to treatment effectiveness"	Beliefs about consequences
Forgetting to use them	"You just don't even remember needing to take the pump"	Memory, attention, and decision processes
Interpretation of respiratory symptoms as normal variation, rather than due to asthma	"Even people in general sometimes have trouble breathing...I don't know what's normal"	Memory, attention, and decision processes [regarding interpreting symptoms] and knowledge [of what's normal]
Using own judgment (over suggestions from professionals/medical tools)	"I might have had a low peak flow for a couple of days. But it kept instructing me to increase the dose, and I did not think it was necessary"	Memory, attention, and decision processes
Significant life event/asthma event prompts use	"Patients tend to be motivated to manage their asthma…when asthma affects a valued activity. Some were not motivated to act until it posed a life-threatening state"	MoA[a] depends on how the person interprets the meaning of that event. For asthma events, the MoA seems likely to be perceived susceptibility/vulnerability.
Stigma/social judgment for using inhalers in public	"Asthma patients never feel comfortable to use asthma medications especially inhaler in public because by this act they feel stigmatized."	Social influences
Opinions of friends/family/media (doubt from others that inhaler is needed	"Attitudes toward taking asthma medications are influenced by friends, family, the media"	Subjective norms
Poor health care professional/patient communication	"When I go to my appointment, they try to rush, get you in there, rush you out, get you some stuff, 'Take this, do this.'"	Social influences
Lack of time to take inhalers (daily life)	"At least one squirt and it does something. Not take two, breathe in for 10, hold that. You haven't got time. You really don't."	Environmental context and resources

XSL•FO
RenderX

Influence on adherence	Example	Mapping of influence to the mechanisms of action defined by Michie et al [24]
Inhalers out of date due to irregular use	"Yes, I've had one of them but by the time I get round to taking it, they're out of date, so…"	Environmental consequences and resources
Not wanting to identify as a 'sick' person	"I hate feeling like I'm different. I have a problem when it comes to that, and I hold off on my medicine."	Social/professional role and identity, also self-image
Level of acceptance of asthma as part of identity	"It depends on whether you've accepted asthma as part of your identity"	Self-image
Not wanting to feel controlled by asthma/dependent on medication	"Restricting use of asthma medicines (to not feel controlled or dependent on medicines)"	Self-image, also beliefs about consequences if the person perceives that a consequence of adherence is feeling controlled/dependent
Doubt about asthma diagnosis	"In terms of her diagnosis, she doesn't believe it addressed the actual condition she had, and she still believes it doesn't"	General attitude/belief, self-image
Confidence in using inhalers	"Medication adherence for asthma is strongly influenced by self-efficacy levels"	Beliefs about capabilities
Perceived efficacy of inhalers	"Adherence is positively affected by the belief that giving inhaled corticosteroids would protect the patient from getting worse"	Beliefs about consequences
Beliefs about side effects	"At least 50% refused to adhere to their prescribed therapy fully because of concerns over side effects"	Beliefs about consequences
Concern that inhalers do more harm than good	"Barriers related to these facts included the wrong beliefs; for instance, asthma medicines may be more harmful than beneficial"	Beliefs about consequences
Perceiving cause and effect of using inhalers on asthma symptoms	"I've had times when I have been on strict medication regimes and I have still had attacks"	Beliefs about consequences
Embarrassment (eg, of using inhalers in public)	"They identified some particular issues such as …embarrassment over medication use in public"	Beliefs about consequences and emotions (embarrassment has both cognitive and emotional components)
Discomfort with long-term use of inhalers	"One factor in their lack of adherence was high discomfort with long-term use of ICS"	Beliefs about consequences and emotions (discomfort has both cognitive and emotional components)
Prioritizing asthma to support	"I've got 2 kids; I cannot afford to be sick, I cannot afford not to have the energy to work and to do all of these things. So, I will take it"	Goals

[a]MoA: mechanism of action.

Step 2: Mapping Influences to Behavior Change Intervention Mechanisms of Action

Each of the influences identified above was then mapped to the definitions of the 26 MoAs used in the Theory and Techniques tool [22,24]. The mappings are shown in the third column of Table 1. Some influences were judged to contain elements of more than one MoA. For example, "Interpretation of respiratory symptoms as normal variation, rather than due to asthma" was judged to map to both memory, attention, and decision process (as interpreting one's symptoms is a decision) and to knowledge (of what is normal variation in respiratory symptoms).

Step 3: Mapping Mechanism of Actions to Behavior Change Techniques That Expert Consensus Suggests Will Change Each Mechanism of Action

Each MoA was linked to between 1 and 11 BCTs that might potentially work to change it (please see Table S3 in Multimedia Appendix 1 for details). In total, 41 BCTs were identified as possible candidates for inclusion in an app.

Step 4: Filtering BCTs in Terms of Their Likely Acceptability and Practicability

The final set of BCTs judged to be acceptable and practicable to change each mechanism of action is shown in Table 2. There were between 0 and 5 BCTs judged practicable and acceptable to change a MoA when delivered as part of a preventer inhaler adherence app for people with asthma, depending on the MoA. The reasoning behind the inclusion or exclusion of each BCT identified in the previous step can be viewed in Table S3 in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Final set of recommended behavior change techniques to change different influences on preventer inhaler use.

Influence on adherence	Closest mechanism of action	Final recommended BCTs[a,b]
Limited awareness that there is a need to take inhalers even if symptoms are controlled	• Knowledge	• 5.1 information about health consequences
Confusion about how much to take and when	• Knowledge	• 4.1 instruction on how to perform the behavior
Confusion over which inhaler to take when if multiple	• Knowledge	• 4.1 instruction on how to perform the behavior • 5.1 information about health consequences
Change environment to accommodate inhalers (eg, change of bag size)	• Environmental context and resources	• 12.1 restructuring the physical environment
Formation of habits as a result of taking regularly	• Behavioral cuing	• 1.4 action planning • 7.1 prompts/cues • 8.3 habit formation • 12.1 restructuring the physical environment
Always have inhaler to hand/deliberately having multiple inhalers in different places	• Environmental context and resources	• 7.1 prompts/cues (for having inhaler to hand) • 12.1 restructuring the physical environment (for having multiple inhalers in useful locations)
Use of prompts and reminders	• Behavioral cuing	• 1.4 action planning • 7.1 prompts/cues • 8.3 habit formation • 12.1 restructuring the physical environment
Symptoms serve as a reminder	• Behavioral cuing	• 1.4 action planning • 7.1 prompts/cues • 8.3 habit formation
Preference to use natural/herbal remedies over medication	• Attitude toward the behavior	• 5.1 information about health consequences • 9.1 credible source • 9.2 pros and cons (when accompanied by at least one of the two other BCTs listed in this column)
Viewing asthma outcomes as having more to do with faith/chance than treatment effectiveness	• Beliefs about consequences	• 5.1 information about health consequences • 5.2 salience of consequences • 5.3 information about social and environmental consequences • 5.6 provide information about emotional consequences • 9.2 pros and cons
Forgetting to use them	• Memory, attention, and decision processes	• 7.1 prompts/cues
Interpretation of respiratory symptoms as normal variation, rather than due to asthma	• Memory, attention, and decision processes (for deciding on the meaning of symptoms) • Knowledge (of what is normal variation)	Memory, attention, and decision processes • N/A[c] Knowledge • 5.1 information about health consequences
Using own judgment (over suggestions from professionals/medical tools)	• Memory, attention, and decision processes	• N/A
Significant life event/asthma event prompts use	• Depends how the person interprets the meaning of that event. Possibly perceived susceptibility/vulnerability.	• 5.2 salience of consequences
Stigma/social judgment for using inhalers in public	• Social influences	• 3.1 social support (unspecified) • 3.2 social support (practical) • 6.3 information about others' approval

XSL•FO
RenderX

Influence on adherence	Closest mechanism of action	Final recommended BCTs[a,b]
Opinions of friends/family/media (doubt from others that inhaler is needed)	• Subjective norms	• 6.3 information about others' approval
Poor health care provider/patient communication	• Social influences	• 3.1 social support (unspecified) • 3.2 social support (practical) • 6.3 information about others' approval
Lack of time to take inhalers (daily life)	• Environmental context and resources	• 3.2 social support (practical) • 12.1 restructuring the physical environment
Inhalers out of date due to irregular use	• Environmental context and resources	• N/A, suggest tackling this barrier by enhancing adherence by other means so that inhalers do not go out of date
Not wanting to identify as a "sick" person	• Social/professional role and identity, possibly also self-image	• 3.1 social support (unspecified) • 6.2 social comparison • 9.1 credible source
Level of acceptance of asthma as part of identity	• Self-image	• 13.1 identification of self as a role model
Not wanting to feel controlled by asthma/dependent on medication	• Possible belief about consequences if a consequence of adherence is feeling controlled/dependent, self-image	Beliefs about consequences: • 5.1 information about health consequences • 5.2 salience of consequences • 5.3 information about social and environmental consequences • 5.6 information about emotional consequences • 9.2 pros and cons • Self-image: N/A
Doubt about asthma diagnosis	• General attitude/belief and self-image	General attitude/belief: • 9.1 credible source • 9.2 pros and cons • Self-image: N/A
Confidence in using inhalers	• Beliefs about capabilities	• 4.1 instruction on how to perform the behavior • 6.1 demonstration of the behaviour • 8.1 behavioural practice/rehearsal • 15.1 verbal persuasion about capability • 15.3 focus on past success
Perceived efficacy of inhalers	• Beliefs about consequences	• 5.1 information about health consequences • 5.2 salience of consequences • 5.3 information about social and environmental consequences • 5.6 information about emotional consequences
Beliefs about side effects	• Beliefs about consequences	• 5.1 information about health consequences • 5.2 salience of consequences (possibly) • 9.2 pros and cons
Concern that inhalers do more harm than good	• Beliefs about consequences	• 5.1 information about health consequences • 5.2 salience of consequences • 9.2 pros and cons
Perceiving cause and effect of using inhalers on asthma symptoms	• Beliefs about consequences	• 5.1 information about health consequences • 5.2 salience of consequences
Embarrassment (eg, of using inhalers in public)	• Beliefs about consequences • Emotions	

XSL•FO
RenderX

Influence on adherence	Closest mechanism of action	Final recommended BCTs[a,b]
		Beliefs about consequences: • 5.2 salience of consequences • 5.3 information about social and environmental consequences • 5.6 information about emotional consequences • 9.2 pros and cons Emotions • 11.2 advise on ways to reduce negative emotions to facilitate performance of the behavior
Discomfort with long-term use of inhalers	• Beliefs about consequences and emotions	• 5.1 information about health consequences • 5.2 salience of consequences • 5.6 information about emotional consequences • 9.2 pros and cons • 11.2 reduce negative emotions
Prioritizing asthma to support	• Goals	• 1.3 goal setting (outcome)

[a]Behavior change techniques are numbered using the numbering system from the BCT Taxonomy v1.

[b]BCT: behavior change technique.

[c]N/A: not applicable.

Discussion

Principal Findings

We have used a systematic approach, grounded in behavioral science, to create a shortlist of BCTs that app developers can use to target particular influences on preventer inhaler adherence in people with asthma. We identified 24 BCTs that might be effective, practical, and acceptable to deliver via an asthma medication adherence app to target specific influences on adherence. In contrast, 1 systematic review of existing adherence apps for patients with asthma identified only 9 BCTs in the included apps [16] while the review of 23 asthma apps by Ramsey et al [17] found a mean number of 4 BCTs per app. However, these BCTs were not linked to specific influences on nonadherence in any logical way. In similar ways, a more recent review of asthma mHealth apps [18] did not consider how appropriate the included BCTs were for altering the influences on asthma self-management targeted by the apps. In contrast, the behavioral influence-targeting approach used in this study can be used more broadly in designing apps and other mHealth interventions targeting both medication adherence and other forms of behavior change, ensuring intervention components are appropriate for the factors influencing the relevant behavior.

There are several strengths to our approach. First, we started with a comprehensive set of the influences on preventer inhaler use based on previous research. Second, we used a tool incorporating the findings of a large expert consensus study to identify BCTs potentially effective at changing specific influences. This increases the chances that an app will be effective at changing the influences on adherence behavior. Our approach, nevertheless, has some limitations. While the set of influences was based on a synthesis of previous research findings, there are a number of limitations to that synthesis. First, time and resource considerations led to the use of rapid review methods and meant that many of the analytic decisions involved only 2 researchers. Second, much of the available evidence came from qualitative studies, which meant it was not possible to identify which of the 31 influences might have the greatest impact on adherence. Third, the only gray literature included was that readily available to Asthma UK. A more extensive search for unpublished literature could have augmented the synthesis results. Finally, the strongest evidence for the impact of a potential influence on adherence comes from an experimental study which shows that changing the levels of the influence (eg, reducing concerns about side effects) alters subsequent adherence. None of the studies included in the research synthesis had this design. Future research in this area should concentrate on generating such evidence.

Ideally, the choice of a BCT to change a particular MoA would have been based on evidence showing that the BCT changed the MoA, which in turn enhanced medication adherence. However, such evidence does not yet exist for every BCT-MoA-adherence combination that app designers might be interested in, hence the present need to rely on expert consensus. Future app evaluation studies can contribute to this evidence base by using experimental designs and assessing changes in the relevant MoAs as well as adherence behavior. Behavioral scientists will need to pursue additional work to develop reliable and valid measures of key MoAs that can be used in such evaluation studies.

We assumed that the app needed to be easy to use with very little input from health care professionals. We may have reached different conclusions about which BCTs were practicable if we had anticipated that the app was going to be used as one component of an adherence intervention that also involved dedicated sessions with a health care provider. The judgments

XSL•FO
RenderX

of the potential practicability and acceptability of the different BCTs will always involve a subjective element [25,26]. However, they were based on the team's experience with behavior change interventions to promote medication adherence and digital technologies. Nevertheless, this approach could have been strengthened further by including the views of people with lived experience of asthma and adherence difficulties.

For app developers, there are some recommendations that arise from our findings. First, since the reasons for nonadherence differ across individuals, we would recommend that an app should start by using a relatively simple assessment of the individual user's reasons for their low inhaler use. This could be in the form of a simple questionnaire [27] or a list of common reasons (eg, forgetting, low belief in need for daily inhaler use, concerns regarding steroids). This could then be linked to an algorithm which would indicate which BCT would be best to target that reason, as per Table 2. For example, if the user indicates that they are not aware that they need to use their preventer inhaler every day (influence 1 in Table 2), then the recommended BCT would be for the app to provide information incorporating the likely health consequences, such as worsening symptoms or possible hospital emergency visits. Alternatively, if the initial screening reveals that the user does not have a regular habit or pattern of daily inhaler use (row 6 of Table 2), then an "action planning" BCT is indicated as an app component. This would require the user to specify a consistent time and place when the inhaler will be used, linked to an existing daily habit such as toothbrushing. If a user indicates a number of reasons for their nonadherence, then the app will need to ask them to prioritize these in order of importance so that the targeted BCTs can be determined and delivered accordingly. For the more practical issue of how to operationalize the BCT within the app, the final column of table S3 provides details of what each BCT entails, as per its definition in the BCT Taxonomy [7]. App developers should clearly state which BCTs their app includes to tackle which influences in its documentation, to allow apps to be appropriately classified in any future repositories of medication adherence technologies [28]. Health care decision-makers concerned with whether to endorse particular asthma adherence apps for patient use in a given health care system can also make use of our findings. For example, they might appraise the extent to which each candidate app tackles the range of influences on adherence, and whether the app applies appropriate BCTs to alter those influences, as per Table 2.

In addition to selecting the most appropriate BCT for targeting a particular cause of nonadherence, apps need to take account of the phase of an individual's inhaler use from early initiation to implementation (regular use as part of a daily routine) and persistence (continued inhaler use even in the absence of symptoms) [29]. Since different influences may be more important at each phase, this could be a part of the initial assessment component of the app. For example, if the app user indicates an initial antipathy toward using an inhaler possibly combined with a preference for "natural" remedies (influence 9 in Table 2), this is a likely cause of noninitiation of treatment, and so the recommended BCTs would be a combination of information about the health consequences of not starting treatment provided by a credible source (eg, expert patient and/or doctor/pharmacist). However, if the individual makes a good start with their inhaler but then decides to stop regular use due to both feeling better and concerns about possible harms from the inhaler (row 29 in Table 2), then the most relevant BCTs would involve a combination of information about the salient consequences of stopping treatment together with ways of enabling them to weigh up the pros and cons of regular inhaler use.

While we have identified a set of BCTs that may be useful to include, many creative decisions remain for app designers about how to implement each BCT. The impact of BCTs on medication adherence is likely to depend on their style of delivery, ie, the manner in which intervention content is communicated [30,31]. Because people are motivated by their needs to feel competent, autonomous, and related to others [32], intervention content that is delivered in a nonjudgmental, collaborative fashion that supports patient autonomy may lead to greater behavior change than the same intervention content delivered in a moralizing, didactic, or controlling style. App designers will also want to consider incorporating features that promote app engagement [33]. Greater daily engagement with a medication adherence app was found to predict greater objectively measured medication adherence in people with asthma and chronic obstructive pulmonary disease in 1 study [34]. Obtaining feedback from people with lived experience of asthma and adherence difficulties will be essential to ensure that the implementation of BCTs within a prototype app is acceptable to and engaging for potential users.

Conclusions

This paper has used a systematic process to identify the optimal BCTs to tackle specific influences on preventer inhaler nonadherence via an app-based intervention. Its findings are useful for app designers and others interested in digital health interventions to promote medication adherence.

Acknowledgments

The open access service fee for publication was funded by Phillips Medisize. One of the authors (IS) is an employee of Phillips Medisize.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Supplementary material including list of selected behavior change techniques.
[DOCX File, 71 KB - mhealth_v13i1e49348_app1.docx]

References

1. Adherence to long-term therapies: evidence for action. : World Health Organization; 2003.
2. Khan R, Socha-Dietrich K. Investing in medication adherence improves health outcomes and health system efficiency: adherence to medicines for diabetes, hypertension, and hyperlipidaemia. OECD Health Working Paper 2018. [doi: 10.1787/8178962c-en]
3. Nieuwlaat R, Wilczynski N, Navarro T, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20;2014(11):CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
4. Conn VS, Ruppar TM, Enriquez M, Cooper P. Medication adherence interventions that target subjects with adherence problems: systematic review and meta-analysis. Res Social Adm Pharm 2016;12(2):218-246. [doi: 10.1016/j.sapharm.2015.06.001] [Medline: 26164400]
5. Wilhelmsen NC, Eriksson T. Medication adherence interventions and outcomes: an overview of systematic reviews. Eur J Hosp Pharm 2019 Jul;26(4):187-192. [doi: 10.1136/ejhpharm-2018-001725] [Medline: 31338165]
6. Armitage LC, Kassavou A, Sutton S. Do mobile device apps designed to support medication adherence demonstrate efficacy? A systematic review of randomised controlled trials, with meta-analysis. BMJ Open 2020 Jan 30;10(1):e032045. [doi: 10.1136/bmjopen-2019-032045] [Medline: 32005778]
7. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
8. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
9. Riley IL, Jackson B, Crabtree D, et al. A scoping review of international barriers to asthma medication adherence mapped to the theoretical domains framework. J Allergy Clin Immunol Pract 2021 Jan;9(1):410-418. [doi: 10.1016/j.jaip.2020.08.021] [Medline: 32861047]
10. Jackson C, Eliasson L, Barber N, Weinman J. Applying COM-B to medication adherence: a suggested framework for research and interventions. Eur Health Psychol 2014;16(1):7-17.
11. Oh J, Kim S, Kim MS, Abate YH, Abd ElHafeez S, Abdelkader A, et al. Global, regional, and national burden of asthma and atopic dermatitis, 1990–2021, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Respir Med 2025;13(5):425-446. [doi: 10.1016/S2213-2600(25)00055-0]
12. Asthma. World Health Organization. 2023 May 4. URL: https://www.who.int/news-room/fact-sheets/detail/asthma [accessed 2023-05-21]
13. Engelkes M, Janssens HM, de Jongste JC, Sturkenboom MCJM, Verhamme KMC. Medication adherence and the risk of severe asthma exacerbations: a systematic review. Eur Respir J 2015 Feb;45(2):396-407. [doi: 10.1183/09031936.00075614] [Medline: 25323234]
14. Dima AL, Hernandez G, Cunillera O, Ferrer M, de Bruin M, ASTRO-LAB group. Asthma inhaler adherence determinants in adults: systematic review of observational data. Eur Respir J 2015 Apr;45(4):994-1018. [doi: 10.1183/09031936.00172114] [Medline: 25504997]
15. Miles C, Arden-Close E, Thomas M, et al. Barriers and facilitators of effective self-management in asthma: systematic review and thematic synthesis of patient and healthcare professional views. NPJ Prim Care Respir Med 2017 Oct 9;27(1):57. [doi: 10.1038/s41533-017-0056-4] [Medline: 28993623]
16. Miller L, Schüz B, Walters J, Walters EH. Mobile technology interventions for asthma self-management: systematic review and meta-analysis. JMIR mHealth uHealth 2017 May 2;5(5):e57. [doi: 10.2196/mhealth.7168] [Medline: 28465281]
17. Ramsey RR, Caromody JK, Voorhees SE, et al. A systematic evaluation of asthma management apps examining behavior change techniques. J Allergy Clin Immunol Pract 2019;7(8):2583-2591. [doi: 10.1016/j.jaip.2019.03.041] [Medline: 30954644]
18. Robinson B, Proimos E, Zou D, Gong E, Oldenburg B, See K. Functionality and quality of asthma mHealth apps and their consistency with international guidelines: structured search and evaluation. JMIR Mhealth Uhealth 2024 Jan 10;12(1):e47295. [doi: 10.2196/47295] [Medline: 38198204]
19. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libraries J 2009 Jun;26(2):91-108. [doi: 10.1111/j.1471-1842.2009.00848.x]
20. Varker T, Forbes D, Dell L, et al. Rapid evidence assessment: increasing the transparency of an emerging methodology. J Eval Clin Pract 2015 Dec;21(6):1199-1204. [doi: 10.1111/jep.12405] [Medline: 26123092]
21. Carey RN, Connell LE, Johnston M, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2018;10. [doi: 10.1093/abm/kay078]

XSL•FO
RenderX

22. Johnston M, Carey RN, Connell Bohlen LE, et al. Development of an online tool for linking behavior change techniques and mechanisms of action based on triangulation of findings from literature synthesis and expert consensus. Transl Behav Med 2021 May 25;11(5):1049-1065. [doi: 10.1093/tbm/ibaa050] [Medline: 32749460]
23. Connell LE, Carey RN, de Bruin M, et al. Links between behavior change techniques and mechanisms of action: an expert consensus study. Ann Behav Med 2019 Jul 17;53(8):708-720. [doi: 10.1093/abm/kay082] [Medline: 30452535]
24. Michie S, Johnston M, Rothman AJ, et al. Developing an evidence-based online method of linking behaviour change techniques and theoretical mechanisms of action: a multiple methods study. Health Serv Deliv Res 2021;9(1):1-168. [doi: 10.3310/hsdr09010]
25. West R, Michie S, Atkins L, Chadwick P, Lorencatto F. Achieving behaviour change: a guide for local government and partners. : Public Health England; 2019 URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/875385/PHEBI_Achieving_Behaviour_Change_Local_Government.pdf [accessed 2025-09-03]
26. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions: Silverback Publishing; 2014.
27. Weinman J, Ali I, Hodgkinson A, Canfield M, Jackson C. Pilot testing of a brief pre-consultation screener for improving the identification and discussion of medication adherence in routine consultations. Patient Prefer Adherence 2019;13:1895-1898. [doi: 10.2147/PPA.S219860] [Medline: 31806938]
28. Nabergoj Makovec U, Goetzinger C, Ribaut J, et al. Developing a medication adherence technologies repository: proposed structure and protocol for an online real-time Delphi study. BMJ Open 2022 Apr;12(4):e059674. [doi: 10.1136/bmjopen-2021-059674]
29. Vrijens B, De Geest S, Hughes DA, et al. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705. [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
30. Wright AJ, Zhang L, Howes E, et al. Specifying how intervention content is communicated: development of a style of delivery ontology. Wellcome Open Res 2023;8:456. [doi: 10.12688/wellcomeopenres.19899.1] [Medline: 39193088]
31. Hagger MS, Hardcastle SJ. Interpersonal style should be included in taxonomies of behavior change techniques. Front Psychol 2014;5:254. [doi: 10.3389/fpsyg.2014.00254] [Medline: 24723904]
32. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
33. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
34. Kaye L, Gondalia R, Thompson A, Stempel DA, Barrett MA. The relationship between objective app engagement and medication adherence in asthma and COPD: a retrospective analysis. Sci Rep 2021 Dec 21;11(1):24343. [doi: 10.1038/s41598-021-03827-2] [Medline: 34934164]

Abbreviations

APEASE: Acceptability, Practicability, Effectiveness and cost-effectiveness, Affordability, Side effects/safety, and Equity
BCT: behavior change technique
MoA: mechanism of action

Edited by L Buis; submitted 13.06.23; peer-reviewed by AL Dima, C Duncan, L Herrle; revised version received 05.01.25; accepted 13.02.25; published 23.09.25.

Please cite as:
Wright AJ, Holland J, Simpson I, Walker S, Bennett-Steele N, Weinman J
Selection of Behavior Change Techniques for Asthma Medication Adherence Apps: Evidence-Based Design Study
JMIR Mhealth Uhealth 2025;13:e49348
URL: https://mhealth.jmir.org/2025/1/e49348
doi:10.2196/49348

© Alison J Wright, Jeremy Holland, Iain Simpson, Samantha Walker, Naomi Bennett-Steele, John Weinman. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of a Digital Peer-Supported App Intervention in Promoting Smoking Cessations: Nonrandomized Controlled Trial

Shota Yoshihara[1,2], MSc, OTR; Kayoko Takahashi[1,3], ScD, OTR; Chiaki Uemura[2], RD; Shin Murakami[2], PhD; Daichi Harada[2], BA; Hiroshi Yamato[4], PhD, MD

[1]Department of Rehabilitation Sciences, Kitasato University Graduate School of Medical Sciences, Kitasato 1-15-1, Sagamihara, Kanagawa, Japan

[2]A10 Lab Inc, Chuo-ku, Tokyo, Japan

[3]Department of Occupational Therapy, School of Allied Health Science, Kitasato University, Sagamihara, Kanagawa, Japan

[4]Department of Health Development, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Japan, Kitakyushu, Fukuoka, Japan

Corresponding Author:
Shota Yoshihara, MSc, OTR
Department of Rehabilitation Sciences, Kitasato University Graduate School of Medical Sciences, Kitasato 1-15-1, Sagamihara, Kanagawa, Japan

Abstract

Background: Smoking cessation has become a global priority, with peer support interventions shown to improve abstinence rates. However, no studies have examined the effectiveness of a group-based digital peer-supported app combined with nicotine gum for smoking cessation among working populations.

Objective: This study aimed to assess whether adding a digital peer-supported app to standard nicotine gums improves 12-week smoking abstinence rates among current working smokers in employment-based settings.

Methods: A nonrandomized comparison trial was conducted with current working smokers in Japan. Eligible participants smoked at least 1 cigarette per day, owned a smartphone (iOS or Android), and were enrolled in their company's health insurance program. Participants were self-selected into one of the two intervention groups (digital peer-supported app + nicotine gums) or a control group (nicotine gums only). The digital peer-supported app creates a group chat for up to 5 people aimed at smoking cessation, where participants can anonymously post counts, photos, and comments daily. Logistic regression analyses adjusted for demographic and smoking-related variables were used to estimate the odds ratios for smoking cessation. Engagement with the app (usage days and posting frequency) was analyzed within the intervention groups.

Results: A total of 451 participants were included in the per-protocol analysis (260 in the intervention groups and 191 in the control group). The 12-week abstinence rate was significantly higher in the digital peer-supported app + nicotine gum group compared to the gum-only group (59.2% [154/260] vs 38.7% [74/191]). The adjusted odds ratio of smoking cessation was 2.41 (95% CI 2.07 - 2.81), indicating a significant impact of digital peer support. Both higher duration of digital peer-supported app usage and increased posting frequency were positively associated with cessation success (*P* for trend <.001).

Conclusions: The addition of a digital peer-supported app to nicotine gum use significantly improved smoking cessation outcomes among working smokers. These findings provide preliminary evidence for the feasibility and effectiveness of integrating group-based digital peer support into smoking cessation interventions.

(JMIR Mhealth Uhealth 2025;13:e68638) doi:10.2196/68638

KEYWORDS

smoking; smoking cessation; digital therapeutics; peer support; digital peer support app; mHealth; mobile health; apps

Introduction

Smoking is the leading cause of noncommunicable disease mortality in Japan, significantly increasing the risk of ischemic heart disease, stroke, chronic obstructive pulmonary disease, and cancer [1-3]. The societal burden of smoking has intensified, marking smoking cessation as a global imperative [4].

Various measures to promote smoking cessation have been implemented in Japan, including a raised tobacco price (taxation: 59.9%) [5] and enforcement of the revised Health Promotion Act that prohibits smoking in public facilities other than designated smoking areas [6]. Under the Japanese national insurance system, patients with nicotine dependence are eligible for covered smoking cessation treatment. This program lasts 12 weeks and includes 5 visits with health care professionals, during which participants receive counseling and pharmacotherapy treatments, such as nicotine patches and gums. Despite these efforts, smoking cessation success rates in Japan

XSL•FO
RenderX

remain modest. According to a report by the Ministry of Health, Labour and Welfare, the average completion rate for the standard 5-session smoking cessation program was 34.5%, with hospitals averaging 43.5% and clinics averaging 32.7% [7].

Peer support–based interventions, which have become a common approach for promoting health-related behavior change [8], have been defined as "a method of teaching or facilitating health promotion that makes use of people sharing specific health messages with members of their own community" [9]. A meta-analysis suggests that interventions involving peer support may significantly increase smoking abstinence rates [10]. Additionally, digital therapies, including smartphone apps, have shown promise. Systematic reviews have confirmed the effectiveness of smoking cessation apps [11,12]. A digital peer-supported app, which facilitates group interactions to motivate users toward quitting smoking, could be particularly beneficial [13-15]. However, no studies have specifically evaluated the effectiveness of this digital peer-supported app for smoking cessation. Investigating the potential benefits of integrating the digital peer-supported app with behavioral support could provide valuable insights.

This study aims to investigate whether adding a digital peer-supported app to nicotine gums can improve the 12-week smoking abstinence rate among current working smokers.

Methods

Study Design

A web-based questionnaire was administered to employees from 3 companies (electronics, insurance, and telecommunications industries) affiliated with a health insurance association from October 2022 to February 2023. This nonrandomized comparison trial assessed the effectiveness of a digital peer-supported app in promoting smoking cessation among current working smokers. Smoking cessation success rates were then compared, with the group using only the nicotine gums serving as the reference.

Ethical Considerations

This study received ethical approval from the Research Ethics Committee of the Institutional Review Board of the School of Allied Health Sciences at Kitasato University (approval number: 2024 - 008). At the time of enrollment, participants were also informed that data obtained through their involvement in the program might be used for research purposes in an anonymized form, and participation was contingent upon their agreement to these terms. An opt-out method was used to obtain informed consent, with study details publicly disclosed on the web. Participants could withdraw at any time. No financial reward was offered. All data were anonymized to ensure privacy.

Participants

This study enrolled employed individuals who were current smokers. Eligibility criteria included (1) ownership of a smartphone compatible with iOS or Android operating systems, (2) enrollment in their company's corporate health insurance program, and (3) self-reported daily smoking of at least 1 cigarette.

Participants self-selected into one of the two intervention groups: (1) nicotine gum only (Nicorette 2 mg) or (2) a combination of the digital peer-supported app and nicotine gum. Participation in the app-based intervention was optional.

Smoking Cessation Program

The smoking cessation program was conducted remotely and lasted 12 weeks (Figure 1). Participant recruitment took place over approximately 20 days, beginning about 1 month before the program's start. Recruitment strategies included both individual outreach and workplace-based promotion. Potential participants were primarily invited via email communications sent by their employers and affiliated health insurance societies. In addition, promotional materials—such as posters and triangular pop-up displays—were distributed and displayed at the participants' workplaces to raise awareness of the program. Interested individuals voluntarily enrolled in the program by independently submitting an application form, typically via a web-based platform.

XSL•FO
RenderX

Figure 1. Timeline of study procedures.

A preprogram online session was held to explain the program's structure and educational content, which was designed to assist participants in quitting smoking. This content included discussions on withdrawal symptoms and how to use the digital peer-supported app effectively. Educational videos featured an ex-smoking physician discussing the health impacts of smoking, the benefits of quitting, coping strategies for withdrawal, methods for creating a smoke-free environment, and planning alternative behaviors.

For those unable to attend the live session, an e-learning course was available. The program also offered participants nicotine gums (96 pieces per box) without a prescription. An instructional video on how to use the smoking cessation aids was provided to all participants before the challenge began.

The Digital Peer-Supported App Intervention

The digital peer-supported app "Minchalle" is a smartphone app developed by A10 Lab Inc, which was initially released in November 2015 and is available for Android and iPhone [13-15]. This app was originally designed to help users build desirable habits, such as exercising, dieting, and learning English conversation, and has been downloaded over 1.6 million times. For this study, the app was adapted to support smoking cessation.

In this smoking cessation program, participants are enrolled by registering and installing the app. They were automatically assigned to teams of up to 5 individuals based on the order of registration. Team members interacted anonymously using nicknames and engaged in group chats, where they could post photos, messages, and stamps to support one other (Figure 2). To ensure privacy, the app was designed to prevent the sharing of personally identifiable information, and participants were instructed not to include such information in their posts.

XSL•FO
RenderX

Figure 2. Examples of app screens. (A) Select a group of 5 members. (B) Upload a photo of the day with a comment related to abstinence. (C) Read responses and applause from the other 4 members of the group. (D) Look back at the records of abstinence by photos.

Two types of photo-posting features were available: Challenge Photo and Normal Photo. Challenge Photo allowed users to post 1 photo with a comment per day to report their progress in quitting smoking. These were recorded in a personal calendar view (Figure 2), supporting self-monitoring and reflection. Normal Photo allowed users to post additional photos without comments, typically after submitting a Challenge Photo for the day, enabling more spontaneous sharing within the team.

Additional key features included (1) peer approval of posts, (2) automated feedback based on team engagement, and (3) automatic removal from the group if a participant exceeded a specified inactivity period of 15 consecutive days. Although participants can post comments or photos more than once a day and interact with other 4 members, such engagement is not mandatory. The digital peer-supported app is available at no cost to the participants. For visual representation, examples of the user interface can be found in Multimedia Appendix 1.

Other Interventions

Throughout the smoking cessation program, participants received supplementary support via a chat platform. This support included guidance on managing withdrawal symptoms in daily life, quizzes designed to educate participants about the adverse health effects of smoking, and information on the physiological changes associated with quitting smoking. Additionally, a consultation service was available, allowing participants to receive personalized responses from medical professionals. Participants could submit individual concerns, such as withdrawal symptoms, proper use of cessation aids, or other issues encountered during the program, through a web-based consultation form. Inquiries were also accepted via email. Medical professionals responded to each consultation individually, typically within 1 week of submission.

Measurements

Baseline Characteristics

Baseline demographic variables, including age and sex, and smoking-related characteristics—such as motivations for quitting, smoking duration, and daily cigarette consumption—were obtained through a web-based survey. These variables were originally collected using predefined categories in the survey instrument. Age was grouped as 20 - 29, 30 - 39, 40 - 49, 50 - 59, and 60 years or older; sex was categorized as male or female. Smoking duration was classified as less than 5, 5 - 9, 10 - 19, 20 - 29, and 30 years or more. Daily cigarette consumption was recorded in categories: 1 - 9, 10 - 19, and 20 or more cigarettes per day. The purpose of smoking cessation was assessed using a multiple-choice format (eg, my health, saving for money, family's health, none, or others). Motivation for smoking cessation was measured on a scale of 1 to 10, with higher scores indicating stronger motivation to quit smoking.

Log Data of the Digital Peer-Supported App

The smoking cessation program was conducted remotely over a 12-week period. Participants were recruited over a 20-day period beginning approximately 1 month prior to the program's start, and they had access to the digital peer-supported app for the full duration of the intervention.

To evaluate participants' engagement with the app, we collected log data automatically recorded by the system. Two behavioral indicators were analyzed: (1) continuous app usage and (2) posting frequency during the 12-week program. Continuous app usage was measured by counting the number of days on which participants logged into the app. Posting frequency was defined as the total number of posts (including messages,

XSL•FO
RenderX

challenge posts, normal photos, and stamps) made by each participant throughout the program.

For both indicators, participants were classified into 3 groups using tertile cutoffs: low, moderate, and high. These categories were used in the analysis to explore the association between different levels of engagement and smoking cessation outcomes.

Outcome

In this program, continuous smoking cessation is defined as self-reported abstinence from smoking for a minimum of 4 weeks. Individuals are classified as "quitters" if they report having abstained from smoking entirely, without consuming a single cigarette, for at least 4 weeks by the end of the program. This information is collected via a web-based form on the program's final day, with cessation determined by the "last smoking date" provided by participants.

Statistical Analysis

A logistic regression analysis was conducted to estimate the odds of smoking cessation. The outcome was treated as a binary variable (1=quit, 0=not quit). The nicotine gum–only group served as the reference, and the odds ratio (OR) for quitting was calculated for the group using both the app and nicotine gum.

The analysis included statistical adjustments made in multiple steps: model 1 adjusted for age (categorized as 20 - 29, 30 - 39, 40 - 49, 50 - 59, and 60 years or older), sex (male or female), and cluster (3 study sites). Model 2 incorporated additional adjustments for the duration of smoking (categorized as 0 - 4, 5 - 9, 10 - 19, 20 - 29, and 30 years or more) and daily cigarette consumption (categorized as 1 - 9, 10 - 19, and 20 cigarettes or more). Model 3 was further adjusted for the purpose of smoking cessation (my health, saving for money, family's health, none, or others) and motivation for smoking cessation (treated as a continuous variable). To account for potential clustering effects within the company, we used robust standard errors clustered by company.

Furthermore, logistic regression was used to evaluate the impact of digital peer-supported app usage duration and posting frequency on smoking cessation success, specifically among participants who used the app. The "low" tertile group was used as the reference category. To assess linear trends across tertiles, *P* values for trend were calculated using orthogonal polynomial contrasts (specifically, the "contrast" command in Stata), with the same covariate adjustments as in model 3.

All statistical analyses were conducted using Stata, version 17.0 (Stata Corp), with statistical significance defined as a 2-tailed *P* value of <0.05. The OR and the corresponding 95% CI were calculated to assess the strength of associations between variables.

Results

Participants

Initially, 542 participants enrolled in the smoking cessation program. Among them, 90 (16.6%) participants did not use the nicotine gum at all during the program, and 1 participant had missing age data. These individuals were excluded from the analysis.

As a result, the final per-protocol analysis set included 451 participants: 260 participants used both the digital peer-supported app and nicotine gums, while 191 used only nicotine gums. Most participants were male (434/451, 96.2%), 61.6% (278/451) were over the age of 50 years, and 68.7% (310/451) reported a smoking history of 20 years or more.

Participant Characteristics

Table 1 presents the baseline characteristics of the participants. Compared to the group using only the nicotine gums, participants using both the digital peer-supported app and the nicotine gums had a higher average age, a longer history of smoking, and were more likely to cite "family health" as their primary motivation for quitting.

XSL•FO
RenderX

Table . Participants' characteristics at baseline.

	Total (n=451)	Nicotine gums–only group (n=191)	Digital peer-supported app and nicotine gums (n=260)
Sex, n (%)			
Male	434 (96.2)	190 (99.5)	244 (93.8)
Female	17 (3.8)	1 (0.5)	16 (6.2)
Age (years), n (%)			
20 - 29	6 (1.3)	4 (2.1)	2 (0.8)
30 - 39	70 (15.5)	39 (20.4)	31 (11.9)
40 - 49	97 (21.5)	45 (23.6)	52 (20)
50 - 59	171 (37.9)	68 (35.6)	103 (39.6)
60＋	107 (23.7)	35 (18.3)	72 (27.7)
Duration of smoking (years), n (%)			
0 - 4	14 (3.1)	8 (4.2)	6 (2.3)
5 - 9	37 (8.2)	20 (10.5)	17 (6.5)
10 - 19	90 (20)	44 (23)	46 (17.7)
20 - 29	167 (37)	65 (34)	102 (39.2)
30＋	143 (31.7)	54 (28.3)	89 (34.2)
Cigarettes/day, n (%)			
1 - 9	72 (16)	29 (15.2)	43 (16.5)
10 - 19	299 (66.3)	126 (66)	173 (66.5)
20＋	80 (17.7)	36 (18.8)	44 (16.9)
Purpose of smoking cessation, n (%)			
My health	250 (55.4)	113 (59.2)	137 (52.7)
Saving for money	129 (28.6)	51 (26.7)	78 (30)
Family's health	46 (10.2)	12 (6.3)	34 (13.1)
None	14 (3.1)	8 (4.2)	6 (2.3)
Others	12 (2.7)	7 (3.7)	5 (1.9)
Motivation for smoking cessation, mean (SD)	7.9 (2.1)	7.7 (2.2)	8.1 (2)

Digital Intervention and Smoking Abstinence Rates

Logistic regression analyses demonstrated that participants using the digital peer-supported app in combination with nicotine gums had significantly higher smoking abstinence rates at 12 weeks compared to those using only the nicotine gums (59.2% [154/260] vs 38.7% [74/191]), as shown in Table 2. Model 1 yielded an OR of 2.40 (95% CI 2.02 - 2.85), model 2 reported an OR of 2.41 (95% CI 2 - 2.9), and model 3 showed an OR of 2.41 (95% CI 2.07 - 2.81), indicating a statistically significant improvement associated with the digital intervention.

XSL•FO
RenderX

Table . The results of the logistic regression analyses on smoking cessation success ratios.

	Nicotine gums–only group	App and nicotine gums group
Participants, n	191	260
Prevalent cases, n (%)	74 (38.7)	154 (59.2)
Model 1,[a,b] OR[c] (95% CI)	1.00 (reference)	2.40 (2.02 - 2.85)
Model 2,[a,d] OR (95% CI)	1.00 (reference)	2.41 (2.00 - 2.90)
Model 3,[a,e] OR (95% CI)	1.00 (reference)	2.41 (2.07 - 2.81)

[a]In all models, companies were incorporated as clusters.

[b]Adjusted for age (in years, continuous) and sex (male or female).

[c]OR: odds ratio.

[d]Adjusted for age (in years, continuous), sex (male or female), duration of smoking (0 - 4, 5 - 9, 10 - 19, 20 - 29, or 30 years or more), and the number of cigarettes smoked per day (1 - 9, 10 - 19, or 20 cigarettes or more).

[e]Adjusted for age (in years, continuous), sex (male or female), duration of smoking (0 - 4, 5 - 9, 10 - 19, 20 - 29, or 30 years or more), the number of cigarettes smoked per day (1 - 9, 10 - 19, or 20 cigarettes or more), purpose of smoking cessation (my health, saving for money, family's health, none, or others), and motivation of smoking cessation (continues).

Smoking Cessation Success and Digital Peer-Supported App Engagement

The relationship between smoking cessation success rates and both the duration of digital peer-supported app usage and posting frequency, restricted to participants in the digital peer-supported app group, is presented in Tables 3 and 4. The analysis restricted to participants using the digital peer-supported app showed that those in the moderate and high tertiles for digital peer-supported app usage duration and posting frequency had significantly higher smoking cessation success rates compared to those in the low tertile (*P* for trend <.001).

Table . Association between the duration of app usage and smoking cessation success ratios among users of digital peer-supported app and nicotine gums.

	The duration of app usage			
	Low (0 - 62 days)	Moderate (63 - 106 days)	High (107 - 136 days)	*P* for trend
Participants, n	95	85	80	
Prevalent cases, n (%)	43 (45.3)	54 (63.5)	57 (71.3)	
Model 3,[a] OR[b] (95% CI)	1.00 (reference)	1.39 (1.31 - 1.48)	3.43 (3.11 - 3.79)	<.001

[a]Adjusted for age (in years, continuous), sex (male or female), duration of smoking (0 - 4, 5 - 9, 10 - 19, 20 - 29, or 30 years or more), the number of cigarettes smoked per day (1 - 9, 10 - 19, or 20 cigarettes or more), purpose of smoking cessation (my health, saving for money, family's health, none, or others), and motivation of smoking cessation (continues). In model 3, companies were incorporated as clusters.

[b]OR: odds ratio.

Table . Association between posting frequency and smoking cessation success ratios among users of a digital peer-supported app and nicotine gums.

	Posting frequency			
	Low (0 - 5 times)	Moderate (6 - 24 times)	High (25 - 449 times)	*P* for trend
Participants, n	90	84	86	
Prevalent cases, n (%)	39 (43.3)	48 (57.1)	67 (77.9)	
Model 3,[a] OR[b] (95% CI)	1.00 (reference)	1.16 (0.88 - 1.53)	3.67 (3.36 - 4.02)	<.001

[a]Adjusted for age (in years, continuous), sex (male or female), duration of smoking (0 - 4, 5 - 9, 10 - 19, 20 - 29, or 30 years or more), the number of cigarettes smoked per day (1 - 9, 10 - 19, or 20 cigarettes or more), purpose of smoking cessation (my health, saving for money, family's health, none, or others), and motivation of smoking cessation (continues). In model 3, companies were incorporated as clusters.

[b]OR: odds ratio.

No adverse events directly attributable to the digital peer-supported app were reported.

Discussion

Principal Results

This study assessed the effectiveness of a digital peer-supported app for smoking cessation within an online group setting, marking a novel therapeutic approach. We observed a significant

XSL•FO
RenderX

increase in smoking abstinence rates among patients with nicotine dependence who used the digital peer-supported app alongside standard cessation products. To our knowledge, this is the first study to evaluate the efficacy of a group-based, digital peer-supported app in improving smoking abstinence rates. This also highlights the feasibility of incorporating such digital tools in smoking cessation interventions and contributes preliminary evidence to support the development of digital health solutions.

Comparison With Previous Studies

Our findings are notably significant. The primary outcome—self-reported smoking abstinence at week 12—was substantially high in the group using both the digital peer-supported app and nicotine gums (154/260, 59.2%), which was higher compared to the group using only the nicotine gums (74/191, 38.7%). Our results coincide with previous research on digital smoking cessation programs using nongroup-based, single-user apps, which typically show success rates about 5% to 15% higher [16-19]. Our findings suggest that a group-type digital peer-supported app, when combined with nicotine gum, may enhance short-term smoking cessation outcomes compared to nicotine gum alone. However, the results should be interpreted with caution due to the reliance on self-reported data and the lack of long-term follow-up.

Our study suggests that a digital peer-supported app, informed by social cognitive theory [20], can significantly aid smoking cessation. To further substantiate these findings, we focused on participants who actively engaged with the digital peer-supported app, examining the relationship between smoking cessation success rates, duration of digital peer-supported app use, and posting frequency. The analysis indicated that the duration of use or posting frequency was significantly greater in the groups with moderate and high cessation success rates compared to the groups with lower success rates. A plausible explanation for this is that individuals who engaged with the digital peer-supported app more extensively and frequently posted updates had increased opportunities to share their success stories, receive recognition, and obtain positive feedback through the chat feature. This likely enhanced their motivation and self-efficacy [21], a phenomenon similar to what has been observed in traditional face-to-face peer support studies [22].

The digital peer-supported app is designed to integrate seamlessly into the daily communication environments of smokers, offering support that is more accessible than traditional methods such as group sessions, telephone calls, or websites. The digital peer-supported app accessibility has likely contributed to increased engagement with quitting efforts [23,24]. In addition, while the app did not have any dedicated features for tracking or managing nicotine gum use, the digital peer support it facilitated may have indirectly encouraged better adherence. This enhanced adherence is one possible factor that might have contributed to the higher smoking cessation success rates observed in the intervention group. We posit that the digital peer-supported app may have mirrored traditional peer-support dynamics, where individuals are known to quit smoking in interconnected groups [25].

Limitations and Future Study

This study has several limitations. First, smoking cessation outcomes were assessed through self-reported data without biochemical validation (eg, carbon monoxide or cotinine testing) [26], which may overestimate true smoking cessation rates. Second, the nonrandomized design and participant self-selection into intervention groups may have introduced bias, thereby limiting the ability to draw causal inferences. Third, the study targeted individuals with smartphone access, potentially introducing selection bias, as more motivated or tech-savvy individuals may have been more likely to enroll. Additionally, the inclusion criterion of smoking at least 1 cigarette per day may limit generalizability to heavy smokers. Fourth, the study did not assess long-term cessation beyond the 12-week intervention, leaving the sustainability of effects unclear [27,28]. Fifth, while participants used anonymous nicknames, some may have known each other due to recruitment from the same or affiliated companies, potentially influencing team dynamics. Sixth, although a consultation service with medical professionals was available throughout the program, its potential impact on smoking cessation could not be evaluated. Only 13 consultation requests were submitted during the intervention period, and all entries were anonymized, making it impossible to identify which participants used the service or to assess its association with cessation outcomes. This introduces a possible confounding factor that could not be statistically controlled. Seventh, nicotine dependence was not formally assessed using standardized assessments (eg, Fagerström test) [29-31], limiting our understanding of baseline addiction severity [32]. Eighth, the use of heated tobacco products and e-cigarettes was not assessed, which may have influenced smoking behaviors and cessation outcomes [33,34]. Ninth, although some participants mentioned withdrawal-like symptoms in group chats, no standardized assessment of withdrawal symptoms was conducted. Nicotine gum usage was self-managed and not tracked; all participants received 2 mg gum (96 pieces) intended for the 12-week program, with no additional supply provided, and no choice of higher dose. The uniform dosing may have been insufficient for heavier smokers. Future studies should consider tailoring nicotine replacement therapy based on individual smoking behaviors. Tenth, no formal sample size calculation was performed due to the lack of previous studies on this specific digital peer support smoking cessation program. Finally, the participants in this study consisted primarily of male employees from large Japanese companies, limiting generalizability to other populations, including women and nonworking individuals. Future studies should explore the sustainability of these outcomes and assess whether prolonged digital peer-supported app engagement or follow-up support is necessary to maintain smoking cessation.

Conclusions

Our findings suggest that using a digital peer-supported app may boost smoking abstinence rates over 12 weeks in current working smokers, compared to using standard nicotine gums alone.

XSL•FO
RenderX

Acknowledgments

This study was conducted with the financial support of A10 Lab Inc., the developer and operator of the habit-formation application. Grammar in portions of this manuscript was checked using ChatGPT [35] and Grammarly [36]. The final content was reviewed and approved by all authors, who take full responsibility for its accuracy and integrity.

Data Availability

The datasets generated and analyzed in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: SY, KT
Methodology: SY
Formal analysis: SY
Investigation: CU, SM, DH
Writing – original draft: SY
Writing – review & editing: SY, KT, CU, SM, DH, HY
Supervision: HY

Conflicts of Interest

SY, CU, SM, and DH were employees of A10 Lab Inc at the time of the research. We paid HY 50,000 JPY as the program supervision fee.

Multimedia Appendix 1
Examples of the user interface.
[DOCX File, 60270 KB - mhealth_v13i1e68638_app1.docx]

References

1. Ikeda N, Inoue M, Iso H, et al. Adult mortality attributable to preventable risk factors for non-communicable diseases and injuries in Japan: a comparative risk assessment. PLoS Med 2012 Jan;9(1):e1001160. [doi: 10.1371/journal.pmed.1001160] [Medline: 22291576]
2. Ikeda N, Saito E, Kondo N, et al. What has made the population of Japan healthy? Lancet 2011 Sep 17;378(9796):1094-1105. [doi: 10.1016/S0140-6736(11)61055-6] [Medline: 21885105]
3. Okui T. An age-period-cohort analysis of the difference in smoking prevalence between urban and non-urban areas in Japan (2004-2019). Epidemiol Health 2020;42:e2020072. [doi: 10.4178/epih.e2020072] [Medline: 33285059]
4. GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. The Lancet 2017 May;389(10082):1885-1906. [doi: 10.1016/S0140-6736(17)30819-X] [Medline: 28390697]
5. WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke. World Health Organization. 2023 Jul 31. URL: https://www.who.int/publications/i/item/9789240077164 [accessed 2025-07-26]
6. Ueda H, Armada F, Kashiwabara M, Yoshimi I. Street smoking bans in Japan: a hope for smoke-free cities? Health Policy 2011 Sep;102(1):49-55. [doi: 10.1016/j.healthpol.2011.05.013] [Medline: 21684620]
7. Special survey on the effectiveness of smoking cessation treatment under the nicotine dependence management fee: results from the 2017 survey related to the FY2016 revision of medical service fees [Article in Japanese]. : Central Social Insurance Medical Council, Ministry of Health, Labour and Welfare of Japan; 2018 URL: https://www.mhlw.go.jp/file/05-Shingikai-12404000-Hokenkyoku-Iryouka/0000192293.pdf [accessed 2025-07-31]
8. Posavac EJ, Kattapong KR, Dew DEJ. Peer-based interventions to influence health-related behaviors and attitudes: a meta-analysis. Psychol Rep 1999 Dec;85(3 Pt 2):1179-1194. [doi: 10.2466/pr0.1999.85.3f.1179] [Medline: 10710973]
9. Webel AR, Okonsky J, Trompeta J, Holzemer WL. A systematic review of the effectiveness of peer-based interventions on health-related behaviors in adults. Am J Public Health 2010 Feb;100(2):247-253. [doi: 10.2105/AJPH.2008.149419] [Medline: 20019321]
10. Yuan P, Westmaas JL, Thrul J, Toussaert S, Hilton JF, White JS. Effectiveness of peer-support interventions for smoking cessation: a systematic review and meta-analysis. Nicotine Tob Res 2023 Aug 19;25(9):1515-1524. [doi: 10.1093/ntr/ntad059] [Medline: 37042206]
11. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299. [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
12. Ghorai K, Akter S, Khatun F, Ray P. mHealth for smoking cessation programs: a systematic review. J Pers Med 2014 Jul 18;4(3):412-423. [doi: 10.3390/jpm4030412] [Medline: 25563359]

XSL•FO
RenderX

13. Tabira K, Oguma Y, Yoshihara S, et al. Digital peer-supported app intervention to promote physical activity among community-dwelling older adults: nonrandomized controlled trial. JMIR Aging 2024 May 30;7:e56184. [doi: 10.2196/56184] [Medline: 38814686]
14. Okamoto M, Saito Y, Nakamura S, et al. Smartphone-based digital peer support for a walking intervention among public officers in Kanagawa Prefecture: single-arm pre- and postintervention evaluation. JMIR Form Res 2024 Sep 24;8:e53759. [doi: 10.2196/53759] [Medline: 39316793]
15. Yoshihara S, Takahashi K, Kawaguchi H, Harai N, Tsuchiya K. Effects of using a digital peer-supported app on glycated hemoglobin changes among patients with type 2 diabetes: prospective single-arm pilot study. JMIR Form Res 2025 May 20;9:e72659. [doi: 10.2196/72659] [Medline: 40393047]
16. Masaki K, Tateno H, Nomura A, et al. A randomized controlled trial of a smoking cessation smartphone application with a carbon monoxide checker. NPJ Digit Med 2020;3:35. [doi: 10.1038/s41746-020-0243-5] [Medline: 32195370]
17. BinDhim NF, McGeechan K, Trevena L. Smartphone Smoking Cessation Application (SSC App) trial: a multicountry double-blind automated randomised controlled trial of a smoking cessation decision-aid “app”. BMJ Open 2018 Jan 21;8(1):e017105. [doi: 10.1136/bmjopen-2017-017105] [Medline: 29358418]
18. Iacoviello BM, Steinerman JR, Klein DB, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR Mhealth Uhealth 2017;5(4):e56. [doi: 10.2196/mhealth.7226]
19. Bricker JB, Mull KE, Kientz JA, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 1;143:87-94. [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
20. Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE. Effects of eHealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021 Feb 22;23(2):e23786. [doi: 10.2196/23786] [Medline: 33616534]
21. Rajani NB, Mastellos N, Filippidis FT. Self-efficacy and motivation to quit of smokers seeking to quit: quantitative assessment of smoking cessation mobile apps. JMIR Mhealth Uhealth 2021 Apr 30;9(4):e25030. [doi: 10.2196/25030] [Medline: 33929336]
22. Ford P, Clifford A, Gussy K, Gartner C. A systematic review of peer-support programs for smoking cessation in disadvantaged groups. Int J Environ Res Public Health 2013 Oct 28;10(11):5507-5522. [doi: 10.3390/ijerph10115507] [Medline: 24169412]
23. Gerber BS, Stolley MR, Thompson AL, Sharp LK, Fitzgibbon ML. Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. Health Informatics J 2009 Mar;15(1):17-25. [doi: 10.1177/1460458208099865] [Medline: 19218309]
24. Suffoletto B, Callaway C, Kristan J, Kraemer K, Clark DB. Text-message-based drinking assessments and brief interventions for young adults discharged from the emergency department. Alcohol Clin Exp Res 2012 Mar;36(3):552-560. [doi: 10.1111/j.1530-0277.2011.01646.x] [Medline: 22168137]
25. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008 May 22;358(21):2249-2258. [doi: 10.1056/NEJMsa0706154] [Medline: 18499567]
26. Gorber SC, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob Res 2009 Jan;11(1):12-24. [doi: 10.1093/ntr/ntn010]
27. Hollis JF, McAfee TA, Fellows JL, Zbikowski SM, Stark M, Riedlinger K. The effectiveness and cost effectiveness of telephone counselling and the nicotine patch in a state tobacco quitline. Tob Control 2007 Dec;16(Suppl 1):i53-i59. [doi: 10.1136/tc.2006.019794] [Medline: 18048633]
28. Stead LF, Hartmann-Boyce J, Perera R, Lancaster T. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2013 Aug 12(8):CD002850. [doi: 10.1002/14651858.CD002850.pub3] [Medline: 23934971]
29. Borland R, Yong HH, O’Connor RJ, Hyland A, Thompson ME. The reliability and predictive validity of the Heaviness of Smoking Index and its two components: findings from the International Tobacco Control Four Country study. Nicotine Tob Res 2010 Oct;12(Suppl 1):S45-S50. [doi: 10.1093/ntr/ntq038] [Medline: 20889480]
30. Brown J, Michie S, Raupach T, West R. Prevalence and characteristics of smokers interested in internet-based smoking cessation interventions: cross-sectional findings from a national household survey. J Med Internet Res 2013 Mar 18;15(3):e50. [doi: 10.2196/jmir.2342] [Medline: 23506944]
31. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
32. Caponnetto P, Polosa R. Common predictors of smoking cessation in clinical practice. Respir Med 2008 Aug;102(8):1182-1192. [doi: 10.1016/j.rmed.2008.02.017] [Medline: 18586479]
33. Nomura A, Ikeda T, Fujimoto T, et al. Outcomes of a telemedicine smoking cessation programme for heated tobacco product users in Japan: a retrospective cohort study. BMJ Open 2022 Dec 20;12(12):e063489. [doi: 10.1136/bmjopen-2022-063489] [Medline: 36600419]

XSL•FO
RenderX

34. Noda Y, So R, Sonoda M, Tabuchi T, Nomura A. The usefulness of a smartphone app-based smoking cessation program for conventional cigarette users, heated tobacco product users, and dual users: retrospective study. J Med Internet Res 2023 Mar 17;25:e42776. [doi: 10.2196/42776] [Medline: 36930197]
35. GPT-4. OpenAI. 2023 Mar 14. URL: https://openai.com/index/gpt-4-research/ [accessed 2025-05-28]
36. Grammarly. URL: https://app.grammarly.com/ [accessed 2025-05-28]

Abbreviations

OR: odds ratio

Edited by L Buis; submitted 11.11.24; peer-reviewed by A Eisingerich, H Tateno, J Ferron; revised version received 09.07.25; accepted 10.07.25; published 19.08.25.

Please cite as:
Yoshihara S, Takahashi K, Uemura C, Murakami S, Harada D, Yamato H
Effectiveness of a Digital Peer-Supported App Intervention in Promoting Smoking Cessations: Nonrandomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e68638
URL: https://mhealth.jmir.org/2025/1/e68638
doi:10.2196/68638

© Shota Yoshihara, Kayoko Takahashi, Chiaki Uemura, Shin Murakami, Daichi Harada, Hiroshi Yamato. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of Information Length and Implementation Intentions on Adherence to Weight Management Strategies: Experimental Study

Khaleda Ahmadyar, PhD; Joanna Szypula, PhD; Angeliki Bogosian, PhD; Katy Tapper, PhD

School of Health and Medical Sciences, City St George's, University of London, Northampton Square, London, United Kingdom

Corresponding Author:
Katy Tapper, PhD
School of Health and Medical Sciences, City St George's, University of London, Northampton Square, London, United Kingdom

Abstract

Background: Adherence to weight management strategies may be undermined where lengthy strategy explanations limit engagement and understanding, weakening intervention efficacy. By contrast, implementation intentions have been shown to promote adherence across various health behaviors.

Objective: This study aimed to investigate the impact of explanation length and implementation intentions on adherence to brief weight management strategies.

Methods: Participants (N=200) with a BMI above 25 and an interest in losing weight were recruited from a commercial digital weight management service provider. Participants received information about 1 of 4 weight management strategies on a smartphone app in either a brief or detailed format and were asked to plan their use of the strategy with implementation intentions or were given tips on strategy use. Participants received daily prompts over a 2-week period to report whether they used their assigned strategy. Proposed moderators (need for cognition and planning skills) were measured at baseline.

Results: Strategy adherence was greater with brief information (mean 74%, SD 23%) compared with detailed information (mean 69%, SD 23%); however, this small effect size (Cohen *d*=0.24) was not statistically significant (*P*=.13). There was no moderation by need for cognition (*P*=.25). Adherence did not differ significantly between implementation intentions (mean 71%, SD 27%) and tips (mean 72%, SD 21%; *P*=.73); however, there was moderation by planning skills (*P*=.04). As predicted, adherence was greater with implementation intentions compared with tips among those with poorer planning skills.

Conclusions: Shorter explanation length and implementation intentions (in poorer planners) may enhance adherence to brief weight management strategies, and further investigation is required to confirm these effects.

(JMIR Mhealth Uhealth 2025;13:e65260) doi:10.2196/65260

KEYWORDS

weight management; behaviour change; adherence; implementation intention; experimental study; weight management strategy; weight loss; weight; strategy; engagement; understanding; intervention; efficacy; brief weight management; linear regression; mobile phone

Introduction

Successful weight loss and weight loss maintenance require adjustments to food intake and physical activity; the key behaviors targeted by most weight management interventions [1]. These interventions include a variety of behavioral and psychological strategies, ranging from calorie counting [2] to mindful eating [3] and intermittent fasting [4]. While the efficacy of many of these interventions is well-established [5], their impact on weight management remains contingent upon individual adherence. However, evidence shows that adherence is often suboptimal and can be as low as 10%, undermining intervention effectiveness [6]. There is also some evidence that adherence is lower in individuals with lower socioeconomic status [7], which may contribute to health inequalities. Therefore, it is crucial to explore methods by which adherence can be increased to maximize the potential benefits of interventions.

One important factor to consider is the way the information is presented. Communication in health promotion is crucial [8], and effective communication involves tailoring messages for the intended audience [9]. The information-motivation-strategy model [10] states that one of the main reasons individuals do not adhere to behavior change advice is that they are not given adequate, understandable information. One aspect that could influence understanding is the length of the written information provided. Although longer and more detailed information may enhance understanding, it is possible that longer material reduces engagement, which in turn limits understanding and implementation. By contrast, briefer information may enhance engagement, leading to increased understanding and implementation.

XSL•FO
RenderX

This idea is supported by research on attentional processes, which play a critical role in how individuals process information [11]. In recent years, there has been a decline in attention span, attributed in part to increased information overload due to the rapid rise of digital technology [12]. Yeykelis et al [13] found that individuals tend to switch between different types of web-based content as frequently as every 19 seconds, and 75% of all content is typically viewed for less than 1 minute. This type of media multitasking is associated with worse performance on cognitive tasks requiring long-term attention [14]. This trend emphasizes the importance of capturing and maintaining individuals' attention to enhance their engagement with the content of weight management strategies. Given the reduction in attention span, individuals may be more likely to engage with shorter, more focused content, while longer information may lead to cognitive overload and skim-reading or disengagement entirely.

Nevertheless, preference for longer versus shorter information may vary from person to person. For example, the Need for Cognition Scale (NCS) [15] assesses individuals' inclination to engage in and enjoy cognitive activities. Therefore, those with a higher need for cognition may be more willing to engage with lengthier written material and more interested in learning the rationale behind a particular weight management strategy. By contrast, those with a low need for cognition may be put off by lengthier material, preferring information that is more succinct and to the point. This hypothesis has been partially supported by Williams-Piehota et al [16], who found that women who were high in need for cognition were more likely to follow more detailed (compared with more succinct) mammography advice. However, the format made no difference for those who were low in need for cognition.

Another key factor that may influence adherence is difficulty translating intentions into action, that is, the intention-behavior gap [17]. Many individuals may understand and be motivated to use strategy information, yet still fail to implement it. The Rubicon model of action phases [18] posits 2 key phases of goal pursuit: a motivational (predecisional) phase when the individual forms the intention to perform the behavior, and a volitional (postdecisional) phase when the behavior is implemented. The theory suggests that behavior change can be promoted by targeting motivation in the predecisional phase and implementation of the behavior in the postdecisional phase [19]. A powerful strategy for the latter phase is the formation of an implementation intention, which involves the development of a specific plan of action in the form of an if-then statement, for example, "If situation X is encountered, then I will initiate behavior Y" [20]. Systematic reviews have found implementation intentions to be effective in improving general goal attainment [21], and adherence to a range of health behaviors such as healthy eating [22,23], physical activity [24-26], and smoking cessation [27]. The use of implementation intentions in weight loss interventions has also been associated with greater weight loss [28,29]. However, findings are mixed; Benyamini et al [30] and Hayes et al [31] found that implementation intentions resulted in similar weight loss outcomes as simple goal intentions. In addition, Knäuper et al [32] found that the addition of implementation intentions to the National Institutes of Health–developed Diabetes Prevention Program did not result in greater weight loss.

This discrepancy may be due to the different ways implementation intentions are used across studies. A range of variables may moderate the effects of implementation intentions, including the quantity [33] and specificity of the intentions [34]. Individual differences in self-regulation may be another potential moderator; whether implementation intentions are helpful for an individual could be contingent upon their proficiency in planning skills. Allan et al [35] instructed participants to complete a web-based food diary to monitor their snack intake, and half were also instructed to generate an implementation intention to help them achieve this goal. The implementation intention intervention was significantly associated with higher completion rates in poorer planners, but not in skilled planners, suggesting that adherence can be enhanced by tailoring interventions to individuals' planning skill abilities.

In light of the above, this study investigated whether adherence to brief weight management strategies over a 2-week period is influenced by the length of strategy information and the use of implementation intentions. The aims were to (1) explore the effect of information length on adherence, and whether this differs depending on need for cognition, and (2) examine the effect of implementation intentions on adherence, and whether this differs depending on planning skills. Given an absence of previous research, we did not formulate any confirmatory hypotheses related to the first aim. However, for the second aim, we predicted that use of implementation intentions would increase adherence, and that this increase would be larger for those with poorer planning skills.

Methods

Sample Size

A minimum sample size of 128 participants was calculated on G*Power based on a medium effect size, 80% power, and 5% α for a 2×2 ANOVA. To account for participant attrition and exclusions, and the testing of exploratory hypotheses, the target sample size was set at 200 participants.

Participants

Participants were patients recruited from Oviva, a digital commercial weight-management service provider. Individuals may access the service via self-referral or general practitioner referral, depending on the program they would like to enroll in. The cost of the service is covered by the UK National Health Service. Individuals on the 9-month Diabetes Prevention Program, 12-week tier 2 weight management, and 12-week diabetes structured education programs were invited to take part in the study by email. To prevent the study from interfering with program engagement, participants were invited when they had completed the program or were close to completion. Additional eligibility criteria were age 18 years or older, access to a smartphone, a BMI of more than 25, an interest in losing weight or avoiding weight gain, and not on a meal replacement diet.

XSL•FO
RenderX

Ethical Considerations

The study received ethical approval from the City, University of London Psychology Department Research Ethics Committee (ETH2223-2482). The method and analysis strategy were preregistered with the Open Science Framework [36].

Participants provided informed consent (Multimedia Appendix 1) before enrolling in the study. Study data were anonymized to ensure that individual participants could not be identified. Participants received Amazon vouchers worth up to £20 (approximately US $25) based on participation duration, with an extra £5 (approximately US $6.25) for completing an optional qualitative part of the study.

Study Design

The study used a 2×2×4 between-groups experimental design. Participants were randomly assigned to 1 of the 16 groups, which varied by information length (short or long), implementation intentions (present or absent), and strategy content (sensory eating, attending to fullness, eating vegetables first, or increasing physical activity). The dependent variable was the percentage of days participants adhered to their assigned strategy.

The study also examined whether higher NCS scores were associated with a preference for longer information, and whether use of implementation intentions was associated with increased strategy automaticity during the 2-week period. In addition, the study investigated whether briefer, more lay-friendly measures may be adequate substitutes for the longer, standardized measures of need for cognition and planning skills. This was considered important since although these measures may prove useful for increased personalization of interventions, their length may make them too burdensome and impractical for digital interventions, especially where multiple characteristics are being assessed.

Effects of information length on ease of understanding and remembering strategy content, as well as memory of strategy rationale, were also investigated. In addition, differences in adherence to mental and physical strategies were explored. To identify potential differences in adherence between different types of mindful eating strategies, 2 distinct mindful eating strategies were examined (attending to feelings of fullness versus attending to sensory properties of food). In addition, 2 further evidence-based, nonmindfulness weight management strategies (eating vegetables before other food groups and performing physical activity immediately following a meal) were included in the study to identify any differences between strategies that require mental effort as opposed to those that require physical effort. Finally, the research also aimed to gain qualitative insights into participants' views on the weight management strategies and their experiences during the study.

Experimental Manipulation

Strategy Content

Evidence-based written information was provided for 1 of the 4 brief weight management strategies: paying attention to the sensory properties of food while eating [37], paying attention to feelings of fullness while eating [38], eating vegetables or salad before the rest of the meal [39], and doing 5 minutes of physical activity following a meal [40]. The information provided the rationale behind the strategy as well as instructions on how to practice it (refer to Multimedia Appendix 2 for full strategy information).

Information Length

The strategy information was either presented in a short format (approximately 70 - 100 words) with a focus on action (eg, pay attention to the taste and texture of food in your mouth) or a long format (approximately 600 - 700 words) with a focus on outcome (eg, how to slow down your eating) and additional detail on the strategy rationale.

Implementation Intentions

Participants were either presented with planning prompts to help them form implementation intentions (present), or they received tips on strategy use (absent). The planning prompts involved first indicating when they would use the strategy (eg, If I am eating breakfast), followed by how they would use it (eg, then I will keep reminding myself to notice the taste, texture, and temperature of the food). The same content was presented in the tips condition in the form of tips on strategy use (eg, when you are eating a meal or snack, keep reminding yourself to notice the taste, texture, and temperature of the food).

It is important to note that we compared implementation intentions with the type of standard behavioral advice typically used by health care providers, such as Oviva, which consisted of information, followed by a series of "tips." Like implementation intentions, these tips also linked an environmental cue with a specific action. However, unlike implementation intentions, they did not follow an "if-then" format, had slightly less cue specificity, and were received passively by the participant, rather than being actively selected. We were interested in whether this type of "gold standard" implementation intentions would outperform the type of behavioral advice already being used by the health care provider.

Materials

The study was delivered on the Avicenna Research (formerly Ethica Data) smartphone app (Avicenna Research Inc) [41]. A baseline survey, a schedule of 14 daily surveys, and a follow-up survey were triggered upon enrollment. Participants were notified to complete the surveys via phone notifications.

Measures

Baseline Measures

Demographics

Participants indicated their age, gender, ethnicity, and education level.

Height and Weight

Self-report measures were provided in kilograms or pounds and centimeters or feet and inches.

Weight Loss Intentions

Participants responded to "Which of the following best describes you?" with "I'm trying to lose some weight," "I'm not trying

XSL•FO
RenderX

to lose weight but I'm trying to avoid gaining weight," or "I'm not currently trying to lose weight."

Planning Skills

Planning skills were assessed using 10 items from the "goal setting" subscale of the Short-form Self-Regulation Questionnaire (SSRQ) [42]. Items were rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The total SSRQ score was computed by summing the 10 items; higher scores indicated greater planning skills.

Alternative Measure of Planning Skills

Participants responded to "Which of the following best describes you?" with "I'm good at making plans and sticking to them. If I set myself a goal, I'll spend time figuring out exactly how to reach it. If I'm not making good progress toward a goal, I'll go back to my plans and think again;" or "I struggle to make plans and stick to them. I often find myself forgetting to do things I'd planned to do or getting distracted with other things." This 1-item measure was developed by simplifying and consolidating the items on the SSRQ using more accessible, lay-friendly language.

Need for Cognition

Need for cognition was assessed using the short form of the NCS [43], which consisted of 6 items, each rated on a 5-point Likert scale ranging from 1 (extremely uncharacteristic of me) to 5 (extremely characteristic of me). A mean NCS score was computed from the 6 ratings, with higher scores indicating greater need for cognition.

Alternative Measure of Need for Cognition

Participants responded to "Which of the following best describes you?" with "If a doctor gives me advice, I like to understand the reasoning behind that advice. I'll ask questions or search the internet until I feel I really understand the issue;" or "If a doctor gives me advice, I'm usually happy to simply take that advice. I don't feel the need for lengthy explanations and justifications." This 1-item measure was developed by simplifying and consolidating the items on the NCS using more accessible, lay-friendly language.

Daily Measures

Adherence

Each day, for 14 days, participants were asked, "Did you use the strategy today?" with response options "Yes, I used it at least once," "No, I forgot to use it," or "No, I didn't have time or didn't use it for another reason." Adherence was calculated as the percentage of days participants indicated they used the strategy out of the total number of daily surveys they completed.

Follow-Up Measures

Ease of Understanding and Remembering Strategy Information

Participants responded to "How easy was it to understand the instructions for the strategy?" and "How easy was it to remember the information you were given about the strategy over the 2-week period?" on a 100 mm visual analog scale anchored from 0 (very difficult) to 100 (very easy).

Views on Length of Strategy

Participants were asked, "When we gave you the strategy, we also explained why it might be helpful. Was this explanation…" with response options "too short?," "about right?," or "too long?"

Automaticity of Strategy Use

The 4-item Self-Report Behavioral Automaticity Index (SRBAI) [44] assessed whether the assigned strategy became a habit over the 2-week period. Items were rated on a 7-point scale ranging from 1 (strongly agree) to 7 (strongly disagree). The 4 ratings were reverse-coded and summed to form a total score with higher scores indicating higher levels of automaticity.

Memory of Strategy Rationale

Four multiple-choice questions, one relating to each of the strategies, were used to assess participants' knowledge of the rationale behind each strategy (Multimedia Appendix 3).

Qualitative Questions

Furthermore, 4 optional open-ended questions were administered to explore participants' experiences during the study. Refer to Multimedia Appendix 4 for details of the qualitative aspect of the study.

Procedure

Data collection took place between August and November 2023. Eligible Oviva patients were emailed the study advert, and interested participants first completed a screening survey on Qualtrics to confirm eligibility. Participants then completed the consent form and were given enrollment instructions. Once enrolled, participants completed the baseline measures and were presented with their assigned strategy, followed by planning prompts or tips on strategy use. The short and long versions of each of the 4 strategies were independently paired with the corresponding planning prompts and tips and presented in separate sections of the survey. This created 16 unique combinations, resulting in 16 sections. Participants were randomly assigned to view one of these sections using simple randomization via the Avicenna Research app. It took approximately 30 minutes to complete baseline measures and read the strategy information. Participants then selected their preferred notification time (between 6 PM and 11 PM) for the daily surveys and were notified at their chosen time each day over the next 14 days to complete the surveys, which took less than a minute. To avoid participants mistakenly recording answers for the wrong day, each daily survey expired within 24 hours.

At 9 AM on the day following the last daily survey, participants were prompted to complete the follow-up measures. This took approximately 30 minutes. Following this, participants were given the option to complete the qualitative survey, which took an additional 30 minutes. Participants were then provided with a written debrief, and payment vouchers were issued via email.

Data Analysis

Data were analyzed in the IBM SPSS statistical analysis package (version 28). All analyses were subject to bootstrapping at 5000 resamples. Linear regression models were used to test the effects

XSL•FO
RenderX

of information length and implementation intentions on strategy adherence. The moderating effects of NCS and SSRQ scores were examined using multiple regression models (model 1) via the PROCESS macro [45]. The association between NCS score and participants' preference for information length was explored with an ordinal logistic regression model. An independent *t* test examined the effect of implementation intentions on SRBAI scores.

Point-biserial correlation tests were used to test the association between the 2 need for cognition measures and the association between the 2 planning skills measures. The moderating effect of the alternative need for cognition measure on the association between information length and adherence was explored using ANOVA, as was the moderating effect of the alternative planning skills measure on the association between implementation intentions and adherence. ANOVA was also used to explore the differences in adherence across the 4 strategies.

The effect of information length on ease of understanding and remembering strategy information was explored using linear regression, and PROCESS model 1 tested the moderating effect of NCS score. A logistic regression model evaluated the effect of information length on the likelihood that participants remember the rationale for the strategy they were assigned to, with NCS score entered as a moderator. Qualitative data were analyzed using content analysis.

Results

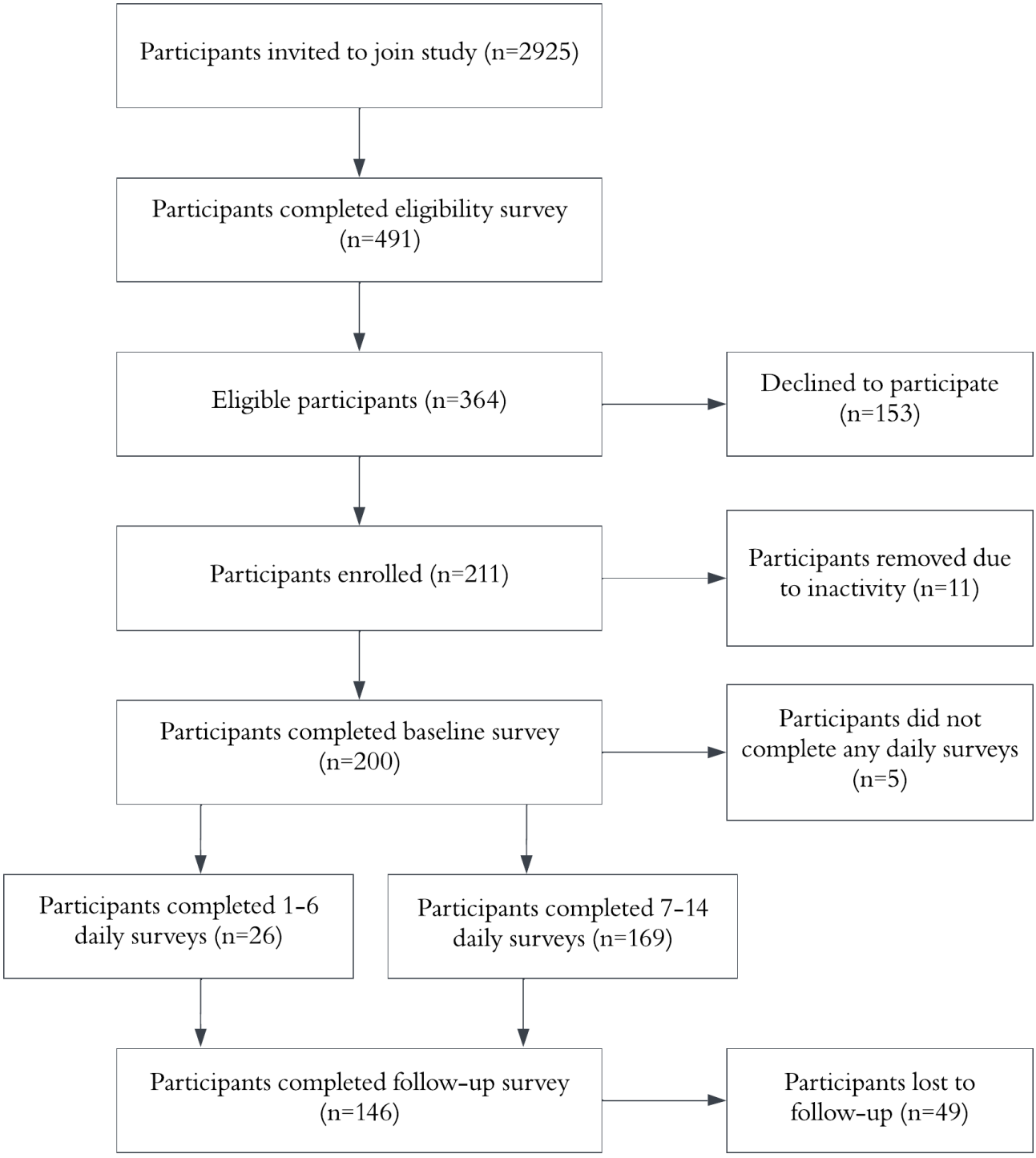
Participant Characteristics

In total, 200 participants enrolled in the study. Participants' age ranged from 23 to 79 years, with a mean of 52 (SD 11.2) years. The sample consisted of 63% (126/199) women. BMI ranged from 25.3 to 64.9 kg/m^2 with a mean of 35.5 (SD 7.3) kg/m^2. Most participants (155/197, 78%) were White, 12% (23/197) were Asian or Asian British, 5% (10/197) were Black, African, Caribbean, or Black British, 4% (7/197) were Mixed or from multiple ethnic groups, and 1% (2/197) were from other ethnic groups. Half of the sample (102/200, 51%) had an undergraduate degree or higher, 22% (43/200) were educated to General Certificate of Secondary Education (GCSE) level, 11% (22/200) had a Business and Technology Education Council (BTEC) qualification, 11% (21/200) had Advanced levels (A-levels), 4% (7/200) had no formal education, and 3% (5/200) had another form of qualification. In the UK education system, students typically undertake GCSEs at age 16. A-levels are typically taken at age 18 as a qualification for university entry, while BTEC qualifications provide vocational training.

Figure 1 presents the flow of participants through the study. As per the preregistration, only participants with data for 7 or more daily surveys were included in the main analyses, exploring the effects of information length and implementation intentions (n=169). Analyses relating to follow-up measures were restricted to participants who completed the follow-up survey (n=140). For all other analyses, the full sample was used (N=200). Participant characteristics of those included in the main analyses (n=169) were well-matched across conditions (Table 1).

XSL•FO
RenderX

Figure 1. Flow chart of participants through the study.



XSL•FO
RenderX

Table . Participant characteristics as a function of condition.

Characteristic	Implementation intentions (n=80)	Tips (n=89)	Short format (n=85)	Long format (n=84)	All (n=169)
Age, mean (SD)	53 (12)	51 (10)	51 (11)	53 (11)	52 (11)
Gender (women), n (%)	52 (66)[a]	58 (65)	53 (63)[b]	57 (68)	110 (65)[c]
BMI, mean (SD)	35.3 (6.5)	35.9 (8.3)	35.7 (7.2)	35.6 (7.8)	35.6 (7.5)
Education, n (%)					
No formal education	2 (3)	4 (5)	2 (2)	4 (5)	6 (4)
GCSEs[i], O-levels, or equivalent	18 (23)	20 (23)	17 (20)	21 (25)	38 (23)
A-levels or equivalent	7 (9)	7 (8)	6 (7)	8 (10)	14 (8)
BTEC[e] or equivalent	10 (13)	9 (10)	9 (11)	10 (12)	19 (11)
Undergraduate degree or equivalent	23 (29)	33 (37)	28 (33)	28 (33)	56 (33)
Master's degree or equivalent	16 (20)	15 (17)	18 (21)	13 (16)	31 (18)
Doctoral degree or equivalent	1 (1)	0 (0)	1 (1)	0 (0)	1 (1)
Other	3 (4)	1 (1)	4 (5)	0 (0)	4 (2)
Ethnicity[f], n (%)					
Arab	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Asian or Asian British	6 (8)	11 (13)	10 (12)	7 (8)	17 (10)
Black, Black British, Caribbean, or African	3 (4)	4 (5)	5 (6)	2 (2)	7 (4)
Mixed or multiple ethnic groups	2 (3)	2 (2)	2 (2)	2 (2)	4 (2)
White	67 (85)	70 (81)	66 (80)	71 (86)	137 (81)
Other ethnic group	1 (1)	0 (0)	0 (0)	1 (1)	1 (1)
Percentage trying to lose weight, n (%)	78 (98)	84 (94)	82 (97)	80 (95)	162 (96)
Percentage trying to avoid weight gain, n (%)	2 (3)	5 (6)	3 (4)	4 (5)	7 (4)
SSRQ[g] score, mean (SD)	34.3 (6.8)	35.3 (5.8)	35.3 (6.5)	34.3 (6.1)	34.8 (6.3)
NCS[h] score, mean (SD)	3.4 (0.8)	3.6 (0.7)	3.6 (0.8)	3.5 (0.8)	3.5 (0.8)

[a]n=79 due to missing data.

[b]n=84 due to missing data.

[c]n=168 due to missing data.

[d]GCSE: General Certificate of Secondary Education.

[e]BTEC: Business and Technology Education Council.

[f]n=79, 87, 83, 83, 166, respectively due to missing data.

[g]SSRQ: Short-form Self-Regulation Questionnaire.

[h]NCS: Need for Cognition Scale.

XSL•FO
RenderX

Main Analyses

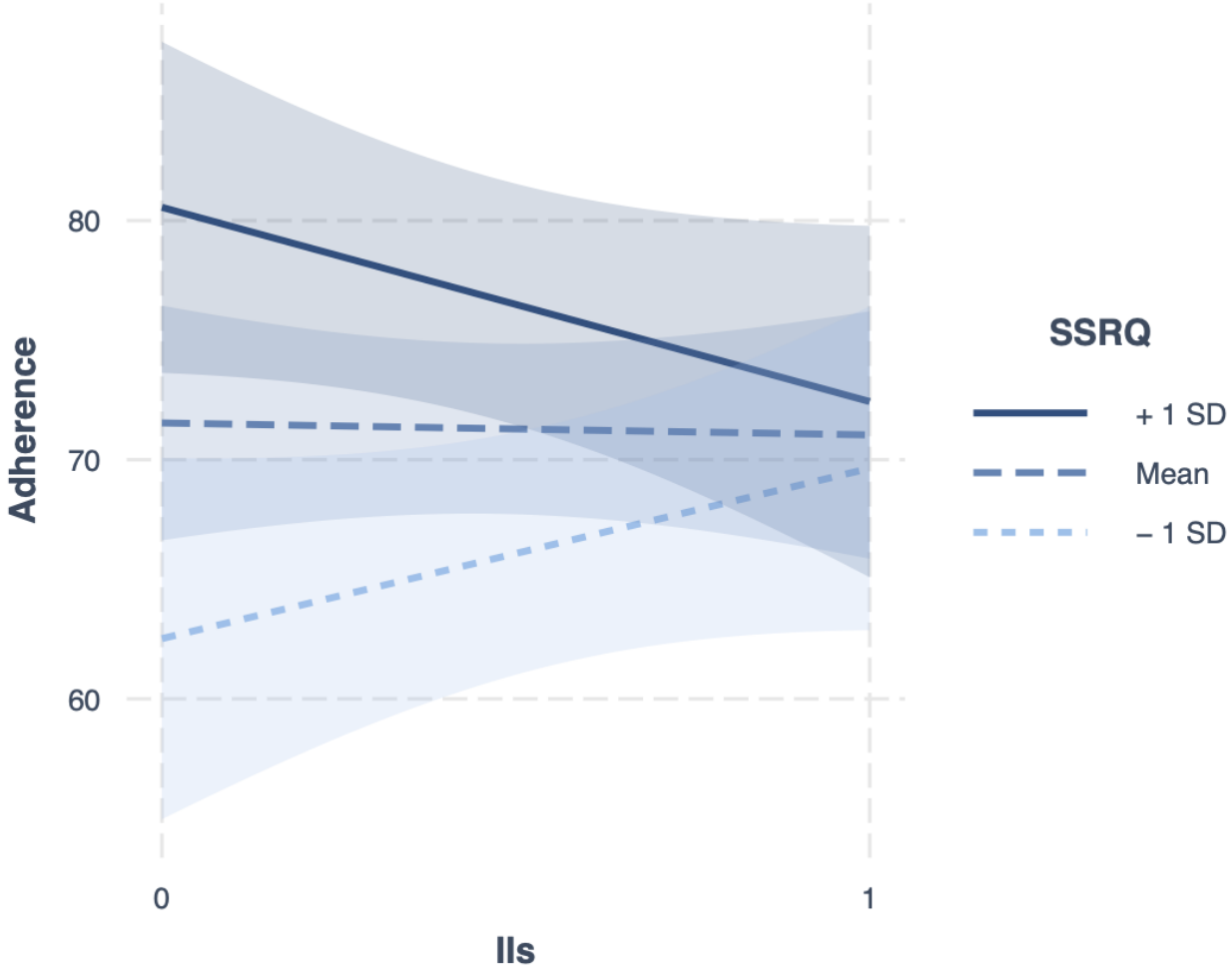
Effect of Information Length on Adherence

Mean adherence to the strategy was 74% (SD 25%) in the short information group and 69% (SD 23%) in the long information group. This difference was not statistically significant (b=−5.70, SE 3.74, 95% CI −13.00 to 1.77, P=.13); however, Cohen d indicated a small effect size (0.24). There was no significant moderation by need for cognition (b=5.50, SE 4.72, 95% CI −3.82 to 14.82, P=.25).

Effect of Implementation Intentions on Adherence

Mean adherence to the strategy was 71% (SD 27%) in the implementation intentions group and 72% (SD 21%) in the group who received tips on strategy use. Contrary to predictions, there was no significant association between forming implementation intentions and strategy adherence (b=−1.34, SE 3.71, 95% CI −8.61 to 5.84, P=.73) but as predicted, there was a significant interaction between implementation intentions and SSRQ (b=−1.21, SE 0.58, 95% CI −2.34 to −0.07, P=.04). The Johnson-Neyman technique (Multimedia Appendix 5) revealed that implementation intentions (as opposed to tips) significantly increased adherence among those who scored below 13.01 on the SSRQ (poorer planners); however, this only represented 0.6% (1/169) of the sample. In addition to our predictions, implementation intentions decreased adherence among those who scored above 48.88 (skilled planners), but again, this only represented 0.6% (1/169) of the sample. This was further explored with simple slopes analysis, which revealed a similar pattern (Figure 2); implementation intentions increased adherence among participants with poor planning skills and decreased adherence among participants with good planning skills. These results suggest that planning skills may moderate the effect of implementation intentions on adherence, but the effect may be small.

Figure 2. Simple slopes for the moderation effect of planning skills on the association between implementation intentions and adherence. II: implementation intention (0 absent, 1 present); SSRQ: Short-form Self-regulation Questionnaire.



Effect of Need for Cognition on Preference for Information Length

Contrary to predictions, there was no significant association between NCS score and odds of preference for shorter information length (odds ratio [OR] 0.96, 95% CI 0.45 - 2.05, P=.92).

Effect of Implementation Intentions on Automaticity

Also contrary to predictions, there was no significant difference in SRBAI scores between participants who formed implementation intentions (mean 17.0, SD 7.1) and those who were given tips (mean 17.4, SD 6.6; t_{138}=0.35, P=.73, 2-tailed).

XSL•FO
RenderX

Additional Exploratory Analyses

Alternative Need for Cognition Measure

As expected, there was a significant positive correlation between the two need for cognition measures (r_{pb}(198)=0.19, *P*=.01). Participants who reported that they like to understand the reasoning behind doctors' advice (n=155) had a greater NCS score (mean 3.6, SD 0.8) than those who reported that they are happy to simply take doctors' advice (n=45, mean 3.2, SD 0.8).

Contrary to the confirmatory analyses, the ANOVA exploring the moderating effect of the alternative need for cognition measure revealed a significant main effect of information length on adherence; adherence was greater in the short format group (mean 74%, SD 25%) than the long format group (mean 69%, SD 23%), $F_{1,165}$=4.97, *P*=.03, η_p^2=0.03. For individuals with a low need for cognition (n=37), mean adherence was 83% (SD 16%) in the short format group and 66% (SD 29) in the long format group. For individuals with a high need for cognition (n=132), mean adherence was 72% (SD 26%) in the short format group and 69% (SD 21%) in the long format group. These figures are in line with expectations; however, the interaction effect between information length and the alternative need for cognition measure was not significant, $F_{1,165}$=2.61, *P*=.11, η_p^2=0.02.

Alternative Planning Skills Measure

In addition, as expected, there was a significant positive correlation between the 2 planning skills measures, r_{pb}(198)=0.58, *P*<.001. Participants who reported that they were good at making plans (n=84) had a greater SSRQ score (mean 38.8, SD 5.4) than those who reported that they struggled to make plans (n=116, mean 31.5, SD 5.0).

Participants who reported being good at making plans also reported significantly greater adherence (mean 76%, SD 25%) than those who struggled to make plans (mean 68%, SD 23%), $F_{1,165}$=3.97, *P*=.048, η_p^2=0.02. For individuals with good planning skills (n=74), mean adherence was 71% (SD 30%) in the implementation intentions group and 81% (SD 17%) in the tips group. For individuals with poor planning skills (n=95), mean adherence was 71% (SD 24%) in the implementation intentions group and 66% (SD 22%) in the tips group. The interaction between implementation intentions and the alternative planning skills measure was significant, $F_{1,165}$=4.66, *P*=.03, η_p^2=.03. However, 2-tailed independent samples *t* tests revealed no significant effect of implementation intentions on adherence in individuals with good planning skills (t_{72}=1.80, *P*=.09) or individuals with poor planning skills (t_{93}=−1.19, *P*=.25).

Effect of Strategy Content on Adherence

Mean strategy adherence was 72% (SD 28%) for sensory eating, 68% (SD 25%) for attending to fullness, 61% (SD 30%) for vegetables first, and 70% (SD 29%) for physical activity. There was no significant difference between the 4 strategies ($F_{3,196}$=1.53, *P*=.21). The pattern of results remained the same when participants with less than 7 days of data were excluded (n=169).

Effect of Information Length on Self-Reported Ease of Understanding and Remembering

There was no significant association between information length and self-reported ease of understanding the strategy information (*b*=−2.62, SE 2.01, 95% CI −6.77 to 1.41, *P*=.22) and no significant interaction between information length and NCS score (*b*=0.72, SE 2.80, 95% CI −4.81 to 6.25, *P*=.80). However, ease of remembering was significantly higher among participants in the short format group (mean 83%, SD 22%) compared with those in the long format group (mean 74%, SD 25%; *b*=−9.08, SE 3.98, 95% CI −16.83 to −1.14, *P*=.02). There was no evidence for a moderation effect of NCS score (*b*=−0.84, SE 5.33, 95% CI −11.39 to 9.71, *P*=.87).

Effect of Information Length on Memory for the Strategy

Participants in the long format group were more likely to correctly remember the rationale for their assigned strategy than those in the short format group (OR 3.08, 95% CI 1.18 - 8.05, *P*=.02). There was no moderation by NCS score (OR 0.81, 95% CI 0.22 - 2.94, *P*=.75). Refer to Multimedia Appendix 6 for details of additional exploratory analyses.

Sensitivity Analyses

The analyses using the adherence variable were repeated using data from all 200 participants, where adherence was calculated over the full 2-week period (missing data were replaced with "non-adherent"). The pattern of results for all analyses remained unchanged apart from the ANOVA model exploring the moderating effect of the alternative need for cognition measure, where the main effect of information length became nonsignificant (*P*=.47). Refer to Multimedia Appendix 6 for details on additional exploratory analyses.

Qualitative Analyses

A total of 120 participants completed the qualitative survey at the end of the study. A summary of participant characteristics and full details of the analyses are provided in Multimedia Appendix 4. Responses generally centered around 4 key themes: app usability, strategy practicality, perceived benefits and effectiveness, and discussion of personal barriers to strategy use. The Avicenna Research app was generally well-received, with participants finding it easy to use and appreciating the daily survey notifications. However, many noted that reminders to use their assigned strategy would have been helpful. Most participants found their strategy easy to use, though many struggled to remember to apply it. Participants reported improvements in eating habits, physical activity, and overall well-being, and a small number of participants also reported weight loss. Family and work responsibilities often interfered with strategy use, along with other challenges such as health issues, travel, and financial constraints.

XSL•FO
RenderX

Discussion

Principal Findings

The novel key aim of this study was to explore whether adherence to brief weight management strategies over a 2-week period could be enhanced by manipulating information length, and whether this varied for individuals with different levels of need for cognition. The findings revealed no significant effect of information length on strategy adherence; however, the observed means were in the expected direction, with adherence 5 percentage points higher among those who viewed the shorter information. This represented a small effect size, which the study had not been powered to detect; while the planned regression analyses revealed a nonsignificant effect, an ANOVA, conducted as part of additional exploratory analyses, showed it as significant (P=.048). Because digital health interventions are often delivered to large numbers, an effect of this size may still be clinically significant. As such, these results warrant further exploration with an appropriately powered sample. More pronounced effects may also emerge over a longer intervention period.

Contrary to the findings of Williams-Piehota et al [16], the study failed to show that the effect of information length on adherence was influenced by need for cognition. Indeed, the overall pattern of results seems to suggest that even those with a high need for cognition may benefit from shorter information. In line with this finding, the need for cognition also failed to predict preference for information length. Findings from the exploratory analyses provide some insights into why shorter information may lead to better adherence; although memory for strategy rationale was better among those given the longer information, those given the shorter information reported that it was easier to remember the strategy. Thus, it is possible that shorter information leads to greater adherence simply because it helps people remember the strategy. A key implication of these findings is that digital health interventions may enhance adherence by limiting information length to no more than 100 words of action-oriented text. Optional links to additional details could then be provided for those who would prefer extra information.

In contrast to ease of remembering, ease of understanding was relatively high across both the short and long information groups (93% and 90%, respectively). It was also not influenced by the need for cognition, suggesting that need for cognition did not impact engagement with strategy content. It is possible that strategy understanding is better predicted by other participant characteristics, such as health literacy and cognitive ability. Health literacy encompasses skills in understanding and applying information about health issues [46], and higher levels have been associated with better health behaviors [47]. Likewise, engagement in health-promoting behaviors has been associated with greater cognitive ability as measured by general intelligence [48], processing speed [49], and analytic reasoning [50]. Given these correlations, it is possible that participants in this study had relatively high health literacy skills and cognitive abilities, which could have contributed to the high-reported ease of understanding, regardless of information length or need for cognition level. However, without direct measures of health literacy and cognitive ability, these interpretations remain speculative. Future studies may consider incorporating these measures in addition to the need for cognition.

The study's second aim was to examine whether adherence could be enhanced with implementation intentions, and whether this varied for individuals with different levels of planning skills. In contrast to previous research [22-26], we found no evidence for the benefits of implementation intentions. This discrepancy may be due to our use of an active control condition that also communicated possible cue-action links. In a systematic review by Adriaanse et al [22], most studies used a passive control group where participants received no instructions or considerably fewer instructions than the experimental group. Effects of implementation intentions were stronger across studies with these weaker control groups compared with studies with active control conditions, which administered identical instructions to both control and experimental groups (apart from the manipulation). In these latter studies, the active control condition itself may promote goal achievement to some extent, thus reducing (but not entirely eliminating) the relative advantage of implementation intentions. As in our study, they may also sometimes specify cue-action links, further reducing their advantage. Thus, while implementation intentions may be effective in promoting goal achievement, other types of behavioral advice that specify cue-action links may be just as effective.

Furthermore, contrary to predictions, the study failed to find evidence to support the notion that implementation intentions achieve their effect by increasing automaticity. Automaticity of the strategies over the 2-week period did not differ significantly between those who formed implementation intentions and those who were given "tips," with both groups reporting high automaticity (17.0 and 17.4 out of 20, respectively). Although it is plausible that the 2-week duration of the study was not sufficient to allow for any noticeable differences in automaticity to manifest, the high scores suggest that the "tips" were just as effective as implementation intentions in promoting automaticity. Again, this may be due to the inclusion of cue-action links in the "tips" condition. Further research with an additional control group would be needed to confirm this.

Nevertheless, in line with our predictions and with previous research [35], we did find that individuals with poorer planning skills reported greater adherence when given implementation intentions instead of "tips." This finding highlights the potential for digital health interventions to enhance outcomes by encouraging the use of implementation intentions in individuals with poor self-regulation and planning ability. Unexpectedly, we also found that those with better planning skills reported greater adherence when provided with "tips" rather than implementation intentions. One possible explanation for this finding is that by asking good planners to form implementation intentions, we prevented them from using their own well-rehearsed, more flexible implementation strategies. Nevertheless, the moderation effect size was relatively small. Given our sample comprised weight management program patients who had volunteered to take part in our study, their

XSL•FO
RenderX

motivation to try the strategies we provided them with may have been higher than the whole cohort of weight management program patients. As such, larger effect sizes may emerge across the whole population or over longer time frames, during which motivation may wane. The next key step in evaluating the effectiveness of personalized interventions based on planning skills would be to test this type of tailored content directly within a digital weight management program.

Another aim of the study was to explore whether simplified and more user-friendly versions of established measures of need for cognition and planning skills were effective substitutes. As expected, the short-form NCS was positively correlated with the alternative need for cognition measure (r=0.19), and the SSRQ goal setting subscale was positively correlated with the alternative planning skills measure (r=0.58). Furthermore, the alternative measures produced similar results to the standardized measures when examining their moderating effects on the influence of information length and implementation intentions on adherence. These findings are key, as both the short-form NCS and SSRQ goal-setting subscales consist of several questions (6 and 10, respectively) and are more burdensome to complete than our alternative measures, which consist of 1 question each, yet both measures yield similar results. The use of these brief alternative measures could facilitate personalized intervention by digital health companies based on need for cognition and planning ability, since they make it easier to capture differences in these measures among individuals.

Strengths and Limitations

A key strength of this study is the sample, as participants were all patients who had been referred to a digital weight management program. They are therefore representative of the types of participants who would typically be targeted by this type of intervention. However, a key limitation is that participants had already completed the weight management program and thus may have been more motivated and had more practice at implementing behavioral advice. As participants were recruited at the end of the program, the sample we obtained is likely to represent individuals with higher adherence rates, since individuals with poor adherence may have dropped out before completing the program. This is evident in the high adherence levels observed across both mental and physical strategies during the 2-week period, despite the different levels of effort and skill required for these 2 strategy types. As noted above, significant differences may have emerged with an alternative sample, such as patients at the start of a weight management program, due to differences in motivation level and experience. It is therefore recommended to further investigate the effects of information length and implementation intentions within real-world digital weight management programs. These initiatives would enhance our understanding of what fosters adherence, enabling the development of more effective interventions.

Conclusions

In conclusion, participants' adherence to brief weight management strategies over a 2-week period appeared to be greater with shorter information compared with longer information. However, the study was not adequately powered to detect the observed small effect size. This effect was not moderated by need for cognition, suggesting that shorter information may be effective for both individuals with low and high need for cognition. There was no evidence that implementation intentions improved adherence compared with the same advice presented in the form of tips. There was some evidence suggesting that implementation intentions enhanced adherence for individuals with poorer planning skills, while the use of tips improved adherence in skilled planners, highlighting the need for personalization of behavior change interventions. The study demonstrated that simplified versions of standardized measures of need for cognition and planning skills may potentially be used as suitable substitutes, offering the practitioners tools to assess user characteristics more easily. While these findings offer valuable insights, they are preliminary and necessitate replication in future research.

Acknowledgments

This work was supported by Innovate UK funding (grant 13221) awarded to KT and AB. The authors would like to thank Lucy Jones, Aoife Barron, Silvia Wels, and Philippa Hendry for help and advice regarding question content and project administration.

Authors' Contributions

The study was conceived by KT. KA and KT contributed to the methodology and study design. KA carried out the data collection and data analysis and wrote the manuscript. JS supported with participant recruitment and reviewed the manuscript. AB supported with data analysis and reviewed the manuscript. KT reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study consent.
[PDF File, 42 KB - mhealth_v13i1e65260_app1.pdf]

Multimedia Appendix 2

XSL•FO
RenderX

Full strategy information.
[DOCX File, 29 KB - mhealth_v13i1e65260_app2.docx]

Multimedia Appendix 3
Multiple choice questions.
[DOCX File, 22 KB - mhealth_v13i1e65260_app3.docx]

Multimedia Appendix 4
Qualitative analyses.
[DOCX File, 28 KB - mhealth_v13i1e65260_app4.docx]

Multimedia Appendix 5
Johnson-Neyman plot.
[DOCX File, 48 KB - mhealth_v13i1e65260_app5.docx]

Multimedia Appendix 6
Additional analyses.
[DOCX File, 25 KB - mhealth_v13i1e65260_app6.docx]

References

1. Chao AM, Quigley KM, Wadden TA. Dietary interventions for obesity: clinical and mechanistic findings. J Clin Invest 2021 Jan 4;131(1):33393504. [doi: 10.1172/JCI140065] [Medline: 33393504]
2. Hartmann-Boyce J, Johns DJ, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Effect of behavioural techniques and delivery mode on effectiveness of weight management: systematic review, meta-analysis and meta-regression. Obes Rev 2014 Jul;15(7):598-609. [doi: 10.1111/obr.12165] [Medline: 24636238]
3. Carrière K, Khoury B, Günak MM, Knäuper B. Mindfulness-based interventions for weight loss: a systematic review and meta-analysis. Obes Rev 2018 Feb;19(2):164-177. [doi: 10.1111/obr.12623] [Medline: 29076610]
4. Welton S, Minty R, O'Driscoll T, et al. Intermittent fasting and weight loss: systematic review. Can Fam Physician 2020 Feb;66(2):117-125. [Medline: 32060194]
5. Twells LK, Harris Walsh K, Blackmore A, et al. Nonsurgical weight loss interventions: a systematic review of systematic reviews and meta-analyses. Obes Rev 2021 Nov;22(11):e13320. [doi: 10.1111/obr.13320] [Medline: 34378849]
6. Lemstra M, Bird Y, Nwankwo C, Rogers M, Moraros J. Weight loss intervention adherence and factors promoting adherence: a meta-analysis. Patient Prefer Adherence 2016;10:1547-1559. [doi: 10.2147/PPA.S103649] [Medline: 27574404]
7. Birch JM, Jones RA, Mueller J, et al. A systematic review of inequalities in the uptake of, adherence to, and effectiveness of behavioral weight management interventions in adults. Obes Rev 2022 Jun;23(6):e13438. [doi: 10.1111/obr.13438] [Medline: 35243743]
8. Rimal RN, Lapinski MK. Why health communication is important in public health. Bull World Health Organ 2009 Apr;87(4):247-247a. [doi: 10.2471/blt.08.056713] [Medline: 19551226]
9. Ngigi S, Busolo DN. Behaviour change communication in health promotion: appropriate practices and promising approaches. ijird 2018;7(9):84-93. [doi: 10.24940/ijird/2018/v7/i9/SEP18027]
10. Martin LR, DiMatteo MR, editors. The Oxford Handbook of Health Communication, Behavior Change, and Treatment Adherence: Oxford Library of Psychology; 2014.
11. Cohen RA. Focused and sustained attention. In: The Neuropsychology of Attention: Springer; 2014.
12. Carr N. The Shallows: What the Internet Is Doing to Our Brains: WW Norton & Company; 2020.
13. Yeykelis L, Cummings JJ, Reeves B. Multitasking on a single device: arousal and the frequency, anticipation, and prediction of switching between media content on a computer. J Commun 2014 Feb;64(1):167-192. [doi: 10.1111/jcom.12070]
14. Uncapher MR, Wagner AD. Minds and brains of media multitaskers: current findings and future directions. Proc Natl Acad Sci U S A 2018 Oct 2;115(40):9889-9896. [doi: 10.1073/pnas.1611612115] [Medline: 30275312]
15. Cacioppo JT, Petty RE. The need for cognition. J Pers Soc Psychol 1982;42(1):116. [doi: 10.1037//0022-3514.42.1.116]
16. Williams-Piehota P, Schneider TR, Pizarro J, Mowad L, Salovey P. Matching health messages to information-processing styles: need for cognition and mammography utilization. Health Commun 2003;15(4):375-392. [doi: 10.1207/S15327027HC1504_01] [Medline: 14527864]
17. Sheeran P. Intention—behavior relations: a conceptual and empirical review. Eur Rev Soc Psychol 2002;12(1):1-36. [doi: 10.1080/14792772143000003]
18. Heckhausen H, Gollwitzer PM. Thought contents and cognitive functioning in motivational versus volitional states of mind. Motiv Emot 1987 Jun;11(2):101-120. [doi: 10.1007/BF00992338]
19. Gollwitzer PM. Mindset theory of action phases. In: Handbook of Theories of Social Psychology 2012, Vol. 1:526-545.

XSL•FO
RenderX

20. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493. [doi: 10.1037//0003-066x.54.7.493]
21. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: a meta-analysis of effects and processes. In: Adv Exp Soc Psychol 2006, Vol. 38:69-119. [doi: 10.1016/S0065-2601(06)38002-1]
22. Adriaanse MA, Vinkers CDW, De Ridder DTD, Hox JJ, De Wit JBF. Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. Appetite 2011 Feb;56(1):183-193. [doi: 10.1016/j.appet.2010.10.012] [Medline: 21056605]
23. Carrero I, Vilà I, Redondo R. What makes implementation intention interventions effective for promoting healthy eating behaviours? A meta-regression. Appetite 2019 Sep 1;140:239-247. [doi: 10.1016/j.appet.2019.05.024] [Medline: 31125588]
24. Bélanger-Gravel A. A meta-analytic review of the effect of implementation intentions on physical activity. Health Psychol Rev 2013;7(1):23-25. [doi: 10.1080/17437199.2011.560095]
25. Silva MD, São-João TM, Brizon VC, Franco DH, Mialhe FL. Impact of implementation intentions on physical activity practice in adults: a systematic review and meta-analysis of randomized clinical trials. PLoS ONE 2018;13(11):e0206294. [doi: 10.1371/journal.pone.0206294] [Medline: 30427874]
26. Kompf J. Implementation intentions for exercise and physical activity: who do they work for? A systematic review. J Phys Act Health 2020 Mar 1;17(3):349-359. [doi: 10.1123/jpah.2018-0720] [Medline: 31923898]
27. Hagerman CJ, Hoffman RK, Vaylay S, Dodge T. Implementation intentions to reduce smoking: a systematic review of the literature. Nicotine Tob Res 2021 Jun 8;23(7):1085-1093. [doi: 10.1093/ntr/ntaa235] [Medline: 33220048]
28. Luszczynska A, Sobczyk A, Abraham C. Planning to lose weight: randomized controlled trial of an implementation intention prompt to enhance weight reduction among overweight and obese women. Health Psychol 2007 Jul;26(4):507-512. [doi: 10.1037/0278-6133.26.4.507] [Medline: 17605571]
29. Armitage CJ, Norman P, Noor M, Alganem S, Arden MA. Evidence that a very brief psychological intervention boosts weight loss in a weight loss program. Behav Ther 2014 Sep;45(5):700-707. [doi: 10.1016/j.beth.2014.04.001] [Medline: 25022780]
30. Benyamini Y, Geron R, Steinberg DM, Medini N, Valinsky L, Endevelt R. A structured intentions and action-planning intervention improves weight loss outcomes in a group weight loss program. Am J Health Promot 2013;28(2):119-127. [doi: 10.4278/ajhp.120727-QUAN-365] [Medline: 23621700]
31. Hayes JF, Balantekin KN, Graham AK, Strube MJ, Bickel WK, Wilfley DE. Implementation intentions for weight loss in college students with overweight and obesity: a proof-of-concept randomized controlled trial. Transl Behav Med 2021 Mar 16;11(2):359-368. [doi: 10.1093/tbm/ibaa038] [Medline: 32359068]
32. Knäuper B, Shireen H, Carrière K, et al. The effects of if-then plans on weight loss: results of the 24-month follow-up of the McGill CHIP Healthy Weight Program randomized controlled trial. Trials 2020 Jan 7;21(1):40. [doi: 10.1186/s13063-019-4014-z] [Medline: 31910891]
33. Verhoeven AAC, Adriaanse MA, de Ridder DTD, de Vet E, Fennis BM. Less is more: the effect of multiple implementation intentions targeting unhealthy snacking habits. Euro J Social Psych 2013 Aug;43(5):344-354 [FREE Full text] [doi: 10.1002/ejsp.1963]
34. de Vet E, Oenema A, Brug J. More or better: do the number and specificity of implementation intentions matter in increasing physical activity? Psychol Sport Exerc 2011 Jul;12(4):471-477. [doi: 10.1016/j.psychsport.2011.02.008] [Medline: 28783502]
35. Allan JL, Sniehotta FF, Johnston M. The best laid plans: planning skill determines the effectiveness of action plans and implementation intentions. Ann Behav Med 2013 Aug;46(1):114-120. [doi: 10.1007/s12160-013-9483-9] [Medline: 23456214]
36. Ahmadyar K, Szypula J, Bogosian A, Tapper K. Improving adherence to weight management strategies: information length and implementation intentions. OSF. 2023. URL: https://doi.org/10.17605/OSF.IO/NX3GJ [accessed 2025-07-14]
37. Seguias L, Tapper K. The effect of mindful eating on subsequent intake of a high calorie snack. Appetite 2018 Feb 1;121:93-100. [doi: 10.1016/j.appet.2017.10.041] [Medline: 29104106]
38. Jordan CH, Wang W, Donatoni L, Meier BP. Mindful eating: trait and state mindfulness predict healthier eating behavior. Pers Individ Dif 2014 Oct;68:107-111. [doi: 10.1016/j.paid.2014.04.013]
39. Nishino K, Sakurai M, Takeshita Y, Takamura T. Consuming carbohydrates after meat or vegetables lowers postprandial excursions of glucose and insulin in nondiabetic subjects. J Nutr Sci Vitaminol (Tokyo) 2018;64(5):316-320. [doi: 10.3177/jnsv.64.316] [Medline: 30381620]
40. Buffey AJ, Herring MP, Langley CK, Donnelly AE, Carson BP. The acute effects of interrupting prolonged sitting time in adults with standing and light-intensity walking on biomarkers of cardiometabolic health in adults: a systematic review and meta-analysis. Sports Med 2022 Aug;52(8):1765-1787. [doi: 10.1007/s40279-022-01649-4] [Medline: 35147898]
41. Patient-centric clinical trial and health research software platform. Avicenna Research. URL: https://avicennaresearch.com [accessed 2025-07-14]
42. Neal DJ, Carey KB. A follow-up psychometric analysis of the self-regulation questionnaire. Psychol Addict Behav 2005 Dec;19(4):414-422. [doi: 10.1037/0893-164X.19.4.414] [Medline: 16366813]
43. Lins de Holanda Coelho G, H P Hanel P, J Wolf L. The very efficient assessment of need for cognition: developing a six-item version. Assessment 2020 Dec;27(8):1870-1885. [doi: 10.1177/1073191118793208] [Medline: 30095000]

XSL•FO
RenderX

44. Gardner B, Abraham C, Lally P, de Bruijn GJ. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int J Behav Nutr Phys Act 2012 Aug 30;9:102. [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
45. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach: Guilford publications; 2017.
46. Ishikawa H, Kiuchi T. Health literacy and health communication. Biopsychosoc Med 2010 Nov 5;4:18. [doi: 10.1186/1751-0759-4-18] [Medline: 21054840]
47. Šulinskaitė K, Zagurskienė D, Blaževičienė A. Patients' health literacy and health behaviour assessment in primary health care: evidence from a cross-sectional survey. BMC Prim Care 2022 Sep 5;23(1):223. [doi: 10.1186/s12875-022-01809-5] [Medline: 36064351]
48. Auld MC, Sidhu N. Schooling, cognitive ability and health. Health Econ 2005 Oct;14(10):1019-1034. [doi: 10.1002/hec.1050] [Medline: 16167318]
49. Anstey KJ, Low LF, Christensen H, Sachdev P. Level of cognitive performance as a correlate and predictor of health behaviors that protect against cognitive decline in late life: the path through life study. Intelligence 2009 Nov;37(6):600-606. [doi: 10.1016/j.intell.2008.10.001]
50. Junger M, van Kampen M. Cognitive ability and self-control in relation to dietary habits, physical activity and bodyweight in adolescents. Int J Behav Nutr Phys Act 2010 Mar 23;7:22. [doi: 10.1186/1479-5868-7-22] [Medline: 20331887]

Abbreviations

A-level: Advanced level
BTEC: Business and Technology Education Council
GCSE: General Certificate of Secondary Education
NCS: Need for Cognition Scale
OR: odds ratio
SRBAI: Self-report Behavioral Automaticity Index
SSRQ: Short-form Self-Regulation Questionnaire

Edited by L Buis; submitted 10.08.24; peer-reviewed by JF Hayes, WJ He; revised version received 19.03.25; accepted 06.05.25; published 08.08.25.

Please cite as:
Ahmadyar K, Szypula J, Bogosian A, Tapper K
Effects of Information Length and Implementation Intentions on Adherence to Weight Management Strategies: Experimental Study
JMIR Mhealth Uhealth 2025;13:e65260
URL: https://mhealth.jmir.org/2025/1/e65260
doi:10.2196/65260

© Khaleda Ahmadyar, Joanna Szypula, Angeliki Bogosian, Katy Tapper. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

User Archetypes of a Well-Being–Promoting Mobile App Among Adults: Cross-Sectional Study and Cluster Analysis of Usage Patterns

Hanna Rekola[1,2], MSc; Tommi Tolmunen[3,4], Prof Dr Med; Elina Mattila[5], DSc; Juho Strömmer[6], PhD; Timo A Lakka[7,8,9], Prof Dr Med; Helena Länsimies[10], PhD; Tomi Mäki-Opas[1,2], Prof Dr

[1]Department of Social Sciences, Faculty of Social Sciences and Business Studies, University of Eastern Finland, PO Box 1627, Kuopio, Finland
[10]City of Kuopio, Kuopio, Finland
[2]Social, Wellbeing, and Rescue Research Centre, Wellbeing Services County of North Savo, Kuopio, Finland
[3]Institute of Clinical Medicine, School of Medicine, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland
[4]Department of Adolescent Psychiatry, Kuopio University Hospital, Kuopio, Finland
[5]VTT Technical Research Centre of Finland Ltd, Espoo, Finland
[6]Wellpro Impact Solutions Oy, Jyväskylä, Finland
[7]Institute of Biomedicine, School of Medicine, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland
[8]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland
[9]Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Excercise Medicine, Kuopio, Finland

Corresponding Author:
Hanna Rekola, MSc
Department of Social Sciences, Faculty of Social Sciences and Business Studies, University of Eastern Finland, PO Box 1627, Kuopio, Finland

Abstract

Background: A healthy lifestyle is associated with mental well-being, and digital lifestyle interventions can be effective in promoting a healthy lifestyle. However, they do not appear to work for all, and we have limited knowledge of how users' background characteristics affect their tendency to adopt well-being–promoting digital apps and actively use them.

Objective: This study aimed to explore the association of the study participants' characteristics and current well-being with their likelihood of using a well-being–promoting mobile app.

Methods: The BitHabit web app (Wellpro Impact Solutions Ltd) was available for a 2-month trial in spring 2023 after completing a short cross-sectional digital questionnaire with questions about well-being, life satisfaction, and lifestyle. Individuals aged 15 years or younger were excluded from the analysis. We used logistic regression to assess how individual characteristics were associated with the initiation of BitHabit app use. To assess user archetypes among those who initiated app use, and k-means clustering analysis and multinomial logistic regression to assess user archetypes among those who initiated app use.

Results: A total of 1646 eligible individuals responded to the questionnaire, and 863 initiated app use. Lower odds of initiating app use were detected among males (odds ratio [OR] 0.66, 95% CI 0.51 - 0.85; *P*<.001), the unemployed (OR 0.68, 95% CI 0.48 - 0.97; *P*=.03), those with higher general life satisfaction (OR 0.94, 95% CI 0.89 - 1.00; *P*=.04), and those reporting fewer life challenges (OR 1.13, 95% CI 1.02 - 1.24; *P*=.02). We identified (1) thriving non-active users, (2) struggling non-active users, and (3) active users as archetypes based on app use activity, life satisfaction, and reported life challenges. Older participants had lower odds of being thriving nonactive (OR 0.96, 95% CI 0.94 - 0.99; *P*=.01) or struggling nonactive users (OR 0.93, 95% CI 0.90 - 0.96; *P*<.001) than active users. Retired participants had higher odds of being struggling nonactive than active users (OR 4.06, 95% CI 1.44 - 11.42; *P*=.01) and unemployed lower odds of being thriving nonactive than active users (OR 0.2, 95% CI 0.08 - 0.51; *P*<.001). Those who were physically more active had higher odds of being thriving nonactive than active users (OR 2.71, 95% CI 1.00 - 7.32; *P*=.05). Participants with higher alcohol consumption had higher odds of being struggling nonactive users than active users (OR 3.22, 95% CI 1.16 - 8.99; *P*=.03).

Conclusions: While lower general life satisfaction and less favorable health behavior appeared to increase the likelihood of trying the app, those who eventually actively used the app were more satisfied with their lives at baseline. In addition, among nonactive users, there were recognizable user profiles of thriving and struggling nonactive users, which were associated with various individual characteristics. Further research is needed to develop digital apps to attract more potential users and meet the needs of those with an unhealthy lifestyle and poor mental health.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e68982) doi:10.2196/68982

KEYWORDS

user archetypes; adherence; user characteristics; digital well-being promotion; mobile app; lifestyle; behavior modification; behavioral change

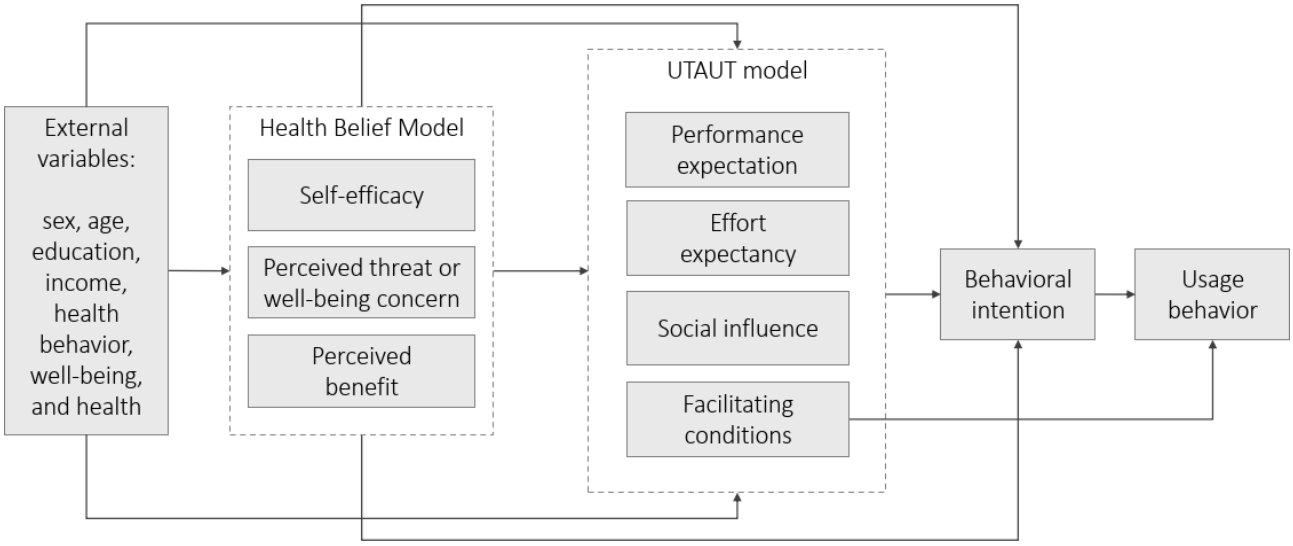
Introduction

Mental health challenges pose a burden on well-being and public health, partly due to digitalization and changes in lifestyle and working life [1]. On the other hand, in social and health services, new digital apps and innovations are becoming increasingly common, and the promotion of health and well-being is becoming more important than ever. Digital interventions (ie, eHealth and mobile health [mHealth] tools) might have the potential to tackle these challenges.

Based on extensive evidence, a healthy lifestyle is associated with mental health and well-being [1]. Digital lifestyle interventions are effective in promoting not only a healthy lifestyle but also mental health [2-4]. More specifically, empirical evidence supports the effectiveness of lifestyle modification based on self-determination theory [5] and a self-regulation approach [6]. However, little research has been conducted on digital well-being–promoting interventions in healthy populations [4]. Furthermore, these interventions do not appear to produce the same outcomes for everyone, and there is still limited evidence on how user characteristics affect the tendency to adopt digital apps to promote well-being and how people use these apps [7].

We use a combination of health behavior (the Health Belief Model [HBM]) [8] and technology acceptance perspectives (the Unified Theory of Acceptance and Use of Technology [UTAUT]) [9] to understand the mechanisms underlying the selection and adoption of well-being–promoting apps (Figure 1). The HBM suggests that perceived susceptibility to health issues and perceived benefits of action influence individuals' motivation to adopt preventive strategies, such as well-being apps. In addition, users with high self-efficacy, or confidence in their ability to manage their well-being, may also be more likely to use the app. Based on the UTAUT model, app engagement is more likely when users expect it to improve their well-being, find it easy to use, perceive social endorsement, and have adequate access to technology. In this context, demographic and psychosocial factors may shape how likely individuals are to engage with well-being–promoting apps.

Figure 1. The combined theoretical framework of the Health Belief Model and the Unified Theory of Acceptance and Use of Technology. UTAUT: Unified Theory of Acceptance and Use of Technology.



According to existing evidence, women are more likely to seek health information on the internet compared with men [10], and women also appear to be more willing to adopt digital tools to promote their health and well-being [11,12], but contradictory results have been obtained in some studies [13]. In addition, among working-age adults, older individuals have been found to use digital well-being–promoting apps more frequently than younger people [11-15]. Social determinants, such as employment, education, or income, have been associated in some studies with the use of digital health promotion tools, but the results have also been contradictory [11-13,16]. While in some studies, the unemployed have been more likely to use digital health interventions as recommended [11], in others, they have been less likely to initiate use of a digital lifestyle intervention [13].

It has been suggested that digital well-being–promoting interventions are particularly attractive to people with a greater need for health information. Some evidence suggests that people with lower life satisfaction, an unhealthy lifestyle, health issues, and challenges are more likely to initiate the use of digital health–promoting apps and use them more frequently compared with the general population [7,13,17]. One reason behind this attraction could be the perceived anonymity behind digital interventions, which might make them more appealing, for example, to people with a lower level of physical activity, poorer

XSL•FO
RenderX

health, and mental health problems [18,19]. Conversely, some researchers have proposed that behavioral intervention programs may better reach those who already have healthier lifestyles, higher life satisfaction, and better well-being at baseline [11,14,15,20], and who thus need these interventions the least.

This study was a part of the Feeling Good North Savo program [21]. The aim of the study was to explore the associations between the characteristics of participants and their likelihood of using a mobile app promoting well-being. Our assumption, based on the limited existing evidence, was that women, older people, and employed people—indicating a higher socioeconomic status—would be more interested in trying and using the app. In addition, we expected that people with healthier lifestyles, fewer reported life challenges, and better life satisfaction would be more likely to use the app.

Our specific research questions were: (1) Are individual characteristics, behavioral factors, and the well-being of the participants associated with the probability of initiating the use of a well-being–promoting app? and (2) How do user archetypes, identified based on activity in using the app, reported life challenges, and life satisfaction, differ in terms of individual characteristics, behavioral factors, and well-being among the participants who initiated app use?

Methods

Study Design and Recruitment

This observational cross-sectional study is reported following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) [22] (the STROBE checklist is provided in Checklist 1) and mERA (mHealth Evidence Reporting and Assessment) [23] guidelines (the mERA checklist is provided in Checklist 2).

The web-based questionnaire and intervention were available to the public from late February until early May 2023. The study was promoted in the North Savo region of Eastern Finland through local newspapers, municipalities' communication channels, and the Feeling Good North Savo program's social media and website. Everyone who came across the invitation could participate in the study. We had no definitive inclusion or exclusion criteria, as the study was designed to act as a feasibility study for a further definitive trial. However, due to ethical reasons, one individual under the age of 15 years was excluded from the analysis. In addition, as the study invitation, digital questionnaire, and intervention were conducted in Finnish, proficiency in the Finnish language was an implicit inclusion criterion. Completing the MySeula digital questionnaire took approximately 5 minutes. Afterward, participants had the option to register for the BitHabit app and use it for 1-2 months, depending on when they joined the study.

BitHabit Web App

Our intervention model used the MySeula digital questionnaire and the BitHabit web application provided by Wellpro Impact Solutions [24]. The MySeula digital survey was used to collect the study data, including background characteristics as well as well-being and lifestyle information. Each time participants completed the survey, they received an automatic report on their well-being status. In addition to collecting the study data, the digital survey was intended to steer the use of the BitHabit web app and raise awareness of well-being promotion among the participants.

The BitHabit web app was originally developed in the Finnish Stop Diabetes study to motivate and facilitate adults to improve their health behavior by providing valuable information and recommending small habits to be integrated into daily routines [2,25]. The BitHabit web app combined with group meetings was found effective in improving the diet quality of the study participants [2]. In addition, a web app without the group meetings was associated with increased physical activity and decreased sedentary time, but only among participants with good adherence to the intervention.

The principal concept of the app is based on self-determination behavior change theory, which supports autonomous motivation and competence, emphasizing freedom of choice and drawing on users' existing knowledge, skills, and habits [5]. Furthermore, the app relies on the habit formation theory, which promotes behavior change through repetition of simple, contextualized, and frequent behaviors to achieve habit automaticity [26-28]. In addition, app development is based on behavior change techniques from self-regulation theory, such as self-monitoring, goal setting, and action planning [6,29].

BitHabit is a web application accessible from any modern browser on smartphones and computers. Infrastructure to support technology operations in North Savo is well developed, with widespread internet access and networks. The app user has access to a comprehensive library of well-being–promoting habits, as described in detail earlier [25]. The application is based on the idea that users engage with it on their own initiative: the user explores the library of habits, plans how to integrate them into their daily life based on their personal needs and the results of the MySeula survey, and logs completed habits in the application. Users can monitor their progress by keeping track of the habits they have completed. The app sends prompts to the user via text message. These prompts are generic reminders to use the app, particularly when it has not been used for a while. The application does not directly suggest specific small actions to the user.

In this study, the app was iteratively tested by trained experts to better fit the target group's needs. Across 3 rounds with 8 users, technical and content adjustments were made, including the addition of several mental health–related daily habits. These included relaxation and breathing exercises, mindfulness, fostering positive self-talk, and writing down personal thoughts, among others.

Measures

Behavioral Factors

Physical activity of the study participants was measured using questions adapted from the Stop Diabetes study [2], the Finnish National FINRISK Study [30], and the International Physical Activity Questionnaire [31]. We inquired with structural questions about how many hours per week the participants carried out conditioning physical activity. We also inquired, with a structural question, how many hours the respondents

XSL•FO
RenderX

engaged in light physical activity or spent sitting or lying during a day (excluding sleeping). The satisfaction of participants with their current physical condition was also assessed with a self-reported measure on a scale from 1 to 10.

Alcohol consumption was assessed with a self-reported measure of how often the participants had consumed 6 or more drinks on one occasion during the past year. The item is included in the Alcohol Use Disorders Identification Test-Concise (AUDIT-C), developed by the World Health Organization (WHO) [32]. For our short questionnaire, we exclusively used the item assessing the frequency of habitual intoxication-seeking behavior.

Well-Being

Life satisfaction was assessed using a self-reported measure of general satisfaction with life at the moment on a scale from 1 to 10, based on the Organisation for Economic Co-operation and Development (OECD) Better Life Index. Life satisfaction subjectively measures how people evaluate their life as a whole rather than their current feelings [33]. We also assessed, with a similar scale, how satisfied the participants were with their physical condition and social relationships.

Life challenges were assessed using a structured multiple-choice question. Possible response alternatives were physical impairment, psychological impairment, cognitive impairment, difficulty in social interaction, financial challenges, or none of these. All of these were used in the analysis as binary variables. A summary variable indicating the number of life challenges reported by the participant was also computed.

Participant Characteristics

We included questions about age, gender, and employment status as individual characteristics. These variables were used to describe the study sample and explore subgroup differences.

BitHabit App Use

BitHabit app use was assessed with comprehensive user log data. All events in the log data included timestamps. For this study, we analyzed the initiation of app use, frequency of sessions, and the selection and performance of habits. “Initiation of app use” refers to the participant creating an account in the app. A “session” refers to the user launching the mobile app on their device. We calculated an “active days ratio,” which is the ratio of days with sessions to all days during the time of study participation. Performing a habit refers to the user marking a specific habit as selected or completed in the app. We also calculated the “habit ratio,” referring to the average number of habits selected or completed per day during the user’s participation in the study.

Statistical Analysis

The data were processed and analyzed using RStudio software (R Core Team, Posit Team) [34,35]. To assess the association between the initiation of app use and user characteristics, simple unweighted 2-way crosstab tables were generated with the chi-square test of association as well as correlation matrices. For continuous variables, we tested mean differences using the Mann-Whitney *U* test. We also tested logistic regression models that included the background characteristics at the same time, accounting for dependencies among the variables to better understand their influence on the initiation of app use. We included in the logistic regression models all confounders analyzed in the univariate assessments, with two exceptions: only the physical activity indicator with the highest statistical significance was retained, and only general life satisfaction was included due to its high correlation with satisfaction in social relationships and physical condition. Missing information is presented in the crosstabulation tables and excluded from the logistic regression models. A cutoff of $P<.05$ was used to determine statistical significance in all analyses.

Different user archetypes among participants who initiated app use were recognized using k-means clustering analysis. The methodology was similar to that of Aziz et al [36]. The k-means clustering method defines clusters so that the total within-cluster variation is minimized. The “kmeans” function from the R package “stats” [35] was used with the default clustering algorithm by Hartigan and Wong [37]. In the Hartigan-Wong algorithm, each participant is assigned to a cluster so that the sum of squared Euclidean distances between observations and the cluster centroid, that is, the total within-cluster variation, is minimized. For the visualizations, we used the R packages “factoextra” [38] and “ggplot2” [39].

The clustering analysis included 4 variables: general life satisfaction, number of reported life challenges or impairments, ratio of active app usage days to no usage days, and average number of reported habits per day. The clustering variables were chosen based on the findings from previous studies indicating that the number of health issues and life satisfaction would be associated with app usage patterns, but with conflicting results [7,11,13-15,17-20].

The optimal number of clusters was determined using the elbow method, which assessed the total within-cluster sum of squares across solutions with 2 to 10 clusters (Figure 2). The plot showed a clear inflection point at k=3, beyond which additional clusters yielded only marginal improvements in model fit. This was supplemented by manual inspection of cluster interpretability with models based on 2-5 clusters, leading to the selection of 3 clusters. One of the clusters accounted for 65% (474/727) of users, and the smallest 8% (55/727). The unequal cluster size distribution was anticipated based on our hypotheses and characteristics of the sample. It was expected that a substantial proportion of participants would demonstrate low engagement with the app, reflecting typical patterns in previous studies [2].

To further assess the classification bias, we examined the principal component analysis plots by cluster color to assess separation and dominance patterns and detected no overlapping (Figure 3). We also assessed the stability of the clustering by executing the algorithm several times with different random initializations, and the results remained stable across runs. Finally, we used univariate assessments and multinomial logistic regression to assess how different user characteristics were represented among the clusters. The confounders included in the multinomial regression model were the same as those used in the logistic regression model for app use initiation, excluding the variables involved in the clustering analysis.

XSL•FO
RenderX

Figure 2. Elbow plot of the optimal number of clusters among participants who initiated BitHabit app use.

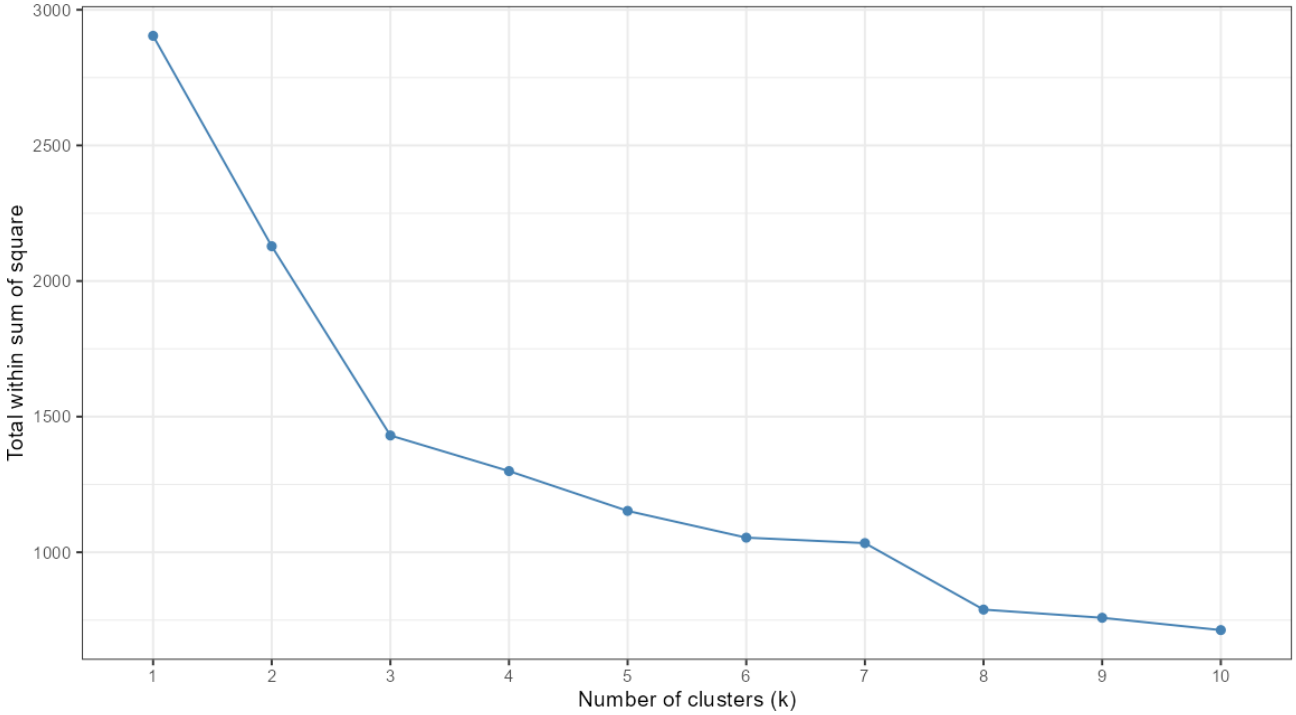
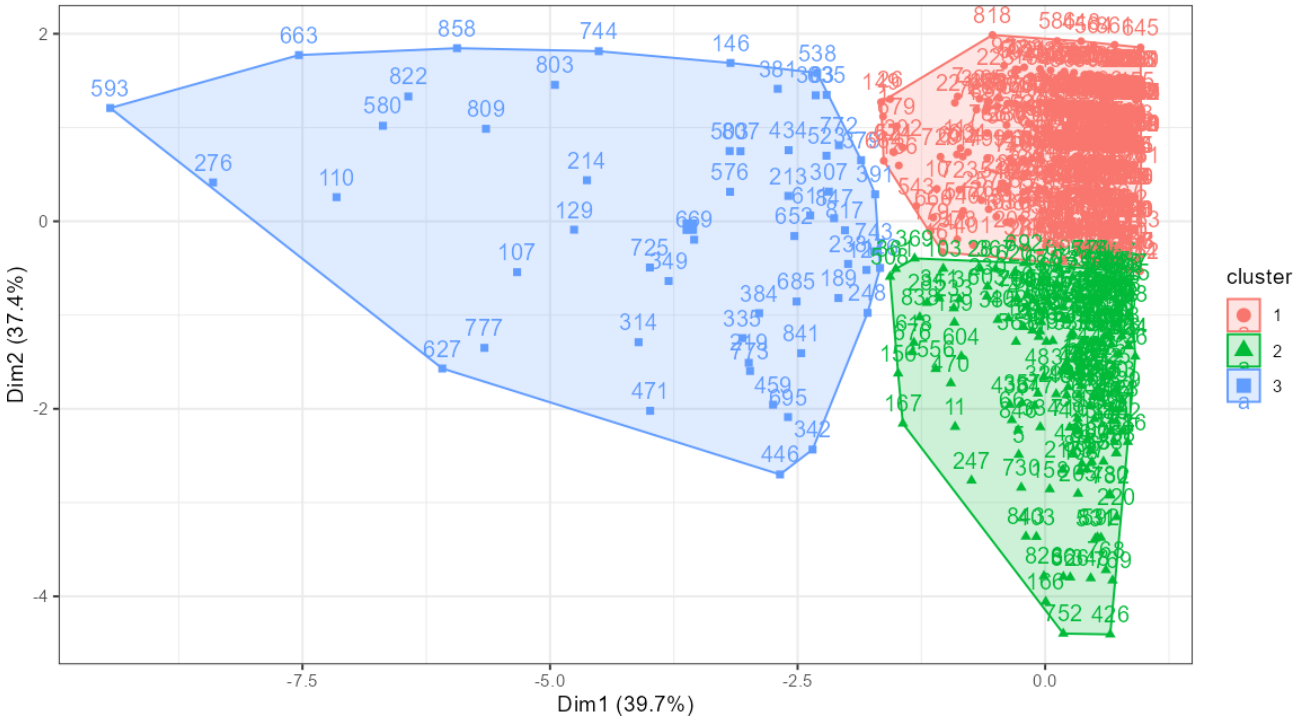


Figure 3. Principal component analysis plot colored by cluster assignment.



Ethical Considerations

The University of Eastern Finland is committed to adhering to the ethical guidelines set forth by the Finnish National Board on Research Integrity (TENK) [40]. These guidelines apply to research involving human participants, except for medical research as defined by the Finnish Medical Research Act (488/1999) [41] and other study types for which ethical review is mandated separately by law. Following TENK's principles, research must be conducted in a manner that upholds the dignity and autonomy of participants and avoids causing significant risk, harm, or damage to individuals, communities, or other participants of research.

An ethical review must be sought before data collection if the study includes any of the following conditions: (1) participation in the study deviates from the principle of informed consent, such as when participation is not voluntary or when participants are not adequately informed about the research; (2) the research involves interventions that affect the physical integrity of participants; (3) the study focuses on children aged 15 years or younger without obtaining separate consent from a parent or

XSL•FO
RenderX

guardian, or without providing them with sufficient information to allow them to prevent the child's participation; (4) participants are exposed to exceptionally strong stimuli; (5) the research carries a risk of causing mental harm beyond the typical stresses of daily life to participants, their family members, or others close to them; and (6) the research may pose a threat to the safety of participants, researchers, or their close contacts. None of these conditions were met in this study; hence, an ethical review was not required.

Informed consent was considered to have been given when participants, after reading the study invitation, chose to respond to the questionnaire. The invitation stated how the collected data would be used in the study conducted by the University of Eastern Finland. It was also stated that no individual user could be identified or traced based on the data collected.

Personal data, such as email addresses, was collected by the service provider during app registration for customer management purposes. No national identification numbers or other sensitive personal data were collected. All data transfers outside the European Union and the European Economic Area, if any, were conducted following Article 46 of the General Data Protection Regulation [42], using appropriate safeguards such as the European Commission's adequacy decisions or standard contractual clauses. The data used by the research team were deidentified, and they only had access to pseudonymized datasets that did not include any directly identifiable information. Participants received no compensation for their participation in the study.

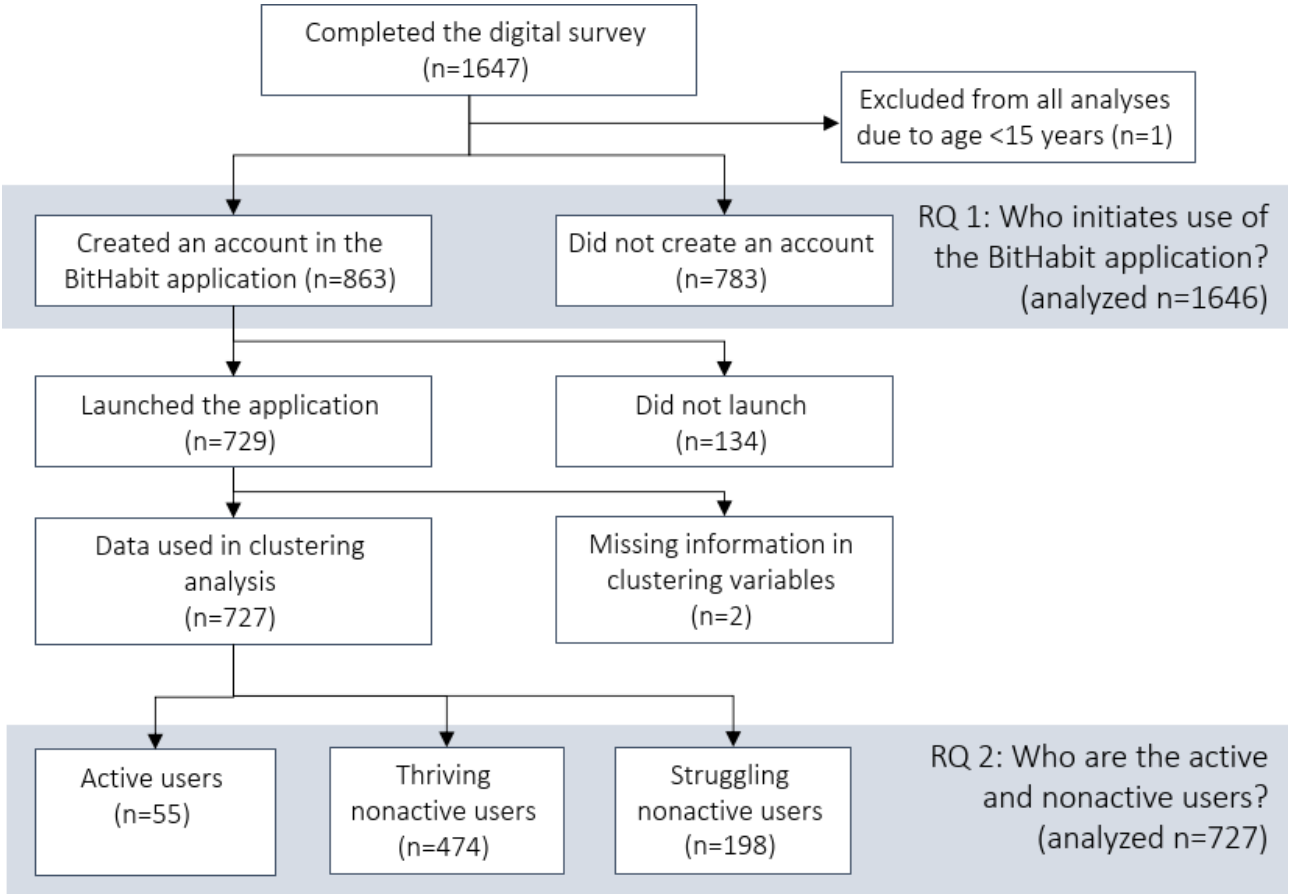
Results

Descriptive Statistics

Figure 4 shows the participant flow of the study. The total number of participants was 1646 (Figure 4), and 77% (1263/1642) of those with nonmissing gender information were women. The mean age among participants was 46.7 (SD 14.2) years. The majority of the participants were used (57%, 937/1646) or retired (18%. 289/1646; Table 1). Most participants (69%, 1135/1642) reported engaging in conditioning exercise for more than 1 hour per week. Approximately 10% (182/1641) of participants consumed 6 or more doses of alcohol every week, and 23% (370/1641) monthly or more often. Almost 80% (1271/1641) reported consuming 6 or more doses less than once a month. Men reported consuming alcohol more often than women.

Participants reported having, on average, at least 1 life challenge; 34% (553/1646) of the participants did not report any. The most commonly reported life challenges were psychological impairment (31%, 516/1646) and physical impairment (29%, 477/1646). Most rarely, participants reported having cognitive impairment (17%, 273/1646). General life satisfaction among the participants was 6.6 on average.

Figure 4. Flowchart of participant selection for the analyses. RQ: research question.



XSL•FO
RenderX

Table . Individual characteristics of study participants (N=1646), stratified by initiation of BitHabit app use.'

Characteristic	All participants (N=1646)	Initiated app use (n=863)	Did not initiate app use (n=783)	*P* value[a]
Age (years), mean (SD)	46.7 (14.2)	46.5 (13.8)	46.9 (14.5)	.51
Missing, n[b]	0	0	0	
Gender, n (%)				<.001
Male	363 (22.1)	159 (18.4)	204 (26.2)	
Female	1263 (76.9)	694 (80.4)	569 (73.0)	
Other	16 (1.0)	10 (1.2)	6 (0.8)	
Missing, n	4	0	4	
Employment status, n (%)				.47
Employed	937 (56.9)	507 (58.7)	430 (54.9)	
Retired	289 (17.6)	146 (16.9)	143 (18.3)	
Student	148 (9.0)	75 (8.7)	73 (9.3)	
Entrepreneur	108 (6.6)	53 (6.1)	55 (7.0)	
Unemployed	91 (5.5)	41 (4.8)	50 (6.4)	
Other	73 (4.4)	41 (4.8)	32 (4.1)	
Missing, n	0	0	0	
Six or more doses of alcohol on one occasion, n (%)				.15
Never	526 (32.0)	267 (30.9)	259 (33.1)	
Less than once a month	745 (45.3)	412 (47.7)	333 (42.5)	
Once a month	188 (11.4)	96 (11.1)	92 (11.7)	
Once a week	158 (9.6)	71 (8.2)	87 (11.1)	
Daily or almost daily	24 (1.5)	13 (1.5)	11 (1.4)	
Missing, n	5	4	1	
Sedentary per day, n (%)				.36
5≥ hours	517 (31.4)	263 (30.5)	254 (32.4)	
6-8 hours	747 (45.4)	389 (45.1)	358 (45.7)	
>8 hours	378 (23.0)	210 (24.3)	168 (21.5)	
Missing, n	4	1	3	
Conditioning exercise per week, n (%)				.02
<1 hour	507 (30.8)	276 (32.0)	231 (29.5)	
1-3 hours	757 (46.0)	409 (47.4)	348 (44.4)	
4≤ hours	378 (23.0)	174 (20.2)	204 (26.1)	
Missing, n	4	4	0	
Light physical activity per day, n (%)				.006
<2 hours	460 (27.9)	245 (28.4)	215 (27.5)	
2-3 hours	739 (44.9)	411 (47.6)	328 (41.9)	
>3 hours	445 (27.0)	205 (23.8)	240 (30.7)	
Missing, n	2	2	0	
Life challenges or impairments, n (%)[c]				
Physical impairment	477 (29.0)	265 (30.7)	212 (27.1)	.10
Psychological impairment	516 (31.3)	296 (34.3)	220 (28.1)	.007
Cognitive impairment	273 (16.6)	159 (18.4)	114 (14.6)	.04

XSL•FO
RenderX

Characteristic	All participants (N=1646)	Initiated app use (n=863)	Did not initiate app use (n=783)	*P* value[a]
Difficulty in social interaction	332 (20.2)	193 (22.4)	139 (17.8)	.02
Financial challenges	394 (23.9)	218 (25.3)	176 (22.5)	.19
None of these	553 (33.6)	265 (30.7)	288 (36.8)	.009
Number of reported life challenges or impairments, mean (SD)	1.21 (1.2)	1.31 (1.2)	1.10 (1.2)	<.001
Missing, n	0	0	0	
Satisfaction with physical condition, mean (SD)	5.18 (2.6)	4.99 (2.6)	5.38 (2.5)	.002
Missing, n	3	1	2	
Satisfaction with social relationships, mean (SD)	6.70 (2.3)	6.57 (2.3)	6.84 (2.4)	.004
Missing, n	4	3	1	
General life satisfaction, mean (SD)	6.64 (2.1)	6.50 (2.1)	6.80 (2.2)	<.001
Missing, n	3	2	1	

[a]*P* values were calculated using the Mann-Whitney *U* test for continuous variables and the chi-square test for categorical variables.

[b]The "Missing" row for each variable refers to the number of participants with incomplete data for that variable. These cases were excluded from the corresponding statistical tests.

[c]For life challenges or impairments, each item was analyzed as a binary variable. Percentages for their frequency refer to the prevalence among participants. The chi-square test has been separately calculated for each binary variable.

Initiation of BitHabit App Use

In total, 863 of 1646 (52%) participants initiated BitHabit app use. Based on univariate assessments (Table 1), the initiation of BitHabit app use was positively associated with female gender, less conditioning exercise and light physical activity per week, psychological impairment, cognitive impairment, difficulty in social interaction, a higher number of reported life challenges or impairments, lower satisfaction with physical condition and social relationships, as well as lower general life satisfaction. Furthermore, initiation of BitHabit app use was negatively associated with not reporting any of the listed life challenges or impairments.

Men were also less likely than women to start using the BitHabit app based on the logistic regression model (Table 2; odds ratio [OR] 0.66, 95% CI 0.51 - 0.85). This association was persistent in all the models generated, indicating that gender has a strong independent effect on the initiation of BitHabit app use, even after accounting for other variables in the data.

XSL•FO
RenderX

Table . Logistic regression of determinants for initiation of app use compared with not initiating (N=1614 due to missingness in independent variables).

Covariate	OR[a] (95% CI)	*P* value
Intercept	1.64 (0.89-3.02)	.11
Male (ref: female)	0.66 (0.51-0.85)	<.001
Age (years)	1.00 (0.99-1.01)	.76
Retired (ref: employed or entrepreneur)	0.85 (0.61-1.17)	.32
Student (ref: employed or entrepreneur)	0.81 (0.55-1.19)	.28
Unemployed, full or part-time disability pension, or other situation (ref: employed or entrepreneur)	0.68 (0.48-0.97)	.03
Conditioning exercise 1 - 3 hrs/wk (ref: <1)	1.03 (0.82-1.31)	.79
Conditioning exercise ≥4 hrs/wk (ref: <1)	0.86 (0.64-1.15)	.30
Consumes ≥6 doses of alcohol on one occasion monthly or more frequently (ref: no)	0.91 (0.71-1.16)	.44
General life satisfaction	0.94 (0.89-1.00)	.04
Number of reported life challenges or impairments	1.13 (1.02-1.24)	.02

[a]OR: odds ratio comparing the likelihood of initiating app use relative to not initiating.

When age was assessed as a continuous variable in the logistic regression model, no statistically significant difference was detected (Table 2). When differences were assessed separately according to age groups, we found in univariate assessments that participants aged 30-50 years were somewhat more likely to start using BitHabit, but these differences were also statistically nonsignificant.

Employment status was not associated with the initiation of BitHabit app use in the univariate assessments, but in the logistic regression model, unemployed participants had lower odds (Table 2; OR 0.68, 95% CI 0.48 - 0.97) of initiating app use than employed participants or entrepreneurs. A student status or retirement did not predict the initiation of app use.

Even though physical activity was negatively associated with the initiation of BitHabit app use in the univariate assessments, no such associations were found in the logistic regression model. Furthermore, alcohol use was not associated with the initiation of app use in either univariate assessments or the logistic regression model.

Participants who were less satisfied with some aspect of their life appeared to be more likely to start using the BitHabit app (Table 1 and Table 2). Those with higher general life satisfaction had lower odds of initiating app use (Table 2; OR 0.94, 95% CI 0.89 - 1.00). In addition, a higher number of reported life challenges or impairments was associated with increased odds of initiating app use (Table 2; OR 1.13, 95% CI 1.02 - 1.24).

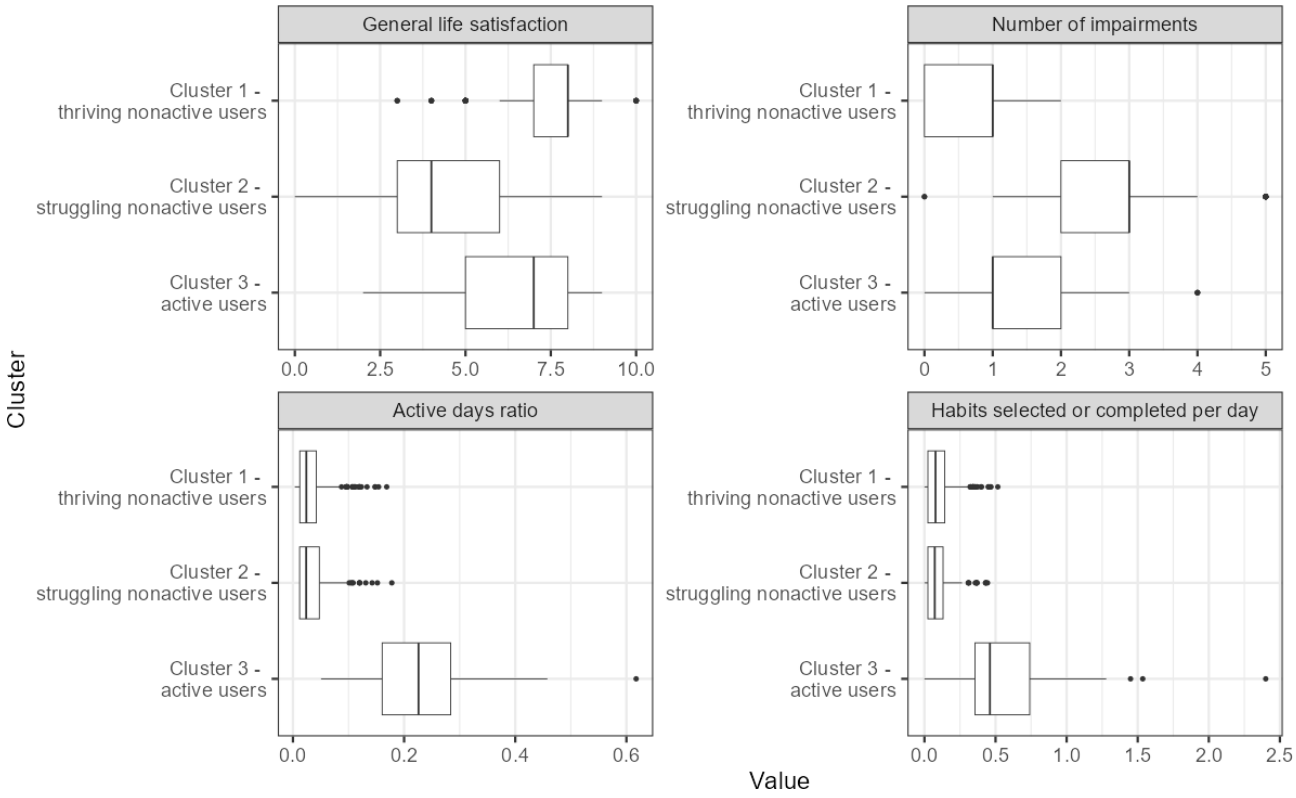
User Archetypes of the BitHabit App

In the clustering analysis for participants who launched the BitHabit app, we recognized 3 different user archetypes based on general life satisfaction, the number of reported impairments or life challenges, the active days ratio (proportion of days with sessions in the app), and the habits ratio (number of habits selected or completed per day). We labeled these archetypes as “thriving nonactive users” (n=474), “struggling nonactive users” (n=198), and “active users” (n=55; Figure 5). In the first cluster, participants had high life satisfaction and a small average number of impairments or life challenges, and they used the app infrequently. In the second cluster, on the other hand, participants had significantly lower life satisfaction and multiple reported impairments on average, but they still used the app infrequently. The struggling nonactive users especially reported financial difficulties (57%, 113/198) and psychological impairment (77%, 153/198; Table 3). In the third cluster, the participants’ life satisfaction and reported number of impairments were at similar levels as in cluster 1, but they used the BitHabit app more frequently. In other words, the participants who actively used the app were thriving at baseline, but among nonactive users, we could recognize participants who were thriving and those who were struggling.

Based on univariate descriptive statistics (Table 3) and the multinomial regression model (Table 4), individual characteristics and behavioral factors were associated with the participants’ user archetype. Compared with active users, thriving nonactive users were more often younger participants, employed or entrepreneurs, engaged in more physical activity, and they reported fewer life challenges. Struggling nonactive users were also more often younger participants, retired, engaged in less physical activity, and consumed alcohol more often. The struggling nonactive users, in particular, reported cognitive impairment, and very few of them reported physical impairment (Table 3).

XSL•FO
RenderX

Figure 5. Boxplot of clustering variables for the different user archetype clusters among participants. The horizontal line represents the median, the box hinges represent the first and third quartiles, and whiskers extend from the hinge to the largest and smallest values, but not further than 1.5 times the IQR from the hinge.



XSL•FO
RenderX

Table . Individual characteristics of participants who initiated use of the BitHabit app (N=727), stratified by user archetype.'

Characteristic	All participants who initiated app use (N=727)	Thriving nonactive users (n=474)	Struggling nonactive users (n=198)	Active users (n=55)	*P* value[a]
Age (years), mean (SD)	46.3 (13.3)	47.2 (13.5)	42.9 (12.5)	51.0 (12.7)	<.001
Missing, n[b]	0	0	0	0	
Gender, n (%)					.11
Male	127 (17.5)	79 (16.7)	41 (20.7)	7 (12.7)	
Female	591 (81.3)	392 (82.7)	153 (77.3)	46 (83.6)	
Other	9 (1.2)	3 (0.6)	4 (2.0)	2 (3.6)	
Missing, n	0	0	0	0	
Employment status, n (%)					<.001
Employed or entrepreneur	479 (65.9)	349 (73.6)	98 (49.5)	32 (58.2)	
Retired	116 (16.0)	74 (15.6)	32 (16.2)	10 (18.2)	
Student	62 (8.5)	34 (7.2)	24 (12.1)	4 (7.3)	
Unemployed, full or part-time disability pension, or other situation	70 (9.6)	17 (3.6)	44 (22.2)	9 (16.4)	
Missing, n	0	0	0	0	
Six or more doses of alcohol on one occasion, n (%)					.007
Less than once a month or never	585 (80.7)	388 (82.2)	147 (74.2)	50 (90.9)	
At least once a month	140 (19.3)	84 (17.8)	51 (25.8)	5 (9.1)	
Missing, n	2	2	0	0	
Sedentary per day, n (%)					.006
5 hours or less	216 (29.8)	143 (30.2)	50 (25.3)	23 (41.8)	
6 to 8 hours	331 (45.6)	228 (48.2)	82 (41.4)	21 (38.2)	
More than 8 hours	179 (24.7)	102 (21.6)	66 (33.3)	11 (20.0)	
Missing, n	1	1	0	0	
Conditioning exercise per week, n (%)					<.001
Less than 1 hour	236 (32.6)	121 (25.7)	98 (49.5)	17 (30.9)	
1 to 3 hours	342 (47.2)	231 (49.0)	79 (39.9)	32 (58.2)	
4 or more hours	146 (20.2)	119 (25.3)	21 (10.6)	6 (10.9)	
Missing, n	3	3	0	0	
Light physical activity per day, n (%)					.06
Less than 2 hours	205 (28.2)	120 (25.4)	70 (35.4)	15 (27.3)	
2 to 3 hours	352 (48.5)	241 (51.0)	88 (44.4)	23 (41.8)	
More than 3 hours	169 (23.3)	112 (23.7)	40 (20.2)	17 (30.9)	
Missing, n	1	1	0	0	
Life challenges or impairments, n (%)[c]					
Physical impairment	227 (31.2)	107 (22.6)	95 (48.0)	25 (45.5)	<.001
Psychological impairment	250 (34.4)	77 (16.2)	153 (77.3)	20 (36.4)	<.001

XSL•FO
RenderX

Characteristic	All participants who initiated app use (N=727)	Thriving nonactive users (n=474)	Struggling nonactive users (n=198)	Active users (n=55)	*P* value[a]
Cognitive impairment	129 (17.7)	42 (8.9)	73 (36.9)	14 (25.5)	<.001
Difficulty in social interaction	156 (21.5)	43 (9.1)	97 (49.0)	16 (29.1)	<.001
Financial challenges	171 (23.5)	45 (9.5)	113 (57.1)	13 (23.6)	<.001
None of these	229 (31.5)	213 (44.9)	3 (1.5)	13 (23.6)	<.001
Missing, n	0	0	0	0	
Number of reported life challenges or impairments, mean (SD)	1.28 (1.2)	0.66 (0.7)	2.68 (1.1)	1.60 (1.2)	<.001
Missing, n	0	0	0	0	
Satisfaction with physical condition, mean (SD)	5.01 (2.6)	5.66 (2.3)	3.44 (2.6)	5.13 (2.4)	<.001
Missing, n	1	1	0	0	
Satisfaction with social relationships, mean (SD)	6.11 (2.4)	6.53 (2.1)	5.15 (2.8)	6.00 (2.2)	<.001
Missing, n	1	1	0	0	
General life satisfaction, mean (SD)	6.55 (2.1)	7.49 (1.2)	4.34 (2.0)	6.42 (1.9)	<.001
Missing, n	0	0	0	0	

[a]*P* values were calculated with the Mann-Whitney *U* test for continuous variables and the chi-square test for categorical variables. The Fisher exact test was used when the assumptions of the chi-square test were not met.

[b]The "Missing" row for each variable refers to the number of cases with incomplete data for that variable. These cases were excluded from statistical tests.

[c]For the life challenges or impairments, each item was analyzed as a binary variable, and the percentages for their frequency refer to the prevalence among participants. The chi-square test was calculated separately for each binary variable.

XSL•FO
RenderX

Table . Multinomial logistic regression of predictors for user archetype clusters (N=713 due to missingness in independent variables).

Covariate	Thriving nonactive users vs active users, OR[a] (95% CI)	*P* value	Struggling nonactive users vs active users, OR (95% CI)	*P* value
Intercept	47.55 (10.23-220.92)	<.001	97.08 (19.28-488.87)	<.001
Male (ref: female)	0.91 (0.37-2.22)	.84	1.55 (0.6-4.01)	.36
Age (years)	0.96 (0.94-0.99)	.01	0.93 (0.9-0.96)	<.001
Retired (ref: employed or entrepreneur)	1.38 (0.53-3.56)	.51	4.06 (1.44-11.42)	.008
Student (ref: employed or entrepreneur)	0.71 (0.2-2.57)	.60	0.93 (0.24-3.59)	.92
Unemployed, full or part-time disability pension, or another disability pension situation (ref: employed or entrepreneur)	0.2 (0.08-0.51)	<.001	1.82 (0.74-4.45)	.19
Conditioning exercise 1 - 3 hrs/wk (ref: <1)	0.99 (0.52-1.91)	.99	0.4 (0.2-0.81)	.01
Conditioning exercise ≥4 hrs/wk (ref: <1)	2.71 (1.00-7.32)	.049	0.56 (0.19-1.67)	.30
Consumes ≥six doses of alcohol on one occasion monthly or more frequently (ref: no)	1.94 (0.72-5.21)	.19	3.22 (1.16-8.99)	.03

[a]OR: odds ratio comparing the likelihood of initiating app use relative to not initiating.

Although female gender was associated with a higher probability of initiating app use, gender was not associated with whether participants were thriving or struggling nonactive users compared with active users. Older participants had higher odds of being active users compared with both thriving and struggling nonactive users (Table 4).

Compared with employed participants or entrepreneurs, unemployed participants had lower odds of being thriving nonactive users compared with active users, but there was no significant association with whether participants were struggling nonactive users compared with active users. Retired people had higher odds of being struggling nonactive users compared with active users, but retirement was not associated with being a thriving nonactive user compared with an active user. Student status did not predict the user archetype for any of the clusters (Table 4).

High engagement in conditioning exercise was associated with participants being thriving nonactive users compared with active users. Participants engaging in 4 or more hours of conditioning exercise per week had the highest odds of being thriving nonactive users (Table 4; OR 2.7, 95% CI 1.00 - 7.32). Similarly, participants engaging in 1 - 3 hours of exercise per week compared with less than 1 hour had lower odds of being struggling nonactive users compared with active users (Table 4; OR 0.4, 95% CI 0.2 - 0.81).

Participants consuming 6 or more doses of alcohol on 1 occasion monthly or more frequently compared with those who did not have higher odds of being struggling nonactive users compared with active users. There was no association with participants being thriving nonactive users compared with active users.

Discussion

Principal Findings

Based on this study, individual characteristics, behavioral factors, and the well-being of the participants were associated with the probability of starting to use the well-being–promoting app BitHabit. In particular, we found that initiation of BitHabit app use was positively associated with female gender, being employed, lower general life satisfaction, and a higher number of reported life challenges.

We also identified 3 separate user archetypes based on app usage, life satisfaction, and reported life challenges. Generally, active users of the app were thriving at the baseline, but among nonactive users, we could recognize distinct thriving and struggling user archetypes. The probability of the participants belonging to the different user archetypes was associated with user characteristics. Older participants appeared to be more likely to be active users, but retired participants were more likely to be struggling nonactive users than active users. Unemployed participants were more likely to be active users than thriving nonusers. Furthermore, physically active participants were more likely to be thriving nonactive users or active users, and alcohol consumption appeared to be more common among struggling nonactive users.

XSL•FO
RenderX

Comparison With Previous Work

Based on our results, women were more likely to initiate use of the BitHabit app compared to men. However, we detected no associations between gender and user archetypes. Previous evidence suggests that women are more likely to use digital tools to promote their health and well-being [10,11]. Our results are consistent in this regard for the initiation of app use, but not for actual use of the app. In a previous study assessing the user trajectories of the BitHabit app [14], we also detected no gender differences with regard to activity in using the app. However, the lack of statistical significance in this study could be due to the small number of both men and active users among our participants.

In our study, age was not associated with the probability of initiating use of the BitHabit app, but it was associated with the user archetype. After the initiation of app use, older participants were more likely to be active users compared to both struggling and thriving nonactive users. In other words, older participants with higher life satisfaction were more likely to be active users than to find the app unnecessary and dismiss it. This is partly consistent with existing evidence suggesting that older individuals are likely to initiate use of digital well-being–promoting apps and use them more frequently compared to younger people [11,13].

Regarding social determinants, we found that unemployed participants were less likely to initiate BitHabit app use compared with employed participants. Unemployed participants were more likely to be active users than thriving nonactive users. However, there was no association with whether unemployed participants were more likely to be struggling nonactive users than active users, indicating that it is equally likely for an unemployed participant to belong in either user archetype. Previous evidence on the participant is unclear, but our findings support the assumption that the unemployed are less likely to initiate the use of well-being–promoting digital interventions [13]. However, based on our results, after the initiation of app use, there appears to be some potential for the unemployed to continue using. Unemployed people generally appear to have lower life satisfaction and more life challenges and impairments compared with employed people [43].

Retirement was not associated with the probability of starting to use the BitHabit app, but in the user archetype analysis, retired participants were more likely to be struggling nonactive users than active users. Retired participants were more often aged 65 years or older compared with employed participants, so this finding is inconsistent with the results regarding participant age. Furthermore, no associations were found for students with regard to either the initiation of app use or the user archetype analysis.

General life satisfaction among the study participants, measured with the life satisfaction item of the OECD Better Life Index [33], was 6.6. This is lower than the average of 8.1 in Finland, which ranks highest among the OECD member countries and is much higher than the OECD average of 7.4 in 2018 [44]. Our results indicate that those who initiated app use were generally less satisfied with their life and reported more life challenges and impairments. Based on existing evidence, people with an unhealthy lifestyle and health issues would be more likely to initiate the use of digital health promotion apps [7,13,17], which is consistent with our findings.

However, in the user archetype analysis of our study, we found that the active users of the BitHabit app actually had relatively high life satisfaction compared with both nonuser groups. In other words, while life challenges and lower general life satisfaction appeared to increase the likelihood of trying out the BitHabit app, in the user archetype analysis, we noticed that those who started using the app more actively were initially more satisfied with their lives. Furthermore, among the nonactive users of the BitHabit app, we recognized distinct user archetypes for both thriving and struggling nonactive users. Our findings are to some extent aligned with studies proposing that behavioral intervention programs may primarily appeal to those who already have healthier lifestyles and thus need the interventions the least [11,20].

These findings can be interpreted through the lenses of the HBM [8] and the UTAUT [9]. Those experiencing lower life satisfaction and a higher number of life challenges may have perceived a higher susceptibility to health problems and greater urgency to act [8], which could explain their initial uptake of the app. However, while initial adoption may be driven by a sense of need, continued use requires the users to find the app beneficial and easy to use [9], as well as for them to have the necessary support and infrastructure to continue engaging with the app. It is possible that individuals with higher baseline well-being had more favorable perceptions in these areas, contributing to more consistent use. Sustained engagement may also depend on users' self-efficacy [8], their confidence in their ability to improve their well-being and use the app effectively, which tends to be lower among individuals in more disadvantaged socioeconomic positions [45].

Strengths and Limitations

Our results bring novel insights to the discussion of how people's individual characteristics, behavioral factors, and well-being influence their likelihood of initiating the use of well-being–promoting digital apps and actively using them. The key strengths of our study include the relatively large baseline sample. We also encountered considerable interest in our digital well-being survey and report, as well as the BitHabit app.

However, due to resource and timing restrictions and the fact that our study was planned to act as a pilot study for a subsequent randomized controlled trial, our research design was not rigorous, and we allowed the sample to self-select without definitive inclusion criteria. We can assume that the sample is biased, and we are aware that the results of the study can at most be generalized to people who are already interested in their well-being and inclined to promote it.

We compared our sample with national statistics of the Finnish general adult population and found that older individuals, and especially men, were underrepresented in our self-selected sample. Most of the sample consisted of women, which needs to be considered when interpreting the results. Based on our previous results, men are more difficult to recruit in lifestyle intervention studies compared with women [19]. In addition, the proportion of unemployed participants aged 15-64 years in

XSL•FO
RenderX

our sample was approximately 6% (91/1448), which is fairly close to the national unemployment rate among the economically active population aged 15-64 years of 7% (200,000/2,732,000), and the proportion unemployed among the total population aged 15-64 years of 6% (200,000/3,442,000) in Finland in 2023 [46]. In future studies, we could address sample selection bias through randomized or weighted sampling, as well as by marketing the study in a way that is more appealing to underrepresented population groups.

Only a small number of participants finally ended up actively using the app, which limited the sample size available for the user archetype analysis. We considered this limitation when interpreting the results of the analysis. With a larger sample size, we could have increased the number of analyzed user archetypes, and with more active users, we would have been able to conduct more detailed analysis of the profiles of both active and nonactive users in the current analysis.

Another limitation of this study is the lack of comprehensive and validated measures in the digital survey. This study was designed as a pilot study for a subsequent randomized controlled trial. To maximize the sample size in our self-selective study, we aimed to make the digital survey as short as possible and only included single-item questions. Validated multi-item questionnaires are generally preferred in most study fields, since the measurement reliability and validity of a single-item question is difficult to estimate [47]. However, the items used in this study have been widely used in previous studies or were relatively unambiguous, such as the self-reported life challenges or health behavior questions.

Conclusions

Our results indicate that while lower general life satisfaction and less favorable health behavior appeared to increase the likelihood of trying the well-being–promoting digital app BitHabit, those who started using the app more actively were initially more satisfied with their lives. Our findings also suggest that among nonactive users, there are recognizable user profiles distinguishing thriving and struggling nonactive users. Individual characteristics such as gender, age, and employment appear to be variously associated with both initiating the well-being–promoting app and actively using the app. Our findings need to be verified in a future randomized controlled trial, preferably weighted by gender. In the future, it would also be relevant to assess the effectiveness of well-being–promoting digital apps for different user profiles and target groups.

Acknowledgments

This work was financially supported by the European Social Fund (grant no 2014/11114/09020101/2020) as a part of the Feeling Good North Savo program and by the Research Council of Finland (363107) as a part of the PREWELL project. During the writing and revision process of the manuscript, we used ChatGPT, a large language model developed by OpenAI, for language refinement.

Data Availability

The datasets generated and analyzed in the current study are available from the corresponding author upon reasonable request for research collaboration.

Authors' Contributions

HR contributed to conceptualization, methodology, visualization, formal analysis, and wrote the original draft. HL contributed to project administration and TMO performed supervision.

Conflicts of Interest

JS is associated with Wellpro Impact Solutions. However, the research was conducted independently, and the authors have no other conflicts of interest to declare.

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[PDF File, 106 KB - mhealth_v13i1e68982_app1.pdf]

Checklist 2
mERA (mHealth Evidence Reporting and Assessment) checklist.
[PDF File, 134 KB - mhealth_v13i1e68982_app2.pdf]

References

1. Mental health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response [accessed 2025-05-15]

XSL•FO
RenderX

2. Lakka TA, Aittola K, Järvelä-Reijonen E, et al. Real-world effectiveness of digital and group-based lifestyle interventions as compared with usual care to reduce type 2 diabetes risk - A stop diabetes pragmatic randomised trial. Lancet Reg Health Eur 2023 Jan;24(24):100527. [doi: 10.1016/j.lanepe.2022.100527] [Medline: 36620354]
3. Lee M, Lee H, Kim Y, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018 Dec 13;15(12):2838. [doi: 10.3390/ijerph15122838] [Medline: 30551555]
4. Dale H, Brassington L, King K. The impact of healthy lifestyle interventions on mental health and wellbeing: a systematic review. Ment Health Rev J 2014 Mar 5;19(1):1-26. [doi: 10.1108/MHRJ-05-2013-0016]
5. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
6. Mann T, de Ridder D, Fujita K. Self-regulation of health behavior: social psychological approaches to goal setting and goal striving. Health Psychol 2013 May;32(5):487-498. [doi: 10.1037/a0028533] [Medline: 23646832]
7. Bregenzer A, Wagner-Hartl V, Jiménez P. Who uses apps in health promotion? A target group analysis of leaders. Health Informatics J 2019 Sep;25(3):1038-1052. [doi: 10.1177/1460458217738121] [Medline: 29113522]
8. Champion VL, Skinner CS. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice, 5th edition: Jossey-Bass; 2015:75-94.
9. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
10. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020]
11. Reinwand DA, Schulz DN, Crutzen R, Kremers SP, de Vries H. Who follows eHealth interventions as recommended? A study of participants' personal characteristics from the experimental arm of a randomized controlled trial. J Med Internet Res 2015 May 11;17(5):e115. [doi: 10.2196/jmir.3932] [Medline: 25963607]
12. Mattila E, Lappalainen R, Välkkynen P, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90. [doi: 10.2196/mhealth.5241] [Medline: 27468653]
13. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012 Mar 1;14(2):e40. [doi: 10.2196/jmir.1939] [Medline: 22382037]
14. Lavikainen P, Mattila E, Absetz P, et al. Digitally supported lifestyle intervention to prevent type 2 diabetes through healthy habits: secondary analysis of long-term user engagement trajectories in a randomized controlled trial. J Med Internet Res 2022 Feb 24;24(2):e31530. [doi: 10.2196/31530] [Medline: 35200147]
15. Edney S, Ryan JC, Olds T, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645. [doi: 10.2196/14645] [Medline: 31774402]
16. Ye J, Ma Q. The effects and patterns among mobile health, social determinants, and physical activity: a nationally representative cross-sectional study. AMIA Jt Summits Transl Sci Proc 2021;2021:653-662. [Medline: 34457181]
17. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening via a smartphone app: cross-country user characteristics and feasibility. J Am Med Inform Assoc 2015 Jan;22(1):29-34. [doi: 10.1136/amiajnl-2014-002840] [Medline: 25326599]
18. Correa T, Hinsley AW, de Zúñiga HG. Who interacts on the Web?: The intersection of users' personality and social media use. Comput Human Behav 2010 Mar;26(2):247-253. [doi: 10.1016/j.chb.2009.09.003]
19. Peels DA, Bolman C, Golsteijn RHJ, et al. Differences in reach and attrition between web-based and print-delivered tailored interventions among adults over 50 years of age: clustered randomized trial. J Med Internet Res 2012 Dec 17;14(6):e179. [doi: 10.2196/jmir.2229] [Medline: 23246790]
20. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007 Jan 22;9(1):e1. [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
21. This is what it's all about. Mieliteko. URL: https://www.mielitekomedia.fi/fi/hankekuvaus [accessed 2024-10-31]
22. von Elm E, Altman DG, Egger M, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808. [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
23. Agarwal S, LeFevre AE, Lee J, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
24. Wellpro impact solutions. BitHabit. URL: https://wellpro.fi [accessed 2024-10-31]
25. Pihlajamäki J, Männikkö R, Tilles-Tirkkonen T, et al. Digitally supported program for type 2 diabetes risk identification and risk reduction in real-world setting: protocol for the StopDia model and randomized controlled trial. BMC Public Health 2019 Mar 1;19(1):255. [doi: 10.1186/s12889-019-6574-y] [Medline: 30823909]

XSL•FO
RenderX

26. Wood W, Neal DT. Healthy through habit: interventions for initiating & maintaining health behavior change. Behavioral Science & Policy 2016 Apr;2(1):71-83. [doi: 10.1177/237946151600200109]
27. Wood W, Rünger D. Psychology of habit. Annu Rev Psychol 2016;67(1):289-314. [doi: 10.1146/annurev-psych-122414-033417] [Medline: 26361052]
28. Lally P, Gardner B. Promoting habit formation. Health Psychol Rev 2013 May;7(sup1):S137-S158. [doi: 10.1080/17437199.2011.603640]
29. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
30. Borodulin K, Kärki A, Laatikainen T, Peltonen M, Luoto R. Daily sedentary time and risk of cardiovascular disease: the national FINRISK 2002 study. J Phys Act Health 2015 Jul;12(7):904-908. [doi: 10.1123/jpah.2013-0364] [Medline: 25153761]
31. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
32. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
33. How's Life?: Measuring Well-Being: OECD Publishing; 2011. [doi: 10.1787/9789264121164-en]
34. RStudio: integrated development environment for r. Posit Software, PBC formerly RStudio, PBC. 2023. URL: http://www.posit.co [accessed 2025-08-07]
35. The r project for statistical computing. The R Foundation. 2023. URL: https://www.R-project.org [accessed 2025-08-03]
36. Aziz M, Erbad A, Belhaouari SB, Almourad MB, Altuwairiqi M, Ali R. Who uses mHealth apps? Identifying user archetypes of mHealth apps. Digit Health 2023;9(9):20552076231152175. [doi: 10.1177/20552076231152175] [Medline: 36714545]
37. Hartigan JA, Wong MA. Algorithm AS 136: a k-means clustering algorithm. Appl Stat 1979;28(1):100. [doi: 10.2307/2346830]
38. Kassambara A, Mundt F. Factoextra: extract and visualize the results of multivariate data analyses. CRAN. 2020. URL: https://CRAN.R-project.org/package=factoextra [accessed 2025-08-03]
39. Wickham H. Ggplot2. ggplot2. 2016. URL: https://ggplot2.tidyverse.org [accessed 2025-08-03]
40. Guidelines for ethical review in human sciences. Finnish National Board on Research Integrity TENK. 2021. URL: https://tenk.fi/en/advice-and-materials/guidelines-ethical-review-human-sciences [accessed 2025-05-16]
41. Finnish Medical Research Act (488/1999). Ministry of Social Affairs and Health. 1999. URL: https://finlex.fi/en/legislation/translations/1999/eng/488 [accessed 2025-06-27]
42. Regulation (EU) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation). OJ L 119. 2016 URL: http://data.europa.eu/eli/reg/2016/679/oj [accessed 2025-08-07]
43. Society at a Glance 2019: OECD Social Indicators: OECD Publishing; 2019. [doi: 10.1787/soc_glance-2019-en]
44. How's Life? 2020: Measuring Well-Being: OECD Publishing; 2020. [doi: 10.1787/9870c393-en]
45. Mäki-Opas T, Pieper R, Vaarama M. Exploring the capability approach to quality of life in disadvantaged population groups. Sci Rep 2022 Sep 15;12(1):15248. [doi: 10.1038/s41598-022-18877-3] [Medline: 36109542]
46. Labour force survey. Statistics Finland. 2024. URL: https://stat.fi/en/statistics/tyti [accessed 2025-02-15]
47. Schmidt H. The single-item questionnaire. Health Prof Educ 2018 Mar;4(1):1-2. [doi: 10.1016/j.hpe.2018.02.001]

Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test-Concise
HBM: Health Belief Model
mERA: mHealth Evidence Reporting and Assessment
mHealth: mobile health
OECD: Organisation for Economic Co-operation and Development
OR: odds ratio
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TENK: Finnish National Board on Research Integrity
UTAUT: Unified Theory of Acceptance and Use of Technology
WHO: World Health Organization

XSL•FO
RenderX

Edited by L Buis; submitted 17.12.24; peer-reviewed by B Cahill, P Liu; revised version received 19.05.25; accepted 22.06.25; published 18.08.25.

Please cite as:
Rekola H, Tolmunen T, Mattila E, Strömmer J, Lakka TA, Länsimies H, Mäki-Opas T
User Archetypes of a Well-Being–Promoting Mobile App Among Adults: Cross-Sectional Study and Cluster Analysis of Usage Patterns
JMIR Mhealth Uhealth 2025;13:e68982
URL: https://mhealth.jmir.org/2025/1/e68982
doi:10.2196/68982

© Hanna Rekola, Tommi Tolmunen, Elina Mattila, Juho Strömmer, Timo A Lakka, Helena Länsimies, Tomi Mäki-Opas. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Social Media to Recruit a Diverse Sample of Participants for a Mobile Health (mHealth) Intervention to Increase Physical Activity: Exploratory Study

Laura Pathak[1], BA, MSW; Rosa Hernandez-Ramos[2], BA, MA; Karina Rosales[1], BA; Jose Miramontes-Gomez[3], BA; Faviola Garcia[3], BA; Vivian Yip[4], BA; Suchitra Sudarshan[5], BA; Anupama Gunshekar Cemballi[3], BA, MA; Courtney Lyles[6], PhD; Adrian Aguilera[1,7], PhD

[1]School of Social Welfare, University of California, Berkeley, 120 Haviland Hall, Berkeley, CA, United States

[2]Department of Psychological Science, University of California, Irvine, Irvine, CA, United States

[3]Department of the History of Medicine, Johns Hopkins University, Baltimore, MD, United States

[4]Whole Person Integrated Care, San Francisco Department of Public Health, San Francisco, CA, United States

[5]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, United States

[6]Center for Healthcare Research and Policy, University of California Davis School of Medicine, Davis, CA, United States

[7]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Laura Pathak, BA, MSW
School of Social Welfare, University of California, Berkeley, 120 Haviland Hall, Berkeley, CA, United States

Abstract

Background: Recruitment of demographically diverse samples in clinical research is often challenging and even more so during the COVID-19 pandemic when traditional in-person recruitment methods could not be implemented. Social media platforms offer an alternative approach for recruiting diverse samples of participants for clinical trials, including those testing digital health interventions. This approach allowed for a quicker recruitment process without the physical constraints associated with traditional in-person methods.

Objective: This study aimed to detail the online and social media campaigns used to recruit participants for "Diabetes and Mental Health Adaptive Notification Tracking and Evaluation" (DIAMANTE), a randomized controlled trial testing a smartphone-based intervention (a text messaging system that uses machine learning to personalize content) to increase physical activity for patients with diabetes and depression. In describing the recruitment process, we seek to offer insights to the research community on recruitment through online and social media advertisements for diverse communities.

Methods: This study sought to recruit demographically diverse individuals in the United States through social media, including paid advertisements on Craigslist and Facebook (Meta). For the DIAMANTE project recruitment, we created 18 personas that mapped into the population's target demographics using a user-centered design methodology. We deployed targeted English and Spanish ads on Craigslist and Facebook in 78 cities based on county-level demographics and diabetes prevalence data to target diverse individuals aged 18 - 75 years old, who had been diagnosed with diabetes and had documented depressive symptoms.

Results: A total of 1379 individuals completed the study's initial screening survey. Of those, 71 respondents on Facebook and 508 on Craigslist were interested in enrolling in our study. In total, 26 out of 58 (45%) eligible respondents from Facebook and 50 out of 235 (21.3%) eligible respondents from Craigslist were eventually recruited in the randomized controlled trial. In all, both platforms showed poor performance in recruiting Spanish speakers, with Facebook advertisements accounting for 0 and Craigslist for 4 (5.3%) of such participants. When it came to English speakers, Craigslist proved to be the better performing platform compared to Facebook, both in terms of reach (579 vs 71) and cost-effectiveness (US $67.61 in average cost per recruited participant vs US $80.16). While Craigslist ads reached more people, resulting in more completed screening surveys than Facebook ads, there was a higher number of ineligible and incomplete enrollment from Craigslist compared with Facebook, leading to a relatively lower conversion rate (9.4% vs 36.6%). Importantly, participants recruited through Craigslist were more ethnically and racially diverse than those recruited from Facebook.

Conclusions: Results from this study revealed that it is possible to recruit diverse sample sets using social media and online advertisements. However, despite targeted recruitment efforts, social media recruitment of Spanish speakers proved especially challenging and costly. Further research is needed to determine systematic, online methods for recruiting marginalized communities.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-034723

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e56329) doi:10.2196/56329

KEYWORDS

mHealth; social media; research subject recruitment; Spanish speakers; Spanish; physical activity; clinical research; social media platform; smartphone; text messaging; machine learning; diabetes; depression; Facebook; depressive symptoms; screening survey; online platform; cost-effectiveness; mobile health

Introduction

Health Disparities and mHealth

Research on the use of information technologies, particularly mobile health (mHealth), as a service delivery paradigm for health care systems has proliferated in the last decade due to the potential of these tools to help bridge gaps in care for underserved individuals [1,2]. Coupled with early evidence supporting its efficacy [3-5], the undisputed ubiquity of mobile technology, and strong interest among patients [6], there is a clear potential for harnessing mHealth to widen health care access. However, marginalized communities, particularly African Americans and Latinos, are underrepresented in mHealth research [7,8], and there are known barriers to recruiting samples of historically underrepresented communities in studies [9]. As a result, there is limited science to support mHealth treatment recommendations for these communities, which arguably worsens existing racial and ethnic health disparities.

Social Media Recruitment for mHealth Research

Recruitment challenges with hard-to-reach populations in mHealth clinical trials were aggravated in the wake of the COVID-19 pandemic when reliance on some traditional recruitment methods, such as using flyers, word-of-mouth, or in-person recruitment by a health care provider, was either not feasible or fully implementable. Using social media platforms such as Facebook (Meta), Craigslist, Google Ad Words, Instagram (Meta), and Twitter (subsequently rebranded X) for recruitment offered a potentially faster, more affordable, and efficient alternative to traditional recruitment methods of national samples. Online recruitment was emphasized in multiple mHealth research areas such as depression [8,10-14], anxiety [10], smoking and tobacco use [15-17], diabetes [18], cancer survivor support [19], and healthy weight gain in pregnancy [20].

While there are a few promising studies leveraging social media to reach historically marginalized communities for surveys or broader outreach [10,21,22], the evidence on recruiting marginalized populations through social media for mHealth randomized controlled trials (RCTs) is more limited. Pratap et al [8] conducted a fully remotely run RCT testing the feasibility of using mental health apps to assess and treat moderate depressive symptoms among Latino adults. They found that it was substantially more expensive and labor-intensive to recruit Spanish-speaking participants through social media relative to English-speaking Latino and other groups, with participant acquisition costs differing dramatically between Spanish (US $31 per enrolled participant) and English speakers (US $1.49 per enrolled participant). Bunge et al [17] reported successful Facebook recruitment of Spanish-speaking smokers through Facebook in a clinical trial of a smoking cessation web app. However, the authors also described concerns about the high costs of recruiting such participants compared with English speakers. These data suggest that, though feasible, recruiting Spanish-speaking respondents for mHealth RCTs through social media can be expensive. Therefore, testing more targeted recruitment strategies on social media platforms has the potential to achieve better affordability.

Furthermore, there are also challenges when it comes to understanding effective social media advertising content for research recruitment. For example, although researchers have reported on Facebook ad development practices for health research recruitment [17,23], the platform's frequently changing ad management interface, abstruse content posting rules, and proprietary algorithm make it challenging to standardize best practices for advertising content tailored to specific populations. However, an experimental study by Choi et al [24] found that recruitment rates, engagement, and participant characteristics varied as a function of changes in Facebook advertisement content. The authors thus emphasized the importance of the content of social media advertisements being carefully crafted for the target population [24]. The literature contains fewer reports regarding Craigslist ad design, possibly due to the platform's less restricted options for content display, which resemble newspaper advertisements with a headline, body text, a map view, and optional picture(s).

Given the state of the evidence, it is clear that understanding the best methods for effective recruitment of hard-to-reach populations in mHealth studies is paramount to advancing digital health science. In this paper, we aim to detail the social media campaigns we used to recruit participants for "Diabetes and Mental Health Adaptive Notification Tracking and Evaluation" (DIAMANTE), an RCT testing a smartphone-based intervention (described below) to increase physical activity for individuals with diabetes and depression. Specifically, we aim to evaluate the performance of social media platforms for research recruitment and the associated costs by population and platform. In describing the process and reporting our recruitment results, we seek to offer the clinical research community insights on best practices for recruiting marginalized communities through social media advertisements.

Methods

Study Overview

The recruitment methods and data reported herein are part of the DIAMANTE RCT (NCT 03490253), which seeks to evaluate the effectiveness of a mobile app that uses reinforcement learning algorithms to predict which categories of health text-messages are most effective in increasing a participant's physical activity based on previous physical activity data

XSL•FO
RenderX

(number of steps) and other demographic and contextual variables [25]. Importantly, our recruitment methods were informed by lessons learned in using a similar recruitment approach in another study aiming to deliver a text-messaging program to individuals struggling with social distancing during the COVID-19 pandemic (StayWell at Home) [26].

This study sought to recruit individuals in the United States who were low-income, spoke English or Spanish, and were from diverse racial and ethnic backgrounds through paid advertisements on Craigslist and Facebook. We deployed targeted English and Spanish ads on Craigslist and Facebook in 78 cities based on county-level demographic and diabetes prevalence data (detailed below). Craigslist is a marketplace website used for posting and searching free and paid local classified ads in different categories, including housing rentals, job postings, for-sale items, and community events. Facebook is a social networking website used primarily for connecting with friends and colleagues through direct messaging and content sharing. We chose these 2 platforms due to their wide reach (Craigslist and Facebook have 250 million and 2.9 billion monthly active users, respectively) and based on the limited literature suggesting they were effective in recruiting diverse samples for digital health studies [8,17].

Participants

Eligible participants for the DIAMANTE trial included individuals aged 18 - 75 years who had been diagnosed with diabetes and had documented depressive symptoms (score>5 on the eight-item Patient Health Questionnaire depression scale [PHQ-8]). Before and early on in the COVID-19 pandemic, we had recruited 15 (5 English and 10 Spanish) participants for the trial using traditional methods (ie, health care provider referrals and flyers) in a large safety-net health care system located in the Western United States. Starting in the summer of 2020, we pivoted our approach to using online and social media platforms as a means to recruit DIAMANTE participants from the broader community due to shelter-in-place and social distancing protocols. We later expanded to a hybrid approach and used traditional (implemented remotely) and social media recruitment methods [27].

To ensure that social media recruits were similar to our clinical participants, the research team designed social media ads to target US residents with both type 2 diabetes and clinical depression or depressive symptoms, who were low-income, English or Spanish-speaking, and from diverse racial and ethnic backgrounds. We decided to focus on major metropolitan regions with higher rates of type 2 diabetes prevalence and higher concentrations of populations demographically similar to our clinical participants. We analyzed Centers for Disease Control and Prevention county-level data for type 2 diabetes prevalence throughout the United States and Census Bureau county-level demographic data to determine which metropolitan regions were most similar to the area in the Western US state where the DIAMANTE trial was originally set in. If a metropolitan region was not demographically similar to our study area, we placed a greater emphasis on regions with higher rates of type 2 diabetes prevalence [28].

Development of Recruitment Ads

Facebook

In June 2020, we created 18 personas that mapped into our target populations' demographics, which comprised of different combinations of sex (male, female), language (Spanish, English), age (younger, older), and race or ethnicity (African American, White, Hispanic or Latino, Asian) categories. A persona is a fictional characterization of a user that includes specific characteristics and demographics found in the target population [29]. A design method commonly used under the User Centered Design framework and more broadly in the fields of human factors engineering and human-computer interaction, personas help think through the early design aspects of a new product that aims to be culturally appropriate, relevant, and user-friendly to the target user or user base [30]. To create the personas (refer to Figure 1 for an example), the research team harnessed personal and professional experiences working with our target populations and subpopulations to write up brief, fictional descriptions of demographic information, interests, motivations, and environmental factors. We paired the descriptions with stock pictures of people who embodied the physical characteristics of each persona.

XSL•FO
RenderX

Figure 1. Sample persona of an older Latino man.

Martin Jimenez

Demographics:

- 62 years old.
- Completed some high school in Nicaragua.
- Migrated to the United States in the late 1990s.
- Disabled from a traumatic brain injury after experiencing a workplace accident 15 years ago.
- Lives with his wife of 30 years in a house in Miami.
- Has one adult son who lives in New York.

Interests: attending church, reading the Bible, listening to music, smoking.

Motivations: his family, religion, and cultural background.

Environment:

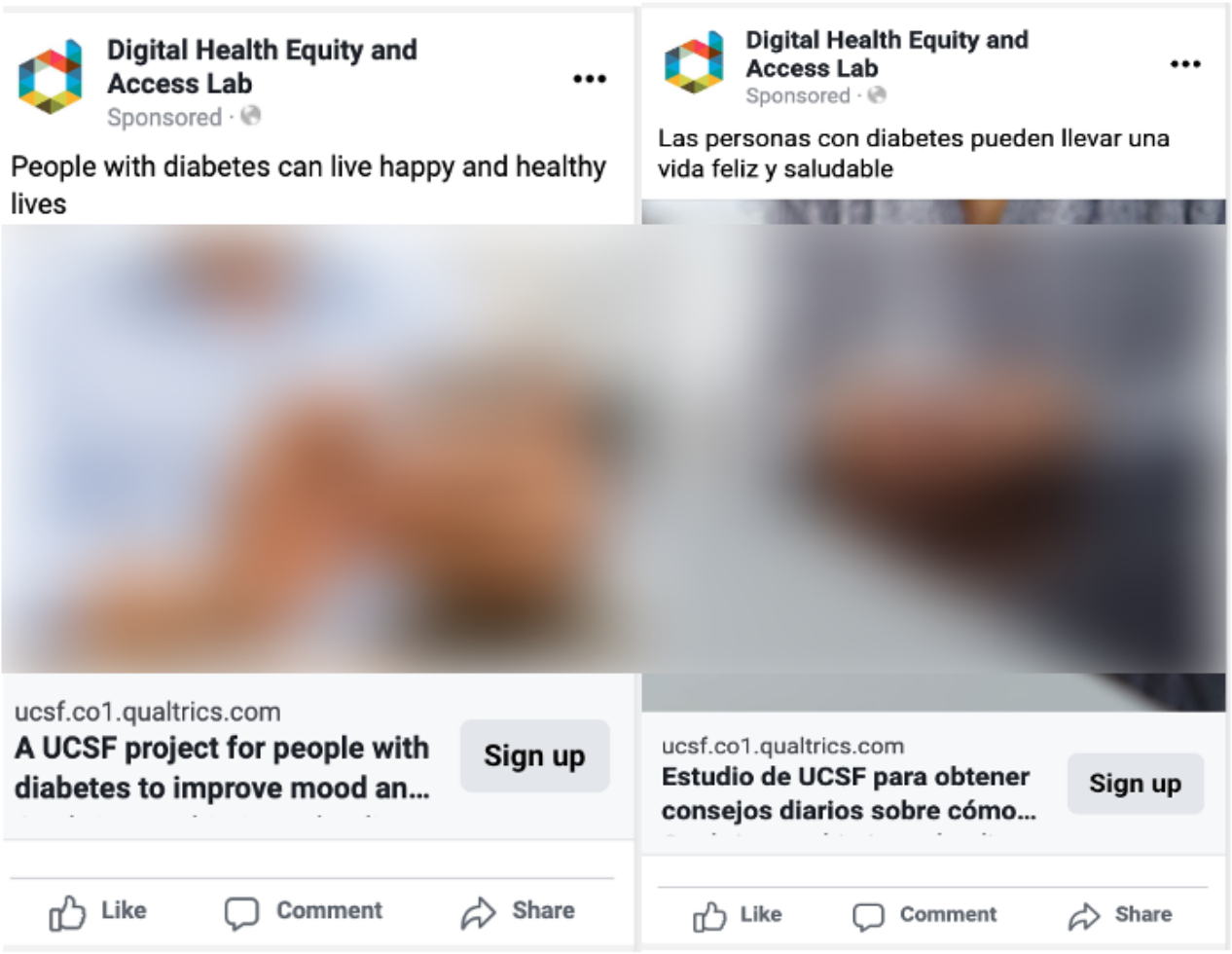
- Long history of depression: diagnosed with major depressive disorder after his tramautic brain injury and has experienced several major depressive episodes ever since.
- Diagnosed with Type 2 diabetes 5 years ago. Has tried to improve his eating behaviors and is mostly following a healthy diet. However, he struggles with finding motivation to exercise regularly.
- Enjoys Miami's weather but doesn't have many friends to go out with, so he and his wife stay home most of the time.
- Likes to login to Facebook to keep up with old friends in Nicaragua and read national and international news.

In August 2020, we used the 18 personas as inspiration to craft 60 Facebook ads following the COM-B (Capability, Opportunity, Motivation and Behavior) model [31] as a framework for content development (refer to Multimedia Appendix 1 for a table of ad copy by persona). The COM-B model posits that behavior change entails an interacting system consisting of 3 constructs: capability, opportunity, and motivation. Effective behavior change occurs when one or more of these components are changed as it rearranges the system and minimizes the risk of default [30]. Hence, each persona had a corresponding set of 3 ads: an ad with a "capability" orientation, another with an "opportunity" orientation, and a third with a "motivation" orientation. The "capability" construct refers to the skills and knowledge necessary for change; in this case, we framed this subset of ads based on the positive physical and social outcomes of exercise. The "opportunity" construct taps into the socio environmental factors that enable change, so we created this subset of ads with cues for exercise opportunities or suggestions on minimizing barriers to exercise. Finally, as the "motivation" construct refers to the internal cognitive processes that direct behavior, this group of ads contained messaging to boost self-efficacy, including self-management skills and confidence to exercise.

In addition, we created 6 broad ads (targeting a broad audience, not specific to any demographic group besides diabetes and depression diagnoses). Three of those 6 ads had vague wording around "well-being," and 3 had more concrete wording on diabetes and depression (again, each type of wording adapted to the 3 COM-B theoretical constructs). Because Facebook requires that ads not contain "content that asserts or implies personal attributes" [32], we used indirect, as opposed to direct, phrasing ("people with diabetes" vs "do you have diabetes?"). Furthermore, we made references to "low mood" rather than "depression," given the stigmatizing nature of this condition, and, based on findings in previous literature [17], framed the content to induce self-efficacy for behavior change, as opposed to fear. In general, the ad copy consisted of a brief sentence or phrase written in terms of the corresponding COM-B construct and persona, a brief study description, a link or button to take the screening survey, and a nondescript stock picture of a person holding a smartphone (refer to Figure 2 for Facebook ad examples).

XSL•FO
RenderX

Figure 2. Example ads on Facebook. UCSF: University of California, San Francisco.



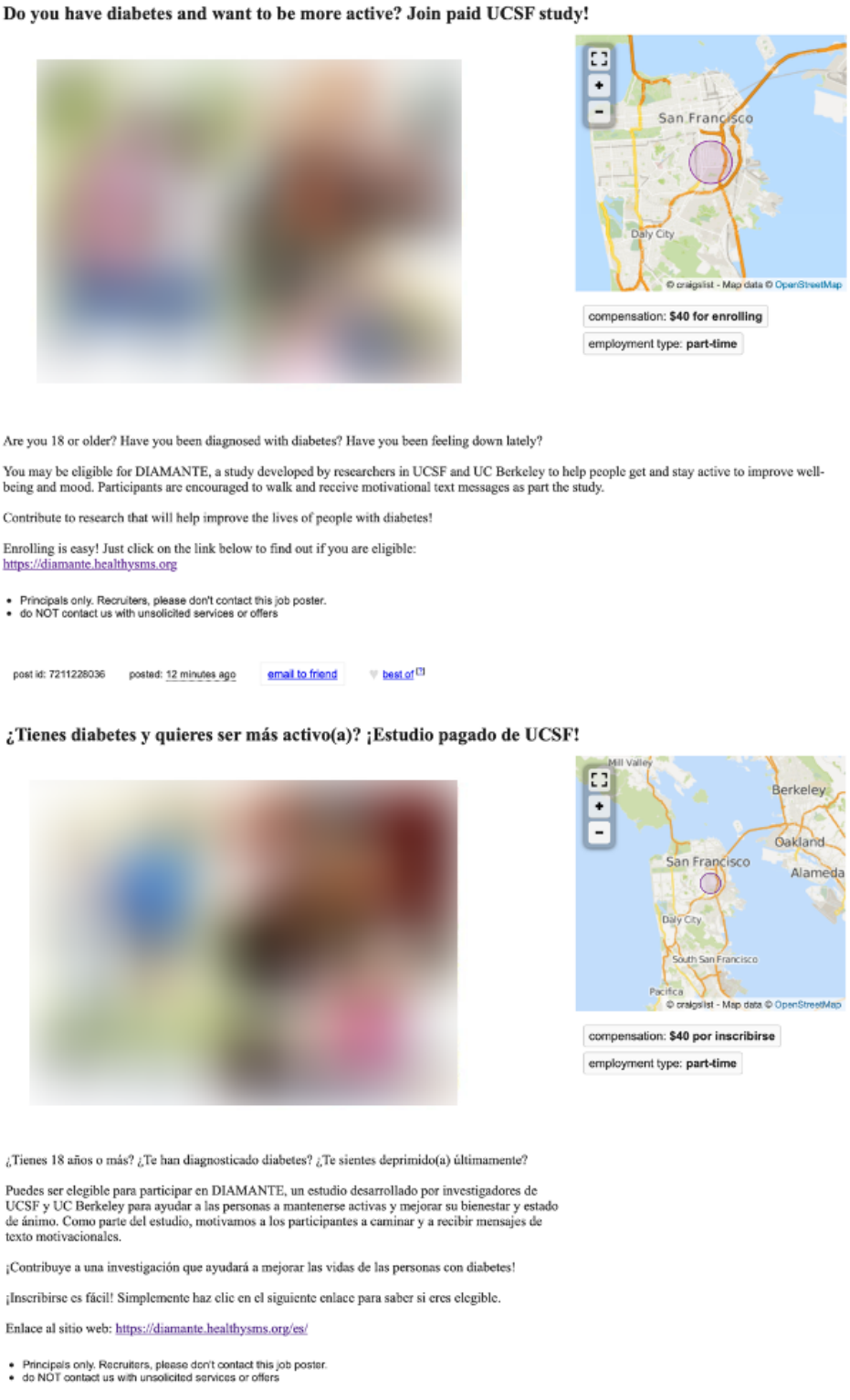
Although we created the copy for 66 Facebook ads based on our 18 personas and general wording around health conditions or well-being, given budget constraints, we chose a subset of 7 ads for posting on Facebook (Multimedia Appendix 1). In this subset we chose to include some of the broad ads (4) in order to achieve high reach and some of the persona ads (3) that best encapsulated the demographic characteristics of the populations we were targeting (ie, older Latina females, older Latino males, older Black females, and older Black males). We then randomly selected the theoretical construct for each chosen ad category.

Craigslist

We developed Craigslist ad content concurrently with Facebook ads using similar wording and stock photographs. Due to budget constraints and the relatively inflexible and time-consuming ad management interface in Craigslist, we limited this recruitment method to only 2 ad versions, 1 English and 1 Spanish (Figure 3).

XSL•FO
RenderX

Figure 3. Craigslist posts. DIAMANTE: Diabetes and Mental Health Adaptive Notification Tracking and Evaluation; UC Berkeley: University of California, Berkeley; UCSF: University of California, San Francisco.



Recruitment

Procedures

We recruited US-based individuals using advertisements on Craigslist and Facebook. To identify eligible individuals, we included a Qualtrics (Silver Lake) link on the social media ads to verify inclusion criteria and to determine human identity using a built-in CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart). If eligible, research staff emailed individuals a one-time-use personalized Qualtrics link to a baseline survey. At baseline, participants provided electronic informed consent and filled out a battery of assessments, including demographics and measures of interest. We sent video instructions to participants who completed the baseline survey on downloading the DIAMANTE mobile app based on their mobile operating system (ie, Android

XSL•FO
RenderX

or iOS). Upon app download, participants were automatically enrolled into the DIAMANTE text-messaging program. Because the DIAMANTE app monitored how many steps participants took each day, throughout the study, we emailed participants if the app did not capture their steps to address any potential technical difficulties.

Facebook

We deployed Facebook ads for approximately 10 weeks (from January 21, 2021, to April 6, 2021). We initially posted 2 English and 2 Spanish ads in the United States with the same set of audience demographics, interests, and education level selected (Textbox 1), except for the ad language ("English [United States]" was set for English ads and "Spanish [All]" for Spanish ads). We saved the English and Spanish demographics sets as reusable "audience" templates for future ads. Based on the selected audience options, Facebook estimated the potential "reaches" for the ads to be 140 million people (English) and 25 million people (Spanish). We set all ads to a daily budget of US $10. Two weeks later, we posted 2 additional English ads. We also activated the "Detailed Targeting Expansion Feature," which, according to Facebook, "expand(s) your detailed targeting to reach more people when it is likely to improve performance." After 2 weeks, we discontinued the Spanish ads due to the much higher price to post ads with lower rates of participants completing our study eligibility screener (and no participants enrolled). Five weeks later, we increased the English daily ad budget to US $25 due to a slow recruitment flow. We also shifted to a targeted approach by narrowing the geographical area from the United States to the San Francisco Bay Area and Los Angeles regions.

Textbox 1. Audience selection for Facebook ads.

Demographics and occupation:

- Cleaning and maintenance
- Construction and extraction
- Healthcare and medical services
- Installation and repair
- Food and restaurants
- Parents (all)
- Ages: 18 - 65 years or older
- Genders (all)

Interests:

- Community mental health
- Diabetes health
- Mental health awareness month
- Diabetes wellbeing
- Diabetes daily
- Diabetes mellitus II awareness

Education:

- Some high school
- High school graduate
- Some college
- Unspecified

Craigslist

We deployed Craigslist ads for 32 weeks (from October 9, 2020, to May 20, 2021). Throughout this period, a research team member monitored the ads daily to ensure they were still live (ie, had not expired, been flagged, or removed). Craigslist requires research studies to be posted as paid employment adverts (either in the "Jobs" or "Gigs" sections), so we posted paid ads with a link to the DIAMANTE study landing page (to find out more information about and fill out an eligibility survey for the study) in the "[Location]>Jobs>Et cetera" section. We used a graded approach wherein we started with a set of ads in a specific region, followed by various cities in a specific state, and finally, simultaneous deployment in several cities across multiple states, prioritizing denser areas with large populations of Spanish speakers.

Data Analysis

We used descriptive statistics to summarize participants' demographic characteristics for both recruiting methods. For ad performance metrics, we tabulated the duration, total costs, and number of participants recruited through each recruitment

XSL•FO
RenderX

platform. We then calculated the cost for enrolled participants for each platform by dividing the total US dollar amount spent in each platform by the number of participants recruited (average cost per participant), number of English speakers (cost per English speaking participant) and number of Spanish speakers (cost per Spanish speaking participant). Cost figures only include direct costs of advertising fees on Facebook and Craigslist; they do not include staff time to develop, implement, and monitor the ads. We compared participant demographics and ad performance metrics between platforms using Pearson chi-square test or Fisher exact tests for categorical variables and one-way analysis of variance for the continuous variable (age). Analyses were carried out on Stata/SE (version 16.1, StataCorp [33].

Ethical Considerations

The study was approved by the Institutional Review Board Committee at the University of California, San Francisco (IRB: 17 - 22608). Participants who screened as eligible provided electronic informed consent in their preferred language, either English or Spanish, were enrolled in the study, and informed of their right to opt out of the study at any time. The demographic data analyzed for this paper are aggregated and thus deidentified. To ensure participant privacy and confidentiality, only institutional review board–approved research staff had access to the data collection platforms. Data were then downloaded from the respective servers and securely stored in a University of California, San Francisco's Box folder. Participants were not paid for completing the screening survey, but were paid US $40 for completing the baseline questionnaire and US $70 for completing the 6-month follow-up questionnaire.

Results

Facebook

Table 1 presents detailed ad performance metrics by platform. Over the 10-week active ads period, our ads were displayed 418,173 times (also known as "impressions"), received 4756 clicks and had a click-through rate of 1.14% (number of impressions ÷ number of clicks), with an average of US $0.44 per click. In the weeks in which both English and Spanish Facebook ads were running, the click-through rate for English ads was 0.98% and the click-through rate for Spanish ads was 2.08%.

Table . Performance counts by language for Craigslist and Facebook ads (n=76).

Metric	Overall	Facebook	Craigslist	*P* value
Clicked on a study ad and completed study enrollment screener, n	579	71	508	.02
English	507	68	439	
Spanish	72	3	69	
Completed and passed eligibility survey, n	293	58	235	.23
English	263	55	208	
Spanish	30	3	27	
Completed baseline survey and downloaded DIAMANTE app, enrolled in study, n	76	26	50	.29
English	72	26	46	
Spanish	4	0	4	
Ad conversion rate, %	13.12	36.62	9.84	
English	14.20	35.62	10.48	—[a]
Spanish	5.56	0	5.8	

[a]Not applicable.

Of the individuals who were initially interested from the Facebook ads, 71 went onto the study Qualtrics website and completed study screening (68 from English ads and 3 from Spanish ads). Out of those 71 interested individuals who completed our Qualtrics screening survey, 58 (55 English speakers and 3 Spanish speakers) were found eligible for the study. Out of those 58 eligible participants, 26 (26 English speakers and 0 Spanish speakers) were eventually recruited into the RCT upon completing the baseline survey, downloaded the DIAMANTE app, and randomized into 1 of the 3 study arms. In all, Facebook advertisements recruited 26 participants, averaging US $80.16 per participant (Table 2), with a conversion rate of 36.62%. Notably, we were unable to recruit Spanish speakers through Facebook.

XSL•FO
RenderX

Table . Recruitment duration, costs, and number of recruits by recruitment platform.

Recruitment platform	Total cost, US $	Campaign duration (days), n	Number of participants recruited, n	Participants per day (number of participants/duration)	Average cost per participant, US $
Facebook	2084.07	76	26	0.34	80.16
English	1703.91	76	26	0.34	80.16
Spanish	380.16	27	0	0	—[a]
Craigslist	7914	224	50	0.22	158.28
English	3110	224	46	0.21	67.61
Spanish	4804	124	4	0.03	1201
Overall total	9998.07	300	76	0.25	131.55
English	4813.91	—	72	—	66.86
Spanish	5184.16	—	4	—	1296.04

[a]Not applicable.

Craigslist

We found that it was challenging to determine the effectiveness of recruitment ads in Craigslist for our study because, unlike the more robust ad analytics framework offered by Facebook, Craigslist offers a simple ad management interface for posting content that does not include ad performance metrics such as number of ad views or click-throughs. However, we could track (from participant self-reports within our Qualtrics platform) that 508 individuals (439 English speakers and 69 Spanish speakers) saw an ad-on Craigslist and completed a screener. Of those, 235 (208 English speakers and 27 Spanish speakers) were found eligible for the study after completing the short Qualtrics survey. Out of those 235 eligible participants, 50 were recruited into the study (46 English speakers and 4 Spanish speakers), averaging US $158.28 per participant (US $67.61 for English speakers and US $1201 for Spanish speakers; Table 2). The conversion rate for Craigslist advertisements was, therefore, 9.84% (10.48% for English advertisements and 5.80% for Spanish advertisements). When broken down by language, the recruitment costs varied substantially, with the average cost for English speakers being US $68 and US $1201 for Spanish speakers. This means that the cost to recruit one Spanish speaker was equivalent to recruiting approximately 18 English speakers.

Participant Characteristics

Table 3 presents the sociodemographic characteristics of the Facebook and Craigslist participants enrolled and randomized in the RCT. Descriptive statistics of respondents across the 2 recruitment platforms revealed similar demographic trends, albeit with some notable differences. The average age (45 years) was the same in both samples. Most respondents were female in both samples (19/26, 73% from Facebook and 32/50, 64% from Craigslist). Craigslist participants tended to be from a higher socioeconomic status than the Facebook participants: a graduate degree (4/26, 15% from Facebook vs 9/50, 18% from Craigslist), currently employed (12/26, 46% from Facebook vs 33/50, 66% from Craigslist), self-reported lower income (16/26, 62% from Facebook vs 23/50, 46% from Craigslist). Race and ethnicity were significantly different between the Craigslist and Facebook samples, with the former being more racially and ethnically diverse than the latter. For example, Whites were overrepresented in the Facebook sample vis-à-vis the general population, with over two-thirds (20/26, 77%) of participants identifying as such. On the other hand, out of 50 Cragslist participants, 18 (36%) identified as White, 14 (28%) as Latino, 8 (16%) as Black, 6 (12%) as Other or Multiracial, and 4 (8%) as Asian or Pacific Islander.

XSL•FO
RenderX

Table . Participant characteristics (n=76).

Characteristic	Overall (n=76)	Facebook (n=26)	Craigslist (n=50)	*P* value
Age (years), mean (SD), range	45 (11.74), 24-68	45 (13.24), 26 - 68	45 (11.02), 24 - 67	.83
Gender, n (%)				.50
Male	23 (30)	6 (23)	17 (34)	
Female	51 (67)	19 (73)	32 (64)	
Other	2 (3)	1 (4)	1 (2)	
Language n (%)				.29
English	72 (95)	26 (100)	46 (92)	
Spanish	4 (5)	0 (0)	4 (8)	
Education, n (%)				.72
Some high school or less	4 (5)	0 (0)	4 (8)	
High school graduate or general educational development	10 (13)	4 (15)	6 (12)	
Some college or tech school	20 (26)	7 (27)	13 (26)	
College graduate	29 (38)	11 (42)	18 (36)	
Graduate degree	13 (17)	4 (15)	9 (18)	
Race and ethnicity, n (%)				.01
White or Caucasian	38 (50)	20 (77)	18 (36)	
Black or African American	11 (14)	3 (12)	8 (16)	
Asian or Pacific Islander	4 (5)	0 (0)	4 (8)	
Latino or Hispanic	15 (20)	1 (4)	14 (28)	
Other or multiracial	8 (11)	2 (8)	6 (12)	
Employment, n (%)				.16
Full time	32 (42)	11 (42)	21 (42)	
Part time	13 (17)	1 (4)	12 (24)	
Homemaker	6 (8)	3 (12)	3 (6)	
Unemployed	13 (17)	5 (19)	8 (16)	
Disabled or on disability	8 (11)	5 (19)	3 (6)	
Retired	3 (4)	1 (4)	2 (4)	
Other	1 (1)	0 (0)	1 (2)	
Low income, n (%)[a]				.20
Yes	39 (51)	16 (62)	23 (46)	
No	37 (49)	10 (38)	27 (54)	

[a]We labeled responses to the subjective income ladder question as: 1 - 5 for low income and 6 - 10 for not low income.

Discussion

Principal Findings

Our results support the feasibility of recruiting diverse respondents for a digital health trial through social media. While Craigslist study recruitment ads reached more people, resulting in more completed screening surveys than Facebook ads, there was a higher number of ineligible and incomplete enrollment from Craigslist compared with Facebook. Yet, despite the overall ad conversion rate being higher on Facebook than on Craigslist, Facebook ads proved more expensive (when only accounting for English speaking participants) and, unlike previous recruiting outcomes by Bunge et al [17], ineffective for recruiting Spanish speakers. Importantly, participants recruited through Craigslist were more ethnically and racially diverse than those recruited from Facebook. We did manage to recruit a small number of Spanish speakers through Craigslist;

XSL•FO
RenderX

this finding aligns with the work by Pratap et al [8], which found this platform to be more effective in recruiting Spanish speakers than targeted advertising through Facebook and Twitter. However, participants recruited through Craigslist only accounted for 5.3% (4/76) of our total sample, even though we intended to have at least a sample of Spanish speakers proportionate to that found in our prepandemic clinical sample recruited through traditional methods (Spanish speakers accounted for 67% (10/15 of our clinical sample). With the cost of recruiting Spanish speakers being substantially higher than English speakers, even more than that reported in previous literature, we confirm the need for substantial financial resources to be injected into digital recruitment efforts in order to sample from this marginalized population more effectively. Budgets of studies that seek to include this group should consider this premium expense. It also appears that an online-only recruitment strategy alone may be ineffective for Spanish speakers. Other mHealth studies have successfully recruited Spanish speakers by coordinating with community organizations to develop trust [34], indicating the potential merit of applying a variety of approaches to recruit hard-to-reach populations in such studies [35].

This study had some limitations. First of all, because we had limited resources yet wanted to create engaging advertisements for Craigslist and Facebook users to recruit from for our study sample in a short amount of time, we conceptualized the ads themselves as a "product to design for" under the UCD framework without stakeholder input. Ideally, we would have asked for feedback directly from targeted audiences in developing such ads, but we were constrained logistically and financially during the COVID-19 pandemic when the online recruitment for this study launched. Co-design with stakeholders strategy could have enhanced the ads' effectiveness, particularly when finding the right balance of advertising specificity (to target specific groups like low-income Spanish speakers) versus reach (to attain a representative sample). For example, although we produced many ads using personas, we ultimately selected a small number of such ads; not including input from end users, particularly of Spanish speakers, in this selection process is a clear limitation. Another limitation is that the timeframe and money spent in both campaigns differed. For example, the fact that the Facebook ads were futile in recruiting Spanish speakers could be due to the shorter campaign duration compared with the Craigslist campaign duration. Our sense, however, was that the Spanish ads on Facebook were simply ineffective (ie, the price post them was very high despite the low rates of participants completing the eligibility screener), regardless of the timeframe, given the much lower engagement they generated among the target population versus the English ads. Furthermore, our cost calculations are not comprehensive, as we did not factor in staff time to develop, implement, and monitor the ads in our cost calculations, beyond the advertising fees charged by the social media platforms. Although this is not inconsistent with cost reporting found in relevant previous literature [8,17], future studies should carefully record and account for staff implementation costs in order to generate more precise cost-effectiveness results of social media study recruitment ads. In general, having a more consistent recruitment, enrollment, and cost reporting standards would allow for more data comparability across mHealth studies [35,36]. Finally, we found it difficult to draw upon best practices in the mHealth literature for recruitment procedures through social media; while there are guides on the use of social media for health research recruitment [37], most studies provide minimal details of their social media recruitment protocols and strategies. Thus, to advance the top public health imperatives of mitigating health service inequities, the field of digital health must adopt more rigorous reporting standards and data transparency of recruitment methods.

Our study proved that it is challenging, albeit possible, to recruit individuals from marginalized groups for digital health trials through social media channels. This methodology allowed us to recruit a diverse sample in terms of socioeconomic and racial characteristics. However, the low number of recruited Spanish speakers implies that this population may have additional recruitment barriers to pursue. Understanding the best methods for effectively recruiting such a hard-to-reach group is crucial for advancing digital health science. Future research should directly ask targeted audiences about barriers and facilitators to enrolling in a study online to help address the challenges by modifying recruitment procedures. Trust may be a critical factor in recruiting Spanish speakers who may be wary of consenting online without having a previous relationship with researchers, especially for individuals who started or completed the study screening but did not ultimately enroll in the trial. Indeed, mHealth studies with other marginalized groups have reported higher outreach outcomes of eligible participants who identified as a racial or ethnic minority through in-person methods as opposed to social media advertising [22].

It may also be worthwhile to test further whether other online methods, beyond paid advertisements, are needed to recruit diverse samples. Using posts and other paid and free methods on Facebook (ie, pages, sending friend invites to prospective participants, and untargeted ads) has been found effective in previous studies [10,36]. Furthermore, content-sharing platforms such as YouTube (Google), TikTok (ByteDance), and Instagram (Meta) may have something beneficial to offer in the targeted recruitment of individuals belonging to marginalized communities for research purposes, particularly when it comes to the use of video. Future studies should continue to test ad customization and content development to advance the research community's understanding of best practices on the recruitment of diverse populations.

Finally, it will be important to continue to report data from online recruitment strategies to refine methods and approaches as technology evolves. This is especially true as we seek to confirm participant identity (using CAPTCHA and other methods to remove bots), given that the large number of participants from Craigslist that did not convert into the study could have been linked to computer-generated responses to our social media ads.

Conclusion

Results from this study revealed that it is possible to recruit diverse sample sets using social media advertisements. However, despite targeted recruitment efforts, social media recruitment of Spanish speakers proved especially challenging and costly.

XSL•FO
RenderX

Further research is needed to determine systematic, online methods for recruiting marginalized communities into digital health studies to advance the public health goal of reducing service inequities. Recruiting samples that reflect the racial and ethnic diversity of the US population in rigorous studies assessing mobile mental health interventions, particularly RCTs, is thus crucial for the mHealth promise of ameliorating racial and ethnic service inequities to unfold as intended.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Advertisement copy by persona.
[DOCX File, 14 KB - mhealth_v13i1e56329_app1.docx]

References

1. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882. [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
2. Price M, Yuen EK, Goetter EM, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436. [doi: 10.1002/cpp.1855] [Medline: 23918764]
3. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
4. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug;218:15-22. [doi: 10.1016/j.jad.2017.04.046]
5. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336. [doi: 10.1002/wps.20673] [Medline: 31496095]
6. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756. [doi: 10.1111/acem.12692] [Medline: 25997375]
7. James DC, Harville C, Sears C, Efunbumi O, Bondoc I. Participation of African Americans in e-Health and m-Health studies: a systematic review. Telemed J E Health 2017 May;23(5):351-364. [doi: 10.1089/tmj.2016.0067] [Medline: 27792475]
8. Pratap A, Renn BN, Volponi J, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 9;20(8):e10130. [doi: 10.2196/10130] [Medline: 30093372]
9. Paskett ED, Reeves KW, McLaughlin JM, et al. Recruitment of minority and underserved populations in the United States: the Centers for Population Health and Health Disparities experience. Contemp Clin Trials 2008 Nov;29(6):847-861. [doi: 10.1016/j.cct.2008.07.006] [Medline: 18721901]
10. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. Internet Interv 2016 May;4:1-10. [doi: 10.1016/j.invent.2016.01.001]
11. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013 Feb 12;15(2):e31. [doi: 10.2196/jmir.2262] [Medline: 23403043]
12. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
13. Arean PA, Hallgren KA, Jordan JT, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330. [doi: 10.2196/jmir.6482] [Medline: 27998876]
14. Pratap A, Anguera JA, Renn BN, et al. The feasibility of using smartphones to assess and remediate depression in Hispanic/Latino individuals nationally. Presented at: UbiComp '17: The 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 11-15, 2017; Maui, Hawaii. [doi: 10.1145/3123024.3127877]
15. Brodar KE, Hall MG, Butler EN, et al. Recruiting diverse smokers: enrollment yields and cost. Int J Environ Res Public Health 2016 Dec 16;13(12):1251. [doi: 10.3390/ijerph13121251] [Medline: 27999280]
16. Akers L, Gordon JS. Using Facebook for large-scale online randomized clinical trial recruitment: effective advertising strategies. J Med Internet Res 2018 Nov 8;20(11):e290. [doi: 10.2196/jmir.9372] [Medline: 30409765]
17. Bunge EL, Taylor LA, Bond M, et al. Facebook for recruiting Spanish- and English-speaking smokers. Internet Interv 2019 Sep;17:100238. [doi: 10.1016/j.invent.2019.02.002]

XSL•FO
RenderX

18. Fletcher S, Clarke J, Sanatkar S, et al. Recruiting to a randomized controlled trial of a web-based program for people with type 2 diabetes and depression: lessons learned at the intersection of e-Mental health and primary care. J Med Internet Res 2019 May 24;21(5):e12793. [doi: 10.2196/12793] [Medline: 31127718]
19. Layi G, Albright CA, Berenberg J, et al. UH Cancer Center Hotline: recruiting cancer survivors for an online health-behavior change intervention: are different strategies more beneficial? Hawaii Med J 2011 Oct;70(10):222-223. [Medline: 22162599]
20. Adam LM, Manca DP, Bell RC. Can Facebook be used for research? Experiences using Facebook to recruit pregnant women for a randomized controlled trial. J Med Internet Res 2016 Sep 21;18(9):e250. [doi: 10.2196/jmir.6404] [Medline: 27655184]
21. Carlini BH, Safioti L, Rue TC, Miles L. Using Internet to recruit immigrants with language and culture barriers for tobacco and alcohol use screening: a study among Brazilians. J Immigr Minor Health 2015 Apr;17(2):553-560. [doi: 10.1007/s10903-013-9934-1] [Medline: 24563138]
22. Zlotorzynska M, Bauermeister JA, Golinkoff JM, Lin W, Sanchez TH, Hightow-Weidman L. Online recruitment of youth for mHealth studies. Mhealth 2021;7:27. [doi: 10.21037/mhealth-20-64] [Medline: 33898596]
23. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64. [doi: 10.1016/j.invent.2014.05.001]
24. Choi I, Milne DN, Glozier N, Peters D, Harvey SB, Calvo RA. Using different Facebook advertisements to recruit men for an online mental health study: engagement and selection bias. Internet Interv 2017 Jun;8:27-34. [doi: 10.1016/j.invent.2017.02.002] [Medline: 30135825]
25. Aguilera A, Figueroa CA, Hernandez-Ramos R, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. BMJ Open 2020 Aug 20;10(8):e034723. [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]
26. Figueroa CA, Hernandez-Ramos R, Boone CE, et al. A text messaging intervention for coping with social distancing during COVID-19 (StayWell at Home): protocol for a randomized controlled trial. JMIR Res Protoc 2021 Jan 14;10(1):e23592. [doi: 10.2196/23592] [Medline: 33370721]
27. Hernandez-Ramos R, Aguilera A, Garcia F, et al. Conducting internet-based visits for onboarding populations with limited digital literacy to an mHealth intervention: development of a patient-centered approach. JMIR Form Res 2021 Apr 29;5(4):e25299. [doi: 10.2196/25299] [Medline: 33872184]
28. National diabetes statistics report. Centers for Disease Control and Prevention Diabetes. URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2025-03-29]
29. Miaskiewicz T, Kozar KA. Personas and user-centered design: how can personas benefit product design processes? Design Studies 2011 Sep;32(5):417-430. [doi: 10.1016/j.destud.2011.03.003]
30. Matthews T, Judge T, Whittaker S. How do designers and user experience professionals actually perceive and use personas? Presented at: 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12; May 5-10, 2012; Austin, Texas, USA. [doi: 10.1145/2207676.2208573]
31. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implementation Sci 2011 Dec;6(1):42. [doi: 10.1186/1748-5908-6-42]
32. Privacy violations and personal attributes. Meta. URL: https://transparency.meta.com/en-gb/policies/ad-standards/objectionable-content/privacy-violations-personal-attributes/ [accessed 2025-04-12]
33. Stata Statistical Software: Release 16: StataCorp LLC; 2019.
34. Haro-Ramos AY, Rodriguez HP, Aguilera A. Effectiveness and implementation of a text messaging intervention to reduce depression and anxiety symptoms among Latinx and Non-Latinx white users during the COVID-19 pandemic. Behav Res Ther 2023 Jun;165:104318. [doi: 10.1016/j.brat.2023.104318] [Medline: 37146444]
35. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022 Aug 4;24(8):e38015. [doi: 10.2196/38015] [Medline: 35925655]
36. Darmawan I, Bakker C, Brockman TA, Patten CA, Eder M. The role of social media in enhancing clinical trial recruitment: scoping review. J Med Internet Res 2020 Oct 26;22(10):e22810. [doi: 10.2196/22810] [Medline: 33104015]
37. Pozzar R, Hammer MJ, Underhill-Blazey M, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020 Oct 7;22(10):e23021. [doi: 10.2196/23021] [Medline: 33026360]

Abbreviations

CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart
COM-B: Capability, Opportunity, Motivation and Behavior
DIAMANTE: Diabetes and Mental Health Adaptive Notification Tracking and Evaluation
mHealth: mobile health
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by L Buis; submitted 17.01.24; peer-reviewed by E Jean Buckler, M Loades, TA Reddy Sure; revised version received 12.12.24; accepted 18.02.25; published 28.04.25.

Please cite as:
Pathak L, Hernandez-Ramos R, Rosales K, Miramontes-Gomez J, Garcia F, Yip V, Sudarshan S, Cemballi AG, Lyles C, Aguilera A Using Social Media to Recruit a Diverse Sample of Participants for a Mobile Health (mHealth) Intervention to Increase Physical Activity: Exploratory Study
JMIR Mhealth Uhealth 2025;13:e56329
URL: https://mhealth.jmir.org/2025/1/e56329
doi:10.2196/56329

© Laura Pathak, Rosa Hernandez-Ramos, Karina Rosales, Jose Miramontes-Gomez, Faviola Garcia, Vivian Yip, Suchitra Sudarshan, Anupama Gunshekar Cemballi, Courtney Lyles, Adrian Aguilera. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Combined Use of Digital and Analog Physical Therapy in Patients With Musculoskeletal Disorders and Indicators of Chronicity: German Claims Data Analysis

Silke Frey[1], MSc; Annika Schmitz[1], MA; Udo Schneider[2], PD, PhD; Linda Kerkemeyer[3], PhD; Birgitta Weltermann[1], MPH, Prof Dr

[1]Institute of General Practice and Family Medicine, University Hospital Bonn, University of Bonn, Venusberg-Campus 1, Bonn, Germany

[2]Techniker Krankenkasse, Hamburg, Germany

[3]LiKe Healthcare Research GmbH, Berlin, Germany

Corresponding Author:
Silke Frey, MSc
Institute of General Practice and Family Medicine, University Hospital Bonn, University of Bonn, Venusberg-Campus 1, Bonn, Germany

Abstract

Background: Musculoskeletal disorders are highly prevalent worldwide and contribute significantly to the overall burden of disease. Regular physical therapy with trained physiotherapists is recommended in the guidelines. Recently, digital physical therapy offered by digital health interventions was shown to be effective. However, the evidence on its real-world usage in health care systems is limited.

Objective: Based on claims data, this study examined the current usage of digital health applications (DiGAs) for musculoskeletal disorders in the German health care system. Patients with standalone digital physical therapy were compared to those with a combination of analog and digital physical therapy. In addition, predictors for concomitant use were identified.

Methods: This retrospective cohort study analyzed claims data from Germany's largest statutory health insurance. Patients who used DiGA for musculoskeletal disorders at least once were included. Sociodemographic and medical characteristics of patients receiving standalone and concomitant physical therapy were compared. Statistical analyses comprised univariate analyses and binomial logistic regression.

Results: Of the 6090 individuals, 58.2% (3543/6090) were prescribed physical therapy within 6 months before or after DiGA prescription. In this population, 36.3% (2210/6090) used DiGA and analog physical therapy at the same time. Concomitant physical therapy was significantly more likely in patients with chronicity risk (odds ratio [OR] 1.49, 95% CI 1.31 - 1.69; *P*<.001) or established chronicity (OR 2.76, 95% CI 2.22 - 3.47; *P*<.001), female gender (OR 1.48, 95% CI 1.33 - 1.66; *P*<.001), and higher age (OR 1.02, 95% CI 1.02 - 1.02; *P*<.001).

Conclusions: The findings highlight the diverse utilization patterns of DiGAs among patients with musculoskeletal disorders. Chronicity emerged as an important predictor for combined digital and analog physical therapy. These findings support considerations on integrating digital health interventions into current guidelines.

(JMIR Mhealth Uhealth 2025;13:e63935) doi:10.2196/63935

KEYWORDS

claims analysis; digital health interventions; digital health applications; DiGA; mHealth; utilization; musculoskeletal disorders; physical therapy; exercise therapy; chronicity

Introduction

Musculoskeletal disorders (MSDs), back pain in particular, contribute to disability, with approximately 1.71 billion people affected worldwide [1]. The burden of MSDs has increased significantly over recent decades, resulting in 107,885,832.6 disability-adjusted life years in 2015 globally [2]. In Europe, the burden of MSDs is even higher than in other continents [2]. The prevalences of lower back pain and neck pain in Germany reached 61.3% and 45.7% over a 12-month period, with women being significantly more often affected than men [3]. Similarly, 17.3% of women and 15.1% of men aged 18-79 years experienced knee pain within the last 24 hours in a German random sample (N=8151) [4]. The increasing incidence of MSDs has resulted in considerable economic costs for individuals and society due to lost workdays and conservative or surgical interventions [2,5].

XSL•FO
RenderX

Exercise therapy is effective in reducing pain and enhancing functionality in individuals suffering from MSDs, including chronic lower back pain, knee osteoarthritis, and fibromyalgia. Due to its demonstrated benefits, this approach is recommended in medical guidelines and reviews [6-8]. Recently, digital health interventions (DHIs) have become a new therapeutic option to guide patients in regular physical activity. DHIs cover various medical conditions and offer features such as symptom tracking, educational content, video-guided therapeutic exercises, and feedback. The following textbox displays the content and therapeutic approaches of digital and analog physical therapy in comparison (Textbox 1).

Textbox 1. Comparison of digital and analog physical therapy.

Digital physical therapy - digital health intervention (DHI)

- Structured exercise programs with video, audio, or app-based guidance
- Algorithm-driven progression and adaptation
- Automated reminders & frequent exercise sessions
- Monitoring of pain levels and physical functionality based on patient-reported outcome measures (PROMs) and self-tests
- Educational content

The content of DHIs is based on information in the DiGA directory of the Federal Institute for Drugs and Medical Devices, Germany [9].

Analog physical therapy

- Supervised exercise therapy with individual feedback and correction
- Therapist-guided progression and adaptation
- Scheduled in-person sessions
- Initial assessment & monitoring of pain levels and physical functionality documented by the therapist
- Patient education and home exercises

The content of analog physical therapy is presented as published in the German Remedy Guideline (Helmittel-Richtlinie) [10].

Analog physical therapy provides direct professional supervision and individualized feedback, but is resource-intensive and limited to scheduled sessions. In contrast, digital physical therapy facilitates frequent exercises independent of time and location, but relies on patients' self-motivation and adherence. Both approaches offer distinct advantages and challenges, making them suitable for different patient populations.

Evidence indicates that DHIs are effective in managing MSDs. A recent systematic review by Nagel and colleagues [11], which included 16 high-quality, randomized controlled trials with a total of 1840 patients with MSDs, confirmed that digital physical therapy reduces back pain significantly. There was moderate or limited evidence for positive effects on other body parts affected by MSDs [11]. These findings are in line with previous studies confirming the efficacy of DHIs in treating MSDs [12-14]. Furthermore, a systematic review and meta-analysis by Fatoye and colleagues [15], referring to 10 studies from 6 countries, concluded that DHIs are cost-effective for people affected by MSDs.

Because of the promising effects of DHIs and the ongoing digital transformation of the health care sector, Germany introduced so-called digital health applications (*Digitale Gesundheitsanwendung* [DiGA]) in 2020. These software-based medical devices are prescribed by physicians and psychotherapists and are reimbursed by statutory health insurances (SHIs). Patients with a confirmed diagnosis can apply for an activation code for a DiGA through their SHI, even in the absence of a prescription (self-application). According to the Digital Healthcare Act (*Digitale-Versorgung-Gesetz* [DVG]), each DiGA is evaluated and gets a (temporal) approval by the Federal Institute for Drugs and Medical Devices (*Bundesinstitut für Arzneimittel und Medizinprodukte* [BfArM]). This fast-track process ensures compliance with quality standards, such as product safety, data protection and user-friendliness, and assesses health care efficacy. Currently, nearly 60 DiGAs are listed in the directory of the BfArM, with MSDs being a significant focus in addition to mental illnesses [16]. The number of DiGA users has increased concurrently with the growing number of prescribable DiGAs, reaching a total of 374,377 individuals by September 2023 [17].

Despite the growing interest in DiGAs, their utilization currently falls short of expectations. This disparity arises from health care providers' skepticism concerning factors such as the efficacy and data security of DiGAs. In addition, DiGAs are not yet part of medical guidelines [18]. In general, DiGAs can be used as substitutes, as a complementary add-on to existing health care services, or as partial replacements, for example, to bridge delays in health care provision, such as waiting times for physical therapy [19].

Previous publications from authorities, interest groups, or SHI companies have primarily examined and described the use of DiGAs in health care [17,20-22]. However, there are no comparisons to analog therapies nor tracking of individual patient journeys in a real-world setting. Therefore, this study aimed to analyze the status quo of DiGA utilization for MSDs based on claims data, with a focus on comparing standalone digital therapy to concomitant digital and analog physical therapy.

XSL•FO
RenderX

Methods

Study Design and Cohort

A retrospective cohort study analyzing DiGA users with MSDs was conducted. Nationwide claims data of the largest sickness fund in Germany, the Techniker Krankenkasse (TK), covering approximately 11 million people, were provided in pseudonymized form. All adults (at least 18 y) insured by TK who received and activated at least one DiGA for MSD (Vivira [Vivira Health Lab GmbH], Companion patella [PrehApp GmbH], or Mawendo [Mawendo GmbH]) since September 2020 were included in the study. Eligible patients were identified via the pharmaceutical central number of the DiGAs.

Given delays of up to 9 months between the occurrence, processing, and availability of claims data especially for outpatient care, services provided until March 31, 2023, were analyzed. In addition, each patient's claims data required a minimum of 6 months of follow-up, restricting inclusion to patients prescribed a DiGA before October 1, 2022. At the time, three DiGAs for MSDs were listed in the DiGA directory: Vivira for unspecific lower back pain and osteochondrosis, as well as Companion patella and Mawendo for anterior or patella-related knee pain. Detailed descriptions of the functionalities of these DiGAs are presented in the Multimedia Appendix 1. During the observation period, patients had to be insured continuously with TK, with a tolerated gap of up to 30 days. Data of individuals with a privacy protection indicator were not transferred; these included persons who objected to the use of their data for research purposes as well as employees of the TK.

Using either the DiGA prescription date or the receipt of the self-application at TK, a tailored preobservation and follow-up period were determined for each patient. Each patient-individual quarter was defined by 90 days, the duration of the DiGA prescription. The preobservation comprised 8 quarters before the DiGA prescription, while the follow-up ideally encompassed eight quarters, with a minimum of 2. The analysis included people who died during the follow-up.

Operationalization of Physical Therapy and Chronicity

The sample was divided into 2 groups based on the utilization of physical therapy in addition to DiGA. The group with DiGA and analog physical therapy comprised patients who received at least one physical therapy session within the last 2 preobservation quarters and the first 2 follow-up quarters, while the group receiving DiGA as standalone therapy comprised patients who did not receive physical therapy within the same time frame.

Additional subgroups were created based on the degree of chronicity, using proxies such as sick leave days and medication patterns due to the absence of clinical information in claims data. This approach follows the methodology of the claims data analysis on back pain and the significance of chronicity by Freytag and colleagues [23], categorizing individuals as follows:

1. Those incapacitated for more than 42 days in the last 2 preobservation and the first 2 follow-up quarters or who received opioid prescriptions at least twice were classified as "with indication of chronicity."
2. Insured individuals were included in the "with indication of chronicity risk" group if the outpatient diagnoses included at least one of the following psychological comorbidities in the last 2 preobservation and the first 2 follow-up quarters: F32* (depressive episode), F33* (recurrent depressive disorder), F34.1 (dysthymia), F34.8 (other persistent affective disorder), F34.9 (unspecified persistent affective disorder), F38* (other affective disorder), F41.2 (mixed anxiety and depressive disorder), F45.4 (persistent somatoform pain disorder), F48.0 (neurasthenia), F43.20 (adjustment disorder), F43.21 (adjustment disorder with prolonged depressive reaction), F43.22 (adjustment disorder with anxiety and depressive reaction), F54 (psychological behavioral factors associated with disorders or diseases classified elsewhere), F62.80 (personality change due to chronic pain syndrome).
3. All other insured individuals who did not fall into the above-mentioned groups were categorized as "without indication of chronicity."

Ethical Considerations

All methods conducted in this study followed relevant guidelines and regulations, including the German Reporting Standard for Secondary Data Analyses (*Standardisierte Berichtsroutine für Sekundärdaten Analysen* [STROSA]) and Good Practice in Secondary Data Analysis (*Gute Praxis Sekundärdatenanalyse* [GPS]) guidelines. The data were transmitted with permission of the German Federal Office for Social Security under § 75 of the Social Code (*Zehntes Buch Sozialgesetzbuch - SGB X*). The Ethics Committee of the Medical Faculty at the University of Bonn has reviewed the study proposal and raised no ethical or legal concerns (054/23-EP). The data protection officer of the University Hospital Bonn actively contributed to the data protection concept and provided a positive assessment of the project. Individual informed consent for data usage was not required, and no experiments or surveys were conducted to gather the data.

Data Analysis

Plausibility and validity checks were performed for all variables in the datasets. Subsequently, the individual preobservation and follow-up were determined for each insured person. Descriptive analyses were conducted for the entire cohort as well as for the physical therapy subgroups. Absolute and relative frequencies are reported for categorical variables. Differences between physical therapy groups were assessed using the χ^2 test or Fisher exact test. The visualization of all patient pathways, the 10 most common pathways, and the state distribution in each quarter was created using the TraMineR package for sequence analysis in R [24].

A binomial logistic regression analysis was performed to investigate the potential factors associated with using DiGA as a standalone therapy or in conjunction with physical therapy. Chronicity risk, age, gender, and DiGA characteristics were incorporated into the model via a forward variable selection process, considering the significance of added variables, the Akaike information criterion, and Nagelkerke R^2. Odds ratios (ORs), along with their corresponding 95% CI and *P* values, were computed to evaluate the presence and strength of the

XSL•FO
RenderX

associations between the covariates and the utilization of physical therapy. Multicollinearity in the final model was assessed by calculating the variance inflation factor. The significance level was predetermined at 5%. All analyses were conducted using R Studio (Posit, PBC) [25].

Results

Sociodemographic Characteristics

After applying inclusion and exclusion criteria, 6090 patients with at least one DiGA for MSD were eligible for the analysis. Within the last 2 preobservation quarters, the DiGA quarter itself, and the subsequent quarter, 3543 individuals (3543/6090, 58.2%) received at least one physical therapy prescription. For 36.3% (2210/6090) of the study population, there was an overlap in the actual duration of physical therapy and the usage of DiGA. Except for the number of patients who died during follow-up, the highest professional training, and the membership of the individuals, all other sociodemographic and medical characteristics differed significantly between patients who did and did not receive physical therapy (Tables 1 and 2). For instance, the proportion of women was greater among patients receiving physical therapy (χ^2_1=75.1; P<.001), and these patients were significantly older (χ^2_6=166.1; P<.001). Further details regarding sociodemographic characteristics can be found in Table 1.

XSL•FO
RenderX

Table . Sociodemographic characteristics of the study population.

Item and category	Total, n (%)	Without physical therapy, n (%)	With physical therapy, n (%)	χ2 (*df*)	*P* value
All patients	6090 (100)	2547 (41.8)	3543 (58.2)	—[a]	—
Gender				75.1 (1)	<.001
Women	3970 (65.2)	1501 (58.9)	2469 (69.7)		
Men	2120 (34.8)	1046 (41.1)	1074 (30.3)		
Age, years				166.1 (6)	<.001
15 to <20	60 (1)	35 (1.4)	25 (0.7)		
20 to <30	771 (12.7)	442 (17.4)	329 (9.3)		
30 to <40	1406 (23.1)	658 (25.8)	748 (21.1)		
40 to <50	1219 (20)	517 (20.3)	702 (19.8)		
50 to <60	1604 (26.3)	562 (22.1)	1042 (29.4)		
60 to <65	530 (8.7)	168 (6.6)	362 (10.2)		
65 to >65	500 (8.2)	165 (6.5)	335 (9.5)		
Death				0.0 (1)	.99, >.99[b]
Died during follow-up	6 (0.1)	2 (0.1)	4 (0.1)		
Membership				4.3 (1)	.04
Individual member	5662 (93)	2347 (92.1)	3315 (93.6)		
Family insured member	428 (7)	200 (7.9)	228 (6.4)		
Highest professional training				12.3 (6)	.06
No professional qualification	196 (3.2)	93 (3.7)	103 (2.9)		
Vocational training	1839 (30.2)	743 (29.2)	1096 (30.9)		
Bachelor’s degree	304 (5)	147 (5.8)	157 (4.4)		
Master’s degree	1119 (18.4)	482 (18.9)	637 (18)		
Technician or master	203 (3.3)	90 (3.5)	113 (3.2)		
Doctorate	101 (1.7)	48 (1.9)	101 (2.9)		
N/A[c]	2328 (38.2)	944 (37.1)	1336 (37.7)		
Domicile				17.6 (3)	<.001
City (≥100,000 inhabitants)	2742 (45)	1198 (47)	1544 (43.6)		
Urban counties (≥150 inhabitants per km^2)	2359 (38.7)	988 (38.8)	1371 (38.7)		
Densely populated rural counties (100 - 150 inhabitants per km^2)	589 (9.7)	222 (8.7)	367 (10.4)		
Sparsely populated rural counties (<100 inhabitants per km^2)	395 (6.5)	135 (5.3)	260 (7.3)		
N/A	5 (0.1)	4 (0.2)	1 (0)		

[a]Not applicable.

[b]This *P* value is determined by Fisher exact test.

[c]N/A: not available.

XSL•FO
RenderX

Table . Medical characteristics of the study population.

Item and category	Total, n (%)	Without physical therapy, n (%)	With physical therapy, n (%)	$\chi 2$ (*df*)	*P* value
All patients	6090 (100)	2547 (41.8)	3543 (58.2)	—[a]	—
Chronicity risk				170.8 (2)	<.001
With indication of chronicity	506 (8.3)	109 (4.3)	397 (11.2)		
With indication of chronicity risk	1490 (24.5)	507 (19.9)	983 (27.7)		
Without indication of chronicity	4094 (67.2)	1931 (75.8)	2163 (61)		
Most common outpatient diagnosis in DiGA[b] quarter – main categories				348.2 (33)	<.001, <.001[c]
Factors influencing health status or contact with health services (examination, investigation, and vaccination)	3704 (60.8)	1432 (56.2)	2272 (64.1)		
Musculoskeletal disorders	3139 (51.5)	1044 (41)	2095 (59.1)		
Disorders and injuries of the spine	3022 (49.6)	940 (36.9)	2082 (58.8)		
Ear, nose, and throat disorders	1894 (31.1)	753 (29.6)	1141 (32.2)		
Psychological disorders	1724 (28.3)	567 (22.3)	1157 (32.7)		
Metabolic disorders	1709 (28.1)	595 (23.4)	1114 (31.4)		
Symptoms and abnormal clinical and laboratory findings, not elsewhere classified	1708 (28)	597 (23.4)	1111 (31.4)		
Gynecological and andrological disorders	1525 (25)	555 (21.8)	970 (27.4)		
Heart diseases	1236 (20.3)	396 (15.5)	840 (23.7)		
Gastrointestinal diseases	1232 (20.2)	435 (17.1)	797 (22.5)		
Most common outpatient diagnosis in DiGA quarter – *ICD*[d] codes				2094.7 (1044)	<.001, <.001[c]
Z01 – Examinations and investigations without complaints or reported diagnosis	2459 (40.4)	938 (36.8)	1521 (42.9)		
M54 – Dorsalgia (Back pain)	2208 (36.3)	667 (26.2)	1541 (43.5)		
I10 – Essential (primary) hypertension	1025 (16.8)	328 (12.9)	697 (19.7)		
M99 – Biomechanical lesions, not classified elsewhere	921 (15.1)	280 (11)	641 (18.1)		
M51 – Other intervertebral disc disorders	834 (13.7)	173 (6.8)	661 (18.7)		
Z30 – Contraceptive management	825 (13.5)	349 (13.7)	476 (13.4)		
Z12 – Special screening examination for neoplasms	823 (13.5)	327 (12.8)	496 (14)		

XSL•FO
RenderX

Item and category	Total, n (%)	Without physical therapy, n (%)	With physical therapy, n (%)	χ2 (*df*)	*P* value
M47 – Spondylosis	768 (12.6)	204 (8)	564 (15.9)		
J06 – Acute upper respiratory infections of multiple and unspecified sites	749 (12.3)	329 (12.9)	420 (11.9)		
F45 – Somatoform disorders	692 (11.4)	158 (6.2)	534 (15.1)		
Incapacity days				38.0 (6)	<.001
None	4477 (73.5)	1920 (75.4)	2557 (72.2)		
1 - 3 days	319 (5.2)	134 (5.3)	185 (5.2)		
4 - 7 days	436 (7.2)	197 (7.7)	239 (6.7)		
8 - 14 days	425 (7)	174 (6.8)	251 (7.1)		
15 - 28 days	256 (4.2)	73 (2.9)	183 (5.2)		
29 - 42 days	85 (1.4)	26 (1)	59 (1.7)		
More than 42 days	92 (1.5)	23 (0.9)	69 (1.9)		
Medication				220.6 (10)	<.001, <.001[c]
M01A – Nonsteroidal anti-inflammatory drugs (NSAIDs)	881 (14.5)	250 (9.8)	631 (17.8)		
N06A –Antidepressants	434 (7.1)	138 (5.4)	296 (8.4)		
N02B – Other analgesics and antipyretics	431 (7.1)	112 (4.4)	319 (9)		
M03B – Muscle relaxants, centrally acting agents	216 (3.5)	52 (2)	164 (4.6)		
N02A – Opioids	145 (2.4)	26 (1)	119 (3.4)		
N03A – Antiepileptics	76 (1.2)	19 (0.7)	57 (1.6)		
N05A – Antipsychotics	64 (1.1)	26 (1)	38 (1.1)		
N05C – Hypnotics and sedatives	44 (0.7)	16 (0.6)	28 (0.8)		
N01B – Local anesthetics	14 (0.2)	3 (0.1)	11 (0.3)		
M03A – Muscle relaxants, peripherally acting agents	3 (0)	0 (0)	3 (0.1)		
None of these	4489 (73.7)	2049 (80.4)	2440 (68.9)		

[a]Not applicable.

[b]DiGA: digital health application.

[c]This *P* value is determined by Fisher exact test

[d]*ICD*: *International Classification of Diseases*.

Medical Characteristics

Among the most common diagnoses in the DiGA quarter were M54 – back pain, M99 – biomechanical dysfunction, not elsewhere classified, and M51 – other intervertebral disc disorders, as displayed in Table 2. Almost 15% (881/6090) of the patients used prescribed nonsteroidal anti-inflammatory drugs; over-the counter medications were not included in the SHI claims data. Patients with indications of chronicity or chronicity risks were significantly more likely to undergo physical therapy than those without such indications (χ^2_2=170.8; *P*<.001).

DiGA Characteristics

The most commonly used DiGA in the study population was Vivira for back pain, followed by Companion patella and

XSL•FO
RenderX

Mawendo. Most DiGAs were prescribed by physicians (5717/6090, 93.9%). The leading prescribers were orthopedic surgeons (4395/6090) with 72.2% of prescriptions, with an average of 5.2 DiGA prescriptions per health care provider. General practitioners were the second most common prescribing physicians (629/6090) with 10.3% of prescriptions. Approximately 15% (793/5065) of Vivira users continued using the DiGA beyond the intended 90-day period, opting for consecutive usage cycles, with some engaging in the same DiGA for up to a maximum of 8 consecutive quarters. Details regarding the utilization of DiGAs are presented in Table 3.

Table . Digital health applications (DiGAs) utilization.

Item and category	n/N (%)	Average number of prescriptions per physician
DiGA utilization		
Vivira	5067/6090 (83.2)	N/A[a]
Companion patella	930/6090 (15.3)	N/A
Mawendo	93/6090 (1.5)	N/A
Access to DiGA		
Prescription by a doctor or physiotherapist	5717/6090 (93.9)	N/A
Self-applications at the SHI[b]	373/6090 (6.1)	N/A
Proportion of self-applications per DiGA		
Vivira self-applications	320/6090 (5.3)	N/A
Companion patella self-applications	50/6090 (0.8)	N/A
Mawendo self-applications	3/6090 (0.1)	N/A
Prescribing physician		
Orthopedic surgeon	4395/6090 (72.2)	5.2
General practitioner or primary care internist	629/6090 (10.3)	1.8
Trauma surgeon or neurosurgeon	280/6090 (4.6)	2.3
Physiatrist	188/6090 (3.1)	4.3
Rheumatologist	122/6090 (2)	4.9
Anesthetist (including pain therapist)	37/6090 (0.6)	1.9
Other specialists	66/6090 (1.1)	1.3
N/A	373/6090 (6.1)	N/A
Vivira: quantity of DiGAs used in follow-up period per patient (follow-up prescriptions)		
Once	4272/5067 (84.3)	N/A
More than once	793/5067 (15.7)	N/A
Companion patella: quantity of DiGAs used in follow-up period per patient (follow-up prescriptions)		
Once	860/930 (92.5)	N/A
More than once	70/930 (7.5)	N/A
Mawendo: quantity of DiGAs used in follow-up period per patient (follow-up prescriptions)		
Once	92/93 (98.9)	N/A
More than once	1/93 (1.1)	N/A

[a]N/A: not applicable

[b]SHI: statutory health insurance.

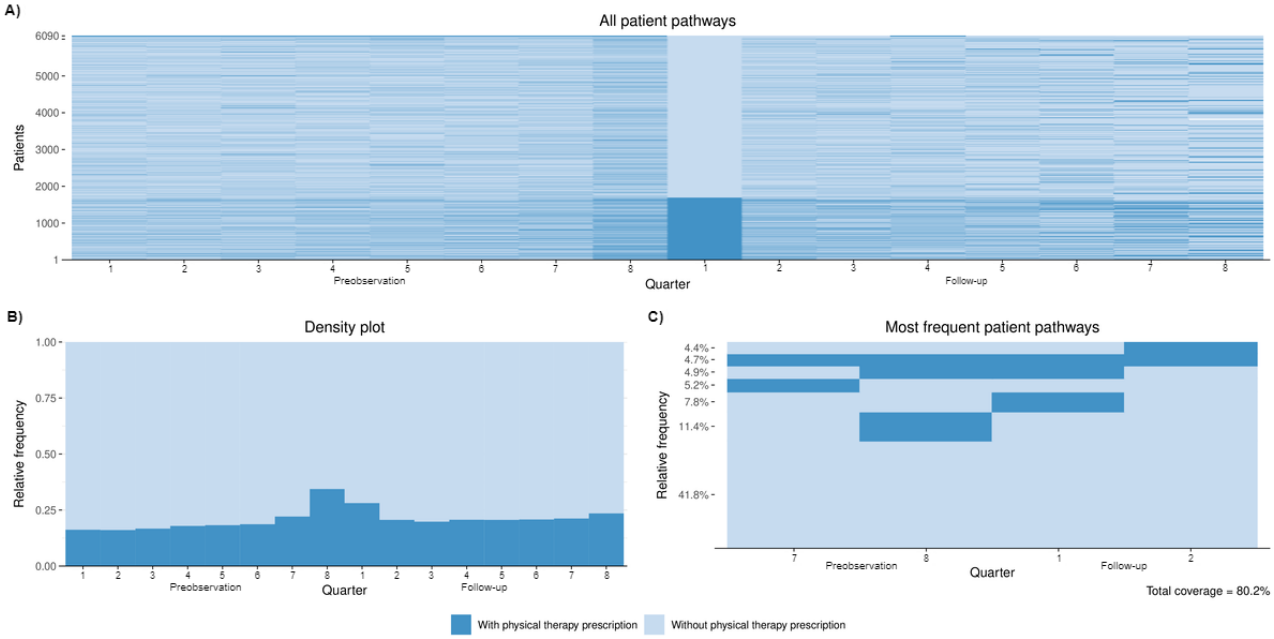
Usage Patterns of Digital and Analog Physical Therapy

During the observation period, there were noticeable alternations between quarters with and without physical therapy (see Figure 1A). Due to varying data periods, insured individuals have different lengths of follow-up periods, but at least two quarters. In the quarter of the DiGA prescription or application, 28.1% (1714/6090) of insured individuals also concurrently had a prescription for physical therapy (see Figure 1B). In most quarters, between 16.1% and 23.5% of insured persons received physical therapy, with a maximum of 34.4% (2096/6090) in the quarter leading up to DiGA use (preobservation 8). The most frequent treatment pathways indicate the concentration of

XSL•FO
RenderX

physical therapy prescriptions around using the DiGA (see Figure 1C). Figure 1 provides insights into patient pathways regarding the utilization of physical therapy.

Figure 1. Patient pathways regarding physical therapy in addition to digital health application utilization. Quarter duration: 90 days.



Multivariate Analysis

Binomial logistic regression was conducted to analyze associations between chronicity, gender, age, type of DiGA, access to DiGA and the risk of using concomitant analog physical therapy. The model yielded statistically significant results (χ^2_7=380.6; *P*<.001), albeit with a relatively modest amount of explained variance, as indicated by Nagelkerke R^2=.08. Notably, significant effects were observed for chronicity, gender, age and type of DiGA used. For instance, patients with indications of chronicity risk (OR 1.48 95% CI 1.31 - 1.67; *P*<.001) or those with indications of chronicity (OR 2.76, 95% CI 2.22 - 3.47; *P*<.001) had higher ORs to undergo physical therapy than patients without such indications. Details of all model coefficients are displayed in Table 4.

Table . Results of the binomial logistic regression model.

Variable[a]	B (SE)	OR[b] values (95% CI)	*P* value
Intercept	−0.88 (0.10)	0.42 (0.34-0.51)	<.001
With indication of chronicity risk	0.40 (0.06)	1.49 (1.31-1.69)	<.001
With indication of chronicity	1.02 (0.11)	2.76 (2.22-3.47)	<.001
Female gender	0.39 (0.06)	1.48 (1.33-1.66)	<.001
Age (in years)	0.02 (0.00)	1.02 (1.02-1.02)	<.001
DiGA[c] – Mawendo	−0.47 (0.22)	0.63 (0.41-0.96)	.03
DiGA – Companion patella	−0.42 (0.07)	0.66 (0.57-0.76)	<.001
DiGA access – self-application	0.08 (0.11)	1.08 (0.87-1.35)	.50

[a]Variable of interest: DiGA use without physical therapy (0) versus with physical therapy (1).

[b]OR: odds ratio.

[c]DiGA: digital health application.

Discussion

Principal Findings

This analysis of claims data from 6090 patients with MSDs from the largest German sickness fund revealed that more than half of the DiGA users (3543/6090, 58.2%) received supplementary analog physical therapy. Orthopedic surgeons were the primary prescribers, and patients receiving DiGAs for MSDs were typically females aged between 50 and 60 years. Indicators of chronicity (risk), female gender, higher age and the DiGA used emerged as significant predictors for concomitant physical therapy use.

XSL•FO
RenderX

Comparison With Prior Work

Our findings align with published reports, such as those of the German National Association of Statutory Health Insurances *(Spitzenverband Bund der Krankenkassen – GKV-Spitzenverband)* [17], which describe the current utilization of all DiGAs in Germany on the basis of data from over 374,000 users. In addition, the report of the TK itself includes data from over 19,000 DiGA (2022) and almost 69,000 users in 2024, and the report of the DiGA manufacturers covers data from 41 DiGA applications [20-22]. The finding that nearly 60% of the patients (3543/6090, 58.2%) received additional physical therapy is in line with a survey of 214 insured patients from another German SHI. In that publication, 54.9% reported either current use or expected use of DiGA for MSDs in addition to other health care services [26]. No further claims data studies were currently available for comparison.

The increasing adoption of DHIs, especially for managing chronic illnesses such as MSDs, highlights the need for comprehensive evaluations of their effectiveness in real-world settings. While systematic reviews have underscored the positive treatment effects of DHIs on pain and functionality [11-14], evidence regarding their long-term impact and integration into health care systems remains limited. An observational post-market study of 3629 Vivira users showed positive effects in pain reduction over 2, 4, 8, and 12 weeks in the real-world setting (T_{3628}=5308; P<.001) especially for back pain. However, this result may be biased due to missing outcome data: 51.1% (1853/3629) of patients were lost to follow-up at 2 weeks and 87.4% (3171/3629) at 12 weeks [27]. Notably, a scoping review encompassing 87 studies demonstrated that only 17 studies (19.5%) analyzed general health care utilization of patients using DHIs [28]. Further research based on real-world data, such as in the current analysis, is warranted to complement the clinical trial-based assessment of DHIs [29,30].

Blended Approaches

Blended therapeutic approaches integrating both digital and analog therapies offer promising evidence for optimizing health care resources and improving patient outcomes, particularly in complex cases of MSDs. These care models are relatively new to the German health care system but are evaluated in neighboring European countries. For example, a Dutch randomized controlled study of 208 patients with hip and knee osteoarthritis compared physical therapy alone (on average 12 sessions) to a hybrid of an e-exercise intervention with five face-to-face physical therapy sessions. Physical functioning and free-living physical activity were similar in both groups [31]. Although intervention costs and medication costs in the e-exercise group were significantly lower, the cost-effectiveness from the societal perspective (total costs of osteoarthritis, irrespective of the payer; eg, absenteeism and presenteeism), as well as the total health care costs (intervention costs+primary health care costs+secondary health care costs+medication costs), could not be demonstrated [32]. Blended therapy may help save health care resources, which is especially important in Germany, where provider-specific budgeting and a shortage of physical therapists can limit access to physical therapy [29].

Although the efficacy of blended therapeutic approaches appears comparable to that of analog therapy, the optimal combination of analog and digital components tailored to individual patient needs must be specified. Stepped care models, such as the one described above, which involve initiating treatment with an unguided internet-delivered intervention and subsequently extending it with physical therapist guided exercises, offer a possible approach to determine the optimal balance [31]. This strategy is supported by the findings of our claims data analysis, which showed that patients with indicators of chronicity and chronicity risk were more likely to receive physical therapy in addition to DiGA. In contrast, less complicated cases typically received DiGA as a standalone therapy. Supporting this approach, preliminary results of a multicenter, cluster-randomized controlled trial (by Koppenaal and colleagues [33]) of 208 patients with lower back pain suggest that patients with a high risk of chronicity might benefit particularly from blended therapy approaches compared to analog physical therapy. While that study did not show any difference between the treatment groups regarding physical functioning over a 3-month period, secondary outcomes such as fear-avoidance beliefs and self-reported adherence revealed significant between-group effects in favor of blended therapy [33]. Future studies need to investigate the extent to which these blended care approaches are suitable and relevant for the German health care system.

Requirements for implementing blended therapy approaches in the German health care system include the development of interfaces between digital and analog therapy, as well as a patient-centered, multidisciplinary exchange of information. Currently, for the DiGAs Vivira, Companion patella, and Mawendo, a single nonmandatory follow-up check by the prescribing physician, consisting of a PDF extraction of follow-up data from the app and a discussion with the patient, is intended. Active changes to treatment by the physician or regular provider intervention are currently not included. Additional evidence shows that feedback from health care professionals enhances self-management and adherence to DHIs, especially in musculoskeletal physical therapy [34]. Hence, the Federal Ministry of Health's *(Bundesministerium für Gesundheit – BMG)* digitalization strategy envisages developing and establishing genuinely blended care models in the future [35].

Strengths and Limitations

The DiGA*move* study is one of the first longitudinal real-world evidence analyses investigating DiGA utilization in comparison with conventional health care services. This analysis benefits from extensive data covering over 6000 patient trajectories in up to four years. This large sample allowed for meaningful subgroup comparisons, reflecting real-world care across various health sectors. However, our design lacked a control group of patients with MSDs who did not use DiGA. Due to the inclusion of TK-insured patients only and the inclusion criterion of using a DiGA at least one time, our sample might be biased towards tech-savvy patients who were among the first to try DiGAs in Germany. The generalizability of the findings may be limited because only three DiGAs were included, with Vivira for treating back pain being overrepresented. Incomplete data,

XSL•FO
RenderX

regarding DiGA-related metrics of early prescriptions in 2020, may have impacted results.

Conclusion

Our analysis highlights the essential role of DiGAs in comprehensive care for MSDs. Chronicity indicators emerged as significant factors influencing the concomitant use of digital and analog physical therapy, emphasizing the relevance of blended care approaches for complex cases. When developing and implementing blended care models in practical settings, interfaces between analog and digital therapies, as well as interdisciplinary collaboration among health care professionals, are necessary. Further research should identify specific patient groups that benefit most from blended therapy approaches.

Acknowledgments

We gratefully acknowledge Techniker Krankenkasse for providing the data and the Wissenschaftliches Institut der AOK (WIdO) for granting permission to use their medication database. We also thank the Zentralinstitut für die kassenärztliche Versorgung (Zi) for funding this study. The funder had no role in the study design, data transfer, data analysis, data interpretation, writing of the manuscript or the decision to submit for publication. This work was supported by the Open Access Publication Fund of the University of Bonn.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to protection of data privacy. Access to the data can be obtained from the authors upon reasonable request and with permission from the TK. The data protection concept of the DiGA*move* study must be adhered to.

Authors' Contributions

Conceptualization: SF, US, LK, BW
Data curation: SF, AS, US
Formal analysis: SF, AS, US
Funding acquisition: SF, US, LK, BW
Investigation: SF, AS, US
Methodology: all authors
Project Administration: SF, AS, BW
Resources: US, BW
Software: SF, AS
Supervision: US, LK, BW
Validation: all authors
Visualization: SF, AS
Writing - original draft: SF
Writing - review & editing: all authors

Conflicts of Interest

None declared.

Multimedia Appendix 1
Functionalities of the three digital health applications used by patients with musculoskeletal disorders in our study.
[PDF File, 420 KB - mhealth_v13i1e63935_app1.pdf]

References

1. Musculoskeletal health 2022. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions [accessed 2025-02-21]
2. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: a systematic analysis of the World Health Organization Burden of Diseases Database. Ann Rheum Dis 2019 Jun;78(6):844-848. [doi: 10.1136/annrheumdis-2019-215142] [Medline: 30987966]
3. von der Lippe E, Krause L, Prost M, et al. Prävalenz von Rücken- und Nackenschmerzen in Deutschland. Ergebnisse der Krankheitslast-Studie BURDEN 2020. Prevalence of back and neck pain in Germany. Results from the BURDEN of Disease Study 2020. J Health Monit 2021;6(S3):2-14. [doi: 10.25646/7855]
4. Fuchs J, Pütz F. Prävalenz von Gelenkschmerzen in Deutschland. Prevalence of joint pain in Germany. J Health Monit 2017;2(3):66-71. [doi: 10.17886/RKI-GBE-2017-056]
5. Bevan S. Economic impact of musculoskeletal disorders (MSDs) on work in Europe. Best Pract Res Clin Rheumatol 2015 Jun;29(3):356-373. [doi: 10.1016/j.berh.2015.08.002] [Medline: 26612235]

XSL•FO
RenderX

6. Hayden JA, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. Cochrane Database Syst Rev 2005 Jul 20;2005(3):CD000335. [doi: 10.1002/14651858.CD000335.pub2] [Medline: 16034851]
7. Reimers N, Reimers C. Sportliches Training bei Lumbalgien, Kox- und Gonarthrose sowie Fibromyalgie-Syndrom: Effekt auf die Schmerzen - Eine Literaturübersicht. Physical training for lumbar pain, coxarthrosis, gonarthrosis, and fibromyalgia syndrome: effect on pain - a literature review. Akt Rheumatol 2012 Jun;37(3):174-188. [doi: 10.1055/s-0032-1308962]
8. Bundesärztekammer (BÄK), Kassenärztliche Bundesvereinigung (KBV), Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF). Nationale VersorgungsLeitlinie Nicht-spezifischer Kreuzschmerz: Kurzfassung. 2. Auflage. National Care Guideline for Non-Specific Lower Back Pain. Short Version. AWMF Leitlinienregister. 2017. URL: https://register.awmf.org/assets/guidelines/nvl-007k_S3_Kreuzschmerz_2018-02-abgelaufen.pdf [accessed 2025-05-29]
9. DiGA Verzeichnis. DiGA directory. Federal Institute for Drugs and Medical Devices. 2025. URL: https://diga.bfarm.de/de/verzeichnis [accessed 2025-02-21]
10. Heilmittel-Richtlinie. Remedy guideline. Gemeinsamer Bundesausschuss. 2011. URL: https://www.g-ba.de/downloads/62-492-3676/HeilM-RL_2024-10-29_iK-2025-01-01.pdf [accessed 2025-02-21]
11. Nagel J, Wegener F, Grim C, Hoppe MW. Effects of digital physical health exercises on musculoskeletal diseases: systematic review with best-evidence synthesis. JMIR Mhealth Uhealth 2024 Jan 23;12:e50616. [doi: 10.2196/50616] [Medline: 38261356]
12. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 5;22(6):e15617. [doi: 10.2196/15617] [Medline: 32501277]
13. Valentijn PP, Tymchenko L, Jacobson T, et al. Digital health interventions for musculoskeletal pain conditions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Sep 6;24(9):e37869. [doi: 10.2196/37869] [Medline: 36066943]
14. Lewkowicz D, Slosarek T, Wernicke S, et al. Digital therapeutic care and decision support interventions for people with low back pain: systematic review. JMIR Rehabil Assist Technol 2021 Nov 19;8(4):e26612. [doi: 10.2196/26612] [Medline: 34807837]
15. Fatoye F, Gebrye T, Mbada C, et al. Economic evaluations of digital health interventions for the management of musculoskeletal disorders: systematic review and meta-analysis. J Med Internet Res 2023 Jul 6;25:e41113. [doi: 10.2196/41113] [Medline: 37410542]
16. Interesting facts. Federal Institute for Drugs and Medical Devices. 2025. URL: https://www.bfarm.de/EN/Medical-devices/Tasks/DiGA-and-DiPA/Digital-Health-Applications/Interesting-facts/_node.html [accessed 2025-02-21]
17. Spitzenverband Bund der Krankenkassen. Bericht des GKV-Spitzenverbandes über die Inanspruchnahme und Entwicklung der Versorgung mit digitalen Gesundheitsanwendungen gemäß § 33a Absatz 6 SGB V: Berichtszeitraum: 01.09.2020 - 30.09.2023. Report of the national association of statutory health insurance funds on the utilization and development of digital health applications according to § 33a paragraph 6 SGB V. : Spitzenverband Bund der Krankenkassen; 2024 URL: https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/telematik/digitales/2023_DiGA_Bericht_GKV-Spitzenverband.pdf [accessed 2025-02-21]
18. Schlieter H, Kählig M, Hickmann E, et al. Digital health applications (DiGA) in the area of tension between progress and criticism: discussion paper from the “Digital health” specialist group of the German Informatics Society. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2024 Jan;67(1):107-114. [doi: 10.1007/s00103-023-03804-2] [Medline: 38086924]
19. Arnold M, Hasse F, Püschner F, et al. DiGA-Zwischenbilanz und Entwicklungspotenziale. Interim assessment of digital health applications (DiGA) and development potential. Market Access & Health Policy 2022;3:26-31.
20. DiGA-Report 2022. : Techniker Krankenkasse; 2022 URL: https://www.tk.de/resource/blob/2126090/778e6135918696524cccc1b7be39fd1b/diga-report-data.pdf [accessed 2025-02-21]
21. Woher kommt dieser Protest? Aktuelle Debatte rund um DiGA und ihre Hintergründe. Where does the protest come from? Current debate surrounding digital health applications (DiGA) and their background. Spitzenverband Digitale Gesundheitsversorgung e.V. 2021. URL: https://digitalversorgt.de/wp-content/uploads/2021/08/SVDGV-Beitrag-DiGA-Debatte.pdf [accessed 2025-02-21]
22. DiGA Report II. : Techniker Krankenkasse; 2024 URL: https://www.tk.de/resource/blob/2170850/e7eaa59ecbc0488b415409d5d3a354cf/tk-diga-report-2-2024-data.pdf [accessed 2025-02-21]
23. Freytag A, Thiede M, Schiffhorst G, et al. Versorgungskosten von Rückenschmerzen und die Bedeutung der Schmerzchronifizierung - Ergebnisse einer GKV-Routinedatenanalyse. Treatment costs of back pain and the significance of pain chronification - Results of a statutory health insurance routine data analysis. Gesundh ökon Qual manag 2012 Apr;17(2):79-87. [doi: 10.1055/s-0031-1281578]
24. Gabadinho A, Ritschard G, Müller NS, Studer M. Analyzing and visualizing state sequences in R with TraMineR. J Stat Softw 2011;40(4):1-37. [doi: 10.18637/jss.v040.i04]
25. The R Project for Statistical Computing. R Foundation for Statistical Computing. 2018. URL: https://www.R-project.org/ [accessed 2025-05-22]

XSL•FO
RenderX

26. Grobe TG, Weller L, Braun A, et al. BARMER Arztreport 2024: Digitale Gesundheitsanwendungen - DiGA. BARMER physician report 2024 - digital health application (DiGA). Barmer. 2024. URL: https://www.barmer.de/resource/blob/1255894/d58c1bd7c1a99c78a54d410c490920c8/dl-arztreport-2024-data.pdf [accessed 2025-02-21]
27. Teepe GW, Kowatsch T, Hans FP, Benning L. Postmarketing follow-up of a digital home exercise program for back, hip, and knee pain: retrospective observational study with a time-series and matched-pair analysis. J Med Internet Res 2023 Feb 27;25:e43775. [doi: 10.2196/43775] [Medline: 36848211]
28. Kelly M, Fullen B, Martin D, McMahon S, McVeigh JG. eHealth interventions to support self-management in people with musculoskeletal disorders, “eHealth: It’s TIME”-a scoping review. Phys Ther 2022 Apr 1;102(4):pzab307. [doi: 10.1093/ptj/pzab307] [Medline: 35079826]
29. Weise H, Zenner B, Schmiedchen B, et al. The effect of an app-based home exercise program on self-reported pain pntensity in unspecific and degenerative back pain: pragmatic open-label randomized controlled trial. J Med Internet Res 2022 Oct 28;24(10):e41899. [doi: 10.2196/41899] [Medline: 36215327]
30. Stern AD, Brönneke J, Debatin JF, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health 2022 Mar;4(3):e200-e206. [doi: 10.1016/S2589-7500(21)00292-2] [Medline: 35216754]
31. Kloek CJJ, Bossen D, Spreeuwenberg PM, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a blended physical therapist intervention in people with hip osteoarthritis, knee osteoarthritis, or both: a cluster-randomized controlled trial. Phys Ther 2018 Jul 1;98(7):560-570. [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
32. Kloek CJJ, van Dongen JM, de Bakker DH, Bossen D, Dekker J, Veenhof C. Cost-effectiveness of a blended physiotherapy intervention compared to usual physiotherapy in patients with hip and/or knee osteoarthritis: a cluster randomized controlled trial. BMC Public Health 2018 Aug 31;18(1):1082. [doi: 10.1186/s12889-018-5975-7] [Medline: 30170586]
33. Koppenaal T, Pisters MF, Kloek CJ, Arensman RM, Ostelo RW, Veenhof C. The 3-month effectiveness of a stratified blended physiotherapy intervention in patients with nonspecific low back pain: cluster randomized controlled trial. J Med Internet Res 2022 Feb 25;24(2):e31675. [doi: 10.2196/31675] [Medline: 35212635]
34. Agnew JMR, Hanratty CE, McVeigh JG, Nugent C, Kerr DP. An investigation into the use of mHealth in musculoskeletal physiotherapy: scoping review. JMIR Rehabil Assist Technol 2022 Mar 11;9(1):e33609. [doi: 10.2196/33609] [Medline: 35275089]
35. Digitalisierungsstrategie für das Gesundheitswesen und die Pflege. Digitalization strategy for health care and nursing. Federal Ministry of Health, Germany. 2023. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/D/Digitalisierungsstrategie/BMG_Broschuere_Digitalisierungsstrategie_bf.pdf [accessed 2025-02-21]

Abbreviations

BfArM: Federal Institute for Drugs and Medical Devices (*Bundesinstitut für Arzneimittel und Medizinprodukte*)
DHI: digital health intervention
DiGA: digital health application (*Digitale Gesundheitsanwendung)*
DVG: Digitale-Versorgung-Gesetz
GPS: Good Practice in Secondary Data Analysis
MSD: musculoskeletal disorder
OR: odds ratio
SHI: statutory health insurance
STROSA: Standardisierte Berichtsroutine für Sekundärdaten Analysen
TK: Techniker Krankenkasse

Edited by G Badicu; submitted 04.07.24; peer-reviewed by G Whelan, R Vincent; revised version received 26.02.25; accepted 27.02.25; published 09.06.25.

Please cite as:
Frey S, Schmitz A, Schneider U, Kerkemeyer L, Weltermann B
Combined Use of Digital and Analog Physical Therapy in Patients With Musculoskeletal Disorders and Indicators of Chronicity: German Claims Data Analysis
JMIR Mhealth Uhealth 2025;13:e63935
URL: https://mhealth.jmir.org/2025/1/e63935
doi:10.2196/63935

© Silke Frey, Annika Schmitz, Udo Schneider, Linda Kerkemeyer, Birgitta Weltermann. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.6.2025. This is an open-access article distributed under the terms of the Creative Commons

XSL•FO
RenderX

Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Sociodemographic Differences in Logins and Engagement With the Electronic Health Coach Messaging Feature of a Mobile App to Support Opioid and Stimulant Use Recovery: Results From a 1-Month Observational Study

Lindsey M Filiatreau[1], MPH, PhD; Hannah Szlyk[2], MSSW, PhD; Alex T Ramsey[2], MA, PhD; Erin Kasson[2], MS, MSW; Xiao Li[2], MS; Zhuoran Zhang[2], MA; Patricia Cavazos-Rehg[2], PhD

[1]Division of Infectious Diseases, School of Medicine, Washington University in St Louis, St Louis, MO, United States

[2]Department of Psychiatry, School of Medicine, Washington University in St Louis, St Louis, MO, United States

Corresponding Author:
Lindsey M Filiatreau, MPH, PhD
Division of Infectious Diseases
School of Medicine
Washington University in St Louis
620 S Taylor Ave
St Louis, MO, 63110
United States
Phone: 1 3142737579
Email: flindsey@wustl.edu

Abstract

Background: Mobile health apps can serve as a critical tool in supporting the overall health of uninsured and underinsured individuals and groups who have been historically marginalized by the medical community and may be hesitant to seek health care. However, data on uptake and engagement with specific app features (eg, in-app messaging) are often lacking, limiting our ability to understand nuanced patterns of app use.

Objective: This study aims to characterize sociodemographic differences in uptake and engagement with a smartphone app (uMAT-R) to support recovery efforts in a sample of individuals with opioid and stimulant use disorders in the Greater St. Louis area.

Methods: We enrolled individuals into the uMAT-R service program from facilities providing recovery support in the Greater St. Louis area between January 2020 and April 2022. Study participants were recruited from service project enrollees. We describe the number of logins and electronic health coach (eCoach) messages participants sent in the first 30 days following enrollment using medians and IQRs and counts and proportions of those who ever (vs never) logged in and sent their eCoach a message. We compare estimates across sociodemographic subgroups, by insurance status, and for those who did and did not participate in the research component of the project using Wilcoxon rank-sum tests and Pearson chi-square tests.

Results: Of all 695 participants, 446 (64.2%) logged into uMAT-R at least once during the 30 days following enrollment (median 2, IQR 0-8 logins). Approximately half of those who logged in (227/446) used the eCoach messaging feature (median 1, IQR 0-3 messages). Research participants (n=498), who could receive incentives for app engagement, were more likely to log in and use the eCoach messaging feature compared to others (n=197). Younger individuals, those with higher educational attainment, and White, non-Hispanic individuals were more likely to log in at least once compared to their counterparts. The median number of logins was higher among women, and those who were younger, employed, and not on Medicaid compared to their counterparts. Among those who logged in at least once, younger individuals and those with lower educational attainment were more likely to send at least one eCoach message compared to others.

Conclusions: Mobile apps are a viable tool for supporting individuals in recovery from opioid and stimulant use disorders. However, older individuals, racial and ethnic minorities, and those with lower educational attainment may need additional login support, or benefit from alternative mechanisms of recovery support. In addition, apps may need to be tailored to achieve sustained engagement (ie, repeat logins) among men, and individuals who are older, unemployed, or on Medicaid. Older individuals and

XSL•FO
RenderX

those with higher educational attainment who may be less likely to use eCoach messaging features could benefit from features tailored to their preferences.

(JMIR Mhealth Uhealth 2025;13:e54753) doi:10.2196/54753

KEYWORDS

substance misuse; substance use recovery; opioid use disorder; stimulant use disorder; uptake; engagement; mHealth; digital health intervention; sociodemographic; mobile app; stimulant use; observational study; mobile health; smartphone; St. Louis; eCoach messaging; Wilcoxon rank-sum tests; Pearson chi-square; recovery; app

Introduction

Unmet health care needs in the United States have increased over the past 2 decades despite overall increases in insurance coverage [1]. The most recent national estimates (2017) suggest the proportion of individuals who were unable to see a physician because of cost was over 15% in the population overall and over 30% among uninsured individuals [1]. In the last 15 years, mobile health (mHealth) technologies have been leveraged to address this issue and support improved physical and mental well-being (eg, weight loss, diabetes management, depression symptom management, substance misuse recovery) nationally [2-5]. mHealth apps are often heralded as a critical tool in supporting the mental and physical health of un- and underinsured individuals and groups, such as racial, ethnic, gender, and sexual minorities who have been historically marginalized by the medical community and may be hesitant to seek care from formal care systems [6-10].

Results from numerous randomized controlled trials have demonstrated the efficacy of mHealth tools in improving a range of health outcomes, including weight loss [11], stress reduction [12], smoking cessation [13], and recovery from substance use disorders (SUDs) [4,14]. For example, a systematic review of digital interventions to support people in recovery from SUDs demonstrated that 70%-90% of individuals found these interventions to be useful and easy to use, and more than half of the studies with control groups demonstrated efficacy in improving recovery-related outcomes [4]. Often, mHealth tools integrate a host of features to allow users to engage in diverse activities such as identifying local community-based resources to support their health; goal and reminder setting; and symptom monitoring, each of which is designed to support healthy behaviors [5,15]. Yet, existing findings suggest uptake and engagement with mHealth apps can vary dramatically [5,15]. Moreover, data on the efficacy, acceptability, and utilization of distinct app features is often lacking, thereby limiting our ability to understand what works for whom, when, and under what settings; importantly, this also limits our understanding of which groups of individuals are missed by unique mHealth strategies [16-18]. As we continue to refine mHealth tools to support overall public health and well-being, understanding the profiles of app users, including users of distinct app features such as in-app messaging, can help fill an important gap in the literature.

The national opioid epidemic has reached new heights with over 105,000 overdose deaths reported in 2021– 80,000 of which were opioid-related [19]. To address this continued and increasing public health crisis, the Substance Abuse and Mental Health Services Administration (SAMHSA) developed the State Targeted Response to the Opioid Crisis Grant in 2017, which funded 57 states and territories to support the provision of prevention, treatment, and recovery programs for people with opioid use disorders (OUDs) [20]. In Missouri, these funds were used to increase access to medication-assisted therapy for OUD, provider-focused overdose prevention initiatives, recovery support services for people with OUD, and to improve the sustainability of such programs through policy implementation [20]. One such recovery support tool funded through the Missouri Opioid State Targeted Response grant is a smartphone app, coined uMAT-R, that was designed to support individuals in recovery from OUDs and stimulant use disorders.

In this study, we aimed to characterize uptake of and overall engagement with the uMAT-R app and in-app electronic health coach (eCoach) messaging feature among those linked to services through the Missouri Opioid State Targeted Response grant funding. We also explore sociodemographic differences in uptake and engagement patterns in distinct subgroups of users to elucidate what types of interventions or adaptations might be needed to improve use in distinct groups of individuals.

Methods

Study Procedures

Recruitment for the uMAT-R mobile app intervention program was conducted by study staff in partnership with facilities providing recovery support services for individuals with OUD and other SUDs (eg, SUD treatment programs, Federally Qualified Health Centers) and through the justice system between January 1, 2020, and April 1, 2022. While uMAT-R program staff aimed to foster positive collaborations with recovery facility staff so these individuals would encourage participation in uMAT-R, participation in uMAT-R was voluntary and separate from facility-based recovery services. To be eligible for inclusion in the uMAT-R intervention program, individuals had to: (1) report a history of or current opioid or stimulant misuse; (2) be in treatment at one of the included recruitment facilities; (3) be 18 years or older; (4) be a US resident; (5) be fluent in English; and (6) own or have access to a smartphone with an iOS or Android operating system. Those enrolled in the intervention program were also invited to participate in a 1-month study to ascertain the effect of uMAT-R engagement on recovery outcomes. All individuals enrolled in the service program were provided access to uMAT-R regardless of participation in the research component of the program. Upon enrollment, individuals were provided a unique username and link to access uMAT-R by study staff.

XSL•FO
RenderX

uMAT-R Intervention

The uMAT-R intervention app was developed based on content from the handbook "Decisions in Recovery: Treatment for Opioid Use Disorders" created by SAMHSA, as previously described [21]. This handbook provides information on OUD treatment options, including information regarding costs and potential side effects of distinct treatments (ie, buprenorphine, methadone, and naltrexone) [22]. The app is comprised of several unique features (Figure 1) including over 40 distinct educational courses on topics ranging from "What is Opioid Use Disorder" to "Boundary Setting and Medication for Addiction Treatment" and an extensive directory of community resources throughout Missouri. Additional uMAT-R features to support recovery include goal setting; appointment and medication reminders; sober day tracking; daily motivational quotes; and an in-app messaging feature to communicate with a personal eCoach who is a trained member of the project team and provides near real-time support focused on recovery or essential needs (eg, housing, health care, transportation). The eCoach is also responsible for helping participants navigate any difficulties accessing the app or specific app features.

Figure 1. Example images of the uMAT-R app (home page, messaging feature, educational library, resource directory).

Survey Data

Participants providing verbal or written consent were sent a unique link to an electronic copy of a baseline survey which included questions on sociodemographic background, substance use history, previous recovery efforts, and mental health symptoms. All surveys were self-administered and completed in the Research Electronic Data Capture (REDCap, Vanderbilt University) system, a secure, web-based application.

Uptake and Engagement Data

Uptake and engagement data were ascertained for all individuals assigned a uMAT-R username regardless of participation in the research component of the program. Data included information on the number and time each user logged into the app and the date and time of messages exchanged with an eCoach through the app. We characterized uptake by creating a dichotomous variable to represent any vs no logins. We characterized engagement using three separate measures, a dichotomous variable representing any vs no eCoach messages sent by participants, and two continuous measures representing the total number of logins and eCoach messages sent. Uptake and engagement data were obtained from iTether, the technology company supporting the design and maintenance of the uMAT-R app during the observation period.

Sociodemographic Characteristics

We used self-reported survey data on gender (men vs women), age (<50 vs ≥50 years), race and ethnicity (non-Hispanic White people vs non-White or Hispanic people), educational attainment (<high school diploma vs high school diploma, General Education Development—a high school equivalency exam in the United States, or greater), employment status (unemployed vs employed part- or full-time), housing status (stably housed [living in own or someone else's home or apartment] vs unstably housed [homeless or living in a treatment facility]), and insurance status (private insurance vs no private insurance; Medicaid vs no Medicaid; and insurance coverage vs no insurance coverage) among those who chose to engage in the research component of this service project.

Analyses

Eligible individuals were excluded if their uptake and engagement data were unavailable. Among included individuals, basic uptake and engagement activity was characterized using login and eCoach message data in the 30-day period following enrollment into the service project. Counts and proportions were used to describe app uptake and those who ever versus never used the eCoach messaging feature. Other engagement outcomes related to the total number of logins and eCoach messages sent by participants were characterized using medians and IQRs. Data on the number of eCoach messages sent by participants

XSL•FO
RenderX

were restricted to include individuals who logged into uMAT-R at least once.

We then linked uptake and engagement data with baseline participant survey data for those who enrolled in the research component of the project and completed a baseline survey within 7 days of enrollment (to ensure measurement of exposures of interest preceded outcomes of interest) using unique user IDs assigned by iTether. We explored differences in uptake and engagement among those who did and did not consent to participate in the research component of the program and, among those who completed a baseline survey, differences in uptake and engagement in distinct sociodemographic subgroups. We compared estimates of uptake and engagement for distinct sociodemographic subgroups and the group of individuals who did and did not participate in the research component of the program using Wilcoxon rank-sum tests for continuous measures and Pearson chi-square tests for categorical measures.

Ethical Considerations

For individuals enrolled in the service component of this project who were also interested in participating in the research component, the research staff reviewed the informed consent document via phone or in person. Verbal or written consent was documented for all research participants. To uphold participant privacy and confidentiality, all engagement data was deidentified before analysis. Only members of the study team (eg, eCoaches) involved in participant enrollment and care provision had access to deidentified survey and messaging data. Through a contingency management component of this program, research participants could earn gift card incentives for completing the baseline survey (US $30) and specific activities in the app (eg, send your eCoach a message, review specific in-app course content) over the 1-month study period. All program participants (ie, research and nonresearch uMAT-R users) were informed of contingency management incentive structures at the time of enrollment. This study was approved by the Washington University Institutional Review Board (IRB 201805132, 201910161).

Results

A total of 712 individuals were granted access to the uMAT-R application between January 1, 2020, and April 1, 2022, 702 of whom were recruited from general SUD recovery organizations (98.6%) and 10 of whom were recruited through the justice center (1%). A total of 17 individuals (2%) were excluded from the analysis because their uptake and engagement data were unavailable from iTether. Of the 695 included individuals, 446 (64.2%) logged into the uMAT-R application at least once during the 30 days following enrollment into the service project with a median number of logins of 2 (IQR 0-8; Table 1). The median number of logins among those who ever logged in was 5 (IQR 2-13; Table 1). Approximately half of those who had ever logged into the app (227/446; 50.9%) sent at least one eCoach message. The median number of messages sent among those who ever logged in was 1 (IQR 0-3; Table 1).

Table 1. Engagement in the uMAT-R app to support recovery from opioid and stimulant use disorders among individuals enrolled in the Missouri Opioid State Targeted Response Service Grant project between January 2020 and April 2022.

Engagement metric	All enrolled individuals (N=695)	Enrolled individuals with baseline survey data (n=498)	Enrolled individuals without baseline survey data (n=197)
At least one login, n (%)	446 (64.2)	379 (76.1)[a]	67 (34)[a]
Total logins, median (IQR) in total sample	2 (0-8)	4 (1-11)[a]	0 (0-1)[a]
Total logins, median (IQR) among those who logged in at least once	5 (2-13)	6 (2-16)[a]	1 (1-5)[a]
At least one eCoach message sent, n (% of those who logged in at least once)	227 (50.9)	211 (55.7)[a]	16 (24)[a]
Total eCoach messages sent, median (IQR) in total sample	0 (0-1)	0 (0-2)[a]	0 (0-0)[a]
Total eCoach messages sent, median (IQR) among those who logged in at least once	1 (0-3)	1 (0-4)[a]	0 (0-0)[a]

[a]These data indicate statistical significance at $P<.001$.

Baseline survey data completed within 7 days of enrollment into the service program was available for 498 of the 695 included individuals (71.7%). App uptake and engagement were greater among those who participated in the research component of the project compared to those who did not (Table 1). Of the 498 individuals with survey data, 379 (76.1%) logged into the uMAT-R app at least once during the 30 days following enrollment with a median number of logins of 4 (IQR 1-11; Table 1). The median number of logins among those who ever logged in was 6 (IQR 2-16; Table 1). Approximately 56% of those with survey data who had ever logged into the app (211/379) sent at least one message to their eCoach. The median number of messages sent was 1 (IQR 0-4) among research participants who ever logged in (Table 1).

Of the 498 individuals with baseline survey data, most were women (275/498; 55.2%), less than 50 years of age (435/491; 88.6%), White, non-Hispanic (382/493; 77.5%), unemployed (292/492; 59.3%), and stably housed (361/491; 73.5%; Table 2). Approximately 40% of the population was on Medicaid (211/492; 42.9%), 15% had private insurance (73/492; 15%), and just over half had any form of insurance coverage (275/491; 56%; Table 2).

XSL•FO
RenderX

Table 2. Login data from the uMAT-R application to support recovery from opioid and stimulant use disorders among 498 research participants enrolled in the Missouri Opioid State Targeted Response Service Grant between January 2020 and April 2022, overall and stratified by sociodemographic subgroup.

Sociodemographic characteristics	Total (N=498), n (%)	No logins (n=119), n (%)	At least one (n=379), n (%)	Total (n=497), median (IQR)
Gender[a]				
Men	222 (44.6)	56 (25)	166 (74.8)	3 (0-8)[c]
Women	275 (55.2)	63 (23)	212 (77.1)	5 (1-14)[c]
Age (years)[b]				
<50	436 (88.6)	98 (22)[c]	338 (77.5)[c]	4 (1-11)[c]
≥50	56 (11)	20 (36)[c]	36 (64)[c]	2 (0-6)[c]
Race and ethnicity[b]				
Non-Hispanic White	382 (77.5)	82 (21)[c]	300 (78.5)[c]	4 (1-11)
Non-White or Hispanic	111 (22.5)	37 (33)[c]	74 (67)[c]	2 (0-11)
Educational attainment[b]				
<High school diploma	97 (20)	37 (38)[c]	60 (62)[c]	3 (0-12)
≥High school diploma	396 (80.3)	82 (21)[c]	314 (79.3)[c]	4 (1-10)
Employment[b]				
Unemployed	292 (59.3)	79 (27)	213 (73)	3 (0-10)[c]
Employed	200 (40.7)	40 (20)	160 (80)	5 (1-13)[c]
Housing[b]				
Unstable	130 (26.5)	36 (28)	94 (72)	3 (0-11)
Stable	361 (73.5)	83 (23)	278 (77)	4 (1-11)
Medicaid insurance coverage[b]				
No	281 (57.1)	59 (21)	222 (79)	4 (1-13)[c]
Yes	211 (42.9)	60 (28)	151 (40.5)	3 (0-10)[c]
Private insurance coverage[b]				
No	419 (85.2)	106 (25.3)	313 (74.7)	4 (0-11)
Yes	73 (15)	12 (16)	61 (16)	3 (1-9)
Any insurance coverage[b]				
No	216 (44)	49 (23)	167 (77.3)	4 (1-13)
Yes	275 (56)	69 (25)	206 (74.9)	3 (0-10)

[a]Data suppression: one individual identified as gender nonconforming so engagement data were suppressed.

[b]Missing: age=5; race and ethnicity=5; educational attainment=5; employment=6; housing=7; Medicaid=6; private insurance=5; any insurance=6.

[c]These values indicate statistical significance at $P<.05$.

Overall, uptake (ie, ever vs never logging in) varied by age, race and ethnicity, and educational attainment. Engagement varied by gender, age, race and ethnicity, educational attainment, employment, and Medicaid coverage (Table 2). Specifically, those under 50 (77.5% vs 64.3% among those 50 and older), White, non-Hispanic individuals (78.5% vs 66.7% among non-White or Hispanic individuals), and those with a high school diploma or higher (79.3% vs 61.9% among those with less than a high school diploma) were more likely to log into the app at least once compared to their counterparts (Table 2). The median number of logins was higher among women (median 5, IQR 1-14) compared to men (median 3, IQR 0-8), those under the age of 50 (median 4, IQR 1-11) compared to those aged 50 years and older (median 2, IQR 0-6), employed individuals (median 5, IQR 1-13) compared to unemployed individuals (median 3, IQR 0-10), and those who were not on Medicaid (median 4,

XSL•FO
RenderX

IQR 1-13) compared to those who were (median 3, IQR 0-10; Table 2). Among those who logged into the app at least once, those under the age of 50 were more likely to send at least one message to their eCoach (195/338; 57.7%) compared to those aged 50 years or older (13/36; 36.1%), as were those with lower educational attainment (41/60; 68.3%) compared to those with higher educational attainment (167/314; 53.2%; Table 3). No sociodemographic differences were observed in the median number of eCoach messages sent among those who logged into the app at least once.

Table 3. Use of the electronic health coach messaging feature in the uMAT-R app to support recovery from opioid and stimulant use disorders among 379 research participants enrolled in the Missouri Opioid State Targeted Response Service Grant project between January 2020 and April 2022 who logged into uMAT-R at least once during the 30 days following study enrollment, overall and stratified by sociodemographic subgroup.

Sociodemographic characteristic	Total (N=379), n (%)	No eCoach message sent (n=171), n (%)	At least one eCoach message sent (n=208), n (%)	Total (n=379), median (IQR)
Gender[a]				
Men	166 (43.8)	76 (46)	90 (54)	1 (0-3)
Women	212 (55.9)	91 (43)	121 (57.1)	1 (0-4)
Age (years)[b]				
<50	338 (90.4)	143 (42.3)[c]	195 (57.7)[c]	1 (0-4)
≥50	36 (10)	23 (64)[c]	13 (36)[c]	0 (0-2)
Race and ethnicity[b]				
White, non-Hispanic	300 (80.2)	138 (46)	162 (54)	1 (0-3)
Non-White or Hispanic	74 (20)	28 (38)	46 (62)	1 (0-4)
Educational attainment[b]				
<High school[d] diploma	60 (16)	19 (32)[c]	41 (68)[c]	2 (0-4)
≥High school diploma	314 (84)	147 (46.8)[c]	167 (53.2)[c]	1 (0-3)
Employment[b]				
Unemployed	213 (57.1)	99 (46)	114 (53.5)	1 (0-4)
Employed	160 (42.9)	66 (41)	94 (59)	1 (0-3)
Housing[b]				
Unstable	94 (25)	41 (44)	53 (56)	1 (0-3)
Stable	278 (74.7)	125 (45)	153 (55)	1 (0-4)
Medicaid coverage[b]				
No	222 (58.5)	96 (43)	126 (56.8)	1 (0-4)
Yes	151 (40.5)	70 (46)	81 (54)	1 (0-3)
Private insurance coverage[b]				
No	313 (83.7)	132 (42.2)	181 (57.8)	1 (0-4)
Yes	61 (16)	34 (56)	27 (44)	0 (0-3)
Any insurance coverage[b]				
No	167 (44.8)	65 (39)	102 (61.1)	1 (0-4)
Yes	206 (55.2)	101 (49)	105 (51)	1 (0-3)

[a]Data suppression: one individual identified as gender nonconforming so engagement data were suppressed.

[b]Missing: age=5; race and ethnicity=5; educational attainment=5; employment=6; housing=7; Medicaid=6; private insurance=5; any insurance=6.

[c]These values indicate statistical significance at $P<.05$.

[d]HS: high school.

XSL•FO
RenderX

Discussion

Principal Findings

We explored uptake and engagement in a mobile app to support recovery among individuals who misuse opioids and stimulants in the Greater St. Louis area. Overall, approximately two-thirds of individuals who obtained access to the app logged in at least once within 30 days of gaining access. Of those who logged in, around half sent at least one message to their eCoach through the mobile app messaging feature. Uptake and engagement with the app were higher among those who chose to participate in the research component of this service project, which offered monetary incentives for engagement, compared to those who did not. Variability in uptake or engagement was also observed with respect to gender, age, race and ethnicity, educational attainment, employment, and Medicaid, but not any insurance coverage.

Comparison to Previous Work

A majority of individuals given access to uMAT-R (446/695, 64%) logged into the app at least once within 30 days of being granted access, and approximately half of those who logged in at least once sent at least one in-app message to their eCoach (227/446, 51%), demonstrating the short-term viability of this app as a resource to those in recovery for opioid and stimulant use disorders in the Greater St. Louis area. Broadly, uptake and engagement with mobile apps to support healthy behaviors, including SUD recovery, have been shown to vary widely [5,15,23-29]. For example, in a study in which nearly 2000 individuals with an SUD were offered access to a smartphone app to support substance misuse recovery, just 40% of participants downloaded the app, and just 18% engaged with the app [24]. In another study, patients who were given access to the Addiction-Comprehensive Health Enhancement Support System, a smartphone app designed to provide continuing care to patients with alcohol use disorders, used the application on 40% of the days during the 8-month intervention period [29]. On the other hand, in a smaller pilot study (n=25) of the HOPE application, a smartphone app designed to support medication for OUD treatment during COVID-19 lockdowns, 88% of participants used a provider messaging feature during the first month following enrollment into the study [30]. While messaging use fell to approximately 50% by month six, 76% were still engaging with other app features at this time point [30]. Acceptable levels of early engagement with uMAT-R suggest user habits should be explored longitudinally to assess how engagement patterns might decline or otherwise change over time.

Individuals who chose to participate in the research component of this project were more likely to log in to and engage with uMAT-R compared to those who did not participate. Studies exploring the variability in the app or digital intervention engagement between those who are and are not under observation (eg, those who are and are not enrolled in a research study) have also demonstrated variability [31]. For example, in a study exploring engagement with the See Me Smoke-Free mHealth app, research participants were equally as likely as nonresearch participants who installed the app to open the app and update smoking cessation dates but more likely to engage with audio features within the app [31]. While our findings are in line with a phenomenon originally coined the “Hawthorne effect,” which, broadly, posits that people under observation behave differently than those who are not under observation [32], these results may also be attributable to the fact that those willing to enroll in research may differ from those who are unwilling to do so (eg, historically marginalized groups who have been systematically oppressed or harmed by the medical or research community may be less likely to enroll in research and less likely to take up or engage with mHealth apps) [31-33]. Alternatively, incentives for app engagement may have increased uptake and use among those enrolled in the research component of this program, as has been documented in other work [33]. While incentives and related forms of contingency management are common in usability studies and were provided for this real-world substance use recovery support project through funding from the Missouri Department of Health, the long-term feasibility of incorporating incentives for app or intervention use or engagement in other real-world settings should be considered [34].

Among those with baseline survey data, uptake and the types and degree of engagement varied by sociodemographic subgroups, in line with findings from numerous other health promotion mobile app studies [35-37]. Specifically, we observed variation in uptake by age, race and ethnicity, and educational attainment, and variation in engagement by gender, age, race and ethnicity, educational attainment, employment, and Medicaid coverage specifically. Though some discrepancies in findings have been observed in the literature [35,36,38], a growing body of research supports our results that older individuals and men are less likely to engage with mHealth apps compared to their counterparts [35-37]. It is also well-documented in the literature that those with lower (versus higher) educational attainment are less likely to engage with mHealth apps [35-37]; while we saw a similar trend in the uptake of uMAT-R, the opposite trend was observed in eCoach messaging among those who logged in (ie, those with lower educational attainment were more likely to send at least one message to their eCoach compared to those with higher educational attainment). It is possible those with lower educational attainment had less access to resources and were, therefore, more inclined to seek support and resources from an eCoach compared to those with higher educational attainment who logged in, or that lifestyle differences between these groups influenced uptake and engagement in distinct ways. Qualitative data exploring why these differences emerged is needed. Racial and ethnic differences in app use have also been observed in the literature (though less commonly), with White, non-Hispanic individuals more likely to demonstrate sustained engagement in mHealth apps compared to others [35,39]. Differences in uptake of and engagement with mHealth apps by employment and insurance coverage have been less commonly explored and documented [35,40].

Ultimately, the variation in findings related to sociodemographic correlates of mHealth app uptake and engagement, coupled with other research related to mHealth app uptake and engagement over time [10,41,42] suggests that mobile apps can be tailored

XSL•FO
RenderX

to increase acceptability, appropriateness, or usability of apps, or improve app engagement in a range of sociodemographic subgroups. For example, ecological momentary assessment was found to be feasible in older adults when it was designed to be responsive to the unique needs and preferences of that specific subgroup of individuals [42]. User co-design (eg, human-centered design methods and other methods for incorporating end-users in app development) [43-46] has shown promise as a method for improving engagement in substance-use recovery apps across a range of settings. The success of these methods has been well-documented [47-49] and should be used to further refine uMAT-R to increase sustained and comprehensive engagement in older individuals, men, racial and ethnic minorities, those with higher educational attainment, unemployed individuals, and those on Medicaid who either logged in less frequently or were less likely to use the eCoach messaging feature of uMAT-R. Potential alterations may include more comprehensive or geographically diverse resource information; participant-preferred or option-based rewards or incentives for sustained engagement; use of eCoaches with diverse specialties (eg, cognitive behavioral therapy, dialectical behavior therapy) or backgrounds (eg, racial and ethnic minorities, those with personal substance misuse recovery experience, men) who can be individually paired with participants based on participant preferences or needs; and personalized lessons or highlighted app features based on individual participant preferences or needs.

Alternative opioid and stimulant misuse recovery support approaches and interventions (eg, peer navigation) or additional support in downloading, logging into, or accessing specific features of a mHealth app may also be needed to support older individuals, non-White or Hispanic individuals, and those with lower educational attainment who were less likely to ever log in to uMAT-R compared to their counterparts. In qualitative work exploring the implementation of a substance use recovery support mobile phone app, Lord and colleagues identified several facilitators of the implementation and uptake of the app in 4 service settings. Facilitators included staff training promoting organizational awareness about the app and emphasizing app priority as a clinical care tool; client encouragement related to app testing and use; client expectation setting about app use; and use of peer coaches and recurrent client-centered messaging to promote engagement [50]. This suggests partnering more closely with staff at existing recruitment centers may improve uptake and engagement outcomes. A systematic review of barriers and facilitators to user engagement with digital mental health interventions also found that individuals were more likely to uptake or engage with these types of interventions if they were trained on how to use them or received automated or therapist-delivered reminders to use the intervention [51]. Such strategies could also be used to increase the uptake, use, and usability of uMAT-R as well.

Limitations

This study has limitations worth nothing. First, sociodemographic data were missing for approximately 30% of individuals who were granted access to the uMAT-R application but chose not to participate in the research component of the project. Second, variability in-app engagement was observed with respect to participation in the research component with less engagement among those who did not participate, potentially due to incentives for engagement among those enrolled in the research component of the program. As a result, the generalizability of findings regarding sociodemographic differences in engagement is limited, with findings also suggesting incentives or contingency management may be a necessary component to ensure engagement in mHealth apps aimed to support opioid and stimulant misuse recovery efforts. Third, other metrics of mobile application engagement (eg, duration of engagement, content accessed) were not adequately captured by iTether, the technology company supporting the design and maintenance of the uMAT-R application during the observation period. Fourth, we only explored engagement during a relatively brief follow-up period (ie, 1 month following enrollment into the program) and engagement has been shown to decline over time, potentially indicating estimates reported here are a “best-case scenario.” Finally, end-user perspectives on other app features that could have influenced app uptake and engagement (eg, acceptability, appropriateness, usability) were not measured, nor were reasons for engagement in substance use treatment.

Future Directions

Future studies exploring mobile app engagement should aim to capture diverse perspectives on overall app design, reasons for enrollment into recovery programs and uMAT-R specifically, and engagement metrics over an extended follow-up period (eg, one year) to establish a more holistic picture of what works for whom, when, and for how long.

Conclusions

A majority of individuals granted access to the uMAT-R application logged in at least once in the first 30 days and of those who logged in, approximately half sent at least one in-app message to an eCoach. This suggests mobile apps, and uMAT-R in particular, may be a viable tool for supporting individuals in recovery from opioid and stimulant use disorders in Missouri. However, older individuals, racial and ethnic minorities, and those with lower educational attainment may be less likely to initiate the use of such interventions and may benefit from alternative mechanisms of substance use recovery support. Moreover, older individuals, men, racial and ethnic minorities, those with higher educational attainment, unemployed individuals, and those on Medicaid may be less likely to fully engage with such tools, or to engage over a sustained period and may benefit from app features tailored more specifically to their wants and needs. Incentives might also play an important role in increasing uptake and engagement with mHealth apps for improving SUD outcomes. Future studies should aim to explore how uptake and engagement in these types of tools vary over longer periods of time, how incentives influence app use, and the effect of engagement on recovery outcomes overall and in distinct socioeconomic subgroups.

XSL•FO
RenderX

Acknowledgments

The authors would like to acknowledge Andrea Fentem, Shawn Fraine, and Amanda Montayne for their support of this project. This work was supported by the National Institutes of Health (grant numbers T37 MD014218, KL2 TR002346, K02 DA043657, and R25 DA037190) and the Substance Abuse and Mental Health Services Administration (grant numbers 1H79TI080271 and 1H79TI081697).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the sensitive nature of this research but are available from the corresponding author on reasonable request.

Authors' Contributions

LMF contributed to conceptualization, methodology, data curation, formal analysis, writing – original draft, and writing – review and editing. HS performed project administration, supervision, and writing – review and editing. ATR contributed to conceptualization and writing – review and editing. EK conducted project administration and writing – review and editing. XL handled data curation and writing – review and editing. ZZ performed investigation, project administration, and writing – review and editing. PC-R managed funding acquisition, conceptualization, and writing – review and editing.

Conflicts of Interest

HS is a paid consultant for Google Health and PCR is a paid consultant for Rissana, LLC, and PredictView.

References

1. Hawks L, Himmelstein DU, Woolhandler S, Bor DH, Gaffney A, McCormick D. Trends in unmet need for physician and preventive services in the United States, 1998-2017. JAMA Intern Med 2020;180(3):439-448 [FREE Full text] [doi: 10.1001/jamainternmed.2019.6538] [Medline: 31985751]
2. Messner EM, Probst T, O'Rourke T, Stoyanov S, Baumeister H. mHealth Applications: Potentials, Limitations, Current Quality and Future Directions. Cham: Springer; 2019.
3. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012;17 Suppl 1:5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]
4. Nesvåg S, McKay JR. Feasibility and effects of digital interventions to support people in recovery from substance use disorders: systematic review. J Med Internet Res 2018;20(8):e255 [FREE Full text] [doi: 10.2196/jmir.9873] [Medline: 30139724]
5. Neale J, Bowen AM. Lessons for uptake and engagement of a smartphone app (SURE Recovery) for people in recovery from alcohol and other drug problems: interview study of app users. JMIR Hum Factors 2022;9(1):e33038 [FREE Full text] [doi: 10.2196/33038] [Medline: 35258474]
6. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR mHealth uHealth 2018;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
7. Friis-Healy EA, Nagy GA, Kollins SH. It is time to REACT: opportunities for digital mental health apps to reduce mental health disparities in racially and ethnically minoritized groups. JMIR Ment Health 2021;8(1):e25456 [FREE Full text] [doi: 10.2196/25456] [Medline: 33406050]
8. Bowen D, Jabson J, Kamen C. mHealth: an avenue for promoting health among sexual and gender minority populations? Mhealth 2016;2:36 [FREE Full text] [doi: 10.21037/mhealth.2016.09.01] [Medline: 28293609]
9. Johnston DC, Mathews WD, Maus A, Gustafson DH. Using Smartphones to Improve Treatment Retention Among Impoverished Substance-Using Appalachian Women: A Naturalistic Study. London, England: SAGE Publications; 2019:13.
10. Lunn MR, Capriotti MR, Flentje A, Bibbins-Domingo K, Pletcher MJ, Triano AJ, et al. Using mobile technology to engage sexual and gender minorities in clinical research. PLoS One 2019;14(5):e0216282 [FREE Full text] [doi: 10.1371/journal.pone.0216282] [Medline: 31048870]
11. Hales S, Turner-McGrievy GM, Wilcox S, Fahim A, Davis RE, Huhns M, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: a randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
12. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app 'Calm' to reduce stress among college students: randomized controlled trial. JMIR mHealth uHealth 2019;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
13. Pallejà-Millán M, Rey-Reñones C, Barrera Uriarte ML, Granado-Font E, Basora J, Flores-Mateo G, et al. Evaluation of the Tobbstop mobile app for smoking cessation: cluster randomized controlled clinical trial. JMIR Mhealth Uhealth 2020;8(6):e15951 [FREE Full text] [doi: 10.2196/15951] [Medline: 32589153]

XSL•FO
RenderX

14. Leightley D, Williamson C, Rona RJ, Carr E, Shearer J, Davis JP, et al. Evaluating the efficacy of the Drinks:Ration mobile app to reduce alcohol consumption in a help-seeking military veteran population: randomized controlled trial. JMIR mHealth uHealth 2022;10(6):e38991 [FREE Full text] [doi: 10.2196/38991] [Medline: 35724966]
15. Patterson Silver Wolf DA, Ramsey AT, Epstein J, Beeler-Stinn S, Deer AAB. Bridges to sobriety: testing the feasibility and acceptability of a mobile app designed to supplement an adolescent substance use disorder treatment program. Clin Soc Work J 2022;50(3):308-315 [FREE Full text] [doi: 10.1007/s10615-020-00765-w] [Medline: 36420447]
16. Hendrikoff L, Kambeitz-Ilankovic L, Pryss R, Senner F, Falkai P, Pogarell O, et al. Prospective acceptance of distinct mobile mental health features in psychiatric patients and mental health professionals. J Psychiatr Res 2019;109:126-132. [doi: 10.1016/j.jpsychires.2018.11.025] [Medline: 30530207]
17. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
18. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
19. Ahmad F, Cisewski J, Rossen L, Sutton P. Provisional drug overdose death counts. CDC. 2022. URL: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm [accessed 2024-02-21]
20. High PM, Marks K, Robbins V, Winograd R, Manocchio T, Clarke T, et al. State targeted response to the opioid Crisis grants (opioid STR) program: preliminary findings from two case studies and the national cross-site evaluation. J Subst Abuse Treat 2020;108:48-54. [doi: 10.1016/j.jsat.2019.06.008] [Medline: 31303359]
21. Cavazos-Rehg PA, Krauss MJ, Costello SJ, Ramsey AT, Petkas D, Gunderson S, et al. Delivering information about medication assisted treatment to individuals who misuse opioids through a mobile app: a pilot study. J Public Health (Oxf) 2020;42(1):149-154 [FREE Full text] [doi: 10.1093/pubmed/fdy207] [Medline: 30445639]
22. Decisions in recovery: treatment for opioid use disorders. SAMHSA Publications and Digital Products. URL: https://store.samhsa.gov/product/decisions-recovery-treatment-opioid-use-disorders/sma16-4993 [accessed 2024-04-14]
23. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
24. Wilde JA, Zawislak K, Sawyer-Morris G, Hulsey J, Molfenter T, Taxman FS. The adoption and sustainability of digital therapeutics in justice systems: a pilot feasibility study. Int J Drug Policy 2023;116:104024 [FREE Full text] [doi: 10.1016/j.drugpo.2023.104024] [Medline: 37086698]
25. Richterman A, Ghadimi F, Teitelman AM, Moore K, Acri T, North H, et al. Acceptability and feasibility of a mobile phone application to support HIV pre-exposure prophylaxis among women with opioid use disorder. AIDS Behav 2023;27(10):3460-3467 [FREE Full text] [doi: 10.1007/s10461-023-04060-w] [Medline: 37036593]
26. Wang G, Wu B, Chen J, Yu G, Lin D, Wang G, et al. A novel mHealth App (RyPros) for prostate cancer management: an accessibility and acceptability study. Transl Androl Urol 2021;10(10):3723-3736 [FREE Full text] [doi: 10.21037/tau-21-459] [Medline: 34804816]
27. Xiong X, Braun S, Stitzer M, Luderer H, Shafai G, Hare B, et al. Evaluation of real-world outcomes associated with use of a prescription digital therapeutic to treat substance use disorders. Am J Addict 2023;32(1):24-31 [FREE Full text] [doi: 10.1111/ajad.13346] [Medline: 36264211]
28. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR mHealth uHealth 2020;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
29. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
30. Hodges J, Waselewski M, Harrington W, Franklin T, Schorling K, Huynh J, et al. Six-month outcomes of the HOPE smartphone application designed to support treatment with medications for opioid use disorder and piloted during an early statewide COVID-19 lockdown. Addict Sci Clin Pract 2022;17(1):16 [FREE Full text] [doi: 10.1186/s13722-022-00296-4] [Medline: 35255965]
31. Schmidt CA, Romine JK, Bell ML, Armin J, Gordon JS. User participation and engagement with the See Me Smoke-Free mHealth app: prospective feasibility trial. JMIR mHealth uHealth 2017;5(10):e142 [FREE Full text] [doi: 10.2196/mhealth.7900] [Medline: 28993302]
32. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. BMJ 2015;351:h4672 [FREE Full text] [doi: 10.1136/bmj.h4672] [Medline: 26341898]
33. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. JMIR mHealth uHealth 2016 Mar 24;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
34. Nuijten R, Van Gorp P, Khanshan A, Le Blanc P, Kemperman A, van den Berg P, et al. Health promotion through monetary incentives: evaluating the impact of different reinforcement schedules on engagement levels with a mHealth app. Electronics 2021;10(23):2935. [doi: 10.3390/electronics10232935]

XSL•FO
RenderX

35. Szinay D, Forbes CC, Busse H, DeSmet A, Smit ES, König LM. Is the uptake, engagement, and effectiveness of exclusively mobile interventions for the promotion of weight-related behaviors equal for all? A systematic review. Obes Rev 2023;24(3):e13542. [doi: 10.1111/obr.13542] [Medline: 36625062]
36. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
37. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
38. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of user engagement with mobile- and web-delivered self-care interventions for adults with T2DM: a review of the literature. Curr Diab Rep 2016;16(7):66 [FREE Full text] [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
39. Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an incentive-based mHealth app: process evaluation of the carrot rewards app. JMIR mHealth uHealth 2017;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]
40. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/?tabId=tab-cf1ede57-7a4b-489e-8bcb-de1f4a6e643c [accessed 2023-11-12]
41. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
42. Ramsey AT, Wetherell JL, Depp C, Dixon D, Lenze E. Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. J Technol Hum Serv 2016;34(2):209-223 [FREE Full text] [doi: 10.1080/15228835.2016.1170649] [Medline: 27683018]
43. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
44. Bartholomew Eldrigde LK, Parcel GS, Kok G, Gottlieb NH, Schaalma H, Markham C, et al. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco: Jossey-Bass/Wiley; 2016.
45. Boy GA. Orchestrating Human-Centered Design. London: Springer; 2013.
46. Waselewski ME, Flickinger TE, Canan C, Harrington W, Franklin T, Otero KN, et al. A mobile health app to support patients receiving medication-assisted treatment for opioid use disorder: development and feasibility study. JMIR Form Res 2021;5(2):e24561 [FREE Full text] [doi: 10.2196/24561] [Medline: 33620324]
47. Saunders CH, Sierpe A, Stevens G, Elwyn G, Cantrell M, Engel J, CONFIDENT Study Long-Term Care Partners, et al. Co-development of a web application (COVID-19 Social Site) for long-term care workers ('Something for Us'): user-centered design and participatory research study. J Med Internet Res 2022;24(9):e38359 [FREE Full text] [doi: 10.2196/38359] [Medline: 35926074]
48. Hardy A, Wojdecka A, West J, Matthews E, Golby C, Ward T, et al. How inclusive, user-centered design research can improve psychological therapies for psychosis: development of SlowMo. JMIR Ment Health 2018;5(4):e11222 [FREE Full text] [doi: 10.2196/11222] [Medline: 30518514]
49. Hemingway C, Baja ES, Dalmacion GV, Medina PMB, Guevara EG, Sy TR, et al. Development of a mobile game to influence behavior determinants of HIV service uptake among key populations in the Philippines: user-centered design process. JMIR Serious Games 2019;7(4):e13695 [FREE Full text] [doi: 10.2196/13695] [Medline: 31859673]
50. Lord S, Moore SK, Ramsey A, Dinauer S, Johnson K. Implementation of a substance use recovery support mobile phone app in community settings: qualitative study of clinician and staff perspectives of facilitators and barriers. JMIR Ment Health 2016;3(2):e24 [FREE Full text] [doi: 10.2196/mental.4927] [Medline: 27352884]
51. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]

Abbreviations

eCoach: electronic health coach
OUD: opioid use disorders
SAMHSA: Substance Abuse and Mental Health Services Administration
SUD: substance use disorders

XSL•FO
RenderX

Edited by L Buis; submitted 20.11.23; peer-reviewed by C Saunders, M Waselewski, J Hodges; comments to author 13.02.24; revised version received 20.05.24; accepted 06.02.25; published 10.04.25.

Please cite as:
Filiatreau LM, Szlyk H, Ramsey AT, Kasson E, Li X, Zhang Z, Cavazos-Rehg P
Sociodemographic Differences in Logins and Engagement With the Electronic Health Coach Messaging Feature of a Mobile App to Support Opioid and Stimulant Use Recovery: Results From a 1-Month Observational Study
JMIR Mhealth Uhealth 2025;13:e54753
URL: https://mhealth.jmir.org/2025/1/e54753
doi:10.2196/54753
PMID:

©Lindsey M Filiatreau, Hannah Szlyk, Alex T Ramsey, Erin Kasson, Xiao Li, Zhuoran Zhang, Patricia Cavazos-Rehg. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Prevention Needs and Target Behavior Preferences in an App-Based Addiction Prevention Program for German Vocational School Students: Cluster Randomized Controlled Trial

Diana Guertler[1,2,3], PhD; Elaine Kraft[1], MD; Dominic Bläsing[4], PhD; Anne Möhring[2,3], PhD; Christian Meyer[1,2], PhD, Prof Dr; Hannah Schmidt[5,6], PhD; Florian Rehbein[7], PhD, Prof Dr; Merten Neumann[8], Dipl Psych; Arne Dreißigacker[8], Dipl Soz; Anja Bischof[5], PhD; Gallus Bischof[5], PhD; Svenja Sürig[5], MSc; Lisa Hohls[5], MSc; Susanne Wurm[1], PhD, Prof Dr; Stefan Borgwardt[5], MD, Prof Dr Med; Severin Haug[9], PhD; Hans-Jürgen Rumpf[5], PhD, Prof Dr

[1]Department of Prevention Research and Social Medicine, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
[2]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany
[3]Department of Methods of Community Medicine, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
[4]Helmholtz Institute for One Health, Greifswald, Germany
[5]Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany
[6]Department of Pediatrics, University of Lübeck, Lübeck, Germany
[7]Department of Social Work, FH Münster University of Applied Sciences, Münster, Germany
[8]Criminological Research Institute of Lower Saxony (KFN), Hannover, Germany
[9]Swiss Research Institute for Public Health and Addiction, University of Zürich, Zürich, Switzerland

Corresponding Author:
Diana Guertler, PhD
Department of Methods of Community Medicine
Institute for Community Medicine
University Medicine Greifswald
Walther-Rathenau-Str. 48
Greifswald, 17475
Germany
Phone: 49 3834867765
Email: diana.guertler@med.uni-greifswald.de

Abstract

Background: Vocational school students exhibit a high prevalence of addictive behaviors. Mobile phone–based prevention programs targeting multiple addictive behaviors and promoting life skills are promising. Tailoring intervention content to participants' preferences, such as allowing them to choose behavior modules, may increase engagement and efficacy. There is limited understanding of how personal characteristics relate to module choices.

Objective: This study examined the prevention needs of German vocational school students as well as their prevention preferences through self-determined module choice in the multibehavior app-based addiction prevention program ready4life.

Methods: A 2-arm cluster randomized controlled trial recruited German vocational school students aged ≥16 years. Among 376 classes from 35 schools, ready4life was introduced during a school lesson. Students were invited to download the ready4life app and completed an anonymous screening with individualized risk and competence feedback in the form of a traffic light system. Informed consent was provided by 2568 students. Intervention classes received individual app-based coaching with weekly chat contacts with a conversational agent over 4 months. They could choose 2 of 6 modules: alcohol, tobacco, cannabis, social media and gaming, stress, and social competencies. Control group classes received a link to health behavior information and could access coaching after 12 months.

Results: Prevention need was high. For 86.2% (2213/2568), ≥2 risks were reported based on yellow or red traffic light feedback. Within the intervention group, stress (818/1236, 66.2%) and social media and gaming (625/1236, 50.6%) were the most chosen topics, followed by alcohol (360/1236, 29.1%), social competencies (306/1236, 24.8%), tobacco (232/1236, 18.8%), and cannabis (131/1236, 10.6%). Module choices closely aligned with received traffic light feedback among those with 1 or 2 risks. Multilevel regression models showed that women were significantly more likely to choose the stress module (odds ratio [OR] 2.38, 95% CI

XSL•FO
RenderX

1.69-3.33; *P*<.001); men preferred social media and gaming (OR 0.52, 95% CI 0.40-0.69; *P*<.001), alcohol (OR 0.50, 95% CI 0.37-0.67; *P*<.001), and cannabis (OR 0.37, 95% CI 0.21-0.63; *P*<.001) when holding age, educational track, and prevention need for the corresponding behavior constant. Younger students were significantly more likely to choose the cannabis module (OR 0.81, 95% CI 0.74-0.90; *P*<.001). Educational track also influenced module choice (eg, those with a lower educational level were more likely to choose alcohol and cannabis, suggesting a positive equity impact). Students' prevention needs significantly influenced choice of the module (eg, higher alcohol consumption increased the likelihood of choosing the alcohol module; OR 1.31, 95% CI 1.20-1.43; *P*<.001).

Conclusions: Our study confirms vocational school students' high prevention needs regarding addictive behaviors. Students' module choices were highly congruent to their demonstrated needs, with most students being interested in the stress module. Module choice also differed by age, gender, and educational track.

Trial Registration: German Clinical Trials Register DRKS00022328; https://drks.de/search/en/trial/DRKS00022328

International Registered Report Identifier (IRRID): RR2-10.1024/0939-5911/a000811

(JMIR Mhealth Uhealth 2025;13:e59573) doi:10.2196/59573

KEYWORDS

eHealth; computer tailoring; multiple addictive behaviors; vocational school students; module choice

Introduction

Addictive Behaviors in Vocational School Students

Adolescence and young adulthood are periods of life associated with a number of challenges, including entering the world of work and increasing independence from parents. Addictive behaviors such as consumption of alcohol, tobacco, and cannabis [1] but also problematic internet use [2] are particularly widespread in this phase of life.

Many adolescents and young adults in Germany attend vocational schools to start an apprenticeship or attend vocational grammar school or vocational preparation courses. Vocational schools are promising settings for prevention efforts [3], especially as there is evidence that students in vocational schools report even higher rates of substance use than other adolescents and young adults. For example, in a survey of 5688 German vocational school students, the reported smoking rate was more than twice as high as that of their peers in the general population [4]. Representative data from German vocational school students in Mecklenburg-Western Pomerania also show particularly high rates of problematic alcohol consumption (62%), smoking (62%), and cannabis use (26%). The study [5] also provided a first insight into the prevalence of media-related problems among German vocational school students. Problematic internet use was reported by 17% of vocational school students, problematic gaming was reported by 10%, and potential smartphone addiction was reported by 17% [5]. This is supported by findings from an intervention trial to reduce problematic use of the internet in vocational school students [6].

In vocational school settings, achieving prevention goals typically requires a shift away from primary prevention strategies. Instead, indicated prevention programs are considered more appropriate as the vast majority of vocational school students have already initiated addictive behavior. A high proportion of vocational school students also engage in multiple health risk behaviors, especially multiple addictive behaviors [5,7]. However, they differ widely in their risk profiles (ie, in the combinations and number of addictive behaviors). Vocational school students also differ in other factors that may contribute to addictive behaviors, such as low social competence or high work stress. This means that complex interventions are needed to address different risk profiles.

Tailored Mobile Interventions

Mobile interventions are best suited to address various risk profiles as they can be easily adapted to the needs of the users and allow for timely and anonymous participation. Mobile interventions have been shown to be effective in changing multiple addictive behaviors in vocational school students [8,9].

Mobile interventions have the advantage of automated delivery of highly individualized content through computer tailoring. Tailoring is a key component in ensuring the effectiveness of an intervention [10]. With tailoring, feedback that is relevant to the individual is derived from an individual assessment based on prespecified rules. Examples include content or wording tailored to users' gender or motivation to change. It has been shown that tailored feedback compared to generic feedback is more likely to be processed and remembered [11]. In addition to automatic computer tailoring, some interventions include forms of self-tailoring, such as allowing participants to choose the medium [12], dose [13], or content (eg, specific behavioral targets or modules [14-17]). In line with theoretical frameworks such as self-determination theory, self-tailoring is assumed to increase autonomy [18].

Although mobile health interventions are effective, adherence is a major challenge [19], including approaches among vocational school students [20]. Low adherence may partly explain the lack of long-term effects and small effect sizes in mobile health interventions [21]. Allowing participants the autonomy to choose which behavioral modules to engage with could positively influence engagement and efficacy [22,23]. Within the vocational context, offering a choice of intervention modules could be valuable, particularly as many vocational school students exhibit multiple addictive behaviors [5,7]. Attempting to address all these behaviors simultaneously might be overwhelming [24], especially considering the decreasing intention to change all behaviors as their number increases [25].

XSL•FO
RenderX

Mechanisms of Module Choice

While some research has explored participants' module choices in multibehavior interventions [14,15,26-31], there is limited understanding of the underlying choice mechanisms, such as which personal characteristics are related to choice. Investigating this is crucial as diverse subgroups may use interventions differently and some may not use it as recommended (eg, not choosing the recommended modules). Such insights can inform tailored strategies to align interventions with participant needs, reducing nonadherence and enhancing intervention exposure and effectiveness.

Most of the evidence on module choice comes from multiple-behavior interventions in adults. These studies have shown that certain modules are preferred over others, for example, physical activity over diet [14,15,26]. There is also consistency that health-promoting behaviors such as physical activity and diet are preferred over substance-related modules [14,15,26-31]. In total, 2 studies among vocational school students reported on preferences within multiple–addictive behavior interventions. They showed that stress and non–substance-related behaviors such as social media use were chosen more often than substance-related behaviors such as smoking, alcohol or cannabis use [8,17].

With regard to potential determinants of module choice, the results are less consistent [14-16,27] and are likely to vary depending on the number and topics of available behavior modules to choose from, whether module recommendations were given, and whether any restrictions were placed on module choice (eg, how many modules could be chosen or whether modules that were not recommended could be chosen). Providing feedback or recommending modules may enhance the alignment between participants' module choice and their prevention needs [30]. Studies show that people are more inclined to choose modules that were recommended to them based on their demonstrated risk profile [17,29]. Consequently, most previous studies have used feedback or recommendations to guide module choice [14-17,26,29,32], with only a few exceptions [27,28,30,33].

In online multiple-behavior interventions for adults addressing a range of health-promoting behaviors and substance use, choice of the physical activity– or diet-related modules was higher among women [14] and those with healthier lifestyles [15]. In the study by Schulz et al [15], choice of the alcohol module was linked to being a man, lower income, and higher psychosocial distress, whereas medium (vs high) educational level, lower quality of life, and lower psychosocial distress were linked to choice of the smoking module. Reinwand et al [34] showed that adherence to module recommendations in online multibehavior interventions (ie, starting all recommended modules vs fewer) also depended on individual characteristics. Individuals were more likely to start all the recommended modules if they were older, a woman, unemployed, ill, and in a relationship.

The main gap in the existing research is that no study has reported on the determinants of module preferences among vocational school students or statistically analyzed the alignment between prevention needs and module preferences in this group. This is particularly important given that a high proportion of vocational school students engage in multiple addictive behaviors. Furthermore, although there have been previous studies on the prevalence of addictive behaviors among German vocational school students [5,35], these studies did not analyze prevention needs in terms of life skills and their determinants. This has only been investigated in a Swiss sample [17].

Given the fact that lower educational attainment among vocational school students is associated with higher prevalence [35] and clustering [5] of addictive behaviors, as well as lower engagement with mobile interventions [36], it is particularly important to analyze educational track as a potential determinant of module choice.

Aims

This study examined the prevention needs of German vocational school students as well as their module choices within the app-based addiction prevention intervention ready4life. The ready4life intervention offers 6 behavior modules: alcohol, tobacco, cannabis, social media and gaming, stress, and social competencies. After receiving feedback on their individual risks and competencies in the form of a traffic light system to guide choice, students could freely choose 2 out of the 6 modules to be coached in. In terms of module choice, we aimed to examine (1) how often each intervention module was chosen, (2) how module choice aligned with the traffic light feedback received, and (3) whether individual characteristics (age, gender, educational track, and prevention needs) were associated with module choice. In addition, the relationship between sociodemographic factors and module choice is likely to be influenced by the prevention needs of the students in this sample. Therefore, we also analyzed how sociodemographics (age, gender, and educational track) were associated with prevention needs. This would provide a clearer understanding of how these factors may shape students' intervention preferences.

The prevention needs of vocational school students in terms of addictive behaviors were expected to be high [5]. On the basis of previous literature on module choice [8,14,15,17,26-31] and the generally low intentions to change addictive behaviors [5], it was expected that life skill modules (stress or social competencies) would be preferred over modules related to addictive behaviors. It was also expected that module choice would show a high alignment with the traffic light feedback received [8,17,29]. Given the differences in prevalence of substance use between the genders [37], it was expected that men would choose modules related to substance use more often than women.

Methods

Design

The data came from a cluster randomized controlled trial among German vocational school students testing the efficacy of the app-based addiction prevention program ready4life [38]. The trial was registered in the German Clinical Trials Register (DRKS00022328).

XSL•FO
RenderX

Participants and Procedures

Recruitment

Details on recruitment can be found elsewhere [39]. Participants in this study were vocational school students from 5 German federal states who were enrolled in vocational training, vocational preparation, or vocational grammar school. Randomization was conducted at the class level within computer-generated blocks of 4 and stratified by school (allocation ratio: 1:1). Within the 35 participating schools, 376 classes were randomized to either the intervention (n=186, 49.5%) or control (n=190, 50.5%) condition.

We used a facilitated access approach by proactively offering intervention participation during school hours. Combining mobile interventions with proactive contacting is crucial to maximize the reach and representativeness of the sample [40], particularly for those not yet planning to change their behavior [41]. This is of high importance as most vocational school students do not yet intend to change their addictive behaviors [5].

Class Introductions

During 1 to 2 school hours, students from both study groups were introduced to the study and were invited to download the ready4life app. During the introduction, students were given detailed information about the study, including its procedures, randomization process, data protection measures, and the content and features of the app. To minimize stigma associated with addiction, introductions largely focused on stress management. Class introductions were conducted for a total of 376 classes between October 2020 and March 2022. Of these, 63.6% (239/376) of classes were introduced by members of the project team, 14.9% (56/376) by trained schoolteachers, 9% (34/376) by addiction prevention experts, 5.9% (22/376) by school social workers, and 6.6% (25/376) by vocational school students. The introductions were delivered via face-to-face (187/376, 49.7%), online streaming (162/376, 43.1%), or an emailed YouTube video link (27/376, 7.2%).

Study Participation

Vocational school students aged ≥16 years with smartphones and who provided contact details (email or phone) for follow-up data collection were eligible for study participation. After downloading the app, all students participated in an anonymous in-app screening on their prevention needs regarding alcohol, tobacco, and cannabis consumption; internet use; social competencies; and stress. Students then provided digital informed consent. A total of 4225 app downloads were recorded, and 2568 students provided informed consent. Participation rate was 46.7% (2508/5370) among students aged ≥16 years who had the correct app version, and belonged to classes where the number of students present during the introduction was known.

All study participants received feedback on their individual risks and competencies in the form of a traffic light system and were then informed of their group allocation. The intervention group then received app-based coaching for 2 self-selected behavior modules over 16 weeks (see the Interventions section for details).

Follow-Up Assessment

Students from both study groups were invited to 2 follow-up sessions at 6 and 12 months via SMS text message or email with a link to an online chat conversation with a virtual coach. If students did not respond after reminders, they were contacted by the project team to provide the option to complete the follow-up during a telephone interview. Prizes were raffled after the 12-month follow-up to increase participant engagement. Prizes were drawn in cash instead of vouchers due to procurement constraints. Participants in both the intervention and control groups with available contact information were eligible. The raffle was based on credits earned on the app, with separate raffles for each group to avoid disadvantaging the control group. Participants were randomly selected for the prizes, which ranged from €10 to €500 (US $11.33 to US $566.39).

Instruments

Class Registration Data

Class-level information was gathered during the registration process for classes, encompassing the federal state of the school, the classes' educational track (vocational preparation, vocational training, and vocational grammar school), year of education, class size, and data concerning the introduction (eg, number of students present, mode of introduction, or introducing person). For students undergoing vocational training, the official label of their training was assessed, and occupations were classified using the German Classification of Occupations [42] (Table S1 in Multimedia Appendix 1). A recode table was used to convert these into the International Standard Classification of Occupations 2008 (ISCO-8) [43], which was used in the main analysis to ensure international compatibility (Table S2 in Multimedia Appendix 1).

Vocational preparation classes are designed for students with basic education to qualify for further vocational training. Vocational grammar schools, on the other hand, allow students to obtain a university entrance qualification. Vocational training typically combines part-time classroom study with on-the-job experience. Entry requirements vary, leading to differences in the socioeconomic backgrounds of students on different tracks. Tracks leading to operational and support roles (ie, clerical support workers, service and sales workers, craft-related trade workers, plant and machine operators and assemblers, and elementary occupations) generally require lower academic qualifications and attract students of a lower socioeconomic status. For example, individuals from lower-income backgrounds often choose these tracks because of financial constraints, limited access to higher education, or expectations of early entry into the labor market. The lower socioeconomic status of students in these tracks also partly results from the lower prestige and job security associated with these tracks, compared to tracks leading to professional, technical, and associate professional roles.

Individual Screening Data

Within the app-based screening, data at the individual level were collected, including sociodemographics, addictive behaviors (alcohol, tobacco, and cannabis consumption and

XSL•FO
RenderX

problematic internet use), perceived stress, and social competencies.

Sociodemographics

Age was calculated from the date of birth given by the students. Gender was assessed as man, woman, or other. When students selected "other," we assessed their tendency to feel more masculine or feminine on a scale from 1 to 6. Due to the limited number of students in the category of "other" (40/2568, 1.6%), those who tended to be more masculine (scores of 1-3; 17/40, 42%) were categorized as men, and those who tended to be more feminine (scores of 4-6; 23/40, 58%) were categorized as women.

Alcohol Consumption

Using questions from the Alcohol Use Disorders Identification Test–Consumption scale [44], students reported their frequency of alcohol consumption in the previous 30 days (0-30). For those who reported consumption, details of the type of drink, the number of drinks consumed on a typical day, and the day with the most alcoholic drinks consumed in the last month were collected using a digital bar displaying common portion sizes (eg, 0.5-L beer or 2-cL shot). Standard drinks were calculated from the selected drinks (1 standard drink=12 g of alcohol). A quantity-frequency index was calculated by multiplying drinking days by typical standard drinks per day and dividing by 30.

Tobacco Consumption

Frequency of smoking tobacco or using nicotine products in the previous 30 days was assessed, with response options including "(almost) daily," "occasionally but not daily," or "never." For those who reported smoking, the number of consumption days (0-30) and cigarettes smoked per day in the previous month was assessed. A quantity-frequency index was calculated by multiplying days of use by cigarettes smoked per day and dividing by 30.

Cannabis Consumption

Lifetime use of cannabis containing tetrahydrocannabinol was assessed using the response options "no, never" and "yes." For those who reported cannabis use, frequency in the previous 6 months was assessed ("not at all," "once a month or less," "2-4 times a month," "2-3 times a week," and "4 times a week or more"), as well as the number of days of use in the previous month (0-30).

Problematic Internet Use

The short version of the Compulsive Internet Use Scale (CIUS) [45] measured problematic internet use on a scale from 0 ("never") to 4 ("very often"). A total score was calculated with a possible range of 0 to 20. Sum scores of ≥7 indicate a problematic use of the internet.

Stress

Stress was assessed using a single-item questionnaire [46]: "Stress is a state in which a person feels tense, restless, nervous, or anxious, or is unable to sleep at night due to disturbing thoughts. How much do you currently feel this type of stress?" Response options ranged from 1 ("not at all strong") to 5 ("very strong").

Social Competencies

On the basis of the Assertion Inventory [47], 8 items measured social competence in approaching others, expressing needs, resisting group pressure, and standing up for oneself. Response options ranged from 1 ("very uncertain") to 5 ("very certain"). A sum score ranging from 8 to 40 was calculated, with higher values indicating higher social competence.

Individual Risk and Competence Feedback and Module Choice

The received traffic light feedback (green, yellow, and red) for each behavior and the 2 chosen behavior modules were automatically recorded on the app for each student. For each module, 1 dichotomous variable was created indicating whether it was chosen (see the Interventions section for details).

Interventions

Development

Details on intervention development have been published elsewhere [38]. The coaching app ready4life was initially developed and evaluated in Switzerland aiming to prevent or reduce addictive behaviors and promote life skills among vocational school students [48]. The latest version of ready4life was further enhanced for this study. In particular, in addition to the 4 existing modules on alcohol, tobacco, stress management, and social competencies, 2 new modules focusing on cannabis and social media and gaming were developed [38]. The adapted version of the ready4life app used in this study consisted of the app-based screening followed by the individual risk and competence feedback and the app-based coaching.

Efficacy

We have shown that the adapted version of ready4life is effective in reducing problematic internet use, stress, and tobacco use, as well as in improving social competencies over 12 months in this study's sample, which will be published elsewhere. Similarly, among Swiss vocational school students, the 6-module ready4life version significantly reduced at-risk drinking and problematic internet use over 6 months [9].

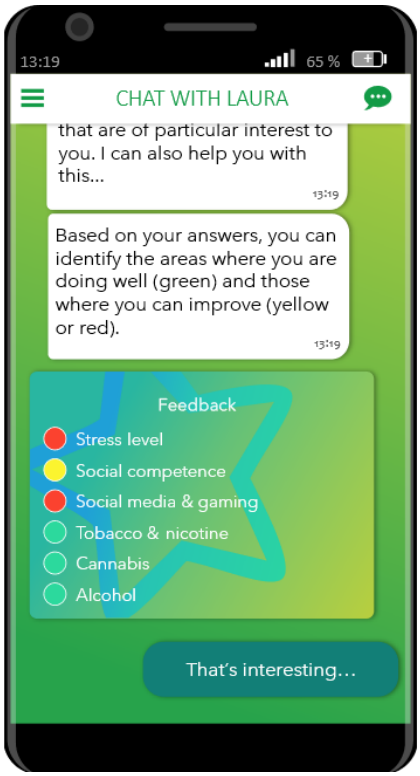
Individual Risk and Competence Feedback

All study participants received in-app feedback on their individual risks and competencies in the form of a traffic light system for each assessed behavior of the in-app screening (Figure 1). Students received a green traffic light when they reported no or minor risks (eg, did not smoke tobacco or use nicotine products in the previous 30 days or perceived no or low stress) or a high competence (eg, feeling secure in social situations) concerning the specific behavior. A yellow traffic light was given if students reported a moderate risk or competence (eg, occasional tobacco smoking or nicotine product use in the previous 30 days, medium perceived stress, and feeling insecure in social situations), and a red traffic light was given if students reported a high risk or low competence (eg, daily or almost daily tobacco smoking or nicotine product use in the previous 30 days, high perceived stress, and feeling highly insecure in social situations). For social media and gaming, students received yellow traffic light feedback if they had a Short CIUS score of ≥7 [45], indicating a problematic internet

XSL•FO
RenderX

use. The red traffic light used a cutoff of ≥9, which has been shown to have a higher specificity. Feedback on alcohol use was based on age- and gender-specific cutoffs derived from established guidelines on alcohol consumption [49]. Topic-specific thresholds for receiving a green, yellow, or red traffic light are shown in Table S3 in Multimedia Appendix 1.

Figure 1. Individual risk and competence feedback provided by the app (English translation).



Module Choice and App-Based Coaching

Students from the intervention group chose 2 out of 6 available behavior modules (alcohol, tobacco, cannabis, social media and gaming, stress, and social competencies) to be coached in for 16 weeks via the ready4life app as described in this section. Because many vocational school students engage in multiple addictive behaviors, addressing all at-risk behaviors could be overwhelming [24]. Therefore, the choice of modules was limited to 2. Previous research suggests that this may optimize adherence to module recommendations [34] and improve intervention efficacy [24]. Given the comprehensive approach of the intervention, we expected that focusing on 2 selected modules would positively influence unselected topics by increasing students' motivation and self-efficacy to subsequently address other behaviors in the long term. In addition, we anticipated that changes in one behavior would positively influence other behaviors (eg, by improving resilience factors such as social competence and stress management, which are known to protect against addiction [50,51]).

The students were advised to use their individual risk and competence feedback to guide their module choice (eg, choosing modules on behaviors for which they had received yellow or red traffic light feedback). However, module choice was completely self-determined, including the option to choose modules on behaviors for which students had received green traffic light feedback.

Each module included 8 weeks of coaching. The 2 selected modules could not be started simultaneously. Instead, we chose to present the modules sequentially for the following reasons: (1) gradual introduction of behaviors may reduce participants' burden and resistance, thereby improving engagement [23] and efficacy [52]; (2) success in changing one behavior may increase students' confidence in tackling others subsequently; (3) this approach is consistent with goal setting theory by breaking goals down into manageable steps [53]; and (4) while simultaneous approaches may work well with populations with a high willingness to change multiple behaviors, a sequential approach may be particularly beneficial for populations with lower motivation, such as vocational school students [5]. The 2 chosen modules were presented in a random order.

The app-based coaching involved weekly chat interactions (once a week for approximately 5 min) over a 4-month period. A rule-based conversational agent (ie, without artificial intelligence) provided individualized feedback within the selected intervention modules. Chats also included media (videos, images, and links), quizzes, and contests. The content of the chats was constructed using the theoretical frameworks of social cognitive theory [54] and the health action process approach [55]. Each weekly session addressed specific behavioral constructs and goals, such as self-monitoring, risk perception, goal setting, or coping strategies. The detailed weekly content can be found in the study by Schmidt et al [38].

Predefined decision rules guided the chat interactions, with participant input during the conversations determining the subsequent responses of the conversational agent. For example, participants actively chose whether to receive additional information on certain topics, and the agent's responses varied

XSL•FO
RenderX

depending on the accuracy of the participant's answers to quiz questions. Feedback was further tailored using participant data collected during the app-based screening. These data included demographic variables (eg, age and gender), behavioral indicators (eg, level of alcohol consumption and social competencies), and other variables such as motivation to change the targeted behavior or perceived pros and cons of behavior change. In addition, progress in behavior change was assessed at weeks 8 and 16 to allow for ipsative feedback.

In addition to the weekly chats, users could self-initiate chat dialogues at any time. Push notifications (eg, "Hi <nickname>") were used to inform the students of new chat content once a week. Further reminders were used for contest participation or if the app was not used for an extended period. Personal support in the form of an "ask-the-expert" function was integrated into the app.

Control Group

After participating in the in-app screening and receiving the individual risk and competence feedback, students from the control group only received a link to information on enhancing health behaviors. However, they could access module choice and app-based coaching after 12 months.

Reporting checklists can be found in Multimedia Appendices 2-4.

Data Analyses

Description of the Sample and Prevention Needs

Descriptive statistics were used to describe the sample using numbers and percentages, means and SDs, or medians and IQRs for highly skewed data. Prevention needs were described in terms of continuous addictive behaviors, stress, and social competencies. Age, gender, and educational track were analyzed as potential determinants of students' prevention needs using multiple multilevel regression models.

Individual Risk and Competence Feedback

The percentages of students receiving green, yellow, or red traffic light feedback were graphically displayed for each behavior. The percentages of students showing different numbers of risks were reported based on yellow or red traffic light feedback and red traffic light feedback only.

Module Choice

Of the 1286 participants randomized to the intervention group, 20 (1.6%) had no access to the intervention due to mistyping of their class password, and furthermore, 30 (2.3%) did not choose modules. Thus, module choice was analyzed for 96.1% (1236/1286) of the intervention group participants.

The percentages of participants who chose each behavior module were reported for all intervention participants, intervention participants who received yellow or red traffic light feedback for that behavior, and intervention participants who received red traffic light feedback. As only 2 out of 6 behavior modules could be chosen, percentages of intervention participants who chose each behavior module were also reported based on the number of shown risks (1 or 2 risks or >2 risks).

Congruence between presented risk and module choice was measured using the proportion of students with yellow or red traffic light feedback for a particular behavior (eg, alcohol) among those who selected the corresponding behavior module (eg, alcohol module).

Potential Determinants of Module Choice

Multiple multilevel logistic regression analyses were conducted to model choice of each module separately (1=module chosen; 0=not chosen). Potential determinants included age, gender, educational track, and the corresponding prevention need (ie, problematic internet use was included as the independent variable for modeling choice of the social media and gaming module, social competencies were included as the independent variable when modeling choice of the social competencies module, and so on). All regression models included random intercepts at the class level [56] to account for the clustered structure of the data. Intraclass correlation coefficients (ICCs) were computed using an intercept-only model. ICCs express the percentage of the total variance in the outcome (eg, module choice) that is attributed to class membership [56]. Partly low numbers of classes per school prohibited modeling of random intercepts at the school level. Significance was set at $P<.05$. All statistical analyses were conducted using Stata SE (version 17.0; StataCorp).

Use of the German Classification of Occupations to Analyze Determinants of Prevention Needs and Module Choice

A sensitivity analysis was conducted using the German Classification of Occupations [42]. The German Classification of Occupations differs from the ISCO-8 in that it classifies occupations based on the specific content, industry, or sector in which the work is performed compared to the ISCO-8, which categorizes occupations based on the tasks, skills, and responsibilities associated with them. It considers the German labor market's unique characteristics and is more detailed in classifying occupations within the context of Germany.

Ethical Considerations

Ethics approval was granted by the ethics committees of the University of Lübeck (19-419) and the University Medicine Greifswald (BB 024/20). Digital informed consent was obtained from vocational school students through the app, with no additional parental consent necessary in accordance with the European Union General Data Protection Regulation. Underaged participants were encouraged to inform their parents or guardians about study participation. Students completed the app-based screening anonymously. No names were collected; instead, students chose nicknames. Data protection was ensured through encrypted data transmission and storage within the app. For those who provided informed consent, research data were stored in a pseudonymized format, and contact information (email or phone number) was securely stored on a separate server. Participants were not offered any compensation other than the chance to win prizes.

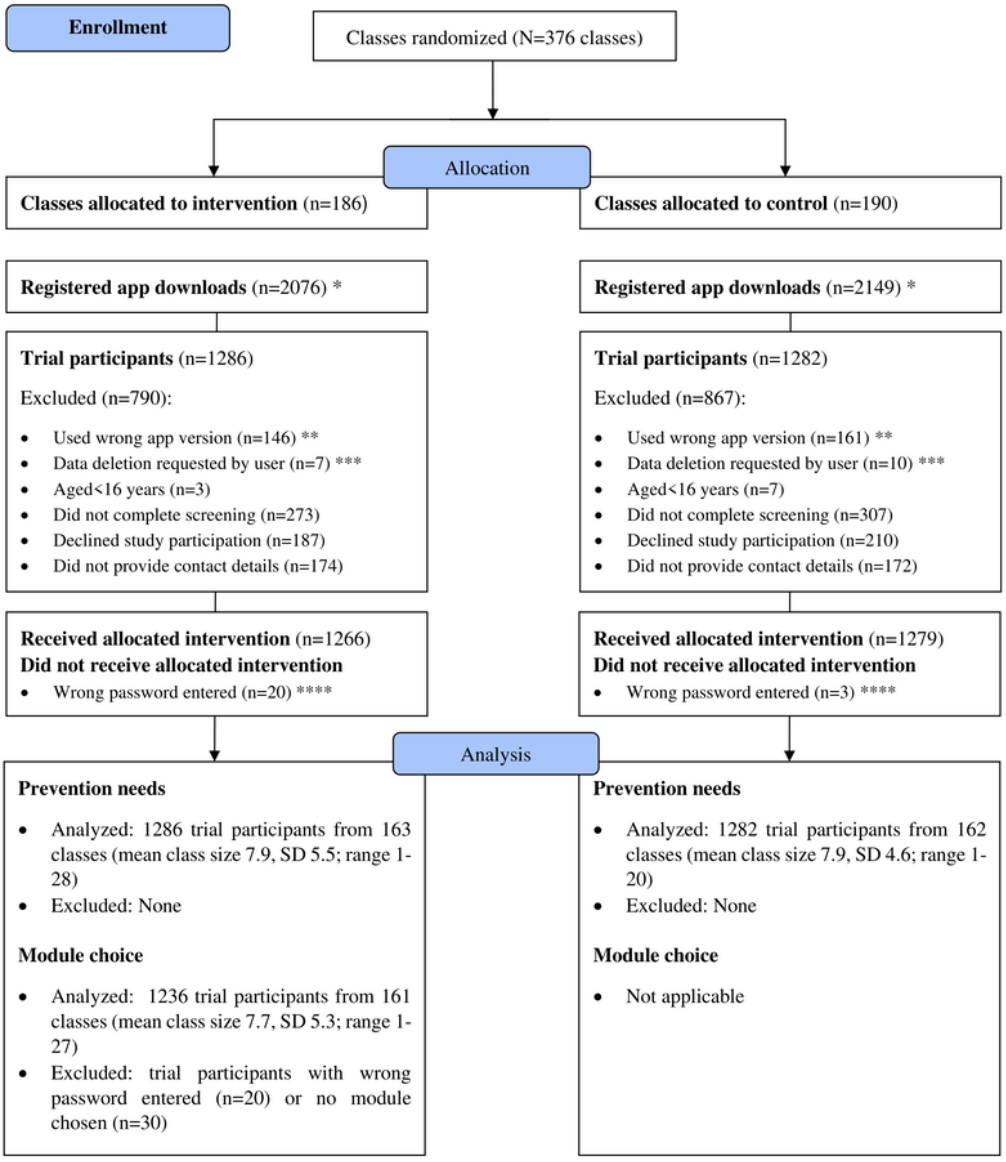
XSL•FO
RenderX

Results

Sample Description and Prevention Needs

The flow of the participants is depicted in Figure 2. Study participants are described in Table S4 in Multimedia Appendix 1. Among the total sample, the mean age was 19.7 (SD 3.7) years, and 45.6% (1170/2568) were women. Most study participants (1648/2568, 64.2%) were in vocational training, followed by vocational grammar school (582/2568, 22.7%) and vocational preparation (282/2568, 11%). Within those students in vocational training, the most prevalent occupations were technicians and associate professionals (511/1648, 31%), as well as craft-rated trade workers (422/1648, 25.6%). Most students (1785/2568, 69.5%) were in their first or second year of education.

Figure 2. Flowchart of clusters and participants. *n=11 students were identified who had registered for the study twice; these were already subtracted from the download count. **The app included multiple versions tailored to different countries or languages. Due to a technical issue, between September 2, 2021, and December 10, 2021, a total of 307 students were mistakenly directed to a different version of the app instead of the German version. As a result, these students used a version of the app that was not part of this study, and informed consent could not be obtained from them. ***App users were able to initiate the deletion of all data collected at any time point regardless of whether they had just started the screening or provided informed consent and were using the intervention already. ****Group allocation was conducted at the class level based on class-specific passwords; due to mistyping of passwords, 1.6% (20/1286) of the students from the intervention group did not receive access to the intervention, and 0.2% (3/1282) of the students from the control group received access to the intervention.



In terms of prevention needs, students on average perceived a moderate level of stress (mean 3.29, SD 1.20 on a scale from 1 to 5) and reported a moderate level of social competencies (mean 29.50, SD 4.69 on a scale from 8 to 40). The mean score on the Short CIUS was 9.14 (SD 4.18), which is above the threshold for problematic use of the internet. The median number of standard alcoholic drinks consumed per day was 0.20 (IQR 0.00-1.00). A proportion of 31.7% (814/2568) and 18.1% (464/2568) had used tobacco and cannabis in the previous month, respectively.

Of the total sample, 45.1% (1158/2568) of the students used the app on Android, and 54.9% (1410/2568) used it on iOS.

XSL•FO
RenderX

Notably, iOS users reported significantly higher prevention needs across all behavioral domains except social competencies, where they had lower needs compared to Android users. In addition, iOS users were more likely to select the tobacco module (data not shown).

When analyzing prevention needs as outcomes, ICCs were highest for alcohol consumption, indicating that 13.5% of the variance in the number of standard alcoholic drinks per day was attributable to class membership (Table S5 in Multimedia Appendix 1). Being a woman was associated with significantly higher stress (regression coefficient=0.70, 95% CI 0.60-0.80; *P*<.001), higher problematic internet use (regression coefficient=0.74, 95% CI 0.39-1.09; *P*<.001), and lower social competencies (regression coefficient=−2.27, 95% CI −2.67 to −1.88; *P*<.001). In contrast, being a woman was also associated with lower substance use (lower alcohol consumption: regression coefficient=−0.40, 95% CI −0.55 to −0.24 and *P*<.001; lower odds of tobacco use: odds ratio [OR] 0.79, 95% CI 0.65-0.97 and *P*=.02; lower odds of cannabis use: OR 0.60, 95% CI 0.48-0.75 and *P*<.001) holding age and educational track constant.

Higher problematic internet use was significantly associated with younger age (regression coefficient=−0.10, 95% CI −0.15 to −0.06; *P*<.001), whereas higher odds of tobacco use were associated with older age (OR 1.04, 95% CI 1.01-1.06; *P*=.006). Stress (regression coefficient=0.01, 95% CI −0.001 to 0.03; *P*=.06), social competencies (regression coefficient=0.05, 95% CI −0.005 to 0.10; *P*=.07), odds of cannabis use (OR 1.02, 95% CI 0.99-1.05; *P*=.18), and alcohol consumption (regression coefficient=0.01, 95% CI −0.004 to 0.03; *P*=.12) were not significantly associated with age.

Wald tests indicated that social competencies (*P*=.02), problematic internet use (*P*=.002), and the odds of tobacco use (*P*<.001) were significantly associated with educational track after adjusting for gender and age, whereas perceived stress (*P*=.26), odds of cannabis use (*P*=.68), and alcohol consumption (*P*=.54) did not differ significantly between educational tracks. Clerical support workers (regression coefficient=1.10, 95% CI 0.36-1.84; *P*=.004), service and sales workers (regression coefficient=0.79, 95% CI 0.05-1.52; *P*=.04), students in vocational grammar school (regression coefficient=0.89, 95% CI 0.27-1.52; *P*=.005), and students in vocational preparation (regression coefficient=1.02, 95% CI 0.26-1.79; *P*=.009) reported significantly higher social competencies than the reference group of professionals, technicians, and associate professionals. The other educational tracks did not differ significantly in their social competencies from each other (data not shown).

Problematic internet use was significantly lower among craft-related trade workers (regression coefficient=−0.76, 95% CI −1.35 to −0.17; *P*=.01), and the odds of tobacco use were significantly higher among service and sales workers (OR 1.61, 95% CI 1.10-2.37; *P*=.02) than in the reference group.

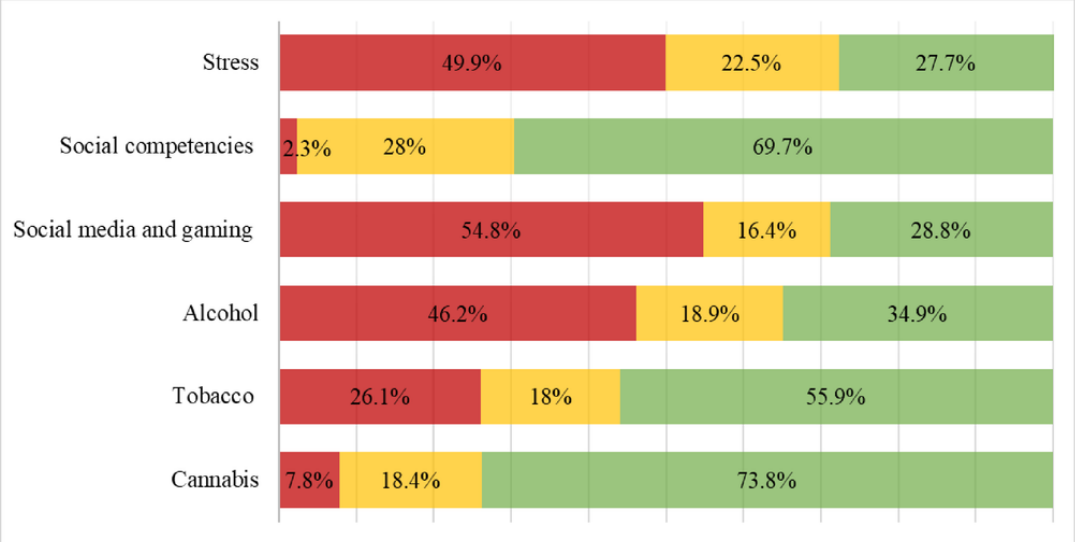
Comparisons between the other educational tracks revealed that service and sales workers (regression coefficient=−0.81, 95% CI −1.57 to −0.06; *P*=.03), as well as craft-related trade workers, plant and machine operators and assemblers, and elementary occupations (regression coefficient=−1.26, 95% CI −1.93 to −0.58; *P*<.001), reported significantly lower problematic internet use than clerical support workers, whereas students in vocational grammar school (regression coefficient=0.95, 95% CI 0.38-1.52; *P*=.001) and vocational preparation (regression coefficient=1.00, 95% CI 0.30-1.69; *P*=.005) reported significantly higher problematic internet use than craft-related trade workers, plant and machine operators and assemblers, and elementary occupations. Vocational grammar school students also had lower odds of tobacco use than service and sales workers (OR 0.46, 95% CI 0.31-0.68; *P*<.001); craft-related trade workers, plant and machine operators and assemblers, and elementary occupations (OR 0.52, 95% CI 0.37-0.74; *P*<.001); and students in vocational preparation (OR 0.64, 95% CI 0.43-0.95; *P*=.03). Finally, service and sales workers had higher odds of tobacco use than clerical support workers (OR 1.61, 95% CI 1.03-2.49; *P*=.04).

Individual Risk and Competence Feedback

Figure 3 shows the percentage of study participants who received red, yellow, or green traffic light feedback for each behavior. Approximately two-thirds of the study participants received either red or yellow traffic light feedback for stress (1858/2568, 72.4%), social media and gaming (1828/2568, 71.2%), or alcohol (1672/2568, 65.1%). For tobacco, less than half (1133/2568, 44.1%) of the study participants received red or yellow traffic light feedback, whereas for cannabis (673/2568, 26.2%) and social competencies (778/2568, 30.3%), only approximately one-third received red or yellow traffic light feedback.

XSL•FO
RenderX

Figure 3. Percentage of students who received red, yellow, or green traffic light feedback by topic (N=2568). Traffic light feedback was based on self-reported behaviors and gave advice on potentially useful modules.



On average, students received red or yellow traffic light feedback for 3.09 (SD 1.42) of the 6 behaviors and red traffic light feedback only for 1.87 (SD 1.17) of the 6 behaviors. Only 2.7% (70/2568) of the study participants received neither red nor yellow traffic light feedback, and 11.1% (286/2568) received no red traffic light feedback for any of the 6 behaviors. The number of participants receiving red or yellow traffic light feedback for 1, 2, 3, 4, 5, and 6 behaviors were 11.1% (285/2568), 21.42% (550/2568), 24.8% (637/2568), 22.6% (580/2568), 13.6% (350/2568), and 3.7% (96/2568), respectively. The number of participants receiving red traffic light feedback for 1, 2, 3, 4, 5, and 6 behaviors were 30% (769/2568), 30.5% (783/2568), 19% (488/2568), 7.9% (204/2568), 1.4% (37/2568), and 0.04% (1/2568), respectively.

Module Choice

Among students who made a module choice, the most frequently chosen coaching topic was stress (818/1236, 66.2%), followed by social media and gaming (625/1236, 50.6%), alcohol (360/1236, 29.1%), social competencies (306/1236, 24.8%), tobacco (232/1236, 18.8%), and cannabis (131/1236, 10.6%). The frequency of module choice was higher among those students who had received red or yellow traffic light feedback for the corresponding behavior than among the total group (Table 1). This trend was highest for stress, with 80.3% (727/905) of students with yellow or red traffic light feedback and 87.7% (541/617) of students with red traffic light feedback only choosing the stress module. The lowest choice rate was found for alcohol consumption. Even within those with red traffic light feedback, only 46.6% (274/588) chose the alcohol module.

Table 1. Frequency of participants' module choice by traffic light feedback. Traffic light feedback was based on self-reported behaviors and gave advice on potentially useful modules; however, vocational school students could freely select any 2 of the 6 modules (N=1236).

Module	Any feedback (n=1236), n (%)	Yellow or red traffic light feedback, n/N (%)	Red traffic light feedback, n/N (%)
Stress	818 (66.2)	727/905 (80.3)	541/617 (87.7)
Social competencies	306 (24.8)	188/368 (51.1)	22/27 (81.5)
Social media and gaming	625 (50.6)	541/882 (61.3)	451/687 (65.7)
Alcohol	360 (29.1)	344/820 (42)	274/588 (46.6)
Tobacco	232 (18.8)	226/552 (40.9)	189/327 (57.8)
Cannabis	131 (10.6)	117/323 (36.2)	68/94 (72.3)

Table S6 in Multimedia Appendix 1 shows that, among those students with only 1 or 2 instances of yellow or red traffic light feedback, the choice rates for the recommended modules ranged from 68% (13/19) to 98% (40/41). For those with >2 instances of yellow or red traffic light feedback, the choice rates were much lower as students had to choose between several recommended modules. For example, choice rate for social media and gaming among those with >2 instances of yellow or red traffic light feedback (193/206, 93.7%) was approximately 30% lower than among those with 1 or 2 instances of yellow or red traffic light feedback (542/884, 61.3%). However, the number for some of the subsamples was rather small.

Most students chose modules that were congruent with their presented risks. For example, 88.9% (727/818) of the students who chose the stress module had yellow or red traffic light feedback in this behavior area. The congruence rates for the other behaviors were 61.4% (188/306) for social competencies, 86.6% (541/625) for social media and gaming, 95.6% (344/360)

XSL•FO
RenderX

for alcohol, 97.4% (226/232) for tobacco, and 89.3% (117/131) for cannabis.

Potential Determinants of Module Choice

Table S7 in Multimedia Appendix 1 shows the association between gender, age, educational track, and prevention needs and module choice. The ICCs for the modules ranged from 2.1% to 11%, indicating a modest influence of class membership on module choice.

Women were significantly more likely to select the stress module than men (OR 2.38, 95% CI 1.69-3.33; *P*<.001), who were more likely to select social media and gaming (OR 0.52, 95% CI 0.40-0.69; *P*<.001), alcohol (OR 0.50, 95% CI 0.37-0.67; *P*<.001), and cannabis (OR 0.37, 95% CI 0.21-0.63; *P*<.001) than women when holding age, educational track, and prevention need for the corresponding behavior constant. Men and women did not differ significantly with respect to their frequency of module choice for the tobacco (OR 1.01, 95% CI 0.70-1.46; *P*=.97) and social competencies (OR 1.13, 95% CI 0.83-1.54; *P*=.44) modules. Younger students were significantly more likely to choose the cannabis module (OR 0.81, 95% CI 0.74-0.90; *P*<.001), whereas there were no significant differences in module choice by age for the other 5 modules.

Although Wald tests did not reach significance, there were individual educational tracks that differed in their odds of choosing the social competencies, social media and gaming, alcohol, and cannabis modules when compared to the reference category. Craft-related trade workers, plant and machine operators and assemblers, and elementary occupations had a significantly lower likelihood of choosing social media and gaming when compared to the reference group of professionals, technicians, and associate professionals (OR 0.61, 95% CI 0.39-0.97; *P*=.04), whereas the other tracks did not differ significantly from the reference group in their likelihood of choosing social media and gaming. Clerical support workers (OR 1.85, 95% CI 1.02-3.38; *P*=.04); service and sales workers (OR 1.77, 95% CI 1.00-3.15; *P*=.049); and craft-related trade workers, plant and machine operators and assemblers, and elementary occupations (OR 1.98, 95% CI 1.16-3.38; *P*=.01) were significantly more likely to choose alcohol compared to the reference group, whereas students in vocational grammar school (OR 1.23, 95% CI 0.73-2.09; *P*=.43) and vocational preparation (OR 1.57, 95% CI 0.86-2.87; *P*=.14) did not differ significantly from the reference group in their likelihood of choosing alcohol. Clerical support workers were most likely to select the cannabis module (OR 3.66, 95% CI 1.33-10.03; *P*=.01), and service and sales workers were least likely to select the social competencies module (OR 0.51, 95% CI 0.30-0.88; *P*=.02) when professionals, technicians, and associate professionals were used as the reference group.

Prevention needs for each behavior were significantly related to the corresponding module choice (eg, the higher the alcohol consumption, the more likely it was that the alcohol module would be chosen; OR 1.31, 95% CI 1.21-1.43; *P*<.001).

Use of the German Classification of Occupations to Analyze Determinants of Prevention Needs and Module Choice

The prevalence of the occupational tracks based on the German Classification of Occupations is shown in Table S8 in Multimedia Appendix 1.

Wald tests indicate that only problematic internet use (*P*=.002) and the odds of tobacco use (*P*=.002) were significantly associated with educational track after adjusting for gender and age (Table S9 in Multimedia Appendix 1). All educational tracks showed numerically higher problematic internet use than the reference category of production and construction. However, these differences did not reach significance for the tracks of natural science, geography, and computer science (regression coefficient=0.44, 95% CI −0.43 to 1.31; *P*=.32) and trading and hotel and tourism (regression coefficient=0.65, 95% CI −0.04 to 1.34; *P*=.06).

The odds of tobacco use were significantly lower in natural science, geography, and computer science (OR 0.58, 95% CI 0.34-0.97; *P*=.04); organization and accounting and media (OR 0.59, 95% CI 0.38-0.92; *P*=.02); and vocational grammar school (OR 0.52, 95% CI 0.37-0.72; *P*<.001) than in the reference category of production and construction.

Comparisons between the other educational tracks revealed that the odds of tobacco use were lower in vocational grammar school than in traffic, logistics, and security and safety (OR 0.51, 95% CI 0.30-0.89; *P*=.02); trading and hotel and tourism (OR 0.51, 95% CI 0.34-0.76; *P*<.001); health, social affairs, teaching, and education (OR 0.65, 95% CI 0.44-0.97; *P*=.03); and vocational preparation (OR 0.64, 95% CI 0.43-0.95; *P*=.03). Similarly, the odds were also lower in organization and accounting and media than in trading and hotel and tourism (OR 0.58, 95% CI 0.35-0.95; *P*=.03).

Choice of the social competencies module could not be modeled using multilevel logistic regression as the inclusion of the independent variables led to nonconvergence of the model. Instead, multiple logistic regression was used. Wald tests indicated that only the odds of choosing the social competencies (*P*<.001) and alcohol (*P*=.01) modules were significantly associated with educational track after adjusting for gender, age, and the corresponding prevention needs (Table S10 in Multimedia Appendix 1). The social competencies module was chosen significantly more often by students from the natural science, geography, and computer science tracks (OR 3.40, 95% CI 1.71-6.73; *P*<.001) and significantly less often by students from the traffic, logistics, and security and safety tracks (OR 0.32, 95% CI 0.11-0.99; *P*=.048) compared to the reference group. Comparisons between the other educational tracks revealed that all educational tracks were significantly less likely to choose the social competencies module compared to students in natural science, geography, and computer science. Students in organization, accounting and media, and vocational grammar school were significantly more likely to choose the social competencies module than students in traffic, logistics, and security and safety, as well as students in trading and hotel and tourism, whereas the other educational tracks did not differ

XSL•FO
RenderX

significantly from each other in their odds of choosing the social competencies module (data not shown).

The alcohol module was selected significantly less often by students from the natural science, geography, and computer science tracks (OR 0.22, 95% CI 0.09-0.59; *P*=.002) compared to the reference group. The other educational tracks did not differ significantly from each other in their odds of choosing the alcohol module.

Although Wald tests did not reach significance, there were individual educational tracks that differed in their odds of choosing the stress, social media and gaming, and cannabis modules when compared to the reference category. Stress was chosen significantly more often by students from the health, social affairs, teaching, and education tracks (OR 2.10, 95% CI 1.11-3.98; *P*=.02), and social media and gaming was chosen significantly more often by students from the organization and accounting and media tracks (OR 2.08, 95% CI 1.16-3.72; *P*=.01). The cannabis module was chosen significantly more often by students from the traffic, logistics, and security and safety tracks (OR 4.41, 95% CI 1.54-11.13; *P*=.005). The results for the other potential determinants in the multiple regression models are consistent with those of the main analysis.

Discussion

Principal Findings

This study examined the prevention needs of German vocational school students and their module choices within the app-based addiction prevention intervention ready4life. Key findings included, first, that prevention needs among vocational school students were high. Almost 80% reported 2 or more risks according to yellow or red traffic light feedback. The highest prevention needs were identified for stress, social media and gaming, and alcohol. Second, in line with this, the most frequently chosen modules were stress and social media and gaming. Third, alignment of module choice with the risk feedback received was high, especially for those with 1 or 2 risks. Fourth, gender, age, educational background, and the corresponding prevention needs were associated with module choice.

Interpretation of the Results and Comparison With Prior Work

Prevention Needs

Our findings regarding the prevention needs related to substance and internet use are consistent with previous epidemiological data showing a high prevalence of addictive behaviors among vocational school students [5,35,57,58], higher levels of substance use among men than among women, and high levels of substance use among students in construction tracks [5,57].

The proportions of students who received green, yellow, and red traffic light feedback, as well as the determinants of prevention needs, are congruent with those found in the Swiss evaluation study of ready4life [17]. For example, Paz Castro et al [17] also found that more than half (57.2%) of the students showed high-risk internet use, approximately a quarter (23.9%) showed high-risk tobacco use, and <10% (9.6%) showed high-risk cannabis consumption. Our study yielded approximately 12.5% more students with a high risk (red traffic light) than the study by Paz Castro et al [17] only in the case of alcohol consumption, probably due to a higher average age in our study. The high need for prevention in relation to stress is in line with previous research showing that vocational school students have high levels of job-related stress and that this is associated with substance use [50].

In line with our study, Paz Castro et al [17] found that women reported more stress and problematic internet use but less social competencies and substance use than men, problematic internet use was associated with younger age, and tobacco use was associated with older age. However, in contrast to our findings, Paz Castro et al [17] also found slightly higher rates of risky alcohol use in women than in men and higher stress as well as higher rates of risky cannabis use in older than in younger students. These differences may be explained by the different definitions of outcomes or potential determinants (eg, continuous vs categorical substance use or age) or different adjustment variables.

Comparisons of prevention needs according to educational track between these studies are difficult because different systems were used to classify occupations. However, both studies are consistent in that the areas of construction and production were among those with the highest levels of tobacco use and those with the lowest levels of problematic internet use. In both studies, cannabis consumption did not differ by educational track.

Module Choice

Stress and social media and gaming were the most frequently selected modules. This is congruent with the high proportion of students in our study with prevention needs in this area. The results regarding module choice were also congruent with the results of the Swiss evaluation study of ready4life [17], which had the highest choice rates for stress and social media and gaming as well.

Substance-related modules such as alcohol, tobacco, and cannabis were less frequently chosen. This is consistent with previous multibehavior interventions among adults showing that health-promoting behaviors such as physical activity and diet are preferred over substance-related modules [14,15,26,28-31] but also with interventions among vocational school students showing that stress and non–substance-related behaviors such as social media use were chosen more often than substance-related behaviors such as smoking, alcohol or cannabis use [8,17].

However, as it was intended, the traffic light feedback did affect module choice. When a module was recommended to the student based on the presence of a moderate or high risk, the choice rates became much higher. This was particularly true for those students with only 1 or 2 risks. This is consistent with previous research showing that individuals are more likely to choose modules that are recommended to them based on their risk profile [17,29]. This finding is not trivial because vocational school students usually have low motivation to change addictive behaviors [5]. This could lead students to choose life skill

XSL•FO
RenderX

modules over addiction-related modules even when given feedback on their high prevention needs in terms of addictive behaviors. Furthermore, in mobile phone–based substance use interventions, those with higher severity have sometimes shown lower engagement [59]. Finally, a similar intervention found that only half of the students chose modules congruent with their risks [60] when no feedback on prevention needs was provided.

In this study, the proportion of students choosing modules congruent with their presented risk was higher (188/306, 61.4%-226/232, 97.4%). However, due to the design of the study, we cannot exclude the possibility that module choice would have been the same without module recommendation. Future experimental research should compare the effects of a choice-based system with those of a similar automated referral system on nonuse attrition and intervention effects. Previous studies have not been adequately designed to draw valid conclusions on this [26,28,30].

Choice rate was lowest for the alcohol module. Allowing users to self-select their intervention topics may leave alcohol use largely unaddressed, especially if multiple risks are present. This is in line with data from adults showing that especially at-risk alcohol users with multiple risks tend to have lower motivation to change and less interest in behavioral feedback [61]. In addition to the low motivation to change, this finding could be explained by the stigmatization of alcohol dependence [62], which may lead to a preference for less stigmatized topics. Furthermore, given the low intention to change addictive behaviors among vocational school students [5], offering the option to choose nonsubstance areas such as life skills might increase interest in participation and intervention use.

Potential Determinants of Module Choice

There have been some studies in adults that have analyzed potential determinants of module choice in multiple-behavior interventions [14-16]. This is the first study to statistically investigate potential determinants of module choice in vocational school students. The gender differences found in module choice are partly consistent with the prevalence differences between genders found in this study and other studies of vocational school students [17]. Furthermore, this is in line with the findings of a previous online intervention on multiple health-promoting behaviors and substance use for adults that showed higher interest of women than men in nonsubstance modules [15]. Specifically, as found in our study, in the study by Schulz et al [15], the choice of the alcohol module was related to being a man, whereas the choice of the tobacco module was not related to gender. Using the German Classification of Occupations, it is evident that differences in module choice in relation to educational track could be partly explained by the specific relevance of the modules for these tracks (eg, high stress levels in the health and education sector [63] or the influence of cannabis use on safe driving in the traffic and logistic sector). However, it is important to emphasize that associations between determinants and module choice exist independently of or in addition to the students' prevention needs and, therefore, may also reflect how effectively at-risk groups are being reached, highlighting the issue of social equity in mobile health interventions. Many mobile interventions fail to equitably engage subpopulations, particularly those of a lower socioeconomic status [36], which can exacerbate health disparities. In this study, gender disparities in module choice were evident. For example, men were less likely to choose the stress module, whereas women were less likely to choose modules on social media and gaming, alcohol, and cannabis. Research suggests that stress is more strongly associated with substance use among men [64], making them more likely to externalize stress through substances, particularly alcohol. Failure to address stress through appropriate modules could lead to the development or continuation of these harmful coping mechanisms among men. The lower likelihood of women choosing the alcohol or cannabis modules compared to men is concerning given the increasing trends in substance use among women [65] and evidence that women are more vulnerable to the negative consequences of substance use than men [66]. This disparity in module choice may reflect perceived gender norms or social stigma.

Interestingly, students from educational tracks typically associated with lower socioeconomic status (eg, clerical support workers, service and sales workers, craft-related trade workers, and vocational preparation) were equally or even more likely to choose modules related to stress and substance use than their peers from educational tracks typically associated with higher socioeconomic status (eg, professionals, technicians, and vocational grammar school). Thus, our mobile intervention approach had a positive impact on equity in that students of a lower socioeconomic status were more likely to respond to alcohol and cannabis modules than students of a higher socioeconomic status [67]. This is a significant finding given that previous research has consistently shown that individuals of a lower socioeconomic status tend to benefit less from mobile interventions due to lower participation and intervention use [36,68,69] despite higher rates of substance use [35], clustering of addictive behaviors [5], and greater substance-related harm [70]. Some lower-status educational tracks were less likely to choose the social competencies or social media and gaming modules, such as service and sales workers or craft-related trade workers. While exposure to substance use modules may have an immediate positive impact on the health of these subgroups, not choosing social skill modules may limit students' opportunities to develop essential skills such as communication, conflict resolution, and emotional regulation, which are valuable in both professional and personal contexts.

The fact that a higher need for prevention increased the likelihood of choosing the corresponding module is in line with the results of previous studies [8,17]. Although these studies did not statistically investigate potential determinants, higher prevention needs (indicated by yellow or red traffic light feedback) led to numerically higher choices of the corresponding modules compared to the overall average.

Strengths

The strengths of this study include (1) the focus on vocational school students, an underserved group in mobile intervention research; (2) the use of a novel multibehavioral approach that combined life skill training with addressing multiple addictive

XSL•FO
RenderX

behaviors; (3) the use of a facilitated access method that ensured a more representative sample compared to relying on online or media recruitment [40]; and (4) being the first study to statistically analyze the determinants of module choice among vocational school students.

Limitations

The following limitations of this study should be considered when interpreting the results. First, the data may not be fully representative of German vocational school students due to selection bias (eg, additional analyses showed that men and students from nonprofessional backgrounds were less likely to participate in this study). Furthermore, not all educational tracks were represented in the sample (eg, the agriculture, forestry, or military tracks were not included). Second, data collection during the COVID-19 pandemic may have influenced participation willingness, addictive behaviors, or module selection. For example, gaming and social media use were elevated during the pandemic in adolescents [71]. Third, all assessments were based on self-report measures. Thus, there is the potential for over- or underreporting, which could have affected the accuracy of the risk and competence feedback. Fourth, this study focused on module choice rather than module completion or frequency of app use. However, completion rates are likely to differ between modules [15,31]. Fifth, we cannot exclude the possibility that the high interest in the stress module was partly triggered by the classes' introductions, which were largely focused on stress. Finally, the number of potential determinants assessed was limited. Other variables such as students' personal values and goals, subjective importance of health, perceptions of social norms, or migration background could also be related to their prevention needs or module choices.

Conclusions

Our study supports the existing evidence that vocational school students have high prevention needs in relation to addictive behaviors. An important finding was that the students' module choices were highly congruent with their identified prevention needs. This is important because intervention effects are likely to be higher when modules that are more relevant to the participant are chosen (eg, when smokers choose to work on the smoking module [60]). Our findings also highlight that incorporating self-tailoring options into mobile health interventions works well to address multiple risks and different combinations of risks for vocational school students. This is particularly important given that a high proportion of vocational school students engage in multiple addictive behaviors and may not be able to address all their issues at once. In terms of educational track, our mobile intervention approach was able to achieve a positive equity impact (ie, those in tracks typically associated with a lower socioeconomic status were more likely to choose modules related to alcohol and cannabis). The identified module preferences may inform the further development of mobile interventions to improve adherence and effectiveness within specific subgroups. To promote more gender-equitable engagement, mobile interventions need to address disparities in module selection. Future interventions for vocational school students could tailor module content or descriptions to gender to ensure that both genders engage with modules that meet their prevention needs. For example, framing stress management around work performance for men or linking substance use prevention to emotional well-being for women may increase engagement. Alternatively, embedding cross-links within modules, such as mentioning substance use in life skill modules, could promote more holistic behavior change across diverse groups. Future research should also investigate whether the choice of a particular module influences change in the behaviors addressed in other unchosen modules (eg, the effect of choosing the stress module on addictive behaviors).

Acknowledgments

This trial is funded by the German Federal Ministry of Health (grant ZMVI1-2519DSM217). The funding institution did not influence the design and conduct of the study; the management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript. The authors thank the Swiss Lung Association and the Swiss Research Institute for Public Health and Addiction (University of Zurich) for their contribution to program development and implementation, as well as Pathmate Technologies for the technical program development. They thank Dominique Brandt for her contributions to the development of the cannabis and internet use modules, the testing of the modified ready4life app, the development of the standard operating procedure, and the training of teachers to introduce the program to their classes. The authors appreciate all teachers, prevention staff, and students for contributing to the study. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

DG wrote the manuscript draft; conducted the statistical analysis; and was responsible for data management, data monitoring, data validation, data cleansing, contribution to the 2 newly developed ready4life modules, testing of the modified ready4life app, recruitment of vocational schools, and presentation of the study in vocational school classes. EK assisted with the classification of occupations for students in vocational training, follow-up data collection, and the preliminary statistical analysis. DB was

XSL•FO
RenderX

responsible for the data management and programming of participant management software, follow-up reminders, and assessment. AM contributed to study planning and the development of the 2 new ready4life modules, as well as the testing of the modified ready4life app. CM contributed to study design, and obtained the data-gathering process. HS was responsible for recruiting vocational schools, students, and trained teachers to introduce the program to their classes. FR was responsible for conducting focus group discussions with students, teachers, and prevention experts; a phase-1 Delphi study to inform the modification of the ready4life app; and the development of the standard operating procedures of implementation. MN and AD were involved in the recruitment of vocational schools, students, and trained teachers to introduce the program to their classes. They also carried out the primary efficacy outcome analysis of the trial. AB and SS were involved in the development of the additional modules of cannabis and internet use. GB contributed to the study design and obtained the data-gathering process. LH was involved in the recruitment of vocational schools and students. SW provided supervision and resources for data management and follow-up assessments. SB provided resources. SH was responsible for the development of the ready4life app. HJR is the principal investigator of this study and was involved in all steps of the study process. All authors contributed to and have approved the final manuscript.

Conflicts of Interest

SH played a leading role in the initial development of ready4life. DG, AM, CM, FR, AB, SS, HJR, and SH were involved in the further development of the latest version of ready4life tested in this study.

Multimedia Appendix 1
Supplementary tables.
[DOC File , 303 KB - mhealth_v13i1e59573_app1.doc]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1028 KB - mhealth_v13i1e59573_app2.pdf]

Multimedia Appendix 3
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[DOC File , 84 KB - mhealth_v13i1e59573_app3.doc]

Multimedia Appendix 4
TIDieR (Template for Intervention Description and Replication) checklist.
[DOC File , 68 KB - mhealth_v13i1e59573_app4.doc]

References

1. Orth B, Merkel C. Der substanzkonsum jugendlicher und junger erwachsener in Deutschland. Ergebnisse des alkoholsurveys 2021 zu alkohol, rauchen, cannabis und trends [The substance use of adolescents and young adults in Germany. Results of the Alcohol Survey 2021 on alcohol, smoking, cannabis and trends]. Bundeszentrale für Gesundheitliche Aufklärung. 2022. URL: https://www.bioeg.de/fileadmin/user_upload/PDF/studien/BZgA_Alkoholsurvey_2021.pdf [accessed 2025-05-31]
2. Wartberg L, Kriston L, Thomasius R. The prevalence and psychosocial correlates of internet gaming disorder. Dtsch Arztebl Int 2017 Jun 23;114(25):419-424 [FREE Full text] [doi: 10.3238/arztebl.2017.0419] [Medline: 28683852]
3. Onrust SA, Otten R, Lammers J, Smit F. School-based programmes to reduce and prevent substance use in different age groups: what works for whom? Systematic review and meta-regression analysis. Clin Psychol Rev 2016 Mar;44:45-59 [FREE Full text] [doi: 10.1016/j.cpr.2015.11.002] [Medline: 26722708]
4. Montag J, Hanewinkel R, Morgenstern M. Verbreitung und Korrelate des Substanzkonsums unter 5 688 Auszubildenden an beruflichen Schulen [Prevalence and correlates of substance use in 5688 vocational school students]. Gesundheitswesen 2015 Jun 22;77(6):411-417. [doi: 10.1055/s-0034-1382043] [Medline: 25244547]
5. Meyer C, Jahnel T, Freyer-Adam J, Ulbricht S, Hanke M, John U. Konsum von Glücksspielen, Medien, Cannabis, Alkohol und Tabak bei Jugendlichen und Jungen Erwachsenen in Beruflichen Schulen und Produktionsschulen Mecklenburg-Vorpommerns: Eine Landesrepräsentative Querschnittserhebung [Gambling, media, cannabis, alcohol and tobacco consumption among adolescents and young adults in vocational schools and production schools in Mecklenburg-Vorpommern: A state-representative cross-sectional survey]. Greifswald, Germany: Institut für Sozialmedizin und Prävention; 2016.
6. Schmidt H, Brandt D, Meyer C, Bischof A, Bischof G, Trachte A, et al. Motivational brief interventions for adolescents and young adults with internet use disorders: a randomized-controlled trial. J Behav Addict 2022 Sep 26;11(3):754-765 [FREE Full text] [doi: 10.1556/2006.2022.00049] [Medline: 36112487]
7. Atorkey P, Byaruhanga J, Paul C, Wiggers J, Bonevski B, Tzelepis F. Multiple health risk factors in vocational education students: a systematic review. Int J Environ Res Public Health 2021 Jan 13;18(2):637 [FREE Full text] [doi: 10.3390/ijerph18020637] [Medline: 33451108]

XSL•FO
RenderX

8. Pietsch B, Arnaud N, Lochbühler K, Rossa M, Kraus L, Gomes de Matos E, et al. Effects of an app-based intervention program to reduce substance use, gambling, and digital media use in adolescents and young adults: a multicenter, cluster-randomized controlled trial in vocational schools in Germany. Int J Environ Res Public Health 2023 Jan 20;20(3):1970 [FREE Full text] [doi: 10.3390/ijerph20031970] [Medline: 36767337]
9. Haug S, Boumparis N, Wenger A, Schaub MP, Paz Castro R. Efficacy of a mobile app-based coaching program for addiction prevention among apprentices: a cluster-randomized controlled trial. Int J Environ Res Public Health 2022 Nov 25;19(23):15730 [FREE Full text] [doi: 10.3390/ijerph192315730] [Medline: 36497804]
10. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
11. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
12. Greaney ML, Puleo E, Bennett GG, Haines J, Viswanath K, Gillman MW, et al. Factors associated with choice of web or print intervention materials in the healthy directions 2 study. Health Educ Behav 2014 Feb;41(1):52-62 [FREE Full text] [doi: 10.1177/1090198113486803] [Medline: 23720532]
13. Crutzen R, de Nooijer J, Candel MJ, de Vries NK. Adolescents who intend to change multiple health behaviours choose greater exposure to an internet-delivered intervention. J Health Psychol 2008 Oct;13(7):906-911. [doi: 10.1177/1359105308095064] [Medline: 18809641]
14. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an Internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595. [doi: 10.1093/her/cyp063] [Medline: 19897515]
15. Schulz DN, Kremers SP, De Vries H. Tailored eHealth lifestyle promotion: which behavioral modules do users prefer? J Health Commun 2015;20(6):663-672. [doi: 10.1080/10810730.2015.1012243] [Medline: 25928161]
16. Coumans JM, Bolman CA, Oenema A, Lechner L. Predictors of self-determined module choice in a web-based computer-tailored diet and physical activity intervention: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2020 Jul 23;22(7):e15024 [FREE Full text] [doi: 10.2196/15024] [Medline: 32706659]
17. Paz Castro R, Meyer C, Rumpf HJ, Schaub MP, Wenger A, Haug S. Präventionsbedarfe und -interessen bei Berufslernenden [Prevention needs and interests among vocational students]. Präv Gesundheitsf 2022 Mar 16;18(1):99-110 [FREE Full text] [doi: 10.1007/s11553-022-00936-2]
18. Keller PA, Harlam B, Loewenstein G, Volpp KG. Enhanced active choice: a new method to motivate behavior change. J Consum Psychol 2011 Jul 22;21(4):376-383. [doi: 10.1016/j.jcps.2011.06.003]
19. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
20. Stassen G, Grieben C, Froböse I, Schaller A. Engagement with a web-based health promotion intervention among vocational school students: a secondary user and usage analysis. Int J Environ Res Public Health 2020 Mar 25;17(7):2180 [FREE Full text] [doi: 10.3390/ijerph17072180] [Medline: 32218251]
21. Ondersma SJ, Ellis JD, Resko SM, Grekin E. Technology-delivered interventions for substance use among adolescents. Pediatr Clin North Am 2019 Dec;66(6):1203-1215. [doi: 10.1016/j.pcl.2019.08.009] [Medline: 31679608]
22. Andersson G, Estling F, Jakobsson E, Cuijpers P, Carlbring P. Can the patient decide which modules to endorse? An open trial of tailored internet treatment of anxiety disorders. Cogn Behav Ther 2011;40(1):57-64. [doi: 10.1080/16506073.2010.529457] [Medline: 21337215]
23. Schulz DN, Schneider F, de Vries H, van Osch LA, van Nierop PW, Kremers SP. Program completion of a web-based tailored lifestyle intervention for adults: differences between a sequential and a simultaneous approach. J Med Internet Res 2012 Mar 08;14(2):e26 [FREE Full text] [doi: 10.2196/jmir.1968] [Medline: 22403770]
24. Wilson K, Senay I, Durantini M, Sánchez F, Hennessy M, Spring B, et al. When it comes to lifestyle recommendations, more is sometimes less: a meta-analysis of theoretical assumptions underlying the effectiveness of interventions promoting multiple behavior domain change. Psychol Bull 2015 Mar;141(2):474-509 [FREE Full text] [doi: 10.1037/a0038295] [Medline: 25528345]
25. Atorkey P, Paul C, Wiggers J, Bonevski B, Mitchell A, Tzelepis F. Intention to change multiple health risk behaviors and predictors of behavior change in vocational education students. J Am Coll Health 2024 May 12;72(4):1135-1143. [doi: 10.1080/07448481.2022.2068960] [Medline: 35549632]
26. Hughes SL, Seymour RB, Campbell RT, Shaw JW, Fabiyi C, Sokas R. Comparison of two health-promotion programs for older workers. Am J Public Health 2011 May;101(5):883-890. [doi: 10.2105/ajph.2010.300082]
27. Campbell MK, Tessaro I, DeVellis B, Benedict S, Kelsey K, Belton L, et al. Tailoring and targeting a worksite health promotion program to address multiple health behaviors among blue-collar women. Am J Health Promot 2000 Aug 25;14(5):306-313. [doi: 10.4278/0890-1171-14.5.306] [Medline: 11009857]
28. Conner M, Wilding S, Prestwich A, Hutter R, Hurling R, Harreveld FV, et al. Goal prioritization and behavior change: evaluation of an intervention for multiple health behaviors. Health Psychol 2022 May;41(5):356-365. [doi: 10.1037/hea0001149] [Medline: 35467903]

XSL•FO
RenderX

29. Kanera IM, Willems RA, Bolman CA, Mesters I, Zambon V, Gijsen BC, et al. Use and appreciation of a tailored self-management eHealth intervention for early cancer survivors: process evaluation of a randomized controlled trial. J Med Internet Res 2016 Aug 23;18(8):e229. [doi: 10.2196/jmir.5975] [Medline: 27554525]
30. Quintiliani LM, Campbell M, Bowling JM, Steck S, Haines PS, DeVellis BM. Results of a randomized trial testing messages tailored to participant-selected topics among female college students: physical activity outcomes. J Phys Act Health 2010 Jul;7(4):517-526 [FREE Full text] [doi: 10.1123/jpah.7.4.517] [Medline: 20683094]
31. Voncken-Brewster V, Amoureus M, de Vries H, Nagykaldi Z, Winkens B, van der Weijden T, et al. The impact of participant characteristics on use and satisfaction of a web-based computer-tailored chronic obstructive pulmonary disease self-management intervention: a process evaluation. JMIR Form Res 2017 Aug 21;1(1):e1. [doi: 10.2196//formative.6585] [Medline: 30684399]
32. Allegrante JP, Peterson JC, Boutin-Foster C, Ogedegbe G, Charlson ME. Multiple health-risk behavior in a chronic disease population: what behaviors do people choose to change? Prev Med 2008 Mar;46(3):247-251. [doi: 10.1016/j.ypmed.2007.09.007] [Medline: 17996930]
33. Graham AK, Munson SA, Reddy M, Neubert SW, Green EA, Chang A, et al. Integrating user-centered design and behavioral science to design a mobile intervention for obesity and binge eating: mixed methods analysis. JMIR Form Res 2021 May 10;5(5):e23809. [doi: 10.2196/23809] [Medline: 33970114]
34. Reinwand DA, Schulz DN, Crutzen R, Kremers SP, de Vries H. Who follows eHealth interventions as recommended? A study of participants' personal characteristics from the experimental arm of a randomized controlled trial. J Med Internet Res 2015 May 11;17(5):e115. [doi: 10.2196/jmir.3932] [Medline: 25963607]
35. Lochbühler K, Rossa M, Ebert C, Morgenstern M, Arnaud N, Kraus L. Substance use and the usage of social media, computer games, and gambling among apprentices at vocational schools. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2024 Apr;67(4):465-474 [FREE Full text] [doi: 10.1007/s00103-024-03854-0] [Medline: 38526678]
36. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
37. McHugh RK, Votaw VR, Sugarman DE, Greenfield SF. Sex and gender differences in substance use disorders. Clin Psychol Rev 2018 Dec;66:12-23. [doi: 10.1016/j.cpr.2017.10.012] [Medline: 29174306]
38. Schmidt H, Brandt D, Bischof A, Bischof G, Sürig S, Gürtler D, et al. App-based coaching to prevent addictive behaviors among young adults. Sucht 2023;69(2) [FREE Full text] [doi: 10.1024/0939-5911/a000811]
39. Guertler D, Bläsing D, Moehring A, Meyer C, Brandt D, Schmidt H, et al. App-based addiction prevention at German vocational schools: implementation and reach for a cluster-randomized controlled trial. Prev Sci 2024 Jul;25(5):849-860 [FREE Full text] [doi: 10.1007/s11121-024-01702-w] [Medline: 38958917]
40. Krause K, Guertler D, Moehring A, Batra A, Eck S, Rumpf HJ, et al. Feasibility and acceptability of an intervention providing computer-generated tailored feedback to target alcohol consumption and depressive symptoms in proactively recruited health care patients and reactively recruited media volunteers: results of a pilot study. Eur Addict Res 2019;25(3):119-131 [FREE Full text] [doi: 10.1159/000499040] [Medline: 30917380]
41. Velicer WF, Prochaska JO, Fava JL, Rossi JS, Redding CA, Laforge RG, et al. Using the Transtheoretical Model for population-based approaches to health promotion and disease prevention. Homeost Health Dis 2000;40(5):174-195 [FREE Full text]
42. Band 1: systematischer und alphabetischer teil mit erläuterungen [Federal Employment Agency, Classification of Occupations 2010 - revised version 2020, Volume 1: Systematic and alphabetical part with explanations]. Bundesagentur für Arbeit. URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Printausgabe-KldB-2010-Fassung2020/Generische-Publikationen/KldB2010-PDF-Version-Band1-Fassung2020.html?__blob=publicationF [accessed 2025-05-23]
43. Ganzeboom HB. International standard classification of occupations ISCO-08 with ISEI-08 scores. In: Proceedings of the Annual Conference of International Social Survey Programme. 2010 Presented at: ISSP 2010; May 1, 2010; Lisbon, Portugal URL: http://www.harryganzeboom.nl/ISCO08/isco08_with_isei.pdf
44. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
45. Besser B, Rumpf HJ, Bischof A, Meerkerk GJ, Higuchi S, Bischof G. Internet-related disorders: development of the short compulsive internet use scale. Cyberpsychol Behav Soc Netw 2017 Nov;20(11):709-717. [doi: 10.1089/cyber.2017.0260] [Medline: 29125788]
46. Elo AL, Leppänen A, Jahkola A. Validity of a single-item measure of stress symptoms. Scand J Work Environ Health 2003 Dec;29(6):444-451 [FREE Full text] [doi: 10.5271/sjweh.752] [Medline: 14712852]
47. Gambrill ED, Richey CA. An assertion inventory for use in assessment and research. Behav Ther 1975 Jul;6(4):550-561. [doi: 10.1016/S0005-7894(75)80013-X]

XSL•FO
RenderX

48. Haug S, Castro RP, Wenger A, Schaub MP. Efficacy of a smartphone-based coaching program for addiction prevention among apprentices: study protocol of a cluster-randomised controlled trial. BMC Public Health 2020 Dec 14;20(1):1910-1918 [FREE Full text] [doi: 10.1186/s12889-020-09995-6] [Medline: 33317488]
49. Botschaften für eine orientierungshilfe zum alkoholkonsum kurzversion. Schweizerische Eidgenossenschaft. URL: https://www.aktionstag-alkoholprobleme.ch/files/simple-file/EKAL-Orientierungshilfe-zum-Alkoholkonsum_Kurzversion.pdf [accessed 2025-05-23]
50. Tomczyk S, Pedersen A, Hanewinkel R, Isensee B, Morgenstern M. Polysubstance use patterns and trajectories in vocational students--a latent transition analysis. Addict Behav 2016 Jul;58:136-141. [doi: 10.1016/j.addbeh.2016.02.027] [Medline: 26943488]
51. de Jonge MC, Bukman AJ, van Leeuwen L, Onrust SA, Kleinjan M. Latent classes of substance use in young adults - a systematic review. Subst Use Misuse 2022;57(5):769-785 [FREE Full text] [doi: 10.1080/10826084.2022.2040029] [Medline: 35188870]
52. Schulz D, Kremers S, Vandelanotte C, van Adrichem MJ, Schneider F, Candel M, et al. Effects of a web-based tailored multiple-lifestyle intervention for adults: a two-year randomized controlled trial comparing sequential and simultaneous delivery modes. J Med Internet Res 2014 Jan 27;16(1):e26 [FREE Full text] [doi: 10.2196/jmir.3094] [Medline: 24472854]
53. Locke E, Latham G. Goal-setting theory. In: Organizational Behavior. Milton Park, UK: Routledge; 2005:159-183.
54. Bandura A, National Institute of Mental Health. Social Foundations of Thought and Action: A Social Cognitive Theory. Hoboken, NJ: Prentice-Hall; 1986.
55. Schwarzer R. Self-regulatory processes in the adoption and maintenance of health behaviors. J Health Psychol 1999 Mar;4(2):115-127 [FREE Full text] [doi: 10.1177/135910539900400208] [Medline: 22021474]
56. Twisk JW. Applied Multilevel Analysis: A Practical Guide for Medical Researchers. Cambridge, UK: Cambridge University Press; 2006.
57. Hanke M, Ulbricht S, Freyer-Adam J, John U, Meyer C, Haug S. Tabakrauchen und Alkoholkonsum bei Auszubildenden an beruflichen Schulen in Vorpommern [Tobacco smoking and alcohol consumption among apprentices at vocational schools in West Pomerania, Germany]. Gesundheitswesen 2013 Apr 3;75(4):216-224. [doi: 10.1055/s-0032-1311621] [Medline: 22864848]
58. Haug S, Schaub MP, Salis Gross C, John U, Meyer C. Predictors of hazardous drinking, tobacco smoking and physical inactivity in vocational school students. BMC Public Health 2013 May 15;13:475-479 [FREE Full text] [doi: 10.1186/1471-2458-13-475] [Medline: 23672294]
59. Paz Castro R, Haug S, Debelak R, Jakob R, Kowatsch T, Schaub MP. Engagement with a mobile phone-based life skills intervention for adolescents and its association with participant characteristics and outcomes: tree-based analysis. J Med Internet Res 2022 Jan 19;24(1):e28638 [FREE Full text] [doi: 10.2196/28638] [Medline: 35044309]
60. Grahlher K, Morgenstern M, Pietsch B, Gomes de Matos E, Rossa M, Lochbühler K, et al. Mobile app intervention to reduce substance use, gambling, and digital media use in vocational school students: exploratory analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2024 Jul 23;12:e51307 [FREE Full text] [doi: 10.2196/51307] [Medline: 39042436]
61. Spielmann M, Krolo-Wicovsky F, Tiede A, Krause K, Baumann S, Siewert-Markus U, et al. Patient motivation and preferences in changing co-occurring health risk behaviors in general hospital patients. Patient Educ Couns 2023 Sep;114:107841. [doi: 10.1016/j.pec.2023.107841] [Medline: 37354731]
62. Schomerus G, Lucht M, Holzinger A, Matschinger H, Carta MG, Angermeyer MC. The stigma of alcohol dependence compared with other mental disorders: a review of population studies. Alcohol Alcohol 2011;46(2):105-112. [doi: 10.1093/alcalc/agq089] [Medline: 21169612]
63. Johnson S, Cooper C, Cartwright S, Donald I, Taylor P, Millet C. The experience of work‐related stress across occupations. J Manag Psychol 2005;20(2):178-187 [FREE Full text] [doi: 10.1108/02683940510579803]
64. Horta Esper L, Furtado EF. Gender differences and association between psychological stress and alcohol consumption: a systematic review. J Alcohol Drug Depend 2013;1:116. [doi: 10.4172/2329-6488.1000116]
65. Zakiniaeiz Y, Potenza MN. Gender-related differences in addiction: a review of human studies. Curr Opin Behav Sci 2018 Oct;23:171-175 [FREE Full text] [doi: 10.1016/j.cobeha.2018.08.004] [Medline: 39896826]
66. Nolen-Hoeksema S. Gender differences in risk factors and consequences for alcohol use and problems. Clin Psychol Rev 2004 Dec;24(8):981-1010 [FREE Full text] [doi: 10.1016/j.cpr.2004.08.003] [Medline: 15533281]
67. Jolles MP, Fort MP, Glasgow RE. Aligning the planning, development, and implementation of complex interventions to local contexts with an equity focus: application of the PRISM/RE-AIM Framework. Int J Equity Health 2024 Feb 27;23(1):41 [FREE Full text] [doi: 10.1186/s12939-024-02130-6] [Medline: 38408990]
68. Littlejohn C. Does socio-economic status influence the acceptability of, attendance for, and outcome of, screening and brief interventions for alcohol misuse: a review. Alcohol Alcohol 2006;41(5):540-545 [FREE Full text] [doi: 10.1093/alcalc/agl053] [Medline: 16855002]
69. Paz Castro R, Haug S, Kowatsch T, Filler A, Schaub MP. Moderators of outcome in a technology-based intervention to prevent and reduce problem drinking among adolescents. Addict Behav 2017 Sep;72:64-71 [FREE Full text] [doi: 10.1016/j.addbeh.2017.03.013] [Medline: 28371696]

XSL•FO
RenderX

70. Katikireddi SV, Whitley E, Lewsey J, Gray L, Leyland AH. Socioeconomic status as an effect modifier of alcohol consumption and harm: analysis of linked cohort data. Lancet Public Health 2017 Jun;2(6):e267-e276 [FREE Full text] [doi: 10.1016/S2468-2667(17)30078-6] [Medline: 28626829]
71. Paschke K, Austermann MI, Simon-Kutscher K, Thomasius R. Adolescent gaming and social media usage before and during the COVID-19 pandemic. SUCHT 2021;67(1) [FREE Full text] [doi: 10.1024/0939-5911/a000694]

Abbreviations

CIUS: Compulsive Internet Use Scale
ICC: intraclass correlation coefficient
ISCO-8: International Standard Classification of Occupations 2008
OR: odds ratio

Edited by L Buis; submitted 16.04.24; peer-reviewed by A Garcia-Perez, TTO Kwok; comments to author 14.10.24; revised version received 04.11.24; accepted 28.04.25; published 24.06.25.

Please cite as:
Guertler D, Kraft E, Bläsing D, Möhring A, Meyer C, Schmidt H, Rehbein F, Neumann M, Dreißigacker A, Bischof A, Bischof G, Sürig S, Hohls L, Wurm S, Borgwardt S, Haug S, Rumpf HJ
Prevention Needs and Target Behavior Preferences in an App-Based Addiction Prevention Program for German Vocational School Students: Cluster Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e59573
URL: https://mhealth.jmir.org/2025/1/e59573
doi:10.2196/59573
PMID:40553509

©Diana Guertler, Elaine Kraft, Dominic Bläsing, Anne Möhring, Christian Meyer, Hannah Schmidt, Florian Rehbein, Merten Neumann, Arne Dreißigacker, Anja Bischof, Gallus Bischof, Svenja Sürig, Lisa Hohls, Susanne Wurm, Stefan Borgwardt, Severin Haug, Hans-Jürgen Rumpf. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Parental and Demographic Predictors of Engagement in an mHealth Intervention: Observational Study From the Let's Grow Trial

Johanna Sandborg[1], PhD; Brittany Reese Markides[1], PhD; Savannah Simmons[1], MDiet; Katherine L Downing[1], PhD; Jan M Nicholson[2], PhD; Liliana Orellana[3], PhD; Harriet Koorts[1], PhD; Valerie Carson[4], PhD; Jo Salmon[1], PhD; Kylie D Hesketh[1], PhD

[1]Institute for Physical Activity and Nutrition (IPAN), Faculty of Health, Deakin University, Geelong, Victoria, Australia

[2]Judith Lumley Centre, La Trobe University, Bundoora, Australia

[3]Faculty of Health Biostatistics Unit, Deakin University, Geelong, Australia

[4]Faculty of Kinesiology, Sport, and Recreation, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Johanna Sandborg, PhD
Institute for Physical Activity and Nutrition (IPAN)
Faculty of Health
Deakin University
221 Burwood Highway, Burwood
Geelong, Victoria, 3125
Australia
Phone: 46 733285912
Email: johanna.sandborg@ki.se

Abstract

Background: Parents are integral in shaping early childhood health behaviors, and mobile health (mHealth) interventions offer an accessible method of supporting them in this role. Optimizing participant engagement is key to mHealth effectiveness and impact; however, research examining personal predictors of engagement remains underexplored.

Objective: We aimed to describe participant engagement with a novel parental mHealth intervention (Let's Grow) during the first 25 weeks of use and investigate whether engagement levels varied by family demographics and parental cognitions and behaviors relevant to the intervention.

Methods: We used data from parents in the intervention group of the Let's Grow trial. The intervention targeted toddlers' movement behaviors, and the program (a purpose-designed progressive web app) was delivered via self-paced modules. The content was built around 3 main components (behavior change activities, information provision, and social support). Engagement data (web app analytics) collected across the first 25 weeks of the intervention were summarized as study-specific metrics (time using the app, proportion of accessed features and pages, clicks in the main parts of the app) and overall engagement measures (composite engagement index [EI], individual subindexes [click depth, loyalty, recency, and diversity]). The baseline measures included family demographics (main carer, child and family characteristics, and postcode) and parental cognitions and behaviors relevant to the intervention (coping, concern, and information seeking). Linear regression was used to assess associations between baseline and engagement measures.

Results: All parents allocated to the intervention group (n=682) were included. Most parents (609/682, 89.3%) logged in and used at least 1 app feature; those who never used the app were excluded from subsequent analyses. App access declined from 90.6% (552/609) in the first week to 31.2% (190/609) at 25 weeks. For users active during weeks 12 to 25, EI remained consistent and was nearly identical to the average EI (28%, range 3%-50%). More work hours, parents living together, having siblings in the family, and living in a regional or remote area were each associated with lower engagement on 10 out of 12 indicators (β=−31.37 to −0.01; all P≤.046). Higher education level was associated with higher engagement on 9 indicators (β=0.77-18.59; all P≤.02). Of the parental characteristics, only higher coping was positively associated with engagement (β=1.25; P=.003).

Conclusions: Our findings indicate that time and sociodemographic factors might be the most relevant predictors of engagement and highlight the characteristics of parents who may benefit from more active strategies to support their engagement with digital

XSL•FO
RenderX

interventions. The uptake and continued engagement with this app exceeded what is generally reported for apps, but it is unknown whether this is sufficient for behavior change. Individual and composite engagement measures yielded similar results, indicating that simpler, more feasible metrics can be useful for reporting engagement in digital interventions.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2021-057521

(JMIR Mhealth Uhealth 2025;13:e60478) doi:10.2196/60478

KEYWORDS

digital intervention; parenting; movement behaviors; sleep; screen time; early childhood

Introduction

Background

Early childhood (age 0-5 years) has been identified as a key time to promote healthy behaviors (eg, physical activity) [1]. During these formative years, parents play an important role in providing opportunities for their children to develop these behaviors, and parental encouragement and support have been shown to positively influence children's health behaviors [2,3]. Parental involvement in childhood interventions has also been shown to be effective in promoting a healthy diet and physical activity in children [4], with greater effects from interventions that include parenting skill training and behavior change strategies [5]. Mobile health (mHealth) interventions (eg, mobile apps, websites, or messaging services) offer a convenient and accessible way to provide support and timely information to parents and can enable participation regardless of geographic location. This delivery mode has been shown to be well accepted among parents and effective in promoting health behaviors in children [6-8].

The digital format of mHealth interventions also enables the use of built-in tools for the objective measurement of user engagement (ie, *how* and *to what extent* users engage with the intervention), which is directly linked to intervention effectiveness [9]. Examining user engagement is a critical component of evaluating mHealth interventions as it provides information on levels of intervention exposure and uptake and can facilitate the identification of factors that may optimize engagement [10]. Understanding who and how participants engage with digital interventions can also provide insight into the mechanisms that drive intervention effectiveness and contribute to the design of more impactful interventions. However, there is a notable lack of reporting on engagement and its predictors in mHealth interventions targeting parents of young children, with only a few studies (eg, [11-15]) presenting parent engagement using objective data rather than relying on parental reports. Among the studies that have investigated predictors of engagement in mHealth parental interventions [11-14], only family demographics (parental education [11-14], number of children in the household [11,13], and work status [12]) have been examined. No previous studies have investigated parental cognitions and behaviors despite evidence that parental constructs, such as parental confidence, can predict dropout rate in traditional family-based interventions [16]. Thus, little is known about how parents of young children engage with mHealth interventions or about the potential influence that family demographics and parental cognitions and behaviors may have on parents' engagement with these interventions.

Let's Grow is a novel mHealth intervention designed to support parents in improving the movement behaviors (ie, physical activity, sedentary behavior, and sleep) of their children aged 2 years via a purpose-built app [17]. The program is delivered via modules that parents work through over time, and the content is built around 3 main components, that is, behavior change activities, information provision, and social support, all of which have been identified as key strategies to promote health behaviors [18].

Objectives

To the best of our knowledge, no previous study has comprehensively investigated engagement and predictors of engagement in a parental mHealth intervention. Thus, we aimed to use data from the Let's Grow trial [17] to (1) describe parental engagement with the Let's Grow app during the first 25 weeks of use and (2) investigate whether engagement levels varied by family demographics and parental cognitions and behaviors relevant to the intervention (hereafter referred to as parental characteristics).

Methods

Study Design

We used data from the Let's Grow randomized controlled trial (ACTRN12620001280998; U1111-1252-0599) [17] in this observational study. The trial design and rationale have been described previously [17]. In brief, the trial investigated a 12-month mHealth intervention (delivered through a progressive web app, the Let's Grow app) aimed at supporting parents of children aged 2 years to improve their child's movement behaviors [17]. This study is reported according to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.

Ethical Considerations

The Let's Grow trial received ethics approval from the Deakin University Human Research Ethics Committee, Australia (2020-077). The main carer was provided with details of the study commitment before providing consent for their own participation and that of their child via an online form. Participants received an Aus $20 (US $13) retail gift voucher upon completion of the study assessments (survey and activity monitors) at each time point; no compensation was provided for engaging with the app. All participant data were deidentified and securely stored in accordance with the Deakin University protocols.

XSL•FO
RenderX

Recruitment, Randomization, and Procedures

Participants were recruited through social media (eg, Facebook and parenting blogs) between March 2021 and June 2022. Parents aged ≥18 years, residing in Australia, owning a mobile phone, having internet access, literate in English, and with a child aged 18 to 35 months walking independently were eligible. Following baseline assessment, participants were randomized in a 1:1 ratio stratified by geographical location (urban, or outer regional and remote), yielding 16 strata across the 8 Australian states and territories. The intervention group (n=682) was then provided access to the Let's Grow app. The baseline survey data and app analytics metrics of the intervention group participants collected during the first 25 weeks of app access were used for this study.

The Let's Grow App

The development and content of the Let's Grow app have been described in detail elsewhere [17]. In brief, the app was built around 3 main components, including active learning targeting changes in specific combinations of movement behaviors (learning modules), general information provision (toolkit), and social support (community chat forum) to help parents improve their child's physical activity, sedentary behavior (including screen time), and sleep. The active learning component consisted of 8 modules: "Switch off and play," "Play, sleep, repeat," "Avoid light sleep tight," "Play skills for life," "Build your best day," "Parents provide, kids decide," "Rocking routines," and "Calm families". Each module contained information and strategies (in text, pictorial, podcast, and video format) on child sleep, play, screen time, and general parenting, as well as a set of behavior change activities, including both online and offline components. The modules were not mutually exclusive to 1 behavior as all modules covered parenting and behavior change strategies for at least 2 out of the 3 target child behaviors, with a focus on the interplay between child movement behaviors (eg, increasing physical activity to decrease sedentary time). To further support parents, each module also included SMS text messaging (eg, interactive messages that facilitate goal setting and self-monitoring).

While the parents could complete the modules in any order, they were only able to access 1 module at a time and were required to complete all activities within that module before moving on to another. As activities took varying amounts of time to complete (eg, self-monitoring occurred across a week), the minimum amount of time to complete a module before moving on to the next ranged from 2 to 5 weeks. The completed modules remained accessible for the participants to revisit information. However, if they did not complete a module, the information contained in the other modules would be inaccessible to them. The intervention duration was 12 months; the minimum timeframe to complete all the 8 modules was 22 weeks, but participants were not expected to complete all the modules within this timeframe. Upon completion of all the modules, the participants received SMS text messaging notifications once a week to refresh their knowledge and encourage revisiting of the app. The toolkit and community chat forum remained freely accessible throughout the intervention period. The toolkit included resources such as outdoor and indoor play ideas, screen-free activities, a podcast providing help with children's common sleep issues, and videos explaining different parenting strategies in relation to managing difficult child behavior. The forum provided the parents with the opportunity to anonymously share ideas and connect with other participants. The parents were prompted with SMS text message reminders to reengage with the app if no activity was detected for an extended period (eg, if the app had not been opened for 3 weeks). An overview with screenshots of the Let's Grow app is presented in Multimedia Appendix 1.

Factors Related to Engagement

At baseline, the main carer completed an online survey (using REDCap [Research Electronic Data Capture; Vanderbilt University]) capturing family demographics and parental characteristics. Main carer characteristics considered in this study included age (years), relationship with the child (mother or father), birth country (Australian-born, yes or no), education level (university degree vs no university degree), and work status (currently working, yes or no). Participant child characteristics included sex and temperament in comparison to other children (single item; scale 1-5 ranging from much easier than average to much more difficult than average) [19]. Family characteristics included parents living together (yes or no) and siblings in the family (yes or no). Postcode was used to derive information on remoteness, that is, major cities versus inner or outer regional and rural Australia, which was based on the Australian Bureau of Statistics Australian Statistical Geography Standard Remoteness Structure [20], as well as area-level socioeconomic status derived using the Australian Bureau of Statistics Socio-Economic Indexes for Areas Index of Relative Socio-economic Advantage and Disadvantage [21].

Three parental characteristics relevant to the intervention were considered: parent coping in general, concern about the targeted topic (movement behaviors), and information seeking related to the targeted topic. Parents reported using a 5-point scale (1=not at all, 5=extremely well, possible range 1-5) how they were coping with life at present [22]. In terms of concern with child movement behaviors, the parents reported about their child's sitting time, physical activity or active play, screen use, and sleep to indicate whether they considered it a problem using a 5-point scale (1=not a problem at all to 5=a serious problem) [23]. A combined movement behavior concern score was calculated as the sum of the individual scores (possible range 4-20; higher scores indicating greater concern) [23]. Finally, the parents reported how much time in a usual week they spend using apps or websites to access information on children's physical activity, sitting time or screen use, and sleep, respectively. The total reported time searching for information on child movement behaviors and child health topics was calculated (h per wk; possible range 0-40 h per wk).

Metrics of Engagement

Data on users' activities and actions in the Let's Grow app were logged to a dedicated database connected to the app's backend system. These data were used to report user engagement descriptively, such as the proportion of participants who initiated app use (ie, intervention uptake), the types of features they used (ie, learning modules, toolkit, and community chat forum), the

XSL•FO
RenderX

number of modules they completed, the total time spent in the app (min), the total days of app use, the proportion of pages (%) and features (%) visited at least once throughout the 25 weeks, and the number of actions (posts, likes, and comments) in the community chat forum.

To assess overall engagement among the Let's Grow app users, an adapted version of a previously developed engagement index (EI) [11,24] was calculated. This index assessed different aspects of engagement compared to the descriptive engagement measures and is more broadly applicable across different apps. As per previous studies [11,14,24], the 4 subindexes relevant to this program, from the original 7 in the index, were used. These included the following: (1) click depth (average number of app pages viewed per day); (2) loyalty (total days with any use data for a participant); (3) recency (average number of days between app use; reversed so a higher score indicates greater engagement); and (4) diversity (the number of different app features used per day, ie, activities within modules, toolkit, community chat forum, and frequently asked questions). A detailed description of the individual subindexes is presented in Multimedia Appendix 2. The subindexes were normalized by rescaling values between 0% and 100% to assign equal weight to each of the subindexes. The EI was calculated as the average of the 4 subindexes. In addition, a weekly EI (range 0%-100%) was calculated for those with engagement data in any particular week to enable a more granular investigation into participant engagement over the 25 weeks. All calculations were carried out using Python (version 3; Python Software Foundation).

Covariates

The Let's Grow trial was conducted during the COVID-19 pandemic (baseline data collection from March 2021 to June 2022; the data for this paper were collected up to January 2023). There were varying pandemic restrictions across different states and territories in Australia, which potentially impacted the everyday life of the participating families and their ability to engage with the program. Thus, the data on the number of days in lockdown during the 6-month engagement period were calculated based on individual postcode and publicly available data on local restrictions. In addition, due to technical issues, the app was not functioning at certain periods; nevertheless, due to the staggered recruitment, not all participants experienced equal app downtime, which might have differentially affected the engagement. The app downtime was calculated for each participant using web app analytics.

Data Analysis

All statistical analyses were carried out using the statistical software R (version 4.0.3; R Foundation for Statistical Computing). Chi-square tests for categorical and *t* test for numerical variables were used to investigate differences between those who logged in at least once and used at least 1 app feature (ie, users: 609/682, 89.3%) and those who never logged in or logged in but never used an app feature (ie, nonusers: 73/682, 10.7%). Further analyses involved only the app users. The associations between the potential predictors (family demographics and parental characteristics) and the individual (total time spent in the app and clicks in the main parts of the app: learning modules, toolkit, and community chat forum) and the overall measures of engagement (EI and subindexes) at 25 weeks were estimated using linear regression. Sensitivity analyses were performed to adjust for the lockdown and app downtime by including the number of days in lockdown and the total app downtime (days) in the models. Estimates from the sensitivity analyses are presented alongside the main analyses where results differed.

Results

Participant Characteristics

The baseline characteristics of the 682 participants randomized to the intervention group are reported in Table 1. The participants were, on average, aged 34.9 (SD 4.5; range 21-48) years; the majority were mothers (670/682, 98.2%), Australian-born (522/681, 76.7%), university educated (490/682, 71.8%), lived in major cities (490/682, 71.8%), and lived together with a partner or spouse (629/681, 92.4%). In terms of intervention uptake, 67 (9.8%) of the 682 participants never logged in, and a further 6 (0.9%) did not progress past the log-in page to use any of the app features. Of the 682 participants, the 609 (89.3%) who logged in at least once and used at least 1 app feature were referred to as *users* and were included in the subsequent analyses. The users did not differ from the nonusers in terms of the potential predictors of engagement at baseline, with the exception of reporting a higher coping score (mean score 3.2 vs 3.0; *P*=.01) and a higher level of concern with their child's movement behaviors (mean score 7.4 vs 6.8; *P*=.04).

XSL•FO
RenderX

Table 1. Family demographics and parental characteristics of the participants in the intervention group of the Let's Grow trial (n=682)[a].

Family demographics and parental characteristics	Values
Parent[b]	
Age (y), mean (SD)	34.9 (4.5)
Mother, n (%)	670 (98.2)
Australian-born, n (%)	522 (76.7)
University degree, n (%)	490 (71.8)
Currently working, n (%)	442 (64.9)
Parental coping, mean (SD)	3.2 (0.7)
Concern with their child's movement behaviors, mean (SD)	7.3 (2.6)
Information seeking (h/wk), mean (SD)	2.2 (4.6)
Child	
Female, n (%)	337 (49.4)
Child temperament, n (%)	
Much easier or easier	221 (32.4)
Average	361 (52.9)
More or much more difficult	100 (14.7)
Family, n (%)	
Living with a partner or spouse	629 (92.4)
Siblings in the family	361 (58.8)
Living area and area-level socioeconomic status	
Remoteness level, n (%)	
Major cities	490 (71.8)
Inner regional Australia	131 (19.2)
Outer regional or remote Australia	61 (8.9)
Socio-Economic Indexes for Areas category, n (%)	
Low advantaged/high disadvantaged	136 (20)
Middle advantaged/disadvantaged	223 (32.8)
High advantaged/low disadvantaged	321 (47.2)

[a]Due to missing data, n varied slightly for some variables, with missing data for individual variables ranging from n=0 to n=68.

[b]These data are from the self-selected main carer of the child.

User Engagement With the Let's Grow App

Table 2 presents the different measures of engagement for those classified as users, including overall use and engagement with the different components of the app. Across the 25 weeks, users spent an average of 31 (SD 36) minutes in the app and accessed the app on an average of 18 (SD 15) separate days over this time. The users visited on average, 13% (SD 13%) of the pages and 40% (SD 18%) of the app features, noting that a large proportion of the content was contained within the learning modules and, thus, remained inaccessible until the modules were completed.

XSL•FO
RenderX

Table 2. Objective measures of the app engagement after 6 months in the participants defined as app users (n=609).

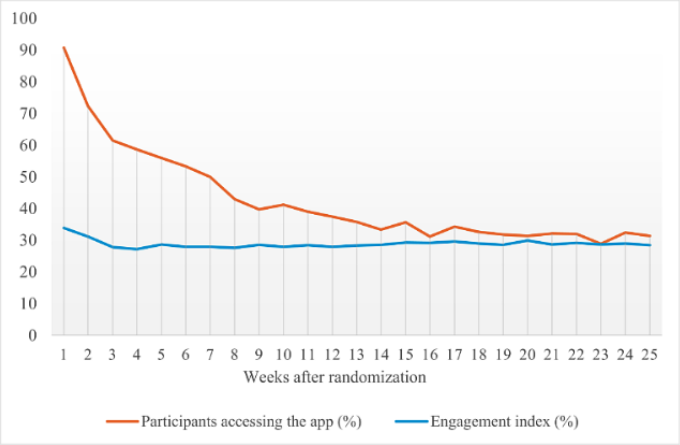
Engagement metrics	Values
Individual engagement metrics	
Overall use	
Total time spent in the app (min), mean (SD; range)	31 (36; 0-270)
App use (d), mean (SD; range)	18 (15; 1-82)
Proportion of content viewed (%; range)	
Pages visited	12.8 (13; 1-77)
Features visited	39.6 (18.5; 5-100)
Engagement in the main parts of the app	
At least 1 time access, n (%)	
Learning modules	505 (82.9)
Toolkit	335 (55)
Community chat forum	177 (29.1)
Clicks (n), mean (SD; range)	
Learning modules	58.8 (74.8; 0-497)
Toolkit	17.6 (40.2; 0-356)
Community chat forum	1.8 (4.8; 0-42)
Composite engagement metrics	
Total engagement index, mean (SD; range)	28 (6.6; 3.2-50)
Subindexes, mean (SD; range)	
Click depth	2.4 (2.4; 0.1-17.8)
Loyalty	9.2 (7.6; 0-46.4)
Recency	87.4 (14.9; 0-100)
Diversity	13 (9.5; 0-66.7)

The highest level of engagement was observed in the learning modules, with 82.9% (505/609) of the users accessing the modules at least once. The average number of completed modules was 2.4 (SD 3.0; range 0-8) out of the possible 8. More than half of the users (329/609, 54%) completed at least 1 module: 18.1% (110/609) completed 1 module, 8.2% (50/609) completed 2 modules, 7.6% (46/609) completed 3 modules, 6.4% (39/609) completed 4 modules, 6.7% (41/609) completed 5 modules, 3.9% (24/609) completed 6 modules, 2.1% (13/609) completed 7 modules, and 1% (6/609) completed all 8 modules by 25 weeks (the mid-intervention point). The information provision component was the second most used part of the app, with 55% (335/609) of the users accessing the toolkit at least once. The lowest level of engagement was observed in the social support component, with 29.1% (177/609) of the users accessing it at least once. The users of the social support feature made between 0 and 37 posts (mean 0.5, SD 2.7), 0 to 96 likes (mean 0.5, SD 4.6), and 0 to 17 comments (mean 0.4, SD 1.7).

The average EI over the 25 weeks was 28% (SD 7%) out of 100%. Regarding the patterns of engagement, the level of engagement (EI) varied throughout the 25 weeks (Figure 1). For those who progressed past the log-in page, the engagement was highest after initial access to the app (552/609, 90.6% of users accessed the app in the first week; average EI in the first week was 34%, SD 12%) and declined during the first 25 weeks of the intervention period (227/609, 37.3% of users accessed the app at 12 weeks and 190/609, 31.2% at 25 weeks) although in a nonlinear fashion. For users with engagement data at both 12 and 25 weeks, the EI was similar at the 2 time points (mean 28%, SD 7%; 217/609, 35.6% and mean 28%, SD 9%; 190/609, 31.2%, respectively).

XSL•FO
RenderX

Figure 1. Average weekly engagement index (EI) over the use period for the participants defined as users (n=609). The line for the participants accessing the app (%), as indicated in the legend, shows the proportion of participants who accessed the app out of the 609 users. The EI (%) comprises 4 subindexes: click depth (the average number of app pages viewed per day), loyalty (the total number of days a user engaged with the app), recency (the average number of days between app use: reversed so higher score indicates greater engagement), and diversity (the number of different app features used per day). The subindexes were normalized by rescaling values between 0 and 100 to assign equal weight to each of the subindexes (each element was equally important in contributing to the measurement of engagement). The EI was then calculated as the arithmetic mean of the 4 subindexes (click depth, loyalty, recency, and diversity).



Associations Between the Predictors and User Engagement

The estimated associations between the potential predictors and the individual (time spent in the app, proportion of pages visited, clicks in the main parts of the app) and the composite (EI and subindexes) indicators of engagement from the unadjusted regression models are presented in Tables 3, 4, and 5, respectively. Overall, the same pattern of associations was observed for the individual and composite measures of engagement.

Higher education level was associated with higher levels of engagement on 4 of the 7 individual indicators and all 5 of the composite indicators. Higher education level was associated with more days of app use (b=3.14, 95% CI 0.50-5.78; *P*=.02); accessing a higher proportion of pages (b=3.27, 95% CI 0.96-5.59; *P*=.006) and features (b=4.33, 95% CI 1.05-7.61; *P*=.01); and more clicks in the modules (b=18.59, 95% CI 5.33-31.86; *P*=.006), as well as higher EI (b=2.61, 95% CI 1.46-3.77; *P*<.001); click depth (b=0.77, 95% CI 0.36-1.19; *P*<.001); loyalty (b=2.48, 95% CI 1.14-3.83; *P*<.001); recency (b=4.64, 95% CI 2.00-7.27; *P*=.001); and diversity index (b=2.56, 95% CI 0.89-4.24; *P*=.003). The parent coping score was positively associated with engagement on 1 composite indicator (ie, loyalty score: b=1.25, 95% CI 0.43-2.08; *P*=.003).

In contrast, higher working hours were associated with lower levels of engagement on 4 of the individual indicators and 1 of the composite indicators. Number of working hours was inversely associated with total time spent in the app (b=−0.24, 95% CI −0.43 to −0.05; *P*=.01); proportion of pages visited (b=−0.08, 95% CI −0.15 to −0.01; *P*=.02); and clicks in the modules (b=−0.40, 95% CI −0.79 to −0.01; *P*=.046) and toolkit (b=−0.27, 95% CI −0.48 to −0.06; *P*=.01), as well as click-depth index (b=−0.01, 95% CI −0.03 to −0.00; *P*=.046). Parents living together were associated with lower levels of engagement on 5 of the individual indicators. Parents living together with a partner spent less time (b=−13.58, 95% CI −24.39 to −2.77; *P*=.01) and days using the app (b=−5.07, 95% CI −9.531 to −0.6; *P*=.03); accessed a lower proportion of pages (b=−4.30, 95% CI −8.22 to −0.38; *P*=.03); and, on average, clicked less times in the modules (b=−31.37, 95% CI −53.79 to −8.95; *P*=.006) and toolkit (b=−12.72, 95% CI 24.82-0.63; *P*=.04). Having siblings in the family was associated with lower engagement on 5 of the individual indicators and 4 of the composite indicators. Having siblings in the family was inversely associated with total time spent in the app (b=−7.11, 95% CI −13.40 to -0.82; *P*=.03); number of days of app use (b=−3.04, 95% CI −5.59 to −0.49; *P*=.02); and clicks in the toolkit (b=−7.75, 95% CI −14.82 to −0.68; *P*=.03), as well as lower EI (b=−1.56, 95% CI −2.65 to −0.48; *P*=.005); click depth (b=−0.43, 95% CI −0.83 to −0.04; *P*=.03); loyalty (b=−2.4, 95% CI −3.68 to −1.12; *P*<.001); and recency index (b=−2.77, 95% CI −5.17 to −0.36; *P*=.02). Living outside a metropolitan area was associated with lower engagement on 1 of the composite indicators (ie, click-depth index: b=−0.47; 95% CI −0.88 to −0.05; *P*=.03).

No associations were observed between any of the predictor variables and clicks in the community chat forum. Parental characteristics (age and birth country), information seeking and concern, and area-level socioeconomic status were not associated with any of the considered indicators of engagement.

The sensitivity analyses yielded similar results, with a few exceptions. The associations between education level and total time spent in the app were significant after adjustment for app downtime (β=6.61, 95% CI 0.21-13.00; *P*=.04) and days in lockdown (β=6.75, 95% CI 0.35-13.16; *P*=.04). In contrast, associations between work status and click-depth index (β=−0.01, 95% CI −0.02 to −0.00; *P*=.05); between siblings in the family and the proportion of features visited (β=−2.90, 95% CI −6.00 to 0.24; *P*=.07); and between remoteness (major city vs regional or remote area) and click-depth index (β=−0.32, 95% CI −0.73 to 0.10; *P*=.14) were attenuated after adjustment for app downtime and days in lockdown.

XSL•FO
RenderX

Table 3. Associations between family demographics and parental characteristics with the individual engagement indicators for overall use (n=609)[a].

Predictor variables	Total time spent in the app (min)	App use (d)	Proportion of pages visited (%)	Proportion of features visited (%)
	β[b] (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Parent age (y)	0.45 (−0.19 to 1.09)	0.10 (−0.16 to 0.37)	0.13 (−0.10 to 0.36)	0.12 (−0.21 to 0.45)
Australian-born	−0.91 (−7.71 to 5.90)	0.22 (−2.58 to 3.03)	−0.14 (−2.61 to 2.32)	0.64 (−2.84 to 4.13)
University degree[c]	6.30 (−0.11 to 12.71)	*3.14 (0.50 to 5.78)* [d]	*3.27 (0.96 to 5.59)*	*4.33 (1.05 to 7.61)*
Work status (h/wk)	*−0.24 (−0.43 to −0.05)*	*−0.04 (−0.12 to 0.04)*	*−0.08 (−0.15 to −0.01)*	−0.08 (−0.18 to 0.02)
Parents living together	*−13.58 (−24.39 to −2.77)*	*−5.07 (−9.53 to −0.61)*	*−4.30 (−8.22 to −0.38)*	−4.51 (−10.06 to 1.05)
Siblings in the family	*−7.11 (−13.4 to −0.82)*	*−3.04 (−5.59 to −0.49)*	*−2.66 (−4.92 to −0.40)*	*−3.22 (−6.38 to −0.07)*
Living in a regional area[e]	−1.62 (−7.97 to 4.72)	−0.95 (−3.57 to 1.66)	−1.00 (−3.29 to 1.30)	−1.49 (−4.75 to 1.76)
SEIFA[f] score	0.01 (−0.04 to 0.05)	0.01 (−0.01 to 0.02)	0.01 (−0.01 to 0.02)	0.01 (−0.01 to 0.03)
Parent coping	1.70 (−2.23 to 5.62)	0.99 (−0.62 to 2.61)	0.32 (−1.11 to 1.74)	0.56 (−1.45 to 2.58)
Concern with child MB[g]	−0.53 (−1.61 to 0.55)	−0.20 (−0.65 to 0.24)	−0.20 (−0.59 to 0.19)	−0.28 (−0.01 to 0.30)
Information seeking	0.33 (−0.31 to 0.98)	−0.04 (−0.31 to 0.23)	−0.03 (−0.21 to 0.26)	−0.06 (−0.39 to 0.27)

[a]Due to missing data, n varied slightly for some models, ranging from 1 to 62 missing observations.

[b]β: unstandardized regression coefficient.

[c]University degree or higher education versus no university degree.

[d]Italics indicate statistical significance.

[e]Major cities versus inner or outer regional and remote Australia.

[f]SEIFA: Socio-Economic Indexes for Areas.

[g]MB: movement behaviors.

XSL•FO
RenderX

Table 4. Associations between family demographics and parental characteristics with the individual indicators for engagement in the main parts of the app (ie, clicks in the modules, toolkit, and community chat forum; n=609)[a].

Predictor variables	Learning module clicks (n)	Toolkit clicks (n)	Community chat forum clicks (n)
	β[b] (95% CI)	β (95% CI)	β (95% CI)
Parent age (y)	0.71 (−0.62 to 2.04)	0.64 (−0.08 to 1.35)	0.00 (−0.08 to 0.09)
Australian-born	−3.22 (−17.4 to 10.9)	−0.40 (−8.00 to 7.21)	−0.01 (−0.92 to 0.90)
University degree[c]	*18.59* [d] *(5.32 to 31.86)*	5.34 (−1.83 to 12.51)	0.74 (−0.12 to 1.60)
Work status (h/wk)	*−0.40 (−0.79 to −0.01)*	*−0.27 (−0.48 to −0.06)*	−0.02 (−0.04 to 0.01)
Parents living together	*−31.37 (−53.79 to −9.00)*	*−12.72 (−24.82 to −0.63)*	−0.50 (−1.95 to 0.95)
Siblings in the family	−10.92 (−23.76 to 1.92)	*−7.75 (−14.82 to −0.68)*	−0.58 (−1.41 to 0.24)
Living in a regional area[e]	−6.37 (−19.55 to 6.81)	−3.51 (−10.60 to 3.58)	−0.13 (−0.98 to 0.72)
SEIFA[f] score	0.03 (−0.05 to 0.12)	0.01 (−0.04 to 0.05)	0.00 (−0.00 to 0.01)
Parent coping	5.89 (−2.26 to 14.03)	−2 (−6.39 to 2.39)	−0.10 (−0.63 to 0.42)
Concern with child MB[g]	−1.07 (−3.31 to 1.17)	−0.63 (−1.84 to 0.58)	0.01 (−0.14 to 0.15)
Information seeking	0.46 (−0.88 to 1.80)	0.04 (−0.65 to 0.73)	0.01 (−0.07 to 0.10)

[a]Due to missing data, n varied slightly for some models, ranging from 1 to 62 missing observations.

[b]β: unstandardized regression coefficient.

[c]University degree or higher education versus no university degree.

[d]Italics indicate statistical significance.

[e]Major cities versus inner or outer regional and remote Australia.

[f]SEIFA: Socioeconomic Indexes for Areas.

[g]MB: movement behaviors.

XSL•FO
RenderX

Table 5. Associations of family demographics and parental characteristics with the composite indicators of engagement (n=609)[a].

Predictor variables	Total score	Click depth	Loyalty	Recency	Diversity
	β[b] (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Parent age (y)	0.07 (−0.05 to 0.18)	0.02 (−0.02 to 0.06)	0.04 (−0.10 to 0.17)	0.11 (−0.16 to 0.38)	0.10 (−0.07 to 0.27)
Australian-born	−0.99 (−2.21 to 0.23)	−0.32 (−0.77 to 0.12)	−0.99 (−2.43 to 0.44)	−1.87 (−4.62 to 0.88)	−0.78 (−2.56 to 1)
University degree[c]	*2.61 (1.46 to 3.77)* [d]	*0.77 (0.36 to 1.19)*	*2.48 (1.14 to 3.83)*	*4.64 (2 to 7.27)*	*2.56 (0.89-4.24)*
Work status (h/wk)	−0.02 (−0.06 to 0.01)	*-0.01 (-0.03 to -0.00)*	−0.03 (−0.07 to 0.02)	−0.04 (−0.12 to 0.04)	−0.02 (−0.07 to 0.03)
Parents living together	−0.29 (−2.27 to 1.68)	−0.44 (−1.15 to 0.27)	−2.15 (−4.43 to 0.15)	1.55 (−2.94 to 6.04)	−0.14 (−2.99 to 2.71)
Siblings in the family	*−1.56 (−2.65 to −0.48)*	*−0.43 (−0.83 to −0.04)*	*−2.40 (−3.68 to −1.12)*	*−2.77 (−5.17 to −0.36)*	−0.65 (−2.24 to 0.93)
Living in a regional area[e]	−0.20 (−1.36 to 0.95)	*−0.47 (−0.88 to −0.05)*	0.07 (−1.27 to 1.41)	−0.14 (−2.77 to 2.48)	−0.27 (−1.93 to 1.40)
SEIFA[f] score	−0.00 (−0.01 to 0.01)	0.00 (−0.00 to 0.00)	−0.00 (−0.01 to 0.01)	0.00 (−0.01 to 0.02)	−0.00 (−0.02 to 0.01)
Parent coping	0.56 (−0.15 to 1.27)	0.13 (−0.13 to 0.39)	*1.25 (0.43-2.08)*	0.62 (−1.01 to 2.24)	0.24 (−0.79 to 1.27)
Concern with child MB[g]	0.03 (−0.17 to 0.23)	0.00 (−0.07 to 0.08)	−0.07 (−0.30 to 0.16)	−0.06 (−0.50 to 0.39)	0.25 (−0.04 to 0.53)
Information seeking	0.04 (−0.08 to 0.16)	0.03 (−0.02 to 0.07)	−0.04 (−0.17 to 0.10)	0.07 (−0.20 to 0.34)	0.09 (−0.08 to 0.26)

[a]Due to missing data, n varied slightly for some models, ranging from 1 to 62 missing observations.

[b]β: unstandardized regression coefficient.

[c]University degree or higher education versus no university degree.

[d]Italics indicates statistical significance.

[e]Major cities versus inner or outer regional and remote Australia.

[f]SEIFA: Socioeconomic Indexes for Areas.

[g]MB: movement behaviors.

Discussion

Principal Findings

To the best of our knowledge, this is the first mHealth intervention study targeting parents of young children to examine user engagement and predictors of engagement beyond simple demographics. In addition to including family demographics, our study also considered parental cognitions and behaviors relevant to the intervention, which are crucial for interventions targeting parents as agents of change for child behavior. Thus, our findings contribute to an increased understanding of which factors influence parental engagement in digital interventions and who may need more active strategies to support their engagement in digital interventions. The use of the Let's Grow app was initially high, with declining levels; however, among those who continued to use the app across the 25-week period, the level of engagement remained consistent across this timeframe. We identified several demographic predictors of engagement. More work hours, parents living together, having siblings in the family, and living in a regional or remote area were associated with lower engagement, while having a higher education level was associated with higher engagement. Of the parental characteristics examined, higher parent coping was positively associated with engagement, while parent concern and information seeking in relation to child movement did not appear to impact engagement.

Comparison With Previous Findings

User engagement in digital interventions has gained more interest in recent years, and the link between intervention uptake and efficacy has been established [25]. However, the conceptualization and reporting of engagement in digital interventions targeting parents and other populations (eg, Saleem et al [26]) remain scarce, and there are no universal criteria for defining *good* engagement. Nevertheless, engagement has been the focus of commercial apps and website companies for a long time. A report on engagement in all available apps in the Apple and Google Play stores suggests that 25% of all downloaded apps are used only once in the first 6 months [27], and an engagement rate between 1% and 5% has been described as acceptable app engagement [28,29]. In this context, initial engagement in our study can be described as high, which is also similar to what has been observed in previous studies investigating engagement in digital health interventions (eg, [29,30]). For instance, Røed et al [13] reported that 13.5% (20/148) of the enrollees in their study did not enter the website at all over the 6-month intervention period, similar to 10.7% (73/682) in our study. Moreover, our results showed that, on average, users accessed the app for more than 18 days (equivalent to approximately once a week), which is similar to the findings of the MINISTOP trial [31] in which parents of children aged 2 to 3 years used the intervention app to register child health behaviors on average once per week during the 6-month intervention period. Furthermore, overall engagement

XSL•FO
RenderX

with our app, as assessed by the mean composite EI (4 subindexes), was moderate and almost identical to that reported in 2 other studies that investigated engagement in the Growing Healthy app among expectant parents and parents with an infant [11] and the Milk Man app among fathers with an infant [14]. Both the Growing Healthy and the Milk Man studies also used an adapted version of the Web Analytics Demystified visitor index (5 subindexes) used in this study and reported a composite mean EI of 30% (SD 12%; range 2%-58%) and 30% (SD 20%; range 1%-80%), respectively. Similarly, a study involving parents (n=42) of children aged 2 to 5 years reported a slightly higher index score (median 56, SD 21; range 4-81; scale 0-100) with an mHealth parent-training program (ezPARENT tablet app); however, they used a different index based on only 3 engagement measures (ie, the completion of app modules, number of visits, and time between visits) [12].

Although efforts to conceptualize engagement with digital health interventions have been made [32-34], to date, there is no standard measurement approach for engagement in mHealth interventions. This makes it difficult to synthesize the evidence and compare engagement across studies. Considering the differences in intervention design and thus purposive tailoring and calculation of these EIs, direct comparison between studies is impractical. Nevertheless, quantifying engagement should still be a priority, and for that purpose, it can be argued that a composite EI can serve as a study-specific indicator of user engagement and may be a valuable tool for statistical analysis purposes (eg, investigating intervention effects). However, careful deliberation is required to include the most relevant indicators for that intervention. Moreover, individual indicators of engagement, such as total time spent using the app and frequency of engagement, may be better for direct comparison between apps, as they are indicators that all apps can extract and are likely more feasible for researchers to use. A further complication in assessing engagement is that some apps, such as the Let's Grow app, act as tools to promote offline engagement in behavior change, which is not measurable through app use metrics. For example, most of the activities in the behavior change component of the app were designed to prompt parents to engage their children in active play opportunities offline. Engagement in these offline components of the intervention, which form a key aspect of intervention dose and are likely to impact intervention efficacy, is not captured in app engagement statistics. Thus, future studies should aim to develop and assess measures beyond objective engagement metrics to enable a complete assessment of engagement in mHealth interventions that promote offline activities.

Another important aspect of participant engagement is use over time, and retaining user's engagement in digital health interventions has been reported to be challenging [35,36]. Our pattern analyses indicate that use was not linear. Rather, we found an initial decline during the first few weeks, followed by steady levels of engagement among the subgroup who were still using the intervention at 12 and 25 weeks. In contrast, Taki et al [11] showed a downward trend in engagement subindexes score from 12 to 25 weeks, which aligns with studies in other populations in which continual linear decline in use over time has been reported [37,38]. One possible reason for the difference between our findings and the consistent declines reported for the Growing Healthy app [11] and other mHealth interventions (eg, the HealthyMoms app [39]) may lie in a key design difference. To illustrate, in the Growing Healthy app, all content was freely available whenever the user wished to access it [40], and new intervention content was delivered regularly (ie, biweekly) in the HealthyMoms app regardless of participant use [39]. In contrast, Let's Grow participants had to work through modules to unlock the next one. This may have encouraged more regular, ongoing use among those highly engaged parents with continued use across the study and could be viewed as a form of gamification (users were encouraged to keep coming back to unlock new content), which has been described as somewhat successful in achieving maintained parental engagement in child health promotion apps [36]. However, it is also possible that not having all content freely available or the need to undertake behavior change activities (rather than only viewing content) in Let's Grow may have been a barrier to engagement for some participants. A better understanding of engagement with these and other app designs, such as individually tailored apps (eg, using a just-in-time adaptive intervention design), is needed. Moreover, formal testing is required to understand what format best optimizes engagement, as well as to develop implementation strategies that support and sustain user engagement.

Furthermore, a better understanding of the participant characteristics that impact engagement in parental interventions can assist in clarifying how to design interventions to maximize behavior change. Nevertheless, a few previous studies [11-14] have reported predictors of engagement in parental mHealth interventions, and they have mostly focused on parent and family demographic characteristics. In our study, we expanded this to include 3 measures of parental characteristics, including parent coping in general as well as concern and information seeking specific to the targeted topic (child movement behaviors). We identified several predictors related to time and socioeconomic pressures (education level, work hours, remoteness, and family structure), as well as parent coping in general (not specific to the targeted intervention topic). In more detail, socioeconomic pressures, such as lower education level, working more hours, and living in a remote or outer regional area, as well as family structure (living with a partner or spouse and siblings in the family), were associated with *lower* engagement. This is consistent with previous studies investigating predictors of engagement in parental mHealth interventions, which have shown lower levels of engagement among parents with lower education levels [13] and those with more than 1 child [11,13]. Indeed, more work hours and having more than 1 child could be seen as indicators of less available time, which has previously been described as a barrier to app engagement [37]. Moreover, poorer parental coping in general was identified as a predictor of *lower* engagement. Similar to the lack of available time, this may reflect a lack of parent capacity to take on new learning when they are struggling to cope with their daily commitments. In contrast, perceived need, motivation, and utility have been described as factors that promote engagement in health apps for adults [35,41,42]. Similarly, a mixed methods assessment in a subsample from

XSL•FO
RenderX

the Growing Healthy study showed that participants defined as having low or moderate engagement felt they had sufficient knowledge and that the app did not provide further information [43]. In our study, parents who did not engage with the app (defined as nonusers) reported lower coping scores and less concern about movement behaviors compared to those defined as users. Unexpectedly and notably, parental concern about child movement behaviors and parent information seeking about topics related to the intervention content were also not associated with engagement in our study. This indicates that being concerned and coping with life may predict actually starting the intervention (ie, initial engagement), but beyond enrolling in the study, it seems that other factors drive how much parents engage. Altogether, this suggests that time constraints and other pressures may outweigh the content relevance in determining actual engagement with the intervention, and parents under time or other pressures might need more support to initiate and sustain engagement.

Strengths and Limitations

This study has several strengths, including the objective and accurate assessment of the outcome (ie, participant engagement) and the relatively large sample size. In addition, the use of several engagement metrics provides a more comprehensive understanding of how users interact with mHealth interventions [44-46], and greater insights into engagement can be achieved by aggregating data across different domains (ie, frequency, intensity, time, and type) [46]. Thus, another strength of the study is the reporting of both the individual measures of engagement (representing engagement in 3 key strategies for promoting health behaviors: behavior change, information provision, and social support) and the overall engagement (EI). The study also has some limitations to acknowledge. First, we used mid-intervention data; thus, our results do not capture engagement over the entire intervention period (ie, 12 months). However, the actual minimum time to complete the intervention program was 22 weeks, which is a similar intervention duration compared to previous studies (eg, Røed et al [13]). Second, the data on the potential predictors were self-reported, and some measures, for example, parental concern, were subjective; thus, results may be subject to reporting biases (ie, social desirability). However, we observed a wide variation in the sample in terms of outcomes and predictor variables. Similar to many previous studies (eg, Røed et al [13]), another limitation of the study is that most participants (98.2%) were mothers. Although most participants lived in major cities (71.8%), participants from inner regional (19.2%) and outer regional or remote Australia (8.9%) were also represented, as well as representation across different socioeconomic areas. In addition, a larger proportion of the participants in the study had a university degree or a higher level of education compared to women in the same age group (25-44 years) in the general population (72% vs 50%) [47]. Nevertheless, our study sample was similar to the general population in terms of birth country (77% vs 71% Australian-born) [47] and work status among women in this age group (65% vs 58%) [47]. Third, we did not measure enactment (ie, the participants' "real-world" use of intervention skills); thus, we do not have information on intervention engagement beyond use within the app, which was actually designed to encourage parents to enact changes offline. Future studies should aim to capture offline engagement as well, potentially through check-in questions during the intervention period. Fourth, a potential limitation of the study is that we only assessed app engagement data for those who used the app and not for all enrollees (ie, excluding the 73/682, 10.7% of the participants who never used the app), which might have exaggerated engagement. However, the differences in engagement were small, with only slightly lower scores on engagement metrics when all participants were included (eg, mean EI for the whole group was 25.2, SD 10.5 vs 28.0, SD 6.6 for the users only). Nevertheless, nonusers may be qualitatively different from those who have any engagement with the app and are likely to need different strategies to engage them in interventions; our results demonstrate patterns of engagement for those who initiated the use of the app, which is important for informing future app development. Finally, the Let's Grow app had a nonlinear, dynamic intervention design, where parents could choose the order and timing of their progress, and not all content was freely available at once. This is different from most previously reported studies; thus, some engagement measures may not readily apply. On the same note, engagement scores need to be tailored for individual interventions, and consequently, results are not directly comparable across studies. Acknowledging the tailoring required to specific mHealth interventions, certain metrics, such as the EI used in this study, may provide some level of generalizability across interventions. Therefore, it is ideal for studies to include generalizable and tailored metrics to enable both nuanced assessment of the intervention under investigation and comparison across interventions. However, in cases where such metrics are not available, reporting individual metrics alone should still be a priority.

Implications and Clinical Relevance

Although the use of mHealth to deliver parental interventions is an established field, and the importance of engagement in interventions has been recognized, engagement in parental mHealth interventions is rarely reported. Thus, our study adds to the scarce literature in several aspects. First, to the best of our knowledge, there is no best practice for assessing engagement in digital interventions, and considering the heterogeneity of intervention designs and delivery, a universal approach may not be feasible, but the inclusion of some generalizable indicators is desirable. Moreover, although the digital format offers a great opportunity to objectively assess engagement, it also represents challenges in deciding which metrics are of importance. In this context, our study reports 2 different approaches to assess participant engagement using both individual metrics and an overall EI and highlights the importance of investigating not only total engagement but also looking at use patterns over time. This information will be useful in investigating potential intervention effects (eg, dose-response relationship), guiding future refinement of the Let's Grow app and development of similar mHealth interventions, as well as informing the development of guidelines for the assessment and evaluation of user engagement in parental mHealth interventions. Second, reporting of participant engagement and predictors of engagement is important for informing the design

XSL•FO
RenderX

and tailoring of future interventions to enhance their use. Moreover, understanding who engaged with the intervention and their patterns of engagement can provide a better understanding of the effect or lack of effect on the targeted outcomes. Third, engagement in digital interventions is often compared to engagement in more traditional face-to-face interventions, which are inherently not comparable due to the differences in intervention delivery and reliance on human interaction. A common assumption is that the digital format (eg, apps) improves reach and makes interventions more accessible for parents, which theoretically should lead to greater engagement. However, as demonstrated in this study, the improved reach may come at the expense of intervention engagement. Although our results show relatively high engagement in comparison to industry reports of app use, they are lower than attendance rates in group-based face-to-face programs, which typically range from 35 to 50% [27,48]. It is interesting to speculate that the lack of reporting of app engagement in mHealth interventions to date could be due to a perception of low engagement relative to what is typically considered *good* engagement through more traditional forms of intervention delivery. It is important for this field of research to report engagement in mHealth interventions so we can accurately gauge any compromise between reach and engagement in this type of intervention delivery. Altogether, our study provides valuable evidence in the field of intervention engagement, which can be of use for the evaluation and design of future parental mHealth interventions.

Conclusions

This study provides novel data on engagement and predictors of engagement in a parental digital intervention. While engagement data were comparable to other published data, a key difference was the relatively consistent level of engagement noted among individuals showing continued use across the period under investigation, rather than the typically observed continued decline. Our findings show that both family demographics and parental characteristics can influence engagement, and time and socioeconomic pressures might outweigh content relevance when it comes to engagement, although content relevance appears important to initial uptake. Our results highlight the characteristics of parents who may benefit from more active engagement strategies to support their use of digital interventions, specifically those with time and capacity stressors. Finally, our results showed similar outcomes for both the individual and composite measures of engagement, indicating that simpler engagement metrics, which might be more feasible and attainable for researchers to collect, process, and analyze, can be useful in the reporting of engagement in digital interventions. Altogether, our findings will be useful when investigating intervention effects (dose-response relationship) and might be important to consider for future refinement of the Let's Grow app and the development of similar interventions.

Acknowledgments

The authors would like to gratefully thank all the participating families in the Let's Grow trial, as well as the members of the Let's Grow team, for their invaluable work on recruitment and data collection. The Let's Grow trial was supported by the National Health and Medical Research Council (GNT 1162980). In addition, KDH is supported by a Heart Foundation Future Leader Fellowship (105929). JS is supported by 2 postdoctoral fellowships from the Henning and Johan Throne-Holst Foundation and the Erik and Edith Fernström Foundation for Medical Research.

Authors' Contributions

KDH is the principal investigator of the Let's Grow trial. KDH, KLD, and JMN contributed to the development of the intervention content. BRM analyzed the app analytics to produce an engagement index and contributed to the data analysis planning and interpretation together with J Sandborg, SS, LO, HK, VC, J Salmon, and KDH. J Sandborg analyzed the data and drafted the manuscript. All authors approved the final draft and the submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PowerPoint showing screenshots of the Let's Grow app.
[PDF File (Adobe PDF File), 1582 KB - mhealth_v13i1e60478_app1.pdf]

Multimedia Appendix 2
Description of the individual subindexes included in the engagement index.
[DOCX File , 17 KB - mhealth_v13i1e60478_app2.docx]

References

1. WHO Guidelines on Physical Activity, Sedentary Behaviour and Sleep for Children Under 5 Years of Age. Geneva, Switzerland: World Health Organization; 2018.

XSL•FO
RenderX

2. Xu H, Wen LM, Rissel C. Associations of parental influences with physical activity and screen time among young children: a systematic review. J Obes 2015;2015:546925 [FREE Full text] [doi: 10.1155/2015/546925] [Medline: 25874123]
3. Lee EY, Hesketh KD, Rhodes RE, Rinaldi CM, Spence JC, Carson V. Role of parental and environmental characteristics in toddlers' physical activity and screen time: Bayesian analysis of structural equation models. Int J Behav Nutr Phys Act 2018 Feb 09;15(1):17 [FREE Full text] [doi: 10.1186/s12966-018-0649-5] [Medline: 29426324]
4. Yavuz HM, van Ijzendoorn MH, Mesman J, van der Veek S. Interventions aimed at reducing obesity in early childhood: a meta-analysis of programs that involve parents. J Child Psychol Psychiatry 2015 Jun 08;56(6):677-692. [doi: 10.1111/jcpp.12330] [Medline: 25292319]
5. Ling J, Robbins LB, Wen F, Zhang N. Lifestyle interventions in preschool children: a meta-analysis of effectiveness. Am J Prev Med 2017 Jul;53(1):102-112. [doi: 10.1016/j.amepre.2017.01.018] [Medline: 28237633]
6. Ajay K, Azevedo LB, Haste A, Morris AJ, Giles E, Gopu BP, et al. App-based oral health promotion interventions on modifiable risk factors associated with early childhood caries: a systematic review. Front Oral Health 2023 Mar 10;4:1125070. [doi: 10.3389/froh.2023.1125070] [Medline: 36968137]
7. Nyström CD, Sandin S, Henriksson P, Henriksson H, Trolle-Lagerros Y, Larsson C, et al. Mobile-based intervention intended to stop obesity in preschool-aged children: the MINISTOP randomized controlled trial. Am J Clin Nutr 2017 Jun;105(6):1327-1335. [doi: 10.3945/ajcn.116.150995] [Medline: 28446496]
8. Bonvicini L, Pingani I, Venturelli F, Patrignani N, Bassi MC, Broccoli S, et al. Effectiveness of mobile health interventions targeting parents to prevent and treat childhood obesity: systematic review. Prev Med Rep 2022 Oct;29:101940 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101940] [Medline: 36161123]
9. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
10. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
11. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing user engagement of an mHealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
12. Breitenstein SM, Brager J, Ocampo EV, Fogg L. Engagement and adherence with ezPARENT, an mHealth parent-training program promoting child well-being. Child Maltreat 2017 Nov 05;22(4):295-304. [doi: 10.1177/1077559517725402] [Medline: 28870112]
13. Røed M, Vik FN, Hillesund ER, Van Lippevelde W, Medin AC, Øverby NC. Process evaluation of an eHealth intervention (Food4toddlers) to improve toddlers' diet: randomized controlled trial. JMIR Hum Factors 2020 Jul 06;7(3):e18171 [FREE Full text] [doi: 10.2196/18171] [Medline: 32628612]
14. White BK, Burns SK, Giglia RC, Dhaliwal SS, Scott JA. Measuring user engagement with a socially connected, gamified health promotion mobile app. Int J Environ Res Public Health 2022 May 05;19(9):5626 [FREE Full text] [doi: 10.3390/ijerph19095626] [Medline: 35565015]
15. Laws RA, Cheng H, Rossiter C, Kuswara K, Markides BR, Size D, et al. Perinatal support for breastfeeding using mHealth: a mixed methods feasibility study of the My Baby Now app. Matern Child Nutr 2023 Apr;19(2):e13482 [FREE Full text] [doi: 10.1111/mcn.13482] [Medline: 36725007]
16. Gunnarsdottir T, Njardvik U, Olafsdottir AS, Craighead LW, Bjarnason R. The role of parental motivation in family-based treatment for childhood obesity. Obesity (Silver Spring) 2011 Aug 31;19(8):1654-1662 [FREE Full text] [doi: 10.1038/oby.2011.59] [Medline: 21455125]
17. Hesketh KD, Downing KL, Galland BC, Nicholson JM, Taylor R, Orellana L, et al. Protocol for the Let's Grow randomised controlled trial: examining efficacy, cost-effectiveness and scalability of a m-Health intervention for movement behaviours in toddlers. BMJ Open 2022 Mar 28;12(3):e057521 [FREE Full text] [doi: 10.1136/bmjopen-2021-057521] [Medline: 35351726]
18. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
19. Prior M, Sanson A, Smart D, Oberklaid F. Pathways from infancy to adolescence: Australian temperament project 1983–2000. Australian Institute of Family Studies. 2001. URL: https://www.theactgroup.com.au/documents/AustralianTemperamentProject1983-2000.pdf [accessed 2025-06-20]
20. 1270.0.55.005 - Australian Statistical Geography Standard (ASGS): volume 5 - remoteness structure, July 2016. Australian Bureau of Statistics. 2018. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.005 [accessed 2023-09-25]
21. 2039.0 - Information paper: an introduction to socio-economic indexes for areas (SEIFA), 2006. Australian Bureau of Statistics. 2008. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/2039.0 [accessed 2023-09-25]

XSL•FO
RenderX

22. Laws R, Love P, Hesketh KD, Koorts H, Denney-Wilson E, Moodie M, et al. Protocol for an effectiveness-implementation hybrid trial to evaluate scale up of an evidence-based intervention addressing lifestyle behaviours from the start of life: INFANT. Front Endocrinol (Lausanne) 2021 Nov 8;12:717468 [FREE Full text] [doi: 10.3389/fendo.2021.717468] [Medline: 34975744]
23. Haines J, Downing KL, Tang L, Campbell KJ, Hesketh KD. Associations between maternal concern about child's weight and related behaviours and maternal weight-related parenting practices: a cross-sectional study. Int J Behav Nutr Phys Act 2018 Oct 24;15(1):104 [FREE Full text] [doi: 10.1186/s12966-018-0738-5] [Medline: 30355337]
24. Peterson ET, Carrabis J. Measuring the immeasurable: visitor engagement. Web Analytics Demystified. 2008 Sep 7. URL: https://www.verticalstudio.com/hs-fs/hub/74398/file-15425996-pdf/docs/web_analytics_demystified_and_nextstage_global_-_measuring_the_immeasurable_-_visitor_engagement.pdf [accessed 2024-01-16]
25. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
26. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]
27. Percentage of mobile apps that have been used only once from 2010 to 2019. Statista. 2019. URL: https://www.statista.com/statistics/271628/percentage-of-apps-used-once-in-the-us/ [accessed 2025-06-20]
28. App engagement. Storyly. 2024 Jun 4. URL: https://www.storyly.io/glossary/app-engagement [accessed 2025-06-20]
29. Røed M, Medin AC, Vik FN, Hillesund ER, Van Lippevelde W, Campbell K, et al. Effect of a parent-focused eHealth intervention on children's fruit, vegetable, and discretionary food intake (Food4toddlers): randomized controlled trial. J Med Internet Res 2021 Feb 16;23(2):e18311 [FREE Full text] [doi: 10.2196/18311] [Medline: 33591279]
30. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
31. Alexandrou C, Henriksson H, Henström M, Henriksson P, Delisle Nyström C, Bendtsen M, et al. Effectiveness of a smartphone app (MINISTOP 2.0) integrated in primary child health care to promote healthy diet and physical activity behaviors and prevent obesity in preschool-aged children: randomized controlled trial. Int J Behav Nutr Phys Act 2023 Feb 21;20(1):22 [FREE Full text] [doi: 10.1186/s12966-023-01405-5] [Medline: 36810069]
32. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020 May 27;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
33. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
34. Milne-Ives M, Homer S, Andrade J, Meinert E. The conceptualisation and measurement of engagement in digital health. Internet Interv 2024 Jun;36:100735 [FREE Full text] [doi: 10.1016/j.invent.2024.100735] [Medline: 38558760]
35. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
36. Blakeslee SB, Vieler K, Horak I, Stritter W, Seifert G. Planting seeds for the future: scoping review of child health promotion apps for parents. JMIR Mhealth Uhealth 2023 Jul 20;11:e39929 [FREE Full text] [doi: 10.2196/39929] [Medline: 37471125]
37. Sandborg J, Henriksson P, Larsen E, Lindqvist AK, Rutberg S, Söderström E, et al. Participants' engagement and satisfaction with a smartphone app intended to support healthy weight gain, diet, and physical activity during pregnancy: qualitative study within the HealthyMoms trial. JMIR Mhealth Uhealth 2021 Mar 05;9(3):e26159 [FREE Full text] [doi: 10.2196/26159] [Medline: 33666554]
38. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
39. Henriksson P, Migueles JH, Söderström E, Sandborg J, Maddison R, Löf M. User engagement in relation to effectiveness of a digital lifestyle intervention (the HealthyMoms app) in pregnancy. Sci Rep 2022 Aug 13;12(1):13793 [FREE Full text] [doi: 10.1038/s41598-022-17554-9] [Medline: 35963935]
40. Denney-Wilson E, Laws R, Russell CG, Ong KL, Taki S, Elliot R, et al. Preventing obesity in infants: the growing healthy feasibility trial protocol. BMJ Open 2015 Nov 30;5(11):e009258 [FREE Full text] [doi: 10.1136/bmjopen-2015-009258] [Medline: 26621519]
41. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]

XSL•FO
RenderX

42. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
43. Taki S, Russell CG, Lymer S, Laws R, Campbell K, Appleton J, et al. A mixed methods study to explore the effects of program design elements and participant characteristics on parents' engagement with an mHealth program to promote healthy infant feeding: the growing healthy program. Front Endocrinol (Lausanne) 2019 Jun 25;10:397 [FREE Full text] [doi: 10.3389/fendo.2019.00397] [Medline: 31293515]
44. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
45. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
46. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
47. Australia's population by country of birth. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/people/population/australias-population-country-birth/latest-release [accessed 2023-08-22]
48. Heinrichs N, Bertram H, Kuschel A, Hahlweg K. Parent recruitment and retention in a universal prevention program for child behavior and emotional problems: barriers to research and program participation. Prev Sci 2005 Dec;6(4):275-286. [doi: 10.1007/s11121-005-0006-1] [Medline: 16075192]

Abbreviations

EI: engagement index
mHealth: mobile health
REDCap: Research Electronic Data Capture
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 13.05.24; peer-reviewed by S Breitenstein, M Klapow; comments to author 08.10.24; revised version received 28.11.24; accepted 21.05.25; published 15.07.25.

Please cite as:
Sandborg J, Markides BR, Simmons S, Downing KL, Nicholson JM, Orellana L, Koorts H, Carson V, Salmon J, Hesketh KD
Parental and Demographic Predictors of Engagement in an mHealth Intervention: Observational Study From the Let's Grow Trial
JMIR Mhealth Uhealth 2025;13:e60478
URL: https://mhealth.jmir.org/2025/1/e60478
doi:10.2196/60478
PMID:

©Johanna Sandborg, Brittany Reese Markides, Savannah Simmons, Katherine L Downing, Jan M Nicholson, Liliana Orellana, Harriet Koorts, Valerie Carson, Jo Salmon, Kylie D Hesketh. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.07.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Motivation as a Predictor of mHealth Engagement Across BMI: Cross-Sectional Survey

Shao-Hsuan Chang[1,2*], PhD; Lung-Kun Yeh[2,3*], MD, PhD; Daishi Chen[4], MD, PhD; Kae-Kuen Hu[5], PhD; Mei-Ching Yu[6], MD, PhD; Ching-Mao Chang[7,8], MD, PhD

[1]Department of Biomedical Engineering, Chang Gung University, Taoyuan, Taiwan

[2]Department of Ophthalmology, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

[3]College of Medicine, Chang Gung University, Taoyuan, Taiwan

[4]Department of Otolaryngology, Shenzhen People's Hospital, Shenzhen, China

[5]Professional Master's Program of Biotechnology Management, School of Professional Education and Continuing Studies, National Taiwan University, Taipei, Taiwan

[6]Department of Pediatric Nephrology, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

[7]Center for Traditional Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

[8]School of Chinese Medicine and Institute of Traditional Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Ching-Mao Chang, MD, PhD
Center for Traditional Medicine
Taipei Veterans General Hospital
No 201, Sector 2, Shih-Pai Road, Beitou District
Taipei, 112201
Taiwan
Phone: 886 2 2875 7453 ext 333
Fax: 886 2 2875 7452
Email: magicbjp@gmail.com

Abstract

Background: Digital health tools, such as mobile apps and wearable devices, have been widely adopted to support self-management of health behaviors. However, user engagement remains inconsistent, particularly among populations with varying BMI. While digital health technologies have the potential to promote healthier behaviors, little is known about how psychological and behavioral factors interact with BMI to influence use patterns.

Objective: This study aimed to explore the relationship between BMI and digital health technology use and to examine how factors such as health awareness, self-efficacy, and health motivation contribute to technology engagement.

Methods: A cross-sectional online survey was conducted from January 2024 to April 2024. A total of 184 valid questionnaire participants were included in this study. The questionnaire was measured on a 5-point Likert scale. Descriptive statistics, chi-square tests, and multiple regression analyses were applied.

Results: Of the participants, 38.6% (71/184) had a BMI<24 kg/m^2, 42.4% (78/184) had a BMI between 24 and 29.9 kg/m^2, and 19% (35/184) had a BMI≥30 kg/m^2. Significant BMI differences were observed based on sex (P<.001) and age (P<.001) but not based on prior digital health tool use. Use rates for Bluetooth or Wi-Fi devices, wearables, and mobile apps were 32.1% (59/184), 38.6% (71/184), and 39.1% (72/184), respectively. A negative correlation between BMI and mobile app use frequency was identified (P=.02). Multiple regression analysis indicated that health motivation significantly predicted digital health use (P<.001), whereas health awareness, lifestyle, and self-efficacy did not.

Conclusions: Individuals with higher BMI reported a lower frequency of digital health tool use, potentially due to lower health motivation in the studied population. Health motivation was the strongest predictor of digital health engagement. Integrating personalized medical records into apps may enhance health motivation, thereby improving user engagement and promoting healthier behaviors in individuals with higher BMI.

(JMIR Mhealth Uhealth 2025;13:e71625) doi:10.2196/71625

XSL•FO
RenderX

KEYWORDS

digital health; BMI; health motivation; mobile health; self-efficacy; user engagement; Mandarin-speaking population

Introduction

The digital era has brought an explosion of health-related data from clinical databases, social networks, wearables, and connected medical devices. Among these digital health innovations, mobile health (mHealth) apps and wearable devices stand out, with an expected 3.7 billion downloads globally by consumers and health care professionals [1]. The increasing adoption of digital health is driven by patient convenience, technological advancements, and persistent health care concerns. The transition from traditional to digital and information-based medicine has contributed to the rise of personalized health care, enabling a broader population to engage in health management. This includes those who may not have a formal diagnosis but experience early symptoms or notice early warning signs, often referred to as subhealth status [2]. Sharing health information through digital devices has given rise to a new form of algorithmic surveillance. These technologies now play an increasing role in the medicalization of health. Researchers have proposed different frameworks to explain the medicalization of everyday life through digital technologies, such as quantification, where individuals track health data using devices, and gamification, where playful elements such as points or badges are incorporated to promote user motivation and adherence [3].

Digital health technologies come in various forms, including mobile platforms, apps, wearables, and Bluetooth- or Wi-Fi–connected devices. Prior research has examined their impact on health outcomes. A study by Vansimaeys et al [4] found that frequent users of multiple digital health felt more empowered and reported stronger patient-doctor relationships. Another randomized controlled trial by Ross and Wing [5] demonstrated that mHealth apps, combined with digital devices, resulted in greater weight loss compared to conventional methods. These findings support the positive role of digital health technologies in disease management [6]. However, despite their promise, many digital devices face challenges with user engagement and adherence [7]. Grounding these tools in behavioral science frameworks and enhancing social validity and user experience are essential [8]. Features such as goal setting, social support, and real-time monitoring may improve motivation and long-term use [9]. Behavioral frameworks such as the health belief model and the unified theory of acceptance and use of technology provide useful lenses to understand what drives users to adopt and sustain digital health practices [10,11].

BMI plays a critical role in disease prevention. Individuals classified as overweight (BMI 24-29.9 kg/m^2) or obese (BMI≥30 kg/m^2) are at increased risk of chronic conditions such as diabetes and cardiovascular diseases [12]. While many studies have examined how digital health technologies affect chronic illness management [13-16], few have explored their role in BMI regulation [16-19]. Furthermore, little is known about how individuals without diagnosed conditions, but with varying BMI, use digital tools for health management.

Therefore, understanding how people with different BMI categories engage with digital health technologies is essential for designing effective, tailored interventions. This study aimed to address this gap by (1) investigating the proportion of participants using digital health tools across BMI groups, (2) identifying the types of digital technologies used for health management, and (3) determining the key predictors associated with digital health use. The study analyzed the association between participant characteristics and BMI categories using chi-square tests and regression models while also discussing how usability and design features influence engagement within a medical self-monitoring framework.

Methods

Ethical Considerations

This study was approved by the institutional review board committee of National Taiwan University Hospital (registration: 202307210RINC). The study was conducted from January 1, 2024, to April 30, 2024. To protect respondents' privacy, all survey responses were anonymous, and IP addresses were hidden. This online anonymous survey was determined to fall under the exempt review criteria. As such, no written informed consent was required. Participation was voluntary, and respondents were informed about the purpose of the study at the beginning of the survey. The collected data were used solely for research purposes. Participants were not compensated for their participation in this study.

Design of the Study

An online anonymous survey was conducted to investigate how people perceive and are aware of their health condition or status, as well as their use of digital devices or mHealth tools (digital health use) for health management. The questionnaire was presented in a fixed sequence determined by logical flow and skip logic, tailored to participants' responses. Randomization of item order was not applicable due to the structure of conditional branching and mandatory completion of relevant questions. All questions were mandatory unless deemed nonapplicable through skip logic. Respondents were not allowed to skip questions or revise answers once submitted. Trap questions were embedded to identify inattentive or invalid responses. The first module of the questionnaire asked respondents to report their health status or health problems. The second module asked respondents about their health-promoting lifestyle. The third module asked respondents about their self-management behaviors and efficacy in health management. The fourth module focused on the digital health use for monitoring of their health conditions. Data for this study were collected from the survey about the respondents' demographic characteristics, health perception or awareness, health promotion lifestyle, health motivation, self-efficacy, digital health use, and adherence to mHealth tools.

XSL•FO
RenderX

Procedure and Participants

The questionnaire designed for the study was hosted on SurveyCake. This platform ensures data integrity, data security, and privacy by using encryption protocols such as Secure Socket Layer to protect data transmission. All collected data were stored on secure servers and are accessible only by authorized personnel. SurveyCake also complies with data protection regulations, ensuring that all personal information is handled in accordance with relevant privacy laws. Participants' responses were anonymized, and IP addresses were not collected to ensure confidentiality. The link to access the questionnaire was initially disseminated through social networks and emails. Participants were recruited via convenience and snowball sampling methods. The actual scope of dissemination through social media and email forwarding may have reached a broader audience beyond the researcher's immediate network. The online questionnaire was available on the internet from October 2023 to December 2023 for the pilot run. The target population included Mandarin-speaking adults aged 20 years and older, residing in urban areas. The sample was drawn from a local community where preventive health services are available, but their use may vary across different population groups. For this work, a total of 215 questionnaires were fully completed anonymously. Completion times ranged from 156 to 2292 seconds, with an average of approximately 6.5 (SD 4.2) minutes. Responses with fewer than 90 seconds or more than 1 hour were flagged for review and excluded if deemed invalid. Duplicate responses were not detected directly as IP addresses were not collected, and no cookies were used to prevent multiple entries. Response consistency was checked through logical review and trap questions. Invalid or inconsistent entries were excluded from the analysis. Only fully completed responses were used for the analysis. Therefore, the valid questionnaires were corresponding to 184 respondents (184/215×100%: 85.6%).

Measures and Variables

Given the online-based approach of the study, a closed-ended questionnaire was used, which included personal factors such as sociodemographic characteristics and health-related information. The questionnaire was developed by referencing validated and reliable instruments to ensure content validity. Expert opinions were also sought during the initial design phase to further refine the questionnaire. The overall design and content were reviewed and approved by the institutional review board prior to conducting the pilot study, which further refined the questionnaire. A pilot study with 35 participants from the target population was conducted to test the feasibility and clarity of the questionnaire, with subsequent modifications made based on feedback. During this process, items with low relevance were removed, and the remaining questions were adjusted to create a new questionnaire tailored to the study's objectives. This refinement enhanced the internal consistency of the questionnaire, resulting in a Cronbach α exceeding 0.80. The final version included items assessing 5 health-related variables (Table S1 in Multimedia Appendix 1): health awareness, health promotion lifestyle, self-efficacy, health motivation, and digital health use to evaluate respondents' use of digital health. Participants completed the survey anonymously via an online platform. The questionnaire had been verified by factor analysis and had good reliability (overall Cronbach α=0.90, 66 items; Table S2 in Multimedia Appendix 1). The items in this questionnaire were adapted and simplified from multiple validated instruments, including components of the health belief model and unified theory of acceptance and use of technology–based surveys [10,11].

We surveyed participants about their experience with digital devices, their preferences regarding use frequency, and the parameters they monitor related to their health status. We then examined the percentage distribution of various factors across BMI categories, followed by chi-square analysis. Additionally, the study explored whether respondents had developed a self-management plan, the types of mHealth tools they use, the factors influencing their decision-making, and their satisfaction with and the usability of specific devices or mHealth apps.

Demographic Characteristics

Demographic characteristics included age (20-29, 30-39, 40-49, 50-59, 60-69, and above 70 years), sex, BMI, educational level (high school or below, college degree, and postgraduate degree), and experience with digital technology.

Health-Related Variables

Health Awareness

Health awareness was measured by assessing respondents' understanding of the risks associated with obesity, diabetes, and other health conditions. The health awareness score reflects how well individuals recognize the health risks of various diseases and their awareness of the impact of lifestyle choices on overall health [20,21].

Health Promotion Lifestyle

A health-promoting lifestyle refers to a set of behaviors that support physical, emotional, and social well-being. In this study, it was assessed using an 11-item simplified scale adapted from the Health Promoting Lifestyle Profile-II [22], covering 6 dimensions: self-actualization, stress management, health responsibility, exercise, interpersonal support, and nutrition [22]. Respondents were asked to rate each statement on a 5-point Likert scale, representing the degree of agreement with each statement. A higher mean score indicates a lifestyle that aligns more closely with an ideal health-promoting lifestyle.

Self-Efficacy

Self-efficacy was assessed by measuring respondents' prior experiences with self-management and their perceived effectiveness in achieving health-related goals [23,24]. Participants rated their confidence in performing specific health management activities using a 5-point Likert scale, with higher scores indicating greater confidence in their ability to successfully manage their health.

Health Motivation

This section addresses the factors that influence an individual's decision to initiate and maintain health-related behaviors, with health motivation defined as the individual's willingness to invest effort and maintain it in order to achieve health-related goals [25,26]. Motivation includes intrinsic factors, such as the need for disease prevention, as well as extrinsic factors, such

XSL•FO
RenderX

as expert advice and social support. This section assesses an individual's perceived importance of various factors, such as physical health, mental health, disease prevention, and expert advice. Respondents rate the importance of these factors on a scale from 1 to 5, and the total score reflects their overall health motivation.

The Digital Health Use

This section evaluated the use of digital health tools, including connected devices, wearable devices, and mobile apps for health management [27-29]. It assessed the types of tools used, the frequency of use, the parameters measured (eg, blood pressure, heart rate, and blood oxygen), and the perceived health improvements associated with these tools [30,31]. For the analysis, this ordinal categorical variable was recoded into a discrete numerical scale, with level 1 representing "less than once per month" and level 5 representing "daily use." Finally, 3 frequency categories were defined: less than once per week, 1-3 times per week, and more than 3 times per week.

Statistical Analysis

We used SPSS (version 23.0; IBM Corp) and Python (Python Software Foundation) to conduct all statistical analyses. A *P* value <.05 was considered as statistically significant. The chi-square test or the independent samples 2-tailed *t* test was performed to examine the differences in demographic characteristics and compare each variable of the total population by BMI. The Pearson correlation was analyzed to determine the correlations between the use of digital technology and health-related variables. Moreover, we conducted multiple regression and logistic regression analyses to identify predictors for each health-related variable. Logistic regression was used to analyze the frequency of device use, while multiple regression was applied to analyze the overall variables related to digital health use.

Results

Comparison of Sociodemographic Characteristics Across Groups Based on BMI Classification

Table 1 presents the distribution of demographic characteristics, including sex, age, education level, and experience with digital technology, across 3 BMI categories: BMI<24 kg/m^2, 24≤BMI<29.9 kg/m^2, and BMI≥30 kg/m^2. Of the 184 respondents to the valid questionnaires recovered in this study, 71 participants had a BMI<24 kg/m^2, accounting for 38.6% of the sample; 78 (42.4%) participants had a BMI between 24 and 29.9 kg/m^2; and 35 (19%) participants had a BMI≥30 kg/m^2. Chi-square analysis revealed significant differences in sex (*P*<.001) and age (*P*=.02) across the different BMI categories. However, no significant differences were observed in education level or experience with digital technology.

XSL•FO
RenderX

Table 1. Demographic characteristics of the sample population by BMI.

Variables	Total (N=184, 100%), n (%)	BMI<24 kg/m^2 (n=71, 38.6%), n (%)	24≤BMI<29.9 kg/m^2 (n=78, 42.4%), n (%)	BMI≥30 kg/m^2 (n=35, 19%), n (%)	*P* value[a]
Sex					<.001[b]
Male	75 (40.8)	18 (25.4)	34 (43.6)	23 (65.7)	
Female	109 (59.2)	53 (74.6)	44 (56.4)	12 (34.3)	
Age (years)					.02[c]
20-29	61 (33.2)	35 (49.3)	19 (24.4)	7 (20)	
30-39	104 (56.5)	29 (40.9)	48 (61.5)	27 (77.1)	
40-49	10 (5.4)	4 (5.6)	5 (6.4)	1 (2.9)	
50-59	6 (3.3)	2 (2.8)	4 (5.1)	0 (0)	
60-69	3 (1.6)	1 (1.4)	2 (2.6)	0 (0)	
Education					.38
High school or below	14 (7.6)	6 (8.5)	4 (5.1)	4 (11.4)	
College degree	134 (72.8)	54 (76)	54 (69.2)	26 (74.3)	
Postgraduate degree	36 (19.6)	11 (15.5)	20 (25.7)	5 (14.3)	
Experience with digital technology					
Years of using smartphone					.41
>3	179 (97.3)	69 (97.1)	77 (98.7)	33 (94.3)	
1-3	5 (2.7)	2 (2.9)	1 (1.3)	2 (5.7)	
None	0 (0)	0 (0)	0 (0)	0 (0)	
Weekly frequency of using smartphone					.45
>3	183 (99.5)	70 (98.6)	78 (100)	35 (100)	
1-3	0 (0)	0 (0)	0 (0)	0 (0)	
<1	1 (0.5)	1 (1.4)	0 (0)	0 (0)	
Years of using Bluetooth- or Wi-Fi– connected device					.88
>3	165 (89.7)	62 (87.3)	71 (91)	32 (91.4)	
1-3	10 (5.4)	5 (7)	3 (3.9)	2 (5.7)	
None	9 (4.9)	4 (5.7)	4 (5.1)	1 (2.9)	
Weekly frequency of using Bluetooth- or Wi-Fi– connected device					.15
>3	155 (84.2)	58 (81.7)	70 (89.7)	27 (77.1)	
1-3	14 (7.6)	4 (5.6)	5 (6.4)	5 (14.3)	
<1	15 (8.2)	9 (12.7)	3 (3.9)	3 (8.6)	

[a]Differences based on chi-square test.

[b]*P*<.01.

[c]*P*<.05.

The distribution of personal health information and health perception across different BMI categories (N=184) is shown in Table 2. There were no significant differences in the reported number of personal health problems or family medical history (*P*=.16) or in the types of health-related information accessed in the past 12 months (*P*=.11). Regarding physical activity, most respondents (128/184, 69.4%) engaged in more than 150 minutes per week, with no significant differences between BMI groups. However, chi-square analysis found that the annual frequency of personal medical checkup (*P*=.03) and the self-rating of personal health status (*P*=.02) significantly differed among BMI categories, indicating that participants' perceptions of their own health varied depending on their BMI. The significance suggested that BMI was associated with how individuals assess their own health. Specifically, as BMI increases, a higher proportion of participants rated their health as "fair to good" (17/35, 48.6% with scores 2-3) or "very poor" (3/35, 8.6% with score 1). Those with a BMI≥30 kg/m^2 tended to rate their health

XSL•FO
RenderX

more negatively compared to those with a lower BMI. However, this did not imply that people with higher BMI were more self-aware about their health, as their annual medical checkup frequency showed an opposite trend (18/35, 51.4% of those with a BMI≥30 kg/m^2 had fewer than 1 checkup per year). Rather, it may suggest that individuals with higher BMI face more health challenges, which could lead them to perceive their health more negatively. Therefore, it is not necessary that individuals with higher BMI have a stronger sense of self-awareness, but rather that their higher BMI is likely associated with more health issues, which influence how they perceive their overall health.

Table 2. Personal health information and health perception of the sample population by BMI (N=184).

Variables	Total, n (%)	BMI<24 kg/m^2, n (%)	24≤BMI<29.9 kg/m^2, n (%)	BMI≥30 kg/m^2, n (%)	*P* value[a]
Personal health problems or family medical history					.16
1-3	70 (38)	24 (33.8)	35 (44.9)	11 (31.4)	
>3	4 (2.2)	0 (0)	2 (2.6)	2 (5.7)	
None	110 (59.8)	47 (66.2)	41 (52.6)	22 (62.9)	
In 12 months, types of health-related information accessed					.11
1-3	90 (48.9)	30 (42.2)	45 (57.7)	15 (42.9)	
>3	32 (17.4)	10 (14.1)	13 (16.7)	9 (25.7)	
None	62 (33.7)	31 (43.7)	20 (25.6)	11 (31.4)	
In 12 months, the attended health education sessions					.67
1-3	29 (15.8)	12 (16.9)	10 (12.8)	7 (20)	
>3	1 (0.5)	0 (0)	1 (1.3)	0 (0)	
None	154 (83.7)	59 (83.1)	67 (85.9)	28 (80)	
Weekly physical activity (minutes)					.86
<150	56 (30.4)	21 (29.6)	23 (29.5)	12 (34.3)	
≥150	128 (69.4)	50 (70.4)	55 (70.5)	23 (65.7)	
Weekly frequency of personal health tracking					.15
<1 or none	151 (82.1)	57 (80.3)	62 (79.5)	32 (91.4)	
1-3	11 (6)	2 (2.8)	7 (9)	2 (5.7)	
>3	22 (11.9)	12 (16.9)	9 (11.5)	1 (2.9)	
Annual frequency of personal medical checkup					.03[b]
<1 or none	77 (41.8)	27 (38)	32 (41)	18 (51.4)	
1	101 (54.9)	43 (60.6)	45 (57.7)	13 (37.2)	
≥2	6 (3.3)	1 (1.4)	1 (1.3)	4 (11.4)	
The self-rating of personal health status					.02[b]
1	4 (2.2)	0 (0)	1 (1.3)	3 (8.6)	
2-3	79 (43.9)	26 (36.6)	36 (46.2)	17 (48.6)	
4-5	101 (54.9)	45 (63.4)	41 (52.5)	15 (42.8)	

[a]Differences based on chi-square test.

[b]*P*<.05.

The typology of participants based on their use of multiple digital technologies for their health management was analyzed according to BMI classification (Table 3). Initially, 32.1% (59/184) of respondents reported using Bluetooth- or Wi-Fi–connected devices for their health management. Among them, the use rates for the BMI<24 kg/m^2, 24≤BMI<29.9 kg/m^2, and BMI≥30 kg/m^2 were 33.8% (24/71), 33.3% (26/78), and 25.7% (9/35), respectively, with a *P* value of .67. Furthermore, 49.2% (29/59) of users reported using connected devices less than once per week, 25.4% (15/59) used them 1-3 times per week, and another 25.4% (15/59) used them more than 3 times per week. The BMI<24 kg/m^2 and 24≤BMI<29.9 kg/m^2 groups had higher use frequencies, with 50% (12/24 and 13/26) in each group, while the BMI≥30 kg/m^2 group showed a lower frequency of 44.4% (4/9). However, this difference did not reach statistical significance (*P*=.13). Regarding the number of parameters measured by connected devices, 57.6% (34/59) of users measured 1 or none parameters, 33.9% (20/59) measured

XSL•FO
RenderX

2-4 parameters, and 8.5% (5/59) measured more than 4 parameters. No significant difference was observed among BMI categories (P=.36). Additionally, 38.6% (71/184) of respondents reported using wearable devices. The use rates in the BMI categories were 38% (27/71) for BMI<24 kg/m^2, 39.7% (31/78) for 24≤BMI<29.9 kg/m^2, and 37.1% (13/35) for BMI≥30 kg/m^2, with a P value of .96. Among the 71 respondents using wearable devices, 56.3% (40/71) used them more than 3 times per week, 21.3% (15/71) used them 1-3 times per week, and 22.5% (16/71) used them less than once per week. Further analysis revealed that the proportion of respondents using wearable devices more than 3 times per week was higher in the BMI<24 kg/m^2 group (20/27, 74.1%) compared to the other 2 groups (15/31, 48.4% for 24≤BMI<29.9 kg/m^2 and 5/13, 38.4% for BMI≥30 kg/m^2), with this difference reaching statistical significance (P=.04). Overall, while BMI did not significantly influence the use behavior of digital health devices, respondents with lower BMI demonstrated higher frequency of use of wearable devices.

Table 3. The use behavior of digital health devices of the sample population by BMI (N=184).

Variables	Total, n (%)	BMI<24 kg/m^2, n (%)	24≤BMI<29.9 kg/m^2, n (%)	BMI≥30 kg/m^2, n (%)	P value[a]
The use of connected devices for health management					.67
Yes	59 (32.1)	24 (33.8)	26 (33.3)	9 (25.7)	
None	125 (67.9)	47 (66.2)	52 (66.7)	26 (74.3)	
Weekly frequency of using connected devices (for "Yes" respondents)					.13
<1	29 (49.2)	12 (50)	13 (50)	4 (44.4)	
1-3	15 (25.4)	4 (16.7)	10 (38.5)	1 (11.1)	
>3	15 (25.4)	8 (33.3)	3 (11.5)	4 (44.4)	
The parameters measured using connected devices (for "Yes" respondents)					.36
1 or none	34 (57.6)	14 (58.3)	15 (57.7)	5 (55.6)	
2-4	20 (33.9)	6 (25)	10 (38.5)	4 (44.4)	
>4	5 (8.5)	4 (16.7)	1 (3.8)	0 (0)	
The use of wearable devices for health management					.96
Yes	71 (38.6)	27 (38)	31 (39.7)	13 (37.1)	
None	113 (61.4)	44 (62)	47 (60.3)	22 (62.9)	
Weekly frequency of using wearable devices (for "Yes" respondents)					.04[b]
<1	16 (22.5)	4 (14.8)	6 (19.4)	6 (46.2)	
1-3	15 (21.3)	3 (11.1)	10 (32.3)	2 (15.4)	
>3	40 (56.3)	20 (74.1)	15 (48.4)	5 (38.4)	
The parameters measured using wearable devices (for "Yes" respondents)					.63
1 or none	13 (18.3)	4 (14.8)	8 (25.8)	1 (7.7)	
2-4	51 (71.8)	20 (74.1)	20 (64.5)	11 (84.6)	
>4	7 (9.9)	3 (11.1)	3 (9.7)	1 (7.7)	
The use of mobile apps for health management					.61
Yes	72 (39.1)	34 (47.9)	27 (35.5)	9 (25.7)	
None	112 (60.9)	37 (52.1)	49 (64.5)	26 (74.3)	
Weekly frequency of using mobile apps (for "Yes" respondents)					.24
<1	43 (59.7)	17 (50)	20 (69)	6 (66.7)	
1-3	21 (29.2)	12 (35.3)	7 (24.1)	2 (22.2)	
>3	8 (11.1)	5 (14.7)	2 (6.9)	1 (11.1)	

[a]Differences based on chi-square test.

[b]P<.05.

Similarly, 39.1% (72/184) of respondents reported using mobile apps for health management. The use rates were 47.9% (34/71) for BMI<24 kg/m^2, 35.5% (27/78) for 24≤BMI<29.9 kg/m^2, and 25.7% (9/35) for BMI≥30 kg/m^2, with no significant difference observed (P=.61). Among mobile app users, 59.7% (43/72) used apps less than once per week, 29.2% (21/72) used

XSL•FO
RenderX

them 1-3 times per week, and only 11.1% (8/72) used them more than 3 times per week. Although the BMI<24 kg/m^2 group had a higher proportion of frequent users (5/34, 14.7% with >3 times per week) compared to other groups, this trend did not reach statistical significance (*P*=.24).

While the initial chi-square analysis provided valuable insights into the categorical relationship between BMI categories and the use of digital health devices, it did not account for other potential confounding factors. To gain a more comprehensive understanding of the impact of BMI on digital health device use frequency, we conducted a logistic regression analysis based on all respondents. Table 4 reveals that BMI had no statistically significant impact on the frequency of use for connected devices (BMI coefficient=–0.084; *P*=.76) or wearable devices (BMI coefficient=–0.26; *P*=.25). The finding suggested that BMI did not substantially influence the frequency of use of these devices, as indicated by the low pseudo R^2 and the likelihood ratio (LLR) test *P* values of .76 and .25, respectively. However, a significant negative association was found between BMI and the frequency of mobile app use (BMI coefficient=–0.67; *P*=.02). Specifically, individuals with higher BMI were less likely to use mHealth apps frequently. The model for mobile app use demonstrated a higher pseudo R^2 value of 0.035, and the LLR *P* value was .02, suggesting a moderate explanatory power of the model in predicting mobile app use frequency. By using logistic regression, we were able to evaluate the influence of BMI and identify trends that were not apparent in the initial chi-square analysis, particularly the relationship between higher BMI and reduced frequency of mobile app use.

Table 4. Logistic regression analysis of BMI and digital health device use frequency classification results.

Logistic regression model	BMI coefficient (β)	*P* value[a]	Pseudo R^{2}[b]	LLR[c] *P* value
Connected devices	–0.08	.76	0.001	.76
Wearable devices	–0.26	.25	0.006	.25
Mobile apps	–0.67	.02[d]	0.035	.02[d]

[a]Differences based on the regression model.

[b]Pseudo R^2: pseudo coefficient of determination.

[c]LLR: likelihood ratio.

[d]*P*<.05.

Correlations Between Clusters Based on Health-Related Variables

Table 5 presents the Pearson correlation analysis, which was conducted to examine the relationships between various psychological and behavioral factors, including health awareness, health promotion lifestyle, self-efficacy, health motivation, and digital health use. The results of the analysis provide valuable insights into how these factors are interrelated. Health awareness was positively correlated with self-efficacy (*r*=0.206; *P*=.06), digital health use (*r*=0.333; *P*=.002), and health motivation (*r*=0.525; *P*<.001). These results suggested that individuals with higher health awareness were more likely to demonstrate greater self-efficacy, use digital health technologies more frequently, and exhibit stronger health motivation. Conversely, health promotion lifestyle had a significant negative correlation with self-efficacy (*r*=–0.339; *P*=.002), which may imply that individuals engaged in more health-promoting behaviors may have a diminished sense of personal control over their health outcomes. A similar negative correlation was observed between health promotion lifestyle and health motivation (*r*=–0.214; *P*=.007). This finding suggests that adopting a health-conscious lifestyle does not necessarily lead to higher motivation to improve health. Furthermore, no significant relationship was found between health promotion lifestyle and digital health use (*r*=–0.136; *P*=.03). Self-efficacy correlated positively with both digital health use (*r*=0.359; *P*=.001) and health motivation (*r*=0.469; *P*<.001), indicating that individuals who believe in their ability to manage their health are more likely to engage with digital health tools and pursue health-related goals. Additionally, health motivation displayed a strong positive correlation with digital health use (*r*=0.535; *P*<.001), further emphasizing the role of motivation in driving the adoption of digital health technologies. The Pearson correlation analysis revealed strong associations between self-efficacy, health motivation, and digital health use. The correlation coefficients represent the strength and direction of the relationships between the variables, while the corresponding *P* values indicate the statistical significance of these associations. These results underscore the tendency for individuals who feel confident in managing their health and are highly motivated to improve it to engage more with digital health solutions. In contrast, the health promotion lifestyle showed weaker or negative correlations with other variables, suggesting a more complex interaction that warrants further investigation.

XSL•FO
RenderX

Table 5. Pearson correlation between variables (N=184).

	Health awareness	Health promotion lifestyle	Self-efficacy	Digital health use	Health motivation
Health awareness					
r	1	–0.149	0.206	0.333	0.525
P value	.001	.04	.06	.001	.001
Health promotion lifestyle					
r	–0.149	1	–0.339	–0.136	–0.214
P value	.04	.001	.002	.07	.004
Self-efficacy					
r	0.206	–0.339	1	0.359	0.469
P value	.06	.002	.001	.001	.001
Digital health use					
r	0.333	–0.136	0.359	1	0.535
P value	.001	.07	.001	.001	.001
Health motivation					
r	0.525	–0.214	0.469	0.535	1
P value	.001	.004	.001	.001	.001

Predictors of Digital Health Use

The multiple regression analysis was conducted to examine the predictors of digital health use. This analysis aimed to explore how these factors, including health awareness, health promotion lifestyle, self-efficacy, and health motivation, contributed to the likelihood of engaging with digital health technologies. The results of the regression model are present in Table 6, highlighting the regression coefficients (B), standardized coefficients (β), and the associated *P* values for each predictor. Among the predictors, health awareness (B=0.107; β=0.101; *P*=.32), health promotion lifestyle (B=–0.096; β=–0.054; *P*=.58), and self-efficacy (B=0.122; β=0.098; *P*=.36) were found to have no statistically significant influence on digital health use. This suggested that, by accounting for other variables in the model, these factors did not contribute meaningfully to explaining the variance in digital health use in this particular dataset. In contrast, health motivation (B=0.597; β=0.474; *P*<.001) was found to have a significant positive effect on digital health use, supporting the conclusion that health motivation is a key predictor of digital health use. Data indicated that individuals with higher levels of health motivation were more likely to engage with digital health technologies.

Table 6. Multiple regression analysis of overall variables for digital health use[a].

Predicators	Digital health use		
	B[b]	β[c]	*P* value[d]
Health awareness	0.107	0.101	.32
Health promotion lifestyle	–0.096	–0.054	.58
Self-efficacy	0.122	0.098	.36
Health motivation	0.597	0.474	.001[e]

[a]Adjusted R^2=0.324; *P* value (model) based on *F* test <.001.

[b]B: regression coefficients.

[c]β: standardized coefficients.

[d]2-tailed *t* test for regression coefficients.

[e]*P*<.01.

The overall model's goodness-of-fit was indicated by the adjusted R^2 value of 0.324, which suggested that approximately 32.4% of the variance in digital health use can be explained by the combination of the predictors included in the model. Furthermore, the overall model was statistically significant, demonstrating that the predictors as a group significantly contribute to explaining digital health use. While health awareness, health promotion lifestyle, and self-efficacy were not significant predictors of digital health use in this analysis, health motivation emerged as a crucial factor influencing engagement with digital health technologies. These findings suggested that interventions aimed at increasing health

XSL•FO
RenderX

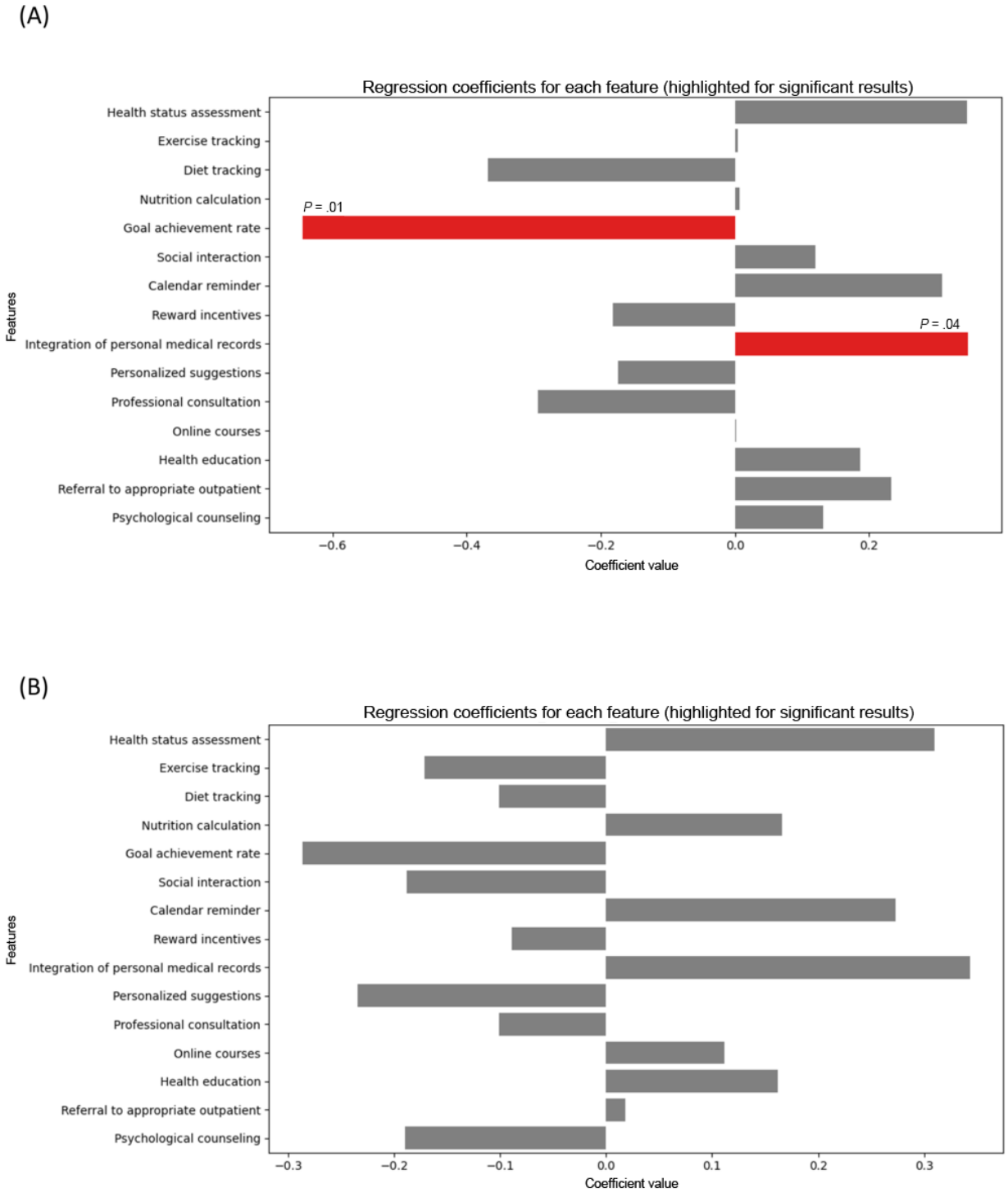
motivation may be particularly effective in encouraging the use of digital health tools.

User Engagement With Mobile App Features Across Different BMI Categories

To understand which digital health features (particularly those of mobile apps) might impact user engagement, we further identified which app features were likely to impact individuals with higher BMI (ie, 24≤BMI<29.9 kg/m^2 and BMI≥30 kg/m^2). We applied multinomial logistic regression to model the relationship between multiple app features (as independent variables) and BMI categories (as the dependent variable). Using BMI<24 kg/m^2 as the reference category, the model estimated the differences between the other BMI categories (24≤BMI<29.9 kg/m^2 and BMI≥30 kg/m^2) and the BMI<24 kg/m^2 category. This approach was similar to creating dummy variables for categorical data in regression models, where a categorical variable is represented by binary variables (0 or 1). The coefficient for the reference category (BMI<24 kg/m^2) was treated as 0, and the coefficients for the other categories (24≤BMI<29.9 kg/m^2 and BMI≥30 kg/m^2) reflect their relative differences from this baseline. The results from the multinomial logistic regression model revealed several important aspects about the relationship between app features and BMI categories (Figure 1 and Table S6 in Multimedia Appendix 1). The pseudo R^2 value of 0.128 indicated that approximately 12.8% of the variance in the BMI categories was explained by the model, suggesting a moderate fit. The statistically significant LLR *P* value of .01 further emphasized the relevance of the app features in differentiating the BMI categories. Notably, the feature of integrating personal medical records showed a positive association with individuals in the 24≤BMI<29.9 kg/m^2 category. For individuals with a BMI≥30 kg/m^2, while the app features did not show statistically significant differences (*P*=.06), further research will be needed to explore whether individuals in the higher BMI range are more likely to engage with digital health technologies offering similar features.

XSL•FO
RenderX

Figure 1. Regression coefficients for (A) 24≤BMI<29.9 kg/m^2 and (B) BMI≥30 kg/m^2 with BMI<24 kg/m^2 as the reference category.



Discussion

Principal Findings

These results indicated that BMI showed significant differences with sex and age, but no significant differences were found in terms of education level and experience with digital technology. Additionally, BMI was negatively correlated with personal health assessment, with participants with higher BMI tending to rate their health more negatively. While BMI did not significantly affect the use of connected and wearable devices for health management, it negatively impacted the frequency of mobile app use, with higher BMI being associated with less frequent use. This analysis not only confirmed the lack of significant relationships between BMI and connected or wearable devices but also highlighted a significant negative association between BMI and the frequency of mobile app use. Additionally, in the multiple regression analysis, health motivation was found to have a significant positive effect on digital health use, whereas health awareness, health promotion lifestyle, and self-efficacy did not significantly influence digital health use. The goodness-of-fit of the overall model indicated

XSL•FO
RenderX

that health motivation was an important factor influencing digital health use.

Initially, chi-square analysis was conducted to examine the associations between BMI and device use frequencies, revealing a significant relationship between BMI and wearable device use. However, this analysis did not account for the continuous nature of the data or potential confounding factors. To address these limitations, logistic regression was used to model the likelihood of frequent device use as a function of BMI, providing a more nuanced exploration of the relationship. While no additional confounders were included in the model, the results from logistic regression revealed a significant negative association between BMI and the frequency of mobile app use. This suggested that individuals with higher BMIs may engage less with mHealth apps, possibly due to lifestyle or health-related factors not captured in the initial analysis. The logistic regression also indicated a stronger model fit for mobile app use compared to other devices, highlighting that BMI had a more pronounced effect on mobile app use frequency than on the use of wearable or other connected devices. This difference may be partially explained by the varying levels of user engagement required across device types. Previous research has demonstrated that wearable devices often feature passive data collection and embed behavior change techniques, such as goal setting, self-monitoring, and user feedback [32]. In contrast, mHealth apps generally require more active engagement. For individuals with higher BMI, this may be further complicated by stigma-related discomfort when logging sensitive data (eg, diet or weight). According to the health stigma and discrimination framework proposed by Stangl et al [33], such self-monitoring may intensify awareness of weight-related issues and discourage sustained use. Additionally, apps often incorporate normative comparisons that may heighten psychological burden, whereas wearables operate unobtrusively in the background, minimizing such stressors. These user experience differences may help explain why individuals with higher BMI in our study reported lower app use but no corresponding decline in wearable use.

In this study, we did not find a correlation between the use of digital health tools and health promotion lifestyle, suggesting that digital health technologies may not significantly improve health-related lifestyles or behaviors in this population. In addition, our results indicated that individuals with higher levels of health motivation were more likely to engage with digital health technologies. However, when health motivation was removed from the model, the direct effects of self-efficacy and health awareness became significant predictors of the digital health use (Table S3 in Multimedia Appendix 1). This prompted a deeper investigation into the role of health motivation through mediation analysis, which helps explain the underlying mechanism or pathway by which the independent variables affect the dependent variable. To explore this, a regression analysis was conducted to examine the effects of health awareness and self-efficacy on health motivation. Both factors significantly predicted health motivation, explaining 34.4% of its variance. Further analysis showed that health motivation significantly influenced the digital health use, accounting for 33.7% of the variance in the dependent variable (Tables S4 and S5 in Multimedia Appendix 1). These findings suggested that health motivation partially mediates the relationship between health awareness, self-efficacy, and digital health use.

This mediation was theoretically meaningful because it emphasized that awareness and confidence alone may not automatically lead to digital health engagement without intrinsic motivation to act [23]. Health motivation acted as a bridge that translated an individual's awareness of health risks and confidence in managing health into actual engagement with digital tools. This mechanism also helped clarify an observation from our data that individuals with higher BMI (≥30 kg/m^2) showed greater health awareness but lower self-efficacy. Prior literature indicated that while people with obesity may have a high level of awareness regarding their health risks, they often faced complex psychological and behavioral barriers that reduced their confidence to initiate or sustain change [34]. Despite high awareness, individuals with obesity may experience internalized weight stigma and frustration from previous unsuccessful attempts. These emotional burdens, along with feelings of helplessness, could undermine self-efficacy, thereby weakening the link between awareness and proactive behavior [33]. These results suggested that heightened awareness alone was insufficient to drive higher engagement with digital health tools among individuals with higher BMI (≥30 kg/m^2). To effectively support this population, interventions or app designs should include features that actively cultivate health motivation, the key mediator linking awareness and self-efficacy into action. Features such as personal goal setting, perceived relevance, and meaningful feedback may help reinforce this pathway and encourage sustained health behaviors [35,36].

Therefore, health motivation played a crucial mediating role in driving individuals' engagement with digital health technologies, suggesting the importance of building both confidence and motivation to encourage proactive health management. mHealth apps and digital devices can serve as a transformative role by empowering individuals to adopt and sustain healthy behaviors. Their interactive features reinforce adherence to health management plans and enhance communication between users and their health care providers [37-39]. This strong predictive power of health motivation may stem from the fact that motivated individuals are more inclined to actively seek out tools that support their health goals. As suggested in previous research, users often perceived digital health technologies as useful tools for achieving behavioral change and self-management [32,35]. Health motivation fostered a proactive mindset, which not only enhanced the perceived usefulness of digital apps but also strengthened the intention to adopt them. Behavioral change is often influenced by perceived outcomes, and digital tools that effectively demonstrate the benefits of specific actions can boost motivation and engagement [40]. Self-efficacy directly influences an individual's likelihood of engaging in health behaviors and their ability to envision and achieve desired outcomes [41,42]. Higher self-efficacy strengthens the link between intent and action, significantly shaping how individuals adopt and use digital technologies. Consequently, these tools not only enhance health management and disease prevention but also empower individuals to take greater control of their health [43,44]. This empowerment is

XSL•FO
RenderX

closely tied to motivation and self-efficacy, both of which are essential for informed decision-making and active participation in health management. The World Health Organization defines empowerment as "a process through which people gain greater control over decisions and actions affecting their health" [45]. This highlights the critical role that self-efficacy and motivation play in helping individuals actively engage in their health management. Furthermore, digital technologies amplify this process by enabling individuals to evaluate and communicate health information more effectively, which requires digital literacy [43]. As a result, the frequency of digital health use also serves as a key indicator of an individual's self-determination and empowerment.

Our findings indicated that individuals with a higher BMI (≥30 kg/m^2) demonstrated greater awareness of their health status (eg, lower scores on self-rating of personal health), reflecting relatively stronger health consciousness. However, this heightened awareness did not translate into proactive health behaviors, such as regular annual checkups. The group of BMI≥30 kg/m^2 exhibited lower confidence in their ability to improve or manage their health, suggesting reduced self-efficacy. While stronger health awareness is evident, the lack of self-efficacy may explain their limited engagement in preventive health measures like regular health checkups. The results align with previous literature and our regression model, both of which highlight health awareness and self-efficacy as key factors influencing health behaviors and motivation [23,32]. Furthermore, studies focused on adolescents diagnosed with overweight or obesity and highlighted that extrinsic motivators like peer support and app-based incentives often led to short-term improvements in BMI [46-48]. This underscored that while digital tools or extrinsic motivators may elicit initial engagement, their long-term effectiveness could be limited. Sustained behavior change, particularly among adults with low self-efficacy, required intrinsic factors such as health motivation, which our findings identified as a critical mediator in digital health adoption.

By fostering a more proactive patient-provider dynamic, mHealth tools encourage shared responsibility and greater engagement in health management [49-51]. Building on this, we further explored which app features are most likely to resonate with individuals with higher BMI (ie, 24≤BMI<29.9 kg/m^2 and BMI≥30 kg/m^2). Among these, the integration of personal medical records was positively associated with higher BMI groups, suggesting that this feature may promote greater engagement in health management and potentially lead to improved health outcomes. This association could be attributed to their heightened focus on health management. Individuals with higher BMI often exhibit greater awareness of their health status, which may motivate them to actively track and manage their health information. In this context, the integration of personal medical records could serve as a critical tool, enabling individuals with higher BMI to make more informed health decisions while further reinforcing their health awareness. Conversely, the goal achievement rate exhibited a negative correlation with the likelihood of being classified within the 24≤BMI<29.9 kg/m^2 category. As the goal achievement rate increased, the probability of individuals being in this BMI range decreased, revealing an inverse relationship between goal achievement and BMI within this group. This suggested that individuals with higher goal achievement rates were less likely to fall into the 24≤BMI<29.9 kg/m^2 category, potentially reflecting improved health outcomes or more proactive engagement in health behaviors.

These findings are further supported by prior intervention studies, which have demonstrated goal achievement as a key determinant of weight-related outcomes. Burke et al [52] observed that adults who regularly tracked diet and activity through digital platforms achieved clinically significant weight loss (≥5%). Participants who actively set and achieved physical activity goals via apps have shown improved BMI and weight outcomes. The mechanism may function through positive reinforcement and increased self-efficacy, which were associated with sustained behavior change and weight maintenance [53]. In our study, higher goal achievement may reflect more consistent and effective engagement in health-promoting behaviors, thereby supporting BMI levels below 24 kg/m^2. These insights underscored the importance of building digital health tools that not only enabled goal-setting but also facilitated goal achievement through features like adaptive difficulty or real-time feedback. Such design considerations could be especially beneficial in interventions targeting individuals at risk of overweight or obesity. While the association between goal achievement rate and BMI≥30 kg/m^2 was not significant, this may be due to the more complex challenges faced by individuals with higher BMIs in terms of health management [52,54]. These individuals often encountered greater barriers, including psychological factors like health stigma and societal pressure, which may undermine their motivation to engage with digital health tools. Additionally, the difficulty in setting and achieving realistic goals may lead to frustration and disengagement, further contributing to the lack of a significant association. Overall, our findings demonstrated that app features may influence individuals across different BMI categories, potentially enhancing engagement with digital health technologies, particularly among those with higher BMI. However, further research is needed to explore the relationship between psychological and motivational factors and the underlying mechanisms driving these patterns.

Limitations

While this study provided valuable insights, several limitations should be noted. The sample size of 184 participants and the focus on a regional population may limit the generalizability of the findings, particularly concerning BMI subgroups and regional or ethnic factors. The participants were recruited from an urban area, a region where preventive health services were accessible but not uniformly used. Health management in this setting was often influenced by cultural emphasis on body image, familial expectations, and access to publicly funded health care services. These factors may shape both health awareness and behavioral engagement in ways that differ from populations in Western countries or other regions. Additionally, the study relied on self-reported data for health assessments and digital health use, which may be subject to biases such as social

XSL•FO
RenderX

desirability or recall bias. These biases could have led to overreporting of healthy behaviors or engagement with digital tools, potentially affecting the observed associations. To address this, future research could incorporate objective measures (eg, app use logs or wearable device data) to validate self-reports. Moreover, expanding the sample size and including a more diverse population would strengthen the findings and improve the external validity of the results. Future studies could adopt more rigorous sampling strategies, including stratified or probability sampling to ensure better representation across BMI categories, regions, and socioeconomic backgrounds. Longitudinal studies will be needed to assess the directionality of the relationships between BMI, health behaviors, and digital health technology use. This study focused on psychological factors such as health awareness, self-efficacy, and motivation. Other unmeasured confounding variables such as socioeconomic status, physical activity levels, or comorbid conditions could also influence the relationships observed between BMI and health-related behaviors. These factors were not controlled in the current design and may partially explain the variability in digital health engagement and health behavior adoption. Finally, while this study focused on wearables and mobile apps and their features, future research should explore other digital health tools, such as telemedicine platforms, to investigate their role in health management across different BMI categories. These tools often provide additional functionalities, such as real-time virtual consultations, personalized feedback, and long-term health monitoring, which may be particularly valuable for individuals with higher BMI who require continuous support or tailored interventions. Moreover, to better understand the underlying psychological or emotional factors influencing digital health engagement, qualitative research such as in-depth interviews could provide richer insights into user experiences, perceived barriers, and motivational drivers.

Conclusions

The results of this study indicate a relationship between BMI, health motivation, and the use of digital health technologies, particularly in the use of mHealth apps. Although BMI did not significantly affect the frequency of digital health device use, a higher BMI was associated with a lower frequency of mobile app use, suggesting that individuals with higher BMI are less likely to use mHealth apps frequently. This implies that individuals with higher BMI may exhibit differences in health motivation, which in turn impacts their frequency of digital health tool use. Furthermore, in the multiple regression analysis, health motivation was found to have a significant positive influence on digital health use. This suggests that health motivation plays a crucial role in influencing the use of digital health technologies, with individuals exhibiting higher health motivation showing greater engagement. Communication strategies should also consider users' preferences by offering flexible notification options, such as message, email, or in-app settings, allowing individuals to tailor the frequency and type of communication they receive. By emphasizing the integration of personalized health data, health care providers could potentially increase user engagement, particularly among individuals with higher BMIs, by providing more tailored and relevant health insights. This could improve their frequency of use and ultimately encourage healthier behaviors. Additionally, improving the accessibility and usability of mHealth tools may help foster greater engagement across different BMI groups.

Acknowledgments

The authors extend special thanks to Ms Hsieh Fang-Chou and Dr Tieng-Chun Chang for expert input during the questionnaire development phase. The authors also thank Mr Nianfeng Ai for providing the licensed software. Generative artificial intelligence (ChatGPT) was used to assist with language polishing, but all scientific content was generated and validated by the authors.

Funding

This study was supported by the Center for Traditional Medicine, Taipei Veterans General Hospital, Taipei, Taiwan (V108B-031, V109B-034, V112C-190, V113C-087, V114C-078), National Yang Ming Chiao Tung University, Taipei, Taiwan, and National Science and Technology Council, Taipei, Taiwan (NSTC-111-2320-B-A49A-501-MY2 and NSTC-113-2320-B-A49-029).

Data Availability

The datasets used and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: SHC, CMC
Methodology: SHC, CMC, DC, KKH
Formal analysis: SHC, DC
Writing—original draft preparation: SHC, LKY, MCY
Writing—review and editing: SHC, CMC, DC
Supervision: SHC, LKY, CMC

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Supplementary tables summarizing measurement tools, reliability and validity analyses, and regression models used in this study.
[DOCX File , 29 KB - mhealth_v13i1e71625_app1.docx]

References

1. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
2. Shih P, Prokopovich K, Degeling C, Street J, Carter SM. Direct-to-consumer detection of atrial fibrillation in a smartwatch electrocardiogram: medical overuse, medicalisation and the experience of consumers. Soc Sci Med 2022;303:114954. [doi: 10.1016/j.socscimed.2022.114954] [Medline: 35569232]
3. Floryan M, Chow PI, Schueller SM, Ritterband LM. The model of gamification principles for digital health interventions: evaluation of validity and potential utility. J Med Internet Res 2020;22(6):e16506 [FREE Full text] [doi: 10.2196/16506] [Medline: 32519965]
4. Vansimaeys C, Benamar L, Balagué C. Digital health and management of chronic disease: a multimodal technologies typology. Int J Health Plann Manage 2021;36(4):1107-1125. [doi: 10.1002/hpm.3161] [Medline: 33786849]
5. Ross KM, Wing RR. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: a randomized pilot study. Obesity (Silver Spring) 2016;24(8):1653-1659 [FREE Full text] [doi: 10.1002/oby.21536] [Medline: 27367614]
6. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
7. Leong D, Roberts E. Social determinants of health and the affordable care act. R I Med J 2013;96(7):20-22. [Medline: 23819136]
8. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
9. Brindal E, Hendrie GA, Freyne J, Noakes M. Incorporating a static versus supportive mobile phone app into a partial meal replacement program with face-to-face support: randomized controlled trial. JMIR Mhealth Uhealth 2018;6(4):e41 [FREE Full text] [doi: 10.2196/mhealth.7796] [Medline: 29669704]
10. Alyafei A, Easton-Carr R. The Health Belief Model of Behavior Change. Treasure Island, FL: StatPearls; 2025.
11. Yang M, Al Mamun A, Gao J, Rahman MK, Salameh AA, Alam SS. Predicting m-health acceptance from the perspective of unified theory of acceptance and use of technology. Sci Rep 2024;14(1):339 [FREE Full text] [doi: 10.1038/s41598-023-50436-2] [Medline: 38172184]
12. Scherer PE, Hill JA. Obesity, diabetes, and cardiovascular diseases: a compendium. Circ Res 2016;118(11):1703-1705. [doi: 10.1161/circresaha.116.308999]
13. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
14. Moses JC, Adibi S, Shariful Islam SM, Wickramasinghe N, Nguyen L. Application of smartphone technologies in disease monitoring: a systematic review. Healthcare (Basel) 2021;9(7):889 [FREE Full text] [doi: 10.3390/healthcare9070889] [Medline: 34356267]
15. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
16. Ajayi KV, Wachira E, Onyeaka HK, Montour T, Olowolaju S, Garney W. The use of digital health tools for health promotion among women with and without chronic diseases: insights from the 2017-2020 Health Information National Trends survey. JMIR Mhealth Uhealth 2022;10(8):e39520 [FREE Full text] [doi: 10.2196/39520] [Medline: 35984680]
17. Lin X, Xu Y, Pan X, Xu J, Ding Y, Sun X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. Sci Rep 2020;10(1):14790 [FREE Full text] [doi: 10.1038/s41598-020-71908-9] [Medline: 32901098]
18. Chao DP. Health-promoting lifestyle and its predictors among health-related and non-health-related university students in Taiwan: a cross-sectional quantitative study. BMC Public Health 2023;23(1):827 [FREE Full text] [doi: 10.1186/s12889-023-15760-2] [Medline: 37147650]
19. Stevens G, Mascarenhas M, Mathers C. Global health risks: progress and challenges. Bull World Health Organ 2009;87(9):646-646 [FREE Full text] [doi: 10.2471/blt.09.070565] [Medline: 19784438]
20. Alsisi EA, Al-Ashaab A, Abualfaraa WA. The development of a smart health awareness message framework based on the use of social Media: quantitative study. J Med Internet Res 2020;22(7):e16212 [FREE Full text] [doi: 10.2196/16212] [Medline: 32459627]
21. Stühmann LM, Paprott R, Heidemann C, Baumert J, Hansen S, Zahn D, et al. Health app use and its correlates among individuals with and without type 2 diabetes: nationwide population-based survey. JMIR Diabetes 2020;5(2):e14396 [FREE Full text] [doi: 10.2196/14396] [Medline: 32432555]

XSL•FO
RenderX

22. Teng HL, Yen M, Fetzer S. Health Promotion Lifestyle Profile-II: Chinese version short form. J Adv Nurs 2010;66(8):1864-1873. [doi: 10.1111/j.1365-2648.2010.05353.x] [Medline: 20557380]
23. Niu Z, Willoughby J, Zhou R. Associations of health literacy, social media use, and self-efficacy with health information-seeking intentions among social media users in China: cross-sectional survey. J Med Internet Res 2021;23(2):e19134 [FREE Full text] [doi: 10.2196/19134] [Medline: 33629955]
24. Chamorro-Koc M, Peake J, Meek A, Manimont G. Self-efficacy and trust in consumers' use of health-technologies devices for sports. Heliyon 2021;7(8):e07794 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e07794] [Medline: 34458625]
25. Choudry M, Ganti L. Exploration of the motivational factors that influence the maintenance of health. Health Psychol Res 2024;12(1):115356 [FREE Full text] [doi: 10.52965/001c.115356] [Medline: 38533343]
26. Xu X, Zhang G, Xia Y, Xie H, Ding Z, Wang H, et al. Influencing factors and implementation pathways of adherence behavior in intelligent personalized exercise prescription: qualitative study. JMIR Mhealth Uhealth 2024;12:e59610 [FREE Full text] [doi: 10.2196/59610] [Medline: 39636668]
27. Rosman L, Lampert R, Zhuo S, Li Q, Varma N, Burg M, et al. Wearable devices, health care use, and psychological well‐being in patients with atrial fibrillation. J Am Heart Assoc 2024;13(15):e033750. [doi: 10.1161/jaha.123.033750]
28. Moorthy P, Weinert L, Schüttler C, Svensson L, Sedlmayr B, Müller J, et al. Attributes, methods, and frameworks used to evaluate wearables and their companion mHealth apps: scoping review. JMIR Mhealth Uhealth 2024;12:e52179 [FREE Full text] [doi: 10.2196/52179] [Medline: 38578671]
29. Pulantara IW, Wang Y, Burke LE, Sereika SM, Bizhanova Z, Kariuki JK, et al. Data collection and management of mHealth, wearables, and internet of things in digital behavioral health interventions with the Awesome Data Acquisition Method (ADAM): development of a novel informatics architecture. JMIR Mhealth Uhealth 2024;12:e50043-e50043 [FREE Full text] [doi: 10.2196/50043] [Medline: 39113371]
30. Kim H, Schnall R, Yoon N, Koh S, Lee J, Cheon JH. Development and validation of a Mobile-Centered Digital Health Readiness Scale (mDiHERS): health literacy and equity scale. J Med Internet Res 2024;26:e58497 [FREE Full text] [doi: 10.2196/58497] [Medline: 39137409]
31. Oba T, Takano K, Katahira K, Kimura K. Use patterns of smartphone apps and wearable devices supporting physical activity and exercise: large-scale cross-sectional survey. JMIR Mhealth Uhealth 2023;11:e49148-e49148 [FREE Full text] [doi: 10.2196/49148] [Medline: 37997790]
32. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
33. Stangl AL, Earnshaw VA, Logie CH, van Brakel W, C Simbayi L, Barré I, et al. The health stigma and discrimination framework: a global, crosscutting framework to inform research, intervention development, and policy on health-related stigmas. BMC Med 2019;17(1):31 [FREE Full text] [doi: 10.1186/s12916-019-1271-3] [Medline: 30764826]
34. Teixeira PJ, Carraça EV, Marques MM, Rutter H, Oppert J, De Bourdeaudhuij I, et al. Successful behavior change in obesity interventions in adults: a systematic review of self-regulation mediators. BMC Med 2015;13:84 [FREE Full text] [doi: 10.1186/s12916-015-0323-6] [Medline: 25907778]
35. Ezeigwe OJ, Nwosu KOS, Afolayan OK, Ojaruega AA, Echere J, Desai M, et al. Technological-based interventions in cancer and factors associated with the use of mobile digital wellness and health apps among cancer information seekers: cross-sectional study. J Med Internet Res 2025;27:e63403 [FREE Full text] [doi: 10.2196/63403] [Medline: 39909418]
36. Anand VV, Zhe ELC, Chin YH, Lim WH, Goh RSJ, Lin C, et al. Barriers and facilitators to engagement with a weight management intervention in Asian patients with overweight or obesity: a systematic review. Endocr Pract 2023;29(5):398-407. [doi: 10.1016/j.eprac.2022.10.006] [Medline: 36396016]
37. Sepah SC, Jiang L, Peters AL. Long-term outcomes of a web-based diabetes prevention program: 2-year results of a single-arm longitudinal study. J Med Internet Res 2015;17(4):e92 [FREE Full text] [doi: 10.2196/jmir.4052] [Medline: 25863515]
38. Brath H, Morak J, Kästenbauer T, Modre-Osprian R, Strohner-Kästenbauer H, Schwarz M, et al. Mobile health (mHealth) based medication adherence measurement—a pilot trial using electronic blisters in diabetes patients. Br J Clin Pharmacol 2013;76(Suppl 1):47-55 [FREE Full text] [doi: 10.1111/bcp.12184] [Medline: 24007452]
39. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
40. Rhodes RE, Yao CA. Models accounting for intention-behavior discordance in the physical activity domain: a user's guide, content overview, and review of current evidence. Int J Behav Nutr Phys Act 2015;12(1):9 [FREE Full text] [doi: 10.1186/s12966-015-0168-6] [Medline: 25890238]
41. Tetri B, Juujärvi S. Self-efficacy, internet self-efficacy, and proxy efficacy as predictors of the use of digital social and health care services among mental health service users in Finland: a cross-sectional study. Psychol Res Behav Manag 2022;15:291-303 [FREE Full text] [doi: 10.2147/PRBM.S340867] [Medline: 35210878]

XSL•FO
RenderX

42. Li Z, Zuo T, Wei X, Ding N. ICT self-efficacy scale: the correlations with the age of first access to the internet, the age at first ownership of a personal computer (PC), and a smartphone. Med Educ Online 2023;28(1):2151068 [FREE Full text] [doi: 10.1080/10872981.2022.2151068] [Medline: 36440825]
43. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
44. Lemire M, Sicotte C, Paré G. Internet use and the logics of personal empowerment in health. Health Policy 2008;88(1):130-140. [doi: 10.1016/j.healthpol.2008.03.006] [Medline: 18436330]
45. Smith BJ, Tang KC, Nutbeam D. WHO health promotion glossary: new terms. Health Promot Int 2006;21(4):340-345. [doi: 10.1093/heapro/dal033] [Medline: 16963461]
46. Andrade A, Correia CK, Coimbra DR. The psychological effects of exergames for children and adolescents with obesity: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2019;22(11):724-735. [doi: 10.1089/cyber.2019.0341] [Medline: 31697604]
47. Staiano AE, Abraham AA, Calvert SL. Adolescent exergame play for weight loss and psychosocial improvement: a controlled physical activity intervention. Obesity (Silver Spring) 2013;21(3):598-601 [FREE Full text] [doi: 10.1002/oby.20282] [Medline: 23592669]
48. Chen JL, Guedes CM, Cooper BA, Lung AE. Short-term efficacy of an innovative mobile phone technology-based intervention for weight management for overweight and obese adolescents: pilot study. Interact J Med Res 2017;6(2):e12 [FREE Full text] [doi: 10.2196/ijmr.7860] [Medline: 28768612]
49. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
50. Lamprinos I, Demski H, Mantwill S, Kabak Y, Hildebrand C, Ploessnig M. Modular ICT-based patient empowerment framework for self-management of diabetes: design perspectives and validation results. Int J Med Inform 2016;91:31-43. [doi: 10.1016/j.ijmedinf.2016.04.006] [Medline: 27185507]
51. Qudah B, Luetsch K. The influence of mobile health applications on patient-healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
52. Burke LE, Bizhanova Z, Conroy MB, Cheng J, Beatrice B, Kariuki JK, et al. Adherence to self-monitoring and behavioral goals is associated with improved weight loss in an mHealth randomized-controlled trial. Obesity (Silver Spring) 2025;33(3):478-489. [doi: 10.1002/oby.24234] [Medline: 39962997]
53. Wren GM, Koutoukidis DA, Scragg J, Whitman M, Jebb S. The association between goal setting and weight loss: prospective analysis of a community weight loss program. J Med Internet Res 2023;25:e43869 [FREE Full text] [doi: 10.2196/43869] [Medline: 37405833]
54. Nezami BT, Lang W, Jakicic JM, Davis KK, Polzien K, Rickman AD, et al. The effect of self-efficacy on behavior and weight in a behavioral weight-loss intervention. Health Psychol 2016;35(7):714-722 [FREE Full text] [doi: 10.1037/hea0000378] [Medline: 27183306]

Abbreviations

LLR: likelihood ratio
mHealth: mobile health

Edited by L Buis; submitted 22.01.25; peer-reviewed by K Mouloudj, L Bovsh; comments to author 22.04.25; revised version received 16.06.25; accepted 04.11.25; published 01.12.25.

Please cite as:
Chang SH, Yeh LK, Chen D, Hu KK, Yu MC, Chang CM
Health Motivation as a Predictor of mHealth Engagement Across BMI: Cross-Sectional Survey
JMIR Mhealth Uhealth 2025;13:e71625
URL: https://mhealth.jmir.org/2025/1/e71625
doi:10.2196/71625
PMID:

©Shao-Hsuan Chang, Lung-Kun Yeh, Daishi Chen, Kae-Kuen Hu, Mei-Ching Yu, Ching-Mao Chang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 01.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is

XSL•FO
RenderX

properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Interpretable Machine Learning Models for Analyzing Determinants Affecting the Use of mHealth Apps Among Family Caregivers of Patients With Stroke in Chinese Communities: Cross-Sectional Survey Study

Yun Du[1], BSc; Jun-Ying Fan[1,2], MSc; Guang-Zhi Liu[1], BSc; Zi-Yue Yang[1], BSc; Yang Lei[3], PhD; Yu-Fang Guo[1], PhD

[1]School of Nursing and Rehabilitation, Shandong University, 44 Wen Hua Xi Road, Jinan, Shandong, China

[2]Department of Neurology, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

[3]School of Nursing, Nanjing Medical University, Nanjing, China

Corresponding Author:
Yu-Fang Guo, PhD
School of Nursing and Rehabilitation, Shandong University, 44 Wen Hua Xi Road, Jinan, Shandong, China

Abstract

Background: Mobile health (mHealth) apps are believed to be an effective method to support family caregivers to better care for patients with stroke. This study's purpose was to explore the status and the influencing factors of mHealth app use among family caregivers of patients with stroke via machine learning (ML) models.

Objective: This study aimed to understand the status quo of mHealth app use among community family caregivers of patients with stroke and the factors influencing their use behavior. Six ML models were used to construct the classifier, and the Shapley Additive Explanations (SHAP) algorithm was introduced to interpret the best ML model.

Methods: In this cross-sectional study, family carers of patients with stroke were recruited. Data on their basic profile and mHealth app use were obtained through face-to-face questionnaires. Hedonic motivation, usage habits, and other relevant information were additionally measured among app users. A total of 12 models were constructed using six ML algorithms. The top-performing logistic regression and random forest models were further analyzed with SHAP to interpret key influencing factors.

Results: A total of 360 family caregivers of patients with stroke were included in this study from March 2023 to November 2023, of which 206 (57.2%) reported having used mHealth apps. Of the 6 ML models, the logistic regression model performed the best in terms of whether caregivers used the mHealth app, with an area under the receiver operating characteristic curve of 0.753 (95% CI 0.698 - 0.802), accuracy of 0.694 (95% CI 0.647 - 0.742), sensitivity of 0.748 (95% CI 0.688 - 0.806), and specificity of 0.623 (95% CI 0.547 - 0.698). SHAP analysis showed that the top 5 most influencing factors were educational level, age, the patient's self-care ability, the relationship with the cared-for individual, and the duration of illness. The random forest model performed best in terms of use behavior with an area under the receiver operating characteristic curve of 0.773 (95% CI 0.725 - 0.818), accuracy of 0.602 (95% CI 0.534 - 0.665), sensitivity of 0.476 (95% CI 0.420 - 0.533), and specificity of 0.769 (95% CI 0.738 - 0.797). The SHAP analysis revealed that hedonic motivation, habits, occupation, convenience conditions, and effort expectations were the 5 most significant influencing factors.

Conclusions: The research results indicate that the software developers and policymakers of mHealth apps should take the abovementioned influencing factors into consideration when developing and promoting the software. We should focus on the older adults with lower educational levels, lower the threshold for software use, and provide more convenient conditions. By grasping the hedonistic tendencies and habitual usage characteristics of users, they can provide them with more concise and accurate health information, which will enhance the popularity and effectiveness of mHealth apps.

(JMIR Mhealth Uhealth 2025;13:e73903) doi:10.2196/73903

KEYWORDS

caregivers; mHealth; community; influencing factors; machine learning model; stroke; mobile health

XSL•FO
RenderX

Introduction

Stroke is a central nervous system disorder characterized by acute vascular damage, which is a leading cause of disability and death globally [1,2]. In recent years, despite advancements in medical care that have contributed to reduced stroke mortality rates, more than 75% of surviving patients with stroke still have moderate or severe functional disabilities in limbs, cognition, and language [3-5], which result in patients having limited ability to perform daily life activities and needing prolonged rehabilitation and treatment. Nevertheless, owing to economic and medical limitations, over 80% of patients with stroke opt for home rehabilitation once their condition stabilizes [6], with care predominantly provided by family members [7]. Notably, 25% of stroke survivors were readmitted to the hospital within 90 days, and 73% of patients with stroke experienced falls within 6 months, with 4% sustaining fractures during these incidents, attributed to improper care practices or lack of professional guidance in caregiving [8,9]. Family caregivers of patients with stroke experienced a deficiency in disease knowledge, inadequate caregiving skills, heightened caregiving stress, poor role adjustment, and insufficient social support [10-12]. This adversely impacted the rehabilitation process, diminished patients' quality of life, and escalated the burden on caregivers [13]. Consequently, assisting family caregivers of patients with stroke in obtaining knowledge about the condition and efficient management strategies has emerged as a pressing issue.

The widespread adoption of smartphones and advancements in digital health technology present significant possibilities for the management of stroke through telemedicine and mobile health (mHealth) technologies [14,15]. mHealth apps, a primary manifestation of mHealth, are apps for mobile phones or tablets designed for health care research and services aimed at enhancing patient health outcomes [16]. mHealth apps offer users affordable, high-quality, round-the-clock access to evidence-based health information worldwide and improve patient disease management through behavioral change models [17-19]. Numerous studies supported that mHealth apps used by family caregivers of patients with stroke for disease management enhanced caregivers' understanding of the condition, refined their caregiving skills and efficiency, and positively influenced their psychological well-being, thereby benefiting the patients' recovery process [20,21].

However, most of the studies focused on the validation of the effects of mHealth app use by family carers of patients with stroke, and the study design was also mostly intervention-based, with limitations in external validity. It is important to recognize that utilization thresholds exist for the use of mHealth apps. Disparities in the quantity and quality of knowledge obtained among users during their use are evident [22]. Multiple contributory factors interact with discrepancy, including personal, familial, economic, and other real-world elements [23]. These factors interact to ascertain that the mHealth app can deliver its full potential and genuinely benefit patients and users. However, there is still a lack of clarity regarding the real-world use of mHealth apps by family caregivers of patients with stroke in nonintervention states and the factors influencing whether they use them or not.

In this study, machine learning (ML) algorithms were used to explore the factors influencing their use and use behavior, and the ML approach demonstrated significant advantages over traditional linear regression models. These algorithms can capture more complex nonlinear relationships and have stronger feature selection capabilities and better predictive performance [24]. Therefore, the choice of ML algorithms enables a more comprehensive exploration of influencing factors. However, there is still the disadvantage of limited model interpretability, so this study applied the Shapley Additive Explanations (SHAP) algorithm based on ML algorithms to enhance model interpretability and facilitate a better understanding of the model and results.

In summary, this study applied ML algorithms to investigate the current status of mHealth app use among family caregivers of patients with stroke in the community and the factors influencing usage behavior. The objectives were to provide empirical data to support health care providers and technology developers in focusing on populations who underuse mHealth apps. By gaining a deeper understanding of the real-world barriers these groups face, we expected to be able to provide them with more targeted assistance to increase their usage of mHealth apps and optimize use patterns. This will help ensure that mHealth apps reach a wider audience of patients with stroke and their family caregivers, improving the effectiveness of disease management.

Methods

Study Design and Participants

This study was a cross-sectional survey. The convenience sampling method was used to select family carers of patients with stroke who met the inclusion criteria for the study. A total of 360 eligible family carers of patients with stroke were included.

The following patient inclusion criteria were used: (1) patients with stroke who met the diagnostic criteria of stroke [25] and were diagnosed by cranial computed tomography or magnetic resonance imaging, (2) aged ≥18 years, (3) conscious and with the ability to use words or clearly express their wishes, and (4) informed consent and voluntary participation in the survey. Those who are too ill to cooperate with the investigation are excluded.

Inclusion criteria for family caregivers were as follows:

1. Primary caregiver of the patient with stroke: the person who is related to the patient and who is responsible for the patient's primary care (the person who cares for the patient for the longest time during the day). If there are several primary caregivers at the same time, the patient will designate one of them as the primary caregiver.
2. Aged ≥18 years
3. Informed consent and voluntary participation in the survey

Paid professional caregivers were excluded.

Sample Size Calculation

This study adopted the sample size calculation method based on the 10-fold events per variable (EPV) principle [26,27].

XSL•FO
RenderX

When constructing the model regarding the use of mHealth apps, a total of 10 variables were included. According to the survey results, the use rate of mHealth apps was 57.22%. According to the EPV principle, at least 175 samples were required. When constructing the model for the use behavior of mHealth apps, a total of 8 variables were included. Based on the clustering results, the expected occurrence rate (cluster 3) was 51.5% (Figure S2 in Multimedia Appendix 1). According to the EPV principle, at least 156 samples were needed. This study ultimately included 360 family caregivers of patients with stroke, among whom 206 had used mHealth apps.

Data Collection

Data collection for this study was conducted between March 2023 and November 2023 in 4 community health centers in Jinan, Shandong Province and Xuzhou, Jiangsu Province. Patients with stroke who had ever been diagnosed by neurological specialists were selected among the population attending the community hospitals, and their primary caregivers who met the criteria were included in the study after reviewing the electronic records, resulting in the inclusion of 360 study participants. For those who used mHealth apps, after completing the basic questionnaire, a further questionnaire was completed to investigate their use behaviors. Data were collected face-to-face by a uniformly trained research team using a standardized paper questionnaire, during which the questionnaire entries were interpreted item by item, respondents' questions were uniformly explained, and answers were recorded after verbal confirmation was obtained. The completeness of each questionnaire was verified immediately after completion, and omissions were found to be added on the spot, resulting in 360 questionnaires being recovered and zero invalid questionnaires being eliminated, giving an effective recovery rate of 100%. This quality control process ensures the integrity and reliability of the data.

Influencing Factors

The influencing factors included general sociodemographic data and disease-related data, where the disease-related data were divided into 2 parts: the patient part included the type of stroke (ischemic and hemorrhagic), whether or not they were taking medication, whether or not they had adjusted their lifestyle after the illness, whether or not they were undergoing rehabilitation, whether or not they had been regularly reexamined, the number of sequelae of the stroke, the number of diseases other than stroke, the number of years of the patient's life with a diagnosis of stroke, the number of years since the stroke, and the patient's ability to care for themselves as measured by the Barthel Index [28]. The carer section investigated the number of illnesses the carer had themselves. In addition to this, the type of mHealth app used was captured using the multiple-choice question "Which mHealth App or Apps do you usually use?" In continuing to explore the influences on users' use behavior, standardized maturity scales measuring indicators of performance expectations, effort expectations, habituation, value, and hedonic motivation were included (Multimedia Appendix 1).

Outcomes

Use of mHealth apps was measured using the question "Have you ever used mHealth software?" with responses specified as a dichotomous variable: "yes" and "no." In addition to this, the frequency, duration, and familiarity of carers' use of mHealth apps were measured using questions, which were used to represent the study participants' use behavior in a combined manner.

Statistical Analysis

Python (version 3.11.3; Python Software Foundation) was used. Data were described and analyzed statistically, with categorical variables represented as frequencies and percentages.

The k-modes clustering algorithm was used to categorize the study participants using an mHealth app. Detailed information on the clustering algorithm can be found in Multimedia Appendix 1. The variables entering k-modes consensus clustering included frequency of use, length of use, and familiarity of use. Through multiple iterations, the data points were grouped into clusters with similar characteristics.

Subsequently, we used the elbow method to determine the optimal number of clusters (Multimedia Appendix 1). After determining the optimal clusters, study participants using mHealth apps were divided into groups with different use behaviors according to the clustering results.

ML algorithms were used to explore the factors influencing whether or not an mHealth app is used and its use behavior [29]. Six main classifiers were selected for testing in this study: logistic regression (LR), random forest (RF), support vector machine, extreme gradient boosting, light gradient boosting machine, and naive Bayes (Multimedia Appendix 1).

This study used the nested cross-validation method suitable for small-sample ML [30] (Multimedia Appendix 1). The outer loop is divided into 5 folds to evaluate performance, and the inner loop (divided into 5 layers) is used for hyperparameter tuning. To ensure the reliability of the evaluation results, reduce the randomness of the model, and improve the generalization ability of the model, different strategies are selected for feature selection based on different models in this study (Table S1 in Multimedia Appendix 1). The model's overall performance was assessed using 4 metrics: accuracy, area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity. The Brier score and calibration curve were reported to evaluate the reliability of the predicted probabilities. For each indicator, we combined the prediction results from all the outer test sets to calculate the overall indicator. The 95% CI was calculated through 1000 bootstrap samplings.

For the models exhibiting optimal performance, we used the SHAP algorithm to elucidate the prediction outcomes of the ML models (Multimedia Appendix 1). Among them, the SHAP value represents the minimum contribution of the feature to the model. The higher the SHAP, the greater the influence of the feature on the model. Consequently, using this method facilitates the interpretation of prediction results for each observation sample within the dataset and enhances our comprehension of the ML model. We used the bootstrap method to evaluate the

XSL•FO
RenderX

stability of the SHAP values. We conducted 100 bootstrap iterations and recalculated the SHAP values after each sampling, thereby quantifying the range of fluctuations in feature importance.

A learning curve was drawn to assess the impact of sample size on model performance and diagnose whether the model is underfitting or overfitting. Highly correlated and low-correlated features were removed to evaluate the model's robustness to feature changes and verify the stability of the feature selection strategy.

Ethical Considerations

This study was approved by the local ethics committee (2023-R-036). Throughout the study, we strictly adhered to the principle of informed consent, fully informed the study participants about the study, and respected the respondents' right to independently choose to participate in or withdraw from the study at any time. After the investigation is completed, we will give small gifts (such as towels and soaps) to the participants as a token of our gratitude.

Results

Sample Characteristics

Table 1 presents the demographic and disease-related characteristics of the participants. The age of participants ranged from 23 to 85 years, with a mean age of 61.93 (SD 15.08) years. Among the 360 participants, 54.4% (n=196) were women, 63.6% (n=229) had attained junior high or high school education, 97.2% (n=350) were married, 95% (n=342) resided in urban areas, 58.9% (n=212) were retired, and 85.6% (n=308) had a monthly income below CN ¥6000 (US $836.52). Furthermore, 92.8% (n=334) of caregivers cared for patients with ischemic stroke, and 51.9% (n=187) of carers cared for mildly dependent patients.

XSL•FO
RenderX

Table . Characteristics among caregivers (N=360).

Characteristics	Total, n (%)	Use (n=206), n (%)	Nonuse (n=154), n (%)	Chi-square (*df*)	*P* value
Demographic characteristics					
Gender				0.1 (1)	.72
Man	164 (45.6)	96 (46.6)	68 (44.2)		
Woman	196 (54.4)	110 (53.4)	86 (55.8)		
Age (y)				45.9 (1)	<.001
≤60	139(38.6)	111 (53.9)	28 (18.2)		
>60	221(61.4)	95 (46.1)	126 (81.8)		
Degree of education				61.7 (4)	<.001
Primary and below	62 (17.2)	15 (7.3)	47 (30.5)		
Junior high school	100 (27.8)	64 (31.1)	65 (42.2)		
Senior high school (technical secondary school)	129 (35.8)	66 (32)	34 (22.1)		
Junior college	40 (11.1)	34 (16.5)	6 (3.9)		
Bachelor’s degree and higher	29 (8.1)	27 (13.1)	3 (1.3)		
Marital status				3.2 (1)	.07
Spinsterhood	10 (2.8)	9 (4.4)	1 (0.7)		
Married	350 (97.2)	197 (95.6)	153 (99.4)		
Family place of residence				0.2 (1)	.70
Cities and towns	342 (95)	197 (95.6)	145 (94.2)		
Village	18 (5)	9 (4.4)	9 (5.8)		
Occupational status				29.0 (3)	<.001
Unemployed	46 (12.8)	22 (10.7)	24 (15.6)		
On job	91 (25.3)	74 (35.9)	17 (11)		
Retired	212 (58.9)	105 (51)	107 (69.5)		
Others	11 (3.1)	5 (2.4)	6 (3.9)		
Monthly income (¥)[a]				11.0 (4)	.03
<2000	69 (19.2)	29 (14.1)	40 (25.9)		
2000 - 3999	105 (29.2)	63 (30.6)	42 (27.3)		
4000 - 5999	134 (37.2)	77 (37.4)	57 (37)		
6000 - 7999	31 (8.6)	22(10.7)	9 (5.8)		
>8000	21 (5.8)	15 (7.3)	6 (3.9)		
Patient-caregiver relationship				42.0 (2)	<.001
Spouse	256 (71.1)	121 (58.7)	135 (87.7)		
Parent	98 (27.2)	83 (40.3)	15 (9.7)		
Child	6 (1.7)	2 (1)	4 (2.6)		
Disease characteristics					
Stroke type				0.4 (1)	.50
Hemorrhagic	26 (7.2)	17 (8.2)	9 (5.8)		
Ischemic	334 (92.8)	189 (91.8)	145 (94.2)		
Rehabilitation measures				0.6 (1)	.43
Yes	350 (97.2)	202 (98.1)	148 (96.1)		

XSL•FO
RenderX

Characteristics	Total, n (%)	Use (n=206), n (%)	Nonuse (n=154), n (%)	Chi-square (*df*)	*P* value
No	10 (2.8)	4 (1.9)	6 (3.9)		
Take drug				3.2 (1)	.07
Yes	30 (8.3)	12 (5.8)	18 (11.7)		
No	330 (91.7)	194 (94.2)	136 (88.3)		
Adjust lifestyle				2.3 (1)	.13
Yes	72 (20)	35 (17)	37 (24)		
No	288 (80)	171 (83)	117 (76)		
Rehabilitation therapy				0.1 (1)	.79
Yes	346 (96.1)	197 (95.6)	149 (96.8)		
No	14 (3.9)	9 (4.4)	5 (3.2)		
Regular reexamination				0.1 (1)	.71
Yes	348 (96.7)	198 (96.2)	150 (97.4)		
No	12 (3.3)	8 (3.8)	4 (2.6)		
Number of sequelae				0.6 (2)	.74
≤1	295 (81.9)	171 (83)	124 (80.5)		
2	41 (11.4)	23 (11.2)	18 (11.7)		
≥3	24 (6.7)	12 (5.8)	12 (7.8)		
Number of complications				3.8 (2)	15
≤1	154 (42.8)	93 (45.1)	61 (39.6)		.
2	117 (32.5)	70 (34)	47 (30.5)		
≥3	89 (24.7)	43 (20.9)	46 (29.9)		
Number of illnesses in caregivers				6.3 (3)	.10
≤1	219 (60.8)	136 (66)	83 (53.9)		
2	61 (16.9)	28 (13.6)	33 (21.4)		
3	44 (12.2)	22 (10.7)	22 (14.3)		
≥4	36 (10)	20 (9.7)	16 (10.4)		
Years of illness				6.8 (2)	.03
<10	252 (70)	154 (74.8)	98 (63.6)		
Approximately 10-19	77 (21.4)	40 (19.4)	37 (24)		
≥20	31 (8.6)	12 (5.8)	19 (12.3)		
Patient self-care ability				6.8 (3)	.08
Complete self-care	122 (33.9)	81 (39.3)	41 (26.6)		
Mild dependence	187 (51.9)	100 (48.5)	87 (56.5)		
Moderate dependence	38 (10.6)	18 (8.7)	20 (13)		
Heavily depend on	13 (3.6)	7 (3.4)	6 (3.9)		

[a]The exchange rate between the Chinese Yuan and US dollar is US $1=CN ¥7.1726. This rate is based on the midpoint of the interbank foreign exchange market on February 25, 2025, as published by the People's Bank of China.

Use of mHealth Apps Among Participants

Among the 360 participants, 206 of them had used mHealth apps, with a use rate of 57.2%. Among the caregivers who had used mHealth apps, 82.5% (n=170) caregivers used sports and health apps for monitoring steps, sleep, and obtaining relevant exercise guidance; 51.9% (n=107) caregivers used hospital-related platforms, mainly for registering, enquiring about fees, and checking examination results; 3.9% (n=8) caregivers used medical and health service apps, including learning medical knowledge, buying medicine, and checking examination results; 1% (n=2) of the caregivers used e-Medicare apps, mainly for inquiring about medical insurance information; and 1% (n=2) of the caregivers used apps connected with

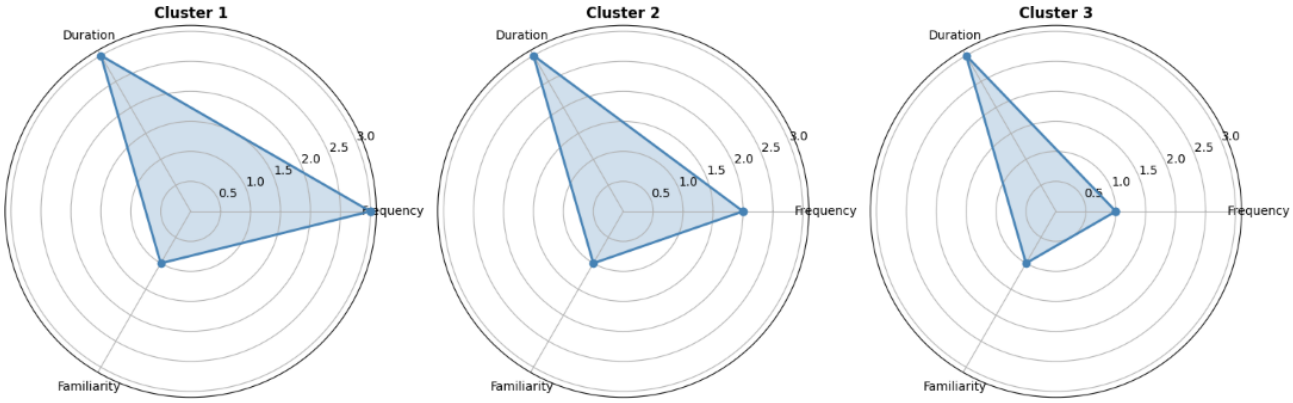
XSL•FO
RenderX

exercise equipment, for checking and analyzing the use of the equipment and adjusting the parameters.

Cluster Analysis and 3-Cluster Features

The position of the elbow in the scree plot indicates the optimal number of clusters (k=3; Figure S1 in Multimedia Appendix 1). The silhouette coefficient is 0.608, the Calinski-Harabasz index is 127.679, and the ratio of intercluster/intracluster distance is 3.15. This indicates a reasonable clustering structure with significant differences between clusters. Therefore, the final value of k is determined to be 3. Furthermore, we used a radar map to represent these features, which is shown in Figure 1. There were no significant differences in the usage duration and familiarity between the 3 clusters. However, there were notable differences in the usage frequency. Cluster 1, which included 23.8% (48/203) of caregivers who used mHealth apps, had the lowest use frequency, while cluster 3, which included 51.5% (105/203) of caregivers who used mHealth apps, had the highest use frequency (Figure S2 in Multimedia Appendix 1).

Figure 1. Radar map of the 3 clusters for characteristics.



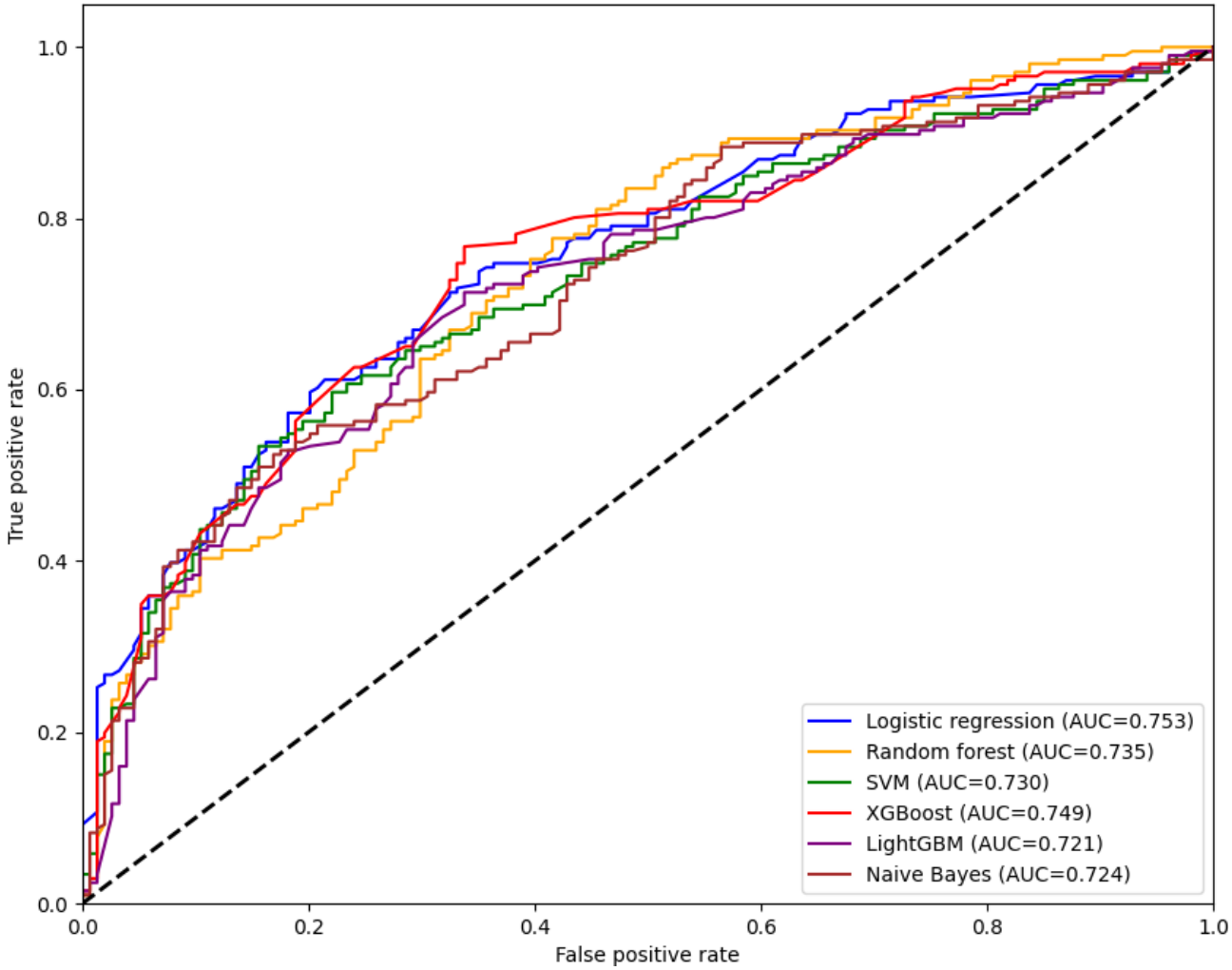
Factors Influencing Whether to Use mHealth Apps

The results of ML indicate that among the 6 models, LR demonstrates the best performance (Table S2 in Multimedia Appendix 1). Compared with other models, LR had the highest AUROC value (0.753), indicating that LR has the strongest comprehensive ability in classification decisions. Moreover, the accuracy, sensitivity, and specificity are all at a relatively high level, and they also balance sensitivity and specificity well. The calibration curve and the Brier score also suggest that LR is the best model (Figure S3 in Multimedia Appendix 1). Therefore, after considering all the indicators comprehensively, this study subsequently selects LR for interpretability analysis.

The results of the learning curve for the LR model show that the performance of the training set and the validation set is good and close, indicating that the model fits well (Figure S4 in Multimedia Appendix 1). The final selection of the LR model is 10 features, namely gender, age, marital status, education level, place of residence, whether taking medication, whether the caregiver has a disease, the duration of the patient's illness, the patient's self-care ability, and the relationship between the patient and the caregiver. The results of the sensitivity analysis also confirmed the rationality of the feature selection results (Figures S5-S6 in the Multimedia Appendix 1). Figure 2 presents the combined receiver operating characteristic curves of these models on the combined validation set.

XSL•FO
RenderX

Figure 2. Receiver operating characteristic curves for machine learning models of factors influencing the use of mobile health apps by family caregivers of patients with stroke (combined test set). AUC: area under the curve; LightGBM: light gradient boosting machine; SVM: support vector machine; XGBoost: extreme gradient boosting.



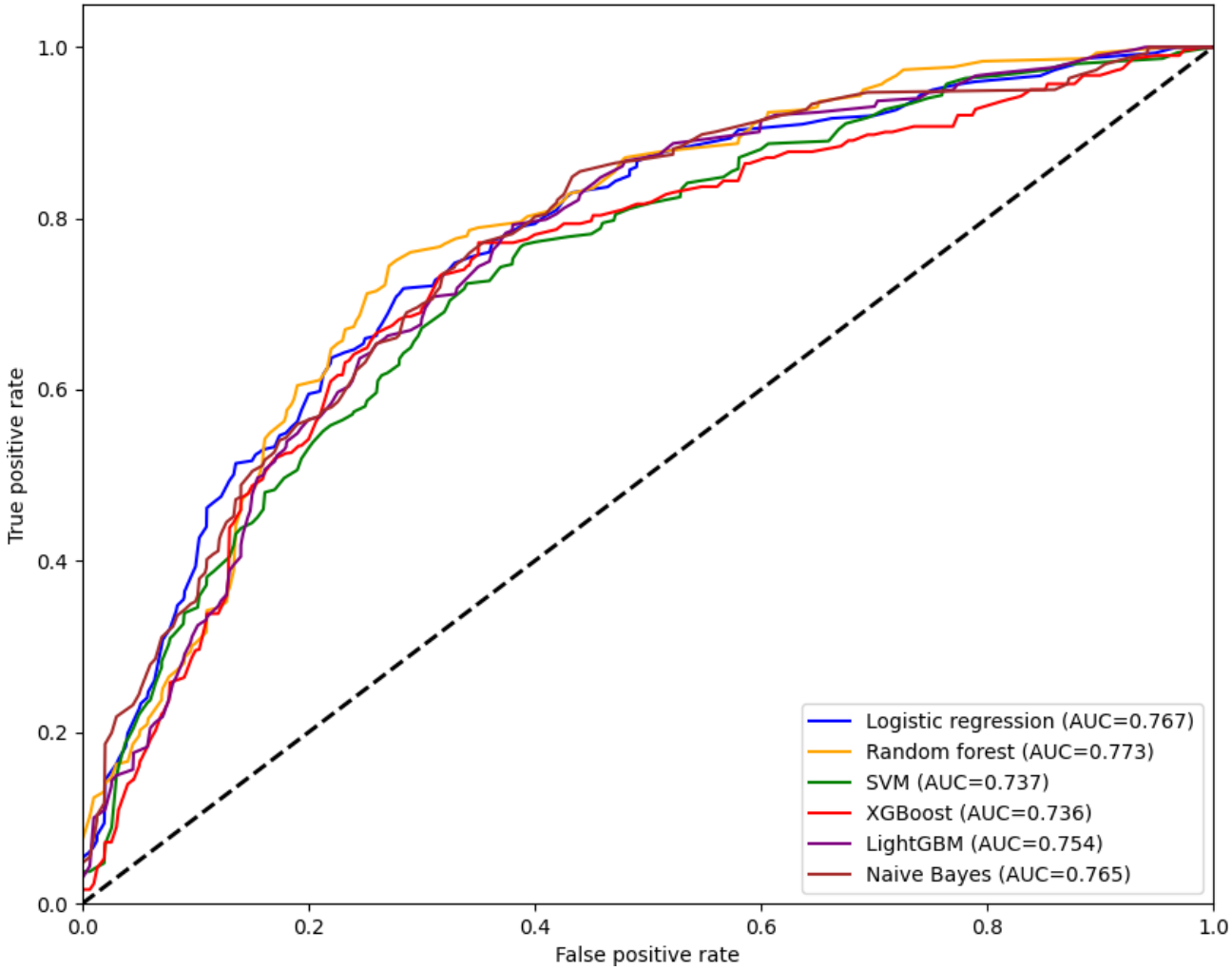
Factors Influencing mHealth App Use Behavior

The ML results show that among the 6 models, RF performed the best (Table S3 in Multimedia Appendix 1). Compared with other models, the RF model had the highest AUROC value (0.773), indicating that this model has the most stable overall ability in classifying categories. RF also has a high accuracy rate and achieves a balance between sensitivity and specificity. The calibration curve and the Brier score also suggest that RF is the best model (Figure S7 in Multimedia Appendix 1). Therefore, after considering all the indicators comprehensively, this study subsequently chose RF for interpretability analysis.

The results of the learning curve for the RF model show that the model has gradually converged, the results in the validation set tend to be stable, and there is a small difference from the training set, indicating that the model fits well (Figure S8 in Multimedia Appendix 1). The final selection of the RF model is 8 features, namely income, performance expectation, effort expectation, convenience conditions, hedonic motivation, price value, perceived risk, and habit. The results of the sensitivity analysis also confirmed the rationality of the feature selection results (Figures S9-S10 in the Multimedia Appendix 1). Figure 3 shows the combined receiver operating characteristic curve of these models on the combined test set.

XSL•FO
RenderX

Figure 3. Receiver operating characteristic (ROC) curves for a machine learning model of factors influencing mobile health (mHealth) app use behavior among family caregivers of stroke patients (combined test set). AUC: area under the curve; LightGBM: light gradient boosting machine; SVM: support vector machine; XGBoost: extreme gradient boosting.

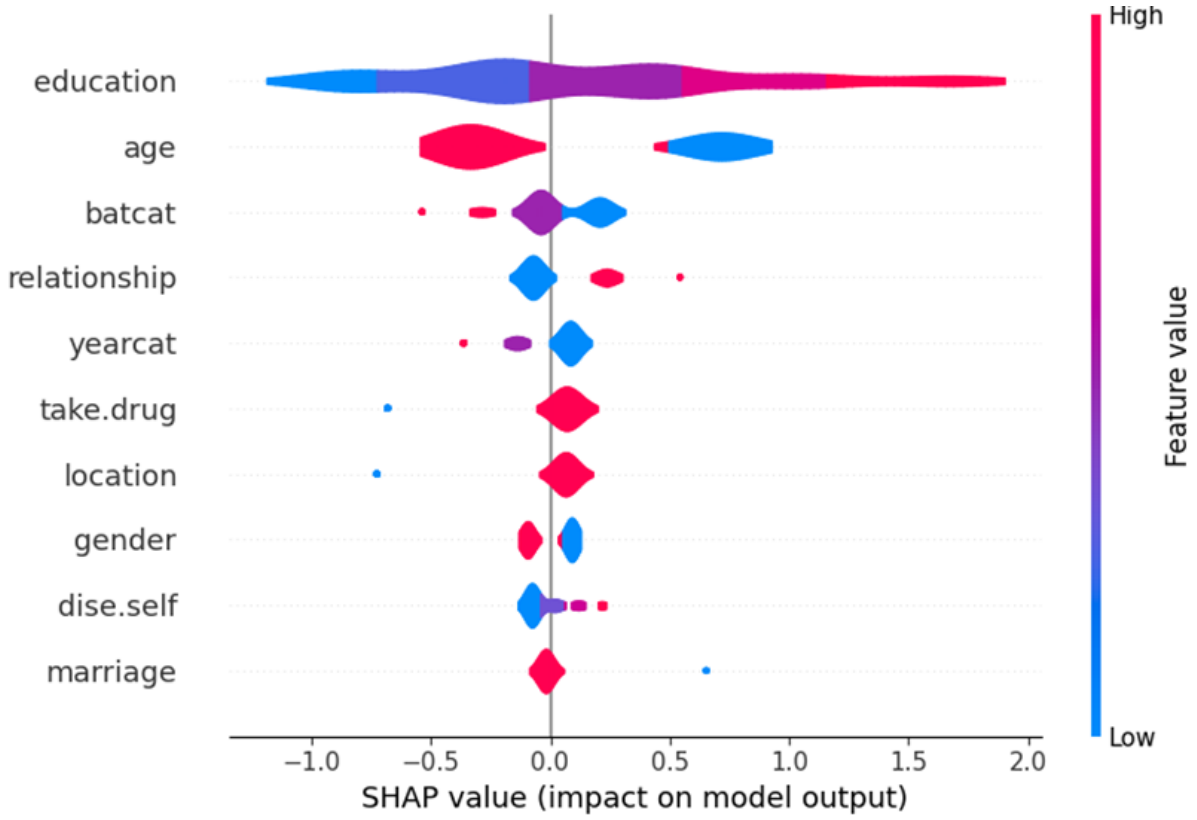


Interpretability Analysis of Whether to Use mHealth Apps (LR Model)

We used the SHAP method to interpret the LR model, calculating SHAP values to identify influential features for the overall classification model. The top 5 most influential characteristics were educational level, age, the patient's self-care ability, the relationship with the cared-for individual, and the duration of illness (Figure S11 in Multimedia Appendix 1). Figure 4 is a SHAP summary plot. The colors in the figure represent the eigenvalue sizes of each participant. Blue represents lower eigenvalues, purple represents medium eigenvalues, and red represents higher eigenvalues. The positive correlation is left blue and right red, while the negative correlation is left red and right blue [31]. The results show that a lower age of the caregiver, a higher educational level of the caregiver, and the caregiver being the patient's child will have a positive impact on the use of mHealth apps by family caregivers. On the other hand, a lower self-care ability of the patient and a shorter duration of illness will have a negative impact on the use of apps by family caregivers.

XSL•FO
RenderX

Figure 4. Shapely Additive Explanations (SHAP) summary plots of logistic regression models on whether to use mobile health (mHealth) apps.

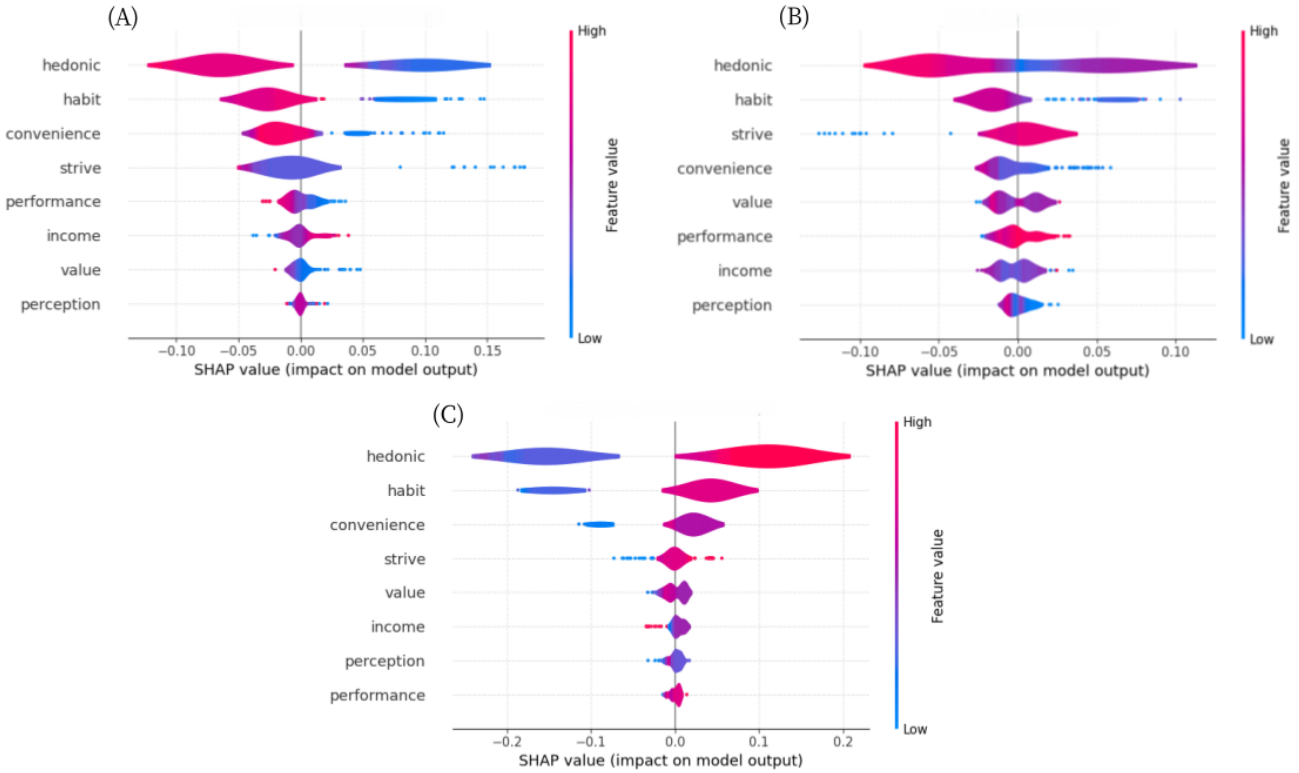


Interpretability Analysis of the Use Behavior of mHealth Apps (RF Model)

We used the SHAP approach to interpret the RF model. The top 5 most influential characteristics were hedonic motivation, habits, convenience conditions, strive expectations, and value (Figure S12 in Multimedia Appendix 1). As the use behaviors were grouped into 3 clusters, summary plots were drawn separately for each cluster to facilitate an understanding of the mechanisms of influence of the different influences within each cluster, as illustrated in Figure 5.

XSL•FO
RenderX

Figure 5. Shapely Additive Explanations (SHAP) summary plots for random forest models of factors influencing mobile health app use behavior. (A) SHAP swarm diagram of cluster 1; (B) SHAP swarm diagram of cluster 2; and (C) SHAP swarm diagram of cluster 3.



Cluster 3 is the cluster with the most ideal usage behavior among the 3 clusters (ie, the one with the highest use frequency). In order to conduct more targeted research on the influencing factors that promote the use behavior of caregivers and provide precise improvement strategies, this study will take cluster 3 as the basis and combine the research results of cluster 1 to comprehensively determine the 5 most notable influencing factors. The results show that the decline in the caregivers' enjoyment motivation, use habits, convenience conditions, effort expectations, and performance expectations had a negative impact on the use frequency.

Discussion

Overview

The study investigated the status of mHealth app use and factors influencing the use behavior of community-dwelling family caregivers of patients with stroke. We found that the usage rate of mHealth apps among family caregivers of patients with stroke in the community was 57.22%. In terms of whether to use mHealth apps, educational level, age, self-care ability, relationship with the cared-for person, and years of illness were several of the most influential factors. Hedonism, usage habits, convenient conditions, effort expectations, and performance expectations were important influencing factors of use behavior.

Family caregivers provide the core support for the health management of patients with stroke, and they have the responsibility of finding information about stroke care, rehabilitation training, and medication management [32]. An mHealth app can provide them with a source of information to help them better care for patients with stroke [33]. In addition to this, long-term caregiving can lead to emotional and psychological stress for carers [32]. Features including psychological support, stress management, and social connection integrated into mHealth apps can help to improve the psychological state of caregivers and enhance the effectiveness of caregiving. Despite the numerous benefits of mHealth apps, in reality, nearly half of the caregivers have not used any. This indicates that in promoting mHealth apps, we still face many challenges. Therefore, it is crucial to explore the factors that affect the use of mHealth apps.

Lower age and higher literacy levels of caregivers positively influenced the use of mHealth apps among family caregivers. This may occur because younger, more literate caregivers tend to show greater acceptance and familiarity with mHealth apps [34]. This positive attitude toward technology and knowledge of emerging health management tools can help promote their use of mHealth apps [35,36]. Not only that, but this group also tends to be children of patients. Previous research has found that care-dependent older adults and their adult children are important determinants of predicting and supporting the use of smart technology [37]. Our study validates this idea, which may be because younger-generation offspring are more tolerant and adaptive to new technologies and devices compared to the patient's partner or parents. In contrast, partners and parents who are relatively older and less educated are affected by cognitive decline and the deterioration of their learning and memory abilities, and hold fearful and skeptical attitudes toward new technologies such as mHealth apps [38].

A shorter duration of illness and poorer self-care ability can have a negative impact on the use of mHealth apps by caregivers. Previous studies have found that a lack of understanding of mHealth apps and insufficient time are significant obstacles to their use [39]. The results of this study

XSL•FO
RenderX

support this view. Patients with shorter diagnosis periods are in the adaptation stage, and caregivers may focus more on basic treatment rather than preventive management, thus having a lower willingness to use mHealth apps. Poorer self-care can negatively affect caregiver use of mHealth apps. Patients with poor self-care skills require a significant investment of time and energy from their caregivers. Caregivers may be busy all day long helping patients with daily living activities, such as washing, feeding, and turning, and they also need to keep an eye on changes in the patient's condition. This leaves them little free time to learn and use mHealth apps.

The study results reveal the existence of the digital divide. These vulnerable groups in the digital society cannot benefit from technological progress. If digital technology–based intervention measures fail to fully consider the obstacles faced by vulnerable groups, such as technological fears and lack of knowledge, when being designed and promoted, they may unintentionally exacerbate health inequalities [40-42]. This may lead to the groups that are most in need of support (such as low-income, high-burden, and older caregivers) being excluded from the benefits brought by digital health, resulting in "digital exclusion." To alleviate this phenomenon, in addition to designing more intuitive and user-friendly apps with lower costs, it is also necessary to explore hybrid support models (such as combining apps with the guidance of community workers or nurses), provide digital training, and develop alternative solutions with low technical barriers (such as SMS text messaging services and voice interaction), so that they can obtain digital technology services even in resource-poor environments.

Hedonic motivation emphasizes the pleasurable experience and satisfaction gained through the use of mHealth apps [43]. When carers experience pleasure, perceived caregiving stress, and reduced psychological burden through the use of mHealth apps, they tend to use these tools more frequently, thus contributing to increased frequency of use. Habit formation significantly influences decision-making processes and behavioral patterns [44]. Once a habit is formed, individuals tend to persist in that pattern of behavior because changing established habits requires more cognitive effort and willpower [45]. If caregivers have developed the habit of seeking help and accessing health information on mHealth apps whenever they encounter health problems related to the patient and themselves, this behavior becomes part of their lives. Therefore, when promoting mHealth apps, it is crucial to value and promote the formation of positive habits of use, which can help to transform the use of mHealth apps and incorporate them into the routines of daily life [46].

Convenient conditions refer to the available resources and support levels for using mHealth apps [47]. Factors such as the lack of equipment, insufficient training, and social support can all increase the barriers to using mHealth apps [48,49]. Therefore, it is suggested that when promoting mHealth apps, we should take into account external factors such as increasing facilities to improve the use behavior. Performance expectations, on the other hand, focus on carers' assessment of the expected benefits from using mHealth apps. If carers believe that using mHealth apps will significantly help them reduce the burden of caring for patients, such as receiving personalized rehabilitation advice, facilitating communication with doctors, and obtaining psychological relief and social support, their use behavior would improve in all 3 areas. Effort expectancy refers to the degree of cognitive and behavioral effort an individual perceives is required to use a particular technology [50]. If carers recognize their abilities and believe they can master and effectively use mHealth apps, they are likely to have a higher willingness to use the tools and be more motivated to use them [51].

In addition, considerations related to data privacy should also be taken seriously. mHealth apps usually need to collect sensitive medical and personal data. If these data are leaked or improperly used, it not only infringes upon personal privacy but may even lead to discrimination or psychological harm [52,53]. Therefore, developers and service providers must attach importance to and provide privacy protection mechanisms (such as anonymization processing, clear data use, and sharing agreements) to ensure that users can conveniently control their data access and deletion permissions.

Limitations

This study has several limitations that should be considered when interpreting the results. First, to maximize participant recruitment, the study used convenience sampling, which may limit sample representativeness. To enhance generalizability and verify the broad applicability of the conclusions, it is recommended that future studies adopt a more diverse and comprehensive sampling strategy to ensure that the sample covers individuals with different characteristics and backgrounds. Second, this study just focused on the overall use and the influencing factors of mHealth apps by family carers of patients with stroke and did not conduct in-depth analyses for specific types of apps. Future research could consider exploring the specific use and influencing factors of family carers of patients with stroke on various types of apps to better meet the specific needs of users. It is worth noting that although some stable analytical results have been achieved so far, the sample of this study is limited to the northern region of China. In the future, data from multicenter large samples will be further collected to conduct more in-depth and extensive research.

Although this study innovatively adopted ML algorithms for data analysis, it is necessary to acknowledge the inherent limitations of ML methods. First, despite the use of the nested cross-validation method, the risk of overfitting still cannot be completely avoided, especially in small sample datasets. Second, although models trained based on data from a single region exhibit internal stability, their external generalization ability may be limited by regional factors, thereby affecting the model's applicability. Furthermore, ensemble algorithms such as RFs, while providing highly accurate predictions, still pose challenges to their interpretability due to their "black box" decision-making characteristics. Although we incorporated SHAP analysis, the complexity of this post hoc explanation is still higher than that of models that are inherently transparent. Therefore, it is suggested that future research should conduct multicenter studies and expand the scale of the dataset to support the research of process-explainable large-scale ML methods, such as the Kolmogorov-Arnold networks.

XSL•FO
RenderX

Conclusions

This study focused on the use of mHealth apps by family caregivers of patients with stroke in the community and the factors influencing their use behavior. The research found that caregivers with advanced age, low education, shorter illness durations, and heavy care burden are vulnerable groups in digital exclusion, and explored the influencing factors of use behavior, providing a theoretical basis for further optimizing mHealth apps. The results suggest that community workers should take digitally vulnerable groups as the key support targets, actively recommend high-quality mHealth apps to them, emphasize the core value of the apps, and provide practical introductory training to help caregivers overcome initial technical fears and operational difficulties. Software developers should focus on the age-friendly design of software and carry out cost control. Under the premise of ensuring professionalism, feedback mechanisms and interactive elements added should be to enhance the pleasure and satisfaction of use and promote the formation of habits. Meanwhile, importance is attached to data privacy and security, enabling mHealth apps to truly benefit patients with stroke and their family caregivers.

Acknowledgments

The authors thank all the investigators who participated in the data collection. The authors also thank all the patients who generously responded to the questionnaires.

Funding

This study was funded by the Shandong Province Social Science Planning Project (23CSHJ13).

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials regarding the data analysis, figures, and tables involved in the research.
[DOCX File, 1132 KB - mhealth_v13i1e73903_app1.docx]

References

1. Sacco RL, Kasner SE, Broderick JP, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2013 Jul;44(7):2064-2089. [doi: 10.1161/STR.0b013e318296aeca] [Medline: 23652265]
2. Feigin VL, Stark BA, Johnson CO. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820. [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
3. Yang Y, Shi YZ, Zhang N, et al. The disability rate of 5-year post-stroke and its correlation factors: a national survey in China. PLoS ONE 2016;11(11):e0165341. [doi: 10.1371/journal.pone.0165341] [Medline: 27824877]
4. Arba F, Ali M, Quinn TJ, et al. Lacunar infarcts, depression, and anxiety symptoms one year after stroke. J Stroke Cerebrovasc Dis 2016 Apr;25(4):831-834. [doi: 10.1016/j.jstrokecerebrovasdis.2015.12.018] [Medline: 26778600]
5. Stinear CM, Lang CE, Zeiler S, Byblow WD. Advances and challenges in stroke rehabilitation. Lancet Neurol 2020 Apr;19(4):348-360. [doi: 10.1016/S1474-4422(19)30415-6] [Medline: 32004440]
6. Dan S, Zheng L. Interventions to reduce caregiving burden of family caregivers of stroke patients: a literature review. J Nurs Sci 2016;31(17):107-110. [doi: 10.3870/j.issn.1001-4152.2016.17.107]
7. Huimin Z, Jie G, Nan Z, Yanyan L, Feng R, Qin P, et al. Nursing experiences of family caregivers of stroke patients: a systematic review of qualitative studies. Chin Nurs Res 2019;33(19):3308-3324. [doi: 10.12102/j.issn.1009-6493.2019.19.005]
8. Forster A, Young J. Incidence and consequences of falls due to stroke: a systematic inquiry. BMJ 1995 Jul 8;311(6997):83-86. [doi: 10.1136/bmj.311.6997.83] [Medline: 7613406]
9. Prvu Bettger J, McCoy L, Smith EE, Fonarow GC, Schwamm LH, Peterson ED. Contemporary trends and predictors of postacute service use and routine discharge home after stroke. J Am Heart Assoc 2015 Feb 23;4(2):e001038. [doi: 10.1161/JAHA.114.001038] [Medline: 25713291]
10. Qiong D, Chence Z, Jiangmei X, Li M, Xiang L, Xiaoping Z. Current situation and influencing factors of family caregiver burden of stroke patients. J Clin Pathol 2022;42(6):1490-1495. [doi: 10.3978/j.issn.2095-6959.2022.06.035]
11. Siqin L, Xiaoyan W, Zhiqiang D, Linxue H, Rong Y. Research progress on assessment tools for the care needs of primary caregivers of stroke patients. Chin Nurs Res 2022;36(1):100-104. [doi: 10.12102/j.issn.1009-6493.2022.01.019]

XSL•FO
RenderX

12. Lutz BJ, Young ME, Creasy KR, et al. Improving stroke caregiver readiness for transition from inpatient rehabilitation to home. Gerontologist 2017 Oct 1;57(5):880-889. [doi: 10.1093/geront/gnw135] [Medline: 27816914]
13. Wu S, Wu B, Liu M, et al. Stroke in China: Advances and challenges in epidemiology, prevention, and management. Lancet Neurol 2019 Apr;18(4):394-405. [doi: 10.1016/S1474-4422(18)30500-3] [Medline: 30878104]
14. Blaya JA, Fraser HSF, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251. [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
15. Taha AR, Shehadeh M, Alshehhi A, et al. The integration of mHealth technologies in telemedicine during the COVID-19 era: a cross-sectional study. PLoS ONE 2022;17(2):e0264436. [doi: 10.1371/journal.pone.0264436] [Medline: 35202424]
16. Ming LC, Hameed MA, Lee DD, et al. Use of medical mobile applications among hospital pharmacists in Malaysia. Ther Innov Regul Sci 2016 Jul;50(4):419-426. [doi: 10.1177/2168479015624732] [Medline: 30227023]
17. An Q, Kelley MM, Hanners A, Yen PY. Sustainable development for mobile health apps using the human-centered design process. JMIR Form Res 2023 Aug 25;7:e45694. [doi: 10.2196/45694] [Medline: 37624639]
18. Cozad MJ, Crum M, Tyson H, et al. Mobile health apps for patient-centered care: review of United States rheumatoid arthritis apps for engagement and activation. JMIR Mhealth Uhealth 2022 Dec 5;10(12):e39881. [doi: 10.2196/39881] [Medline: 36469397]
19. Wu W, Graziano T, Salner A, et al. Acceptability, effectiveness, and roles of mHealth applications in supporting cancer pain self-management: integrative review. JMIR Mhealth Uhealth 2024 Jul 18;12:e53652. [doi: 10.2196/53652] [Medline: 39024567]
20. Kamal A, Khoja A, Usmani B, et al. Effect of 5-minute movies shown via a mobile phone app on risk factors and mortality after stroke in a low- to middle-income country: randomized controlled trial for the stroke caregiver dyad education intervention (Movies4Stroke). JMIR Mhealth Uhealth 2020 Jan 28;8(1):e12113. [doi: 10.2196/12113] [Medline: 32012080]
21. Yan W, Da-bin Z, Xiu-yao L, Yun-hui X, Geng-hui C. Effects of rehabilitation guidance based on “WeChat platform” on themental health status and care ability of caregivers for the stroke patients after discharge. J Community Med 2021;19(11):689-692. [doi: 10.19790/j.cnki.JCM.2021.11.10]
22. Jiehua L, Xiaodan W. Analysis framework, concept, and pathways of digital divide governance for older adults: from the perspective of digital divide and knowledge gap theory. Popul Res 2021;45(3):17-30 [FREE Full text]
23. Modzelewski KL, Stockman MC, Steenkamp DW. Rethinking the endpoints of mHealth intervention research in diabetes care. J Diabetes Sci Technol 2018 Mar;12(2):389-392. [doi: 10.1177/1932296817733007] [Medline: 28952379]
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: KDD ’16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016. [doi: 10.1145/2939672.2939785]
25. Jinsheng Z, Ming L, Liying C. Diagnostic criteria of cerebrovascular diseases in China (version 2019). Chin J Neurol 2019;52(9):710-715. [doi: 10.3760/cma.j.issn.1006-7876.2019.09.003]
26. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014 Oct;11(10):e1001744. [doi: 10.1371/journal.pmed.1001744] [Medline: 25314315]
27. Pavlou M, Ambler G, Seaman S, De Iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. Stat Med 2016 Mar 30;35(7):1159-1177. [doi: 10.1002/sim.6782] [Medline: 26514699]
28. Barthel DW, Mahony FI. Functional evaluation: The Barthel Index. Md State Med J 1965 Feb;14(1):61-65. [Medline: 14258950]
29. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4] [Medline: 31044724]
30. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019;14(11):e0224365. [doi: 10.1371/journal.pone.0224365] [Medline: 31697686]
31. Da C, Wu C, Ji Z, et al. Features influencing surface acting of different clusters of nursing students in vocational college based on interpretable machine learning: a cross-sectional study. Nurse Educ Pract 2025 Jan;82:104204. [doi: 10.1016/j.nepr.2024.104204] [Medline: 39580978]
32. Yan J, Zhu Z, Wang W, et al. Qualitative study on psychological experience of family primary caregivers of elderly stroke patients. Nurs Sci 2023;12(4):578-583. [doi: 10.12677/NS.2023.124082]
33. Haji Mukhti MI, Ibrahim MI, Tengku Ismail TA, et al. Exploring the need for mobile application in stroke management by informal caregivers: a qualitative study. Int J Environ Res Public Health 2022 Oct 10;19(19):12959. [doi: 10.3390/ijerph191912959] [Medline: 36232257]
34. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients’ adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
35. Dermody G, Fritz R, Glass C, Dunham M, Whitehead L. Family caregiver readiness to adopt smart home technology to monitor care-dependent older adults: a qualitative exploratory study. J Adv Nurs 2024 Feb;80(2):628-643. [doi: 10.1111/jan.15826] [Medline: 37614010]

XSL•FO
RenderX

36. Mendez KJW, Budhathoki C, Labrique AB, Sadak T, Tanner EK, Han HR. Factors associated with intention to adopt mHealth apps among dementia caregivers with a chronic condition: cross-sectional, correlational study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e27926. [doi: 10.2196/27926] [Medline: 34463637]
37. Demiris G. Chapter 12 - information technology to support aging: implications for caregiving. In: Kane RL, Gaugler JE, editors. Family Caregiving in the New Normal: Academic Press; 2015:211-223. [doi: 10.1016/B978-0-12-417046-9.00012-X]
38. Jingjing Y, Xu J, Weidong H. Status quo and countermeasures of self management for elderly patients with chronic diseases based on digital medical care. Chin Nurs Res 2022;36(15):2732-2735. [doi: 10.12102/j.issn.1009-6493.2022.15.019]
39. Hailiye Teferi G, Tadele MM, Tizazu G, et al. Utilization of mobile health applications and determinant factors among health science students at Debre Markos University, northwest Ethiopia in 2022. PLoS One 2023;18(7):e0275689. [doi: 10.1371/journal.pone.0275689] [Medline: 37440563]
40. Birati Y, Yefet E, Perlitz Y, Shehadeh N, Spitzer S. Cultural and digital health literacy appropriateness of app- and web-based systems designed for pregnant women with gestational diabetes mellitus: scoping review. J Med Internet Res 2022 Oct 14;24(10):e37844. [doi: 10.2196/37844] [Medline: 36240008]
41. Giebel GD, Abels C, Plescher F, et al. Problems and barriers related to the use of mHealth apps from the perspective of patients: focus group and interview study. J Med Internet Res 2024 Apr 23;26:e49982. [doi: 10.2196/49982] [Medline: 38652508]
42. Schroeder T, Haug M, Georgiou A, Seaman K, Gewald H. Evidence of how physicians and their patients adopt mHealth apps in Germany: exploratory qualitative study. JMIR Mhealth Uhealth 2024 Jan 17;12:e48345. [doi: 10.2196/48345] [Medline: 38231550]
43. Brown SA. Model of adoption of technology in households: a baseline model test and extension incorporating household life cycle. MIS Q 2005;29(3):399. [doi: 10.2307/25148690]
44. Wood W, Mazar A, Neal DT. Habits and goals in human behavior: separate but interacting systems. Perspect Psychol Sci 2022 Mar;17(2):590-605. [doi: 10.1177/1745691621994226] [Medline: 34283681]
45. Marien H, Custers R, Aarts H. Understanding the formation of human habits: an analysis of mechanisms of habitual behaviour: theory, mechanisms, change, and contexts. In: Verplanken B, editor. The Psychology of Habit: Theory, Mechanisms, Change, and Contexts: Springer Cham; 2018:51-69. [doi: 10.1007/978-3-319-97529-0_4]
46. Wu P, Zhang R, Zhu X, Liu M. Factors influencing continued usage behavior on mobile health applications. Healthcare (Basel) 2022 Jan 21;10(2):208. [doi: 10.3390/healthcare10020208] [Medline: 35206823]
47. Venkatesh V, Thong J, Xu X. Consumer acceptance and use of information technology: extending the Unified Theory of Acceptance and Use of Technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
48. Wong AKC, Wong FKY, Bayuo J, Chow KKS, Wong SM, Lee AYL. A randomized controlled trial of an mHealth application with nursing interaction to promote quality of life among community-dwelling older adults. Front Psychiatry 2022;13(978416):978416. [doi: 10.3389/fpsyt.2022.978416] [Medline: 36329920]
49. Lee T, Seohyun K, Park S. Mobile health applications for communication between caregivers of community-dwelling older adults: a scoping review. Geriatr Nurs 2023;52:172-180. [doi: 10.1016/j.gerinurse.2023.06.011] [Medline: 37356295]
50. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
51. Ghahramani F, Wang J. Intention to adopt mHealth apps among informal caregivers: cross-sectional study. JMIR Mhealth Uhealth 2021 Mar 17;9(3):e24755. [doi: 10.2196/24755] [Medline: 33729166]
52. Chow E, Virani A, Pinkney S, et al. Caregiver and youth characteristics that influence trust in digital health platforms in pediatric care: mixed methods study. J Med Internet Res 2024 Oct 28;26:e53657. [doi: 10.2196/53657] [Medline: 39467279]
53. Rezaee R, Khashayar M, Saeedinezhad S, Nasiri M, Zare S. Critical criteria and countermeasures for mobile health developers to ensure mobile health privacy and security: mixed methods study. JMIR Mhealth Uhealth 2023 Mar 2;11:e39055. [doi: 10.2196/39055] [Medline: 36862494]

Abbreviations

AUROC: area under the receiver operating characteristic
EPV: events per variable
LR: logistic regression
mHealth: mobile health
ML: machine learning
RF: random forest
SHAP: Shapley Additive Explanations

XSL•FO
RenderX

Edited by SD Bhattacharjee; submitted 13.03.25; peer-reviewed by D Vlahovic, X Ding; accepted 17.10.25; published 24.11.25.

Please cite as:
Du Y, Fan JY, Liu GZ, Yang ZY, Lei Y, Guo YF
Interpretable Machine Learning Models for Analyzing Determinants Affecting the Use of mHealth Apps Among Family Caregivers of Patients With Stroke in Chinese Communities: Cross-Sectional Survey Study
JMIR Mhealth Uhealth 2025;13:e73903
URL: https://mhealth.jmir.org/2025/1/e73903
doi:10.2196/73903

© Yun Du, Jun-Ying Fan, Guang-Zhi Liu, Zi-Yue Yang, Yang Lei, Yu-Fang Guo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Clinical Efficacy of Multimodal Exercise Telerehabilitation Based on AI for Chronic Nonspecific Low Back Pain: Randomized Controlled Trial

Chongwu Xiao[1,2,3,4,5*], MD; Yijin Zhao[1,2*], MD; Gege Li[1,2*], MS; Zhuodong Zhang[1,2], MS; Siyu Liu[6], BS; Weichao Fan[1,7], MS; Jinjing Hu[1,2], MS; Qiuru Yao[1,7], MS; Chengduan Yang[1,2], PhD; Jihua Zou[1,2,8], MS; Qing Zeng[1,2], MD; Guozhi Huang[1,2], MD

[1]Center for Rehabilitation Medicine, Zhujiang Hospital, Southern Medical University, No.253, Industrial Avenue Middle Guangzhou, Guangzhou, China

[2]School of Rehabilitation Sciences, Southern Medical University, Guangzhou, China

[3]Department of Rehabilitation Medicine, The Second Affiliated Hospital of Guangxi Medical University, Guangxi Medical University, Nanning, China

[4]GuangDong Engineering Technology Research Center of Brain Function Assessment and Neuroregulation Rehabilitation, Guangzhou, China

[5]Institute of Exercise and Rehabilitation Science, Zhujiang Hospital, Southern Medical University, Guangzhou, China

[6]School of Sport Medicine and Physical Therapy, Beijing Sport University, Beijing, China

[7]School of Nursing, Southern Medical University, Guangzhou, China

[8]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China

[*]these authors contributed equally

Corresponding Author:
Guozhi Huang, MD
Center for Rehabilitation Medicine, Zhujiang Hospital, Southern Medical University, No.253, Industrial Avenue Middle Guangzhou, Guangzhou, China

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2025/1/e78188

Abstract

Background: Exercise therapy is strongly recommended as a treatment for chronic nonspecific low back pain (CNSLBP). However, therapist-guided exercise therapy requires significant medical resources. Ordinary digital telerehabilitation affects efficacy due to a lack of guidance and dynamic support. Artificial intelligence (AI)–assisted interactive health promotion systems may solve these problems.

Objective: We aimed to explore whether AI-assisted multimodal exercise telerehabilitation is superior to conventional telerehabilitation in the treatment of people with CNSLBP.

Methods: This study was a prospective, double-arm, open-label, randomized clinical controlled trial. People with CNSLBP were randomly allocated to either the AI or video group, receiving AI-assisted multimodal exercise therapy or conventional video guidance, respectively, via a WeChat application add-in. The multimodal exercise consisted of deep core muscle, flexibility, Mackenzie, and breathing exercises. The exercises were performed for 30 - 45 minutes per session, 3 times a week, for 4 weeks. Participants underwent face-to-face assessment at baseline and week 4, and web-based assessment at weeks 2 and 8. The primary outcome was the change in Numerical Rating Scale (NRS) relative to baseline at week 4. Secondary outcomes included changes in the Roland-Morris Disability Questionnaire (RMDQ), Oswestry Disability Index (ODI), Pain Castastrophizing Scale (PCS), Timed Up-and-Go (TUG) test, and thickness of the transverse abdominus (TrA) and multifidus (MF) muscles relative to baseline at week 4. Generalized estimating equation and covariance were used to examine the efficacy of the interventions.

Results: A total of 38 participants (19 participants per group) were recruited. Eighteen participants in the AI group and 16 participants in the video group completed and were included in the final analysis. There was a significant difference in NRS at week 4 between the AI group and video group (most severe NRS: −3.00 vs −1.50; adjusted mean difference −1.08, 95% CI −1.68 to −0.49; P<.001; mean NRS: −2.61 vs −1.62; adjusted mean difference −0.67, 95% CI −1.19 to −0.15; P=.01). The difference in most severe NRS persisted until week 8 (−3.06 vs −1.69; adjusted mean difference −0.95, 95% CI −1.73 to −0.18; P=.02). Compared with the video group at week 4, the AI group showed significant improvement in secondary outcomes, including RMDQ, PCS, and core muscle thickness of left TrA, right TrA, left MF, and right MF.

XSL•FO
RenderX

Conclusions: We showed that 4 weeks of telerehabilitation based on AI-assisted multimodal exercise has better therapeutic effects compared to conventional exercise telerehabilitation in people with CNSLBP. This study provides guidance for developing effective real-time home-based exercise therapies for people with CNSLBP, which may help reduce economic and human resource costs associated with treatment.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300073185; https://www.chictr.org.cn/showproj.html?proj=198413

(JMIR Mhealth Uhealth 2025;13:e56176) doi:10.2196/56176

KEYWORDS

low back pain; artificial intelligence; exercise; telerehabilitation

Introduction

Low back pain (LBP) is usually defined as pain between the costal margins and the subgluteal fold region [1], is one of the most common health problems worldwide, and causes a high socioeconomic impact [2]. The probability of developing LBP in a person's lifetime is as high as 75% - 84% [3]. Pain lasting for more than 3 months is called chronic LBP (CLBP).

Studies have shown that 80% - 90% of CLBP is nonspecific and incurable without any specific pathological cause, which is called chronic nonspecific LBP (CNSLBP) [4]. In China, the incidence of CNSLBP is high. In the Departments of Orthopedics, Rehabilitation Medicine, and Pain Medicine, CNSLBP is the most common disease, which accounts for one-third of daily outpatient visits, second only to upper respiratory tract infections [5]. The high incidence makes a huge burden on the health care budget, which is greater than that of coronary heart disease, diabetes, arthritis, and cerebrovascular disease [6]. Thus, it is necessary to conduct in-depth research on CNSLBP and to determine improved treatment methods.

There are many current treatments for CNSLBP. Nonpharmacological therapies are the first-line treatment for CLBP [1]. Among these therapies, high-quality evidence suggests that exercise therapy is more effective in reducing pain compared to no treatment or other conservative treatments [7]. Exercise therapy, as an effective means of treating LBP, is not only effective but also cost-effective compared to other medical methods [8]. Therefore, clinical guidelines strongly recommend exercise therapy for people with CLBP [9].

There are various forms of exercise therapy, such as core stability training, Pilates, McKenzie exercises, breathing training, trunk muscle strength training, suspension training, and Tai Chi. To improve the effectiveness of exercise, specialists usually recommend using a combination of different conservative treatments [10]. In exercise therapy, the combination of flexibility and strength training (the basic composition of multimodal exercise) has been shown to be the most effective intervention for people with CLBP [11,12].

Exercise therapy can be practiced in medical institutions or at home. However, due to insufficient resources, it is difficult to carry out exercise therapy in medical institutions on a large scale. In most cases, exercise therapy is conducted through home-based training, which leads to problems such as a lack of guidance and dynamic support and difficulties in contacting care providers [13]. This reduces the effectiveness of this therapy. Therefore, exploring new home-based multimodal exercise training methods, which can offer feedback and guidance, is necessary.

The effectiveness of application-based therapy has been proven in exploring new home-based multimodal exercise training methods [14]. Using application-based therapy as a therapy supplement could help promote the implementation of home-based exercise protocols. The application features include exercise tracking of prefabricated or adaptable workout programs, educational aspects, artificial intelligence (AI)–based therapy, or workout programs, and motion detection [15]. In terms of AI-based applications, most previous research has focused on pushing personalized self-management to people with LBP [16-18], but this has not always achieved better therapeutic outcomes compared to conventional clinical care [17,18]. Few AI-based applications focused on motion detection and guidance. With the development of technology, AI human key point identification technology can accurately determine body surface key points and guide individuals during movement, which has been validated in knee or hip osteoarthritis [19]. However, there is no research on applying this technology to exercise therapy for CNSLBP.

Characterized by real-time guidance, simplification of instructions, and convenience of doctor-patient communication, multimodal exercise telerehabilitation based on AI human key point identification technology may improve the effectiveness of exercise therapy. However, studies regarding home-based multimodal exercise therapy based on AI human key point identification technology among people with CNSLBP are lacking. We hypothesized that AI-assisted exercise therapy would have a positive effect on therapeutic efficacy. We, therefore, aimed to explore the effects of AI-assisted multimodal exercise on pain intensity, body function, psychology, and the core muscles in CNSLBP telerehabilitation.

Methods

Trial Design

This was a prospective, double-arm, open-label, randomized controlled clinical study conducted in Guangzhou, China, from March to October 2023 at Zhujiang Hospital of Southern Medical University. All the participants were randomly assigned to either the AI-assisted exercise therapy group (AI group) or the video-assisted exercise therapy group (video group).

Ethical Considerations

Protocol approval was obtained from the Ethics Committee of Zhujiang Hospital of Southern Medical University (number

XSL•FO
RenderX

2023-KY-017). The study was prospectively registered in the Chinese Clinical Trial Registry (registration number ChiCTR2300073185) and conducted in compliance with local policies and regulations. All participants were provided with an informed consent form based on the principles of the Helsinki Declaration, which included comprehensive details of the study. Participants were informed that they could withdraw from the study at any time, and the form also included clarifications on relevant issues to ensure they were fully informed. To protect privacy, the data of participants were anonymized using letter codes, and only the principal research team had access to this data. Participants were explicitly informed that their data would only be used for anonymous article publication. In the event of any adverse reactions, participants could contact the researchers through communication tools at any time. The researchers would then assess the situation to determine whether the participant should be withdrawn from the study and provided with alternative treatments, such as physical therapy or analgesics.

Study Population

The sample size was determined by G*Power software (version 3.1.9.2; Kiel University). The effect size was determined based on a previous study using an AI-based application for CLBP exercise intervention [18]. In this study, the mean Numerical Rating Scale (NRS) of the exercise group decreased by 1.1 (SD 0.3) points. The mean NRS of conventional group decreased by 0.9 (SD 0.4) points. The average SD was 0.35. The effect size was calculated to be 0.57. The correlation between repeated measurements was set as 0.5. There were 2 groups with 4 measurements. With a statistical power of 0.95 and an α level of 0.05, the total sample size was calculated to be 28. Considering a 20% shedding rate, the recruitment target was at least 36 participants.

Inclusion criteria followed the nonspecific LBP diagnosis guidelines of the American College of Physicians and the American Pain Society [20]. Participants who met the following criteria were included in the study: (1) clinical diagnosis of nonspecific LBP or discomfort for >3 months, with a mean NRS≥3 points; (2) age 18 to 75 years; (3) right-handed; and (4) possession of a smartphone and the skill to operate WeChat. The key exclusion criteria were as follows: (1) pregnancy; (2) a history of waist trauma or waist or abdominal surgery in the past 2 years; (3) a history of nerve root symptoms, spine fracture, infection, lumbar malignancy, or LBP caused by any other disease; and (4) participants suffering from hypertension, heart disease, Parkinson's disease, and other conditions that were not suitable for intense exercise. Participants were recruited through the pain management clinic of Zhujiang Hospital, WeChat friend circles, and recruitment posters.

All interested patients with LBP who presented for consultation were provided with basic information about the study, and a preliminary screening questionnaire was completed. The number of participants who were invited to participate, refused to participate, and were not eligible (and their related reasons) were recorded by a research assistant.

Randomization and Blinding

Using a random number sequence generated by SPSS 20.0 software (IBM), all eligible participants were randomly assigned to the AI group or video group in a 1:1 ratio by a research assistant who was not involved in the assessments and treatment. Random numbers were hidden in opaque envelopes. Each envelope was successively numbered, and the screening number was attached to the surface. The research assistant opened the envelope to assign participants to the corresponding groups.

There were 6 people in the study team. A research assistant was responsible for the recruitment and another for the random assignment of participants. Two therapists with many years of treatment experience were responsible for the intervention exercises of the 2 groups. A rehabilitation physician was specifically responsible for the assessment of the outcomes. A statistical analyst was responsible for the statistical analysis. The statistical analyst and rehabilitation physician were blind to group allocation. As adequate informed consent was required, all participants were informed about what exercise intervention they would receive. Thus, this study was an open-label study.

Intervention Measures

The intervention measures included 2 parts: patient education and exercise therapy [21]. Through the form of community management and graphic publicity, patient education could improve patient enthusiasm for exercise and provide participants with rehabilitation knowledge about correct posture and reasonable exercise modes. The multimodal exercises included deep core muscle [11,22], flexibility [23], Mackenzie [24], and breathing exercises [25] to improve patient core strength and increase joint flexibility, as shown in Multimedia Appendix 1.

The telerehabilitation of participants was realized through an application called Yun-fu (Beijing Yinshan Future Health Technology Co, Ltd.), a WeChat application add-in used for smartphones and tablets. The application add-in was used to collect patient evaluation information and complete the exercise training.

The angle and length ratios among the 30 key points of the human body were identified by AI for all the exercise movements involved, and were set in the application add-in to complete the standards and common errors of the movements. The specific procedures for participants using the application add-in were as follows. Participants opened the application add-in by scanning a QR code and completed the self-assessment, which included the basic information of CLBP, Oswestry Disability Index (ODI) questionnaire, STarT back questionnaire, Generalized Anxiety Disorder questionnaire, and the AI-assisted acquisition of lumbar spine mobility and core muscle endurance. According to the self-assessment results, the application add-in classified participants into 5 categories of LBP based on treatment-based classifications [26], including flexion intolerance, extension or rotation intolerance, stability deficiency, muscle tension, and nerve compression. Finally, the application add-in sent each patient exercise plans of 30 - 45 minutes per session, 3 times a week, for a duration of 4 weeks [27]. In order to ensure consistency in the intervention exercise prescription between the 2 groups, the subjects included in this

XSL•FO
RenderX

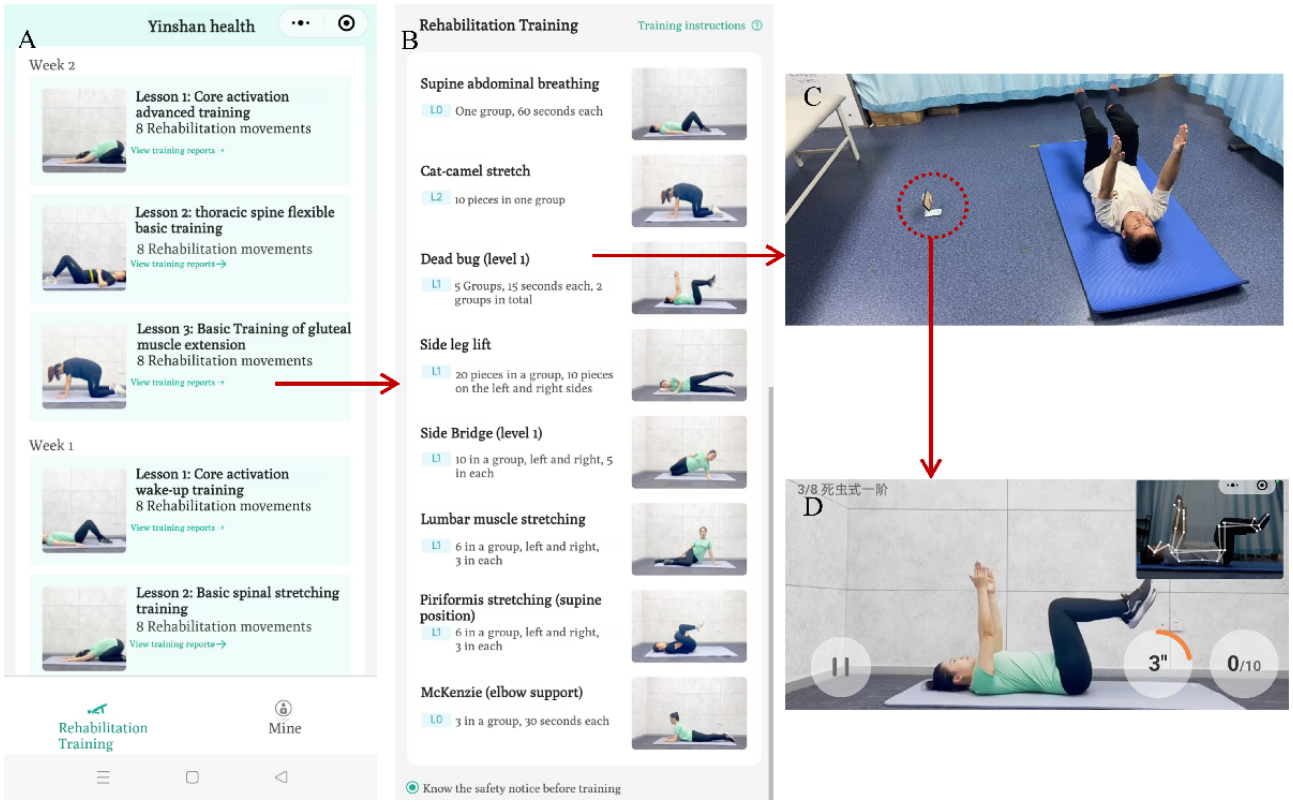
study were all subjects with LBP classified as stability deficiency.

When the participants wanted to exercise, they simply needed to open the application add-in in WeChat (Figure 1A), select the corresponding prescription course (Figure 1B), stand or lie on the yoga mat, and place the phone on a phone stand 1.5 m away (Figure 1C). The application add-in would automatically broadcast the training content and movement points, and movement video demonstrations were presented on the screen. Furthermore, an AI recognition box in the upper right corner of the screen automatically recognized the movements and provided real-time guidance to ensure standardization of movements (Figure 1D). When a movement was complete, the application add-in immediately provided an evaluation and prompt, including “qualified,” “good,” “excellent,” and “super excellent,” to motivate participants to complete the following exercise.

After exercising, the application add-in sent a rating of perceived exertion and training reports. Participants then posted a screenshot of the training report in a prejoined WeChat group and received incentives from other participants and the rehabilitation therapists in the group. At the beginning of each week, participants received video and graphic education about CLBP from the therapists, including correct posture, pain management, causes of CLBP, and lumbar spine structure. At 2 and 4 weeks of exercise, participants completed the same initial assessment in the application add-in, and their program was appropriately adjusted.

The video group received the same education, evaluation, and exercise prescription as the AI group. The only difference was that, during the exercise process, the video group completed the exercise by watching a training video, and the AI group received training video and AI guidance during the process. The groups did not receive additional treatments throughout the entire study period.

Figure 1. Example of artificial intelligence (AI)–assisted exercise in the Yunfu application add-in. (A) Weekly course schedule; (B) specific exercise prescriptions for each course; (C) scene arrangement during exercise; and (D) video and AI guidance on the mobile screen.



Outcomes and Assessments

Demographic information, including age, gender, BMI, years of education, and duration of pain (<6 mo, 6 - 12 mo, >12 mo), was obtained using self-reported questionnaires. To describe the pain changes of participants more comprehensively, pain (current, mean, and most severe in the past week) was measured with the NRS [28]. The primary outcome of the study was the change in NRS relative to baseline at week 4. The secondary outcomes were changes in NRS at week 8, and the scores of ODI, Roland-Morris Disability Questionnaire (RMDQ), Pain Catastrophizing Scale (PCS), time of Timed Up-and-Go (TUG) test, and thickness of core muscles (transverse abdominus [TrA] and multifidus [MF]) at week 4, relative to baseline. The baseline and week 4 data were collected face-to-face in the laboratory. The remaining data were collected through web-based questionnaires.

The NRS uses a scale with numbers 0-10 to quantify pain intensity, with a higher number representing a greater degree of pain. A score of 0 indicates no pain; 1 - 3, mild pain; 4 - 6, moderate pain; and 7 - 10, severe pain [29]. The RMDQ was developed to evaluate low back function with good reliability and validity [30]. The questionnaire has 24 options with a total score of 24 points. The higher the score, the greater the patient’s functional limitations. The ODI, comprising 10 questions

XSL•FO
RenderX

evaluating pain intensity, daily activities, lifting, walking, sitting, standing, sleep, sexual life, social life, and travel, was used to assess the disability of participants. Each question has 6 options, corresponding to 0-5 points. Higher points represent more severe dysfunction [31]. The PCS was used to assess the catastrophic pain thoughts. The questionnaire consists of 13 items, rated on a 5-point scale (with 0=never, 1=almost never, 2=sometimes, 3=fairly often, and 4=very often). Higher scores reflect higher pain catastrophizing [32].

The TUG test was used to assess physical function. Participants began by sitting in a standard armchair with a height of approximately 46 cm. They were then asked to stand up from the armchair, walk 3 m at a comfortable speed, then immediately turn around, walk back to the armchair, and sit down again. The time was recorded between when they left the armchair and when they returned to a seated position [33]. The test was performed 3 times, and the results were averaged.

The muscle thickness of TrA and MF was measured by using a Terason uSmart 3300 ultrasound (Terason). TrA and MF are considered to be the most critical muscles in maintaining lumbar spine stability and are closely related to the development of LBP [34,35]. For the thickness of TrA, participants lay in the supine position with hips flexed to approximately 135° and knees flexed to 90°. They were then instructed to take a deep breath, exhale fully, and maintain a relaxed abdomen for 5 seconds. A linear probe was used to measure the thickness above the iliac crest in the anterior axillary line [36]. Muscle thickness was recorded as the distance between the upper and lower hyperechoic myofascial. For the thickness of MF, participants lay in the prone position, with a thin pillow under the abdomen to straighten the lumbar spine. The head of the participant was positioned in the treatment bed hole. A curved probe was used to measure the thickness at the fourth lumbar vertebra (L4). Muscle thickness was recorded as the distance from the tip of the articular process to the lower margin of subcutaneous fat [37]. The measurement was performed on the left and right sides, 3 times, with the average used for data analysis.

Statistical Analysis

The mean (SD) or median (interquartile range) was used to describe continuous variables, whereas numbers (percentages) were used for categorical variables. All analyses were conducted based on the per-protocol principle. An independent *t*-test, Wilcoxon rank-sum test, or Fisher's exact test was used to test the differences of baseline characteristics between groups, as appropriate. Normal distribution was examined by Shapiro-Wilk test and QQ plots. Analyses of variables changing from baseline at week 2 and week 4 (most severe NRS, mean NRS, current NRS, ODI, RMDQ, and PCS) were performed by generalized estimating equations and adjusted for the respective baseline value. Analyses of variables changing from baseline at week 8 (most severe NRS, mean NRS, and current NRS) were performed by generalized estimating equations and adjusted for the respective baseline value. Variables changing from baseline at week 4 (TUG and muscle thickness) were tested by covariance analysis adjusted for the respective baseline value. The significance level was set at $P<.05$ for all statistical tests. All analyses were performed in R (version 4.3.1; R Foundation for Statistical Computing) and SPSS (version 20.0; IBM).

Results

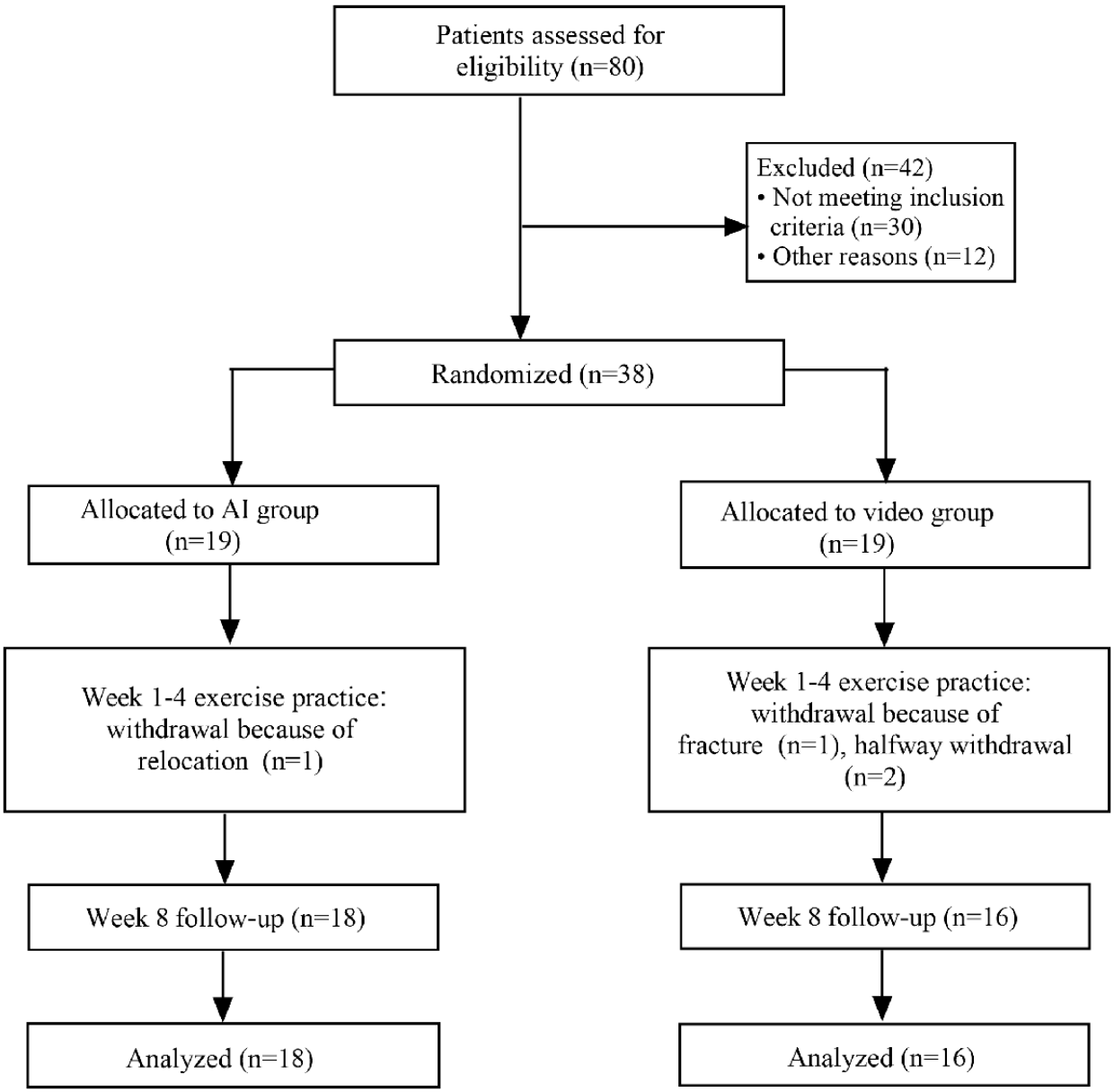
Study Population

As shown in Figure 2, a total of 38 participants with CNSLBP were randomly assigned to either the AI or video group. Finally, 34 participants completed the intervention and follow-up (n=18, 47% analyzed for efficacy in the AI group; n=16, 42% analyzed for efficacy in the video group; Checklist 1).

Table 1 shows the baseline characteristics of participants included in the analysis. The mean age of the participants was 28.9 (SD 9.1) years and 29.3 (SD 7.4) years for the video and AI groups, respectively. Both groups were predominantly female (n=14 (88%) for the video group; n=12 (67%) for the AI group). The education duration of the video group was 17.0 (IQR 16.0-17.0) years, which was similar to the AI group with 16.0 (IQR 16.0-18.0) years. The 2 groups also had similar BMI results (21.0 (SD 2.9) kg/m^2 for the video group; 21.3 (SD 2.4) kg/m^2 for the AI group). A pain duration of 32.6 (SD 27.8) months was reported for the participants in the video group, which was less than that of the AI group, a pain duration of 53.3 (SD 61.2) months. These demographic results and baseline measurement data, including NRS, RMDQ, ODI, PCS, TUG, and thickness of core muscles (TrA and MF), showed no significant differences between the 2 groups.

XSL•FO
RenderX

Figure 2. Flowchart of included participants. AI, artificial intelligence.



XSL•FO
RenderX

Table . Baseline characteristics.

Characteristic	Statistic	Video group (n=16)	AI[a] group (n=18)	*P* value
Age (years)	Mean (SD)	28.9 (9.1)	29.3 (7.4)	.91
Gender				.23
Male	n (%)	2 (13%)	6 (33%)	
Female	n (%)	14 (88%)	12 (67%)	
Education (years)	Median (IQR)	17.0 (16.0-17.0)	16.0 (16.0-18.0)	.91
BMI (kg/m^2)	Mean (SD)	21.0 (2.9)	21.3 (2.4)	.78
Pain duration (months)	Mean (SD)	32.6 (27.8)	53.3 (61.2)	.21
Most severe NRS[b]	Median (IQR)	4.0 (4.0-5.0)	5.0 (4.0-5.8)	.15
Mean NRS	Median (IQR)	4.0 (3.0-4.0)	4.0 (3.0-5.0)	.21
Current NRS	Median (IQR)	3.0 (3.0-4.0)	3.0 (3.0-3.8)	.83
ODI[c]	Mean (SD)	7.8 (2.8)	8.4 (3.1)	.50
RMDQ[d]	Mean (SD)	4.5 (2.0)	5.1 (1.7)	.39
PCS[e]	Mean (SD)	13.9 (9.1)	18.6 (9.1)	.14
Timed up-and-go (s)	Mean (SD)	10.2 (1.2)	10.4 (1.1)	.69
LTrA[f] thickness (mm)	Mean (SD)	3.8 (0.5)	4.1 (0.8)	.19
RTrA[g] thickness (mm)	Mean (SD)	3.9 (0.6)	4.0 (0.6)	.56
LMF[h] thickness (mm)	Mean (SD)	26.8 (3.2)	27.4 (4.1)	.67
RMF[i] thickness (mm)	Mean (SD)	25.7 (3.3)	26.6 (4.4)	.50

[a]AI, artificial intelligence.
[b]NRS, Numerical Rating Scale.
[c]ODI, Oswestry Disability Index.
[d]RMDQ, Roland-Morris Disability Questionnaire.
[e]PCS, Pain Catastrophizing Scale.
[f]LTrA, left transverse abdominal muscle.
[g]RTrA, right transverse abdominal muscle.
[h]LMF, left multifidus muscle.
[i]RMF, right multifidus muscle.

Changes in NRS

At week 4, the mean change (SD) in the most severe NRS in the AI group and video group was −3.00 (SD 1.41) and −1.50 (SD 1.03), respectively (adjusted mean difference −1.08, 95% CI −1.68 to −0.49; *P*<.001). The mean change in mean NRS in the AI group and video group was −2.61 (SD 0.92) and −1.62 (SD 0.96), respectively (adjusted mean difference −0.67, 95% CI −1.19 to −0.15; *P*=.01). There was no significant difference between the AI group (mean −2.17, SD 1.29) and video group (mean −2.19, SD 1.22; adjusted mean difference −0.07, 95% CI −0.71 to 0.57; *P*=.83) in current NRS. At week 2, no significant difference in the reduction of NRS (most severe, mean, and current) was observed between the 2 groups. At follow-up of week 8, the change in the most severe NRS in the AI group and video group was −3.06, SD 1.51, and −1.69, SD 1.54, respectively (adjusted mean difference −0.95, 95% CI −1.73 to −0.18; *P*=.02). No significant difference was observed in mean NRS and current NRS between the 2 groups (Table 2). The 3-dimensional NRS changes between the AI group and the video group during the study period are shown in Figure 3.

XSL•FO
RenderX

Table . Primary and secondary outcomes change from baseline.

Outcome	AI[a] group, mean (SD)	Video group, mean (SD)	Adjusted mean difference (95% CI)	*P* value[b]
Most severe NRS[c]				
Week 2	−2.06 (1.47)	−1.31 (1.30)	−0.33 (−1.05 to 0.40)	.38
Week 4	−3.00 (1.41)	−1.50 (1.03)	−1.08 (−1.68 to −0.49)	<.001
Week 8	−3.06 (1.51)	−1.69 (1.54)	−0.95 (−1.73 to −0.18)	.02
Mean NRS				
Week 2	−1.67 (1.19)	−1.69 (0.79)	0.34 (−0.19 to 0.87)	.20
Week 4	−2.61 (0.92)	−1.62 (0.96)	−0.67 (−1.19 to −0.15)	.01
Week 8	−2.50 (1.10)	−1.50 (1.55)	−0.68 (−1.53 to 0.17)	.12
Current NRS				
Week 2	−1.22 (1.63)	−1.62 (1.15)	0.31 (−0.34 to 0.96)	.35
Week 4	−2.17 (1.29)	−2.19 (1.22)	−0.07 (−0.71 to 0.57)	.83
Week 8	−2.06 (1.11)	−1.88 (1.86)	−0.27 (−1.15 to 0.60)	.54
Timed up-and-go (s)				
Week 4	−1.05 (0.78)	−0.85 (0.86)	−0.15 (-0.65 to 0.35)	.55
Left TrA[d] thickness (mm)				
Week 4	0.30 (0.45)	−0.14 (0.31)	0.47 (0.19 to 0.75)	.002
Right TrA thickness (mm)				
Week 4	0.57 (0.41)	−0.09 (0.47)	0.69 (0.39 to 0.99)	<.001
Left MF[e] thickness (mm)				
Week 4	0.90 (1.15)	−0.01 (1.11)	0.94 (0.14 to 1.74)	.02
Right MF thickness (mm)				
Week 4	0.84 (0.85)	−0.03 (0.98)	0.92 (0.27 to 1.56)	.007
ODI[f]				
Week 2	−2.06 (2.21)	−1.56 (3.16)	−0.16 (−1.66 to 1.34)	.83
Week 4	−3.67 (2.59)	−2.25 (1.69)	−1.09 (−2.26 to 0.09)	.07
RMDQ[g]				
Week 2	−2.39 (2.23)	−2.00 (1.71)	−0.01 (−0.98 to 0.96)	.99
Week 4	−3.17 (1.82)	−1.88 (1.02)	−0.91 (−1.48 to −0.34)	.002
PSC[h]				
Week 2	−4.33 (7.77)	−1.12 (6.83)	−1.65 (−6.14 to 2.84)	.47
Week 4	−9.06 (6.80)	−3.56 (6.15)	−3.94 (−7.69 to −0.19)	.04

[a]AI, artificial intelligence.

[b]The significance level was set at $P<.05$

[c]NRS, Numerical Rating Scale.

[d]TrA, transverse abdominal muscle.

[e]MF, multifidus muscle.

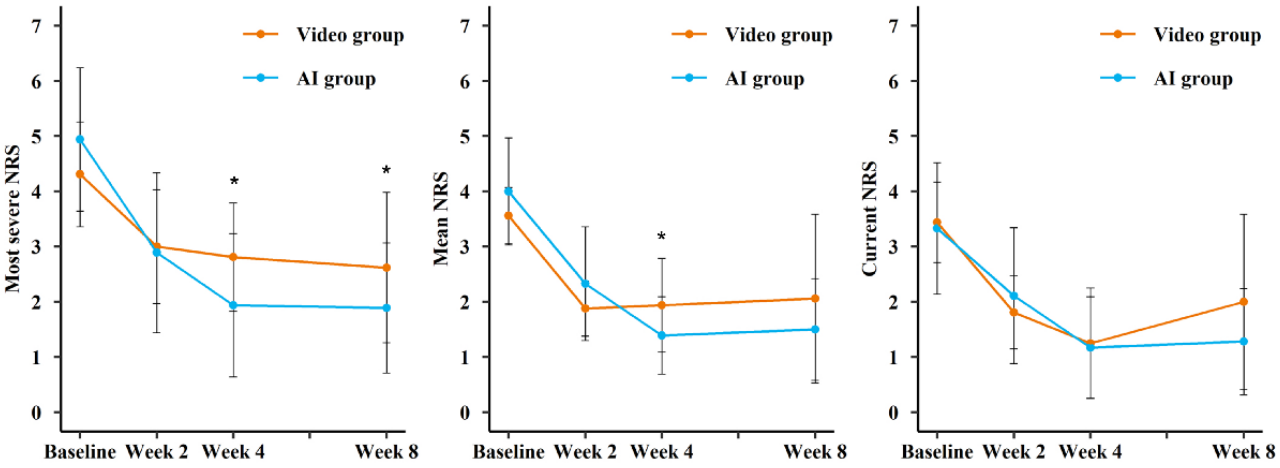
[f]ODI, Oswestry Disability Index.

[g]RMDQ, Roland-Morris Disability Questionnaire.

[h]PCS, Pain Catastrophizing Scale. Each outcome variable was analyzed after adjusting the baseline value.

XSL•FO
RenderX

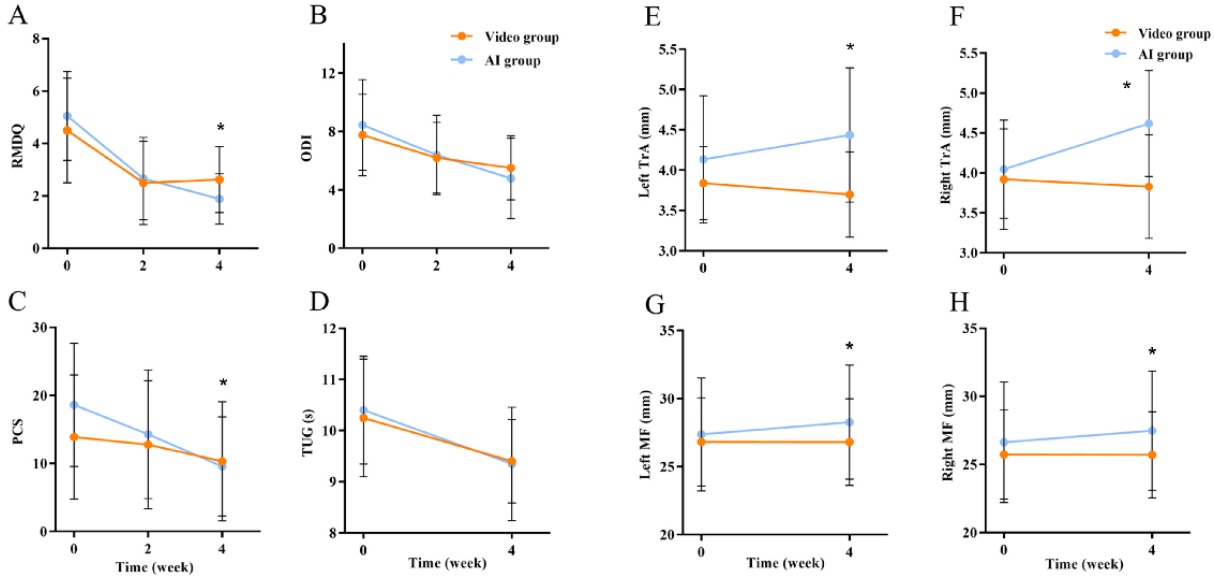
Figure 3. Numerical Rating Scale (NRS) at baseline and changes in 2 weeks, 4 weeks, and 4 weeks of follow-up. AI, artificial intelligence.



Changes in RMDQ

Compared with the baseline, the RMDQ scores of the AI group and video group showed a downward trend at week 2 and week 4 (Figure 4A). A significant difference was observed in the AI group (mean −3.17, SD 1.82) compared with that in the video group (mean −1.88, SD 1.02; adjusted mean difference −0.91, 95% CI −1.48 to −0.34; *P*=.002) at week 4. There was no significant difference between the AI group (mean −2.39, SD 2.23) and video group (mean −2.00, SD 1.71; adjusted mean difference −0.01, 95% CI −0.98 to 0.96; *P*=.99) at week 2 (Table 2).

Figure 4. Secondary outcomes in baseline and changes from baseline to 4 weeks. (A) RMDQ (Roland-Morris Disability Questionnaire); (B) ODI (Oswestry Disability Index); (C) PCS (Pain Catastrophizing Scale); (D) TUG (Timed Up-and-Go); (E) left TrA (transverse abdominus); (F) right TrA; (G) left MF (multifidus); (H) right MF.



Changes in ODI

The ODI scores of the AI group and video group showed a downward trend at week 2 and week 4 compared with the baseline (Figure 4B). At week 4, no significant difference was observed in the AI group (mean −3.67, SD 2.59) and video group (mean −2.25, SD 1.69); adjusted mean difference −1.09, 95% CI −2.26 to 0.09; *P*=.07). Similarly, there was no significant difference between the AI group (mean −0.26, SD 2.21) and video group (mean −1.56, SD 3.16); adjusted mean difference −0.16, 95% CI −1.66 to 1.34; *P*=.83) at week 2 (Table 2).

Changes in PCS

An improvement of PCS was observed at week 2 and week 4 compared with the baseline (Figure 4C). There was a significant difference between the AI group (mean −9.06, SD 6.80) and video group (mean −3.56, SD 6.15; adjusted mean difference −3.94, 95% CI −7.69 to −0.19; *P*=.04) at week 4. However, no significant difference was observed between the 2 groups (AI, mean −4.33, SD 7.77; video, mean −1.12, SD 6.83; adjusted mean difference −1.65, 95% CI −6.14 to 2.84; *P*=.47) at week 2 (Table 2).

XSL•FO
RenderX

Changes in TUG

Compared with the baseline, the TUG test time for both groups decreased (Figure 4D). No difference was observed in the AI group (mean −1.05, SD 0.78) and video group (mean −0.85, SD 0.86; adjusted mean difference −0.15, 95% CI −0.65 to 0.35; *P*=.55) at week 4 (Table 2).

Changes in Core Muscle Thickness

Figure 4E–H shows the thickness changes of core muscles (TrA and MF) in the 2 groups throughout the study period. At week 4, significant differences were observed for the thickness of TrA and MF on both sides between the AI group and video group (left TrA: 0.30 vs −0.14; adjusted mean difference 0.47, 95% CI 0.19 to 0.75; *P*=.002; right TrA: 0.57 vs −0.09; adjusted mean difference 0.69, 95% CI 0.39 to 0.99; *P*<.001; left MF: 0.90 vs −0.01; adjusted mean difference 0.94, 95% CI 0.14 to 1.74; *P*=.02; right MF: 0.84 vs −0.03; adjusted mean difference 0.92, 95% CI 0.27 to 1.56; *P*=.007) (Table 2).

Discussion

Principal Findings

This study examined the therapeutic effect of AI real-time guidance exercise therapy compared with conventional video guidance exercise therapy in young people with CNSLBP. We found that after 4 weeks of AI-assisted exercise therapy, the pain intensity, body function, psychology, and core muscle thickness had higher improvement compared to conventional exercise therapy.

The Efficacy of AI-Assisted Exercise Telerehabilitation for Participants With CNSLBP

Exercise therapy improves symptoms in participants with CNSLBP. However, as an active exercise and mostly in the form of telerehabilitation, exercise therapy has the problems of insufficient compliance, an inability to receive guidance, and reduced efficacy [13]. AI-based telemedicine has been expected to solve these challenges. After 4 weeks of exercise, the NRS showed a downward trend in 2 groups. Furthermore, the change of most severe NRS and mean NRS in the AI group was significantly different compared to the conventional video group, suggesting that AI-assisted exercise therapy could reduce pain to a greater extent. There have been several previous similar studies based on AI-based exercise therapy for CLBP [16-18,38,39], but not characterized by motion detection and guidance. Some of these studies have shown similar results in pain as our research [38,39]. The comparison demonstrates that AI-based telerehabilitation focusing on motion detection and guidance is feasible in the treatment of CNSLBP. Moreover, a previous study has established that a minimal clinical important difference (MCID) on the NRS is 2/10 points [40]. In our study, the AI group reached MCID at week 2 on the most severe NRS and reached MCID at week 4 on the mean NRS and current NRS, while the video group only reached MCID on the current NRS at week 4. As we know, MCID is not a statistically significant difference, but it is an indicator of the clinical benefits to patients. The difference in the time to achieve MCID reflects that AI-assisted exercise training has better therapeutic effects compared to traditional training. The MCID of the AI group still existed at the week 8 follow-up, while the video group did not reach the MCID, further reflecting the efficacy stability of AI-assisted exercise training. People with CNSLBP received standardized exercise training through AI guidance and increased their motivation to engage in exercise training through real-time feedback, which we believe is the reason for the greater pain reduction in the AI group.

AI-assisted exercise training could improve the physical function of participants with CNSLBP more than traditional training. For physical function in the secondary outcomes, we used ODI and RMDQ for subjective evaluation and used the TUG test for objective evaluation. All 3 indicators showed a downward trend at week 4. Only RMDQ showed a significant difference between the 2 groups. A previous study has shown that RMDQ is more suitable for people with mild to moderate disability, and ODI suitable for people with severe disability [41]. In our study, the majority of participants with CNSLBP were middle-aged and young people with mild to moderate disability, they were still in the early stage of LBP, which are the reasons why only RMDQ showed a significant difference. A previous study has established that an MCID on RMDQ is a 30% change from baseline [42]. Both groups reached MCID on RMDQ at week 2 and week 4, but the change amplitude of the AI group was greater, suggesting that AI-assisted exercise training was superior in certain aspects in reducing physical disabilities. The result is consistent with a previous study [16].

Previous theory has shown that catastrophizing psychology is a prerequisite for the adverse outcomes of chronic musculoskeletal pain [43]. Amplifying pain perception when faced with chronic musculoskeletal pain will increase pain sensitivity, ultimately leading to further exacerbation of pain. Thus, improving catastrophic psychology is equally important for the treatment of CNSLBP. In our study, both groups showed a downward trend on PCS. A previous study investigated the effects of exercise therapy on catastrophizing psychology in people with CNSLBP. The results indicated that 8 weeks of core stability training could effectively improve catastrophizing psychology [44], which is consistent with our findings. Moreover, there was a significant difference between the 2 groups, suggesting that AI-assisted exercise training could better improve the catastrophizing psychology of people with CNSLBP.

Different from previous studies, we measured the thickness of TrA and MF at the end of the 4 weeks of exercise therapy. The occurrence of LBP is usually related to the morphological changes of TrA and MF. These changes usually manifest as fat infiltration, abnormal changes in type I and type II fibers, and asymmetric muscle atrophy [45]. Compared to normal individuals, a decrease in the 2 muscles' thickness can be observed through imaging examinations among people with CNSLBP [46,47]. In our study, AI-assisted exercise therapy significantly increased the thickness of the 2 muscles, while the results of the video group did not show any significant changes. A previous study investigated the effect of routine remote core stability training over 4 weeks on TrA, and the results showed that there was no significant change in muscle thickness [48]. However, 2 other studies with 4 weeks of core stability training guided by therapists showed significant changes in the thickness

XSL•FO
RenderX

of TrA and MF [36,49]. We thought that the reason for this difference was the presence of timely guidance. In our study, AI human key point identification technology replaced therapists to present real-time guidance and feedback. To be precise, guided exercise therapy, rather than simple exercise therapy, has been recommended as the first-line treatment for CNSLBP [50].

To our knowledge, this is the first study to apply AI human key point identification technology to specific exercise training movements among participants with CNSLBP. Through timely guidance of exercise therapy, participants can accurately exercise the target core muscle groups, enhance core strength, increase spinal stability, and ultimately alleviate CNSLBP. Overall, our study demonstrated the effectiveness of AI-assisted exercise therapy compared to conventional therapy from subjective and objective perspectives.

Limitations

Our study had some limitations. First, most of our participants had mild to moderate pain, whereas few had severe pain. The efficacy of AI-assisted exercise therapy for severe LBP is still unclear. Second, we only conducted 4 weeks of follow-up, with lasting efficacy. Sustained efficacy beyond this period remains to be explored. Finally, our participants were middle-aged or young. Whether our findings could be extrapolated to older adults is unknown.

Conclusion

To address the limitations of a lack of guidance and dynamic support in conventional home-based exercise therapy telerehabilitation, we explored a new treatment based on AI human key point identification technology. We found that AI-assisted multimodal exercise therapy had better therapeutic effects compared to conventional telerehabilitation on pain intensity, body function, psychology, and the core muscles. In the future management of CNSLBP, further consideration can be given to applying this treatment model to maximize the effectiveness of telerehabilitation.

Acknowledgments

Jihua Zou, Qing Zeng, and Guozhi Huang are corresponding authors and contributed equally to this work.

We thank Beijing Yinshan Future Health Technology Co, Ltd. (Beijing, China) for providing artificial intelligence technical support. This work was supported by the National Natural Science Foundation of China (grant numbers 82072528, 82002380), Natural Science Foundation of Guangdong Province (grant number 2022A1515012460), and National Health Commission multi-center collaborative horizontal research project (grant number DCMST-NHC-2019-AHT-01).

Authors' Contributions

GH, QZ and YZ designed the study. CX, GL and ZZ drafted the manuscript. CX, ZZ, GL and QY recruited participants. WF was responsible for the random assignment of participants. SL and JH were responsible for the intervention training. JZ collected the data and CY analyzed the data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exercise prescription list.
[XLSX File, 8295 KB - mhealth_v13i1e56176_app1.xlsx]

Checklist 1
CONSORT-EHEALTH checklist (V 1.6).
[PDF File, 543 KB - mhealth_v13i1e56176_app2.pdf]

References

1. Knezevic NN, Candido KD, Vlaeyen JWS, Van Zundert J, Cohen SP. Low back pain. Lancet 2021 Jul 3;398(10294):78-92. [doi: 10.1016/S0140-6736(21)00733-9] [Medline: 34115979]
2. Hoy D, Brooks P, Blyth F, Buchbinder R. The epidemiology of low back pain. Best Pract Res Clin Rheumatol 2010 Dec;24(6):769-781. [doi: 10.1016/j.berh.2010.10.002] [Medline: 21665125]
3. Thiese MS, Hegmann KT, Wood EM, et al. Prevalence of low back pain by anatomic location and intensity in an occupational population. BMC Musculoskelet Disord 2014 Aug 21;15:283. [doi: 10.1186/1471-2474-15-283] [Medline: 25146722]
4. Savigny P, Watson P, Underwood M, on behalf of the Guideline Development Group. Early management of persistent non-specific low back pain: summary of NICE guidance. BMJ 2009 Jun 4;338(jun04 3):b1805-b1805. [doi: 10.1136/bmj.b1805]

XSL•FO
RenderX

5. Ma K, Zhuang ZG, Wang L, et al. The Chinese Association for the Study of Pain (CASP): consensus on the assessment and management of chronic nonspecific low back pain. Pain Res Manag 2019;2019:8957847. [doi: 10.1155/2019/8957847] [Medline: 31511784]
6. Maniadakis N, Gray A. The economic burden of back pain in the UK. Pain 2000 Jan;84(1):95-103. [doi: 10.1016/S0304-3959(99)00187-6] [Medline: 10601677]
7. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017 Feb 18;389(10070):736-747. [doi: 10.1016/S0140-6736(16)30970-9] [Medline: 27745712]
8. Weiner DK, Kim YS, Bonino P, Wang T. Low back pain in older adults: are we utilizing healthcare resources wisely? Pain Med 2006;7(2):143-150. [doi: 10.1111/j.1526-4637.2006.00112.x] [Medline: 16634727]
9. Bernstein IA, Malik Q, Carville S, Ward S. Low back pain and sciatica: summary of NICE guidance. BMJ 2017 Jan 6;356:i6748. [doi: 10.1136/bmj.i6748] [Medline: 28062522]
10. Kamper SJ, Apeldoorn AT, Chiarotto A, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. BMJ 2015 Feb 18;350:h444. [doi: 10.1136/bmj.h444] [Medline: 25694111]
11. Searle A, Spink M, Ho A, Chuter V. Exercise interventions for the treatment of chronic low back pain: a systematic review and meta-analysis of randomised controlled trials. Clin Rehabil 2015 Dec;29(12):1155-1167. [doi: 10.1177/0269215515570379] [Medline: 25681408]
12. Hayden JA, van Tulder MW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. Ann Intern Med 2005 May 3;142(9):776-785. [doi: 10.7326/0003-4819-142-9-200505030-00014] [Medline: 15867410]
13. Palazzo C, Klinger E, Dorner V, et al. Barriers to home-based exercise program adherence with chronic low back pain: patient expectations regarding new technologies. Ann Phys Rehabil Med 2016 Apr;59(2):107-113. [doi: 10.1016/j.rehab.2016.01.009] [Medline: 27050664]
14. Moman RN, Dvorkin J, Pollard EM, et al. A systematic review and meta-analysis of unguided electronic and mobile health technologies for chronic pain—is it time to start prescribing electronic health applications? Pain Med 2019 Nov 1;20(11):2238-2255. [doi: 10.1093/pm/pnz164] [Medline: 31386151]
15. Ulrich L, Thies P, Schwarz A. Availability, quality, and evidence-based content of mHealth apps for the treatment of nonspecific low back pain in the German language: systematic assessment. JMIR Mhealth Uhealth 2023 Sep 13;11:e47502. [doi: 10.2196/47502] [Medline: 37703072]
16. Rughani G, Nilsen TIL, Wood K, et al. The selfBACK artificial intelligence-based smartphone app can improve low back pain outcome even in patients with high levels of depression or stress. Eur J Pain 2023 May;27(5):568-579. [doi: 10.1002/ejp.2080] [Medline: 36680381]
17. Marcuzzi A, Nordstoga AL, Bach K, et al. Effect of an artificial intelligence-based self-management app on musculoskeletal health in patients with neck and/or low back pain referred to specialist care: a randomized clinical trial. JAMA Netw Open 2023 Jun 1;6(6):e2320400. [doi: 10.1001/jamanetworkopen.2023.20400] [Medline: 37368401]
18. Itoh N, Mishima H, Yoshida Y, Yoshida M, Oka H, Matsudaira K. Evaluation of the effect of patient education and strengthening exercise therapy using a mobile messaging app on work productivity in Japanese patients with chronic low back pain: open-label, randomized, parallel-group trial. JMIR Mhealth Uhealth 2022 May 16;10(5):e35867. [doi: 10.2196/35867] [Medline: 35576560]
19. Biebl JT, Rykala M, Strobel M, et al. App-based feedback for rehabilitation exercise correction in patients with knee or hip osteoarthritis: prospective cohort study. J Med Internet Res 2021 Jul 13;23(7):e26658. [doi: 10.2196/26658] [Medline: 34255677]
20. Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007 Oct 2;147(7):478-491. [doi: 10.7326/0003-4819-147-7-200710020-00006] [Medline: 17909209]
21. Hajihasani A, Rouhani M, Salavati M, Hedayati R, Kahlaee AH. The influence of cognitive behavioral therapy on pain, quality of life, and depression in patients receiving physical therapy for chronic low back pain: a systematic review. PM R 2019 Feb;11(2):167-176. [doi: 10.1016/j.pmrj.2018.09.029] [Medline: 30266349]
22. Byström MG, Rasmussen-Barr E, Grooten WJA. Motor control exercises reduces pain and disability in chronic and recurrent low back pain. Spine (Phila Pa 1986) 2013 Mar;38(6):E350-E358. [doi: 10.1097/BRS.0b013e31828435fb]
23. Keane LG. Comparing AquaStretch with supervised land based stretching for chronic lower back pain. J Bodyw Mov Ther 2017 Apr;21(2):297-305. [doi: 10.1016/j.jbmt.2016.07.004] [Medline: 28532872]
24. Lam OT, Strenger DM, Chan-Fee M, Pham PT, Preuss RA, Robbins SM. Effectiveness of the McKenzie method of mechanical diagnosis and therapy for treating low back pain: literature review with meta-analysis. J Orthop Sports Phys Ther 2018 Jun;48(6):476-490. [doi: 10.2519/jospt.2018.7562] [Medline: 29602304]
25. Anderson BE, Bliven KCH. The use of breathing exercises in the treatment of chronic, nonspecific low back pain. J Sport Rehabil 2017 Sep;26(5):452-458. [doi: 10.1123/jsr.2015-0199] [Medline: 27632818]
26. George SZ, Fritz JM, Silfies SP, et al. Interventions for the management of acute and chronic low back pain: revision 2021. J Orthop Sports Phys Ther 2021 Nov;51(11):CPG1-CPG60. [doi: 10.2519/jospt.2021.0304] [Medline: 34719942]

XSL•FO
RenderX

27. Niederer D, Pfeifer AC, Engel T, et al. Dose-response relationship and effect modifier of stabilisation exercises in nonspecific low back pain: a project-wide individual patient data re-analysis on 1483 intervention participants. Pain 2023 May 1;164(5):1087-1095. [doi: 10.1097/j.pain.0000000000002801] [Medline: 36515485]
28. Morone NE, Greco CM, Moore CG, et al. A mind-body program for older adults with chronic low back pain: A randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):329-337. [doi: 10.1001/jamainternmed.2015.8033] [Medline: 26903081]
29. Herr KA, Spratt K, Mobily PR, Richardson G. Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. Clin J Pain 2004;20(4):207-219. [doi: 10.1097/00002508-200407000-00002] [Medline: 15218405]
30. Fan S, Hu Z, Hong H, Zhao F. Cross-cultural adaptation and validation of simplified Chinese version of the Roland-Morris Disability Questionnaire. Spine (Phila Pa 1986) 2012 May;37(10):875-880. [doi: 10.1097/BRS.0b013e31823b0460]
31. Liu H, Tao H, Luo Z. Validation of the simplified Chinese version of the Oswestry Disability Index. Spine (Phila Pa 1986) 2009 May;34(11):1211-1216. [doi: 10.1097/BRS.0b013e31819e2b34]
32. Shen B, Wu B, Abdullah TB, et al. Translation and validation of simplified Chinese version of the Pain Catastrophizing Scale in chronic pain patients: education may matter. Mol Pain 2018;14:1744806918755283. [doi: 10.1177/1744806918755283] [Medline: 29353539]
33. Browne W, Nair BKR. The Timed Up and Go test. Med J Aust 2019 Jan;210(1):13-14. [doi: 10.5694/mja2.12045] [Medline: 30636313]
34. Southwell DJ, Hills NF, McLean L, Graham RB. The acute effects of targeted abdominal muscle activation training on spine stability and neuromuscular control. J Neuroeng Rehabil 2016 Feb 27;13:19. [doi: 10.1186/s12984-016-0126-9] [Medline: 26922079]
35. Goubert D, Oosterwijck JV, Meeus M, Danneels L. Structural changes of lumbar muscles in non-specific low back pain: a systematic review. Pain Physician 2016;19(7):E985-E1000. [Medline: 27676689]
36. Ge L, Huang H, Yu Q, et al. Effects of core stability training on older women with low back pain: a randomized controlled trial. Eur Rev Aging Phys Act 2022 Apr 15;19(1):10. [doi: 10.1186/s11556-022-00289-x] [Medline: 35428169]
37. Skeie EJ, Borge JA, Leboeuf-Yde C, Bolton J, Wedderkopp N. Reliability of diagnostic ultrasound in measuring the multifidus muscle. Chiropr Man Therap 2015;23:15. [doi: 10.1186/s12998-015-0059-6] [Medline: 25878771]
38. Bates NA, Huffman A, Goodyear E, et al. Physical clinical care and artificial-intelligence-guided core resistance training improve endurance and patient-reported outcomes in subjects with lower back pain. Clin Biomech (Bristol, Avon) 2023 Mar;103:105902. [doi: 10.1016/j.clinbiomech.2023.105902]
39. Anan T, Kajiki S, Oka H, et al. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2021 Sep 24;9(9):e27535. [doi: 10.2196/27535] [Medline: 34559054]
40. Childs JD, Piva SR, Fritz JM. Responsiveness of the Numeric Pain Rating Scale in patients with low back pain. Spine (Phila Pa 1986) 2005 Jun;30(11):1331-1334. [doi: 10.1097/01.brs.0000164099.92112.29]
41. Roland M, Fairbank J. The Roland–Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine (Phila Pa 1986) 2000 Dec;25(24):3115-3124. [doi: 10.1097/00007632-200012150-00006]
42. Díaz-Arribas MJ, Fernández-Serrano M, Royuela A, et al. Minimal clinically important difference in quality of life for patients with low back pain. Spine (Phila Pa 1986) 2017;42(24):1908-1916. [doi: 10.1097/BRS.0000000000002298]
43. Vlaeyen JWS, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000;85(3):317-332. [doi: 10.1016/S0304-3959(99)00242-0]
44. Akodu AK, Akindutire OM. The effect of stabilization exercise on pain-related disability, sleep disturbance, and psychological status of patients with non-specific chronic low back pain. Korean J Pain 2018 Jul;31(3):199-205. [doi: 10.3344/kjp.2018.31.3.199] [Medline: 30013734]
45. Wong AYL, Parent EC, Funabashi M, Stanton TR, Kawchuk GN. Do various baseline characteristics of transversus abdominis and lumbar multifidus predict clinical outcomes in nonspecific low back pain? A systematic review. Pain 2013 Dec;154(12):2589-2602. [doi: 10.1016/j.pain.2013.07.010] [Medline: 23867731]
46. Teyhen DS, Williamson JN, Carlson NH, et al. Ultrasound characteristics of the deep abdominal muscles during the active straight leg raise test. Arch Phys Med Rehabil 2009 May;90(5):761-767. [doi: 10.1016/j.apmr.2008.11.011] [Medline: 19406295]
47. Parkkola R, Rytökoski U, Kormano M. Magnetic resonance imaging of the discs and trunk muscles in patients with chronic low back pain and healthy control subjects. Spine (Phila Pa 1986) 1993 Jun;18(7):830-836. [doi: 10.1097/00007632-199306000-00004]
48. Park SD, Yu SH. The effects of abdominal draw-in maneuver and core exercise on abdominal muscle thickness and Oswestry disability index in subjects with chronic low back pain. J Exerc Rehabil 2013 Apr;9(2):286-291. [doi: 10.12965/jer.130012] [Medline: 24278873]
49. Hlaing SS, Puntumetakul R, Khine EE, Boucaut R. Effects of core stabilization exercise and strengthening exercise on proprioception, balance, muscle thickness and pain related outcomes in patients with subacute nonspecific low back pain: a randomized controlled trial. BMC Musculoskelet Disord 2021 Nov 30;22(1):998. [doi: 10.1186/s12891-021-04858-6] [Medline: 34847915]

XSL•FO
RenderX

50. Airaksinen O, Brox JI, Cedraschi C, et al. Chapter 4 European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006 Mar;15(S2):s192-s300. [doi: 10.1007/s00586-006-1072-1]

Abbreviations

AI: artificial intelligence
CLBP: chronic low back pain
CNSLBP: chronic nonspecific low back pain
LBP: low back pain
MCID: minimal clinical important difference
MF: multifidus
NRS: Numerical Rating Scale
ODI: Oswestry Disability Index
PCS: Pain Catastrophizing Scale
RMDQ: Roland-Morris Disability Questionnaire
TrA: transverse abdominus
TUG: Timed Up-and-Go

Edited by L Buis; submitted 09.01.24; peer-reviewed by C Sitges, J Moral-Munoz, Q Chen; revised version received 24.02.25; accepted 06.03.25; published 22.05.25.

Please cite as:
Xiao C, Zhao Y, Li G, Zhang Z, Liu S, Fan W, Hu J, Yao Q, Yang C, Zou J, Zeng Q, Huang G
Clinical Efficacy of Multimodal Exercise Telerehabilitation Based on AI for Chronic Nonspecific Low Back Pain: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e56176
URL: https://mhealth.jmir.org/2025/1/e56176
doi:10.2196/56176

© Chongwu Xiao, Yijin Zhao, Gege Li, Zhuodong Zhang, Siyu Liu, Weichao Fan, Jinjing Hu, Qiuru Yao, Chengduan Yang, Jihua Zou, Qing Zeng, Guozhi Huang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Community-Based Intelligent Blood Glucose Management for Older Adults With Type 2 Diabetes Based on the Health Belief Model: Randomized Controlled Trial

Anqi Zhang[1,2,3*], MM; Jinsong Wang[1,2*], PhD; Xiaojuan Wan[2], PhD; Ziyi Zhang[2], MSN; Shuhan Zhao[2], MPH; Shuo Bai[2], MM; Yamin Miao[2], MSN; Shuang Yang[2], MPH; Xue Jiang[2], MPH

[1]The Affiliated Hospital of Yangzhou University, Yangzhou University, No. 368, Hanjiang Middle Road, Yangzhou City, Yangzhou, China

[2]School of Nursing and Public Health, Yangzhou University, Yangzhou, China

[3]Mufu Mountain Community Health Service Center,Nanjing Gulou District Health Commission, Nan jing, China

[*]these authors contributed equally

Corresponding Author:
Jinsong Wang, PhD
The Affiliated Hospital of Yangzhou University, Yangzhou University, No. 368, Hanjiang Middle Road, Yangzhou City, Yangzhou, China

Abstract

Background: The number of older patients with type 2 diabetes (T2D) is increasing, and effective self-management is crucial for controlling disease progression and its complications.

Objective: We designed a home telemedicine intervention that combines telemedicine with health education based on the Health Belief Model (HBM). This study evaluated its effectiveness on self-management in older patients with T2D.

Methods: Between March and April 2022, we recruited 198 community-dwelling patients with T2D aged 65 years and older. Patients were randomly assigned to either a control group, which received a conventional diabetes management program, or an intervention group, which received a home telemedicine intervention with a health education program based on the HBM. The intervention lasted 6 months. The primary outcome measured was glycosylated hemoglobin (HbA_{1c}); secondary outcomes included diabetes self-management capacity, self-efficacy, and health beliefs. We collected outcome metrics at baseline, 3 months, and 6 months. Generalized estimating equations were used to compare changes in outcomes.

Results: A total of 96.5% (191/198) of patients completed the study. From baseline to 6 months, HbA_{1c} decreased by mean −0.99% (95% CI −1.60% to −0.60%) in the intervention group and mean −0.42% (95% CI −0.90% to 0.90%) in the control group. The intervention group experienced a significantly greater reduction of 0.42% compared to the control group (95% CI 0.12%-0.73%). Furthermore, compared to the control group, the intervention group showed significant improvements in diabetes self-management skills (mean 5.88, 95% CI 4.98-6.79), self-efficacy (mean 9.40, 95% CI 8.15-10.66), and health beliefs (mean 19.54, 95% CI 17.71-21.36) at both 3 and 6 months.

Conclusions: Home telemedicine interventions incorporating health education based on the HBM can provide significant benefits for community-dwelling older patients with T2D, potentially offering new avenues for chronic disease prevention and management. However, future large-scale studies are required to further assess their effectiveness and feasibility.

Trial Registration: China Clinical Trial Registry ChiCTR2200057428; https://www.chictr.org.cn/showproj.html?proj=152850

(JMIR Mhealth Uhealth 2025;13:e60227) doi:10.2196/60227

KEYWORDS

older adults; Type 2 diabetes; telemedicine; health belief model; glycosylated hemoglobin; blood glucose; glucose; glucose management; community; community based; home intelligent; home telemedicine; health education; community dwelling; randomized controlled trial

Introduction

The global diabetes population has quadrupled in the last 30 years, making diabetes the ninth leading cause of death [1]. In China, the prevalence of diabetes has rapidly increased from less than 1% in 1980 to 11.9% in 2022 [2], ranking it the second highest in the world [3]. The incidence of diabetes mellitus in China is positively correlated with aging, with a prevalence of 11.5% among the middle-aged population (45 - 59 y old) and over 20% among the older individuals (≥65 y old) [4]. Older adults with type 2 diabetes (T2D) are at higher risk of comorbidities and complications. Approximately one-fifth of

XSL•FO
RenderX

older patients with T2D develop serious complications [5], such as cardiovascular disease, psychiatric disorders, and physical dysfunction [1]. This imposes a significant burden on both society and the individual [6]. Consequently, providing adequate health care services to enable effective diabetes management has become critical.

Emerging telemedicine practices may address these needs by facilitating the delivery of health care services and health education to patients across geographies via the internet [7]. Particularly for chronic diseases like T2D, which require repeated consultations with physicians, telemedicine offers a viable alternative for patients seeking medical guidance [8]. Telemedicine typically supports four functions: (1) health promotion and knowledge dissemination, (2) timely medical and nursing support, (3) remote disease monitoring, and (4) decision support systems [9]. Given the limitations of health care service systems and the use of 5G networks, remote medical interventions represent a new opportunity to improve chronic disease management for patients with T2D [10]. Since the majority of patients with T2D are treated in community hospitals, telemedicine can help patients benefit from long-term monitoring and regular self-management [11].

Telemedicine interventions guided by health theories may yield better results [12]. The Health Belief Model (HBM) is an important theoretical framework that uses social psychological approaches to explain and predict changes in health-related behaviors, and it is widely used in health behavior research [13]. Studies show that applying the HBM to diabetes management can enhance knowledge [14], improve self-care behaviors [15], promote healthy lifestyles [14], and encourage active participation in disease management [16]. HBM has been used as a theoretical foundation for developing mobile health or telemedicine interventions [17], and many of its constructs have been demonstrated to effectively predict a range of health behaviors in meta-analyses [18]. The model is designed to help patients understand the severity of their disease, provide strategies for management, increase self-efficacy, enhance health behaviors, and achieve better health outcomes [19]. Therefore, given the importance of behavior change in telemedicine interventions, HBM may be a particularly relevant theory to enhance the impact of such interventions on self-management behaviors in older adults with T2D.

However, the effectiveness of such telemedicine interventions in community-dwelling older adults with T2D has not yet been established. Studies have indicated that telemedicine users are generally younger, while older or less-educated individuals have lower usage rates, placing them at a relative disadvantage [20]. Moreover, mismatches between app design and user capabilities, along with the poor quality of diabetes apps, pose significant challenges to telemedicine implementation [21]. To date, there is a notable lack of directly applicable clinical trial data on telemedicine interventions for community-dwelling older adults with T2D.

To provide clinical evidence to inform treatment recommendations for community-dwelling older adults, we conducted a randomized controlled trial of a telemedicine intervention among this group with T2D. We evaluated the effectiveness of a comprehensive, personalized education, and behavioral intervention program, based on the HBM and telemedicine, in improving home self-management for these older adults.

Methods

Recruitment

From March 2022 to April 2022, we recruited participants with T2D from a community hospital in Yangzhou City.

The inclusion criteria were as follows: (1) age 65 years and older; (2) type 2 diabetes mellitus: diabetes was diagnosed based on self-reports with verification (current treatment, medical records, confirmation from health care providers, fasting plasma glucose of ≥126 mg/dL, or symptoms of hyperglycemia with a plasma glucose level of ≥200 mg/dL); (3) informed consent and voluntary participation in the study; and (4) capability for self-care, no communication barriers, and no cognitive impairment.

The exclusion criteria were as follows: (1) serious physical illnesses (eg, malignant tumors; severe cardiac, hepatic, or renal insufficiency; and cerebrovascular accidents), (2) serious diabetes mellitus–related comorbidities, and (3) current participation in other clinical trials or plans to participate in other clinical trials.

Researchers screened eligible participants through electronic health records within the community hospital health care system and contacted them by phone. Participants who agreed to participate in the study were invited to the community hospital, where researchers met them face to face to discuss the details of the intervention program again, following informed consent. Laboratory testing and questionnaire collection were completed for eligible participants. A total of 198 participants were recruited. At 3 and 6 months of intervention, nurses reminded all participants to visit the community hospital for laboratory examinations and to complete the corresponding questionnaire collection.

Randomization and Masking

Using Microsoft Excel to generate random numbers, eligible study participants were randomly divided into an intervention group and a control group, each comprising 99 individuals, thus maintaining a 1:1 ratio. Recruitment personnel, data collectors, and data analysts were kept unaware of the intervention allocation. Recruitment, data collection, and data analysis were conducted by the respective research assistants, researchers, and statisticians. The physician conducting the lab tests was also unaware of the group assignments. Due to the nature of the intervention, the health care provider delivering the intervention was not blinded.

Control Group

Routine community-based diabetes management was implemented, including a free health consultation yearly and random blood glucose testing every 3 months. Participants also received diabetes health management handbooks.

XSL•FO
RenderX

Intervention Group

Overview

In addition to routine community diabetes management and receiving diabetes health management handbooks, participants in this group were managed by a team comprising 8 doctors, 2 pharmacists, 20 nurses, and 1 information engineer. The interventions included (1) telemedicine smart devices and their manuals for information collection and enhancing doctor-patient communication, (2) provision of a glucose meter for self-monitoring of blood glucose, (3) telemedicine guidance from a multidisciplinary team, and (4) personalized diet and exercise plans.

Home Telemedicine Interventions

The "Xiaodu Smart Screen X8 (Baidu)" telemedicine smart device connects via Bluetooth to a glucometer. Health information collected from the smart device, including blood sugar levels, exercise, diet, weight, and blood pressure, is automatically transmitted over wireless networks to a centralized website. This setup allows a multidisciplinary team to easily access and analyze the data, motivating participants to maintain self-management through ongoing personal monitoring and intervention.

Participants self-measured their blood sugar at least twice a week during the study. They chose the timing of each measurement, either fasting or 2 hours post meal. If a patient's blood glucose value deviates from the normal range, the system triggers an automatic alert on the physician's side. Within 3 minutes of receiving the alert, the physician contacts the patient by phone to assess whether medical attention is needed. The doctor also determines the cause of the abnormal values and provides appropriate medical guidance. If a patient's blood glucose value has not been uploaded for more than a week, a nurse calls the patient to remind them.

Pharmacists provide guidance to participants on improving blood glucose levels and offer medication recommendations based on individual glucose control. Participants in the intervention group used a smart device to voice-record their daily medication intake, and pharmacists received these records along with blood glucose data for each patient via the physician. Based on data analysis, the health care team sent a monthly medication management program to participants in the intervention group.

Participants could also input health records, including exercise, diet, weight, and blood pressure, through voice or manual entry on their smart devices. The device then generated personalized diet and exercise regimens based on their conditions. The multidisciplinary team accessed the website every 2 weeks to review the participants' diet and exercise plans and sent nutritional education information through the smart devices.

In addition, the smart devices featured a one-click medical alert and enabled bidirectional communication between participants and the health care team via a chat function.

Health Guidance by the Multidisciplinary Team

Upon their initial inclusion in the study, all participants received a diabetes self-management handbook based on the HBM. The multidisciplinary team, consisting of physicians, nurses, and pharmacists, had undergone specialized training in diabetes management intervention based on the HBM, equipping them with expertise in diabetes management and health coaching. Each group of 25 intervention participants was assigned to a team that provided face-to-face group education sessions every 2 weeks for 3 consecutive months (40 - 50 min each). These sessions took place in the training room of the community hospital. The intervention group was divided into 4 groups, each comprising approximately 25 individuals. Each group attended educational sessions scheduled for the same week. The educational content was structured around the dimensions of the HBM and covered topics, such as the definition, symptoms, and risk factors of T2D (perceived susceptibility); benefits of using smart devices for diabetes management (perceived benefits); impact of T2D on quality of life and physical health (perceived severity); diabetes self-management through smart devices (cues to action); discussion of the patient's experience in self-management and use of smart devices, providing solutions to reduce and remove barriers (perceived barriers); and mobilizing family support, showcasing role models, and increasing self-efficacy (self-efficacy). Each session featured a lecture and discussion on the week's topic, incorporating photos, videos, presentations, and other relevant materials.

Outcomes and Measures

Participants' baseline data were collected using a researcher-designed individual health information form, which included age, sex, living arrangements, educational level, marital status, medical insurance coverage, employment status, monthly household income, duration of diabetes, treatment methods, complications, comorbidities, and other demographic information.

Primary Outcome

The primary outcome was the level of glycated hemoglobin (HbA_{1c}), assessed at baseline, 3 months, and 6 months. A nurse collected a fasting venous blood sample through venipuncture. The samples were processed by the diagnostic testing laboratory at the community hospital and measured using high-performance liquid chromatography on a Tosoh Automatic Glycohemoglobin Analyzer HLC-723G8.

Secondary Outcomes

The Summary of Diabetes Self-Care Activities Measure

The summary of diabetes self-care activities measure (SDSCA) was used to measure the frequency of self-management practices in individuals with diabetes over the past 7 days. This scale, designed by Toobert et al [22], includes 11 items divided into six dimensions: general diet, special diet, blood glucose monitoring, physical activity, foot care, and smoking. It uses a 7-point Likert scoring system, where the score for each dimension is the mean score of the items within that dimension. A higher total score indicates better self-management behavior.

XSL•FO
RenderX

The Self-Efficacy for Diabetes

The study used the Chinese version of the self-efficacy for diabetes (SED), originally developed by Lorig and translated and revised by Wei [23]. This scale assesses participants' self-efficacy in managing their glycemia, diet, exercise, and disease monitoring. It includes 9 items scored on a 5-point Likert scale, with higher scores indicating greater self-efficacy.

Health Beliefs Questionnaire

The study used a health belief questionnaire designed by Chen [24], based a previously proposed HBM framework. Tailored for individuals with diabetes, this questionnaire measures their health beliefs across 20 items, distributed among 5 dimensions: perceived benefits, perceived severity, perceived susceptibility, perceived barriers, and cues to action. Scoring used a 5-point Likert scale, where higher scores indicate stronger health beliefs.

Sample Size Calculation

This study used HbA_{1c} values as the primary outcome indicator. We set the reference values based on previous literature [25]. The mean HbA_{1c} in the treatment group was 9% (SD 1.9%). In the control group, the mean HbA_{1c} was 10% (SD 2.7%). We set a 2-sided α of 0.05 with 80% power. Using G*Power 3.1 software (Heinrich-Heine-Universität Düsseldorf), the sample size for the treatment group (N1) was calculated to be 87 cases, and for the control group (N2), it was 87 cases. Accounting for a potential dropout and refusal rate of 10%, we determined that a minimum of 96 participants in each of the intervention and control groups was needed, totaling a minimum of 192 participants.

Statistical Analysis

Statistical analysis in this study was conducted using SPSS 27.0 (IBM Corp), R (version 4.0.3, R Core Team), and GraphPad Prism (version 9.4.1). A *P* value <.05 was considered statistically significant.

Descriptive statistics were used to describe the demographic distribution and scores for the SDSCA, SED, and HBM. Variables with normal distribution were expressed as mean (SD), categorical variables were expressed as frequencies (percentages), and non-normally distributed variables were expressed as medians (quartiles). Continuous values were compared to baseline data using the independent samples *t* test or Mann-Whitney *U* test.

The study used an intention-to-treat analysis to analyze the data. With no pattern in the missing data, the 10% missing values were assumed to be missing at random. The predicted values of the missing data were estimated using Markov Chain Monte Carlo multiple imputation to obtain a complete dataset. For the primary outcome (6-month change in HbA_{1c}), linear mixed-effects models were used with HbA_{1c} as the dependent variable. This model included a random intercept with an unstructured covariance matrix, indicator variables for time and study group, and group-by-time interaction terms. The same approach was used for the secondary outcomes.

Ethical Considerations

The study protocol adhered to the Declaration of Helsinki and received approval from the Ethics Committee of the School of Nursing and Public Health at Yangzhou University (approval number YZUHL2022001). The trial has been developed and reported in agreement with the CONSORT (Consolidated Standards for Reporting Clinical Trials) checklist (Checklist 1). Written informed consent was obtained from all participants.

To ensure privacy and security during the use of the device, each participant must log in to the device with a unique account and password. After data collection, it is anonymized and stored in the cloud server, strictly following advanced data protection regulations. At the same time, the system will perform encrypted backup operations as needed to ensure that data and programs are protected.

Results

Participant Characteristics

A total of 219 potential participants were assessed for eligibility in this study. Out of the 210 who were eligible, 12 were excluded due to scheduling conflicts and a lack of interest. Consequently, 198 individuals agreed to participate and were randomly assigned to either the intervention group (n=99) or the control group (n=99) after providing informed consent. A total of 198 out of 191 (96.5%) participants completed the follow-up assessments (Figure 1). The ages of the 198 participants ranged from 65 to 87 years. The duration of diabetes varied from 1 to 41 years, with a mean HbA_{1c} level of 7.05% (SD 2.41%, range 6.20%-8.25%). In addition, 76.8% (198/152) of the participants were taking oral hypoglycemic drugs to control their blood glucose, 22.7% (198/45) had complications, and 29.3% (198/58) had 2 or more comorbidities. There were no significant differences in baseline characteristics between the 2 groups (*P*>.05), as shown in Table 1.

XSL•FO
RenderX

Table . Demographic and clinical characteristics of the participants (N=198).

Variables	Intervention group (n=99)	Control group (n=99)	*t* test[a]/Mann-Whitney χ^2[b] (degrees of freedom for *t* test=196)	*P* value
Age (years), mean (SD)	72.82 (4.85)	74.11 (5.49)	–1.621[a]	.11
Disease duration (years), mean (SD)	12.67 (7.96)	13.61 (9.83)	–0.266[a]	.79
Sex, n (%)			–1.703[b]	.09
Male	53 (53.54)	44 (44.44)	—[c]	—
Female	46 (46.46)	55 (55.56)	—	—
Living arrangements, n (%)			–0.097[b]	.92
Living alone	7 (7.07)	8 (8.08)	—	—
Living with spouse	30 (30.30)	31 (31.31)	—	
Living with family	62 (62.62)	60 (60.61)	—	—
Educational level, n (%)			–1.871[b]	.06
Elementary school or below	32 (32.32)	47 (47.47)	—	—
Junior high school	38 (38.38)	29 (29.29)	—	—
High school	24 (24.24)	18 (18.18)	—	—
University and above	5 (5.06)	5 (5.06)	—	—
Marital status, n (%)			–0.420[b]	.68
Married	85 (85.85)	87 (87.88)	—	—
Divorced or widowed	14 (14.15)	12 (12.12)	—	—
Medical insurance coverage, n (%)			–1.219[b]	.22
Employees health insurance	82 (82.82)	76 (76.77)	—	—
Residential health insurance	10 (10.10)	8 (8.08)	—	—
Commercial health insurance	7 (7.08)	15 (15.15)	—	—
Employment status, n (%)			–1.790[b]	.07
Employed	1 (1.06)	4 (4.12)	—	—
Unemployed	4 (4.04)	8 (8.08)	—	—
Retired	94 (94.95)	87 (87.88)	—	—
Monthly household income (RMB[d]), n (%)			–0.064[b]	.95
<3000	15 (15.15)	21 (21.21)	—	—
3000-5000	33 (33.33)	24 (24.24)	—	—
5000-10000	45 (45.45)	46 (46.46)	—	—
>10000	6 (6.07)	8 (8.09)	—	—
Treatment method, n (%)			–1.330[b]	.18
None	0	10 (10.10)	—	—
Oral medication	81 (81.82)	71 (71.72)	—	—
Insulin	6 (6.06)	5 (5.05)	—	—
Oral medication+ insulin	12 (12.12)	13 (13.13)	—	—

XSL•FO
RenderX

Variables	Intervention group (n=99)	Control group (n=99)	*t* test[a]/Mann-Whitney χ^2[b] (degrees of freedom for *t* test=196)	*P* value
Comorbidities, n (%)			–1.583[b]	.11
No	11 (11.11)	25 (25.25)	—	—
One	58 (58.59)	46 (46.47)	—	—
Two	26 (26.26)	25 (25.25)	—	—
Three	4 (4.04)	3 (3.03)	—	—
Complications, n (%)			–1.867[b]	.06
No	71 (71.72)	82 (82.83)	—	—
One	26 (26.26)	16 (16.16)	—	—
Two	2 (2.02)	1 (1.01)	—	—
HbA_{1c}[e], mean (95% CI)	7.3 (6.30-8.50)	6.90 (6.10-8.90)	–0.514[a]	.61
SDSCA[f], mean (SD)	16.86 (5.02)	17.18 (4.83)	–1.321[a]	.19
SED[g], mean (SD)	25.37 (5.01)	26.28 (3.34)	–1.197[a]	.23
HBM[h], mean (SD)	69.53 (7.56)	71.59 (4.71)	–1.110[a]	.27

[a]*t* test.

[b]Chi-square values derived from the Mann-Whitney *U* test.

[c]Not applicable.

[d]RMB: Renminbi; RMB 1=US $0.14.

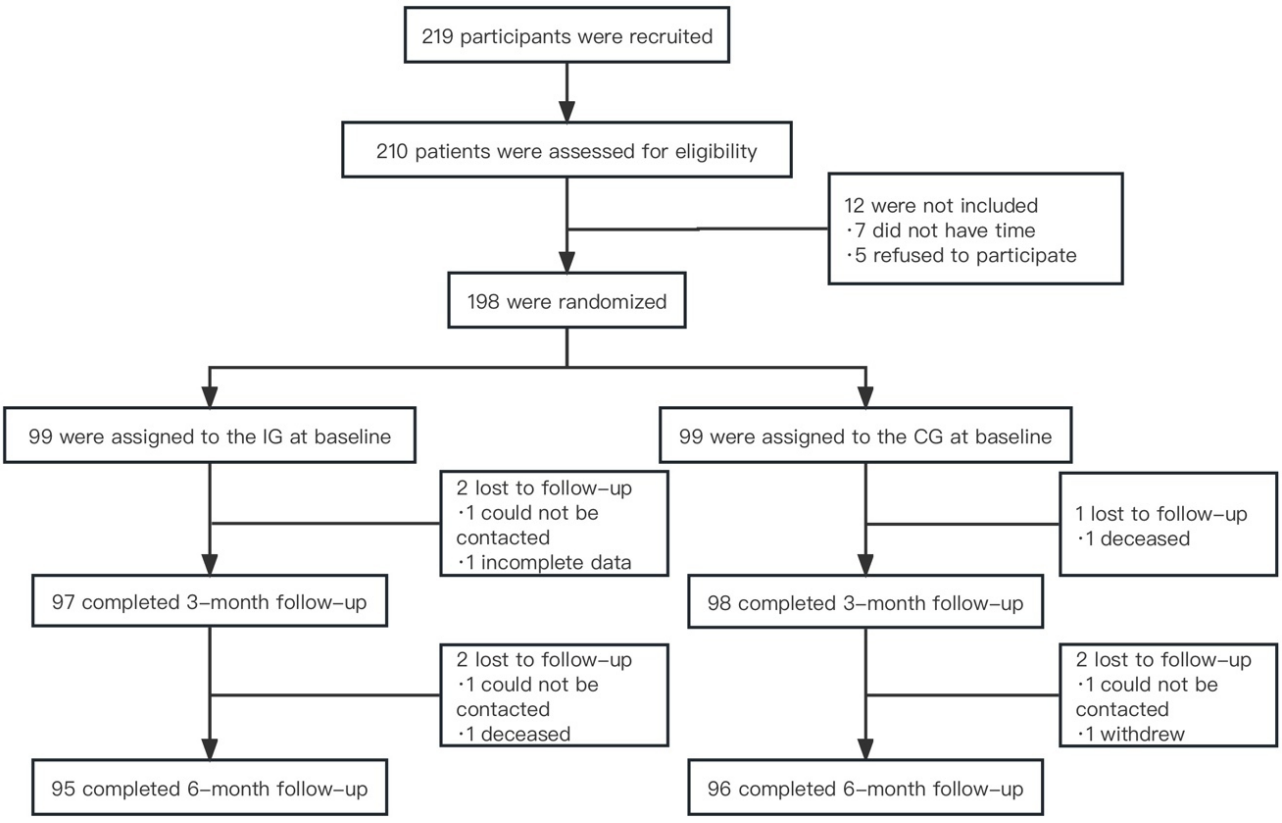
[e]HbA_{1c}: glycated hemoglobin.

[f]SDSCA: The Summary of Diabetes Self-Care Activities Measure.

[g]SED: The Self-efficacy for Diabetes.

[h]HBM: health belief model.

Figure 1. Flowchart of the study. IG: Intervention group; CG: Control group



XSL•FO
RenderX

Primary Outcome

Participants' HbA_{1c} levels decreased after 3 months of intervention (P=.02), with a more significant improvement noted at 6 months (P<.001). In the intervention group, participants' HbA_{1c} levels showed better control during follow-up (P<.001). The decrease in HbA_{1c} levels at 6 months in the control group was not statistically significant (P=.06). The interaction between group and time was not statistically significant (P=.51), as shown in Table 2. In the intervention group, HbA_{1c} levels decreased from a baseline of 7.87% to 7.28% at 3 months and further to 6.88% at 6 months. In the control group, HbA_{1c} levels decreased from a baseline of 7.72% to 7.30% at 6 months. These changes are illustrated in Figure S1 in Multimedia Appendix 1.

Table . The impact of home telemedicine interventions based on the Health Belief Model on glycated hemoglobin levels in older adults with type 2 diabetes.

Outcome measurement	[a]	SE (95% CI)	*P* value
HbA_{1c} (%)[b]			
Group[c] effect	8.02	0.404 (7.228 to 8.812)	<.001
Time[d] 2	−1.04	0.451 (−1.924 to −0.156)	.02
Time[d] 3	−1.56	0.451 (−2.444 to −0.676)	<.001
Group time[e] 2	0.45	0.285 (−0.109 to 1.009)	.11
Group time[e] 3	0.57	0.285 (0.011 to 1.129)	.05

[a] : regression coefficient.

[b]HbA_{1c}: glycated hemoglobin.

[cc]group effect: difference between groups at baseline.

[d]time 2, time 3: time effects at 3 and 6 months compared to baseline.

[e]group*time effect: difference in change at 3 and 6 months postintervention compared to baseline.

Secondary Outcomes

At 3 months of intervention, mean SDSCA scores increased from baseline levels in both the intervention (17.13, SD 24.50 vs 19.59, SD 21.54) and control groups (16.97, SD 22.52 vs 18.43, SD 17.58). At 6 months, mean SDSCA scores were higher than baseline levels in the intervention group (17.13, SD 24.50 vs 22.93, SD 14.59; P<.001), whereas mean SDSCA scores were lower than baseline levels in the control group (16.97, SD 22.52 vs 16.93, SD 22.49). The difference in scores between the 2 groups was statistically significant (P<.001).

Mean SED scores also improved from baseline in both groups at 3 months: the intervention (25.38, SD 25.38 vs 29.51, SD 28.25) and control groups (26.18, SD 11.16 vs 26.44, SD 7.55). At 6 months, mean SED scores were higher than baseline levels in the intervention group (25.38, SD 25.38 vs 34.99, SD 17.95; P<.001), while mean SED scores were slightly higher than baseline levels in the control group (26.18, SD 11.16 vs 26.24, SD 8.75). There was a statistically significant difference between the 2 groups (P<.001) at both 3 and 6 months.

At 3 months, HBM scores were higher than baseline in both the intervention group (69.91, SD 56.44 vs 77.89, SD 29.95) and the control group (1.55, SD 22.12 vs 73.35, SD 19.73). At 6 months, HBM scores were significantly higher than baseline levels in both the intervention group (89.65, SD 18.44) and the control group (77.04, SD 23.05), and the intervention group showed a greater improvement in HBM scores. There was a statistically significant difference between the 2 groups (P<.001) at both 3 and 6 months, as detailed in Table 3 and Figure S2 in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Changes in secondary outcomes during the home telemedicine intervention compared to baseline.

Variables and time	Intervention group, mean (95% CI)	Control group, mean (95% CI)	Difference between groups, mean (95% CI)
SDSCA [a]-compared to Baseline			
Change in 3 months	2.58 (1.27 to 3.88)	1.34 (0.17 to 2.51)[b]	–0.97 (–2.19 to 0.26)
Change in 6 months	5.88 (4.98 to 6.79)[c]	–0.11 (–1.34 to 1.14)	–5.72 (–6.89 to –4.55)[c]
SED [d]-compared to baseline			
Change in 3 months	3.98 (2.72 to 5.24)[c]	0.18 (–0.23 to 0.60)	–3.15 (–4.22 to –2.07)[c]
Change in 6 months	9.40 (8.15 to 10.66)[c]	0.05 (–0.49 to 0.59)	−8.70 (−9.68 to −7.66)[c]
HBM-compared to Baseline			
Change in 3 months	8.42 (6.54 to 10.30)[c]	1.71 (1.02 to 2.39)[c]	–4.89 (–6.21 to –3.57)[c]
Change in 6 months	19.54 (17.71 to 21.36)[c]	5.52 (4.43 to 6.60)[c]	–12.19 (–13.57 to –10.81)[c]

[a]SDSCA: The Summary of Diabetes Self-Care Activities Measure.

[b]$P<.05$.

[c]$P<.001$.

[d]SED: The Self-efficacy for Diabetes.

[e]HBM: Health Belief Model.

Adverse Events

No serious adverse clinical events were reported from enrollment to the 6-month follow-up. However, some operational issues with the devices, such as Bluetooth pairing problems between blood glucose meters and smart devices, were reported as inconvenient by participants. Indeed, it has been demonstrated that such issues can lead to lower satisfaction and reduced use of remote technology [26]. Before the start of the study, all participants were informed about the potential for these problems and the availability of telephone assistance.

Discussion

Principal Findings

In this study, we tested a home telemedicine intervention and health education program based on the HBM among community-recruited older patients with T2D. Screening, recruitment, and participation outcomes indicated that, although the number of participants was small, the vast majority were willing to participate and complete the program. While adherence was high in both groups, adherence to self-management varied significantly. Overall, this program has shown potential in helping older patients with T2D achieve safe and effective home diabetes self-management under the supervision and guidance of health care providers.

Longitudinal analyses demonstrated significant improvements in HbA_{1c} levels at 3 and 6 months, as well as significant enhancements in diabetes self-management, self-efficacy, and health belief scores in the intervention group. In contrast, in the control group, only health beliefs showed significant improvement at 6 months. HbA_{1c} levels were significantly lower following the intelligent blood glucose management in the intervention group compared to the control group. This study underscores the effectiveness of intelligent blood glucose management as a means of delivering behavioral health interventions to older adults with T2D [27]. At 3 months, HbA_{1c} levels in the intervention group decreased significantly by 0.59% from baseline, compared to 0.13% in the control group; however, this difference was not statistically significant. At 6 months, HbA_{1c} levels in the intervention group further decreased by 0.40% compared to 0.29% in the control group, indicating sustained improvement. This finding is consistent with research by Sun et al [28], suggesting that older adults may require time to familiarize themselves with telemedicine systems. Furthermore, after training and remote guidance from the health care team, participants were better able to use smart devices independently [29]. In this study, glycemic management improved in both groups, which may be attributed to several factors. First, the Hawthorne effect [30] could explain improvements, as control participants also had frequent contact with researchers and received additional attention, including regular tests of their glycosylated hemoglobin every 3 months. Second, all participants received a diabetes self-management handbook and specialized care from health care professionals, potentially allowing for more intensive glucose management than usual [31]. However, sustained and effective long-term effects are challenging without ongoing telemedicine oversight and support [32]. Therefore, long-term follow-up of older adults with T2D is crucial.

The significant and long-lasting improvement in self-management in the intervention group, compared to the control group, suggests that our intervention may promote improved self-management behaviors, aligning with findings from previous studies [33]. However, one study indicated that the difference between groups using intelligent blood glucose management alone was not significant, possibly due to older adults' lack of experience with and incomplete knowledge of smart devices [34]. After incorporating group health education based on the HBM, participants in the intervention group

XSL•FO
RenderX

significantly improved their self-management skills. This underscores the vital role that face-to-face diabetes health education plays in helping older adults overcome the challenges of using smart devices and enhancing their self-management. A systematic evaluation noted that an essential component of diabetes self-management is the active participation of the patient in their care [8]. One effective method is through health education, which facilitates the learning process of self-management. Ahmad et al [35] found that older adults with T2D may require regular face-to-face counseling to address the challenges they encounter during telemedicine interventions. Providing support and education through a dedicated community-based diabetes health care team can help participants achieve optimal self-management [31]. Therefore, combining telemedicine with face-to-face health education appears to be a promising approach for enhancing diabetes self-management.

In this study, the health belief scores of participants in the intervention group were higher than those in the control group. The longer the intervention period, the more pronounced the differences in health belief scores became. This suggests that telemedicine-based health education is more effective in improving the health beliefs of older participants with T2D compared to traditional health education [36]. Health education makes participants aware of the complications, severity, and harms associated with their condition, as well as the benefits of adopting a healthy lifestyle. This encourages them to take the initiative in managing their health and deepens their understanding of diabetes. In addition, the feedback and reminders provided by telemedicine play a positive role in enhancing participants' health beliefs [37], which strengthens their health motivation and helps them establish a correct and comprehensive set of health beliefs. The results of this study show that health education based on the HBM within the context of telemedicine has a positive impact, increasing patient self-efficacy, reducing perceived barriers in the intervention group, and facilitating an increase in patient health behaviors.

We also found that the intervention improved self-efficacy in older adults with T2D, similar to the results of other studies [38]. Given that increased health beliefs are associated with higher diabetes self-efficacy [20], participants in the intervention group experienced higher self-efficacy as their level of health beliefs increased. Elevated levels of self-efficacy are often linked with decreased barriers to engaging in health behaviors, which in turn leads to increased motivation to engage in desired activities [39].

Limitations

First, while studies have shown the short-term impact of telemedicine on glycemic management in older adults with T2D, the long-term validity of these effects remains unassessed. Second, the participants who chose to join this study likely demonstrated some level of readiness for digital health literacy change. Therefore, strategies have yet to be developed for those uninterested individuals who did not participate in the study. Finally, the study relied on structured questionnaires for participants to self-assess their self-efficacy, health beliefs, and diabetes self-management skills, which may have introduced recall bias into the results.

Conclusions

The findings of this study suggest that health education and telemedicine interventions based on the HBM can improve HbA_{1c} levels and other related outcomes in older adults with T2D. The telemedicine system provided timely access to the health status of older participants with T2D, facilitated communication and feedback between doctors and participants, and enabled the inclusion of health care provider counseling to enhance patient motivation. This, in turn, improved the effectiveness of medical care, leading to better glycemic management in the intervention group. Future research should further evaluate the long-term effects of this approach compared to routine community diabetes care. In addition, it should focus on providing better support for lifestyle changes through enhanced interventions and counseling to improve patient motivation and confidence, ultimately offering more comprehensive diabetes management for participants.

Acknowledgments

Thanks to the School of Nursing of Yangzhou University, the Affiliated Hospital of Yangzhou University, and the Community Health Service Center of Wenfeng Street, Guangling District, Yangzhou City.

Data Availability

The datasets generated during and/or analyzed in the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

AZ, JW, and XW contributed to conceptualization, validation, writing—review and editing. AZ performed data curation. AZ and JW were involved in formal analysis. AZ, YM, SB, SZ, SY, and ZZ led investigation. AN and JW designed the methodology, handled project administration, and performed visualization. JW and XW were involved in supervision. AN managed writing–original draft.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Mean andCIs of glycosylated hemoglobin, summary of diabetes self-care activities measure, self-efficacy for diabetes and Health Belief Model scores for the 2 groups in the study.
[DOCX File, 122 KB - mhealth_v13i1e60227_app1.docx]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File, 1385 KB - mhealth_v13i1e60227_app2.pdf]

References

1. Ali MK, Pearson-Stuttard J, Selvin E, et al. Interpreting global trends in type 2 diabetes complications and mortality. Diabetologia 2022 Jan;65(1):3-13. [doi: 10.1007/s00125-021-05585-2] [Medline: 34837505]
2. Society CD, Care NOfPD. National guidelines for the prevention and control of diabetes in primary care. Zhonghua Nei Ke Za Zhi 2022;61(3):249-262. [doi: 10.3760/cma.j.cn112138-20220120-000063] [Medline: 30486556]
3. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119. [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
4. Overbeek JA, Nijpels G, Swart KMA, et al. Type 2 Diabetes, the epidemic: trends in prevalence and incidence, 2004-2020. Diabetes Metab Syndr Obes 2024;17:1503-1509. [doi: 10.2147/DMSO.S445288] [Medline: 38562280]
5. Guldemond N. What is meant by “integrated personalized diabetes management”: a view into the future and what success should look like. Diabetes Obes Metab 2024 Mar;26 Suppl 1:14-29. [doi: 10.1111/dom.15476] [Medline: 38328815]
6. Nauck MA, Wefers J, Meier JJ. Treatment of type 2 diabetes: challenges, hopes, and anticipated successes. Lancet Diabetes Endocrinol 2021 Aug;9(8):525-544. [doi: 10.1016/S2213-8587(21)00113-3] [Medline: 34181914]
7. Sood S, Mbarika V, Jugoo S, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
8. Altabtabaei R, Alhuwail D. Challenges and opportunities of the use and adoption of telemedicine for diabetes care and management during the COVID-19 pandemic and beyond. iproc 2023;9:e41285. [doi: 10.2196/41285]
9. Fleming GA, Petrie JR, Bergenstal RM, et al. Diabetes digital app technology: benefits, challenges, and recommendations. A Consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetes Care 2020 Jan;43(1):250-260. [doi: 10.2337/dci19-0062] [Medline: 31806649]
10. Hood KK, Bergenstal RM, Cushman T, et al. Patient-reported outcomes improve with a virtual diabetes care odmel that includes continuous glucose monitoring. Telemed J E Health 2025 Jan;31(1):75-84. [doi: 10.1089/tmj.2024.0093] [Medline: 39166322]
11. Lozano Hernández CM, Medina-García R, de Hoyos-Alonso MC, et al. Improvement in quality of life with the use of a technological system among patients with chronic disease followed up in primary care (TeNDER Project): protocol for a randomized controlled trial. JMIR Res Protoc 2023 Jul 3;12:e47331. [doi: 10.2196/47331] [Medline: 37399054]
12. Del-Valle-Soto C, López-Pimentel JC, Vázquez-Castillo J, et al. A comprehensive review of behavior change techniques in wearables and IoT: implications for health and well-being. Sensors (Basel) 2024 Apr 10;24(8):2429. [doi: 10.3390/s24082429] [Medline: 38676044]
13. Du QH, Zhang ZC, Yang Y, et al. How health seeking behavior develops in patients with type 2 diabetes: a qualitative study based on health belief model in China. Front Public Health 2024;12:1414903. [doi: 10.3389/fpubh.2024.1414903] [Medline: 39045167]
14. Owolabi EO, Boakye MDS, Omololu SO, et al. Technology-based interventions to promote diabetes self-management behaviors for persons newly diagnosed with Type 2 Diabetes: a scoping review. Curr Diab Rep 2024 Dec;24(12):257-272. [doi: 10.1007/s11892-024-01553-3] [Medline: 39352613]
15. Melkamu L, Berhe R, Handebo S. Does patients’ perception affect self-care practices? The perspective of health belief model. Diabetes Metab Syndr Obes 2021;14(2145-54):2145-2154. [doi: 10.2147/DMSO.S306752] [Medline: 34012280]
16. Hashim SA, Mohd Yusof BN, Abu Saad H, et al. Effectiveness of simplified diabetes nutrition education on glycemic control and other diabetes-related outcomes in patients with type 2 diabetes mellitus. Clin Nutr ESPEN 2021 Oct;45(141-9):141-149. [doi: 10.1016/j.clnesp.2021.07.024] [Medline: 34620310]
17. Sharifat R, Borazjani F, Araban M, et al. Nutritional education on health beliefs, metabolic profiles, and quality of life among high-risk pregnant women for gestational diabetes mellitus: a randomized controlled trial. Sci Rep 2024 Nov 12;14(1):27712. [doi: 10.1038/s41598-024-78447-7] [Medline: 39532953]
18. Skovlund SE, Troelsen LH, Noergaard LM, et al. Feasibility and acceptability of a digital patient-reported outcome tool in routine outpatient diabetes care: mixed methods formative pilot study. JMIR Form Res 2021 Nov 3;5(11):e28329. [doi: 10.2196/28329] [Medline: 34730545]

XSL•FO
RenderX

19. Liu YW, Tang CC. Health beliefs, protective behaviors, and information-seeking. Nurs Res 2024;73(2):158-165. [doi: 10.1097/NNR.0000000000000712] [Medline: 38193910]
20. Guo SHM, Lin JL, Hsing HC, et al. The effect of mobile eHealth education to improve knowledge, skills, self-care, and mobile eHealth literacies among patients with diabetes: development and evaluation study. J Med Internet Res 2023 Dec 6;25:e42497. [doi: 10.2196/42497] [Medline: 38055321]
21. Toschi E, Adam A, Atakov-Castillo A, et al. Use of telemedicine in adults with Type 1 Diabetes: do age and use of diabetes-related technology matter? Telemed J E Health 2023 Sep;29(9):1374-1382. [doi: 10.1089/tmj.2022.0397] [Medline: 36695656]
22. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
23. Wei J. A Study on the Correlation between Knowledge, Self-Efficacy and Self-Management Behavior of Elderly Diabetic Patients in Rural Areas: Hangzhou Normal University; 2013.
24. Chen Y. A Survey of Compliance and Its Influencing Factors in Readmitted Diabetic Patients: Central South University; 2007.
25. Egede LE, Dawson AZ, Walker RJ, et al. Randomized controlled trial of technology-assisted case management in low-income adults with type 2 diabetes: effect on quality of life and blood pressure. J Telemed Telecare 2024 Jan;30(1):107-115. [doi: 10.1177/1357633X211028491] [Medline: 34251865]
26. Bradway M, Pfuhl G, Joakimsen R, et al. Analysing mHealth usage logs in RCTs: explaining participants' interactions with type 2 diabetes self-management tools. PLoS ONE 2018;13(8):e0203202. [doi: 10.1371/journal.pone.0203202] [Medline: 30161248]
27. Chong CJ, Makmor-Bakry M, Hatah E, et al. Factors infuencing type 2 diabetes mellitus patients' readiness, acceptance and barriers towards mobile apps adoption for medication adherence. Int J Diabetes Dev Ctries 2024 Oct 11;PMID:WOS. [doi: 10.1007/s13410-024-01413-6]
28. Sun C, Sun L, Xi S, et al. Mobile phone-based telemedicine practice in older Chinese patients with Type 2 Diabetes Mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 4;7(1):e10664. [doi: 10.2196/10664] [Medline: 30609983]
29. Altabtabaei R, Alhuwail D. Exploring the challenges and opportunities of adopting and using telemedicine for diabetes care and management: qualitative semistructured interview study among health care providers and patients with diabetes. JMIR Hum Factors 2023 Sep 7;10:e46324. [doi: 10.2196/46324] [Medline: 37676711]
30. McCarney R, Warner J, Iliffe S, et al. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007 Jul 3;7(30):17608932. [doi: 10.1186/1471-2288-7-30] [Medline: 17608932]
31. Sankoda A, Nagae Y, Waki K, et al. Glycemic control, renal progression, and use of telemedicine phone consultations among Japanese patients with Type 2 diabetes mellitus during the COVID-19 pandemic: retrospective cohort study. JMIR Diabetes 2023 Nov 21;8:e42607. [doi: 10.2196/42607] [Medline: 37315193]
32. Lewinski AA, Walsh C, Rushton S, et al. Telehealth for the longitudinal management of chronic conditions: systematic review. J Med Internet Res 2022 Aug 26;24(8):e37100. [doi: 10.2196/37100] [Medline: 36018711]
33. Barnard-Kelly K. Utilizing eHealth and telemedicine technologies to enhance access and quality of consultations: it's not what you say, it's the way you say it. Diabetes Technol Ther 2019 Jun;21(S2):S241-S247. [doi: 10.1089/dia.2019.0015] [Medline: 31169431]
34. Klooster IT, Kip H, van Gemert-Pijnen L, et al. A systematic review on eHealth technology personalization approaches. iScience 2024 Sep 20;27(9):110771. [doi: 10.1016/j.isci.2024.110771] [Medline: 39290843]
35. Ahmad F, Wysocki RW, Fernandez JJ, et al. Patient perspectives on telemedicine during the COVID-19 pandemic. Hand (N Y) 2023 May;18(3):522-526. [doi: 10.1177/15589447211030692] [Medline: 34515550]
36. Webb TL, Joseph J, Yardley L, et al. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
37. Cho YM, Lee S, Islam SMS, et al. Theories applied to m-Health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018 Oct;24(10):727-741. [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
38. Guo J, Wang H, Ge L, et al. Effectiveness of a nurse-led mindfulness stress-reduction intervention on diabetes distress, diabetes self-management, and HbA1c levels among people with type 2 diabetes: a pilot randomized controlled trial. Res Nurs Health 2022 Feb;45(1):46-58. [doi: 10.1002/nur.22195] [Medline: 34741544]
39. Izquierdo M, Merchant RA, Morley JE, et al. International Exercise Recommendations in Older Adults (ICFSR): Expert Consensus Guidelines. J Nutr Health Aging 2021;25(7):824-853. [doi: 10.1007/s12603-021-1665-8] [Medline: 34409961]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
HbA_{1c}: glycosylated hemoglobin
HBM: Health Belief Model

XSL•FO
RenderX

SDSCA: summary of diabetes self-care activities measure
SED: self-efficacy for diabetes
T2D: type 2 diabetes

Edited by L Buis; submitted 05.05.24; peer-reviewed by C Yang, D Dinh, M Karimi; revised version received 24.11.24; accepted 15.04.25; published 19.06.25.

Please cite as:
Zhang A, Wang J, Wan X, Zhang Z, Zhao S, Bai S, Miao Y, Yang S, Jiang X
Community-Based Intelligent Blood Glucose Management for Older Adults With Type 2 Diabetes Based on the Health Belief Model: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e60227
URL: https://mhealth.jmir.org/2025/1/e60227
doi:10.2196/60227

©Anqi Zhang, Jinsong Wang, Xiaojuan Wan, Ziyi Zhang, Shuhan Zhao, Shuo Bai, Yamin Miao, Shuang Yang, Xue Jiang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of 12-Month mHealth Home Telemonitoring on Clinical Outcomes in Older Individuals With Hypertension and Type 2 Diabetes: Multicenter Randomized Controlled Trial

Matic Mihevc[1,2], MD; Majda Mori Lukančič[2], RN; Črt Zavrnik[1,2], MD; Tina Virtič Potočnik[2,3], MD, PhD; Nina Ružić Gorenjec[2,4], PhD; Marija Petek Šter[1], MD, PhD; Zalika Klemenc-Ketiš[1,2,3], MD, PhD; Antonija Poplas Susič[1,2], MD, PhD

[1]Department of Family Medicine, Medical Faculty University of Ljubljana, Ljubljana, Slovenia

[2]Primary Healthcare Research and Development Institute, Community Health Centre Ljubljana, Metelkova ulica 9, Ljubljana, Slovenia

[3]Department of Family Medicine, Medical Faculty University of Maribor, Maribor, Slovenia

[4]Institute for Biostatistics and Medical Informatics, Medical Faculty University of Ljubljana, Ljubljana, Slovenia

Corresponding Author:
Matic Mihevc, MD
Department of Family Medicine, Medical Faculty University of Ljubljana, Ljubljana, Slovenia

Abstract

Background: As the population ages, the prevalence of chronic diseases such as arterial hypertension (AH) and type 2 diabetes (T2D) is increasing, posing challenges for effective management in primary care settings. Although mobile health (mHealth) home telemonitoring offers promising support, evidence regarding its clinical impact on older patients is limited.

Objective: The objective of this paper was to evaluate the impact of 12-month telemonitoring on clinical outcomes in older individuals with AH and T2D compared to standard care in a primary care setting.

Methods: In a multicenter, open-label, randomized controlled trial, individuals aged 65 years and older with AH and T2D were randomly assigned in a 1:1 ratio to either a telemonitoring group or a standard care group. The telemonitoring group received mHealth support in addition to standard care. Over 12 months, participants measured blood pressure (BP) twice weekly with 2 consecutive readings each morning and evening, using the second reading as valid. Blood glucose (BG) was measured monthly, both fasting and 90 minutes after meals. Abnormal results triggered a 7-day BP or 1-day BG profile or a teleconsultation with a general practitioner. Meanwhile, the control group received routine care based on integrated care protocols at community health centers. Primary outcomes were the differences between groups in the change in systolic blood pressure (SBP) and HbA_{1c} levels at 12 months after inclusion from baseline. Secondary outcomes included changes in diastolic blood pressure, fasting BG, lipid profile, body mass index, appraisal of diabetes, and behavioral risk factors.

Results: Initially, 128 patients were enrolled, with 117 (91.4%) completing the 12-month follow-up. The mean age was 71.3 (SD 4.7) years, with a mean SBP of 136.7 (SD 14.1) mmHg and mean HbA_{1c} of 7.2% (SD 1.0%). There were no significant sociodemographic or clinical differences between groups at baseline. At 12 months, the telemonitoring group experienced significant reductions in SBP (−9.7 mmHg, 95% CI −12.6 to −6.8; P<.001) and HbA_{1c} (−0.5%, 95% CI −0.8 to −0.3; P<.001), whereas the control group exhibited nonsignificant changes in SBP (−2.8 mmHg, 95% CI −5.9 to 0.2; P=.07) and HbA_{1c} (0%, 95% CI −0.3 to 1.9; P=.75). The difference between groups at 12 months was significant for both SBP (−6.9 mmHg, 95% CI −11 to −2.7; P=.001) and HbA_{1c} (−0.5%, 95% CI −0.8 to −0.2; P=.002), with no significant differences observed in secondary outcomes.

Conclusions: Telemonitoring effectively improves AH and T2D control in older people but has no impact on other cardiovascular risk factors and diabetes-related quality of life. Future research should explore combining educational and behavioral interventions with telemonitoring to enhance overall health outcomes. However, complex interventions may pose challenges for the elderly, suggesting the need for careful patient selection to ensure that benefits outweigh potential burdens.

Trial Registration: ISRCTN Registry ISRCTN31471852; https://doi.org/10.1186/ISRCTN31471852

International Registered Report Identifier (IRRID): RR2-10.2478/sjph-2022-0029

(JMIR Mhealth Uhealth 2025;13:e59733) doi:10.2196/59733

XSL•FO
RenderX

KEYWORDS

mobile health; aged; HbA1c; blood pressure; primary health care; multimorbidity; sustainability; health education; patient-centered care

Introduction

The growing challenges in managing chronic diseases, specifically arterial hypertension (AH) and type 2 diabetes (T2D), highlight the need for innovative solutions. Recent changes in age distribution, health care approaches, and technological integration, along with projections of increased prevalence, particularly among the elderly, underscore the importance of developing effective strategies to address these health issues [1,2].

In response to these challenges, health care providers are increasingly adopting telemonitoring as a strategic approach to optimize integrated care processes and extend patient care beyond traditional health care settings. This approach entails the use of medical devices for the real-time collection of physiological data, including metrics such as blood pressure (BP) and blood glucose (BG) [3-6].

When coupled with teleconsultations, telemonitoring facilitates seamless communication, enabling timely adjustments to treatment plans. By emphasizing a proactive and collaborative paradigm, this approach contributes to the development of a well-informed and engaged patient population. This, in turn, has the potential to improve adherence to treatment plans and lifestyle modifications, ultimately leading to better health outcomes [4,7,8].

Combining telemonitoring and teleconsultations for at least 12 months optimizes outcomes in controlling BP and BG. After 12 months of telemonitoring, individuals with AH experienced additional mean reductions in systolic blood pressure (SBP) ranging from −4.7 to −9.7 mmHg and mean reductions in diastolic blood pressure (DBP) ranging from −1.3 to −5.1 mmHg [9-12]. For individuals with T2D, mean reductions in HbA_{1c} levels ranged from −0.03% to −0.33% at the 12-month mark [13-20]. However, the impact on lipid profile, body mass index (BMI) and quality of life was considered clinically irrelevant [13,15-18].

A notable limitation of previous research lies in the oversight of the central role of primary care in managing chronic diseases, overseeing nearly 80% of all cases of AH and T2D [6,21]. Clinical pathways, typically developed for larger clinical settings, have often overlooked the potential impact on smaller health centers. In addition, research studies have primarily focused on younger, motivated individuals dealing with isolated AH or T2D, neglecting older multimorbid patients [22].

Consequently, there is limited evidence on the clinical effectiveness of telemonitoring among the elderly in primary care. Older people face challenges such as multimorbidity, cognitive impairment, visual impairment, limited mobility, environmental dependency, and reluctance to adopt modern technologies, which may reduce its effectiveness [2,5,23,24]. Considering demographic shifts, it is imperative to evaluate the benefits of telemonitoring in this age group and to adapt interventions to address the specific needs of older people.

The objective of this study was to assess the impact of a 12-month telemonitoring intervention on clinical outcomes in older individuals with AH and T2D, comparing it with standard care in primary care settings.

Methods

Study Design

We conducted a multicenter, prospective, open-label, parallel, randomized controlled study as part of the SCUBY (Scale up an integrated care package for diabetes and hypertension for vulnerable people in Cambodia, Slovenia, and Belgium) project between March 2021 and June 2023. The study adhered to the published protocol [4] and was registered in the ISRCTN registry.

Ethical Considerations

The study followed the ethical standards of the Declaration of Helsinki, with ethical approval granted by the Medical Ethics Committee of the Republic of Slovenia (0120-219/2019/4) before initiation. Informed written consent was obtained from all participants before enrollment, encompassing both primary and secondary analyses. Participants were informed of their right to withdraw from the study at any time without consequence. To safeguard privacy and confidentiality, all data were fully anonymized, with each participant assigned a unique identification code upon enrollment. No compensation was offered for participation.

Study Setting

The research was conducted in 3 primary health care centers (PHCs) in Slovenia, each representing a different development context. PHC Ljubljana, serving a population of approximately 300,000 residents, presented an urban profile with a Gross Domestic Product (GDP) of 108% of the EU average in 2021. In contrast, PHC Trebnje and PHC Slovenj Gradec, serving approximately 50,000 residents in Eastern Slovenia, represented a rural setting with a GDP of 74% of the EU average in 2021 [25].

Study Population, Sampling Strategy, and Randomization

The study enrolled participants aged 65 years and older with both AH and T2D. Convenient sampling was used, with patients invited by their general practitioners (GPs) during regular checkups. Upon agreement, participants were randomly assigned in a 1:1 ratio to either the telemonitoring or standard care group. Randomization was performed by a third party independent of the study.

The inclusion criteria were (1) ≥65 years of age, (2) confirmed diagnosis of AH with a 7-day mean home BP values ≥135/85 mmHg, (3) confirmed diagnosis of T2D with fasting BG value

XSL•FO
RenderX

≥7 mmol/l or venous plasma glucose ≥11.1 mmol/l 2 hours after glucose tolerance test with 75 g glucose intake or any random opportunity, (4) diagnosis of AH and T2D for at least a year, and (5) ability to use telemonitoring equipment.

The exclusion criteria were (1) <65 years of age, (2) T2D requiring insulin treatment at inclusion, (3) gestational diabetes or type 1 diabetes, (4) dementia, and (5) inability to use telemonitoring equipment for various reasons.

Standard Care

In Slovenia, a team comprising a GP and a registered nurse delivers integrated care for patients with AH and T2D at the primary care level. The care follows established chronic integrated care protocols. Annual check-ups encompass physical examinations, laboratory tests, disease assessments, and treatment adjustments overseen by the GP. The nurse conducts screenings for complications, provides education on nonpharmacological measures, and refers patients to health promotion activities within health education centers staffed by nurses, nutritionists, psychologists, and kinesiologists. At least once a year, patients are also screened for diabetic retinopathy at retinopathy screening centers [4,26,27].

Telemonitoring Intervention

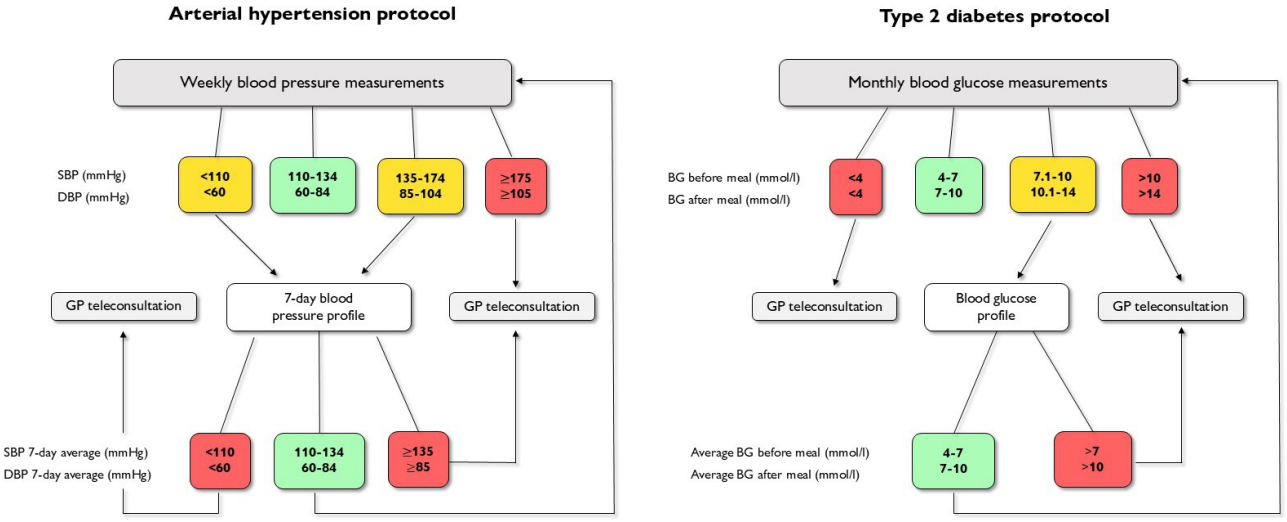
We supported standard care by introducing a telemonitoring intervention, detailed in another paper [4]. At enrollment, each patient received an hour of training from a nurse on the measurement protocol and how to use the telemonitoring package. The package included an LG K22 smartphone and monitors for BP (A&D 651 BLE) and BG (Contour Next One). Data from these sensors were transmitted to a mobile app through Bluetooth technology, acting as a central hub. The application then transferred results to the telemedicine cloud platform using a 4G or 5G mobile standard. All received data were encrypted and adhered to the General Data Protection Regulation (GDPR).

The telemedicine center was coordinated by a GP who reviewed the recorded values and acted as a link between included patients and their GPs. Patients were managed according to recent guidelines [1,28], aiming for home BP <135/85 mmHg, fasting BG <7 mmol/l, and postprandial BG <10 mmol/l, with treatment goals adjusted for the elderly.

Over the 12-month period, participants measured their BP twice a week and BG once a month. For BP, they took readings in the morning and evening, after resting for 5 minutes while seated with back and arm supported. A total of 2 readings were taken with a programmed 1-minute interval between them, with the second considered valid. For BG, they took two readings: one in the morning on an empty stomach and another 90 minutes after lunch.

The transmitted data were color-coded to indicate their status (see Figure 1). Green signaled normal values with no need to change the monitoring protocol. Yellow indicated abnormal values, triggering more intensive monitoring to better assess participants' health. This included a 7-day BP profile with morning and evening readings for AH or a 1-day BG profile with 6 readings for T2D (values on an empty stomach, 90 minutes after breakfast, before lunch, 90 minutes after lunch, before dinner, and 90 minutes after dinner). Red indicated critically abnormal readings that required immediate teleconsultation with a GP, further evaluation, or a change in treatment plan.

Figure 1. Mobile health care pathways for managing arterial hypertension and type 2 diabetes. BG: blood glucose; DBP: diastolic blood pressure; GP: general practitioner; SBP: systolic blood pressure.



Data Collection

Demographic and Behavioral Profile

Participants completed a structured questionnaire providing information on demographic and behavioral characteristics, including age, gender, education, marital status, smoking, daily activity level, salt intake habits, daily number of meals, daily vegetable consumption, and alcohol consumption.

For our study, we considered physical activity adequate if it exceeded 150 minutes of moderate-intensity or 75 minutes of high-intensity per week, aligning with the World Health Organization (WHO) recommendations [4].

XSL•FO
RenderX

For alcohol consumption, we used the Slovenian version of the Alcohol Use Disorders Identification Test (AUDIT-C) to identify hazardous drinking habits. Scores over 6 points for men and 5 points for women indicated hazardous drinking habits [4].

Clinical History

Clinical history data included duration of AH and T2D, current treatment regimen, and accompanying medical conditions. Data were initially collected from patients and later verified in medical records.

Chronic kidney disease was defined according to the Kidney Disease Improving Global Outcomes (KDIGO) criteria as abnormalities in kidney structure (ie, albuminuria, urine sediment abnormalities, structural abnormalities detected by imaging, and history of kidney transplantation) or function (glomerular filtration rate [GFR] <60 ml/min/1.73 m^2) that persist for at least 3 months and have significant implications for health [29].

Obesity was defined according to the WHO criteria as BMI ≥30 kg/m^2 [30].

Systolic and Diastolic Blood Pressure

SBP and DBP values were measured at baseline, 6 months, and 12 months using validated monitors (A&D 651 BLE). Patients took BP readings every morning and evening, with two readings 1 minute apart. The second morning and evening reading was averaged over 7 days to establish the reference value. For telemonitored patients, data were automatically transmitted to the telemonitoring platform, while the control group manually recorded their readings. If data were missing, patients were contacted to repeat the measurements.

Glycemic Control, Lipid Profile, and Glomerular Filtration Rate

HbA_{1c} level, fasting BG value, lipid profile, and glomerular filtrate rate values were determined from peripheral venous blood sampling at baseline and after 12 months. In addition, fasting BG value was checked after 6 months using a validated finger-stick BG monitor (Contour Next One).

Body Mass Index

Body mass index was assessed at baseline and after 12 months with the measurement of body weight and height.

Quality of Life

Diabetes-related quality of life was measured using the Slovenian Appraisal of Diabetes Scale (ADS-S) at baseline and after 12 months. This 7-item questionnaire assesses distress, diabetes control, uncertainty, coping, and life interference, with lower scores indicating a more positive appraisal strategy [31].

Outcomes

The primary outcomes were the differences between groups in the change in SBP and HbA_{1c} levels at 12 months after inclusion from baseline.

Secondary outcomes were differences between groups in a change in DBP, fasting BG, lipid profile, GFR, BMI, and ADS score up to 12 months after inclusion from baseline. In addition, we explored changes in selected behavioral risk factors at 12 months after inclusion from baseline (exploratory analysis).

Sample Size and Power of the Research

Before the study began, we set a total sample size of 120 participants, with 60 in the intervention group, due to limited telemonitoring equipment. Consequently, rather than calculating the sample size required to demonstrate a clinically significant reduction (10 mmHg in SBP and 0.5% in HbA_{1c}), we determined the minimum detectable difference between groups based on this planned sample size [4].

With this sample size, the study had 80% statistical power at a 0.05 significance level to detect differences in SBP changes between the telemonitoring and standard care groups at 12 months from baseline using an independent samples *t* test when the population mean difference between groups was at least 6.2 mmHg, assuming an SD of 12 in both groups, or 10.8 mmHg with a SD of 21 [10,17,32]. For changes in HbA_{1c} at 12 months from baseline, this was true when the population mean difference between groups was at least 0.7% (assuming an SD of 1.3) or 1.2% (assuming an SD of 2.4) [14,18,19,33].

Statistical Analysis

Initially, sample distribution characteristics were assessed using the Shapiro-Wilks test. Numerical variables were summarized with mean (SD), or with median (IQR) in case of an asymmetric distribution. Categorical variables were summarized with frequencies and percentages. Between-group differences were explored using the independent samples *t* test for numerical variables with a normal distribution and the Mann-Whitney *U*-test for variables with a non-normal distribution. For categorical variables, the χ^2 test was used. Within-group differences were explored using the paired-samples *t* test for numerical variables and McNemar's test for categorical variables. We used the Holm method to adjust *P* values for multiple comparisons when reporting secondary outcomes.

Furthermore, we used mixed ANOVA analysis to examine the influence of 2 key factors—time and group—on the outcomes measured in 3 time points, namely SBP, DBP, and fasting BG. This allowed us to explore how these variables change over time within the same subjects (time factor) and how these changes differ between telemonitoring and standard care groups (group factor) within one model.

An (adjusted) $P<.05$ was considered as statistically significant. All analyses were performed using IBM SPSS Statistics for Windows (version 25.0).

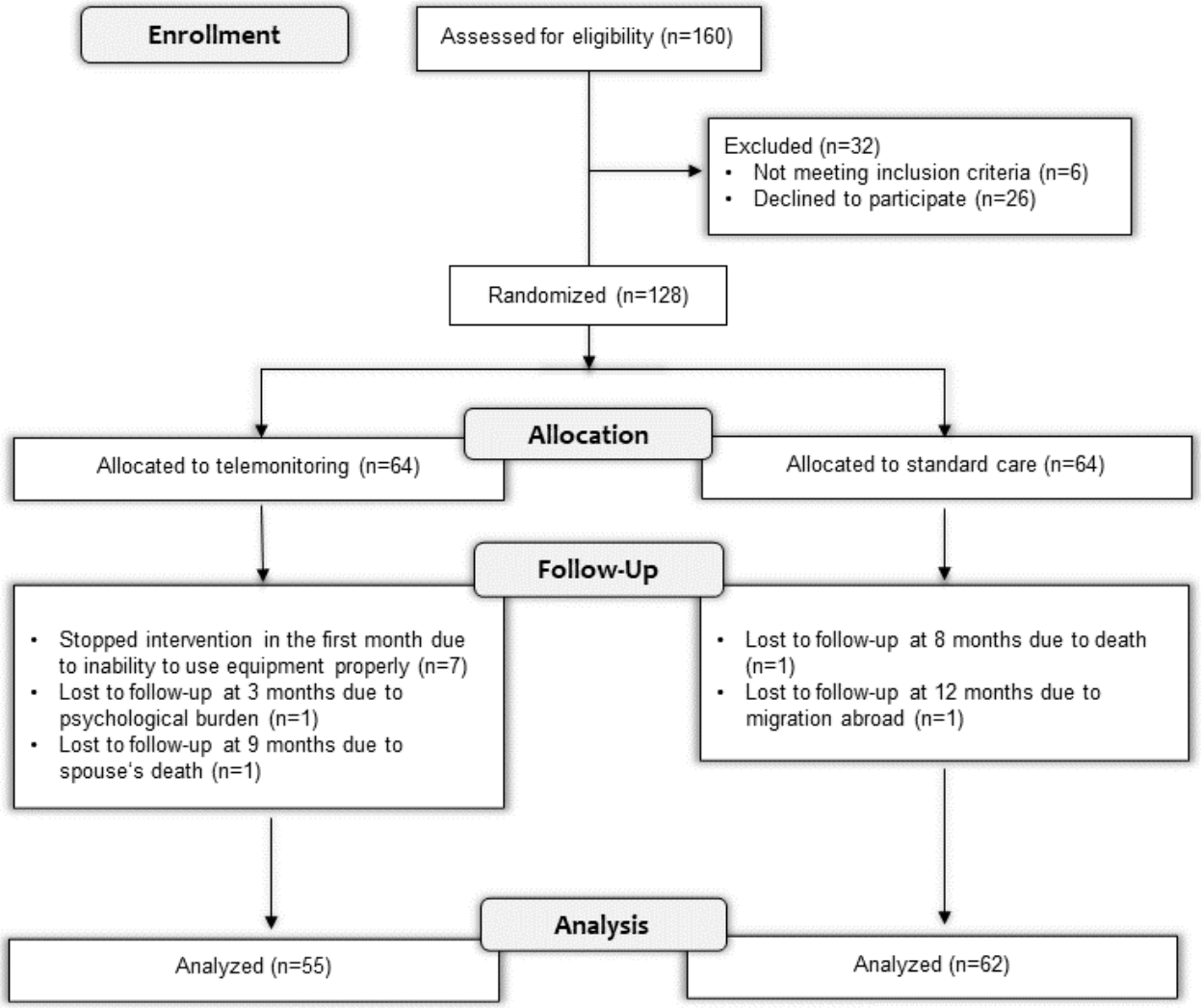
Results

Randomization

Figure 2 shows the randomization process. Among the 128 patients assigned to either the telemonitoring or standard care groups, 93.8% (120) attended the 6-month follow-up, and 91.4% (117) attended the 12-month follow-up.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram showing randomization process.



Sample Characteristics

The study included 117 participants, with a mean age of 71.3 (SD 4.7 years; range 65 - 88), of whom 60.7% (71/117) were male. Most patients had completed primary or vocational school education, were married, engaged in regular physical activity, were nonsmokers, and did not report hazardous drinking habits or a tendency to add extra salt at the table. There were no significant differences between groups in terms of their sociodemographic and clinical characteristics (Table 1).

For the treatment of AH, most patients received ACE inhibitors (64/117, 54.7%), followed by calcium channel blockers (53/117, 45.3%), diuretics (51/117, 43.6%), beta blockers (34/117, 29.1%), angiotensin II receptor blockers (23/117, 19.7%), and alpha-1 antagonists (8/117, 6.8%). Conservative measures were used in 10.3% (12/117) of cases. There were no significant differences between the groups in terms of their AH treatment (see Table 2).

For the treatment of T2D, most patients received metformin (77/117, 65.8%), followed by sulphonyl urea (31/117, 26.5%), SGLT-2 inhibitor (17/117, 14.5%), DPP-4 inhibitor (6/117, 5.1%), and GLP-1 agonist (3/117, 2.6%). Conservative measures were used in 26.5% (31/117) of cases. Patients receiving standard care were significantly more commonly treated with DPP-4 inhibitor compared to telemonitoring group, whereas there were no significant differences in terms of other treatment regimens (see Table 2).

XSL•FO
RenderX

Table . Baseline sociodemographic and clinical characteristics across groups.

Characteristic	All patients (n=117)	Telemonitoring group (n=55)	Standard care group (n=62)	*P* value
Age, years, mean (SD)	71.3 (4.7)	70.6 (4.3)	72 (5)	.12
Gender				.81
Male, n (%)	71 (60.7)	34 (61.8)	37 (59.7)	
Female, n (%)	46 (39.3)	21 (38.2)	25 (40.3)	
Education				.52
Primary care, n (%)	18 (15.4)	7 (12.7)	11 (17.7)	
Vocational school, n (%)	69 (59)	33 (60)	36 (58.1)	
High school, n (%)	16 (13.7)	10 (18.2)	6 (9.7)	
Bachelor's degree, n (%)	10 (8.5)	3 (5.5)	7 (11.3)	
Master's degree, n (%)	4 (3.4)	2 (3.6)	2 (3.2)	
Region				.55
Urban, n (%)	54 (46.2)	27 (49.1)	27 (43.6)	
Rural, n (%)	63 (53.8)	28 (50.9)	35 (56.4)	
Marital status				.53
Married, n (%)	86 (73.5)	41 (74.5)	45 (72.6)	
Divorced, n (%)	8 (6.8)	2 (3.6)	6 (9.7)	
Widowed, n (%)	5 (4.3)	2 (3.6)	3 (4.8)	
Single, n (%)	18 (15.4)	10 (18.2)	8 (12.9)	
Duration of T2D[a] (years), mean (SD)	9.5 (7.6)	9.8 (6.3)	9.2 (8.6)	.67
Duration of AH[b] (years), mean (SD)	14.5 (10.6)	13.7 (10.6)	15.2 (10.6)	.46
Systolic blood pressure (mmHg), mean (SD)	136.7 (14.1)	135.8 (14.9)	137.7 (13.1)	.47
Diastolic blood pressure (mmHg), mean (SD)	76.5 (8.2)	75.6 (7)	77.2 (9)	.28
Glycated hemoglobin (%), mean (SD)	7.2 (1)	7.2 (1.2)	7.1 (0.8)	.41
Non-HDL[c] (mmol/l), mean (SD)	3.2 (0.9)	3.1 (0.9)	3.3 (0.9)	.31
Body mass index (kg/m^2), mean (SD)	30.4 (4.8)	30.2 (4.6)	30.5 (5.1)	.70
Glomerular filtration rate, (ml/min/1.73 m^2), mean (SD)	77.9 (14)	80 (13.8)	76 (14.1)	.13
Chronic kidney disease, n (%)	24 (20.5)	9 (16.4)	15 (24.2)	.30
Obesity, n (%)	56 (47.9)	26 (47.3)	30 (48.4)	.90
Regular physical activity, n (%)	84 (71.8)	42 (76.3)	42 (67.7)	.30
Current smoker, n (%)	14 (12)	9 (16.3)	5 (8)	.17
Hazardous drinking habits, n (%)	16 (13.7)	10 (18.2)	6 (9.7)	.18
Salting at the table, n (%)	12 (10.3)	7 (12.7)	5 (8.1)	.39

XSL•FO
RenderX

Characteristic	All patients (n=117)	Telemonitoring group (n=55)	Standard care group (n=62)	*P* value
Appraisal of diabetes score, mean (SD)	15.8 (3.6)	15.8 (3.3)	15.9 (3.4)	.84

[a]T2D: type 2 diabetes.

[b]AH: arterial hypertension.

[c]Non-HDL: non–high-density lipoprotein.

Table . Baseline treatment regimens across groups.

	All patients (n=117)	Telemonitoring group (n=55)	Standard care group (n=62)	*P* value
Treatment of arterial hypertension, n (%)				
ACE[a] inhibitor	64 (54.7)	28 (50.9)	36 (58.1)	.44
Angiotensin II receptor blocker	23 (19.7)	12 (21.8)	11 (17.7)	.58
Diuretic	51 (43.6)	21 (38.2)	30 (48.4)	.27
Calcium channel blocker	53 (45.3)	23 (41.8)	30 (48.4)	.48
Beta blocker	34 (29.1)	14 (25.5)	20 (32.3)	.42
Alpha-1 antagonist	8 (6.8)	5 (9.1)	3 (4.8)	.36
Conservative	12 (10.3)	8 (14.5)	4 (6.5)	.15
Number of antihypertensive medication classes, median (IQR)	2 (1-3)	2 (1-3)	2 (1-3)	.32
Treatment of type 2 diabetes, n (%)				
Metformin	77 (65.8)	36 (65.5)	41 (66.1)	.94
Sulphonyl urea	31 (26.5)	13 (23.6)	18 (29)	.51
GLP-1[b] agonist	3 (2.6)	1 (1.8)	2 (3.2)	.63
SGLT-2[c] inhibitor	17 (14.5)	8 (14.5)	9 (14.5)	>.99
DPP-4[d] inhibitor	6 (5.1)	0 (0)	6 (9.7)	.02
Conservative	31 (26.5)	15 (27.3)	16 (25.8)	.86
Number of antidiabetic medication classes, median (IQR)	1 (0 - 2)	1 (0 - 2)	1 (0 - 2)	.45

[a]ACE: angiotensin converting enzyme.

[b]GLP-1: glucagon-like peptide-1.

[c]SGLT-2: sodium-glucose transport protein 2.

[d]DPP-4: dipeptidyl peptidase 4.

Primary Outcomes

The telemonitoring group demonstrated significant reductions in both SBP and HbA_{1c} levels compared to the baseline up to 12 months after enrollment (see Table 3). The mean differences between telemonitoring and control group for SBP change from baseline were –10.5 mmHg (95% CI −14.6 to −6.5; *P*<.001) at 6 months and −6.9 mmHg (95% CI −11 to −2.7; *P*=.001) at 12 months. For HbA_{1c} change from baseline, the mean intergroup difference at 12 months was −0.5% (95% CI −0.8 to −0.2; *P*=.002).

Mixed ANOVA indicated significant temporal changes in SBP across both groups ($F_{2,114}$=26.93; *P*<.001; η^2=0.19). In addition, a significant time-group interaction ($F_{2,114}$=15.08; *P*<.01; η^2=0.12) highlighted differing effects on SBP between telemonitoring and control groups.

XSL•FO
RenderX

Table . Change in primary outcomes over 12 months with effect sizes.

Group	Baseline		6 months		12 months		Mean 6-month change within groups (95% CI)	*P* value	Mean 12-month change within groups (95% CI)	*P* value	Mean 6-month change between groups (95% CI)	*P* value	Mean 12-month change between groups (95% CI)	*P* value
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)								
Systolic blood pressure (mmHg)														
Intervention	55	135.8 (14.9)	55	124.7 (11.2)	55	126.2 (10.4)	–11.2 (–14.1 to –8.3)	<.001	–9.7 (–12.6 to –6.8)	<.001	–10.5 (–14.6 to –6.5)	<.001	–6.9 (–11 to –2.7)	.001
Control	62	137.7 (13.1)	62	137.1 (9.4)	61	134.9 (9.6)	–0.7 (–3.5 to 2.2)	.66	–2.8 (–5.9 to 0.2)	.07		N/A[a]		N/A
Glycated hemoglobin (%)														
Intervention	55	7.2 (1.2)	N/A	N/A	55	6.7 (0.8)	N/A	N/A	–0.5 (–0.8 to –0.3)	<.001	N/A	N/A	–0.5 (–0.8 to –0.2)	.002
Control	62	7.1 (0.8)	N/A	N/A	55	6.7 (0.8)	N/A	N/A	0 (–0.3 to 1.9)	.75		N/A		N/A

[a]N/A: Not assessed.

Secondary Outcomes

The difference between groups in fasting BG change from baseline was −1 mmol/l at 6 months (95% CI −1.5 to −0.4; *P*<.001) and −0.8 mmol/l at 12 months (95% CI −1.4 to −0.2; *P*=.006). However, after adjustment for multiple comparisons, the 12-month result narrowly missed statistical significance with *P*=.06.

Although telemonitoring group demonstrated statistically significant reductions in DBP at 6 and 12 months, they were not substantial. The difference between groups in DBP change from baseline was −3.4 mmHg (95% CI −6 to −0.9; *P*=.007) at 6 months and −1.7 mmHg (95% CI −4.2 to 0.7; *P*=.15) at 12 months with adjusted *P*=.81.

In terms of other secondary outcomes, there was a significant between-group difference in low-density lipoprotein (LDL) levels after 12 months compared to baseline (0.3 mmol/l, 95% CI 0.1 to 0.6), while between-group differences in total cholesterol, high-density lipoprotein, triglycerides, BMI, ADS score, and GFR were not significant. However, after adjusting *P* values for multiple comparisons, the difference in LDL was also not significant (see Table 4).

Finally, in exploratory assessment of changes in behavioral risk factors control there were no significant changes in either group (Table 5).

XSL•FO
RenderX

Table . Change in secondary outcomes over 12 months with effect sizes.

Group	Baseline		6 months		12 months		Mean 6-month change within groups (95% CI)	*P* value	Mean 12-month change within groups (95% CI)	*P* value	Mean 6-month change between groups (95% CI)	*P* value	Mean 12-month change between groups (95% CI)	*P* value	*P* value (adjusted for 12-month change between groups[a]
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)									
Diastolic blood pressure (mmHg)															
Intervention	55	75.6 (7)	55	71.2 (6.4)	55	71.6 (5.8)	−4.4 (−5.7 to −3)	<.001	−4 (−5.4 to −2.6)	<.001	−3.4 (−6 to −0.9)	.007	−1.7 (−4.2 to 0.7)	.15	.81
Control	62	77.2 (9)	62	76.3 (8.1)	61	75 (7.6)	−0.9 (−3 to 1.2)	.38	−2.3 (−4.2 to −0.3)	.02		N/A[b]		N/A	
Fasting blood glucose (mmol/l)															
Intervention	55	8.3 (1.5)	55	7.3 (1.3)	55	7.5 (1.6)	−1 (−1.3 to −0.7)	<.001	−0.8 (−1.1 to −0.6)	<.001	−1 (−1.5 to −0.4)	<.001	−0.8 (−1.4 to −0.2)	.006	.06
Control	62	7.9 (1.4)	60	8 (1.9)	61	8 (2.2)	0 (−0.4 to 0.4)	.91	+0.1 (−0.5 to 0.6)	.76		N/A		N/A	
Total cholesterol (mmol/l)															
Intervention	55	4.5 (1)	N/A	N/A	55	4.5 (1.1)	N/A	N/A	0 (−0.2 to 0.2)	.95	N/A	N/A	+0.3 (−0.1 to 0.6)	.12	.81
Control	61	4.7 (1)	N/A	N/A	61	4.3 (1)	N/A	N/A	-0.4 (−0.6 to −0.1)	.002		N/A		N/A	
Low-density lipoprotein (mmol/l)															
Intervention	55	2.5 (0.7)	N/A	N/A	55	2.6 (1)	N/A	N/A	+0.1 (−0.1 to 0.3)	.48	N/A	N/A	+0.3 (0.1 to 0.6)	.03	.21
Control	60	2.7 (0.7)	N/A	N/A	61	2.4 (0.8)	N/A	N/A	−0.3 (−0.5 to −0.1)	.002		N/A		N/A	
High-density lipoprotein (mmol/l)															
Intervention	55	1.4 (0.3)	N/A	N/A	55	1.3 (0.3)	N/A	N/A	-0.1 (−0.2 to −0.1)	<.001	N/A	N/A	0.0 (−0.1 to 0.1)	.68	≥.99
Control	61	1.4 (0.5)	N/A	N/A	61	1.3 (0.4)	N/A	N/A	−0.1 (−0.2 to −0.1)	<.001		N/A		N/A	
Triglycerides (mmol/l)															

XSL•FO
RenderX

Group	Baseline		6 months		12 months		Mean 6-month change within groups (95% CI)	*P* value	Mean 12-month change within groups (95% CI)	*P* value	Mean 6-month change between groups (95% CI)	*P* value	Mean 12-month change between groups (95% CI)	*P* value	*P* value (adjusted for 12-month change between groups[a]
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)									
Intervention	55	1.6 (0.7)	N/A	N/A	55	1.5 (0.7)	N/A	N/A	−0.1 (−0.2 to 0.1)	.38	N/A	N/A	−0.2 (−0.5 to 0.1)	.29	≥.99
Control	61	1.6 (0.8)	N/A	N/A	61	1.7 (1.3)	N/A	N/A	+0.1 (−0.2 to 0.3)	.54		N/A		N/A	
Body mass index (kg/m^2)															
Intervention	54	30.2 (4.6)	N/A	N/A	54	30.6 (4.9)	N/A	N/A	+0.3 (−0.1 to 0.8)	.10	N/A	N/A	+0.5 (-0.1 to 1)	.12	.81
Control	59	30.5 (5.1)	N/A	N/A	59	30.4 (4.9)	N/A	N/A	−0.1 (−0.5 to 0.3)	.56		N/A		N/A	
Glomerular filtration rate (ml/min/173 m^2)															
Intervention	54	80 (13.8)	N/A	N/A	55	78.5 (13.5)	N/A	N/A	−1.7 (−3.1 to −0.2)	.03	N/A	N/A	0 (−4.8 to 4.9)	.98	≥.99
Control	61	76 (14.1)	N/A	N/A	62	74.7 (13.8)	N/A	N/A	−1.5 (−3 to −0.1)	.049		N/A		N/A	
Appraisal of diabetes score															
Intervention	55	15.8 (3.3)	N/A	N/A	55	15.9 (3)	N/A	N/A	+0.1 (−0.9 to 1.1)	.88	N/A	N/A	+0.2 (−1.2 to 1.5)	.80	≥.99
Control	62	15.9 (6.4)	N/A	N/A	60	15.9 (3.4)	N/A	N/A	−0.1 (−1 to 0.8)	.83		N/A		N/A	

[a]Adjusted *P* values for multiple comparisons between groups at 12 months regarding all secondary outcomes in Table 4.

[b]Not assessed; indicates that at this time point the secondary outcome was not assessed. We assessed it only at baseline and after 12 months as per study protocol.

XSL•FO
RenderX

Table . Change in behavioral risk factors over 12 months with effect sizes (exploratory analysis).

Group	Baseline		12 months		Relative 12-month change within groups, %	*P* value
	n[a] (%)	n[b]	n[a] (%)	n[b]		
Regular physical activity						
Intervention	42 (76.3)	55	36 (65.5)	55	−10.8	.11
Control	42 (67.7)	62	32 (54.2)	59	−13.5	.07
Harmful drinking habits						
Intervention	7 (13)	55	8 (14.5)	55	+1.5	≥.99
Control	5 (8.1)	62	5 (8.3)	60	+0.2	.75
Salting at the table						
Intervention	10 (18.2)	54	6 (11.1)	54	−7.1	≥.99
Control	6 (9.7)	62	4 (6.8)	59	−2.9	≥.99
Having 3 - 5 meals/day						
Intervention	49 (89.1)	55	48 (87.3)	55	−1.8	≥.99
Control	59 (95.2)	62	56 (93.3)	60	−1.9	≥.99
Eating vegetables daily						
Intervention	39 (70.9)	55	37 (67.3)	55	−3.6	.83
Control	44 (71)	62	38 (63.3)	60	−7.7	.49
Smoking						
Intervention	9 (16.4)	55	11 (20)	55	+3.6	≥.99
Control	5 (8.1)	62	8 (13.3)	60	+5.2	≥.99

[a] Represents the number of participants at inclusion.

[b]Represents the number of participants remaining after 12 months.

Discussion

Principal Findings and Comparison With the Existing Literature

The major finding of this study is that telemonitoring has been found to be a valuable strategy for improving BP and BG control in older people with AH and T2D in primary care settings. However, there were no significant improvements in control over other cardiovascular risk factors or diabetes-related quality of life.

The difference in SBP change between groups was significant with a decrease of −10.5 mmHg (95% CI −14.6 to −6.5) after 6 months and −6.9 mmHg (95% CI −11 to −2.7) after 12 months compared to the control group. From a clinical standpoint, a 6-month SBP reduction correlates with a 20% lower risk of major cardiovascular events, whereas a 12-month SBP reduction correlates to a 10% - 15% lower risk of cardiovascular events [33]. The observed 12-month SBP difference aligns with findings from previous studies, where a similar difference at 12 months ranged from −4.7 to −9.7 mmHg in patients with AH [10,34] and from −0.2 to −8.2 mmHg in patients with AH and T2D [11,17,18].

In contrast, the difference in DBP change between groups was significant after 6 months, with a decrease of −3.4 mmHg (95% CI −6 to −0.9) but became nonsignificant after 12 months, showing a decrease of −1.7 mmHg (95% CI −4.2 to +0.7). Similar patterns were observed in previous studies, where the DBP between-group difference at 12 months ranged from −1.3 to −5.1 mmHg in patients with AH [10,34], and from −0.4 to −0.7 mmHg in patients with AH and T2D [17,18]. Clinically, both outcomes are considered irrelevant, failing to meet the 5 mmHg DBP decrease threshold associated with a 20% reduction in the risk of major cardiovascular events [35].

In terms of glycemic control, we found a significant between-group difference in HbA_{1c} change of −0.5% (95% CI −0.8 to −0.2) after 12 months. In addition, a clinically meaningful between-group difference in fasting BG of −0.8 mmol/l (95% CI −1.4 to −0.2) was found after 12 months. These findings contrast with previous studies, where HbA_{1c} reductions at 12 months were less pronounced, ranging from −0.03% to −0.32% in patients with isolated T2D [13-16,19] and from +0.14 to −0.33% in patients with both AH and T2D [11,17,18,36]. Recent meta-analysis also indicates that telemonitoring in primary care produced a smaller HbA_{1c} reduction of −0.21% (95% CI −0.37 to −0.06) in patients with AH and T2D after 12 months [37]. A 0.5% reduction in HbA_{1c} is clinically significant and is associated with an estimated 17% reduction in cardiovascular disease risk. The extent of risk reduction is,

XSL•FO
RenderX

however, curvilinear and dependent on baseline HbA_{1c} status and treatment regimens [38,39].

Furthermore, there were no significant differences between groups in terms of lipid profile, BMI, and GFR after 12 months. This aligns with earlier studies in patients with T2D and/or AH, where intergroup reductions for LDL ranged from −0.16 to +0.04 mmol/l [13,15-18,36]; for triglycerides, it ranged from −0.15 to +0.07 mmol/l [13,16-19]; for BMI, it remained stable or slightly increased [15,17,18]; and for GFR, it improved up to 3 ml/min/1.73 m^2 [40].

Ultimately, telemonitoring alone did not significantly improve diabetes-related quality of life or lead to positive lifestyle changes. This finding is important, as patients achieved significant BP and BG control without making meaningful lifestyle changes. The positive outcomes may be attributed to other factors such as improved communication, medication adherence, early detection of exacerbations, and therapy adjustments based on accurate telemonitoring data [5,8,41]. However, this finding should be interpreted with caution as lifestyle changes were self-reported by participants. In-depth interviews with 29 telemonitoring participants provided several examples of reported lifestyle changes and a better understanding of how lifestyle choices affect BG control [42]. Notably, our intervention lacked a health education cointervention. Previous studies incorporating such interventions reported significant improvements in physical activity, alcohol intake, and diet quality, while smoking cessation rates remained unchanged [42-44].

Implications for Practice and Sustainability

Establishing the Gold Standard for Telemonitoring Frequency

In our study, patients monitored their BP twice weekly and BG once monthly, with more frequent checks in case of derailments, resulting in a significant reduction in SBP and HbA_{1c} levels. Nevertheless, uncertainty persists regarding the optimal telemonitoring routine. Low-intensity BP monitoring, such as once weekly, boasts high adherence but offers less significant long-term benefits [45,46], while high-intensity BP monitoring, such as 5 to 6 times weekly, demonstrates better long-term outcomes but lower adherence [46-48]. When monitoring BG levels in insulin-naive patients, 1 measurement per week was found as effective as four measurements per week to maintain HbA_{1c} [49]. In a previous feasibility study [6], we found that less frequent monitoring (ie, 7-day BP profile and one-day BG profile at baseline, 2 weeks, 3 months, 6 months) was clinically irrelevant, leading to a more intensive approach in this study. However, even in this study patients exceeded the expected number of BP and BG measurements by 46.5% and 252.8%, respectively, which resulted in total mean of 2.7 BP measurements per week and 1.2 BG measurements per week [8]. Accordingly, based on results of this study and previous studies, a regimen in which BP is measured two times weekly with morning and evening readings and BG once weekly with pre- and postprandial readings could be considered as one of the standard regimens for future research.

Bridging Gaps in Knowledge and Long-Term Lifestyle Modification

Our study aimed to improve BP and BG control through telemonitoring, excluding health education training, which is freely available at health education centers. Despite significant improvements in primary outcomes, there were no significant changes in behavioral risk factors, highlighting the need for a more comprehensive approach. Integrating telemonitoring with tailored health education and behavioral interventions such as group sessions, individual education, or telehealth tools such as phone counseling, text messaging, health applications, and web-based resources could increase effectiveness [50]. In addition, peer support groups may be a valuable alternative, allowing individuals to share experiences, provide mutual encouragement, and foster a community-based approach to health management [51].

Personalizing mHealth Interventions for Elderly and Multimorbid Individuals

Telemonitoring introduces an external monitoring system that may be perceived as burdensome by older patients and their caregivers, potentially leading to unintentional dropouts. In our study, older participants encountered technical challenges such as difficulty with touch screens, forgotten passwords, and device pairing. To enhance the acceptability for this demographic, telemonitoring devices should include larger screens, clear icons, voice commands, and alternative login options such as fingerprints or lock patterns. In addition, transition to continuous glucose monitoring systems may be beneficial for insulin-dependent patients [5,6,42].

For multimorbid patients, telemonitoring should take a holistic approach that considers the interplay between different conditions, medications, and treatments. Conducting an early patient assessment focused on health risks, technological literacy, and personal preferences can help tailor interventions and improve acceptability [5,6,23,42]. Integrating comprehensive monitoring of multiple health metrics, such as weight and activity levels, along with multidisciplinary support from professionals like nutritionists and kinesiologists, could further improve patient outcomes [52].

Optimizing Telemonitoring Costs for Sustainable Health Care Integration

For the effective and sustainable integration of telemonitoring in health care, optimizing key cost drivers is crucial. In our parent study, we conducted a bottom-up analysis to identify these factors and define variables for future cost-effectiveness analyses [8].

To comprehensively assess costs, we considered both health care provider and patient perspectives. For providers, this included infrastructure and operational costs, while for patients, it involved participation costs and changes in out-of-pocket costs. The telemonitoring intervention incurred an annual infrastructure cost of €489.4 and operational costs of €97.3 (95% CI 85.7 to 109) per patient. For patients, annual participation costs amounted to €215.6 (95% CI 190.9 to 241.1), with average annual out-of-pocket costs for both groups reported at €345 (95% CI 221 to 469). Notably, after 12 months, the

XSL•FO
RenderX

telemonitoring group experienced significantly lower out-of-pocket costs (€132 vs €545; *P*<.001), largely due to reduced spending on food, dietary supplements, medical equipment, and specialist checkups compared to the standard care group [8]. All euro (€) amounts are based on the exchange rate at the time of the study, with €1 equivalent to approximately US $1.10.

One potential cost-reduction strategy involves shortening the monitoring period to focus on critical periods without compromising patient outcomes. In the 12-month follow-up, significant BP and BG improvements were seen within the first 6 months, with minimal changes afterward. This pattern aligns with previous pilot study findings [6]. Furthermore, patients reported monitoring fatigue due to regular measurements [6,42,53]. For these reasons, we propose a future model where telemonitoring is discontinued after 6 months and replaced by a less intensive self-monitoring protocol, with monthly nurse practitioner teleconsultations to review results and restart telemonitoring if needed due to exacerbations [6,10].

Furthermore, establishing a national platform could standardize infrastructure and reduce third-party costs. Developing a health app that patients can download on their personal phones, rather than providing them with additional devices for telemonitoring could further reduce equipment expenses while ensuring accessibility, however, this may raise privacy and security risks [54]. Integrating artificial intelligence into telemonitoring systems could improve data analysis, automate routine tasks, and decrease staffing requirements [55]. Finally, employing nurse practitioners as telemedicine coordinators offers a more cost-effective option than specialist-led models, allowing specialists to focus on more complex cases [5,6,8].

Strengths and Limitations

The strength of this study is highlighted by its multicenter, randomized controlled design, the complexity of the outcomes observed, and the novelty of the population studied. In addition, we adapted the inclusion criteria to include patients with varying degrees of difficulty and control of their AH and T2D. This deliberate approach aimed to capture a diverse population typically managed at the primary care level, thereby increasing the transferability of our findings to real-world settings.

However, it is important to consider limitations that may affect the generalizability of our findings. These include convenience sampling, the exclusion of patients requiring insulin because they are not commonly treated in primary care settings, and the reliance on self-reported behavioral data, which may affect the reliability of lifestyle-related outcomes. The inclusion criterion requiring participants to operate telemonitoring devices may have introduced a bias toward individuals with greater technical skills. However, despite training, some patients experienced technical challenges, suggesting possible limitations in practical implementation. Next, we did not apply advanced imputation methods for missing data due to the dataset's high quality and minimal missing values. For full transparency, we reported the actual sample size for each analysis and made the dataset freely accessible online. Finally, the study primarily focused on short-term outcomes, neglecting the importance of long-term (>12 months) considerations for future expansion.

Conclusions

The 12-month telemonitoring intervention effectively improved BP and BG control in older people with AH and T2D. Despite this improvement, it did not lead to better management of other cardiovascular risk factors or an improvement in diabetes-related quality of life. To improve overall health outcomes, future research should explore integrating educational and behavioral components into telemonitoring. Furthermore, there is a need to investigate new models of remote care, such as combining short-term telemonitoring with long-term self-measurement routines while maintaining regular teleconsultations. However, the complexity of these interventions may be challenging for the elderly, highlighting the importance of careful patient selection to ensure that the benefits outweigh potential burdens.

Acknowledgments

The authors would like to thank the participating community health centers and health workers for their help in recruiting patients and cooperating with the telemedicine center: Tatjana Lekšan, Ana Perdih, Andrej Škrinjar, Tamara Fras-Stefan, Andreja Beg Krasnič, Barbara Mazej Poredoš, Elizabeta Žlajpah, Špela Jenkole, Janja Glavan, Sonja Peternel, Branka Rajter, and Mojca Krajnc. The research was financially supported by the SCUBY project (Scale up an integrated care package for diabetes and hypertension for vulnerable people in Cambodia, Slovenia, and Belgium), an international research initiative co-financed by the European Union through the H2020 - Health Program (H2020-SC1) and identified by contract 825432 - SCUBY. The authors also acknowledge the financial support from the Slovenian Research Agency (research core funding P3-0339).

Data Availability

The datasets generated or analyzed during this study are available in the Harvard Dataverse repository [56].

Conflicts of Interest

None declared.

Checklist 1
CONSORT eHeatlh checklist.

XSL•FO
RenderX

[PDF File, 1226 KB - mhealth_v13i1e59733_app1.pdf]

References

1. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2021 Sep 7;42(34):3227-3337. [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
2. Mihevc M, Virtič Potočnik T, Zavrnik Č, Šter MP, Klemenc-Ketiš Z, Poplas Susič A. Beyond diagnosis: investigating factors influencing health-related quality of life in older people with type 2 diabetes in Slovenia. Prim Care Diabetes 2024 Apr;18(2):157-162. [doi: 10.1016/j.pcd.2024.01.010] [Medline: 38320938]
3. Mihevc M, Podgoršek D, Gajšek J, et al. A remote care model for patients at high risk of hospital admission due to COVID-19 deterioration: who makes it at home? - a multicenter follow-up case from Slovenia. Croat Med J 2023 Jun 30;64(3):170-178. [doi: 10.3325/cmj.2023.64.170] [Medline: 37391914]
4. Mihevc M, Zavrnik Č, Mori Lukančič M, et al. Telemonitoring of elderly with hypertension and type 2 diabetes at the primary care level: protocol for a multicentric randomized controlled pilot study. Zdr Varst 2022 Dec;61(4):216-223. [doi: 10.2478/sjph-2022-0029] [Medline: 36348965]
5. Mihevc M, Podgoršek D, Gajšek J, et al. The rise of telemedicine in primary care: understanding patients' and healthcare workers' perspectives on acceptability of the COVID-19 remote care model. fmpcr 2023;25(3):297-301. [doi: 10.5114/fmpcr.2023.130091]
6. Mihevc M, Puntar Š, Peter Šter M. Telemonitoring of patients with arterial hypertension and/or type 2 diabetes in a family medicine practice: results of a mixed-methods pilot project. Med Razgl 2023;62(1):3-16.
7. Bingham JM, Black M, Anderson EJ, et al. Impact of telehealth interventions on medication adherence for patients with type 2 diabetes, hypertension, and/or dyslipidemia: a systematic review. Ann Pharmacother 2021 May;55(5):637-649. [doi: 10.1177/1060028020950726] [Medline: 32815400]
8. Mihevc M, Zavrnik Č, Mori Lukančič M, et al. Bottom-up analysis of telemonitoring costs: a case study in Slovenian primary care. Zdr Varst 2024 Mar;63(1):5-13. [doi: 10.2478/sjph-2024-0002] [Medline: 38156340]
9. Margolis KL, Asche SE, Dehmer SP, et al. Long-term outcomes of the effects of home blood pressure telemonitoring and pharmacist management on blood pressure among adults with uncontrolled hypertension: follow-up of a cluster randomized clinical trial. JAMA Netw Open 2018 Sep 7;1(5):e181617. [doi: 10.1001/jamanetworkopen.2018.1617] [Medline: 30646139]
10. McManus RJ, Mant J, Franssen M, et al. Efficacy of self-monitored blood pressure, with or without telemonitoring, for titration of antihypertensive medication (TASMINH4): an unmasked randomised controlled trial. Lancet 2018 Mar 10;391(10124):949-959. [doi: 10.1016/S0140-6736(18)30309-X] [Medline: 29499873]
11. Wakefield BJ, Koopman RJ, Keplinger LE, et al. Effect of home telemonitoring on glycemic and blood pressure control in primary care clinic patients with diabetes. Telemed J E Health 2014 Mar;20(3):199-205. [doi: 10.1089/tmj.2013.0151] [Medline: 24404819]
12. Wakefield BJ, Holman JE, Ray A, et al. Effectiveness of home telehealth in comorbid diabetes and hypertension: a randomized, controlled trial. Telemed J E Health 2011 May;17(4):254-261. [doi: 10.1089/tmj.2010.0176] [Medline: 21476945]
13. Bohingamu Mudiyanselage S, Stevens J, Watts JJ, et al. Personalised telehealth intervention for chronic disease management: a pilot randomised controlled trial. J Telemed Telecare 2019 Jul;25(6):343-352. [doi: 10.1177/1357633X18775850] [Medline: 29793387]
14. Steventon A, Bardsley M, Doll H, Tuckey E, Newman SP. Effect of telehealth on glycaemic control: analysis of patients with type 2 diabetes in the Whole Systems Demonstrator cluster randomised trial. BMC Health Serv Res 2014 Aug 6;14:334. [doi: 10.1186/1472-6963-14-334] [Medline: 25100190]
15. Tang PC, Overhage JM, Chan AS, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 1;20(3):526-534. [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
16. Lee JY, Chan CKY, Chua SS, et al. Telemonitoring and team-based management of glycemic control on people with type 2 diabetes: a cluster-randomized controlled trial. J Gen Intern Med 2020 Jan;35(1):87-94. [doi: 10.1007/s11606-019-05316-9] [Medline: 31512187]
17. Nicolucci A, Cercone S, Chiriatti A, Muscas F, Gensini G. A randomized trial on home telemonitoring for the management of metabolic and cardiovascular risk in patients with type 2 diabetes. Diabetes Technol Ther 2015 Aug;17(8):563-570. [doi: 10.1089/dia.2014.0355] [Medline: 26154338]
18. Karhula T, Vuorinen AL, Rääpysjärvi K, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. J Med Internet Res 2015 Jun 17;17(6):e153. [doi: 10.2196/jmir.4059] [Medline: 26084979]
19. Jia W, Zhang P, Zhu D, et al. Evaluation of an mHealth-enabled hierarchical diabetes management intervention in primary care in China (ROADMAP): a cluster randomized trial. PLoS Med 2021 Sep;18(9):e1003754. [doi: 10.1371/journal.pmed.1003754] [Medline: 34547030]

XSL•FO
RenderX

20. Ettehad D, Emdin CA, Kiran A, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet 2016 Mar 5;387(10022):957-967. [doi: 10.1016/S0140-6736(15)01225-8] [Medline: 26724178]
21. Hodgson S, Morgan-Harrisskitt J, Hounkpatin H, Stuart B, Dambha-Miller H. Primary care service utilisation and outcomes in type 2 diabetes: a longitudinal cohort analysis. BMJ Open 2022 Jan 31;12(1):e054654. [doi: 10.1136/bmjopen-2021-054654] [Medline: 35105641]
22. Dugard A, Giraudeau B, Dibao-Dina C. Do patients with type 2 diabetes mellitus included in randomised clinical trials differ from general-practice patients? a cross-sectional comparative study. BMJ Open 2024 Apr 11;14(4):e077582. [doi: 10.1136/bmjopen-2023-077582] [Medline: 38604640]
23. Mao A, Tam L, Xu A, et al. Barriers to telemedicine video visits for older adults in independent living facilities: mixed methods cross-sectional needs assessment. JMIR Aging 2022 Apr 19;5(2):e34326. [doi: 10.2196/34326] [Medline: 35438648]
24. Aïdoud A, Gana W, Poitau F, et al. High prevalence of geriatric conditions among older adults with cardiovascular disease. J Am Heart Assoc 2023 Jan 17;12(2):e026850. [doi: 10.1161/JAHA.122.026850] [Medline: 36628962]
25. Gross domestic product by region. Statistical Office of the Republic of Slovenia. 2023 Dec 15. URL: https://www.stat.si/StatWeb/en/News/Index/11537 [accessed 2024-04-15]
26. Klemenc-Ketis Z, Stojnić N, Zavrnik Č, et al. Implementation of integrated primary care for patients with diabetes and hypertension: a case from Slovenia. Int J Integr Care 2021;21(3):15. [doi: 10.5334/ijic.5637] [Medline: 34690619]
27. Zavrnik Č, Stojnić N, Mori Lukančič M, et al. Facilitators and barriers to scaling-up integrated care for arterial hypertension and type 2 diabetes in Slovenia: qualitative study. Zdr Varst 2024 Mar;63(1):38-45. [doi: 10.2478/sjph-2024-0006] [Medline: 38156335]
28. Mancia G, Kreutz R, Brunström M, et al. 2023 ESH Guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension. J Hypertens (Los Angel) 2023;41(12):1874-2071. [doi: 10.1097/HJH.0000000000003480]
29. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Apr 2024;105(4S):S117-S314. [doi: 10.1016/j.kint.2023.10.018]
30. Lin X, Li H. Obesity: epidemiology, pathophysiology, and therapeutics. Front Endocrinol (Lausanne) 2021;12:706978. [doi: 10.3389/fendo.2021.706978] [Medline: 34552557]
31. Mihevc M, Miroševič Š, Lukančič MM, et al. Assessing the reliability and validity of the Slovenian version of the Appraisal of Diabetes Scale (ADS-S) in type 2 diabetes patients. PLoS ONE 2024;19(3):e0300797. [doi: 10.1371/journal.pone.0300797] [Medline: 38527033]
32. Margolis KL, Bergdall AR, Crain AL, et al. Comparing pharmacist-led telehealth care and clinic-based care for uncontrolled high blood pressure: the hyperlink 3 pragmatic cluster-randomized trial. Hypertension 2022 Dec;79(12):2708-2720. [doi: 10.1161/HYPERTENSIONAHA.122.19816] [Medline: 36281763]
33. Weinstock RS, Teresi JA, Goland R, et al. Glycemic control and health disparities in older ethnically diverse underserved adults with diabetes: five-year results from the Informatics for Diabetes Education and Telemedicine (IDEATel) study. Diabetes Care 2011 Feb;34(2):274-279. [doi: 10.2337/dc10-1346] [Medline: 21270184]
34. Margolis KL, Asche SE, Bergdall AR, et al. Effect of home blood pressure telemonitoring and pharmacist management on blood pressure control: a cluster randomized clinical trial. JAMA 2013 Jul 3;310(1):46-56. [doi: 10.1001/jama.2013.6549] [Medline: 23821088]
35. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet 1990 Mar 31;335(8692):765-774. [doi: 10.1016/0140-6736(90)90878-9] [Medline: 1969518]
36. Shea S, Weinstock RS, Teresi JA, et al. A randomized trial comparing telemedicine case management with usual care in older, ethnically diverse, medically underserved patients with diabetes mellitus: 5 year results of the IDEATel study. J Am Med Inform Assoc 2009;16(4):446-456. [doi: 10.1197/jamia.M3157] [Medline: 19390093]
37. Mihevc M, Virtič Potočnik T, Zavrnik Č, Klemenc-Ketiš Z, Poplas Susič A, Petek Šter M. Managing cardiovascular risk factors with telemedicine in primary care: A systematic review and meta-analysis of patients with arterial hypertension and type 2 diabetes. Chronic Illn 2025 Mar;21(1):3-24. [doi: 10.1177/17423953241277896] [Medline: 39194352]
38. Fralick M, Colacci M, Odutayo A, Siemieniuk R, Glynn RJ. Lowering of hemoglobin A1C and risk of cardiovascular outcomes and all-cause mortality, a meta-regression analysis. J Diabetes Complications 2020 Nov;34(11):107704. [doi: 10.1016/j.jdiacomp.2020.107704] [Medline: 32888788]
39. Chen J, Yin D, Dou K. Intensified glycemic control by HbA1c for patients with coronary heart disease and Type 2 diabetes: a review of findings and conclusions. Cardiovasc Diabetol 2023 Jun 22;22(1):146. [doi: 10.1186/s12933-023-01875-8] [Medline: 37349787]
40. Ramallo-Fariña Y, García-Bello MA, García-Pérez L, et al. Effectiveness of internet-based multicomponent interventions for patients and health care professionals to improve clinical outcomes in type 2 diabetes evaluated through the INDICA study: multiarm cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 2;8(11):e18922. [doi: 10.2196/18922] [Medline: 33136059]

XSL•FO
RenderX

41. Timpel P, Oswald S, Schwarz PEH, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020 Mar 18;22(3):e16791. [doi: 10.2196/16791] [Medline: 32186516]
42. Mihevc M, Lukančič MM, Črt Z, et al. Towards integrated care for the elderly: exploring the acceptability of telemonitoring for hypertension and type 2 diabetes management. Int J Integr Care 2024;24(2):16. [doi: 10.5334/ijic.7621] [Medline: 38765055]
43. David CN, Iochpe C, Harzheim E, et al. Effect of mobile health interventions on lifestyle and anthropometric characteristics of uncontrolled hypertensive participants: secondary analyses of a randomized controlled trial. Healthcare (Basel) 2023 Apr 8;11(8):1069. [doi: 10.3390/healthcare11081069] [Medline: 37107903]
44. Palmer M, Sutherland J, Barnard S, et al. The effectiveness of smoking cessation, physical activity/diet and alcohol reduction interventions delivered by mobile phones for the prevention of non-communicable diseases: a systematic review of randomised controlled trials. PLoS ONE 2018;13(1):e0189801. [doi: 10.1371/journal.pone.0189801] [Medline: 29304148]
45. Jo SH, Kim SA, Park KH, Kim HS, Han SJ, Park WJ. Self-blood pressure monitoring is associated with improved awareness, adherence, and attainment of target blood pressure goals: prospective observational study of 7751 patients. J Clin Hypertens (Greenwich) 2019 Sep;21(9):1298-1304. [doi: 10.1111/jch.13647] [Medline: 31393062]
46. Gazit T, Gutman M, Beatty AL. Assessment of hypertension control among adults participating in a mobile technology blood pressure self-management program. JAMA Netw Open 2021 Oct 1;4(10):e2127008. [doi: 10.1001/jamanetworkopen.2021.27008] [Medline: 34652447]
47. Kerby TJ, Asche SE, Maciosek MV, O'Connor PJ, Sperl-Hillen JM, Margolis KL. Adherence to blood pressure telemonitoring in a cluster-randomized clinical trial. J Clin Hypertens (Greenwich) 2012 Oct;14(10):668-674. [doi: 10.1111/j.1751-7176.2012.00685.x] [Medline: 23031143]
48. Ewald S, vor dem Esche J, Uen S, Neikes F, Vetter H, Mengden T. Relationship between the frequency of blood pressure self-measurement and blood pressure reduction with antihypertensive therapy : results of the OLMETEL (OLMEsartan TELemonitoring blood pressure) study. Clin Drug Investig 2006;26(8):439-446. [doi: 10.2165/00044011-200626080-00002] [Medline: 17163276]
49. Scherbaum WA, Ohmann C, Abholz HH, Dragano N, Lankisch M. Effect of the frequency of self-monitoring blood glucose in patients with type 2 diabetes treated with oral antidiabetic drugs-A multi-centre, randomized controlled trial. PLoS One 2008 Aug 28;3(8):e3087. [doi: 10.1371/journal.pone.0003087] [Medline: 18769484]
50. Chatterjee S, Davies MJ, Heller S, Speight J, Snoek FJ, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Feb;6(2):130-142. [doi: 10.1016/S2213-8587(17)30239-5] [Medline: 28970034]
51. Virtič Potočnik T, Mihevc M, Zavrnik Č, et al. Evaluation of a specialist nurse-led structured self-management training for peer supporters with type 2 diabetes mellitus with or without comorbid hypertension in Slovenia. BMC Nurs 2024 Aug 15;23(1):567. [doi: 10.1186/s12912-024-02239-7] [Medline: 39148069]
52. Mannoubi C, Kairy D, Menezes KV, Desroches S, Layani G, Vachon B. The key digital tool features of complex telehealth interventions used for type 2 diabetes self-management and monitoring with health professional involvement: scoping review. JMIR Med Inform 2024 Mar 13;12:e46699. [doi: 10.2196/46699] [Medline: 38477979]
53. Monahan M, Jowett S, Nickless A, et al. Cost-effectiveness of telemonitoring and self-monitoring of blood pressure for antihypertensive titration in primary care (TASMINH4). Hypertension 2019 Jun;73(6):1231-1239. [doi: 10.1161/HYPERTENSIONAHA.118.12415] [Medline: 31067190]
54. Chettri S, Wang V, Balkin EA, Rayo MF, Lee CN. Development of a mobile app for clinical research: challenges and implications for investigators. JMIR Mhealth Uhealth 2022 Apr 1;10(4):e32244. [doi: 10.2196/32244] [Medline: 35363154]
55. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mhealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]
56. SCUBY mhealth 12-month randomised controlled trial data. Harvard Dataverse. URL: https://doi.org/10.7910/DVN/RJF2EK [accessed 2025-04-26]

Abbreviations

ACE: angiotensin converting enzyme
ADS: Appraisal of Diabetes Scale
AH: arterial hypertension
AUDIT-C: Alcohol Use Disorders Identification Test
BG: blood glucose
BP: blood pressure
CONSORT: Consolidated Standards of Reporting Trials
DBP: diastolic blood pressure
DPP-4: dipeptidyl peptidase 4

XSL•FO
RenderX

EU: European Union
GDP: gross domestic product
GDPR: General Data Protection Regulation
GFR: glomerular filtration rate
GLP-1: glucagon-like peptide-1
GP: general practitioner
KDIGO: Kidney Disease Improving Global Outcomes
LDL: low-density lipoprotein
mHealth: mobile health
Non-HDL: non-high-density lipoprotein
PHC: primary health care center
SBP: systolic blood pressure
SGLT-2: sodium-glucose transport protein 2
T2D: type 2 diabetes
WHO: World Health Organization

Edited by L Buis; submitted 21.04.24; peer-reviewed by AN Ali, G Stiglic, N Laichuthai; revised version received 10.11.24; accepted 17.03.25; published 29.05.25.

Please cite as:
Mihevc M, Mori Lukančič M, Zavrnik Č, Virtič Potočnik T, Ružić Gorenjec N, Petek Šter M, Klemenc-Ketiš Z, Poplas Susič A
Impact of 12-Month mHealth Home Telemonitoring on Clinical Outcomes in Older Individuals With Hypertension and Type 2 Diabetes: Multicenter Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e59733
URL: https://mhealth.jmir.org/2025/1/e59733
doi:10.2196/59733

© Matic Mihevc, Majda Mori Lukančič, Črt Zavrnik, Tina Virtič Potočnik, Nina Ružić Gorenjec, Marija Petek Šter, Zalika Klemenc-Ketiš, Antonija Poplas Susič. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Barriers and Facilitators in Implementing a Telemonitoring Application for Patients With Chronic Kidney Disease and Health Professionals: Ancillary Implementation Study of the NeLLY (New Health e-Link in the Lyon Region) Stepped-Wedge Randomized Controlled Trial

Marion Delvallée[1], PhD; Abdallah Guerraoui[2], MD, MSc; Lucas Tchetgnia[3,4], PhD; Jean-Pierre Grangier[2], LPN; Nassira Amamra[5], MSc; Anne-Laure Camarroque[2], MSc; Julie Haesebaert[1,5], MD, PhD; Agnès Caillette-Beaudoin[2], MD

[1]Research on Healthcare Performance RESHAPE, INSERM U1290, Université Claude Bernard Lyon 1, Lyon, France

[2]Calydial, Vienne, France

[3]Collectif en Sciences Sociales Appliquées, Paris, France

[4]Université Bourgogne Europe, Laboratoire Interdisciplinaire de Recherches - Sociétés, Sensibilités, Soin, Dijon, France

[5]Service Recherche et Epidémiologie Cliniques, Pôle de Sante Publique, Hospices Civils de Lyon, Lyon, France

Corresponding Author:
Julie Haesebaert, MD, PhD
Research on Healthcare Performance RESHAPE, INSERM U1290
Université Claude Bernard Lyon 1
8 avenue Rockefeller
Lyon, 69008
France
Phone: 33 06 51 42 97 43
Email: julie.haesebaert01@chu-lyon.fr

Abstract

Background: The use of telemonitoring to manage renal function in patients with chronic kidney disease (CKD) is recommended by health authorities. However, despite these recommendations, the adoption of telemonitoring by both health care professionals and patients faces numerous challenges.

Objective: This study aims to identify barriers and facilitators in the implementation of a telemonitoring program for patients with CKD, as perceived by health care professionals and patients, and to explore factors associated with the adoption of the program. This study serves as a process evaluation conducted alongside the cost-effectiveness NeLLY (New Health e-Link in the Lyon Region) trial.

Methods: A mixed methods approach combining a quantitative questionnaire and semistructured interviews was conducted among nurses, nephrologists, and patients with stages 3 and 4 CKD across 10 renal care centers in France that have implemented telemonitoring. The Technology Acceptance Model (TAM) and the Consolidated Framework for Implementation Research (CFIR) were used to design the questionnaires and interview guides. The dimensions investigated included ease of use, perceived usefulness, and intention to use (TAM), as well as characteristics of the intervention, local and general context, individual factors, and processes (CFIR). The adoption of telemonitoring was assessed based on the frequency with which patients connected to the telemonitoring device. Determinants of telemonitoring use were analyzed using nonparametric tests, specifically the Wilcoxon-Mann-Whitney and Kruskal-Wallis tests. Thematic analysis was conducted on the transcriptions of semistructured interviews. Both quantitative and qualitative results, including data from patients and professionals, were integrated to provide a comprehensive understanding of the factors associated with the use of remote monitoring in CKD.

Results: A total of 42 professionals and 128 patients with CKD responded to our questionnaire. Among these, 11 professionals and 13 patients participated in interviews. Nurses, who were responsible for patient follow-up, regularly used telemonitoring (8/13, 62%, at least once a month), while nephrologists, who were responsible for prescribing it, were primarily occasional users (5/8, 63%, using it less than once a month). Among professionals, the main obstacles identified were the heavy workload generated

XSL•FO
RenderX

by telemonitoring, lack of training, and insufficient support for nurses. Among the 128 patients, 46 (35.9%) reported using the application at least once a week. The main barriers for patients were issues related to computer use, as well as the lack of feedback and communication with health care professionals. The main facilitators identified by both professionals and patients for using telemonitoring were the empowerment of patients in managing their health and the reduction of the burden associated with CKD.

Conclusions: Improving adherence to telemonitoring in the context of CKD requires collaborative efforts from both professionals and patients. Our results provide insights that can inform the design of effective, theory-driven interventions aimed at improving telemonitoring adoption and usage.

(JMIR Mhealth Uhealth 2025;13:e50014) doi:10.2196/50014

KEYWORDS

telehealth; telemonitoring; chronic kidney disease; implementation; Consolidated Framework for Implementation Research; Technology Acceptance Model

Introduction

Chronic kidney disease (CKD) is a generally asymptomatic condition characterized by a progressive decline in kidney function. The prevalence of CKD is approximately 11% in high-income countries [1,2]. The primary challenge for health professionals is to preserve patients' renal function for as long as possible to delay progression to stage 5, where kidney function must be compensated by dialysis or transplantation. To achieve this, it is necessary to closely monitor patients' health status—such as blood pressure, creatinine levels, or weight [3]—from the onset of the disease [4]. This adds to the burden of the disease for patients and may lead to challenges with adherence and engagement in their care. In the context of CKD, telemonitoring, which involves remotely tracking a patient's health status through regular collection of health data, has been internationally recommended [5] and in France "to slow the progression of the disease by setting up early detection of kidney disease in patients at risk and appropriate therapeutic management" [6]. Indeed, in patients with CKD, telemonitoring has been shown to improve data sharing between patients and health professionals, enhance patient autonomy (eg, blood pressure control) [7], facilitate care coordination [5], and support the monitoring of patients in remote areas [8]. Additionally, some studies have reported better compliance in these patients [9-12].

Although telemonitoring is recommended for CKD in many countries to adapt patient care [13], several studies have raised concerns about its feasibility [14,15]. Bonner et al [16] highlighted that factors such as health literacy, sociodemographic characteristics, and ease of use of technology could influence telehealth adoption in patients with CKD. In a recent review, Jacob et al [17] classified the factors influencing telehealth adoption by professionals into 3 categories. These categories include (1) technical and material factors (eg, ease of use, user experience), (2) social and personal factors (eg, personal characteristics), and (3) organizational and policy factors (eg, workflow-related and patient-related considerations). Another review by O'Connor et al [18] explained patient engagement factors in telehealth. These factors were divided into 4 categories: (1) personal agency and motivation, (2) personal life and values, (3) engagement and recruitment approaches, and (4) quality of the digital health intervention. Several theoretical frameworks have been proposed to organize and group these factors into domains to guide telehealth implementation [19].

Studies have shown that telehealth implementation impacts patients and professionals both positively and negatively [17,19-23]. However, these impacts may vary depending on cultural and organizational contexts. Currently, no data are available in France regarding the acceptability and adoption of telemonitoring for the follow-up of patients with CKD. A French national multicenter stepped-wedge randomized controlled trial (SW-RCT) (NeLLY [New Health e-Link in the Lyon Region] Trial NCT03348839) is ongoing to evaluate the cost-effectiveness of a program that combines multiprofessional telemonitoring with support for a personalized care pathway for patients with severe predialysis CKD. In the NeLLY program, telemonitoring is delivered through the apTeleCare application (Multimedia Appendix 1). This application enables the planning and monitoring of scheduled or unscheduled activities, as well as the management of alerts triggered when patients enter concerning health data. Patients and health professionals can access the NeLLY program when their center is in the intervention phase of the SW-RCT.

The aim of our study was to identify the barriers and facilitators to implementing the NeLLY telemonitoring program by gathering insights from health professionals and patients with CKD participating in the NeLLY trial.

Methods

Study Design

We conducted a mixed methods implementation survey, combining quantitative questionnaires and semistructured interviews, following an explanatory sequential design [24]. This survey was conducted alongside the NeLLY national multicenter SW-RCT. The aim of our implementation survey was to assess the use of remote monitoring by patients and professionals in the units participating in the trial. Our survey was conducted in the 10 centers (out of a total of 15 participating centers) that were in the intervention phase at the time of the survey.

Ethics Approval

The Institutional Review Board CPP OUEST II-Angers approved this study, project no. 2017/37 (2017-A00091-52). The reporting of the results is guided by the Checklist for

XSL•FO
RenderX

Reporting Results of Internet E-Surveys (CHERRIES) guidelines [25]. All data were anonymized with a unique anonymity number. No compensation was provided to participants for the study.

Theoretical Framework

We based the design and conduct of our study on 2 complementary theoretical frameworks: the Technology Acceptance Model (TAM), specific to telemedicine adoption, supplemented by additional domains from the general implementation framework—the Consolidated Framework for Implementation Research (CFIR)—to achieve a comprehensive understanding of the factors shaping implementation.

The TAM focuses on understanding why users accept or reject technology and how acceptance can be improved [26]. It is based on 2 key principles: perceived usefulness and perceived ease of use. These concepts shape users' attitudes toward technology, which, in turn, influence behavioral intention. This framework can help identify factors for optimizing the implementation of telehealth.

The CFIR examines 5 areas: intervention characteristics, outer settings, inner settings, characteristics of individuals, and the implementation process [27]. It helps in understanding implementation and can be used to guide formative assessments. This framework has been widely applied in studies to explore factors influencing the implementation of telehealth [28].

TAM and CFIR share common domains, but TAM includes domains that are more specific to telemonitoring. CFIR, by contrast, incorporates additional domains related to intervention characteristics, outer settings, inner settings, individuals, and processes, which we integrated to develop our framework.

Participants

Our study targeted 2 populations: patients and health professionals participating in the NeLLY trial at centers in the intervention phase during our study.

The patients included in the NeLLY trial were those with CKD at stage 4 or a glomerular filtration rate between 30 and 37 ml/minute, as determined based on the nephrologist's assessment. Eligible patients also had at least one comorbidity (cardiovascular disease, diabetes, or both) and access to an internet connection at home. All patients enrolled in the intervention phase of PRME (Medical and Economic Research Program) NeLLY at the start of our survey (March 2021) were invited to participate and complete the questionnaire. A total of 305 patients were invited to the study, and all voluntary respondents were included in the survey.

In each center, health professionals—including nephrologists and nurses—who were involved in recruiting patients for the NeLLY trial or monitoring patients at the time of the study were eligible to participate. A total of 103 eligible professionals were contacted, and all voluntary respondents were included in the survey.

Development of Questionnaires and Interview Guides

Questionnaires and interview guides were developed based on the 3 categories of factors described by Jacob et al [17] and informed by the CFIR and TAM frameworks [26-29]. The first question of the survey addressed participants' nonopposition to study participation. The patient questionnaire comprised 29 items divided into 3 groups: sociodemographic factors (4 items), technological and material factors (18 items), and social and personal factors (5 items). The professionals' questionnaire consisted of 45 items divided into 4 sections: sociodemographic factors (8 items), technological and material factors (14 items), social and personal factors (6 items), and organizational and policy factors (15 items). A table detailing the questions by domains of the theoretical framework is available in Multimedia Appendix 2.

The questions were of 2 types: single or multiple-choice questions, and questions assessing the respondent's agreement with proposals written in the affirmative form, rated on a scale from 0 (not agree at all) to 10 (totally agree). Questions with a score were grouped by domains of the theoretical frameworks (Multimedia Appendix 2).

The questionnaires were tested by 3 patients who were not eligible for the study but had benefited from remote monitoring, and by 3 professionals, focusing on form and content. The estimated time to complete the questionnaire was 15 minutes for both patients and professionals. At the end of the questionnaire, participants were invited to take part in semistructured interviews that explored their use of apTeleCare in detail, based on the 3 categories of factors developed by Jacob et al [17]. We conducted 13 interviews with patients and 11 interviews with health professionals.

Outcome: Telemonitoring Uptake/Adoption

Among professionals, telemonitoring adoption was defined by the responses to the question, "How often do you consult the telemonitoring application (apTeleCare)?" with 4 response options: "once or several times a day," "once or several times a week," "once or several times a month," and "less often." The initial recommendation in the NeLLY trial was to use telemonitoring several times a week. Following these recommendations, professionals who reported using the application once a day or once a week were categorized as "frequent users," while those who used it at least once a month or less often were classified as "average users." Finally, those who reported not using the telemonitoring application were categorized as "nonusers." For patients, the type of user (frequent, average, or on-off) was determined after analyzing their application usage data, which was directly extracted from apTeleCare. The initial recommendation in the NeLLY trial was to use telemonitoring at least once a week. Based on this, we considered "frequent users" to be patients with a frequency of use (the number of log-ins divided by application usage time, with usage time defined as the data extraction date minus the training date) greater than or equal to 4 connections per month. Patients with a frequency of use strictly less than 1 connection per month were categorized as "on-off" users. The remaining patients were classified as "average" users.

Data Collection

An information letter about the study was sent along with the questionnaire to both patients and professionals. Only

XSL•FO
RenderX

participants who indicated that they did not object to the research (question 1) were considered.

Patient questionnaire data were initially collected via apTeleCare from March 22 to April 6, 2021. Follow-up emails were sent to patients who did not respond to the application between April 9 and April 26, 2021. Patients had until May 14, 2021, to return the questionnaire. To maintain patient anonymity, questionnaires were first sent to the investigative centers. The clinical research officers at these centers were responsible for entering the patient's anonymization number on the questionnaire before sending it. Data collected on apTeleCare were securely transmitted to the data controller via TMM Software, the developer of apTeleCare, using a secure file transfer.

Professional questionnaire data were collected using the REDCap database (Research Electronic Data Capture, version 11.0.1; Vanderbilt University) from March 22 to April 17, 2021 [30]. Recruitment was conducted via email, with a link providing access to an online questionnaire hosted by the REDCap database.

Patient and health professional interviews were conducted and transcribed by the first author (MD) between June and August 2021.

The interviews were conducted via videoconference or phone call, based on the participant's preference, and were audio recorded for accuracy. Nonobjection was obtained orally at the start of each interview.

Data Analysis

Questionnaire analysis was conducted using R software (R Foundation) [31]. For both populations (patients and professionals), we followed the same analysis plan. First, we described respondent characteristics and the scores obtained for each item. We calculated the number and frequency for categorical variables, and for quantitative variables, we calculated the mean and SD or median, along with the first and third quartiles, depending on the distribution. We also calculated average scores for each domain of our theoretical framework (CFIR and TAM) by summing the scores for the items and dividing by the total number of items in the domain. Once this calculation was performed for all respondents, we calculated the median and first/third quartiles for each domain. Subsequently, we analyzed the consistency of the items by calculating Cronbach α, both globally and by grouping items according to the domains of the theoretical frameworks. We then studied factors associated with telemonitoring adoption using bivariate comparisons with the Wilcoxon Mann-Whitney test among health professionals, grouping them into 2 categories: frequent users and average/on-off users. For patients, we used the Kruskal-Wallis test, grouping them into 3 categories (frequent users, average users, and on-off users). In the professional sample, we analyzed responses according to occupation: nurses or nephrologists, as they were differently involved in telemonitoring. Nurses were responsible for responding to patient alerts, while nephrologists were in charge of including patients and proposing telemonitoring to them.

Interview analysis was conducted by the first and third authors (MD and LT) following a theoretical thematic analysis method [32]. The analysis involved identifying the themes and subthemes raised in the interviews, and then reclassifying them into the different domains of our theoretical frameworks (CFIR and TAM). The analysis was initially conducted separately for interviews with professionals and patients, after which the results were cross-checked between the 2 groups to compare opinions. Verbatims were coded as Nu for nurses, MD for nephrologists, and Pa for patients, with the interview number added (eg, Nu2 corresponds to the second nurse interviewed). These analyses were conducted using NVivo software (Lumivero) [33]. The qualitative analysis report was written by the third author (LT). Key findings from different sources (quantitative questionnaires and semistructured interviews) and different populations (patients and health professionals) were triangulated to understand behavioral determinants and identify barriers and facilitators regarding telemonitoring use in CKD [34].

Results

Insights From Health Professionals

Sample Characteristics

A total of 43 professionals out of the 103 contacted responded to the questionnaire, representing a response rate of 41.7%. One respondent was excluded because they did not specify whether they were a nurse or a nephrologist. The respondents (nurses and nephrologists) practiced in all 10 centers targeted by the study. Among the 42 respondents, 13 (31%) were nurses and 29 (69%) were nephrologists (Table 1). As shown in Table 1, 8 out of 13 (62%) nurses were classified as "frequent users," compared with only 1 out of 29 (3%) nephrologists.

XSL•FO
RenderX

Table 1. Demographics characteristics of professionals (N=42).

Variables	Total population (N=42)	Nurses (n=13)	Nephrologists (n=29)
Number of respondents	42 (100)	13 (31)	29 (69)
Gender, n (%)			
Male	15 (36)	1 (8)	14 (48)
Female	27 (64)	12 (92)	15 (52)
Age (years), n (%)			
[30; 40[	16 (38)	4 (31)	12 (41)
[40; 50[	15 (36)	5 (38)	10 (34)
>50	11 (26)	4 (31)	7 (24)
Exercise time in the center (months), n (%)			
<1	0 (0)	0 (0)	0 (0)
[2; 6[	1 (2)	1 (8)	0 (0)
[6; 12[	0 (0)	0 (0)	0 (0)
>12	41 (98)	12 (92)	29 (100)
Number of patients included in the NeLLY[a] trial, n (%)			
0	6 (14)	2 (15)	4 (14)
[1; 5[	15 (36)	3 (23)	12 (41)
[5; 20[	11 (26)	2 (15)	9 (31)
>20	10 (24)	6 (46)	4 (14)
Time of use of the NeLLY service (months), n (%)			
<1	1 (2)	1 (8)	0 (0)
[1; 3[	2 (5)	1 (8)	1 (3)
>3	18 (43)	11 (85)	7 (24)
Nonuser	21 (50)	0 (0)	21 (72)
Total of user	21 (50)	13 (100)	8 (28)
Use of remote monitoring other than NeLLY, n (%)			
Yes, the service uses it	11 (26)	1 (8)	10 (34)
Yes, I use it	9 (21)	6 (46)	3 (10)
No	22 (52)	6 (46)	16 (55)
Frequency of consultation of the application (if used), n/N (%)			
At least once a day	4/21 (19)	4/13 (31)	0/8 (0)
At least once a week	5/21 (24)	4/13 (31)	1/8 (13)
At least once a month	2/21 (10)	0/13 (0)	2/8 (25)
Less often	10/21 (48)	5/13 (38)	5/8 (63)
Type of user, n (%)			
Frequent user	9 (21)	8 (62)	1 (3)
Average user	12 (29)	5 (38)	7 (24)
One-off user	21 (50)	0 (0)	21 (72)

[a]NeLLY: New Health e-Link in the Lyon Region.

Characteristics by user profile are described in Table 2. The mean age was approximately 44 years in both user groups. Some differences between the 2 user profiles regarding gender and the number of patients included in the NeLLY trial could be

XSL•FO
RenderX

linked to the distribution between nurses and nephrologists in the 2 groups. It is noteworthy that 8 of 9 (89%) good users were nurses. Of the 43 respondents, 11 agreed to participate in a semistructured interview, including 4 nephrologists and 7 nurses.

Table 2. Description of sociodemographic characteristics according to the user profile in health professionals.

Variables	Frequent users (n=9)	Average/one-off users (n=33)
Gender, n (%)		
Male	1 (11)	14 (42)
Female	8 (89)	19 (58)
Age (years), mean (SD)	44 (15.95)	44.2 (8.87)
Type of professional, n (%)		
Nephrologists	1 (11)	28 (85)
Nurses	8 (89)	5 (15)
Number of patients included in the NeLLY[a] trial, n (%)		
0	0 (0)	6 (18)
[1; 5[	1 (11)	14 (42)
[5; 20[	3 (33)	8 (24)
>20	5 (56)	5 (15)
Time of use of the NeLLY service (in months) , n (%)		
<1	0 (0)	1 (3)
[1; 3[	1 (11)	1 (3)
>3	8 (89)	10 (30)
Nonuser	0 (0)	21 (64)

[a]NeLLY: New Health e-Link in the Lyon Region.

Consistency of the Questionnaire

The overall Cronbach α coefficient for the professionals' questionnaire was 0.97, indicating good consistency among all items. All questions related to the TAM showed strong consistency, with coefficients greater than 0.90. Regarding CFIR, 3 domains exhibited weaker consistency: intervention characteristics (Cronbach α=0.55), individual (Cronbach α=0.62), and process (Cronbach α=0.67).

Results According to Framework Dimensions

The distribution of the median scores by subdomain of the theoretical framework, profession, and user profile can be found in Tables 3 and 4. The highest scores were observed among respondents in the "Data Security" trust and "Self-Efficacy" perception subdomains of the CFIR (individual field), with respective medians of 8 (IQR 8-10) and 7 (IQR 6-8; Table 3). Indeed, during the interviews, it emerged that health professionals reported using technologies on a daily basis. Nu5 noted the following about using technologies: "In general I manage quite well [...]in fact we use (Information technologies) all the time for patient medical records, telemonitoring, e-mail communication". The least-rated subdomains were Perceived Adaptability, Perceived Response to Patients' Needs, Peer Support, and Planning. Adaptability, referring to the integration of telemonitoring into daily workflow, was perceived by the majority of nurses as an additional workload (Nu2: "It takes work you, don't realize how long it takes to manage an alert"). However, in a center where they dedicated a nurse to the telemonitoring activity, the experience was different (Nu7: "Frankly, no, I don't think so (increase in workload). You have to contact those who don't log on, but I think that's part of the role of the professional behind the application. I don't see that as a constraint, but as my role"). Peer support for using the application was perceived as insufficient, as mentioned by Nu4 "I asked for support from my colleagues who had also been trained in NeLLY, so that I could take over, except that it fell apart over time, [...] At first my colleagues took over (checking alerts when Nu4 was off) and lately there's no longer any relay." Regarding the Planning subdomain (organizational processes and protocols to integrate the application into practice), interviewees expressed regret over the lack of a structured protocol to support them in providing information and training to patients on how to use the telemonitoring system.

XSL•FO
RenderX

Table 3. Description of scores according to the type of professional (nurse or nephrologist) and the field of theoretical frameworks (N=42).

Domain and subdomain (CFIR[a])	Total population (N=42), median (Q1-Q3)	Nurses (n=13), median (Q1-Q3)	Nephrologists (n=29), median (Q1-Q3)	*P* value[b]
Intervention characteristics				
Adaptability	5 (4-7)	5 (4-6)	6 (3-7)	.94
Outer settings				
Patients' Needs and Resources	5 (4-6)	5 (4-5)	5 (4-6)	.35
Peer Pressure	5 (3-6)	4 (3-6)	5 (3-6)	.54
Inner settings				
Readiness for Implementation	5 (4-6)	5 (3-7)	5 (4-6)	.78
Individual				
Data Security	8 (8-10)	8 (8-10)	9 (8-10)	.97
Self-Efficacy	7 (6-8)	7 (7-8)	7 (6-8)	.49
Process				
Planning	5 (3-9)	4 (3-5)	8 (6-10)	<.01
TAM[c]				
Perceived Ease of Use	6 (5-8)	6 (5-7)	6 (5-8)	.93
Perceived Usefulness	7 (5-7)	7 (6-8)	7 (5-7)	.35
Behavioral Intention to Use	6 (3-8)	6 (4-8)	7 (2-8)	.60

[a]CFIR: Consolidated Framework for Implementation Research.

[b]*P* value for the comparison between the 2 users' profiles (Wilcoxon test).

[c]TAM: Technology Acceptance Model.

Table 4. Descriptions of scores based on user profile and areas of theoretical frameworks in health professionals (N=42).

Field and subdomain (CFIR[a])	Frequent user (n=9), median (Q1-Q3)	Average/on-off user (n=33), median (Q1-Q3)	*P* value[b]
Characteristics of the intervention			
Adaptability	6 (5-8)	4 (3-6)	.02
Outer settings			
Patients' Needs and Resources	5 (5-6)	5 (3-6)	.92
Peer Pressure	6 (4-7)	5 (3-6)	.18
Inner settings			
Readiness for Implementation	7 (5-8)	5 (4-6)	.07
Individual			
Data Security	9 (8-10)	8 (7-10)	.45
Self-Efficacy	8 (7-9)	7 (5-8)	.05
Process			
Planning	5 (4-9)	5 (3-8)	.89
Field (TAM[c])			
Perceived Ease of Use	7 (6-8)	6 (5-8)	.38
Perceived Usefulness	8 (7-8)	6 (5-7)	.008
Behavioral Intention to Use	8 (7-9)	5 (2-7)	.02

[a]CFIR: Consolidated Framework for Implementation Research.

[b]*P* value for the comparison between the 2 users' profile (Wilcoxon test).

[c]TAM: Technology Acceptance Model.

XSL•FO
RenderX

Results According to Profession and User Profile

Looking at the scores by profession (Table 3), nurses generally had equal or lower ratings than nephrologists. Two domains were rated below 5/10 by nurses: Peer Support and Planning. The Planning subdomain showed a significantly lower score for nurses compared with nephrologists, with a median of 4 (IQR 3-5) versus 8 (IQR 6-10), respectively (P<.01). Concerns in the Planning subdomain were explained in interviews regarding the collaboration between nurses and nephrologists in managing telemonitoring alerts. Nu1 mentioned, “I monitor, I alert but if I don’t get feedback from the medical team I can’t do much more.” Peer support was less well perceived among nurses (Table 3) who considered that telemonitoring was not a priority for the medical team, Nu2 “We sent an alert to a doctor, by e-mail for example, and got no response. [...] It (telemonitoring) is not yet recognized as a priority.”, and that only few professionals really got involved, Nu3 “In a 10-person medical team, only one person is really involved.”

Table 4 shows the median scores among health care professionals by domain according to the user profile. The details of the scores by question are available in Multimedia Appendix 3. Significant differences according to user profiles were observed in the Adaptability subdomain of CFIR (median 6, IQR 5-8 vs median 4, IQR 3-6, P=.02), Perceived Usefulness (median 8, IQR 7-8 vs median 6, IQR 5-8, P=.008), and Behavioral Intention to Use (median 8, IQR 7-9 vs median 5, 2-7, P=.02) of the TAM, with median scores 2-3 points lower for average users compared with frequent users. Questions in the Adaptability subdomain of CFIR focused on workload and changing practices. In the interviews, professionals linked the workload to the management of apTeleCare. Nu4 mentioned, when discussing technical issues, “It worked well at first, but then the problems started with a large number of patient calls.” The perceived usefulness of telemonitoring for patients was mentioned during the interviews. Nu7 said, “It’s really something that can be very beneficial for the patient, allowing us to have contact with them, a very close follow-up [...] in the kidney disease, these are the patients with whom we need to stay in regular contact to be able to avoid any problems.”

To a lesser extent, average users tended to have lower scores for the Readiness for Implementation (median 7, IQR 5-8 vs median 5, IQR 4-6, P=.07) and Self-Efficacy (median 8, IQR 7-9 vs median 7, IQR 5-8, P=.05) domains of the CFIR. The readiness for implementation questions focused on the management of the application from a technical point of view.

Insights From Patients

Sample Characteristics

A total of 128 patients out of 305 contacted responded to the questionnaire, representing a response rate of 42.0%. No statistically significant differences were found between the demographic characteristics of patients who responded to the survey and nonrespondents from the source population (Multimedia Appendix 4). Respondents were from all 10 centers in the intervention period, which had already included patients. Of these respondents, 75 (58.6%) completed the survey via apTeleCare and 53 (41.4%) via postal mail. The mean age of the population was 70.8 (SD 10.2) years. As shown in Table 5, 46 (35.9%) patients were classified as “frequent users” (≥4 connections/month), 39 (30.5%) as “average users,” and 15 (11.7%) as “on-off users” (<1 connection/month). The mean age of frequent users was approximately 74.3 years, which is similar to that of the other 2 groups. Some differences were observed regarding gender, center, and the presence of a caregiver across the profiles.

XSL•FO
RenderX

Table 5. Demographic characteristics of patients in total population and by type of user (N=128).

Variables	Total population (N=128)	Frequent user (n=46)	Average user (n=39)	One-off user (n=15)
Questionnaire format, n (%)				
apTeleCare	75 (58.6)	41 (89.1)	26 (66.7)	1 (6.7)
Mail	53 (41.4)	5 (10.9)	13 (33.3)	14 (93.3)
Gender, n (%)				
Male	85 (66.4)	33 (71.7)	22 (56.4)	10 (66.7)
Female	43 (33.6)	13 (28.3)	17 (43.6)	5 (33.3)
Age (years)				
Mean (SD)	71.9 (10.2)	74.3 (9.87)	68.85 (10.9)	71.5 (11.1)
Range	39-94	47-94	39-86	50-87
Monitoring center, n (%)				
Centre Hospitalier Universitaire (University hospitals centers)	61 (47.7)	21 (45.7)	16 (41.0)	9 (60.0)
Nonuniversity hospitals/association center	67 (52.3)	25 (54.3)	23 (59.0)	6 (40.0)
NeLLY[a] service proposal date (months), n (%)				
<1	4 (3.1)	1 (2.2)	0 (0)	1 (6.7)
[1; 3[	7 (5.5)	4 (8.7)	0 (0)	0 (0)
[3; 6[	20 (15.6)	5 (10.9)	6 (15.4)	4 (26.7)
≥6	86 (67.2)	36 (78.3)	33 (84.6)	9 (60.0)
Frequency of use of a computer/tablet/smartphone, n (%)				
Daily	101 (78.9)	38 (82.6)	30 (76.9)	13 (86.7)
Weekly	12 (9.4)	3 (6.5)	5 (12.8)	1 (6.7)
Less often	10 (7.8)	4 (8.7)	2 (5.1)	0 (0)
Never	5 (3.9)	1 (2.2)	2 (5.1)	1 (6.7)
Frequency of e-mailbox consultation, n (%)				
Daily	98 (76.6)	35 (76.1)	29 (74.4)	11 (73.3)
Weekly	17 (13.3)	4 (8.7)	6 (15.4)	4 (26.7)
Less often	9 (7.0)	6 (13.0)	2 (5.1)	0 (0.0)
Never	4 (3.1)	1 (2.2)	2 (5.1)	0 (0.0)
Internet connection speed, n (%)				
Very good	50 (39.1)	20 (43.5)	13 (33.3)	7 (46.7)
Pretty good	41 (32.0)	16 (34.8)	13 (33.3)	3 (20.0)
Medium	25 (19.5)	8 (17.4)	9 (23.1)	3 (20.0)
Bad	10 (7.8)	2 (4.3)	3 (7.7)	2 (13.3)
Help of caregiver, n (%)				
Yes	32 (25.0)	11 (23.9)	15 (38.5)	3 (20.0)
No	67 (52.3)	32 (69.6)	21 (53.8)	6 (40.0)
If yes, what kind of caregivers?[b], n (%)				
Family members or relatives	16 (50.0)	7 (63.6)	8 (53.3)	1 (33.3)
Caregiver of nephrology service	16 (50.0)	4 (36.4)	7 (46.7)	2 (66.7)
Remote monitoring improves, n (%)				
Relationships with health professionals[c]	96 (75.0)	41 (89.1)	30 (76.9)	8 (53.3)
Quality of medical monitoring	58 (45.3)	21 (45.7)	25 (64.1)	6 (40.0)

XSL•FO
RenderX

Variables	Total population (N=128)	Frequent user (n=46)	Average user (n=39)	One-off user (n=15)
The course of kidney disease	57 (44.5)	28 (60.9)	15 (38.5)	4 (26.7)
The attention I pay to my state of health	53 (41.4)	24 (52.2)	16 (41.0)	4 (26.7)
Quality of care	34 (26.6)	15 (32.6)	10 (25.6)	4 (26.7)
Considering the experience of my illness	31 (24.2)	7 (15.2)	12 (30.8)	3 (20.0)
Taking my medication	17 (13.3)	6 (13.0)	6 (15.4)	1 (6.7)
Reasons which help to continue the use of remote monitoring, n (%)				
I see an interest in my health	61 (47.7)	1 (2.2)	1 (2.6)	4 (26.7)
A health professional at my center recommended it to me	52 (40.6)	22 (47.8)	24 (61.5)	7 (46.7)
It saves time	5 (3.9)	4 (8.7)	13 (33.3)	6 (40.0)
My relatives recommended it to me	8 (6.3)	4 (8.7)	1 (2.6)	0 (0)
I do not use remote monitoring	4 (3.1)	5 (10.9)	3 (7.7)	0 (0)

[a]NeLLY: New Health e-Link in the Lyon Region.

[b]Proportion calculated on patients who said yes in different groups.

[c]Proportion calculated on patients who said yes to either improve communication with a nephrologist or a nurse or relationship with health professionals in general.

Regarding patients' perceptions of telemonitoring, 96 (75%) patients felt that it improved their relationship with health professionals. Additionally, across the total population and within each user group, patients agreed that telemonitoring enhanced their attention to their disease. Furthermore, approximately half of the patients believed that telemonitoring improved the quality of their medical monitoring (58/128, 45.3%) and the course of their kidney disease (57/128, 44.5%). The main reasons for nonuse, as reported by on-off users, were a lack of interest in telemonitoring (8/25, 32%), difficulties in using the telemonitoring application (7/25, 28%), and internet connection issues (6/25, 24%). Otherwise, in the total patient population, 32 (25%) reported using telemonitoring with the help of a caregiver. No statistically significant differences were identified between users regarding the variables associated with the use of remote monitoring (Multimedia Appendix 5). Further general demographics can be found in Table 5.

Of the 128 respondents, 13 accepted to participate in a semistructured interview (8 men). The mean age of the participants was 67.5 years. Regarding the user profiles, 5 were frequent users, 2 were average users, 1 was an on-off user, and 4 were not categorized because their usage data were unavailable.

Consistency of the Questionnaire

The overall Cronbach α coefficient for the patient questionnaire was 0.91, indicating good consistency between all items. Questions relating to TAM showed good consistency with coefficients around 0.90. Similar to the professionals' questionnaire, the domains with less strong consistency were from the CFIR framework: Individual (Cronbach α=0.25) and Outer Settings (Cronbach α=0.59).

Results According to Framework Dimensions

The median scores by domains and subdomains of our 2 theoretical frameworks in the total population were generally high (median ≥7). Higher scores were observed for the Data Security (knowledge and beliefs; median 10) and Peer Support (median 8) domains (Table 6). During the interviews, every patient mentioned receiving support in the use of telemonitoring from their relatives, although not necessarily related to their disease. Indeed, some patients do not discuss their disease out of fear of bothering their loved ones. As Pa2 mentioned, "No, if I have an internet problem I call one of my daughters, but no, no I never talk about my health. You can't leave with your illness all the time." Similarly, Pa11 said, "I try not to spread my illness and then my entourage is bored enough." Regarding the knowledge and belief aspect, patients were generally unconcerned about the use of their data, as they understood it was being used by their health care professionals. As Pa2 mentioned, "because it helps Professor Y (nephrologist) every time he sees me."

XSL•FO
RenderX

Table 6. Descriptions of scores based on user profile and areas of theoretical frameworks in patients (N=100).

Domain and subdomain (CFIR[a])	Total population (N=128), median (Q1-Q3)	Frequent user (n=46), median (Q1-Q3)	Average user (n=39), median (Q1-Q3)	One-off user (n=15), median (Q1-Q3)	*P* value[b]
Outer Settings					
Peer Pressure	8 (8-9)	7 (5-9)	8 (4-10)	8 (8-9)	.55
Inner Settings					
Readiness for Implementation	8 (7-9)	8 (5-8)	8 (5-10)	8 (6-9)	.97
Individual					
Data Security	10 (9-10)	9 (5-10)	10 (9-10)	8 (0-10)	.09
Self-Efficacy	8 (6-8)	7 (5-10)	8 (5-9)	8 (1-8)	.51
TAM[c]					
Perceived Ease of Use	8 (5-9)	9 (8-10)	9 (7-10)	5 (4-9)	.44
Perceived Usefulness	7 (5-9)	7 (6-9)	8 (5-9)	6 (5-7)	.87
Behavioral Intention to Use	8 (3-9)	7 (5-10)	7 (5-10)	7 (0-9)	.38

[a]CFIR: Consolidated Framework of Implementation Research.

[b]*P* value for the comparison between the three users' profile (Kruskal-Wallis test).

[c]TAM: Technology Acceptance Model.

Results According to the User Profile

Looking at the scores by user profile (Table 6), median scores across all user groups were close and high. Although the differences were not statistically significant, some variations were observed between the user profiles (Table 6). On-off users, in particular, had wider score distributions compared with the other groups, with lower first quartile values in the Data Security and Self-Efficacy subdomains of the CFIR, as well as in the Perceived Ease of Use and Behavioral Intention to Use domains of the TAM. Questions regarding self-efficacy focused on the ease of use of the internet. In the interviews, 6 out of 13 patients mentioned feeling uncomfortable with computers. Pa13 shared his experience with using a new computer: "It may have been simpler for the others, but for me it was complicated, I had trouble, I still have trouble using it so in the end I'm going to use it less and less." The perceived ease of use of telemonitoring was a common theme in the interviews; 9 out of 13 patients reported having difficulty using apTeleCare, and 10 out of 13 encountered technical problems. However, only 3 out of 13 had their technical issues resolved. Technical issues mentioned during the interviews included bugs and problems with log-ins and passwords. Pa4 explained, "at one point when I tried to set the measurements such as blood pressure, each time it did not work so I had to disconnect, then reconnect again." Pa11 shared, "there was an update that caused a system bug and as I had contacted the local correspondent [...], she sent them to the hotline who had to call me back to solve the problem and then I never got a reply."

Insights Crossed Between Professionals and Patients

Several topics were discussed by both health care professionals and patients during the interviews. Both groups highlighted the perceived psychological benefits of telemonitoring for patients and the overall improvement in patient care. Regarding the psychological benefits, professionals reported that patients often expressed feeling reassured through their interactions. As MD3 mentioned, "Feedbacks from patients is that they feel reassured." Nu2 also shared, "Yeah, like the words some people say (Nu2cites patients feedback that she collected during telemonitoring calls): 'it's comforting', 'it's a positive approach', 'it's a weekly constraint that encourages us to listen better, detect the warning signs of a possible evolution of our pathology'." This sense of safety expressed by professionals was also echoed by patients, who reported a reduction in anxiety related to their disease: "It's already more reassuring, since you know that you have someone immediately in case of a problem." [Pa6]. The evolution of patient care was also highlighted, with several mentions of the responsiveness in case of issues. As Pa8 said, "in case of difference, the doctor can intervene." Another key point raised by both professionals and patients was patient empowerment. The adoption of a greater awareness of their disease was noted, with Pa5 stating, "it forces you to be vigilant, it forces you to understand a little about your disease too, [...] to know what you can do and what you cannot do. [...] we feel responsible for ourselves." However, the perceived ease of use of telemonitoring was impacted by technical issues such as "bugs," reconnection problems, and password issues, which were mentioned frequently during the interviews. As Pa1 noted, "I don't find it very ergonomic. [...] It is not easy to use, you can very easily end up bothering with passwords. [...] There are two portals one behind the other [...] in reality the second portal disappears automatically but it disappears after 30 seconds. Then it's over." Finally, one factor mentioned by health care professionals that may have impacted patients was the adaptation of the application's usage frequency in their centers. As Nu2 explained, "We negotiated with them, we tried to find out why they weren't logging on every week, they told us it was burdensome; so, we suggested they lighten up, fill in once every 15 days."

XSL•FO
RenderX

Discussion

Principal Findings

This ancillary study aimed to identify the barriers and facilitators in implementing a telemonitoring application for patients with CKD and health professionals. Although telemonitoring adoption was generally positive, through interviews with our populations and using both quantitative and qualitative approaches, we identified several areas for improvement.

Professionals were divided into 2 groups with distinct roles: nephrologists prescribed the telemonitoring, while nurses were responsible for following up with telemonitoring patients. This confirms that telemonitoring use depends on the occupation—nephrologist or nurse [20]. Indeed, the majority of nephrologists (28/29) were average or nonusers of telemonitoring, whereas most nurses (8/13) were frequent users. Some authors [20,21] distinguished the roles that nurses and physicians should play in the implementation of telemonitoring. Nurses were considered more adaptable to this new "working method" and were seen as key players due to their close relationship with patients. Our results reveal several factors that can be interpreted as facilitators for improving the use of telemonitoring among professionals. These facilitators include confidence in the security of the data collected by the application, its ease of use, and its perceived usefulness for patients. Differences were also observed across professional roles and user types in other domains. The scores in the Adaptability subdomain of CFIR, which related to workload and the time required to integrate telemonitoring into existing practice, were lower in the average/nonuser group. This issue was also mentioned repeatedly during the interviews and appears to particularly impact nurses' workloads. Modifying workload is a key factor in the adoption of telemonitoring. Indeed, both the time required for use and the ease of use can significantly affect workload, either positively (by decreasing it) or negatively (by increasing it) [17,20,21]. This issue can be linked to other subdomains of CFIR, such as readiness for implementation and planning. The results indicated that nurses felt there was no well-defined nursing protocol in their center for the use of telemonitoring. Several studies highlight the importance of having proper training and dedicated spaces for telemonitoring [20,22]. Implementing a training and support protocol for professionals, in alignment with their workload and availability, is a crucial point to consider if we aim to improve adherence to telemonitoring among health professionals [20,21,28].

The patient population in our study had a high average age, which may explain some difficulties in using the telemonitoring application [16,21]. However, there was no significant age difference between the various user profiles in our study. This effect warrants further investigation in a larger population or with a wider age range. The instructions for using the NeLLY service were to complete 1 questionnaire per week, but based on the qualitative interviews, health professionals adapted the frequency based on the needs of their patients in their centers. A total of 46 (35.9%) patients completed their questionnaires at least once a week as recommended for the telemonitoring program. Several barriers and facilitators were identified in our results, which align with existing literature. Our findings indicated that the ease of use of the telemonitoring system was perceived as high, which could serve as a facilitator for its use [22,35]. Additionally, the perceived usefulness of telemonitoring, particularly if patients believe it can improve their health or help them gain more independence in managing their disease, appears to have a positive impact on its adoption [12]. Moreover, almost half of the patients across all user groups felt that telemonitoring helped them pay more attention to their condition. Several studies have shown that telemonitoring can enhance patients' self-management skills and empower them [23,35,36]. Peer support also emerged as a potential lever for encouraging the use of the telemonitoring application. All patients interviewed during the qualitative phase mentioned receiving technical assistance or support from their relatives. Additionally, the support from professionals and family members was highlighted in the questions. Consistent with this, the literature emphasizes that the engagement of health care professionals with their patients plays a crucial role in the successful implementation of telemonitoring programs [22,37]. Finally, our results showed that some on-off users were less comfortable with the computer tool (self-efficacy) and lacked motivation to use it (behavioral intention to use). These patients also had lower scores for the ease of use of the telemonitoring system; bugs, connection issues, and password problems were frequently mentioned during the interviews. Such technical issues could have further diminished their intention to continue using telemonitoring. Technical problems were also identified among professionals, impacting their workload. This aspect has been previously recognized in the literature as a barrier, as it affects the ease of use of the device [17]. However, the results are exploratory, and the questionnaires developed did not capture the personal context of the participants, which could influence self-efficacy and, consequently, the behavior of the participants [38]. Therefore, in this context, it is valuable to combine mixed methods and 2 theoretical frameworks (CFIR and TAM) to obtain a more comprehensive understanding and explore the real-life context of the participants through qualitative interviews.

The results suggest the need for tailored implementation strategies that address the specific issues faced by both patients and professionals, as they share common and distinct challenges. However, professional acceptance is crucial, as health care professionals' perceptions and the support they offer directly influence patients' adoption of remote monitoring. Training should also emphasize the patient-professional relationship and the professional's approach when prescribing telemonitoring. The role of a dedicated facilitator, as proposed in the integrated-Promoting Action on Research Implementation in Health Services (iPARIHS) implementation framework [39], could be beneficial in offering continuous reinforcement and assisting professionals in integrating the innovation into their daily practice. This approach could include strategies tailored to both professionals and patients.

Limitations

This study has 2 key limitations. First, due to the ongoing COVID-19 health crisis, we were unable to validate our questionnaires through focus groups. Bringing together patients

XSL•FO
RenderX

with CKD during this time posed an unnecessary risk to their health. To address this challenge, we tested the patient questionnaire individually with 3 patients and 3 professionals, which allowed us to make necessary adjustments to both the form and content. Our second limitation was the small sample size of professionals. While the response rate was relatively high compared with the literature, it was too low to allow for meaningful comparisons or to provide sufficient statistical power for multivariate regression models. As a result, we were unable to quantify the strength of the associations between the identified barriers and facilitators and the use of telemonitoring. These factors were explored during semistructured interviews with nephrologists and nurses. Additionally, the survey predominantly reached frequent and average users, with only 15 participants identified as irregular users. Therefore, we can assume that the barriers identified in our sample are likely to be more significant for the broader population of irregular users.

Conclusions

This study highlighted that patients in the NeLLY trials generally adhered to telemonitoring. Among health care professionals, however, adoption varied, with nurses showing higher levels of adherence compared with nephrologists.

Nurses and nephrologists were involved in telemonitoring with patients with CKD in very different ways. Despite these distinct roles, both groups faced common barriers to the implementation of telemonitoring, such as the additional workload generated by remote monitoring. Addressing these challenges is essential to improve the adoption of telemonitoring among health care professionals. Given that support from health care professionals is crucial for patient adoption of telemonitoring, it is important to target both health care professionals and patients to enhance the implementation of telemonitoring in CKD. These factors should be taken into account when interpreting the results of the main cost-effectiveness study (PRME NeLLY) and underscore the need for trials that also evaluate the efficacy of implementation strategies.

Acknowledgments

The authors thank all the investigators of the NeLLY (New health e-Link in the Lyon region) trial: Professor Éric Alamartine, nephrologist, CHU St Etienne, Saint-Priest-en-Jarez, France; Dr Philippe Dubot, nephrologist, Service Néphrologie-Hémodialyse (Pôle 4), CH William Morey de Chalon-Sur-Saône, Chalon-sur-Saône, France; Professor Philippe Rieu, Service de Néphrologie-Transplantation, CNRS/URCA UMR 7369, CHU Reims, Reims, France; Professor Cécile Vigneau, nephrologist, CHU de Rennes, Rennes, France; Dr Emmanuel Villar, nephrologist, St Joseph St Luc, Lyon Epidémiologiste et Chargé de Recherche Laboratoire biostatistiques Santé UMR CNRS UCLB1 5558, Lyon, France; Professor Thierry Lobbedez, nephrologist, Centre Universitaire des Maladies Rénales, CHU Caen, Caen, France; Professor Denis Fouque, nephrologist, Service de néphrologie – dialyse – nutrition, Centre Hospitalier Lyon Sud, 69495 Pierre Bénite, Lyon, France.

Authors' Contributions

MD and JH conceptualized the original draft of the manuscript. MD, JH, AG, NA, JPC, ALC, and ACB designed the protocol of the study. MD collected the data. MD, LT, and JH participated in the study analysis. AG, LT, NA, JPG, ALC, and ACB critically reviewed and revised the manuscript and made appropriate edits or changes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
apTelecare log-in page.
[DOCX File , 1500 KB - mhealth_v13i1e50014_app1.docx]

Multimedia Appendix 2
Details of questions by theoretical framework in patient’s and professional's questionnaire.
[DOCX File , 18 KB - mhealth_v13i1e50014_app2.docx]

Multimedia Appendix 3
Details of scores by questions among health professionals.
[DOCX File , 16 KB - mhealth_v13i1e50014_app3.docx]

Multimedia Appendix 4
Complementary analysis.
[DOCX File , 20 KB - mhealth_v13i1e50014_app4.docx]

Multimedia Appendix 5

XSL•FO
RenderX

Comparison of variables linked to technological literacy between different types of users.
[DOCX File , 19 KB - mhealth_v13i1e50014_app5.docx]

References

1. Ayodele OE, Alebiosu CO. Burden of chronic kidney disease: an international perspective. Adv Chronic Kidney Dis 2010 May;17(3):215-224. [doi: 10.1053/j.ackd.2010.02.001] [Medline: 20439090]
2. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. Lancet 2017 Mar 25;389(10075):1238-1252. [doi: 10.1016/S0140-6736(16)32064-5] [Medline: 27887750]
3. Guide du parcours de soins – maladie rénale chronique de l'adulte (MRC). Haute Autorité de Santé. 2021 Jul. URL: https://www.has-sante.fr/upload/docs/application/pdf/2021-09/guide__mrc.pdf [accessed 2021-05-13]
4. Black C, Sharma P, Scotland G, McCullough K, McGurn D, Robertson L, et al. Early referral strategies for management of people with markers of renal disease: a systematic review of the evidence of clinical effectiveness, cost-effectiveness and economic analysis. Health Technol Assess 2010 Apr;14(21):1-184. [doi: 10.3310/hta14210] [Medline: 20441712]
5. Gordon EJ, Fink JC, Fischer MJ. Telenephrology: a novel approach to improve coordinated and collaborative care for chronic kidney disease. Nephrol Dial Transplant 2013 Apr;28(4):972-981. [doi: 10.1093/ndt/gfs552] [Medline: 23243040]
6. Efficience de la télémédecine: état des lieux de la littérature internationale et cadre d'évaluation. Haute Autorité de Santé. 2013 Jul. URL: https://www.has-sante.fr/jcms/c_1622477/fr/efficience-de-la-telemedecine-etat-des-lieux-de-la-litterature-internationale-et-cadre-d-evaluation [accessed 2021-02-19]
7. Rifkin DE, Abdelmalek JA, Miracle CM, Low C, Barsotti R, Rios P, et al. Linking clinic and home: a randomized, controlled clinical effectiveness trial of real-time, wireless blood pressure monitoring for older patients with kidney disease and hypertension. Blood Press Monit 2013 Feb 18;18(1):8-15 [FREE Full text] [doi: 10.1097/MBP.0b013e32835d126c] [Medline: 23275313]
8. Blinkhorn TM. Telehealth in nephrology health care: a review. Renal Society of Australasia Journal 2012 Nov;8(3):132-139 [FREE Full text]
9. McManus RJ, Mant J, Haque MS, Bray EP, Bryan S, Greenfield SM, et al. Effect of self-monitoring and medication self-titration on systolic blood pressure in hypertensive patients at high risk of cardiovascular disease: the TASMIN-SR randomized clinical trial. JAMA 2014 Aug 27;312(8):799-808. [doi: 10.1001/jama.2014.10057] [Medline: 25157723]
10. Ong SW, Jassal SV, Miller JA, Porter EC, Cafazzo JA, Seto E, et al. Integrating a smartphone-based self-management system into usual care of advanced CKD. Clin J Am Soc Nephrol 2016 Jun 06;11(6):1054-1062 [FREE Full text] [doi: 10.2215/CJN.10681015] [Medline: 27173169]
11. Caillette-Beaudoin A, Zaoui P, Colin C, Touzet S, Kuentz F, Fabre J, et al. Évaluation de l'efficacité d'un télésuivi à domicile de patients insuffisants rénaux chroniques, stade 3 et 4, sur la fréquence d'événements porteurs de risques et la progression de la maladie rénale chronique. European Research in Telemedicine / La Recherche Européenne en Télémédecine 2016 Dec;5(4):119-135 [FREE Full text] [doi: 10.1016/J.EURTEL.2016.10.002]
12. Tuot DS, Boulware LE. Telehealth applications to enhance CKD knowledge and awareness among patients and providers. Adv Chronic Kidney Dis 2017 Jan 24;24(1):39-45 [FREE Full text] [doi: 10.1053/j.ackd.2016.11.017] [Medline: 28224941]
13. Hailey D. Telehealth in nephrology care-promises and challenges. Am J Kidney Dis 2016 Jul;68(1):5-7. [doi: 10.1053/j.ajkd.2016.04.008] [Medline: 27343807]
14. Rocha PN. Are we ready to embrace telemedicine in nephrology? J Bras Nefrol 2015 Jul;37(3):287-288 [FREE Full text] [doi: 10.5935/0101-2800.20150046] [Medline: 26398636]
15. Soman SS, Yee J. Nephrology and telehealth: Now? or Now!. Advances in Chronic Kidney Disease 2017 Jan;24(1):1-3. [doi: 10.1053/j.ackd.2016.12.006]
16. Bonner A, Gillespie K, Campbell KL, Corones-Watkins K, Hayes B, Harvie B, et al. Evaluating the prevalence and opportunity for technology use in chronic kidney disease patients: a cross-sectional study. BMC Nephrol 2018 Dec 02;19(1):28 [FREE Full text] [doi: 10.1186/s12882-018-0830-8] [Medline: 29394930]
17. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
18. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
19. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
20. Bagot K, Moloczij N, Arthurson L, Hair C, Hancock S, Bladin CF, et al. Nurses' role in implementing and sustaining acute telemedicine: a mixed-methods, pre-post design using an extended technology acceptance model. J Nurs Scholarsh 2020 Jan;52(1):34-46. [doi: 10.1111/jnu.12509] [Medline: 31508882]

XSL•FO
RenderX

21. Paulsen MM, Varsi C, Paur I, Tangvik RJ, Andersen LF. Barriers and facilitators for implementing a decision support system to prevent and treat disease-related malnutrition in a hospital setting: qualitative study. JMIR Form Res 2019 May 09;3(2):e11890 [FREE Full text] [doi: 10.2196/11890] [Medline: 31094333]
22. Batsis JA, McClure AC, Weintraub AB, Sette D, Rotenberg S, Stevens CJ, et al. Barriers and facilitators in implementing a pilot, pragmatic, telemedicine-delivered healthy lifestyle program for obesity management in a rural, academic obesity clinic. Implement Sci Commun 2020;1:83 [FREE Full text] [doi: 10.1186/s43058-020-00075-9] [Medline: 33015640]
23. McGillicuddy JW, Weiland AK, Frenzel RM, Mueller M, Brunner-Jackson BM, Taber DJ, et al. Patient attitudes toward mobile phone-based health monitoring: questionnaire study among kidney transplant recipients. J Med Internet Res 2013;15(1):e6 [FREE Full text] [doi: 10.2196/jmir.2284] [Medline: 23305649]
24. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med 2015 Nov;13(6):554-561 [FREE Full text] [doi: 10.1370/afm.1865] [Medline: 26553895]
25. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
26. Ammenwerth E. Technology acceptance models in health informatics: TAM and UTAUT. Stud Health Technol Inform 2019 Jul 30;263:64-71. [doi: 10.3233/SHTI190111] [Medline: 31411153]
27. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4(50):1-15 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
28. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
29. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
31. R: The R Project for Statistical Computing. R Foundation. URL: https://www.r-project.org/ [accessed 2022-12-21]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. NVivo version 161 (1137). QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2022-12-21]
34. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. BMJ 2010;341:c4587. [doi: 10.1136/bmj.c4587] [Medline: 20851841]
35. Abdullah A, Liew SM, Hanafi NS, Ng CJ, Lai PSM, Chia YC, et al. What influences patients' acceptance of a blood pressure telemonitoring service in primary care? A qualitative study. Patient Prefer Adherence 2016;10:99-106 [FREE Full text] [doi: 10.2147/PPA.S94687] [Medline: 26869773]
36. Fuji KT, Abbott AA, Galt KA. A qualitative study of how patients with type 2 diabetes use an electronic stand-alone personal health record. Telemed J E Health 2015 Apr;21(4):296-300. [doi: 10.1089/tmj.2014.0084] [Medline: 25614996]
37. Record JD, Ziegelstein RC, Christmas C, Rand CS, Hanyok LA. Delivering personalized care at a distance: how telemedicine can foster getting to know the patient as a person. J Pers Med 2021 Feb 17;11(2):137 [FREE Full text] [doi: 10.3390/jpm11020137] [Medline: 33671324]
38. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY: W H Freeman/Times Books/Henry Holt & Co; 1997.
39. Harvey G, Kitson A. Implementing Evidence-Based Practice in Healthcare. London, UK: Taylor and Francis Group; 2015.

Abbreviations

CFIR: Consolidated Framework for Implementation Research
CKD: chronic kidney disease
iPARIHS: integrated-Promoting Action on Research Implementation in Health Services
NeLLY: New Health e-Link in the Lyon Region
PRME: Medical and Economic Research Program
REDCap: Research Electronic Data Capture
SW-RCT: stepped-wedge randomized controlled trial
TAM: Technology Acceptance Model

XSL•FO
RenderX

Edited by L Buis; submitted 16.06.23; peer-reviewed by H Witteman, A Rotondi, AA Sulayman; comments to author 26.03.24; revised version received 25.06.24; accepted 23.09.24; published 22.01.25.

Please cite as:
Delvallée M, Guerraoui A, Tchetgnia L, Grangier JP, Amamra N, Camarroque AL, Haesebaert J, Caillette-Beaudoin A
Barriers and Facilitators in Implementing a Telemonitoring Application for Patients With Chronic Kidney Disease and Health Professionals: Ancillary Implementation Study of the NeLLY (New Health e-Link in the Lyon Region) Stepped-Wedge Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e50014
URL: https://mhealth.jmir.org/2025/1/e50014
doi:10.2196/50014
PMID:

©Marion Delvallée, Abdallah Guerraoui, Lucas Tchetgnia, Jean-Pierre Grangier, Nassira Amamra, Anne-Laure Camarroque, Julie Haesebaert, Agnès Caillette-Beaudoin. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of mHealth on Postoperative Quality of Life, Self-Management, and Dysfunction in Patients With Oral and Maxillofacial Tumors: Nonrandomized Controlled Trial

Yufei Li[1*], PhD; Yueping Wang[2*], RN; Yifan Wu[2], PhD, RN; Hong Yu[3], RN; Hua Yao[4], RN; Yuqun Wang[5], RN; Yulan Yin[6], RN; Lan Wang[7], RN; Lili Hou[2], PhD, RN

[1]School of Nursing, Shanghai JiaoTong University, Shanghai, China

[2]Department of Nursing, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Bldg 8, 2nd Fl, 639 Manufacturing Bureau Road, Huangpu district, Shanghai, China

[3]Department of Nursing, Anhui Stomatological Hospital, Anhui Medical University, Hefei, China

[4]Department of Oral & Maxillofacial Head & Neck Oncology, Shanghai Ninth People's Hospital, Huangpu Branch, Shanghai Ninth People's Hospital, Shanghai, China

[5]Department of Nursing, Nanjing Stomatological Hospital, Nanjing Medical University, Nanjing, China

[6]Department of Oral & Maxillofacial Head & Neck Oncology, Shanghai Ninth People's Hospital, Fengcheng Branch, Shanghai Ninth People's Hospital, Shanghai, China

[7]Department of Head and Neck Oncology, Zhejiang Cancer Hospital, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Lili Hou, PhD, RN
Department of Nursing, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Bldg 8, 2nd Fl, 639 Manufacturing Bureau Road, Huangpu district, Shanghai, China

Abstract

Background: With a focus on postoperative dysfunctions that may occur after maxillofacial tumor resection and the difficulties faced during home rehabilitation, we developed a mobile health app based on nurse-patient cooperation. The app extends rehabilitation care from hospital to home with the help of artificial intelligence, Internet of Things, and other technologies, thus helping patients to better carry out their home functional rehabilitation and meet their health needs.

Objective: The primary objective of this quasi-experimental study is to evaluate the impact of the Intelligent Home Rehabilitation Care Platform on the quality of life, self-management, and functional impairment in patients with oral and maxillofacial head and neck tumors. We aim to determine whether the intervention through this platform can lead to significant improvements in these areas compared with traditional postdischarge care methods.

Methods: In this study, patients with oral and maxillofacial head and neck tumors who had undergone surgery were recruited from allied hospitals in the Yangtze River Delta region, divided into an experimental group (n=138) and a control group (n=123), and received either the Intelligent Home Rehabilitation Care Platform intervention or the conventional health care and guidance, respectively. The intervention lasted 3 months, and the patients' quality of life, self-management efficacy, and improvement in dysfunction were assessed at 1 week (baseline), 1 month (T1), and 3 months (T2) postoperatively. SPSS software (IBM Corp) was used to perform the chi-square test, rank sum test, *t* test, repeated-measures ANOVA, and generalized estimation equation for data analysis.

Results: We analyzed the effects of the quality of life and self-management using the generalized estimating equation method. The generalized estimating equation results showed that after adjusting for age, sex, pathological histology, cancer stage, and primary site, the intervention group had a significantly higher improvement in quality of life than the control group at the T2 (regression coefficient, β=−68.020, 95 % CI −116.639 to −19.412; *P*=.006) stage. The degree of improvement in self-management efficacy was significantly higher in the T1 (regression coefficient, β =−7.030, 95 % CI −9.540 to −4.520; *P*<.001) and T2 (regression coefficient, β =−13.245, 95 % CI −16.923 to −9.566; *P*<.001) stages than in the control group. The results of repeated-measures ANOVA and rank sum test showed that the experimental group showed improvements in shoulder function, dysphagia, and trismus after mHealth intervention; however, the differences were not significant.

Conclusions: The Intelligent Home Rehabilitation Care Platform interventions can effectively enhance patients' self-management efficacy, improving quality of life and facilitate recovery from dysfunctions. Therefore, mHealth may be used in oncology care to provide smarter home self-care for patients.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300073518; https://www.chictr.org.cn/showproj.html?proj=200364

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e59926) doi:10.2196/59926

KEYWORDS

head and neck cancer; quality of life; self-management efficiency; mHealth; functional rehabilitation; maxillofacial tumor; postoperative dysfunction; artifical intelligence

Introduction

Background

Oral and maxillofacial head and neck malignant tumors are highly fatal and among the top 10 malignant tumors worldwide, accounting for approximately 7% of systemic malignant tumors, with approximately 300,000 new cases and 128,000 related deaths yearly, of which oral cancer is the most common cause [1,2]. The 2024 American Cancer Society's Cancer Statistics report revealed that the incidence of oral and pharyngeal cancers has increased by an average of 1% per year, and the mortality rate from human papillomavirus–related oral cancers has shown an increase of 2% per year [3].

The treatment and care of oral and maxillofacial head and neck tumors is a comprehensive process that includes several aspects, such as surgical treatment, radiotherapy, chemotherapy, and supportive care; surgical treatment is currently the most dominant and effective method of treating oral and maxillofacial head and neck tumors [4]. The surgical approach and surgical area involved in oral and maxillofacial head and neck tumor treatment are more complex and extensive, and surgical resection inevitably damages parts of the normal tissues. In addition to the specificity of the disease, patients often experience various complications and physiological dysfunctions in the postoperative period, such as swallowing difficulties, shoulder dysfunction, and trismus, which greatly affect survival quality [5].

Therefore, it is particularly important to provide comprehensive management to patients as it can improve their functional disorders and quality of life. During hospitalization, a professional nursing team provides patients with full care and early guidance on functional rehabilitation exercises. However, functional rehabilitation is a long-term process, and studies have shown that it usually takes 3 - 6 months to recover from the symptoms and dysfunctions that occur in patients with head and neck cancer postoperatively. Therefore, patients need to continue home-based self-exercise to promote functional rehabilitation after discharge from the hospital [6]. However, some studies have shown that patients are deprived of systematic care management services after being discharged due to time and space requirements of traditional follow-ups. Notably, most patients are not well equipped to continue self-management and functional rehabilitation training in the home environment because of their lack of medical background and knowledge and poor knowledge of predischarge health education, resulting in a prolonged recovery period, poor rehabilitation effect, and high readmission rate, greatly affecting postoperative survival quality [7-9].

With the recent development and advancement of mHealth, mHealth software has been increasingly applied in the self-management and functional rehabilitation of patients with cancer and chronic disease [10]. Mobile health care supports medical and public health practices using mobile devices, such as mobile phones, personal digital assistants, patient-monitoring devices, and other wireless devices [11]. Mobile health care establishes a more convenient, comprehensive, and detailed service and communication bridge between doctors, nursing staff, and patients in hospitals, communities, and at home, which further promotes and develops the continuity of care services and, to some extent, facilitates patient self-monitoring, management, and treatment adherence; however, it is difficult to provide objective measurements based only on the software level. The popular application and technological development of the Internet of Things (IoT) also provide a boost for more convenient, accurate, and objective remote collection of medical data [12]. IoT literally means interconnected network of physical objects or "Things" integrated to exchange data between devices and systems using the web. In this regard, the Internet of Medical Things, on the other hand, is a specific app of IoT in the medical field, where remote patient monitoring, screening, and treatment are realized through wearables, sensors, and other devices telemedicine, which has been successfully adopted by caregivers or health care providers and patients [13].

The Study

Patients with oral and maxillofacial tumors need self-management and functional rehabilitation at home after operation [5], but at present, there is insufficient coordination between China Medical Union and community rehabilitation centers, which leads to the lack of professional guidance at home [14]. Most of the existing mobile medical apps support only one-way information transmission and mainly rely on patients to fill in data subjectively, which makes it difficult to guarantee the accuracy of data and seriously affects the user experience and compliance of patients [15]. Therefore, this study developed an intelligent home rehabilitation care platform integrating software and IoT equipment, aiming at expanding the scope of nursing, providing convenience for nurse-patient communication through monitoring and recording medical data, and helping patients recover after operation. The platform adopts the professional nursing process of "Assessment-Planning-Implementation-Evaluation" to help patients recover their functions and improve their quality of life. To the best of our knowledge, there is no similar comprehensive rehabilitation care platform for postoperative rehabilitation of patients with oral and maxillofacial tumors. The goal of this study is to guide and supervise patients' rehabilitation through mobile medical technology, enhance their self-management ability, and reduce the incidence of family adverse events.

XSL•FO
RenderX

Methods

Study Design and Sample

This is a quasi-experimental study of a multicenter collaboration. Four hospitals conducted the data collection. The ethical approval process and trial registration were completed before the first patient was formally enrolled in the study. The inclusion criteria were age ≥18 years; basic cognitive ability; history of oral and maxillofacial head and neck tumors (tumors originating from the oral cavity, oropharynx, hypopharynx, larynx, and other regions of the head and neck) treated with surgery only; tumor, node, and metastasis clinical stage I-III; absence of distant metastases; and at least one of the following postoperative dysfunction complications or other special care needs: dysphagia, trismus, malnutrition, shoulder dysfunction, or having been discharged with a tracheostomy tube after tracheotomy. The exclusion criteria included the presence of psychiatric disorders or cognitive dysfunction, comorbid serious cardiovascular and cerebrovascular underlying diseases (eg, cardiac infarction, cerebral infarction, and coronary artery disease), pregnancy, and other cancers or severe comorbidities.

Features of the Intelligent Home Rehabilitation Platform

The Intelligent Home Rehabilitation Care Platform developed in this study comprised 2 parts: an Android-based mobile app and Bluetooth-connected wireless IoT devices. The software allowed login access through 2 separate portals. The web-based port was available for health care professionals to log in. After logging in, health care professionals could execute modules such as patient screening and inclusion, symptom assessment, plan development, scale management, communication channels, and exception handling on the platform to develop a personalized home rehabilitation plan for patients. Each patient was given a free-to-use smart tablet before discharge (to be returned at the end of the intervention period) and logged in to the app software port on the tablet through a personal account added by the system administrator. The digital platform was intended for use with study participants only, as a tool for data collection and interaction with study participants, and did not involve any form of participant monitoring beyond the parameters of the study. After logging in, patients could access medical and health care knowledge, view and execute plans, and contact health care professionals. The platform had a suite of outreach devices, including a wireless oximeter with Bluetooth connectivity and a food scale, which can automatically collect and upload the patient's oxygen saturation metrics and daily nutritional intake.

The app software part of the platform contained 6 modules: rehabilitation calendar, training video upload, message reminder, communication channel, health promotion media library, and follow-up scale filling, which could provide 5 major home rehabilitation nursing guidelines and training: mouth opening, swallowing, shoulder exercise, nutritional intake, and tracheotomy. According to different home rehabilitation nursing needs, we developed, designed, and filmed the corresponding professional training guidance videos, which patients could follow for home rehabilitation exercises and upload according to their personalized plan schedule created by health care professionals through the platform and configured IoT devices. We also built a media database of health knowledge containing 20 live actions, animated home rehabilitation, and nursing videos for patients to watch. In addition, the platform had a communication channel through which patients could directly establish text or voice contact with health care personnel in case of emergency. Furthermore, the developed Intelligent Home Rehabilitation Care Platform was an integrated platform that could be jointly used by the Yangtze River Delta Consortium units. An additional figure file showed screenshots of the different ports on the platform (Multimedia Appendix 1).

Development, Training, and Implementation of the Intervention

The study consisted of a total of 43 multidisciplinary experts and clinical caregivers from 4 hospitals, including clinical surgeons, nutritional managers, rehabilitation therapists, nurses, and so forth. All mHealth-based interventions were constructed based on evidence-based results from the project team's previous work [16]. For example, the diet and nutrition program included recipes for total nutrient solution nasal feeding, nutritional recipes for patients on a normal diet, timing of implementation of nutrition, nutritional pathways, and nutritional energy and types; swallowing included dimensions such as oral muscle training methods and airway protection maneuvers; and the rehabilitative care program for shoulder exercise included dimensions such as passive exercise, active exercise, frequency of exercise, and duration of exercise.

Before the platform was officially used, several rounds of centralized training and guidance on its use were provided to the project implementation members. Each patient will complete a functional assessment and quality-of-life and self-efficacy survey on a one-to-one basis prior to discharge. The control and experimental groups received routine health education, home rehabilitation nursing guidance, and rehabilitation plan suggestions before discharge, whereas the experimental group further received 30 minutes of additional guidance via the mHealth platform. The extra services included helping participants and their caregivers recognize and use the Intelligent Home Rehabilitation Care Platform and its outreach smart devices and explaining the interventions that need to be completed when using the platform at home: the implementation of an intervention plan individually tailored to them.

The intervention period was for 3 months. At each follow-up node, when the patient's dysfunctions had returned to a normal level, the rehabilitation program was terminated, but the other platform functions continued to be used until the end of the 3-month period.

In this paper, we focus on the impact of mHealth on postoperative quality of life, self-management efficacy, and functional improvement of shoulder function, dysphagia, and trigeminal neuralgia in patients with oral and maxillofacial tumors. The analysis of nutritional improvement and tracheotomy care and the effects and economic benefits of the Intelligent Home Rehabilitation Care Platform will be elaborated in detail in a subsequent paper.

XSL•FO
RenderX

Data Collection

We conducted clinical data collection in 12 oral and maxillofacial head and neck tumor wards in 4 regional units of the Yangtze River Delta Consortium. To minimize the interference of confounding factors in the design of nonrandomized controlled study, and considering the resource allocation and total number of patients in different wards in multicenter, wards were naturally allocated. Different wards are not allocated according to disease risk level or tumor-site preference. To be considered for the symptom management program and rehabilitation strategies we offered, patients were required to have 1 or more of the following postoperative complications at the sample inclusion stage: malnutrition, dysphagia, shoulder dysfunction, trismus, or having been discharged with a tube after tracheotomy. For patients with multiple symptoms, we established a corresponding intervention plan for each symptom and included them in different symptomatic subgroups that were analyzed at the same time.

Before discharge, each patient was individually instructed by the researcher on disease-related health knowledge, postdischarge functional rehabilitation exercise methods and points, airway home care, and follow-up review reminders. In this study, the data collected from patients at 1 week postoperatively (T0) were used as baseline data, and electronic questionnaires were used to investigate the outcome of the intervention in the 2 patient groups at the first (T1) and third (T2) postoperative months.

Questionnaires

The questionnaire used in this study was a general information questionnaire that included questions about patients' demographic characteristics (age, sex, marital status, and education level), clinical characteristics (cancer stage, primary site of cancer, and days of hospitalization), and the University of Washington Quality of Life Scale (UW-QOL; version 4), which is a patient-specific quality-of-life questionnaire for head and neck cancers [17]. The UW-QOL consists of 15 items with 12 disease-specific questions on pain, appearance, activity, recreation, swallowing, chewing, speech, shoulder, taste, saliva, mood, and anxiety, and 3 composite questions. The scale has been translated into several versions, and the Chinese version also has good reliability, with a Cronbach coefficient of 0.90 for the UW-QOL. Strategies Used by People to Promote Health (SUPPH) was used to assess patients' self-management efficacy [18]. The Chinese version of the SUPPH scale consists of 28 items in the areas of self-determination, stress relief, and positive attitude. The SUPPH evaluation index is calculated as (the actual score on the scale/the total score on the scale) × 100%, with ≥80%, 61% - 79%, and ≤60% categorized as high, medium, and low levels, respectively. The Cronbach coefficient of the SUPPH (Chinese version) is 0.97. All patients were assessed using the UW-QOL and SUPPH scales, which was the primary outcome indicator in this study.

For symptom assessment, symptomatic patient groups were individually assessed using targeted scales. The Constant-Murley Shoulder Score was used to evaluate shoulder dysfunction, the Kubota Water Swallowing Test was used to evaluate dysphagia, and the Subjective Objective Management and Analytic criteria were used to evaluate trismus.

Simple Size and Power

With no previous experiments from which to obtain expected significant differences or sample SDs, we conducted a small pilot trial to calculate the sample size before the start of the formal study, which collected preintervention-postintervention changes in quality-of-life and self-management efficacy scores using the UW-QOL and the SUPPH. We calculated the pooled sample SD using the pooled results, and the sample size was calculated using G*Power (version 3.1; Heinrich-Heine-Universität Düsseldorf), with test levels chosen as α=.01 and β=.1. The calculations showed that at least 104 patients should be collected in each of the experimental and control groups. Thus, at an assumed attrition rate of 20%, the final formal experimental sample sizes of the experimental group and the control group should each recruit at least 130 patients.

Statistical Analysis

Continuous variables with normally distributed data are expressed as mean (SD), nonnormal data are expressed as median with quartiles (median, P25-P75), and categorical variables are expressed as n (%). Differences in categorical variables were tested by chi-square test, rank sum test, or Fisher exact test, and differences between groups for continuous variables at each time point were tested by independent *t* test. Differential changes in primary outcomes (quality of life and self-management) at T1 and T2 compared with T0 between the 2 groups were assessed using generalized estimating equations, and repeated-measures ANOVA was used to assess the improvement in functioning (shoulder functioning) between the 2 groups at T1 and T2. Generalized estimating equation analyses were used because the method takes into account internal correlations between repeated measures. Adjusted generalized estimating equation analyses were performed using covariates, including age, sex, pathological histology, cancer stage, and primary site. All statistical assessments were double-tested, and statistical significance was set at $P<.05$. Statistical analyses were performed using IBM SPSS (version 25.0; IBM Corp).

Ethical Considerations

This study was approved by the Institute in the Hospital Ethics Committee (SH9H-2023-T157-1) and abided by the principles of Declaration of Helsinki. All patients who participated in the study participated voluntarily and obtained the informed consent of each patient in writing. Patients have the right to withdraw from the study at any time. Due to the principle of confidentiality, each participant's information will be anonymized and given a code during the data analysis stage. Individuals cannot be identified in any results presentation. Additionally, no compensation was provided to study participants.

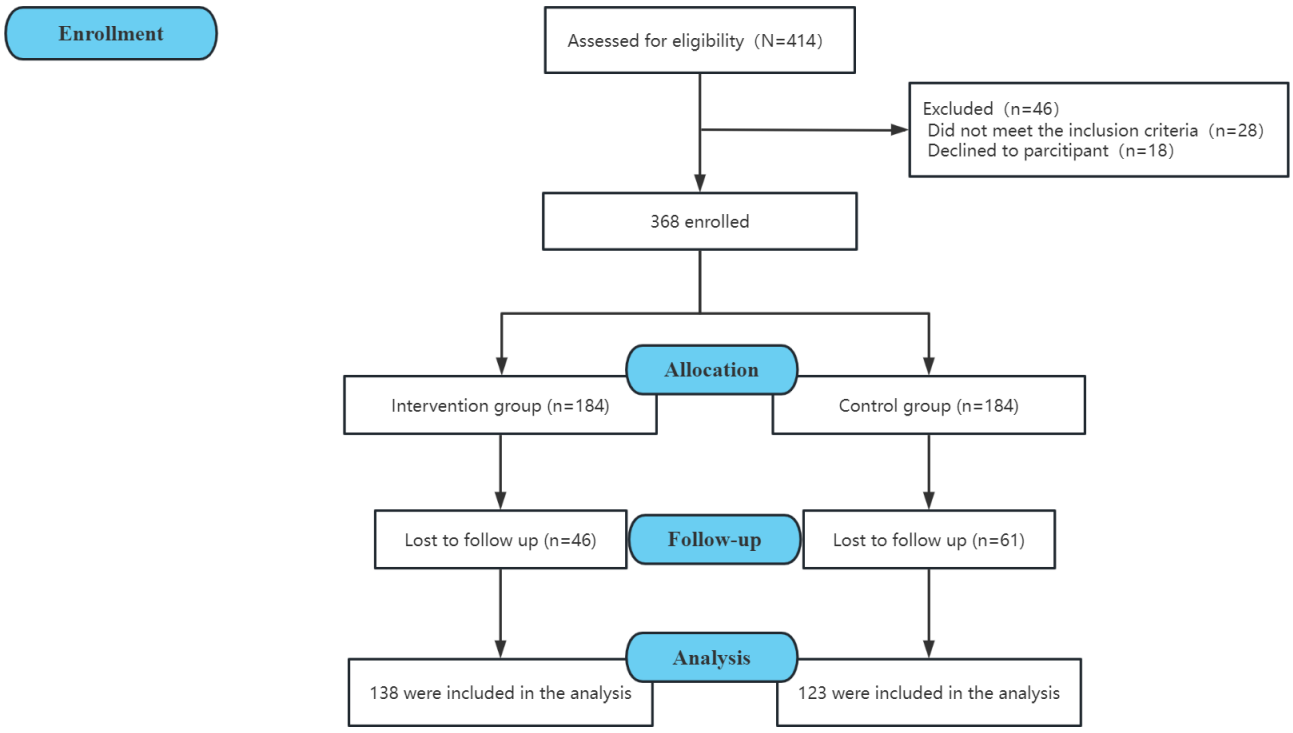
XSL•FO
RenderX

Results

Analytical Sample

A total of 368 participants (184 in the experimental group and 184 in the control group) were enrolled in this study. After a 3-month intervention period with 2 follow-up measurement time points, 25% (46/184) of patients in the experimental group and 33% (61/184) of patients in the control group were lost to visit or dislodging. The final sample included 138 and 123 patients in the experimental and control groups, respectively (Figure 1).

Figure 1. Flowchart of included participants.



Participants' Background Characteristics

A total of 261 patients with oral and maxillofacial head and neck tumors who had undergone surgery were recruited in this study, including 162 men and 99 women, with a mean age of 54.84 (SD 14.45) years. The marital status of most patients was married (234/261, 89.7%), and the more predominant education level was middle school (82/261, 31.4%). Furthermore, most patients had stage II tumors (109/261, 41.8%), the predominant primary site of the tumor was the tongue (116/261, 44.4%), and the median length of hospitalization was 18 days (Table 1).

XSL•FO
RenderX

Table . Participants' baseline demographics and clinical characteristics (N=261).

Variable	Total (N=261）	Experimental group (n=138）	Control group (n=123）	*P* value
Demographics				
Sex, n (%)				<.01[a]
Male	162 (62.1)	97 (70.3)	65 (52.8)	
Female	99 (37.9)	41 (29.7)	58 (47.2)	
Age (years), mean (SD)	54.84 (14.45)	55.34 (13.17)	54.28 (15.80)	.56
Marital status, n (%)				.09
Married	234 (89.7)	122 (88.4)	112 (91.1)	
Unmarried	12 (4.6)	4 (2.9)	8 (6.5)	
Other (widowed/divorced）	15 (5.7)	12 (8.7)	3 (2.4)	
Education, n (%)				.14
Below primary school	60 (23.0)	28 (20.3)	32 (26.0)	
Middle school	82 (31.4)	52 (37.7)	30 (24.4)	
High school	57 (21.8)	28 (20.3)	29 (23.6)	
Undergraduate and above	62 (23.8)	30 (21.7)	32 (26.0)	
Career status, n (%)				.58
In-service	113 (43.3)	56 (40.6)	57 (46.3)	
Separated/retired	128 (49.0)	70 (50.7)	58 (47.2)	
Unemployed	20 (7.7)	12 (8.7)	8 (6.5)	
Smoking status, n (%)				.36
Smoking	39 (14.9)	22 (15.9)	17 (13.8)	
Quit smoking	77 (29.5)	45 (32.6)	32 (26.0)	
Never smoking	145 (55.6)	71 (51.4)	74 (60.2)	
Alcohol status, n (%)				.08
Regular alcohol	27 (10.3)	15 (10.9)	12 (9.8)	
Quit alcohol	89 (34.1)	55 (39.9)	34 (27.6)	
Never had alcohol	145 (55.6)	68 (49.3)	77 (47.1)	
Clinical characteristic				
Pathological histology, n (%)				.63
Squamous cell carcinomas	209 (80.1)	112 (81.2)	97 (78.9)	
Adenoid cystic carcinoma	23 (8.8)	13 (9.4)	10 (8.1)	
Others	29 (11.1)	13 (9.4)	16 (13.0)	
Cancer stage, n (%)				.38
I	83 (31.8)	40 (29.0)	43 (35.0)	
II	109 (41.8)	63 (45.7)	46 (37.4)	
III	69 (26.4)	35 (25.4)	34 (27.6)	
Primary site, n (%)				.09
Tongue	116 (44.4)	66 (47.8)	50 (40.7)	
Buccal side	49 (18.8)	24 (17.4)	25 (20.3)	

XSL•FO
RenderX

Variable	Total (N=261）	Experimental group (n=138）	Control group (n=123）	*P* value
Hard palate	16 (6.1)	7 (5.1)	9 (7.3)	
Gums	24 (9.2)	13 (9.4)	11 (8.9)	
Parotid and submaxillary glands	10 (3.8)	1 (0.7)	9 (7.3)	
Throat	17 (6.5)	12 (8.7)	5 (4.1)	
Lip	2 (0.8)	0 (0)	2 (1.6)	
Mouth floor	3 (1.2)	2 (1.5)	1 (0.9)	
Jaws	24 (9.2)	13 (9.4)	11 (8.9)	
Length of hospitalization (days), median (IQR)	18 (15-21)	19 (16-22)	18 (14-21)	.08

[a]*P*<.05.

mHealth Intervention Improved Quality of Life

The UW-QOL scores of the patients are shown in Table 2; higher scores are associated with better quality of life. At baseline, the total quality-of-life scores of the experimental and control groups were 625.38 and 647.11. There were no statistical intergroup differences at baseline (mean difference −21.73, SD 21.90; 95 % CI −64.86 to 21.4; *P*=.322). However, after 1 month of intervention (T1), the total quality-of-life scores of the experimental and control groups increased to 826.38 and 875.04, respectively, and there was a significant difference in the quality of life between the 2 groups at this stage (*P*<.01). After 3 months of intervention (T2), the total quality-of-life scores of the experimental group and the control group improved to 957.88 and 911.59, respectively, and there was a significant difference between the 2 groups in terms of improvement in quality of life (*P*<.01), and the overall improvement in quality of life of the experimental group was higher than that of the control group (mean difference 46.29, SD 16.83; 95 % CI 13.14-79.44; *P*=.006).

Table . Follow-up of primary outcomes among 2 groups based on University of Washington Quality of Life Scale.

Outcome by group	T0 (baseline)			T1 (1 month)			T2 (3 months)		
	Mean (SD)	Difference between groups (MD,95% CI)	*P* value	Mean (SD)	Difference between groups (MD,95% CI)	*P* value	Mean (SD)	Difference between groups (MD,95% CI)	*P* value
Experimental group	625.38 (170.74)	−21.73, −64.86 to 21.40	.322	826.38 (129.33)	−48.66, −84.10 to −13.23	.007[a]	957.88 (911.59)	46.29, 13.14 to 79.44	.006[a]
Control group	647.11 (183.00)			875.04 (161.02)			911.59 (141.07)		

[a]*P*<.05.

Furthermore, the generalized estimating equation models adjusted for age, sex, pathological histology, cancer stage, and primary site showed significant interaction terms (group × time) at 3 months (β=−68.020, 95 % CI −116.639 to −19.412; *P*=.006; Table 3), which validating that Intelligent Home Rehabilitation Platform–based intervention was effective in improving patients' quality of life.

XSL•FO
RenderX

Table . Modeling changes in the total score of University of Washington Quality of Life Scale over time by adjusting side variables using the generalized estimating equation model (N=261).

UW-QOL[a]	β	SE	95% CI	Wald χ^2	*P* value[b]
Group	15.109	22.438	−28.869 to 59.086	0.453	.501
Time					
Baseline	Reference				
T1	201.000	14.185	173.199 to 228.801	200.795	<.001[c]
T2	332.500	15.329	302.455 to 362.545	470.474	<.001[c]
Group × time					
Group × T0	Reference				
Group × T1	26.935	23.391	−18.911 to 72.781	1.326	.250
Group × T2	−68.020	24.801	−116.639 to −19.412	7.522	.006[c]

[a]UW-QOL: University of Washington Quality of Life Scale.

[b] Adjusted with age, sex, pathological histology, cancer stage, and primary site.

[c]*P*<.05.

mHealth Intervention Improved Self-Management Efficacy

Table 4 shows the statistical results of the patients' SUPPH scores, with higher scores indicating higher self-management efficacy. At baseline, the self-management efficacy scores of the experimental and control groups were 64.51 and 65.79, respectively. There were no statistical intergroup differences at baseline (mean difference −1.274, SD 2.50; 95 % CI −6.19 to 3.64; *P*=.610). After 1 month of intervention, the self-management efficacy scores of the experimental and control groups increased to 83.54 and 77.78, respectively, and the difference between the 2 groups was significant at this stage (mean difference 5.76, SD 2.22; 95 % CI 1.38-10.13; *P*=.01). After 3 months of intervention, the self-management efficacy scores of the experimental group and the control group further improved to 101.44 and 89.47, respectively, with the experimental group's self-management efficacy scores being significantly higher than those of the control group (mean difference 11.97, SD 2.01; 95 % CI 8.01-15.93; *P*<.001).

Table . Participants' (N=261) Strategies Used by People to Promote Health–measured self-management efficacy scores before and after the mHealth intervention.

Outcome by group	T0 (baseline)			T1 (1 month)			T2 (3 months)		
	Mean (SD)	Difference between groups (MD,95% CI)	*P* value	Mean (SD)	Difference between groups (MD,95% CI)	*P* value	Mean (SD)	Difference between groups (MD,95% CI)	*P* value
SUPPH[a]									
Experimental group	64.51 (19.63)	−1.27, −6.19 to 3.64	.610	83.54 (18.15)	5.76, 1.38 to 10.13	.01[b]	101.44 (15.61)	11.97, 8.01 to 15.93	<.001[b]
Control group	65.79 (20.70)			77.78 (17.65)			89.47 (16.86)		

[a]SUPPH: Strategies Used by People to Promote Health.

[b]*P*<.05.

In addition, the interaction term (group× time) of the 1-month (β=−7.030, 95% CI −9.540 to −4.520; *P*<.001) and 3-month (β=−13.245, 95% CI −16.923 to −9.566; *P*<.001) adjusted models showed that, over time, patients who received the smart platform intervention had a self-management efficacy improved further (Table 5). These results suggest that the Intelligent Home Rehabilitation Platform is effective in improving patients' self-management abilities during the home rehabilitation period.

XSL•FO
RenderX

Table . Modeling changes in the total score of Strategies Used by People to Promote Health over time by adjusting side variables using the generalized estimating equation model (N=261).

SUPPH[a]	β	SE	95% CI	Wald χ^2	*P* value[b]
Group	.610	2.547	−4.381 to 5.601	0.057	.811
Time					
Baseline	Reference				
T1	19.022	0.930	17.200 to 20.844	418.706	<.001[c]
T2	36.928	1.376	34.231 to 39.625	720.181	<.001[c]
Group × time					
Group × T0	Reference				
Group × T1	−7.030	1.281	−9.540 to −4.520	30.132	<.001[c]
Group × T2	−13.245	1.877	−16.923 to −9.566	49.799	<.001[c]

[a]SUPPH: Strategies Used by People to Promote Health.

[b]Adjusted with age, sex, pathological histology, cancer stage, and primary site.

[c]*P*<.05.

mHealth Intervention Improved Dysfunctions

To assess patients' symptoms and develop a personalized intervention plan, we additionally assessed the patients with dysfunctions (reduced shoulder mobility, dysphagia, and trismus) in both groups at baseline and, based on the assessment results, we developed a rehabilitation program for these patients and evaluated them using the same measurement tools at 1 and 3 months postoperatively to confirm the effectiveness of the platform intervention. Tables S1-S5 in Multimedia Appendix 2 present the data on improvements in shoulder function, swallowing function, and mouth opening in both groups after the Intelligent Home Rehabilitation Care Platform intervention.

Shoulder Function

Among the recruited patients, 24 patients who developed shoulder dysfunction in the postoperative period were included (13 in the experimental group and 11 in the control group). Shoulder function was scored on a total point scale, with higher total scores associated with better shoulder function. At baseline, the mean shoulder function scores of the experimental and control groups were 60.77 and 55.64, respectively, with no statistical difference between the 2 groups. After 1 month of intervention, the mean shoulder function scores of the experimental and control groups increased to 76.38 and 70.09, respectively; after 3 months of intervention, the mean shoulder function scores of the experimental and control groups further increased to 84.46 and 77.45, respectively. There was no significant difference in the functional status of the shoulder joint between the 2 groups at the 2 measurement points. Repeated-measures ANOVA results showed that the 2 groups were statistically different in terms of the time effect at *P*<.001, indicating that the change in shoulder function scores over time at each time point was statistically different.Detailed results are shown in Table S1 in Multimedia Appendix 2.

Dysphagia

Among the recruited patients, 43 patients who developed dysphagia in the postoperative period were included (19 in the experimental group and 24 in the control group). The swallowing test was assessed on a scale of 5 grades: I-V, where the smaller the grade, the better the swallowing function. In the evaluation of outcome indicators, "Resolved" indicated that dysphasia had disappeared, and the level of the water-swallowing test was classified as grade I; "Significantly effective" indicated that dysphasia had significantly improved, and the level of the water-swallowing test was increased by 2 grades; "Improved" indicated that dysphasia had improved, and the level of the water-swallowing test was increased by 1 grade; and "Ineffective" indicated that the improvement of dysphasia was insignificant, and the level of the water-swallowing test was unchanged. These improvement grades were considered effective.

At baseline, patients in the experimental group and the control group had different degrees of dysphagia, with no significant difference between the 2 groups. After 1 month of intervention, the water-swallowing test grades of patients in the control and experimental groups improved. After 3 months of intervention, most patients were resolved, which was a significant improvement from baseline. Treatment effectiveness rates in the experimental and control groups were 94.74% and 87.50%, respectively. There was no significant difference in the effect of treatment on swallowing function between the 2 groups; however, the overall effectiveness rate in the experimental group was higher than that in the control group, and the specific data are shown in detail in Tables S2 and S3 in Multimedia Appendix 2.

Trismus

Among the recruited patients, 73 patients who developed postoperative trismus were included (38 in the experimental group and 35 in the control group) . The degree of trismus was evaluated according to the Subjective Objective Management and Analytic criteria: 0 indicated normal mouth opening, with an incisor distance of >3.5 cm; grade I indicated restricted mouth opening, with an incisor distance of 2.1 - 3.0 cm; grade II indicated difficulty in dry food intake, with an incisor distance

XSL•FO
RenderX

of 1.1 - 2.0 cm; grade III indicated difficulty in soft food intake, with an incisor distance of 0.5 - 1.0 cm; and grade IV indicated the need for nasal feeding, with an incisor distance of <0.5 cm. The higher the grade, the more severe the degree of trismus. Regarding efficacy evaluation, the same criteria as that used in the dysphagia component were used to evaluate efficacy.

At baseline, patients in the experimental and control groups had grades I-III trismus, with no significant difference between the 2 groups. After 1 month of intervention, both the control and experimental groups showed improvement in the degree of trismus. After 3 months of intervention,>50% of the patients were resolved, which was a significant improvement from baseline. The treatment effectiveness in the experimental and control groups was 84.21% and 91.43%, respectively. The effect of treatment on trismus in the 2 groups was not significant between the 2 groups, as shown in Tables S4 and S5 in Multimedia Appendix 2.

Discussion

Principal Findings

Previous research has shown that patients with oral and maxillofacial head and neck tumors experience deterioration in their health status and symptoms during the acute treatment phase within 3 months. This deterioration may be attributed to the adverse effects of treatments such as surgery and radiotherapy. However, symptoms tend to improve and gradually stabilize between 3 and 6 months after treatment [6]. A prospective study revealed that physiological status (fatigue, pain, nausea, and vomiting) and physical activity (resuming leisure activities and sustainable work) significantly worsened within the first 3 months after oral cancer surgery [19]. Postoperatively, patients experience a significant decline in social functioning, body image, and financial status. Therefore, we administered an intervention using the Intelligent Home Rehabilitation Care Platform to address these challenges within the first 3 months postoperatively, and the results indicate a significant improvement in shoulder function, dysphagia, and trismus in patients after 3 months postoperatively, with some returning to normal function. These findings were consistent with those of a previous study [6].

Reports suggest that the quality of life of survivors of head and neck cancer is generally lower during different postoperative periods, affecting their survival rate [20,21]. This may be due to disease-related and treatment-related factors, leading to physiological impairment and body image issues [22]. Symptom control and functional improvement are long-term tasks. Health care professionals typically provide health education through verbal instructions or knowledge manuals upon discharge; however, patients often have limited understanding of and compliance with these methods and lack medical and health knowledge [8,23-25]. Therefore, mobile health care provides a solution for improving patients' access to medical knowledge and providing scientific guidance for family rehabilitation.

Previous studies explored the impact of different forms of mobile health care on improving quality of life, care needs, psychological and physiological functions, and other aspects of patients with head and neck cancer [26]. However, these mobile health care forms are primarily mobile apps, websites, or other wearable devices, and the involvement of health care professionals is limited. In contrast, the developed Intelligent Home Rehabilitation Care Platform in this study closely connects patients with health care professionals. Before discharge, health care professionals assess the patients and personalize home-based rehabilitation training plans based on each patient's symptoms and characteristics. Patients follow these plans and receive reminders when they forget. Furthermore, patients can obtain medical and health knowledge and contact health care professionals on the platform, which, to a certain extent, also increases patients' adherence to rehabilitation training and helps them build up their confidence and self-management ability.

This study's results showed that using the Intelligent Home Rehabilitation Care Platform significantly improved patients' quality of life at 3 months postoperatively compared with the baseline level; however, such difference was not observed at 1 month postoperatively, possibly due to the shorter duration of the intervention. There was a difference in the improvement of quality of life between the treatment and control groups at 3 months postoperatively, suggesting that the effect of the intervention becomes progressively more significant as the intervention time increases.

We also obtained better feedback on self-management efficacy, which was not significantly different between the 2 groups at baseline, and was evaluated based on the SUPPH scores. We observed that self-management efficacy was low in both groups at baseline, possibly because the patients had just received surgical treatment, experienced physiological discomfort, and had somatic limitations. At 1 month postoperatively, the self-management efficacy scores in both groups improved; however, the overall score was still low. At 3 months postoperatively, the self-management efficacy of both groups improved significantly, with the efficacy scores in the control group increasing to a medium level and that in the experimental group increasing to a high level, suggesting that mHealth interventions can significantly increase patients' self-management information and efficacy, which has not been reported in previous studies.

Shoulder dysfunction is usually due to functional impairment caused by clearance of the operative area, and studies have shown that adherence to functional shoulder exercises can effectively improve postoperative shoulder dysfunction [27]. We developed a comprehensive rehabilitation program for patients with shoulder dysfunction to improve shoulder function through pendulum exercises and resistance training. In this study, there was no significant difference in the shoulder functional status between the 2 groups at the 2 measurement points. Overall, the experimental group demonstrated better improvements than the control group. Similarly, we did not observe significant differences in the results of dysphagia and trismus measurements; however, the overall efficacy rate in the experimental group was higher than that in the control group, possibly due to the small sample size of this study. In addition, the objective indicators did not demonstrate statistically significant differences compared with the subjective indicators,

XSL•FO
RenderX

possibly indicating that the intervention implemented in this study was more effective at the level of subjective perception improvement compared with the objective measures.

The results of this study are important for guiding the postoperative care of patients with oral and maxillofacial tumors. The implementation of the Smart Home Rehabilitation Nursing Platform demonstrates that it has the potential to greatly enhance patient self-management and improve quality of life during the critical postoperative period. By using mHealth technology, the platform provides a new way to extend rehabilitation care from the hospital to the home environment, thereby bridging a critical gap in the continuum of care. The combination of artificial intelligence and the IoT makes remote collection of health care data easier, more accurate, and objective, thereby informing personalized care plans and facilitating timely medical interventions. In addition, the platform provides patients with accessible medical knowledge and ongoing support from health care professionals, which may increase patient adherence to rehabilitation programs and improve outcomes. Therefore, we believe that the adoption of this technology could change the traditional model of postoperative care by emphasizing the importance of patient-centered technology–assisted rehabilitation strategies in improving patient autonomy and overall well-being.

Limitations and Recommendations

This study has some limitations. First, this study did not use a rigorous randomized controlled trial design and did not set up a blind method, which may lead to confounding factors such as placebo effect, so we need to use randomized clinical trials to further verify our results in the future. Second, in practice, patients do not present with symptoms individually but usually in clusters; however, in this study, the strategy for symptom management was to intervene separately for multiple symptoms, which may have increased the workload of health care professionals and limited the efficiency of symptom management. Third, owing to time constraints, we implemented only a 3-month follow-up intervention and did not continue the follow-up to a longer time point after the intervention period; although we achieved relatively positive results, there is a need to expand the sample size as well as validate the impact of the intervention on long-term efficacy and quality of life in the future through larger-scale replication and longer follow-up interventions. In addition, the attrition rate of both groups in this study is high. Although the potential impact has been reduced as much as possible by expanding recruitment, future research should further optimize the research process and improve the platform experience to better cope with the challenge of sample loss. Finally, this study involved multicenter data collection in the Yangtze River Delta region of China, and the results may not apply to patients with oral and maxillofacial head and neck tumors in other regions (western or northern regions).

Conclusions

This study's results suggest that compared with conventional follow-up measures, interventions using the Intelligent Home Rehabilitation Care Platform can effectively enhance patients' self-management efficacy, promoting dysfunctional rehabilitation and improving postoperative survival quality.

Acknowledgments

This study was supported by the Shanghai Association for Science and Technology (21002411300), the Innovation Research Team of High-level Local Universities in Shanghai (SHSMU-ZDCX20212802), and Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine: Excellent Nursing Talent Program (JYHRC22-J04).

Authors' Contributions

LH was responsible for the design, implementation, and management of the entire project. YL was responsible for the development of the software, project promotion and management, data analysis, and the writing of the paper. YW was responsible for the implementation of the project and the collection of clinical data and the cooperation and communication between multicenters. YW was responsible for the development of the software, management of the project equipment, and data analysis. YW, HY, YY, HY, and LW were responsible for the collection of clinical data subsite managers. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the interfaces of the app.
[DOCX File, 5454 KB - mhealth_v13i1e59926_app1.docx]

Multimedia Appendix 2
Data on improvements in shoulder function, swallowing function, and mouth opening.
[DOCX File, 24 KB - mhealth_v13i1e59926_app2.docx]

Checklist 1

XSL•FO
RenderX

CONSORT-EHEALTH checklist.
[PDF File, 1050 KB - mhealth_v13i1e59926_app3.pdf]

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020 Jan;70(1):7-30. [doi: 10.3322/caac.21590] [Medline: 31912902]
2. Lin HH, Lin YH, Hwang TZ, Wang CC. The effects of a diet education program on nutritional status and quality of life in oral cancer patients who underwent surgery. Hu Li Za Zhi 2020 Feb;67(1):33-43. [doi: 10.6224/JN.202002_67(1).06] [Medline: 31960395]
3. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74(1):12-49. [doi: 10.3322/caac.21820] [Medline: 38230766]
4. Shanti RM, O'Malley BW. Surgical management of oral cancer. Dent Clin North Am 2018 Jan;62(1):77-86. [doi: 10.1016/j.cden.2017.08.005] [Medline: 29126495]
5. Parke SC, Langelier DM, Cheng JT, Kline-Quiroz C, Stubblefield MD. State of rehabilitation research in the head and neck cancer population: functional impact vs. impairment-focused outcomes. Curr Oncol Rep 2022 Apr;24(4):517-532. [doi: 10.1007/s11912-022-01227-x] [Medline: 35182293]
6. Ren J, Pang W, Hueniken K, et al. Longitudinal health utility and symptom-toxicity trajectories in patients with head and neck cancers. Cancer 2022 Feb 1;128(3):497-508. [doi: 10.1002/cncr.33936] [Medline: 34597435]
7. Nund RL, Rumbach AF, Debattista BC, et al. Communication changes following non-glottic head and neck cancer management: the perspectives of survivors and carers. Int J Speech Lang Pathol 2015 Jun;17(3):263-272. [doi: 10.3109/17549507.2015.1010581] [Medline: 25764915]
8. Oswald LB, Arredondo B, Geiss C, et al. Considerations for developing supportive care interventions for survivors of head and neck cancer: a qualitative study. Psychooncology 2022 Sep;31(9):1519-1526. [doi: 10.1002/pon.5985] [Medline: 35726386]
9. Ellis J, von Mücke Similon M, Korman MB, et al. Using the Delphi method to elucidate patient and caregiver experiences of cancer care. J Patient Exp 2022;9:23743735221092633. [doi: 10.1177/23743735221092633] [Medline: 35450086]
10. Zan S, Agboola S, Moore SA, Parks KA, Kvedar JC, Jethwani K. Patient engagement with a mobile web-based telemonitoring system for heart failure self-management: a pilot study. JMIR Mhealth Uhealth 2015 Apr 1;3(2):e33. [doi: 10.2196/mhealth.3789] [Medline: 25842282]
11. Smith R, Menon J, Rajeev JG, Feinberg L, Kumar RK, Banerjee A. Potential for the use of mHealth in the management of cardiovascular disease in Kerala: a qualitative study. BMJ Open 2015 Nov 17;5(11):e009367. [doi: 10.1136/bmjopen-2015-009367] [Medline: 26576813]
12. Kong HJ, An S, Lee S, et al. Usage of the Internet of Things in medical institutions and its implications. Healthc Inform Res 2022 Oct;28(4):287-296. [doi: 10.4258/hir.2022.28.4.287] [Medline: 36380426]
13. Dwivedi R, Mehrotra D, Chandra S. Potential of Internet of Medical Things (IoMT) applications in building a smart healthcare system: a systematic review. J Oral Biol Craniofac Res 2022;12(2):302-318. [doi: 10.1016/j.jobcr.2021.11.010] [Medline: 34926140]
14. Hong Y, Wang J, Zhang X, Zhao D, He H, Sun M. The effectiveness of home rehabilitation based on BPMpathway exercise rehabilitation system for patients after total knee arthroplasty in China. Int J Orthop Trauma Nurs 2024 May;53:101062. [doi: 10.1016/j.ijotn.2023.101062] [Medline: 37891073]
15. da Silva HEC, Santos GNM, Ferreira Leite A, et al. The feasibility of telehealth in the monitoring of head and neck cancer patients: a systematic review on remote technology, user adherence, user satisfaction, and quality of life. Support Care Cancer 2022 Oct;30(10):8391-8404. [doi: 10.1007/s00520-022-07109-z] [Medline: 35524146]
16. Gu F, Wang Y, Yang W, et al. Expert consensus on postoperative rehabilitation nursing of patients with head and neck cancer. J Shanghai Jiao Tong Univ (Med Sci) 2023;43(10):1289-1296. [doi: 10.3969/j.issn.1674-8115.2023.10.010]
17. Rogers SN, Gwanne S, Lowe D, Humphris G, Yueh B, Weymuller EA Jr. The addition of mood and anxiety domains to the University of Washington quality of life scale. Head Neck 2002 Jun;24(6):521-529. [doi: 10.1002/hed.10106] [Medline: 12112548]
18. Lev EL, Owen SV. A measure of self-care self-efficacy. Res Nurs Health 1996 Oct;19(5):421-429. [doi: 10.1002/(SICI)1098-240X(199610)19:5<421::AID-NUR6>3.0.CO;2-S] [Medline: 8848626]
19. Dunnack Yackel H, Xu W, Lee JW, et al. Symptom patterning across the cancer care trajectory for patients receiving chemoradiation for head and neck cancer: a retrospective longitudinal study using latent transition analysis. Cancer Nurs 2024;47(4):261-270. [doi: 10.1097/NCC.0000000000001227] [Medline: 36881642]
20. Thompson TL, Pagedar NA, Karnell LH, Funk GF. Factors associated with mortality in 2-year survivors of head and neck cancer. Arch Otolaryngol Head Neck Surg 2011 Nov;137(11):1100-1105. [doi: 10.1001/archoto.2011.179] [Medline: 22106233]
21. Chaukar DA, Walvekar RR, Das AK, et al. Quality of life in head and neck cancer survivors: a cross-sectional survey. Am J Otolaryngol 2009;30(3):176-180. [doi: 10.1016/j.amjoto.2008.05.001] [Medline: 19410123]

XSL•FO
RenderX

22. So WKW, Chan RJ, Chan DNS, et al. Quality-of-life among head and neck cancer survivors at one year after treatment—a systematic review. Eur J Cancer 2012 Oct;48(15):2391-2408. [doi: 10.1016/j.ejca.2012.04.005] [Medline: 22579456]
23. Wang TF, Huang RC, Yang SC, Chou C, Chen LC. Evaluating the effects of a mobile health app on reducing patient care needs and improving quality of life after oral cancer surgery: quasiexperimental study. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18132. [doi: 10.2196/18132] [Medline: 32716303]
24. Wang TF, Li YJ, Chen LC, Chou C, Yang SC. Correlation between postoperative health-related quality of life and care needs of oral cancer patients. Cancer Nurs 2020;43(1):12-21. [doi: 10.1097/NCC.0000000000000677] [Medline: 30433894]
25. So WKW, Wong CL, Choi KC, et al. A mixed-methods study of unmet supportive care needs among head and neck cancer survivors. Cancer Nurs 2019;42(1):67-78. [doi: 10.1097/NCC.0000000000000542] [Medline: 28945632]
26. Caputo MP, Rodriguez CS, Padhya TA, Mifsud MJ. Telehealth interventions in head and neck cancer patients: a systematic review. Cancer Nurs 2023;46(5):E320-E327. [doi: 10.1097/NCC.0000000000001130] [Medline: 37607382]
27. Chen YH, Liang WA, Hsu CY, et al. Functional outcomes and quality of life after a 6-month early intervention program for oral cancer survivors: a single-arm clinical trial. PeerJ 2018;6:e4419. [doi: 10.7717/peerj.4419] [Medline: 29492348]

Abbreviations

IoT: Internet of Things
SUPPH: Strategies Used by People to Promote Health
UW-QOL: University of Washington Quality of Life Scale

Edited by L Buis; submitted 04.05.24; peer-reviewed by A Abbasi, D Bracken-Clarke, L Weinert; revised version received 07.02.25; accepted 15.04.25; published 25.06.25.

Please cite as:
Li Y, Wang Y, Wu Y, Yu H, Yao H, Wang Y, Yin Y, Wang L, Hou L
Impact of mHealth on Postoperative Quality of Life, Self-Management, and Dysfunction in Patients With Oral and Maxillofacial Tumors: Nonrandomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e59926
URL: https://mhealth.jmir.org/2025/1/e59926
doi:10.2196/59926

©Yufei Li, Yueping Wang, Yifan Wu, Hong Yu, Hua Yao, Yuqun Wang, Yulan Yin, Lan Wang, Lili Hou. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Outcomes of an App-Based Intervention to Target Naming Among Individuals With Poststroke Aphasia: Virtual Randomized Controlled Trial

Esther S Kim[1], PhD; Laura Laird[2,3], BA; Carlee Wilson[1], PhD; Steven Stewart[3], PhD; Philip Mildner[4], PhD; Sebastian Möller[5], PhD; Raimund Schatz[6], PhD; Robert P Spang[5], PhD; Jan-Niklas Voigt-Antons[5,7], PhD; Elizabeth Rochon[2,3], PhD

[1]Department of Communication Sciences and Disorders, University of Alberta, 2-70 Corbett Hall, Edmonton, AB, Canada

[2]Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada

[3]KITE Research Institute, Toronto Rehabilitation - University Health Network, University of Toronto, Toronto, ON, Canada

[4]Nuromedia GmbH, Cologne, Germany

[5]Quality and Usability Lab, Technische Universitat Berlin, Berlin, Germany

[6]Center for Technology Experience, AIT Austrian Institute of Technology, Vienna, Austria

[7]Immersive Reality Lab, Hamm-Lippstadt University of Applied Sciences, Hamm, Germany

Corresponding Author:
Esther S Kim, PhD
Department of Communication Sciences and Disorders, University of Alberta, 2-70 Corbett Hall, Edmonton, AB, Canada

Abstract

Background: People with aphasia present with language and communication deficits, most notably in lexical retrieval (naming). Although positive outcomes in naming have been observed following speech-language treatment, many individuals with aphasia continue to face impairments after the acute phase of rehabilitation. Mobile app–based therapies are increasingly being used by speech-language pathologists in the rehabilitation of people with aphasia as an adjunct to or in lieu of traditional in-person therapy approaches. These apps can increase the intensity of treatment and have been shown to result in meaningful outcomes across several domains.

Objective: VoiceAdapt is a mobile therapy app addressing naming impairments, designed within a user-centered design framework. The VoiceAdapt app uses two evidence-based lexical retrieval treatments to engage people with aphasia to improve their naming abilities through interaction with the app. The purpose of this study was to conduct a randomized controlled trial to examine the preliminary clinical efficacy of training with VoiceAdapt on the language and communication outcomes of people with aphasia.

Methods: A two-arm, waitlist-controlled, crossover group randomized controlled trial was conducted at two sites within Canada. During the intervention phase, participants completed 5 weeks of independent training with the app, which involved naming practice using Semantic Features Analysis and Phonological Components Analysis. The primary outcome measure was naming performance (Boston Naming Test); secondary outcomes included measures of overall language and naming (Western Aphasia Battery-Revised), communication (Communication Effectiveness Index), and quality of life (Stroke and Aphasia Quality of Life Scale-39).

Results: A total of 37 people with aphasia in the chronic stages (average 4.6 y postonset of aphasia) participated in this study. Participants used the app for an average of 20 hours over the 5-week intervention phase. Training with VoiceAdapt resulted in an increase of 1.6 points on the Boston Naming Test (Cohen *d*=0.3). Evidence for improved naming was also observed on trained items, as well as subtests of naming or word-finding on the WAB-R. Training with the app also resulted in a significant increase in participants' perceptions of their communication quality of life (increase of 0.1 points; Cohen *d*=0.3), but no other measures (WAB-R Aphasia Quotient, Communicative Effectiveness Index) were significant.

Conclusions: Individuals with aphasia who used the VoiceAdapt app for 5 weeks to target naming skills demonstrated measurable gains in naming and communication-based quality of life. Notably, these changes were observed in a remotely delivered program, in participants who were in the chronic stages of aphasia. These findings inform the profession on the use of app-based home therapy programs as an accessible, cost-effective option for individuals in the chronic stages of recovery who often have limited options for rehabilitation.

Trial Registration: ClinicalTrials.gov NCT04108364; https://clinicaltrials.gov/ct2/show/NCT04108364

International Registered Report Identifier (IRRID): RR2-10.2196/30621

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e67711) doi:10.2196/67711

KEYWORDS

aphasia; rehabilitation; speech-language pathology; app-based therapy; user-centered design; mHealth; adaptive software; technology-based intervention; speech; language; communication; stroke; randomized controlled trial; mobile app; efficacy; quality of life

Introduction

Background

Aphasia, an acquired language disorder most commonly caused by stroke, impacts linguistic expression and comprehension in both written and oral modalities. People with aphasia may have impairments in their ability to speak, understand, read, and write. Thus, aphasia can severely limit activities and restrict participation in everyday activities [1]. Nearly one-third of individuals who experience a stroke have aphasia [2,3], and its presence has been associated with poorer outcomes in both acute and chronic periods following stroke [4,5].

Speech-language pathologists (SLPs) provide rehabilitation services for people with aphasia, often focusing on remediation of the language impairment or training of compensatory strategies. Such interventions have been shown to improve outcomes for people with aphasia [6,7], but constraints on health care resources often prevent people with aphasia from receiving these services, especially in the chronic stages. Recently, the proliferation of mobile app–based therapies available to deliver speech-language interventions has meant that people with aphasia can receive prolonged rehabilitation in an efficient, cost-effective manner. SLPs have increasingly been using such apps as adjuvant therapy approaches, or even in lieu of traditional (face-to-face) therapy, with positive outcomes observed across several domains [8,9]. One benefit of these apps is that people with aphasia can increase the intensity of their rehabilitation, a factor that is associated with greater long-term recovery [7,10-12]. In addition, app-based interventions can be self-delivered by people with aphasia and still yield positive language and communication outcomes [13-15]. Moreover, app-based interventions allow for therapy content to be adapted according to the user or patient state, due to the ability to measure multiple parameters of user interactions with touch-based devices.

The research evidence investigating the feasibility, acceptability, and clinical utility of tablet-based speech-language therapy apps for aphasia has increased in recent years, with inconclusive results. In part, this may relate to the chronicity of individuals with aphasia included in these studies. For instance, Ruiz et al [16] found no difference between an intervention group provided with tablet-based therapy in the first week poststroke and a no-treatment control group. However, Mallet et al [17,18] reported that the use of mobile therapy apps for speech-language therapy was feasible in the very early stages of acute stroke recovery and could potentially be used to bridge the gap between discharge from acute care and the start of outpatient therapy. On the other hand, evidence is emerging on the utility of app-based therapy for adults in the chronic stages of acquired aphasia. Across a growing number of studies, individuals with aphasia have shown improvement across a variety of domains, including word production, reading, naming, spoken discourse, quality of life, and overall language [13-15,19-24].

App-based therapy has the potential to increase treatment dosage (total hours), intensity (hours per week), frequency (days per week), and duration (total weeks)—important factors that contribute to positive speech-language therapy outcomes [25]. For instance, participants in a study by Alam et al [22] were found to benefit more from therapy in the condition where an app was provided as an adjunct to therapy (the “study” arm) than in the condition where there was no adjunct (the “control” arm). However, it should be noted that participants in the study arm received more therapy overall due to the addition of the app. Individuals with aphasia who self-managed their own participation using therapy apps in the absence of face-to-face therapy have also demonstrated improved language outcomes [13-15,21]. Indeed, app-based therapy in lieu of traditional therapy, particularly for individuals in the chronic stages of aphasia, could be an important component of optimizing recovery in the context of limited rehabilitation resources. However, further research to understand the impact of apps (alone and in combination with face-to-face rehabilitation) is warranted.

A recent review by Szeto et al [26] on the effect of mobile apps for stroke rehabilitation reported that aphasia therapy apps that resulted in the greatest improvements were ones that incorporated characteristics of face-to-face therapy. These included massed practice (repetition) of goal-oriented, task-specific exercises, user interactivity, and multisensory stimulation. A similar review identifying and evaluating mobile apps for speech therapy in adults with communication disorders (not limited to aphasia) revealed that there was a general lack of engaging and interactive elements in the apps [27]. Increasing the usability of therapy apps could be enhanced by designing apps with input from relevant stakeholders, following the principles of user-centered design (UCD). UCD is a systematic approach to meeting user experience goals and usability through integrating the needs and abilities of end users during the design process [28]. Although the number of therapy apps designed for aphasia is steadily increasing [27,29], very few have been designed following the principles of UCD [30,31], despite increasing evidence supporting the beneficial impacts of UCD on user experience in digital health technologies [32]. Indeed, in their review of 70 therapy apps for adults with communication disorders, Vaezipour et al [27] recommend using a co-design process involving end users (ie, people with aphasia), subject matter experts (ie, SLPs), and incorporating interactive elements and gamification to tailor the app to individual needs.

Development of the VoiceAdapt App

The VoiceAdapt app, developed by an international team of interdisciplinary researchers from the fields of computer science,

XSL•FO
RenderX

usability design, and speech-language pathology, is a tablet-based app focused on improving naming abilities in individuals with aphasia. VoiceAdapt instantiates two evidence-based lexical retrieval protocols, Phonological Components Analysis (PCA) [33,34] and Semantic Feature Analysis (SFA) [35,36], into an adaptive, voice-responsive therapy app that people with aphasia can use to improve their naming skills.

Using UCD principles, the team engaged 75 end users (including people with aphasia, clinicians, and caregivers) in designing the app. Over the course of three iteration cycles, end users participated in interviews to elicit user requirements, provided feedback on mock-ups and prototypes, and evaluated the designs with a focus on user experience of multimodal speech and adaptivity [32]. The end result was a mobile speech-language training app for iOS and Android that supported regular training sessions, adaptivity, personalization, and addressed multimodal language goals (reading, understanding, speech production).

Objective

The purpose of this study is to report the efficacy of a beta version of the VoiceAdapt app, within a randomized controlled trial (RCT; ClinicalTrials.gov NCT04108364). Specifically, we examined the impact of training with the VoiceAdapt app on language and communication outcomes in individuals with chronic aphasia. Primary outcomes were naming ability, and secondary outcomes included measures of overall language, communication effectiveness, and quality of life.

Methods

Recruitment

Participants were recruited offline between August 2020 and August 2022 in two geographic regions in Canada through the Language Sciences Lab (Department of Speech-Language Pathology, University of Toronto) and Aphasia Research Lab (Department of Communication Sciences and Disorders, University of Alberta), based in Toronto, Ontario, and Edmonton, Alberta, respectively. Information about the study was disseminated to clinicians and programs serving individuals with aphasia. Interested participants reached out directly to study personnel (CW, Edmonton; LL, Toronto) and provided informed consent prior to being enrolled in the study. Communicatively accessible consent forms were used, and all study personnel were trained in supported communication techniques.

Participants enrolled in the study were at least 6 months postonset of a stroke in the left hemisphere, had aphasia of mild to moderate severity (Aphasia Quotient ≥30 on Western Aphasia Battery-Revised [WAB-R] [37]), scored below 75% correct on the Boston-Naming Test [38] at initial assessment, had a prominent verbal expression impairment (based on assessment scores and SLP determination), passed screenings of basic sensory (vision and hearing) and cognitive functioning, and spoke English as a primary language. Participants with severe aphasia were excluded from the study due to significant verbal expression impairment and limited naming abilities in this population. As the app was designed for self-directed treatment, individuals with severe aphasia profiles were unlikely to benefit from this type of intervention. Participants also had access to a computer or tablet and the internet in order to participate in web-based assessment sessions. In addition to meeting all inclusion criteria, participants were excluded from participation if they were currently engaged in individual (one-on-one) speech-language therapy or were currently using speech-language therapy apps. All study procedures (intake, informed consent, and assessments) took place using videoconferencing software, in line with recent advancements indicating the feasibility and reliability of virtual assessment protocols [39,40].

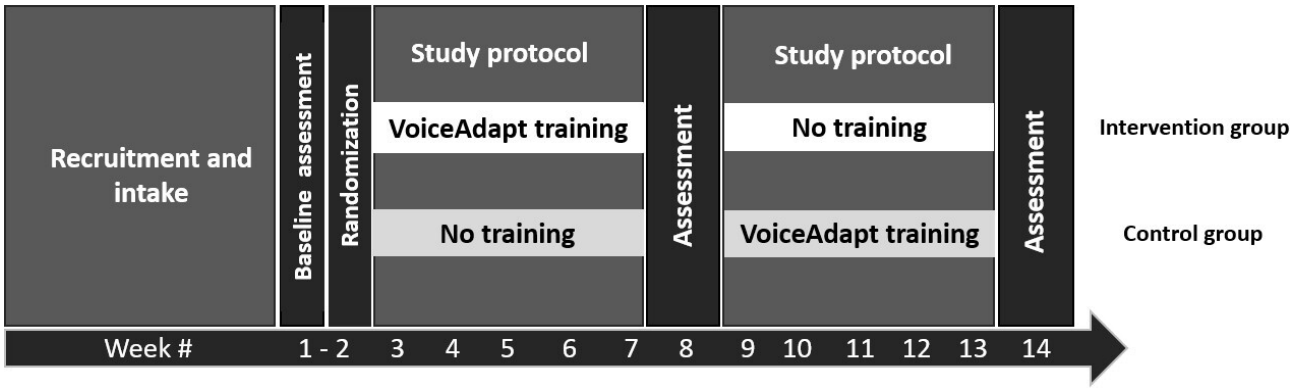
Planned enrollment for the study was 80 participants based on sample size calculations, with interim analyses planned when quarterly enrollment targets were met (ie, 20, 40, and 60 participants) [39]. The sample size calculation was based on medium effect sizes estimated in this study according to the results shown by Varley et al [41]. Group sample sizes of 72 are needed to achieve 80% power to detect a difference in proportions of 0.25 between the null hypothesis and the alternative hypothesis using a 1-tailed *t* test with a probable α error of 0.05. Accounting for a dropout rate of approximately 10% (based on the dropout rate of Varley et al [41]) from baseline to follow-up, we estimated we would need 79 participants (rounded up to 80). It was determined that the trial would be stopped early if evidence of a treatment effect on the primary outcome measure was observed at the time of interim analyses.

Design

The RCT was a two-arm, randomized, waitlist-controlled, crossover group design. All participants completed initial assessments and then were randomized into intervention or treatment-deferred control groups. Participants in the intervention group completed training with the VoiceAdapt for 5 weeks during Phase 1 of the study, while those in the control group continued with usual care. Following midpoint assessments for all participants, those in the treatment-deferred control group crossed over into receiving treatment, while those in the intervention group crossed over into the control phase. Final assessments took place following Phase 3, approximately 13 weeks following study enrollment (Figure 1). Although the study was originally intended to be conducted in person, the COVID-19 pandemic necessitated a move to all procedures being conducted virtually.

XSL•FO
RenderX

Figure 1. Study design.



An individual not affiliated with the study created a computer-generated randomization sequence using Excel (version 2007; Microsoft Corp), stratified by site with 1:1 allocation using random block sizes of 10. Each cell containing the allocation status was concealed using black fill in the spreadsheet. One member of the study team (CW) revealed allocation status to participants (Edmonton site) or to the research coordinator at the Toronto site (LL), who then related the status to participants following baseline assessment. Assessments were completed by SLPs (Edmonton site) or speech-language pathology graduate students under the supervision of a SLP (Toronto site) who were blinded to the treatment condition of participants. Participants were told not to reveal to assessors which group they were in.

Intervention

The VoiceAdapt app intervention comprises naming exercises based on SFA and PCA, two evidence-based anomia treatments, delivered on a mobile tablet. A color picture is presented along with a series of prompts designed to engage the people with aphasia in providing semantic features (eg, “What is it used for?” [Action]; “What does it look like?” [Properties]), or phonological components (eg, “What sound does it start with?” [First sound]; “What is a rhyming word?” [Rhyme]). People with aphasia randomized to the VoiceAdapt app treatment group were instructed to practice using the app for 1 hour per day every weekday (5 days per week) for 5 weeks for a total of 25 hours of intervention. Participants were provided with tablets with the app loaded, or could download the app from a link provided by the researchers onto their own devices. Participants selected interests at initial app setup, which then tailored the practice items to each individual. Research staff provided orientation to the app and monitored participants’ app use (number of minutes per week) through the app portal. Research staff (CW and LL) checked in with participants through email, or phone, or video calls on a weekly basis, providing practice reminders and answering questions.

Outcome Measures

The primary outcome measure was the Boston Naming Test (BNT) [38] to measure naming, secondary outcome measures were WAB-R [37] and Stroke and Aphasia Quality of Life Scale (SAQOL-39g) [42] to measure quality of life, and Communication Effectiveness Index (CETI) [43] to measure the person with aphasia’s perceptions of communication effectiveness. Both the person with aphasia and their caregiver (where applicable) completed the CETI. These outcomes were assessed at baseline and after intervention phases (Figure 1). In addition, at the final assessment session, the Study Stimulus Naming Test (SSNT), a measure of performance on trained and untrained items on the VoiceAdapt app, the System Usability Scale (SUS) [44], and a 5-question posttreatment questionnaire specific to app use were administered to all participants.

Fidelity and Reliability

To ensure fidelity of procedures across both sites, a standardized intake and assessment training protocol was developed. Assessors at both sites completed all steps in the training protocol including reading background papers [39,40], reviewing intake and consent forms and manuals for all assessments, observing the study coordinator or primary investigators (licensed SLPs) doing a complete assessment session, and finally, being observed by the study coordinators or primary investigators to ensure protocol adherence prior to completing assessments on their own.

Administration of the primary outcome measure (BNT) was video recorded through the Zoom (Zoom Video Communications, Inc) videoconferencing platform. A total of 16 assessment sessions (just under 15% of the sample) were randomly selected for assessing the reliability of scoring. Point-to-point reliability was high overall (Cohen κ=0.91).

Data Management

Study data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) electronic data capture tools hosted at the University of Alberta [45,46]. REDCap is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for data integration and interoperability with external sources.

Ethical Considerations

This project was approved by the University of Toronto Health Sciences Research Ethics Board (Protocol Number 37980) and the University of Alberta Health Research Ethics Board (Pro00091587). All participants provided informed consent following an explanation of research procedures in a communicatively accessible manner by research staff who were

XSL•FO
RenderX

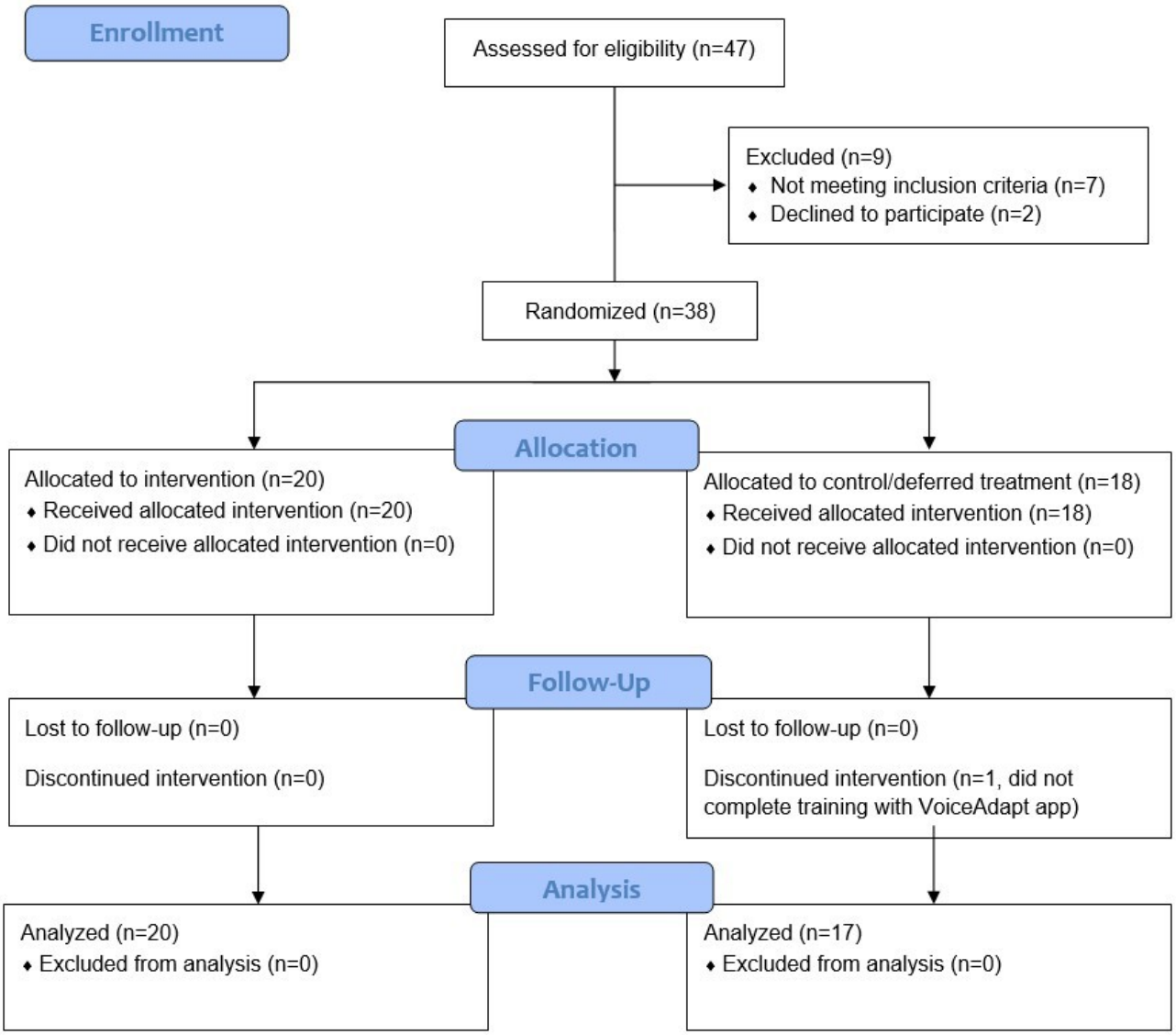
trained in supported communication techniques. All data were anonymized and deidentified. Participants were not provided with monetary compensation for their participation in this study.

Results

Participants

A total of 38 participants with aphasia (13 female) were recruited and randomized across the two sites. In total, 20 individuals were randomly assigned to receive training with VoiceAdapt first, while 18 individuals were randomly assigned to the deferred treatment (control) condition. One individual randomized to the control group dropped out of the study shortly after starting treatment due to frustrations with the automatic speech recognition (ASR) aspects of the app. As the withdrawn participant only contributed two baseline assessments and did not contribute any treatment information, their data were not included in analyses. A total of 37 participants completed both phases of the study (Figure 2). The trial ended early due to results from interim analyses, as well as slower-than-expected recruitment due to challenges following the COVID-19 pandemic (see details below). Participants had a mean age of 62.2 (SD 13.9) years and 15.1 (SD 3.2) years of education and were an average of 4.6 (SD 4.2) years postonset of aphasia. No differences in age, education, years postonset, or baseline WAB-AQ/BNT were observed between groups (Table 1). Site differences were present for age and education across sites, with participants recruited through the University of Alberta being younger and having less education than those recruited through the University of Toronto. The number of participants across allocation conditions and mean demographic information are presented in Table 1.

Figure 2. Participant flowchart.



XSL•FO
RenderX

Table . Participant mean age, education, years post-onset, WAB-AQ[a], and BNT[b] scores across allocations[c].

Characteristic	Intervention (n=20; 6F), mean (SD)	Control (n=17; 7F), mean (SD)	Total (n=37; 13F), mean (SD)
Age (years)	59.9 (15.3)	64.8 (11.9)	62.2 (13.9)
Education	14.7 (2.6)	15.5 (3.7)	15.1 (3.2)
Years postonset	3.5 (2.5)	5.8 (5.3)	4.6 (4.2)
WAB-AQ[a]	68.3 (15.2)	73.6 (14.0)	70.7 (14.7)
BNT[b]	25.0 (12.1)	23.5 (12.5)	24.3 (12.1)

[a]WAB-AQ: Western Aphasia Battery-Revised Aphasia Quotient.

[b]BNT: Boston Naming Test.

[c]Intervention refers to participants randomized to receive VoiceAdapt training first; control refers to participants randomized to the treatment-deferred control group.

Analytic Strategy

The 2×2 crossover trial with one baseline was analyzed using Bayesian inference to estimate the parameters of a mixed-effects regression model. Using mixed-effects regression can account for hierarchical sampling of participants within sites. Using this crossover design allowed estimation of treatment effect, time effect, sequence effect, and carryover effect. Bayesian inference was chosen for accurate multiple comparisons [47]. For each outcome, we modeled treatment effect, time effect, and the outcome at baseline covariate as a normal distribution with unknown variance using a noninformative flat prior for the mean parameter and a noninformative inverse-gamma prior for the variance parameter that was sampled independently of the other parameters. This Bayesian model was fit using the adaptive Metropolis-Hastings Markov chain Monte Carlo simulation method implemented in Stata (version 16; StataCorp LLC). The proportion of the posterior distribution of the treatment effect that was greater than zero was interpreted as the probability that the outcome increased. For such interval hypothesis testing, a probability of 0.95 or higher was considered strong evidence of improvement during VoiceAdapt training. In these data, we found no statistical evidence for a sequence effect or a carryover effect, so these were dropped from our model. We also checked for an effect of app use and found no statistical evidence for an effect, so it was also dropped from the model. All analyses were carried out on the 37 participants who completed both phases of the study unless otherwise indicated.

Our initial sample size calculation was based on using frequentist approaches. There were several realities that necessitated a change to Bayesian analyses. Due to recruitment challenges in the virtual context, we needed to do interim analyses more frequently than outlined in our protocol. Bayesian analysis allows researchers to conduct interim analyses without being penalized because decisions at analysis points are based on the posterior probabilities of treatment effects [47]. Bayesian analysis also provides a straightforward means of analyzing multiple outcomes multiple times. While our power analysis was based on the effect size of one strong primary outcome, BNT, in reality, our different outcomes had differential effects based on our interim analyses. Given that we found significant results with our reduced sample size, and the recruitment challenges referenced above, we decided to end the trial early.

Primary Outcome

The primary outcome was the BNT to measure change in naming performance as a result of training with the VoiceAdapt app. Participants increased an average of 1.6 points on the BNT (equal-tailed 95% credible interval −0.3 to 3.6; Cohen *d*=0.3) when training with VoiceAdapt (Table 2).

Table . Median and equal-tailed 95% credible interval of the posterior probability distribution of each parameter of the cross-over model for the primary outcome (BNT[a]).

Parameter	BNT, median (95% credible interval)
Fixed-effect parameters[b]	
VoiceAdapt effect	1.6 (−0.3 to 3.6)[c]
Time effect	1.5 (−0.4 to 3.3)
Outcome at baseline covariate	1.0 (0.8 to 1.1)
Constant	2 (−3 to 6)
Random-effect parameters	
Among participant variance	21 (8 to 39)
Residual variance	17 (10 to 28)

[a]BNT: Boston Naming Test.

[b]Proportion of Bayesian posterior probability distribution of VoiceAdapt effect greater than zero, Cohen *d:* 0.954.

[c]Cohen *d*=0.3.

XSL•FO
RenderX

Other Outcomes: Naming Ability

Additional analyses with other measures related specifically to naming abilities were examined further using the same analysis procedures. These included the Object Naming subtest on the WAB-R (20 item naming test); the Naming and Word Finding subscore of the WAB-R (comprising combined scores from Object Naming, Word Fluency, Sentence Completion, and Responsive Speech subsets); and the SSNT, a 322 item naming measure comprising both trained and untrained words programmed into the VoiceAdapt app. As seen in Table 3, training with VoiceAdapt resulted in an increase of 3 points on the Object Naming subtest (equal-tailed 95% credible interval 1-5) and an increase of 2.6 points on the Naming and Word Finding subtest (equal-tailed 95% credible interval 0.15-5), representing small effects (Cohen *d*=0.4 and 0.3, respectively). Interestingly, the SSNT demonstrated the greatest change as a result of training with VoiceAdapt, with a median change score of 13 (equal-tailed 95% credible interval 7-19; Cohen *d*=0.9).

Table . Median and equal-tailed 95% credible interval of the posterior probability distribution of each parameter of the cross-over model for additional outcomes (SSNT[a], WAB-R[b] Object Naming, WAB-R Naming, and Word Finding total).

Parameter	SSNT	Object Naming (WAB-R)	Naming and Word Finding (WAB-R)[b]
Fixed-effect parameters			
VoiceAdapt effect, median (95% credible interval)	13 (7 to 19)[c]	3 (1 to 5)[d]	2.6 (0.15 to 5)[e]
Proportion of Bayesian posterior probability distribution of VoiceAdapt effect greater than zero, Cohen *d*	0.999	0.998	0.980
Time effect, median (95% credible interval)	13 (6 to 19)	−0.6 (−2.6 to 1.4)	−0.6 (−3.0 to 1.9)
Outcome at baseline covariate, median (95% credible interval)	1.05 (0.9 to 1.2)	1.0 (0.8 to 1.1)	1.0 (0.9 to 1.1)
Constant, median (95% credible interval)	1 (20 to 18)	1 (−5 to 7)	1 (9 to 6)
Random-effect parameters, median (95% credible interval)			
Among participant variance	375 (188 to 668)	23 (10 to 43)	31 (11 to 60)
Residual variance	168 (103 to 277)	19 (11 to 30)	28 (17 to 45)

[a]SSNT: Study Stimulus Naming Test.

[b]WAB-R: Western Aphasia Battery-Revised.

[c]Cohen *d*=0.9.

[d]Cohen *d*=0.4.

[e]Cohen *d*=0.3.

Secondary Outcomes

There was no significant effect of VoiceAdapt training on overall language abilities, as evidenced by a mean change in WAB-R Aphasia Quotient of 0.4 (Cohen *d*=0.01). VoiceAdapt training resulted in a significant increase in participants' perceptions on the SAQOL-39g Communication Quality of Life subscale (increase of 0.1 points; Cohen *d*=0.3), representing a small effect, but not on overall Quality of Life (SAQOL-39g mean score decrease of 0.01 points; Cohen *d*=0.1). Measures of communication effectiveness as measured by the participant (CETI-participant mean score=−5; Cohen *d*=0.1) or their partners (CETI-partner mean score=0.1; Cohen *d*=0.2) were not found to be significantly improved by VoiceAdapt training (Table 4).

XSL•FO
RenderX

Table . Median and equal-tailed 95% credible interval of the posterior probability distribution of each parameter of the cross-over model for the secondary outcomes (WAB-R[a] aphasia quotient, SAQOL-39g[b] mean score, SAQOL-39g Communication score, CETI[c], and CETI-Partner)[d].

Parameter	WAB-R AQ	SAQOL-39g mean	SAQOL-39Comm	CETI	CETI-Partner
Fixed-effect parameters					
VoiceAdapt effect, median (95% credible interval)	0.4 (−1.1 to 1.8)[e]	−0.005 (−0.1 to 0.1)[f]	0.1 (−0.1 to 0.3)[g]	−5 (−12 to 4)[f]	0.1 (−7 to 7)[h]
Proportion of Bayesian posterior probability distribution of VoiceAdapt effect greater than zero, Cohen *d*	0.707	0.468	0.846	0.142	0.506
Time effect, median (95% credible interval)	0.4 (−1.1 to 1.8)	0.01 (−0.1 to 0.1)	−0.05 (−0.3 to 0.1)	1 (−7 to 10)	9 (2 to 16)
Outcome at baseline covariate, median (95% credible interval)	1.0 (0.9 to 1.1)	0.9 (0.7 to 1.1)	0.8 (0.5 to 1.0)	0.6 (0.4 to 0.9)	0.8 (0.5 to 1)
Constant, median (95% credible interval)	−1 (−7 to 4)	0.4 (−0.4 to 1.2)	0.9 (0.1 to 1.6)	44 (22 to 66)	18 (−3 to 38)
Random-effect parameters, median (95% credible interval)					
Among participant variance	7 (1 to 15)	0.1 (0.05 to 0.2)	0.2 (0.1 to 0.3)	231 (37 to 493)	137 (12 to 311)
Residual variance	10 (6 to 17)	0.06 (0.03 to 0.1)	0.2 (0.1 to 0.3)	309 (175 to 526)	151 (76 to 289)

[a]WAB-R: Western Aphasia Battery-Revised.

[b]SAQOL-39g: Stroke and Aphasia Quality of Life Scale.

[c]CETI: Communication Effectiveness Index.

[d]Sample includes three longitudinal assessments of 37 participants, except for CETI Partner outcome because 7 participants did not have a participating partner.

[e]Cohen *d*=0.01.

[f]Cohen *d*=0.1.

[g]Cohen *d*=0.3.

[h]Cohen *d*=0.2.

Participant Perceptions of VoiceAdapt App

Participants used the app for an average of 19.5 (SD 8.6) hours over the course of the 5-week VoiceAdapt intervention phase. The SUS [44], a 10-item questionnaire, was used to evaluate participant perceptions of the ease of using the VoiceAdapt app. Each question has one of five response options, ranging from strongly disagree to strongly agree; items are scored (accounting for positive or negative valence) and multiplied by 2.5 to yield an overall usability score out of 100. The average SUS score for all participants was 61.2 (SD 18.9). The mean score for each question ranged from 2.2 to 2.6, indicating participants felt neutral or agreed with statements pertaining to the usability of the app. In addition to the SUS, a 5-question posttreatment questionnaire was administered to all participants, asking about their experiences with the app. Mean responses to the posttreatment questionnaire are provided in Table 5.

Table . Average responses to posttreatment questionnaires by study participants.

Question	Average response (1=strongly disagree; 5=strongly agree)
I really liked the app	4.1
The app was easy to use	3.9
Training with the app helped my communication	3.9
I would use this app if it were still available	3.8
I would be willing to pay a fee for the VoiceAdapt app	3.2

XSL•FO
RenderX

Discussion

Principal Findings

This RCT was designed to assess the efficacy of the VoiceAdapt app, a tablet-based naming therapy program for people with aphasia. The app implemented evidence-based treatment principles into an adaptive, personalizable app that addressed multimodal language stimulation with the aim of stimulating language production, specifically single-word naming.

Training with the VoiceAdapt app resulted in improved single-word naming performance. The change in naming after treatment on the primary outcome measure (ie, BNT) was minimal, though statistically robust. Converging evidence for improved naming also came from observed improvements on subtests of the WAB-R related specifically to naming abilities (Naming and Word Finding, Object Naming), as well as the SSNT, a set of 322 items (including trained and untrained words). Notably, improvements on the SSNT were accompanied by a large effect size. These results extend the literature base on the efficacy of SFA and PCA for improving naming [48,49], as well as the growing number of studies reporting positive outcomes from app-based instantiations of these therapy protocols [15,18,23].

Importantly, participants' perception of their quality of life as it related specifically to their communication abilities (SAQOL-39g Communication score) also revealed a small but significant improvement after VoiceAdapt training, indicating that over and above changes in actual numbers of words produced after training, participants felt more positive about their communicative abilities (eg, speaking in general, finding the right word, and being understood by others). In contrast to the participants' perceived communication quality of life, communicative effectiveness as measured by the CETI was not perceived to have improved after VoiceAdapt treatment by either the participants or their family members. The concept of communicative effectiveness extends beyond performance on linguistic tasks to situational interactions in the real world [50]. Implementing exercises or activities in future apps that provide the opportunity to improve participation (eg, in conversations) and social integration, perhaps by incorporating digital networking for peer support [51] or virtual reality [52], may in the future lead to generalization to domains of communication effectiveness, overall functioning, and well-being.

Despite improvements in naming and communication quality of life, secondary outcomes encompassing overall language and communication or quality of life did not change significantly. In part, this may be related to the broad domains they were measuring. For instance, the WAB-R AQ calculation includes language comprehension, repetition, and spoken discourse in addition to naming. Similarly, the mean SAQOL-39g score includes evaluation of activities of daily living (eg, preparing food and getting dressed), physical activities (eg, walking and standing), and emotional or mental health items, in addition to communication items. Therefore, despite current recommendations for reporting outcomes for aphasia treatment research [53], it may have been unrealistic to expect change on these two comprehensive measures when only one domain, naming, was implemented in the training app. Incorporating different and varied exercises in the future that target additional language domains may lead to generalization to other language domains.

Recent reviews of the effectiveness of speech-language therapy have reported that intensity (dosage, frequency, and duration) of treatment is associated with the largest gains [25], particularly for individuals in the chronic phases of aphasia. However, in this study, app use was not found to be associated with improvement. In addition, although participants were tasked with practicing for 1 hour daily, average app use was less than that. Nonetheless, our participants, who were an average of 4.6 years (SD 4.2) postonset of aphasia, self-managed their therapy with the app with periodic check-ins from the intervention team, supporting prior reports [15,21,26] that tablet-based therapy apps can be an effective option for people with aphasia in the chronic stages to continue to work on their language [54].

One distinctive feature of the VoiceAdapt app was in its design, incorporating UCD principles with input from end users (therapists, people with aphasia, etc), consistent with findings demonstrating that UCD can significantly enhance user engagement and intervention outcomes in technology-based health apps [27,55]. The app included several features recommended in Vaezipour et al's [27] review of mobile apps. Specifically, the app was tailored to participants' interests in that they were able to select categories of items they wished to practice. Participants also interacted with the app to select multimodal cues as needed, providing individualized support according to their performance. Despite this, participants' ratings of app favorability were in the midrange (mean SUS score 61.2, SD 18.9). Other studies investigating therapy apps have reported higher scores [56,57], but there may be other factors contributing to these higher scores (eg, more clinician involvement and different app content). Nonetheless, improvements could be made to further enhance the effectiveness and usability of the app. One limitation of the app was that the ASR software built into the app was not always accurate, particularly given the common co-occurrence of accompanying motor speech impairments with poststroke aphasia. Indeed, frustrations with ASR contributed to the one participant's withdrawal we had in this study. As this is a rapidly evolving area of study and development, future iterations of the app may perform better in this regard. Incorporation of additional features demonstrated to improve engagement and interactivity [26,27], such as social interaction and gamification, could also result in an improved app in the future. Importantly, future additions and designs should incorporate the wishes, needs, and recommendations of people with aphasia [31].

Limitations and Future Directions

The planned methods, including recruitment activities, assessments, and treatment, were all modified to be delivered virtually due to the COVID-19 pandemic. This may have affected recruitment, as our sample size was less than expected. However, positive results in naming were still obtained. Future research will have to determine whether greater and more extensive improvements might be obtained with a larger sample size. In addition, people with aphasia, on average, did not

XSL•FO
RenderX

practice for the full dosage prescribed. This, too, leaves a question for future research of whether greater gains might be obtained with more practice.

The study design did not include a maintenance assessment for all participants postintervention (although individuals in the treatment condition had one maintenance assessment at approximately 6 wk postintervention). This is recognized as a limitation, as long-term maintenance of treatment gains is an important outcome in aphasia rehabilitation. Finally, the nature of this study did not allow for participants to be blinded to the condition. However, all individuals completing assessments at each time point were blinded to participant condition. Future studies could examine the efficacy of training using the VoiceAdapt app compared to other speech-language therapy apps on improving naming performance in individuals with aphasia.

Conclusions

In this RCT, 5 weeks of training in naming single words with the VoiceAdapt app resulted in measurable gains in naming and communication-based quality of life, in contrast to no training in a group of participants in the chronic stages of poststroke aphasia. These changes were seen in a remotely delivered program, in participants who were no longer receiving speech therapy and who practiced their exercises at home using the app.

The findings inform the profession on telerehabilitation practices and monitoring app–based home therapy programs, with a strong recommendation to incorporate UCD methodologies for improved user adherence and therapeutic efficacy [32,55]. In addition, the improvements seen in naming and communicative functioning suggest that instantiating evidence-based treatments such as the one done with this app and which can be found in some others (eg, Tactus and Constant Therapy) can be recommended as an efficacious, easily accessible, and cost-effective method for individuals who are in the chronic stages of recovery and who often have limited options for rehabilitation.

Acknowledgments

The authors would like to acknowledge Anna-Maria Bieber, Stephanie Schwarz, and Till Bieg for their support of the user-centered design of the app, Jevgenijs Danilins for his contributions to app development, and March of Dimes Canada for their support of participant recruitment. This research was funded under the JTC 2017 launched by the Joint Programming Initiative “More Years Better Lives” (JPI MYBL), which was supported by Horizon2020, the EU Framework for Research and Innovation (grant 643850). The VoiceAdapt Consortium was funded by the Canadian Institutes of Health Research grant #155164; the Austrian Research Promotion Agency (FFG), project 861960, and the German Federal Ministry of Education and Research (grant 16SV7995K).

Authors' Contributions

ESK and ER designed the study with the assistance of LL, JNV-A, SM, RS, and RPS. All authors designed the VoiceAdapt app. PM developed the VoiceAdapt app and backend. ESK, ER, LL, and CW recruited participants and managed the randomized controlled trial. SS completed statistical analyses in consultation with ESK and ER. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Checklist 1
CONSORT eHEALTH (V 1.6.1) checklist.
[PDF File, 10772 KB - mhealth_v13i1e67711_app1.pdf]

References

1. Hilari K. The impact of stroke: are people with aphasia different to those without? Disabil Rehabil 2011;33(3):211-218. [doi: 10.3109/09638288.2010.508829] [Medline: 20712416]
2. Dickey L, Kagan A, Lindsay MP, Fang J, Rowland A, Black S. Incidence and profile of inpatient stroke-induced aphasia in Ontario, Canada. Arch Phys Med Rehabil 2010 Feb;91(2):196-202. [doi: 10.1016/j.apmr.2009.09.020] [Medline: 20159121]
3. Flowers HL, Silver FL, Fang J, Rochon E, Martino R. The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic stroke. J Commun Disord 2013;46(3):238-248. [doi: 10.1016/j.jcomdis.2013.04.001] [Medline: 23642855]
4. Lazar RM, Boehme AK. Aphasia as a predictor of stroke outcome. Curr Neurol Neurosci Rep 2017 Sep 19;17(11):83. [doi: 10.1007/s11910-017-0797-z] [Medline: 28929424]
5. Stransky ML, Jensen KM, Morris MA. Adults with communication disabilities experience poorer health and healthcare outcomes compared to persons without communication disabilities. J Gen Intern Med 2018 Dec;33(12):2147-2155. [doi: 10.1007/s11606-018-4625-1] [Medline: 30143977]

XSL•FO
RenderX

6. Breitenstein C, Grewe T, Flöel A, et al. Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. Lancet 2017 Apr 15;389(10078):1528-1538. [doi: 10.1016/S0140-6736(17)30067-3] [Medline: 28256356]
7. Ali M, VandenBerg K, Williams LJ, et al. Predictors of poststroke aphasia recovery. Stroke 2021. [doi: 10.1161/STROKEAHA.120.03116]
8. Des Roches CA, Kiran S. Technology-based rehabilitation to improve communication after acquired brain injury. Front Neurosci 2017;11:382. [doi: 10.3389/fnins.2017.00382] [Medline: 28804443]
9. Braley M, Pierce JS, Saxena S, et al. A virtual, randomized, control trial of a digital therapeutic for speech, language, and cognitive intervention in post-stroke persons with aphasia. Front Neurol 2021;12:626780. [doi: 10.3389/fneur.2021.626780] [Medline: 33643204]
10. Bhogal SK, Teasell R, Speechley M. Intensity of aphasia therapy, impact on recovery. Stroke 2003 Apr;34(4):987-993. [doi: 10.1161/01.STR.0000062343.64383.D0] [Medline: 12649521]
11. Brady MC, Kelly H, Godwin J, Enderby P, Campbell P. Speech and language therapy for aphasia following stroke. Cochrane Database Syst Rev 2016 Jun 1;2016(6):CD000425. [doi: 10.1002/14651858.CD000425.pub4] [Medline: 27245310]
12. Cherney LR, Patterson JP, Raymer A, Frymark T, Schooling T. Evidence-based systematic review: effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. J Speech Lang Hear Res 2008 Oct;51(5):1282-1299. [doi: 10.1044/1092-4388(2008/07-0206)] [Medline: 18812489]
13. Braley M, De Oliveira E, Munsell M, et al. A phase ii randomized, virtual, clinical trial of speech therapy app for speech, language, and cognitive intervention in stroke. Arch Phys Med Rehabil 2020 Nov;101(11):e62. [doi: 10.1016/j.apmr.2020.09.186]
14. Palmer R, Dimairo M, Cooper C, et al. Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): a multicentre, single-blinded, randomised controlled trial. Lancet Neurol 2019 Sep;18(9):821-833. [doi: 10.1016/S1474-4422(19)30192-9]
15. Stark BC, Warburton EA. Improved language in chronic aphasia after self-delivered iPad speech therapy. Neuropsychol Rehabil 2018 Jul;28(5):818-831. [doi: 10.1080/09602011.2016.1146150] [Medline: 26926872]
16. Ferrete Ruíz EJ, de Torres Chacón MDLR, Perez Sánchez S, et al. Breaking the communication barrier using new technologies in aphasic patients post-stroke. Revista Científica de la Sociedad de Enfermería Neurológica (English ed) 2023 Jul;58:14-20. [doi: 10.1016/j.sedeng.2022.06.002]
17. Mallet KH, Shamloul RM, Corbett D, et al. RecoverNow: feasibility of a mobile tablet-based rehabilitation intervention to treat post-stroke communication deficits in the acute care setting. PLoS One 2016;11(12):e0167950. [doi: 10.1371/journal.pone.0167950] [Medline: 28002479]
18. Mallet KH, Shamloul RM, Lecompte-Collin J, Winkel J, Donnelly B, Dowlatshahi D. Telerehab at home: mobile tablet technology for patients with poststroke communication deficits—a pilot feasibility randomized control trial. J Speech Lang Hear Res 2023 Feb 13;66(2):648-655. [doi: 10.1044/2022_JSLHR-21-00616] [Medline: 36634230]
19. Des Roches CA, Balachandran I, Ascenso EM, Tripodis Y, Kiran S. Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform. Front Hum Neurosci 2014;8:1015. [doi: 10.3389/fnhum.2014.01015] [Medline: 25601831]
20. Kurland J, Liu A, Stokes P. Effects of a tablet-based home practice program with telepractice on treatment outcomes in chronic aphasia. J Speech Lang Hear Res 2018 May 17;61(5):1140-1156. [doi: 10.1044/2018_JSLHR-L-17-0277] [Medline: 29710115]
21. Heide J, Netzebandt J, Ahrens S, Brüsch J, Saalfrank T, Schmitz-Antonischki D. Improving lexical retrieval with LingoTalk: an app-based, self-administered treatment for clients with aphasia. Front Commun 2023;8:1210193. [doi: 10.3389/fcomm.2023.1210193]
22. Alam N, Munjal S, Panda NK, Kumar R, Gupta S. Efficacy of Jellow app as an adjunct to stimulation therapy in improvement in language and quality of life in patients with chronic Broca's Aphasia. Disabil Rehabil Assist Technol 2023 Jul;18(5):596-602. [doi: 10.1080/17483107.2021.1892844] [Medline: 33689529]
23. Ballard KJ, Etter NM, Shen S, Monroe P, Tien Tan C. Feasibility of automatic speech recognition for providing feedback during tablet-based treatment for apraxia of speech plus aphasia. Am J Speech Lang Pathol 2019 Jul 15;28(2S):818-834. [doi: 10.1044/2018_AJSLP-MSC18-18-0109]
24. Routhier S, Bier N, Macoir J. Smart tablet for smart self-administered treatment of verb anomia: two single-case studies in aphasia. Aphasiology 2014 Nov 14;30(2-3):1-21. [doi: 10.1080/02687038.2014.973361]
25. Brady MC, Myzoon A, Vandenberg K, REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators. Dosage, intensity, and frequency of language therapy for aphasia: a systematic review-based, individual participant data network meta-analysis. Stroke 2022 Mar;53(3):956-967. [doi: 10.1161/STROKEAHA.121.035216] [Medline: 34847708]
26. Szeto SG, Wan H, Alavinia M, Dukelow S, MacNeill H. Effect of mobile application types on stroke rehabilitation: a systematic review. J Neuroeng Rehabil 2023 Jan 24;20(1):12. [doi: 10.1186/s12984-023-01124-9] [Medline: 36694257]

XSL•FO
RenderX

27. Vaezipour A, Campbell J, Theodoros D, Russell T. Mobile apps for speech-language therapy in adults with communication disorders: review of content and quality. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e18858. [doi: 10.2196/18858] [Medline: 33118953]
28. Bevan N, Carter J, Earthy J, Geis T, Harker S. New ISO standards for usability, usability reports and usability measures. 2016 Presented at: International Conference on Human Computer Interaction 2016; Toronto, Canada p. 268-278. [doi: 10.1007/978-3-319-39510-4_25]
29. Repetto C, Paolillo MP, Tuena C, Bellinzona F, Riva G. Innovative technology-based interventions in aphasia rehabilitation: a systematic review. Aphasiology 2020 Sep;25:1-24. [doi: 10.1080/02687038.2020.1819957]
30. Griffel J, Leinweber J, Spelter B, Roddam H. Patient-centred design of aphasia therapy apps: a scoping review. Aphasie Verwandte Geb 2019;46(2):6-21 [FREE Full text]
31. Kamaruddin AI, Masrom M, Aziz MAA, Mohd Noor N. End-user needs and perspectives on speech therapy apps for aphasia. Presented at: 2023 IEEE 2nd National Biomedical Engineering Conference (NBEC); Melaka, Malaysia p. 127-132. [doi: 10.1109/NBEC58134.2023.10352627]
32. Schatz R, Kim E, Bieber A, et al. User-centered design of a multimodal training app for people with aphasia. 2020 Presented at: Annual Convention of the American Speech-Language-Hearing Association (ASHA); Nov 19-21, 2020; San Diego, CA URL: https://voigt-antons.de/publication/2020-11-01-OC17 [accessed 2025-08-11]
33. Leonard C, Rochon E, Laird L. Treating naming impairments in aphasia: findings from a phonological components analysis treatment. Aphasiology 2008 Sep;22(9):923-947. [doi: 10.1080/02687030701831474]
34. Simic T, Leonard C, Laird L, Stewart S, Rochon E. The effects of intensity on a phonological treatment for anomia in post-stroke aphasia. J Commun Disord 2021;93:106125. [doi: 10.1016/j.jcomdis.2021.106125] [Medline: 34166970]
35. Boyle M, Coelho CA. Application of semantic feature analysis as a treatment for aphasic dysnomia. Am J Speech Lang Pathol 1995 Nov;4(4):94-98. [doi: 10.1044/1058-0360.0404.94]
36. Efstratiadou EA, Papathanasiou I, Holland R, Archonti A, Hilari K. A Systematic review of semantic feature analysis therapy studies for aphasia. J Speech Lang Hear Res 2018 May 17;61(5):1261-1278. [doi: 10.1044/2018_JSLHR-L-16-0330] [Medline: 29710193]
37. Kertesz A, Raven JC. WAB-R: Western Aphasia Battery-Revised: PsychCorp; 2007.
38. Goodglass H, Kaplan E, Weintraub S. Boston Naming Test: Lea & Febiger; 1983. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft27208-000 [accessed 2025-08-11]
39. Kim ES, Laird L, Wilson C, et al. Implementation and effects of an information technology-based intervention to support speech and language therapy among stroke patients with aphasia: protocol for a virtual randomized controlled trial. JMIR Res Protoc 2021 Jul 2;10(7):e30621. [doi: 10.2196/30621] [Medline: 34255727]
40. Dekhtyar M, Braun EJ, Billot A, Foo L, Kiran S. Videoconference administration of the western aphasia battery-revised: feasibility and validity. Am J Speech Lang Pathol 2020 May 8;29(2):673-687. [doi: 10.1044/2019_AJSLP-19-00023] [Medline: 32191122]
41. Varley R, Cowell PE, Dyson L, Inglis L, Roper A, Whiteside SP. Self-administered computer therapy for apraxia of speech: two-period randomized control trial with crossover. Stroke 2016 Mar;47(3):822-828. [doi: 10.1161/STROKEAHA.115.011939] [Medline: 26797664]
42. Hilari K, Byng S, Lamping DL, Smith SC. Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): evaluation of acceptability, reliability, and validity. Stroke 2003 Aug;34(8):1944-1950. [doi: 10.1161/01.STR.0000081987.46660.ED] [Medline: 12855827]
43. Lomas J, Pickard L, Bester S, Elbard H, Finlayson A, Zoghaib C. The communicative effectiveness index: development and psychometric evaluation of a functional communication measure for adult aphasia. J Speech Hear Disord 1989 Feb;54(1):113-124. [doi: 10.1044/jshd.5401.113] [Medline: 2464719]
44. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McLelland IL, editors. Usability Evaluation In Industry: Taylor and Francis; 1996:189-194.
45. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
46. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
47. Gelman A, Tuerlinckx F. Type S error rates for classical and Bayesian single and multiple comparison procedures. Comput Stat 2000 Sep;15(3):373-390. [doi: 10.1007/s001800000040]
48. Quique YM, Evans WS, Dickey MW. Acquisition and generalization responses in aphasia naming treatment: a meta-analysis of semantic feature analysis outcomes. Am J Speech Lang Pathol 2019 Mar 11;28(1S):230-246. [doi: 10.1044/2018_AJSLP-17-0155]
49. Sze WP, Hameau S, Warren J, Best W. Identifying the components of a successful spoken naming therapy: a meta-analysis of word-finding interventions for adults with aphasia. Aphasiology 2021 Jan 2;35(1):33-72. [doi: 10.1080/02687038.2020.1781419]

XSL•FO
RenderX

50. Doedens WJ, Meteyard L. Measures of functional, real-world communication for aphasia: a critical review. Aphasiology 2020 Apr 2;34(4):492-514. [doi: 10.1080/02687038.2019.1702848]
51. Kurfess C, Corsten S, Nickel MT, Knieriemen M, Kreiter D, Lauer N. Peer-to-peer support: digital networking in aphasia to improve quality of life (PeerPAL). Front Commun 2023 Sep 7;8:1187233. [doi: 10.3389/fcomm.2023.1187233]
52. Devane N, Behn N, Marshall J, Ramachandran A, Wilson S, Hilari K. The use of virtual reality in the rehabilitation of aphasia: a systematic review. Disabil Rehabil 2023 Nov 6;45(23):3803-3822. [doi: 10.1080/09638288.2022.2138573]
53. Wallace SJ, Worrall L, Rose T, et al. A core outcome set for aphasia treatment research: the ROMA consensus statement. Int J Stroke 2019 Feb;14(2):180-185. [doi: 10.1177/1747493018806200] [Medline: 30303810]
54. Macoir J, Lavoie M, Routhier S, Bier N. Key factors for the success of self-administered treatments of poststroke aphasia using technologies. Telemed J E Health 2019 Aug;25(8):663-670. [doi: 10.1089/tmj.2018.0116] [Medline: 30142032]
55. Greinacher R, Kojic T, Meier L, Parameshappa RG, Moller S, Voigt-Antons JN. Impact of tactile and visual feedback on breathing rhythm and user experience in VR exergaming. 2020 Presented at: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX); Athlone, Ireland p. 635-640. [doi: 10.1109/QoMEX48832.2020.9123141]
56. Best B, Campbell J, Roxbury T, Worthy P, Copland DA. Exploring the usability and feasibility of a mobile music listening application for people living in the community with post-stroke aphasia. Disabil Rehabil 2024 Jan;46(2):344-353. [doi: 10.1080/09638288.2022.2161646] [Medline: 36593725]
57. Simic T, Leonard C, Laird L, Cupit J, Höbler F, Rochon E. A usability study of internet-based therapy for naming deficits in aphasia. Am J Speech Lang Pathol 2016 Nov 1;25(4):642-653. [doi: 10.1044/2016_AJSLP-15-0030] [Medline: 27784033]

Abbreviations

ASR: automatic speech recognition
BNT: Boston Naming Test
CETI: Communication Effectiveness Index
PCA: Phonological Components Analysis
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SAQoL-39g: Stroke and Aphasia Quality of Life Scale - 39
SFA: Semantic Features Analysis
SLP: speech-language pathologist
SSNT: Study Stimulus Naming Test
SUS: System Usability Scale
UCD: user-centered design
WAB-R: Western Aphasia Battery-Revised

Edited by L Buis; submitted 18.10.24; peer-reviewed by J Pierce, L Zhang; revised version received 16.06.25; accepted 22.06.25; published 18.08.25.

Please cite as:
Kim ES, Laird L, Wilson C, Stewart S, Mildner P, Möller S, Schatz R, Spang RP, Voigt-Antons JN, Rochon E
Outcomes of an App-Based Intervention to Target Naming Among Individuals With Poststroke Aphasia: Virtual Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e67711
URL: https://mhealth.jmir.org/2025/1/e67711
doi:10.2196/67711

© Esther S Kim, Laura Laird, Carlee Wilson, Steven Stewart, Philip Mildner, Sebastian Moller, Raimund Schatz, Robert P Spang, Jan-Niklas Voigt-Antons, Elizabeth Rochon. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone App–Guided Pulmonary Rehabilitation in Chronic Respiratory Diseases: Randomized Controlled Trial

Chiwook Chung[1,2*], MD, PhD; Deog Kyeom Kim[3*], MD, PhD; Jung-Kyu Lee[3], MD, PhD; Eun Young Heo[3], MD, PhD; Hee Kwon[4], BS; Dongbum Kim[4], MS; Woo Jin Kim[5], PhD, MD; Sei Won Lee[2], MD, PhD

[1]Division of Pulmonary, Allergy and Critical Care Medicine, Department of Internal Medicine, Hallym University Dongtan Sacred Heart Hospital, Hallym University College of Medicine, Hwaseong, Republic of Korea

[2]Department of Pulmonary and Critical Care Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul, Republic of Korea

[3]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University College of Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul, Republic of Korea

[4]LifeSemantics Corp, Seoul, Republic of Korea

[5]Department of Internal Medicine and Environmental Health Center, Kangwon National University, Chuncheon, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Sei Won Lee, MD, PhD
Department of Pulmonary and Critical Care Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul, Republic of Korea

Abstract

Background: Pulmonary rehabilitation improves exercise capacity, dyspnea, quality of life, and survival in patients with chronic respiratory disease. However, center-based pulmonary rehabilitation programs remain unavailable in many health care facilities due to several barriers. To address this, we developed a smartphone app that enabled individuals to perform pulmonary rehabilitation at home.

Objective: We aimed to evaluate the efficacy of smartphone app–guided pulmonary rehabilitation in improving exercise capacity in individuals with chronic respiratory diseases.

Methods: This was a multicenter prospective, single-blind, randomized controlled trial conducted in 2022. A total of 100 participants with chronic respiratory disease, including chronic obstructive pulmonary disease, asthma, and lung cancer, were recruited, with equal distribution (50:50) between the intervention group and the control group. The intervention group followed a 12-week app-guided rehabilitation program, while the control group received standard outpatient treatment. The primary outcome was the 6-minute walk test distance (6MWD) after the 12-week rehabilitation period. Secondary outcomes included quality of life questionnaires and health care usage.

Results: Among the 100 participants included, 88 completed the follow-up visit (41 in the intervention group and 47 in the control group). Their median age was 68.0 years, and 72 (81.8%) were men. Most participants (n=70, 79.5%) had a smoking history, with a median of 40.0 pack-years. Their forced expiratory volume in 1 second was a median of 63.0% (IQR 50.5 - 71.5). Most participants (n=85, 96.6%) had chronic obstructive pulmonary disease. After the 12-week rehabilitation program, 6MWD was not different between the intervention and control group (median 490.0, IQR 468.8 - 556.3 vs 485.0, IQR 440.0 - 527.3 m). Assuming a clinically minimal effective change of 25 meters in 6MWD, only 7 out of 41 participants among the intervention group achieved the minimal clinically important differences after the rehabilitation program. Quality of life questionnaire scores, including the St George's Respiratory Questionnaire and Hospital Anxiety and Depression Scale, did not differ between groups. In addition, none of the participants experienced hospitalization or emergency room visits during the study period. Regarding the service satisfaction questionnaire, more than 3-quarters of the intervention group (34/41) rated their scores as ≥17/20.

Conclusions: In this study, smartphone app–guided pulmonary rehabilitation failed to improve exercise capacity and quality of life in patients with chronic respiratory diseases. However, the results indicated that older adults with chronic respiratory conditions can safely use smartphone app–guided pulmonary rehabilitation. Thus, smartphone app–guided pulmonary rehabilitation may be a feasible option for older adults with chronic respiratory disease.

Trial Registration: ClinicalTrials.gov NCT05299385; https://clinicaltrials.gov/study/NCT05299385

(JMIR Mhealth Uhealth 2025;13:e64884) doi:10.2196/64884

XSL•FO
RenderX

KEYWORDS

pulmonary rehabilitation; chronic respiratory diseases; COPD; lungs; respiratory; mHealth; apps; smartphones; walking; randomized controlled trial

Introduction

Chronic respiratory diseases are significant contributors to global mortality and morbidity [1,2]. In 2019, chronic obstructive lung disease (COPD), lung cancer, and lower respiratory infections ranked among the top 10 leading causes of disability-adjusted life years for adults aged 50 years and older [1]. Individuals with chronic respiratory diseases experience various clinical deteriorations, including poor quality of life and reduced exercise capacity [3,4]. Previous studies have shown that lower exercise capacity and physical activity levels are associated with higher risks of exacerbation and mortality in individuals with COPD [5,6]. Exercise capacity and physical activity serve as independent predictors of mortality in individuals with advanced lung cancer [7].

Pulmonary rehabilitation, a comprehensive intervention, aims to enhance the physical and psychological well-being of individuals with chronic respiratory disease through exercise training, education, and behavior change [4]. Pulmonary rehabilitation increased exercise capacity, improved quality of life, and reduced dyspnea in individuals with chronic respiratory diseases [3,4,8]. Previous studies have highlighted that muscle wasting and dysfunction, known as respiratory sarcopenia, were commonly found in individuals with chronic respiratory diseases, which is associated with decreased exercise capacity and respiratory function [9,10]. By combining pulmonary rehabilitation with nutritional management, we can potentially prevent and treat respiratory sarcopenia and further enhance respiratory function [10].

Previous landmark studies presented pulmonary rehabilitation programs that include exercise training lasting 30 - 45 minutes per day, 3 - 5 days per week, and at least 8 - 12 weeks [11,12]. However, these studies also highlighted challenges associated with center-based pulmonary rehabilitation, such as limited facilities, transportation barriers, low social support, and lack of motivation [13,14]. Furthermore, the COVID-19 pandemic necessitated alternatives to face-to-face center-based rehabilitation for individuals with chronic respiratory diseases [15]. Consequently, we designed a pulmonary rehabilitation program and developed a smartphone app called Redpill Breath (LifeSemantics), enabling individuals to perform pulmonary rehabilitation at home. Our study aimed to evaluate the efficacy of smartphone app–guided pulmonary rehabilitation in improving exercise capacity in individuals with chronic respiratory diseases.

Methods

Study Design

This multicenter prospective, single-blind, randomized controlled trial aimed to evaluate the clinical effectiveness and safety of pulmonary rehabilitation using the smartphone app Redpill Breath for patients with chronic respiratory disease, including COPD, asthma, and lung cancer. In 2022, a total of 100 participants (50 in the intervention group and 50 in the control group) were recruited from the outpatient pulmonology clinics at Asan Medical Center (Seoul, Republic of Korea), Kangwon National University Hospital (Chuncheon, Republic of Korea), and Seoul Metropolitan Government Seoul National University Boramae Medical Center (Seoul, Republic of Korea). Participants were recruited from each health care center through poster advertisements from January 2022 to September 2022. The trial started in January 2022 and finished in November 2022. Eligible participants were interviewed by trained clinical research coordinators to confirm their eligibility. After the eligibility screening, participants were randomly assigned to either the intervention or control group using a prepared randomization sequence.

At the time of enrollment, clinical research coordinators collected participants' baseline information, including age, sex, height, weight, BMI, smoking history, and underlying comorbidities. The study outcomes were measured twice, before and after the 12-week rehabilitation program. In addition, the intervention group completed a service satisfaction questionnaire after the end of the study. The intervention group participated in a 12-week app–guided rehabilitation program. They were equipped with a Bluetooth-connected wrist pulse oximeter (either the WristOx 2 3150, Nonin, or the MD300W628, ChoiceMMed) to monitor heart rate and oxygen saturation during exercise. In contrast, the control group received standard outpatient treatment, which included medical treatment using bronchodilator inhalers according to current guidelines, along with brief advice on the importance of physical activity. They were also provided with a leaflet containing instructions on pulmonary rehabilitation (center-based rehabilitation programs were not provided). The leaflet provided static illustrations and text-based instructions for walking exercise, muscle training, and an exercise diary [16]. In contrast, the app provided interactive guidance, video demonstrations, progress tracking, and motivational messages, offering a more engaging and personalized experience. This study was conducted in line with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines (the CONSORT-EHEALTH checklist is provided in Checklist 1).

Inclusion and Exclusion Criteria

The inclusion criteria for this study were as follows: (1) participants aged 19-80 years; (2) presence of dyspnea symptoms (at least 1 point on the modified Medical Research Council Dyspnea Scale [mMRC]) and impaired lung function, defined as [1] either forced vital capacity (FVC) or forced expiratory volume in 1 second (FEV1) <80% predicted for patients with lung cancer or [2] postbronchodilator FEV1/FVC <0.7 and postbronchodilator FEV1 <80% predicted for patients with other respiratory disease; (3) ownership of an Android smartphone (OS 8.0 or higher, with internal memory of 32 GB or higher) or an iPhone (iOS 13.0 or higher, with internal

XSL•FO
RenderX

memory of 32 GB or higher); (4) ability to effectively use a smartphone and mobile software, and willingness to bring their smartphone for a walking test during the 12-week clinical trial; and (5) voluntary informed consent.

The exclusion criteria were as follows: (1) respiratory distress syndrome caused by neuromuscular diseases, spinal cord injury, or thoracic deformities; (2) participation in pulmonary rehabilitation treatment within 6 months of the screening date; (3) disease exacerbation requiring antibiotics or steroid treatment within 2 weeks of the screening date; (4) unstable cardiovascular diseases (such as unstable angina pectoris, acute myocardial infarction, or severe coarctation of aorta); (5) uncontrolled pulmonary arterial hypertension; (6) physical disabilities or cognitive impairments; (7) pregnancy or breastfeeding; (8) illiteracy; (9) participation in other clinical trials within 90 days of the screening date; and (10) deemed inappropriate to participate based on the clinical research associate's opinion.

Smartphone App and Rehabilitation Program

The smartphone app Redpill Breath was developed collaboratively by the investigators and LifeSemantics Corp. Available on both Android and iPhone platforms (requiring at least Android 8.0 or iOS 13.0), the app incorporates an exercise program based on established guidelines [3,17,18]. Key features include a 6-minute walk test (6MWT), a daily walking exercise regimen (25 - 35 minutes) combined with muscle training (20 - 25 minutes), exercise records, and real-time heart rate and oxygen saturation monitoring during walks. Users benefit from audio guidance for walking exercises and video instruction for muscle training. In addition, the exercise program's intensity can be tailored based on the 6-minute walk test distance (6MWD; Figure 1A-I).

Figure 1. Screenshots of the Redpill Breath app. Shown are the following: (**A**) opening screen; (**B**) account creation; (**C**) the 6-minute walk test, which displays step counts and test time; (**D**) walking distance and timer for walking exercises; (**E**) completion of walking exercise, which displays walking distance, heart rate, oxygen saturation, and rating of perceived exertion scale evaluated after exercise; (**F**) exercise level change; (**G**) instruction videos for muscle training exercise program; (**H**) completion of muscle training exercise; and (**I**) exercise report, which displays exercise completion rate per week and day.

Study Outcome

The original study protocol was designed to measure the study outcomes twice, at 8 and 12 weeks after the baseline. However, this was modified to assess the outcomes only once, 12 weeks after the baseline. The primary outcome measure was the 6MWD after the 12-week rehabilitation program. The test was conducted on a 30-meter track using previously described methods [19]. Secondary outcomes included assessments using the mMRC dyspnea scale, the St. George's Respiratory Questionnaire (SGRQ) [20], the Hospital Anxiety and Depression Scale (HADS) [21], as well as tracking hospitalizations, emergency room visits, and a service satisfaction questionnaire (Patient Global Assessment, applicable to the intervention group only). The satisfaction questionnaire consisted of 4 questions with scores ranging from 1 to 5 for each question (total score range 4 - 20). Within the intervention group, participants with compliance >70% were categorized into the per-protocol analysis set (PP set) according to the study protocol.

Randomization and Sample Size Calculation

Participants were evenly divided into the intervention and control groups at a 1:1 ratio (50 participants in each group).

XSL•FO
RenderX

Upon enrollment, participants were randomly assigned to either group before undergoing baseline tests, and the randomization sequence was concealed until the time of randomization. The random sequence was generated using stratified block randomization with the Proc plan procedure in SAS software version 9.4 (SAS Institute).

The sample size estimation was based on the minimal important difference in the 6MWD observed in previous studies [22-24]. We assumed a 25-meter difference in 6MWD between the groups and an SD of 52.5 meters. With an alpha level of .05 and a power of 80%, to achieve a 95% CI with the difference between groups >0 meters, we needed a minimum of 40 participants in each group. To account for a potential dropout rate of 20%, we aimed for 100 participants (50 in each group).

Statistical Analysis

Continuous variables were reported as median (IQR) and compared using the Mann-Whitney *U* test. Categorical variables were presented as numbers (percentages) and compared using the chi-square or Fisher exact test. All *P* values were 2-tailed, with statistical significance set at $P<.05$. The statistical analyses were conducted using MedCalc Statistical Software version 22.021 (MedCalc Software Ltd).

Ethical Considerations

The study protocol was approved by the institutional review board of each hospital (Asan Medical Center: 2021‐1108, Kangwon National University Hospital: A-2021-07-009, and Seoul Metropolitan Government Seoul National University Boramae Medical Center: 30-2021-103). Written informed consent was obtained from all participants prior to inclusion. The study complied with the guidelines stipulated in the Declaration of Helsinki, and all methods were performed in accordance with the relevant guidelines. This study was registered in the ClinicalTrials.gov database (NCT05299385). Data were deidentified and participants did not receive compensation.

Results

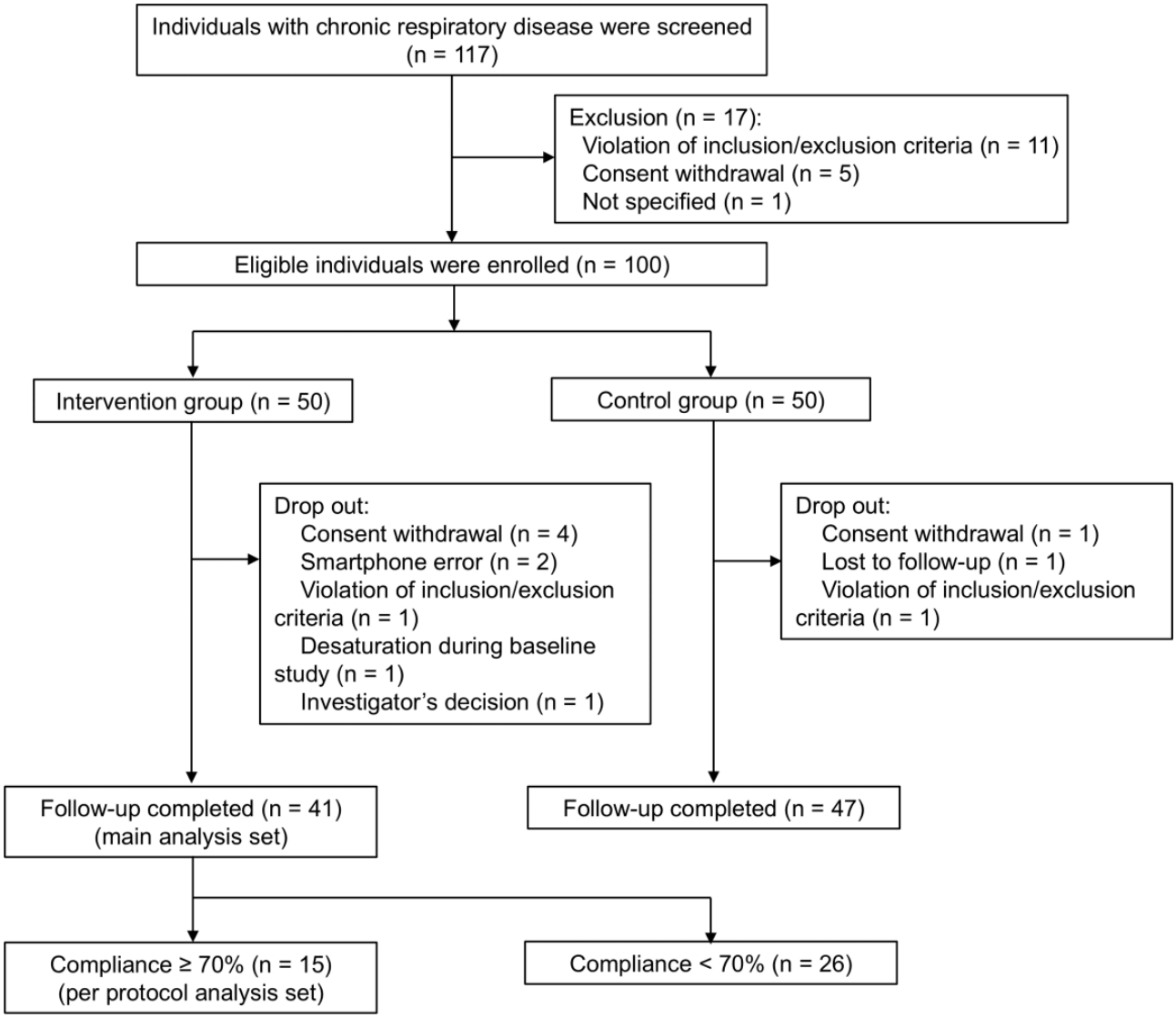
Participant Baseline Characteristics

Figure 2 illustrates the study process. Out of 117 individuals screened, 100 were eligible and included in the study. In the intervention group, 9 participants dropped out, and 41 completed the follow-up visit. Among them, 15 achieved exercise program compliance of ≥70% and were categorized into the per-protocol analysis set as follows: 0% ‐ 20% compliance (11 participants), 21% ‐ 40% (8 participants), 41% ‐ 60% (5 participants), 61% ‐ 80% (3 participants), and 81% ‐ 100% (14 participants). In the control group, 3 participants dropped out, and 47 completed the follow-up visit. No participants experienced musculoskeletal injury related to the rehabilitation program during the study period.

Table 1 presents the baseline characteristics of the study participants. Their median age was 68.0 (IQR 62.0 ‐ 73.0) years, and 72 (81.8%) were men. Most of the participants (n=85, 96.6%) had COPD as their respiratory condition, with a median disease duration of 4.9 years. Furthermore, 70 (79.5%) participants had a smoking history, with a median of 40.0 pack-years. Their FEV1 was reduced, with a median of 63.0% (IQR 50.5% ‐ 71.5%).

XSL•FO
RenderX

Figure 2. Study flowchart.



XSL•FO
RenderX

Table . Baseline characteristics of study participants.

Variables	Total (n=88)	Intervention (n=41)	Control (n=47)	*P* value
Age (years), median (IQR)	68.0 (62.0 - 73.0)	67.0 (59.0 - 71.0)	69.0 (63.3 - 73.0)	.10
Male sex, n (%)	72 (81.8)	36 (87.8)	36 (76.6)	.14
Height (cm), median (IQR)	165.0 (159.1 - 169.0)	166.0 (159.1 - 170.1)	164.0 (159.3 - 168.0)	.28
Body weight (kg), median (IQR)	65.0 (56.0 - 73.0)	67.0 (56.5 - 77.0)	64.0 (6.0 - 71.0)	.25
BMI (kg/m^2), median (IQR)	24.1 (21.7 - 26.7)	24.9 (21.9 - 27.5)	24.0 (21.7 - 26.4)	.56
Respiratory disease, n (%)				.26
COPD[a]	85 (96.6)	41 (100.0)	44 (93.6)	
Asthma	2 (2.3)	0 (0.0)	2 (4.3)	
Bronchiectasis	1 (1.1)	0 (0.0)	1 (2.1)	
Duration of disease (years), median (IQR)	4.9 (2.1 - 9.7)	4.8 (1.9 - 9.4)	5.0 (2.2 - 10.1)	.51
Ever-smokers, n (%)	70 (79.5)	34 (82.9)	36 (76.6)	.47
Amount, pack-years (n=70), median (IQR)	40.0 (20.0 - 45.0)	35.0 (22.5 - 45.0)	40.0 (20.0 - 45.0)	.72
Lung function, median (IQR)				
FEV1[b] (% predicted)	63.0 (50.5 - 71.5)	63.0 (53.0 - 71.3)	63.0 (48.3 - 71.8)	.98
FVC[c] (% predicted)	83.0 (73.5 - 88.5)	84.0 (75.5 - 89.3)	81.0 (73.0 - 87.8)	.57
FEV1/FVC	0.54 (0.44 - 0.62)	0.54 (0.45 - 0.62)	0.54 (0.43 - 0.64)	.88

[a]COPD: chronic obstructive pulmonary disease.

[b]FEV1: forced expiratory volume in one second.

[c]FVC: forced vital capacity.

Clinical Parameters of Participants

Analysis of the clinical outcomes at follow-up indicated no significant differences in 6MWD, the mMRC dyspnea scale, SGRQ, and HADS between the intervention and control groups in both the main analysis set and PP set (Tables 2 and 3). Assuming a clinically minimal effective change of 25 m in 6MWD, only 7 out of 41 participants among the intervention group achieved the minimal clinically important differences after the rehabilitation program.

None of the participants required hospitalization or emergency room visits related to their underlying respiratory disease or rehabilitation program during the study period. Regarding the service satisfaction questionnaire, more than half of the intervention group rated scores ≥17 (4 - 8: 1 participant; 9 - 12: 6 participants; 13 - 16: 11 participants; 17 - 20: 23 participants).

XSL•FO
RenderX

Table . Comparison of clinical outcomes of study participants at baseline and follow-up (main analysis).

Variables	Total (n=88)	Intervention (n=41)	Control (n=47)	*P* value
Baseline, median (IQR)				
mMRC[a] dyspnea scale	1.0 (1.0 - 2.0)	1.0 (1.0 - 2.0)	1.0 (1.0 - 1.8)	.40
6MWD[b] (m)	483.5 (444.0 - 520.0)	495.0 (460.0 - 541.3)	480.0 (440.0 - 503.8)	.10
HADS[c]	9.0 (6.0 - 14.0)	10.0 (7.0 - 14.3)	8.5 (5.0 - 14.0)	.16
SGRQ[d]				
Total	30.3 (20.1 - 39.6)	26.5 (18.8 - 36.5)	32.8 (20.6 - 41.2)	.18
Symptoms	44.1 (33.5 - 54.9)	41.7 (30.1 - 53.8)	48.5 (33.6 - 57.5)	.13
Activity	45.2 (32.2 - 57.9)	44.4 (30.5 - 53.9)	50.7 (32.4 - 63.5)	.24
Impacts	15.5 (7.3 - 28.3)	13.0 (5.9 - 25.8)	18.3 (8.3 - 30.7)	.26
Follow-up, median (IQR)				
mMRC dyspnea scale	1.0 (1.0 - 1.0)	1.0 (1.0 - 1.0)	1.0 (1.0 - 1.0)	.35
6MWD (m)	485.0 (460.0 - 544.0)	490.0 (468.8 - 556.3)	485.0 (440.0 - 527.3)	.35
HADS	10.0 (5.0 - 14.0)	10.0 (5.0 - 13.3)	10.0 (4.3 - 15.0)	.70
SGRQ				
Total	30.9 (18.5 - 44.6)	27.9 (17.6 - 43.7)	33.5 (22.1 - 44.8)	.21
Symptoms	44.8 (32.4 - 61.2)	44.9 (25.2 - 54.1)	44.6 (33.2 - 65.4)	.34
Activity	50.7 (31.5 - 64.0)	44.4 (25.7 - 57.8)	51.6 (40.1 - 64.9)	.12
Impacts	16.9 (5.9 - 30.1)	13.8 (5.1 - 27.2)	18.0 (7.8 - 31.2)	.34

[a]mMRC: modified Medical Research Council.

[b]6MWD: 6-minute walk test distance.

[c]HADS: Hospital Anxiety and Depression Scale.

[d]SGRQ: St George's Respiratory Questionnaire.

XSL•FO
RenderX

Table . Comparison of clinical outcomes of study participants at baseline and follow-up (compliance ≥70%, per-protocol analysis).

Variables	Total (n=62)	Intervention (n=15)	Control (n=47)	*P* value
Baseline, median (IQR)				
mMRC[a] dyspnea scale	1.0 (1.0 - 2.0)	1.0 (1.0 - 1.8)	1.0 (1.0 - 1.8)	>.99
6MWD[b] (m)	480.0 (440.0 - 520.0)	480 (421.3 - 551.8)	480.0 (440.0 - 503.8)	.65
HADS[c]	9.0 (5.8 - 14.0)	9.0 (6.3 - 16.0)	8.5 (5.0 - 14.0)	.54
SGRQ[d]				
Total	31.4 (20.2 - 39.5)	26.5 (19.5 - 36.0)	32.8 (20.6 - 41.2)	.26
Symptoms	46.8 (33.5 - 54.1)	38.0 (19.5 - 53.4)	48.5 (33.6 - 57.5)	.16
Activity	50.7 (33.0 - 58.6)	44.4 (44.4 - 56.2)	50.7 (32.4 - 63.5)	.66
Impacts	16.3 (7.8 - 30.4)	12.1 (6.1 - 17.5)	18.3 (8.3 - 30.7)	.23
Follow-up, median (IQR)				
mMRC dyspnea scale	1.0 (1.0 - 1.0)	1.0 (1.0 - 1.0)	1.0 (1.0 - 1.0)	.29
6MWD (m)	485.0 (440.0 - 538.0)	485.0 (463.3 - 558.8)	485.0 (440.0 - 527.3)	.54
HADS	10.5 (5.0 - 15.0)	11.0 (5.0 - 15.0)	10.0 (4.3 - 15.0)	.90
SGRQ				
Total	31.2 (18.5 - 44.8)	27.9 (15.2 - 42.4)	33.5 (22.1 - 44.8)	.22
Symptoms	41.0 (30.6 - 62.8)	36.8 (21.2 - 53.1)	44.6 (33.2 - 65.4)	.12
Activity	51.4 (33.1 - 64.1)	44.4 (22.6 - 56.2)	51.6 (40.1 - 64.9)	.13
Impacts	17.5 (6.7 - 31.5)	12.8 (4.3 - 33.7)	18.0 (7.8 - 31.2)	.44

[a]mMRC: modified Medical Research Council.

[b]6MWD: 6-minute walk test distance.

[c]HADS: Hospital Anxiety and Depression Scale.

[d]SGRQ: St George's Respiratory Questionnaire.

Discussion

Principal Findings

In this study, we evaluated the effectiveness of smartphone app–guided pulmonary rehabilitation in patients with chronic respiratory diseases. These results did not indicate a significant improvement in clinical outcomes through the use of the smartphone app–guided pulmonary rehabilitation program. Nevertheless, no rehabilitation-related musculoskeletal injuries, hospitalizations, or emergency room visits were recorded during the study period. Taken together, these results suggest that smartphone app–guided pulmonary rehabilitation may be feasible for older adults with chronic respiratory disease. Further study will be needed to clarify the clinical efficacy of this approach.

Clinical Usefulness of Pulmonary Rehabilitation

Improvements in exercise capacity and physical activity are associated with favorable clinical outcomes in patients with chronic respiratory diseases [5,25]. Exercise capacity has been found to be a significant predictor of mortality in patients with COPD [5]. In addition, low levels of physical activity were associated with increased risks for acute exacerbation or mortality in patients with COPD [6]. In addition, the BODE index, which includes BMI, airflow limitation (FEV1), dyspnea severity (measured by the mMRC dyspnea scale), and exercise capacity (6MWD), serves as a robust predictor of mortality in patients with COPD [26]. Furthermore, disease-specific health status and quality of life, measured by SGRQ, predicted disease exacerbation and mortality in patients with COPD [27]. Anxiety or depression, as measured using HADS, has been associated with increased mortality in patients with COPD [28]. Furthermore, positive changes in HADS scores persisted even one year after the 6-week rehabilitation program for patients with COPD [29]. Therefore, although our findings did not demonstrate significant improvements in clinical outcomes, including 6MWD, SGRQ, and HADS, these findings indicate that a vigorous rehabilitation program should be recommended for patients with chronic respiratory diseases, pending further investigation [3,4,8,30].

Compliance With Rehabilitation Programs

Previous studies have emphasized compliance with rehabilitation programs in home-based and smartphone app–based pulmonary rehabilitation. Participants who demonstrated good compliance significantly improved clinical parameters [16,31,32]. Specifically, a previous study estimated a decrease of −0.22 (95% CI −0.74 to 0.31) in the COPD assessment test for every 7-day increase in app use during pulmonary rehabilitation [33]. However, another study also indicated that the number of smartphone app users declined as the study progressed [34]. In

XSL•FO
RenderX

our study, approximately one-third of participants completed exercise programs at least ≥70% of the scheduled sessions. Regarding the lack of motivation being a key factor for poor compliance in home-based pulmonary rehabilitation [35], physicians should consistently educate patients on maintaining steady rehabilitation practices during routine clinical care. Moreover, text messages or telephone contacts from health care providers may enhance motivation [36].

Strengths

Despite most of the participants being older adults, many of them expressed satisfaction with the app and exercise program. Furthermore, we did not observe any health care usage or musculoskeletal injuries related to the rehabilitation program during the study period. These findings support the feasibility of applying smartphone app–guided pulmonary rehabilitation to older adults with chronic respiratory diseases.

Limitations and Future Perspectives

This study also has limitations. First, although we designed the app to measure daily step counts using a smartphone-mounted pedometer, technical issues prevented us from obtaining these data. Because exercise capacity, represented by physical activity, is related to the prognosis of patients with chronic respiratory diseases (such as COPD) [5,6], further app development is necessary. Second, the participants' preserved exercise capacity at baseline and the inherent limitations of the 6MWT may have contributed to the failure to achieve the primary outcome of improving the 6MWD. A previous study reported that healthy participants aged 60 - 69 years had a 6MWD of 559 (80) meters, while those aged 70 - 80 had a 6MWD of 514 (71) meters [37]. Our study participants had a median age of 68.0 (IQR 62.0 - 73.0) years, and their baseline 6MWD was a median of 483.5 meters (IQR 444.0 - 520.0), suggesting minimal impairments in exercise capacity. Notably, the 6MWT may have a "ceiling effect," making it less responsive than the incremental shuttle walk test for detecting changes in COPD patients with preserved exercise capacity [38,39]. Similarly, clinical trials in pulmonary arterial hypertension, which traditionally use 6MWD as the primary endpoint, have excluded patients with preserved exercise capacity (eg, baseline 6MWD >450 m) [40]. Therefore, we recommend that future research consider participants' baseline exercise capacity during recruitment. Moreover, the incremental shuttle walk test or cardiopulmonary exercise test should be considered for a more accurate assessment of exercise capacity [31,41,42]. Third, although we did not observe health care usage during the 12-week study period, long-term follow-up after rehabilitation may reveal the impact on disease progression or mortality. Finally, participants who dropped out after the enrollment did not undergo the follow-up measurements for the study outcomes. Thus, an intention-to-treat analysis could not be performed.

Conclusions

In conclusion, this study highlighted a lack of improvement in clinical outcomes using smartphone app–guided pulmonary rehabilitation. However, the results did indicate that older adults with chronic respiratory conditions can safely use smartphone app–guided pulmonary rehabilitation. Thus, smartphone app–guided pulmonary rehabilitation may be feasible for older adults with chronic respiratory disease.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; RS-2023-NR077159), the Bio and Medical Technology Development Program of the NRF funded by the MSIT (RS-2022-NR067421 and RS-2023-00222687), the National Institute of Health research project (2024ER080601), and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (HI23C1234). The funder did not have any role in the design of the study and did not have any role in collection, analysis, and interpretation of data and in writing the manuscript.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

CC contributed to conceptualization, investigation, formal analysis, methodology, data curation, writing—original draft, and writing—review and editing. DKK, J-KL, EYH, HK, and DK contributed to conceptualization, investigation, methodology, data curation, and writing—review and editing. WJK and SWL contributed to conceptualization, investigation, formal analysis, methodology, data curation, writing—original draft, writing—review and editing, funding acquisition, and supervision. WJK is the co-corresponding author and can be reached via email at pulmo2@kangwon.ac.kr or phone at (82) 33-610-5355.

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File, 1080 KB - mhealth_v13i1e64884_app1.pdf]

XSL•FO
RenderX

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020;396(10258). [doi: 10.1016/S0140-6736(20)30925-9]
2. Viegi G, Maio S, Fasola S, Baldacci S. Global burden of chronic respiratory diseases. J Aerosol Med Pulm Drug Deliv 2020 Aug;33(4):171-177. [doi: 10.1089/jamp.2019.1576] [Medline: 32423274]
3. Bolton CE, Bevan-Smith EF, Blakey JD, et al. British Thoracic Society guideline on pulmonary rehabilitation in adults. Thorax 2013 Sep;68 Suppl 2:ii1-i30. [doi: 10.1136/thoraxjnl-2013-203808] [Medline: 23880483]
4. Spruit MA, Singh SJ, Garvey C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. Am J Respir Crit Care Med 2013 Oct 15;188(8):e13-e64. [doi: 10.1164/rccm.201309-1634ST] [Medline: 24127811]
5. Oga T, Nishimura K, Tsukino M, Sato S, Hajiro T. Analysis of the factors related to mortality in chronic obstructive pulmonary disease: role of exercise capacity and health status. Am J Respir Crit Care Med 2003 Feb 15;167(4):544-549. [doi: 10.1164/rccm.200206-583OC] [Medline: 12446268]
6. Garcia-Rio F, Rojo B, Casitas R, et al. Prognostic value of the objective measurement of daily physical activity in patients with COPD. Chest 2012 Aug;142(2):338-346. [doi: 10.1378/chest.11-2014] [Medline: 22281798]
7. Jones LW, Hornsby WE, Goetzinger A, et al. Prognostic significance of functional capacity and exercise behavior in patients with metastatic non-small cell lung cancer. Lung Cancer (Auckl) 2012 May;76(2):248-252. [doi: 10.1016/j.lungcan.2011.10.009] [Medline: 22112290]
8. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2015 Feb 23;2015(2):CD003793. [doi: 10.1002/14651858.CD003793.pub3] [Medline: 25705944]
9. Wüst RCI, Degens H. Factors contributing to muscle wasting and dysfunction in COPD patients. Int J Chron Obstruct Pulmon Dis 2007;2(3):289-300. [Medline: 18229567]
10. Nagano A, Wakabayashi H, Maeda K, et al. Respiratory sarcopenia and sarcopenic respiratory disability: concepts, diagnosis, and treatment. J Nutr Health Aging 2021;25(4):507-515. [doi: 10.1007/s12603-021-1587-5] [Medline: 33786569]
11. Casaburi R, Patessio A, Ioli F, Zanaboni S, Donner CF, Wasserman K. Reductions in exercise lactic acidosis and ventilation as a result of exercise training in patients with obstructive lung disease. Am Rev Respir Dis 1991 Jan;143(1):9-18. [doi: 10.1164/ajrccm/143.1.9] [Medline: 1986689]
12. Maltais F, LeBlanc P, Simard C, et al. Skeletal muscle adaptation to endurance training in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1996 Aug;154(2 Pt 1):442-447. [doi: 10.1164/ajrccm.154.2.8756820] [Medline: 8756820]
13. Johnston KN, Young M, Grimmer KA, Antic R, Frith PA. Barriers to, and facilitators for, referral to pulmonary rehabilitation in COPD patients from the perspective of Australian general practitioners: a qualitative study. Prim Care Respir J 2013 Sep;22(3):319-324. [doi: 10.4104/pcrj.2013.00062] [Medline: 23797679]
14. Augustine A, Bhat A, Vaishali K, Magazine R. Barriers to pulmonary rehabilitation - A narrative review and perspectives from a few stakeholders. Lung India 2021;38(1):59-63. [doi: 10.4103/lungindia.lungindia_116_20] [Medline: 33402639]
15. Tsutsui M, Gerayeli F, Sin DD. Pulmonary rehabilitation in a Post-COVID-19 world: telerehabilitation as a new standard in patients with COPD. Int J Chron Obstruct Pulmon Dis 2021;16(379-391):379-391. [doi: 10.2147/COPD.S263031] [Medline: 33642858]
16. Lee JH, Lee HY, Jang Y, et al. Efficacy of unsupervised home-based pulmonary rehabilitation for patients with chronic ostructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 2020;15(2297-2305):2297-2305. [doi: 10.2147/COPD.S268683] [Medline: 33061347]
17. Holland AE, Cox NS, Houchen-Wolloff L, et al. Defining modern pulmonary rehabilitation. an official American Thoracic Society workshop report. Annals ATS 2021 May;18(5):e12-e29. [doi: 10.1513/AnnalsATS.202102-146ST]
18. Consensus document on pulmonary rehabilitation in korea 2015. The Korea Academy of Tuberculosis and Respiratory Diseases. 2015. URL: https://www.lungkorea.org/bbs/index.html?code=guide&category=&gubun=&page=2&number=3484&mode=view&keyfield=&key= [accessed 2025-07-18]
19. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med Jul 2002;166(1):111-117. [doi: 10.1164/ajrccm.166.1.at1102]
20. Kim YS, Byun MK, Jung WY, et al. Validation of the Korean version of the St. George's Respiratory Questionnaire for patients with chronic respiratory disease. Tuberc Respir Dis 2006;61(2):121. [doi: 10.4046/trd.2006.61.2.121]
21. Oh S, Min K, Park D. A study on the standardization of the hospital anxiety and depression scale for Koreans: a comparison of normal, depressed and anxious groups. J Korean Neuropsychiatr Assoc 1999;38(2):289-296.
22. Granger CL, Holland AE, Gordon IR, Denehy L. Minimal important difference of the 6-minute walk distance in lung cancer. Chron Respir Dis 2015 May;12(2):146-154. [doi: 10.1177/1479972315575715] [Medline: 25749346]
23. Holland AE, Hill CJ, Rasekaba T, Lee A, Naughton MT, McDonald CF. Updating the minimal important difference for six-minute walk distance in patients with chronic obstructive pulmonary disease. Arch Phys Med Rehabil 2010 Feb;91(2):221-225. [doi: 10.1016/j.apmr.2009.10.017] [Medline: 20159125]

XSL•FO
RenderX

24. Park S, Kim JY, Lee JC, et al. Mobile phone app-based pulmonary rehabilitation for chemotherapy-treated patients with advanced lung cancer: pilot study. JMIR Mhealth Uhealth 2019 Feb 4;7(2):e11094. [doi: 10.2196/11094] [Medline: 30714943]
25. Vaes AW, Garcia-Aymerich J, Marott JL, et al. Changes in physical activity and all-cause mortality in COPD. Eur Respir J 2014 Nov;44(5):1199-1209. [doi: 10.1183/09031936.00023214] [Medline: 25063247]
26. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med 2004 Mar 4;350(10):1005-1012. [doi: 10.1056/NEJMoa021322] [Medline: 14999112]
27. Müllerova H, Gelhorn H, Wilson H, et al. St George's Respiratory Questionnaire score predicts outcomes in patients with COPD: analysis of individual patient data in the COPD biomarkers qualification consortium database. Chronic Obstr Pulm Dis 2017 Mar 28;4(2):141-149. [doi: 10.15326/jcopdf.4.2.2017.0131] [Medline: 28848923]
28. Vikjord SAA, Brumpton BM, Mai XM, Vanfleteren L, Langhammer A. The association of anxiety and depression with mortality in a COPD cohort. The HUNT study, Norway. Respir Med 2020 Sep;171(106089):106089. [doi: 10.1016/j.rmed.2020.106089] [Medline: 32799059]
29. Griffiths TL, Burr ML, Campbell IA, et al. Results at 1 year of outpatient multidisciplinary pulmonary rehabilitation: a randomised controlled trial. Lancet 2000 Jan 29;355(9201):362-368. [doi: 10.1016/s0140-6736(99)07042-7] [Medline: 10665556]
30. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. Am J Respir Crit Care Med 2017 Mar 1;195(5):557-582. [doi: 10.1164/rccm.201701-0218PP] [Medline: 28128970]
31. Chung C, Kim AR, Kim D, et al. Smartphone application-based rehabilitation in patients with chronic respiratory and cardiovascular diseases. Sci Rep 2024 Feb 6;14(1):3018. [doi: 10.1038/s41598-024-53583-2] [Medline: 38321153]
32. Chung C, Lee JW, Lee SW, Jo MW. Clinical efficacy of mobile app-based, self-directed pulmonary rehabilitation for patients with chronic obstructive pulmonary disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2024 Jan 4;12:e41753. [doi: 10.2196/41753] [Medline: 38179689]
33. Crooks MG, Elkes J, Storrar W, et al. Evidence generation for the clinical impact of myCOPD in patients with mild, moderate and newly diagnosed COPD: a randomised controlled trial. ERJ Open Res 2020 Oct;6(4):33263052. [doi: 10.1183/23120541.00460-2020] [Medline: 33263052]
34. North M, Bourne S, Green B, et al. A randomised controlled feasibility trial of E-health application supported care vs usual care after exacerbation of COPD: the RESCUE trial. NPJ Digit Med 2020;3(145):145. [doi: 10.1038/s41746-020-00347-7] [Medline: 33145441]
35. Li Y, Qian H, Yu K, Huang Y. Nonadherence in home-based pulmonary rehabilitation program for COPD patients. Can Respir J 2020;2020(5146765):5146765. [doi: 10.1155/2020/5146765] [Medline: 31998425]
36. Demeyer H, Louvaris Z, Frei A, et al. Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. Thorax 2017 May;72(5):415-423. [doi: 10.1136/thoraxjnl-2016-209026] [Medline: 28137918]
37. Casanova C, Celli BR, Barria P, et al. The 6-min walk distance in healthy subjects: reference standards from seven countries. Eur Respir J 2011 Jan;37(1):150-156. [doi: 10.1183/09031936.00194909] [Medline: 20525717]
38. Pepin V, Brodeur J, Lacasse Y, et al. Six-minute walking versus shuttle walking: responsiveness to bronchodilation in chronic obstructive pulmonary disease. Thorax 2007 Apr;62(4):291-298. [doi: 10.1136/thx.2006.065540] [Medline: 17099077]
39. Chae G, Ko EJ, Lee SW, et al. Stronger correlation of peak oxygen uptake with distance of incremental shuttle walk test than 6-min walk test in patients with COPD: a systematic review and meta-analysis. BMC Pulm Med 2022 Mar 24;22(1):102. [doi: 10.1186/s12890-022-01897-0] [Medline: 35331171]
40. Frost AE, Langleben D, Oudiz R, et al. The 6-min walk test (6MW) as an efficacy endpoint in pulmonary arterial hypertension clinical trials: demonstration of a ceiling effect. Vascul Pharmacol 2005 Jun;43(1):36-39. [doi: 10.1016/j.vph.2005.03.003] [Medline: 15890561]
41. Chung C, Kim AR, Jang IY, et al. Smartphone application-based rehabilitation in patients with chronic respiratory and cardiovascular diseases: a randomised controlled trial study protocol. BMJ Open 2023 Sep 20;13(9):e072698. [doi: 10.1136/bmjopen-2023-072698] [Medline: 37730392]
42. Wang CH, Chou PC, Joa WC, et al. Mobile-phone-based home exercise training program decreases systemic inflammation in COPD: a pilot study. BMC Pulm Med 2014 Aug 30;14(142):25175787. [doi: 10.1186/1471-2466-14-142] [Medline: 25175787]

Abbreviations

6MWD: 6-minute walk test distance
6MWT: 6-minute walk test

XSL•FO
RenderX

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
COPD: chronic obstructive lung disease
FEV1: forced expiratory volume in 1 second
FVC: forced vital capacity
HADS: Hospital Anxiety and Depression Scale
mMRC: modified Medical Research Council Dyspnea Scale
PP: per-protocol analysis
SGRQ: St. George's Respiratory Questionnaire

Edited by L Buis; submitted 30.07.24; peer-reviewed by A Holland, GP Tinoco Areas; revised version received 27.03.25; accepted 11.06.25; published 25.08.25.

Please cite as:
Chung C, Kim DK, Lee JK, Heo EY, Kwon H, Kim D, Kim WJ, Lee SW
Smartphone App–Guided Pulmonary Rehabilitation in Chronic Respiratory Diseases: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e64884
URL: https://mhealth.jmir.org/2025/1/e64884
doi:10.2196/64884

© Chiwook Chung, Deog Kyeom Kim, Jung-Kyu Lee, Eun Young Heo, Hee Kwon, Dongbum Kim, Woo Jin Kim, Sei Won Lee. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Smartwatch Device on Adherence to Home-Based Cardiac Rehabilitation in Patients With Coronary Heart Disease: Randomized Controlled Trial

Sisi Zhang[1], MD, PhD; Yuehui Wang[2], MD, PhD; Jiahui Wu[1], MD, PhD; Changsheng Ma[1*], MD, PhD; Xiaoping Meng[3*], MD, PhD

[1]Department of Cardiology, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

[2]Department of Geriatrics, Jilin Geriatrics Clinical Research Center, The First Hospital of Jilin University, Changchun, China

[3]Department of Cardiovascular Disease, Affiliated Hospital of Changchun University of Chinese Medicine, No.1478 Gongnong Road, Chaoyang District, Changchun, China

[*]these authors contributed equally

Corresponding Author:
Xiaoping Meng, MD, PhD
Department of Cardiovascular Disease, Affiliated Hospital of Changchun University of Chinese Medicine, No.1478 Gongnong Road, Chaoyang District, Changchun, China

Abstract

Background: Digital technologies have the potential to overcome many of the limitations associated with traditional center-based cardiac rehabilitation (CBCR), such as limited accessibility, transportation barriers, and low adherence. In this context, home-based cardiac rehabilitation (HBCR) has emerged as a promising alternative. However, maintaining adherence and providing continuous supervision in remote settings remain a major challenge. Smartwatch-based interventions may offer a novel solution to support and monitor patients in HBCR programs, yet robust clinical evidence is still limited.

Objective: This study was designed to investigate the effectiveness of a smartwatch-facilitated HCBR model in improving exercise adherence and health-related outcomes in patients with coronary heart disease (CHD), aiming to improve adherence and other outcomes related to the secondary prevention of cardiovascular disease.

Methods: We conducted a prospective, single-center, randomized, parallel-controlled, non-blinded trial. Eligible participants were adults (≥18 years) with a confirmed diagnosis of CHD, recruited from a tertiary hospital in Jilin Province, China. Participants were randomly assigned in a 1:1 ratio to either the intervention group (smartwatch-facilitated HBCR) or the control group (standard HBCR) for a duration of 3 months. The intervention group received a comprehensive program delivered via a smartwatch, including real-time feedback, remote supervision, physical activity monitoring, and educational content. The control group received conventional HBCR without technological assistance. The primary outcome was adherence to the HBCR program, assessed using the Home-Based Cardiac Rehabilitation Exercise Adherence Scale. Secondary outcomes included cardiopulmonary function (peak VO_2 measured via cardiopulmonary exercise testing), anxiety (Generalized Anxiety Disorder-7), depression (Patient Health Questionnaire-9), and health-related quality of life (36-Item Short Form Survey, SF-36), evaluated at baseline and at 3 months.

Results: Between January 1 and December 30, 2023, a total of 62 patients (mean [SD] age 59.93 [10.06] years; 40.4% women [25/62]) were enrolled and randomized to the intervention group (n=32) or control group (n=30). Baseline characteristics were well balanced between the groups. At 3 months, participants in the smartwatch group demonstrated significantly higher adherence scores compared to the control group (P<.01). Additionally, the smartwatch group showed significant improvements in peak VO_2 (P<.01), anxiety (GAD-7, P<.01), depression (PHQ-9, P<.01), and selected domains of SF-36 (P<.05). No serious adverse events related to the intervention were reported, and user engagement with the smartwatch platform was high throughout the study period.

Conclusions: This study demonstrates that a smartwatch-facilitated HBCR model is both feasible and effective in enhancing adherence and improving clinical outcomes among patients with CHD. These findings support the integration of wearable technology into routine HBCR and lay the groundwork for future large-scale, multicenter trials.

Trial Registration: Chinese Clinical Trial Registry; ChiCTR2400088039; https://www.chictr.org.cn/bin/project/edit?pid=215602

(JMIR Mhealth Uhealth 2025;13:e70848) doi:10.2196/70848

XSL•FO
RenderX

KEYWORDS

smartwatch; digital technology; home-based cardiac rehabilitation; adherence; anxiety; depression; quality of life

Introduction

Background

Coronary heart disease (CHD), as the leading cause of death globally, presents significant public health and economic challenges [1-3]. Exercise-based cardiac rehabilitation (CR), an important management strategy for CHD, is strongly recommended by clinical practice guidelines from the American College of Cardiology, American Heart Association, and European Society of Cardiology [4,5]. Despite its recognized benefits, access to these services remains limited for many patients in China due to barriers such as accessibility, time constraints, and uneven economic and technological development across regions [6]. As an alternative, home-based cardiac rehabilitation (HBCR) is similarly effective in improving clinical outcomes [7]. However, the benefits of HBCR largely depend on long-term adherence. Previous studies have reported inconsistent findings regarding patients' adherence to HBCR [8,9].

With the development of mobile health (mHealth) technology, remote cardiac rehabilitation programs now use the capabilities of internet-connected devices, wearable monitors, smartphones, and smartwatches. These tools enable the delivery of comprehensive rehabilitation services that were previously available only through in-person sessions at medical facilities [10]. The American Heart Association has noted that mHealth technologies play a crucial role in enhancing physical activity, supporting smoking cessation, promoting weight loss, managing blood glucose levels, and controlling hypertension [11]. The World Health Organization has emphasized the role of digital health technologies in improving health care delivery, particularly for chronic disease management and rehabilitation [12]. Studies across various countries have demonstrated the effectiveness of digital platforms, mobile applications, and wearable devices in supporting CR. For instance, a study in Canada reported that tele-rehabilitation programs provided comparable outcomes to in-person CR in terms of functional recovery and patient satisfaction [13]. Additionally, mHealth technologies now allow for the near real-time monitoring of patient performance data at home. This capability covers various aspects of physical activity, such as intensity, duration, distance traveled, number of steps, and sedentary periods, as well as vital signs such as heart rate and blood pressure, thus facilitating remotely supervised HBCR programs.

Among various wearable devices, smartwatches stand out due to their ability to provide continuous, real-time monitoring of vital signs and physical activity metrics such as heart rate, steps, exercise intensity, and sedentary time. This makes them uniquely suited for remotely supervised HBCR programs, offering timely feedback and enhanced patient engagement. Despite this potential, research specifically evaluating smartwatch-facilitated HBCR remains limited, especially in China where mHealth-based HBCR programs are still in early development stages and data on patient acceptance and adherence are scarce. In China, the development of mHealth HBCR programs is still in the nascent stages; research on patient acceptance and adherence to these mHealth-enabled HBCR programs remains sparse.

Objectives

This randomized controlled trial was designed to assess the effectiveness of a remote HBCR program supported by a smartwatch device and to determine whether this innovative mHealth HBCR model enhances adherence to CR and improves other outcomes related to the secondary prevention of cardiovascular disease.

Method

Study Design and Participants

This study was a single-center, 2-arm parallel design randomized controlled trial performed in a tertiary hospital in Jilin Province, China, from January 2023 to December 2023.

Inclusion Criteria

The inclusion criteria were as follows: (1) age of more than 18 years old; (2) diagnosed with CHD (including myocardial infarction and stable angina), with stable clinical manifestations and cardiovascular-related examination indicators; (3) a documented diagnosis of CHD and if the patient had either undergone percutaneous coronary intervention, experienced a recent myocardial infarction, or were referred for secondary prevention following a clinical evaluation. The need for cardiac rehabilitation was based on established clinical guidelines and individualized physician recommendations.

Exclusion Criteria

The exclusion criteria included (1) unstable angina pectoris, acute myocardial infarction, uncontrolled atrial tachycardia, asthma, obstructive hypertrophic cardiomyopathy, severe pulmonary hypertension, acute myocarditis, severe valvular heart disease, and other conditions where cardiopulmonary exercise testing is contraindicated (nerve, skeletal muscle, and rheumatic diseases); (2) uncontrolled hypertension (systolic blood pressure >160 mmHg or diastolic blood pressure >100 mmHg at rest); (3) uncontrolled heart failure (New York Heart Association class III-IV); (4) have severe motor, cognitive, mental illnesses, or visual impairment limiting the smartphone and smartwatch use; (5) high - risk for falls; and (6) being unwilling or unable to provide informed consent.

Randomization

The participants enrolled were randomized 1:1 to receive the smartwatch device-based CR or usual care CR alone. Participants were assigned to either the control or the intervention group at a ratio of 1:1 through computerized randomization. Due to the nature of the intervention, neither the participants nor the outcome assessors were blinded to the group assignments.

XSL•FO
RenderX

Care providers involved in the study were assigned based on their roles to ensure consistency and minimize potential bias. A dedicated team of cardiac rehabilitation specialists and research staff was responsible for participant enrollment, baseline assessments, and follow-up evaluations. To maintain protocol adherence, health care providers delivering the intervention were specifically trained in the use of the smartwatch-based rehabilitation program, while those managing the control group followed standard cardiac rehabilitation guidelines. Importantly, the care providers administering the intervention did not overlap with those assessing outcomes to reduce potential bias. Additionally, all care providers were blinded to the study hypotheses and instructed to follow standardized procedures for patient interactions.

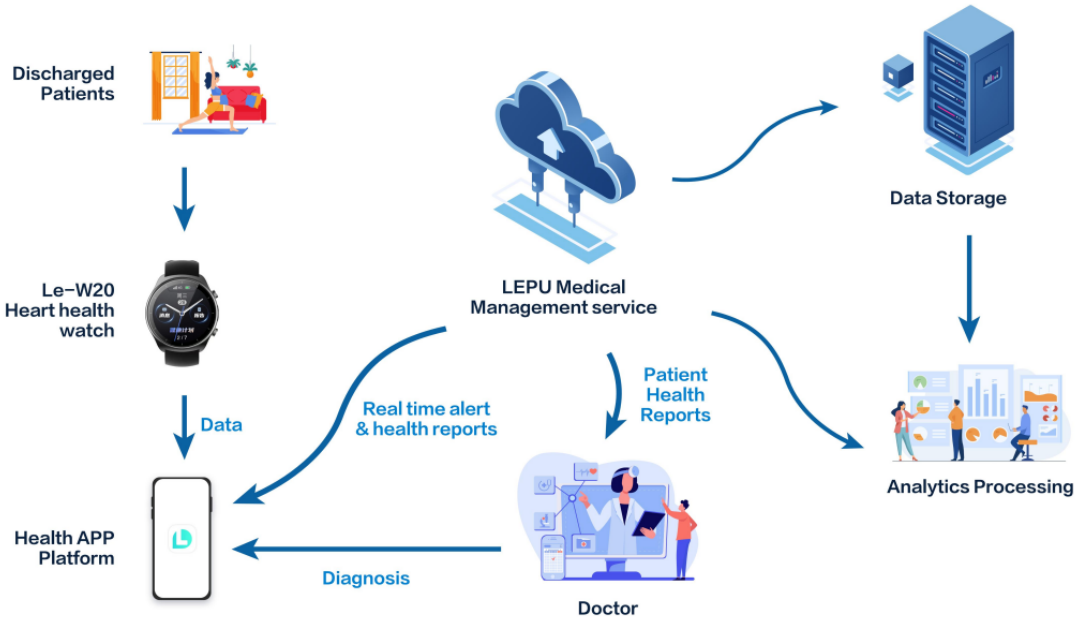
Control Group

Participants in the control group received a standard HBCR program based on the secondary prevention guidelines [14].

Intervention Group

The smartwatch-based HBCR program was delivered through a smartphone app. At the time of discharge, participants in the intervention group were provided with a smartwatch pre-installed with the intervention application. Additionally, they were instructed to download the corresponding mobile app onto their personal smartphones via the designated app stores. A step-by-step user guide was provided, and trained health care staff conducted a one-on-one demonstration to ensure participants could successfully log in, navigate the interface, and use key features such as exercise tracking, medication reminders, and data sharing. Technical support was available through a dedicated helpline to assist with any issues related to app access or usage during the study period. The intervention consists of real-time monitoring, automatic reminders, health education, data sharing, and individualized follow-up (Figure 1).

Figure 1. Diagram showing the flow of data for the intervention group.



Real-time Monitoring

With high-precision sensors, the smartwatch continuously tracks the patient's heart rate, blood pressure, and blood oxygen levels during exercise. Furthermore, the smartwatch could calculate the number of calories burned during the exercise session, helping patients understand their energy expenditure and adjust their exercise intensity accordingly. This comprehensive real-time monitoring enables both patients and health care providers to track progress and make informed decisions to optimize the effectiveness and safety of the HBCR program.

Automatic Reminders

Specific prescriptions set by the system are transmitted to the patient's smartwatch, providing daily reminders at designated times. For example, the doctor inputs the medication prescription into the system and sets reminder times. At 8 AM each day, the patient receives a reminder to take antihypertensive medication, including the specific name, dosage, and other important instructions. At 8 PM each evening, the smartwatch reminds the patient to take their statin medication (such as simvastatin, atorvastatin, rosuvastatin, or pravastatin) with the specific dosage. Additionally, every Monday, Wednesday, and Friday (or at patient-chosen intervals), the patient receives a notification of their exercise plan for the day, based on their exercise prescription. At 6 AM each day, the patient is reminded of their daily hydration goal, such as drinking 1500 - 2000 ml of water. If the smartwatch shuts down due to low battery or other reasons, it will send any missed reminders once it is powered back on. After the patient completes a task, they can mark it as "Completed," which will automatically be sent into the hospital information system.

XSL•FO
RenderX

Health Education

After enrollment, the eligible participants will receive app-based messages regarding dietary habits, physical activity, prevention of hypertension, diseases, cardiovascular disease, and knowledge of smoking cessation. The app also included educational resources for symptom management and other CR-related topics in the form of text, images, and videos in the native language.

Data Sharing

Each day, we share with patients and their caregivers a summary of their motion data via the app installed on their smartphones. By sharing these results, the health care providers give feedback, which is then sent to the caregivers for supervision and guidance.

Individualized Follow-Up

The app would pop up the goal achievement according to pre-set intervals automatically. The participants had to reply depending on whether they had achieved the tasks or not, in addition to the Borg scale score to evaluate the exercise intensity.

Data Collection, Outcomes, and Measures

Data were collected at baseline and 3 months. The demographic and clinical data were collected from questionnaires or electronic medical records.

The primary outcome measure was HBCR adherence as assessed using the home-based exercise training adherence questionnaire (HETAQ), which is composed of 17 items across the following domains: program comprehension, patient motivation, family support, and peer support. The total score ranged from 0 to 100 (higher scores indicating greater HBCR adherence), and ≥60 was considered to adhere well to the HBCR program [15].

Secondary outcomes included exercise capacity, quality of life, and psychological distress. Exercise capacity was measured using the cardiopulmonary exercise testing (CPET). The parameters obtained from the CPET, including oxygen consumption (VO_2), minute ventilation (VE), carbon dioxide output (VCO_2), peak metabolic equivalents (METs), peak VE/VCO_2 ratio, ventilatory efficiency (VE/VCO_2 slope), work efficiency ($deltaVO_2/deltaWR$), O_2 plus, and anaerobic threshold (AT).

The quality of life was assessed with the 36-item Short-Form Health Survey(SF-36), with the score ranging from 0 to 100, and higher scores reflect better quality of life [16]. Depression and anxiety symptoms were measured using the Patient Health Questionnaire 9-item scale (PHQ-9) and the Generalized Anxiety Disorder 7-item scale (GAD-7) [17,18]. Scores on the PHQ-9 range from 0 to 27, with results of no depression, or mild, moderate, moderately severe, and severe depression. Scores on the GAD-7 range from 0 to 21 points, with result of mild, moderate, and severe levels of anxiety symptoms. Higher points on these scales were closely related to psychological distress.

Sample Size

Sample size calculation was performed using the results of the preliminary study, which showed that the mean (SD) HETAQ score was 85.99 (6.68) in the intervention group versus 72.68 (12.93) in the control group. With the 95% confidence interval, 80% power, 5% type I error, and 10% dropout rate, we determined a minimum of 16 participants per group at recruitment. The calculation was based on two independent *t* tests, using PASS, version 15 (NCSS LLC).

Statistical Analysis

Data were analyzed using the intention-to-treat principles using the Statistical Package for Social Sciences software (SPSS, version 26.0). Statistical analyses were conducted by an independent statistician who was blinded to group assignments. The dataset was anonymized, with group labels concealed during statistical processing to prevent bias in interpretation.

The statistical analysis followed the intention-to-treat principle, meaning all randomized participants were included in their originally assigned groups, regardless of adherence to the intervention. Participants who remained in the trial but did not use the application as prescribed were still included in the analysis. Their adherence data were categorized separately in exploratory analyses to assess potential differences in outcomes based on actual app usage.

Baseline comparisons were performed using descriptive statistics. The categorical variables were summarized as frequencies and percentages, and the quantitative variables were expressed as means (standard deviation) or medians (interquartile range) depending on the distribution of the values. Before conducting group comparisons, all continuous variables were assessed for normality using the Shapiro-Wilk test. Variables conforming to a normal distribution were analyzed using independent-sample *t* tests for between-group comparisons and paired-sample *t* tests for within-group comparisons. For variables that did not meet the normality assumption, non-parametric tests such as the Mann-Whitney *U* test (between groups) and Wilcoxon signed-rank test (within groups) were applied. Within-group analyses compared pre- and post-intervention outcomes to evaluate changes over time in each group separately. Between-group analyses at follow-up assessed the differences in outcomes between the intervention and control groups.The *P* value ≤.05 was considered significant.

Ethical Considerations

This study protocol adhered to the Declaration of Helsinki and received approval from the institutional review board (IRB) of the Affiliated Hospital of Changchun Traditional Chinese Medicine (approval no. CCZYFYKYLL2023-164). All participants provided written informed consent before enrollment in the study. The consent process ensured that participants were fully informed about the study's purpose, procedures, potential risks, and benefits.

To safeguard user privacy and ensure data security, each participant is required to log into the device using a unique account and password. Collected data are anonymized before being uploaded to a secure cloud server in full compliance with advanced data protection standards.

All participants who completed the trial received appropriate compensation for their time and commitment. Specifically, each participant was allowed to retain the smartwatch device used

XSL•FO
RenderX

during the study, which has an estimated retail value of approximately US $200. No additional monetary compensation was provided. This non-cash compensation was disclosed during the informed consent process and was provided equally to participants in both the intervention and control groups to ensure fairness and transparency.

Results

Overview

Figure 2 illustrates the CONSORT diagram of the flow of participants (CONSORT-eHEALTH checklist is provided in Checklist 1). A total of 70 eligible participants were randomly assigned to the smartwatch group (n=35; intervention group) or usual care (n=35). During the follow-up period, 3/35 participants dropped out in the intervention group and 5/35 dropped out in the control group. In the intervention group, 2 participants withdrew due to technical difficulties with the application, and 1 due to loss of interest. In the control group, 3 participants were lost to follow-up, and 2 withdrew due to personal reasons unrelated to the study.

Figure 2. CONSORT diagram of the flow of participants.

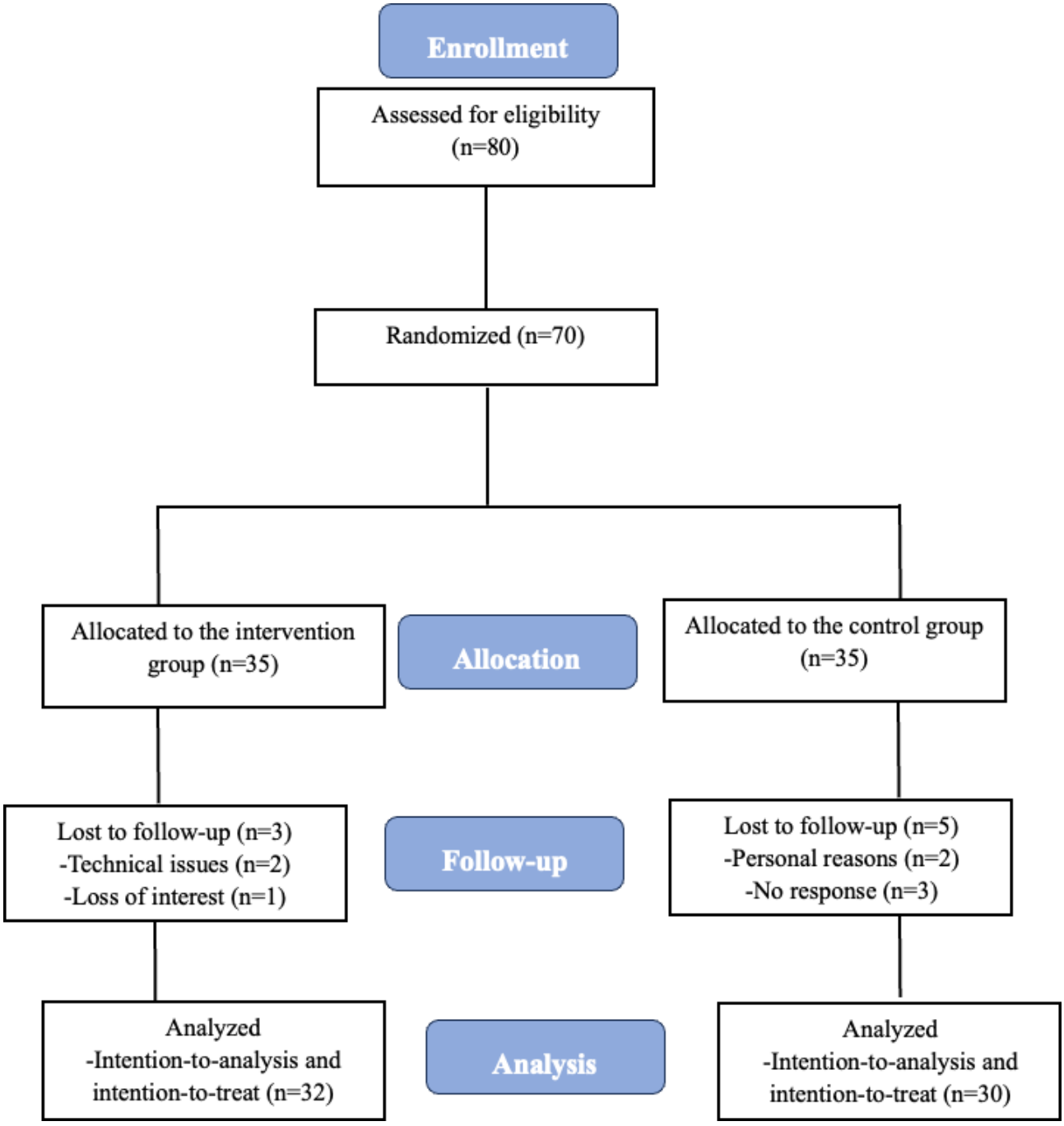


Table 1 shows the baseline characteristics of patients included in this study. There were no significant differences between the two groups at baseline considering the sociodemographic and clinical variables. The overall mean (SD) age of the participants

XSL•FO
RenderX

in the intervention and control groups was 59.93 (10) years and 62.67 (4.89) years, respectively.

XSL•FO
RenderX

Table . Baseline characteristics.

Variable	Total (n=62)	Intervention group (n=32）	Control group (n=30)	Chi-square test (*df*)	Two-tailed *t* test *(df)*	*P* value
Gender, n (%)				0	—[a]	>.99
Male	37 (59.6)	18 (56.2）	19 (63.3）			
Female	25 (40.4)	14 (43.8）	11 (36.7）			
Age (y), mean (SD)	59.93 (10.06)	62.67 (4.89)	57.2 (13)	—	1.52 (37)	.14
Height (cm), mean (SD)	167.3 (7.25)	167.07 (6.96)	167.53 (7.76)	—	−0.17 (58)	.86
Weight (kg), mean (SD)	69.4 (14.1)	69.3 (17.89)	69.5 (9.58)	—	−0.03 (48)	.97
BMI (kg/m^2), mean (SD)	25.47 (3.96)	25.97 (3.9)	24.9 (4.08)	—	0.67 (59)	.5
Marital status, n (%)				1.034 (1)	—	>.99
Married	60 (96.8)	32 (100）	28 (93.3）			
Single, divorced, or windowed	2 (3.22)	0 (0）	2 (6.7）			
Employment status, n (%)				1.64 (1)	—	.44
Retired	20 (32.3)	10 (31.2）	10 (33.3）			
Employed	32 (51.6)	18 (56.2）	14 (46.7）			
Others	10 (16.1)	4 (12.5）	6 (20）			
Medicine, n (%)						
Antiplatelets	58 (93.5)	28 (87.5）	30 (100）	1.15 (1)	—	.59
Statins	42 (67.7)	18 (56.2）	24 (80）	1.29 (1)	—	.45
β-blocker	4 (6.5)	2 (6.25）	2 (6.7）	0 (1)	—	1
ACEI[b] or ARB[c]	16 (25.8)	10 (31.3）	6 (20）	1.67 (1)	—	.39
CCB[d]	12 (19.4)	6 (18.8）	6 (20）	0.18 (1)	—	1
White blood cell count (×10^9/L), mean (SD)	5.76 (1.11)	5.56 (1.04)	5.94 (1.18)	—	−0.91 (58)	.36
Red blood cell count (×10^{12}/L), mean (SD)	4.68 (0.69)	4.43 (0.49)	4.91 (0.77)	—	−1.99 (46)	.06
Platelet count (×$10^{9/}$L), mean (SD)	210.48 (65.4)	196.5 (85)	223.53 (38.2)	—	−1.11 (38)	.27
Hemoglobin level (g/L), mean (SD)	149.72 (38.35)	157.64 (52.73)	142.33 (15.44)	—	1.07 (31)	.29
Low-density lipoprotein level (mmol/L), mean (SD)	2.66 (0.78)	2.6 (0.76)	2.71 (0.82)	—	−0.38 (58)	.7
Total cholesterol level (mmol/L), mean (SD)	4.51 (0.8)	4.32 (0.82)	4.69 (0.77)	—	−1.27 (57)	.21
High-density lipoprotein level (mmol/L), mean (SD)	1.16 (0.29)	1.1 (0.21)	1.21 (0.34)	—	−1.0 (50)	.30

XSL•FO
RenderX

Variable	Total (n=62)	Intervention group (n=32)	Control group (n=30)	Chi-square test (*df*)	Two-tailed *t* test *(df)*	*P* value
Triglyceride level (mmol/L), mean (SD)	1.54 (0.72)	1.59 (0.85)	1.49 (0.58)	—	0.39 (46)	.69
Blood urea nitrogen level (mmol/L), mean (SD)	6.17 (2)	6.49 (2.5)	5.86 (1.29)	—	0.86 (43)	.39
Serum creatinine level (μmol/L), mean (SD)	82.4 (29.99)	84.6 (40.85)	80.2 (13.59)	—	0.39 (35)	.69
Blood glucose level (mmol/L), mean (SD)	6.15 (1.75)	6.48 (2.39)	5.81 (0.63)	—	1.04 (31)	.3
Left ventricular measurement (mm), mean (SD)	47.12 (4.67)	47.29 (5.95)	46.92 (2.74)	—	0.19 (60)	.84

[a]not applicable.

[b]ACEI: angiotensin-converting enzyme inhibitor.

[c]ARB: angiotensin II receptor blocker.

[d]CCB: calcium channel blocker.

Primary Outcome

At 3 months, the mean (SD) HETAQ score was 89.17 (9.29) for the intervention group and 72.50 (17.17) for the control group (*P*=.003). However, there was no overall significant difference between the groups in the domains of peer support (*P*=.311; Table 2).

Table . HETAQ[a] score (mean, SD) differences between the groups at 3 months.

Category	Patient's perception	Patient's motivation	Family support	Peer support	Total score
Intervention group (n=32)	90.66 (12.91)	88.33 (11.70)	75.56 (9.65)	56.94 (13.42)	89.17 (9.29)
Control group (n=30)	50.83 (19.17)	58.05 (22.41)	59.00 (18.75)	50.83 (18.58)	72.50 (17.17)
Two-tailed *t* test *(df)*	7.68 (50)	4.639 (43)	3.041 (43)	1.033 (53)	3.307 (44)
P value	<.001	<.001	<.001	.311	<.001

[a]HETAQ: home-based exercise training adherence questionnaire.

Secondary Outcomes

Regarding the CPET parameters, only the peak VO_2 differed within and between groups postintervention. After intervention, the mean (SD) peak VO_2 of the intervention group (19.8 [4.3]) was significantly superior to that of the control group (15.88 [2.3], *P*<.05; Table 3).

XSL•FO
RenderX

Table . Comparing cardiopulmonary exercise testing (CPET) parameters at baseline and at 3 months after the intervention.

CPET	Intervention group (n=32）		*P* value[a]	Control group (n=30)		*P* value[a]	*P* value[b]	*P* value[c]
	Baseline	3 months		Baseline	3 months			
Peak VO_2[d] (ml/kg/min), mean (SD)	15.66 (3.3)	19.8 (4.3)	.002	16 (1.6)	15.88 (2.3)	.91	.81	<.001
Peak Mets[e], mean (SD)	4.37 (0.64)	5.17 (0.55)	.000	4.88 (0.74)	5.55 (1.7)	.23	.21	.55
Peak O_2 plus (ml/beat), mean (SD)	9.46 (2.47)	10.41 (2.59)	.004	10.15 (1.94)	9.2 (2.11)	.47	.58	.36
Peak ventilation (ml/kg/min), mean (SD)	35.51 (5.74)	38.58 (6.88)	.192	39.71 (7.96)	41.32 (9.52)	.56	.27	.54
AT[f] (ml/kg/min), median (IQR)	11.5 (10.1-12.85)	13.6 (11.95-14.9)	.012	12.4 (5.88-13.65)	12.4 (11.35-16.4)	.75	.73	.84
VE/VCO_2[g], median (IQR)	26.61 (23.48-30.33)	26.18 (24.2-28.25)	.069	25.32 (22.39-25.98)	23.43 (20.59-25.94)	.11	.12	.12
$\triangle VO_2/\triangle WR$[h] (ml/min/w), mean (SD)	8.31 (1.58)	8.71 (1.54)	.422	8.28 (1.38)	9.11 (2.53)	.29	.97	.71

[a]*P* value derived from test T1 (3 months) vs T0 (baseline).

[b]*P* value derived from the test T0 intervention group vs T0 control group.

[c]*P* value derived from the test T1 intervention group vs T1 control group.

[d]VO_2: oxygen consumption.

[e]Mets: metabolic equivalent of tasks.

[f]AT: anaerobic threshold.

[g]VE/VCO_2: ventilatory equivalent for carbon dioxide.

[h]$\triangle VO_2/\triangle WR$: functional gain or the increase in VO_2 per W.

Changes in PHQ-9 and GAD-7 scores in the two groups are shown in Table 4. At baseline, there was no significant difference in the mean PHQ-9 and GAD-7 scores between the two groups (*P*>.05). After the 3-month intervention, the mean PHQ-9 and GAD-7 scores of the intervention group were lower than those at baseline (*P*<.05), whereas the scores for the control group showed no differences over the same period Figure 3.

XSL•FO
RenderX

Table . Comparison of anxiety, depression, and quality of life between the two groups before and after intervention.

	Intervention group (n=32）		Control group (n=30）	
	Baseline	3 months	Baseline	3 months
Anxiety and depression				
PHQ-9[a] score, mean (SD)	7.6 (3.4)	3.2 (1.01)[b, c]	6.67 (3.66)	5.2 (3.57)
GAD-7[d] score, mean (SD)	6.33 (2.55)	1.93 (0.96)[b, c]	5.87 (4.09)	4.13 (3.15)
Quality of life (SF-36)				
Physical function, median (IQR)	90 (90-95)	95 (90-95)	90 (90-93.75)	95 (90-95)
Bodily pain, median (IQR)	52 (0-87.5)	62 (49-74)[b, c]	0 (0-68.75)	80 (80-88.5)
Social function, mean (SD)	67.52 (13.95)	61.53 (17.34)[b, c]	69.44 (10.72)	83.61 (11.69)
General health, median (IQR)	45 (40-52)	50 (43-52)[b]	45 (41.25-50)	52 (45-60.75)
Energy, mean (SD)	52.31 (17.27)	62.46 (12.02)[b, c]	52.08 (12.33)	72.25 (9.23)
Physical role, median (IQR)	50 (0-87.5)	25 (25-62.5)[b, c]	0 (0-68.75)	100 (75-100)
Emotional function, median (IQR)	33.33 (16.66-66.67)	33.33 (16.66-66.67)[b]	33.33 (0-58.33)	83.33 (33.33-100)
Mental health(MH), mean (SD)	53.23 (20.35)	69.23 (17.46)[b]	59.67 (12.92)	78.33 (7.71)
Health transition, median (IQR)	25 (25-50)	36 (25-75)[b, c]	25 (25-43.75)	75 (50-75)
Total score, mean (SD)	971.29 (116)	523.51 (69.72)[b, c]	1087.64 (93.57)	678.3 (62.48)[b]

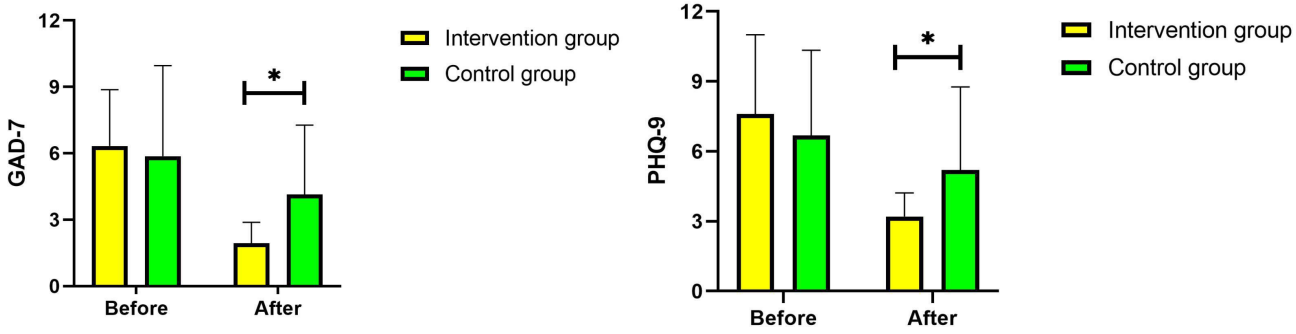
[a]PHQ-9: Patient Health Questionnaire 9-item scale

[b]Compared with the same group after intervention, $P<.05$.

[c]Compared with the control group after intervention, $P<.05$.

[d]GAD-7: Generalized Anxiety Disorder 7-item scale.

Figure 3. Changes in the Generalized Anxiety Disorder 7-item (GAD-7) scale and Patient Health Questionnaire 9-item (PHQ-9) scale scores before and after intervention.



The mean (SD) SF-36 scores of the intervention and control groups at baseline were 971.29 (116) points and 1087.64 (93.57) points, respectively. The scores did not differ significantly between the two groups ($P>.05$). After the intervention, the scores for the following domains (bodily pain, social function, general health, energy, physical role, emotional function, mental health, and health transition) improved significantly in the intervention group, whereas the control group showed no change. The between-group changes after intervention showed significant differences between the bodily pain, social function, energy, physical role, and health transition domains (Table 4).

Discussion

Principal Findings

Our findings suggest that the smartwatch-assisted HBCR program may help improve cardiac rehabilitation adherence and peak VO_2 , with potential benefits in quality of life, depression,

XSL•FO
RenderX

and anxiety symptoms. These improvements, although statistically significant, should be interpreted with caution given the modest sample size and short follow-up duration. Nonetheless, the high acceptance and engagement rates indicate that the proposed wearable device is well tolerated and feasible for use in HBCR programs, supporting its potential for broader clinical application.

Recently, the development of wearable devices has attracted attention given the advances in health care. Previous studies on the technology of HBCR have mainly included telephone, smartphone applications, and remote monitoring systems [19-21]. We have provided evidence that patients who underwent comprehensive HBCR based on the smartwatch application demonstrated better adherence and improved peak VO_2 after 3 months.

Non-adherence to HBCR is a major problem, especially for women, older adults, and ethnic minority populations [22]. As previous studies reported, unlike the center-based CR, the adherence for HBCR was usually in the lower range [23]. The adherence to HBCR is mainly determined by external factors, such as environmental factors, social factors, transportation factors, and patient-self factors. We believe that the smartwatch's feature of automatic reminder function has led to changes in patients' exercise behavior, thereby improving their adherence to physical activity. The reminder function has been reported to encourage patients' behavior toward better adherence [24].

One of the unique features of the intervention described here was the use of a commercially available smartwatch device to monitor participants' real-time ECG, heart rate, oxygen saturation, physical activity, energy consumption, exercise time, and sleep-related parameters. Indeed, such observations were helpful for patients to manage their activity levels and allow health care providers to tailor the individual intensity of exercise through their performance. One of the major advantages of digital health interventions is their ability to deliver care remotely, which is particularly beneficial for patients who may face geographical, mobility, or time-related barriers to attending traditional, in-person rehabilitation programs. The convenience of participating in rehabilitation from home has been shown to increase program participation rates and improve patient satisfaction [13]. This accessibility can be especially important in regions where health care resources are limited or for patients with chronic conditions who find frequent hospital visits challenging.

In particular, individualized follow-up led the patients to continue the exercise training and medicine therapy and thus resulted in better adherence in the intervention group. Convenience is another reason for achieving better adherence among the intervention group; it is more convenient to access the smartwatch rather than the smartphone or other wearable devices. However, in this study, we observed no overall significant difference between the intervention and control groups in the peer support domain of the HETAQ scale. This finding suggests that while digital health interventions can enhance self-management and adherence to treatment, their impact on social support may be limited in the absence of structured peer interaction components. Unlike structured cardiac rehabilitation programs that facilitate in-person interactions, our intervention was primarily technology-driven, with limited direct peer engagement. While participants had access to digital educational resources and self-monitoring tools, the intervention did not include dedicated features for fostering peer interaction, which may explain the lack of significant improvement in this domain. Future iterations of the intervention may benefit from incorporating virtual community support or group-based engagement features.

Peak oxygen uptake (peak VO_2), considered the gold standard for assessing aerobic fitness, is an important prognostic factor in patients with CHD. In our study, we observed a significantly greater improvement in peak VO_2 in the intervention group at 3 months than in the control group. Our findings are similar to a previous study, which reported that wearable monitoring devices significantly improved the peak VO_2 compared with interventions not using wearable devices [25]. As suggested by a large observational study, a 3.5 ml/kg/min increase in exercise capacity has been associated with an improvement in survival of ≈12% [26]. The increase in the smartwatch group is probably of clinical relevance, and if this improvement could be sustained over time, digital remote HBCR will likely lead to improved patient outcomes.

In our study, the smartwatch intervention also improved anxiety, depressive symptoms, and the quality of life. Depression and anxiety are very common in patients with coronary artery disease and are predictors of poorer outcomes. Additionally, improving quality of life is also a key component in the secondary prevention of coronary artery disease [27]. Notably, the European Society of Cardiology has recommended the use of e-health solutions, including mobile apps, to enhance psychosocial health as part of secondary prevention for patients with coronary artery disease [28]. However, some studies demonstrate that remote digital CR does not seem to improve patients' psychological status. A recent meta-analysis, which included 20 studies and a total of 4535 patients, found improvements in both the physical and mental dimensions of quality of life in the eHealth group. However, the differences in overall scores did not achieve statistical significance (P=.15) [28]. In this study, the personalized, interactive nature of the smartwatch program may have contributed to improved mental health by promoting engagement in regular physical activity and enhancing the patient's sense of social connectedness and support. Additionally, educational resources and real-time guidance may have provided patients with coping mechanisms to deal with the emotional challenges associated with their condition.

Limitations

Despite the promising findings of this feasibility study, several limitations should be considered when interpreting the results. First, the single-center recruitment from Jilin Province (China) may introduce geographic selection bias, as participants shared similar cultural and socioeconomic profiles (ie, urban residents with access to smart devices). Rural populations were underrepresented, potentially inflating adherence rates compared to regions with digital literacy gaps.This selection bias could

XSL•FO
RenderX

affect the applicability of the smartwatch-facilitated HBCR model in more diverse settings. Future multi-center studies involving diverse populations would help validate the applicability of the smartwatch-facilitated HBCR model across different settings. Second, the study did not account for potential confounding factors that could have influenced the outcomes, such as patients' prior experience with digital health technologies, socio-economic status, or comorbidities. These factors may affect both adherence to the HBCR program and the patient's overall health outcomes. A more comprehensive analysis considering such confounders could offer deeper insights into the factors influencing the success of smartwatch-based rehabilitation. Third, in China, family involvement in health care decisions is often more pronounced than in many Western countries, which could have influenced patient adherence to the intervention. Furthermore, the acceptance and familiarity with digital health technologies vary across cultures. While digital health solutions are increasingly popular in urban areas of China, older adults or those in rural regions may face challenges in using such technologies. These factors may limit the generalizability of our findings to other populations. Lastly, the duration of the intervention (3 months) may be too short to fully evaluate the long-term benefits of the smartwatch-based HBCR program. Further research with extended follow-up periods is needed to assess the sustainability of the observed improvements in cardiopulmonary function, anxiety, depression, and quality of life beyond the immediate post-intervention period.

Conclusions

This study demonstrates that a smartwatch-facilitated HBCR model is both feasible and effective for patients with CHD. The intervention significantly improved exercise adherence, cardiopulmonary function, and psychosocial health outcomes, highlighting its potential as an accessible and scalable alternative to traditional center-based cardiac rehabilitation programs. These findings underscore the value of mHealth technologies in overcoming barriers to rehabilitation participation, promoting patient engagement, and improving overall health outcomes. Future studies with larger sample sizes and longer follow-up periods are warranted to further validate these results and explore the broader implications of digital health interventions in cardiac care.

Acknowledgments

We would like to express our sincere gratitude to all participants who took part in this study. Their commitment and cooperation made this research possible. Special thanks to Lepu Medical for their technical support and expertise in the use of the smartwatch devices.

No generative AI tools were used in the writing or preparation of this manuscript.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to patient privacy concerns and institutional data-sharing policies. However, de-identified data may be made available by the corresponding author upon reasonable request and with appropriate institutional approval.

Authors' Contributions

SZ, HW, and XM contributed to conceptualization, validation, writing—review and editing. JW performed data curation. SZ and JW were involved in formal analysis. SZ, HW, and CM led investigation. SZ and JW designed the methodology, handled project administration, and performed visualization. XM and CM were involved in supervision. SZ managed writing–original draft.

Conflicts of Interest

None declared.

Checklist 1
CONSORT e-HEALTH checklist (V 1.6.1).
[PDF File, 1432 KB - mhealth_v13i1e70848_app1.pdf]

References

1. Timmis A, Vardas P, Townsend N, Atlas Writing Group, European Society of Cardiology. Erratum to: European Society of Cardiology: cardiovascular disease statistics 2021. Eur Heart J 2022 Feb 22;43(8):799-799. [doi: 10.1093/eurheartj/ehac064]
2. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics-2023 update: a report from the american heart association. . 2023 Feb 21(8) p. e93-e621. [doi: 10.1161/CIR.0000000000001123] [Medline: 36695182]
3. Stone PH, Libby P, Boden WE. Fundamental pathobiology of coronary atherosclerosis and clinical implications for chronic ischemic heart disease management-the plaque hypothesis: a narrative review. JAMA Cardiol 2023 Feb 1;8(2):192-201. [doi: 10.1001/jamacardio.2022.3926] [Medline: 36515941]

XSL•FO
RenderX

4. Writing Committee Members, Gulati M, Levy PD, Mukherjee D, et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the Evaluation and Diagnosis of CHEST Pain: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol 2021;78(22):e187-e285. [doi: 10.1016/j.jacc.2021.07.053]
5. Knuuti J, Wijns W, Saraste A, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J 2020 Jan 14;41(3):407-477. [doi: 10.1093/eurheartj/ehz425] [Medline: 31504439]
6. Zhang S, Ding R, Chen S, et al. Availability and trend of dissemination of cardiac rehabilitation in China: report from the multicenter national registration platform between 2012 and 2021. Front Cardiovasc Med 2023;10:1210068. [doi: 10.3389/fcvm.2023.1210068]
7. Anderson L, Sharp GA, Norton RJ, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2017 Jun 30;6(6):CD007130. [doi: 10.1002/14651858.CD007130.pub4] [Medline: 28665511]
8. Jolly K, Taylor R, Lip GY, et al. The Birmingham Rehabilitation Uptake Maximisation Study (BRUM). Home-based compared with hospital-based cardiac rehabilitation in a multi-ethnic population: cost-effectiveness and patient adherence. Health Technol Assess 2007 Sep;11(35):1-118. [doi: 10.3310/hta11350] [Medline: 17767899]
9. McKelvie RS, Teo KK, Roberts R, et al. Effects of exercise training in patients with heart failure: the Exercise Rehabilitation Trial (EXERT). Am Heart J 2002 Jul;144(1):23-30. [doi: 10.1067/mhj.2002.123310] [Medline: 12094184]
10. Golbus JR, Gupta K, Stevens R, et al. Abstract 159: understanding baseline physical activity in cardiac rehabilitation enrollees using mobile health technologies. Circ: Cardiovascular Quality and Outcomes 2022 May;15(Suppl_1):e009182. [doi: 10.1161/circoutcomes.15.suppl_1.159]
11. Burke LE, Ma J, Azar KM, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and Stroke Council, et al. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the american heart association. Circulation 2015;132:e233. [doi: 10.1161/CIR.0000000000000335]
12. World health organization. Recommendations on digital interventions for health system strengthening. 2021. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2025-08-18]
13. Moulson N, Bewick D, Selway T, et al. Cardiac rehabilitation during the COVID-19 era: guidance on implementing virtual care. Can J Cardiol 2020 Aug;36(8):1317-1321. [doi: 10.1016/j.cjca.2020.06.006] [Medline: 32553606]
14. Thomas RJ, Beatty AL, Beckie TM, et al. Home-based cardiac rehabilitation: a scientific statement from the American association of cardiovascular and pulmonary rehabilitation, the American heart association, and the American college of cardiology. J Am Coll Cardiol 2019 Jul 9;74(1):133-153. [doi: 10.1016/j.jacc.2019.03.008] [Medline: 31097258]
15. Ge C, Ma J, Xu Y, et al. Predictors of adherence to home-based cardiac rehabilitation program among coronary artery disease outpatients in China. J Geriatr Cardiol 2019 Oct;16(10):749-755. [doi: 10.11909/j.issn.1671-5411.2019.10.003] [Medline: 31700514]
16. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). Med Care 1992 Jun;30(6):473-483. [doi: 10.1097/00005650-199206000-00002]
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
18. Schopfer DW, Whooley MA, Allsup K, et al. Effects of home-based cardiac rehabilitation on time to enrollment and functional status in patients with ischemic heart disease. J Am Heart Assoc 2020 Oct 20;9(19):e016456. [doi: 10.1161/JAHA.120.016456] [Medline: 32954885]
19. Maddison R, Rawstorn JC, Stewart RAH, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan;105(2):122-129. [doi: 10.1136/heartjnl-2018-313189] [Medline: 30150328]
20. Dorje T, Zhao G, Tso K, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019 Nov;1(7):e363-e374. [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]
21. Beswick AD, Rees K, Griebsch I, et al. Provision, uptake and cost of cardiac rehabilitation programmes: improving services to under-represented groups. Health Technol Assess 2004 Oct;8(41):iii-iiv. [doi: 10.3310/hta8410] [Medline: 15461879]
22. McDonald MT, Siebert S, Coulter EH, McDonald DA, Paul L. Level of adherence to prescribed exercise in spondyloarthritis and factors affecting this adherence: a systematic review. Rheumatol Int 2019 Feb;39(2):187-201. [doi: 10.1007/s00296-018-4225-8] [Medline: 30649569]
23. Danbjørg DB, Villadsen A, Gill E, Rothmann MJ, Clemensen J. Usage of an exercise app in the care for people with osteoarthritis: user-driven exploratory study. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e11. [doi: 10.2196/mhealth.7734] [Medline: 29326092]
24. Hannan AL, Harders MP, Hing W, Climstein M, Coombes JS, Furness J. Impact of wearable physical activity monitoring devices with exercise prescription or advice in the maintenance phase of cardiac rehabilitation: systematic review and meta-analysis. BMC Sports Sci Med Rehabil 2019;11:14. [doi: 10.1186/s13102-019-0126-8] [Medline: 31384474]
25. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. N Engl J Med 2002 Mar 14;346(11):793-801. [doi: 10.1056/NEJMoa011858] [Medline: 11893790]

XSL•FO
RenderX

26. Abreu A, Frederix I, Dendale P, et al. Standardization and quality improvement of secondary prevention through cardiovascular rehabilitation programmes in Europe: The avenue towards EAPC accreditation programme: A position statement of the Secondary Prevention and Rehabilitation Section of the European Association of Preventive Cardiology (EAPC). Eur J Prev Cardiol 2021 May 14;28(5):496-509. [doi: 10.1177/2047487320924912] [Medline: 33611459]
27. Frederix I, Caiani EG, Dendale P, et al. ESC e-cardiology working group position paper: overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiolog 2019 Jul;26(11):1166-1177. [doi: 10.1177/2047487319832394]
28. Cruz-Cobo C, Bernal-Jiménez M, Vázquez-García R, Santi-Cano MJ. Effectiveness of mHealth interventions in the control of lifestyle and cardiovascular risk factors in patients after a coronary event: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Dec 2;10(12):e39593. [doi: 10.2196/39593] [Medline: 36459396]

Abbreviations

AT: anaerobic threshold
CEPT: cardiopulmonary exercise testing
CHD: coronary heart disease
CR: cardiac rehabilitation
GAD-7: Generalized Anxiety Disorder 7-item scale
HBCR: home-based cardiac rehabilitation
HETAQ: home-based exercise training adherence questionnaire
MET: peak metabolic equivalent
PHQ-9: Patient Health Questionnaire 9-item scale
VCO_2: carbon dioxide output
VE: minute ventilation
VE: minute ventilation
VO_2: oxygen consumption

Edited by L Buis, M Dorsch; submitted 05.01.25; peer-reviewed by A Jamal, T Guetterman, Z Liang; revised version received 06.07.25; accepted 30.07.25; published 18.09.25.

Please cite as:
Zhang S, Wang Y, Wu J, Ma C, Meng X
Effectiveness of Smartwatch Device on Adherence to Home-Based Cardiac Rehabilitation in Patients With Coronary Heart Disease: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e70848
URL: https://mhealth.jmir.org/2025/1/e70848
doi:10.2196/70848

© Sisi Zhang, Yuehui Wang, Jiahui Wu, Changsheng Ma, Xiaoping Meng. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effect of Lifestyle Modification Through Web-Based Telerehabilitation Monitoring Combined With Supervised Sensorimotor Training After Total Knee Arthroplasty: Randomized Controlled Trial

Samreen Sadiq[1*], BSPT, tDPT, MPhil; Rabiya Noor[1*], BSPT, PPDPT, PhD; Rizwan Akram[2*], MBBS

[1]Department of Physical Therapy, Riphah College of Rehabilitation and Allied Health Sciences, Riphah International University, 25 Raza Saeed Rd, Bhabra Block M Gulberg III, Lahore, Pakistan

[2]Department of Orthopedic and Spine Center, Ghurki Trust Teaching Hospital, Lahore, Pakistan

[*]all authors contributed equally

Corresponding Author:
Rabiya Noor, BSPT, PPDPT, PhD
Department of Physical Therapy, Riphah College of Rehabilitation and Allied Health Sciences, Riphah International University, 25 Raza Saeed Rd, Bhabra Block M Gulberg III, Lahore, Pakistan

Abstract

Background: Total knee arthroplasty (TKA) is commonly performed to manage end-stage knee osteoarthritis, yet postsurgical recovery varies significantly among patients. Lifestyle modification and rehabilitation interventions play a critical role in optimizing outcomes. While telerehabilitation has shown promise in enhancing accessibility and compliance, its role in supporting lifestyle behavior change alongside supervised sensorimotor training remains underexplored.

Objective: This study aimed to evaluate the effects of a home-based lifestyle modification program delivered through web-based telerehabilitation monitoring in addition to supervised sensorimotor training, in improving physical function, pain, balance, quality of life (QOL), and adherence in patients undergoing TKA.

Methods: A single-blinded randomized controlled trial was conducted among 52 participants undergoing primary TKA, who were randomly assigned to either the intervention group (IG) (supervised sensorimotor training plus a telerehabilitation-supported lifestyle modification program) or the control group (CG) (supervised sensorimotor training alone and a traditional home exercise plan). The intervention lasted 22 weeks, and participants were assessed at baseline (presurgery), 14 weeks, and 22 weeks postsurgery. Outcome measures included joint position sense (JPS), musculoskeletal ultrasound of the rectus femoris muscle, Berg Balance Scale, knee function using the Knee Injury and Osteoarthritis Outcome Score, and QOL via EuroQol 5-dimension 5-level questionnaire.

Results: Significant improvements were observed in the IG across all outcomes compared with the CG. Notably, the IG showed greater improvements in musculoskeletal ultrasound thickness. JPS showed superior accuracy in the experimental group (baseline [3.2 degrees] to 22 wk postsurgery [0.05 degrees]) compared with the CG (baseline [3.1 degrees] to 22 wk postsurgery [1.8 degrees]), with significant improvements noted (P=.001, Cohen d=3.1 vs 0.7), Knee Injury and Osteoarthritis Outcome Score subscales (pain, symptoms, activities of daily living, sport, and QOL), and JPS (mean absolute error 0.05 vs 1.8 degrees). Berg Balance Scale demonstrated significant gains in balance for the experimental group (baseline [34] to 22 wk postsurgery [53]) relative to the CG (baseline [37] to 22 wk postsurgery [48]), with substantial differences observed (P=.001, Cohen d=1.8 vs 0.4). The EuroQol 5-dimension 5-level questionnaire health-related QOL scores were markedly higher for the experimental group (baseline [45.4] to 22 wk postsurgery [88.1]) compared with the CG (baseline [42.8] to 22 wk postsurgery [70.9]), indicating substantial gains in overall health status (P=.001, Cohen d=2.4 vs 1.3). The IG also reported higher compliance, with 81.8% (18/22) achieving over 90% adherence compared with 68.18% (15/22) in the CG.

Conclusions: Home-based lifestyle modification program through telerehabilitation monitoring significantly improved functional and patient-reported outcomes in individuals following TKA. These findings support the integration of lifestyle modification programs through telerehabilitation monitoring into post-TKA recovery pathways to optimize rehabilitation outcomes.

Trial Registration: Clinical Trials.gov NCT05018494, https://clinicaltrials.gov/study/NCT05018494

(JMIR Mhealth Uhealth 2025;13:e64643) doi:10.2196/64643

XSL•FO
RenderX

KEYWORDS

prescribed home-based program; supervised training; total knee replacement; telerehabilitation; web based portal; monitoring; lifestyle modification; total knee arthroplasty; knee; knee osteoarthritis; osteoarthritis; low back pain; Asia; quality of life; digital rehabilitation; rehabilitation; single-center; randomized controlled trial; mobile health; telehealth; digital health

Introduction

Knee osteoarthritis (OA) and low back pain are prevalent rheumatic disorders in Asian regions, posing significant challenges to patients and health care systems [1]. Total Knee Arthroplasty (TKA) is widely recognized as the gold standard management strategy for relieving symptoms associated with degenerative knee arthritis [2]. Rehabilitation following TKA plays a crucial role in improving functional performance and quality of life (QOL) for individuals in this population. However, access to rehabilitation services, health care facilities, and follow-up appointments can be limited, particularly in regions with resource constraints or remote areas [3]. Telerehabilitation, a branch of telehealth, presents a potential solution by using technology to deliver monitoring and rehabilitation services to patients from a distance, ensuring adequate follow-up and support [4].

One important aspect of post-TKA rehabilitation is sensorimotor training, which aims to improve neuromuscular control, balance, and coordination to restore functional abilities and reduce the risk of falls. Sensorimotor training focuses on optimizing central nervous system function and promoting appropriate muscular firing patterns for joint stability [5]. By incorporating sensorimotor training into telerehabilitation programs, patients can receive personalized exercises and real-time feedback from health care professionals, even when they are at home. Another crucial component missing from many post-TKA rehabilitation programs is lifestyle modification. Lifestyle modifications encompass a range of interventions that focus on promoting healthy behaviors and habits to improve overall well-being [6]. In the context of post-TKA rehabilitation, incorporating lifestyle modification programs into telerehabilitation monitoring can have several benefits for patients. Providing patients with educational resources and information about lifestyle factors that can impact their joint health is essential. Telerehabilitation can offer interactive educational materials, including videos, articles, and digital sessions, to educate patients about the importance of maintaining a healthy lifestyle. Topics may include proper nutrition, weight management, joint protection techniques, and strategies to reduce joint stress in daily activities [7]. Regular exercise is a cornerstone of lifestyle modification and plays a crucial role in managing joint health after TKA. Telerehabilitation platforms can provide guided exercise programs tailored to individual patients, including specific exercises targeting muscle strengthening, flexibility, and cardiovascular fitness. Through video demonstrations and remote monitoring, health care professionals can ensure that patients perform exercises correctly and modify the program as needed to accommodate their progress [8]. Nutrition plays a vital role in joint health and overall well-being. Telerehabilitation can integrate dietary guidance by providing personalized nutrition plans and recommendations [9].

Excessive body weight can contribute to increased joint stress and deteriorate the outcomes of TKA. Telerehabilitation monitoring can include weight management programs that track patients' weight, provide nutritional guidance, and offer support in setting realistic weight loss goals [10]. Lifestyle modifications through telerehabilitation can address behavior change techniques to help patients adopt and maintain healthy habits. This may include setting goals, self-monitoring techniques, problem-solving strategies, and relapse prevention [11].

Integrating lifestyle modification programs into telerehabilitation monitoring for post-TKA patients can have a comprehensive impact on their overall outcomes [12]. By addressing education, exercise, dietary interventions, weight management, and behavior change, patients can experience improved joint health, reduced joint stress, better functional outcomes, enhanced QOL, and increased long-term success of the TKA procedure. Furthermore, telerehabilitation provides a convenient and accessible platform for patients to receive ongoing support and guidance, promoting adherence to lifestyle modifications even after formal rehabilitation sessions are completed [13]. Patients can access resources remotely, receive reminders and encouragement, and engage in digital consultations to address any concerns or challenges they may face. While digital rehabilitation solutions are well-established in developed nations, there is a pressing need to introduce technological advancements in countries like Pakistan. Telerehabilitation, if incorporated into standard therapy, has the potential to benefit the community by reducing dependence on limited human resources while ensuring better clinical outcomes [14]. By embracing telerehabilitation, health care providers in Pakistan can overcome geographical barriers, increase access to rehabilitation services, and enhance patient care in a cost-effective manner.

TKA is commonly followed by structured rehabilitation programs to restore joint function and QOL. Though various rehabilitation protocols, including supervised sensorimotor training, have been extensively studied, a vital component termed lifestyle modification in the postoperative period remains principally overlooked. Current literature lacks the incorporation of home-based lifestyle interventions that address patients' daily habits, activity levels, and long-term behavioral changes, which are crucial for continual recovery and enhanced outcomes after TKA. This study aims to fill this critical gap by evaluating the combined effect of supervised sensorimotor training and structured lifestyle modification, delivered through telerehabilitation monitoring. The significance of this study lies in its unique approach to combine digitally monitored physical rehabilitation with behavior-driven home-based strategies, targeting not only physical but also the psychosocial and practical facets of recovery. This integrated model could offer a more holistic and sustainable rehabilitation pathway, particularly in underserved settings.

XSL•FO
RenderX

Therefore, the main aim of our study was to evaluate the effects of lifestyle modifications through telerehabilitation monitoring combined with supervised sensorimotor training on various outcomes, including joint position sense (JPS), balance, muscle architecture, knee joint function, and QOL among patients with TKA.

Methods

Study Setting and Design

We conducted a single-centered randomized controlled trial with parallel groups between 2021 and 2023 aimed to compare the effect of a home-based lifestyle modification program delivered through telerehabilitation monitoring, combined with supervised sensorimotor training for patients undergoing elective TKA for unilateral OA. This study received approval from the Research and Ethics Committee at Riphah College of Rehabilitation and Allied Health Sciences, Riphah International University, Lahore, Pakistan (REC/RCR&AHS/2033) and was conducted in accordance with the Declaration of Helsinki [15]. The protocol of the study has been officially registered with the clinical study identifier NCT05018494 in the ClinicalTrials.gov Protocols Registration and Results System on May 8, 2021. The participating center in this study was the Physical Therapy department of Ghurki Trust Teaching Hospital, Lahore, which was following the rehabilitation protocol for patients post-knee replacement.

Sample Size

Sample size calculation was conducted using G*Power (version 3.1.9.4; Heinrich Heine University Düsseldorf) based on the Knee Injury and Osteoarthritis Outcome Score (KOOS) pain dimension, which had demonstrated a large effect size (Cohen *d*=2.0) in previous literature [16], demonstrating substantial differences between the intervention groups (IGs). To ensure adequate power to detect significant changes, the calculation was performed with a significance level (α) of .05 (2-tailed) and a statistical power of 80%. The pooled SD for the KOOS pain dimension was calculated as 11.26, and the mean difference between the experimental and control groups (CGs) was 13.4 points. Based on these parameters, the required sample size was calculated to be 26 participants per group, resulting in a total sample size of 52 participants. This sample size was sufficient to detect meaningful differences in pain outcomes between the groups. Moreover, considering potential participant dropout rates of approximately 10% - 15%**,** a total of 52 participants were estimated to be adequate to maintain the study's statistical power.

Study Participants

Eligibility for the study was assessed for all patients who had received an anesthesiologist's approval and were already scheduled for surgery. Patients were invited to participate in the study, and detailed oral and written explanations were provided about the trial. Those willing to participate were required to sign a written informed consent form. Prior to admission to the hospital, patients completed preoperative outcome assessment questionnaires, including the knee joint function, Berg Balance Scale (BBS), QOL questionnaire, along with objective measures of JPS and cross-sectional area of rectus femoris (RF) through diagnostic musculoskeletal ultrasound (MSK-US).

The inclusion criteria were (1) patients who were men or women, (2) had undergone primary TKA surgery, (3) had the ability to walk (with a walking aid or unaided), (4) were aged between 45 and 75 years, (5) had active knee flexion of 80° and active knee extension of −10° upon discharge, and (6) had the availability of a smartphone and internet service in the residing area.

The exclusion criteria were (1) the presence of health-related medical conditions that could interfere with tests or the rehabilitation program, (2) neurological conditions and vestibular disorders that might affect balance, (3) inability to attend rehabilitation services, revised knee arthroplasty, and (4) blindness and any condition incompatible with 30 minutes of light to moderate physical activity.

Randomization, Group Allocation, and Blinding

Participants were randomized to groups on a 1:1 basis through computer-generated randomization performed by an independent statistician. Following the initial assessment, a volunteer rehabilitation house officer distributed envelopes based on the randomization sequence. These envelopes, which were consecutively numbered, sealed, and opaque, were stored in a secure location accessible only to the unblinded lead researcher-physiotherapist (SS). This physiotherapist (SS) then scheduled the participants' training session to introduce the web-based rehabilitation lifestyle modification program. Participants reviewed and signed the informed consent form, which adhered to the guidelines of the Helsinki Declaration and outlined the study's objectives.

Intervention and Control Group

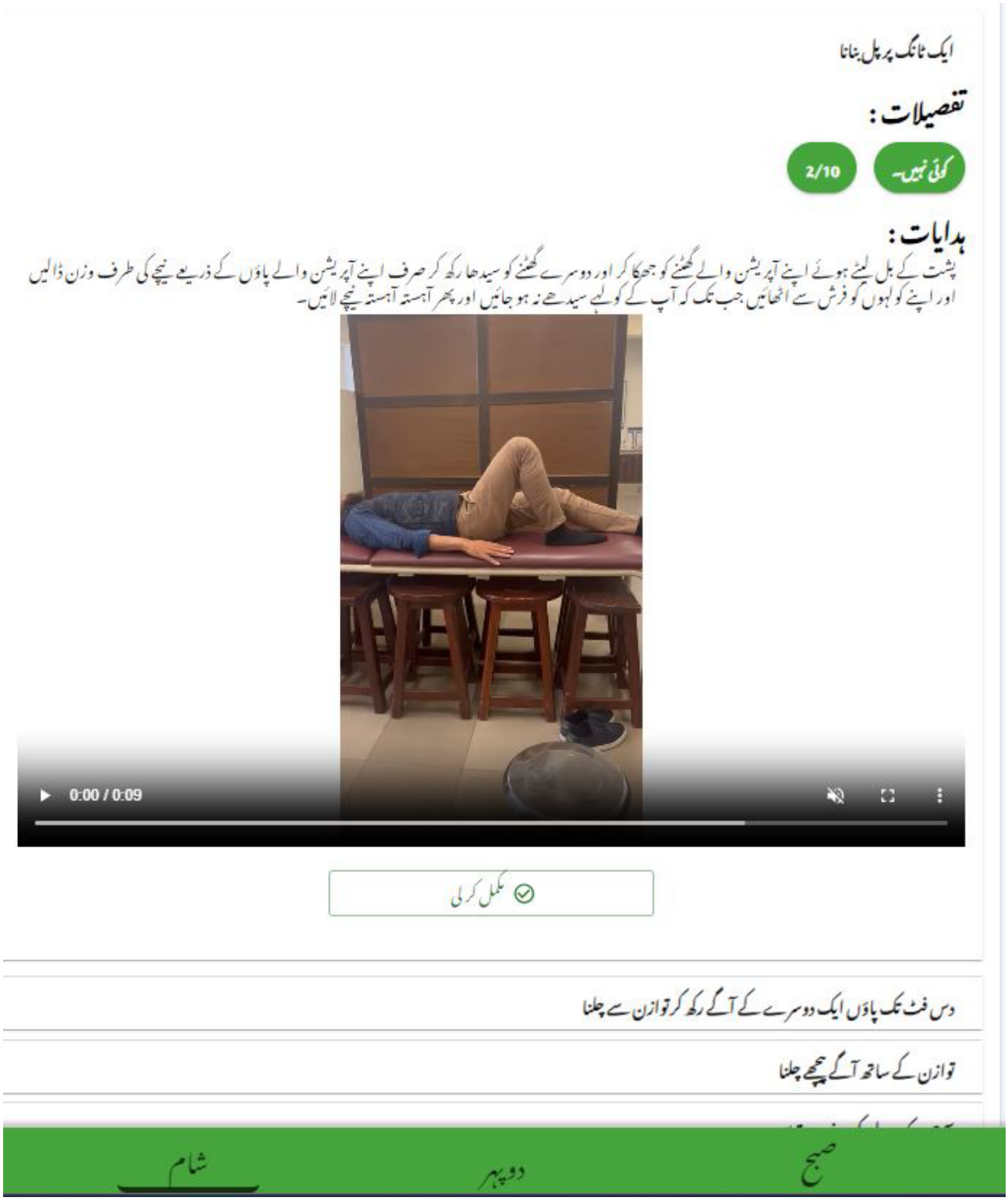
Once included in the study, post 2 weeks of knee replacement, patients were randomized into either the IG or the CG. The randomization process was blinded to patients. Patients in the IG and CG received one-on-one 12 weeks of supervised strengthening exercises, neuromuscular electrical stimulation, and sensorimotor training, 3 times a week for 45 minutes to learn strength and sensory-motor exercises. Patients who received any rehabilitation limitations from the surgeon were excluded from the study. At the time of discharge, patients in IG received a web portal link that could be viewed on any smartphone or device with a complete manual on lifestyle modification, constituting 3 sections (Figure 1). The first section included exercises comprising videos of morning, afternoon, and evening, along with clear instructions in the local Urdu language for better understanding (Figure 2). The second section was about a daily diet plan of morning, afternoon, mid-evening, and night meals, considering the optimal caloric balance (Figure 3). The third section was the education section, giving instruction points about safety and daily activities. The CG patients received standard written instructions commonly prescribed as home exercises with oral guidance.

XSL•FO
RenderX

Figure 1. Digital lifestyle modification process.

XSL•FO
RenderX

Figure 2. Exercise with instructional videos and Urdu guidance.



XSL•FO
RenderX

Figure 3. Diet plan view.

XSL•FO
RenderX

Regular messages and phone calls every week were conducted to follow up with all patients, addressing any possible complications, completion of the training diary, and exercise clarity. The patients had scheduled hospital visits at 3 and 5 months. During the follow-up visit, the physical therapist performed assessments, and patients were asked to fill out questionnaires. Postoperative cross-sectional area of the RF muscle was taken to compare the muscle architecture presurgery, after sensorimotor training, and post home-based exercises. Three readings were taken for all the measures (presurgery, 12 weeks after sensorimotor training, and a further 8 weeks after home-based lifestyle modification implementation). At follow-up, the physiotherapist inquired about any adverse effects or reasons for training cessation from the patients. This comprehensive study design allowed us to assess the effectiveness, safety, and long-term outcomes of a home-based lifestyle modification manual delivered through telerehabilitation monitoring compared with the standard home-based exercises for patients undergoing elective TKA.

Participants maintained compliance through their portal by checking off their exercise, educational, and dietary activities. Any adverse events were promptly reported to the therapist via the WhatsApp button available on the portal. Exercise assurance videos were also sent through WhatsApp for additional guidance. They recorded each exercise performed daily, made notes about any additional sports activities such as walking or stationary bicycling, documented any adverse effects experienced, and provided reasons for skipping any program element if applicable. Participants also used the Visual Analog Scale (VAS) to record their pain levels throughout the day, before and after training sessions, and at night.

Standard Rehabilitation

All patients received standard physiotherapy care during their hospitalization. This began on the day of surgery and included mobilization using a walking aid (typically 2 crutches, occasionally a walker), exercises to prevent deep vein thrombosis, lower limb range of motion exercises, and an isometric strengthening program. Patients were advised to perform these exercises twice daily, aiming for approximately 10 repetitions of each exercise, and to walk as much as possible. The exercises were conducted in various positions, including supine, on the healthy side, on the abdomen, sitting, and standing (eg, buttock squeezes, leg sliding motions, straight leg raises, bridges, and postural exercises).

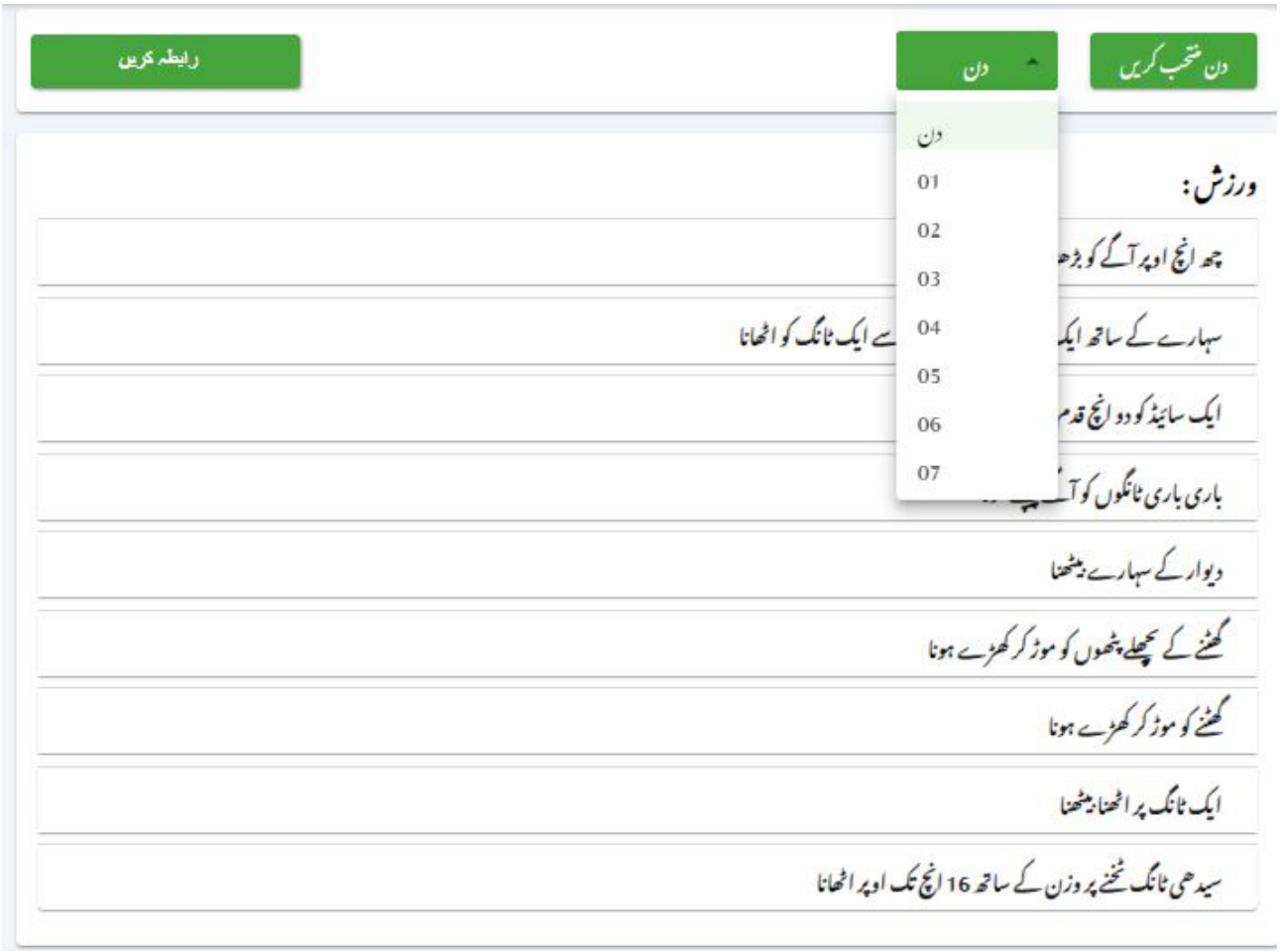
The treatment protocol involved a structured regimen starting with 2 weeks of hospital-based bedside and preparatory physiotherapy. This was followed by 12 weeks of supervised strengthening exercises, neuromuscular electrical stimulation, and sensorimotor training, conducted 3 times a week for 45 minutes each session. Additionally, an 8-week lifestyle modification program encompassing education, exercise, and diet was provided to the IG. Common treatments consisted of strengthening exercises, neuromuscular electrical stimulation, and sensorimotor training, all carried out over a 12-week period with sessions 3 times a week for 45 minutes. Each session included routine elements such as warm-up activities involving range-of-motion exercises, stretching, or flexibility exercises; mobility exercises for the lower extremities; and strengthening exercises, including isometric and concentric exercises for the knee quadriceps, hamstrings, and hip abductors. Functional, task-oriented exercises such as stair climbing, squats, stationary cycling, treadmill walking, or regular walking were also part of the regimen, followed by cool-down activities. Concentric mode strengthening exercises were performed using a medium-resistance elastic theraband provided to all patients. The primary focus of the exercises was enhancing sensorimotor functioning. Innovative agility and perturbation training techniques were incorporated, such as side-stepping, backward stepping, and using irregular foam laminations to offer targeted sensorimotor stimulation over unstable surfaces [17]. Exercise challenges and progression were achieved using regular pillows to simulate unstable surfaces, plastic cups to create obstacles, and strategies such as transitioning from bipedal to monopedal stance and from eyes open to eyes closed to increase the difficulty of maintaining or achieving balance. Patients in the CG were given written instructions, handouts with pictorial exercises, and a portal login that provides access to exercise videos along with a checklist to mark daily exercise compliance.

Prescribed Digital Home-Based Lifestyle Modification Program

Before discharge, physiotherapists instructed patients in the IG on specialized strength and sensorimotor exercises using various exercise aids along with educational points and a diet plan. These patients also received a web portal login containing exercise videos, written instructions, a diet plan with pictorial representation, and a checklist to track their daily lifestyle routine (Figure 4). The new training protocol included exercises designed to enhance hip stability and decrease local stress on the knee prosthesis. Safety considerations regarding knee, hip, and lumbar spine load; fall prevention; and prosthesis dislocation were integral to the training design [18]. The program encompassed hip and knee muscle strengthening exercises (emphasizing the abductors), hip and pelvic stabilization exercises, ankle and knee muscle strengthening exercises (to better dissipate impact forces and control femoral internal rotation), and trunk muscle strengthening exercises (to stabilize the pelvis and lumbar spine, thereby reducing local loads). The exercises were designed to be simple and easy to understand, with home exercise videos that are straightforward to follow. Additionally, the training program avoided the need for expensive equipment. Focusing on education, physical activity, dietary modifications, weight management, and behavioral changes might lead to better joint health, decreased joint stress, improved functional outcomes, enhanced QOL, and increased long-term success of the TKA procedure. The detailed lifestyle modification plan was attached in the Multimedia Appendix 1.

XSL•FO
RenderX

Figure 4. Select day and time.



Outcome Measures

The primary objective of the study was to evaluate the effectiveness of the chosen protocol, which was assessed through JPS, muscle thickness of the RF, and balance. Secondary outcomes included knee function, and QOL was assessed through patient questionnaires, as well as consistent reporting of any adverse event.

Knee Joint Position Sense

Knee JPS was assessed using a reliable and validated method, incorporating measurements of joint angles through digital photography and nonreflective markers, supported by specialized software. The JPS assessment was performed in a quiet, isolated room. Patients were seated in a high sitting position on a treatment plinth, and testing was conducted on the operated knee. Four square nonreflective markers, each 4 cm in diameter, were placed at specific anatomical landmarks on the lateral side of the limb: First, proximal to a quarter of the distance along a line from the greater trochanter to the lateral joint line of the knee, second, over the neck of the fibula, third over the proximal part of the lateral malleolus, and fourth on the iliotibial tract adjacent to the superior border of the patella. With eyes closed, the patient's knee was passively moved by the examiner to one of 2 target reference angles, 30° or 60° of knee extension, and held in that position for approximately 5 seconds to allow the patient to internalize the joint position. The limb was then returned to the resting position for 7 seconds. The patient was subsequently asked to actively reproduce the same knee angle. Each angle was tested 3 times, and the average value of the reproduced angles was calculated. A side-view photograph was taken for each trial using an Apple iPhone 14 Pro Max (48 MP camera), positioned 185 cm from the side of the patient and 65 cm above the floor. The angles formed by the markers were analyzed using Paint software by drawing lines between points to measure joint angles, and final values, including the absolute angular error, were calculated and documented in Microsoft Excel. Absolute error was defined as the absolute difference between the target and perceived angles for each of the 3 trials at specified angles for each limb of each participant. Relative error indicates the variability of errors between trials, reflecting the consistency of proprioceptive performance [19].

Muscle Thickness of Rectus Femoris

MSK-US was used using a curvilinear probe operating at frequencies ranging from 7.1 to 14 MHz, along with a 55 mm linear probe. The probe was positioned ventrally in the transverse plane, perpendicular to the skin over the RF muscle. The RF muscle's location was marked at a point 3/5 of the distance from the anterior superior iliac spine to the superior border of the patella, ensuring that the entire cross-section of the RF is visible within a single field for all patients [20].

Imaging was conducted with patients seated. Gentle contraction and relaxation maneuvers were used to enhance visualization of muscle septa before capturing images. Cross-sectional area,

XSL•FO
RenderX

muscle thickness, and width were measured at this specific muscle point, with images taken both in complete relaxation and during contraction.

Balance

The original 14-item BBS was used to evaluate balance in older individuals. This scale has been validated, proven reliable, and shown to be responsive in assessing balance. It comprises 14 straightforward balance tasks, each scored on a 5-point ordinal scale (0, 1, 2, 3, and 4), with a maximum possible score of 56. Higher scores on the BBS indicate better balance. This assessment tool is widely used clinically, with all tasks focusing on balance, and has previously been used to evaluate balance in patients recovering from hip fractures and knee arthritis [21].

Knee Function

Knee function was assessed using the KOOS. Originating from the WOMAC Osteoarthritis Index in 1995, the KOOS is a reliable instrument for evaluating symptoms and function in individuals with knee injuries or OA. This self-administered questionnaire takes approximately 10 minutes to complete and encompasses 5 subscales: pain, symptoms other than pain, function in activities of daily living (ADL), function in sports and recreation (Sport/Rec), and knee-related QOL. Scores are reported as percentages ranging from 0 to 100. This comprehensive tool enables both short-term and long-term assessment of various knee conditions, including OA. The Urdu version of the KOOS was used for this evaluation [22].

Quality of Life

The EuroQol 5-dimension 5-level questionnaire (EQ-5D-5L) was used to assess health-related quality of life (HRQOL). Its conceptual framework takes a holistic approach to health, encompassing not only the medical definition but also the essential aspects of independent physical, emotional, and social functioning. The EQ-5D-5L considers health as a composite of positive elements (well-being) and negative aspects [23]. It comprises a questionnaire and a VAS (EuroQol VAS), where the EuroQol VAS allows individuals to rate their own current overall health status using a VAS [24]. This scale enables continuous monitoring of health changes over time. The self-assessment questionnaire involves the participant describing their current health status across 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Participants are asked to assess their level of function in each dimension as severe, moderate, or none.

Telerehabilitation Monitoring

Participants maintained compliance through their portal by checking off their exercise, educational, and dietary activities. Any adverse events were promptly reported to the therapist via the WhatsApp button available on the portal. Exercise assurance videos were also sent through WhatsApp for additional guidance. They recorded each exercise performed daily, made notes about any additional sports activities such as walking or stationary bicycling, documented any adverse effects experienced, and provided reasons for skipping any program element if applicable. Participants also used the VAS to record their pain levels throughout the day, before and after training sessions, and at night.

All of this information was analyzed to evaluate patient satisfaction, assess the feasibility of the protocols, and ensure their safety.

Statistical Analysis

All data was collected and recorded in hard copy format by the primary author, who bears responsibility for maintaining data confidentiality. Check-up forms were printed and completed preoperatively and at each subsequent check-up by the physiotherapist and orthopedic surgeon. Patients also filled out printed questionnaires during these sessions. The author entered the anonymized data into tables for statistical analysis. Statistical processing was conducted using IBM SPSS Statistics (version 25.0; IBM Corp).

An intention-to-treat analysis was performed to mitigate potential biases arising from missing data in randomized controlled trials. This analytic approach involved analyzing participants based on their original randomized treatment assignment, regardless of whether they completed the treatment as intended. Sensitivity analyses were conducted and reported to address any missing data [25].

Descriptive statistics were presented as mean (SD), and intergroup differences in baseline characteristics will be assessed using an independent *t* test. The normality of the sampling distribution was evaluated using the Shapiro-Wilk W-test for inferential analysis [26]. Assumptions of sphericity (Mauchly test) and homogeneity of variances (Levene test) were checked prior to analysis. Repeated Measures ANOVA was used to compare mean score changes both between and within groups to evaluate the effects of the intervention. The significance level for all analyses was set at $P<.050$.

Mixed effects ANOVA assessed the impact of Group (experimental and control), Time (baseline, 14 wk, and 22 wk), and their interaction on each dependent variable. Effect sizes were computed using partial eta squared (η^2p), with values between 0.1 and 0.24 indicating a moderate effect, values from 0.25 to 0.36 indicating a medium effect, and values of 0.37 or higher indicating a large effect [27]. To control for potential inflation of Type I error due to multiple outcome measures, the Holm-Bonferroni correction and Benjamini-Hochberg (FDR) methods were applied to *P* values of all outcomes.

Ethical Considerations

This study was approved by the Research and Ethics Committee at Riphah College of Rehabilitation and Allied Health Sciences, Riphah International University, Lahore, Pakistan (REC/RCR&AHS/2033), and was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. The study protocol was officially registered with ClinicalTrials.gov (Identifier: NCT05018494) on May 8, 2021. Written informed consent was obtained from all participants prior to their inclusion in the study. Participants were fully informed about the study's purpose, procedures, potential risks, and benefits. To maintain privacy and confidentiality, all collected data were anonymized and securely stored, with access to the telerehabilitation platform restricted to authorized personnel through password-protected accounts. No financial compensation was provided for participation, and all

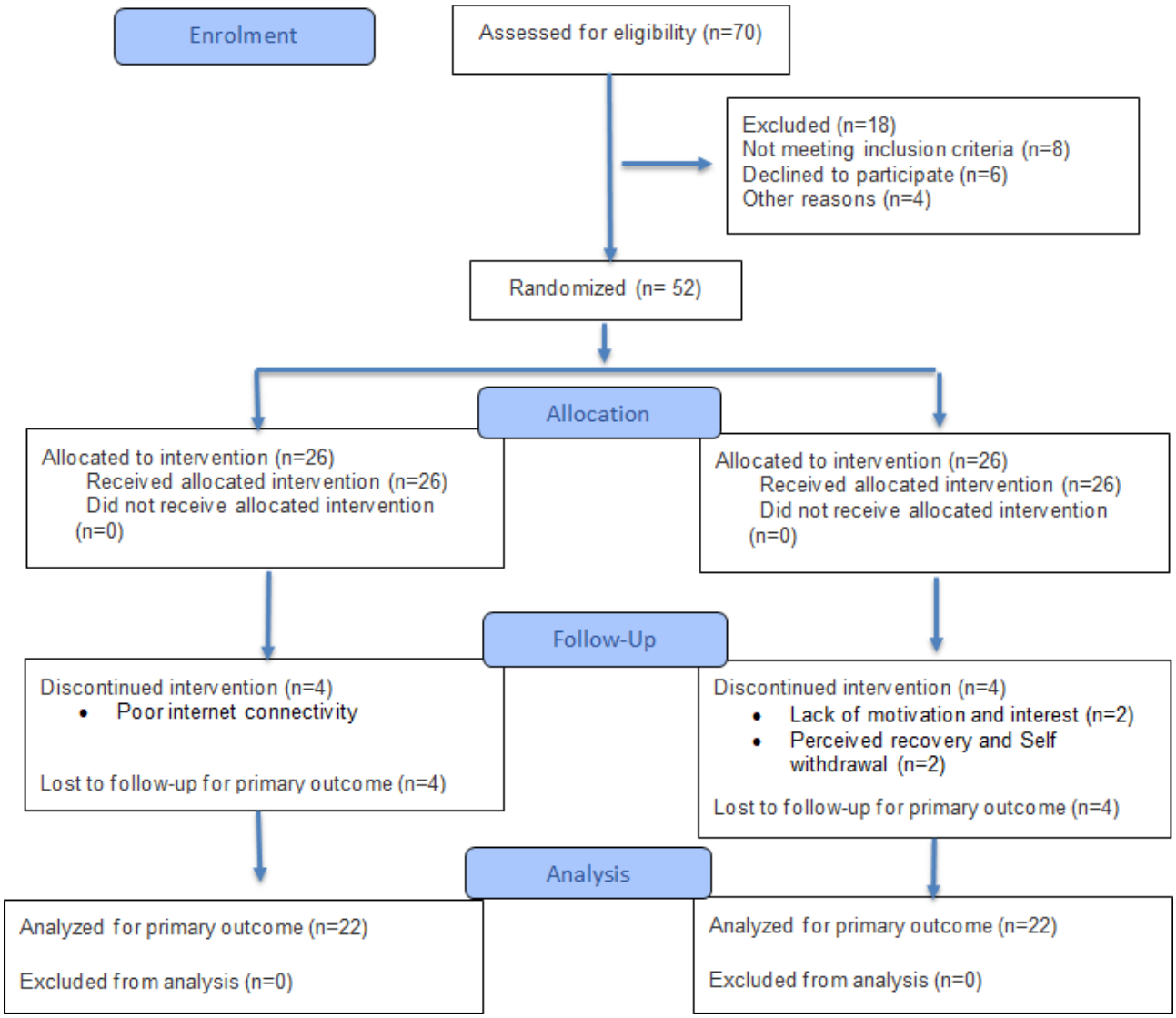
XSL•FO
RenderX

rehabilitation services, materials, portal logins, and digital access were offered free of charge.

Results

Out of 70 participants initially assessed for eligibility, 52 were randomly allocated—26 to the IG and 26 to the standard care group. However, due to participant dropouts, the final assessment comprised 22 individuals from the experimental group and 22 from the CG. The different stages of the study were outlined in the CONSORT flowchart presented in Figure 5. The data in Table 1 illustrate the baseline characteristics of both the interventional group (n=22) and the CG (n=22). Statistically significant differences were not observed between the two groups (P>.50), suggesting that both groups were well-aligned at baseline.

Figure 5. Consort flow diagram.



XSL•FO
RenderX

Table . Participants' demographic characteristics.

Characteristics	Experimental group (n=22)	Control group (n=22)	*P* value
Age (years), mean (SD)	64.1 (5.4)	63.5 (5.5)	.63
Body mass (kg), mean (SD)	74.5 (4.8)	75.0 (5.5)	.98
Height (m), mean (SD)	1.68 (0.1)	1.64 (0.1)	.54
Sex, n (%)			.62
Female	15 (68.1)	16 (72.7)	
Male	7 (31.9)	6 (27.3)	
Number of falls presurgery (within 1 year), mean (SD)	1.4 (0.6)	1.6 (0.7)	.88
NPRS[a] (cm), mean (SD)	7.35 (1.1)	7.04 (1.2)	.72

[a]NPRS: Numeric Pain Rating Scale.

Table 2 shows inter-group comparisons of variables at baseline, 14th week, and 22nd week. The experimental group in this study demonstrated significant improvements across a range of outcome measures compared with the CG, reflecting substantial benefits in functional recovery and HRQOL. JPS showed superior accuracy in the experimental group (baseline [3.2 degrees] to 22 wk postsurgery [0.05 degrees]) compared with the CG (baseline [3.1 degrees] to 22 wk postsurgery [1.8 degrees]), with significant improvements noted (*P*=.001, Cohen *d*=3.1 vs 0.7). MSK-US measurements showed marked improvements in the experimental group (baseline [450.3 units] to 22 wk postsurgery [663.9 units]) compared with the CG (baseline [422.3 units] to 22 wk postsurgery [562.3 units]), indicating enhanced musculoskeletal health with a large effect size (Cohen *d*=3.7 vs 2.0, *P*=.001). Moreover, assessments using the BBS demonstrated significant gains in balance for the experimental group (baseline [34] to 22 wk postsurgery [53]) relative to the CG (baseline [37] to 22 wk postsurgery [48]), with substantial differences observed (*P*=.001, Cohen *d*=1.8 vs 0.4). The KOOS revealed consistent improvements across its various subscales (Pain, Symptoms, ADL, Sports/Recreation, QOL) for the experimental group, reflecting better knee function and QOL outcomes compared with the CG (all *P*=.001, Cohen *d* ranging from 1.6 to 4.9). Last, the EQ-5D-5L HRQOL scores were markedly higher for the experimental group (baseline [45.4] to 22 wk postsurgery [88.1]) compared with the CG (baseline [42.8] to 22 wk postsurgery [70.9]), indicating substantial gains in overall health status (*P*=.001, Cohen *d*=2.4 vs 1.3). Post-hoc correction for multiple comparisons using Holm-Bonferroni and Benjamini-Hochberg (FDR) methods was conducted, and all outcomes remained statistically significant after adjustment (*P*<.05) as shown in Multimedia Appendix 2. In conclusion, these findings exhibited greater improvements in the experimental group receiving supervised sensorimotor training with a lifestyle modification program than the CG, indicating potential benefits.

XSL•FO
RenderX

Table . Group-wise comparisons at baseline, 14 weeks postsurgery, and 22 weeks postsurgery.

Variable and group	Baseline (presurgery), mean (SD)	14 weeks post-surgery, mean (SD)	22 weeks post-surgery, mean (SD)	*P* value	*F* test (*df*)	Effect size	Cohen *d*
JPS[a] (absolute error)				.001	7.1 (1, 42)	3.1	0.7
Intervention	3.2 (0.5)	1.8 (0.7)	0.05 (0.2)				
Control	3.1 (0.3)	2.1 (0.2)	1.8 (0.2)				
MSK-US[b]				.001	19.6 (1, 42)	3.7	2.0
Interventional	450.3 (52.4)	556.8 (53.6)	663.9 (60.2)				
Control	422.3 (51.5)	497.9 (48.5)	562.3 (42.9)				
BBS[c]				.001	9.6 (1, 42)	1.8	0.4
Interventional	34 (5)	48 (3)	53 (3)				
Control	37 (4)	45 (3)	48 (2)				
KOOS[d]							
KOOS (P)[e]				.001	19.0 (1, 42)	2.7	1.6
Interventional	38.3 (6.9)	55.3 (13.3)	81.2 (9.5)				
Control	32.3 (5.2)	61.0 (8.5)	66.9 (9.3)				
KOOS (S)[f]				.001	17.2 (1, 42)	3.8	1.3
Interventional	40.3 (9.0)	57.4 (10.2)	77.5 (7.0)				
Control	38.0 (13.4)	48.2 (11.8)	54.7 (10.3)				
KOOS (ADL)[g]				.001	11.2 (1, 42)	3.2	1.7
Interventional	57.6 (8.7)	65.0 (7.4)	70.8 (10.1)				
Control	60.1 (9.2)	66.3 (12.0)	67.5 (10.6)				
KOOS (SP)[h]				.001	12.6 (1, 42)	2.7	1.4
Interventional	26.6 (11.9)	35.7 (10.4)	44.5 (12.6)				
Control	30.3 (12.5)	38.0 (6.8)	40.2 (8.3)				
KOOS (QOL)[i]				.001	13.3 (1, 42)	4.9	2.6
Interventional	27.6 (8.1)	44.0 (11.7)	72.3 (9.8)				
Control	25.4 (9.6)	41.6 (10.8)	48.9 (8.5)				
EQ-5D-5L[j]				.001	8.4 (1, 42)	2.4	1.3
Interventional	45.4 (12.5)	68.2 (8.9)	88.1 (9.4)				
Control	42.8 (13.2)	64.6 (10.1)	70.9 (8.7)				

[a]JPS: joint position sense.

[b]MSK-US: musculoskeletal ultrasound.

[c]BBS: Berg Balance Scale.

[d]KOOS: Knee Injury and Osteoarthritis Outcome Score.

[e]KOOS (P): Knee Injury and Osteoarthritis Outcome Score (Pain).

[f]KOOS (S): Knee Injury and Osteoarthritis Outcome Score (Symptoms).

[g]KOOS (ADL): Knee Injury and Osteoarthritis Outcome Score (Activities of Daily Living).

[h]KOOS(SP): Knee Injury and Osteoarthritis Outcome Score (Sports).

[i]KOOS (QOL): Knee Injury and Osteoarthritis Outcome Score (Quality of life).

[j]EQ-5D-5L: EuroQol 5-Dimension 5-Level.

Table 3 demonstrates significant main effects and interaction effects across various outcome measures, indicating substantial differences between groups, over time, and the interaction of group and time.

XSL•FO
RenderX

Table . Main effects and interaction effects for outcome variables.

Dependent variables and source	Mean square	Partial 2	*P* value
JPS[a]			
Group	7.89	0.19	<.001
Time	4.56	0.12	<.001
Group * Time	6.78	0.18	<.001
MSK-US[b]			
Group	9.54	0.23	<.001
Time	5.67	0.16	.002
Group * Time	8.21	0.22	.001
BBS[c]			
Group	11.87	0.28	<.001
Time	6.78	0.18	.001
Group * Time	9.01	0.24	.001
KOOS[d]			
Group	9.32	0.22	.001
Time	5.12	0.14	.003
Group * Time	7.45	0.19	.001
EQ-5D-5L[e]			
Group	8.45	0.20	.001
Time	5.23	0.15	.002
Group * Time	7.12	0.19	<.001

[a]JPS: joint position sense.

[b]MSK-US: musculoskeletal ultrasound.

[c]BBS: Berg Balance Scale.

[d]KOOS: Knee Injury and Osteoarthritis Outcome Score.

[e]EQ-5D-5L: EuroQol 5-Dimension 5-Level.

JPS measurements showed significant main effects for the group (Mean square=7.89, Partial η^2=0.19, *P*<.001) and time (mean square=4.56, Partial η^2=0.12, *P*<.001), indicating improvements in JPS due to the intervention and over time. The interaction effect (Mean square=6.78, Partial η^2=0.18, *P*<.001) supports that the experimental group had more significant gains in JPS.

MSK-US results showed significant main effects for the group (Mean square=9.54, Partial η^2=0.23, *P*<.001) and time (Mean square=5.67, Partial η^2=0.16, *P*=.002), suggesting significant improvements in musculoskeletal health due to the intervention and over time. The interaction effect (Mean square=8.21, Partial η^2=0.22, *P*=.001) further supports that these improvements were more significant in the experimental group.

For balance, as measured by the BBS, there were significant main effects for the group (Mean square=11.87, Partial η^2=0.28, *P*<.001) and time (Mean square=6.78, Partial η^2=0.18, *P*=0.001), indicating substantial improvements in balance due to the intervention and over time. The interaction effect (Mean square=9.01, Partial η^2=0.24, *P*=.001) highlights that the experimental group experienced more significant improvements in balance.

The KOOS results revealed significant main effects for the group (Mean square=9.32, Partial η^2=0.22, *P*=.001) and time (Mean square=5.12, Partial η^2=0.14, *P*=.003), indicating notable improvements in knee function and symptoms due to the intervention and over time. The interaction effect (Mean square=7.45, Partial η^2=0.19, *P*=.001) suggests that these improvements were greater in the experimental group.

Finally, the EQ-5D-5L HRQOL scores showed significant main effects for the group (Mean square=8.45, Partial η^2=0.20, *P*=.001) and time (Mean square=5.23, Partial η^2=0.15, *P*=.002), demonstrating that the intervention significantly improved QOL and that these improvements were sustained over time. The interaction effect (Mean square=7.12, Partial η^2=0.19, *P*<.001) indicates that the experimental group experienced more significant improvements in QOL.

Overall, these results demonstrate that the intervention had a substantial positive impact on functional mobility, pain reduction, musculoskeletal health, balance, knee function, JPS, and QOL. The significant interaction effects highlight that the experimental group benefited more from the intervention

XSL•FO
RenderX

compared with the CG, with these benefits being sustained and increasing over time.

The telerehabilitation monitoring results demonstrated that self-compliance with the home-based lifestyle modification program varied between the intervention and CGs, with notable differences in adherence levels. Overall, 33 out of 44 participants (75%) exhibited more than 90% compliance with the program. Specifically, 18 out of 22 participants (81%) in the IG achieved over 90% compliance, compared with 15 out of 22 participants (68%) in the CG. Moderate compliance was observed in 10 out of 44 participants (23%) of the total sample, with 4 out of 22 participants (18%) of the IG and 6 out of 22 (27%) of the CG falling into this category. Only 1 out of 22 participants (4%) in the CG demonstrated low compliance, while the IG had no participants in this range. Importantly, no participants in either group had compliance levels below 51%, indicating that the majority were highly engaged with the program. These results suggest that the IG, supported by telerehabilitation monitoring, showed better adherence to the home-based lifestyle modification program compared with the CG, highlighting the potential benefits of telerehabilitation in enhancing compliance and engagement in post-TKA rehabilitation.

Discussion

Principal Findings

This study aimed to evaluate the effectiveness of supervised sensorimotor training with and without a lifestyle modification manual through telerehabilitation monitoring in patients undergoing TKA. The main findings indicate that both groups demonstrated improvements in pain, physical function, balance, joint proprioception, and QOL over time; however, the group receiving the combined intervention of sensorimotor training with lifestyle modification showed significantly greater improvements across almost all outcomes. Notably, enhancements were observed in the muscle thickness (MSK-US), balance (BBS), and all subscales of the KOOS, as well as JPS and HRQOL (EQ-5D-5L). The CG did not receive a structured lifestyle plan emphasizing exercise, diet, and education instructions via a web portal; instead, they were instructed to maintain their usual daily routine. Consistent communication was maintained through the telerehabilitation mode, providing continuous guidance and exercise instructions. Both groups received telerehabilitation monitoring, but the lifestyle modification plan was provided exclusively to the IG. Both groups demonstrated improvement in various outcome measures; however, the experimental group had superior outcomes. Specifically, the experimental group showed greater improvements in functional mobility, pain reduction, musculoskeletal health, balance, knee function, JPS, and overall QOL compared with the CG.

Participants in the experimental group, who received the lifestyle modification plan, showed significantly greater improvements in knee function as measured by the KOOS. The improvements spanned multiple subscales, including Pain, Symptoms, ADL, Sports/Recreation, and QOL. This aligns with previous studies demonstrating the efficacy of comprehensive rehabilitation programs in enhancing knee function postsurgery [28]. However, the addition of lifestyle modifications appears to amplify these benefits, suggesting that addressing behavioral and lifestyle factors plays a crucial role in optimizing recovery [29]. The significant increase in the musculoskeletal cross-sectional area, as measured by MSK-US, further supports the advantage of the combined approach. The experimental group exhibited greater increases in cross-sectional area compared with the CG, indicating enhanced musculoskeletal health and strength. This finding contrasts with earlier studies that focused solely on exercise interventions, which reported modest improvements in muscle mass and strength [30]. The inclusion of lifestyle modifications likely contributed to the observed gains by promoting overall physical health and activity levels. JPS, a critical component of proprioception, showed marked improvements in the experimental group. The reduction in absolute error in JPS was significantly greater for participants receiving lifestyle modifications, suggesting that this comprehensive approach enhances proprioceptive accuracy. Previous research has highlighted the benefits of sensorimotor training in improving proprioception post-TKR. This study extends these findings by demonstrating that lifestyle modifications can further refine proprioceptive function [31]. Similarly, balance, as measured by the BBS, improved more significantly in the experimental group. The substantial gains in balance scores indicate that lifestyle modifications, in conjunction with sensorimotor training, provide a more holistic approach to rehabilitation [32]. These findings are consistent with previous literature that emphasizes the role of balance training in improving postural stability postsurgery [33]. However, this study suggests that incorporating lifestyle changes may lead to even greater improvements in balance and stability.

Pain reduction, as measured by the Numeric Pain Rating Scale, was significantly more pronounced in the experimental group. Participants reported lower pain levels throughout the study period, underscoring the effectiveness of the combined intervention in managing postsurgical pain. This aligns with existing studies that have shown exercise and rehabilitation programs can reduce pain levels [34]. The additional benefit observed in this study may be attributed to lifestyle modifications that address inflammation, diet, and overall well-being, contributing to pain reduction [35]. QOL, assessed using the EQ-5D-5L, showed substantial improvements in the experimental group. The significant gains in HRQOL scores indicate that the combined approach not only enhances physical recovery but also positively impacts overall well-being and satisfaction [36]. These findings resonate with previous research that has demonstrated the positive impact of comprehensive rehabilitation programs on QOL post-TKR [37]. The added value of lifestyle modifications highlights the importance of addressing holistic health factors in rehabilitation programs.

The telerehabilitation monitoring results demonstrated higher adherence rates in the experimental group, suggesting that lifestyle modifications may enhance engagement and compliance with rehabilitation programs [34]. The majority of participants in the experimental group achieved over 90% compliance, indicating that the combined approach was well-received and effectively implemented. These findings are in line with studies emphasizing the role of telerehabilitation in supporting

XSL•FO
RenderX

adherence to home-based rehabilitation programs [38]. This study adds to this body of evidence by showing that lifestyle modifications, supported by telerehabilitation, can further improve adherence and outcomes.

To our knowledge, this is the first study of lifestyle modification programs conducted among TKA patients and using the telerehabilitation monitoring method in Pakistan. A noteworthy feature of this study is the translation of the lifestyle modification manual into Urdu, the native language of the Pakistani population. This translation ensured that participants fully understood the lifestyle recommendations and guidelines, thereby enhancing the effectiveness of the intervention. The translation addressed language barriers and made the program more accessible and culturally relevant to the participants, contributing to the high compliance and positive outcomes observed in the experimental group.

However, the study also has certain limitations. The use of a single clinical care ensured superior standardization of care, which can be regarded as a notable benefit in the context of conducting this exploratory and innovative trial. This standardization allowed for consistent and controlled conditions, thereby enhancing the reliability of the results. Nevertheless, the capacity to generalize the findings from this study to other settings, joint systems, and surgical procedures is expected to be limited in comparison to results derived from more diverse multicenter trials. An additional limitation of this innovative study is that, although it was hypothesized that the lifestyle program would enhance participants' confidence to safely engage in a more physically active lifestyle, physical activity was not directly assessed. Future studies should also focus on assessing the cost-effectiveness of telerehabilitation, predominantly in the framework of developing nations like Pakistan. Evaluating economic feasibility, possible savings, and application barriers could benefit in larger-scale adoption.

Conclusions

Both groups that received sensorimotor training demonstrated improvements in their outcomes. However, the incorporation of a lifestyle modification program in the experimental group yielded significantly better results compared with those who received the standard intervention alone. The experimental group showed superior improvements in knee function, performance, and overall QOL. These findings suggest that integrating a lifestyle modification program with sensorimotor training can enhance rehabilitation outcomes for patients undergoing total knee replacement. In addition, the IG, supported by telerehabilitation monitoring, showed better adherence to the home-based lifestyle modification program compared with the CG, highlighting the potential benefits of telerehabilitation in enhancing compliance and engagement in lifestyle modification rehabilitation programs.

Acknowledgments

The author would like to thank Ghurki Trust Teaching Hospital, Lahore, Pakistan, for their continuous support. The study received no external funding. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Data Availability

The data generated or analyzed throughout this study are outlined in this paper and can be made available by the corresponding author upon a reasonable request.

Authors' Contributions

SS and RN contributed to the conceptualization of the study and to the methodology. SS prepared the original draft of the manuscript. SS, RN, and RA contributed to the review and editing of the manuscript. RN provided supervision. SS, RN, and RA contributed to project administration. All authors have read and agreed to the published version of the manuscript.

Multimedia Appendix 1
Lifestyle modification plan.
[PDF File, 1360 KB - mhealth_v13i1e64643_app1.pdf]

Multimedia Appendix 2
Post-hoc adjusted P values.
[DOCX File, 51 KB - mhealth_v13i1e64643_app2.docx]

Checklist 1
CONSORT eHealth checklist.
[PDF File, 11312 KB - mhealth_v13i1e64643_app3.pdf]

References

XSL•FO
RenderX

1. Yang G, Wang J, Liu Y, et al. Burden of knee osteoarthritis in 204 countries and territories, 1990–2019: results from the global burden of disease study 2019. Arthritis Care Res (Hoboken) 2023 Dec;75(12):2489-2500. [doi: 10.1002/acr.25158]
2. Birmingham TB, Primeau CA, Shariff SZ, et al. Incidence of total knee arthroplasty after arthroscopic surgery for knee osteoarthritis: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2024 Apr 1;7(4):e246578. [doi: 10.1001/jamanetworkopen.2024.6578] [Medline: 38635272]
3. Kladny B. Rehabilitation following total knee replacement. Orthopade 2021 Nov;50(11):894-899. [doi: 10.1007/s00132-021-04175-9] [Medline: 34654935]
4. Alyacoubi OK, Ulfa M, Dewi A. Assessing the major aspects and implications of telerehabilitation services: a bibliometric analysis. JMMR 2023;12(2):123-144. [doi: 10.18196/jmmr.v12i2.10]
5. Blasco JM, Hernández-Guillen D, Domínguez-Navarro F, Acosta-Ballester Y, Alakhdar-Mohmara Y, Roig-Casasús S. Sensorimotor training prior total knee arthroplasty and effects on functional outcome: a systematic review and meta-analysis. Gait Posture 2021 May;86:83-93. [doi: 10.1016/j.gaitpost.2021.03.001] [Medline: 33711615]
6. Lippman D, Stump M, Veazey E, et al. Foundations of lifestyle medicine and its evolution. Mayo Clin Proc Innov Qual Outcomes 2024 Feb;8(1):97-111. [doi: 10.1016/j.mayocpiqo.2023.11.004] [Medline: 38304165]
7. Barlow J. How to use education as an intervention in osteoarthritis. Best Pract Res Clin Rheumatol 2001 Oct;15(4):545-558. [doi: 10.1053/berh.2001.0172] [Medline: 11567538]
8. Leech RD, Eyles J, Batt ME, Hunter DJ. Lower extremity osteoarthritis: optimising musculoskeletal health is a growing global concern: a narrative review. Br J Sports Med 2019 Jul;53(13):806-811. [doi: 10.1136/bjsports-2017-098051] [Medline: 30030282]
9. Choukou MA, Zhu X, Malwade S, Dhar E, Abdul SS. Digital health solutions transforming long-term care and rehabilitation. In: Healthcare Information Management Systems: Cases, Strategies, and Solutions: Springer; 2022:301-316. [doi: 10.1007/978-3-031-07912-2_19]
10. Vassilev I, Rowsell A, Pope C, et al. Assessing the implementability of telehealth interventions for self-management support: a realist review. Implement Sci 2015 Apr 24;10:59. [doi: 10.1186/s13012-015-0238-9] [Medline: 25906822]
11. Di Martino G, Della Valle C, Centorbi M, et al. Enhancing behavioural changes: a narrative review on the effectiveness of a multifactorial app-based intervention integrating physical activity. Int J Environ Res Public Health 2024 Feb 16;21(2):233. [doi: 10.3390/ijerph21020233] [Medline: 38397722]
12. Hamilton DF, Akhtar S, Griffiths B, Prior Y, Jones RK. The use of technology to support lifestyle interventions in knee osteoarthritis: a scoping review. Osteoarthr Cartil Open 2023 Jun;5(2):100344. [doi: 10.1016/j.ocarto.2023.100344] [Medline: 36852286]
13. Brennan DM, Mawson S, Brownsell S. Telerehabilitation: enabling the remote delivery of healthcare, rehabilitation, and self management. In: Advanced Technologies in Rehabilitation: IOS Press; 2009:231-248. [doi: 10.3233/978-1-60750-018-6-231]
14. Nuara A, Fabbri-Destro M, Scalona E, Lenzi SE, Rizzolatti G, Avanzini P. Telerehabilitation in response to constrained physical distance: an opportunity to rethink neurorehabilitative routines. J Neurol 2022 Feb;269(2):627-638. [doi: 10.1007/s00415-021-10397-w] [Medline: 33449202]
15. Agha RA, Altman DG, Rosin D. The SPIRIT 2013 statement – defining standard protocol items for trials. Int J Surg 2015 Jan;13(12):288-291. [doi: 10.1016/j.ijsu.2014.12.007]
16. Moutzouri M, Gleeson N, Coutts F, Tsepis E, Gliatis J. Early self-managed focal sensorimotor rehabilitative training enhances functional mobility and sensorimotor function in patients following total knee replacement: a controlled clinical trial. Clin Rehabil 2018 Jul;32(7):888-898. [doi: 10.1177/0269215518757291] [Medline: 29473481]
17. Moutzouri M, Gleeson N, Billis E, Panoutsopoulou I, Gliatis J. What is the effect of sensori-motor training on functional outcome and balance performance of patients' undergoing TKR? a systematic review. Physiotherapy 2016 Jun;102(2):136-144. [doi: 10.1016/j.physio.2015.11.001] [Medline: 26712529]
18. Mears SC, Kates SL. A guide to improving the care of patients with fragility fractures, edition 2. Geriatr Orthop Surg Rehabil 2015 Jun;6(2):58-120. [doi: 10.1177/2151458515572697] [Medline: 26246957]
19. Noor R, Olyaei G, Hadian MR, Talebian S, Salman Bashir M. A reliable and accurate system of joint position sense measurement. Biomed Res 2018;29(12):2528-2531. [doi: 10.4066/biomedicalresearch.29-18-410]
20. Sconfienza LM, Orlandi D, Silvestri E, editors. Ultrasound-Guided Musculoskeletal Procedures: The Lower Limb: Springer; 2015. [doi: 10.1007/978-88-470-5764-7]
21. La Porta F, Caselli S, Susassi S, Cavallini P, Tennant A, Franceschini M. Is the Berg Balance Scale an internally valid and reliable measure of balance across different etiologies in neurorehabilitation? a revisited rasch analysis study. Arch Phys Med Rehabil 2012 Jul;93(7):1209-1216. [doi: 10.1016/j.apmr.2012.02.020] [Medline: 22521926]
22. Roos EM, Toksvig-Larsen S. Knee injury and osteoarthritis outcome score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003 May 25;1:1-0. [doi: 10.1186/1477-7525-1-17] [Medline: 12801417]
23. Kuharić M, Pickard AS, Mukuria C, Finch AP. The measurement properties of the EQ health and well-being and EQ health and well-being short form in Italian population: a comparative study with EQ-5D-5L. Value Health 2024 Mar 14. [doi: 10.1016/j.jval.2024.03.002]

XSL•FO
RenderX

24. Cha AS, Law EH, Shaw JW, Pickard AS. A comparison of self-rated health using EQ-5D VAS in the United States in 2002 and 2017. Qual Life Res 2019 Nov;28(11):3065-3069. [doi: 10.1007/s11136-019-02249-y] [Medline: 31321671]
25. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide. Stat Med 2020 Sep 20;39(21):2815-2842. [doi: 10.1002/sim.8569] [Medline: 32419182]
26. Liang J, Tang ML, Chan PS. A generalized Shapiro–Wilk statistic for testing high-dimensional normality. Comput Stat Data Anal 2009 Sep;53(11):3883-3891. [doi: 10.1016/j.csda.2009.04.016]
27. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012 Feb;141(1):2-18. [doi: 10.1037/a0024338] [Medline: 21823805]
28. Peer MA, Rush R, Gallacher PD, Gleeson N. Pre-surgery exercise and post-operative physical function of people undergoing knee replacement surgery: a systematic review and meta-analysis of randomized controlled trials. J Rehabil Med 2017 Apr 6;49(4):304-315. [doi: 10.2340/16501977-2210] [Medline: 28352936]
29. Ditton E, Johnson S, Hodyl N, et al. Improving patient outcomes following total knee arthroplasty: identifying rehabilitation pathways based on modifiable psychological risk and resilience factors. Front Psychol 2020;11:1061. [doi: 10.3389/fpsyg.2020.01061] [Medline: 32670136]
30. Jakobsen TL, Jakobsen MD, Andersen LL, Husted H, Kehlet H, Bandholm T. Quadriceps muscle activity during commonly used strength training exercises shortly after total knee arthroplasty: implications for home-based exercise-selection. J Exp Orthop 2019 Jul 2;6(1):29. [doi: 10.1186/s40634-019-0193-5] [Medline: 31267365]
31. Moutzouri M, Coutts F, Gliatis J, Billis E, Tsepis E, Gleeson N. Early initiation of home-based sensori-motor training improves muscle strength, activation and size in patients after knee replacement: a secondary analysis of a controlled clinical trial. BMC Musculoskelet Disord 2019 May 17;20(1):231. [doi: 10.1186/s12891-019-2575-3] [Medline: 31101039]
32. Morri M, Vigna D, Raffa D, Donati DM, Benedetti MG. Effect of game based balance exercises on rehabilitation after knee surgery: a controlled observational study. J Med Syst 2019 Apr 12;43(5):1-6. [doi: 10.1007/s10916-019-1271-z] [Medline: 30980213]
33. Doma K, Grant A, Morris J. The effects of balance training on balance performance and functional outcome measures following total knee arthroplasty: a systematic review and meta-analysis. Sports Med 2018 Oct;48(10):2367-2385. [doi: 10.1007/s40279-018-0964-7] [Medline: 30117054]
34. Mistry JB, Elmallah RDK, Bhave A, et al. Rehabilitative guidelines after total knee arthroplasty: a review. J Knee Surg 2016 Apr;29(3):201-217. [doi: 10.1055/s-0036-1579670] [Medline: 26963074]
35. Abe M, Abe H. Lifestyle medicine – an evidence based approach to nutrition, sleep, physical activity, and stress management on health and chronic illness. Personalized Medicine Universe 2019 Jul;8:3-9. [doi: 10.1016/j.pmu.2019.05.002]
36. Pano O, Sayón-Orea C, Hershey MS, Bes-Rastrollo M, Martínez-González MA, Martínez JA. Development of a general health score based on 12 objective metabolic and lifestyle items: the lifestyle and well-being index. Healthcare (Basel) 2022 Jun 11;10(6):1088. [doi: 10.3390/healthcare10061088] [Medline: 35742139]
37. Bhatia S, Karvannan H, Prem V. The effect of bio psychosocial model of rehabilitation on pain and quality of life after total knee replacement: a randomized controlled trial. Journal of Arthroscopy and Joint Surgery 2020 Oct;7(4):177-183. [doi: 10.1016/j.jajs.2020.09.005]
38. Zhang H, Wang J, Jiang Z, Deng T, Li K, Nie Y. Home-based tele-rehabilitation versus hospital-based outpatient rehabilitation for pain and function after initial total knee arthroplasty: a systematic review and meta-analysis. Medicine (Baltimore) 2023;102(51):e36764. [doi: 10.1097/MD.0000000000036764]

Abbreviations

ADL: activities of daily living
BBS: Berg Balance Scale
CG: control group
EQ-5D-5L: EuroQol 5-dimension 5-level questionnaire
HRQOL: health-related quality of life
IG: intervention group
JPS: joint position sense
KOOS: Knee Injury and Osteoarthritis Outcome Score
MSK-US: musculoskeletal ultrasound
OA: osteoarthritis
QOL: quality of life
RF: rectus femoris
TKA: total knee arthroplasty
VAS: Visual Analog Scale

XSL•FO
RenderX

Edited by L Buis; submitted 23.07.24; peer-reviewed by B Unver, S Lu; revised version received 05.05.25; accepted 10.07.25; published 02.10.25.

Please cite as:
Sadiq S, Noor R, Akram R
Effect of Lifestyle Modification Through Web-Based Telerehabilitation Monitoring Combined With Supervised Sensorimotor Training After Total Knee Arthroplasty: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e64643
URL: https://mhealth.jmir.org/2025/1/e64643
doi:10.2196/64643

© Samreen Sadiq, Rabiya Noor, Rizwan Akram. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Innovating Care for Postmenopausal Women Using a Digital Approach for Pelvic Floor Dysfunctions: Prospective Longitudinal Cohort Study

Ana P Pereira[1], PhD; Dora Janela[1], PT; Anabela C Areias[1], PhD; Maria Molinos[1], PhD; Xin Tong[2], PhD; Virgílio Bento[1], PhD; Vijay Yanamadala[1,3,4], MD; Jennesa Atherton[1], PT, DPT; Fernando Dias Correia[1,5], MD, PhD; Fabíola Costa[1], PhD

[1]Sword Health, Inc, Draper, UT, United States

[2]Department of Psychology, University of Virginia, Charlottesville, VA, United States

[3]Department of Surgery, Frank H Netter School of Medicine, Quinnipiac University, Hamden, CT, United States

[4]Department of Neurosurgery, Hartford Healthcare Medical Group, Westport, CT, United States

[5]Neurology Department, Centro Hospitalar e Universitário do Porto, Porto, Portugal

Corresponding Author:
Fabíola Costa, PhD
Sword Health, Inc, Draper, UT, United States

Abstract

Background: The menopause transition is a significant life milestone that impacts quality of life and work performance. Among menopause-related conditions, pelvic floor dysfunctions (PFDs) affect ~40% - 50% of postmenopausal women, including urinary or fecal incontinence, genito-pelvic pain, and pelvic organ prolapse. While pelvic floor muscle training (PFMT) is the primary treatment, access barriers leave many untreated, advocating for new care delivery models.

Objective: This study aims to assess the outcomes of a digital pelvic program, combining PFMT and education, in postmenopausal women with PFDs.

Methods: This prospective, longitudinal study evaluated engagement, safety, and clinical outcomes of a remote digital pelvic program among postmenopausal women (n=3051) with PFDs. Education and real-time biofeedback PFMT sessions were delivered through a mobile app. The intervention was asynchronously monitored and tailored by a physical therapist specializing in pelvic health. Clinical measures assessed pelvic floor symptoms and their impact on daily life (Pelvic Floor Impact Questionnaire–short form 7, Urinary Impact Questionnaire–short form 7, Colorectal-Anal Impact Questionnaire–short form 7, and Pelvic Organ Prolapse Impact Questionnaire–short form 7), mental health, and work productivity and activity impairment. Structural equation modeling and minimal clinically important change response rates were used for analysis.

Results: The digital pelvic program had a high completion rate of 77.6% (2367/3051), as well as a high engagement and satisfaction level (8.6 out of 10). The safety of the intervention was supported by the low number of adverse events reported (21/3051, 0.69%). The overall impact of pelvic floor symptoms in participants' daily lives decreased significantly (−19.55 points, 95% CI −22.22 to −16.88; P<.001; response rate of 59.5%, 95% CI 54.9%-63.9%), regardless of condition. Notably, nonwork-related activities and productivity impairment were reduced by around half at the intervention-end (−18.09, 95% CI −19.99 to −16.20 and −15.08, 95% CI −17.52 to −12.64, respectively; P<.001). Mental health also improved, with 76.1% (95% CI 60.7%-84.9%; unadjusted: 97/149, 65.1%) and 54.1% (95% CI 39%-68.5%; unadjusted: 70/155, 45.2%) of participants with moderate to severe symptomatology achieving the minimal clinically important change for anxiety and depression, respectively. Recovery was generally not influenced by the higher baseline symptoms' burden in individuals with younger age, high BMI, social deprivation, and residence in urban areas, except for pelvic health symptoms where lower BMI levels (P=.02) and higher social deprivation (P=.04) were associated with a steeper recovery.

Conclusions: This study demonstrates the feasibility, safety, and positive clinical outcomes of a fully remote digital pelvic program to significantly improve PFD symptoms, mental health, and work productivity in postmenopausal women while enhancing equitable access to personalized interventions that empower women to manage their condition and improve their quality of life.

Trial Registration: ClinicalTrials.gov NCT05513417; https://clinicaltrials.gov/study/NCT05513417

(JMIR Mhealth Uhealth 2025;13:e68242) doi:10.2196/68242

KEYWORDS

women's health; pelvic floor muscle training; physical therapy; menopause; digital therapeutics; biofeedback; mobile phone

XSL•FO
RenderX

Introduction

The menopause transition is a major health milestone women face, often associated with a significant psychological, behavioral, and social impact, that can deeply decrease productivity and overall quality of life [1-4]. Every year, about 27 million women in the United States workforce experience menopause, corresponding to approximately 20% of the workforce [5]. Menopause-related symptoms, ranging from cognitive or mental to pelvic floor dysfunctions (PFDs), often emerge between the ages of 44 and 56 years (average age of 51 years in the United States) [1,3], just when women are likely to move into leadership positions [5,6]. Menopause has therefore been pointed as a risk factor for female talent loss [5-7]. Notably, 1 in 5 women have either quit or considered leaving their job due to the severity of menopause symptoms [5,6]. Overall costs attributable to menopause symptoms were estimated to amount to up to US $26 billion per year in the United States alone, including US $1.8 billion stemming from indirect costs associated with impaired work productivity [6].

Among menopause-related symptoms, PFDs are particularly prevalent, affecting ~40% - 50% of postmenopausal women [8,9]. Due to the decline in estrogen, the pelvic floor muscles and tissues become thinner, drier, less elastic, and weaker, which can result in conditions such as urinary or fecal incontinence, genito-pelvic pain, and pelvic organ prolapse [10-14].

First-line interventions for PFDs encompass pelvic floor muscle training (PFMT), education, and behavioral change strategies [15-21]. PFMT has proven effectiveness in the management of urinary incontinence [22-28], bowel symptoms [26], pelvic organ prolapse [24,26], and genito-pelvic pain [29,30] in postmenopausal women.

Despite high symptom prevalence and their negative impact, many women do not seek care at all (up to 61%) [31-33]. Low treatment rates are related to multiple barriers to care, including geographical and time constraints, persistent social stigma, the perception that symptoms are a “normal” part of aging [31-33], and lack of access to trained providers [34-36] or awareness of available treatment options [31-33]. Consequently, the disease burden experienced by these women might increase. As an example, in a randomized controlled trial, postmenopausal women (n=48) who did not receive PFMT experienced a 50% increase in urinary leakage after one year, while those who performed PFMT showed significant improvement in their condition [27], underscoring the need for appropriate care for this population.

Digital pelvic rehabilitation solutions have arisen as a potential alternative to improve both access and adherence to treatment. Digital interventions have already been reported to be effective against urinary incontinence [37,38], even when compared to in-person PFMT [39]. Additionally, in a study including mostly postmenopausal women, a self-management mobile app for urinary incontinence has shown higher improvement in quality of life and urinary-related symptoms than an information-only intervention [40]. However, further evidence is lacking on the specific needs of postmenopausal women, including those with PFDs other than urinary incontinence.

This study aims to assess the feasibility, engagement, and clinical outcomes (namely the impact of PFDs on physical, mental, and work productivity) of postmenopausal women after a fully remote digital pelvic program, combining education with PFMT using real-time biofeedback. This digital pelvic program is asynchronously managed by a physical therapist (PT), and it has previously shown significant clinical improvements in managing urinary incontinence in women [41]. This study hypothesizes that participants would report improvement in all outcomes after the program. The findings collected could inform health care providers, researchers, and policy makers about the potential of remote digital pelvic programs to improve health outcomes and accessibility for postmenopausal women with PFDs.

Methods

Study Design

This is a real-world, prospective single-arm observational cohort study. The recruitment period was from December 1, 2022, until July 29, 2024, and the home-based digital pelvic program was conducted between December 1, 2022, and August 8, 2024. This study was reported in accordance with STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Table S1 in Multimedia Appendix 1).

Ethical Considerations

This research was conducted in accordance with the Declaration of Helsinki and all applicable ethical guidelines and regulations. The protocol was prospectively approved by the Advarra Institutional Review Board (Pro00064510) and registered on ClinicalTrials.gov (NCT05513417) on August 24, 2022. Electronic informed consent was obtained from all participants. Participants were informed of their right to withdraw from the study at any time without any adverse consequences. All collected data underwent a rigorous anonymization process to safeguard the privacy of the individuals involved in the research. The data collection and analysis procedures complied with established guidelines and regulations. Participants were not offered any form of compensation.

Participants

Female beneficiaries of employers or health plans covered by the Sword Bloom program (Draper, Utah, United States) across 50 states in the United States and the District of Columbia were invited to apply. Individuals in the postmenopausal phase, defined as at least 1 year after their final menstrual period [42], who reported PFDs encompassing conditions such as urinary conditions (including incontinence and voiding dysfunctions), pelvic organ prolapse, bowel conditions, and genito-pelvic pain were included in this study. This diverse array of PFDs was included considering common conditions among postmenopausal women [3,14,43], and to evaluate the feasibility and outcomes of the intervention across multiple conditions for which pelvic rehabilitation is recommended as first-line treatment [15-21,30,44].

Exclusion criteria comprised: (1) inability to perform 20 minutes of light to moderate exercise; (2) active cancer or under

XSL•FO
RenderX

treatment for cancer; (3) surgery, significant trauma, or other conditions where mobilization is contraindicated; (4) clinical red flags suggestive of serious underlying conditions not cleared by their attending physician; (5) signs of acute, serious neurologic compromise; (6) clinical conditions (eg, dementia) precluding compliance with autonomous home-based exercise; (7) pelvic infection or suspicion of inflammatory bowel disease; (8) contraindication for the use of an intravaginal device; and (9) allergy to silicone. These criteria were selected in order to exclude participants that require medical referral and conditions that prevent autonomous engagement with the program, guaranteeing participant's safety.

Participants were considered dropouts in the absence of exercise sessions for 30 consecutive days. Those who did not complete reassessment surveys but were compliant with the intervention were not excluded.

Intervention

The intervention had an average duration of 10 weeks and consisted of biofeedback-mediated PFMT, functional exercises, and education (Table S2 in Multimedia Appendix 1), as previously described [41]. Briefly, after enrollment, participants completed an onboarding form with demographic and clinical information and selected a PT specializing in pelvic health who oversaw and continuously tailored the treatment. During the onboarding video call, the PT conducted the initial clinical evaluation, leveraging the information provided by the participant upon enrollment. During anamnesis, PTs confirmed the absence of red flags indicating potential serious conditions requiring medical screening and assessed the participant's clinical presentation (including the participant's ability to contract the PFM and the presence of comorbid pelvic floor conditions). Additionally, PTs provided education (about PFDs, PFMT, diaphragmatic breathing training, behavioral modifications, and relevant information regarding the use and maintenance of the biofeedback device), and established the intervention goals collaboratively with the patient, based on the clinical practice guidelines [15-21,30,44].

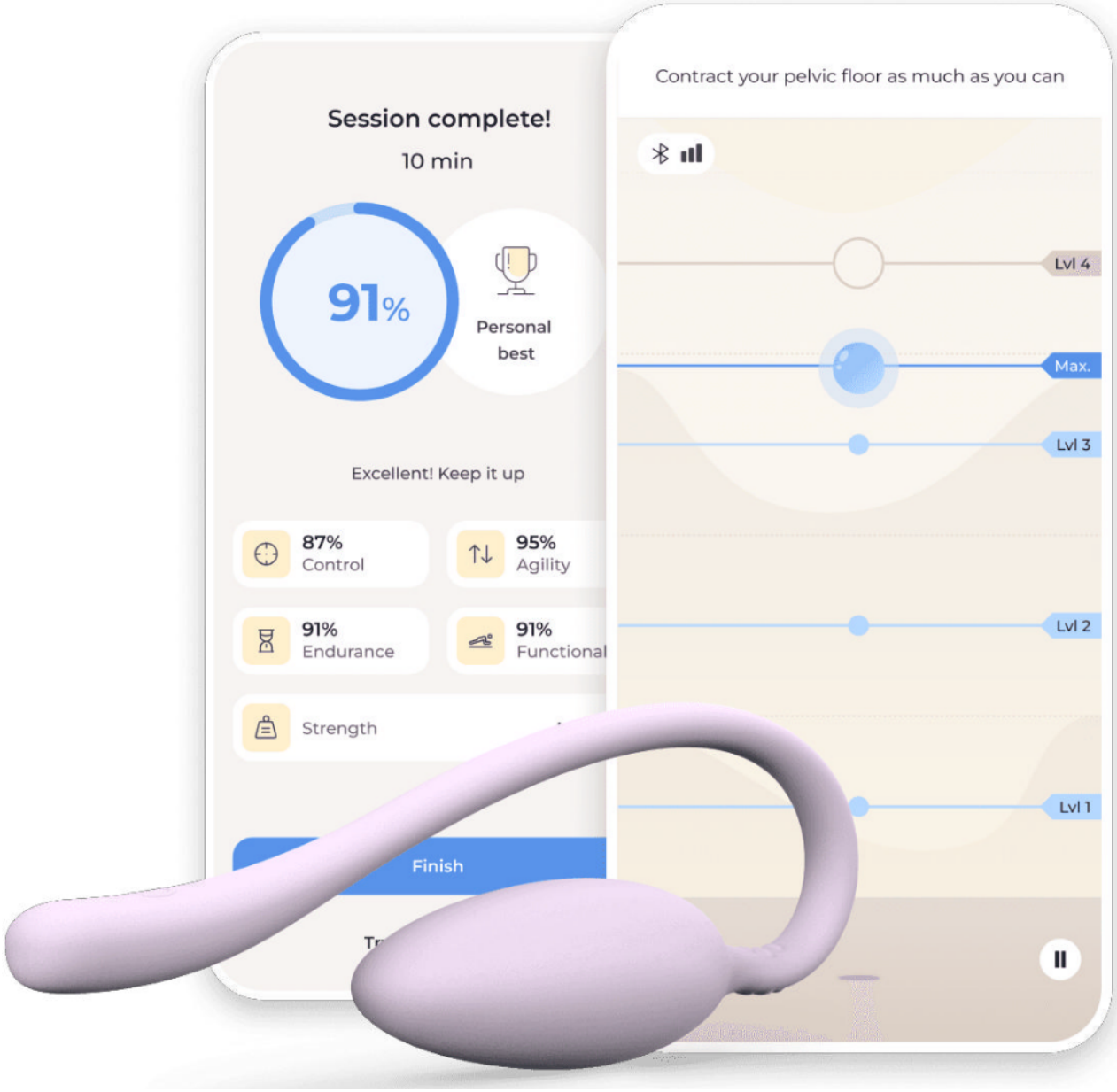
This digital pelvic program used a Food and Drug Administration–listed medical device composed of an intravaginal sensor (shipped to each participant), a dedicated mobile app (to be downloaded to the participant's smartphone), and a cloud-based platform, allowing remote monitoring and care (Figure 1). The intravaginal sensor included a force transducer to monitor pelvic floor muscle activity (contraction and relaxation) and an accelerometer to assess pelvic floor muscle motion. The data captured by the intravaginal sensor was accessed remotely by the PT through a cloud-based portal, allowing the asynchronous assessment and monitoring of the pelvic floor muscles function in different aspects (eg, maximal strength and endurance).

Gamified exercise sessions (4 sessions per week as the default recommendation) were displayed in the app, accompanied by written instructions, with the intravaginal sensor providing real-time biofeedback based on the prescribed contraction and relaxation targets.

The educational content on pelvic health followed current clinical guidelines and research [15,16,44] and included the role of mental health and topics tailored to each specific PFD (eg, bladder retraining and toilet habits). These were selected by the attending PT and were delivered in the form of written studies and videos via the app throughout the digital pelvic program. Both exercise performance data and education content engagement were stored on a cloud-based platform, being continuously monitored by the PT to asynchronously assess progress and adjust the treatment plan.

XSL•FO
RenderX

Figure 1. Representation of the intravaginal pod and interventional mobile app screenshots, featuring an exercise session.



Outcomes

Outcomes included engagement, satisfaction, adverse events, and clinical outcomes, which are described in Table 1. Clinical outcomes assessment surveys were completed at baseline, session 9, session 15, and session 21 as long as participants reached that milestone.

XSL•FO
RenderX

Table . Study engagement, satisfaction, and clinical outcomes.

Outcome and measure	Description
Engagement metrics	
Completion rates	• Percentage of participants completing the program
Sessions	• Number of total sessions performed
Education	• Number of educational pieces read or watched
Interaction	• Number of text interactions exchanged between participant and PT[a]
Program satisfaction	
Satisfaction	• Through the question "On a scale from 0 to 10, how likely are you to recommend Bloom's programs to a friend or family member?"; range: 0 - 10 (higher scores indicate higher satisfaction)
Safety	
Adverse events rate	• Self-reported adverse events
Clinical outcomes	
Pelvic floor symptoms	• Pelvic Floor Impact Questionnaire–short form 7 [45]; range: 0 - 300 (higher scores indicate greater impact). MCIC[b] threshold: 12% [45] • Urinary Impact Questionnaire–short form 7 [44-46]; range: 0 - 100 (higher scores indicate greater impact) • Colorectal-Anal Impact Questionnaire–short form 7 [45]; range: 0 - 100 (higher scores indicate greater impact) • Pelvic Organ Prolapse Impact Questionnaire–short form 7 [45]; range: 0 - 100 (higher scores indicate greater impact) • Through the question "In the past 7 days, how would you rate the severity of your pelvic health symptoms?"; range: 0 - 10 (higher scores indicate greater severity)
Mental health	• Depression through Patient Health 9-item Questionnaire [47]; range: 0 - 27 (higher scores indicate greater symptoms). MCIC: ≥5 [48] • Anxiety through Generalized Anxiety Disorder 7-item scale [49]; range: 0 - 21 (higher scores indicate greater symptoms). MCIC: ≥3.8 [50]
WPAI[c]	• WPAI questionnaire–General Health v2.0 [51]; range: 0% - 100% (higher scores indicate greater impairment)

[a]PT: physical therapist.

[b]MCIC: minimal clinically important change.

[c]WPAI: work productivity and activity impairment.

Safety and Adverse Events

Participants were asked to report any adverse events to their assigned PT. An adverse event was defined as any undesirable experience associated with the use of a medical product by a participant [52]. An adverse event was considered serious if it was life-threatening, required hospitalization, or led to disability, permanent damage, congenital anomaly, or birth defect. Reported events were classified as related, not related, or of unknown relationship to the intervention. Regular and on-demand contact between participants and PTs was maintained via a secure in-app chat for support, feedback, motivation, and safety assurance. Additionally, symptoms and fatigue levels (rated on a 0 - 10 scale) were also self-reported after each exercise session and remotely monitored by the PT.

Statistical Analysis

A latent growth curve analysis (LGCA) was used to model clinical outcome trajectories across sessions following an intention-to-treat approach. LGCA is a type of structural equation model that calculates overall change based on individual trajectories, considering time as continuous [53]. Advantages of this methodology include the provision of model fit measures, and the handling of missing data through full information maximum likelihood, which outperforms listwise deletion and other imputation models [54,55]. A robust sandwich estimator was used for standard errors. A conditional LGCA was conducted to assess the influence of covariates—age, BMI, hormone replacement therapy (yes or no), social deprivation index [56], and rurality (rural or urban)—on intercept, slope, and curve, fitted as random effects.

XSL•FO
RenderX

Latent-basis growth analysis (LBGA), following an intention-to-treat approach, was conducted for Work Productivity and Activity Impairment Questionnaire (WPAI) and Pelvic Health Symptoms outcomes, as LGCA did not result in a satisfactory model fit. LBGA provided a better model fit as it accounts for variations in the pace and timing of change across individuals [57]. These analyses considered those with baseline scores of >0. A conditional LBGA was also performed with the same mentioned covariates.

Response rates for Pelvic Floor Impact Questionnaire–short form-7 (PFIQ-7), Patient Health 9-item Questionnaire (PHQ-9), and Generalized Anxiety Disorder 7-item (GAD-7) were calculated for participants who completed the program and reported baseline clinically relevant scores (ie, PHQ-9 ≥10 and GAD-7 ≥10, respectively), using the respective minimal clinically important change (MCIC) thresholds (Table 1). These response rates were adjusted for the aforementioned covariates.

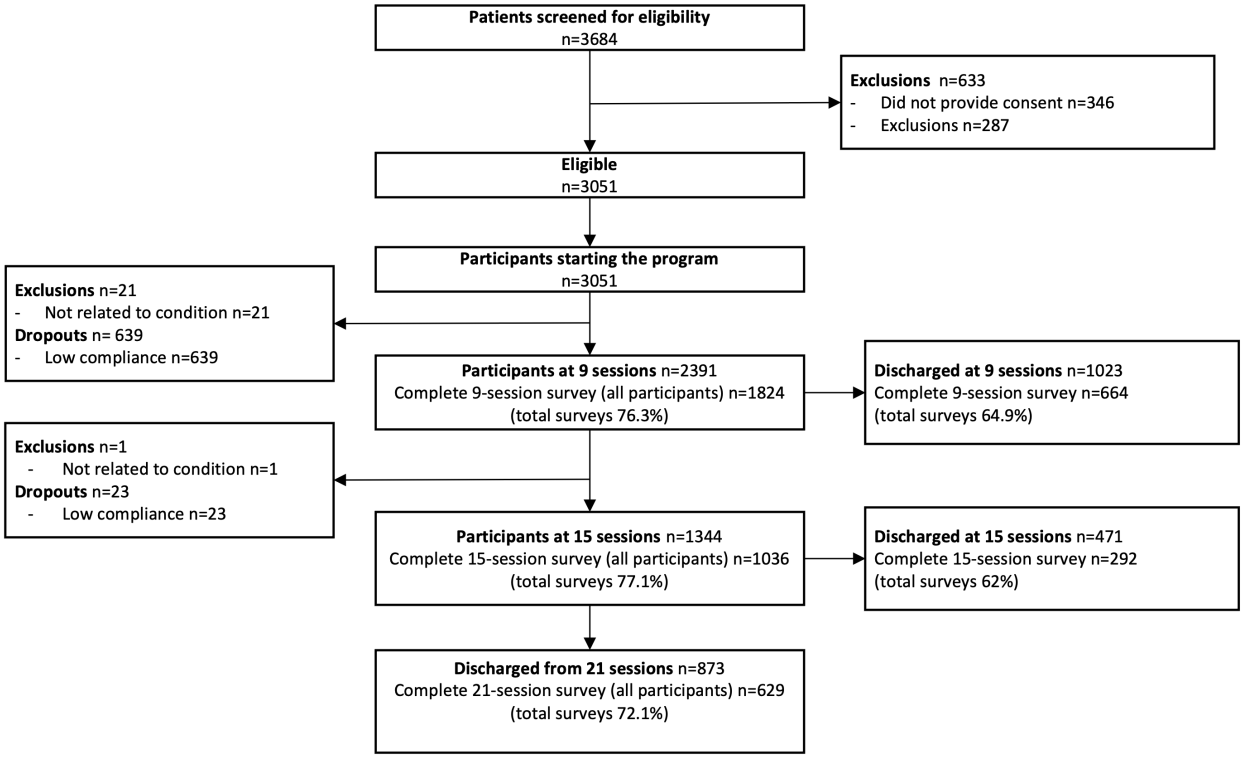
Statistical analyses were performed using R Studio (version 2023.09.1+494; Posit, PBC). Statistical significance was defined as $P<.05$ considering a 2-sided hypothesis test.

Results

Overview

Of the 3684 participants screened for eligibility, 633 participants were excluded (346 declined participation and 287 did not meet eligibility criteria; Figure 2). The program started with 3051 participants, of which 2367 participants completed it, translating into a completion rate of 77.6%.

Figure 2. Study flowchart.



Baseline Characteristics

Overall, the cohort was on average 54.3 (SD 6.5) years old, with an average BMI of 29.4 (SD 6.8) kg/m^2, having a high proportion of participants with obesity (1197/3051, 39.2%), of those with higher education levels (1826/3051, 59.9%) and full-time employed (2332/3051, 76.4%). Regarding races or ethnicities, 18.5% (563/3051) were from minority groups (Table 2). Of those who reported parity, 87.1% (1763/2025) were parous or multiparous (Table 2). Urinary conditions were the most common PFD (2174/3051, 71.3%), followed by genito-pelvic pain (597/3051, 19.6%; Table 2).

XSL•FO
RenderX

Table . Baseline characteristics of study participants: entire cohort (n=3051).

Characteristic	Entire cohort (n=3051)
Age (years), mean (SD)	54.3 (6.5)
Age categories (years), n (%)	
35 - 40	45 (1.5)
41 - 54	1595 (52.3)
≥55	1411 (46.2)
Gender, n (%)	
Women	2879 (94.4)
Nonbinary	12 (0.4)
Prefer not so to specify or not available	160 (5.2)
BMI (kg/m^2; n=3038), mean (SD)	29.4 (6.8)
BMI categories (kg/m^2; n=3038), n (%)	
Underweight (<18.5)	24 (0.8)
Normal (18.5 - 25)	879 (28.9)
Overweight (≥25 - 30)	938 (30.9)
Obesity (≥30 - 40)	971 (32)
Severe obesity (>40)	226 (7.4)
Employment status, n (%)	
Employed full-time	2332 (76.4)
Employed part-time	253 (8.3)
Unemployed	211 (6.9)
Retired	198 (6.5)
Prefer not to specify	57 (1.9)
Education level, n (%)	
Less than high school	11 (0.4)
High school diploma	273 (8.9)
Some college	903 (29.6)
Bachelor’s degree	1149 (37.7)
Graduate degree	677 (22.2)
Prefer not to specify	38 (1.2)
Race or ethnicity, n (%)	
American Indian or Alaska Native	9 (0.3)
Asian or Pacific Islander	73 (2.4)
Black or African American	219 (7.2)
Hispanic or Latino	255 (8.4)
Multiracial or biracial	43 (1.4)
Not listed	7 (0.2)
Prefer not to specify	47 (1.5)
White or Caucasian	2398 (78.6)
Geographic location, n (%)	
Rural	437 (14.3)
Urban	2608 (85.5)

XSL•FO
RenderX

Characteristic	Entire cohort (n=3051)
Not available	6 (0.2)
Social Deprivation Index (n=3045), n (%):	
1 - 20 (less deprived)	1025 (33.6)
21 - 40	728 (23.9)
41 - 60	576 (18.9)
61 - 80	428 (14)
81 - 100 (more deprived)	288 (9.4)
Pelvic floor dysfunction, n (%)	
Genito-pelvic pain or penetration disorder	597 (19.6)
Bowel conditions	58 (1.9)
Pelvic organ prolapse	198 (6.5)
Urinary conditions	2174 (71.3)
Multiple conditions	24 (0.8)
Parity, n (%)	
Nulliparous	262 (8.5)
Parous or multiparous	1763 (57.8)
Not available	1026 (33.6)
Self-reported use of hormone replacement therapy (n=260), n (%)	260 (8.5)
Clinical outcomes, mean (SD)	
PFIQ-7[a]	49.5 (51.9)
UIQ-7[b] (n=3043)	24.7 (22.8)
CRAIQ-7[c] (n=2887)	11.3 (19.8)
POPIQ-7[d] (n=2981)	14.5 (20.6)
Pelvic Health Symptoms >0 (n=2855)	4.81 (2.01)
GAD-7[e]	3.59 (4.4)
GAD-7 ≥10 (n=289)	13.8 (3.4)
PHQ-9[f]	2.64 (4.6)
PHQ ≥10 (n=297)	13.8 (3.6)
WPAI[g] Overall >0 (n=1048)	31.2 (22.6)
WPAI Work >0 (n=990)	28.1 (19.3)
WPAI Activities >0 (n=1571)	33.2 (22.3)

[a] PFIQ-7: Pelvic Floor Impact Questionnaire–short form 7.
[b]UIQ-7: Urinary Impact Questionnaire–short form 7.
[c]CRAIQ-7: Colo-Rectal-Anal Impact Questionnaire–short form 7.
[d]POPIQ-7: Pelvic Organ Prolapse Impact Questionnaire–short form 7.
[e]GAD-7: Generalized Anxiety Disorder 7-item scale.
[f]PHQ-9: Patient Health 9-item Questionnaire.
[g]WPAI: Work Productivity and Activity Impairment Questionnaire.

Engagement Outcomes

Treatment duration lasted 9.9 (SD 8.98) weeks on average, with participants completing a mean of 19.1 (SD 16.2) PFMT biofeedback-assisted sessions. Participants also demonstrated high engagement with the educational component, consulting on average 14.1 (SD 18.4) educational materials. Communication between participants and PTs was frequent throughout the program, amounting to a mean of 22.1 (SD 9.9) text interactions. Satisfaction with the program was 8.6 (SD 2.0) out of 10.

XSL•FO
RenderX

Safety

Throughout the intervention, a total of 21 (out of 3051, 0.69%) adverse events were reported. Of these, 1 vaginal abrasion was reported, 1 case of vaginal bleeding that was determined to be unrelated to the intervention, and 19 cases of yeast infection or urinary tract infections were reported with unknown relation to the intervention. No serious adverse events were reported.

Clinical Outcomes

Model estimates derived from longitudinal analyses are presented in Table 3, showing a good model fit (Table S3 in Multimedia Appendix 1). Estimates from the conditional models are described in Table S4 in Multimedia Appendix 1 and the respective model fit in Table S5 in Multimedia Appendix 1.

Table . Model estimates of clinical outcome measures following an unconditional intent-to-treat approach.

Outcome	Baseline, mean (95% CI)	Program-end, mean (95% CI)	Change, mean (95% CI)	*P* value
PFIQ-7[a] (n=3051)	49.59 (47.75 to 51.43)	30.04 (27.45 to 32.63)	−19.55 (−22.22 to −16.88)	<.001
UIQ-7[b] (n=3043)	24.70 (23.90 to 25.51)	14.35 (13.33 to 15.36)	−10.36 (−11.41 to −9.30)	<.001
CRAIQ-7[c] (n=2887)	11.33 (10.60 to 12.05)	7.57 (6.54 to 8.59)	−3.76 (−4.83 to −2.69)	<.001
POPIQ-7[d] (n=2981)	14.58 (13.85 to 15.32)	8.96 (7.92 to 10.00)	−5.62 (−6.73 to −4.51)	<.001
Pelvic Health Symptoms >0 (n=2855)	4.81 (4.74 to 4.88)	2.84 (2.70 to 2.98)	−1.97 (−2.12 to −1.83)	<.001
WPAI[e] Overall >0 (n=1048)	31.21 (29.84 to 32.58)	16.13 (13.78 to 18.48)	−15.08 (−17.52 to −12.64)	<.001
WPAI Work >0 (n=990)	28.13 (26.93 to 29.33)	13.73 (11.69 to 15.78)	−14.40 (−16.57 to −12.22)	<.001
WPAI Activities >0 (n=1571)	33.14 (32.04 to 34.24)	15.05 (13.23 to 16.86)	−18.09 (−19.99 to −16.20)	<.001

[a]PFIQ-7: Pelvic Floor Impact Questionnaire–short form 7.

[b]UIQ-7: Urinary Impact Questionnaire–short form 7.

[c]CRAIQ-7: Colo-Rectal-Anal Impact Questionnaire–short form 7.

[d]POPIQ-7: Pelvic Organ Prolapse Impact Questionnaire–short form 7.

[e]WPAI: Work Productivity and Activity Impairment Questionnaire.

Pelvic Floor Symptoms

The impact of pelvic floor symptoms in participants' daily activities was significantly decreased by the program-end. Specifically, PFIQ-7 scores decreased by 19.55 points (95% CI −22.22 to −16.88; *P*<.001; Table 3) from a baseline score of 49.59 points (95% CI 47.75-51.43; *P*<.001), resulting in a treatment response rate of 59.5% (95% CI 54.9%-63.9%; unadjusted: 1125/1843, 61%).

Similarly, significant improvements were observed in bladder symptoms (Urinary Impact Questionnaire–short form 7 [UIQ-7]), with a reduction of 10.36 (95% CI −11.41 to −9.30; *P*<.001), as well as in bowel symptoms (Colorectal-Anal Impact Questionnaire–short form 7 [CRAIQ-7]: −3.76, 95% CI −4.83 to −2.69; *P*<.001), pelvic organ prolapse symptoms (Pelvic Organ Prolapse Impact Questionnaire–short form 7 [POPIQ-7]: −5.62, 95% CI −6.73 to −4.51; *P*<.001), and pelvic health symptoms (−1.97, 95% CI −2.12 to −1.83).

Additionally, a change of 18.09 points (95% CI −19.99 to −16.20; *P*<.001) was also observed in WPAI daily activities, reflecting a 54.6% (18.09/33.14) overall improvement in the cohort's ability to perform nonwork-related activities.

Mental Health

At baseline, 9.5% (289/3051) of participants reported moderate-to-severe anxiety symptoms (GAD-7 ≥10), whereas 9.7% (297/3051) participants reported moderate-to-severe depression symptoms (PHQ-9 ≥10). Mental distress was significantly reduced at program-end among these participants, with 76.1% (95% CI 60.7%-84.9%) reaching the MCIC for anxiety and 54.1% (95% CI 39%-68.5%) reaching the MCIC for depression (unadjusted: 97/149, 65.1% for anxiety and 70/155, 45.2% for depression).

Work Productivity

Work productivity impairment had a mean score of 31.21 (95% CI 29.84-32.58) at baseline, which decreased to nearly half at intervention-end (mean change: −15.08, 95% CI −17.52 to −12.64; *P*<.001), representing a 48.3% (15.08/31.21) change.

Impact of Covariates

At baseline, higher social deprivation was associated with a greater pelvic floor symptoms severity (*P*<.001) and impact on daily living (PFIQ-7; *P*<.001), including both bladder (UIQ-7; *P*<.001), bowel (CRAIQ-7; *P*=.001), and pelvic organ prolapse (POPIQ-7; *P*=.004 and *P*=.009, respectively; Table S4 in Multimedia Appendix 1). Additionally, participants with older age and higher BMI also reported more severe pelvic floor symptoms (*P*<.001) and higher symptoms impact on daily living

XSL•FO
RenderX

(PFIQ-7, UIQ-7, CRAIQ-7, and POPIQ-7; all P=.001; Table S4 in Multimedia Appendix 1), respectively. Those living with higher BMI also reported a greater impairment in daily activities (WPAI activities; P=.003). A lower impairment in nonwork-related activities was observed among participants residing in rural locations (P=.045). Younger age was related to both greater impact of pelvic organ prolapse symptoms on daily living (P=.003) and impairment in nonwork-related activities (P=.006) at baseline.

Covariates did not affect the recovery of any of the clinical outcomes assessed (P≥.05; Table S4 in Multimedia Appendix 1), except for pelvic health symptoms where lower BMI levels (P=.02) and higher social deprivation (P=.04) were associated with a steeper recovery.

Discussion

Principal Results

This study provides real-world evidence to support the feasibility, safety, and positive clinical outcomes of a digital pelvic program for postmenopausal women. A high completion rate of 77.6% (2367/3051) was observed, alongside high engagement and satisfaction levels with the program (8.6/10, SD 2.0). Adverse events were low (21/3051, 0.69%), suggesting the safety of this digital pelvic program. Pelvic floor symptoms were significantly reduced at program-end (−1.97 points, 95% CI −2.12 to −1.83; P<.001), as well as their impact on daily lives (−19.55 points, 95% CI −22.22 to −16.88; P<.001), with a treatment adjusted response rate of 59.5% (95% CI 54.9%-63.9%; unadjusted: 1125/1843, 61%). Pelvic floor symptoms were reduced at intervention-end regardless of the specific condition—bladder, bowel, pelvic organ prolapse, or genito-pelvic symptoms (all P<.001). This also translated into a substantial reduction in impairment in carrying out nonwork-related activities (18.09/33.14, 54.6%; P<.001). Mental health outcomes among participants with moderate-to-severe anxiety or depression symptoms were improved at program-end, with 76.1% (95% CI 60.7%-84.9%; unadjusted: 97/149, 65.1%) and 54.1% (95% CI 39%-68.5%; unadjusted: 70/155, 45.2%) of participants reaching the MCIC for anxiety and depression, respectively. Work productivity impairment was decreased by nearly half at intervention-end (15.08/31.21, 48.3%; P<.001). Outcomes were not generally impacted by age, BMI, hormone replacement therapy, social deprivation, or rurality. This study showcases for the first time the benefit of a digital rehabilitation program, including education and biofeedback-guided PFMT, to improve physical, mental, and work productivity outcomes in postmenopausal women.

Comparison With Prior Work

Engagement With the Program

Herein, we explore the potential of digital interventions in providing effective and scalable avenues that facilitate access to care, and importantly, stimulate compliance in postmenopausal women. In this study, a diverse cohort was evaluated, encompassing participants from various socioeconomic, rural, and racial or ethnic backgrounds, following the diversity observed in the US general population [58]. The program completion rate was high (2367/3051, 77%) and was accompanied by high adherence as translated by the number of biofeedback-assisted PFMT sessions performed (mean 19.1, SD 16.2) and the educational content consulted (mean 14.1, SD 18.4), reinforcing the role of remote interventions in promoting equitable access. Prior studies using PFMT for postmenopausal women have often combined in-person sessions with home-based prescriptions for unsupervised PFMT [23,26,27,59], to promote higher treatment frequency, which is not achievable through in-person settings alone. The high engagement in this digital pelvic program matches the range reported in past controlled studies [22]. The observed engagement may stem from privacy, lack of stigma, and convenience of digital care, enabling women to engage with the program at their own pace and fit sessions into their daily routines [26]. Furthermore, participants could openly discuss their condition with PTs specialized in pelvic health throughout the program as needed (mean text interactions 22.1, SD 9.9). This likely contributed to building a strong therapeutic alliance and fostering an empathetic, nonjudgmental environment, as reflected by the high satisfaction levels of 8.6 (SD 2.0) out of 10, similar to the reported previous in-person PFMT studies [60-62]. Additionally, the biofeedback feature may have also contributed to the high adherence observed. Previous studies have shown a positive impact of biofeedback on the autonomous performance of PFMT, guiding women in correctly contracting their pelvic floor muscles [63], promoting PFMT self-efficacy [64,65], higher adherence [66], and recovery likelihood (by 3 times compared to no biofeedback) [67].

Safety

A significant concern with any intervention is the possibility of the occurrence of side effects that could negatively impact a participant's health. As such, the monitoring of potential adverse events is one of the mandatory requests in health care. Herein, the number of adverse events was low and no serious adverse events occurred, consistent with current literature for in-person interventions [26,27,59], suggesting the safety of this fully remote intervention.

Clinical and Productivity Outcomes

This digital pelvic program led to significant improvements in pelvic floor symptomatology, reflected by a high treatment response rate (59.5%, 95% CI 54.9%-63.9%; unadjusted: 1125/1843, 61%), which was observed regardless of the condition. The findings of this study are consistent with prior research showing the benefits of in-person PFMT in reducing pelvic floor symptoms and their impact on daily living among postmenopausal women [22-28,59]. Moreover, PFIQ-7 improvements were achieved regardless of the impact of age, BMI, self-reported hormone replacement therapy use, social deprivation, and rurality, which is particularly encouraging given that these factors contribute to PFD onset [68] or are typically associated with poorer outcomes after PFMT [69-71]. According to previous reports from patients with PFDs, mental distress is frequently associated with menopause, namely through feelings of discomfort, helplessness, and mood deterioration, all of which subsequently contribute to a diminished quality of life [72-76]. Herein, participants with

XSL•FO
RenderX

moderate-to-severe mental distress reported significant improvements, with treatment response rates reaching 76.1% (95% CI 60.7%-84.9%; unadjusted: 97/149, 65.1%) for anxiety and 54.1% (95% CI 39%-68.5%; unadjusted: 70/155, 45.2%) for depression. These results align with a previous systematic review focused on postmenopausal women who reported significant effects of exercise on mental outcomes [77]. This underscores the importance of considering mental health in managing PFDs, as mental distress predicts poorer recovery following PFMT, and improvements in depression are linked to better PFD symptom outcomes [78].

Adding to the positive clinical outcomes observed, productivity was also significantly improved after this fully remote digital pelvic program (15.08/31.21, 48.3%; $P<.001$). The overarching impact of menopause on productivity is well-known [3,79], encompassing tremendous socioeconomic consequences and even loss of female talent [5-7]. Therefore, preliminary studies have been delving into workplace strategies to support women in dealing with menopause-related symptoms and mitigate the associated challenges [80]. A self-help cognitive behavioral therapy booklet promoted improvements in work presenteeism by 15% and in work and social adjustment (32%) [81]. Although measured by different patient-reported outcomes, our results exceed these findings, showcasing the potential of multimodal digital pelvic programs in helping women manage symptoms and thrive in their professional lives, especially when offered as an employee benefit. Overall, these results showcase, for the first time, the benefits of a digital pelvic program, including education and biofeedback-guided PFMT, to improve physical, mental, and work productivity outcomes, while providing equitable access to care, in postmenopausal women from all socioeconomic backgrounds.

Limitations

This study presents some limitations that warrant discussion: (1) the observational study design that lacks a control group precludes the establishment of a causal effect; (2) the lack of pelvic floor muscle function outcomes, such as strength and endurance, preventing the assessment of objective measures in this domain; and (3) the concomitant use of other medication besides hormone replacement therapy by some participants (eg, anticholinergics), or other nonpharmacological interventions (eg, pessaries), which may be a source of potential confounding. Nevertheless, the use of hormone replacement therapy was collected and used as a covariate when analyzing outcomes. Future research is needed to determine the maintenance of the observed improvements through long-term follow-ups.

Conclusions

This study demonstrates the feasibility, safety, and positive clinical outcomes, namely on pelvic floor symptoms, mental health status, and work productivity, after a fully remote, biofeedback-assisted digital pelvic program in postmenopausal women with PFDs. These findings highlight the potential of digital care to enhance equitable access to timely, scalable, and personalized pelvic health interventions. By empowering postmenopausal women to manage their pelvic health disorders more effectively, this approach enhances their overall quality of life, while leveling the opportunities to thrive in the workplace.

Acknowledgments

The authors acknowledge the Bloom team of physical therapists and their support to guarantee the successful delivery of care. The authors also acknowledge the contributions of João Tiago Silva, Margarida Morais, Daniel Rodrigues, and Guilherme Freches in data processing (all employees of Sword Health). This work was supported by Sword Health Inc. and also developed within the scope of project nº 62—"Responsible AI," financed by European Funds, namely the Recovery and Resilience Plan—"Componente 5: Agendas Mobilizadoras para a Inovação Empresarial," included in the NextGenerationEU funding program. The study sponsor, Sword Health, was involved in the study design, data collection and interpretation, and writing of the manuscript.

Data Availability

The datasets analyzed during this study are not publicly available due to privacy restrictions but are available from the corresponding author upon reasonable request.

Authors' Contributions

JA, FDC, and FC conceptualized the study. MM and ACA curated the data, while ACA and XT conducted the formal analysis. APP, DJ, ACA, MM, JA, and FC contributed to the interpretation of the data. VB provided the necessary resources for the research. APP, DJ, ACA, and FC drafted the original manuscript, and all authors participated in reviewing and editing the final version. All authors have read and approved the published version of the manuscript.

Conflicts of Interest

The authors declare the following competing interests: APP, DJ, ACA, MM, VB, VY, FDC, JA, and FC are employees of Sword Health, the sponsor of this study. VB, VY and FDC also hold equity in Sword Health, and VB is the CEO of the same company. XT received a scientific advisor honorarium from Sword Health.

XSL•FO
RenderX

Multimedia Appendix 1
Supplementary tables.
[PDF File, 338 KB - mhealth_v13i1e68242_app1.pdf]

References

1. Santoro N, Roeca C, Peters BA, Neal-Perry G. The menopause transition: signs, symptoms, and management options. J Clin Endocrinol Metab 2021 Jan 1;106(1):1-15. [doi: 10.1210/clinem/dgaa764]
2. Gatenby C, Simpson P. Menopause: physiology, definitions, and symptoms. Best Pract Res Clin Endocrinol Metab 2024 Jan;38(1):101855. [doi: 10.1016/j.beem.2023.101855]
3. OlaOlorun FM, Shen W. Menopause: Oxford University Press; 2020.
4. Nelson HD. Menopause. Lancet 2008 Mar 1;371(9614):760-770. [doi: 10.1016/S0140-6736(08)60346-3] [Medline: 18313505]
5. Women in the workplace survey. Biote. 2022. URL: https://biote.com/learning-center/biote-women-in-the-workplace-survey [accessed 2024-10-01]
6. Faubion SS, Enders F, Hedges MS, et al. Impact of menopause symptoms on women in the workplace. Mayo Clin Proc 2023 Jun;98(6):833-845. [doi: 10.1016/j.mayocp.2023.02.025]
7. Pierce H, Perry L, Gallagher R, Chiarelli P. Urinary incontinence, work, and intention to leave current job: a cross sectional survey of the Australian nursing and midwifery workforce. Neurourol Urodyn 2017 Sep;36(7):1876-1883. [doi: 10.1002/nau.23202] [Medline: 28144981]
8. Kołodyńska G, Zalewski M, Rożek-Piechura K. Urinary incontinence in postmenopausal women—causes, symptoms, treatment. Prz Menopauzalny 2019 Apr;18(1):46-50. [doi: 10.5114/pm.2019.84157] [Medline: 31114458]
9. Weintraub AY, Glinter H, Marcus-Braun N. Narrative review of the epidemiology, diagnosis and pathophysiology of pelvic organ prolapse. Int Braz J Urol 2020;46(1):5-14. [doi: 10.1590/S1677-5538.IBJU.2018.0581] [Medline: 31851453]
10. Alperin M, Burnett L, Lukacz E, Brubaker L. The mysteries of menopause and urogynecologic health: clinical and scientific gaps. Menopause 2019 Jan;26(1):103-111. [doi: 10.1097/GME.0000000000001209] [Medline: 30300297]
11. Lo SS, Lim EJ, Ng LG, Kuo TLC. The role of estrogen status in the causation of female lower urinary tract and pelvic floor dysfunction. Curr Bladder Dysfunct Rep 2019 Sep;14(3):139-143. [doi: 10.1007/s11884-019-00523-w]
12. Gandhi J, Chen A, Dagur G, et al. Genitourinary syndrome of menopause: an overview of clinical manifestations, pathophysiology, etiology, evaluation, and management. Am J Obstet Gynecol 2016 Dec;215(6):704-711. [doi: 10.1016/j.ajog.2016.07.045] [Medline: 27472999]
13. Marino JM. Genitourinary syndrome of menopause. J Midwifery Womens Health 2021 Nov;66(6):729-739. [doi: 10.1111/jmwh.13277] [Medline: 34464022]
14. Angelou K, Grigoriadis T, Diakosavvas M, Zacharakis D, Athanasiou S. The genitourinary syndrome of menopause: an overview of the recent data. Cureus 2020 Apr 8;12(4):e7586. [doi: 10.7759/cureus.7586] [Medline: 32399320]
15. Abrams P, Andersson KE, Apostolidis A, et al. 6th International Consultation on Incontinence. Recommendations of the International Scientific Committee: evaluation and treatment of urinary incontinence, pelvic organ prolapse and faecal incontinence. Neurourol Urodyn 2018 Sep;37(7):2271-2272. [doi: 10.1002/nau.23551] [Medline: 30106223]
16. NICE guideline NG123—urinary incontinence and pelvic organ prolapse in women: management. : National Institute for Health and Care Excellence (NICE); 2019 Apr 2 URL: https://www.ncbi.nlm.nih.gov/books/NBK577730/bin/bm3.pdf [accessed 2025-03-24] [doi: 10.1111/bju.14763]
17. Dumoulin C, Cacciari LP, Hay-Smith EJC. Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in women. Cochrane Database Syst Rev 2018 Oct 4;10(10):CD005654. [doi: 10.1002/14651858.CD005654.pub4] [Medline: 30288727]
18. Brækken IH, Majida M, Engh ME, Bø K. Can pelvic floor muscle training reverse pelvic organ prolapse and reduce prolapse symptoms? An assessor-blinded, randomized, controlled trial. Am J Obstet Gynecol 2010 Aug;203(2):170. [doi: 10.1016/j.ajog.2010.02.037]
19. van Reijn-Baggen DA, Han-Geurts IJM, Voorham-van der Zalm PJ, Pelger RCM, Hagenaars-van Miert CHAC, Laan ETM. Pelvic floor physical therapy for pelvic floor hypertonicity: a systematic review of treatment efficacy. Sex Med Rev 2022 Apr;10(2):209-230. [doi: 10.1016/j.sxmr.2021.03.002] [Medline: 34127429]
20. Cardozo L, Rovner E, Wagg A, Wein A, Abrams P, editors. Incontinence, 7th edition: ICI-ICS. International Continence Society.
21. Fall M, Baranowski AP, Elneil S, et al. EAU guidelines on chronic pelvic pain. Eur Urol 2010 Jan;57(1):35-48. [doi: 10.1016/j.eururo.2009.08.020]
22. López-Pérez MP, Afanador-Restrepo DF, Rivas-Campo Y, et al. Pelvic floor muscle exercises as a treatment for urinary incontinence in postmenopausal women: a systematic review of randomized controlled trials. Healthcare (Basel) 2023 Jan 11;11(2):216. [doi: 10.3390/healthcare11020216] [Medline: 36673584]
23. Bertotto A, Schvartzman R, Uchôa S, Wender MCO. Effect of electromyographic biofeedback as an add-on to pelvic floor muscle exercises on neuromuscular outcomes and quality of life in postmenopausal women with stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2017 Nov;36(8):2142-2147. [doi: 10.1002/nau.23258] [Medline: 28508398]

XSL•FO
RenderX

24. Alves FK, Riccetto C, Adami DBV, et al. A pelvic floor muscle training program in postmenopausal women: a randomized controlled trial. Maturitas 2015 Jun;81(2):300-305. [doi: 10.1016/j.maturitas.2015.03.006] [Medline: 25862491]
25. Pereira VS, de Melo MV, Correia GN, Driusso P. Long-term effects of pelvic floor muscle training with vaginal cone in post-menopausal women with urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2013 Jan;32(1):48-52. [doi: 10.1002/nau.22271] [Medline: 22674639]
26. Wiegersma M, Panman C, Kollen BJ, Berger MY, Lisman-Van Leeuwen Y, Dekker JH. Effect of pelvic floor muscle training compared with watchful waiting in older women with symptomatic mild pelvic organ prolapse: randomised controlled trial in primary care. BMJ 2014 Dec 22;349:g7378. [doi: 10.1136/bmj.g7378] [Medline: 25533442]
27. Sran M, Mercier J, Wilson P, Lieblich P, Dumoulin C. Physical therapy for urinary incontinence in postmenopausal women with osteoporosis or low bone density. Menopause 2016;23(3):286-293. [doi: 10.1097/GME.0000000000000594]
28. Malinauskas AP, Bressan EFM, de Melo AMZRP, Brasil CA, Lordêlo P, Torelli L. Efficacy of pelvic floor physiotherapy intervention for stress urinary incontinence in postmenopausal women: systematic review. Arch Gynecol Obstet ;308(1):13-24. [doi: 10.1007/s00404-022-06693-z]
29. Jorge CH, Bø K, Catai CC, Brito LGO, Driusso P, Tennfjord MK. Pelvic floor muscle training as treatment for female sexual dysfunction: a systematic review and meta-analysis. Am J Obstet Gynecol 2024 Jul;231(1):51-66. [doi: 10.1016/j.ajog.2024.01.001] [Medline: 38191016]
30. Torosis M, Carey E, Christensen K, et al. A treatment algorithm for high-tone pelvic floor dysfunction. Obstet Gynecol 2024 Apr 1;143(4):595-602. [doi: 10.1097/AOG.0000000000005536] [Medline: 38387036]
31. Waetjen LE, Xing G, Johnson WO, Melnikow J, Gold EB, Study of Women s Health Across the Nation (SWAN). Factors associated with reasons incontinent midlife women report for not seeking urinary incontinence treatment over 9 years across the menopausal transition. Menopause 2018 Jan;25(1):29-37. [doi: 10.1097/GME.0000000000000943] [Medline: 28763399]
32. Krychman M, Graham S, Bernick B, Mirkin S, Kingsberg SA. The women's EMPOWER survey: women's knowledge and awareness of treatment options for vulvar and vaginal atrophy remains inadequate. J Sex Med 2017 Mar;14(3):425-433. [doi: 10.1016/j.jsxm.2017.01.011] [Medline: 28202319]
33. Toye F, Barker KL. A meta-ethnography to understand the experience of living with urinary incontinence: "is it just part and parcel of life?". BMC Urol 2020 Jan 16;20(1):1. [doi: 10.1186/s12894-019-0555-4] [Medline: 31941470]
34. A Physical Therapy Profile: Wages Earned in the Profession, 2021–2022: American Physical Therapy Association; 2023.
35. Lamin E, Parrillo LM, Newman DK, Smith AL. Pelvic floor muscle training: underutilization in the USA. Curr Urol Rep 2016 Feb;17(2):26757904. [doi: 10.1007/s11934-015-0572-0] [Medline: 26757904]
36. Dieter AA, Wilkins MF, Wu JM. Epidemiological trends and future care needs for pelvic floor disorders. Curr Opin Obstet Gynecol 2015 Oct;27(5):380-384. [doi: 10.1097/GCO.0000000000000200] [Medline: 26308198]
37. Hou Y, Feng S, Tong B, Lu S, Jin Y. Effect of pelvic floor muscle training using mobile health applications for stress urinary incontinence in women: a systematic review. BMC Womens Health 2022 Oct 3;22(1):400. [doi: 10.1186/s12905-022-01985-7] [Medline: 36192744]
38. Bernard S, Boucher S, McLean L, Moffet H. Mobile technologies for the conservative self-management of urinary incontinence: a systematic scoping review. Int Urogynecol J 2020 Jun;31(6):1163-1174. [doi: 10.1007/s00192-019-04012-w] [Medline: 31267139]
39. Santiago M, Cardoso-Teixeira P, Pereira S, Firmino-Machado J, Moreira S. A hybrid-telerehabilitation versus a conventional program for urinary incontinence: a randomized trial during COVID-19 pandemic. Int Urogynecol J 2023 Mar;34(3):717-727. [doi: 10.1007/s00192-022-05108-6] [Medline: 35596802]
40. Wadensten T, Nyström E, Franzén K, Lindam A, Wasteson E, Samuelsson E. A mobile app for self-management of urgency and mixed urinary incontinence in women: randomized controlled trial. J Med Internet Res 2021 Apr 5;23(4):e19439. [doi: 10.2196/19439] [Medline: 33818395]
41. Janela D, Areias AC, Molinos M, et al. Digital care program for urinary incontinence in females: a large-scale, prospective, cohort study. Healthcare (Basel) 2024 Jan 8;12(2):141. [doi: 10.3390/healthcare12020141] [Medline: 38255031]
42. Menopause. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/menopause [accessed 2025-01-08]
43. Bohle B, Belvis F, Vial M, et al. Menopause and obstetric history as risk factors for fecal incontinence in women. Dis Colon Rectum 2011 Aug;54(8):975-981. [doi: 10.1097/DCR.0b013e31821c404a] [Medline: 21730786]
44. NICE guideline NG210 - pelvic floor dysfunction: prevention and non-surgical management. : National Institute for Health and Care Excellence (NICE); 2021 Dec 9 URL: https://www.nice.org.uk/guidance/ng210 [accessed 2025-03-24] [Medline: 35438874]
45. Barber MD, Walters MD, Bump RC. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). Am J Obstet Gynecol 2005 Jul;193(1):103-113. [doi: 10.1016/j.ajog.2004.12.025] [Medline: 16021067]
46. de Figueiredo VB, Ferreira CHJ, da Silva JB, et al. Responsiveness of Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Floor Impact Questionnaire (PFIQ-7) after pelvic floor muscle training in women with stress and mixed urinary incontinence. Eur J Obstet Gynecol Reprod Biol 2020 Dec;255(129-33):129-133. [doi: 10.1016/j.ejogrb.2020.10.039]

XSL•FO
RenderX

47. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
48. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006]
49. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
50. Toussaint A, Hüsing P, Gumz A, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). J Affect Disord 2020 Mar 15;265:395-401. [doi: 10.1016/j.jad.2020.01.032] [Medline: 32090765]
51. Ospina MB, Dennett L, Waye A, Jacobs P, Thompson AH. A systematic review of measurement properties of instruments assessing presenteeism. Am J Manag Care 2015 Feb 1;21(2):e171-e185. [Medline: 25880491]
52. What is a serious adverse event. US Food and Drug Administration. 2023. URL: https://www.fda.gov/safety/reporting-serious-problems-fda/what-serious-adverse-event [accessed 2025-01-08]
53. McNeish D, Matta T. Differentiating between mixed-effects and latent-curve approaches to growth modeling. Behav Res Methods 2018 Aug;50(4):1398-1414. [doi: 10.3758/s13428-017-0976-5] [Medline: 29067672]
54. Xiao J, Bulut O. Evaluating the performances of missing data handling methods in ability estimation from sparse data. Educ Psychol Meas 2020 Oct;80(5):932-954. [doi: 10.1177/0013164420911136] [Medline: 32855565]
55. Pfaffel A, Schober B, Spiel C. A comparison of three approaches to correct for direct and indirect range restrictions: a simulation study. Pract Assess Res Eval 2016;21(1):6. [doi: 10.7275/x4ep-fv42]
56. Butler DC, Petterson S, Phillips RL, Bazemore AW. Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery. Health Serv Res 2013 Apr;48(2 Pt 1):539-559. [doi: 10.1111/j.1475-6773.2012.01449.x] [Medline: 22816561]
57. McNeish D. Relaxing the proportionality assumption in latent basis models for nonlinear growth. Struct Equ Model 2020 Sep 2;27(5):817-824. [doi: 10.1080/10705511.2019.1696201]
58. S0201 selected population profile in the United States. United States Census Bureau. 2022. URL: https://data.census.gov/table?t=-00:-01:Race%20and%20Ethnicity [accessed 2024-10-17]
59. Mercier J, Dumoulin C, Carrier-Noreau G. Pelvic floor muscle rehabilitation for genitourinary syndrome of menopause: why, how and when? Climacteric 2023 Aug;26(4):302-308. [doi: 10.1080/13697137.2023.2194527] [Medline: 37019140]
60. Fitz FF, Gimenez MM, de Azevedo Ferreira L, Matias MMP, Bortolini MAT, Castro RA. Pelvic floor muscle training for female stress urinary incontinence: a randomised control trial comparing home and outpatient training. Int Urogynecol J 2020 May;31(5):989-998. [doi: 10.1007/s00192-019-04081-x] [Medline: 31444535]
61. Sherburn M, Bird M, Carey M, Bø K, Galea MP. Incontinence improves in older women after intensive pelvic floor muscle training: an assessor-blinded randomized controlled trial. Neurourol Urodyn 2011 Mar;30(3):317-324. [doi: 10.1002/nau.20968] [Medline: 21284022]
62. Pereira VS, de Melo MV, Correia GN, Driusso P. Vaginal cone for postmenopausal women with stress urinary incontinence: randomized, controlled trial. Climacteric 2012 Feb;15(1):45-51. [doi: 10.3109/13697137.2011.593211] [Medline: 22066898]
63. Vermandel A, De Wachter S, Beyltjens T, D'Hondt D, Jacquemyn Y, Wyndaele JJ. Pelvic floor awareness and the positive effect of verbal instructions in 958 women early postdelivery. Int Urogynecol J 2015 Feb;26(2):223-228. [doi: 10.1007/s00192-014-2483-x] [Medline: 25062656]
64. Harper RC, Sheppard S, Stewart C, Clark CJ. Exploring adherence to pelvic floor muscle training in women using mobile apps: scoping review. JMIR Mhealth Uhealth 2023 Nov 30;11:e45947. [doi: 10.2196/45947] [Medline: 38032694]
65. Hagen S, Elders A, Stratton S, et al. Effectiveness of pelvic floor muscle training with and without electromyographic biofeedback for urinary incontinence in women: multicentre randomised controlled trial. BMJ 2020 Oct 14;371:m3719. [doi: 10.1136/bmj.m3719] [Medline: 33055247]
66. Kannan P, Cheing GLY, Fung BKY, et al. Effectiveness of pelvic floor muscle training alone or combined with either a novel biofeedback device or conventional biofeedback for improving stress urinary incontinence: a randomized controlled pilot trial. Contemp Clin Trials 2022 Dec;123:106991. [doi: 10.1016/j.cct.2022.106991] [Medline: 36332826]
67. Fitz FF, Stüpp L, da Costa TF, Bortolini MAT, Girão M, Castro RA. Outpatient biofeedback in addition to home pelvic floor muscle training for stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2017 Nov;36(8):2034-2043. [doi: 10.1002/nau.23226] [Medline: 28169458]
68. Liu H, Wu W, Xiang W, Yuan J. Lifestyle factors, metabolic factors and socioeconomic status for pelvic organ prolapse: a Mendelian randomization study. Eur J Med Res 2023;28(1):37287058. [doi: 10.1186/s40001-023-01148-w]
69. Wiegersma M, Panman C, Hesselink LC, et al. Predictors of success for pelvic floor muscle training in pelvic organ prolapse. Phys Ther 2019 Jan 1;99(1):109-117. [doi: 10.1093/ptj/pzy114] [Medline: 30329105]
70. Antônio FI, Herbert RD, Bø K, et al. Pelvic floor muscle training increases pelvic floor muscle strength more in post-menopausal women who are not using hormone therapy than in women who are using hormone therapy: a randomised trial. J Physiother 2018 Jul;64(3):166-171. [doi: 10.1016/j.jphys.2018.05.002]

XSL•FO
RenderX

71. Hendriks EJM, Kessels AGH, de Vet HCW, Bernards ATM, de Bie RA. Prognostic indicators of poor short-term outcome of physiotherapy intervention in women with stress urinary incontinence. Neurourol Urodyn 2010 Mar;29(3):336-343. [doi: 10.1002/nau.20752] [Medline: 19475574]
72. Alblooshi S, Taylor M, Gill N. Does menopause elevate the risk for developing depression and anxiety? Results from a systematic review. Australas Psychiatry 2023 Apr;31(2):165-173. [doi: 10.1177/10398562231165439] [Medline: 36961547]
73. Bromberger JT, Epperson CN. Depression during and after the perimenopause: impact of hormones, genetics, and environmental determinants of disease. Obstet Gynecol Clin North Am 2018 Dec;45(4):663-678. [doi: 10.1016/j.ogc.2018.07.007] [Medline: 30401549]
74. Ai F, Deng M, Mao M, Xu T, Zhu L. Depressive symptoms screening in postmenopausal women with symptomatic pelvic organ prolapse. Menopause 2018 Mar;25(3):314-319. [doi: 10.1097/GME.0000000000001006] [Medline: 29040216]
75. Rodríguez-Almagro J, Hernández Martínez A, Martínez-Vázquez S, Molina RAP, Bermejo-Cantarero A, Martínez-Galiano JM. A qualitative exploration of the perceptions of women living with pelvic floor disorders and factors related to quality of life. J Clin Med 2024 Mar 25;13(7):1896. [doi: 10.3390/jcm13071896] [Medline: 38610661]
76. Molina RAP, Vázquez SM, Galiano JMM, Izquierdo MR, Khan KS, Cano‐Ibáñez N. Prevalence of depression and anxiety in women with pelvic floor dysfunctions: a systematic review and meta‐analysis. Intl J Gynecol Obstet 2024 Nov;167(2):507-528. [doi: 10.1002/ijgo.15719]
77. Han B, Duan Y, Zhang P, et al. Effects of exercise on depression and anxiety in postmenopausal women: a pairwise and network meta-analysis of randomized controlled trials. BMC Public Health 2024;24(1):38977980. [doi: 10.1186/s12889-024-19348-2]
78. Osborne LA, Mair Whittall C, Emery S, Reed P. Change in depression predicts change in bladder symptoms for women with urinary incontinence undergoing pelvic-floor muscle training. Eur J Obstet Gynecol Reprod Biol 2023 Jan;280(54-9):54-59. [doi: 10.1016/j.ejogrb.2022.11.010] [Medline: 36402023]
79. Break through the stigma: menopause in the workplace. Bank of America. URL: https://business.bofa.com/en-us/content/workplace-benefits/menopause-in-the-workplace.html [accessed 2024-10-10]
80. Rodrigo CH, Sebire E, Bhattacharya S, Paranjothy S, Black M. Effectiveness of workplace-based interventions to promote wellbeing among menopausal women: a systematic review. Post Reprod Health 2023 Jun;29(2):99-108. [doi: 10.1177/20533691231177414] [Medline: 37207326]
81. Hardy C, Griffiths A, Norton S, Hunter MS. Self-help cognitive behavior therapy for working women with problematic hot flushes and night sweats (MENOS@Work): a multicenter randomized controlled trial. Menopause 2018 May;25(5):508-519. [doi: 10.1097/GME.0000000000001048] [Medline: 29315132]

Abbreviations

CRAIQ-7: Colo-Rectal-Anal Impact Questionnaire—short form 7
GAD-7: Generalized Anxiety Disorder 7-item scale
LBGA: latent-basis growth analysis
LGCA: latent growth curve analysis
MCIC: minimal clinically important change
PFD: pelvic floor dysfunction
PFIQ-7: Pelvic Floor Impact Questionnaire—short form 7
PFMT: pelvic floor muscle training
PHQ-9: Patient Health 9-item Questionnaire
POPIQ-7: Pelvic Organ Prolapse Impact Questionnaire—short form 7
PT: physical therapist
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
UIQ-7: Urinary Impact Questionnaire—short form 7
WPAI: Work Productivity and Activity Impairment Questionnaire

Edited by L Buis; submitted 31.10.24; peer-reviewed by C Prevett, S Dufour; revised version received 24.02.25; accepted 04.03.25; published 02.04.25.

Please cite as:
Pereira AP, Janela D, Areias AC, Molinos M, Tong X, Bento V, Yanamadala V, Atherton J, Dias Correia F, Costa F
Innovating Care for Postmenopausal Women Using a Digital Approach for Pelvic Floor Dysfunctions: Prospective Longitudinal Cohort Study
JMIR Mhealth Uhealth 2025;13:e68242
URL: https://mhealth.jmir.org/2025/1/e68242
doi:10.2196/68242

XSL•FO
RenderX

© Ana P Pereira, Dora Janela, Anabela C Areias, Maria Molinos, Xin Tong, Virgílio Bento, Vijay Yanamadala, Jennesa Atherton, Fernando Dias Correia, Fabíola Costa. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Effectiveness of the Be Prepared mHealth App on Recovery of Physical Functioning After Major Elective Surgery: Multicenter Randomized Controlled Trial

Miriam van der Velde[1,2], PT, MSc; Marike van der Leeden[3,4], PT, PhD; Edwin Geleijn[3], PT; Cindy Veenhof[1,2], PT, Phd; Karin Valkenet[1,2], PT, PhD; Be Prepared Group[5]

[1]Research Group Innovation of Human Movement Care, Center for Healthy and Sustainable Living, HU University of Applied Sciences, Heidelberglaan 7, Utrecht, The Netherlands

[2]Department of Rehabilitation, Physical Therapy Science and Sports, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[3]Department of Rehabilitation Medicine, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[4]Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, The Netherlands

[5]

Corresponding Author:
Miriam van der Velde, PT, MSc
Research Group Innovation of Human Movement Care, Center for Healthy and Sustainable Living, HU University of Applied Sciences, Heidelberglaan 7, Utrecht, The Netherlands

Abstract

Background: Patients undergoing major surgery are at risk of complications and delayed recovery. Prehabilitation has shown promise in improving postoperative outcomes. Offering prehabilitation by means of mHealth can help overcome barriers to participating in prehabilitation and empower patients prior to major surgery. We developed the Be Prepared mHealth app, which has shown potential in an earlier pilot study.

Objective: This study aims to evaluate the effectiveness of the Be Prepared app on postoperative recovery of physical functioning (PF) in patients undergoing major elective surgery.

Methods: This study was a multicenter randomized controlled trial with 2 arms. Adults scheduled for major elective surgery were randomly assigned to the control (usual care) or intervention group (Be Prepared app in addition to usual care). The Be Prepared app is a smartphone app with pre- and postoperative information and instructions on changing risk behavior for patients undergoing major elective surgery. The primary outcome was recovery of postoperative PF up to 12 weeks after hospital discharge measured with the Computer Adaptive Test Patient-Reported Outcomes Measurement Information System-PF. Secondary outcomes included social participation, self-reported recovery, health-related quality of life, postoperative outcomes, and patient satisfaction. Measurements were performed at 5 time points: before random assignment and 1, 3, 6, and 12 weeks after hospital discharge.

Results: A total of 369 patients were analyzed, 181 in the control group and 188 in the intervention group. The result of the linear mixed effects model showed a mean slope difference in recovery of PF over 12 weeks of 2.97 (95% CI 0.90-5.02) in favor of the intervention group. However, this effect was not clinically relevant and was negated by the significantly lower PF score 1 week after hospital discharge in the intervention group (mean difference –1.72, 95% CI –3.38 to –0.07). Most secondary outcome measures did not show significantly greater improvements in the intervention group compared to the control group. Patient satisfaction with overall perioperative care was significantly higher in the intervention group compared to the control group and satisfaction with the Be Prepared app was high.

Conclusions: The use of the Be Prepared app as a stand-alone intervention does not seem beneficial for improving postoperative recovery in patients undergoing major surgery. However, satisfaction with perioperative care was higher in patients using the app. Given the advantages of digital technology in health care, it can be considered a basis for prehabilitation care pathways, complemented by guidance from health care professionals as needed.

Trial Registration: CCMO Netherlands NL-OMON53078; https://onderzoekmetmensen.nl/en/trial/53078

(JMIR Mhealth Uhealth 2025;13:e58703) doi:10.2196/58703

KEYWORDS

perioperative care; mHealth; prehabilitation; surgery; digital health

XSL•FO
RenderX

Introduction

Patients undergoing major surgery are at risk of adverse postoperative health outcomes such as complications and delayed or poor recovery [1,2]. There is growing evidence showing that prehabilitation can improve postoperative recovery and reduce the incidence of postoperative complications [1-6]. Prehabilitation aims to enhance the patients' general health and well-being prior to major surgery by modifying behavioral and lifestyle risk factors [3]. With prehabilitation, patients can better withstand the forthcoming stressor of surgery and thus minimize the risk of postoperative complications and prolonged hospital stay and enhance recovery after surgery [2-4,6]. For patients, recovery of physical functioning (PF) and well-being are the most highly valued goals following surgery [7].

Prehabilitation is preferably multimodal and involves the optimization of the physical, nutritional, and mental status of patients undergoing major elective surgery [2,4,6]. Prehabilitation interventions can vary widely, including in terms of modalities, context (home-, community-, or hospital-based), target population, and degree of supervision. Qualitative evidence shows that many patients undergoing major surgery prefer home-based prehabilitation [4,8,9]. Home-based prehabilitation helps overcome various barriers to participation in prehabilitation: it resolves transportation and parking issues, it makes it easier for patients to combine prehabilitation with their everyday life, and patients with physical or psychological symptoms might be more comfortable at home [8]. Using a mobile health (mHealth) app as a home-based strategy for prehabilitation could be an effective approach and is in line with current developments in health care in which digital technologies play a crucial role in providing sustainable, efficient, and patient-centered health care [10,11]. In addition, mHealth can positively influence self-efficacy and empower patients prior to major surgery [12-14].

We developed the Be Prepared mHealth app for patients undergoing major surgery. The Be Prepared app is a smartphone app with pre- and postoperative information and instructions to prepare patients for major surgery. The app is multimodal and focuses on changing risk behavior (ie, smoking and alcohol cessation, increasing physical activity and muscle strengthening activities, and protein-rich food consumption). The first version of the Be Prepared app has proven potential in a pilot study in terms of usability, patient satisfaction, and modifying risk behavior prior to surgery [15]. Several points of improvement were identified in this pilot study and have been addressed as part of the further development of the app prior to this study. These included expanding the preoperative content, extending the content postoperatively, and further personalization of the Be Prepared app [15].

The primary aim of this multicenter randomized controlled trial (RCT) was to evaluate the effectiveness of the Be Prepared mHealth app on the recovery of postoperative PF, when added to usual care, in a population of patients undergoing major elective surgery. The secondary aim was to evaluate the effectiveness of the Be Prepared mHealth app on social participation, self-reported recovery, health-related quality of life, postoperative outcomes, and patient satisfaction.

Methods

Design and Participants

This study was conducted as a single-blind parallel multicenter RCT. Patients were eligible for participation if they were scheduled for elective surgery with an expected postoperative hospital stay of at least 2 nights. Patients had to be aged 18 years or older and have one or more risk behaviors (ie, currently smoking, ≥7 alcohol consumptions per week, moderate-intensity physical activity <30 min every day, muscle-strengthening activities on <2 days a week, and unintentional weight loss of >3 kg in the last month). Patients were excluded if they had insufficient comprehension of the Dutch language, had no access to a mobile device, had to undergo brain surgery, had their surgery scheduled within 7 days, were participating in conflicting studies, or in an intensive preoperative care pathway (eg, an exercise program).

Recruitment and Blinding

Four centers participated in the trial: 3 university hospitals and 1 general hospital in the Netherlands. Potentially eligible patients were identified at the preoperative assessment at the participating hospitals and received an invitation letter to participate in the study. After providing informed consent and completing the baseline questionnaire, patients were randomly assigned in a 1:1 ratio to the control group (usual care) or the intervention group (usual care plus the Be Prepared mHealth app) using a web-based randomization system. All patients received an automated email of their group allocation. Patients were aware of their group allocation, but researchers analyzing the data and health care professionals at the hospital were kept blinded to the allocation. Patients in the intervention group filled in the Physical Activity Readiness Questionnaire (PAR-Q) before starting the intervention to determine whether they could safely perform physical activity [16,17]. Patients who were scheduled to receive instructions on increasing physical activity and muscle strengthening through the app were contacted by a member of the research team when answering yes to one or more questions in the PAR-Q. These patients were advised, when necessary, to contact their physician before commencing the intervention.

Intervention

All patients received usual care. Usual care varied widely, depending on hospital guidelines and the respective care pathway. In most cases, preoperative care consisted of a preoperative consultation with the anesthesiologist (digital, by telephone, or face-to-face) and information about general preoperative preparation through leaflets. A single care pathway, for example, the esophageal cancer care pathway, includes a preoperative consultation with a dietician or a physiotherapist. Counseling on smoking or alcohol cessation is not a standard part of any care pathway.

Patients in the intervention group had access to the Be Prepared mHealth app (Patient Journey platform by Interactive Studios BV), in addition to usual care. After randomization, patients

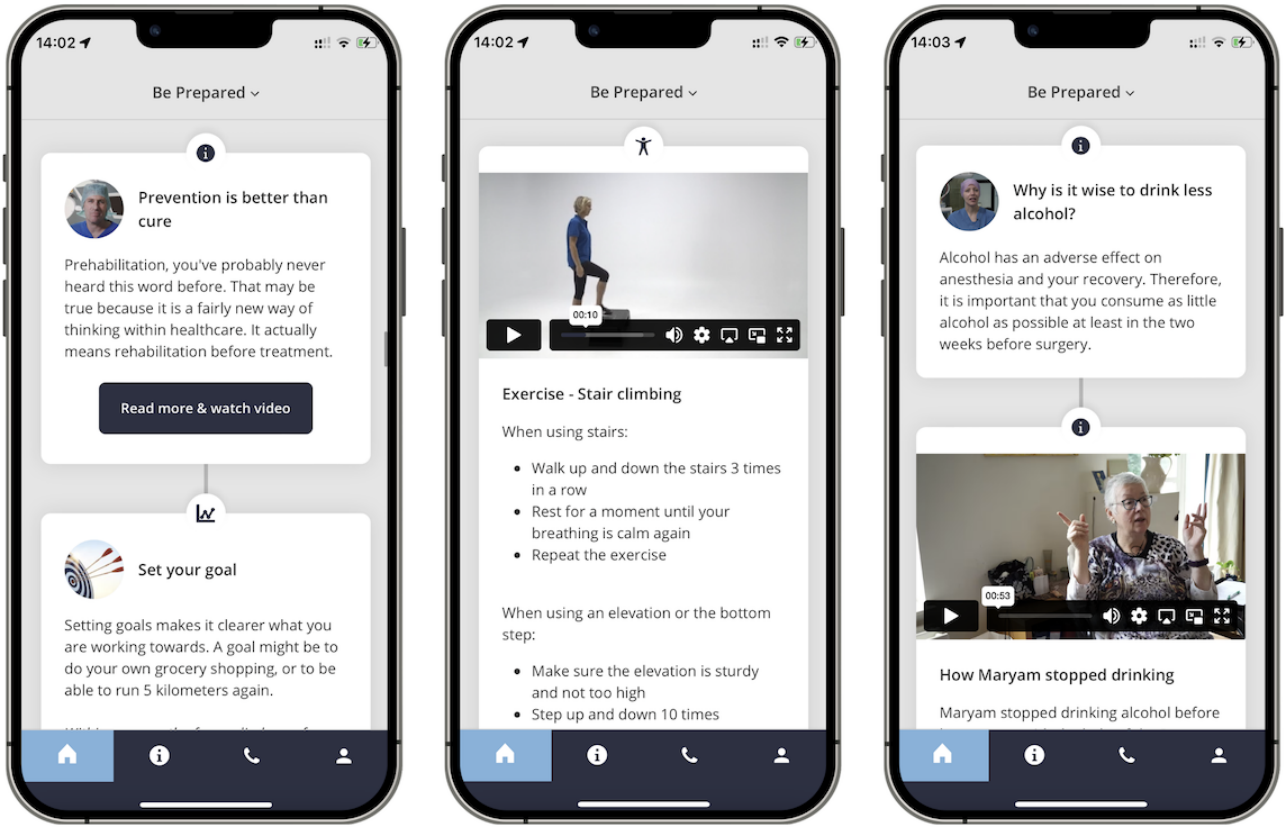
XSL•FO
RenderX

received an automated email with a link to download the app. The Be Prepared app is a smartphone app with pre- and postoperative information and instructions on changing risk behavior for patients undergoing major surgery. The Be Prepared app was deployed as a stand-alone intervention, and no health care professionals were involved during the use of the app. The content presented in the app was tailored to the patients' risk profile, meaning patients only received content on the risk behaviors present. At the start of the intervention and again after hospital discharge, patients answered screening questions in the app, regarding risk behavior, to determine their risk profile. The content of the app differed according to the risk profile, and focused on smoking and alcohol cessation, increasing physical activity and muscle-strengthening activities according to the recommendations of the Dutch Health Council, and protein-rich food consumption. Furthermore, the mHealth app included content on general preparation (eg, preoperative fasting and medication intake prior to surgery), stress, and practical issues (eg, transportation and postoperative care).

Patients were advised to use the app in the run-up to their surgery, during their hospital stay, and in the first weeks after hospital discharge. The day-to-day information and instructions were presented on a dynamic timeline based on the patient's operation date. The timeline provided written information, information videos, patient-experience videos, tips on healthy behavior and changing risk behavior, quiz questions, and exercise videos. Several behavior change techniques were used in this app, such as goal setting and feedback on behavior [18]. Push notifications informed the patient about available new content. The amount of content depended on the patients' risk profile.

Furthermore, participants were asked whether they succeeded in following the instructions and received automated feedback based on their responses. When patients repeatedly were unsuccessful in following the instructions, they were advised to contact a health care professional of their choice for personal counseling; there were no health care professionals affiliated with the study. In addition, the patient's feedback was used to tailor exercises to the patient's level. Based on the patient's feedback, the level of exercises was maintained, scaled up, or scaled down. Data collected by the app and app usage are stored anonymously in a database of Interactive Studios (ISO 27001 and NEN7510 certified). There was no integration with the hospital's electronic health record. Screenshots of the app can be found in Figure 1. An overview of the intervention content and features can be found in Multimedia Appendix 1.

Figure 1. Screenshots of the Be Prepared app (translated from Dutch).



Compliance

Through user authentication, the use of the mHealth intervention was registered by date and time: for example, opening the app, answering in-app questions, and opening information. Self-reported app use was registered by asking patients in the web-based questionnaire if they had used the app regularly, occasionally, or not at all.

Outcome Measures

Patients completed online questionnaires at 5 time points: before random assignment (T0), 1 week (T1), 3 weeks (T2), 6 weeks (T3), and 12 weeks (T4) after hospital discharge. The timing of

XSL•FO
RenderX

the baseline measurements (T0) relative to the surgery date varied. At baseline, demographic and clinical characteristics were collected through questionnaires and electronic health records (ie, age, sex, BMI, the American Society of Anesthesiologists (ASA) physical status classification, surgical specialty, waiting time for surgery, risk behaviors, level of education, and Charlson Comorbidity Index).

The primary outcome was the recovery of PF (from 1 to 12 weeks after hospital discharge) assessed by the Computer Adaptive Test (CAT) Patient-Reported Outcomes Measurement Information System-Physical Function (PROMIS-PF). The CAT PROMIS-PF is a computer-administered measure with questions that are selected by a computer algorithm, based on a patient's response and their estimated health state, from the PROMIS-PF item bank (version 1.2). The item bank contains 121 items, covering a wide range of (everyday) activities from self-care to more complex activities. Standard PROMIS CAT stopping rules were used [19]. The minimal important difference (MID) was estimated using a distribution-based model (0.5 SD at baseline) [20]. The estimated MID for this sample was 4.18.

The secondary outcomes included PF, social participation, self-reported recovery, and health-related quality of life at the different time points after surgery, the course of social participation, self-reported recovery, and health-related quality of life, in-hospital physical and mental symptoms, self-reported risk behavior change, in-hospital mobilization, postoperative outcomes, and patient satisfaction (Table 1). Scores on CAT PROMIS-PF, CAT PROMIS Ability to Perform Social Roles and Activities (APS), and European Quality of Life 5 Dimensions 3 Level Version were calculated according to published scoring algorithms [19,21-23].

XSL•FO
RenderX

Table . Primary and secondary outcome measures.

Outcome	Instrument or method	Item information	Scoring	Interpretation	Time points[a]
Primary outcome measure					
PF[b]	CAT[c] PROMIS-PF[d] (v1.2) [19,21]	• 4-12 questions from the item bank containing 121 items • 5-point response scale from 1 to 5	T-score metric (mean 50, SD 10)	Higher is better outcome	T0, T1, T2, T3, T4
Secondary outcome measures					
Social participation	CAT PROMIS-APS[e] (version 2.0) [21]	• 4-12 questions from the item bank containing 35 items • 5-point response scale from 1 to 5	T-score metric (mean 50, SD 10)	Higher is better outcome	T0, T1, T2, T3, T4
Self-reported recovery	Self-reported recovery (study specific)	• 11-point response scale from 0=not recovered to 10=fully recovered	0 - 10	Higher is better outcome	T1, T2, T3, T4
Health-related quality of life	EQ-5D-3L[f] [22,23]	• 5 items • Response scale from 1=no problems to 3=extreme problems	Health state index score 0 - 1	Higher is better outcome	T0, T1, T2, T3, T4
In-hospital physical and mental symptoms	In-hospital physical and mental symptoms (study specific)	• 3 items on physical symptoms • 5-point response scale from never to extremely often	0 - 12	Lower is better outcome	T1
		• 4 items on mental symptoms • 5-point response scale from never to extreme often	0 - 16		
In-hospital mobilization	In-hospital mobilization (study specific)	• 3 items: sitting in chair, hospital room ambulation, and hallway ambulation • 6-point response scale from 0=day of surgery to 5=day 5 or later.	0 - 15	Lower is better outcome	T1
Self-reported risk behavior change	Risk behavior change: stopped smoking, stopped drinking alcohol, increased physical activity, increased muscle strengthening activities, and increased protein rich food consumption. (2-point scale; yes-no; study specific)	• 5 items • Binary response scale (yes or no)			T1

XSL•FO
RenderX

Outcome	Instrument or method	Item information	Scoring	Interpretation	Time points[a]
Postoperative outcomes	Total and severe postoperative complication rates (EHR[g])	• Severity of complications are graded according to Clavien-Dindo classification • Grade I (least severe) to grade V (death) [24]	I-V	Clavien-Dindo grade III or higher is classified as a severe complication	Within 30-days after surgery
	Length of hospital stay (EHR)	• Days between admission and discharge from hospital			
	Hospital readmission rates (EHR)				Within 30-days after surgery
Patient satisfaction	Satisfaction with perioperative care	• NRS[h] from not satisfied at all to very satisfied	0 - 10	Higher is better outcome	T1
	Satisfaction with the Be Prepared app (intervention group)	• NRS from not satisfied at all to very satisfied	0 - 10	Higher is better outcome	T1
	Net Promotor Score [25] (intervention group)	• Rating between 0 and 10 • The score is calculated by subtracting % of detractors (0 - 6) from % of promotors (9-10)	−100 to 100	Positive total NPS is good	T1

[a]T0: before random assignment, T1: 1 week, T2: 3 weeks, T3: 6 weeks, and T4: 12 weeks after hospital discharge.

[b]PF: physical functioning.

[c]CAT: Computer Adaptive Test.

[d]PROMIS-PF: Patient-Reported Outcomes Measurement Information System-Physical Functioning.

[e]PROMIS-APS: Patient-Reported Outcomes Measurement Information System-Ability to Perform Social Roles and Activities.

[f]EQ-5D-3L: European Quality of Life 5 Dimensions 3 Level Version.

[g]EHR: electronic health record.

[h]NRS: Numeric Rating Scale.

Sample Size

A sample size calculation was performed on the primary outcome: the postoperative recovery of PF. To achieve a power of 0.85, with a 2-sided significance level of .05, 402 patients would be needed (201 per arm) to detect a mean standardized effect of 0.30 (ie, small to moderate effect) between the intervention and usual care according to an independent Student *t* test (2-tailed). As patients are assessed at multiple time points and analyzed accordingly, thus increasing the amount of information, the estimated sample size is conservative. In addition, we assumed a dropout rate of 0.15 and aimed to include 480 patients.

Statistical Analysis

Characteristics of the study population are summarized per treatment group using descriptive statistics. Primary and secondary outcomes were analyzed according to the modified intention-to-treat principle, meaning that we only included participants who completed at least 1 follow-up measurement [26,27].

The differences between groups on the postoperative course of the primary and secondary longitudinal outcomes were analyzed using linear mixed effects models. The fixed part of the models included a term for the baseline score of the respective outcome (except for the model for self-reported recovery), time, group allocation, and the interaction between group allocation and time. The random part of the model contained a random intercept and slope for time per individual. Treatment effectiveness was determined by the statistical significance of the time by group interaction, using a likelihood ratio test with 1 degree of freedom, indicating whether the rate of change over time was different in the intervention or control group. The difference in rate of change over time has been expressed as a difference in change over 12 weeks to ease interpretation. No imputation

XSL•FO
RenderX

techniques were used, as missing data were addressed by the chosen modeling strategies.

Three sensitivity analyses were conducted for the primary outcome. To explore the impact of missing data, the primary outcome, in participants without any follow-up measurement, was once imputed with the lowest score and once with the highest score measured at that time point. To assess possible center effects, we performed a sensitivity analysis, including a 3-way interaction between group allocation, time, and center, which was assessed using a likelihood ratio test [28]. In addition, we performed a per-protocol analysis for the primary outcome in which only data from participants who activated the app, answered the in-app screening questions, and used the app at least once after the first login, were included.

A mixed-effects model for repeated measures was used to analyze the between-group differences in the longitudinal outcomes at each postoperative time point using similar random and fixed effects.

We estimated odds ratios (ORs) and their CIs between the intervention and control group of the binary outcomes (overall complications, severe complications, and hospital readmissions) using logistic regression. A Cox proportional hazards model was used to estimate hazard ratios and their CIs for length of hospital stay. A Mann-Whitney *U* test was performed to compare the in-hospital physical and mental symptoms sum scores, in-hospital mobilization sum scores, and patient satisfaction with perioperative care. A chi-square test was used to compare self-reported risk behavior change between groups.

Subgroup analyses were performed, on the primary outcome, to explore if the effects of treatment might differ across specific groups of patients. Subgroup analyses were specified a priori in the study protocol and are exploratory and hypothesis-generating. We compared the following subgroups: intervention duration (<14 days, 14 - 28 days, or >28 days), surgery type (oncological or nononcological), and number of risk behaviors (≤2 risk behaviors or ≥3 risk behaviors).

Analyses were performed using IBM SPSS for Windows version 28.0 and RStudio (version 2022.07.2; R Foundation for Statistical Computing). A significance level of α=.05 was applied for all tests.

Ethical Considerations

Written informed consent was obtained from all participating patients. This study has been approved by the Medical Ethical Committee of Amsterdam University Medical Center, location VUmc, (registration NL61503.029.18), and has been registered at the Overview of Medical Research in the Netherlands (NL-OMON53078). Overview of Medical Research in the Netherlands is an official data provider to the International Clinical Trial Registry Platform of the World Health Organization. All data collected in the trial were deidentified and stored in a secured cloud storage environment. There was no compensation for the participants. The CONSORT (Consolidated Standards of Reporting Trials) guidelines were followed.

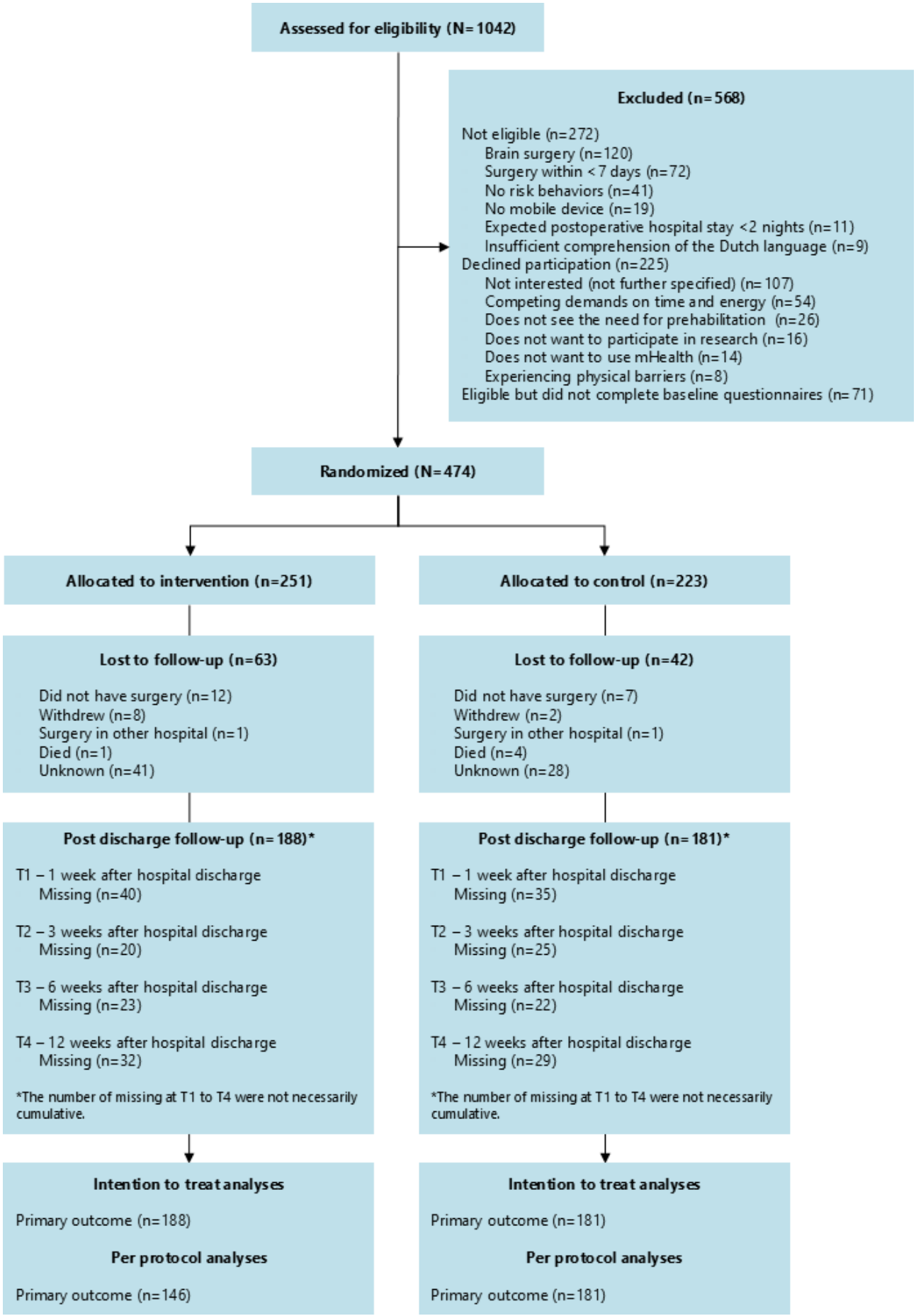
Results

Participant Flow

Between June 2020 and June 2022, 1042 patients were assessed for eligibility. In total, 225 patients declined participation and 272 patients did not meet the inclusion criteria. The reasons for nonparticipation are mentioned in Figure 2. Furthermore, 71 patients were eligible but did not complete the baseline questionnaire and were therefore not randomized. Finally, 474 patients were randomly assigned to the intervention (n=251) or control (n=223) group of whom 369 (78%) completed at least 1 follow-up questionnaire and were therefore included in the final analyses (188 patients in the intervention group and 181 in the control group). When comparing characteristics of participants included in the analysis (N=369) to those who only completed baseline questionnaires (n=105), the participants not included in the final analyses scored lower on the baseline scores of the CAT PROMIS-PF (mean difference 3.21, 95% CI 1.36-5.03) and CAT PROMIS-APS (mean difference=3.57, 95% CI 1.26-5.87) and more frequently had an ASA classification of 4 (7% vs 1%).

XSL•FO
RenderX

Figure 2. Flow of participants through the trial.



XSL•FO
RenderX

Patient Characteristics

Participants had a mean age of 62 (SD 11.9) years, 60.4 % were male (223/369), and 41.5% (153/369) had a high level of education. Most participants underwent gastrointestinal surgery (124/369, 33.6%) or cardiothoracic surgery (103/369, 27.9%). The median waiting time for surgery was 27 (IQR 14 - 51) days. Patient characteristics per group are shown in Table 2. The number of included patients per participating center varied between 7 and 246. No intervention-related adverse events were reported during the trial.

XSL•FO
RenderX

Table . Baseline characteristics of participants.

	Modified intention-to-treat	
Characteristics	Intervention (n=188)	Control (n=181)
Age (years), mean (SD)	63.3 (11.3)	60.1 (12.5)
Sex (male), n (%)	111 (59)	112 (61.9)
BMI (kg/m^2), median (IQR)	25.4 (23.2-29.1)	25.5 (23.4-28.7)
ASA[a] classification, n (%)		
I	14 (7.4)	14 (7.7)
II	87 (46.3)	93 (51.4)
III	76 (40.4)	63 (34.8)
IV	3 (1.6)	2 (1.1)
Unknown	8 (4.3)	9 (5)
Surgical specialty, n (%)		
Gastrointestinal	58 (31)	66 (37)
Cardiothoracic	55 (29.3)	48 (26.5)
Urology and gynecology	29 (15.4)	20 (11)
Orthopedic and spine	21 (11.2)	13 (7.2)
Head and neck	10 (5.3)	20 (11.0)
Vascular	10 (5.3)	9 (5)
Plastic	4 (2.1)	4 (2.2)
Neurosurgical	1 (0.5)	1 (0.6)
Surgical oncology, n (%)	99 (52.7)	101 (55.8)
Waiting time for surgery[b] (days), median (IQR)	26 (14-53)	27 (14-50)
Risk behaviors[c], n (%)		
Smoking	16 (8.5)	16 (8.8)
Alcohol consumption	20 (10.6)	12 (6.6)
Physical inactivity	145 (77.1)	144 (79.6)
No muscle strengthening activities	173 (92)	173 (95.6)
Unintentional weight loss	25 (13.3)	24 (13.3)
Number of risk behaviors, n (%)		
1	29 (15.4)	25 (13.8)
2	133 (70.7)	130 (71.8)
3	21 (11.2)	20 (11)
4	4 (2.1)	6 (3.3)
5	1 (0.5)	—[d]
Level of education[e], n (%)		
Low	46 (24.5)	50 (27.6)
Intermediate	68 (36.2)	45 (24.9)
High	70 (37.2)	83 (45.9)
Unknown	4 (2.1)	3 (1.7)
CCI[f], n (%)		
Mild (1-2)	96 (51.1)	93 (51.4)
Moderate (3-4)	35 (18.6)	24 (13.3)

XSL•FO
RenderX

	Modified intention-to-treat	
Characteristics	Intervention (n=188)	Control (n=181)
Severe (≥5)	15 (8)	18 (9.9)
CAT[g] PROMIS-PF[h], mean (SD)	44.94 (8.36)	45.68 (8.35)
CAT PROMIS-APS[i], mean (SD)	50.50 (11.08)	49.91 (11.02)
EQ-5D-3L[j], mean (SD)	0.79 (.20)	0.83 (.18)

[a]ASA: American Society of Anesthesiologists.

[b]Days between baseline measurement and surgery.

[c]Multiple response options.

[d]Not applicable.

[e]Low: preschool, primary school, lower vocational education; intermediate: secondary education, intermediate vocational education; high: higher vocational education, university, postgraduate.

[f]CCI: Charlson Comorbidity Index.

[g]CAT: Computer Adaptive Test.

[h]PROMIS-PF: Patient-Reported Outcomes Measurement Information System-Physical Functioning.

[i]PROMIS-APS: Patient-Reported Outcomes Measurement Information System-Ability to Perform Social Roles and Activities.

[j]EQ-5D-3L: European Quality of Life 5 Dimensions 3 Level Version.

Compliance

Of the 188 patients in the intervention group, 157 (83.5%) activated the app. In total, 155 patients (82.4%) answered the in-app screening questions, and 146 patients (77.7%) used the app at least once after first login. Therefore, 146 patients in the intervention group were included in the per-protocol analysis (Figure 2). The median number of days on which patients opened the app during the perioperative period was 25 days (IQR 7.5-39.0). When looking at the self-reported app use, 158 patients completed the questionnaire. A total of 91 patients (57.6%) indicated to have used the app regularly, 37 patients (23.4%) used the app occasionally, and 30 patients (19%) did not use the app.

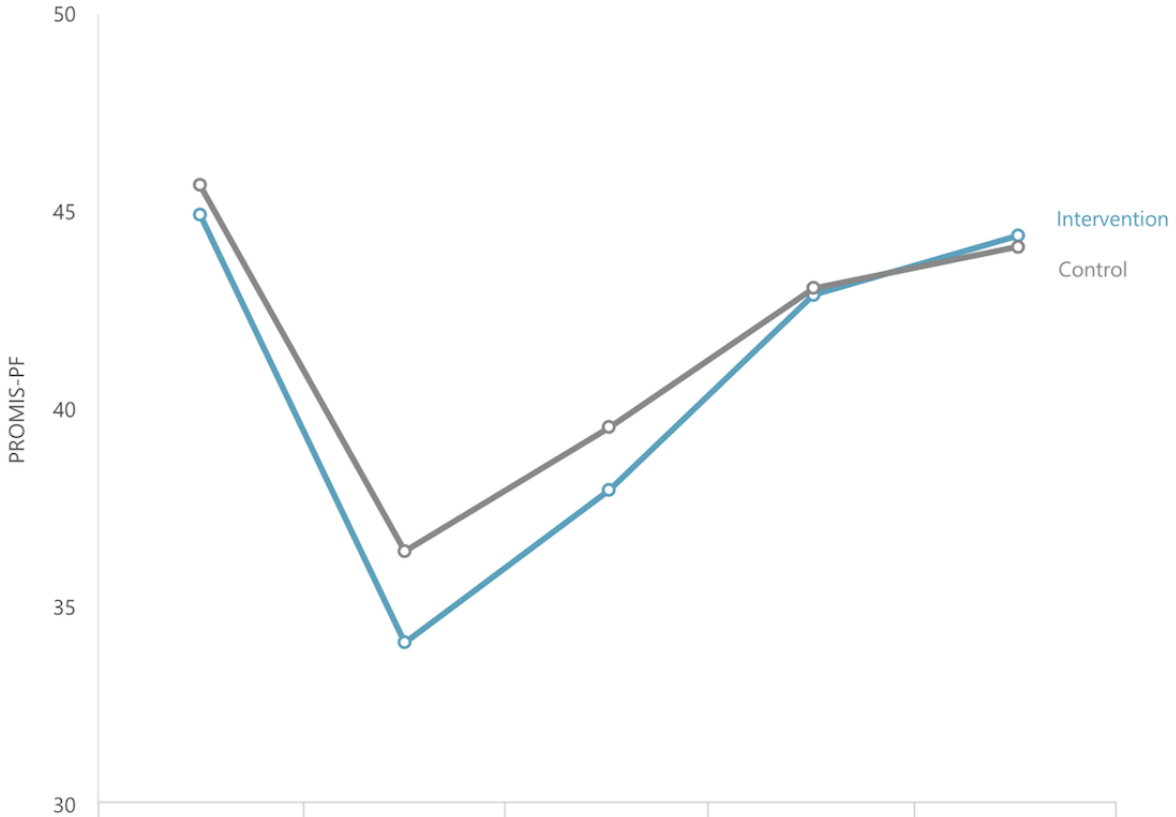
Physical Functioning

The course of PF per group and the means and SDs per time point are shown in Figure 3. Results of the linear mixed effects models are shown in Table 3. The linear mixed effects models revealed that in both groups PF improved from 1 week after hospital discharge to 12 weeks after hospital discharge. The intervention group increased by 10.93 points over 12 weeks versus 7.96 points in the control group (time effect), resulting in a mean slope difference between groups (time by group interaction) of 2.97 over 12 weeks (95% CI 0.90-5.02; *P*=.005) in favor of the intervention group (Table 3 and Multimedia Appendix 2). The sensitivity analyses exploring the impact of missing data did not lead to different results. Imputation of missing data with the lowest score measured at that time point led to a mean slope difference of 2.60 (95% CI 0.53-4.65; *P*=.014) while imputation with the highest score led to a mean slope difference of 2.42 (95% CI 0.32-4.51; *P*=.023). In addition, sensitivity analysis for possible center effects showed that the time by group interaction was not different for each center (*P*=.749). The per-protocol analysis displayed similar results as the modified intention-to-treat analysis (mean slope difference=3.58, 95% CI 1.39-5.76; *P*=.001).

Per time point, the mean PF score at 1 week after hospital discharge (adjusted for baseline) was significantly lower in the intervention group than in the control group (mean slope difference=−1.72, 95% CI −3.38 to −0.07; *P*=.042). No significant differences at other time points were found (Multimedia Appendix 2).

XSL•FO
RenderX

Figure 3. The course of PF (PROMIS-PF) across all time points (observed mean values and SDs). PF: physical functioning; PROMIS-PF: Patient-Reported Outcomes Measurement Information System-Physical Functioning; T0: before random assignment; T1: 1 week; T2: 3 weeks; T3: 6 weeks; T4: 12 weeks after hospital discharge.



	T0	T1	T2	T3	T4
Control	45.68 (8.35) n=181	36.43 (8.84) n=146	39.56 (8.52) n=156	43.08 (8.18) n=159	44.13 (8.84) n=152
Intervention	44.94 (8.36) n=188	34.12 (7.80) n=148	37.98 (8.17) n=168	42.91 (8.10) n=165	44.41 (7.65) n=156

XSL•FO
RenderX

Table . Mean slope differences between groups (time by group interaction) over 12-weeks for the primary and secondary longitudinal outcome measures.

	Intervention (n=188), mean	Control (n=181), mean	Mean slope difference between groups over 12 weeks (95% CI)	*P* value
Primary outcome				
CAT[a] PROMIS-PF[b] (version 1.2)[c]	10.92	7.95	2.97 (0.90 to 5.02)	.005[d]
Secondary outcome				
CAT PROMIS-APS[e] (version 2.0)[c]	12.37	8.86	3.51 (−0.60 to 7.61)	.094
Self-reported functional recovery (0 - 10)	3.29	2.63	0.66 (0.10 to 1.23)	.022[d]
EQ-5D-3L[fc]	0.16	0.14	0.02 (−0.03 to 0.08)	.391

[a]CAT: Computer Adaptive Test.

[b]PROMIS-PF: Patient-Reported Outcomes Measurement Information System-Physical Functioning

[c]Adjusted for baseline value of the outcome measure.

[d]Statistically significant *P*<.05.

[e]PROMIS-APS: Patient-Reported Outcomes Measurement Information System-Ability to Perform Social Roles and Activities

[f]EQ-5D-3L: European Quality of Life 5 Dimensions 3 Level Version

Secondary Longitudinal Outcomes

No significant differences were found between groups on the postoperative recovery of social participation (3.51, 95%CI −0.60 to 7.61; *P*=.094) and health-related quality of life (0.02, 95% CI −0.03 to 0.08; *P*=.391). On the course of postoperative self-reported recovery (scale 0 - 10), a significant difference was found favoring the intervention group (0.66, 95% CI 0.10 to 1.23, *P*=.022) (see Table 3).

The mean scores for social participation and self-reported recovery were significantly lower in the intervention group than in the control group at 1 and 3 weeks after hospital discharge, indicating a higher level of functioning in the control group. Social participation scores were 2.89 points (95%CI −4.93 to −0.86; *P*=.006) lower in the intervention group at 1 week and −3.12 points (95% CI −5.14 to −1.11; *P*=.003) lower at 3 weeks after hospital discharge. Self-reported recovery scores were −0.49 points (95% CI −0.98 to −0.01; *P*=.048) lower in the intervention group at 1 week and −0.58 points (95% CI −1.03 to −0.13; *P*=.012) lower at 3 weeks after hospital discharge. No significant differences at other time points were found for any of the secondary longitudinal outcomes (Multimedia Appendix 2).

Postoperative Outcomes

No significant between-group differences were found for overall complications (OR 1.10, 95% CI 0.73-1.67; *P*=.658) and severe complications (OR 0.57, 95% CI 0.30-1.08, *P*=.085). In addition, no between-group differences were identified for length of stay (hazard ratio 0.95, 95% CI 0.78-1.17; *P*=.628), and hospital readmissions (OR 0.89, 95% CI 0.42-1.90; *P*=.764; Multimedia Appendix 2).

No significant differences were found for in-hospital physical and mental symptoms and mobilization (Multimedia Appendix 2). Preoperatively, a larger proportion of participants in the intervention group increased their muscle-strengthening activities compared to the control group (77% vs 35%; *P*<.001). Preoperative changes in other risk behaviors were similar for both the intervention and control groups (Multimedia Appendix 2).

Patient Satisfaction

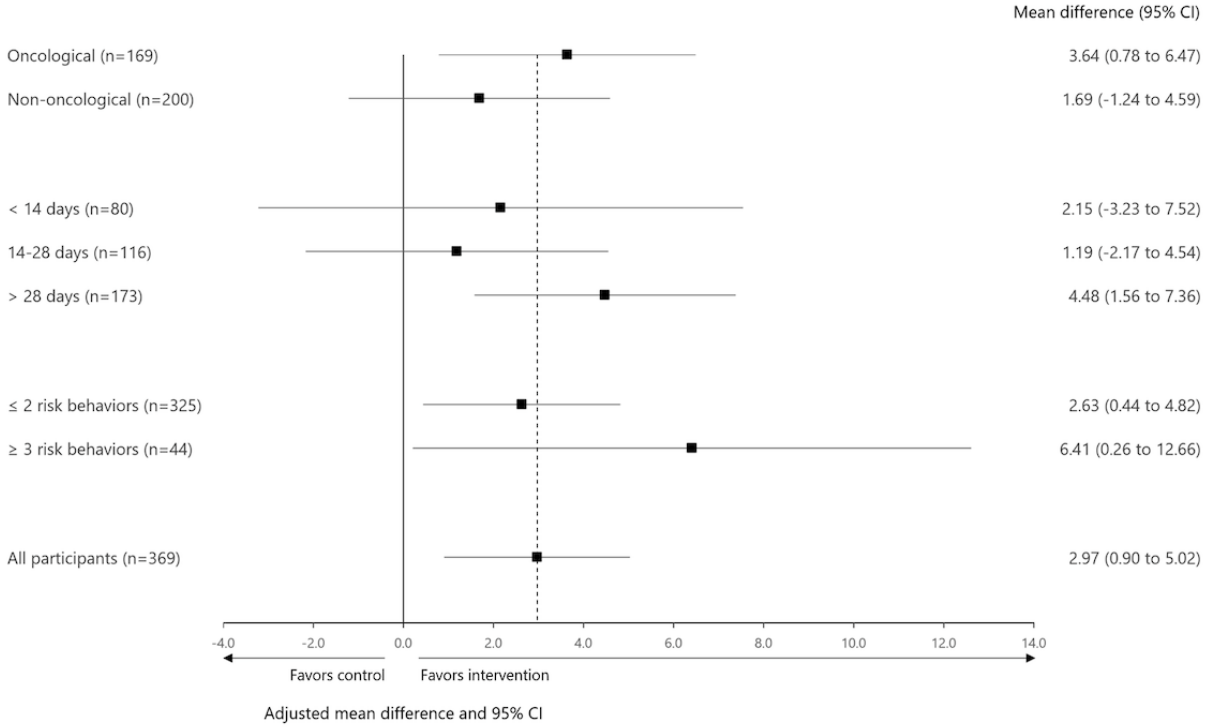
Participants in the intervention group gave a higher rating for satisfaction with overall perioperative care than participants in the control group (median 8.2, IQR 7.3-9.0 vs mean 7.8, IQR 6.6-8.7; *P*=.041). The median satisfaction score with the Be Prepared app was 8.1 (IQR 6.6-9.1). The Net Promotor Score for the Be Prepared app was 19, which can be considered good.

Subgroup Analyses

Subgroup analyses on the primary outcome resulted in similar mean slope differences between groups (Figure 4). However, the subgroup analyses do show that treatment effects seem larger in some subgroups (ie, participants undergoing oncological surgery, participants with a preoperative preparation of more than 28 days, and participants with 3 or more risk behaviors).

XSL•FO
RenderX

Figure 4. Forest plot of the subgroup analyses on the primary outcome of recovery of PF measured with Promis-PF. Bold line shows no effect point and dotted line shows overall treatment effect point. PF: physical functioning; PROMIS-PF: Patient-Reported Outcomes Measurement Information System-Physical Functioning.



Discussion

Principal Findings

In this multicenter RCT, the effectiveness of the Be Prepared mHealth intervention was evaluated in comparison with usual care in people undergoing major elective surgery. This study shows that the use of the Be Prepared app led to a small statistically significant improvement in the postoperative recovery of PF from 1 week to 12 weeks after hospital discharge, compared to patients receiving usual care. However, as this improvement is smaller than the estimated MID of 4.18, it can not be considered clinically meaningful. In addition, the intervention group showed a significantly greater decline in PF at 1 week after hospital discharge which seems to negate the effect on the postoperative recovery of PF. Most secondary outcome measures did not show significantly greater improvements in the intervention group compared to the control group, but patients in the intervention group scored significantly higher on self-reported recovery. Patient satisfaction with the overall perioperative care was significantly higher in the intervention than in the control group and participants in the intervention group expressed high satisfaction with the Be Prepared app.

Our hypothesis was that the Be Prepared intervention would improve the preoperative level of PF of patients and that they would therefore be able to better withstand the impact of major surgery, resulting in a smaller decrease of PF and a better and faster recovery after surgery. Our results, however, are in contrast with this hypothesis as the analysis of differences between groups at the individual postoperative time points (adjusted for baseline scores) showed that patients in the intervention group had a significantly lower level of PF at 1 week after hospital discharge than patients in the control group.

The limited effect found in this study could possibly be attributed to the characteristics of the Be Prepared intervention. First, the Be Prepared intervention was deployed as a stand-alone intervention without the involvement or guidance of health care professionals. Evidence suggests that supervised prehabilitation has a greater effect and higher adherence than unsupervised prehabilitation programs [5]. The Be Prepared intervention relies heavily on the initiative and discipline of the patient. In this study, only 58% of the patients indicated to have used the app regularly, and 19% did not use the app at all. With this evidence, it is questionable whether this approach, without the involvement of a health care professional, is sufficient to improve the postoperative course of recovery in every patient [29].

A second reason could be that the exercise component of the Be Prepared intervention does not provide sufficient stimuli to actually increase functional capacity and with that postoperative recovery. A recently published randomized clinical trial, the PREHAB trial, found beneficial effects on complications and postoperative recovery of a 4-week supervised multimodal prehabilitation program, consisting of high-intensity exercise, nutritional and psychological support, and smoking cessation, compared to usual care before colorectal cancer surgery [4]. While the Be Prepared intervention is also a multimodal intervention, the exercise component of the program focuses on increasing physical activity and muscle-strengthening activities according to the recommendations of the Dutch Health Council as opposed to high-intensity exercise training, as used in the aforementioned study. Given the current results, we can

XSL•FO
RenderX

question whether the total stimulus in the Be Prepared app was sufficient to actually improve functional capacity, and with that improve the postoperative recovery in this patient group.

Third, within the Be Prepared app, no objective monitoring of progression, for example by activity trackers, was used for personalization and adjustment of treatment. Literature shows that adequate personalization at commencement of the program, and objective monitoring of progression to adjust treatment are essential [5,30]. Within the Be Prepared app, personalization was pursued based on the screening questions answered by the patient on commencement. Further adjustment of treatment based on progression was limited. For example, muscle strengthening exercises were scaled up or down based on patient-reported feedback in terms of perceived exertion on the proposed exercises in the app. The limited personalization and monitoring could have influenced adherence to the intervention and with that the effectiveness [29].

Besides the characteristics of the intervention, the selection of participants may also have played a role in the limited effect found in this study. The fact that mHealth might not be suitable for all patients [31,32] and that this suitability was not used as an inclusion criterion, may have resulted in the nonuse of the intervention and a smaller contrast between the groups than expected. Given this insight, a patient preference trial, instead of an RCT, might have been a better alternative [33,34].

Unfortunately, the COVID-19 pandemic impacted this study. The changed surgical planning resulted in longer waiting times or, on the contrary, short-term surgical planning, and surgeries were sometimes postponed several times. This might have impacted the effectiveness of the intervention as the app was not designed to adapt to changing surgery dates. As the information is presented in the app based on the surgery date, adjusting the surgery date would lead to the patient being presented with previously shown information.

The use of digital technologies is considered to have great potential and is an increasingly researched topic in the field of health care [10,11]. To our knowledge, this is the first study to perform a multicenter RCT evaluating the effect of a multimodal prehabilitation app on the postoperative course of PF in people undergoing major surgery. A recent systematic review on the use of digital technologies to support home-based prehabilitation prior to major surgery concluded that the use of technologies is feasible and has high acceptability [10]. However, none of the included studies evaluated postoperative recovery outcomes and the results of PF outcomes were inconclusive. A possible explanation for these inconclusive findings is the heterogeneity in the studied characteristics and outcomes, which hampers a clear comparison. For example, in our study, we used a smartphone app while other RCTs in this research area have mainly focused on telehealth or web-based interventions. Only one of the RCTs included in this review investigated a smartphone app as part of an eHealth intervention [35]. The main focus of the intervention was on managing recovery expectations and providing postoperative guidance. The study showed that the eHealth intervention was effective in decreasing the time taken to return to normal activities after intermediate-grade abdominal surgery. The magnitude of the effect on physical function over time, however, was comparable with the results found in our study.

An interesting finding, though one that is based on subgroup analyses, was the trend that patients with a more complex profile (patients with three or more risk behaviors and patients undergoing oncological surgery) seem to have a larger effect on the intervention. This argues for including especially high-risk patients for this type of intervention. In addition, patients with a preparation time longer than 28 days seemed to have a better effect of the intervention. Even though findings from previous research suggest that people's engagement with technology wanes over time [36], the result from this subgroup analysis suggests that sufficient preparation time is important to reach an intervention effect. This also raises the question of whether the minimum intervention duration of 7 days in this study was adequate. The results of these subgroup analyses are all exploratory but can inform future research.

Despite the limited effect of the Be Prepared app on postoperative outcomes, patient satisfaction in the intervention group is high. Patients using the Be Prepared app are more satisfied with overall perioperative care and express high satisfaction with the app, which is an important finding as patient satisfaction is known to be a critical indicator of health care quality and the expected outcomes of care [14,37]. The usability and satisfaction with the app were explored using quantitative and qualitative data in the pilot RCT of the Be Prepared app [15]. In addition, the Be Prepared intervention can be considered a safe intervention for patients undergoing major elective surgery. Prior to using the Be Prepared mHealth intervention, patients were screened using the PAR-Q to determine whether they could safely participate in physical activity. Based on the PAR-Q results no individual adjustments to the intervention have been made. During the intervention, there was no supervision or monitoring by a health care professional and no intervention-related adverse events have taken place.

Strengths and Limitations

This study has several important strengths, including a thorough process of designing, prototyping and evaluating the intervention, a large sample size, multicenter participation, and the use of both objective and self-reported outcomes.

Nevertheless, this study also has limitations.

First, almost a third of the eligible patients declined to participate in this study. While this is not unique and it is a commonly reported finding in (exercise) trials, it does raise questions about the generalizability of the results of this study to the target population [38]. As we could not collect any data on nonparticipants, except for reasons for nonparticipation, we are not able to compare their characteristics to the study participants.

Second, the engagement metrics available in this study were limited to basic usage logs, such as the date and time of app openings. Due to these limitations, we were unable to conduct a comprehensive analysis of app use which could have helped to understand the impact of app use on outcomes.

In addition, we did not reach our prespecified sample size and the dropout rate was higher than expected. Dropout could have

XSL•FO
RenderX

introduced a selection bias in the results. Participants who only completed baseline questionnaires differed in their baseline score on the PROMIS-PF, CAT PROMIS-APS, and ASA classification, from participants who completed at least 1 follow-up. This may indicate that participants with a lower level of functioning were more likely to drop out.

In addition, given the increasing focus on the importance of preoperative preparation and prehabilitation in research, health care but also in mainstream media in the Netherlands, we cannot rule out the possibility that this has influenced usual care and diluted the contrast between the 2 groups in our trial, which could have led to an underestimation of the effect of the intervention.

Conclusions

In this study, the Be Prepared app shows no added value regarding postoperative recovery and clinical outcomes. The Be Prepared app led to a statistically significant but not clinically relevant improvement in the postoperative recovery of PF up to 12 weeks after hospital discharge. In addition, we found a significantly greater decline in PF in the intervention group at 1 week after surgery. In light of these findings, implementation of the Be Prepared app as a stand-alone intervention does not appear to be beneficial for improving postoperative recovery. However, the Be Prepared app is a safe intervention, and patients using the Be Prepared app are more satisfied with overall perioperative care and express high satisfaction with the app. Given these benefits, mHealth can be considered a foundation for prehabilitation care pathways, but to be effective in improving postoperative recovery, preoperative mHealth apps should probably include objective monitoring and integration of mHealth with supervised prehabilitation according to patients' needs and preferences. In addition, careful selection of patients with a preference for mHealth could increase adherence, and with that effectiveness.

Acknowledgments

The authors thank all the patients who participated in this study. The authors thank everyone who contributed to the content of the app. The authors would also like to thank all the involved staff from the participating centers: University Medical Centre Utrecht, Utrecht, the Netherlands; Dijklander Hospital, Hoorn, the Netherlands; Amsterdam University Medical Center, location AMC, Amsterdam, the Netherlands; and Amsterdam University Medical Center, location VUmc (the coordinating center), Amsterdam, the Netherlands. This study received funding from the Taskforce for Applied Research SIA. The funder had no role in the design or conduct of the study; collection, management, analysis, or interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

The Be Prepared Group includes: Ruben PA van Eijk, Marije Marsman, Juul WM van Grootel, Martine Botjes, Lisa A Cohen, Liedewij MJ Janssen, Donald L van der Peet, Jens-Peter Hering, Jelle P Ruurda, and Jeroen Hermanides.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MvdV contributed to conceptualization, data curation, methodology, formal analysis, investigation, visualization, writing–original draft, and writing–review and editing. KV and MvdL contributed to conceptualization, methodology, supervision, funding acquisition, and writing–review and editing. EG contributed to conceptualization, methodology, investigation, funding acquisition, and writing–review and editing. CV contributed to conceptualization, methodology, funding acquisition, and writing–review and editing. RvE contributed to formal analysis, visualization, and writing–review and editing. MM, JvG, MB, JPH, and LC contributed to investigation, resources, and writing–review and editing. LJ, DvdP, JR, and JH contributed to resources and writing–review and editing. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of the intervention content and features.
[PDF File, 213 KB - mhealth_v13i1e58703_app1.pdf]

Multimedia Appendix 2
Supplementary tables.
[DOCX File, 30 KB - mhealth_v13i1e58703_app2.docx]

Checklist 1
CONSORT-eHEALTH checklist (V1.6.1).

XSL•FO
RenderX

[PDF File, 1127 KB - mhealth_v13i1e58703_app3.pdf]

References

1. Tew GA, Ayyash R, Durrand J, Danjoux GR. Clinical guideline and recommendations on pre-operative exercise training in patients awaiting major non-cardiac surgery. Anaesthesia 2018 Jun;73(6):750-768. [doi: 10.1111/anae.14177] [Medline: 29330843]
2. Scheede-Bergdahl C, Minnella EM, Carli F. Multi-modal prehabilitation: addressing the why, when, what, how, who and where next? Anaesthesia 2019 Jan;74 Suppl 1(Suppl 1):20-26. [doi: 10.1111/anae.14505] [Medline: 30604416]
3. Durrand J, Singh SJ, Danjoux G. Prehabilitation. Clin Med (Lond) 2019 Nov;19(6):458-464. [doi: 10.7861/clinmed.2019-0257] [Medline: 31732585]
4. Molenaar CJL, Minnella EM, Coca-Martinez M, et al. Effect of multimodal prehabilitation on reducing postoperative complications and enhancing functional capacity following colorectal cancer surgery: the PREHAB randomized clinical trial. JAMA Surg 2023 Jun 1;158(6):572-581. [doi: 10.1001/jamasurg.2023.0198] [Medline: 36988937]
5. Thomas G, Tahir MR, Bongers BC, Kallen VL, Slooter GD, van Meeteren NL. Prehabilitation before major intra-abdominal cancer surgery: a systematic review of randomised controlled trials. Eur J Anaesthesiol 2019 Dec;36(12):933-945. [doi: 10.1097/EJA.0000000000001030] [Medline: 31188152]
6. McDonald S, Yates D, Durrand JW, et al. Exploring patient attitudes to behaviour change before surgery to reduce peri-operative risk: preferences for short- vs. long-term behaviour change. Anaesthesia 2019 Dec;74(12):1580-1588. [doi: 10.1111/anae.14826] [Medline: 31637700]
7. Myles PS. More than just morbidity and mortality – quality of recovery and long-term functional recovery after surgery. Anaesthesia 2020 Jan;75 Suppl 1(Suppl 1):e143-e150. [doi: 10.1111/anae.14786] [Medline: 31903564]
8. van der Velde M, van der Leeden M, Geleijn E, Veenhof C, Valkenet K. What moves patients to participate in prehabilitation before major surgery? A mixed methods systematic review. Int J Behav Nutr Phys Act 2023 Jun 21;20(1):75. [doi: 10.1186/s12966-023-01474-6] [Medline: 37344902]
9. Ferreira V, Agnihotram RV, Bergdahl A, et al. Maximizing patient adherence to prehabilitation: what do the patients say? Support Care Cancer 2018 Aug;26(8):2717-2723. [doi: 10.1007/s00520-018-4109-1] [Medline: 29478189]
10. Blumenau Pedersen M, Saxton J, Birch S, Rasmussen Villumsen B, Bjerggaard Jensen J. The use of digital technologies to support home-based prehabilitation prior to major surgery: a systematic review. Surgeon 2023 Dec;21(6):e305-e315. [doi: 10.1016/j.surge.2023.05.006] [Medline: 37336649]
11. Senbekov M, Saliev T, Bukeyeva Z, et al. The recent progress and applications of digital technologies in healthcare: a review. Int J Telemed Appl 2020;2020:8830200. [doi: 10.1155/2020/8830200] [Medline: 33343657]
12. Elbert NJ, van Os-Medendorp H, van Renselaar W, et al. Effectiveness and cost-effectiveness of eHealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110. [doi: 10.2196/jmir.2790] [Medline: 24739471]
13. Vonk Noordegraaf A, Anema JR, van Mechelen W, et al. A personalised eHealth programme reduces the duration until return to work after gynaecological surgery: results of a multicentre randomised trial. BJOG 2014 Aug;121(9):1127-1135. [doi: 10.1111/1471-0528.12661] [Medline: 24511914]
14. Timmers T, Janssen L, van der Weegen W, et al. The effect of an app for day-to-day postoperative care education on patients with total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323. [doi: 10.2196/15323] [Medline: 31638594]
15. van der Velde M, Valkenet K, Geleijn E, et al. Usability and preliminary effectiveness of a preoperative mHealth app for people undergoing major surgery: pilot randomized controlled trial. JMIR Mhealth Uhealth 2021 Jan 7;9(1):e23402. [doi: 10.2196/23402] [Medline: 33410758]
16. Goodman JM, Thomas SG, Burr J. Evidence-based risk assessment and recommendations for exercise testing and physical activity clearance in apparently healthy individuals. Appl Physiol Nutr Metab 2011 Jul;36 Suppl 1(Suppl 1):S14-S32. [doi: 10.1139/h11-048] [Medline: 21800940]
17. Thomas SG, Goodman JM, Burr JF. Evidence-based risk assessment and recommendations for physical activity clearance: established cardiovascular disease. Appl Physiol Nutr Metab 2011 Jul;36 Suppl 1(Suppl 1):S190-S213. [doi: 10.1139/h11-050] [Medline: 21800942]
18. Michie S, Carey RN, Johnston M, et al. From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. Ann Behav Med 2018 May 18;52(6):501-512. [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
19. Abma IL, Butje BJD, Ten Klooster PM, van der Wees PJ. Measurement properties of the Dutch-Flemish patient-reported outcomes measurement information system (PROMIS) physical function item bank and instruments: a systematic review. Health Qual Life Outcomes 2021 Feb 24;19(1):62. [doi: 10.1186/s12955-020-01647-y] [Medline: 33627157]
20. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003 May;56(5):395-407. [doi: 10.1016/s0895-4356(03)00044-1] [Medline: 12812812]
21. Terwee CB, Roorda LD, de Vet HCW, et al. Dutch–Flemish translation of 17 item banks from the Patient-Reported Outcomes Measurement Information System (PROMIS). Qual Life Res ;2014(23):1733-1741. [doi: 10.1007/s11136-013-0611-6]

XSL•FO
RenderX

22. Group E. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9]
23. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
24. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg 2004 Aug;240(2):205-213. [doi: 10.1097/01.sla.0000133083.54934.ae] [Medline: 15273542]
25. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54. [Medline: 14712543]
26. Dossing A, Tarp S, Furst DE, et al. Modified intention-to-treat analysis did not bias trial results. J Clin Epidemiol 2016 Apr;72(72):66-74. [doi: 10.1016/j.jclinepi.2015.11.003] [Medline: 26562052]
27. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ 2010 Jun 14;340(340):c2697. [doi: 10.1136/bmj.c2697] [Medline: 20547685]
28. Thabane L, Mbuagbaw L, Zhang S, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med Res Methodol 2013 Jul 16;13:92. [doi: 10.1186/1471-2288-13-92] [Medline: 23855337]
29. Deslippe AL, Soanes A, Bouchaud CC, et al. Barriers and facilitators to diet, physical activity and lifestyle behavior intervention adherence: a qualitative systematic review of the literature. Int J Behav Nutr Phys Act 2023 Feb 14;20(1):14. [doi: 10.1186/s12966-023-01424-2] [Medline: 36782207]
30. de Leeuwerk ME, Bor P, van der Ploeg HP, et al. The effectiveness of physical activity interventions using activity trackers during or after inpatient care: a systematic review and meta-analysis of randomized controlled trials. Int J Behav Nutr Phys Act 2022 May 23;19(1):59. [doi: 10.1186/s12966-022-01261-9] [Medline: 35606852]
31. Swinkels ICS, Huygens MWJ, Schoenmakers TM, et al. Lessons learned from a living lab on the broad adoption of eHealth in primary health care. J Med Internet Res 2018 Mar 29;20(3):e83. [doi: 10.2196/jmir.9110] [Medline: 29599108]
32. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
33. Torgerson DJ, Sibbald B. Understanding controlled trials: what is a patient preference trial? BMJ 1998 Jan 31;316(7128):360. [doi: 10.1136/bmj.316.7128.360] [Medline: 9487173]
34. Kowalski CJ, Mrdjenovich AJ. Patient preference clinical trials: why and when they will sometimes be preferred. Perspect Biol Med 2013;56(1):18-35. [doi: 10.1353/pbm.2013.0004] [Medline: 23748524]
35. van der Meij E, Anema JR, Leclercq WKG, et al. Personalised perioperative care by e-health after intermediate-grade abdominal surgery: a multicentre, single-blind, randomised, placebo-controlled trial. Lancet 2018 Jul 7;392(10141):51-59. [doi: 10.1016/S0140-6736(18)31113-9] [Medline: 29937195]
36. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883. [doi: 10.1177/2055207619839883] [Medline: 30944728]
37. Sixma HJ, Kerssens JJ, Campen CV, Peters L. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. Health Expect 1998 Nov;1(2):82-95. [doi: 10.1046/j.1369-6513.1998.00004.x] [Medline: 11281863]
38. Chinn DJ, White M, Howel D, Harland JOE, Drinkwater CK. Factors associated with non-participation in a physical activity promotion trial. Public Health (Fairfax) 2006 Apr;120(4):309-319. [doi: 10.1016/j.puhe.2005.11.003] [Medline: 16473376]

Abbreviations

APS: Ability to Perform Social Roles and Activities
ASA: American Society of Anesthesiologists
CAT: Computer Adaptive Test
MID: minimal important difference
OR: odds ratio
PAR-Q: Physical Activity Readiness Questionnaire
PF: physical functioning
PROMIS: Patient-Reported Outcomes Measurement Information System
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by L Buis, TDA Cardoso; submitted 22.03.24; peer-reviewed by HR Abdullah, T Groenveld; revised version received 06.02.25; accepted 27.03.25; published 30.05.25.

Please cite as:
van der Velde M, van der Leeden M, Geleijn E, Veenhof C, Valkenet K, Be Prepared Group
The Effectiveness of the Be Prepared mHealth App on Recovery of Physical Functioning After Major Elective Surgery: Multicenter Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e58703
URL: https://mhealth.jmir.org/2025/1/e58703
doi:10.2196/58703

© Miriam van der Velde, Marike van der Leeden, Edwin Geleijn, Cindy Veenhof, Karin Valkenet, Be Prepared Group. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy of a Digital Postoperative Rehabilitation Intervention in Patients With Primary Liver Cancer: Randomized Controlled Trial

Kaitao Yu[1*], MD; Baobing Yin[2,3*], PhD; Ying Zhu[2*], PhD; Hongdao Meng[4], PhD; Wenwei Zhu[2], PhD; Lu Lu[2], PhD; Junqiao Wang[5], Prof Dr; Shugeng Chen[1], PhD; Jun Ni[6], PhD; Yifang Lin[1], PhD; Jie Jia[1,6], PhD

[1]Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, Shanghai, China

[2]Department of General Surgery, Hepatobiliary Surgery, Huashan Hospital, Fudan University, Shanghai, China

[3]Department of Hepatobiliary Surgery, National Regional Medical Center, Binhai Campus of the First Affiliated Hospital, Fujian Medical University, Fuzhou, China

[4]College of Behavioral & Community Sciences, School of Aging Studies, University of South Florida, Florida, FL, United States

[5]School of Nursing, Fudan University, Shanghai, China

[6]Department of Rehabilitation Medicine, National Regional Medical Center, Binhai Campus of the First Affiliated Hospital, Fujian Medical University, Fuzhou, China

[*]these authors contributed equally

Corresponding Author:
Jie Jia, PhD
Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, Shanghai, China

Abstract

Background: Rehabilitation is considered a fundamental component of cancer treatment, especially for patients undergoing cancer surgery. In contrast to conventional rehabilitation education, digital rehabilitation has the potential to improve patients' access to postoperative rehabilitation programs. While digital health has rapidly emerged to aid patients with various diseases, their clinical efficacy in the recovery of patients with primary liver cancer (PLC) undergoing hepatectomy remains inadequately investigated.

Objective: This study aims to evaluate whether a digital postoperative rehabilitation intervention is efficient in improving physical fitness, enhancing exercise adherence, and alleviating fatigue among patients with PLC after hepatectomy.

Methods: A randomized controlled trial was undertaken across 2 university-affiliated hospitals in Eastern China. A total of 100 participants were enrolled in this study and were allocated randomly to either the digital health (intervention group, n=50) or the rehabilitation manual-based group (control group, n=50) at a 1:1 ratio. Patients were unblinded and prospectively followed for the intervention of 3 weeks. Outcome measures included physical fitness, exercise adherence, and status of fatigue.

Results: Overall, 91 out of 100 patients completed the research and were evaluated after 3 weeks of intervention. The digital health group showed better cardiopulmonary endurance than the control group. The mean difference in the change of 6-minute walk test distance from baseline between the groups was 70.21 (95% CI 0.730-82.869) m (P=.05). No statistically significant effects were found for grip strength, 5-repetition-sit-to-stand test time, and fatigue. The exercise adherence in the digital health group was higher than that in the control group (χ^2_2=15.871, P<.001).

Conclusions: The findings suggested that the implementation of digital health had a positive impact on recovery in exercise capacity after hepatectomy. In addition, rehabilitation exercise mode based on digital health has the potential to improve the exercise adherence of patients with PLC compared to conventional manual-based rehabilitation guidance.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2100052911; https://www.chictr.org.cn/showproj.html?proj=135351

(JMIR Mhealth Uhealth 2025;13:e59228) doi:10.2196/59228

KEYWORDS

digital health; surgery; exercise rehabilitation; randomized controlled trial; primary liver cancer

Introduction

Primary liver cancers (PLCs), including hepatocellular carcinoma and intrahepatic cholangiocarcinoma, are associated with significant morbidity and mortality [1]. Over the years, advanced hepatic surgical techniques have led to the development of safe and effective therapies for patients with PLC. Nonetheless, the complex procedure of hepatic resection results in a decline in physical fitness and a high incidence of postoperative adverse physical symptoms such as fatigue, which hampers quality of life and disrupts daily activities [2,3].

XSL•FO
RenderX

Increasing evidence suggests that a combination of multimodal exercise interventions can improve physical fitness and function for patients with PLC after hepatectomy [4]. In addition, enhanced recovery after surgery is mostly confined to the perioperative period. Patients face problems in self-management after discharge and thus have difficulty in exercise adherence [5], which is the extent to which individuals undertake prescribed behavior accurately and at the agreed frequency, intensity, and duration [6]. Physical inactivity has been described as the greatest public health threat. It is estimated that as few as 26% of men and 19% of women adhere to the guidelines [7]. Despite this, patients with PLC may also feel ambiguous about the content of rehabilitation and have difficulty understanding how to balance exercise and discomfort due to the complexity of surgical incisions and side effects [8]. In patients undergoing hepatectomy, the type, intensity, and amount of exercise rehabilitation must be adapted to their individual performance levels [9]. Therefore, advanced interventions of rehabilitation and real-time feedback are important to increase exercise adherence, which may eventually contribute to improving physical fitness and reducing adverse symptoms. In 2021, over 1.028 billion Chinese individuals, accounting for 73% of the Chinese population, accessed the internet via mobile phones. Digital health is an important medium for the delivery of rehabilitation interventions based on mobile devices, providing health education to improve the awareness of liver cancer rehabilitation, assisting in self-management and improving compliance with liver cancer rehabilitation, real-time exercise monitoring to improve the safety of discharged liver cancer rehabilitation, which could serve as a viable and accessible support platform. It significantly enhances outreach to the extensive oncological patient population [10].

There have been no studies that examined the clinical efficacy of digital health-based exercise programs for patients with PLC after hepatectomy during the discharge transition period in China. To measure the impact of exercise rehabilitation based on digital health, we hypothesized that exercise rehabilitation supervised through digital health would be more efficient than traditional unsupervised interventions like manual-based exercise education in improving physical outcomes for patients with PLC.

Methods

Study Design

A prospective, randomized controlled trial was conducted in 2 university-affiliated hospitals (Huashan Hospital of Fudan University in Shanghai and The First Hospital of Fujian Medical University Hospital in Fujian, China). Randomization was 1:1 into the digital health group (intervention group) or the rehabilitation manual-based group (control group). They were requested to complete 2 circles of assessments with baseline as the first time before surgery and subsequent assessments after 3 weeks' intervention.

Study Participants

This study was intended for patients with PLC after hepatectomy. Participants aged 35 to 75 years who met the following inclusion criteria were approached to participate: (1) patients with PLC expected to undergo hepatectomy 2 - 4 liver segments, progressive PLC (stage 2 or 3) for open or laparoscopic hepatectomy; (2) Child-Pugh A; (3) patients with an oncology history of PLC that has not recurred within 3 years after surgery; and (4) ability to operate a mobile phone. Notably, patients were excluded if they had exercise, psychological, or cognitive disorders before hepatectomy.

Appropriate participants were jointly identified by the chief rehabilitation therapists and chief surgeons. Participants were given ample time to consider and voluntarily choose to participate in the study, with written informed consent obtained the day before surgery. Before obtaining consent, researchers thoroughly explained the research procedures and detailed the preoperative and follow-up data to be collected in the ward. The methods and specific data content were clearly outlined in the informed consent form. All participants received standard postoperative care in the operating room, postanesthesia care unit, and ward.

Intervention

The R Plus Health app (Recovery Plus), a medical device cleared by the Chinese FDA, served as the underlying software and hardware for the intervention in this trial, including the basic app structure and the target heart rate belt to guide and supervise training exercises. Based on this, a 3-week exercise program specifically designed for patients with PLC after hepatectomy was developed by an oncology exercise rehabilitation team, which consisted of 3 hepatobiliary surgical oncologists, 2 rehabilitation doctors, 2 nurses, and 3 physiotherapists. After the preliminary literature review and expert demonstration meeting, a set of exercise rehabilitation plans for patients with PLC surgery after discharge was discussed [11]. Reference to the framework in the guide for "FITT Principle" (Exercise prescriptions include the 4 elements of exercise frequency, intensity, time, and type), aerobic exercise, resistance exercise, and stretching training are in the exercise type. The exercise frequency was 30 - 50 minutes each time, 3 - 5 times per week, and 90 - 150 minutes each week. The oncology rehabilitation program consists of 4 parts: warm-up movements, whole-body exercises, stretching exercises, and finishing exercises. Exercise intensity is recommended to choose low-intensity movements during the warm-up and finishing stages, and moderate intensity during full-body activities. The details of the 4 modules of the oncology rehabilitation program are described in Table 1. This program was subsequently integrated into the app by the software manufacturer.

XSL•FO
RenderX

Table . Objectives of intervention sessions for exercise rehabilitation.

Stage	Rehabilitation
Warm-up movements	(1）Sitting abdominal breathing: 1 set for 60 s; rest (30 s)
	(2）Still squat against the wall: 1 set for 60 s; rest (30 s)
	(3）Step in place: 1 set for 60 s; rest (30 s)
	(4）Front kick stepping: 1 set for 60 s; rest (30 s)
	(5）Side leg raises with chair: 1 set for 60 s; rest (30 s)
Stretching sessions	(1) Seated quadriceps stretch: 2 sets for 60 s; rest (30 s)
Whole body sessions	(1）Brisk walking training: 2 sets for 1120 s; rest (30 s)
	(2）Standing elastic band curls: 2 sets for 160 s; rest (30 s)
	(3）Micro squats against the wall: 2 sets for 160 s; rest (30 s)
	(4）Sitting-standing core stabilization: 2 sets for 160 s; rest (30 s)
	(5）Chair alternating side leg raises: 2 sets for 160 s; rest (30 s)
	(6）Sitting deltoid stretch (both sides): 2 sets 1120 s; rest (30 s)
	(7）Seated elastic band shoulder flexion: 2 sets for 160 s; rest (30 s)
	(8）Brisk walking training: 2 set for 1120 s; rest (30 s)
Finishing sessions	(1）Sitting abdominal breathing: 1 set for 60 s; rest (30 s)
	(2）Step in place: 1 set for 60 s; rest (30 s)
	(3）Micro squats against the wall: 1 set for 60 s; rest (30 s)
	(4）Side leg raises with a chair: 1 set for 60 s; rest (30 s)

Digital Rehabilitation Group

Participants in the intervention group were provided with a QR code to download the app onto their mobile phones (available for both Android and iOS) before discharge. Upon approval by the researchers, participants were wirelessly connected to a chest-worn heart rate monitor, which measured exercise frequency, intensity, duration, and progression and ensured safety. The exercise intensity was performed at 64% - 76% of the age-predicted maximal heart rate (220-age) [12]. When participants first logged into the app, a tutorial guide video was displayed and saved in the tutorial section of their personal account. The exercise prescription was posted to each participant before discharge, teaching them how to follow the video action training. The content of the training was consistent with intervention sessions for exercise rehabilitation (Table 2). Every time when patients finished the rehabilitation prescription, the APP provided the questionnaire scale (Borg Breathing Scale) and the option for the patient to record the discomfort symptoms and feelings, combined with the record of the heart rate during exercise, which was transmitted to the doctor's platform in time. The doctors adjusted the order and number of movements in the exercise program according to the patient's information (Figures 1 and 2). In this trial, different patients had varying exercise capacities. Weaker patients could choose to skip movements they were unable to complete, and the system informed the doctor to make appropriate adjustments. Patients initially unable to complete 30 - 50 minutes could exercise in smaller increments throughout the day but were encouraged to exercise for 90 - 150 minutes per week. The software automatically uploaded the amount of exercise and daily exercise time.

XSL•FO
RenderX

Table . Baseline characteristics of study participants by treatment.

	Control (N=50)	Digital health (N=50)	Total (N=100)	*P* value
Age (years)				.39
N	50	50	100	
Mean (SD)	56.4 (9.7)	54.6 (10.6)	55.5 (10.2)	
Female, n (%)				.16
0[a]	40 (80.0)	45 (90.0)	85 (85.0)	
1[b]	10 (20.0)	5 (10.0)	15 (15.0)	
BMI (kg/m^2)				.62
N	50	50	100	
Mean (SD)	24.2 (2.7)	23.9 (3.4)	24.0 (3.1)	
Hypertension, n (%)				.33
0	41 (82.0)	37 (74.0)	78 (78.0)	
1	9 (18.0)	13 (26.0)	22 (22.0)	
Diabetes, n (%)				>.99
0	45 (90.0)	45 (90.0)	90 (90.0)	
1	5 (10.0)	5 (10.0)	10 (10.0)	
HBV positive, n (%)				.66
0	15 (30.0)	13 (26.0)	28 (28.0)	
1	35 (70.0)	37 (74.0)	72 (72.0)	
Tobacco, n (%)				.69
0	21 (42.0)	23 (46.0)	44 (44.0)	
1	29 (58.0)	27 (54.0)	56 (56.0)	
Alcohol, n (%)				.31
0	26 (52.0)	31 (62.0)	57 (57.0)	
1	24 (48.0)	19 (38.0)	43 (43.0)	
Education, n (%)				.33
Illiterate	4 (8.0）	3 (6.0）	7 (7.0)	
Preliminary	16 (32.0）	10 (20.0）	26 (26.0)	
Junior	21 (42.0）	28 (56.0）	49 (49.0)	
Senior	5 (10.0）	2 (4.0）	7 (7.0)	
College or above	4 (8.0）	7 (14.0）	11 (11.0)	
Income (US $/month)，n (%)				.73
<420	20 (20.0）	17 (17.0)	37 (37.0)	
421-840	20 (20.0)	20 (20.0)	40 (40.0)	
>840	10 (10.0)	13 (13.0)	23 (23.0)	
HCC[c], n (%)				.83
0	15 (30.0)	16 (32.0)	31 (31.0)	
1	35 (70.0)	34 (68.0)	69 (69.0)	
Open surgery, n (%)				.23
0	27 (54.0)	21 (42.0)	48 (48.0)	
1	23 (46.0)	29 (58.0)	52 (52.0)	

XSL•FO
RenderX

	Control (N=50)	Digital health (N=50)	Total (N=100)	*P* value
Surgery duration (min)				.14
N	50	49	99	
Mean (SD)	147.2 (63.3)	160.9 (53.0)	154.0 (58.5)	
IBL[d] (mL)				.20
N	50	49	99	
Mean (SD)	297.0 (252.2)	428.1 (471.6)	361.9 (380.9)	
LOS[e] (days)				.27
N	50	50	100	
Mean (SD)	7.6 (3.4)	8.5 (4.6)	8.1 (4.1)	
Maximum tumor diameter (cm)				.22
N	46	46	92	
Mean (SD)	4.6 (3.6)	5.0 (2.7)	4.8 (3.2)	

[a]0=No.

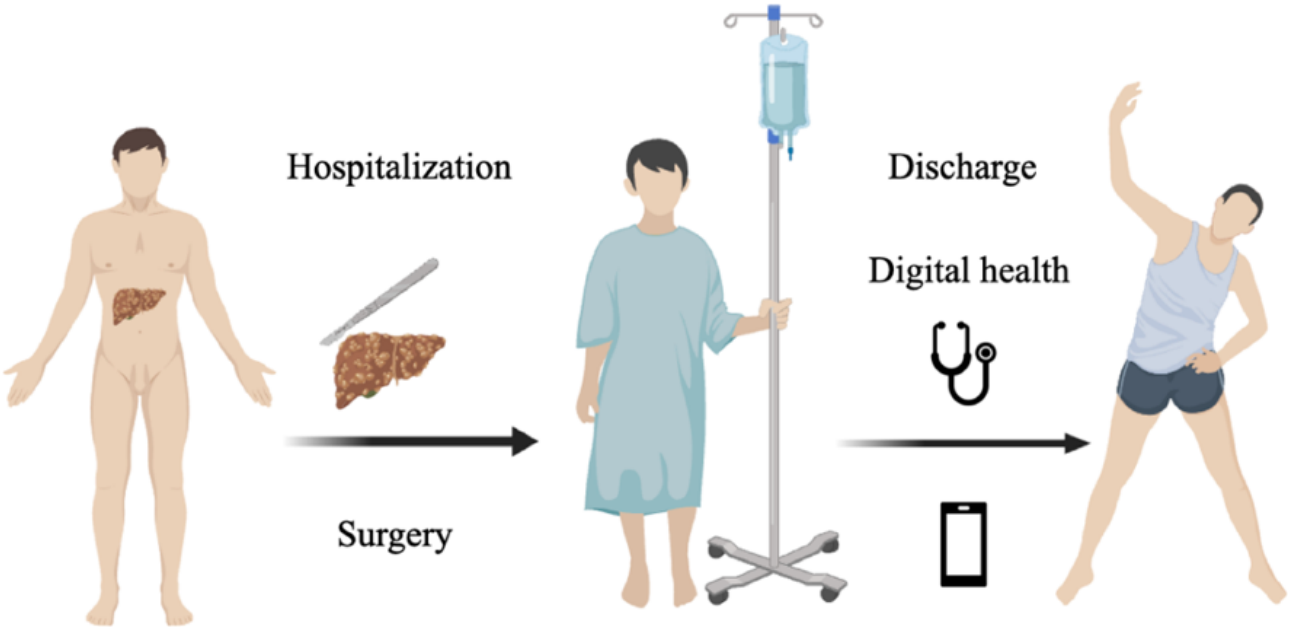
[b]1=Yes.

[c]HCC: hepatocellular carcinoma.

[d]IBL: intraoperative blood loss.

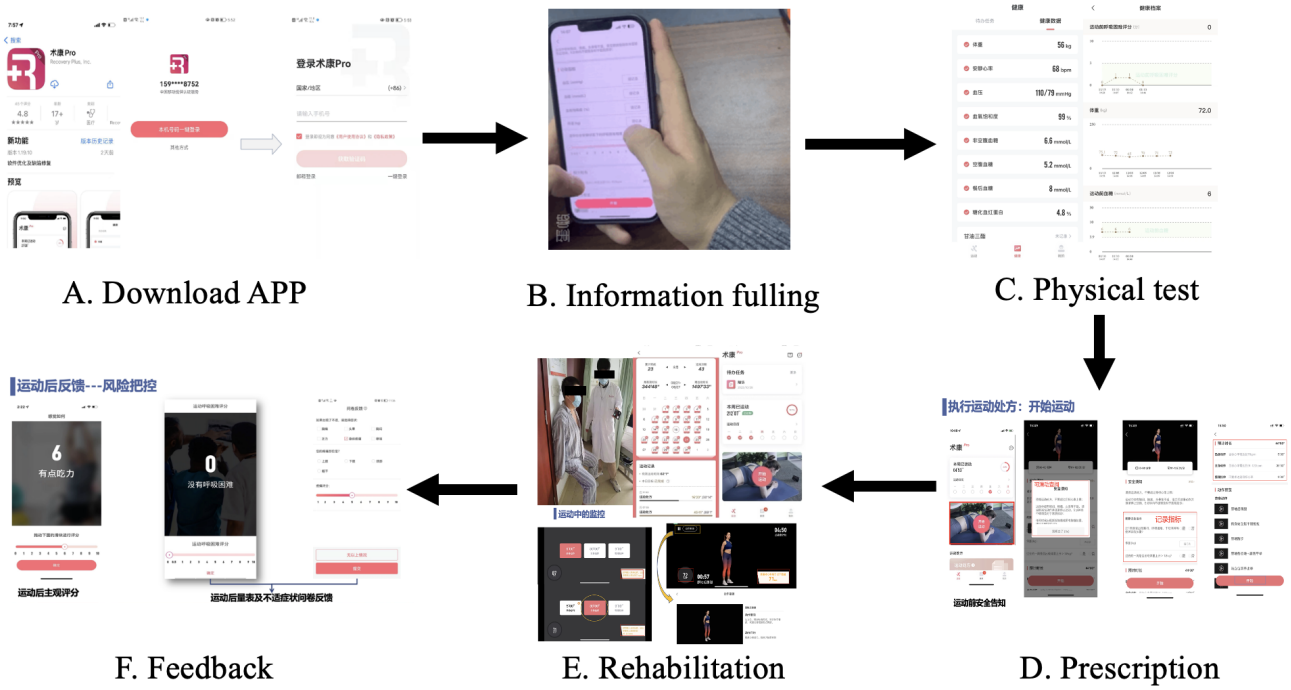
[e]LOS: length of stay.

Figure 1. Liver cancer care pathway.



XSL•FO
RenderX

Figure 2. Digital health intervention process.



Rehabilitation Education Group

From discharge from the hospital to 3 weeks until follow-up, patients received a paper-format rehabilitation manual in exercise rehabilitation, the content of rehabilitation training was consistent with that of the digital health group, and rehabilitation maneuvers were guided before discharge patiently until patients could master them. Patients were asked to write down their daily exercise content in the exercise diary. After discharge, patients were supervised via the web once a week. If there were any problems with the exercise, participants could contact the researchers by phone and receive instructions until the outpatient follow-up visit. Participants were followed up in the outpatient clinic at the end of the 3 weeks after discharge from the hospital.

Outcome Measures

Patient Characteristics

General demographic and clinical characteristics measurements were designed by the research team. Demographic characteristics included age, sex, educational level, and history of diabetes, hypertension, smoking, drinking, and HBV. Clinical characteristics included classification and stage of cancer, surgical procedure, length of hospital stay, and intraoperative blood loss. Demographic characteristics were recorded before surgery, and clinical characteristics were recorded through the electronic medical system one day after surgery.

Primary Outcome Measure

The 6-minute walk test (6MWT) was used to evaluate the cardiopulmonary fitness level of the patients with PLC. The test method was to select a 50-meter-long hard surface in the corridor outside the ward and place markers at both ends. Patients were asked to walk back and forth along the marked line as far as possible, ensuring that there was no interference in the surrounding area, talking, running, and jumping were prohibited during the walk, and there was no hesitation at the folding point, and the total distance traveled was calculated after timing the walk for 6 minutes [13]. The 6MWT was selected for its ease of administration, greater tolerability, and superior reflection of cardiopulmonary fitness in patients undergoing abdominal cancer surgery [14].

Secondary Outcome Measure

Upper Limb Muscle Strength

Grip strength is the sum of the forces produced when holding a target object tightly with the hand in each situation, and its force is produced by the joint contractile activity of the lateral forearm muscle group and the intrinsic hand muscle group [15]. In this study, the Jamar + grip dynamometer was used, which was measured by maintaining a seated position, adjusting the gripping position according to the size of the hand, and grasping the grip dynamometer with force. Each hand was repeated 3 times, and the tester was required to record all the test results, and the maximum value was taken as the final measurement [16]. Patients undergoing major abdominal surgery endure a significant reduction in muscle mass, which exacerbates postoperative outcomes [17-19].

Lower Limb Muscle Strength

The Five Times Sit-to-Stand test was used to assess the lower limb muscle strength, balance, and locomotor ability of individuals [20]. This was done by preparing a chair with a height of 48 centimeters, and subjects were instructed to sit with their arms crossed over their chests and their backs against the chair. The sitting and standing test maneuver was repeated 5 times at the fastest speed under the instructor's command, and the timing was stopped when the subject's body touched the chair after the 5th time, and the time was recorded [21]. Five Times Sit-to-Stand test is a valid and reliable objective screening tool for muscle strength and frailty of cancer survivors.

XSL•FO
RenderX

Fatigue

The fatigue was assessed using the Multidimensional Fatigue Inventory-20, which was developed by Dutch in 1995, contains 20 subjective entries, and was used to assess the severity of fatigue in patients. Higher scores on the scale indicate more severe fatigue. The application of the Chinese version in a cancer population showed a Cronbach α coefficient of 0.867 and good internal consistency [22].

Exercise Adherence

There is no standardized scale for exercise adherence. Recommended exercise time is at least 90 minutes per week. In this study, according to the constructed exercise rehabilitation program, sensor- and app-based log files were individually analyzed for each exercise session to quantify exercise adherence. The digital health group was evaluated according to the degree of the patients' achievement of exercise through the feedback of the exercise record time in the app and the exercise diary of the control group, and the patients' achievement of exercise amount was evaluated, and the patients' exercise adherence was regarded to be high if they reached 80% and above of the total recommended exercise time, 40% - 80% was regarded as medium exercise adherence; and below 40% was regarded as low exercise adherence.

Sample Size, Randomization, and Blinding

To determine the sample size, we calculated the primary outcome based on 6-minute walking distance in the pretest, we obtained a 6-minute walking distance of 480 (SD 100) m after 3 weeks of intervention for the digital health intervention group and 400 (SD 100) m for the control group. A total of 68 participants was required, calculated with a statistical power of 0.80 and a significance level of .05. Anticipating a 15% dropout rate, the adjusted sample size for our study was set at 80 participants. The web-based research randomizer was used, and a blocked design was selected to ensure balanced allocation across the 2 groups. For each hospital, participants were randomized to the intervention group or the control group to ensure an allocation ratio of 1:1.

All eligible patients provided informed consent according to the protocol and agreed to participate in the study. They were then centrally randomized into either the digital health group (intervention group) or the manual-based group (control group) at a 1:1 ratio. The randomization code was generated using a computer random number generator by researchers who were not involved in patient treatment. Stratified randomization was conducted by performing the randomization procedure separately for each of the 2 centers. Participants were not blinded to group assignment or type of study intervention, but study personnel responsible for data collection and analysis were blinded.

Data Collection

Data were collected between January 2022 and October 2023. After signing the paper consent form, all participants completed the demographic and baseline physical outcomes before randomization (T0). Participants in both groups also completed the assessment at 3 weeks (T1) when they followed up. If participants did not return to the hospital, researchers contacted them by phone to collect related information.

Statistical Methods

Data were analyzed using IBM SPSS (version 26.0; IBM Corp) and Stata BE (version 17; Stata Corp). Primary and secondary outcomes were analyzed as follows: using the Kolmogorov-Smirnov test to determine whether continuous data follows a normal distribution. Count data were presented as frequencies and percentages (%). Continuous data following normal distribution were expressed as mean and SD, while nonnormal distributions were expressed as median (quartile spacing). Because the outcome variables were measured twice before and after intervention, a difference in differences approach (linear regression) was used to control for baseline levels. The independent samples *t* test was used for comparison between groups of normally distributed continuous data. The Mann-Whitney *U* test was used for comparison of nonnormally distributed continuous data. Count data were expressed as the number of cases (rate) and tested using the chi-square test or Fisher exact test. Mean differences in the change from baseline to week 3 were also calculated. If subsequent multi-factor analysis is required, multiple linear regression or logistic regression models are selected according to continuous-type or subtype dependent variables. Multiple interpolations were used for the analysis of missing data. $P<.05$ was considered significant.

The study used mixed-effects linear regression models to examine the preliminary impact of the intervention while adjusting baseline characteristics. The model incorporated fixed effects for age, sex, education levels, hypertension status, drinking status, and time. To account for the hierarchical structure of the data (patients nested within 2 hospitals), the model included random intercept and slope at the hospital level. At the patient level, random slope and intercept for a time were included to allow variations between individual patients in terms of initial status and trajectories over time on the outcome variables. The model is estimated using maximum likelihood estimation [23,24].

Ethical Considerations

This study was approved by the institutional review board of Huashan Hospital, Fudan University (approval number 2021-794). The research procedures were conducted in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and the World Medical Association Declaration of Helsinki. All participants provided written informed consent prior to their participation in the study. To ensure privacy and confidentiality, all collected data were anonymized and securely stored. Access to the study application was protected by individual private usernames and passwords assigned to each participant. No financial compensation was provided to participants for their involvement in the study, and all study materials were provided free of charge. The study was registered at ClinicalTrials.gov (Identifier: ChiCTR2100052911).

XSL•FO
RenderX

Results

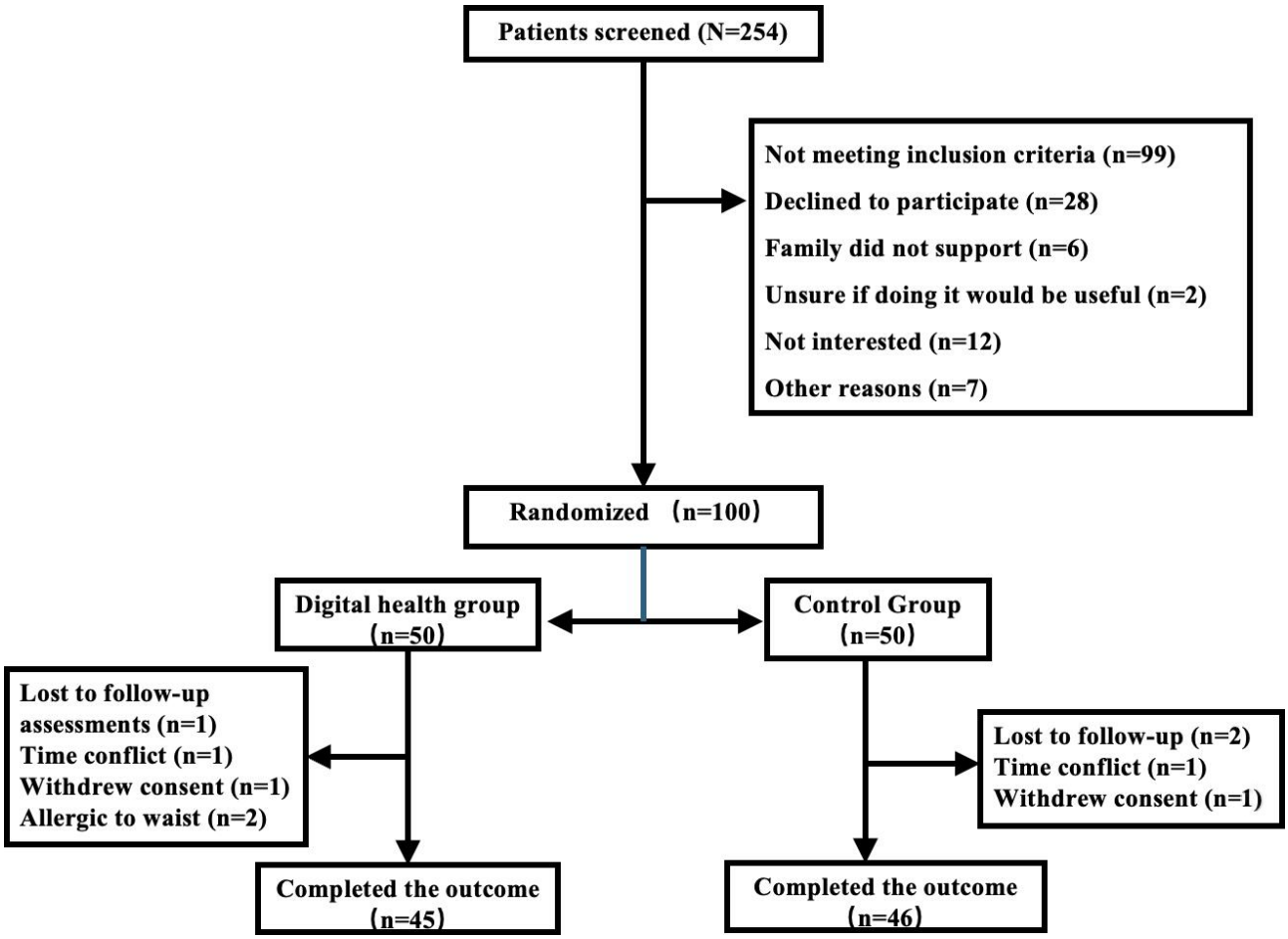
Participant Timeline

Of the 254 patients screened, 99 patients did not meet the inclusion criteria. Specifically, 72 patients were diagnosed with benign liver tumors, 17 patients were expected to undergo hepatectomy involving fewer than 2 liver segments, and 10 patients were unable to operate a mobile phone. Additionally, 28 patients declined to participate without specifying reasons, 6 patients were not enrolled due to lack of family support, 2 patients refused because they were uncertain about the trial's usefulness, 12 patients lacked interest, and 7 patients declined for other unspecified reasons.

Of the 100 patients who met the criteria, 50 patients were randomly assigned to the digital health group (n=45, 90% male patients; the mean age was 54.6, SD 10.6 years) and 50 to the control group (n=40, 80%) male patients; the mean age was 56.4, SD 9.7 years). At follow-up, 5 patients in the digital health group were lost: 1 patient lost contact, 1 patient cited time constraints, 1 patient withdrew voluntarily, and 2 patients withdrew due to an allergic reaction to the heart rate monitor. In the control group, 4 patients were lost to follow-up: 2 patients lost contacts, 1 patient died, and 1 patient withdrew voluntarily.

A total of 91 participants completed all assessments, with a dropout rate of 10% (5/50) in the digital health group and 8% (4/50) in the control group. No significant differences were detected between those who were lost to follow-up and those who completed the assessments, as none of the *P* values reached statistical significance. For more details, refer to Figure 3 for the participant timeline.

Figure 3. Participant timeline.



Baseline Data

Participants' ages ranged from 35 to 75 (average 55.5, SD 10.2) years. No significant differences were detected in the baseline characteristics between the digital health and control groups (*P*>.05). The majority (69/100, 69%) of the patients were diagnosed with hepatocellular carcinoma. No significant differences were detected in the surgery characteristics between the digital health and control groups (*P*>.05). Details are listed in Table 2.

Primary Outcome

The 6MWT demonstrated a significant improvement in the intervention group compared with the control group. At baseline, the mean 6MWT distance was 515.3 (SD 127.6) m in the intervention group and 520.0 (SD 73.7) m in the control group. After 3 weeks, the mean 6MWT distance in the intervention group decreased to 501.0 (SD 114.1) m, whereas the control group showed a more pronounced decline to 464.0 (SD 100.7) m. The mean difference in the change from baseline between the groups was 70.21 (95% CI 0.730-82.869) m (*P*=.05),

XSL•FO
RenderX

indicating a statistically significant improvement in the intervention group over the control group.

Secondary Outcomes

Grip Strength

The mean grip strength at baseline was comparable between groups, with the intervention group showing 34.8 (SD 9.3) kg and the control group 33.4 (SD 9.8) kg. Posttreatment, the intervention group had a slight improvement to 34.9 (SD 9.0) kg, while the control group decreased to 31.8 (SD 8.6) kg. The mean difference between the groups was 1.76 (95% CI −0.288 to 3.814) kg. However, this difference was not statistically significant (*P*=.09).

5-Repetition Sit-to-Stand Test

Baseline scores for the 5-repetition-sit-to-stand test were 9.2 (SD 3.0) seconds in the intervention group and 9.7 (SD 4.7) seconds in the control group. Posttreatment, the intervention group improved to 8.9 (SD 2.4) seconds, whereas the control group worsened slightly to 10.0 (SD 3.1) seconds. The mean difference between groups was 0.62 (95% CI −2.236 to 0.934) seconds, but the result was not statistically significant (*P*=.42).

Fatigue Score (Multidimensional Fatigue Inventory-20)

Fatigue scores were measured at baseline and posttreatment. The intervention group showed a mean baseline fatigue score of 44.8 (SD 9.8) points, identical to the control group (mean 44.8, SD 9.8). However, after treatment, the intervention group's fatigue score increased to 47.4 (SD 10.8) points, while the control group saw a more substantial increase to 55.9 (SD 11.6) points. The mean difference in the change from baseline was −2.77 (95% CI −7.919 to 1.271) points, which was not statistically significant (*P*=.16). Details are listed in Table 3.

Table . Comparison of physical function between 2 groups.

Change from baseline at week 3	Pretreatment				Posttreatment				Mean difference from baseline to week 3, (95% CI)	*P* values
	Patients number	Intervention, mean (SD)	Patients number	Control, mean (SD)	Patients number	Intervention, mean (SD)	Patients number	Control, mean (SD)		
6MWT[a] (meters)	47	515.3 (127.6)	49	520.0 (73.7)	47	501.0 (114.1)	49	464.0 (100.7)	70.21 (0.730 to 82.869)	<.05
Grip (Kg)	46	34.8 (9.3)	49	33.4 (9.8)	46	34.9 (9.0)	49	31.8 (8.6)	1.76 (−0.288 to 3.814)	.09
5R-STS[b] (s)	48	9.2 (3.0)	48	9.7 (4.7)	48	8.9 (2.4)	48	10.0 (3.1)	0.62 (−2.236 to 0.934)	.42
Fatigue score (points)	50	44.8 (9.8)	49	44.8 (9.8)	49	47.4 (10.8)	49	55.9 (11.6)	−2.77 (−7.919 to 1.271)	.16

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

[b]5R-STS 5-repetition-sit-to-stand test.

Intervention Adherence and Completion

The chi-square test was used to compare the exercise compliance between the 2 groups, and the results showed that compared with the control group, the exercise compliance of the experimental group was significantly higher than that of the control group at 3 weeks after discharge (χ^2_2=15.87, *P*<.001), and the difference was statistically significant, as shown in Table 4.

Table . Comparison of exercise adherence between 2 groups.

Adherence	Control group (n=50)	Digital health (n=50)	Chi-square (*df*)	*P* value
High adherence, n (%)	6 (12.0)	18 (36.0)	15.87 (2)	<.001
Medium adherence, n (%)	17 (34.0)	3 (6.0)	—[a]	—
Low adherence, n (%)	27 (54.0)	29 (58.0)	—	—

[a]Not available.

XSL•FO
RenderX

Discussion

Overview

In recent years, rapid advancements in the treatment of PLC have significantly improved overall survival rates for patients [25]. Rehabilitation, increasingly recognized as a vital component of postoperative care, is delivered through a multidisciplinary team. The benefits of exercise are well-documented by authoritative bodies [26]. However, exercise rehabilitation remains a valuable yet underused treatment strategy for patients with PLC. In this trial, the digital health rehabilitation program was rigorously validated by a multidisciplinary team and specifically designed for patients with PLC posthepatectomy in China. This pilot study aimed to compare the efficacy of a digital health intervention with a conventional rehabilitation manual on physical fitness and symptoms in patients with PLC following hepatectomy over a 3-week intervention period. The findings demonstrated a statistically significant improvement in cardiopulmonary fitness (as measured by the 6-minute walk distance) in the digital health group compared with the control group. However, no statistically significant effects were observed for secondary outcomes. Importantly, exercise adherence was significantly higher in the digital health group.

Principal Findings

Performance Measures

The results showed that the digital health program significantly improved the cardiopulmonary fitness (primary outcome) of patients after hepatobiliary surgery at 3 weeks. The comparison between groups' mean differences value in 6-minute walking distance was 70.21 (95% CI 0.730-82.869) m (P=.05). Although the minimal clinically important difference for 6MWD in patients with PLC after hepatectomy remains undetermined. In contrast to previous research indicating that home-based exercise or telerehabilitation interventions in patients with cancer have demonstrated only minor to moderate effects on 6-minute walking distance or oxygen uptake [27,28]. The improvement in this trial is in the range of the conservative estimate of minimal clinically important difference, indicating a statistically significant improvement in the digital health group [29]. Significant changes in cardiorespiratory performance resulted in an inverse and clinically relevant change in mortality risk and contributed to improved health [30]. Golbus et al [31] and Kim et al [32] also validated the statistically significant improvements in 6MWD using a digital health app.

Regarding secondary outcomes, there were no statistically significant differences between the 2 groups in terms of upper and lower limb muscle strength and fatigue. Although the digital health group exhibited a slight increase in grip strength and a decrease in the time required to complete the five sit-to-stand tests compared with the control group, these changes did not reach statistical significance. However, considering that interventions such as quadriceps stretching, seated deltoid stretches, elastic band exercises, and wall squats are designed to enhance overall muscle mass, improvements in muscle strength and fatigue over time may hold clinical significance.

Adherence

In this trial, the digital health program significantly increased exercise adherence compared with the control group (P<.001). This is consistent with that reported in Snoek's research [33]. Advanced supervision and technology, strong professional support and relatedness communication and feedback are the key advantages of digital health to enhance adherence. Digital health is an important medium for the delivery of exercise rehabilitation intervention, it not only integrates a variety of sensory systems to simulate movement situations, restores the realism and effectiveness of rehabilitation training, but also builds a social support network to support patients' contact with physicians when concerns about rehabilitation arise after discharged [34,35]. However, in our study, exercise adherence was lower than expected, likely due to several contributing factors. One potential explanation is latency in the perceived benefits of exercise and a lack of self-efficacy. The participants might have higher expectations of recovery, but improvements in fitness levels gained through exercise are not immediate. When they failed to perceive recovery or encountered adverse symptoms, they chose to withdraw, affecting their persistence, enthusiasm, and adherence [36,37]. Self-efficacy is also an important determinant in exercise adherence [38]. And the socioeconomic and educational disadvantages among participants are negative to self-efficacy. Notably, 82 (82%) of the patients had an education level of junior high school or below, and 77 (77%) reported monthly incomes of less than US $840. These factors may have limited their understanding and engagement, consistent with existing literature that associates lower education and income with poor adherence [39-41]. In this trial, the stability of the connection between software and hardware products also needed to be improved and 2 patients had allergies to the hardware products, all of which need to be improved in the later studies.

Limitations

Our findings are limited by a short-term follow-up and the adherence was lower than expected. Further studies of the digital health program could use recruitment strategies aimed at extending the intervention period and follow-up, ensuring the effectiveness of rehabilitation programs, and maximizing the exercise adherence for patients' physical fitness. Another limitation of this study is the absence of muscle mass measurement. Future research should incorporate muscle mass measurements to provide a more comprehensive assessment of the impact of exercise rehabilitation programs on physical fitness and overall health.

Conclusion

The findings of this randomized controlled trial conducted in China suggested that novel digital intervention in exercise rehabilitation could improve exercise capacity and physical fitness among patients with PLC after hepatectomy. The intervention has the potential to increase access to rehabilitation treatment and reduce adverse symptoms in patients in the postsurgical recovery period.

XSL•FO
RenderX

Acknowledgments

The study received funding from the National Key Research and Development Program of China (grant/award number: 2018YFC2002300 and 2018YFC2002301), the National Natural Innovation Research Group Project (82021002), the National Natural Science Foundation of China (82472602), the National Natural Science Foundation of China (82272606), and the National Natural Integration Project (91948302) the National Key Research.

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH Checklist (V 1.6.1).
[PDF File, 477 KB - mhealth_v13i1e59228_app1.pdf]

References

1. Benson AB, Venook AP, Al-Hawary MM, et al. Colon cancer, version 2.2021, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2021 Mar 2;19(3):329-359. [doi: 10.6004/jnccn.2021.0012] [Medline: 33724754]
2. Steel JL, Kim KH, Dew MA, et al. Cancer-related symptom clusters, eosinophils, and survival in hepatobiliary cancer: an exploratory study. J Pain Symptom Manage 2010 May;39(5):859-871. [doi: 10.1016/j.jpainsymman.2009.09.019] [Medline: 20471546]
3. Burke EC, Jarnagin WR, Hochwald SN, Pisters PW, Fong Y, Blumgart LH. Hilar Cholangiocarcinoma: patterns of spread, the importance of hepatic resection for curative operation, and a presurgical clinical staging system. Ann Surg 1998 Sep;228(3):385-394. [doi: 10.1097/00000658-199809000-00011] [Medline: 9742921]
4. Rock CL, Doyle C, Demark-Wahnefried W, et al. Nutrition and physical activity guidelines for cancer survivors. CA Cancer J Clin 2012;62(4):243-274. [doi: 10.3322/caac.21142] [Medline: 22539238]
5. Okunrintemi V, Gani F, Pawlik TM. National trends in postoperative outcomes and cost comparing minimally invasive versus open liver and pancreatic surgery. J Gastrointest Surg 2016 Nov;20(11):1836-1843. [doi: 10.1007/s11605-016-3267-z] [Medline: 27613735]
6. Bailey DL, Holden MA, Foster NE, Quicke JG, Haywood KL, Bishop A. Defining adherence to therapeutic exercise for musculoskeletal pain: a systematic review. Br J Sports Med 2020 Mar;54(6):326-331. [doi: 10.1136/bjsports-2017-098742] [Medline: 29875278]
7. Blair SN. Physical inactivity: the biggest public health problem of the 21st century. Br J Sports Med 2009 Jan;43(1):1-2. [Medline: 19136507]
8. Jones C, Kelliher L, Dickinson M, et al. Randomized clinical trial on enhanced recovery versus standard care following open liver resection. Br J Surg 2013 Jul;100(8):1015-1024. [doi: 10.1002/bjs.9165] [Medline: 23696477]
9. Patel S, Goldsack JC, Cordovano G, et al. Advancing digital health innovation in oncology: priorities for high-value digital transformation in cancer care. J Med Internet Res 2023 Jan 4;25:e43404. [doi: 10.2196/43404] [Medline: 36598811]
10. Li J, Wei D, Liu S, et al. Efficiency of an mHealth app and chest-wearable remote exercise monitoring intervention in patients with type 2 diabetes: a prospective, multicenter randomized controlled trial. JMIR Mhealth Uhealth 2021 Feb 9;9(2):e23338. [doi: 10.2196/23338] [Medline: 33560244]
11. O'Neill L, Guinan E, Doyle S, et al. Rehabilitation strategies following oesophagogastric and Hepatopancreaticobiliary cancer (ReStOre II): a protocol for a randomized controlled trial. BMC Cancer 2020 May 13;20(1):415. [doi: 10.1186/s12885-020-06889-z] [Medline: 32404096]
12. Riebe D, Pescatello LA. In: Thompson PD, editor. ACSM's Guidelinesfor Exercise Testing and Prescription, 9th edition: Baltimore, MD: Lippincott, Williams &Wilkins; 2013.
13. Agarwala P, Salzman SH. Six-minute walk test: clinical role, technique, coding, and reimbursement. Chest 2020 Mar;157(3):603-611. [doi: 10.1016/j.chest.2019.10.014] [Medline: 31689414]
14. Hijazi Y, Gondal U, Aziz O. A systematic review of prehabilitation programs in abdominal cancer surgery. Int J Surg 2017 Mar;39:156-162. [doi: 10.1016/j.ijsu.2017.01.111] [Medline: 28161527]
15. Mitsionis G, Pakos EE, Stafilas KS, Paschos N, Papakostas T, Beris AE. Normative data on hand grip strength in a Greek adult population. Int Orthop 2009 Jun;33(3):713-717. [doi: 10.1007/s00264-008-0551-x] [Medline: 18414855]
16. Budziareck MB, Pureza Duarte RR, Barbosa-Silva MCG. Reference values and determinants for handgrip strength in healthy subjects. Clin Nutr 2008 Jun;27(3):357-362. [doi: 10.1016/j.clnu.2008.03.008] [Medline: 18455840]
17. Zenith L, Meena N, Ramadi A, et al. Eight weeks of exercise training increases aerobic capacity and muscle mass and reduces fatigue in patients with cirrhosis. Clin Gastroenterol Hepatol 2014 Nov;12(11):1920-1926. [doi: 10.1016/j.cgh.2014.04.016] [Medline: 24768811]

XSL•FO
RenderX

18. Otsuji H, Yokoyama Y, Ebata T, et al. Surgery-related muscle loss and its association with postoperative complications after major hepatectomy with extrahepatic bile duct resection. World J Surg 2017 Feb;41(2):498-507. [doi: 10.1007/s00268-016-3732-6] [Medline: 27718001]
19. Yang J, Chen K, Zheng C, et al. Impact of sarcopenia on outcomes of patients undergoing liver resection for hepatocellular carcinoma. J Cachexia Sarcopenia Muscle 2022 Oct;13(5):2383-2392. [doi: 10.1002/jcsm.13040] [Medline: 35854105]
20. van Cappellen-van Maldegem SJM, Hoedjes M, Seidell JC, et al. Self-performed five times sit-to-stand test at home as (pre-)screening tool for frailty in cancer survivors: reliability and agreement assessment. J Clin Nurs 2023 Apr;32(7-8):1370-1380. [doi: 10.1111/jocn.16299] [Medline: 35332600]
21. Klukowska AM, Staartjes VE, Vandertop WP, Schröder ML. Five-repetition sit-to-stand test performance in healthy individuals: reference values and predictors from 2 prospective cohorts. Neurospine 2021 Dec;18(4):760-769. [doi: 10.14245/ns.2142750.375] [Medline: 35000330]
22. Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995 Apr;39(3):315-325. [doi: 10.1016/0022-3999(94)00125-o] [Medline: 7636775]
23. Grilli L, Rampichini C. Review of multilevel and longitudinal modeling using stata, fourth edition, by Sophia Rabe-Hesketh and Anders Skrondal. Stata J 2023 Sep;23(3):901-904. [doi: 10.1177/1536867X231196518]
24. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence: Oxford University Press; 2003. URL: https://doi.org/10.1093/acprof:oso/9780195152968.001.0001
25. Firkins J, Hansen L, Driessnack M, Dieckmann N. Quality of life in "chronic" cancer survivors: a meta-analysis. J Cancer Surviv 2020 Aug;14(4):504-517. [doi: 10.1007/s11764-020-00869-9] [Medline: 32162194]
26. Johnson JM, Ballin SD. Surgeon General's report on physical activity and health is hailed as a historic step toward a healthier nation. Circulation 1996 Nov 1;94(9):2045. [doi: 10.1161/01.cir.94.9.2045] [Medline: 8901647]
27. Filakova K, Janikova A, Felsoci M, et al. Home-based cardio-oncology rehabilitation using a telerehabilitation platform in hematological cancer survivors: a feasibility study. BMC Sports Sci Med Rehabil 2023 Mar 23;15(1):38. [doi: 10.1186/s13102-023-00650-2] [Medline: 36959613]
28. Lenneman CG, Sawyer DB. Cardio-oncology: an update on cardiotoxicity of cancer-related treatment. Circ Res 2016 Mar 18;118(6):1008-1020. [doi: 10.1161/CIRCRESAHA.115.303633] [Medline: 26987914]
29. Wise RA, Brown CD. Minimal clinically important differences in the six-minute walk test and the incremental shuttle walking test. COPD 2005 Mar;2(1):125-129. [doi: 10.1081/copd-200050527] [Medline: 17136972]
30. Falz R, Bischoff C, Thieme R, et al. Effect of home-based online training and activity feedback on oxygen uptake in patients after surgical cancer therapy: a randomized controlled trial. BMC Med 2023 Aug 8;21(1):293. [doi: 10.1186/s12916-023-03010-6] [Medline: 37553660]
31. Golbus JR, Gupta K, Stevens R, et al. A randomized trial of a mobile health intervention to augment cardiac rehabilitation. NPJ Digit Med 2023 Sep 14;6(1):173. [doi: 10.1038/s41746-023-00921-9]
32. Kim Y, Seo J, An SY, Sinn DH, Hwang JH. Efficacy and safety of an mHealth app and wearable device in physical performance for patients with hepatocellular carcinoma: development and usability study. JMIR Mhealth Uhealth 2020 Mar 11;8(3):e14435. [doi: 10.2196/14435] [Medline: 32159517]
33. Snoek JA, Meindersma EP, Prins LF, et al. The sustained effects of extending cardiac rehabilitation with a six-month telemonitoring and telecoaching programme on fitness, quality of life, cardiovascular risk factors and care utilisation in CAD patients: the TeleCaRe study. J Telemed Telecare 2021 Sep;27(8):473-483. [doi: 10.1177/1357633X19885793] [Medline: 31760855]
34. Brennan L, Kessie T, Caulfield B. Patient experiences of rehabilitation and the potential for an mHealth system with biofeedback after breast cancer surgery: qualitative study. JMIR Mhealth Uhealth 2020 Jul 29;8(7):e19721. [doi: 10.2196/19721] [Medline: 32687476]
35. Su Z, Zhang L, Lian X, Guan M. Virtual reality-based exercise rehabilitation in cancer-related dysfunctions: scoping review. J Med Internet Res 2024 Feb 26;26:e49312. [doi: 10.2196/49312] [Medline: 38407951]
36. Jones F, Harris P, Waller H, Coggins A. Adherence to an exercise prescription scheme: the role of expectations, self-efficacy, stage of change and psychological well-being. Br J Health Psychol 2005 Sep;10(Pt 3):359-378. [doi: 10.1348/135910704X24798] [Medline: 16238853]
37. Morris TP, Burzynska A, Voss M, et al. Brain structure and function predict adherence to an exercise intervention in older adults. Med Sci Sports Exerc 2022 Sep 1;54(9):1483-1492. [doi: 10.1249/MSS.0000000000002949] [Medline: 35482769]
38. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Feb 19;18(4):2023. [doi: 10.3390/ijerph18042023] [Medline: 33669679]
39. Lin B, Zhang Z, Mei Y, Liu L, Ping Z. The influential factors of adherence to physical activity and exercise among community-dwelling stroke survivors: a path analysis. J Clin Nurs 2022 Sep;31(17-18):2632-2643. [doi: 10.1111/jocn.16091] [Medline: 34664325]

XSL•FO
RenderX

40. Katzmarzyk PT, Friedenreich C, Shiroma EJ, Lee IM. Physical inactivity and non-communicable disease burden in low-income, middle-income and high-income countries. Br J Sports Med 2022 Jan;56(2):101-106. [doi: 10.1136/bjsports-2020-103640] [Medline: 33782046]
41. Almutary H, Tayyib N. Factors associated with exercise self-efficacy among people with chronic diseases. Appl Nurs Res 2020 Aug;54:151275. [doi: 10.1016/j.apnr.2020.151275] [Medline: 32650891]

Abbreviations

6MWT: 6-minute walk test
PLC: primary liver cancer

Edited by L Buis; submitted 05.04.24; peer-reviewed by F Wu, KL Du; revised version received 27.10.24; accepted 26.02.25; published 07.04.25.

Please cite as:
Yu K, Yin B, Zhu Y, Meng H, Zhu W, Lu L, Wang J, Chen S, Ni J, Lin Y, Jia J
Efficacy of a Digital Postoperative Rehabilitation Intervention in Patients With Primary Liver Cancer: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e59228
URL: https://mhealth.jmir.org/2025/1/e59228
doi:10.2196/59228

©Kaitao Yu, Baobing Yin, Ying Zhu, Hongdao Meng, Wenwei Zhu, Lu Lu, Junqiao Wang, Shugeng Chen, Jun Ni, Yifang Lin, Jie Jia. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 7.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Clinical, Psychological, Physiological, and Technical Parameters and Their Relationship With Digital Tool Use During Cardiac Rehabilitation: Comparison and Correlation Study

Fabian Wiesmüller[1,2,3], MSc; David Haag[1,2,4], MSc; Mahdi Sareban[2,5], PhD; Karl Mayr[6], PhD; Norbert Mürzl[7], Mag; Michael Porodko[7], PhD; Christoph Puelacher[8], MD; Lisa-Marie Moser[8], Mag; Marco Philippi[9]; Heimo Traninger[10], Mag; Stefan Höfer[11], PhD; Josef Niebauer[2,5], MD, PhD, MBA; Günter Schreier[1], PhD; Dieter Hayn[1,2], PhD, MBA

[1]Center for Health & Bioresources, AIT Austrian Institute of Technology GmbH, Reininghausstraße 13/1, Graz, Austria

[10]ZARG Zentrum für ambulante Rehabilitation GmbH, Graz, Austria

[11]Department of Psychiatry II, Medizinische Universität Innsbruck, Innsbruck, Austria

[2]Ludwig Boltzmann Institute for Digital Health and Prevention, Ludwig Boltzmann Gesellschaft, Salzburg, Austria

[3]Institute of Neural Engineering, Faculty of Computer Science and Biomedical Engineering, Graz University of Technology, Graz, Austria

[4]Department of Psychology, Paris-Lodron-University of Salzburg, Salzburg, Austria

[5]University Institute of Sports Medicine, Prevention and Rehabilitation, Paracelsus Medical University, Salzburg, Austria

[6]CARDIOMED Kardiologisches Rehabilitationszentrum GmbH, Linz, Austria

[7]Institut für Präventiv- und Rehabilitationsmedizin, Cardio Vital Wels, Wels, Austria

[8]Reha Innsbruck, REHAmed-tirol GmbH, Innsbruck, Austria

[9]MedReha GmbH, Feldkirch, Austria

Corresponding Author:
Fabian Wiesmüller, MSc
Center for Health & Bioresources, AIT Austrian Institute of Technology GmbH, Reininghausstraße 13/1, Graz, Austria

Abstract

Background: Home and telehealth-based interventions are increasingly used in cardiac rehabilitation, a multidisciplinary model of health care. Digital tools such as wearables or digital training diaries are expected to support patients to adhere to recommended lifestyle changes, including physical exercise programs. As previously published, the EPICURE study (effect of digital tools in outpatient cardiac rehabilitation including home training) analyzed the effects of digital tools, that is, a digital training diary, adherence monitoring, and wearables, on exercise capacity during outpatient cardiac rehabilitation phase III (OUT-III) which includes an approximately 12-week home-training phase. The study encompassed 149 Austrian patients, of which 50 used digital tools.

Objective: The present paper takes a deeper look into the EPICURE data to better understand the relation between the use of digital tools and various psychological, clinical, and physiological parameters, and the relation between these parameters and the improvement of exercise capacity during cardiac rehabilitation.

Methods: For this work, we analyzed questionnaires concerning the patients' cardiac rehabilitation. On all these parameters we performed 2 analyzes: (1) Comparison of the 2 groups with and without digital tools and (2) correlation with the change in the maximum workload as achieved during the exercise stress test. If data pre- and post OUT-III were available, the change in the respective parameter during OUT-III was determined and group analysis and correlation were applied on data pre OUT-III, data post OUT-III, and the change during OUT-III.

Results: We found significant improvements in quality of life in both groups, with no discernible differences between patients with or without digital tools (P=.53). Patients with digital tools perceived significantly higher competence during cardiac rehabilitation (P=.05), and they anticipated higher cardiac risks if nonadherent to physical activity (P=.03). Although, the overall subjectively reported adherence was not significantly different in the 2 groups (P=.50), specific items differed. Patients with digital tools were significantly more likely to do their exercises even when they were tired (P=.01) and less likely to forget their exercises (P=.01). Concerning reasons for (non-) adherence, patients with digital tools reported significantly more often to do their exercises because they enjoyed them (P=.01), whereas they were significantly less likely to stop exercising when muscular pain was worse (P=.01) and to continue doing their exercises when muscular pain improved (P=.02). Finally, patients who reported a high level of concrete planning achieved significantly higher improvements in exercise capacity (r=0.14, P=.04).

Conclusions: This comprehensive analysis provides valuable insights into the multifaceted impact of digital tools on outpatient cardiac rehabilitation including home training, shedding light on the importance of digital tools for increased competence and a

XSL•FO
RenderX

higher risk perception during cardiac rehabilitation. In addition, the impact of digital tools on adherence and their influence on patient outcomes were assessed in the evolving landscape of digital health interventions.

Trial Registration: ClincalTrials.gov NCT04458727; https://clinicaltrials.gov/study/NCT04458727

(JMIR Mhealth Uhealth 2025;13:e57413) doi:10.2196/57413

KEYWORDS

mHealth; telehealth; cardiac rehabilitation; wearable; adherence; health-related quality of life; intrinsic motivation; self-efficacy; health action process approach; cardiac; rehabilitation; quality of life; efficacy; psychological; physiological; digital tools; home training; monitoring; questionnaire; cardiac risk

Introduction

Rationale

Cardiovascular diseases (CVD) are the leading cause of death worldwide [1]. The burden of CVD, considering the number of disability-adjusted life years and deaths, and especially the CVD burden attributable to modifiable risk factors increase globally. Therefore, countries are recommended to implement cost-effective public health programs and interventions which target modifiable risks, promote healthy aging across the lifespan, and reduce disability and premature death due to CVD [2].

Background

Cardiac rehabilitation is a multidisciplinary model of health care that consists of 4 phases. Phase I starts during in-hospital treatment and focusses on early mobilization. Phase II can either be performed as in-clinic or outpatient cardiac rehabilitation, depending on the availability of outpatient cardiac rehabilitation and the patient's individual needs and preferences. Outpatient cardiac rehabilitation phase III (OUT-III) can follow both in-clinic and outpatient cardiac rehabilitation and consists of weekly visits at outpatient cardiac rehabilitation facilities. While during phase II patients are being introduced to the management of cardiovascular risk factors and subsequent lifestyle changes, phases III targets at further improving or at least maintaining results of lifestyle changes achieved during phase II [3,4]. Home- and telehealth-based interventions are increasingly being used in cardiac rehabilitation, and depending on the chosen cohort, outcomes for home-based rehabilitation may well compare with center-based cardiac rehabilitation programmes in terms of hospitalizations, quality of life (QoL), and cost [5].

The success of home- and telehealth-based cardiac rehabilitation is associated with various clinical, psychological, and physiological parameters that influence and are influenced by the patients' adherence to the cardiac rehabilitation programme. Various studies have analyzed the effect of wearables and home-base cardiac rehabilitation on the QoL. In a recent review including 57 articles, Jones et al [6] conclude that home-based cardiac rehabilitation leads to an improved QoL and exercise capacity.

So far, there is little knowledge concerning the relationship of patients' intrinsic motivation and digital tools during home-based cardiac rehabilitation. In a small-scale study with 23 patients, Lu et al [7] identified patients' intrinsic motivation as a key facilitator for successful rehabilitation with remote activity sensing. The rehabilitation program, however, did not focus on cardiac patients and none of the patients were reported to have a cardiac disease. Nonetheless, considering the broader context of health behavior change, which ties in closely with multiple of the above-mentioned cardiac rehabilitation components, there is a considerable body of research supporting the importance of (intrinsic) motivation [8]. Many current psychological models of health behavior, such as the Health Action Process Approach (HAPA), place motivation and resulting intentions (motivational phase) as a prerequisites of health behavior engagement, distinguishing it from a second, volitional phase, which translates intentions into actual behavior [9]. In the motivational phase, the HAPA proposes risk perception and outcome expectancy as determinants of intentions. Consecutively, action planning, self-efficacy, and action control are proposed to affect the enactment of these health behavior intentions. Therefore, all of these are promising targets for health interventions [10-12]. Various studies analyzed self-efficacy and behavioral driving models during home-based rehabilitation [13-16], indicating that there is a huge potential for addressing psychological drivers in digital technologies used in cardiac rehabilitation. When implementing digital tools to support cardiac rehabilitation, user-friendly interfaces are crucial, since poor usability is a main barrier of digital technologies in cardiac rehabilitation, as recently highlighted by McGowan et al [17].

The multicentered EPICURE study analyzed the effect of digital tools in outpatient cardiac rehabilitation including home training. Both patient groups concluded the OUT-III according to the latest national and international recommendations for center- and home-based exercise training [18,19]. Patients aged ≥18 years who concluded OUT-III at 1 of the 5 participating institutions, who had any documented cardiovascular disease (not further specified), and who were able to provide their informed consent were eligible for the EPICURE study if they gave their written informed consent [18]. The study duration was 6 to 12 months, depending on the duration of OUT-III which is flexibly adjusted according to the underlying disease [19]. EPICURE's primary outcomes did not support the hypothesis that the additional use of digital tools during home training lead to further improvement in the exercise capacity (P_{max}) during and after OUT-III. P_{max} was assessed during exercise stress test on an ergometer under supervision of health care professionals in the cardiac rehabilitation facilities. Patients who self-reportedly used either 1 or multiple of the following technologies during OUT-III were categorized as a patient using digital tools: phone-based assessments by the attending cardiac

XSL•FO
RenderX

rehabilitation facility and digital training diaries (with and without adherence monitoring done by the cardiac rehabilitation facility and with and without wearables).

This information was reported retrospectively by the patient and no protocol had been put in place to assign patient as users of digital tools (eg, minimum number of phone-based assessments) [18]. So far, the relationship between clinical, psychological, and physiological parameters as recorded in the EPICURE study had not been analyzed in detail.

Objectives

The present paper takes a deeper look into the EPICURE data to better understand the relationship between the use of digital tools and aforementioned psychological, clinical, and physiological parameters, and the relationship between digital tools, psychological, clinical, and physiological parameters and the improvement of exercise capacity during cardiac rehabilitation. Results are discussed in relationship to the primary outcome of the EPICURE study as published in [18] and concerning their accordance with the state-of-the-art.

Methods

Dataset

Data were taken from the multicentric EPICURE study [18]. The dataset contained 149 patients, who were asked to answer a questionnaire concerning their cardiac rehabilitation, and to grant access to data acquired during cardiac rehabilitation. No blinding or randomization was applied since patients were recruited retrospectively. Patients were assigned to 2 groups, depending on whether they reported to have used any of the following digital tools during OUT-III: phone-based assessments, digital training diaries with and without adherence monitoring and with and without wearables. At baseline (before OUT-III), patients using digital tools were significantly younger, fitter in terms of the maximum power during ergometry (P_{max}), had a lower BMI and body weight, and reported a higher QoL in all four aspects of the MacNew questionnaire prior to OUT-III [20]. More details concerning the EPICURE study population at baseline are shown in Table 1. The data in Table 1 are presented as mean (SD) and P-values were calculated with a *t* test. MacNew quality-of-life scores were assessed based on a German version of the MacNew quality-of-life questionnaire with subcategories regarding physical limitations in daily living (physical), emotional state of the patient in regard to the cardiovascular disease (emotional), effects of the cardiovascular disease on the social life and social activities (social) and an overall score over all subcategories (global) [21]. All questionnaires were recorded retrospectively post OUT-III, except for MacNew, which was recorded before and after the study duration. The structure of the analyzed data is described in the following.

XSL•FO
RenderX

Table . Study population at baseline (before outpatient cardiac rehabilitation phase III) as described in a previous study [18].

Characteristics	Without digital tools	With digital tools	*P* value
Sex, n			
Male	107	50	—[a]
Female	30	11	—
Age (years), mean (SD)	62 (9)	55 (13)	<.001
Maximum power during ergometry (P_{max}) (W), mean (SD)	142 (41)	186 (53)	<.001
BMI (kg/m²), mean (SD)	27.9 (4.7)	26.4 (4.4)	.04
Body weight (kg), mean (SD)	86 (16)	82 (13)	.13
Blood pressure (mmHg), mean (SD)			
Systolic	118 (11)	120 (17)	.77
Diastolic	77 (8)	75 (8)	.45
Blood levels (mg/dL), mean (SD)			
Glucose	105 (19)	99 (30)	.23
Cholesterol			.23
LDL	85 (32)	77 (35)	—
HDL	48 (12)	52 (12)	.04
Triglycerides	115 (53)	99 (59)	.16
Smoking, n			.06
Nonsmoker	28	18	—
Ex-smoker	46	23	—
Smoker	14	1	—
MacNew score, mean (SD)			
Physical	5.64 (0.93)	6.15 (0.81)	.01
Emotional	5.57 (0.91)	6.01 (0.8)	.02
Social	5.84 (0.91)	6.23 (0.87)	.04
Global	5.68 (0.88)	6.11 (0.76)	.02

[a] —: not available.

Quality of Life

We used the German version of the MacNew heart disease health-related QoL questionnaire [20] to assess QoL before and after OUT-III. The MacNew questionnaire was designed to evaluate how daily activities and physical, emotional, and social functioning are affected by coronary heart disease and its treatment. The questionnaire consists of 27 items concerning 3 domains (physical limitations, emotional function, and social function), with a maximum score of 7 per domain. The MacNew questionnaire is routinely assessed before and after cardiac rehabilitation by the centers participating in the EPICURE study. Therefore, unlike most other parameters, even differences of QoL pre versus post OUT-III and their relationship to the use of digital tools during OUT-III were analyzed.

Intrinsic Motivation

We used the German version of the Intrinsic Motivation Inventory [22] to analyze relations between intrinsic motivation, digital tools, and outcomes of cardiac rehabilitation. Intrinsic Motivation Inventory is a standardized multidimensional measurement device intended to assess participants' subjective experience related to a target activity in laboratory experiments. The target activity was defined as physical activity during the home training phase. Intrinsic Motivation Inventory consists of 22 items concerning the categories interest and enjoyment, perceived competence, perceived choice, and felt pressure and tension.

Health Action Process Approach

Self-efficacy was determined based on the German version of the Self-Efficacy Scale by Jerusalem and Schwarzer [23]. In addition, we included the following parts of the HAPA described by Schwarzer [24] in the questionnaire, intention, action planning (planning), coping planning (control), outcome expectations, and risk perception. In total, 7 questions concerning intentions before OUT-III, 10 questions concerning concrete planning of physical activity during OUT-III, and 6 questions concerning control during OUT-III were provided to the participants. In addition, 1 question concerning expected

XSL•FO
RenderX

health benefits of physical activity (outcome expectations) and 1 question concerning expected cardiac risk of nonadherence to physical activity (risk perception) were retrieved. All these questions were implemented in a 4-fold Likert scale.

Self-Reported Adherence to the Exercise Program

Patients were asked to answer 17 questions concerning their adherence to cardiac rehabilitation, which were based on the Exercise Adherence Rating Scale (EARS) [25]. Eleven out of the 17 questions concerned reasons for nonadherence, while the remaining 6 questions focused on adherence. EARS was translated to German and slightly adapted according to the study needs (refer to the self-reported adherence to the exercise program section for the detailed questions). As compared with the original EARS which applies a 5-fold Likert scale, binary responses (yes or no) were implemented.

Statistical Analysis

The maximum mechanical power P_{max} (W) as achieved by patients during the exercise stress test at their regular assessments in the study centers pre- and post OUT-III was analyzed. The stress test was conducted through ergometry following the Austrian guidelines for exercise testing [26]. The difference ΔP_{max} (W) between P_{max} at the end of OUT-III minus P_{max} at the beginning of OUT-III was determined. If no data from the assessment before OUT-III was available, data from the assessment post phase II were taken instead since, in Austria OUT-III is expected to immediately follow phase II and, therefore, the patients P_{max} and questionnaire results are not expected to change between the end of phase II and start of OUT-III [27-29].

For each of the parameters described in Dataset section, we performed 2 analyses, (1) comparison of the 2 groups with and without digital tools and (2) correlation with $\Delta P_{max.}$ Where data for both, pre- and post OUT-III, were available, the change in the respective parameter during OUT-III was determined and group analysis and correlation were applied on data pre OUT-III, data post OUT-III, and the change during OUT-III.

Normal distribution of the data was tested by the Shapiro-Wilk test. A Student *t* test was applied to test for global differences between pre- to post cardiac rehabilitation (dependent *t* test) and for differences between the groups (independent *t* test). A value of $\alpha<.05$ was considered significant. Where appropriate, we included confounding variables in the ANCOVA (analysis of covariance) models as covariates, to control for theses confounders. Confounding variables were all variables which significantly differed between patients without versus patients with digital tools determined by *t* tests. Changes pre- versus post cardiac rehabilitation were analyzed with boxplots and 2-tailed *t* tests based on matched pairs. The correlation coefficients "*r*" were calculated using Pearson's correlation coefficient and the significance of the influence was calculated using ANCOVA. If not stated otherwise, the cofounders considered in the ANCOVA were the age, P_{max}, body mass index, and the global MacNew score, all measured at the start of OUT-III.

Ethical Considerations

The EPICURE study protocol was approved by the ethics committee of Upper Austria (vote nr. 1165/2019) and registered at ClincalTrials.gov (Identifier: NCT04458727). All analysis performed during this work are covered by this ethics approval. All participants received oral and written information before the study entry and provided written informed consent to participate in this study. This informed consent also covered the analysis of secondary study objectives performed in this paper. All data were recorded in a pseudonymized way and anonymity was ensured for all participants during presentation of the findings. Patients did not receive any financial incentives for participating in the study.

Results

Table 2 shows the mean (SD) of the quality-of-life as reported by the MacNew questionnaire (mean per subcategory, 1…strongly disagree, 7… strongly agree). The number of data points available for each analysis is shown in brackets. Subcategories (physical, emotional, and social) scores and an overall score aggregated over all categories (global), pre- and post OUT-III, as well as the difference between pre- and post OUT-III are shown (*P* value pre vs post). Differences between the groups with and without digital tools are provided (*P* value between groups). As illustrated in Table 2, the QoL according to the MacNew score aggregated over all subcategories (global), significantly improved in patients without ($P=.01$) as well as patients with digital tools ($P=.01$). A significant increase in the physical subquestionnaire (physical) has been identified for the group without ($P=.01$) and with digital tools ($P=.04$). The group with digital tools showed a significant improvement concerning the emotional QoL subsection ($P=.01$) while the group without digital tools achieved a significant increase in the social QoL score ($P=.01$).

XSL•FO
RenderX

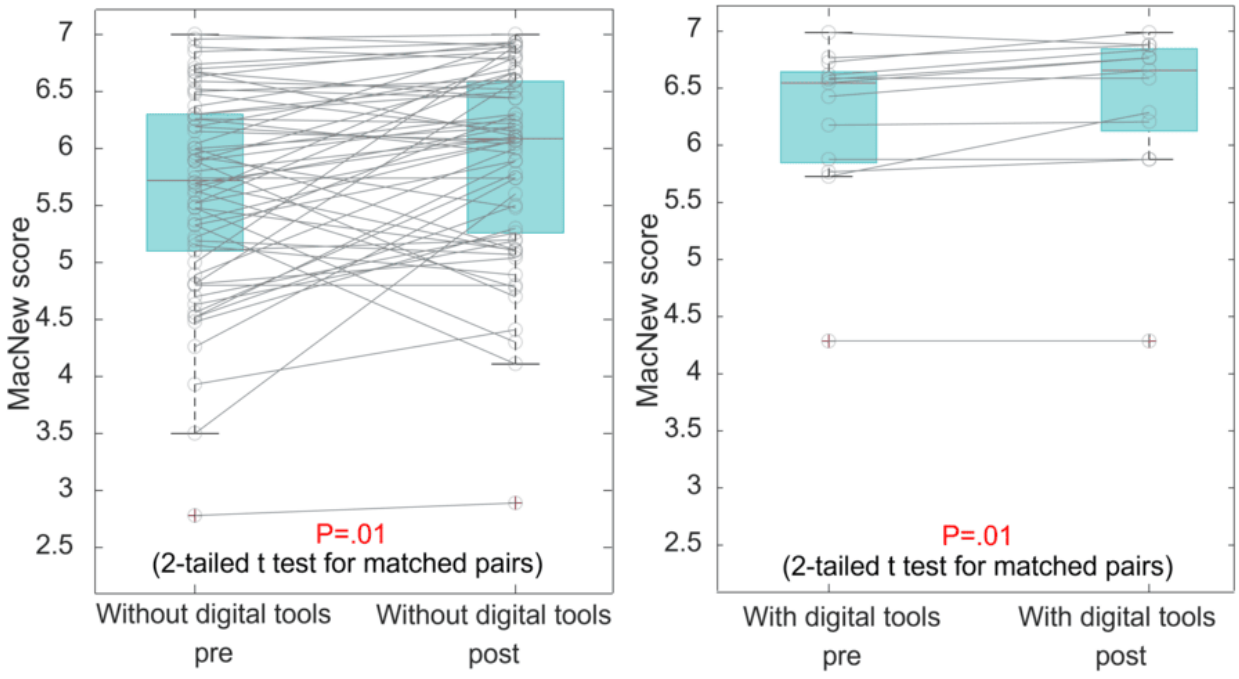
Table . Mean (SD) of the quality of life as reported by the MacNew questionnaire (mean per subcategory, 1…strongly disagree, 7… strongly agree).

Subcategories	Without digital tools		With digital tools		*P* value between groups
	Mean (SD)	Number of data points, n	Mean (SD)	Number of data points, n	
Physical					
Pre	5.64 (0.93)	74	6.15 (0.81)	32	.02
Post	5.92 (0.82)	63	6.38 (0.8)	13	.04
P value pre versus post	.01	62	.04	13	.21
Emotional					
Pre	5.57 (0.91)	74	6.01 (0.8)	32	.01
Post	5.71 (0.95)	63	6.4 (0.7)	13	.01
P value pre versus post	.15	62	.01	13	.94
Social					
Pre	5.84 (0.91)	74	6.23 (0.87)	32	.07
Post	6.11 (0.84)	63	6.42 (0.79)	13	.06
P value pre versus post	.01	62	.84	13	.78
Global					
Pre	5.68 (0.88)	74	6.11 (0.7)	32	.03
Post	5.91 (0.85)	63	6.39 (0.73)	13	.02
P value pre versus post	.01	62	.01	13	.53

The change of the global QoL pre- versus post OUT-III for both groups is illustrated in (Figure 1). Although both groups improved, no significant differences in the improvement between the 2 groups were found according to the ANCOVA analysis, neither for the global MacNew score nor for any of the subcategories. Markers in pre- and post which are connected by lines correspond to one and the same patient.

As shown in Table 3, there were no correlations between changes in QoL scores and ΔP_{max}.

Figure 1. Quality of life according to the global MacNew score pre- versus post outpatient cardiac rehabilitation phase III for patients without digital tools (left) and with digital tools (right), illustrated as boxplots with matched pairs.



XSL•FO
RenderX

Table . Pearson's correlation coefficient *r* (*P* value calculated by analysis of covariance) between changes of quality of life in subcategories physical, emotional, and social as well as the global overall score as reported by the MacNew questionnaire and the change in exercise capacity ΔP_{max}.

Subcategories	Correlations with ΔP_{max}	
	r value	*P* value
ΔPhysical	0.2	.16
ΔEmotional	0.09	.84
ΔSocial	0.18	.28
ΔGlobal	0.19	.20

Intrinsic Motivation Scale

Table 4 shows results per subsection of the Intrinsic Motivation Scale (IMS) as well as the aggregated score over all 4 categories. A significant difference in the item group concerning perceived competence was found between patients with versus without digital tools (*P*=.05), however, no significant correlation with ΔP_{max} was identified (*P*=.57). In addition, the IMS score for interest and enjoyment and the total IMS score showed borderline significant differences between the groups. In all significant and borderline significant cases, the group with digital tools achieved higher scores.

Table . Mean (SD) of the Intrinsic Motivation Scale (1…strongly disagree, 5…strongly agree) and differences between the groups and Pearson's correlation coefficient *r* (*P* value calculated by analysis of covariance) between the Intrinsic Motivation Scale and change in exercise capacity ΔP_{max}.

Subcategories	Influence of the group			Correlation with ΔP_{max}	
	Without digital tools, mean (SD)	With digital tools, mean (SD)	*P* values	*r* values	*P* values
Interest and enjoyment	2.58 (0.83)	3.16 (0.68)	.08	0.27	.41
Perceived choice	3.15 (0.78)	3.49 (0.69)	.50	0.20	.39
Perceived competence	2.58 (0.78)	3.24 (0.78)	.05	0.23	.57
Felt pressure and tension	2.80 (0.84)	3.10 (0.91)	.66	0.06	.31
Total	2.77 (0.62)	3.25 (0.57)	.09	0.25	.74

Health Action Process Approach

As summarized in Table 5, we did not identify a significant difference between the patients without and with digital tools concerning the self-efficacy (*P*=.95). In addition, there was no significant correlation between self-efficacy and the change in exercise capacity ΔP_{max} (*P*=.16). We did not identify any significant differences between patients without and with digital tools regarding the intention, control, and planning scores. However, the planning score was significantly correlated with ΔP_{max} (*P*=.04, *r*=0.14). Table 5 also shows results concerning the patients' expected health benefit of physical activity (outcome expectations) and cardiac risk of nonadherence to physical activity (risk perception). Borderline significance (*P*=.06) was found between outcome expectations and ΔP_{max}. Patients with digital tools reported to expect significantly higher cardiac risks, if they were adherent to physical activity (*P*=.03).

Table . Mean (SD) of health action process approach scores (Self-efficacy: sum of 10 items, each item 1…strongly disagree, 4…strongly agree; all other parameters: mean per category, 1…strongly disagree, 4…strongly agree) per group and difference between group including *P* value and Pearson's correlation coefficient *r* (*P* value calculated by analysis of covariance) between the scores and change in exercise capacity ΔP_{max} during OUT-III.

Subcategories	Influence of the group			Correlation with ΔP_{max}	
	Without digital tools, mean (SD)	With digital tools, mean (SD)	*P* values	*r* values	*P* values
Self-efficacy	31.80 (3.91)	32.20 (5.55)	.95	0.08	.16
Intention	3.33 (0.49)	3.46 (0.5)	.92	0.11	.36
Control	3.22 (0.59)	3.51 (0.52)	.21	0.17	.13
Planning	2.90 (0.57)	3.15 (0.62)	.10	0.14	.04
Outcome expectations	3.75 (0.54)	3.88 (0.33)	.82	0.13	.06
Risk perception	2.95 (0.83)	3.31 (0.85)	.028	0.02	.57

XSL•FO
RenderX

Self-Reported Adherence to the Exercise Program

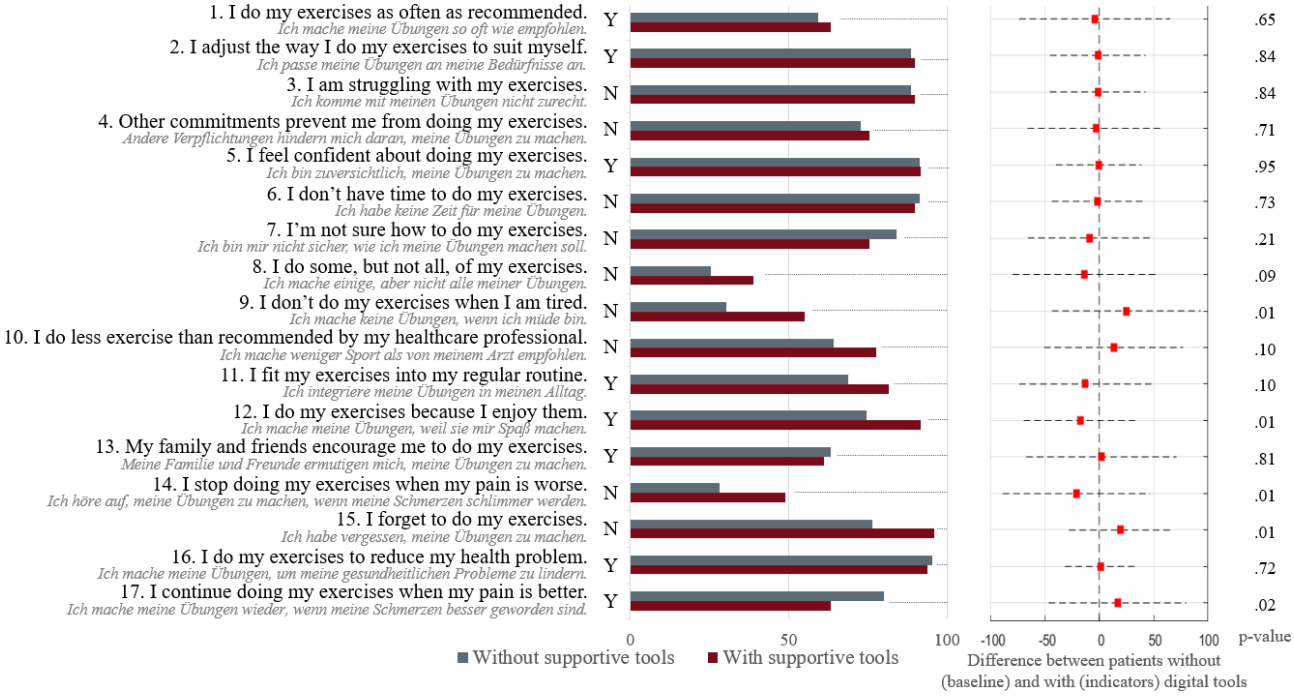
As shown in Figure 2, patients with versus without digital tools significantly differed in five out of 17 EARs related questions. The left of Figure 2 shows the percentage of patients per group, answering with Yes or No as indicated left to the bars. Yes or No are aligned alternatively, depending on positive or negative phrasings of the respective questions, so that end-to-end bars relate to positive behavior. The right of Figure 2 shows the difference of the score per item between patients with versus without digital tools, including the corresponding *P* value.

Concerning adherence-related items, patients without digital tools reported to do their exercises significantly less often when tired (item 9). Patients with digital tools reported to forget to do their exercises significantly less often (item 15).

Concerning reasons for nonadherence, item 12 ("I do my exercises because I enjoy them") was significantly more important in patients with digital tools. Pain-related items were more important in the group without digital tools (item 14 and item 17), although high scores of item 14 represent low adherence, while high scores for item 17 relate to high adherence.

The overall EARs score showed no significant difference between the patients without (11.80 [2.7]) and patients with (12.80 [2.79]) digital tools (*P*=.5). No significant correlation between the overall score and ΔP_{max} was found (*P*=.21, *r*=0.07). Also, the adherence specific score as calculated by items 1, 3, 8, 10, 11, and 15 resulted in no significant difference between patients without (3.79 [1.71]) and patients with (4.38 [1.58]) digital tools (*P*=.57). No correlation between the adherence specific score and ΔP_{max} was found (*P*=.27, *r*=0.01). These scores were calculated with the sum of the selected items (0…no, 1…yes).

Figure 2. Results of the subjective adherence questionnaire per group with the original questions in German (gray) and the English translation (black).



Discussion

Principal Findings

In this work, we analyzed the secondary hypotheses of the EPICURE study (primary outcomes have been published previously [18]), that is, the effects of digital tools in OUT-III. Our main findings are that, according to the MacNew questionnaire, QoL improved significantly during OUT-III, with similar improvement in patients without and with digital tools indicating that the use of digital tools did not significantly influence the increase of QoL during OUT-III. Furthermore, we found that patients with and without digital tools significantly differed in the following aspects:

1. Patients using digital tools perceived a significantly higher competence during cardiac rehabilitation compared with patients without digital tools, suggesting that either digital tools improve the competence or patients feeling more competent tend to use digital tools more often.
2. Patients with digital tools reported significantly more often to expect an increased risk for future cardiac events if they were not adequately physically active in the future as determined by the health action process approach.
3. Patients with digital tools were significantly less likely "not to do their exercises when they are tired" and to "forget to do their exercises." Concerning reasons for (non) adherence, patients with digital tools significantly more often reported to do their exercises because they enjoyed them. In addition, patients with digital tools were significantly less likely to stop exercising when pain got worse and to continue doing their exercises when there was a relieve in pain. However, the subjectively reported overall adherence was not significantly different in the 2 groups.

XSL•FO
RenderX

4. Patients who reported a high level of concrete planning to perform exercise training during cardiac rehabilitation achieved significantly higher improvements in ΔP_{max}.

Comparison With Previous Work

The previously published EPICURE paper [18] summarized that, overall, exercise capacity improved both in patients with and without digital tools. The exercise capacity was measured during ergometry in 5 different Austrian rehabilitation centres. No data concerning the inter-rater reliability between the centres was available for analysis. When comparing changes with *t* tests, patients with digital tools improved significantly more than patients without digital tools. However, this change did no longer reach statistical significance when correcting for confounders with an ANCOVA analysis. In this paper, wherever suitable, *P* values were based on ANCOVA corrected for confounders.

Quality of Life

Jones et al [6] reported that digital tools resulted in a significantly higher improvement of QoL for the home-based compared with the center-based cardiac rehabilitation group. Although we identified a significant improvement in both groups, no significant difference in the improvement of the QoL between patients with or without digital tools was found. In both groups, QoL was already high before OUT-III, with a mean of 5.68 (without) and 6.11 (with digital tools) out of 7 possible points. These high QoL scores before OUT-III in combination with a small sample size of patients with MacNew data pre- and post OUT-III, especially in the group with digital tools (n=13), might explain the difference between Jones et al [6] and our results. Jones et al [6] also analyzed multiple studies of which the majority (44 out of 55) were randomized controlled trials, which differed from the retrospective EPICURE setting.

Intrinsic Motivation

The higher IMS scores in the group with digital tools are in line with Lu et al [7], who identified patients' intrinsic motivation as a key facilitator for successful rehabilitation of chronically ill patients and patients on the verge of being chronically ill with remote activity sensing. Similarly to the EPICURE study, Lu et al [7] conducted retrospective analysis on a smaller patient cohort of 23 patients which combined a mixed-methods approach, of retrospective questionnaires and interviews with real time activity sensing using wearable devices. Our results indicate that digital tools can also help to motivate cardiac patients, without exposing patients to additional distress. However, since IMS was asked retrospectively, patients who knew that they were adherent may have been influenced by this knowledge when retrospectively evaluating their motivation before the program. Therefore, and since there is only little literature on this topic so far, prospective studies might be indicated to get a deeper insight into the relationship between intrinsic motivation and digital tools during cardiac rehabilitation.

Health Action Process Approach

Unlike a previous study by Salah Eldin Saad et al [30] which showed significant differences in self-efficacy, our study indicates that self-efficacy neither differed between patients with and without digital tools, nor was there a significant correlation with the achieved change in exercise capacity. In part, this result can be explained by the retrospective nature of the EPICURE study compared to a prospective setting of Salah Eldin Saad et al [30] who divided the patients equally before the cardiac rehabilitation. However, the relationship between self-efficacy, digital tools and outcome of cardiac rehabilitation might further be studied in the future. Additionally, the intervention of Salah Eldin Saad et al [30]was based on regular assessments of patients' progress and provided cardiac rehabilitation relevant information weekly to patients using digital tools, ensuring a personalized and tailored rehabilitation process. Even though personal phone-based assessments were also part of the EPICURE study, these were not structured and routinely performed [30].

The change in exercise capacity ΔP_{max} correlated significantly with the level of planning, as retrieved from ten items in a respective questionnaire (Table 5). No significant difference between patients with and without digital tools was identified. In addition, there was no significant difference or correlation concerning the intention and control dimension of the questionnaire. As Scholz et al [10] suggested in their 5-week longitudinal on-line study with 354 patients, planning predicts behavior when intention scores are high. The study by Scholz et al [10] focused more thoroughly on the planning stage the EPICURE study since patients were first encouraged to articulate their intentions regarding physical activity and subsequently outlined specific plans that detail the context and timing of their intended activities [10]. Even though this process was not implemented to this extent in EPICURE, the influence of the planning on the execution was confirmed by our findings that showed a significant correlation between the planning score and ΔP_{max}.

The outcome expectation showed a marginally significant influence on ΔP_{max}, therefore, it might be beneficial to strengthen the patients' expectation during OUT-III. Even though the risk perception did not influence the outcome ΔP_{max} we identified a significant difference between patients with and without digital tools. Since patients with digital tools had a higher score on the risk question this might be an indicator that patients who are more aware of their risk are more willing to participate in home-based cardiac rehabilitation.

Self-Reported Adherence to the Exercise Program

As previously discussed in a meta-analysis by Li et al [31] telehealth in a cardiac rehabilitation setting, which is based on digital tools (a combination of instant communication tools, smartphone applications, and wearable devices) improves exercise capacity. Home-based and community-based telehealth cardiac rehabilitation in the secondary phase were eligible interventions. These were compared with a combination of traditional cardiac rehabilitation or usual care [31]. This corresponds to our findings, since results from the exercise adherence questionnaire revealed that the use of digital tools in cardiac rehabilitation is associated with higher exercise adherence and to support patients to do their exercises even when they are tired. In addition, patients with digital tools

XSL•FO
RenderX

reported more often to do their exercises because they enjoy them. This is in line with the results from the IMS questionnaire, which also showed higher interest and enjoyment scores in the group with digital tools. Differences in the IMS scores did, however, not reach statistical significance. Since the data were analyzed retrospectively, people who decided to use digital tools might have been more motivated in general and, therefore, better in adhering to their exercise plans.

Significant differences in the subjective adherence scores related to pain are expected to be caused by different baseline characteristics concerning pain the 2 groups. However, since pain itself was not recorded in our study, these results might require further analyses in the future.

Limitations

Most questionnaire data were recorded at the time of study entry of the patients, that is, post OUT-III. Therefore, except for those questionnaires that were routinely recorded during cardiac rehabilitation (eg, MacNew questionnaire prior to OUT-III), all questions dealing with data before the end of OUT-III need to be interpreted with care, since the answers might have been influenced by the patients' experiences during OUT-III and different answers might have been given in a prospective setting.

During sample number calculations in the study preparation phase, we assumed that approx. A total of 50% of patients would report to have performed cardiac rehabilitation with and without digital tools, respectively. However, questionnaire data revealed that approx. One third did and 2 thirds did not use digital tools, which reduced the power of our analyses. Since especially the number of patients using digital tools was rather low, additional studies with larger sample size may be indicated. In addition, not all patients had valid data at the start of OUT-III. For these patients data from the end of phase II was used with the assumption that the relevant physiological and psychological characteristics did not change in the short period between phase II and OUT-III. For future work, a prospective study is warranted that randomizes patients to either a group with or without digital tools, preferably including stratification on those baseline parameters that were found to have significant influence on cardiac rehabilitation outcomes, that is, age, P_{max}, BMI, body weight, and QoL.

Conclusion

We conclude that digital tools used during OUT-III including home-training can increase perceived confidence and improve joy during cardiac rehabilitation. In addition, digital tools can help to not forget exercises and to do the exercises even when tired. Concrete planning is correlated with better improvement in exercise capacity and should be fostered during OUT-III.

Acknowledgments

This study was funded by the Austrian Research Promotion Agency (FFG) under the grant number 865531 and by the Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria. This study is supported by TU Graz Open Access Publishing Fund.

Data Availability

The data sets generated during and analyzed during this study are not publicly available due to privacy concerns and based our ethics approval but are available from the corresponding author on reasonable request.

Authors' Contributions

FW had the main responsibility for preparing the final manuscript and for the statistical analyses. DH contributed to preparing the final manuscript and interpreting the results. MS contributed to the study design, and study preparation. JN contributed to the study design, study preparation, statistical analyses and interpreting the results. KM, NM, MP, CP, L-MM, MP, HT contributed to the study design and conducted the study. MP and L-MM conducted the study. SH contributed to the study design, study preparation, statistical analyses, and interpreting the results. GS contributed to preparing the final manuscript, interpreting the results and the statistical analyses. DH had the main responsibility for setting up and conducting the study and contributed to preparing the final manuscript and the statistical analyses. All authors discussed the results, supervised the findings and contributed to the article and approved the submitted version.

Conflicts of Interest

Author Karl Mayer is employed by heartfish GmbH and CARDIOMED Kardiologisches Rehabilitationszentrum GmbH and holds 20% shares in both companies. Author Norbert Mürzl is employed by Cardio Vital Wels. Author Christoph Puelacher is employed by Reha Innsbruck, REHAmed-tirol GmbH and holds 100% of the company's shares. Author Lisa-Marie Moser is employed by Reha Innsbruck, REHAmed-tirol GmbH. Author Marco Philippi is employed by MedReha GmbH and owns 100% shares of the company. Author Heimo Traninger is employed by ZARG Zentrum für ambulante Rehabilitation GmbH and holds 10% shares of the company.

References

XSL•FO
RenderX

1. World Health Organization. Global health estimates 2020: deaths by cause, age, sex, by country and by region, 2000-2019. 2020.
2. Roth GA, Mensah GA, Johnson CO, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol 2020 Dec 22;76(25):2982-3021. [doi: 10.1016/j.jacc.2020.11.010] [Medline: 33309175]
3. Ambrosetti M, Abreu A, Corrà U, et al. Secondary prevention through comprehensive cardiovascular rehabilitation: From knowledge to implementation. 2020 update. A position paper from the Secondary Prevention and Rehabilitation Section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2021 May 14;28(5):460-495. [doi: 10.1177/2047487320913379] [Medline: 33611446]
4. Piepoli MF, Corrà U, Adamopoulos S, et al. Secondary prevention in the clinical management of patients with cardiovascular diseases. Core components, standards and outcome measures for referral and delivery. Eur J Prev Cardiolog 2014 Jun;21(6):664-681. [doi: 10.1177/2047487312449597]
5. Dalal HM, Doherty P, McDonagh ST, Paul K, Taylor RS. Virtual and in-person cardiac rehabilitation. BMJ 2021 Jun 3;373:n1270. [doi: 10.1136/bmj.n1270] [Medline: 34083376]
6. Jones AK, Yan CL, Rivera Rodriquez BP, Kaur S, Andrade-Bucknor S. Role of wearable devices in cardiac telerehabilitation: A scoping review. PLoS One 2023;18(5):e0285801. [doi: 10.1371/journal.pone.0285801] [Medline: 37256878]
7. Lu Z, Signer T, Sylvester R, Gonzenbach R, von Wyl V, Haag C. Implementation of remote activity sensing to support a rehabilitation aftercare program: observational mixed methods study with patients and health care professionals. JMIR Mhealth Uhealth 2023 Dec 8;11:e50729. [doi: 10.2196/50729] [Medline: 38064263]
8. Ntoumanis N, Ng JYY, Prestwich A, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. Health Psychol Rev 2021 Jun;15(2):214-244. [doi: 10.1080/17437199.2020.1718529] [Medline: 31983293]
9. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors - The health action process approach. Eur Psychol 2008;13(2):141-151. [doi: 10.1027/1016-9040.13.2.141]
10. Scholz U, Schüz B, Ziegelmann JP, Lippke S, Schwarzer R. Beyond behavioural intentions: planning mediates between intentions and physical activity. Br J Health Psychol 2008 Sep;13(Pt 3):479-494. [doi: 10.1348/135910707X216062] [Medline: 17553212]
11. Scholz U, Sniehotta FF, Schüz B, Oeberst A. Dynamics in self‐regulation: plan execution self‐efficacy and mastery of action plans. J Applied Social Pyschol 2007 Nov;37(11):2706-2725 [FREE Full text] [doi: 10.1111/j.1559-1816.2007.00277.x]
12. Haag D, Carrozzo E, Pannicke B, Niebauer J, Blechert J. Within-person association of volitional factors and physical activity: Insights from an ecological momentary assessment study. Psychol Sport Exerc 2023 Sep;68:102445. [doi: 10.1016/j.psychsport.2023.102445] [Medline: 37665897]
13. Ghisi GDM, Aultman C, Vanzella L, Konidis R, Sandison N, Oh P. Effectiveness of a virtual vs. in-person group-based education curriculum to increase disease-related knowledge and change health behaviour among cardiac rehabilitation participants. Patient Educ Couns 2024 Jan;118:108021. [doi: 10.1016/j.pec.2023.108021] [Medline: 37866071]
14. Yang Z, Jia H, Zhang F, Huang H, Hao X, Wang A. A behavioural driving model of adherence to home-based cardiac rehabilitation exercise among patients with chronic heart failure: A mixed-methods study. J Clin Nurs 2024 Feb;33(2):531-542. [doi: 10.1111/jocn.16901] [Medline: 37881110]
15. Yang Z, Zheng X, Hu N, Zhang F, Wang A. “Challenges to Normalcy”- perceived barriers to adherence to home-based cardiac rehabilitation exercise in patients with chronic heart failure. Patient Prefer Adherence 2023;17:3515-3524. [doi: 10.2147/PPA.S440984] [Medline: 38146500]
16. Pierobon A, Zanatta F, Granata N, et al. Psychosocial and behavioral correlates of self-efficacy in treatment adherence in older patients with comorbid hypertension and type 2 diabetes. Health Psychol Rep 2023;11(3):188-199. [doi: 10.5114/hpr/159284] [Medline: 38084262]
17. McGowan H, Gutenberg J, Leitner V, et al. Use and perceptions of digital technology for daily life, physical activity and health information among former cardiac rehabilitation patients in Austria: A cross-sectional survey. Digit Health 2023;9:20552076231219437. [doi: 10.1177/20552076231219437] [Medline: 38089166]
18. Hayn D, Sareban M, Höfer S, et al. Effect of digital tools in outpatient cardiac rehabilitation including home training-results of the EPICURE study. Front Digit Health 2023;5:1150444. [doi: 10.3389/fdgth.2023.1150444] [Medline: 37519897]
19. Niebauer J, Mayr K, Tschentscher M, Pokan R, Benzer W. Outpatient cardiac rehabilitation: the Austrian model. Eur J Prev Cardiol 2013 Jun;20(3):468-479. [doi: 10.1177/2047487312446137] [Medline: 22508693]
20. Oldridge N, Guyatt G, Jones N, et al. Effects on quality of life with comprehensive rehabilitation after acute myocardial infarction. Am J Cardiol 1991 May 15;67(13):1084-1089. [doi: 10.1016/0002-9149(91)90870-q] [Medline: 2024598]
21. Höfer S, Benzer W, Brandt D, et al. MacNew heart disease Lebensqualitätsfragebogen nach Herzinfarkt. Zeitschrift für Klinische Psychologie und Psychotherapie 2004 Oct;33(4):270-280. [doi: 10.1026/1616-3443.33.4.270]
22. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
23. Schwarzer R, Jerusalem M. General self-efficacy scale (GSE). : APA PsycTests; 1995.

XSL•FO
RenderX

24. Schwarzer R, Sniehotta FF, Lippke S, et al. On the assessment and analysis of variables in the health action process approach conducting an investigation. : Freie Universität; 2003.
25. Newman-Beinart NA, Norton S, Dowling D, et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the Exercise Adherence Rating Scale (EARS). Physiotherapy 2017 Jun;103(2):180-185. [doi: 10.1016/j.physio.2016.11.001] [Medline: 27913064]
26. Primus C, Wonisch M, Berent R, Auer J. Praxisleitlinien Ergometrie und Spiroergometrie//Practice guidelines for exercise testing. J Cardiol 2022;29(1):17-26.
27. Herzinfarkt: rehabilitation & nachsorge. : Bundesministerium für Soziales G, Pflege und Konsumentenschutz [Article in German]; 2024.
28. Reich B, Benzer W, Harpf H, et al. Efficacy of extended, comprehensive outpatient cardiac rehabilitation on cardiovascular risk factors: A nationwide registry. Eur J Prev Cardiol 2020 Jul;27(10):1026-1033. [doi: 10.1177/2047487319898958] [Medline: 31937125]
29. Niebauer J, Mayr K, Harpf H, et al. Long-term effects of outpatient cardiac rehabilitation in Austria: a nationwide registry. Wien Klin Wochenschr 2014 Mar;126(5-6):148-155. [doi: 10.1007/s00508-014-0527-3] [Medline: 24615677]
30. Salah Eldin Saad N, El Ashery Ashery Asker R. Effect of digital cardiac rehabilitation program on self efficacy of patients with coronary artery diseases. Egyptian Journal of Health Care 2020 Mar 1;11(1):400-416. [doi: 10.21608/ejhc.2020.177145]
31. Li R, Wang M, Chen S, Zhang L. Comparative efficacy and adherence of telehealth cardiac rehabilitation interventions for patients with cardiovascular disease: A systematic review and network meta-analysis. Int J Nurs Stud 2024 Oct;158:104845. [doi: 10.1016/j.ijnurstu.2024.104845] [Medline: 39032245]

Abbreviations

ANCOVA: analysis of covariance
CVD: cardiovascular diseases
EARS: Exercise Adherence Rating Scale
HAPA: Health Action Process Approach
IMS: Intrinsic Motivation Scale
OUT-III: outpatient cardiac rehabilitation phase III
QoL: quality of life

Edited by L Buis; submitted 23.02.24; peer-reviewed by D Filos, G Pepera, M Hollings; revised version received 02.12.24; accepted 25.01.25; published 08.04.25.

Please cite as:
Wiesmüller F, Haag D, Sareban M, Mayr K, Mürzl N, Porodko M, Puelacher C, Moser LM, Philippi M, Traninger H, Höfer S, Niebauer J, Schreier G, Hayn D
Clinical, Psychological, Physiological, and Technical Parameters and Their Relationship With Digital Tool Use During Cardiac Rehabilitation: Comparison and Correlation Study
JMIR Mhealth Uhealth 2025;13:e57413
URL: https://mhealth.jmir.org/2025/1/e57413
doi:10.2196/57413

© Fabian Wiesmüller, David Haag, Mahdi Sareban, Karl Mayr, Norbert Mürzl, Michael Porodko, Christoph Puelacher, Lisa-Marie Moser, Marco Philippi, Heimo Traninger, Stefan Höfer, Josef Niebauer, Günter Schreier, Dieter Hayn. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of a Computer Vision–Based Exercise Application for People With Knee Osteoarthritis: Randomized Controlled Trial

Dian Zhu[1], PhD; Jianan Zhao[2], PhD; Tong Wu[1], PhD; Beiyao Zhu[3], MD; Mingxuan Wang[1], PhD; Ting Han[1], PhD

[1]School of Design, Shanghai Jiao Tong University, Dong Chuan rd, No 800, Shanghai, China

[2]College of Fashion and Design, Donghua University, Donghua University, Shanghai, China

[3]Department of Plastic and Reconstructive Surgery, Shanghai Jiao Tong University Ninth People's Hospital, Shanghai, China

Corresponding Author:
Ting Han, PhD
School of Design, Shanghai Jiao Tong University, Dong Chuan rd, No 800, Shanghai, China

Abstract

Background: Exercise is a primary recommended treatment for knee osteoarthritis (KOA), as it helps alleviate symptoms and improves joint functionality. Personalized exercise programs, tailored to individual patient needs, have demonstrated promising results in maintaining physical fitness and enhancing overall well-being. In recent years, digital health applications have emerged as innovative tools for supervising and facilitating rehabilitation programs. Leveraging computer vision (CV) technology, these applications offer the potential to provide precise feedback and support personalized exercise interventions for patients with KOA in a scalable and accessible manner.

Objective: This study aims to evaluate the impact of a CV–graded exercise intervention application over a 6-week period on clinical outcomes in patients with KOA . The outcomes were compared to those achieved through conventional exercise education by videos.

Methods: A randomized controlled trial was conducted with 60 participants aged 60 - 80 years, recruited through community administrators between July 2023 and September 2023. Participants were randomly assigned to one of two groups: the graded exercise application group (n=32) and the exercise education brochure group (n=28). The primary outcomes assessed were short-term changes in pain, physical function, and stiffness as measured by the Western Ontario and McMaster Universities Arthritis Index (WOMAC). Secondary outcomes included assessments of participants' affective state, self-efficacy, quality of life, and user experience.

Results: The study recruited 60 participants, including 26 males and 34 females. Analysis revealed statistically significant improvements in physical function (P=.02) and self-efficacy (P=.04) in the graded exercise application group compared to the exercise education brochure group after the intervention. While improvements in pain and stiffness were observed in both groups, these changes were not statistically significant. In addition, participants in the graded exercise application group reported a positive user experience, highlighting the application's usability and engagement features as beneficial to their rehabilitation process.

Conclusions: The findings suggest that the CV-based graded exercise intervention application effectively improves physical function and self-efficacy among patients with KOA . This digital tool demonstrates the potential to enhance the quality and personalization of exercise rehabilitation compared to traditional methods. Future studies should explore the application's long-term efficacy and replicability in larger community-based populations, with a focus on sustained engagement and adherence to rehabilitation programs.

Trial Registration: ClinicalTrials.gov NCT06220565; https://clinicaltrials.gov/study/NCT06220565

(JMIR Mhealth Uhealth 2025;13:e63022) doi:10.2196/63022

KEYWORDS

knee osteoarthritis; behavior change theory; application; digital health; randomized controlled trial; knee; osteoarthritis; behavior; vision; exercise; app; applications; exercise rehabilitation; rehabilitation; older adults; self-efficacy; physical function

Introduction

Knee osteoarthritis (KOA) is a common and chronic joint disorder that is growing in prevalence as the world's population ages [1,2]. Exercise has been acknowledged as a nonpharmacological intervention modality for the treatment and prevention of musculoskeletal disorders, including osteoarthritis, osteoporosis, back pain, and rheumatoid arthritis [3]. Specifically, it has been empirically demonstrated that participating in appropriate physical exercise while being monitored by a physiotherapist effectively maintains physical health and athletic ability [4]. In order to optimize joint

XSL•FO
RenderX

flexibility, strengthen muscles, and reduce strain, they develop individualized training plans that are prescribed in accordance with the patient's condition [5]. Nevertheless, this procedure imposes considerable financial burdens on the patients and requires a significant time commitment from the caregivers [6].

Digital health interventions have the potential to mitigate the time and resource limitations faced by patients with KOA by offering education and self-management through web-based platforms or applications that are scalable, inexpensive, readily available, and high coverage [7-9]. Digital health interventions have been shown to potentially offer benefits in the management of musculoskeletal disorders [10,11]. Nevertheless, there are still several digital intervention tools that demonstrate inconsistent levels of effectiveness. For instance, physical therapy sessions and digital applications were shown to be successful in improving physical function [12], while a program using wearable devices to deliver physical activity counseling was deemed ineffective in promoting physical function [13]. This is attributed to the varying functionalities that are accessible and the absence of support for these functionalities in influencing user behavior.

Research has indicated that the integration of behavior change theory offers the potential for both positive behavior modification and amelioration of negative emotions [14]. Behavior modification approaches are comprised of three essential components: goals and planning, feedback and monitoring, and repetition and replacement [15]. Goals and planning promote increased general physical activity by setting short-term goals and developing evidence-based progressive individualized exercise plans [16]. Monitoring and feedback oversee progress and acquire immediate feedback on the accomplishment of objectives [17]. The application primarily consists of recurring and substituted generalized and graded tasks, which incorporate behavioral exercises and target behaviors [18,19]. A systematic review revealed that exercise programs through positive feedback, effort reinforcement, motivating techniques, and graded interventions had a higher likelihood of adhering to therapeutic exercise [20].

Previous research has established that patients can improve their management of illnesses through the implementation of goal-setting and planning strategies (monitoring, documenting, etc) as well as the consistent practice and substitution of behaviors (performing repetitive movements, engaging in daily timed exercise, etc) [21]. Clinical treatment is facilitated through the utilization of sensor technology, which provides objective monitoring data for noninvasive evaluation of knee function [22,23]. Despite the availability of various feedback programs formulated to assist patients in adhering to exercise regimens, a subset of patients diagnosed with KOA continue to encounter suboptimal exercise outcomes [9,24,25]. Previous reviews have highlighted that one of the primary objectives of CV-based applications is to calculate joint angles, enabling the identification and counting of correct postures during rehabilitation training. In addition, accurately assessing whether exercises are performed correctly is another critical goal, ensuring that the movement sequence aligns with rehabilitation standards [26,27]. Given the requirement for enhanced and immediate visual input to enable the execution of rehabilitation exercises, the application of computer vision (CV)–algorithms for body position tracking appears to hold considerable promise [28,29]. Nevertheless, clinical investigations evaluating most of these CV-based rehabilitation programs have not yet been conducted.

Given the tremendous potential of these applications for rehabilitation exercises, a program based on CV was developed. The application facilitates the development of a graded exercise rehabilitation program for patients and aids them in self-monitoring the program's implementation. The main objective was to assess the effects of using a CV-graded exercise intervention application (after 6 weeks) on clinical outcomes (pain and physical function) among patients diagnosed with KOA. The secondary outcome was to investigate the effects of application implementation on the affective state and self-efficacy of patients with KOA.

Methods

Study Design

This parallel, 2-arm, unbalanced randomized, single-blinded controlled trial was conducted following the CONSORT (Consolidated Standards of Reporting Trials) Checklist 1 [30] and CONSORT-EHEALTH (Electronic and Mobile Health Applications and Online Telemedicine Reporting Trials) [31] guidelines. The trial was carried out from January 2024 to May 2024 in community activity centers in Shanghai. Participants in the intervention group used CV-based exercise assessment and intervention system designed specifically for patients with KOA, while those in the control group used an exercise rehabilitation education video.

However, during the pilot study of 4 participants, it was observed that some of them lacked adequate digital health literacy and were unable to independently configure and operate the experimental devices. To address this issue, the research team modified the experimental protocol, organizing in-person sessions at community centers 3 times per week, where trained personnel facilitated the intervention tasks for participants.

Participants Recruitment

Participants underwent a 6-week intervention comprising CV-based graded exercise sessions 3 times per week, designed to deliver personalized rehabilitation exercise plans and instructional videos. Recruitment was conducted from July 2023 to January 2024[d1] in Shanghai through invitations extended by community managers and posters distributed in nursing homes.

Inclusion criteria were as follows: (1) age ≥50 years, (2) radiologically confirmed diagnosis (KL grade ≥2, pain in affected joints) in accordance with American College of Rheumatology clinical criteria, (3) mean overall pain severity ≥4-point numeric rating scale (NRS), (4) good communication and comprehension skills without significant cognitive deficits, and (5) the ability to operate an electronic device with some proficiency.

Exclusion criteria were as follows: (1) participants with severe knee pain and discomfort, (2) participants with severe organic

XSL•FO
RenderX

lesions of vital organs such as heart, brain, and kidney, and (3) participants who had been undergoing knee arthroplasty.

Only participants who finished the final 6-week assessment of the experiment and did not withdraw were included in this study.

Interventions

Intervention Group (Applications)

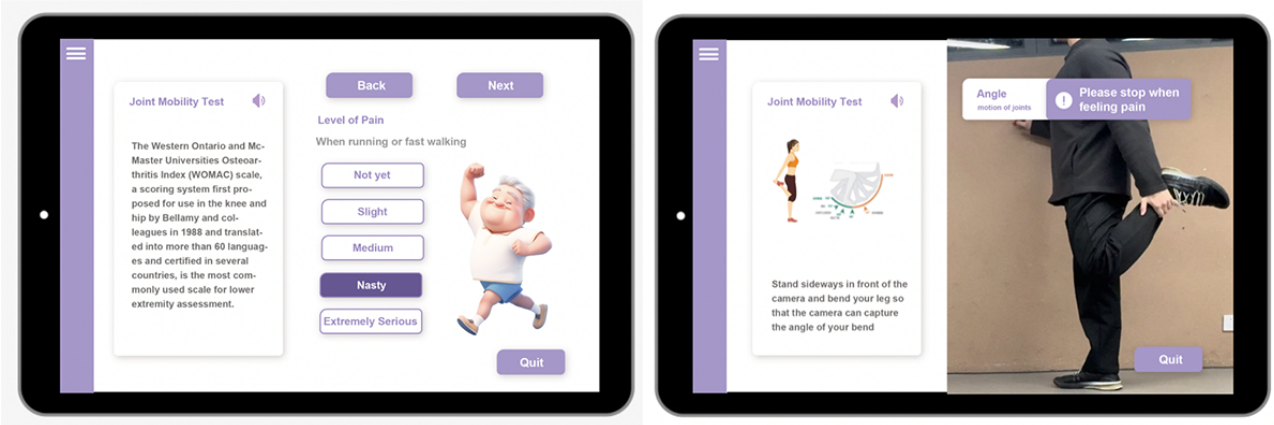
This study developed an exercise evaluation and intervention system for individuals suffering from KOA (Multimedia Appendix 1). The objective was to assist participants in preserving and enhancing their knee joint motor function by implementing a tailored progressive rehabilitation regimen and providing immediate feedback. The intervention consists of three components: (1) establishment of a comprehensive knee function assessment system; (2) delivery of secure and precise exercise interventions for participants and offer immediate feedback on the impact of exercise intervention by continuously collecting knee function metrics in real-time; (3) issuing timed tasks to encourage participants to exercise accordingly.

Status Assessment

A comprehensive knee function assessment system was developed using CV technology, an electronic Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaire, joint mobility detection, and mobile gait data monitoring. The tool offers a scientific and all-encompassing evaluation instrument for individuals afflicted with knee joint disorders (Figure 1).

First, a methodical evaluation of the participants' joint pain, rigidity, and dyskinesia was conducted with the WOMAC scale. Furthermore, CV technology was used to identify and document the frequency and quantity of movements executed by the participants throughout the physical function assessments. A physical function test involves the participants repeatedly performing a designated movement, and physical function is evaluated by tally-marking the number of movements completed within a specified time interval. CV is also used to dynamically gather the utmost angle of knee flexion. The remote gait data tracking feature relies on Apple's HealthKit, which enables the system to collect gait data over a period of time. This includes measurements such as step length, step frequency, walking speed, bipedal symmetry, and bipedal support duration. However, the user must provide authorization for the system to access this data. Continuous and extended monitoring of this data allows for evaluation of the advancement of the condition and offers participants insights into their walking patterns and behavioral habits.

Figure 1. Knee function evaluation system.



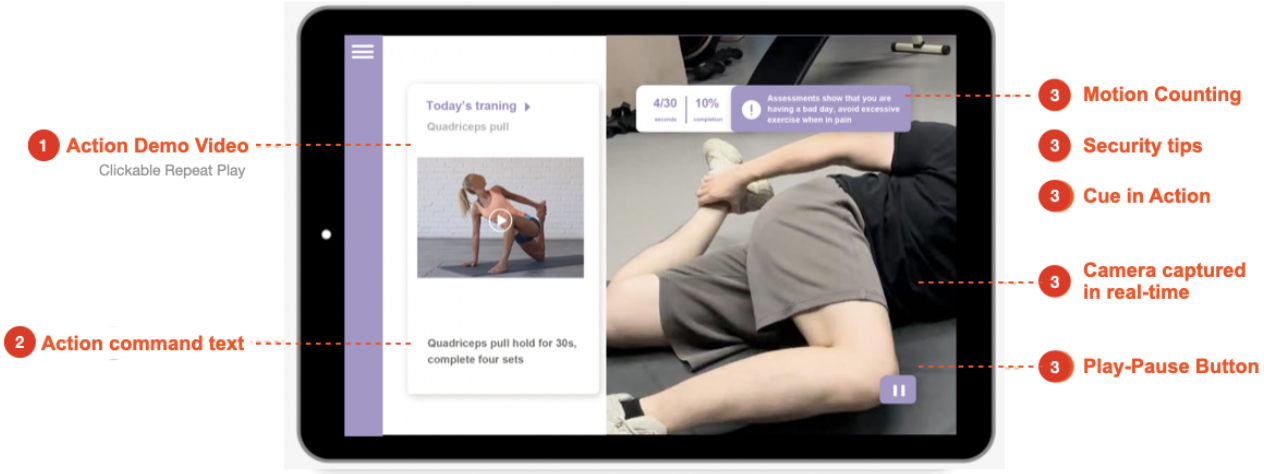
Action Detection

Standardizing movements during exercise interventions has an impact on the body's dynamics during exercise and physical function testing. Hence, aside from the completion status of the action, they also prioritize factors such as the precision and efficacy of the movement (Figure 2).

Through the application of CV technology, the system is able to capture the participants' movement posture. The system collects data from posture capture to create a real-time analysis of movement completion as the participants conduct the scheduled exercise. This enables immediate feedback to be given to the participants during the training regimen. The screen displays a physical function test demonstration, which the older adults observe and subsequently replicate. The software generates a tally of the older adults' completed motions within a specified time frame. Throughout the entire procedure, the program is equipped with a function that can assess hazardous scenarios. If a fall is imminently possible, the physical function exercise program will promptly halt and emit an audible alert. Upon completion of the test, the older adults could access and review the results displayed on the screen. By conducting regular physical function assessments, individuals can acquire an impartial comprehension of the condition of their knee joint and obtain an accurate assessment of their physical mobility that is supported by the knee joint.

XSL•FO
RenderX

Figure 2. Application of computer vision technology.



Exercise Rehabilitation

Following the evaluation process, participants are granted the ability to implement a customized exercise regimen that aligns with their rehabilitation strategy and integrates digital evaluation techniques to obtain immediate feedback and suggestions for modifications. The goal is to prioritize the rehabilitation requirements of the participants through the implementation of a customized rehabilitation program and consistent remote care.

The primary purpose of the exercise intervention is to implement an interactive exercise program that captures and assesses the participants' movements by CV during exercise. The user performs the prescribed movements in a sequential manner in order to achieve the daily workout objectives. The subsequent interface will present the exhibited actions, and older individuals can imitate each movement by following the demonstration video (Table 1). A brief interval is observed between each set of movements, and the exercise routine concludes once the task of the designated movements has been completed. The degree of standardization of the movement is denoted by the range of motion, and feedback pertaining to this degree of standardization is additionally acquired following each movement completion.

Table . Description of rehabilitation exercise movements.

Action	Action instructions
Calf raises	With this calf raise exercise, calf muscles can be strengthened so that patients can walk and climb stairs with ease.
Lateral raise (leg)	This side leg raise will strengthen the hip abductors and make everyday tasks like getting in and out of the car easier.
Knee lift exercise	Strengthen the muscles around the buttocks with this knee lift exercise.
Seated knee lift	It can be used to improve the ability to get up from a chair, get in and out of bed or a car. Patients can start with the easiest and then slowly progress to the hardest.
Seated rise	Increase the flexibility and strength of the hip extensors with this standing hip extension exercise.
Hip stretches	Maintain range of motion in the legs with heel and toe-tapping exercises.
Heel and toe tapping	Improved balance and coordination through toe-tapping exercises.

Remote Follow-Up

The primary functions also encompass the formation of a rehabilitation community and conducting remote follow-up. In the context of rehabilitation, incentives like as competition and leader-boards are implemented to motivate participants to actively engage in remote therapy and foster mutual contact within the rehabilitation community. Another purpose of the remote follow-up feature is to establish remote communication channels with outpatient doctors in order to minimize the time and effort participants spend on medical treatment.

Control Group (Exercise Rehabilitation Education Video)

Participants in the control group received a video with instructions for daily exercise movements, which were similar to those in the application (Table 1).

Outcomes

Participants received validated digital questionnaires at baseline, after 6 weeks of intervention, or face-to-face questionnaire completion at offline follow-up. Participants did not receive (financial) incentives or other compensation for completing the

XSL•FO
RenderX

questionnaire or the study. Demographic data were collected at baseline and described in detail in the outcome measures section.

Main Outcomes

Pre and postintervention status was assessed by the WOMAC scale [32]. Respondents reported the severity of their usual arthritis pain from 0 (none) to 20 (severe), functional impairment due to osteoarthritis from 0 (none) to 68 (severe), and knee stiffness from 0 (none) to 8 (severe).

Secondary Outcomes

The Arthritis Self-Efficacy Scale

The Arthritis Self-Efficacy Scale (ASES-8) assesses participants' confidence in their ability to manage arthritis pain and its impact on function [33,34]. Scores range from 1 (very uncertain) to 8 (very certain). Responses are averaged, with higher scores indicating higher self-efficacy.

Geriatric Depression Scale (GDS)

Negative emotions were measured by the Geriatric Depression Scale (GDS), whose 30 entries represent the symptoms of depression in old age and contain low mood, decreased activity, irritability, withdrawal from painful thoughts, and negative appraisal of the past, present, and future [35]. Scores range from 0 (none) to 30 (extremely). Higher scores indicate greater positive or negative affect.

Quality of Life

Quality of Life (AQoL-6D [Assessment of Quality of Life-6D] version) ranges from -0.04 to 1.00, with higher scores indicating better quality of life [34].

Range of motion

Range of motion [36], consisting of both extension (ie, the ability to straighten the leg) and flexion (ie, the ability to bend the leg), with scores ranging from 0 to 135 (higher scores indicate better joint mobility).

User Experience Questionnaire

The User Experience Questionnaire (UEQ), which is used to quickly assess the user experience of an interactive product, includes the traditional metrics of ease-of-use aspects: efficiency, comprehensibility, trustworthiness, attractiveness, motivation, and freshness [37].

Sample Size

The sample size for this study was determined using a "Sample Size Calculation for Two-Group Mean Comparison." Considering the characteristics of the intervention and control groups (digital application group and education group, respectively), the calculation was informed by the parameters reported in previous studies [38,39].

The sample size calculation was conducted assuming a difference between the intervention and control groups ($\mu_1 - \mu_2$) of 8.8 (SD 9) Using a 2-sided *t* test with a significance level (α) of .05 and a statistical power (1-β) of 80%, the minimum required sample size was estimated to be 36 participants, with 18 participants in each group. This ensured adequate power to detect meaningful differences between the groups.

Based on previous experience from our research team's extensive work in community-based digital rehabilitation, we have observed a high dropout rate among older adults in such settings. To address this, we implemented over-recruitment, increasing the sample size by at least 50% beyond the calculated minimum to ensure sufficient statistical power for final analyses.

Randomization, Allocation Concealment, and Blinding

The randomization process will involve stratifying the 60 participants into 15 blocks, each containing 4 participants. Within each block, participants will be randomly allocated to either the intervention group or the control group (received the exercise rehabilitation education video). Experienced statisticians used SAS (SAS Institute) for randomization grouping to ascertain that the random assignment of researchers was concealed for treatment assignment. Due to the nature of the study, participants were not blinded as they knew whether they received the program during the intervention period, while the statisticians were blinded to the group assignment.

Following the assignment, individuals in the intervention group were sent an email by the researcher (JZ) containing instructions on how to access the computerized visual grading exercise application and details regarding the field trial arrangements. Participants in the control group were notified via email about their assignment and received an exercise rehabilitation education video. Throughout the trial, participants were provided with the option to contact the researchers by phone or email if they had any inquiries regarding the application or the study.

Statistical Methods

The analyses were performed by statisticians (MW and TW) on SPSS 26.0 (IBM) software and following the intention-to-treat principle. Baseline characteristics of participants who provided and did not provide the primary outcome were compared using *t* tests or X^2 tests. The Mann-Whitney *U* test was employed to examine between-group and within-group data pre and post experiment. For this work, a significance level (α) of .05 was established for all statistical analyses. Differences were deemed statistically significant for the purposes of this study if the *P* value was less than .05.

Ethical Considerations

The clinical trial was registered (NCT06220565) and approved by the Ethics Committee of Shanghai Jiao Tong University (H20240039I). All participants were required to complete an informed consent form and were informed that their data would remain anonymous. After the completion of the experiment, each participant will receive a US $50 financial compensation.

Results

Baseline

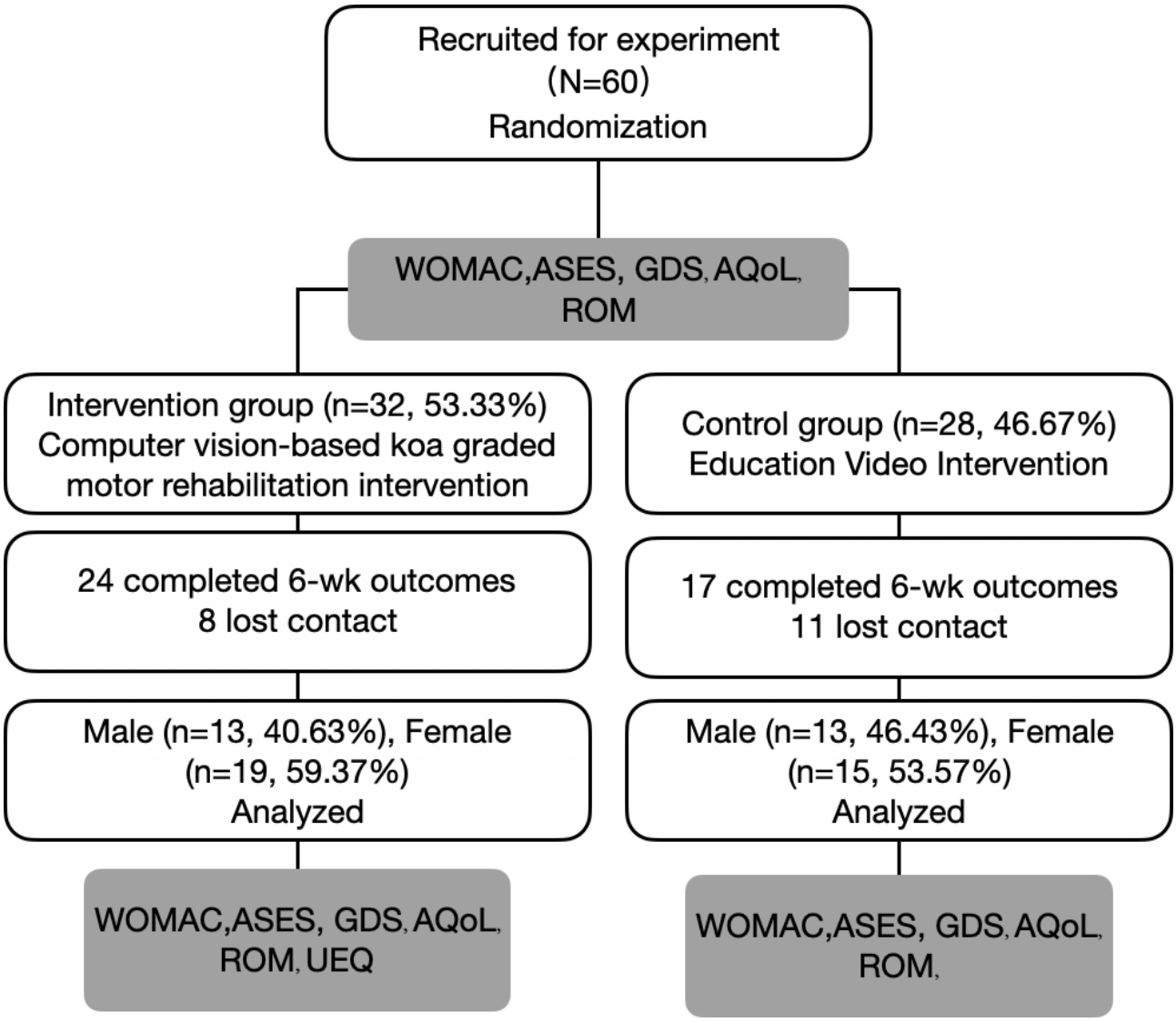
Between June and September 2023, a randomization process involving a total of 60 people was conducted who fulfilled the recruiting criteria (32 in the application group and 28 in the educational video group) (Figure 3). A total of 41 (68.33%)

XSL•FO
RenderX

participants successfully fulfilled the major outcome measure after 6 weeks, including 24/32 (75%) participants from the application group and 17/28 (60.7%) participants from the instructional video group. The average (SD) age of the sample was 68.8 (SD 6.4), and most of the sample was female (37/60, 61.67%). There were no significant differences between the groups at baseline (Table 2).

Initially, with the exception of 2 individuals, all participants maintained the belief that the rehabilitation application would aid them in the more precise rehabilitation of their motions. In addition, 87% (135/155) of participants believed that the application would aid them in recovering physical function more effectively. A considerable percentage indicated that they were not presently engaged in active therapy, despite experiencing knee discomfort. There were very few negative incidents reported, with only one instance reporting a negative incident.

Figure 3. The flowchart of the experiment. AQol: Assessment of Quality of Life; ASES: Arthritis Self‐Efficacy Scale; GDS: Geriatric Depression Scale; ROM: range of motion; UEQ: User Experience Questionnaire; WOMAC: Western Ontario and McMaster Universities Arthritis Index.



XSL•FO
RenderX

Table . Baseline characteristics of participants.

Outcome	Intervention (n=32）	Control (n=28)	All (N=60）	*P* value
Age (years), mean (SD)	68.59 (7.16)	67.64 (7.85)	68.15 (7.28)	.82
Female, n (%）	19 (60)	15 (54)	35 (57)	.79
Medication, n (%）				
Yes	11 (34)	10 (36)	21 (35)	.43
No	21 (66)	18 (64)	39 (65)	
Symptom duration (years), n (%）				
<2	7 (22)	6 (23)	13 (22)	.98
2 - 10	21 (66)	18 (63)	39 (65)	
>10	4 (12)	4 (14)	8 (13)	
Primary outcome (WOMAC)[a], mean (SD)				
Physical function	24.13 (2.7)	25.39 (2.99)	24.72 (2.9)	.12
Pain	8.87 (1.59)	8.64 (1.52)	8.76 (1.57)	.98
Symptoms	3.47 (1.19)	3.39 (1)	3.43 (1.1)	.9
Secondary outcome, mean (SD)				
ASES-8[b]	49.97 (7.49)	48.32 (7.25)	49.97 (7.37)	.41
GDS[c]	6.28 (1.84)	6.32 (0.08)	6.30 (1.8)	.9
AQoL-6D[d]	0.65 (0.1)	0.58 (0.08)	0.62 (0.09)	.06
ROM[e]	104.75 (7.16)	103.93 (7.93)	104.37 (7.48)	.57

[a]WOMAC: Western Ontario and McMaster Universities Arthritis Index.

[b]ASES-8: Arthritis Self-Efficacy Scale–8.

[c]GDS: Geriatric Depression Scale.

[d]AQoL-6D: 6-Dimensional Assessment of Quality of Life scale.

[e]ROM: range of motion.

Primary Outcomes

Table 3 and Figure 4 describe the application program in the intervention and control groups. At week 6, the application group demonstrated improvement in physical activity impairment (*P*=.02) compared to the control group, and 15/24 participants reported subsequent adherence to home exercise or community fitness. Although there was a significant within-group difference in the intervention group before and after the intervention (*P*<.001), there was no evidence of a significant improvement in pain (*P*=.96) between the intervention group and the control group. In addition, there were no within-group (*P*=.21) or between-group (*P*=.49) differences in stiffness before or after the intervention group experiment.

XSL•FO
RenderX

Table . Primary and secondary intervention outcomes.

	Intervention group (n=24）			Within group	Control group (n=17）			Within group	Between group
	Baseline	6-wk	Difference (95% CI)	*P* value	Baseline	6-wk	Difference 95% CI	*P* value	*P* value
Primary outcomes (WOMAC)[a]									
Physical function	24.17 (3.24)	17.08 (2.7)	−7.08 (−9.98 to 4.18)	.00[b]	24.76 (3.27)	18.71 (2.99)	−6.06 (−7.97 to −4.15)	.42[c]	.02[c]
Pain	8.83 (1.79)	7.21 (1.58)	−1.63 (−3.38 to 0.06)	.00[b]	8.94 (1.82)	7.76 (1.52)	−1.18 (−3.3 to 0.94)	.00[b]	.96
Symptoms	3.5 (1.25)	3.25 (1.19)	−0.25 (−1.19 to 0.69)	.21	3.41 (1)	3.35 (0.99)	−0.06 (−0.81 to 0.69)	.74	.49
Secondary outcomes, mean (SD)									
ASES-8[d]	49.79 (6.96)	51.75 (6.54)	1.96 (-2.3 to 6.22)	.003[c]	48.32 (7.25)	49.47 (7.37)	−0.47 (−4.44 to 3.5)	.79	.04[c]
GDS[e]	6.33 (1.86)	5.58 (1.83)	−0.75 (−3.06 to 1.56)	.13	6.24 (1.3)	5.65 (1.8)	−0.59 (−2.9 to 1.72)	.42	.98
AQoL[f]	0.45 (0.18)	0.48 (0.19)	0.05 (−0.09 to 0.19)	.08	0.55 (0.19)	0.52 (0.21)	−0.03 (−0.23 to 0.17)	.56	.09
ROM[g]	104.63 (7.72)	104.96 (7.79)	0.33 (−8.58 to 9.24)	.76	103.29 (7.82)	104.23 (9.52)	0.94 (−8.62 to 8.62)	.59	.58

[a]WOMAC: Western Ontario and McMaster Universities Arthritis Index.

[b]P<.001.

[c]P<0.05.

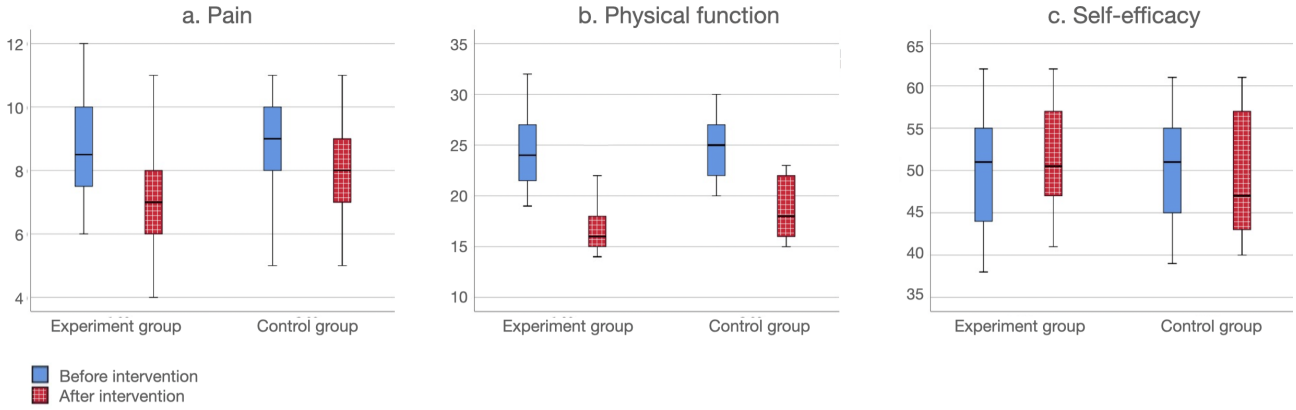
[d]ASES-8: Arthritis Self-Efficacy Scale–8.

[e]GDS: Geriatric Depression Scale.

[f]AQoL: Assessment of Quality of Life.

[g]ROM: range of motion.

Figure 4. Box plots illustrating pre and postintervention comparisons of the outcomes.



Secondary Outcomes

The secondary outcome data showed a significant difference between the 2 groups in terms of the change in self-efficacy after 6 weeks (*P*=.04). Nevertheless, the study found no evidence to suggest that the application program intervention led to a noteworthy enhancement in geriatric depression (*P*=.93) and range of motion (*P*=.58) as compared to the control group. The intervention group showed a slight improvement in quality of life. However, this improvement was not statistically significant overall (*P*=.09).

User Experience

The scoring results of the 6 dimensions of the UEQ are shown in Table 4. According to the scoring results, it can be visualized that “Attractiveness,” “Clarity” and “Novelty” received high scores (Figure 5). The exercise intervention and functional assessment interface used in this study received the highest

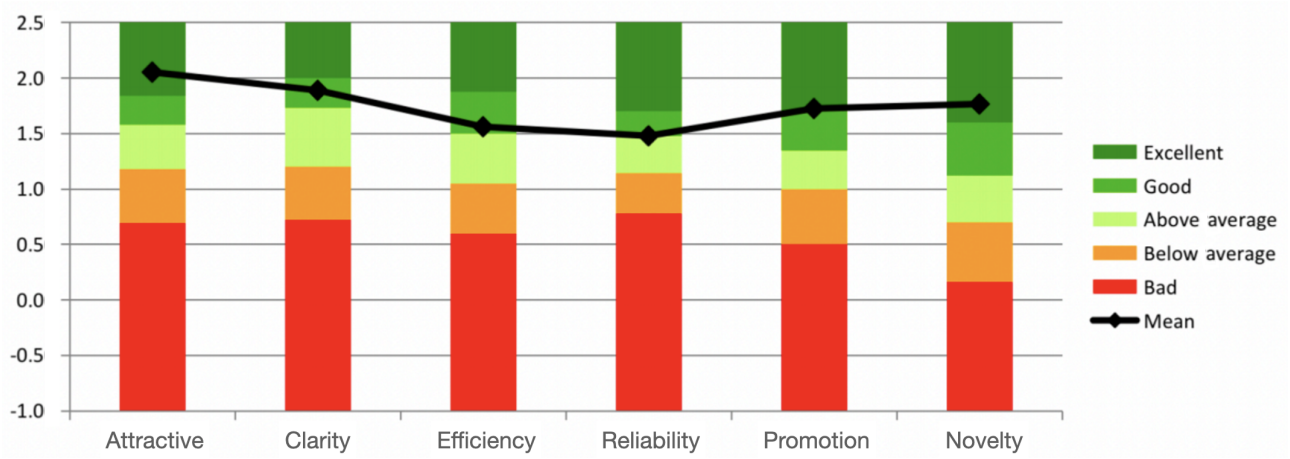
XSL•FO
RenderX

score for "attractiveness," indicating that it really appealed to the patients in the experimental group. The "Clarity" metric, representing the second highest score, suggests that the product presentation is straightforward and easily comprehensible for the user. The elevated score of "Novelty" suggests that the product possesses a commendable level of innovation and captivation, hence enticing customers to engage with it. However, Conversely, "Reliability" receives the lowest rating.

Table . User Experience Questionnaire score.

Dimension	Mean (SD)	Rating
Attractive	2.05 (0.33)	Excellent
Clarity	1.89 (0.25)	Good
Efficiency	1.56 (0.17)	Good
Reliability	1.48 (0.48)	Above average
Promotion	1.72 (0.2)	Excellent
Novelty	1.76 (0.35)	Excellent

Figure 5. Quality rating score chart.



Discussion

Principal Findings

The extensive incorporation of digital applications into the rehabilitation process for patients with KOA has been driven by the overarching goal of minimizing the duration of time for them to access specialized rehabilitation experts [5]. More precisely, it is developed based on the thorough enhancement of the widely used framework called "Behavior Change Technique," as it has a greater propensity to support patients in actively participating in the day-to-day management of their chronic illness [16,25,39]. In addition, the upward trend in the proportion of older individuals who access the Internet demonstrates the potential for internet-based interventions designed to assist this demographic [40]. In addition, studies have yielded inconsistent results concerning the effectiveness of remote rehabilitation for individuals with KOA, primarily attributable to the lack of oversight and guidance [41]. Consequently, a suite of CV-based applications was created to aid patients in the execution of rehabilitation maneuvers with accuracy and to track their advancements through the utilization of noninvasive technology. The results of the study indicated that individuals with KOA who made use of the application reported significantly improved physical function and self-efficacy when compared to the control group.

Research has shown that CV technology–based applications can enhance participants' physical functioning and alleviate discomfort. This is due to the fact that the capability of integrating behavioral change into digital intervention technologies has been designed to facilitate and encourage adjustments to the patients' decision-making framework, thereby promoting behavioral change among participants with KOA [42]. The educational aspect of the digital tool encompasses knowledge about the pathology and etiology of osteoarthritis, treatment based on established standards, exercise for osteoarthritis, and methods to alleviate pain and symptoms through adopting healthy behaviors [43,44]. This readily accessible, semisupervised intervention has the potential to be an effective means of alleviating pain and physical dysfunction [39]. Furthermore, our study revealed that the application group experienced a significantly greater improvement in physical functioning compared to the control group. The finding is reminiscent of a study conducted by Tore et al [41], which showed that the quality of physical therapy obtained by telerehabilitation was notably superior to self-management. The difference in quality could perhaps be attributed to the kind and length of the training activities undertaken [45,46]. By means of CV, our application tracks and evaluates the motion of users, thereby facilitating the integration of a customized exercise regimen and aerobic regimen. Contrary to previous research, the investigation did not identify any correlation between the

XSL•FO
RenderX

application program and an enhancement in rigidity. Variations in the duration of the studies and the quality and scope of functional support provided may account for this [39]. The reduction in stiffness can be attributed to the fortification of the leg and abdominal muscles; however, a short-term intervention fails to yield substantial outcomes [47,48].

Furthermore, our research revealed that the utilization of a CV-based application resulted in a favorable impact on self-efficacy. Consistent with findings from previous research [11,49,50], digital self-monitoring programs have been associated with relative increases in self-efficacy activation. The promotion of self-management behaviors is facilitated by the provision of information and individualized instruction pertaining to arthritis care, which encourages and assists in the resolution of individual obstacles [51]. Since the factors that influence exercise behavior are motivational and volitional in nature [52,53]. it can be concluded that the application's monitoring features provided users with the capability to observe their progress and receive prompt feedback on the degree to which their goals were achieved. Conversely, the intervention did not yield any discernible statistically significant effects on negative affect or quality of life. This discovery is consistent with previous inquiries [54], which have shown that digital health interventions have minimal to no effect on adverse emotions and quality of life. Future work should underlie mechanisms responsible for these feelings in patients.

In conclusion, the researchers determined, based on the responses to the user experience questionnaire, that the design principles' objectives for our application had been effectively met. A majority of the respondents indicated that their encounter with the e-exercise intervention was positive. They believed that the interactive design and real-time feedback enhanced the exercisers' enjoyment and promoted rehabilitation initiatives. Furthermore, apart from guaranteeing a heightened level of individualized attention, the participants perceived the e-enabled exercise intervention to be more tailored to their rehabilitation needs, thus exceeding apprehensions regarding comfort. The participants articulated that the real-time monitoring effectively captured the motion of the joints, thereby augmenting the efficacy of the measurements. Nevertheless, further investigation is necessary to determine the dependability and accuracy of the sensor-based measurement medium in comparison to the traditional measurement method.

Limitations

This study has several limitations that warrant discussion. First, the small sample size and the recruitment of participants from communities in close proximity may have led to a concentration of the population distribution, thereby limiting the generalizability of the findings. Second, the pre-experiment revealed that older individuals with insufficient digital health literacy were unable to independently configure and operate the experimental devices. As a result, the experimental protocol was adjusted to involve experimenters facilitating tri-weekly centralized sessions in collaboration with multiple community centers.

While the experimenters' involvement was limited to device debugging and their interaction with participants was minimal, their presence may have inadvertently influenced the experimental group's motivation and engagement with the intervention.

The broad age range of 60-80 years is another limitation, as physical capabilities vary significantly within this demographic, potentially introducing variability in the intervention's outcomes. Future studies should consider stratifying participants into narrower age ranges to enhance sample homogeneity and provide more precise insights into the intervention's effects across age subgroups.

In addition, the intervention duration was relatively short at 6 weeks, which may not have been sufficient to capture the long-term impact of the application. Future research should extend the study duration, further refine the application design to better accommodate older adults' needs, and explore tailored strategies to improve digital health literacy. These adjustments would enhance the application's utility and scalability in diverse community settings, enabling broader adoption and more robust evaluations of its effectiveness.

To address these limitations and enhance future research outcomes, it is recommended to extend the experiment duration, refine the application design to better accommodate older adults' needs, and explore strategies to improve digital health literacy among the target population. This would enable a more comprehensive evaluation of the application's capabilities and its potential impact on long-term rehabilitation outcomes.

Conclusions

The findings of this study demonstrated that the application based on CV technology effectively improved the physical functioning and self-efficacy of participants compared to conventional interventions. This suggests that the application holds promise for replication and implementation within community environments for patients with KOA.

A key novelty of this research is its validation, through a randomized controlled trial, of an application designed according to behavioral change theory principles. The study confirmed that such an application can partially substitute the guidance of a rehabilitation therapist, thereby enhancing exercise rehabilitation outcomes for participants.

In addition, the potential scalability and adaptability of the application in diverse settings warrant further exploration. Its implementation could significantly support older adults in community health management, particularly in addressing the varying levels of digital health literacy among the population. Future research should focus on extending the application's reach to broader demographics, optimizing its design to better accommodate older adults, and tailoring strategies to improve engagement and usability. This will maximize the application's practical impact and facilitate its adoption in real-world scenarios.

XSL•FO
RenderX

Acknowledgments

We thank the members of our steering committee for their oversight and expert guidance. We also thank the stakeholders who contributed to our project management group and people with lived experience of receiving or providing exercise therapy for KOA patients who took part in the stakeholder workshops and online survey for their useful feedback. The study was funded by National Key R&D Program of China (Grant No. 2022YFB3303303) Fundamental Research Funds for the Central Universities (Grant No. YG2023ZD10) and Lushan Lab Research Funding. We would also like to extend our special thanks to Tian Xia, Bochen Cao, and Wenhui Zhang for their contributions to the product design.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

DZ contributed to writing–original draft, conceptualization, data curation, and methodology. JZ was responsible was conceptualization, data curation, software, and methodology. TH handled writing–review and editing, project administration, and funding acquisition. TW conducted formal analysis and visualization. MW was involved in software and validation. BZ performed investigation and data curation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the Motion Tracking and Evaluation Algorithm in the CV Application.
[DOCX File, 1653 KB - mhealth_v13i1e63022_app1.docx]

Checklist 1
CONSORT checklist.
[PDF File, 1235 KB - mhealth_v13i1e63022_app2.pdf]

References

1. Uivaraseanu B, Vesa CM, Tit DM, et al. Therapeutic approaches in the management of knee osteoarthritis (Review). Exp Ther Med 2022 May;23(5):328. [doi: 10.3892/etm.2022.11257] [Medline: 35386619]
2. Jormand H, Mohammadi N, Khani Jeihooni A, et al. Self-care behaviors in older adults suffering from knee osteoarthritis: Application of theory of planned behavior. Front Public Health 2022;10:958614. [doi: 10.3389/fpubh.2022.958614] [Medline: 36408046]
3. Arden NK, Perry TA, Bannuru RR, et al. Non-surgical management of knee osteoarthritis: comparison of ESCEO and OARSI 2019 guidelines. Nat Rev Rheumatol 2021 Jan;17(1):59-66. [doi: 10.1038/s41584-020-00523-9] [Medline: 33116279]
4. Sharma L. Osteoarthritis of the knee. N Engl J Med 2021 Jan 7;384(1):51-59. [doi: 10.1056/NEJMcp1903768] [Medline: 33406330]
5. Kraus VB, Sprow K, Powell KE, et al. Effects of physical activity in knee and hip osteoarthritis: a systematic umbrella review. Med Sci Sports Exerc 2019 Jun;51(6):1324-1339. [doi: 10.1249/MSS.0000000000001944] [Medline: 31095089]
6. Kampmeijer R, Pavlova M, Tambor M, et al. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 5;16 Suppl 5(Suppl 5):290. [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
7. Wilson R, Chua J, Briggs AM, et al. The cost-effectiveness of recommended adjunctive interventions for knee osteoarthritis: results from a computer simulation model. Osteoarthr Cartil Open 2020 Dec;2(4):100123. [doi: 10.1016/j.ocarto.2020.100123] [Medline: 36474885]
8. Kostic AM, Leifer VP, Gong Y, et al. Cost-effectiveness of surgical weight-loss interventions for patients with knee osteoarthritis and class III obesity. Arthritis Care Res (Hoboken) 2023 Mar;75(3):491-500. [doi: 10.1002/acr.24967] [Medline: 35657632]
9. Bennell KL, Nelligan R, Dobson F, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 4;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
10. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 5;22(6):e15617. [doi: 10.2196/15617] [Medline: 32501277]

XSL•FO
RenderX

11. Shah N, Costello K, Mehta A, et al. Applications of digital health technologies in knee osteoarthritis: narrative review. JMIR Rehabil Assist Technol 2022 Jun 8;9(2):e33489. [doi: 10.2196/33489] [Medline: 35675102]
12. Kloek CJJ, Bossen D, Spreeuwenberg PM, et al. Effectiveness of a blended physical therapist intervention in people with hip osteoarthritis, knee osteoarthritis, or both: a cluster-randomized controlled trial. Phys Ther 2018 Jul 1;98(7):560-570. [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
13. Li LC, Feehan LM, Xie H, et al. Effects of a 12-week multifaceted wearable-based program for people with knee osteoarthritis: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 3;8(7):e19116. [doi: 10.2196/19116] [Medline: 32618578]
14. Dobson F, Hinman RS, French S, et al. Internet-mediated physiotherapy and pain coping skills training for people with persistent knee pain (IMPACT - knee pain): a randomised controlled trial protocol. BMC Musculoskelet Disord 2014 Aug 13;15(1):279. [doi: 10.1186/1471-2474-15-279] [Medline: 25125068]
15. Bouma AJ, van Wilgen P, Dijkstra A. The barrier-belief approach in the counseling of physical activity. Patient Educ Couns 2015 Feb;98(2):129-136. [doi: 10.1016/j.pec.2014.10.003] [Medline: 25455791]
16. Bennell KL, Campbell PK, Egerton T, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017 Jan;69(1):84-94. [doi: 10.1002/acr.22915] [Medline: 27111441]
17. Bailey JF, Agarwal V, Zheng P, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250. [doi: 10.2196/18250] [Medline: 32208358]
18. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
19. Zhu D, Zhao J, Wang M, et al. Rehabilitation applications based on behavioral therapy for people with knee osteoarthritis: systematic review. JMIR Mhealth Uhealth 2024 May 2;12:e53798. [doi: 10.2196/53798] [Medline: 38696250]
20. Bowden JL, Hunter DJ, Deveza LA, et al. Core and adjunctive interventions for osteoarthritis: efficacy and models for implementation. Nat Rev Rheumatol 2020 Aug;16(8):434-447. [doi: 10.1038/s41584-020-0447-8] [Medline: 32661322]
21. Bossen D, Veenhof C, Dekker J, et al. The usability and preliminary effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis. BMC Med Inform Decis Mak 2013 May 28;13:61. [doi: 10.1186/1472-6947-13-61] [Medline: 23714120]
22. Cui X, Zhao Z, Ma C, et al. A gait character analyzing system for osteoarthritis pre-diagnosis using RGB-D camera and supervised classifier. In: World Congress on Medical Physics and Biomedical Engineering 2018: Springer; 2018:297-301 URL: https://link.springer.com/chapter/10.1007/978-981-10-9035-6_53#citeas [accessed 2025-04-16] [doi: 10.1007/978-981-10-9035-6_53]
23. Bolink S, van Laarhoven SN, Lipperts M, et al. Inertial sensor motion analysis of gait, sit-stand transfers and step-up transfers: differentiating knee patients from healthy controls. Physiol Meas 2012 Nov;33(11):1947-1958. [doi: 10.1088/0967-3334/33/11/1947] [Medline: 23110821]
24. Fitzgibbon ML, Tussing-Humphreys L, Schiffer L, et al. Fit and Strong! Plus: Twelve and eighteen month follow-up results for a comparative effectiveness trial among overweight/obese older adults with osteoarthritis. Prev Med 2020 Dec;141:106267. [doi: 10.1016/j.ypmed.2020.106267] [Medline: 33022324]
25. Bennell K, Nelligan RK, Schwartz S, et al. Behavior change text messages for home exercise adherence in knee osteoarthritis: randomized trial. J Med Internet Res 2020 Sep 28;22(9):e21749. [doi: 10.2196/21749] [Medline: 32985994]
26. Debnath B, O'Brien M, Yamaguchi M, et al. A review of computer vision-based approaches for physical rehabilitation and assessment. Multimed Syst 2022 Feb;28(1):209-239. [doi: 10.1007/s00530-021-00815-4]
27. Eswaran U, Khang A. Artificial intelligence (AI)-aided computer vision (CV) in healthcare system. In: Computer Vision and AI-Integrated IoT Technologies in the Medical Ecosystem: CRC Press; 2024:125-137. [doi: 10.1201/9781003429609-8]
28. Sanford S, Liu M, Selvaggi T, et al. Effects of visual feedback complexity on the performance of a movement task for rehabilitation. J Mot Behav 2021;53(2):243-257. [doi: 10.1080/00222895.2020.1770670] [Medline: 32496974]
29. Bing F, Wang Y, Chen SF, et al. Effects of cycling rehabilitation training on patients with knee osteoarthritis: a systematic review and meta-analysis. Research Square. Preprint posted online on Jan 6, 2023. [doi: 10.21203/rs.3.rs-2425616/v1]
30. Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008 Feb 19;148(4):295-309. [doi: 10.7326/0003-4819-148-4-200802190-00008] [Medline: 18283207]
31. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
32. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988 Dec;15(12):1833-1840. [Medline: 3068365]
33. Brady TJ. Measures of self‐efficacy: arthritis self‐efficacy scale (ASES), arthritis self‐efficacy scale‐8 item (ASES‐8), children's arthritis self‐efficacy scale (CASE), chronic disease self‐efficacy scale (CDSES), parent's arthritis self‐efficacy

XSL•FO
RenderX

scale (PASE), and rheumatoid arthritis self‐efficacy scale (RASE). Arthritis Care Res (Hoboken) 2011 Nov;63(S11):S473-S485. [doi: 10.1002/acr.20567]
34. Allen J, Inder KJ, Lewin TJ, et al. Construct validity of the assessment of quality of life - 6D (AQoL-6D) in community samples. Health Qual Life Outcomes 2013;11:1. [doi: 10.1186/1477-7525-11-61] [Medline: 3639231]
35. Kurlowicz L, Greenberg SA, Kurlowicz L, Greenberg SA. The geriatric depression scale (GDS). Am J Nurs 2007;107(10):67-68. [doi: 10.1097/01.NAJ.0000292207.37066.2f]
36. Li T, Ma J, Zhao T, et al. Application and efficacy of extracorporeal shockwave treatment for knee osteoarthritis: a systematic review and meta-analysis. Exp Ther Med 2019 Oct;18(4):2843-2850. [doi: 10.3892/etm.2019.7897] [Medline: 31555375]
37. Schrepp M, Hinderks A, Thomaschewski J. Construction of a benchmark for the user experience questionnaire (UEQ). IJIMAI 2017;4(4):40. [doi: 10.9781/ijimai.2017.445]
38. Fernandes L, Storheim K, Sandvik L, et al. Efficacy of patient education and supervised exercise vs patient education alone in patients with hip osteoarthritis: a single blind randomized clinical trial. Osteoarthr Cartil 2010 Oct;18(10):1237-1243. [doi: 10.1016/j.joca.2010.05.015]
39. Nelligan RK, Hinman RS, Kasza J, et al. Effects of a self-directed web-based strengthening exercise and physical activity program supported by automated text messages for people with knee osteoarthritis: a randomized clinical trial. JAMA Intern Med 2021 Jun 1;181(6):776-785. [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
40. Seiferth C, Vogel L, Aas B, et al. How to e-mental health: a guideline for researchers and practitioners using digital technology in the context of mental health. Nat Mental Health 2023;1(8):542-554. [doi: 10.1038/s44220-023-00085-1]
41. Tore NG, Oskay D, Haznedaroglu S. The quality of physiotherapy and rehabilitation program and the effect of telerehabilitation on patients with knee osteoarthritis. Clin Rheumatol 2023 Mar;42(3):903-915. [doi: 10.1007/s10067-022-06417-3] [Medline: 36279075]
42. Tack C. A model of integrated remote monitoring and behaviour change for osteoarthritis. BMC Musculoskelet Disord 2021 Aug 9;22(1):669. [doi: 10.1186/s12891-021-04555-4] [Medline: 34372822]
43. Jönsson T, Dell'Isola A, Lohmander LS, et al. Comparison of face-to-face vs digital delivery of an osteoarthritis treatment program for hip or knee osteoarthritis. JAMA Netw Open 2022 Nov 1;5(11):e2240126. [doi: 10.1001/jamanetworkopen.2022.40126] [Medline: 36326763]
44. Cenamor J. Use of health self-management platform features: the case of a specialist ehealth app. Technol Forecast Soc Change 2022 Dec;185:122066. [doi: 10.1016/j.techfore.2022.122066]
45. Batrakoulis A, Jamurtas AZ, Metsios GS, et al. Comparative efficacy of 5 exercise types on cardiometabolic health in overweight and obese adults: a systematic review and network meta-analysis of 81 randomized controlled trials. Circ Cardiovasc Qual Outcomes 2022 Jun;15(6):e008243. [doi: 10.1161/CIRCOUTCOMES.121.008243] [Medline: 35477256]
46. Solis-Navarro L, Gismero A, Fernández-Jané C, et al. Effectiveness of home-based exercise delivered by digital health in older adults: a systematic review and meta-analysis. Age Ageing 2022 Nov 2;51(11):afac243. [doi: 10.1093/ageing/afac243] [Medline: 36346736]
47. Assar S, Gandomi F, Mozafari M, et al. The effect of Total resistance exercise vs. aquatic training on self-reported knee instability, pain, and stiffness in women with knee osteoarthritis: a randomized controlled trial. BMC Sports Sci Med Rehabil 2020;12(1):27. [doi: 10.1186/s13102-020-00175-y] [Medline: 32368344]
48. Chen H, Zheng X, Huang H, et al. The effects of a home-based exercise intervention on elderly patients with knee osteoarthritis: a quasi-experimental study. BMC Musculoskelet Disord 2019 Apr 9;20(1):160. [doi: 10.1186/s12891-019-2521-4] [Medline: 30967131]
49. Seppen BF, den Boer P, Wiegel J, et al. Asynchronous mHealth interventions in rheumatoid arthritis: systematic scoping review. JMIR Mhealth Uhealth 2020 Nov 5;8(11):e19260. [doi: 10.2196/19260] [Medline: 33151161]
50. Baumeister H, et al. Persuasive E-health design for behavior change. In: Montag C, Baumeister H, editors. Digital Phenotyping and Mobile Sensing, New Developments in Psychoinformatics: Springer; 2023:347-364. [doi: 10.1007/978-3-030-98546-2_20]
51. Olsson CB, Ekelund J, Degerstedt Å, et al. Change in self-efficacy after participation in a supported self-management program for osteoarthritis - an observational study of 11 906 patients. Disabil Rehabil 2020 Jul;42(15):2133-2140. [doi: 10.1080/09638288.2018.1555616] [Medline: 30686131]
52. Durst J, Roesel I, Sudeck G, et al. Effectiveness of human versus computer-based instructions for exercise on physical activity-related health competence in patients with hip osteoarthritis: randomized noninferiority crossover trial. J Med Internet Res 2020 Sep 28;22(9):e18233. [doi: 10.2196/18233] [Medline: 32985991]
53. Xie SH, Wang Q, Wang LQ, et al. Effect of internet-based rehabilitation programs on improvement of pain and physical function in patients with knee osteoarthritis: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Jan 5;23(1):e21542. [doi: 10.2196/21542] [Medline: 33399542]
54. Valentijn PP, Tymchenko L, Jacobson T, et al. Digital health interventions for musculoskeletal pain conditions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Sep 6;24(9):e37869. [doi: 10.2196/37869] [Medline: 36066943]

XSL•FO
RenderX

Abbreviations

ASES: Arthritis Self-Efficacy Scale
CONSORT: Consolidated Standards of Reporting Trials
CV: computer vision
GDS: Geriatric Depression Scale
KOA: knee osteoarthritis
NRS: numeric rating scale
ROM: Range of motion
UEQ: User Experience Questionnaire
WOMAC: Western Ontario and McMaster Universities Arthritis Index

Edited by L Buis; submitted 08.06.24; peer-reviewed by MA Shaphe, R Mansoorizade; revised version received 19.12.24; accepted 23.02.25; published 12.05.25.

Please cite as:
Zhu D, Zhao J, Wu T, Zhu B, Wang M, Han T
Effects of a Computer Vision–Based Exercise Application for People With Knee Osteoarthritis: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e63022
URL: https://mhealth.jmir.org/2025/1/e63022
doi:10.2196/63022

© Dian Zhu, Jianan Zhao, Tong Wu, Beiyao Zhu, Mingxuan Wang, Ting Han. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Wearable Devices in Remote Cardiac Rehabilitation With and Without Weekly Online Coaching for Patients With Coronary Artery Disease: Randomized Controlled Trial

Ryota Nishio[1], PhD; Tomotaka Dohi[1], PhD; Miho Yokoyama[1], PhD; Taisuke Nakade[1], MD; Norihito Takahashi[1], PhD; Yuichi Chikata[1], PhD; Hirohisa Endo[2], PhD; Hiroki Nishiyama[1], PhD; Iwao Okai[1], PhD; Hiroshi Iwata[1], PhD; Shinya Okazaki[1], PhD; Katsumi Miyauchi[1], PhD; Hiroyuki Daida[1], PhD; Tohru Minamino[1,3], PhD

[1]Department of Cardiovascular Biology and Medicine, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo, Japan

[2]Department of Cardiovascular Medicine, Juntendo University Shizuoka Hospital, Izunokuni, Shizuoka, Japan

[3]Japan Agency for Medical Research and Development, Japan Agency for Medical Research and Development-Core Research for Evolutionary Medical Science and Technology (AMED-CREST), Tokyo, Japan

Corresponding Author:
Tomotaka Dohi, PhD
Department of Cardiovascular Biology and Medicine, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

Background: Cardiac rehabilitation (CR) is effective in preventing cardiovascular diseases; however, participation in CR programs remains limited due to the associated challenges. The integration of wearable devices and real-time monitoring offers a potential solution to enhance adherence to remote CR programs and their outcomes.

Objective: This study aimed to evaluate the efficacy of a remote CR program using wearable devices and real-time monitoring with or without online coaching (OLC) in improving exercise capacity in patients with coronary artery disease (CAD).

Methods: We enrolled 50 patients with CAD in a remote CR program in this randomized, open-label, single-center pilot trial (phase III). After baseline cardiopulmonary exercise tests (CPETs), all patients were assigned a CPET-based home exercise program and were provided with a wearable device (Fitbit Sense; Fitbit, Inc) with a real-time monitoring system (Recoval; SapplyM, Inc). The patients were randomly assigned to an intervention group with OLC (n=25) or a control wearable device (CON; n=25) group. The primary outcomes were changes in peak oxygen consumption (peak VO_2) and anaerobic threshold VO_2 (oxygen consumption) at 12 weeks. The secondary outcomes were changes in CPET parameters, daily activity, anxiety levels, and health-related quality of life.

Results: Peak VO_2 and anaerobic threshold VO_2 increased significantly from baseline to 12 weeks in the OLC (+1.6 mL/kg/min, *P*<.001; +1.0 mL/kg/min, *P*=.001) and CON (+0.6 mL/kg/min, *P*=.008; +1.3 mL/kg/min, *P*=.002) groups with no significant between-group differences (*P*=.65 and *P*=.90). In the latter half of the intervention, the daily distance and highly active time in the OLC group were significantly increased compared with those in the CON group (all *P*<.05). Mental status and health-related quality of life scores showed no significant differences between the groups. No major adverse cardiac events were reported. One patient in the OLC group dropped out due to lower limb muscle strain.

Conclusions: Remote CR using wearable devices and a real-time monitoring system significantly improved exercise capacity in patients with CAD over a 12-week intervention program. The addition of regular OLC to the intervention program further enhanced the physical activity levels, particularly in high-intensity activities. Future studies are warranted to validate these findings in more diverse populations and over longer intervention periods to optimize the benefits and safety of remote CR programs.

Trial Registration: UMIN Clinical Trial Registry UMIN000047789; https://center6.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000054483

(JMIR Mhealth Uhealth 2025;13:e63797) doi:10.2196/63797

KEYWORDS

cardiac rehabilitation; cardiopulmonary exercise test; coronary artery disease; wearable device; wearable; cardiopulmonary; artery; rehabilitation; exercise; randomized trial; online coaching; cardiovascular; exercise program; home exercise program; monitoring system; monitor; anxiety; quality of life; VO2; anaerobic; coaching; mobile phone; oxygen consumption

XSL•FO
RenderX

Introduction

Cardiac rehabilitation (CR) is beneficial not only in the aftermath of acute myocardial infarction but also in patients with coronary artery disease (CAD) after percutaneous coronary intervention for prior myocardial infarction, unstable angina, stable angina, and coronary artery bypass graft surgery [1-3]. Despite its numerous benefits, CR use remains insufficient, with various contributing factors such as limited access to rehabilitation facilities, financial constraints, busy schedules, and insufficient patient education [4,5]. A meta-analysis evaluating interventions to enhance adherence found that unmonitored CR and reduced copayments significantly improved adherence [6]. Recently, several studies have reported a negative correlation between cardiovascular disease risk and physical activity measured using wearable devices and a smartphone [7-9]. Furthermore, smartphone-based home CR programs have been shown to enhance participation and adherence and improve both physiological and psychological health outcomes [10].

Recent meta-analyses reported that wearable device-based interventions effectively increase daily activity levels and physical capacity in cardiovascular patients, especially when paired with feedback mechanisms [11]. However, there remains considerable heterogeneity with regard to the methods, frequency, and quality of feedback used in these wearable device-based interventions. For instance, while some studies have used weekly feedback sessions, others have implemented daily or real-time feedback, reflecting variability in intervention designs [12-15]. Additionally, the specific content and personalization of feedback have been underexplored, further highlighting the need for a standardized framework. Therefore, to address the limitations and gaps in prior research, this study aimed to provide insights into the most effective strategies for wearable device-based CR interventions. In this study, we developed a system for real-time monitoring of patient activity levels by integrating wearable devices and smartphones (Recoval; SapplyM, Inc). This system combines real-time data collection on physical activity and biometric metrics with online coaching (OLC) to deliver personalized feedback and support to the users. The primary objectives of this pilot study were to evaluate the efficacy of this system in improving exercise capacity and adherence through continuous and interactive communication between health care providers and patients and to clarify the synergistic effects of OLC on the use of wearable devices. We hypothesized that the combination of wearable devices and OLC would result in greater improvements in exercise capacity and adherence than those achieved through the use of wearable devices alone.

Methods

Eligibility Criteria

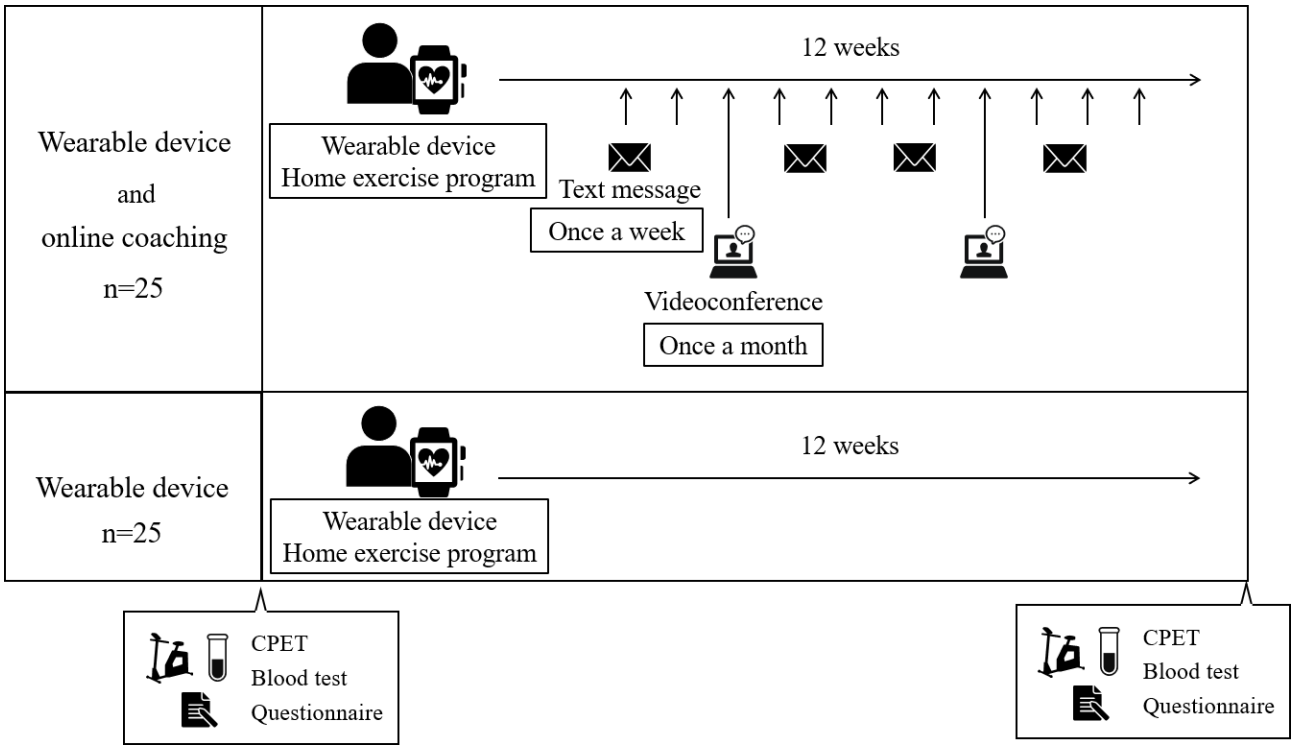
We enrolled 50 patients with CAD who were eligible for outpatient CR (phase III) for the secondary prevention of CAD at Juntendo University Hospital between April 30, 2022, and January 21, 2023. The exclusion criteria were as follows: (1) implantable medical device users such as patients with cardiac pacemakers and defibrillators; (2) patients diagnosed with acute myocardial infarction, uncontrolled angina pectoris not stabilized by medical treatment, uncontrolled arrhythmias causing subjective symptoms or hemodynamic abnormalities, uncontrolled heart failure, symptomatic severe aortic stenosis, or a psychiatric disorder impeding communication; and (3) participants deemed unsuitable for this study by the principal investigator and research physician.

Study Design

This was a randomized, open-label, comparative, parallel-group, single-center interventional study. This study followed the CONSORT (Consolidated Standards of Reporting Trials) reporting guidelines. In this study, patients were randomly divided into a wearable device plus OLC group and a control wearable device (CON) group, stratified by age (<60 years and ≥60 years) and sex (male and female), using a stratified permuted block design as the allocation method. The smartphones used by the patients were personal devices, and the wearable device was a Fitbit Sense. Patient data were collected using a wearable device linked to a smartphone for real-time monitoring of step count, activity level, and heart rate (Recoval; for detailed information about the product, please see [16]). A proprietary algorithm converted the acceleration signals from a wearable device into step counts and metabolic-equivalent tasks (METs). The intervention lasted for 12 weeks, and all patients underwent a cardiopulmonary exercise test (CPET)-based home exercise program. This program aims to engage in exercise at anaerobic threshold (AT) heart rate and a Borg scale of 12 - 13, for at least 30 minutes per day, 5 days a week. Additionally, the patients in the OLC group were provided with exercise guidance based on their activity levels recorded by their wearable device. Guidance was provided via weekly text messages and monthly videoconferences. The participants in the CON group wore only the wearable device and did not receive any additional instructions (Figure 1).

XSL•FO
RenderX

Figure 1. Timeline of this study. The patients were divided into the wearable device and online coaching group and the wearable device group. The intervention lasted for 12 weeks, and during this time, all patients were given a home exercise program. The wearable device and online coaching group received exercise guidance based on their activity levels recorded by their wearable device. This guidance was given through weekly text messages and monthly videoconferences. The wearable device group only wore the wearable device and did not receive any additional instruction. CPET: cardiopulmonary exercise test.



Ethical Considerations

The Ethics Committee of the Juntendo Clinical Research and Trial Centre approved this study (IRB number E21-0353). Patients were enrolled by outpatient primary care physicians who were not involved in the analysis, and external clinical research coordinators allocated them into the 2 groups. All patients provided written informed consent to participate in this study and had the option to opt out of this study at their discretion. The collected data were deidentified and managed accordingly. Additionally, no compensation was provided to study participants.

Data Collection

Overview

Blood samples were collected and blood pressure (BP) was measured for all the participants at the beginning and end of this study. Patients with BP >140/90 mm Hg or those taking antihypertensive medications were considered hypertensive. Dyslipidemia was defined based on the values of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglycerides, (≥140, ≤40, and ≥150 mg/dL, respectively) or use of statins or lipid-lowering agents [17]. Diabetes mellitus was defined as either a hemoglobin A_{1c} level of ≥6.5% or the use of medication with insulin or oral hypoglycemic drugs. Chronic kidney disease was defined as an estimated glomerular filtration rate of <60 mL/min/1.73 m^2, calculated using the Modification of Diet in Renal Disease equation modified with a Japanese coefficient using the follow-up serum creatinine level [18].

Measurements

We assessed the anthropometric parameters of patients and exercise tolerance at the beginning and end of this study. Briefly, anthropometric parameters, including the percentage of body fat, lean body weight, and muscle mass, were measured using bioelectrical impedance analysis (MC-780A; Tanita). The bioelectrical impedance analysis measurements were conducted 2 - 3 hours after a meal and before CPET. Exercise capacity was assessed using CPET on a cycle ergometer (Strength Ergo 8, Mitsubishi Electric) with an expiratory gas analysis machine (AE-310S, Minato Medical Science Co, Ltd). After 4 minutes of rest in the sitting position, the patient warmed up for several minutes at 20 W, followed by ramp loading (10 W/min) until they felt exhausted or experienced progressive angina, ST-segment depression (≥2 mm), or sustained tachyarrhythmia [19]. A standard 12-lead electrocardiogram was continuously recorded, and heart rate and BP were recorded every minute during the exercise test. A satisfactory end point of CPET was a respiratory exchange ratio greater than 1.10 [20]. Peak oxygen consumption (peak VO_2) was defined as the highest oxygen consumption (VO_2) value recorded during CPET, and the AT point was determined by the V-slope method [21].

Anxiety levels were assessed using a self-administered State Trait Anxiety Inventory (STAI) form at baseline and the end of this study. This inventory consisted of 40 statements about the feelings of the participants and was divided into 2 parts. In part I (comprising 20 statements), patients were instructed to rate the intensity of their current feelings of anxiety (indicating state anxiety) on a scale ranging from 1 (absolutely not) to 4 (very much). In part II (the remaining 20 statements), patients reported

XSL•FO
RenderX

the frequency of their general symptoms of anxiety (indicating trait anxiety) on a scale ranging from 1 (hardly ever) to 4 (often). The total score for each part ranges from 20 to 80, with higher scores indicating higher levels of anxiety. The Japanese version of the STAI was used in this study. Health-related quality of life (HR-QOL) was assessed using the Japanese version of the 36-item short form health survey (SF-36) at baseline and at the end of our study [22,23]. SF-36 measures 8 health domains: physical function, physical role, body pain, general health, vitality, social function, emotional role, and mental health. Each domain was scored separately from 0 (indicating the lowest level of functioning) to 100 (the highest level). Activity level was categorized as sedentary (<1.5 METs), lightly active (1.5 - 3 METs), moderately active (3 - 6 METs), or highly active (>6 METs or ≥145 steps/min sustained for at least 10 min).

Recoval System

The Recoval system provides three specific functionalities: (1) visualization of activity level and vital signs during use, (2) personalized activity goal setting, and (3) message-based intervention. The system enables medical personnel to remotely monitor patients' activity and vital data, including heart rate, steps, and calories burned, through data acquired from supported wearable devices. This data can be visualized on both the medical personnel's and the patient's devices, allowing for continuous tracking and analysis. Medical personnel can set individualized exercise goals for each patient, such as target exercise duration, target days per week, target heart rate during exercise, maximum allowable heart rate, and target step count per day. Patients can view these personalized goals and track their progress. Furthermore, the system provides a weekly summary that shows the number of target days versus actual days achieved, as well as comparison of actual step counts to daily targets. The system also facilitates asynchronous communication between patients and medical personnel via a dedicated message screen, where medical personnel can provide tailored feedback, guidance, and motivational messages, while patients can report their conditions and concerns in real time. This communication fosters continuous interaction, enhancing patient engagement and adherence (Multimedia Appendix 1).

End Points

The primary outcomes for this study included changes in peak VO_2 and AT VO_2 at 12 weeks. The secondary outcomes for this study were changes in CPET parameters, daily activity, anxiety level, and HR-QOL.

Statistical Analysis

The sample size calculation was based on anticipated changes in peak VO_2 between the OLC and CON groups, as reportedly previously by comparable studies. We assumed a mean improvement in peak VO_2 of 2.0 mL/kg/min in the OLC group and 1.0 mL/kg/min in the CON group, estimating this as approximately 80% of the peak VO_2 increase observed in comparable studies [24]. To determine the sample size, we used a pooled SD of 1.1 mL/kg/min, a 2-sided significance level of 5%, and a power of 80%, which led to an initial requirement of 22 participants per group. To account for an anticipated 10% dropout rate, we adjusted the final sample size to 25 participants per group.

Categorical data are presented as numbers and percentages and were compared using the chi-square test. Continuous variables are expressed as mean (SD) or median (IQR). Between-group comparisons were performed using Student *t* test, while within-group comparisons were performed using a paired *t* test. A *P* value <.05 was considered statistically significant. Effect sizes for the group×time interaction were reported using Cohen *f*, interpreted as small (*f*≥0.10), medium (*f*≥0.25), and large (*f*≥0.40). Missing values were not imputed, and multiplicity was not considered. In this pilot trial, the outcome assessors were blinded to group allocations. All statistical analyses were performed using R (version 4.4.2; R Foundation).

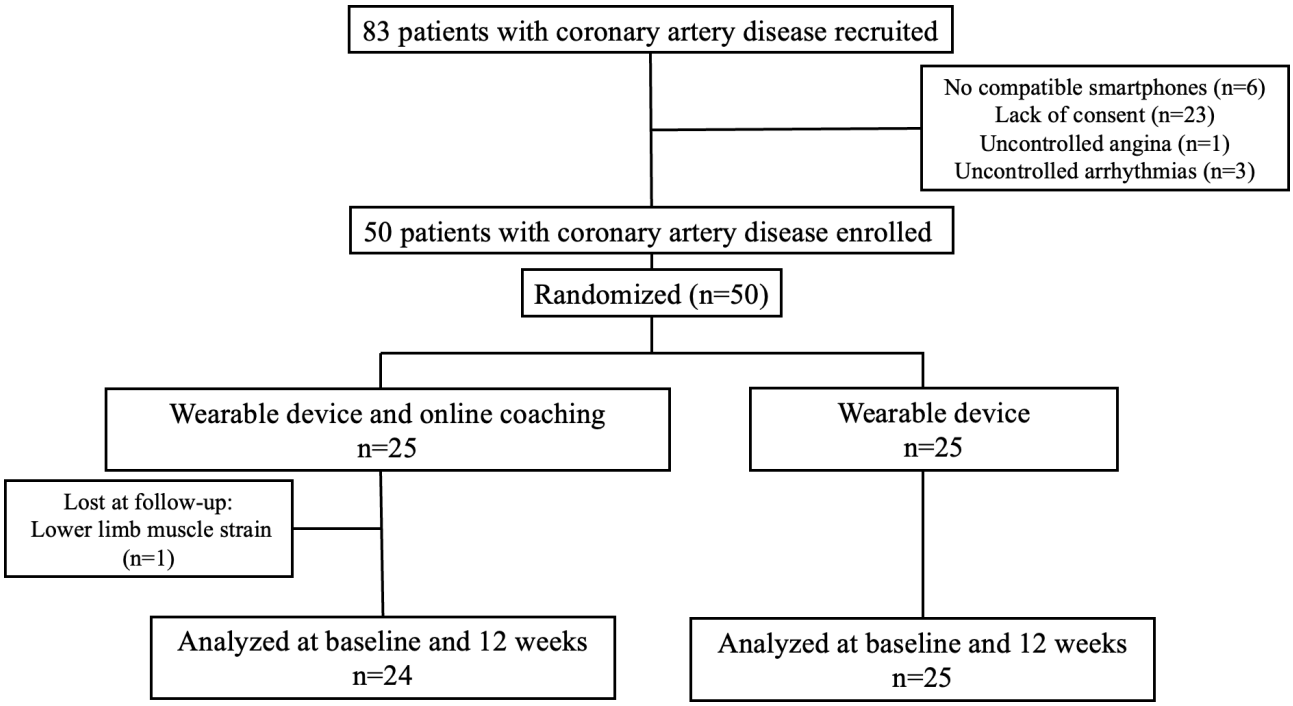
Results

Baseline Patient Clinical Characteristics and Blood Test Results

A total of 83 patients with CAD were enrolled in this study. Of these, 6 patients were excluded because they did not have compatible smartphones, 23 were excluded due to lack of consent, 1 was excluded due to uncontrolled angina, and 3 patients were excluded due to uncontrolled arrhythmias. As a result, 50 patients with CAD were finally analyzed and classified into 2 groups, OLC and CON, with randomization yielding 25 participants per group. One patient in the OLC group dropped out during the follow-up period because of lower limb muscle strain. No major adverse cardiac events occurred during this study's period; these were defined as a composite of cardiovascular death, nonfatal myocardial infarction, nonfatal stroke, and admission for heart failure. Finally, baseline data and data at 12 weeks following intervention from a total of 49 patients were included in the analysis (Figure 2). Table 1 presents the patient backgrounds. The mean age (63.8, SD 6.4 vs 62.6, SD 7.9 years) and sex distribution (n=23, 95.8% vs n=22, 92% male) between the OLC and CON groups were similar. No significant differences were observed between the 2 groups regarding baseline characteristics and coronary risk factors (all *P*>.05). However, the use of β-blockers was significantly higher in the OLC group (n=19, 76%) than in the CON group (n=11, 45.8%; *P*=.03). There were no significant differences in baseline blood data between the 2 groups.

XSL•FO
RenderX

Figure 2. Study flowchart. A total of 83 patients with CAD were initially recruited for this study. Of these, 33 patients were excluded due to various reasons, such as not having compatible smartphones (n=6), lack of consent (n=23), uncontrolled angina (n=1), and uncontrolled arrhythmias (n=3). Finally, 50 patients with CAD were enrolled and classified into 2 groups: the wearable device and online coaching group and the wearable device group. Randomization yielded 25 participants per group. One patient in the wearable device and online coaching group dropped out during the follow-up period. We finally analyzed baseline and 12-week data from a total of 49 patients.



XSL•FO
RenderX

Table . Baseline clinical characteristics of patients.

		Wearable device and online coaching (n=24)	Wearable device (n=25)	*P* value
Baseline characteristic				
	Age (years), mean (SD)	63.8 (6.4)	62.6 (7.9)	.57
	Sex (male), n (%)	23 (95.8)	22 (92)	.57
	BMI (kg/m^2), mean (SD)	26.2 (3.8)	25.6 (5.2)	.59
	Appendicular skeletal muscle mass (kg), mean (SD)	25 (4.3)	24.2 (4.1)	.51
	Body fat percentage (%), mean (SD)	23.7 (5.3)	23.1 (11.4)	.81
	Lean body mass (kg), mean (SD)	55.7 (7.5)	53.4 (6.7)	.26
	Acute coronary syndrome, n (%)	12 (50)	13 (52)	.57
	Post PCI[a], n (%)	21 (87.5)	21 (84)	.73
	Post CABG[b], n (%)	3 (12.5)	4 (16)	.23
	Hypertension, n (%)	21 (87.5)	3 (12.5)	.60
	Dyslipidemia, n (%)	23 (95.8)	22 (88)	.31
	Diabetes, n (%)	7 (29.2)	12 (48)	.17
	Chronic kidney disease, n (%)	13 (54.2)	18 (72)	.19
Medication				
	Aspirin, n (%)	16 (66.7)	17 (68)	.69
	P2Y12 receptor antagonist, n (%)	11 (45.8)	14 (56)	.48
	β-blocker, n (%)	11 (45.8)	19 (76)	.03
	Calcium channel blocker, n (%)	8 (33.3)	13 (52)	.10
	ACE-I/ARB[cd], n (%)	10 (41.7)	13 (52)	.31
	Statin, n (%)	22 (91.7)	19 (76)	.29
Baseline blood data				
	HbA_{1c}[e] (%), median (IQR)	6.7 (5.9 - 6.9)	6.2 (5.9 - 6.9)	.51
	Glucose (mg/dL), median (IQR)	100 (92 - 126)	106 (93 - 130)	.83
	Triglycerides (mg/dL), median (IQR)	118 (91 - 148)	98 (76 - 148)	.85
	HDL-C[f] (mg/dL), median (IQR)	46 (42 - 51)	51 (44 - 60)	.43
	LDL-C[g] (mg/dL), median (IQR)	69 (60 - 79)	62 (58 - 69)	.16
	NT-proBNP[h] (pg/mL), median (IQR)	57 (32 - 102)	131 (44 - 258)	.12
	High-sensitivity C-reactive protein (mg/mL), median (IQR)	0.05 (0.02 - 0.07)	0.07 (0.03 - 0.17)	.09
	Creatinine (mg/mL), median (IQR)	0.96 (0.86 - 1.03)	0.86 (0.76 - 0.91)	.87

XSL•FO
RenderX

	Wearable device and online coaching (n=24)	Wearable device (n=25)	*P* value
eGFR[i] (mL/min/1.73m^2), median (IQR)	61 (58 - 70)	70 (58 - 80)	.19

[a]PCI: percutaneous coronary intervention.

[b]CABG: coronary artery bypass grafting.

[c]ACE-I: angiotensin-converting enzyme inhibitors.

[d]ARB: angiotensin receptor blockers.

[e]HbA_{1c}: hemoglobin A_{1c}.

[f]HDL-C: high-density lipoprotein cholesterol.

[g]LDL-C: low-density lipoprotein cholesterol.

[h]NT-proBNP: N-terminal probrain natriuretic peptide.

[i]eGFR: estimated glomerular filtration rate.

Baseline CPET Parameters, Mental Status, and HR-QOL

A baseline comparison showed no significant differences between the 2 groups regarding CPET parameters, including peak VO_2 (21.2, SD 3.5 vs 20, SD 5.9; *P*=.39) and AT VO_2 (13.1, SD 2.3 vs 12.4, SD 3.9; *P*=.40). Similarly, no significant between-group differences were observed concerning mental status measured by STAI and in HR-QOL measured by SF-36 (Table S1 in Multimedia Appendix 2).

Postintervention Patients' Clinical Characteristics and Blood Data

After 12 weeks of intervention, both groups exhibited changes in their health metrics (Table 2). In the OLC group, there was a slight reduction in BMI (*P*=.003). Both groups exhibited a decrease in appendicular skeletal muscle mass and a slight increase in body fat percentage; however, these changes were not significant. In terms of blood data, high-density lipoprotein cholesterol levels significantly increased in both groups (*P*=.04).

XSL•FO
RenderX

Table . Comparison of patient clinical characteristics and blood data at baseline and after 12 weeks in wearable device and online coaching group and wearable device group.

	Wearable device and online coaching				Wearable device				Between-group changes *P* value	Cohen *f*
	Baseline	After 12 weeks	Within-group changes	Within-group *P* value	Baseline	After 12 weeks	Within-group changes	Within-group *P* value		
Patient characteristic										
BMI (kg/m^2), mean (SD)	26.2 (3.8)	25.9 (3.8)	–0.2 (–0.5 to 0)[a]	.003	25.6 (5.2)	25.5 (5.2)	–0.2 (–0.43 to –0.3)[a]	.003	.16	0.01
Appendicular skeletal muscle mass (kg), mean (SD)	25 (4.3)	24.6 (4.1)	–0.3 (0.2 to –1.1)[a]	.06	24.2 (4.1)	23.8 (3.8)	–0.2 (–1 to 0)[a]	.11	.87	0.004
Body fat percentage (%), mean (SD)	23.7 (5.3)	23.5 (5.7)	0.3 (–1.6 to 1.2)[a]	.58	23.1 (11.4)	23.7 (11.7)	0.7 (–1 to 2.3)[a]	.18	.17	0.02
Blood data										
HbA_{1c}[b] (%), median (IQR)	6.7 (5.9 to 6.9)	6.4 (6.1 to 6.9)	0 (–0.1 to 0)	.32	6.2 (5.9 to 6.9)	6.2 (5.9 to 6.8)	0 (–0.1 to 0.1)	.56	.29	0.05
Glucose (mg/dL), median (IQR)	100 (92 to 126)	111 (100 to 130)	7 (1 to 15)	.11	106 (93 to 130)	105 (96 to 132)	0 (–9 to 15)	.59	.27	0.05
Triglycerides (mg/dL), median (IQR)	118 (91 to 148)	103 (90 to 127)	–5 (–41 to 14)	.29	98 (76 to 148)	110 (80 to 152)	0 (–14 to 4)	.66	.67	0.03
HDL-C[c] (mg/dL), median (IQR)	46 (42 to 51)	50 (45 to 56)	2 (0 to 3)	.04	51 (44 to 60)	53 (45 to 60)	2 (0 to 7)	.04	.91	0.004
LDL-C[d] (mg/dL), median (IQR)	69 (60 to 79)	65 (60 to 78)	–2 (–10 to 6)	.35	62 (58 to 69)	65 (55 to 78)	0 (–2 to 10)	.17	.11	0.07
NT-proBNP[e] (pg/mL), median (IQR)	57 (32 to 102)	58 (25 to 86)	–1 (–39 to 11)	.46	131 (44 to 258)	71 (33 to 195)	–12 (–28 to 0)	.23	.23	0.05
High-sensitivity C-reactive protein (mg/mL), median (IQR)	0.05 (0.02 - 0.07)	0.06 (0.03 - 0.10)	0 (0 to 0.05)	.81	0.07 (0.03 to 0.17)	0.08 (0.04 to 0.14)	0 (–0.03 to 0.03)	.63	.71	0.04

XSL•FO
RenderX

	Wearable device and online coaching				Wearable device				Between-group changes *P* value	Cohen *f*
	Baseline	After 12 weeks	Within-group changes	Within-group *P* value	Baseline	After 12 weeks	Within-group changes	Within-group *P* value		
Creatinine (mg/mL), median (IQR)	0.96 (0.86 - 1.03)	0.95 (0.90 - 1.04)	0 (–0.1 to 0.06)	.45	0.86 (0.76 to 0.91)	0.94 (0.80 to 1.01)	0.1 (0 to 0.11)	.30	.97	0.001
eGFR[f] (mL/min/1.73m^2), median (IQR)	61 (58 - 70)	63 (56 - 68)	0 (–3 to 7)	.33	70 (58 to 80)	62 (57 to 80)	2 (–2 to 8)	.08	.58	0.02

[a]Median (IQR).

[b]HbA_{1c}: hemoglobin A_{1c}.

[c]HDL-C: high-density lipoprotein cholesterol.

[d]LDL-C: low-density lipoprotein cholesterol.

[e]NT-proBNP: N-terminal probrain natriuretic peptide.

[f]eGFR: estimated glomerular filtration rate

Postintervention CPET Parameters, Mental Status, and HR-QOL

Over a period of 12 weeks, changes in CPET parameters and mental status outcomes were observed in both groups (Table 3). In the OLC group, there was a significant increase in peak VO_2 with a change of 9 (SD 10.3%; *P*<.001), AT VO_2 with a change of 11.8 (SD 16%; *P*=.001), peak VO_2/HR with a median change of 0.5 (range −0.2 to 0.7, *P*=.04), and $\Delta VO_2/\Delta$load with a mean range 1 (SD 0.8 range 0.4 to 3, *P*=.001). Similarly, in the CON group, significant improvements were noted in peak VO_2 with a mean change of 8.1 (SD 14%; *P*=.008), AT VO_2 with a mean change of 12.9 (SD 19.3%; *P*=.002), and $\Delta VO_2/\Delta$load with a mean change of 2.0 (range 0.1 to 3.2, *P*=.001). The changes in peak VO_2, AT VO_2, peak VO_2/HR, and $\Delta VO_2/\Delta$load over 12 weeks did not show a significant difference between the 2 groups. Concerning mental status, the OLC group showed a significant increase in body pain scores. Conversely, the CON group did not show significant changes in body pain levels.

XSL•FO
RenderX

Table . Comparison of CPET[a] parameters and mental status outcomes at baseline and after 12 weeks in wearable device and online coaching group and wearable device group.

		Wearable device and online coaching				Wearable device				Between-group changes *P* value	Cohen *f*
		Baseline, mean (SD)	After 12 weeks, mean (SD)	Within-group changes, median (IQR)	Within-group *P* value	Baseline, mean (SD)	After 12 weeks, mean (SD)	Within-group changes, median (IQR)	Within-group *P* value		
CPET parameters											
Peak VO_2[b] (mL/kg/min)		21.2 (3.5)	23 (3.9)	9% (10.3%)[c]	<.001	20 (5.9)	21.5 (6.3)	8.1% (14%)[c]	.008	.80	0.02
AT[d] VO_2 (mL/kg/min)		13.1 (2.3)	14.4 (2.1)	11.8% (16%)[c]	.001	12.4 (3.9)	13.8 (3.3)	12.9% (19.3%)[c]	.002	.82	0.006
Resting HR[e] (beats/min)		69 (11.4)	74 (12.8)	5 (–0.3 to 9.3)	.002	70.9 (13.9)	69.5 (15)	–2 (–5 to 1)	.25	.001	0.12
Peak HR (beats/min)		140.3 (16.4)	146.3 (17.2)	5.5 (0 to 11.8)	.03	127.6 (27.4)	132.2 (27.5)	6 (0 to 9)	.07	.06	0.02
Peak VO_2/HR		11 (2)	11.4 (2)	0.5 (–0.2 to 0.7)	.04	11 (2.3)	11.4 (2.4)	0.4 (–0.6 to 1.1)	.12	.94	0.002
Ventilation versus VCO_2[f] slope		31.8 (4.5)	32.9 (3.9)	1.8 (–0.4 to 3)	.16	31.1 (4.5)	31.8 (5)	0.3 (–0.4 to 1.5)	.22	.70	0.02
Minimum ventilation/VCO_2		33.1 (4.4)	33.7 (4.5)	0.5 (0 to 1.2)	.09	33.1 (3.9)	33.3 (4)	0.1 (–0.8 to 1.3)	.75	.43	0.03
$\Delta VO_2/\Delta$load (mL/min/W)		9.2 (1.4)	9.2 (1.4)	1 (0.4 to 3)	.001	8.2 (2)	9.8 (1.6)	2 (0.1 to 3.2)	.001	.53	0.05
ΔHR/Δload ×100 (beats/W)		60.1 (15)	61.7 (17.3)	1.2 (–5.8 to 6)	.57	52.4 (19.2)	56.3 (20.6)	5.7 (–2.6 to 11)	.09	.38	0.03
Mental status											
STAI[g]											
	State anxiety	32.4 (7.1)	35.7 (7.8)	4 (0 to 10)	.05	35.1 (8.2)	36 (8.3)	2 (–4 to 6)	.61	.29	0.05
	Trait anxiety	38.8 (10.1)	38.5 (9.3)	1 (–4.3 to 3)	.82	38 (8.5)	39.3 (11.5)	0 (–1 to 6)	.30	.78	0.03
HR-QOL[h]											
SF-36[i]											
	Physical function	87.3 (12.9)	89.8 (9)	0 (0 to 5)	.26	83.2 (18.3)	85 (19)	0 (0 to 5)	.37	.81	0.01
	Physical role	85.4 (15.5)	88 (15.7)	0 (–12.5 to 18.7)	.46	84.3 (22.5)	84.5 (19.7)	0 (–12.5 to 6.3)	.95	.68	0.03
	Body pain	77.7 (21.7)	67.7 (21.1)	–1 (–21.3 to 0)	.02	66 (26.7)	71.6 (24.2)	0 (–11 to 16)	.33	.03	0.17
	General health	56.5 (21.6)	57.5 (17.5)	0 (–5.8 to 10.5)	.73	54.7 (15.8)	53.8 (19.2)	0 (–5 to 5)	.76	.64	0.03
	Vitality	65.1 (15)	68.2 (14.3)	0 (–1.6 to 12.5)	.30	59.3 (12.2)	62.5 (16.6)	6.3 (–6.2 to 12.5)	.22	.97	0.002
	Social function	95.8 (8.8)	95.3 (8.1)	0 (0 to 0)	.77	89.5 (12.8)	90 (16.9)	0 (0 to 12.5)	.89	.80	0.02
	Emotional role	88.2 (14.3)	93.8 (10.8)	0 (0 to 25)	.13	79.3 (26.7)	87.3 (17.9)	0 (–8.3 to 8.3)	.10	.68	0.03

XSL•FO
RenderX

		Wearable device and online coaching				Wearable device				Between-group changes *P* value	Cohen *f*
		Baseline, mean (SD)	After 12 weeks, mean (SD)	Within-group changes, median (IQR)	Within-group *P* value	Baseline, mean (SD)	After 12 weeks, mean (SD)	Within-group changes, median (IQR)	Within-group *P* value		
	Mental health	79.6 (11.9)	78.1 (15.6)	0 (–5 to 5)	.65	74 (14.1)	75.2 (15.9)	0 (–5 to 5)	.64	.51	0.05

[a]CPET: cardiopulmonary exercise test.

[b]VO2: oxygen consumption.

[c]Mean (SD).

[d]AT: anaerobic threshold.

[e]HR: heart rate.

[f]VCO_2: carbon dioxide production.

[g]STAI: State-Trait Anxiety Inventory.

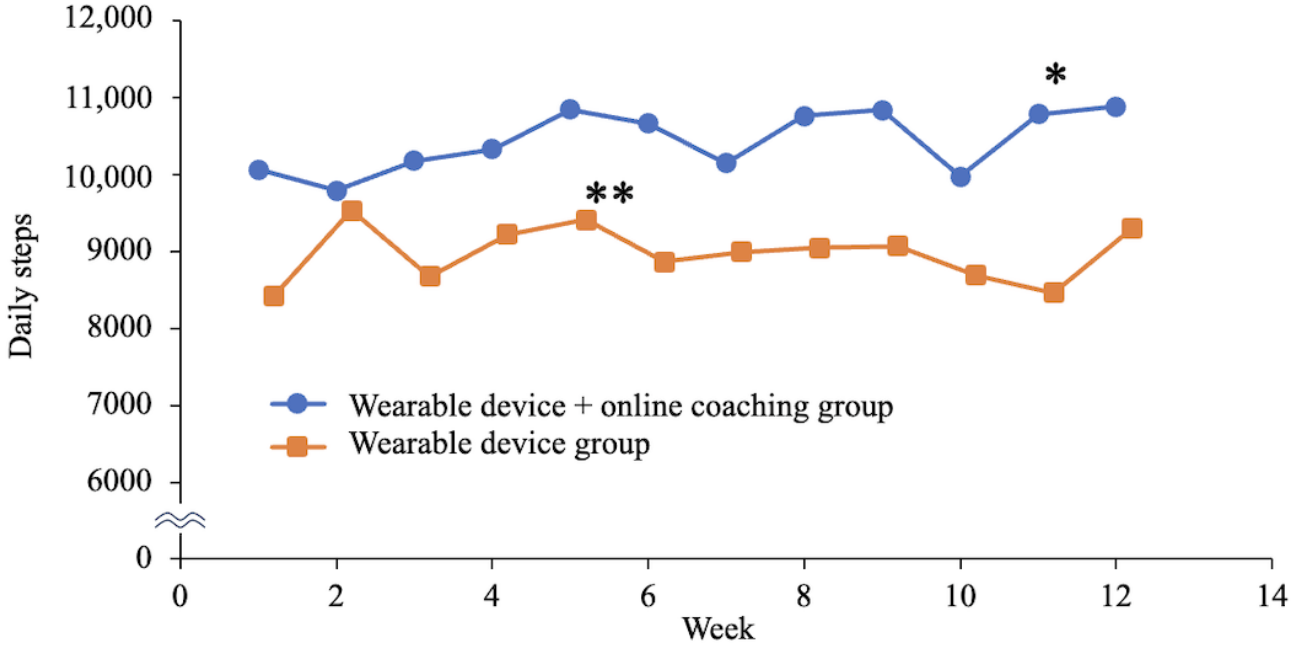
[h]HR-QOL: health-related quality of life.

[i]SF-36: 36-Item short form health survey.

Daily Activity

The parameters measured by the wearable device are shown in Figures 3-6. Figure 3 displays the average number of daily steps taken each week. Although the increase in steps in the OLC group was not significant, a notable increase in the number of steps was observed. After 11 weeks, there was a significant increase in the number of steps compared with that in the CON group (10,785 steps vs 8463 steps; *P*=.03).Figure 4 displays the average daily travel distance per week. The distance covered was significantly longer for OLC group, especially in the latter period, compared with that in the CON group (8 km vs 6.5 km; *P*=.003). Of activity level, the daily duration of high activity was significantly longer for the OLC group after 9 weeks than that for the CON group (37 min vs 21.7 min; *P*=.03). However, there was no significant difference in the duration of moderate activity between the 2 groups (Figures 5 and 6).

Figure 3. The average daily steps taken each week. * *P*<.05 Student *t* test (between-groups), ** *P*<.05 paired *t* test (within-group).



XSL•FO
RenderX

Figure 4. The average daily distance taken each week. * *P*<.05 Student *t* test (between-groups), ** *P*<.05 paired *t* test (within-group).

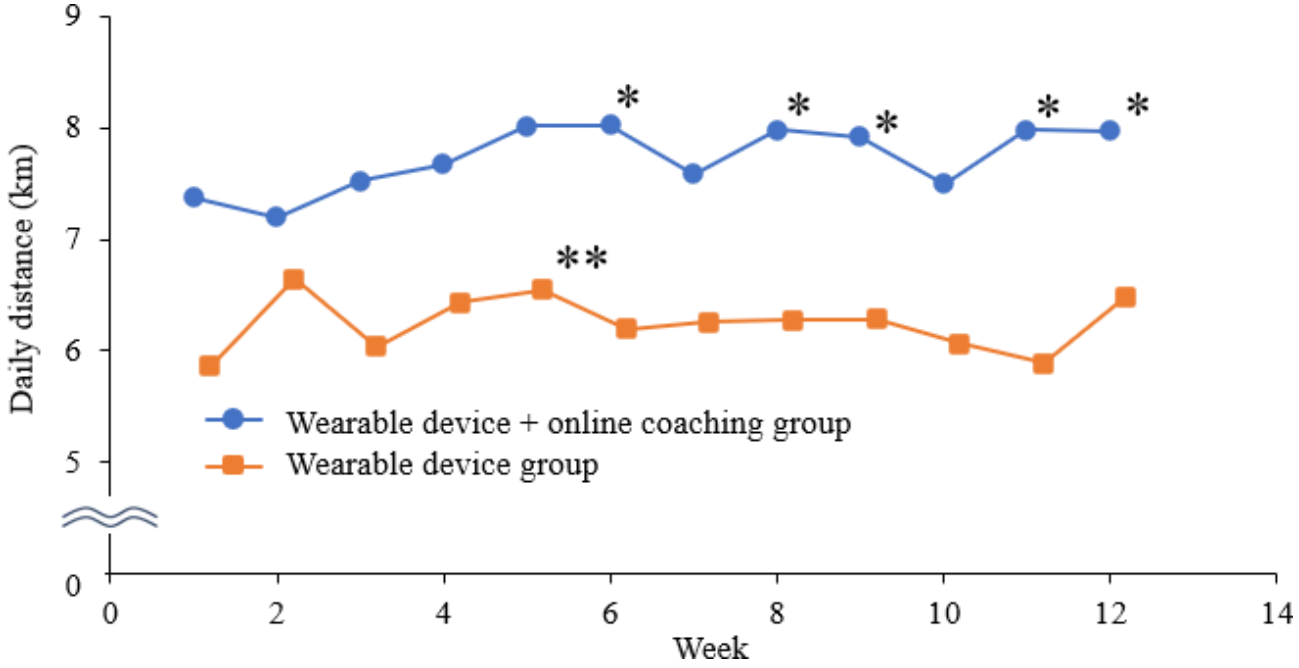
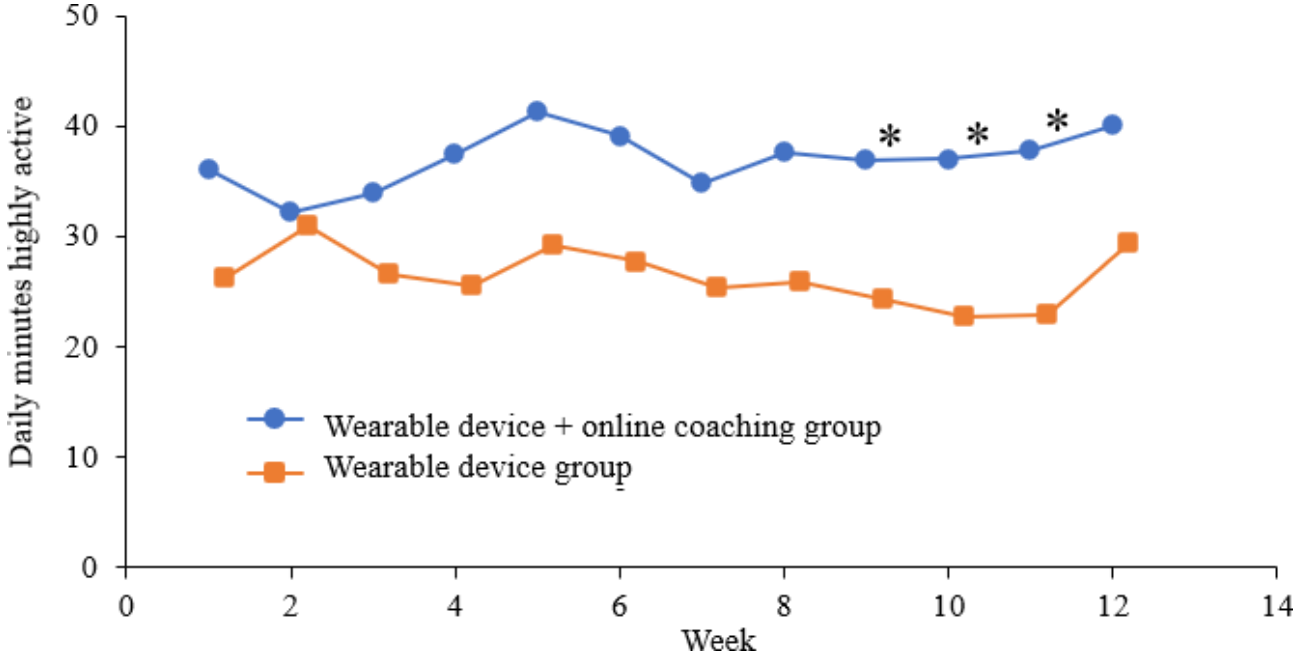
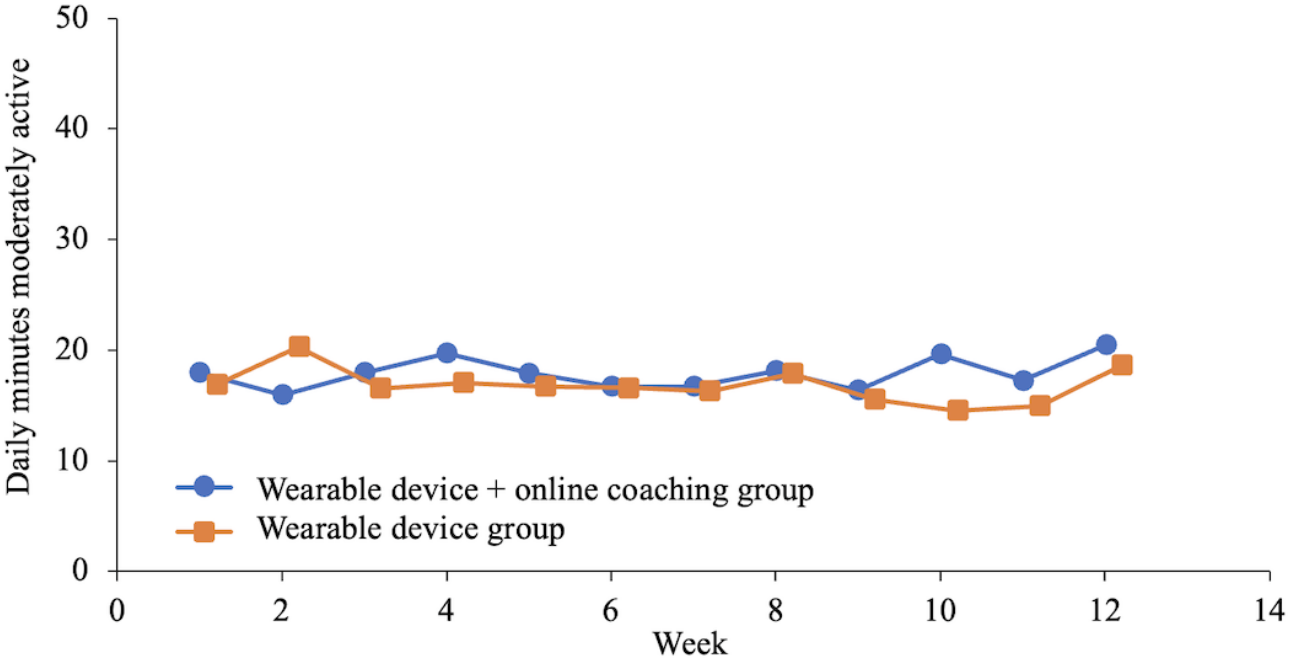


Figure 5. The average daily minutes of highly active taken each week. * *P*<.05 Student *t* test (between-groups).



XSL•FO
RenderX

Figure 6. The average daily minutes of moderately active taken each week.



Discussion

Principal Results

We investigated the effects of remote CR using wearable devices with and without OLC in patients with CAD. The major findings of this study were as follows: (1) using wearable devices and providing a home exercise program based on CPET for 3 months significantly improved peak VO_2 and AT VO_2; (2) the OLC group demonstrated significantly higher physical activity (daily distance and daily duration of high activity) in the latter half of the intervention than did the CON group; and (3) during the research period, there were no dropouts owing to the use of wearable device or OLC.

Limitations

This study had some limitations. First, there was a potential selection bias in the cases enrolled in this study, as it may have included many individuals with high health consciousness and tolerance for wearable devices and desire for OLC. These factors may limit the generalizability of our findings to population groups with lower levels of health consciousness or technological proficiency. Second, this was a single-center study conducted primarily among Japanese patients, with a small sample size and a short follow-up period of 12 weeks. Regional differences and cultural factors may have influenced the results. Furthermore, the predominance of the male sex in this study's cohort may also limit the applicability of the findings to females, older adults, or different ethnic groups. Additionally, varying levels of digital literacy and access to technology could impact the effectiveness of remote CR programs, both within Japan and internationally. Third, the power analysis indicates that this study is sufficiently powered to detect large effect size differences between the OLC group and CON group, specifically a Cohen *d* of 0.8 with a sample size of 50. These effect sizes are rather large and may be unrealistic for the intervention being studied, suggesting that this study is underpowered for detecting more moderate or smaller effect sizes. This study serves primarily as a proof of concept. Future interventional studies with larger sample sizes, longer follow-up durations, and more diverse populations are needed to confirm the findings and evaluate long-term adherence and sustained benefits of such programs.

Comparison With Prior Work

Peak VO_2, measured using the CPET, represents the maximal ability of the body to transport and use oxygen. It is widely recognized as a definitive measure of exercise capacity [25-28]. In patients with CAD, peak VO_2 is associated with all-cause mortality and cardiovascular events [29-32]. AT VO_2, similar to peak VO_2, is considered a prognostic indicator [33,34]. It is recommended to prescribe exercise below the AT to establish a safe exercise tolerance range of intensity for patients [5]. In this study, by providing a home exercise program based on AT VO_2 and using wearable devices, we identified significant enhancements in both peak VO_2 and AT VO_2 over a 12-week intervention period, irrespective of the presence or absence of OLC. Additionally, we observed a significant improvement in $\Delta VO_2/\Delta load$. This parameter serves as an indicator of oxygen transport to the peripheral muscles, reflecting the degree of increase in cardiac output during exercise [35]. It has been reported that $\Delta VO_2/\Delta load$ is an independent prognostic factor in patients with cardiovascular disease [36]. In previous studies, it has been reported that the use of physical activity trackers for more than 12 weeks improves peak VO_2 in patients with CAD, however there is no mention of CPET parameters other than peak VO_2, such as AT VO_2 or $\Delta VO_2/\Delta load$ [15,37]. Additionally, the patient group had already participated in a phase II CR program. Our study targeted patients in phase III and did not consider whether they had participated in the CR program, suggesting that a combined approach using wearable devices and a CPET-based home exercise regimen offers a promising strategy for remote CR in phase III patients with

XSL•FO
RenderX

CAD, potentially leading to improved prognosis and enhanced exercise capacity. However, this study was a short-term intervention of 12 weeks, and determining its effects requires a long-term intervention and follow-up.

The World Health Organization guidelines recommend at least 150 - 300 minutes of moderate physical activity weekly, at least 75 - 150 minutes of vigorous physical activity weekly, or an equivalent combination of both [38]. It has been reported that moderate-to-vigorous-intensity activity and vigorous physical activity, as measured by wearable devices, have a significant negative correlation with the risk of all-cause mortality, cardiovascular death, and onset of heart failure [8,9,39]. In this study, the OLC group had a significantly increased daily distance travelled compared with that by the CON group in the latter period. Additionally, while there was no significant difference in the time spent moderately active (3 - 6 METs), the time spent highly active (>6 METs) significantly increased in the OLC group during the latter half of the intervention. To date, no study has examined the significance of OLC using wearable devices. In this study, weekly OLC had a positive effect on daily activity, particularly vigorous activity, in the latter half of the remote CR program. One possible reason is that OLC may help maintain the motivation to continue exercise habits and provide guidance on appropriate exercise methods. Notably, 1 participant in the OLC group dropped out owing to calf muscle strain, and the OLC group experienced increased body pain as measured by HR-QOL. This suggests that an increase in activity may lead to musculoskeletal problems. To safely continue remote-CR, it is important to not only increase vigorous activity but also to enhance moderate activity. Patient education and OLC are crucial in this regard.

With the advancement of digital devices, including wearable devices and smartphones, it is anticipated that remote-CR will continue to evolve in the future [40]. Prior research on remote CR using these digital platforms has primarily focused on relatively healthy patients in stable condition. Nonetheless, reported dropout rates in studies with a duration under 12 weeks ranged 3% - 13%, even among patients who demonstrated a degree of digital device usage tolerance [41-43]. In this study, the patients were predominantly male (>90%), with an average age of 63.1 (SD 7.2) years, indicating that the cohort was less likely to be affected by the digital divide. Future research must prioritize the validation of remote CR in older and more vulnerable populations, while also assessing the efficacy and safety of interventions over extended periods.

Conclusions

Remote CR using wearable devices and a real-time monitoring system significantly improved exercise capacity in patients with CAD over a 12-week intervention period. Physical activity improved significantly with regular OLC. Future research should focus on diverse patient populations and longer intervention duration to validate these findings.

Acknowledgments

The authors thank the staff of the Department of Cardiovascular Biology and Medicine, Juntendo University. The authors are grateful for the secretarial assistance provided by graphic designers Ms Yumi Nozawa and Ms Shino Nishio. The authors did not receive support from any organization for the submitted work.

Authors' Contributions

RN did the analysis. RN and TD handled the concept and design. RN, TD, NT, YC, TN, HE, HN, IO, HI, SO, KM, HD, and TM interpreted the data. RN drafted of this paper. RN, TN, and MY acquired clinical data. TD edited the writing. TM reviewed and edited the writing. TD and TM supervised the study.

Conflicts of Interest

RN and TD received research grants from M3 Inc. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Overview of the Recoval system. This supplemental material provides an overview of the Recoval system, a platform designed to enable medical personnel to remotely monitor and manage patients' activity and vital data through wearable devices. By integrating patient data from smart devices, the system facilitates remote interventions and personalized care.
[DOCX File, 612 KB - mhealth_v13i1e63797_app1.docx]

Multimedia Appendix 2
Baseline comparison of CPET parameters and mental status. CPET: cardiopulmonary exercise test.
[DOCX File, 19 KB - mhealth_v13i1e63797_app2.docx]

Checklist 1
CONSORT-EHEALTH checklist.
[PDF File, 1288 KB - mhealth_v13i1e63797_app3.pdf]

XSL•FO
RenderX

References

1. Fihn SD, Gardin JM, Abrams J, et al. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012 Dec 18;60(24):2564-2603. [doi: 10.1016/j.jacc.2012.07.012] [Medline: 23182124]
2. Task Force Members, Montalescot G, Sechtem U, et al. 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology. Eur Heart J 2013 Oct;34(38):2949-3003. [doi: 10.1093/eurheartj/eht296] [Medline: 23996286]
3. Roffi M, Patrono C, Collet JP, et al. 2015 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the management of acute coronary syndromes in patients presenting without persistent st-segment elevation of the European Society of Cardiology (ESC). Eur Heart J 2016 Jan 14;37(3):267-315. [doi: 10.1093/eurheartj/ehv320] [Medline: 26320110]
4. Turk-Adawi KI, Grace SL. Narrative review comparing the benefits of and participation in cardiac rehabilitation in high-, middle- and low-income countries. Heart Lung Circ 2015 May;24(5):510-520. [doi: 10.1016/j.hlc.2014.11.013] [Medline: 25534902]
5. Makita S, Yasu T, Akashi YJ, et al. JCS/JACR 2021 guideline on rehabilitation in patients with cardiovascular disease. Circ J 2022 Dec 23;87(1):155-235. [doi: 10.1253/circj.CJ-22-0234] [Medline: 36503954]
6. Ades PA, Keteyian SJ, Wright JS, et al. Increasing cardiac rehabilitation participation from 20% to 70%: a road map from the million hearts cardiac rehabilitation collaborative. Mayo Clin Proc 2017 Feb;92(2):234-242. [doi: 10.1016/j.mayocp.2016.10.014] [Medline: 27855953]
7. Varnfield M, Karunanithi M, Lee CK, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779. [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
8. Khurshid S, Weng LC, Nauffal V, et al. Wearable accelerometer-derived physical activity and incident disease. NPJ Digit Med 2022 Sep 2;5(1):131. [doi: 10.1038/s41746-022-00676-9] [Medline: 36056190]
9. Stamatakis E, Ahmadi MN, Gill JMR, et al. Association of wearable device-measured vigorous intermittent lifestyle physical activity with mortality. Nat Med 2022 Dec;28(12):2521-2529. [doi: 10.1038/s41591-022-02100-x] [Medline: 36482104]
10. Hamaya R, Mori M, Miyake K, Lee IM. Association of smartphone-recorded steps over years and change in cardiovascular risk factors among working-age adults. J Am Heart Assoc 2022 Jul 19;11(14):e025689. [doi: 10.1161/JAHA.121.025689] [Medline: 35861838]
11. Heizmann AN, Chapelle C, Laporte S, Roche F, Hupin D, Le Hello C. Impact of wearable device-based interventions with feedback for increasing daily walking activity and physical capacities in cardiovascular patients: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2023 Jul 11;13(7):e069966. [doi: 10.1136/bmjopen-2022-069966] [Medline: 37433730]
12. Lindman BR, Gillam LD, Coylewright M, et al. Effect of a pragmatic home-based mobile health exercise intervention after transcatheter aortic valve replacement: a randomized pilot trial. Eur Heart J Digit Health 2021 Mar;2(1):90-103. [doi: 10.1093/ehjdh/ztab007] [Medline: 34048509]
13. Izawa KP, Watanabe S, Hiraki K, et al. Determination of the effectiveness of accelerometer use in the promotion of physical activity in cardiac patients: a randomized controlled trial. Arch Phys Med Rehabil 2012 Nov;93(11):1896-1902. [doi: 10.1016/j.apmr.2012.06.015] [Medline: 22750166]
14. Houle J, Doyon O, Vadeboncoeur N, Turbide G, Diaz A, Poirier P. Innovative program to increase physical activity following an acute coronary syndrome: randomized controlled trial. Patient Educ Couns 2011 Dec;85(3):e237-e244. [doi: 10.1016/j.pec.2011.03.018] [Medline: 21546203]
15. Duscha BD, Piner LW, Patel MP, et al. Effects of a 12-week mHealth program on peak VO2 and physical activity patterns after completing cardiac rehabilitation: a randomized controlled trial. Am Heart J 2018 May;199:105-114. [doi: 10.1016/j.ahj.2018.02.001] [Medline: 29754647]
16. Recoval. m3.com. URL: https://m3comlp.m3.com/lp/sapplym/recoval [accessed 2025-04-05]
17. Teramoto T, Sasaki J, Ishibashi S, et al. Diagnostic criteria for dyslipidemia. J Atheroscler Thromb 2013;20(8):655-660. [doi: 10.5551/jat.17152] [Medline: 23892528]
18. Matsuo S, Imai E, Horio M, et al. Revised equations for estimated GFR from serum creatinine in Japan. Am J Kidney Dis 2009 Jun;53(6):982-992. [doi: 10.1053/j.ajkd.2008.12.034] [Medline: 19339088]
19. Adachi H. Cardiopulmonary exercise test. Int Heart J 2017 Oct 21;58(5):654-665. [doi: 10.1536/ihj.17-264] [Medline: 28966333]
20. Nakade T, Adachi H, Murata M, Oshima S. Characteristics of patients with a relatively greater minimum VE/VCO2 against peak VO2% and impaired exercise tolerance. Eur J Appl Physiol 2018 Aug;118(8):1547-1553. [doi: 10.1007/s00421-018-3884-1] [Medline: 29761310]

XSL•FO
RenderX

21. Beaver WL, Wasserman K, Whipp BJ. Bicarbonate buffering of lactic acid generated during exercise. J Appl Physiol (1985) 1986 Feb;60(2):472-478. [doi: 10.1152/jappl.1986.60.2.472] [Medline: 3949651]
22. Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. J Clin Epidemiol 1998 Nov;51(11):1037-1044. [doi: 10.1016/s0895-4356(98)00095-x] [Medline: 9817121]
23. Fukuhara S, Ware JE, Kosinski M, Wada S, Gandek B. Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. J Clin Epidemiol 1998 Nov;51(11):1045-1053. [doi: 10.1016/s0895-4356(98)00096-1] [Medline: 9817122]
24. Valkeinen H, Aaltonen S, Kujala UM. Effects of exercise training on oxygen uptake in coronary heart disease: a systematic review and meta-analysis. Scand J Med Sci Sports 2010 Aug;20(4):545-555. [doi: 10.1111/j.1600-0838.2010.01133.x] [Medline: 20492590]
25. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019 Sep 10;140(11):e563-e595. [doi: 10.1161/CIR.0000000000000677] [Medline: 30879339]
26. Mancini DM, Eisen H, Kussmaul W, Mull R, Edmunds LH Jr, Wilson JR. Value of peak exercise oxygen consumption for optimal timing of cardiac transplantation in ambulatory patients with heart failure. Circulation 1991 Mar;83(3):778-786. [doi: 10.1161/01.cir.83.3.778] [Medline: 1999029]
27. Keteyian SJ, Patel M, Kraus WE, et al. Variables measured during cardiopulmonary exercise testing as predictors of mortality in chronic systolic heart failure. J Am Coll Cardiol 2016 Feb 23;67(7):780-789. [doi: 10.1016/j.jacc.2015.11.050] [Medline: 26892413]
28. Nakanishi M, Takaki H, Kumasaka R, et al. Targeting of high peak respiratory exchange ratio is safe and enhances the prognostic power of peak oxygen uptake for heart failure patients. Circ J 2014;78(9):2268-2275. [doi: 10.1253/circj.cj-14-0047] [Medline: 25056425]
29. Hagnäs MJ, Kurl S, Rauramaa R, et al. The value of cardiorespiratory fitness and exercise-induced ST segment depression in predicting death from coronary heart disease. Int J Cardiol 2015 Oct 1;196(31-33):31-33. [doi: 10.1016/j.ijcard.2015.05.134] [Medline: 26070181]
30. Van de Veire NR, De Winter O, Philippé J, et al. Maximum oxygen uptake at peak exercise in elderly patients with coronary artery disease and preserved left ventricular function: the role of inflammation on top of tissue Doppler-derived systolic and diastolic function. Am Heart J 2006 Aug;152(2):297. [doi: 10.1016/j.ahj.2006.04.029] [Medline: 16875912]
31. Abulimiti A, Nishitani-Yokoyama M, Shimada K, et al. Prognostic impact of peak oxygen uptake and heart rate reserve in patients after off-pump coronary artery bypass grafting. Clin Cardiol 2021 Apr;44(4):580-587. [doi: 10.1002/clc.23579] [Medline: 33634477]
32. Laukkanen JA, Mäkikallio TH, Rauramaa R, Kiviniemi V, Ronkainen K, Kurl S. Cardiorespiratory fitness is related to the risk of sudden cardiac death: a population-based follow-up study. J Am Coll Cardiol 2010 Oct 26;56(18):1476-1483. [doi: 10.1016/j.jacc.2010.05.043] [Medline: 20951323]
33. Gitt AK, Wasserman K, Kilkowski C, et al. Exercise anaerobic threshold and ventilatory efficiency identify heart failure patients for high risk of early death. Circulation 2002 Dec 10;106(24):3079-3084. [doi: 10.1161/01.cir.0000041428.99427.06] [Medline: 12473555]
34. Hughson RL, Green HJ, Sharratt MT. Gas exchange, blood lactate, and plasma catecholamines during incremental exercise in hypoxia and normoxia. J Appl Physiol 1995 Oct 1;79(4):1134-1141. [doi: 10.1152/jappl.1995.79.4.1134]
35. Itoh H, Nakamura M, Ikeda C, et al. Changes in oxygen uptake-work rate relationship as a compensatory mechanism in patients with heart failure. Jpn Circ J 1992 May;56(5):504-508. [doi: 10.1253/jcj.56.504] [Medline: 1602599]
36. Koike A, Itoh H, Kato M, et al. Prognostic power of ventilatory responses during submaximal exercise in patients with chronic heart disease. Chest 2002 May;121(5):1581-1588. [doi: 10.1378/chest.121.5.1581] [Medline: 12006447]
37. Frederix I, Van Driessche N, Hansen D, et al. Increasing the medium-term clinical benefits of hospital-based cardiac rehabilitation by physical activity telemonitoring in coronary artery disease patients. Eur J Prev Cardiol 2015 Feb;22(2):150-158. [doi: 10.1177/2047487313514018] [Medline: 24249840]
38. Okely AD, Kontsevaya A, Ng J, Abdeta C. 2020 WHO guidelines on physical activity and sedentary behavior. Sports Med Health Sci 2021 Jun;3(2):115-118. [doi: 10.1016/j.smhs.2021.05.001] [Medline: 35782159]
39. Strain T, Wijndaele K, Dempsey PC, et al. Wearable-device-measured physical activity and future health risk. Nat Med 2020 Sep;26(9):1385-1391. [doi: 10.1038/s41591-020-1012-3] [Medline: 32807930]
40. Scherrenberg M, Wilhelm M, Hansen D, et al. The future is now: a call for action for cardiac telerehabilitation in the COVID-19 pandemic from the secondary prevention and rehabilitation section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2021 May 14;28(5):524-540. [doi: 10.1177/2047487320939671] [Medline: 32615796]
41. Daskapan A, Arikan H, Caglar N, Tunali N, Ataman S. Comparison of supervised exercise training and home-based exercise training in chronic heart failure. Saudi Med J 2005 May;26(5):842-847. [Medline: 15951880]
42. Karapolat H, Demir E, Bozkaya YT, et al. Comparison of hospital-based versus home-based exercise training in patients with heart failure: effects on functional capacity, quality of life, psychological symptoms, and hemodynamic parameters. Clin Res Cardiol 2009 Oct;98(10):635-642. [doi: 10.1007/s00392-009-0049-6] [Medline: 19641843]

XSL•FO
RenderX

43. Gordon NF, English CD, Contractor AS, et al. Effectiveness of three models for comprehensive cardiovascular disease risk reduction. Am J Cardiol 2002 Jun 1;89(11):1263-1268. [doi: 10.1016/s0002-9149(02)02323-8] [Medline: 12031725]

Abbreviations

AT: anaerobic threshold
BP: blood pressure
CAD: coronary artery disease
CON: control wearable device
CONSORT: Consolidated Standards of Reporting Trials
CPET: cardiopulmonary exercise test
CR: cardiac rehabilitation
HR-QOL: health-related quality of life
MET: metabolic-equivalent task
OLC: online coaching
SF-36: 36-item short form health survey
STAI: State Trait Anxiety Inventory
VO_2: oxygen consumption

Edited by L Buis; submitted 30.06.24; peer-reviewed by M Hollings, S Okita; revised version received 21.01.25; accepted 25.02.25; published 12.05.25.

Please cite as:
Nishio R, Dohi T, Yokoyama M, Nakade T, Takahashi N, Chikata Y, Endo H, Nishiyama H, Okai I, Iwata H, Okazaki S, Miyauchi K, Daida H, Minamino T
Wearable Devices in Remote Cardiac Rehabilitation With and Without Weekly Online Coaching for Patients With Coronary Artery Disease: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e63797
URL: https://mhealth.jmir.org/2025/1/e63797
doi:10.2196/63797

© Ryota Nishio, Tomotaka Dohi, Miho Yokoyama, Taisuke Nakade, Norihito Takahashi, Yuichi Chikata, Hirohisa Endo, Hiroki Nishiyama, Iwao Okai, Hiroshi Iwata, Shinya Okazaki, Katsumi Miyauchi, Hiroyuki Daida, Tohru Minamino. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Therapeutics–Based Cardio-Oncology Rehabilitation for Lung Cancer Survivors: Randomized Controlled Trial

Guangqi Li[1,2], MD; Xueyan Zhou[3,4], MD; Junyue Deng[2], MD; Jiao Wang[5], PT; Ping Ai[6], PhD; Jingyuan Zeng[2], BD; Xuelei Ma[1*], PhD; Hu Liao[7*], PhD

[1]Department of Biotherapy, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, China

[2]West China School of Medicine, Sichuan University, Chengdu, China

[3]Department of Biotherapy, State Key Laboratory of Biotherapy, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

[4]Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China

[5]Department of Rehabilitation Medicine, West China Hospital, Sichuan University, Chengdu, China

[6]Department of Head and Neck Oncology, Cancer Center, West China Hospital, Sichuan University, Chengdu, China

[7]Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Hu Liao, PhD
Department of Thoracic Surgery
West China Hospital
Sichuan University
No. 37 GuoXue Alley
Chengdu, 610041
China
Phone: 86 18980605130
Email: drliaohu@scu.edu.cn

Abstract

Background: Lung cancer ranks as the leading cause of cancer-related deaths. For lung cancer survivors, cardiopulmonary fitness is a strong independent predictor of survival, while surgical interventions impact both cardiovascular and pulmonary function. Home-based cardiac telerehabilitation through wearable devices and mobile apps is a substitution for traditional, center-based rehabilitation with equal efficacy and a higher completion rate. However, it has not been widely used in clinical practice.

Objective: The objective of this study was to broaden the use of digital health care in the cardiopulmonary rehabilitation of lung cancer survivors and to assess its impact on cardiopulmonary fitness and quality of life (QOL).

Methods: Early-stage nonsmall cell lung cancer survivors aged 18-70 years were included. All the participants received surgery 1-2 months before enrollment and did not require further antitumor therapy. Participants were randomly assigned to receive cardiac telerehabilitation or usual care for 5 months. Artificial intelligence–driven exercise prescription with a video guide and real-time heart rate (HR) monitoring was generated based on cardiopulmonary exercise testing. Aerobic exercise combining elastic band–based resistance exercises were recommended with a frequency of 3-5 d/wk and a duration of 90-150 min/wk. The effective exercise duration was recorded when patients' HR reached the target zone ($HR_{resting}$ + [HR_{max} – $HR_{resting}$] × [≈40%-60%]), representing the duration under the target intensity. The prescription used a gradual progression in duration and action intensity based on the exercise data and feedback. Outcome measurements included cardiopulmonary fitness; lung function; cardiac function; tumor marker; safety; compliance; and scales assessing symptoms, psychology, sleep, fatigue, and QOL.

Results: A total of 40 (85%) out of 47 patients finished the trial. The average prescription compliance rate of patients in the telerehabilitation group reached 101.2%, with an average exercise duration of 151.4 min/wk and an average effective exercise duration of 92.3 min/wk. The cardiac telerehabilitation was associated with higher improvement of maximal oxygen uptake peak (3.66, SD 3.23 mL/Kg/min vs 1.09, SD 3.23 mL/Kg/min; *P*=.02) and global health status or QOL (16.25, SD 23.02 vs 1.04, SD 13.90; *P*=.03) compared with usual care. Better alleviation of affective interference (–0.88, SD 1.50 vs 0.21, SD 1.22; *P*=.048), fatigue (–8.89, SD 15.96 vs 1.39, SD 12.09; *P*=.02), anxiety (–0.31, SD 0.44 vs –0.05, SD 0.29; *P*=.048), and daytime dysfunction

XSL•FO
RenderX

(–0.55, SD 0.69 vs 0.00, SD 0.52; *P*=.02) was also observed in the telerehabilitation group. No exercise-related adverse events were identified during the intervention period.

Conclusions: The 5-month, digital therapeutics–based telerehabilitation improved cardiorespiratory fitness in lung cancer survivors with good compliance and safety. Patients receiving telerehabilitation also reported improved QOL with reduced levels of fatigue, anxiety, and daytime dysfunction.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200064000; https://www.chictr.org.cn/showproj.html?proj=180594

(JMIR Mhealth Uhealth 2025;13:e60115) doi:10.2196/60115

KEYWORDS

cardio-oncology rehabilitation; digital therapeutics; telerehabilitation; non-small cell lung cancer; exercise prescription; cardiology; oncology; rehabilitation; cardiorespiratory fitness; cardiopulmonary; cancer; physical activity; digital health; digital technology; randomized controlled trial; wearable; app; quality of life; survivor

Introduction

Cardio-oncology is a specialized field dedicated to studying and managing cardiovascular health in cancer therapy. It has emerged in response to an increasing awareness of the intricate interplay between the 2 disciplines [1]. Globally, the incidence of lung cancer is 64.1 cases per 100,000 individuals in men and 50.3 cases per 100,000 individuals in women. Lung cancer ranks as the second most commonly diagnosed type of cancer overall and the leading cause of cancer-related deaths. For lung cancer survivors, cardiopulmonary fitness is of paramount importance due to its multifaceted impact on overall health, quality of life (QOL), and long-term outcomes [2-7]. Parameters such as maximal oxygen uptake peak (VO_2 peak), predicted postoperative forced expiratory volume in 1 second, and predicted postoperative diffusing capacity of the lungs for carbon monoxide are commonly used to assess surgical risks and are correlated with the prognosis of lung cancer survivors [2,7]. As the primary therapeutic intervention for localized nonsmall cell lung cancer (NSCLC), surgical interventions impact both cardiovascular and pulmonary function [8-13] and increase cardiovascular disease (CVD) risk [14-16]. Therefore, incorporating cardio-oncology rehabilitation into postoperative care for NSCLC survivors should be considered an indispensable component [17]. In addition, previous studies have also reported that exercise could reduce psychological symptoms [18], sleep dysfunction [19], and fatigue [20] and improve QOL [21] in lung cancer survivors.

The effectiveness and safety of traditional medical monitoring under center-based cardiac rehabilitation (CBCR) have been thoroughly validated. Numerous studies indicate that CBCR significantly reduces the incidence of secondary cardiovascular events, lowers hospital readmission rates, and decreases mortality among patients with CVDs [22-25]. Cardiac rehabilitation in China has rapidly expanded with multilevel hospitals gradually offering rehabilitation services. However, the significant investment in terms of manpower, equipment, and facilities, as well as time costs and transportation limitations, have constrained the participation rate, completion rate, and economic efficiency of CBCR. The scarcity of specialized cardiac rehabilitation personnel is a primary factor contributing to low adherence in overpopulated countries like China [26]. In the United States, less than 20% of individuals requiring cardiac rehabilitation participate in such programs [27,28]. The low adherence of cancer survivors to exercise rehabilitation and the need for individualized prescriptions further hinder the progress of CBCR. Home-based cardiac rehabilitation (HBCR) is a viable alternative to the traditional CBCR model. Numerous clinical studies have demonstrated that the effectiveness and safety of HBCR are comparable to CBCR, particularly in cases of stable coronary heart disease, chronic heart failure, acute myocardial infarction, and patients undergoing vascular reconstruction [29,30]. Notably, participation and completion rates in HBCR surpass those in CBCR [29,30]. In a meta-review, 63% of the 25 virtual reality–based exercise rehabilitation programs for cancer-related dysfunctions reached a patient-reported compliance rate of more than 85% [31], although data on cardiopulmonary fitness in NSCLC survivors are still insufficient. An additional advantage lies in the cost-effectiveness of HBCR, making it a more economical option. The European Society of Cardiology, the American Association of Cardiovascular and Pulmonary Rehabilitation, and the Chinese Association of Rehabilitation Medicine have each provided recommendations and guidance on HBCR in the years 2016, 2019, and 2021 [32-34].

Digital health, often associated with the use of information technology, serves to promote health and deliver health care services. This encompasses a range of platforms such as websites, mobile text messages, smartphone apps, and wearable devices embedded with various sensors [35]. For cardio-oncology telerehabilitation, wearable devices using Bluetooth or Wi-Fi technology facilitate remote monitoring of metrics like heart rate (HR), respiratory rate, electrocardiogram, physical activity, and other health indicators [36]. These technologies can enhance patient engagement and adherence in HBCR, strengthen communication between health care providers (HCPs) and patients, assist in managing CVD risk factors, and reduce health care expenses [33,37-39]. Digital therapeutics (DTx)–based rehabilitation has been shown to improve cardiopulmonary fitness, particularly among patients with cardiac disease and those infected with COVID-19 [40-42]. Cardio-oncology telerehabilitation has also been studied to relieve treatment toxicity and prevent CVD [43-45]. Despite this, there are limited studies specifically addressing its effects on patients with lung cancer after surgery. Recovery Plus Health is a software as medical device approved by the Chinese National Medical Products Administration for remote cardiac rehabilitation and prehabilitation. With the app for HCPs, they

XSL•FO
RenderX

can (1) check, modify, or confirm the artificial intelligence (AI)–driven tailored exercise prescription and send it to patients; and (2) check the feedback information from patients and optimize the exercise prescription dynamically. With the app for patients, they can (1) do exercise rehabilitation following the tailored exercise prescription video guide and AI-driven instructions; (2) receive real-time HR monitoring with individualized safety alerts; and (3) provide exercise-specific feedback. This study aimed to evaluate the efficacy, safety, and compliance of a 5-month cardio-oncology rehabilitation program based on the Recovery Plus Health app and wearable devices for NSCLC survivors after surgery.

Methods

Trial Design and Randomization

This is a single-blind, prospective, randomized controlled trial (RCT) to determine the effectiveness, safety, and compliance of a DTx-based cardiopulmonary rehabilitation program. Random numbers were generated before the start of the enrollment by a researcher unrelated to this trial. The main researcher enrolled participants and assigned them to interventions according to the random number and the entry order. The grouping was blinded to patient identity. Outcomes were analyzed blinded to the study group. The reporting of this trial adheres to the CONSORT (Consolidated Standards of Reporting Trials) 2010 statement (Multimedia Appendix 1).

Ethical Considerations

This trial was approved by the Research Ethics Board of the West China Hospital of Sichuan University (2022-1317) and registered at the Chinese Clinical Trial Register (ChiCTR2200064000). The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participating patients and they were allowed to withdraw the consent anytime during the study for any reason. Data were collected at West China Hospital. Patients' data are anonymized. Each patient received a subsidy of RMB 800 (≈US $110).

Inclusion and Exclusion Criteria

The patients were recruited from West China Hospital. Patients who met all of the following criteria were included:

1. Early-stage NSCLC survivors after surgery who did not require radiotherapy or chemotherapy
2. Patients who underwent surgery 1-2 months before enrollment
3. Patients aged 18-70 years
4. Patients who own and are proficient in the use of a smartphone

Patients who met any of the following criteria were excluded:

1. Patients who had engaged in organized aerobic training within the preceding 3 months
2. Patients who had severe orthopedic or cardiopulmonary disease (including severe COVID-19 infection)
3. Patients with cognitive dysfunction or psychological disorders resulting in an inability to understand the content of the study

Sample Size Calculation

According to previous research from Messaggi-Sartor et al [46] on the improvement of VO_2 peak in postoperative exercise intervention for NSCLC, the sample size in this study was estimated to detect a difference of 2.13 mL/Kg/min in VO_2 peak, with a SD of 2.3 mL/Kg/min, a type I error rate α of 0.05, a power of 0.90 ($1-\beta$), and an intraclass correlation coefficient of 0.10 for the multicenter design. It was determined that 45 participants would be required for an independent 2-sample *t* test comparison. Considering a 10% attrition rate, the intended total sample size was 50 (25 in the intervention group and 25 in the control group). PASS (version 15.0; NCSS) was used for sample size calculation.

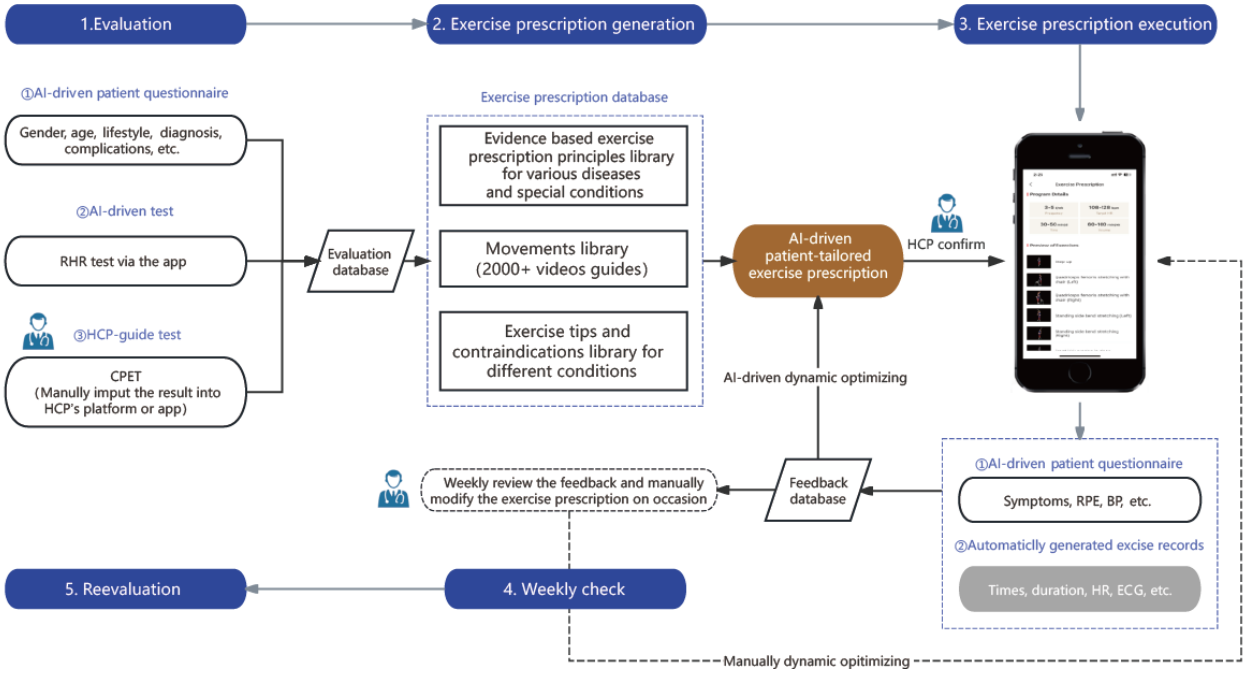
Intervention

Participants in the usual care group received routine postoperative rehabilitation guidance including condition monitoring, medication guidance, dietary guidance, oral exercise guidance, and psychological care. Besides the usual care, participants in the intervention group received exercise prescriptions with a video guide and real-time monitoring using the Recovery Plus Health app (Recovery Plus USA Inc), a software as medical device approved by the National Medical Products Administration for remote cardiac rehab and prehab. They were instructed to download the app for patients (Recovery Plus Inc; Multimedia Appendix 2), which was wirelessly connected to a chest-worn HR band (Recovery Plus Inc) to monitor exercise frequency, intensity, time, volume, and progression. These participants were shown how to operate the app and the equipment and were required to adhere to the prescribed exercises available on the app, wear the HR band during exercise, and provide feedback about the specific exercise after each rehabilitation session. The exercise prescriptions were generated according to the result of the cardiopulmonary exercise testing (CPET) and patients' previous exercise habits. The prescription principle was based on suggestions for patients with cancer in American College of Sports Medicine's guidelines [47] and Expert Consensus on Exercise Rehabilitation for Chinese Cancer Patients Centered on Functional Impairments [48]. Generally, aerobic exercise combining elastic band-based resistance exercises was recommended with a frequency of 3-5 d/wk and a duration of 90-150 min/wk. For each participant, the prescription was dynamically changed and highly personalized. Patient-reported, disease-related symptoms were considered during action selection. The effective exercise duration was recorded when patients' HR reached the target HR zone, representing the duration under the target intensity. The lower number of the target HR zone was calculated as $HR_{resting} + (HR_{max} - HR_{resting}) \times 40\%$. The higher number of the target HR zone was calculated as $HR_{resting} + (HR_{max} - HR_{resting}) \times 60\%$. The prescription used a gradual progression and automatic adjustment in duration and action intensity based on the exercise data and feedback. All the adjustments automatically generated by the system were approved by the physician. The HCP reviewed the exercise records and feedback information via the Recovery Plus Health app for HCPs (Multimedia Appendix 3) weekly and made necessary modifications to the exercise regimen, while the app

XSL•FO
RenderX

automatically recorded compliance with the exercise prescription. Regardless of the group, all participants were instructed to discontinue exercising and promptly report to their clinician or exercise professional if any cardiac-related symptoms occurred. Moreover, HCPs were available 24/7 to receive and respond to alerts from the participants. The intervention began 1-2 months after the surgery. Figure 1 illustrates the operational diagram of the AI-driven exercise prescription system.

Figure 1. Diagram of the AI-driven exercise prescription system. AI: artificial intelligence; BP: blood pressure; CPET: cardiopulmonary exercise test; ECG: electrocardiogram; HCP: health care provider; HR: heart rate; RHR: resting heart rate; RPE: rating of perceived exertion.



Measurements

Outcome measurements included cardiopulmonary fitness (VO_2 peak; primary outcome); lung function (forced expiratory volume in 1 second and diffusing capacity of the lungs for carbon monoxide); cardiac function (tricuspid annular plane systolic excursion and ejection fraction); cytokeratin 19 fragmentation antigen; safety; compliance (completion rate, exercise times per week, effective exercise duration, and prescription adherence rate); and scales assessing symptoms (MD Anderson Symptom Inventory for Patients With Lung Cancer [49,50]), psychology (Hospital Anxiety and Depression Scale [51]), sleep (Pittsburgh Sleep Quality Index [52]), fatigue (Multidimensional Fatigue Symptom Inventory—Short Form [53,54]), and QOL (European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 [55]). All the scales have been studied for validity and reliability. The assessment was conducted at baseline (1-2 months after surgery) and 5 months after the intervention, respectively. All the examinations were performed at West China Hospital and the scales were collected via the web. Exercise data including frequency, duration, HR, and subjective feedback were recorded by the app.

Statistical Analysis

Descriptive statistics are presented as frequency distributions, means, and SD. Fisher exact test was used for unordered qualitative data. Independent 2-sample *t* tests were used for normally distributed continuous quantitative data with equal variance. The Mann-Whitney *U* test was used for ordered qualitative data and nonnormally distributed or heteroscedastic continuous quantitative data. The distribution of data was analyzed by the Shapiro-Wilk test. The equality of variances was tested using the Levene test. Statistical significance was established at a 2-tailed *P* value of less than .05. All analyses were conducted using an intention-to-treat approach.

Results

Participants and Compliance

From September 2022 to September 2023, a total of 47 patients in West China Hospital were enrolled in the study, with 40 (85%) individuals completing the trial (22 in the telerehabilitation group and 18 in the usual care group). Seven participants withdrew from the trial. One participant in the telerehabilitation group was diagnosed with asthma and the other 6 withdrew their consent because they were unable to come back to the hospital for examination on time due to living far away (Figure 2). The baseline data of participants who completed the study are presented in Table 1. The adherence rate of the telerehabilitation group was 88% (22/25). Patients in this group exercised 3.33 times a week on average, with an average exercise duration of 151.4 min/wk. The average effective exercise duration (reaching the prescribed target HR zone) was 92.25 min/wk. The mean prescription adherence rate (effective exercise duration to the lower limit of prescription duration) was 101.2%.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) study flow diagram.

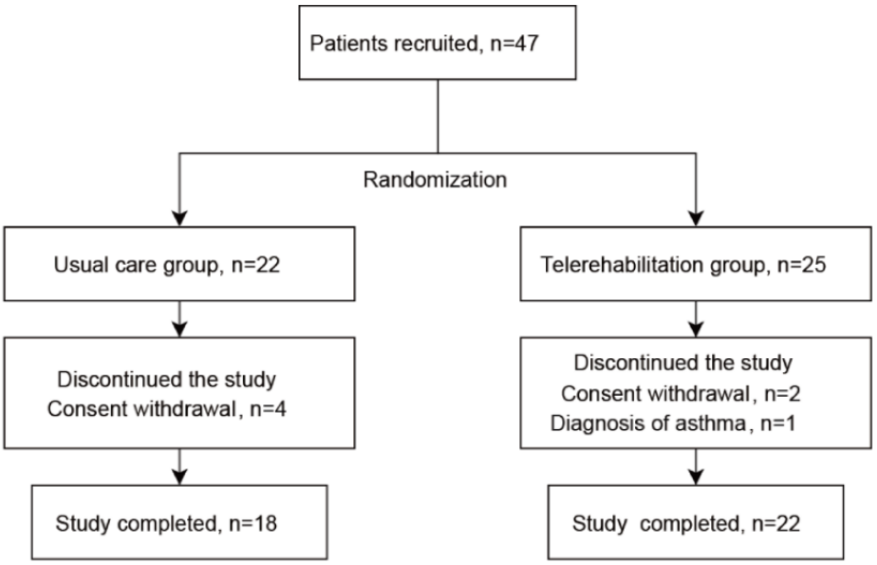


Table 1. Baseline data of participants completed the study.

Characteristics	Total (n=40)	Usual care (n=18)	Telerehabilitation (n=22)	*P* value
Age (years), mean (SD)	51.33 (11.32)	53.50 (10.71)	49.50 (11.66)	.27
Sex, n (%)				.25
Male	9 (22.5)	6 (33.3)	3 (13.6)	
Female	31 (77.5)	12 (66.7)	19 (86.4)	
American Joint Committee on Cancer stage, n (%)				.75
IA1	16 (40)	7 (38.9)	9 (40.9)	
IA2	13 (32.5)	7 (38.9)	6 (27.3)	
IA3	7 (17.5)	3 (16.7)	4 (18.2)	
IB	4 (10)	1 (5.6)	3 (13.6)	
Pathology, n (%)				>.99
Adenocarcinoma	39 (97.5)	18 (100)	21 (95.5)	
Squamous-cell carcinoma	1 (2.5)	0 (0)	1 (4.5)	
VO_2 peak[a] (mL/Kg/min), mean (SD)	19.43 (4.51)	19.98 (4.18)	18.97 (4.82)	.49
MET[b], mean (SD)	6.12 (1.33)	6.29 (1.27)	5.99 (1.4)	.53
FEV_1[c] (L), mean (SD)	2.2 (0.49)	2.16 (0.5)	2.23 (0.5)	.69
DLCO[d] (mmol/min/kPa), mean (SD)	20.44 (4.3)	20.7 (4.92)	20.23 (3.88)	.75
TAPSE[e] (mm), mean (SD)	20.86 (2.97)	21.5 (2.94)	20.32 (2.96)	.25
EF[f], mean (SD)	67.63 (4.92)	67 (4.07)	68.14 (5.56)	.47
CYFRA21[g] (ng/mL), mean (SD)	1.85 (0.7)	1.9 (0.78)	1.81 (0.65)	.72

[a]VO_2 peak: maximal oxygen uptake peak.

[b]MET: metabolic equivalent of task.

[c]FEV_1: forced expiratory volume in 1 second.

[d]DLCO: diffusing capacity of the lungs for carbon monoxide.

[e]TAPSE: tricuspid annular plane systolic excursion.

[f]EF: ejection fraction.

[g]CYFRA21: cytokeratin 19 fragmentation antigen.

XSL•FO
RenderX

Efficacy and Safety

Four participants did not receive a CPET after the intervention and another 4 participants did not complete the scales on time. A total of 36 participants accepted both baseline and postintervention CPETs, with 20 individuals in the telerehabilitation group and 16 in the usual care group. Additionally, 36 participants responded to the baseline and postintervention scales, with 20 individuals in the telerehabilitation group and 16 in the usual care group. The cardiac telerehabilitation was associated with higher improvement of VO_2 peak (3.66, SD 3.23 mL/Kg/min vs 1.09, SD 3.23 mL/Kg/min; *P*=.02) and global health status or QOL (16.25, SD 23.02 vs 1.04, SD 13.90; *P*=.03) compared with usual care. Better alleviation of affective interference (relationships with others, enjoyment of life, and emotions; (–0.88, SD 1.50 vs 0.21, SD 1.22; *P*=.048), fatigue (–8.89, SD 15.96 vs 1.39, SD 12.09; *P*=.02), anxiety (–0.31, SD 0.44 vs –0.05, SD 0.29; *P*=.048), and daytime dysfunction (–0.55, SD 0.69 vs 0.00, SD 0.52; *P*=.02) were also observed in the telerehabilitation group (Table 2). No significant differences in other efficacy indicators were observed between the 2 groups (all *P*>.05). No exercise-related adverse events were identified during the intervention period.

XSL•FO
RenderX

Table 2. Effectiveness outcomes changing for both groups.

Outcomes	Usual care (n=16), mean (SD)	Telerehabilitation (n=20), mean (SD)	*P* value
VO_2 peak[a] (mL/Kg/min)	1.09 (3.23)	3.66 (3.23)	.02
FEV_1[b] (L)	0.20 (0.17)	0.25 (0.23)	.53
DLCO[c] (mmol/min/kPa)	1.74 (2.10)	0.84 (1.87)	.23
TAPSE[d] (mm)	–2.00 (4.40)	0.00 (3.24)	.27
EF[e] (%)	–0.60 (4.90)	–2.33 (5.86)	.37
CYFRA21[f] (ng/mL)	0.13 (1.04)	0.44 (0.67)	.42
MDASI-LC[g]			
Mean core symptom severity	0.05 (1.12)	–0.70 (1.53)	.11
Mean module symptom severity	–0.26 (1.47)	–0.63 (1.16)	.40
Mean total symptom severity	–0.21 (2.29)	–1.33 (2.35)	.16
Mean interference	–0.06 (1.44)	–0.79 (1.64)	.08
Mean activity interference	–0.33 (1.84)	–0.70 (2.07)	.17
Mean affective interference	0.21 (1.22)	–0.88 (1.50)	.048
EORTC[h] QLQ-C30[i]			
Functional scales			
Physical functioning	–1.67 (15.3)	6.00 (8.06)	.11
Role functioning	9.38 (21.05)	6.67 (13.68)	.58
Emotional functioning	–0.52 (16.52)	4.58 (13.10)	.24
Cognitive functioning	–6.25 (13.44)	0.00 (10.81)	.13
Social functioning	3.13 (20.38)	11.67 (18.81)	.32
Symptom scales or items			
Fatigue	1.39 (12.09)	–8.89 (15.96)	.02
Nausea and vomiting	1.04 (11.33)	–1.67 (11.97)	.53
Pain	–6.25 (20.07)	–8.33 (17.52)	.97
Dyspnea	–6.25 (18.13)	–6.67 (25.59)	.97
Insomnia	8.33 (22.77)	0.00 (28.61)	.31
Appetite loss	–6.25 (18.13)	0.00 (15.29)	.26
Constipation	6.25 (25.00)	0.00 (15.29)	.31
Diarrhea	2.08 (14.75)	1.67 (7.45)	.88
Financial difficulties	0.00 (17.21)	–1.67 (22.88)	.60
Global health status or quality of life	1.04 (13.90)	16.25 (23.02)	.03
HADS[j]			
Anxiety	–0.05 (0.29)	–0.31 (0.44)	.048
Depression	0.01 (0.37)	–0.22 (0.38)	.07
PSQI[k]			
Subjective sleep quality	0.06 (0.44)	–0.10 (0.64)	.47
Sleep latency	–0.38 (1.41)	–0.20 (1.15)	.92
Sleep duration	–0.44 (1.36)	0.05 (0.76)	.22
Habitual sleep efficiency	0.19 (0.98)	–0.12 (0.86)	.33
Sleep disturbances	–0.25 (0.86)	–0.20 (0.83)	.92

XSL•FO
RenderX

Outcomes	Usual care (n=16), mean (SD)	Telerehabilitation (n=20), mean (SD)	*P* value
Use of sleeping medication	0.00 (0.63)	0.25 (0.79)	.71
Daytime dysfunction	0.00 (0.52)	–0.55 (0.69)	.02
Global score	–0.81 (3.64)	–0.94 (3.19)	.57
MSFI-SF[l]			
General Fatigue	–0.05 (0.49)	–0.10 (0.89)	.51
Physical Fatigue	–0.13 (0.63)	–0.06 (0.65)	.76
Emotional Fatigue	–0.20 (0.48)	–0.29 (0.84)	.65
Mental Fatigue	0.11 (0.43)	0.00 (0.52)	.49
Vigor	–0.20 (1.04)	0.10 (0.98)	.41
Total Score	–0.06 (2.29)	–0.55 (2.66)	.57

[a]VO_2 peak: maximal oxygen uptake peak.

[b]FEV_1: forced expiratory volume in 1 second.

[c]DLCO: diffusing capacity of the lungs for carbon monoxide.

[d]TAPSE: tricuspid annular plane systolic excursion.

[e]EF: ejection fraction.

[f]CYFRA21: cytokeratin 19 fragmentation antigen.

[g]MDASI-LC: MD Anderson Symptom Inventory for Patients With Lung Cancer.

[h]EORTC: European Organization for Research and Treatment of Cancer.

[i]QLQ-C30: Quality of Life Questionnaire-Core 30.

[j]HADS: Hospital Anxiety and Depression Scale.

[k]PSQI: Pittsburgh Sleep Quality Index.

[l]MSFI-SF: Multidimensional Fatigue Symptom Inventory—Short Form.

Discussion

Strengths and Clinical Implications

This RCT shows the effects of DTx-based telerehabilitation for NSCLC survivors. Both cardiopulmonary fitness and QOL were significantly improved by telerehabilitation compared with usual care. Previous research has demonstrated that exercise-based rehabilitation can enhance cardiopulmonary fitness across a diverse range of disease conditions. A similar RCT in early-stage NSCLC survivors has shown that 8-week aerobic exercise combining high-intensity respiratory muscle training improved VO_2 peak (2.13 mL/Kg/min). However, in this CBCR program, only 69% (11/16) participants persisted in the exercise group with a compliance (session completion rate) of 80%, and no significant QOL improvement was observed [46]. Similar low adherence has been reported in many rehabilitation studies for NSCLC [56,57], especially in advanced NSCLC survivors after surgery [58,59]. As for the impact of exercise interventions on the QOL, conflicting findings have been reported in lung cancer survivors [60]. A meta-analysis involving 34 RCTs demonstrated that unsupervised HBCR did not improve QOL in cancer survivors compared with supervised programs [61]. Patients in unsupervised programs often fail to reach the recommended exercise dosage, which is the main reason for the observed absence of QOL improvement. Maintaining high adherence to exercise routines is a crucial approach to ensure the benefits [60]. Integrating wearable devices into HBCR presents a flexible solution that overcomes the limitations of traditional CBCR, offering patients a more adaptable and convenient way to access rehabilitation programs. In this study, 88% (22/25) of the patients in the telerehabilitation group completed the trial with an impressive average prescription compliance rate of 101.2%, indicating a strong engagement and adherence to the prescribed exercise regimen. DTx enables patients to actively participate in their health management by delivering ongoing feedback and encouragement, motivating them to maintain regular physical activity.

Exercise type affects the outcome of rehabilitation. Aerobic training is particularly effective in improving VO_2 peak and associated cardiopulmonary variables while resistance training primarily focuses on enhancing overall muscular health. An RCT reported that aerobic training rather than resistance training improved VO_2 peak in lung cancer survivors with poor cardiorespiratory fitness [62]. However, a combination of aerobic exercise and resistance exercise can enhance various health outcomes commonly affected by cancer [63]. Another factor that affects the outcome of rehabilitation is the intensity. High-intensity exercise, characterized by a metabolic equivalent of task greater than 6, is defined as an activity that induces sweating and elevates heart and respiratory rates. It has been shown to decrease cancer progression and enhance survival rates in breast, colon, and prostate cancers, with emerging evidence suggesting similar benefits in lung cancer [64]. The exercise prescription in this program combined aerobic and resistance exercises and displayed real-time HR on the guidance interface to help patients maintain an effective exercise intensity.

XSL•FO
RenderX

In addition to good compliance and effectiveness, the following 2 features of this system make it promising for clinical applications. First, the tag-based action recommendation system allows for a high degree of individualization. This is reflected not only in action selection based on patient symptoms and duration settings according to previous exercise habits but also in the ability to adjust action intensity and duration dynamically based on exercise data and feedback. Second, this system enables clinicians to manage a large number of patients simultaneously via the web, viewing patient exercise data and feedback intuitively, thus achieving high management efficiency. These characteristics are difficult to achieve with existing CBCR systems.

Limitations

This study has certain limitations that warrant consideration. First, the relatively short duration of the intervention, which was limited to 5 months, did not allow us to observe long-term outcomes such as prognosis. Implementing the intervention through an app introduces a learning curve for patients. The novelty of digital therapeutics can also be a limitation. Therefore, a longer follow-up period is needed to understand the sustained effects on disease progression and participants' awareness, knowledge, attitudes, intentions to change, help-seeking behaviors, and behavior change, providing a fuller picture of the long-term benefits of HBCR. In addition, a rehabilitation regimen that encompasses a broader time frame, including the perioperative period, could potentially yield greater benefits. Second, the focus of this study was on exercise prescription, which, while a central aspect of cardiac rehabilitation, is only one part of a complex intervention. Comprehensive cardiac rehabilitation programs should also include nutrition guidance, psychological support, medication management, smoking cessation efforts, patient education, and diligent monitoring of key health indicators. Incorporating these multifaceted components could significantly enhance the program's effectiveness and the overall wellness of lung cancer survivors. Third, this study is centered on patients with early-stage, postoperative NSCLC due to their extended survival expectancy, which correlates with a heightened risk of CVDs in the future. They can tolerate most of the exercise prescriptions and intensities, which positions them as the ideal candidates for telerehabilitation following lung surgery. However, for patients with advanced NSCLC, additional treatments such as radiotherapy and chemotherapy can lead to more significant side effects and cardiopulmonary damage. It is essential to conduct further studies to investigate telerehabilitation modalities and their effectiveness for these patients.

Conclusions

This study aimed to expand the application of digital health care in cardiopulmonary rehabilitation after lung cancer surgery. The 5-month, DTx-based telerehabilitation improved cardiorespiratory fitness in lung cancer survivors with good compliance and safety. Patients receiving telerehabilitation also reported improved QOL with reduced levels of fatigue, anxiety, and daytime dysfunction. The findings indicate that incorporating DTx-based cardiac rehabilitation into postoperative care may benefit early-stage NSCLC survivors. More work should be done to address the long-term impact on prognosis.

Acknowledgments

This study was sponsored by Recovery Plus USA Inc.

Authors' Contributions

XL (drmaxuelei@gmail.com) and HL are co-corresponding authors. GL, XZ, JD, and JW contributed equally to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 162 KB - mhealth_v13i1e60115_app1.pdf]

Multimedia Appendix 2
App for patients.
[PDF File (Adobe PDF File), 570 KB - mhealth_v13i1e60115_app2.pdf]

Multimedia Appendix 3
App for health care providers.
[PDF File (Adobe PDF File), 1995 KB - mhealth_v13i1e60115_app3.pdf]

References

XSL•FO
RenderX

1. Sturgeon KM, Deng L, Bluethmann SM, Zhou S, Trifiletti DM, Jiang C, et al. A population-based study of cardiovascular disease mortality risk in US cancer patients. Eur Heart J 2019;40(48):3889-3897 [FREE Full text] [doi: 10.1093/eurheartj/ehz766] [Medline: 31761945]
2. Lee JH, Song EM, Sim YS, Ryu YJ, Chang JH. Forced expiratory volume in one second as a prognostic factor in advanced non-small cell lung cancer. J Thorac Oncol 2011;6(2):305-309 [FREE Full text] [doi: 10.1097/JTO.0b013e318201884b] [Medline: 21150675]
3. Peel AB, Thomas SM, Dittus K, Jones LW, Lakoski SG. Cardiorespiratory fitness in breast cancer patients: a call for normative values. J Am Heart Assoc 2014;3(1):e000432 [FREE Full text] [doi: 10.1161/JAHA.113.000432] [Medline: 24419734]
4. Friedenreich CM, Wang Q, Neilson HK, Kopciuk KA, McGregor SE, Courneya KS. Physical activity and survival after prostate cancer. Eur Urol 2016;70(4):576-585 [FREE Full text] [doi: 10.1016/j.eururo.2015.12.032] [Medline: 26774959]
5. Avancini A, Sartori G, Gkountakos A, Casali M, Trestini I, Tregnago D, et al. Physical activity and exercise in lung cancer care: will promises be fulfilled? Oncologist 2020;25(3):e555-e569 [FREE Full text] [doi: 10.1634/theoncologist.2019-0463] [Medline: 32162811]
6. Berry MF, Yang CJ, Hartwig MG, Tong BC, Harpole DH, D'Amico TA, et al. Impact of pulmonary function measurements on long-term survival after lobectomy for stage I non-small cell lung cancer. Ann Thorac Surg 2015;100(1):271-276 [FREE Full text] [doi: 10.1016/j.athoracsur.2015.02.076] [Medline: 25986099]
7. Jones LW, Watson D, Herndon JE, Eves ND, Haithcock BE, Loewen G, et al. Peak oxygen consumption and long-term all-cause mortality in nonsmall cell lung cancer. Cancer 2010;116(20):4825-4832 [FREE Full text] [doi: 10.1002/cncr.25396] [Medline: 20597134]
8. Licker M, de Perrot M, Höhn L, Tschopp JM, Robert J, Frey JG, et al. Perioperative mortality and major cardio-pulmonary complications after lung surgery for non-small cell carcinoma. Eur J Cardiothorac Surg 1999;15(3):314-319. [doi: 10.1016/s1010-7940(99)00006-8] [Medline: 10333029]
9. Amar D, Roistacher N, Burt M, Reinsel RA, Ginsberg RJ, Wilson RS. Clinical and echocardiographic correlates of symptomatic tachydysrhythmias after noncardiac thoracic surgery. Chest 1995;108(2):349-354. [doi: 10.1378/chest.108.2.349] [Medline: 7634865]
10. Matyal R, Mahmood F, Hess P, Zhao X, Mitchell J, Maslow A, et al. Right ventricular echocardiographic predictors of postoperative supraventricular arrhythmias after thoracic surgery: a pilot study. Ann Thorac Surg 2010;90(4):1080-1086. [doi: 10.1016/j.athoracsur.2010.05.019] [Medline: 20868791]
11. Granger CL, McDonald CF, Irving L, Clark RA, Gough K, Murnane A, et al. Low physical activity levels and functional decline in individuals with lung cancer. Lung Cancer 2014;83(2):292-299. [doi: 10.1016/j.lungcan.2013.11.014] [Medline: 24360323]
12. Zhang Y, Gao Y. Effects of VATS lobectomy, VATS anatomic segmentectomy, and open thoracotomy on pulmonary function of patients with non-small cell lung cancer. Zhongguo Fei Ai Za Zhi 2016;19(10):700-704 [FREE Full text] [doi: 10.3779/j.issn.1009-3419.2016.10.15] [Medline: 27760602]
13. Kim HK, Lee YJ, Han KN, Choi YH. Pulmonary function changes over 1 year after lobectomy in lung cancer. Respir Care 2016;61(3):376-382 [FREE Full text] [doi: 10.4187/respcare.04284] [Medline: 26604331]
14. Yoon DW, Shin DW, Cho JH, Yang JH, Jeong S, Han K, et al. Increased risk of coronary heart disease and stroke in lung cancer survivors: a Korean nationwide study of 20,458 patients. Lung Cancer 2019;136:115-121. [doi: 10.1016/j.lungcan.2019.08.025] [Medline: 31493668]
15. Lowres N, Mulcahy G, Jin K, Gallagher R, Neubeck L, Freedman B. Incidence of postoperative atrial fibrillation recurrence in patients discharged in sinus rhythm after cardiac surgery: a systematic review and meta-analysis. Interact Cardiovasc Thorac Surg 2018;26(3):504-511. [doi: 10.1093/icvts/ivx348] [Medline: 29161419]
16. Semeraro GC, Meroni CA, Cipolla CM, Cardinale DM. Atrial fibrillation after lung cancer surgery: prediction, prevention and anticoagulation management. Cancers (Basel) 2021;13(16):4012 [FREE Full text] [doi: 10.3390/cancers13164012] [Medline: 34439166]
17. Akhter N, Dent S. Cardio-oncology rehabilitation programs-the next phase in improving care for cancer survivors. JAMA Cardiol 2023;8(12):1128-1130. [doi: 10.1001/jamacardio.2023.3568] [Medline: 37819675]
18. Chen H, Tsai C, Wu Y, Lin K, Lin C. Randomised controlled trial on the effectiveness of home-based walking exercise on anxiety, depression and cancer-related symptoms in patients with lung cancer. Br J Cancer 2015;112(3):438-445 [FREE Full text] [doi: 10.1038/bjc.2014.612] [Medline: 25490525]
19. Chen H, Tsai C, Wu Y, Lin K, Lin C. Effect of walking on circadian rhythms and sleep quality of patients with lung cancer: a randomised controlled trial. Br J Cancer 2016;115(11):1304-1312 [FREE Full text] [doi: 10.1038/bjc.2016.356] [Medline: 27811855]
20. Quist M, Sommer MS, Vibe-Petersen J, Stærkind MB, Langer SW, Larsen KR, et al. Early initiated postoperative rehabilitation reduces fatigue in patients with operable lung cancer: a randomized trial. Lung Cancer 2018;126:125-132. [doi: 10.1016/j.lungcan.2018.10.025] [Medline: 30527176]

XSL•FO
RenderX

21. Gerritsen JKW, Vincent AJPE. Exercise improves quality of life in patients with cancer: a systematic review and meta-analysis of randomised controlled trials. Br J Sports Med 2016;50(13):796-803. [doi: 10.1136/bjsports-2015-094787] [Medline: 26719503]
22. Abraham LN, Sibilitz KL, Berg SK, Tang LH, Risom SS, Lindschou J, et al. Exercise-based cardiac rehabilitation for adults after heart valve surgery. Cochrane Database Syst Rev 2021;5(5):CD010876 [FREE Full text] [doi: 10.1002/14651858.CD010876.pub3] [Medline: 33962483]
23. Anderson L, Oldridge N, Thompson DR, Zwisler A, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease: cochrane systematic review and meta-analysis. J Am Coll Cardiol 2016;67(1):1-12 [FREE Full text] [doi: 10.1016/j.jacc.2015.10.044] [Medline: 26764059]
24. Lawler PR, Filion KB, Eisenberg MJ. Efficacy of exercise-based cardiac rehabilitation post-myocardial infarction: a systematic review and meta-analysis of randomized controlled trials. Am Heart J 2011;162(4):571-584.e2. [doi: 10.1016/j.ahj.2011.07.017] [Medline: 21982647]
25. Dalal HM, Doherty P, Taylor RS. Cardiac rehabilitation. BMJ 2015;351:h5000 [FREE Full text] [doi: 10.1136/bmj.h5000] [Medline: 26419744]
26. Zhao D, Liu J, Wang M, Zhang X, Zhou M. Epidemiology of cardiovascular disease in China: current features and implications. Nat Rev Cardiol 2019;16(4):203-212. [doi: 10.1038/s41569-018-0119-4] [Medline: 30467329]
27. Ades PA, Keteyian SJ, Wright JS, Hamm LF, Lui K, Newlin K, et al. Increasing cardiac rehabilitation participation from 20% to 70%: a road map from the million hearts cardiac rehabilitation collaborative. Mayo Clin Proc 2017;92(2):234-242 [FREE Full text] [doi: 10.1016/j.mayocp.2016.10.014] [Medline: 27855953]
28. Beatty AL, Truong M, Schopfer DW, Shen H, Bachmann JM, Whooley MA. Geographic variation in cardiac rehabilitation participation in medicare and veterans affairs populations: opportunity for improvement. Circulation 2018;137(18):1899-1908 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.029471] [Medline: 29305529]
29. Anderson L, Sharp GA, Norton RJ, Dalal H, Dean SG, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2017;6(6):CD007130 [FREE Full text] [doi: 10.1002/14651858.CD007130.pub4] [Medline: 28665511]
30. Buckingham SA, Taylor RS, Jolly K, Zawada A, Dean SG, Cowie A, et al. Home-based versus centre-based cardiac rehabilitation: abridged cochrane systematic review and meta-analysis. Open Heart 2016;3(2):e000463 [FREE Full text] [doi: 10.1136/openhrt-2016-000463] [Medline: 27738516]
31. Su Z, Zhang L, Lian X, Guan M. Virtual reality-based exercise rehabilitation in cancer-related dysfunctions: scoping review. J Med Internet Res 2024;26:e49312 [FREE Full text] [doi: 10.2196/49312] [Medline: 38407951]
32. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
33. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. J Am Coll Cardiol 2019;74(1):133-153 [FREE Full text] [doi: 10.1016/j.jacc.2019.03.008] [Medline: 31097258]
34. Committee of Cardiac Rehabilitation Prevention of Chinese Association of Rehabilitation Medicine, Committee of Cardiovascular Disease of China Association of GerontologyGeriatrics. China expert consensus on center guided home-based cardiac rehabilitation. Zhonghua Nei Ke Za Zhi 2021 Mar 01;60(3):207-215. [doi: 10.3760/cma.j.cn112138-20200629-00630] [Medline: 33663168]
35. Burke LE, Ma J, Azar KMJ, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American heart association. Circulation 2015;132(12):1157-1213 [FREE Full text] [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
36. Lear SA. The delivery of cardiac rehabilitation using communications technologies: the "Virtual" cardiac rehabilitation program. Can J Cardiol 2018;34(10 Suppl 2):S278-S283. [doi: 10.1016/j.cjca.2018.07.009] [Medline: 30274638]
37. Whittaker F, Wade V. The costs and benefits of technology-enabled, home-based cardiac rehabilitation measured in a randomised controlled trial. J Telemed Telecare 2014;20(7):419-422. [doi: 10.1177/1357633X14552376] [Medline: 25400004]
38. Xu L, Li F, Zhou C, Li J, Hong C, Tong Q. The effect of mobile applications for improving adherence in cardiac rehabilitation: a systematic review and meta-analysis. BMC Cardiovasc Disord 2019;19(1):166 [FREE Full text] [doi: 10.1186/s12872-019-1149-5] [Medline: 31299903]
39. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]

XSL•FO
RenderX

40. Pierucci A, de Almeida NS, Lemes, Milanez VF, Oliveira CB, Kretli Winkelströter L, et al. M-health with cardiac rehabilitation improves functional capacity: a systematic review with meta-analysis. Comput Methods Programs Biomed 2025;260:108551. [doi: 10.1016/j.cmpb.2024.108551] [Medline: 39700688]
41. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: a randomized controlled trial. Eur J Prev Cardiol 2020;27(16):1782-1792 [FREE Full text] [doi: 10.1177/2047487320905717] [Medline: 32106713]
42. da Silva MMC, Viana DR, Colucci MG, Gonzaga LA, Arcuri JF, Frade MCM, et al. Effects of a cardiopulmonary telerehabilitation using functional exercises in individuals after COVID-19 hospital discharge: a randomized controlled trial. J Telemed Telecare 2023:1357633X231188394. [doi: 10.1177/1357633X231188394] [Medline: 37559399]
43. Batalik L, Filakova K, Radkovcova I, Dosbaba F, Winnige P, Vlazna D, et al. Cardio-oncology rehabilitation and telehealth: rationale for future integration in supportive care of cancer survivors. Front Cardiovasc Med 2022;9:858334 [FREE Full text] [doi: 10.3389/fcvm.2022.858334] [Medline: 35497988]
44. Batalik L, Chamradova K, Winnige P, Dosbaba F, Batalikova K, Vlazna D, et al. Effect of exercise-based cancer rehabilitation via telehealth: a systematic review and meta-analysis. BMC Cancer 2024;24(1):600 [FREE Full text] [doi: 10.1186/s12885-024-12348-w] [Medline: 38760805]
45. Batalik L, Filakova K, Dosbaba F, Pepera G, Papathanasiou J, Felsoci M, et al. Efficacy of remotely monitored cardiac rehabilitation in patients after cancer treatment: a pilot study. Eur J Prev Cardiol 2022;29(14). [doi: 10.1093/eurjpc/zwac169] [Medline: 35962569]
46. Messaggi-Sartor M, Marco E, Martínez-Téllez E, Rodriguez-Fuster A, Palomares C, Chiarella S, et al. Combined aerobic exercise and high-intensity respiratory muscle training in patients surgically treated for non-small cell lung cancer: a pilot randomized clinical trial. Eur J Phys Rehabil Med 2019;55(1):113-122 [FREE Full text] [doi: 10.23736/S1973-9087.18.05156-0] [Medline: 29984565]
47. ACSM's guidelines for exercise testing and prescription, 11th edition. American College of Sports Medicine. URL: https://www.acsm.org/education-resources/books/guidelines-exercise-testing-prescription [accessed 2024-08-13]
48. Chinese Association of Rehabilitation Medicine, Jiangsu Integrative Medicine Research Society. Expert consensus on exercise rehabilitation for Chinese cancer patients centered on functional impairment. Chin J Rehabil Med 2023 Aug 24;38(1):1-7. [doi: 10.3969/j.issn.1001-1242.2023.01.001]
49. Lili Z, Yu Z. Revision and evaluation of the MD Anderson symptom assessment scale lung cancer module. Tumor 2013 Jun 24;33(5):434-438 [FREE Full text] [doi: 10.3781/j.issn.1000-7431.2013.05.009]
50. Mendoza TR, Wang XS, Lu C, Palos GR, Liao Z, Mobley GM, et al. Measuring the symptom burden of lung cancer: the validity and utility of the lung cancer module of the M. D. Anderson symptom inventory. Oncologist 2011;16(2):217-227 [FREE Full text] [doi: 10.1634/theoncologist.2010-0193] [Medline: 21285393]
51. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
52. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
53. Xiujuan X, Cuiping X, Lin X, Qingzhi L, Hao Z, Yan L, et al. The reliability and validity of the Chinese version of multidimensional fatigue symptom inventory short form. Chin J Pract Nurs 2013 Mar 08;29(7):43-45 [FREE Full text] [doi: 10.3760/cma.j.issn.1672-7088.2013.07.017]
54. Stein KD, Martin SC, Hann DM, Jacobsen PB. A multidimensional measure of fatigue for use with cancer patients. Cancer Pract 1998;6(3):143-152. [doi: 10.1046/j.1523-5394.1998.006003143.x] [Medline: 9652245]
55. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
56. Cavalheri V, Jenkins S, Cecins N, Gain K, Phillips MJ, Sanders LH, et al. Exercise training for people following curative intent treatment for non-small cell lung cancer: a randomized controlled trial. Braz J Phys Ther 2017;21(1):58-68. [doi: 10.1016/j.bjpt.2016.12.005] [Medline: 28442076]
57. Andersen AH, Vinther A, Poulsen LL, Mellemgaard A. A modified exercise protocol may promote continuance of exercise after the intervention in lung cancer patients--a pragmatic uncontrolled trial. Support Care Cancer 2013;21(8):2247-2253. [doi: 10.1007/s00520-013-1781-z] [Medline: 23508894]
58. Quist M, Adamsen L, Rørth M, Laursen JH, Christensen KB, Langer SW. The impact of a multidimensional exercise intervention on physical and functional capacity, anxiety, and depression in patients with advanced-stage lung cancer undergoing chemotherapy. Integr Cancer Ther 2015;14(4):341-349 [FREE Full text] [doi: 10.1177/1534735415572887] [Medline: 25800229]
59. Temel JS, Greer JA, Goldberg S, Vogel PD, Sullivan M, Pirl WF, et al. A structured exercise program for patients with advanced non-small cell lung cancer. J Thorac Oncol 2009;4(5):595-601 [FREE Full text] [doi: 10.1097/JTO.0b013e31819d18e5] [Medline: 19276834]

XSL•FO
RenderX

60. Ma RC, Yin YY, Liu X, Wang YQ, Xie J. Effect of exercise interventions on quality of life in patients with lung cancer: a systematic review of randomized controlled trials. Oncol Nurs Forum 2020;47(3):E58-E72. [doi: 10.1188/20.ONF.E58-E72] [Medline: 32301933]
61. Buffart LM, Kalter J, Sweegers MG, Courneya KS, Newton RU, Aaronson NK, et al. Effects and moderators of exercise on quality of life and physical function in patients with cancer: an individual patient data meta-analysis of 34 RCTs. Cancer Treat Rev 2017;52:91-104 [FREE Full text] [doi: 10.1016/j.ctrv.2016.11.010] [Medline: 28006694]
62. Scott JM, Thomas SM, Herndon JE, Douglas PS, Yu AF, Rusch V, et al. Effects and tolerability of exercise therapy modality on cardiorespiratory fitness in lung cancer: a randomized controlled trial. J Cachexia Sarcopenia Muscle 2021;12(6):1456-1465 [FREE Full text] [doi: 10.1002/jcsm.12828] [Medline: 34658160]
63. Campbell KL, Winters-Stone KM, Wiskemann J, May AM, Schwartz AL, Courneya KS, et al. Exercise guidelines for cancer survivors: consensus statement from international multidisciplinary roundtable. Med Sci Sports Exerc 2019;51(11):2375-2390 [FREE Full text] [doi: 10.1249/MSS.0000000000002116] [Medline: 31626055]
64. Assi M, Dufresne S, Rébillard A. Exercise shapes redox signaling in cancer. Redox Biol 2020;35:101439. [doi: 10.1016/j.redox.2020.101439] [Medline: 31974046]

Abbreviations

AI: artificial intelligence
CBCR: center-based cardiac rehabilitation
CPET: cardiopulmonary exercise test
CONSORT: Consolidated Standards of Reporting Trials
CVD: cardiovascular disease
DTx: digital therapeutics
HBCR: home-based cardiac rehabilitation
HCP: health care provider
HR: heart rate
NSCLC: nonsmall cell lung cancer
QOL: quality of life
RCT: randomized controlled trial
VO2 peak: maximal oxygen uptake peak

Edited by L Buis; submitted 02.05.24; peer-reviewed by G Pepera, M Fraser; comments to author 18.12.24; revised version received 09.01.25; accepted 04.02.25; published 25.02.25.

Please cite as:
Li G, Zhou X, Deng J, Wang J, Ai P, Zeng J, Ma X, Liao H
Digital Therapeutics–Based Cardio-Oncology Rehabilitation for Lung Cancer Survivors: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e60115
URL: https://mhealth.jmir.org/2025/1/e60115
doi:10.2196/60115
PMID:

©Guangqi Li, Xueyan Zhou, Junyue Deng, Jiao Wang, Ping Ai, Jingyuan Zeng, Xuelei Ma, Hu Liao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Recovery of Patient-Reported Outcome Measures vs Gait Parameters Obtained by Instrumented Insoles After Tibial and Malleolar Fractures: Prospective Longitudinal Observational Study

Elke Warmerdam[1], PhD; Marianne Huebner[2], PhD; Caroline Stoll[1]; Andrey Ivanovic Lange[1]; Bergita Ganse[1], MD, PhD

[1]Innovative Implant Development (Fracture Healing), Departments and Institutes of Surgery, Saarland University, Homburg, Germany

[2]Department of Statistics, Michigan State University, East Lansing, MI, United States

Corresponding Author:
Bergita Ganse, MD, PhD
Innovative Implant Development (Fracture Healing)
Departments and Institutes of Surgery
Saarland University
Kirrberger Str. 1
Building 57
Homburg, 66421
Germany
Phone: 49 (0) 6841 ext 1631570
Email: bergita.ganse@uks.eu

Abstract

Background: New technologies from the field of mobile health (mHealth) are increasingly used to improve patient monitoring during rehabilitation. While in recent years, mobile phones, health apps, personal digital assistants, and smartwatches opened up new diagnostic and monitoring opportunities for patients, the development of innovative sensor devices, such as instrumented insoles, has now reached a sufficient level of usability with promising opportunities for clinical practice. According to research on the best method for monitoring recovery after musculoskeletal injury or surgery, the Patient-Reported Outcome Measurement Information System (PROMIS) and wearables such as instrumented insoles are among the most promising newer options. However, it is unknown how a patient's health perception and improvements in instrumented insole-derived gait parameters correlate after surgery for tibial or malleolar fractures.

Objective: This study aimed to compare the longitudinal trajectories in separate PROMIS (sub)scores with gait and further patient-specific parameters, as well as associations between PROMIS scores and gait parameters. It was also aimed to determine the influence of anthropometric parameters and comorbidities.

Methods: A total of 85 patients (39 women and 46 men; average age 50.8, SD 17.1 years) requiring surgery after tibial or malleolar fractures were included in this prospective longitudinal observational study. In the hospital and during follow-up visits, the patients completed the PROMIS Global Health and Pain Interference questionnaires. During the same visits, individually fitted instrumented insoles with 16 pressure sensors, an accelerometer, and a gyroscope each were used to assess the maximal force, pressure distribution, and angular velocity during walking with data being recorded at 100 Hz. Statistical analyses were conducted using linear mixed effect models, pairwise Spearman correlation coefficients, and generalized additive models.

Results: The gait parameters assessed via the instrumented insoles quickly improved during the first 3 months after surgery, followed by a slowing of further improvement. After surgery, the PROMIS scores increased or decreased to extrema that were reached after 6 weeks to 3 months, followed by a return to preinjury values. Between 3 and 6 months, no significant improvements in PROMIS scores were observed. Between 6 months and 1 year, the Physical Health and Mental Health scores still improved significantly (P=.003 in both cases). Men had better Physical Health and lower Pain Interference scores than women (P=.01 and P=.03, respectively). Hypertension had a negative effect on the Physical Health score (P=.03). The associations between the PROMIS score and gait parameters were strongest at approximately 3 months after surgery, predominantly between the Pain Interference score and gait parameters.

Conclusions: The patients' perception improved later than the objective gait parameters obtained by instrumented insoles did. When the gait pattern improved, pain perception correlated with the gait parameters.

XSL•FO
RenderX

Trial Registration: German Clinical Trials Registry DRKS00025108; https://drks.de/search/en/trial/DRKS00025108

(JMIR Mhealth Uhealth 2025;13:e71022) doi:10.2196/71022

KEYWORDS

fracture; gait parameters; generalized additive model; hypertension; instrumented insoles; lower leg; mixed-effects models; patient-reported outcome measures; PROMIS; wearable

Introduction

Wearable devices such as instrumented insoles have been shown to provide excellent objective measures for monitoring rehabilitation progress after surgery and injury, whereas responses obtained from questionnaires reflect subjective patient perspectives [1-3]. It is currently unknown whether and how the trajectories of these separate measures differ throughout recovery. This information, however, is crucial for drawing correct conclusions from such data in clinical practice and research.

Instrumented insoles are equipped with several pressure sensors and often an inertial measurement unit (IMU) containing triaxial accelerometers and gyroscopes. The total force and pressure distribution curves, as well as the IMU data of the gait cycle, can be used to extract parameters to monitor the progress of healing of lower limb fractures [4,5]. Among the most useful parameters for monitoring fracture healing progress with instrumented insoles are the pressure distribution parameters [6,7]. The greatest improvements in the gait pattern are observed throughout the first 3 months after surgery, with additional, smaller improvements up to at least 6 months [8-10].

In addition, patients' perspectives obtained in a standardized manner, such as with the Patient-Reported Outcome Measurement Information System (PROMIS), can provide valuable information. Several questionnaires are being increasingly used in orthopedic care [2,11,12]. The Pain Interference and Physical Function subscales are among the most popular PROMIS subscales [11,13]. The Global Health Questionnaire includes a Mental Health component and a Physical Health component and has been used in several studies after fracture treatment [14-18]. In these studies, the Mental and Physical Health scores for patients with lower extremity fractures were worse than those for patients with upper extremity fractures [15,18]. Mental and Physical Health scores are also associated with fracture severity in patients with malleolar fractures [16]. The Mental Health score was correlated with the Physical Function score [14].

In studies in which the PROMIS (Physical Health or Physical Function) was completed more than twice after surgery with a follow-up time of more than 1 year, improvements in the Physical Function score were observed up to 6 months [19]. Improvements in the Physical Health score were observed up to 1 year after surgery [15]. The PROMIS, which measures mobility and physical function, is correlated with the RAND-36 Physical Function Scale, the Short Musculoskeletal Function Assessment, the Foot and Ankle Ability Measures, and the University of California Los Angeles Activity Scale [20]. However, the associations of the PROMIS Global Health and Pain Interference scores after surgery remain largely unknown. Therefore, it is unclear whether longitudinally collected patient-reported outcome data after fractures can inform health care providers and how these data can be compared with data obtained from instrumented insoles.

To answer these questions, the primary hypothesis of this study was that the longitudinal trajectories of the PROMIS scores for Physical Health, Mental Health, and Pain Interference differ from those of gait parameters assessed with instrumented insoles among patients with tibial and malleolar fractures, with sex, age, fracture type, implant type, BMI, diabetes, hypertension, cancer, and previous injuries as explanatory variables. The secondary hypothesis was that PROMIS scores are associated with changes in gait parameters of the injured leg after surgery.

Methods

Participants

Patients who underwent surgery for tibial and malleolar fractures at Saarland University Hospital, Germany, were prospectively recruited between February 2022 and May 2024. The exclusion criteria were being younger than 18 years, inability to give consent, requiring walking aids before the fracture, other injuries, or disorders that affect mobility and pregnancy. Patients were recruited during their hospital stay after surgery. The study size was determined by the length of the recruitment period and not based on a sample size calculation. The length of the recruitment period was determined by the ending nonpermanent work contracts of the staff.

Data Collection

During the hospital stay, anthropometric variables, such as sex, age, weight, and height, as well as information about comorbidities, previous injuries, and smoking status, were recorded. Details on the fracture and implant types (nail or plate) were extracted from the electronic hospital system.

During each follow-up visit, patients were asked to fill out the PROMIS Global Health questionnaire (version 1.2) and the Pain Interference short form 8a questionnaire (Item Bank version 1.0), which were developed for adults [21]. The paper-based questionnaires were written in German and completed by each patient without assistance. The baseline PROMIS forms were collected during the hospital stay after surgery, at which time patients were asked to assess their preinjury condition. Patients then returned for outpatient follow-up visits 6 weeks, 3 months, 6 months, 9 months, and 1 year after surgery. Each time, they completed the same questionnaires to document their current condition.

Gait assessments were performed with instrumented insoles containing 16 pressure sensors, a triaxial accelerometer, and a

XSL•FO
RenderX

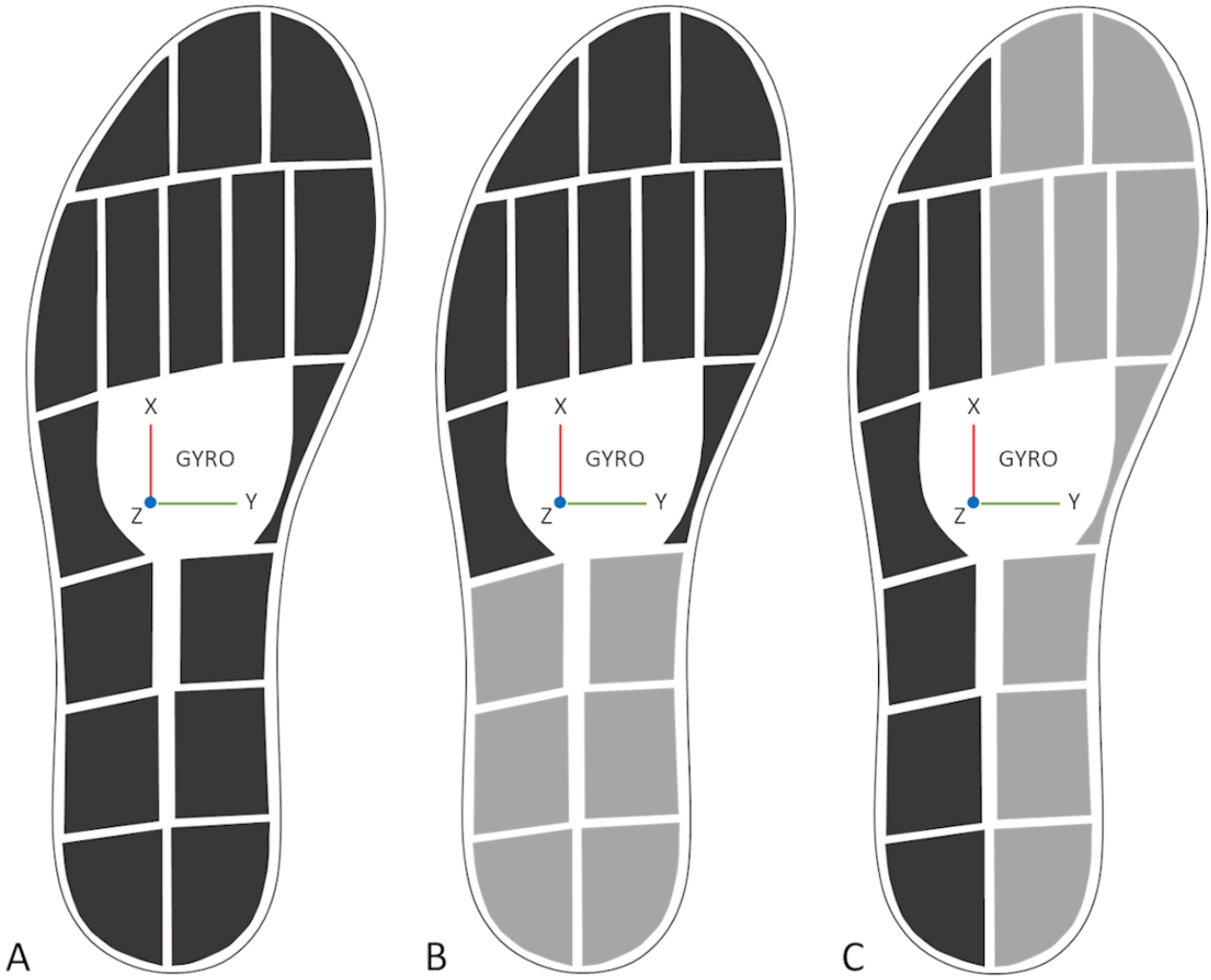
triaxial gyroscope (OpenGO insoles, Moticon GmbH). The insoles were fitted to the shoe soles of each patient, and then calibrated to their body weight. Data were recorded on a 10-m walkway during regular gait at each patient's own preferred speed at 100 Hz. Gait assessments took place during the initial hospital stay and during each outpatient follow-up visit together with the PROMIS assessments.

Data Processing

The PROMIS Global Health items were used to calculate the raw Global Physical Health score and the Global Mental Health score. The raw scores of the Global Physical Health, Global Mental Health, and Pain Interference scores were used to calculate T scores for further analysis [22]. If there were missing responses to one specific score, only this score was excluded for that visit of the patient.

The insole software (OpenGO, Moticon GmbH) provided the total vertical force, the single pressure sensor data, and the gyroscope data (Figure 1A). A custom-written MATLAB (The MathWorks, Inc) code was used to process the data further. To remove noise from the data, a fourth-order Butterworth filter with a cutoff frequency of 6 Hz was applied to all the data. The detection of the initial and final contact of a stride was based on an arbitrary threshold of 30 N in the vertical force data. The first instant at which the force exceeded the threshold and remained above this threshold for at least 300 milliseconds was defined as initial contact. The first instant after the initial contact when the force dropped under 30 N was defined as the final contact. The force data were normalized by first dividing them by the body weight (in Newtons) and then multiplying them by 100 to obtain the force data as a percentage of body weight. The pressure data were normalized to the percentage of body weight and then divided by the number of pressure sensors that were combined to obtain the pressure underneath the forefoot, hindfoot, lateral, and medial sides of the foot (Figures 1B and 1C). Only the angular velocity around the mediolateral axis was used for further analysis. From the total force, the 4 types of pressure data and the angular velocity data, the maximum per stride of the injured side were extracted. The first and last few strides were discarded, such that only the 5 strides in the middle part of the 10-m walkway were kept for analysis to minimize the effects of gait initiation and termination. The average of these 5 strides on the injured side was used for analysis.

Figure 1. Layout of pressure sensors and location and orientation of the gyroscope in the instrumented insole. (A). Total force, which is calculated from all the pressure sensor data pooled. (B). Forefoot (dark gray) and hindfoot (light gray) pressure sensor combination. (C). Lateral (dark gray) and medial (light gray) pressure sensor combinations.



XSL•FO
RenderX

Whether patients experienced fracture union or nonunion was judged on the basis of radiographs taken 6 months after surgery as part of the regular clinical care procedure. Nonunion was defined as a lack of visible callus bridging on radiographs and was determined by an experienced specialist orthopedic surgeon. The lack of visible callus bridging was chosen as an objective criterion to avoid subjective bias. Similarly, the questionnaires and treadmill recordings were objective measures.

Statistical Analysis

Continuous variables are summarized with means and SD, and categorical variables are summarized with frequencies and percentages in the tables. Wilcoxon or Pearson tests were used to compare these variables between males and females, as appropriate.

Disjoint time intervals were defined as 0-15 days (preinjury level for PROMIS and first week for gait analysis), 16-55 days (6 weeks), 56-110 days (3 months), 111-200 days (6 months), 201-300 days (9 months), and more than 300 days (1 year). If more than 1 measurement occurred during the same time interval for a particular patient, only the last measurement was included in the analysis set for that patient.

A 5-point score change in the T scores of the PROMIS Physical Health, PROMIS Mental Health, and PROMIS Pain Interference between visits was considered clinically meaningful. This corresponds to a moderate effect size change of 0.5 SD, as shown in other studies with orthopedic patient populations [23,24].

Linear mixed effects models (LMMs) were used to estimate the time course of the PROMIS components (physical health, mental health, and pain interference) after surgery. Random effects accounted for variation at the individual level. Patient data from all laboratory visits (up to 6 visits) were used for the estimation. Separate models were developed for each of the 3 PROMIS constructs. The explanatory variables were sex, age, fracture type, implant type, BMI, diabetes, hypertension, cancer, and previous injuries. These variables were included in a full model, and the variables with the largest coefficients (absolute size larger than 1) were included in all final models (sex and hypertension). LMMs were also used to estimate the time course of gait parameters (total force, forefoot, hindfoot, lateral, and medial pressure, and angular velocity) after surgery in a similar fashion as the PROMIS.

Pairwise Spearman correlation coefficients were calculated between PROMIS components (physical health, mental health, and pain interference) and gait parameters (maximal force, medial, lateral, forefoot, hindfoot pressure, and angular velocity) at 3 time points, 6 weeks, 3 months, and 6 months, since during these time points, the largest changes in both PROMIS and gait parameters were expected.

Generalized additive models for location, scale, and shape (GAMLSS) [25] were used to estimate the distribution of gait parameters on the injured leg for a given score in the PROMIS Physical Health, Mental Health, and Pain Interference (Model 1) below. Separate models were constructed for each gait parameter (maximal force, medial, lateral, forefoot, hindfoot, and angular velocity). GAMLSSs are regression analyses that allow modeling of the outcome with a parametric distribution whose moments are estimated as smooth curves for the covariate PROMIS scores. We used the Box-Cox Cole Green (BCCG) distribution for the outcome [26]. The moments for these distributions correspond to the median, coefficient of variation, and Box-Cox power transformation needed to adjust for skewness in the distribution. The associations with the gait parameters were estimated using penalized B-splines for each PROMIS subscore.

Model 1:

$$\mu = a_{\mu} + pb(\text{Physical Health}) + pb(\text{Mental health}) + pb(\text{Pain Interference})$$

$$\log(\sigma) = a_{\sigma} + pb(\text{Pain Interference})$$

$$\nu = a_{\nu}$$

All analyses were conducted using R (version 4.0.0; R Core Team) [27]. The lmerTest package was used for the mixed effects models. The GAMLSS package was used to estimate the associations of PROMIS scores with the gait parameters of the injured legs. *P* values of <.05 indicate significance in all tests.

Ethical Considerations

The study protocol was approved by the institutional review board of Saarland Medical Board (Ärztekammer des Saarlandes, Germany, application number 30/21). Written informed consent was obtained from all participants before the start of the measurements. Participants had the opportunity to opt out at any time without needing to give reasons. This observational, prospective longitudinal cohort study was conducted in accordance with the Declaration of Helsinki. The data were pseudonymized. The participants did not receive any financial or nonfinancial compensation for their study participation.

Results

Overview

A total of 102 patients were initially recruited. Out of this dataset, patients were excluded in the case of the inability to assess gait or PROMIS (missingness of observational units). Accordingly, 85 patients (39 women and 46 men) were included in the analysis dataset (Figure 2). Smoking and diabetes were more prevalent in male patients than in female patients (Table 1). Age, comorbidities, fracture type, implant type, and previous injuries did not differ between men and women. There were no sex differences in the PROMIS scores at the preinjury level or at 6 months or 1 year after surgery (Table 2).

XSL•FO
RenderX

Figure 2. Flowchart of patients included in this study. PROMIS: Patient-Reported Outcome Measurement Information System.

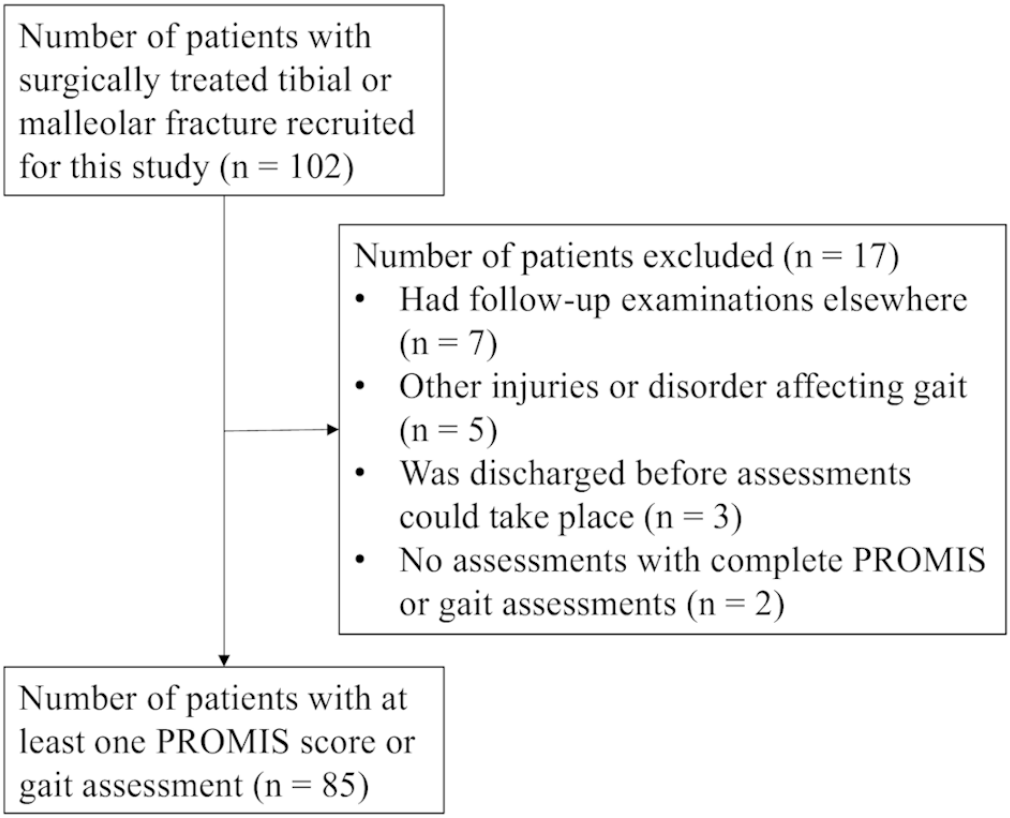


Table 1. Demographics and diagnoses by sex.

Demographics[a]	Total (N)	Female (n=39)	Male (n=46)	Combined (n=85)	Test statistics		
					F test (*df*)	Chi-square (*df*)	*P* value
Age (years), mean (SD)	85	52.8 (17.9)	49.0 (16.4)	50.8 (17.1)	1.11 (1, 83)	—[b]	.29
BMI (kg/m^2), mean (SD)	85	26.43 (4.89)	27.35 (3.47)	26.93 (4.18)	1.80 (1, 83)	—	.18
Implant type: n nail (%)	85	5 (13)	12 (26)	17 (20)	—	2.32 (1)	.13
Smoking, n (%)	84	3 (8)	16 (35)	19 (23)	—	8.60 (1)	.003
Hypertension, n (%)	85	6 (15)	7 (15)	13 (15)	—	0 (1)	.98
Diabetes, n (%)	85	1 (3)	7 (15)	8 (9)	—	3.96 (1)	.05
Cancer, n (%)	85	2 (5)	0 (0)	2 (2)	—	2.42 (1)	.12
Previous injuries, n (%)	85	6 (15)	7 (15)	13 (15)	—	0 (1)	.98
Fracture type, n (%)							
Proximal tibia	85	10 (26)	10 (22)	20 (24)	—	4.21 (2)	.12
Tibial shaft	—	6 (15)	16 (35)	22 (26)	—	—	—
Malleolar	—	23 (59)	20 (44)	43 (51)	—	—	—

[a]For continuous variables, means (SDs) and *F* test scores of the Wilcoxon tests are reported. For categorical variables, n (percentages) and chi-square values from Pearson tests are reported.

[b]Not applicable.

XSL•FO
RenderX

Table 2. Patient-Reported Outcome Measurement Information System scores preinjury and 6 months and 1 year after surgery, separated by sex.

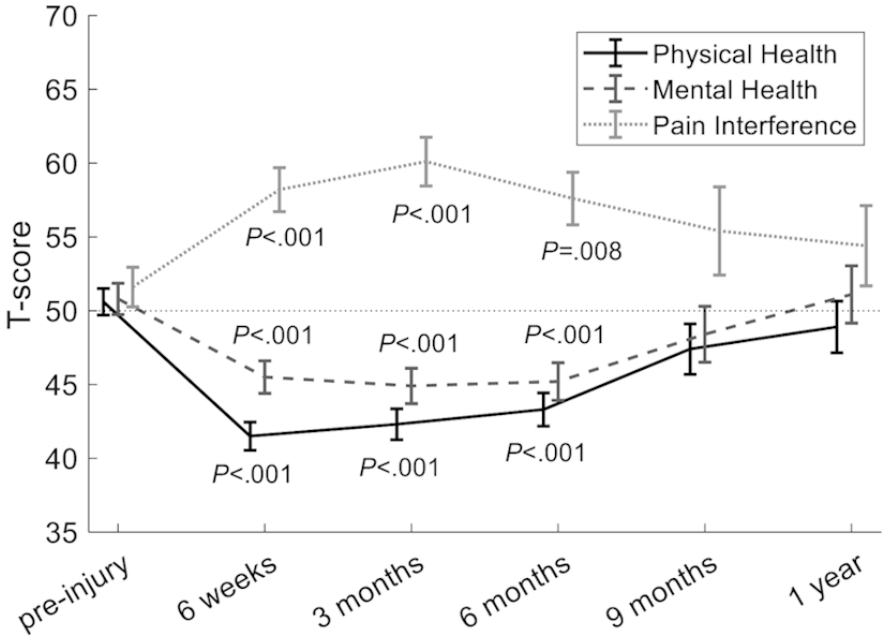
PROMIS system	Total (N)	Female, mean (SD)	Male, mean (SD)	Combined, mean (SD)	Test statistics	
					F test (*df*)	*P* value
PROMIS[a] T score preinjury level						
Physical Health	70	49.60 (10.25)	51.68 (7.39)	50.90 (8.82)	0.43 (1, 67)	.51
Mental Health	69	50.91 (10.53)	51.24 (9.22)	51.21 (9.72)	0.02 (1, 66)	.90
Pain Interference	51	55.0 (13.6)	48.3 (10.3)	51.2 (12.1)	3.20 (1, 48)	.08
PROMIS T score 6 months after surgery						
Physical Health	39	41.84 (6.30)	44.45 (6)	43.33 (6.12)	2.47 (1, 36)	.13
Mental Health	40	44.36 (8.36)	45.82 (7.30)	45.16 (7.64)	0.22 (1, 37)	.65
Pain Interference	26	57.02 (7.24)	58.08 (8.15)	57.62 (7.43)	0.14 (1, 23)	.71
PROMIS T score 1 year after surgery						
Physical Health	13	46.0 (8.4)	54.7 (10.0)	49.3 (9.7)	3.05 (1, 11)	.11
Mental Health	13	48.6 (12.2)	54.4 (11.5)	50.9 (11.8)	0.91 (1, 11)	.36
Pain Interference	10	54.35 (11.53)	53.15 (9.84)	53.87 (10.32)	0.09 (1, 8)	.77

[a]PROMIS: Patient-Reported Outcome Measurement Information System.

PROMIS Trajectories Throughout the Healing Process

After surgery, the Physical Health, Mental Health and Pain Interference (PROMIS) scores increased or decreased to extrema that were reached after 6 weeks to 3 months, followed by a return to preinjury values (Figure 3). All scores were significantly worse than they were before surgery during the first 6 months after surgery (Table S1 in Multimedia Appendix 1). No significant differences were found between preinjury, 9 months, and 1 year after surgery. Between 3 and 6 months, no significant improvements in PROMIS scores were observed. Between 6 months and 1 year, the Physical Health and Mental Health scores improved significantly (*P*=.003 in both cases). Compared with women, men had better Physical Health and lower Pain Interference scores (*P*=.01, *P*=.03, respectively). Hypertension had a negative effect on the Physical Health score (*P*=.04, Table S1 in Multimedia Appendix 1).

Figure 3. The trajectories of the Patient-Reported Outcome Measurement Information System scores and the means with SE are presented. A decrease in physical health and mental health and an increase in pain interference indicate a worse score.

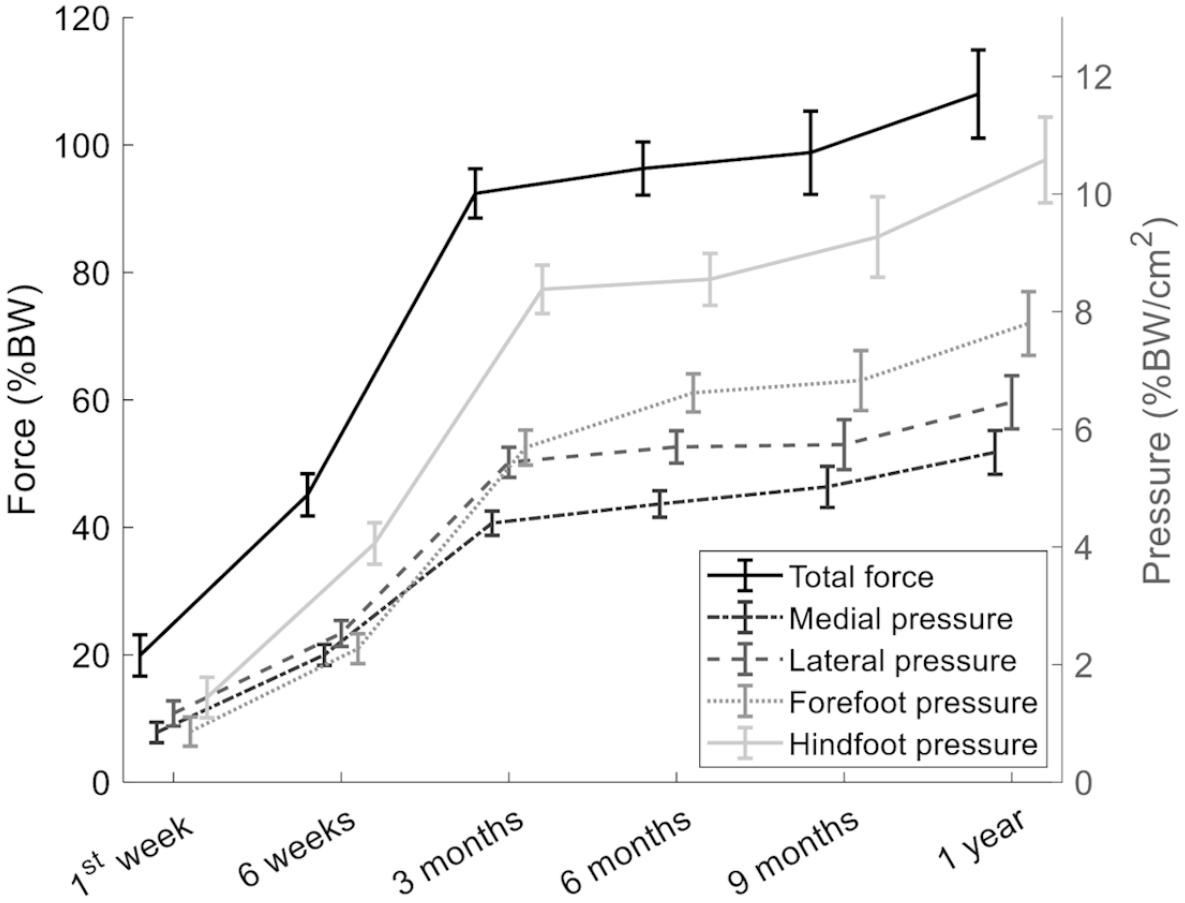


XSL•FO
RenderX

Gait parameter Trajectories Throughout the Healing Process

Compared with the first assessment shortly after surgery, all gait parameters were significantly better at all follow-up assessments (Figure 4). After an initial fast improvement, no significant difference was found between 3 and 6 months for any of the gait parameters, except for the improvement seen in forefoot pressure. After 6 months, there was no further improvement in total force or lateral or forefoot pressure, whereas the medial and hindfoot pressures continued to improve significantly for another 6 months. Neither sex nor hypertension affected gait parameters (Table S2 in Multimedia Appendix 1).

Figure 4. Trajectories of insole-derived gait parameters' means with SE. All the parameters at all the time points were significantly different from those at the first week. BW: body weight.



Associations Between the PROMIS Score and Gait

The associations between the PROMIS score and gait parameters were calculated for patients whose data were collected at the same time intervals. At 6 weeks, the correlation coefficients of the PROMIS scores and gait parameters were small (absolute values less than 0.3). The largest correlation coefficients (defined as absolute correlation coefficients close to 0.5) between all PROMIS scores and gait parameters except for the maximal total force were observed at 3 months. At 6 months, the corresponding correlation coefficients were again smaller (Table 3).

XSL•FO
RenderX

Table 3. Correlation coefficients (Spearman) of Patient-Reported Outcome Measurement Information System scores with gait parameters (n is due to missingness of either gait or patient-reported outcome measurement information system).

Time post surgery	Total force	Medial pressure	Lateral pressure	Forefoot pressure	Hindfoot pressure	Angular velocity
PROMIS[a] Physical Health						
6 weeks (n=37)	0.200	0.101	0.246	0.054	0.237	0.153
3 months (n=28)	0.257	0.379	0.452	0.282	0.477	0.453
6 months (n=26)	0.301	0.247	0.407	0.336	0.373	0.454
PROMIS Mental Health						
6 weeks (n=37)	0.246	0.218	0.291	0.194	0.268	0.241
3 months (n=28)	0.252	0.422	0.526	0.461	0.486	0.470
6 months (n=26)	0.253	0.184	0.372	0.264	0.297	0.478
PROMIS Pain Interference						
6 weeks (n=37)	–0.124	–0.057	–0.192	–0.022	–0.155	–0.065
3 months (n=28)	–0.328	–0.550	–0.538	–0.437	–0.513	–0.550
6 months (n=26)	–0.368	–0.410	–0.456	–0.466	–0.382	–0.545

[a]PROMIS: Patient-Reported Outcome Measurement Information System.

Generalized additive models [25] were used to estimate the associations of PROMIS scores (Mental Health and Physical Health, and Pain Interference) with each gait parameter on the injured leg for a given score on the PROMIS at 3 months (Table 4). This time point was chosen because of the greater correlations between PROMIS scores and gait parameters and because of the larger sample size than those at later time points. Among the PROMIS subscores, the mean pain interference score was significantly associated with all gait parameters. Physical health was associated with medial and forefoot pressure. Mental health was not associated with gait data.

Table 4. Coefficients of generalized additive models for location, scale, and shape for models with Patient-Reported Outcome Measurement Information System scores as explanatory variables and each gait parameter as the dependent variable at 3 months (n=28). The estimates (SE) and *P* values are presented.

	Dependent variable: gait parameters											
	Total force		Medial pressure		Lateral pressure		Forefoot pressure		Hindfoot pressure		Angular velocity	
	Mean (SD)	*P* value	Mean (SD)	*P* value	Mean (SD)	*P* value	Mean (SD)	*P* value	Mean (SD)	*P* value	Mean (SD)	*P* value
Mu link function: identity												
Intercept	365.46 (86.91)	<.001	21.55 (2.03)	.001	17.04 (5.87)	.009	22.97 (2.95)	<.001	21.37 (5.48)	<.001	656.65 (245.55)	.01
Physical Health	–2.14 (1.89)	.27	–0.15 (0.05)	.004	–0.03 (0.08)	.70	–0.37 (0.10)	.001	0.04 (0.10)	.72	–2.82 (2.91)	.34
Mental Health	0.96 (1.07)	.38	0.03 (0.03)	.29	0.02 (0.04)	.64	0.13 (0.07)	.06	0.05 (0.04)	.26	3.12 (2.41)	.21
Pain Interference	–3.83 (0.79)	<.001	–0.20 (0.02)	<.001	–0.19 (0.08)	.02	–0.096 (0.02)	<.001	–0.30 (0.07)	<.001	–5.87 (2.45)	.03
Sigma link function: log												
Intercept	–7.08 (2.74)	.02	–9.77 (2.17)	<.001	–5.49 (3.98)	.18	–10.59 (2.01)	<.001	–8.85 (5.33)	.11	–3.83 (1.58)	.02
Pain Interference	0.10 (0.05)	.05	0.14 (0.04)	.001	0.08 (0.07)	.30	0.16 (0.04)	<.001	0.14 (0.10)	.19	0.04 (0.03)	.19
Nu link function: identity												
Intercept	1.20 (0.52)	.03	1.75 (0.55)	.005	1.51 (0.55)	.01	1.17 (0.39)	.007	1.80 (0.59)	.006	1.40 (1.82)	.45

XSL•FO
RenderX

Discussion

Principal Findings

The PROMIS scores and gait parameters revealed distinct trajectories throughout the healing process of the lower leg and malleolar fractures. The gait parameters showed the most pronounced improvements during the first 3 months, whereas the PROMIS scores lagged and showed worsening to an extremum, followed by improvements at approximately 6 to 9 months. Nen had better Physical Health and lower Pain Interference scores than women. Hypertension had a negative effect on the Physical Health score.

PROMIS

The Mental Health and Pain Interference scores were at their worst 3 months after surgery, indicating that, according to the patient's perspective, it takes longer for their health to improve than indicated by the insole gait measurements. Patients might have expected to be free of symptoms at this time and may be disappointed about their slower-than-expected progress. Similar psychological effects are known from populations in confined environments, such as when over-wintering on Antarctic stations, where the worst mood, called the third-quarter phenomenon, is usually reported after 3 quarters of the mission length [28]. Unmet expectations have been shown to be a mental burden for patients [29]. Different patient experiences and recovery challenges have been reported after ankle fracture surgery, with the key themes of "understanding the recovery journey," "navigating the healthcare system," "understanding personal physical capabilities," "building confidence for weight-bearing," and "resuming daily activities" [30]. In addition, mobility, loss of independence, health care, psychological effects, social and family life, ankle symptoms, sleep disturbance and fatigue, and activities of daily living have been identified as majorly relevant factors for patients with ankle fractures [31]. These patients considered 5 factors to be particularly important: ability to regain independence, sleep quality and quantity, ability to drive, ability to walk without walking aids or weight-bearing restrictions, and radiological union [31]. The worst PROMIS scores 3 months after surgery in this study may therefore likely be explained by the persistence of at least some of these factors. Thus, in future studies, researchers should try to identify the detailed underlying reasons to determine whether interventions may be available to address at least some of them.

Correlations Between Gait and PROMIS

In this study, a novel statistical approach was used to model the associations of PROMIS scores with gait parameters. The new approach considers that the variability of a variable across a range of measured values can change (heteroscedasticity) and can model complex distributions. The greatest correlations were found approximately 3 months after surgery. These values were between those of pain interference and gait parameters, indicating that the amount of pain is associated with gait patterns. This finding aligns with a study on hip arthroscopy, where moderate correlations were found 3 months after surgery between gait speed and the PROMIS Physical Function and Pain scores [32]. In other patient groups with chronic pain conditions, such as knee osteoarthritis, patellofemoral pain, low back pain, and neck pain, patients walk more slowly, with shorter strides, and smaller joint angles and moments [33-36]. Therefore, pain and gait seem to be associated with each other. Pain reliever use and partial weight bearing may be the cause of the lack of significant correlations between gait and pain interference 6 weeks after surgery.

According to the generalized additive models, PROMIS Mental Health scores were not associated with gait 3 months after surgery. In people with mental health problems but without physical problems, mental health is associated with gait [37,38]. It is possible that in our cohort, mental health also influenced gait parameters, but the accompanying physical health problems might have had a greater effect on gait.

The trajectories of PROMIS scores and gait parameters can be used to inform patients what they can expect after their surgery. This might provide them with a better understanding of how quickly different health aspects improve over time. In the patients with nonunion analyzed in the present study (n=7), the PROMIS score, and mainly the Pain Interference score, returned to preinjury levels more slowly between 6 months and one year after surgery.

Effects of Sex and Hypertension

In this study, men reported better physical health and lower pain interference than women. Similar sex differences were found in PROMIS scores of patients with joint arthroplasty [39] and populations with lupus erythematosus [40], neuromuscular disease, multiple sclerosis, postpolio syndrome, or spinal cord injury [41]. This finding is therefore in line with the literature and not specifically related to this particular patient collective.

Arterial hypertension had a negative effect on the Physical Health score. This finding is certainly related to the fact that arterial hypertension is usually associated with further comorbidities of the metabolic syndrome, including obesity, insulin resistance and diabetes, dyslipidemia [42], as well as in women with polycystic ovarian syndrome [43]. Similar to the sex difference found in the PROMIS results of this study, the finding that arterial hypertension was associated with a lower Physical Health score is not surprising.

Implications for Clinical Practice

In this study, the PROMIS trajectories provided information about the healing process of ankle or tibial fractures. However, the usability of these subjective measures to monitor fracture healing might be limited because of the potential influence of secondary gain. In clinical practice, patients may score questionnaires worse than they actually perceive them to receive longer paid sick leave or further rehabilitation prescriptions [44]. Since long-term limitations after work accidents in some countries result in permanent payments of disability pensions [45], to avoid abuse, the decision of whether problems persist must not only depend on questionnaire data but also apply more objective outcomes. Since it was clear to the patients in this study that the data were used for research only and that they had no effect on their treatment or potential secondary benefit, we expect that secondary gain played no or only a minor role in this study. However, the results from questionnaires in

XSL•FO
RenderX

patients with injuries need to be handled with care and, if indicated, combined with more objective measures to avoid abuse. Among such more objective measures, gait analyses certainly play an important role. Various technologies and products exist to perform gait analyses in patients with fracture [4].

Impact on the Field of Mobile Health

While the development of instrumented insoles has only recently reached a level that is suitable for daily use in a clinical environment, this study demonstrates a new benefit of this technology. The same data could also have been recorded with a ground-based force plate. However, insoles are inexpensive and with them, gait data can be recorded in any outpatient clinic or private practice, just as in settings where less financial resources are available. Another benefit of instrumented insoles over laboratory-based measurements is that they allow continuous monitoring of the daily lives of patients [5]. As patients after injury are initially unable to walk on a treadmill, especially if they still use walking aids, instrumented treadmills are less suitable for this purpose. Other than peak forces and acceleration data, the use of ground-reaction force-derived parameters recorded with instrumented insoles is a recent development with basic studies on the nature of the ground-reaction force curve obtained via insoles only having been reported recently [46-48]. The clinical usability to monitor and predict recovery based on these parameters has been demonstrated in several studies [4,5]. Instrumented insoles can also be used to provide immediate patient feedback on the weight bearing [49]. This study showed a potential benefit of this new technology for rehabilitation monitoring that should be followed up with more studies that compare traditional monitoring methods with instrumented insole monitoring. This way, the true clinical value of this new mobile health (mHealth) technology for clinical practice can be determined. It needs to be pointed out, that the development and evaluation of parameters derived from the raw data is of crucial importance. To be able to conduct this parameter development, large data sets from many patients are needed. Similar parameters may be derived from the data obtained by smart shoes and socks [50,51], which are further new wearable mHealth devices with great potential.

Strengths and Limitations

A limitation of this study is the incompleteness of the longitudinal data. The reason is that some patients did not return for follow-up visits because they either switched to a local orthopedic surgeon, had no or minor symptoms remaining (3 months after surgery), or were lost to follow-up for unknown reasons. This led to a reduction in sample size over time and a small sample size at 1 year post surgery. Another weakness is that countless gait parameters exist that have not been measured or analyzed in this study. The findings of this study are of course only valid for the few parameters reported. A strength of this study is that the follow-up time of both the PROMIS and gait measurements was one year. The PROMIS scores indicate that it takes up to a year for the scores to return to normal values. If patients who improved did not return, this may have led to an underestimation of the healing progression in the trajectories in our study.

Different types of fractures were pooled: proximal tibial, tibial shaft, distal tibial, and malleolar fractures. This is a limitation, but the authors expect only minor differences in gait and PROMIS between the different fracture types.

Conclusion

The gait parameters improved during the first 3 months post surgery, whereas the improvements in PROMIS scores lagged, with meaningful differences observed at approximately 6 to 9 months. Thus, the patients' perceptions improved later than their objective parameters assessed by instrumented insoles did. Pain perception was associated with the gait parameters, but only if they had already improved. Compared with women, men had better physical health and lower pain interference scores. Hypertension had a negative effect on the physical health score.

Acknowledgments

This research was funded by the Werner Siemens Foundation (project Smart Implants 2.0) and by the European Union under the Horizon Europe grant (101136376; project SmILE). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union nor the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor HaDEA can be held responsible for them.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables.
[DOCX File , 20 KB - mhealth_v13i1e71022_app1.docx]

References

1. Balu A, Gensler R, Liu J, Grady C, Brennan D, Cobourn K, et al. Single-center pilot study of remote therapeutic monitoring in patients with operative spinal pathologies. Clinical Neurology and Neurosurgery 2024 Jul;242:108346. [doi: 10.1016/j.clineuro.2024.108346]

XSL•FO
RenderX

2. Gilat R, Mitchnik IY, Patel S, Dubin JA, Agar G, Tamir E, et al. Pearls and pitfalls of PROMIS clinically significant outcomes in orthopaedic surgery. Arch Orthop Trauma Surg 2023;143(11):6617-6629. [doi: 10.1007/s00402-023-04983-y] [Medline: 37436494]
3. North K, Simpson GM, Stuart AR, Kubiak EN, Petelenz TJ, Hitchcock RW, et al. Early postoperative step count and walking time have greater impact on lower limb fracture outcomes than load-bearing metrics. Injury 2023;54(7):110756. [doi: 10.1016/j.injury.2023.04.043] [Medline: 37202224]
4. Warmerdam E, Orth M, Pohlemann T, Ganse B. Gait analysis to monitor fracture healing of the lower leg. Bioengineering (Basel) 2023;10(2):255 [FREE Full text] [doi: 10.3390/bioengineering10020255] [Medline: 36829749]
5. Warmerdam E, Wolff C, Orth M, Pohlemann T, Ganse B. Long-term continuous instrumented insole-based gait analyses in daily life have advantages over longitudinal gait analyses in the lab to monitor healing of tibial fractures. Front Bioeng Biotechnol 2024;12:1355254 [FREE Full text] [doi: 10.3389/fbioe.2024.1355254] [Medline: 38497053]
6. Becker HP, Rosenbaum D, Kriese T, Gerngross H, Claes L. Gait asymmetry following successful surgical treatment of ankle fractures in young adults. Clin Orthop Relat Res 1995(311):262-269. [Medline: 7634584]
7. Lajevardi-Khosh A, Bamberg S, Rothberg D, Kubiak E, Petelenz T, Hitchcock R. Center of pressure in a walking boot shifts posteriorly in patients following lower leg fracture. Gait Posture 2019;70:218-221. [doi: 10.1016/j.gaitpost.2019.03.010] [Medline: 30904788]
8. Bennett KJ, Millar SC, Fraysse F, Arnold JB, Atkins GJ, Solomon LB, et al. Postoperative lower limb joint kinematics following tibial plateau fracture: A 2-year longitudinal study. Gait Posture 2021;83:20-25. [doi: 10.1016/j.gaitpost.2020.10.005] [Medline: 33069125]
9. Kröger I, Müßig J, Brand A, Pätzold R, Wackerle H, Klöpfer-Krämer I, et al. Recovery of gait and function during the first six months after tibial shaft fractures. Gait Posture 2022;91:66-72 [FREE Full text] [doi: 10.1016/j.gaitpost.2021.09.199] [Medline: 34653876]
10. Thewlis D, Callary SA, Fraysse F, Solomon LB. Peak loading during walking is not associated with fracture migration following tibial plateau fracture: A preliminary case series. J Orthop Res 2015;33(9):1398-1406 [FREE Full text] [doi: 10.1002/jor.22905] [Medline: 25820829]
11. Houwen T, de Munter L, Lansink KWW, de Jongh MAC. There are more things in physical function and pain: A systematic review on physical, mental and social health within the orthopedic fracture population using PROMIS. J Patient Rep Outcomes 2022;6(1):34 [FREE Full text] [doi: 10.1186/s41687-022-00440-3] [Medline: 35384568]
12. Wong LH, Meeker JE. The promise of computer adaptive testing in collection of orthopaedic outcomes: An evaluation of PROMIS utilization. J Patient Rep Outcomes 2022;6(1):2 [FREE Full text] [doi: 10.1186/s41687-021-00407-w] [Medline: 34982280]
13. Ott NA, Barber CL, Jordan SE. Trends in patient-reported outcomes measurement information system (PROMIS) utilization in orthopedic sports medicine: A scoping review. Ochsner J 2023;23(4):304-314 [FREE Full text] [doi: 10.31486/toj.23.0054] [Medline: 38143549]
14. Bhashyam AR, Ochen Y, van der Vliet QM, Leenen LP, Hietbrink F, Houwert RM, et al. Association of patient-reported outcomes with clinical outcomes after distal humerus fracture treatment. JAAOS Glob Res Rev 2020;4(2):e19.00122. [doi: 10.5435/jaaosglobal-d-19-00122]
15. Gregory JJ, Werth PM, Reilly CA, Lucas AP, Bessen SY, Rode JB, et al. Trajectory of recovery in PROMIS global physical health after surgical fracture fixation. J Am Acad Orthop Surg 2021;30(3):e434-e443. [doi: 10.5435/jaaos-d-21-00308]
16. Rajan L, Eble S, Kim J, Kukadia S, Kumar P, Day J, et al. Risk factors associated with worse clinical outcomes of ankle fractures involving the posterior malleolus. Foot Ankle Orthop 2023;8(1):24730114231154217 [FREE Full text] [doi: 10.1177/24730114231154217] [Medline: 36860801]
17. Smith KS, Drexelius K, Challa S, Moon DK, Metzl JA, Hunt KJ. Outcomes following ankle fracture fixation with or without ankle arthroscopy. Foot Ankle Orthop 2020;5(1):2473011420904046 [FREE Full text] [doi: 10.1177/2473011420904046] [Medline: 35097364]
18. Tran SK, Yeager MT, Rutz RW, Mohammed Z, Johnson JP, Spitler CA. Resilience improves patient-reported outcomes after orthopaedic trauma. J Orthop Trauma 2024;38(5):e163-e168. [doi: 10.1097/BOT.0000000000002785] [Medline: 38506510]
19. Kellam PJ, Cutshall ZA, Dekeyser GJ, Rothberg DL, Higgins TF, Haller JM, et al. Recovery curve for patients with pilon fractures using patient-reported outcome measurement information system. Foot Ankle Int 2023;44(4):317-321. [doi: 10.1177/10711007231156424] [Medline: 36932665]
20. Rothrock NE, Kaat AJ, Vrahas MS, O Toole RV, Buono SK, Morrison S, et al. Validation of PROMIS physical function instruments in patients with an orthopaedic trauma to a lower extremity. J Orthop Trauma 2019;33(8):377-383. [doi: 10.1097/BOT.0000000000001493] [Medline: 31085947]
21. Website of the PROMIS Health Organization. URL: https://www.promishealth.org/ [accessed 2025-01-07]
22. Website HealthMeasures. URL: https://www.healthmeasures.net/ [accessed 2025-01-07]
23. Houwen T, Theeuwes HP, Verhofstad MHJ, de Jongh MAC. From numbers to meaningful change: Minimal important change by using PROMIS in a cohort of fracture patients. Injury 2023;54 Suppl 5:110882 [FREE Full text] [doi: 10.1016/j.injury.2023.110882] [Medline: 37923506]

XSL•FO
RenderX

24. Zhang W, Singh SP, Clement A, Calfee RP, Bijsterbosch JD, Cheng AL. Improvements in physical function and pain interference and changes in mental health among patients seeking musculoskeletal care. JAMA Netw Open 2023;6(6):e2320520 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.20520] [Medline: 37378984]
25. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. J R Stat Soc Ser C Appl Stat 2005;54:507-554. [doi: 10.1111/j.1467-9876.2005.00510.x]
26. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11(10):1305-1319. [doi: 10.1002/sim.4780111005] [Medline: 1518992]
27. Website of the statistics software R. URL: https://www.R-project.org/ [accessed 2025-01-07]
28. Sandal GM, van deVijver FJR, Smith N. Psychological hibernation in antarctica. Front Psychol 2018;9:2235 [FREE Full text] [doi: 10.3389/fpsyg.2018.02235] [Medline: 30524340]
29. Devkota HR, Bhandari B, Adhikary P. Perceived mental health, wellbeing and associated factors among Nepali male migrant and non-migrant workers: A qualitative study. J Migr Health 2021;3:100013 [FREE Full text] [doi: 10.1016/j.jmh.2020.100013] [Medline: 34405181]
30. Bretherton CP, Kotwal T, Sandhu H, Baird J, Griffin XL. Patient experiences and desires for recovery after ankle fracture surgery. Injury 2024;55(10):111763 [FREE Full text] [doi: 10.1016/j.injury.2024.111763] [Medline: 39106533]
31. McKeown R, Kearney RS, Liew ZH, Ellard DR. Patient experiences of an ankle fracture and the most important factors in their recovery: a qualitative interview study. BMJ Open 2020;10(2):e033539 [FREE Full text] [doi: 10.1136/bmjopen-2019-033539] [Medline: 32024789]
32. Allahabadi S, Fenn TW, Larson JH, Zhu D, Malloy P, Nho SJ. A smartphone app shows patients return to preoperative gait metrics 6 weeks after hip arthroscopy, and gait metrics have low to moderate correlations with a hip-specific patient-reported outcome measure. Arthrosc Sports Med Rehabil 2023;5(5):100779 [FREE Full text] [doi: 10.1016/j.asmr.2023.100779] [Medline: 37680367]
33. Bazett-Jones DM, Neal BS, Legg C, Hart HF, Collins NJ, Barton CJ. Kinematic and kinetic gait characteristics in people with patellofemoral pain: A systematic review and meta-analysis. Sports Med 2023;53(2):519-547. [doi: 10.1007/s40279-022-01781-1] [Medline: 36334239]
34. Burton W, Ma Y, Manor B, Hausdorff JM, Kowalski MH, Bain PA, et al. The impact of neck pain on gait health: a systematic review and meta-analysis. BMC Musculoskelet Disord 2023;24(1):618 [FREE Full text] [doi: 10.1186/s12891-023-06721-2] [Medline: 37516827]
35. Marriott KA, Birmingham TB, Leitch KM, Pinto R, Giffin JR. Strong independent associations between gait biomechanics and pain in patients with knee osteoarthritis. J Biomech 2019;94:123-129. [doi: 10.1016/j.jbiomech.2019.07.015] [Medline: 31421807]
36. Smith JA, Stabbert H, Bagwell JJ, Teng H, Wade V, Lee SP. Do people with low back pain walk differently? A systematic review and meta-analysis. J Sport Health Sci 2022;11(4):450-465 [FREE Full text] [doi: 10.1016/j.jshs.2022.02.001] [Medline: 35151908]
37. Miao B, Liu X, Zhu T. Automatic mental health identification method based on natural gait pattern. Psych J 2021;10(3):453-464. [doi: 10.1002/pchj.434] [Medline: 33569920]
38. Nagano H, Sarashina E, Sparrow W, Mizukami K, Begg R. General mental health is associated with gait asymmetry. Sensors (Basel) 2019;19(22):4908 [FREE Full text] [doi: 10.3390/s19224908] [Medline: 31717634]
39. Steinbeck V, Bischof AY, Schöner L, Langenberger B, Kuklinski D, Geissler A, et al. Gender health gap pre- and post-joint arthroplasty: identifying affected patient-reported health domains. Int J Equity Health 2024;23(1):44 [FREE Full text] [doi: 10.1186/s12939-024-02131-5] [Medline: 38413981]
40. Jiang TE, Mackey S, Darnall BD, Simard JF, Falasinnu T. The problem of pain in systemic lupus erythematosus: A comprehensive analysis of pain distribution using the CHOIR body map and PROMIS measures. Lupus 2025;34(1):47-56. [doi: 10.1177/09612033241301176] [Medline: 39584468]
41. Thakral M, Lacroix AZ, Molton IR. Sex/gender disparities in health outcomes of individuals with long-term disabling conditions. Rehabil Psychol 2019;64(2):221-228 [FREE Full text] [doi: 10.1037/rep0000248] [Medline: 30299140]
42. Peterseim CM, Jabbour K, Kamath Mulki A. Metabolic syndrome: An updated review on diagnosis and treatment for primary care clinicians. J Prim Care Community Health 2024;15:21501319241309168 [FREE Full text] [doi: 10.1177/21501319241309168] [Medline: 39714021]
43. Macut D, Mladenović V, Bjekić-Macut J, Livadas S, Stanojlović O, Hrnčić D, et al. Hypertension in polycystic ovary syndrome: Novel insights. Curr Hypertens Rev 2020;16(1):55-60. [doi: 10.2174/1573402115666190531071422] [Medline: 31146668]
44. Callison K, Pesko MF. The effect of paid sick leave mandates on coverage, work absences, and presenteeism. J Hum Resour 2022;57(4):1178-1208. [doi: 10.3368/jhr.57.4.1017-9124r2] [Medline: 35812986]
45. Farrants K, Alexanderson K. Sickness absence after work accidents and post-traumatic stress among white-collar workers in the retail and wholesale industry; a longitudinal Swedish cohort study. BMC Public Health 2024;24(1):2389 [FREE Full text] [doi: 10.1186/s12889-024-19865-0] [Medline: 39227810]

XSL•FO
RenderX

46. Wolff C, Steinheimer P, Warmerdam E, Dahmen T, Slusallek P, Schlinkmann C, et al. Effects of age, body height, body weight, body mass index and handgrip strength on the trajectory of the plantar pressure stance-phase curve of the gait cycle. Front Bioeng Biotechnol 2023;11:1110099 [FREE Full text] [doi: 10.3389/fbioe.2023.1110099] [Medline: 36873371]
47. Warmerdam E, Burger LM, Mergen DF, Orth M, Pohlemann T, Ganse B. The walking surface influences vertical ground reaction force and centre of pressure data obtained with pressure-sensing insoles. Front Digit Health 2024;6:1476335 [FREE Full text] [doi: 10.3389/fdgth.2024.1476335] [Medline: 39583997]
48. Wolff C, Steinheimer P, Warmerdam E, Dahmen T, Slusallek P, Schlinkmann C, et al. Characteristic changes of the stance-phase plantar pressure curve when walking uphill and downhill: Cross-sectional study. J Med Internet Res 2024;26:e44948 [FREE Full text] [doi: 10.2196/44948] [Medline: 38718385]
49. Ma T, Xu X, Chai Z, Wang T, Shen X, Sun T. A wearable biofeedback device for monitoring tibial load during partial weight-bearing walking. IEEE Trans. Neural Syst. Rehabil. Eng 2023;31:3428-3436. [doi: 10.1109/tnsre.2023.3305205]
50. Daugulis P, Kataševs A, Okss A. Estimation of the knee joint load using plantar pressure data measured by smart socks: A feasibility study. THC 2023;31(6):2423-2434. [doi: 10.3233/thc-235008]
51. Xia H, Xu J, Wang J, Hunt MA, Shull PB. Validation of a smart shoe for estimating foot progression angle during walking gait. J Biomech 2017;61:193-198. [doi: 10.1016/j.jbiomech.2017.07.012] [Medline: 28780187]

Abbreviations

BCCG: Box-Cox Cole Green
GAMLSS: generalized additive models for location, scale, and shape
IMU: inertial measurement unit
LMM: linear mixed-effects model
mHealth: mobile health
PROMIS: Patient-Reported Outcome Measurement Information System

Edited by L Buis; submitted 08.01.25; peer-reviewed by P Do, VB Semwal; comments to author 05.03.25; revised version received 26.03.25; accepted 02.04.25; published 16.06.25.

Please cite as:
Warmerdam E, Huebner M, Stoll C, Lange AI, Ganse B
Recovery of Patient-Reported Outcome Measures vs Gait Parameters Obtained by Instrumented Insoles After Tibial and Malleolar Fractures: Prospective Longitudinal Observational Study
JMIR Mhealth Uhealth 2025;13:e71022
URL: https://mhealth.jmir.org/2025/1/e71022
doi:10.2196/71022
PMID:

©Elke Warmerdam, Marianne Huebner, Caroline Stoll, Andrey Ivanovic Lange, Bergita Ganse. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of a Mobile Multidisciplinary Digital Therapeutics App for Patellofemoral Pain: Randomized Controlled Trial

Sanghee Lee[1*], MD; Chan Yoon[1*], MD, MS; Chi-Hyun Choi[1], MD, MS; Tae Hyun Park[1], MD; Sang Jin Yang[2], PhD; Ha Ri Cha[1], MPH; Tae Woo Kim[3], MD, PhD; Jae Hyeon Park[4,5*], MD, PhD; Moon Jong Chang[3*], MD, PhD; Chong Bum Chang[6], MD, PhD

[1]EverEx, Seoul, Republic of Korea

[2]Department of Health and Exercise Management, University of Tongwon, Kwangju, Republic of Korea

[3]Department of Orthopedic Surgery, Seoul National University College of Medicine, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

[4]Department of Rehabilitation Medicine, Hanyang University College of Medicine, Seoul, Republic of Korea

[5]Department of Rehabilitation Medicine, Hanyang University Guri Hospital, Guri-si, Republic of Korea

[6]Department of Orthopedic Surgery, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Gyeonggi-do, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Moon Jong Chang, MD, PhD
Department of Orthopedic Surgery
Seoul National University College of Medicine
SMG-SNU Boramae Medical Center
5 Gil 20
Boramae-road, Dongjak-gu
Seoul, 07061
Republic of Korea
Phone: 82 1041885264
Email: moonjongchang@gmail.com

Abstract

Background: Patellofemoral pain (PFP) is a common musculoskeletal disorder characterized by persistent knee pain, often without any structural abnormalities. Conservative treatments, particularly exercise therapy, are widely recommended; however, adherence remains generally low, and full recovery is often not achieved. Psychological interventions can aid in symptom management; however, studies integrating cognitive behavioral therapy (CBT), which is known to be effective for chronic pain, with exercise therapy for patients with PFP are limited. This study examined the impact of MORA Cure (PFP), a multidisciplinary digital therapeutics (DTx) app that integrates exercise and CBT, in comparison with conventional treatments for PFP management.

Objective: This study aimed to evaluate the efficacy and safety of an 8-week DTx intervention incorporating exercise and CBT compared with in-person exercise education in patients with PFP.

Methods: A parallel-group randomized controlled trial was conducted with 35 patients diagnosed with PFP recruited from orthopedic outpatient clinics. Participants were randomly assigned to either the DTx group (n=18, 51%) or the control group (n=17, 49%). The DTx group received an 8-week intervention via the MORA Cure (PFP) app incorporating home-based exercises and weekly CBT modules with daily worksheets. The control group received conventional treatment, including disease education, a single in-person exercise education session conducted by a medical professional, and encouragement to continue self-exercising throughout the study period. The outcome measures included pain severity (usual and worst, assessed using the numeric pain rating scale), functional disability (Anterior Knee Pain Scale), knee extension strength (measured using an isokinetic dynamometer), health-related quality of life (EQ-5D), and mental health status (9-item Patient Health Questionnaire). Assessments were conducted from baseline at 4-week intervals for up to 12 weeks.

Results: The DTx group showed significant reductions in usual pain at each time point (4 weeks: mean score 2.2, SD 1.5, and P=.006; 8 weeks: mean 2.3, SD 1.7, and P=.003; 12 weeks: mean 1.2, SD 1.8, and P=.008), whereas the control group exhibited no changes. The knee extension strength in the DTx group increased significantly at both 8 and 12 weeks (P<.001), with greater improvement than that in the control group at 8 weeks (P=.04). Both groups showed significant improvements in functional

XSL•FO
RenderX

disability at 12 weeks (DTx: mean score 85.2, SD 12.7, and *P*=.006; control: mean 84.5, SD 13.0, and *P*=.01). Health-related quality of life (EQ-5D) also improved in the DTx group at 8 and 12 weeks, whereas the control group showed improvement only at 12 weeks.

Conclusions: This multidisciplinary DTx intervention was associated with significant pain reduction, improved functional disability, and increased knee extension strength in patients with PFP. These findings underscore the promise of DTx in PFP management and their potential to enhance patient outcomes.

Trial Registration: ClinicalTrials.gov NCT05614583; https://clinicaltrials.gov/study/NCT05614583

(JMIR Mhealth Uhealth 2025;13:e69627) doi:10.2196/69627

KEYWORDS

patellofemoral pain; knee pain; app; smartphone; digital therapeutics; exercise; cognitive behavioral therapy; mobile health; mHealth; eHealth; digital health

Introduction

Background

Patellofemoral pain (PFP) is one of the most common musculoskeletal disorders causing knee pain, with an annual prevalence of 23% in the general population [1,2]. Although patients with PFP do not exhibit significant structural abnormalities, symptoms often persist for more than 2 years without spontaneous resolution [3,4]. Furthermore, PFP may contribute to the development of patellofemoral osteoarthritis [5-7]. Therefore, active management of PFP is imperative.

Conservative interventions are generally recommended for managing PFP, with therapeutic exercise targeting the hip and lower-extremity muscles being considered the primary intervention [1,7,8]. Despite these recommendations, a considerable proportion of patients in clinical practice exhibit poor adherence to exercise therapy and often fail to achieve complete perceived recovery even after undergoing exercise [9-11]. In addition, some studies have explored the efficacy of psychological interventions in managing PFP because of the association between psychological factors and pain [12-14]. However, few studies have examined the use of cognitive behavioral therapy (CBT), which has demonstrated beneficial effects on chronic pain [15], in combination with exercise therapy for patients with PFP. Therefore, research into interventions that can enhance the effectiveness of exercise therapy and incorporate CBT is required.

In recent years, digital therapeutics (DTx) have emerged as promising treatment options for managing various musculoskeletal conditions [16-20]. A previous meta-analysis demonstrated that DTx enhanced adherence to therapeutic exercises in patients with musculoskeletal disorders [21]. Furthermore, DTx can offer not only therapeutic exercise but also a variety of treatments, including exercise, education, and CBT. Multidisciplinary DTx have been shown to be effective in treating chronic low back and shoulder pain [18,20,22].

Objectives

However, to our knowledge, research on the effectiveness of multidisciplinary DTx interventions in patients with PFP is lacking. Therefore, this study aimed to investigate the efficacy of an 8-week multidisciplinary mobile app that provides therapeutic exercise, education, and CBT compared to in-person detailed exercise education in patients with PFP in a randomized controlled trial (RCT). We hypothesized that this digital intervention, compared to conventional in-person exercise education, would lead to symptom improvement in patients with PFP after a 12-week follow-up period.

Methods

Study Design and Randomization

This study was designed as a parallel-group, individual RCT to evaluate the safety and efficacy of a multidisciplinary DTx app, MORA Cure (PFP), compared with conventional treatment for PFP. The trial was conducted at the Department of Orthopedics, Seoul National University Bundang Hospital, and Boramae Medical Center. The study protocol, along with the statistical analysis plan, has been previously published and registered at ClinicalTrials.gov (NCT05614583).

The participants were recruited from outpatient clinics between November 2022 and July 2023. Study-related posters were displayed in clinic waiting areas, and individuals who responded to the information provided in the posters were screened by the research team for eligibility. Those who met the inclusion criteria and provided written informed consent were enrolled in the study. Participants were randomly assigned in a 1:1 ratio to either the DTx group, which received the MORA Cure (PFP) DTx intervention, or the control group, which received conventional treatment. Randomization was performed using a computer-generated block randomization method. An independent statistician generated the randomization sequence, and allocation concealment was ensured using sealed, opaque envelopes. To minimize bias, the outcome assessors were blinded to group allocation.

Study Participants

The inclusion criteria for this study were as follows: adults aged 19 to 49 years with a history of PFP lasting for at least 3 months but less than 2 years affecting at least one knee joint. The diagnosis of PFP was made by orthopedic surgeons based on clinical symptoms, including anterior or retropatellar knee pain, in conjunction with physical examination findings and plain radiographic imaging to exclude structural abnormalities. Standardized anteroposterior weight-bearing knee radiographs with approximately 10° of flexion were obtained to rule out conditions such as osteoarthritis or bony lesions. If both knees

XSL•FO
RenderX

were affected, the knee with the highest pain score on the numeric pain rating scale (NPRS) was selected as the target joint for evaluation; this selection was maintained throughout the trial. The participants were also required to report pain during squatting and at least 2 of the following activities: prolonged sitting, cycling, running, stair climbing, kneeling, patellar compression, and palpation of the patellar facets.

The exclusion criteria were knee osteoarthritis classified as Kellgren-Lawrence grade 2 or higher on radiographs, history of significant knee injuries (including fractures, ligament injuries, cartilage damage, or patellar dislocation), or previous knee surgery. Participants diagnosed with patellar tendinitis or structural knee abnormalities (such as Osgood-Schlatter disease), who had conditions that could limit physical activity, who were pregnant or breastfeeding, who used pain or anti-inflammatory medications for other conditions, and who had a history of substance or alcohol abuse were also excluded. Participants with psychiatric conditions that could affect pain perception, such as somatoform disorders, were excluded.

During the trial, intra-articular injections, surgery, narcotic analgesics, topical agents, and any treatment intended to alleviate joint pain outside the trial protocol were prohibited. However, acetaminophen (650 mg) was allowed as a rescue medication at a dosage of 2 tablets every 8 hours (up to a maximum of 6 tablets per day). The participants were instructed to avoid acetaminophen within 24 hours before each visit and until all assessments were completed.

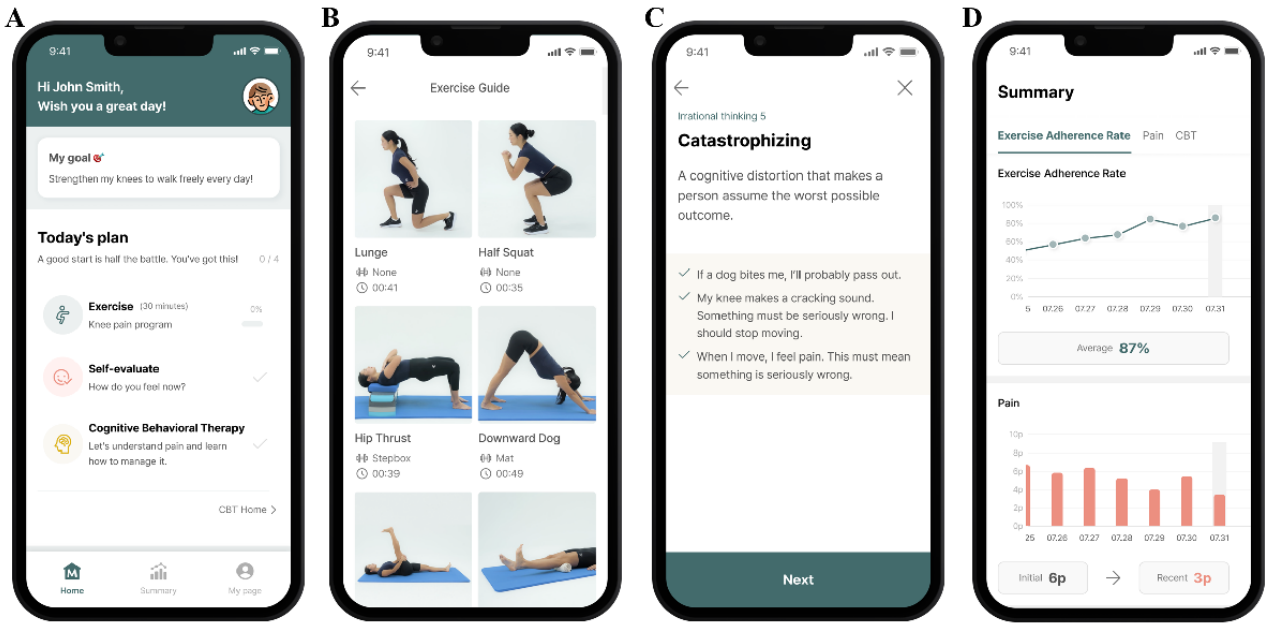
Ethical Considerations

This study was approved by the institutional review boards of Seoul National University Bundang Hospital (E-2209-780) and Boramae Medical Center (10-2023-33). This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Written informed consent was obtained from all participants before enrollment, and the investigators ensured adherence to the trial procedures and protocol. Participants were informed of their right to voluntarily participate and withdraw from the study at any time without penalty. All data were deidentified before analysis and stored in password-protected files accessible only to the research team in accordance with institutional data protection policies. No personally identifiable information was used in the analysis or dissemination of results. Participants received a transportation allowance of 50,000 KRW (approximately US $36) at each visit. No additional financial compensation was offered for participation.

Intervention and Control Groups

The DTx group used the MORA Cure (PFP) DTx app for 8 weeks. Participants were provided with free access to the app as part of the study and were guided by the research team during enrollment. Figure 1 shows a schematic of the MORA Cure (PFP) multidisciplinary DTx model. The intervention included a structured home-based exercise program incorporating combined hip- and knee-targeted muscle strengthening and stretching exercises as recommended by the clinical guidelines for PFP [23,24]. The exercises were designed to progressively strengthen the knee and hip muscles, with intensity adjusted according to the participants' reported pain levels as assessed using the Borg rating of perceived exertion scale and the NPRS. Participants completed daily exercise sessions lasting approximately 30 minutes.

Figure 1. Multidisciplinary digital therapeutics MORA Cure (PFP), incorporating personalized exercise therapy and self-guided cognitive behavioral therapy for patellofemoral pain.



In addition to exercise therapy, the participants received weekly CBT modules. These modules covered topics such as identifying and managing negative emotions and pain-related thoughts, developing positive coping strategies, and practicing relaxation techniques. Throughout with the 8 weekly modules, participants were required to complete daily worksheets to reinforce their understanding and practice of the CBT content.

XSL•FO
RenderX

Following the 8-week treatment period, the exercise therapy materials remained accessible via the app for continued use. The participants were encouraged to continue their exercise routine throughout the observation period.

The control group received conventional treatment, including disease education materials and a 30-minute, in-person exercise education session with a medical professional specializing in musculoskeletal disorders. The participants were provided with educational materials outlining the instructed content and were encouraged to perform rehabilitation exercises independently at home.

Outcome Measures

The primary outcomes of this study were changes in the usual pain severity and functional disability. Usual pain severity was assessed using the NPRS, in which the participants rated their average daily pain over the previous week on a scale from 0 (no pain) to 10 (worst pain imaginable). The second primary outcome, functional disability, was measured using the Anterior Knee Pain Scale (AKPS), also known as the Kujala score. This disease-specific disability scale ranges from 0 (severe disability) to 100 (no disability) and assesses the limitations specific to PFP. Outcome measures were collected through in-person evaluations conducted at the clinical sites.

Secondary outcomes included assessments of the worst pain experienced in the previous week and pain during specific activities, such as running, sit-to-stand transitions, stair ascent, and stair descent. Participants' health-related quality of life (QoL) was measured using the EQ-5D. Psychological outcomes were assessed using the 9-item Patient Health Questionnaire (PHQ-9) to evaluate depressive symptoms. The knee extensor strength was measured using a Biodex isokinetic dynamometer (Biodex Medical Systems, Inc). Usual pain, worst pain, functional disability, and health-related QoL were assessed at baseline and at 4, 8, and 12 weeks. Knee extensor strength and psychological outcomes were assessed at baseline and 8 and 12 weeks.

In addition to the primary and secondary outcomes, baseline demographic and structural alignment variables, including age, sex, BMI, pain duration, and the hip-knee-ankle (HKA) angle, were collected at enrollment. The HKA angle, defined as the angle formed by lines connecting the centers of the femoral head, knee joint, and talus, was measured using standardized weight-bearing full-length anteroposterior radiographs [25]. Malalignment in the coronal plane, as quantified by the HKA angle, has been associated with patellofemoral joint degeneration and instability and may play a role as a biomechanical contributor to PFP [26-28].

Compliance and Safety Monitoring

Treatment compliance was monitored and compared between the DTx and control groups. In the DTx group, treatment compliance was evaluated at weeks 4 and 8 based on the actual time participants spent exercising, as recorded by the app, relative to the prescribed target time. In the control group, the participants documented their daily exercise completion and session duration in a log, which was reviewed at each visit to monitor adherence throughout the study. Participants were advised to complete 30 minutes of exercise 6 times per week. Compliance in the control group was evaluated based on the total weekly exercise time relative to the recommended exercise duration. Safety was monitored throughout the study by recording adverse events or symptom exacerbations reported by participants. At each visit (weeks 4, 8, and 12), participants were asked about any new or worsening symptoms related to the intervention and were encouraged to report adverse events at any time during the trial.

Sample Size

Estimating the target sample size for this trial was challenging due to the lack of previous studies comparing the efficacy of DTx integrating both exercise and CBT for PFP. To guide the calculation, we referred to a previous prospective study that evaluated the effectiveness of in-person therapy combining exercise therapy with psychological intervention in a similar patient population [13]. On the basis of this study's results, where the primary outcome of usual pain showed a significant difference between groups (mean difference −6.5, 95% CI −10.9 to −2.1; P=.001; effect size=1.11), we calculated the sample size.

Sample size calculation was conducted using the G*Power software with the following parameters: type I error (α) of .05, 80% power, and the effect size derived from a previous study. This calculation yielded the required sample size of 14 participants per group. Considering previous studies with an 8-week intervention period and accounting for a potential dropout rate of up to 30%, we aimed to recruit a total of 40 participants, with 20 in each group [29].

Statistical Analysis

The primary analysis was conducted on the per-protocol population, which comprised 13 participants per group (DTx and control). Statistical analyses were conducted using the R software (version 4.4.1; R Foundation for Statistical Computing).

Baseline characteristics were compared between the groups using appropriate statistical tests to assess differences. For continuous variables, either the 2-tailed t test or Mann-Whitney U test was used depending on the normality of the data distribution. As sex was the only categorical variable, the Fisher exact test was used to assess between-group differences. Between-group comparisons of change scores (eg, baseline to 8 weeks) were conducted using the Mann-Whitney U test.

Within-group differences from baseline to the postintervention time point were assessed using the Friedman test for repeated measures. Post hoc pairwise comparisons were conducted using the Wilcoxon signed rank test, with the Bonferroni correction applied to adjust the significance threshold for multiple comparisons.

A P value of <.05 was considered statistically significant for all analyses. When the Bonferroni correction was applied to account for multiple comparisons, the significance threshold was adjusted by dividing the conventional α level (.05) by the number of comparisons, controlling for familywise errors, and reducing the risk of type I errors.

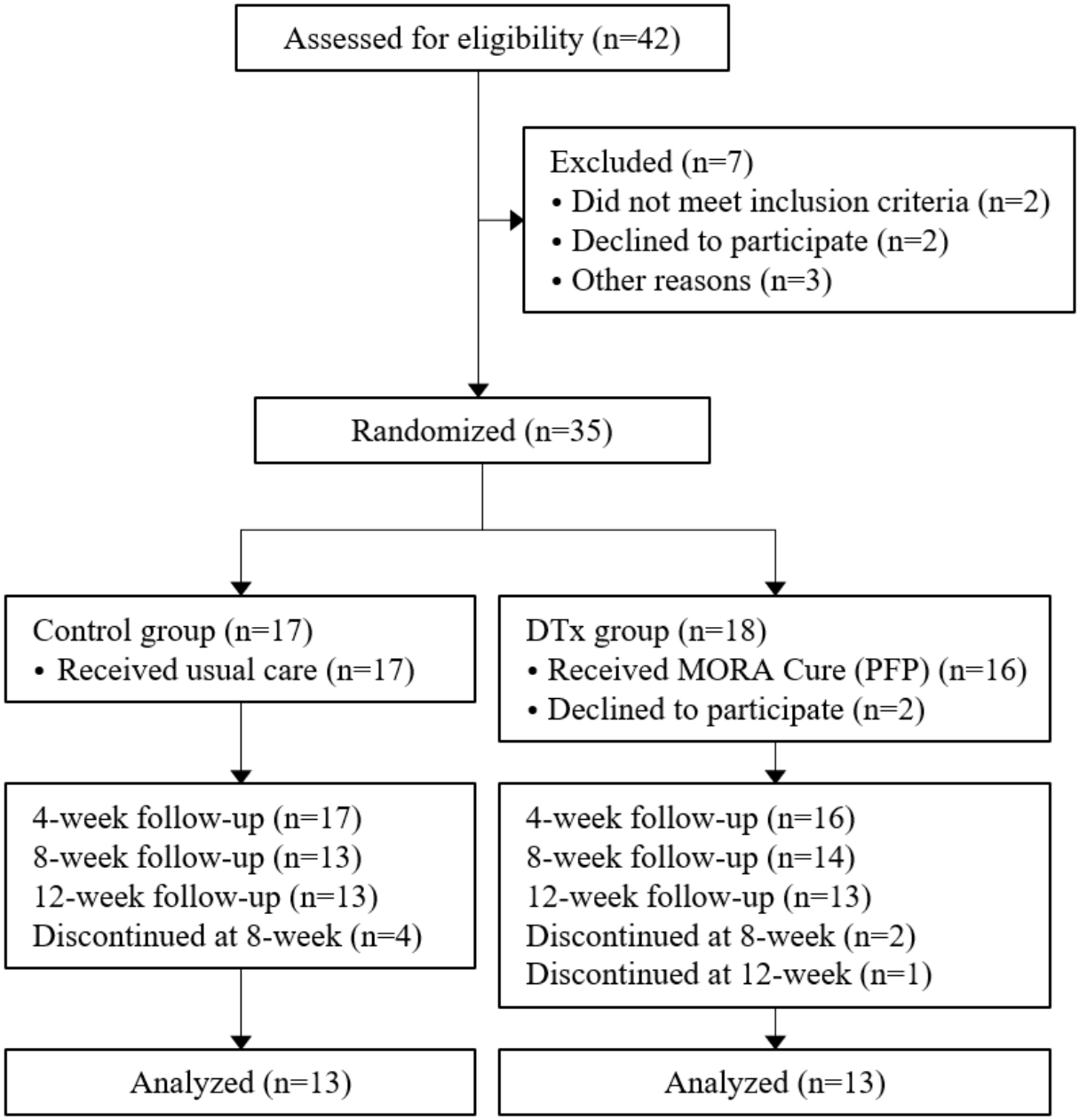
XSL•FO
RenderX

Results

Study Population

Of the 42 individuals screened for eligibility, 7 (17%) were excluded: 2 (29%) for not meeting the inclusion criteria (one due to PFP duration and the other due to knee osteoarthritis); 2 (29%) for declining participation; and 3 (43%) for other reasons, including vulnerability and follow-up issues (Figure 2). This resulted in 35 participants randomized into the control (n=17, 49%) and DTx (n=18, 51%) groups. In the DTx group, of the 18 participants, 16 (89%) began the intervention, with 2 (11%) declining to participate after randomization. By the 12-week follow-up, 72% (13/18) of the participants remained in the DTx group for analysis after 3 withdrawals. In the control group, all 17 participants remained up to the 4-week follow-up; however, 4 (24%) discontinued their participation by the 8-week follow-up due to consent withdrawal and worsening knee pain, resulting in 13 (76%) participants for the final analysis. Ultimately, 13 participants in each group completed all the assessments over the 12-week period.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram. DTx: digital therapeutics.



A total of 26 participants (n=13, 50% per group) were included in the final analysis. The baseline characteristics of the 2 groups were largely comparable across key clinical variables, including age, sex, BMI, HKA angle, pain scores, and knee function, except for PHQ-9 scores (Table 1). PHQ-9 scores, which assess depressive symptoms, were significantly higher in the control group than in the DTx group (P=.02). Participants were mostly women in their mid-30s, and the average duration of knee pain was approximately 1 year (DTx group: mean 48.4, SD 22.0 weeks; control group: mean 49.7, SD 26.3 weeks). The mean usual pain score at baseline was 3.9 (SD 2.0) in the DTx group

XSL•FO
RenderX

and 3.7 (SD 1.5) in the control group, whereas the mean worst pain score was 4.8 (SD 2.5) and 5.2 (SD 1.9), respectively.

Table 1. Baseline characteristics and symptom-related variables of the study participants[a].

Variable	DTx[b] group (n=13)	Control group (n=13)	*P* value
Age (y), mean (SD)	33.6 (11.0)	34.8 (10.9)	.76
Sex (female), n (%)	11 (84.6)	9 (69.2)	.65
BMI (kg/m^2), mean (SD)	21.3 (1.9)	21.3 (2.0)	.80
HKA[c] angle (°), mean (SD)	−0.6 (3.4)	−0.7 (3.8)	.92
Usual pain score (0-10), mean (SD)	3.9 (2.0)	3.7 (1.5)	.75
Worst pain score (0-10), mean (SD)	4.8 (2.5)	5.2 (1.9)	.98
Duration of pain (wk), mean (SD)	48.4 (22.0)	49.7 (26.3)	.84
AKPS[d] score (0-100), mean (SD)	74.5 (11.8)	73.5 (12.8)	.84
Knee extension peak torque (Biodex isokinetic dynamometer; Nm), mean (SD)	78.0 (19.4)	68.5 (32.2)	.09
EQ-5D score (–0.07 to –1.00), mean (SD)	0.8 (0.05)	0.8 (0.05)	.37
PHQ-9[e] score (0-27), mean (SD)	1.5 (2.2)	3.8 (3.4)	.02

[a]*P* values for continuous variables, including age, BMI, worst pain score, pain duration, knee extension peak torque (measured using a Biodex isokinetic dynamometer), EQ-5D score, and PHQ-9 score, were calculated using the Mann-Whitney *U* test, whereas *P* values for HKA angle, usual pain score, and AKPS score were calculated using the Student *t* test. The *P* value for the sex variable was calculated using the Fisher exact test.

[b]DTx: digital therapeutics.

[c]HKA: hip-knee-ankle.

[d]AKPS: Anterior Knee Pain Scale.

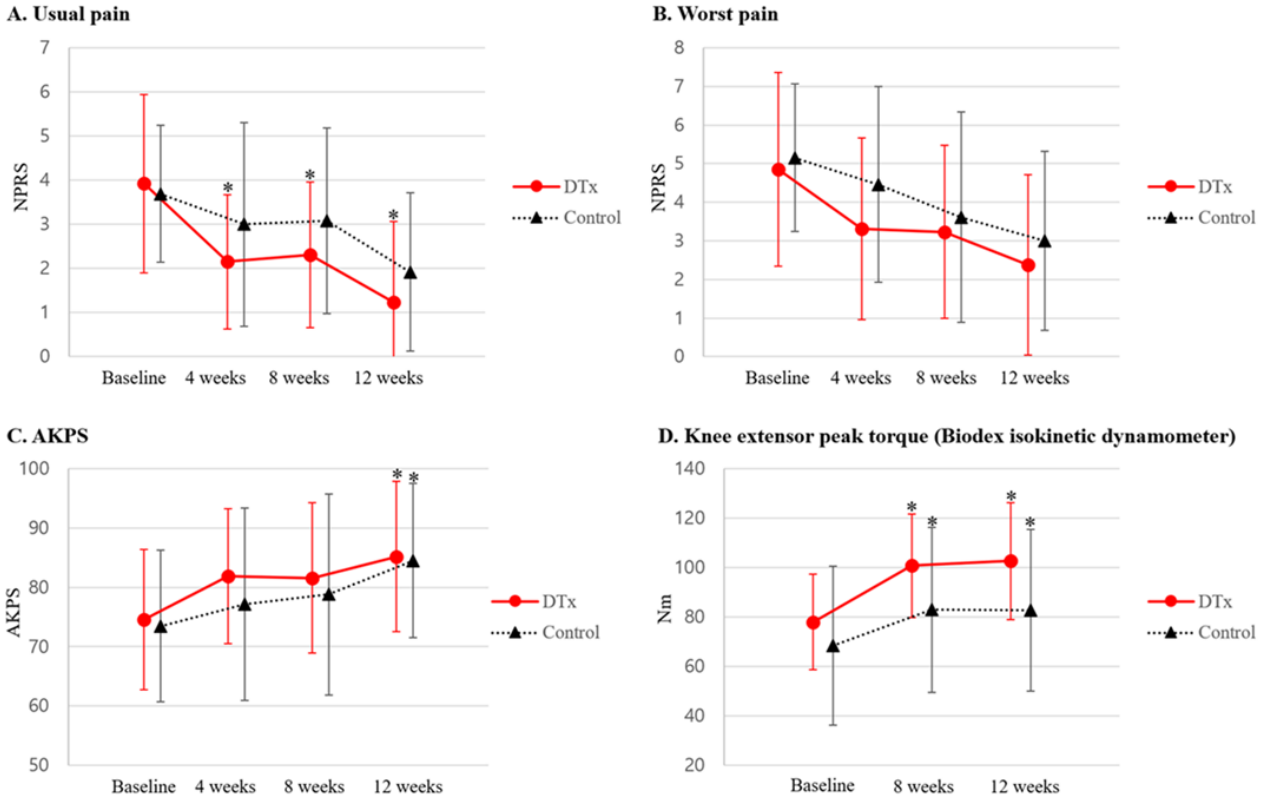
[e]PHQ-9: 9-item Patient Health Questionnaire.

Outcome Measures

Figure 3 illustrates the changes in usual and worst pain, functional disability, and knee extension strength. After applying Bonferroni correction, statistical significance was set at $P<.02$ for usual pain, worst pain, and AKPS and at $P<.03$ for knee extension peak torque. The DTx group showed a statistically significant reduction in usual pain from baseline at all time points. Specifically, usual pain decreased significantly at 4 (mean 2.2, SD 1.5; $P=.006$), 8 (mean 2.3, SD 1.7; $P=.003$), and 12 (mean 1.2, SD 1.8; $P=.008$) weeks in the DTx group. In contrast, the control group did not show significant reductions in pain at 4 (mean 3.0, SD 2.3; $P=.15$), 8 (mean 3.1, SD 2.1; $P=.32$), or 12 (mean 1.9, SD 1.8; $P=.03$) weeks.

XSL•FO
RenderX

Figure 3. Comparison of clinical outcomes between the digital therapeutics (DTx) and control groups over 8 weeks of treatment. AKPS: Anterior Knee Pain Scale; NPRS: numerical pain rating scale.



For the worst pain, both groups showed a gradual reduction over time; however, after Bonferroni correction, this reduction was not statistically significant from baseline, although the DTx group exhibited borderline significance. In the DTx group, pain reduction was observed at 4 (mean 3.3, SD 2.4; *P*=.03), 8 (mean 3.2, SD 2.2; *P*=.03), and 12 (mean 2.4, SD 2.3; *P*=.04) weeks. The control group showed reductions at 4 (mean 4.5, SD 2.5; *P*=.27), 8 (mean 3.6, SD 2.7; *P*=.12), and 12 (mean 3.0, SD 2.3; *P*=.049) weeks; however, these were not statistically significant.

AKPS scores showed gradual improvement in both groups over time, with significant improvements from baseline observed only at 12 weeks in the DTx (mean 85.2, SD 12.7; *P*=.006) and control (mean 84.5, SD 13.0; *P*=.01) groups. For pain and knee function, between-group comparisons of change scores from baseline showed no statistically significant differences in usual pain (*P*=.19, .69, and .36), worst pain (*P*=.55, .80, and .78), or AKPS scores (*P*=.36, .92, and .72) at 4, 8, and 12 weeks, respectively.

In the DTx group, knee extension peak torque increased significantly from baseline at both 8 (mean 100.8, SD 20.9; *P*<.001) and 12 (mean 102.7, SD 23.7; *P*<.001) weeks. The control group also exhibited significant improvements at 8 (mean 82.9, SD 33.5; *P*=.01) and 12 (mean 82.8, SD 32.7; *P*=.02) weeks. Between-group comparisons in change from baseline revealed that the DTx group showed significantly greater improvements at 8 weeks (*P*=.04), with borderline significance at 12 weeks (*P*=.05).

Table 2 summarizes the changes in pain scores during stair activities, sit-to-stand transitions, and running, as well as the EQ-5D and PHQ-9 scores over time for each group. In the DTx group, pain during stair ascent decreased significantly at both 4 and 12 weeks compared to baseline, whereas stair descent pain showed consistent reductions at all time points. Pain during running was significantly lower at 8 weeks in the DTx group, whereas the control group showed a significant reduction at both 4 and 12 weeks. Regarding health-related QoL, the EQ-5D scores in the DTx group improved significantly at 8 and 12 weeks, whereas the control group showed a similar significant improvement only at 12 weeks. The PHQ-9 scores were significantly reduced in the control group at 12 weeks, with no corresponding changes in the DTx group.

XSL•FO
RenderX

Table 2. Comparison of baseline and posttreatment outcomes in the digital therapeutics (DTx) and control groups at multiple time points[a].

Variable and group	Baseline, mean (SD)	4 wk, mean (SD)	8 wk, mean (SD)	12 wk, mean (SD)
Pain during stair ascent (0-10)				
DTx	3.7 (2.4)	2.3 (2.0)	2.2 (2.0)	1.3 (1.6)
P value	—[b]	.01	.03	.008
Control	3.4 (2.1)	2.5 (2.1)	2.2 (1.8)	1.4 (1.4)
P value	—	.19	.12	.02
Pain during stair descent (0-10)				
DTx	4.0 (2.2)	2.5 (2.1)	2.1 (1.6)	1.7 (1.5)
P value	—	.007	.006	.008
Control	4.0 (2.1)	3.2 (2.3)	2.8 (2.4)	2.2 (2.4)
P value	—	.17	.09	.03
Pain during sit-to-stand transition (0-10)				
DTx	2.8 (2.6)	1.9 (1.9)	1.3 (1.4)	0.8 (1.3)
P value	—	.13	.03	.02
Control	3.7 (2.3)	2.2 (2.3)	2.4 (1.9)	1.4 (1.8)
P value	—	.07	.09	.02
Pain during running (0-10)				
DTx	4.5 (2.7)	2.9 (2.2)	2.5 (1.7)	2.2 (2.0)
P value	—	.03	.009	.03
Control	4.7 (2.0)	3.5 (2.0)	3.1 (2.1)	2.5 (2.1)
P value	—	.01	.03	.01
EQ-5D (–0.07 to 1.00)				
DTx	0.84 (0.05)	0.86 (0.04)	0.87 (0.04)	0.88 (0.04)
P value	—	.04	.01	.01
Control	0.82 (0.05)	0.85 (0.05)	0.85 (0.04)	0.88 (0.04)
P value	—	.02	.19	.009
PHQ-9[c] (0-27)				
DTx	1.5 (2.2)	—	1.7 (2.4)	1.5 (1.6)
P value	—	—	>.99	.80
Control	3.8 (3.4)	—	2.5 (3.2)	1.4 (1.4)
P value	—	—	.07	.003

[a]Wilcoxon signed rank tests with Bonferroni corrections were applied for post hoc pairwise comparisons. The adjusted significance level was set at *P*<.02 (.05/3) except in the case of the PHQ-9, where the significance level was *P*<.03 (.05/2).

[b]Baseline values served as the reference for within-group comparisons; therefore, no *P* values were calculated for baseline columns. The PHQ-9 was measured only at baseline, 8 weeks, and 12 weeks; thus, data were not available for the 4-week time point.

[c]PHQ-9: 9-item Patient Health Questionnaire.

Compliance and Safety Monitoring

At the 4-week assessment, average treatment compliance was 81.4% (SD 24.1%) and 83.7% (SD 74.8%) in the DTx and control groups, respectively (*P*=.84); at 8 weeks, average compliance was 61.9% (SD 29.9%) and 58.4% (SD 42%) in the DTx and control groups, respectively (*P*=.81). No significant differences were observed in compliance between groups at either time point.

No serious adverse events were observed. Mild musculoskeletal symptoms, such as arthralgia, extremity pain, joint stiffness, and muscle spasms, were observed as adverse device effects, with no statistically significant differences between the groups. These symptoms occurred in 15% (2/13) of the participants in the DTx group and in none in the control group, and all events were resolved without complications.

XSL•FO
RenderX

Discussion

Principal Findings

The effectiveness of an 8-week multidisciplinary DTx intervention was evaluated in patients with PFP and compared with usual care. This study demonstrated that the DTx app improved pain, functional disability, QoL, and knee extensor strength. Although functional disability, QoL, and knee extensor strength also improved in the control group, significant pain reduction was not observed. However, no significant differences were observed between the 2 groups except for knee extensor strength at week 8.

The results of this study were consistent with those of previous studies demonstrating the efficacy of digital health interventions on musculoskeletal disorders, with a positive impact observed following their application [30,31]. Among these, home-based physical exercise programs were the most commonly used interventions, followed by education, CBT, and monitoring of outcomes or physical activity [30]. A study that implemented a multidisciplinary DTx intervention with the same components (combining therapeutic exercise and CBT) as those used in this study for patients with chronic low back pain also observed reductions in pain and functional disability, although the differences were not statistically significant compared to the control group [32]. However, in previous studies, digital interventions for the knee have primarily been applied to knee osteoarthritis or postarthroplasty rehabilitation, whereas the application of multidisciplinary DTx interventions to PFP remains limited.

For PFP, exercise is the key component of treatment [8,33,34]. Although the optimal exercise for PFP has not yet been determined, combining hip and knee exercises has been reported to be more effective than knee exercises alone [34,35]. A previous meta-analysis reported pain reduction with a mean difference of −1.71 on a 10-point visual analogue scale in patients with PFP following hip and knee exercise programs [35]. In this study, the mean difference in usual pain improved significantly by −1.6 and −2.7 after 8 and 12 weeks, respectively, in the DTx group. The hip and knee exercise studies included in the aforementioned meta-analysis implemented supervised exercise [35]. Therefore, the effect of DTx on reducing usual pain in this study appears comparable to that of supervised exercise. The proposed physiological mechanisms of exercise include reducing pressure on the knee through improved leg strength, promoting tissue adaptation through appropriate loading, and enhancing movement control [24,36]. Although there was no significant difference between the groups, the control group did not show a significant improvement in usual pain. The control group also received 30 minutes of in-person exercise education, and their compliance was adequate. This finding is consistent with those of previous studies that demonstrated that supervised progressive exercise was more effective [37,38]. Adjusting exercise intensity and type using DTx rather than simply performing the same prescribed exercises may have contributed to better pain outcomes.

Patients with PFP often exhibit impaired physical function [39]. Exercise has positive effects on physical disability and strength [34,35]. Although functional disability and pain intensity have been reported to have a moderate relationship [40], both groups showed significant improvement in functional disability by week 12, with no significant difference between the groups. However, the control group did not experience a significant pain reduction. These findings are consistent with those of a previous study that showed that, although there was a significant long-term difference in pain improvement between the control group (who only performed isometric quadriceps contractions and received education) and supervised exercise therapy group, long-term differences in functional disability were not observed [9]. In addition, in terms of knee extensor strength, which is an objectively measured physical function, the DTx group in our study demonstrated significantly greater improvement than the control group by week 8.

Mental distress, including depression, anxiety, and pain catastrophizing, is prevalent among patients with PFP and is often associated with increased pain, impaired physical function, and poor outcomes [12,41,42]. In addition, a previous meta-analysis reported that patients with PFP have a reduced QoL [43]. However, the efficacy of psychological interventions in patients with PFP remains inconclusive [13,14,29]. CBT, one of the most studied psychological approaches in chronic pain management, has shown efficacy in altering maladaptive pain-related thoughts and reducing pain-related avoidant behaviors [15]. These effects are thought to occur via mechanisms such as reduced pain catastrophizing and increased pain coping strategies. In this study, despite randomization, baseline PHQ-9 scores differed between the groups; however, most participants were clinically categorized as having minimal or no depression (score of 0-4; DTx group: 12/13, 92%; control group: 9/13, 69%) [44]. Furthermore, improvement in PHQ-9 scores was observed only in the control group, which may be attributed to baseline differences. Given the potential benefits of psychological interventions, future studies are required to evaluate the effectiveness of DTx for patients with musculoskeletal disorders, particularly for patients with PFP experiencing mental distress.

Given the advances in digital and communication technologies, DTx have been investigated for various musculoskeletal disorders [18,20,45,46]. DTx could enhance overall health outcomes, accessibility, adherence, and cost-effectiveness [21,47,48]. Previous studies have evaluated the effectiveness of digital health technology in knee osteoarthritis and other nonoperative knee conditions and reported positive effects on pain, physical function, and QoL [46,49]. However, RCTs investigating the effects of multidisciplinary DTx, specifically in patients with PFP, are still lacking. Considering the disease characteristics of PFP and our results, multidisciplinary DTx appear promising for the treatment of PFP.

Limitations

This study has several limitations. First, although the sample size calculation accounted for a 30% dropout rate, the final per-protocol analysis included only 13 participants per group (reflecting a 5/40, 13% prerandomization dropout rate and a

XSL•FO
RenderX

9/35, 26% postrandomization dropout rate) due to additional attrition, which may have reduced the statistical power and limited the generalizability of the findings. Second, participant blinding was not feasible due to the nature of the intervention. Third, many participants reported low baseline pain and PHQ-9 scores, which could have influenced their responsiveness to treatment. Fourth, despite randomization, there was a baseline imbalance in PHQ-9 scores between groups, potentially confounding the psychological outcomes. Fifth, hip strength, which plays an important role in PFP management, was not assessed. Sixth, because the intervention combined exercise and CBT, it is difficult to determine the relative contribution of each component to the observed effects. Finally, the 12-week follow-up period may have been insufficient to evaluate the long-term sustainability of treatment outcomes.

Conclusions

This study demonstrated that an 8-week multidisciplinary DTx intervention incorporating exercise and CBT effectively reduced pain, enhanced knee extensor strength, and improved QoL in patients with PFP. These findings support the efficacy and safety of a multidisciplinary DTx approach for PFP, highlighting its potential to enhance patient adherence and facilitate symptom management through accessible home-based interventions. Further large-scale studies are required to explore the broad applicability and sustained effectiveness of DTx for PFP.

Acknowledgments

This study was funded by the Ministry of Trade, Industry, and Energy and by local governments (Gangwon special self-governing province and Chuncheon City) in 2022, with support from Gangwon Techno Park's Digital Therapeutics Development Support Center construction project (P0018138). Statistical analysis was conducted by Zarathu Co, Ltd, to ensure methodological rigor and minimize bias. The authors thank Byung-Hoon Kim for technical advice and Haeryeon Jang for project coordination of the MORA Cure (PFP) app. No generative artificial intelligence tools were used to generate or ideate the text or visual materials included in this manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SL contributed to writing—original draft preparation, formal analysis, and visualization. CY contributed to conceptualization, software, funding acquisition, methodology, and writing—review and editing. CHC and HRC contributed to software and writing—review and editing. THP contributed to methodology and data curation. SY contributed to software and supervision. TWK contributed to investigation, resources, and writing—review and editing. JHP contributed to writing—original draft preparation, formal analysis, and methodology. MJC contributed to conceptualization, methodology, investigation, resources, and writing—review and editing. CBC contributed to conceptualization, methodology, investigation, and resources. SL and CY served as co–first authors, and JHP and MJC served as co–corresponding authors. All authors reviewed and approved the final manuscript.

Conflicts of Interest

JHP and CBC served as clinical advisors for EverEx. SL, CY, CHC, THP, and HRC were affiliated with EverEx.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1024 KB - mhealth_v13i1e69627_app1.pdf]

References

1. Petersen W, Ellermann A, Gösele-Koppenburg A, Best R, Rembitzki IV, Brüggemann GP, et al. Patellofemoral pain syndrome. Knee Surg Sports Traumatol Arthrosc 2014 Oct;22(10):2264-2274 [FREE Full text] [doi: 10.1007/s00167-013-2759-6] [Medline: 24221245]
2. Smith BE, Selfe J, Thacker D, Hendrick P, Bateman M, Moffatt F, et al. Incidence and prevalence of patellofemoral pain: a systematic review and meta-analysis. PLoS One 2018 Jan 11;13(1):e0190892 [FREE Full text] [doi: 10.1371/journal.pone.0190892] [Medline: 29324820]
3. Crossley KM, Stefanik JJ, Selfe J, Collins NJ, Davis IS, Powers CM, et al. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. Br J Sports Med 2016 Jul;50(14):839-843 [FREE Full text] [doi: 10.1136/bjsports-2016-096384] [Medline: 27343241]
4. Rathleff MS, Rathleff CR, Olesen JL, Rasmussen S, Roos EM. Is knee pain during adolescence a self-limiting condition? Prognosis of patellofemoral pain and other types of knee pain. Am J Sports Med 2016 May;44(5):1165-1171. [doi: 10.1177/0363546515622456] [Medline: 26792702]

XSL•FO
RenderX

5. Thomas MJ, Wood L, Selfe J, Peat G. Anterior knee pain in younger adults as a precursor to subsequent patellofemoral osteoarthritis: a systematic review. BMC Musculoskelet Disord 2010 Sep 09;11:201 [FREE Full text] [doi: 10.1186/1471-2474-11-201] [Medline: 20828401]
6. Crossley KM. Is patellofemoral osteoarthritis a common sequela of patellofemoral pain? Br J Sports Med 2014 Mar;48(6):409-410. [doi: 10.1136/bjsports-2014-093445] [Medline: 24569144]
7. Witvrouw E, Callaghan MJ, Stefanik JJ, Noehren B, Bazett-Jones DM, Willson JD, et al. Patellofemoral pain: consensus statement from the 3rd International Patellofemoral Pain Research Retreat held in Vancouver, September 2013. Br J Sports Med 2014 Mar;48(6):411-414. [doi: 10.1136/bjsports-2014-093450] [Medline: 24569145]
8. Collins NJ, Barton CJ, van Middelkoop M, Callaghan MJ, Rathleff MS, Vicenzino BT, et al. 2018 Consensus statement on exercise therapy and physical interventions (orthoses, taping and manual therapy) to treat patellofemoral pain: recommendations from the 5th International Patellofemoral Pain Research Retreat, Gold Coast, Australia, 2017. Br J Sports Med 2018 Sep;52(18):1170-1178 [FREE Full text] [doi: 10.1136/bjsports-2018-099397] [Medline: 29925502]
9. van Linschoten R, van Middelkoop M, Berger MY, Heintjes EM, Verhaar JA, Willemsen SP, et al. Supervised exercise therapy versus usual care for patellofemoral pain syndrome: an open label randomised controlled trial. BMJ 2009 Oct 20;339:b4074 [FREE Full text] [doi: 10.1136/bmj.b4074] [Medline: 19843565]
10. Collins NJ, Bierma-Zeinstra SM, Crossley KM, van Linschoten RL, Vicenzino B, van Middelkoop M. Prognostic factors for patellofemoral pain: a multicentre observational analysis. Br J Sports Med 2013 Mar;47(4):227-233. [doi: 10.1136/bjsports-2012-091696] [Medline: 23242955]
11. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther 2010 Jun;15(3):220-228 [FREE Full text] [doi: 10.1016/j.math.2009.12.004] [Medline: 20163979]
12. Piva SR, Fitzgerald GK, Irrgang JJ, Fritz JM, Wisniewski S, McGinty GT, et al. Associates of physical function and pain in patients with patellofemoral pain syndrome. Arch Phys Med Rehabil 2009 Feb;90(2):285-295 [FREE Full text] [doi: 10.1016/j.apmr.2008.08.214] [Medline: 19236982]
13. Bagheri S, Naderi A, Mirali S, Calmeiro L, Brewer BW. Adding mindfulness practice to exercise therapy for female recreational runners with patellofemoral pain: a randomized controlled trial. J Athl Train 2021 Aug 01;56(8):902-911 [FREE Full text] [doi: 10.4085/1062-6050-0214.20] [Medline: 33237990]
14. Selhorst M, Rice W, Jackowski M, Degenhart T, Coffman S. A sequential cognitive and physical approach (SCOPA) for patellofemoral pain: a randomized controlled trial in adolescent patients. Clin Rehabil 2018 Dec;32(12):1624-1635. [doi: 10.1177/0269215518787002] [Medline: 30045632]
15. Williams AC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2020 Aug 12;8(8):CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub4] [Medline: 32794606]
16. Anan T, Kajiki S, Oka H, Fujii T, Kawamata K, Mori K, et al. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2021 Sep 24;9(9):e27535 [FREE Full text] [doi: 10.2196/27535] [Medline: 34559054]
17. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: a systematic review. Internet Interv 2016 May;4:17-34 [FREE Full text] [doi: 10.1016/j.invent.2015.12.001] [Medline: 30135787]
18. Rintala A, Rantalainen R, Kaksonen A, Luomajoki H, Kauranen K. mHealth apps for low back pain self-management: scoping review. JMIR Mhealth Uhealth 2022 Aug 26;10(8):e39682 [FREE Full text] [doi: 10.2196/39682] [Medline: 36018713]
19. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
20. Pak SS, Janela D, Freitas N, Costa F, Moulder R, Molinos M, et al. Comparing digital to conventional physical therapy for chronic shoulder pain: randomized controlled trial. J Med Internet Res 2023 Aug 18;25:e49236 [FREE Full text] [doi: 10.2196/49236] [Medline: 37490337]
21. Zhang ZY, Tian L, He K, Xu L, Wang XQ, Huang L, et al. Digital rehabilitation programs improve therapeutic exercise adherence for patients with musculoskeletal conditions: a systematic review with meta-analysis. J Orthop Sports Phys Ther 2022 Nov;52(11):726-739. [doi: 10.2519/jospt.2022.11384] [Medline: 35960507]
22. Rohaj A, Bulaj G. Digital therapeutics (DTx) expand multimodal treatment options for chronic low back pain: the nexus of precision medicine, patient education, and public health. Healthcare (Basel) 2023 May 18;11(10):1469 [FREE Full text] [doi: 10.3390/healthcare11101469] [Medline: 37239755]
23. Wallis JA, Roddy L, Bottrell J, Parslow S, Taylor NF. A systematic review of clinical practice guidelines for physical therapist management of patellofemoral pain. Phys Ther 2021 Mar 03;101(3):pzab021. [doi: 10.1093/ptj/pzab021] [Medline: 33533400]
24. Willy RW, Hoglund LT, Barton CJ, Bolgla LA, Scalzitti DA, Logerstedt DS, et al. Patellofemoral pain. J Orthop Sports Phys Ther 2019 Sep;49(9):CPG1-CP95. [doi: 10.2519/jospt.2019.0302] [Medline: 31475628]

XSL•FO
RenderX

25. Hirschmann MT, Hess S, Behrend H, Amsler F, Leclercq V, Moser LB. Phenotyping of hip-knee-ankle angle in young non-osteoarthritic knees provides better understanding of native alignment variability. Knee Surg Sports Traumatol Arthrosc 2019 May 09;27(5):1378-1384. [doi: 10.1007/s00167-019-05507-1] [Medline: 30968238]
26. Wang X, Duan L, Liu H, Ge H, Dong Z, Chen X, et al. The influence of varus and valgus deviation on the contact area of patellofemoral joint in healthy knees. BMC Musculoskelet Disord 2023 Oct 31;24(1):857 [FREE Full text] [doi: 10.1186/s12891-023-06976-9] [Medline: 37907910]
27. Felson DT, Niu J, Gross KD, Englund M, Sharma L, Cooke TD, et al. Valgus malalignment is a risk factor for lateral knee osteoarthritis incidence and progression: findings from the Multicenter Osteoarthritis Study and the Osteoarthritis Initiative. Arthritis Rheum 2013 Feb 28;65(2):355-362 [FREE Full text] [doi: 10.1002/art.37726] [Medline: 23203672]
28. Only AJ, Arendt EA, Hinckel BB. Anatomic risk factors for lateral patellar instability. Arthroscopy 2024 Nov;40(11):2642-2644. [doi: 10.1016/j.arthro.2024.08.009] [Medline: 39477653]
29. Kedroff L, Li Ko Lun A, Shimoni D, Bearne LM. Cognitive behavioural therapy-informed physiotherapy for patellofemoral pain: a feasibility study. Musculoskeletal Care 2019 Dec;17(4):382-389. [doi: 10.1002/msc.1429] [Medline: 31691524]
30. Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I. mHealth intervention for improving pain, quality of life, and functional disability in patients with chronic pain: systematic review. JMIR Mhealth Uhealth 2023 Feb 02;11:e40844 [FREE Full text] [doi: 10.2196/40844] [Medline: 36729570]
31. Nagel J, Wegener F, Grim C, Hoppe MW. Effects of digital physical health exercises on musculoskeletal diseases: systematic review with best-evidence synthesis. JMIR Mhealth Uhealth 2024 Jan 23;12:e50616 [FREE Full text] [doi: 10.2196/50616] [Medline: 38261356]
32. Kang DH, Park JH, Yoon C, Choi CH, Lee S, Park TH, et al. Multidisciplinary digital therapeutics for chronic low back pain versus in-person therapeutic exercise with education: a randomized controlled pilot study. J Clin Med 2024 Dec 04;13(23):7377 [FREE Full text] [doi: 10.3390/jcm13237377] [Medline: 39685835]
33. Crossley KM, van Middelkoop M, Callaghan MJ, Collins NJ, Rathleff MS, Barton CJ. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). Br J Sports Med 2016 Jul 31;50(14):844-852 [FREE Full text] [doi: 10.1136/bjsports-2016-096268] [Medline: 27247098]
34. van der Heijden RA, Lankhorst NE, van Linschoten R, Bierma-Zeinstra SM, van Middelkoop M. Exercise for treating patellofemoral pain syndrome. Cochrane Database Syst Rev 2015 Jan 20;1(1):CD010387 [FREE Full text] [doi: 10.1002/14651858.CD010387.pub2] [Medline: 25603546]
35. Manojlović D, Kozinc Ž, Šarabon N. Trunk, hip and knee exercise programs for pain relief, functional performance and muscle strength in patellofemoral pain: systematic review and meta-analysis. J Pain Res 2021;14:1431-1449 [FREE Full text] [doi: 10.2147/JPR.S301448] [Medline: 34079359]
36. Neal BS, Lack SD, Bartholomew C, Morrissey D. Best practice guide for patellofemoral pain based on synthesis of a systematic review, the patient voice and expert clinical reasoning. Br J Sports Med 2024 Dec 18;58(24):1486-1495. [doi: 10.1136/bjsports-2024-108110] [Medline: 39401870]
37. Arslan T, Gültekin MZ. The effect of a supervised online group exercise program on symptoms associated with patellofemoral pain syndrome in women. Technol Health Care 2023;31(2):771-782. [doi: 10.3233/THC-220533] [Medline: 36442169]
38. Hart L. Supervised exercise versus usual care for patellofemoral pain syndrome. Clin J Sport Med 2010 Mar;20(2):133. [doi: 10.1097/01.jsm.0000369405.77182.29] [Medline: 20215901]
39. Nunes GS, de Oliveira Silva D, Crossley KM, Serrão FV, Pizzari T, Barton CJ. People with patellofemoral pain have impaired functional performance, that is correlated to hip muscle capacity. Phys Ther Sport 2019 Nov;40:85-90. [doi: 10.1016/j.ptsp.2019.08.010] [Medline: 31499400]
40. Bennell K, Bartam S, Crossley K, Green S. Outcome measures in patellofemoral pain syndrome: test retest reliability and inter-relationships. Phys Ther Sport 2000 May;1(2):32-41. [doi: 10.1054/ptsp.2000.0009]
41. Glaviano NR, Holden S, Bazett-Jones DM, Singe SM, Rathleff MS. Living well (or not) with patellofemoral pain: a qualitative study. Phys Ther Sport 2022 Jul;56:1-7. [doi: 10.1016/j.ptsp.2022.05.011] [Medline: 35691244]
42. Maclachlan LR, Collins NJ, Matthews ML, Hodges PW, Vicenzino B. The psychological features of patellofemoral pain: a systematic review. Br J Sports Med 2017 May;51(9):732-742. [doi: 10.1136/bjsports-2016-096705] [Medline: 28320733]
43. Coburn SL, Barton CJ, Filbay SR, Hart HF, Rathleff MS, Crossley KM. Quality of life in individuals with patellofemoral pain: a systematic review including meta-analysis. Phys Ther Sport 2018 Sep;33:96-108. [doi: 10.1016/j.ptsp.2018.06.006] [Medline: 30059951]
44. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
45. Baroni MP, Jacob MF, Rios WR, Fandim JV, Fernandes LG, Chaves PI, et al. The state of the art in telerehabilitation for musculoskeletal conditions. Arch Physiother 2023 Jan 04;13(1):1 [FREE Full text] [doi: 10.1186/s40945-022-00155-0] [Medline: 36597130]
46. Xiang XN, Wang ZZ, Hu J, Zhang JY, Li K, Chen QX, et al. Telehealth-supported exercise or physical activity programs for knee osteoarthritis: systematic review and meta-analysis. J Med Internet Res 2024 Aug 02;26:e54876 [FREE Full text] [doi: 10.2196/54876] [Medline: 39094114]

XSL•FO
RenderX

47. Lewkowicz D, Wohlbrandt AM, Bottinger E. Digital therapeutic care apps with decision-support interventions for people with low back pain in Germany: cost-effectiveness analysis. JMIR Mhealth Uhealth 2022 Feb 07;10(2):e35042 [FREE Full text] [doi: 10.2196/35042] [Medline: 35129454]
48. Seo YC, Yong SY, Choi WW, Kim SH. Meta-analysis of studies on the effects of digital therapeutics. J Pers Med 2024 Jan 30;14(2):157 [FREE Full text] [doi: 10.3390/jpm14020157] [Medline: 38392592]
49. Gruner MP, Hogaboom N, Hasley I, Hoffman J, Gonzalez-Carta K, Cheville AL, et al. Prospective, single-blind, randomized controlled trial to evaluate the effectiveness of a digital exercise therapy application compared with conventional physical therapy for the treatment of nonoperative knee conditions. Arch Rehabil Res Clin Transl 2021 Dec;3(4):100151 [FREE Full text] [doi: 10.1016/j.arrct.2021.100151] [Medline: 34977534]

Abbreviations

AKPS: Anterior Knee Pain Scale
CBT: cognitive behavioral therapy
DTx: digital therapeutics
HKA: hip-knee-ankle
NPRS: numeric pain rating scale
PFP: patellofemoral pain
PHQ-9: 9-item Patient Health Questionnaire
QoL: quality of life
RCT: randomized controlled trial

Edited by L Buis; submitted 10.12.24; peer-reviewed by E Demirdel, JY Park; comments to author 20.05.25; revised version received 08.07.25; accepted 22.09.25; published 17.10.25.

Please cite as:
Lee S, Yoon C, Choi CH, Park TH, Yang SJ, Cha HR, Kim TW, Park JH, Chang MJ, Chang CB
Efficacy of a Mobile Multidisciplinary Digital Therapeutics App for Patellofemoral Pain: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e69627
URL: https://mhealth.jmir.org/2025/1/e69627
doi:10.2196/69627
PMID:

©Sanghee Lee, Chan Yoon, Chi-Hyun Choi, Tae Hyun Park, Sang Jin Yang, Ha Ri Cha, Tae Woo Kim, Jae Hyeon Park, Moon Jong Chang, Chong Bum Chang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Completion and Compliance Rates for an Intensive mHealth Study Design to Promote Self-Awareness and Self-Care Among Care Partners of Individuals With Traumatic Brain Injury: Secondary Analysis of a Randomized Controlled Trial

Noelle E Carlozzi[1,2,3], BA, MS, PhD; Jonathan Troost[4], PhD; Wendy L Lombard[1], MPH; Jennifer A Miner[1], MBA; Christopher M Graves[1], BA; Sung Won Choi[5], MD; Zhenke Wu[6,7], PhD; Srijan Sen[8], MD; Angelle M Sander[9,10], PhD

[1]Department of Physical Medicine and Rehabilitation, University of Michigan, 2800 Plymouth Road, Ann Arbor, MI, United States

[10]Brain Injury Research Center, TIRR Memorial Hermann, Houston, TX, United States

[2]Department of Surgery, University of Michigan, Ann Arbor, MI, United States

[3]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, United States

[4]Michigan Institute for Clinical and Health Research, University of Michigan, Ann Arbor, MI, United States

[5]Department of Pediatrics, University of Michigan, Ann Arbor, MI, United States

[6]Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States

[7]Michigan Institute for Data and Artificial Intelligence in Society, University of Michigan, Ann Arbor, MI, United States

[8]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States

[9]H. Ben Taub Department of Physical Medicine and Rehabilitation, Baylor College of Medicine, Houston, TX, United States

Corresponding Author:
Noelle E Carlozzi, BA, MS, PhD
Department of Physical Medicine and Rehabilitation, University of Michigan, 2800 Plymouth Road, Ann Arbor, MI, United States

Abstract

Background: Compliance rates for mobile health (mHealth) studies that involve intensive study designs are highly variable. Both person- and study-specific factors likely contribute to this variability. We were interested in understanding the impact that care partner characteristics and demographics have on study engagement, given that engagement is critical to the success of mHealth interventions.

Objective: The primary objective of this report was to analyze the overall and component-specific completion and compliance rates for an intensive 6-month mHealth intervention (CareQOL app) designed to promote self-awareness and self-care among care partners of individuals with traumatic brain injury.

Methods: This randomized controlled trial was designed to test the CareQOL app, an mHealth app designed to promote care partner self-awareness (through self-monitoring) and self-care (through personalized self-care push notifications). The study design consisted of a baseline assessment, a 6-month home-monitoring period that included 3 daily ecological momentary assessment (EMA) questions, monthly patient-reported outcome (PRO) surveys, continuous activity and sleep monitoring using a Fitbit, and 2 follow-up PRO surveys at 3 and 6 months posthome monitoring. Three participants withdrew prior to the initiation of the home-monitoring period, resulting in a final analytical sample size of 254. All participants had access to a self-monitoring dashboard (CareQOL app) that included graphical displays of the daily survey scores, as well as daily steps and sleep data from the Fitbit.

Results: Overall compliance for the different aspects of the study was high. On average, the full-sample daily EMA PRO completion rate was 84% (SD 19%), Fitbit-based step count compliance was 90% (SD 21%), and Fitbit-based sleep duration compliance was 75% (SD 32%); there was no difference between the study arms for daily EMA PROs and Fitbit compliance rates. Completion rates for monthly and follow-up PRO surveys were even higher, with average end-of-month completion rates ranging from 97% to 100%, and follow-up completion rates of 95% for both time points. Again, these rates did not differ by study arm. The data were represented by 3 engagement groups: high-compliance—all data; high-compliance—PROs and steps only; and moderate PRO compliance—low Fitbit compliance. Group membership was predicted by both race (P<.001) and relationship to the care recipient (P=.001), but not by the other person-specific variables.

Conclusions: The compliance rates for this intensive study design are consistent, but at the high end, with what has been reported previously in the literature for studies with shorter time durations. Except for race and relationship to the care recipient,

XSL•FO
RenderX

person-specific factors did not appear to be significantly associated with the engagement group. As such, we anticipate that the high compliance rates observed in this study are likely due to several study-specific design elements that were used to encourage study engagement.

Trial Registration: ClinicalTrial.gov NCT04570930; https://clinicaltrials.gov/study/NCT04570930

International Registered Report Identifier (IRRID): RR2-10.1371/journal.pone.0268726

(JMIR Mhealth Uhealth 2025;13:e73772) doi:10.2196/73772

KEYWORDS

caregivers; mobile apps; self-care; ecological momentary assessment; quality of life

Introduction

Survivors of traumatic brain injury (TBI) experience varying degrees of cognitive and motor function and require a range of acute, intermediate, and long-term supportive care based on the nature of their injury [1,2]. Much of this caring burden falls to informal (family) care partners, whose lives are upended by both dramatic changes in their loved one's health and functioning and the sudden need to assume a medical caregiving role for which they may feel ill-equipped. The major lifestyle impacts and associated distress imposed by these changes increase these care partners' risks for depression, anxiety, fatigue, social isolation, and sleep disturbance [3-7]. Not only do these problems negatively impact the care partners, but they also affect the health and well-being of the survivor of TBI who relies on them for care [3,5,8-11].

The time and energy demands of caring for someone with a TBI make it difficult for care partners to prioritize their own needs [12]. Thus, interventions targeting this group of care partners must be low burden, simple to use, and easy to incorporate into their day-to-day lives. Mobile health (mHealth) interventions are well-suited for this purpose due to widespread acceptance of mobile technologies and devices, as well as convenient access to mHealth-delivered content. In a recent systematic review of cognitive and behavioral digital health interventions among people with TBI and their care partners, mHealth interventions were found to be feasible, and all included studies demonstrated positive outcomes [13]. However, it is important to consider how differences in care partner characteristics and demographics may influence compliance rates for mHealth studies that involve ecological momentary assessments (EMAs), which have been shown to be highly variable.

In general, compliance for brief, intensive mHealth studies (ie, 2 wk or less) is moderately high, ranging from 57% to 87% [14-20]; these rates tend to progressively worsen as the study duration increases (ranging from 55% to 90% for studies lasting 2 to 4 weeks [21-30] and 77% for the single study we found that included a study duration of greater than 3 mo [31]). These high rates of variability are due to demographic factors [32], as well as person-specific [33] and protocol-specific factors [34]. For example, the literature has shown that racial and ethnic minorities tend to have lower compliance and no-show rates relative to their White or non-Hispanic counterparts [35-37]. Specifically, a recent review of barriers and facilitators to engaging minority populations in health care research identified mistrust (of health care professionals or systems, researchers, and research in general), cultural and language barriers, socioeconomic and logistical challenges, lack of research information and awareness, external influences (family and friends), and perceived bias from health care providers or researchers as the primary barriers to participation in research [38]. Furthermore, there are lower compliance rates among individuals with lower socioeconomic status, as well as individuals with less education [39,40]. Reasons for this include financial, time, and resource constraints associated with participation (eg, access to transportation, child and older people care, time off from work, or unanticipated costs), and disparities in health literacy [41-44]. Rates for age are mixed [45,46], but most studies show a positive association between age and compliance [47-54]. Furthermore, protocol-specific factors [34], which include survey length, study duration, and compensation amount, as well as compensation criteria, are highly variable across EMA studies and likely contribute to compliance rates (Table 1).

XSL•FO
RenderX

Table . Summary of compliance rates for published EMA[a] studies.

Authors	Sample population	Sample size, n	Study length	EMA assessments, n	Average compliance rates for EMAs	Compensation
Jia et al [14]	Existing users of online food delivery	102 (53 in EMA group)	3 days	5 per day	72.5%	US $29 for completion of the entire study, US $9 if they completed 1/3 of the study, and US $19 for 2/3 of the study
Trang et al [15]	Men who have sex with men	46	7 days	6 - 8 per day	61.8%	US $0.30 for each completed EMA
Solk et al [16]	People with breast cancer	75	10 days	4 per day	86%	US $100 for each of 3 check-ins, regardless of compliance
Kratz et al [17]	People with spinal cord injury and chronic pain	131	7 days	5 per day	81%	US $5 per day for days 1 - 3, US $10 per day for days 4 - 5, and US $20 per day for days 6 - 7 of home monitoring
Smiley et al [18]	Gay and bisexual men	25	14 days	3 per day	57.3%	US $25 for <50% of EMAs completed; US $50 for ≥50% of EMAs completed
Ponnada et al [19]	University students or staff	17	7 days	72 per day	87%	Not reported
Elbin et al [20]	Adolescents and young adults with concussion	116	7 days	4 per day	64%	Not reported
Tonkin et al [21]	Daily cigarette users	81	9 weeks	1 scheduled and 4 random per day	58% for random and 86% for scheduled	US $1 per assessment completed
Niznik et al [22]	Young adult cannabis and tobacco users	97	28 days	3 random per day	55%	US $20 per week for 4 weeks of IVR[b] monitoring, an additional US $1 for each random assessment completed (maximum of US $3 per day), and a bonus of US $2 per week for completing assessments 6 of 7 days or US $5 per week for completing assessments for all 7 days
Juengst et al [23]	People with traumatic brain injury	20	8 weeks	2 - 4 per day	73.4%	Not reported
Krohn et al [24]	People with postpartum depression	26	6 weeks	2 per day	67%	Not reported
Slade et al [25]	College students	145	30 days	1 per day	78%	Not reported

XSL•FO
RenderX

Authors	Sample population	Sample size, n	Study length	EMA assessments, n	Average compliance rates for EMAs	Compensation
Joo et al [26]	People with chronic pain with cannabis and opioid use	133	30 days	4 random prompts and 1 scheduled	89.7% for scheduled and 63.3% for randomly prompted	US $2 for completing at least one EMA each day, plus US $60 bonus if they achieved a 75% or greater EMA compliance rate over the study duration
Yang et al [27]	Men who have sex with men	16	4 weeks	3 random prompts and 1 morning prompt	74% of random prompts and 80.7% of morning prompts	US $50 every week for answering 80% of their alarms or US $25 every week for answering 60% of their alarms. They received no payments for answering less than 60% of their alarms or if their phone was uncharged
Beres et al [28]	People in Uganda	50	90 days	1 fixed and 1 random	66.5%	~US $30 for their time for responding to ≥50% of prompts
Sanjuan et al [29]	Pregnant women with prior trauma exposure	33	28 days	3 per day	74%	US $10 - 40 for each weekly EMA download (dependent upon EMA completion rate), and US $25 extra for the final EMA download.
Laborde et al [30]	Older adults with knee osteoarthritis	27	~2 weeks	4 per day	83%	Not reported
Howard and Lamb [31]	Undergraduate alcohol drinkers	196	14 weeks	1.9 per day (13 per week)	76.5%	US $1 in Amazon cash for each survey; daily draws awarding US $25 in Amazon cash to one winner per day (Friday through Sunday; completion of at least one survey on the previous day to be eligible). After weeks 5 and 10, two unannounced US $5 bonuses to students who remained enrolled

[a]EMA: ecological momentary assessment.

[b]IVR: interactive voice response.

In our own work, we have demonstrated compliance rates at the high end of the above ranges. Specifically, in a 1-week study that included 3 daily EMA questions, we found an average response rate of 83% [55]. In another, longer study (3 mo duration), we found an average response rate of 90% for once-daily EMA questions, and 96% and 85% for daily step count data and sleep duration estimates, respectively (derived from continuous monitoring with a wearable device) [56]. We believe that these high compliance rates can be attributed, at least in part, to several study-specific design elements focused on fostering participant engagement (including survey brevity, monetary compensation, and app-specific customization

XSL•FO
RenderX

features, as well as regular reminders following a 3-day lapse in responding [57].

Regardless of our previous findings in other care partner populations, care partners of individuals with TBI are a population that may potentially have difficulty engaging with EMA over longer time periods due to elevated rates of TBI in racial and ethnic minority groups, established racial and ethnic differences in TBI caregiving styles (Black care partners have been found to include more nonspouse relatives, spend considerably more time providing direct care, and care for significantly more disabled survivors than White care partners [58]), and the overall time commitment required of the caregiver role. We engaged a sample of care partners of individuals with TBI in an intensive study design that was of a longer duration (ie, a 12 mo study that included 6 mo of EMA PROs) [59]. In this study, participants completed a once-daily EMAs (3 items), wore a Fitbit, completed monthly surveys for 6 months, and completed follow-up surveys at 9 and 12 months. Previous examination of this dataset indicated high rates of compliance, with average completion rates of 84% (SD 19%) for once-daily EMA questions, and 90% (SD 21%) and 75% (SD 32%) for daily step count and sleep duration estimates, respectively (derived from continuous monitoring with a Fitbit) [59]. These study-specific race or ethnicity participation rates (20%) were comparable to or exceeded established US Census rates [60]. For the purposes of this analysis, we wished to explore these rates more closely and determine what, if any, additional demographic factors were related to compliance rates in our sample.

Methods

Participants

We enrolled care partners (ie, informal caregivers) of people living with TBI in this study between December 2020 and February 2023. Participants were recruited through two academic medical centers using clinical databases [61], site-specific registries, and community outreach. Recruitment often included contacting a person with a known TBI for their care partner referrals. Care partners needed to be at least 18 years of age, able to read and understand English, and caring for an adult at least 1 year postinjury who had sustained a medically documented complicated mild, moderate, or severe TBI. The injury must have occurred when the care recipient was aged 16 years or older. A care partner was defined as an individual who provided assistance to a person with a TBI (indicated by a response greater than 0 on the following eligibility rating question: On a scale of 0 - 10, where 0 is “no assistance” and 10 is “assistance with all activities,” how much assistance does the person you care for require from you to complete activities of daily living due to problems resulting from his or her TBI?). Care partners were excluded if they did not have access to resources for participating in an mHealth intervention, including a personal mobile device capable of downloading the study apps for this study. Participants had to be willing to download the CareQOL and Fitbit apps to their device and be willing to complete all study assessments. In addition, we excluded professional, paid caregivers.

Ethical Considerations

All study activities were conducted in accordance with institutional review board (IRB) approvals, and the study was registered with ClinicalTrials.gov (NCT04570930). The protocol, informed consent document, and all participant materials have received approval from IRBMED, which is the IRB of record for both data collection sites (IRBMED Multisite Application Approval HUM00181282; IRBMED University of Michigan Site Application Approval HUM00186921; IRBMED Baylor College of Medicine Site Application Approval SITE0000087; Baylor College of Medicine/Memorial Hermann IRB H-48478). All participants provided informed consent prior to the engagement in study activities. Participants were assigned a participant ID by the study team, which was used to avoid the inclusion of other identifying information. The electronic systems used to store the data collected in this study were secure systems with password protection and restricted access. Paper documents related to participation were stored in a locked cabinet or office. No names or other identifying information have been used in any report or publication of this study. This research was also covered by a Certificate of Confidentiality from the National Institutes of Health. Participants were compensated up to US $310 (US $20 each at baseline, 6 months, 3 months post, and 6 months post assessments; US $10 for each end of month survey for months 1 - 5; and US $1 per day for each day that they had EMA or Fitbit data during the 6-month home monitoring period) and were allowed to keep the Fitbit. Compensation occurred monthly to encourage prompt responding. Participation was confidential, and study data were deidentified.

Measures

A detailed description of the study protocol is provided elsewhere [62]. Briefly, participants completed a baseline assessment assessing demographic variables, proxy-reported measures of the care recipient’s functional and emotional status (Supervision Rating Scale [SRS] [63], Mayo-Portland Adaptability Inventory-Fourth Edition [64], and the Posttraumatic Stress Disorder Checklist for DSM-5 [65]), and 12 care partner self-reported health-related quality of life (HRQOL) PROs (Caregiver Strain [66,67], Caregiver-Specific Anxiety [67,68], Sleep-Related Impairment [69], Fatigue [69], Anxiety [69], Depression [69], Anger [69], Self-Efficacy-General [70], Positive Affect and Well-Being [69], Perceived Stress [70], Ability to Participate in Social Roles and Activities [69], and Global Health [71]). This was followed by a 6-month home monitoring period that included 3 care partner self-reported daily EMA PRO questions (single-item assessments of Caregiver Strain [66,67], Anxiety [69], and Depression [69]) and monthly self-reported (ie, PRO) surveys (again assessing the 12 HRQOL domains), as well as continuous monitoring of physical activity and sleep using a Fitbit. The 3- and 6-month follow-up HRQOL PROs were identical to the end-of-month PROs. In addition, a feasibility and acceptability survey was administered at the end of month 6 [56].

Study Procedures

Participants were randomized to either a self-monitoring alone arm, which included completion of the daily EMA questions;

XSL•FO
RenderX

baseline, monthly, and follow-up PRO surveys; and 6 months of continuous activity and sleep monitoring with a Fitbit, or to a self-monitoring plus self-care push notifications arm, which included self-monitoring plus self-care push notifications that involved a 50/50 chance each day of receiving a self-care prompt in addition to the other assessments. All participants had access to a self-monitoring dashboard (CareQOL app) that included graphical displays of the daily EMA scores, as well as daily step count and sleep duration data from the Fitbit. The proxy-reported measures were administered at baseline only through a REDCap (Research Electronic Data Capture; Vanderbilt University) survey, and self-reported PROs were administered through the CareQOL app at baseline, monthly, and both follow-ups. In addition, several electronic data capture and management platforms were used, including REDCap, CareQOL, Qualtrics, Fitbit, the University of Michigan Health Information Technology and Services server, and the Google Cloud.

Statistical Analysis

First, we examined completion rates for daily EMA questions and monthly surveys, separately by study arm (calculated as the percentage of days with data over the number of days in the study for daily, monthly, and follow-up surveys), and then conducted a series of linear regression analyses to determine whether there were differences by study arm. Next, we examined the interrelationships among different types of completion (EMA questions, monthly survey responses) and compliance (Fitbit-based estimates of daily step count and sleep duration). We then used a backward selection process to determine which variables in the dataset related to compliance rates. For this analysis, we conducted a series of linear regression analyses to determine which variables (including demographic variables; baseline PROs, SRS, Mayo-Portland Adaptability Inventory-Fourth Edition, and Posttraumatic Stress Disorder Checklist for DSM-5; and feasibility and acceptability questions [assessed at 6 mo]) predicted completion or compliance rates.

Following this selection process, we used k-means clustering to identify latent (ie, “unobserved”) categorical subgroups of respondents based on their daily completion and compliance rates. The optimal number of clusters was determined based on assessment of model fit and parsimony (pseudo F statistic, approximate R-squared, cubic clustering criterion). Once the class number was determined, respondents were classified into latent classes based on maximum posterior probability. We also examined whether or not different descriptive variables (care partner age, care partner gender, care recipient age, care recipient gender, care partner race, care partner ethnicity, duration providing care, time spent caregiving, relationship to care recipient, work status, SRS score, PCL score, and functional ability of the care recipient) were able to predict the identified clusters.

Results

Although 257 participants were initially randomized, 3 participants withdrew prior to the initiation of the study home monitoring period, resulting in a final sample size of 254 care partners of people with TBI. Of these 254 participants, 236 (92.9%) completed the 12-month study (n=118 for each of the individual study arms). Details describing the demographic data for the different study arms have been published previously (Carlozzi et al [59]). Table 2 provides an abbreviated summary of the demographic data for the full sample.

Overall compliance for the different aspects of the study was high (Tables 3 and 4). Specifically, on average, the full-sample daily EMA completion rate was 84.5% (SD 19%), Fitbit-based step count compliance was 90.4% (SD 21%), and Fitbit-based sleep duration compliance was 74.9% (SD 32%); there was no difference between the study arms for daily completion and compliance rates (Table 3). Completion rates for monthly and follow-up surveys were even higher, with average end-of-month completion rates ranging from 97.2% to 100.0%, and follow-up completion rates of 95.1% for both 3-month and 6-month post time points; again, these rates did not differ by study arm (Table 4). Compliance rates were moderately correlated for the two Fitbit-based measures (r=0.65), and the magnitude of the correlations was less robust between Fitbit-based compliance data and the daily EMAs (r=0.38 between steps and EMAs and r=0.29 between sleep and EMAs).

A backward selection process indicated that: (1) several feasibility and acceptability questions related to ease, satisfaction, and burden of completing the EMAs predicted daily EMA completion rates; (2) race, relation to caregiver, and several of the feasibility and acceptability questions related to using the Fitbit predicted daily Fitbit-based step count compliance rates; and (3) race, relationship status, eligibility rating, and some of the feasibility and acceptability questions related to using the Fitbit predicted daily Fitbit-based sleep duration compliance rates (Multimedia Appendix 1). More specifically, positive perceptions about feasibility and acceptability were related to better completion and compliance rates for EMAs and monthly surveys, steps, and sleep; being Black was associated with lower Fitbit-based compliance rates for both steps and sleep; being a friend or other family member was associated with lower Fitbit-based compliance rates for daily steps; and being single and caring for someone with more functional deficits was associated with lower Fitbit-based compliance rates for sleep (Table 5).

Results from the cluster analysis did not indicate clear support for either a 2-, 3-, or 4-cluster model (Multimedia Appendix 2). The 2-cluster model indicated a “good” versus a “bad” compliance group, the 3-cluster model indicated: (1) a “high-compliance group—all data” cluster (cluster 1), where compliance across all data types (EMA and survey PROs, steps, and sleep) was high (n=182, 71.7% of participants); (2) a “high-compliance group—PROs and steps only” cluster (cluster 2), where compliance was high for EMA and survey PROs and Fitbit-based compliance rates for steps, but not for sleep (n=55, 21.7% of participants); and (3) a “moderate PRO compliance, low Fitbit compliance group” cluster (cluster 3), where monthly survey completion was good, but not excellent, and Fitbit compliance for both steps and sleep was low (n=17, 6.7% of people). The 4-cluster model further splits the “moderate PRO compliance, low Fitbit compliance group.” We elected to further explore the 3-cluster model (Figure 1) given the fact that this clustering was best aligned with the previous analyses (where

XSL•FO
RenderX

we consider survey, sleep, and step compliance separately), as well as the overall number of people represented by each cluster (ie, we combined a data-driven approach with clinical interpretability for final cluster selection [72]).

Cluster membership was predicted by both race (P<.001), with a larger proportion of White caregivers relative to the other racial groups in cluster 1 (high compliance group—all data) relative to the other 2 clusters (high compliance group—PROs and steps only and moderate PRO and low Fitbit compliance group), and relationship to the care recipient (P=.001), with the moderate PRO and low Fitbit compliance group having a higher proportion of parents and a lower proportion of partners than the other two clusters, but not to other demographic factors or other characteristics (Table 6).

XSL•FO
RenderX

Table . Sample descriptive data (n=254).

	Value (n=254)
Age (years), mean (SD)	52.0 (14.7)
Age of person cared for (years), mean (SD)	43.0 (17.7)
Sex (female), n (%)	201 (79.1)
Sex of person cared for (female), n (%)	68 (26.8)
Race, n (%)	
American Indian or Alaska Native	3 (1.2)
Asian	10 (3.9)
Black or African American	36 (14.2)
More than 1 race	8 (3.1)
Native Hawaiian or Other Pacific Islander	1 (0.4)
White or Caucasian	195 (76.8)
Missing	1 (0.4)
Hispanic ethnicity, n (%)	41 (16.1)
Length of caregiving (years), mean (SD)	6.5 (5.7)
Relation to caregiver, n (%)	
Partner	104 (40.9)
Child	38 (15.0)
Parent	81 (31.9)
Sibling	20 (7.9)
Other family	3 (1.2)
Friend	3 (1.2)
Missing	5 (2.0)
Same household, n (%)	
Yes, all the time	182 (71.7)
Yes, but only a few days a week	13 (5.1)
No	59 (23.2)
Work status, n (%)	
Employed full time (at least 40 h/wk)	111 (43.7)
Employed part-time	29 (11.4)
Homemaker	17 (6.7)
Student	11 (4.3)
Retired	56 (22.0)
Retired early due to disability	2 (0.8)
Unemployed <1 year, and looking for work	4 (1.6)
Unemployed <1 year, not looking for work	1 (0.4)
Unemployed >1 year, looking for work	3 (1.2)
Unemployed >1 year, not looking for work	7 (2.8)
Unable to work or disabled	8 (3.1)
Other	5 (2.0)
How much assistance does the person you care for require from you to complete activities of daily living due to problems resulting from his or her TBI[a]? (0 - 10), mean (SD)	6.3 (2.4)

XSL•FO
RenderX

	Value (n=254)
Supervision Rating Scale	
Mean (SD)	4.0 (3.1)
Range	1 - 12
Posttraumatic stress symptoms (PCL-5[b])	
Mean (SD)	20.0 (15.6)
Range	0 - 74
Time spent caregiving, n (%)	
1 to 2 hours per day or less	121 (47.6)
3 to 4 hours per day (ie, half of a working day)	54 (21.3)
5 to 8 hours per day (ie, full working day)	23 (9.1)
9 to 12 hours per day	13 (5.1)
>12 hours per day or round-the-clock care	43 (16.9)

[a]TBI: traumatic brain injury.

[b]PCL-5: Posttraumatic Stress Disorder Checklist for the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*.

Table . Daily compliance data for PROs[a] and Fitbit data (n=254)[b].

	Self-monitoring only		Self-monitoring plus self-care push notifications		Overall			
Assessment	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	β (95% CI)	*P* value
Daily EMA[c] questions	1 - 100	86 (21)	4 - 100	83 (18)	1 - 100	84 (19)	−2.4 (−7.2 to 2.4)	.33
Daily steps	4 - 100	90 (21)	2 - 100	90 (21)	2 - 100	90 (21)	0.00 (−5.3 to 5.3)	.99
Daily sleep	1 - 100	74 (34)	1 - 100	76 (31)	1 - 100	75 (32)	2.4 (−5.7 to 10.4)	.56

[a]PRO: patient-reported outcome.

[b]Overall data are shown for comparative purposes only; it is reproduced from Carlozzi et al [59].

[c]EMA: ecological momentary assessment.

Table . Survey compliance data for baseline, end of month, and follow-up surveys (n=254).

	Self-monitoring only	Self-monitoring plus self-care push notifications	Overall	Chi-square (*df*)	*P* value
Baseline	100	100	100	—[a]	—
Month 1	100	100	100	—	—
Month 2	98	99	99	0.4 (1)	.56
Month 3	99	98	99	0.3 (1)	.57
Month 4	98	98	98	0.2 (1)	.66
Month 5	100	94	97	7.1 (1)	.008
Month 6	100	98	99	2.0 (1)	.16
3-month post	95	95	95	0.0 (1)	.98
6-month post	96	94	95	0.3 (1)	.57

[a]Not applicable.

XSL•FO
RenderX

Table . Variables with a statistically significant relationship with compliance rates.

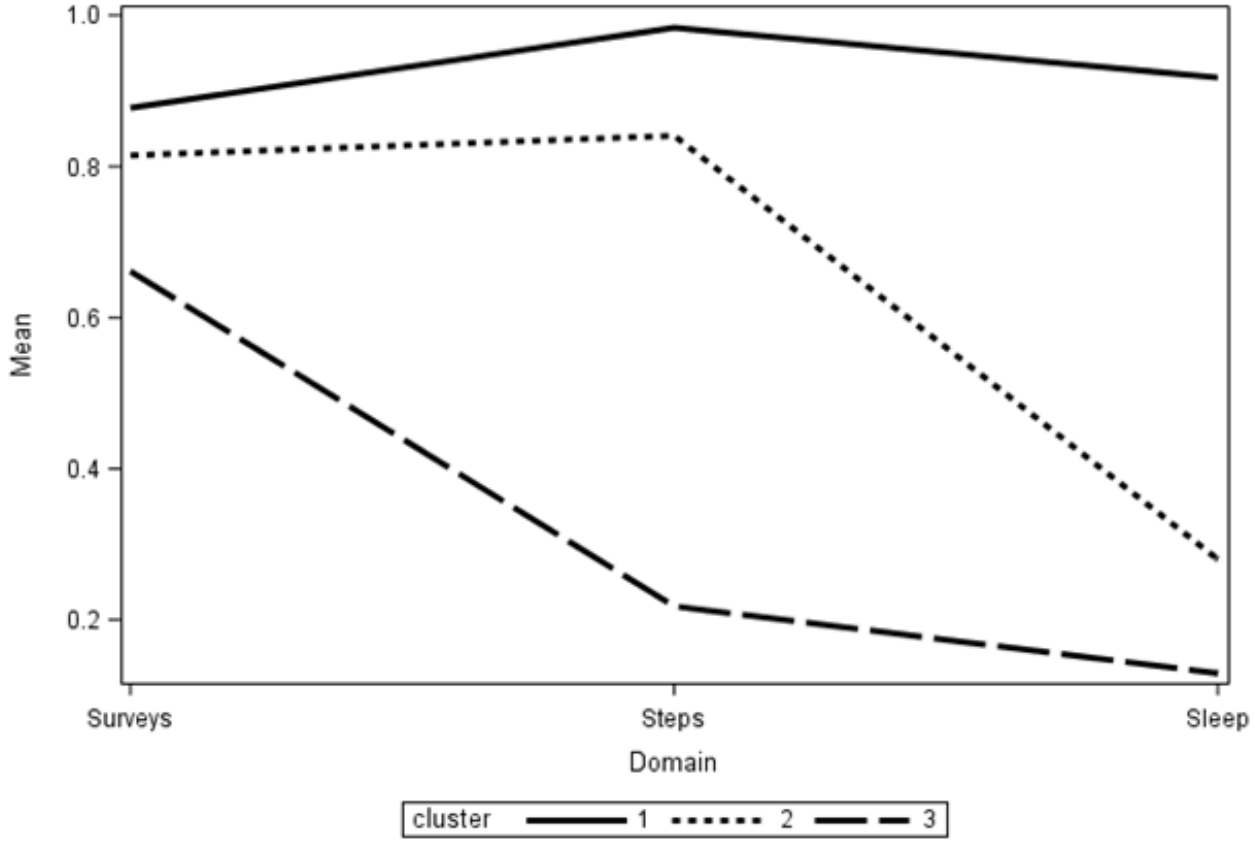
	Daily EMAs[a]		Steps		Sleep	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Race				.045		<.001
Black	__[b]	—	−9.2 (−16.7 to −1.8)	.02	−31.2 (−42.2 to −20.0)	<.001
Other	—	—	−4.2 (−13.5 to 5.0)	.37	9.0 (−22.8 to 4.8)	.20
White	—	—	—	Reference	Reference	Reference
Relation to caregiver				.01	—	—
Child	—	—	−2.4 (−10.1 to 5.4)	.55	—	—
Friend	—	—	−30.6 (−54.7 to −6.6)	.01	—	—
Other family	—	—	−36.0 (−60.0 to −12.0)	.004	—	—
Parent	—	—	−2.5 (−8.6 to 3.5)	.41	—	—
Sibling	—	—	−4.1 (−14.1 to 5.9)	.42	—	—
Partner	—	—	Reference	Reference	Reference	Reference
Marital status (married vs single [reference])	—	—	—	—	18.1 (8.8 to 27.5)	<.001
Eligibility rating	—	—	—	—	−1.8 (−3.4 to −0.1)	.04
The Fitbit data were easy to sync with my phone (per 1 point on a 5-point Likert scale)	2.7 (0.1 to 4.6)	.005	6.2 (3.9 to 8.4)	<.001	7.1 (3.0 to 11.1)	<.001
I was confident using the Fitbit (per 1 point on a 5-point Likert scale)	2.7 (0.1 to 5.1)	.03	9.2 (6.4 to 12.1)	<.001	9.1 (3.9 to 14.2)	<.001

[a]EMA: ecological momentary assessment.

[b]Not applicable.

XSL•FO
RenderX

Figure 1. The solid line (cluster 1) represents a "high compliance group—all data" where compliance, across all data types (PROs, steps, and sleep) was high (72% of participants); the small dashed line (cluster 2) indicates a "high compliance group—PROs and steps only" where compliance was high for PROs and Fitbit-based steps, but not Fitbit-based sleep (22% of participants); and the large dashed line (cluster 3) indicates a "moderate PRO and low Fitbit compliance group" where monthly survey compliance was good, but not excellent, and Fitbit data were very low (7% of people).



XSL•FO
RenderX

Table . Variables with statistically significant relationships with compliance rates.

	Cluster 1 (high compliance group—all data), mean (SD)	Cluster 2 (high compliance group—PROs[a] and steps only), mean (SD)	Cluster 3 (moderate PRO and low Fitbit compliance group), mean (SD)	*P* value
Age (years; care partner)	49.9 (17.6)	52.8 (14.0)	49.9 (15.7)	.45
Age (years; care recipient)	36.9 (18.1)	44.1 (17.7)	41.0 (17.3)	.12
Sex				.50
Male (care partner)	2 (12)	41 (23)	10 (18)	
Female (care partner)	15 (88)	141 (77)	45 (82)	
Sex				.14
Male (care recipient)	9 (53)	137 (75)	40 (73)	
Female (care recipient)	8 (47)	45 (25)	15 (27)	
Race[b]				<.001
Black	4 (24)	12 (7)	20 (36)	
Other	2 (12)	14 (8)	6 (11)	
White	11 (65)	155 (86)	29 (53)	
Ethnicity				.21
Not Hispanic or Latino	15 (88)	148 (81)	50 (91)	
Hispanic or Latino	2 (12)	34 (19)	5 (9)	
Relation to caregiver[c]				.001
Child	3 (18)	26 (15)	9 (16)	
Friend	1 (6)	0 (0)	2 (4)	
Other family	2 (12)	1 (1)	0 (0)	
Parent	7 (41)	58 (33)	16 (29)	
Sibling	1 (6)	12 (7)	7 (13)	
Partner	3 (18)	80 (45)	21 (38)	
Living in the same household				.58
Yes, all of the time	12 (71)	134 (74)	36 (65)	
Yes, but only a few days a week	1 (6)	7 (4)	5 (9)	
No	4 (24)	41 (23)	14 (25)	
Length of time caregiving (years)	4.0 (3.4)	6.8 (5.9)	6.6 (5.35)	.09
Work status				.10
Employed full time (at least 40 h/wk)	5 (29)	80 (44)	26 (47)	
Employed part-time	2 (12)	22 (12)	5 (9)	
Homemaker	3 (18)	9 (5)	5 (9)	
Student	2 (12)	7 (4)	2 (4)	
Retired	3 (18)	45 (25)	8 (15)	
Retired early due to disability	0 (0)	1 (1)	1 (2)	
Unemployed less than 1 year, and looking for work	0 (0)	1 (1)	3 (5)	

XSL•FO
RenderX

	Cluster 1 (high compliance group—all data), mean (SD)	Cluster 2 (high compliance group—PROs[a] and steps only), mean (SD)	Cluster 3 (moderate PRO and low Fitbit compliance group), mean (SD)	*P* value
Unemployed less than 1 year, not looking for work	0 (0)	1 (1)	0 (0)	
Unemployed for more than 1 year, looking for work	1 (6)	0 (0)	2 (4)	
Unemployed for more than 1 year, not looking for work	1 (6)	6 (3)	0 (0)	
Unable to work or disabled	0 (0)	7 (4)	1 (2)	
Other	0 (0)	3 (2)	2 (4)	
How much assistance does the person you care for require from you to complete activities of daily living due to problems resulting from his or her TBI[d]? (0 - 10)	6.8 (2.51)	6.1 (2.44)	6.7 (2.22)	.13
Supervision Rating Scale	5.2 (3.9)	4.0 (3.1)	3.7 (2.9)	.47
Posttraumatic stress symptoms (PCL-5[e])	20.9 (16.6)	18.9 (14.6)	23.3 (17.9)	.23
Time spent caregiving				.21
1 to 2 hours per day or less	4 (24)	93 (51)	24 (44)	
3 to 4 hours per day (ie, half of a working day)	3 (18)	38 (21)	13 (24)	
5 to 8 hours per day (ie, full working day)	3 (18)	14 (8)	6 (11)	
9 to 12 hours per day	1 (6)	11 (6)	1 (2)	
>12 hours per day or round-the-clock care	6 (35)	26 (14)	11 (20)	
Functional status of the person with TBI (MPAI-4[f])	47.9 (20.31)	46.7 (17.49)	46.5 (20.60)	.78

[a]PRO: patient-reported outcome.

[b]One participant in cluster 2 was missing race.

[c]Five participants in cluster 2 were missing relationship to caregiver.

[d]TBI: traumatic brain injury.

[e]PCL-5: Posttraumatic Stress Disorder Checklist for DSM-5.

[f]MPAI-4: Mayo-Portland Adaptability Inventory-Fourth Edition.

Discussion

Principal Findings

Overall, there were high rates of compliance for a 6-month intensive home monitoring protocol that involved daily EMA ratings of HRQOL and continuous monitoring of physical activity and sleep by a Fitbit. Average compliance rates were highest (90%) for activity (step) monitoring (ie, wearing the Fitbit during the day), followed by EMA completion (84%), and lowest (75%) for sleep monitoring (ie, wearing the Fitbit overnight). In addition, the completion rates for the monthly HRQOL surveys were also high (ranging from 92% to 98%). Compliance and completion rates did not differ by study arm. There was only a moderate relationship between wearing the Fitbit during the day versus at night, and there was a less robust (ie, small) relationship between EMA completion rates and Fitbit compliance rates. These rates are consistent, at the high end, with what has been reported previously in the literature for studies with shorter time durations [14-31,33]. The data on compliance rates for longer study durations in the literature, such as we report from this study, is sparse, but would be

XSL•FO
RenderX

expected to be lower than what we found here, given typical patterns of decline in compliance rates with longer study duration.

We were also interested in better understanding the participant-specific variables (eg, personality, comfort with technology, and wearable devices [33]) that might have impacted compliance rates. Not surprisingly, we found that positive perceptions about feasibility and acceptability were related to better compliance rates for EMA questions, surveys, steps, and sleep. This finding is consistent with the primary findings of this study, which showed that participants who were more positive about the study itself (regardless of study arm) were more likely to show HRQOL improvements [59], as well as those reported in the general literature that showed compliance rates were higher among participants who found the research study to be more favorable [73,74]. We also found that demographic variables impacted compliance. We found that being Black was associated with lower compliance rates for Fitbit data (sleep and steps), but not for EMA data. Historically, Black participants are underrepresented in research [75,76], and consistent with our findings, they have higher rates of missing data than their White counterparts [77-83]. Furthermore, being a friend or other family member of the person with TBI (vs a spouse or adult child of the person with TBI) was associated with lower compliance rates for Fitbit daytime wear data (steps). To our knowledge, while some meta-analytic work has examined caregiver populations (eg, caregivers of people living with dementia [84]), this work has not examined caregiver type in consideration of the differential factors that might influence missing data rates. Given this, we postulate that nontraditional caregivers (ie, friends and other family members) may feel less obligated to provide care than those in more traditional caregiver groups (ie, spousal and adult child caregivers). Future work to better understand these relationships, how they influence care, and how this may be related to study compliance rates is needed. We also found that being single or caring for someone with more functional deficits was associated with lower Fitbit nighttime wear data (sleep). While we are unaware of work that looks explicitly at these factors and nighttime compliance rates with wearables, we hypothesize that these types of caregivers may already be experiencing fragmented or disturbed sleep and therefore are more likely to find nighttime Fitbit wear uncomfortable and prohibitive [85].

While there were a handful of person-specific factors that contributed to compliance, our overall high rates of compliance for the different study elements emphasize the need for a focus on study-specific design elements that can help to mitigate these person-specific factors. For example, in this study, we postulate that the brevity of overall assessments, customization of administration windows for the EMA questions, and regular reminders following a 3-day lapse in responding may also have contributed to the higher response rates. Study staff completed regular checks for completion of the EMA questions and the presence of daily step and sleep data from the Fitbit and monthly surveys. In instances where participants were missing greater than 3 days' worth of EMA or Fitbit data, we contacted those participants directly. In addition, automatic reminders for completion of the EMA questions and surveys were sent via the CareQOL app, and study staff contacted participants at least once per month during the home monitoring period to foster engagement. Additionally, we provided monetary compensation for the different elements of the study, including separate payments of US $20 for completing the baseline and 3-month and 6-month postmonitoring period follow-up surveys; US $10 compensation for completion of the monthly surveys during the home monitoring period (with the exception of the final monthly survey, for which participants were paid US $20); and US $1 per day for daily completion of the EMA questions or any daily data from wearing the Fitbit (either day or night). Participants could also keep the study-provided Fitbit after they completed the study. Participants were also able to customize what time they received their three EMA questions each day and were offered different wristband options for the Fitbit to maximize their comfort and style preferences.

These proposed factors align with literature that identifies a number of protocol-specific factors (eg, total number of questions, study duration, compensation rates, and criteria for compensation [34]) that appear to be characteristic of different study designs (Table 1).

Furthermore, when trying to maximize engagement for underrepresented groups, adopting a personalized approach to recruitment (eg, build rapport and use culturally sensitive communication styles), providing culturally appropriate financial incentives (eg, reimburse for travel costs), reducing language barriers (eg, representative research teams and culturally or linguistically appropriate study materials and communication strategies), engaging community champions in the recruitment process, addressing accessibility and logistical barriers to participation (eg, flexibility with location, timing, childcare, and transportation), participating in the community, and building cultural competence of the study team are strategies that have been shown to foster research participation and compliance among these groups [38,86].

Finally, we explored whether there were meaningful subgroups for different patterns of compliance rates. To this end, we found a "high compliance group—all data" that was compliant with all of the different study elements; a "high compliance group—PROs and steps only" where compliance was high for EMA and survey PROs and Fitbit-based compliance rates for steps, but not for sleep; and a "moderate PRO compliance, low Fitbit compliance group" where monthly survey completion was good, but not excellent, and Fitbit wear-time data were low. These subgroups were predicted by race and relationship to the care recipient, but not to other demographic, clinical, or behavioral characteristics, once again supporting the premise that study-specific enhancements to foster engagement would be beneficial for these groups.

Limitations

While these results support the feasibility of care partner participation in studies that use intensive study designs (including EMAs over a 6-mo period, wearing of a wrist-worn device that provides continuous monitoring, and completion of end-of-month surveys), it is also important to acknowledge several study limitations. For example, while we have postulated about the reasons the observed completion rates were so high

XSL•FO
RenderX

in this study, we did not systematically assess the impact that any of these factors had; future work is needed to explore the impact of factors, such as compensation and survey length, on participant completion and compliance rates. In addition, many participants in this sample did not endorse poor HRQOL at baseline, nor high levels of supervision required, nor high levels of assistance with activities required. This could mean that this sample is not experiencing high levels of strain as we had anticipated, and therefore, they may have been higher functioning than the general care partner population, such that they had less room for improvement, or they had more capacity or time to complete the intensive study activities. In addition, there was no clear front-runner for the clustering analyses. As such, we elected to explore the model that we felt best represented the data, but we acknowledge that this selection might represent an overfitting of the data.

Conclusions

Overall, the results from this study indicated that although person-specific factors influence completion and compliance rates, it is still reasonable to expect high rates of compliance with an associated thoughtful study design that uses flexibility, tailoring, and financial incentives. Teasing apart the differential impact of the different study design elements on study participation will be a focus of future work. Furthermore, consistent with other literature, we found that disadvantaged groups (such as racial or ethnic minorities, single, or nontraditional caregivers) were more likely to have higher rates of missing data than their majority counterparts, further exemplifying the need for more focused work on understanding the reason for these lower rates and using methods to improve compliance among these groups.

Acknowledgments

We thank the investigators, coordinators, and research associates or assistants who worked on this study, the study participants, and the organizations who supported recruitment efforts. Site Investigators and Coordinators: NEC, SWC, ZW, SS, CMG, Angela Lyden, Nikki Hubbard, Abigail Biddix, JAM (University of Michigan, Ann Arbor, MI); AMS (Baylor College of Medicine and TIRR Memorial Hermann, Houston, TX), Jay Bogaards (TIRR Memorial Hermann, Houston, TX). Work on this manuscript was supported by grants R01NR013658 from the National Institutes of Health, National Institute of Nursing Research, R01HL146354 and K24HL156896 from the National Heart, Lung, and Blood Institute, and UL1TR002240 from the National Center for Advancing Translational Sciences. This work was also supported by the University of Michigan Institute for Healthcare Policy and Innovation. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

NEC was the principal investigator for this project and was responsible for conceptualization, oversight of the statistical analysis, and drafted methods, results, and discussion sections of the initial manuscript; she also collated and addressed coauthor feedback. JT, SWC, ZW, SS, and AMS were coinvestigators on this project and provided input into project conceptualization and funding acquisition. JT conducted the analyses and drafted all manuscript tables and figures. ZW gave specific input to data analysis and interpretation, including data visualization. WLL drafted the introduction and assisted with compiling and addressing coauthor feedback. SWC, AMS, SS, JAM, and CMG critically scrutinized the first draft and provided comments. JAM and CMG gave specific input to the methods section and provided oversight to the study staff. All authors read, commented on, and approved the final manuscript prior to submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Linear regression analyses (each cell indicates the *P* value for a separate model; italics indicate $P<.05$)
[DOCX File, 26 KB - mhealth_v13i1e73772_app1.docx]

Multimedia Appendix 2
Assessment of model fit and parsimony.
[DOCX File, 14 KB - mhealth_v13i1e73772_app2.docx]

Checklist 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.

XSL•FO
RenderX

[PDF File, 1686 KB - mhealth_v13i1e73772_app3.pdf]

References

1. Sosin DM, Sniezek JE, Waxweiler RJ. Trends in death associated with traumatic brain injury, 1979 through 1992. Success and failure. JAMA 1995 Jun 14;273(22):1778-1780. [Medline: 7769773]
2. Thurman D, Coronado V, Selassie AW. Epidemiology of traumatic brain injury. In: Zasler ND, Katz DI, Zafonte RD, editors. Brain Injury Medicine: Principles and Practice: Demos: New York; 2007:373-405.
3. Carlozzi NE, Brickell TA, French LM, et al. Caring for our wounded warriors: a qualitative examination of health-related quality of life in caregivers of individuals with military-related traumatic brain injury. J Rehabil Res Dev 2016;53(6):669-680. [doi: 10.1682/JRRD.2015.07.0136] [Medline: 27997672]
4. Marsh NV, Kersel DA, Havill JA, Sleigh JW. Caregiver burden during the year following severe traumatic brain injury. J Clin Exp Neuropsychol 2002 Jun;24(4):434-447. [doi: 10.1076/jcen.24.4.434.1030] [Medline: 12187457]
5. Carlozzi NE, Kratz AL, Sander AM, et al. Health-related quality of life in caregivers of individuals with traumatic brain injury: development of a conceptual model. Arch Phys Med Rehabil 2015 Jan;96(1):105-113. [doi: 10.1016/j.apmr.2014.08.021] [Medline: 25239281]
6. Ponsford J, Schönberger M. Family functioning and emotional state two and five years after traumatic brain injury. J Int Neuropsychol Soc 2010 Mar;16(2):306-317. [doi: 10.1017/S1355617709991342] [Medline: 20128950]
7. Kreutzer JS, Rapport LJ, Marwitz JH, et al. Caregivers' well-being after traumatic brain injury: a multicenter prospective investigation. Arch Phys Med Rehabil 2009 Jun;90(6):939-946. [doi: 10.1016/j.apmr.2009.01.010] [Medline: 19480869]
8. Sander AM, Caroselli JS, High WM Jr, Becker C, Neese L, Scheibel R. Relationship of family functioning to progress in a post-acute rehabilitation programme following traumatic brain injury. Brain Inj 2002 Aug;16(8):649-657. [doi: 10.1080/02699050210128889] [Medline: 12182162]
9. Schönberger M, Ponsford J, Olver J, Ponsford M. A longitudinal study of family functioning after TBI and relatives' emotional status. Neuropsychol Rehabil 2010 Dec;20(6):813-829. [doi: 10.1080/09602011003620077] [Medline: 20859828]
10. Vangel SJJ, Rapport LJ, Hanks RA. Effects of family and caregiver psychosocial functioning on outcomes in persons with traumatic brain injury. J Head Trauma Rehabil 2011;26(1):20-29. [doi: 10.1097/HTR.0b013e318204a70d] [Medline: 21209560]
11. Sander AM, Maestas KL, Sherer M, Malec JF, Nakase-Richardson R. Relationship of caregiver and family functioning to participation outcomes after postacute rehabilitation for traumatic brain injury: a multicenter investigation. Arch Phys Med Rehabil 2012 May;93(5):842-848. [doi: 10.1016/j.apmr.2011.11.031] [Medline: 22417896]
12. Bendixen RM, Fairman AD, Karavolis M, Sullivan C, Parmanto B. A user-centered approach: understanding client and caregiver needs and preferences in the development of mhealth apps for self-management. JMIR Mhealth Uhealth 2017 Sep 26;5(9):e141. [doi: 10.2196/mhealth.7136] [Medline: 28951378]
13. Avramović P, Rietdijk R, Kenny B, Power E, Togher L. Developing a digital health intervention for conversation skills after brain injury (convers-ABI-lity) using a collaborative approach: mixed methods study. J Med Internet Res 2023 Aug 9;25:e45240. [doi: 10.2196/45240] [Medline: 37556179]
14. Jia SS, Allman-Farinelli M, Roy R, et al. Using mobile ecological momentary assessment to understand consumption and context around online food delivery use: pilot feasibility and acceptability study. JMIR Mhealth Uhealth 2023 Nov 29;11:e49135. [doi: 10.2196/49135] [Medline: 38019563]
15. Trang K, Le LX, Brown CA, et al. Feasibility, acceptability, and design of a mobile ecological momentary assessment for high-risk men who have sex with men in Hanoi, Vietnam: qualitative study. JMIR Form Res 2022 Jan 27;6(1):e30360. [doi: 10.2196/30360] [Medline: 35084340]
16. Solk P, Gavin K, Fanning J, et al. Feasibility and acceptability of intensive longitudinal data collection of activity and patient-reported outcomes during chemotherapy for breast cancer. Qual Life Res 2019 Dec;28(12):3333-3346. [doi: 10.1007/s11136-019-02278-7] [Medline: 31493269]
17. Kratz AL, Kalpakjian CZ, Hanks RA. Are intensive data collection methods in pain research feasible in those with physical disability? A study in persons with chronic pain and spinal cord injury. Qual Life Res 2017 Mar;26(3):587-600. [doi: 10.1007/s11136-016-1494-0] [Medline: 28097459]
18. Smiley SL, Elmasry H, Webb Hooper M, Niaura RS, Hamilton AB, Milburn NG. Feasibility of ecological momentary assessment of daily sexting and substance use among young adult African American gay and bisexual men: a pilot study. JMIR Res Protoc 2017 Feb 2;6(2):e9. [doi: 10.2196/resprot.6520] [Medline: 28153816]
19. Ponnada A, Thapa-Chhetry B, Manjourides J, Intille S. Measuring criterion validity of microinteraction ecological momentary assessment (Micro-EMA): exploratory pilot study with physical activity measurement. JMIR Mhealth Uhealth 2021 Mar 10;9(3):e23391. [doi: 10.2196/23391] [Medline: 33688843]
20. Elbin RJ, Durfee KJ, Womble MN, et al. Documenting compliance and symptom reactivity for ambulatory assessment methodology following concussion in adolescents and young adults: feasibility of the mobile neurocognitive health (MNCH) project. J Head Trauma Rehabil 2025;40(1):E87-E95. [doi: 10.1097/HTR.0000000000000977] [Medline: 38922030]

XSL•FO
RenderX

21. Tonkin S, Gass J, Wray J, et al. Evaluating declines in compliance with ecological momentary assessment in longitudinal health behavior research: analyses from a clinical trial. J Med Internet Res 2023 Jun 22;25:e43826. [doi: 10.2196/43826] [Medline: 37347538]
22. Niznik T, Ehlke SJ, Mermelstein R, et al. Parameters of EMA compliance and self-reported reactivity in a longitudinal study of young adult cannabis and tobacco co-users. Cannabis 2023;6(2):47-61. [doi: 10.26828/cannabis/2023/000139] [Medline: 37484048]
23. Juengst SB, Graham KM, Pulantara IW, et al. Pilot feasibility of an mHealth system for conducting ecological momentary assessment of mood-related symptoms following traumatic brain injury. Brain Inj 2015;29(11):1351-1361. [doi: 10.3109/02699052.2015.1045031] [Medline: 26287756]
24. Krohn H, Guintivano J, Frische R, Steed J, Rackers H, Meltzer-Brody S. App-based ecological momentary assessment to enhance clinical care for postpartum depression: pilot acceptability study. JMIR Form Res 2022 Mar 23;6(3):e28081. [doi: 10.2196/28081] [Medline: 35319483]
25. Slade C, Benzo RM, Washington P. Design guidelines for improving mobile sensing data collection: prospective mixed methods study. J Med Internet Res 2024 Nov 18;26:e55694. [doi: 10.2196/55694] [Medline: 39556828]
26. Joo J, Devkota J, Stone BM, et al. Predictors of participant compliance with ecological momentary assessment among individuals with chronic pain who are using cannabis and opioids. Internet Interventions 2024 Dec;38:100784. [doi: 10.1016/j.invent.2024.100784] [Medline: 39611051]
27. Yang C, Linas B, Kirk G, et al. Feasibility and acceptability of smartphone-based ecological momentary assessment of alcohol use among African American men who have sex with men in Baltimore. JMIR Mhealth Uhealth 2015 Jun 17;3(2):e67. [doi: 10.2196/mhealth.4344] [Medline: 26085078]
28. Beres LK, Mbabali I, Anok A, et al. Acceptability and feasibility of mobile phone-based ecological momentary assessment and intervention in Uganda: a pilot randomized controlled trial. PLoS One 2022;17(8):e0273228. [doi: 10.1371/journal.pone.0273228] [Medline: 36018846]
29. Sanjuan PM, Pearson MR, Poremba C, Amaro HDLA, Leeman L. An ecological momentary assessment study examining posttraumatic stress disorder symptoms, prenatal bonding, and substance use among pregnant women. Drug Alcohol Depend 2019 Feb 1;195:33-39. [doi: 10.1016/j.drugalcdep.2018.11.019] [Medline: 30572290]
30. Laborde CR, Cenko E, Mardini MT, et al. Satisfaction, usability, and compliance with the use of smartwatches for ecological momentary assessment of knee osteoarthritis symptoms in older adults: usability study. JMIR Aging 2021 Jul 14;4(3):e24553. [doi: 10.2196/24553] [Medline: 34259638]
31. Howard AL, Lamb M. Compliance trends in a 14-week ecological momentary assessment study of undergraduate alcohol drinkers. Assessment 2024 Mar;31(2):277-290. [doi: 10.1177/10731911231159937] [Medline: 36914966]
32. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
33. Martinez GJ, Mattingly SM, Robles-Granda P, et al. Predicting participant compliance with fitness tracker wearing and ecological momentary assessment protocols in information workers: observational study. JMIR Mhealth Uhealth 2021 Nov 12;9(11):e22218. [doi: 10.2196/22218] [Medline: 34766911]
34. Businelle MS, Hébert ET, Shi D, et al. Investigating best practices for ecological momentary assessment: nationwide factorial experiment. J Med Internet Res 2024 Aug 12;26:e50275. [doi: 10.2196/50275] [Medline: 39133915]
35. Shimotsu S, Roehrl A, McCarty M, et al. Increased likelihood of missed appointments (“No Shows”) for racial/ethnic minorities in a safety net health system. J Prim Care Community Health 2016 Jan;7(1):38-40. [doi: 10.1177/2150131915599980] [Medline: 26286688]
36. Xie Z, St Clair P, Goldman DP, Joyce G. Racial and ethnic disparities in medication adherence among privately insured patients in the United States. PLoS One 2019;14(2):e0212117. [doi: 10.1371/journal.pone.0212117] [Medline: 30763400]
37. Fabian AJ, Balado RL, Chase MG, Nemec EC. Patient-provider race concordance and medication adherence: a systematic review. J Racial Ethn Health Disparities 2025 Feb 27. [doi: 10.1007/s40615-025-02330-y] [Medline: 40016592]
38. Pardhan S, Sehmbi T, Wijewickrama R, Onumajuru H, Piyasena MP. Barriers and facilitators for engaging underrepresented ethnic minority populations in healthcare research: an umbrella review. Int J Equity Health 2025 Mar 12;24(1):70. [doi: 10.1186/s12939-025-02431-4] [Medline: 40075407]
39. Su J, Dugas M, Guo X, Gao GG. Influence of personality on mHealth use in patients with diabetes: prospective pilot study. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e17709. [doi: 10.2196/17709] [Medline: 32773382]
40. Hermosa JLR, Gomila AF, Maestu LP, et al. Compliance and utility of a smartphone app for the detection of exacerbations in patients with chronic obstructive pulmonary disease: cohort study. JMIR Mhealth Uhealth 2020 Mar 19;8(3):e15699. [doi: 10.2196/15699] [Medline: 32191213]
41. Unger JM, Gralow JR, Albain KS, Ramsey SD, Hershman DL. Patient income level and cancer clinical trial participation: a prospective survey study. JAMA Oncol 2016 Jan;2(1):137-139. [doi: 10.1001/jamaoncol.2015.3924] [Medline: 26468994]
42. Unger JM, Hershman DL, Albain KS, et al. Patient income level and cancer clinical trial participation. J Clin Oncol 2013 Feb 10;31(5):536-542. [doi: 10.1200/JCO.2012.45.4553] [Medline: 23295802]

XSL•FO
RenderX

43. Clark LT, Watkins L, Piña IL, et al. Corrigendum to "Increasing diversity in clinical trials: overcoming critical barriers''. [Current Problems in Cardiology, Volume 44, Issue 5 (2019) 148–172]. Curr Probl Cardiol 2021 Mar;46(3):100647. [doi: 10.1016/j.cpcardiol.2020.100647]
44. Williams CP, Everson NS, Shelburne N, Norton WE. Demographic and health behavior factors associated with clinical trial invitation and participation in the United States. JAMA Netw Open 2021 Sep 1;4(9):e2127792. [doi: 10.1001/jamanetworkopen.2021.27792] [Medline: 34586365]
45. Harzand A, Witbrodt B, Davis-Watts ML, et al. Feasibility of a smartphone-enabled cardiac rehabilitation program in male veterans with previous clinical evidence of coronary heart disease. Am J Cardiol 2018 Nov 1;122(9):1471-1476. [doi: 10.1016/j.amjcard.2018.07.028] [Medline: 30217377]
46. Chen YS, Wong JE, Ayob AF, Othman NE, Poh BK. Can Malaysian young adults report dietary intake using a food diary mobile application? A pilot study on acceptability and compliance. Nutrients 2017 Jan 13;9(1):62. [doi: 10.3390/nu9010062] [Medline: 28098770]
47. Crafoord MT, Fjell M, Sundberg K, Nilsson M, Langius-Eklöf A. Correction: engagement in an interactive app for symptom self-management during treatment in patients with breast or prostate cancer: mixed methods study. J Med Internet Res 2021 Oct 12;23(10):e33140. [doi: 10.2196/33140] [Medline: 34637396]
48. Crafoord MT, Fjell M, Sundberg K, Nilsson M, Langius-Eklöf A. Engagement in an interactive app for symptom self-management during treatment in patients with breast or prostate cancer: mixed methods study. J Med Internet Res 2020 Aug 10;22(8):e17058. [doi: 10.2196/17058] [Medline: 32663140]
49. Min YH, Lee JW, Shin YW, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014 May 23;16(5):e135. [doi: 10.2196/jmir.3421] [Medline: 24860070]
50. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020 Nov 6;8(11):e22212. [doi: 10.2196/22212] [Medline: 32975198]
51. Hendrie GA, Hussain MS, Brindal E, James-Martin G, Williams G, Crook A. Impact of a mobile phone app to increase vegetable consumption and variety in adults: large-scale community cohort study. JMIR Mhealth Uhealth 2020 Apr 17;8(4):e14726. [doi: 10.2196/14726] [Medline: 32301739]
52. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138. [doi: 10.1037/ocp0000118] [Medline: 29723001]
53. Göransson C, Wengström Y, Hälleberg-Nyman M, Langius-Eklöf A, Ziegert K, Blomberg K. An app for supporting older people receiving home care—usage, aspects of health and health literacy: a quasi-experimental study. BMC Med Inform Decis Mak 2020 Sep 15;20(1):226. [doi: 10.1186/s12911-020-01246-3] [Medline: 32933500]
54. Seng EK, Prieto P, Boucher G, Vives-Mestres M. Anxiety, incentives, and adherence to self-monitoring on a mobile health platform: a naturalistic longitudinal cohort study in people with headache. Headache 2018 Nov;58(10):1541-1555. [doi: 10.1111/head.13422] [Medline: 30334248]
55. Carlozzi NE, Schilling S, Freedman J, Kalpakjian CZ, Kratz AL. The reliability of end of day and ecological momentary assessments of pain and pain interference in individuals with spinal cord injury. Qual Life Res 2018 Nov;27(11):3003-3012. [doi: 10.1007/s11136-018-1952-y] [Medline: 30073468]
56. Carlozzi NE, Choi SW, Wu Z, et al. An app-based just-in-time-adaptive self-management intervention for care partners: the CareQOL feasibility pilot study. Rehabil Psychol 2022 Nov;67(4):497-512. [doi: 10.1037/rep0000472] [Medline: 36355640]
57. Baroudi L, Zernicke RF, Tewari M, Carlozzi NE, Choi SW, Cain SM. Using wear time for the analysis of consumer-grade wearables' data: case study using Fitbit data. JMIR Mhealth Uhealth 2025 Mar 21;13:e46149. [doi: 10.2196/46149] [Medline: 40116717]
58. Hart T, O'Neil-Pirozzi TM, Williams KD, Rapport LJ, Hammond F, Kreutzer J. Racial differences in caregiving patterns, caregiver emotional function, and sources of emotional support following traumatic brain injury. J Head Trauma Rehabil 2007;22(2):122-131. [doi: 10.1097/01.HTR.0000265100.37059.44] [Medline: 17414314]
59. Carlozzi NE, Troost JP, Sen S, et al. Improving outcomes for care partners of individuals with traumatic brain injury: results for a mHealth randomized control trial of the CareQOL app. Arch Phys Med Rehabil 2025 Apr;106(4):548-561. [doi: 10.1016/j.apmr.2024.12.022] [Medline: 39798892]
60. 2020 census results. United States Census Bureau. 2020. URL: https://www.census.gov/programs-surveys/decennial-census/decade/2020/2020-census-results.html [accessed 2025-07-14]
61. Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting information retrieval from electronic health records: a report of University of Michigan's nine-year experience in developing and using the electronic medical record search engine (EMERSE). J Biomed Inform 2015 Jun;55:290-300. [doi: 10.1016/j.jbi.2015.05.003] [Medline: 25979153]
62. Carlozzi NE, Sander AM, Choi SW, et al. Improving outcomes for care partners of persons with traumatic brain injury: protocol for a randomized control trial of a just-in-time-adaptive self-management intervention. PLoS ONE 2022;17(6):e0268726. [doi: 10.1371/journal.pone.0268726] [Medline: 35679283]

XSL•FO
RenderX

63. Boake C. Supervision rating scale: a measure of functional outcome from brain injury. Arch Phys Med Rehabil 1996 Aug;77(8):765-772. [doi: 10.1016/s0003-9993(96)90254-3] [Medline: 8702369]
64. Malec J. Mayo-Portland Adaptability Inventory. The Center for Outcome Measurement in Brain Injury. 2005. URL: http://www.tbims.org/combi/mpai [accessed 2025-07-14]
65. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
66. Carlozzi NE, Kallen MA, Ianni PA, et al. The development of a new computer-adaptive test to evaluate strain in caregivers of individuals with TBI: TBI-CareQOL caregiver strain. Arch Phys Med Rehabil 2019 Apr;100(4S):S13-S21. [doi: 10.1016/j.apmr.2018.05.033] [Medline: 29966647]
67. Carlozzi NE, Kallen MA, Hanks R, et al. The TBI-CareQOL measurement system: development and preliminary validation of health-related quality of life measures for caregivers of civilians and service members/veterans with traumatic brain injury. Arch Phys Med Rehabil 2019 Apr;100(4):S1-S12. [doi: 10.1016/j.apmr.2018.08.175]
68. Carlozzi NE, Kallen MA, Sander AM, et al. The development of a new computer adaptive test to evaluate anxiety in caregivers of individuals with traumatic brain injury: TBI-CareQOL caregiver-specific anxiety. Arch Phys Med Rehabil 2019 Apr;100(4):S22-S30. [doi: 10.1016/j.apmr.2018.05.027]
69. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194. [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
70. Salsman JM, Butt Z, Pilkonis PA, et al. Emotion assessment using the NIH Toolbox. Neurology (ECronicon) 2013 Mar 12;80(11 Suppl 3):S76-S86. [doi: 10.1212/WNL.0b013e3182872e11] [Medline: 23479549]
71. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009 Sep;18(7):873-880. [doi: 10.1007/s11136-009-9496-9] [Medline: 19543809]
72. Wagstaff KL, Cardie C, Rogers S, Schroedl S. Constrained k-means clustering with background knowledge. Presented at: Eighteenth International Conference on Machine Learning (ICML); Jun 28 to Jul 1, 2001; Williams College, Williamstown, MA, United States.
73. Edwards PJ, Roberts I, Clarke MJ, DiGuiseppi C, Woolf B, Perkins C. Methods to increase response to postal and electronic questionnaires. Cochrane Database Syst Rev 2023 Nov 30;11(11):MR000008. [doi: 10.1002/14651858.MR000008.pub5] [Medline: 38032037]
74. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. Cochrane Database Syst Rev 2009 Jul 8;2009(3):MR000008. [doi: 10.1002/14651858.MR000008.pub4] [Medline: 19588449]
75. Shavers-Hornaday VL, Lynch CF, Burmeister LF, Torner JC. Why are African Americans under-represented in medical research studies? Impediments to participation. Ethn Health 1997;2(1-2):31-45. [doi: 10.1080/13557858.1997.9961813] [Medline: 9395587]
76. Fisher JA, Kalbaugh CA. Challenging assumptions about minority participation in US clinical research. Am J Public Health 2011 Dec;101(12):2217-2222. [doi: 10.2105/AJPH.2011.300279] [Medline: 22021285]
77. Arnold KB, Hermos JA, Anderson KB, et al. Retention of Black and White participants in the selenium and vitamin E cancer prevention trial (SWOG-coordinated intergroup study S0000). Cancer Epidemiol Biomarkers Prev 2014 Dec;23(12):2895-2905. [doi: 10.1158/1055-9965.EPI-14-0724] [Medline: 25242051]
78. Kim H, Cutter GR, George B, Chen Y. Understanding and preventing loss to follow-up: experiences from the spinal cord injury model systems. Top Spinal Cord Inj Rehabil 2018;24(2):97-109. [doi: 10.1310/sci2402-97] [Medline: 29706754]
79. Feigelson HS, McMullen CK, Madrid S, et al. Optimizing patient-reported outcome and risk factor reporting from cancer survivors: a randomized trial of four different survey methods among colorectal cancer survivors. J Cancer Surviv 2017 Jun;11(3):393-400. [doi: 10.1007/s11764-017-0596-1] [Medline: 28084606]
80. Bolland AC, Tomek S, Bolland JM. Does missing data in studies of hard-to-reach populations bias results? Not necessarily. OJS 2017;07(02):264-289. [doi: 10.4236/ojs.2017.72021]
81. Hewitson I, White AE, Walter ES, Jervis RH. Using online surveys for routine campylobacter case investigations in Colorado, September 2020-December 2021. J Public Health Manag Pract 2024;30(5):718-727. [doi: 10.1097/PHH.0000000000001953] [Medline: 39041766]
82. Fontil V, Pacca L, Bellows BK, et al. Association of differences in treatment intensification, missed visits, and scheduled follow-up interval with racial or ethnic disparities in blood pressure control. JAMA Cardiol 2022 Feb 1;7(2):204-212. [doi: 10.1001/jamacardio.2021.4996] [Medline: 34878499]
83. Bulkley JE, O'Keeffe-Rosetti M, Wendel CS, et al. The effect of multiple recruitment contacts on response rates and patterns of missing data in a survey of bladder cancer survivors 6 months after cystectomy. Qual Life Res 2020 Apr;29(4):879-889. [doi: 10.1007/s11136-019-02379-3]
84. Atefi GL, Koh WQ, Kohl G, et al. Adherence to online interventions for family caregivers of people with dementia: a meta-analysis and systematic review. Am J Geriatr Psychiatry 2024 Oct;32(10):1271-1291. [doi: 10.1016/j.jagp.2024.04.008]

XSL•FO
RenderX

85. Byun E, Lerdal A, Gay CL, Lee KA. How adult caregiving impacts sleep: a systematic review. Curr Sleep Med Rep 2016;2(4):191-205. [doi: 10.1007/s40675-016-0058-8] [Medline: 31080704]
86. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/AJPH.2013.301706] [Medline: 24328648]

Abbreviations

EMA: ecological momentary assessment
HRQOL: health-related quality of life
IRB: institutional review board
mHealth: mobile health
PRO: patient-reported outcome
REDCap: Research Electronic Data Capture
SRS: Supervision Rating Scale
TBI: traumatic brain injury

Edited by G Badicu; submitted 11.03.25; peer-reviewed by J Mira, S Tonkin; revised version received 02.06.25; accepted 05.06.25; published 21.08.25.

Please cite as:
Carlozzi NE, Troost J, Lombard WL, Miner JA, Graves CM, Choi SW, Wu Z, Sen S, Sander AM
Completion and Compliance Rates for an Intensive mHealth Study Design to Promote Self-Awareness and Self-Care Among Care Partners of Individuals With Traumatic Brain Injury: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e73772
URL: https://mhealth.jmir.org/2025/1/e73772
doi:10.2196/73772

© Noelle E Carlozzi, Jonathan Troost, Wendy L Lombard, Jennifer A Miner, Christopher M Graves, Sung Won Choi, Zhenke Wu, Srijan Sen, Angelle M Sander. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Current Practice and Expert Perspectives on Cultural Adaptations of Digital Health Interventions: Qualitative Study

Vasileios Nittas[1], PhD; Sarah J Chavez[2], PhD; Paola Daniore[3], PhD

[1]Department of Behavioral and Social Sciences, Brown University, 121 South Main Street, Providence, RI, United States

[2]Brown University Center for Alcohol and Addiction Studies, Brown University, Providence, RI, United States

[3]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

Corresponding Author:
Vasileios Nittas, PhD
Department of Behavioral and Social Sciences, Brown University, 121 South Main Street, Providence, RI, United States

Abstract

Background: Some people are less likely to benefit from digital health interventions (DHIs) than others. Culture, along with other factors, contributes to these differences. DHIs that do not address a population's cultural norms or concerns are likely to be less effective. One way to create culturally sensitive DHIs is through cultural adaptations. Yet, there is currently little evidence-based guidance on when and how to adapt DHIs.

Objective: We aimed to capture the experiences of experts to understand the (1) current practices, (2) challenges, and (3) recommendations around culturally adapting DHIs.

Methods: We conducted semistructured interviews (n=15) via Zoom (Zoom Video Communications, Inc) between May and August 2023, with academic experts who have previously undertaken cultural adaptations of DHIs. Experts were identified through publications and snowball sampling. We used a thematic analytical approach, beginning with a preliminary deductive codebook and then following a three-stage analysis. All transcripts were coded with the MAXQDA (VERBI Software GmbH) software. Codes were reviewed, and similar or related codes were categorized into broader themes, consolidating one or multiple codes into a single topic.

Results: Our analysis produced 30 codes, which were categorized into (1) defining culture, (2) justifying the adaptation, (3) choosing the adaptation elements, (4) implementing the adaptation, (5) understanding the challenges, and (6) recommendations. Based on their experiences, experts recommended that (1) the adaptation team is multiprofessional, digitally competent, and culturally sensitive; (2) DHI users and (3) all other relevant stakeholders are continuously involved; and (4) the adaptations incorporate evaluations and knowledge exchange. They further emphasized that culturally adapted DHIs must be understandable, relatable, appealing, and easy to adhere to, ensuring that health technology and content reflect the target population's lived experiences, sociodemographic characteristics, and digital literacy. When asked which elements of cultural DHI adaptations, the most common responses were language, lived experience, and technology. Responses revealed five common DHI-relevant challenges, including (1) technology, (2) uncertainty, (3) user involvement, (4) communication, and (5) evaluation and sustainability.

Conclusions: The cultural adaptation of DHIs was described as an iterative, often unstructured, and resource-intensive process that requires careful justification and a solid understanding of the culture and the specific cultural group for which it is implemented. Our interviews confirmed the absence of technology-specific frameworks to guide the cultural adaptations of DHIs. Based on our findings, such a framework should guide the choice of the correct definition of culture and the criteria for assessing the need to adapt. It should also offer tools to drive stakeholder engagement, prioritize adaptation elements, and address common challenges.

(JMIR Mhealth Uhealth 2025;13:e59965) doi:10.2196/59965

KEYWORDS

cultural adaptation; digital health; health equity; intervention; interventions; practice; practices; expert; experts; perspective; perspectives; experience; experiences; qualitative study; interview; interviews; thematic analysis; web-based; mobile health; mHealth; cultural tailoring; digital health intervention

Introduction

Digital health interventions (DHIs) use information and communication technologies (ICTs) to prevent, monitor, manage, and treat health conditions [1]. DHIs are often used to support individuals in developing healthy habits (eg, smoking cessation and healthy eating) and improve outcomes for those living with chronic conditions [1]. A key feature of DHIs is that they can provide access to health information and treatments

XSL•FO
RenderX

without the need for traditional health care intermediaries such as insurance companies and clinicians [1,2]. DHIs are spearheaded by web and mobile apps, social networking platforms, and self-monitoring devices like fitness trackers, sleep monitors, and wearables (electronic health devices work on the body, such as smartwatches) [2]. Recent systematic reviews indicate that DHIs can be effective and cost-effective in reducing risk and enhancing health outcomes across a variety of conditions, such as diabetes, cardiovascular health, mental health, and addictions [3-5].

Despite increasing access to the internet, mobile devices, and other technologies, medically underserved populations are less likely to benefit from DHIs than others [6-9]. The term "medically underserved" describes those (eg, racial or ethnic minorities) who have limited access to primary health care, primarily due to sociocultural, economic, and geographic vulnerabilities [10]. Two recent systematic reviews on DHIs for physical activity and other weight-related behaviors have demonstrated that the use and effectiveness of these interventions vary across several characteristics, including race, ethnicity, and socioeconomic status [11,12]. Other studies found that African American and Hispanic individuals are less likely to engage with DHIs than their White counterparts [13,14]. Together with other factors, culture has been shown to play a role in these differences because most DHIs are most often designed for dominant cultural groups, limiting their appropriateness for marginalized populations. DHIs that do not address a population's cultural norms or concerns, thus, are not culturally sensitive, are likely to be less accepted by that population, and therefore, underused or engaged with [10,13,15]. Although culture can influence how people use and engage with a DHI, very few DHIs are designed to be culturally sensitive [8,15]. This often results in DHIs that inherently exclude certain population groups, further exacerbating the digital and health inequities [8]. Although no single definition of culture exists, Castro et al [16] defined it as a group of people's shared perspectives and practices.

Ensuring that DHIs address the sociocultural norms of a population is challenging. One way to create culturally sensitive DHIs is through adaptations. Cultural adaptations can be defined as interventions that tailor existing, often evidence-based technologies to align with a specific group's cultural identity, including cultural norms, beliefs, values, and practices [16]. In our previous work, we argued that cultural adaptations may increase DHIs' reach, engagement, and outcomes for medically underserved subgroups [17]. However, we emphasized that there is currently little evidence-based guidance on when and how to adapt DHIs [17]. The broader literature on cultural adaptations offers various models designed to guide the adaptation process for evidence-based health care interventions [18,19]. These models specify various steps in the cultural adaptation process, such as (1) assessing the evidence base, (2) creating a preliminary adaptation design, (3) testing preliminary adaptations, and (4) refining them into final adaptations. While many of these models offer broad guidance applicable to the cultural adaptation of DHIs, they do not sufficiently recognize and tackle the unique challenges associated with their use of technology. For example, creating a preliminary adaptation design of a mental health intervention will differ drastically when the intervention is delivered physically compared to it being delivered through a mobile phone app that might require expensive and time-consuming software changes. Nevertheless, there have been many noteworthy attempts to adapt DHIs described in the literature to date, culturally, generating essential knowledge and experience that has the potential to contribute to the development of future guidance [20,21].

In this study, we interviewed researchers who have conducted cultural adaptations of DHIs. Through these interviews, we aimed to capture current practices, challenges, gaps, and recommendations for future adaptations. We specifically asked researchers about (1) why and how they adapted their DHIs and defined culture, (2) how they implemented those adaptations, (3) which elements they chose for adaptation and the criteria for their selection, (4) the challenges they faced during adaptation, and (5) their recommendations for future cultural adaptations of DHIs. We hoped that understanding these experiences might help us identify general principles that could be used to develop broader recommendations to help guide future cultural adaptations of DHIs.

Methods

Study Design

We conducted semistructured 45-minute interviews (n=15) with researchers who have previously undertaken cultural adaptations of DHIs. Researchers (hereafter referred to as experts) were invited to participate in this study if they (1) developed and implemented at least one cultural adaptation of a DHI (globally) and (2) either currently hold an academic appointment or have had a past academic appointment.

Participants and Recruitment

We initially identified eligible experts through academic publications. In March 2023, we searched PubMed and Google Scholar using keywords related to DHIs, culture, and adaptations. We used the following keywords: "digital health," "digital health intervention," "mobile health," "web-based health," "e-health," "culture," "socio-cultural," "cultural tailoring," "cultural adaptation," and "cultural sensitivity." Relevant publications had to address an actual cultural adaptation of a DHI, irrespective of the health field or population. One reviewer (VN) screened and selected all relevant studies. We screened 138 studies, of which 37 were deemed relevant and appropriate and were used for extracting all corresponding authors' names and email addresses. Invitations were emailed to participants, where we briefly introduced the project's aims and participation requirements. In case of nonresponse, we sent reminder emails 2 - 3 weeks after initial contact. Those who expressed interest were emailed a bulleted summary of the project and asked to provide their feedback on availability. We identified additional experts through snowball sampling, which involved asking experts at the end of their interview if they knew of other researchers who may be eligible. Inclusion criteria were verified by screening a recommended expert's publication list to ensure experts had (1) at least one publication specific to the cultural adaptation of a DHI and (2) a past or current academic appointment. We

XSL•FO
RenderX

purposively chose interviews over a literature review because the current DHI literature primarily provides superficial descriptions of the adaptation processes. Instead, interviewing experts allows for a deeper and more nuanced understanding of current practice, challenges, and gaps, and provides experience-based recommendations. Interviews were concluded once we reached thematic saturation. We defined thematic saturation as reaching a state where conducting new research does not provide substantially new information or additional conceptual depth but rather a repetition of themes. We stopped interviewing once we felt that a significant amount of information was repeated, which occurred after 13 interviews. We then conducted two additional interviews to confirm that thematic saturation was reached.

Data Collection

We used a semistructured interview guide developed by the research team and informed by a literature review that we previously conducted (Multimedia Appendix 1) [17]. We used the main topics that emerged in the literature review to develop the interview sections and questions. Interviews were divided into four parts that were specific to DHIs. The first part was an introduction (eg, "tell us about your background and overall experience in cultural adaptations"); the second part focused on the justification of cultural DHI adaptations (eg, what made you decide to adapt?). Furthermore, the third part consisted of questions around adaptation elements (eg, what elements did you adapt and how did you select these elements?), and the fourth part addressed challenges, solutions, and recommendations (eg, what challenges did you face during the adaptation process?"). Based on the first two interviews (pilot), we iteratively adjusted the interview guide. Specifically, the wording of parts three and four was adjusted to be more targeted and more precise, with a series of narrower subquestions.

One interviewer (VN; epidemiologist) with previous experience in qualitative research and interviewing conducted the interviews via Zoom (Zoom Video Communications, Inc) between May and July 2023. Each interview began with a verbal summary of basic information about the study and its objectives. Experts had the opportunity to ask questions. All sessions were recorded and transcribed with Zoom's transcription software. Additional notes were taken by the interviewer during each interview. Each transcription underwent quality checking and was corrected by one of the research team members (VN). Before analyzing the data, we deidentified all transcripts and removed any references to names, institutions, and project names. Each expert was given a unique identification number linked to their transcripts. A week after transcribing, we deleted all recordings.

Analysis and Synthesis

We applied an iterative, thematic analytical approach [22]. We initially created a preliminary deductive codebook based on our prior work [17]. Specifically, we used the challenges identified in our review to create initial codes (eg, adaptation justification). We then followed a three-stage analysis to ensure themes were understood and developed thoroughly. First, we read and reread all the interviews to familiarize ourselves with the data. That stage led to inductive modifications of the codebook. Second, all transcripts were coded by VN using the MAXQDA software (VERBI Software GmbH). Third, codes were reviewed, and similar or related codes were categorized into broader themes, consolidating one or more codes into a single unit topic. We then described and narratively synthesized these themes. Following standard practice, a second coder (PD) independently coded about 15% (n=2) of our interviews (PD) to ensure that no critical codes were missed [23]. Finally, we synthesized our data and created themes that represented what most experts were stating. Additionally, we elaborated on emerging findings that warrant further investigation.

Ethical Considerations

An ethics approval and informed consent were not required as this study does not meet Brown University's institutional definition of human participant research [24]. An ethics application was submitted and was dismissed as nonhuman participant research. All study data have been anonymized and deidentified. Recordings have been permanently deleted. Each expert received a US $100 voucher.

Results

Overview

We invited 34 researchers to participate in our study. A total of 25 (74%) researchers responded to our email, and 15 (44%) researchers agreed to be interviewed. Of those, 11 (70%) researchers were directly identified through publications, and 4 (30%) were through referrals. In total, 13 of the experts were female. Participant characteristics are listed in Table 1.

XSL•FO
RenderX

Table . Participant characteristics.

ID	Country	Area of expertise	Adapted technology	Target population
1	Brazil	Public health	Mobile app	Pregnant woman
2	United States	Oncology	Web-based decision aid	Latino men
3	Germany	Psychology	Mobile interventions	Low-income setting
4	United Kingdom	Caregiving	eHealth training	Young carers
5	Switzerland	Physical activity	Mobile app	English speakers
6	New Zealand	Public Health	Text-based app	Indigenous
7	Canada	Nursing	Web-based program	Low-income setting
8	United Kingdom	Psychology	eHealth training	Minority carers
9	Sweden	Psychology	Web-based intervention	Minority youth
10	Spain	Caregiving	eHealth training	Informal caregivers
11	Switzerland	Psychology	Mobile app	Migrants
12	Germany	Psychology	Digital intervention	Refugees
13	Germany	Psychology	Mobile app	Refugees
14	Greece	Psychology	Web-based program	Informal caregivers
15	Spain	Public health	Web-based program	Adolescents

Our analysis yielded 30 codes, which we categorized into six broader themes aligned with our interview guide. Table 2 provides a description of all themes and subthemes, which includes: (1) defining culture, (2) justifying the adaptation, (3) choosing the adaptation elements, (4) implementing the adaptation, (5) understanding the challenges, and (6) recommendations. The first theme addresses topics related to all preparatory steps before a cultural adaptation is implemented. Themes 2 to 5 address issues directly related to implementation.

XSL•FO
RenderX

Table . Definitions of themes and subthemes.

Themes and subthemes	Description
1. Defining culture	Main theme that includes all statements referring to the definition, perception, and conceptualization of culture.
2. Justifying the adaptation	Main theme that includes all statements referring to reasons for culturally adapting DHIs[a].
2.1. Pragmatic reasons	Subtheme that includes all opportunity-driven reasons to culturally adapt DHIs.
2.2. Equal access	Subtheme that includes reasons to culturally adapt DHIs that refer to equal access to technology.
2.3. Other ethical reasons	Subtheme that includes all other ethical reasons to culturally adapt DHIs.
3. Implementing the adaptation	Main theme that includes all statements referring to implementing cultural adaptations of DHIs.
3.1. The adaptation team	Subtheme that describes the role of adaptation team members in the implementation of adaptations.
3.2. User involvement	Subtheme that describes the role of user involvement in the implementation of adaptations.
3.3. Expert or stakeholder involvement	Subtheme that describes the consulting role of experts or stakeholders in the implementation of adaptations.
4. Choosing the adaptation elements	Main theme that includes all statements referring to selecting and prioritizing what DHI elements to adapt.
4.1. Language	Subtheme that describes language as a critical adaptation element.
4.2. Lived experience	Subtheme that describes lived experience as a critical adaptation element.
4.3. Technology	Subtheme that describes technology as a critical adaptation element.
5. Understanding the challenges	Main theme that includes all statements referring to challenges when adapting DHIs.
5.1. Technology	Subtheme that describes adaptation challenges related to technology and its functions.
5.2. Uncertainty	Subtheme that describes adaptation challenges related to missing knowledge or evidence and uncertainty
5.3. User involvement	Subtheme that describes challenges related to involving and engaging users in the adaptation process.
5.4. Communication	Subtheme that describes challenges related to communication during the adaptation process.
5.5. Evaluation and sustainability	Subtheme that describes challenges related to evaluating and sustainably maintaining an adapted DHI.
6. Recommendations	Main theme that includes all statements referring to expert recommendations for future cultural adaptations of DHIs.

[a]DHI: digital health intervention.

Defining Culture

We asked participants how they define culture, primarily in the context of their work. Most describe culture as a complex concept beyond ethnicity and place of origin. Instead, they emphasized (1) community, (2) shared experience, (3) shared values, (4) family, and (5) sociodemographics, as well as (6) shared context and living conditions. Of those experts who provided more details, defining culture (in the context of DHI adaptations) was described as an intuitive and pragmatic process often grounded in common knowledge (eg, what people think they know about culture) or the adaptation team's expertise of the target culture (eg, the culture of the focal population that a DHI is adapted for). Overall, many experts described culture as an influencer of one's perception of the world, which includes the perception of health and technology. One expert summarized that as "cultural and context," and described cultural adaptations as "contextual adaptations." One expert who has adapted a mobile mental health app for refugees described:

> *But what you have to do is you have to do contextual adaptation, you have to understand the context in which the people grew up in which they live now, and in which they will use the app.* [ID13]

XSL•FO
RenderX

Justifying the Adaptation

Overview

We asked participants about valid reasons to culturally adapt to a DHI, as well as the reasons for their adaptations. Replies led to three subthemes: (1) pragmatic reasons, (2) providing equal access to DHIs for underserved populations, and (3) other ethical reasons.

Pragmatic Reasons

Some experts mainly mentioned the lack of guidance specific to cultural adaptations of DHIs. A participant (ID9) who has previously adapted a web-based health intervention for minority youth confirmed: "I don't think there is much in the existing research that can guide us in that." Therefore, their decisions to adapt were often described as intuitive and common sense–based (eg, they believed they were needed and they made sense), as well as the result of arising opportunities, such as available funding for transferring a DHI to another community. One expert (ID8) who has worked on the adaptation of an eHealth training for young carers in the United Kingdom said: "Our adaptation came about through a recommendation from our funders."

Equal Access

One consistent theme that arose when justifying their adaptations was the provision of equal access. It was described as making digital health resources available to people who are expected to benefit but tend to be medically underserved and would not be reached otherwise. This was described as the case either because the nonadapted DHI is (1) not relevant or appealing, (2) not understandable, (3) offensive, (4) incompatible with prevailing norms and values, and (5) not aligned with the context and lived experiences. An expert who has adapted an eHealth training for minority young carers described:

> *There was very, very little in terms of resources for them, support in the caring world [...] there is a lot of evidence south Asian minorities, communities in the UK have a different look at what caring means, and because their family context is quite different, their openness or the way they like to receive the help is quite specific.* [ID4]

Other Ethical Reasons

One expert emphasized the ethical reasons for adaptation, including (1) eliminating Western biases from DHIs before implementing them in more culturally diverse environments and (2) enhancing the visibility of individuals who are otherwise socially excluded groups. One expert said:

> *But we knew there wasn't anything for younger carers of people living with dementia [...] we could do something that at least was going to help increase the visibility of the group.* [ID4]

Another common theme was the improvement of DHI engagement and adherence, specifically for people who have otherwise previously shown low engagement intervention rates. Here, adaptations were described as synonymous with an elevated user experience, which ensures that people interact with a DHI long enough for it to have an effect. An expert who had previously adapted a mental health app for refugees described:

> *So, the ultimate goal for cultural adaptation is really to make sure that the core of the intervention can even take effect. And people do not drop out before they get to the point.* [ID13]

Two additional adaptation reasons that experts gave were (1) to understand whether a DHI is suitable for implementation across sociocultural settings and (2) to implement a DHI that is initially generic. For instance, a DHI that has been successfully adapted at least once indicates that it is transferable and potentially scalable across sociocultural contexts. Similarly, many DHIs were designed to be adapted (eg, some generic interventions developed by the World Health Organization [WHO]), which was the case for many of our experts. One mental health expert said:

> *The intervention in itself is set up to be adapted. So it is developed by the World Health Organization that is recommending that researchers and research groups pick this up and adapt it.* [ID8]

Implementing the Adaptation

Overview

Experts' views and experiences on the implementation of the cultural adaptation focused on (1) the adaptation team, (2) user involvement, and (3) expert and stakeholder involvement. The adaptation process of DHIs was described as dynamic, iterative, often unstructured, and trial-and-error-based.

The Adaptation Team

Overall, responses revealed two critical characteristics of a good cultural adaptation team. First, a multiprofessional background, which was described as a team with a diverse skill set and professional expertise, including experts in health and technology, such as hardware and software designers. An expert who adapted a mobile health intervention for refugees described:

> *So, you need lots of people with very different backgrounds, and you need them to work together efficiently [...] If you want to have a proper user experience, you need a designer. You need people who know something about user interfaces.* [ID13]

Second, cross-cultural competence, which includes language skills (written and verbal), is another factor in successful DHI adaptations. Experts described that this could be experts (eg, researchers, citizens, or community members) with the same cultural background (eg, lived experience) as the focal population. Cross-cultural competence was described as enhancing cultural credibility, which facilitates the reach and engagement of the target population in the adaptation process. Additional specific roles and expertise requirements were unclear, aside from providing cultural expertise and facilitating communication with the focal point population. Some experts highlighted that cultural adaptation teams often lack these characteristics due to a lack of time and resources.

XSL•FO
RenderX

User Involvement

Overall, experts viewed user involvement (of the focal cultural group) as the gold standard approach to culturally adapting DHIs. Good user involvement was described as (1) timely (from the start of the adaptation), (2) iterative and stepwise (allowing for multiple rounds of feedback), and (3) continuous (throughout the adaptation process). Yet it was also mentioned that there are no clear guidelines on how to involve users. An expert (ID12) who worked on the adaptation of the DHI for refugees mentioned:

> *The field is new ... there is no blueprint, so you have to make it participatively and try to see what really works within that culture in the sense of what is their understanding.* [ID12]

Responses highlighted several benefits of user involvement, including (1) a deeper understanding of technology use within a specific culture, (2) capturing lived experiences and comprehending the extent of necessary adaptation, (3) identifying and recognizing diversity within cultural groups (eg, subcultures) and adjusting accordingly, (4) increased chances that the adapted DHI aligns culturally with the target group, (5) a higher likelihood of ongoing use, and (6) fewer errors (eg, adapting something that requires no change). Avoiding mistakes was highlighted as particularly relevant to DHIs by one of our experts, who adapted an eHealth training for informal caregivers, primarily because changing software or hardware is associated with lower cost-effectiveness of the adaptation process. They mentioned:

> *Well, advantages, many in terms of that you probably will get a product that is going to be hopefully used [...] you will reduce or limit the amount of mistakes that you will have if you don't do it. And that is important for cost-effectiveness.* [ID10]

Another expert highlighted that recruiting must be culturally tailored and supported by relevant community groups and organizations to reach and engage users in the adaptation process.

Responses revealed three methods for user involvement, primarily through interviews and focus groups. The first approach was expert-centered, in which the adaptation team and consulted experts made all initial decisions on which DHI elements must be adapted and how (top-down approach). The second approach included simultaneous expert and user feedback, mainly in the form of interviews and focus groups (hybrid top-down and bottom-up approach). The third approach was user-centered, commonly agreed to be best practice, in which users were continuously involved in the adaptation process (through surveys, interviews, and focus groups), and expert opinions were mainly confirmatory (bottom-up approach). An expert who adapted a text-based app for Indigenous communities in New Zealand said:

> *Always start with the people. So, we generally always start with some formative research, focus groups, or other kind of qualitative research with that group you know, kind of representing the target population.* [ID6]

Overall, this approach was often described as the most resource-intensive.

One expert highlighted that even if the adaptation of the technology is behind (eg, the software has not been adapted yet), it is essential to ensure users are provided with regular input (mock-up versions, examples) to ensure that user feedback is as helpful as possible. Yet, experts highlighted that due to a lack of time and resources, in practice, user involvement often plays a more minor role than it should in cultural adaptations of DHIs. One expert in the cultural adaptation of DHIs in middle and low-income settings mentioned:

> *I didn't have a lot of money. So, it would be nice to include more participants [...]. So, like If I had more time, I definitely would have done a qualitative study to get feedback on the adapted version before conducting the pilot's study.* [ID3]

Overall, involving DHI users, the adaptation process was described as iterative.

Expert and Stakeholder Involvement

Our participants also addressed the integration of experts and other stakeholders (eg, health care professionals and software designers) into the adaptation process, in addition to user involvement. This was distinguished from having experts or stakeholders within the adaptation team, and it mostly referred to consulting additional ones throughout the adaptation process. An expert said:

> *So all the right stakeholders, the funders, the referrers, the service providers in that kind of area that you're in [...] just try and think of them all up front and then include them in your project, if you can.* [ID6]

Overall, responses highlighted that experts and stakeholders can provide critical knowledge (eg, technical), facilitate user involvement in the adaptation process (eg, doctors referring patients), and contribute to an adaptation's long-term success. One expert described the role of experts and stakeholders as supportive, by reducing the overall uncertainty around cultural adaptations.

Choosing the Adaptation Elements

When asked which elements of cultural DHI require adaptations, the most common responses were: (1) language, (2) lived experience, and (3) technology.

Language

First, experts explained that the language of a DHI, including its tone, narrative, and complexity, must be clear and relatable to the focal population. One expert in cultural adaptations for young carers described:

> *[...] just the language and not necessarily meaning translation. Just the whole tone of the language, the simplicity, clarity of the language.* [ID4]

Lived Experience

Second, they viewed that adapted DHI content must reflect the lived experience of the focal cultural group, which includes

XSL•FO
RenderX

their reality, past and current living conditions, values, beliefs, and preferences. An expert in caregiver DHIs described:

> *We then definitely wanted to collect a number of lived experiences that we could then use to create test case scenarios.* [ID4]

Technology

Third, experts noted that culturally sensitive changes in technology and its functions must take into account the focal group's digital literacy and skills, particularly if the group's cultural influences play a role. An expert on the adaptation of mobile mental health interventions for refugees said:

> *We made sure that it looks and works similar to what people are already using [...]. So people are already familiar with the interfaces and how it works to support those with low tech literacy.* [ID3]

That ensures that the adapted DHI is easy to adhere to. Experts also mentioned that it is essential to carefully consider the sociodemographic characteristics of the focal population, such as age and gender, as these often interact with one's cultural background, influencing how health technologies are perceived and used.

When asked how to choose what to adapt, responses varied and did not suggest any specific models or frameworks to help prioritize which DHI elements to adapt, except for those who conducted implementations of generic DHIs, where the developers of the initial intervention provided some adaptation guidance.

Understanding the Challenges

Responses identified five common DHI-relevant challenges, including (1) technology, (2) uncertainty, (3) user involvement, (4) communication, and (5) evaluation and sustainability.

Technology

DHIs were described as more complex to adapt than non-DHIs because changing hardware and software often requires more resources (expertise, time, and money), including higher up-front investments than usually expected or available. An expert described:

> *So a couple of months ago, we needed to update one of our digital tools. You swear we were trying to put somebody on the moon. We're changing something that's very basic. But it's going to be $6000. It is taking so much longer. It was just changing a few words.* [ID7]

Experts also described that if resources are limited, the cultural adaptation risks being insufficient or superficial, which was the case for many of the examples mentioned during the interviews. One expert said:

> *We were very limited in budget, so ideally, the adaptation would have been a lot more radical, and it would have had animations.* [ID4]

Also, two experts emphasized that current health technologies are based on, and therefore biased toward, Western cultures. Adapting these technologies without any influence of these biases and preconceptions was described as a complex yet crucial task.

Uncertainty

Uncertainty was implied in most interviews and primarily referred to missing knowledge of the required cultural adaptation. Experts often described not knowing the right balance between adapting a DHI, keeping costs contained, and ensuring that the adapted DHI is effective. An expert in text-based DHIs for Indigenous communities mentioned:

> *You can never fully account for everything that you need to in terms of cultural adaptation. So there are always things that you don't know.* [ID6]

Adapting too much was described as too costly, leading to DHIs that are too specific, and thus, exclude some people who might otherwise have benefited. Adapting too little was described as being insufficiently targeted, which may result in ineffective adaptations and wasted effort and resources. Many aspects outlined in the previous paragraphs (eg, ensuring that the adaptation team is multiprofessional and culturally competent, involving users, experts, and other stakeholders, and evaluating adapted DHIs) were mentioned as potentially reducing some of that uncertainty. Ultimately, one expert described that uncertainty is inherent and cannot be avoided entirely.

User Involvement

Overall, responses highlighted that user involvement in the adaptation process can be challenging, mainly when conducted digitally and remotely (often in DHIs). Experts mentioned that it is difficult to (1) reach people and get them engaged, (2) bring experts and users to collaborate, and (3) secure enough time and money for thorough user involvement. An expert in the adaptation of maternity mobile health apps in Brazil said:

> *I tried it, but there are categories of power, and I was less powerful. So, to get these people (experts) together and convince them that this was important was already very difficult. Imagine having the users involved.* [ID1]

As explained by the experts, these challenges, in addition to limited resources (money and time), often lead to cultural adaptations that have minimal to moderate (eg, small sample sizes and few feedback loops) user involvement, often after the adaptation team and experts have decided on what and how much needs to be adapted.

Communication

Experts emphasized that cultural adaptations, especially of DHIs, are a multidisciplinary effort and rely on feedback from various stakeholders across disciplines (eg, technology, medicine, and public health). According to their experience, communication was viewed as a common challenge. This primarily involved communication among IT experts, the rest of the adaptation team (eg, health researchers), and other stakeholders, including users. For example, one expert in informal caregiver DHIs mentioned that translating user and expert feedback in understandable ways for hardware and software designers was described as particularly challenging, saying:

XSL•FO
RenderX

> *So, I was in between the target population and the engineering developers of the platform. And it's very difficult to have a communication with both and try to put them together.* [ID10]

Maintaining transparency and including team members with interprofessional skills (eg, health and engineering knowledge) were mentioned as potential solutions to communication problems. Yet, some responses indicated that the lack of resources and time does not always allow for that.

Evaluation and Sustainability

Experts mentioned that robust evaluations (eg, randomized controlled trials) of how well culturally adapted DHIs perform are rare. Two common themes that arose were (1) the scarcity of additional resources to conduct evaluations and (2) the difficulty of engaging often hard-to-reach individuals in evaluations (eg, in longitudinal studies). Finally, culturally adapted DHIs were described as dynamic and may need to be regularly updated to remain relevant. Our participants described that as a challenge because (1) technology or software changes can be complex and costly, (2) lacking postadaptation funding, and (3) lacking evaluations that prove the initial cultural adaptation was effective. Overall, responses suggested evidence of the effectiveness of cultural adaptations of DHIs is currently lacking, which explains the lack of systematic guidance (eg, DHI-specific models or frameworks). An expert (ID6) in text-based DHIs described:

> *[...] they generally don't even put enough money into maintaining it because they just think once you've done a digital program, it's there. And so then they tend to put funding in for things like upgrades of digital systems.* [ID6]

Recommendations

Based on all our findings, we have compiled a list of recommendations for future cultural adaptations of DHIs. All stem from the experiences, mistakes, and lessons learned by the interviewed experts. We summarized all recommendations in Textbox 1.

Textbox 1. Recommendations for future cultural adaptations of digital health interventions (DHIs).

- Define the focal culture or cultural group clearly from the start
- Learn from the experience of those who have previously culturally adapted DHIs
- Involve users continuously, adapt only what is relevant to them, and consider their digital literacy
- Think about the DHI design and functionality from the start, and prioritize what users want
- Do not underestimate the money and time needed to adapt technology and expect setbacks
- Stay open-minded and free of assumptions (ie, do not assume what the focal population wants)

The first recommendation highlights the importance of a clear, appropriate definition, chosen well before an adaptation starts. An expert on mobile mental health intervention mentioned:

> *You really need to define culture as something trans-cultural, something diverse, and you have different categories that you have to keep in mind.* [ID11]

The second recommendation addresses current evidence gaps and urges future adaptations to learn from existing knowledge, either through reviews or by talking to people who have adapted DHIs. One expert in the cultural adaptation of DHIs in middle and low-income settings said:

> *There are some projects doing cultural adaptations of the digital interventions and like in a in a very systematic way. So, I think it's good to talk to them to get their experiences.* [ID3]

The third recommendation emphasizes the need to involve users throughout the adaptation and ensure that all changes to a DHI reflect their needs and preferences. While involving users in the adaptation process, it is important to understand and address their digital support needs. One expert highlighted: "But I mean, even involve them at the very, very earlier stage of the research. I think that would be the best thing." Related to that, thinking early about technological changes (eg, changing certain functionalities for better cultural fit) is the fourth recommendation. One mental health expert (ID8) said:

> *So thinking about the web build and the functionality from the outset rather than as a second thought. Because, yeah, it's kind of the cultural adaptation, the content, and like I said that it could work.* [ID8]

The fifth recommendation urges adaptors to consider the time and money that adaptations require before adapting. One expert (ID13) who has adapted a mobile mental health app for refugees mentioned:

> *And then, another thing is, it is really important not to underestimate the money and time it takes to create software.* [ID13]

The final recommendation refers to adapting without assuming what the focal population wants, but listening carefully and only acting on that. An expert who has adapted an eHealth training for minority young carers described:

> *I just think it's getting rid of your preconceptions and your assumptions about the cultural group that you are targeting.* [ID4]

Discussion

Principal Results

Cultural adaptations are emerging and have the potential to improve DHI access for medically underserved population groups [25]. This work represents one of the first attempts to learn from the experiences of academic experts who have

XSL•FO
RenderX

previously culturally adapted DHIs. Our findings outline current practice and underline existing gaps. Overall, the current practice emphasizes culturally competent and multiprofessional adaptation teams that involve IT experts (eg, hardware and software designers), and user, expert, and other stakeholder involvement in the adaptation process. There is also an emphasis on three core adaptation elements, highlighting that a culturally adapted DHI must be understandable (eg, using appropriate language, tone, and narrative), relatable (eg, reflecting lived experience and sociodemographics), and appealing and easy to adhere to (eg, functions, design, and layout that consider digital literacy and preferences). However, many challenges remain, and although the field is emerging, responses have not revealed any systematic ways for the cultural adaptations of DHIs.

Comparison With Prior Work

Our findings are partially aligned with general adaptation frameworks, such as FRAME (framework for reporting adaptations and modifications to evidence-based interventions) [26]. Our findings, and FRAME, highlight the importance of justifying adaptations, clearly outlining adaptation elements, involving a broad range of stakeholders (from experts to community members), and conducting evaluations [26]. Our findings also align with the 17 cultural adaptation considerations provided by the systematic review of Spanhel et al [27] for internet- and mobile-based mental health interventions [26]. In particular, they confirm the importance of adapting in ways that reflect the lived experience of the target population. That includes, as also identified by Spanhel et al [27], habits, gender norms, religious beliefs, values and traditions, stressors, and socioeconomic factors (eg, education and political context) [27]. Spanhel et al [27] also underlined the importance of language tailoring and visualization, ensuring that text is simple, culture-specific, and accompanied by visual cues [27]. This aligns well with our findings, highlighting that adaptation goes well beyond simple translations. Experts described that language, including its tone, needs to be understandable, providing a relatable and culturally familiar narrative. Our work contributes beyond that and provides insights into implementers' challenges when adapting digital interventions. Securing enough resources to conduct thorough hardware and software adjustments, reaching and engaging users of the focal cultural group (digitally or remotely) throughout the adaptation process, and ensuring good communication between ICT and non-ICT members of the adaptation team were some of the technology-specific challenges that experts described.

Uncertainty was emphasized as one of the most prominent and cross-cutting adaptation challenges. That means adaptations often start with little knowledge of how much to adapt to ensure DHI effectiveness without wasting time and resources. Uncertainty is not exclusive to DHIs but to cultural adaptations in general [28]. Several well-described challenges might lead to uncertainty. Castro et al [16] describe some of these and include (1) defining culture, (2) segmenting and understanding the population for which the adaptation is conducted, and (3) understanding heterogeneity and subcultures within that population. Although experts provided recommendations to mitigate uncertainty when adapting DHIs (eg, continuously involving the user and expert feedback), the lack of evidence, guidance, frameworks, and systematic practice makes uncertainty inherent. That uncertainty was repeatedly voiced during the interviews and is reflected in our findings overall.

When asked how to choose or prioritize what to adapt, responses varied and did not provide any systematic approach. User-centeredness, although not always possible if resources are lacking or the target population is hard to reach, was highlighted as the best approach for identifying (or prioritizing) which DHI elements are culturally relevant and worth adapting. This aligns with previous work highlighting the value of user-centered adaptations, particularly when adapting DHIs [29]. The acceptance and use of health technologies depend on many factors, such as self-efficacy, prevailing norms and values, and trust [30]. These factors are often related to culture, and therefore, highly relevant to the adaptation process. Without direct feedback from prospective users and members of the focal cultural group, these factors cannot be fully captured. Based on the experience of these experts, engaging users should start early in the adaptation process, be iterative, allow for flexibility, be continuous yet stepwise, and allow for feedback loops. However, there is still little systematic guidance on how to involve DHI users in the cultural adaptation process, especially if conducted digitally or remotely.

Implications, Current Practice, and Existing Gaps

Our findings highlight the importance of cultural adaptations for reaching and engaging systematically underserved communities. Despite overall agreement on their value, current practice often lags, and the needed support to advance the field has not been prioritized yet. First, there seems to be a lack of dedication to resources to complete well-designed and thorough adaptations, which leads to adapted DHIs that are only partially adapted and not adequately evaluated. This is aligned with the findings of a recent literature review by Naderbagi et al [31], which highlights the lack of rigorous DHI adaptations, arguing that budgets and timelines should facilitate instead of limiting the adaptation process.

Second, the lack of thorough evaluations likely has a negative impact on the field as it limits available evidence on which future DHI adaptations could be built. Existing frameworks, such as the Ecological Validity Framework, the Cultural Accommodation Model, and the circular model of cultural tailoring, offer valuable higher-level guidance on cultural adaptations, yet do not sufficiently cover aspects specific to DHIs, such as technological and digital literacy barriers [32,33]. Recent work, such as the systematic review on the cultural adaptation of internet- and mobile-based interventions for mental disorders by Spanhel et al [27] and the related editorial by Marwaha and Kvedar [25], provides valuable steps toward filling that gap. However, a comprehensive framework focusing on the cultural adaptation of DHIs is still lacking. That feeds an uncertainty on what to prioritize to maintain a balance between available resources and addressing the population's needs, the adaptations are aimed at. Our findings contribute to closing the practice gaps by (1) highlighting actual practice gaps and DHI-specific challenges, (2) providing examples of commonly agreed-upon good practices, (3) identifying core adaptation

XSL•FO
RenderX

elements, and (4) highlighting the need for DHI-specific cultural adaptation frameworks.

Toward DHI-Specific Cultural Adaptation Frameworks

Our methodological approach does not allow for the development of a DHI-specific cultural adaptation framework because it is a single qualitative study with a limited range of participant perspectives. Developing a robust and transferable framework typically requires insights from multiple studies and across a diverse range of participants, as well as data triangulation [34]. However, our findings contribute toward a future DHI-specific cultural adaptation framework. First, our findings highlight the importance of defining culture and justifying the adaptation early on. A future framework should provide stepwise guidance on how to identify relevant cultural elements in ways that align with a specific DHI and the targeted cultural group, as well as tools to identify cultural nuances beyond ethnicity and place of origin. It should also, as our findings highlight, provide criteria for assessing the need to adapt.

Second, our findings indicate the importance of diverse adaptation teams, as well as user and stakeholder engagement. This is aligned with prior work for nondigital interventions that emphasize the role of community and stakeholder involvement in sustainable and efficient adaptations [35]. A DHI-specific framework should (1) guide the creation of adaptation teams with adequate digital expertise, and (2) provide digital approaches to reach and engage users and other stakeholders. This would complement existing work, which highlights the importance of involving the target group in the adaptation process but offers little guidance on digital ways to do so [27].

Third, our findings present essential elements of adaptation but emphasize that these decisions are context-specific. A future framework should offer support for decision-making related to prioritizing which elements need adaptation, including both technical and user-centered aspects. Finally, we identified the core DHI-specific challenges. Based on these, a framework should guide overcoming technological limitations and offer strategies for effective communication, resource allocation, and evaluations.

Limitations

This study has some limitations. First, because this is an emerging field, we only interviewed academic experts since they were easy to identify (through publications). Thus, our findings might not be fully transferable to cultural adaptations of DHIs in nonacademic settings. The wealth of knowledge outside of academia (eg, private sector) would add to the understanding of cultural adaptations and should be prioritized in future research. Yet, we kept our questions general and not specific to academic settings, and most of the findings are likely applicable to cultural adaptations of technology in general. Second, we used snowball sampling as a complementary recruitment method. While snowball sampling is a widely used and practical recruitment method in qualitative research, it can lead to selection bias (eg, too homogeneous sample), primarily because participants often share common professional networks, which may hinder thematic saturation. Third, culture is a multifaceted concept. This study approached culture in the context of health and technology, which is only one of many approaches to explain and understand it. Fourth, our interviewees were predominantly health and digital health experts. Including experts with a pure technology or ICT background would have likely added a valuable perspective that might be currently missing.

Conclusions

Our work is one of the first to explore the cultural adaptation of DHIs through an academic lens. The adaptation process was described as an iterative, often unstructured, and resource-intensive process that needs careful justification and a good understanding of culture and the cultural group for which it is conducted. Based on their experiences, experts recommended that (1) the adaptation team is multiprofessional, digitally competent, and culturally sensitive (2) DHI users and (3) all other relevant stakeholders are continuously involved, and (4) that the adaptations integrate evaluations and knowledge exchange. They further emphasized that culturally adapted DHIs must be understandable, relatable, appealing, and easy to adhere to, ensuring that health technology and content reflect the target population's lived experience, sociodemographic characteristics, and digital literacy. Yet, beyond those broader recommendations, our interviews confirmed the lack of systematic, detailed approaches to culturally adapting DHIs, including how to assess whether an adaptation is needed, how to structure the adaptation process, how to choose and prioritize adaptation elements, as well as how much to adapt to ensure that adaptations are cost-effective. That uncertainty and the added complexity that technology adds to cultural adaptations were described as critical challenges. To identify general recommendations for guiding cultural adaptations of DHIs, future research should focus on thoroughly evaluating adapted DHIs and exploring systematic ways to mitigate core challenges.

Acknowledgments

VN's work was funded by the Swiss National Science Foundation (grant 202914). SJC's contribution to the manuscript was supported by NIAAA T32 AA07459. All funding sources had no role in the study design; collection, analysis, or interpretation of the data; or writing the manuscript. We thank Prof Dr Taylor Wray for his continuous support and feedback.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

VN contributed to the study's conceptualization, methodology, data curation, formal analysis, and writing of the original draft. PD contributed to the study's formal analysis, writing, review, and editing. SJC contributed to the study's formal analysis, writing, review, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guide.
[DOCX File, 14 KB - mhealth_v13i1e59965_app1.docx]

References

1. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
2. Cohen AB, Mathews SC, Dorsey ER, Bates DW, Safavi K. Direct-to-consumer digital health. Lancet Digital Health 2020 Apr;2(4):e163-e165. [doi: 10.1016/S2589-7500(20)30057-1] [Medline: 33328077]
3. Leblalta B, Kebaili H, Sim R, Lee SWH. Digital health interventions for gestational diabetes mellitus: a systematic review and meta-analysis of randomised controlled trials. PLOS Digital Health 2022 Feb;1(2):e0000015. [doi: 10.1371/journal.pdig.0000015] [Medline: 36812531]
4. Jiang X, Ming WK, You JH. The cost-effectiveness of digital health interventions on the management of cardiovascular diseases: systematic review. J Med Internet Res 2019 Jun 17;21(6):e13166. [doi: 10.2196/13166] [Medline: 31210136]
5. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digital Health 2021;3(764079):764079. [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
6. Eruchalu CN, Pichardo MS, Bharadwaj M, et al. The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York City. J Urban Health 2021 Apr;98(2):183-186. [doi: 10.1007/s11524-020-00508-9] [Medline: 33471281]
7. Karri K, Yarra P. Inequities still exist in the use of digital health technology across different sociodemographic subgroups. Evid Based Nurs 2022 Jan;25(1):23-23. [doi: 10.1136/ebnurs-2020-103355]
8. Brewer LC, Fortuna KL, Jones C, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512. [doi: 10.2196/14512] [Medline: 31934874]
9. Krukowski RA, Ross KM, Western MJ, et al. Digital health interventions for all? Examining inclusivity across all stages of the digital health intervention research process. Trials 2024 Jan 30;25(1):98. [doi: 10.1186/s13063-024-07937-w] [Medline: 38291539]
10. Huh J, Koola J, Contreras A, et al. Consumer health informatics adoption among underserved populations: thinking beyond the digital divide. Yearb Med Inform 2018 Aug;27(1):146-155. [doi: 10.1055/s-0038-1641217] [Medline: 30157518]
11. Szinay D, Forbes CC, Busse H, DeSmet A, Smit ES, König LM. Is the uptake, engagement, and effectiveness of exclusively mobile interventions for the promotion of weight-related behaviors equal for all? A systematic review. Obes Rev 2023 Mar;24(3):e13542. [doi: 10.1111/obr.13542] [Medline: 36625062]
12. Western MJ, Armstrong MEG, Islam I, Morgan K, Jones UF, Kelson MJ. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021 Nov 9;18(1):148. [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
13. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 9;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
14. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 4;18(3):e50. [doi: 10.2196/jmir.5105] [Medline: 26944212]
15. Perski O, Short CE. Acceptability of digital health interventions: embracing the complexity. Transl Behav Med 2021 Jul 29;11(7):1473-1480. [doi: 10.1093/tbm/ibab048] [Medline: 33963864]
16. Castro FG, Barrera M, Holleran Steiker LK. Issues and challenges in the design of culturally adapted evidence-based interventions. Annu Rev Clin Psychol 2010;6(213-39):213-239. [doi: 10.1146/annurev-clinpsy-033109-132032] [Medline: 20192800]
17. Nittas V, Daniore P, Chavez SJ, Wray TB. Challenges in implementing cultural adaptations of digital health interventions. Commun Med (Lond) 2024 Jan 5;4(1):7. [doi: 10.1038/s43856-023-00426-2] [Medline: 38182750]
18. Barrera M, Castro FG, Strycker LA, Toobert DJ. Cultural adaptations of behavioral health interventions: a progress report. J Consult Clin Psychol 2013 Apr;81(2):196-205. [doi: 10.1037/a0027085] [Medline: 22289132]

XSL•FO
RenderX

19. Fendt-Newlin M, Jagannathan A, Webber M. Cultural adaptation framework of social interventions in mental health: evidence-based case studies from low- and middle-income countries. Int J Soc Psychiatry 2020 Feb;66(1):41-48. [doi: 10.1177/0020764019879943] [Medline: 31580173]
20. Spanhel K, Schweizer JS, Wirsching D, et al. Cultural adaptation of internet interventions for refugees: results from a user experience study in Germany. Internet Interventions 2019 Dec;18(100252):100252. [doi: 10.1016/j.invent.2019.100252] [Medline: 31890608]
21. Garabiles MR, Harper Shehadeh M, Hall BJ. Cultural adaptation of a scalable world health organization E-mental health program for overseas Filipino workers. JMIR Form Res 2019 Mar 29;3(1):e11600. [doi: 10.2196/11600] [Medline: 30924784]
22. Pope C, Mays N. Qualitative Research in Health Care: John Wiley & Sons Ltd; 2019. [doi: 10.1002/9781119410867]
23. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 1;19. [doi: 10.1177/1609406919899220]
24. Does my project need HRPP/IRB review? Brown University. URL: https://division-research.brown.edu/research-cycle/conduct-research/human-subjects-research/need-hrppirb-review [accessed 2025-03-20]
25. Marwaha JS, Kvedar JC. Cultural adaptation: a framework for addressing an often-overlooked dimension of digital health accessibility. NPJ Digital Med 2021 Oct 1;4(1):143. [doi: 10.1038/s41746-021-00516-2] [Medline: 34599270]
26. Stirman SW, Baumann AA, Miller CJ. The FRAME: an expanded framework for reporting adaptations and modifications to evidence-based interventions. Implement Sci 2019 Jun 6;14(1):58. [doi: 10.1186/s13012-019-0898-y] [Medline: 31171014]
27. Spanhel K, Balci S, Feldhahn F, Bengel J, Baumeister H, Sander LB. Cultural adaptation of internet- and mobile-based interventions for mental disorders: a systematic review. NPJ Digital Med 2021 Aug 25;4(1):128. [doi: 10.1038/s41746-021-00498-1] [Medline: 34433875]
28. Rodríguez MMD, Baumann AA, Schwartz AL. Cultural adaptation of an evidence based intervention: from theory to practice in a Latino/a community context. Am J Community Psychol 2011 Mar;47(1-2):170-186. [doi: 10.1007/s10464-010-9371-4] [Medline: 21116707]
29. Burchert S, Alkneme MS, Bird M, et al. User-centered app adaptation of a low-intensity e-mental health intervention for syrian refugees. Front Psychiatry 2018;9(663):663. [doi: 10.3389/fpsyt.2018.00663] [Medline: 30740065]
30. Tao D, Wang T, Wang T, Zhang T, Zhang X, Qu X. A systematic review and meta-analysis of user acceptance of consumer-oriented health information technologies. Comput Hum Behav 2020 Mar;104:106147. [doi: 10.1016/j.chb.2019.09.023]
31. Naderbagi A, Loblay V, Zahed IUM, et al. Cultural and contextual adaptation of digital health interventions: narrative review. J Med Internet Res 2024 Jul 9;26:e55130. [doi: 10.2196/55130] [Medline: 38980719]
32. Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: issues for the cultural adaptation and development of psychosocial treatments with Hispanics. J Abnorm Child Psychol 1995 Feb;23(1):67-82. [doi: 10.1007/BF01447045] [Medline: 7759675]
33. Leong FT, Lee SH. A cultural accommodation model for cross-cultural psychotherapy: Illustrated with the case of Asian Americans. Psychotherapy (Chic) 2006;43(4):410-423. [doi: 10.1037/0033-3204.43.4.410] [Medline: 22122133]
34. Natow RS. The use of triangulation in qualitative studies employing elite interviews. Qual Res 2020 Apr;20(2):160-173. [doi: 10.1177/1468794119830077]
35. Perera C, Salamanca-Sanabria A, Caballero-Bernal J, et al. No implementation without cultural adaptation: a process for culturally adapting low-intensity psychological interventions in humanitarian settings. Confl Health 2020;14(46):46. [doi: 10.1186/s13031-020-00290-0] [Medline: 32684948]

Abbreviations

DHI: digital health intervention
FRAME: Framework for Reporting Adaptations and Modifications to Evidence-based Interventions
ICT: information and communication technology
WHO: World Health Organization

Edited by L Buis; submitted 27.04.24; peer-reviewed by E Rosenfeld, S Choemprayong; revised version received 23.04.25; accepted 20.05.25; published 18.07.25.

Please cite as:
Nittas V, Chavez SJ, Daniore P
Current Practice and Expert Perspectives on Cultural Adaptations of Digital Health Interventions: Qualitative Study
JMIR Mhealth Uhealth 2025;13:e59965
URL: https://mhealth.jmir.org/2025/1/e59965
doi:10.2196/59965

XSL•FO
RenderX

© Vasileios Nittas, Sarah J Chavez, Paola Daniore. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patterns of Engagement With the mHealth Component of a Sexual and Reproductive Health Risk Reduction Intervention for Young People With Depression: Latent Trajectory Analysis

Lydia A Shrier[1,2], MD, MPH; Carly E Milliren[3], MPH; Brittany Ciriello[1], MA; Madison M O'Connell[1], MPH; Sion Kim Harris[1,2], MSN, PhD, PMHNP-BC, CPH

[1]Division of Adolescent/Young Adult Medicine, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA, United States

[2]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

[3]Biostatistics and Research Design Center, Boston Children's Hospital, Boston, MA, United States

Corresponding Author:
Lydia A Shrier, MD, MPH
Division of Adolescent/Young Adult Medicine, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA, United States

Abstract

Background: Mobile health (mHealth) interventions are increasingly used to reduce risk and promote health in real-time, real-life contexts. Engagement is critical for effectiveness of mHealth interventions but may be challenging for young people experiencing depressive symptoms.

Objective: We examined engagement with the 4-week mHealth component of a counseling-plus-mHealth intervention to reduce sexual and reproductive health (SRH) risk among young people with depression (Momentary Affect Regulation – Safer Sex Intervention [MARSSI]) to determine (1) mHealth engagement patterns over time and (2) how sociodemographic characteristics, SRH risks, and depressive symptom severity were associated with these engagement patterns.

Methods: We undertook secondary analysis of data collected from June 2021 to September 2023 in a randomized controlled trial of MARSSI versus a breast health podcast. Eligibility included age 16 - 21 years, ability to become pregnant, smartphone ownership, English fluency, past-3-month penile-vaginal sex ≥1x/week and ≥1 SRH risk, and Patient Health Questionnaire-8 item score ≥8. Intervention participants received one-on-one telehealth counseling and then used an app for 4 weeks, responding to surveys (3 prompted at quasi-random, 1 scheduled daily) about affect, effective contraception and condom use self-efficacy, sexual and pregnancy desire, and recent sex, and receiving tailored messages reinforcing the counseling. We computed mHealth engagement days (responding to ≥1 app survey) by week and overall. Latent trajectory analysis identified engagement patterns over the 4 mHealth weeks among participants with any engagement. Using regression analysis, we examined the associations of sociodemographic characteristics, SRH risks, and depressive symptom severity with mHealth engagement patterns and examined moderation by depressive symptom severity. Of the 201 intervention participants, 194 (96.5%) enrolled in the app.

Results: Among those responding to app surveys (167/194, 86.1%), the median engagement was 14 (IQR 4-23) days; 32.9% (55/167) responded on ≥20 days. Overall app engagement (median) declined from 5 (IQR 3-7) days in week 1 to 1 (IQR 0 - 5) day in week 4. On latent trajectory analysis, 4 patterns of app engagement emerged: high-throughout (48/167, 28.7%), high-then-declining (40/167, 23.9%), mid-then-declining (47/167, 28.1%), and low-throughout (33/167, 19.7%). Participants identifying gender other than female and those perceiving higher socioeconomic status were more likely to have high-throughout or high-then-declining engagement. Asian or Black non-Hispanic participants and those using low-effectiveness contraception were more likely to have no engagement. In the multivariable model, Asian (adjusted odds ratio [AOR] 0.28, 95% CI 0.10-0.81), Black non-Hispanic (AOR 0.28, 95% CI 0.12-0.66), and higher perceived socioeconomic status (AOR 1.24, 95% CI 1.05-1.48) remained significantly associated with engagement. Engagement patterns showed no differences by depressive symptom severity and no significant moderation.

Conclusions: Young people with depressive symptoms showed initial high engagement with the intervention's mHealth app to reduce adverse SRH outcomes. Methods to increase and sustain mHealth engagement and differences in engagement by sociodemographic characteristics warrant further studies to optimize the reach of mHealth interventions.

Trial Registration: ClinicalTrials.gov NCT04798248; https://tinyurl.com/zxybv4vn

(JMIR Mhealth Uhealth 2025;13:e70219) doi:10.2196/70219

KEYWORDS

mHealth; ecological momentary assessment; engagement; depressive symptoms; intervention

XSL•FO
RenderX

Introduction

Adolescents and young adults have high rates of unintended pregnancy and sexually transmitted infections (STIs) compared to other age groups [1,2]. Adolescents and young adults who have depressive symptoms are at particularly high risk of these adverse sexual and reproductive health (SRH) outcomes. Although depressed young women may have difficulty forming and sustaining intimate relationships and experience less interest in sexual intercourse, they may have sexual intercourse to manage their depressive symptoms, often under the influence of substances [3] and without condoms or hormonal contraception [4]. These functional encounters may lead to lower self-esteem and bad mood [4]. As a result, depressive symptoms and SRH risk may be self-reinforcing and difficult to change.

The observed associations are consistent with a functional model of sexual behavior, in which motives for sexual behavior can be understood according to motivational (approach and avoidance) and social (intrapersonal and interpersonal) dimensions [5]. Efforts to regulate affect in depression through sexual behavior may aim to enhance positive affect, reduce or cope with negative affect, or both. Individuals with depressive symptoms may also seek validation and affirmation through social interactions. Thus, sexual behavior in depression must be understood in the contexts of both individual affect regulation motives and interpersonal relationships and that these contexts and related experiences influence future motives [6].

The bidirectional, longitudinal relationship of depressive symptoms with sexual behavior and SRH risk is dynamic and can unfold on a moment-to-moment basis. For example, among adolescents and young adults who are depressed, momentary positive affect is associated with having sex [7] and momentary negative affect is associated with having sex with a non–main partner [8]. Furthermore, for adolescents and young adults with depressive symptoms, sexual intercourse events may occur on impulse, without planning [4], and greater impulsiveness is associated with incorrect condom use, if a condom is used [9]. Interrupting the cycle of poor mental health and SRH risk behavior as it develops moment-by-moment in the natural environment may reduce unplanned pregnancy and STIs for these at-risk youth.

Mobile health (mHealth) interventions are well-suited to address the interplay between poor mental health and SRH risk behaviors. Although the definition is subject to some debate [10], in general, mHealth interventions use mobile devices, including phones, tablets, wearable sensors, communication technologies, and other wireless devices, to improve health. As described in several reviews and meta-analyses, mHealth interventions have been applied to the treatment of psychiatric disorders (eg, as a way to strengthen the effects of counseling [11]), contraceptive use [12,13], STI prevention [14], and other SRH outcomes [15]. One form of an mHealth intervention, ecological momentary intervention (EMI) [16], delivers treatment in real-time, real-life contexts, which can be especially helpful for addressing momentary dynamics such as those observed between affective states, substance use, impulsiveness, and SRH risk [4,8,9,17].

To receive mHealth interventions, participants must engage with the mobile device platform, which requires carrying their device, having it powered and on, and with active interventions in which participants provide survey data, responding to queries or initiating reporting. Thus, if participants do not engage with and therefore receive an intervention, the intervention's effect cannot be assessed [18]. Furthermore, there is a presumption of a dose-response or threshold relationship, such that higher engagement is associated with improved outcomes [19] or that sufficient engagement is required to achieve the intended outcomes (ie, effective engagement) [20].

For individuals with depression, intervention engagement may be difficult to sustain owing to symptoms such as low motivation, poor organization, low energy, behavioral avoidance, negative cognition, difficulty concentrating, and impaired information processing [21-23]. Individuals with psychiatric disorders (eg, schizophrenia) [24] have demonstrated engagement with mHealth interventions focused on mental health [25-27]. However, engagement rates are highly variable and may decline over time for mHealth interventions delivered for weeks to months [24,25,28]. Additionally, individual demographic characteristics and illness severity may be associated with varying levels of engagement with an mHealth intervention [24,27]. Limited information is known about depressed persons' engagement with mHealth interventions for SRH risk [28]. To fully evaluate an mHealth intervention for adolescents and young adults with the dual challenges of depressive symptoms and SRH risk [28], we must understand engagement with the intervention and the differences in engagement by the level of depressive symptoms and other characteristics of participants from this unique population.

We developed the Momentary Affect Regulation – Safer Sex Intervention (MARSSI) to reduce the risk for unplanned pregnancy and STIs in at-risk adolescents and young adults with depressive symptoms. MARSSI is comprised of a one-on-one main counseling session, 4 weeks of EMI, and a one-on-one booster counseling session at the end of EMI. The counseling uses motivational interviewing to support behavior change and teaches skills from cognitive behavioral therapy for depression. EMI extends the effects of counseling by assessing in-the-moment SRH risk, and if risk is reported, providing personalized, tailored messages to prompt healthy behaviors and encourage cognitive behavioral skill use. We sought to determine (1) patterns of engagement with the 4-week mHealth component of MARSSI and (2) the level of depressive symptoms, SRH risks, and sociodemographic characteristics associated with the engagement patterns.

Methods

Ethical Considerations

The parent study (Catherine Henley, PhD, MPH, et al. unpublished data, October. 2024) received institutional review board approval from Sterling institutional review board (approval 8687). All participants provided informed consent.

XSL•FO
RenderX

The original institutional review board approval covered secondary analysis without additional consent.

Participants and Setting

We undertook a secondary analysis of anonymized data collected from June 2021 to September 2023 in a randomized controlled trial of MARSSI versus a breast health podcast (ClinicalTrials.gov NCT04798248). Participants were aged 16 - 21 years; biologically able to become pregnant; patients of an SRH care provider, with a visit in the past 2 years; a smartphone owner; fluent in English; not pregnant, desiring pregnancy, or within 6 months post partum; not married or engaged to be married; reporting penile-vaginal sex at least once a week, on average; endorsing at least one SRH risk in the past 3 months; and reporting elevated depressive symptoms as measured by Patient Health Questionnaire (PHQ)-8 score ≥8 [29]. SRH risks in the past 3 months included more than one sexual partner, inconsistent or no condom use, no contraception or low-effectiveness method (ie, condoms, diaphragm, withdrawal, rhythm) as primary form of contraception, penile-anal sex, sex under the influence of alcohol or other drugs, and treatment of an STI. Participants were remunerated up to US $230 in electronic gift cards for completed study activities.

To recruit participants, we partnered with 9 Planned Parenthood affiliates serving 14 US states: Planned Parenthood Great Northwest, Hawai'i, Alaska, Indiana, Kentucky (includes Washington and Idaho); Planned Parenthood of Wisconsin; Planned Parenthood Southeastern Pennsylvania; Planned Parenthood Gulf Coast (Texas and Louisiana); Planned Parenthood of the St. Louis Region and Southwest Missouri; Planned Parenthood Southwest Ohio Region; Planned Parenthood of Illinois; Planned Parenthood Mar Monte (California); and Planned Parenthood of the Pacific Southwest (California). Active recruitment methods at the Planned Parenthood affiliates included study staff sending SMS text messages and patient portal messages directly to patients through the electronic medical record system and affiliate staff mentioning the study to the patients at their appointments. Passive recruitment methods included hanging posters and distributing flyers, as well as posting information on the affiliate website and study-specific or study-affiliated social media accounts on Google, Instagram, and Facebook. Recruitment materials contained a QR code and link to a web-based self-screener for eligibility. In the trial, 405 individuals were enrolled and randomly assigned 1:1 to MARSSI (n=201) versus the podcast (n=204). The randomization scheme used blocking by study coordinator or intervention counselor and by state to account for the potential differences in the availability, quality, and access to reproductive health care services across states. The study coordinator or intervention counselor staff (n=6) could enroll and, per assigned study condition, deliver the intervention counseling to eligible individuals from any US state. Data from participants assigned to the intervention were included in the analyses reported herein.

Intervention

Participants assigned to MARSSI met with a study coordinator or an intervention counselor (master's level) for a one-on-one telehealth counseling session in which they chose the SRH risk behavior that they wished to change, developed a change plan, and discussed other SRH risks in brief [28]. Subsequently, for the mHealth component, participants used a smartphone app (MetricWire) for 4 weeks to respond to surveys (3 prompted at quasi-random times, 1 scheduled each day) about their affective states, effective contraception self-efficacy, condom use self-efficacy, desire to have sex, desire to be pregnant, and recent sexual intercourse, and then receive messages based on their responses. After the initial development, the intervention materials were revised to enhance gender inclusivity by limiting the use of gendered pronouns and including the counselor and the participant sharing their preferences for name and pronouns at the outset of the main counseling session.

Measures

We defined mHealth app engagement as responding to at least 1 survey (momentary or daily) in a day. Because MARSSI emphasizes autonomy over health behavior decisions, we did not expect or require responding to every app survey prompt. Thus, we did not consider response rate as an accurate indicator of app engagement [24,28]. The eligibility survey included PHQ-8, a self-reported measure of depressive symptoms comprising 8 of the 9 Likert-type items (each scored 0 - 3) in the well-validated PHQ-9 [30]. The PHQ-8 excludes the question on thoughts of suicide and therefore is recommended in research studies in which further evaluation (ie, with a clinical interview) may not be possible. The score range for PHQ-8 is 0 - 24, and the cutoffs for diagnosis of a depressive disorder and the level of symptom severity are the same as those used for PHQ-9 [30]. Consistent with scoring interpretation guidelines [29], we categorized PHQ-8 scores as mild (8-9), moderate (10-14), and moderately severe or severe (15-24) [15-23,31]. Each SRH risk was assessed on the eligibility survey with a single yes-no item. The baseline survey included age (years), race, ethnicity (select all that apply: American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino/a, Native Hawaiian or other Pacific Islander, White, some other race or ethnicity), gender identity (female, male, transgender female, transgender male, nonbinary or genderqueer, unsure or questioning, I do not identify as any of these), mother's highest education (did not finish high school, received a high school diploma or general equivalency diploma, completed some college, finished college, finished graduate school, law school, or medical school, don't know), and self-perceived socioeconomic status (SES) relative to other people in the United States from 10 (most well off) to 1 (least well off) [32].

Statistical Analyses

All analyses were performed in SAS (version 9.4; SAS Institute) using an α level of .05. We report mean (SD) or median (IQR) for continuous variables and frequency (%) for categorical variables. We examined engagement rates by week and overall. We used group-based latent trajectory modeling to cluster participants by their level of engagement with the mHealth app over 4 weeks [33]. These models were implemented using Proc Traj (SAS Institute) among the participants who ever engaged with the app to identify and group participants into distinct patterns of engagement over the study period. The best-fitting

XSL•FO
RenderX

model with respect to the number of engagement groups and polynomial order was selected using the model with the lowest Akaike Information Criterion. Participants who never engaged with the app during the study period formed a "no engagement" category. The identified engagement trajectory patterns comprised a trajectory group variable with categories that were ordered by the initial level of engagement.

We conducted bivariate analyses to examine the associations of sociodemographic characteristics, SRH risks, and severity of depressive symptoms with the identified patterns of mHealth app engagement by using one-way analysis of variance for continuous variables and chi-square tests for categorical variables. We then used multivariable ordinal logistic regression to examine the associations in an adjusted model and to compute the adjusted odds ratios (AORs) and 95% CIs for all predictor variables (AOR>1 indicates greater odds of being in a higher engagement trajectory, AOR<1 indicates lower odds of being in a higher engagement trajectory and thus greater odds of a lower engagement trajectory). In addition, we explored the potential for the effect modification of depressive symptom severity with sociodemographic characteristics and SRH risks by testing the interaction terms in a multivariable regression model.

Results

Baseline Characteristics of the Participants

Of the 201 intervention participants, 194 (96.5%) enrolled in the app and were included in the analyses herein. Participant characteristics at baseline are shown in Table 1.

XSL•FO
RenderX

Table . Sociodemographic characteristics of the participants using the mobile health app component of the Momentary Affect Regulation – Safer Sex Intervention—overall and by pattern of app engagement.

Characteristic	Engagement pattern						*P* value
	Overall (n=194)	No engagement (n=27)	Low-throughout (n=33)	Mid-then- declining (n=46)	High-then- declining (n=40)	High-throughout (n=48)	
Age (years), mean (SD)	19.1 (1.1)	19.2 (1.1)	19.4 (1.1)	19.1 (1.1)	18.9 (1.1)	19.1 (1.1)	.45
Gender, n (%)							.04
Female	169 (87)	23 (85)	31 (94)	39 (85)	39 (98)	37 (77)	
Other	25 (13)	4 (15)	2 (6)	7 (15)	1 (2)	11 (23)	
Race and ethnicity, n (%)							.006
Asian	14 (7)	4 (15)	4 (12)	1 (2)	4 (10)	1 (2)	
Black or African American, non-Hispanic	22 (11)	8 (29)	5 (15)	3 (7)	3 (7)	3 (6)	
Hispanic or Latino/a	22 (11)	4 (15)	2 (6)	10 (22)	3 (7)	3 (6)	
White, non-Hispanic	94 (49)	10 (37)	17 (52)	19 (41)	21 (53)	27 (56)	
Another race or ethnicity[a]	42 (22)	1 (4)	5 (15)	13 (28)	9 (23)	14 (29)	
Mother's highest education (n=193), n (%)							.62
Less than college graduate	129 (66)	19 (70)	23 (70)	34 (74)	26 (65)	27 (56)	
College degree or beyond	64 (33)	8 (30)	10 (30)	12 (26)	14 (35)	20 (42)	
Perceived socioeconomic status (1-lowest to 10-highest), mean (SD)	5.0 (1.6)	4.9 (1.6)	4.6 (1.8)	4.8 (1.6)	5.1 (1.5)	5.6 (1.4)	.06
Sexual and reproductive health risk in the past 3 months, n (%)							
Sex without condom	180 (93)	25 (93)	32 (97)	40 (87)	38 (95)	45 (94)	.47
Low-effectiveness contraception as primary method[b]	113 (58)	21 (78)	13 (39)	24 (52)	29 (73)	26 (54)	.009
More than 1 partner	70 (36)	10 (37)	10 (30)	18 (39)	18 (45)	14 (29)	.55
Sex within 2 hours after using alcohol or drugs (n=193)	125 (64)	17 (63)	22 (67)	31 (67)	24 (60)	31 (65)	.68
Treated for an STI[c]	39 (20)	6 (22)	7 (21)	8 (17)	8 (20)	10 (21)	.99
Number of risks, mean (SD)	2.7 (1.1)	2.9 (1.1)	2.5 (1.1)	2.6 (1.1)	2.9 (1.0)	2.6 (1.2)	.42
Depressive symptoms, mean (SD)							
PHQ-8[d] score (0 - 24)	13.0 (4.2)	13.3 (3.6)	13.3 (4.8)	12.1 (3.9)	13.5 (4.9)	13.1 (3.9)	.54
Severity (score range), n (%)							.80

XSL•FO
RenderX

Characteristic	Engagement pattern						*P* value
	Overall (n=194)	No engagement (n=27)	Low-throughout (n=33)	Mid-then- declining (n=46)	High-then- declining (n=40)	High-throughout (n=48)	
Mild (8-9)	49 (25)	4 (15)	10 (30)	14 (30)	9 (22)	12 (25)	
Moderate (10-14)	84 (43)	15 (55)	11 (33)	20 (44)	18 (45)	20 (42)	
Moderately severe to severe (≥15)	61 (31)	8 (30)	12 (36)	12 (26)	13 (33)	16 (33)	

[a]Includes American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, other race, and multiple race or ethnicity.

[b]Used condoms, diaphragm, cervical cap, spermicide, sponge, fertility awareness, or withdrawal as the primary form of contraception.

[c]STI: sexually transmitted infection.

[d]PHQ-8: Patient Health Questionnaire-8 items.

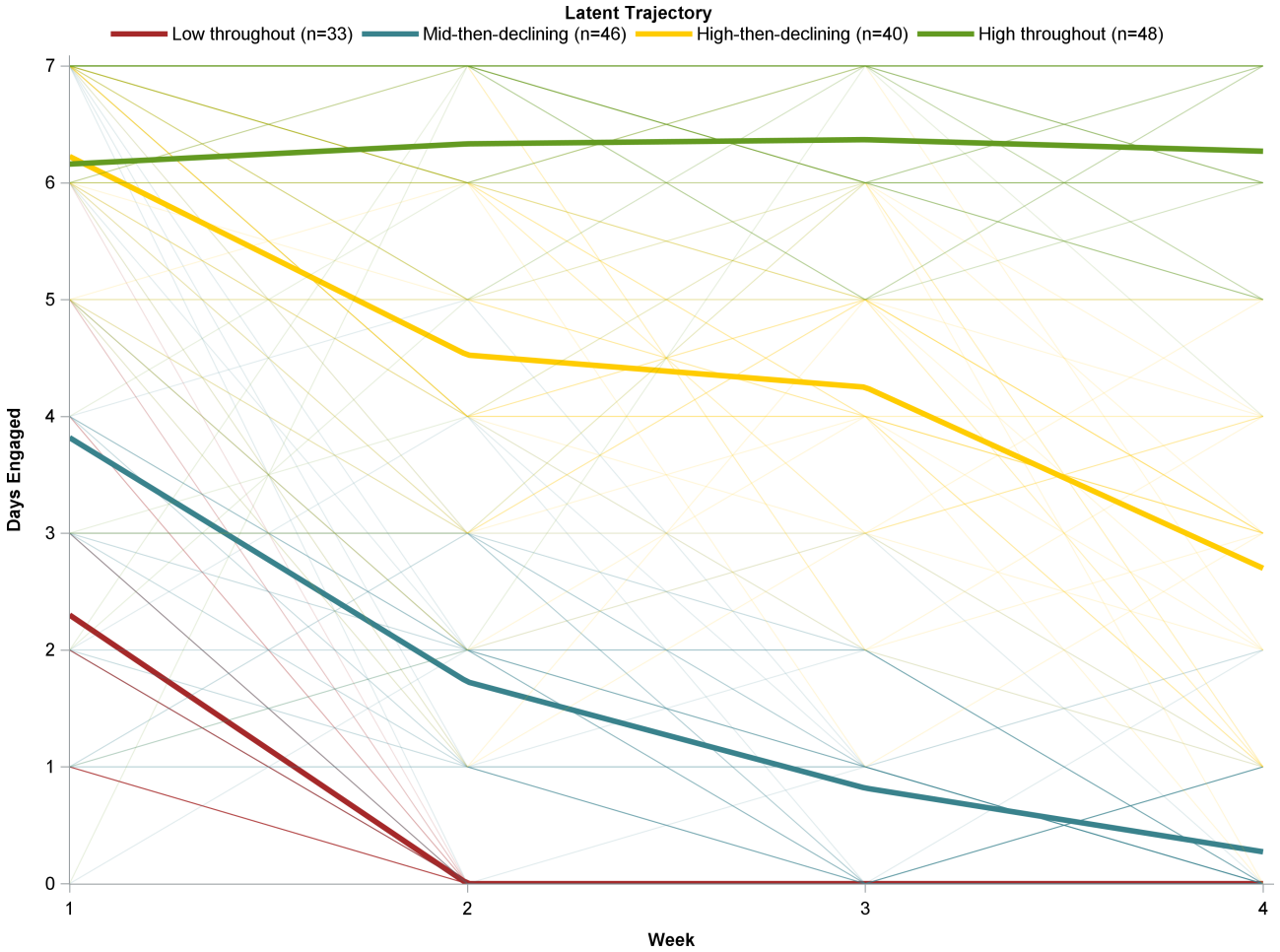
App Engagement

Of the 194 participants, 167 (86.1%) responded to at least one app survey (ie, had at least one day of engagement with the app) and 27 (13.9%) did not respond to any survey during the 4-week study period (no engagement). Among participants who ever engaged with the app, median engagement was 14 (IQR 4 - 23) days out of 28 days; 32.9% (55/167) responded on ≥20 days. The overall median engagement among app users declined from 5 days in week 1 to one day in week 4.

The latent trajectory analysis of the app users revealed 4 patterns of any engagement across the 4 weeks: high-throughout (55/167, 32.9% of participants with any engagement), high-then-declining (40/167, 23.9%), mid-then-declining (47/167, 28.1%), and low-throughout (33/167, 19.7%) (Figure 1). Participants in the high-throughout trajectory tended to use the app on 6 or 7 days every week for all 4 weeks. The high-then-declining group also used the app nearly every day in week 1 but declined in use each week thereafter to around 3 days in week 4. Participants in the mid-then-declining group began use on about half the days in week 1 and then declined over the weeks to use on 0 or 1 day in week 4. In the low-throughout group, participants engaged with the app only in week 1 and then discontinued use.

XSL•FO
RenderX

Figure 1. Latent trajectories for days of app engagement across 4 weeks among participants who had any app use in the Momentary Affect Regulation – Safer Sex Intervention.



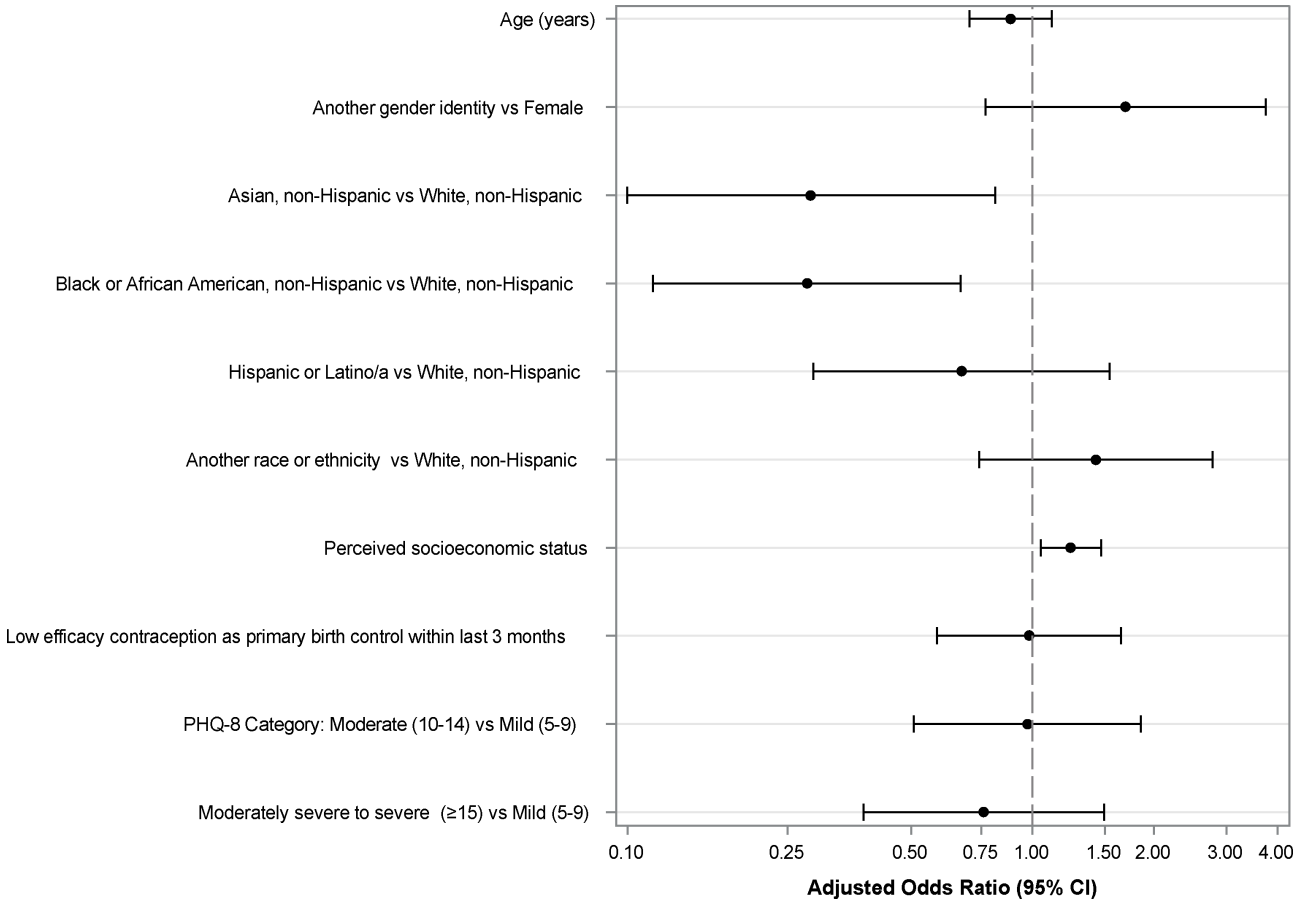
Associations of Sociodemographic Characteristics, Sexual Risk, and Depressive Symptom Severity With App Engagement

In the bivariate analysis (Table 1), there were statistically significant differences in the level of engagement by gender identity (*P*=.04), race and ethnicity (*P*=.006), and report of low-effectiveness contraception as the primary method in last 3 months (*P*=.009). Participants identifying gender other than female were more likely to have high-throughout engagement. Those in the high-then-declining engagement group tended to perceive a higher SES score than other groups. Asian and Black or African American, non-Hispanic participants, and those using low-effectiveness contraception were more likely to have no engagement. There were no differences in the engagement patterns by depressive symptom severity.

In the multivariable model (Figure 2), race and ethnicity remained significantly associated with lower engagement with participants identifying as Asian (AOR=0.28; 95% CI: 0.10, 0.81) or Black or African American, non-Hispanic (AOR=0.28; 95% CI: 0.12, 0.66) less likely to have higher engagement compared to those identifying as White, non-Hispanic. Higher perceived SES was associated with higher engagement (AOR 1.24, 95% CI 1.05-1.48). Gender identity and use of low-effectiveness contraception were no longer statistically significant after adjustment for other covariates (gender identity: AOR 1.70, 95% CI 0.77-3.76; low-effectiveness contraception: AOR 0.98, 95% CI 0.58-1.65). The severity of the depressive symptoms was not associated with the engagement pattern and did not moderate associations between the sociodemographic characteristics and the level of engagement.

XSL•FO
RenderX

Figure 2. Multivariable ordinal logistic regression model examining associations of sociodemographic characteristics, sexual and reproductive health risk, and depressive symptom severity with trajectory of engagement with the Momentary Affect Regulation – Safer Sex Intervention app. PHQ-8: Patient Health Questionnaire-8 items.



Discussion

Main Findings

This study investigates the patterns of mHealth engagement among adolescents and young adults with depressive symptoms assigned to a counseling-plus-mHealth intervention to reduce their risk for unintended pregnancy and STIs. Despite all adolescents and young adults in the study having at least mild depressive symptoms, most participants engaged with the 4-week mHealth component at least once, one-half of the participants engaged on at least 50% of the days, and one-third of the participants engaged on more than 70% of the days. Importantly, for MARSSI, the severity of the depressive symptoms did not appear to influence engagement. Women with depression may benefit from risk reduction interventions to a degree similar to women without depression [34,35]. Additionally, interventions may be more effective in reducing the risk for young women with depression if they address factors related to emotional distress [36-38]. The results of our study add evidence that mHealth interventions to reduce SRH risk that include strategies to manage psychiatric symptoms engage participants, and thus, the effectiveness of these types of interventions can be evaluated across the full range of depressive symptoms.

Across all participants with any mHealth engagement, engagement declined substantially over 4 weeks, as seen with engagement in other longitudinal studies of mHealth interventions [24,39]. However, using latent trajectory analysis, we identified 4 distinct patterns of engagement among those with any use of the mHealth app, demonstrating variability in use over time that was not evident in the overall analysis. More than one-fourth of the participants with any app engagement sustained their engagement for all 4 weeks, and another nearly one-fourth had high initial engagement that declined but remained moderate (on about half of days) through week 4 of EMI; in total, these participants represented nearly one-half of all the participants who enrolled in the app, including those who did not engage at all. Thus, overall declining engagement with an mHealth app may still be associated with continued intervention exposure for a substantial proportion of participants. Furthermore, declining engagement may be a desirable or an undesirable effect in a multiweek mHealth intervention [40]. For example, participants may have decreased app use as a result of habituation or accumulated burden—an undesirable effect of repeated app use. However, declining engagement may have been a function of perceived or actual intervention effect, a desirable consequence of app exposure; in other words, if a participant experiences improvement, that person may determine that the app is not needed as much or at all and elect to engage with it less or discontinue use altogether. Future research will need to determine the thresholds for intervention effectiveness and the reasons for app disengagement. mHealth studies should take an expansive approach to evaluating app engagement, including both querying participants regarding their reasons for

XSL•FO
RenderX

app use or nonuse and examining participant characteristics for associations with engagement.

In this study, we found differences in engagement over time by sociodemographic characteristics. First, participants identifying as Asian or as Black or African American, non-Hispanic, compared to those identifying as non-Hispanic White tended to have lower engagement trajectories. Other studies of mHealth interventions for behavior change [41] and in individuals with psychiatric disorders have found lower app engagement among participants from minoritized racial and ethnic groups, compared to White participants [24,27,42]. Smartphone usage patterns and dependency for online access differ among racial and ethnic groups [43,44]. For example, compared to White teens, Black teens report higher rates of using most common social media platforms and of being on social media almost constantly [43], and Black and Asian adults report higher rates of smartphone dependency compared to White adults [44]. Participants who identified as Asian or as Black may have been lesser engaged with the study app than White participants because they were more likely to have been using their smartphones for other purposes such as accessing social media sites. However, individuals from minoritized racial and ethnic groups have also demonstrated lower engagement with in-person interventions [24,42], suggesting that there may be alternative or additional explanations for the observed differences in engagement, such as race as proxy for social determinants of health or racism [45].

Second, we found that intervention participants with higher perceived SES tended to have higher engagement trajectories. Financial resources may also influence engagement with mHealth interventions. Although nearly all adolescents and young adults in the United States have access to or own a smartphone [46], there may be differences in smartphone usage related to SES. Adults with lower household incomes and those with lower levels of formal education are more likely than their peers to be dependent on a smartphone for online access [44] and thus may have less time free from other smartphone activities to attend to the intervention app. As with the associations of engagement with race and ethnicity, other factors may explain the positive correlation between engagement and SES, such as social context, time use, or mobile phone data plan. It will be important to understand and mitigate disparities in app engagement by sociodemographic characteristics to fully evaluate the effectiveness of mHealth interventions across racial, ethnic, and economic groups.

Limitations

Our findings contribute to the limited literature on mHealth interventions for adolescents and young adults with SRH risk that address emotional distress or depressive symptoms [47], with some caveats. In recruiting participants from across the United States and using both active and passive recruitment methods, this study yields a sample diverse in gender, race, ethnicity, and perceived SES. This diversity permitted analyses into differences in mHealth engagement for groups at particularly high risk, yet underrepresented in mHealth research [48]. However, because there were several distinct engagement patterns, a larger sample was needed to find small differences in engagement by sociodemographic factors that may be important for assessing and improving intervention effectiveness (eg, characteristics of participants who sustained high engagement vs those whose engagement was initially high but then declined). Additionally, engagement was specific to the MARSSI mHealth component, and the patterns and associations may not be generalizable to other interventions. Although MARSSI development employed a human-centered design [28,49], further tailoring of the intervention to address the unique needs of adolescent and young adult participants from minoritized and underresourced groups, rooted in their individual and cultural contexts, may increase their mHealth engagement [50].

Conclusions

In a counseling-plus-mHealth intervention to reduce SRH risk, young people with depressive symptoms demonstrated varied patterns of engagement over 4 weeks of mobile app use. It will be important to establish what level of engagement with the intervention is required to achieve positive outcomes (ie, constitutes effective engagement) [20]. Patterns of mHealth engagement did not differ by level of depressive symptoms, suggesting that mHealth interventions for SRH risk may be applied in those with and without depressive symptoms. mHealth engagement patterns did differ by racial group and perceived SES. Future research is needed to understand and address disparities in engagement with mHealth interventions as a necessary step toward improving evaluations of effectiveness and expanding reach for this rapidly growing method of intervention delivery [20].

Acknowledgments

This research was supported by grant 5TP2AH000076 (author LAS, Principal Investigator) from the Department of Health and Human Services Office of Population Affairs as part of the Teen Pregnancy Prevention Program. Contents are solely the responsibility of Boston Children's Hospital and do not necessarily represent the official views of the Department of Health and Human Services or the Office of Population Affairs. The Policy and Research Group, which conducted the independent evaluation of the intervention, provided the data analyzed herein.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Rossen LM, Hamilton BE, Abma JC, Gregory ECW, Beresovsky V, Resendez AV, et al. Updated methodology to estimate overall and unintended pregnancy rates in the United States. Centers for Disease Control and Prevention. 2023 Apr 12. URL: https://stacks.cdc.gov/view/cdc/124395 [accessed 2025-06-09]
2. Sexually transmitted infections surveillance, 2023. CDC STI Statistics. 2024 Nov 12. URL: https://www.cdc.gov/sti-statistics/annual/index.html [accessed 2024-12-20]
3. Lehrer JA, Shrier LA, Gortmaker S, Buka S. Depressive symptoms as a longitudinal predictor of sexual risk behaviors among US middle and high school students. Pediatrics 2006 Jul;118(1):189-200. [doi: 10.1542/peds.2005-1320] [Medline: 16818565]
4. Burke PJ, Katz-Wise SL, Spalding A, Shrier LA. Intimate relationships and sexual behavior in young women with depression. J Adolesc Health 2018 Oct;63(4):429-434. [doi: 10.1016/j.jadohealth.2018.05.022] [Medline: 30120062]
5. Cooper ML, Shapiro CM, Powers AM. Motivations for sex and risky sexual behavior among adolescents and young adults: a functional perspective. J Pers Soc Psychol 1998 Dec;75(6):1528-1558. [doi: 10.1037//0022-3514.75.6.1528] [Medline: 9914665]
6. Cooper ML, Barber LL, Zhaoyang R, Talley AE. Motivational pursuits in the context of human sexual relationships. J Pers 2011 Dec;79(6):1333-1368. [doi: 10.1111/j.1467-6494.2010.00713.x] [Medline: 21204845]
7. Shrier LA, Feldman HA, Black SK, et al. Momentary affective states surrounding sexual intercourse in depressed adolescents and young adults. Arch Sex Behav 2012 Oct;41(5):1161-1171. [doi: 10.1007/s10508-011-9787-4] [Medline: 21755382]
8. Blood EA, Shrier LA. The temporal relationship between momentary affective states and condom use in depressed adolescents. Arch Sex Behav 2013 Oct;42(7):1209-1216. [doi: 10.1007/s10508-012-0048-y] [Medline: 23297149]
9. Shrier LA, Walls C, Lops C, Feldman HA. Correlates of incorrect condom use among depressed young women: an event-level analysis. J Pediatr Adolesc Gynecol 2011 Feb;24(1):10-14. [doi: 10.1016/j.jpag.2010.04.002] [Medline: 20493734]
10. Istepanian RSH. Mobile health (m-Health) in retrospect: the known unknowns. Int J Environ Res Public Health 2022 Mar 22;19(7):3747. [doi: 10.3390/ijerph19073747] [Medline: 35409431]
11. Lindhiem O, Bennett CB, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. Behav Modif 2015 Nov;39(6):785-804. [doi: 10.1177/0145445515595198] [Medline: 26187164]
12. Smith C, Gold J, Ngo TD, Sumpter C, Free C. Mobile phone-based interventions for improving contraception use. Cochrane Database Syst Rev 2015 Jun 26;2015(6):CD011159. [doi: 10.1002/14651858.CD011159.pub2] [Medline: 26115146]
13. Zulu EM, Sukwa T. Impact of mHealth on contraceptive use among women and men of reproductive age in low- and middle-income countries: a systematic review and meta-analysis. Trop Med Int Health 2020 Oct;25(10):1182-1197. [doi: 10.1111/tmi.13464] [Medline: 33463875]
14. Berendes S, Gubijev A, McCarthy OL, Palmer MJ, Wilson E, Free C. Sexual health interventions delivered to participants by mobile technology: a systematic review and meta-analysis of randomised controlled trials. Sex Transm Infect 2021 May;97(3):190-200. [doi: 10.1136/sextrans-2020-054853] [Medline: 33452130]
15. Chandler R, Guillaume D, Parker A, Wells J, Hernandez ND. Developing culturally tailored mHealth tools to address sexual and reproductive health outcomes among Black and Latina Women: a systematic review. Health Promot Pract 2022 Jul;23(4):619-630. [doi: 10.1177/15248399211002831] [Medline: 33771045]
16. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39. [doi: 10.1348/135910709X466063] [Medline: 19646331]
17. Shrier LA, Walls C, Lops C, Kendall AD, Blood EA. Substance use, sexual intercourse, and condom nonuse among depressed adolescents and young adults. J Adolesc Health 2012 Mar;50(3):264-270. [doi: 10.1016/j.jadohealth.2011.07.005] [Medline: 22325132]
18. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
19. Choi SK, Muessig KE, Hightow-Weidman LB, Bauermeister JA. Paradata: measuring engagement in digital HIV interventions for sexual and gender minorities. Curr HIV/AIDS Rep 2023 Dec;20(6):487-501. [doi: 10.1007/s11904-023-00679-5] [Medline: 37930613]
20. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
21. Beck AT. The evolution of the cognitive model of depression and its neurobiological correlates. Am J Psychiatry 2008 Aug;165(8):969-977. [doi: 10.1176/appi.ajp.2008.08050721] [Medline: 18556777]
22. Rock PL, Roiser JP, Riedel WJ, Blackwell AD. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol Med 2014 Jul;44(10):2029-2040. [doi: 10.1017/S0033291713002535] [Medline: 24168753]
23. DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. Arch Intern Med 2000 Jul 24;160(14):2101-2107. [doi: 10.1001/archinte.160.14.2101] [Medline: 10904452]
24. Ben-Zeev D, Scherer EA, Gottlieb JD, et al. mHealth for schizophrenia: patient engagement with a mobile phone intervention following hospital discharge. JMIR Ment Health 2016 Jul 27;3(3):e34. [doi: 10.2196/mental.6348] [Medline: 27465803]

XSL•FO
RenderX

25. Cormack F, McCue M, Taptiklis N, et al. Wearable technology for high-frequency cognitive and mood assessment in major depressive disorder: longitudinal observational study. JMIR Ment Health 2019 Nov 18;6(11):e12814. [doi: 10.2196/12814] [Medline: 31738172]
26. Molloy A, Anderson PL. Engagement with mobile health interventions for depression: a systematic review. Internet Interv 2021 Dec;26:100454. [doi: 10.1016/j.invent.2021.100454] [Medline: 34621626]
27. Arean PA, Hallgren KA, Jordan JT, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330. [doi: 10.2196/jmir.6482] [Medline: 27998876]
28. Shrier LA, Burke PJ, Parker S, et al. Development and pilot testing of a counseling-plus-mHealth intervention to reduce risk for pregnancy and sexually transmitted infection in young women with depression. Mhealth 2020;6:17. [doi: 10.21037/mhealth.2019.11.05] [Medline: 32270009]
29. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
30. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
31. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012 Feb 21;184(3):E191-E196. [doi: 10.1503/cmaj.110829] [Medline: 22184363]
32. Goodman E, Adler NE, Kawachi I, Frazier AL, Huang B, Colditz GA. Adolescents' perceptions of social status: development and evaluation of a new indicator. Pediatrics 2001 Aug;108(2):e31-e31. [doi: 10.1542/peds.108.2.e31] [Medline: 11483841]
33. Jones BL, Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociol Methods Res 2007 May;35(4):542-571. [doi: 10.1177/0049124106292364]
34. Holden AEC, Shain RN, Miller WB, et al. The influence of depression on sexual risk reduction and STD infection in a controlled, randomized intervention trial. Sex Transm Dis 2008 Oct;35(10):898-904. [doi: 10.1097/OLQ.0b013e31817d7a33] [Medline: 18607311]
35. Sales JM, Lang DL, Hardin JW, Diclemente RJ, Wingood GM. Efficacy of an HIV prevention program among African American female adolescents reporting high depressive symptomatology. J Womens Health (Larchmt) 2010 Feb;19(2):219-227. [doi: 10.1089/jwh.2008.1326] [Medline: 20109119]
36. Brown LK, Nugent NR, Houck CD, et al. Safe Thinking and Affect Regulation (STAR): human immunodeficiency virus prevention in alternative/therapeutic schools. J Am Acad Child Adolesc Psychiatry 2011 Oct;50(10):1065-1074. [doi: 10.1016/j.jaac.2011.06.018] [Medline: 21961780]
37. Sterk CE, Theall KP, Elifson KW. The impact of emotional distress on HIV risk reduction among women. Subst Use Misuse 2006;41(2):157-173. [doi: 10.1080/10826080500391639] [Medline: 16393740]
38. Compton WM, Cottler LB, Ben-Abdallah A, Cunningham-Williams R, Spitznagel EL. The effects of psychiatric comorbidity on response to an HIV prevention intervention. Drug Alcohol Depend 2000 Mar 1;58(3):247-257. [doi: 10.1016/S0376-8716(99)00097-6] [Medline: 10759035]
39. Egilsson E, Bjarnason R, Njardvik U. Usage and weekly attrition in a smartphone-based health behavior intervention for adolescents: pilot randomized controlled trial. JMIR Form Res 2021 Feb 17;5(2):e21432. [doi: 10.2196/21432] [Medline: 33481750]
40. Nahum-Shani I, Yoon C. Towards the science of engagement with digital interventions. Curr Dir Psychol Sci 2024 Aug;33(4):239-246. [doi: 10.1177/09637214241254328] [Medline: 39552747]
41. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146. [doi: 10.2196/jmir.2665] [Medline: 23859884]
42. van Aubel E, Vaessen T, Uyttebroek L, et al. Engagement and acceptability of acceptance and commitment therapy in daily life in early psychosis: secondary findings from a multicenter randomized controlled trial. JMIR Form Res 2024 Nov 21;8:e57109. [doi: 10.2196/57109] [Medline: 39570655]
43. Anderson M, Faverio M, Gottfried J. Teen, social media and technology 2023. Pew Research Center. 2023 Dec 11. URL: https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/ [accessed 2024-12-20]
44. Mobile fact sheet. Pew Research Center. 2024 Nov 13. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-12-20]
45. Snyder CR, Mohammed SA. A critical race theory analysis of mental health disparities research. J Racial Ethn Health Disparities 2024 Dec;11(6):3900-3906. [doi: 10.1007/s40615-023-01840-x] [Medline: 37884854]
46. Vogels EA, Gelles-Watnick R, Massarat N. Nearly all teens in 2022 have access to a smartphone, up from 73% in 2014-15. Pew Research Center. 2022 Aug 3. URL: https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/pj_2022-08-10_teens-and-tech_0-04-png/ [accessed 2024-12-20]
47. Hammett JF, Stappenbeck CA. Preliminary efficacy of a web-based alcohol and emotion regulation intervention on sexual risk behaviors among college women. J Stud Alcohol Drugs 2024 Sep;85(5):643-647. [doi: 10.15288/jsad.23-00420] [Medline: 38647105]
48. Litke SG, Resnikoff A, Anil A, et al. Mobile technologies for supporting mental health in youths: scoping review of effectiveness, limitations, and inclusivity. JMIR Ment Health 2023 Aug 23;10:e46949. [doi: 10.2196/46949] [Medline: 37610818]

XSL•FO
RenderX

49. Shrier LA, Burke PJ, Jonestrask C, Katz-Wise SL. Applying systems thinking and human-centered design to development of intervention implementation strategies: an example from adolescent health research. J Public Health Res 2020 Oct 14;9(4):1746. [doi: 10.4081/jphr.2020.1746] [Medline: 33117757]
50. Huh J, Koola J, Contreras A, et al. Consumer health informatics adoption among underserved populations: thinking beyond the digital divide. Yearb Med Inform 2018 Aug;27(1):146-155. [doi: 10.1055/s-0038-1641217] [Medline: 30157518]

Abbreviations

AOR: adjusted odds ratio
EMI: ecological momentary intervention
MARSSI: Momentary Affect Regulation – Safer Sex Intervention
mHealth: mobile health
PHQ: Patient Health Questionnaire
SES: socioeconomic status
SRH: sexual and reproductive health
STI: sexually transmitted infection

Edited by M Sobolev; submitted 24.12.24; peer-reviewed by R Eckhoff, Y Li; revised version received 31.03.25; accepted 19.05.25; published 11.07.25.

Please cite as:
Shrier LA, Milliren CE, Ciriello B, O'Connell MM, Harris SK
Patterns of Engagement With the mHealth Component of a Sexual and Reproductive Health Risk Reduction Intervention for Young People With Depression: Latent Trajectory Analysis
JMIR Mhealth Uhealth 2025;13:e70219
URL: https://mhealth.jmir.org/2025/1/e70219
doi:10.2196/70219

© Lydia A Shrier, Carly E Milliren, Brittany Ciriello, Madison M O'Connell, Sion Kim Harris. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Integrating CEN ISO/TS 82304-2 in the Catalan Health App Assessment Framework: Comparative Case Study

Berta Llebot Casajuana[1], BSc; Petra Hoogendoorn[2], MSc; Maria Villalobos-Quesada[2], PhD; Carme Pratdepàdua Bufill[1], MSc

[1]Fundació TIC Salut Social, Ministry of Health, Government of Catalonia, Aragó 330-332, Barcelona, Spain

[2]National eHealth Living Lab, Public Health and Primary Care Department, Leiden University Medical Centre, Albinusdreef 2, Leiden, The Netherlands

Corresponding Author:
Berta Llebot Casajuana, BSc
Fundació TIC Salut Social, Ministry of Health, Government of Catalonia, Aragó 330-332, Barcelona, Spain

Abstract

Background: Health apps are increasingly being used to promote health, manage diseases, and deliver health care services. Still, there is scarce objective information regarding their quality beyond the required Conformité Européenne mark for medical apps, leading to potential risks for users. To address these challenges, several authorities have developed health app assessment frameworks. In 2017, the TIC Salut Social Foundation (FTSS) in Catalonia developed its own health app assessment framework, which has been in use since that year. The publication of CEN ISO/TS 82304‐2 (abbreviated as 82304‐2)—a Technical Specification for assessing health apps—and the cocreation of the Label2Enable 82304‐2 handbook for certified assessment organizations provide a unique opportunity to harmonize app assessments across the European Union.

Objective: This study aimed to perform a comparative analysis of the FTSS assessment framework with 82304‐2 to explore the integration of 82304‐2 in Catalonia. Our broader aim was to provide this methodology for health authorities elsewhere to consider integrating 82304‐2 or other evaluation frameworks.

Methods: For the comparative analysis, a mixed methods approach was used, combining a qualitative case study with a quantitative analysis of the 2 frameworks. The qualitative evaluation covered rationale for assessment, framework characteristics, governance, workflows, quality aspects, and quality requirements. For the quantitative analysis, all FTSS and 82304‐2 requirements were translated into concepts and subconcepts. A scoring system identified matches of the frameworks with these subconcepts, with scores ranging from 0 (no match) to 0.5 (partial match) and 1 (full match). Integration was evaluated considering several scenarios, including adopting the Label2Enable 82304‐2 handbook, adopting the 82304‐2 requirements, adapting the 82304‐2 requirements to local needs, and maintaining the current FTSS framework.

Results: The main difference between the frameworks was the app usage–based assessment (FTSS) versus evidence- and app usage–based assessment (82304‐2). All 120 FTSS requirements and 74 quality aspect–related 82304‐2 requirements were translated into 78 concepts and 97 subconcepts. Overall, 48% (47/97) of the subconcepts were found in both frameworks, 39% (37.5/97) were specific to 82304‐2, and 13% (12.5/97) were specific to FTSS. All 82304‐2-specific subconcepts and thus all 82304‐2 requirements were found to be relevant to FTSS. FTSS decided to integrate (adopt and adapt) all 74 82304‐2 requirements. In total, 5 FTSS-specific requirements were included in the Label2Enable 82304‐2 handbook, while another 4 rigor-enhancing requirements, 1 scope-expanding requirement, and 1 context-specific requirement would be assessed on top.

Conclusions: The comprehensive comparative analysis of the FTSS framework and 82304‐2 enabled FTSS decision-making to integrate all 82304‐2 quality requirements and adopt the Label2Enable 82304‐2 handbook in the future. The many new and all relevant 82304‐2 concepts, the rigor of the handbook, and the few remaining FTSS-specific requirements are expected to be indicative of 82304‐2's potential to make harmonized, robust health app assessments common in Catalonia and elsewhere. FTSS encourages other authorities to perform a similar evaluation.

(JMIR Mhealth Uhealth 2025;13:e67858) doi:10.2196/67858

KEYWORDS

mHealth; mobile health; mobile applications; assessment; criteria; quality; regulatory; comparative analysis; health app; comparative case study; Catalonia; health apps assessment; ISO; certification; CEN ISO/TS 82304-2; mixed methods; framework; TIC Salut Social Foundation

XSL•FO
RenderX

Introduction

Background

Health apps refer to apps designed for managing, maintaining, or improving the health and well-being of individual persons or the delivery of care, and thus include both medical and wellness apps [1,2]. A great number of health apps are already being used by millions of people, and several reports have pointed out the potential of these apps and digital technologies to deliver the right information to the right people at the right time. Health apps can enhance the safety, effectiveness, and efficiency of care by promoting health, contributing to the prevention and treatment of diseases, facilitating patients' equal access to personalized health care, and engaging patients in their own care. They can also respond to unmet patient needs, enable better coordination of care, exploit the possibilities of remote health promotion and remote care, and strengthen the resilience and sustainability of health systems [3-5]. However, not all health apps provide sufficient evidence to support their claims, putting users at risk and calling into question their quality [6]. This creates a need for robust evaluation and regulatory oversight [3,4].

The European Union (EU) Medical Device Regulation (MDR) mandates a conformity assessment and related Conformité Européenne (CE) mark for all medical devices, including software, and is therefore applicable to medical apps [7-9]. However, while the MDR ensures patient safety, it produces merely a CE mark, not a quality report. Additionally, Notified Bodies are only involved in the evaluation of class IIa, IIb, and III medical devices; class I medical devices are not subject to third-party assessment. Wellness apps, on the other hand, fall outside the scope of the MDR, and digital marketplaces do not require robust evaluations [6]. As a result, health apps available on the market largely lack essential quality information apart from the user star ratings in app stores, which are known to be a poor indicator of quality [10]. To address this situation, several countries and regions in Europe and worldwide have developed assessment frameworks for health apps [11]. These frameworks assess characteristics beyond the MDR, such as data privacy, and provide more assessment information than a CE mark to support decision-making. The TIC Salut Social Foundation (FTSS, Catalan abbreviation), within the Ministry of Health of Catalonia (Spain), was in 2017 one of the first European health authorities to develop such an assessment framework [12,13]. With the emergence of more frameworks, a lack of cross-country harmonization has been introduced, leading to a significant duplication of app assessment efforts for manufacturers and health care systems, causing confusion among app users [2]. In addition, despite the widespread interest in the use of health apps across Europe, even authorities with relatively well-established frameworks encounter challenges with efficient implementation [11].

The World Health Organization (WHO) has called on health authorities and researchers to establish a common methodology for evaluating health apps, highlighting the need for these evaluations to become a standard practice rather than an exception [3]. Founded upon the premise that international harmonization could highly benefit the integration of apps in health care systems [2], the European Commission commissioned the European Committee for Standardization (CEN) to develop a Technical Specification (TS) for quality and reliability of health and wellness apps, which was achieved in collaboration with the International Organization for Standardization (ISO). CEN ISO/TS 82304‐2 "Health software—Part 2: Health and wellness apps—Quality and Reliability" (henceforth referred to as 82304‐2 or the TS) was published in 2021 [1,2,14]. As a next step, the EU-funded Label2Enable project (2022‐2024) has been dedicated to facilitating the adoption of this TS as the common assessment framework within the EU and potentially beyond. The project has iteratively cocreated the Label2Enable handbook for app assessment with 82304‐2 and the Label2Enable certification scheme [15].

Objective

The objective of this study was to generate a systematic comparative analysis of the FTSS assessment framework for health apps with the 82304‐2 framework, to identify all new and relevant concepts distinct to 82304‐2 and all Catalonia-specific concepts addressed by the FTSS framework, and to adequately consider and allow proper integration of 82304‐2 in Catalonia. Our broader aim was to provide this methodology to health authorities in other regions and countries to make informed decisions about the adoption and integration of 82304‐2 or other evaluation frameworks.

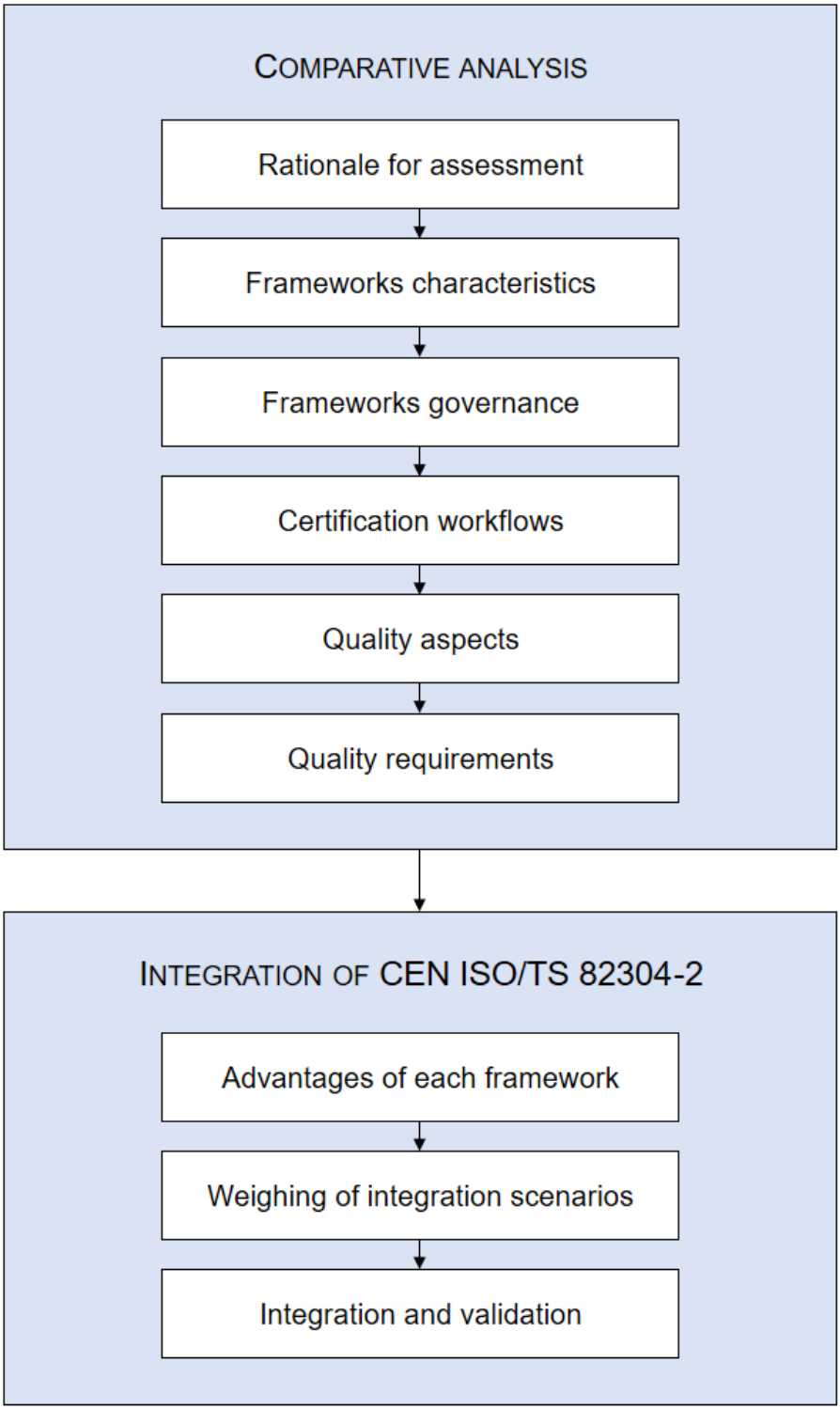
Methods

Study Design

The project was carried out in 2 parts, as depicted in Figure 1. The first part, a comprehensive comparative analysis of both assessment frameworks, is divided into 6 sections. This part aimed to identify the characteristic traits and differences of each assessment framework. The second part is divided into 3 sections and involves considering the integration of 82304‐2 in Catalonia, weighing its advantages and disadvantages, and carrying out the decision made.

XSL•FO
RenderX

Figure 1. Overview of the methodology used in this study.



Comparative Analysis

The comparative analysis was conducted using a mixed methods approach, comprising a qualitative case study, given its capacity to capture complexity, and a quantitative analysis. This methodology [16] was adapted from a comparative analysis of assessment frameworks for sustainability [17]. The analysis addressed qualitatively and descriptively the (1) rationale for assessment, (2) assessment framework characteristics, (3) assessment framework governance, (4) certification workflows and techniques used, (5) quality aspects and subaspects of each framework, and (6) quality requirements (or criteria).

To enable a quantitative analysis of the FTSS and 82304‐2 requirements shared across the 2 assessment frameworks (6), first, all concepts and subconcepts assessed in each of the 82304‐2 and FTSS requirements were identified by 2 researchers (BLC and CPB). The concepts and subconcepts were organized within the 82304‐2 subaspects (“Health

XSL•FO
RenderX

requirements," "Health risks," "Ethics," "Health benefit," "Societal benefit," "Accessibility," "Usability," "Privacy," "Security," "Technical robustness," and "Interoperability"). In cases where multiple requirements addressed the same overarching concept, distinctions were made to recognize when they assessed different subconcepts. For the case of FTSS requirements, concepts were extracted both from the FTSS quality requirements and from the FTSS assessment submission form. In some instances, individual requirements were broken down into 2 distinct subconcepts to account for a proper comparison. Similarly, multiple requirements were grouped under the same concept and subconcept when applicable.

To compare the concepts and subconcepts of both frameworks, first, the obtained 82304‐2 subconcepts were searched within the FTSS requirements. The process was repeated in the opposite direction for FTSS subconcepts in 82304‐2 requirements. When finding matches, the original requirements of 82304‐2 and FTSS were compared. In total, 3 outcomes were possible: 0="no match," 0.5="partial match," and 1="full match," considering aspects such as the intended assessment goal, the level of detail required, and any differences in scope. Further detail as to the degree and type of partial match was deemed unnecessary for this paper, as the distinction between "no match" and "partial match" was primarily used to determine which FTSS requirements to discuss with the lead expert of 82304‐2 (PH) to ensure content was not accidentally overlooked and subconcepts were adequately identified as "partial match" or "no match." Consensus on this identification was reached for all concepts and subconcepts.

Then, overall match scores were computed. The sum of all full matching subconcepts and partial matching subconcepts*0.5 was labeled "Both" (meaning that both frameworks assess this subconcept), the sum of all not matching subconcepts and partial matching subconcepts*0.5 applicable to 82304‐2 as "82304‐2 only," and the sum of all not matching subconcepts and partial matching subconcepts*0.5 applicable to FTSS as "FTSS only." Overall match scores were calculated as the sum of all full matching subconcepts and partial matching subconcepts*0.5 divided by the total number of subconcepts. The values were then converted to a percentage and plotted with R.

Integration of CEN ISO/TS 82304-2

The second part of our study consisted of an analysis of the results of the comparative analysis to decide on the integration of 82304‐2 into the FTSS assessment framework. The analysis was carried out by 2 researchers (BLC and CPB). Integration would imply using the concepts and subconcepts identified in 82304‐2 for the assessment of health apps' quality in Catalonia.

First, the advantages and disadvantages of each assessment framework and their certification workflow were identified and discussed. Next, the integration scenarios were considered—whether (1) to adopt the Label2Enable 82304‐2 certification scheme and related handbook adding only Catalonia-specific requirements, (2) to adopt 82304‐2 requirements into the FTSS framework (using the phrasing of 82304‐2), (3) to adapt 82304‐2 requirements according to FTSS needs, or (4) to maintain the current FTSS assessment framework in Catalonia.

Ethical Considerations

Approval from an ethics committee was not required under Catalan and Spanish regulations, as the study did not involve human participants, personal health data, or clinical interventions. According to Law 14/2007 on Biomedical Research (Ley 14/2007 de Investigación Biomédica) and the guidelines of the Research Ethics Committee of Catalonia (Comitè d'Ètica de la Investigació amb Medicaments de Catalunya, CEIm), studies based solely on document analysis and framework comparison do not require ethical approval. Additionally, no identifiable personal or patient data were processed.

Results

Comparative Analysis

Overview

Key findings from the comparative analysis included a significant alignment between the FTSS assessment framework and 82304‐2, with quite similar rationales for assessment, framework governance, and certification workflows. The main difference between the frameworks was the app usage–based assessment (FTSS) versus evidence- and app usage–based assessment (82304‐2). Also, the 82304‐2 framework provided many concepts new to FTSS, which were all considered relevant.

Rationale for Assessment

In Catalonia, at the beginning of 2017, the Government requested the establishment of the FTSS assessment framework to ensure the quality of health apps used within the Catalan health care system and to promote the use of these health apps [18]. The development of 82304‐2 was commissioned by the European Commission to tackle the scattered assessment landscape with its many different criteria that struggle to scale, duplicate work, produce contradicting results, create challenges for manufacturers, and generally ignore international standards applicable to health apps [1,2].

Both the FTSS and 82304‐2 rationale for assessment included the large scale (hundreds of thousands of health and wellness apps, millions of downloads), concerns about the risks (many apps collect sensitive personal information and provide health advice that may not be supported by any evidence; app stores have limited evaluations; only some apps fall under the MDR), and opportunities missed (apps that are proven effective in addressing, for example, unhealthy lifestyles, chronic diseases, access to and affordability of health and care are not necessarily adopted at scale and reimbursed).

Framework Characteristics

Table 1 summarizes and compares the main characteristics of each assessment framework [1,2,12,15,18-22]. More detailed information can be found in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Comparison of the main characteristics of the TIC Salut Social Foundation (FTSS) assessment framework and CEN ISO 82304‐2 Technical Specification (TS or 82304‐2).

	FTSS assessment framework	CEN ISO/TS 82304‐2
Year of publication	2017	2021
Scope	Catalan health care system	International (initiative Europe)
Scientific evidence	The FTSS framework was based on clinical guidelines and research papers (with input of relevant stakeholders in Catalonia).	The 82304‐2 framework was based on a Delphi study with 83 experts from 8 stakeholder groups and 6 continents. The Label2Enable 82304‐2 handbook "pass/fails," and subquestions are informed by international scientific studies on health app quality and risks. The Label2Enable project included many studies to assess and fine=tune compatibility of 82304‐2 related products with mostly European multi-stakeholder needs.
Consensus	Relevant stakeholders in Catalonia	Key stakeholders, mostly European
Product information	Obtained from the FTSS assessment submission form	1 aspect: product information
Quality aspects	4 quality aspects: • Clinical contents and functionality • Usability, accessibility, and design • Security and privacy of data • Technological robustness	4 quality aspects: • Healthy and safe • Easy to use • Secure data • Robust build
Quality requirements	120 quality requirements	74 quality requirements within the 4 quality aspects (81 requirements with the additional 7 requirements in "Product information")
Score-impacting requirements	Maximum 120	Maximum 67
Structure of each requirement	Affirmative sentence with: • Description • Levels of obligatoriness	Question with: • Condition • Purpose • Response options • Evidence • Notes
Assessed apps	20 for testing, 55 for assessment, 23 as smoke test	35 for testing (11 in development of 82304‐2, 24 in development of the Label2Enable 82304‐2 handbook)
Certified apps	14 out of 55 assessed	0 (given testing phase)
Results communication	Certification seal + internal report	Quality label + quality report

Framework Governance

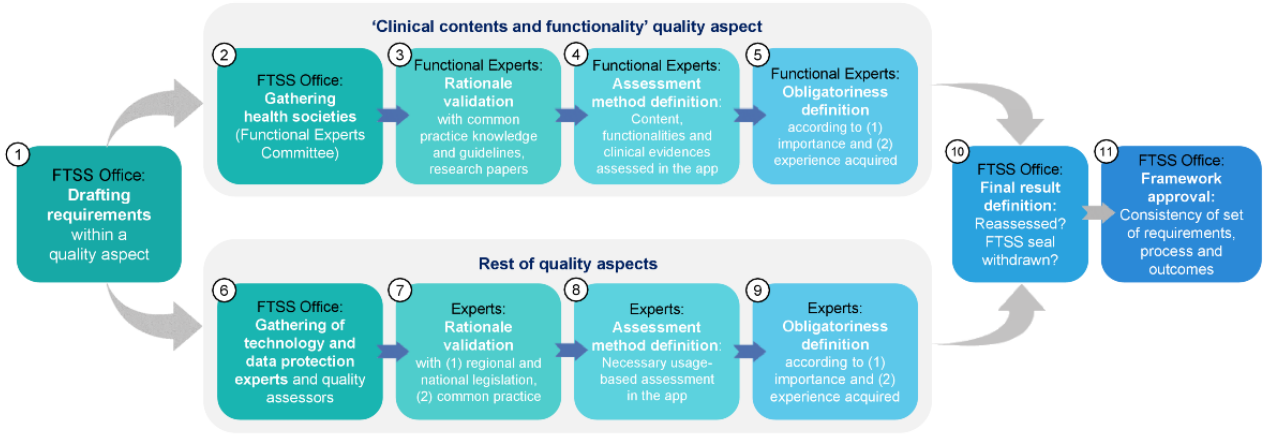
Both FTSS and Label2Enable have a stakeholders and experts organization to help approve, maintain, and improve their framework. Both FTSS and Label2Enable have defined a methodology of maintenance, with many steps overlapping (see Figure 2 for FTSS and Figure 3 for Label2Enable). Table 2 outlines and compares these maintenance methodologies. A broader explanation of each governance process can be found in Multimedia Appendix 2.

XSL•FO
RenderX

Table . Comparison of the main traits of the maintenance process of the TIC Salut Social Foundation (FTSS) assessment framework and the CEN ISO/TS 82304 - 2 (82304 - 2) Label2Enable handbook, outlined in Figures 2 and 3 (and referenced in this table).

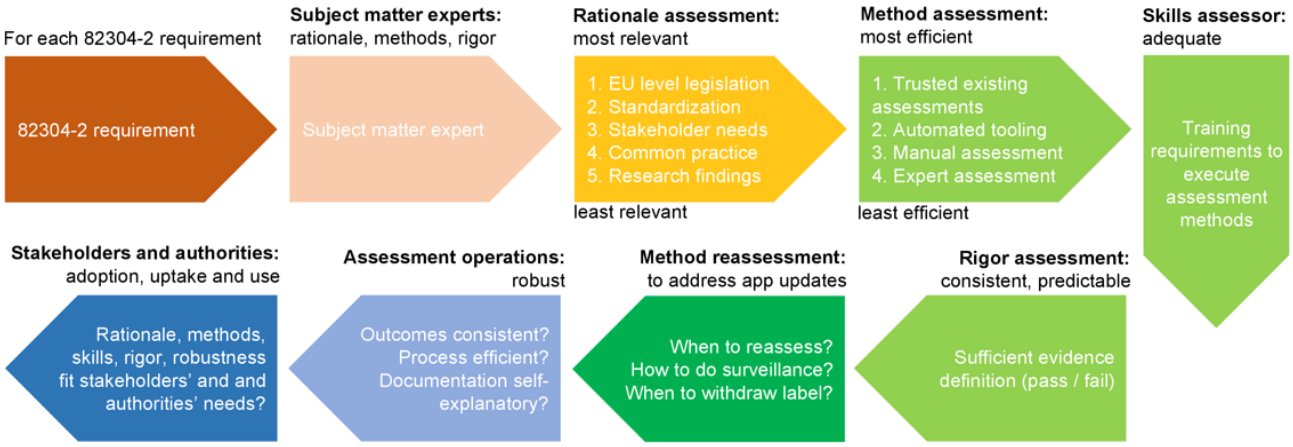
	FTSS assessment framework	CEN ISO/TS 82304 - 2: Label2Enable handbook
Maintenance cycle	Maintained every 2-3 years	Intent to maintain annually and more frequently if needed
Methodology (steps 1 and 2 of Figures 2 and 3)	Experts to consult recruited per quality aspect	Experts to consult recruited for each quality requirement
Experts (step 2 of Figures 2 and 3)	Societies of health professionals [23], technology and data protection experts, quality assessors	Subject matter experts and key stakeholders
Rationale validation (step 3 of Figures 2 and 3)	Legislation, stakeholder needs, and common practice	EU level legislation, stakeholder needs, common practice, standardization, and scientific research findings
Assessment method (step 4 of Figures 2 and 3)	Decided for each requirement. Can include expert and trusted existing assessments	Evolves the assessment method and efficiency from expert to manual, where applicable, using automated and trusted existing assessments
Obligatoriness definition (step 5 of Figure 2)	Establishes which requirements are mandatory to obtain the certificate, which can be reviewed over time	Establishes which requirements are mandatory to qualify for the quality label, which can be reviewed over time
Assessors training (step 5 of Figure 3)	Not contemplated by FTSS as assessors have previous knowledge and experience	Considers the needed skills of the assessor and its training based on the assessment methods
Evidence definition (step 6 of Figure 3)	No	Defines the specific evidence needed to pass or fail for each requirement
Impact (step 9 of Figure 3)	Not evaluated by FTSS per se, but the FTSS framework was built by considering its impact	Explores the "compatibility" with what key stakeholders and authorities consider useful (and proportionate)

Figure 2. Methodology of definition and maintenance of each quality aspect within the TIC Salut Social Foundation (FTSS) assessment framework.



XSL•FO
RenderX

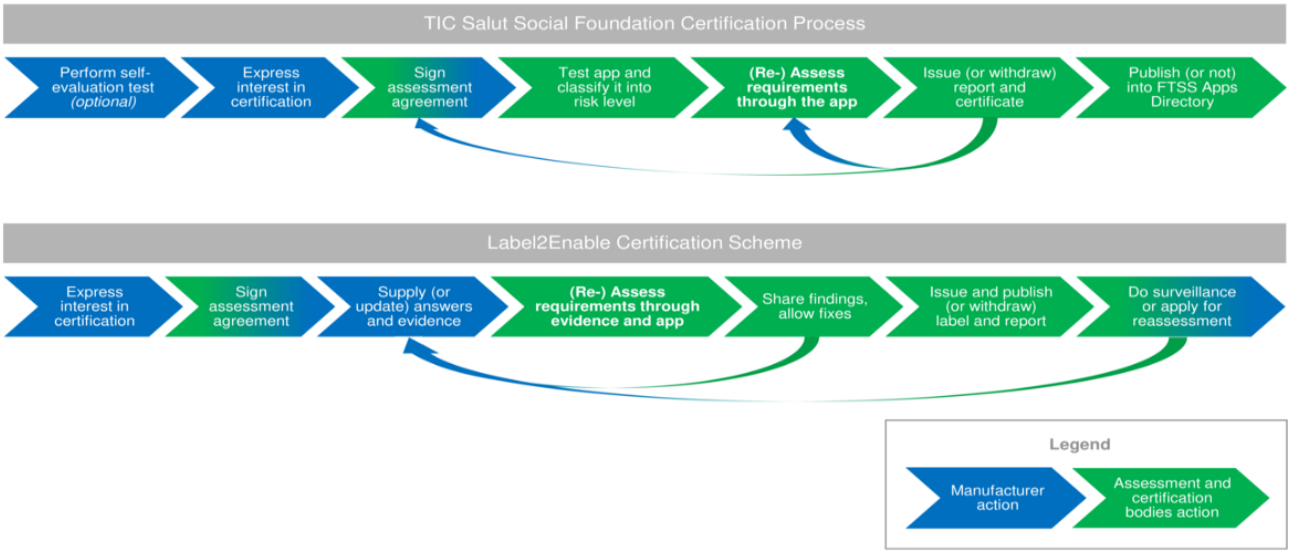
Figure 3. Methodology to cocreate the Label2Enable handbook for CEN ISO/TS 82304‐2 (82304‐2) for use by certified app assessment organizations. EU: European Union.



Certification Workflows

The certification workflows of FTSS [21] and 82304‐2 (as defined by Label2Enable) are quite comparable (Figure 4). The main difference between the two is that the FTSS assessment is mostly app usage–based, while the Label2Enable 82304‐2 assessment is evidence- and app usage–based. In FTSS, assessors use the app and analyze its content and functionalities to decide on the pass or fail of each requirement. Following the Label2Enable 82304‐2 handbook, the manufacturer must supply evidence for each quality requirement that they claim to meet. Subsequently, a conformity assessment body assesses the evidence to determine the pass or fail, also checking the app. Other differences are that FTSS includes an optional self-evaluation test [24] and publishes certified apps in the FTSS directory. Label2Enable intends to provide a verification service of the label it supplies and proposes to generate a directory. For more detail about the certification workflows, see Multimedia Appendix 3.

Figure 4. Overview of the certification workflows of TIC Salut Social Foundation (FTSS) (above) and the Label2Enable certification scheme (below).

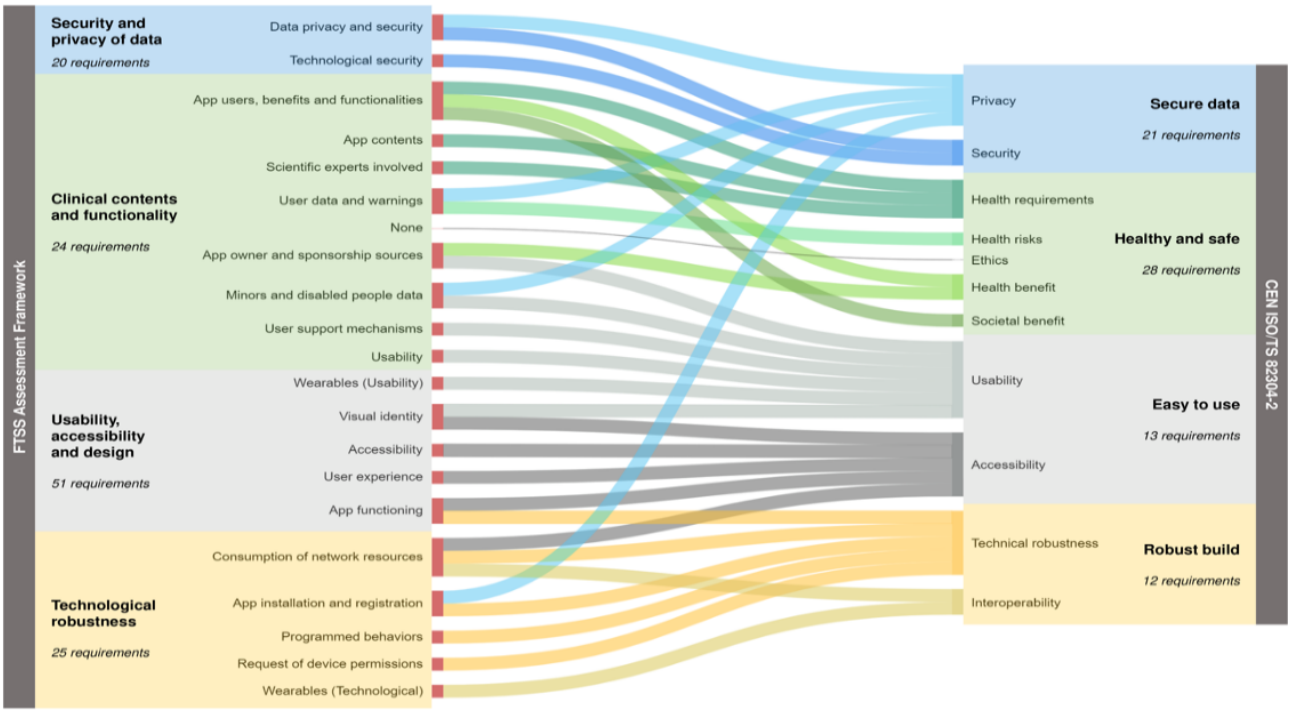


Quality Aspects

The FTSS and the 82304‐2 assessment frameworks both contain 4 overarching quality aspects [2,20]. Based on the semantics of each aspect and the concepts and subconcepts assessed in each, a match between the aspects of the 2 frameworks was observed (Figure 5). On a more detailed level, differences in distribution applied, which highlighted a need to consider assessment frameworks in its entirety in a comparative analysis, instead of merely considering semantically similar subsections. The organization of 82304‐2 seems a bit more logical; FTSS has some scattered requirements when compared to 82304‐2. For example, the quality aspect "Clinical contents and functionality" of FTSS is broader than "Healthy and safe" of 82304‐2, with many of FTSS' subaspects connecting to 82304‐2's "Easy to use" and some to "Secure data." "Technological robustness" is also broader than "Robust build." The concepts gathered in 82304‐2 subaspect "Privacy" are distributed among 4 FTSS subaspects, from 3 different quality aspects. All in all, concepts assessed are fairly common in both frameworks, but wording used for aspects and subaspects within FTSS and 82304‐2 tends to differ, with some exceptions, such as "Usability" and "Accessibility."

XSL•FO
RenderX

Figure 5. Matches found between the quality aspects and subaspects of the TIC Salut Social Foundation (FTSS) and the CEN ISO/TS 82304 - 2 (82304 - 2) assessment frameworks. The 4 quality aspects of FTSS (eg, "Security and privacy of data") are displayed in bold on the left, and the 4 quality aspects of 82304 - 2 (eg, "Secure data") in bold on the right. The colored boxes also contain the number of quality requirements within the quality aspect and the subaspects in which these requirements are clustered. The similar background colors (blue, green, grey, and yellow) indicate a match of the quality aspects between the 2 frameworks. The lines in the middle connect the concepts within the FTSS and 82304 - 2 quality requirements that were similar. The color coding of the connecting lines differs per 82304 - 2 subaspect it originates from and is merely used to support easy recognition of the distribution of the 82304 - 2 concepts across FTSS subaspects. Note that this is a qualitative analysis; thus, the thickness of the connection lines is not numerically representative.



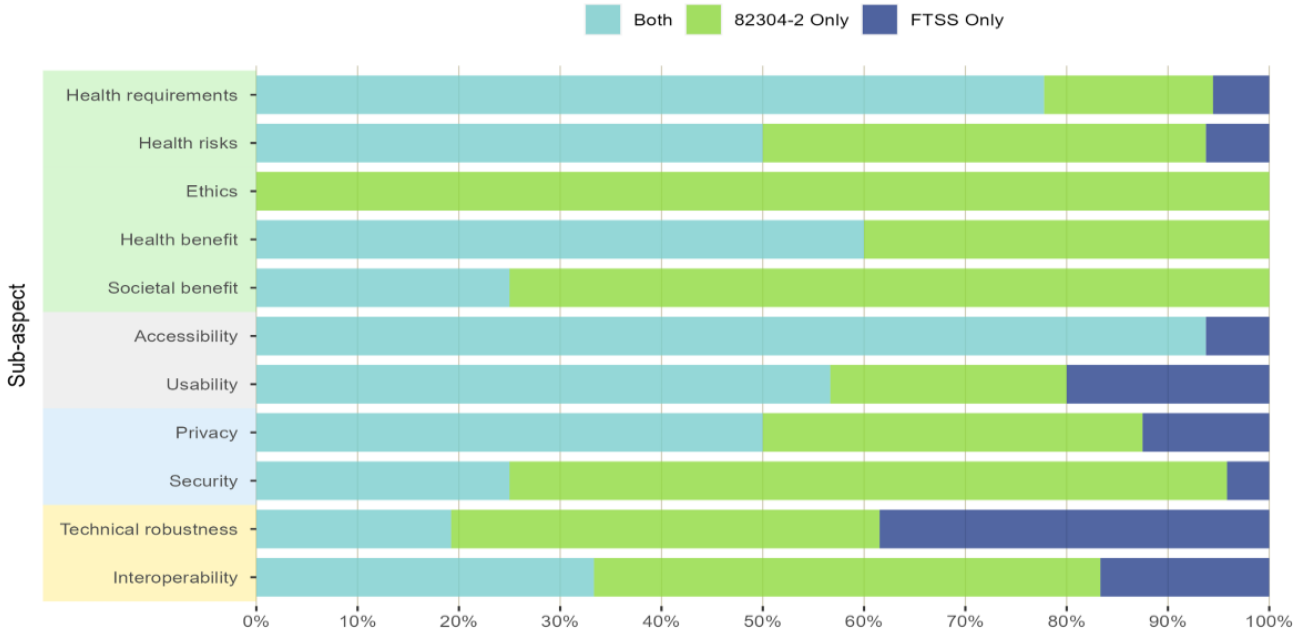
Quality Requirements

To enable a comparison of all the 82304 - 2 and FTSS quality requirements, in total, 78 different concepts and 97 subconcepts were extracted from all 74 quality requirements of the 82304 - 2 framework and all 120 quality requirements of the FTSS framework combined. Overall, 48% (47/97) of the subconcepts of both frameworks were a partial or full match, 39% (37.5/97) were 82304 - 2-specific ("82304 - 2 only") and 13% (12.5/97) FTSS-specific ("FTSS only"). The subaspect "Accessibility" had the highest match, with 94% (7.5/8) of subconcepts assessed in both frameworks, and FTSS addressing another 6% (0.5/8) (see Figure 6). Other subaspects such as "Health requirements," "Health risks," "Health benefit," and "Usability" had at least 50% of subconcepts covered by both frameworks. The 82304 - 2 requirements were found to cover most of the FTSS requirements, ranging from subaspects such as "Health benefit" and "Societal benefit" missing no concepts (0%) to "Technical robustness" missing a maximum of 38% (5/13). The most relevant "82304 - 2 only" subaspect was "Ethics," which is not assessed by FTSS. For other subaspects ("Societal benefit," "Security," and "Interoperability"), 82304 - 2 covered at least 50% of the subconcepts.

For more detail, Table 3 provides the match scores obtained for each concept assessed in the frameworks. Note that this comparison has been performed with the original 82304 - 2 rather than with the more detailed Label2Enable 82304 - 2 handbook for certified app assessment organizations, which provides standardized subquestions to assess the 82304 - 2 quality requirements and thus may add more subconcepts.

XSL•FO
RenderX

Figure 6. Percentage match of subconcepts between the CEN ISO/TS 82304‐2 (82304‐2) and TIC Salut Social Foundation (FTSS) assessment framework clustered according to the 82304‐2 subaspects. Colored boxes framing the subaspects indicate the 82304-2 aspect to which they originally belong to: “Healthy and safe” (green), “Easy to use” (gray), “Secure data” (blue), and “Robust build” (yellow).



XSL•FO
RenderX

Table . Evaluation of the degree to which each concept and subconcept identified in the CEN ISO/TS 82304 - 2 (82304 - 2) and TIC Salut Social Foundation (FTSS) assessment framework was matched in the other framework. In total, 3 outcomes were possible: 0=no match, 0.5=partial match, and 1=full match. From the 2 frameworks combined, in total 78 concepts and 97 subconcepts were extracted.[a]

Concepts extracted within each subaspect	Subconcepts extracted (n)	Concepts match (sum of subconcepts match scores)		
		Both	82304 - 2 only	FTSS only
Health requirements				
Intended users[b]	1	1	0	0
Age restrictions	1	1	0	0
Health issues[b]	1	0	1	0
Intended use[b,c]	1	1	0	0
Medical device[c]	1	1	0	0
Health professionals' collaboration	2	1.5	0	0.5[d]
Literature used	2	1.5	0.5	0
Health risks				
Risks analysis	1	0	1	0
Risks control	2	0	2	0
Professional approval before use[b]	1	0.5	0.5	0
Risks communication	1	1	0	0
Contraindications and limitations	2	1.5	0	0.5[d]
Safety concerns and incidents	1	1	0	0
Ethics				
Ethical challenges	1	0	1	0
Approval from ethics board	1	0	1	0
Health benefit				
Health benefit[b]	1	1	0	0
Health interventions	1	0	1	0
Financial cost for users	1	0	1	0
Need of health professional support	1	0.5	0.5	0
Evidence for health benefit	2	1.5	0.5	0
Maintenance of health information	1	0.5	0.5	0
Sources for health information	1	1	0	0
Funding	1	1	0	0
Advertising	1	0.5	0.5	0
Societal benefit				
Societal benefit	1	0.5	0.5	0
Evidence for societal benefit	1	0	1	0
Accessibility				
Perceptibility	2	2	0	0

XSL•FO
RenderX

Concepts extracted within each subaspect	Subconcepts extracted (n)	Concepts match (sum of subconcepts match scores)		
		Both	82304 - 2 only	FTSS only
Operativity	2	2	0	0
Understandability	2	2	0	0
Robustness	1	0.5	0	0.5[d]
Age appropriateness	1	1	0	0
Usability				
Design taking users into account	2	1.5	0	0.5[d]
Users involved in design	1	0	1	0
User-centered evaluation	1	0	1	0
Measures to avoid misuse	3	2	0	1[e]
App information communication	3	2	0.5	0.5[d]
Instructions for use	2	1.5	0	0.5[d]
User support when experiencing problems	1	1	0	0
Mechanisms for network and systems problems avoidance	1	0.5	0	0.5[d]
Usability research for improvements	1	0	1	0
Privacy				
Processing of personal data[b,c]	1	0.5	0.5	0
Processing of personal health data[b]	1	0	1	0
Data minimization	1	1	0	0
Data retention policy	2	1.5	0	0.5[f]
Privacy statement for users	1	1	0	0
Data processing of minors and severely disabled people	1	0.5	0	0.5[d]
Secure contracts with data processors	1	0	1	0
User approval for data use and opt-in default	2	1.5	0	0.5[d]
Data protection officer	1	0	1	0
Security incident response procedures	1	0	1	0
Security				
Security management implementation	1	0	1	0
Information security risk assessment	1	0	1	0
Secure by design process	1	0	1	0
Security of third-party software	1	0	1	0

XSL•FO
RenderX

Concepts extracted within each subaspect	Subconcepts extracted (n)	Concepts match (sum of subconcepts match scores)		
		Both	82304 - 2 only	FTSS only
Security of source code	1	0.5	0.5	0
Organizational measures for legitimate processing of data	1	0	1	0
Security with user authentication	2	1.5	0	0.5[d]
Encryption of data stored	1	1	0	0
Management of security vulnerabilities	1	0	1	0
Regular testing of app security	1	0	1	0
Security policy for users	1	0	1	0
Technical robustness				
Product requirements	1	0	1	0
Software development with standards and methods	1	0	1	0
Secure coding standard	1	0.5	0.5	0
Configuration management plan	1	0	1	0
Dealing with increase in demand	1	1	0	0
Validation and verification plan	1	0	1	0
Release and deployment process	1	0.5	0.5	0
Maintenance process	1	0.5	0.5	0
General functioning	2	0	0	2[d]
Functioning with device events	1	0	0	1[e]
Use of device resources	1	0	0	1[e]
Use of network resources	1	0	0	1[e]
Interoperability				
Functioning with wearables	2	1	0	1[g]
Guides and specifications for all APIs[h]	1	0.5	0.5	0
Guides and specifications for terminologies	1	0	1	0
Validation of data transferred via APIs	1	0.5	0.5	0

XSL•FO
RenderX

Concepts extracted within each subaspect	Subconcepts extracted (n)	Concepts match (sum of subconcepts match scores)		
		Both	82304 - 2 only	FTSS only
Export of app data	1	0	1	0

[a]Concepts and subconcepts of each requirement in each assessment framework have been extracted and organized according to 82304 - 2 subaspects but include both 82304 - 2 and FTSS requirements.

[b]This subconcept is not score-impacting in 82304 - 2.

[c]This subconcept is not score-impacting in the FTSS framework.

[d]This "FTSS only" subconcept is now included in the subquestions of the Label2Enable 82304 - 2 handbook.

[e]This "FTSS only" subconcept is now partially included in the subquestions of the Label2Enable 82304 - 2 handbook. This means that the subconcept is included, but with a lower level of detail. For example, the Label2Enable 82304 - 2 handbook checks whether measures to avoid misuse are in place, but not all measures that FTSS assesses have been included. In most cases, the Label2Enable consortium considered this level of detail not relevant.

[f]This "FTSS only" subconcept is context-specific for Catalonia or Spain. It is not for 82304 - 2 uptake.

[g]This "FTSS only" subconcept is scope-expanding, that is, it goes beyond the definition of a health app assessment framework, at least for 82304 - 2. Currently, it is not for 82304 - 2 uptake.

[h]API: application programming interface.

Both for the "82304 - 2 only" and the "FTSS only" subconcepts, it is important to consider that these subconcepts may not be relevant or may be of lesser importance in the other assessment framework. Of the initial 13% (12.5/97) "FTSS only" subconcepts, 7% (7/97) were in the meantime included in the subquestions of the Label2Enable 82304 - 2 handbook as considered relevant, 1% (1/97) was scope-expanding and thus was not for 82304 - 2 uptake, and 1% (0.5/97) was related to specific Catalan and Spanish regulations, also not for 82304 - 2 uptake. The remaining 4% (4/97) "FTSS only" subconcepts were specific to app usage–based assessment of the technical robustness of the app, labeled as rigor-enhancing subconcepts, and not included in the Label2Enable 82304 - 2 handbook, as the evidence-based assessment was considered sufficient.

Integration of CEN ISO/TS 82304-2

Advantages of Each Framework

In general, both assessment frameworks were found to be quite similar. The international applicability, foundation in standardization, scientific evidence, concepts covered, and maintenance of 82304 - 2 and Label2Enable make the framework distinctive in rigor and capable of enabling harmonization and preventing a duplication of efforts. In addition, the 82304 - 2 health app quality label and report, the more detailed version of the 82304 - 2 health app quality label that was cocreated with health care professionals to support decision-making on a health app, are expected to be useful for potential users, health professionals, and health systems to increase the willingness to use health apps. It is remarkable of FTSS; nevertheless, that the assessment of their framework is supported by societies of health professionals collaborating in the Functional Experts Committee, and that they have more years of experience in their assessment. On the other hand, Label2Enable tested with medical societies the usefulness of the health app quality report in providing guidance on health apps. An article with the European Society of Cardiology established that 82304 - 2 covered all specified requirements for apps in the 4 intended uses and 3 health issues investigated [25].

Regarding the assessment itself, combining evidence- and app usage–based assessment, as defined in the Label2Enable 82304 - 2 handbook, was considered by the researchers to provide the most advantages in rigor. The objectivity and consistency of the assessment are higher in Label2Enable due to the specific terminologies (eg, 'intended use,' 'intended user'), and the clarity for each quality requirement of what is needed to pass, although it is valuable of FTSS that their framework calculates the risk level of the app to determine the number of mandatory requirements. CEN ISO/TS 82304 - 2 has a similar approach, with at most 67 score-impacting requirements, depending on, among other intended uses, whether an app includes health information, processes personal data, and is interoperable. On a global level, 4 requirements are mandatory to qualify for an 82304 - 2 label; on an EU level; more requirements are mandatory based on an alignment with EU-level legislation and values. Additionally, FTSS has a self-evaluation (readiness) test for manufacturers, which was recommended to be implemented similarly for 82304 - 2.

Considering the quality requirements themselves, the most important dimension to decide on integration, 82304 - 2 covered 87% (84.5/97) of the subconcepts of both frameworks (47/97 (48%) "Both" plus 37.5/97 (39%) "82304 - 2 only"). By having added 7 "FTSS only" subconcepts to the subquestions, the current Label2Enable 82304 - 2 handbook contains 100% of the EU-relevant requirements of 82304 - 2 and FTSS assessment frameworks. The remaining 5.5 FTSS subconcepts (1 scope-expanding, 0.5 context-specific, and 4 rigor-enhancing) are specific to the FTSS assessment framework and can be assessed separately on top in Catalonia. Overall, 82304 - 2 and the Label2Enable certification scheme provide many advantages; in short, a broader, more objective, pan-European, evidence-based assessment framework that was considered to provide an enhanced value.

Weighing of Integration Scenarios

Given the advantages that 82304 - 2 proved to provide, scenario (4) was disregarded, making the remaining viable integration scenarios (1) to adopt the Label2Enable 82304 - 2 certification scheme and related handbook, adding Catalonia-specific additional requirements on top, (2) to adopt the 82304 - 2

XSL•FO
RenderX

requirements into the FTSS framework (using the phrasing of 82304 - 2), or (3) to adapt 82304 - 2 requirements according to our needs.

The Label2Enable consortium proposes an EU-level assessment via the Label2Enable certification scheme and 82304 - 2 handbook, with each country (or region) then assessing only the requirements specific to the country (or region), being context-specific, scope-expanding, and rigor-enhancing requirements. For Catalonia and manufacturers involved, cross-country recognition would simplify and accelerate the certification process. Thus, FTSS sees value in directly using the Label2Enable certification scheme (scenario 1) once it is operational and 82304 - 2 health app quality labels and reports are issued.

However, as the Label2Enable certification scheme is not yet operational, and the Government of Catalonia needs to carry on with app assessments and is interested in already improving its framework, we decided to start by implementing 82304 - 2 ourselves and adapting some requirements according to our needs (mix of scenarios 2 and 3). This allowed the Label2Enable certification scheme to follow its process of implementation in Europe while updating the FTSS framework.

Integration and Validation

All 81 82304 - 2 requirements were incorporated in the new FTSS framework, using different integration mechanisms based on the match scores. For unmatched 82304 - 2 concepts and subconcepts, the 82304 - 2 requirements were adopted as is. For full matches, FTSS-related subtleties were added to the 82304 - 2 requirements. Of the partially matched FTSS subconcepts, 8 could be added to existing 82304 - 2 requirements. Next, the remaining 3 partially matched FTSS subconcepts, the 7 unmatched "FTSS only" subconcepts, and 1 new concept originating from a consultation of the newest regional guidelines [26] and Catalan subject matter experts, as per the methodology outlined in Figure 2, were phrased as a question to create 11 new quality requirements for the new FTSS framework. The entire set of requirements was translated to Catalan using plain language.

Subsequently, 5 of these 11 new requirements were included in the Label2Enable 82304 - 2 handbook. The 6 remaining FTSS-specific requirements included 1 scope-expanding requirement (wearables), 4 rigor-enhancing requirements (additional app usage–based assessment), and 1 context-specific requirement (ensuring that apps developed within the Ministry of Health of Catalonia contain the required visual identity).

FTSS has already developed a new web-based form for submitting the necessary information and evidence for assessment. Assessment results will be tested in the following months, and the intent is to present the results with a quality label very similar to that of 82304 - 2. Once the Label2Enable 82304 - 2 handbook is operational and 82304 - 2 quality labels and reports are issued, scenario 1 will become feasible. Then, the new Catalan assessment framework would consist of the 82304 - 2 label and report with 81 requirements and its added Label2Enable 82304 - 2 handbook subquestions, plus 6 additional requirements assessed by FTSS. That would mean the 82304 - 2 label covers 93% (81/87) of the Catalan assessment, with FTSS having to assess only 7% (6/87).

Discussion

Principal Findings

We successfully achieved our objective of generating a systematic and comprehensive comparative analysis of the FTSS assessment framework for health apps with the 82304 - 2 framework. We identified 78 concepts and 97 subconcepts; 37.5 of these were new or partially new subconcepts from 82304 - 2, all relevant to Catalonia, and 12.5 were FTSS-specific. All 81 82304 - 2 requirements were incorporated in the new FTSS framework. FTSS subtleties and partially matched subconcepts were added to the existing 82304 - 2 requirements where suitable. The remaining FTSS-specific concepts and subconcepts were rephrased as questions to create 11 additional quality requirements. Of these 11 requirements, 5 were integrated in the Label2Enable 82304 - 2 handbook. Once the Label2Enable 82304 - 2 handbook is operational and can be adopted, the remaining 6 requirements will be assessed on top of the existing 81 82304 - 2 requirements. Of these 6, 1 requirement is scope-expanding, 4 are rigor-enhancing, and 1 is context-specific.

The methodology used enabled us to adequately consider and achieve proper integration of 82304 - 2 in Catalonia. We believe that this comparative methodology can serve as a useful model for health authorities in other regions and countries considering or seeking to integrate 82304 - 2 into their own assessment framework and adopt the Label2Enable 82304 - 2 handbook. In such cases, we expect that the comparison of the quality requirements is key. In our case, extracting the concepts and subconcepts from the FTSS and 82304 - 2 quality requirements highly increased the reliability of the percentage matches obtained. The percentages obtained for each subaspect, the relevance of all "82304 - 2 only" concepts to Catalonia, as well as the low percentage of Catalonia-specific requirements to be assessed on top, were particularly useful for decision-making and necessary for actual 82304 - 2 integration.

The Results From a Theoretical Perspective

Rogers' diffusion of innovation theory [27] was foundational in the Label2Enable project and is of value given our broader aim to provide this methodology for health authorities elsewhere to consider integrating 82304 - 2 or other evaluation frameworks. Rogers' theory acknowledges 5 attributes that influence rate of adoption of an innovation. In this case, 82304 - 2 can be considered an innovation. The first and main attribute, according to Rogers, and in our experience, is relative advantage, the degree to which 82304 - 2 is perceived as better than the current FTSS framework. Overall, 82304 - 2 has been observed to provide a positive relative advantage in comparison with the current FTSS framework. First, 82304 - 2 entails a more structured, better maintained, standards- and evidence-based framework, with additional value-adding concepts and subconcepts to assess the quality of a health app compared to the FTSS framework. Second, 82304 - 2 is an international standardization effort, and the Label2Enable 82304 - 2 handbook adopts a pan-European perspective, making

XSL•FO
RenderX

it a more robust and scalable solution for app assessment across borders. This broader scope allows for the harmonization of assessment requirements, which is especially beneficial given the current lack of cross-country standardization and efficiency and scalability challenges in health app assessments. This advantage cannot be reached with a regional framework. Thirdly, the Label2Enable 82304 - 2 handbook's evidence-based assessment approach on top of app usage–based assessment with clearly defined terminologies and definitions of what is needed to pass was evaluated favorably by FTSS. For this reason, the Label2Enable 82304 - 2 handbook will be adopted, and all 82304 - 2 requirements were incorporated in the new FTSS framework. FTSS subconcepts not covered by 82304 - 2 were maintained as additional requirements, as suggested by Frey et al [28] and already operationalized by the Australian eHealth Agency's framework [29].

Rogers' second attribute is compatibility, the degree to which integration of 82304 - 2 is perceived as consistent with the existing known values, experiences, and needs of Catalonia. Beyond our decisions detailed in the previous paragraphs, we propose multistakeholder feedback to assess compatibility and specify expected benefits, challenges, and the practical implications of integration from a multistakeholder perspective, enhancing existing knowledge on the compatibility of 82304 - 2 from a health authority perspective [30]. In June 2023, FTSS invited the members of the Functional Experts Committee, formed by different societies of health professionals, to discuss the value of 82304 - 2. Careers of its 8 members can be found on the FTSS website [23]. They concluded that this certification could become greatly useful if it became mandatory for all health apps in the European market, and especially if it appeared in the apps marketplaces. This way, the quality of all published health apps would be ensured, and patients, professionals, and health care systems would use and recommend health apps with more confidence and reliability. Additionally, in a later meeting, they valued the Label2Enable health app quality report as a powerful tool to obtain relevant quality information about health apps and thus allow them to prescribe these apps. In the Label2Enable project, the manufacturer perspective was obtained through a discrete choice experiment with 41 manufacturers and a pilot study with 24 manufacturers. The first study showed that manufacturer willingness to pay for app assessment was primarily associated with the potential for integration of the health app into clinical guidelines and reimbursement or procurement [31]. In the second study, after having participated in a pilot 82304 - 2 assessment, 87.5% (21/24) of the manufacturers responded positively to willingness to "do it again" [32]. The Catalan health care system currently considers making this new assessment free of charge for apps of interest.

Rogers' third attribute is complexity, the degree to which 82304 - 2 integration is perceived as relatively difficult to understand and use. Once the Label2Enable 82304 - 2 handbook is operational and 82304 - 2 quality labels and reports are issued, the new Catalan assessment framework would consist of the 82304 - 2 quality label and report plus only 6 additional requirements assessed by FTSS (7%, 6/87). A further expansion of the 82304 - 2 scope, as currently advocated for the upcoming 82304 - 2 revision, would likely eliminate the need for assessing the scope-expanding requirement by FTSS, further reducing the Catalonia-specific workload to 6% (5/87). Uptake of 82304 - 2 thus avoids considerable duplication of assessment efforts with similar requirements. Currently, however, as the Label2Enable 82304 - 2 handbook was not yet available, we used an extensive analysis of 82304 - 2 to consider and achieve proper integration. To operationalize the new framework, training of the assessors will be required. Although it is expected that this integration may carry more workload for assessors, at least in the beginning, the time investment was considered acceptable in view of the enhanced value of the new framework and quality of certified apps that combined promote health app use. To end with, it will be necessary to properly decide how the quality labels are used; for instance, which scores are considered sufficient for apps to be used within the Catalan health care system.

The final 2 of Rogers' attributes, trialability, the extent to which 82304 - 2 can be tested before its integration, and observability, the level to which 82304 - 2 integration provides tangible results, were considered of lesser importance at this stage. Taking into account that Label2Enable has already performed tests with 24 apps with the handbook [32], trials were considered mostly covered. However, also considering the handbook evolved after testing with 24 apps, Catalonia will test the resulting assessment framework with a minimum of 6 apps (hospital and private apps) by mid-2025. This test will include an analysis of the time investment for both manufacturers and assessors and the elapsed time between the assessment request and the availability of the assessment results. Finally, regarding observability, we especially perceive great potential of 82304 - 2 in increasing the trust in and willingness to use health apps among all stakeholders, supported by recent [25,30] and upcoming 82304 - 2-related publications. Once certificates are issued, it will be possible to observe the actual increase in the degree of trust and—gradually—use. For now, with the new 82304 - 2 integrated assessment framework in Catalonia in place, entities within the Ministry of Health of Catalonia have gained more trust in the assessment framework. They are offering guidelines on health app development [26,33] to all health care providers (primary care and specialized care) within the Catalan health care system, developed by FTSS in parallel with the new assessment framework. They are also using the new assessment framework to ensure the quality of all apps integrated within the electronic health record patient portal (La Meva Salut).

Future Developments

This study was one of many studies within Label2Enable; some have already been published [25,30-32,34], and others are expected to be published later in 2025. Most of these studies aimed at assessing the needs of a specific stakeholder group, the compatibility of 82304 - 2 with these needs, and how to further increase related outcomes. In total, 4 multistakeholder backcasting workshops were used to arrive at a multistakeholder road map toward the preferred future of digital health and care and the role of digital health quality labeling in it [35]. Part of this road map is a demonstrator phase in which the first 50 - 100 health app quality labels and related, more detailed reports can

XSL•FO
RenderX

be issued. In this effort, the inter-rater reliability, efficiency, clarity, and scalability of the Label2Enable 82304‐2 handbook for certified app assessment organizations and the multistakeholder usefulness of the 82304‐2 health app quality report, both cocreated in the Label2Enable project, can be tested at a larger scale and further fine-tuned. Multistakeholder effectiveness is expected to further support decision-making by authorities and other stakeholders on adopting 82304‐2.

Practical Recommendations

Following findings from this study, the Label2Enable project, and the WHO [3,35,36], we made recommendations for other authorities and stakeholders who seek to capture the value of digital health products (Textbox 1).

Textbox 1. Recommendations to other health authorities and stakeholders who seek to capture the value of digital health products.

1. Build your digital health policy based on international policies and applicable regulations, and an understanding of how digital health products can support the aims and address the challenges of your health care system or context.
2. Assign and equip an entity for regulation, oversight, and implementation of health apps or the larger scope of digital health products. Avoid, if possible, scattering responsibilities across authorities.
3. Determine your common assessment framework, assessment organization, and assessment process instead of starting from scratch. Identify with local stakeholders and experts the potential need for additional context-specific, scope-expanding, and rigor-enhancing requirements. Consider using 82304‐2 and the Label2Enable 82304‐2 handbook, and, if you have a framework in place, consider following the methodologies used in this study to evaluate and integrate 82304‐2. If more semantic rigor is preferred for analyzing concept matches, semantic knowledge tools can be considered. Consider also comparing other traits of the frameworks if applicable.
4. Engage with stakeholders, ranging from related authorities, health care providers, health professionals, and patients to digital health manufacturers, to assess their needs in integrating health apps. For instance, manufacturers' needs in meeting the assessment requirements; related authorities' needs in meeting health care system goals; health care providers', professionals', and patients' needs in selecting, integrating, and using apps and the related data; and enabling, capturing, and measuring the value of apps and their data for the individuals, health care system, and society at large.
5. Facilitate the use of quality health apps with transformation and support. For example, by integrating them into health care pathways and developing guidelines. Measure progress in integrating health apps, multistakeholder value added, and what is required to enhance that value.

Conclusions

The methodology presented has allowed a systematic, comprehensive comparative analysis of Catalonia's FTSS health app assessment framework and 82304‐2. The results informed the FTSS decision to integrate the 82304‐2 quality requirements to achieve a more comprehensive, evidence-based, standardized approach to health app assessment, and adopt the Label2Enable 82304‐2 handbook in the future. CEN ISO/TS 82304‐2 included many subconcepts new to FTSS, which were all considered relevant for Catalonia. The Label2Enable 82304‐2 handbook was found to be value-adding in rigor and consistency, and the remaining FTSS-specific requirements were few. These findings are expected to be indicative of the potential of 82304‐2 to contribute to solving the heterogeneous and complex landscape of assessment frameworks for health apps, making robust health app assessments common. FTSS highly encourages other authorities to undertake a similar evaluation or to wait until the Label2Enable 82304‐2 handbook is operational to adopt it in full.

Acknowledgments

This study was initiated and led by the TIC Salut Social Foundation (FTSS), within the Ministry of Health of Catalonia, and later integrated and conducted as part of the Label2Enable project and funded by the European Union (Project: 101057522—Label2Enable—HORIZON-HLTH-2021-IND-07). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Publication costs and part of the development costs were covered by FTSS. Although FTSS was part of the Label2Enable consortium, the actions, findings, and conclusions presented in this study have not been instructed by Label2Enable. This study has been conducted following the willingness and needs identified by the Ministry of Health of Catalonia, based on their existing knowledge on health apps assessment.

Data Availability

Most of the data used in this study are publicly available and can be accessed or purchased online via the links provided in the references. Additional data regarding the methodology and results may be made available upon reasonable request to the corresponding author.

Authors' Contributions

Conceptualization: CPB (lead), BLC, PH, MVQ (supporting)
Formal analysis: BLC (lead), CPB (equal), PH (supporting)

XSL•FO
RenderX

Funding acquisition: CPB (lead), PH (equal)
Methodology: CPB (lead), BLC (equal)
Project administration: CPB
Supervision: CPB
Validation: CPB
Visualization: BLC
Writing – original draft: BLC
Writing – review & editing: BLC (lead), PH (equal), CPB, MVQ (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Fundamental information and characteristics of the CEN ISO/TS 82304-2 and TIC Salut Social Foundation assessment frameworks.
[DOCX File, 316 KB - mhealth_v13i1e67858_app1.docx]

Multimedia Appendix 2
Governance and maintenance of the CEN ISO/TS 82304-2 and TIC Salut Social Foundation assessment frameworks.
[DOCX File, 179 KB - mhealth_v13i1e67858_app2.docx]

Multimedia Appendix 3
Certification workflows of the CEN ISO/TS 82304-2 and TIC Salut Social Foundation assessment frameworks.
[DOCX File, 183 KB - mhealth_v13i1e67858_app3.docx]

References

1. Hoogendoorn P, Versluis A, van Kampen S, et al. What makes a quality health app-developing a global research-based health app quality assessment framework for CEN-ISO/TS 82304-2: Delphi study. JMIR Form Res 2023 Jan 23;7(1):e43905. [doi: 10.2196/43905] [Medline: 36538379]
2. ISO/TS 82304-2:2021 - Health software — Part 2: Health and wellness apps — Quality and reliability. International Organization for Standardization. URL: https://www.iso.org/standard/78182.html [accessed 2024-09-19]
3. The ongoing journey to commitment and transformation - digital health in the WHO European Region 2023. World Health Organization. URL: https://cdn.who.int/media/docs/librariesprovider2/data-and-evidence/english-ddh-260823_7amcet.pdf?sfvrsn=4c674522_2&download=true [accessed 2024-09-19]
4. Empowering the health workforce to make the most of the digital revolution. Organisation for Economic Co-operation and Development (OECD). 2021. URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/07/empowering-the-health-workforce-to-make-the-most-of-the-digital-revolution_4a9f0ad5/37ff0eaa-en.pdf [accessed 2024-09-19]
5. European Commission. Shaping Europe's digital future communication on enabling the digital transformation of health and care in the digital single market; empowering citizens and building a healthier society. EUR-Lex. 2018. URL: https://digital-strategy.ec.europa.eu/en/library/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering [accessed 2024-09-19]
6. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 5;43:117-134. [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
7. Medical devices. European Medicines Agency (EMA). URL: https://www.ema.europa.eu/en/human-regulatory-overview/medical-devices [accessed 2024-09-19]
8. MDCG 2019-11 Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR. Medical Device Coordination Group. 2019. URL: https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2019_11_guidance_qualification_classification_software_en_0.pdf [accessed 2024-09-19]
9. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR Mhealth Uhealth 2022 Oct 31;10(10):e37980. [doi: 10.2196/37980] [Medline: 36315221]
10. Singh K, Drouin K, Newmark LP, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578]
11. Essén A, Stern AD, Haase CB, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):31. [doi: 10.1038/s41746-022-00573-1] [Medline: 35304561]
12. López Seguí F, Pratdepàdua Bufill C, Rius Soler A, et al. Prescription and integration of accredited mobile apps in Catalan health and social care: protocol for the AppSalut site design. JMIR Res Protoc 2018 Dec 21;7(12):e11414. [doi: 10.2196/11414] [Medline: 30578234]

XSL•FO
RenderX

13. Lopez Segui F, Pratdepadua Bufill C, Abdon Gimenez N, Martinez Roldan J, Garcia Cuyas F. The prescription of mobile apps by primary care teams: a pilot project in Catalonia. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10701. [doi: 10.2196/10701] [Medline: 29929948]
14. About CEN ISO/TS 82304-2. Label2Enable. URL: https://label2enable.eu/about-CEN ISO-ts-82304-2 [accessed 2024-09-19]
15. About Label2Enable. Label2Enable. URL: https://label2enable.eu/about-the-project [accessed 2024-09-19]
16. Knight CG. Human–environment relationship: comparative case studies. In: International Encyclopedia of the Social & Behavioral Sciences Pergamon: Elsevier Ltd; 2001:7039-7045. [doi: 10.1016/B0-08-043076-7/04195-4]
17. Fischer D, Jenssen S, Tappeser V. Getting an empirical hold of the sustainable university: a comparative analysis of evaluation frameworks across 12 contemporary sustainability assessment tools. Assess Eval High Educ 2015 Aug 18;40(6):785-800. [doi: 10.1080/02602938.2015.1043234]
18. Generalitat de Catalunya El Govern aprova el Pla de Mobilitat “mHealth.cat”, que aproparà serveis sanitaris i de benestar a través de tecnologies mòbils. Govern. 2015. URL: https://govern.cat/salapremsa/notes-premsa/279950/govern-aprova-pla-mobilitat-healthcat-que-apropara-serveis-sanitaris-benestar-traves-tecnologies-mobils [accessed 2024-09-19]
19. Desenvolupament d’aplicacions mòbils en l’àmbit de la salut i l’atenció a la persona. Fundació TIC Salut Social. 2021. URL: https://ticsalutsocial.cat/wp-content/uploads/2021/07/DesAppsAmbitSalutIAtPersona.pdf [accessed 2024-09-19]
20. Criteris de certificació d’Apps de salut o benestar. Fundació TIC Salut Social. 2021. URL: https://ticsalutsocial.cat/wp-content/uploads/2021/09/2021_Criteris-certificacio_CAT_v3.0.pdf [accessed 2024-09-19]
21. Guia del procés d’acreditació per Apps de salut o benestar. Fundació TIC Salut Social. 2021. URL: https://ticsalutsocial.cat/wp-content/uploads/2021/07/guia-proces-acreditacio-fts.pdf [accessed 2024-09-19]
22. Digital assets directory. Fundació TIC Salut Social. URL: https://ticsalutsocial.cat/en/que-fem/digital-assets-for-citizens/assets/ [accessed 2024-09-19]
23. Expert committee. Fundació TIC Salut Social. URL: https://ticsalutsocial.cat/en/que-fem/digital-assets-for-citizens/comite-dexperts/ [accessed 2024-09-19]
24. Self-assessment test. Fundació TIC Salut Social. URL: https://ticsalutsocial.cat/en/que-fem/digital-assets-for-citizens/test-dautoavaluacio/ [accessed 2024-09-19]
25. Caiani EG, Kemps H, Hoogendoorn P, Asteggiano R, et al. Standardized assessment of evidence supporting the adoption of mobile health solutions: a clinical consensus statement of the ESC Regulatory Affairs Committee. Eur Heart J 2004:1-15. [doi: 10.1093/EHJDH/ZTAE042]
26. Guia De Bones Pràctiques Per Desenvolupar Actius Digitals per a la Ciutadania. Fundació TIC Salut Socia. 2024. URL: https://ticsalutsocial.cat/wp-content/uploads/2024/04/Guia-de-Bones-Practiques-per-desenvolupar-Actius-Digitals-per-la-Ciutadania.pdf [accessed 2024-09-19]
27. Rogers EM. Diffusion of Innovations, 5th edition: Free Press; 2003.
28. Frey AL, Phillips B, Baines R, et al. Domain coverage and criteria overlap across digital health technology quality assessments: a systematic review. OSF. Preprint posted online on May 11, 2024. [doi: 10.31219/OSF.IO/QG9VD]
29. Assessment framework for mHealth apps. Australian Digital Health Agency. 2022. URL: https://www.digitalhealth.gov.au/about-us/strategies-and-plans/assessment-framework-for-mhealth-apps [accessed 2024-09-19]
30. Hoogendoorn P, Shokralla M, Wilemsen R, Guldemond N, Villalobos-Quesada M. Reducing uncertainty about the adoption of CEN ISO/TS 82304-2: a qualitative interview study on the compatibility of a health app assessment framework with catalan and italian health authorities’ needs. JMIR Preprints. Preprint posted online on Oct 22, 2024. [doi: 10.2196/preprints.67855]
31. Frey AL, Leigh S, Toro C, et al. Preferences and willingness to pay for health app assessments among healthcare stakeholders: a discrete choice experiment. JMIR Preprints. Preprint posted online on Feb 17, 2024. [doi: 10.2196/preprints.57474]
32. Frey AL, Matei D, Phillips B, et al. Testing and iterative improvement of the CEN ISO/TS 82304-2 health app quality assessment: pilot interrater reliability study. JMIR Form Res 2025 Mar 10;9:e64565. [doi: 10.2196/64565] [Medline: 40063936]
33. Relació amb ciutadania i professionals. Departament de Salut. URL: https://salutweb.gencat.cat/ca/ambits-actuacio/linies/tic/solucions-siscat/model-adhesio/relacio-ciutadania-professionals/ [accessed 2025-01-23]
34. Biliunaite I, van Gestel L, Hoogendoorn P, Adriaanse M. Value of a quality label and European healthcare professionals’ willingness to recommend health apps: an experimental vignette study. J Health Psychol 2024 Aug 2:13591053241258205. [doi: 10.1177/13591053241258205] [Medline: 39096027]
35. The preferred future of health & care in 5-10 years and the role of digital health quality labelling. Label2Enable. URL: https://label2enable.eu/assets/downloads/label2enable_preferred-future-health-and-care-visual-1716569529.pdf [accessed 2025-01-23]
36. Recommendations for the reimbursement of health apps. Label2Enable. URL: https://label2enable.eu/assets/downloads/la2en-infographic-recommendations-v5-1716569726.pdf [accessed 2025-01-23]

XSL•FO
RenderX

Abbreviations

82304-2: CEN ISO/TS 82304-2
CE: Conformité Européenne
CEN: European Committee for Standardization
EU: European Union
FTSS: TIC Salut Social Foundation (Catalan abbreviation)
ISO: International Organization for Standardization
MDR: Medical Device Regulation
TS: Technical Specification
WHO: World Health Organization

Edited by L Buis; submitted 22.10.24; peer-reviewed by CM McCay, I Grau-Corral, MDSP López; revised version received 30.01.25; accepted 02.04.25; published 04.06.25.

Please cite as:
Llebot Casajuana B, Hoogendoorn P, Villalobos-Quesada M, Pratdepàdua Bufill C
Integrating CEN ISO/TS 82304-2 in the Catalan Health App Assessment Framework: Comparative Case Study
JMIR Mhealth Uhealth 2025;13:e67858
URL: https://mhealth.jmir.org/2025/1/e67858
doi:10.2196/67858

© Berta Llebot Casajuana, Petra Hoogendoorn, Maria Villalobos-Quesada, Carme Pratdepàdua Bufill. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Leveraging Social Media and Crowdsourcing to Recruit and Retain Military Veterans With Posttraumatic Stress Disorder or Experience of Harmful Gambling for mHealth Interventions: Descriptive Study

Conor Heath[1], MSc; Jess M Williams[1], PhD; Daniel Leightley[2], PhD; Dominic Murphy[3,4], PhD; Simon Dymond[1,5], PhD

[1]Centre for Military Gambling Research, Swansea University, Singleton Park, Swansea, United Kingdom

[2]Department of Population Health, Faculty of Life Science and Medicine, King's College London, London, United Kingdom

[3]King's Centre for Military Health Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[4]Combat Stress Centre for Applied Military Health Research, Leatherhead, United Kingdom

[5]Department of Psychology, Reykjavík University, Reykjavík, Iceland

Corresponding Author:
Simon Dymond, PhD
Centre for Military Gambling Research, Swansea University, Singleton Park, Swansea, United Kingdom

Abstract

Background: Military veterans may be at increased risk of posttraumatic stress disorder (PTSD) compared to the general population. PTSD is often comorbid with harmful and problematic patterns of gambling. Behavioral therapies such as acceptance and commitment therapy have shown promise in treating these co-occurring disorders, especially if combined with mobile health (mHealth) interventions to circumvent known help-seeking barriers faced by veterans. However, to date, recruitment for mHealth interventions has been challenging and may impact intervention feasibility.

Objective: In this paper, our objectives were to describe the strategies used to recruit UK military veterans with PTSD or experience of harmful gambling to a pilot study of a smartphone-based digital intervention, ACT Vet.

Methods: We used several recruitment strategies, such as direct mailing, paid study advertising on social media (Facebook) and an online research platform (Prolific), study-specific website management, in-person event hosting with veterans' charities, snowball sampling, and incentives for completion.

Results: Results showed that, over 27 days, recruitment through Facebook accounted for 21 eligible veterans (n=7, 33% through unpaid advertising and n=14, 67% through paid advertising), whereas Prolific accounted for 50 veterans. Additional strategies recruited 8 eligible veterans. In total, 79 eligible military veterans were recruited for ACT Vet, with 24 (30%) completing the final steps of the study.

Conclusions: Difficulties such as low advertisement conversion rate and participant and data attrition arose throughout this study. Our findings illustrate the relative effectiveness of social media– and online platform–based initiatives in recruiting veterans with PTSD or harmful gambling. Future research should consider establishing an online presence for effective digital intervention recruitment with diverse branding to attract representative samples of veterans for mHealth research.

(JMIR Mhealth Uhealth 2025;13:e73706) doi:10.2196/73706

KEYWORDS

military veteran; gambling; posttraumatic stress disorder; PTSD; acceptance and commitment therapy; ACT; mobile health; smartphone; social media; crowdsourcing; recruitment

Introduction

Military veterans may face barriers when accessing help and support for difficulties related to mental health up to 11 years after service [1-3]. Among veterans, posttraumatic stress disorder (PTSD) is often prevalent alongside challenges related to anxiety, alcohol misuse, and depression that contribute to poorer well-being [4,5]. Moreover, veterans are also more likely to engage in and experience avoidance behaviors, including harmful gambling behaviors motivated by escape or avoidance of distress [6-9]. Recently, Dighton et al [9] highlighted that veterans may engage in gambling as a coping mechanism for distressing PTSD symptoms. They argue that increased gambling behaviors may lead to harmful gambling (gambling disorder), defined as problematic gambling behaviors that cause ongoing distress [10], and are likely to have an impact on the individual and the wider health and well-being systems that aim to provide support [11]. Therefore, there is an urgent need for

XSL•FO
RenderX

improved access to efficacious treatment or interventions that may help manage comorbid mental health challenges such as PTSD and harmful gambling.

Promising therapeutic interventions for military veterans experiencing PTSD and comorbid harmful gambling include third-wave, mindfulness approaches such as acceptance and commitment therapy (ACT) [12-15]. ACT involves 6 core principles, defined as being present in the moment, identifying important values, committing to action, the self as context, cognitive defusion, and acceptance [16]. These principles focus on building psychological flexibility, the capacity to adapt and manage distress in accordance with valued goals [16,17]. There is also growing evidence of the feasibility and acceptability of mobile health (mHealth) interventions for PTSD, depression, alcohol misuse, and mental health [18-20]. Such apps embed forms of therapeutic support (eg, ACT) using smartphones, which allows for quick access to independent support supplemented by regular notifications, which may help with therapy engagement [21]. A meta-analysis by Linardon [22] revealed that smartphone-based apps could be effective in building acceptance and mindfulness skills (key elements of ACT). In addition, Farzandipour et al [18] suggested that apps may reduce self-reported PTSD symptoms. However, while Hawker et al [23] suggested that gambling-related apps could be accepted as an intervention, there is a lack of evidence-based treatment testing for many apps that are currently available [24,25]. Thus, evidence-based ACT mHealth apps may provide timely and easy-to-access help for military veterans with comorbid gambling-related harms and mental health difficulties.

When designing and evaluating interventions, feasibility trials, including those involving treatment for gambling in military veterans and nonveterans using mental health mHealth apps, tend to experience recruitment challenges and show high rates of attrition [26-28]. Combined recruitment strategies such as advertising on social media, emailing, printing flyers, offering incentives, and collaborating with relevant organizations are recommended when recruiting and retaining participants in evaluations of mHealth interventions [28-30]. In-person strategies should not be excluded where possible. For example, in-person strategies such as attending local events could provide participants with needed human interaction [28], acting as a guide to the mHealth intervention [29,30] and mitigating any concerns. However, advertising using social media platforms such as Facebook may be particularly useful in reaching individuals considered “hard to reach,” such as veterans [30-32]. Previous studies have shown that, while high rates of attrition still occur, recruitment for mHealth interventions using paid social media advertising may be more cost-effective and can target “hard-to-reach” populations effectively [32-34]. Thus, social media could help counter recruitment challenges faced during feasibility trials.

In addition, Pickering and Blaszczynski [35] have raised concerns regarding the quality of data and increased risk of bias from gambling research using online crowdsourcing platforms such as Amazon Mechanical Turk [36], Qualtrics [37], and Prolific [38]. However, Russell et al [39] argue that they can be effective in recruiting specific hard-to-reach groups given that stringent data quality checks are conducted to reduce potential biases [39]. These checks could include eligibility screening and validation of data quality such as military service numbers, which are identifiers unique to military populations. Prolific can provide high-quality data while being a cost-effective recruitment strategy [40,41]. Therefore, the combination of online crowdsourcing, social media advertising, and traditional recruitment (face-to-face meetings, emails, and flyers) may be effective for recruiting hard-to-reach and vulnerable populations such as military veterans in the United Kingdom with comorbid diagnoses of PTSD and harmful gambling (gambling disorder). However, to date, little is known about the relative effectiveness of these recruitment strategies.

In this study, we sought to compare these recruitment and retention methods for a pilot study designed to evaluate a 10-week app-based ACT intervention for UK military veterans experiencing PTSD and harmful gambling. Veterans were screened for PTSD severity using the PTSD Checklist for the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) [42], with scores of 20 or above conferring study eligibility. Veterans experiencing harmful gambling were screened using the Problem Gambling Severity Index [43], with scores of 1 or above conferring eligibility, indicating some form of harmful gambling. Drawing on the work of Williamson et al [32] conducted on UK military veterans and alcohol use, the aim of this paper is to define and reflect on the recruitment strategies and make recommendations for future studies recruiting and retaining military veterans with co-occurring PTSD and harmful gambling.

Methods

Ethical Considerations

The recruitment strategies used in this study formed part of a larger study approved by the Swansea University School of Psychology Research Ethics Committee (1-2023-6798-6614). All participants provided written informed consent to take part. The data obtained were fully anonymized by the research team, and all participants were compensated with either Prolific funds or online shopping vouchers on completion of the larger study.

ACT Vet

We developed the ACT Vet app using a co-design approach with veterans with lived experience of harmful gambling or PTSD and then adopted the Ionic Capacitor framework to achieve a research-viable product standard. ACT Vet is a self-guided app addressing the core components of ACT through seven modules: (1) an introduction that provides an overview of ACT Vet and ACT as a therapeutic practice, (2) Exploring thoughts around coping (cognitive defusion), (3) Control as the problem (acceptance), (4) Contact with the present (present moment awareness), (5) The self as the observer (self as context), (6) Living by your values (values), and (7) Committing to action (committed action). The modules took approximately 15 minutes to complete, with a mandatory break period of 5 days taking place between completed and new modules. The ACT-specific module content took approximately 4 weeks to complete, and participants were encouraged to use the app for the remainder of the 10-week study.

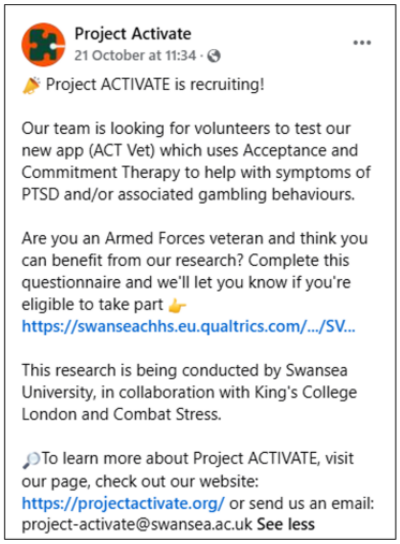
XSL•FO
RenderX

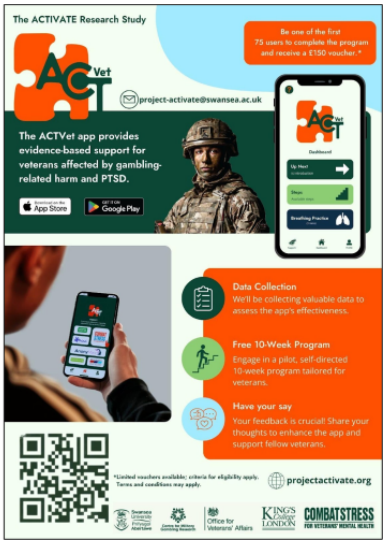
Online Recruitment

The primary recruitment and retention strategies included online engagement, providing study details and links across social media sites: Facebook, X (formerly known as Twitter), and Instagram. Individual project pages across each site (Facebook, X, and Instagram) were created (in May 2024) with the purpose of information sharing using blog-type posts and project recruitment (Figure 1, left panel). Paid Facebook advertisements were used as they have been considered a useful tool in similar veteran-specific app-based recruitment efforts [32]. These advertisements were run for a total of 27 days. The textual keywords that the target audience needed to have on their profiles to be included are as follows: Armed Forces Day (United Kingdom), veterans' benefits (military) or army men, school or university, British Army and employers, Royal Air Force Regiment, Royal Marines, veteran, His Majesty's Armed Forces, and Royal Navy or British Army. Figure 1, middle panel shows the imagery and copy used for the advertisement.

Additionally, Prolific, which is an online recruitment platform where researchers advertise studies among a large pool of vetted participants, was used to screen potential participants who were UK military veterans aged ≥18 years. A dedicated project-specific website was also launched that housed information and links related to participating in the study.

Figure 1. Screenshots from the ACT Vet app: (left panel) Facebook project page, (middle panel) paid advertisement, and (right panel) flyer for ACT Vet.





Other Recruitment Strategies

Email addresses were collected from organization websites, and study details were sent. The research team carried out face-to-face meetings and visits with organizations such as the Forces in Mind Trust, National Health Service West Midlands Gambling Harms Clinic, the Royal British Legion, Combat Stress, Help for Heroes, Beacon Counselling Trust, GamFam, and GamCare. Physical and digital flyers (Figure 1, right panel) were created and circulated among organizations. Members of the research team also had personal contacts they used, including details of veterans who had previously agreed to participate in the study. Any veteran who agreed to take part was asked to encourage others who may have benefited from participating (snowball sampling).

The information regarding the recruitment source for each participant was documented and assigned a unique identifier, allowing the research team to track where participants had viewed and accessed the prescreening questionnaire (eg, Facebook, Prolific, or website). If eligible, previous military service numbers were also confirmed for the identification of spam or bots to ensure the integrity of the data. Unless provided by the recruitment platform, demographic information was collected at the baseline time point. Finally, it was advertised that, on completion of the 10-week program, each participant would be offered a high street voucher of £150 (US $189.77) for their time.

Measures and Analysis

Our primary outcome measures included advertisement duration in days, cost in GBP (and the USD equivalent), online market reach or armed forces audience criteria, online post link clicks, and online post shares and reactions. Please note that online market reach differs depending on when digital marketing campaigns are conducted because of the underlying Facebook or Meta analytics. Means and SDs were reported for the continuous variable (respondents' ages), and percentages and frequencies were reported for categorical variables.

Results

Online Recruitment

Facebook

From the dedicated Facebook page, posts were circulated in veteran and gambling-related groups. Paid advertisements were also circulated separately for a total of 27 days to reach veterans or those with connections to veterans aged 18 to 65 years in the United Kingdom. Facebook campaign key performance indicators provide the performance and cost breakdown of each advertisement campaign and are shown in Table 1.

XSL•FO
RenderX

Table . Key performance indicators (KPIs) for ACT Vet paid Facebook advertising.

KPI	Paid advertisement campaign number		
	1	2	3
Advertisement duration (days)	3	14	10
Cost, £ (US $)	5.00 (6.33)	100.00 (126.51)	177.70 (224.81)
Total reach (number of people)	1132	14,682	117,085
Number of link clicks	61	506	854
Cost per click, £ (US $)	0.08 (0.10)	0.20 (0.25)	0.21 (0.27)
Number of post shares	2	14	1
Number of post reactions	1	12	2

The first paid campaign was run as a test for 3 days and reached 1132 people (n=112, 9.89% female; n=1020, 90.1% male), of whom 340 (30.03%) were aged between 35 and 44 years, with 1114 (98.41%) viewing it on mobile devices. The second advertisement campaign reached a total of 14,682 people (n=3656, 24.9% female; n=11,026, 75.1% male), of whom 4405 (30%) were aged between 55 and 64 years, with 12,263 (83.52%) viewing it on mobile devices. The third advertisement campaign reached a total of 117,085 people (n=70,836, 60.5% female; n=46,249, 39.5% male); 35,126 (30%) were aged ≥65 years, with 67,326 (57.5%) being reached using the audience network (third-party apps such as Instagram or mobile games).

Facebook accounted for 7 participants who were eligible and consented through organic (without paid advertising) page or post interactions, with an additional 14 who were eligible and consented through paid advertising, which cost £20.19 (US $25.55) per consenting individual. Recruitment outcomes from paid Facebook advertising are shown in Table 2. In total, Facebook accounted for 21 eligible participants. Unless provided by the recruitment platform, demographic information was only collected at the baseline time point.

Table . Total recruitment outcomes across paid Facebook advertising and Prolific.

	Paid Facebook advertising	Prolific
Total cost, £ (US $)	282.70 (357.65)	157.50 (199.25)
Number of clicks or submissions	1421	71
Cost per click or submission, £ (US $)	0.49 (0.62)	0.75 (0.95)
Number of eligible consenting individuals	14	50

X (Formerly Known as Twitter)

Additionally, a dedicated X page was created and amassed 170 followers. The page consisted of 32 posts related to recruitment advertising and blog-type posts showing the research team's activities. The highest engagement was with a blog-type post with 419 views, followed by a recruitment advertisement with 328 views. No participants were directly recruited.

Instagram

An Instagram page was also created, which resulted in 16 followers. The page consisted of 4 posts designed to raise awareness of the project and recruit participants. The highest engagement was through 6 likes on a post circulating project awareness. No participants were directly recruited.

Prolific

The study also used Prolific. The built-in screening criteria of "UK military veteran" and "aged ≥18 years" were applied, yielding a total of 287 eligible participants who had been active in the previous 90 days on Prolific. The study recruitment target was set at 150 participants. Sensitive content warnings were given as the study involved topics related to mental health. Participants were paid £0.75 (US $0.95), the equivalent to £9.00 (US $11.39) per hour, for the 5 minutes it took to complete the prescreening questionnaire. The total cost (including fees and value-added tax) for the screening was £157.50 (US $199.25). From this screening, 71 submissions were made (n=63, 89% male; n=8, 11% female), with a mean age of 45 (SD 13.24) years. Of these 71 participants, 50 (70%) met the eligibility criteria and consented to be sent app codes, which cost, on average, £3.15 (US $3.99) per consented individual. Table 2 shows the total recruitment outcomes from Prolific. Recruitment ceased without meeting the target of 150 participants due to time constraints.

Website

The project-specific website (including domain hosting) was used as a platform for project dissemination and recruitment and was launched on January 8, 2024, and continued to function until the project end. Four eligible participants completed the prescreening.

Email

After collecting email addresses from organization websites, nearly 200 emails were sent. One eligible participant was recruited and consented to be sent app codes.

XSL•FO
RenderX

Face to Face

Throughout recruitment, a total of 15 veteran, charity, or clinical meetings (in person and online) related to military veterans or gambling across the United Kingdom were conducted. Contact with organizations was vital to raise awareness of the study and interact with the veteran community. One eligible participant consented to be sent app codes directly from a face-to-face meeting, whereas another participant consented to be sent app codes via word of mouth after a meeting.

Unknown

One participant accessed the prescreening questionnaire directly, and the recruitment strategy could not be identified.

Study Attrition and Sample Characteristics

Figure 2 shows the attrition from initial screening to week 10, and the completed time points are split by recruitment source. A total of 35 participants (n=31, 89% male; n=3, 9% female; and n=1, 3% undisclosed) completed in-app baseline time points, and 1 (3%) participant formally withdrew from the study. The mean age was 44.77 (SD 10.89) years, and all participants (35/35, 100%) were of White British ethnicity, with 49% (n=38) having a diagnosis of PTSD and none having a *DSM-5* diagnosis of gambling disorder. A total of 24 participants (n=21, 88% male; n=2, 8% female; and n=1, 4% undisclosed) with a mean age of 45.29 (SD 10.70) years finished the week 10 time points and had a mean PTSD Checklist for the DSM-5 score of 40.92 (SD 20.12) and mean Problem Gambling Severity Index score of 5.47 (SD 3.59), indicating moderate risk of problem gambling. According to responses to the mHealth App Usability Questionnaire, participants provided high ratings of usability (mean 6.09, SD 0.76; range 4.18-7.00), indicating that the app was liked by veterans.

XSL•FO
RenderX

Figure 2. Study attrition and recruitment sources across project time points. “Facebook advertisement” indicates paid advertising compared to unpaid Facebook recruitment.

XSL•FO
RenderX

Discussion

Principal Findings

This paper outlines recruitment and retention strategies used for ACT Vet, a smartphone digital intervention for UK military veterans with PTSD and harmful gambling. Overall, Facebook (both unpaid and paid advertising) and Prolific were the most effective strategies for recruiting eligible participants. The combined alternative recruitment methods (hub visits, meetings, emails, and website) allowed for building interdisciplinary networks among researchers, service users, and practitioners but resulted in few participants. Challenges throughout recruitment remained and included low conversion rates from social media advertising, spam or bots, ineligible participants, participant attrition, and missing questionnaire time points. However, app usability ratings were high, and the app was well received.

Social media was effective in recruiting eligible participants, aligning with previous findings [30,32-34]. In this study, although we only ran brief advertisement campaigns, Facebook outperformed traditional strategies in the recruitment of eligible veterans. However, there was a noteworthy difference in link clicks through to eligible consent using paid Facebook advertising. Most of those who showed initial interest by clicking the advertisement link (n=1421) did not proceed through to eligible consent (n=1407, 99.01%). Even so, as evidenced by these results and highlighted by Aily et al [33], a low conversion rate from Facebook paid advertising can still contribute to a cost-effective recruitment strategy.

Prolific was also effective in recruiting and retaining eligible military veterans. Most of those who completed final questionnaire time points were recruited via Prolific (n=19, 79%) and helped the study access secondary categories of groups that may be hard to reach [39]. Prolific was also the most cost-effective compared to all other strategies. This is in line with previous work that argues that Prolific can be cost-effective and reliable for high-quality data [40,41]. Hence, our study shows that the combination of Facebook and Prolific is important in recruiting for clinical research, particularly military veterans with co-occurring PTSD and harmful gambling.

Challenges

Notwithstanding the aforementioned advantages to the approaches used, we did encounter some recruitment and retention challenges. Few veterans were recruited through strategies such as face-to-face meetings, emails, or visits to organizations. Veterans can take up to 11 years to access support services and may face significant barriers and stigmatization when accessing treatment [1,3], which may have contributed to the limited interaction observed in this study. However, due to the anonymous nature of this study, it was difficult to measure whether more veterans who later accessed the online information were recruited because of our in-person visits. For this reason, it is advisable that future digital or online interventions maintain a physical presence among service users, charities, and hubs to build in-person networks and raise further awareness. Moreover, future recruitment initiatives should consider ensuring that long-term strategies and resources are dedicated to developing engaging content and other in-person activities with veterans.

In addition, the reliance on social media resulted in many spam responses and unreliable data, which may have contributed to the large attrition rate. Even among eligible participants, close to half (n=39, 49%) did not respond after requesting and being sent individualized app codes and did not go on to complete baseline measurements. These participants could not be contacted and, therefore, were considered to have withdrawn. Similar instances of attrition have been noted in UK-based military veteran digital interventions that have used social media for recruitment [19,20] and are considered a key challenge in clinical mHealth research. As recommended by Parkes et al [20], future work using social media for recruitment should also ensure that additional strategies and resources are made available to counteract any attrition from social media recruitment.

Data attrition inevitably resulted when participants did not complete required questionnaires when requested. This led to variation between program time and data points. In this study, there was an overall decrease in completion of all measures from baseline to each time point, yet the final week, week 10, had more completions than interim weeks such as weeks 4 and 6. One reason for this could be that participants stopped using ACT Vet after all steps had been completed within the app by the week 4 time point. While push notifications and email reminders were sent, this study still experienced a decrease in completed measures during those weeks. The opportunity to provide week 10 data was provided using an email link after a lengthy period despite also being the final time point, which may have resulted in an increased response rate. Thus, if the program length extends past the intervention length (ie, in study designs with extended follow-up), attrition is likely to occur without targeted initiatives to increase engagement.

A final retention challenge may have inadvertently stemmed from offering remuneration in a single large amount at the end of the study. That is, attrition may have been reduced had we adopted a progressive or escalating schedule of payments similar to those used in research on incentivized treatment attendance or adherence to interventions for related addictive behaviors [44]. This possibility warrants consideration in future mHealth intervention research.

Limitations

There were some limitations to the recruitment and retention methods we used. First, participants recruited through Prolific were self-selected, which may increase the likelihood of bias and decrease the likelihood of obtaining a representative sample [35]. Despite this, data integrity measures were implemented throughout, such as requiring service numbers as evidence of veteran status and validating data quality. Second, as found in similar recruitment efforts, the use of online media platforms may have excluded veterans without social media or with limited experience using technology [32]. However, this study implemented mitigating strategies to reduce these issues [30]. Third, the range of colors, imagery, and designs used for the project-specific branding may have resembled army branding (ie, army green and camouflage patterns). There is a possibility that this branding impacted recruitment of veterans from other

XSL•FO
RenderX

services such as the Royal Navy and the Royal Air Force, both of which have unique colors and branding. Finally, 89% (n=31) of all eligible recruited participants were male, which resulted in an underrepresentation of female veterans. Although female representation in the regular UK armed forces increased by 11.7% in 2024 [45], the veteran criteria for ACT Vet likely resulted in a sample that reflected representation in the UK armed forces, as previously noted by Williamson et al [32]. Furthermore, men are considered at high risk of harmful gambling [46], which may have contributed to the sex variance observed.

Conclusions

In conclusion, our findings showed the effectiveness of social media and online platforms compared to additional recruitment strategies in line with previous findings and recommendations for mHealth recruitment [30,33,34], particularly for military veterans [32]. Furthermore, these findings can be extended to populations with comorbid conditions, such as veterans with both PTSD and harmful gambling. In this paper, we offer some key recommendations for future recruitment efforts. First, face-to-face meetings and visits should be organized as early as possible to strengthen interdisciplinary relationships and networks, but they should not be solely relied upon for recruitment. Second, a strong online presence should be built across various platforms, and unpaid and paid strategies should be used. This will assist in project awareness and recruitment from populations that may be hard to reach, such as those with PTSD and harmful gambling. Finally, increased efforts should be made to diversify branding and advertising. For military veteran research, this should reflect all 3 UK regular armed forces services (Army, Royal Navy, and Royal Air Force), which, in turn, may also lead to an increase in participation from female veterans with comorbid disorders. It is hoped that these findings inform future recruitment initiatives in veteran and gambling research and mHealth research more broadly.

Acknowledgments

The authors are grateful to the national veterans' service charities and Dr Catherine Hitch for their input.

Funding

This study was funded by an award made from the Office for Veterans' Affairs (Veterans' Health Innovation Fund) in 2023 to SD, DL, and DM as part of an independent research initiative by the then UK government. The views, findings, and conclusions expressed herein are those of the authors and do not reflect the policies or positions of the UK government.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the sensitive nature of the data but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: CH, JMW, DL, DM, SD
Data curation: CH, JMW
Funding acquisition: SD, DL, DM
Investigation: CH, JMW
Methodology: CH, JMW
Supervision: SD
Writing – original draft: CH
Writing – review & editing: CH, JMW, DL, DM, SD

Conflicts of Interest

DL is a reservist in the UK armed forces. This work was undertaken as part of his civilian employment. All other authors declare no other conflicts of interest.

References

1. Clary KL, Pena S, Smith DC. Masculinity and stigma among emerging adult military members and veterans: implications for encouraging help-seeking. Curr Psychol 2023 Feb;42(6):4422-4438. [doi: 10.1007/s12144-021-01768-7]
2. Gribble R, Wessely S, Klein S, et al. Who is a "veteran"? Understanding definitions of the term among the British public: a research note. RUSI J 2019;164(7):10-17. [doi: 10.1080/03071847.2019.1700683]
3. Murphy D, Ashwick R, Palmer E, Busuttil W. Describing the profile of a population of UK veterans seeking support for mental health difficulties. J Ment Health 2019 Dec;28(6):654-661. [doi: 10.1080/09638237.2017.1385739] [Medline: 28982279]
4. Binks E, Cambridge S. The transition experiences of British military veterans. Polit Psychol 2018 Feb;39(1):125-142 [FREE Full text] [doi: 10.1111/pops.12399]

XSL•FO
RenderX

5. Stevelink SAM, Jones M, Hull L, et al. Mental health outcomes at the end of the British involvement in the Iraq and Afghanistan conflicts: a cohort study. Br J Psychiatry 2018 Dec;213(6):690-697. [doi: 10.1192/bjp.2018.175]
6. Etuk R, Shirk SD, Grubbs J, Kraus SW. Gambling problems in US military veterans. Curr Addict Rep 2020 Jun;7(2):210-228. [doi: 10.1007/s40429-020-00310-2]
7. Stefanovics EA, Potenza MN, Tsai J, Pietrzak RH. Prevalence and clinical characteristics of recreational and at-risk/problematic gambling in a national sample of U.S. military veterans. J Gambl Stud 2023 Sep;39(3):1077-1097. [doi: 10.1007/s10899-022-10165-z] [Medline: 36378356]
8. Roberts E, Dighton G, Fossey M, et al. Gambling problems and military- and health-related behaviour in UK armed forces veterans. Mil Behav Health 2020 Apr 2;8(2):212-221. [doi: 10.1080/21635781.2019.1644263]
9. Dighton G, Wood K, Armour C, et al. Gambling problems among United Kingdom armed forces veterans: associations with gambling motivation and posttraumatic stress disorder. Int Gambl Stud 2023 Jan 2;23(1):35-56. [doi: 10.1080/14459795.2022.2063923]
10. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Association; 2013. [doi: 10.1176/appi.books.9780890425596]
11. Wardle H, Reith G, Langham E, Rogers RD. Gambling and public health: we need policy action to prevent harm. BMJ 2019 May 8;365:l1807. [doi: 10.1136/bmj.l1807] [Medline: 31068335]
12. Donahue ML, Fruge JE, Andresen FJ, Twohig MP. Acceptance and commitment therapy (ACT) among U.S. veterans: a systematic review. J Contextual Behav Sci 2024 Apr;32:100731. [doi: 10.1016/j.jcbs.2024.100731]
13. Goldberg SB, Riordan KM, Sun S, Kearney DJ, Simpson TL. Efficacy and acceptability of mindfulness-based interventions for military veterans: a systematic review and meta-analysis. J Psychosom Res 2020 Nov;138:110232. [doi: 10.1016/j.jpsychores.2020.110232] [Medline: 32906008]
14. Hitch C, Leightley D, Murphy D, Trompeter N, Dymond S. Acceptance and commitment therapy for co-occurring gambling disorder and posttraumatic stress disorder in veterans: a narrative review. Eur J Psychotraumatol 2023;14(1):2178203. [doi: 10.1080/20008066.2023.2178203] [Medline: 37052089]
15. Romaniuk M, Saunders-Dow E, Brown K, Batterham PJ. Feasibility, acceptability, and initial outcomes of a psychological adjustment and reintegration program for transitioned military veterans. BMC Psychol 2024 Oct 29;12(1):597. [doi: 10.1186/s40359-024-02097-7] [Medline: 39468675]
16. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 2006 Jan;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
17. Cherry KM, Hoeven EV, Patterson TS, Lumley MN. Defining and measuring "psychological flexibility": a narrative scoping review of diverse flexibility and rigidity constructs and perspectives. Clin Psychol Rev 2021 Mar;84:101973. [doi: 10.1016/j.cpr.2021.101973] [Medline: 33550157]
18. Farzandipour M, Sharif R, Anvari S. Effects of mHealth applications on military personnel's physical and mental health: a systematic review. Mil Psychol 2025;37(3):199-207. [doi: 10.1080/08995605.2024.2336640] [Medline: 38551462]
19. Leightley D, Williamson C, Rona RJ, et al. Evaluating the efficacy of the drinks: ration mobile app to reduce alcohol consumption in a help-seeking military veteran population: randomized controlled trial. JMIR Mhealth Uhealth 2022 Jun 20;10(6):e38991. [doi: 10.2196/38991] [Medline: 35724966]
20. Parkes S, Croak B, Brooks SK, et al. Evaluating a smartphone app (MeT4VeT) to support the mental health of UK armed forces veterans: feasibility randomized controlled trial. JMIR Ment Health 2023 Aug 28;10(1):e46508. [doi: 10.2196/46508] [Medline: 37639295]
21. Williamson C, White K, Rona RJ, et al. Smartphone-based alcohol interventions: a systematic review on the role of notifications in changing behaviors toward alcohol. Subst Abus 2022;43(1):1231-1244. [doi: 10.1080/08897077.2022.2074595] [Medline: 35670777]
22. Linardon J. Can acceptance, mindfulness, and self-compassion be learned by smartphone apps? A systematic and meta-analytic review of randomized controlled trials. Behav Ther 2020 Jul;51(4):646-658. [doi: 10.1016/j.beth.2019.10.002] [Medline: 32586436]
23. Hawker CO, Merkouris SS, Thomas AC, Rodda SN, Cowlishaw S, Dowling NA. The general acceptability and use of smartphone app-delivered interventions for gambling in Australia. J Gambl Stud 2025 Jun;41(2):593-613. [doi: 10.1007/s10899-024-10373-9] [Medline: 39786522]
24. Brownlow L. A review of mHealth gambling apps in Australia. J Gambl Issues 2021(47). [doi: 10.4309/jgi.2021.47.1]
25. McCurdy LY, Loya JM, Hart-Derrick VR, Young GC, Kiluk BD, Potenza MN. Smartphone apps for problem gambling: a review of content and quality. Curr Addict Rep 2023 Jun;10(2):178-186. [doi: 10.1007/s40429-023-00479-2] [Medline: 38463534]
26. Goetter EM, Bui E, Ojserkis RA, Zakarian RJ, Brendel RW, Simon NM. A systematic review of dropout from psychotherapy for posttraumatic stress disorder among Iraq and Afghanistan combat veterans. J Trauma Stress 2015 Oct;28(5):401-409. [doi: 10.1002/jts.22038] [Medline: 26375387]
27. Pfund RA, Peter SC, McAfee NW, Ginley MK, Whelan JP, Meyers AW. Dropout from face-to-face, multi-session psychological treatments for problem and disordered gambling: a systematic review and meta-analysis. Psychol Addict Behav 2021 Dec;35(8):901-913. [doi: 10.1037/adb0000710] [Medline: 34881915]

XSL•FO
RenderX

28. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
29. Kidd B, McCormack JC, Newcombe D, Garner K, O'Shea A, Humphrey G. Recruitment and retention in mHealth interventions for addiction and problematic substance use: a systematic review. J Addict Res Ther 2023;14:582. [doi: 10.1101/2023.09.09.23295094]
30. Moungui HC, Nana-Djeunga HC, Anyiang CF, Cano M, Ruiz Postigo JA, Carrion C. Dissemination strategies for mHealth apps: systematic review. JMIR Mhealth Uhealth 2024 Jan 5;12:e50293. [doi: 10.2196/50293] [Medline: 38180796]
31. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using Facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015 Jun 1;4(2):e63. [doi: 10.2196/resprot.3996] [Medline: 26033209]
32. Williamson C, Rona RJ, Simms A, et al. Recruiting military veterans into alcohol misuse research: the role of social media and Facebook advertising. Telemed J E Health 2023 Jan;29(1):93-101. [doi: 10.1089/tmj.2021.0587] [Medline: 35544055]
33. Aily JB, Copson J, Voinier D, et al. Understanding recruitment yield from social media advertisements and associated costs of a telehealth randomized controlled trial: descriptive study. J Med Internet Res 2023 May 18;25:e41358. [doi: 10.2196/41358] [Medline: 37200067]
34. Wozney L, Turner K, Rose-Davis B, McGrath PJ. Facebook ads to the rescue? Recruiting a hard to reach population into an Internet-based behavioral health intervention trial. Internet Interv 2019 Sep;17:100246. [doi: 10.1016/j.invent.2019.100246] [Medline: 31080751]
35. Pickering D, Blaszczynski A. Paid online convenience samples in gambling studies: questionable data quality. Int Gambl Stud 2021 Sep 2;21(3):516-536. [doi: 10.1080/14459795.2021.1884735]
36. Amazon Mechanical Turk. Amazon Web Services. URL: https://www.mturk.com [accessed 2025-11-04]
37. Qualtrics XM platform. Qualtrics. URL: https://www.qualtrics.com [accessed 2025-11-04]
38. Quality data. From real people. For faster breakthroughs. Prolific. URL: https://www.prolific.com/ [accessed 2025-11-04]
39. Russell AM, Browne M, Hing N, Rockloff M, Newall P. Are any samples representative or unbiased? Reply to Pickering and Blaszczynski. Int Gambl Stud 2022 Jan 2;22(1):102-113. [doi: 10.1080/14459795.2021.1973535]
40. Douglas BD, Ewell PJ, Brauer M. Data quality in online human-subjects research: comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. PLoS ONE 2023;18(3):e0279720. [doi: 10.1371/journal.pone.0279720] [Medline: 36917576]
41. Stanton K, Carpenter RW, Nance M, Sturgeon T, Villalongo Andino M. A multisample demonstration of using the prolific platform for repeated assessment and psychometric substance use research. Exp Clin Psychopharmacol 2022 Aug;30(4):432-443. [doi: 10.1037/pha0000545] [Medline: 35025585]
42. Weathers FW, Litz B, Keane T, Palmieri P, Marx B, Schnurr P. The PTSD checklist for DSM-5 (PCL-5). US Department of Veterans Affairs. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2025-11-04]
43. Ferris J, Wynne H. The Canadian Problem Gambling Index: final report. Canadian Centre on Substance Abuse. URL: https://www.greo.ca/Modules/EvidenceCentre/files/Ferris%20et%20al(2001)The_Canadian_Problem_Gambling_Index.pdf [accessed 2025-11-07]
44. Pfund RA, Ginley MK, Rash CJ, Zajac K. Contingency management for treatment attendance: a meta-analysis. J Subst Abuse Treat 2022 Feb;133:108556. [doi: 10.1016/j.jsat.2021.108556] [Medline: 34210566]
45. UK armed forces biannual diversity statistics. Ministry of Defence. 2024 Apr. URL: https://www.gov.uk/government/statistics/uk-armed-forces-biannual-diversity-statistics-april-2024/uk-armed-forces-biannual-diversity-statistics-april-2024 [accessed 2025-11-04]
46. Tran LT, Wardle H, Colledge-Frisby S, et al. The prevalence of gambling and problematic gambling: a systematic review and meta-analysis. Lancet Public Health 2024 Aug;9(8):e594-e613. [doi: 10.1016/S2468-2667(24)00126-9] [Medline: 39025095]

Abbreviations

ACT: acceptance and commitment therapy
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
mHealth: mobile health
PTSD: posttraumatic stress disorder

XSL•FO
RenderX

Edited by L Buis; submitted 10.03.25; peer-reviewed by C Mills-Finnerty, T Harpel; revised version received 13.10.25; accepted 14.10.25; published 11.11.25.

Please cite as:
Heath C, Williams JM, Leightley D, Murphy D, Dymond S
Leveraging Social Media and Crowdsourcing to Recruit and Retain Military Veterans With Posttraumatic Stress Disorder or Experience of Harmful Gambling for mHealth Interventions: Descriptive Study
JMIR Mhealth Uhealth 2025;13:e73706
URL: https://mhealth.jmir.org/2025/1/e73706
doi:10.2196/73706

© Conor Heath, Jess M Williams, Daniel Leightley, Dominic Murphy, Simon Dymond. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cluster-Based Predictive Modeling of User Ratings for Physical Activity Apps Using Mobile App Rating Scale (MARS) Dimensions: Model Development and Validation

Ayush Bhattacharya, MS; Jose Fernando Florez-Arango, MSc, MD, PhD

Department of Population Health Sciences, Weill Cornell Medicine, 575 Lexington Ave, Room FP 1025, New York, NY, United States

Corresponding Author:
Jose Fernando Florez-Arango, MSc, MD, PhD
Department of Population Health Sciences, Weill Cornell Medicine, 575 Lexington Ave, Room FP 1025, New York, NY, United States

Abstract

Background: The expansion of mobile health app or apps has created a growing need for structured and predictive tools to evaluate app quality before deployment. The Mobile App Rating Scale (MARS) offers a standardized, expert-driven assessment across 4 key dimensions—engagement, functionality, aesthetics, and information—but its use in forecasting user satisfaction through predictive modeling remains limited.

Objective: This study aimed to investigate how k-means clustering, combined with machine learning models, can predict user ratings for physical activity apps based on MARS dimensions, with the goal of forecasting ratings before production and uncovering insights into user satisfaction drivers.

Methods: We analyzed a dataset of 155 MARS-rated physical activity apps with user ratings. The dataset was split into training (n=111) and testing (n=44) subsets. K means clustering was applied to the training data, identifying 2 clusters. Exploratory data analysis included box plots, summary statistics, and component+residual plots to visualize linearity and distribution patterns across MARS dimensions. Correlation analysis was performed to quantify relationships between each MARS dimension and user ratings. In total, 5 machine learning models—generalized additive models, k-nearest neighbors, random forest, extreme gradient boosting, and support vector regression—were trained with and without clustering. Models were hypertuned and trained separately on each cluster, and the best-performing model for each cluster was selected. These predictions were combined to compute final performance metrics for the test set. Performance was evaluated using correct prediction percentage (0.5 range), mean absolute error, and R^2. Validation was performed on 2 additional datasets: mindfulness (n=85) and older adults (n=55) apps.

Results: Exploratory data analysis revealed that apps in cluster 1 were feature-rich and scored higher across all MARS dimensions, reflecting comprehensive and engagement-oriented designs. In contrast, cluster 2 comprised simpler, utilitarian apps focused on basic functionality. Component+residual plots showed nonlinear relationships, which became more interpretable within clusters. Correlation analysis indicated stronger associations between user ratings and engagement and functionality, but weaker or negative correlations with aesthetics and information, particularly in cluster 2. In the unclustered dataset, k nearest neighbors achieved 79.55% accuracy, mean absolute error=0.26, and R^2=0.06. The combined support vector regression (cluster 1)+k-nearest neighbors (cluster 2) model achieved the highest performance: 88.64% accuracy, mean absolute error=0.27, and R^2=0.04. Clustering improved prediction accuracy and enhanced alignment between predicted and actual user ratings. Models also generalized well to the external datasets.

Conclusions: The combined clustering and modeling approach enhances prediction accuracy and reveals how user satisfaction drivers vary across app types. By transforming MARS from a descriptive tool into a predictive framework, this study offers a scalable, transparent method for forecasting user ratings during app development—particularly useful in early-stage or low-data settings.

(JMIR Mhealth Uhealth 2025;13:e70987) doi:10.2196/70987

KEYWORDS

Mobile App Rating Scale; user rating prediction; machine learning; k means clustering; predictive modelling; MARS

XSL•FO
RenderX

Introduction

The Mobile App Rating Scale (MARS) is a tool designed to assess the quality of mobile health apps [1]. It was developed to meet the increasing demand for a standardized and reliable method for app evaluation. This 23-item, 5-point rating system evaluates aspects such as user engagement (engagement), feature performance (functionality), visual design (aesthetics), content accuracy (information), and overall satisfaction, providing a holistic assessment to identify high-quality apps [2].

Although MARS has been widely used to quantify app quality across various categories [3-11], no study has yet explored its potential to predict app user ratings—a key indicator of an app's popularity and perceived value. User ratings are meaningful because they reflect the aggregated experiences of the user base and often influence app visibility, downloads, and credibility in app store ecosystems. Exploring how MARS scores relate to these ratings could offer insights into the underlying drivers of user satisfaction and support more informed design and evaluation decisions.

This work builds upon our previous research in which we applied k-means clustering to MARS-rated drug-drug interaction apps [12]. That study revealed that the correlation between MARS dimensions and user ratings varied significantly across app clusters. In some groups, functionality was most strongly associated with user ratings, while in others, dimensions like engagement or information played a larger role. These findings highlighted that relationships between expert scores and user feedback are not uniform but highly context-dependent. Clustering enabled the identification of feature-based subgroups aligned with distinct satisfaction profiles—reinforcing the value of segmentation when interpreting MARS-based evaluations.

Complementary work by Aziz et al [13] and Mertens et al [14] also demonstrated the value of clustering in understanding user behavior and engagement in mobile interventions. Aziz et al identified behavioral archetypes based on app usage and personality traits, while Mertens et al showed that user impact varies depending on engagement profiles over time. Although these studies focus on clustering users rather than apps, they reinforce a common insight: segmentation enhances our understanding of satisfaction drivers.

Alternatively, other researchers have pursued bottom-up strategies for quality assessment that rely on postlaunch user data. For example, Haoues et al [15] used machine learning and natural language processing on user reviews within the International Organization for Standardization or International Electrotechnical Commission 25010 framework to assess app quality. While these methods can capture sentiment and usability themes, they require large volumes of user feedback—often unavailable in early-stage development.

Building on these findings, our study combines unsupervised clustering and supervised machine learning to predict app store user ratings based on expert-evaluated app quality dimensions. By segmenting apps into meaningful groups and training cluster-specific models, we aimed to improve prediction accuracy and uncover which MARS dimensions matter most in each context. This top-down approach enables forecasting of user satisfaction before release, making it particularly valuable in low-data settings. Ultimately, it offers developers a proactive, scalable framework to anticipate user expectations and optimize app design before launch.

Methods

Reporting Standards

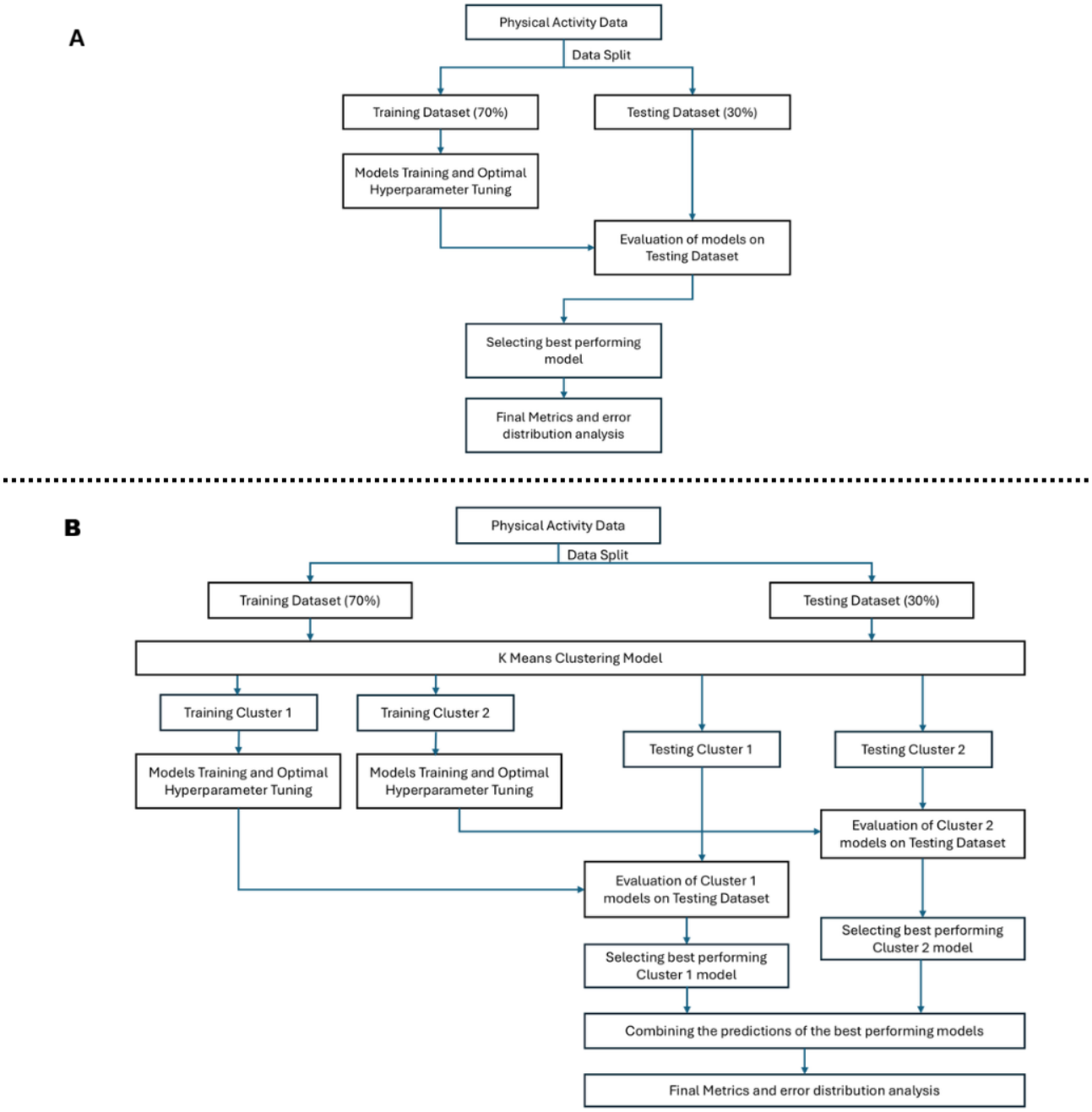
This study adheres to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) [16] guidelines to ensure comprehensive and transparent reporting of prediction model development and evaluation.

Data Source and Eligibility Criteria

This study uses physical activity app data originally compiled by Paganini et al [3] and made accessible by Terhorst et al [17]. The dataset included item-level MARS scores previously evaluated and published by the original authors as part of their systematic app evaluation studies, with complete assessments by trained experts. No new MARS evaluations were conducted. Since the objective of this study was to examine the predictive utility of MARS scores in relation to user ratings, we worked directly with the original reported scores to maintain analytical neutrality and avoid introducing evaluator bias. Of the 312 apps in the dataset, 155 had complete MARS evaluations and corresponding user ratings and were included in the analysis. Apps with missing data for any dimension or user rating were excluded using a complete-case approach, with no imputation performed. Figure 1 illustrates the overall workflow of the study, with Figure 1A depicting the modeling process without clustering and Figure 1B showing the process incorporating clustering.

XSL•FO
RenderX

Figure 1. Workflow for predictive modeling (1A) without clustering (1B) with K-means clustering.



Outcome and Predictor Variables

The outcome variable was the average user rating on a scale from 1 to 5. Predictor variables included the 4 MARS dimensions: engagement, functionality, aesthetics, and information. No transformations or standardization were applied to the predictor variables.

Exploratory Data Analysis

The dataset was first divided into training (111/155, 70%) and testing (44/155, 30%) sets. The training dataset was used for exploratory data analysis (EDA) and training the k-means and prediction models, while the testing dataset was held out for final model evaluation. During the EDA phase, we examined the relationships between user ratings and the 4 MARS dimensions. We visualized distribution patterns using box plots, which provided insights into the central tendencies and variability across the dimensions. In addition, we created a summary table to present descriptive statistics for each dimension, including the mean, median, 25th and 75th percentiles, range, minimum, and maximum values [12,18]. We then generated component+residual plots for each MARS dimension to better understand the data patterns and guide the selection of appropriate modeling techniques [19]. In these plots, the blue line represents the fitted values showing the main trend between the predictor and user ratings, while the pink line provides a smoother fit, highlighting any deviations or nonlinear relationships. Finally, we conducted a correlation analysis to quantify the strength and direction of associations between each MARS dimension and user ratings, highlighting dimensions with the most significant impact on user satisfaction [12].

XSL•FO
RenderX

Model Development and Hyperparameter Tuning

We applied 5 machine learning models: generalized additive models (GAM), k-nearest neighbors (KNN), random forest (RF), extreme gradient boosting (XGBoost), and support vector regression (SVR). Each model was selected for its unique approach to capturing different data patterns: GAM for handling potential nonlinear relationships [20], KNN for proximity-based prediction by averaging ratings of similar apps [21], RF as an ensemble approach combining multiple decision trees to improve accuracy [22], XGBoost for its boosting techniques that enhance prediction accuracy [23], and SVR for determining the optimal hyperplane to separate groups effectively [24].

Hyperparameter tuning for each of the 5 models was conducted using 5-fold cross-validation on the training dataset [25], optimizing the models for the best predictive performance. For GAM, tuning involved adjusting the number of knots in the smoothing splines to capture different levels of nonlinearity [26]. For KNN, a range of k values was evaluated to find the optimal number of neighbors [27]. For RF, the number of randomly selected features at each split was tuned to balance model depth and generalization [23]. For XGBoost, the number of boosting rounds, learning rate, and tree depth were optimized [23]. SVR tuning involved testing different combinations of cost and gamma parameters to optimize the margin of tolerance and the influence of individual data points [24]. The best hyperparameters were selected based on achieving the highest percentage of correct predictions (within a 0.5 range), the lowest mean absolute error (MAE), and the highest R^2 values [28]. After identifying the optimal hyperparameters, the models were evaluated on the testing dataset using the same metrics. All analyses were conducted using the latest version of RStudio.

Clustering Analysis

To further refine the predictive modeling, we incorporated k-means clustering as a preprocessing step, grouping the apps into clusters based on their MARS scores. This approach aimed to capture more nuanced relationships between app characteristics and user ratings [12]. The optimal number of clusters was determined using an elbow plot, which identifies the point where the within-cluster sum of squares (WCSS) shows a diminishing return with increasing clusters [12,29]. Although widely used, this method involves a degree of subjectivity, as identifying the inflection point in the curve relies on visual interpretation.

Cluster-Specific Modeling

To characterize the clusters identified through k-means, we performed a second round of EDA, mirroring the initial steps conducted on the entire training dataset. Following the EDA, the machine learning models were applied separately to each cluster, with hyperparameter tuning conducted individually for each cluster to ensure optimal model performance. This cluster-specific approach aimed to better capture the unique relationships between MARS scores and user ratings within each group.

The testing dataset was segmented using the trained k-means clustering model to allocate each data point to the appropriate cluster. Each cluster-specific model was then tested on its respective cluster within the testing dataset, using the same evaluation metrics as before.

Combined Prediction and Performance Evaluation

To assess the overall effectiveness of the clustering-based approach, predictions from the best-performing models were consolidated to compute the overall metrics for the entire testing dataset. The best model for each cluster was selected based on the highest correct prediction percentage, the lowest MAE, and the highest R^2 value. Predictions from each cluster's best model were combined to calculate the final overall metrics for the entire testing dataset. This enabled the direct comparison of outcomes of the model's performance with and without the clustering step.

Scatter plots of actual versus predicted user ratings, with a 0.5 tolerance band, were used to visually assess alignment across the test set for both unclustered and clustered models. The diagonal line in each plot indicates where predicted ratings equal actual ratings. We also analyzed the distribution of prediction accuracy across error ranges, which provided deeper insights into how many apps were accurately predicted within each range. These distributions were visualized using bar plots.

Model Generalizability Across App Categories

To validate our model's generalizability, we applied the same clustering and modeling approach to 2 additional datasets: mindfulness [4] and older adults [6] apps. The process of evaluation remained the same, with clustering, hyperparameter tuning, and model evaluation performed as previously described.

Ethical Considerations

This study did not involve human participants, patient data, or identifiable information. The analysis was limited to publicly available mobile applications and datasets. Institutional review board approval was not required, and informed consent was not applicable.

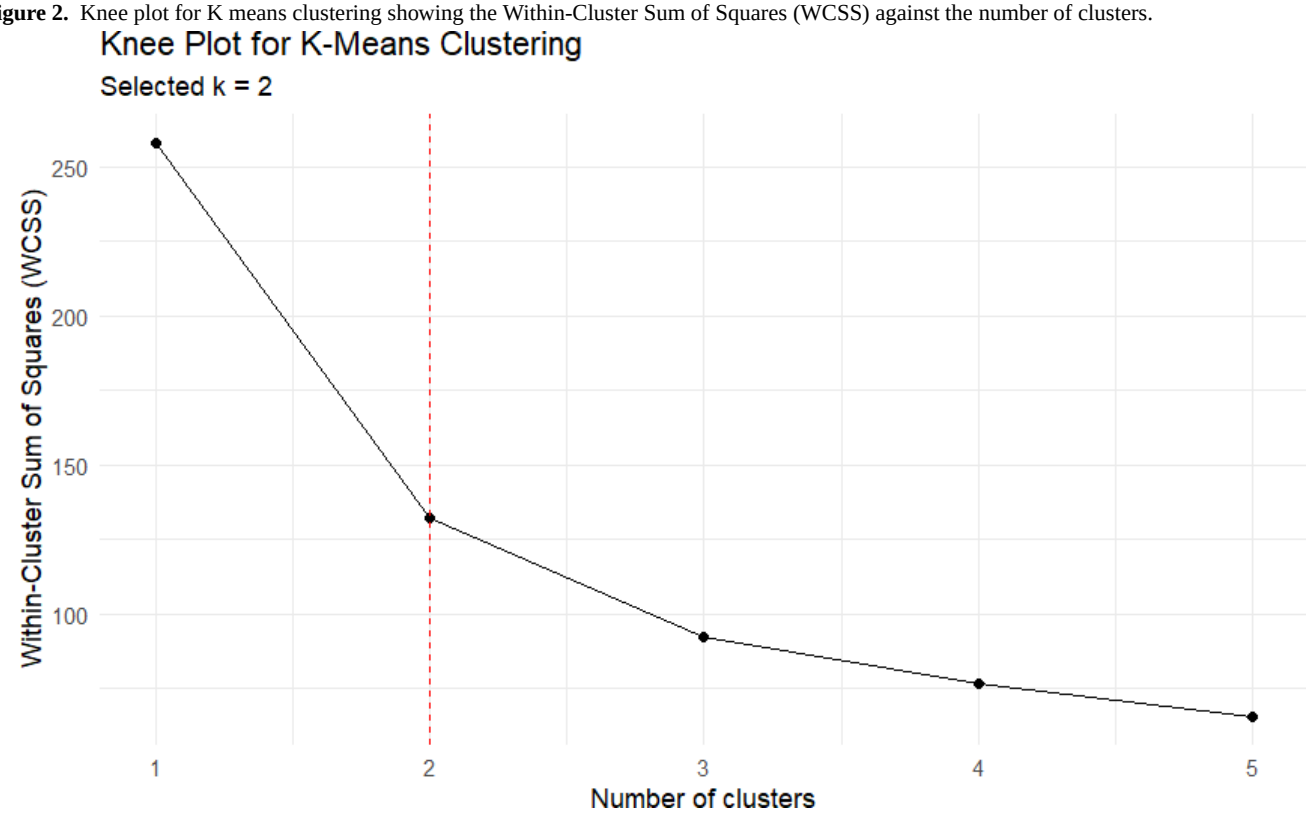
Results

K Means Clustering

The knee plot, shown in Figure 2, illustrates the relationship between the number of clusters and the WCSS. As the number of clusters increases, the WCSS decreases, indicating that data points are more tightly grouped within each cluster [12]. However, beyond 2 clusters, the rate of decrease in WCSS slows significantly, marking an "elbow" point. This suggests that adding more clusters beyond 2 provides diminishing returns in compactness. Thus, selecting 2 clusters was deemed optimal, as it balances the trade-off between data cohesion within clusters and the simplicity of the model structure.

XSL•FO
RenderX

Figure 2. Knee plot for K means clustering showing the Within-Cluster Sum of Squares (WCSS) against the number of clusters.



EDA

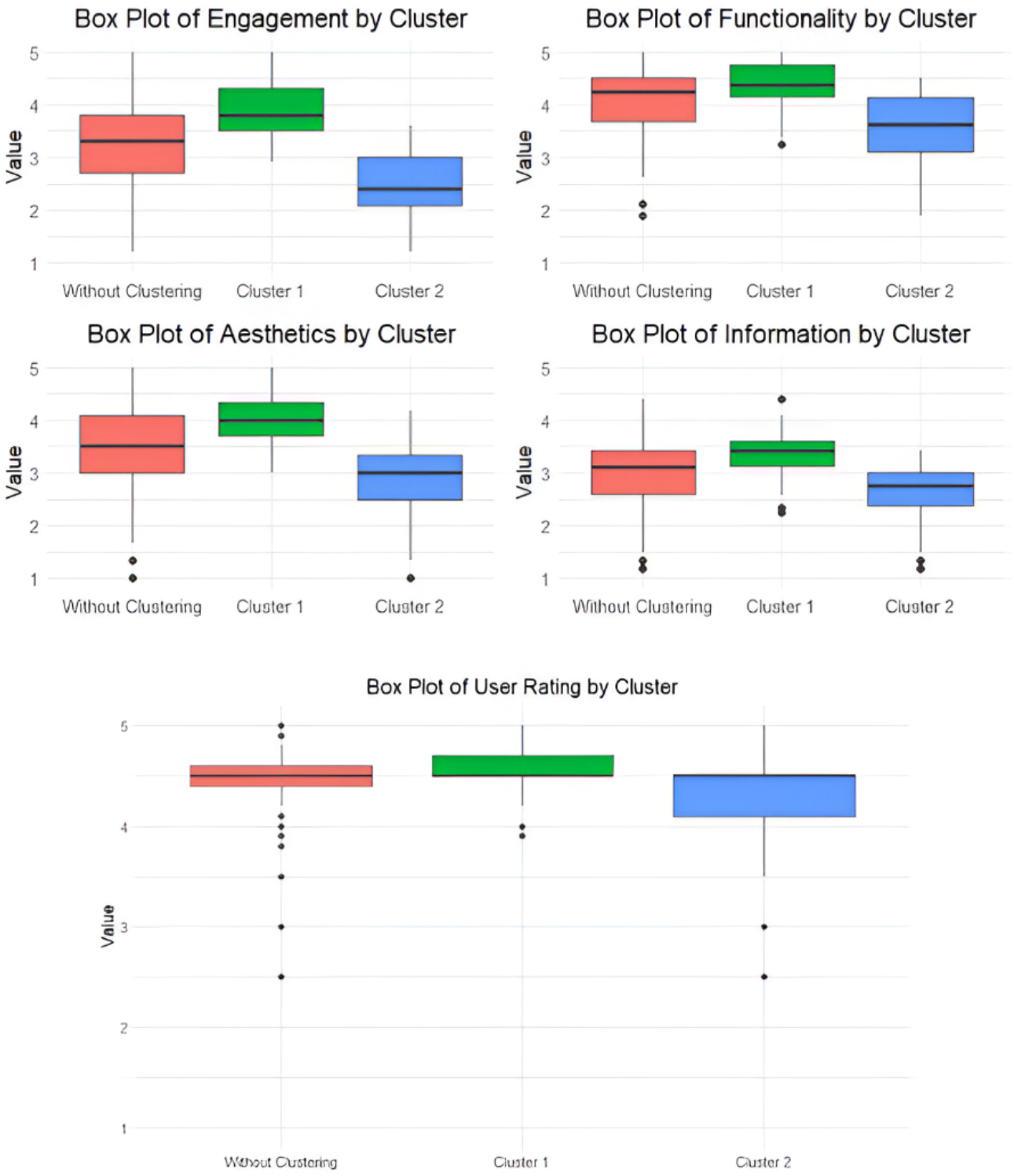
Data Distribution

The EDA began by examining the distributions of the 4 MARS dimensions and user ratings without applying clustering. The box plots (Figure 3) for the "Without Clustering" group showed centralized distributions across all dimensions, with some variability, particularly in engagement (range=3.80) and aesthetics (range=4.00). These findings provided a baseline understanding of the training dataset's characteristics.

Application of k-means clustering resulted in 2 clusters: cluster 1 with 62 apps (55.9%) and cluster 2 with 49 apps (44.1%). The trained clustering model was then used to assign the testing apps to clusters, resulting in 21 apps in cluster 1 and 23 in cluster 2. Although the distribution is slightly imbalanced, both clusters retained sufficient size in both the training and testing sets to support reliable model training and evaluation. We then re-evaluated the distributions for each dimension and user rating within each cluster, as shown in the box plots (Figure 3) and summary statistics table (Table 1). Cluster 1 consistently exhibited higher medians across all dimensions—engagement (3.80), functionality (4.38), aesthetics (4.00), and information (3.42)—compared to cluster 2, which had corresponding medians of 2.40, 3.62, 3.00, and 2.75, respectively. These results suggest that cluster 1 apps, based on MARS evaluations, demonstrate stronger quality features than cluster 2, which includes lower-rated apps. The "Without Clustering" analysis produced average scores that fell between the 2 clusters.

XSL•FO
RenderX

Figure 3. Box plots of MARS dimensions and user ratings across unclustered and clustered data. MARS: Mobile App Rating Scale.



XSL•FO
RenderX

Table . Summary statistics for MARS dimensions and user ratings across unclustered and clustered data.

Clusters and Dimensions	Values		
	Mean (SD)	Median (IQR)	Range (Min-Max)
Training data sample size (n=111), without clustering			
Engagement	3.23 (0.90)	3.30 (2.70-3.80)	3.80 (1.20-5.00)
Functionality	4.06 (0.65)	4.25 (3.69-4.50)	3.12 (1.88-5.00)
Aesthetics	3.50 (0.85)	3.50 (3.00-4.09)	4.00 (1.00-5.00)
Information	3.01 (0.61)	3.12 (2.60-3.43)	3.23 (1.17-4.40)
User rating	4.46 (0.37)	4.50 (4.40-4.60)	2.50 (2.50-5.00)
Training data sample size (Cluster 1), (n=62)			
Engagement	3.84 (0.55)	3.80 (3.50-4.30)	2.10 (2.90-5.00)
Functionality	4.42 (0.43)	4.38 (4.15-4.75)	1.75 (3.25-5.00)
Aesthetics	4.05 (0.51)	4.00 (3.71-4.33)	2.00 (3.00-5.00)
Information	3.32 (0.47)	3.42 (3.13-3.60)	2.15 (2.25-4.40)
User ratings	4.56 (0.27)	4.50 (4.50-4.70)	1.10 (3.90-5.00)
Training data sample size (Cluster 2), (n=49)			
Engagement	2.46 (0.61)	2.40 (2.10-3.00)	2.40 (1.20-3.60)
Functionality	3.62 (0.62)	3.62 (3.12-4.12)	2.62 (1.88-4.50)
Aesthetics	2.80 (0.67)	3.00 (2.50-3.33)	3.17 (1.00-4.17)
Information	2.62 (0.55)	2.75 (2.38-3.00)	2.25 (1.17-3.42)
User ratings	4.32 (0.44)	4.50 (4.10-4.50)	2.50 (2.50-5.00)

Component+Residual Plots

From the component+residual plots of the unclustered analysis (Figure 4A), residuals appear widely dispersed (–2.0 to +1.0) across all dimensions, with only subtle trends. Engagement (1.20 to 5.00) shows a slight positive relationship with user ratings, while functionality (1.88 to 5.00) displays an initial upward trend in residuals up to around 4.0, then plateaus—suggesting diminishing returns at higher scores. In contrast, aesthetics (1.00 to 5.00) and information (1.17 to 4.40) show a negative trend, indicating inverse associations with user ratings. These generalized, nonlinear patterns reflect the broad variability of app characteristics in the training dataset, especially for aesthetics (range=4.00) and engagement (range=3.80).

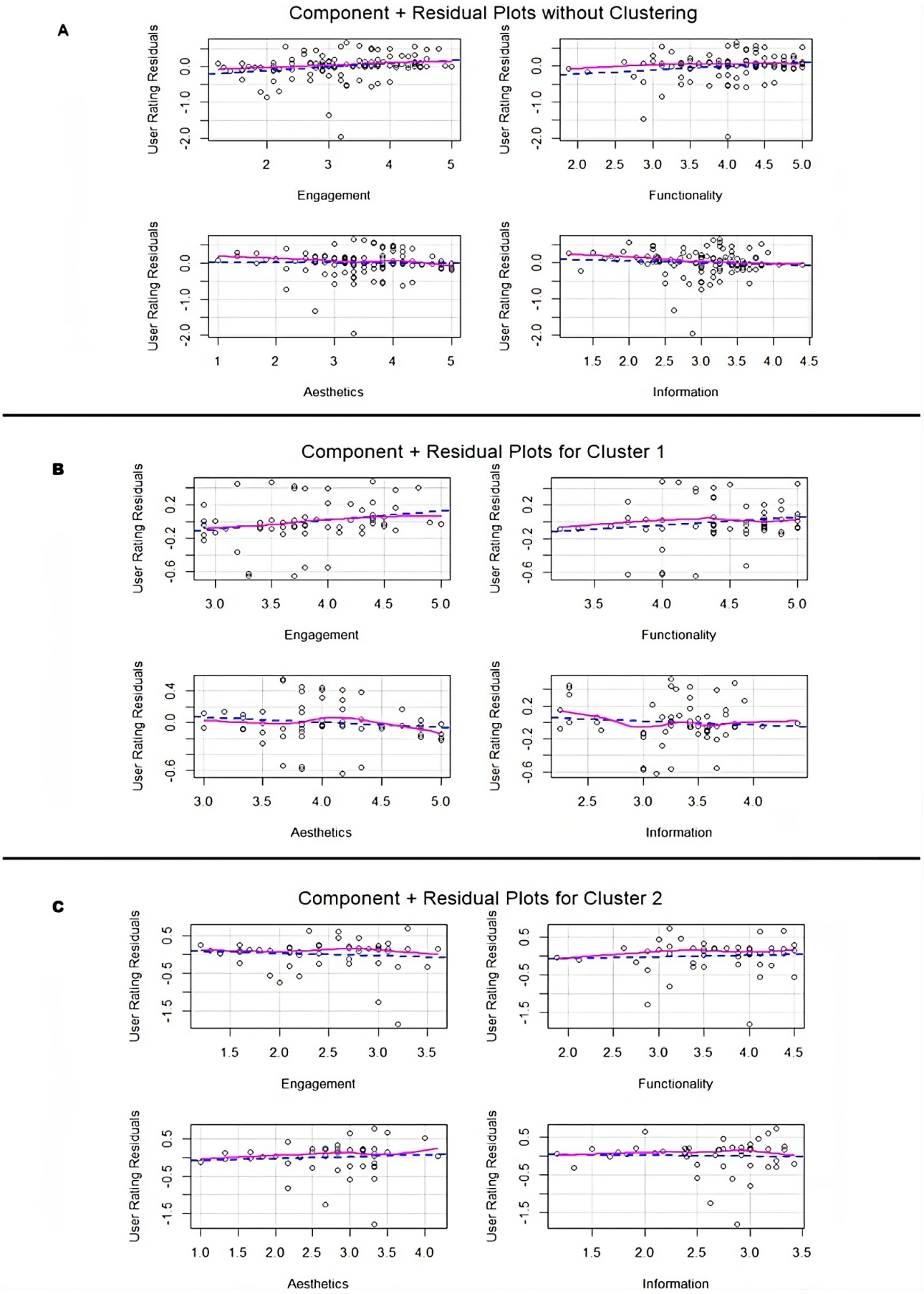
In Cluster 1 (Figure 4B), residuals were more consistent and compact across all dimensions (–0.6 to +0.4), reflecting the group's overall uniformity. Engagement (2.90 to 5.00) showed an initial upward trend that plateaued beyond approximately 4.3, while functionality (3.25 to 5.00) also rose early before flattening after 4.5—indicating nonlinear associations with user ratings. Aesthetics (3.00 to 5.00) displayed a complex curve, dipping, then rising, then falling again. Information (2.25 to 4.40) followed a similar nonlinear pattern, declining first, then increasing, and eventually leveling off around 3.5 - 4.0. These results suggest that although cluster 1 is more cohesive, relationships between MARS dimensions and user ratings remain complex.

In cluster 2 (Figure 4C), residuals showed greater variability with values ranging from approximately –1.5 to +0.5. Engagement (1.20 to 3.60) showed a downward trend, between 2.0 and 3.2, indicating a weak negative association with user ratings. Functionality (1.88 to 4.50) and aesthetics (1.00 to 4.17) exhibited slight upward trends, between 2.5 - 4.0 and 2.0 - 3.5, respectively, indicating modest positive associations. Information (1.17 to 3.42) showed a nonlinear pattern—increasing initially up to 2.6, then declining beyond 2.8—suggesting that high Information scores may correspond to lower user ratings. These findings highlight greater heterogeneity and less stable patterns in cluster 2, reinforcing the need for distinct modeling approaches between clusters.

In summary, the unclustered data shows weak and scattered associations. Cluster 1 reveals cohesive but complex nonlinear patterns across all dimensions. Cluster 2 shows more variability, with modest positive trends in functionality and aesthetics.

XSL•FO
RenderX

Figure 4. Component+residual plots for MARS dimensions and user ratings (4A) without clustering (4B) for cluster 1 (4C) for cluster 2. MARS: Mobile App Rating Scale

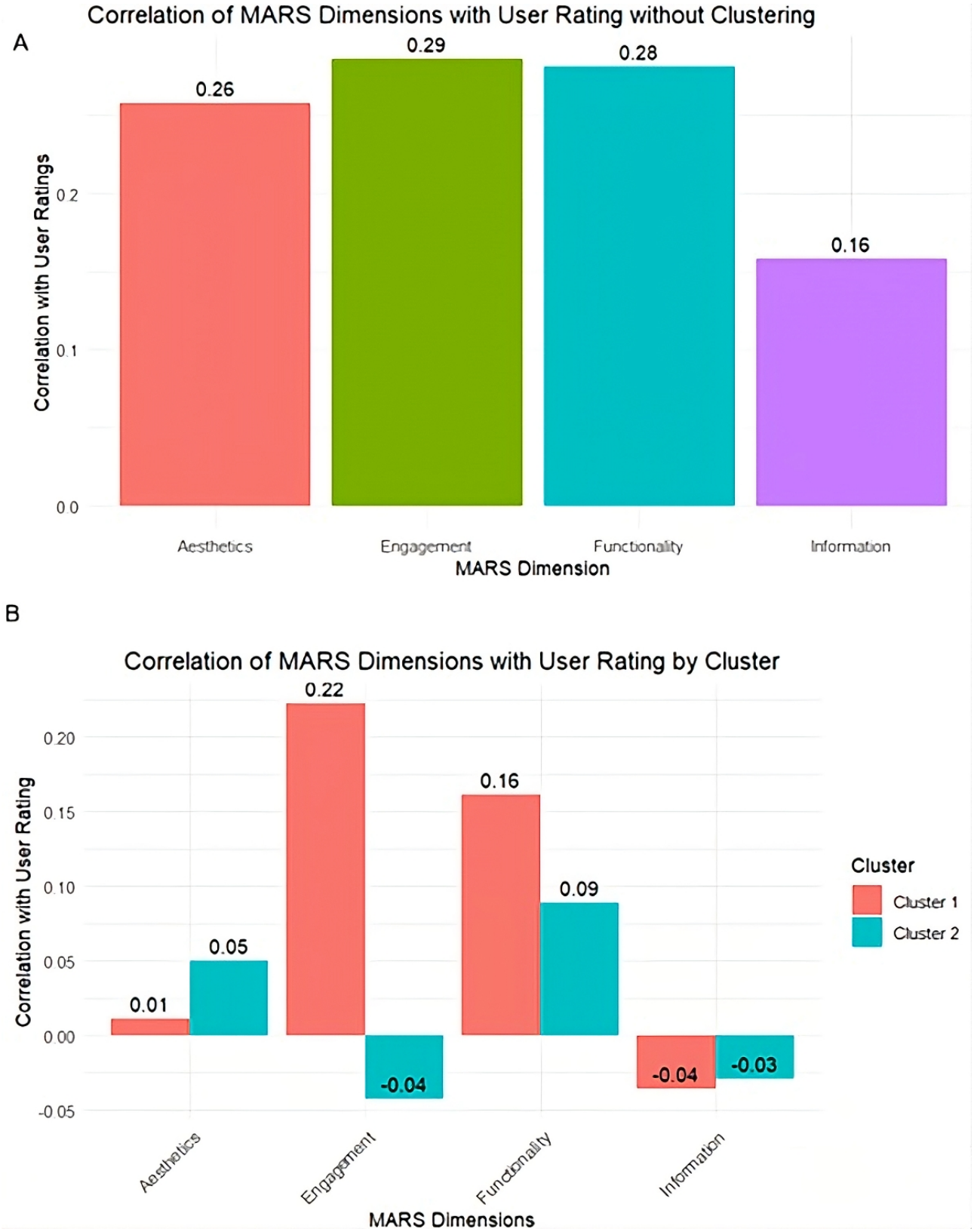


XSL•FO
RenderX

Correlation Analysis

Correlation analysis between MARS dimensions and user ratings (Figure 5) revealed different patterns when comparing the unclustered dataset to the segmented clusters. In the unclustered analysis (Figure 5A), engagement and functionality show positive correlations with user ratings, at 0.29 and 0.28, respectively, indicating a moderate association. Aesthetics follows with a correlation of 0.26, while information has the lowest correlation with user ratings at 0.16. These findings suggest that, across the entire dataset, engagement and functionality are the primary drivers of user satisfaction.

Figure 5. Correlation of MARS dimensions with user rating (5A) without clustering (5B) with clusters. MARS: Mobile App Rating Scale.



XSL•FO
RenderX

Within the clusters (Figure 5B), correlations appear less pronounced. In cluster 1, engagement has the highest correlation (0.22), followed by functionality (0.16). Aesthetics and information show minimal or no correlation, at 0.01 and -0.04, respectively. This indicates that while engagement and functionality remain relevant in cluster 1, their influence is weaker compared to the unclustered data.

In cluster 2, correlations are uniformly weak. Engagement (–0.04) and information (–0.03) show negative correlations, while aesthetics (0.05) and functionality (0.09) exhibit weak positive associations. These findings suggest that in cluster 2, no single MARS dimension strongly influences user ratings, pointing to a user base whose satisfaction may depend on factors outside the scope of MARS, or potential misalignment between expert evaluations and user perceptions.

Predictions

Table 2 presents the final hyperparameter settings selected for each model and cluster. Table 3 summarizes the predictive performance of the models across the testing dataset and within each cluster.

Without clustering, KNN and SVR achieved the highest correct prediction percentage at 79.55%, both maintaining relatively low MAEs (0.26 and 0.29, respectively). GAM also performed reasonably well with 77.27% accuracy and the lowest MAE (0.25). RF and XGB showed weaker performance, with lower correct prediction percentages (72.72%) and negative R^2 values (−0.19), indicating poor model fit for these models. KNN was selected as the best-performing model without clustering due to its lower MAE of 0.26 and higher *R*-square of 0.06, compared to SVR's MAE of 0.29 and *R*-square of −0.13.

Table . Tuned hyperparameters by model type and cluster.

Clusters and models	Selected parameters after hypertuning
Without clustering	
Generalized additive model	Number of spline knots=4
K nearest neighbors	Number of neighbors=7
Random forest	Number of features selected at each split=2
Extreme gradient boosting	• Number of boosting rounds=150 • Maximum tree depth=1 • Learning rate=0.05
Support vector regression	• Cost=4 • Gamma=4
Cluster 1	
Generalized additive model	Number of spline knots=3
K nearest neighbors	Number of neighbors=7
Random forest nearest neighbors	Number of features selected at each split=4
Extreme gradient boosting	• Number of boosting rounds=150 • Maximum tree depth=1 • Learning rate=0.05
Support vector regression	• Gamma=1 • Cost=2
Cluster 2	
Generalized additive model	Number of spline knots=3
K nearest neighbors	Number of neighbors=8
Random forest	Number of features selected at each split=2
Extreme gradient boosting	• Number of boosting rounds=50 • Maximum tree depth=1 • Learning rate=0.1
Support vector regression	• Cost=0.5 • Gamma=4

XSL•FO
RenderX

Table . Model performance metrics with and without clustering.

Clusters and models	Correct predictions percentage	MAE[a]	R^2
Without clustering, (n=44)			
GAM[b]	77.27	0.25	0.10
KNN[c]	79.55	0.26	0.06
RF[d]	72.72	0.30	−0.19
XGB[e]	72.72	0.28	−0.19
SVR[f]	79.55	0.29	−0.13
Cluster 1, (n=27)			
GAM	85.19	0.24	0.12
KNN	85.19	0.25	0.14
RF	81.48	0.26	−0.03
XGB	77.78	0.25	0.01
SVR	88.89	0.27	0.05
Cluster 2, (n=17)			
GAM	88.24	0.29	−0.05
KNN	88.24	0.28	0.03
RF	82.35	0.34	−0.45
XGB	64.71	0.43	−1.26
SVR	88.24	0.28	−0.04
Combined cluster model, (n=44) SVR+KNN	88.64	0.27	0.04

[a]MAE: mean absolute error.

[b]GAM: generalized additive models.

[c]KNN: K-nearest neighbors

[d]RF: random forest.

[e]XGB: extreme gradient boosting.

[f]SVR: support vector regression.

In Cluster 1, SVR delivered the highest accuracy, achieving an 88.89% correct prediction rate and an MAE of 0.27. This model was closely followed by GAM and KNN, both achieving 85.19% accuracy, with GAM having a slightly lower MAE (0.24). For RF and XGB, the performance dropped, with RF showing a correct prediction percentage of 81.48% and XGB at 77.78%. The R^2 values indicate a better model fit within this cluster, especially for GAM and KNN, which displayed positive R^2 values, while RF and XGB struggled, showing low or negative values. SVR was selected as the best-performing model for cluster 1.

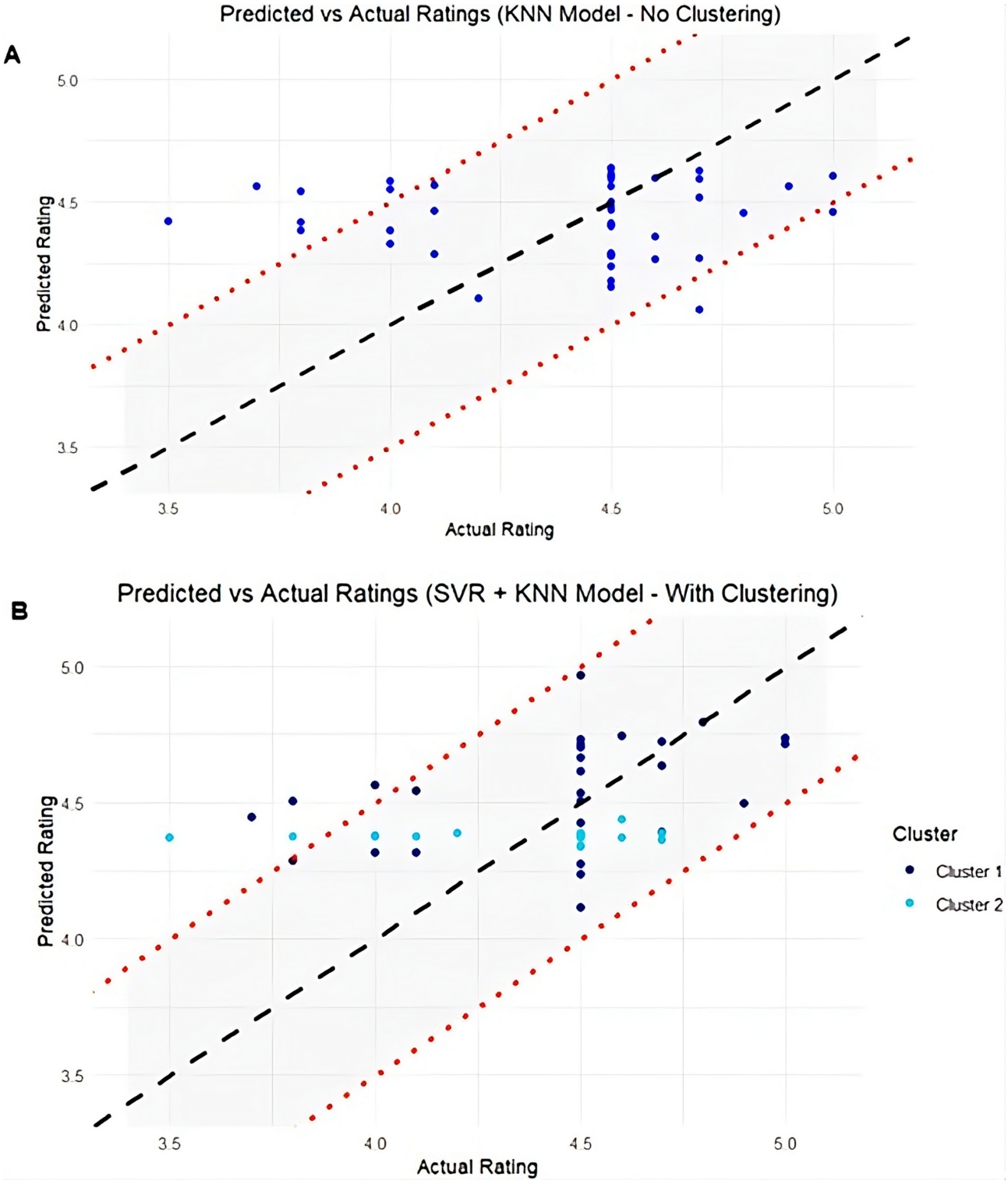
In cluster 2, SVR, KNN, and GAM achieved the same prediction accuracy, each at 88.24%. RF and XGB showed notable drops in performance, with XGB achieving the lowest accuracy (64.71%) and a high negative R^2 (−1.26), indicating poor fit and low predictive reliability within this cluster. KNN was selected as the best-performing model due to its lower MAE (0.28) and a positive R^2 (0.03), compared to GAM's MAE of 0.29 and R^2 of −0.05, and SVR's MAE of 0.28 and R^2 of −0.04.

The best-performing models from each cluster—SVR from cluster 1 and KNN from cluster 2—were selected, and their predictions were combined to compute the final metrics for the entire testing dataset. This combined approach achieved an improved overall correct prediction percentage of 88.64% outperforming the best unclustered data model, KNN, which achieved 79.55% accuracy.

To evaluate prediction alignment, we plotted actual versus predicted user ratings across the full test set for both the unclustered and clustered approaches (Figure 6). The unclustered KNN model (Figure 6A) exhibited wider deviations, particularly for lower and higher actual ratings. Two apps were underpredicted, and seven were overpredicted. In contrast, the cluster-based SVR+KNN model (Figure 6B) showed closer alignment with the actual values, with a greater number of predictions falling within the 0.5 tolerance band. Five apps were overpredicted, with 3 belonging to cluster 1 and 2 to cluster 2. This indicates improved accuracy and consistency following the clustering step.

XSL•FO
RenderX

Figure 6. Comparison of actual versus predicted user ratings (6A) KNN model predictions without clustering (6B) Combined SVR and KNN model predictions following clustering. KNN: k-nearest neighbors; SVR: support vector regression.



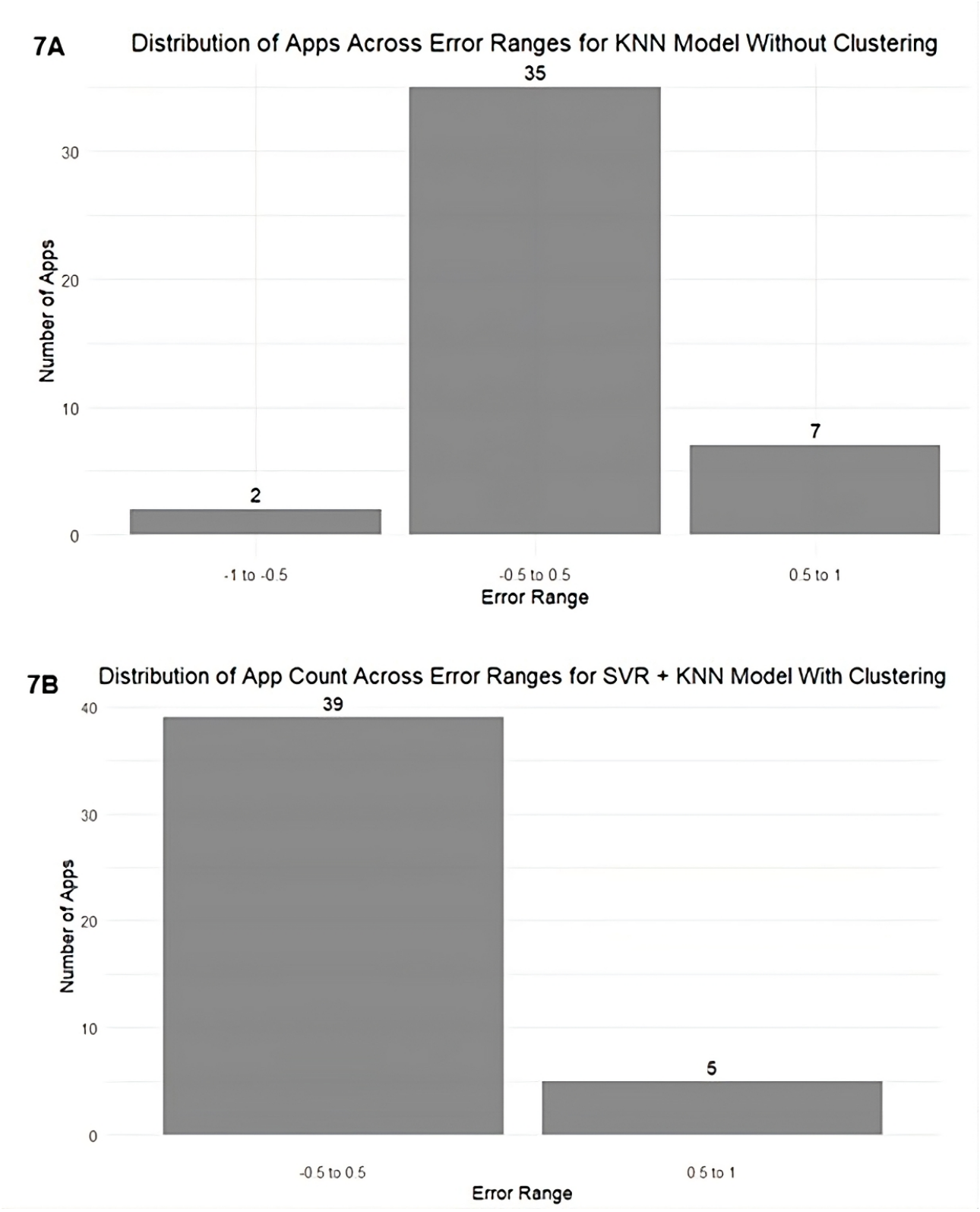
The error distribution plots (Figure 7) show the extent of prediction deviations across models. In the unclustered KNN model (Figure 7A), 2 apps had errors between −1 to −0.5, 35 were within the −0.5 to 0.5 range, and 7 had errors between 0.5 to 1. In comparison, the clustered SVR+KNN model (Figure 7B) showed improved precision, with 39 apps within the −0.5 to 0.5 range and only 5 apps falling in the 0.5 to 1 range.

Further to assess generalizability, the clustering-based modeling framework was validated on 2 external datasets: mindfulness apps (n=85) and older adult apps (n=55). In both cases, clustering either matched or outperformed the best-performing unclustered models. For the mindfulness dataset, the combined SVR+XGB cluster model achieved the highest overall accuracy (66.67%), a lower MAE (0.46), and a positive R^2 (0.07)—improving upon the best unclustered model (SVR), which had a negative R^2 (–0.06). Notably, cluster 2 models outperformed cluster 1, with XGB achieving an R^2 of 0.40, indicating that meaningful patterns emerged when segmenting

XSL•FO
RenderX

the apps. For the Elderly dataset, the combined KNN-based cluster model produced the best results, with 73.34% accuracy, an MAE of 0.48, and a positive R^2 (0.12)—surpassing the unclustered RF model, which had an R^2 of –0.08. Full results of both datasets are detailed in the Multimedia Appendix 1.

Figure 7. Error distribution of predictions (7A) using KNN model without clustering (7B) using SVR+KNN model with clustering. KNN: k-nearest neighbors; SVR support vector regression.



XSL•FO
RenderX

Discussion

Principal Findings

This study evaluated whether MARS scores could be used to predict user ratings through a clustering-based machine learning approach. The combined SVR+KNN model outperformed unclustered models, with higher accuracy and lower error. Clustering revealed distinct satisfaction patterns across app types and improved prediction alignment. These findings suggest that expert-driven MARS evaluations, when paired with data-driven modeling, can effectively forecast user satisfaction and ratings before app release.

In the unclustered dataset, MARS scores tended to average out, potentially obscuring the influence of specific features. Clustering enabled a deeper examination by isolating distinct subgroups, revealing more nuanced relationships between app characteristics and satisfaction—reflecting the diverse feature priorities among physical activity app users.

Cluster 1, characterized by higher mean scores across all MARS dimensions, appears to represent apps that align closely with user expectations by delivering a balanced, comprehensive experience. Users in this cluster may value features such as personalized workout recommendations, progress tracking, interactive challenges, or goal-setting tools [30-33]. These attributes contribute to a well-rounded app experience, where multiple dimensions—such as engagement through interactive features and functionality through reliable tracking—enhance overall satisfaction as reflected in the correlation analysis (Figure 5) with user ratings.

Conversely, cluster 2 exhibited lower scores across all MARS dimensions, likely representing simpler, function-focused apps. Residual plots (Figure 4C) showed negative patterns for engagement and information and modest positive trends for functionality and aesthetics—indicating that only certain dimensions influenced user ratings. Correlation analysis (Figure 5B) further supported this, revealing weak or negative associations for most dimensions, with functionality showing the strongest correlation. These patterns suggest that users in this segment prioritize basic features such as step counting, activity logging, or workout timers, with less emphasis on visual design or interactive elements [30-33].

Notably, the median user rating for the unclustered dataset, as well as for both clusters, remained consistent at 4.5 (Table 1), indicating a shared baseline of high satisfaction across groups, despite the lower MARS dimension scores observed in cluster 2. This indicates that functional adequacy alone may drive satisfaction for utilitarian apps, even when MARS scores are lower. This observation aligns with previous research showing that user expectations vary by app context. For example, Portenhauser et al [6] found that older adults prioritize simplicity, while Mani et al [4] noted that mindfulness app users emphasize content credibility. These findings also resonate with our earlier work on drug-drug interaction apps [12], which showed that the relationship between MARS dimensions and user satisfaction varied across clusters. Collectively, these results reinforce that no single MARS dimension universally predicts user satisfaction; instead, the relevance of each dimension depends on the app's functional scope and the expectations of its target audience.

From the correlation analysis, we observe that only certain MARS dimensions—particularly engagement and functionality—consistently correlate with user ratings, though the strength of these associations varies by cluster. This raises the broader question of how well expert evaluations align with real-world user perceptions [34,35]. While MARS is designed to objectively assess app quality, our findings suggest that its dimensions do not contribute equally to user satisfaction. For example, in cluster 2, all dimensions show weak or negligible correlations, offering little explanatory value regarding what drives user satisfaction in this group. This underscores the importance of context-aware evaluations—an insight that only became apparent after clustering revealed how user expectations vary across app groups.

In predictive modeling, KNN emerged as the top performer in the unclustered dataset, achieving an accuracy of 79.55%, with a relatively low MAE of 0.26 and an R^2 of 0.06. The low MAE suggests that KNN's predictions were generally close to actual user ratings, but the modest R^2 reflects limited explanatory power.

The prediction outcomes improved when clustering was applied, as models could be tailored to more similar groups of apps. In Cluster 1, SVR achieved the highest accuracy at 88.89%, with an MAE of 0.27 and a modest R^2 of 0.05, while KNN and GAM also showed improved performance with R^2 values above 0.10. This suggests that in apps with higher MARS scores, expert-assessed quality tends to align more closely with what users value—which is reflected in the improved R^2 values.

In contrast, Cluster 2 posed greater challenges. Although KNN and SVR both achieved high accuracy (88.24%), R^2 values for most models remained low or negative—notably −1.26 for XGB and –0.45 for RF. Even the best-performing models in this cluster, like KNN, achieved only a 0.03 R^2, indicating that while predictions were accurate in terms of MAE, they explained very little of the variance in user ratings. This suggests that user satisfaction in these simpler apps may stem from minimal but sufficient functionality rather than broader quality dimensions emphasized by MARS. The weak positive correlation for functionality further supports this interpretation, highlighting a possible disconnect between expert-assessed quality and user-perceived value in utility-focused apps.

The combined approach—using SVR for cluster 1 and KNN for cluster 2—achieved the highest overall accuracy at 88.64%, with a consistent MAE of 0.27 and R^2 of 0.04. While this confirms the value of clustering, the modest R^2 values across both settings underscore that MARS alone cannot fully explain user ratings. User satisfaction may be influenced by external factors such as marketing, demographics, or social dynamics—elements not captured by MARS. This highlights the need for future studies to incorporate contextual and user-centered variables (eg, downloads and review sentiment) to improve generalizability.

XSL•FO
RenderX

Further, an interesting observation emerged that even though the correlations between individual MARS dimensions and user ratings were weaker within clusters, the prediction accuracy improved. This highlights a limitation of traditional correlation analysis, which only captures simple, one-to-one relationships. In contrast, machine learning models can identify more complex patterns by analyzing multiple variables together, especially after clustering separates apps into more similar groups. This explains why weak correlations can still lead to strong predictions—an effect also noted in previous research that supports using model-based approaches over basic correlation in complex data contexts [36,37].

The comparison of actual versus predicted ratings shows that the unclustered KNN model (Figure 6A) exhibited underfitting, particularly at the extremes—likely a result of averaging across diverse app types. In contrast, the clustered SVR+KNN model (Figure 6B) demonstrated closer alignment with actual ratings and fewer large deviations. This improvement can be attributed to clustering, which grouped similar apps together, enabling the models to learn clearer and more relevant patterns. The error distribution plots (Figure 7) further support this, showing a reduction in the number and extent of prediction errors. Importantly, the overpredicted apps in the SVR+KNN model were distributed across both clusters—3 from cluster 1 and 3 from cluster 2—indicating that residual errors were not driven by any single subgroup but instead reflect natural prediction variability.

The evaluation of our framework on 2 external datasets—mindfulness and older adults—further supports the model's generalizability. As reported in the Results, clustering-based models consistently outperformed or matched the best unclustered models across both domains.

Taken together, our findings highlight the value of integrating structured expert evaluations with unsupervised clustering and machine learning to forecast user satisfaction in mobile health apps. By transforming MARS scores into actionable predictors, this framework enables early-stage evaluation of app quality, especially in the absence of large-scale user feedback. The improved predictive performance of cluster-specific models—validated across both physical activity and external domains—demonstrates that user satisfaction is best understood as context-specific, shaped by both app features and audience expectations. As the mobile health ecosystem continues to expand, such scalable and interpretable modeling approaches can play a pivotal role in guiding user-centered app design, regulatory review, and quality benchmarking before release. For real-world implementation, developers can first assign MARS scores to a representative set of apps within their domain to build context-specific clusters. They can then score their own app using the MARS framework and input these scores into our modeling pipeline to forecast its projected user rating. In future work, we aim to develop a standardized, user-friendly tool that will allow developers and evaluators to input MARS scores and receive predicted user satisfaction outcomes, streamlining early-stage decision-making.

Conclusion

This study demonstrates the effectiveness of combining unsupervised k-means clustering with supervised machine learning models to predict user ratings based on MARS dimensions in physical activity apps. The clustering approach revealed distinct user segments—cluster 1 users preferred feature-rich, engaging apps, while cluster 2 users gravitated toward simpler, utility-focused tools. By tailoring models to these groups, our combined approach (SVR for cluster 1 and KNN for cluster 2) outperformed the best unclustered model (KNN) in accuracy and reliability, offering more context-aware and consistent predictions.

This framework enables developers to forecast user satisfaction before app release using MARS-based expert evaluations. By identifying their app's alignment with specific user clusters, teams can make informed design choices, benchmark against competitors, and tailor strategies—even in low-data environments.

While prior applications of MARS have largely been descriptive, this study presents a predictive framework that transforms expert scores into actionable, prelaunch indicators of user satisfaction. Unlike postmarket sentiment analysis, our method supports proactive, data-informed decision-making that bridges the gap between expert evaluation and real-world user expectations.

By demonstrating improved predictive performance across clusters and validating the approach on external datasets, this work lays the foundation for a new generation of app evaluation tools that are not only interpretable and resource-efficient but also predictive. Future research should build on this by incorporating richer behavioral signals and contextual app metadata to further improve personalization and model precision.

Limitations

This study has several limitations. First, MARS scores offer a static, expert-driven evaluation and may not capture changes in app quality over time. Second, user satisfaction is shaped by contextual and behavioral factors that are not included in the MARS framework. Third, the sample size may limit the robustness of clustering and reduce generalizability across broader app ecosystems. Fourth, the accuracy of predictions is contingent on high-quality and consistent MARS scoring, as variability in rating practices could undermine model performance. Fifth, there is a risk of overfitting, particularly within cluster-specific models trained on relatively small subsets. Although hyperparameter tuning and cross-validation were used to minimize this, models such as RF and XGB may still capture spurious patterns specific to the training data. Finally, while the models were hypertuned using cross-validation, tuning choices are often context-dependent, and future applications may require re-optimization to fit specific app categories or user groups.

Future Scope of Work

Future work will be to quantify how well expert-driven MARS evaluations align with real-world user ratings, especially across diverse user segments. Incorporating additional contextual

XSL•FO
RenderX

data—such as user demographics, review sentiment, or app usage patterns—may help explain residual variance and improve model generalizability beyond MARS dimensions alone. To support practical adoption, we also plan to develop a standardized, user-friendly tool that enables developers to input MARS scores and apply our modeling framework to forecast projected user ratings, thereby streamlining early-stage decision-making.

Acknowledgments

This research received no external funding. No generative artificial intelligence or large language models were used in the writing, editing, or analysis of this manuscript.

Data Availability Mobile App Rating Scale

The study uses previously published Mobile App Rating Scale (MARS) data made available by Terhorst et al [3]. The data analysis and model development code are available from the corresponding author upon request.

Authors' Contributions

AB and JFFA jointly contributed to the conceptualization and methodology of the study. AB conducted the formal analysis, data processing, and model development and prepared the original draft of the manuscript. Both authors contributed to the review and editing of the manuscript. JFFA provided overall supervision and guidance throughout the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Validation of model using mindfulness data.
[DOCX File, 63 KB - mhealth_v13i1e70987_app1.docx]

Checklist 1
TRIPOD checklist.
[DOCX File, 89 KB - mhealth_v13i1e70987_app2.docx]

References

1. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
2. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. The mobile application rating scale (MARS): a new tool for assessing the quality of health apps. Presented at: In Young and Well Cooperative Research Centre (CRC) Connect National Conference; Feb 20-21, 2014.
3. Paganini S, Terhorst Y, Sander LB, et al. Quality of physical activity apps: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2021 Jun 9;9(6):e22587. [doi: 10.2196/22587] [Medline: 34106073]
4. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82. [doi: 10.2196/mhealth.4328] [Medline: 26290327]
5. Sucala M, Cuijpers P, Muench F, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
6. Portenhauser AA, Terhorst Y, Schultchen D, et al. Mobile apps for older adults: systematic search and evaluation within online stores. JMIR Aging 2021 Feb 19;4(1):e23313. [doi: 10.2196/23313] [Medline: 33605884]
7. Messner EM, Terhorst Y, Barke A, et al. The German Version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479. [doi: 10.2196/14479] [Medline: 32217504]
8. Messner EM, Sturm N, Terhorst Y, et al. Mobile apps for the management of gastrointestinal diseases: systematic search and evaluation within app stores. J Med Internet Res 2022 Oct 5;24(10):e37497. [doi: 10.2196/37497] [Medline: 36197717]
9. Terhorst Y, Messner EM, Schultchen D, et al. Systematic evaluation of content and quality of English and German pain apps in European app stores. Internet Interv 2021 Apr;24:100376. [doi: 10.1016/j.invent.2021.100376] [Medline: 33718002]
10. Domhardt M, Messner EM, Eder AS, et al. Mobile-based interventions for common mental disorders in youth: a systematic evaluation of pediatric health apps. Child Adolesc Psychiatry Ment Health 2021 Sep 13;15(1):49. [doi: 10.1186/s13034-021-00401-6] [Medline: 34517896]

XSL•FO
RenderX

11. Schoeppe S, Alley S, Rebar AL, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83. [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
12. Bhattacharya A, Florez-Arango JF. Investigating the quality of mobile apps for drug-drug interaction management using the mobile app rating scale and K-means clustering: systematic search of app stores. JMIR Mhealth Uhealth 2025 Oct 13;13:e65927. [doi: 10.2196/65927] [Medline: 41081618]
13. Aziz M, Erbad A, Belhaouari SB, Almourad MB, Altuwairiqi M, Ali R. Who uses mHealth apps? Identifying user archetypes of mHealth apps. Digit Health 2023;9:20552076231152175. [doi: 10.1177/20552076231152175] [Medline: 36714545]
14. Mertens ECA, Siezenga AM, van der Schalk J, van Gelder JL. A novel smartphone-based intervention aimed at increasing future orientation via the future self: a pilot randomized controlled trial of a prototype application. Prev Sci 2024 Feb;25(2):392-405. [doi: 10.1007/s11121-023-01609-y] [Medline: 37976010]
15. Haoues M, Mokni R, Sellami A. Machine learning for mHealth apps quality evaluation: an approach based on user feedback analysis. Softw Qual J 2023 May 23:1-31. [doi: 10.1007/s11219-023-09630-8] [Medline: 38625157]
16. Collins GS, Reitsma JB, Altman DG, Moons KGM, TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. Circulation 2015 Jan 13;131(2):211-219. [doi: 10.1161/CIRCULATIONAHA.114.014508] [Medline: 25561516]
17. Terhorst Y, Philippi P, Sander LB, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS ONE 2020;15(11):e0241480. [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
18. Agiro A, Bhattacharya A, Khan Z, Tung A, Brahmbhatt Y. PUK13 characteristics of patients receiving sodium zirconium cyclosilicate during hospital-based treatment of hyperkalemia. Value Health 2021 Jun;24:S236. [doi: 10.1016/j.jval.2021.04.1181]
19. Cook RD. Exploring partial residual plots. Technometrics 1993 Nov;35(4):351-362. [doi: 10.1080/00401706.1993.10485350]
20. Murase H, Nagashima H, Yonezaki S, Matsukura R, Kitakado T. Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: a case study in Sendai Bay, Japan. ICES J Mar Sci 2009 Jul 1;66(6):1417-1424. [doi: 10.1093/icesjms/fsp105]
21. Guo G, Wang H, Bell D, Bi Y, Greer K. KNN model-based approach in classification. Presented at: On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE; Nov 3-7, 2003; Catania, Sicily, Italy p. 986-996. [doi: 10.1007/978-3-540-39964-3_62]
22. Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. WIREs Data Min & Knowl 2019 May;9(3):e1301. [doi: 10.1002/widm.1301]
23. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Presented at: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13, 2016 p. 785-794. [doi: 10.1145/2939672.2939785]
24. Jakkula V. Tutorial on support vector machine (SVM). : School of EECS, Washington State University; 2006.
25. Schratz P, Muenchow J, Iturritxa E, Richter J, Brenning A. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol Modell 2019 Aug;406:109-120. [doi: 10.1016/j.ecolmodel.2019.06.002]
26. Maindonald J. Smoothing terms in GAM models. : The Australian National University URL: https://maths-people.anu.edu.au/~johnm/r-book/xtras/autosmooth.pdf [accessed 2025-11-06]
27. Zhang S, Li X, Zong M, Zhu X, Wang R. Efficient kNN classification with different numbers of nearest neighbors. IEEE Trans Neural Netw Learn Syst 2018 May;29(5):1774-1785. [doi: 10.1109/TNNLS.2017.2673241] [Medline: 28422666]
28. Hodson TO. Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not. Climate and Earth system modeling Preprint posted online on Mar 11, 2022. [doi: 10.5194/gmd-2022-64]
29. Kodinariya TM, Makwana PR. Review on determining number of Cluster in K-Means Clustering. Int J 2013 Nov;1(6):90-95 [FREE Full text]
30. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A review of smartphone applications for promoting physical activity. Jacobs J Community Med 2016;2(1):021. [Medline: 27034992]
31. He Z, Hassan MA, Saiz-González P, Ryu S, Wang R, Gao Z. Smartphone app-based interventions on physical activity behaviors and psychological correlates in healthy young adults: a systematic review. PLoS ONE 2024;19(4):e0301088. [doi: 10.1371/journal.pone.0301088] [Medline: 38578729]
32. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97. [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
33. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
34. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766. [doi: 10.2196/30766] [Medline: 35099398]

XSL•FO
RenderX

35. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849]
36. Shmueli G. To explain or to predict? SSRN Journal 2010:289-310. [doi: 10.2139/ssrn.1351252]
37. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003:1157-1182.

Abbreviations

EDA: exploratory data analysis
GAM: generalized additive models
KNN: k-nearest neighbors
MAE: mean absolute error
MARS: Mobile App Rating Scale
RF: random forest
SVR: support vector regression
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis
WCSS: within-cluster sum of squares
XGB: extreme gradient boosting

Edited by L Buis; submitted 07.01.25; peer-reviewed by M Gasmi, N Mungoli, Y Jiang; revised version received 14.06.25; accepted 30.07.25; published 06.11.25.

Please cite as:
Bhattacharya A, Florez-Arango JF
Cluster-Based Predictive Modeling of User Ratings for Physical Activity Apps Using Mobile App Rating Scale (MARS) Dimensions: Model Development and Validation
JMIR Mhealth Uhealth 2025;13:e70987
URL: https://mhealth.jmir.org/2025/1/e70987
doi:10.2196/70987

©Ayush Bhattacharya, José F Flórez-Arango. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of Heat Adaptation Behaviors on Resting Heart Rate Response to Summer Temperatures in Older Adults: Wearable Device Panel Study

Chi-Hsien Chen[1], MD, PhD; Feipei Lai[2], PhD; Yu-Lin Chen[3]; Yue Leon Guo[1,4,5], MD, PhD

[1]Department of Environmental and Occupational Medicine, National Taiwan University College of Medicine and National Taiwan University Hospital, Rm 339, 17 Syujhou Road, Taipei, Taiwan

[2]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

[3]School of Environmental and Forest Sciences, University of Washington, Seattle, Washington, United States

[4]Institute of Environmental and Occupational Health Sciences, National Taiwan University, Taipei, Taiwan

[5]National Institute of Environmental Sciences, National Health Research Institutes, Zhunan Township, Miaoli County, Taiwan

Corresponding Author:
Yue Leon Guo, MD, PhD
Department of Environmental and Occupational Medicine, National Taiwan University College of Medicine and National Taiwan University Hospital, Rm 339, 17 Syujhou Road, Taipei, Taiwan

Abstract

Background: The health impact of summer heat on older adults is a growing public concern, yet the physiological responses, particularly changes in resting heart rate (RHR), and the role of personal heat adaptation behaviors remain underexplored. Wearable devices offer an opportunity to objectively monitor physiological responses and evaluate the effectiveness of adaptation strategies in real-world settings.

Objective: This study aimed to quantify the short-term association between summer temperatures and RHR in older adults and to examine how individual heat adaptation behaviors modify this relationship, with additional consideration of personal characteristics such as age, sex, BMI, and chronic disease status.

Methods: We conducted a panel study among 83 community-dwelling older adults (≥65 y) in Taipei City during the summer of 2021 (May to September). Participants wore Garmin smartwatches to continuously monitor heart rate. Daily RHR was defined as the lowest 30-minute average heart rate. In September, heat adaptation behaviors were assessed via structured telephone interviews. Ambient temperature and relative humidity were obtained from a nearby monitoring station. Linear mixed-effect models were used to estimate temperature-RHR associations, and interaction terms were included to examine behavioral modifications. Subgroup analyses were conducted to explore effect modification by individual characteristics such as age, sex, BMI, and chronic disease status.

Results: Each 1 °C increase in daily mean temperature over lag days 0 - 1 was associated with a 0.11 (95% CI 0.07-0.15; *P*<.001) beats/min increase in RHR. After mutual adjustment for behaviors, several heat adaptation strategies showed significant protective effects, including reducing physical activity (β=−.15, *P*=.001), drinking cold beverages (β=−.24, *P*<.001), increasing naps or sleep duration (β=−.28, *P*=.003), drinking additional water ≥500 mL (β=−.10, *P*=.02), using air conditioner (AC) before (β=−.15, *P*=.002) and during sleep (β=−.13, *P*=.007), and using electric fans during sleep (β=−.12, *P*=.01). Subgroup analyses revealed stronger effects for certain behaviors in vulnerable populations: reduced physical activity was particularly beneficial for those with higher BMI; AC use and cold beverage intake were more effective in people with diabetes; increased naps yielded the largest benefits in individuals with hypertension; and the use of AC or fans during sleep was especially protective for older adults and females.

Conclusions: Summer heat is associated with elevated RHR in older adults, but this effect can be mitigated through targeted heat adaptation behaviors. Smartwatch monitoring provides a feasible and informative approach for capturing physiological changes, supporting the development of personalized heat-health recommendations for aging populations in a warming climate.

(JMIR Mhealth Uhealth 2025;13:e67721) doi:10.2196/67721

KEYWORDS

temperature; heat; adaptation behavior; resting heart rate; wearable device; modification effect; mobile phone

XSL•FO
RenderX

Introduction

Many epidemiological studies have demonstrated the impact of extreme ambient heat exposure on all-cause and cardiovascular mortality [1-3]. The specific causes of cardiovascular mortality and morbidity include heart failure, stroke, ischemic heart disease, hypertensive heart disease, arrhythmias, and cardiac arrest [4,5]. In the context of global warming, the risk of heat-related cardiovascular mortality is increasing [6,7]. Older adult individuals are more susceptible to heat-related deaths [5], likely due to impaired cardiopulmonary reserve and thermoregulatory capability, as well as relative socioeconomic disadvantages [8,9].

On an individual level, identifying personal risk factors and implementing appropriate adaptive behaviors may help mitigate heat-related risks [8]. However, existing studies on the effectiveness of heat adaptation behaviors have notable limitations, including their reliance on self-reported symptoms, cross-sectional designs that limit causal inference, and inconsistent findings regarding the effectiveness of specific adaptation measures. For example, a study in Guangdong, China, demonstrated an association between drinking more water and wearing light clothing with a reduced risk of heatstroke [10]. In contrast, other studies have shown that the implementation of heat adaptation behaviors is linked to an elevated risk of heat-related illness or symptoms [11-14]. These inconsistencies highlight the need for more robust, longitudinal studies.

Most current studies depend on self-reported symptoms and behaviors, introducing potential reporting bias and reverse causality issues where individuals experiencing symptoms may be more likely to adopt adaptation behaviors. Our previous research has addressed some of these limitations by using effect modification analysis to demonstrate the benefits of heat adaptation behaviors in reducing heat-related symptoms associated with obesity [14]. Given that symptom reporting is inherently subjective, using objective physiological indicators remains critical to comprehensively assess the health impacts of heat exposure and validate the benefits of heat adaptation behaviors.

Heart rate plays a central role in thermoregulatory response and is affected by autonomic nerve and neurohormone regulations [15]. Resting heart rate (RHR) has been known as a significant predictor of mortality and morbidity [15]. According to a meta-analysis of 87 prospective studies, an increase of 10 beats per minute in baseline RHR is linked to a 15% rise in cardiovascular mortality and a 17% rise in all-cause mortality [16]. This suggests that even a single beat per minute increase in RHR is not clinically negligible.

However, RHR measurements can vary significantly under different conditions (eg, higher during the daytime than at nighttime and higher in a hospital setting than at home), which limits its clinical application and interpersonal comparison [17]. This challenge has been addressed by the increasing prevalence of wearable devices. Wearable devices using photoplethysmography sensors to measure heart rate usually provide good precision, especially in resting conditions [18-20]. Nevertheless, these devices present several limitations in real-world applications, including motion artifacts during physical activity, reduced signal accuracy in individuals with darker skin pigmentation or tattoos, variability caused by inconsistent wrist positioning and contact pressure, interference from ambient light and low temperatures, as well as interdevice variability in measurement precision [21-23]. Despite these limitations, wearable technology continues to serve as a valuable tool for objective, long-term ambulatory heart rate monitoring in nonclinical settings.

The objectives of this study are twofold—first, to investigate the impact of summer temperatures on the RHR of older adult individuals using wearable devices and time-series data analysis; and second, to evaluate the benefits of personal heat adaptation behaviors by examining the effect modification of heart rate. By integrating physiological data from wearable technology with behavioral adaptation patterns, this research aims to provide evidence-based recommendations for personal heat adaptation strategies. The findings can inform public health interventions targeting heat-vulnerable populations, particularly older adults, and support the development of personalized guidance to reduce heat-related cardiovascular risks in aging societies under climate change.

Methods

Study Design and Participants

This investigation was devised as a prospective, longitudinal cohort study aimed at examining the older adult population residing in the Taipei Basin, a region in northern Taiwan characterized by a humid subtropical climate (Köppen climate classification: Cfa). Between 1987 and 2009, Taipei City noted an average temperature increase of 1.57 °C per century, surpassing the global average warming rate of 0.6 °C per century, as indicated in the National Oceanic and Atmospheric Administration's 2023 Annual Climate Report [24,25]. The Taipei Basin, which includes the densely populated cities of Taipei and New Taipei, is marked by a significant urban heat island effect due to its extensive urbanization. The Surface Urban Heat Island Index in Taipei during the summer reaches as high as 9 °C during the day and 3 °C at night, the highest among all cities in Taiwan [26].

Participants were recruited from an established older adult cohort, originally assembled between 2016 and 2018, to investigate the health impacts of air pollution [27]. The initial cohort was recruited from hospitals across 5 regions in Taiwan, where older adults attended government-funded health examinations. For this study, further enrollment took place between May and November 2020 [28]. Eligible participants were selected from the original cohort based on the following criteria: residence in the Taipei Basin, ownership of a smartphone compatible with the Garmin app, and the ability and willingness to wear a smartwatch consistently. Individuals with severe language, cognitive, or communication impairments were excluded. While leveraging an existing cohort provided a robust foundation, the requirement for smartphone ownership and wearable device usage may have introduced selection bias toward participants with higher socioeconomic status or greater technological proficiency, potentially limiting generalizability.

XSL•FO
RenderX

However, the sample reflects a growing segment of older adults who are increasingly adopting mobile health technologies. Of the 97 cohort members eligible for this study, 14 were lost to follow-up due to refusal to continuously wear the device, resulting in 83 participants who provided sufficient wearable data for analysis. The study size was determined by the number of eligible and consenting participants; no formal sample size calculation was performed due to the exploratory panel design, which used repeated daily measurements to enhance statistical power. From September to October 2021, the participants were interviewed via telephone to gather information on their personal background, health condition, and behaviors related to heat adaptation. To explore the impact of summer temperatures on heart rate, the study extracted heart rate data from May to September 2021 for statistical analysis.

Ethical Considerations

The study was conducted in accordance with the principles outlined in the Declaration of Helsinki and received approval from the Institutional Review Board of National Taiwan University Hospital (IRB 202003049RINB). All participants were recruited voluntarily and provided written informed consent before their participation. They were informed of their right to withdraw from the study at any time without consequence. No financial compensation was provided for participation. All collected data were deidentified before analysis to ensure participant privacy and confidentiality.

Questionnaire Assessment

A structured questionnaire was developed to gather data on various aspects, including personal habits (such as cigarette smoking and alcohol consumption), comorbidities (like diagnosed diabetes mellitus, hypertension, and heart diseases), educational levels, and adaptation strategies for dealing with heat and its related symptoms. Telephone interviews were conducted by an experienced interviewer who had previously administered face-to-face interviews with the participating older adult individuals using the same questionnaire [14]. A specific question, “What was the method from the following list that you used to adapt to heat during the extremely hot days in the past year?” was used to identify 15 adaptation behaviors.

For behaviors involving the use of air conditioners (ACs) or electric fans, participants were asked to report whether they used these devices during specific time periods of the day (morning, noon, afternoon, before sleep, and during sleep) and to indicate usage if it typically totaled more than 1 hour within each period, including intermittent use.

As the data relied on self-reported behaviors, particularly over a 2‐4 month recall period, responses may have been subject to recall bias. Evidence from recall error studies indicates that a 90-day recall period yields the lowest bias, compared with the tendency for overreporting with shorter recall periods (eg, 30 d) and underreporting with longer periods (eg, 365 d) [29]. To help mitigate this, interviews were conducted by a familiar interviewer using a standardized protocol to improve consistency and participant comfort. In future studies, the incorporation of real-time behavior logging through mobile apps may further improve data accuracy and reduce recall-related limitations.

RHR Measurement

Study participants were directed to wear a Garmin Vivosmart 4 smartwatch (Garmin International Inc) on their preferred wrist for continuous monitoring of their heart rate. At the initiation of the wearable device setup, participants’ height and weight were measured. The device uses photoplethysmography sensors, which detect changes in blood volume by measuring variations in light absorption through the skin. The heart rate data were uploaded to the Garmin Connect platform via a smartphone-internet connection and then retrieved from the National Taiwan University Medical Genie Platform [30], integrated via the Garmin Health application programming interface. Participants were instructed to wear the smartwatch consistently throughout the day and night, ensuring a minimum daily usage of 8 hours, with the daily RHR determined from the lowest 30-minute average within a 24-hour span [31].

In older adult individuals, the Garmin Vivosmart series has shown good accuracy for RHR monitoring, with a concordance correlation coefficient of 0.93 compared with electrocardiogram measurements from the Polar H7 chest-strapped heart rate monitor [32]. However, despite its strengths, photoplethysmography-based heart rate monitoring may be affected by age-related skin changes—such as wrinkles, roughness, pigmentation, dilated blood vessels, and reduced skin tone—which may alter light absorption and degrade signal quality [33,34]. In addition, vascular aging and increased arterial stiffness in older adults may influence pulse waveforms and reduce measurement precision [35]. These limitations should be considered when interpreting heart rate data in older populations.

Exposure Assessment

In this study, the hourly values of air temperature and relative humidity were obtained from the Environmental Protection Agency air pollution monitoring stations located closest to the participants’ residences. The monitoring stations, positioned 13 to 15 meters above ground level, used Metone 083R instruments to measure temperature and humidity with accuracies of 0.1 °C and ±1% at 23 °C, respectively. The distance between each participant’s residence and the nearest monitoring station ranged from 0.23 to 4.29 kilometers, with a mean distance of 1.74 kilometers. For analysis, air temperature and relative humidity measurements were averaged on a daily basis. The heat index was calculated using these daily average values based on the algorithm released by the US National Weather Service in 2011 [36], with the detailed algorithm and equation provided in Method S1 in Multimedia Appendix 1.

Statistical Analysis

A linear mixed-effect (LME) model was used to evaluate the association between personal characteristics, meteorological conditions, and RHR, incorporating random intercepts for individual participants to account for repeated measurements within participants. For the assessment of the impact of temperature and humidity on RHR, the LME model included adjustments for age (continuous variable), sex (male or female), hypertension (yes or no), diabetes (yes or no), heart diseases (yes or no), and educational attainment (<13 or ≥13 y). Days

XSL•FO
RenderX

with missing RHR data (11,016/12,699, 13.3%) due to nonwear were excluded from the analysis.

To assess the modifying effects of personal characteristics and adaptive behaviors, subgroup analysis was performed based on age (≤ or > median), sex (female or male), hypertension (no or yes), diabetes (no or yes), heart disease (no or yes), BMI (≤ or > median), educational attainment (<13 or ≥13 y), and each adaptive behavior (no or yes). The *P* value for the interaction term in the LME model was used to evaluate the statistical significance of effect modification.

To better evaluate the independent impact of each adaptive behavior on heart rate response to heat, we performed 2 mutually adjusted models for various heat adaptation behaviors. In the first model, AC and electric fan use across different time periods were mutually adjusted to identify which specific usage patterns had significant modifying effects. In the second model, the significant AC and electric fan usage behaviors identified in the first model were further adjusted alongside other adaptive behaviors that had shown significant modifying effects when tested individually.

Furthermore, a 3-way interaction term was applied in the LME model to examine whether the modifying effect of heat adaptation behaviors depends on personal characteristics.

To assess the robustness of the association between ambient temperature and RHR, we conducted a series of sensitivity analyses using progressively adjusted models. Model 1 was unadjusted (crude); model 2 adjusted for age, sex, and BMI; model 3 additionally included hypertension, diabetes, and heart disease; and model 4 (the final fully adjusted model) further accounted for educational attainment. All models were estimated using LME models with random intercepts for individuals. In addition, the aforementioned subgroup analyses were used to examine whether specific demographic or clinical subgroups exhibited differential RHR responses to heat exposure, further supporting the consistency and robustness of the observed associations. All analyses were executed using JMP 17 (JMP Statistical Discovery LLC), with all statistical tests being 2-sided and a *P* value of <.05 deemed statistically significant.

Results

Participant Characteristics

In this study, 83 older adult participants wore wearable devices throughout the warm season from May to October 2021, averaging 132.7 days of use and providing 11,016 RHR measurements. Details on their demographics, duration of device use, and heat adaptation methods are outlined in Table 1. The participants, with an average age of 73.7 (SD 3.6) years, had chronic disease rates of 15% (12/83) for diabetes, 42% (35/83) for hypertension, and 15% (12/83) for heart disease. These rates are slightly lower than previously reported national estimates among Taiwanese older adults, where the prevalence of hypertension and diabetes has been found to range from 48% to 60% and 20% to 24%, respectively [37,38]. This may suggest that our sample represents a relatively healthier segment of the older adult population, possibly because participants were initially recruited from individuals who attended older adult health check-up programs, reflecting greater health awareness and engagement.

They reported an average of 8.4 different heat adaptation behaviors, with the most prevalent including wearing light clothes (79/83, 95%), using AC (78/83, 94%) and electric fans (77/83, 93%), and using shade strategies such as umbrellas or hats (67/83, 81%), and staying indoors (60/83, 72%). Drinking more water (58/83, 70%) and increasing bathing or showering frequency (50/83, 60%), along with opening windows (49/83), were also common practices.

XSL•FO
RenderX

Table . Distribution of characteristics and wearable device indicators among 83 participants.

Characteristics and wearable device indicators	Value
Age (y)	
Mean (SD)	73.7 (3.6)
Median (IQR)	73.3 (70.1-75.0)
Sex (male), n (%)	39 (40)
Hypertension, n (%)	35 (42)
Diabetes, n (%)	12 (15)
Heart diseases, n (%)	12 (15)
BMI	
Mean (SD)	23.7 (2.5)
Median (IQR)	23.4 (21.6-25.6)
Education <13 years, n (%)	16 (19)
Adaptation behaviors for heat	
Use electric fans, n (%)	77 (93)
Morning	49 (59)
Noon	70 (84)
Afternoon	70 (84)
Night before sleep	70 (84)
Night during sleep	59 (71)
Wear light cloth, n (%)	79 (95)
Drink more water (extra ≥500 mL), n (%)	58 (70)
Use air conditioner, n (%)	78 (94)
Morning	7 (8)
Noon	37 (45)
Afternoon	35 (42)
Night before sleep	38 (46)
Night during sleep	56 (68)
Use public air conditioner, n (%)	22 (27)
Days per week using air conditioner	
Mean (SD)	5.6 (2.1)
Median (IQR)	7 (5-7)
Use umbrella or hat, n (%)	67 (81)
Stay in the shade, n (%)	61 (74)
Open window, n (%)	49 (59)
Increase frequency of bathing or shower, n (%)	50 (60)
Heated water for showering, n (%)	47 (57)
Stay indoors, n (%)	60 (72)
Reduce physical activity, n (%)	34 (41)
Drink cold water or beverage, n (%)	18 (22)
Eat more fruit, n (%)	35 (42)
Increase time for nap or sleep, n (%)	5 (6)
Number of adaptive behaviors	
Mean (SD)	8.4 (1.6)

XSL•FO
RenderX

Characteristics and wearable device indicators	Value
Median (IQR)	8 (7-10)
Duration of wearable device (d)	
Mean (SD)	132.7 (36.9)
Median (IQR)	153 (132-153)
Daily resting heart rate (beats/min)	
Mean (SD)	58.4 (7.1)
Median (IQR)	59 (54-63)

Device Adherence and Missing Data

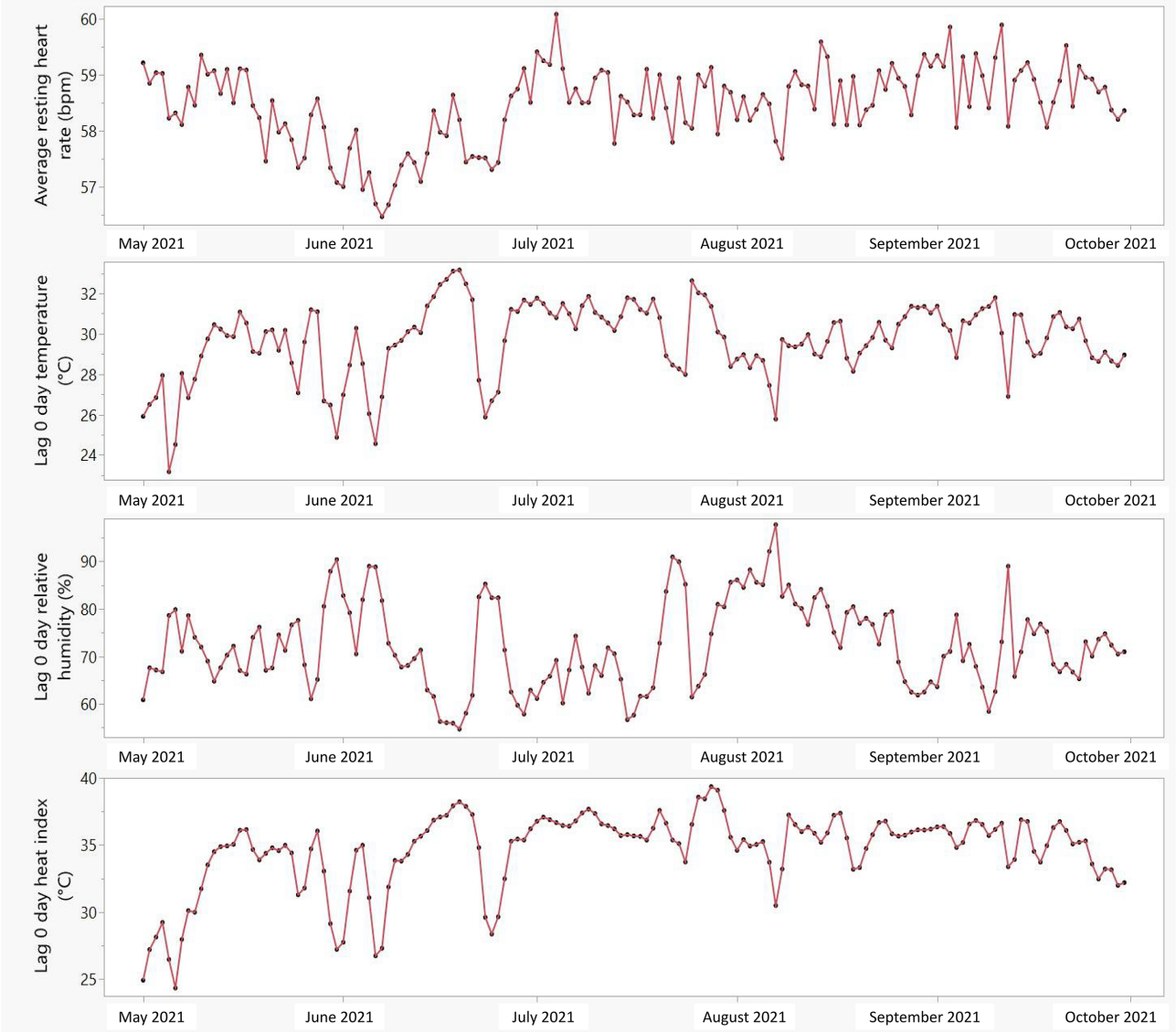
The number of participants providing RHR values each day is illustrated in Figure S1 in Multimedia Appendix 1. A gradual decline in device wear over the study period (from 83 to 63 participants) suggests potential adherence fatigue or reduced engagement. Missing RHR data were primarily due to smartwatch nonwear, which occurred more frequently on hotter days, likely due to discomfort or device removal during heat exposure (Figure S2 in Multimedia Appendix 1). Missingness was not associated with age (Figure S3 in Multimedia Appendix 1), but this pattern may have resulted in a mildly underrepresentation of high-temperature days.

Environmental Conditions

The distribution of temperature and humidity during the study period is summarized in Multimedia Appendix 2 and illustrated in Figure 1. The average temperature was 29.6 °C (SD 1.9 °C), with a mean humidity of 72.1% (SD 9.3%), and the heat index averaged 34.2 °C (SD 3.1 °C). Temperature and humidity were strongly inversely correlated (Spearman $\rho=-0.82$, $P<.001$), while temperature and heat index were positively correlated ($\rho=0.84$, $P<.001$).

XSL•FO
RenderX

Figure 1. Daily resting heart rate, temperature, humidity, and heat index during the study period.



Associations Between Individual Characteristics and RHR

Multimedia Appendix 3 outlines the results from univariate regression analyses linking individual traits with RHR. Diabetes was associated with a higher RHR, with a regression coefficient of 3.95 (95% CI 0.29 - 7.60; *P*=.04), while hypertension and heart diseases did not show significant effects. No significant correlations were found with age, sex, BMI, or education level. The morning use of electric fans initially correlated with higher heart rates (coefficient=2.76, 95% CI 0.14 - 5.37; *P*=.04), but this association became nonsignificant after adjusting for personal characteristics (data not shown). Other heat adaptation behaviors did not significantly affect RHR.

Associations Between Temperature and RHR

Table 2 shows that temperature and the heat index positively correlate with RHR on the same and following day, with effects diminishing and turning negative by Lag 4 and 5. Humidity negatively correlates with RHR initially, with significance lasting through Lag 2, but becomes nonsignificant thereafter. For example, each 1.9 °C increase in ambient temperature (1 SD) was associated with a 0.19 (95% CI 0.12-0.26; *P*<.001) beats per minute increase in RHR on Lag 0 . Model fit statistics favored ambient temperature over heat index as the exposure metric: both Corrected Akaike Information Criterion and Bayesian Information Criterion were consistently smaller for temperature-based models, supporting its use as the primary indicator of heat exposure in this study.

XSL•FO
RenderX

Table . Association between temperature and humidity and resting heart rate in various lag days. Coefficients were estimated for an SD increase in temperature (1.9 ˚C), relative humidity (9.3%), and heat index (3.1 ˚C) by mixed effect models with random intercept for individuals and with adjustments for age, sex, BMI, hypertension, diabetes, heart diseases, and education attainment.

Condition and lag day	Coefficient (95% CI)	*P* value
Air temperature		
Lag 0 day	0.19 (0.12 to 0.26)	<.001
Lag 1 day	0.19 (0.11 to 0.26)	<.001
Lag 2 day	0.08 (0.01 to 0.15)	.03
Lag 3 day	−0.03 (−0.10 to 0.04)	.43
Lag 4 day	−0.09 (−0.16 to −0.02)	.02
Lag 5 day	−0.09 (−0.16 to −0.02)	.02
Lag 6 day	−0.06 (−0.13 to 0.02)	.13
Lag 7 day	−0.07 (−0.14 to 0.01)	.08
Relative humidity		
Lag 0 day	−0.21 (−0.28 to −0.13)	<.001
Lag 1 day	−0.23 (−0.30 to −0.16)	<.001
Lag 2 day	−0.17 (−0.24 to −0.1)	<.001
Lag 3 day	−0.06 (−0.13 to 0.02)	.13
Lag 4 day	−0.003 (−0.08 to 0.07)	.94
Lag 5 day	−0.03 (−0.1 to 0.05)	.49
Lag 6 day	−0.05 (−0.12 to 0.03)	.21
Lag 7 day	−0.04 (−0.12 to 0.03)	.25
Heat index		
Lag 0 day	0.19 (0.12 to 0.27)	<.001
Lag 1 day	0.18 (0.11 to 0.25)	<.001
Lag 2 day	0.06 (−0.01 to 0.14)	.09
Lag 3 day	−0.03 (−0.11 to 0.04)	.41
Lag 4 day	−0.08 (−0.16 to −0.01)	.03
Lag 5 day	−0.08 (−0.16 to −0.01)	.03
Lag 6 day	−0.05 (−0.13 to 0.02)	.18
Lag 7 day	−0.06 (−0.14 to 0.01)	.09

Sensitivity Analyses

Sensitivity analyses using progressively adjusted models (Table S1 in Multimedia Appendix 1) demonstrated that the association between ambient temperature and RHR was consistent across all models, from the unadjusted model to the fully adjusted model, including age, sex, BMI, comorbidities, and educational attainment. This indicates that the observed associations are robust and not substantially influenced by confounding from the included covariates.

Effect Modification by Personal Characteristics and Heat Adaptation Behaviors

Subgroup analyses demonstrated that the association between ambient temperature and RHR varied significantly by personal characteristics and heat adaptation behaviors (Table 3). Sex modified the response, with males showing a greater RHR increase per 1 °C rise in temperature (coefficient=0.20, 95% CI 0.14 - 0.27) than females (coefficient=0.06, 95% CI 0.003 - 0.11; for interaction *P*<.001). A similar pattern was seen for BMI, where those above the median had stronger responses than those below (0.13 vs 0.10; for interaction *P*=.02). Participants without diabetes had a significant RHR increase (0.14 beats per minute [bpm]), while those with diabetes showed no clear change (–0.09 bpm; for interaction *P*<.001). Lower education levels were also associated with stronger RHR responses.

XSL•FO
RenderX

Table . Effect modification of personal conditions and adaptation behaviors on the relationship of lag 0 - 1 day temperature and resting heart rate. Subgroup analyses were done by mixed effect models with random intercept for individuals and with adjustments for age, sex, BMI, hypertension, diabetes, heart diseases, and education. Coefficients were estimated for 1℃ increase in temperature.

Conditions and adaptation behaviors	Coefficient (95% CI)	*P* value	*P* for interaction
Age (y)			
≤Median	0.10 (0.04 to 0.15)	<.001	.42[a]
>Median	0.13 (0.07 to 0.19)	<.001	.16[b]
Sex			
Male	0.20 (0.14 to 0.27)	<.001	<.001
Female	0.06 (0.003 to 0.11)	.04	—[c]
Hypertension			
No	0.12 (0.06 to 0.17)	<.001	.75
Yes	0.10 (0.05 to 0.16)	<.001	—
Diabetes			
No	0.14 (0.1 to 0.19)	<.001	<.001
Yes	−0.09 (−0.2 to 0.03)	.15	—
Heart diseases			
No	0.10 (0.05 to 0.14)	<.001	.14
Yes	0.19 (0.09 to 0.28)	<.001	—
BMI			
≤Median	0.10 (0.04 to 0.15)	<.001	.48[a]
>Median	0.13 (0.07 to 0.19)	<.001	.02[b]
Education <13 y			
No	0.14 (0.09 to 0.18)	<.001	.02
Yes	0.01 (−0.08 to 0.1)	.80	—
Adaptation behaviors for heat			
Use electric fans			
No	0.28 (0.10 to 0.46)	.002	.03
Yes	0.10 (0.06 to 0.14)	<.001	—
Morning			
No	0.10 (0.04 to 0.16)	<.001	.77
Yes	0.12 (0.06 to 0.17)	<.001	—
Noon			
No	0.09 (−0.02 to 0.19)	.10	.63
Yes	0.12 (0.07 to 0.16)	<.001	—
Afternoon			
No	0.07 (−0.04 to 0.19)	.20	.45
Yes	0.12 (0.08 to 0.16)	<.001	—
Night before sleep			
No	0.08 (−0.02 to 0.19)	.13	.55
Yes	0.12 (0.07 to 0.16)	<.001	—
Night during sleep			
No	0.20 (0.13 to 0.26)	<.001	.008
Yes	0.08 (0.03 to 0.13)	.003	—

XSL•FO
RenderX

Conditions and adaptation behaviors	Coefficient (95% CI)	*P* value	*P* for interaction
Wear light cloth			
No	0.19 (0.03 to 0.34)	.02	.39
Yes	0.11 (0.07 to 0.15)	<.001	—
Drink more water (extra ≥500 mL)			
No	0.24 (0.17 to 0.30)	<.001	<.001
Yes	0.05 (0.00 to 0.11)	.04	—
Use air conditioner			
No	0.33 (0.14 to 0.52)	<.001	.01
Yes	0.10 (0.06 to 0.14)	<.001	—
Morning			
No	0.13 (0.08 to 0.17)	<.001	.03
Yes	−0.03 (−0.15 to 0.09)	.59	—
Noon			
No	0.17 (0.11 to 0.22)	<.001	<.001
Yes	0.05 (−0.01 to 0.11)	.98	—
Afternoon			
No	0.15 (0.09 to 0.20)	<.001	.06
Yes	0.07 (0.01 to 0.13)	.03	—
Night before sleep			
No	0.20 (0.14 to 0.26)	<.001	<.001
Yes	0.01 (−0.05 to 0.06)	.84	—
Night during sleep			
No	0.19 (0.11 to 0.26)	<.001	.01
Yes	0.07 (0.03 to 0.12)	.002	—
Use public air conditioner			
No	0.16 (0.11 to 0.21)	<.001	<.001
Yes	−0.02 (−0.10 to 0.06)	.61	—
Days per week using air conditioner			
0-6 d	0.22 (0.15 to 0.28)	<.001	<.001[a]
7 d	0.04 (−0.01 to 0.09)	.14	<.001[b]
Use umbrella or hat			
No	0.03 (−0.06 to 0.12)	.48	.07
Yes	0.13 (0.08 to 0.18)	<.001	—
Stay in the shade			
No	0.06 (−0.04 to 0.16)	.22	.17
Yes	0.13 (0.09 to 0.17)	<.001	—
Open window			
No	−0.004 (−0.07 to 0.06)	.90	<.001
Yes	0.19 (0.14 to 0.24)	<.001	—
Increase frequency of bathing or shower			

XSL•FO
RenderX

Conditions and adaptation behaviors	Coefficient (95% CI)	*P* value	*P* for interaction
No	0.13 (0.06 to 0.19)	<.001	.58
Yes	0.10 (0.05 to 0.16)	.0002	—
Heated water for showering			
No	0.07 (0.00 to 0.14)	.04	.09
Yes	0.14 (0.09 to 0.19)	<.001	—
Stay indoors			
No	0.09 (0.01 to 0.16)	.02	.48
Yes	0.12 (0.07 to 0.17)	<.001	—
Reduce physical activity			
No	0.16 (0.11 to 0.22)	<.001	.005
Yes	0.04 (−0.02 to 0.11)	.15	—
Drink cold water or beverage			
No	0.16 (0.12 to 0.21)	<.001	<.001
Yes	−0.09 (−0.19 to 0.05)	.06	—
Eat more fruit			
No	0.16 (0.10 to 0.22)	<.001	.01
Yes	0.05 (−0.01 to 0.10)	.08	—
Increase time for nap or sleep			
No	0.13 (0.08 to 0.17)	<.001	.005
Yes	−0.12 (−0.27 to 0.02)	.09	—
Number of adaptive behaviors			
<8	0.21 (0.14 to 0.28)	<.001	<.001[a]
≥8	0.06 (0.01 to 0.11)	.03	<.001[b]

[a]Represented categorical variables in interaction terms.

[b]Represented continuous variables in interaction terms.

[c]Not available.

Several adaptation behaviors were linked to reduced RHR increases. Using AC (0.07 vs 0.19 bpm; *P*=.01), electric fans during sleep (0.08 vs 0.20 bpm; for interaction *P*=.008), and drinking extra water (0.05 vs 0.24 bpm; for interaction *P*<.001) were protective. Other beneficial behaviors included drinking cold beverages (−0.09 vs 0.16 bpm; for interaction *P*<.001), eating more fruit (0.05 vs 0.16 bpm; for interaction *P*=.01), and increasing sleep or nap time (−0.12 vs 0.13 bpm; for interaction *P*=.005). In contrast, opening windows was associated with greater RHR increases (0.19 bpm; *P*<.001). Notably, participants engaging in 8 or more adaptation behaviors had significantly lower RHR responses compared with those with fewer behaviors (0.06 vs 0.21 bpm; for interaction *P*<.001), suggesting a cumulative protective effect. These results highlight the modifying roles of individual health factors and behavioral adaptations in shaping physiological responses to heat.

Multivariate Models of Heat Adaptation Behaviors

To assess the independent effects of specific heat adaptation behaviors, we conducted 2 mutually adjusted models evaluating their interaction with temperature on RHR (Table 4). Model m1 examined the effect of AC and electric fan use across different times of day. Among these, using AC before sleep (coefficient=−0.19, 95% CI –0.29 to −0.08; *P*<.001) and during sleep (coefficient=−0.15, 95% CI −0.24 to −0.05; *P*=.002) were significantly associated with attenuated RHR responses to temperature. Electric fan use during sleep also demonstrated a protective effect (coefficient=−0.19, 95% CI −0.31 to −0.07; *P*=.002). Fan use during other time periods and daytime AC use did not show statistically significant effects.

XSL•FO
RenderX

Table . Regression coefficients of interaction terms with mutual adjustment of various heat adaptation behaviors to assess modification effects. Coefficients denote the regression coefficients of interaction terms between each adaptation behavior and temperature on resting heart rate. Coefficients were estimated for 1°C increase in temperature.

Adaptation behaviors	Coefficient (95% CI)	*P* value
Model m1: mutual adjustment of electric fan and AC[a] usage patterns over various time periods		
Use AC in the morning	−0.12 (−0.3 to 0.06)	.18
Use AC at noon	−0.03 (−0.13 to 0.06)	.50
Use AC in the afternoon	−0.04 (−0.14 to 0.06)	.41
Use AC before sleep	−0.19 (−0.29 to −0.08)	<.001
Use AC during sleep	−0.15 (−0.24 to −0.05)	.002
Use electric fans in the morning	0.03 (−0.07 to 0.13)	.57
Use electric fans at noon	0.02 (−0.18 to 0.23)	.83
Use electric fans in the afternoon	0.19 (−0.06 to 0.45)	.14
Use electric fans before sleep	−0.16 (−0.4 to 0.08)	.19
Use electric fans during sleep	−0.19 (−0.31 to −0.07)	.002
Model m2: mutual adjustment for significant adaptation behaviors		
Open window	0.05 (−0.06 to 0.15)	.39
Reduce physical activity	−0.15 (−0.23 to −0.06)	.001
Drink cold water or beverage	−0.24 (−0.35 to −0.13)	<.001
Increase naps or sleep	−0.28 (−0.46 to −0.1)	.003
Drink more water (extra ≥500 mL)	−0.10 (−0.19 to −0.01)	.02
Eat more fruit	−0.06 (−0.15 to 0.02)	.13
Use AC before sleep	−0.15 (−0.24 to −0.05)	.002
Use AC during sleep	−0.13 (−0.22 to −0.03)	.007
Use electric fans during sleep	−0.12 (−0.22 to −0.03)	.01

[a]AC: air conditioner.

Model m2 included multiple significant behaviors from previous subgroup analyses, allowing for mutual adjustment. The most protective effects were observed for increasing nap or sleep time (−0.28, 95% CI −0.46 to −0.10; *P*=.003), and drinking cold beverages (−0.24, 95% CI −0.35 to −0.13; *P*<.001). Other behaviors with significant effects included reducing physical activity (−0.15, 95% CI −0.23 to −0.06; *P*=.001), drinking extra water (−0.10, 95% CI −0.19 to −0.01; *P*=.02), and continued support for AC use before or during sleep and fan use during sleep.

In contrast, opening windows and eating more fruit did not significantly modify RHR response to temperature after adjustment for other behaviors. These findings highlight that certain behaviors, particularly those that directly cool the body or reduce exertion, have independent and additive effects in mitigating physiological stress during heat exposure.

Effect Modification of Heat Adaptation Behaviors by Participant Characteristics

Subgroup analyses revealed that the effectiveness of heat adaptation behaviors in mitigating RHR response to temperature varied significantly across individual characteristics (Table 5). Opening windows was associated with a greater increase in RHR among individuals with higher BMI compared with those with lower BMI (0.30 vs 0.10 bpm, for interaction *P*=.02), suggesting that this behavior may expose more heat-vulnerable individuals to excessive ambient temperatures.

Reducing physical activity was more protective in participants with heart disease (−0.33 vs −0.08 bpm, for interaction *P*=.03) and those with higher BMI (−0.28 vs 0.03 bpm, for interaction *P*<.001), indicating that limiting exertion may be particularly beneficial for those with cardiometabolic risks. Similarly, drinking cold beverages led to a stronger reduction in RHR among older adults (−0.41 vs −0.14 bpm, for interaction *P*=.02), males (−0.39 vs −0.16 bpm, for interaction *P*=.048), and individuals with diabetes (−0.46 vs −0.12 bpm, for interaction *P*=.007).

XSL•FO
RenderX

Table . Subgroup analysis by individual characteristics for the modification effects of several adaptation behaviors. Coefficients denote the regression coefficients of interaction terms between each adaptation behavior and temperature on resting heart rate. Coefficients were estimated for 1°C increase in temperature.

Characteristics	Coefficient (95% CI)	*P* for interaction[a]
Open window		
Age (y), <median	0.21 (0.10 to 0.32)	.64
Age (y), ≥median	0.17 (0.05 to 0.29)	—[b]
Sex, male	0.25 (0.12 to 0.38)	.29
Sex, female	0.16 (0.05 to 0.26)	—
HTN[c], no	0.13 (0.01 to 0.25)	.08
HTN, yes	0.28 (0.17 to 0.39)	—
DM[d], no	0.17 (0.08 to 0.26)	.34
DM, yes	0.28 (0.05 to 0.52)	—
Heart disease, no	0.21 (0.12 to 0.30)	.11
Heart disease, yes	−0.02 (−0.26 to 0.22)	—
BMI (kg/m^2), <median	0.10 (−0.02 to 0.22)	.02
BMI (kg/m^2), ≥median	0.30 (0.19 to 0.42)	—
Reduce physical activity		
Age (y), <median	−0.05 (−0.16 to 0.06)	.13
Age (y), ≥median	−0.18 (−0.30 to −0.06)	—
Sex, male	−0.18 (−0.31 to −0.05)	.17
Sex, female	−0.06 (−0.17 to 0.05)	—
HTN, no	−0.19 (−0.31 to −0.07)	.047
HTN, yes	−0.02 (−0.13 to 0.09)	—
DM, no	−0.11 (−0.20 to −0.03)	.71
DM, yes	−0.16 (−0.40 to 0.08)	—
Heart disease, no	−0.08 (−0.17 to 0.01)	.03
Heart disease, yes	−0.33 (−0.52 to −0.14)	—
BMI (kg/m^2), <median	0.03 (−0.08 to 0.14)	<.001
BMI (kg/m^2), ≥median	−0.28 (−0.41 to −0.16)	—
Drink cold water or beverage		
Age (y), <median	−0.14 (−0.27 to −0.02)	.02
Age (y), ≥median	−0.41 (−0.58 to −0.24)	—
Sex, male	−0.39 (−0.56 to −0.23)	.04
Sex, female	−0.16 (−0.29 to −0.04)	—
HTN, no	−0.32 (−0.46 to −0.18)	.12
HTN, yes	−0.16 (−0.30 to −0.01)	—
DM, no	−0.12 (−0.24 to 0.01)	.007
DM, yes	−0.46 (−0.70 to −0.23)	—
Heart disease, no	−0.25 (−0.35 to −0.14)	—
Heart disease, yes	—	—
BMI (kg/m^2), <median	−0.35 (−0.49 to −0.21)	.06
BMI (kg/m^2), ≥median	−0.15 (−0.30 to 0.00)	—

XSL•FO
RenderX

Characteristics	Coefficient (95% CI)	*P* for interaction[a]
Increase naps or sleep		
Age (y), <median	−0.17 (−0.42 to 0.09)	.39
Age (y), ≥median	−0.32 (−0.55 to −0.08)	—
Sex, male	0.03 (−0.24 to 0.29)	.01
Sex, female	−0.43 (−0.65 to −0.20)	—
HTN, no	−0.05 (−0.26 to 0.16)	<.001
HTN, yes	−1.06 (−1.40 to −0.72)	—
DM, no	−0.29 (−0.46 to −0.12)	—
DM, yes	—	—
Heart disease, no	−0.37 (−0.58 to −0.17)	.03
Heart disease, yes	0.07 (−0.23 to 0.37)	—
BMI (kg/m^2), <median	−0.32 (−0.51 to −0.13)	.08
BMI (kg/m^2), ≥median	0.07 (−0.33 to 0.48)	—
Drink more water (extra ≥500 mL)		
Age (y), <median	−0.26 (−0.39 to −0.12)	.15
Age (y), ≥median	−0.13 (−0.25 to −0.01)	—
Sex, male	−0.09 (−0.24 to 0.05)	.09
Sex, female	−0.25 (−0.36 to −0.14)	—
HTN, no	−0.23 (−0.35 to −0.11)	.17
HTN, yes	−0.10 (−0.24 to 0.03)	—
DM, no	−0.14 (−0.23 to −0.05)	.08
DM, yes	−0.48 (−0.88 to −0.08)	—
Heart disease, no	−0.23 (−0.33 to −0.14)	.007
Heart disease, yes	0.10 (−0.09 to 0.29)	—
BMI (kg/m^2), <median	−0.31 (−0.43 to −0.20)	.002
BMI (kg/m^2), ≥median	−0.03 (−0.17 to 0.12)	—
Eat more fruit		
Age (y), <median	−0.18 (−0.29 to −0.07)	.06
Age (y), ≥median	−0.02 (−0.14 to 0.10)	—
Sex, male	−0.09 (−0.22 to 0.04)	.84
Sex, female	−0.10 (−0.21 to 0.00)	—
HTN, no	−0.13 (−0.24 to −0.01)	.64
HTN, yes	−0.09 (−0.20 to 0.03)	—
DM, no	−0.17 (−0.26 to −0.09)	<.001
DM, yes	0.26 (0.02 to 0.50)	—
Heart disease, no	−0.11 (−0.20 to −0.02)	.62
Heart disease, yes	−0.05 (−0.25 to 0.14)	—
BMI (kg/m^2), <median	−0.04 (−0.15 to 0.07)	.10
BMI (kg/m^2), ≥median	−0.18 (−0.31 to −0.06)	—
Use AC before sleep		
Age (y), <median	−0.12 (−0.24 to −0.01)	.09

XSL•FO
RenderX

Characteristics	Coefficient (95% CI)	*P* for interaction[a]
Age (y), ≥median	−0.27 (−0.39 to −0.15)	—
Sex, male	−0.27 (−0.4 to −0.14)	.15
Sex, female	−0.15 (−0.25 to −0.04)	—
HTN, no	−0.25 (−0.36 to −0.14)	.11
HTN, yes	−0.12 (−0.23 to 0.00)	—
DM, no	−0.13 (−0.21 to −0.04)	.009
DM, yes	−0.47 (−0.73 to −0.21)	—
Heart disease, no	−0.20 (−0.29 to −0.11)	.82
Heart disease, yes	−0.17 (−0.36 to 0.02)	—
BMI <median	−0.19 (−0.31 to −0.07)	.64
BMI (kg/m^2), ≥median	−0.23 (−0.35 to −0.11)	—
Use AC during sleep		
Age (y), <median	−0.17 (−0.29 to −0.06)	.21
Age (y), ≥median	−0.06 (−0.19 to 0.07)	—
Sex, male	0.07 (−0.07 to 0.20)	.001
Sex, female	−0.22 (−0.34 to −0.11)	—
HTN, no	−0.14 (−0.26 to −0.02)	.44
HTN, yes	−0.07 (−0.19 to 0.04)	—
DM, no	−0.10 (−0.20 to −0.01)	.31
DM, yes	−0.23 (−0.47 to 0.01)	—
Heart disease, no	−0.10 (−0.20 to 0.00)	.86
Heart disease, yes	−0.12 (−0.31 to 0.06)	—
BMI (kg/m^2), <median	−0.13 (−0.25 to −0.01)	.67
BMI (kg/m^2), ≥median	−0.09 (−0.22 to 0.03)	—
Use electric fans during sleep		
Age (y), <median	0.09 (−0.03 to 0.22)	<.001
Age (y), ≥median	−0.30 (−0.42 to −0.17)	—
Sex, male	0.06 (−0.08 to 0.19)	.002
Sex, female	−0.23 (−0.35 to −0.11)	—
HTN, no	−0.25 (−0.37 to −0.12)	.002
HTN, yes	0.03 (−0.09 to 0.15)	—
DM, no	−0.08 (−0.17 to 0.01)	—
DM, yes	—	—
Heart disease, no	−0.13 (−0.24 to −0.03)	.335
Heart disease, yes	−0.02 (−0.20 to 0.17)	—
BMI (kg/m^2), <median	−0.25 (−0.37 to −0.14)	.002
BMI (kg/m^2), ≥median	0.04 (−0.10 to 0.19)	—

[a]The *P* value for the interaction refers to the effects of the 3-way interaction between personal characteristics, adaptation behaviors, and temperature on resting heart rate.

[b]Not applicable.

[c]HTN: hypertension.

[d]DM: diabetes mellitus.

XSL•FO
RenderX

Increasing nap or sleep time showed sex-specific and disease-related differences in effect. The behavior led to greater RHR reductions in females than males (−0.43 vs 0.03 bpm, for interaction *P*=.01), in participants with hypertension versus without hypertension (−1.06 vs −0.05 bpm, for interaction *P*<.001), and in those without heart disease compared with those with heart disease (−0.37 vs 0.07 bpm, for interaction *P*=.03). Drinking additional water (≥500 mL) was also more effective in participants with lower BMI (−0.31 vs −0.03 bpm, for interaction *P*=.002) and those without heart disease (−0.23 vs 0.10 bpm, for interaction *P*=.007).

Fruit consumption showed opposing effects between people with diabetes and without diabetes: while it was associated with a decrease in RHR among people without diabetes (−0.17 bpm), it was paradoxically linked to an increase among people with diabetes (0.26 bpm, for interaction *P*<.001). This finding suggests potential metabolic differences in behavioral responses.

The use of AC also showed heterogeneity. AC use before sleep was more protective among people with diabetes (–0.47 vs –0.13 bpm, *P*=.009), while AC use during sleep had a stronger effect in females than males (–0.22 vs 0.07 bpm, *P*=.001). Similarly, the use of electric fans during sleep was associated with a greater reduction in RHR among older adults (–0.30 vs 0.09 bpm, *P*<.001), females (–0.23 vs 0.06 bpm, *P*=.002), participants without hypertension (–0.25 vs 0.03 bpm, *P*=.002), and those with lower BMI (–0.25 vs 0.04 bpm, *P*=.002).

These results underscore that not all adaptation behaviors are equally effective across individuals. Tailoring behavioral recommendations based on age, sex, BMI, and chronic disease status may enhance the protective effects of personal strategies during heat exposure.

Discussion

Principal Findings and Personalized Heat Adaptation

The results of this longitudinal study highlight the significant influence of summer temperatures on the RHR of older adult individuals, with clear evidence demonstrating that heat adaptation behaviors can effectively mitigate these impacts. By leveraging wearable device technology, we captured objective physiological data, revealing lag and interaction patterns in heart rate responses to varying temperatures and humidity levels. Our findings underscore the importance of adaptive behaviors such as using AC, using electric fans during sleep, increasing water intake, reducing physical activity, increasing rest time, and drinking cold water, all of which were shown to have a protective effect on RHR. Notably, these adaptive strategies appeared to be more beneficial for certain subgroups: reducing physical activity for individuals with higher BMI, using electric fans during sleep and drinking cold water for the older age group, increasing rest time for participants with hypertension, and drinking cold water and using AC before sleep for participants with diabetes. However, eating more fruit may not be an ideal heat adaptation behavior for patients with diabetes. These results emphasize the need for personalized heat adaptation recommendations.

Comparison With Previous Research and Mechanistic Insights

The observed RHR response to summer temperature (0.1 beats/min per 1 °C) was consistent with findings from a large cross-sectional study in China (0.133 beats/min per 1 °C) [39]. Heart rate can serve as an early indicator of heat effects, as the increase in heart rate occurs before the rise in core body temperature at the inflection point of heat response [40]. The physiological mechanism involves an increase in heart rate due to heat stimulating the sympathetic nervous system, while stroke volume is limited due to reduced venous return, as blood pools in the skin vessels to dissipate more heat [41,42]. Another follow-up study showed that temperature increases during the warm season decrease parasympathetic modulation of heart rate and modify baroreflex sensitivity in as little as 6 hours [43]. The immediate response to heat aligns with our findings of peak RHR response to heat on lag 0 - 1 days.

Interpretation of Humidity-Related Effects

In this study, humidity exhibited a negative correlation with both temperature and RHR. This phenomenon may be related to the frequent afternoon rainfall during summer in Taiwan. In addition to increasing humidity, the evaporation process during rainfall absorbs a significant amount of heat energy, leading to a decrease in air temperature [44]. Furthermore, the cloud cover before and after the rainfall blocks sunlight, reducing solar radiation reaching the ground and further lowering the temperature [45]. Solar radiation significantly influences human thermal sensation by elevating skin temperature and perceived heat load, thereby reducing self-regulated outdoor exercise intensity [46-48]. Although high humidity impairs the body's ability to dissipate heat through sweat evaporation, thereby increasing heat strain and reducing self-paced exercise performance under hot conditions [49,50], the observed negative association between humidity and RHR may reflect its collinear relationship with concurrent reductions in temperature and solar radiation exposure.

Behavioral Adaptation Versus Physiological Response

Previous research [10-13], including our own [14], has shown a significant positive correlation between personal heat adaptation behaviors and self-reported heat-related symptoms. However, in this study, no direct significant correlation was found between personal heat adaptation behaviors and RHR. This suggests that individuals adjust their behaviors in response to perceived heat-related symptoms but are unable to perceive cardiovascular impacts in advance to take preventive measures.

Through the analysis of effect modification, we can clearly demonstrate the impact of heat adaptation behaviors on heart rate response to heat. In our previous research on heat-related symptoms, significant effects were only observed when combining 2 adaptive behaviors [14]. However, for RHR, a single adaptive behavior alone has an independently significant modifying effect, demonstrating beneficial effects on physiological response.

Effectiveness of AC, Especially at Night

The benefits of AC stem from maintaining a cooler, more stable indoor environment, thereby reducing thermal stress. Studies

XSL•FO
RenderX

from various countries, including the United States [51,52], Spain [52,53], China [54], Canada [52], and Japan [52,55], have shown that AC use can lower heat-related mortality. A randomized controlled trial demonstrated that providing targeted heat health messages increased AC use and reduced the risk of self-reported heat stress among older adults [56]. Our study extends these findings by demonstrating the specific impact of AC use on an objective physiological measure—RHR—highlighting its potential role in preventing heat-related cardiovascular events. Furthermore, the stronger protective effect of AC use at night highlights the significance of nighttime heat, which aligns with previous literature showing the impact of nighttime warming on mortality, especially under the urban heat island effect [57]. For patients with diabetes, nighttime AC is even more beneficial in lowering RHR, likely due to their increased vulnerability to heat exposure caused by impaired thermoregulatory and autonomic nerve responses [58].

Protective Role and Limitations of Electric Fan Use

Electric fans enhance convective and evaporative heat loss from the skin, significantly improving thermal comfort while consuming up to 50 times less electricity than AC [59]. However, at extremely high temperatures, such as around 36 °C with 80% relative humidity or 42 °C with 50% relative humidity, electric fans can exacerbate physiological heat strain [60]. In addition, the cooling efficacy diminishes with age and in conditions that impair sweating efficiency [61]. According to biophysiological modeling, the recommended safe temperature thresholds for electric fan use are 39 °C for healthy adults aged 18 - 40 years, 38 °C for those older than 65 years, and 37 °C for older adults on anticholinergic medications [59]. During the study period, nighttime temperatures ranged between 28 °C and 32 °C (average 28 °C), which is below the recommended safe temperature thresholds for electric fan use. This may explain the observed protective effects of nighttime electric fan use for older adults.

The protective effect of electric fan use during sleep on RHR was more pronounced in senior participants and women, potentially because these groups are more susceptible to heat-related hazards and have reduced sudomotor function [5,62]. Conversely, no protective effect was observed in individuals with higher levels of obesity, likely due to subcutaneous fat reducing the surface area-to-mass ratio, limiting heat exchange with the environment [63]. Furthermore, sweat gland density is lower in areas with higher adipose content, weakening heat dissipation through sweat evaporation [64]. Similarly, individuals with hypertension did not show a protective effect of electric fan use on RHR during sleep, unlike people without hypertension. This may be attributed to heightened sympathetic nervous system activity in people with hypertension [65], leading to skin vasoconstriction and diminished water vapor evaporation from the skin [66]. In addition, certain antihypertensive medications, such as beta-blockers, calcium-channel blockers, and diuretics, reduce heart rate and cardiac contractility, while also increasing the risk of dehydration [67].

Modifying Physical Activity and Sleep During Heat Exposure

Reducing physical activity lowers the RHR response to heat exposure because physical exertion increases metabolic heat production. Combined with environmental heat, this exacerbates heat stress [8]. The protective effect of reduced physical activity is likely more pronounced in individuals with heart disease and those with a higher BMI, as the added internal heat may increase strain on their cardiovascular system. In these populations, compromised cardiac function and an already elevated cardiac workload make them more vulnerable to heat stress [68].

In addition to the protective benefit of reducing physical activity during heatwaves, increasing naps or sleep also demonstrated independent protective effects, likely due to the enhancement of autonomic balance through increased parasympathetic activity and reduced sympathetic activity associated with longer daytime naps [69] and extended nighttime sleep [70]. The stronger effect observed in females may be linked to generally lower physical fitness levels and reduced thermoregulatory capacity [71], while the heightened response in individuals with hypertension could be attributed to elevated sympathetic nerve activity [65]. The observation that individuals with heart disease experienced less benefit from increased naps or sleep compared with those without heart disease may be explained by the high prevalence of sleep-disordered breathing in patients with heart conditions [72].

Effects of Cold Water Consumption and Hydration

Our study found that drinking cold water significantly reduced the RHR response to heat, with a more pronounced effect observed in males, those with diabetes, and older adults. This aligns with previous research, including a randomized crossover study demonstrating that cold and room-temperature water increase vagal activity, leading to a greater reduction in heart rate compared with body-temperature water [73]. A systematic review further supports these findings, noting that cold beverages help attenuate core temperature and improve performance in the heat [74]. The stronger effects observed in our subgroups may be explained by physiological differences: men have been shown to experience greater core temperature increases after repeated exertion [75] and stronger parasympathetic activation during cooling tasks [76]. People with diabetes and older adults often exhibit impaired thermoregulation [8], heightened sympathetic activity, and reduced parasympathetic tone [77,78], which may amplify the benefits of cold water on RHR. Further research is necessary to clarify the mechanisms behind these effects and their broader implications for mitigating heat-related health risks.

In our mutually adjusted model, both increased water intake and cold water exhibited independent protective effects. Hydration replenished fluids, maintaining blood volume, cardiac preload, and output, thereby enhancing skin blood flow, facilitating heat dissipation, and promoting the restoration of heart rate variability [79,80]. However, participants with heart disease and higher BMI showed a blunted response to increased water consumption. In patients with heart disease, maintaining fluid balance can be challenging, and the ability to augment cardiac output is less efficient through fluid supplement,

XSL•FO
RenderX

potentially due to the limitations imposed by the Frank-Starling mechanism [81]. Similarly, increased BMI is associated with impaired left ventricular diastolic function, which reduces the effectiveness of fluid supplementation on improving cardiac function [82].

Reconsidering the Role of Window Opening

Opening windows during heatwaves was a common adaptation in our study, with 59% (49/83) of participants reporting this behavior. However, we found that opening windows increased the RHR response to heat. This finding aligns with our previous research, which showed that this behavior intensified the association between body fat and heat-related symptoms (not opening windows: odds ratio [OR] 0.91, 95% CI 0.07-11.12; opening windows: OR 1.71, 95% CI 1.06-1.29; interaction *P*=.19), suggesting a potentially harmful effect. On hot days, opening windows can increase the inflow of hot air, raising indoor temperatures more effectively than heat conducted through the building [14]. In the mutually adjusted model, the effect of window opening became nonsignificant, likely due to confounding effects from other adaptation behaviors, such as the tendency to close windows when using AC [83]. In summary, although opening windows is a commonly used adaptation during hot weather, it does not appear to reduce heat-related health risks.

Fruit Consumption and Heat Response in Patients With Diabetes

In our study, the heat adaptation behavior of consuming more fruits among patients with diabetes was found to increase the RHR response to heat. A meta-analysis of 19 randomized controlled trials has shown that, without exceeding total energy intake, increased fresh and dried fruit consumption can aid in glycemic control [84]. However, whether the observed phenomenon in our study is due to excessive intake remains uncertain. A limitation of this study is the lack of information on individual total caloric intake, as well as whether the fruits were consumed as whole fruit or fruit juice. Taiwan's summer is characterized by an abundance of high-sugar fruits such as watermelon, passion fruit, pineapple, mango, longan, and sugar apple. Given that increased fruit consumption is a common adaptation behavior (reported by 35/83, 42% of participants), further research is warranted to clarify the impact of this dietary behavior on heat-related health risks in patients.

Strengths

This study has several strengths. First, continuous measurement of RHR via a smartwatch provides an objective assessment of physiological responses to heat. Second, compared with studies focusing on subjective heat-related symptoms [14], integrating RHR with adaptation behaviors minimizes the risk of reverse causality, as individuals generally do not consciously perceive their heart rate response to heat, which might otherwise influence their behavior. Third, the use of moderation and mutually adjusted statistical analysis strategies allowed us to delineate the independent health benefits of different heat adaptation behaviors.

Limitations

Several limitations should be acknowledged. First, the relatively small sample size of 83 participants, all from the Taipei Metropolitan Area, may limit the generalizability of our findings to the broader older adult population, particularly those in different geographic regions or climates. However, our study may serve as a model for assessing the health impacts of summer heat and developing region-specific personal adaptation strategies. Second, the collection of heat adaptation behavior data through telephone interviews conducted in September introduces the potential for recall bias, as participants were asked to recall behaviors from the preceding months. This reliance on self-reported data may also lead to inaccuracies in reporting. Third, the cross-sectional nature of the adaptation behavior data prevents the assessment of changes in behaviors over time or in response to specific heat events. Fourth, smartwatch nonwear was more common on hotter days, possibly due to discomfort or device removal during heat exposure. This pattern of missingness may lead to underrepresentation of high-temperature days in the dataset and result in conservative estimates of the heat's effect on RHR. In addition, using meteorological data from monitoring stations may introduce some exposure misclassification, as it does not account for indoor conditions and outdoor activities that affect individual exposure. Wearable thermosensors could offer a future solution [18]. Nonetheless, such misclassifications are likely nondifferential and may result in an underestimation of the observed associations.

Conclusion

By integrating objective physiological monitoring with self-reported behavioral data, this study provides a novel contribution to the field of environmental health by moving beyond subjective symptom reporting to evaluate the real-time cardiovascular effects of heat and the modifying role of personal adaptation strategies. This combination enables more accurate assessments of heat stress responses and intervention effectiveness in older adult populations.

Our findings demonstrate that common heat adaptation behaviors can significantly reduce cardiovascular strain during hot weather, particularly in older adults. Based on subgroup-specific responses, several behaviors should be prioritized for health promotion: using AC (especially before and during sleep), drinking cold water or beverages, reducing physical activity during hot periods, and increasing rest or nap time. These targeted strategies may be especially effective for vulnerable subgroups, including individuals with diabetes, hypertension, higher BMI, or advanced age.

This information may guide public health education and risk communication campaigns aimed at older adult populations, particularly in regions experiencing intensified heat due to climate change. Future research using enhanced time-series tracking and wearable technologies may further refine personalized adaptation guidance and support climate-resilient health systems.

XSL•FO
RenderX

Acknowledgments

The authors thank the field workers for their assistance in the data collection, as well as the participants of this study. This study was supported by grants from the National Health Research Institute (Taiwan) (Projects: EM-109-GP-08), and partially by the National Science and Technology Council (Taiwan) (NSTC 113-2314-B-002-180 and NSTC 114-2314-B-002-104). The funding source had no involvement in the study design; data collection, analysis, and interpretation; manuscript writing; and decision to submit the article for publication. The authors confirm that no generative artificial intelligence tools were used to generate content, tables, figures, or perform data analysis.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: CHC, YLG
Formal analysis: CHC, YLC
Funding acquisition: FL, YLG
Investigation: CHC
Methodology: CHC, FL, YLG
Project administration: CHC
Resources: FL
Software (National Taiwan University Medical Genie Platform): FL
Supervision: YLG
Writing - original draft: CHC, Y-LC
All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Heat index formula; distribution of weather conditions; univariate regression coefficients to resting heart rate; sensitivity analysis; and participation rate by age, time, and temperature.
[DOCX File, 122 KB - mhealth_v13i1e67721_app1.docx]

Multimedia Appendix 2
Distribution of temperature and humidity during the study period.
[DOCX File, 14 KB - mhealth_v13i1e67721_app2.docx]

Multimedia Appendix 3
Univariate regression coefficients to resting heart rate.
[DOCX File, 17 KB - mhealth_v13i1e67721_app3.docx]

References

1. Gasparrini A, Guo Y, Hashizume M, et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. Lancet 2015 Jul 25;386(9991):369-375. [doi: 10.1016/S0140-6736(14)62114-0] [Medline: 26003380]
2. Guo Y, Gasparrini A, Armstrong BG, et al. Heat wave and mortality: a multicountry, multicommunity study. Environ Health Perspect 2017 Aug 10;125(8):087006. [doi: 10.1289/EHP1026] [Medline: 28886602]
3. Burkart KG, Brauer M, Aravkin AY, et al. Estimating the cause-specific relative risks of non-optimal temperature on daily mortality: a two-part modelling approach applied to the Global Burden of Disease Study. Lancet 2021 Aug 21;398(10301):685-697. [doi: 10.1016/S0140-6736(21)01700-1] [Medline: 34419204]
4. Alahmad B, Khraishah H, Royé D, et al. Associations between extreme temperatures and cardiovascular cause-specific mortality: results from 27 countries. Circulation 2023 Jan 3;147(1):35-46. [doi: 10.1161/CIRCULATIONAHA.122.061832] [Medline: 36503273]
5. Liu J, Varghese BM, Hansen A, et al. Heat exposure and cardiovascular health outcomes: a systematic review and meta-analysis. Lancet Planet Health 2022 Jun;6(6):e484-e495. [doi: 10.1016/S2542-5196(22)00117-6] [Medline: 35709806]
6. Vicedo-Cabrera AM, Guo Y, Sera F, et al. Temperature-related mortality impacts under and beyond Paris Agreement climate change scenarios. Clim Change 2018 Oct;150(3-4):391-402. [doi: 10.1007/s10584-018-2274-3] [Medline: 30405277]

XSL•FO
RenderX

7. Gu S, Zhang L, Sun S, et al. Projections of temperature-related cause-specific mortality under climate change scenarios in a coastal city of China. Environ Int 2020 Oct;143(105889):105889. [doi: 10.1016/j.envint.2020.105889] [Medline: 32619913]
8. Epstein Y, Yanovich R. Heatstroke. N Engl J Med 2019 Jun 20;380(25):2449-2459. [doi: 10.1056/NEJMra1810762] [Medline: 31216400]
9. Worfolk JB. Heat waves: their impact on the health of elders. Geriatr Nurs 2000;21(2):70-77. [doi: 10.1067/mgn.2000.107131] [Medline: 10769330]
10. Liu T, Xu YJ, Zhang YH, et al. Associations between risk perception, spontaneous adaptation behavior to heat waves and heatstroke in Guangdong province, China. BMC Public Health 2013 Oct 2;13(1):913. [doi: 10.1186/1471-2458-13-913] [Medline: 24088302]
11. Hass AL, Ellis KN. Motivation for heat adaption: how perception and exposure affect individual behaviors during hot weather in Knoxville, Tennessee. Atmosphere (Basel) 2019;10(10):591. [doi: 10.3390/atmos10100591]
12. Kemen J, Schäffer-Gemein S, Grünewald J, Kistemann T. Heat perception and coping strategies: a structured interview-based study of elderly people in Cologne, Germany. Int J Environ Res Public Health 2021 Jul 14;18(14):7495. [doi: 10.3390/ijerph18147495] [Medline: 34299944]
13. Shih WY, Lung SC, Hu SC. Perceived heat impacts and adaptive behaviours in different socio-demographic groups in the subtropics. Int J Disaster Risk Reduct 2022 Mar;71:102799. [doi: 10.1016/j.ijdrr.2022.102799]
14. Chen CH, Lin WY, Lee KY, Sun WJ, Huang LY, Guo YLL. Adaptation behaviors modify the effects of body fat on heat-related symptoms among Taiwanese elderly. Int J Hyg Environ Health 2024 Jan;255:114296. [doi: 10.1016/j.ijheh.2023.114296] [Medline: 37979230]
15. Olshansky B, Ricci F, Fedorowski A. Importance of resting heart rate. Trends Cardiovasc Med 2023 Nov;33(8):502-515. [doi: 10.1016/j.tcm.2022.05.006] [Medline: 35623552]
16. Aune D, Sen A, ó'Hartaigh B, et al. Resting heart rate and the risk of cardiovascular disease, total cancer, and all-cause mortality - a systematic review and dose-response meta-analysis of prospective studies. Nutr Metab Cardiovasc Dis 2017 Jun;27(6):504-517. [doi: 10.1016/j.numecd.2017.04.004] [Medline: 28552551]
17. Lequeux B, Uzan C, Rehman MB. Does resting heart rate measured by the physician reflect the patient's true resting heart rate? White-coat heart rate. Indian Heart J 2018;70(1):93-98. [doi: 10.1016/j.ihj.2017.07.015] [Medline: 29455795]
18. Tsou MCM, Lung SCC, Cheng CH. Demonstrating the applicability of smartwatches in PM2.5 health impact assessment. Sensors (Basel) 2021 Jul 4;21(13):4585. [doi: 10.3390/s21134585] [Medline: 34283134]
19. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]
20. Shcherbina A, Mattsson CM, Waggott D, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):3. [doi: 10.3390/jpm7020003] [Medline: 28538708]
21. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3(18):18. [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
22. Koerber D, Khan S, Shamsheri T, Kirubarajan A, Mehta S. Accuracy of heart rate measurement with wrist-worn wearable devices in various skin tones: a systematic review. J Racial Ethn Health Disparities 2023 Dec;10(6):2676-2684. [doi: 10.1007/s40615-022-01446-9] [Medline: 36376641]
23. Fine J, Branan KL, Rodriguez AJ, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel) 2021 Apr 16;11(4):126. [doi: 10.3390/bios11040126] [Medline: 33923469]
24. Bai Y, Juang JY, Kondoh A. Urban warming and urban heat islands in Taipei, Taiwan. In: Groundwater and Subsurface Environments: Human Impacts in Asian Coastal Cities 2011:231-246. [doi: 10.1007/978-4-431-53904-9_12]
25. Lindsey R, Dahlman L. Climate change: global temperature. Climate.gov. 2025. URL: https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature [accessed 2025-10-22]
26. Liou YA, Tran DP, Nguyen KA. Spatio-temporal patterns and driving forces of surface urban heat island in Taiwan. Urban Climate 2024 Jan;53:101806. [doi: 10.1016/j.uclim.2024.101806]
27. Chen CH, Wu CD, Lee YL, et al. Air pollution enhance the progression of restrictive lung function impairment and diffusion capacity reduction: an elderly cohort study. Respir Res 2022 Jul 14;23(1):186. [doi: 10.1186/s12931-022-02107-5] [Medline: 35836168]
28. Chen CH, Lai F, Huang LY, Guo YLL. Short- and medium-term cumulative effects of traffic-related air pollution on resting heart rate in the elderly: a wearable device study. Ecotoxicol Environ Saf 2024 Oct 15;285(117140):117140. [doi: 10.1016/j.ecoenv.2024.117140] [Medline: 39368154]
29. Kjellsson G, Clarke P, Gerdtham UG. Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. J Health Econ 2014 May;35(34-46):34-46. [doi: 10.1016/j.jhealeco.2014.01.007] [Medline: 24595066]
30. Wu CT, Li GH, Huang CT, et al. Acute exacerbation of a chronic obstructive pulmonary disease prediction system using wearable device data, machine learning, and deep learning: development and cohort study. JMIR Mhealth Uhealth 2021 May 6;9(5):e22591. [doi: 10.2196/22591] [Medline: 33955840]

XSL•FO
RenderX

31. Garmin support center. Garmin. 2024. URL: https://support.garmin.com/en-US/?faq=F8YKCB4CJd5PG0DR9ICV3A [accessed 2025-10-13]
32. Chow HW, Yang CC. Accuracy of optical heart rate sensing technology in wearable fitness trackers for young and older adults: validation and comparison study. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e14707. [doi: 10.2196/14707] [Medline: 32343255]
33. Delgado-Gonzalo R, Parak J, Tarniceriu A, Renevey P, Bertschi M, Korhonen I. Evaluation of accuracy and reliability of pulseon optical heart rate monitoring device. Presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 25-29, 2015. [doi: 10.1109/EMBC.2015.7318391]
34. McCullough JL, Kelly KM. Prevention and treatment of skin aging. Ann N Y Acad Sci 2006 May;1067(323-31):323-331. [doi: 10.1196/annals.1354.044] [Medline: 16804006]
35. Allen J, Murray A. Age-related changes in the characteristics of the photoplethysmographic pulse shape at various body sites. Physiol Meas 2003 May;24(2):297-307. [doi: 10.1088/0967-3334/24/2/306] [Medline: 12812416]
36. Anderson GB, Bell ML, Peng RD. Methods to calculate the heat index as an exposure metric in environmental health research. Environ Health Perspect 2013 Oct;121(10):1111-1119. [doi: 10.1289/ehp.1206273] [Medline: 23934704]
37. Pan WH, Chang HY, Yeh WT, Hsiao SY, Hung YT. Prevalence, awareness, treatment and control of hypertension in Taiwan: results of Nutrition and Health Survey in Taiwan (NAHSIT) 1993-1996. J Hum Hypertens 2001 Nov;15(11):793-798. [doi: 10.1038/sj.jhh.1001268] [Medline: 11687924]
38. Chang CH, Shau WY, Jiang YD, et al. Type 2 diabetes prevalence and incidence among adults in Taiwan during 1999-2004: a national health insurance data set study. Diabet Med 2010 Jun;27(6):636-643. [doi: 10.1111/j.1464-5491.2010.03007.x] [Medline: 20546280]
39. Madaniyazi L, Zhou Y, Li S, et al. Outdoor temperature, heart rate and blood pressure in Chinese adults: effect modification by individual characteristics. Sci Rep 2016 Feb 15;6:21003. [doi: 10.1038/srep21003] [Medline: 26876040]
40. Cottle RM, Fisher KG, Wolf ST, Kenney WL. Onset of cardiovascular drift during progressive heat stress in young adults (PSU HEAT project). J Appl Physiol 2023 Aug 1;135(2):292-299. [doi: 10.1152/japplphysiol.00222.2023]
41. Liu C, Yavar Z, Sun Q. Cardiovascular response to thermoregulatory challenges. Am J Physiol Heart Circ Physiol 2015 Dec 1;309(11):H1793-H1812. [doi: 10.1152/ajpheart.00199.2015] [Medline: 26432837]
42. No M, Kwak HB. Effects of environmental temperature on physiological responses during submaximal and maximal exercises in soccer players. Integr Med Res 2016 Sep;5(3):216-222. [doi: 10.1016/j.imr.2016.06.002] [Medline: 28462121]
43. Wasserman EB, Zareba W, Utell MJ, et al. Acute changes in ambient temperature are associated with adverse changes in cardiac rhythm. Air Qual Atmos Health 2014 Sep 1;7(3):357-367. [doi: 10.1007/s11869-014-0244-0] [Medline: 25368681]
44. Liu W, Dong S, Zheng J, et al. Quantifying the rainfall cooling effect: the importance of relative humidity in Guangdong, South China. Journal of Hydrometeorology 2022;23(6):875-889. [doi: 10.1175/JHM-D-21-0155.1]
45. Jadhav AV, Bhawar RL, Dumka UC, Rahul PRC, Kumar PP. Impacts of meteorological conditions on the plummeting surface-reaching solar radiation over a sub-tropical station – Pune, India. Energy Sustain Dev 2024 Jun;80:101444. [doi: 10.1016/j.esd.2024.101444]
46. Hodder SG, Parsons K. The effects of solar radiation on thermal comfort. Int J Biometeorol 2006 Dec 19;51(3):233-250. [doi: 10.1007/s00484-006-0050-y]
47. Otani H, Kaya M, Tamaki A, Goto H, Maughan RJ. Exposure to high solar radiation reduces self-regulated exercise intensity in the heat outdoors. Physiology & Behavior 2019 Feb;199:191-199. [doi: 10.1016/j.physbeh.2018.11.029]
48. Li G, Liu D, Hu A, et al. Effect of solar radiation on human thermal sensation and physiological parameters in a convection–radiation air conditioning environment. Build Simul 2024 Aug;17(8):1359-1377. [doi: 10.1007/s12273-024-1133-6]
49. Beigtan M, Gonçalves M, Weon BM. Heat transfer by sweat droplet evaporation. Environ Sci Technol 2024 Apr 16;58(15):6532-6539. [doi: 10.1021/acs.est.4c00850] [Medline: 38538556]
50. Bright FM, Clark B, Jay O, Périard JD. Elevated humidity impairs evaporative heat loss and self-paced exercise performance in the heat. Scand J Med Sci Sports 2025 Mar;35(3):e70041. [doi: 10.1111/sms.70041] [Medline: 40107869]
51. Eisenman DP, Wilhalme H, Tseng CH, et al. Heat death associations with the built environment, social vulnerability and their interactions with rising temperature. Health Place 2016 Sep;41(89-99):89-99. [doi: 10.1016/j.healthplace.2016.08.007] [Medline: 27583525]
52. Sera F, Hashizume M, Honda Y, et al. Air conditioning and heat-related mortality: a multi-country longitudinal study. Epidemiology (Sunnyvale) 2020 Nov;31(6):779-787. [doi: 10.1097/EDE.0000000000001241] [Medline: 33003149]
53. López-Bueno JA, Díaz J, Sánchez-Guevara C, et al. The impact of heat waves on daily mortality in districts in Madrid: the effect of sociodemographic factors. Environ Res 2020 Nov;190:109993. [doi: 10.1016/j.envres.2020.109993] [Medline: 32745539]
54. Li C, Gu H. Climate change and mortality evolution in China. J Environ Manage 2020 Aug 1;267(110622):110622. [doi: 10.1016/j.jenvman.2020.110622] [Medline: 32364132]
55. Chua PLC, Takane Y, Ng CFS, et al. Net impact of air conditioning on heat-related mortality in Japanese cities. Environ Int 2023 Nov;181:108310. [doi: 10.1016/j.envint.2023.108310] [Medline: 37951014]

XSL•FO
RenderX

56. Nitschke M, Krackowizer A, Hansen AL, Bi P, Tucker GR. Heat health messages: a randomized controlled trial of a preventative messages tool in the older population of South Australia. Int J Environ Res Public Health 2017 Aug 31;14(9):992. [doi: 10.3390/ijerph14090992] [Medline: 28858262]
57. He C, Kim H, Hashizume M, et al. The effects of night-time warming on mortality burden under future climate change scenarios: a modelling study. Lancet Planet Health 2022 Aug;6(8):e648-e657. [doi: 10.1016/S2542-5196(22)00139-5] [Medline: 35932785]
58. Vallianou NG, Geladari EV, Kounatidis D, et al. Diabetes mellitus in the era of climate change. Diabetes Metab 2021 Jul;47(4):101205. [doi: 10.1016/j.diabet.2020.10.003] [Medline: 33127474]
59. Morris NB, Chaseling GK, English T, et al. Electric fan use for cooling during hot weather: a biophysical modelling study. Lancet Planet Health 2021 Jun;5(6):e368-e377. [doi: 10.1016/S2542-5196(21)00136-4] [Medline: 34119011]
60. Ravanelli NM, Hodder SG, Havenith G, Jay O. Heart rate and body temperature responses to extreme heat and humidity with and without electric fans. JAMA 2015 Feb 17;313(7):724-725. [doi: 10.1001/jama.2015.153] [Medline: 25688786]
61. Gagnon D, Romero SA, Cramer MN, Jay O, Crandall CG. Cardiac and thermal strain of elderly adults exposed to extreme heat and humidity with and without electric fan use. JAMA 2016 Sep 6;316(9):989-991. [doi: 10.1001/jama.2016.10550] [Medline: 27599335]
62. D'Souza AW, Notley SR, Kenny GP. The relation between age and sex on whole-body heat loss during exercise-heat stress. Med Sci Sports Exerc 2020 Oct;52(10):2242-2249. [doi: 10.1249/MSS.0000000000002373] [Medline: 32496737]
63. Kenney WL. Physiological correlates of heat intolerance. Sports Med 1985;2(4):279-286. [doi: 10.2165/00007256-198502040-00005] [Medline: 2862686]
64. Bar-Or O, Magnusson LI, Buskirk ER. Distribution of heat-activated sweat glands in obese and lean men and women. Hum Biol 1968 May;40(2):235-248. [Medline: 5664189]
65. Valensi P. Autonomic nervous system activity changes in patients with hypertension and overweight: role and therapeutic implications. Cardiovasc Diabetol 2021 Aug 19;20(1):170. [doi: 10.1186/s12933-021-01356-w] [Medline: 34412646]
66. Kitada K, Nishiyama A. Potential role of the skin in hypertension risk through water conservation. Hypertension 2024 Mar;81(3):468-475. [doi: 10.1161/HYPERTENSIONAHA.123.20700] [Medline: 37942635]
67. Sorensen C, Hess J. Treatment and prevention of heat-related illness. N Engl J Med 2022 Oct 13;387(15):1404-1413. [doi: 10.1056/NEJMcp2210623] [Medline: 36170473]
68. Messerli FH. Cardiovascular effects of obesity and hypertension. Lancet 1982 May 22;1(8282):1165-1168. [doi: 10.1016/s0140-6736(82)92234-6] [Medline: 6122945]
69. AlQatari AA, Alturki JA, Abdulali KA, et al. Changes in heart rate variability and baroreflex sensitivity during daytime naps. Nat Sci Sleep 2020;12(661-9):661-669. [doi: 10.2147/NSS.S270191] [Medline: 33061723]
70. Whitehurst LN, Naji M, Mednick SC. Comparing the cardiac autonomic activity profile of daytime naps and nighttime sleep. Neurobiol Sleep Circadian Rhythms 2018 Jun;5:52-57. [doi: 10.1016/j.nbscr.2018.03.001] [Medline: 31236511]
71. Yanovich R, Ketko I, Charkoudian N. Sex differences in human thermoregulation: relevance for 2020 and beyond. Physiology (Bethesda) 2020 May 1;35(3):177-184. [doi: 10.1152/physiol.00035.2019] [Medline: 32293229]
72. Cowie MR, Linz D, Redline S, Somers VK, Simonds AK. Sleep disordered breathing and cardiovascular disease: JACC State-of-the-Art review. J Am Coll Cardiol 2021 Aug 10;78(6):608-624. [doi: 10.1016/j.jacc.2021.05.048] [Medline: 34353537]
73. Girona M, Grasser EK, Dulloo AG, Montani JP. Cardiovascular and metabolic responses to tap water ingestion in young humans: does the water temperature matter? Acta Physiol (Oxf) 2014 Jun;211(2):358-370. [doi: 10.1111/apha.12290] [Medline: 24684853]
74. Burdon CA, Johnson NA, Chapman PG, O'Connor HT. Influence of beverage temperature on palatability and fluid ingestion during endurance exercise: a systematic review. Int J Sport Nutr Exerc Metab 2012 Jun;22(3):199-211. [doi: 10.1123/ijsnem.22.3.199] [Medline: 22693241]
75. Anderson CAJ, Stewart IB, Stewart KL, Linnane DM, Patterson MJ, Hunt AP. Sex-based differences in body core temperature response across repeat work bouts in the heat. Appl Ergon 2022 Jan;98:103586. [doi: 10.1016/j.apergo.2021.103586]
76. Ruiz JM, Uchino BN, Smith TW. Hostility and sex differences in the magnitude, duration, and determinants of heart rate response to forehead cold pressor: parasympathetic aspects of risk. Int J Psychophysiol 2006 Jun;60(3):274-283. [doi: 10.1016/j.ijpsycho.2005.07.005] [Medline: 16125263]
77. Hadad R, Akobe SF, Weber P, et al. Parasympathetic tonus in type 2 diabetes and pre-diabetes and its clinical implications. Sci Rep 2022 Oct 26;12(1):18020. [doi: 10.1038/s41598-022-22675-2] [Medline: 36289393]
78. Giunta S, Xia S, Pelliccioni G, Olivieri F. Autonomic nervous system imbalance during aging contributes to impair endogenous anti-inflammaging strategies. Geroscience 2024 Feb;46(1):113-127. [doi: 10.1007/s11357-023-00947-7] [Medline: 37821752]
79. Trangmar SJ, González-Alonso J. Heat, hydration and the human brain, heart and skeletal muscles. Sports Med 2019 Feb;49(Suppl 1):69-85. [doi: 10.1007/s40279-018-1033-y] [Medline: 30671905]
80. Manning CN, Morrissey MC, Langan SP, et al. Fluid replacement strategies and heart rate variability recovery following prolonged exercise in the heat and mild dehydration. Physiologia 2023;3(1):98-112. [doi: 10.3390/physiologia3010008]
81. Delicce AV, Makaryus AN. StatPearls. In: Physiology, Frank Starling Law: StatPearls Publishing; 2024.

XSL•FO
RenderX

82. Russo C, Jin Z, Homma S, et al. Effect of obesity and overweight on left ventricular diastolic function: a community-based study in an elderly cohort. J Am Coll Cardiol 2011 Mar 22;57(12):1368-1374. [doi: 10.1016/j.jacc.2010.10.042] [Medline: 21414533]
83. Yan B, Long E, Meng X, Zhang Y, Hou D, Du X. Influence of user behavior on unsatisfactory indoor thermal environment. Energy Convers Manag 2014 Oct;86:1-7. [doi: 10.1016/j.enconman.2014.05.004]
84. Ren Y, Sun S, Su Y, Ying C, Luo H. Effect of fruit on glucose control in diabetes mellitus: a meta-analysis of nineteen randomized controlled trials. Front Endocrinol (Lausanne) 2023;14:1174545. [doi: 10.3389/fendo.2023.1174545] [Medline: 37214237]

Abbreviations:

AC: air conditioner
bpm: beats per minute
LME: linear mixed-effect
OR: odds ratio
RHR: resting heart rate

Edited by L Buis; submitted 19.10.24; peer-reviewed by H Otani, K Mouloudj; revised version received 17.06.25; accepted 25.09.25; published 14.11.25.

Please cite as:
Chen CH, Lai F, Chen YL, Guo YL
Effects of Heat Adaptation Behaviors on Resting Heart Rate Response to Summer Temperatures in Older Adults: Wearable Device Panel Study
JMIR Mhealth Uhealth 2025;13:e67721
URL: https://mhealth.jmir.org/2025/1/e67721
doi:10.2196/67721

© Chi-Hsien Chen, Feipei Lai, Yu-Lin Chen, Yue Leon Guo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Reliability and Accuracy of the Fitbit Charge 4 Photoplethysmography Heart Rate Sensor in Ecological Conditions: Validation Study

Maxime Ceugniez[1], MSc; Hervé Devanne[1], PhD; Eric Hermand[1,2], PhD

[1]ULR 7369 - URePSSS - Unité de Recherche Pluridisciplinaire Sport Santé Société, Univ. Littoral Côte d'Opale, Univ. Lille, Univ. Artois, 189b, Avenue Maurice Schumann, Centre Universitaire des Darses, Dunkerque, France

[2]UMR INSERM U1272 Hypoxie & Poumon, Département STAPS, Université Sorbonne Paris Nord, Bobigny, France

Corresponding Author:
Eric Hermand, PhD
ULR 7369 - URePSSS - Unité de Recherche Pluridisciplinaire Sport Santé Société, Univ. Littoral Côte d'Opale, Univ. Lille, Univ. Artois, 189b, Avenue Maurice Schumann, Centre Universitaire des Darses, Dunkerque, France

Abstract

Background: Wrist-worn photoplethysmography (PPG) sensors allow for continuous heart rate (HR) measurement without the inconveniences of wearing a chest belt. Although green light PPG technology reduces HR measurement motion artifacts, only a limited number of studies have investigated the reliability and accuracy of wearables in non–laboratory-controlled conditions with actual specific and various physical activity movements.

Objective: The purpose of this study was to (1) assess the reliability and accuracy of the PPG-based HR sensor of the Fitbit Charge 4 (FC4) in ecological conditions and (2) quantify the potential variability caused by the nature of activities.

Methods: We collected HR data from participants who performed badminton, tennis, orienteering running, running, cycling, and soccer while simultaneously wearing the FC4 and the Polar H10 chest belt (criterion sensor). Skin tone was assessed with the Fitzpatrick Skin Scale. Once data from the FC4 and criterion data were synchronized, accuracy and reliability analyses were performed, using intraclass correlation coefficients (ICCs), Lin concordance correlation coefficients (CCCs), mean absolute percentage errors (MAPEs), and Bland-Altman tests. A linear univariate model was also used to evaluate the effect of skin tone on bias. All analyses were stratified by activity and pooled activity types (racket sports and running sports).

Results: A total of 77.5 hours of HR recordings from 26 participants (age: mean 21.1, SD 5.8 years) were analyzed. The highest reliability was found for running sports, with ICCs and CCCs of 0.90 and 0.99 for running and 0.80 and 0.93 for orienteering running, respectively, whereas the ICCs and CCCs were 0.37 and 0.78, 0.42 and 0.88, 0.65 and 0.97, and 0.49 and 0.81 for badminton, tennis, cycling, and soccer, respectively. We found the highest accuracy for running (bias: 0.1 beats per minute [bpm]; MAPE 1.2%, SD 4.6%) and the lowest for badminton (bias: −16.5 bpm; MAPE 16.2%, SD 14.4%) and soccer (bias: −16.5 bpm; MAPE 17.5%, SD 20.8%). Limit of agreement (LOA) width and artifact rate followed the same trend. No effect of skin tone was observed on bias.

Conclusions: LOA width, bias, and MAPE results found for racket sports and soccer suggest a high sensitivity to motion artifacts for activities that involve "sharp" and random arm movements. In this study, we did not measure arm motion, which limits our results. However, whereas individuals might benefit from using the FC4 for casual training in aerobic sports, we cannot recommend the use of the FC4 for specific purposes requiring high reliability and accuracy, such as research purposes.

(JMIR Mhealth Uhealth 2025;13:e54871) doi:10.2196/54871

KEYWORDS

photoplethysmography; physical activity; ecological conditions; accuracy; reliability; Fitbit Charge 4; Fitbit; exercise; ecological; wrist-worn device; device; sensor; wearables; usefulness; variability; sensitivity; heart rate; heart rate sensor

Introduction

Over the recent years, connected bracelet and watch sales have been regularly increasing [1]. These tools allow individuals to monitor various active life parameters, such as time of activity, sleep quality, step numbers, and energy expenditure, among others. Most of them are computed through algorithms that use accelerometry and heart rate (HR) data. Although these new devices are used for training by both recreational athletes and elite athletes, chest belts remain the gold standard, especially among high-level athletes, for measuring HR, as they are based on R peak detection from the QRS electrocardiogram (ECG) complex [2,3]. In opposition to chest belts, other wearables perform additional measurements (eg, accelerometry and

XSL•FO
RenderX

positioning measurements via a GPS), which, when combined with HR data and algorithm processing, provide data on other parameters, such as energy expenditure and quality of sleep. Moreover, wrist-worn devices could reduce the tolerance and acceptability issues observed with chest belts [4]. Wrist-worn devices usually estimate continuous HR through the photoplethysmography (PPG) technique, which was first used in the late 1930s [5]. PPG involves measuring light absorption through tissues of interest [6]; red and infrared lights are emitted by an LED through the skin, and a photoreceptor captures the remaining emissions after tissue absorption [7]. However, even if the concept remains similar, connected watches usually come with a green light PPG sensor for its ability to reduce motion artifacts, contrary to the red ones commonly used in the medical field for blood oxygen saturation evaluation [8,9]. The reason for this is that the deeper the light penetrates the tissue (eg, red wavelength), the more the pulse wave is affected by limb movements [10,11]. As light penetration depends on light wavelength, the shorter wavelength of green light provides less information from deeper nonpulsatile tissues [9,12]. Considering this, green light is less prone to motion artifacts during normal daily life [13-15]. In the case of HR monitoring via watches or bracelets during activities, signal accuracy and reliability may vary according to numerous factors. Among them, gear placement on skin, strap tightening (which induces skin compression), skin tone, and activity type and intensity can affect HR recording [15-21]. PPG HR sensor accuracy has been investigated during physical activity across a spectrum of intensities. However, researchers tend to measure PPG HR sensor accuracy with treadmill running or cycling ergometers in laboratory-controlled conditions [17,22-24]. To our knowledge, only a few evaluations of connected device accuracy (eg, accuracy of a smartwatch, as evaluated in this paper) were performed in ecological conditions across different physical activity types. As this type of device is meant to be used in non–laboratory-controlled conditions or free-living conditions, this study aimed to evaluate the accuracy and validity of the PPG HR data from the Fitbit Charge 4 (FC4; Fitbit LLC) across multiple physical activity types. Therefore, the objectives of this study were to (1) assess the accuracy and reliability of FC4 HR measurement in ecological conditions and (2) quantify the potential impact of activity type on accuracy.

Methods

Participants

A total of 26 healthy young adults from the Sport Sciences University of Calais, France, who were practicing physical activities on a weekly basis, volunteered and were included in this study, which was advertised on the university campus and via social networks. No inclusion or exclusion criteria were used.

A minimum HR sample size was calculated with G*Power (version 3.1.9.6) [25] by using a significance level (α) of 5%, a statistical power of $1 - \beta = 80\%$, and an effect size of 0.075 (computed from the expected HR mean and SD). The number of necessary HR samples was below 6000, representing 100 minutes of recording (sample rate=1 s^{-1}). Measurements were performed during participants' regular training sessions for soccer, badminton, orienteering running, basketball, tennis, and road biking.

Ethical Considerations

This study was approved by the National Commission for Data Protection and Liberties (CNIL-France; registration number: 2224247). All participants gave their written informed consent, with the possibility to opt out of the protocol at any point. Data collected throughout the protocol were deidentified for privacy and confidentiality reasons. Finally, although participants could not be financially compensated for their participation, which did not impact their usual routine, each of them personally received an individualized analysis of their HR data, so that they could receive information about cardiac demand during various phases of their training sessions (intensity levels, duration, and cardiac work zone) and adjust their sessions' contents if needed.

Data Collection

PPG HR signals from the FC4 were compared to those from the Polar H10 thoracic belt (Polar Electro Oy), which was used as the criterion sensor [26]. The assessment of skin tone was performed with the Fitzpatrick Skin Scale, which ranges from 1 (lightest tone) to 6 (darkest tone) [27].

Participants were asked to wear both sensors simultaneously at each session. The FC4 was placed on the wrist of the nondominant arm (ie, around 2 cm away [proximal] from the ulnar styloid process), whereas the Polar H10 thoracic belt was placed under the thorax (ie, on the xyphoid process) and paired with Polar V800 wristwatches for HR recording. Each device was placed firmly against the skin, as recommended by the manufacturer's instructions.

Data Extraction and Analysis

Data were extracted through both companies' web services (ie, the Fitbit app [28] and Polar Flow website [29] for the FC4 and Polar H10, respectively). A MATLAB script (The MathWorks Inc) was then used to synchronize the two devices' HR measurements. Record alignments were performed by using a least square method to minimize squared deviation between FC4 and Polar H10 records, and they were smoothed over a 10-second window to calibrate the sample rate from both devices, as previously described [18,30]. Data normality was verified by a Kolmogorov-Smirnov test.

FC4 artifact data were defined as values that deviated from the criterion data by 20 beats per minute (bpm). Bland-Altman tests were performed on smoothed data to assess the accuracy of FC4 HR data by participant and by activity [31]. Means (bias) and SDs of the differences between the FC4 and H10 values were used to evaluate upper and lower limits of agreement (LOAs), per the following formula: upper/lower LOA = bias ± 1.96 × SD. Mean absolute error (MAE) and mean absolute percentage error (MAPE) were calculated to quantify mean differences between FC4 and Polar H10 HR data. Two tests were performed to evaluate the reliability of the FC4: (1) 2-way random intraclass correlation coefficients (ICCs) with an absolute consistency type were calculated and interpreted according to

XSL•FO
RenderX

current guidelines (ICC<0.5: poor; 0.5<ICC<0.75: moderate; 0.75<ICC<0.90: good; ICC>0.90: excellent reliability) [32], and (2) a computation of Lin concordance correlation coefficients (CCCs) was performed, interpreted following McBride's [33] recommendations (CCC<0.90: poor; 0.90<CCC<0.95: moderate; 0.95<CCC<0.99: very good; CCC>0.99: almost perfect strength of agreement).

All statistical analyses were stratified by activity type, and a Kruskal-Wallis test was performed to compare activity bias, with activity types as independent groups. Mann-Whitney *U* tests were implemented to compare bias from 0 among activity types (independent groups). Further, a linear univariate model was used to estimate the effect of skin tone on bias while controlling the impact of activity type. Statistics were performed using IBM SPSS statistics 25 software (IBM Corp).

Results

Participants' Characteristics

A total of 26 young adults (11 women and 15 men) were included in this study. Their characteristics are compiled in Table 1. In total, 77.5 hours of practice were recorded, distributed across 55 sessions (Table 2).

Table . Participants' characteristics.

	Male (n=15), mean (SD)	Female (n=11), mean (SD)	All participants (N=26), mean (SD)
Age (y)	21.2 (7.0)	20.8 (3.7)	21.1 (5.8)
Weight (kg)	75.8 (9.6)	57.6 (8.9)	68.1 (12.9)
Height (cm)	183 (6)	166 (8)	176 (11)
BMI (kg/m^2)	22.6 (1.8)	20.9 (1.7)	21.9 (1.9)
Skin tone (Fitzpatrick Skin Scale score)	2.9 (0.6)	2.8 (0.6)	2.9 (0.6)

Table . Descriptive data of recorded sessions.

Activity		Sessions, n	Recorded time, h	Participants, n
Racket sports				
	Badminton	10	15.07	7
	Tennis	3	4.90	2
	Total	13	19.96	9
Running sports				
	Orienteering running	5	7.02	5
	Run	11	14.09	3
	Total	16	21.10	8
Other sports				
	Bike	13	18.39	2
	Soccer	13	18.01	12

Accuracy and Artifact Percentage

Biases, LOAs, and artifact percentages are shown in Table 3.

XSL•FO
RenderX

Table . Bland-Altman analyses, mean absolute error (MAE), and mean absolute percentage error (MAPE) by activity.

Activity		Bland-Altman analyses, bpm[a]			Artifact ratios, %	bpm, MAE (SD)	bpm, MAPE (SD)
		Bias	Upper LOA[b]; lower LOA	LOA width			
Racket sports							
	Badminton	−16.5	35.2; −68.2	103.5	39.3	21.7 (22.3)	16.2 (14.4)
	Tennis	−6.2	24.8; −37.2	62.0	22.7	12.8 (11.2)	8.9 (7.4)
	Total	−14.0	34.3; −62.3	96.6	35.2	19.5 (20.5)	14.4 (13.4)
Running sports							
	Orienteering running	−8.6	26.0; −43.3	69.4	17.0	11.7 (15.9)	9.5 (10.4)
	Run	0.1	10.9; −10.7	21.6	2.5	1.7 (5.2)	1.2 (4.6)
	Total	−2.8	20.5; −26.1	46.6	7.3	5.0 (11.1)	4.0 (8.1)
Other sports							
	Bike	4.8	36.8; −27.3	64.1	18.1	10.4 (10.4)	8.1 (11.0)
	Soccer	−16.5	26.7; −59.6	86.3	35.5	19.2 (19.7)	17.5 (20.8)

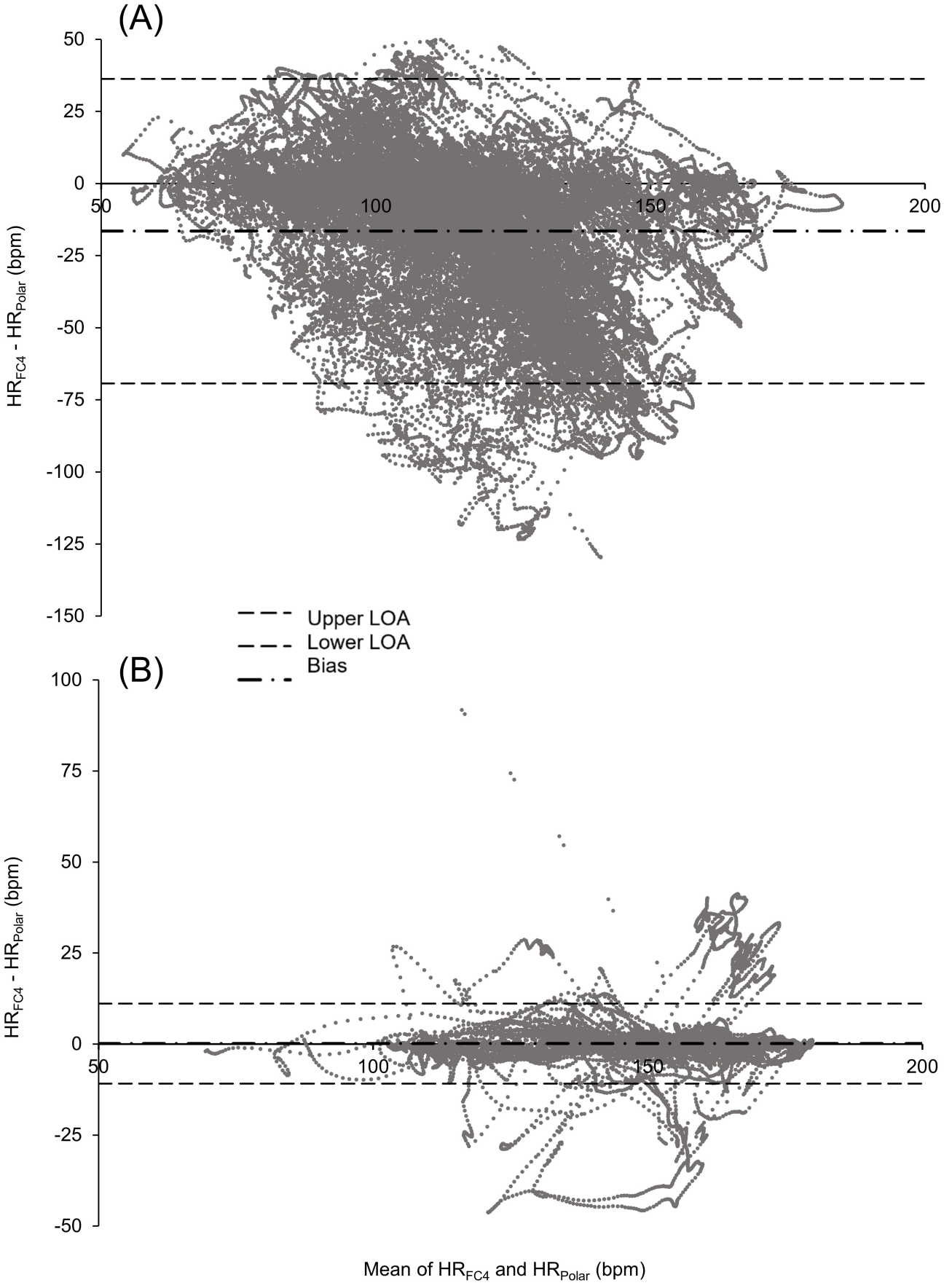
[a]bpm: beats per minute.

[b]LOA: limit of agreement.

Biases were different between each activity (P<.001), with all of them also being different from 0 (P<.001). The lowest bias values, MAEs, and MAPEs were found for running and cycling, and the highest ones were found for badminton and soccer. Furthermore, the narrowest LOA width was found for running, and the widest was found for badminton (Figure 1). We found similar results by grouping activities; the lowest bias, MAE, MAPE, and LOA width were found for running activities (running and orienteering running), and the largest ones were found for racket sports (badminton and tennis; Table 3).

XSL•FO
RenderX

Figure 1. Bland-Altman plots for (A) badminton and (B) running. Badminton exhibits the highest bias, and the bias for running is the closest to the origin. Each activity represents, respectively, 14 and 15.07 hours of recording. bpm: beats per minute; FC4: Fitbit Charge 4; HR: heart rate; LOA: limit of agreement; Polar: Polar H10.



XSL•FO
RenderX

Reliability

ICCs and CCCs are presented in Table 4. ICCs and CCCs indicated poor reliability (<0.50 and <0.90, respectively) for racket sports and soccer but excellent reliability for running overall (the total ICC and CCC for all running activities overall were >0.90 and >0.99, respectively). The largest percentages of artifact HR data were found for badminton and soccer, whereas running exhibited the lowest (Table 3). The results were similar when pooling activities; the highest rate of artifact HR data was found for racket sports, and the lowest was found for running sports (Table 3).

Table . Intraclass correlation coefficients (ICCs) and Lin concordance correlation coefficients (CCCs) by single and pooled activities.

Activity		ICC (95% CI)	CCC
Racket sports			
	Badminton	0.365 (0.517-0.154)	0.778
	Tennis	0.421 (0.320-0.421)	0.884
	Total	0.435 (0.237-0.574)	0.883
Running sports			
	Orienteering running	0.801 (0.865-0.801)	0.932
	Run	0.900 (0.898-0.900)	0.999
	Total	0.926 (0.915-0.935)	0.996
Other sports			
	Bike	0.658 (0.702-0.603)	0.971
	Soccer	0.487 (0.158-0.487)	0.809

Effect of Skin Tone

Mean Fitzpatrick Skin Scale scores are shown in Table 1. No overall interaction and no interaction in each activity were found between bias and skin tone, while being standardized by activity type.

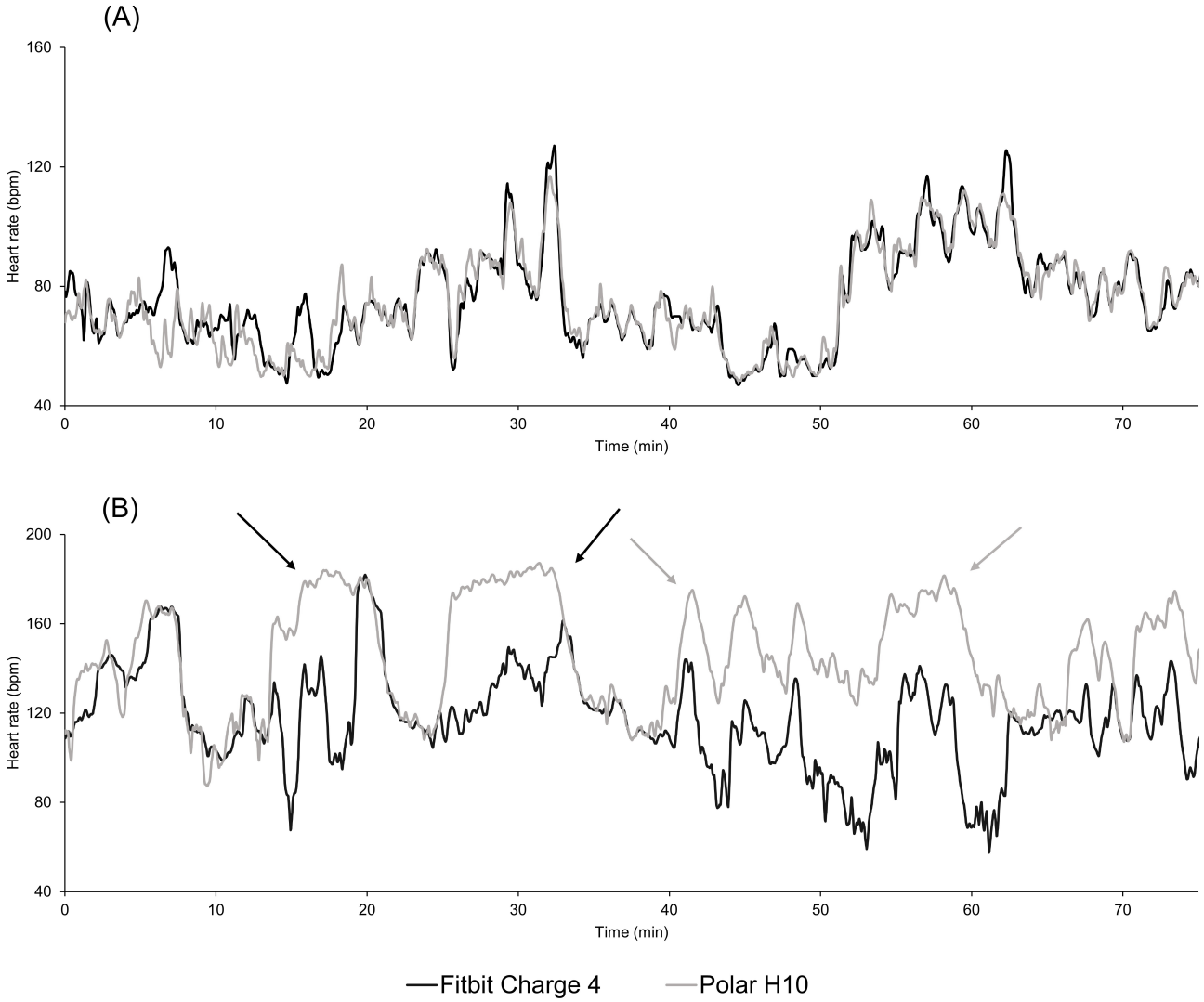
Discussion

Main Results

In this study, we evaluated the HR accuracy and reliability of the FC4 by comparing it to the Polar H10 chest belt (criterion sensor) in non–laboratory-controlled conditions. Our results showed negative biases for most activities (except running and cycling; Table 3), the presence of artifact data, and HR underevaluation by the FC4 (Figure 2). These results are similar to earlier findings that show the tendency of PPG wrist sensors to overestimate or underestimate HR [34-36]. ICCs and CCCs were fluctuant, mainly depending on the activity type. The FC4 shows good reliability for running activities, with almost perfect and moderate CCCs and excellent and good ICCs for running and orienteering running, respectively. Additionally, running was the activity with the lowest bias, the lowest MAPE, and the smallest LOA width. On the other hand, we found lower ICCs for badminton, soccer, and cycling, which also showed higher artifact ratios. Thus, we suggest that the excessive amount of arm movement in these activities could affect HR recording, as previously shown [37]. Badminton and tennis are characterized by “sharp” movements and rotations of the nondominant arm, whereas soccer can induce some instability of the sensor due to random arm and wrist actions, which can result in watches sliding over skin and the transient loss of HR signals (Figure 2). Furthermore, although cycling shows an overall good CCC (0.971), the ICC (0.658), MAPE (8.1%, SD 11.0%), and LOA width (64.03 bpm) indicate a lack of reliability during this activity. Since cycling remains a lower limb cyclic activity, there are little to no arm movements that may affect sensor placement. However, wrist position on handlebars (eg, during road cycling) and the contractility of wrist muscle flexors and extensors could alter vascular arteriovenous system detection and lower signal quality while enhancing compression forces [6,8,16,17]. ICCs and CCCs were calculated for pooled activities—running sports (running and orienteering running) and racket sports (tennis and badminton). These activities mostly rely on the same corporal pattern, and grouping them allowed us to equilibrate the time of practice between other activities. Our data showed no differences in ICC or CCC parameters, with those for running sports and racket sports showing excellent and poor reliability, respectively (ICC: 0.926 vs 0.435; CCC: 0.996 vs 0.883; Table 4).

XSL•FO
RenderX

Figure 2. Examples of synchronized heart rate signals (Polar H10: light grey; Fitbit Charge 4: black; A: orienteering running; B: soccer). Soccer heart rate data show recurrent sudden uncoupling between Fitbit Charge 4 and Polar H10 data, as well as heart rate underevaluations by the Fitbit Charge 4 throughout the recording session. bpm: beats per minute.



Overall, artifacts and random movements could highly affect a sensor's precision. Some studies tried to reduce motion artifacts by using novel techniques, such as accelerometry coupling or algorithm-based processing, but there is no consensus yet on which one should be used [38-42]. Even by varying numbers of diodes, lights colors, and algorithms, the technical aspect for measuring HR remains similar, suggesting that some FC4 characteristics could increase the presence of motion artifacts. For example, the materials used for the strap was slippery on skin, amplified by exercise-induced sudation. In addition, from a general standpoint, algorithms used by manufacturers can also affect the recordings, but we did not have access to these proprietary processing scripts.

Limitations

We observed no influence of skin tone on bias, unlike previous works that highlighted some effects on HR error rates [21]. However, our study is in line with another paper that was based on analyses of the signal to noise ratio, which showed no effect of skin tone when using a proper PPG wavelength (520 nm) [15]. More recently, another study did not find an effect of skin tone on beat-to-beat interval quality while separating skin types into two major groups (group 1: 1 to 4 on the Fitzpatrick Skin Scale; group 2: 5 and 6 on the Fitzpatrick Skin Scale) [43]. Even if our study findings are consistent with these results, no generalization can be made, since participants' skin types ranged from 2 (n=6) to 4 (n=3) on the Fitzpatrick Skin Scale. Moreover, we did not take into account physiological and environment factors, such as local temperature, humidity, and sudation, which may impact peripheral vasomotricity and therefore increase or decrease PPG signal intensity [44-46]. It would also have been relevant to assess the influence of motion parameters, such as acceleration measured at the wrist, on HR accuracy and reliability of the FC4; however, we could not access raw data produced by the proprietary processes for further analyses, and it was not possible to use inertial units because these could have resulted in discomfort for the participants, and inertial units are not a commodity among the public. Furthermore, we chose to place the FC4 on the wrist of the nondominant arm, as this placement is recurrent in daily life and research, although we knew that the movements would be inferior to those of the dominant wrist and therefore would reduce the artifact rate. Finally, the low number of participants included for the running activity (n=3) should be considered while interpreting our

XSL•FO
RenderX

results, although the pooled analysis lowered this potential bias while providing similar reliability and accuracy results.

Comparison With Previous Works

A recent study evaluated the accuracy of the FC4 HR sensor against an ECG Holter monitor for activities of daily living (sitting, walking, typing, lying down, etc) and showed acceptable HR measurement capabilities [47]. The added value of our protocol comes from the ecological approach to the gear sensor validation, which included additional physical activities with various intensities and limb movements. To our knowledge, no study has evaluated the FC4 in this manner during multiple sports or activities yet. However, further studies should be conducted to measure the reliability of the FC4 and its parameters for activities of daily living (number of steps, number of stairs climbed, number of calories burned, and quality of sleep), especially among persons or patients whose physical activities are restricted, as observed in sedentary individuals, people with obesity, people with heart failure, etc.

Considering our results (ie, the lack of precision and reliability in specific activities), the FC4 should not be used for research or athlete training purposes, including those related to running, which showed the lowest LOA width and artifact ratio. However, the FC4 could be useful for tracking HR during daily activities, which does not require such accurate monitoring, and it may be considered for patients' reeducation. However, for the latter, further studies should inspect the FC4's reliability according to the characteristics of the population. For example, obesity affects the physiological factors necessary for proper PPG signal intensity and quality, such as capillary density and recruitment, blood flow, and skin thickness [44].

Conclusion

The FC4 shows excellent reliability for measuring HR during activities with slow and predictive arms movements, such as running. However, it should not be used for activities with "sharp" and random arm and wrist movements, such as soccer and racket sports, due to its sensitivity to motion artifacts. Hence, in ecological conditions, this device should not be used for research or training purposes due to the high artifact rate and LOA width.

Authors' Contributions

MC and EH designed the research. MC collected and analyzed the data. MC wrote the manuscript. EH and HD revised the manuscript. All authors have read and agreed to publish this version of the paper.

Conflicts of Interest

None declared.

References

1. Costello K. Gartner says worldwide wearable device sales to grow 26 percent in 2019. Gartner. 2018 Nov 29. URL: https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow-#:~:text=Gartner%2C%20Inc.,billion%20will%20be%20on%20smartwatches [accessed 2024-12-12]
2. Gilgen-Ammann R, Schweizer T, Wyss T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. Eur J Appl Physiol 2019 Jul;119(7):1525-1532. [doi: 10.1007/s00421-019-04142-5] [Medline: 31004219]
3. Speer KE, Semple S, Naumovski N, McKune AJ. Measuring heart rate variability using commercially available devices in healthy children: a validity and reliability study. Eur J Investig Health Psychol Educ 2020 Jan 10;10(1):390-404. [doi: 10.3390/ejihpe10010029] [Medline: 34542492]
4. Andre D, Wolf DL. Recent advances in free-living physical activity monitoring: a review. J Diabetes Sci Technol 2007 Sep;1(5):760-767. [doi: 10.1177/193229680700100522] [Medline: 19885145]
5. Hertzman AB. Photoelectric plethysmography of the fingers and toes in man. Exp Biol Med (Maywood) 1937 Dec 1;37(3):529-534. [doi: 10.3181/00379727-37-9630]
6. Alian AA, Shelley KH. Photoplethysmography. Best Pract Res Clin Anaesthesiol 2014 Dec;28(4):395-406. [doi: 10.1016/j.bpa.2014.08.006] [Medline: 25480769]
7. Bartels K, Thiele RH. Advances in photoplethysmography: beyond arterial oxygen saturation. Can J Anesth 2015 Dec;62(12):1313-1328. [doi: 10.1007/s12630-015-0458-0] [Medline: 26286382]
8. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-R39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
9. Maeda Y, Sekine M, Tamura T, Moriya A, Suzuki T, Kameyama K. Comparison of reflected green light and infrared photoplethysmography. Presented at: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 20-25, 2008; Vancouver, BC p. 2270-2272. [doi: 10.1109/IEMBS.2008.4649649]
10. Anderson RR, Parrish JA. The optics of human skin. J Invest Dermatol 1981 Jul;77(1):13-19. [doi: 10.1111/1523-1747.ep12479191] [Medline: 7252245]
11. Cui WJ, Ostrander LE, Lee BY. In vivo reflectance of blood and tissue as a function of light wavelength. IEEE Trans Biomed Eng 1990 Jun;37(6):632-639. [doi: 10.1109/10.55667] [Medline: 2354845]

XSL•FO
RenderX

12. Giltvedt J, Sira A, Helme P. Pulsed multifrequency photoplethysmograph. Med Biol Eng Comput 1984 May;22(3):212-215. [doi: 10.1007/BF02442745] [Medline: 6738126]
13. Maeda Y, Sekine M, Tamura T. Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. J Med Syst 2011 Oct;35(5):969-976. [doi: 10.1007/s10916-010-9505-0] [Medline: 20703691]
14. Lee J, Matsumura K, Yamakoshi KI, Rolfe P, Tanaka S, Yamakoshi T. Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. Presented at: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 3-7, 2013; Osaka, Japan p. 1724-1727. [doi: 10.1109/EMBC.2013.6609852]
15. Fallow BA, Tarumi T, Tanaka H. Influence of skin type and wavelength on light wave reflectance. J Clin Monit Comput 2013 Jun;27(3):313-317. [doi: 10.1007/s10877-013-9436-7] [Medline: 23397431]
16. Rafolt D, Gallasch E. Influence of contact forces on wrist photoplethysmography--prestudy for a wearable patient monitor. Biomed Tech (Berl) 2004;49(1-2):22-26. [doi: 10.1515/BMT.2004.005] [Medline: 15032494]
17. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Aug 5;15(3):540-547. [Medline: 27803634]
18. Hermand E, Cassirame J, Ennequin G, Hue O. Validation of a photoplethysmographic heart rate monitor: Polar OH1. Int J Sports Med 2019 Jul;40(7):462-467. [doi: 10.1055/a-0875-4033] [Medline: 31189190]
19. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
20. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34. [doi: 10.2196/mhealth.7043] [Medline: 28302596]
21. Shcherbina A, Mattsson CM, Waggott D, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):3. [doi: 10.3390/jpm7020003] [Medline: 28538708]
22. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016 Apr 25;2(1):e000106. [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
23. Hough P, Glaister M, Pledger A. The accuracy of wrist-worn heart rate monitors across a range of exercise intensities. J Phys Act Res 2017 Nov 25;2(2):112-116. [doi: 10.12691/jpar-2-2-8]
24. Lee CM, Gorelick M. Validity of the Smarthealth watch to measure heart rate during rest and exercise. Meas Phys Educ Exerc Sci 2011 Jan 29;15(1):18-25. [doi: 10.1080/1091367X.2011.539089]
25. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
26. Schaffarczyk M, Rogers B, Reer R, Gronwald T. Validity of the Polar H10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women. Sensors (Basel) 2022 Aug 30;22(17):6536. [doi: 10.3390/s22176536] [Medline: 36081005]
27. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol 1988 Jun;124(6):869-871. [doi: 10.1001/archderm.124.6.869] [Medline: 3377516]
28. Fitbit. Google Play. URL: https://play.google.com/store/apps/details?id=com.fitbit.FitbitMobile&hl=en [accessed 2024-12-23]
29. Polar Flow. URL: https://flow.polar.com/ [accessed 2024-12-23]
30. Mühlen JM, Stang J, Lykke Skovgaard E, et al. Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE Network. Br J Sports Med 2021 Jul;55(14):767-779. [doi: 10.1136/bjsports-2020-103148] [Medline: 33397674]
31. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. Journal of the Royal Statistical Society: Series D (The Statistician) 1983 Sep;32(3):307-317. [doi: 10.2307/2987937]
32. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
33. McBride GB. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. : National Institute of Water and Atmospheric Research; 2005 NIWA client report: HAM2005-062.
34. Gillinov S, Etiwy M, Wang R, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
35. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018 Feb 28;13(2):e0192691. [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
36. Boudreaux BD, Hebert EP, Hollander DB, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
37. Ahmadi AK, Moradi P, Malihi M, Karimi S, Shamsollahi MB. Heart rate monitoring during physical exercise using wrist-type photoplethysmographic (PPG) signals. Presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 25-29, 2015; Milan, Italy p. 6166-6169. [doi: 10.1109/EMBC.2015.7319800]

XSL•FO
RenderX

38. Zhu S, Tan K, Zhang X, Liu Z, Liu B. MICROST: a mixed approach for heart rate monitoring during intensive physical exercise using wrist-type PPG signals. Presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 25-29, 2015; Milan, Italy p. 2347-2350. [doi: 10.1109/EMBC.2015.7318864]
39. Mashhadi MB, Asadi E, Eskandari M, Kiani S, Marvasti F. Heart rate tracking using wrist-type photoplethysmographic (PPG) signals during physical exercise with simultaneous accelerometry. IEEE Signal Process Lett 2016 Feb;23(2):227-231. [doi: 10.1109/LSP.2015.2509868]
40. Fujita Y, Hiromoto M, Sato T. PARHELIA: particle filter-based heart rate estimation from photoplethysmographic signals during physical exercise. IEEE Trans Biomed Eng 2018 Jan;65(1):189-198. [doi: 10.1109/TBME.2017.2697911] [Medline: 28459679]
41. Biswas D, Simoes-Capela N, Van Hoof C, Van Helleputte N. Heart rate estimation from wrist-worn photoplethysmography: a review. IEEE Sensors J 2019 Aug 15;19(16):6560-6570. [doi: 10.1109/JSEN.2019.2914166]
42. Arunkumar KR, Bhaskar M. Heart rate estimation from wrist-type photoplethysmography signals during physical exercise. Biomed Signal Process Control 2020 Mar;57:101790. [doi: 10.1016/j.bspc.2019.101790]
43. Puranen A, Halkola T, Kirkeby O, Vehkaoja A. Effect of skin tone and activity on the performance of wrist-worn optical beat-to-beat heart rate monitoring. Presented at: 2020 IEEE SENSORS; Oct 25-28, 2020; Rotterdam, Netherlands p. 1-4. [doi: 10.1109/SENSORS47125.2020.9278523]
44. Fine J, Branan KL, Rodriguez AJ, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosens (Basel) 2021 Apr 16;11(4):126. [doi: 10.3390/bios11040126] [Medline: 33923469]
45. Jeong IC, Yoon H, Kang H, Yeom H. Effects of skin surface temperature on photoplethysmograph. J Healthc Eng 2014;5(4):429-438. [doi: 10.1260/2040-2295.5.4.429] [Medline: 25516126]
46. Johnson JM. Exercise in a hot environment: the skin circulation. Scand J Med Sci Sports 2010 Oct;20 Suppl 3:29-39. [doi: 10.1111/j.1600-0838.2010.01206.x] [Medline: 21029188]
47. Nissen M, Slim S, Jäger K, et al. Heart rate measurement accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: device evaluation study. JMIR Form Res 2022 Mar 1;6(3):e33635. [doi: 10.2196/33635] [Medline: 35230250]

Abbreviations

bpm: beats per minute
CCC: Lin concordance correlation coefficient
ECG: electrocardiogram
FC4: Fitbit Charge 4
HR: heart rate
ICC: intraclass correlation coefficient
LOA: limit of agreement
MAE: mean absolute error
MAPE: mean absolute percentage error
PPG: photoplethysmography

Edited by L Buis; submitted 25.11.23; peer-reviewed by L Ardigò, M Nissen, M Etiwy, S Okita; revised version received 10.10.24; accepted 24.10.24; published 08.01.25.

Please cite as:
Ceugniez M, Devanne H, Hermand E
Reliability and Accuracy of the Fitbit Charge 4 Photoplethysmography Heart Rate Sensor in Ecological Conditions: Validation Study
JMIR Mhealth Uhealth 2025;13:e54871
URL: https://mhealth.jmir.org/2025/1/e54871
doi:10.2196/54871

© Maxime Ceugniez, Hervé Devanne, Eric Hermand. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Trade-Offs Between Simplifying Inertial Measurement Unit–Based Movement Recordings and the Attainability of Different Levels of Analyses: Systematic Assessment of Method Variations

Manu Airaksinen[1], DSc; Okko Räsänen[2], DSc; Sampsa Vanhatalo[1,3], MD, PhD

[1]BABA Center, University of Helsinki and Helsinki University Hospital, Haartmaninkatu 8, Helsinki, Finland

[2]Signal Processing Research Centre, Tampere University, Tampere, Finland

[3]Department of Physiology, University of Helsinki, Helsinki, Finland

Corresponding Author:
Manu Airaksinen, DSc
BABA Center, University of Helsinki and Helsinki University Hospital, Haartmaninkatu 8, Helsinki, Finland

Abstract

Background: Human movement activity is commonly recorded with inertial measurement unit (IMU) sensors in many science disciplines. The IMU data can be used for an algorithmic detection of different postures and movements, which may support more detailed assessments of complex behaviors, such as daily activities. Studies on human behavior in real-life environments need to strike a balance between simplifying the recording settings and preserving sufficient analytic gains. It is poorly understood, however, what the trade-offs are between alternative recording configurations and the attainable analyses of naturalistic behavior at different levels of inspection, or with respect to achievable scientific questions.

Objective: This study assessed systematically the effects of IMU recording configurations (placement and number of IMU sensors, sampling frequency, and sensor modality) on the high temporal resolution detections of postures and movements, and on their lower temporal resolution derivative statistics when the data represents naturalistic daily activity without excessively repetitive movements.

Methods: We used a dataset from spontaneously moving infants (N=41; age range 4 - 18 months) recorded with a multisensor wearable suit. The analysis benchmark was obtained using human annotations of postures and movements from a synchronously recorded video, and the reference IMU recording configuration included 4 IMU sensors collecting triaxial accelerometer and gyroscope modalities at 52 Hz. Then, we systematically tested how the algorithmic classification of postures (N=7), and movements (N=9), as well as their distributions and a derivative motor performance score, are affected by reducing IMU data sampling frequency, sensor modality, and sensor placement.

Results: Our results show that reducing the number of sensors has a significant effect on classifier performance, and the single sensor configurations were nonfeasible (posture classification Cohen kappa<0.75; movement<0.45). Reducing sensor modalities to accelerometer only, that is, dropping gyroscope data, leads to a modest reduction in movement classification performance (kappa=0.50-0.53). However, the sampling frequency could be reduced from 52 to 6 Hz with negligible effects on the classifications (posture kappa=0.90-0.92; movement=0.56-0.58).

Conclusions: The present findings highlight the significant trade-offs between IMU recording configurations and the attainability of sufficiently reliable analyses at different levels. Notably, the single-sensor recordings employed in most of the literature and wearable solutions are of very limited use when assessing the key aspects of real-world movement behavior at relevant temporal resolutions. The minimal configuration with an acceptable classifier performance includes at least a combination of one upper and one lower extremity sensor, at least 13 Hz sampling frequency, and at least an accelerometer, but preferably also a gyroscope (posture kappa=0.89-0.91; movement=0.50-0.53). These findings have direct implications for the design of future studies and wearable solutions that aim to quantify spontaneously occurring postures and movements in natural behaviors.

(JMIR Mhealth Uhealth 2025;13:e58078) doi:10.2196/58078

KEYWORDS

human activity recognition; recording configuration; wearable; movement sensors; gross motor development; neurodevelopment; neural networks; MAIJU; IMU; sensor; algorithm; detection; balance; motility; activity recognition; motor; motor development; posture; motility assessment; accelerometer

XSL•FO
RenderX

Introduction

There is a wide interest in studying human activity with inertial measurement unit (IMU)-sensors in a variety of disciplines, including health care, developmental psychology, education, and sociology [1-6]. The objective physical measures of acceleration and gyration from the IMU sensors are considered to provide a significant complement, or alternative, to the traditionally used questionnaires [7] and expert assessments [8], both of which are compromised by subjectiveness and inherent ambiguities. Moreover, the expert-based assessments in the lab or hospital setting are costly and labor-intensive, whereas IMU sensors can be used as wearables to measure human activity in the real world, such as in participants' homes, supporting substantially improved ecological validity of the findings.

Compared with the many other methods used in movement studies, such as surface electromyography, video recordings, or 3D motion capture (3D movement analysis), the currently available IMU sensors allow for an easy-to-use [9], cost-efficient, privacy-preserving [10], and domain-relevant collection of movement data that can be easily deployed for long-form recording. IMU-sensor recordings are well accepted by the participants and their caregivers (in the case of minors), and they are relatively easy to use in long-term naturalistic measurements [11,12]. Moreover, the analyses and interpretability of IMU signals may be considerably easier than analyzing electromyography signals or video recordings obtained in uncontrolled environments.

The traditional approach to IMU-sensor data collection is to measure accelerometry ("actigraphy") signals from a single wrist-, hip, or ankle-worn accelerometer sensor [13], followed by quantifying the amounts of movements as total, with break-down to different levels of activity (eg sedentary, light, or moderate to vigorous), or to indirectly estimate the related energy expenditure [1,14-16]. The single-sensor setups are used in many commercially available wellness products to categorize sustained and repetitive activities (eg, sitting, lying, walking, running, climbing, and cycling [17,18]). Further, the recently introduced parallel gyroscope modality has increased its accuracy [19-21]. This methodologically simple approach reduces movement analysis to amounts of movements or detections of repetitive activities only, both of which fail to account for the frequently changing postures and movements, the key components of all natural human behavior. At the other extreme of recording setups, a very high number of IMU sensors may support detailed skeletal pose modeling [22-24]. However, the technical and analytic complexity of such measurements has limited their use to dedicated laboratory environments only.

An intermediate form between these extremes would be multi-IMU recording setups with a handful (N=2 - 6) of sensors, which have been used increasingly to improve the analytic range of human activity recognition (HAR); however, studies with such setups have used acted-out data collection with predefined activity types (eg, [18,25,26]), which inherently limits their generalizability to real-life measurements [27]. For at-home measurements of naturalistic activity, we have recently developed a multisensor wearable suit for infants (MAIJU, motor ability assessment of infants with a jumpsuit [28-32]), using 4 IMU sensors that collect triaxial accelerometer and gyroscope signals at 52 Hz. The data analysis uses a validated, deep learning–based analytical pipeline to provide a human-equivalent level classification of postures (7 categories) and generic movements (9 categories) for each second of the recording [29]. These high-temporal-resolution/low-analytic-level detections can be used further to build low-temporal-resolution/high-analytic-level assessments of daily activity or motor development [29,30].

In order to maximize the scalability and utility of the novel IMU technology, it would be essential to know the exact trade-offs between simplifying the recording setup and the losses of achievable analyses at different levels of inspection. For example, out-of-lab recordings would benefit from a lower number of sensors that simplify the wearable gear design, or lower data rates that increase the maximum recording lengths and reduce errors in data logistics; meanwhile, well-controlled environments with brief recordings would benefit from redundancy in the setup that increases analytical accuracy. Prior studies have assessed selected aspects of these trade-offs primarily on acted-out or otherwise controlled (ie, nonspontaneous) activities. They have reported that at least 2 sensors with either thigh and ankle [6] or thigh and shoulder [28] positioning are required to achieve adequate posture tracking with a high temporal resolution, and the sampling rate is likely not a significant analytic bottleneck between 10 and 100Hz [18,26,33].

Here, we studied systematically how the recording configurations in naturalistic measurements would affect the accuracy of analytic details at different levels of inspection, ranging from high-temporal-resolution and low-analytic-level detections to low-temporal-resolution and high-analytic-level assessments of spontaneous behavior. The overarching aim was to define which sensor placements, sampling frequencies, and sensor modalities are necessary for different levels of study questions. We used a multisensor IMU dataset with second-by-second level expert annotations of well-defined movements and postures in naturalistic settings, which together provide a rich, previously published and fully generalizable (ie, age-independent) presentation of spontaneously occurring movement behaviors. The novel contribution of this work is the systematic evaluation of the effects of (1) the number of sensors over 4 specific proximal limb locations, (2) sensor modality, and (3) sampling frequency, in the area of classifying posture and movement and their higher-order aggregate and derivative summary metrics for 4 - 18-month-old infants.

Methods

Overview

The overall goal was to study the effect of the recording configuration on the automated, quantified movement and posture analyses. The classifier outputs were assessed at high time resolution using previously developed second-by-second outputs [28,29], followed by assessing summary representations over the recording session (category distributions and scalar projections [29,30]). The ground truth of postures and

XSL•FO
RenderX

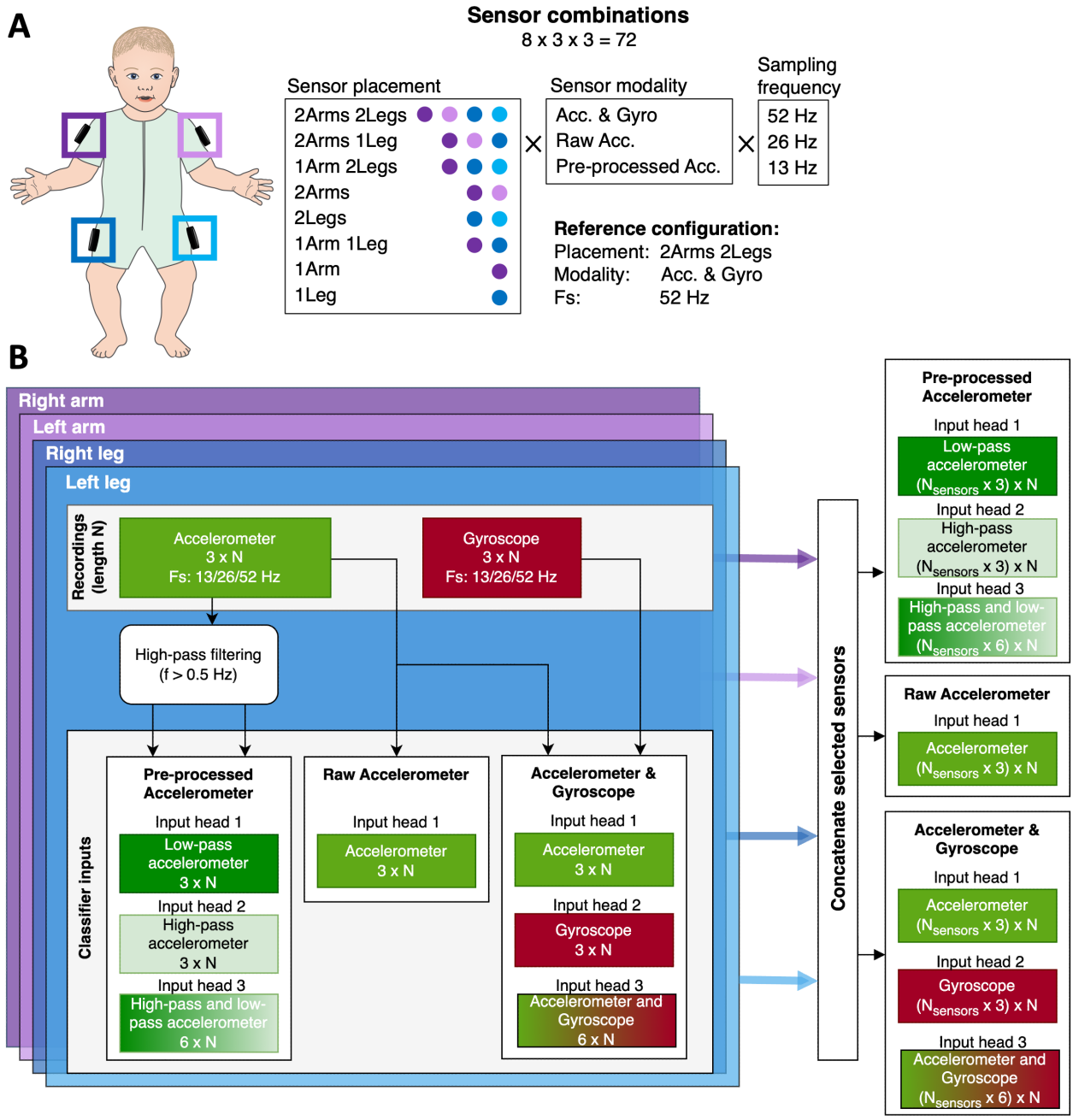
movements was taken from the human inter-rater agreement obtained from independent annotations using IMU-synchronized video recordings.

Data Collection and Recordings

MAIJU Wearable

MAIJU (Figure 1 [28-30]) is a full-body overall akin to a commonplace infant swimsuit that is equipped with 4 movement sensors (Movesense [34]), one proximally placed in each limb. The sensors stream IMU data via a Bluetooth 5.0 low-energy connection into an iOS-based mobile device with a custom-built application, "Maiju logger" (Kaasa GmbH, Düsseldorf, Germany). Data collection is performed with a 52 Hz sampling rate using a triaxial accelerometer (measuring linear acceleration at m/s^2; sensitivity 8 g) and gyroscope (measuring angular velocity at deg/s; sensitivity ±500 deg/s).

Figure 1. Overview of the study. (**A**) Illustration of the studied recording configurations that are obtained by subsampling from the reference configuration data. (**B**) Block diagram depicting the data preprocessing for each configuration. The "raw accelerometer" and "accelerometer+gyroscope" configurations take in the raw inertial measurement unit (IMU) signals (gyroscope with bias removal), whereas the "pre-processed accelerometer" configuration preprocesses the raw accelerometer signal into low- and high-pass parts, after which they are treated as independent data streams.



IMU Recording Configurations

Prior work with the MAIJU recordings has been published with the highest available configuration setup (4 sensors; accelerometer and gyroscope; 52 Hz sampling). This reference configuration was shown to enable classifier performance at a human equivalent level, ie, the agreement between the classifier and a human expert is comparable to human inter-rater

XSL•FO
RenderX

agreement [28,29]. Our previous experiments have shown that classifier robustness can be improved with simulated sensor dropout [31], and other studies on infants have shown that 2 or 3 IMU sensors may provide successful posture classification [4], if the sensors are attached to the unilateral hip, ankle, and waist. Initial experiments with MAIJU demonstrated that an incomplete list of postures is detectable with 2 sensors [28]. These results together suggest that the number of IMU sensors could be reduced from 4. Regarding the sensor locations, the proximal thigh placement (laterally in MAIJU wearable) is identified as the most robust single sensor location for reliable capture of different human activities [17].

In this work, we studied systematically how reductions in recording configurations affect the movement analyses at different levels of inspection, ranging from high-resolution (1 s level) classifications to summary measures of the whole recording session. The tested 3 domains of varying recording configurations were (Figure 1): sensor placement, sensor modality, and sampling frequency. Sensor placement refers to the number of sensors used (from 1 to 4) at the different (proximal-lateral-) limb locations. The recording modality of the sensors can be selected as either an accelerometer, a gyroscope, or both. Since the gyroscope requires orders of magnitude more power compared with the accelerometer, and it does not contain information about the absolute state of orientation, we omitted the gyroscope-only configuration from the study. Since we are using an end-to-end neural network for our classifier, we tested 2 variants of accelerometer-only configurations to reflect the 2 distinct physical sources of acceleration—the gravity component (limb orientation) and limb acceleration—which were obtained by splitting the accelerometer signal into low-pass and high-pass components, respectively.

The “raw accelerometer” variant uses solely the measured accelerometer as such, and in the configuration “pre-processed accelerometer” the low- and high-pass filtered components are fed into separate input heads (Figure 1B). Finally, we test all of the configuration variants with 3 different sampling frequencies supported by the Movesense sensor (52, 26, and 13 Hz), all of which are obtained from the original reference configuration recording by decimation followed by upsampling and low-pass filtering.

Dataset Details

This study uses 29.3 hours (91,449 frames) of previously published data [29] that includes 41 recordings from infants aged 4.5 - 16.6 months. The data, alongside parallel video recordings, were obtained at the infants’ homes (N=17) or at the lab within a home-like setting (N=24) during minimally structured play sessions that lasted from 18 minutes to 74 minutes.

Annotations

We used a previously reported set of human annotations [29] from 3 (N=34) or 2 (N=9) independent annotators. The annotations were performed with the free Anvil software [35], which presents synchronized video and raw signal waveforms to annotate 3 separate tracks: posture, movement, and auxiliary (for details, see [29]). The posture track contains 7 categories (prone, supine, side left, side right, crawl posture, sitting, and standing), and the movement track contains 9 categories (still, proto, elementary, fluent, transition, roll left, roll right, pivot left, and pivot right) (Figure 2). The auxiliary track contained segments where posture and movement annotation could not be performed, viz, infant being carried instead of independent movement and out-of-camera segments. These segments were not included in training or evaluation and have not been counted in the reported dataset size. The baseline Cohen kappa scores for the multiclass inter-rater agreement for the posture and movement annotations were 0.93 (95% CI 0.92 to 0.94) and 0.59 (95% CI 0.57 to 0.62), respectively. As discussed in detail in [28,29], the posture and movement tracks are designed to represent complementary aspects of human movement behavior, and due to the differences in their conceptual clarity and temporal dynamics (Figure 3), the kappa levels between the tracks should not be directly compared. Specifically, the accuracy metrics of the movement track suffer from short bout lengths because the true category transitions cannot be perfectly aligned with the analytic time frames. This unavoidable compromise in temporal accuracy arises from real-world behavior. Notably, our prior work has demonstrated how it is effectively mitigated by using aggregate statistics over the measurement session, yielding unbiased movement track accuracies that are highly comparable with human visual annotations [29] and that have direct utility in describing the gross motor developmental level [36]. Further, our prior work has demonstrated that the movement categories can be meaningfully combined, for example, to denote a dichotomy between “locomotion” and “non-locomotion” [37].

To estimate suitable window lengths for short-term analysis, we studied durations of the posture and movement segments in the data for different categories. In principle, the choice of window length is a compromise of 2 factors: more data in the longer windows would typically improve classifications, however, a too-long window length leads to contamination by different activity types inside the same window. The cumulative distributions for the counts of variable-length annotation segments are presented in Figure 3. The figure shows that almost none of the annotations are less than 1 second in length, whereas especially for movement, there are substantial amounts of segments with duration <2 seconds (≈10% - 40%, depending on the movement category; corresponding roughly to ≈5% - 20% of total time). This suggests that the length of analysis windows should be within 2 seconds to capture all annotated activities. Thus, we used the previously published [28-32], window length of 2.3 seconds (120 samples at 52 Hz) with 50% overlap (1.15 s; 60 samples at 52 Hz).

XSL•FO
RenderX

Figure 2. Example of the raw accelerometer signals and the video-based annotations for posture and movement within a 10-minute recording segment. For each track, the 3 parallel annotations are depicted alongside the training target (obtained with iterative annotation refinement [28]). The classifier outputs are presented with 4 different configurations: 2Arms 2Legs (52 Hz, acc and gyro), 1Arm 1Leg (13 Hz, pre-processed acc), 1 Leg (13 Hz, pre-processed acc), and 1 Arm (13 Hz, pre-processed acc). Note the increasing classification errors when reducing the recording configurations, and the apparently random classification with 1 arm recording.

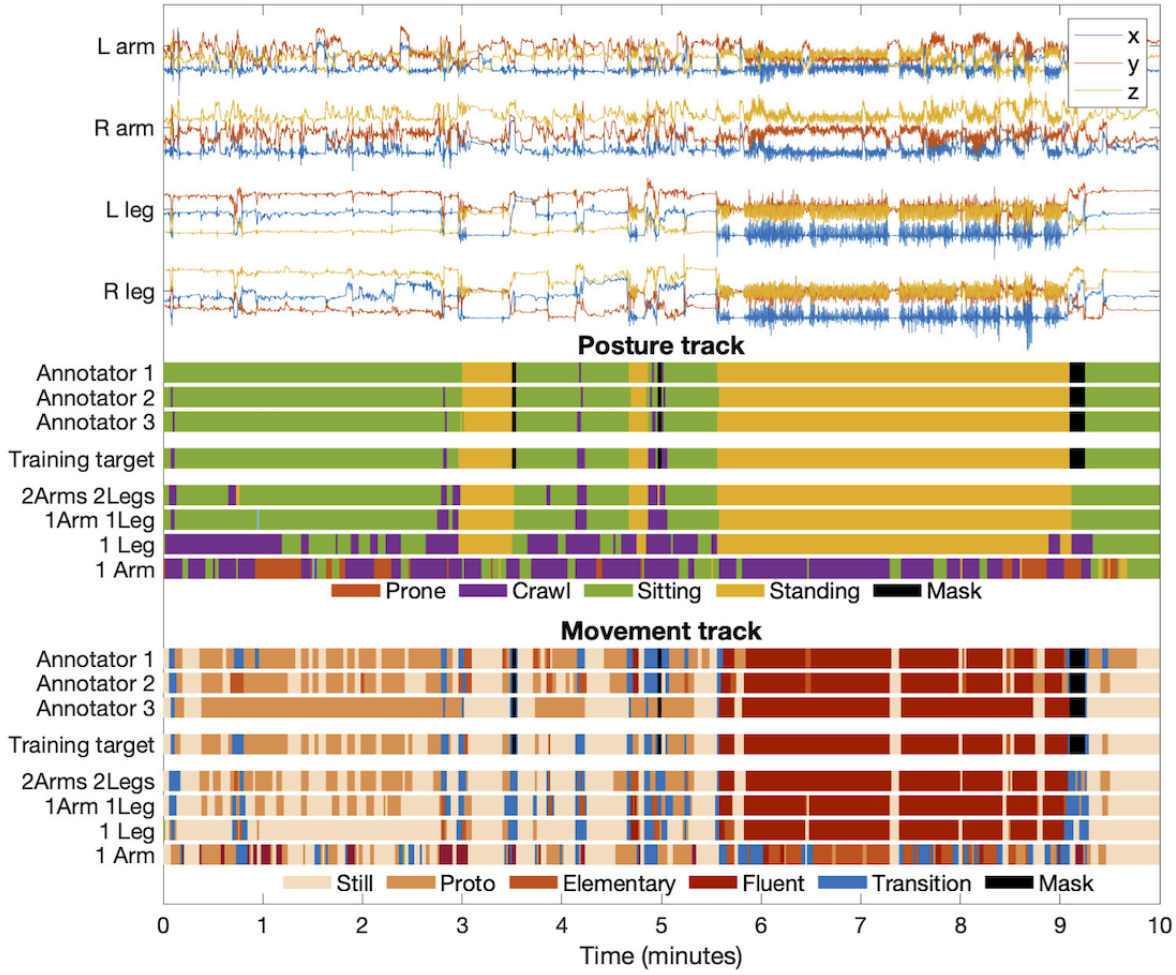
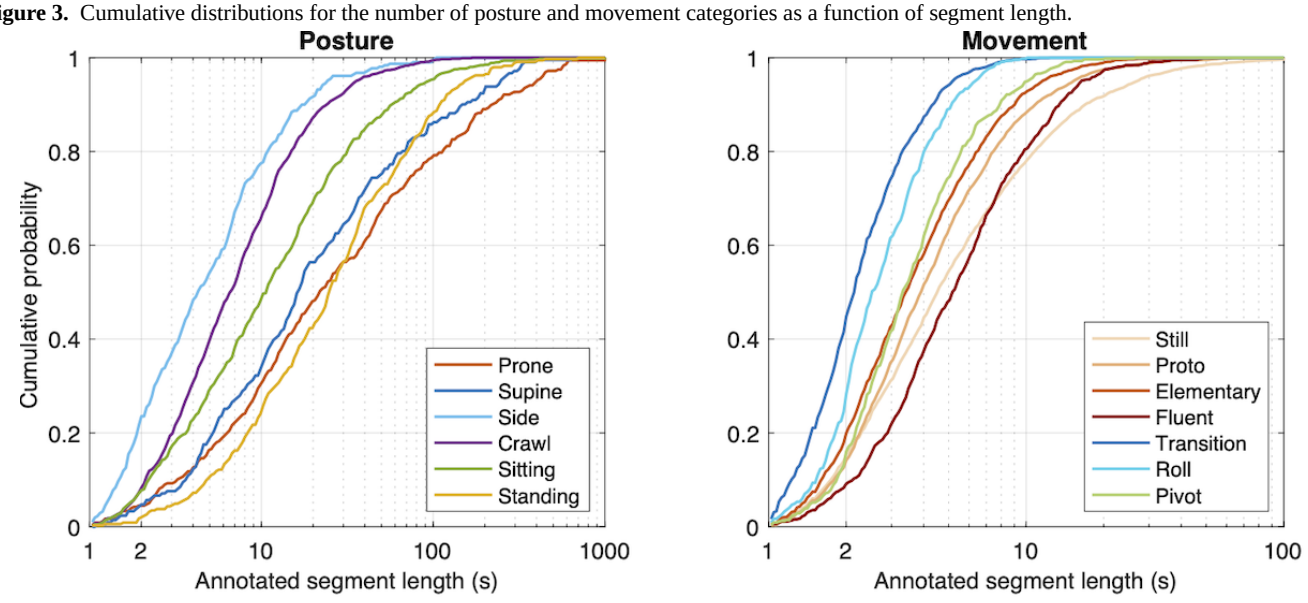


Figure 3. Cumulative distributions for the number of posture and movement categories as a function of segment length.



XSL•FO
RenderX

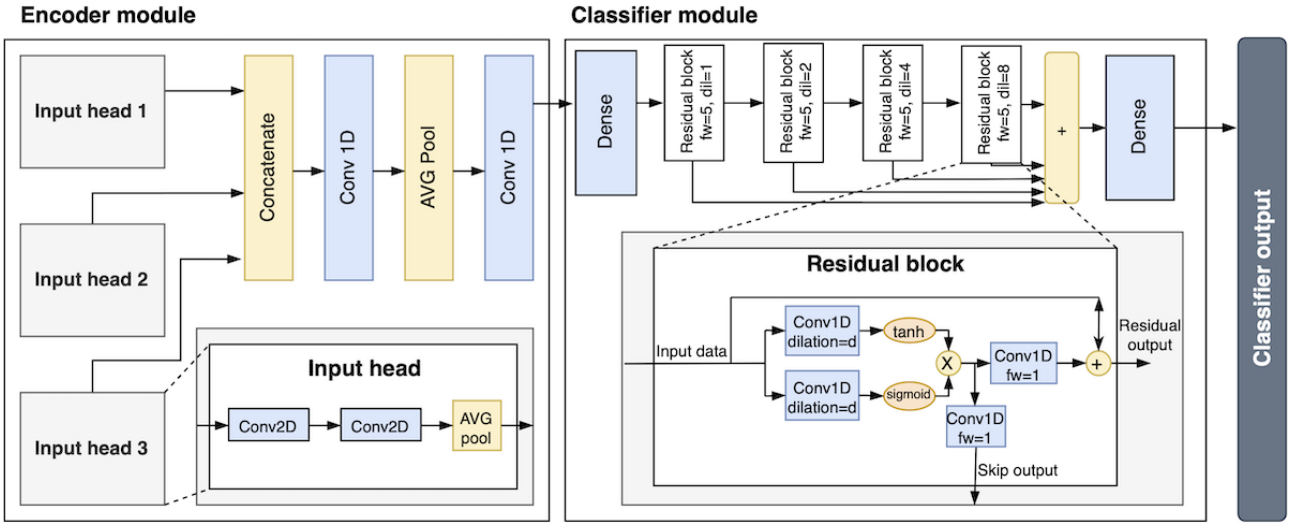
Data Preprocessing and Neural Network-Based Classifier

Data Preprocessing

Windowing

The raw IMU-sensor data is segmented into 2.3 s (120 samples per channel) frames with 50% overlap. The frames are processed and concatenated based on modality (Figure 1B) and then fed into the classifier (Figure 4).

Figure 4. Block diagram depicting the neural network structure for all of the classifiers.



Downsampling

The used Movesense sensor firmware supports measurement at sampling rates starting from 13 Hz. We simulated sampling frequencies lower than 52 Hz (reference configuration) by decimating the original signal by a factor of 2 or 4 (for 26 Hz and 13 Hz, respectively) and then upsampling by the same factor followed by forward-backward low-pass filtering with a sixth order Butterworth filter (with a cut-off frequency of 0.5 * decimated sampling frequency). This was done to keep the classifier architectures identical (ie, each variant takes in frames of 120 samples).

Pre-Processed Accelerometer

In the pre-processed accelerometer modality, the obtained raw accelerometer signal is split into low-pass and high-pass components that are processed in parallel blocks (Figure 4). The low-pass component in accelerometers is primarily caused by the earth's gravity, and it contains information about the absolute orientation of the sensor, whereas the high-pass component primarily contains information about the instantaneous movement. The pre-processed accelerometer signals were obtained by forward-backward filtering the raw accelerometer signals with an eighth-order Butterworth filter with a cut-off frequency of 0.5 Hz (for the low-pass component). The high-pass component was obtained by subtracting the low-pass component from the original signal.

Classifier Architecture

The overall architecture of the end-to-end neural network classifier is presented in Figure 4, and is identical to our earlier experiments [28-32]. The architecture consists of an encoder module (responsible for condensing frame-level information) and a classification module (which performs time-series modeling over the frame-level features). The encoder module has a varying amount of input heads, depending on the chosen modality (1 for raw accelerometer, 3 for the rest). Each input head has a similar architecture but takes as input the concatenated channels of the different sensor modalities (Figure 1B). The modality-level information is combined within the input head to latent representations that are concatenated and further processed in the final layers of the encoder module to obtain a 160-dimensional latent representation for each frame. The classifier module contains stacked residual blocks that do dilated convolutions over the frame-level representations of the data tensor (looking symmetrically into the past and future with a total receptive field of 30 frames=34.6 s).

Training Details

All classifier training experiments were performed with recording-level 10-fold cross-validation with a fixed random seed. Minibatch gradient descent with the ADAM optimization algorithm was used with a weighted categorical cross-entropy loss (batch size=1×100 consecutive frames, learning rate 10^{-4}, $beta_1$=.9, $beta_2$=.999, epsilon=10^{-8}). To mitigate the effects of unbalanced category distributions, each frame's error in the loss functions was weighted with the inverse probability of the target class occurrence. Sensor dropout (P=.3), sample dropout (P=.3), and random Euler rotation (P=.3, ±15 deg) were applied randomly during training to the input signals to enhance model robustness. The training was run for 200 epochs, and held-out validation data (20% of training data) was used to select the best-performing model in terms of the Cohen kappa score.

Levels of Analysis Outputs

Overview

The classifiers for posture and movement produce frame-level information at a rate of 0.87 frames/s. Many use cases, such as health care or developmental research, are interested in holistic

XSL•FO
RenderX

estimates of overall movement performance, and they need summary metrics from the whole recording with a fixed-length feature vector or a scalar value representing the recording. In this study, we used 2 such summary metrics at the level of individual infants: concatenated category distribution vectors and a holistic scalar value, BABA infant motor score (BIMS) [29].

Recording-Level Category Distributions

The classifiers produce a one-hot coded vector (1, N_{cats}) for each frame, where N_{cats} is the number of output categories in the classifier. Thus, for each track in a recording, the classifiers output a matrix of shape (N_{frames}, N_{cats}), where N_{frames} is the number of frames in the recording (at 1.15 s intervals). An automatic carrying detection classifier [32] produces a binary mask of shape (N_{frames}, 1), which is applied as a filter to the posture and movement classifier outputs before computing the distributions. The distribution for each track is computed by taking the mean of the automatic carrying detection-filtered one-hot matrices along the frame axis to yield a vector of shape (1, N_{cats}). It is possible to combine the posture and movement classifications into a single "posture-conditioned movement" representation (ie, prone-still, prone-fluent, standing-still, etc), which yields a total of 27 sensible combinations of motor activity (ie, standing-fluent is a sensible combination, but supine-fluent is not). After obtaining the recording-level distribution value $p'_{c(m_i)}$, where *c* is the category in question, *i* is the recording number, and *m* is the MAIJU recording, the values were concatenated across all recordings into a vector $p'=[p'_{c(m_i)}]$ and compared with the annotation-based reference values $[p_{c(m_i)}]$ with Pearson *r*. 95% CIs were obtained with bootstrapping (10,000 iterations) at the recording level.

Recording-Level Holistic Assessment

One of the targets of MAIJU recordings is to track the development of gross motor abilities. A novel index BIMS was introduced [29] as a holistic scalar measure for gross motor maturity. BIMS is a normalized (to a scale of [0, 100]), unitless measure defined as the statistical expectation of age that has most likely generated the given measurement. The concatenated recording-level posture and posture-conditioned movement distributions are used as feature vectors within Gaussian process regression [38] for the BIMS computation [30]. Our previous experiments have shown that BIMS correlates well with a conventional assessment battery of gross motor ability (Alberta Infant Motor Scale) [8].

Performance Metrics

Performance for frame-level classification is computed from a compounded confusion matrix where each individual human annotation is compared with the classifier output. For inter-rater agreement, the individual annotations are compared with each other in all combinations. The confusion matrices over all recordings are summed into a single compound confusion matrix, from which the number of true positives (tp), true negatives (tn), false positives (fp), and false negatives (fn) are obtained. Overall performance is reported as Cohen kappa with the formula:

$$k=2(tp*tn-fn*fp)(tp+fp)(fp+tn)+(tp+fn)(fn+tn)$$

Due to the unbalanced nature of the category distributions, kappa gives a more informative (compared with accuracy) correlation-like score (in [–1,1]), where a score of 0 denotes chance-level performance.

Statistical Methods

Paired t Test

Two-tailed paired *t* tests are used to test for the null hypothesis that the differences between otherwise equivalent systems' performances are zero when accounting for a single configuration variable. The results are reported with significance levels of (*P<.05; **P<.001; ***P<.0001). To justify the use of the *t* test, the Gaussianity of the differences is verified with a single-sample Kolmogorov-Smirnov test with P>.05.

Pearson r

The aggregate metrics are compared between the annotation-based and classifier-based variants with the Pearson correlation coefficient. This is due to the naturally uneven distribution of category occurrence, which makes straightforward loss metrics such as L1 or L2 loss hard to interpret. Aggregate scores include the category distributions and the BIMS score. The use of Pearson *r* is justified by a strong linear dependency between the tested variables.

Bootstrapping

The basic bootstrapping method is used to obtain 95% CIs for the accuracy, Cohen kappa, and Pearson *r* values. In each case, the sampling (10,000 iterations) is performed at the recording level when computing the compounded result. The *t* test is not used in connection with bootstrapping, where statistical significance is interpreted in terms of the overlap of the CIs.

Ethical Considerations

The study complies with the Declaration of Helsinki, and the research was approved by the Ethics Committee of Children's Hospital (number HUS-1331-2016-8; September 14, 2015), Helsinki University Hospital. A written informed consent was obtained from one of the parents of each participant.

Results

Overview

We present the results of 3 experiments to assess the effects of reducing recording configurations (placement, amount, modality, sampling frequency) to (1) high-resolution classification of posture and movement (1.15 frames per second), (2) to recording-level summary measures, and (3) to holistic assessment on infants' gross motor development. For all experiments, the classifier performance was assessed with recording-level 10-fold cross-validation.

Experiment 1: Effect of Reducing Recording Configuration to High-Resolution Classification

Overview

We trained separate posture/movement classifiers for all reduced IMU sensor inputs to study how the classifier performance

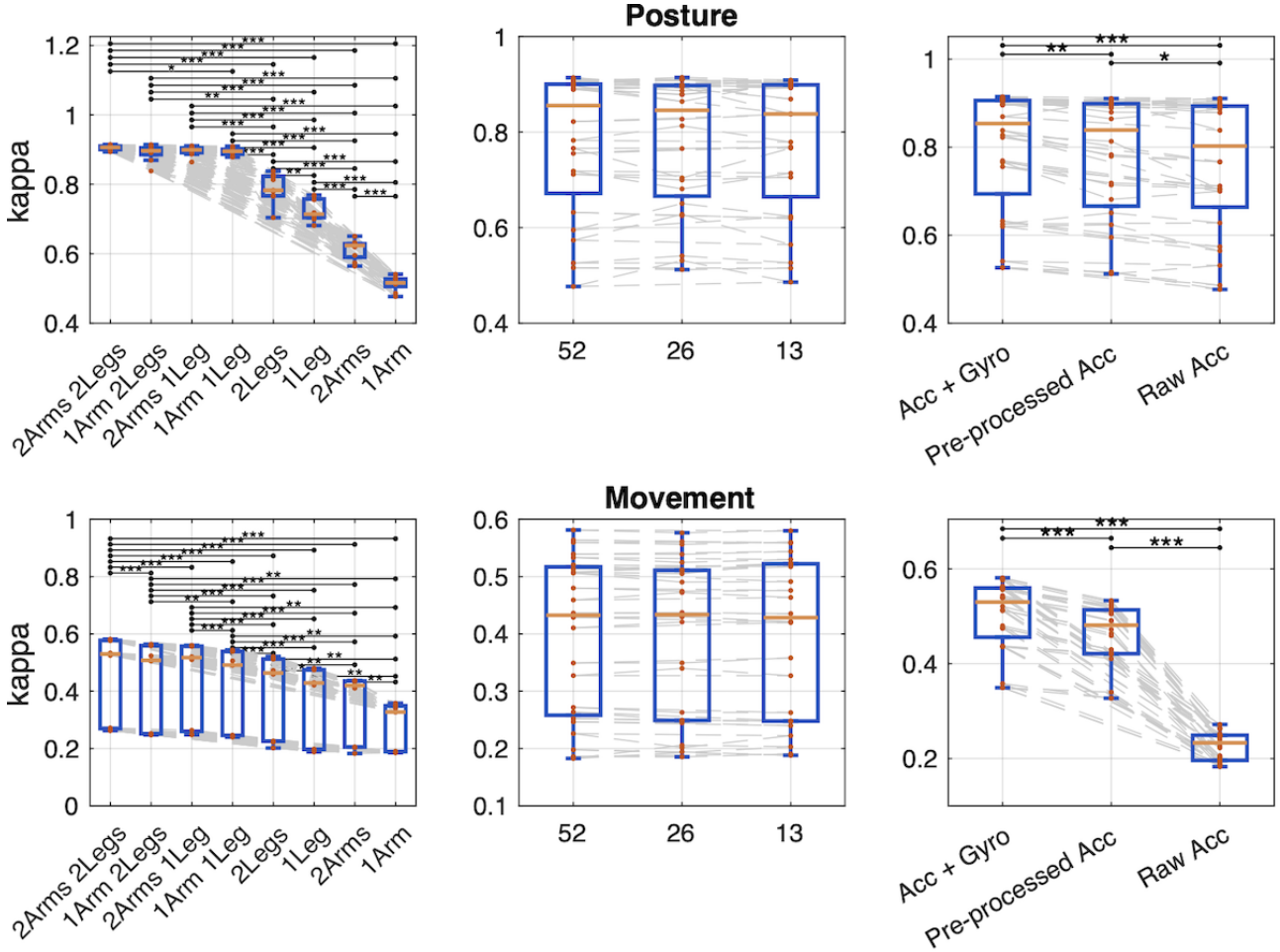
XSL•FO
RenderX

degrades as a function of compromising the IMU configurations. A total number of (8x3×3) 72 classifiers for both tasks (Figure 1) were trained with varying recording configurations, including 8 variants of sensor placement, 3 sampling frequencies (52, 26, and 13 Hz), and 3 alternatives of measurement modality (acc and gyro, raw acc, pre-processed acc) (see Methods section).

Sensor Placement

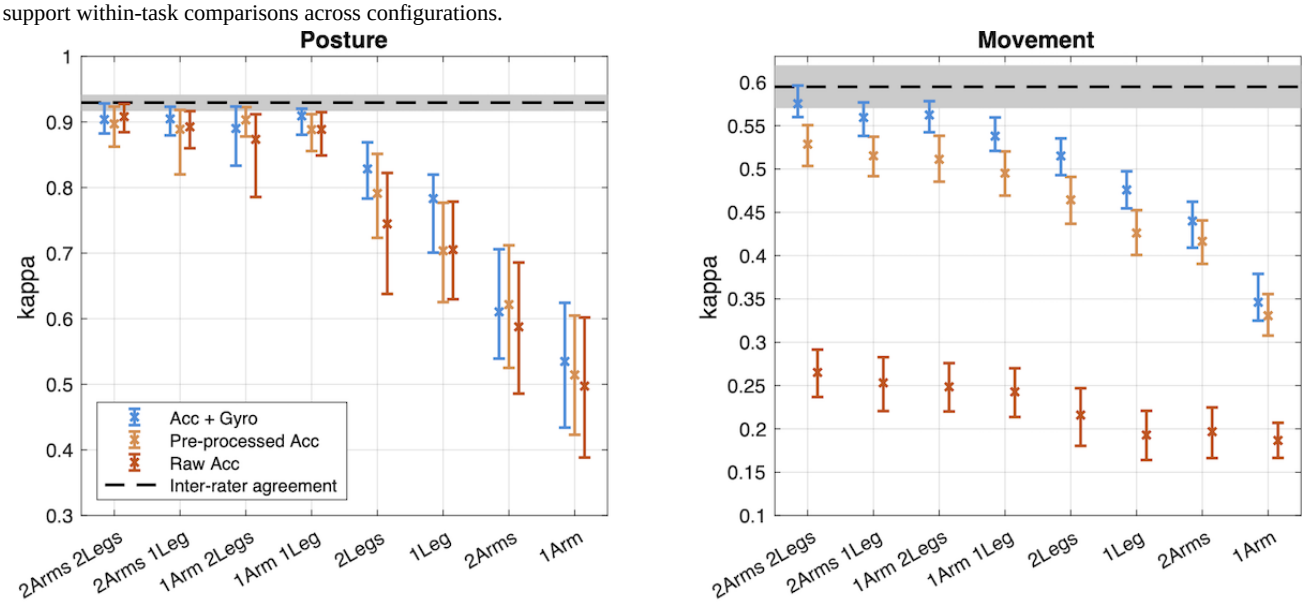
The full classifier performance results are shown in Figures 5 and 6, ordered by descending order of mean kappa levels. Notably, the same rank hierarchy of recording configurations emerges for both posture and movement detections: 2Arm2Leg > 1Arm2Leg > 2Arm1Leg > 1Arm1Leg > 2Leg > 1Leg > 2Arm > 1Arm. For postures, however, the 3 best-performing configurations are not significantly different (t_8 range 0.4 - 2; P=.08-.59). These results together indicate clearly that a combination of upper and lower limb IMU sensors yields the best classification, while IMU sensors in the legs support better classification than IMU sensors in the arms.

Figure 5. Effect of recording configuration to frame-level classification. Effect of sensor placement (left), sampling frequency (middle), and sensor modality (right) on the classification performance (Cohen kappa) for posture (top) and movement (bottom). The boxplots show the median, IQR, and range of experiment values for the cases where the given attribute is held constant. The gray lines connect systems with otherwise equivalent parameters. Two-tailed paired *t* tests are used to test for the null hypothesis that the differences between otherwise equivalent systems' performances are zero (*P<.05; **P<.005; ***P<.001). Please note the different y-axes for posture and movement to support within-task comparisons across configurations.



XSL•FO
RenderX

Figure 6. Effects of sensor modalities on the classifier performance. The findings extend the data presented in this figure by indicating sensor modalities by different colors. The 95% CIs of the performance values (kappa) are drawn with whiskers. The level of human inter-rater agreement (dashed line, 95% CIs depicted with gray) indicates the kappa levels needed to claim human-level accuracy. Please note the different y-axes for posture and movement to support within-task comparisons across configurations.



Sensor Modality

The effect of the sensor modality has the highest discrepancy between posture and movement: For posture, all 3 modalities perform rather acceptably, even though systematic differences can be observed (t_{23} range 2.2 - 3.8; P=.035 to <.001). However, for movement, the performance of the "raw accelerometer" modality plummets (t_{23} range 28 - 30; P<.001; 4 sensors from kappa=0.58 to kappa=0.27). In contrast, the pre-processed accelerometer modality manages to maintain performance in a more robust way when sensors are being reduced from the full setup, although the use of a gyroscope still results in superior performance (t_{23}=14; P<.001).

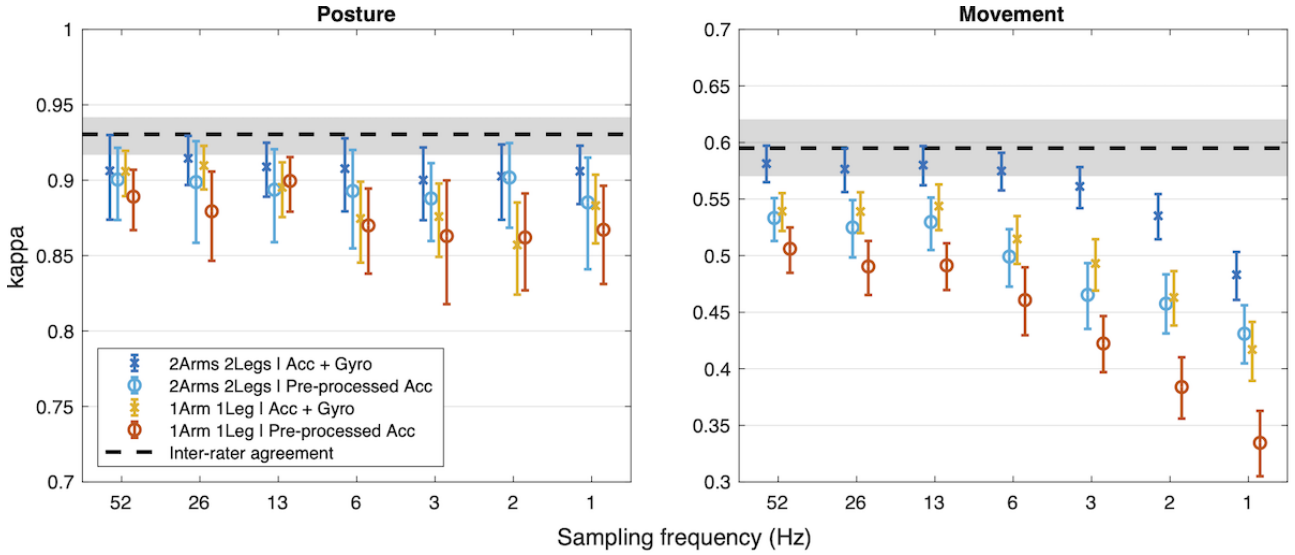
An alternative representation of results is shown in Figure 6. Based on the above findings, the sampling frequencies could be pooled together, and the means of the kappa values and their 95% CIs (obtained by bootstrapping) are presented for recording configuration and modality. For posture detections, the classifier performance remains stable until the 1Arm1Leg sensor placement, and it deteriorates when only one limb type is recorded. For movement detections, however, the overall performance deteriorates gradually with all sensor placements. The CIs overlap for the 4- and 3-sensor placements when the recordings included both acc and gyro modalities. Notably, the reference configuration with acc and gyro modalities is nonoverlapping with all of the pre-processed accelerometer modality recordings.

Sampling Frequency

There were no systematic effects of sampling frequency either for posture or movement classifications within the supported range of the used Movesense sensors (13+ Hz). The median performance ($kappa_{pos}$=0.85 and $kappa_{mov}$=0.43) was slightly higher with 26 Hz, but there were no significant differences (t_{23} range 0.1 - 1.1; P=.30-.90). Because the default sampling rate range did not present a differentiating effect for the posture and movement classification tasks, we conducted a further experiment to assess the effects of extreme downsampling with 4 candidate systems: 2Arms2Legs and 1Arm1Leg with either acc+gyro or pre-processed acc. The results are presented in Figure 7. Strikingly, the 2Arm2Leg system with a gyroscope retained unchanged performance down to a sampling rate of 6 Hz and was comparable to other full-sampled systems even at a 2 Hz sampling rate. We reason this to be the case due to 2 factors. First, even at a 1 Hz sampling rate, the raw accelerometer and gyroscope readings are, in fact, a good feature representation for orientation and movement. The accelerometer always includes the gravity vector, and the gyroscope reading is a good estimate of movement intensity. Second, the used neural network architecture takes consecutive frame-level features into account from a receptive field of 35 seconds, which significantly increases the amount of information available for each frame-level inference (ie, classification of single-frame features without surrounding context would drastically decrease performance).

XSL•FO
RenderX

Figure 7. Effect of extreme sampling rate decimation on selected configurations. Notice the surprisingly high performance in posture detection for sampling rates as low as 1Hz. Please note the different y-axes for posture and movement to support within-task comparisons across configurations.



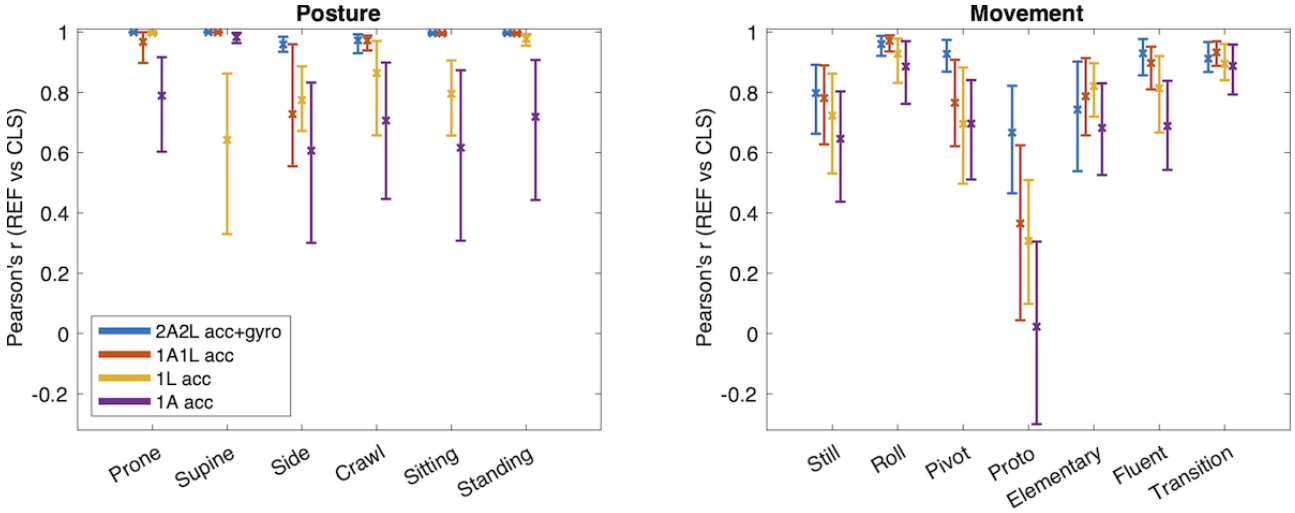
Experiment 2: Effect of Recording Configuration on Recording Level Category Distributions

Based on the result of experiment 1, we identified the following 4 recording configurations to be the most appealing in terms of maximizing some of the several measurement properties. First, (2Arm2Leg, acc and gyro, 13 Hz) as the best-performing system with the lowest supported sampling rate. Second, (1Arm1Leg, pre-processed acc, 13 Hz) as the best-performing 2-sensor configuration. Third and fourth, (1Arm and 1Leg, pre-processed acc, 13 Hz) as single-sensor configurations with large differences.

Posture and movement category distributions provide a straightforward and explainable summary of the whole recording session. The results are presented in Figure 8. For the distributions of all posture categories, the reference recording configuration was very highly correlated to the distributions obtained from human expert annotations (range r=0.96-0.999; n=41). The sensor placement with 1Arm1Leg showed also a very high correlation (r=0.96-0.999) for all but the Side posture (r=0.73). The single-sensor recordings yielded a considerably lower performance: 1Leg has range (r=0.64-0.996), with most 4 of the categories having r<0.9, and 1Arm has the worst performance with range (r=0.61-0.94), with 5 categories having r<0.8.

The correlation between classifier-based and human expert-based movement category distributions was more variable, and there was also a more gradual deterioration of performance when compromising the recording configuration. Our reference configuration showed a range of 0.67-0.96, with 4 (out of 7) movement categories having r>0.9. For 1Arm1Leg the same metrics are in the range 0.34-0.97 (2 out of 7 categories with r>0.9). Overall, the movement categories “roll”, “fluent”, and “transition” seem to be the most robust in performance, whereas the “proto” category was least correlated to the human annotations.

Figure 8. Effect of recording configuration on recording-level summary metrics. The correlations (Pearson r) between posture (left) and movement (right) distributions obtained using human annotation (REF) and those obtained from IMU recordings and classifier-based (CLS) detections.

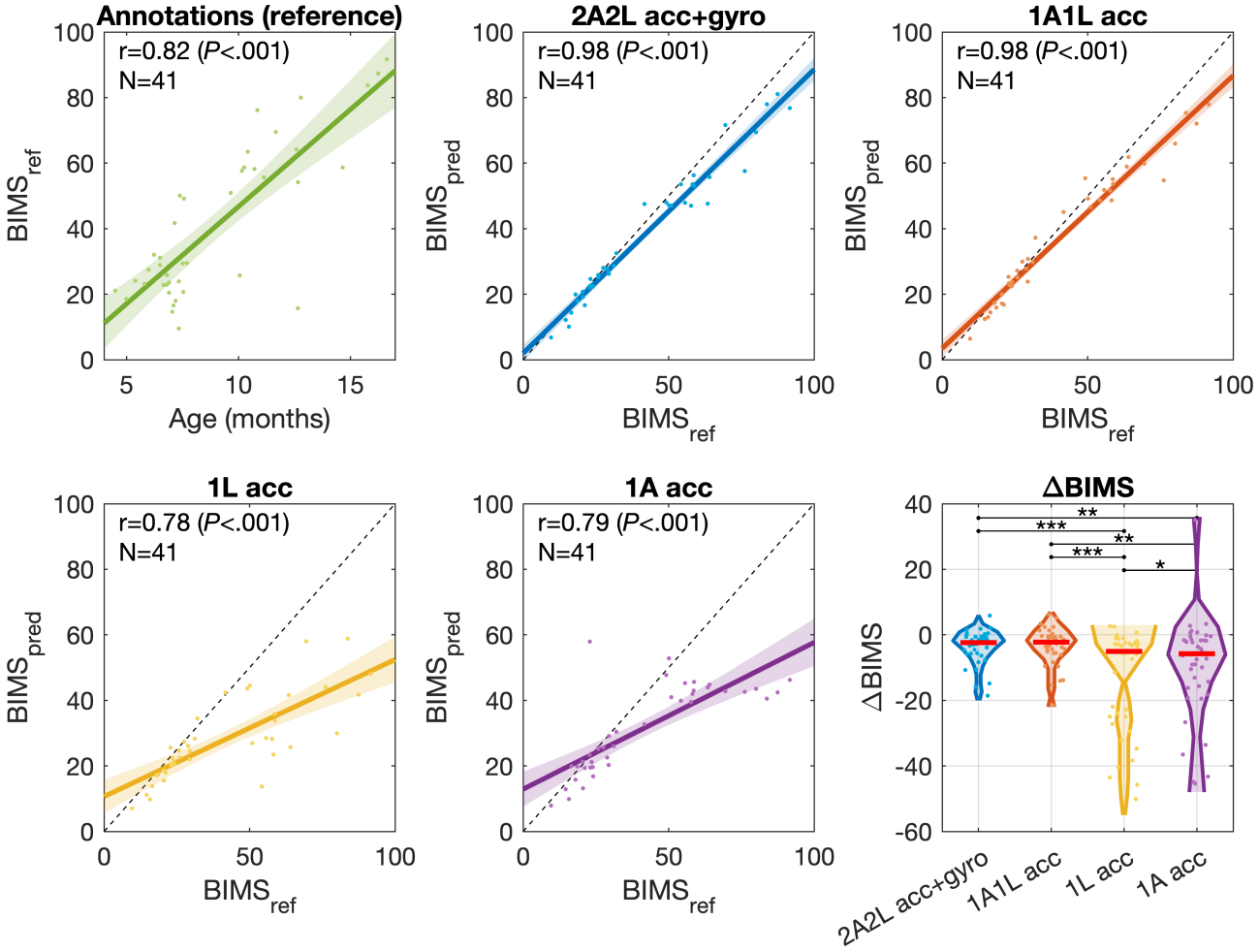


XSL•FO
RenderX

Experiment 3: Effect of Recording Configuration on Holistic Assessment of Motor Development

Figure 9 presents the effect of reducing recording configurations to a holistic measure, here the scalar metric of a motor maturity score (BIMS). The first panel (green) indicates the reference score $BIMS_{ref}$ that is obtained from the annotated distributions as a function of chronological age. The subsequent panels show the correlation of the IMU-recording-based and classifier-derived BIMS scores ($BIMS_{pred}$) to $BIMS_{ref}$. The dashed diagonal lines represent the ideal case where the scores are equal. The results show that the BIMS scores obtained from the 2Arm2Leg configuration are indistinguishable from those of 1Arm1Leg (r=0.98 for both), whereas the single-sensor configurations have clearly lower performance (r=0.78-0.79). Compared with the human expert annotations, the BIMS scores related to older age groups (higher baseline BIMS scores) are more difficult to predict, and this effect is most clearly seen in the 1-sensor variants (see bottom row of Figure 9): 2Arm2Leg median ΔBIMS=–2.4 (range –18 to 5.8), 1Arm1Leg –2.3 (range –21 to 6.8), 1Leg –5.0 (range –50 to 2.7), and 1Arm -5.7 (range –45 to 35). We assume this effect to be caused by the drastic qualitative shifts in gross motor abilities as the infants progress from lower (BIMS<50) to higher motor maturation (BIMS>50); this developmental time window is characterized by a shift from floor-based postures to increasingly upright postures and high postural variability.

Figure 9. Effect of classifier-based inputs on the BIMS score ($BIMS_{pred}$) compared with the reference ($BIMS_{ref}$). Green: Scatter plot and correlation (Pearson r) of $BIMS_{ref}$ compared with the chronological age. Blue, red, yellow, purple: Scatter plots and correlations of $BIMS_{pred}$ compared with $BIMS_{ref}$. Bottom right: Violin plots showing the $\Delta BIMS=BIMS_{ref}-BIMS_{pred}$ for the studied recording configurations. The red lines denote the median difference. Two-tailed paired t tests are used to test for the null hypothesis that the differences between systems' performances are zero (*P<.05; **P<.005; ***P<.001). BIMS: BABA infant motor score.



Summary of Results

Finally, Tables 1 and 2 present the most relevant conclusions from the experiments: Sampling frequency did not have a significant effect on the classifier performance. Sensor modality did affect classifier performance, showing the best results with a combined accelerometer and gyroscope modality. However, a pre-processed accelerometer-only recording with separate low- and high-pass heads may produce reasonably good results with a minimalist measurement setup. For sensor placement, all 4- and 3-sensor configurations reach human equivalence in all measures of results. The best-performing 2-sensor configuration is obtained with a combined 1 arm and 1 leg placement, which has slightly lower-than-human equivalent performance in all other metrics except BIMS estimation. Finally, the one-sensor configurations perform poorly in all comparisons, while the proximal leg position is clearly better than an arm placement.

XSL•FO
RenderX

Table . The key takeaways of the experiments for sampling frequency and sensor modality for posture and movement classification. The effect of the sampling frequency is negligible. Ideally, both accelerometer and gyroscope are used, but relatively good performance can be reached with the pre-processed accelerometer system.

	Posture classification	Movement classification
Sampling frequency	52 ≈ 26 ≈ 13Hz	52 ≈ 26 ≈ 13 Hz
Sensor modality	Acc and gyro ≈ Pre-pr. acc ≈ Raw acc	Acc and gyro Pre-pr. acc >> Rawacc

Table . The key takeaways of the experiments for sensor placement at various levels of analysis. The minimal system is the variant with 1Arm1Leg sensor placement, 13 Hz sampling rate, and pre-processed accelerometer modality. HEQ: human equivalence in terms of inter-rater agreement.

Sensor placement	High-resolution		Overall statistics		Holistic
	Posture	Movement	Posture	Movement	BIMS
4 (2A2L)	HEQ	HEQ	HEQ	HEQ	HEQ
3 (1A2L or 2A1L)	HEQ	HEQ	HEQ	HEQ	HEQ
2 (1A1L)	↓	↓	(↓)	↓	HEQ
1 (1L)	↓↓↓	↓↓↓	↓↓↓	↓↓↓	↓↓↓↓
1 (1A)	↓↓↓↓	↓↓↓↓	↓↓↓↓	↓↓↓	↓↓↓↓

Discussion

Principal Results

Overview

This study shows that reductions in the IMU recording configurations lead to systemic and predictable compromise at all analytical levels, ranging from high temporal resolution analyses to recording-level statistics and a holistic assessment of motor performance. (Tables 1 and 2). While the present dataset is collected from young infants during their spontaneous activity, the studied movement behaviors [28] present a natural repertoire of well-defined movements and postures at any age. We, therefore, posit that the findings on high-resolution posture and movement detections are generalizable to other age groups as well, though the lower-resolution movement behavior in the older participants is more structured and predictable. Thus, the highly variable infant data provides a stress test for the overall methodology. Most importantly, (1) sensor placement is critical and at least one lower and one upper extremity sensor is essential for all high-resolution detections of posture or movement categories, (2) gyroscope data is significant for classifying movement types, and (3) data sampling rate has no meaningful effect on analyses.

When the analyses are based on complex pipelines, such as machine learning-based classifiers, the links between compromises in data versus results escape heuristic predictions. Instead, direct experimental evidence is needed, preferably in the context of an authentic end-to-end solution, such as the MAIJU recordings for infants' motor assessment used in our work. The strong effects of recording configuration, in particular the very low analytic accuracy of single sensor recordings, may seemingly conflict with the long history of actigraphy recordings [13,17,20]. Our study targeted high-fidelity detections of qualitatively different postures and movements, whereas actigraphy analyses measure the intensity (or amount) of movements to be categorized into different activity levels, such as sedentary, light, or moderate to vigorous activities [15,16].

Our present work does not aim to claim a universal superiority of any recording configurations. Instead, our findings emphasize the tight interplay, or trade-off, between the chosen recording configurations and the attainable analyses and study questions. Another potential comparison is the range of recently launched sport and wellness wearables that use one sensor to provide the user with logs of selected activity types, such as running, walking, or cycling. The apparently successful performance of those applications is based on using very long time-epochs (minutes to tens of minutes) and analyzing only monotonous and rhythmic movement patterns. Our results from human video annotations confirmed that the behaviorally relevant time resolution needs to be in seconds rather than minutes, and a significant part of the natural movement and posture categories of daily activity are not monotonous or rhythmic, at least in the studied infant population. While the present results and the presently studied movement measures are directly generalizable to any age group, it is important to note that studies using other movement categories (ie, other analysis targets) would benefit from their respective trade-off analyses.

The main purpose of reducing the recording configuration is to simplify movement studies for wider scalability and easier analytics:

Sampling Frequency and Data Rate

We show that the data rate can be reduced without loss of analytic gain from the original setup to 13 Hz, 2-sensor acc only recordings, corresponding to a compression factor of 16 (from 1248 numbers per second; 5.0 kilobytes/s with 32-bit floats to 78 numbers per second; 0.4 kilobytes/s). Furthermore, our results suggest that an even more radical reduction of the sampling rate to 2 - 6 Hz is feasible, especially if a higher number of sensors can be maintained. This is a novel finding that might have significant ramifications for future research directions, however, the current findings are directly applicable to the specific movement and posture categories used in our study pipeline. Other HAR classification targets with a wider dynamic range of physical movements may be more sensitive to lowering the

XSL•FO
RenderX

sampling rate. Lower sampling rates improve the pipeline at all levels: First, smaller IMU data files per time enable storage of longer recordings in the sensor memory [39]. Compared with real-time data streaming, local storage is easier for the user, and it yields substantial savings on the sensor battery (–96.8 µA) [40]. Second, the transmission of the recorded data over a wireless connection becomes much faster, and in the case of continuous streaming, it improves the stability of the wireless connection (here Bluetooth 5.0 low-energy was used).

Sensor Modalities

Reduction in sensor modalities to accelerometer only has direct effects on power consumption because gyroscopes consume considerably more power than accelerometers (approximately 5 times in the used Movesense sensor; 46 µA for 13 Hz acc, 227 µA for 52 Hz acc+gyro [40]). Additionally, dropping the gyroscope will also halve the data rate.

Sensor Number and Placement

Reduction in the number of sensors without significant effects on analysis performance gives 2 kinds of advantages: First, data recorded with initially higher sensor numbers can be used even in the event of sensor dropouts, either due to an actual recording failure of individual sensors or due to rejection by quality control afterward. Second, a lower number of sensors will also imply cheaper construction and easier use of the wearables. Here, we showed that a 2-sensor configuration is feasible in most situations if the data is reliably collected from at least one upper and one lower limb; yet, additional sensors are always improving the robustness by providing redundancy in case of technical failures that may be unpredictable in real-world environments.

Limitations

While our results contain readily translatable interpretations to wider use cases, the exact experimental design implies some limitations. First, our reference IMU data were obtained using a recording setup that was previously optimized for infant studies. Other anatomic locations (eg, distal limbs, trunk, or head [41,42]) could add further insights for use cases outside of the infant gross motor assessment. Our own unpublished studies indicated that distal sensor placements would not improve analytic reliability enough to justify the extra practical challenges coming from adding distal IMU sensors in wearable suits. Additionally, distal placements would come at the cost of far higher movement complexity in most activities. Regarding trunk and hip placement, our data does not inform it, but the lateral proximal sensor location in the leg comes close to hip placement.

Second, our work focused on posture and movement classifications of infants from the age they begin moving until they walk fluently. The apparently canonical postures and movements used in our classifier task [29] are likely applicable to any age group, but other groups were not explicitly assessed. However, we posit that naturalistic IMU datasets from adults including substantial sedentary behavior could well lead to better performance with more reduced recordings [27]. Nevertheless, we suggest that our present results and the conclusions thereof should be taken primarily as recommendations for a “safely sufficient” recording setup for the universally applicable posture and movement categorization task. Finally, it is important to note that our classifiers did not extend to more subtle movement characteristics, such as symmetry or synchrony between limbs, which may be important in the medical context [43]. Such analyses, however, require symmetrical sensor placements and preferably high sampling rates, such as our 4-sensor reference configuration.

Comparison With Prior Work

Our findings are fully in line with a larger body of prior literature (both for infants [6,28] and for adults [41,42,44]) and with the intuitive reasoning that reductions in the recorded data would compromise the analytically extractable information. However, we are not aware of prior studies performing so an extensive and systematic study, including so widely ranging recording configurations and diverse analytical outputs, where data collection has been based on naturalistic observation of behavior. Our results need to be compared with prior work separately for each study dimension. Regarding sensor number and placement, our findings are fully compatible with prior studies showing that 3 to 4 sensors in infants [4,6,28], or at least 2 sensors in adults [41,42], may provide sufficient analysis of postures at lower time resolutions. Our findings extend prior work by showing that 2-arm or 2-leg only setups are not adequate. We also extend our analysis outputs to movement classification, where the gyroscope supports clearly better analytic accuracy. Regarding the resolutions of our movement classifier, our study extends significantly the prior works by using a very high temporal resolution as well as discrimination between widely varying movement types [29]. Our algorithms operate at the second-by-second level for both posture and movements, whereas prior studies have provided multisecond [4,6], or more often multiple-minute resolution [45], or even coarser than that in many consumer electronic applications. It is clear that averaging movement data over longer time windows will improve classification, but this comes at the cost of losing the natural behavioral variability. Here, we showed empirically using video-based annotations that natural behavior is characterized by altering postures and movements at the level of seconds. Thus, an ecologically valid and informative movement analysis must also work at the same time resolution. It is striking in this context that most of the published HAR research has used “acted out” data collection, where the classifiers are trained (and validated) using movement recordings where participants’ performance is instructed (eg, [4,17,42,46,47]). Such classifiers may perform seemingly well, but they grossly simplify the naturalistic human movement activity behavior, leading to significant concerns about the lack of real-world validity and applicability when using the common recording configurations and analytic approaches [27]. The need for higher quality validation protocols, improved gold-standard reference methods (eg, video), and more realistically assigned classification targets [29] is imperative when the movement recordings are applied to assessing real-world behavior using wearable systems. A recent thorough review by Giurgiu et al [27] concludes with the statement (sic): “Future research should strongly aim at biological state and posture or activity outcomes and strive for standardized protocols embedded in a validation framework”, which is precisely addressed by this article.

XSL•FO
RenderX

Conclusions

This study identified the minimalist setup for adequate posture and movement tracking as: at least a combination of one upper and one lower extremity sensor, 13 Hz sampling frequency, and at least an accelerometer, but preferably also a gyroscope. However, optimizing recording configurations always comes at a price, and striking the balance between gains and losses must be done with respect to each analysis target and the study question. For instance, a posture-based behavioral assessment may benefit from a very long recording time, even up to a week, supporting the use of the minimalist setup. However, thorough characterization of gross motor performance needs high accuracy in both posture and movement measures, while it does not benefit from excessively long recording times [29,30]. For the particular purpose of infant recordings with optimal analytic gain, it is recommended to use the full reference configuration (as in the MAIJU wearable setup [29]), including 4 sensors with both accelerometer and gyroscope modalities. The systematic assessment of analytic compromises provides future IMU sensor users with an evidence-based guide to optimize their recording configurations for each particular use case.

Acknowledgments

The authors wish to thank the Finnish CSC-IT Center for Science for the scientific computing resources used to perform the experiments. This work was supported by the Finnish Academy (314602, 335788, 335872, 332017, 343498), Finnish Pediatric Foundation (Lastentautiensäätiö), Suomen Aivosäätiö, Sigrid Juselius Foundation, and HUS Children's Hospital/HUS diagnostic center research funds. Open access was funded by the Helsinki University Library. The funding organizations had no role in the design or conduct of the study; in the collection, analysis, or interpretation of the data; in the decision to submit the article for publication; or in the interpretation, review, or approval of the article. Generative artificial intelligence was not used in any form to generate or ideate the text or the visualizations within the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. Please contact SV (sampsa.vanhatalo@helsinki.fi) for requests.

Authors' Contributions

MA, OR, and SV conceptualized the study. MA performed all the experiments, analyses, and visualizations. MA wrote the initial draft, and OR and SV critically edited and revised the final manuscript.

Conflicts of Interest

None declared.

References

1. Bruijns BA, Truelove S, Johnson AM, Gilliland J, Tucker P. Infants' and toddlers' physical activity and sedentary time as measured by accelerometry: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2020 Feb 7;17(1):14. [doi: 10.1186/s12966-020-0912-4] [Medline: 32028975]
2. de Barbaro K. Automated sensing of daily activity: a new lens into development. Dev Psychobiol 2019 Apr;61(3):444-464. [doi: 10.1002/dev.21831] [Medline: 30883745]
3. Mendoza JK, Fausey CM. Quantifying everyday ecologies: principles for manual annotation of many hours of infants' lives. Front Psychol 2021;12:710636. [doi: 10.3389/fpsyg.2021.710636] [Medline: 34552533]
4. Franchak JM, Scott V, Luo C. A contactless method for measuring full-day, naturalistic motor behavior using wearable inertial sensors. Front Psychol 2021;12:701343. [doi: 10.3389/fpsyg.2021.701343] [Medline: 34744865]
5. Salo VC, Pannuto P, Hedgecock W, et al. Measuring naturalistic proximity as a window into caregiver-child interaction patterns. Behav Res Methods 2022 Aug;54(4):1580-1594. [doi: 10.3758/s13428-021-01681-8] [Medline: 34505995]
6. Franchak JM, Tang M, Rousey H, Luo C. Long-form recording of infant body position in the home using wearable inertial sensors. Behav Res Methods 2024 Aug;56(5):4982-5001. [doi: 10.3758/s13428-023-02236-9] [Medline: 37723373]
7. Libertus K, Landa RJ. The early motor questionnaire (EMQ): a parental report measure of early motor development. Infant Behav Dev 2013 Dec;36(4):833-842. [doi: 10.1016/j.infbeh.2013.09.007] [Medline: 24140841]
8. Piper MC, Darrah J. Motor Assessment of the Developing Infant: Saunders; 1994.
9. Jablecki CK. Electromyography in infants and children. J Child Neurol 1986 Oct;1(4):297-318. [doi: 10.1177/088307388600100402] [Medline: 3036935]
10. Reich S, Zhang D, Kulvicius T, et al. Novel AI driven approach to classify infant motor functions. Sci Rep 2021 May 10;11(1):9888. [doi: 10.1038/s41598-021-89347-5] [Medline: 33972661]
11. Levin HI, Egger D, Andres L, Johnson M, Bearman SK, de Barbaro K. Sensing everyday activity: parent perceptions and feasibility. Infant Behav Dev 2021 Feb;62:101511. [doi: 10.1016/j.infbeh.2020.101511] [Medline: 33465730]

XSL•FO
RenderX

12. Fish LA, Jones EJH. A survey on the attitudes of parents with young children on in-home monitoring technologies and study designs for infant research. PLOS ONE 2021;16(2):e0245793. [doi: 10.1371/journal.pone.0245793] [Medline: 33544777]
13. Godfrey A, Conway R, Meagher D, ÓLaighin G. Direct measurement of human movement by accelerometry. Med Eng Phys 2008 Dec;30(10):1364-1386. [doi: 10.1016/j.medengphy.2008.09.005]
14. Chen KY, Acra SA, Majchrzak K, et al. Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers. Diabetes Technol Ther 2003;5(6):1023-1033. [doi: 10.1089/152091503322641088] [Medline: 14709206]
15. Clanchy K, Stanfield M, Smits E, Liimatainen J, Ritchie C. Calibration and validation of physical behaviour cut-points using wrist-worn ActiGraphs for children and adolescents: a systematic review. J Sci Med Sport 2024 Feb;27(2):92-104. [doi: 10.1016/j.jsams.2023.11.008] [Medline: 38087661]
16. Quante M, Kaplan ER, Rueschman M, Cailler M, Buxton OM, Redline S. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. Sleep Health 2015 Dec;1(4):275-284. [doi: 10.1016/j.sleh.2015.09.002] [Medline: 29073403]
17. Skotte J, Korshøj M, Kristiansen J, Hanisch C, Holtermann A. Detection of physical activity types using triaxial accelerometers. J Phys Act Health 2014 Jan;11(1):76-84. [doi: 10.1123/jpah.2011-0347] [Medline: 23249722]
18. Niazi AH, Yazdansepas D, Gay JL, et al. Statistical analysis of window sizes and sampling rates in human activity recognition. Presented at: 10th International Conference on Health Informatics; Feb 21-23, 2017; Porto, Portugal p. 319-325. [doi: 10.5220/0006148503190325]
19. Smith BA, Trujillo-Priego IA, Lane CJ, Finley JM, Horak FB. Daily quantity of infant leg movement: wearable sensor algorithm and relationship to walking onset. Sensors (Basel) 2015 Aug 4;15(8):19006-19020. [doi: 10.3390/s150819006] [Medline: 26247951]
20. Abrishami MS, Nocera L, Mert M, et al. Identification of developmental delay in infants using wearable sensors: full-day leg movement statistical feature analysis. IEEE J Transl Eng Health Med 2019;7:2800207. [doi: 10.1109/JTEHM.2019.2893223] [Medline: 30800535]
21. Wilson RB, Vangala S, Elashoff D, Safari T, Smith BA. Using wearable sensor technology to measure motion complexity in infants at high familial risk for autism spectrum disorder. Sensors (Basel) 2021 Jan 17;21(2):616. [doi: 10.3390/s21020616] [Medline: 33477359]
22. Roetenberg D, Luinge H, Slycke P. Full 6DOF human motion tracking using miniature inertial sensors. Xsens Motion Technol BV Tech Rep 2009 Jan.
23. McGrath T, Stirling L. Body-worn IMU human skeletal pose estimation using a factor graph-based optimization framework. Sensors (Basel) 2020 Dec 2;20(23):6887. [doi: 10.3390/s20236887] [Medline: 33276492]
24. Southerland JB. Activity recognition and crawling assistance using multiple inexpensive inertial measurement units. Master's thesis.: University of Oklahoma; 2012.
25. Arif M, Kattan A. Physical activities monitoring using wearable acceleration sensors attached to the body. PLOS ONE 2015;10(7):e0130851. [doi: 10.1371/journal.pone.0130851] [Medline: 26203909]
26. Gao L, Bourke AK, Nelson J. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. Med Eng Phys 2014 Jun;36(6):779-785. [doi: 10.1016/j.medengphy.2014.02.012]
27. Giurgiu M, Timm I, Becker M, et al. Quality evaluation of free-living validation studies for the assessment of 24-hour physical behavior in adults via wearables: systematic review. JMIR Mhealth Uhealth 2022 Jun 9;10(6):e36377. [doi: 10.2196/36377] [Medline: 35679106]
28. Airaksinen M, Räsänen O, Ilén E, et al. Automatic posture and movement tracking of infants with wearable movement sensors. Sci Rep 2020 Jan 13;10(1):169. [doi: 10.1038/s41598-019-56862-5] [Medline: 31932616]
29. Airaksinen M, Gallen A, Kivi A, et al. Intelligent wearable allows out-of-the-lab tracking of developing motor abilities in infants. Commun Med (Lond) 2022;2(1):69. [doi: 10.1038/s43856-022-00131-6] [Medline: 35721830]
30. Airaksinen M, Taylor E, Gallen A, et al. Charting infants' motor development at home using a wearable system: validation and comparison to physical growth charts. EBioMedicine 2023 Jun;92:104591. [doi: 10.1016/j.ebiom.2023.104591] [Medline: 37137181]
31. Airaksinen M, Vanhatalo S, Räsänen O. Comparison of end-to-end neural network architectures and data augmentation methods for automatic infant motility assessment using wearable sensors. Sensors (Basel) 2023 Apr 6;23(7):3773. [doi: 10.3390/s23073773] [Medline: 37050833]
32. Airaksinen M, Vaaras E, Haataja L, Räsänen O, Vanhatalo S. Automatic assessment of infant carrying and holding using at-home wearable recordings. Sci Rep 2024 Feb 28;14(1):4852. [doi: 10.1038/s41598-024-54536-5] [Medline: 38418850]
33. Murphy SJ, Vitali RV. Assessment of IMU configurations for human activity recognition using ensemble learning. IEEE Access 2024;12:111433-111442. [doi: 10.1109/ACCESS.2024.3440215]
34. Movesense - open wearable tech platform. Movesense Ltd. URL: https://www.movesense.com [accessed 2025-05-07]
35. Kipp M. Anvil software. URL: https://www.anvil-software.de [accessed 2025-05-07]
36. Airaksinen M, Gallen A, Taylor E, et al. Assessing infant gross motor performance with an at-home wearable. Pediatrics 2025 Apr 1;155(4):e2024068647. [doi: 10.1542/peds.2024-068647] [Medline: 40049221]

XSL•FO
RenderX

37. Gallen A, Taylor E, Salmi J, Haataja L, Vanhatalo S, Airaksinen M. Early gross motor performance is associated with concurrent prelinguistic and social development. Pediatr Res 2025 Jan 17. [doi: 10.1038/s41390-025-03832-5] [Medline: 39824938]
38. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning: The MIT Press; 2005. [doi: 10.7551/mitpress/3206.001.0001]
39. Taylor E, Airaksinen M, Gallen A, et al. Quantified assessment of infant's gross motor abilities using a multisensor wearable. J Vis Exp 2024 May 17;207(207). [doi: 10.3791/65949] [Medline: 38829118]
40. Movesense sensor power optimization. Movesense Ltd. URL: https://www.movesense.com/docs/system/power_consumption/ [accessed 2024-08-11]
41. Fortune E, Lugade VA, Kaufman KR. Posture and movement classification: the comparison of tri-axial accelerometer numbers and anatomical placement. J Biomech Eng 2014 May;136(5):051003. [doi: 10.1115/1.4026230] [Medline: 24337255]
42. Tang QU, John D, Thapa-Chhetry B, Arguello DJ, Intille S. Posture and physical activity detection: impact of number of sensors and feature type. Med Sci Sports Exerc 2020 Aug;52(8):1834-1845. [doi: 10.1249/MSS.0000000000002306] [Medline: 32079910]
43. Haataja L, Mercuri E, Regev R, et al. Optimality score for the neurologic examination of the infant at 12 and 18 months of age. J Pediatr 1999 Aug;135(2 Pt 1):153-161. [doi: 10.1016/s0022-3476(99)70016-8] [Medline: 10431108]
44. Sikandar T, Rabbi MF, Ghazali KH, Altwijri O, Almijalli M, Ahamed NU. Minimum number of inertial measurement units needed to identify significant variations in walk patterns of overweight individuals walking on irregular surfaces. Sci Rep 2023 Sep 27;13(1):16177. [doi: 10.1038/s41598-023-43428-9] [Medline: 37758958]
45. Jayasinghe U, Janko B, Hwang F, Harwin WS. Classification of static postures with wearable sensors mounted on loose clothing. Sci Rep 2023 Jan 4;13(1):131. [doi: 10.1038/s41598-022-27306-4] [Medline: 36599887]
46. Lugade V, Fortune E, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - part I: posture and movement detection. Med Eng Phys 2014 Feb;36(2):169-176. [doi: 10.1016/j.medengphy.2013.06.005] [Medline: 23899533]
47. Mokhlespour Esfahani MI, Nussbaum MA. Classifying diverse physical activities using "Smart Garments". Sensors (Basel) 2019 Jul 16;19(14):3133. [doi: 10.3390/s19143133] [Medline: 31315261]

Abbreviations

BIMS: BABA infant motor score
HAR: human activity recognition
IMU: inertial measurement unit
MAIJU: motor ability assessment of infants with a jumpsuit

Edited by L Buis, TDA Cardoso; submitted 06.03.24; peer-reviewed by B Smith, J Franchak; revised version received 03.04.25; accepted 07.04.25; published 03.06.25.

Please cite as:
Airaksinen M, Räsänen O, Vanhatalo S
Trade-Offs Between Simplifying Inertial Measurement Unit–Based Movement Recordings and the Attainability of Different Levels of Analyses: Systematic Assessment of Method Variations
JMIR Mhealth Uhealth 2025;13:e58078
URL: https://mhealth.jmir.org/2025/1/e58078
doi:10.2196/58078

© Manu Airaksinen, Okko Räsänen, Sampsa Vanhatalo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A New Wearable System for Personal Air Pollution Exposure Estimation: Pilot Observational Study

Sara Bernasconi, PhD; Alessandra Angelucci, PhD; Andrea Rossi, MSc; Andrea Aliverti, PhD

Department of Electronics, Information and Bioengineering, Politecnico di Milano, 32 Piazza Leonardo Da Vinci, Milan, Italy

Corresponding Author:
Sara Bernasconi, PhD
Department of Electronics, Information and Bioengineering, Politecnico di Milano, 32 Piazza Leonardo Da Vinci, Milan, Italy

Abstract

Background: Air pollution is a major environmental cause of premature deaths, responsible for around 7 million deaths annually. In this context, personal air pollution exposure (PAPE), the product of pollutant concentration and minute ventilation (*V'm*), is a crucial measure for understanding individual health risks. Standard exposure techniques do not address the space-time variability of air pollution, both indoor and outdoor, and the intra- and intersubject variability in *V'm*.

Objective: This study evaluates the feasibility of using a wearable body sensor network (BSN) to estimate PAPE in real-life settings, assess its capability to detect spatiotemporal variations in pollution levels, and compare inhaled dose estimates from the BSN with those from fixed monitoring stations and standard *V'm* values. The study also examines the system's usability.

Methods: The system, a BSN capturing physiological (pulse rate [PR] and respiratory rate [RR]) and environmental data, including health-affecting pollutants (particulate matter [PM] 1, PM2.5, PM10, CO_2, CO, total volatile organic compounds, and NO_2), was tested in a 4.5 km walk in Milan by 20 healthy volunteers. PR and RR collected by the system were used, together with biometric data and forced vital capacity estimations, in a model for *V'm* estimation to compute PAPE. Pollution levels were compared between morning and afternoon measurements, as well as between indoor and outdoor settings.

Results: Variations in RR were found among volunteers and at different locations for the same participant. Significant differences (P<.001) in pollutant concentrations were observed between morning and afternoon for CO_2 (higher in the morning) and PM (higher in the afternoon). Spatial variability along the walking path was also detected, highlighting the system's high spatiotemporal resolution. Indoor environments showed high variability in CO_2 and total volatile organic compounds, while outdoor settings exhibited elevated and variable PM levels. The mean PAPE to PM2.5 estimated with tabulated *V'm* and fixed station data was 13.31 (SD 4.16) µg while the one estimated with the BSN was 16.27 (SD 9.78) µg, 2.96 µg higher (22.3%; 95% CI −6.55 to 0.63; P=.05) than the former one, and with a broader IQR. Nevertheless, the 2 estimation methods show a good and strongly significant correlation (r=0.665; P<.001). The system's usability was generally rated as good.

Conclusions: The BSN provides high-resolution spatiotemporal data on personal exposure, capturing differences in pollution levels dependent on time, location, and surrounding environment, along with individual physiological variations. It offers a more accurate estimation of inhaled dose in real-life settings, supporting personalized exposure assessments and potential applications in activity planning and complementing epidemiological research.

(JMIR Mhealth Uhealth 2025;13:e60426) doi:10.2196/60426

KEYWORDS

air quality; body sensor network; personal exposure; cardiorespiratory monitoring; wearable system

Introduction

Air pollution is the main environmental trigger for premature deaths worldwide. A total of 9 out of 10 people breathe highly polluted air, as stated by the Organization for Economic Co-operation and Development Outdoor and household air pollution causes around 7 million deaths every year, 90% of which occurs in low- and middle-income countries because of open fires and stoves used for cooking purposes [1].

Reducing environmental pollution is a widespread concern and the United Nations has included it among the Sustainable Development Goals, more specifically in relation to the third goal about good health and well-being.

In this alarming scenario, accurate monitoring of the air we breathe is fundamental, both from an individual point of view, as a way of enhancing awareness and promoting a healthy lifestyle, and from a collective point of view, to foster new public policy oriented towards improving air quality. For instance, in the context of public intervention, monitoring and control over air pollution are among the foundation stones of smart cities [2,3].

XSL•FO
RenderX

Nowadays, there are various technologies for monitoring air pollution. Fixed stations are located in various geographical areas to assess the environment, which in turn is characterized by different predominant emission sources and different distribution and density of buildings. In addition, in some countries, such as Italy, mobile stations are used for a predetermined period (eg, 2 weeks) to investigate pollution variability during the year and examine specific situations, following notification by the municipal administrations [4]. Moreover, satellite observations help identify large areas of pollution. For instance, the Sentinel-5 Precursor satellite from the European Space Agency produces one image per day for different pollutants. These measurements are crucial for protecting public health, preserving the environment, guiding policy decisions, mitigating climate change, and promoting sustainable development, but operate at a macro-geographical and population level and give little insight into more specific data at the individual level.

Generally, studies conducted at the individual level approximate a person's level of exposure to various pollutants to the outdoor pollution level at their residential address. As demonstrated by Lu [5], ignoring human motion and spatiotemporal variability of air pollution could lead to differential exposure misclassification, potentially biasing health risk assessments. This happens because standard monitoring techniques are not sufficient to account for the great space-time variability of pollution [6]. Indeed, air quality can drastically change over short distances, lower than 100 m [7], and short time intervals [8] due to its pronounced dependence on both pollution sources and air circulation. As an example, "urban street canyons", that is, streets lined with tall buildings that impede air circulation, can give rise to a pollution scenario that can be completely different from the surrounding one.

As mentioned before, indoor air quality must be monitored as well, as demonstrated in multiple research studies [9,10]. In fact, people spend between 80% and 90% of their time in closed environments and the World Health Organization reports 1.6 million premature deaths caused by indoor air pollution [11].

However, focusing solely on indoor exposure results in a significant underestimation of exposure levels [5]. For example, particulate matter [PM] 2.5 exposure measured exclusively in indoor environments is estimated to be 61% lower than actual exposure for commuting distances exceeding 30 km. For these reasons, a personal air quality monitor is suggested to accurately assess exposure to pollutants, taking into account the participant's mobility between indoor and outdoor environments [12].

Various companies have developed portable systems capable of monitoring key air pollutants [6]. Among them, it is worth mentioning the Atmotube PRO (Atmotech Inc), which provides continuous monitoring of PM and volatile organic compounds (VOCs), and Flow (Plume Labs) which measures PM, VOCs, and nitrogen dioxide (NO_2). Further examples include the Wynd Air Quality Tracker (WYND), which specializes in PM monitoring, and Aeroqual's portable air quality monitor (Aeroqual), which supports the integration of different sensors in a limited configuration.

Nevertheless, for an accurate assessment of the pollution that effectively infiltrates human organisms, the respiratory activity of the participant assumes a crucial role. To consider the whole phenomenon, personal air pollution exposure (PAPE) is the appropriate measure, as proposed by Arano et al [8]. PAPE is evaluated considering a respiratory factor and an environmental factor. It is calculated with the following formula:

$$(1) PAPE(p)=SZ(p)*V'm$$

Where V′m is minute ventilation in m^3/min and SZp is the fully aggregated pollution value in a certain period *p*, calculated as:

$$(2) SZp=\sum t=1n12(Zti+1+Zti)(ti+1-ti)$$

Where Zti+1 and Zti are the specific pollutant concentrations in the time instants of ti+1 and ti, with ti+1−ti should be lower than 10 minutes.

Hence, a wearable system following the person everywhere, monitoring both indoor and outdoor environments and the cardio-respiratory activity of the person, can represent a robust solution for the accurate assessment of PAPE. In this study, we have carefully structured our experimental design to systematically address multiple research questions. Our primary objective is to evaluate the feasibility of using a body sensor network (BSN) of wearable devices to estimate personal pollutant exposure in real-life settings. To validate the performance of the BSN, we compare the inhaled dose [13] calculated from the system's measurements with the dose derived from fixed monitoring station data and tabulated Vm′ values. Furthermore, we assess the system's ability to detect variations in pollution levels across different times and locations. Additionally, we seek to determine the usability of the BSN in both indoor and outdoor environments. These distinct yet interconnected objectives are central to advancing our understanding of the system's potential applications and limitations. In section II, the novel system and the experimental protocol involving 20 healthy participants are detailed. Results about pollutant distributions and evaluation of PAPE to air pollutants are reported and discussed in section III. In section IV conclusions are drawn, highlighting limitations and future developments.

Methods

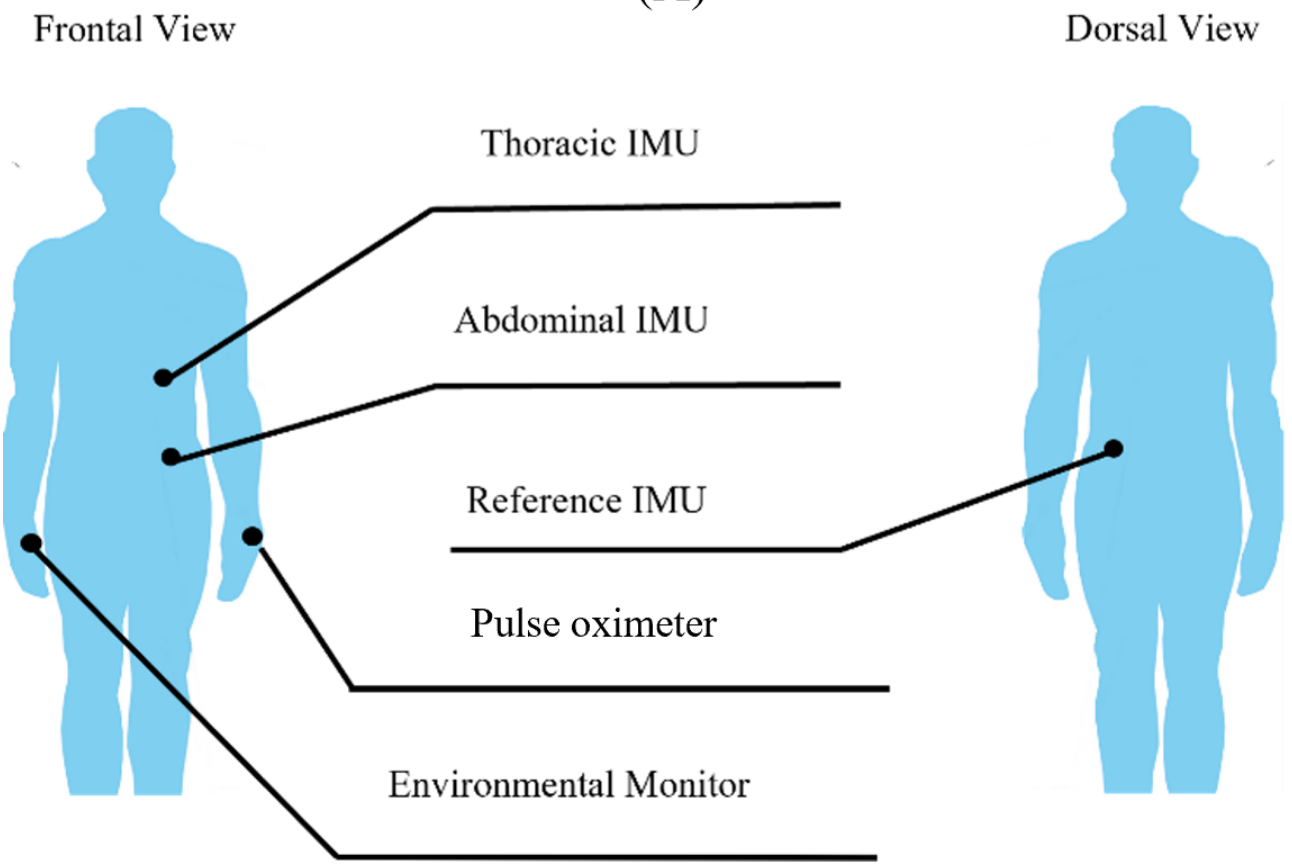
Environmental Monitor and Cardio-Pulmonary Monitoring Platform

The developed wearable environmental monitor, shown in Figure 1A, is designed to be worn on the wrist or on the arm to track the pollutant concentrations to which the participant is exposed [14].

XSL•FO
RenderX

Figure 1. (A) Wearable device for environmental monitoring; (B) Body sensor network composed of 3 IMUs, a wrist-worn pulse oximeter, and a wrist-worn environmental monitor. IMU: inertial measurement unit.

(A)



(B)

The monitored pollutants are PM (specifically PM1, PM2.5, and PM10; where the number indicates the diameter of the particle in µm), CO_2, total volatile organic compounds (tVOCs), carbon monoxide (CO), and NO_2. The PM sensor is the SPS30 by Sensirion (accuracy of ±10 µg/m^3 for PM1 and PM2.5, ±25 µg/m^3 for PM10) [15], the CO sensor is the SCD41 (±40 ppm of accuracy) [16], the tVOC sensor is the SGP30 (1.3% of measured value with Ethanol and H_2) [17], and the CO and NO_2 sensor is the MICS-6814 by SGX [18]. Atmospheric parameters such as temperature, humidity, and pressure are also detected using the Bosch BME280 sensor. The environmental monitor node is integrated into a BSN, shown in Figure 1B, aimed at collecting cardio-respiratory data. The BSN is composed of 3

XSL•FO
RenderX

devices each embedding an inertial measurement unit (IMU), positioned one on the thorax, one on the abdomen, and a reference one on the back [19-21], and a wrist-worn pulse oximeter [22], in addition to the environmental monitor.

During the experimental campaign, data recorded by all the nodes of the BSN were exploited to evaluate the PAPE. The sample frequency of the 3 IMUs and the pulse oximeter is 10 Hz for each of them and data are sent through the ANT communication protocol in a shared channel network configuration. Instead, the environmental monitor has a sampling frequency of 0.042 Hz and sends a new set of data every 24 s in an independent channel for data transmission to the master which, in our case, is a smartphone. An Android application displays environmental data in real time and geolocates them using the integrated smartphone GPS. Both environmental and physiological data are saved in a cloud-hosted database (ie, Google Firebase) for later analysis.

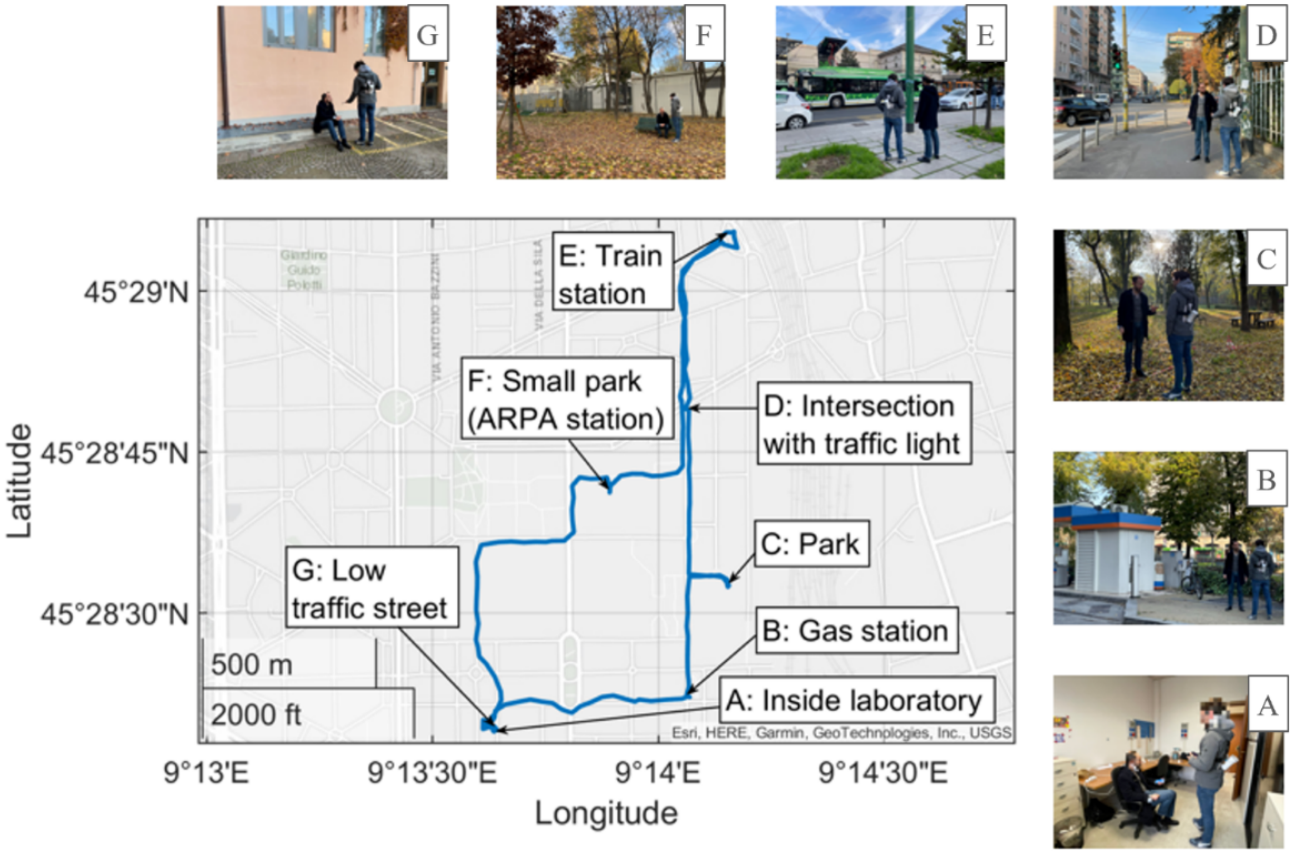
Inclusion and Exclusion Criteria

Volunteers were eligible to participate if they were between 18 and 75 years of age. However, individuals were excluded if they exhibited any physical deformities in the upper limbs that might compromise the validity of the data obtained. Anyone with gait impairments likely to interfere with the ability to complete the study protocol was also excluded.

Experimental Campaign

The experimental protocol consisted of a walk over 4.5 km in the university area of Politecnico di Milano, passing through some specific places. The full path is reported in Figure 2. Two different types of acquisition were performed as follows: static and dynamic. Static acquisitions consisted of 5 minutes of recording of the overall system in a static condition in 7 acquisition points. Each point represents a distinct microenvironment: a laboratory (A), a gas station (B), a park (C), an intersection with traffic light (D), a train station (E), a small park (F) in which there is also a fixed station of the Lombardy regional agency for environmental protection (Agenzia Regionale per la Protezione Ambientale; ARPA) and a low traffic street (G). Participants were seated in points A, F, and G, and standing in points B, C, D, and E. Dynamic acquisitions involved only the environmental monitor and consisted of the recording of pollutant concentration while walking from one static point to the other. Most acquisitions were performed in November 2022. Two time slots per day were available, one in the morning around 10:00 AM and one in the afternoon around 2:30 PM. This choice was made to avoid peak traffic hours and keep as much as possible the same conditions during all the acquisitions.

Figure 2. Route of the experimental campaign with static acquisition points highlighted. They are (A) inside a laboratory, (B) near a gas station, (C) in a park, (D) at an intersection with traffic lights, (E) near a train station, (F) in a small park where there is also an ARPA station, and (G) in a low traffic street. ARPA: Agenzia Regionale per la Protezione Ambientale.



XSL•FO
RenderX

Preprocessing and Normalization of Environmental Data for Comparative Analysis

Environmental data were resampled with a sampling frequency of 1 Hz. Then, the entire acquisition was segmented using time stamps registered during the acquisition protocol when arriving at and leaving each acquisition point. This process generates 7 arrays for static conditions and 7 arrays for dynamic acquisition for each participant, serving as a means of space normalization to facilitate the comparison of acquisitions across different volunteers. In fact, given the hypothesis of variability in pacing among volunteers and that an uncontrolled environment is susceptible to random events (eg, red traffic lights or crowds affecting walking speed), this approach ensures a more consistent basis for comparison.

A linear interpolation that takes 3 samples per minute is applied to obtain the same number of samples for each participant in each segmented array.

After these processing steps, different pollutants and their variation along the path were plotted and visually inspected. Considering that ARPA provides a single daily value for PM2.5 and PM10, our comparison with data recorded by the fixed station (F) focused on acquisitions conducted in the morning and afternoon of the same day.

Personal Air Pollution Exposure

The assessment of the inhaled dose, or PAPE, involved the use of Equation (1) and was computed for PM1, PM2.5, PM10, tVOC, and CO_2. The minute ventilation in (1) is computed as the product of the respiratory rate (RR) and tidal volume. Since tidal volume information are not provided by the system, a model proposed by Greenwald et al [23] was instead deployed for the respiratory factor. This model is in the form of a power function, reported in (3), based on a large, pooled dataset, with a median percent error of 1.20% (IQR 37.9%).

$$(3)\quad V'm = e^{-8.75} HR^{1.72} RR^{0.611} age^{0.298} sex^{-0.206} FVC^{0.614}$$

Where V′m is the value of minute ventilation expressed in L/min, HR is the heart rate in beats per minute, and RR is the respiratory rate in breaths per minute, and FVC is the forced vital capacity in L. Pulse rate (PR) data are recorded by the pulse oximeter and are used instead of HR, whereas the 3 IMUs monitor RR. In this paper, FVC was evaluated using the Global Lung Function Initiative method from the European Respiratory Society [24]. The remaining variable is sex (1 for men, 2 for women).

For what concerns the environmental factor, it consists of a mean of the values acquired during 1 minute of measurement for all 5 minutes in the different static acquisition points. PAPE to PM2.5, assessed using data obtained from wearable sensors, was compared with what is hereafter referred to as the “Standard Method.” This method estimates PAPE based on tabulated values for V′m, and daily pollutant concentration data from ARPA fixed monitoring stations. Specifically, in the Standard Method, V′m values were derived from tables considering age range, sex, and activity level [25].

In our study, we opted to align with an activity level of 3, characterized by METs typically falling within the range of 2 to 4, as an approximation to our experimental conditions. In fact, the trial encompassed brisk walking, an activity associated with a MET value of 5, and standing and seated positions, corresponding to MET values of 1.8 and 1.3, respectively. The tabulated minute ventilation for an activity level of 3 is 13.26 L/min for women and 15.14 L/min for men. These values were used as respiratory factors in the Standard Method.

Questionnaire

At the end of the experimental protocol, a questionnaire, reported in Table 1, was verbally administered by the principal investigator or his collaborators to the participants to assess satisfaction with device usability. Answers were recorded, transcribed, and then the audio recordings were deleted. The recording was convenient as it allowed the participants to walk while answering, ensuring that the process did not interrupt the experimental activities and reducing the overall time required for the protocol.

Table . Items of the wearability questionnaire.

Item number	Item	Ideal scores
I1	I think the device is simple to wear	5
I2	I think I would be able to use the device autonomously	5
I3	I found the fixing method uncomfortable	1
I4	I would have preferred to remove the device during the test	1
I5	I think I could wear the device for a long time	5
I6	I think using the device would negatively affect my daily activities	1
I7	I think I would be able to use the smartphone app autonomously	5
I8	I think the data shown are easy to read	5
I9	I think the device has a functional design	5

XSL•FO
RenderX

Participants were asked to give a score between 1 (completely disagree) and 5 (completely agree) to the 9 statements. The collected responses were compared with the ideal scores (reported in Table 1) to evaluate satisfaction with the device, and the results were visualized using a radar plot.

Statistical Analysis

First, the statistical difference in pollutants' concentration and physiological parameters between the 7 static acquisition points was assessed. After checking the normality of the distributions through the Shapiro-Wilk test, a Friedman test was performed. If the null hypothesis was rejected, a post-hoc test was performed to assess between which acquisition points there was a statistical difference in pollutants' concentration and physiological parameters and the entity of this difference.

Then, concentrations of the morning and afternoon acquisitions of the same day were compared by means of a Wilcoxon signed rank test, aiming to assess potential statistical differences between these time intervals.

RR of the different participants was analyzed using Kruskal-Wallis one-way ANOVA on Ranks to determine the presence of any statistical differences. This assessment aimed to gauge whether a standard value of RR could effectively serve as a proxy for the actual real-time RR recorded during the protocol.

Ethical Considerations

The experimental protocol was approved by Politecnico di Milano Ethics Committee (approval number 39/2022; date of approval October 10, 2022) and involved 20 healthy volunteers, who gave their informed consent. Participants retained the right to withdraw from the study protocol at any time, as outlined in the informed consent. Participant data were safeguarded through a process of pseudonymization, whereby each participant was assigned a unique identifier, with the key linking identifiers to individual identities securely maintained by the principal investigator. Participants were recruited through word of mouth and did not receive any financial compensation for their participation.

Results

Participants

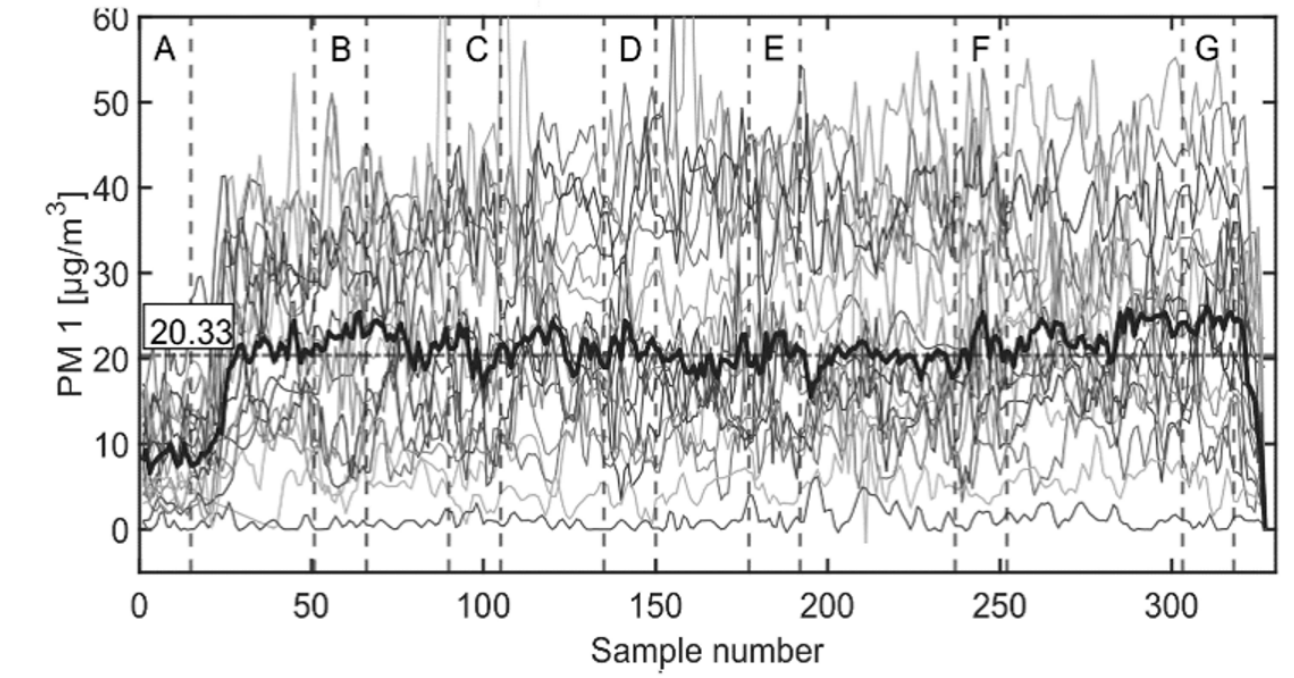
The population consisted of 14 male and 6 female participants. The median of the age distribution was 25.5 years with an IQR of 3 years, with 3 outliers at 32, 46, and 55 years. The median body weight was 70 (IQR 67.5-76.5) kg and the median height was 179 (IQR 168.0-180.5) cm.

Pollutant Concentration Trends: Indoor Versus Outdoor and Temporal Comparisons

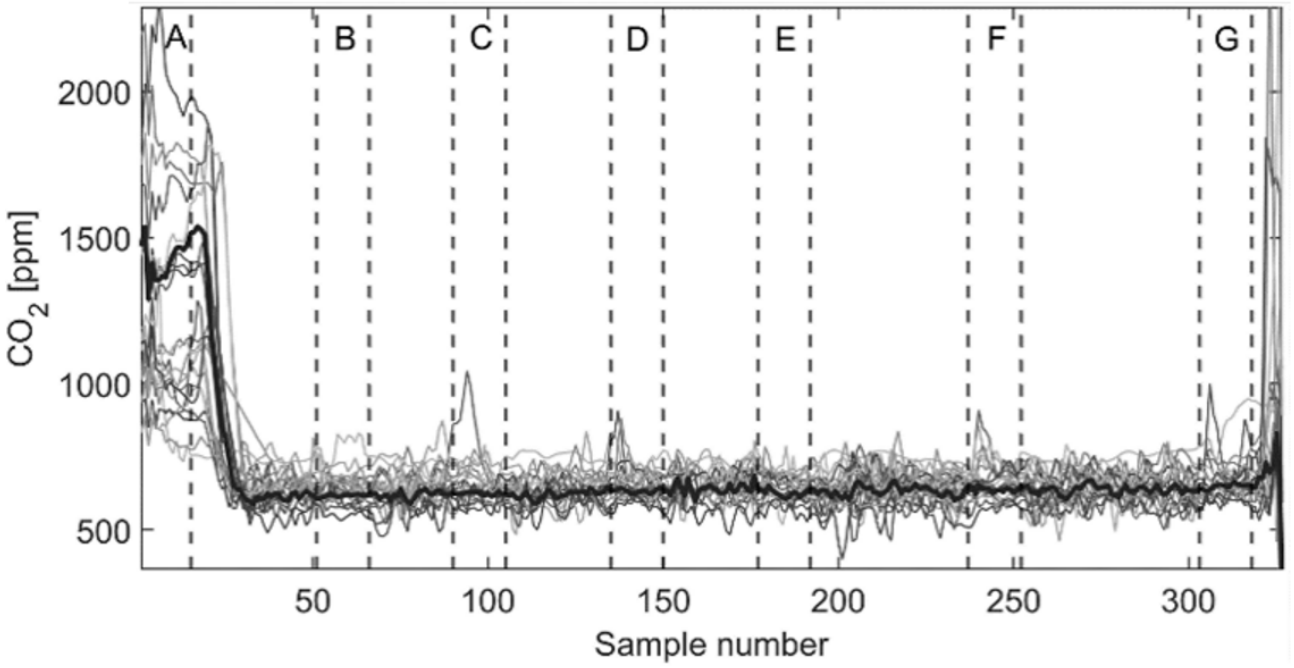
In Figure 3A and B, PM1 and CO_2 concentrations, respectively, of the whole acquisitions are reported, with the median curve represented by the bold line. As expected, there is a strong variation moving from the indoor environment (acquisition point A, inside a laboratory) to the outdoor environment (all other acquisition points). Specifically, PM1 exhibits a significant increase outdoors, whereas CO_2 shows the opposite trend. Inside the laboratory, significant variability and numerous high peaks of CO_2 can be observed.

XSL•FO
RenderX

Figure 3. PM1 (A) and CO_2 (B) concentration along the route for the 20 participants. The median is represented by the bold line. The vertical dashed lines show the 7 specific acquisition points, denoted by capital letters A through G: point A, inside a laboratory; point B, near a gas station; point C, in a park; point D, at an intersection with traffic lights; point E, near a train station; point F, in a small park where there is also an ARPA station; and point G, in a low traffic street. ARPA: Agenzia Regionale per la Protezione Ambientale; PM: particulate matter.



(A)



(B)

Substantial variability in PM was observed for the outdoor segment of the acquisition path. This variability is evident in Figure 4, which illustrates PM2.5 concentrations for 2 participants tested on the same day during the morning (Figure 4A) and the afternoon (Figure 4B). For instance, it can be noted that along the same street, from acquisition point B to E, concentrations span from very low (0 - 12 µg/m^3) to very high (55 - 150 µg/m^3) values. Moreover, Figure 4 also shows a generally lower concentration of PM2.5 in the morning with respect to the afternoon for a representative day. All these considerations are true also for PM1 and PM10.

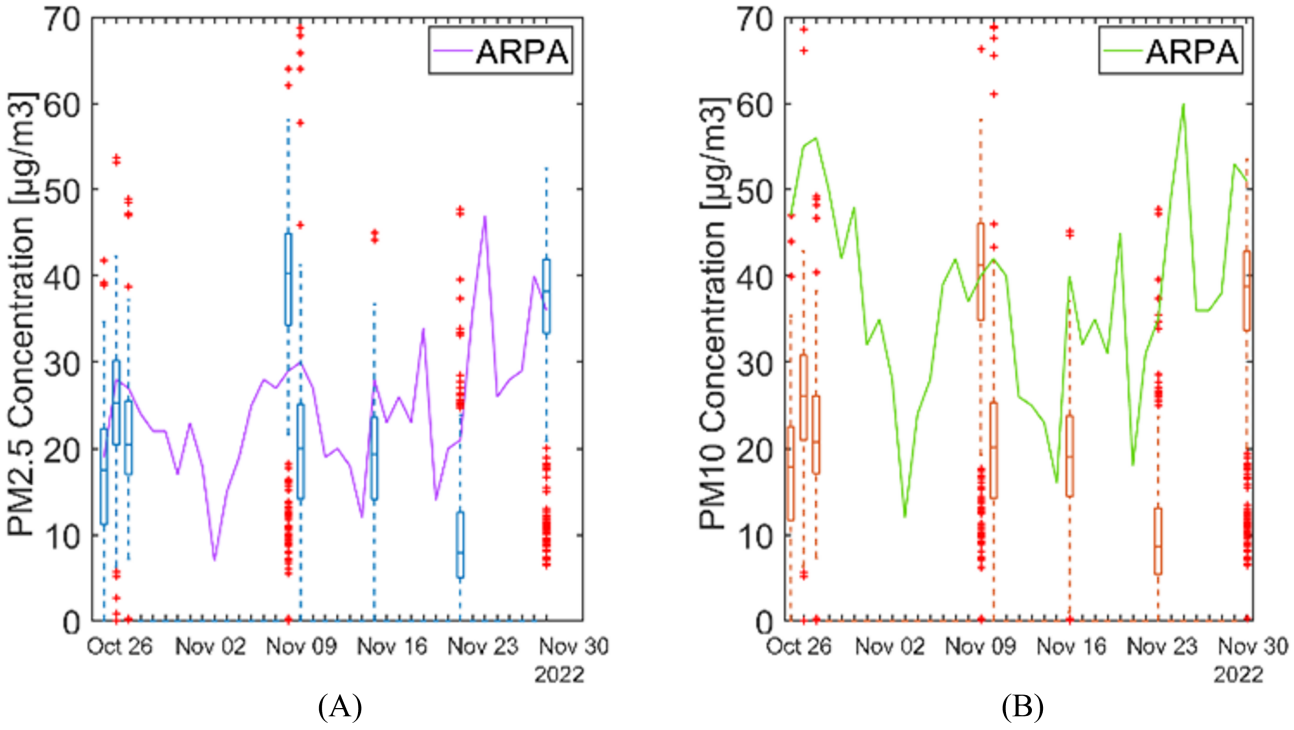
XSL•FO
RenderX

In Figure 5, boxplots of the new device measurements and a solid line for ARPA values are reported. It can be observed that concentrations captured by the new device are often lower than the mean value of the day reported by ARPA.

NO_2 and CO levels never overcame the detection limit of 20 ppb during our acquisitions.

Figure 4. Map of the PM2.5 concentration during (A) morning acquisition and (B) afternoon acquisition of 2 participants, tested the same day. The 7 specific acquisition points are denoted by capital letters A through G: point A, inside a laboratory; point B, near a gas station; point C, in a park; point D, at an intersection with traffic lights; point E, near a train station; point F, in a small park where there is also an ARPA station; and point G, in a low traffic street. ARPA: Agenzia Regionale per la Protezione Ambientale; PM: particulate matter

Figure 5. PM2.5 (A) and PM10 (B) concentrations' comparison between ARPA values taken from the fixed station in point F (continuous line) and data recorded by our device when 2 participants were tested on the same day (boxes). ARPA: Agenzia Regionale per la Protezione Ambientale; PM: particulate matter.



Quantitative Analysis of Pollutant Concentration and Physiological Parameters

For each pollutant, the Friedman test showed a statistical difference between acquisition points. In fact, for each gas and particulate, there was a statistical difference between the indoor environment and the outdoor environment. A higher value of CO_2 and tVOCs was found inside the laboratory with respect to all other outdoor acquisitions. Instead, PM was higher outdoors and lower indoor. Medians and IQRs for each pollutant together with RR and PR in the different acquisition points are reported in Table 2.

XSL•FO
RenderX

Table . Concentrations of CO_2, total volatile organic compounds (tVOCs), particulate matter (PM) 1, PM2.5, PM10, respiratory rate (RR), and pulse rate (PR) in the 7 specific acquisition points.

	CO_2 (ppm), median (IQR)	tVOCs (ppb), median (IQR)	PM1 (µg/m^3), median (IQR)	PM2.5 (µg/m^3), median (IQR)	PM10 (µg/m^3), median (IQR)	RR (breaths/min), median (IQR)	PR (beats/min), median (IQR)
A	1010.4 (214.2)	586.1 (207.5)	8.4 (7.8)	9.9 (8.6)	10.4 (8.9)	24.0 (16.9)	86.7 (13.1)
B	615.9 (56.3)	167.7 (215.0)	22.7 (22.8)	24.2 (16.0)	24.6 (16.9)	17.1 (16.6)	98.8 (19.2)
C	628.7 (73.5)	161.8 (254.2)	19.7 (23.3)	23.0 (17.5)	23.3 (17.3)	14.3 (9.67)	98.1 (18.6)
D	619.4 (74.7)	160.4 (259.9)	21.5 (24.8)	21.7 (16.9)	22.0 (18.9)	14.6 (9.6)	97.7 (19.0)
E	628.5 (72.0)	86.1 (134.7)	20.7 (23.5)	23.3 (21.9)	23.7 (21.6)	14.3 (9.7)	105.1 (21.0)
F	626.6 (64.6)	134.5 (319.6)	21.7 (20.2)	21.8 (17.0)	22.1 (17.1)	11.1 (8.6)	99.0 (21.5)
G	650.0 (79.1)	205.8 (440.0)	24.4 (17.8)	26.2 (15.1)	26.6 (15.1)	11.3 (9.6)	101.0 (25.0)

In addition to the significantly different results between indoor and outdoor acquisition points, other statistically significant differences were detected between the distributions. For CO_2, the acquisitions in point G, outside the laboratory in the parking lot near a low-traffic street, differ with a $P<.05$ from all the other acquisition points. The great variability of the indoor CO_2 concentration, assessed visually in Figure 1, was confirmed also by a quantitative analysis, showing a wider IQR in the indoor acquisition of all participants with respect to all the other outdoor acquisitions. For what concerns tVOCs, point E near the train station, which is highly ventilated, shows significantly lower values with respect to point G ($P<.001$). PM concentration was significantly higher in point G ($P<.001$ for all PMs) with respect to the gas station (B), the park (C), and the intersection with traffic lights (D).

About the comparison between morning and afternoon acquisitions, significantly higher values of concentration were found in the afternoon ($P<.001$) for all the pollutants except CO_2, which was higher in the morning ($P<.001$).

Regarding physiological parameters at various acquisition points, several noteworthy observations emerge.

The median RR at the initial acquisition point (A) was notably higher compared with all others, exhibiting a significant difference from points C, E, F, and G (with P values respectively $P_{AC}=.002$, $P_{AE}=.001$, $P_{AF}<.001$, $P_{AG}=.001$). In points C, D, and E, RR values are slightly higher than those in points F and G, although not significantly except for the couple CF ($P_{CF}=.049$, $P_{CG}=.33$, $P_{DF}=.26$, $P_{DG}=.30$, $P_{EF}=.647$, $P_{EG}=.94$).

Turning our attention to PR data, a remarkable disparity exists between point A and all other acquisition points ($P<.001$). Both RR and PR values showed great variability among participants, confirmed also by the Kruskal-Wallis one-way ANOVA on Ranks ($P<.001$ in both cases).

Inhaled Dose

A participant-specific comparison of PAPE to PM2.5 computed with tabulated data for V′m of an activity level of 3 (13.26 L/min for women and 15.14 L/min for men) and ARPA daily concentration values (indicated by the subscript “STD”), and the assessment derived from the wearable system (indicated by the subscript “NEW”), is reported in Table 3.

XSL•FO
RenderX

Table . Particulate matter 2.5 (PM2.5) inhaled dose evaluated with the standard method (STD) and with the wearable system (NEW).

P#	$V'm^{NEW}$ (L/min), median (IQR)	$V'm_{STD}$ (L/min)	C^{a}_{NEW} (µg/m^3), median (IQR)	C_{STD} (µg/m^3)	ID^{b}_{NEW} (µg)	ID_{STD} (µg)
P01	20.24 (7.96)	13.26	36.52 (5.05)	17	24.93	7.89
P02	13.64 (6.64)	15.14	9.84 (9.88)	19	5.59	10.07
P03	20.35 (6.1)	15.14	20.77 (5.73)	19	15.18	10.07
P04	14.43 (4.09)	13.26	20.87 (4.36)	28	10.41	12.99
P05	24.75 (13.23)	15.14	31.07 (6.27)	28	31.41	14.84
P06	15.71 (9.14)	13.26	19.35 (7.4)	27	12.33	12.53
P07	23.32 (14.97)	15.14	22.27 (6.72)	27	23.81	14.31
P08	20.05 (12.86)	15.14	1.24 (0.96)	7	1.07	3.71
P09	21.31 (10.99)	15.14	37.04 (8.22)	29	26.81	15.37
P10	14.77 (7.14)	15.14	44.91 (7.24)	29	24.32	15.37
P11	18.82 (7.84)	15.14	12.51 (4.36)	30	9.34	15.90
P12	18.19 (6.71)	13.26	26.73 (8.9)	30	16.00	13.92
P13	22.73 (5.71)	15.14	20.14 (7.65)	28	14.83	14.84
P14	30.9 (12.45)	15.14	14.06 (10.2)	28	16.11	14.84
P15	8.72 (3.73)	13.26	9.7 (4.05)	20	3.06	9.28
P16	16.3 (8.44)	15.14	4.6 (1.6)	21	3.11	11.13
P17	15.06 (7.12)	13.26	11.77 (4.29)	21	6.32	9.75
P18	18.2 (5.27)	15.14	46.24 (6.84)	40	26.73	21.20
P19	26.11 (11.68)	15.14	37.06 (6.48)	36	31.92	19.08
P20	16.9 (5.61)	15.14	39.64 (9.05)	36	22.11	19.08

[a]C: concentration.

[b]ID: inhaled dose.

Computed V′m has a great variability among the different participants as highlighted by median values, but also intrasubject variability as can be observed analyzing the reported IQRs. It can be observed that the IQR values of C_{NEW} are very high, reaching a maximum of 10.2 µg/m^3. This proves the great spatial and temporal variability of PM2.5 along the 4.5 km path and the ability of the developed wearable environmental monitor to account for that.

It is calculated that using ARPA values of concentration and tabulated values of minute ventilation led to a mean PAPE to PM2.5 (ID_{STD} in Table 3) of 13.31 µg with a SD of 4.16 µg, whereas the PAPE obtained with our BSN (ID_{NEW} in Table 3) had a mean of 16.27 µg and a SD of 9.78 µg. They are strongly positively correlated (r=0.665, P<.001), indicating that when the dose estimated from wearable devices increases, the dose calculated from standard data also tends to increase; similarly, when there is a decrease in one measure, the other decreases as well. This alignment suggests a consistent relationship between the 2 methods, with changes in wearable-derived data generally reflecting those in the standard measurements, however, the method with wearable data exhibits higher variability in the estimations. The paired *t* test applied to the 2 normal distributions reveals that on average, the PAPE to PM2.5 with standard data underestimates by 2.96 µg the PAPE evaluated with recorded data coming from the proposed wearable system (22.3% difference; 95% CI −6.55 to 0.63, *P*=.05). Boxplots comparing the PAPE of PM2.5 are reported in Figure 6. In Table 4 the medians and IQRs of aggregated data of PAPE for the different acquisition points for CO_2, VOC, PM1, PM2.5, and PM10 are reported.

XSL•FO
RenderX

Figure 6. Boxplots representing PAPE to PM2.5 computed with standard data, which are ARPA values for pollutant's concentration and tabulated values for minute ventilation, and with wearable data detected by the proposed wearable system. ARPA: Agenzia Regionale per la Protezione Ambientale; PAPE: personal air pollution exposure; PM: particulate matter.

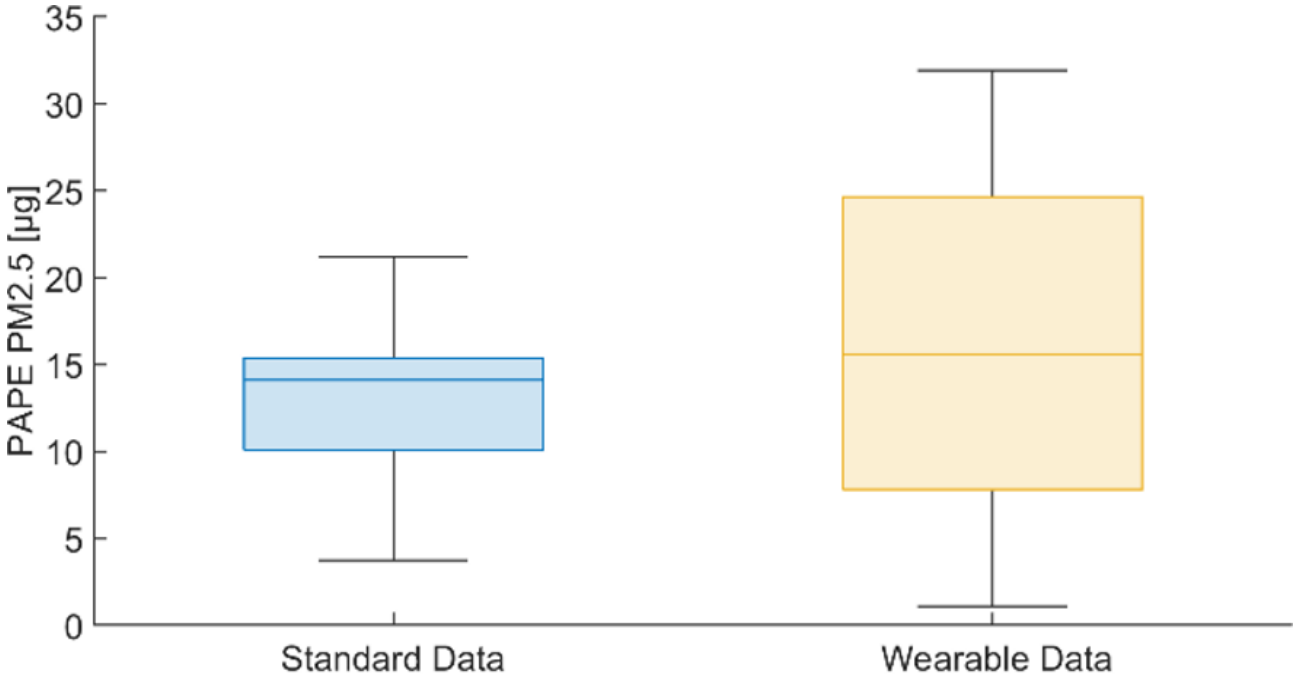


Table . Personal air pollution exposure of CO_2, total volatile organic compounds (tVOCs), particulate matter (PM) 1, PM2.5, and PM10 in the 7 acquisition points.

	CO_2 (pp•mm^3), median (IQR)	tVOCs (pp•bm^3), median (IQR)	PM1 (µg), median (IQR)	PM2.5 (µg), median (IQR)	PM10 (µg), median (IQR)
A	108.3 (59.3)	108.3 (59.3)	0.7 (1.3)	0.8 (1.4)	0.9 (1.5)
B	69.9 (52.0)	69.9 (52.0)	1.8 (3.1)	2.1 (2.9)	2.4 (3.2)
C	80.1 (22.5)	80.1 (22.5)	2.5 (2.6)	2.3 (2.5)	2.6 (2.6)
D	65.8 (18.7)	65.8 (18.7)	2.4 (3.1)	2.3 (2.9)	2.5 (3.2)
E	72.2 (35.0)	72.2 (35.0)	2.5 (3.8)	2.5 (3.5)	2.7 (3.8)
F	67.3 (20.0)	67.3 (20.0)	2.6 (1.9)	2.1 (1.8)	2.4 (1.9)
G	71.2 (20.0)	71.2 (20.02)	2.5 (2.5)	2.2 (2.5)	2.4 (2.8)

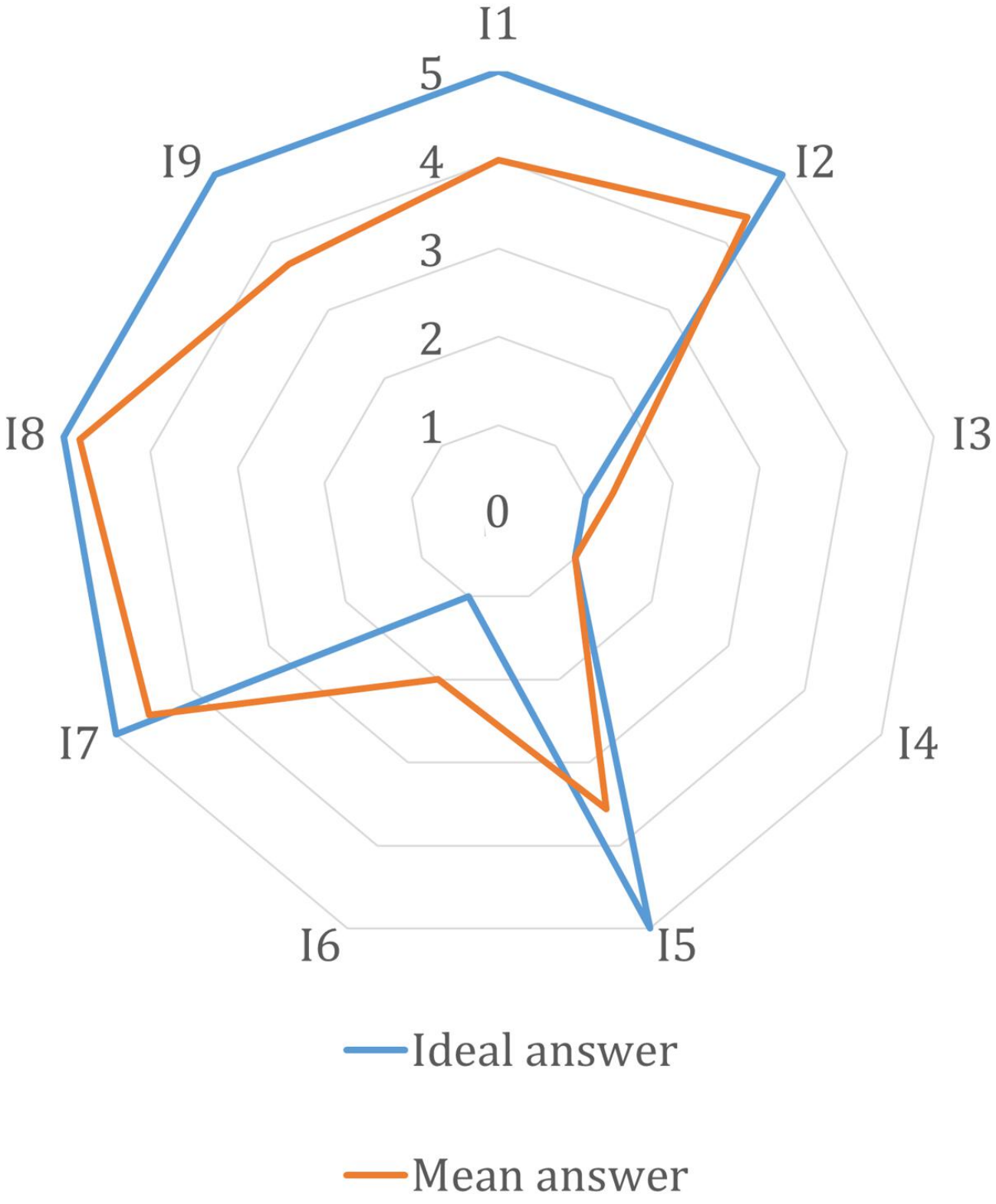
The Friedman test resulted in a significant difference in PAPE between acquisition points for all the pollutants, with a $P<.001$. The post-hoc analysis on PAPE, with the Bonferroni correction, revealed that the only point that is statistically different from all the others is point A, which was expected since it is the only indoor location of the experimental campaign. Some considerations, even though not statistically significant, can be added. PAPE to CO_2 is higher ($P=.12$) in point C (park) with respect to point F (small park with ARPA station). The situation is analogous for tVOCs ($P=.10$), with the additional factor that trees contribute to their emissions. Considering PAPE to PM near the station (E), higher values with respect to the other points were reached, even if statistically significant only compared with point C ($P_{BE}=.88$, $P_{CE}=.03$, $P_{DE}=.19$, $P_{EF}=.58$, $P_{EG}=.91$).

Questionnaire Results

The results of the questionnaire are presented in the radar plot in Figure 7, which displays both the collected scores and the optimal/ideal scores for each item. The lowest scores are obtained for statements regarding wearability over a long period and the functionality of the hardware design. Despite these concerns, high scores were obtained for the question addressing the comfort of the fixing method (I3 in Table 1). Additionally, the way data are presented in the accompanying application was appreciated, as demonstrated in I8.

XSL•FO
RenderX

Figure 7. Radar plot of questionnaire results. Ideal scores are indicated with the blue line while in orange mean scores collected during the campaign.



Discussion

Main Results

This study aimed to determine the feasibility of using the proposed BSN for estimating inhaled pollutant doses in real-life settings. Additionally, it sought to assess whether the system could detect differences in pollution levels between morning and afternoon, evaluate variations across predefined locations with different pollution sources regardless of the time of day, and examine the suitability of the BSN for both indoor and outdoor measurements. The study also compared the system's data with those from an ARPA fixed station and assessed the overall usability and user acceptance of the proposed solution.

The feasibility of using the proposed BSN for estimating the inhaled dose of pollutants in real-life settings is supported by the lightweight of the devices and their complete wearability.

XSL•FO
RenderX

For what concerns the environmental monitor, the sensors chosen are specifically suited for high temporal resolution sampling, making the system well-equipped to track the substantial spatial and temporal variability of air pollution. This is particularly relevant given that pollutant concentrations can vary significantly over small spatial scales (<100 m) and short time periods due to emission dynamics, such as traffic peaks during rush hours. As highlighted in the literature, fixed monitoring stations may fail to capture these variations adequately, reinforcing the need for a high-resolution mobile system [7]. For comparison, commercial environmental monitoring instruments typically offer a measurement every minute at best, and more commonly, one sample every 5 minutes. Our device, which samples environmental parameters nearly 3 times faster than the fastest commercial options, is thus well-suited to capture the detailed spatial and temporal information required in environments like those tested. Moreover, the system includes units capable of synchronously retrieving RR and PR already aligned with the pollution data [26]. These units were already tested both in static [22] and dynamic conditions [27] in previous works. The synchronized data collection of the system allows for an accurate evaluation of the inhaled dose, calculated as the product of pollutant concentration and V′m. Furthermore, the smartphone app has been thoroughly tested and performed reliably throughout the different tests. Therefore, the system is well-suited for estimating the inhaled dose of pollutants in real-life settings. It is worth noting that, although the sample size was determined by practical considerations, it also aligned with the sample sizes typically used in comparable studies focused on wearable prototypes or exposure assessment [22,28,29].

The system demonstrated its capability to study differences in pollution levels between morning and afternoon. CO_2 was found to be higher during morning acquisitions with respect to afternoon acquisitions on the same day, both excluding and including the indoor environment in the analysis. This result is also confirmed in the literature. In fact, outdoor concentrations of CO_2 exhibit daily and diurnal variations, with fluctuations ranging from 50 to 100 ppm in large urban centers, as was observed in a study in Baltimore [30]. CO_2 concentrations tend to be lower in the afternoon and higher in the morning. Instead, PM concentrations were statistically higher in the afternoon compared with morning measurements. This trend aligns with the influence of traffic, which significantly elevates PM levels during peak hours such as lunchtime, coinciding with school dismissal and workers' breaks.

The BSN effectively evaluated differences at predefined points along a walking path, each with potentially distinct pollution sources, regardless of the time of day. During the study, substantial fluctuations in PM were observed within a single data acquisition session, underscoring the limitations of stationary monitoring stations in capturing nuanced air quality dynamics. Indeed, over a short 4.5 km walk, individuals encounter diverse pollution levels, challenging the notion of a single daily concentration value as representative of actual exposure. For instance, in a low-traffic street in a parking lot (acquisition point G), PM levels were unexpectedly high, surpassing those at more conventionally polluted locations such as intersections with traffic lights. This elevated pollution was likely due to the architectural configuration around the parking lot, which impeded air recirculation, emphasizing the value of having a device that provides objective measurements rather than relying on subjective impressions. Physiological parameters further illustrate the system's effectiveness in individually evaluating PAPE. At the initial acquisition point A indoors, participants showed a significantly higher RR, likely due to heightened anxiety due to the start of the experimental acquisition or the high CO_2 concentration in the room. Currently, the system cannot distinguish between these factors because additional sensors, such as those measuring galvanic skin response, are needed to properly assess the participants' stress levels. However, the system shows potential for future developments in this area. Furthermore, the transition from point A to point B, which involved ascending a flight of stairs, contributed to the observed higher, even if not significant, RR at point B with respect to the other acquisition points. RR further decreased at subsequent points, particularly where participants were seated, such as at points F and G. PR data showed lower values at point A, with consistent increases as participants moved and engaged in brisk walking, reflecting expected physiological responses. Additionally, variations in CO_2 and tVOCs levels were observed across different points. Point C, a park with dense tree cover, had significantly higher CO_2 and tVOCs readings compared with point F, a smaller park with less tree coverage. Studies suggest that VOCs from trees may have beneficial effects, such as enhancing cognitive functions and reducing mental fatigue [31]. Consequently, understanding whether a sensor can detect tree-derived VOCs—and at what concentrations—is essential for making informed decisions about exposure and potential benefits. For PM, the highest IQR was recorded at point E, near a train station, influenced by factors such as nearby taxis and varying traffic conditions. This variability, along with differences in physiological responses, underscores the importance of the BSN's ability to provide nuanced data across various environments. Overall, the system's capacity to detect and analyze these spatial and temporal variations confirms its effectiveness in evaluating pollution and exposure differences at multiple predefined points along a walking path.

The system has proven to be suitable for both indoor and outdoor measurements. Indoor measurements indicated significantly higher concentrations of CO_2 and tVOCs, with considerable variability observed even within short, 5-minute intervals. This variability was particularly notable among different participants and can be attributed to varying conditions of air recirculation and fluctuations in the number of individuals in the room. In contrast, PM concentrations were generally lower indoors, consistently with expectations. Outdoors, the system detected notable variability in PM levels across different days and even within the same day, highlighting its capability to capture dynamic changes in air quality. This ability to effectively monitor and account for fluctuations in both indoor and outdoor environments confirms the system's versatility and suitability for comprehensive air quality assessment. Given that urban populations spend most of their time indoors, where pollutant concentrations often exceed outdoor levels, the ability to monitor

XSL•FO
RenderX

both environments is essential for accurately assessing personal exposure [7,32].

The data obtained from the new device show notable variability when compared with values reported by the ARPA fixed station. The new device typically recorded lower pollutant concentrations compared to the ARPA mean values. This discrepancy is primarily due to the experimental protocol, which intentionally avoided peak traffic hours, resulting in data collection during periods of lower pollutant concentrations. This highlights the limitations in time resolution of fixed stations, which may not fully capture the real exposure levels experienced by individuals. Moreover, fixed monitoring stations are often limited in their spatial representativeness, as air pollution levels can vary significantly even over small distances due to local emission sources, urban morphology, and meteorological conditions. The presence of street canyons, for example, can lead to pollutant accumulation, while open areas may experience greater dispersion. Consequently, a mobile system such as the one used in this study provides a more accurate representation of personal exposure by capturing both spatial and temporal variations that fixed stations may not be able to detect [7].

Additionally, PR and RR varied significantly among different participants and even in the same participants at different acquisition points. This variability was reflected in V′m data, which were computed using a model based on these physiological measurements. Unlike previous studies that assign standard Vm′ values based on activity type [33], our approach accounts for individual physiological variations [34], improving the accuracy of inhaled pollutant dose estimation [8,32]. This individualized evaluation of inhaled dose is especially important, as shown in studies of athletes, where factors such as physical activity and ventilation rate significantly impact exposure levels, emphasizing the need for personalized assessment to better reflect real exposure [35]. Exemplary comparison of inhaled doses of PM2.5—calculated as the product of V′m and pollutant concentrations—revealed an average overestimation of 2.96 µg with the new wearable system compared with ARPA data and standard tables for V′m, and also a higher variability in the measurements obtained from the BSN. These observations highlight the system's ability to capture fine temporal and spatial variations in air quality and physiological parameters, demonstrating its potential for more accurate real-time monitoring compared with traditional fixed stations combined with tabulated data.

The usability of the proposed system was positively received by the tested population. The results from usability satisfaction questionnaires indicate an overall favorable perception of the system. Participants found it acceptable and expressed strong interest in its functionality. This positive feedback underscores the system's practicality and user appeal and indicates its potential for broader application and adoption.

Limitations

A limitation of the work is that physiological data were collected only in static positions, while the environmental monitor continuously detects pollution values. In fact, to properly assess PAPE during the testing session, also dynamic acquisitions with physiological parameters need to be performed, and future implementations should account for that. However, the collected dataset was enough to demonstrate significant variability in V′m among different participants, and also for the same participant. Future research should also include activities other than walking, such as cycling to enhance a wider range of minute ventilation values. Additionally, incorporating simultaneous measurements of 2 or more individuals walking the same route could provide further insights into individual variations in exposure to micro-environmental pollution and physiological responses. This approach could be considered for future experimental trials to better highlight the need for personalized environmental exposure and health impact assessments.

In reference to the usability satisfaction questionnaires, some limitations were evident. Wearability over prolonged durations and device functionality emerged as pivotal aspects, emphasizing the necessity for a more compact and lighter device. It is worth noting that the device was consistently worn at the level of the wrist during all acquisitions. However, alternative configurations, such as attaching it to a belt or to an armband, can be implemented for enhanced comfort. Despite differences in pollution levels observed over distances of several meters, the relatively small distance between the mouth or nose and the device in these configurations suggests that any of these placements could accurately measure pollutant exposure and reliably support PAPE assessment.

Conclusions

Traditional air pollution exposure estimation methods fail to account for the dynamic variability in air pollution both indoors and outdoors, as well as the variations in minute ventilation among individuals. To overcome such limitations, our study introduces a novel wearable PAPE evaluation system consisting of a BSN collecting both physiological and environmental parameters, as concentrations of pollutants that have a known adverse effect on human health. Statistically significant differences in pollutant concentration were found between morning and afternoon but also between acquisition points in the same test session.

To estimate V′m, we used a pre-existing model described in the literature. The PAPE was then computed by integrating over time the product of V′m and pollutant concentrations sampled by the developed system. As an example, PAPE to PM2.5, evaluated with fixed station's concentration and standard values of V′m for the 20 participants, was significantly lower with respect to PAPE obtained with the proposed wearable system. Despite the study's participants pool was limited, it effectively highlighted the variability in V′m among individuals and fluctuations in pollutant concentration.

Future improvements involve incorporating an active air recirculation system to mitigate micro-environmental effects inside the device's 3D-printed case over temperature and humidity and advancing device miniaturization.

In conclusion, this system demonstrates high spatial and temporal resolution, positioning it as a valuable tool that may supplement epidemiological studies and empower the public

XSL•FO
RenderX

with data to enhance pollution awareness and acceptability towards policies aimed at reducing air pollution.

Acknowledgments

We gratefully acknowledge Stefano Canali for his invaluable support and insightful suggestions, which significantly enhanced the quality of this work. This work was funded by the National Plan for NRRP Complementary Investments (PNC, established with the decree-law 6 May 2021, number 59, converted by law number 101 of 2021) in the call for the funding of research initiatives for technologies and innovative trajectories in the health and care sectors (Directorial Decree number 931 of 06-06-2022), project number PNC0000003—Advanced Technologies for Human-Centred Medicine (project acronym: ANTHEM). This work reflects only the authors' views and opinions; neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

Authors' Contributions

SB contributed to the conceptualization, data curation, formal analysis, methodology, software development, visualization, writing of the original draft, and review and editing of the manuscript. A Angelucci was involved in the conceptualization, methodology, software development, visualization, and review and editing of the manuscript. AR contributed to data curation, formal analysis, methodology, and software development. A Aliverti contributed to the conceptualization, funding acquisition, methodology, project administration, supervision, provision of resources, visualization, and review and editing of the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

A Angelucci and A Aliverti are coinventors of the patent application "Method for monitoring of respiratory rate of a person." A Aliverti is co-inventor of the granted patent "Wearable device for the continuous monitoring of the respiratory rate." The authors declare no other competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Air pollution. OECD. URL: https://www.oecd.org/environment/air-pollution/ [accessed 2023-02-17]
2. Zhang QS. Environment pollution analysis on smart cities using wireless sensor networks. SPEE 2022 Dec. [doi: 10.13052/spee1048-5236.42112]
3. Honarvar AR, Sami A. Towards sustainable smart city by particulate matter prediction using urban big data, excluding expensive air pollution infrastructures. Big Data Res 2019 Sep;17:56-65. [doi: 10.1016/j.bdr.2018.05.006]
4. Rete di rilevamento [Web page in Italian]. ARPA Lombardia. URL: https://www.arpalombardia.it/temi-ambientali/aria/rete-di-rilevamento/ [accessed 2024-01-19]
5. Lu Y. Beyond air pollution at home: assessment of personal exposure to PM2.5 using activity-based travel demand model and low-cost air sensor network data. Environ Res 2021 Oct;201:111549. [doi: 10.1016/j.envres.2021.111549] [Medline: 34153337]
6. Bernasconi S, Angelucci A, Aliverti A. A scoping review on wearable devices for environmental monitoring and their application for health and wellness. Sensors (Basel) 2022 Aug 11;22(16):5994. [doi: 10.3390/s22165994] [Medline: 36015755]
7. Dias D, Tchepel O. Spatial and temporal dynamics in air pollution exposure assessment. Int J Environ Res Public Health 2018 Mar 20;15(3):3. [doi: 10.3390/ijerph15030558] [Medline: 29558426]
8. Arano KAG, Sun S, Ordieres-Mere J, Gong AB. The use of the internet of things for estimating personal pollution exposure. Int J Environ Res Public Health 2019 Aug 28;16(17):17. [doi: 10.3390/ijerph16173130] [Medline: 31466302]
9. Xu Y, Yi L, Cabison J, et al. The impact of GPS-derived activity spaces on personal PM2.5 exposures in the MADRES cohort. Environ Res 2022 Nov;214:114029. [doi: 10.1016/j.envres.2022.114029]
10. Angelucci A, Kuller D, Aliverti A. A home telemedicine system for continuous respiratory monitoring. IEEE J Biomed Health Inform 2020;25(4):1247-1256. [doi: 10.1109/JBHI.2020.3012621]
11. Saini J, Dutta M, Marques G. Sensors for indoor air quality monitoring and assessment through internet of things: a systematic review. Environ Monit Assess 2021 Jan 15;193(2):66. [doi: 10.1007/s10661-020-08781-6] [Medline: 33452599]
12. Cárdenas M, Vallejo M, Romano-Riquer P, Ruiz-Velasco S, Ferreira-Vidal AD, Hermosillo AG. Personal exposure to PM2.5 air pollution and heart rate variability in subjects with positive or negative head-up tilt test. Environ Res 2008 Sep;108(1):1-6. [doi: 10.1016/j.envres.2008.04.005]
13. Bernasconi S, Angelucci A, Rossi A, Aliverti A. A wearable system for personal air pollution exposure: a walk-about in milan. Eur Respir J 2023 Sep;62(suppl 67):PA2908. [doi: 10.1183/13993003.congress-2023.PA2908]

XSL•FO
RenderX

14. Bernasconi S, Angelucci A, Aliverti A. Evaluation of a new wearable device for indoor and outdoor environmental monitoring. Presented at: 2024 IEEE International Workshop on Sport, Technology and Research (STAR); Jul 8-10, 2024; Lecco, Italy p. 252-257. [doi: 10.1109/STAR62027.2024.10635952]
15. SPS30-PM2.5 sensor for HVAC and air quality applications. Sensirion. URL: https://sensirion.com/products/catalog/SPS30 [accessed 2024-08-19]
16. SCD41-high CO_2 accuracy with extended measurement range and low power modes. Sensirion. URL: https://sensirion.com/products/catalog/SCD41 [accessed 2024-08-19]
17. SGP30-multi-gas (VOC and CO_2 eq) sensor. Sensirion. URL: https://sensirion.com/products/catalog/SGP30 [accessed 2024-08-19]
18. SGX - SGX sensortech sensor. Sensortech. URL: https://www.sgxsensortech.com/sensor/mics-6814 [accessed 2024-08-19]
19. Angelucci A, Camuncoli F, Galli M, Aliverti A. A wearable system for respiratory signal filtering based on activity: a preliminary validation. Presented at: 2022 IEEE International Workshop on Sport, Technology and Research (STAR); Jul 6-8, 2022; Trento - Cavalese, Italy. [doi: 10.1109/STAR53492.2022.9860001]
20. Angelucci A, Camuncoli F, Dotti F, et al. A wireless body sensor network for cardiorespiratory monitoring during cycling. 2023 Presented at: 2023 IEEE International Workshop on Sport, Technology and Research (STAR); Sep 14-16, 2023; Cavalese - Trento, Italy. [doi: 10.1109/STAR58331.2023.10302438]
21. Angelucci A, Aliverti A. An IMU-based wearable system for respiratory rate estimation in static and dynamic conditions. Cardiovasc Eng Technol 2023 Jun;14(3):351-363. [doi: 10.1007/s13239-023-00657-3] [Medline: 36849621]
22. Angelucci A, Bernasconi S, D'Andrea M, et al. Integration of a body sensor network of wearable devices for cardio-respiratory monitoring. Annu Int Conf IEEE Eng Med Biol Soc 2023 Jul;2023:1-4. [doi: 10.1109/EMBC40787.2023.10340495] [Medline: 38083222]
23. Greenwald R, Hayat MJ, Dons E, et al. Estimating minute ventilation and air pollution inhaled dose using heart rate, breath frequency, age, sex and forced vital capacity: a pooled-data analysis. PLoS ONE 2019;14(7):e0218673. [doi: 10.1371/journal.pone.0218673] [Medline: 31287820]
24. GLI lung function calculator. Global Lung Function Initiative. URL: http://gli-calculator.ersnet.org/index.html [accessed 2023-02-18]
25. Allan M, Richardson GM, Jones-Otazo H. Probability density functions describing 24-hour inhalation rates for use in human health risk assessments: an update and comparison. Hum Ecol Risk Assess 2008 Apr 10;14(2):372-391. [doi: 10.1080/10807030801934796]
26. Angelucci A, Aliverti A. A Wearable Body Sensor Network for Personal Monitoring During Daily Life: Technology and Applications 2024:129-139. [doi: 10.1007/978-3-031-63755-1_10]
27. Angelucci A, Camuncoli F, Dotti F, et al. Validation of a body sensor network for cardiorespiratory monitoring during dynamic activities. Biocybern Biomed Eng 2024 Oct;44(4):794-803. [doi: 10.1016/j.bbe.2024.09.002]
28. Bales E, Nikzad N, Quick N, Ziftci C, Patrick K, Griswold W. Citisense: mobile air quality sensing for individuals and communities. design and deployment of the citisense mobile air-quality system. Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2012; San Diego, United States. [doi: 10.4108/icst.pervasivehealth.2012.248724]
29. Carvalho AM, Krecl P, Targino AC. Variations in individuals' exposure to black carbon particles during their daily activities: a screening study in Brazil. Environ Sci Pollut Res Int 2018 Jul;25(19):18412-18423. [doi: 10.1007/s11356-018-2045-8] [Medline: 29696538]
30. Batterman S. Review and extension of CO_2 -based methods to determine ventilation rates with application to school classrooms. Int J Environ Res Public Health 2017 Feb 4;14(2):145. [doi: 10.3390/ijerph14020145] [Medline: 28165398]
31. Antonelli M, Donelli D, Barbieri G, Valussi M, Maggini V, Firenzuoli F. Forest volatile organic compounds and their effects on human health: a state-of-the-art review. Int J Environ Res Public Health 2020 Sep 7;17(18):6506. [doi: 10.3390/ijerph17186506] [Medline: 32906736]
32. Borghi F, Spinazzè A, Mandaglio S, et al. Estimation of the inhaled dose of pollutants in different micro-environments: a systematic review of the literature. Toxics 2021 Jun 12;9(6):140. [doi: 10.3390/toxics9060140] [Medline: 34204794]
33. Asimina S, Chapizanis D, Karakitsios S, et al. Assessing and enhancing the utility of low-cost activity and location sensors for exposure studies. Environ Monit Assess 2018 Mar;190(3):3. [doi: 10.1007/s10661-018-6537-2]
34. Aliverti A, Evangelisti M, Angelucci A. Wearable tech for long-distance runners. In: The Running Athlete 2022:77-89. [doi: 10.1007/978-3-662-65064-6_10]
35. Ribalta C, Garrandes F, Bermon S, et al. Dynamic and stationary monitoring of air pollutant exposures and dose during marathons. Sci Total Environ 2024 Jun 1;927:171997. [doi: 10.1016/j.scitotenv.2024.171997] [Medline: 38565357]

Abbreviations

ARPA: Agenzia Regionale per la Protezione Ambientale
BSN: body sensor network
HR: heart rate

XSL•FO
RenderX

IMU: inertial measurement unit
PAPE: personal air pollution exposure
PM: particulate matter
PR: pulse rate
RR: respiratory rate
tVOC: total volatile organic compound
VOC: volatile organic compound

Edited by L Buis; submitted 10.05.24; peer-reviewed by F Borghi, Y Lu; revised version received 21.03.25; accepted 16.04.25; published 04.07.25.

Please cite as:
Bernasconi S, Angelucci A, Rossi A, Aliverti A
A New Wearable System for Personal Air Pollution Exposure Estimation: Pilot Observational Study
JMIR Mhealth Uhealth 2025;13:e60426
URL: https://mhealth.jmir.org/2025/1/e60426
doi:10.2196/60426

© Sara Bernasconi, Alessandra Angelucci, Andrea Rossi, Andrea Aliverti. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Value of Continuous Plantar Temperature Monitoring for Diabetic Foot Health Management: Observational, Prospective Cohort Study

Maryam Hajizadeh[1], PhD; Emily Matijevich[1], PhD; Emily Bray[1], MBA; Evan Minty[2], MD, MSc; Brock Liden[3], DPM

[1]Orpyx Medical Technologies, 1240 20 Ave SE Suite 205, Calgary, AB, Canada

[2]Department of Medicine, University of Calgary, Calgary, AB, Canada

[3]Cutting Edge Research LLC, Circleville, OH, United States

Corresponding Author:
Maryam Hajizadeh, PhD
Orpyx Medical Technologies, 1240 20 Ave SE Suite 205, Calgary, AB, Canada

Abstract

Background: Diabetic foot ulcers (DFUs) are a life-changing complication of diabetes. There is increasing evidence that remote plantar temperature monitoring can reduce the recurrence of DFUs. Monitoring of foot temperature once a day is the current guideline for identifying early signs of foot inflammation. However, single readings of physiological signals can increase the risk of misdiagnosis when the signals fluctuate throughout the day.

Objective: The aim of this study was to evaluate whether intraday temperature asymmetry signals were stable or varied as a function of time in individuals at risk of DFUs.

Methods: In total, 64 participants with diabetes (mean age 68, SD 13.8 y) were provided with multimodal sensory insoles (Orpyx Sensory Insoles) to monitor continuous temperature data at a frequency of once per minute at 5 plantar locations in a 90-day study window. The augmented Dickey-Fuller test was used to determine whether the temperature asymmetry signals were stationary (ie, indicating constant mean and variance over time) or nonstationary (ie, indicating time-varying behaviors or trends in the signal).

Results: The study included 43 participants, 1080 data days, and 5400 contralateral temperature asymmetry signals. Most (4428/5400, 82%) of the temperature asymmetry signals were nonstationary, with intraday fluctuations likely influenced by physiological and environmental factors. Of the nonstationary signals, nearly half (1948/4428, 44%) fluctuated above and below the concerning asymmetry threshold of 2.2 °C. The intraday variability underscores the potential for false-positive and false-negative hot spot detection with once-daily measurements. Substantial variability was observed in stationarity patterns both within and across participants. Notably, concerning asymmetries in nonstationary signals occurred at different time points across participants, measurement windows, and days.

Conclusions: Our findings highlight the value of continuous plantar temperature monitoring for diabetic foot health management, relative to once-daily measurements. Several repeated measurements throughout the day increase confidence with regard to the accuracy of observed plantar physiology trends. Continuous monitoring may improve the accuracy of plantar temperature measurement, unlock new diagnostic capabilities, and support personalized care.

(JMIR Mhealth Uhealth 2025;13:e73187) doi:10.2196/73187

KEYWORDS

continuous temperature monitoring; wearables; sensor-based monitoring; diabetic foot ulcer; temperature; remote patient monitoring

Introduction

Background

Diabetic foot ulcers (DFUs) are open wounds on the foot caused by neuropathy, poor circulation, or repetitive stress to the foot tissue [1]. DFUs are a devastating foot complication in individuals with diabetes, with 1 in 5 ulcers resulting in severe infection and amputation [2]. Fortunately, 75% of DFUs are preventable when the risk factors are detected and treated early [3]. Remote plantar temperature monitoring has been proposed as a preferred method for DFU prevention, as tissue inflammation (indicated by warmth) can precede wound formation [2].

Current international guidelines suggest once-daily foot temperature monitoring to identify early signs of foot inflammation [3,4]. Once-daily plantar temperature measurements can be captured with handheld thermometry [5] or with sensor-based digital health technologies (eg, sensory

XSL•FO
RenderX

insoles [6], smart mats [7], or sensor-embedded socks [8]). Daily temperature asymmetry measures (difference between similar regions on 2 feet) are typically compared to an established asymmetry threshold of 2.2 °C to identify a trend of concern (ie, a temperature hot spot) [7-10]. While studies informing once-daily plantar temperature measurements have shown most success in DFU risk reduction when managed with marked reductions in activity [5], there are concerns that once-daily plantar temperature measurements may not provide a comprehensive view of foot temperature dynamics. Once-daily measurements may also lack the accuracy needed to limit false negatives (which could result in missed, concerning trends) or false positives (which could unnecessarily recommend sedentary behavior in a diabetic population). Key fluctuations in temperature asymmetry that indicate early-stage complications may occur outside the once-daily measurement [11,12].

Single readings of health measurements increase the risk of misdiagnoses due to false-positive and false-negative test results. Many physiological signals fluctuate throughout the day as a function of behavior, activity, environment, circadian rhythms, and other factors [13-15]. While single health measurements remain valuable in some contexts, repeated or continuous measurements are now being taken to improve accuracy. For example, rather than a single blood pressure measurement, at least 4 daily home blood pressure measurements over 2 days are recommended to accurately diagnose out-of-clinic hypertension [14]. For glycemic management, continuous glucose monitoring has become a widely adopted supplement to intermittent capillary blood glucose testing as it more accurately reflects how lifestyle factors impact glucose fluctuations, reducing the risk of hypo- and hyperglycemia events [16,17]. For outpatient fever monitoring in high-risk patients, continuous temperature monitoring, rather than daily oral temperature assessment, has been shown to detect fever onset more accurately and with greater lead time [18].

Wearable devices that facilitate continuous plantar temperature monitoring have shown that absolute plantar temperatures can vary as a function of foot and body posture, limb perfusion, activity levels, disease status, and other factors [12,19]. For example, Beach et al [12] used a wearable device for continuous plantar foot temperature monitoring to calculate dynamic temperature parameters (eg, rise time) in the sitting and standing positions in participants with diabetes and healthy controls. Foot temperature rise times were significantly higher in participants with diabetes, potentially due to differences in soft tissue biomechanics and vascularization. The authors concluded, “temperature monitoring for DFUs focuses on a single snapshot of temperature differences, which may not represent the more complex dynamics of temperature change present in those with diabetes” [12].

Objective

It remains unknown if plantar temperature *asymmetries*, the key biomarker in remote temperature monitoring for DFU management protocols, vary throughout the day. The aim of this analysis was to evaluate whether intraday plantar temperature asymmetries vary as a function of time in individuals with diabetes and a history of DFUs.

Methods

Study Design

The following sections describe a secondary analysis of a broader clinical study evaluating the use of the Orpyx Sensory Insole (Orpyx Medical Technologies Inc.) system as an adjunct to diabetic peripheral neuropathy standard of care. The primary outcome of the broader clinical study was to observe the clinical advantages of the sensory insoles alongside the standard of care. This study was reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Checklist 1).

Participants

In total, 64 participants with type 1 or 2 diabetes, peripheral neuropathy, and a history of a previous plantar foot ulcer (mean age 68, SD 13.8 y) were enrolled through a prospective, observational cohort study. Participants were recruited from a single podiatry clinic in Ohio, United States, from April 2022 to January 2023. The exclusion criteria included the presence of active ulcers or other open chronic wounds, severe vascular disease (defined as an ankle-brachial index [ABI] of ≤0.6 or an ABI of ≥1.2, or adequate lower extremity perfusion per physician discretion), a history of nonneuropathic foot ulcers (ie, arterial or venous insufficiency ulcers), or significant balance disorders.

Technical Implementation

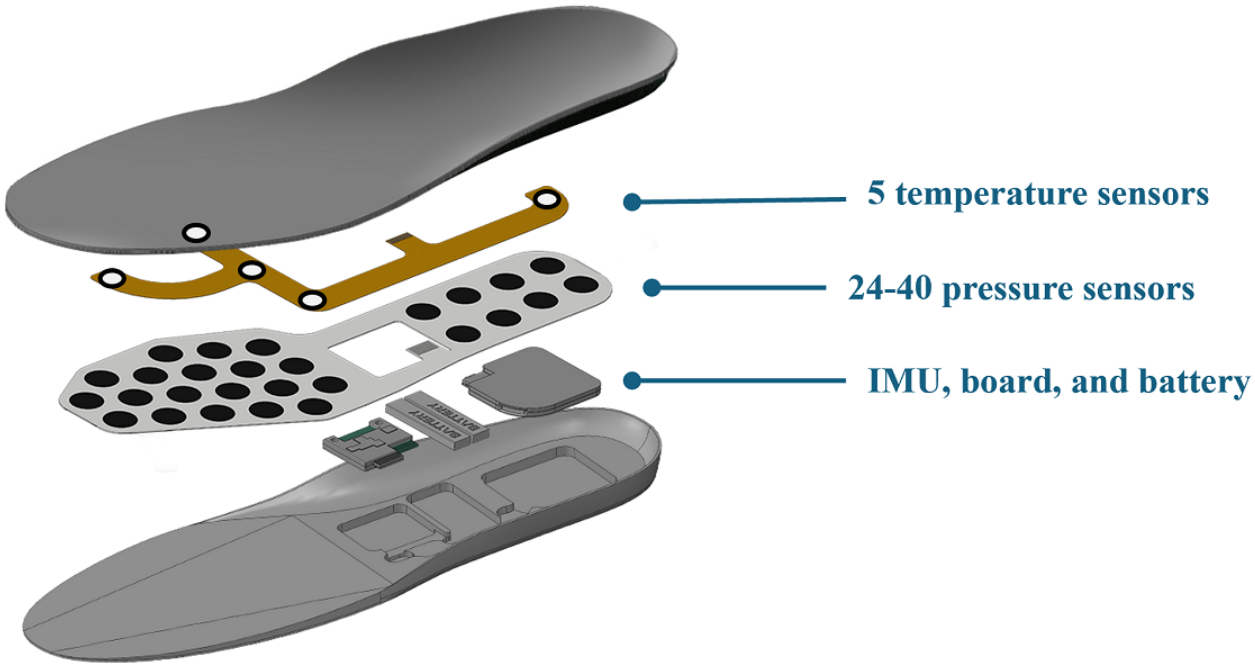
Custom sensory insoles (Orpyx Sensory Insole) were used for data collection. The Orpyx Sensory Insole is a standard-of-care, custom orthotic combined with a technology layer that enables patient-facing biofeedback and remote patient monitoring services.

The custom insoles comprised a custom contoured layer manufactured using a direct milling CAD/CAM (computer-aided design/computer-aided manufacturing) process. When requested by the provider, the insoles were then further modified to accommodate any foot abnormalities. The insole materials and fabrication process meet HCPCS A5514 requirements.

The technology layer included multiple sensors and electronics that enable multimodal, continuous, underfoot data collection (Figure 1). Depending on the size of the insole, the technology layer contains between 24 and 40 pressure sensors. In addition, 5 temperature sensors are embedded and positioned at the big toe; metatarsal heads 1 (Meta 1), 3 (Meta 3), and 5 (Meta 5); and the heel (the areas of bony prominence) that are the most frequent locations for DFU formation. Finally, the electronics assembly includes an inertial measurement unit (IMU) that enables the ability to track steps and wear time.

XSL•FO
RenderX

Figure 1. The multimodal Orpyx Sensory Insole system includes custom-milled insoles that are placed into the participants' shoes. The insoles include a technology layer to enable multimodal monitoring and biofeedback to participants and providers. The technology layer includes temperature sensors, pressure sensors, and an inertial measurement unit (IMU) sensor. For this analysis, only temperature data measured by the 5 temperature sensors were used.



Data Collection and Analysis

Participants wore the sensory insoles for up to 90 days in standardized diabetic footwear, with a recommended daily adherence of 4.5 hours per day [20]. Participation in the study concluded at the end of the 90-day period, upon the development of a plantar foot ulcer within the monitoring window or if the participant withdrew for another reason. Participants were also provided remote patient monitoring services through the in-house remote patient monitoring service at Orpyx Medical Technologies. A case series, focused on the value of multimodal sensing across multiple physiological domains for managing diabetic foot health, was previously published for 3 participants from this study cohort [21].

Plantar temperature data, logged at a frequency of once per minute, were the only parameter used for this analysis. The daily contralateral temperature asymmetry time-series signal was calculated as the difference between continuous temperatures at the corresponding left and right foot regions. Therefore, each sensory insole pair provided 5 temperature asymmetry signals. A series of data-cleaning steps corrected the underlying signal bias and masked portions of the time-series signals that were unreliable measures of foot skin temperature (eg, periods surrounding insole donning and doffing). The remaining signals were smoothed using a rolling window mean (window size=10 minutes).

For this analysis, participants were excluded if they withdrew from the study before insole dispensing, withdrew after minimal device use (defined as no data day with a temperature asymmetry measurement >2 h), or were enrolled as unilateral users, as these participants did not generate plantar temperature asymmetry data. Participants with previous minor lower-limb amputations (eg, removal of the first digit) were also excluded from this specific analysis to control for the potential impact of missing foot anatomy on temperature asymmetry signals. Furthermore, short-duration temperature asymmetry measurements (<2 h) were excluded from this analysis. The short-duration measurement exclusion criteria were specifically to accommodate this exploratory statistical analysis and were more rigorous than what were used in our remote temperature monitoring protocol for DFU management.

Statistical Analysis and Outcomes

A stationarity test was conducted to assess whether the time-series temperature asymmetry signals remained constant over time. Specifically, we used the augmented Dickey-Fuller (ADF) test, performed in Python (version 3.10) with the adfuller function from the statsmodels Python package (version 0.13.2), to evaluate the stationarity of each daily temperature asymmetry time series [22,23]. The ADF test evaluates the null hypothesis that a unit root is present in the time series, indicating nonstationarity due to a stochastic trend [23,24]. In our analysis, we specified regression="c" to assume the data may have a nonzero mean but no trend term. Maximum lag was not manually defined (maxlag=None), while the test automatically selected the optimal number of lags based on the Akaike information criterion (AIC) by setting autolag="AIC," which

XSL•FO
RenderX

balances model fit and complexity. The test yields a test statistic, which is compared against critical values from simulated distributions to assess statistical significance. A *P* value ≤.05 was considered evidence against the null hypothesis, indicating that the time series is "weakly stationary." This infers that certain statistical properties, such as mean, variance, and autocovariance remain constant over that wear period. For this paper, we refer to contralateral temperature time series from the data days in which the ADF test rejects the null hypothesis at a *P* value threshold ≤.05 as "stationary" and in which the ADF test fails to reject the null hypothesis at a *P* value threshold >.05 as "nonstationary."

The stationarity result of all daily signals was aggregated for each participant to calculate the percentage of stationary versus nonstationary signals over their adherence days in the 90-day study period. Individual data days were reviewed to determine if the signals at all 5 plantar locations were consistently stationary, consistently nonstationary, or a mixed result (some stationary and others nonstationary).

Given that nonstationary temperature asymmetry signals will exhibit trend and variability, further analysis was conducted to determine the proportion of measurements within a day that were symmetric versus asymmetric. As no established threshold currently exists for defining concerning asymmetry in *continuous* temperature monitoring, the established threshold for once-daily temperature monitoring (2.2 °C) was used [7-10]. For nonstationary signals with one or more temperature asymmetry measurements exceeding this threshold, the percentage of intraday measurements with "concerning asymmetry" (defined as asymmetry >2.2 °C) was calculated, similar to how continuous glucose measurements are classified as "time in range" and "time above or below range" [17].

Finally, nonstationary signals with at least one concerning asymmetry measurement were qualitatively examined to determine the time of day at which these concerning asymmetrical measurements occurred. This analysis provided insights into whether periods of concerning asymmetry were consistent and repeatable or temporally dispersed between days and across participants.

Ethical Considerations

This study was approved by the WCG Institutional Review Board (20220828) on October 3, 2022. Informed consent was obtained from all participants. The original informed consent permitted secondary analyses without requiring additional consent from participants. Participant data were stored on secure, access-controlled servers. Data were deidentified for this analysis and only accessible by authorized research personnel. Participants did not receive any financial compensation for participation in this study.

Results

Participant Characteristics

A total of 43 participants were included in this analysis (Table 1; Figure 2), of which 21 participants were excluded: 8 participants withdrew from the study before insole dispensing, 2 participants withdrew after minimal device use, 1 participant enrolled as a unilateral user, and 10 participants had previous minor lower-limb amputations.

Across the 43 participants, 1429 data days with temperature asymmetry measurements were considered for further data cleaning and analysis (Figure 2); 313 short-duration data days (<2 h) were discarded, in addition to 36 low-quality data days that had no remaining measurements after data cleaning (Figure 2). Ultimately, 43 participants and 1080 data days provided 5400 temperature asymmetry signals for statistical analysis (Figure 2).

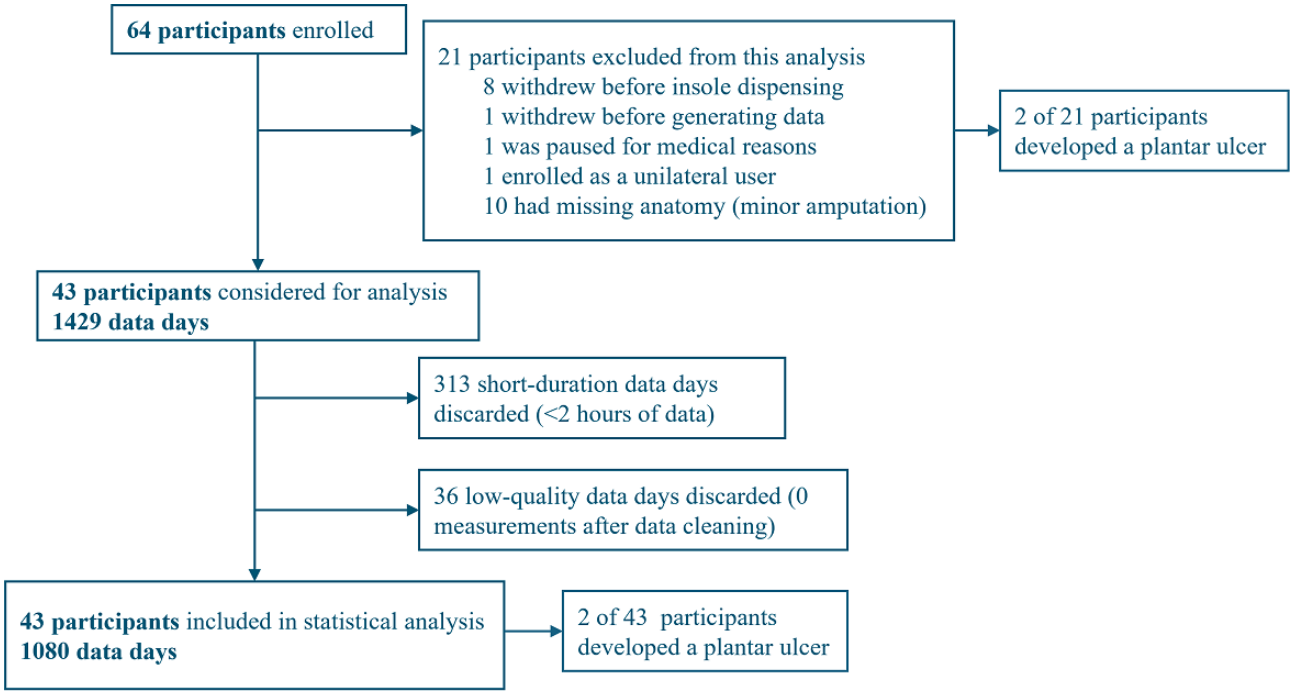
Across the 64 participants enrolled in this study, 4 participants developed a plantar ulcer within the 90-day study period. Of the 4 participants who ulcerated, two were in a cohort excluded from this analysis and two were in the included cohort.

Table . Participant characteristics.

Characteristics	Participants (n=43)
Sex, n (%)	
Male	25 (58)
Female	18 (41)
Age (years)	
Values, mean (SD)	68 (13.8)
Values, median (range; IQR)	70 (39-92; 56-77)
Race, n (%)	
Black or African American	1 (13)
White	42 (97)
Comorbidities, n (%)	
Cardiovascular disease	8 (19)

XSL•FO
RenderX

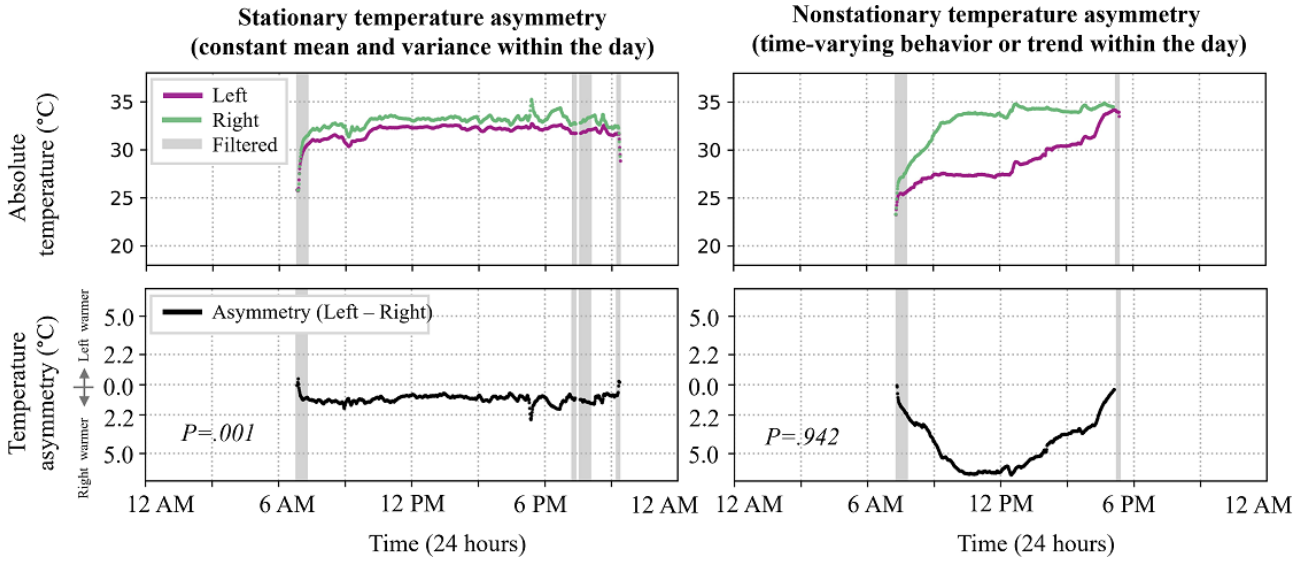
Figure 2. Flowchart of the process used in selecting participants and data days for inclusion in the study analysis.



Summary of Stationary Versus Nonstationary Temperature Asymmetry Signals

Across the 5400 temperature asymmetry signals analyzed, 82% (n=4428) were nonstationary. Figure 3 shows an illustrative stationary and nonstationary temperature asymmetry signal.

Figure 3. Illustrative stationary (indicating constant mean and variance over time) and nonstationary (indicating time-varying behaviors or trends in the signal) plantar temperature signals. Continuous time series for left (purple) and right (green) plantar temperature measurements (upper row); corresponding left versus right plantar temperature asymmetry (lower row). The gray bars show the masked portions of the time-series signals during the data-cleaning steps. The two illustrative days are from two different participants. *P* values indicate the test result from the augmented Dickey-Fuller test on the temperature asymmetry signal. Signals with a *P* value ≤.05 were considered stationary.

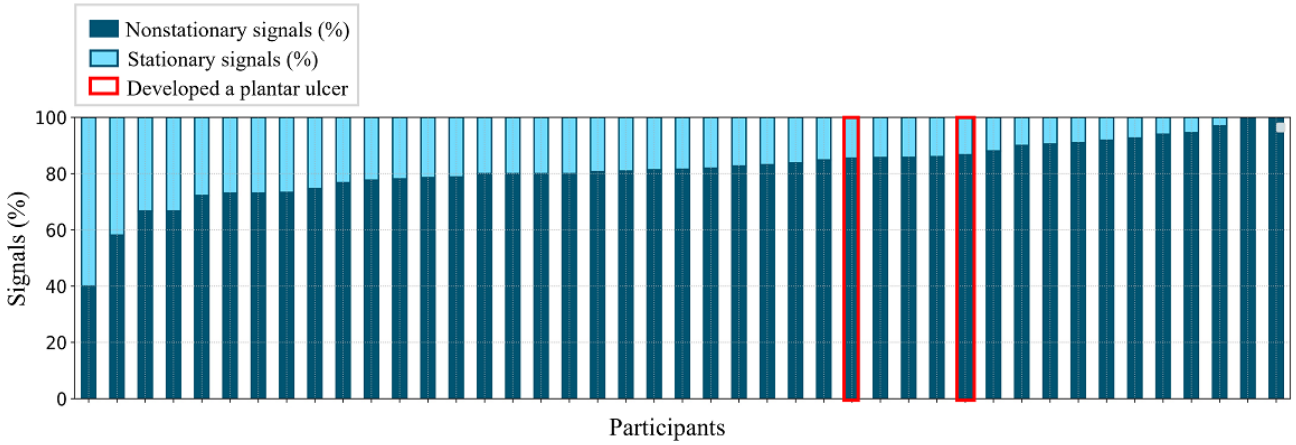


Stationarity of Temperature Asymmetry Signals in Individual Participants

For a given participant, 40% to 100% of their temperature asymmetry signals were nonstationary (Figure 4). No participant exhibited entirely stationary data throughout the 90-day study period. For the 2 participants who developed a plantar ulcer in this study, >80% of their temperature asymmetry signals were nonstationary.

XSL•FO
RenderX

Figure 4. For each participant (n=43), most of the temperature asymmetry signals were nonstationary (indicating time-varying behaviors or trends in the signal). For a given participant, 40% to 100% of their temperature asymmetry signals were nonstationary (dark blue bars) and the remaining signals were stationary (light blue bars). Most of the temperature asymmetry signals were nonstationary for the 2 participants who developed a plantar ulcer (identified with red bars). Participants were visually sorted from lowest to highest percent of nonstationary signals.

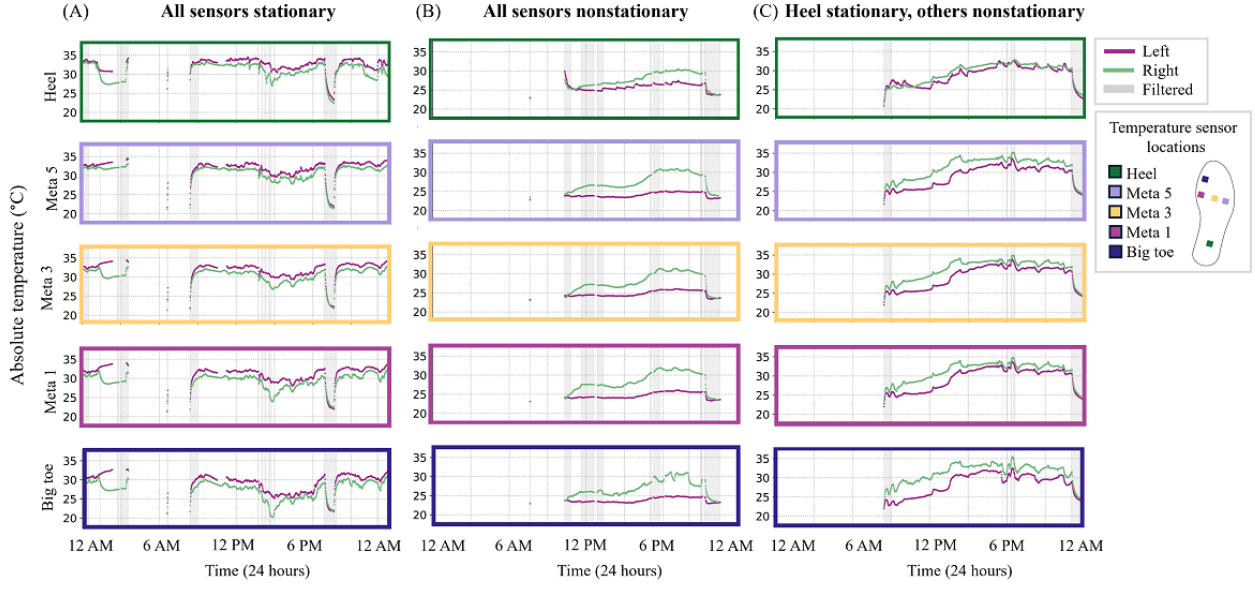


Stationarity of Temperature Asymmetry Signals at Different Plantar Locations

Pooling all participants' data days, the percentage of nonstationary temperature asymmetry signals was similar at each plantar sensor location: 83% of data days at heel, 82% at Meta 1, 82% at Meta 3, 80% at Meta 5, and 82% at the big toe region.

When analyzing individual participant data, some individuals exhibited a consistent stationarity result across all sensor locations (Figure 5A and B). However, others showed a mix of stationary and nonstationary signals across plantar sensor locations (Figure 5C). For example, in 1 participant, the forefoot sensor (Meta 1, Meta 3, Meta 5, and big toe) signals were nonstationary for >90% of data days, whereas heel temperature asymmetry signals were nonstationary for only 64% of data days. Therefore, this participant often exhibited stationary heel temperature asymmetry while other plantar sensor locations showed nonstationary signals.

Figure 5. Stationarity of plantar temperature asymmetry signals sometimes varied across sensor locations. Across sensor locations, some data days displayed consistently stationary temperature asymmetry signals (**A**), some displayed consistently nonstationary temperature asymmetry signals (**B**), and others displayed a mix of stationary and nonstationary signals across sensor locations (**C**). The 3 illustrative data days are from 3 different participants.

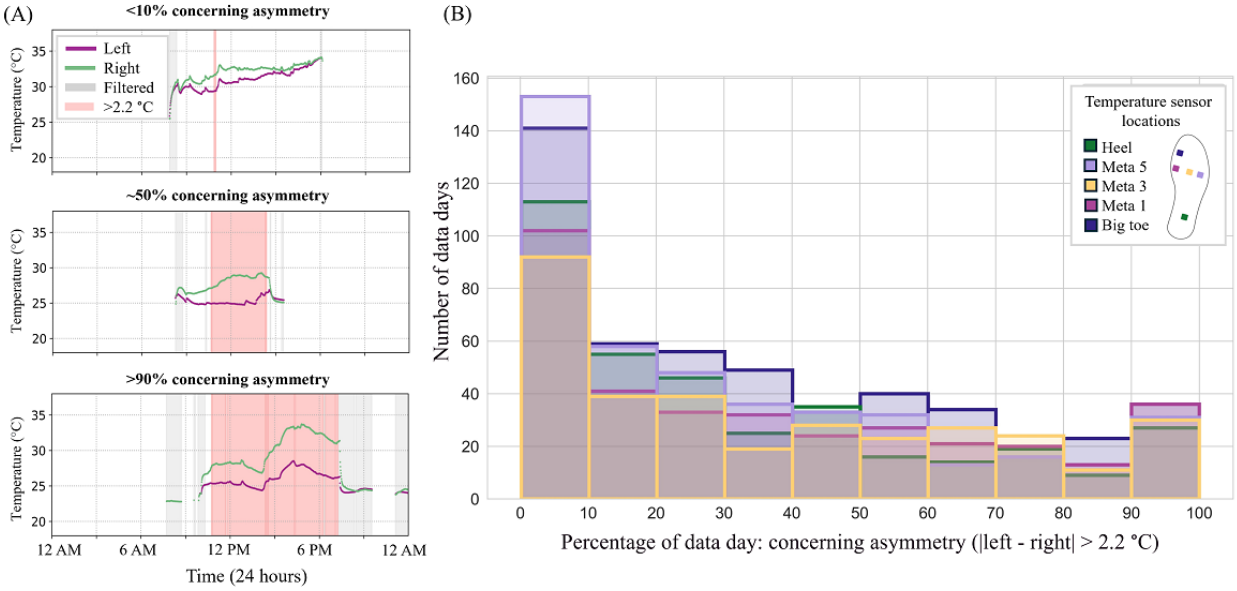


Temperature Trends of Concern in Nonstationary Signals

Of the total, 44% of the nonstationary signals included one or more temperature asymmetry measurements that exceeded a 2.2-°C asymmetry threshold for a data trend of concern. The total number of data days with at least one measurement of concerning asymmetry was 359 for the heel, 431 for Meta 5, 332 for Meta 3, 349 for Meta 1, and 487 for the big toe. The percentage of measurements within a day that exceeded the 2.2-°C threshold varied from very few measurements to nearly all of the daily measurements (Figure 6).

XSL•FO
RenderX

Figure 6. Of the total, 44% of nonstationary signals included one or more concerning temperature asymmetry measurements (>2.2 °C). (**A**) Three illustrative data days showing varying percentage of concerning temperature asymmetry (red shaded region); (**B**) frequency of data days with different percentages of concerning temperature asymmetry measurements at each sensor location.

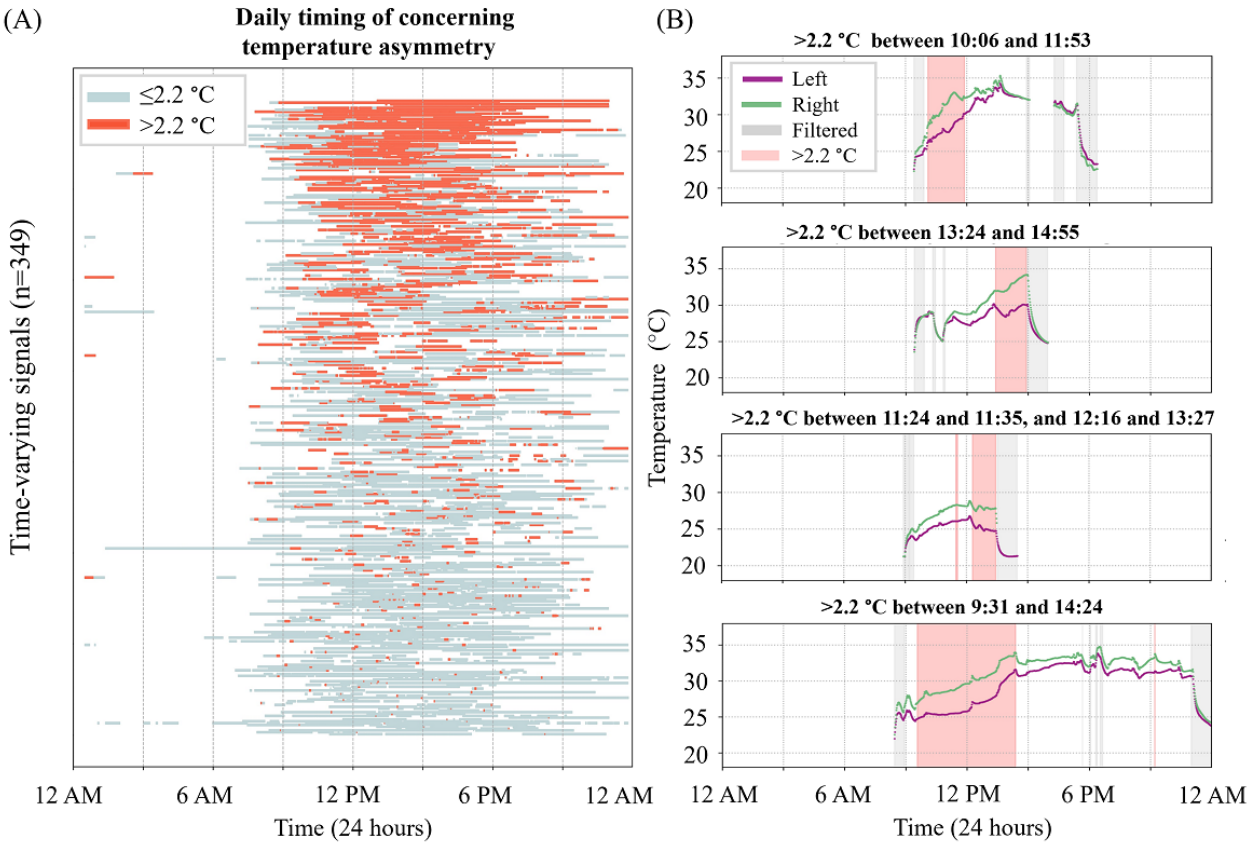


Temporally Dispersed Concerning Temperature Asymmetry

A visual inspection of nonstationary data days with one or more temperature asymmetry measurements exceeding the 2.2-°C threshold revealed that the periods of concerning temperature asymmetry were temporally dispersed across participants, time windows, and data days (Figure 7A and B).

XSL•FO
RenderX

Figure 7. Temporally dispersed concerning temperature asymmetry. (**A**) The distribution of nonconcerning (≤2.2 °C, light blue bands) and concerning (>2.2 °C, red bands) temperature asymmetry measurements across 349 plantar temperature signals that included at least one concerning temperature asymmetry measurement. The results are for one representative sensor location (Meta 1). Each horizontal bar is the usage duration for a data day. The data days were visually ordered based on the duration of the concerning asymmetry. (**B**) Four illustrative data days with at least one concerning temperature asymmetry measurement. Concerning temperature asymmetries occurred at different times of the day.



Discussion

Principal Findings

The primary finding of this study was that most (4428/5400, 82%) of the plantar temperature asymmetry signals were nonstationary. Temperature asymmetry fluctuations occurred throughout the day, likely influenced by underlying physiological and environmental factors. About half (1948/4428, 44%) of the nonstationary temperature signals included a mix of concerning (>2.2 °C) and nonconcerning (≤2.2 °C) temperature asymmetry measurements. Single or intermittent measurements of temperature asymmetry are therefore at risk of over- or underestimating plantar hot spots and DFU risk. These results highlight the value of continuous remote temperature monitoring in the management of diabetic foot disease.

Leveraging Continuous Monitoring to Improve Accuracy

Modest success in DFU risk reduction has been demonstrated with intermittent plantar temperature monitoring [5]. A meta-analysis of 5 randomized controlled trials (n=772) demonstrated that single time-point foot temperature monitoring and reduction of ambulatory activity in response to hot spots (temperature asymmetry >2.2 °C on 2 consecutive days) reduced the risk of developing a DFU by approximately 50% [5]. However, the overall certainty of this evidence was low. Intermittent monitoring protocols may miss transient but clinically relevant changes that occur between measurements.

Of the total, 44% of nonstationary data days included one or more temperature asymmetry measurements that exceeded a 2.2-°C threshold, indicating a data trend of concern. On these days, relying on a single, once-daily snapshot risks both false-negative and false-positive assessments (Figures 6 and 7). Multiple studies have demonstrated the limitations of once-daily temperature monitoring. In a study [7] using a thermometric foot mat to assess once-daily temperature asymmetry (one 20 s measurement), the authors reported a false-positive rate of 57% and false-negative rate of 3% for predicting ulcer development with an average lead time of 37 days. Another study [25] found that in the 2 months preceding ulceration, only 8 (28%) participants had a true hot spot, while 7 (24%) had false hot spots and 14 (48%) had no detectable hot spot, challenging the assumption that once-daily measurements consistently identify impending skin breakdown. Similarly, a post hoc analysis [26] of foot thermograms from 15 individuals with diabetes identified hot spots (asymmetry >2.2 °C) in 20% of regions of interest, yet none of the participants developed ulcers. While several factors likely contribute to the high rate of false positives and false negatives reported, one explanation

XSL•FO
RenderX

may be that single snapshot measurements fail to capture the inherent variability of intraday plantar temperature asymmetries.

Several repeated measurements throughout the day, rather than a once-daily measurement, increase confidence with regard to the accuracy of observed plantar physiology trends. The recommended frequency of daily plantar temperature measurements may need to be increased to provide a more holistic view of a patient's foot health.

Uncovering New Diagnostic Opportunities

Across participants, 40% to 100% of temperature asymmetry signals were nonstationary (Figure 4), and individual participants displayed different distributions of stationary versus nonstationary signals at each plantar location (Figure 5). Numerous factors such as activity, foot care practices, vascular disease status, history of ulceration, foot posture and gait, and foot deformities may contribute to the observed temperature signal patterns [19,27]. While current remote temperature monitoring protocols emphasize the *magnitude* of temperature asymmetry, it is possible the *pattern* of temperature asymmetry also offers diagnostic value. As a secondary, exploratory analysis, we examined 3 illustrative participants to explore potential associations between participant-level factors and patterns of temperature asymmetry.

Case 1 is a participant with comorbid peripheral vascular disease. Continuous plantar temperature monitoring may be particularly valuable for individuals with peripheral vascular disease, a common comorbidity of diabetes, as limb temperature can also serve as a surrogate measure of limb perfusion [28]. The most recent vascular assessment for this participant (approximately 4 mo before enrollment in the study) showed a right ABI of 1.3, suggesting arterial calcification and noncompressibility, and a left ABI measurement of 1.16, within the normal range (Figure 8A). Interestingly, the sensory insole typically measured lower temperatures in forefoot regions on the right foot, aligning with the presumed calcification in that limb. Temperature asymmetry was nonstationary at the forefoot sensor sites (Meta 1, Meta 3, Meta 5, and big toe) but not at the heel. This pattern may reflect the impact of greater vessel length supplying the forefoot compared to the heel, which is served by separate, more proximal arterial branches. It should be noted that this participant was excluded from the primary study analysis due to a right fourth toe amputation (Figure 2), but it is reported here as an exploratory example to illustrate the potential relationship between vascular disease status and temperature asymmetry profile.

Case 2 is a participant with bilateral hammertoes and high daily step counts. The participant exhibited nonstationary temperature asymmetry patterns at all foot regions (5 sensor locations), with the right limb typically presenting as warmer. Interestingly, the magnitude of temperature asymmetry increased progressively with time and cumulative steps throughout the day (Figure 8B). We suspect that the correlation between temperature asymmetry and activity may reflect asymmetrical loading or altered biomechanics from the underlying foot deformities, potentially leading to increased tissue stress and localized inflammation on the right side. This participant was not reported to have underlying vascular disease.

Case 3 is a participant who developed an ulcer at the lateral aspect of the right foot during the study. Notably, the participant exhibited nonstationary temperature asymmetry patterns with episodes of temperature asymmetry exceeding the concerning threshold of 2.2 °C intermittently throughout the day (Figure 8C). The location of intermittent warmth (Meta 3, Meta 5, and big toe) did not exactly correspond to the reported ulcer location (ie, the lateral aspect of the foot). However, these transient temperature elevations, with the right forefoot being relatively warmer, would have been at risk of being missed by once-daily temperature assessments, highlighting the advantage of continuous intraday temperature monitoring. No meaningful data trends were observed for the other participant who developed an ulcer during the study, likely due to poor adherence to the sensory insoles, which were worn for only 3 days.

Exploratory observations from these representative cases highlight the potential for continuous temperature monitoring to offer insights that go far beyond those provided by once-daily measurements. We hypothesize that the ongoing collection of multimodal, continuous plantar data will uncover meaningful associations and previously unrecognized patterns that link temperature asymmetry with vascular status, activity, wound risk, or other factors. Continuous data may enable more personalized and timely interventions based on when and where risk emerges.

XSL•FO
RenderX

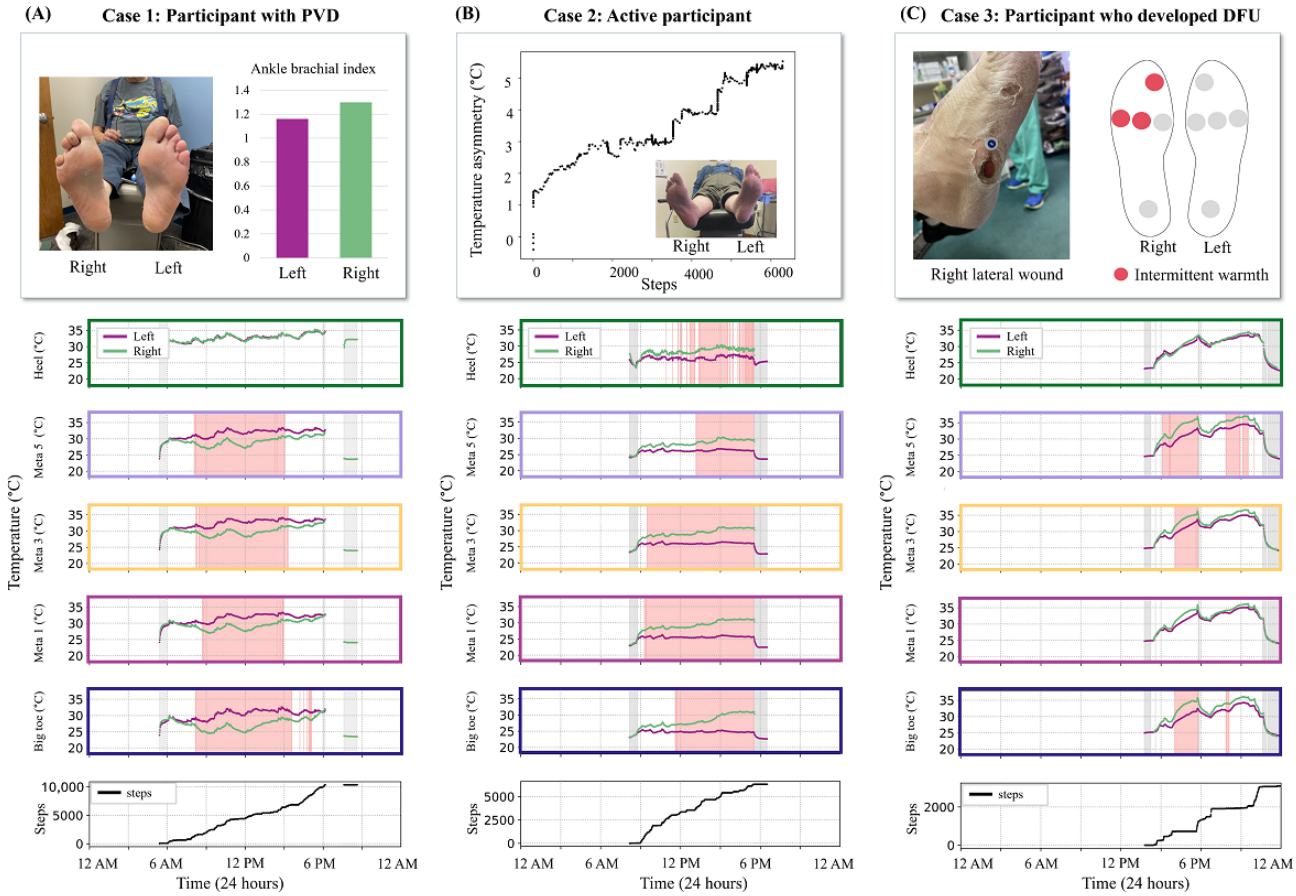


Figure 8. Three illustrative participants to demonstrate potential associations between participant-level factors and patterns of temperature asymmetry. DFU: diabetic foot ulcer; PVD: peripheral vascular disease.

A Step Toward Personalized Monitoring

Once-daily foot skin temperature monitoring is a clear, concise, and achievable guideline [7]. However, such one-size-fits-all guidelines may not be appropriate for all patients. In this study, a review of all data days that included one or more measurements exceeding the threshold of concern suggested there was not a consistent time of day when concerning temperature asymmetries typically occurred (Figure 7). A fixed, single time point for monitoring temperature asymmetry risks under- or overestimating risk.

Greater visibility into the day-by-day temperature asymmetry fluctuations may be meaningful to providers and patients, similar to how continuous glucose monitoring information and a core set of glucose level statistics overcame the limited information provided by isolated glucose readings [29]. Such continuous monitoring may better support health care providers to personalize care for the individual. We anticipate that continuous temperature asymmetry signals will be used increasingly in clinical trials for managing diabetic foot health.

Limitations and Future Opportunities

Many unknowns remain regarding the underlying physiological and environmental factors that contribute to intraday fluctuations in plantar temperature asymmetry. Additional clinical trials are necessary to determine the degree to which continuous temperature monitoring improves DFU detection compared to once-daily temperature measurements. While ongoing research is necessary to better understand these mechanisms, we remain confident that continuous, repeated measurements throughout the day build more confidence for clinical decision-making.

Further research is needed to determine if and how plantar temperature warning thresholds should be adapted for continuous monitoring. Current guidelines, based on clinical trials using once- or twice-daily measurements, recommend a hot spot indicator threshold of 2.2 °C for 2 or more consecutive days. However, with the recent availability of continuous plantar temperature monitoring, it may be necessary to reconsider this fixed threshold. A variable or individualized threshold, accounting for factors such as activity levels, comorbid peripheral vascular disease, previous ulcer locations, and environmental influences, may be more appropriate.

Moreover, continuous temperature monitoring provides an opportunity to move beyond the static threshold-based criteria to other statistical strategies, such as time in range versus time above threshold, intraday trend and variability metrics, or interday temperature profile comparisons. This evolution parallels the development of extended analytics during the transition from discrete self-monitoring of blood glucose to continuous glucose monitoring systems [16,17]. Developing updated reporting standards will require further research and multidisciplinary collaboration to achieve consensus. Ultimately, we envision a future in which continuous temperature monitoring and its associated analytics become as ubiquitous and clinically valuable as continuous glucose monitoring and other continuous physiologic monitoring tools.

XSL•FO
RenderX

Conclusions

This study demonstrates the high prevalence (82% of data days) of nonstationary temperature asymmetry signals in individuals with diabetes and history of ulceration. Our results highlight the value of continuous plantar temperature monitoring for diabetic foot health management, relative to once-daily measurements. Continuous monitoring may improve the accuracy of plantar temperature measurement, unlock new diagnostic capabilities, and support personalized monitoring. Future research is required to better understand the underlying mechanisms of plantar temperature asymmetry fluctuations and to prove the clinical value of continuous monitoring for reducing the risk of DFU formation. Continuous monitoring is an exciting advancement in remote temperature monitoring for diabetes health span extension.

Acknowledgments

The authors thank Orpyx Medical Technologies, Inc., for funding this research. ORPYX is a registered trademark of Orpyx Medical Technologies Inc. in jurisdictions where registered.

Data Availability

The datasets generated and analyzed for this study can be made available upon reasonable request to the authors.

Authors' Contributions

MH, E Matijevich, E Minty, and EB contributed to the conceptualization of the study. The methodology was developed by MH, E Matijevich, E Minty, and EB. Formal analysis and data curation were conducted by MH and E Matijevich. Visualization was carried out by MH and E Matijevich. Writing of this paper was undertaken by MH, E Matijevich, EB, E Minty, and BL. All authors read and contributed to the editing of the paper.

Conflicts of Interest

BL sits on the clinical advisory board of Orpyx Medical Technologies, Inc. E Minty is under a consulting agreement with Orpyx Medical Technologies Inc. as medical director of their remote patient monitoring service. E Minty discloses equity ownership in Orpyx Medical Technologies Inc. and has family members in the employ of Orpyx Medical Technologies Inc. MH, E Matijevich, and EB are employees of Orpyx Medical Technologies Inc.

Checklist 1
STROBE reporting checklist.
[DOCX File, 42 KB - mhealth_v13i1e73187_app1.docx]

References

1. Armstrong DG, Tan TW, Boulton AJM, Bus SA. Diabetic foot ulcers: a review. JAMA 2023 Jul 3;330(1):62-75. [doi: 10.1001/jama.2023.10578] [Medline: 37395769]
2. Armstrong DG, Boulton AJM, Bus SA. Diabetic foot ulcers and their recurrence. N Engl J Med 2017 Jun 15;376(24):2367-2375. [doi: 10.1056/NEJMra1615439] [Medline: 28614678]
3. Bus SA, Lavery LA, Monteiro-Soares M, et al. Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2019 update). Diabetes Metab Res Rev 2020 Mar;36 Suppl 1:e3269. [doi: 10.1002/dmrr.3269] [Medline: 32176451]
4. Bergtholdt HT. Temperature assessment of the insensitive foot. Phys Ther 1979 Jan;59(1):18-22. [doi: 10.1093/ptj/59.1.18] [Medline: 424445]
5. Golledge J, Fernando ME, Alahakoon C, et al. Efficacy of at home monitoring of foot temperature for risk reduction of diabetes-related foot ulcer: a meta-analysis. Diabetes Metab Res Rev 2022 Sep;38(6):e3549. [doi: 10.1002/dmrr.3549] [Medline: 35605998]
6. Najafi B, Chalifoux CB, Everett JB, Razjouyan J, Brooks AE, Armstrong GD. Cost effectiveness of smart insoles in preventing ulcer recurrence for people in diabetic foot remission. Wound Care Manage 2018;1(1). [doi: 10.15761/WCM.1000102]
7. Frykberg RG, Gordon IL, Reyzelman AM, et al. Feasibility and efficacy of a smart mat technology to predict development of diabetic plantar ulcers. Diabetes Care 2017 Jul;40(7):973-980. [doi: 10.2337/dc16-2294] [Medline: 28465454]
8. Reyzelman AM, Koelewyn K, Murphy M, et al. Continuous temperature-monitoring socks for home use in patients with diabetes: observational study. J Med Internet Res 2018 Dec 17;20(12):e12460. [doi: 10.2196/12460] [Medline: 30559091]
9. Gordon IL, Rothenberg GM, Lepow BD, et al. Accuracy of a foot temperature monitoring mat for predicting diabetic foot ulcers in patients with recent wounds or partial foot amputation. Diabetes Res Clin Pract 2020 Mar;161:108074. [doi: 10.1016/j.diabres.2020.108074] [Medline: 32109516]
10. Armstrong DG, Lavery LA, Liswood PJ, Todd WF, Tredwell JA. Infrared dermal thermometry for the high-risk diabetic foot. Phys Ther 1997 Feb;77(2):169-175. [doi: 10.1093/ptj/77.2.169] [Medline: 9037217]

XSL•FO
RenderX

11. van Netten JJ, Prijs M, van Baal JG, Liu C, van der Heijden F, Bus SA. Diagnostic values for skin temperature assessment to detect diabetes-related foot complications. Diabetes Technol Ther 2014 Nov;16(11):714-721. [doi: 10.1089/dia.2014.0052] [Medline: 25098361]
12. Beach C, Cooper G, Weightman A, Hodson-Tole EF, Reeves ND, Casson AJ. Monitoring of dynamic plantar foot temperatures in diabetes with personalised 3D-printed wearables. Sensors (Basel) 2021 Mar 2;21(5):1717. [doi: 10.3390/s21051717] [Medline: 33801346]
13. Gautam N, Ghanta SN, Mueller J, et al. Artificial intelligence, wearables and remote monitoring for heart failure: current and future applications. Diagnostics (Basel) 2022 Nov 26;12(12):2964. [doi: 10.3390/diagnostics12122964] [Medline: 36552971]
14. Bello NA, Schwartz JE, Kronish IM, et al. Number of measurements needed to obtain a reliable estimate of home blood pressure: results from the improving the detection of hypertension study. J Am Heart Assoc 2018 Oct 16;7(20):e008658. [doi: 10.1161/JAHA.118.008658] [Medline: 30371272]
15. Mendez M, Bianchi AM, Villantieri O, Cerutti S. Time-varying analysis of the heart rate variability during REM and non REM sleep stages. Presented at: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 30 to Sep 3, 2006; New York, NY p. 3576-3579. [doi: 10.1109/IEMBS.2006.260067]
16. Rodbard D. Continuous glucose monitoring: a review of successes, challenges, and opportunities. Diabetes Technol Ther 2016 Feb;18 Suppl 2(Suppl 2):S3-S13. [doi: 10.1089/dia.2015.0417] [Medline: 26784127]
17. Rodbard D. The ambulatory glucose profile: opportunities for enhancement. Diabetes Technol Ther 2021 May;23(5):332-341. [doi: 10.1089/dia.2020.0524] [Medline: 33107758]
18. Verma N, Haji-Abolhassani I, Ganesh S, et al. A novel wearable device for continuous temperature monitoring & fever detection. IEEE J Transl Eng Health Med 2021;9:2700407. [doi: 10.1109/JTEHM.2021.3098127] [Medline: 34765323]
19. Leister KR, Wurdeman SR. A walking bout reveals altered foot thermodynamics associated with unilateral transtibial amputation among individuals with type 2 diabetes. Prosthetics & Orthotics International 2021;45(2):178-183. [doi: 10.1177/0309364620968642]
20. Abbott CA, Chatwin KE, Foden P, et al. Innovative intelligent insole system reduces diabetic foot ulcer recurrence at plantar sites: a prospective, randomised, proof-of-concept study. Lancet Digit Health 2019 Oct;1(6):e308-e318. [doi: 10.1016/S2589-7500(19)30128-1] [Medline: 33323253]
21. Matijevich E, Minty E, Bray E, Bachus C, Hajizadeh M, Liden B. A multi-faceted digital health solution for monitoring and managing diabetic foot ulcer risk: a case series. Sensors (Basel) 2024 Apr 23;24(9):2675. [doi: 10.3390/s24092675] [Medline: 38732781]
22. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. SciPy 2010;7(1). [doi: 10.25080/Majora-92bf1922-011]
23. MacKinnon JG. Approximate asymptotic distribution functions for unit-root and cointegration tests. J Bus Econ Stat 1994 Apr;12(2):167-176. [doi: 10.1080/07350015.1994.10510005]
24. Mushtaq R. Augmented Dickey Fuller Test. SSRN J. Preprint posted online on Aug 17, 2011. [doi: 10.2139/ssrn.1911068]
25. aan de Stegge WB, Van Netten JJ, Bus SA. Does the skin heat up before it breaks down in diabetic foot ulceration? Diabetes Metabolism Res 2023 Jul;39(5):e3621. [doi: 10.1002/dmrr.3621]
26. Liew H, Tang W, Plassmann P, et al. Infrared thermography shows that a temperature difference of 2.2°C (4°F) or greater between corresponding sites of neuropathic feet does not always lead to a diabetic foot ulcer. J Diabetes Sci Technol 2024 May 6:19322968241249970. [doi: 10.1177/19322968241249970] [Medline: 38708580]
27. Reddy PN, Cooper G, Weightman A, Hodson-Tole E, Reeves ND. Walking cadence affects rate of plantar foot temperature change but not final temperature in younger and older adults. Gait Posture 2017 Feb;52:272-279. [doi: 10.1016/j.gaitpost.2016.12.008] [Medline: 28012341]
28. Carabott M, Formosa C, Mizzi A, Papanas N, Gatt A. Thermographic characteristics of the diabetic foot with peripheral arterial disease using the angiosome concept. Exp Clin Endocrinol Diabetes 2021 Feb;129(2):93-98. [doi: 10.1055/a-0838-5209] [Medline: 30873574]
29. Simonson GD, Holt EH, Grady M, Hurrell G, Gaudiani LM, Bergenstal RM. Unleashing the potential of blood glucose monitoring data with the Ambulatory Glucose Profile Report. Clin Diabetes 2024;42(4):550-560. [doi: 10.2337/cd23-0092] [Medline: 39429460]

Abbreviations:

ABI: ankle-brachial index
ADF: augmented Dickey-Fuller
AIC: Akaike information criterion
CAD/CAM: computer-aided design/computer-aided manufacturing
DFU: diabetic foot ulcer
IMU: inertial measurement unit
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by L Buis; submitted 27.02.25; peer-reviewed by CP Agu, C Ashmore; revised version received 13.06.25; accepted 21.07.25; published 12.09.25.

Please cite as:
Hajizadeh M, Matijevich E, Bray E, Minty E, Liden B
Exploring the Value of Continuous Plantar Temperature Monitoring for Diabetic Foot Health Management: Observational, Prospective Cohort Study
JMIR Mhealth Uhealth 2025;13:e73187
URL: https://mhealth.jmir.org/2025/1/e73187
doi:10.2196/73187

© Maryam Hajizadeh, Emily Matijevich, Emily Bray, Evan Minty, Brock Liden. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Temporal Trajectories in Sleep, Temperature Trends, Cardiorespiratory, and Activity Metrics Measured via Oura Ring During Pregnancy: Large-Scale Observational Analysis

Rebecca Adaimi[1], PhD; Nina Thigpen[1], PhD; Alicia Clausel[1], MPH, MS; Neta Gotlieb[1], PhD; Ketan Patel[1], PhD; Massimiliano de Zambotti[1], PhD

Ouraring Inc., San Francisco, CA, United States

Corresponding Author:
Massimiliano de Zambotti, PhD
Ouraring Inc.
222 Kearny Street, 7th Floor
San Francisco, CA, 94108
United States
Phone: 1 415 818 9430
Email: max@ouraring.com

Abstract

Background: Pregnancy and childbirth involve significant health challenges, including preventable maternal deaths, severe complications, and disparities tied to social determinants, emphasizing the need for improved maternal care. Pregnancy could benefit from a more comprehensive, continuous care model that captures dynamic changes and enhances maternal-fetal outcomes.

Objective: This large-scale, real-world, high-density study aims to use wearable data to investigate maternal biobehavioral trajectories for pregnancies leading to loss, preterm, and term births, exploring how demographic factors like age and body mass index (BMI) affect these trajectories.

Methods: Retrospective observational analysis of pregnancies from a sample of 10,318 and 18- to 51-year-old female Oura Ring users (324 preterm births, 5039 term births, 4955 pregnancies ending in loss before 20 weeks of gestation). Oura biobehavioral data were analyzed across a 64-week window encompassing 8 weeks prepregnancy, through pregnancy, and post partum, via generalized estimating equation (GEE) statistical modeling.

Results: Gestational age emerged as a significant factor across all domains among term pregnancies (P<.001). During the first trimester, participants experienced marked sleep changes, peaking around week 9 and characterized by more time in bed (+30 min), asleep (+15 min), and awake (+15 min) compared with prepregnancy. Metrics declined and stabilized in the second trimester; by the third trimester, time in bed returned to baseline, while sleep remained reduced and wakefulness elevated. At birth, time in bed and wakefulness peaked, and sleep duration reached its minimum, with nighttime wake exceeding 3 SDs above baseline. Temperature changes were more pronounced, sustained, and occurred earlier than sleep changes—becoming evident by week 4, peaking at +0.3 °C above baseline by week 9, and showing a steady decline until birth. A secondary, modest increase (+0.1 °C) was observed near birth, followed by a decline postpartum. Heart rate (HR) increased steadily, peaking at +10 bpm above baseline at week 32, while HR variability declined by >15 milliseconds in a mirrored pattern. Respiratory rate peaked around week 9 and declined thereafter. Step count declined in the first trimester, with a ≈2000-step reduction at around week 8. After a slight rebound midpregnancy, activity declined again, reaching its lowest point near birth, with >2500 fewer steps than prepregnancy. Age and BMI showed significant but modest interaction effects (all P<.01). In pregnancies ending in loss, deviations emerged up to 2 weeks prior. Time in bed decreased starting ≈2 weeks before loss (P<.001), followed by reductions in sleep duration (P<.001), temperature trends (P<.001), respiratory rate (P=.019), HR (P=.005), and awake time (P=.033).

Conclusions: These findings highlight the complex dynamics underlying changes in participants' sleep, temperature trends, cardiorespiratory, and activity data throughout pregnancy, involving extensive adaptations. A deeper focus on normative changes can advance maternal-fetal medicine, improve clinical outcomes, and address scientific gaps.

(JMIR Mhealth Uhealth 2025;13:e80213) doi:10.2196/80213

KEYWORDS

pregnancy; consumer wearables; sleep; activity; heart rate; postpartum

XSL•FO
RenderX

Introduction

More than 4 births occur every second of every day worldwide. In the United States alone, more than 10,000 babies are born (live births) each day [1]. According to the World Health Organization, approximately 287,000 women died during pregnancy or childbirth in 2020, equating to one maternal death every 2 minutes [2]. Data from the 2017-2019 Maternal Mortality Review Committees in the United States showed that over 80% of pregnancy-related deaths (occurring during pregnancy, delivery, and up to a year postpartum) were determined to be preventable [3]. Beyond mortality, millions of women experience severe maternal morbidity, including life-threatening complications such as hemorrhage, preeclampsia, and infections. For example, in the United States, as many as 60,000 women per year are affected by severe maternal morbidity, most of which could have been avoided with timely interventions [4]. Miscarriage (loss of a fetus before the 20th week of pregnancy) and stillbirth (loss of a fetus after 20 weeks of pregnancy) are common and represent significant emotional and physical trauma. Importantly, maternal health outcomes are deeply tied to social determinants of health. Women in underserved communities face higher risks due to limited access to health care, systemic inequalities, and a lack of prenatal care.

Pregnancy represents a highly complex physiological state, characterized by extensive biological, hormonal, and behavioral adaptations that optimize maternal and fetal development. Current prenatal care is largely centered on assessing health and screening for disease at discrete time points during scheduled visits, focusing on key physical markers such as blood pressure, weight, and fetal growth. In the United States, standard prenatal care varies based on a woman's risk profile. Standard prenatal care globally includes relatively few touchpoints compared with the length of pregnancy and the significant physiological and emotional changes that occur. Postpartum, care and connection often drop off, with sometimes only a single visit after delivery. Those considered high-risk may require more frequent visits and ad hoc interventions depending on the severity and progression of pregnancy-associated factors, but the care, nevertheless, is brief and infrequent [5].

While valuable, this episodic approach often fails to capture the dynamic and continuous nature of physiological, behavioral, and psychosocial changes throughout pregnancy. Critical fluctuations or early signs of atypical trajectories may go unnoticed between visits. Notably, it overlooks individual trajectories relative to one's baseline, potentially missing significant deviations that, while within population norms, may still indicate important risks or shifts in health status. This limitation underscores the need for a more comprehensive, individualized, longitudinal model of prenatal care, incorporating continuous monitoring and integrating emerging biobehavioral and contextual data. Such an approach would provide a deeper understanding of pregnancy as an evolving process and improve maternal-fetal outcomes.

Traditional maternal health research has been instrumental in shaping our understanding of pregnancy, providing critical insights into fetal development, maternal physiology, and the factors contributing to pregnancy outcomes. Through clinical assessments and structured studies, this body of work has established foundational knowledge and improved care practices.

Thanks to the advancements in sensor technology and digital health, multisensor consumer wearables (ie, smartwatches, smart rings, and other health trackers) in particular, present an opportunity to complement and extend traditional research approaches. However, this line of work is still in its infancy, and more research is needed to understand their role in supporting the reproductive life course [6-8]. By enabling passive, large-scale monitoring of biobehavioral data—such as sleep, temperature trends, cardiorespiratory, and activity patterns—wearables could provide a dynamic and real-time perspective on maternal health throughout pregnancy. Wearables demonstrate feasibility in estimating ovulation compared with calendar-based fertility tracking [9], presenting a potential avenue to support women's fertility journeys. Grant and Smarr [10] showcase the possibility of passive early pregnancy detection through continuous analysis of distal skin temperature, while others suggest the use of a wearable to proxy the clinically estimated delivery date [11,12]. Other studies showcase the potential to use wearables in relation to pregnancy risks (eg, classification of healthy pregnant vs nonpregnant women) [13] and complications, including gestational diabetes [14], gestational hypertension [15], preterm labor [11,16], and postpartum depression [17].

The maternal health market is evolving rapidly, with a growing focus on digital health solutions and consumer-centric innovations to address the diverse needs of pregnant women and new mothers. While the role of consumer wearable in maternal health is still unclear, initial feasibility for continuous high-resolution passive monitoring of pregnancy using a multisensor consumer wearable device has been shown by Keeler et al [18]. In that study, biobehavioral data collected via the Oura Ring Gen2 and questionnaire data were analyzed from 120 pregnancies, including 97 full-term pregnancies and 23 cases of pregnancy loss. Two additional recent studies evaluating signals before, during, and after pregnancy highlight changes across 3 distinct physiological states. A retrospective analysis of sleep data from 2540 Fitbit users in the US and Canada examined patterns 12 weeks before pregnancy, throughout gestation, and during the first 20 weeks postpartum [19]. Similarly, an evaluation of 757 pregnancies for Apple Watch users described trends in sleep, as well as exercise minutes, steps, and physiologic signals such as heart rate (HR) and respiratory rate (RR) [20]. Overall, these studies demonstrated that high-resolution data throughout pregnancy offer valuable insights into temporal and individual variability, as well as differences across different biobehavioral data, such as, sleep, temperature trends, cardiorespiratory, and activity.

This work aims to extend the study by Keeler et al [18] by conducting a large-scale mapping of biobehavioral changes across pregnancy—from preconception to postpartum—for pregnancies ending in loss (before 20 weeks), preterm birth, and term birth, using data from a large sample of 10,318 Oura Ring users. This retrospective observational study—the largest compilation of pregnancies using wearable data to date—aims

XSL•FO
RenderX

to delineate the physiological and behavioral characteristics of pregnancy and to evaluate whether factors such as advanced maternal age and overweight status influence biobehavioral data trajectories during pregnancy. For example, existing literature indicates that advanced maternal age [21] and overweight/obesity [22] significantly increase the risk of adverse maternal outcomes; however, this study uniquely leverages wearable technology to provide continuous, real-world biobehavioral data across pregnancy, offering novel insights into how age and BMI impact sleep, activity, and cardiorespiratory patterns throughout gestation.

Methods

Dataset

A cohort of Oura Ring users was included in the study. Users were eligible if they self-identified as female (ie, sex assigned at birth), were pregnant between May 2023 and November 2024, and were aged 18 to 51 years. The date range was selected to ensure that participant data were collected using the latest Oura algorithms and Gen3 hardware. The final sample included 10,318 unique pregnancies. Among these, 5363 pregnancies led to live births (324 preterm and 5039 at term), while 4955 ended in loss before 20 weeks of gestation (eg, miscarriage or other early pregnancy losses).

Ethical Considerations

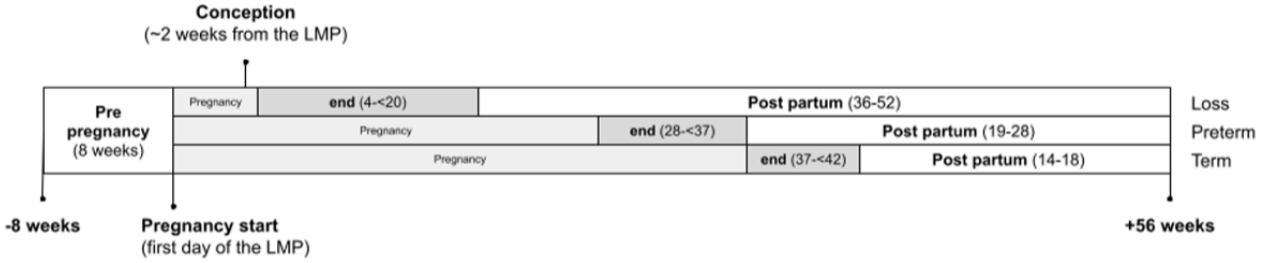
This study involved secondary analysis of data from adult users of the Oura Ring, collected by Oura Health Oy (Oulu, Finland) in accordance with its Terms of Use [23] and Privacy Policy [24], which users consent to prior to data collection. By agreeing to the Privacy Policy, users consented to the use of their data for research and analytical purposes. All data used in this study were fully deidentified and aggregated in accordance with Oura's internal data governance protocols. The research protocol was reviewed and approved by Oura's internal science and legal teams, which oversee regulatory and ethical compliance for research activities. The study was deemed exempt from formal ethics board review as it involved only secondary analysis of anonymized, nonidentifiable data. Additional safeguards included restricted data access solely for verification and analytical accuracy. No compensation was provided to participants.

Procedure

This is a retrospective observational study on Oura users who opted to use the Oura App's Pregnancy Insights feature (see below) during pregnancy. Sleep, temperature trends, cardiorespiratory, and activity data from Oura Ring were accessed for a 64-week window encompassing the period from 8 weeks prior to the estimated start of pregnancy (the first day of the last menstrual period [LMP]) to 56 weeks following that.

Participants were grouped based on self-reported pregnancy outcomes and the gestational age at pregnancy end, that is, pregnancies ending in loss versus live births (with live births further categorized into preterm and term births (Figure 1).

Figure 1. Schematics of the data collection timeline across pregnancies resulting in loss, preterm birth, and term birth are provided. The start and end of pregnancy, as well as the estimated time of conception, are highlighted in the graph. LMP: last menstrual period.



Gestational age was also used to classify pregnancy into 3 trimesters. Trimester 1 included weeks 0 days to 13 weeks 6 days of gestation. Trimester 2 ranged from 14 weeks 0 days to 27 weeks 6 days. Trimester 3 covered 28 weeks 0 days until 40 weeks 6 days of gestation.

Oura Bio-Behavioral Data Collection

All participants wore an Oura Ring Gen3 (Oura Health Oy, Oulu, Finland), a noninvasive, water-resistant, and commercially available multisensor device that continuously measures finger blood flow, peripheral oxygen saturation, skin temperature, and movement via photoplethysmography green, red, and infrared LEDs, a negative skin temperature coefficient thermistor (±0.1 °C resolution), and 3D accelerometry. Oura Ring connects to mobile platforms via Bluetooth Low Energy, and data are transmitted to secure cloud-based servers for processing and storage. Oura Ring battery lasts up to 7 days.

By using proprietary algorithms, Oura Ring measures several health-related indices. Please refer to de Zambotti et al [25] for an overview of consumer wearable sleep metrics and their operationalization. A list of Oura Ring metrics used in this study is provided in Table 1.

XSL•FO
RenderX

Table 1. Oura Ring Gen3 metrics.

Measure	Description
Sleep	• Daily Time in bed (hours), asleep (hours), awake (hours), and time spent in "light" sleep (equivalent of PSG[a] N1+N2 sleep; % time asleep), "deep" sleep (equivalent of PSG N3 sleep; % time asleep) and rapid-eye-movement (REM) sleep (% time asleep), for the Oura classified main nocturnal sleep (>3h sleep duration). For each participant, single daily measures of overnight sleep measures were obtained. These metrics are calculated between the Oura Ring's automatically detected onset and offset of nocturnal sleep windows, approximating individuals' bedtimes and wakeup times via proprietary algorithms that combine features from PPG[b], temperature trends, and accelerometer sensors, trained on expert-scored PSG sleep records. The performance of Oura Ring for sleep measurement has been evaluated across multiple peer-reviewed studies across different demographics, providing overall good agreement with the gold standard PSG [26-31]. In the most recent Oura-sponsored performance evaluation of Oura Ring Gen3 [29], Oura Ring achieved a sensitivity of 73.0% for wake, 76.7% for "light" sleep, 64.0% for "deep" sleep, and 72.6% for REM sleep, in healthy individuals.
Temperature	• Daily peak sleep skin temperature (°C). For each of the participants, a single daily measurement of peak sleep skin temperature during sleep was obtained. Oura internal validation (unpublished company data) [32] shows good concordance of the skin temperature sensor with reference standards in both laboratory and free-living conditions.
Cardiorespiratory	• Daily average sleep PPG-based HR[c] (bpm) HRV[d] (ms). For each of the participants and night, HR and HRV have been averaged across the Oura Ring's classified main nocturnal sleep to provide a daily measure. Oura Ring Gen3 estimates HR at 30s resolution and time-domain HRV, that is, the root mean square of successive differences between normal heartbeats (RMSSD) for consecutive 5-minute windows during the nocturnal sleep. RMSSD captures the high-frequency component of HRV reflecting vagal autonomic nervous system function. Please see Shaffer and Ginsberg [33] for an overview of HRV assessment and meaning. As a note, the term HRV is used for consistency, considering that Oura Ring measures pulse rate variability from PPG signal, which reflects variation in blood flow and does not directly measure heart's electrical depolarization signal. Average-per-night test comparing Oura Ring versus reference standard electrocardiogram-based HR and HRV metrics, shows good level of concordance with low mean biases [34]. • Daily average respiratory rate (RR; breaths/min). RR is measured by computing PPG-HR time series data during nocturnal sleep. For each of the participants and night, RRs have been averaged across Oura classified main nocturnal sleep to provide a daily measure.
Activity	• Daily total steps (number). Oura Ring calculates the number of daily steps within the time period spanning across the 24h. For each of the participants, a single daily measure of total steps was obtained. • Oura activity data show good concordance for laboratory and free-living testing when compared with reference standards, particularly when considering group-level comparisons [35].

[a]PSG: polysomnography.

[b]PPG: photoplethysmography.

[c]HR: heart rate.

[d]HRV: heart rate variability.

Oura App and the Pregnancy Insights Feature

The Oura App Pregnancy Insights feature is designed to provide data-driven information (how sleep and other biometrics change with pregnancy) and in-app education to help users navigate their pregnancy. The feature tracks gestational age, providing weekly updates about physiological changes detected by the Oura Ring. Pregnancy Insights requires a user to turn on the feature in the "Women's Health" settings of the app, confirming pregnancy. The feature enables self-report of estimated due date and first day of the last period, as well as date of birth or pregnancy loss.

Age, prepregnancy height and weight, and country of residence have been extracted from the Oura App users' profiles. BMI has been calculated as weight (kg) × height $(m)^{-2}$.

Data Preprocessing

Oura Ring Data

This step included the generation of "local" date-time indexing, removal of duplicate time points, and for sleep, temperature trends, and cardiorespiratory data, time points not belonging to main sleep periods (eg, naps, diurnal sleep) have been removed. In addition, pregnancies with less than 40% of Oura Ring data available [18], for prepregnancy and for each trimester during pregnancy, were excluded.

Identifying Pregnancy Start and End Date for Pregnancies Leading to Birth and Early Pregnancy End

Pregnancy start (standardized point for estimating gestational age) is operationalized as the first day of the LMP, which differs from conception (biological event), that is, the moment when the sperm fertilizes the egg, typically occurring around 14 days after the start of the LMP in a standard 28-day cycle. Pregnancy start is computed based on self-reported expected due date if known (expected due date minus 280 days; 5090/10,318, 49.3% pregnancies) or directly based on their self-reported first day of their last period (5228/10,318, 50.7% pregnancies). Pregnancy end was operationalized based on self-reported data, either as a birth date or as an early pregnancy loss.

Statistical Analysis

A first statistical analysis was conducted to evaluate patterns of change in Oura bio-behavioral data throughout pregnancy,

XSL•FO
RenderX

specifically focusing on pregnancies leading to term births. All bio-behavioral data were first aligned to the estimated start of pregnancy. The analysis was performed on average weekly bio-behavioral data from Oura, encompassing both the pregnancy and postpartum periods. To account for inherent interindividual variability in baseline physiological levels, each metric was normalized to a prepregnancy baseline. This baseline was defined as the average value over the 4 weeks preceding the estimated pregnancy start. Normalization was performed by calculating Z-scores, representing the number of SDs a value deviated from the mean of its respective prepregnancy baseline. This standardization allowed for a more focused capture and comparison of trends and patterns of physiological change throughout pregnancy across the cohort.

To appropriately account for the longitudinal within-subject correlation structure inherent in repeated measurements from the same individual, a generalized estimating equation (GEE) statistical modeling framework was used [36]. An exchangeable working correlation structure was specified for the GEE. This choice assumes a constant correlation between any 2 measurements within the same individual, regardless of the time interval between them, and is commonly used for its robustness. The GEE framework provides consistent estimates of population-averaged effects and computes robust standard errors, which are valid even if the assumed working correlation structure is not perfectly specified.

The nonlinear relationship between each normalized physiological metric and gestational age (our time variable) was modeled using B-spline basis functions. The specific flexibility of the curve was determined by using 8 degrees of freedom for the B-spline. This choice of degrees of freedom was guided by a balance between capturing complex nonlinear patterns (assessed through visual inspection of predicted fits) and maintaining statistical stability.

Subsequently, the impact of continuous covariates, specifically maternal age and prepregnancy BMI, as well as their interactions with gestational age, was investigated. This was achieved by extending the primary GEE model to include corresponding B-spline terms for these covariates and tensor product spline interaction terms with gestational age. These tensor product splines allowed for the characterization of how the nonlinear trend of each physiological metric over gestational age might vary across different levels of age and BMI.

The statistical significance of effects was assessed using Wald tests. First, for the gestational age-only model, a joint Wald test was performed on the set of B-spline coefficients for gestational age. A significant *P* value indicated a statistically significant nonlinear population-level trend in the physiological metric throughout pregnancy. Then, for models incorporating maternal age and BMI, joint Wald tests were conducted on the sets of coefficients corresponding to their main effect B-spline terms and interaction terms with gestational age. A statistically significant *P* value for these interaction terms indicated that maternal age or BMI had a significant moderating effect on the physiological trend over gestational age.

All tested effects (including the main trends over gestational age, and the main and interaction effects of age and BMI) were found to be statistically significant (*P* values <.05). However, it was observed that the magnitude of some of these effects (ie, the effect size) was small in practical terms, which was further explored through visualization of predicted trends and quantification of absolute differences over each trimester.

A second statistical analysis was conducted to evaluate differential trajectories in Oura bio-behavioral data between pregnancies ending before 20 weeks and those resulting in term live births, focusing on the period immediately preceding early pregnancy end. Specifically, we investigated how pregnancy outcome influenced physiological metrics during the 28 days leading up to the day of early pregnancy end. Prior to analysis, data underwent preprocessing, including linear interpolation for missing values and subsequent normalization via z-scoring with respect to each individual's prepregnancy baseline (defined as the average over the 4 weeks preceding pregnancy start). For each pregnancy that ended before 20 weeks, we extracted bio-behavioral data for the 28 days immediately preceding the end of pregnancy. To establish a time-matched control, data from a randomly sampled pregnancy resulting in a term live birth were selected for the equivalent gestational age (eg, if a pregnancy end occurred at 7 weeks 1 day estimated gestational age, data from 3 weeks 2 days to 7 weeks 1 day were extracted for both the early pregnancy end case and its matched control).

A GEE model, like that used in the first analysis, was applied. In this model, the day relative to pregnancy end served as the primary independent variable to predict each normalized physiological metric, while simultaneously controlling for pregnancy outcome (pregnancy end before 20 weeks or term live birth). The nonlinear relationship between each normalized physiological metric and the day relative to a pregnancy ending before 20 weeks was modeled using a B-spline with 3 degrees of freedom. The model also included a linear term for pregnancy outcome and a linear interaction term between pregnancy outcome and the day relative to early pregnancy end. To determine if the physiological trajectory for pregnancies resulting in loss was statistically different from the trajectory for pregnancies resulting in a term live birth, we assessed the overall effect of pregnancy outcome using a joint Wald test on both the linear term for pregnancy outcome and the linear interaction term. Effect sizes were quantified as the mean difference between the modeled responses for pregnancies resulting in term live birth and those ending before 20 weeks, calculated across 2 distinct periods: the entire 28 days preceding early pregnancy end, and specifically during the 7 days immediately prior to the early pregnancy's end.

All models were implemented using the Python package statsmodels [37]. This study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines. A completed STROBE checklist is provided in Multimedia Appendix 1.

XSL•FO
RenderX

Results

Demographics and Pregnancy Characteristics

Participants in the study had a mean age of 32.2 (SD 4.5) years and a mean BMI of 25.3 (SD 5.4) kg/m², based on data collected prior to pregnancy.

Participants in the study represented various regions, including the United States (7398/10,318, 71.7% participants), the European Union (1104/10,318, 10.7% participants), the United Kingdom (351/10,318, 3.4% participants), Canada (320/10,318, 3.1% participants), Australia (155/10,318, 1.5% participants), and other countries (371/10,318, 3.6% participants). For 619 out of 10,318 (6.0%) participants, the location was unknown.

Distribution of Oura Ring sizes in the study population were the following: Size 6 (856/10,318, 8.3% participants), Size 7 (2270/10,318, 22% participants), Size 8 (3281/10,318, 31.8% participants), Size 9 (2384/10,318, 23.1% participants), Size 10 (1011/10,318, 9.8% participants), Size 11 (382/10,318, 3.7% participants), Size 12 (103/10,318, 1% participants), and Size 13 (31/10,318, 0.3% participants).

The distribution of pregnancies as a function of the estimated gestational age at which the pregnancy ended, for pregnancies ending in loss, preterm birth, and term birth, is provided in Multimedia Appendix 2. Self-reports of early pregnancy end did not differentiate between types of loss, such as spontaneous abortion, ectopic pregnancy, molar pregnancy, or pregnancy termination; therefore, this study does not adjust for the specific etiology of pregnancy loss. The distribution of participants' age and BMI for pregnancies ending in loss, preterm birth, and term birth is provided in Multimedia Appendix 3.

Prepregnancy age (Kruskal-Wallis, P=.44) and BMI (Kruskal-Wallis, P=.08) did not significantly differ between pregnancies ending in loss, preterm, or term birth.

Pattern of Changes in Oura Bio-Behavioral Data Throughout Pregnancy in Pregnancies Leading to Term Births

Overview

Absolute changes in Oura data throughout pregnancy are provided in Multimedia Appendix 4. The number of women contributing valid data at each gestational week is shown in Multimedia Appendix 5. Detailed results for the GEE model are provided in Multimedia Appendices 6 and 7. Changes in normalized Oura bio-behavioral data throughout pregnancy are displayed in Figure 2 (sleep), Figure 3 (temperature trends), Figure 4 (cardiorespiratory), and Figure 5 (activity).

XSL•FO
RenderX

Figure 2. Changes in Oura sleep data throughout pregnancy. Data (Z-scores - normalized data based on prepregnancy individuals' baseline) are plotted as the weekly median (thick line) and the 50th percentile range, separately for pregnancies leading to loss (N=4955), preterm births (N=324), and term (N=5039) births. Gestational age 0 (the start of pregnancy) is defined as the first day of the last menstrual period. The red, purple, and green vertical bars delineate the mean pregnancy end and one SD for pregnancy loss, preterm birth, and term birth, respectively.

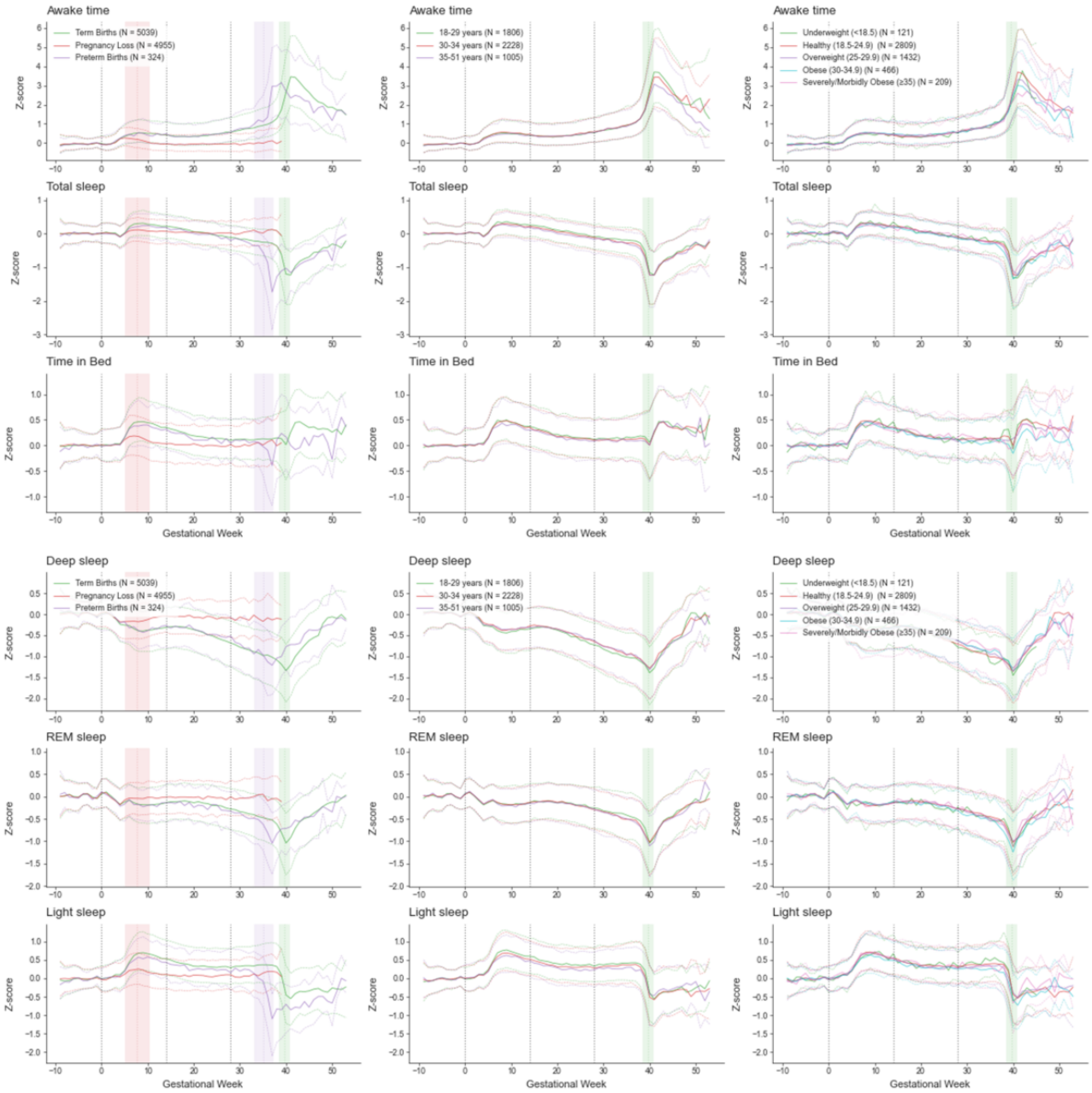
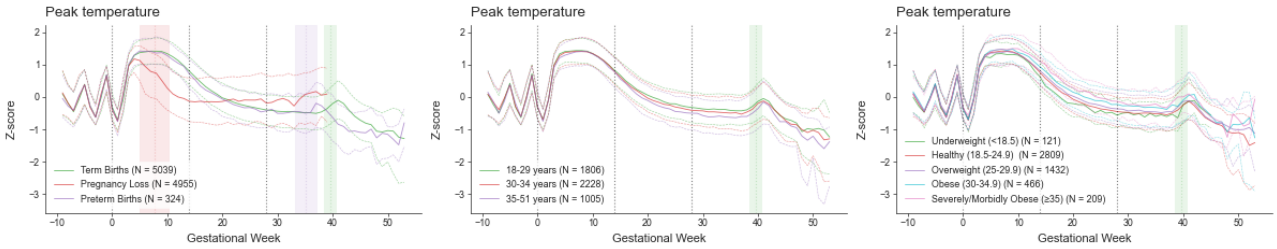


Figure 3. Changes in Oura temperature trends data throughout pregnancy. Data (Z-scores - normalized data based on prepregnancy individuals' baseline) are plotted as the weekly median (thick line) and the 50th percentile range, separately for pregnancies ending in loss (N=4955), preterm births (N=324), and term (N=5039) births. Gestational age 0 (the start of pregnancy) is defined as the first day of the last menstrual period. The red, purple, and green vertical bars delineate the mean pregnancy end and one SD for pregnancy loss, preterm birth, and term birth, respectively.



XSL•FO
RenderX

Figure 4. Changes in Oura cardiorespiratory data throughout pregnancy. Data (Z-scores - normalized data based on pre-pregnancy individuals' baseline) are plotted as the weekly median (thick line) and the 50th percentile range, separately for pregnancies ending in loss (N=4955), preterm births (N=324), and term (N=5039) births. Gestational age 0 (the start of pregnancy) is defined as the first day of the last menstrual period. The red, purple, and green vertical bars delineate the mean pregnancy end and one SD for pregnancy loss, preterm birth, and term birth, respectively.

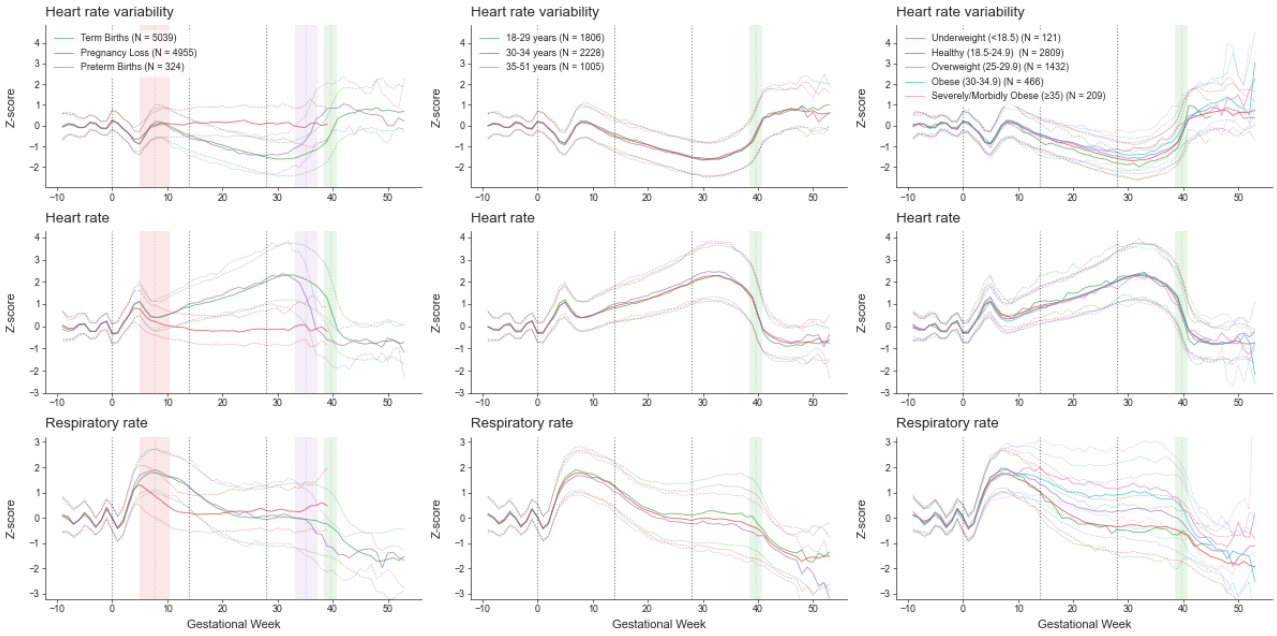
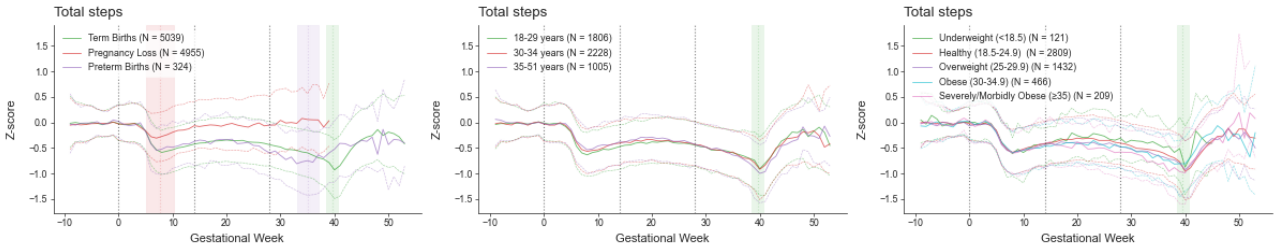


Figure 5. Changes in Oura activity data throughout pregnancy. Data (Z-scores - normalized data based on prepregnancy individuals' baseline) are plotted as the weekly median (thick line) and the 50th percentile range, separately for pregnancies ending in loss (N=4955), preterm births (N=324), and term (N=5039) births. Gestational age 0 (the start of pregnancy) is defined as the first day of the last menstrual period. The red, purple, and green vertical bars delineate the mean pregnancy end and one SD for pregnancy loss, preterm birth, and term birth, respectively.



While analyses focus on term births, the figures provide a visual comparison with pregnancies ending in loss and preterm birth. Pregnancies ending in loss (occurring between estimated gestational ages of 4 to 20 weeks) exhibit significant deviations from the trajectories of pregnancies resulting in birth at the same gestational age. In contrast, pregnancies resulting in preterm birth closely mirror Oura data trajectories of term births, except that the physiological changes associated with birth occur at an earlier gestational age.

Sleep

GEE results indicate a significant gestational age main effect for time in bed, time asleep, and time awake at night (all *P* values <.001). From the onset of pregnancy, time in bed, time asleep, and time awake all rise and reach a first peak around week 9. They then decrease and stabilize in the second trimester. In the third trimester, while time in bed returns to approximately pre-pregnancy level, time awake remains elevated and time asleep reduced. Around birth, sleep patterns change, with time in bed and time awake increasing to their peaks, while time asleep decreases to its lowest point, all occurring at week 40. Notably, participants' time spent awake at around birth exceeds 3 SDs over baseline. GEE results indicate statistically significant gestational age-by-maternal-age and gestational age-by BMI interaction for each sleep metric (all *P* values <.01). However, the effect sizes are relatively modest except for differences following birth. We do see a slightly increased time asleep in younger compared with older participants from week 9 up to birth, with respect to prepregnancy level.

Similarly, significant gestational age main effects were observed for sleep stages (all *P* values <.001). Deep and rapid-eye-movement sleep both decreased throughout pregnancy, reaching a nadir around birth before recovering postpartum. Light sleep increased during the first trimester, reaching a peak around week 9, remained stable for most of pregnancy, and then declined postpartum. While statistically significant gestational age-by-maternal-age and gestational age-by BMI interactions were also found for these sleep stages (all *P* values <.01), their effect sizes were relatively modest. A subtle increase in light sleep and a decrease in deep sleep were noted in younger participants compared with older participants from week 9 until birth, relative to prepregnancy levels. There was a slight increase in rapid-eye-movement sleep from around

XSL•FO
RenderX

week 33 until birth for younger participants compared with older participants.

Temperature Trends

GEE results indicate a significant gestational age main effect (*P* value <.001). From the start of pregnancy, temperature trends increase sharply at around conception (around 2-week gestational age), until it reaches a max peak around 9-week gestational age. It then gradually decreases to below prepregnancy levels. Around birth, temperature trends slightly increase, peaking at week 41, and afterward, it drops again below prepregnancy baseline levels. There are statistically significant but modest effects associated with age and BMI (all *P* values <.001). While the peak in temperature around week 9 is roughly the same across age groups, we do see a slightly less pronounced downward trend in temperature during the second and third trimesters for younger compared with older participants. The temperature peak around week 9 is slightly lower for participants with lower BMI compared with those with higher BMI, and there is a more significant drop in temperature during the second trimester for lower compared with higher BMI participants. This difference between lower and higher BMI participants persists through the third trimester.

Cardiorespiratory

GEE results indicate a significant gestational age main effect for HR and heart rate variability (HRV), as well as statistically significant interactions with age and BMI (all *P* values <.001). From the beginning of pregnancy, HR rises and reaches a first peak at week 5. After a small trough at week 9, it continues rising, peaking at week 32, then drops sharply at birth and stabilizes in the postpartum period. The HRV trajectory mirrors the opposite pattern of HR. Older participants exhibit a slightly larger rise in HR in the second and third trimesters (until approximately week 33), and have a slightly larger drop in HRV between the end of the first trimester and the first half of the second trimester (approximately from week 12 through week 24). Participants with lower BMI experienced a greater drop in HRV during the second trimester compared with those with higher BMI.

GEE results indicate significant gestational age main effect, gestational age by maternal age interaction, and gestational age by BMI interaction for RR (all *P* values <.001). From the onset of pregnancy, RR increases, peaking at 9 weeks of gestational age. It then steadily declines, stabilizing during the second and third trimesters. Following birth, RR experiences a sharp drop, falling below baseline levels. Interaction effects indicate that after reaching the first peak, RR decreases less in participants with a higher BMI and in younger participants.

Activity

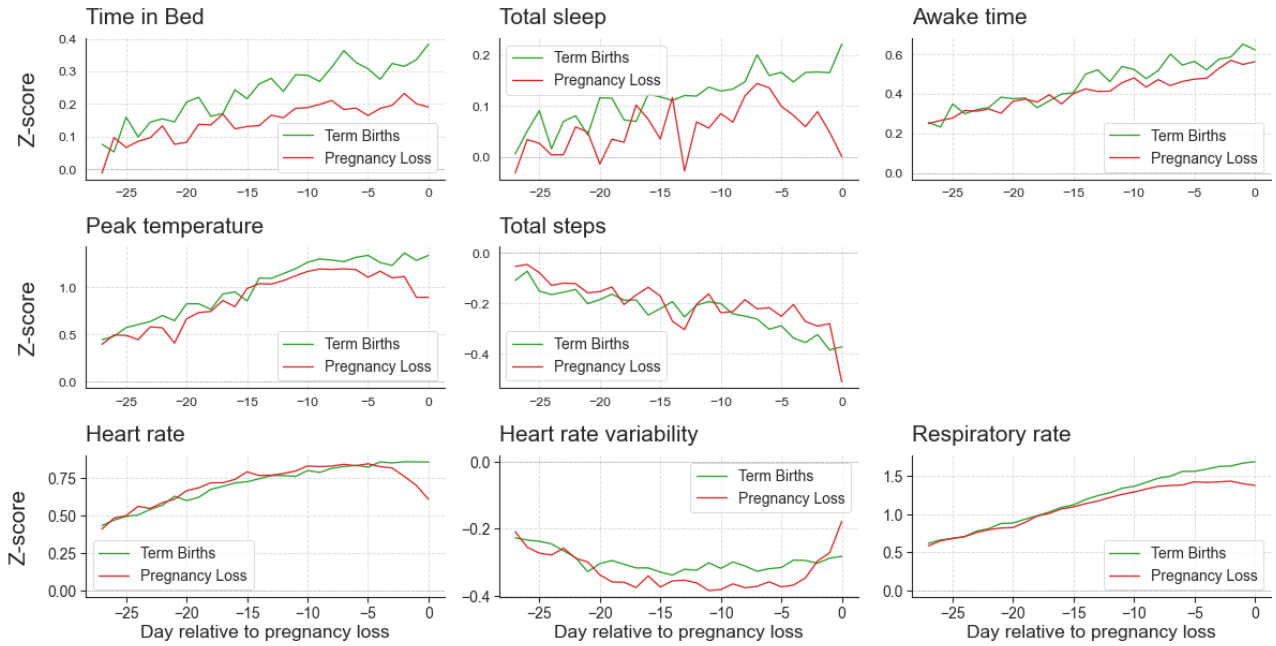
GEE results indicate significant gestational age effects for total numbers of steps (*P* values <.001). From the start of pregnancy, the number of steps decreases, reaching a trough at gestational week 8 and then increases slightly until 21 weeks of gestational age (middle of the second trimester) before decreasing again until birth, reaching the min level of activity. GEE results indicate significant interactions between gestational age and both maternal age and BMI for the total number of steps (all *P* values <.001). Though statistically significant, the effect size for the maternal age interaction is modest. There is a greater decrease in steps starting in the second trimester for participants with higher BMI compared with those with lower BMI. There is also a slightly larger drop in steps for younger participants starting at the trough at gestational week 8, with the difference between younger and older participants disappearing by the end of the second trimester.

Oura Biobehavioral Signatures Preceding Pregnancy Loss

Detailed results for GEE are provided in Multimedia Appendix 8. Changes in Oura bio-behavioral data for 28 days preceding and following the end of pregnancy are displayed in Figure 6. The number of women contributing valid data at each gestational week is shown in Multimedia Appendix 5.

XSL•FO
RenderX

Figure 6. Changes in Oura sleep, temperature trends, activity, and cardiorespiratory data across 28 days preceding the end of pregnancy, for pregnancies ending in loss (N=4955), compared with Oura data trajectories matching the same gestational age, for a randomly chosen pregnancy leading to birth (N=4955).



The GEE model, comparing the loss versus nonloss curves, shows that the time in bed starts decreasing about 2 weeks prior to pregnancy loss (*P* values <.001), while total sleep time starts to decrease one week before (*P* values <.001). Awake time drops around the same period (*P*=.033). Temperature trends also show a significant effect (*P* values <.001), with peak temperature trends decreasing a week before pregnancy loss. Average HR drops a few days prior to pregnancy loss (*P*=.005), and RR drops one week before loss (*P*=.019). Steps and HRV effects were not statistically significant in the GEE models.

Discussion

In this comprehensive, real-world, observational, retrospective study, we analyzed data from 10,318 pregnancies, providing the first large-scale examination of multimodal physiological and behavioral metrics recorded across the entire pregnancy continuum—from prepregnancy through conception and into the postpartum period. The novelty of the study lies in the high-density analysis of key biobehavioral aspects of pregnancy (sleep, temperature trends, cardiorespiratory, and activity data), mapping their trajectories over time and highlighting the complex, domain-specific temporal dynamics that characterize the multitude of biological and behavioral changes accompanying a woman's pregnancy journey.

Changes that women experience throughout pregnancy are largely driven by hormonal shifts and can be viewed as adaptive responses designed to support a successful pregnancy and birth while protecting the mother [38,39]. Our results indicate that during the first trimester, participants experienced significant sleep alteration, peaking at week 9 and being characterized by increased time in bed (≈ +30 min), asleep (≈ +15 min), and awake (≈ +15 min) compared with prepregnancy levels. These changes align with rising progesterone levels, which have both thermogenic and soporific effects and can promote sleepiness and increase total sleep time, but also contribute to sleep disruption. While sleep duration returned to baseline (or lower) in the second trimester, wake time remained elevated, likely due to physiological adaptation to pregnancy as well as known factors such as increased frequency of nocturnal urination [40]. In the third trimester, sleep disturbances worsened, possibly driven by increased physical discomfort, fetal movements, or additional increases in nocturnal urination. After birth, sleep deteriorated dramatically, with wake time rising by ≈90 min and time spent asleep dropping ≈70 min below prepregnancy levels, possibly reflecting infant care demands, postpartum hormonal shifts, and disrupted circadian rhythms [41].

Temperature changes were more pronounced, sustained, and occurred earlier than sleep changes, becoming evident by 4 weeks estimated gestational age (≈+0.3 °C above baseline), mapping early hormonal shifts, particularly rising progesterone, which is known to increase basal skin temperature via thermogenic effects. These changes persisted throughout the first trimester. In the second trimester, temperature trends declined as maternal adaptation mechanisms, such as increased blood volume and improved thermoregulation (eg, response to increased metabolic heat produced by the fetus) [38], took effect, stabilizing below baseline in the third trimester. At birth, a slight skin temperature rise (≈+0.1 °C) may reflect labor-associated metabolic demands (eg, increased physical activity with uterine contractions) and hormonal fluctuations, followed by a drop postpartum as pregnancy-related thermogenic effects subside [38]. The postpartum return to monthly cyclical temperature oscillations (ie, the resumption of menstrual cycles) is not observed within the analyzed time window, which is limited to a few months after birth for live births. Given that menstruation gradually resumes, varying among individuals and influenced by factors such as lactation practices and hormonal

XSL•FO
RenderX

fluctuations—none of which are captured in the current dataset—this process remains beyond the scope of our analysis.

Changes in cardiorespiratory data are consistent with the dynamic cardiovascular adaptations that occur throughout pregnancy to support the increasing metabolic demands of both the developing fetus and the mother [42]. Oura HR and HRV, that is, indices reflecting autonomic modulation, show profound changes across gestational age; they exhibit inverse patterns, characterized by a biphasic fluctuation in the first trimester—an initial HR peak exceeding 4 bpm above baseline and an HRV drop of >10 milliseconds at week 5, followed by a trough at week 9. These changes suggest early significant maternal physiologic preparatory changes to pregnancy, that is, happening before the end of the embryonic period, with HR rises following systemic vasodilation [43-45]. During the second trimester, HR increases linearly, peaking at >10 bpm above baseline, while HRV declines, reaching a minimum of >15 milliseconds below baseline, possibly reflecting decreases in parasympathetic cardiovascular control, as evaluated via standardized autonomic tests [46]. In the third trimester, these trends begin shifting earlier, with a marked drop in HR and a rise in HRV around birth, ultimately approaching prepregnancy levels.

The pulmonary system undergoes significant mechanical and functional changes during pregnancy. Notably, ventilation and tidal volume increase by approximately 20%-50% from prepregnancy levels [47]. Changes in RR are less well-defined in the literature; when reported, increases are typically modest, ranging from 1 to 2 breaths per minute. In our study, RR shows a slight increase during the first trimester, peaking around 8 weeks at approximately 1 breath per minute above baseline. This initial rise may reflect the stimulating effect of progesterone. Subsequently, the RR decreases and stabilizes by the mid-second trimester, returning to prepregnancy baseline levels, and remains unchanged until birth. After delivery, the RR may drop below baseline levels as the body readjusts to the nonpregnant state.

Our data indicate a decline in step count during the first trimester of pregnancy, with a notable dip around 8 weeks—a reduction of approximately 2000 steps compared with prepregnancy levels. This decline may be attributed to a combination of physiological, psychological, and social factors driven by hormonal fluctuations and early maternal adaptations. For example, elevated levels of human chorionic gonadotropin contribute to common pregnancy symptoms such as nausea, dizziness, and fatigue, which may lead to reduced physical activity [48]. Additionally, other factors such as worsening sleep quality and increased psychological stress may further contribute to decreased activity levels during this period. Cardiovascular adjustments can also impact energy levels. Following a slight increase in the second trimester, physical activity declines again, reaching its lowest point near birth, with step counts dropping by more than 2500 steps. This reduced activity in the third trimester may be explained by factors such as reduced mobility due to the mechanical strain, such as increased lordosis and laxity of joints, as well as increased effort on the cardiorespiratory system to sustain exercise under peak cardiometabolic demand as the maternal body supports rapid fetal growth, increased oxygen consumption, and metabolic adaptations to prepare for labor [42].

The use of passive, continuous longitudinal data collection in free-living conditions during pregnancy remains limited. However, our findings align with recent studies using consumer-grade devices. For example, 2 studies—one involving a limited sample of 17 (88%) US Hispanic women wearing Oura Ring devices [49] and another with a convenience sample of 20 Finnish women using Garmin Vivosmart bands [50]—revealed similar overall trajectories in sleep duration, activity levels, and HR increases through pregnancy. Focusing on photoplethysmography-based HR and a comprehensive set of time-domain and frequency-domain HRV metrics, Sarhaddi et al [51] analyzed data collected from 58 Finnish pregnant women, including those with high-risk pregnancies, using the Samsung Gear Sport smartwatch. Like our results, their findings reveal measure-specific changes in HR and HRV. Additional similar studies include pregnancy analyses based on Ava Fertility Tracker data in 23 women [52] and WHOOP 2.0 data in 18 healthy women [53]. These studies vary in sample characteristics, methodology, and purpose (eg, narrow focus on specific events during pregnancy), relying on small sample sizes and often lacking early pregnancy and prepregnancy baseline data, making direct outcome comparisons challenging. Additionally, many of these studies were conducted during the COVID-19 pandemic, which may have influenced pregnant women's behaviors throughout pregnancy.

Notably, recent analyses of wearable biobehavioral data in 2540 Fitbit users [19], primarily focused on sleep, and an analysis of 733 Apple Watch users [20] revealed similar temporal patterns across the health metrics examined. Likewise, Keeler et al [18] analyzed Oura-derived biobehavioral trajectories in 97 full-term pregnancies and 23 early pregnancy losses. Our analysis builds upon and extends these findings by offering significantly greater temporal resolution, analyzing a substantially larger and more diverse sample, and uniquely including pregnancies ending in loss or preterm birth. These contributions allow for a more comprehensive characterization of physiological changes across pregnancy and highlight early, modality-specific deviations associated with nonviable outcomes. Importantly, similar to others [18,20], our results did not show large variation in Oura biobehavioral trajectories throughout pregnancy as a function of factors such as maternal age and prepregnancy BMI status. This aspect needs to be further explored. Notably, our evaluation of BMI was limited to prepregnancy status, and it is likely that different weight trajectories during pregnancy—such as the rate and pattern of weight gain—may affect pregnancy biometrics.

In our observations, we detected changes in sleep and downward shifts in temperature, RR, and HR preceding the reported early end of pregnancy, with most signal deflections beginning approximately 7 days before the event. These biobehavioral changes suggest the potential utility of continuous monitoring for identifying early indicators of pregnancy failure. Specifically, we were unable to determine the underlying cause or classification of these outcomes (eg, spontaneous miscarriage, ectopic pregnancy, molar pregnancy) due to the lack of diagnostic or medical metadata. As such, while we observed biobehavioral deviations in pregnancies not resulting in live

XSL•FO
RenderX

birth, these findings should be interpreted as preliminary and hypothesis-generating. Future studies incorporating clinical validation and diagnostic classification are critical to elucidate whether specific biobehavioral patterns reliably differentiate among types of pregnancy loss, and to explore their potential for early detection of adverse outcomes.

Several limitations should be acknowledged. A major limitation of this work is the lack of detailed information on sample characteristics, pregnancy complications, and other biopsychosocial factors, which limits our ability to fully understand the complex influences on trajectories of sleep, temperature trends, cardiorespiratory function, and activity patterns throughout pregnancy. In addition, the observational design of the study does not allow for control of external variables or inference of causal relationships between biobehavioral changes and maternal outcomes. This is a retrospective analysis of data from Oura Ring users who engaged with an in-app feature specifically developed to support women during pregnancy. The sample likely overrepresents individuals with access to wearable technology and higher levels of health literacy, introducing demographic bias and limiting generalizability—particularly to underserved or marginalized populations. As such, the sample may not reflect the characteristics of the general population. For instance, while 71.7% of participants are based in the United States, only 13% meet the criteria for obesity—substantially lower than the 40.3% prevalence reported by the US Centers for Disease Control and Prevention among adults aged 20 years and older [54]. Furthermore, several key variables were self-reported, including prepregnancy BMI and the date of pregnancy end (eg, birth or loss), increasing the potential for recall bias and misclassification. The absence of clinical validation limits the applicability of these findings in medical settings and underscores the need for complementary studies that integrate health records and diagnostic assessments to enhance the translational utility of wearable-based monitoring.

This study also has inherent technological limitations. Wearable devices may be subject to measurement error and limited clinical validation—including during pregnancy, which involves profound physiological changes. Device accuracy may fluctuate across the pregnancy continuum, potentially masking some underlying physiological effects. These and other limitations of consumer wearable technology have been extensively discussed in recent literature [25,55,56]. In this study, in an attempt to control for potential confounders, we limited variability in device model and algorithm version by including only Oura Ring Gen3 users following the latest algorithm updates.

Future studies would also benefit from a multimodal data collection approach that combines passive wearable monitoring with ecological momentary assessments in more diverse populations. This would enable contextualization through variables such as reproductive history, lifestyle habits, occupational and relationship status, perceived support, health conditions, light exposure, physical activity not captured by the wearable (eg, activities performed when the wearable is not worn or not detected by the device), medication use, weight changes, and the presence or treatment of sleep disorders (eg, sleep-disordered breathing, restless legs syndrome). At the same time, this will be essential to improve generalizability and ensure that findings are applicable to racially, socioeconomically, and geographically diverse populations. Such enriched datasets would allow for nuanced characterization of the individual pregnancy experience and support the development of predictive models for both adverse pregnancy outcomes and natural events such as labor onset or early indicators of abnormalities—factors known to influence maternal and infant morbidity and mortality.

While outside the scope of this study, emerging evidence also highlights the potential relevance of circadian biology in shaping pregnancy outcomes. The reproductive system is tightly coupled to circadian regulation, and disruptions in maternal circadian timing—such as those caused by irregular light-dark cycles or shift work—are associated with negative pregnancy outcomes [57,58]. The placenta exhibits circadian gene expression, and disturbances in maternal circadian alignment may interfere with fetal-maternal signaling, nutrient transfer, and developmental timing [59]. There is evidence of a relationship between disturbed sleep during and after pregnancy and negative pregnancy outcomes, such as depression [60], anxiety [61], and gestational diabetes [62]. Going forward, incorporating circadian-focused assessments using wearable devices may provide novel insight into modifiable risk factors affecting both maternal well-being and fetal development. Although this study did not examine circadian-related metrics, wearable-derived data streams—such as nighttime temperature rhythms, intraindividual variability in sleep timing, and rest-activity rhythm—may offer promising noninvasive proxies for circadian alignment [25]. Future work should explore whether these features can serve as digital biomarkers of circadian misalignment and help identify individuals at higher risk for adverse pregnancy outcomes.

To conclude, this comprehensive exploration provides valuable insights into the physiological trajectories of pregnancy within a real-world population—information that until now, has not been widely available. By integrating passive sensing data with clinical information on reproductive history, pregnancy progression, and outcomes, this approach has the potential to advance our understanding of maternal health, enabling more personalized monitoring and improved pregnancy care. Beyond the continuous, automatic collection of relevant health metrics, this method allows for the examination of patterns relative to an individual's baseline, rather than solely determining whether a single data point falls within a population range, irrespective of personal variation.

XSL•FO
RenderX

Acknowledgments

The authors used the generative AI tool ChatGPT (OpenAI) to support language editing of the manuscript. We are grateful to Chris Curry, MD, PhD, and Shelby Bachman, PhD, for their valuable clinical insights, broader contributions, and support during the review of the manuscript. This study was funded by Oura Health Oy.

Data Availability

The dataset analyzed during this study is not publicly available due to privacy protections for participants, but may be made available, in deidentified form, at Oura's sole discretion.

Authors' Contributions

RA, MdZ, NT, KP, AC, and NG contributed to the conceptualization of the study. RA, MdZ, NT, and KP contributed to the methodology. RA and KP contributed to data curation and formal analysis, with KP also responsible for validation. KP and MdZ contributed to the visualization. MdZ, RA, and KP drafted the original manuscript. RA, NT, AC, NG, KP, and MdZ reviewed and edited the manuscript. MdZ provided supervision.

Conflicts of Interest

AC, NG, KP, and MdZ are employees of Oura Health Oy or any of its affiliates. RA and NT were employees at Oura Health Oy at the time of their contribution to this study, which predates their transition to Apple Inc (RA) and Bayesian Health (NT). While the authors have a professional affiliation with these companies, they have conducted this research with the highest standards of scientific integrity and transparency. Oura has a portfolio of patents and patent applications [63] that includes coverage of women's reproductive health, which may be broadly relevant to the topic of this study. The authors confirm that this intellectual property interest has not influenced the study design, data collection, analysis, or interpretation of the results. MdZ has received research funding unrelated to this work from Noctrix Health Inc and Verily Life Science LLC.

Multimedia Appendix 1
STROBE checklist - cohort studies.
[DOCX File , 27 KB - mhealth_v13i1e80213_app1.docx]

Multimedia Appendix 2
Distribution of volume of pregnancies as a function of gestational age at which pregnancies end, for pregnancies ending in loss, preterm births and term births.
[PNG File , 21 KB - mhealth_v13i1e80213_app2.png]

Multimedia Appendix 3
Distribution of age and body mass index (BMI) of participants with pregnancies ending in loss, preterm births and term births.
[PNG File , 129 KB - mhealth_v13i1e80213_app3.png]

Multimedia Appendix 4
Absolute changes in Oura data (sleep, temperature trends, cardiorespiratory, and activity) throughout pregnancy. Data are plotted as the weekly median (thick line) and the 50th percentile range, separately for pregnancies ending in loss (N = 4955), preterm births (N = 324), and term (N = 5039) births. Gestational age 0 (the start of pregnancy) is defined as the first day of the last menstrual period. The red, purple and green vertical bars delineate the mean pregnancy end ± one standard deviation for pregnancy loss, preterm birth, and term birth, respectively.
[PNG File , 345 KB - mhealth_v13i1e80213_app4.png]

Multimedia Appendix 5
The number of women contributing valid data at each gestational week for pregnancies ending in loss, preterm births and term births.
[PNG File , 22 KB - mhealth_v13i1e80213_app5.png]

Multimedia Appendix 6
Detailed results for Generalized Estimating Equation (GEE) model for analyses evaluating pattern of changes in Oura bio-behavioral data throughout pregnancy, in pregnancies leading to term births. REM, Rapid-Eyes-Movement-Sleep. Maximum effect sizes are summarized per trimester; however, the statistical model was fit over the entire pregnancy duration.
[DOCX File , 21 KB - mhealth_v13i1e80213_app6.docx]

XSL•FO
RenderX

Multimedia Appendix 7
Detailed results for Generalized Estimating Equation (GEE) model for analyses evaluating effect of Body Mass Index (BMI) and maternal age on Oura bio-behavioral data throughout pregnancy, in pregnancies leading to term births. Maximum effect sizes are summarized per trimester; however, the statistical model was fit over the entire pregnancy duration.
[DOCX File , 25 KB - mhealth_v13i1e80213_app7.docx]

Multimedia Appendix 8
Detailed results for Generalized Estimating Equation (GEE) model for analyses evaluating pattern of changes in Oura bio-behavioral data for the 28 days preceding pregnancies ending in loss. The effect size is calculated over the 28 and 7 days preceding the pregnancy ending before 20 weeks and represents the average difference with respect to corresponding biometrics over the same period from randomly selected full-term pregnancies.
[DOCX File , 19 KB - mhealth_v13i1e80213_app8.docx]

References

1. Osterman M, Hamilton B, Martin J, Driscoll A, Valenzuela C. Births: Final Data for 2022. Hyattsville, MD: National Center for Health Statistics; 2024.
2. Trends in maternal mortality 2000 to 2020: estimates by WHO, UNICEF, UNFPA, world bank group and UNDESA/population division. In: World Health Organization. Geneva: World Health Organization; 2020.
3. Trost S, Beauregard J, Chandra G, Njie F, Berry J, Harvey A. Pregnancy-Related Deaths: Data from Maternal Mortality Review Committees in 36 US States, 2017-2019. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services; 2022.
4. Declercq E, Zephyrin L. Severe maternal morbidity in the United States: a primer. The Commonweealth Fund. 2021 Oct 28. URL: https://www.commonwealthfund.org/sites/default/files/2021-10/Declercq_severe_maternal_morbidity_in_US_primer_db.pdf [accessed 2025-06-28]
5. Kriebs JM. Guidelines for perinatal care, sixth edition: by the American Academy of Pediatrics and the American College of Obstetricians and Gynecologists. J Midwife Womens Health 2010 Dec 24;55(2):e37-e37. [doi: 10.1016/j.jmwh.2009.12.013]
6. Maugeri A, Barchitta M, Agodi A. How wearable sensors can support the research on foetal and pregnancy outcomes: a scoping review. J Pers Med 2023;13(2):218 [FREE Full text] [doi: 10.3390/jpm13020218] [Medline: 36836452]
7. Alim A, Imtiaz MH. Wearable sensors for the monitoring of maternal health-a systematic review. Sensors (Basel) 2023;23(5):2411. [doi: 10.3390/s23052411] [Medline: 36904615]
8. Liu L, Pu Y, Fan J, Yan Y, Liu W, Luo K, et al. Wearable sensors, data processing, and artificial intelligence in pregnancy monitoring: a review. Sensors (Basel) 2024;24(19):6426. [doi: 10.3390/s24196426] [Medline: 39409471]
9. Thigpen N, Patel S, Zhang X. Oura ring as a tool for ovulation detection: validation analysis. J Med Internet Res 2025;27:e60667. [doi: 10.2196/60667] [Medline: 39889300]
10. Grant A, Smarr B. Feasibility of continuous distal body temperature for passive, early pregnancy detection. PLOS Digit Health 2022;1(5):e0000034. [doi: 10.1371/journal.pdig.0000034] [Medline: 36812529]
11. Erickson EN, Gotlieb N, Pereira LM, Myatt L, Mosquera-Lopez C, Jacobs PG. Predicting labor onset relative to the estimated date of delivery using smart ring physiological data. NPJ Digit Med 2023;6(1):153. [doi: 10.1038/s41746-023-00902-y] [Medline: 37598232]
12. Basavaraj C, Grant AD, Aras SG, Erickson EN. Deep learning model using continuous skin temperature data predicts labor onset. BMC Pregnancy Childbirth 2024;24(1):777. [doi: 10.1186/s12884-024-06862-9] [Medline: 39587525]
13. Bester M, Almario Escorcia MJ, Fonseca P, Mollura M, van Gilst MM, Barbieri R, et al. The impact of healthy pregnancy on features of heart rate variability and pulse wave morphology derived from wrist-worn photoplethysmography. Sci Rep 2023;13(1):21100 [FREE Full text] [doi: 10.1038/s41598-023-47980-2] [Medline: 38036597]
14. Kytö M, Koivusalo S, Tuomonen H, Strömberg L, Ruonala A, Marttinen P, et al. Supporting the management of gestational diabetes mellitus with comprehensive self-tracking: mixed methods study of wearable sensors. JMIR Diabetes 2023;8:e43979. [doi: 10.2196/43979] [Medline: 37906216]
15. Penders J, Altini M, Van Hoof C, Dy E. Wearable sensors for healthier pregnancies. Proc IEEE 2015;103(2):179-191. [doi: 10.1109/jproc.2014.2387017]
16. Feli M, Azimi I, Sarhaddi F, Sharifi-Heris Z, Niela-Vilen H, Liljeberg P, et al. Preterm birth risk stratification through longitudinal heart rate and HRV monitoring in daily life. Sci Rep 2024;14(1):19896 [FREE Full text] [doi: 10.1038/s41598-024-70773-0] [Medline: 39191907]
17. Hurwitz E, Butzin-Dozier Z, Master H, O'Neil ST, Walden A, Holko M, et al. Harnessing consumer wearable digital biomarkers for individualized recognition of postpartum depression using the all of us research program data set: cross-sectional study. JMIR Mhealth Uhealth 2024;12:e54622 [FREE Full text] [doi: 10.2196/54622] [Medline: 38696234]
18. Keeler Bruce L, González D, Dasgupta S, Smarr BL. Biometrics of complete human pregnancy recorded by wearable devices. NPJ Digit Med 2024;7(1):207 [FREE Full text] [doi: 10.1038/s41746-024-01183-9] [Medline: 39134787]

XSL•FO
RenderX

19. Young-Lin N, Heneghan C, Liu Y, Schneider L, Niehaus L, Haney A, et al. Insights into maternal sleep: a large-scale longitudinal analysis of real-world wearable device data before, during, and after pregnancy. EBioMedicine 2025;114:105640. [doi: 10.1016/j.ebiom.2025.105640] [Medline: 40147048]
20. Mishra A, Park J, Shapiro I, Fisher-Colbrie T, Baird DD, Suharwardy S, et al. Trends in sensor-based health metrics during and after pregnancy: descriptive data from the apple women's health study. AJOG Glob Rep 2024;4(4):100388 [FREE Full text] [doi: 10.1016/j.xagr.2024.100388] [Medline: 39296604]
21. Lean SC, Derricott H, Jones RL, Heazell AEP. Advanced maternal age and adverse pregnancy outcomes: a systematic review and meta-analysis. PLoS One 2017;12(10):e0186287. [doi: 10.1371/journal.pone.0186287] [Medline: 29040334]
22. Vats H, Saxena R, Sachdeva MP, Walia GK, Gupta V. Impact of maternal pre-pregnancy body mass index on maternal, fetal and neonatal adverse outcomes in the worldwide populations: a systematic review and meta-analysis. Obes Res Clin Pract 2021;15(6):536-545. [doi: 10.1016/j.orcp.2021.10.005] [Medline: 34782256]
23. Oura Health Oy. URA Terms of Use. 2024. URL: https://ouraring.com/terms-and-conditions [accessed 2024-04-02]
24. URA health privacy policy. Oura Health Oy. 2025. URL: https://ouraring.com/privacy-policy [accessed 2025-04-21]
25. de Zambotti M, Goldstein C, Cook J, Menghini L, Altini M, Cheng P, et al. State of the science and recommendations for using wearable technology in sleep and circadian research. Sleep 2024;47(4):zsad325. [doi: 10.1093/sleep/zsad325] [Medline: 38149978]
26. Altini M, Kinnunen H. The promise of sleep: a multi-sensor approach for accurate sleep stage detection using the oura ring. Sensors (Basel) 2021;21(13):4302 [FREE Full text] [doi: 10.3390/s21134302] [Medline: 34201861]
27. Robbins R, Weaver MD, Sullivan JP, Quan SF, Gilmore K, Shaw S, et al. Accuracy of three commercial wearable devices for sleep tracking in healthy adults. Sensors (Basel) 2024;24(20):6532. [doi: 10.3390/s24206532] [Medline: 39460013]
28. Chee NI, Ghorbani S, Golkashani HA, Leong RL, Ong JL, Chee MW. Multi-night validation of a sleep tracking ring in adolescents compared with a research actigraph and polysomnography. Nat Sci Sleep 2021;13:177-190. [doi: 10.2147/nss.s286070]
29. Svensson T, Madhawa K, Nt H, Chung U, Svensson AK. Validity and reliability of the oura ring generation 3 (Gen3) with oura sleep staging algorithm 2.0 (OSSA 2.0) when compared to multi-night ambulatory polysomnography: a validation study of 96 participants and 421,045 epochs. Sleep Med 2024;115:251-263 [FREE Full text] [doi: 10.1016/j.sleep.2024.01.020] [Medline: 38382312]
30. de Zambotti M, Rosas L, Colrain IM, Baker FC. The sleep of the ring: comparison of the ŌURA sleep tracker against polysomnography. Behav Sleep Med 2019;17(2):124-136. [doi: 10.1080/15402002.2017.1300587] [Medline: 28323455]
31. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors (Basel) 2022;22(16):6317 [FREE Full text] [doi: 10.3390/s22166317] [Medline: 36016077]
32. Kryder C. How accurate is Oura's temperature data? Oura. 2020. URL: https://ouraring.com/blog/temperature-validated-accurate/?srsltid=AfmBOoqBQ5mgDx0n9eLlyT-H2Gf9eeIEMwtEVBB3zShQkePEjIimVsj- [accessed 2025-09-27]
33. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
34. Cao R, Azimi I, Sarhaddi F, Niela-Vilen H, Axelin A, Liljeberg P, et al. Accuracy assessment of oura ring nocturnal heart rate and heart rate variability in comparison with electrocardiography in time and frequency domains: comprehensive analysis. J Med Internet Res 2022;24(1):e27487. [doi: 10.2196/27487] [Medline: 35040799]
35. Kristiansson E, Fridolfsson J, Arvidsson D, Holmäng A, Börjesson M, Andersson-Hall U. Validation of Oura ring energy expenditure and steps in laboratory and free-living. BMC Med Res Methodol 2023;23(1):50 [FREE Full text] [doi: 10.1186/s12874-023-01868-x] [Medline: 36829120]
36. LIANG KY, ZEGER SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
37. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. SciPy 2010;7(1):92-96. [doi: 10.25080/majora-92bf1922-011]
38. Georgescu T. The role of maternal hormones in regulating autonomic functions during pregnancy. J Neuroendocrinol 2023;35(12):e13348. [doi: 10.1111/jne.13348] [Medline: 37936545]
39. Smiley KO, Ladyman SR, Gustafson P, Grattan DR, Brown RS. Neuroendocrinology and adaptive physiology of maternal care. Curr Top Behav Neurosci 2019;43:161-210. [doi: 10.1007/7854_2019_122] [Medline: 31808002]
40. Khosla L, Huang AJ, Kasarla N, Monaghan TF, Weiss JP, Kabarriti AE. Association between pregnancy and nocturia: a national health and nutrition examination survey analysis. Neurourol Urodyn 2022;41(6):1505-1510. [doi: 10.1002/nau.24998] [Medline: 35731180]
41. Balserak BI, O'Brien LM, Bei B. Sleep and sleep disorders associated with pregnancy. In: Principles and Practice of Sleep Medicine: Sleep in Women. Netherlands: Elsevier; 2022:1751-1763.
42. Sanghavi M, Rutherford JD. Cardiovascular physiology of pregnancy. Circulation 2014;130(12):1003-1008. [doi: 10.1161/circulationaha.114.009029]
43. Clapp JF, Seaward BL, Sleamaker RH, Hiser J. Maternal physiologic adaptations to early human pregnancy. Am J Obstet Gynecol 1988;159(6):1456-1460. [doi: 10.1016/0002-9378(88)90574-1] [Medline: 3207124]

XSL•FO
RenderX

44. Duvekot JJ, Cheriex EC, Pieters FA, Menheere PP, Peeters LL. Early pregnancy changes in hemodynamics and volume homeostasis are consecutive adjustments triggered by a primary fall in systemic vascular tone. Am J Obstet Gynecol 1993;169(6):1382-1392. [doi: 10.1016/0002-9378(93)90405-8] [Medline: 8267033]
45. Mulder EG, de Haas S, Mohseni Z, Schartmann N, Abo Hasson F, Alsadah F, et al. Cardiac output and peripheral vascular resistance during normotensive and hypertensive pregnancy - a systematic review and meta-analysis. BJOG 2022;129(5):696-707. [doi: 10.1111/1471-0528.16678] [Medline: 33630403]
46. Ekholm EM, Piha SJ, Antila KJ, Erkkola RU. Cardiovascular autonomic reflexes in mid-pregnancy. Br J Obstet Gynaecol 1993;100(2):177-182. [doi: 10.1111/j.1471-0528.1993.tb15217.x] [Medline: 8476812]
47. Hegewald MJ, Crapo RO. Respiratory physiology in pregnancy. Clin Chest Med 2011;32(1):1-13. [doi: 10.1016/j.ccm.2010.11.001] [Medline: 21277444]
48. Furneaux EC, Langley-Evans AJ, Langley-Evans SC. Nausea and vomiting of pregnancy: endocrine basis and contribution to pregnancy outcome. Obstet Gynecol Surv 2001;56(12):775-782. [doi: 10.1097/00006254-200112000-00004] [Medline: 11753180]
49. Guo Y, Xu Q, Dutt N, Kehoe P, Qu A. Longitudinal changes in objective sleep parameters during pregnancy. Womens Health (Lond) 2023;19:17455057231190952 [FREE Full text] [doi: 10.1177/17455057231190952] [Medline: 37650368]
50. Saarikko J, Niela-Vilen H, Ekholm E, Hamari L, Azimi I, Liljeberg P, et al. Continuous 7-month internet of things-based monitoring of health parameters of pregnant and postpartum women: prospective observational feasibility study. JMIR Form Res 2020;4(7):e12417 [FREE Full text] [doi: 10.2196/12417] [Medline: 32706696]
51. Sarhaddi F, Azimi I, Axelin A, Niela-Vilen H, Liljeberg P, Rahmani AM. Trends in heart rate and heart rate variability during pregnancy and the 3-month postpartum period: continuous monitoring in a free-living context. JMIR Mhealth Uhealth 2022;10(6):e33458 [FREE Full text] [doi: 10.2196/33458] [Medline: 35657667]
52. Bossung V, Singer A, Ratz T, Rothenbühler M, Leeners B, Kimmich N. Changes in heart rate, heart rate variability, breathing rate, and skin temperature throughout pregnancy and the impact of emotions-a longitudinal evaluation using a sensor bracelet. Sensors (Basel) 2023;23(14):6620. [doi: 10.3390/s23146620] [Medline: 37514915]
53. Rowan SP, Lilly CL, Claydon EA, Wallace J, Merryman K. Monitoring one heart to help two: heart rate variability and resting heart rate using wearable technology in active women across the perinatal period. BMC Pregnancy Childbirth 2022;22(1):887 [FREE Full text] [doi: 10.1186/s12884-022-05183-z] [Medline: 36451120]
54. Emmerich SD, Fryar CD, Stierman B, Ogden CL. Obesity and severe obesity prevalence in adults: United States, august 2021-august 2023. NCHS Data Brief 2024(508):2024. [doi: 10.15620/cdc/159281] [Medline: 39808758]
55. de Zambotti M, Cellini N, Menghini L, Sarlo M, Baker FC. Sensors capabilities, performance, and use of consumer sleep technology. Sleep Med Clin 2020;15(1):1-30 [FREE Full text] [doi: 10.1016/j.jsmc.2019.11.003] [Medline: 32005346]
56. Chee MW, Baumert M, Scott H, Cellini N, Goldstein C, Baron K, World Sleep Society Sleep Tracker Task Force. World sleep society recommendations for the use of wearable consumer health trackers that monitor sleep. Sleep Med 2025;131:106506. [doi: 10.1016/j.sleep.2025.106506] [Medline: 40300398]
57. Zhu JL, Hjollund NH, Andersen AN, Olsen J. Shift work, job stress, and late fetal loss: the national birth cohort in Denmark. J Occup Environ Med 2004;46(11):1144-1149. [doi: 10.1097/01.jom.0000145168.21614.21] [Medline: 15534501]
58. Xu X, Ding M, Li B, Christiani DC. Association of rotating shiftwork with preterm births and low birth weight among never smoking women textile workers in China. Occup Environ Med 1994;51(7):470-474 [FREE Full text] [doi: 10.1136/oem.51.7.470] [Medline: 8044246]
59. Gotlieb N, Moeller J, Kriegsfeld LJ. In: Wray S, Blackshaw S, editors. Development and Modulation of Female Reproductive Function by Circadian Signals. Cham: Springer Developmental Neuroendocrinology; 2020:413-446.
60. Maghami M, Shariatpanahi SP, Habibi D, Heidari-Beni M, Badihian N, Hosseini M, et al. Sleep disorders during pregnancy and postpartum depression: a systematic review and meta-analysis. Int J Dev Neurosci 2021;81(6):469-478. [doi: 10.1002/jdn.10118] [Medline: 33942364]
61. Rubio E, Levey EJ, Rondon MB, Friedman L, Sanchez SE, Williams MA, et al. Poor sleep quality and obstructive sleep apnea are associated with maternal mood, and anxiety disorders in pregnancy. Matern Child Health J 2022;26(7):1540-1548 [FREE Full text] [doi: 10.1007/s10995-022-03449-8] [Medline: 35596848]
62. Eleftheriou D, Athanasiadou KI, Sifnaios E, Vagiakis E, Katsaounou P, Psaltopoulou T, et al. Sleep disorders during pregnancy: an underestimated risk factor for gestational diabetes mellitus. Endocrine 2024;83(1):41-50. [doi: 10.1007/s12020-023-03537-x] [Medline: 37740834]
63. Oura intellectual property notice. Oura Health Oy. 2025. URL: https://ouraring.com/intellectual-property-notice?srsltid=AfmBOopcokH5MSk_w2u7L_xMn_Th7dObpWm_8QHESgJxrngvfxpwo71T [accessed 2025-06-08]

Abbreviations

GEE: generalized estimating equation
HR: heart rate
HRV: heart rate variability

XSL•FO
RenderX

LMP: last menstrual period
RR: respiratory rate
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 08.07.25; peer-reviewed by AM Babker, C Leite da Silva; comments to author 24.07.25; revised version received 07.08.25; accepted 15.09.25; published 27.10.25.

Please cite as:
Adaimi R, Thigpen N, Clausel A, Gotlieb N, Patel K, de Zambotti M
Temporal Trajectories in Sleep, Temperature Trends, Cardiorespiratory, and Activity Metrics Measured via Oura Ring During Pregnancy: Large-Scale Observational Analysis
JMIR Mhealth Uhealth 2025;13:e80213
URL: https://mhealth.jmir.org/2025/1/e80213
doi:10.2196/80213
PMID:40996267

©Rebecca Adaimi, Nina Thigpen, Alicia Clausel, Neta Gotlieb, Ketan Patel, Massimiliano de Zambotti. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Multimodal Sleep Measurement and Alignment Analysis in Outpatients With Major Depressive Episode: Observational Study

Afrooz Mahir[1], BSc; Nguyen Luong[1], MSc; Ilya Baryshnikov[2,3], MD, PhD; Annasofia Martikkala[2,3], MD; Erkki Isometsä[2,3], MD, PhD; Talayeh Aledavood[1], PhD

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Department of Psychiatry, University of Helsinki, Helsinki, Finland

[3]Helsinki and Uusimaa Hospital District, Helsinki, Finland

Corresponding Author:
Afrooz Mahir, BSc
Department of Computer Science
Aalto University
Tietotekniikantalo, Konemiehentie 2,
Espoo, 02150
Finland
Phone: 358 403739808
Email: fathima.abdulmahir@aalto.fi

Abstract

Background: Sleep is essential for overall health and plays a critical role in the diagnosis of psychiatric disorders. Although polysomnography remains the gold standard for measuring sleep, its reliance on laboratory settings limits its feasibility for long-term, naturalistic monitoring, particularly for patients with mental disorders.

Objective: This study assesses sleep-tracking reliability and alignment in healthy individuals and patients with mood disorders using wearables, nearables, and ecological momentary assessment, while examining measurement biases and the impact of seasonal and demographic factors on discrepancies across methods.

Methods: We conducted a 14-day study in Finland and enrolled a total of 201 participants, comprising patients with a major depressive episode and healthy controls. Of these, 169 participants with sufficient observations were retained for further analyses. Participants' sleep patterns (onset, offset, and total sleep time [TST]) were gathered daily from an actigraph (Actiwatch 2), a bed sensor (Murata SCA11H), mobile screen events, and a daily survey. The alignment between sleep measurement methods was evaluated using Bland-Altman plots and Pearson correlation. Linear mixed models were used to assess the effects of demographics, season, and disorder type on the sleep measures alignment.

Results: Patients exhibited greater variability in sleep measures than healthy controls. For sleep onset, mean biases between devices were small and not statistically significant in either group, with moderate to strong correlations. In contrast, sleep offset showed significantly larger biases in patients: actigraph versus bed (+34.9 minutes; P=.01), smartphone versus bed (–45.3 minutes; P=.004), and actigraph versus smartphone (+78.7 minutes; P<.001), while controls exhibited minimal and nonsignificant differences. For TST, smartphone underestimates sleep compared to both bed sensors (–0.71 minutes; P<.001) and actigraphs (–1.35 minutes; P<.001). Across devices, TST correlations remained low, spanning r=0.12 (P=.58) to r=0.55 (P=.10) in controls and r=0.17 (P=.19) to r=0.43 (P=.002) in patients. Mixed models showed that older age was linked to better sleep offset alignment between actigraphy and bed sensors (β=–0.02, 95% CI –0.04 to 0.00; P=.048), as well as smartphone and bed sensor (β=–0.03, 95% CI –0.06 to 0.00; P=.03). Patients with bipolar/borderline personality disorder showed lower TST alignment, and alignment between smartphone and bed sensor was worse in females (β=−1.03, 95% CI −1.74 to −0.33, P=.004). Longer daylight duration was also associated with improved alignment in sleep offset and TST.

Conclusions: This study demonstrates the feasibility of using actigraphy, smartphone data, and bed sensors for sleep tracking in naturalistic settings with patients. It highlights measurement biases across devices, the impact of seasonal variations on sleep research in unique geographical regions like Finland, and key demographic factors influencing sleep measurement discrepancies.

(JMIR Mhealth Uhealth 2025;13:e82465) doi:10.2196/82465

XSL•FO
RenderX

KEYWORDS

actigraphy; bed sensors; depression; ecological momentary assessment; EMA; sleep monitoring; sleep quantification; smartphone data

Introduction

Sleep is a fundamental aspect of mental and physical well-being, playing a crucial role in emotional regulation, cognitive function, and overall health [1-3]. Insufficient and excessive sleep are both linked to cognitive impairment, heightened stress, and increased risk of chronic diseases [4-6]. In psychiatric populations, sleep abnormalities are especially pronounced, with insomnia and hypersomnia being core symptoms of major depressive disorder (MDD) [7], and hypersomnia is frequently observed in bipolar disorder (BD) [8]. Moreover, deviations from typical sleep are associated with elevated suicide risk, mood instability, and early relapse indicators in many psychiatric disorders, including mood disorders [9-11]. Given these profound implications, accurately capturing sleep patterns is essential for psychiatric assessments. However, sleep is typically measured in laboratory settings, which limits its applicability for real-world monitoring. This limitation makes obtaining reliable long-term measurements challenging.

Polysomnography (PSG) is the gold standard for sleep assessment but is typically limited to controlled environments [12-14]. Alternative methods, such as sleep diaries and questionnaires, offer greater flexibility but are prone to recall bias [15], often leading to discrepancies between subjective reports and objective measurements [16,17]. Wearable technology, such as actigraphy, provides a more practical solution with reasonable accuracy in tracking some sleep parameters, such as onset and offset [18]. However, actigraphy has limitations, including potential discomfort and the inability to measure physiological markers like heart rate and respiration [19]. More recent advancements in sleep monitoring have introduced contactless methods, such as ballistocardiography-based bed sensors and smartphone-based sleep tracking [20]. While these newer modalities show promise, their validity and comparability, especially in assessing patients with depressive disorders in real-world settings, remain underexplored. To address this gap, we use a multimodal approach, combining actigraphy, bed sensors, smartphone-based tracking, and the ecological momentary assessment (EMA) [21] to assess sleep.

A crucial issue in sleep research is the discrepancy between different measurement modalities. Studies comparing PSG with actigraphy and self-reports have consistently shown variation in onset and offset measurements [13,14]. The divergence is particularly pronounced in psychiatric populations, where sleep misperception—the tendency to perceive one's sleep inaccurately—is common [15,22]. While actigraphy has been frequently validated against PSG, the role of newer modalities like bed sensors and smartphone data requires further investigation [23].

Beyond measurement discrepancies, demographic, clinical, and environmental factors introduce further variance. For example, chronotype, gender, and age, as well as external variables such as seasons, are associated with sleep patterns [24,25], with women generally reporting longer sleep [26,27]. Using smartphones at bedtime, a behavior more common in younger participants, can decrease agreement between smartphone and actigraphy estimates because actigraphy may register sleep while the smartphone detects ongoing wakefulness [28]. Psychological conditions such as depression further complicate measurement accuracy due to circadian rhythm disruptions [29]. Seasonal fluctuations in daylight and temperature are consistently linked to variation in sleep timing and duration [30], consequently influencing cross-modal alignment. Evidence for these seasonal influences appears across modalities: nearable sensors [31], wearables [24,25,32-34], and smartphone [35,36].

Given these complexities, this study investigates how sleep onset, offset, and total sleep time (TST) measurements compare across actigraphy, smartphone data, bed sensors, and EMA. Additionally, it examines whether these measurements differ between healthy controls and patients with depression, providing insights into the extent of alignment in sleep assessments. Finally, this research explores the demographic, psychological, and environmental factors contributing to discrepancies between measurement modalities, highlighting the sources of error and bias. By addressing these questions, this study contributes to the area of naturalistic sleep research, offering a deeper understanding of sleep monitoring technologies and their implications for psychiatric evaluation, monitoring, and intervention.

Methods

Ethical Considerations

The study was approved by the Helsinki and Uusimaa Hospital District Ethics Committee and by Helsinki and Uusimaa Hospital District Psychiatry (approval number §125/2018). Data were anonymized, and participants received four movie tickets as compensation. Participation was voluntary, and individuals were informed that they could withdraw without consequence. All participants received clear information about the study and provided written informed consent.

Data Collection

The MoMo-Mood study [37] explored the effectiveness of wearable technology in monitoring sleep and mood in individuals with psychiatric disorders as well as finding behavioral markers of depression from passively collected data [37,38]. The study included 164 participants, of which 133 were diagnosed with a current major depressive episode, including 85 patients with MDD, 27 with borderline personality disorder (BPD), and 21 with BD, alongside 31 healthy controls. Additionally, a pilot study was conducted with 37 participants, including 14 patients with MDD and 23 healthy controls. The main and the pilot studies had a similar design and data from both studies were included in the analysis.

XSL•FO
RenderX

Participants were recruited voluntarily from mood disorder outpatient services at Helsinki University Hospital Mood Disorder Division, Turku University Central Hospital Department of Psychiatry, and City of Espoo Mental Health Services, and healthy controls were recruited from the general community. Eligibility required a confirmed diagnosis of MDD, BD, or BPD, and individuals with psychotic symptoms, current substance use disorder, or acute suicide risk were excluded. Healthy controls were screened to ensure the absence of clinical depression. In the pilot study, patients were asked to participate for 6 months and healthy controls for 1 year, while in the MoMo-Mood study participants were invited to take part for up to 1 year. Participation length varied, resulting in differences in the amount of data contributed. Enrollment occurred on an ongoing basis, with participants joining and leaving at different times. Participation was voluntary, and individuals were informed that they could withdraw without consequence. All participants received clear information about the study and provided written informed consent. The study was approved by the Helsinki and Uusimaa Hospital District Ethics Committee and by Helsinki and Uusimaa Hospital District Psychiatry. Data were anonymized, and participants received 4 movie tickets as compensation. Further details on study design, participants recruitment, and data collection procedure can be found in previous studies [37,38].

The study period was divided into two phases: an initial 2-week active phase and a subsequent passive phase lasting up to 1 year. In this work, we focused exclusively on the active phase of data collection. During this phase, participants wore wrist-worn actigraphy devices and used bed sensors to monitor sleep patterns. Concurrently, smartphone and EMA data was gathered through the AWARE app [39], which tracked various behavioral metrics [37]. Participants were also prompted 5 times daily to report on their mood, energy levels, and other psychological states using a 7-point Likert scale. Participants returned the actigraphy devices and bed sensors after the active phase, but smartphone data collection continued during the passive phase. The duration of the active phase was determined based on the amount of time actigraphs could collect data without needing to be recharged. Charging the actigraphs would have significantly increased the burden on participants and was not deemed feasible. A more detailed description of the questions can be found elsewhere [40].

Sleep and Activity Monitoring Protocol

Participants, were provided with actigraphy devices (Phillips Actiwatch 2, Philips Respironics) to measure activity levels and sleep for 2 weeks. The devices can operate during this period without recharging to ensure that participants did not have to remove the device during the monitoring session. Participants were required to wear the devices on the wrist for as long as they could and were instructed to remove the device only during sauna use, which is a common activity in Finland.

In addition, nearable devices, including bed sensors and smartphones, were also used for sleep monitoring. The Murata SCA11H (Murata Electronics) nodes, based on ballistocardiography technology, were used for bed sensing. Each participant received a preconfigured Wi-Fi router to enable automatic data transfer to the study server. The participants were asked to place the Murata SCA11H node under their mattress, positioned close to but not directly beneath them. Alternatively, they could be attached to the bed frame near the participant. The sensors processed information on site and produced metrics such as pulse rate, heart rhythm variability, breathing rate, cardiac output, and signal intensity at a 1 Hz rate [41].

Smartphones were used for sleep tracking due to their noninvasive nature and ability to monitor behaviors [20]. We collected communication timestamps (calls and texts), anonymous contact identifiers, smartphone screen activity events (screen on, off, lock, and unlock), location data, app usage patterns, and battery state. In this study, we only analyzed screen events to identify sleep periods.

Furthermore, smartphones facilitated the use of EMA, which was used to gather real-time data on sleep and participants' emotional states throughout the study [37]. During the active phase, participants received multiple EMA prompts daily on their smartphones. These included a morning questionnaire to assess sleep from the previous night and an evening questionnaire to track the activities of the day. Additionally, a total of 3 randomly timed prompts were sent throughout the afternoon within specific time ranges. The morning questionnaire included a single item asking participants to report how long they had slept during the previous night. The item was answered using categorical response options rather than an open text field. The response options were: under 5 hours, 5-6 hours, 6-7 hours, 7-8 hours, 8-9 hours, 9-10 hours, and over 10 hours. Participants were prompted to complete the morning questionnaire at a fixed time in the morning rather than immediately upon awakening, and they were given a specific time window to respond. Similarly, the evening questionnaire was completed within a designated evening window. This schedule allowed for collecting subjective data on sleep, mood, energy levels, physical activity, and other psychological factors [40]. Data from all modalities were gathered through the Niima data collection platform [42] and was preprocessed with the Niimpy behavior analysis toolbox [43].

Device Data Export Procedures and Analysis

Sleep data were derived using proprietary algorithms from both the actigraph and bed sensor, which were then transformed into sleep parameters for analysis. For the actigraph, raw data were aggregated into 30-second intervals, with each interval assigned 1 of 3 possible labels based on activity levels. The "ACTIVE" label indicated high activity, indicating the user was awake and moving. The "REST" label represented low activity, suggesting the user was resting but not necessarily asleep. The "REST-S" label denoted sustained low activity, indicating that the user was likely asleep. To identify sleep episodes, the data was sorted chronologically, and continuous REST-S intervals were grouped together and treated as a single sleep episode.

The bed sensor collected raw data on heart rate and respiratory rate at 1-second intervals. This dataset was processed using the manufacturer's algorithm to generate labels indicating bed occupancy and physiological states [37]. Sleep periods were identified based on the status variable, where a status of "1"

XSL•FO
RenderX

indicated sleep, and "0," "2," and "3" represented nonsleep states (eg, not in bed, awake in bed, or signal overload). Continuous status "1" intervals were treated as sleep episodes and grouped together as part of the same sleep period.

For the smartphone data, sleep was inferred based on the lock and unlock status as indicators of inactivity, following the procedure described by Aledavood et al [37] and used in previous studies [44,45]. Periods of inactivity, identified when the device was locked, were assessed to determine sleep periods. The longest stretch of inactivity was classified as nocturnal sleep. The lock and unlock status was converted into a binary format, and the longest inactivity periods were tracked to calculate sleep onset, sleep offset, and TST.

In addition, the EMA questions and responses, initially in Finnish, were translated into English. Data were selected based on responses to the question, "How many hours did you sleep last night?" to examine the relationship between subjective and sensor-based sleep data. Categorical TST ranges were transformed into numerical values (eg, "6-7 hours" was converted to 6.5 hours).

Transforming Raw Data to Sleep Measures

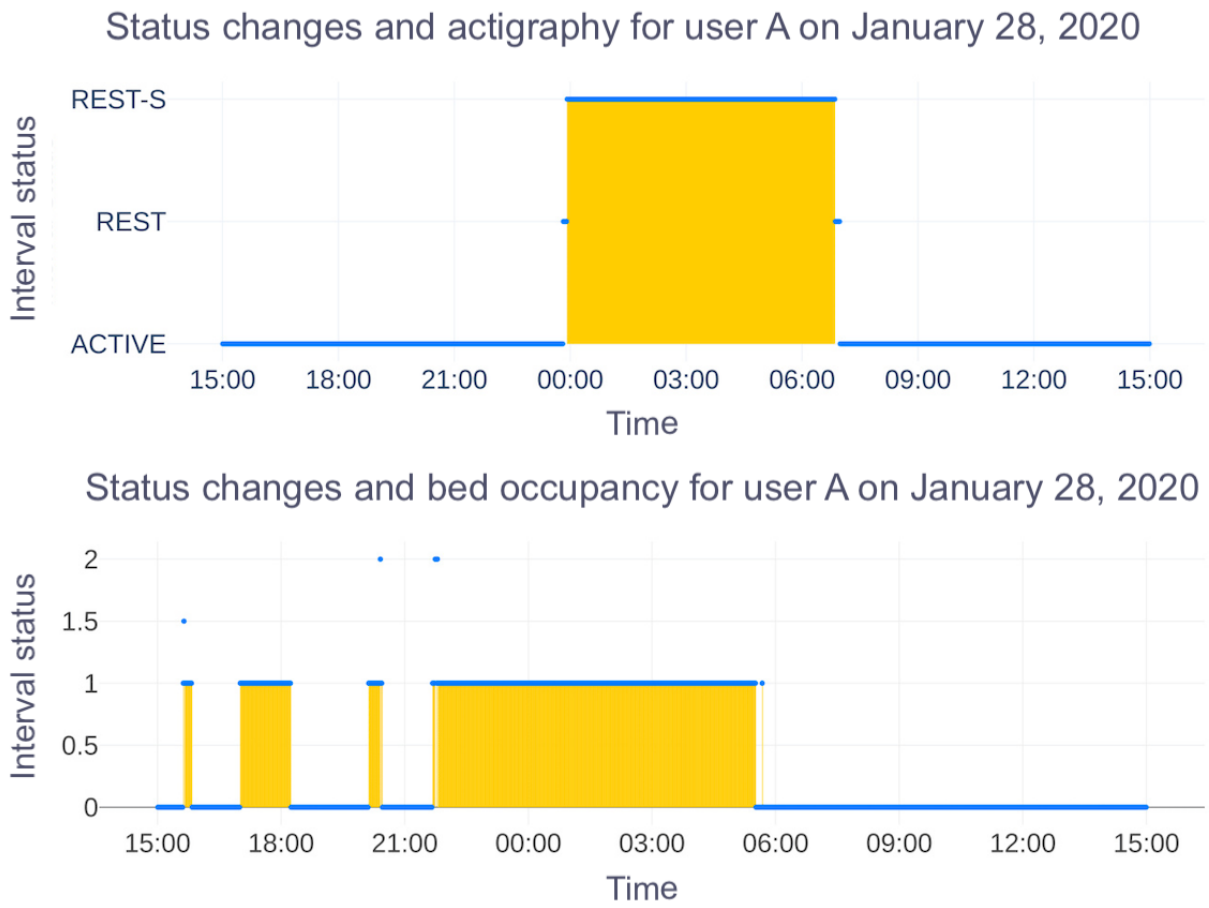
Data preprocessing was performed on the smartphone, actigraph, and bed sensor data. First, missing data were removed to ensure completeness. Second, timestamps were standardized to the Europe/Helsinki time zone to maintain consistency across all data sources. Third, sleep data were aligned using a 3 PM to 3 PM time frame, meaning each 24-hour period started and ended at 3 PM each day. This approach was chosen to account for variations in sleep schedules and to standardize sleep cycle measurements across participants.

A 5-minute threshold was applied to identify sleep periods, allowing gaps of up to 5 minutes between consecutive sleep intervals to be treated as part of the same sleep episode. TST outside the 3-13 hour range was excluded, following established guidelines [46]. This threshold was set to exclude implausible TST values, as unusually short or long sleep episodes could indicate naps, data errors, or atypical patterns.

Finally, sleep onset and offset were determined by identifying the start and end times of the longest detected sleep episode within each 24-hour period. Participants with missing values for these parameters were excluded from the analysis. Figure 1 illustrates how sleep periods were identified using the sleep labels provided by the manufacturer for both the actigraph and bed sensor.

XSL•FO
RenderX

Figure 1. Status changes identified by proprietary algorithms of the actigraph and bed sensor. For the actigraph, the sleep period began at 23:55 and ended at 06:51, marked by "Rest-S," while the bed sensor indicated "Status 1," showing the bed was occupied and the user was asleep from 21:40 to 05:30. This displays differences in sleep period detection and wakefulness between the two modalities.



Missing Data and Outlier Handling Procedures

Although 164 participants were initially recruited for the main study, 13 (7.9%) participants did not submit passive data and were removed from all evaluations. Additionally, missing data resulted from various factors, including device noncompliance, such as failing to wear or charge the device, user dropout, inconsistent usage, and data cleaning procedures. Technical issues, such as battery depletion, sensor malfunctions, or connectivity failures, may have also contributed to data loss. After filtering out missing data, the sample included 173 unique participants from both the main and pilot studies. The datasets included 102 actigraph users, 108 bed sensor users, and 148 smartphone users.

Additionally, the 2 largest outliers from both the onset and offset columns were identified and removed to limit the impact of extreme values. Due to the overlap between the outliers in both columns, a total of 3 outliers were removed from the bed sensor and actigraph datasets and 2 from the smartphone dataset, resulting in the exclusion of 8 observations in total. Manual checks revealed that these outliers were due to device-related issues, including improper device wear for the actigraph, placement or configuration errors for the bed sensor, and technical problems with the smartphone data collection. Following data cleaning, the final sample size consisted of 172 participants and was as follows: actigraph (29 healthy controls, 70 patients; n=99), bed sensor (34 healthy controls, 71 patients; n=105), smartphone (46 healthy controls, 100 patients; n=146), and EMA (48 healthy controls, 109 patients; n=157). Furthermore, a total of 3 participants from the pilot study did not provide information on age or gender and, therefore, their data were excluded from further analyses, resulting in a final sample of 169 participants.

Statistical Analysis

To assess the agreement between different sleep measurement modalities, Bland-Altman plots [47] were created by plotting the differences between paired measurements against their mean values. This approach provides a visual assessment of the differences between two modalities, revealing any systematic bias as well as the limit of agreement (LoA). Sleep onset, offset, and TST comparisons were made across modalities using this approach, with 95% LoA calculated. To account for day-to-day variability within each participant, data were aggregated by averaging individuals' data across multiple days. Pearson correlation analysis was conducted to quantify the relationships between sleep parameters from different modalities. Normality was first assessed using the Shapiro-Wilk test [48]. If deviations

XSL•FO
RenderX

from normality were observed, quantile-quantile plots were used for cross-checking [49]. The size of the correlation coefficients was interpreted based on the thresholds from the guidelines suggested by Mukaka [50]. A high correlation ranged from 0.70 to 1.00, a moderate correlation ranged from 0.50 to 0.70, a low correlation was between 0.30 and 0.50, and a negligible correlation was considered when ranged from 0.00 to 0.30. Finally, 95% CIs were calculated for the regression slopes.

Linear mixed models [51,52] were used to investigate factors associated with the alignment of sleep parameters between modalities and to account for the repeated measurements within each participant. The alignment of sleep parameters between measurement modality pairs was used as the dependent variable, represented as the absolute difference between the two modalities. The models adjusted for demographic factors, including age and gender. Chronotype was included as a fixed effect, measured by the Morning-Eveningness Questionnaire (MEQ), with higher scores indicating morningness preferences and lower scores indicating eveningness preferences. Given the unique geographical position of Finland, the seasonal factor was controlled for using day length duration. Between-group differences were evaluated by comparing each of the 3 patient groups with the healthy control group, which was used as the reference category. Participants were included as the random effect. The model was performed in the R Statistical Software (version 4.3.1; R Core Team) [53] using the *lme4* package [54]. All other analyses were performed in Python (version 3.12, Python Software Foundation) [55].

Results

Descriptive Statistics

The demographic characteristics of the participants are outlined in Table 1. Additionally, descriptive statistics for sleep onset, offset, and TST across modalities are presented in Table 2, based on daily averages to account for daily fluctuations. In the control group, sleep onset varies between modalities. The earliest average sleep onset was recorded by the actigraph (11:48 PM, SD 70 minutes), while the smartphone reported the latest (12:01 AM, SD 106 minutes). The bed sensor recorded a mean onset time of 11:58 PM (SD 67 minutes). In the patient group, the mean onset times were generally later, as the bed sensor recorded the earliest average onset at 12:19 AM (SD 100 minutes) and the actigraph reported the latest at 12:35 AM (SD 92 minutes). Lastly, the smartphone recorded a mean onset time of 12:30 AM (SD 101 minutes) for patients.

Table 1. Demographic characteristics of participants (N=169).

Characteristic	Value
Age (years), mean (SD)	35.1 (12.8)
Female, n (%)	125 (73.9)
MEQ[a] score, mean (SD)	39.2 (5.2)
Group, n (%)	
Major depressive disorder	78 (46.2)
Control	50 (29.6)
Bipolar disorder	21 (12.4)
Borderline personality disorder	20 (11.8)

[a]MEQ: Morning-Eveningness Questionnaire.

Table 2. Mean sleep parameters (onset, offset, and total sleep time) and SDs across different assessment methods for control and patient groups.

Parameter	Group	Bed sensor, mean (SD)	Actigraph, mean (SD)	Smartphone, mean (SD)	EMA[a], mean (SD)
Onset					
	Control	23:58 (67 min)	23:48 (70 min)	00:01 (106 min)	—[b]
	Patients	00:19 (100 min)	00:35 (92 min)	00:30 (101 min)	—
Offset					
	Control	07:33 (87 min)	07:44 (74 min)	07:06 (72 min)	—
	Patients	08:24 (119 min)	08:58 (91 min)	07:32 (92 min)	—
Total sleep time (duration)					
	Control	7 h 35 min (78 min)	7 h 55 min (62 min)	7 h 10 min (87 min)	7 h 14 min (35 min)
	Patients	8 h 5 min (107 min)	8 h 22 min (60 min)	7 h 7 min (78 min)	7 h 26 min (68 min)

[a]EMA: ecological momentary assessment.

[b]Not applicable.

XSL•FO
RenderX

Sleep offset showed a similar pattern, with differences between groups and measurement modalities observed. In the control group, the smartphone reported the earliest mean offset time (7:06 AM, SD 72 minutes), while the actigraph recorded the latest (7:44 AM, SD 74 minutes). The bed sensor estimated a mean offset time of 7:33 AM (SD 87 minutes). In the patient group, offsets occur later, with the smartphone recorded the earliest mean offset time at 7:32 AM (SD 92 min) and the actigraph reported the latest at 8:58 AM (SD 91 min). Finally, the bed sensor reported a mean offset of 8:24 AM (SD 119 min). All modalities indicate a trend of delayed offset times among patients compared to controls. Furthermore, patients show greater variability in sleep offset times, suggesting more inconsistency in their wake-up patterns.

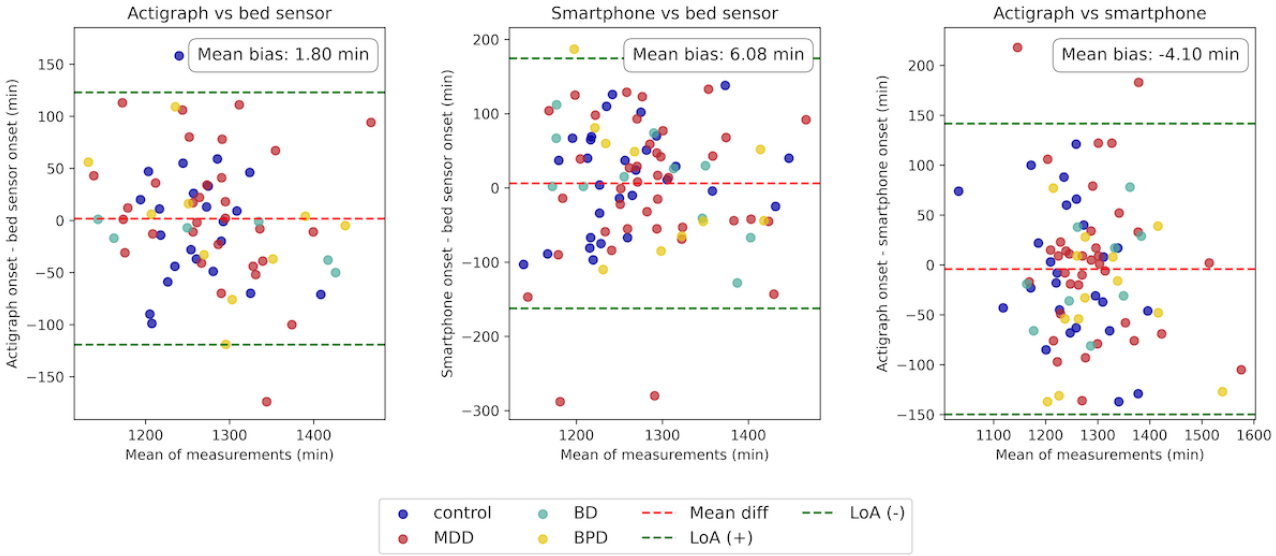
Finally, TST was evaluated using both objective measurements and subjective reports from the EMA questionnaires. In the control group, the smartphone recorded the shortest mean TST (7 hours 10 minutes, SD 87 minutes), while the actigraph recorded the longest (7 hours 55 minutes, SD 62 minutes). The bed sensor estimated a mean TST of 7 hours 35 minutes (SD 78 minutes), and subjective reports from the EMA questionnaires indicated a mean of 7 hours 14 minutes (SD 35 minutes). For the patient group, the mean TST was generally longer, with the smartphone recording the shortest mean TST at 7 hours 7 minutes (SD 78 minutes) and the actigraph recording the longest at 8 hours 22 minutes (SD 60 minutes). The bed sensor estimated a mean TST of 8 hours 5 minutes (SD 107 minutes), and EMA responses showed a mean TST of 7 hours 26 minutes (SD 68 minutes). Overall, the findings showed relatively consistent trends across the modalities in measuring TST.

Alignment of Sleep Parameters Across Assessment Modalities

Sleep Onset

Bland-Altman plots comparing sleep onset times across modalities are presented in Figure 2. The mean difference between sleep onset was not statistically significant (Table 3). However, sleep onset showed significant positive correlations across modalities, as displayed in Figure 3. Actigraph and bed sensor onset times had the strongest correlation (r=0.70, 95% CI 0.58-0.96; P<.001), indicating a high level of agreement between these modalities. The correlation between actigraph and smartphone onset times was moderate (r=0.69, 95% CI 0.60-0.95; P<.001), indicating that these two modalities capture similar sleep onset patterns. The correlation between smartphone and bed sensor onset times was also moderate (r=0.50, 95% CI 0.32-0.69; P<.001), suggesting greater variability in smartphone-based measurements.

Figure 2. Bland-Altman plot comparing sleep onset times across modalities, with data points color-coded by group. The red dashed line indicates the mean bias, and the green dashed lines represent the 95% limit of agreement (LoA). The smallest mean bias is observed between the actigraph and bed sensor at 1.80 minutes.



XSL•FO
RenderX

Figure 3. Scatter plots comparing sleep onset times across different modalities, with each point representing a paired dataset. Data points are color-coded to distinguish control and patient groups. All comparisons show positive correlations, with the strongest correlation observed between the actigraph and bed sensor.

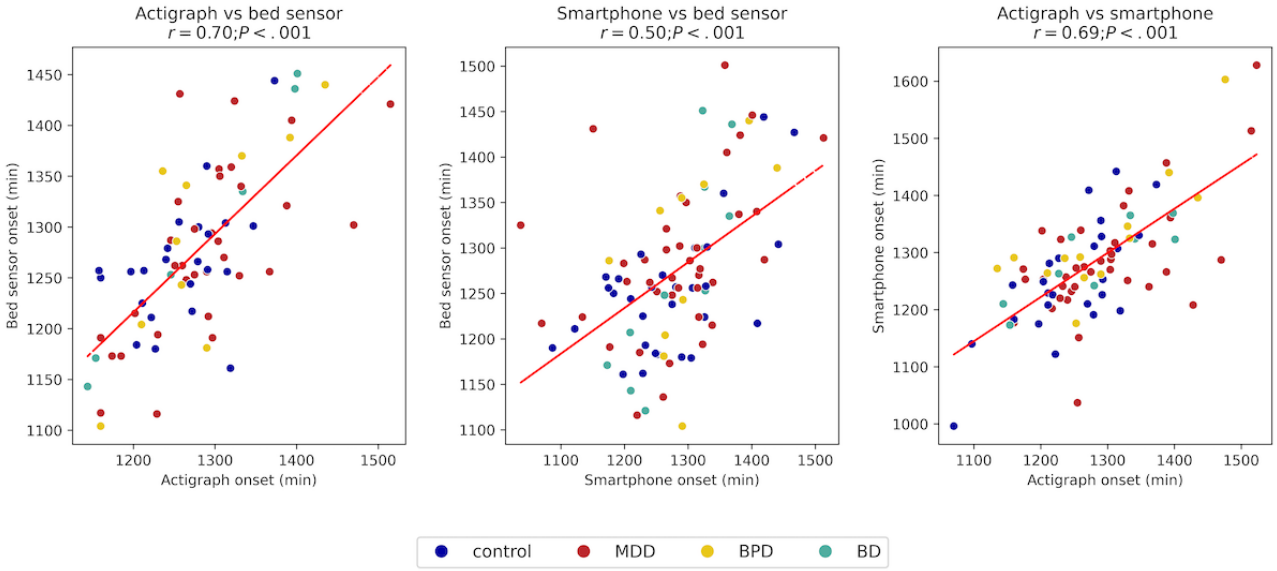


Table 3. Mean bias and statistical comparison of sleep metrics across assessment methods.

Sleep metric and comparison	Mean bias (SD)	*t* test (*df*)	*P* value
Onset			
Actigraph vs bed sensor	1.80 (62.23)	0.24 (70)	.81
Smartphone vs bed sensor	6.08 (86.42)	0.67 (89)	.51
Actigraph vs smartphone	–4.10 (74.78)	–0.50 (83)	.62
Offset			
Actigraph vs bed sensor	30.39 (85.26)	3.00 (70)	<.001
Smartphone vs bed sensor	–37.86 (107.19)	–3.35 (89)	<.001
Actigraph vs smartphone	64.85 (86.78)	6.85 (83)	<.001
Total sleep time			
Actigraph vs bed sensor	28.57 (103.03)	1.51 (98)	.13
Actigraph vs smartphone	59.21(87.56)	6.29 (98)	<.001
Actigraph vs EMA[a]	58.51 (66.42)	5.43 (98)	<.001
EMA vs bed sensor	–24.9 (93.15)	–2.49 (104)	.01
EMA vs smartphone	22.61 (79.74)	1.81 (120)	.07
Smartphone vs bed sensor	–34.55 (111.37)	–3.57 (104)	<.001

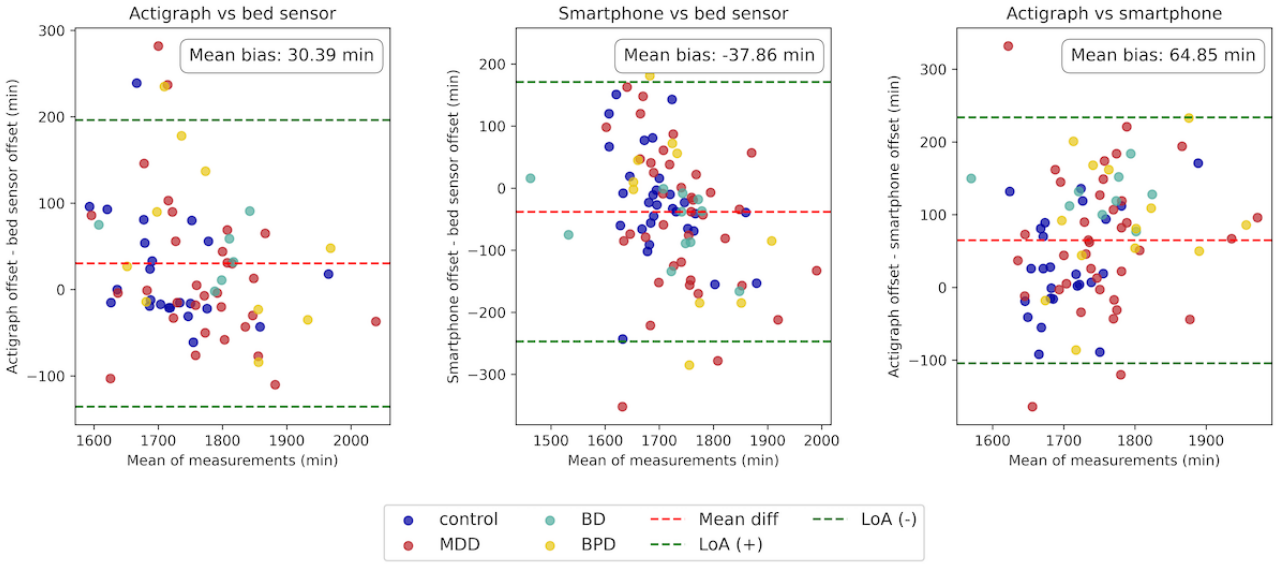
[a]EMA: ecological momentary assessment.

Sleep Offset

Figure 4 presents Bland-Altman plots comparing sleep offset times across the different measurement modalities. The mean bias for sleep offset across all modalities was statistically significant (P<.001). In particular, the mean bias between the actigraph and bed sensor offsets was 30.39 (SD 85.26; 95% CI 10.36-50.43) minutes, indicating a positive bias, where the actigraph tends to report later onset times compared to the bed sensor. The smartphone offset differed from the bed sensor by a mean of –37.86 (SD 107.19; 95% CI –60.18 to –15.53) minutes, showing a negative bias, with the smartphone tending to report earlier offset times compared to the bed sensor. Lastly, the mean difference between the actigraph and smartphone offsets was 64.85 (SD 86.78; 95% CI 46.13-83.57) minutes, reflecting a larger positive bias, with the actigraph reporting significantly later onset times compared to the smartphone.

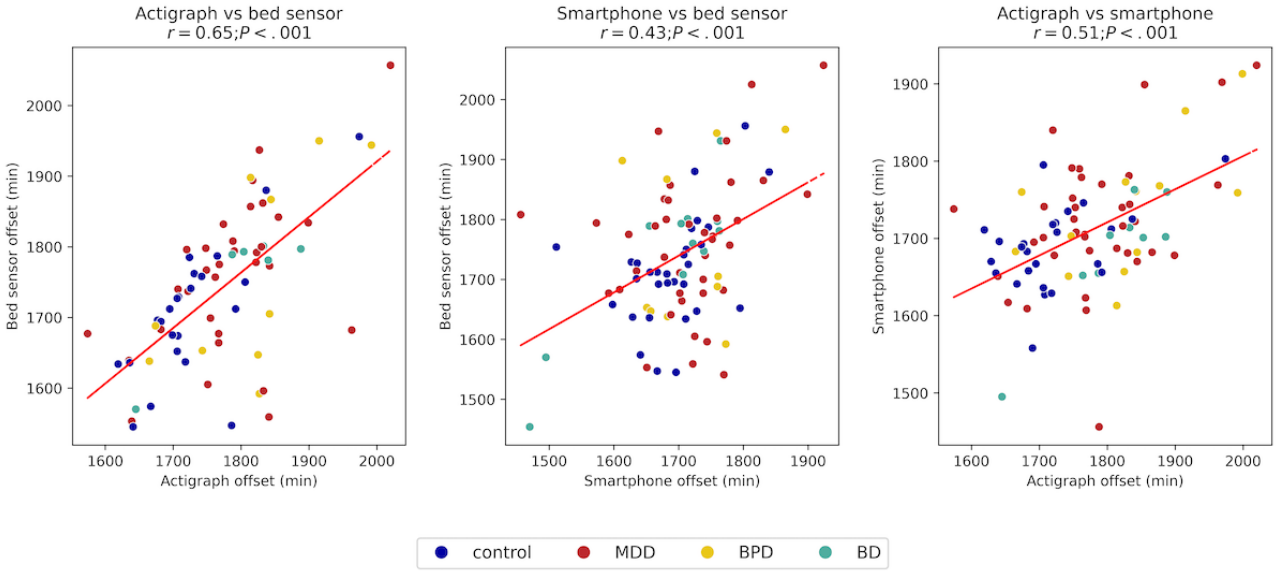
XSL•FO
RenderX

Figure 4. Bland-Altman plot comparing sleep offset times across modalities, with data points color-coded by group. The red dashed line indicates the mean bias, and the green dashed lines represent the 95% limit of agreement (LoA). The actigraph records later offsets, while the smartphone records earlier offsets relative to other modalities. The largest mean bias is observed between the actigraph and smartphone at 64.85 minutes.



Moreover, all modalities showed moderate to low correlation in measuring sleep offset (Figure 5). The highest correlation was observed between the actigraph and bed sensor (r=0.65, 95% CI 0.57-1.01; P<.001), followed by the correlation between the actigraph and smartphone (r=0.51, 95% CI 0.27-0.59; P<.001). The smartphone and bed sensor showed the lowest correlation (r=0.43, 95% CI 0.34-0.88; P<.001). Results indicated moderate to low alignment in sleep offset measurements across modalities, with greater variability observed in smartphone-based measurements.

Figure 5. Scatter plots comparing sleep offset times across different modalities. All offsets show a mild positive correlation, with the strongest correlation observed between the actigraph and bed sensor.

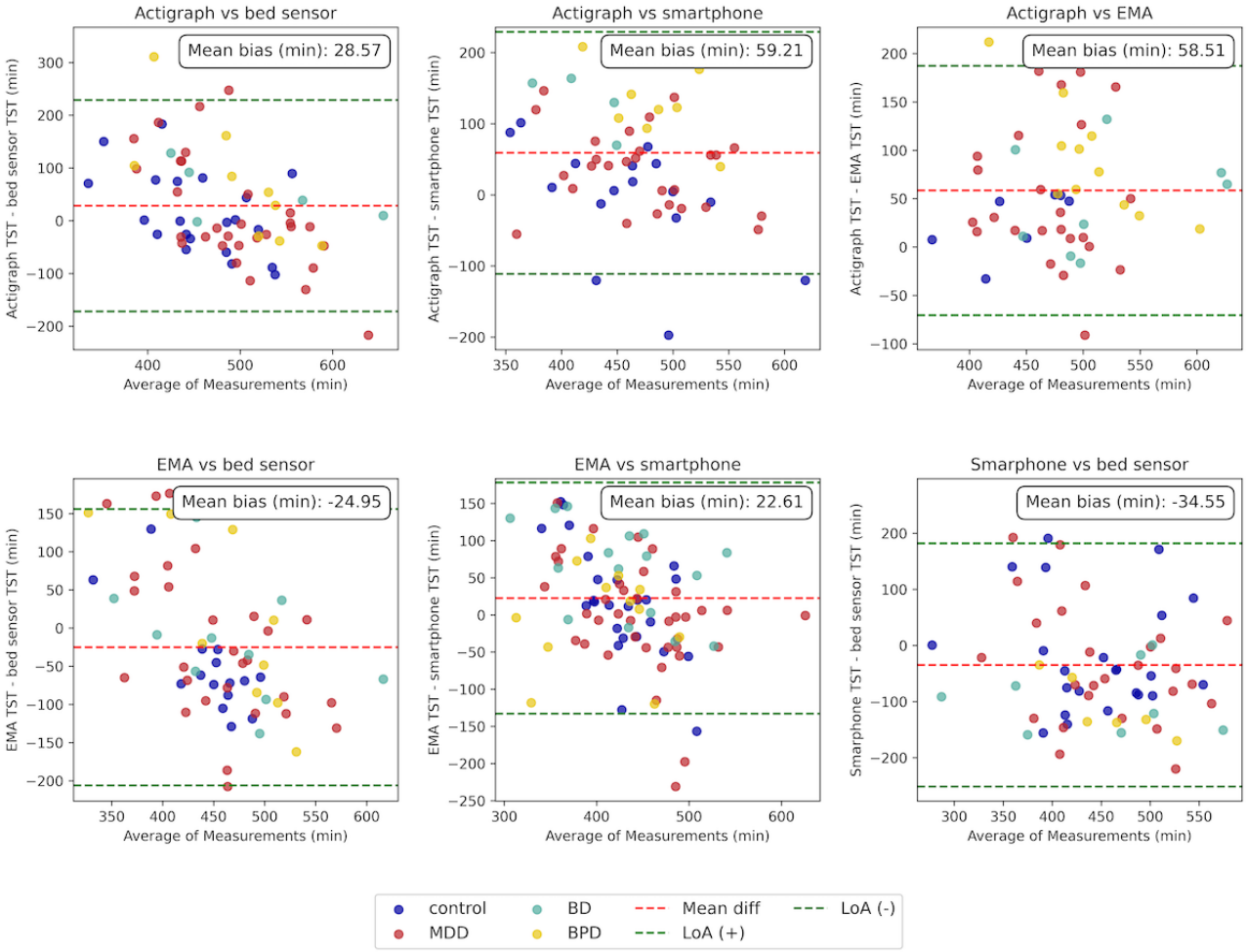


Total Sleep Time

The Bland-Altman plot in Figure 6 showed that the actigraph consistently reported higher TST compared to the other modalities. The mean difference between actigraph and bed sensor in TST was 28.57 (SD 103.03; 95% CI 10.83-46.31) minutes, reflecting a small positive bias, which was not statistically significant (P=.13). The actigraph also overestimated TST compared to the smartphone by 59.21 (SD 87.56; 95% CI 44.41-74.00) minutes and by 58.51 (SD 66.42; 95% CI 47.68-69.34) minutes when compared to the EMA data, both of which were statistically significant (P<.001).

XSL•FO
RenderX

Figure 6. Bland-Altman plot comparing total sleep time (TST) across modalities, with data points color-coded by group. The red dashed line indicates the mean bias, and the green dashed lines represent the 95% limit of agreement (LoA). The actigraph reported higher TST compared to other modalities, while the smartphone reported lower values.

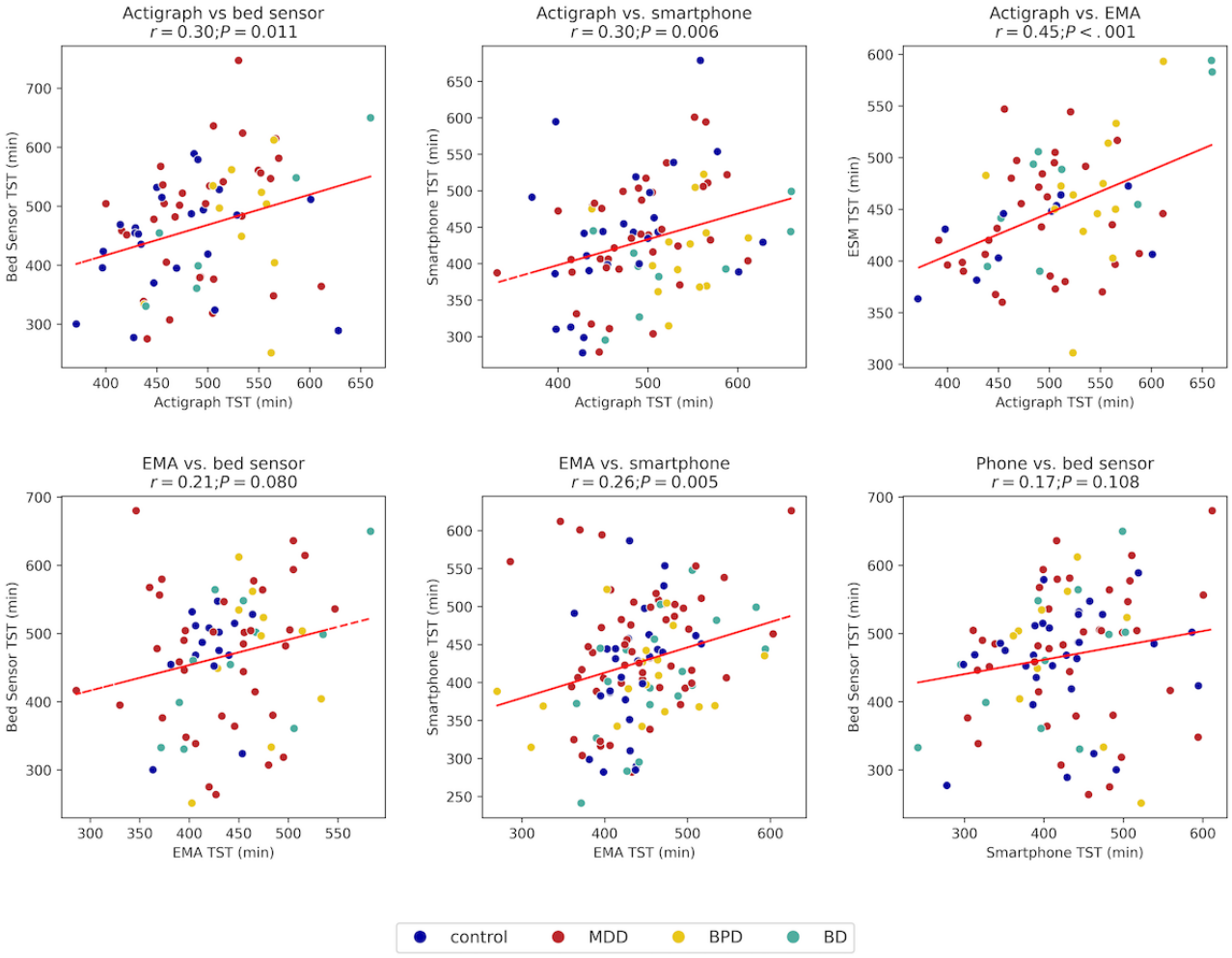


EMA slightly underestimated TST relative to the bed sensor and actigraph but overestimated it compared to the smartphone. The mean difference between EMA and the bed sensor was –24.95 (SD 93.15; 95% CI –40.08 to –9.82) minutes, reflecting a small negative bias, which was statistically significant (*P*=.01). In comparison to the smartphone, EMA had a mean difference of 22.61 (SD 79.74; 95% CI 9.52-35.69) minutes, indicating a very small positive bias, though this difference was not marginally significant (*P*=.07). Lastly, the smartphone underestimated TST compared to the bed sensor, with a mean difference of –34.55 (SD 111.37; 95% CI –52.82 to –16.26) minutes, which was statistically significant (*P*<.001). These results suggest that there are significant biases between the actigraph and both EMA and the smartphone. The actigraph overestimated TST, likely because it can misclassify wakefulness as sleep. EMA, based on subjective recall, slightly underestimated TST. The smartphone showed the greatest underestimation, which may be due to screen activity shortly before sleep onset and after waking.

Pearson correlations revealed varying levels of agreement between modalities for TST (Figure 7). The strongest correlation was observed between actigraph and EMA (*r*=0.45, 95% CI 0.20-0.63; *P*<.001), Actigraph and bed sensor sleep measurements demonstrated a low positive correlation (*r*=0.30, 95% CI 0.12-0.91; *P*=.01), while actigraph and smartphone measurements showed a similar low correlation (*r*=0.30, 95% CI 0.10-0.60; *P*=.006). The correlation between EMA and bed sensor data was less pronounced and marginally significant (*r*=0.21, 95% CI –0.05 to 0.79; *P*=.08). EMA and smartphone measurements showed a negligible correlation (*r*=0.26, 95% CI 0.10-0.56; *P*=.005). Similarly, smartphone and bed sensor data did not exhibit a significant correlation (*r*=0.17, 95% CI –0.04 to 0.47; *P*=.11), suggesting greater variability in smartphone-based sleep estimates.

XSL•FO
RenderX

Figure 7. Scatter plots comparing total sleep time (TST) across different modalities. All comparisons show weak positive correlations, with ecological momentary assessment (EMA) versus bed sensor and smartphone versus bed sensor not exhibiting statistically significant correlations.



Sleep Measurement Discrepancies in Patient and Control Groups

Paired *t* tests (2-tailed) were used to compare sleep onset times between devices and methods for controls and patients separately, with differences reported in minutes. Linear regression with Pearson correlation was also performed to assess the relationship between device pairs.

Sleep Onset

Analysis of sleep onset times using paired *t* tests (2-tailed) revealed minimal mean biases (Table 4) between measurement modalities in both control and patient groups. For instance, the mean bias between actigraph and bed sensor was –4.57 minutes in controls (t_{22}=–0.37; *P*=.71) and 4.85 minutes in patients (t_{47}=0.52; *P*=.60), indicating no systematic difference in onset detection timing. Similarly, the smartphone versus bed sensor comparison showed a mean bias of 18.2 minutes in controls (*P*=.20) but was near zero in patients (0.02 minutes; *P*=.99). Despite these small biases, correlations (Table 5) revealed stronger agreement in patients than controls for some modality pairs. Notably, the actigraph versus bed sensor correlation was 0.74 (*P*<.001) in patients, compared to 0.49 (*P*=.02) in controls, suggesting more consistent relative timing across these modalities in patients. The smartphone versus bed sensor and actigraph versus smartphone onset times also showed significant positive correlations in both groups. These findings suggest that although absolute onset time differences between devices are small and nonsignificant, the consistency between modalities is greater in patient groups.

XSL•FO
RenderX

Table 4. Mean bias and statistical comparison of sleep metrics across methods and groups.

Sleep metric, comparison, and group	Mean bias (SD)	*t* test (*df*)	*P* value
Onset			
Actigraph vs bed sensor			
Control	–4.57 (58.9)	–0.37 (22)	.71
Patients	4.85 (64.2)	0.52 (47)	.60
Smartphone vs bed sensor			
Control	18.2 (75.2)	1.32 (29)	.20
Patients	0.02 (91.5)	0.0014 (59)	.99
Actigraph vs smartphone			
Control	–8.00 (68.2)	–0.59 (24)	.56
Patients	–2.44 (77.9)	–0.24 (58)	.81
Offset			
Actigraph vs bed sensor			
Control	20.9 (65.8)	1.52 (22)	.14
Patients	34.9 (93.5)	2.59 (47)	.01
Smartphone vs bed sensor			
Control	–22.9 (86.0)	–1.46 (29)	.15
Patients	–45.3 (116.3)	–3.02 (59)	.004
Actigraph vs smartphone			
Control	32.2 (70.4)	2.29 (24)	.03
Patients	78.7 (89.9)	6.72 (58)	<.001
Total sleep time			
Actigraph vs bed sensor			
Control	33.9 (105.44)	1.74 (28)	.09
Patients	28.6 (104.83)	1.10 (69)	.27
Actigraph vs EMA			
Control	53.3 (61.65)	3.11 (27)	.004
Patients	61.6 (67.28)	5.10 (69)	<.001
Actigraph vs smartphone			
Control	8.26 (98.54)	1.69 (28)	.10
Patients	80.9 (78.04)	6.65 (69)	<.001
EMA vs bed sensor			
Control	–42.5 (75.975)	–0.70 (27)	.49
Patients	–20.9 (97.54)	–2.51 (70)	.01
EMA vs smartphone			
Control	3.65 (80.44)	0.05 (27)	.96
Patients	18.9 (90.22)	1.65 (92)	.10
Smartphone vs bed sensor			
Control	–14.5 (100.62)	–0.92 (33)	.36
Patients	–42.5 (116.72)	–3.57 (70)	<.001

XSL•FO
RenderX

Table 5. Correlations across assessment methods between control group and patients.

Sleep metric, comparison, and users	Correlation, *r* (95% CI)	*P* value
Onset		
Actigraph vs bed sensor		
Control	0.49 (0.10 to 0.95)	.02
Patients	0.74 (0.61 to 1.05)	<.001
Smartphone vs bed sensor		
Control	0.58 (0.21 to 0.68)	<.001
Patients	0.48 (0.27 to 0.77)	<.001
Actigraph vs smartphone		
Control	0.73 (0.59 to 1.38)	<.001
Patients	0.66 (0.50 to 0.92)	<.001
Offset		
Actigraph vs bed sensor		
Control	0.73 (0.54 to 1.32)	<.001
Patients	0.59 (0.44 to 1.04)	<.001
Smartphone vs bed sensor		
Control	0.44 (0.13 to 1.13)	.02
Patients	0.42 (0.25 to 0.92)	<.001
Actigraph vs smartphone		
Control	0.46 (0.05 to 0.59)	.02
Patients	0.49 (0.25 to 0.68)	<.001
Total sleep time		
Actigraph vs bed sensor		
Control	0.12 (–0.30 to 0.55)	.54
Patients	0.33 (0.06 to 0.61)	.02
Actigraph vs EMA		
Control	0.55 (–0.02 to 1.13)	.097
Patients	0.43 (0.18 to 0.68)	.002
Actigraph vs smartphone		
Control	0.36 (–0.02 to 0.74)	.08
Patients	0.30 (0.05 to 0.54)	.02
EMA vs bed sensor		
Control	0.31 (–0.21 to 0.83)	.26
Patients	0.21 (–0.05 to 0.47)	.12
EMA vs smartphone		
Control	0.34 (–0.02 to 0.71)	.07
Patients	0.25 (0.05 to 0.45)	.02
Smartphone vs bed sensor		
Control	0.17 (0.53 to –0.20)	.38
Patients	017 (–0.08 to 0.43)	.19

XSL•FO
RenderX

Sleep Offset

Significant mean biases were observed across modalities for sleep offset, particularly in the patient group (Table 4). In controls, the mean difference between actigraph and bed sensor was 20.91 minutes but was not significant (*P*=.14). In contrast, this difference increased to 34.94 minutes in patients and was statistically significant (t_{47}=2.59; *P*=.01), indicating that patient characteristics such as movement patterns or restlessness affect actigraph readings more strongly. Comparisons involving smartphone also showed larger and significant discrepancies in patients. The smartphone underestimated sleep offset by 45.3 minutes relative to the bed sensor (*P*=.004), whereas in controls the difference was nonsignificant (*P*=.16). The largest offset bias was observed between actigraph and smartphone in patients, with a mean difference of 78.7 minutes (*P*<.001), compared to 32.2 minutes in controls *(P*=.03).

Despite these differences, correlation analyses showed that the relative agreement across modalities remained comparable between groups (Table 5). Actigraph and bed sensor offset times correlated strongly in both controls (*r*=0.73, *P*<.001) and patients (*r*=0.59, *P*<.001). Smartphone and bed sensor correlations were moderate and similar across groups (controls: *r*=0.44; *P*=.02 and patients: *r*=0.42; *P*<.001). Actigraph and smartphone also showed comparable correlations (controls: *r*=0.46; *P*=.02 and patients: *r*=0.49; *P*<.001). These results indicate that absolute disagreement in sleep offset timing was greater among patients, especially involving smartphone data. However, the consistency of measurements across individuals, as reflected by correlation coefficients, was largely preserved.

Total Sleep Time

TST comparisons demonstrated the largest mean biases, with notable differences between controls and patients (Table 4). Controls exhibited small, marginally significant biases between modalities, such as 33.9 minutes for actigraph versus bed sensor (*P*=.09) and 8.26 minutes for actigraph versus smartphone (*P*=.10). Conversely, patients showed significantly larger biases: actigraph versus smartphone differed by 80.9 minutes (*P*<.001), actigraph versus EMA by 61.6 minutes (*P*<.001), and actigraph versus bed sensor by 28.6 minutes (*P*=.27). In Table 5, correlations for TST were generally weak and nonsignificant in controls (eg, actigraph vs bed sensor *r*=0.12; *P*=.54), but patients exhibited moderate significant correlations, including actigraph versus bed sensor (*r*=0.33; *P*=.02), actigraph versus smartphone (*r*=0.30; *P*=.02), and actigraph versus EMA (*r*=0.43; *P*=.002). EMA also correlated moderately with smartphone in patients (*r*=0.25; *P*=.02), but not significantly in controls (*r*=0.34; *P*=.07). These findings suggest that while absolute differences in TST between modalities are more pronounced in patients, the modalities capture shared variance in sleep duration better in patient than in control groups.

Demographic and Seasonal Factors Predicting the Alignments of Sleep Measurement

For each sleep measure, 3 linear mixed-effects models were fitted, using the difference between modalities as the dependent variable and including demographic and seasonal factors as fixed effects. The results are summarized in Tables 6-11.

Table 6. Total sleep time (TST) alignment.

Predictors	Actigraph – bed		Smartphone – bed		Actigraph – smartphone	
	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value
Age	0.00 (–0.03 to 0.03)	.97	–0.02 (–0.05 to 0.01)	.11	–0.02 (–0.04 to –0.00)	.048
Sex (female)	–0.61 (–1.51 to 0.29)	.18	–1.03 (–1.74 to –0.33)	.004	0.52 (–0.05 to 1.09)	.08
MEQ[a]	–0.04 (–0.12 to 0.03)	.23	–0.04 (–0.11 to 0.02)	.19	–0.04 (–0.08 to 0.01)	.10
Day length (minutes)	–0.09 (–0.19 to 0.02)	.10	-0.04 (–0.12 to 0.05)	.37	–0.07 (–0.14 to –0.01)	.03
Group: BD[b]	-0.86 (–2.27 to 0.54)	.23	–0.50 (–1.57 to 0.58)	.36	0.89 (0.11 to 1.68)	.03
Group: BPD[c]	0.00 (–1.34 to 1.35)	>99	–0.03 (–1.15 to 1.08)	.95	0.83 (0.01 to 1.64)	.05
Group: MDD[d]	–0.08 (–1.21 to 1.06)	.90	–0.60 (–1.52 to 0.32)	.20	–0.16 (–0.77 to 0.46)	.62

[a]MEQ: Morning-Eveningness Questionnaire.

[b]BD: bipolar disorder.

[c]BPD: borderline personality disorder.

[d]MDD: major depressive disorder.

XSL•FO
RenderX

Table 7. Random-effects variance components for total sleep time (TST) alignment.

Metric	Actigraph – bed	Smartphone – bed	Actigraph – smartphone
σ^2	2.47	2.84	2.86
τ (user)	1.27	0.82	0.74
ICC[a]	0.34	0.22	0.21
Marginal R^2/Conditional R^2	0.045/0.369	0.073/0.281	0.111/0.294

[a]ICC: intraclass correlation coefficient.

Table 8. Sleep offset alignment.

Predictors	Actigraph – Bed		Smartphone – Bed		Actigraph – Smartphone	
	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value
Age	0.01 (–0.02 to 0.04)	.59	–0.03 (–0.06 to –0.00)	.03	–0.01 (–0.04 to 0.01)	.20
Sex (female)	–0.57 (–1.49 to 0.36)	.23	–0.18 (–0.95 to 0.59)	.64	0.51 (–0.13 to 1.15)	.12
MEQ[a]	–0.05 (–0.12 to 0.03)	.22	0.01 (–0.06 to 0.08)	.68	–0.02 (–0.07 to 0.03)	.46
Daylight duration	–0.11 (–0.22 to –0.00)	.045	–0.11 (–0.20 to –0.01)	.03	–0.07 (–0.14 to 0.00)	.07
Group: BD[b]	–0.33 (–1.78 to 1.11)	.65	–0.32 (–1.47 to 0.83)	.59	0.66 (–0.20 to 1.53)	.13
Group: BPD[c]	0.84 (–0.55 to 2.23)	.24	0.10 (–1.10 to 1.29)	.88	0.83 (–0.07 to 1.74)	.07
Group: MDD[d]	0.67 (–0.50 to 1.84)	.26	–0.13 (–1.12 to 0.86)	.80	–0.15 (–0.84 to 0.53)	.66

[a]MEQ: Morning-Eveningness Questionnaire.
[b]BD: bipolar disorder.
[c]BPD: borderline personality disorder.
[d]MDD: major depressive disorder.

Table 9. Random-effects variance components for sleep offset alignment.

Metric	Actigraph – Bed	Smartphone – Bed	Actigraph – Smartphone
σ^2	3.15	7.18	7.15
τ (user)	1.32	0.67	0.72
ICC[a]	0.30	0.09	0.09
Marginal R^2/Conditional R^2	0.075/0.348	0.042/0.124	0.042/0.130

[a]ICC: intraclass correlation coefficient.

XSL•FO
RenderX

Table 10. Sleep onset alignment.

Predictors	Actigraph – Bed		Smartphone – BedEstimate		Actigraph – Smartphone	
	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value
Age	0.03 (–0.01 to 0.07)	.20	–0.01 (–0.05 to 0.02)	.56	0.01 (–0.02 to 0.03)	.45
Sex (female)	–0.14 (–1.28 to 1.00)	.81	–0.52 (–1.42 to 0.38)	.26	0.53 (–0.16 to 1.22)	.13
MEQ[a]	–0.01 (–0.10 to 0.08)	.88	–0.04 (–0.12 to 0.05)	.40	–0.04 (–0.09 to 0.02)	.17
Daylight duration	–0.02 (–0.15 to 0.12)	.81	0.00 (–0.10 to 0.11)	.93	–0.03 (–0.11 to 0.04)	.41
Group: BD[b]	–0.40 (–2.18 to 1.39)	.66	–0.19 (–1.57 to 1.19)	.79	0.01 (–0.94 to 0.97)	.98
Group: BPD[c]	0.41 (–1.30 to 2.12)	.64	0.67 (–0.77 to 2.10)	.36	0.69 (–0.29 to 1.67)	.17
Group: MDD[d]	0.68 (–0.75 to 2.12)	.35	–0.25 (–1.42 to 0.93)	.68	0.01 (–0.74 to 0.75)	.99

[a]MEQ: Morning-Eveningness Questionnaire.

[b]BD: bipolar disorder.

[c]BPD: borderline personality disorder.

[d]MDD: major depressive disorder.

Table 11. Random-effects variance components for sleep onset alignment.

Metric	Actigraph – Bed	Smartphone – Bed	Actigraph – Smartphone
σ^2	2.97	2.47	2.97
τ (user)	2.12	1.54	1.16
ICC[a]	0.42	0.38	0.28
Marginal R^2/Conditional R^2	0.036/0.437	0.045/0.413	0.045/0.313

[a]ICC: intraclass correlation coefficient.

Older age was associated with higher alignment in TST between actigraphy and the bed sensor (β=−0.02, 95% CI −0.04 to 0.00; *P*=.048), as well as higher alignment in sleep offset between the smartphone and the bed sensor (β=−0.03, 95% CI −0.06 to 0.00; *P*=.03). Males exhibited higher alignment compared with females in TST between smartphone and bed sensor measurements (β=−1.03, 95% CI −1.74 to 0.33; *P*=.004). Longer daylight duration consistently predicted higher alignment in sleep offset for both actigraphy–bed sensor (β=–0.11, 95% CI –0.22 to 0.00; *P*=.045) and smartphone–bed sensor comparisons (β=–0.11, 95% CI −0.20 to −0.01; *P*=.03), highlighting the impact of weather on the alignment between measurement modalities. Patients diagnosed with BD (β=0.89, 95% CI 0.11-1.68; *P*=.03) and BPD (β=0.83, 95% CI 0.01-1.64; *P*=.047) demonstrated lower alignment in TST between actigraphy and bed sensor compared to controls. No significant predictors of sleep onset alignment were identified.

Discussion

Principal Findings

Our results showed that patients exhibited greater variability in sleep onset, offset, and TST compared to healthy controls. While actigraphy generally reported longer TST, smartphones typically reported lower values. Sleep onset and offset showed stronger correlations across modalities, whereas TST correlations were weaker. The largest measurement biases were observed for sleep offset timing and TST in patients, with the greatest discrepancies involving smartphone-derived data. However, correlation analyses suggest that relative agreement across devices remains intact within both groups. Further analysis with linear mixed models revealed that older individuals showed better alignment between modalities, for example in TST between actigraphy and the bed sensor and in sleep offset between the smartphone and the bed sensor. In contrast, females and patients with BD or BPD exhibited poorer alignment in TST between modalities. Additionally, longer daylight duration was linked to better alignment in both sleep offset and TST across all modalities.

Comparison With Previous Work

Accurate sleep monitoring outside of laboratory settings is increasingly relevant for mental health research, particularly in the absence of PSG as ground truth [3]. As evidenced in this study, sleep measurements among individuals with depressive disorders exhibited greater between-individual variability compared to the control group, as reflected by the wider CIs. Patients with depression may experience either insomnia, such as trouble falling asleep, frequent awakenings, or early morning awakenings, or hypersomnia. Hypersomnia is often linked to atypical depression, a subtype of major depressive episodes identified in the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) [56], and contributes significantly to the variation in sleep patterns among individuals with depression. These findings align with prior research, including Ho et al [34], which reported notable disturbances in both day and night activities among individuals experiencing MDD or symptoms of depression. Their meta-analysis review highlighted disruptions in sleep onset and sleep continuity,

XSL•FO
RenderX

poorer sleep efficiency, and greater variability in circadian sleep-wake patterns in relation to healthy controls.

Our results reveal systematic biases in sleep estimation in all measurement modalities. The higher values of TST as measured by actigraphy may be attributed to its limited ability to detect wakefulness during sleep, leading to an overestimation of TST compared to other modalities, a finding consistent with previous studies that have shown similar patterns of overestimation by actigraphy [13,57]. On the other hand, smartphone data may have limitations in accurately capturing sleep onset and offset times. The Bland-Altman plots showed that the differences in sleep onset between the smartphone and other modalities were significantly larger, which could contribute to the smartphone's underestimation of TST. This aligns with findings from Ciman and Wac [58], who noted that smartphone algorithms, while more accurate in estimating TST, struggle to detect precise sleep onset and offset. They noted that individuals often stay in bed after waking up before using their smartphones, making it difficult for the smartphone to detect the precise offset time. Similarly, smartphone use before sleep may delay the detected sleep onset. Moreover, the smartphone's underestimation of TST in this study is consistent with previous research, such as Natale et al [59], which found that smartphone generally underestimated TST in healthy individuals, particularly during longer TSTs.

Our findings indicate that smartphone-based measurements exhibited the lowest mean bias and a positive correlation of TST with EMA data. These results partly align with those of Maynard et al [60], who found that a smartphone app strongly correlated with self-reported sleep diaries. However, while their study reported that the smartphone overestimated TST compared to actigraphy in healthy adults, our results indicated that actigraphy consistently recorded higher TST, with healthy controls showing a small, marginally significant bias (8.26 minutes; P=.10) and patients exhibiting a much larger, significant bias (80.89 minutes; P<.001). Some discrepancy may stem from methodological differences, as Maynard's study used a proprietary algorithm to convert accelerometer data into sleep metrics, while this study relied on a simpler approach based on smartphone screen activity events tracking. Given the limited number of studies comparing smartphone-based sleep measurements, Maynard's study represents the closest methodological comparison. This further highlights the novelty of this study and enhances the validity of our results by showing that smartphone-based sleep measurements can provide meaningful insights even in clinical populations.

Regarding psychiatric populations, Staples et al [61] conducted one of the few studies to compare smartphone-based sleep monitoring in individuals with schizophrenia. They found that smartphone accelerometer estimates of TST were moderately correlated with self-reported TST (r=0.69). However, their study did not compare smartphone estimates directly with actigraphy or other modalities, making it difficult to fully interpret their findings. This study, which directly compared smartphone-derived TST estimates with those from other objective modalities in psychiatric populations, contributes to the sparse literature on this topic [20]. We emphasize the necessity for additional research to gain a deeper understanding of the accuracy of smartphone sleep monitoring in clinical populations, particularly when compared to other objective modalities.

Furthermore, the weak correlation in TST across different methods suggests that each device captures distinct aspects of sleep, leading to measurement inconsistencies. Such direct comparisons across multiple modalities within patient groups, without using PSG as a ground truth, are rarely conducted, making this study a unique contribution. One of the few studies in this area is Piantino et al [62], which identified a strong relationship between a bed sensor and an actigraphy device for metrics such as bed entry time, sleep onset, sleep offset, bed exit time, and TST in healthy individuals. However, Piantino et al also acknowledged that the consistency of the two modalities could be reduced in specific groups, such as individuals with depression or sleep disorders, who may be more likely to lie motionless in bed without achieving sleep. Additionally, in this study, some psychiatric patients were recruited from psychiatric units rather than primary care, where individuals often present with more severe symptoms. As a result, they may be more likely to be overmedicated. Sedative medications can also have muscle-relaxing effects, which may contribute to this phenomenon.

Factors Associated With the Alignments of Sleep Measurements

Demographic factors such as age and gender were associated with sleep measurement alignment. Older individuals exhibited better alignment between smartphone and actigraphy estimates, possibly because of more stable sleep patterns. Another explanation could be the reduced smartphone use before bedtime by older participants [63], which may align with actigraphy's overestimation tendencies. This also aligns with findings from Jonasdottir et al [64], who analyzed a large global dataset of over 11 million nights from wearable devices across 47 countries. Their study confirmed that older adults exhibit shorter TST, earlier sleep timing, and lower intraindividual variability compared with younger adults, which could contribute to more consistent sleep measurement alignment. Gender differences also played a role, as women showed more discrepancies between smartphone and bed sensor measurements. Jonasdottir et al [64] also found that women experience more frequent nighttime awakenings than men, particularly in early to middle adulthood. This suggests that increased sleep fragmentation in women may lead to greater inconsistencies between sensor-based measurements.

Additionally, day length appeared to correlate with better alignment, likely because extended daylight promotes more consistent sleep-wake schedules. These findings align with prior research suggesting that external cues, such as light exposure, play a significant role in regulating sleep timing [30-34]. This is particularly important given that the study was conducted in Finland, a country prone to extreme fluctuations in daylight. Such fluctuations may have a substantial impact on sleep patterns, making it critical to study and track these changes closely to better understand their effects on sleep and circadian rhythms.

XSL•FO
RenderX

Lastly, the greater alignment observed in patients with BD and those with BPD patients reinforces the established link between psychiatric disorders and sleep disturbances. Our findings also emphasize the comorbidity between these conditions, consistent with previous literature [65,66]. This is also supported by research linking these conditions to irregular sleep patterns. For instance, a meta-analysis found significant variations between patients with BPD and healthy participants in sleep stability and architecture indicators, including sleep efficiency, sleep onset delay, TST, and rapid eye movement sleep parameters [12]. Similarly, studies examining BD have highlighted disruptions in both quantity and quality, which are often influenced by the manic and depressive episodes characteristic of the disorder [67].

When using sleep monitoring to assess clinical status or detect early changes, it is important to recognize that there is no definitive standard for directly comparing different modalities. However, we can compare TST with subjective reports from EMA. Although no single device can be definitively identified as the best, we can identify which devices align closely with subjective reports. Our findings showed that, in terms of bias, smartphone had the lowest bias compared to EMA, although the difference was not statistically significant and may be due to random measurement error. Following this, the bed sensor exhibited the next lowest bias, while the actigraph had the highest bias. In contrast, when considering correlation, the actigraph and EMA showed the strongest alignment, suggesting that actigraphy may show a more consistent correlation with subjective sleep reports. However, because actigraphy has been shown to overestimate TST, it is possible that individuals also tend to overestimate their own TST. We note that additional context may contribute to device misalignment. For instance, co-sleeping can distort heart rate signals and the sleep metrics derived from them. Conversely, habitual snoozers usually experience brief awakenings to turn off an alarm, thus underestimating smartphone-based sleep estimates. Clinicians should take these nuances into account, along with the various factors that influence agreement between devices, when choosing a method for sleep monitoring.

Limitations and Future Work

A key strength of this study is its naturalistic approach, allowing for real-world sleep monitoring without interfering with participants' routines. This enhances ecological validity and provides insights into sleep measurement in everyday settings, particularly for individuals with depressive disorders. Additionally, using multiple modalities allows for a comprehensive comparison of different methodologies, highlighting their relative strengths and weaknesses. However, there are important limitations to consider. First, this study did not include PSG or sleep logs, as these could have disrupted participants' natural sleep patterns. Incorporating PSG would have posed significant challenges due to the need for frequent laboratory visits, which could have increased participant burden and potentially led to higher dropout rates, especially among patients. While this approach improves feasibility, it also means that we lack a clinical gold standard for validation. Without PSG data, determining the absolute accuracy of each device remains challenging. Second, although EMA provides subjective estimates of TST that align with some objective modalities, it remains susceptible to sleep misperception, where individuals may misjudge their sleep quality or duration [22]. Third, this study includes a relatively small sample size and short follow-up period, which may impact the generalizability of the results. Additionally, the sleep/wake labels used in the study were generated by proprietary algorithms provided by the manufacturers, which may limit the accuracy and comparability of the sleep data. Moreover, differences in sleep patterns may be partly attributable to the use of sleep medication, which is common among psychiatric patients but not typically used by healthy participants. TST outside the 3-13 hour range were excluded based on standard guidelines, which are suitable for healthy individuals but may overlook cases of hypersomnia in patients with depression who may sleep more than 13 hours. Finally, the study was conducted in a single location, resulting in a lack of population diversity. These limitations should be considered when interpreting the results, particularly in clinical contexts.

Future research could include replication of the study in countries with different climates or daylight patterns to determine whether places with different seasonal patterns also influence sleep characteristics in individuals. Studies incorporating PSG or validated sleep logs could help improve measurement accuracy while exploring ways to minimize participant burden. Larger and more diverse samples, as well as longer follow up periods, would allow for stronger conclusions regarding the stability of sleep patterns over time. Future work could also compare proprietary device algorithms with standardized scoring methods to evaluate the consistency of measurements. Examining sleep patterns among participants who use or do not use sleep medication may help clarify how medication influences digital sleep measures. In addition, more detailed clinical symptom data should be collected, as the present analyses lacked detailed information on specific sleep symptoms, such as insomnia or hypersomnia. Using validated measures to distinguish these sleep disturbances would improve understanding of how particular sleep problems relate to different sleep modalities and rest–activity patterns.

Conclusions

This study highlights the feasibility of using actigraphy, smartphone data, and bed sensors for sleep tracking in naturalistic settings for patients with depressive disorders. However, it also reveals systematic biases in sleep offset across these modalities, which contribute to discrepancies in TST. Furthermore, our findings highlight that factors such as age, sex, clinical diagnosis, and daylight duration can influence the alignment of sleep measurements across different modalities. This is particularly important in psychiatric populations, where accurate and reliable sleep tracking is essential for research and clinical care.

XSL•FO
RenderX

Acknowledgments

The authors gratefully acknowledge the computational resources provided by the Aalto Science-IT project. We used the generative artificial intelligence tool ChatGPT (OpenAI) to assist with proofreading and minor language editing. All content, interpretations, and conclusions were developed and verified by the authors.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to containing highly sensitive personal information and restrictions in the research permit, but are available from the corresponding author on reasonable request for purposes of scientific collaboration.

Conflicts of Interest

None declared.

References

1. Palagini L, Hertenstein E, Riemann D, Nissen C. Sleep, insomnia and mental health. J Sleep Res 2022;31(4):e13628. [doi: 10.1111/jsr.13628] [Medline: 35506356]
2. Dzierzewski J, Hendren L, Langer G. Sleep health in America: update from the National Sleep Foundation's 2022 Sleep Health Index®. Sleep 2023;46(Suppl 1):A129-A130. [doi: 10.1093/sleep/zsad077.0293]
3. Pan Q, Brulin D, Campo E. Current status and future challenges of sleep monitoring systems: systematic review. JMIR Biomed Eng 2020;5(1):e20921. [doi: 10.2196/20921]
4. Suarez EC, Goforth H. Sleep and inflammation: a potential link to chronic diseases. In: The Psychoneuroimmunology of Chronic Disease: Exploring the Links Between Inflammation, Stress, and Illness. Washington: American Psychological Association; 2010.
5. Medic G, Wille M, Hemels ME. Short- and long-term health consequences of sleep disruption. Nat Sci Sleep 2017;9:151-161 [FREE Full text] [doi: 10.2147/NSS.S134864] [Medline: 28579842]
6. Xu L, Jiang CQ, Lam TH, Liu B, Jin YL, Zhu T, et al. Short or long sleep duration is associated with memory impairment in older Chinese: the Guangzhou Biobank Cohort Study. Sleep 2011;34(5):575-580 [FREE Full text] [doi: 10.1093/sleep/34.5.575] [Medline: 21532950]
7. Zhao J, Liu H, Wu Z, Wang Y, Cao T, Lyu D, et al. Clinical features of the patients with major depressive disorder co-occurring insomnia and hypersomnia symptoms: a report of NSSD study. Sleep Med 2021;81:375-381. [doi: 10.1016/j.sleep.2021.03.005] [Medline: 33813234]
8. Nofzinger EA, Thase ME, Reynolds CF, Himmelhoch JM, Mallinger A, Houck P, et al. Hypersomnia in bipolar depression: a comparison with narcolepsy using the multiple sleep latency test. Am J Psychiatry 1991;148(9):1177-1181. [doi: 10.1176/ajp.148.9.1177] [Medline: 1882995]
9. Brower KJ, Perron BE. Sleep disturbance as a universal risk factor for relapse in addictions to psychoactive substances. Med Hypotheses 2010;74(5):928-933 [FREE Full text] [doi: 10.1016/j.mehy.2009.10.020] [Medline: 19910125]
10. Bernert RA, Joiner TE. Sleep disturbances and suicide risk: a review of the literature. Neuropsychiatr Dis Treat 2007;3(6):735-743 [FREE Full text] [doi: 10.2147/ndt.s1248] [Medline: 19300608]
11. Romier A, Maruani J, Lopez-Castroman J, Palagini L, Serafini G, Lejoyeux M, et al. Objective sleep markers of suicidal behaviors in patients with psychiatric disorders: a systematic review and meta-analysis. Sleep Med Rev 2023;68:101760. [doi: 10.1016/j.smrv.2023.101760] [Medline: 36706699]
12. Winsper C, Tang NKY, Marwaha S, Lereya ST, Gibbs M, Thompson A, et al. The sleep phenotype of borderline personality Disorder: a systematic review and meta-analysis. Neurosci Biobehav Rev 2017;73:48-67 [FREE Full text] [doi: 10.1016/j.neubiorev.2016.12.008] [Medline: 27988314]
13. Van de Water ATM, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography: a systematic review. J Sleep Res 2011;20(1 Pt 2):183-200. [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
14. McCall C, McCall WV. Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. J Sleep Res 2012;21(1):122-127 [FREE Full text] [doi: 10.1111/j.1365-2869.2011.00917.x] [Medline: 21447050]
15. Tang NKY, Harvey AG. Time estimation ability and distorted perception of sleep in insomnia. Behav Sleep Med 2005;3(3):134-150. [doi: 10.1207/s15402010bsm0303_2] [Medline: 15984915]
16. Trimmel K, Eder HG, Böck M, Stefanic-Kejik A, Klösch G, Seidel S. The (mis)perception of sleep: factors influencing the discrepancy between self-reported and objective sleep parameters. J Clin Sleep Med 2021;17(5):917-924 [FREE Full text] [doi: 10.5664/jcsm.9086] [Medline: 33393901]
17. Maltezos A, Perrault AA, Walsh NA, Phillips E, Gong K, Tarelli L, et al. Methodological approach to sleep state misperception in insomnia disorder: comparison between multiple nights of actigraphy recordings and a single night of

XSL•FO
RenderX

polysomnography recording. Sleep Med 2024;115:21-29 [FREE Full text] [doi: 10.1016/j.sleep.2024.01.027] [Medline: 38325157]
18. Acebo C, LeBourgeois MK. Actigraphy. Respir Care Clin N Am 2006;12(1):23-30. [doi: 10.1016/j.rcc.2005.11.010] [Medline: 16530645]
19. Hendriks MMS, van Lotringen JH, Vos-van der Hulst M, Keijsers NLW. Bed sensor technology for objective sleep monitoring within the clinical rehabilitation setting: observational feasibility study. JMIR Mhealth Uhealth 2021;9(2):e24339 [FREE Full text] [doi: 10.2196/24339] [Medline: 33555268]
20. Aledavood T, Torous J, Triana Hoyos AM, Naslund JA, Onnela JP, Keshavan M. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Curr Psychiatry Rep 2019;21(7):49 [FREE Full text] [doi: 10.1007/s11920-019-1043-y] [Medline: 31161412]
21. Kubey R, Larson R, Csikszentmihalyi M. Experience sampling method applications to communication research questions. J Commun 1996;46(2):99-120. [doi: 10.1111/j.1460-2466.1996.tb01476.x]
22. Krishnamurthy V, Mukherjee D, Reider A, Seaman S, Singh G, Fernandez-Mendoza J, et al. Subjective and objective sleep discrepancy in symptomatic bipolar disorder compared to healthy controls. J Affect Disord 2018;229:247-253. [doi: 10.1016/j.jad.2017.12.100] [Medline: 29329056]
23. van Rijssen IM, Hulst RY, Gorter JW, Gerritsen A, Visser-Meily JMA, Dudink J, et al. Device-based and subjective measurements of sleep in children with cerebral palsy: a comparison of sleep diary, actigraphy, and bed sensor data. J Clin Sleep Med 2023;19(1):35-43. [doi: 10.5664/jcsm.10246] [Medline: 35975545]
24. Luong N, Mark G, Kulshrestha J, Aledavood T. Sleep during the COVID-19 pandemic: longitudinal observational study combining multisensor data with questionnaires. JMIR Mhealth Uhealth 2024;12:e53389 [FREE Full text] [doi: 10.2196/53389] [Medline: 39226100]
25. Mattingly SM, Grover T, Martinez GJ, Aledavood T, Robles-Granda P, Nies K, et al. The effects of seasons and weather on sleep patterns measured through longitudinal multimodal sensing. NPJ Digit Med 2021;4(1):76 [FREE Full text] [doi: 10.1038/s41746-021-00435-2] [Medline: 33911176]
26. Mallinson DC, Kamenetsky ME, Hagen EW, Peppard PE. Subjective sleep measurement: comparing sleep diary to questionnaire. Nature and Science of Sleep 2019;11:197-206. [doi: 10.2147/nss.s217867]
27. Zou H, Zhou H, Yan R, Yao Z, Lu Q. Chronotype, circadian rhythm, and psychiatric disorders: recent evidence and potential mechanisms. Front Neurosci 2022;16:811771 [FREE Full text] [doi: 10.3389/fnins.2022.811771] [Medline: 36033630]
28. Borger JN, Huber R, Ghosh A. Capturing sleep-wake cycles by using day-to-day smartphone touchscreen interactions. NPJ Digit Med 2019;2:73 [FREE Full text] [doi: 10.1038/s41746-019-0147-4] [Medline: 31372507]
29. Kim KM, Han SM, Heo K, Kim W, Chu MK. Sex differences in the association between chronotype and risk of depression. Sci Rep 2020;10(1):18512 [FREE Full text] [doi: 10.1038/s41598-020-75724-z] [Medline: 33116223]
30. Potter GDM, Skene DJ, Arendt J, Cade JE, Grant PJ, Hardie LJ. Circadian rhythm and sleep disruption: causes, metabolic consequences, and countermeasures. Endocr Rev 2016;37(6):584-608 [FREE Full text] [doi: 10.1210/er.2016-1083] [Medline: 27763782]
31. Hashizaki M, Nakajima H, Shiga T, Tsutsumi M, Kume K. A longitudinal large-scale objective sleep data analysis revealed a seasonal sleep variation in the Japanese population. Chronobiol Int 2018;35(7):933-945 [FREE Full text] [doi: 10.1080/07420528.2018.1443118] [Medline: 29589960]
32. Hand AJ, Stone JE, Shen L, Vetter C, Cain SW, Bei B, et al. Measuring light regularity: sleep regularity is associated with regularity of light exposure in adolescents. Sleep 2023;46(8):zsad001 [FREE Full text] [doi: 10.1093/sleep/zsad001] [Medline: 36625482]
33. Harrison Y. The impact of daylight saving time on sleep and related behaviours. Sleep Med Rev 2013;17(4):285-292. [doi: 10.1016/j.smrv.2012.10.001] [Medline: 23477947]
34. Ho FYY, Poon CY, Wong VWH, Chan K, Law K, Yeung W, et al. Actigraphic monitoring of sleep and circadian rest-activity rhythm in individuals with major depressive disorder or depressive symptoms: a meta-analysis. J Affect Disord 2024;361:224-244. [doi: 10.1016/j.jad.2024.05.155] [Medline: 38851435]
35. Robbins R, Affouf M, Seixas A, Beaugris L, Avirappattu G, Jean-Louis G. Four-year trends in sleep duration and quality: a longitudinal study using data from a commercially available sleep tracker. J Med Internet Res 2020;22(2):e14735. [doi: 10.2196/14735] [Medline: 32078573]
36. Walch OJ, Cochran A, Forger DB. A global quantification of "normal" sleep schedules using smartphone data. Sci Adv 2016;2(5):e1501705 [FREE Full text] [doi: 10.1126/sciadv.1501705] [Medline: 27386531]
37. Aledavood T, Luong N, Baryshnikov I, Darst R, Heikkilä R, Holmén J, et al. Multimodal digital phenotyping study in patients with major depressive episodes and healthy controls (mobile monitoring of mood): observational longitudinal study. JMIR Ment Health 2025;12:e63622 [FREE Full text] [doi: 10.2196/63622] [Medline: 39984168]
38. Ikäheimonen A, Luong N, Baryshnikov I, Darst R, Heikkilä R, Holmen J, et al. Predicting and monitoring symptoms in patients diagnosed with depression using smartphone data: observational study. J Med Internet Res 2024;26:e56874 [FREE Full text] [doi: 10.2196/56874] [Medline: 39626241]
39. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. Front ICT 2015;2. [doi: 10.3389/fict.2015.00006]

XSL•FO
RenderX

40. Baryshnikov I, Aledavood T, Rosenström T, Heikkilä R, Darst R, Riihimäki K, et al. Relationship between daily rated depression symptom severity and the retrospective self-report on PHQ-9: a prospective ecological momentary assessment study on 80 psychiatric outpatients. J Affect Disord 2023;324:170-174 [FREE Full text] [doi: 10.1016/j.jad.2022.12.127] [Medline: 36586594]
41. SCA11H OPERATION MODES. URL: https://www.murata.com/-/media/webrenewal/products/sensor/accel/sca10h_11h/product-specification-1324-rev1-sca11h-operation-modes-eng-id-15741.ashx?la=en-us&cvid=20160209090049206400 [accessed 2025-11-25]
42. Aledavood T, Hoyos AMT, Alakörkkö T, Kaski K, Saramäki J, Isometsä E, et al. Data collection for mental health studies through digital platforms: requirements and design of a prototype. JMIR Res Protoc 2017;6(6):e110 [FREE Full text] [doi: 10.2196/resprot.6919] [Medline: 28600276]
43. Ikäheimonen A, Triana AM, Luong N, Ziaei A, Rantaharju J, Darst R, et al. Niimpy: a toolbox for behavioral data analysis. SoftwareX 2023;23:101472. [doi: 10.1016/j.softx.2023.101472]
44. Aledavood T, Kivimäki I, Lehmann S, Saramäki J. Quantifying daily rhythms with non-negative matrix factorization applied to mobile phone data. Sci Rep 2022;12(1):5544 [FREE Full text] [doi: 10.1038/s41598-022-09273-y] [Medline: 35365710]
45. Aledavood T, Lehmann S, Saramäki J. Social network differences of chronotypes identified from mobile phone data. EPJ Data Sci 2018;7(1):46. [doi: 10.1140/epjds/s13688-018-0174-4]
46. Roenneberg T, Allebrandt KV, Merrow M, Vetter C. Social jetlag and obesity. Curr Biol 2012;22(10):939-943 [FREE Full text] [doi: 10.1016/j.cub.2012.03.038] [Medline: 22578422]
47. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-310. [Medline: 2868172]
48. Kwak SG, Park SH. Normality test in clinical research. J Rheum Dis 2019;26(1):5. [doi: 10.4078/jrd.2019.26.1.5]
49. Yang S, Berdine G. Normality tests. The Southwest Respir Med Crit Care Chronicles 2021;9(37):87-90. [doi: 10.12746/swrccc.v9i37.805]
50. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012;24(3):69-71 [FREE Full text] [Medline: 23638278]
51. Galecki A, Burzykowski T. Linear Mixed-Effects Models Using R: A Step-by-Step Approach. New York: Springer New York; 2013:245-273.
52. Meteyard L, Davies RAI. Best practice guidance for linear mixed-effects models in psychological science. J Mem Lang 2020;112:104092. [doi: 10.1016/j.jml.2020.104092]
53. R: A language and environment for statistical computing [technical report]. URL: https://www.r-project.org/ [accessed 2025-11-28]
54. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
55. Rossum G, Drake FL. No Pt 2 in Python documentation manual. The Python Language Reference. 2010. URL: https://docs.python.org/3/reference/ [accessed 2025-11-28]
56. Łojko D, Rybakowski JK. Atypical depression: current perspectives. Neuropsychiatr Dis Treat 2017;13:2447-2456 [FREE Full text] [doi: 10.2147/NDT.S147317] [Medline: 29033570]
57. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med 2001;2(5):389-396. [doi: 10.1016/s1389-9457(00)00098-8] [Medline: 14592388]
58. Ciman M, Wac K. Smartphones as sleep duration sensors: validation of the iSenseSleep Algorithm. JMIR Mhealth Uhealth 2019;7(5):e11930 [FREE Full text] [doi: 10.2196/11930] [Medline: 31115341]
59. Natale V, Drejak M, Erbacci A, Tonetti L, Fabbri M, Martoni M. Monitoring sleep with a smartphone accelerometer: monitoring sleep with a smartphone. Sleep and Biological Rhythms 2012;10(4):292. [doi: 10.1111/j.1479-8425.2012.00575.x]
60. Maynard T, Appleman E, Cronin-Golomb A, Neargarder S. Objective measurement of sleep by smartphone application: comparison with actigraphy and relation to self-reported sleep. Exploration of Targeted Anti-tumor Therapy 2021:382-391. [doi: 10.37349/emed.2021.00057]
61. Staples P, Torous J, Barnett I, Carlson K, Sandoval L, Keshavan M, et al. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. NPJ Schizophr 2017;3(1):37 [FREE Full text] [doi: 10.1038/s41537-017-0038-0] [Medline: 29038553]
62. Piantino J, Luther M, Reynolds C, Lim MM. Emfit bed sensor activity shows strong agreement with wrist actigraphy for the assessment of sleep in the home setting. Nat Sci Sleep 2021;13:1157-1166 [FREE Full text] [doi: 10.2147/NSS.S306317] [Medline: 34295199]
63. Gradisar M, Wolfson AR, Harvey AG, Hale L, Rosenberg R, Czeisler CA. The sleep and technology use of Americans: findings from the National Sleep Foundation's 2011 Sleep in America Poll. J Clin Sleep Med 2013;9(12):1291-1299 [FREE Full text] [doi: 10.5664/jcsm.3272] [Medline: 24340291]
64. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep 2021;44(2):zsaa169. [doi: 10.1093/sleep/zsaa169] [Medline: 32886772]

XSL•FO
RenderX

65. Fornaro M, Orsolini L, Marini S, De Berardis D, Perna G, Valchera A, et al. The prevalence and predictors of bipolar and borderline personality disorders comorbidity: systematic review and meta-analysis. J Affect Disord 2016;195:105-118. [doi: 10.1016/j.jad.2016.01.040] [Medline: 26881339]
66. Socada JL, Söderholm JJ, Rosenström T, Ekelund J, Isometsä E. Presence and overlap of bipolar symptoms and borderline features during major depressive episodes. J Affect Disord 2021;280(Pt A):467-477. [doi: 10.1016/j.jad.2020.11.043] [Medline: 33246197]
67. Gold AK, Sylvia LG. The role of sleep in bipolar disorder. Nat Sci Sleep 2016;8:207-214 [FREE Full text] [doi: 10.2147/NSS.S85754] [Medline: 27418862]

Abbreviations

BD: bipolar disorder
BPD: borderline personality disorder
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
EMA: ecological momentary assessment
LoA: limit of agreement
MDD: major depressive disorder
MEQ: Morning-Eveningness Questionnaire
PSG: polysomnography
TST: total sleep time

Edited by D Majekodunmi, G Badicu; submitted 15.08.25; peer-reviewed by L Cerqueira, C Simons; comments to author 25.10.25; accepted 22.11.25; published 11.12.25.

Please cite as:
Mahir A, Luong N, Baryshnikov I, Martikkala A, Isometsä E, Aledavood T
Multimodal Sleep Measurement and Alignment Analysis in Outpatients With Major Depressive Episode: Observational Study
JMIR Mhealth Uhealth 2025;13:e82465
URL: https://mhealth.jmir.org/2025/1/e82465
doi:10.2196/82465
PMID:

©Afrooz Mahir, Nguyen Luong, Ilya Baryshnikov, Annasofia Martikkala, Erkki Isometsä, Talayeh Aledavood. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Multiple Technology–Based Physical Activity Intervention for Latina Adolescents: Results From the Chicas Fuertes Randomized Controlled Trial

Jacob R Carson[1], MPH; Emily Greenstadt[1], MA, MPH; Brittany Olivera[1], MPH; Shira Dunsiger[2], PhD; Michelle Zive[1], PhD; Michael Higgins[3], MS; Job Godino[1,4], PhD; Bess Marcus[2], PhD; Dawn Meyer[5], PhD; Britta Larsen[1], PhD

[1]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA, United States

[2]Department of Behavioral and Social Sciences, Brown University, Providence, RI, United States

[3]Exercise and Physical Activity Resource Center, University of California, San Diego, La Jolla, CA, United States

[4]Laura Rodriguez Research Institute, Family Health Centers of San Diego, San Diego, CA, United States

[5]Department of Neurosciences, School of Medicine, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Jacob R Carson, MPH
Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA, United States

Abstract

Background: Latina adolescents report low levels of moderate-vigorous physical activity (MVPA) and high lifetime risk of lifestyle-related diseases. There is a lack of MVPA interventions targeted at this demographic despite documented health disparities. Given their high rates of mobile technology use, interventions delivered through mobile devices may be effective for this population.

Objective: This paper examines the efficacy of the Chicas Fuertes intervention in increasing MVPA across 6 months in Latina adolescents.

Methods: Participants were Latina adolescents (aged 13 - 18 years) in San Diego County who reported being underactive (<150 min/wk of MVPA). All participants received a wearable fitness tracker (Fitbit Inspire HR); half were randomly assigned to also receive the multimedia intervention. Intervention components included a personally tailored website, personalized texting based on Fitbit data, and social media. The primary outcome was change in minutes of weekly MVPA from baseline to 6 months, measured by ActiGraph accelerometers and the 7-Day Physical Activity Recall Interview. Changes in daily steps using Fitbit devices were also examined to test intervention efficacy.

Results: Participants (N=160) were 15.85 (SD 1.71) years old on average, and mostly second generation in the United States. For ActiGraph-measured MVPA, participants in the intervention group (n=83) increased from a median of 0 (IQR 0-24) minutes/week at baseline to 64 (IQR 19-72) minutes/week at 6 months compared to control participants, who showed increases from a median of 0 (IQR 0-26) at baseline to 41 (IQR 7-76) minutes/week at 6 months (P=.04). Self-reported MVPA increased in the intervention group from a median of 119 (IQR 62.5-185) minutes/week at baseline to 147 (IQR 96-181) minutes/week at 6 months compared to control participants, who showed increases from a median of 120 (IQR 48.8-235) at baseline to 124 (IQR 100-169) minutes/week at 6 months (P=.03). Steps also increased in both groups, with the intervention group showing significantly greater increases (P=.03).

Conclusions: This intervention was successful in using a tailored technology-based strategy to increase MVPA in Latina adolescents and provides a promising approach for addressing a key health behavior. Given the scalable technology used, future studies should focus on broad-scale dissemination to address health disparities.

Trial Registration: ClinicalTrials.gov NCT04190225; https://clinicaltrials.gov/study/NCT04190225

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-022-06105-2

(JMIR Mhealth Uhealth 2025;13:e71623) doi:10.2196/71623

KEYWORDS

mHealth; health disparities; Fitbit; Latina adolescents; physical activity; mobile health

XSL•FO
RenderX

Introduction

Physical activity (PA) declines by as much as 90% between early childhood and adolescence, particularly in girls [1]. Additionally, adolescent girls of racial and ethnic minorities have the lowest PA rates of any demographic group in the United States [1-3]. According to the National Health and Nutrition Examination Survey, only 2.9% of Mexican American girls meet the national guidelines of 60 minutes of moderate-vigorous physical activity (MVPA) per day [1]. Increasing MVPA in Latina adolescents could improve physical, psychosocial, and cognitive health [4] while producing lifelong habits that mitigate obesity, type 2 diabetes, and other chronic illnesses that disproportionately impact Latina adults [5-7]. To date, no large-scale interventions focused on increasing MVPA in Latina adolescents are documented in the scientific literature.

Previous literature has identified that psychosocial variables may be responsible for PA disparities in Latina teens [8-16]. Adolescent girls frequently report lower self-efficacy, social support, and behavioral capability than boys [8,14-17], and Latina teens specifically report lower positive support and more negative support than non-Latina White girls [10,11]. An intervention grounded in theory to promote psychosocial constructs such as self-efficacy, social support, and perceived competence could be effective for increasing MVPA. Additionally, a previous meta-analysis found the strongest effects on PA for multicomponent interventions that were specifically targeted at adolescent girls (as opposed to adolescents generally) [18]. To date, employment of these intervention strategies in cohorts of Latina adolescents has been limited, specifically interventions that deliver via newer technologies [19].

We conducted a pilot study, Niñas Saludables, to test a web-delivered theory-based MVPA intervention for Latina teens, adapted from an intervention shown to be effective with Latina adults [20]. Self-reported MVPA increased significantly, and the intervention was generally well-received. In follow-up interviews, however, participants highlighted a desire for more visual intervention content in smaller, more frequent doses and the use of more mobile technology channels. This was corroborated in additional focus groups and design workshops, in which Latina teens highlighted a desire for an intervention designed specifically for them, along with use of wearable trackers and social media [21].

The aim of the Chicas Fuertes trial was to test the efficacy of a theory-based, individually tailored, multitechnology intervention for increasing MVPA in Latina adolescents compared to a control group receiving a wearable tracker. Descriptions of the intervention development, study protocol, recruitment strategies, and baseline findings have been published previously [22]. This paper presents the main findings for the Chicas Fuertes trial, which examined changes in MVPA from baseline to 6 months measured by the ActiGraph accelerometer and 7-Day Physical Activity Recall (PAR) Interview, and steps measured by the Fitbit Inspire HR monitor.

Methods

Overall Design

The Chicas Fuertes trial tested an individually tailored, multitechnology intervention aimed at promoting MVPA among Latina adolescents (N=160) in San Diego County. Promotion of MVPA over total activity or light physical activity (LPA) was prioritized to align with current PA guidelines [23]. The intervention is based on multiple behavior change theories, including social cognitive theory [24], the transtheoretical model [25,26], Control Theory, and the Behavior Change Techniques taxonomy [27]. Components of theory integrated into the intervention design included goal setting, self-monitoring, problem-solving barriers, increasing social support, social norms, and rewarding oneself for meeting goals. Intervention components were delivered via multiple media channels over 6 months. The primary outcome was change in minutes of weekly MVPA from baseline to 6 months, measured by ActiGraph accelerometers (objective) and the 7-Day PAR Interview (self-report). Psychosocial and environmental variables were also measured at baseline and follow-up.

Study Population

To participate, individuals had to self-identify as Latina; be 13 - 18 years old; live in San Diego County; read, write, and speak in English; and be underactive (ie, self-reporting less than 150 weekly minutes of MVPA). Additionally, participants needed regular (≥2 times/wk) access to the internet and access to a cell phone; participants who did not have a personal smartphone were provided additional equipment to sync their Fitbit to a computer and access Fitbit through a website instead of using the smartphone app. Participants were excluded if they self-reported a serious heart condition, chest pain during exercise, pregnancy, uncontrolled asthma, or any other serious medical condition that would make unsupervised PA unsafe, unless they provided written permission to participate from their health care provider. Strategies for recruitment, randomization, allocation, and blinding have been published previously [22,28].

Measures

Primary Outcome: PA

ActiGraph Accelerometer

MVPA was measured objectively at baseline (prior to randomization) and 6-month follow-up using the ActiGraph GT3X+ accelerometer (ActiGraph, LLC). The ActiGraph GT3X+ is a hip-worn triaxial accelerometer worn on the hip that measures the movement and activity intensity. It has been validated against heart rate telemetry [29] and total energy expenditure [30], including in children [31]. Accelerometers were delivered via mail with detailed instructions to wear the accelerometer around their waist (as a belt) on their left-side hip for an entire day (approximately 12 hours) for 7 consecutive days. Throughout the week, participants received SMS reminders to wear the ActiGraph. Participants were provided with a prepaid and preaddressed envelope to mail back the accelerometer after use, at which time it was checked for sufficient wear time. Valid wear time was defined as wearing the device for at least 600 minutes per day for 5 days or a total

XSL•FO
RenderX

of at least 3000 minutes over 4 days, and participants were instructed to rewear the device if they did not meet the criteria. Activity was considered part of the total minutes per week only if it occurred in bouts of 10 minutes or longer. The same procedure was followed for the 6-month follow-up visit.

7-Day PAR Interview

The 7-Day PAR, a semistructured interview-administered instrument that assesses the frequency, duration, and intensity of MVPA, served as an additional primary outcome at baseline and 6-month follow-up [32]. The 7-Day PAR Interview asks participants about their activities over the past week and their estimates of their weekly PA in minutes. To improve recall accuracy, it breaks down activities by time of day, asks about the intensity and the purpose of the activity (eg, transportation, leisure, occupation, or school). It has been validated in children as young as age 11 years [33], and consistently shown acceptable validity and internal consistency [32,34], as well as sensitivity to changes over time [35,36]. The 7-Day PAR also measures sleeping time, work hours, and school hours to enhance accuracy. After wearing the accelerometer for a full 7 days (with valid wear time), participants were asked to complete the 7-Day PAR Interview to coincide with the accelerometer wear. All 6-month follow-up PAR interviews were conducted via Zoom (Zoom Communications) video.

Fitbit Inspire HR

Each participant enrolled was provided a Fitbit Inspire HR to be worn daily throughout the study period. Fitbit monitors were worn on the nondominant wrist and continuously captured day-level (steps) and minute-level (activity intensity and heart rate) data, in addition to sleep. The Fitbit app (Fitabase) automatically syncs with the monitor and stores historical data and offers a user-friendly interface to review activity and set goals for the participant. The Fitbit has excellent validity for measuring steps in free-living conditions [37]. Fitbit automatically classifies the intensity of activity based on movement and heart rate. Classification of MVPA has shown moderate validity, and recent studies have shown a high correlation between MVPA measured by the Fitbit and ActiGraph accelerometers [38]. To prevent feedback from influencing baseline activity, participants' Fitbits were initially blinded with tape over the display and no activity information was accessible on the Fitbit app.

Each participant's Fitbit accounts were linked to Fitabase (Small Steps Labs LLC), a third-party research platform that collects continuous Fitbit data whenever data are synced with the Fitbit app. An adherent day was characterized by having either 10 or more hours of heart rate data or 6000 or more steps [22,39]. Participant-level data for analysis were derived by averaging daily summaries across all adherent days.

Secondary Outcomes and Covariates

Prior to their baseline and 6-month visits, participants received a personalized link to complete questionnaires directly into REDCap (Research Electronic Data Capture; Vanderbilt University) software, a Health Insurance Portability and Accountability Act (HIPAA)-compliant clinical trials software database. They were sent 3 reminder emails if the surveys were not completed by the time of the follow-up visit.

The 5-item Self-Efficacy for Physical Activity Scale was used to measure self-efficacy to become physically active across diverse contexts (α=0.82) [40]. We also administered the Social Support for Exercise Scale, which is a 13-item questionnaire with the 3 subscales of family, friends, and rewards and punishments focused on social support in the last 3 months (α=0.61 - 0.91) [41]. To assess enjoyment for PA, the 18-item Physical Activity Enjoyment Scale (PACES) is used to evaluate the level of personal satisfaction from PA (α=0.96) [42]. Neighborhood environment is measured using the Neighborhood Environment Walkability Scale for Youth (NEWS-Y) and the Physical Activity Neighborhood Environment Scale (PANES). The NEWS-Y measures perceptions of neighborhood environment across 67 items across 9 subscales, including recreation facility availability, pedestrian traffic safety, and walking facilities having shown acceptable internal consistencies and correlations with objectively measured PA [43]. The PANES is a 9-item questionnaire including Likert responses to questions about walking, cycling, and safety features to indicate the neighborhood environment of participants [44]. Finally, participants also filled out the Center for Epidemiological Studies Depression Scale (CES-D) at baseline and the 6-month follow-up visit. The CES-D is a 10-item questionnaire related to depression symptoms experienced by the participant over the past week [45].

Two questionnaires were administered to everyone at baseline and monthly to the intervention group to develop tailored content for the website (see "Tailored Mobile Health PA Intervention (Intervention Group)"). Stages of Change for Physical Activity is a 5-item survey that assesses motivational readiness for adopting a consistent PA routine. Participants provide "Yes" or "No" responses to statements like, "I intend to become more physically active in the next 6 months." [46]. Based on their responses, participants are classified as being in precontemplation, contemplation, preparation, action, or maintenance. Processes of Change for Physical Activity contains 40 items across 10 subscales assessing processes related to MVPA change that are either cognitive (eg, learning about the benefits of PA) or behavioral (eg, rewarding oneself for meeting goals) [46].

Protocols

Overview

Once participants completed all baseline questionnaires and sufficient wear time was confirmed for the ActiGraph, the Fitbit monitors and apps were unblinded, and they were randomized 1:1 to a study condition. Randomization was done using REDCap clinical trials software, with participants stratified based on baseline stage of change to ensure equal distribution of motivational readiness for PA adoption across conditions. The randomization protocol was based on a permuted block-randomization protocol, with small, random-sized blocks. Friends or relatives were yoked into conditions together, with yoked groups distributed equally across conditions.

XSL•FO
RenderX

Tailored Mobile Health PA Intervention (Intervention Group)

A detailed description of the intervention has been published previously [28]. Summaries of the intervention components are included below.

Coaching session

The intervention consisted of a baseline counseling session (optional in-person or virtual) to teach key behavior change strategies, followed by theory-based components delivered via a range of remote technology channels over the next 6 months (details have been published previously) [28]. During the baseline visit, participants received a one-on-one health coaching session with a trained interventionist where they learned to set personalized activity goals. The coaching was grounded in motivational interviewing, and participants anticipated barriers and engaged in problem-solving techniques.

Study website

Participants in the intervention were also given an account on the study website, which included a range of personalized features. Participants entered weekly goals and scheduled their weekly activities. Their Fitbit data were synced with the website, which produced a calendar of their past activity and compared their activity with their set goals. Participants also filled out monthly questionnaires regarding their activity and psychosocial factors, which produced an individually tailored report on their activity progress and strategies for behavior change. Nonpersonalized features of the website included links to workout videos, solutions to common barriers, a list of places to be active including low-cost gyms, an interactive message board, and an activity leader board with the top 3 participants for the week based on current Fitbit data.

Text messages

Individually tailored automated algorithm-driven text messages based on participants' Fitbit data were sent 4 times per week that updated them on the progress toward their goals and provided summaries of their MVPA for the week. Texts also prompted interactive goal setting and provided participants with goal progress feedback. Those who met their goals were congratulated and encouraged to continue increasing their goal; those who were not on track to meet their goal were given a Behavior Change Technique, such as goal setting, to try. Tailored texts also incorporated the names of friends and family and local places to be active.

Social Media

Participants were also invited to follow the intervention's private Instagram account. Daily posts incorporated different theoretical constructs of behavior change from social cognitive theory and Behavior Change Techniques. Instagram stories were also used to prompt wearing and charging Fitbits.

Social media use was not required. Instagram posts were also shown on the study website for individuals who did not use social media.

At 1 week and 1 month postbaseline, a trained interventionist conducted a 10-minute phone call with participants to address any questions or concerns regarding the media channels or barriers to activity.

Fitbit Only Group (Control Group)

The control group was told to continue wearing their Fitbit daily, and charge and sync regularly. Reminders to sync were sent via text message if a participant had not synced in 5 days. Participants in the control group received a US $10 Amazon gift card each month they remembered to wear, sync, and charge their Fitbit.

Analysis

As a preliminary step, the study sample was summarized with respect to baseline sociodemographics and PA rates and compared between groups using *t* tests, chi-square tests, and nonparametrics as appropriate. Both participant and parent characteristics at baseline are reported for sociodemographics.

As minutes/week of MVPA were skewed and attempts to transform toward normality were unsuccessful, a series of quantile regression models with bootstrapped SEs (10,000 bootstrapped samples) were used to regress median minutes/week of MVPA on baseline values, group, and covariates. Models were run separately for self-reported MVPA (via the PAR) and objectively measured (using accelerometry). Models for the latter further adjusted for accelerometer wear time. As a small proportion of study participants were yoked together during the randomization process (siblings, cousins, or close friends who enrolled together), we adjusted for clustering by yoked group in each of our models (to ensure SEs were adjusting for nonindependence of outcomes within cluster). All other covariates were balanced at baseline and tested for significance in the model before being dropped.

To understand the group differences in minutes of PA as collected via the Fitbit, we examined data from the weeks corresponding to measurement of PA via the PAR and accelerometer. Using a longitudinal mixed effects model with subject-specific intercepts, we examined group differences adjusting for clustering by yoked pair.

All analyses were carried out in R (R Core Team), with significance level set at .05 a priori.

Ethical Considerations

This study is approved by the University of California, San Diego Human Research Protections Programs (protocol #182070). Written, informed consent and parental permission and adolescent assent to participate are obtained from all participants. All data were deidentified for analyses. Fitbit devices and incentives in the form of cash payments and gift cards are provided to all participants, equating to US $25, US $50, and US $50 for baseline, 6-month, and 12-month follow-up measurement visits with an additional US $25 for each follow-up visit where all measures were completed.

Results

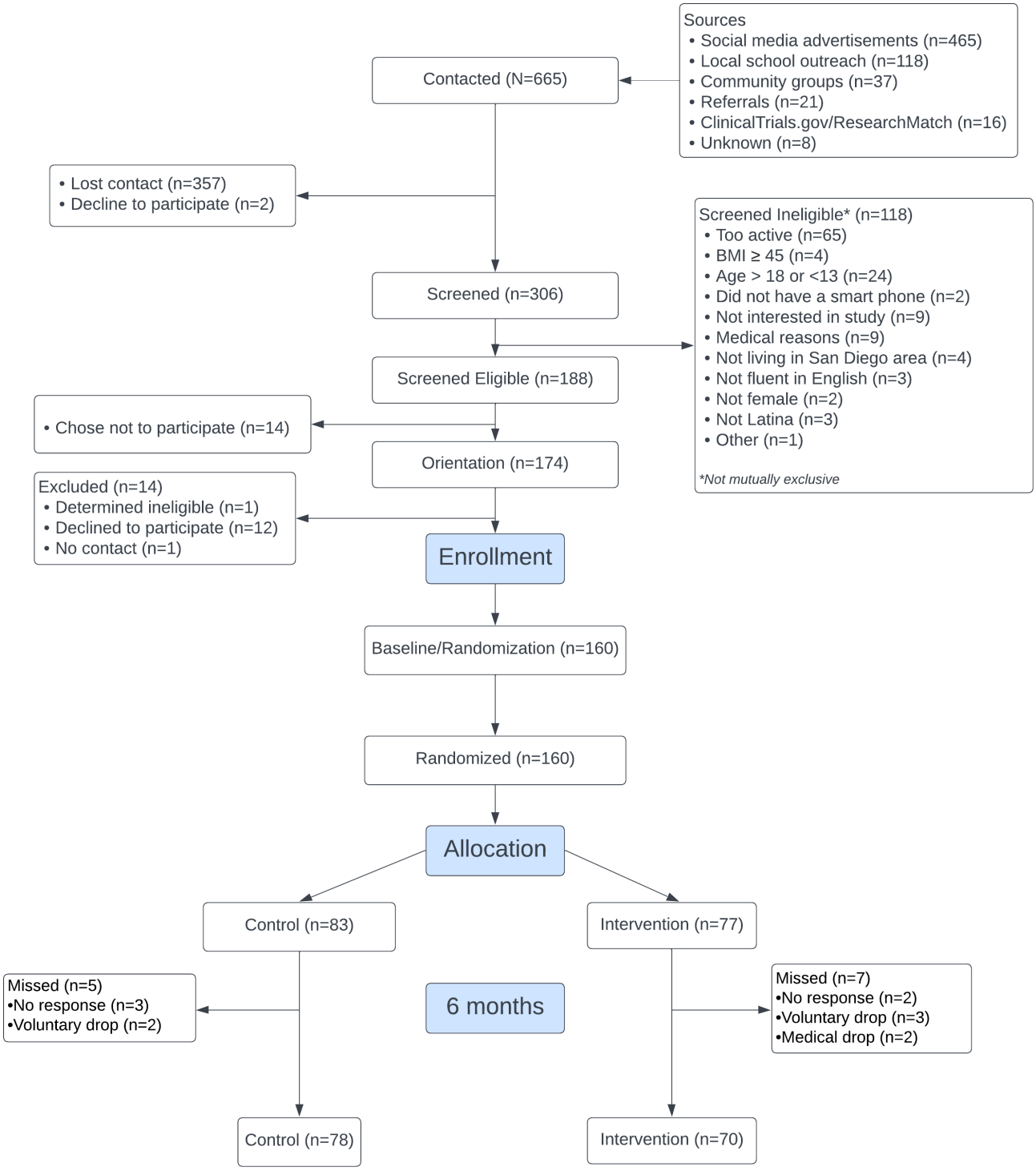
The study sample has been described in a previous publication [22]. In brief, participants were 15.85 years of age on average (SD 1.71), with an average BMI of 25.78 (SD 6.07). About half

XSL•FO
RenderX

(84/160, 53%) reported speaking more English than Spanish, with 69% (110/160) of the aggregate sample reporting second-generation status. When considering motivation to become physically active, 84% (134/160) were considered in preparation. Participant retention at 6 months was high (148/160, 93%), with reasons for loss to follow-up depicted in Figure 1. More information on recruitment and challenges due to the COVID-19 pandemic is published elsewhere [22].

The parents' sociodemographic data indicated that about half (86/160, 54%) of parents had a high school diploma, General Education Development (GED) equivalent, or less than high school level education, and about 25% of the sample reported an average household income of at least US $75,000. There were no between-group differences in baseline sociodemographics or psychosocial variables (all P>.05; see Table 1).

Figure 1. Six-month CONSORT (Consolidated Standards of Reporting Trials) diagram.



XSL•FO
RenderX

Table . Baseline characteristics of study sample by group (N=160).

Variables	Intervention (n=77)	Control (n=83)
Participant characteristics		
Age (years), mean (SD)	15.85 (1.71)	15.90 (1.50)
BMI, mean (SD)	26.16 (5.49)	25.24 (6.57)
Race, n (%)		
White	35 (45.5)	33 (39.8)
Black	1 (1.3)	0 (0.0)
Asian	1 (1.3)	1 (1.2)
American Indian/Alaska Native	1 (1.3)	2 (2.4%
Pacific Islander or Hawaiian	0 (0.0)	0 (0.0)
Multiracial	8 (10.4)	4 (48)
Other	31 (40.3)	43 (51.8)
Primary language, n (%)		
Spanish preferred over English	4 (5.2)	6 (7.2)
Spanish and English equally	21 (27.3)	31 (37.3)
English preferred over Spanish	47 (61.0)	37 (44.6)
Only English	5 (6.5)	9 (10.8)
Generational status, n (%)		
First generation	6 (7.8)	7 (8.4)
Second generation	53 (68.8)	57 (68.7)
Third generation	18 (23.4)	19 (22.9)
Parent characteristics		
Education, n (%)		
<High school	26 (33.8)	25 (30.1)
High school or GED	16 (20.8)	19 (22.9)
Some college or associate degree	14 (18.2)	18 (21.7)
College graduate	13 (16.9)	10 (12.0)
Master’s degree	8 (10.4)	8 (9.6)
Doctoral degree	0 (0)	3 (3.6)
Income (US dollar), n (%)		
<$11,999	12 (15.6)	22 (26.5)
$12,000-$24,999	12 (15.6)	11 (14.3)
$25,000-$49,999	21 (27.3)	21 (27.3)
$50,000-$99,999	21 (27.3)	18 (21.6)
$100,000+	11 (14.3)	11 (14.3)
Number of children in home, mean (SD)	2.01 (1.18)	1.93 (1.20)
Psychosocial constructs		
Neighborhood Environment (NEWS-Y)		
Land-use diversity, mean (SD)	2.47 (0.62)	2.37 (0.81)
Recreation, mean (SD)	2.27 (0.68)	2.21 (0.76)
Residential density, n (%)	112 (39.1)	109 (40.3)
Land-use access, mean (SD)	2.94 (0.50)	3.01 (0.50)
Street connectivity, mean (SD)	2.83 (0.59)	2.84 (0.57)

XSL•FO
RenderX

Variables	Intervention (n=77)	Control (n=83)
Walk or cycle facilities, mean (SD)	3.07 (0.59)	3.05 (0.66)
Neighborhood aesthetics, mean (SD)	2.74 (0.66)	2.89 (0.73)
Safety from traffic, mean (SD)	2.78 (0.45)	2.79 (0.45)
Safety from crime, mean (SD)	2.26 (0.78)	2.19 (0.75)
Neighborhood Environment (PANES), mean (SD)	4.96 (1.16)	4.59 (1.38)
PA Enjoyment (PACES), mean (SD)	4.67 (0.92)	5.03 (0.93)
Depressive symptoms (CES-D), mean (SD)	10.62 (3.87)	11.14 (4.70)
Exercise self-efficacy, mean (SD)	1.98 (0.54)	2.03 (0.46)
Friend social support for exercise, mean (SD)	20.5 (13.4)	24.2 (16.7)
Maternal support for exercise	2.46 (0.75)	2.55 (0.64)
Stage of change, n (%)		
Precontemplation	3 (3.9)	1 (1.2)
Contemplation	10 (13.0)	12 (14.5)
Preparation	64 (83.1)	70 (84.3)
Physical activity		
MVPA[b] (min/wk), median (IQR)		
7-Day PAR[a,c]	119 (62.5-185)	120 (48.8-235)
ActiGraph GT3X+[a]	0 (0-24)	0 (0-26)
Total steps[a], median (IQR)	5196 (3318-6634)	5175 (3536.12-6727.12)

[a]Variable had a skewed distribution. There were no significant between-group differences.

[b]MVPA: moderate-vigorous physical activity.

[c]PAR: Physical Activity Recall.

Figures 2 and 3 depict changes over time by group in unadjusted median minutes/week of MVPA from the PAR and accelerometer, respectively. The median minutes/week of self-reported MVPA among intervention participants increased from 119 (IQR 62.5-185) minutes/week at baseline to 147 (IQR 96-181) minutes/week at 6 months. Control participants showed increases from a median of 120 (IQR 48.8-235) at baseline to 124 (IQR 100-169) minutes/week at 6 months. Adjusted model results indicate intervention participants did an additional 23.5 minutes/week of MVPA at 6 months compared to control (95% CI 8.35 - 32.76). Via accelerometry, intervention participants showed increases in minutes/week of MVPA from a median of 0 (IQR 0-24) minutes/week at baseline to 64 (IQR 19-72) minutes/week at 6 months compared to control participants, who showed increases from a median of 0 (IQR 0-26) at baseline to 41 (IQR 7-76) minutes/week at 6 months. Adjusted model results indicate intervention participants were doing an additional 18.4 minutes/week of MVPA at 6 months compared to control (95% CI 4.67 - 37.28).

Finally, we considered daily steps (as measured by the Fitbit), restricted to weeks in which PAR and accelerometer data were also being collected (for comparability). At baseline, among the aggregated sample, the median daily steps were 5196 (IQR 3318-6634) among intervention participants and 5175 (IQR 3536.12-6727.12) among controls. Steps/day significantly increased in both arms with greater increases among intervention participants relative to control (β=101.25, SE=46.31; *P*=.03) suggesting intervention participants recorded an additional 101 steps/day compared to control at 6 months.

XSL•FO
RenderX

Figure 2. Between-group changes over time in moderate-vigorous physical activity (Physical Activity Recall).

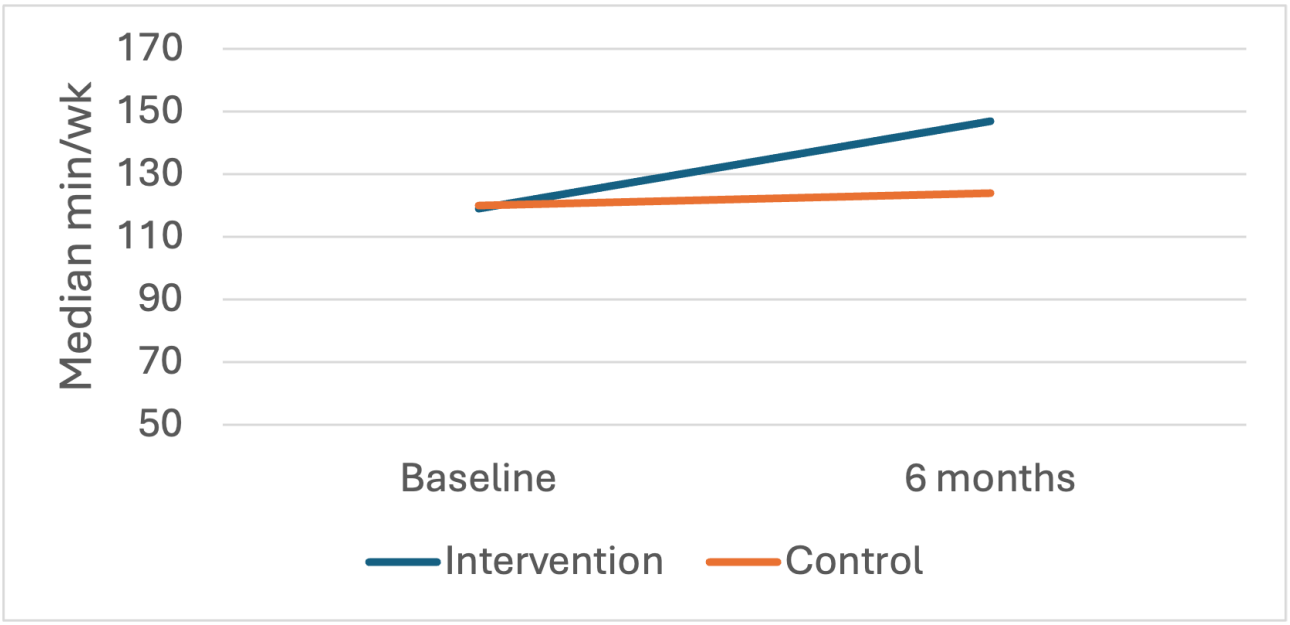
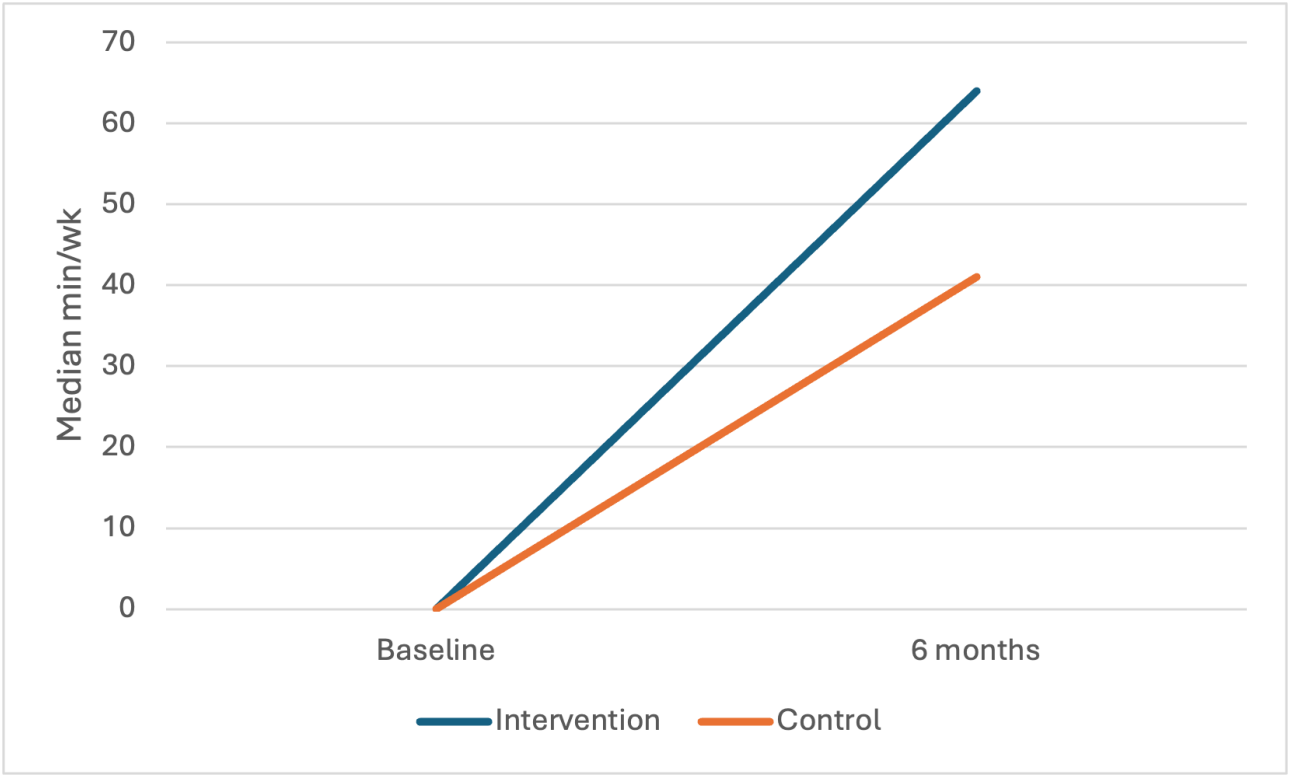


Figure 3. Between-group changes over time in moderate-vigorous physical activity (accelerometer).



Discussion

Principal Results

The aim of the Chicas Fuertes trial was to test the efficacy of a theory-based, individually tailored, multitechnology intervention for increasing MVPA in Latina adolescents. The results at 6 months presented here show statistically significant differences in changes to PA from baseline favoring the intervention group across objectively measured MVPA, self-reported MVPA, and daily steps. While both intervention and control conditions showed an increase in objectively measured PA from baseline, the intervention condition participated in an additional 18.4 minutes/week compared to the control condition. This effect was consistent with the self-reported data, with the intervention group reporting an additional 23.5 minutes/week compared to the controls, reinforcing the robustness of our findings.

Comparison With Prior Work

Consistent with the existing body of literature [47], self-reported MVPA was consistently higher than objectively measured methods. While self-reported MVPA was higher at all time points for both the intervention and control groups, the magnitude of change measured by the ActiGraph was

XSL•FO
RenderX

considerably larger, mostly due to the discrepancy between self-reported and objectively measured MVPA at baseline. This could suggest that participants did not have a clear understanding of what MVPA intensity felt like at baseline. This interpretation is consistent with previous research from this group that found significantly lower self-reported MVPA (which better correlated with Actigraph measured MVPA) when participants first engaged in a 10-minute demonstration walk at a moderate pace prior to conducting the PAR at baseline measurements [48]. Participants in this trial may have developed a better sense of MVPA over the course of the intervention; however, the simultaneous rise in MVPA minutes among the control condition suggests that access to the Fitbit and Fitbit App alone could be providing participants with better knowledge of MVPA.

The primary findings are consistent with previous research using similar intervention channels. Text-messaging interventions have been consistently effective at increasing PA in adults [49]. Specifically, meta-analyses suggested that interventions that used text-messaging in combination with other intervention components, such as in-person counseling sessions and websites, were more effective than those using text-messaging alone [49]. These findings are consistent with other research focused on adolescents with similar recommendations regarding the inclusion of goal setting and the combination with other web interfaces [50]. This paper is consistent with those findings by testing an approach that was both tailored to the participants and deployed multiple intervention components, such as coaching sessions, social media, and the website alongside text-messaging.

Despite health behavior and outcome disparities adversely impacting Latina adolescents, limited health behavior interventions have been explicitly designed for this population. This study begins to address that gap and builds on our previous results. Previous findings from our randomized controlled trials with adult Latina populations suggested that tailored technology-based interventions can be feasible and effective for increasing MVPA in adults [20,51-54]. The results here expand on this literature by further suggesting the efficacy of tailored technology-based interventions in Latina adolescents.

These findings are particularly significant given how sedentary the sample was at baseline, with a median of 0 (IQR 0-24) minutes/week of ActiGraph-measured MVPA. Increasing activity for individuals who were entirely sedentary, even if they do not meet PA guidelines, can have substantial impacts on their health risks [55]. Powell et al [55] posited that the benefits of activity are immediate and need no threshold to reduce the risk of all-cause mortality relative to those who are entirely inactive. Additionally, the greatest benefits to all-cause mortality risk are to those who go from no activity to any activity, further highlighting the importance of identifying interventions that are feasible and efficacious for inactive groups. As the most inactive demographic group, interventions that specifically improve health behaviors for Latina adolescents are needed to reduce health inequities and improve public health more generally [1]. This intervention provides evidence of effective, scalable, and contextually relevant strategies to increase MVPA among Latina adolescents.

Limitations

Generalizability of these findings may be limited. The design was specifically chosen to target Latina adolescents in largely urban San Diego County; this precludes assumptions of this specific intervention strategy among other demographics, particularly among more rural communities where there may be fewer PA resources, thereby reducing intervention efficacy. Nearly 84% of the participants recruited were preparing to engage in more PA, so the effect of this intervention may not be as robust in groups who are less motivated to be active in the first place. Finally, these analyses only focused on MVPA as opposed to including LPA. While this is a common intervention target (and aligns with the 7-Day PAR), these findings preclude the ability to identify how LPA may be important for participants who are trending from inactive to more active lifestyles.

Despite limitations, this study has numerous strengths, notably the randomized controlled trial design. It directly addresses a gap in the literature by testing a technology-based intervention to improve PA in Latina adolescents. The success of this intervention can be partially attributed to multiple steps of iterative codesign with the target population [22,28]. This study also had a high retention rate (93%), further signaling the acceptability and feasibility of our approach. Additionally, the control condition was an active comparator that received a Fitbit monitor, allowing us to compare our intervention to a commercially available program. Our approach used multiple validated measures of PA providing robustness for our primary accelerometer-based findings. Finally, the study was conducted during the COVID-19 pandemic, showing that the intervention can be effective even amid marked interruptions to daily routines.

Conclusions

This intervention was successful in using a tailored technology-based strategy to increase MVPA in Latina adolescents and provides a promising and scalable method for addressing a key health behavior. This study was co-designed with the target population, intentionally using digital media channels with the potential for broad scale-up and remote delivery. Future studies should examine strategies for dissemination and implementation, including partnering with educational and clinical organizations.

Acknowledgments

The use of Research Electronic Data Capture (REDCap) clinical trials software was supported through a grant to the Altman Clinical Research Institute (UL1TR001442). Generative artificial intelligence was not used in any portion of the manuscript.

XSL•FO
RenderX

Funding

This study is supported by a grant from the National Institutes of Health/National Institute of Nursing Research: R01NR017876. JRC's time was supported by 5T32HL079891-18.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to required practices at the time of funding, but deidentified data are available from the primary investigator on reasonable request.

Authors' Contributions

JRC: project administration, writing – original draft, writing – review & editing. EG: investigation, project administration, writing – original draft. BO: Investigation, project administration, writing – original draft. SD: formal analysis, software, visualization, writing – original draft. MZ: writing – review & editing. MH: data curation. JG: conceptualization, data curation. BM: conceptualization. DM: conceptualization. BL: conceptualization, funding acquisition, investigation, supervision, writing – original draft, writing – review & editing.

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 889 KB - mhealth_v13i1e71623_app1.pdf]

References

1. Whitt-Glover MC, Taylor WC, Floyd MF, Yore MM, Yancey AK, Matthews CE. Disparities in physical activity and sedentary behaviors among US children and adolescents: prevalence, correlates, and intervention implications. J Public Health Policy 2009;30 Suppl 1:S309-S334. [doi: 10.1057/jphp.2008.46] [Medline: 19190581]
2. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
3. Wolf AM, Gortmaker SL, Cheung L, Gray HM, Herzog DB, Colditz GA. Activity, inactivity, and obesity: racial, ethnic, and age differences among schoolgirls. Am J Public Health 1993 Nov;83(11):1625-1627. [doi: 10.2105/ajph.83.11.1625] [Medline: 8238692]
4. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40. [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
5. Barnekow-Bergkvist M, Hedberg G, Janlert U, Jansson E. Adolescent determinants of cardiovascular risk factors in adult men and women. Scand J Public Health 2001 Sep;29(3):208-217. [Medline: 11680773]
6. Sacker A, Cable N. Do adolescent leisure-time physical activities foster health and well-being in adulthood? Evidence from two British birth cohorts. Eur J Public Health 2006 Jun;16(3):332-336. [doi: 10.1093/eurpub/cki189] [Medline: 16157610]
7. Schiller JS, Lucas JW, Peregoy JA. Summary health statistics for U.S. Adults: National Health Interview Survey, 2011. Vital Health Stat 10 2012 Dec(256):1-218. [Medline: 25116400]
8. Biddle SJH, Whitehead SH, O'Donovan TM, Nevill ME. Correlates of participation in physical activity for adolescent girls: a systematic review of recent literature. J Phys Act Health 2005;2(4):423-434. [doi: 10.1123/jpah.2.4.423]
9. Grieser M, Vu MB, Bedimo-Rung AL, et al. Physical activity attitudes, preferences, and practices in African American, Hispanic, and Caucasian girls. Health Educ Behav 2006 Feb;33(1):40-51. [doi: 10.1177/1090198105282416] [Medline: 16397158]
10. Grieser M, Neumark-Sztainer D, Saksvig BI, Lee JS, Felton GM, Kubik MY. Black, Hispanic, and white girls' perceptions of environmental and social support and enjoyment of physical activity. J Sch Health 2008 Jun;78(6):314-320. [doi: 10.1111/j.1746-1561.2008.00308.x] [Medline: 18489464]
11. Kelly EB, Parra-Medina D, Pfeiffer KA, et al. Correlates of physical activity in black, Hispanic, and white middle school girls. J Phys Act Health 2010 Mar;7(2):184-193. [doi: 10.1123/jpah.7.2.184] [Medline: 20484757]
12. McKenzie TL, Marshall SJ, Sallis JF, Conway TL. Student activity levels, lesson context, and teacher behavior during middle school physical education. Res Q Exerc Sport 2000 Sep;71(3):249-259. [doi: 10.1080/02701367.2000.10608905] [Medline: 10999262]
13. McKenzie TL, Marshall SJ, Sallis JF, Conway TL. Leisure-time physical activity in school environments: an observational study using SOPLAY. Prev Med 2000 Jan;30(1):70-77. [doi: 10.1006/pmed.1999.0591] [Medline: 10642462]
14. Robbins LB, Pender NJ, Kazanis AS. Barriers to physical activity perceived by adolescent girls. J Midwifery Womens Health 2003;48(3):206-212. [doi: 10.1016/s1526-9523(03)00054-0] [Medline: 12764306]

XSL•FO
RenderX

15. Sallis JF, Zakarian JM, Hovell MF, Hofstetter CR. Ethnic, socioeconomic, and sex differences in physical activity among adolescents. J Clin Epidemiol 1996 Feb;49(2):125-134. [doi: 10.1016/0895-4356(95)00514-5] [Medline: 8606313]
16. Taylor WC, Yancey AK, Leslie J, et al. Physical activity among African American and Latino middle school girls: consistent beliefs, expectations, and experiences across two sites. Women Health 1999;30(2):67-82. [doi: 10.1300/j013v30n02_05] [Medline: 10881759]
17. Verloigne M, Cardon G, De Craemer M, D'Haese S, De Bourdeaudhuij I. Mediating effects of self-efficacy, benefits and barriers on the association between peer and parental factors and physical activity among adolescent girls with a lower educational level. PLoS One 2016;11(6):e0157216. [doi: 10.1371/journal.pone.0157216] [Medline: 27309847]
18. Pearson N, Braithwaite R, Biddle SJH. The effectiveness of interventions to increase physical activity among adolescent girls: a meta-analysis. Acad Pediatr 2015;15(1):9-18. [doi: 10.1016/j.acap.2014.08.009] [Medline: 25441655]
19. Sharma M. Physical activity interventions in Hispanic American girls and women. Obes Rev 2008 Nov;9(6):560-571. [doi: 10.1111/j.1467-789X.2008.00501.x] [Medline: 18507777]
20. Larsen B, Benitez T, Cano M, et al. Web-based physical activity intervention for Latina adolescents: feasibility, acceptability, and potential efficacy of the Niñas Saludables study. J Med Internet Res 2018 May 9;20(5):e170. [doi: 10.2196/jmir.9206] [Medline: 29743151]
21. Larsen B, Greenstadt ED, Olesen BL, Marcus BH, Godino J, Zive MM. An mHealth physical activity intervention for Latina adolescents: iterative design of the Chicas Fuertes study. JMIR Form Res 2021 Jun 15;5(6):e26195. [doi: 10.2196/26195] [Medline: 34128823]
22. Naqvi JB, Olesen B, Greenstadt E, et al. Randomized controlled trial of a multiple technology-based physical activity intervention for Latina adolescents: recruitment strategies and baseline data from the Chicas Fuertes trial. Contemp Clin Trials 2024 Dec;147:107716. [doi: 10.1016/j.cct.2024.107716] [Medline: 39413991]
23. Child activity: an overview. CDC. URL: https://www.cdc.gov/physical-activity-basics/guidelines/children.html [accessed 2025-05-26]
24. Bandura A. Social Foundations of Thought and Action: Englewood Cliffs; 1986.
25. Prochaska JO. Toward a comprehensive model of change. In: Miller WR, editor. Treating Addictive Behaviors: Springer; 1986:3-27. [doi: 10.1007/978-1-4613-2191-0_1]
26. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
27. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
28. Larsen B, Greenstadt E, Olesen B, et al. A multiple technology-based physical activity intervention for Latina adolescents in the USA: randomized controlled trial study protocol for Chicas Fuertes. Trials 2022 Feb 23;23(1):176. [doi: 10.1186/s13063-022-06105-2] [Medline: 35197106]
29. Janz KF. Validation of the CSA accelerometer for assessing children's physical activity. Med Sci Sports Exerc 1994 Mar;26(3):369-375. [Medline: 8183103]
30. Melanson EL, Freedson PS. Validity of the Computer Science and Applications, Inc. (CSA) activity monitor. Med Sci Sports Exerc 1995 Jun;27(6):934-940. [Medline: 7658958]
31. Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC. Using objective physical activity measures with youth: how many days of monitoring are needed? Med Sci Sports Exerc 2000 Feb;32(2):426-431. [doi: 10.1097/00005768-200002000-00025] [Medline: 10694127]
32. Pereira MA, FitzerGerald SJ, Gregg EW, et al. A collection of Physical Activity Questionnaires for health-related research. Med Sci Sports Exerc 1997 Jun;29(6 Suppl):S1-205. [Medline: 9243481]
33. Sallis JF, Buono MJ, Roby JJ, Micale FG, Nelson JA. Seven-day recall and other physical activity self-reports in children and adolescents. Med Sci Sports Exerc 1993 Jan;25(1):99-108. [doi: 10.1249/00005768-199301000-00014] [Medline: 8423762]
34. Rauh MJ, Hovell MF, Hofstetter CR, Sallis JF, Gleghorn A. Reliability and validity of self-reported physical activity in Latinos. Int J Epidemiol 1992 Oct;21(5):966-971. [doi: 10.1093/ije/21.5.966] [Medline: 1468861]
35. Dunn AL, Garcia ME, Marcus BH, Kampert JB, Kohl HW, Blair SN. Six-month physical activity and fitness changes in Project Active, a randomized trial. Med Sci Sports Exerc 1998 Jul;30(7):1076-1083. [doi: 10.1097/00005768-199807000-00009] [Medline: 9662676]
36. Dunn AL, Marcus BH, Kampert JB, Garcia ME, Kohl HW, Blair SN. Comparison of lifestyle and structured interventions to increase physical activity and cardiorespiratory fitness: a randomized trial. JAMA 1999 Jan 27;281(4):327-334. [doi: 10.1001/jama.281.4.327] [Medline: 9929085]
37. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]

XSL•FO
RenderX

38. Hartman SJ, Chen R, Tam RM, Narayan HK, Natarajan L, Liu L. Fitbit use and activity levels from intervention to 2 years after: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2022 Jun 30;10(6):e37086. [doi: 10.2196/37086] [Medline: 35771607]
39. Larsen B, Micucci S, Hartman SJ, Ramos G. Feasibility and acceptability of a counseling- and mHealth-based physical activity intervention for pregnant women with diabetes: the fit for two pilot study. JMIR Mhealth Uhealth 2020 Oct 21;8(10):e18915. [doi: 10.2196/18915] [Medline: 33084584]
40. Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. Res Q Exerc Sport 1992 Mar;63(1):60-66. [doi: 10.1080/02701367.1992.10607557] [Medline: 1574662]
41. Prochaska JJ, Rodgers MW, Sallis JF. Association of parent and peer support with adolescent physical activity. Res Q Exerc Sport 2002 Jun;73(2):206-210. [doi: 10.1080/02701367.2002.10609010] [Medline: 12092896]
42. Mullen SP, Olson EA, Phillips SM, et al. Measuring enjoyment of physical activity in older adults: invariance of the physical activity enjoyment scale (paces) across groups and time. Int J Behav Nutr Phys Act 2011 Sep 27;8(1):103. [doi: 10.1186/1479-5868-8-103] [Medline: 21951520]
43. Rosenberg D, Ding D, Sallis JF, et al. Neighborhood Environment Walkability Scale for Youth (NEWS-Y): reliability and relationship with physical activity. Prev Med 2009;49(2-3):213-218. [doi: 10.1016/j.ypmed.2009.07.011] [Medline: 19632263]
44. Sallis JF, Kerr J, Carlson JA, et al. Evaluating a brief self-report measure of neighborhood environments for physical activity research and surveillance: Physical Activity Neighborhood Environment Scale (PANES). J Phys Act Health 2010 Jul;7(4):533-540. [doi: 10.1123/jpah.7.4.533] [Medline: 20683096]
45. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
46. Marcus BH, Rossi JS, Selby VC, Niaura RS, Abrams DB. The stages and processes of exercise adoption and maintenance in a worksite sample. Health Psychol 1992;11(6):386-395. [doi: 10.1037//0278-6133.11.6.386] [Medline: 1286658]
47. LeBlanc AGW, Janssen I. Difference between self-reported and accelerometer measured moderate-to-vigorous physical activity in youth. Pediatr Exerc Sci 2010 Nov;22(4):523-534. [doi: 10.1123/pes.22.4.523] [Medline: 21242602]
48. Larsen BA, Carr LJ, Dunsiger S, Marcus BH. Effect of a moderate-intensity demonstration walk on accuracy of physical activity self-report. J Exerc Sci Fit 2017 Jun;15(1):1-7. [doi: 10.1016/j.jesf.2016.10.002]
49. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text message interventions for physical activity: a systematic review and meta-analysis. Am J Prev Med 2020 Jan;58(1):142-151. [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]
50. Ludwig K, Arthur R, Sculthorpe N, Fountain H, Buchan DS. Text messaging interventions for improvement in physical activity and sedentary behavior in youth: systematic review. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e10799. [doi: 10.2196/10799] [Medline: 30224335]
51. Hartman SJ, Dunsiger SI, Bock BC, et al. Physical activity maintenance among Spanish-speaking Latinas in a randomized controlled trial of an Internet-based intervention. J Behav Med 2017 Jun;40(3):392-402. [doi: 10.1007/s10865-016-9800-4] [Medline: 27752866]
52. Linke SE, Dunsiger SI, Gans KM, et al. Association between physical activity intervention website use and physical activity levels among Spanish-speaking Latinas: randomized controlled trial. J Med Internet Res 2019 Jul 24;21(7):e13063. [doi: 10.2196/13063] [Medline: 31342902]
53. Marcus BH, Dunsiger SI, Pekmezi DW, et al. The Seamos Saludables study. Am J Prev Med 2013 Nov;45(5):598-605. [doi: 10.1016/j.amepre.2013.07.006]
54. Marcus BH, Hartman SJ, Larsen BA, et al. Pasos Hacia La Salud: a randomized controlled trial of an internet-delivered physical activity intervention for Latinas. Int J Behav Nutr Phys Act 2016 May 28;13(1):62. [doi: 10.1186/s12966-016-0385-7] [Medline: 27234302]
55. Powell KE, Paluch AE, Blair SN. Physical activity for health: What kind? How much? How intense? On top of what? Annu Rev Public Health 2011;32:349-365. [doi: 10.1146/annurev-publhealth-031210-101151] [Medline: 21128761]

Abbreviations

CES-D: Center for Epidemiological Studies Depression Scale
CONSORT: Consolidated Standards for Reporting Trials
GED: General Education Development
HIPAA: Health Insurance Portability and Accountability Act
LPA: light physical activity
MVPA: moderate-vigorous physical activity
NEWS-Y: Neighborhood Environment Walkability Scale for Youth
PA: physical activity
PACES: Physical Activity Enjoyment Scale
PANES: Physical Activity Neighborhood Environment Scale

XSL•FO
RenderX

PAR: Physical Activity Recall
REDCap: Research Electronic Data Capture

Edited by L Buis; submitted 22.01.25; peer-reviewed by M Fraser, O Abasilim; revised version received 16.10.25; accepted 17.10.25; published 11.12.25.

Please cite as:
Carson JR, Greenstadt E, Olivera B, Dunsiger S, Zive M, Higgins M, Godino J, Marcus B, Meyer D, Larsen B
A Multiple Technology–Based Physical Activity Intervention for Latina Adolescents: Results From the Chicas Fuertes Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e71623
URL: https://mhealth.jmir.org/2025/1/e71623
doi:10.2196/71623

© Jacob R Carson, Emily Greenstadt, Brittany Olivera, Shira Dunsiger, Michelle Zive, Michael Higgins, Job Godino, Bess Marcus, Dawn Meyer, Britta Larsen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Minimal Clinically Important Difference of Average Daily Steps Measured Through a Consumer Smartwatch in People With Mild-to-Moderate Parkinson Disease: Cross-Sectional Study

Edoardo Bianchini[1,2,3], MD; Marika Alborghetti[1], MD, PhD; Silvia Galli[1], MD; Clint Hansen[2,4], PhD; Alessandro Zampogna[3,5], MD, PhD; Antonio Suppa[3,5], MD, PhD; Marco Salvetti[1,5], MD; Francesco Ernesto Pontieri[1], MD; Domiziana Rinaldi[1], MD, PhD; Nicolas Vuillerme[2,6], PhD

[1]Department of Neuroscience, Mental Health and Sensory Organs (NESMOS), Sapienza University of Rome, Via di Grottarossa 1035, Rome, Italy

[2]AGEIS, Université Grenoble Alpes, Grenoble, France

[3]Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

[4]Department of Neurology, Kiel University, Kiel, Germany

[5]IRCCS Neuromed Institute, Pozzilli, Italy

[6]Institut Universitaire de France, Paris, France

Corresponding Author:
Edoardo Bianchini, MD
Department of Neuroscience, Mental Health and Sensory Organs (NESMOS), Sapienza University of Rome, Via di Grottarossa 1035, Rome, Italy

Abstract

Background: Recent studies demonstrated the validity, reliability, and accuracy of consumer smartwatches for measuring daily steps in people with Parkinson disease (PD). However, no study to date has estimated the minimal clinically important difference (MCID) for average daily steps (avDS), measured through a consumer smartwatch in people with PD.

Objective: This study aimed to calculate the MCID of avDS, measured through a commercial smartwatch (Garmin Vivosmart 4) in people with PD.

Methods: People with PD with a disease stage <4, without cognitive impairment, and who were able to walk unaided, wore a Garmin Vivosmart 4 smartwatch for 5 consecutive days on the wrist least affected by the disease, allowing the computation of avDS. To define the 3 levels of MCID for avDS, we used an anchor-based method linked to: (1) scales capturing subtle changes in global mobility and motor functions, (2) clinical and health-related measures, and (3) disease-related patient-reported outcomes. Linear regressions, Student *t* test, and ANOVA were used to estimate the minimal change in avDS based on anchors relevant change. For each level, the overall MCID was calculated as the average of the variables included, and the range was reported.

Results: A total of 100 people with PD were enrolled. Participants took on average 5949 (SD 3034) daily steps, ranging from 357 to 12,620. The MCID of avDS anchored to standardized measures of motor symptoms and mobility was 581 steps/day (range 554 - 608) or around 10% of mean avDS in our population. The MCID of avDS anchored to clinical and health-related variables was 1200 steps/day (range 350 - 1683), or around 20% of mean avDS in our population. Finally, the MCID of avDS anchored to disease-related patient-reported outcomes was 1592 steps/day (range 594 - 2589), or around 27% of the mean avDS in our population.

Conclusions: These findings could be relevant for designing future clinical trials involving avDS as a digital mobility outcome in daily life for people with PD and evaluating the effectiveness of the intervention promoting free-living walking in this population.

(JMIR Mhealth Uhealth 2025;13:e64213) doi:10.2196/64213

KEYWORDS

gait; Parkinson's disease; sensors; step count; smartwatch; MCID; wearables; activity monitor; motor symptoms; digital mobility outcomes; digital health; cross sectional study; linear regression; effectiveness; minimal clinically important difference

Introduction

Parkinson disease (PD) is a neurodegenerative disease characterized by both motor and nonmotor symptoms. Among the motor manifestations of PD, gait disturbances are notably prevalent and encompass a variety of issues, such as a shuffling gait, reduced step length and diminished automaticity, decreased arm swing, increased double support time, altered cadence, and periodic episodes such as freezing of gait (FOG) [1]. Walking, a fundamental daily activity, is crucial for maintaining functional

XSL•FO
RenderX

independence and social well-being [2]. Gait disturbances, therefore, represent a particularly disabling category of symptoms in PD, significantly impacting the autonomy and quality of life of patients [3].

Gait characteristics can be examined through various methods, both in controlled and ecological settings. Recently, the use of wearable inertial sensors for gait analysis has gained considerable traction [4-6]. Specifically, the daily step count serves as an indirect yet rather easily obtainable indicator of walking and physical activity [7]. While this metric does not provide detailed gait analysis, there is established evidence linking a reduced daily step count to a range of health issues, including oncological and cardiovascular diseases [8], dementia [9], as well as an elevated risk of overall mortality [10-13]. In the context of PD, studies have consistently demonstrated a correlation between a decrease in daily steps and increased disease severity [14] as well as with levels of physical activity [15,16]. These studies have also proposed a minimum daily step goal of around 4200 to align with physical activity recommendations in the early stages of PD [17].

Step measurement can be accomplished using either research-grade or consumer devices. Research-grade devices are specifically developed for use by professionals in research or clinical settings and are generally recognized for their superior accuracy and reliability [18-21]. Despite this, they come with drawbacks such as higher costs, limited market availability, the need for specialized expertise for operation, and less focus on aesthetic appeal and user-friendliness. In contrast, consumer wearable devices are designed for the general public. These devices are easily accessible in the market and user-friendly, requiring no specialized knowledge for operation. However, they typically lack validation for research purposes and often do not include sophisticated algorithms or analysis software [22]. Despite this, there is an increasing focus in the literature on using consumer wearables for data collection and monitoring in neurological patients [23-25]. Both device types demonstrate reliable step counting in individuals with normal gait patterns [26], though research-grade devices show higher reliability across a wider range of conditions [27]. However, for individuals with nonphysiological gait, such as those with reduced walking speeds or those using walking aids, the performance of consumer wearables tends to diminish [18,28-31].

PD poses a unique challenge for gait analysis, as its associated disturbed gait patterns could adversely impact the reliability of device algorithms. Furthermore, the occurrence of tremors and dyskinesias in patients with PD may introduce additional background noise in recordings, potentially resulting in an overestimation of step counts [31,32]. Conversely, the presence of rigidity and bradykinesia, characteristic of PD, might lead to an underestimation of steps, especially when the activity monitor is worn on the wrist [33].

In our previous in-laboratory study involving 47 people with PD, we demonstrated a good criterion validity (intraclass correlation coefficient=0.76) and an acceptable level of error (mean absolute percentage error=5%) in step counting using a consumer smartwatch (Garmin Vivosmart 4), when worn on the side least affected by the disease and under well-controlled pharmacological conditions [31]. In the context of real-life settings, a separate study underscored the criterion validity of average daily steps (avDS) measured by several wrist-worn consumer devices (Fitbit Alta [Google] and Fitbit Inspire 3 [Google]) in 28 people with PD at home, compared with a research-grade device (Dynaport movemonitor) [34]. We recently replicated these results, using Garmin Vivosmart 4, in a proof-of-concept experiment [35]. Beyond criterion validity, the reliability of a measurement tool—the consistency of repeated measures in unchanged individuals—is equally crucial. In this regard, our recent findings indicate that a minimum of 4 days is necessary to obtain a reliable estimate of avDS recorded by a consumer smartwatch (Garmin Vivosmart 4) in free-living conditions [36].

An essential additional metrological characteristic of an activity monitor is its capability to reliably capture changes in measured parameters and the clinical significance of these changes. In this context, the minimum detectable change (MDC) and the minimal clinically important difference (MCID) are 2 key metrics and are pivotal for calculating clinical trial sample sizes and assessing the effectiveness of interventions [37,38]. On the one hand, MDC denotes the smallest change that can be confidently identified beyond the measurement error of the instrument used for a specific parameter [37]. In a recent work of ours on 56 individuals with mild-to-moderate PD, we found that for a monitoring period of at least 4 days, the MDC for the Garmin Vivosmart 4 ranged between 1198 and 1524 steps/day, averaging at 1366 steps/day [36]. On the other hand, MCID represents the smallest difference in a parameter that is meaningful from a clinical perspective [38]. In fact, a statistically significant change after an intervention may not represent a modification that has a real impact on the patient's condition and, ultimately, might not be perceived by individuals. Therefore, a measure for clinical significance is required. MCID can be estimated in different ways, and there is no definite consensus for the best method [39,40]. Globally, two calculation methods exist: (1) distribution-based and (2) anchor-based. Distribution-based methods rely on the statistical characteristics of the variables of interest within the given population to identify the magnitude of change that would be unlikely to be observed by chance. Spread parameters, such as SD and SE of measurements, are commonly used indices to calculate MCID using distribution-based methods [39,40]. However, several authors underlined the limitation of distribution-based methods in capturing truly meaningful change and considered the use of these techniques only as an addition to anchor-based methods [39,41].

Anchor-based methods compare the change in patients' variable of interest with how patients score on a second, explicit metric of improvement (ie, the anchor) [39,40,42]. Usually, when applying the anchor-based method, improvement is evaluated longitudinally following an intervention and a time period for which a change in the anchor measure is expected [39]. However, cross-sectional evaluation is also possible, as reported by Motl and collaborators [43]. With this second approach, a value, or a range of values, is defined for the anchor that corresponds to the MCID and then the target score of the variable of interest that corresponds to the anchor value is

XSL•FO
RenderX

calculated. With the cross-sectional method, the between-subjects variability in the relationship between the anchor and the new measure is often high; therefore, multiple anchors are often needed to provide a well-triangulated MCID value [39]. However, this could offer the advantage of delineating several levels of MCID depending on their specific goals and expected outcomes, providing a more versatile framework [43]. While generally considered more reliable than distribution-based methods, potential limitations of anchor-based methods include recall bias and the need for at least a moderate correlation between the anchor and the variable of interest [39].

As of now, there are no published studies that have evaluated the MCID for avDS in people with PD. Therefore, considering these identified research gaps and the clinical significance of the psychometric and clinometric properties of daily steps measured by consumer smartwatches in free-living conditions for people with PD, the objective of this study was to determine the MCID of avDS as recorded by a consumer smartwatch in real-world settings for this population of individuals.

Methods

Population

Participants were recruited at the Movement Disorder Outpatient Service of the Sant'Andrea University Hospital, Rome, Italy, in the period between March 2023 and February 2024. Inclusion criteria were (1) diagnosis of idiopathic PD according to Movement Disorder Society criteria [44], (2) aged 18 years or older, (3) disease stage <4 according to the modified Hoeh and Yahr scale (mHY; "severe disability; still able to walk or stand unassisted") [45], and (4) ability to perform the experimental procedure. Exclusion criteria were (1) cognitive impairment as defined by a Montreal Cognitive Assessment score < 21, and (2) orthopedic, rheumatologic, or systemic conditions affecting mobility as judged by the assessor.

Ethical Considerations

This cross-sectional study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments. Approval was granted by the local Ethical Committee of Sapienza, University of Rome (Reference 0372/2022). Data collection and processing followed the current European regulation for data protection. All participants provided written informed consent (Multimedia Appendix 1) before the beginning of measurements. All data were deidentified. Participants did not receive any form of compensation.

Demographic and Clinical Data

Participants were evaluated during scheduled visits. Demographics (age and sex) and anthropometric measures (weight, height, and BMI) were collected. Disease duration, disease stage according to the mHY scale, and levodopa equivalent daily dose (LEDD) [46] were also collected. Participants were then evaluated with the following rating scales:

The sum of the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [47] parts I and II was used to assess the impact of motor and nonmotor symptoms on daily life [48]. MDS-UPDRS part III was used to assess the severity of motor symptoms. From MDS-UPDRS part II and III scores, participants were classified into tremor dominant (TD), postural instability and gait disorder (PIGD), or indeterminate disease subtypes, according to Stebbins et al [49]. Participants were also grouped in those with mild and moderate disease severity based on the MDS-UPDRS score as proposed by Martínez-Martín and colleagues [50]. The Parkinson's Disease Questionnaire-39 was used to evaluate participants' quality of life [51]. The Short Physical Performance Battery (SPPB) was used to assess lower limb function and general physical capacity [52]. The Wearing-Off Questionnaire-19 was used to assess the presence of wearing off. A score ≥2 was used as a cutoff to define the presence of wearing off [53]. The Fatigue Severity Scale was used to assess fatigue severity; a score ≥4.67 was used to define clinically relevant fatigue [54]. The item 1 of the New Freezing of Gait Questionnaire [55] was used to assess the presence of FOG, defined as a score >0. Finally, the MDS-UPDRS item 2.10 was used to assess the absolute presence or absence of tremor as reported by participants.

Experimental Procedure

Participants then received the smartwatch Garmin Vivosmart 4 to be worn at home for 5 consecutive days, including at least 1 weekend day, on the wrist of the body side least affected by the disease [31]. Participants were instructed to wear the smartwatch at all times during the day and night and remove it only when involved in water activities (eg, bathing, showering, and swimming). Participants were also asked to perform daily activities as usual. No additional reminder or instructions were provided to participants during the monitoring period. We chose a 5-day period since we demonstrated previously that a minimum of 4 days of monitoring is needed to reliably estimate daily step count in people with PD [36]. Each smartwatch was configured according to the producer's recommendations by indicating the patient's age, height, weight, and the wrist on which the smartwatch was worn (ie, left or right). During the monitoring period, participants could access information on the smartwatch screen, including the total number of steps, heart rate, and sleep time. The device did not provide instructions, alerts, or other messages to participants.

After 5 days, they returned the smartwatch to the hospital. From the smartwatch dashboard, the total daily number of steps for each day was recorded, and the avDS were calculated [36]. Compliance was assessed based on the participants' dashboard data. We could assess that the device was worn through heart rate and activity data. We considered all recording days with ≥80% wear time while awake to be valid. We estimated the awake time as 24 hours minus the total sleep time as indicated by the device.

Data and Statistical Analysis

The statistical analyses were performed using JASP v0.18.3.0 (JASP Team, University of Amsterdam), R v4.0.3 (R Core Team), and RStudio v2022.07.1+554 for Windows (R Foundation for Statistical Computing). Descriptive statistics were calculated for the examined variables. Normality of

XSL•FO
RenderX

distributions was assessed by histogram and residual plots inspection.

To define MCID, we used an anchor-based method identifying three levels of variables, following the approach of Motl and collaborators [43]: (1) scales capturing subtle changes in global mobility and motor functions (ie, MDS-UPDRS part III and SPPB); (2) clinical and health-related measures including disease duration, mHY, LEDD, PD subtype, disease severity as assessed by MDS-UPDRS, age, sex, and BMI; and (3) disease-related patient-reported outcomes (PROs) including MDS-UPDRS parts I and II, the Parkinson's disease questionnaire-39, presence of FOG, presence of wearing off, presence of fatigue, and presence of tremor.

Such an approach could provide ranges of MCIDs and options for MCID selection that are appropriate for specific populations and circumstances, including both assessor-based and PROs, in clinical trials and interventions involving people with PD [39].

Since previous studies have recommended a correlation coefficient of at least 0.3 - 0.35 between the change score and the anchor [56], we first conducted a correlation analysis between avDS and the included variables using the Spearman test. Only associations with a correlation coefficient >0.3 were considered for MCID estimation.

We then conducted a linear regression between avDS and MDS-UPDRS part I+II and part III, mHY, SPPB, LEDD, and age for estimating the incremental change in daily steps per relevant variable change. To evaluate the relevant change, we used the already known MCID for each scale included in the regression analysis. Scales for which an MCID was not reported were excluded from the analysis. In case of asymmetric MCID (ie, different values for improvement and deterioration), we used the average value rounded to the higher next whole number.

For the remaining variables, we calculated mean and SE differences in avDS between groups by performing between-subjects ANOVA with Bonferroni-Holm post hoc pairwise comparisons in case of overall significance, when the variables included more than 2 classes. For 2-class variables, we used the Student *t* test for independent groups to compute the mean differences and SE in avDS. Cohen *d* was used to estimate the standardized effect size. MCID for the 3 levels was then calculated, and the average of the included variables and range was reported. The level of significance was set at α<.05.

Results

Overview

A total of 100 people with PD were enrolled in the study. All participants were monitored through Garmin Vivosmart 4 at home for a period of 5 consecutive days. No participants or days were excluded based on the prespecified compliance criteria. Participants took on average 5949 (SD 3034) daily steps, ranging from 357 to 12620. Details of demographic, anthropometric, and clinical variables in the enrolled population are illustrated in Table 1.

XSL•FO
RenderX

Table . Clinical, anthropometric, and demographic characteristics of enrolled population.

Characteristics	People with PD[a] (n=100)
Age (y), mean (SD)	67.8 (8.5)
Height (cm), mean (SD)	171 (9.0)
Weight (kg), mean (SD)	76.0 (13.2)
BMI (kg/m^2), mean (SD)	25.8 (3.5)
Sex, n (%)	
Male	67 (67)
Female	33 (33)
Disease duration (y), mean (SD)	6.2 (4.4)
LEDD[b] (mg), mean (SD)	544 (298)
mHY[c], median (IQR)	2 (2 - 2.5)
MDS-UPDRS I+II[d], median (IQR)	11 (6 - 16)
MDS-UPDRS-III, median (IQR)	27 (21 - 32)
PDQ-39[e], mean (SD)	12.1 (9.4)
Disease subtypes, n (%)	
TD[f]	31 (31)
IND[g]	8 (8)
PIGD[h]	61 (61)
WO[i]	28 (28)
FOG[j]	13 (13)
Fatigue	23 (23)
Tremor	55 (55)
Disease severity, n (%)	
Mild	63 (63)
Moderate	37 (37)
SPPB[k], median (IQR)	10 (9 - 12)
avDS[l], mean (SD)	5949 (3034)

[a]PD: Parkinson disease.
[b]LEDD: levodopa equivalent daily dose.
[c]mHY: modified Hoehn and Yahr scale.
[d]MDS-UPDRS: Movement Disorder Society Unified Parkinson Disease Rating Scale.
[e]PDQ-39: Parkinson's disease questionnaire-39.
[f]TD: tremor dominant.
[g]IND: indeterminate.
[h]PIGD: postural instability and gait disorder.
[i]WO: wearing-off.
[j]FOG: freezing of gait.
[k]SPPB: Short Physical Performance Battery.
[l]avDS: average daily steps.

XSL•FO
RenderX

MCID

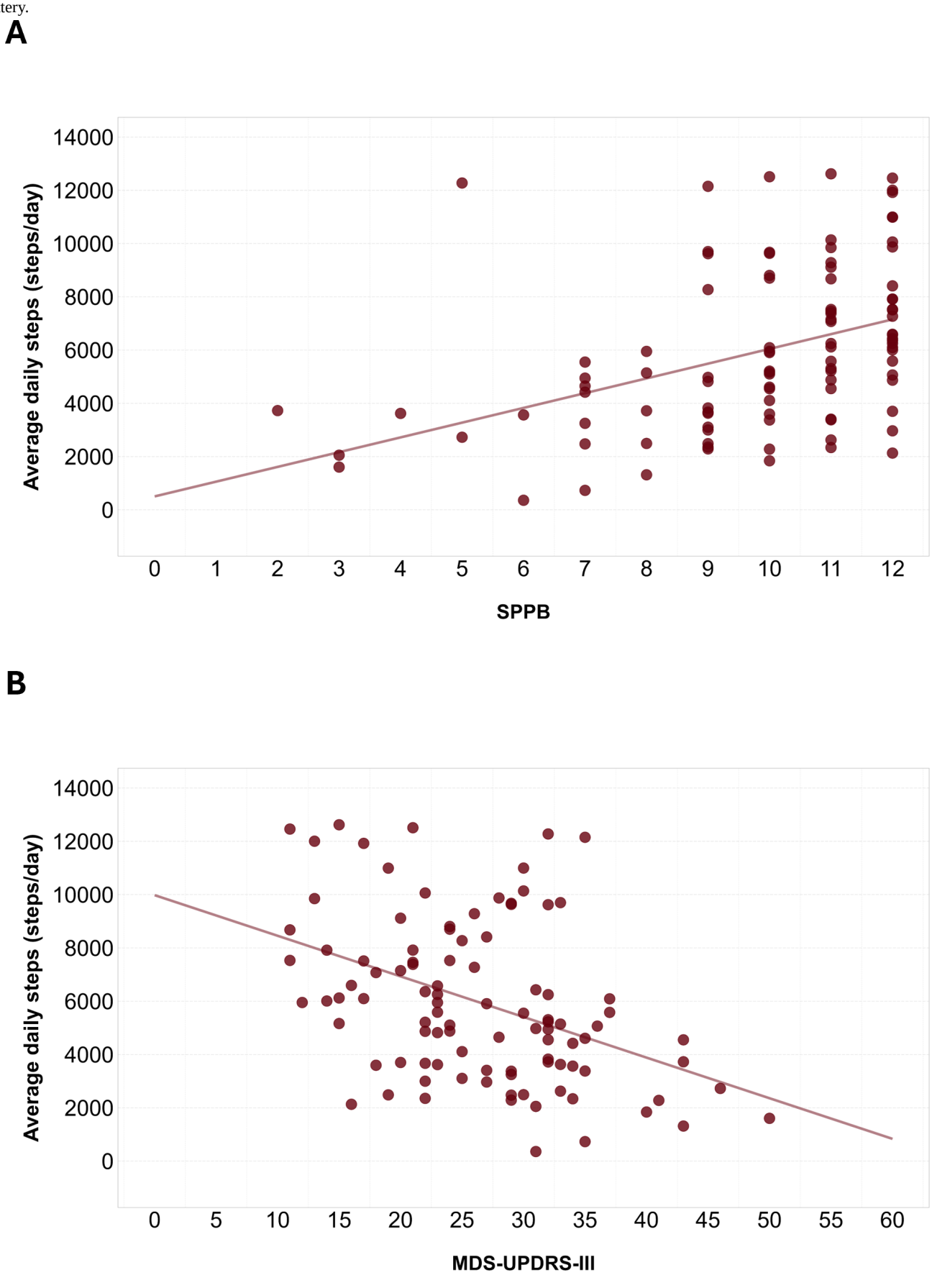
Global Mobility and Motor Functions

All included variables showed a correlation coefficient >0.3 with avDS. The 2 linear regression models between avDS and MDS-UPDRS-III and SPPB were statistically significant (MDS-UPDRS-III: $F_{1,98}$=20.46, P<.001, R^2=0.16; SPPB: $F_{1,98}$=19.84, P<.001; R^2=0.16) (Figure 1).

The resulting regression equations indicated that every 1-point increase in MDS-UPDRS-III corresponded to a 152 steps/day decrease (avDS=9778−152×MDS-UPDRS-III). MCID for MDS-UPDRS was reported as 3.25 for score decrease and 4.63 for score increase [57]; therefore, we calculated a rounded average of 4 points, and we estimated a 608 steps/day modification for each 4-point change in MDS-UPDRS-III. Finally, each 1-point increase in SPPB score corresponded to an increase of 554 steps/day (avDS=503+554×SPPB). A change of 1 point is also the reported MCID for SPPB in older adults; therefore, we adopted this value [58]. Overall, the mean MCID across the 3 clinical scales that capture meaningful changes in motor and mobility condition was 581 steps/day (range 554 - 608), corresponding to 10% of the avDS in our population.

XSL•FO
RenderX

Figure 1. Correlation of average daily steps with Short Physical Performance Battery and Movement Disorder Society Unified Parkinson Disease Rating Scale-III. Scatterplot showing correlation between average daily steps and Short Physical Performance Battery score (panel A) and between average daily steps and Movement Disorder Society Unified Parkinson Disease Rating Scale-III score (panel B). A linear regression line was added to show the linear tendency. MDS-UPDRS: Movement Disorder Society Unified Parkinson Disease Rating Scale; SPPB: Short Physical Performance Battery.



XSL•FO
RenderX

Clinical and Health-Related Measures

All included variables except BMI showed a correlation coefficient >0.3 with avDS. Regression models of mHY, LEDD, age, and disease duration were statistically significant (mHY: $F_{1,98}$=26.26, P<.001, R^2=0.20; LEDD: $F_{1,98}$=12.97, P<.001, R^2=0.11; age: $F_{1,98}$=18.22, P<.001, R^2=0.15; disease duration: $F_{1,98}$=7.01, P=.009, R^2=0.06) (Figure 2).

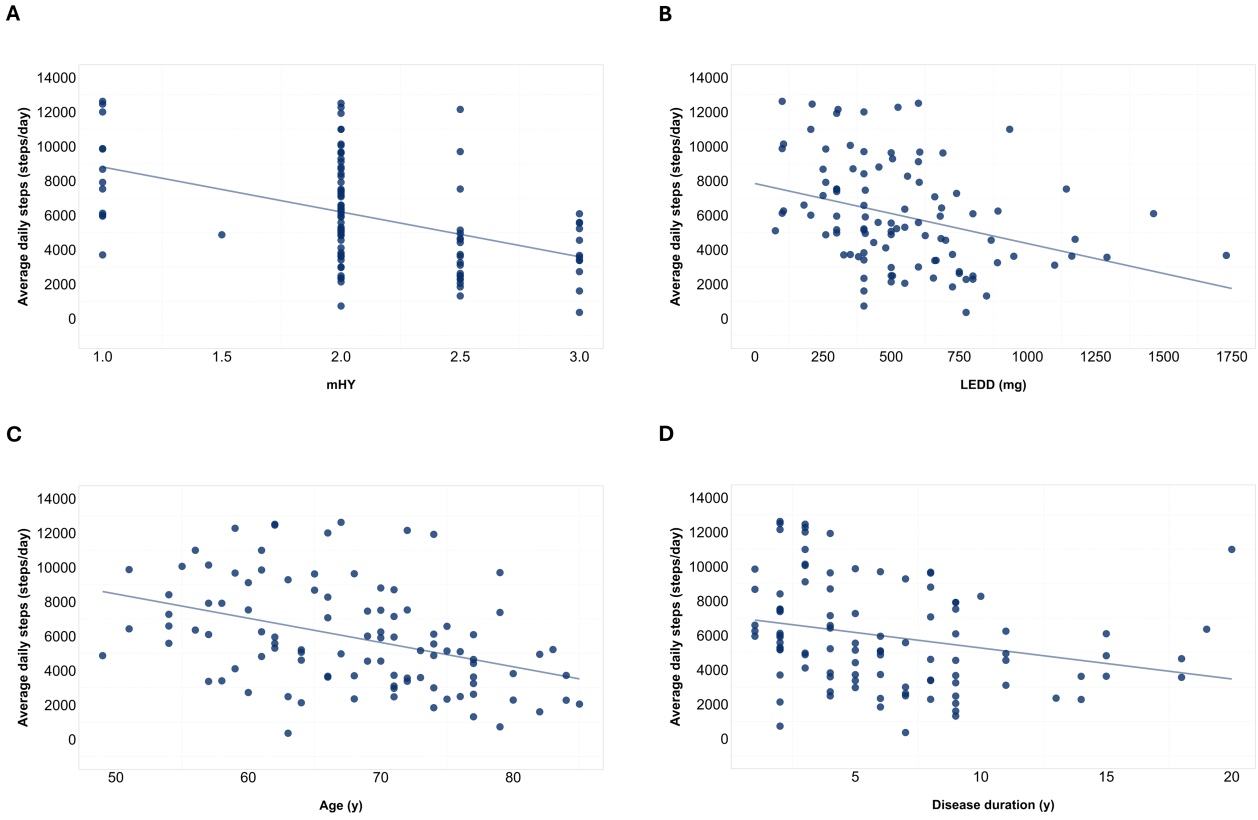
The resulting regression equations indicated that every 1-point increase in mHY stage yielded a reduction of 2607 steps/day (avDS=11,409−2607×mHY). Since mHY included 0.5 steps increases between scores 1 and 3, we calculated that each step corresponded to a 1304 steps/day reduction.

For LEDD, regression equations indicated that every 1 equivalent mg increase led to a reduction of 3.5 steps/day (avDS=7846−3.5×mHY). Since the usual single dose of levodopa corresponds to 100 mg, we used this value as a reference, and we estimated a 350 steps/day modification for each 100 mg change in LEDD.

For age, regression equations indicated that a 1-year increase in age yielded a reduction of 141 steps/day (avDS=15,512−141×age). Furthermore, 10-year increases are usually used to divide participants in age groups; therefore, we estimated a 1410-step/day modification for each 10-year change of age.

For disease duration, regression equations indicated that a 1-year increase in disease duration resulted in a reduction of 180 steps/day (avDS=7071−180×age). We chose 5 years as a significant increment in disease duration since this was a common cutoff to categorize disease duration in previous studies [59-61]. Therefore, we estimated a 900-step/day modification for each 5-year change of disease duration.

Figure 2. Correlation of average daily steps with modified Hoehn and Yahr stage, levodopa equivalent daily dose, age, and disease duration. Scatterplots are reported showing correlation between average daily steps and modified Hoehn and Yahr stage (panel A), levodopa equivalent daily dose (panel B), age (panel C), and disease duration (panel D). A linear regression line was added to show the linear tendency. LEDD: levodopa equivalent daily dose; mHY: modified Hoehn and Yahr stage.



Considering categorical descriptors, ANOVA showed a significant difference in avDS among participants with TD, PIGD, and indeterminate participants ($F_{2,97}$=4.55, P=.01). Post hoc pairwise comparisons showed a significant difference between participants with TD and PIGD (t_{90}=2.60, P=.03) with a mean difference of 1683 (SD 646) steps/day and a moderate effect size (Cohen d=0.57) (Figure 3A).

The Student t test showed a significant difference between mild and moderate disease severity, as categorized based on MDS-UPDRS cutoffs (t_{98}=2.54, P=.01) with a mean difference of 1554 (SD 612) steps/day and moderate effect size (Cohen d=0.53) (Figure 3B).

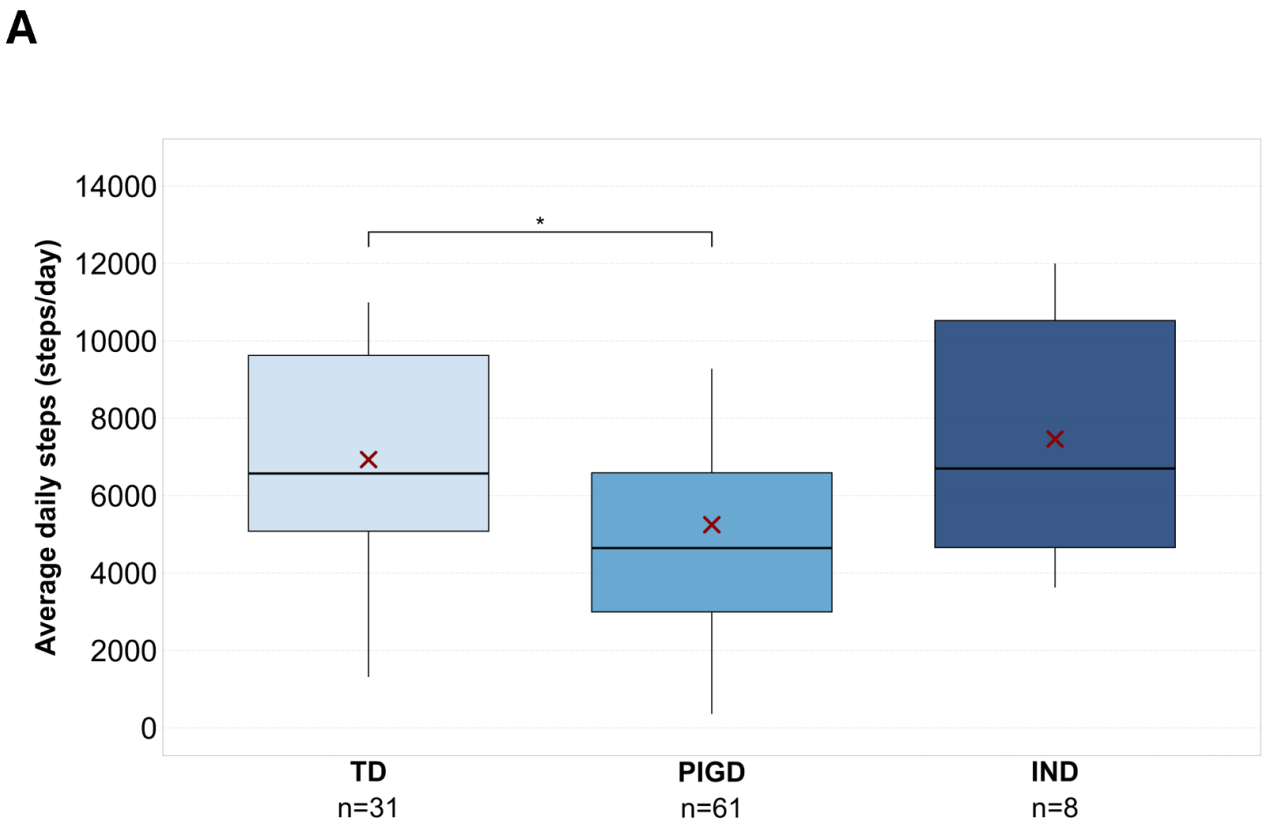
No significant differences were found between males and females (t_{98}=−0.12, P=.90).

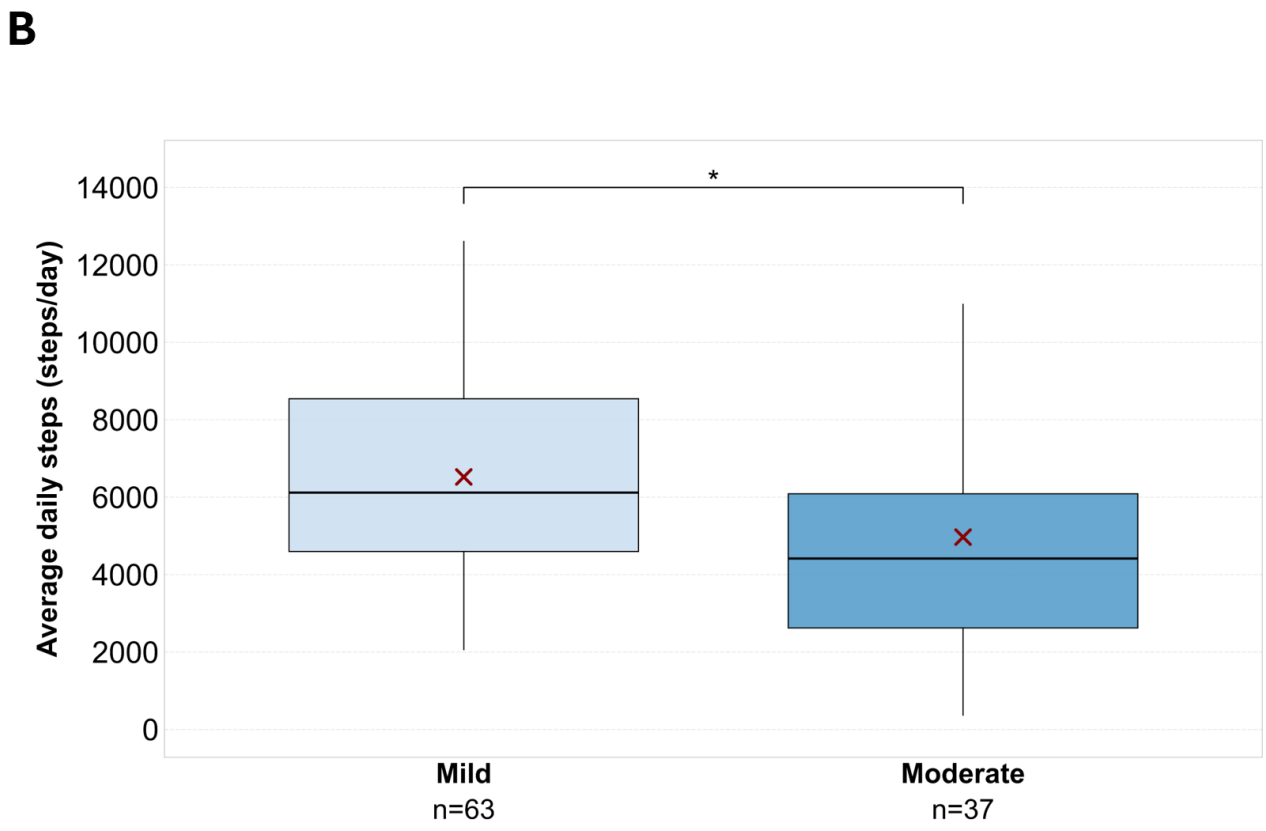
Overall, the mean MCID across the clinical and health-related variables was 1200 steps/day (range 350 - 1683), corresponding to 20% of the avDS in our population.

XSL•FO
RenderX

Figure 3. Differences across disease phenotypes, severity, and duration groups. Boxplots showing average daily steps in people with Parkinson disease across disease phenotypes (panel A) and disease severity (panel B). Thick line in the boxes indicates median; lower and upper box limits indicate first quartile (Q1) and third quartile (Q3), respectively; black vertical lines indicate lower and upper outliers boundaries calculated as Q1−(1.5×IQR) and Q3+(1.5×IQR), respectively. Red X indicates mean values for each disease group. Significant pairwise comparisons are marked by brackets and asterisks. IND: indeterminate; PIGD: postural instability and gait disorder; TD: tremor dominant. *P<.05.

XSL•FO
RenderX





XSL•FO
RenderX

Disease-Related PROs

Only MDS-UPDRS I+II showed a correlation coefficient >0.3 with avDS and was included in the regression analysis. Regression models of MDS-UPDRS part I+II were statistically significant (MDS-UPDRS I+II: $F_{1,98}$=7.03, P=.009, R^2=−0.06) (Figure 4).

For MDS-UPDRS I+II, regression equations indicated that every 1-point increase led to a reduction of 99 steps/day (avDS=7188−99×MDS-UPDRS I+II). For the sum of MDS-UPDRS I and II, the reported MCID is 5.73 points for improvement and 4.70 points for worsening; therefore, we calculated a rounded average of 6 points, and we estimated a 594 steps/day modification for each 6-point change in MDS-UPDRS I+II.

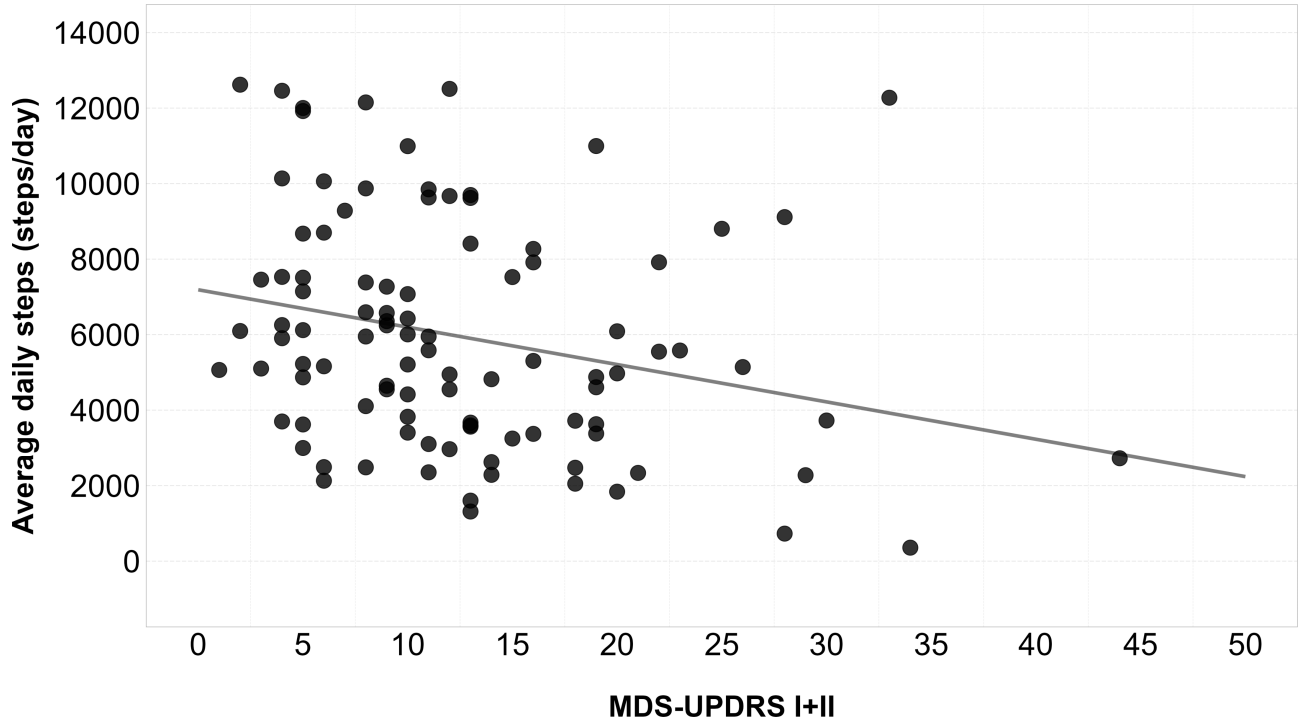
The Student t test showed a significant difference between patients reporting tremor and those not reporting it (t_{98}=−4.67, P<.001) with a mean difference of 2589 (SD 554) steps/day and a large effect size (Cohen d=−0.94) (Figure 5).

No significant differences were found in avDS between participants with and without FOG (t_{98}=0.74, P=.46), between participants with and without fatigue (t_{98}=0.61, P=.55), between participants with and without wearing-off (t_{98}=−0.39, P=.70)

Overall, the mean MCID across the disease-related PROs variables was 1592 steps/day (range 594 - 2589), corresponding to 27% of the avDS in our population.

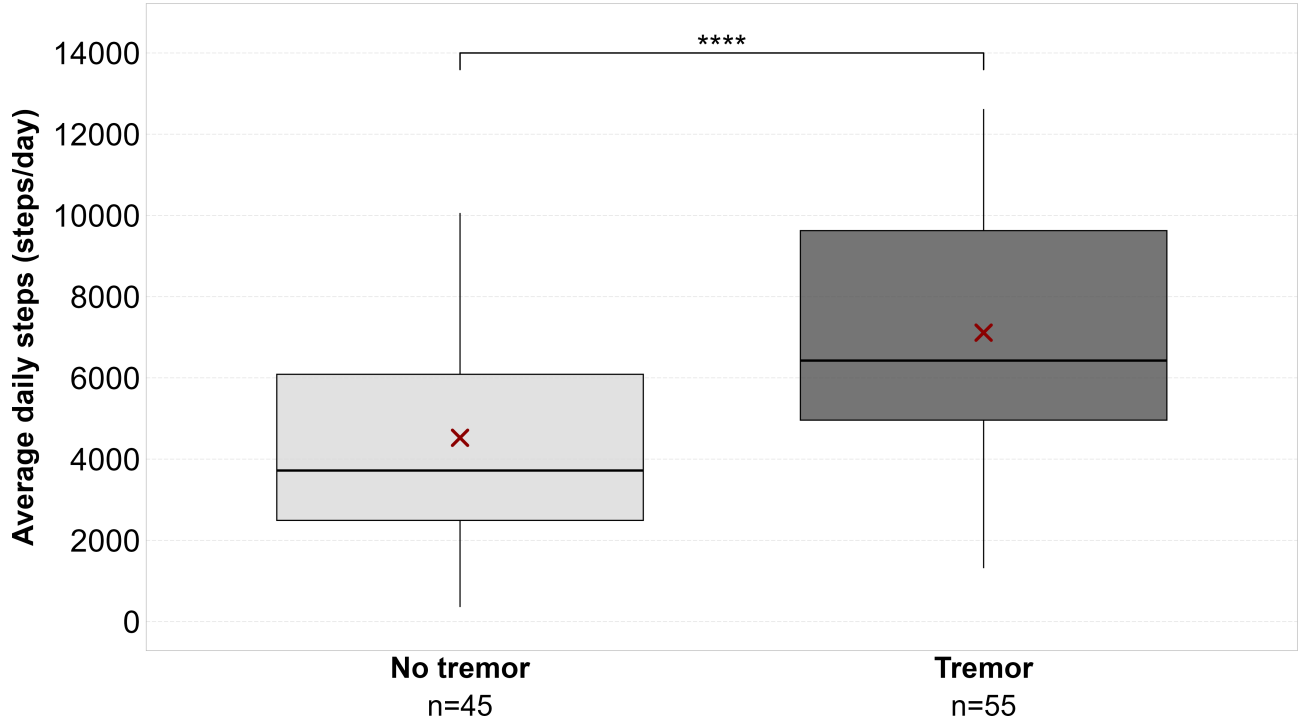
The 3 calculated MCIDs, as well as the anchors and the avDS modification related to anchors significant changes, are reported in Table 2.

Figure 4. Correlation between average daily steps and Movement Disorder Society Unified Parkinson Disease Rating Scale I+II. A linear regression line was added to show the linear tendency. MDS-UPDRS: Movement Disorder Society Unified Parkinson Disease Rating Scale.



XSL•FO
RenderX

Figure 5. Differences between people with Parkinson disease with and without tremor. Boxplots are reported showing average daily steps in people with Parkinson disease with and without tremor. Thick line in the boxes indicates median; lower and upper box limits indicate first quartile (Q1) and third quartile (Q3), respectively; black vertical lines indicate lower and upper outliers boundaries calculated as Q1−(1.5×IQR) and Q3+(1.5×IQR), respectively. Red X indicates mean values for each disease group. Significant pairwise comparisons are marked by brackets and asterisks. PD: Parkinson disease. ****P<.0001.



XSL•FO
RenderX

Table . Minimal clinically important differences for average daily steps for each of the 3 identified levels. The clinically significant anchor change and the corresponding average daily steps variation are reported next to each variable.

Anchors	avDS[a] changes	MCIDs[b]
Global mobility and motor functions		581
MDS-UPDRS-III[c] (4 points)	608	
SPPB[d] (1 point)	554	
Clinical and health-related measures		1200
mHY[e] (0.5 points)	1304	
LEDD[f] (100 mg)	350	
Age (10 y)	1410	
Disease duration (5 y)	900	
Phenotype (TD[g], PIGD[h], or IND[i])	1683	
Symptoms severity (mild or moderate)	1554	
Disease-related PROs[j]		1592
MDS-UPDRS I+II	594	
Tremor or no tremor (MDS-UPDRS 2.10)	2589	

[a]avDS: average daily steps.

[b]MCID: minimal clinically important difference.

[c]MDS-UPDRS: Movement Disorder Society Unified Parkinson Disease Rating Scale.

[d]SPPB: Short Physical Performance Battery.

[e]mHY: modified Hoehn and Yahr scale.

[f]LEDD: levodopa equivalent daily dose.

[g]TD: tremor dominant.

[h]PIGD: postural instability and gait disorder.

[i]IND: indeterminate.

[j]PRO: patient-reported outcome.

Discussion

Principal Findings

In this cross-sectional study, we aimed to calculate, for the first time, the MCID for avDS as an indicator of free-living walking behavior in people with PD using a consumer smartwatch under real-world conditions. We identified 3 distinct levels of MCID for avDS. The first level, of around 580 steps/day, is anchored to standardized measures of motor symptoms and mobility. The second level, of around 1200 steps/day, is based on clinical and health-related variables. Finally, the third level, of around 1590 steps/day, is anchored on disease-related PROs.

Globally, the MCID values ranged from 581 to 1592 steps/day, representing 10% - 27% of the average in our study population. We used the method used by Motl and colleagues [43], which offers the advantage of delineating distinct levels of clinically meaningful changes in real-world walking behavior for people with PD. These 3 thresholds provide a versatile framework for assessing the impact of clinical interventions in people with PD, depending on their specific goals and expected outcomes. Furthermore, the MCID estimates for avDS can be used either independently or in conjunction with other MCIDs, as presented in this manuscript for clinical assessment tools, to offer a more comprehensive evaluation of clinically meaningful changes in individuals with PD.

Our results suggest that a value of around 580 steps/day might reflect the smallest clinically meaningful change in avDS when measured by a consumer smartwatch and should be considered when designing a study including interventions focused on mobility or motor condition improvement. Differently, when global disease and health status are the main outcomes, our results showed that a larger value of around 1200 steps/day might be considered to define a clinically meaningful change.

In the last decades, a growing importance has been attributed to PROs, such as patient-reported scales, quality of life questionnaires, etc, as vital indicators of people's perception of their own condition [62]. PROs are often central outcomes in patients' perspectives and might capture some aspects that cannot be evaluated through objective measurements. The use of these outcomes is encouraged to guide research and clinical practice in the context of a patient-centered model of care, and several PROs were included in a recent consensus on outcome measures for PD [63]. To this end, in this paper, we calculated that a change of around 1590 steps/day might be considered to define a clinically meaningful intervention when patients' self-perception is the main outcome.

XSL•FO
RenderX

It is also vital to consider the feasibility of achieving changes in avDS of the magnitudes identified in this study. Previous research has explored interventions in people with PD that use daily steps as an outcome measure [17,64]. In 1 study by Handlery and colleagues [17], people with PD who initially took fewer than 4200 steps per day managed to increase their daily step count by approximately 1250 steps/day 6 months after participating in high-intensity treadmill exercises [17]. Similarly, another randomized controlled trial by Ellis and colleagues [64] involving 44 people with PD found that participants who initially averaged fewer than 7500 steps per day increased their avDS by about 763 steps/day over a year, following a mobile health–supported exercise program. These findings suggest that achieving changes comparable with the MCID values identified in our study is feasible through interventions aimed at promoting physical activity.

Nevertheless, the capability of consumer devices to accurately detect changes of the magnitude reported here is crucial for their clinical utility. In a recent study of ours, we found that for a monitoring duration of at least 4 days, the MDC for the Garmin Vivosmart 4 ranged between 1198 and 1524 steps/day, averaging at 1366 steps/day [36]. This MDC value is higher than the smaller MCID we calculated, indicating that any observed change exceeding this threshold is not only beyond the realm of measurement error but also clinically significant. However, this also implies that the device might not be sensitive enough to consistently detect subtler yet clinically meaningful changes in avDS, as indicated by the lower MCID limit of approximately 580 steps/day. Consequently, while the Garmin Vivosmart 4 can be effective in measuring significant differences in avDS in people with PD, for smaller anticipated changes, other instruments with higher reliability need to be considered. Future research should therefore investigate the performance of various consumer smartwatches and other devices such as smart socks, insoles, etc, and their applicability across different neurological conditions to validate, refine, and possibly extend our findings.

Limitations

We recognize certain limitations in our study. First, the people with PD included in our study displayed relatively high levels of physical capacity, as evidenced by a median SPPB score of 10 (IQR 9-12), and adequate cognitive functions, due to our exclusion of participants with a Montreal Cognitive Assessment score below 21. In addition, those with more advanced disease stages or requiring walking aids were not part of our study group. This selection criterion potentially limits the generalizability of our findings to only mild-to-moderate people with PD. Indeed, in a population of individuals with worse cognitive and motor conditions, we might expect lower accuracy and reliability of avDS estimates by smartwatches and fewer steps/day due to greater walking impairments [65-67]. Therefore, higher absolute and relative MCID values as well as different anchor combinations could be observed. Nevertheless, the sample in our study can be seen as representative of the typical target demographic for interventions using consumer-grade wearable technology. In fact, including people with PD with a disease stage greater than 3, those using walking aids, or with more severe cognitive impairments poses significant challenges in the use of consumer technology and was beyond our study's scope. Eventually, the evaluation of people using walking aids with commercial out-of-the-shelf devices, without more advanced data processing, might not even be feasible due to the potential inadequate accuracy of a smartwatch for measuring daily steps among this patient population, whose mobility characteristics differ substantially from those of individuals with physiological ambulation [68]. Future research incorporating participants with PD with lower functional scores, increased disease severity, and more pronounced cognitive impairments would be valuable.

In addition, while we incorporated a range of general and disease-specific measurements, other factors that might affect daily life mobility, such as fear of falling, frontal executive cognitive dysfunction, and objectively measured gait parameters (eg, gait speed, cadence, and step length), were not examined. Furthermore, specific subdomains of adopted measurement tools could be further explored (eg, walking speed or balance subitems of SPPB and PIGD subscore of MDS-UPDRS part III). Further studies exploring the relationship between these factors and smartwatch-recordable parameters are recommended to provide a more comprehensive understanding of mobility in people with PD.

In addition, we did not extract the raw data from the smartwatches, but used the information reported by the device's dashboard. While acknowledging the technical limitations of this approach, we designed the study to mimic as closely as possible the actual use of the device in a real-world context and to use parameters as similar as possible to those that a general user would have access to. Finally, in this work, we used an anchor-based cross-sectional method to define different levels of MCID as presented by Motl et al [43]. However, other anchor-based methods are available. In particular, a widely used method consists of the longitudinal calculation of MCID following an intervention [39,40]. However, while this latter method could allow for drawing more sound conclusions, it must be considered that it might be dependent on the type of interventions used, thus hampering the generalizability. Furthermore, previous evidence highlighted the sample-specificity of MCID estimates (particularly for distribution-based methods) [69] and the dependency on the baseline scores of the enrolled population [70]. This further stresses the importance of conducting future longitudinal, multicenter studies in a free-living context, including independent cohorts of people with PD with different clinical baseline conditions, from different clinical centers—and even different countries—and using different evaluation tools to confirm and expand the results hereby presented. In addition, since different methods to calculate MCID could lead to different results [71], using different techniques and comparing them will be important to provide the best scientific evidence on this topic. Nevertheless, we considered it relevant to provide researchers with a first indication of MCID for avDS as measured by a consumer smartwatch in order to build a framework to interpret findings in studies incorporating this outcome measure. In fact, researchers designing a study (eg, a clinical trial that aims at assessing the effectiveness of an intervention specifically designed to increase physical activity in PD) and incorporating smartwatch-based avDS as an outcome

XSL•FO
RenderX

measure of physical activity will be able to define end points that are clinically relevant, beyond statistical significance. Furthermore, based on the other included endpoints (ie, mobility, clinical indicators, or PROs), investigators will be able to select the more appropriate MCID level.

Conclusions

We reported 3 levels of MCID for avDS measured through Garmin Vivosmart 4, for 5 consecutive days, 1 of approximately 580 steps/day (or about 10% of our population's average), anchored to standardized measures of motor symptoms and mobility. A second, of around 1200 steps/day (or about 20% of the mean in our population), anchored to clinical and health-related variables. Finally, a third, or around 1590 steps/day (or about 27% of the mean in our population) anchored to disease-related PROs. Considering that the previously reported MCD for Garmin Vivosmart 4 was 1366 steps/day, this indicates that any observed change exceeding this threshold is not only beyond the realm of measurement error but also clinically meaningful. Furthermore, based on previous studies, we could suppose that achieving changes comparable with the smaller MCID identified in our study might be feasible following interventions aimed at promoting physical activity in people with PD. These findings, together with existing evidence for reliability and validity on consumer smartwatches, could be relevant for designing future clinical trials involving avDS as an outcome measure of real-world walking behavior in PD, and evaluating the effectiveness of intervention promoting free-living walking.

Acknowledgments

The authors sincerely thank all people that were involved in the data collection, from study participants to students who aided and assisted the assessments. This research was performed as part of Edoardo Bianchini's PhD (Sapienza University of Rome, Italy & University Grenoble Alpes, France) and funded through Doctoral Bursary from Sapienza University of Rome (Student 1472455). The research was supported by Sapienza University of Rome as part of the program Sapienza Research Call 2020, grant AR120172B97AF626 and Research Call 2022, grant AR222181683A15AA and by the French National Research Agency (ANR) in the framework of the Investissements d'avenir program (ANR-10-AIRT-05 and ANR-15-IDEX-02), and the MIAI Cluster (ANR-23-IACL-0006). This work also forms part of a broader translational and interdisciplinary GaitAlps research program (NV).

Data Availability

The data presented in this study are available on reasonable request from the corresponding author.

Authors' Contributions

EB, FEP, and NV designed the study. DR, MA, and SG helped with patients' recruitment and data collection. EB performed data analyses. EB, CH, FEP, and NV wrote the first draft of the manuscript. DR, MA, SG, AZ, AS, and MS reviewed manuscript draft. NV supervised the study. All authors contributed to the article and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Informed consent form.
[PDF File, 105 KB - mhealth_v13i1e64213_app1.pdf]

References

1. Mirelman A, Bonato P, Camicioli R, et al. Gait impairments in Parkinson's disease. Lancet Neurol 2019 Jul;18(7):697-708. [doi: 10.1016/S1474-4422(19)30044-4] [Medline: 30975519]
2. MacKinnon CD. Sensorimotor anatomy of gait, balance, and falls. Handb Clin Neurol 2018;159:3-26. [doi: 10.1016/B978-0-444-63916-5.00001-X] [Medline: 30482322]
3. Lee M, Noh Y, Youm C, et al. Estimating health-related quality of life based on demographic characteristics, questionnaires, gait ability, and physical fitness in Korean elderly adults. Int J Environ Res Public Health 2021 Nov 11;18(22):11816. [doi: 10.3390/ijerph182211816] [Medline: 34831575]
4. Correno MB, Hansen C, Carlin T, Vuillerme N. Objective measurement of walking activity using wearable technologies in people with Parkinson disease: a systematic review. Sensors (Basel) 2022 Jun 16;22(12):4551. [doi: 10.3390/s22124551] [Medline: 35746329]
5. Bate GL, Kirk C, Rehman RZU, et al. The role of wearable sensors to monitor physical activity and sleep patterns in older adult inpatients: a structured review. Sensors (Basel) 2023 May 18;23(10):4881. [doi: 10.3390/s23104881] [Medline: 37430796]

XSL•FO
RenderX

6. Soulard J, Carlin T, Knitza J, Vuillerme N. Wearables for measuring the physical activity and sedentary behavior of patients with axial spondyloarthritis: systematic review. JMIR Mhealth Uhealth 2022 Aug 22;10(8):e34734. [doi: 10.2196/34734] [Medline: 35994315]
7. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul;47(7):1303-1315. [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
8. Del Pozo Cruz B, Ahmadi MN, Lee IM, Stamatakis E. Prospective associations of daily step counts and intensity with cancer and cardiovascular disease incidence and mortality and all-cause mortality. JAMA Intern Med 2022 Nov 1;182(11):1139-1148. [doi: 10.1001/jamainternmed.2022.4000] [Medline: 36094529]
9. Del Pozo Cruz B, Ahmadi M, Naismith SL, Stamatakis E. Association of daily step count and intensity with incident dementia in 78 430 adults living in the UK. JAMA Neurol 2022 Oct 1;79(10):1059-1063. [doi: 10.1001/jamaneurol.2022.2672] [Medline: 36066874]
10. Paluch AE, Bajpai S, Bassett DR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022 Mar;7(3):e219-e228. [doi: 10.1016/S2468-2667(21)00302-9] [Medline: 35247352]
11. Inoue K, Tsugawa Y, Mayeda ER, Ritz B. Association of daily step patterns with mortality in US adults. JAMA Netw Open 2023 Mar 1;6(3):e235174. [doi: 10.1001/jamanetworkopen.2023.5174] [Medline: 36976556]
12. Lee IM, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 1;179(8):1105-1112. [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
13. Saint-Maurice PF, Troiano RP, Bassett DR Jr, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020 Mar 24;323(12):1151-1160. [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
14. Skidmore FM, Mackman CA, Pav B, et al. Daily ambulatory activity levels in idiopathic Parkinson disease. J Rehabil Res Dev 2008;45(9):1343-1348. [Medline: 19319758]
15. Mantri S, Wood S, Duda JE, Morley JF. Comparing self-reported and objective monitoring of physical activity in Parkinson disease. Parkinsonism Relat Disord 2019 Oct;67:56-59. [doi: 10.1016/j.parkreldis.2019.09.004] [Medline: 31621608]
16. Liu M, Gan X, Ye Z, et al. Association of accelerometer-measured physical activity intensity, sedentary time, and exercise time with incident Parkinson's disease. NPJ Digit Med 2023 Nov 28;6(1):224. [doi: 10.1038/s41746-023-00969-7] [Medline: 38017114]
17. Handlery R, Stewart JC, Pellegrini C, et al. Physical activity in De Novo Parkinson disease: daily step recommendation and effects of treadmill exercise on physical activity. Phys Ther 2021 Oct 1;101(10):pzab174. [doi: 10.1093/ptj/pzab174] [Medline: 34244805]
18. Hergenroeder AL, Barone Gibbs B, Kotlarczyk MP, Kowalsky RJ, Perera S, Brach JS. Accuracy of objective physical activity monitors in measuring steps in older adults. Gerontol Geriatr Med 2018;4:2333721418781126. [doi: 10.1177/2333721418781126] [Medline: 29977979]
19. Brewer W, Swanson BT, Ortiz A. Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med 2017;3(1):e000254. [doi: 10.1136/bmjsem-2017-000254] [Medline: 29018543]
20. Topalidis P, Florea C, Eigl ES, Kurapov A, Leon CAB, Schabus M. Evaluation of a low-cost commercial actigraph and its potential use in detecting cultural variations in physical activity and sleep. Sensors (Basel) 2021 May 29;21(11):3774. [doi: 10.3390/s21113774] [Medline: 34072347]
21. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The accuracy of smart devices for measuring physical activity in daily life: validation study. JMIR Mhealth Uhealth 2018 Dec 13;6(12):e10972. [doi: 10.2196/10972] [Medline: 30545810]
22. Bianchini E, Maetzler W. Wearable systems in movement disorders. Int Rev Mov Disord 2023;5. [doi: 10.1016/bs.irmvd.2023.04.002]
23. Powers R, Etezadi-Amoli M, Arnold EM, et al. Smartwatch inertial sensors continuously monitor real-world motor fluctuations in Parkinson's disease. Sci Transl Med 2021 Feb 3;13(579):eabd7865. [doi: 10.1126/scitranslmed.abd7865] [Medline: 33536284]
24. Chandrabhatla AS, Pomeraniec IJ, Ksendzovsky A. Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson's disease motor symptoms. NPJ Digit Med 2022 Mar 18;5(1):32. [doi: 10.1038/s41746-022-00568-y] [Medline: 35304579]
25. Löhle M, Timpka J, Bremer A, et al. Application of single wrist-wearable accelerometry for objective motor diary assessment in fluctuating Parkinson's disease. NPJ Digit Med 2023 Oct 17;6(1):194. [doi: 10.1038/s41746-023-00937-1] [Medline: 37848531]
26. Lamont RM, Daniel HL, Payne CL, Brauer SG. Accuracy of wearable physical activity trackers in people with Parkinson's disease. Gait Posture 2018 Jun;63:104-108. [doi: 10.1016/j.gaitpost.2018.04.034] [Medline: 29729611]
27. Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. Validity of different activity monitors to count steps in an inpatient rehabilitation setting. Phys Ther 2017 May 1;97(5):581-588. [doi: 10.1093/ptj/pzx010] [Medline: 28339904]

XSL•FO
RenderX

28. Fokkema T, Kooiman TJM, Krijnen WP, VAN DER Schans CP, DE Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017 Apr;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]
29. Svarre FR, Jensen MM, Nielsen J, Villumsen M. The validity of activity trackers is affected by walking speed: the criterion validity of Garmin Vivosmart® HR and StepWatchTM 3 for measuring steps at various walking speeds under controlled conditions. PeerJ 2020;8:e9381. [doi: 10.7717/peerj.9381] [Medline: 32742766]
30. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site. Gait Posture 2017 Sep;57:199-203. [doi: 10.1016/j.gaitpost.2017.06.012] [Medline: 28666177]
31. Bianchini E, Caliò B, Alborghetti M, et al. Step-counting accuracy of a commercial smartwatch in mild-to-moderate PD patients and effect of spatiotemporal gait parameters, laterality of symptoms, pharmacological state, and clinical variables. Sensors (Basel) 2022 Dec 25;23(1):214. [doi: 10.3390/s23010214] [Medline: 36616812]
32. Kim DW, Hassett LM, Nguy V, Allen NE. A comparison of activity monitor data from devices worn on the wrist and the waist in people with Parkinson's disease. Mov Disord Clin Pract 2019 Nov;6(8):693-699. [doi: 10.1002/mdc3.12850] [Medline: 31745480]
33. Wendel N, Macpherson CE, Webber K, et al. Accuracy of activity trackers in Parkinson disease: should we prescribe them? Phys Ther 2018 Aug 1;98(8):705-714. [doi: 10.1093/ptj/pzy054] [Medline: 29718452]
34. Ginis P, Goris M, De Groef A, et al. Validation of commercial activity trackers in everyday life of people with Parkinson's disease. Sensors (Basel) 2023 Apr 21;23(8):4156. [doi: 10.3390/s23084156] [Medline: 37112496]
35. Bianchini E, Rinaldi D, De Carolis L, et al. Reliability of average daily steps measured through a consumer smartwatch in Parkinson disease phenotypes, stages, and severities: cross-sectional study. JMIR Form Res 2025 Mar 18;9:e63153. [doi: 10.2196/63153] [Medline: 40101170]
36. Bianchini E, Galli S, Alborghetti M, et al. Four days are enough to provide a reliable daily step count in mild to moderate Parkinson's disease through a commercial smartwatch. Sensors (Basel) 2023 Nov 4;23(21):8971. [doi: 10.3390/s23218971] [Medline: 37960670]
37. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil 2005 Sep;84(9):719-723. [doi: 10.1097/01.phm.0000176452.17771.20] [Medline: 16141752]
38. Bloom DA, Kaplan DJ, Mojica E, et al. The minimal clinically important difference: a review of clinical significance. Am J Sports Med 2023 Feb;51(2):520-524. [doi: 10.1177/03635465211053869] [Medline: 34854345]
39. Byrom B, Breedon P, Tulkki-Wilke R, Platko JV. Meaningful change: defining the interpretability of changes in endpoints derived from interactive and mHealth technologies in healthcare and clinical research. J Rehabil Assist Technol Eng 2020;7:2055668319892778. [doi: 10.1177/2055668319892778] [Medline: 32206336]
40. Sedaghat AR. Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. Otolaryngol Head Neck Surg 2019 Oct;161(4):551-560. [doi: 10.1177/0194599819852604] [Medline: 31159641]
41. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes Res 2011 Apr;11(2):163-169. [doi: 10.1586/erp.11.12] [Medline: 21476818]
42. Zhang Y, Xi X, Huang Y. The anchor design of anchor-based method to determine the minimal clinically important difference: a systematic review. Health Qual Life Outcomes 2023 Jul 15;21(1):74. [doi: 10.1186/s12955-023-02157-3] [Medline: 37454099]
43. Motl RW, Pilutti LA, Learmonth YC, Goldman MD, Brown T. Clinical importance of steps taken per day among persons with multiple sclerosis. PLoS One 2013;8(9):e73247. [doi: 10.1371/journal.pone.0073247] [Medline: 24023843]
44. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015 Oct;30(12):1591-1601. [doi: 10.1002/mds.26424] [Medline: 26474316]
45. Goetz CG, Poewe W, Rascol O, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. Mov Disord 2004 Sep;19(9):1020-1028. [doi: 10.1002/mds.20213] [Medline: 15372591]
46. Jost ST, Kaldenbach MA, Antonini A, et al. Levodopa dose equivalency in Parkinson's disease: updated systematic review and proposals. Mov Disord 2023 Jul;38(7):1236-1252. [doi: 10.1002/mds.29410] [Medline: 37147135]
47. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008 Nov 15;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
48. Zou H, Goetz CG, Stebbins GT, Schrag A, Mestre TA, Luo S. Summing MDS-UPDRS Parts 1 + 2 (non-motor and motor experience of daily living): the patient's voice. Mov Disord 2023 Jul;38(7):1363-1364. [doi: 10.1002/mds.29417] [Medline: 37087725]
49. Stebbins GT, Goetz CG, Burn DJ, Jankovic J, Khoo TK, Tilley BC. How to identify tremor dominant and postural instability/gait difficulty groups with the Movement Disorder Society Unified Parkinson's Disease Rating Scale: comparison with the Unified Parkinson's Disease Rating Scale. Mov Disord 2013 May;28(5):668-670. [doi: 10.1002/mds.25383] [Medline: 23408503]

XSL•FO
RenderX

50. Martínez-Martín P, Rodríguez-Blázquez C, et al. Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale. Parkinsonism Relat Disord 2015 Jan;21(1):50-54. [doi: 10.1016/j.parkreldis.2014.10.026] [Medline: 25466406]
51. Peto V, Jenkinson C, Fitzpatrick R, Greenhall R. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. Qual Life Res 1995 Jun;4(3):241-248. [doi: 10.1007/BF02260863] [Medline: 7613534]
52. Tanji H, Gruber-Baldini AL, Anderson KE, et al. A comparative study of physical performance measures in Parkinson's disease. Mov Disord 2008 Oct 15;23(13):1897-1905. [doi: 10.1002/mds.22266] [Medline: 18709681]
53. Mantese CE, Schumacher-Schuh A, Rieder CRM. Clinimetrics of the 9- and 19-item wearing-off questionnaire: a systematic review. Parkinsons Dis 2018;2018:5308491. [doi: 10.1155/2018/5308491] [Medline: 29808113]
54. Siciliano M, Chiorri C, De Micco R, et al. Fatigue in Parkinson's disease: Italian validation of the Parkinson Fatigue Scale and the Fatigue Severity Scale using a Rasch analysis approach. Parkinsonism Relat Disord 2019 Aug;65:105-110. [doi: 10.1016/j.parkreldis.2019.05.028] [Medline: 31147224]
55. Nieuwboer A, Rochester L, Herman T, et al. Reliability of the new freezing of gait questionnaire: agreement between patients with Parkinson's disease and their carers. Gait Posture 2009 Nov;30(4):459-463. [doi: 10.1016/j.gaitpost.2009.07.108] [Medline: 19660949]
56. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008 Feb;61(2):102-109. [doi: 10.1016/j.jclinepi.2007.03.012] [Medline: 18177782]
57. Horváth K, Aschermann Z, Ács P, et al. Minimal clinically important difference on the Motor Examination part of MDS-UPDRS. Parkinsonism Relat Disord 2015 Dec;21(12):1421-1426. [doi: 10.1016/j.parkreldis.2015.10.006] [Medline: 26578041]
58. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006 May;54(5):743-749. [doi: 10.1111/j.1532-5415.2006.00701.x] [Medline: 16696738]
59. De Carolis L, Galli S, Bianchini E, et al. Age at onset influences progression of motor and non-motor symptoms during the early stage of Parkinson's disease: a monocentric retrospective study. Brain Sci 2023 Jan 17;13(2):157. [doi: 10.3390/brainsci13020157] [Medline: 36831700]
60. Adler CH, Beach TG, Zhang N, et al. Clinical diagnostic accuracy of early/advanced Parkinson disease: an updated clinicopathologic study. Neurol Clin Pract 2021 Aug;11(4):e414-e421. [doi: 10.1212/CPJ.0000000000001016] [Medline: 34484939]
61. Prange S, Danaila T, Laurencin C, et al. Age and time course of long-term motor and nonmotor complications in Parkinson disease. Neurology (ECronicon) 2019 Jan 8;92(2):e148-e160. [doi: 10.1212/WNL.0000000000006737] [Medline: 30541866]
62. Martinez-Martin P, Rodriguez-Blazquez C, Frades-Payo B. Specific patient-reported outcome measures for Parkinson's disease: analysis and applications. Expert Rev Pharmacoecon Outcomes Res 2008 Aug;8(4):401-418. [doi: 10.1586/14737167.8.4.401] [Medline: 20528346]
63. de Roos P, Bloem BR, Kelley TA, et al. A consensus set of outcomes for Parkinson's disease from the International Consortium for Health Outcomes Measurement. J Parkinsons Dis 2017;7(3):533-543. [doi: 10.3233/JPD-161055] [Medline: 28671140]
64. Ellis TD, Cavanaugh JT, DeAngelis T, et al. Comparative effectiveness of mHealth-supported exercise compared with exercise alone for people with Parkinson disease: randomized controlled pilot study. Phys Ther 2019 Feb 1;99(2):203-216. [doi: 10.1093/ptj/pzy131] [Medline: 30715489]
65. Hobert MA, Nussbaum S, Heger T, Berg D, Maetzler W, Heinzel S. Progressive gait deficits in Parkinson's disease: a wearable-based biannual 5-year prospective study. Front Aging Neurosci 2019;11(22):22. [doi: 10.3389/fnagi.2019.00022] [Medline: 30814947]
66. Cohen JA, Verghese J. Gait and dementia. Handb Clin Neurol 2019;167:419-427. [doi: 10.1016/B978-0-12-804766-8.00022-4] [Medline: 31753146]
67. Rosenfeldt AB, Jansen AE, Lopez-Lennon C, et al. Physical activity declines over a 12-month period in Parkinson's disease: considerations for longitudinal activity monitoring. Med Sci Sports Exerc 2025 Apr 1;57(4):738-745. [doi: 10.1249/MSS.0000000000003615] [Medline: 39589012]
68. García-Bustillo Á, Valiñas-Sieiro F, Allende-Río M, González-Santos J, Cubo E, and the Multidisciplinary Telemedicine Group. Assistive devices for personal mobility in Parkinson's disease: a systematic review of the literature. Mov Disord Clin Pract 2022 Nov;9(8):1040-1046. [doi: 10.1002/mdc3.13568] [Medline: 36339311]
69. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J Man Manip Ther 2012 Aug;20(3):160-166. [doi: 10.1179/2042618612Y.0000000001] [Medline: 23904756]
70. Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. Phys Ther 2011 May;91(5):675-688. [doi: 10.2522/ptj.20100229] [Medline: 21372203]

XSL•FO
RenderX

71. Franceschini M, Boffa A, Pignotti E, Andriolo L, Zaffagnini S, Filardo G. The minimal clinically important difference changes greatly based on the different calculation methods. Am J Sports Med 2023 Mar;51(4):1067-1073. [doi: 10.1177/03635465231152484] [Medline: 36811558]

Abbreviations

AvDS: average daily steps
FOG: freezing of gait
LEDD: levodopa equivalent daily dose
MCID: minimal clinically important difference
MDC: minimum detectable change
MDS-UPDRS: Movement Disorder Society Unified Parkinson's Disease Rating Scale
mHY: modified Hoeh and Yahr scale
PD: Parkinson disease
PIGD: postural instability and gait disorder
PRO: patient-reported outcome
SPPB: Short Physical Performance Battery
TD: tremor dominant

Edited by L Buis; submitted 11.07.24; peer-reviewed by R Handlery, S Okita; revised version received 05.12.24; accepted 23.05.25; published 29.07.25.

Please cite as:
Bianchini E, Alborghetti M, Galli S, Hansen C, Zampogna A, Suppa A, Salvetti M, Pontieri FE, Rinaldi D, Vuillerme N
Minimal Clinically Important Difference of Average Daily Steps Measured Through a Consumer Smartwatch in People With Mild-to-Moderate Parkinson Disease: Cross-Sectional Study
JMIR Mhealth Uhealth 2025;13:e64213
URL: https://mhealth.jmir.org/2025/1/e64213
doi:10.2196/64213

© Edoardo Bianchini, Marika Alborghetti, Silvia Galli, Clint Hansen, Alessandro Zampogna, Antonio Suppa, Marco Salvetti, Francesco Ernesto Pontieri, Domiziana Rinaldi, Nicolas Vuillerme. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparison of ActiGraph CentrePoint Insight Watch Placement on Dominant and Nondominant Wrists in Young Adults in Free-Living Conditions: Observational Validation Study

Daehyoung Lee[1], PhD; Haley Voermans-Dean[2,3], BSc; Jung Eun Lee[2], PhD; Jong Cheol Shin[4], PhD; Gregory Dominick[1], PhD

[1]Department of Health Behavior and Nutrition Sciences, University of Delaware, 26 North College Avenue (013 Carpenter Sports Building), Newark, Delaware, United States

[2]Department of Applied Human Sciences, University of Minnesota Duluth, Duluth, Minnesota, United States

[3]Palmer College of Chiropractic, Davenport, Iowa, United States

[4]Department of Community and Population Health, Lehigh University, Bethlehem, PA, United States

Corresponding Author:
Daehyoung Lee, PhD
Department of Health Behavior and Nutrition Sciences, University of Delaware, 26 North College Avenue (013 Carpenter Sports Building), Newark, Delaware, United States

Abstract

Background: With the continuous evolution of technology, wearable accelerometers have become one of the most popular means of measuring daily physical activity (PA) levels. Despite the conventional use of the nondominant wrist as a device placement in numerous PA studies, the impact of wrist-worn accelerometer placement on PA data outcomes remains uncertain.

Objective: This study aimed to examine the degree of agreement between accelerometry data collected from CentrePoint Insight Watches (CPIWs; ActiGraph) worn on the dominant and nondominant wrists of young adults in free-living conditions.

Methods: Twenty-nine participants (mean age 20.2, SD 1.6 years; 23 females) simultaneously wore an ActiGraph CPIW on both dominant and nondominant wrists for 7 consecutive days during waking hours. A sampling frequency of 32 Hz and Montoye 2020 cut-points were used to categorize activity intensity based on counts per minute. Data validity criteria included (1) ≥600 minutes per day of monitor wear time for both wrists, (2) a daily wear time difference of <1% of the average wear time between the dominant and nondominant wrists, and (3) a minimum of 3 valid days of monitor wear for both wrists. Bland-Altman plots and intraclass correlation coefficient (ICC) analyses were performed to compare the accelerometry data between the two device placements.

Results: Average daily monitor wear time was 789.6 (SD 86.1) minutes per day for the dominant wrist and 793.0 (SD 91.8) minutes per day for the nondominant wrist. All accelerometer variables, including sedentary time (ST), light PA, moderate-to-vigorous PA (MVPA), steps, triaxial counts, and vector magnitude (VM), showed good-to-excellent levels of reliability between the two measurements (ICC >0.88 for all; P<.001). Bland-Altman analysis calculated mean bias and SD between the two device placements as follows: ST (−18.8, SD 27.6 min/d), light PA (2.7, SD 15.9 min/d), MVPA (12.7, SD 26.7 min/d), steps (218.1, SD 476.6 counts/d), x-axis (99.4, SD 188.8 counts/min), y-axis (73.9, SD 147.0 counts/min), z-axis (107.6, SD 183.5 counts/min), and VM (161.2, SD 273.4 counts/min). Bland-Altman plots revealed that the upper and lower limits of agreement across most variables were considerably wide.

Conclusions: Our findings partially align with previous research, demonstrating higher MVPA and step counts on the dominant wrist, while the nondominant wrist produced a higher level of ST. Despite the acceptable level of reliability between the two placements based on ICC analyses, the dominant wrist tended to produce greater outcomes as the intensity of PA increased, highlighting the need for careful consideration when determining the wear location of CPIWs and interpreting data outcomes.

(JMIR Mhealth Uhealth 2025;13:e63033) doi:10.2196/63033

KEYWORDS

wrist-worn accelerometry; physical activity; sedentary time; device placement; measurement

XSL•FO
RenderX

Introduction

Participation in regular physical activity (PA) contributes substantial benefits to overall health, such as weight control, brain health, bone and muscle strengthening, and emotional regulation [1,2]. For this reason, PA levels primarily serve as an informative gauge of general health status for both the public and health care practitioners [3,4]. The 2018 National PA Guidelines for Americans suggest that adults should participate in at least 150 minutes of moderate-to-vigorous PA (MVPA) or 75 minutes of vigorous-intensity PA per week, along with additional muscle-strengthening activities, to achieve health benefits [5]. Accurately quantifying PA levels is a key component in developing and analyzing public health initiatives to promote regular PA participation in the general population [6,7]. Among numerous approaches for PA measurement and estimation, accelerometer-derived outcomes not only help individuals objectively recognize their PA levels but also build foundational knowledge for organizations and communities to develop health intervention plans for specific populations [8,9].

With the continuous evolution of technology, sensor-based ecological measurement has emerged, and wearable activity trackers have become one of the most popular means of measuring daily PA levels [10]. Global health technology companies, including Apple, Garmin, and Fitbit, have revolutionized the market for commercially available activity or fitness-tracking devices by making them affordable, easily accessible, and compatible with other electronic devices and applications [11]. Despite their widespread use, commercial wearable devices have shown limited accuracy in estimating PA levels in free-living environments, largely due to relatively low intradevice reliability and limited criterion validity. Most available studies have been conducted in controlled laboratory settings, focusing primarily on basic metrics, such as step counts and distance, rather than comprehensive assessments of PA parameters [12]. Moreover, their measurement accuracy varies significantly by device brand and type, and the criterion validity of comparative data is relatively poor [12,13]. ActiGraph triaxial accelerometers (ActiGraph LLC) have been used extensively by researchers across the world to estimate various PA parameters across diverse clinical and nonclinical populations, with high reliability and validity in both laboratory and free-living conditions [14]. Although limited affordability and a lack of real-time feedback are recognized as drawbacks compared to consumer-level devices [15], ActiGraph accelerometers remain among the most widely used research-grade instruments for PA measurement [16].

Due to poor compliance of wear time during the measurement period, which often results in inaccurate classification or misrepresentation of physical behaviors (eg, overrepresentation of MVPA or underrepresentation of sedentary time [ST]), an increasing number of research studies are exploring the use of wrist-worn accelerometers instead of waist-worn devices [17]. Nonetheless, there is a lack of agreement in current literature as to how the placement of wrist-worn devices (eg, dominant vs nondominant or right vs left) affects the accuracy and interdevice reliability of accelerometer data outcomes. A few studies have used earlier models of ActiGraph triaxial accelerometers (eg, GT3X, GT3X+, and GT9X) to compare specific PA outcomes collected from both the dominant and nondominant wrists in free-living settings [18,19]. Buchan et al [18] reported no significant differences between the two device placements for wear time and MVPA; however, they monitored the behavioral outcomes during waking hours for only 1 day. Park et al [20] also compared the two device placements during waking hours in a single-day, free-living setting and found that the dominant wrist tended to yield higher step counts compared with the nondominant wrist. Interestingly, a similarly designed study using comparable data collection protocols (eg, 24-hour monitoring) revealed that there was no substantial difference in PA levels measured by ActiGraph GT3X accelerometers on both wrists [19]. A recent systematic review also examined how ActiGraph accelerometry compares to popular commercial wrist-worn devices for estimating step counts, energy expenditure, and heart rate, highlighting the dearth of validity evidence for their usability in free-living settings [13]. Few attempts have been made to investigate the impact of accelerometer placement on measurement outcomes in free-living conditions using research-grade accelerometry with a 7-day monitoring protocol, which can significantly increase the reliability of predicting and representing an individual's habitual behaviors [21,22]. A recent study by Buchan et al [7] compared 7-day wear time between the dominant and nondominant wrists. The results suggested that the intensity gradient may be equivalent between both wrists, but average acceleration was slightly higher on the dominant wrist.

Given the increasingly dominant use of wearable technologies to monitor and estimate PA levels, it is important to clarify the discrepancy in the literature and verify the conventional use of the nondominant wrist as the preferred device placement in PA measurement studies. The objective of this study was to examine the degree of agreement between PA data collected from accelerometers worn on the dominant and nondominant wrists of young adults in free-living environments. This study used one of ActiGraph's newest models, the CentrePoint Insight Watch (CPIW; ActiGraph LLC), which features a unique user-centered band design and high-resolution raw acceleration data that may improve participant comfort and data reliability. In addition, the relatively new, cross-validated vector magnitude (VM) count cut-points by Montoye et al [23] were used to determine and classify activity intensity, improving the practical implications of the study findings.

Methods

Participants

Twenty-nine young adults (mean age 20.2, SD 1.6 years; 23, 79.3% females) from a Midwestern university in the United States were recruited for this study. Flyers were posted around the university campus during the spring term, and prospective participants were encouraged to contact the research team for the eligibility screening. Inclusion criteria were as follows: (1) aged 18 - 25 years, (2) ability to follow study protocols and independently provide informed consent, and (3) walking as a primary form of ambulation without the use of assistive devices.

XSL•FO
RenderX

Individuals with a mobility impairment (eg, use of a wheelchair or crutches) or a history of cardiovascular disease were excluded from this study. Prospective participants were invited to an individual meeting with the research team, during which study procedures were thoroughly reviewed for all participants.

Procedures and Instruments

Self-administered demographic and general health information was obtained from the participants with the assistance of a trained student investigator. This included date of birth, biological sex, height, weight, and hand dominance, which were used to initialize the ActiGraph CPIWs (dimensions: 1.97×1.35×0.41 in and weight: 14 g) for each participant. Prior to the accelerometry data collection, participants attended a training session on how to wear and care for the device. Specifically, the research team emphasized wearing and removing both accelerometers simultaneously and placing them high on their wrists with appropriate tightness to prevent slippage during movement or daily activities. They were asked to ensure the accelerometers were placed at the same relative positions on the dorsal aspect of their wrists, as variations in proximity to the hand could affect acceleration and activity counts. Participants were also asked to carry out their normal activities of daily living while wearing the accelerometers. To avoid the risk of water damage to the devices, participants were asked to take the devices off during any water-based activities, such as swimming and bathing, as well as during sleep. Participants wore an ActiGraph CPIW on both their dominant and nondominant wrists during waking hours for 7 consecutive days, for at least 10 hours per day. The watch displayed only the date and time without any interpretive information on participants' PA levels, which is an intentional design feature to minimize reactivity to accelerometer-based measurement [24]. In addition, remote synchronization and notification features of the ActiGraph CPIW were disabled to avoid influencing participants' habitual behaviors and PA engagement. ActiGraph CPIW enables PA data collection for up to 30 days without charging, and thus, participants were not instructed on how to charge the devices.

For this study, accelerometers were programmed to record acceleration data at a default sampling frequency of 32 Hz using 60-second epochs, which were subsequently converted to counts per minute (CPM) based on the Montoye et al [23] VM count cut-points to classify the activity intensity. Specifically, the CPM cut-points included sedentary (<2860 CPM), light PA (2860 - 3940 CPM), and MVPA (≥3941 CPM). Accelerometer activity count of three axes (ie, vertical, horizontal, and mediolateral) is a measure that indicates quantified acceleration within a designated epoch length [25,26]. For data processing, accelerometer data were first saved in the CentrePoint portal (ActiGraph LLC) and later exported to ActiLife (version 6.14.0; ActiGraph) to validate the wear time and score the collected datasets based on the predetermined CPM cut-points. The wear criteria threshold for both wrists was set at ≥600 minutes per day, and nonwear time was defined as ≥90 minutes of consecutive zero counts with an allowance of 2-minute spike tolerance [27]. After the 7-day data collection period, participants attended a debriefing meeting with the research team to return the accelerometers. In addition, perceived PA level was assessed using a 4-category classification approach outlined in the Physical Activity Guidelines for Americans (PAGA), 2nd edition, which includes the following PA levels: inactive, insufficiently active, active, and highly active [5].

Statistical Analysis

PA as a variable was measured using the following parameters: step counts per day, time spent at each PA intensity, including light PA and MVPA (min/d), activity counts from 3 axes, and VM CPM. VM was calculated using the following formula: $|V| = \sqrt{(\text{Axis } X^2+\text{Axis } Y^2+\text{Axis } Z^2)}$, to describe in-depth analyses of sampled accelerations of movement detected from both device placements. Considering that reactivity to accelerometer measurement is most pronounced at the beginning of the PA data collection [24], the first day of the 7-day accelerometry data was intentionally removed from the analyses. Data validity criteria included: (1) ≥600 minutes of monitor wear time per day for both wrists, (2) a daily wear time difference of less than 1% of the average wear time between the dominant and nondominant wrists, and (3) a minimum of 3 valid days of monitor wear for both wrists during the same periods. Of the 30 participants who completed the entire study, 29 participants met all 3 data validity criteria and were included in the analyses. One participant was excluded for having only one valid day of monitor wear. Each participant generated 3-6 valid days of monitor wear, and the total number of valid days included in the analyses was 153. We calculated the average of each data metric across valid days for each participant and compared the data outputs from the 2 device locations on the individual level. Paired samples *t* tests were performed to examine differences in PA and ST data collected from dominant and nondominant wrists. Bland-Altman plots and intraclass correlation coefficients (ICCs) were used to determine the degree of agreement between the two device placements. The ICC value ranges from 0 to 1, where <0.5=poor, 0.5 - 0.74=moderate, 0.75 - 0.9=good, and >0.9=excellent reliability [28]. In addition, linear regression analyses were conducted to assess the relationship between the magnitude of measurements and individual bias. The coefficient of determination (R^2) was interpreted as follows: ≤0.01=small, ≈0.09=medium, and ≥0.25=large effect size [29]. As a follow-up analysis, the equivalence test, also known as the two one-sided test (TOST), was performed to determine whether the means of the two different measures were equivalent within a 5% margin and whether any found differences were clinically significant. To conclude significant equivalence within the specified margin, both one-sided tests (ie, upper and lower bounds represented as t_1 and t_2) must reject the null hypothesis of nonequivalence (ie, unequal to zero). All statistical analyses were performed using IBM SPSS Statistics version 28, and the significance level was set at $P<.05$.

Ethical Considerations

Written informed consent was obtained from all participants prior to their participation in this study. All participants received a US $20 prepaid debit card as an incentive at the end of their study participation. The institutional review board at the University of Minnesota approved this study (STUDY00017556).

XSL•FO
RenderX

Results

Participant characteristics are shown in Table 1. The average weight of the participants was 63.6 (SD 8.3; males: 70.7, SD 5.1; females: 61.8, SD 8.0) kg. Except for 2 (6.9%), most participants (n=27; 93.1%) were nonsmokers at the time of study participation. Of the 29 participants, 2 (6.9%) indicated left as their dominant wrist or hand. Average daily monitor wear time was 789.6 (SD 86.1) minutes per day for the dominant and 793.0 (SD 91.8) minutes per day for nondominant wrists, respectively. The average number of valid monitor wear days was 5.3 (SD 1.3; median 6, IQR 2) days.

The descriptive accelerometry data between dominant and nondominant wrists and the results of the paired samples *t* tests are shown in Table 2. According to the ICC analyses, all accelerometer variables, including ST, light PA, MVPA, steps, triaxial counts, and VM, showed good to excellent levels of reliability between the two device placements (ICC >0.88 for all; *P*<.001; refer to Table 3).

Table . Sample characteristics (N=29).

Characteristic	Value
Age (years), mean (SD)	20.2 (1.6)
Female, n (%)	23 (79.3)
Weight (kg), mean (SD)	63.6 (8.3)
Height (cm), mean (SD)	168.3 (10.2)
BMI ($kg*m^{-2}$), mean (SD)	22.5 (2.9)
Ethnicity, n (%)	
White	24 (82.8)
African American	0 (0)
Asian	3 (10.3)
Hispanic	2 (6.9)
Class standing, n (%)	
Freshman	4 (13.8)
Sophomore	7 (24.2)
Junior	15 (51.7)
Senior	3 (10.3)
Proximity to campus, n (%)	
On campus	5 (17.2)
Within 1 mile	16 (55.2)
Within 3 miles	6 (20.8)
Within 5 miles	1 (3.4)
More than 5 miles	1 (3.4)
Perceived PA[a] level, n (%)	
Highly active	17 (58.6)
Active	9 (31.1)
Insufficiently active	3 (10.3)
Inactive	0 (0)

[a]PA: physical activity.

XSL•FO
RenderX

Table . Descriptive accelerometry data from dominant and nondominant wrists (N=29).

Variable	Dominant, mean (SD)	Nondominant, mean (SD)	Average, mean (SD)
ST[a] (min/d)	526.9 (89.6)	545.7 (93.2)[b]	536.3 (91.1)
LPA[c] (min/d)	154.6 (33.9)	151.9 (39.0)	153.3 (36.3)
MVPA[d] (min/d)	108.1 (44.1)	95.4 (41.3)[b]	101.8 (42.8)
Steps (per day)	8304.6 (3337.5)	8086.1 (3193.9)[b]	8195.5 (3239.6)
x-axis (CPM[e])	1601.6 (401.6)	1502.2 (389.9)[b]	1551.9 (395.5)
y-axis (CPM)	1740.8 (570.7)	1666.9 (572.7)[b]	1703.9 (567.9)
z-axis (CPM)	1666.8 (342.0)	1559.2 (365.4)[b]	1613.0 (355.0)
VM[f] (CPM)	2911.1 (727.5)	2749.8 (740.8)[b]	2830.5 (732.2)

[a]ST: sedentary time.

[b]Statistically significant difference compared to dominant wrist; $P<.05$.

[c]LPA: light physical activity.

[d]MVPA: moderate-to-vigorous physical activity.

[e]CPM: counts per minute.

[f]VM: vector magnitude.

Table . Intraclass correlation coefficients between dominant and nondominant accelerometer measures (N=29).

Variable	ICC[ab] (95% CI)	*F* test (*df*)	*P* value
ST[c] (min/d)	0.97 (0.88 - 0.99)	42.88 (28)	<.001
LPA[d] (min/d)	0.95 (0.90 - 0.98)	20.01 (28)	<.001
MVPA[e] (min/d)	0.88 (0.70 - 0.94)	9.23 (28)	<.001
Steps (per day)	0.99 (0.98 - 0.99)	186.87 (28)	<.001
x-axis (CPM)[f]	0.93 (0.81 - 0.97)	16.58 (28)	<.001
y-axis (CPM)	0.98 (0.95 - 0.99)	59.46 (28)	<.001
z-axis (CPM)	0.91 (0.74 - 0.96)	13.88 (28)	<.001
VM[g] (CPM)	0.95 (0.86 - 0.98)	27.85 (28)	<.001

[a]Average measures with a two-way mixed effects model were used.

[b]ICC: intraclass correlation coefficient.

[c]ST: sedentary time.

[d]LPA: light physical activity.

[e]MVPA: moderate-to-vigorous physical activity.

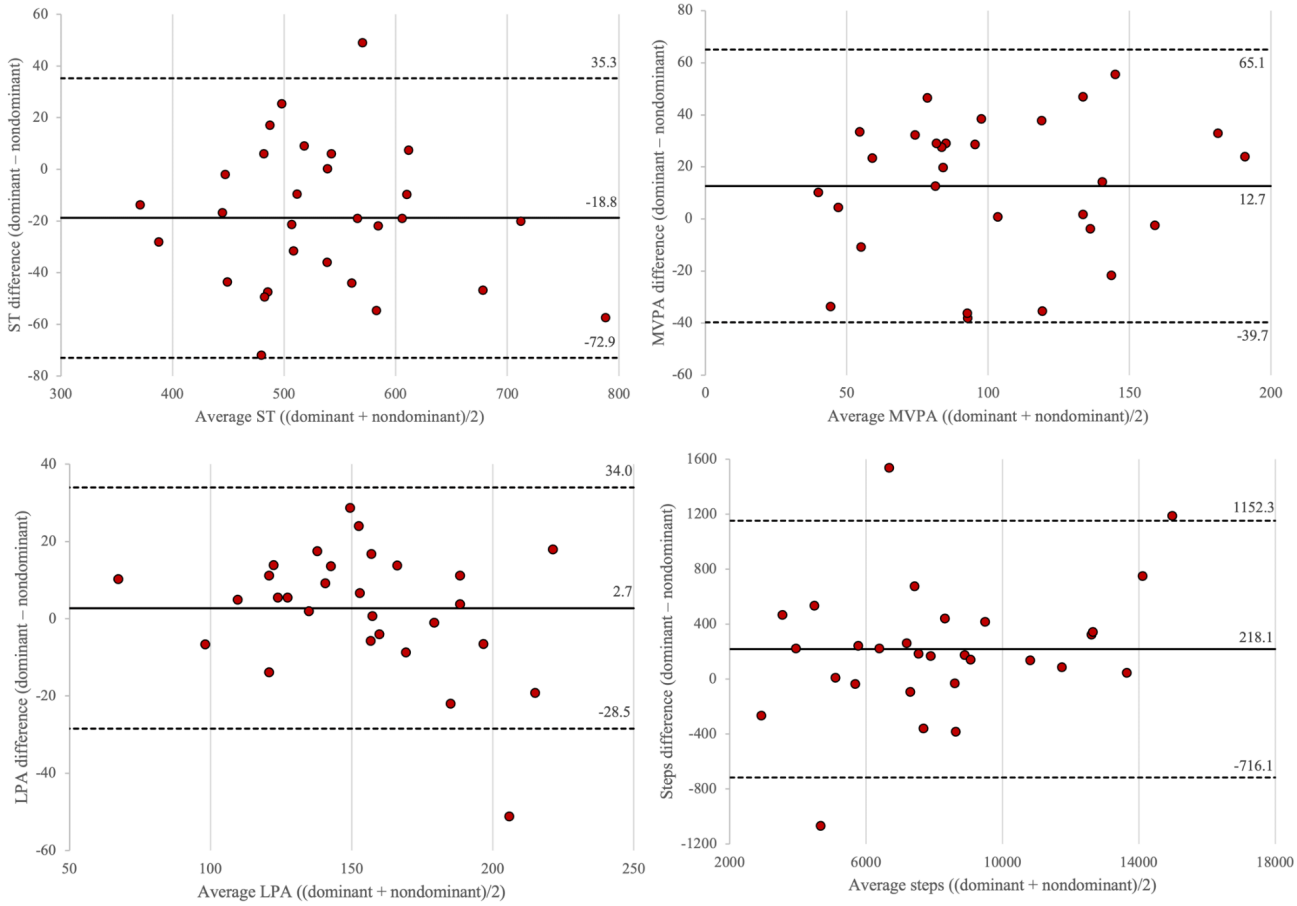
[f]CPM: counts per minute.

[g]VM: vector magnitude.

Bland-Altman analysis calculated mean bias and SD between the two device placements as follows: ST (−18.8, SD 27.6 min/d), light PA (2.7, SD 15.9 min/d), MVPA (12.7, SD 26.7 min/d), steps (218.1, SD 476.6 counts/d), x-axis (99.4 SD 188.8 counts/min), y-axis (73.9, SD 147.0 counts/min), z-axis (107.6 SD 183.5 counts/min), and VM (161.2, SD 273.4 counts/min). Figures 1 and 2 display individual-level biases with upper and lower limits of agreements. It is estimated to have proportional bias when there is a linear tendency in the discrepancies between the two measures, which is often represented by the slope of the regression or R^2 that significantly differs from zero [30]. Proportional bias was considered present in steps (R^2=0.091) and light PA (R^2=0.105), given the medium effect sizes for the regression coefficients, coupled with the progressive divergence in the generated plots. However, follow-up TOSTs with a 5% equivalence margin revealed that none of the accelerometry variables showed statistically significant equivalence between the means of two device placements: t_1 and t_2 for ST=0.33 and −1.90; light PA=1.08 and −0.52; MVPA=1.59 and 0.68; steps=0.73 and −0.22; x-axis=1.70 and 0.21; y-axis=1.06 and −0.08; z-axis=2.03 and 0.29; and VM=1.57 and 0.10 (all $P>.05$).

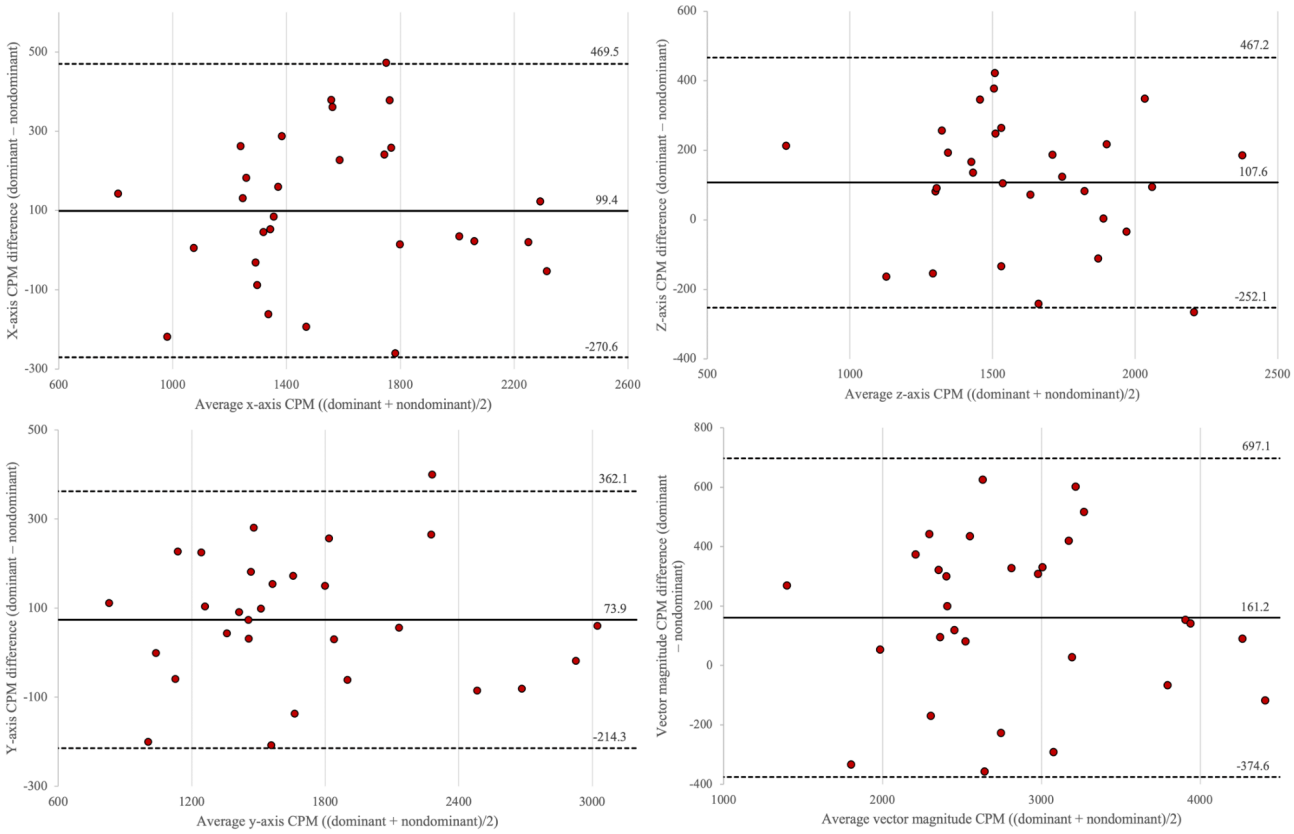
XSL•FO
RenderX

Figure 1. Bland-Altman plots were used to assess data agreement between the dominant and the nondominant wrists in measuring sedentary time, light physical activity, moderate-to-vigorous physical activity (min/d), and steps per day. Each red circle represents the individual-level difference between the two device placements for the 29 participants. Two dotted black lines indicate the limits of agreement (mean difference ±1.96× SD of the differences), and a solid black line indicates the mean bias between the two measures. LPA: light physical activity; MVPA: moderate-to-vigorous physical activity; ST: sedentary time.



XSL•FO
RenderX

Figure 2. Bland-Altman plots were used to assess data agreement between the dominant and the nondominant wrists in measuring x, y, and z axes, as well as vector magnitude (counts per minute). Each red circle represents the individual-level difference between the two device placements for the 29 participants. Two dotted black lines indicate the limits of agreement (mean difference ±1.96× SD of the differences), and the solid black line indicates the mean bias between the two measures. CPM: counts per minute.



Discussion

Principal Findings

The objective of this study was to examine the degree of agreement between accelerometry data collected from ActiGraph CPIWs worn on the dominant and nondominant wrists of young adults in free-living conditions. This study comprehensively compared accelerometry variables, including ST, light PA, MVPA, and steps, as well as activity counts of 3 axes and VM, between the two device placements. Despite statistically significant differences observed through paired *t* tests, all outcome variables between the two placements showed good-to-excellent levels of comparability. However, the limits of agreement were considerably wide across all variables except ST, highlighting the need for cautious interpretation of these mixed findings.

Overall, previous research comparing PA data collected from the dominant and nondominant wrists revealed mixed results with limited consistency. A recent study examined PA data from ActiGraph GT9X accelerometers worn on both wrists for a 1-day data collection period and reported that the dominant wrist consistently produced higher step counts than the nondominant wrist in free-living conditions [20]. This result aligns with our findings, wherein the dominant wrist produced relatively higher outcomes of MVPA, light PA, and steps than the nondominant wrist. Conversely, an earlier study used a nearly identical protocol and found that device placement had no impact on the average daily PA outcomes, regardless of the axis examined in free-living environments [19]. Our study partly supports this finding, as all accelerometer variables, including the 3 axes and VM, exhibited excellent ICC values between the two measurements. However, proportional bias was considered present in steps (R^2=0.091) and light PA (R^2=0.105). This implies that the dominant wrist tended to produce higher step counts as the average step level increased, while the opposite trend was predicted for light PA. Given that 7 consecutive days of monitor wear are recommended for investigating representative and predictable PA behaviors, differences in wear periods (eg, 1 d vs 7 d) and the use of different ActiGraph models (eg, GT9X and GT3X vs CPIW) may contribute to the inconsistent results [31,32].

As explained by the paired *t* test results, MVPA, steps, and all axis-related variables showed statistically significant differences between the two placements, in which the dominant wrist produced relatively higher levels of outcomes compared to the nondominant wrist. Interestingly, the opposite result was observed in ST, whereby the nondominant wrist produced a higher level of outcome than the dominant wrist. Montoye et al [33] compared research-grade accelerometers (ie, GENEActiv; Activinsights Ltd) worn on both wrists for measuring PA and sedentary behavior. Their findings underscored higher sensitivities and specificities (eg, 97%) in predicting sedentary behavior as well as low-intensity activities using the left wrist-worn accelerometer (left hand was indicated as the nondominant hand among most participants), compared to the right wrist-worn accelerometer. Meanwhile, the right

XSL•FO
RenderX

wrist-worn accelerometer (mostly the dominant wrist among participants) produced significantly lower sensitivity and specificity for MVPA, which was overly misclassified as light PA due to underestimation of MVPA. Similarly, Fraysse et al [34] used GENEActiv accelerometers to compare the classification accuracy between dominant and nondominant wrists in free-living environments. While their sample consisted of older adults (ie, ≥70 years of age), significantly higher ST and lower MVPA were linked to the dominant wrist compared to the nondominant wrist. Once again, these mixed results may be attributed to the use of different accelerometer brands (eg, ActiGraph vs GENEActiv) that can result in varying raw acceleration outputs [35]. More comprehensive cross-validation studies are needed to determine the degree of agreement among a variety of research-grade accelerometer brands and models according to specific movement contexts and populations.

It should be noted that the paired *t* test considers only the overall bias of two measures; thus, this approach may not fully address how rejecting the null hypothesis corresponds to the degree of agreement between the two measurements [36]. As such, we implemented follow-up TOSTs with an equivalence margin of 5% to determine whether the observed difference between the two device placements is clinically significant. None of the included variables rejected the null hypothesis of nonequivalence for both one-sided tests, indicating insufficient evidence to declare statistically significant equivalence between the two placements within the specified margin. In line with the TOST results, ICC values for all accelerometer measures indicated good-to-excellent correlation (>0.88), suggesting strong agreement between the two placements. Moreover, Bland-Altman plots of all variables showed that the individual differences between the dominant and nondominant wrists mostly fell within the limits of agreement, which highlights the comparable estimations of variables between the two placements [37]. However, given the volume of average accelerometry outcomes (ie, ST=536.3 min/d, light PA=153.3 min/d, and MVPA=101.8 min/d), the discrepancy between the two measurements may not be at the acceptable level. The mean bias for MVPA (12.7 min/d) was comparable to that for ST (−18.8 min/d), indicating lower interdevice comparability in estimating MVPA. Wrist-worn devices are often recommended to capture lower-intensity activities, especially those mainly involving upper-body movements such as household chores [38]. Therefore, our findings emphasize the practical comparability (eg, sedentary-to-light activities vs MVPA) between the two device placements in free-living conditions, in which the intensity of the activity determines the degree of agreement or discrepancy between the two placements: the higher the intensity of PA, the greater the discrepancy in outcomes between the dominant and nondominant wrists.

The current and previous findings underscore the need for cautious interpretation of ICC and Bland-Altman plots when comparing the PA data of multiple measurements [39,40]. It is critical to note that although the mean bias for most accelerometry variables may seem marginal, the limits of agreement between the two measurements were considerably wide for PA variables, spanning over 104.8 minutes per day range for MVPA (−39.7 to 65.1 min/d), nearly 2000 steps per day for walking steps (−716.1 to 1152.3 steps/d), and more than 1000 VM CPM per day (−374.6 to 697.1 CPM/d). A similar result was observed by Pfister et al [39], in which the research team examined the comparability between the waist-worn ActiGraph GT3X+ and the thigh-worn activPal3 (PAL Technologies Ltd). While the ICCs for all PA and ST variables between the two devices were generally high and consistent, statistically significant—albeit small—differences in MVPA were found between the GT3X+ VM, vertical axis, and activPal3 at the group level. Further, the limits of agreement between the two accelerometer types were generally wide at the individual level, restricting the interchangeable use of these devices [39]. Regardless of the ICC analysis results, the degree of agreement may need to be determined by the overall mean bias in comparison to the range of the limits of agreement.

When it comes to the objective measurement of physical behaviors, the anterior waist or hip has been a conventional placement in earlier studies. Karaca et al [41] compared step counts measured by wGT3X-BT (ActiGraph) in five different body locations, including both right and left wrists, and found that the waist and right upper arm showed low mean absolute percentage error compared to other placements. In addition, waist-worn accelerometers tended to provide more accurate estimates of sedentary behavior that engaged the lower extremities exclusively, compared to wrist-worn devices [42]. However, the shift to using wrist-worn accelerometers over the past decade has demonstrated relative advantages over standard waist-worn devices. Despite the lower-performing algorithm compared to waist or hip location and a lack of criterion validity, particularly in free-living settings, wrist-worn devices can enhance participant acceptability given their familiarity with a wristwatch and enable increased data collection within 24-hour movement and sleep studies [43,44]. The use of the wrist as a device placement is often preferred over the waist because wrist-worn devices are generally less intrusive and often result in higher compliance with study protocol [45,46]. Although there is a lack of scientific consensus on cut-points for categorizing activity intensities in wrist-worn accelerometers, including CPIW, it is understood that devices on the nondominant wrist are likely to ensure higher measurement accuracy than those on the dominant wrist in free-living settings [33]. Given the high variability of movement and participants' unfamiliarity with wearing devices on the dominant wrist, the nondominant wrist may be a more reliable choice for assessing levels of PA and ST in real-world settings.

Limitations and Strengths

It is important to acknowledge the methodological limitations that may have affected the interpretation of our findings. The use of wrist-worn accelerometers may lead to underestimation or overestimation of PA levels, as they are limited in detecting nonambulatory activities (eg, cycling and yoga) and distinguishing between varying postures (eg, sitting vs standing) [38]. This limitation should be considered when interpreting PA estimates derived from wrist-based measurements. Although it was possible for participants to wear the accelerometers in the water, we intentionally advised them to remove the monitors during water activities and sleep to minimize potential malfunction or improper device positioning, which may have

XSL•FO
RenderX

underestimated their PA levels and the discrepancies between the two measures. Longer device wear time, especially during unstructured movements such as taking a shower or pool-based activities, could increase the likelihood of movement discrepancies between the two limbs. According to our follow-up survey, 1 participant consistently wore the monitors and engaged in water-based PA despite the instructions from the research team. Sensitivity analyses excluding this participant revealed that their inclusion had a minimal impact on the results. Without detailed logs of specific activities of daily living, such as brushing teeth, washing dishes, and driving, it is difficult to systematically examine differences related to hand dominance. Another limitation of this study is the use of the nondominant wrist-specific cut-points to classify activity intensities for both wrist placements. While dominant wrist–specific cut-points have been proposed using alternative metrics such as Euclidean Norm Minus One (ENMO) [47], these may not be directly comparable due to fundamental differences in data processing and calibration methods. The lack of validated CPM-based cut-points for the dominant wrist in adults highlights the need for further calibration studies to improve comparability across wear locations between the two wrists. Finally, the sample recruited in this study may also have introduced potential bias. Most participants in this study were women with relatively high levels of PA participation, as many were college students majoring in Exercise and Rehabilitation Sciences. This group included a few student-athletes who regularly engage in intensive training sessions, which may not be representative of average young adults. Prior studies suggest that there exists gender- and age-related variability in arm movement patterns and movement efficiency. Specifically, females tend to exhibit significantly less arm swing asymmetry, while males show more lateralized, asymmetric arm movements during walking [48]. Aging is also associated with reduced arm swing amplitude and less efficient arm movement coordination [49]. Thus, our sample of young adults may not accurately capture the variability in wrist-worn accelerometer–based PA estimates that could frequently occur in older populations. We acknowledge the growing importance of using open-source, monitor-independent metrics, such as ENMO, mean amplitude deviation, and Monitor-Independent Movement Summary, to enhance reproducibility and cross-study comparability in accelerometry research [50-52]. While Migueles et al [47] demonstrated that ENMO may yield greater agreement across wrist placements and proposed dominant wrist-specific cut-points, future research is warranted to directly compare these emerging metrics with VM counts across diverse device placements and evaluate their implications for interpreting accelerometry outputs in surveillance and intervention studies. Post hoc power analyses revealed that both the paired *t* tests and equivalence tests were underpowered, with estimated power ranging from 6% to 47% across primary accelerometry outcomes (eg, ST, MVPA, and steps). As such, these findings should be interpreted with caution given the limited sample size.

Despite the limitations, this study demonstrates notable strengths. First, this study used a standard 7-day data collection protocol in free-living conditions to provide more transferable knowledge for determining the use of the nondominant wrist as device placement. Moreover, the use of a user-friendly, relatively new CPM cut-point method [23] to classify ST and PA intensities is a major study strength. Past PA measurement studies have been interpreted with caution due to the misuse of the cut-points that were developed for waist or hip locations, different age groups (eg, adults vs children), different axial outputs (eg, uniaxial vs triaxial), or structured laboratory settings [53]. Our study findings, which leveraged the cross-validated cut-points, provide practical implications for future studies that adopt research-grade wrist-worn accelerometers.

Conclusions

This study aimed to examine the degree of agreement between accelerometry data collected from ActiGraph CPIWs worn on the dominant and nondominant wrists of young adults in free-living environments. The current findings align partially with previous research, demonstrating higher step counts, MVPA levels, and axis-related outcomes on the dominant wrist, while the nondominant wrist produced a higher level of ST compared with the dominant wrist. Although we observed excellent levels of reliability based on ICC across most accelerometry outcomes, the intervals between the upper and lower limits of agreement were considerably wide, particularly in steps and MVPA, along with the potential proportional bias in steps and light PA. It is estimated that the intensity of PA determines the degree of discrepancy between the dominant and nondominant wrists (eg, the dominant wrist tends to produce greater outcomes as the intensity of PA increases). These findings emphasize caution in interpreting ICC and Bland-Altman plots when comparing PA data from multiple wear locations. Taken together, the nondominant wrist could offer a reliable option for measuring physical behaviors in free-living environments; however, potential underestimation of MVPA and overestimation of ST should be considered when interpreting data outcomes and selecting the device placement.

Acknowledgments

The authors would like to thank the volunteer participants and research staff in the Exercise and Rehabilitation Sciences Program at the University of Minnesota Duluth. This study was partially supported by funding from the University of Minnesota Office of the Vice President for Research.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809. [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Goode AP, Hall KS, Batch BC, et al. The Impact of Interventions that Integrate Accelerometers on Physical Activity and Weight Loss: A Systematic Review. Ann Behav Med 2017 Feb;51(1):79-93. [doi: 10.1007/s12160-016-9829-1] [Medline: 27565168]
3. Kraus WE, Bittner V, Appel L, et al. The National Physical Activity Plan: a call to action from the American Heart Association: a science advisory from the American Heart Association. Circulation 2015 May 26;131(21):1932-1940. [doi: 10.1161/CIR.0000000000000203] [Medline: 25918126]
4. Lane-Cordova AD, Jerome GJ, Paluch AE, et al. Supporting Physical Activity in Patients and Populations During Life Events and Transitions: A Scientific Statement From the American Heart Association. Circulation 2022 Jan 25;145(4):e117-e128. [doi: 10.1161/CIR.0000000000001035]
5. Piercy KL, Troiano RP, Ballard RM, et al. The Physical Activity Guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
6. Quinn TD, Barone Gibbs B. Context Matters: The Importance of Physical Activity Domains for Public Health. Journal for the Measurement of Physical Behaviour 2023;6(4):245-249. [doi: 10.1123/jmpb.2023-0030]
7. Buchan DS, Boddy LM, McLellan G. Comparison of Free-Living and Laboratory Activity Outcomes from ActiGraph Accelerometers Worn on the Dominant and Non-Dominant Wrists. Meas Phys Educ Exerc Sci 2020 Oct 1;24(4):247-257. [doi: 10.1080/1091367X.2020.1801441]
8. Watts EL, Saint-Maurice PF, Doherty A, et al. Association of Accelerometer-Measured Physical Activity Level With Risks of Hospitalization for 25 Common Health Conditions in UK Adults. JAMA Netw Open 2023 Feb 1;6(2):e2256186. [doi: 10.1001/jamanetworkopen.2022.56186] [Medline: 36795414]
9. Baldwin CE, Parry SM, Norton L, Williams J, Lewis LK. A scoping review of interventions using accelerometers to measure physical activity or sedentary behaviour during hospitalization. Clin Rehabil 2020 Sep;34(9):1157-1172. [doi: 10.1177/0269215520932965] [Medline: 32517508]
10. De Man M, Vanderploeg E, Aimers N, MacMahon C, Wise L, Parrington L. Validity and inter-device reliability of dominant and non-dominant wrist worn activity trackers in suburban walking. Sensoria 2016;12(2). [doi: 10.7790/sa.v12i2.442]
11. Germini F, Noronha N, Borg Debono V, et al. Accuracy and Acceptability of Wrist-Wearable Activity-Tracking Devices: Systematic Review of the Literature. J Med Internet Res 2022 Jan 21;24(1):e30791. [doi: 10.2196/30791] [Medline: 35060915]
12. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159. [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
13. Fuller D, Colwell E, Low J, et al. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]
14. Migueles JH, Cadenas-Sanchez C, Ekelund U, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. Sports Med 2017 Sep;47(9):1821-1845. [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
15. Degroote L, Hamerlinck G, Poels K, et al. Low-Cost Consumer-Based Trackers to Measure Physical Activity and Sleep Duration Among Adults in Free-Living Conditions: Validation Study. JMIR Mhealth Uhealth 2020 May 19;8(5):e16674. [doi: 10.2196/16674] [Medline: 32282332]
16. Wijndaele K, Westgate K, Stephens SK, et al. Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus. Med Sci Sports Exerc 2015 Oct;47(10):2129-2139. [doi: 10.1249/MSS.0000000000000661] [Medline: 25785929]
17. Rowlands AV, Rennie K, Kozarski R, et al. Children's physical activity assessed with wrist- and hip-worn accelerometers. Med Sci Sports Exerc 2014;46(12):2308-2316. [doi: 10.1249/MSS.0000000000000365] [Medline: 24781890]
18. Buchan DS, McSeveney F, McLellan G. A comparison of physical activity from Actigraph GT 3X+ accelerometers worn on the dominant and non‐dominant wrist. Clin Physio Funct Imaging 2019 Jan;39(1):51-56. [doi: 10.1111/cpf.12538]
19. Dieu O, Mikulovic J, Fardy PS, Bui‐Xuan G, Béghin L, Vanhelst J. Physical activity using wrist‐worn accelerometers: comparison of dominant and non‐dominant wrist. Clin Physio Funct Imaging 2017 Sep;37(5):525-529. [doi: 10.1111/cpf.12337]
20. Park S, Toth LP, Hibbing PR, et al. Dominant vs. non-dominant wrist placement of activity monitors: impact on steps per day. J Meas Phys Behav Human Kinetics 2019 Jun;2(2):118-123. [doi: 10.1123/jmpb.2018-0060]
21. Dillon CB, Fitzgerald AP, Kearney PM, et al. Number of days required to estimate habitual activity using wrist-worn GENEActiv accelerometer: a cross-sectional study. PLoS ONE 2016;11(5):e0109913. [doi: 10.1371/journal.pone.0109913] [Medline: 27149674]
22. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int J Behav Nutr Phys Act 2011 Jun 16;8:1-7. [doi: 10.1186/1479-5868-8-62] [Medline: 21679426]

XSL•FO
RenderX

23. Montoye AHK, Clevenger KA, Pfeiffer KA, et al. Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. J Sports Sci 2020 Nov 16;38(22):2569-2578. [doi: 10.1080/02640414.2020.1794244]
24. Dössegger A, Ruch N, Jimmy G, et al. Reactivity to accelerometer measurement of children and adolescents. Med Sci Sports Exerc 2014 Jun;46(6):1140-1146. [doi: 10.1249/MSS.0000000000000215] [Medline: 24219978]
25. Neishabouri A, Nguyen J, Samuelsson J, et al. Quantification of acceleration as activity counts in ActiGraph wearable. Sci Rep 2022 Jul 13;12(1):11958. [doi: 10.1038/s41598-022-16003-x] [Medline: 35831446]
26. Smith MP, Horsch A, Standl M, Heinrich J, Schulz H. Uni- and triaxial accelerometric signals agree during daily routine, but show differences between sports. Sci Rep 2018 Oct 10;8(1):15055. [doi: 10.1038/s41598-018-33288-z] [Medline: 30305651]
27. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364. [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
28. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
29. Cohen J. Statistical Power Analysis for the Behavioral Sciences, th edition: Academic Press; 2013.
30. Ludbrook J. Confidence in Altman–Bland plots: A critical review of the method of differences. Clin Exp Pharma Physio 2010 Feb;37(2):143-149. [doi: 10.1111/j.1440-1681.2009.05288.x]
31. Ullrich A, Baumann S, Voigt L, John U, Ulbricht S. Measurement Reactivity of Accelerometer-Based Sedentary Behavior and Physical Activity in 2 Assessment Periods. Journal of Physical Activity and Health 2021;18(2):185-191. [doi: 10.1123/jpah.2020-0331]
32. Skender S, Ose J, Chang-Claude J, et al. Accelerometry and physical activity questionnaires - a systematic review. BMC Public Health 2016 Jun 16;16:515. [doi: 10.1186/s12889-016-3172-0] [Medline: 27306667]
33. Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Validation and Comparison of Accelerometers Worn on the Hip, Thigh, and Wrists for Measuring Physical Activity and Sedentary Behavior. AIMS Public Health 2016;3(2):298-312. [doi: 10.3934/publichealth.2016.2.298] [Medline: 29546164]
34. Fraysse F, Post D, Eston R, Kasai D, Rowlands AV, Parfitt G. Physical Activity Intensity Cut-Points for Wrist-Worn GENEActiv in Older Adults. Front Sports Act Living 2020;2:579278. [doi: 10.3389/fspor.2020.579278] [Medline: 33521631]
35. Rowlands AV. Moving Forward With Accelerometer-Assessed Physical Activity: Two Strategies to Ensure Meaningful, Interpretable, and Comparable Measures. Pediatr Exerc Sci 2018;30(4):450-456. [doi: 10.1123/pes.2018-0201]
36. Morgan CJ, Aban I. Methods for evaluating the agreement between diagnostic tests. J Nucl Cardiol 2016 Jun;23(3):511-513. [doi: 10.1007/s12350-015-0175-7] [Medline: 25995184]
37. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151. [doi: 10.11613/BM.2015.015] [Medline: 26110027]
38. Liu F, Wanigatunga AA, Schrack JA. Assessment of Physical Activity in Adults Using Wrist Accelerometers. Epidemiol Rev 2021 Dec 30;43(1):65-93. [doi: 10.1093/epirev/mxab004]
39. Pfister T, Matthews CE, Wang Q, Kopciuk KA, Courneya K, Friedenreich C. Comparison of two accelerometers for measuring physical activity and sedentary behaviour. BMJ Open Sport Exerc Med 2017;3(1):e000227. [doi: 10.1136/bmjsem-2017-000227] [Medline: 28761711]
40. Hart TL, Ainsworth BE, Tudor-Locke C. Objective and subjective measures of sedentary behavior and physical activity. Med Sci Sports Exerc 2011 Mar;43(3):449-456. [doi: 10.1249/MSS.0b013e3181ef5a93] [Medline: 20631642]
41. Karaca A, Demirci N, Yılmaz V, Hazır Aytar S, Can S, Ünver E. Validation of the ActiGraph wGT3X-BT Accelerometer for Step Counts at Five Different Body Locations in Laboratory Settings. Meas Phys Educ Exerc Sci 2022 Jan 2;26(1):63-72. [doi: 10.1080/1091367X.2021.1948414]
42. Marcotte RT, Petrucci GJ, Cox MF, Freedson PS, Staudenmayer JW, Sirard JR. Estimating Sedentary Time from a Hip- and Wrist-Worn Accelerometer. Med Sci Sports Exerc 2020 Jan;52(1):225-232. [doi: 10.1249/MSS.0000000000002099] [Medline: 31343523]
43. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and Wrist Accelerometer Algorithms for Free-Living Behavior Classification. Med Sci Sports Exerc 2016 May;48(5):933-940. [doi: 10.1249/MSS.0000000000000840] [Medline: 26673126]
44. van Hees VT, Renström F, Wright A, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PLoS ONE 2011;6(7):e22922 [FREE Full text] [doi: 10.1371/journal.pone.0022922] [Medline: 21829556]
45. Kamada M, Shiroma EJ, Harris TB, Lee IM. Comparison of physical activity assessed using hip- and wrist-worn accelerometers. Gait Posture 2016 Feb;44:23-28. [doi: 10.1016/j.gaitpost.2015.11.005] [Medline: 27004628]
46. Fairclough SJ, Noonan R, Rowlands AV, Van Hees V, Knowles Z, Boddy LM. Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. Med Sci Sports Exerc 2016 Feb;48(2):245-253. [doi: 10.1249/MSS.0000000000000771] [Medline: 26375253]
47. Migueles JH, Cadenas-Sanchez C, Rowlands AV, et al. Comparability of accelerometer signal aggregation metrics across placements and dominant wrist cut points for the assessment of physical activity in adults. Sci Rep 2019 Dec 3;9(1):18235. [doi: 10.1038/s41598-019-54267-y] [Medline: 31796778]

XSL•FO
RenderX

48. Killeen T, Elshehabi M, Filli L, et al. Arm swing asymmetry in overground walking. Sci Rep 2018 Aug 24;8(1):12803. [doi: 10.1038/s41598-018-31151-9] [Medline: 30143717]
49. Mirelman A, Bernad-Elazari H, Nobel T, et al. Effects of Aging on Arm Swing during Gait: The Role of Gait Speed and Dual Tasking. PLoS ONE 2015;10(8):e0136043. [doi: 10.1371/journal.pone.0136043] [Medline: 26305896]
50. Choudhury R, Park JH, Thiamwong L, Xie R, Stout JR. Objectively Measured Physical Activity Levels and Associated Factors in Older US Women During the COVID-19 Pandemic: Cross-sectional Study. JMIR Aging 2022 Aug 22;5(3):e38172. [doi: 10.2196/38172] [Medline: 35994346]
51. Vähä-Ypyä H, Vasankari T, Husu P, et al. Validation of Cut-Points for Evaluating the Intensity of Physical Activity with Accelerometry-Based Mean Amplitude Deviation (MAD). PLoS ONE 2015;10(8):e0134813. [doi: 10.1371/journal.pone.0134813] [Medline: 26292225]
52. Choudhury R, Park JH, Banarjee C, et al. Associations between monitor-independent movement summary (MIMS) and fall risk appraisal combining fear of falling and physiological fall risk in community-dwelling older adults. Front Aging 2024;5:1284694. [doi: 10.3389/fragi.2024.1284694]
53. Kim Y, Lee JM, Peters BP, Gaesser GA, Welk GJ. Examination of different accelerometer cut-points for assessing sedentary behaviors in children. PLoS ONE 2014;9(4):e90630. [doi: 10.1371/journal.pone.0090630] [Medline: 24699259]

Abbreviations

CPIW: CentrePoint Insight Watch
CPM: counts per minute
ENMO: Euclidean Norm Minus One
ICC: intraclass correlation coefficient
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PAGA: Physical Activity Guidelines for Americans
ST: sedentary time
TOST: two one-sided test
VM: vector magnitude

Edited by Z Liang; submitted 19.06.24; peer-reviewed by J Gay, R Choudhury; revised version received 14.05.25; accepted 04.06.25; published 11.08.25.

Please cite as:
Lee D, Voermans-Dean H, Lee JE, Shin JC, Dominick G
Comparison of ActiGraph CentrePoint Insight Watch Placement on Dominant and Nondominant Wrists in Young Adults in Free-Living Conditions: Observational Validation Study
JMIR Mhealth Uhealth 2025;13:e63033
URL: https://mhealth.jmir.org/2025/1/e63033
doi:10.2196/63033

© Daehyoung Lee, Haley Voermans-Dean, Jung Eun Lee, Jong Cheol Shin, Gregory Dominick. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Boosting Physical Activity Among Individuals With Low Engagement Through Double-Point Incentives in a Community-Based mHealth Intervention: Retrospective Observational Study

Jungin Joo[1,2,3*], PhD; Mangyeong Lee[1,2,4*], PhD; Junghee Yoon[4,5], PhD; Hyeonjin Cho[1,2], MD; Govind Warrier[6,7], MPH, MD; Johannes Thrul[3,8,9], PhD; Juhee Cho[1,2,4,10], PhD

[1]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

[10]Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Center for Clinical Epidemiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[3]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[4]Department of Clinical Research Design and Evaluation, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, 717, Irwon-dong, Gangnam-gu, Seoul, Republic of Korea

[5]Samsung Medical Center, Patient-Centered Outcomes Research Institute, Seoul, Republic of Korea

[6]Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[7]Bloomberg-Kimmel Institute for Cancer Immunotherapy and Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[8]Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, United States

[9]Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia

[*]these authors contributed equally

Corresponding Author:
Juhee Cho, PhD
Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

Abstract

Background: The administration of incentives to promote physical activity, such as the amount or timing, can vary depending on target health behaviors, research settings, intervention delivery channels, and participants' preferences. Interventions implemented at scale necessitate the consideration of potential fiscal constraints for public health promotion. Since limited funding is a barrier to implementing community-based interventions, examining both immediate and sustained effects of temporary incentive increases on physical activity is important.

Objective: This study aimed to evaluate the effect of a 1-week double-point event on increasing physical activity among low-engaged individuals in the context of a community-based mobile intervention.

Methods: Using retrospective data from a Seoul Metropolitan Government mobile health (mHealth) intervention, we evaluated the effects of a 1-week double-point incentive on participants' physical activity. During 3 registration phases from November to December 2021, a total of 50,145 individuals enrolled. Our analysis focused on the low-engaged group (n=27,833, 55.5%), who averaged fewer than 3 days per week of meeting the daily step challenge (at least 7000 steps) before the intervention. We performed a segmented regression analysis to assess changes in physical activity before and after the event. Multivariable logistic regression and Cox proportional hazards models were used to identify factors associated with improving and maintaining physical activity after starting the intervention.

Results: Of 27,833 low-engaged participants, only 13.7% (n=3835) improved their physical activity. Daily challenge engagements per week increased by 2.53 times, and average daily steps increased by 1924.97 (standardized mean difference 0.55, 95% CI 0.51 - 0.58). In multivariable logistic regression, older age was significantly associated with improved physical activity immediately after starting the intervention. However, 50% (1918/3835) of the improved group was likely to return to low engagement 3 weeks after the intervention ended. Older age and use of certain wearable devices were associated with maintaining physical activity after the intervention.

Conclusions: Double-point incentives in the short term may serve as a cue-to-action to motivate low-engagement targets; however, they do not seem to guarantee long-term maintenance in the context of community-based mHealth interventions. Further research is needed to identify additional strategies beyond monetary incentives to sustain long-term healthy behavior.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e66227) doi:10.2196/66227

KEYWORDS

community-based interventions; physical activity; engagement; financial incentive; relapsed period; mobile phone

Introduction

Physical inactivity is the fourth leading cause of global mortality, contributing significantly to the global health burden [1]. As of 2016, a total of 27.5% of the global population was physically inactive, a prevalence that has persisted since 2001 [2]. This widespread inactivity poses health risks and results in substantial health care costs for governments and other stakeholders [3]. Achieving the recommended levels of physical activity remains challenging; for example, in 2018, a total of 39.5% of Americans aged 18 years and older did not meet the aerobic activity guidelines [4]. To combat this, many countries have implemented community-based mobile interventions, such as the Carrot Rewards program in Canada and the National Steps Challenge Season 3 in Singapore, which have successfully increased physical activity [5-7].

Despite the promising effect of community-based mobile interventions, encouraging properly those whose engagement waned during the intervention remains challenging. To address this challenge, providing financial rewards or incentives is considered a good way, demonstrating up to 2.5-fold greater effectiveness in promoting physical activity engagement compared to nonincentivized conditions [8-10]. A meta-analysis revealed that among studies only targeting physically inactive populations, financial incentives improved a mean increase of 1150 steps per day [6].

The administration of incentives, such as the amount or timing, can vary depending on target health behaviors, research settings, intervention delivery channels, and participants' preferences [5,11-13]. Interventions implemented at scale necessitate the consideration of potential fiscal constraints for public health promotion. Thus, the incentive operation may change during the intervention administration due to business or financial decisions, and these changed rewards may impact user engagement [14]. In a longitudinal study with 55,000 users of a commercial physical activity application, reducing incentive size over time was associated with lowering user engagement [15]. Another longitudinal study revealed that adding a team-based incentive policy to the conventional individual-oriented policy contributed to improving daily step counts [16]. Additionally, a study with 1600 users reported that their physical activity decreased after reducing incentives [14].

The previous studies allowed us to examine the relationship between incentive changes and user behaviors in a real-world context [17]. However, further elucidation is needed regarding the effect of intentional, theory-based changes to incentive policies on user engagement within community-based mobile physical activity interventions. Specifically, as limited funding is a barrier to implementing community-based interventions, examining both immediate and sustained effects of temporary incentive increases on physical activity is important [18]. Therefore, our study aimed to evaluate the effect of a 1-week double-point event on increasing physical activity among low-engaged individuals in the context of a community-based mobile intervention. We also identified which participants increased their activity due to the event and assessed how long these improvements lasted.

Methods

Study Population

This retrospective cohort study used data from "On Seoul, Health On (OSHO)," a 9-month community-based mobile health (mHealth) intervention program organized by the Seoul Metropolitan Government from November 2021 to July 2022. The analysis focused on a subset of participants who were categorized into the low-engaged group in terms of physical activity during the program's initial phase (1 month). The low-engaged group was defined as individuals who engaged in the daily step count challenge (completing at least 7000 steps) for an average of less than 3 days per week. Several studies suggest that taking 7000 steps per day is a minimum recommended level of physical activity to reduce the risk of morbidity or mortality in adults [19-21]. Moreover, the government aimed to facilitate the participants' self-efficacy by setting achievable goals for a wide range of age groups [22,23]. Additionally, setting it to 7000 steps per day can be reasonable considering that the average number of steps for Koreans is 5755 [24], and it is usually recommended to increase the target setting by about 10% - 15% [6]. A total of 50,145 participants registered for the program, and 27,833 (55.5%) individuals in the low-engaged group during the initial phase of the program were included in this study (Figure S1 in Multimedia Appendix 1).

Description of the OSHO Program

The OSHO program aimed to enhance physical activity levels of individuals in Seoul, South Korea. Eligibility for the program required participants to be adults aged 19 - 64 years, residing or working in Seoul, and owning a mobile device compatible with at least Android (version 5.0; Google LLC) or iOS (version 10.0; Apple). Recruitment was conducted through the program's official homepage, media coverage, and advertisements in subway stations. During 3 registration phases spanning November to December 2021, a total of 50,145 individuals registered to participate. Upon registration, participants were provided with 1 of 4 different wearable devices free and asked to install a mobile app for the program. All wearable devices measured steps in real time, but the shape (band-like or watch-like) and the availability of additional functions (body composition, heart rate, etc) varied depending on the brand. Detailed specifications of each wearable brand are presented in the supplementary materials (Table S1 and Figure S2 in Multimedia Appendix 1). The mobile app enabled participants to monitor their step counts as recorded by the wearable device. To encourage participation, individuals received 200 points,

XSL•FO
RenderX

convertible to an equivalent amount of cash (KR ₩200, approximately US $0.20), for every 7000 steps or more per day. The points could be redeemed to promote their health at pharmacies, sports facilities, and other designated health care businesses.

"Spring Walk Campaign": Additional Financial Incentive Event

Approximately 4 months after the start of the program, the "Spring Walk Campaign" (SWC) was introduced as a special 1-week event, starting on March 21, 2022. The SWC was intended to encourage physical activity among participants who remained inactive during the initial intervention by providing additional financial incentives. The SWC was advertised to OSHO participants through the mobile app and the official website. During the event, those who engaged in the daily step challenge were awarded double points (400 points) to encourage physical activity. We anticipated that the SWC, a temporary incentive increase, could motivate the low-engaged participants based on a behavioral mechanism that combines "loss aversion" and "scarcity effects" [25]. Although the intervention offered additional rewards rather than removing existing ones, which is not a classical use of loss aversion, the limited-time nature may have triggered similar urgency effects associated with perceived opportunity costs.

Outcomes

The primary outcome was the immediate change in physical activity during the SWC event. Physical activity was evaluated using steps counted by wearable devices. Individuals who succeeded in the daily step challenge (>7000 steps) for at least 3 days a week were considered the improved group. The secondary outcome was the maintenance of improved physical activity among participants who transitioned to the improved group during the SWC after the removal of additional incentives.

At the time of registration for this program, sociodemographic information, including sex, age, and occupation, was assessed. The BMI and comorbidities were also assessed. Physical activity levels at enrollment were assessed using the International Physical Activity Questionnaire-7 Short Form [26]. This measure is a self-administered tool that focuses on the frequency and duration of walking, moderate to vigorous intensity activities, and sedentary time over the past 7 days. In this study, we adapted the instrument by using 6 items while excluding 1 item that assessed sedentary behavior. Responses were converted to metabolic equivalent of task (MET)–minutes per week, an objective metric that quantifies the energy cost of physical activities [27]. One MET is defined as the amount of oxygen consumed while sitting at rest and equals 3.5 mL O_2 per kg body weight×min [28]. Based on MET-min/wk values, participants' physical activities were categorized into 3 levels—low, moderate, and high. The criteria for these categorizations are detailed in Table S2 in Multimedia Appendix 1. Step count data recorded in real time through wearable devices were summarized and shared as daily cumulative figures. All distributed devices were manufactured by Korean companies and had obtained KC (Korea Certification) for product safety and qualification testing. Only authorized personnel had access to the data on the OSHO administration server, and the research team obtained deidentified data.

Statistical Analyses

Descriptive analyses were performed to summarize participants' baseline characteristics. Significance tests were conducted using the chi-square test (for categorical variables) or *t* test (for continuous variables), and the standardized mean difference (SMD) was estimated to present the magnitude of the difference between the 2 groups, considering this study's population size.

We performed a segmented regression to analyze changes in daily user engagement trajectories before and after the SWC. An equation of the regression model was considered to define levels (eg, y-intercepts) and slopes for each time segment (before vs after starting the SWC). Time started from 0, and the intervention was coded as 0 (before starting the SWC) and 1 (after starting the SWC). Time since the intervention was measured in days post the SWC. The model was fitted using the restricted maximum likelihood method, including the autoregressive error term. The autocorrelation value was calculated using the Durbin-Watson test. After fitting, we presented a graphical depiction of the data points and trend lines. Multivariable logistic regression was used to identify the factors associated with improved physical activity (the improved group) after the event. This model was adjusted for potential confounders, including age, sex, BMI, physical activity level at enrollment, chronic diseases, OS type, and wearable device type. The inclusion of OS type was due to its potential to reflect individual behavior disposition [29] or varying motivations for healthy behavior change among users [30]. Additionally, we considered that heterogeneity across wearable brands may result in different user experiences and satisfaction with the OSHO program [31,32]. This was because certain brands had high dropout rates in the exploratory analysis, which may be a confounder or attrition bias (Table S3 in Multimedia Appendix 1). Technological features are also known to influence user engagement [33]. Kaplan-Meier survival curves were plotted to estimate the time to the first occurrence of relapse to the low engagement level in the improved group, with weekly time points. Differences between curves were assessed using the log-rank test. All statistical analyses were conducted using R (version 4.3.1; R Foundation for Statistical Computing). Statistical significance was defined as a 95% CI and a *P* value of <.05.

Ethical Considerations

This study was approved by the institutional review board of the Samsung Medical Center (SMC 2022-07-073). All individuals who participated in the OSHO program provided informed consent for the program registration. However, our research uses retrospective data containing anonymized personal information and has received a waiver of additional informed consent for the analyses from the participants. Participants received 200 points (equivalent to 200 Korean Won or approximately USD 0.20) for each day they achieved 7,000 steps or more. These points could be redeemed at pharmacies, sports facilities, and other designated health-related businesses.

XSL•FO
RenderX

Results

Comparison of Demographic Characteristics Between Those With Improved Physical Activity During the SWC and Those With No Improvement

Of 27,833 participants included in the analysis, 3835 (13.7%) showed improved physical activity during the SWC (Figure S1 in Multimedia Appendix 1). The improved group members were older (39.4 vs 37.4 years, SMD=0.194) and especially included more adults aged 40 years and over, compared to the nonimproved group (SMD=0.215, Table 1). The BMI (SMD=0.031) and physical activity level at baseline (SMD=0.054) were similar between the 2 groups.

Table . Difference in demographic characteristics by improved physical activity during the Spring Walk Campaign.

	Overall (N=27,833)	Improved physical activity during the Spring Walk Campaign		Standardized mean difference
		Yes (n=3835)	No (n=23,998)	
Age (years), mean (SD)	37.7 (10.3)	39.4 (10.2)	37.4 (10.3)	0.194
Age groups (years), n (%)				
20s	7730 (27.8)	797 (20.8)	6933 (28.9)	0.215
30s	9214 (33.1)	1241 (32.4)	7973 (33.2)	
40s	7034 (25.3)	1168 (30.5)	5866 (24.4)	
50s	3073 (11)	498 (13)	2575 (10.7)	
60s	782 (2.8)	131 (3.4)	651 (2.7)	
Sex (female)	18,470 (66.4)	2644 (68.9)	15,826 (65.9)	0.064
BMI (kg/m^2), mean (SD)	23.9 (18.4)	23.8 (3.7)	23.9 (19.6)	0.013
BMI categories (kg/m^2), n (%)				
Underweight	1433 (5.1)	181 (4.7)	1252 (5.2)	0.031
Normal	12,272 (44.1)	1671 (43.6)	10,601 (44.2)	
Overweight	5525 (19.9)	790 (20.6)	4735 (19.7)	
Obese	8603 (30.9)	1193 (31.1)	7410 (30.9)	
Physical activity level, n (%)				0.054
Inactive	16,028 (57.6)	2132 (55.6)	13,896 (57.9)	
Minimally active	6586 (23.7)	981 (25.6)	5605 (23.4)	
High active	5219 (18.8)	722 (18.8)	4497 (18.7)	
Chronic diseases, yes, n (%)	5016 (18)	719 (18.7)	4297 (17.9)	0.022
OS type (%)[a], n (%)				0.137
iOS	9393 (33.9)	1085 (28.4)	8308 (34.8)	
Android	18,322 (66.1)	2735 (71.6)	15,587 (65.2)	
Wearable device type (%), n (%)				0.154
A	16,297 (58.6)	2424 (63.2)	13,873 (57.8)	
B	5621 (20.2)	751 (19.6)	4870 (20.3)	
C	1156 (4.2)	177 (4.6)	979 (4.1)	
D	4759 (17.1)	483 (12.6)	4276 (17.8)	

[a]N= 27,715, eliminated 118 missing values.

Factors Associated With Moving to the Improved Group During the SWC

In multivariable logistic regression, participants in their aged (years) 30s (adjusted odds ratio [aOR] 1.40, 95% CI 1.27 - 1.54), 40s (aOR 1.79, 95% CI 1.62 - 1.97), 50s (aOR 1.73, 95% CI 1.53 - 1.95), and 60s (aOR 1.82, 95% CI 1.48 - 2.23) were more likely to show improved physical activity during the SWC (Table 2). Additionally, those who had minimal physical activity levels assessed using the International Physical Activity Questionnaire-7 Short Form at baseline (aOR 1.16, 95% CI 1.07 - 1.26) or used android-powered smartphones

XSL•FO
RenderX

compared to iOS (aOR 1.20, 95% CI 1.11 - 1.30) were more likely to show improved physical activity. However, those who received wearable B (aOR 0.89, 95% CI 0.81 - 0.97) or wearable D (aOR 0.63, 95% CI 0.57 - 0.70) were less likely to show improved physical activity.

Table . Factors associated with improved physical activity during the Spring Walk Campaign (N=27,833).

	Adjusted OR (95% CI)[a]
Age (years)	
20s	*Reference*
30s	1.40 (1.27-1.54)
40s	1.79 (1.62-1.97)
50s	1.73 (1.53-1.95)
60s	1.82 (1.48-2.23)
Sex (female)	1.23 (1.14-1.33)
BMI (kg/m^2)	
Underweight	*Reference*
Normal	1.05 (0.89-1.24)
Overweight	1.10 (0.92-1.32)
Obese	1.09 (0.92-1.30)
Physical activity level	
Inactive	*Reference*
Minimally active	1.16 (1.07-1.26)
High active	1.08 (0.99-1.19)
Chronic diseases (yes)	0.93 (0.84-1.02)
OS type (%)[b]	
iOS	*Reference*
Android	1.20 (1.11-1.30)
Wearable device type (%)	
A	*Reference*
B	0.89 (0.81-0.97)
C	1.00 (0.84-1.18)
D	0.63 (0.57-0.70)

[a]Adjusted Model included age group, sex, BMI, physical activity level, chronic diseases, type of operating system, and wearable device type.
[b]N=27,715, eliminated 118 missing values.

Immediate Changes in Physical Activity After Starting the SWC

Daily activity challenge achievement rate (%) tended to decrease by 0.18 units per day at 27.45 (Figure 1). After starting the SWC, the achievement rate immediately increased by 8.16 (P<.001). Especially, those with improved physical activity during the SWC increased their weekly engagement from an average of 1.76 to 4.29 times (SMD=1.81, 95% CI 1.76 - 1.86, Table 3). Concerning average step count, there was an increase of 1924.97 steps in those with improved physical activity during the SWC (SMD=0.55, 95% CI 0.51 - 0.58).

XSL•FO
RenderX

Figure 1. Graphical depiction of interrupted time series analysis of physical activity engagement before and after the Spring Walk Campaign.

XSL•FO
RenderX

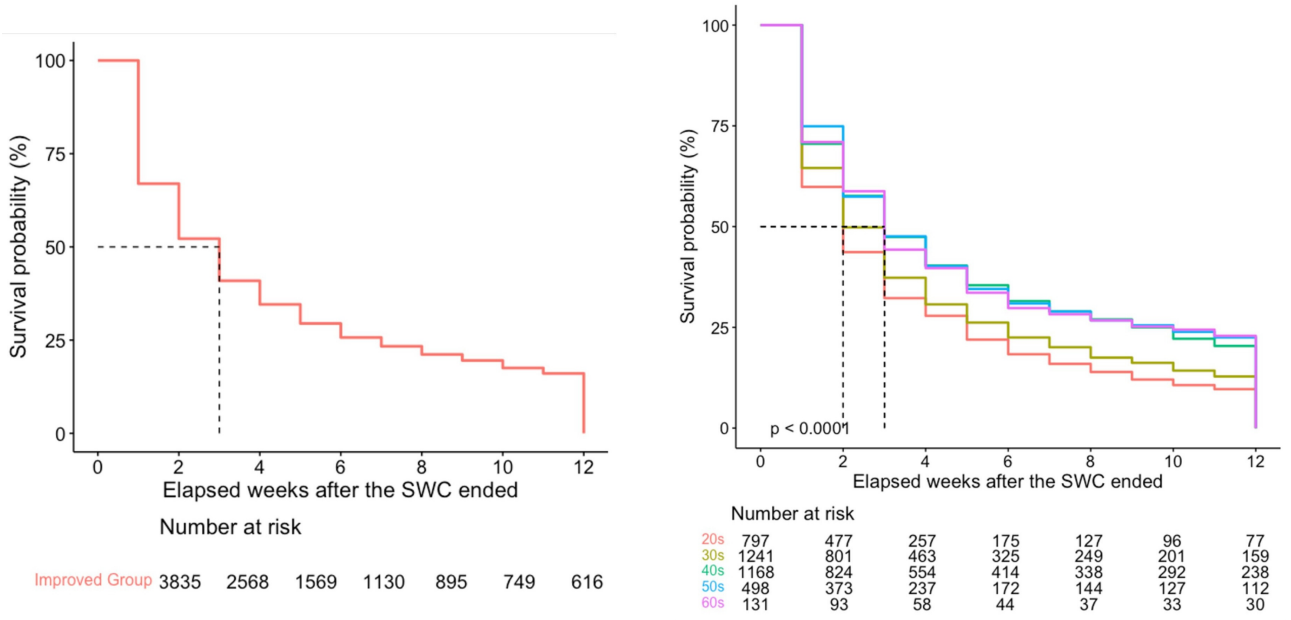
Table . Changes in the number of daily challenge engagement per week and average daily step counts before and during the Spring Walk Campaign.

	Before the Spring Walk Campaign, mean (SD)	During the Spring Walk Campaign, mean (SD)	Mean difference	Standardized mean difference (95% CI)
The number of daily challenge engagements per week				
Overall	0.57 (0.87)	0.77 (1.56)	0.20	0.17 (0.16 to 0.18)
Improved group (N=3835)	1.76 (0.83)	4.29 (1.22)	2.53	1.81 (1.76 to 1.86)
Nonimproved group (N=23,998)	0.38 (0.71)	0.21 (0.54)	−0.17	−0.26 (−0.27 to –0.25)
Average step counts per day				
Overall	2857.22 (3724.84)	2243.80 (3714.70)	−613.42	−0.18 (−0.19 to –0.17)
Improved group (N=3835)	6769.13 (3640.74)	8694.10 (2857.22)	1924.97	0.55 (0.51 to 0.58)
Nonimproved group (N=23,998)	2232.08 (3337.30)	1213.01 (2643.52)	−1019.07	−0.32 (−0.34 to –0.31)

Sustaining Improved Physical Activity After the SWC Ended

After starting the SWC, the decreasing slope of the completion rates improved from −0.18 to −0.02 (P<.001, Figure 1). The Kaplan-Meier curve showed that half of the improved group were likely to return to low engagement in physical activity 3 weeks after the SWC ended (Figure 2). In the multivariable Cox proportional hazards model, participants in their aged (years) 20s (adjusted hazard ratio [aHR] 1.44, 95% CI 1.19 - 1.75) or 30s (aHR 1.27, 95% CI 1.05 - 1.53) had a higher risk of returning to low engagement in physical activity after the SWC ended (Table 4). Moreover, users of wearable B (aHR 1.12, 95% CI 1.03 - 1.21) and wearable D (aHR 1.15, 95% CI 1.04 - 1.27) had a greater risk of returning to low engagement compared to users of wearable A.

Figure 2. Kaplan-Meier curves for time to return to low engagement in physical activity after the Spring Walk Campaign ended among the improved group. SWC: Spring Walk Campaign.



XSL•FO
RenderX

Table . Factors associated with relapse to low physical activity engagement after the Spring Walk Campaign among participants with improved physical activity during the Spring Walk Campaign[a].

	Crude hazard ratio (95% CI)	Adjusted[b] hazard ratio (95% CI)[c,d]
Age group (years)		
20s	1.43 (1.19-1.72)	1.44 (1.19-1.75)
30s	1.26 (1.05-1.51)	1.27 (1.05-1.53)
40s	1.02 (0.85-1.23)	1.03 (0.86-1.24)
50s	0.99 (0.82-1.20)	1.00 (0.82-1.21)
60s	*Reference*	*Reference*
Sex (female)	0.98 (0.91-1.07)	0.99 (0.91-1.08)
BMI (kg/m^2), mean (SD)	1.04 (0.97-1.12)	1.04 (0.96-1.11)
Underweight	*Reference*	*Reference*
Normal	0.92 (0.79-1.07)	0.98 (0.84-1.14)
Overweight	0.96 (0.81-1.12)	1.05 (0.89-1.24)
Obese	0.96 (0.82-1.12)	1.05 (0.89-1.24)
Physical activity level		
Inactive	*Reference*	*Reference*
Minimally active	0.97 (0.90-1.04)	0.98 (0.91-1.06)
High active	0.99 (0.91-1.07)	1.01 (0.92-1.10)
Chronic diseases (yes)	0.93 (0.85-1.00)	1.02 (0.93-1.12)
OS type (%)[d]		
iOS	*Reference*	*Reference*
Android	0.89 (0.83-0.95)	0.96 (0.89-1.03)
Wearable device type (%)		
A	*Reference*	*Reference*
B	1.12 (1.03-1.21)	1.12 (1.03-1.21)
C	0.89 (0.76-1.04)	0.93 (0.80-1.08)
D	1.11 (1.01-1.23)	1.15 (1.04-1.27)

[a]The improved group (N[1]=3820).

[b]Eliminated 15 missing values.

[c]Crude model represents the crude hazard ratio (95% CI).

[d]Adjusted Model included age group, sex, BMI, physical activity level, chronic diseases, OS type, and wearable device type.

Discussion

Principal Findings

In this study, we evaluated the effect of a 1-week double-point event on increasing physical activity among low-engaged individuals in the context of a community-based mobile intervention. We also identified which participants increased their activity due to the event and assessed how long these improvements lasted. We found that 14% (n=3835) of those who were in the low-engaged group with the intervention improved their physical activity during the SWC. However, the effect diminished in 50% (1918/3835) of the participants within 3 weeks, with those in their 20s and 30s showing a more rapid return to preintervention levels.

Immediate Effects of a 1-Week Double-Point Incentive Event on Change of Physical Activity

In our study, the OSHO program provided a stream of small-size incentives (US $0.20 per day) for achieving the daily physical activity challenge. However, participants' physical activity or engagement with the program declined over time. We found that temporarily increasing incentives could help improve intervention engagement, potentially leading to physical activity improvement. Immediately after starting the SWC, the achievement rate of the daily activity challenge increased by 8.16%. In the low-engaged group of 27,833 people, approximately 14% (n=3835) showed an improvement in their physical activity during the SWC. In general, larger absolute amounts of financial incentives tend to ensure greater user engagement with mobile interventions [5,8]. However, recent

XSL•FO
RenderX

evidence highlights that considering how to design an incentive structure is more important than securing a larger amount of incentives. A study using real-world data from 54,817 Canadians found that decreasing incentives over time can lead to lower user engagement in mobile health programs [15]. Another real-world study found that policy change toward increasing the uncertainty of obtaining incentives led to decreased physical activity among users [14]. A randomized controlled trial in which the total amount of incentives was the same but the structure was different found that a constant incentive rate was more effective than increasing or decreasing rates over time [34]. Therefore, considering our research and previous studies, it would be necessary to develop and evaluate effective incentive policies to increase user participation. However, since the overall effect size of SWC was low (SMD=0.17) and certain participants responded to it, multiple incentive structures should be established within an intervention [35].

Lasting Effects of a 1-Week Double-Point Incentive Event on Change of Physical Activity

Although 14% (n=3835) of the low-engaged group became more active during the SWC, 50% (1918/3835) of them relapsed into the low-engaged group within 3 weeks after the event. This suggests that the temporary incentive increase does not ensure long-term effects on physical activity. Participants may feel that the end of SWC was akin to a withdrawal of an incentive. Though attitudinal assessment of participants was not a primary focus of this investigation, study of participants' attitudes toward the SWC will be important in future initiatives.

Eventually, this may lead to lowering their motivation. In previous studies, financial incentives rarely led to long-term behavior change; particularly, people tended to revert to their initial behaviors soon after the incentives stopped [36]. A large randomized clinical trial evaluating the effects of a fitness tracker and various incentive structures on physical activity found that those who received cash incentives were more physically active than the control group for 6 months [5]. However, the effect diminished over the 6 months after the incentives ended. A meta-analysis also indicated that only 4 of the 18 studies exhibited positive effects following the end of the interventions. These interventions include various types of incentives, such as immediate rewards, team-based incentives, loss-framed incentives, etc, based on a randomized controlled trial [6]. In general, monetary incentives could encourage extrinsic motivation while discouraging intrinsic motivation [37]. Therefore, strategies are needed that can quickly convert extrinsic rewards provided within a temporary event into intrinsic motivation [36-38]. For instance, combining wearable technology with regular counseling produced higher engagement rates, indicating that integrating financial incentives with personalized support and feedback may be a good option [39].

Impact of Age and Technology-Related Factors on Physical Activity Changes

In our study, we found that the older age group had greater responsiveness to the SWC and were more likely to maintain their improved physical activity. These findings are supported by previous studies that have noted greater participation among older adults in physical activity interventions [8,40]. According to a study using combined data from 8 studies with 100,000 participants, older age was a significant factor associated with retention in remote digital health studies [41]. Another longitudinal study with 1000 participants found that older age was positively associated with consistent and continued use of mobile health apps after becoming a user [42]. A possible reason is that older people are relatively more responsive to health care services, so their satisfaction and perceived benefits from mHealth may have been well maintained [43]. This may reflect the older population's greater temporal flexibility compared to younger individuals constrained by occupational commitments. Another possible explanation may be that older people placed a higher value on the use of incentives provided in our study. According to the administrative manual of the OSHO program, the places where the points could be redeemed were limited to pharmacies, sports facilities, and other designated health care businesses to improve their health. This may not have been appealing to younger participants. Indeed, an exploratory study with Israeli adults (n=379) aged 20 - 89 years reveals that while participants in the young age group (ages 20 - 39 years) rated 4 factors—affiliation, competition, appearance, and mastery—as more important. In contrast, the older age group (ages 61 - 89 years) rated health benefits as more important than the other age groups [44]. For younger participants, strategies such as group exercises, challenges, and sharing achievements on social media could be effective to encourage intrinsic motivation. Hence, it would be necessary to consider tailored strategies to balance extrinsic and intrinsic motivation for each age group [38].

Although we did not directly evaluate the participants' user experience or investigate technological problems, it is possible that participants had negative experiences with certain brands of wearable devices. The Seoul Metropolitan Government tried to distribute devices with similar specifications, but our exploratory analysis discovered that brands A and B had higher dropout rates. This suggests that compatibility issues with the OSHO app or other technological problems with the devices may have negatively affected the user experience. In fact, heterogeneity between devices and systems for a single service can easily create compatibility problems [45]. This may have affected engagement with the intervention and, eventually, motivation for physical activity [33]. Therefore, rather than solely considering incentives as a means of increasing physical activity, it is important to incorporate design and administrative considerations that ensure consistent performance and user experience across platforms and devices.

Limitations

Limitations of the current study may include confounding by unmeasured variables. The Seoul Metropolitan Government did not measure baseline health conditions such as weight, BMI, or average steps before starting this promotion. We applied retrospective analyses to evaluate a specific intervention with a large population in a real-world setting where experimental manipulations are not feasible. However, this approach may also have limitations in clearly identifying the causal effects of the intervention due to unclear or selective exposure to the intervention. To better validate the efficacy of additional financial incentives, conducting a randomized controlled trial

XSL•FO
RenderX

will be needed. Behavioral aspects related to physical activity, which could impact participants' motivation in the SWC, were also not considered. The use of wearable devices or smartphone apps to track physical activity may result in technological limitations or inaccuracies in data collection. The OSHO program provided all participants with wearable devices to monitor daily activities. However, there may still be a possibility of underestimation of daily steps if participants did not consistently wear the device. Additionally, not all participants had equal access to or familiarity with the technology, potentially affecting the results. Moreover, this study may not have adequately measured or addressed variations in participant engagement and adherence to the intervention, which could have affected the outcomes. Future research should explore methods to improve and measure engagement to enhance the effectiveness of similar interventions. Finally, the sample consisted solely of Seoul citizens, where the walkability index ranks among the top 10 cities in the world [46]. The generalizability of our findings could vary depending on the infrastructure of a given location. Future studies should replicate and extend the current findings using more diverse and representative samples from different regions to enhance the generalizability of the results.

Conclusions

In conclusion, we found that the double-point incentive motivated only a small subset of participants, leading to significant short-term increases in physical activity. However, only a small proportion of participants maintained these behavioral changes, indicating that sustaining such changes typically requires longer or repeated interventions and personalized strategies to enhance both engagement and long-term adherence. Financial incentives alone may not be sufficient to promote sustained physical activity. Other factors, such as personalized support, behavioral change techniques, and environmental modifications, should be considered to effectively enhance physical activity levels. Integrating these elements can provide a more comprehensive approach to motivate individuals and sustain healthy behaviors.

Acknowledgments

There is no declaration for acknowledgments. This research was supported by the Seoul Metropolitan Government (20220435393-00) and the SmartTech Clinical Research Center (SCRC), funded by the Ministry of Health & Welfare, Republic of Korea (RS-2023-KH142023).

Data Availability

Data and codes supporting the findings of this study are available from the corresponding author on reasonable request.

Authors' Contributions

JJ, ML, and JC conceptualized this study. ML, JT, and JC worked on the methodology and validated this study. JJ and ML did the data curation, formal analysis, visualization of this study; and wrote and prepared the original draft. All authors reviewed and edited the writing. JC supervised and obtained funding for this study.

Conflicts of Interest

The authors declare no conflicts of interest. However, JC, ML, and JY participated in this research project, which was conducted with support and funding from the Seoul Metropolitan Government. Therefore, to ensure the transparency and independence of the research project, data curation and analyses were performed and reviewed together with external experts who were not involved in the research project. None of the authors received any financial benefits from the funder, including royalties, incentives, stocks, or stock options.

Multimedia Appendix 1
Additional figures' and tables' final edits: Sankey plot, wearable device description and specifications, IPAQ-7 Short Form Scoring protocol, and devices' attrition rate. IPAQ-7: International Physical Activity Questionnaire - Short Form.
[DOCX File, 344 KB - mhealth_v13i1e66227_app1.docx]

References

1. Kohl HW 3rd, Craig CL, Lambert EV, et al. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 23041199]
2. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart Disease and Stroke Statistics-2022 update: a report from the American Heart Association. Circulation 2022 Feb 22;145(8):e153-e639. [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
3. Ding D, Lawson KD, Kolbe-Alexander TL, et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016 Sep 24;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
4. Elgaddal N, Kramarow EA, Reuben C. Physical activity among adults aged 18 and over: United States, 2020. NCHS Data Brief 2022 Aug(443):1-8. [doi: 10.15620/cdc:120213] [Medline: 36043905]

XSL•FO
RenderX

5. Finkelstein EA, Haaland BA, Bilger M, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
6. Mitchell MS, Orstad SL, Biswas A, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med 2020 Nov;54(21):1259-1268. [doi: 10.1136/bjsports-2019-100633] [Medline: 31092399]
7. Rayward AT, Vandelanotte C, Corry K, Van Itallie A, Duncan MJ. Impact of a social media campaign on reach, uptake, and engagement with a free web- and app-based physical activity intervention: the 10,000 steps Australia program. Int J Environ Res Public Health 2019 Dec 12;16(24):31842383. [doi: 10.3390/ijerph16245076] [Medline: 31842383]
8. Yao J, Lim N, Tan J, et al. Evaluation of a population-wide mobile health physical activity program in 696 907 adults in Singapore. J Am Heart Assoc 2022 Jun 21;11(12):e022508. [doi: 10.1161/JAHA.121.022508] [Medline: 35699174]
9. Giles EL, Robalino S, Sniehotta FF, Adams J, McColl E. Acceptability of financial incentives for encouraging uptake of healthy behaviours: a critical review using systematic methods. Prev Med 2015 Apr;73(145-58):145-158. [doi: 10.1016/j.ypmed.2014.12.029] [Medline: 25600881]
10. Giles EL, Robalino S, McColl E, Sniehotta FF, Adams J. The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. PLoS One 2014;9(3):e90347. [doi: 10.1371/journal.pone.0090347] [Medline: 24618584]
11. Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G. Evaluating the Carrot Rewards app, a population-level incentive-based intervention promoting step counts across two Canadian provinces: quasi-experimental study. JMIR mHealth uHealth 2018 Sep 20;6(9):e178. [doi: 10.2196/mhealth.9912] [Medline: 30148712]
12. Sen AP, Sewell TB, Riley EB, et al. Financial incentives for home-based health monitoring: a randomized controlled trial. J Gen Intern Med 2014 May;29(5):770-777. [doi: 10.1007/s11606-014-2778-0] [Medline: 24522623]
13. Kramer JN, Künzler F, Mishra V, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540. [doi: 10.2196/11540] [Medline: 30702430]
14. McCarthy H, Potts HWW, Fisher A. Tracked physical activity levels before and after a change in incentive strategy among UK adults using a rewards app: retrospective quasi-experimental study. JMIR Form Res 2024 Dec 10;8:e50041. [doi: 10.2196/50041] [Medline: 39657165]
15. Brower J, LaBarge MC, White L, Mitchell MS. Examining responsiveness to an incentive-based mobile health app: longitudinal observational study. J Med Internet Res 2020 Aug 10;22(8):e16797. [doi: 10.2196/16797] [Medline: 32773371]
16. Pearson E, Prapavessis H, Higgins C, Petrella R, White L, Mitchell M. Adding team-based financial incentives to the Carrot Rewards physical activity app increases daily step count on a population scale: a 24-week matched case control study. Int J Behav Nutr Phys Act 2020 Nov 19;17(1):139. [doi: 10.1186/s12966-020-01043-1] [Medline: 33208166]
17. Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. Annu Rev Public Health 2017 Mar 20;38:39-56. [doi: 10.1146/annurev-publhealth-031816-044327] [Medline: 28125392]
18. Cooper J, Murphy J, Woods C, et al. Barriers and facilitators to implementing community-based physical activity interventions: a qualitative systematic review. Int J Behav Nutr Phys Act 2021 Sep 7;18(1):118. [doi: 10.1186/s12966-021-01177-w] [Medline: 34493306]
19. Paluch AE, Gabriel KP, Fulton JE, et al. Steps per day and all-cause mortality in middle-aged adults in the coronary artery risk development in young adults study. JAMA Netw Open 2021 Sep 1;4(9):e2124516. [doi: 10.1001/jamanetworkopen.2021.24516] [Medline: 34477847]
20. Hsueh MC, Lin CY, Lai TF, et al. Is achieving 7,000 steps/day cross-sectionally and prospectively associated with older adults' lower-extremity performance? BMC Geriatr 2021 Jun 12;21(1):359. [doi: 10.1186/s12877-021-02289-5] [Medline: 34118884]
21. Tudor-Locke C, Craig CL, Brown WJ, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8(79):21798015. [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
22. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
23. Stock J, Cervone D. Proximal goal-setting and self-regulatory processes. Cogn Ther Res 1990 Oct;14(5):483-498. [doi: 10.1007/BF01172969]
24. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature New Biol 2017 Jul 20;547(7663):336-339. [doi: 10.1038/nature23018] [Medline: 28693034]
25. Salmani B, Prapavessis H, Vanderloo LM, Mitchell MS. Financial incentives for physical activity in adults: systematic review and meta-analysis update. Prev Med 2025 Mar;192(108237):108237. [doi: 10.1016/j.ypmed.2025.108237] [Medline: 39870214]
26. International physical activity questionnaire - short form. YOUTHREX Research & Evaluation Exchange. 2002. URL: https://youthrex.com/wp-content/uploads/2019/10/IPAQ-TM.pdf [accessed 2025-06-11]
27. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ) – short and long forms. IPAQ Research Committee. 2005. URL: https://sites.google.com/view/ipaq/home?authuser=0 [accessed 2025-06-11]

XSL•FO
RenderX

28. Jetté M, Sidney K, Blümchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 1990 Aug;13(8):555-565. [doi: 10.1002/clc.4960130809] [Medline: 2204507]
29. Götz FM, Stieger S, Reips UD. Users of the main smartphone operating systems (iOS, Android) differ only little in personality. PLoS One 2017;12(5):e0176921. [doi: 10.1371/journal.pone.0176921] [Medline: 28467473]
30. Ubhi HK, Kotz D, Michie S, van Schayck OCP, West R. A comparison of the characteristics of iOS and Android users of a smoking cessation app. Transl Behav Med 2017 Jun;7(2):166-171. [doi: 10.1007/s13142-016-0455-z] [Medline: 28168609]
31. Jia Y, Wang W, Wen D, Liang L, Gao L, Lei J. Perceived user preferences and usability evaluation of mainstream wearable devices for health monitoring. PeerJ 2018;6:e5350. [doi: 10.7717/peerj.5350] [Medline: 30065893]
32. Liang J, Xian D, Liu X, et al. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. JMIR mHealth uHealth 2018 Nov 8;6(11):e11066. [doi: 10.2196/11066] [Medline: 30409767]
33. Yang X, Ma L, Zhao X, Kankanhalli A. Factors influencing user's adherence to physical activity applications: a scoping literature review and future directions. Int J Med Inform 2020 Feb;134(104039):104039. [doi: 10.1016/j.ijmedinf.2019.104039] [Medline: 31865054]
34. Bachireddy C, Joung A, John LK, et al. Effect of different financial incentive structures on promoting physical activity among adults: a randomized clinical trial. JAMA Netw Open 2019 Aug 2;2(8):e199863. [doi: 10.1001/jamanetworkopen.2019.9863] [Medline: 31441936]
35. Rummo PE, Elbel B. Using multiple financial incentive structures to promote sustainable changes in health behaviors. JAMA Netw Open 2019 Aug 2;2(8):e199859. [doi: 10.1001/jamanetworkopen.2019.9859] [Medline: 31441932]
36. Winkler-Schor S, Brauer M. What happens when payments end? Fostering long-term behavior change with financial incentives. Perspect Psychol Sci 2024 May 20:17456916241247152. [doi: 10.1177/17456916241247152] [Medline: 38767968]
37. Promberger M, Marteau TM. When do financial incentives reduce intrinsic motivation? Comparing behaviors studied in psychological and economic literatures. Health Psychol 2013 Sep;32(9):950-957. [doi: 10.1037/a0032727] [Medline: 24001245]
38. Ryan RM, Patrick H. Self-determination theory and physical. Hell J Psychol 2009;6(2):107-124. [doi: 10.1016/j.psychsport.2018.11.010]
39. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled pilot trial. JMIR mHealth uHealth 2017 Mar 6;5(3):e28. [doi: 10.2196/mhealth.6967] [Medline: 28264796]
40. Edney S, Ryan JC, Olds T, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645. [doi: 10.2196/14645] [Medline: 31774402]
41. Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3(21):21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
42. Coughlin JW, Martin LM, Zhao D, et al. Electronic health record-based recruitment and retention and mobile health app usage: multisite cohort study. J Med Internet Res 2022 Jun 10;24(6):e34191. [doi: 10.2196/34191] [Medline: 35687400]
43. Ford KL, Bryant AN, Kim G. Age differences in satisfaction with and perceived benefit from mental health services: results from the collaborative psychiatric epidemiology surveys. Int J Geriatr Psychiatry 2013 Aug;28(8):831-840. [doi: 10.1002/gps.3889] [Medline: 23044699]
44. Raviv S, Netz Y. Age, gender, and level of activity as moderators of personal incentives to physical activity in Israel. J Psychol 2007 May;141(3):241-261. [doi: 10.3200/JRLP.141.3.241-262] [Medline: 17564256]
45. Developing crossplatform software applications to enhance compatibility across devices and systems. Comput Sci IT Res J 2024;5(8):2040-2061. [doi: 10.51594/csitrj.v5i8.1491]
46. Sandle T. South Korea bound: welcome to the world's most social city. Digital Journal. 2025 Mar 21. URL: https://www.digitaljournal.com/social-media/south-korea-bound-welcome-to-the-worlds-most-social-city/article [accessed 2025-06-11]

Abbreviations

aHR: adjusted hazard ratio
aOR: adjusted odds ratio
KC: Korea Certification
MET: metabolic equivalent of task
mHealth: mobile health
OSHO: On Seoul, Health On
SMD: standardized mean difference
SWC: Spring Walk Campaign

XSL•FO
RenderX

Edited by G Badicu; submitted 14.09.24; peer-reviewed by K Yuenyongchaiwat, Q He; revised version received 15.05.25; accepted 16.05.25; published 21.08.25.

Please cite as:
Joo J, Lee M, Yoon J, Cho H, Warrier G, Thrul J, Cho J
Boosting Physical Activity Among Individuals With Low Engagement Through Double-Point Incentives in a Community-Based mHealth Intervention: Retrospective Observational Study
JMIR Mhealth Uhealth 2025;13:e66227
URL: https://mhealth.jmir.org/2025/1/e66227
doi:10.2196/66227

© Jungin Joo, Mangyeong Lee, Junghee Yoon, Hyeonjin Cho, Govind Warrier, Johannes Thrul, Juhee Cho. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Accelerometry-Assessed Physical Activity and Circadian Rhythm to Detect Clinical Disability Status in Multiple Sclerosis: Cross-Sectional Study

Nicole Bou Rjeily[1], MD; Muraleetharan Sanjayan[1], MS; Pratim Guha Niyogi[2], PhD; Blake E Dewey[1], PhD; Alexandra Zambriczki Lee[1], BS; Christy Hulett[1], BA; Gabriella Dagher[1], BS; Chen Hu[1], MS; Rafal D Mazur[1], BS; Elena M Kenney[1], BS; Erin Brennan[1], MSN, PMHNP-BC; Anna DuVal[1], BA, MPH; Peter A Calabresi[1], MD; Vadim Zipunnikov[2], PhD; Kathryn C Fitzgerald[1], ScD; Ellen M Mowry[1], MD, MCR

[1]Department of Neurology, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Pathology 627, Baltimore, MD, United States
[2]Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Nicole Bou Rjeily, MD
Department of Neurology, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Pathology 627, Baltimore, MD, United States

Abstract

Background: Tools for measuring clinical disability status in people with multiple sclerosis (MS) are limited. Accelerometry objectively assesses physical activity and circadian rhythmicity profiles in the real-world environment and may potentially distinguish levels of disability in MS.

Objective: This study aims to determine if accelerometry can detect differences in physical activity and circadian rhythms between relapsing-remitting multiple sclerosis (RRMS) and progressive multiple sclerosis (PMS) and to assess the interplay within person between the 2 domains of physical activity (PA) and circadian rhythm (CR) in relation to MS type.

Methods: This study represents an analysis of the baseline data from the prospective HEAL-MS (home-based evaluation of actigraphy to predict longitudinal function in multiple sclerosis) study. Participants were divided into 3 groups based on the Expanded Disability Status Scale (EDSS) criteria for sustained disability progression: RRMS-Stable, RRMS-Suspected progression, and PMS. Baseline visits occurred between January 2021 and March 2023. Clinical outcome measures were collected by masked examiners. Participants wore the GT9X Link ActiGraph on their nondominant wrists for 2 weeks. After adjusting for age, sex, and BMI, a logistic regression model was fitted to evaluate the association of each accelerometry metric with odds of PMS versus RRMS. We also evaluated the association of accelerometry metrics in differentiating the 2 RRMS subtypes. The Joint and Individual Variation Explained (JIVE) model was used to assess the codependencies between the PA and CR domains and their joint and individual association with MS subtype.

Results: A total of 253 participants were included: 86 with RRMS-Stable, 82 with RRMS-Suspected progression, and 85 with PMS. Compared to RRMS, participants with PMS had lower total activity counts (β=−0.32, 95% CI −0.61 to −0.03), lower time spent in moderate to vigorous physical activity (β=−0.01, 95% CI −0.02 to −0.004), higher active-to-sedentary transition probability (β=5.68, 95% CI 1.86-9.5), lower amplitude (β=−0.0004, 95% CI −0.0008 to −0.0001), higher intradaily variability (β=4.64, 95% CI 1.45-7.84), and lower interdaily stability (β=−4.43, 95% CI −8.77 to −0.10). Using the JIVE model for PA and CR domains, PMS had higher first joint component (β=0.367, 95% CI 0.088-0.656), lower PA-1 component (β=−0.441, 95% CI −0.740 to −0.159), and lower PA-2 component (β=−0.415, 95% CI −0.717 to −0.126) compared to RRMS. No significant differences were detected between the 2 RRMS subtypes except for lower relative amplitude in those with suspected progression (β=−5.26, 95% CI −10.80 to −0.20).

Conclusions: Accelerometry detected differences in physical activity patterns between RRMS and PMS. More advanced analytic techniques may help discern differences between the 2 RRMS subgroups. Longitudinal follow-up is underway to assess the potential for accelerometry to detect or predict disability progression.

(JMIR Mhealth Uhealth 2025;13:e57599) doi:10.2196/57599

KEYWORDS

multiple sclerosis; disability; progressive; physical activity; circadian rhythm; accelerometer; ActiGraph; accelerometry

XSL•FO
RenderX

Introduction

Multiple sclerosis (MS) is a chronic disease affecting the central nervous system in which those affected typically have intermittent neurologic symptoms and signs early in the course (relapsing-remitting multiple sclerosis [RRMS]) but often, subsequently (or, less commonly, from onset), slowly accumulate disability (progressive multiple sclerosis [PMS]). Currently available tools to measure disability in MS are limited. The Expanded Disability Status Scale (EDSS) is the most widely used clinical outcome measure and is considered by the Food and Drug Administration to be the gold standard for phase 3 trials in PMS. The EDSS, however, is semiquantitative, has limited reliability, and only captures a person's state at one short point in time during a clinical visit [1]. People with MS may have different disability states at different times throughout the day, as symptoms can fluctuate with temperature, fatigue, stress, and other factors [2,3]. The limitations of the EDSS lead to long delays in confirming that a person has transitioned from RRMS to PMS and also inflate the sample size and follow-up time required for MS trials for which disability progression is the primary end point [4,5].

Triaxial accelerometry is a safe and relatively inexpensive tool that may offer an objective and sensitive measure of disability in people with MS. With the use of an accelerometer worn on the wrist, real-time information about physical activity and circadian rhythmicity patterns can be collected in a person's natural environment. Such data may allow detection of variation in activity that may be missed during clinical visits [6,7]. Several studies have looked at the use of accelerometry in MS to identify associations with sleep, fatigue, depression, or disability outcomes (EDSS and Patient-Determined Disease Steps); however, whether wrist-worn accelerometry can differentiate between people with RRMS and PMS remains an existing question [8-16]. We hypothesized that people with RRMS and PMS have different physical activity and circadian rhythmicity patterns. Due to the interplay between physical activity and circadian rhythm domains within person, we also aimed to study joint and individual variations in these 2 domains and their associations with MS subtype.

Methods

Participant Selection and Study Design

This study includes an analysis of the baseline data from the longitudinal observational study HEAL-MS (home-based evaluation of actigraphy to predict longitudinal function in multiple sclerosis; Figure S1 in Multimedia Appendix 1). Eligible participants followed at the Johns Hopkins Multiple Sclerosis Precision Medicine Center of Excellence were recruited between January 2021 and March 2023. Participants had to have a diagnosis of MS, were aged ≥40 years, had no apparent comorbidities that may limit physical activity (eg, heart failure and end-stage renal disease), had no MS relapse within the last 6 months prior to enrollment, and had baseline EDSS score ≤6.5. These criteria were chosen so that the participants with RRMS were of the age at which transition to PMS might begin within the next few years, the changes in accelerometry measures were more likely to be due to MS and not comorbid conditions, and the baseline measures were not affected by recovery from a recent relapse. Medical records were reviewed by an MS-trained neurologist to ensure eligibility.

Participants were divided into 3 groups with a target of 85 participants each: those with stable RRMS who had no suspected or confirmed progression (RRMS-Stable), those with PMS who had confirmed disability worsening on EDSS (PMS), and those with RRMS who were suspected to be progressing clinically (inferred by an MS-trained neurologist from medical record notes or personal knowledge of the participant) but did not have sustained disability worsening on exam (RRMS-Suspected progression). Confirmed disability worsening on exam was defined as an EDSS change of ≥1.0 point if baseline EDSS was ≤5.5, or of ≥0.5 points if baseline EDSS was ≥6.0; this change should be sustained for ≥24 weeks, not in the context of an explanatory relapse. The 2 RRMS groups were matched based on age (target ±2 years with prespecified ability to relax the criterion if necessary), sex at birth, race or ethnicity, and efficacy class of current disease-modifying therapy (no treatment, first-line treatment [injectables and oral therapies (except for cladribine or ofatumumab)], or higher-efficacy treatment [infusion therapies and cladribine]). The PMS group was expected to be a little older and to be prescribed different (or even no) medications due to the likely lack of neuroinflammation and relapse activity in this group. However, enrollment of a comprehensive cohort of participants with respect to sex and race or ethnicity was attempted.

Clinical Measures

The EDSS exam was conducted after the fitting of the accelerometer and was performed by a masked EDSS-trained physician. A modified multiple sclerosis functional composite (MSFC), which includes the 9-hole peg test, the timed 25-foot walk test, the high- and low-contrast letter acuity (binocular, 2.5% contrast Sloan charts), and the Symbol Digit Modalities Test, was performed by a masked, trained study team member. Participants also completed patient-reported outcomes such as the International Physical Activity Questionnaire (IPAQ).

Accelerometry Measures

Accelerometry metrics were derived from accelerometry data collected with the GT9X Link ActiGraph using a built-in triaxial accelerometer [17]. All participants were instructed to wear the ActiGraph device on the wrist of their nondominant hand for 24 hours a day over a duration of 2 weeks. Accelerometers were set to capture 3D acceleration at 30 Hz with the acceleration range of ±8 G. The raw acceleration data (.gt3x) were downloaded from the device using ActiLife v6.134 Lite Edition. Binary raw activity data (Hz-level accelerometry data) were read by read.gt3x package into an R data frame (R Foundation for Statistical Computing) and transformed into 60-second epochs activity count data in 1440 minutes per day (12 AM to 11:59 PM) analytic format. The activity counts are vector magnitude-based activity counts.

The following criteria were applied to define valid days: (1) intervals of 90 minutes or longer with all minute-level activity counts equal to 0 were defined as nonwear intervals [18], (2)

XSL•FO
RenderX

valid days were defined as those with total wear-time ≥90% of the day (≥1296 minutes of wear), and (3) each participant should have ≥3 valid days of accelerometry data.

When a wear period overlapped with the transition to or from daylight saving time, the following adjustments were made. If the clock was set 1 hour forward, 1 hour of data were missing, so imputation was used by averaging the same hour from other valid days for this participant. If the clock was set 1 hour back, duplicate data were generated for that hour; in this instance, we used the average of the duplicated hour, and the loss of 1 hour on the final day was imputed by averaging the same hour from the other valid days for this participant. The wear period overlapped with the transition to or from daylight saving time for only 5 participants.

Accelerometry measures in this study are presented as averages over all valid days and included (1) measures of volume and intensity: total activity count (TAC; daily and 2-hour specific), total log-transformed activity count (TLAC), the total daily number of nonactive minutes, the number of minutes spent in moderate to vigorous physical activity (MVPA), and the number of minutes spent in light intensity physical activity (LIPA); (2) measures of composition: MVPA/LIPA (the ratio of the time spent in MVPA over the time spent in LIPA) and MVPA/nonactivity (the ratio of the time spent in MVPA over the time spent nonactive); (3) measures of fragmentation: sedentary-to-active transition probability (SATP) and active-to-sedentary transition probability (ASTP) [19,20]; and (4) diurnal landmarks of rest-activity rhythms: average log acceleration during the most active 10 hours of the day (M10), midpoint of M10, average log acceleration during the least active 5 hours of the day (L5), midpoint of L5, daytime activity ratio estimate (DARE), amplitude, relative amplitude (RA), midline estimating statistic of rhythm (MESOR), acrophase, intradaily variability (IV), and interdaily stability (IS) [21].

The following sample-specific cut points were developed for discriminating between nonactive, LIPA, and MVPA minutes: (1) nonactive minute was defined as a minute with activity counts ≤2000, (2) LIPA minute was defined as activity counts >2000 and ≤6750, and (3) MVPA minute was defined as activity counts >6750. These cut points were specific to our cohort and defined based on an internal calibration procedure that maximized concordance between accelerometry-based estimates and the participants' self-reported physical activity outcomes in the IPAQ (Table S1 and Figure S2 in Multimedia Appendix 1 [22]). Our derived cut points were similar to those defined in other population cohorts [22-24]. Rather than using predetermined cutoffs reported in the literature, we believe this method was more suitable for this study due to different cohort demographics and clinical conditions that may alter average physical activity intensity.

Statistical Analyses

All statistical analyses were conducted using R (version 4.3.1; R Foundation for Statistical Computing). Descriptive statistics of the demographics, clinical data, and accelerometry measures for each of the 3 groups were summarized as mean (SD) or median (IQR). Differences between the combined RRMS group (RRMS-Stable and RRMS-Suspected progression) and PMS group were compared using *t* tests and Pearson χ^2 test as appropriate. After adjusting for age, sex, and BMI, a logistic regression model was fitted to evaluate the association of each accelerometry metric with odds of PMS (vs RRMS). We also evaluated the association of accelerometry metrics in differentiating the RRMS subtypes (RRMS-Suspected progression vs RRMS-Stable). To account for differences in duration and timing of sleep, sensitivity analyses were performed by re-exploring the same accelerometry-derived measures within the most active 10 hours of each day (M10) rather than the 24-hour daily total.

Functional Principal Component Analysis

Functional principal component (fPC) analysis, applied to diurnal rest-activity rhythms (minute-level activity profiles), captures main patterns of temporal allocation of activity and provides important diurnal landmarks. The choice of the first 5 fPCs as measures of circadian rhythmicity is based on 2 factors. First, they capture most of the diurnal variability (76%) in our data. Second, the use of fPC analysis in accelerometry studies of diurnal rest-activity rhythms in large national cohorts such as the National Health and Nutrition Examination Survey (NHANES) and the UK Biobank mostly focuses on the first 4 fPCs since functional principal components beyond the 4th typically capture a relatively low proportion of variance in the data, and their shapes tend to be very difficult to interpret due to their very dynamic shape [25-27]. Nevertheless, in our case, the fifth fPC explained a significant proportion of the total variability (4%) and was also included. Figure S3 in Multimedia Appendix 1 shows that the additional fPCs beyond the 5th did not provide meaningful explanations of variability.

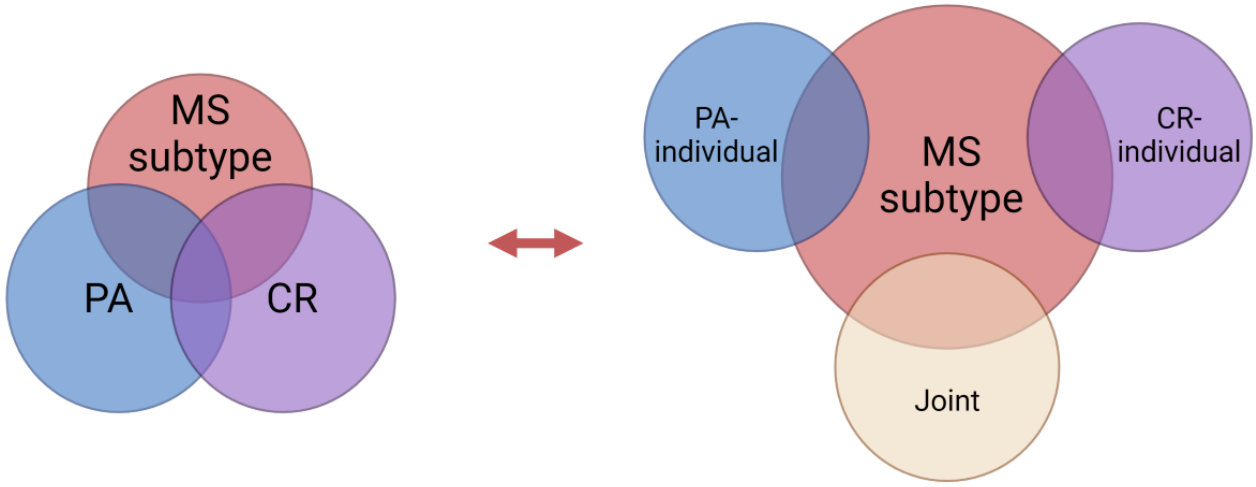
Joint and Individual Variation Explained

The Joint and Individual Variation Explained (JIVE) is an integrative dimension reduction technique that can be applied to multiple features grouped within several domains [28]. The physical activity (PA) domain was characterized by measures of total volume of physical activity, times spent in different intensities of activity and composition ratios of those times, measures of activity fragmentation/continuity, and temporal-local measures of total activity presented in 2-hour bins. The circadian rhythmicity (CR) domain was characterized by measures of the strength of circadian rhythms (RA and related components such as M10 and timing of M10), measures of variability including IS and IV, parametric measures extracted from the cosinor model (MESOR, amplitude, acrophase), and the first 5 fPCs. There is a significant amount of interdependence between the PA and CR domains since the measures belong to the same participants over the same time frame. Variations in PA and CR domains that are independent from each other are also expected. JIVE is a model developed to explore such situations and to decompose the joint and individual effects [29]. Figure 1 demonstrates the conceptual representation of the benefits of JIVE to better understand the association between MS subtype and the joint and individual variation of PA and CR domains. We applied JIVE to the domains of PA and CR to create three groups of latent variables that explain (1) joint variation shared across PA and CR, (2) individual variation

XSL•FO
RenderX

specific to each of the domains, and (3) remaining unexplained variation.

Figure 1. Conceptual representation of associations between multiple sclerosis (MS) subtype and the physical activity (PA) and circadian rhythm (CR) domains performed in standard regression modeling (left) versus the associations between MS subtype and 3 sets of independent (uncorrelated) latent variables representing joint-PA-CR, individual-PA, and individual-CR information after JIVE (Joint and Individual Variation Explained) decomposition (right). Regression analysis with JIVE components as predictors can reveal and distinguish joint and individual associations between the PA and CR domains and MS subtype.



Ethical Considerations

The study was approved by the institutional review board at Johns Hopkins University (IRB00243681). All participants provided written informed consent and had the ability to opt out at any point. All data are stored on REDCap (Research Electronic Data Capture), and only study members who are approved by the institutional review board have access to the data. The data are deidentified to the greatest extent possible. All participants received US $40 as compensation for the baseline visit to offset study-related travel and parking costs. All study procedure costs, including shipping of the ActiGraph, were covered as part of the study.

Results

Demographics and Clinical Outcomes

A total of 275 people with MS completed a baseline study visit as part of the HEAL-MS study; 253 had retrievable accelerometry data and were included in this baseline analysis (16 never returned the device, while 6 had insufficient valid days). Participants were divided into three groups: 86 with RRMS-Stable, 82 with RRMS-Suspected progression, and 85 with PMS. Baseline visits occurred between January 2021 and March 2023. Table 1 shows the demographic and clinical data for the 3 groups.

XSL•FO
RenderX

Table . Demographics and clinical data collected at the baseline visit for HEAL-MS[a] participants.

	RRMS[b]		PMS[c] (n=85)	*P* value[d]
	RRMS-Stable (n=86)	RRMS-Suspected progression (n=82)		
Demographics				
Age (years), mean (SD)	53 (7)	54 (8)	57 (9)	<.001[e]
Sex (female), n (%)	61 (71)	61 (74)	57 (67)	.36
Race, n (%)				.67
White	64 (74)	63 (77)	67 (79)	
Black	19 (22)	14 (17)	13 (15)	
Other	3 (4)	5 (6)	5 (6)	
Hispanic or Latino, n (%)	4 (5)	3 (4)	3 (4)	.79
BMI, mean (SD)	29.4 (6.4)	28.2 (5.6)	28 (6.2)	.35
Employment status, n (%)				.18
Full time	45 (58)	37 (51)	32 (43)	
Part time	4 (5)	6 (8)	7 (9)	
Homemaker	7 (9)	2 (3)	4 (5)	
Retired	13 (17)	13 (18)	18 (24)	
Unemployed	3 (4)	3 (4)	0 (0)	
Disability	6 (8)	12 (16)	14 (19)	
Clinical data				
EDSS[f] score, median (IQR)	2.0 (1.5-3.0)	3.0 (2.0-3.5)	5.0 (3.5-6.0)	<.001[e]
MSFC[g], median (IQR)				
25-foot walk	4.5 (4.0-5.1)	5.3 (4.3-6.2)	7.2 (5.8-9.5)	<.001[e]
D[h]-9hpg[i]	21.3 (19.4-24.0)	23.8 (20.9-27.3)	26.9 (22.1-34.4)	<.001[e]
ND[j]-9hpg	22.7 (20.8-25.2)	25.5 (21.9-28.9)	27.4 (23.5-34.6)	<.001[e]
SDMT[k]	51 (47-59)	48 (38-57)	42 (34-50)	<.001[e]
Tremor[l], median (IQR)	0 (0-0.5)	0 (0-0.5)	0 (0-1)	.20

[a]HEAL-MS: home-based evaluation of actigraphy to predict longitudinal function in multiple sclerosis.
[b]RRMS: relapsing-remitting multiple sclerosis.
[c]PMS: progressive multiple sclerosis.
[d]*P* value from *t* test or Pearson χ^2 test comparing RRMS (RRMS-Stable and RRMS-Suspected progression) vs PMS.
[e]*P*<.05
[f]EDSS: Expanded Disability Status Scale.
[g]MSFC: multiple sclerosis functional composite.
[h]D: dominant hand.
[i]9hpg: 9-hole peg test.
[j]ND: nondominant hand.
[k]SDMT: Symbol Digit Modalities Test.
[l]Bain Score for Tremor Severity (BSTS) was used for tremor evaluation.

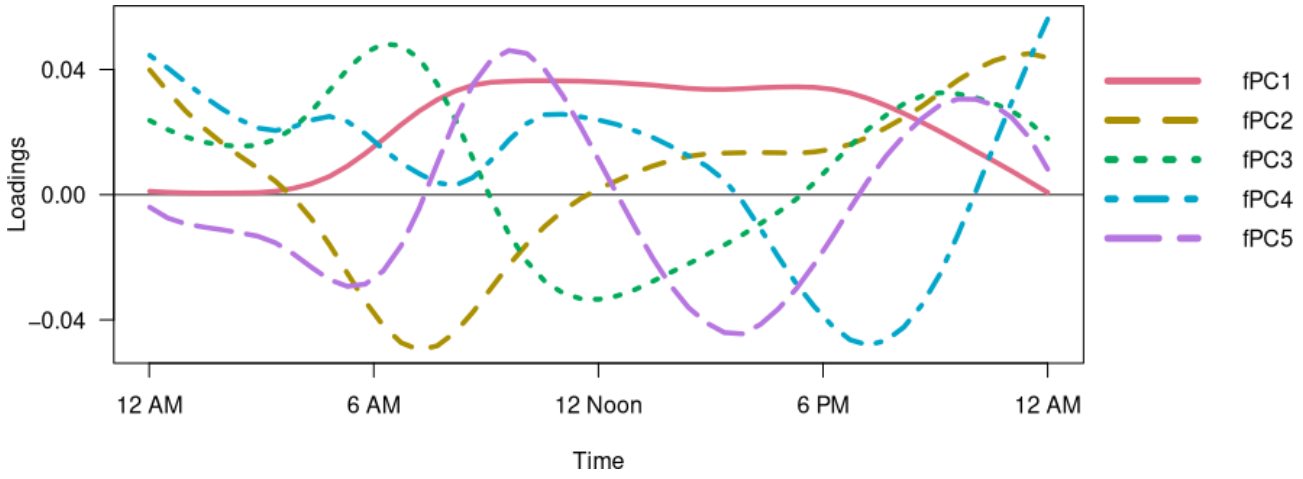
Accelerometry Measures Over 24 Hours

Figure S4 in Multimedia Appendix 1 shows the raw accelerometry data for 1 participant over the course of the week. The fPCs captured important temporal or diurnal landmarks (Figure 2) and are comparable to those captured in large national surveys such as UK Biobank and NHANES [25,27]. Accelerometry metrics for the 3 groups are shown in Table 2. TAC and MVPA, the 2 main measures of daily physical activity volume, as well as M10, amplitude, and MESOR, measures of circadian rhythmicity, were significantly different between the RRMS and PMS groups, with greater intensity of activity in

XSL•FO
RenderX

the RRMS groups compared to the PMS group. ASTP was also significantly different, with higher values (more frequent transitions to sedentary time) in the PMS group. When TAC was divided into 2-hour windows, it was significantly lower in the PMS group at specific times of the day (6 AM to 12 PM and 2 PM to 10 PM; Figure 3). Figure S5 in Multimedia Appendix 1 demonstrates the correlation plot for all accelerometry measures; the green box demonstrates the temporal profiles of the circadian measures associated with TAC over 24 hours. MESOR, amplitude, and fPC1 are highly correlated with TAC between 9 AM and 8 PM.

Figure 2. First 5 functional principal components (fPCs). fPC 2, 3, and 4 values are multiplied by −1.



XSL•FO
RenderX

Table . Mean (SD) baseline accelerometry metrics over 24 hours for HEAL-MS[a] participants.

	RRMS[b]		PMS[c] (n=85)	*P* value[d]
	RRMS-Stable (n=86)	RRMS-Suspected progression (n=82)		
Metrics of physical activity volume and fragmentation				
TAC[f] (x10^3)	2113 (566)	2181 (724)	1909 (717)	.008[e]
TLAC[g]	6277 (868)	6278 (933)	6196 (999)	.51
Nonactive minutes	1067 (102)	1052 (114)	1087 (130)	.08
MVPA[h]	59 (34)	61 (52)	41 (43)	<.001[e]
LIPA[i]	314 (84)	327 (83)	312 (105)	.52
SATP[j]	0.09 (0.03)	0.09 (0.03)	0.09 (0.03)	.21
ASTP[k]	0.27 (0.06)	0.26 (0.06)	0.30 (0.11)	.003[e]
Metrics of circadian rhythm				
M10[l]	2482 (665)	2540 (844)	2161 (806)	<.001[e]
Midpoint of M10[m]	2:20 PM (85)	2:15 PM (86)	2:14 PM (77)	.72
L5[n]	111 (67)	135 (109)	146 (124)	.09
Midpoint of L5[m]	3:58 AM (68)	3:58 AM (80)	4:05 AM (76)	.57
RA[o]	0.91 (0.05)	0.89 (0.07)	0.87 (0.09)	<.001[e]
DARE[p]	0.70 (0.05)	0.70 (0.05)	0.69 (0.05)	.049[e]
Amplitude	2290 (1081)	2218 (847)	1923 (839)	.008[e]
MESOR[q]	1313 (623)	1264 (390)	1120 (429)	.01[e]
Acrophase	14.56 (1.92)	14.61 (1.35)	14.76 (1.31)	.40
IV[r]	0.52 (0.11)	0.49 (0.09)	0.49 (0.08)	.02[e]
IS[s]	0.26 (0.07)	0.27 (0.07)	0.25 (0.06)	.08
fPC1[t]	1912 (17,574)	4242 (22,747)	−5206 (21,441)	.003[e]
fPC2	933 (11,249)	−1596 (11,922)	507 (10,181)	.59
fPC3	235 (9505)	−166 (8702)	−108 (6432)	.89
fPC4	−1489 (6339)	−585 (8694)	451 (4989)	.10
fPC5	49 (6677)	−489 (5772)	677 (5744)	.27

[a]HEAL-MS: home-based evaluation of actigraphy to predict longitudinal function in multiple sclerosis.

[b]RRMS: relapsing-remitting multiple sclerosis.

[c]PMS: progressive multiple sclerosis.

[d]*P* value from *t* test comparing RRMS (RRMS-Stable and RRMS-Suspected progression) vs PMS.

[e]*P*<.05

[f]TAC: total activity count.

[g]TLAC: total log-transformed activity count.

[h]MVPA: moderate to vigorous physical activity.

[i]LIPA: light intensity physical activity.

[j]SATP: sedentary-to-active transition probability.

[k]ASTP: active-to-sedentary transition probability.

[l]M10: average log acceleration during the most active 10 hours of the day.

[m]Midpoints of M10 and L5 are represented as time of the day in hour:minutes (±SD in minutes).

[n]L5: average log acceleration during the least active 5 hours of the day.

XSL•FO
RenderX

[o]RA: relative amplitude.

[p]DARE: daytime activity ratio estimate.

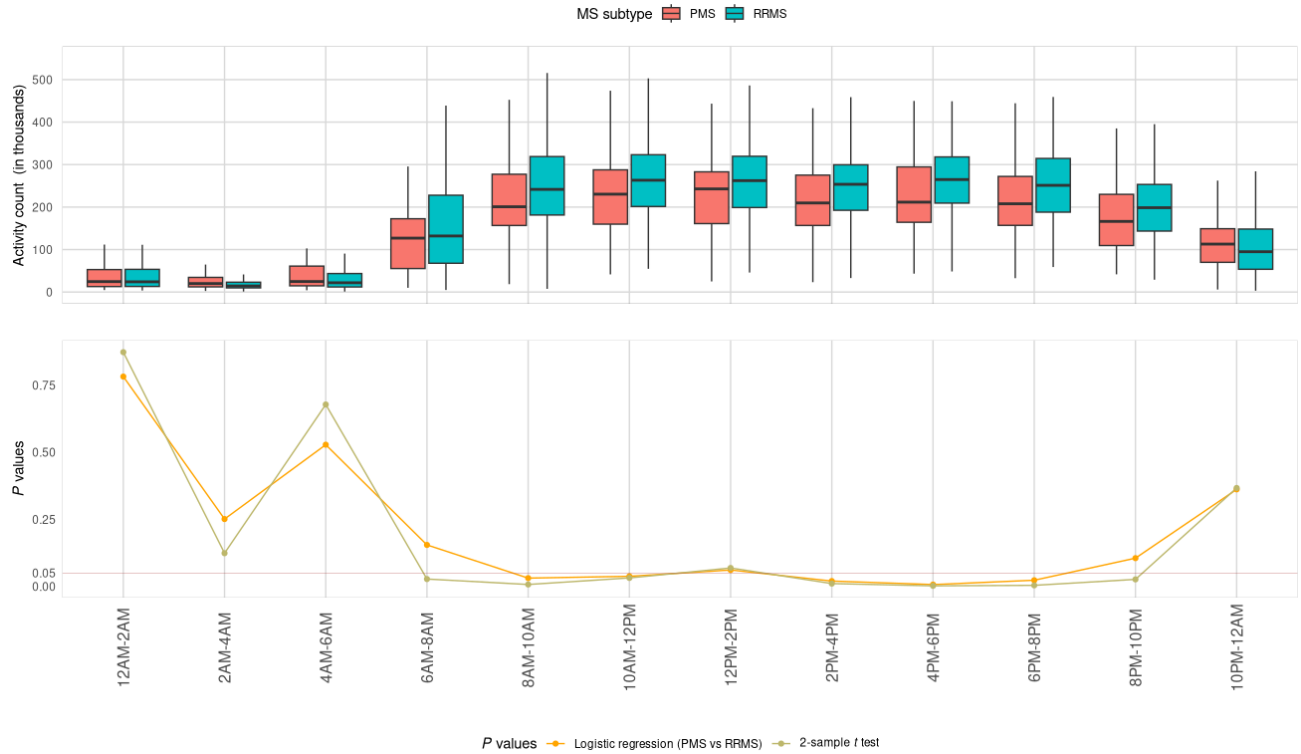
[q]MESOR: midline estimating statistic of rhythm.

[r]IV: intradaily variability.

[s]IS: interdaily stability.

[t]fPC: functional principal component.

Figure 3. Total activity counts (TAC) for each 2-hour interval over the course of 24 hours for PMS versus RRMS. The 2-sample *t* test *P* values show the 2-hour intervals when TAC was significantly different between the 2 groups. As an additional reference, the logistic regression *P* values are shown with the TAC 2-hour intervals used as predictors, after adjusting for age, sex, and BMI. MS: multiple sclerosis; PMS: progressive multiple sclerosis; RRMS: relapsing-remitting multiple sclerosis.



In multivariable-adjusted models (Table 3), participants with PMS had lower TAC (β=−0.32, 95% CI −0.61 to −0.03; *P*=.03), lower MVPA (β=−.01, 95% CI −0.02 to −0.004; *P*=.004), lower M10 (β=−0.0006, 95% CI −0.001 to −0.0002; *P*=.003), and higher ASTP (β=5.68, 95% CI 1.86-9.50; *P*=.004) when compared to those with RRMS. PMS (vs combined RRMS) was also associated with lower TAC from 8 AM to 12 PM and from 2 PM to 8 PM (Figure 3). Lower values of compositional ratios MVPA/LIPA and MVPA/nonactive were also associated with PMS (vs RRMS), indicating that individuals with PMS tend to spend relatively less time in moderate/vigorous activity compared to the time in light intensity activity or nonactivity (Table 3). Participants with PMS also had lower amplitude (β=−0.0004, 95% CI −0.0008 to −0.0001; *P*=.01), lower MESOR (β=−0.0009, 95% CI −0.002 to −0.0002; *P*=.01), higher IV (β=4.64, 95% CI 1.45-7.84; *P*=.004), lower IS (β=−4.43, 95% CI −8.77 to −0.10; *P*=.045), and lower fPC1 (β=−0.39, 95% CI −0.69 to −0.09; *P*=.01) (Table 3). There were no significant associations of accelerometry measures between RRMS subgroups except for a difference in RA (β=−5.26, 95% CI −10.80 to −0.20; *P*=.049), with lower RA in RRMS-Suspected progression (Table 3; Figure S6 in Multimedia Appendix 1).

XSL•FO
RenderX

Table . Multivariable logistic regression model to evaluate the association of each accelerometry metric with multiple sclerosis subtype. The multivariable logistic regression models were adjusted for age, sex, and BMI.

	PMS[a] vs RRMS[b] (reference group)			RRMS-Suspected progression vs RRMS-Stable (reference group)		
	Estimate	*P* value	95% CI	Estimate	*P* value	95% CI
TAC[d,e]	−0.32	.03[c]	−0.61 to −0.03	0.14	.42	−0.20 to 0.50
TLAC[d,f]	−0.04	.79	−0.31 to 0.23	0.05	.77	−0.28 to 0.38
Nonactive minutes	0.002	.16	−0.0007 to 0.004	−0.002	.29	−0.005 to 0.001
MVPA[g]	−0.01	.004[c]	−0.02 to −0.004	0.002	.70	−0.006 to 0.009
LIPA[h]	−0.0006	.69	−0.004 to 0.002	0.002	.25	−0.002 to 0.006
SATP[i]	−3.004	.59	−13.84 to 7.83	1.60	.81	−11.16 to 14.47
ASTP[j]	5.68	.004[c]	1.86 to 9.50	−4.44	.13	−10.47 to 1.05
MVPA/LIPA	−4.25	.002[c]	−6.99 to −1.51	−0.31	.81	−2.91 to 2.27
MVPA/nonactive	−7.90	.02[c]	−14.53 to −1.27	2.35	.46	−3.73 to 9.04
LIPA/nonactive	−0.36	.74	−2.50 to 1.78	1.74	.22	−0.98 to 4.58
M10[k]	−0.0006	.003[c]	−0.001 to −0.0002	0.0001	.63	−0.0003 to 0.0005
Midpoint of M10	−0.0004	.82	−0.004 to 0.003	−0.0006	.75	−0.004 to 0.003
L5[l]	0.002	.11	−0.0005 to 0.005	0.004	.07	0 to 0.008
Midpoint of L5	0.0009	.71	−0.004 to 0.005	−0.004	.16	−0.01 to 0.002
RA[m]	−6.00	.002[c]	−9.73 to −2.26	−5.26	.049[c]	−10.80 to −0.20
DARE[n]	−5.80	.04[c]	−11.25 to −0.34	−0.34	.91	−6.42 to 5.74
Amplitude	−0.0004	.01[c]	−0.0008 to −0.0001	−0.0001	.59	−0.0004 to −0.0002
MESOR[o]	−0.0009	.01[c]	−0.002 to −0.0002	−0.0002	.48	−0.0009 to 0.0004
Acrophase	0.08	.37	−0.10 to 0.27	0.02	.84	−0.17 to 0.20
IV[p]	4.64	.004[c]	1.45 to 7.84	−0.30	.88	−4.21 to 3.62
IS[q]	−4.43	.045[c]	−8.77 to −0.10	0.53	.82	−3.97 to 5.03
fPC1[r]	−0.39	.01[c]	−0.69 to −0.09	0.14	.39	−0.19 to 0.47
fPC2	0.04	.79	−0.23 to 0.31	−0.25	.12	−0.58 to 0.07
fPC3	0.08	.57	−0.19 to 0.35	0.009	.96	−0.30 to 0.32
fPC4	0.19	.18	−0.08 to 0.46	0.11	.51	−0.21 to 0.42
fPC5	0.15	.27	−0.12 to 0.42	−0.05	.74	−0.36 to 0.26

[a]PMS: progressive multiple sclerosis.

[b]RRMS: relapsing-remitting multiple sclerosis.

[c]$P<.05$.

[d]Regression inputs were scaled by dividing TAC and TLAC values by 1 SD.

[e]TAC: total activity count.

[f]TLAC: total log-transformed activity count.

[g]MVPA: moderate to vigorous physical activity.

[h]LIPA: light intensity physical activity.

[i]SATP: sedentary-to-active transition probability.

[j]ASTP: active-to-sedentary transition probability.

[k]M10: average log acceleration during the most active 10 hours of the day.

[l]L5: average log acceleration during the least active 5 hours of the day.

[m]RA: relative amplitude.

XSL•FO
RenderX

[n]DARE: daytime activity ratio estimate.

[o]MESOR: midline estimating statistic of rhythm.

[p]IV: intradaily variability.

[q]IS: interdaily stability.

[r]fPC: functional principal component.

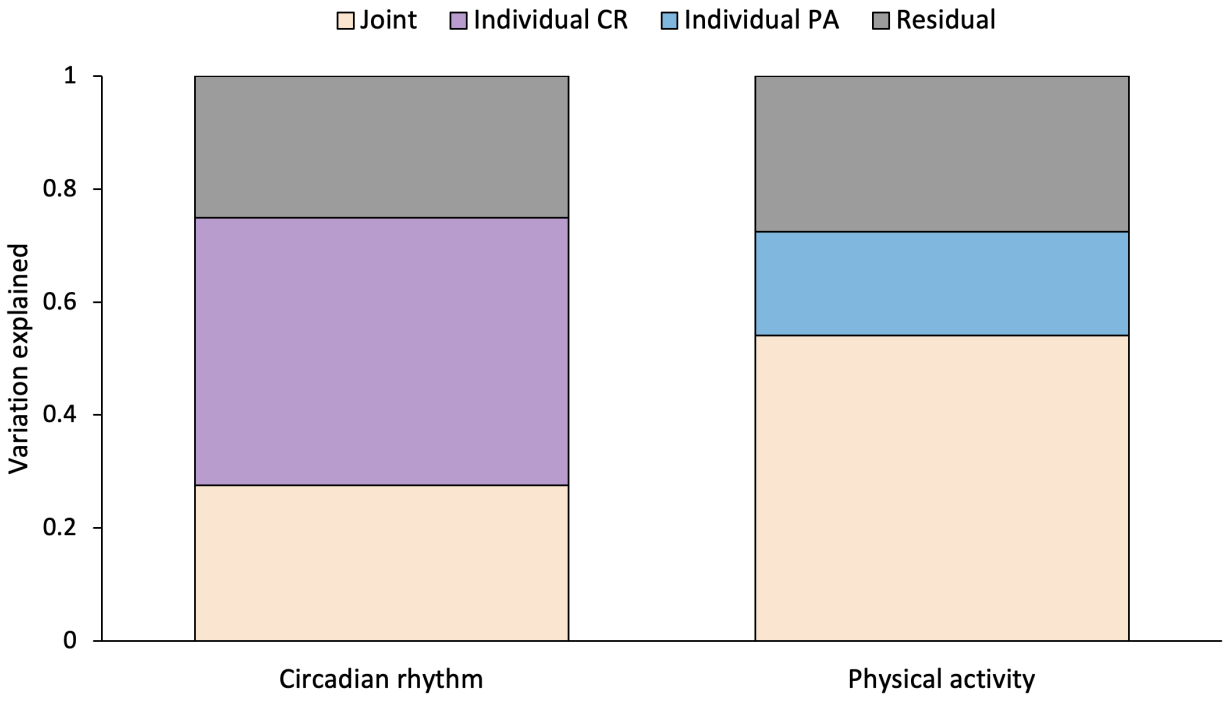
Accelerometry Measures Within M10

To account for differences in duration and timing of sleep, the accelerometry metrics were explored within the M10 period. The results are summarized in Table S2 in Multimedia Appendix 1. TAC, MVPA, and ASTP remained significantly different between the RRMS and PMS groups. Additionally, the PMS group displayed more nonactive time compared to the RRMS group. Logistic regression identified significant associations between the MS subtype (PMS vs RRMS) and TAC, MVPA, nonactive minutes, ASTP, and MVPA/LIPA (Table S3 in Multimedia Appendix 1). No significant associations were identified when comparing the 2 RRMS subgroups.

The JIVE Components

We also applied JIVE, a novel integrative dimension reduction technique. Figure 1 presents the conceptualized diagram of JIVE decomposition and how JIVE can be used to better understand codependencies between PA and CR domains and their joint and individual association with MS subtype. Figure 4 shows the estimated JIVE components obtained from the accelerometry-derived domains of PA and CR. The directions of loadings (+ or −) and squares of loadings for the estimated JIVE joint component and individual PA and CR components are shown in Tables 4 and 5, and Table S4 in Multimedia Appendix 1, respectively. Higher positive scores of the first joint JIVE component primarily represent lower volume of physical activity (captured by TAC) and weaker circadian rhythms (captured by M10, MESOR, amplitude, and fPC1) (Table 4). Higher positive individual PA-1 scores capture less frequent transitions to active behavior (SATP), less time spent in LIPA (LIPA, LIPA/nonactive, TLAC), and more time spent nonactive. Higher positive individual PA-2 scores capture more time spent in MVPA (MVPA, MVPA/LIPA, MVPA/nonactive) and more activity in the morning and evening hours (TAC 8 to 10 AM and TAC 6 PM to 12 AM) (Table 5). Joint JIVE component and individual PA-1 and PA-2 components were found to be significantly associated with MS subtype (Table 6). No significant associations were identified when comparing the 2 RRMS subgroups.

Figure 4. Joint and Individual Variation Explained (JIVE) by the 2 JIVE accelerometry-derived domains of circadian rhythm (CR) and physical activity (PA).



XSL•FO
RenderX

Table . JIVE[a] joint component with variables with more than 5% of proportional variation.

Variable	Joint component loading	Relative magnitude of joint component loading
M10[b]	−0.2665385	0.0710428
fPC1[c]	−0.2608276	0.0680310
MESOR[d]	−0.2519508	0.0634792
Amplitude	−0.2408992	0.0580324
TAC[e]	−0.2256510	0.0509184

[a]JIVE: Joint and Individual Variation Explained.

[b]M10: average log acceleration during the most active 10 hours of the day.

[c]fPC: functional principal component.

[d]MESOR: midline estimating statistic of rhythm.

[e]TAC: total activity count.

XSL•FO
RenderX

Table . Individual JIVE[a] components of physical activity with variables with more than 5% of proportional variation.

Variable	PA component loading	Relative magnitude of PA component loading
Physical activity component 1		
SATP[b]	−0.4493703	0.2019337
LIPA[c]	−0.4141149	0.1714912
LIPA/nonactive	−0.3664975	0.1343204
TLAC[d]	−0.3365927	0.1132946
MVPA/LIPA[e]	0.3293182	0.1084505
Nonactive minutes	0.2448185	0.0599361
Physical activity component 2		
MVPA/LIPA	0.4728996	0.2236340
MVPA	0.3970161	0.1576218
MVPA/nonactive	0.3673462	0.1349432
TAC[f] 8 PM to 10 PM	0.3341579	0.1116615
TAC 8 AM to 10 AM	0.3099402	0.0960629
TAC 10 PM to 12 AM	0.2477134	0.0613620
TAC 6 PM to 8 PM	0.2430202	0.0590588
Physical activity component 3		
TAC 12 AM to 2 AM	0.6035690	0.3642955
TAC 10 PM to 12 AM	0.4550918	0.2071086
TAC 4 AM to 6 AM	−0.3500618	0.1225433
TAC 2 AM to 4 AM	−0.3085089	0.0951777
Physical activity component 4		
ASTP[g]	−0.5134849	0.2636667
TAC 8 PM to 10 PM	0.3803549	0.1446698
TAC 2 PM to 4 PM	−0.3588593	0.1287800
SATP	−0.3283307	0.1078011
TLAC	−0.2739182	0.0750312
TAC 10 AM to 12 PM	0.2516689	0.0633372

[a]JIVE: Joint and Individual Variation Explained.
[b]SATP: sedentary-to-active transition probability.
[c]LIPA: light intensity physical activity.
[d]TLAC: total log-transformed activity count.
[e]MVPA: moderate to vigorous physical activity.
[f]TAC: total activity count.
[g]ASTP: active-to-sedentary transition probability.

XSL•FO
RenderX

Table . Multivariable logistic regression model to evaluate the association of each JIVE[a] component with multiple sclerosis subtype. The multivariable logistic regression models were adjusted for age, sex, and BMI.

	PMS[b] vs RRMS[c] (reference group)			RRMS-Suspected progression vs RRMS-Stable (reference group)		
	Estimate	*P* value	95% CI	Estimate	*P* value	95% CI
Joint component	0.3665	.01[d]	0.0881 to 0.6558	−0.1266	.44	−0.4525 to 0.1897
PA-1[e]	−0.441	.003[d]	−0.7397 to −0.159	−0.0183	.91	−0.3404 to 0.3034
PA-2	−0.4147	.006[d]	−0.7174 to −0.1255	−0.3066	.07	−0.6502 to 0.0207
PA-3	−0.0527	.71	−0.3282 to 0.2204	−0.2622	.11	−0.5887 to 0.0529
PA-4	0.235	.11	−0.046 to 0.5272	0.1849	.25	−0.129 to 0.5098
CR-1[f]	0.2588	.07	−0.0162 to 0.5411	0.0107	.95	−0.3044 to 0.3275
CR-2	−0.2732	.052	−0.5536 to 0.0001	−0.2991	.07	−0.6344 to 0.019
CR-3	−0.142	.32	−0.4266 to 0.135	0.0815	.61	−0.2345 to 0.4031

[a]JIVE: Joint and Individual Variation Explained.

[b]PMS: progressive multiple sclerosis.

[c]RRMS: relapsing-remitting multiple sclerosis.

[d]$P<.05$.

[e]PA: individual physical activity component.

[f]CR: individual circadian rhythm component.

Discussion

Principal Findings

This cross-sectional analysis was conducted to determine if a set of detailed accelerometry-derived summary measures of physical activity and circadian rhythmicity distinguishes between MS subtypes. As an extension of the work we and others had previously published, in which physical activity counts were on average lower in those with MS who were more disabled [9-11], we show here that participants with PMS exerted less total daily volume of physical activity than those with RRMS, and that these differences seemed more pronounced during specific times of the day. Sensitivity analyses to account for differences in timing and duration of sleep using the accelerometry metrics during the M10 period demonstrated very similar findings, which, if confirmed in subsequent studies, might help improve the feasibility of data collection (eg, if participants are not eager to wear the device overnight).

This work newly demonstrates that beyond average overall activity measures, specific PA indices had stronger discriminative properties than others in distinguishing PMS from RRMS. For example, those with PMS were more likely to transition from active to nonactive behavior (higher ASTP) and had less time spent in MVPA, instead favoring a greater amount of time in which they are inactive or only lightly active. ASTP may be a complementary measure of functionality to the measures of volume of daily activity such as TAC and MVPA. In the general population, higher ASTP is associated with worse health and functional status and higher fatiguability in older adults [20]. Our findings are consistent, as people with PMS have overall greater disability and symptoms. Therefore, it may be useful for future studies to explore novel measures of fragmentation of physical activity to determine indices that could better distinguish MS subtype.

Although differences in patterns of activity were demonstrated between the PMS and RRMS subgroups, exploring accelerometry-measured sleep and circadian/diurnal rhythmicity may discern even more significant differences. For example, we previously demonstrated that among people with MS, diurnal variability is greater in those with higher EDSS [9]. Herein, we confirmed that differences in TAC within specific times of the day (8 AM to 12 PM; 2 PM to 8 PM) appeared to drive the daily average activity measures. Participants with PMS also had higher IV and lower IS, further demonstrating the importance of detecting differences in diurnal rhythmicity between the 2 groups. The observed differences in our cohort herein serve as motivation for further exploration with more advanced analytical techniques beyond the use of averages of daily CR measures. Our next steps will include further investigation of more complex measures of fragmentation of physical activity, distributional and temporally local distributional characteristics of activity, and parameters of sleep and circadian disruptions [19,28,30-32].

An application of JIVE demonstrated a significant overlap between variation in physical activity and circadian rhythmicity domains. Joint component 1 was significant and was primarily driven by the total volume of PA along with measures of strength of CR (M10, fPC1, MESOR, amplitude), suggesting that a joint study of these 2 domains may provide a more accurate picture and capture the interdependence between the domains and their combined association with MS subtype. The significant association of MS subtype with the individual PA-1 and PA-2 components confirmed the results from the standalone associations of individual measures with MS subtype. Specifically, more fragmented physical activity profiles with

XSL•FO
RenderX

lower times spent in MVPA and LIPA can be used to differentiate PMS and RRMS.

Accelerometry did not distinguish between the RRMS subgroups cross-sectionally. Correctly classifying people with MS as having transitioned from RRMS to PMS often takes several years and is usually done retrospectively [33]. In designing this cohort, we hypothesized that data acquired in the early phase of cohort follow-up would predict subsequent worsening as measured by EDSS and that accelerometry may help to reduce the time of diagnostic uncertainty. Since we report only cross-sectional data, we did not expect substantial differences between the RRMS groups, especially as these groups were matched for key characteristics (age, sex, race or ethnicity, and disease modifying therapy class). Longitudinal tracking of the participants will allow for tracking within-person changes in disability trajectories. Whether including additional features of accelerometry data, such as more advanced measures of sleep or circadian rhythm, will better distinguish RRMS subpopulations even at a single time point will be an important next step.

Finally, although our findings are currently not immediately applicable in clinical practice, the study suggests it may be possible to alert clinicians about certain patterns in 24-hour behaviors that may affect the lives of people with MS. Though it can be difficult to clinically distinguish PMS from RRMS, accelerometry measures appear to detect distinct patterns of activity. People with PMS seem to have increased variability and fragmentation of their activity patterns, and intermittently reduced function is commonly reported, subjectively, by people with more advanced MS during clinical visits. This increases the confidence that accelerometry can in fact measure clinically meaningful patterns in people with MS. It is important for future efforts to recapitulate these findings with commercial accelerometers in order to support scalability and to translate the findings into results that clinicians and patients can easily understand and monitor over time. For instance, the increased variability in people with PMS indicates that their experience of their disease may be very different from the clinical picture evident during their clinical visits. A discussion with patients regarding their physical activity, supplemented by accelerometry data, could shed light on other aspects of their lives that should be acknowledged. Understanding when patients are likely to be at their peak activity levels may help plan their day better (when to plan effortful tasks or schedule breaks) and may inform lifestyle modifications such as exercise or even physical therapy interventions to evaluate for improvements.

Limitations

There are limitations to the study. The classification of RRMS as stable versus suspected progression is subjective, and it is likely that some of these individuals were misclassified, where a stable person was deemed as having suspected progression or vice versa, which could contribute to the lack of ability to detect differences in accelerometry measures between the 2 RRMS groups. We chose to create this subclassification to mirror the "clinical intuition" whereby physicians and, often, their patients worry they are progressing even if they do not meet the criteria for sustained EDSS worsening, and we wanted to be sure that sufficient numbers of such people were enrolled in the study. The longitudinal trajectories of the participants with RRMS will be informative; since all are at risk, based on age, of transitioning to a PMS phenotype, we can evaluate this risk as it relates to accelerometry measures not only using these subjective subtypes but also using a more agnostic, data-driven approach within the participants with RRMS subgroup as a whole. Another potential bias related to our study is the Hawthorne effect, where participants could potentially alter their physical activity because they were being monitored. Although this risk cannot be eliminated, it was minimized by disabling the features that display activity levels on the ActiGraph monitors (ie, the number of steps); only the time of day was displayed. It is possible that relevant comorbidities were missed during chart review although this is unlikely to affect a substantial enough proportion of individuals to meaningfully have influenced results. Finally, this is a cross-sectional study, and we will be able to better explore the dynamic changes in activity measured by accelerometry as well as fluctuations in the EDSS with the longitudinal follow-up, which will provide deeper insight into the evolution of MS progression measures.

Conclusions

In conclusion, accelerometry-derived measures captured differences in physical activity and circadian rhythm profiles between people with PMS and RRMS. Further analytic approaches that consider more sophisticated approaches to time (such as circadian patterns) may help further refine the ability to detect differences between these subgroups at a single time point, whereas longitudinal follow-up is most likely to allow for the identification of accelerometry patterns that portend subsequent disability worsening as defined by EDSS. We specifically hypothesize that the measures that distinguish PMS from RRMS at baseline will be those to change first in the participants with RRMS who transition to PMS in the longitudinal follow-up of this cohort. Enabling the identification of MS progression or risk may help with clinical decision-making (such as choosing an appropriate disease-modifying therapy for a given person) or may enable advancements in telemedicine by tracking functions remotely. More importantly, these technologies will likely facilitate therapeutic development for PMS by improving disability detection and worsening, alone or in combination with other metrics such as neuroimaging, which will enable faster and smaller trials of promising interventions.

Acknowledgments

This study was funded by the National Institute of Nursing Research, National Institutes of Health (grant 5R01NR018851). The funder played no role in the study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

XSL•FO
RenderX

We would also like to thank MS4MS for their support of our wearables program.

Data Availability

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request and with appropriate approvals in place. The source code associated with this study is available upon request and will be made publicly accessible on GitHub upon the publication of this paper.

Conflicts of Interest

EMM discloses research funding from Biogen and Roche/Genentech, consulting for BeCareLink, LLC, and royalties for editorial duties for UpToDate. PAC is/has been Principal Investigator on grants to Johns Hopkins University from Genentech and Principia. He has served as a paid consultant for Lilly, Avidea Technologies, Idorsia, Nervgen, and Biogen. All other authors have no disclosures.

Multimedia Appendix 1
Illustrations of the study framework along with additional statistical analyses and sensitivity checks.
[DOCX File, 796 KB - mhealth_v13i1e57599_app1.docx]

References

1. Inojosa H, Schriefer D, Ziemssen T. Clinical outcome measures in multiple sclerosis: a review. Autoimmun Rev 2020 May;19(5):102512. [doi: 10.1016/j.autrev.2020.102512] [Medline: 32173519]
2. Albrecht H, Wötzel C, Erasmus LP, Kleinpeter M, König N, Pöllmann W. Day-to-day variability of maximum walking distance in MS patients can mislead to relevant changes in the Expanded Disability Status Scale (EDSS): average walking speed is a more constant parameter. Mult Scler 2001 Apr;7(2):105-109. [doi: 10.1177/135245850100700206] [Medline: 11424630]
3. Christogianni A, Bibb R, Davis SL, et al. Temperature sensitivity in multiple sclerosis: an overview of its impact on sensory and cognitive symptoms. Temperature (Austin) 2018 Jul 3;5(3):208-223. [doi: 10.1080/23328940.2018.1475831]
4. Katz Sand I, Krieger S, Farrell C, Miller AE. Diagnostic uncertainty during the transition to secondary progressive multiple sclerosis. Mult Scler 2014 Oct;20(12):1654-1657. [doi: 10.1177/1352458514521517]
5. Inojosa H, Proschmann U, Akgün K, Ziemssen T. A focus on secondary progressive multiple sclerosis (SPMS): challenges in diagnosis and definition. J Neurol 2021 Apr;268(4):1210-1221. [doi: 10.1007/s00415-019-09489-5] [Medline: 31363847]
6. Block VJ, Lizée A, Crabtree-Hartman E, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. J Neurol 2017 Feb;264(2):316-326. [doi: 10.1007/s00415-016-8334-6] [Medline: 27896433]
7. Dalla-Costa G, Radaelli M, Maida S, et al. Smart watch, smarter EDSS: improving disability assessment in multiple sclerosis clinical practice. J Neurol Sci 2017 Dec 15;383:166-168. [doi: 10.1016/j.jns.2017.10.043] [Medline: 29246607]
8. Klaren RE, Motl RW, Dlugonski D, Sandroff BM, Pilutti LA. Objectively quantified physical activity in persons with multiple sclerosis. Arch Phys Med Rehabil 2013 Dec;94(12):2342-2348. [doi: 10.1016/j.apmr.2013.07.011] [Medline: 23906692]
9. Keller JL, Tian F, Fitzgerald KC, et al. Using real-world accelerometry-derived diurnal patterns of physical activity to evaluate disability in multiple sclerosis. J Rehabil Assist Technol Eng 2022;9. [doi: 10.1177/20556683211067362] [Medline: 35070348]
10. Abbadessa G, Lavorgna L, Miele G, et al. Assessment of multiple sclerosis disability progression using a wearable biosensor: a pilot study. J Clin Med 2021 Mar 10;10(6):1-8. [doi: 10.3390/jcm10061160] [Medline: 33802029]
11. Ezeugwu V, Klaren RE, Hubbard EA, Manns PT, Motl RW. Mobility disability and the pattern of accelerometer-derived sedentary and physical activity behaviors in people with multiple sclerosis. Prev Med Rep 2015;2:241-246. [doi: 10.1016/j.pmedr.2015.03.007] [Medline: 26844077]
12. Cederberg KLJ, Mathison BG, Schuetz ML, Motl RW. Discrepancies between self-reported and device-measured sleep parameters in adults with multiple sclerosis. J Clin Sleep Med 2022 Feb 1;18(2):415-421. [doi: 10.5664/jcsm.9586] [Medline: 34338630]
13. Gulde P, Rieckmann P. The association between actigraphy-derived behavioral clusters and self-reported fatigue in persons with multiple sclerosis: cross-sectional study. JMIR Rehabil Assist Technol 2022 Mar 17;9(1):e31164. [doi: 10.2196/31164] [Medline: 35297774]
14. Neal WN, Cederberg KL, Jeng B, Sasaki JE, Motl RW. Is symptomatic fatigue associated with physical activity and sedentary behaviors among persons with multiple sclerosis? Neurorehabil Neural Repair 2020 Jun;34(6):505-511. [doi: 10.1177/1545968320916159] [Medline: 32340521]
15. Jones CD, Jeng B, Silic P, Motl RW. Do device-measured physical activity and sedentary behavior differ by depression symptom status in persons with multiple sclerosis? Mult Scler Relat Disord 2022 Jul;63:103889. [doi: 10.1016/j.msard.2022.103889] [Medline: 35636270]

XSL•FO
RenderX

16. Macdonald E, Buchan D, Cerexhe L, Renfrew L, Sculthorpe N. Accelerometer measured physical activity and sedentary time in individuals with multiple sclerosis versus age matched controls: a systematic review and meta-analysis. Mult Scler Relat Disord 2023 Jan;69:104462. [doi: 10.1016/j.msard.2022.104462] [Medline: 36521386]
17. User Guide: ActiGraph GT9X Link + ActiLife, Revision 6. Amazon S3. 2020 Feb 18. URL: https://s3.amazonaws.com/actigraphcorp.com/wp-content/uploads/2020/03/05155628/ActiGraph_Link_UserGuide_E.200.6001_Revision6_FINAL.pdf [accessed 2023-09-03]
18. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364. [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
19. Di J, Leroux A, Urbanek J, et al. Patterns of sedentary and active time accumulation are associated with mortality in US adults: the NHANES study. bioRxiv. Preprint posted online on Aug 31, 2017. [doi: 10.1101/182337]
20. Schrack JA, Kuo PL, Wanigatunga AA, et al. Active-to-sedentary behavior transitions, fatigability, and physical functioning in older adults. J Gerontol A Biol Sci Med Sci 2019 Mar 14;74(4):560-567. [doi: 10.1093/gerona/gly243] [Medline: 30357322]
21. Leroux A, Xu S, Kundu P, et al. Quantifying the predictive performance of objectively measured physical activity on mortality in the UK Biobank. J Gerontol A Biol Sci Med Sci 2021 Jul 13;76(8):1486-1494. [doi: 10.1093/gerona/glaa250] [Medline: 33000171]
22. Welk GJ, Lamoureux NR, Zeng C, et al. Equating NHANES monitor-based physical activity to self-reported methods to enhance ongoing surveillance efforts. Med Sci Sports Exerc 2023 Jun 1;55(6):1034-1043. [doi: 10.1249/MSS.0000000000003123] [Medline: 36633833]
23. Montoye AHK, Clevenger KA, Pfeiffer KA, et al. Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. J Sports Sci 2020 Nov;38(22):2569-2578. [doi: 10.1080/02640414.2020.1794244] [Medline: 32677510]
24. Rhudy MB, Dreisbach SB, Moran MD, Ruggiero MJ, Veerabhadrappa P. Cut points of the Actigraph GT9X for moderate and vigorous intensity physical activity at four different wear locations. J Sports Sci 2020 Mar;38(5):503-510. [doi: 10.1080/02640414.2019.1707956] [Medline: 31865845]
25. Xiao Q, Lu J, Zeitzer JM, Matthews CE, Saint-Maurice PF, Bauer C. Rest-activity profiles among U.S. adults in a nationally representative sample: a functional principal component analysis. Int J Behav Nutr Phys Act 2022 Mar 24;19(1):32. [doi: 10.1186/s12966-022-01274-4] [Medline: 35331274]
26. Yeung CHC, Lu J, Soltero EG, Bauer C, Xiao Q. U.S. adolescent rest-activity patterns: insights from functional principal component analysis (NHANES 2011-2014). Int J Behav Nutr Phys Act 2023 Oct 13;20(1):125. [doi: 10.1186/s12966-023-01520-3] [Medline: 37833691]
27. Stein MJ, Baurecht H, Sedlmeier AM, et al. Association between circadian physical activity patterns and mortality in the UK Biobank. Int J Behav Nutr Phys Act 2023 Sep 1;20(1):102. [doi: 10.1186/s12966-023-01508-z] [Medline: 37653438]
28. Di J, Spira A, Bai J, et al. Joint and individual representation of domains of physical activity, sleep, and circadian rhythmicity. Stat Biosci 2019 Jul;11(2):371-402. [doi: 10.1007/s12561-019-09236-4] [Medline: 32440309]
29. Lock EF, Hoadley KA, Marron JS, Nobel AB. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. Ann Appl Stat 2013 Mar 1;7(1):523-542. [doi: 10.1214/12-AOAS597] [Medline: 23745156]
30. Salomon A, Galperin I, Buzaglo D, et al. Fragmentation, circadian amplitude, and fractal pattern of daily-living physical activity in people with multiple sclerosis: is there relevant information beyond the total amount of physical activity? Mult Scler Relat Disord 2022 Dec;68:104108. [doi: 10.1016/j.msard.2022.104108] [Medline: 36063732]
31. Ghosal R, Varma VR, Volfson D, et al. Distributional data analysis via quantile functions and its application to modeling digital biomarkers of gait in Alzheimer's disease. Biostatistics 2023 Jul 14;24(3):539-561. [doi: 10.1093/biostatistics/kxab041] [Medline: 36519565]
32. Ghosal R, Ghosh SK, Schrack JA, Zipunnikov V. Distributional outcome regression and its application to modelling continuously monitored heart rate and physical activity. arXiv. Preprint posted online on Jan 26, 2023. [doi: 10.48550/arXiv.2301.11399]
33. Ziemssen T, Bhan V, Chataway J, et al. Secondary progressive multiple sclerosis: a review of clinical characteristics, definition, prognostic tools, and disease-modifying therapies. Neurol Neuroimmunol Neuroinflamm 2023 Jan;10(1):381-384. [doi: 10.1212/NXI.0000000000200064] [Medline: 36414428]

Abbreviations

ASTP: active-to-sedentary transition probability
CR: circadian rhythm
DARE: daytime activity ratio estimate
EDSS: Expanded Disability Status Scale
fPC: functional principal component
HEAL-MS: home-based evaluation of actigraphy to predict longitudinal function in multiple sclerosis
IPAQ: International Physical Activity Questionnaire

XSL•FO
RenderX

IS: interdaily stability
IV: intradaily variability
JIVE: Joint and Individual Variation Explained
L5: average log acceleration during the least active 5 hours of the day
LIPA: light intensity physical activity
M10: average log acceleration during the most active 10 hours of the day
MESOR: midline estimating statistic of rhythm
MS: multiple sclerosis
MVPA: moderate to vigorous physical activity
NHANES: National Health and Nutrition Examination Survey
PA: physical activity
PMS: progressive multiple sclerosis
RA: relative amplitude
REDCap: Research Electronic Data Capture
RRMS: relapsing-remitting multiple sclerosis
SATP: sedentary-to-active transition probability
TAC: total activity count
TLAC: total log-transformed activity count

Edited by L Buis; submitted 01.03.24; peer-reviewed by C Park, I Vujovic; revised version received 14.12.24; accepted 04.02.25; published 31.03.25.

Please cite as:
Bou Rjeily N, Sanjayan M, Guha Niyogi P, Dewey BE, Zambriczki Lee A, Hulett C, Dagher G, Hu C, Mazur RD, Kenney EM, Brennan E, DuVal A, Calabresi PA, Zipunnikov V, Fitzgerald KC, Mowry EM
Accelerometry-Assessed Physical Activity and Circadian Rhythm to Detect Clinical Disability Status in Multiple Sclerosis: Cross-Sectional Study
JMIR Mhealth Uhealth 2025;13:e57599
URL: https://mhealth.jmir.org/2025/1/e57599
doi:10.2196/57599

© Nicole Bou Rjeily, Muraleetharan Sanjayan, Pratim Guha Niyogi, Blake E Dewey, Alexandra Zambriczki Lee, Christy Hulett, Gabriella Dagher, Chen Hu, Rafal D Mazur, Elena M Kenney, Erin Brennan, Anna DuVal, Peter A Calabresi, Vadim Zipunnikov, Kathryn C Fitzgerald, Ellen M Mowry. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of Integrating Wearable Activity Trackers With a Home-Based Multicomponent Exercise Intervention on Fall-Related Parameters and Physical Function in Older Adults: Randomized Controlled Trial

Yejin Kim[1], MS; Kyung Hee Park[2], MPH, MD, PhD; Hye-Mi Noh[2], MD, PhD

[1]Department of Medical Sciences, College of Medicine, Hallym University, Chuncheon, Republic of Korea

[2]Department of Family Medicine, Hallym University Sacred Heart Hospital, Hallym University, 22, Gwanpyeong-ro 170beon-gil, Anyang, Republic of Korea

Corresponding Author:
Hye-Mi Noh, MD, PhD
Department of Family Medicine, Hallym University Sacred Heart Hospital, Hallym University, 22, Gwanpyeong-ro 170beon-gil, Anyang, Republic of Korea

Abstract

Background: Older adults with a history of falling often encounter challenges in participating in group exercise programs. Recent technological advances, such as activity trackers, can potentially enhance home-based exercise programs by providing continuous physical activity monitoring and feedback.

Objective: The aim of the study is to explore whether integrating wearable activity trackers with a home-based exercise intervention is effective in reducing fear of falling and improving physical function in older adults.

Methods: This was a 12-week, parallel-group, randomized controlled trial involving 30 older adults (≥60 years) with a history of falling. Participants were randomly assigned in a 1:1 ratio to either a group combining an activity tracker with a home-based multicomponent exercise intervention, which included in-person exercise sessions, exercise videos, and objective feedback via phone calls (AT+EX group) or to a group using the activity tracker only for self-monitoring (AT-only group). The primary and secondary outcomes included fall-related parameters (fear of falling assessed by the Activities-Specific Balance Confidence [ABC] and the Falls Efficacy Scale-International [FES-I] scales), depression (Short Geriatric Depression Scale), cognition (Montreal Cognitive Assessment), physical function (grip strength, Short Physical Performance Battery, Timed Up and Go [TUG] test, and 2-Minute Step Test), and body composition. Changes in the average daily step count were monitored and analyzed.

Results: Overall, 28 (mean age 74.0, SD 6.4 years; n=23, 77% female) participants completed the 12-week follow-up period (28/30, 93%). In the activity tracker and exercise group (AT+EX group), significant improvements were observed in fear of falling (15.5 points of ABC: *P*=.002; –5.1 points of FES-I: *P*=.01). The activity tracker alone group (AT-only group) also showed a significant improvement in FES-I score (–5.5 points: *P*=.01). Physical function significantly improved in the AT+EX group (1.1 points of Short Physical Performance Battery: *P*=.004; –1.4 seconds of TUG; *P*=.008; and 26.7 steps of 2-Minute Step Test: *P*=.001), whereas the AT-only group showed significant improvement only in the TUG test (–1.3 seconds: *P*=.002). However, no significant between-group differences were observed in the ABC score, FES-I score, or physical function. Despite no significant increase in daily step counts, both groups maintained close to 10,000 steps per day throughout the 12 weeks.

Conclusions: Both groups showed improvements in the FES-I and TUG test scores without significant between-group differences. Wearable technology, with or without an exercise intervention, seems to be an effective tool in reducing the fear of falling and improving physical function in older adults susceptible to falls.

Trial Registration: Clinical Research Information Service KCT0008230; https://tinyurl.com/3au6kfzu

(JMIR Mhealth Uhealth 2025;13:e64458) doi:10.2196/64458

KEYWORDS

fear of falling; physical function; home exercise; multicomponent exercise; older adults; physical activity intervention; step counts; wearable activity tracker; exercise; intervention; gerontology; geriatrics; older person; aging; activity tracker; physical activity; self-monitoring; randomized controlled trial; wearable technology; fall; falling

XSL•FO
RenderX

Introduction

Falls are a significant health issue in older adults and are among the most prevalent geriatric syndromes [1]. The consequences of falls can be severe, resulting in injuries such as fractures, bleeding, and disability, leading to increased health care costs [2]. Falls increase the likelihood of nursing home admissions among older adults and are a leading cause of accidental death [3]. According to the Centers for Disease Control and Prevention, 27.6% of individuals aged >65 years in the United States reported falling during the previous year, and 38,742 (78.0 per 100,000) died from unintentional falls in 2020 [4]. In Korea, the 2017 National Survey of Korean Elderly reported that 15.9% of older adults experienced falls in the past year, with an average of 2.1 falls per person [5]. A history of falls is a significant risk factor for future falls, as approximately 1 in 2 individuals who have experienced falls may experience recurrent falls [6]. Additionally, the fear of falling, characterized by anxiety concerning falling or a loss of confidence in certain activities, leads to decreased physical activity and an increased risk of further falls [7].

Numerous studies have demonstrated the effectiveness of exercise interventions in preventing falls and improving physical function in older adults [8,9]. Each type of exercise (ie, resistance, aerobic, and balance training) has specific benefits, and incorporating high-challenge balance training in exercise programs is particularly emphasized for fall prevention [9,10]. Resistance training is reportedly effective in preventing age-related decline in muscle function, especially with high-intensity training [11]. Clinical guidelines from leading health and geriatric organizations recommend that older adults participate in multicomponent exercise programs, including balance and resistance training, as well as aerobic and flexibility activities, to improve physical function and reduce the risk of falls [12-14].

Effective doses and consistent exercise are crucial for maintaining health in older adults [9]. However, adherence to home-based exercises is typically lower compared to that for community-based exercises, and self-reported physical activity levels often exceed actual levels [15]. Although many group exercise programs have been conducted for older adults in the community, the onset of the COVID-19 pandemic has made it challenging to continue group exercise sessions [16]. Moreover, factors, such as low confidence in physical activity and self-efficacy, further hinder participation in group exercises [17]. Older adults also face restrictions in participating in physical activity programs because of physical and mental health issues (ie, risk of injury, joint pain, and fear of falling) and structural and organizational barriers (ie, access to facilities, limited infrastructure, and cost) [18].

Recently, various wearable devices have been developed and are increasingly used as popular devices for monitoring physical activity. Activity trackers provide individuals with objective measures of physical activity levels and enable health professionals to offer precise feedback and sustained support [19]. The use of activity trackers, either as the primary element or within a broader behavioral intervention, effectively increases physical activity participation [20]. Moreover, integrating activity trackers with traditional intervention components is more effective than using an activity tracker alone [21].

Despite these advancements, there is a lack of research on the integration of activity trackers with other multifaceted interventions for falls in older adults [22,23]. Furthermore, to date, no study has evaluated the effect of combining activity trackers with home-based multicomponent exercises in older adults. Therefore, our aim was to assess the effectiveness of integrating activity trackers with home-based interventions in older adults with a history of falls. We hypothesized that combining an activity tracker with a home-based multicomponent exercise intervention (activity tracker and exercise group [AT+EX group]) would result in greater improvements in fall-related parameters and physical function compared to using an activity tracker alone (activity tracker alone group [AT-only group]).

Methods

Study Design

This 12-week, parallel-group, randomized controlled trial involved older adults aged ≥60 years with a history of falls and COVID-19 in Anyang, Gyeonggi-do, South Korea. Participants were randomly assigned in a 1:1 ratio to either the AT+EX or AT-only group. The AT+EX group received multicomponent exercise sessions, exercise videos, and objective feedback from an activity tracker, whereas the AT-only group received activity trackers without additional interventions. A CONSORT (Consolidated Standards of Reporting Trials) checklist was completed to guide the design, analysis, and reporting of trial findings (Checklist 1).

Participants

Participants were recruited through poster advertisements, flyers, word-of-mouth referrals, and visits to local senior welfare centers. Eligible individuals met the following criteria: (1) aged ≥60 years, (2) history of COVID-19 infection within 1 year from their first study visit, (3) experience with falls, (4) ownership of a smartphone compatible with wearable devices, (5) ability to walk independently without assistance, and (6) voluntary decision to participate in the study with a signed informed consent form before participation. The exclusion criteria were as follows: (1) diagnosis of cognitive decline or dementia and (2) contraindications to exercise, including myocardial infarction within the last 3 months, unstable angina pectoris, uncontrolled diabetes, or musculoskeletal disease that prevented exercise.

Sample Size Calculation and Randomization

According to Giné-Garriga et al [24], the difference in Activities-Specific Balance Confidence (ABC) scale scores between the functional circuit training group and the routine daily activity control group was 6.65 points [24]. Based on this score difference, assuming a type 1 error of 0.05, statistical power of 80%, and an anticipated dropout rate of 20%, 26 (13 per group) participants were required. Therefore, 30 participants were recruited, 15 of whom were assigned to each group. The research team recruited the participants for the trial. A person

XSL•FO
RenderX

who was not involved in the trial carried out randomization. After completing the baseline surveys, the participants were randomly assigned to the intervention (AT+EX group) or control group (AT-only group) using block randomization and randomly selected block sizes of 2, 4, or 6. Blinded assessors evaluated fall-related parameters, body composition, and the data of mean daily step counts from activity trackers.

Intervention

All participants were provided with a wearable activity tracker (HL5, Seven Elec) and its corresponding application. The participants received an individual information session on how to use, pair, and charge the activity tracker during their first visit and were advised to wear the tracker on their wrists throughout the 12-week study period. The activity tracker data were paired and synchronized with the Health & You app on a mobile device. The daily step goal was set at 10,000 steps, and adjustments were made only if necessary, based on the participant's physical ability and current activity levels. The step goals remained consistent throughout the 12 weeks. Participants could track their daily steps via the activity tracker and mobile phone app and were required to synchronize their device with the app at the end of each day. The researcher could view the participants' activity tracker data on the admin page.

The AT-only group received education on the use of an activity tracker and educational materials for general fall prevention exercises. This group focused on self-monitoring using a wearable device to increase daily physical activity and did not receive any exercise interventions or feedback. Participants in the AT-only group visited the clinic at baseline and after 12 weeks.

The participants randomized to the AT+EX group received an activity tracker and participated in a home-based fall exercise intervention program. The exercise intervention included 4 in-person sessions (weeks 1, 3, 6, and 9) conducted at either a hospital or a senior welfare center. Each session was conducted by a certified exercise trainer and at least 1 assistant, with 1 - 6 participants. The exercise protocol was specifically designed to prevent falls in older adults, including balance, resistance, aerobic, and multitask training, which are effective in preventing falls [9,10,13]. Each 1-hour exercise session consisted of 5 minutes of stretching, 15 minutes of balance training, 15 minutes of strength training, 15 minutes of gait and aerobic exercise, and 10 minutes of multitask training. Over the 12-week intervention period, the exercise protocol consisted of 3 phases (first, second, and third stages), with a gradual increase in intensity between each phase. The detailed exercise program for each level is described in Multimedia Appendix 1. All participants in the AT+EX group were provided exercise equipment (step boxes, TheraBands, and yoga mats). To support consistent exercise at home, the participants received exercise videos that followed the same protocol as the in-person sessions. They were asked to exercise at least thrice a week to meet the physical activity recommendations for older adults [13]. Throughout the 12-week study period, the participants received 6 physical activity counseling phone calls (weeks 2, 4, 5, 7, 10, and 11). Phone calls were conducted by a researcher (YK) with a master's degree in exercise physiology, who followed the same protocol for all participants. Each call included feedback on physical activity levels derived from their activity trackers, aiming to motivate participants to initiate and sustain their physical activity and exercise routines. Additionally, the participants were asked concerning their recent health status, barriers to participation in the study, and any issues with using their wearable devices.

Fall-Related Parameters

The primary outcomes were fall-related parameters, including fear of falling, depression, and cognitive function, all evaluated using the Korean version and measured at baseline and 12 weeks. Fear of falling was measured using the ABC scale and Falls Efficacy Scale-International (FES-I). The ABC scale was used to assess the individuals' confidence in performing various activities without losing balance or falling [25]. The questionnaire included 16 items, with scores ranging from 0%=no confidence at all to 100%=completely confident. The overall score was calculated by averaging the scores for all 16 items. The FES-I measures self-efficacy and fear of using 16 items, each rated on a scale of 1=not at all concerned to 4=very concerned. The total score ranged 16 - 64 points, with higher scores indicating a greater fear of falling [26]. Depression, which is closely related to the fear of falling among older adults, was assessed using the Short Geriatric Depression Scale (SGDS), a reliable and valid self-administered questionnaire designed to screen for depression in older adults [27]. The SGDS comprises 15 yes-or-no questions, with higher scores indicating a more depressed mood. The Montreal Cognitive Assessment, a tool for screening mild cognitive impairment, was used to assess cognitive function. Scores range 0 - 30 points, with scores of ≤22 points indicating mild cognitive impairment [28].

Physical Function

Physical function was assessed at baseline and 12 weeks by the same assessor using the same protocol. Grip strength is a key measure used to assess overall health and muscle strength, particularly in older adults [29]. This was assessed using a calibrated digital dynamometer (Takei 5401-C; Takei Scientific Instruments; in kilograms). Individuals were asked to squeeze the dynamometer with maximum effort, performing 3 alternating measurements on each side. The highest values obtained from these trials were used in this study. According to the Asian Working Group for Sarcopenia, low grip strength is defined as <28 kg for male and <18 kg for female participants [30]. The Short Physical Performance Battery (SPPB) is used to assess lower-extremity function and physical performance in older adults [31]. The SPPB includes 3 components: balance tests (side-by-side, semitandem, and tandem stances), a gait speed test (4-m distance), and a 5-time sit-to-stand test. Each component was scored 0 - 4 points, with a total score ranging 0 - 12 points. Higher scores indicate better performance. A change of ≥1 point in the SPPB score was considered a clinically important difference [32]. The Timed Up and Go (TUG) test was used to assess balance and mobility. Each individual was asked to rise from a chair and walk 3 m, turn around, walk back to the chair, and sit down. The time taken to complete the task was measured in seconds. A shorter time indicated better mobility. Scores of ≥13.5 seconds indicate a high risk of falling

XSL•FO
RenderX

[33]. The 2-Minute Step Test (2MST) was used to assess aerobic capacity and endurance as an alternative aerobic endurance test in environments with limited space [34]. The participants were instructed to march in place for 2 minutes, lifting their knees midway between the patella and the iliac crest as many times as possible. Performance was measured by counting the number of right-sided steps completed within the 2-minute period. Higher step counts indicated better aerobic endurance.

Body Composition

Anthropometric measurements were obtained from the participants at baseline and at 12 weeks while wearing light clothes and no shoes, except for height, which was measured only at baseline. Height measurements were recorded to the nearest 0.1 cm using a stadiometer, while weight measurements were recorded to the nearest 0.1 kg with a digital scale. BMI was calculated by dividing weight by height squared (kg/m^2). Body composition was assessed using a dual-energy x-ray absorptiometry scanner (Lunar Prodigy Advance; GE Healthcare Technologies). The total fat, total lean, and appendicular muscle masses were determined. Appendicular muscle mass was calculated as the sum of the lean mass in both arms and legs divided by the square of height (kg/m^2).

Demographics

Basic demographic factors were assessed at baseline and included age, sex, marital status (single, married, widowed, or divorced), education level (college or higher, high school, middle school, or elementary school), occupation (working or not working), number of falls in the previous year, and the presence of comorbidities (eg, hypertension, diabetes, cancer, and other chronic diseases, such as cardiovascular and musculoskeletal disorders).

Mean Step Counts

The mean daily step count was assessed using data from activity trackers and was calculated based on the number of steps recorded over a specific week. The baseline step count was defined as the mean number of daily steps in the first week (days 1 - 7) after the initial visit. Subsequent steps were recorded as the daily average steps between each exercise session (weeks 1, 3, 6, and 9), and the final average steps were recorded from the fourth exercise session to the final visit (week 12). For the AT-only group, average daily step counts were recorded during the same period as those for the AT+EX group, specifically at weeks 1, 3, 6, 9, and 12. Participants with <50% wear days during these periods were excluded from the analysis. Additionally, days with <500 steps were considered nonwear days and excluded from the step count calculations [35].

Statistical Analysis

Continuous variables are presented as means and SDs, while categorical variables are presented as frequencies and percentages. The baseline characteristics of the AT+EX and AT-only groups were examined using independent 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables. Outcome data, including all available data, were analyzed using the intention-to-treat approach. When data were missing due to dropouts of participants, we used the last observation carried forward method to impute missing values. Initially, independent 2-tailed *t* tests or Mann-Whitney *U* test were used to determine whether there were any significant baseline differences between the 2 groups in fall-related outcomes, physical function, and body composition. Subsequently, to compare changes from baseline to 12 weeks (Δ values) between the groups, independent 2-tailed *t* tests or Mann-Whitney *U* tests were conducted. Paired 2-tailed *t* tests or Wilcoxon signed rank tests were used to evaluate within-group changes from baseline to 12 weeks. The change in average daily step count over the 12-week study period was analyzed using a mixed-effects model for repeated measures to account for within-participant correlations. The change in the average daily step count was calculated by subtracting the mean step count at baseline from that at each subsequent visit (weeks 3, 6, 9, and 12). Subgroup analyses were conducted for individuals with SPPB scores ≤9 (low physical performance). All statistical analyses were conducted using SPSS statistical software (version 26.0; IBM Corp) and R software (version 4.2.0; R Foundation for Statistical Computing), with statistical significance set at $P<.05$.

Ethical Considerations

This study was approved by the institutional review board of Hallym University Sacred Heart Hospital (HALLYM 2022-10-020) and conducted in accordance with the tenets of the Declaration of Helsinki. The study was registered with the Clinical Research Information Service on February 23, 2023 (KCT0008230). All participants provided written informed consent after receiving a comprehensive explanation of the study objectives, procedures, and their right to withdraw at any time. To ensure confidentiality, all data were anonymized and restricted to authorized researchers. Participants received a transportation allowance of US $34.97 as compensation.

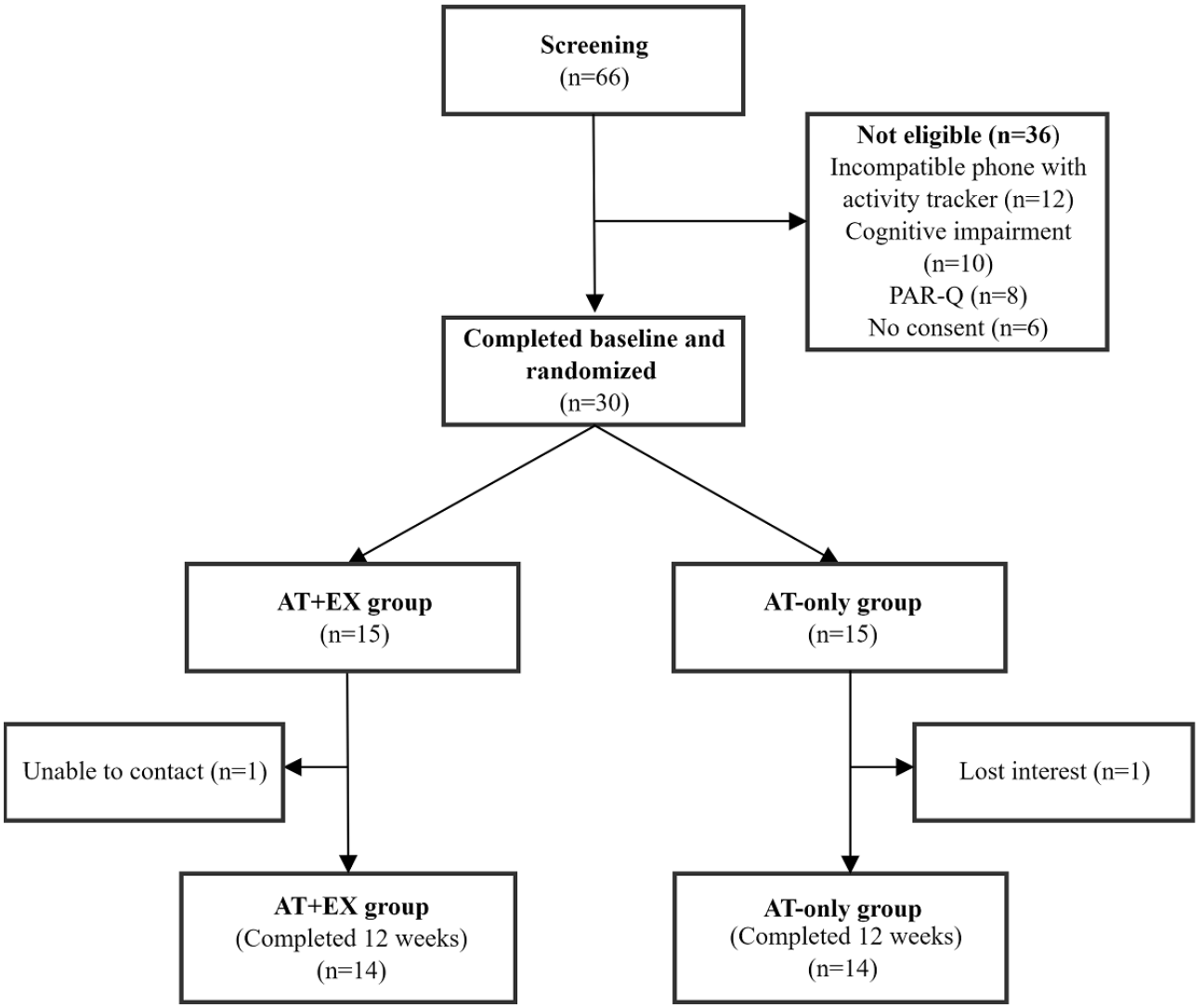
Results

Baseline Characteristics

Recruitment occurred from February to August 2023 until the target sample size was achieved. Overall, 30 participants were enrolled and randomly assigned to either the AT+EX or the AT-only group (Figure 1). The baseline characteristics of the study participants are presented in Table 1. The mean age of the participants was 74.0 (SD 6.4) years, and 77% (23/30) were female participants. A significant difference was observed in body weight, with participants in the AT+EX group (mean 55.1, SD 8.6 kg) having a significantly lower body weight compared to those in the AT-only group (mean 62.0, SD 9.1 kg; $P=.04$). No significant differences were observed in BMI, marital status, educational level, occupation, number of falls during the previous year, or comorbidities between the groups.

XSL•FO
RenderX

Figure 1. Participant flowchart. AT+EX group: activity tracker and exercise group; AT-only group: activity tracker alone group; PAR-Q: Physical Activity Readiness Questionnaire.



XSL•FO
RenderX

Table . Baseline characteristics of the participants.

	Total (n=30)	AT+EX group[a] (n=15)	AT-only group[b] (n=15)	*P* value[c]
Age (years), mean (SD)	74.0 (6.4)	73.5 (5.3)	74.6 (7.4)	.64
Sex, n (%)				>.99
Male	7 (23)	3 (20)	4 (27)	
Female	23 (77)	12 (80)	11 (73)	
Anthropometrics, mean (SD)				
Height (cm)	156.4 (9.0)	155.1 (8.7)	157.7 (9.4)	.45
Weight (kg)	58.6 (9.4)	55.1 (8.6)	62.0 (9.1)	.04
BMI (kg/m^2)	23.9 (3.0)	22.8 (2.7)	24.9 (3.1)	.06
Marital status, n (%)				.68
Single	1 (3)	0 (0)	1 (7)	
Married	24 (80)	12 (80)	12 (80)	
Widowed	3 (10)	2 (13)	1 (7)	
Divorced	2 (7)	1 (7)	1 (6.)	
Education level, n (%)				.15
Less than elementary school	8 (27)	3 (20)	5 (33)	
Middle school	7 (23)	2 (13)	5 (33)	
High school	8 (27)	4 (27)	4 (27)	
College and higher	7 (23)	6 (40)	1 (7)	
Occupation, n (%)				.65
Working	6 (20)	4 (27)	2 (13)	
Not working	24 (80)	11 (73)	13 (87)	
Number of falls during the previous year, mean (SD)	2.4 (3.5)	2.9 (4.8)	1.9 (1.5)	.25
Comorbidity, n (%)				.83
0	6 (20)	4 (27)	2 (13)	
1	9 (30)	4 (27)	5 (33)	
≥2	15 (50)	7 (47)	8 (53)	

[a]AT+EX group: activity tracker and exercise group.

[b]AT-only group: activity tracker alone group.

[c]*P* values represent the differences between the AT+EX group and the AT-only group. Independent 2-tailed *t* tests were used for continuous variables and chi-square tests for categorical variables.

Fall-Related Outcomes

The intention-to-treat analyses of the primary outcomes, physical function, and body composition are presented in Table 2. Changes in fear of falling, assessed using the ABC and FES-I scores, showed significant improvement in the AT+EX group (*P*=.002 and *P*=.01, respectively). The AT-only group also showed improvements in the FES-I score (*P*=.01); the increase in the ABC score was only marginally significant (*P*=.07). Depression levels, assessed using the SGDS, significantly improved in both the AT+EX (*P*=.006) and AT-only groups (*P*=.02). Cognitive function, as measured by the Montreal Cognitive Assessment, showed no significant change in both the AT+EX (*P*=.10) and AT-only groups (*P*=.49). No significant differences were detected between the groups in any of the fall-related outcomes.

XSL•FO
RenderX

Table . Intention-to-treat analysis of changes in fall-related parameters, physical function, and body composition from baseline to 12 weeks.

Measures	AT+EX group[a] (n=15)				AT-only group[b] (n=15)				AT+EX vs AT-only	
	Baseline, mean (SD)	12 weeks, mean (SD)	Δ (12 weeks to baseline), mean (SD)	*P* value[c]	Baseline, mean (SD)	12 weeks, mean (SD)	Δ (12 weeks to baseline), mean (SD)	*P* value[c]	*P* value for baseline[d]	*P* value for 12 weeks to baseline[e]
Fall-related parameters										
ABC[f] (score)	56.5 (17.8)	72.0 (13.6)	15.5 (15.7)	.002	57.9 (18.2)	68.0 (21.4)	10.2 (20.2)	.07	>.99	.43
FES-I[g] (score)	29.5 (8.2)	24.4 (5.3)	–5.1 (6.1)	.01	30.4 (10.3)	24.9 (8.1)	–5.5 (8.0)	.01	.92	.90
SGDS[h] (score)	4.5 (3.6)	2.5 (2.6)	–2.0 (2.3)	.006	5.4 (4.5)	4.0 (4.4)	–1.4 (2.1)	.02	.54	.46
MoCA[i] (score)	23.8 (2.8)	24.7 (3.0)	0.9 (2.1)	.10	23.9 (3.2)	24.4 (2.7)	0.5 (2.4)	.49	.90	.59
Physical function										
Grip strength (kg)	22.0 (6.4)	23.6 (6.1)	1.5 (4.0)	.14	22.5 (5.2)	24.3 (5.1)	1.7 (3.7)	.12	.82	.90
SPPB[j] (score)	10.3 (1.7)	11.5 (1.1)	1.1 (1.1)	.004	9.9 (2.2)	10.7 (1.5)	0.8 (1.7)	.09	.52	.23
Timed Up and Go (seconds)	9.7 (2.6)	8.3 (2.2)	–1.4 (2.1)	.008	9.9 (2.2)	8.7 (1.3)	–1.3 (1.3)	.002	.52	.81
Gait speed (m/s)	0.81 (0.23)	0.91 (0.15)	0.10 (0.23)	.11	0.79 (0.21)	0.85 (0.13)	0.06 (0.14)	.12	.84	.56
2-Minute Step Test (n)	156.0 (35.9)	182.7 (34.3)	26.7 (27.4)	.001	131.7 (41.2)	142.1 (36.3)	10.5 (32.3)	.10	.11	.10
Body composition										
BMI (kg/m²)	22.8 (2.7)	22.6 (2.8)	–0.3 (0.6)	.08	24.9 (3.0)	24.9 (3.0)	0.0 (0.5)	.97	.06	.15
Skeletal muscle index (kg/m²)	6.6 (0.8)	6.5 (0.8)	–0.1 (0.2)	.26	7.0 (0.9)	6.8 (0.9)	–0.2 (0.3)	.07	.16	.36
Total fat mass (kg)	17.2 (4.6)	16.8 (4.7)	–0.4 (1.0)	.09	20.8 (4.8)	20.5 (4.6)	–0.3 (0.9)	.97	.048	.17
Total lean mass (kg)	36.1 (6.0)	35.7 (6.3)	–0.4 (0.8)	.04	39.5 (6.9)	39.0 (6.3)	–0.5 (1.3)	.19	.17	.96

[a]AT+EX group: activity tracker and exercise group.

[b]AT-only group: activity tracker alone group.

[c]*P* values represent the differences between baseline and 12 weeks at each group using the paired 2-tailed *t* test or the Wilcoxon signed rank test.

[d]*P* values represent the differences between the AT+EX group and the AT-only group at baseline using the independent 2-tailed *t* test or the Mann-Whitney test.

[e]*P* values represent the comparison of changes between 2 groups using the independent 2-tailed *t* test or the Mann-Whitney test.

[f]ABC: Activities-Specific Balance Confidence.

[g]FES-I: Falls Efficacy Scale-International.

[h]SGDS: Short Geriatric Depression Scale.

[i]MoCA: Montreal Cognitive Assessment.

[j]SPPB: Short Physical Performance Battery.

XSL•FO
RenderX

Physical Function

Regarding physical function, grip strength showed no significant change in both groups (AT+EX: 1.5 kg; *P*=.14 and AT-only: 1.7 kg; *P*=.12). The SPPB total score showed significant improvement in the AT+EX group (1.1 points; *P*=.004) but was not significant in the AT-only group (0.8 points; *P*=.09). The AT+EX group also demonstrated significant improvements in both the TUG test (1.4 seconds; *P*=.008) and the 2MST (26.7 steps;, *P*=.001), while the AT-only group showed a significant improvement only in the TUG test (1.3 seconds; *P*=.002). No significant differences were observed between the groups in terms of changes in any physical functional variables.

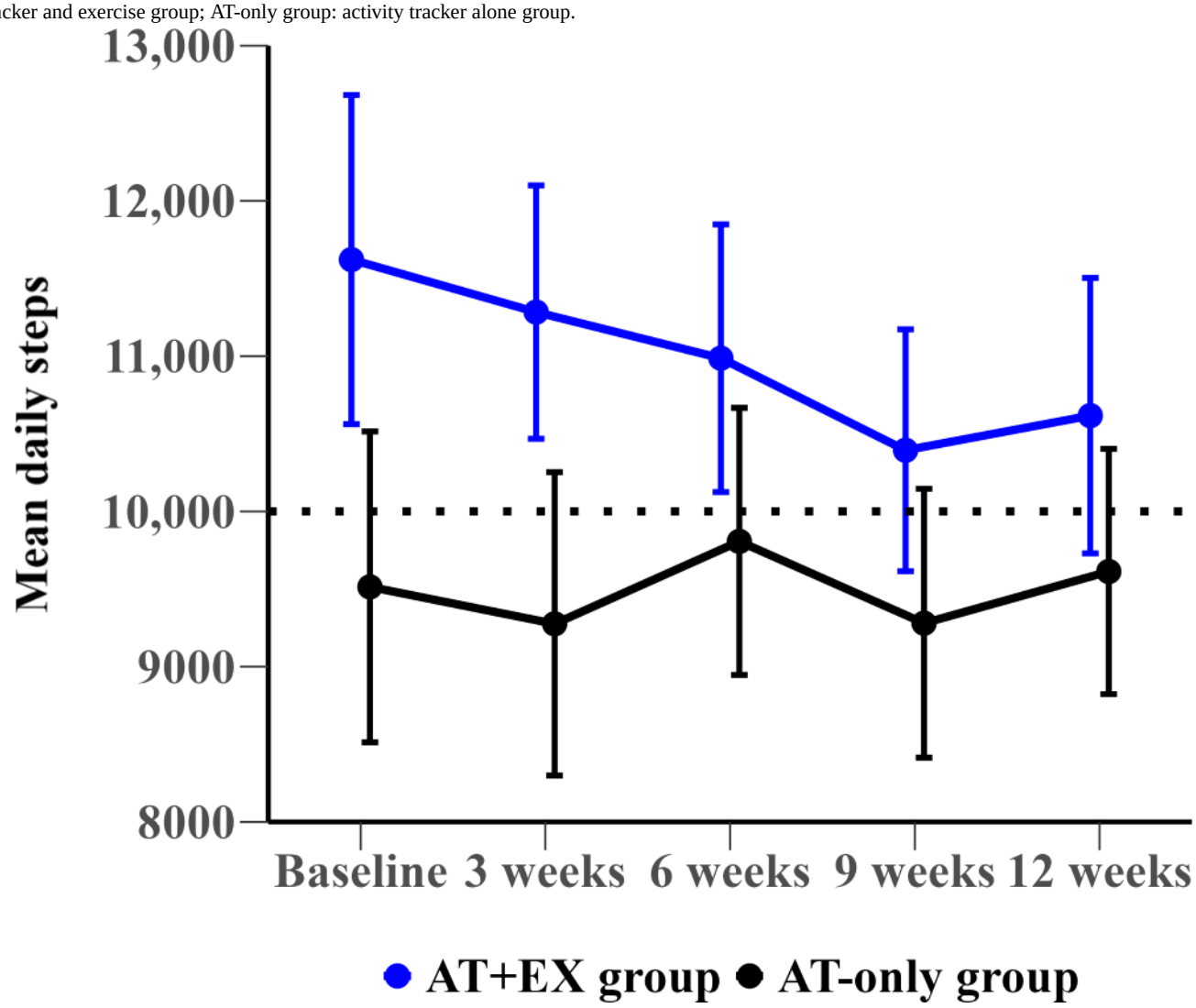
Body Composition

Regarding body composition, the AT+EX group showed reductions, including a significant decrease in total lean mass by 0.4 kg (*P*=.04). In contrast, the AT-only group showed nonsignificant decreases in total lean mass and skeletal muscle index (*P*=.19 and *P*=.07, respectively). No significant differences between the groups were observed in the changes of any body composition variables.

Changes in Mean Daily Steps

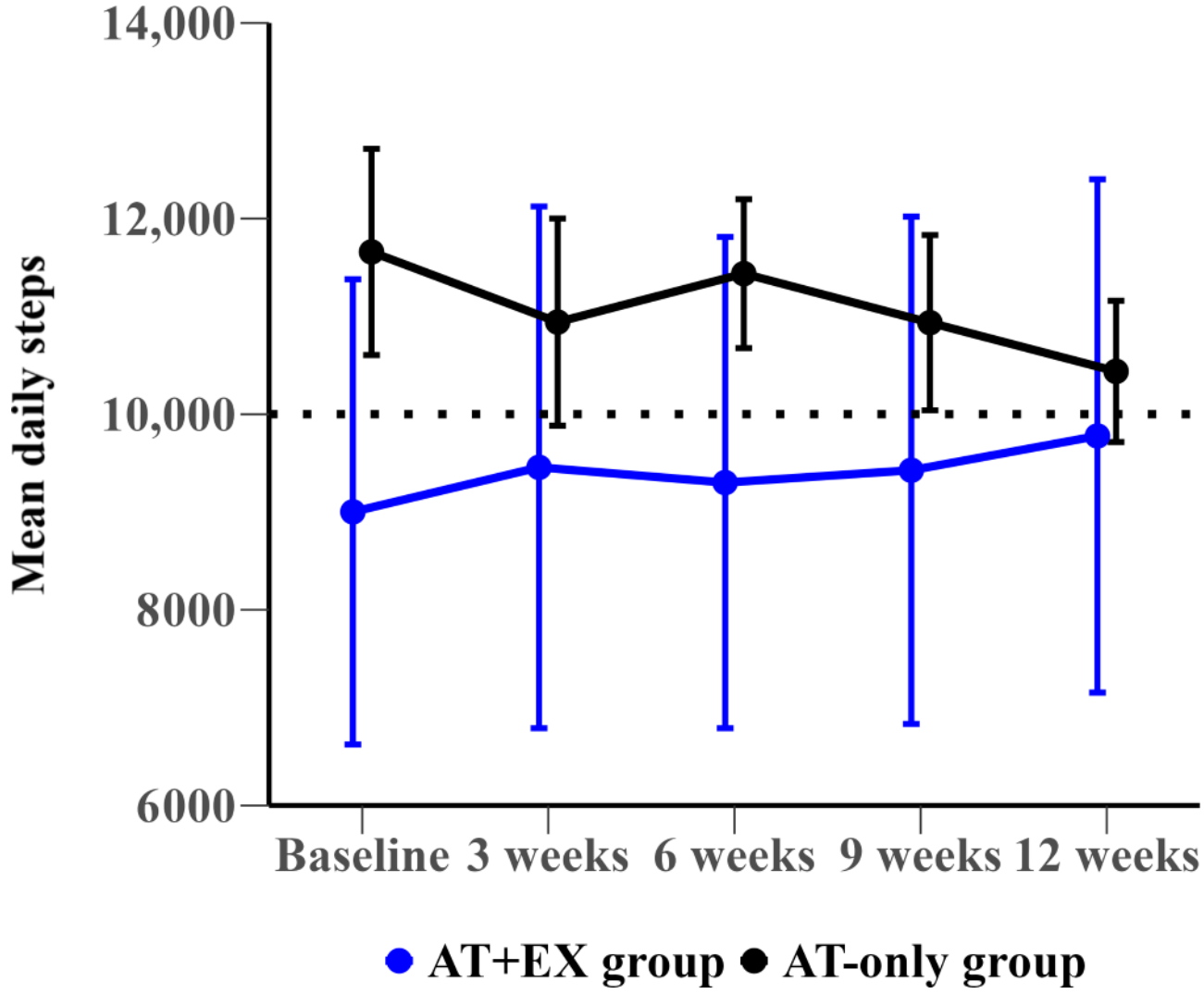
Figures 2 and 3 present the mean daily steps, while Table 3 details changes in mean daily steps over 12 weeks for both the overall group and the subgroup with low physical performance. At baseline, the AT+EX group averaged 11,622 daily steps compared to 9514 steps in the AT-only group, showing a mean difference of 2108 steps (95% CI –691 to 4906; *P*=.14). Throughout the 12-week study period, both groups maintained high levels of daily steps, and differences between them were not statistically significant. Over the study period, the AT+EX group decreased by 1005 (from 11,622 to 10,617) steps, whereas the AT-only group increased by 98 (from 9514 to 9613) steps.

Figure 2. Mean daily steps with SE of the mean at baseline and over specific weeks (1, 3, 6, 9, and 12) for the total group. AT+EX group: activity tracker and exercise group; AT-only group: activity tracker alone group.



XSL•FO
RenderX

Figure 3. Mean daily steps with SE of the mean at baseline and over specific weeks (1, 3, 6, 9, and 12) for the low physical performance subgroup. AT+EX group: activity tracker and exercise group; AT-only group: activity tracker alone group.



XSL•FO
RenderX

Table . Differences in changes in mean daily step count over 12 weeks between groups for the total group and low physical performance subgroup.

	AT+EX group[a], mean (SD)	AT-only group[b], mean (SD)	Mean difference (95% CI)	*P* value[c]
Total study participants, n	15	15	—[d]	—
Mean daily steps				
Baseline	11,622 (4104)	9514 (3878)	2108 (–691 to 4906)	.14
Week 3	11,284 (3161)	9276 (3785)	2007 (–461 to 4476)	.11
Week 6	10,987 (3337)	9807 (3331)	1179 (–1192 to 3550)	.32
Week 9	10,486 (3013)	9280 (3352)	1206 (–1074 to 3486)	.29
Week 12	10,617 (3437)	9613 (3060)	1004 (–1318 to 3325)	.39
Changes in the steps from baseline				
Week 3 to baseline	–338 (2221)	–238 (1563)	–100 (–1537 to 1336)	.88
Week 6 to baseline	–635 (1885)	293 (1633)	–928 (–2248 to 392)	.15
Week 9 to baseline	–1136 (2493)	–235 (2021)	–902 (–2599 to 796)	.27
Week 12 to baseline	–1005 (2860)	98 (2870)	–1104 (–3247 to 1039)	.28
Subgroup of participants with low physical performance (SPPB[e] scores ≤9), n	4	5	—	—
Mean daily steps				
Baseline	9002 (4755)	11,659 (2356)	–2657 (–7457 to 2143)	.31
Week 3	9457 (5023)	10,942 (2368)	–1485 (–6495 to 3525)	.57
Week 6	9301 (5186)	11,437 (1701)	–2137 (–6995 to 2722)	.41
Week 9	9427 (5248)	10,936 (2003)	–1509 (–6534 to 3516)	.57
Week 12	9778 (5334)	10,438 (1614)	–660 (–5610 to 4290)	.80
Changes in the steps from baseline				
Week 3 to baseline	455 (565)	–717 (1039)	1172 (–205 to 2549)	.08
Week 6 to baseline	298 (683)	–222 (1498)	520 (–1411 to 2451)	.54
Week 9 to baseline	425 (790)	–723 (1448)	1148 (–772 to 3068)	.2
Week 12 to baseline	776 (612)	–1221 (2060)	1997 (–553 to 4547)	.11

[a]AT+EX group: activity tracker and exercise group.

[b]AT-only group: activity tracker alone group.

[c]*P* value for comparisons between 2 groups calculated using a mixed model. Mean daily steps were calculated over specific weeks. Baseline steps were the average of the first week (days 1 - 7). Subsequent steps were averaged between exercise sessions (weeks 1, 3, 6, and 9), and from week 9 to the final visit.

[d]Not applicable

[e]SPPB: Short Physical Performance Battery.

In the subgroup with low physical performance (SPPB scores ≤9), the AT+EX group started with 9002 steps, lower than the AT-only group's 11,659 steps, with a mean difference of 2657 steps (95% CI –7457 to 2143; *P*=.31). By week 12, the AT+EX subgroup increased by 776 steps, while the AT-only subgroup decreased by 1221 steps. However, the difference between the groups remained nonsignificant (mean difference 1997 steps, 95% CI –553 to 4547; *P*=.11).

Discussion

Principal Findings

We investigated the effects of integrating activity trackers with a home-based multicomponent exercise intervention (AT+EX group) compared to using activity trackers alone (AT-only group) in older adults (≥60 years) with a history of falls and COVID-19. Our results showed that both groups exhibited significant improvements in fear of falling (FES-I score) and depression (SGDS) as well as enhanced physical function (TUG test). Notably, the AT+EX group showed improvements in the ABC score, SPPB score, and 2MST. However, no significant between-group differences were observed in any outcome

XSL•FO
RenderX

variables. Although there is no consensus on the cutoff value of the ABC score for predicting future falls, previous studies have suggested cutoff values of <58 - 67 [36,37]. In this study, the mean ABC score at baseline was 56.5 (SD 17.8) and 57.9 (SD 18.2) in the AT+EX group and AT-only group, respectively. After the intervention, the ABC score improved significantly in the AT+EX group (15.5 points; *P*=.002) but was not significant in the AT-only group (10.2 points; *P*=.07). An increase in the ABC score reflects a reduction in fear of falling and activity limitation as well as an improvement in the quality of life of the study participants [38]. We also showed that the SPPB score improved significantly in the AT+EX group (1.1 points; *P*=.004) but was not significant in the AT-only group (0.8 points; *P*=.09) after the intervention. Better physical performance reduces the risk of mobility-related disability [32]. Although no significant increase in daily steps was observed in either group, the participants maintained a high activity level of approximately 10,000 steps per day over the 12-week study period.

To date, few trials have evaluated the impact of activity trackers on fall-related outcomes, and the results have been inconsistent. Yamada et al [23] reported that a 6-month pedometer-based behavioral change program effectively improved fear of falling, physical activity, locomotive function, and leg muscle mass compared with the corresponding in the control group (*P*<.05). Another study by Oliveira et al [22] examined the effectiveness of a fall prevention intervention with a Fitbit (Google Inc). Although mobility goal attainment improved at 6 months compared to that in the control group, no between-group differences were noted in fall rates, fear of falling, physical activity, and mobility limitation at 6 or 12 months.

Our study showed significant improvements in fear of falling and depression in both the AT+EX and AT-only groups without significant between-group differences. However, we did not observe any improvements in cognitive function. This may be attributed to the fact that our study included participants with normal cognitive function, which may have underestimated the effect on cognitive function. Given the high prevalence of depression and cognitive impairment and the close interaction between 2 common disorders among older adults [39], it would be useful to conduct future large clinical trials to demonstrate the effectiveness of exercise in patients with these conditions.

Previous studies have compared groups using activity trackers to control groups with no intervention, whereas our study compared the AT+EX group to the AT-only group. Moreover, these studies did not include direct exercise training in their interventions. Although cognitive decline and mood disorders, such as depression and anxiety, are associated with fear of falling, research on how interventions affect these variables is lacking [40,41]. Considering the effectiveness of exercise in improving physical and cognitive functions, further research is needed to explore the impact of integrating activity trackers with home-based exercise interventions on cognitive and mood parameters as well as the fear of falling.

Comparison With Prior Work

Studies on the impact of wearable devices on physical function in older adults are limited and have shown mixed results [22,23,42-48]. A few studies using pedometers or Fitbit devices reported improvements in physical function [42,44,47,48]. For instance, 2 randomized controlled trials with pedometer-based walking programs have found significant improvements in walking speed and physical performance [47,48]. Another study using Fitbit-based interventions showed improvements in the 30-second chair stand test and physical endurance [42,44]. However, some studies using wearable devices have not shown improvements in physical function [43,45,46].

In our study, both groups showed overall improvements in physical function, although no statistically significant differences were observed between the groups. However, the AT+EX group showed significant improvements in the total SPPB score, exceeding the clinically significant threshold of a 1-point improvement [31]. Moreover, the AT+EX group showed significant improvements in mobility and endurance, as measured by the TUG test and 2MST. These findings suggest that a multicomponent exercise intervention may be more effective in enhancing physical function (ie, balance, gait, mobility, and endurance) in older adults than using activity trackers alone.

The improvement in fear of falls and physical function in both groups using activity trackers may be related to the amount of physical activity. Physical activity has been associated with a lower risk of falls and improved physical function in older adults [49]. A comprehensive analysis of systematic reviews showed that the use of activity trackers is effective in increasing physical activity levels across various age groups [50]. However, we did not observe any increase in daily step counts in either the AT+EX or AT-only group over the 12-week period. There are several potential reasons for this lack of increase. First, the participants' baseline activity level was already high, and the target step count was set at a maximum of 10,000 steps per day. Those who met this goal were encouraged to sustain their activity level rather than increase it further. Second, the reactivity effect may have influenced the results; participants may have initially increased their activity levels upon realizing that their physical activity was monitored [51].

Although no significant increases were observed in daily step counts, both groups maintained close to 10,000 steps per day throughout the 12 weeks. Notably, even the AT-only group, which did not receive additional feedback, showed a consistently high step count. This suggests that setting a specific step goal and synchronizing the activity tracker data with the app provided participants with daily visual feedback, allowing them to objectively monitor their physical activity levels and maintain their daily step counts. However, considering that this study was conducted over 12 weeks, further long-term studies are required to determine the sustained effectiveness of activity trackers without additional interventions. Moreover, a subgroup analysis revealed that participants with low physical performance (SPPB scores ≤9) exhibited increased daily step counts in the AT+EX group. This indicates that the effectiveness of activity tracker use may vary based on participants' physical performance. Further research is required to identify the subgroups that benefit the most from using activity trackers.

XSL•FO
RenderX

Regarding body composition, the AT+EX group showed an unexpected substantial decrease in lean body mass. This may be attributed to the lack of nutritional intervention in this study. Inadequate nutritional support for increased physical activity may result in reductions in both fat and muscle mass [52]. Moreover, the multicomponent exercise program focused on enhancing balance and physical function rather than increasing muscle mass. Considering the inadequate protein intake of older people in Korea [53], future research should include both physical activity and nutritional interventions including protein supplementation. Further studies are required to determine the most effective exercise regimens when combined with activity trackers.

Limitations

This study had several limitations. First, the study duration was only 12 weeks, which may not have been sufficient to observe the long-term effects of the intervention. Second, the sample size was relatively small, potentially limiting the generalizability of the results to individuals who are not motivated to increase their physical activity levels. Third, the participants were already relatively active at baseline, which could have constrained the observable benefits of increased physical activity. Fourth, the study lacked a control group with no intervention, which restricted the possibility of comparing intervention effects. Fifth, participants' nutritional intake and other lifestyle changes were not controlled.

Conclusions

Although no significant between-group differences were observed, both activity trackers with and without a home-based multicomponent exercise program effectively improved fall-related parameters (fear of falling and depression) and physical function in older adults compared to baseline. Our findings underscore the potential of digital health technologies to enhance fall prevention strategies.

Acknowledgments

This work was supported by the National Research Foundation of Korea grant funded by the Korean government (Ministry of Science and ICT; RS-2022 - 00165717).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exercise training protocol.
[DOCX File, 23 KB - mhealth_v13i1e64458_app1.docx]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File, 84 KB - mhealth_v13i1e64458_app2.pdf]

References

1. Inouye SK, Studenski S, Tinetti ME, Kuchel GA. Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept. J Am Geriatr Soc 2007 May;55(5):780-791. [doi: 10.1111/j.1532-5415.2007.01156.x] [Medline: 17493201]
2. Kannus P, Parkkari J, Koskinen S, et al. Fall-induced injuries and deaths among older adults. JAMA 1999 May 26;281(20):1895-1899. [doi: 10.1001/jama.281.20.1895] [Medline: 10349892]
3. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, et al. Sarcopenia: European consensus on definition and diagnosis: report of the European Working Group on Sarcopenia in Older People. Age Ageing 2010 Jul;39(4):412-423. [doi: 10.1093/ageing/afq034] [Medline: 20392703]
4. Kakara R, Bergen G, Burns E, Stevens M. Nonfatal and fatal falls among adults aged ≥65 years—United States, 2020-2021. MMWR Morb Mortal Wkly Rep 2023 Sep 1;72(35):938-943. [doi: 10.15585/mmwr.mm7235a1] [Medline: 37651272]
5. 2017 National survey of older Koreans. Ministry of Health and Welfare. 2018. URL: https://www.kihasa.re.kr/publish/report/view?type=others&seq=29648 [accessed 2025-04-24]
6. Tinetti ME, Kumar C. The patient who falls: "It's always a trade-off". JAMA 2010 Jan 20;303(3):258-266. [doi: 10.1001/jama.2009.2024] [Medline: 20085954]
7. Scheffer AC, Schuurmans MJ, van Dijk N, van der Hooft T, de Rooij SE. Fear of falling: measurement strategy, prevalence, risk factors and consequences among older persons. Age Ageing 2008 Jan;37(1):19-24. [doi: 10.1093/ageing/afm169] [Medline: 18194967]
8. Jadczak AD, Makwana N, Luscombe-Marsh N, Visvanathan R, Schultz TJ. Effectiveness of exercise interventions on physical function in community-dwelling frail older people: an umbrella review of systematic reviews. JBI Database Syst Rev Implement Rep 2018 Mar;16(3):752-775. [doi: 10.11124/JBISRIR-2017-003551] [Medline: 29521871]

XSL•FO
RenderX

9. Sherrington C, Tiedemann A, Fairhall N, Close JCT, Lord SR. Exercise to prevent falls in older adults: an updated meta-analysis and best practice recommendations. N S W Public Health Bull 2011 Jun;22(3-4):78-83. [doi: 10.1071/NB10056] [Medline: 21632004]
10. Cadore EL, Rodríguez-Mañas L, Sinclair A, Izquierdo M. Effects of different exercise interventions on risk of falls, gait ability, and balance in physically frail older adults: a systematic review. Rejuvenation Res 2013 Apr;16(2):105-114. [doi: 10.1089/rej.2012.1397] [Medline: 23327448]
11. Peterson MD, Rhea MR, Sen A, Gordon PM. Resistance exercise for muscular strength in older adults: a meta-analysis. Ageing Res Rev 2010 Jul;9(3):226-237. [doi: 10.1016/j.arr.2010.03.004] [Medline: 20385254]
12. Avin KG, Hanke TA, Kirk-Sanchez N, et al. Management of falls in community-dwelling older adults: clinical guidance statement from the Academy of Geriatric Physical Therapy of the American Physical Therapy Association. Phys Ther 2015 Jun;95(6):815-834. [doi: 10.2522/ptj.20140415] [Medline: 25573760]
13. Nelson ME, Rejeski WJ, Blair SN, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Circulation 2007 Aug 28;116(9):1094-1105. [doi: 10.1161/CIRCULATIONAHA.107.185650] [Medline: 17671236]
14. British Geriatrics Society. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. J Am Geriatr Soc 2011 Jan;59(1):148-157. [doi: 10.1111/j.1532-5415.2010.03234.x]
15. Mayoux-Benhamou MA, Roux C, Perraud A, Fermanian J, Rahali-Kachlouf H, Revel M. Predictors of compliance with a home-based exercise program added to usual medical care in preventing postmenopausal osteoporosis: an 18-month prospective study. Osteoporos Int 2005 Mar;16(3):325-331. [doi: 10.1007/s00198-004-1697-z] [Medline: 15726237]
16. Gough C, Barr C, Lewis LK, Hutchinson C, Maeder A, George S. Older adults' community participation, physical activity, and social interactions during and following COVID-19 restrictions in Australia: a mixed methods approach. BMC Public Health 2023 Jan 25;23(1):172. [doi: 10.1186/s12889-023-15093-0] [Medline: 36698158]
17. Freene N, Waddington G, Chesworth W, Davey R, Cochrane T. Community group exercise versus physiotherapist-led home-based physical activity program: barriers, enablers and preferences in middle-aged adults. Physiother Theory Pract 2014 Feb;30(2):85-93. [doi: 10.3109/09593985.2013.816894] [Medline: 24405399]
18. Bethancourt HJ, Rosenberg DE, Beatty T, Arterburn DE. Barriers to and facilitators of physical activity program use among older adults. Clin Med Res 2014 Sep;12(1-2):10-20. [doi: 10.3121/cmr.2013.1171] [Medline: 24415748]
19. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192. [doi: 10.2196/jmir.3469] [Medline: 25131661]
20. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
21. Wu S, Li G, Du L, Chen S, Zhang X, He Q. The effectiveness of wearable activity trackers for increasing physical activity and reducing sedentary time in older adults: a systematic review and meta-analysis. Digit Health 2023;9:20552076231176705. [doi: 10.1177/20552076231176705] [Medline: 37252261]
22. Oliveira JS, Sherrington C, Paul SS, et al. A combined physical activity and fall prevention intervention improved mobility-related goal attainment but not physical activity in older adults: a randomised trial. J Physiother 2019 Jan;65(1):16-22. [doi: 10.1016/j.jphys.2018.11.005] [Medline: 30581138]
23. Yamada M, Mori S, Nishiguchi S, et al. Pedometer-based behavioral change program can improve dependency in sedentary older adults: a randomized controlled trial. J Frailty Aging 2012;1(1):39-44. [doi: 10.14283/jfa.2012.7] [Medline: 27092936]
24. Giné-Garriga M, Guerra M, Unnithan VB. The effect of functional circuit training on self-reported fear of falling and health status in a group of physically frail older individuals: a randomized controlled trial. Aging Clin Exp Res 2013 Jun;25(3):329-336. [doi: 10.1007/s40520-013-0048-3] [Medline: 23740589]
25. Powell LE, Myers AM. The Activities-Specific Balance Confidence (ABC) Scale. J Gerontol A Biol Sci Med Sci 1995 Jan;50A(1):M28-M34. [doi: 10.1093/gerona/50a.1.m28] [Medline: 7814786]
26. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). Age Ageing 2005 Nov;34(6):614-619. [doi: 10.1093/ageing/afi196] [Medline: 16267188]
27. Fountoulakis KN, Tsolaki M, Iacovides A, et al. The validation of the short form of the Geriatric Depression Scale (GDS) in Greece. Aging (Milano) 1999 Dec;11(6):367-372. [doi: 10.1007/BF03339814] [Medline: 10738851]
28. Lee JY, Cho SJ, et al. Brief screening for mild cognitive impairment in elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. J Geriatr Psychiatry Neurol 2008 Jun;21(2):104-110. [doi: 10.1177/0891988708316855] [Medline: 18474719]
29. Bohannon RW. Grip strength: an indispensable biomarker for older adults. Clin Interv Aging 2019;14:1681-1691. [doi: 10.2147/CIA.S194543] [Medline: 31631989]
30. Chen LK, Woo J, Assantachai P, et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc 2020 Mar;21(3):300-307. [doi: 10.1016/j.jamda.2019.12.012] [Medline: 32033882]

XSL•FO
RenderX

31. de Fátima Ribeiro Silva C, Ohara DG, Matos AP, Pinto A, Pegorari MS. Short Physical Performance Battery as a measure of physical performance and mortality predictor in older adults: a comprehensive literature review. Int J Environ Res Public Health 2021 Oct 10;18(20):10612. [doi: 10.3390/ijerph182010612] [Medline: 34682359]
32. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the Short Physical Performance Battery. J Gerontol A Biol Sci Med Sci 2000 Apr;55(4):M221-M231. [doi: 10.1093/gerona/55.4.m221] [Medline: 10811152]
33. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. Phys Ther 2000 Sep;80(9):896-903. [Medline: 10960937]
34. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist 2013 Apr;53(2):255-267. [doi: 10.1093/geront/gns071] [Medline: 22613940]
35. Tudor-Locke C, Giles-Corti B, Knuiman M, McCormack G. Tracking of pedometer-determined physical activity in adults who relocate: results from RESIDE. Int J Behav Nutr Phys Act 2008 Aug 7;5(39):18687137. [doi: 10.1186/1479-5868-5-39] [Medline: 18687137]
36. Moiz JA, Bansal V, Noohu MM, et al. Activities-specific Balance Confidence scale for predicting future falls in Indian older adults. Clin Interv Aging 2017;12(645-51):645-651. [doi: 10.2147/CIA.S133523] [Medline: 28435236]
37. Lajoie Y, Gallagher SP. Predicting falls within the elderly community: comparison of postural sway, reaction time, the Berg balance scale and the Activities-specific Balance Confidence (ABC) scale for comparing fallers and non-fallers. Arch Gerontol Geriatr 2004;38(1):11-26. [doi: 10.1016/s0167-4943(03)00082-7] [Medline: 14599700]
38. Schoene D, Heller C, Aung YN, Sieber CC, Kemmler W, Freiberger E. A systematic review on the influence of fear of falling on quality of life in older people: is there a role for falls? Clin Interv Aging 2019;14(701-19):701-719. [doi: 10.2147/CIA.S197857] [Medline: 31190764]
39. Han S, Gao Y, Gan D. The combined associations of depression and cognitive impairment with functional disability and mortality in older adults: a population-based study from the NHANES 2011-2014. Front Aging Neurosci 2023;15(1121190):1121190. [doi: 10.3389/fnagi.2023.1121190] [Medline: 37213544]
40. Noh HM, Roh YK, Song HJ, Park YS. Severe fear of falling is associated with cognitive decline in older adults: a 3-year prospective study. J Am Med Dir Assoc 2019 Dec;20(12):1540-1547. [doi: 10.1016/j.jamda.2019.06.008] [Medline: 31351857]
41. Painter JA, Allison L, Dhingra P, Daughtery J, Cogdill K, Trujillo LG. Fear of falling and its relationship with anxiety, depression, and activity engagement among community-dwelling older adults. Am J Occup Ther 2012;66(2):169-176. [doi: 10.5014/ajot.2012.002535] [Medline: 22394526]
42. Batsis JA, Petersen CL, Clark MM, et al. Feasibility and acceptability of a technology-based, rural weight management intervention in older adults with obesity. BMC Geriatr 2021 Jan 12;21(1):44. [doi: 10.1186/s12877-020-01978-x] [Medline: 33435877]
43. Brickwood KJ, Ahuja KDK, Watson G, O'Brien JA, Williams AD. Effects of activity tracker use with health professional support or telephone counseling on maintenance of physical activity and health outcomes in older adults: randomized controlled trial. JMIR Mhealth Uhealth 2021 Jan 5;9(1):e18686. [doi: 10.2196/18686] [Medline: 33399541]
44. Liu JYW, Kwan RYC, Yin YH, Lee PH, Siu JYM, Bai X. Enhancing the physical activity levels of frail older adults with a wearable activity tracker-based exercise intervention: a pilot cluster randomized controlled trial. Int J Environ Res Public Health 2021 Sep 30;18(19):34639644. [doi: 10.3390/ijerph181910344] [Medline: 34639644]
45. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Mar 6;5(3):e28. [doi: 10.2196/mhealth.6967] [Medline: 28264796]
46. McDermott MM, Spring B, Berger JS, et al. Effect of a home-based exercise intervention of wearable technology and telephone coaching on walking performance in peripheral artery disease: the HONOR randomized clinical trial. JAMA 2018 Apr 24;319(16):1665-1676. [doi: 10.1001/jama.2018.3275] [Medline: 29710165]
47. Snyder A, Colvin B, Gammack JK. Pedometer use increases daily steps and functional status in older adults. J Am Med Dir Assoc 2011 Oct;12(8):590-594. [doi: 10.1016/j.jamda.2010.06.007] [Medline: 21450191]
48. Talbot LA, Gaines JM, Huynh TN, Metter EJ. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. J Am Geriatr Soc 2003 Mar;51(3):387-392. [doi: 10.1046/j.1532-5415.2003.51113.x] [Medline: 12588583]
49. Dipietro L, Campbell WW, Buchner DM, et al. Physical activity, injurious falls, and physical function in aging: an umbrella review. Med Sci Sports Exerc 2019 Jun;51(6):1303-1313. [doi: 10.1249/MSS.0000000000001942] [Medline: 31095087]
50. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
51. Tudor-Locke C, Bassett DR, Shipe MF, McClain JJ. Pedometry methods for assessing free-living adults. J Phys Act Health 2011 Mar;8(3):445-453. [doi: 10.1123/jpah.8.3.445] [Medline: 21487145]

XSL•FO
RenderX

52. De Stefani FC, Pietraroia PS, Fernandes-Silva MM, Faria-Neto J, Baena CP. Observational evidence for unintentional weight loss in all-cause mortality and major cardiovascular events: a systematic review and meta-analysis. Sci Rep 2018 Oct 18;8(1):15447. [doi: 10.1038/s41598-018-33563-z] [Medline: 30337578]
53. Kim KW, Park HA, Cho YG, Bong AR. Protein intake by Korean adults through meals. KJHP 2021 Jun;21(2):63-72. [doi: 10.15384/kjhp.2021.21.2.63]

Abbreviations

2MST: 2-Minute Step Test
ABC: Activities-Specific Balance Confidence
AT+EX group: activity tracker and exercise group
AT-only group: activity tracker alone group
CONSORT: Consolidated Standards of Reporting Trials
FES-I: Falls Efficacy Scale-International
SGDS: Short Geriatric Depression Scale
SPPB: Short Physical Performance Battery
TUG: Timed Up and Go

Edited by L Buis; submitted 17.07.24; peer-reviewed by M Shi, M Reshma; revised version received 18.02.25; accepted 01.04.25; published 08.05.25.

Please cite as:
Kim Y, Park KH, Noh HM
Effects of Integrating Wearable Activity Trackers With a Home-Based Multicomponent Exercise Intervention on Fall-Related Parameters and Physical Function in Older Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e64458
URL: https://mhealth.jmir.org/2025/1/e64458
doi:10.2196/64458

© Yejin Kim, Kyung Hee Park, Hye-Mi Noh. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Consumer Wearable Usage to Collect Health Data Among Adults Living in Germany: Nationwide Observational Survey Study

Kristin Manz, PhD; Susanne Krug, PhD; Charlotte Kühnelt, MSc; Johannes Lemcke, Diploma in Sociology; Ilter Öztürk, MA; Julika Loss, PhD

Department 2: Epidemiology and Health Monitoring, Robert Koch Institute, Nordufer 20, Berlin, Germany

Corresponding Author:
Kristin Manz, PhD
Department 2: Epidemiology and Health Monitoring, Robert Koch Institute, Nordufer 20, Berlin, Germany

Abstract

Background: The usage of consumer wearables (CWs; eg, fitness trackers and smartwatches) in the population has increased enormously within the last decade. This has resulted in a large amount of digital person-generated health data that could be used to answer vital research questions. However, little is currently known about the usage of CWs to collect health data from the population living in Germany.

Objective: This study aimed to describe the ownership of consumer wearables and their usage for the collection of health data from the adult population living in Germany, as well as the motives for the collection of health data and the average wear times. In addition, this study also aimed to investigate sociodemographic and health- and behavior-related differences between the group of CW users and the group of nonusers.

Methods: We used data from the nationally representative survey "German Health Update," which was conducted through telephone interviews in 2021 and 2022. The final sample comprised 4464 adults aged 18 years and older. We derived weighted prevalences for the usage of CWs, as well as adjusted odds ratios for the ownership and the usage of CWs and their association with sociodemographic and health- and behavior-related variables.

Results: Of the adult population, 19.3% (843/4459) owned a CW, of whom 77.8% (650/842) used their CW to collect health data (which corresponds to 650/4458, 15.0% of the adult population). Older people, people with a low income, and people with a lower level of physical activity (PA) were less likely to own a CW and were less likely to use it for the collection of health data. Of the CW users who collected health data, 47.2% (321/650) wore their CW during nocturnal sleep. The most frequently named motives for the collection of health data with a CW were "to observe my PA" (544/647, 85.0%), "for fun" (508/644, 79.0%), and "for support during exercising" (423/647, 66.3%). Women chose the motive "to observe my PA" and "to increase my PA" more often than men, whereas men chose the motive "to observe health issues" more often than women.

Conclusions: Adults living in Germany owning a CW are younger, have a higher income, and are more physically active than individuals who do not use a CW. This means that the population groups that would be in particular need of health care are not sufficiently represented in these health datasets. Researchers should consider the selectivity of CW users when planning to use CW health data to answer research questions.

(JMIR Mhealth Uhealth 2025;13:e59199) doi:10.2196/59199

KEYWORDS

health survey; fitness tracker; smartwatch; selectivity; physical activity; sleep

Introduction

Background

Individuals can nowadays collect certain health-related data with commercial wearable devices (also called "consumer wearables" [CWs]) like fitness trackers or smartwatches using integrated sensor technology (eg, accelerometers, pulse oximeters, and geolocation). Therefore, CWs can track vital parameters and behaviors, such as heart rate, physical activity (PA), or sleep [1].

The usage of CWs in the population has increased enormously within the last decade. In 2022, around 500 million of these devices were sold worldwide, which means that the sales increased by the factor 16 compared with 2014 [2]. In Germany, 7.2 million CWs were sold in 2022 [3]. However, for Germany, data from adults on the ownership of CWs and their actual use to collect health data are scarce.

CW Data as Part of Public Health Surveillance

The prevalent usage of CWs in the population results in a large amount of digital person-generated health data that could be used to answer vital research questions or to support individual

XSL•FO
RenderX

health care [4,5]. Within the public health context, these CW data could supplement the monitoring of important health behaviors, like PA and sleep, within the national public health surveillance [6,7]. Public health surveillance includes the ongoing systematic collection, analysis, and interpretation of data on the population level by, for example, using questionnaires and integrating secondary data. It aims not to give personal feedback to individuals but to identify the range and extent of population-related risk factors to facilitate goal-oriented public health responses and health policies [8].

CW data as part of public health surveillance could be used to describe the PA level of the population continuously and point out trends over time. For example, it is possible to investigate whether certain events, such as heat waves or an outbreak of a pandemic, have an effect on PA and sleep behavior.

The usage of CW data collected by individuals with their own CWs produces a variety of benefits for the research community, like no expenditures on devices and data collection as well as the availability of objective data, which are often superior to self-reports [7,9]. Furthermore, the data are collected continuously by the CW users, which enables researchers to perform retrospective analyses and to observe health behavior over a longer period of time [7]. In addition, the effort required of the participants is low, as they only have to share their data via a smartphone app.

At the same time, there are challenges in using such data for research, such as privacy and security issues, technical issues and a lack of access to the raw data [9]. The algorithms used by the companies of the CWs to convert the raw data into, for example, steps or the duration of PA at different intensities, are often not disclosed [9]. This limits the comparability of CW data over time, as the algorithm may have changed, and between different devices. In addition, a further challenge is the selectivity of the population providing CW data [10,11]. Selectivity of study samples already exists in questionnaire-based observational studies (eg, lower response rates of people with a low educational level) and may increase for data collection using respondents' own CW [12]. A national survey of the US population showed that adults using a CW to monitor their health are younger, healthier, wealthier, and more educated than the general adult population [10]. The low representativeness of older adults using CWs was confirmed by a study from the United Kingdom based on nationwide datasets [13]. In addition, the authors of the United Kingdom study showed that CW users had a higher PA level than the general population. To the best of our knowledge, no study has so far investigated the usage of CWs in the population of Germany and characterized the user group.

Uncritical usage of CW data might lead to biased results that cannot be generalized to the whole population. Within this context, Ibrahim and colleagues [14] used the term "health data poverty" to describe the underrepresentation of certain groups in health datasets, which can foster health inequalities and create a digital health divide.

The first step to counteract health data poverty is to raise awareness of the existing inequalities in health datasets [14]. Therefore, knowledge about the characteristics of the group of CW users is important to be able to make informed decisions about the possibilities and limitations of such datasets and the interpretation of the results. Furthermore, the rapid development of the CW market and the increasing number of cheaper CWs will constantly change the user group, requiring regular characterization of the group of CW users.

Objective

The aim of this study is to describe the CW ownership and usage among adults living in Germany based on data from a nationally representative survey. The intentions are (1) to identify CW owners and users who collect health data and their sociodemographic and health-related characteristics as compared with the population not owning and not using a CW and (2) to gain insights into the motivations of individuals to use their CW to collect health data and their user behavior (eg, the wear time of the CW per day).

Methods

Study Design

The assessment of CW usage was integrated into the representative population-wide survey "German Health Update" (GEDA) 2021 and 2022. GEDA is a regularly conducted cross-sectional survey with the aim of describing the respondents' health status and health behavior and their influencing factors [12].

GEDA 2021 was conducted from July to December 2021 (n=5030) and GEDA 2022 from February 2022 to mid-January 2023 (n=33,149). In both surveys, telephone interviews were carried out using a programmed, structured questionnaire (computer-assisted telephone interview). The sampling is based on a random sample of mobile and landline telephone numbers (dual-frame method) [15]. The population comprised the German-speaking population aged 16 and older (until December 2021; after February 2022: 18 y and older), living in private households, and having a usual place of residence at the time of data collection in Germany. Measures have been taken to increase the response rate of people who are less likely to participate (eg, people with a low educational level and older people), such as oversampling.

The present analysis is limited to persons aged 18 years and older. The topic "consumer wearable usage" was assessed in GEDA in the periods from the end of October to the end of December 2021 (n=1986) and the beginning of February to mid-June 2022 (n=2478; total sample n=4464).

Ethical Considerations

GEDA 2021 and 2022 are subject to strict compliance with the data protection provisions set out in the EU General Data Protection Regulation and the Federal Data Protection Act. The Ethics Committee of the Charité – Universitätsmedizin Berlin assessed the ethics of the study and approved its implementation (application number EA2/201/21). Participation in the study was voluntary. The participants were informed about the aims and contents of the study and about data protection. Informed consent was obtained verbally. Participant confidentiality was maintained by pseudonymizing data and presenting results in

XSL•FO
RenderX

an aggregated format. No compensation or incentives were given for participating in the telephone interview.

Measures

Consumer Wearable–Related Variables

The questions regarding the usage of CWs were prefaced by short definitions of the terms “wearable,” “fitness tracker” and “smartwatch.” The ownership and the usage of a CW were assessed with the questions “Do you own a wearable?” (answer categories: yes or no) and “Do you currently use your wearable to measure your PA, fitness or other health data such as blood pressure or pulse?” (answer categories: yes or no).

The usage of CWs was assessed via the questions “Do you measure your PA, fitness or other health data such as blood pressure or pulse with a fitness tracker and fitness bracelet?” (answer categories: yes or no), followed by the question “On how many days per week do you wear the fitness tracker or fitness bracelet for a total of at least 8 hours per day?” (answer options: the number of days from 1 to 7 and “no day per week”). The same questions were asked to obtain information about the usage of other types of devices (we used the terms “smartwatch” and “other device” instead of “fitness tracker or bracelet”).

The usage of a CW during nocturnal sleep was assessed with the question “Do you wear any of your devices while sleeping at night?” and the answer categories were “no,” “yes, fewer than 4 nights per week” and “yes, at least 4 nights per week.” The motives for CW usage were assessed via the question “Which of the following statements apply to you? I measure my PA, fitness or other health data with a wearable or app on my smartphone...,” and the following motives could be chosen (multiple responses were possible): “to observe how active or inactive I am,” “for support during sports activities or training,” “to motivate myself to be more physically active,” “to monitor a health problem, for example blood pressure, pulse,” “to lose weight,” “to exchange or compare myself with others,” “to eat and drink more healthily,” “because it’s fun for me,” and “for other reasons”

Explanatory Variables

Gender identity was used to describe gender differences. The participants were able to indicate the gender to which they felt they belonged [16]. In the current study sample, 53.8% (2395/4454) identified themselves as female (Table 1). In the analyses by gender, individuals with a different gender identity or no indication are not shown (n=10).

Participants’ age (in years) resulted from their date of birth and the date of the survey. For the analyses, age was categorized into 7 groups: 18 to 29 years, 30 to 39 years, 40 to 49 years, 50 to 59 years, 60 to 69 years, 70 to 79 years and 80 years and older.

Education was categorized into low International Standard Classification of Education (ISCED 1 - 2), medium (ISCED 3 - 4) and high (ISCED 5 - 8) groups based on the educational and vocational qualifications of the study participants according to the 2011 version of the (ISCED 2011) [17].

Based on the self-reported monthly net income of the study participants’ households, the net equivalent income was calculated using the equivalence scale of the Organization for Economic Cooperation and Development. Missing income information was imputed using regression analysis procedures. Income groups were categorized as low (quintile 1), medium (quintiles 2 - 4), and high (quintile 5) for analyses.

For the evaluation of the participants’ subjective health status, they were asked, “How would you describe your general health status?” The five response categories were summarized in the categories “very good or good” and “medium or poor” (including fair, poor and very poor).

The achievement of the endurance-related part of the World Health Organization (WHO) recommendation for PA (at least 150 min moderate to vigorous PA per week [18]) was assessed via the German version of the European Health Interview Survey – PA Questionnaire [19]. The indicator considers the weekly duration of leisure time PA and weekly cycling for transportation. The recommendation was achieved if the sum of the described activities was at least 150 minutes per week.

XSL•FO
RenderX

Table . Description of the study sample (n=4464).

		Proportion unweighted, n (%)	Proportion weighted, n (%)	Missing values
Gender				10
	Male	2059 (46.2)	2059 (48.8)	
	Female	2395 (53.8)	2395 (51.2)	
Age group (years)				0
	18 - 29	322 (7.2)	322 (16.1)	
	30 - 39	418 (9.4)	418 (15.7)	
	40 - 49	551 (12.3)	551 (14.6)	
	50 - 59	920 (20.6)	920 (19.2)	
	60 - 69	1067 (23.9)	1067 (15.3)	
	70 - 79	704 (15.8)	704 (10.7)	
	80+	482 (10.8)	482 (8.5)	
Educational level				14
	Low	237 (5.3)	237 (17.6)	
	Medium	1904 (42.8)	1904 (56.8)	
	High	2309 (51.9)	2309 (25.6)	
Income				0
	Low	509 (11.4)	509 (18.4)	
	Medium	2599 (58.2)	2599 (60.5)	
	High	1356 (30.4)	1356 (21.1)	

Statistical Analysis

All the analyses were performed using a weighting factor that corrects for deviations of the sample from the population structure. First, a design weighting was carried out for the different selection probabilities (mobile and fixed networks), and then an adjustment was made to the official population figures with regard to age, gender, federal state, and district type (as of December 31, 2020) and in relation to education (Microcensus 2018). The analyses were carried out with Stata 17.0 (StataCorp LLC) using the survey procedures. Prevalences are reported with the corresponding 95% CI to display the range in which the true value falls with 95% probability (95% CIs are reported directly in the text or in the corresponding tables or figures). In bivariate analyses, differences between groups were determined using Pearson χ^2 tests. In addition, multivariable adjusted odds ratios (ORs) with corresponding 95% CIs were calculated using logistic regression analyses to identify relevant associations between the sociodemographic and health-related variables and the CW-related outcome variables ("ownership of a CW" and "collection of health data"). In the regression analysis about the "ownership of a CW" 82 participants (82/4464, 1.8% of the total sample) had to be excluded due to missing values for the included variables, and 83 participants had to be excluded (83/4464, 1.9% of the total sample) from the regression analysis regarding the "collection of health data." Following the recommendations of Statistics Canada, estimations with a coefficient of variation of 16.5% to 33.5% were flagged as having a high degree of uncertainty (estimations with even higher coefficients of variations would not be released; this does not apply to any of our results) [20]. A statistically significant difference between groups was assumed if the corresponding *P* value was less than .05.

Results

Study Sample

The study sample comprises 4464 participants aged 18 years and older. The weighted and unweighted distributions of the sample across gender, age, educational, and income groups are presented in Table 1. Participants younger than 40 years of age and those with a low educational level are underrepresented within the study sample, which is corrected by the weighting factor used.

Ownership of a CW

Among the adult population, 19.3% (843/4459) owned a CW (Table 2). In the bivariate analyses, age, educational level, income, subjective health status and PA behavior were significantly associated with CW ownership (Table 2).

XSL•FO
RenderX

Table . Proportion of consumer wearable ownership stratified by sociodemographic and health- and behavior-related variables.

	Ownership of a CW[a]				
	n	Yes, % (95% CI)	n	No, % (95% CI)	n	*P* value[b]
Total	4459	19.3 (17.6 - 21.1)	843	80.7 (78.9 - 82.4)	3616	—[c]
Gender	4449					.99
Female		19.4 (17.0 - 22.1)	421	80.6 (77.9 - 83.0)	1970	
Male		19.4 (16.9 - 22.1)	422	80.6 (77.9 - 83.1)	1636	
Age group (years)	4459					
18 - 29		20 (15 - 26)	74	79.7 (73.8 - 84.6)	248	
30 - 39		31.2 (25.5 - 37.6)	129	68.8 (62.4 - 74.5)	289	
40 - 49		30.0 (24.8 - 35.9)	163	70.0 (64.1 - 75.2)	388	
50 - 59		20.1 (16.8 - 23.8)	222	79.9 (76.2 - 83.2)	696	
60 - 69		13.0 (10.3 - 16.3)	177	87.0 (83.7 - 89.7)	888	
70 - 79		7 (4-10)	61	93.5 (90.5 - 95.6)	643	
80 years and older		2 (1-4)[d]	17	97.6 (95.6 - 98.7)	464	
Educational level	4447					<.001
Low		9 (6-15) [d]	22	90.6 (85.2 - 94.1)	215	
Medium		20.3 (17.8 - 23.0)	334	79.7 (77.0 - 82.2)	1568	
High		23.6 (21.2 - 26.2)	486	76.4 (73.8 - 78.8)	1822	
Income	4459					<.001
Low		11 (8 - 15)	63	89.0 (84.9 - 92.1)	443	
Medium		18.6 (16.4 - 21.1)	444	81.4 (78.9 - 83.6)	2153	
High		28.4 (24.7 - 32.5)	336	71.6 (67.5 - 75.3)	1020	
Subjective health status	4459					<.001
Good or very good		22.0 (19.9 - 24.3)	695	78.0 (75.7 - 80.1)	2616	
Medium or poor		11.6 (9.1 - 14.6)	148	88.4 (85.4 - 90.9)	1000	
PA [e] as recommended [f]	4402					<.001
Yes		23.3 (20.6 - 26.1)	511	76.7 (73.9 - 79.4)	1774	
No		15.4 (13.2 - 17.9)	325	84.6 (82.1 - 86.8)	1792	

[a]CW: consumer wearable.

[b]Pearson χ^2 test

XSL•FO
RenderX

[c]Not applicable.

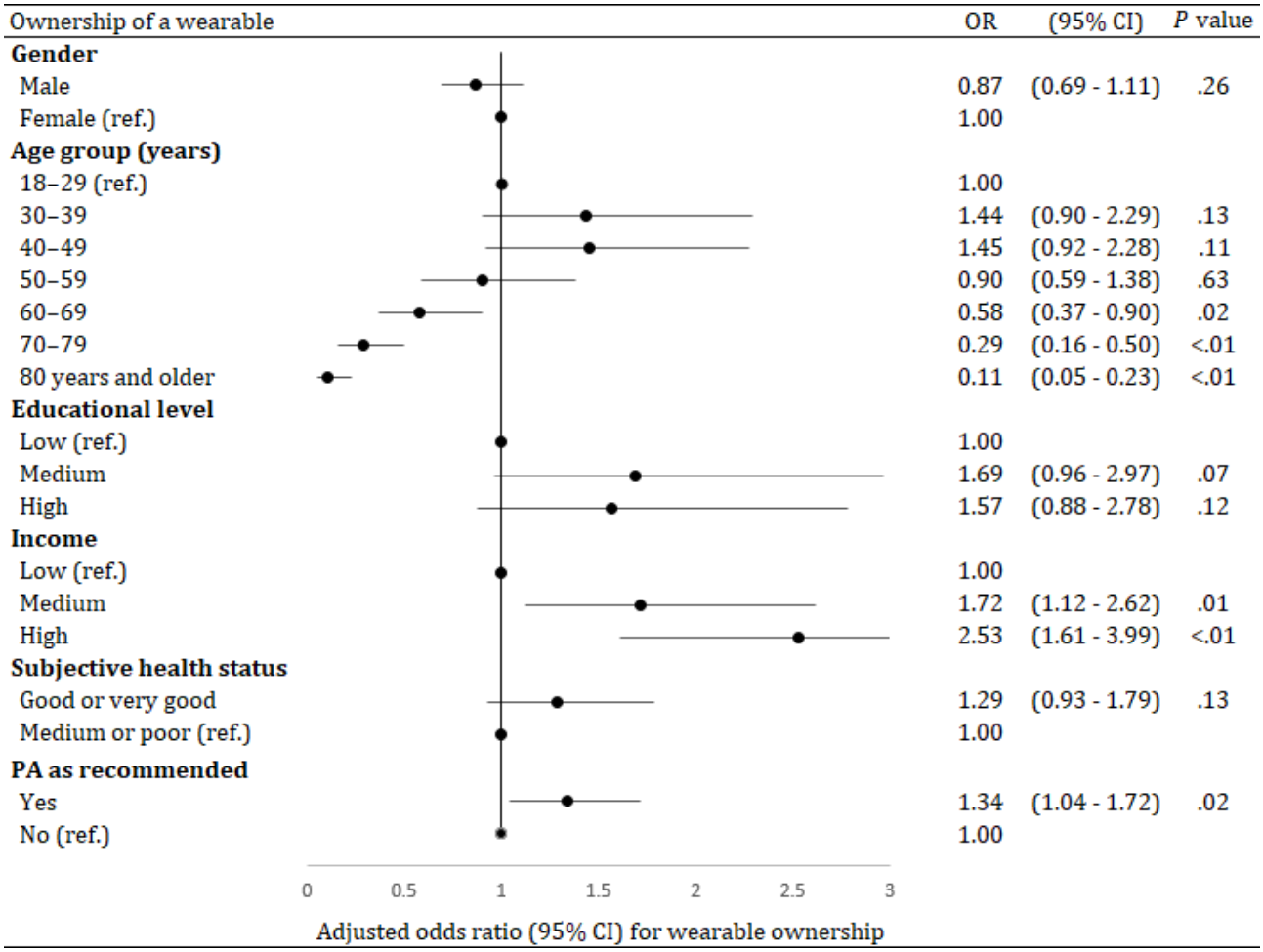
[d]Estimation with high uncertainty due to a small sample size.

[e]PA: physical activity.

[f]At least 150 min PA per week (WHO recommendation).

The multivariable regression analysis, including the dependent variables gender, age, education, income, subjective health status and PA level, revealed that the ownership of a CW was significantly associated with the age, income and PA level of the participants (Figure 1). Adults aged 60 years and older were less likely to own a CW than 18- to 29-year-olds: In 60- to 69-year-olds, the odds of owning a CW was 0.6 times lower than of those aged 18 to 29-year (*P*=.02), in 70- to 79-year-olds the odds was 0.3 times lower (*P*<.01), and of those aged 80 years and older the odds was 0.1 times lower (*P*<.01). Adults with a medium income had 1.7 greater odds to own a CW than adults with a low income (*P*=.01) and adults with a high income had 2.5 greater odds (*P*<.01). In addition, adults who achieved the WHO recommendation for PA had 1.3 greater odds to own a CW than adults not achieving the WHO recommendation (*P*=.02).

Figure 1. Results of a multivariable logistic regression analysis (dependent variable: "Ownership of a consumer wearable," categories: yes or no [reference], adjusted OR with 95% CI, n=4382). "PA as recommended" refers to at least 150 minutes of PA per week (WHO recommendation). OR: odds ratio; PA: physical activity; ref: reference category; WHO: World Health Organization.



Usage of a CW to Collect Health Data

Of the adult population, 15.0% (650/4458) used a CW to collect health data (Table 3). This corresponds to 77.8% (650/842) of the CW owners. In the bivariate analyses, we compared CW users with nonusers, and the respondents' age, education, income, subjective health status, and PA level were significantly associated with the collection of health data with their own CW (Table 3).

In the multivariable regression analysis, the variables age, income and PA level remained significant predictors for the collection of health data (Figure 2). The odds of adults aged 70 to 79 years to collect health data using a CW was 0.3 times lower than in 18- to 29-year-olds (*P*<.01). Participants aged 80 years and older had 0.1 lower odds to collect health data using a CW than the youngest age group (*P*<.01). In addition, participants who achieved the WHO recommendation for PA had 1.5 higher odds to collect health data than participants who did not achieve this recommendation (*P*=.01).

XSL•FO
RenderX

Table . Proportion of collection of health data using respondents' own consumer wearable stratified by sociodemographic and health- and behavior-related variables.

		Collection of health data using respondents' own CW[a]					
		n	Yes, % (95% CI)	n	No, % (95% CI)	n	*P* value[b]
Total		4458	15.0 (13.4 - 16.7)	650	85.0 (83.3 - 86.6)	3808	—[c]
Gender		4448					.99
	Female		15.1 (12.9 - 17.6)	322	84.9 (82.4 - 87.1)	2068	
	Male		15.1 (12.9 - 17.7)	328	84.9 (82.3 - 87.1)	1730	
Age group (years)		4458					<.001
	18 - 29		15 (1 - 21)	53	85.0 (79.5 - 89.3)	269	
	30 - 39		25 (20 - 32)	96	74.7 (68.4 - 80.0)	322	
	40 - 49		23.6 (18.8 - 29.1)	130	76.4 (70.9 - 81.2)	420	
	50 - 59		15.8 (12.9 - 19.1)	177	84.2 (80.9 - 87.1)	741	
	60 - 69		10.3 (7.8 - 13.4)	139	89.7 (86.6 - 92.2)	926	
	70 - 79		4 (3-7)[d]	45	95.6 (92.9 - 97.3)	659	
	80 years and older		2 (1-4)[d]	10	98.5 (96.5 - 99.4)	471	
Educational level		4446					.01
	Low		7 (4-12)[d]	14	93.0 (87.8 - 96.1)	223	
	Medium		16.1 (13.8 - 18.6)	260	83.9 (81.4 - 86.2)	1642	
	High		17.9 (15.7 - 20.2)	375	82.1 (79.8 - 84.3)	1932	
Income		4458					<.001
	Low		9 (6-12)	46	91.4 (87.6 - 94.2)	460	
	Medium		14.7 (12.6 - 17.0)	342	85.3 (83.0 - 87.4)	2255	
	High		21.6 (18.3 - 25.2)	262	78.4 (74.8 - 81.7)	1093	
Subjective health status		4458					<.001
	Good or very good		17.3 (15.3 - 19.4)	541	82.7 (80.6 - 84.7)	2769	
	Medium or poor		8.6 (6.5 - 11.4)	109	91.4 (88.6 - 93.5)	1039	
PA [e] as recommended [f]		4401					<.001
	Yes		18.7 (16.3 - 21.4)	407	81.3 (78.6 - 83.7)	1878	
	No		11.4 (9.4 - 13.6)	237	88.6 (86.4 - 90.6)	1879	

[a]CW: consumer wearable.

XSL•FO
RenderX

[b]Pearson χ^2 test.

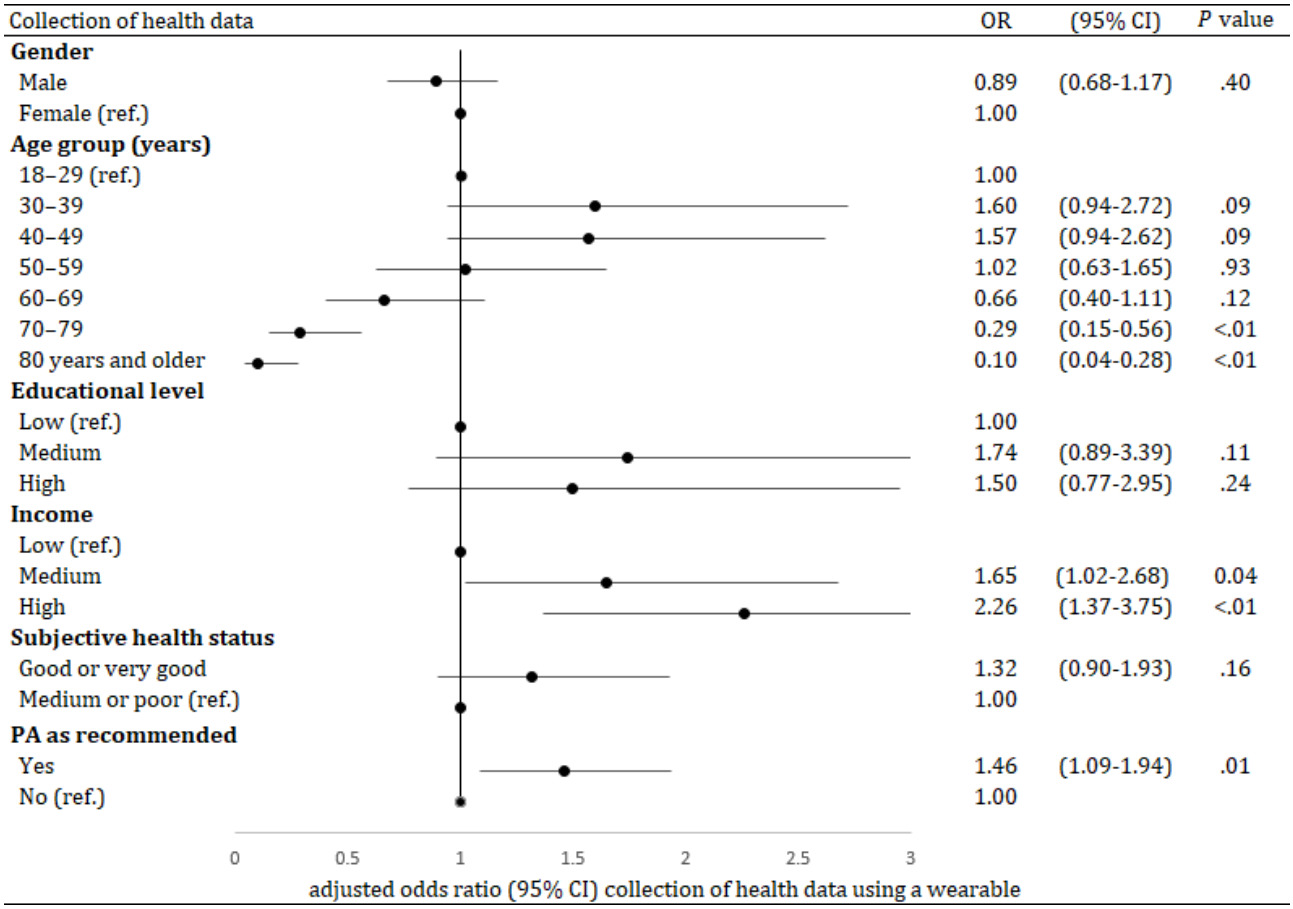
[c]Not applicable.

[d]Estimation with high uncertainty due to a small sample size.

[e]PA: physical activity.

[f]At least 150 min PA per week (WHO recommendation).

Figure 2. Results of a multivariable logistic regression analysis (dependent variable: "Collection of health data using a consumer wearable," categories: yes or no (reference), adjusted OR with 95% CI, n=4381). "PA as recommended" refers to at least 150 minutes of PA per week (WHO recommendation). OR: odds ratio; PA: physical activity; ref: reference category; WHO: World Health Organization.



User Behavior

The most commonly used CW to collect health data was a smartwatch. A total of 74.5% (456/648, 95% CI 69.1% - 79.3%) of the CW users who collected health data used a smartwatch, of whom 71.8% (341/456, 95% CI 65.0% - 77.8%) wore their smartwatch for at least 8 hours per day. Fitness trackers were used by 60.7% (414/650, 95% CI 54.7% - 66.4%), of whom 70.7% (310/413, 95% CI 63.0% - 77.3%) wore it for at least 8 hour per day. Furthermore, 26.8% (214/650, 95% CI 22.1% - 32.2%) used a CW other than a smartwatch or fitness tracker. The results show that some of the participants used more than one device.

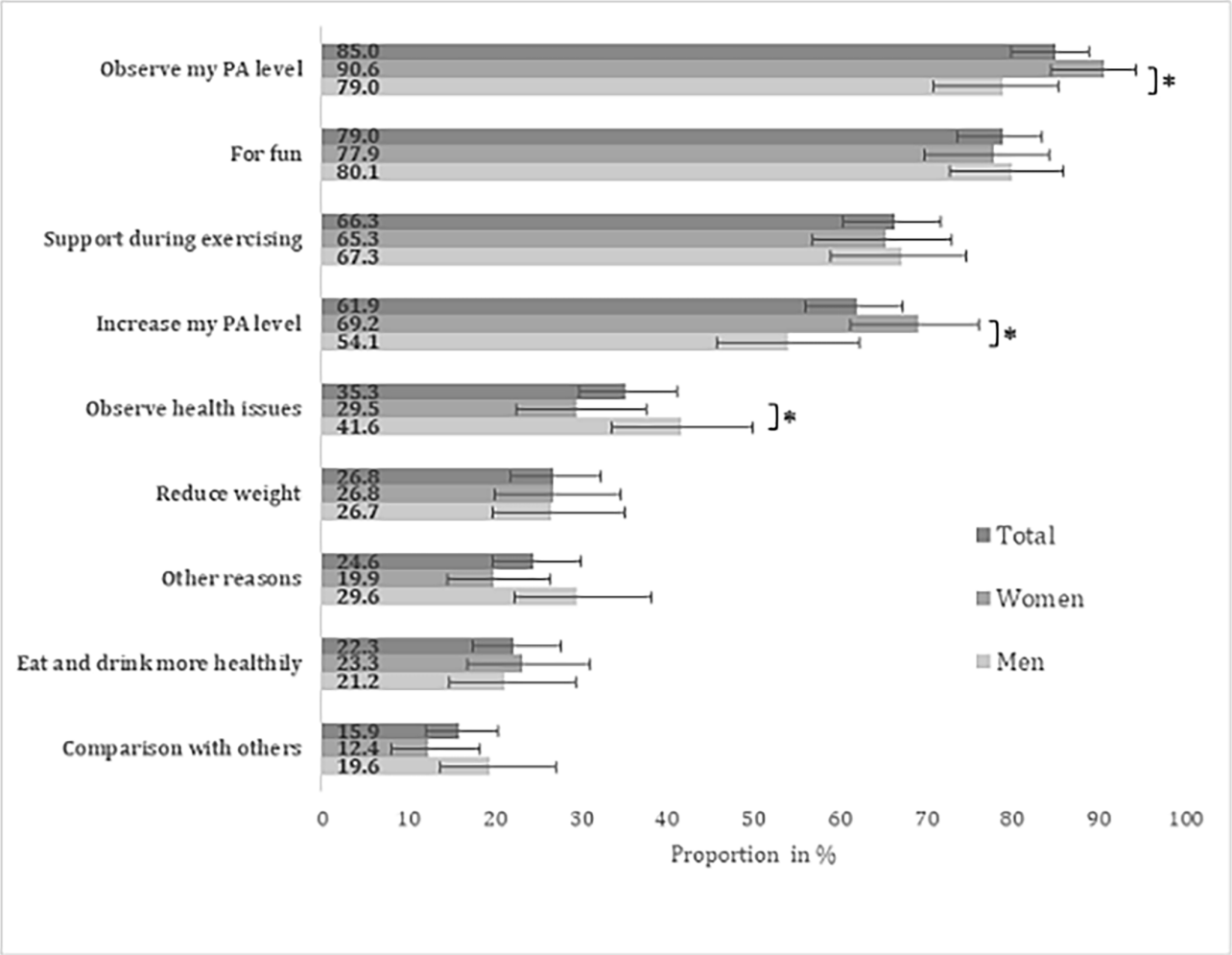
Of the CW users who collected health data, 47.2% (321/650, 95% CI 41.4% - 53.2%) wore their CW during nocturnal sleep for at least 4 nights per week, and an additional 6% (42/650, 95% CI 4.0% - 9.1%) wore it for fewer than 4 nights per week. Consequently, 46.7% (287/650, 95% CI 40.8% - 52.7%) did not collect sleep data.

Motives for CW Usage

The most frequently chosen motives for the usage of a CW were "to observe my PA level" (544/647, 85.0%), "for fun" (508/644, 79.0%), "for support during exercise" (423/647, 66.3%), and "to increase my PA level" (386/647, 61.9%). Women named the motives "to observe my PA level" and "to increase my PA level" more often than men (287/321, 90.6% vs 257/326, 79.0%, *P*=.01; 211/321, 69.2% vs 175/326, 54.1%, *P*=.01, respectively), whereas men chose the motive "to observe health issues" more often than women (137/326, 41.6% vs 103/321, 29.5%, *P*=.04). The described differences between women and men remained statistically significant when the analyses were adjusted for age, educational level, income, subjective health status and PA level (results of multivariable logistic regression analyses, data not shown). The motives for the usage of a CW for the total sample and stratified by gender are shown in Figure 3.

XSL•FO
RenderX

Figure 3. Reported motives for the usage of consumer wearables (multiple responses were possible). Proportions with 95% CIs are shown. Sample: wearable users who collect health data (n=650). PA: physical activity. *P<.05.



Discussion

Principal Findings

To our knowledge, this is the first nationwide study to describe comprehensively the ownership of CWs and the usage of CWs for the collection of health data among the adult population living in Germany. In addition, the selectivity of the user group was investigated. We found that almost one-fifth of the adult population living in Germany owned a CW, 77.8% (650/842) of whom used their CW to collect health data (which corresponds to 650/4458, 15.0% of the adult population). Older people, people with a low income, and people with a lower PA level were less likely to own a CW and to use their own CW to collect health data. Of the 4 most frequently chosen motives for collecting health data using a CW, 3 were related to PA behavior, but using the CW for fun was also a common motive.

Comparison With Previous Work

Compared with our results, a nationwide survey from Canada conducted in 2017 showed a slightly higher prevalence of CW ownership, whereas the prevalence of using the device for the collection of health data in relation to the total population was almost the same: 25% of the Canadian population owned a CW (843/4459, 19% in this study) and 57% of the owners used the device to collect health data [21] (which corresponds to 14% of the population; 15% (650/4458) in our study). A nationwide study from the United Kingdom conducted in 2018 assessed the usage of a CW to collect health data and revealed similar prevalences to those of this study, with 14% using a smartwatch or fitness tracker [13], whereas a nationwide survey from the United States conducted in 2019 revealed a higher prevalence, with 30% using a CW in the past 12 months [10]. However, it should be borne in mind that the data collection for the studies described above took place 3 to 5 years before our study and that the prevalences of the ownership and usage of CWs may have increased in the meantime.

Our results concerning the selectivity of the group of CW owners and users confirm the findings of studies from the United States [10], the United Kingdom [13] and Canada [21] for the population living in Germany. Among CW users, people who are the most in need of health care and health promotion, such as older people and socio-economically disadvantaged people, are underrepresented. Older people (aged 60 y and older) may use CWs less frequently because they are less tech savvy, they may be more reluctant to use new technologies and they may have less fine motor skills or impaired vision, making it difficult to operate a CW or smartphone [22,23]. A study from the United States showed that seniors with higher technical self-efficacy were more likely to use a CW [22]. Thus, the improvement of technical skills in combination with the further tailoring of

XSL•FO
RenderX

devices to the needs of older people could increase the user behavior in this age group.

The lower likelihood of people with low income owning and using a CW in this study is consistent with the findings from the United States and Canada [10,21]. Furthermore, in the nationwide study from Canada, it was shown that costs are a common mentioned reason for not owning a CW [21]. However, due to the increasing proportion of low-cost CWs on the market, the economic aspect will probably become less important in the future. In addition, adults with a lower income often have a poorer state of health and lower PA levels, which might be accompanied by less interest in these behaviors and thus less interest in tracking health and PA indicators using a CW [24,25].

Our results revealed that adults with a higher PA level were more likely to own a CW and to use it for the collection of health data. A higher level of PA in CW users than in nonusers has also been shown in previous observational studies [26,27]. People with a higher level of PA may have a greater interest in their PA behavior and are thus more likely to track it using a CW than people with a lower level of PA [28]. This could also explain our findings on the motives for using a CW: 3 of the 4 most frequently named motives were related to PA behavior. Conversely, the use of CWs can influence the PA behavior of the users and increase their PA level. The authors of an umbrella review concluded that the usage of CWs is effective in increasing PA [29]. However, the positive effects are significantly smaller when the usage of a CW is the only measure compared with multifaceted interventions to promote PA [30]. In addition, little is known about the long-term effects of CW usage on PA behavior.

Health data collected by the population using their own CW can be used for different purposes. Our results indicate that almost 20% (843/4459) of the members of the adult population living in Germany who use a CW and collect health data wear the device regularly during daytime and thus provide data that reflect the general behavior of this group. However, only half of the CW users wear their device during nocturnal sleep, which restricts the possibilities for sleep analyses significantly.

Strengths and Limitations

A strength of our study is that we used a nationally representative survey to investigate the ownership and usage of CWs and the motives in the adult population living in Germany. In addition, we analyzed a variety of sociodemographic and health- and behavior-related predictors of CW ownership and usage. Our results can be used to inform researchers designing research projects including CW data, as well as practitioners. Nevertheless, the following limitations of our study must be considered. First, the analyses of the factors predicting the ownership and usage of CWs might miss predictors of CW ownership or usage because they were not assessed in our study. For example, further societal indicators, like cultural background, might have an impact on CW usage as well as the technical competence of a person. Second, our results provide some initial insights into the motives of using CWs, but we might have missed information about relevant motives because almost one-quarter of the sample used the answer category “other reasons.” Further research using open-ended questions or qualitative research approaches could help to better understand the motives for CW use. Third, small sample sizes of specific subgroups, such as younger adults and adults with a low educational level, resulted in estimations with higher measurement uncertainty. We have marked results with high measurement uncertainty to make this limitation transparent. In addition, we were not able to perform more in-depth analyses, such as further subgroup analyses, due to the small sample size in the subgroups. Fourth, the assessment of data using self-reports is a limitation because we cannot rule out the possibility of biased results due to recall bias or different interpretations of terms like “wearable,” “smartwatch,” or “fitness tracker.” However, we included short definitions of the aforementioned terms in the interviews to minimize the possibility of different interpretations. Finally, the cross-sectional design of the dataset does not allow us to clarify predictive directions. Future longitudinal study designs are needed to contribute to the understanding of the long-term effects of CW use on PA behavior.

Implications for Research and Practice

In our opinion, a major challenge when using health data collected by individuals through their own CW is that the sample is selective and does not represent the general population. Consequently, parts of the population who are particularly likely to need health care and the improvement of the health care system, such as older and socially disadvantaged people, would not benefit from these research approaches or might even be harmed. Thus, the use of data collected with respondents’ own device is, according to the current status, not suitable for the surveillance of health behavior on the populational level. A possibility to reduce the selectivity could be to supplement the sample with a randomly chosen sample of individuals who do not own a CW. This group could be sent a fitness tracker for the period of the study. Oversampling of difficult-to-reach groups of the population, like older people or people with low incomes, as well as a study design and study information adapted to difficult-to-reach target groups can further reduce selectivity.

However, health data collected by individuals using their own CW could be used to achieve other research aims than population-wide surveillance of health behavior. For example, they could be used to describe trends in PA and sleep behavior and to investigate whether certain events, like heat waves, have an impact on these behavior. In addition, these data could be used to describe individual trends in PA behavior in the form of longitudinal analyses to, for example, evaluate how the transition from employment to retirement affects PA and sleep behavior.

Researchers and practitioners should still consider carefully which parts of the population are underrepresented in their CW datasets and which conclusions can be drawn, and they should communicate the selectivity of the sample transparently. Eventually, it depends on the research question whether CW data collected by individuals with their own device are suitable for answering it. The results of our study might support the decision process that researchers should execute when planning a study using CW data.

XSL•FO
RenderX

Conclusion

Our results indicate that one-fifth of the adult population living in Germany owns a CW. Individuals owning a CW are younger, have higher incomes, and are more physically active than the general population. The usage of such data describing the health behavior at the population level will therefore overestimate health-promoting behaviors such as PA.

Acknowledgments

The authors would like to thank the participants in the German Health Update (GEDA) study. We would also like to thank those who funded the surveys, undertook the fieldwork, and processed the data.

GEDA 2021 and 2022 were funded by the Robert Koch Institute and the German Federal Ministry of Health.

Conflicts of Interest

None declared.

References

1. Hicks JL, Althoff T, Sosic R, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2(45):45. [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
2. Absatz von Wearables weltweit in den Jahren 2014 bis 2022 (in Millionen Stück). Staistica. 2023. URL: https://de.statista.com/statistik/daten/studie/515723/umfrage/absatz-von-wearables-weltweit/ [accessed 2025-05-28]
3. Absatz von Wearables in Deutschland in den Jahren 2015 bis 2022 (in Millionen Stück). Statistica. 2023. URL: https://de.statista.com/statistik/daten/studie/551366/umfrage/absatz-von-wearables-in-deutschland/ [accessed 2025-05-28]
4. Dixon WG, van der Veer SN, Ali SM, et al. Charting a course for smartphones and wearables to transform population health research. J Med Internet Res 2023 Feb 7;25:e42449. [doi: 10.2196/42449] [Medline: 36749628]
5. Kang HS, Exworthy M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth 2022 Jul 13;10(7):e35684. [doi: 10.2196/35684] [Medline: 35830222]
6. Strain T, Wijndaele K, Pearce M, Brage S. Considerations for the use of consumer-grade wearables and smartphones in population surveillance of physical activity. J Meas Phys Behav 2022 Mar 1;5(1):8-14. [doi: 10.1123/jmpb.2021-0046]
7. Mair JL, Hayes LD, Campbell AK, Sculthorpe N. Should we use activity tracker data from smartphones and wearables to understand population physical activity patterns? J Meas Phys Behav 2022 Mar 1;5(1):3-7. [doi: 10.1123/jmpb.2021-0012]
8. Shakeri Hossein Abad Z, Kline A, Sultana M, et al. Digital public health surveillance: a systematic scoping review. NPJ Digit Med 2021 Mar 3;4(1):41. [doi: 10.1038/s41746-021-00407-6] [Medline: 33658681]
9. Huhn S, Axt M, Gunga HC, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth 2022 Jan 25;10(1):e34384. [doi: 10.2196/34384] [Medline: 35076409]
10. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443. [doi: 10.2196/22443] [Medline: 33064083]
11. Cho PJ, Yi J, Ho E, et al. Demographic imbalances resulting from the bring-your-own-device study design. JMIR Mhealth Uhealth 2022 Apr 8;10(4):e29510. [doi: 10.2196/29510] [Medline: 34913871]
12. Allen J, Born S, Damerow S, et al. German Health Update (GEDA 2019/2020-EHIS) - background and methodology. J Health Monit 2021 Sep;6(3):66-79. [doi: 10.25646/8559] [Medline: 35146317]
13. Strain T, Wijndaele K, Brage S. Physical activity surveillance through smartphone apps and wearable trackers: examining the UK potential for nationally representative sampling. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e11898. [doi: 10.2196/11898] [Medline: 30694198]
14. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021 Apr;3(4):e260-e265. [doi: 10.1016/S2589-7500(20)30317-4] [Medline: 33678589]
15. Sand M, Gabler S. Gewichtung von (dual-frame-)telefonstichproben. In: Häder S, Häder M, Schmich P, editors. Telefonumfragen in Deutschland: Springer VS; 2019.
16. Pöge K, Rommel A, Starker A, et al. Survey of sex/gender diversity in the GEDA 2019/2020-EHIS study - objectives, procedure and experiences. J Health Monit 2022 Jun;7(2):48-65. [doi: 10.25646/9958] [Medline: 35891940]
17. Internationale Standardklassifikation für das Bildungswesen (ISCED). Eurostat. 2017. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:International_standard_classification_of_education_(ISCED)/de [accessed 2025-05-28]
18. WHO guidelines on physical activity and sedentary behaviour. World Health Organization. 2020 Nov 25. URL: https://www.who.int/publications/i/item/9789240015128 [accessed 2025-05-28]
19. Finger JD, Tafforeau J, Gisle L, et al. Development of the European Health Interview Survey - Physical Activity Questionnaire (EHIS-PAQ) to monitor physical activity in the European Union. Arch Public Health 2015;73(59):59. [doi: 10.1186/s13690-015-0110-z] [Medline: 26634120]

XSL•FO
RenderX

20. Quality level guidelines. Statistics Canada. 2009. URL: https://www150.statcan.gc.ca/n1/pub/13f0026m/2007001/table/tab5p1-eng.htm [accessed 2023-10-09]
21. Paré G, Leaver C, Bourget C. Diffusion of the digital health self-tracking movement in Canada: results of a national survey. J Med Internet Res 2018 May 2;20(5):e177. [doi: 10.2196/jmir.9388] [Medline: 29720359]
22. Chandrasekaran R, Katthula V, Moustakas E. Too old for technology? Use of wearable healthcare devices by older adults and their willingness to share health data with providers. Health Informatics J 2021;27(4):14604582211058073. [doi: 10.1177/14604582211058073] [Medline: 34802315]
23. Chung J, Brakey HR, Reeder B, Myers O, Demiris G. Community-dwelling older adults' acceptance of smartwatches for health and location tracking. Int J Older People Nurs 2023 Jan;18(1):e12490. [doi: 10.1111/opn.12490] [Medline: 35818900]
24. Beenackers MA, Kamphuis CBM, Giskes K, et al. Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: a systematic review. Int J Behav Nutr Phys Act 2012 Sep 19;9:116. [doi: 10.1186/1479-5868-9-116] [Medline: 22992350]
25. Wachtler B, Hoebel J, Lampert T. Trends in socioeconomic inequalities in self-rated health in Germany: a time-trend analysis of repeated cross-sectional health surveys between 2003 and 2012. BMJ Open 2019 Sep 27;9(9):e030216. [doi: 10.1136/bmjopen-2019-030216] [Medline: 31562151]
26. Leung W, Shi L, Jung J. Comparison of participation in strength activity between wearable device users and nonusers: 2017 Behavioral Risk Factors Surveillance System. J Phys Act Health 2022 Oct 1;19(10):673-680. [doi: 10.1123/jpah.2022-0155] [Medline: 36087931]
27. Xie Z, Jo A, Hong YR. Electronic wearable device and physical activity among US adults: an analysis of 2019 HINTS data. Int J Med Inform 2020 Dec;144(104297):104297. [doi: 10.1016/j.ijmedinf.2020.104297] [Medline: 33059242]
28. Jin D, Halvari H, Maehle N, Olafsen AH. Self-tracking behaviour in physical activity: a systematic review of drivers and outcomes of fitness tracking. Behav Inf Technol 2022 Jan 25;41(2):242-261. [doi: 10.1080/0144929X.2020.1801840]
29. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
30. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]

Abbreviations

CW: consumer wearable
GEDA: German Health Update
ISCED: International Standard Classification of Education
OR: odds ratio
PA: physical activity
WHO: World Health Organization

Edited by L Buis; submitted 05.04.24; peer-reviewed by G Christopoulos, H Kao; revised version received 29.11.24; accepted 10.04.25; published 11.06.25.

Please cite as:
Manz K, Krug S, Kühnelt C, Lemcke J, Öztürk I, Loss J
Consumer Wearable Usage to Collect Health Data Among Adults Living in Germany: Nationwide Observational Survey Study
JMIR Mhealth Uhealth 2025;13:e59199
URL: https://mhealth.jmir.org/2025/1/e59199
doi:10.2196/59199

© Kristin Manz, Susanne Krug, Charlotte Kühnelt, Johannes Lemcke, Ilter Öztürk, Julika Loss. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

College Community–Based Physical Activity Support at a Public University During the COVID-19 Pandemic: Retrospective Longitudinal Analysis of Intra- Versus Interpersonal Components for Uptake and Outcome Association

Garrett I Ash[1,2,3*], PhD; Selene S Mak[4*], MPH, PhD; Adrian D Haughton[5], BS; Madilyn Augustine[6], BA; Phillip O Bodurtha[5], MSc; Robert S Axtell[5], PhD; Beatrice Borsari[1,2], PhD; Jason J Liu[1,2], PhD; Shaoke Lou[2], PhD; Xin Xin[7,8], MS; Lisa M Fucito[6,9,10], PhD; Sangchoon Jeon[11], PhD; Matthew Stults-Kolehmainen[12,13], PhD; Mark B Gerstein[1,2,14,15,16], PhD

[1]Program in Computational Biology & Bioinformatics, Yale University, New Haven, CT, United States

[2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, United States

[3]Center for Pain, Research, Informatics, Medical Comorbidities and Education Center (PRIME), VA Connecticut Healthcare System, West Haven, CT, United States

[4]Center for the Study of Healthcare Innovation, Implementation & Policy (CSHIIP), VA Greater Los Angeles Healthcare system, Los Angeles, CA, United States

[5]Department of Health and Movement Sciences Programs, Southern Connecticut State University, New Haven, CT, United States

[6]Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States

[7]Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, CT, United States

[8]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States

[9]Yale Cancer Center, New Haven, CT, United States

[10]Smilow Cancer Hospital, Yale New Haven Hospital, New Haven, CT, United States

[11]Yale School of Nursing, Orange, CT, United States

[12]Center for Weight Management, Yale New Haven Hospital, New Haven, CT, United States

[13]Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, NY, United States

[14]Department of Computer Science, Yale University, New Haven, CT, United States

[15]Department of Statistics and Data Science, Yale University, New Haven, CT, United States

[16]Section of Biomedical Informatics and Data Science, Yale University, New Haven, CT, United States

[*]these authors contributed equally

Corresponding Author:
Mark B Gerstein, PhD
Program in Computational Biology & Bioinformatics
Yale University
Bass 432
266 Whitney Ave
New Haven, CT, 06520
United States
Phone: 1 203 432 8189
Email: mark@gersteinlab.org

Abstract

Background: College students are vulnerable to setting long-term trajectories of low physical activity (PA) but are reachable via mobile health fitness tracking (eg, mobile health step counting) and interpersonal support tailored to the college community. However, no studies have statistically isolated the appeal and influence of these intra- and interpersonal components in college-based PA interventions.

XSL•FO
RenderX

Objective: This study retrospectively examined a college-based PA promotion program at a northeast US public university during the COVID-19 pandemic to (1) test the impact of student status on the use of intervention components and (2) determine whether such use was associated with successful retention and goal achievement in the program.

Methods: The university used a commercial platform for a 30-day PA promotion program during April 2021 with intrapersonal (step-tracker syncing, education, self-monitoring, and motivational messaging) and interpersonal (friend interactions and team games) components. App use was operationalized as intrapersonal (frequency of opening app, education, and self-monitoring) and interpersonal (friends made in-app and team affiliation and size).

Results: Campus-wide emails elicited sign-up by 156 undergraduate students, 57 graduate students, and 126 faculty and staff members. Objective 1 yielded the following results: undergraduates used the app less frequently (median 0.8, IQR 0.4-1.7 times per day) than other groups (graduate students: median 1.4, IQR 0.7-2.7 times per day; *P*=.01; faculty: median 1.3, IQR 0.7-2.7 times per day, H_2=14.5; *P*=.001) but made the same number of friends (median 1-2) and teammates (median 8-9; *P*=.77 for friends and *P*=.93 for teammates). Objective 2 yielded the following results: most participants (313/335, 93.4%; 95% CI 90%-96%) were retained for the first 7 days, but by 30 days, retention dropped, most notably for undergraduate students (82/154, 53.2%; 95% CI 45%-61%), followed by graduate students (39/56, 70%; 95% CI 56%-81%) and faculty and staff (93/125, 74.4%; 95% CI 66%-82%; χ^2_2=12.6; *P*<.001). Retention was associated with app engagement frequency (model hazard ratio 0.56, 95% CI 0.43-0.72; *P*<.001) and affiliation with a team having high median app engagement and a large size (intracluster correlation coefficient 0.064, 95% CI 0.001-0.164, *P*=.05). Meeting a daily step goal was associated with app engagement frequency (β=.72, SE=0.21; *P*=.001), number of friends (β=.40, SE 0.20; *P*=.04), and an initial motive of maintaining or increasing (rather than starting) PA (β=.99, SE=0.21; *P*<.001).

Conclusions: College students, compared with faculty and staff, used the app less frequently, used the app for a shorter duration before abandonment, and met the step goal on fewer days. Engagement with the program was associated with longer retention and better PA outcomes, which were critically modified by the interpersonal engagement. These findings suggest that college students using virtual PA support during times of physical isolation could benefit from more tailored implementation strategies (eg, timed prompts and team reassignments).

(JMIR Mhealth Uhealth 2025;13:e51707) doi:10.2196/51707

KEYWORDS

universities; students; faculty; mobile apps; goals; social networks; education

Introduction

Background

Physical activity (PA) levels are important determinants of long-term health outcomes. For instance, 6000 to 8000 steps per day is associated with a lower risk of health problems such as all-cause mortality, diabetes, and mental health concerns [1,2]. One critical life stage for PA support is during the college years for those who attend. For some people, college is a transition when great positive or negative change from adolescence is possible, and the lifestyle at the end of this transition is a predictor of the long-term health trajectory [3]. Studies have found that around half of college students have low PA levels [4], which further declined by 32% to 366% during the COVID-19 pandemic [5]. Thus, college is a critical window for interventions that maintain or increase PA to have a powerful and sustained impact on an individual's life.

College students are a prime population to reach through mobile health (mHealth; ie, medicine involving mobile devices) because most consider the internet a source of health information [6,7], almost all own a smartphone, and around half use a fitness tracker or other wearable device [8]. mHealth could be a viable strategy to help college students increase their PA because their use of wearable fitness trackers and mobile apps has been associated with higher PA levels.

This association in these studies was mediated by the social aspects of using such technology (eg, real-time sharing of step counts), making it critical to understand the social context in which students are functioning. Starting college presents students with a new set of social circumstances [9]. When they leave the outside world and enter the college community, often one with cultural and geographic differences from where they grew up, there is a desire to belong and feel accepted in this new community [9]. Such an innate want for interpersonal connections is a fundamental determinant of human cognition and health outcomes including PA [10-12]. Fortunately, this new community presents unique opportunities to make social connections. First, students are assigned class instructors with high education levels and experience mentoring young adults, a cross-demographic connection less likely to occur in the outside world [9,11]. Second, students share experiences and concerns with other students such as accessing unique or licensed educational content, feeling eustress or distress about academic performance, and concerns about loneliness because of the restricted size of the college community [9,11]. On the other hand, students often struggle to initially adapt to these aspects of college [9,12]. The COVID-19 pandemic further heightened students' unmet needs and desire for social connection [13,14]. However, when students are able to use the college environment opportunities to join peers in pursuing activities and exploring interests, these connections correlate with increased PA participation and engagement [11]. In fact, social support from friends effectively motivates PA among

XSL•FO
RenderX

college students, more so than family support [15]. Yet, despite these rewards of connecting to peers of like age and education level in college, it can also be overwhelming. For instance, while social support in college correlates with situational interest in PA,1 study found that the college students' environment was a barrier to PA because of feeling overwhelmed by the change in their first year of attendance [16]. For these reasons, college PA support must be considered and tailored to the college environment.

There are several frameworks for understanding determinants of human goal setting and accomplishment such as increasing PA. Goal-setting theory and the theory of planned behavior focus on intrapersonal determinants such as the specificity of the goal and beliefs about results from the goal. The self-determination theory also includes such intrapersonal competence but then extends to the inclusion of interpersonal relationships (ie, autonomy and relatedness) [17]. The social cognitive theory (SCT) extends beyond one-to-one and small-group interpersonal relationships to the broader sociocultural environment [18]. The SCT's intrapersonal determinants include self-efficacy (belief in oneself to perform a behavior) and outcome expectations (the expectation that a desired outcome will result from the behavior). Its interpersonal determinants include sociocultural factors (the influences from an individual's social and physical environment). Importantly, there is a reciprocal determinism (ie, feedback loops) between the intra and interpersonal determinants. For example, social persuasion and support from peers can boost a person's self-efficacy to become a successful athlete. Similarly, when peers see that heightened self-efficacy and athleticism, they are even more inclined to support that person. Therefore, it is critical to study both the intra and interpersonal aspects of PA support programs.

Wearables can support the aforementioned SCT constructs. They can support self-efficacy by conveyance of personal accomplishment (eg, tracking step counts in relation to previous days or a preset goal) and verbal persuasion (eg, encouragement messages to reach a goal), as well as by offering the opportunity to leverage this self-efficacy for goal setting. Self-efficacy has been shown to impact both engagement with wearable technology [19] and behavioral outcomes resulting from its use [20]. Wearables can also support outcome expectations because the metrics they collect and display such as step counts are associated with health benefits [1,2]. Finally, wearables can support sociocultural influences through methods such as group step challenges and social media postings of achievements [21,22]. Overall, this robust applicability to SCT shows that wearables should be explored to deliver both inter and intrapersonal intervention components among college students.

Health-related interventions often have low uptake among college students [23-25]. Some reasons are related to intrapersonal aspects such as students prioritizing more pressing needs and disbelief the information was valid and applicable to them. Other reasons related to interpersonal aspects such as lack of peer involvement in designing the intervention, and the interpersonal interactions occurring over a social media platform they were not accustomed to using. Students also noted the criticality of being able to conveniently select which intervention components to use. Thus, both the intra- and interpersonal components clearly represent areas for improvement. Some preliminary evidence suggests that social support could be incorporated into mHealth PA interventions for the college student population to promote PA adherence. Findings from a systematic review of 19 papers suggest a positive association between PA and social support for PA among college students [26]. A total of 14 primary studies leveraged social interactions in college community populations to promote PA through mobile apps, including social components [21-23,27-38]. Some of these mobile apps provided an "open forum" feature (eg, Facebook-style news feed and WeChat group chat) where participants could share PA performance, interact, and provide social (ie, interpersonal) support [21,23,27-32,35,36]. Some discussed an additional feature in the mobile app which allowed participants to form ongoing "team" relationships. These relationships included an ongoing comparison of step counts with other team members (ie, a social norm) [22,27-30,38], discussions fostering mutual support (eg, complimenting success and empathizing with barriers) [21,22,33-37], and discussing educational materials on PA [22]. These activities were rewarded with team success scores, thus leveraging the established gamification strategies of cooperation [39] and socially oriented persuasion [40]. These studies incorporated a mobile app with these interpersonal components, activity trackers, and standard solitarily (ie, intrapersonally) used behavior change techniques such as text message reminders to be active, health education, goal setting, self-monitoring, and extrinsic incentives (eg, prizes). The utility for tracking devices without support from a multimodal app tailored for college students has yielded mixed results (reviewed in the study by Sultoni et al [41]), thus, multimodal communication is considered best practice for mHealth interventions [42].

One challenge with current research is determining the use and effectiveness of each of the multiple intervention components; this information is crucial to the refinement and success of the intervention [43]. Edney et al [30] found self-monitoring and ongoing team relationships were rated highly by 50% to 70% of participants compared with individual incentives and open forum social interactions, which only 20% to 30% of participants rated highly. Tong et al [29] found the self-monitoring and ongoing team relationship features were used an average of approximately 2 times per week, while open forum use was less than weekly. When the app was subsequently refined to include more personalization of the individual self-monitoring, by tailoring it to participants' weekly progress and barriers, this approach was reported to be well-received [32]. Zhang et al [37] found participants randomly assigned to participate in an interpersonal team did not achieve greater PA increases than those randomly assigned to receive intrapersonal text messages. There is a need to build upon these mixed findings from clinical trials by statistically isolating which of the intrapersonal (ie, education, self-monitoring, and personalized advice) and interpersonal components (ie, open forum and team feature) are associated with successful engagement and PA promotion among college students in real-world settings.

XSL•FO
RenderX

Objectives

This study aimed to build upon prior findings from clinical trials by retrospective observation of an mHealth PA promotion program incorporating both intrapersonal and interpersonal components, which was delivered with campus-wide outreach to students and faculty and staff. The first research question was how the intervention components were used and whether this use was different between students and faculty and staff, and the second question was whether such use was associated with successful retention and goal achievement in the program. We hypothesized that undergraduate students would have lower use of intervention components than graduate students and faculty and staff and that such use would be associated with successful retention and goal achievement in the program. This statistical isolation of specific components builds upon prior studies by providing insight into specific facilitators and barriers of the program, which could inform future refinement of interventions aimed at increasing PA in situations of physical isolation from peers and PA spaces.

Methods

Intervention Design

The American College of Sports Medicine Exercise is Medicine On Campus initiative recognizes approximately 150 higher educational institutions per year for campus PA initiatives. One of their gold-level programs, the Southern Connecticut State University (SCSU) MoveSpring Steps Challenge, was suited to our study purpose because it used mHealth delivery with backend data capture. The university initiated the challenge in response to COVID-19 pandemic policies that restricted use of their fitness and recreational facilities. It operated from April 1 to 30, 2021, and was directed by the campus fitness center, incorporating input from the campus exercise physiology club and health center. It was digitally supported by the MoveSpring mobile app platform. This platform provides customizable fitness tracking and interaction features for individual companies and organizations. SCSU and MoveSpring collaborated to program the targeted intervention with both intra- and interpersonal components (Table 1). For the passive monitoring component, users could synchronize their own device such as a fitness watch or smartphone pedometer.

XSL•FO
RenderX

Table 1. Program components and content.

Type and component	Content
Intrapersonal	
Educational content on demand[a]	• Videos • Articles • Actionable health tips • Motivational quotes
Motivational content as push notifications	• Topic and mini challenge of the week (eg, benefits of nature walking or post a picture of your nature walk): 2-3 times per week • Campus events (eg, group walk): once per month • Prize reminders: 3 times per week • Achievement acknowledgment: once per week
Active monitoring (self-reported yes or no response)[b]	• Ate ≥3-4 servings of fruits and vegetables • Slept ≥7 hours • Drank ≥8 glasses of water
Passive monitoring (device)	• Steps (15,000 limit) in relation to both the personal goal[c] and tiered cutoffs for prize raffles[d] • Minutes of moderate to vigorous physical activity[e] • Distance
Interpersonal	
Community engagement and peer support	• Peer leaders and fitness center director • Chat box[f] • Friend activity (average daily steps of friends) • Team-specific challenges[g] • Team leaderboard (top 4 teams' avatar, name, and average daily steps per member)

[a]Content from MoveSpring proprietary collection and the Southern Connecticut State University student leaders related to physical activity, diet, sleep, hydration, and mindfulness.

[b]Completing ≥67% (20/30) of days ensured entry into a US $50 gift card raffle.

[c]Personal goal options were 6500, 10,000, or 12,500 steps per day.

[d]Average values of 7000, 10,000, or 12,500 steps per day ensured entry into a raffle for US $50, US $75, or US $100, respectively. One entry was awarded for days 1 to 15 and 2 entries were awarded for days 16 to 30.

[e]Meeting ≥20 min on 67% (20/30) of days ensured entry into a US $50 gift card raffle.

[f]Chat topics were wide ranging and sometimes unrelated to physical activity. This feature was only used by a few individuals, whereas the features in the subsequent 3 bullets were used by a majority of the 335 participants.

[g]The top 3 teams by average step count per member received T-shirts, grouped by the academic status of users who founded the team. Participants could actively form their own teams. Those not doing so were randomly assigned to a team in the first half of the first week of the challenge.

Recruitment and Enrollment

Nine days before the start of the challenge (March 22, 2021), the campus fitness center coordinator sent a single advertisement email to all SCSU students and employees with instructions guiding those interested to download the SCSU-customized MoveSpring SCSU app to their personal smartphone, create an account, and synchronize a watch or pedometer for passive monitoring. Upon registration, all participants were required to complete an intake survey of academic status, campus visit frequency, reason for enrolling, daily step goal, and device used for activity tracking (Multimedia Appendix 1). To minimize the onboarding burden, there was no query on sex or other demographics.

Data Capture

The SCSU Fitness Center and MoveSpring removed identifiers and donated the dataset to the Dataverse repository [44] and a research team based at Yale University (GIA, SSM, MA, BB, JL, SL, XX, SJ, LMF, MS-K, and MBG), who performed data cleaning, analysis, interpretation, and write-up. They liaised with several co-designers of the Steps Challenge program content (ADH, PB, and RSA) to summarize the intervention methodology for this manuscript.

Ethical Considerations

No researchers had contact with participants or access to identifiable information, so the Yale University Institutional Review Board determined that the project did not constitute human participants research, meaning the institution waived ethical approval (determination 2000033852). This manuscript follows STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for cohort studies (Multimedia Appendix 2). Participant identifiers were removed before the transfer of the data to Dataverse and Yale.

XSL•FO
RenderX

Data Analysis

Overview

This study was a retrospective analysis of longitudinal observations on campus community members over the 30 days they were exposed to a steps challenge intervention app. Analyses used SPSS (version 26.0; IBM Corp; aim 1) and SAS (version 9.4; aim 2). Significance was set at .05. Study size was based on the available dataset.

Outcomes

The first objective was to determine how the app was used and whether app use differed by academic status. The outcomes for this first objective were app use metrics, operationalized as intrapersonal (average frequency of opening the app, education, and self-monitoring as continuous variables) and interpersonal features (friends made in-app, team affiliation, and team size as discrete variables). The second objective was to determine whether app use was associated with successful retention and goal achievement in the program. The outcomes for this second objective were (1) days of retention quantified as a time-to-event discrete variable censored at the end of the 30-day exposure (ie, the event being the discontinuation of app syncing) and (2) each day's success at meeting the step goal, quantified as a ternary categorical variable (tracking device not worn, tracking device worn but not reaching goal, tracking device worn, and reaching goal). The first objective analyzed outcomes at the person level (ie, daily averages), while the second objective analyzed outcomes at the day level adjusted for within-person correlations. These units were chosen based on the available granularity of the data from the app that both delivered the intervention and tracked its outcomes.

Research Question 1: How Was the App Used, and Did App Use Differ by Academic Status?

Participant characteristics (Table 2) were tabulated as frequencies and compared by academic status using chi-square tests (χ^2). The duration of app use before abandonment (Table 3) was compared by academic status using the Kaplan-Meier estimate of survival probability because it was right-censored at the 30-day length of the program. Most metrics of app use before abandonment (Table 3) were compared by academic status using Kruskal-Wallace (H) tests as they were interval scale variables with distributions that were bimodal (number of teammates) or right skewed (frequency of opening the app before abandonment, friends made in-app, and use of the self-monitoring and education features) and visually similar between the academic statuses. Two metrics of app use before abandonment were even more heavily right skewed—use of the self-monitoring and education features—so they were recoded as binary (<1 time weekly or ≥1 time weekly) and compared by academic status using chi-square tests. All variables with a significant main effect across the 3 academic groups were further tested by post hoc pairwise comparisons with *P* value multiplied by 3 to Bonferroni-adjust for multiple comparisons.

XSL•FO
RenderX

Table 2. Participant descriptive characteristics.

	Overall	Undergraduates	Graduates	Faculty and staff	Chi-square (*df*)	*P* value
Participants, n	339	156	57	126	N/A[a]	N/A
Campus visit frequency, n (%)						
Remote	82 (24)	34 (22)	32 (56)	16 (13)	71.9 (4)	*<.001*[b]
Regularly visit	102 (30)	61 (39)	18 (32)	23 (18)		
Live or work	155 (46)	61 (39)	7 (12)	87 (69)		
Step count mechanism, n (%)					—[c]	—
Apple	207 (61)	110 (71)	41 (72)	56 (44)		
Fitbit	63 (19)	22 (14)	12 (21)	29 (23)		
Garmin	14 (4)	4 (3)	1 (2)	9 (7)		
Withings	1 (0.3)	0 (0)	0 (0)	14 (11)		
Google	25 (7)	11 (7)	0 (0)	1 (1)		
Manual entry	21 (6)	5 (3)	2 (4)	14 (11)		
Dropped out before syncing	8 (2)	4 (3)	1 (2)	3 (2)		
Motive for joining the program, n (%)					2.5 (4)	.65
Start physical activity	68 (20)	31 (20)	10 (18)	27 (21)		
Increase physical activity	102 (30)	44 (28)	22 (39)	36 (29)		
Maintain physical activity	169 (50)	81 (52)	25 (44)	63 (50)		
Step goal, n (%)					0.8 (4)	.94
6500	176 (52)	82 (53)	30 (53)	64 (51)		
10,000	132 (39)	62 (40)	21 (37)	49 (39)		
12,500	31 (9)	12 (8)	6 (11)	13 (10)		

[a]N/A: not applicable.

[b]Italicized values indicate *P*<.05.

[c]Not compared across groups due to small cell sizes.

XSL•FO
RenderX

Table 3. Use of app by academic status.

	Overall	Undergraduates	Graduates	Faculty and staff	Test statistic	*P* value
Participants, n	335	154	56	125	—[a]	—
Overall use						
Days of app use before abandonment, n (%)					12.6 (*df*=2)[b]	.002[c]
>7	313 (92)	145 (93)	51 (90)	117 (93)		
>14	291 (86)	126 (81)	49 (86)	116 (92)		
>21	260 (77)	110 (71)	46 (81)	104 (83)		
>28	226 (67)	88 (56)	42 (74)	96 (76)		
30	214 (63)	82 (53)	39 (68)	93 (74)		
App openings per day before abandonment, median (IQR)	1.1 (0.5-2.3)	0.8 (0.4-1.7)[d]	1.4 (0.7-2.7)	1.3 (0.7-2.7)	14.5 (*df*=2)[e]	<.001[c]
≥1 time weekly, n (%)	304 (91)	138 (90)	54 (96)	112 (90)		
≥1 time daily, n (%)	190 (57)	71 (46)	39 (70)	80 (64)		
≥2 times daily, n (%)	95 (28)	32 (21)	21 (38)	42 (34)		
Friends made, median (IQR)	2 (0-3)	2 (0-3)	1 (0-4)	2 (0-3)	0.5	.77
≥1, n (%)	211 (63)	96 (62)	35 (63)	80 (64)		
≥2, n (%)	168 (50)	77 (50)	27 (48)	64 (51)		
≥3, n (%)	116 (35)	49 (32)	18 (32)	49 (39)		
≥6, n (%)	36 (11)	14 (9)	5 (9)	17 (14)		
Teammates made, median (IQR)	9 (6-10)	9 (6-10)	9 (5-10)	8 (6-10)	0.1	.93
1, n (%)	1 (1)	0 (0)	1 (1)	0 (0)		
3-5, n (%)	70 (21)	31 (20)	15 (27)	24 (19)		
6-9, n (%)	132 (39)	62 (40)	15 (27)	55 (44)		
≥10, n (%)	132 (39)	61 (40)	25 (45)	46 (37)		
Sleep logged[f], n (%)	54 (16)	23 (15)	15 (27)	16 (13)	5.9 (*df*=2)[g]	.053
Diet logged[f], n (%)	40 (12)	18 (12)	9 (16)	13 (10)	1.2 (*df*=2)[g]	.55
Hydration logged[f], n (%)	62 (19)	28 (18)	15 (27)	19 (15)	3.5 (*df*=2)[g]	.18
Education viewed[f], n (%)	9 (3)	4 (3)	3 (5)	2 (2)	2.1 (*df*=2)[g]	.35

[a]Not applicable.

[b]Chi-square value of the Kaplan-Meier log rank.

[c]Italicized values indicate *P*<.05.

[d]Pairwise results: undergraduate was lower than graduate (H_1=43.6, adjusted *P*=.01) and faculty (H_1=35.9, adjusted *P*=.006). Graduate and faculty were not different (H_1=0.5, adjusted *P*>.99).

[e]H value of the Kruskal-Wallace test.

[f]Recoded as binary (<1 time weekly or ≥1 time weekly) due to substantial right-skew.

[g]Chi-square value of binary comparison.

Research Question 2: Was App Use Associated With Successful Retention and Goal Achievement in the Program?

Independent variables were app use metrics (engagement frequency, number of friends made, and number of teammates made) and all available participant characteristics (academic status, campus visit frequency, motive for joining the program, and step count goal). Each of the following analyses tested each independent variable's bivariate association with the retention or goal achievement outcome; those meeting a *P*<.10 cutoff were tested in a multiple variable model, and those appearing to influence that model (*P*<.20) were retained in the final parsimonious model.

Retention (ie, duration until last time the app was opened before abandonment) was estimated using the Kaplan-Meier estimate, and effects of independent variables were examined using a Cox frailty proportional hazard model incorporating correlation

XSL•FO
RenderX

within teams. Besides accounting for the right censorship of the data, these methods can adjust for possible correlations within the dataset; in this case, the correlation between teammates. The lone influential independent variable was visualized using a survival curve (Figure 1), and correlation within teammates was visualized as a bubble plot (Figure 2).

Figure 1. Retention survival curve according to engagements per day, which was the only predictor of retention retained in the parsimonious final model.

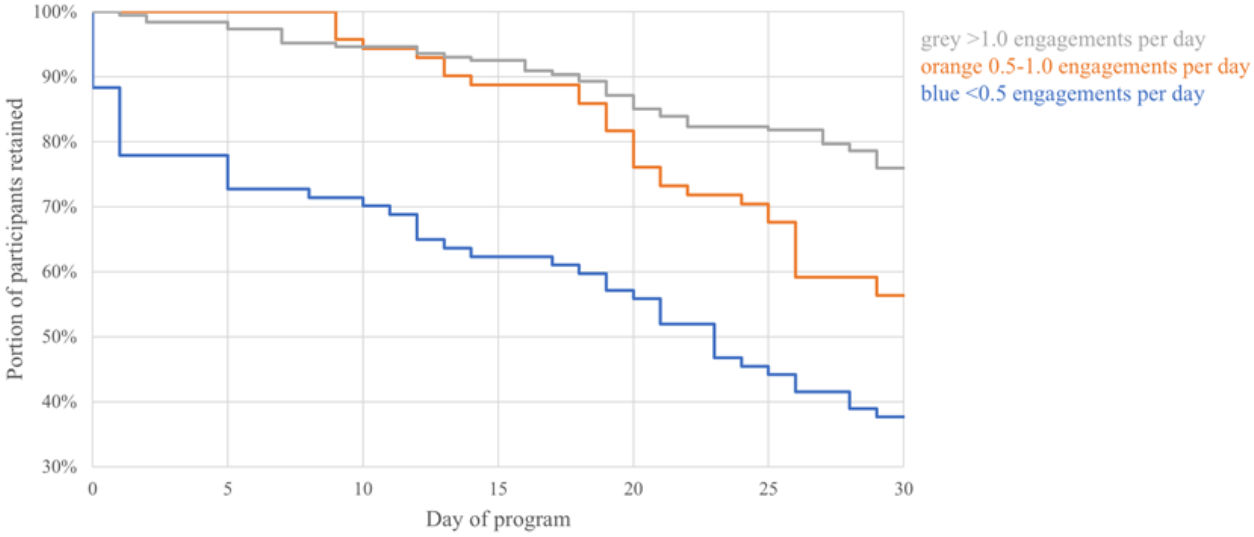
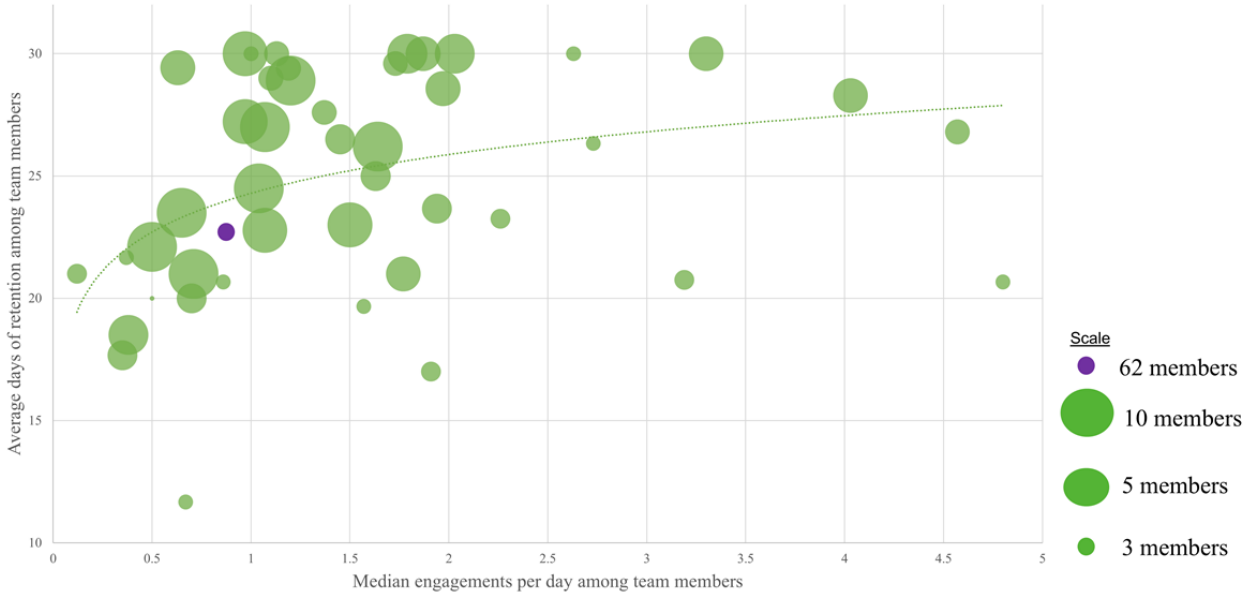


Figure 2. Retention (ie, duration of app use before abandonment) was clustered by team affiliation (intracluster correlation coefficient 0.064, 95% CI 0.001-0.164; *P*=.05). Therefore, we plotted average retention among team members in relation to team characteristics, including team size (bubble size) and median app engagements per day among team members (x-axis). It is visually apparent that the longer-retained individuals were those on teams with a larger size and higher median rate of app engagement. The dotted curve is for visual purposes and approximates retention days as 2.29 × ln(median engagement of team) + 24.29.



Goal achievement was labeled for each person-day as device nonwear (0 steps), wear-below-goal (steps nonzero but below goal), or goal (steps at or above goal). We tested independent variables associated with nonwear versus goal and wear-below-goal versus goal, using a generalized estimating equation that incorporated correlation within participants by a working correlation structure, a random intercept, and team affiliation as a random effect (Table 4). We also ran a supplemental analysis where the outcome variable was the total steps per day among participants meeting the standard minimum representative PA sampling of 4 days [45] (Multimedia Appendix 3).

XSL•FO
RenderX

Table 4. Generalized estimating equation of day-level step-goal achievement among 8287 daily observations occurring before app abandonment.

	Below goal (*k*=4659) vs at goal (*k*=3548)		Nonwear (*k*=80) vs at goal (*k*=3548)	
	Adjusted (SE)[a,b,c]	*P* value	Adjusted (SE)[a,b,c]	*P* value
Baseline characteristics				
Academic status				
Undergraduate student	.35 (0.16)	*.03*[d]	.82 (0.53)	.12
Graduate student	.50 (0.22)	*.02*[d]	−.21 (0.76)	.78
Faculty and staff (reference group)	—[e]	—	—	—
Motive for joining				
Maintain physical activity	−.99 (0.21)	*<.001*[d]	−1.69 (0.70)	*.02*[a]
Increase physical activity	−.63 (0.22)	*.01*[d]	−1.70 (0.70)	*.02*[a]
Start physical activity (reference group)	—	—	—	—
Goal				
12,500	.38 (0.27)	.15	.35 (1.05)	.74
10,000	.35 (0.16)	*.03*[d]	.80 (0.48)	.09
6500 (reference group)	—	—	—	—
Use of app features				
App engagements per d before abandonment				
>1.00	−.72 (0.21)	*.001*[d]	−.81 (0.61)	.19
0.50 to approximately 1.00	−.36 (0.23)	.12	−.60 (0.60)	.31
<0.50 (reference group)	—	—	—	—
Teammates made				
≥9	−.41 (0.27)	.13	−1.08 (0.76)	.16
5-8	−.37 (0.27)	.17	−1.40 (0.81)	.09
0-4 (reference group)	—	—	—	—
Friends made				
≥4	−.40 (0.20)	*.04*[d]	−.69 (0.88)	.44
1-3	−.33 (0.18)	.07	−.24 (0.46)	.60
0 (reference group)	—	—	—	—

[a]Positive coefficient indicates a greater likelihood of "below goal" or "not-wear."

[b]Within-subject correlations were incorporated using a working correlation structure.

[c]The "team affiliation" random effect was ignorable because its estimated variance was very small compared with the variance of the random intercept.

[d]Italicized values indicate *P*<.05.

[e]Not applicable.

Results

Enrollment

Overview

For the 2020 to 2021 school year, there were 11,055 individuals affiliated with SCSU (7047/11,055, 63.74% female, 6287/11,055, 56.87% non-Hispanic White) as either an undergraduate student (7440/11,055, 67.3%), graduate student (1891/11,055, 17.11%), or faculty and staff member (1724/11,055, 15.59%; Multimedia Appendix 4) [46]. Among these individuals, all of whom received a recruitment invitation email, 339 (3.1%) respondents registered for the steps challenge. A total of 156 (2.1%) undergraduate students, 57 (3%) graduate students, and 126 (7.3%) faculty and staff members enrolled in the program relative to the total SCSU population. All 339 enrolled individuals were included in the sample description (Table 2), although 4 were excluded from the subsequent analysis owing to missing data on the MoveSpring backend.

Campus Visit Frequency

Most undergraduate students (122/156, 78.2%) frequently visited campus, of whom about half reported also residing on campus.

XSL•FO
RenderX

Conversely, only 44% (25/57) of graduate students frequently visited campus, of whom the majority reported residing off campus. Nearly all faculty and staff (110/126, 87.3%) worked on campus but lived off campus as the university did not have faculty and staff housing.

Mechanism to Sync Step Count

Most participants (284/339, 83.8%) synchronized a device brand that most studies in a quantitative systematic review found to meet Consumer Technology Association validity standards (<10% mean absolute percent error; ie, Apple, Fitbit, or Garmin) [47]. Some (25/339, 7.5%) used a Google product (eg, Android phone step counter) to collect step count data, with faculty and staff members using this feature more than undergraduate and graduate students. In total, 6.2% (21/339) of participants manually entered step count, but it was unclear how the data were collected; 2.4% (8/339) of participants did not synchronize a device after sign-up and thus did not report any step count data.

Motives and Goals

Among possible motives for program enrollment, 20.1% (68/339) enrolled to initiate PA, 30.1% (102/339) to increase PA, and 49.8% (169/339) for PA maintenance. Among these 169 "maintainers," step goals were as follows: 75 (44.4%) chose 6500 steps, 76 (45%) chose 10,000 steps, and 18 (10.7%) chose 12,500 steps. These proportions were similar to the 102 "increasers:" (n=55, 53.9%, 6500 steps; n=39, 38.2%, 10,000 steps; and n=8, 8%, 12,500 steps; χ^2_2=2.4, *P*=.30) but higher than the 68 "starters" (n=46, 68%, 6500 steps; n=17, 25%, 10,000 steps; and n=5, 7%, 12,500 steps; χ^2_2=10.6; *P*=.005).

Research Question 1: Use of the App Overall and by Academic Status

Duration Before Abandonment

Most (313/335, 93.4%) participants stayed in the program for the first 7 days, and 63.9% (214/335) stayed for the full 30 days. At all time points beyond the first week, retention among undergraduate students tended to be lower than among graduate students and was lower than among faculty and staff (Table 3).

Frequency Before Abandonment

Most participants opened the app with some regular frequency—at least 0.5 times per day—but undergraduates did so less frequently than both faculty and staff and graduate students. Lower values of this metric were associated with both sooner abandonment and fewer step counts logged before abandonment.

Self-Monitoring and Education Features

In addition to regularly opening the app as described earlier, almost all participants (331/335, 98.8%) used at least one of the educational or self-monitoring features at least once to collect data about sleep, diet, and hydration. At the same time, the proportions using them at least weekly were much smaller: accessing education content (9/335, 2.7%), recording sleep data (54/335, 16.1%), recording diet (40/335, 11.9%), and recording hydration (62/335, 18.5%). Use of these features did not differ by academic status, although sleep logging had a nonsignificant tendency to be more common among graduate students than among undergraduates or faculty and staff.

Social Friends and Teams

Most participants made at least 1 friend in the app, and half made multiple friends. All participants were part of a team, which ranged in size from 2 to 62 members. Most teams comprised ≥6 individuals. Faculty and staff members were most often part of a team of 6 to 9 participants, and graduate students were most often part of a team of ≥10 participants. Most undergraduate participants were split between being in a team of 6 to 9 and a team of ≥10 participants. We did not detect statistically significant differences between the number of friends made or the number of teammates by academic status.

Research Question 2: Association of App Use With Successful Retention and Goal Achievement in the Program

Variables Associated With Retention

The final model of variables associated with retention (ie, duration of use before abandonment) included app engagement frequency (Figure 1) with adjustment for correlation within teams. Overall frailty model hazard ratios for abandonment were 0.56, 95% CI 0.43-0.72; *P*<.001 for those in the first and second quartile of engagement frequency (≥1.0/d), and 0.69, 95% CI 0.50-0.96; *P*=.03) for those in the third quartile of engagement frequency (0.5-1.0/d), compared with those in the fourth quartile of engagement frequency (<0.5/d). At 10 days, abandonment was highest among those in the fourth quartile compared with the 3 other quartiles (30% vs 5%). At 30 days, abandonment was again highest among those in the fourth quartile (62%), followed by those in the third quartile (44%) and those in the first and second quartiles (24%).

The intraclass correlation coefficient by teams was 0.064 (0.001-0.164); *P*=.05. Specifically, the longer-retained individuals were those on the teams with larger size and higher median rate of app engagement (Figure 2).

Variables Associated With Step-Goal Achievements

Among the 10,050 participant-days, 1763 occurred after app abandonment. Of the remaining 8287 participant-days, there were 3550 where steps meeting or exceeding goal amount were recorded, 4670 where steps exceeding a reasonable daily value (>250) but not meeting goal amount were recorded, and 80 days (among 28 participants) where no steps were recorded indicating nonwear of the device. On days the device was worn, the median step count was 6756 (IQR 4649-9793).

Both failing to meet the step goal and failing to wear the watch were less likely for individuals whose initial motive was maintaining or increasing PA, compared with those whose initial motive was to start PA (Table 4). Failing to meet the step goal was also less likely for individuals who engaged most frequently with the app and made a high number of friends. Finally, failing to meet the step goal was less likely for faculty and staff compared with students. When this analysis of meeting the step goal was repeated on average step counts, the associated

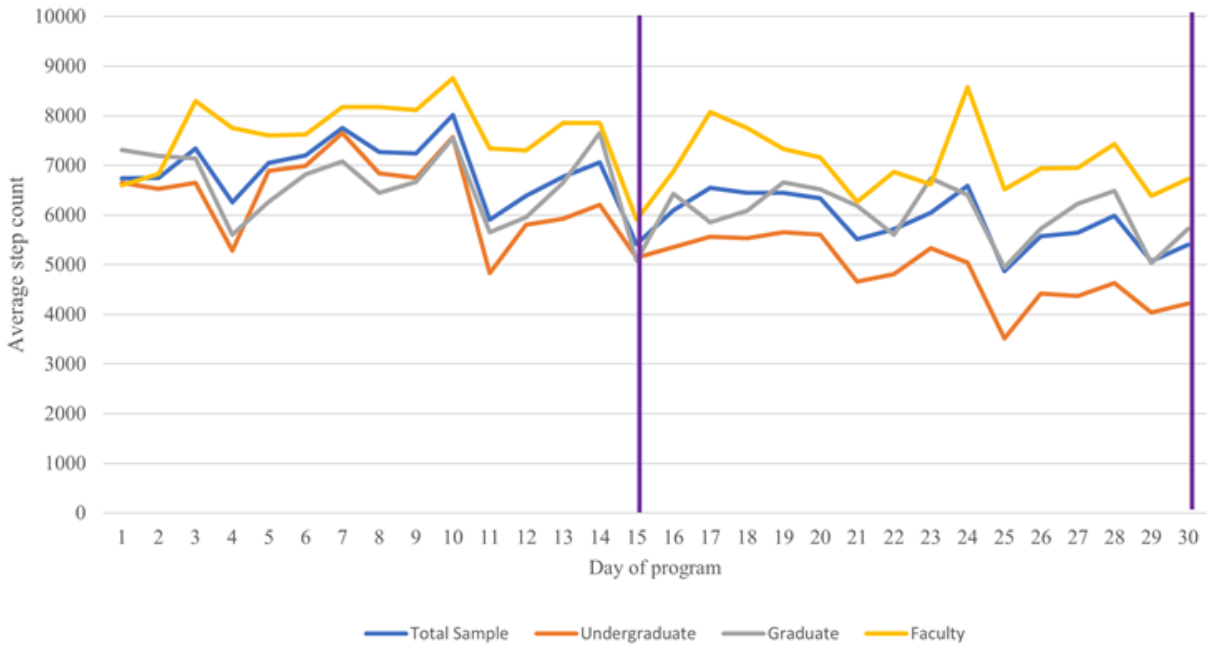
XSL•FO
RenderX

independent variables were the same (Multimedia Appendix 2).

Exploratory Analysis: Role of Gift Card Incentives as Possible Confounder

The escalation of incentive for extrinsic goals in the second half compared with the first half (ie, 2 raffle entries instead of 1) was not associated with a positive change in attaining these goals: among the 182 who did not meet the 7000 steps threshold in the first half, only 11 (6%) met it in the second half, whereas among the 157 who met the 7000 steps threshold in the first half, 36 (23%) did not meet it in the second half. A similar 4- to 10-fold smaller preponderance of upward versus downward transitioners was apparent at the step cutoffs of 10,000 (17/247, 6.9% vs 20/75, 27%) and 12,500 (8/304, 2.6% vs 8/27, 30%). In addition, visually, there was no indication that participants made a "sprint finish" as they approached day 15 or day 30 to qualify for the entries of that period (Figure 3).

Figure 3. Temporal trends in steps per day relative to the cutoffs for the prize drawings (purple).



The incentive of a drawing for completing self-reports on ≥67% (20/30) of days also appeared uninfluential as no participants met this goal. The incentive of a drawing for achieving 20 minutes of moderate to vigorous activity on ≥67% (20/30) of days could not be evaluated for influence because the database did not contain this metric on a day level. The incentive of receiving a T-shirt for being a top 3 team was not separable from the intrinsic motivation associated with being on a successful team.

Discussion

Principal Findings

We assessed intrapersonal and interpersonal factors to better understand facilitators and barriers to college community–based PA support at a northeast public university during the COVID-19 pandemic. For intrapersonal factors, most participants engaged solitarily with the app at least several times per week, and this frequency was positively associated with both retention and step-goal achievement while retained. College students, when compared with faculty and staff, (1) used the app less frequently; (2) used the app for a shorter duration before abandonment; and (3) were less successful in step count goal achievement. For interpersonal factors, students and faculty and staff equally used the open friends forum and the ongoing teams feature, which were positively associated with step-goal achievement and retention before abandonment, respectively. In fact, retention before abandonment clustered among teammates (Figure 2) with a similar intracluster correlation coefficient as various PA metrics cluster among high school classmates [48]. Overall, it is important to note this interaction between intrapersonal (ie, use of app) and interpersonal (ie, social support) characteristics.

College students are vulnerable to establishing a lifelong trajectory of low PA levels. Findings from our study suggest that expanding interventions beyond self-monitoring to a multimodal intervention including interpersonal components may help address engagement and retention. The multicomponent mHealth PA promotion program provided insight into enrollment, use of specific app features, retention, device wear, and achievement of step goals in a real-world setting of a public university during pandemic circumstances, which had presented barriers to PA. Furthermore, the sample was reflective of US 4-year public university averages with respect to the proportion of students living on campus (2.1 million/5.8 million, 37%) [49], studying remotely (2.1 million/9.1 million, 23%) [50], and choosing Apple tracking devices (567/768, 73.8%) [51]. This situation presented an opportunity to retrospectively observe the intervention in the

XSL•FO
RenderX

naturalistic college community environment with individuals having varying degrees and types of participation.

Comparison With Prior Work

Teams were included in this program because small-group team challenges within a fitness program improve individual challenge completion and the degree of participation [52]. Tong et al [29] and Edney et al [30] found that ongoing team relationships were a relatively popular aspect of such an intervention among college students, more so than open forum use. Similarly, we found that most of the social interaction on our app occurred in the "teams" rather than the "friends" feature. Our additional multiple variable and multilevel (person and team) analyses revealed that differential use of both intra- and interpersonal components critically modified outcomes in step counting and retention. This finding builds upon 2 previous studies that revealed correlations between intrapersonal component use (goal setting and content completion) and increased PA [53,54]. There is a need for further research with additional methodologies to provide further insight into how we could better deploy engagement strategies such as "teams" or "open forum friends" features to promote PA.

The SCSU-customized MoveSpring program was highly accessible to the college population because participants were able to participate from anywhere. Accessibility during this time was crucial due to COVID-19 pandemic restrictions creating a very inaccessible environment for PA. The steps challenge enabled participants to stay engaged within a social network, despite being physically isolated from their friends, family, and prior PA partners during the COVID-19 pandemic. Individuals were able to join teams with other individuals they may not have known previously. This virtual connection allowed individuals to immediately have at least one other individual on their campus also participating in the same activity.

However, because using an intensive recruitment and retention strategy would have been challenging during the pandemic, the program adopted a more pragmatic approach to implementation, which revealed a more true-to-life scenario. Naturalistic studies provide an opportunity to observe patterns in college students' PA patterns. In addition to examining participant results from an individual point of view, including teams in the intervention provided additional information about the potential of virtual teams in facilitating program engagement and retention, even in situations where participants are separated from each other and common PA spaces (eg, rural areas, satellite campuses, web-based degree programs). As college students can be a hard-to-engage population, having this knowledge, as well as understanding their PA habits can better inform future implementation and intervention efforts.

First, intervention planners should consider the added value of web-based interaction within an educational community and its ability to withstand barriers of geography and cost. The SCSU-customized MoveSpring program was accessible to those not physically on campus, incurred zero cost of devices (everyone used their own device), and had peer uptake and interactions that were mostly automatically facilitated by the app's team features. The minimal human facilitation required was achievable by student leaders and the fitness center director without extra staffing required in a minimal resource environment.

Second, these findings suggest that college students could benefit from a more tailored implementation strategy. For example, retention prompts could be targeted at those with the lowest overall engagement early in the program, and at those with medium engagement after 3 weeks of the program, in accord with when these respective groups fell below the highest engaged group in retention (Figure 2). Such anticipation of vulnerability to abandonment has been done elsewhere [55]. In addition, these prompts could include team reassignment to teams with high engagement to leverage the clustering of engagement by team affiliation. This strategy aligns with recent recommendations that interventions should be examined not only for their overall effects but also for the contribution of individual components to determine the active versus unnecessary ingredients and achieve tailoring to individuals and contexts [43]. Further tailoring could be attained by a machine learning prediction algorithm to trigger just-in-time adaptive intervention prompts.

Limitations and Strengths

This study had several limitations. First, although the program was offered campus-wide to recruit a representative sample, only 3.1% of the campus participated, which could lead to biased results. However, this representation is higher than the typical clinical trial enrollment of a population where often <3% even click the advertisement and <1% can commit to the logistical burdens of participation [56]. Nonetheless, future programs could make more extensive recruitment efforts to increase representation. Second, because it was a retrospective evaluation of a real-world program rather than a human subjects research study, we did not have survey assessments to test theoretical mechanisms (eg, self-efficacy and outcome expectations to test the SCT) nor participant demographics other than identification as undergraduate, graduate, or faculty and staff. Collecting more demographic and theoretical construct data would help target key program factors and their relationship with PA outcomes, accounting for theoretical mechanisms and diversity of participants. However, in young adulthood, the stage of independence or education is a more meaningful metric than biological age [57]. Third, also to reduce the burden of sign-up, participants used their own step-counting device. Almost all were previously validated up to consumer standards, but they nonetheless varied in brand and likely in model. However, the study outcome was behavioral goal achievement rather than a surrogate of the ground truth step [2]. Fourth, the program had a relatively short duration. Although, in our study and studies conducted by Lattie et al [14] and Adu et al [58] the first month of app use was critical regarding emergent differences between individuals in retention and degree of participation, so it was an appropriate timeframe for our research question. Finally, we isolated the inter- versus intrapersonal component comparison by controlling for within-person correlations statistically and within-environment correlations by using a single site, which held a gold-level designation. Nonetheless, it could also be useful to run a between-site design in which an alternate "control" site offered fewer of the components. It might also be useful to set up a design isolating intrinsic from extrinsic (ie,

XSL•FO
RenderX

prize incentives) motivation, but our brief analysis of the latter indicated a lack of influence.

While prior clinical trials evaluated interpersonal college community–based PA support [27-30,32], a unique strength of our study was addressing its real-world potential for retention, engagement, and association with step-goal achievement. Moreover, we ran this evaluation at the level of individual intervention components across the spectrum from intra to interpersonal. Doing so required detailed statistical analysis including survival curves and models including multiple variables and levels (person and team). This analysis was a critical advancement from prior studies that tested such associations only univariately and only for intrapersonal not interpersonal components [53,54]. Finally, the study was strengthened by its timing: it occurred during the COVID-19 pandemic before vaccine rollout to young adults, which was a time of extensive barriers to both PA and social interactions. Moreover, it was the month before the university's final examinations when college mHealth interventions were vulnerable to abandonment [24]. Thus, the results will inform guidance for coping with difficult times for PA and social interactions.

Future Directions

Clinical trials cannot address the real-world potential for retention and engagement by these applications. App disengagement and abandonment is prevalent in real-world settings according to survey data [59]. In direct assessments of interventions, while these metrics are challenging to standardize across studies, a recent systematic review found only 9% of 62 published studies considered their retention and engagement to be successful [60]. Retention was 10 times less likely if participants were not financially compensated and 4 times less likely if the app was for general lifestyle maintenance rather than for an existing clinical condition [60]. Although none of the included studies had an average age <30 years, other researchers have reported that college students are challenging to retain [23]. As such, despite the availability of pre-post outcomes on retained participants, there is a dearth of information regarding how to achieve retention and engagement, especially for preventive interventions for college students in real-world settings without financial compensation for participation. These retention issues also speak to concerns about the representativeness of the select group that is successfully recruited and retained. While concerns about representativeness apply to any research, the topic of social support approaches is particularly affected because it depends on the environment of these interactions; there is a need for open forums that represent the full university sociocultural environment rather than the subsample who met the criteria and commitment for research procedures.

We also considered that lack of goal achievement can result not only from lower-than-desired step counts but also from nonwear of the tracking device. In our analysis, these 2 occurrences were both associated with a motive for joining the program being the start rather than the increase or maintenance of PA. These findings suggest that the act of wearing and monitoring a fitness wearable has a motivational etiology, rather than simply being a methodological concern as it is usually treated in research contexts. Stricter study procedures with more controlled observation can sometimes cause participants to act differently than they would usually (ie, Hawthorne bias), potentially leading to different outcomes when implemented in a real-world setting. On the other hand, successful step-goal achievement on days the watch was worn was associated with variables that were not relevant to successful watch wear, including high engagement and a high number of friends. Thus, although having similar etiology, these 2 behaviors may require different interventions.

Finally, the intervention and analyses reported herein among several hundred participants from a single campus should be repeated at other campuses to test generalizability. Efforts should be made to engage campuses from the American College of Sports Medicine's Exercise is Medicine On Campus initiative lists and others for means that interventions can collect data and transfer it to research teams. The MoveSpring platform would be one excellent mechanism; many institutions offer it to faculty and staff but should consider offering it to students, as in this study.

Conclusions

Engagement with the program suggests longer retention and better PA outcomes, which were critically modified by the interpersonal aspects of this engagement. Adopting a pragmatic approach to implementation highlights opportunities for future trials and the need for specific strategies tailored to different groups to account for differences in demographics and baseline motives and goals.

Acknowledgments

The authors acknowledge support from the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health (R01AA031959) and from the AL Williams Professorship funds. In addition, author GIA was supported by American Heart Association (grant 852679); Robert E Lee and Clara Guthrie Patterson Trust Mentored Research Award, Bank of America, NA, Trustee; and the National Institute of Diabetes, Digestive, and Kidney Diseases of the National Institutes of Health under a mentored research scientist development award (K01DK129441). These funders played no role in study design, data procurement, analysis and interpretation of data, or the writing of this manuscript.

The Southern Connecticut State University Fitness Center and the MoveSpring app supported the study by donating data from their previously conducted program. They played no role in analysis and interpretation of data or the writing of this manuscript.

XSL•FO
RenderX

Data Availability

All data analyzed in this study are publicly available on Dataverse [44].

Authors' Contributions

GIA, SSM, MS-K, and MBG conceptualized the research design, aims, and methodology with support from LMF, MA, BB, JJL, SL, and XX. MBG and GIA acquired funding. ADH, POB, and RSA summarized intervention methodology. GIA curated the data. SJ conducted formal analysis and data visualization with support from GIA, MS-K, and MBG. GIA validated the analysis. GIA and SSM wrote the first draft of the manuscript with support from ADH, MA, and SJ. MS-K reviewed and edited the first draft of the manuscript. MBG, LMF, SJ, ADH, MA, POB, RSA, BB, JJL, SL, SL, and XX reviewed and edited a subsequent draft of the manuscript, and all authors approved the final manuscript.

Conflicts of Interest

GIA receives professional service fees from Calm.com [61] (nominal fee), Labfront (full fee), and GlucoseZone (full fee). LMF and GIA have a provisional patent filed for a digital system for lifestyle medicine (047162-5346-P1US) outside the submitted work. SSM, ADH, MA, POB, RSA, BB, JJL, SL, XX, SJ, MS-K, and MBG declare no potential conflicts of interest. The authors attest that Calm.com [61], Labfront, GlucoseZone, and a digital system for lifestyle medicine had no influence on the design of this study or its outcomes.

Multimedia Appendix 1
Intake survey.
[DOCX File , 17 KB - mhealth_v13i1e51707_app1.docx]

Multimedia Appendix 2
STROBE checklist.
[PDF File (Adobe PDF File), 124 KB - mhealth_v13i1e51707_app2.pdf]

Multimedia Appendix 3
Supplemental analysis.
[DOCX File , 18 KB - mhealth_v13i1e51707_app3.docx]

Multimedia Appendix 4
Campus demographics.
[DOCX File , 19 KB - mhealth_v13i1e51707_app4.docx]

References

1. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022 Mar;7(3):e219-e228 [FREE Full text] [doi: 10.1016/S2468-2667(21)00302-9] [Medline: 35247352]
2. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al. Association of step counts over time with the risk of chronic disease in the All of Us Research Program. Nat Med 2022 Nov;28(11):2301-2308 [FREE Full text] [doi: 10.1038/s41591-022-02012-w] [Medline: 36216933]
3. The healthy campus framework. American College Health Association. Silver Spring, MD: American College Health Association; 2023. URL: https://www.acha.org/wp-content/uploads/2024/06/The_Healthy_Campus_Framework.pdf [accessed 2025-06-02]
4. Coakley KE, Lardier DT, Holladay KR, Amorim FT, Zuhl MN. Physical activity behavior and mental health among university students during COVID-19 lockdown. Front Sports Act Living 2021 Jul 09;3:682175 [FREE Full text] [doi: 10.3389/fspor.2021.682175] [Medline: 34308346]
5. López-Valenciano A, Suárez-Iglesias D, Sanchez-Lastra MA, Ayán C. Impact of COVID-19 pandemic on university students' physical activity levels: an early systematic review. Front Psychol 2020 Jan 15;11:624567 [FREE Full text] [doi: 10.3389/fpsyg.2020.624567] [Medline: 33519653]
6. Rouvinen H, Turunen H, Lindfors P, Kinnunen JM, Rimpelä A, Koivusilta L, et al. Online health information-seeking behaviour and mental well-being among Finnish higher education students during COVID-19. Health Promot Int 2023 Dec 01;38(6):daad143 [FREE Full text] [doi: 10.1093/heapro/daad143] [Medline: 37935171]
7. Schäfer M, Stark B, Werner AM, Tibubos AN, Reichel JL, Pfirrmann D, et al. Health information seeking among university students before and during the Corona crisis-findings from Germany. Front Public Health 2020 Jan 25;8:616603 [FREE Full text] [doi: 10.3389/fpubh.2020.616603] [Medline: 33585388]

XSL•FO
RenderX

8. Wearable devices - ownership. YouGov. URL: https://today.yougov.com/topics/technology/trackers/wearable-devices-ownership?crossBreak=1829 [accessed 2025-06-02]
9. Gravett K, Winstone NE. Making connections: authenticity and alienation within students' relationships in higher education. High Educ Res Dev 2020 Nov 18;41(2):360-374 [FREE Full text] [doi: 10.1080/07294360.2020.1842335]
10. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol Bull 1995 May;117(3):497-529. [Medline: 7777651]
11. Otundo JO, MacGregor SK. Effect of situational interest and social support on college students' physical activity motivation: a mixed methods analysis. Phys Educ 2019 Mar 19;76(2):502-523. [doi: 10.18666/tpe-2019-v76-i2-8502]
12. Alshehri MA, Kruse-Diehr AJ, McDaniel J, Partridge JA, Null D. Impact of social support on the physical activity behaviors of international college students in the United States. Int J Exerc Sci 2021 Nov 01;14(5):1305-1319 [FREE Full text] [doi: 10.70252/OMTE6688] [Medline: 35096246]
13. Marler EK, Bruce MJ, Abaoud A, Henrichsen C, Suksatan W, Homvisetvongsa S, et al. The impact of COVID-19 on university students' academic motivation, social connection, and psychological well-being. Scholarsh Teach Learn Psychol 2024;10(3):320-330. [doi: 10.1037/stl0000294]
14. Lattie EG, Cohen KA, Hersch E, Williams KD, Kruzan KP, MacIver C, et al. Uptake and effectiveness of a self-guided mobile app platform for college student mental health. Internet Interv 2021 Dec 24;27:100493 [FREE Full text] [doi: 10.1016/j.invent.2021.100493] [Medline: 35036331]
15. Kim GS, Lee CY, Kim IS, Lee TH, Cho E, Lee H, et al. Dyadic effects of individual and friend on physical activity in college students. Public Health Nurs 2015;32(5):430-439. [doi: 10.1111/phn.12176] [Medline: 25565084]
16. Lieffers JR, Quintanilha M, Trottier CF, Johnson ST, Mota JF, Prado CM. Experiences with and perception of a web-based mindfulness, nutrition, and fitness platform reported by first-year university students: a qualitative study. J Acad Nutr Diet 2021 Dec;121(12):2409-18.e3. [doi: 10.1016/j.jand.2021.04.019] [Medline: 34119458]
17. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
18. Conner M, Norman P. Predicting Health Behaviour: Research and Practice with Social Cognition Models. London, UK: Open University Press; 1996.
19. Bao H, Lee EW. Examining theoretical frameworks and antecedents of health apps and wearables use: a scoping review. Health Commun 2024 Nov;39(12):2671-2681. [doi: 10.1080/10410236.2023.2283655] [Medline: 37968803]
20. Kang HS, Exworthy M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth 2022 Jul 13;10(7):e35684 [FREE Full text] [doi: 10.2196/35684] [Medline: 35830222]
21. Hartson KR, Della LJ, King KM, Liu S, Chatman JM, Rhodes RE. A web-based intervention for insufficiently active college students: feasibility and preliminary efficacy. West J Nurs Res 2024 Jul;46(7):541-551. [doi: 10.1177/01939459241253218] [Medline: 38767319]
22. Mo D, Xiang M, Luo M, Dong Y, Fang Y, Zhang S, et al. Using gamification and social incentives to increase physical activity and related social cognition among undergraduate students in Shanghai, China. Int J Environ Res Public Health 2019 Mar 08;16(5):858 [FREE Full text] [doi: 10.3390/ijerph16050858] [Medline: 30857261]
23. Fackler CA, Baugh N, Lovegren AA, Nemeroff C, Whatley Blum J. Technology-enhanced health promotion for college students: a seed development project. Nurs Rep 2021 Mar 01;11(1):143-151 [FREE Full text] [doi: 10.3390/nursrep11010014] [Medline: 34968319]
24. Krzyzanowski MC, Kizakevich PN, Duren-Winfield V, Eckhoff R, Hampton J, Blackman Carr LT, et al. Rams have heart, a mobile app tracking activity and fruit and vegetable consumption to support the cardiovascular health of college students: development and usability study. JMIR Mhealth Uhealth 2020 Aug 05;8(8):e15156 [FREE Full text] [doi: 10.2196/15156] [Medline: 32755883]
25. Melcher J, Camacho E, Lagan S, Torous J. College student engagement with mental health apps: analysis of barriers to sustained use. J Am Coll Health 2022;70(6):1819-1825. [doi: 10.1080/07448481.2020.1825225] [Medline: 33048626]
26. Wang X, Yang X, Juzaily Bin Mohd Nasiruddin N, Wei S, Dong D, Bin Samsudin S. Social support and physical activity in college and university students: a meta-analysis. Health Educ Behav 2024 Aug;51(4):533-543. [doi: 10.1177/10901981231216735] [Medline: 38305027]
27. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
28. Rote AE. Physical activity intervention using Fitbits in an introductory college health course. Health Educ J 2016 Nov 16;76(3):337-348. [doi: 10.1177/0017896916674505]
29. Tong HL, Coiera E, Tong W, Wang Y, Quiroz JC, Martin P, et al. Efficacy of a mobile social networking intervention in promoting physical activity: quasi-experimental study. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12181 [FREE Full text] [doi: 10.2196/12181] [Medline: 30920379]

XSL•FO
RenderX

30. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):e51-e62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
31. Wang M, Guo Y, Zhang Y, Xie S, Yu Z, Luo J, et al. Promoting healthy lifestyle in Chinese college students: evaluation of a social media-based intervention applying the RE-AIM framework. Eur J Clin Nutr 2021 Feb;75(2):335-344. [doi: 10.1038/s41430-020-0643-2] [Medline: 32366994]
32. Tong HL, Quiroz JC, Kocaballi AB, Ijaz K, Coiera E, Chow CK, et al. A personalized mobile app for physical activity: an experimental mixed-methods study. Digit Health 2022;8:20552076221115017 [FREE Full text] [doi: 10.1177/20552076221115017] [Medline: 35898287]
33. Nam SJ, Cha C. Effects of a social-media-based support on premenstrual syndrome and physical activity among female university students in South Korea. J Psychosom Obstet Gynecol 2019 Mar 04;41(1):47-53. [doi: 10.1080/0167482x.2018.1559811]
34. Rote AE, Klos LA, Brondino MJ, Harley AE, Swartz AM. The efficacy of a walking intervention using social media to increase physical activity: a randomized trial. J Phys Act Health 2015 Jun 16;12 Suppl 1:S18-S25 [FREE Full text] [doi: 10.1123/jpah.2014-0279] [Medline: 25599378]
35. Mandic D, Bjegovic-Mikanovic V, Vukovic D, Djikanovic B, Stamenkovic Z, Lalic NM. Successful promotion of physical activity among students of medicine through motivational interview and web-based intervention. PeerJ 2020 Jul 08;8:e9495 [FREE Full text] [doi: 10.7717/peerj.9495] [Medline: 32714663]
36. Todorovic J, Terzic-Supic Z, Djikanovic B, Nesic D, Piperac P, Stamenkovic Z. Can social media intervention improve physical activity of medical students? Public Health 2019 Sep;174:69-73. [doi: 10.1016/j.puhe.2019.05.030] [Medline: 31323599]
37. Zhang J, Brackbill D, Yang S, Centola D. Efficacy and causal mechanism of an online social media intervention to increase physical activity: results of a randomized controlled trial. Prev Med Rep 2015 Aug 13;2:651-657 [FREE Full text] [doi: 10.1016/j.pmedr.2015.08.005] [Medline: 26844132]
38. Sofro ZM, Wibowo RA, Wasityastuti W, Kusumadewi AF, Utomo PS, Ekawati FM, et al. Physical activity virtual intervention for improving mental health among university students during the COVID-19 pandemic: a co-creation process and evaluation using the behavior change wheel. Heliyon 2023 Aug 05;9(8):e18915 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e18915] [Medline: 37588605]
39. Marker AM, Staiano AE. Better together: outcomes of cooperation versus competition in social exergaming. Games Health J 2015 Feb;4(1):25-30 [FREE Full text] [doi: 10.1089/g4h.2014.0066] [Medline: 26181677]
40. Orji R. Why are persuasive strategies effective? Exploring the strengths and weaknesses of socially-oriented persuasive strategies. In: Proceedings of the 12th International Conference on Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviors. 2017 Presented at: PERSUASIVE 2017; April 4–6, 2017; Amsterdam, The Netherlands. [doi: 10.1007/978-3-319-55134-0_20]
41. Sultoni K, Peralta LR, Cotton W. Using a design-based research approach to develop a technology-supported physical education course to increase the physical activity levels of university students: study protocol paper. PLoS One 2022 Dec 01;17(12):e0269759 [FREE Full text] [doi: 10.1371/journal.pone.0269759] [Medline: 36454897]
42. Afshin A, Babalola D, Mclean M, Yu Z, Ma W, Chen CY, et al. Information technology and lifestyle: a systematic evaluation of internet and mobile interventions for improving diet, physical activity, obesity, tobacco, and alcohol use. J Am Heart Assoc 2016 Aug 31;5(9):e003058 [FREE Full text] [doi: 10.1161/JAHA.115.003058] [Medline: 27581172]
43. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland: Springer International Publishing; 2018.
44. Haughton A, Ash G. Southern Connecticut State University MoveSpring (TM) steps challenge. Harvard Dataverse. 2024. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IJNAVQ [accessed 2025-06-02]
45. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
46. Southern Connecticut State University. National Center for Education Statistics. 2020. URL: https://nces.ed.gov/ipeds/reported-data/130493?year=2020&surveyNumber=9 [accessed 2025-06-02]
47. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
48. Steenholt CB, Pisinger VS, Danquah IH, Tolstrup JS. School and class-level variations and patterns of physical activity: a multilevel analysis of Danish high school students. BMC Public Health 2018 Feb 14;18(1):255-266 [FREE Full text] [doi: 10.1186/s12889-018-5155-9] [Medline: 29444648]
49. National postsecondary student aid study: 2020 undergraduate students. National Center for Education Statistics. URL: https://nces.ed.gov/datalab/powerstats/157-national-postsecondary-student-aid-study-2020-undergraduate-students/percentage-distribution [accessed 2025-06-02]

XSL•FO
RenderX

50. Number and percentage of students enrolled in degree-granting postsecondary institutions, by distance education participation, location of student, level of enrollment, and control and level of institution: Fall 2021 and fall 2022. National Center for Education Statistics. 2023. URL: https://nces.ed.gov/programs/digest/d23/tables/dt23_311.15.asp [accessed 2025-06-02]
51. Wearables: Apple users in the United States. Statista. URL: https://tinyurl.com/3ysu4pj7 [accessed 2025-06-02]
52. Tullar JM, Walker TJ, Page MF, Taylor WC, Roman R, Amick BC. Evaluation of a worksite-based small group team challenge to increase physical activity. Am J Health Promot 2019 Feb;33(2):259-266 [FREE Full text] [doi: 10.1177/0890117118784229] [Medline: 29996662]
53. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surviv 2013 Sep;7(3):355-368 [FREE Full text] [doi: 10.1007/s11764-013-0279-5] [Medline: 23532799]
54. Looyestyn J, Kernot J, Boshoff K, Maher C. A web-based, social networking beginners' running intervention for adults aged 18 to 50 years delivered via a Facebook group: randomized controlled trial. J Med Internet Res 2018 Feb 26;20(2):e67 [FREE Full text] [doi: 10.2196/jmir.7862] [Medline: 29483065]
55. Bricker J, Miao Z, Mull K, Santiago-Torres M, Vock DM. Can a single variable predict early dropout from digital health interventions? Comparison of predictive models from two large randomized trials. J Med Internet Res 2023 Jan 20;25:e43629 [FREE Full text] [doi: 10.2196/43629] [Medline: 36662550]
56. Ellington M, Connelly J, Clayton P, Lorenzo CY, Collazo-Velazquez C, Trak-Fellermeier MA, et al. Use of Facebook, Instagram, and Twitter for recruiting healthy participants in nutrition-, physical activity-, or obesity-related studies: a systematic review. Am J Clin Nutr 2022 Feb 09;115(2):514-533 [FREE Full text] [doi: 10.1093/ajcn/nqab352] [Medline: 34669955]
57. Settersten RA, Ottusch TM, Schneider B. Becoming adult: meanings of markers to adulthood. In: Scott R, Kossyn S, editors. Emerging Trends in the Social and Behavioral Sciences. Hoboken, NJ: John Wiley & Sons; 2015.
58. Adu MD, Malabu UH, Malau-Aduli AE, Drovandi A, Malau-Aduli BS. User retention and engagement with a mobile app intervention to support self-management in Australians with type 1 or type 2 diabetes (My Care Hub): mixed methods study. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e17802 [FREE Full text] [doi: 10.2196/17802] [Medline: 32525491]
59. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221 [FREE Full text] [doi: 10.3390/healthcare10020221] [Medline: 35206837]
60. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
61. Calm homepage. Calm. URL: https://www.calm.com/ [accessed 2025-06-04]

Abbreviations

mHealth: mobile health
PA: physical activity
SCSU: Southern Connecticut State University
SCT: social cognitive theory
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by T de Azevedo Cardoso; submitted 16.08.23; peer-reviewed by P Gorp, B Tabuenca; comments to author 15.05.24; revised version received 24.09.24; accepted 09.04.25; published 16.06.25.

Please cite as:
Ash GI, Mak SS, Haughton AD, Augustine M, Bodurtha PO, Axtell RS, Borsari B, Liu JJ, Lou S, Xin X, Fucito LM, Jeon S, Stults-Kolehmainen M, Gerstein MB
College Community–Based Physical Activity Support at a Public University During the COVID-19 Pandemic: Retrospective Longitudinal Analysis of Intra- Versus Interpersonal Components for Uptake and Outcome Association
JMIR Mhealth Uhealth 2025;13:e51707
URL: https://mhealth.jmir.org/2025/1/e51707
doi:10.2196/51707
PMID:

©Garrett I Ash, Selene S Mak, Adrian D Haughton, Madilyn Augustine, Phillip O Bodurtha, Robert S Axtell, Beatrice Borsari, Jason J Liu, Shaoke Lou, Xin Xin, Lisa M Fucito, Sangchoon Jeon, Matthew Stults-Kolehmainen, Mark B Gerstein. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted

XSL•FO
RenderX

use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of Mobile Augmented Reality Counseling on Improving Shared Decision-Making in Thoracic Surgery: Randomized Clinical Crossover Trial

Ying-Shian Chen[1,2,3], MD; Yi-Chen Hsu[4,5], MD; Worachate Romalee[4,6], PhD; Ding-Han Wang[4,5,7], PhD; Jennifer Lai[8], MD; Tsai-Wang Huang[3], PhD; Kuan Hsun Lin[3], MD

[1]School of Medicine, National Defense Medical University, Taipei, Taiwan

[2]Taichung Armed Forces General Hospital, Taichung, Taiwan

[3]Division of Thoracic Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical University, Taipei, Taiwan

[4]College of Dentistry, National Yang Ming Chiao Tung University, Taipei, Taiwan

[5]Oral Medicine Innovation Center (OMIC), National Yang Ming Chiao Tung University, Taipei, Taiwan

[6]Department of Community Dentistry & Gerodontology, Faculty of Dentistry, Thammasat University, Pathum Thani, Thailand

[7]Research Center for Intelligent Sensor Device, National Yang Ming Chiao Tung University, Taipei, Taiwan

[8]Tri-Service General Hospital, National Defense Medical University, Taipei, Taiwan

Corresponding Author:
Kuan Hsun Lin, MD
Division of Thoracic Surgery, Department of Surgery, Tri-Service General Hospital
National Defense Medical University
No. 325, Sec. 2, Chenggong Rd., Neihu District
Taipei, 11490
Taiwan
Phone: 886 2 8792 3311 ext 88080
Email: dontchangela@gmail.com

Abstract

Background: Augmented reality (AR) superimposes virtual objects onto a real-world environment, allowing users to interact in real time. As AR has become widely used, its integration into smartphones or tablets has enabled mobile augmented reality (MAR) experiences. AR has been adopted in many industries, and the literature has highlighted its applications in academic and clinical settings, particularly in enhancing visualization, communication, and learning.

Objective: This study investigated the potential of MAR as a mobile health tool to enhance shared decision-making (SDM) in thoracic surgery by increasing patient understanding and engagement during medical consultations.

Methods: A randomized crossover clinical trial was conducted at the Tri-Service General Hospital in Taiwan. Participants scheduled for thoracic surgery were enrolled and randomized in a crossover design. The MAR intervention incorporated patient-specific 3D anatomical models that were reconstructed from computed tomography imaging to facilitate understanding and support SDM. The impact of each counseling approach on SDM was evaluated using postintervention questionnaires.

Results: A total of 47 participants were enrolled in this study. After analyzing the data, we found that patients in the MAR group showed significantly higher scores compared to those in the traditional counseling group (P<.001) during the SDM process. Moreover, patients reported higher satisfaction levels and found the visual objects helpful for understanding tumor location and surgical procedures.

Conclusions: This study demonstrated that MAR counseling significantly enhanced patients' comprehension of thoracic conditions and increased their active engagement in the SDM process (P<.001). The integration of patient-specific 3D anatomical models into MAR technology provided an intuitive method for critical medical information. This digital approach not only enhanced personalization in medical communication but also reinforced patient education about their health care conditions. These findings suggest that MAR counseling represents a promising approach for promoting patient-centered care in thoracic surgery and has potential applications across various clinical domains.

Trial Registration: ClinicalTrials.gov NCT07062393; https://clinicaltrials.gov/study/NCT07062393

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e79632) doi:10.2196/79632

KEYWORDS

mobile augmented reality; shared decision-making; patient education; thoracic surgery; 3D models

Introduction

Role of Mobile Augmented Reality in Clinical Counseling and Mobile Health

Augmented reality (AR) allows the seamless integration of virtual elements with real-world environments, including the superimposition of digital files, such as 3D models, animations, and videos. This enhances interactive learning, visualization, and communication in medical contexts [1]. A more accessible and scalable evolution of AR is mobile AR (MAR), which operates via smartphones and tablets [2]. Unlike AR headsets or smart glasses, MAR leverages everyday mobile devices to deliver immersive experiences, making it a cost-effective and broadly deployable solution in both clinical and remote settings.

In the medical field, AR has been increasingly applied in thoracic surgery for training, preoperative planning, and intraoperative assistance [3]. However, a noticeable gap exists in the literature regarding the use of MAR as a patient-facing mobile health (mHealth) tool to improve communication and support shared decision-making (SDM). By enabling real-time 3D visualization of patient-specific anatomy through mobile devices, MAR represents a novel strategy within the mHealth paradigm to promote accessibility, patient engagement, and autonomy in surgical counseling.

Surgical Counseling Challenges in Thoracic Surgery

Lung cancer is a global health challenge, as reflected in the 2020 world statistics. The incidence of lung and bronchus cancer was reported at 31.5 cases per 100,000 population for men and 14.6 cases per 100,000 population for women. The widespread prevalence of lung cancer continues to affect a significant portion of the global population [4].

In Taiwan, the situation is even more concerning. According to the latest report by Taiwan's Ministry of Health and Welfare in 2021, the rate of new cases of lung and bronchus cancer was alarmingly high, at 77.39 per 100,000 population for men and 67.13 per 100,000 population for women. Among the 16,880 newly diagnosed cases, a considerable portion (40.4%) received surgical treatment [5], which is one of the essential treatment options and the standard approach for early-stage disease.

The surgical approaches for lung resection could be divided into open thoracotomy and minimally invasive thoracic surgery (ie, video-assisted and robot-assisted thoracoscopic surgery). The various modes of lung resection, including anatomical (pneumonectomy, lobectomy, and segmentectomy) and nonanatomical (wedge resection) [6], pose a significant challenge to patient understanding. Traditional methods of patient counseling, relying on oral explanations, 2D computed tomography scans, and schematic diagrams, often fail to provide a comprehensive understanding. Although informative, these methods do not fully convey the intricacies of each surgical option, leaving patients inadequately equipped to make informed decisions.

This gap in understanding is critical, as patients need to comprehend various aspects of their condition, such as tumor location, its proximity to vital structures, and the specifics of different surgical procedures, each with its unique risks and implications. The limitations of current counseling methods highlight the necessity for innovative approaches that can enhance physician-patient communication, provide better visual support, and enable patients to engage more effectively in their treatment planning.

Communication and SDM

Several factors contribute to the quality of patient counseling and decision-making. These include patient anxiety, cultural background, insufficient medical knowledge, and awareness, all of which can significantly influence clinical decision outcomes [7,8]. Patients tend to participate more actively in treatment planning when they clearly understand their medical information [9]. Ineffective communication, especially when lacking meaningful patient engagement [10], may result in misunderstandings, dissatisfaction, disputes, and even violence against health care providers [11]. SDM has been increasingly promoted to emphasize patient preferences, values, and autonomy in the medical decision-making process [12]. It is now widely endorsed by academic and professional societies as a means to improve decision quality [13]. In addition, assistive tools—especially visual and interactive aids—have also been shown to support this process [14].

Study Objective

To address this unmet need, we developed a MAR-based counseling system that integrates patient-specific 3D visualization into the SDM process for thoracic surgery. This study introduces a novel application of mHealth aimed at bridging the gap between complex surgical information and patient understanding.

Methods

Study Design

This study was a single-center, 2-arm, randomized crossover clinical trial conducted at Tri-Service General Hospital, Taipei, Taiwan, between July 2022 and June 2023. The study compared traditional image-based counseling with MAR-enhanced counseling for patients undergoing thoracic surgery. Participants were randomized to 1 of 2 counseling sequences: traditional followed by MAR (group A) or MAR followed by traditional (group B). Each participant received both formats of counseling during a single encounter. To evaluate the impact of each method, participants completed self-administered questionnaires after each session to assess SDM and counseling preferences. No formal washout period was implemented due to the educational and noninterventional nature of the study.

XSL•FO
RenderX

Patient and Public Involvement

Patients or the public were not involved in the design, conduct, reporting, or dissemination of this study, as the research focused on the technical development and initial validation of a mobile-based counseling tool. Future iterations may incorporate patient feedback to optimize usability and relevance.

Sample Size Calculation

Sample size was calculated using G*Power (version 3.1.9.7; Heinrich Heine University Düsseldorf) to detect a significant difference between 2 independent groups using nonparametric analysis. The calculation was based on unpublished pilot data from our research team, involving 10 patients who received both MAR and traditional counseling. A large effect size (Cohen *d*=0.9) was observed in SDM score differences. Based on a two-tailed test with an alpha level of .05 and statistical power of 0.80, the required sample size was calculated to be 44 participants (22 per group). To accommodate for potential dropouts, 47 patients were ultimately enrolled.

Recruitment

Participants were recruited through outpatient and inpatient referrals from thoracic surgery clinics. Inclusion criteria were as follows: (1) scheduled for thoracic surgery (eg, lobectomy, segmentectomy, or wedge resection); (2) aged 20 to 65 years; (3) capable of using a smart device; (4) able to communicate in Mandarin or Taiwanese; (5) willing and able to provide informed consent; and (6) capable of completing questionnaires independently or with minimal assistance. Exclusion criteria were as follows: (1) emergency surgery, (2) severe cognitive impairment or psychiatric illness interfering with participation, (3) visual or hearing impairments interfering with mobile device interaction, (4) prior participation in similar counseling studies, and (5) inability to complete follow-up.

Randomization

Consenting participants were randomly allocated to either group A (traditional counseling first) or group B (MAR counseling first) using a basic random number generator by a research assistant without allocation concealment or stratification. Due to the nature of the interventions, blinding was not feasible in this study.

Interventions

A standardized counseling protocol was developed to ensure consistency across both groups and was delivered by the same thoracic surgeon to minimize interoperator variability. During crossover sessions, surgeons maintained neutrality and avoided providing additional interpretation. Surgeons adhered to the protocol and followed a structured sequence to explain the tumor location, surgical options, and associated risks.

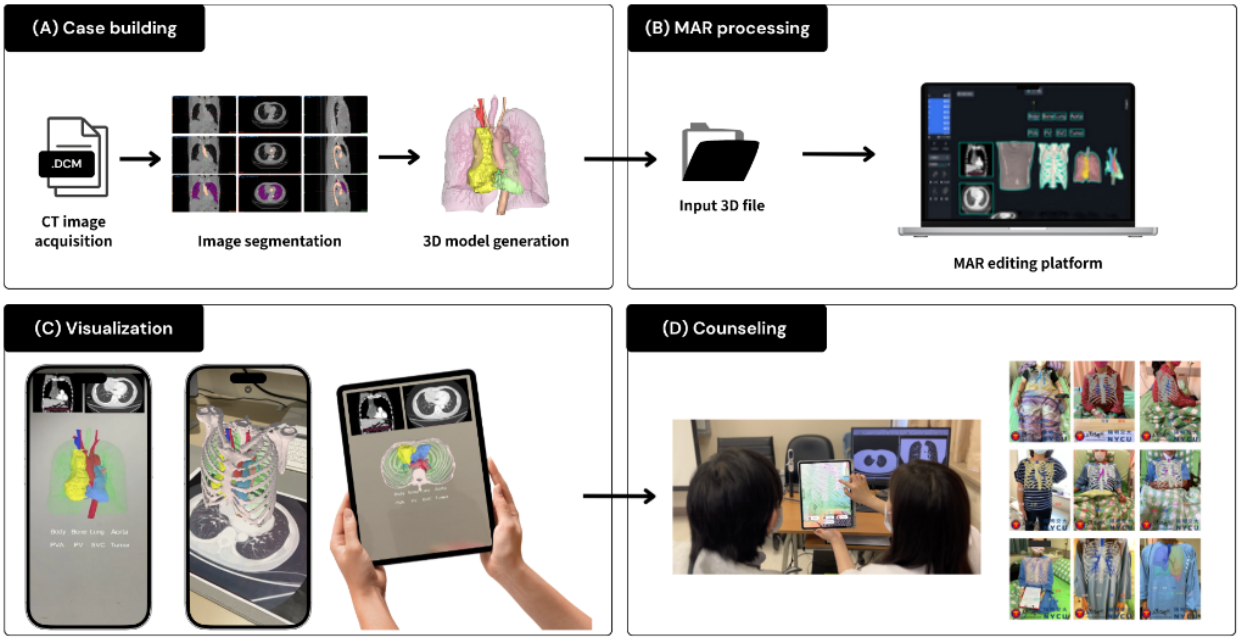
In group A, participants received traditional counseling first, which involved an explanation of their health condition using verbal explanations supported by axial, coronal, and sagittal CT images and a 10 to 15-minute discussion. They were asked to complete a questionnaire related to SDM (questionnaire 1), followed by MAR counseling with interactive, patient-specific 3D anatomical models presented through mobile devices, and then to fill out a second questionnaire (questionnaire 2). In contrast, participants in group B began with MAR counseling. This part included a 5-minute app introduction followed by a 10-to-15-minute personalized anatomical visualization and discussion using the MAR application. After participants (specifically to participants who are in group B) completed the questionnaire 1, they received traditional counseling and proceeded to complete questionnaire 2. All participants experienced both counseling approaches, and their overall preferences and perceptions were evaluated through a questionnaire 3. In Multimedia Appendix 1, demonstration of both traditional and mobile augmented reality (MAR) counseling methods is shown in MP4 format.

Patient-Specific 3D Models via MAR Application

CT images of 47 patients scheduled for thoracic surgery at the Tri-Service General Hospital were collected. Digital images in Digital Imaging and Communications in Medicine format were acquired and imported into Materialise Mimics software (version 22; Materialise NV) for anatomical segmentation and 3D reconstruction of the lung and tumor regions (Figure 1A). Subsequently, we meticulously crafted standard tessellation language 3D models and transformed these Standard Tessellation Language models into GLB format models in Blender software (version 3.6, Blender Foundation) to ensure accurate color representation. The finalized GLB files were then uploaded into the MAR Editor platform (version 3.4.2, Mifly). In the process of editing, we enhanced the visualization by assigning realistic colors, transparency settings, and interactive marks (Figure 1B). The editing platform allowed real-time interaction with the model on tablets or smartphones. Each model was fully interactive, allowing rotation, zooming, and exploration of the anatomical structures from multiple angles (Figure 1C). During counseling sessions, surgeons used these personalized models to explain the disease, treatment options, and the anticipated surgical procedures. (Figure 1D). For standardization, all participants interacted with the same mobile device, an iPad Pro (12.9 inch; fourth generation; Apple Inc), during MAR counseling sessions. In Multimedia Appendix 2, step-by-step demonstration of the mobile augmented reality (MAR) 3D model creation and deployment is shown in MP4 format.

XSL•FO
RenderX

Figure 1. Preparation workflow for building a mobile augmented reality (MAR) system. (A) Acquiring patient-specific computer tomography (CT) data, performing anatomical segmentation, and generating models in a 3D format. (B) Models are optimized and uploaded to the editing platform to adjust settings. (C) Models are embedded into the application's interface, enabling interactive manipulation (zoom, rotation). (D) The finalized interactive model is deployed to the mobile device for physician-patient counseling.



Measurement of Satisfaction Outcomes

To assess the impact of the different counseling methods on SDM and user preference, we designed the self-reported questionnaires for this study. Questionnaire 1 focused on SDM outcomes following the counseling session and included 10 items. Questionnaire 2 was the same as questionnaire 1, allowing the acquired preferences to be determined by comparing the results of questionnaire 2 with those of questionnaire 1. Questionnaire 3 specifically evaluated participants' overall preference for a traditional or an MAR counseling approach. All questionnaire items were rated on a 5-point Likert-type scale, ranging from 1 ("strongly disagree") to 5 ("strongly agree"). A higher score indicated a stronger impact of patient counseling on SDM. To ensure content validity, the questionnaires were reviewed and refined with input from an expert panel.

The primary outcome was the level of SDM assessed in each counseling session (questionnaire 1 and questionnaire 2). The secondary outcome was participant preference (questionnaire 3).

Data Collection and Statistical Analysis

All 47 participants completed the questionnaires, and no one was excluded from the analysis. After the questionnaires were collected and anonymized in Microsoft Excel, relevant statistical analyses were performed using SPSS (version 24.0; IBM Corp). Descriptive statistics were used to summarize the baseline characteristics, demographic data, and outcome variables. Group differences were assessed using Fisher exact and the chi-square tests for categorical variables. To evaluate group comparisons and relationships among variables, independent 2-tailed *t* tests, paired 2-tailed *t* tests, Pearson correlation analyses, and multiple linear regression models were conducted. We recognized that in a crossover study, participants receiving the experimental variable in the first period might continue to be affected in the second period. To minimize carryover effects, we applied a linear mixed model analysis, considering sequence and period effects to adjust the influencing factors. $P<.05$ was considered statistically significant.

Given the noninterventional and educational nature of this study, no interim analyses or stopping guidelines were planned.

Ethical Considerations

The study was reviewed and approved by the institutional review board of Tri-Service General Hospital (A202205082) and was retrospectively registered on ClinicalTrials.gov (NCT07062393). Written informed consent was obtained from all participants before enrollment after being informed of the study purpose, procedures, and voluntary nature of participation. The study posed minimal risk, involved no invasive procedures, and used patients' own clinical data for educational counseling purposes. No additional imaging was required. No concomitant care or cointerventions were provided during the trial period. Participation was voluntary, with the right to withdraw at any time. All data were securely stored and anonymized before the analysis to ensure participant confidentiality. Access to identifiable information was restricted to authorized research personnel. No financial or material incentives were provided to participants.

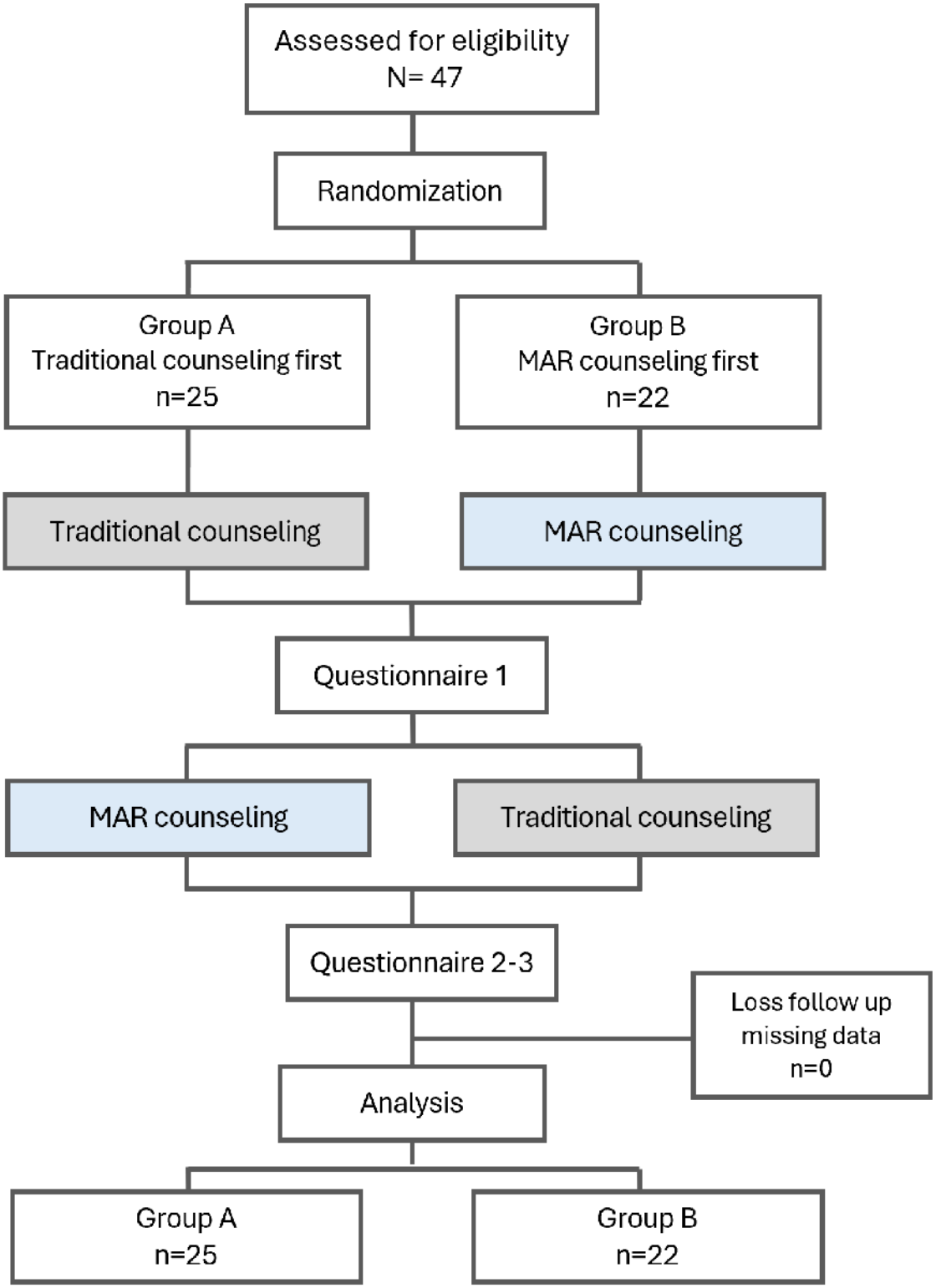
XSL•FO
RenderX

Results

Participant Characteristics

A total of 47 participants fully completed questionnaires and were included in the final analysis, with 25 from counseling group A and 22 from counseling group B. A visual summary of the study flow is provided in Figure 2. No adverse events or unintended effects were reported in either group during or after the counseling sessions. The interventions involved no physical or psychological risks, and all participants completed the study without incident. Most participants in both groups were older than 30 years, held a university degree or higher, had been using smartphones for 10 to 20 years, and had no experience with AR or virtual reality (VR) technologies. No significant differences were observed in participant characteristics between the two groups, as summarized in Table 1.

Figure 2. Study flowchart for participant enrollment, randomization, intervention, and analysis. MAR: mobile augmented reality.



XSL•FO
RenderX

Table 1. Demographic characteristics of participants (N=47).

Characteristics	Group A (n=25)	Group B (n=22)	*P* value
Sex n (%)			.12[a]
Male	8 (32%)	12 (54.6)	
Female	17 (68)	10 (45.4)	
Age (years), n (%)			.20[b]
<30	1 (4)	2 (9.1)	
30-59	12 (48)	15 (68.2)	
≥60	12 (48)	5 (22.7)	
Education, n (%)			.35[a]
High school or lower	7 (28)	9 (40.9)	
University or higher	18 (72)	13 (59.1)	
Duration of mobile phone use (years), n (%)			>.99[b]
≤10	6 (24)	5 (22.7)	
10-20	16 (64)	15 (68.2)	
≥20	3 (12)	2 (9.1)	
Average daily internet use (hours), n (%)			.64[a]
<3	10 (40)	7 (31.8)	
3-5	10 (40)	8 (36.4)	
≥5	5 (20)	7 (31.8)	
Experience of using health care application, n (%)			.91[b]
Every day	7 (28)	7 (31.8)	
Every week	5 (20)	4 (18.2)	
Every month	3 (12)	4 (18.2)	
Less than once a month	4 (16)	4 (18.2)	
Never	6 (24)	3 (13.6)	
Experience of using AR[c] and VR[d], n (%)			.20[b]
Yes	4 (16)	7 (31.8)	
No	21 (84)	15 (68.2)	

[a]Chi-square test.

[b]Fisher exact test.

[c]AR: augmented reality.

[d]VR: virtual reality.

SDM Scores

Table 2 presents the results of the questionnaire related to SDM. Before grouping, the overall mean score for traditional counseling was 37.40 (SD 5.83), and the overall mean score for MAR counseling was 47.98 (SD 2.92). The results related to patient SDM scores showed a significant difference ($P \leq .001$). This finding indicates that MAR counseling was more effective in supporting patients' engagement in the decision-making process. In traditional counseling, the lowest-scoring topics were encouraging participation (mean 3.66, SD 0.70) and enhancing communication (mean 3.66, SD 0.67). In contrast, MAR counseling achieved significantly higher scores in these 2 topics (mean 4.83, SD 0.38 and mean 4.87, SD 0.34; $P < .001$), including thought organization (mean 4.83, SD 0.38) and trusting health care providers' treatment methods (mean 4.83, SD 0.38). MAR counseling also supported participants in understanding medical options (mean 4.74, SD 0.44) and provided more opportunities to ask questions to health care providers (mean 4.77, SD 0.43), thereby supporting decision-making (mean 4.79, SD 0.41).

XSL•FO
RenderX

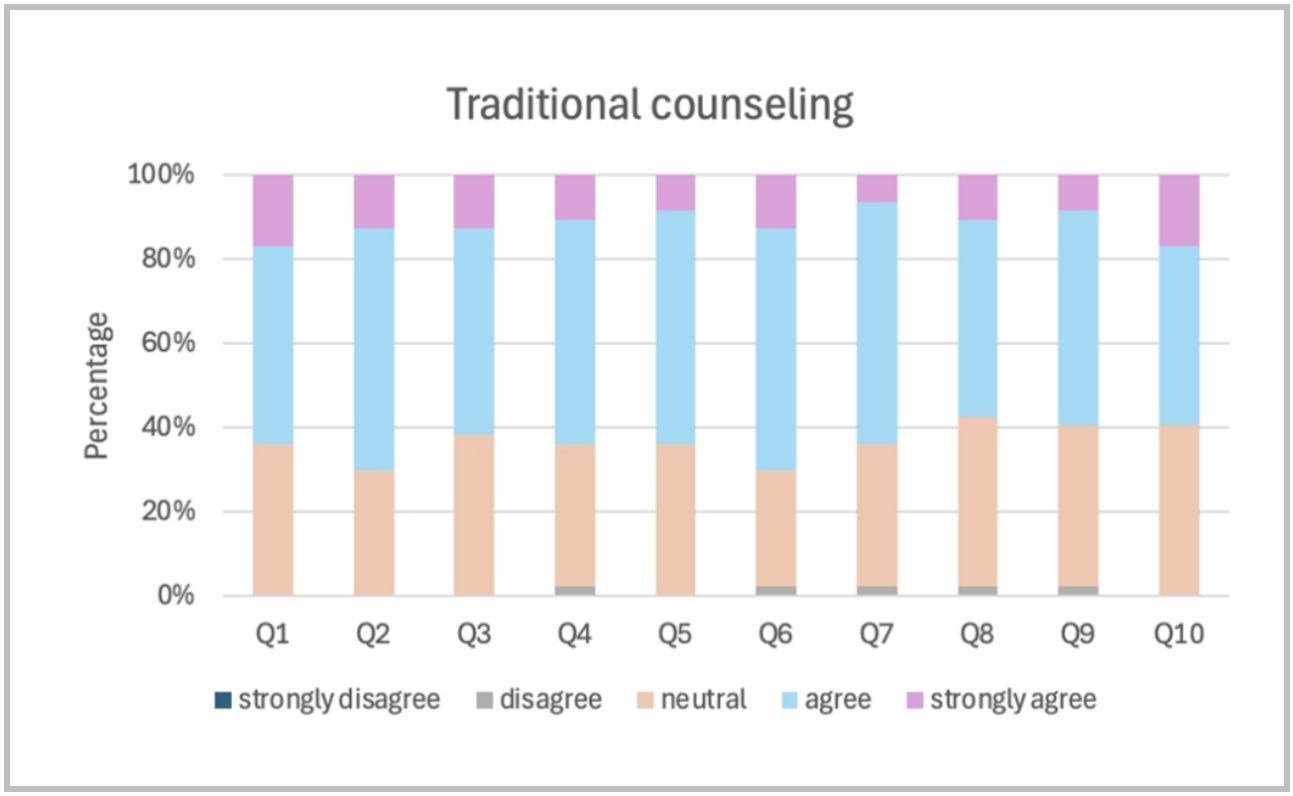
Table 2. Results from the 10-item questionnaire assessing patients' perspectives on shared decision-making (SDM) after traditional versus mobile augmented reality (MAR) counseling.

Questions	Traditional counseling, mean (SD)	MAR counseling, mean (SD)	*P* value[a]
1. Helps me feel more prepared to make decisions.	3.81(0.71)	4.79 (0.41)	≤.001
2. Helps me organize my thoughts about these decisions.	3.83 (0.64)	4.83 (0.38)	≤.001
3. Helps me consider the pros and cons of each option.	3.74 (0.68)	4.77 (0.48)	≤.001
4. Helps me compile the questions I want to ask.	3.69 (0.67)	4.77 (0.43)	≤.001
5. Clarifies and communicates what matters most to me.	3.72 (0.62)	4.79 (0.41)	≤.001
6. Helps me understand the available medical options.	3.81 (0.68)	4.74 (0.44)	≤.001
7. Provides me with the opportunity to ask questions.	3.68 (0.63)	4.77 (0.43)	≤.001
8. Encourages my participation in the health care process.	3.66 (0.70)	4.83 (0.38)	≤.001
9. Enhances communication between me and health care providers	3.66 (0.67)	4.87 (0.34)	≤.001
10. Helps me trust health care providers' treatment methods	3.77 (0.73)	4.83 (0.38)	≤.001

[a]*P* values were obtained using the Wilcoxon rank-sum test. Higher scores indicate greater agreement with each statement (1-5 Likert scale).

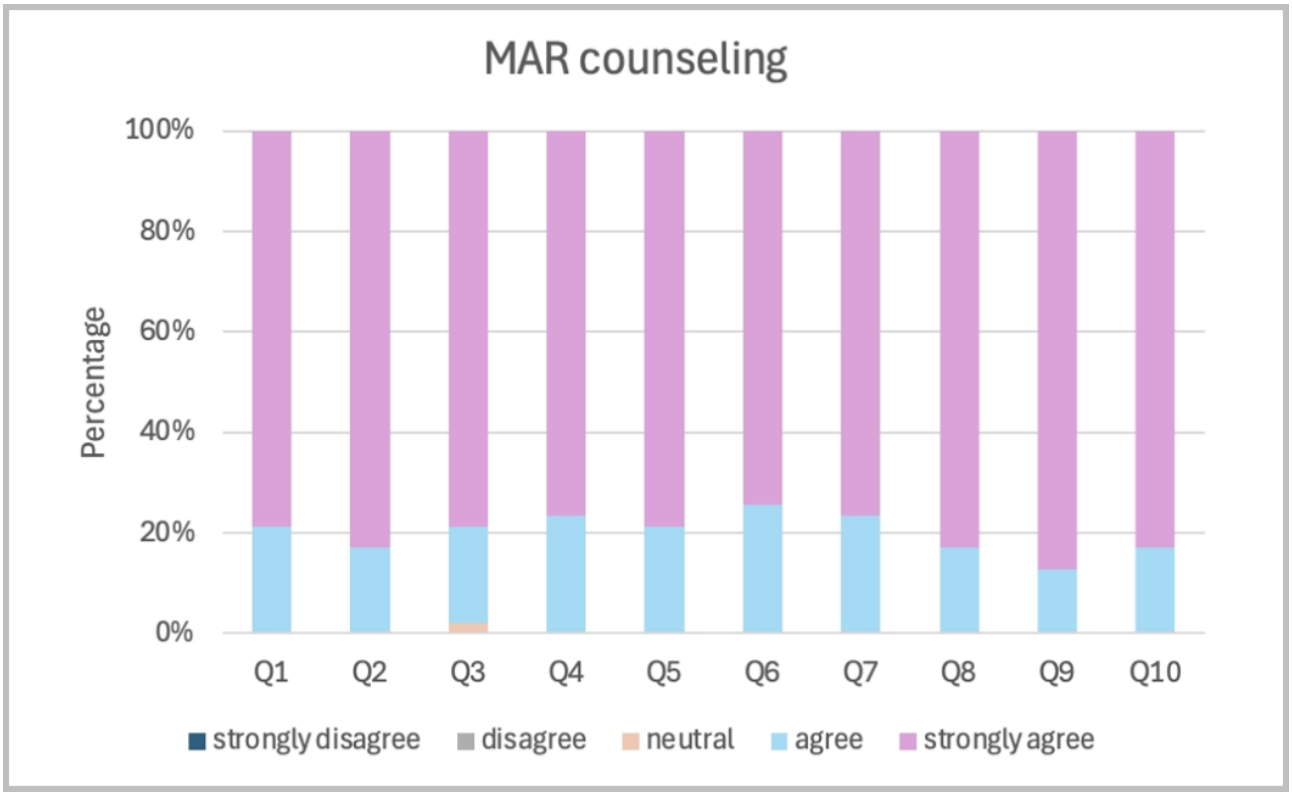
Figures 3 and 4 illustrate the distribution of participant responses for each of the 10 questionnaire items following traditional and MAR counseling. In traditional counseling, most participants rated between "agree" and "neutral," with a few selecting "strongly agree." By contrast, the MAR counseling demonstrated higher agreement across all topics, with nearly all participants selecting "agree or strongly agree." The lowest-scoring item, considering the pros and cons of each treatment option (question 3), was rated at 95.46%. For all other topics, 100% of participants rated either agree or strongly agree, with virtually no neutral or negative responses.

Figure 3. Distribution of participant responses to each item of the 10-item Shared Decision-Making Questionnaire (SDM-Q-10) following traditional counseling. Responses were measured on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree." Item wording is provided in Table 2.



XSL•FO
RenderX

Figure 4. Distribution of participant responses to each item of the 10-item Shared Decision-Making Questionnaire (SDM-Q-10) following mobile augmented reality (MAR) counseling. Responses were measured on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree." Item wording is provided in Table 2.



Interventions in Crossover Design

To evaluate the effect of counseling sequence on participant satisfaction, an independent sample *t* test was conducted to compare the differences between group A and group B in this crossover design. Participants were randomly assigned to either group A, which received traditional counseling followed by MAR counseling, or group B, which received MAR counseling first, followed by traditional counseling. The results showed no significant differences between the 2 sequences. The mean satisfaction score for traditional counseling was 36.92 (SD 6.07) in group A and 37.95 (SD 5.64) in group B (t=−0.60, P=.55), whereas MAR counseling scored 47.72 (SD 2.94) and 48.27 (SD 2.95) in group A and B, respectively (t=−0.64, P=.52) These findings indicated that the order in which traditional and MAR counseling methods were administered did not influence satisfaction outcomes.

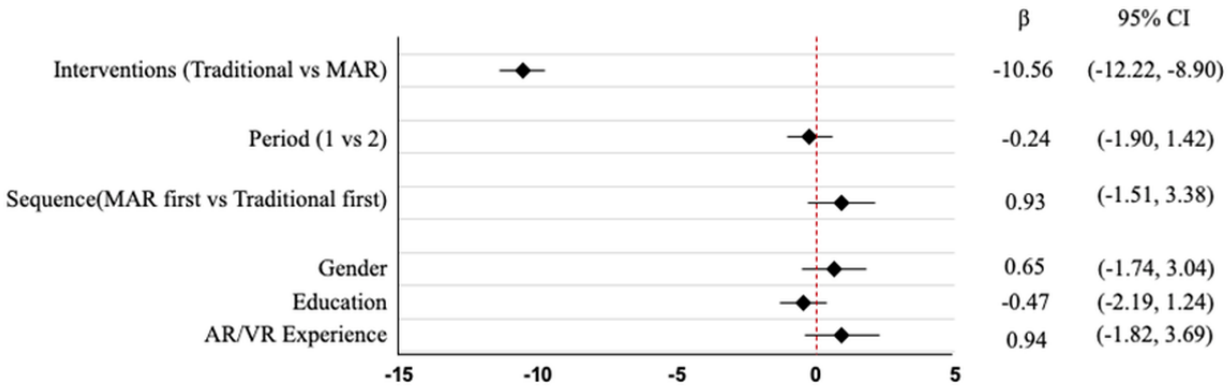
Further within-group analyses revealed that MAR counseling was consistently rated higher than traditional counseling across both sequences. In group A, the mean difference between traditional and MAR counseling was −10.80 (SD 5.65, P<.001, Cohen d=−1.91), and in group B, the difference between MAR and traditional counseling was 10.32 (SD 5.61, P<.001, Cohen d=1.84).

Adjusted Comparison Using Linear Mixed Model

Although the paired samples *t* test provided preliminary evidence that MAR counseling was rated higher than traditional counseling, additional analyses using a linear mixed model were conducted to examine any potential carryover effect. Considering both the sequence (traditional first or MAR first) and the period (first or second), neither the period nor the sequence effect reached statistical significance (P>.05). Demographic variables, including gender, education level, and AR and VR experience, also did not significantly affect the outcomes. However, the analysis showed that the different interventions demonstrated a significant difference; MAR counseling scored 10.55 points higher than traditional counseling, reaching statistical significance (P≤.001). These findings indicate that the type of intervention was the primary factor influencing participants' perceptions; whereas period and sequence effects did not significantly impact counseling outcomes. The fixed-effect estimates are presented in Figure 5.

XSL•FO
RenderX

Figure 5. Fixed effects estimated from linear mixed model assessing the impact of treatment sequence, period, and demographic variables on shared decision-making (SDM) scores. AR: augmented reality; VR: virtual reality; MAR: mobile augmented reality.



Participant Preferences

The participants' overall preferences were assessed with questionnaire 3 (I am willing to use this counseling method to communicate with health care providers; 1=traditional; 2=MAR). The analysis revealed that participants had a strong preference for the MAR counseling method. The mean score for MAR counseling (4.96, SD 0.20) was significantly higher than that for traditional counseling (mean 3.77, SD 0.70; $P \leq .001$). Furthermore, 45 of 47 (96%) participants indicated a strong willingness to use MAR counseling as a communication method in the SDM process.

Influence of Demographic Variables

To further investigate whether the participants' preferences were influenced by demographic variables, a multiple linear regression analysis was conducted (Table 3). The results showed that none of the demographic variables, including gender, education level, duration of mobile phone use, average daily internet use, and AR and VR experience, reached statistical significance ($P > .05$) in the MAR counseling group. Similarly, most of the demographic variables did not demonstrate a significant result in traditional counseling. However, without AR and VR experience showed a positive effect (B=0.583, β=.357; $P=.03$), indicating that the participants without AR and VR experience reported a higher preference for the traditional counseling approaches compared to those who had the experience in AR and VR technology. The adjusted R^2 values for both models were low (MAR=0.12; traditional=0.19), indicating that demographic factors explained a small proportion of the results.

Table 3. Multiple linear regression predicting shared decision-making scores from demographic variables.

Variable	B (95% CI)	β	*t* test (*df*)	*P* value
MAR[a] counseling				
Sex (female)	0.092 (−0.034 to 0.217)	.224	1.476 (40)	.15
Education level (college)	−0.037 (−0.185 to 0.110)	−.092	−0.51 (40)	.61
Education level (graduate)	−0.147 (−0.341 to 0.047)	−.259	−1.532 (40)	.13
Duration of mobile phone use	−0.003 (−0.121 to 0.116)	−.008	−0.048 (40)	.96
Average daily internet use	0.026 (−0.056 to 0.109)	.106	0.641 (40)	.52
AR[b] and VR[c] experience (no)	−0.013 (−0.173 to 0.148)	−.027	−0.161 (40)	.87
Traditional counseling				
Sex (female)	0.198 (−0.215 to 0.611)	.142	0.969 (40)	.34
Education level (college)	0.253 (−0.232 to 0.738)	.183	1.054 (40)	.30
Education level (graduate)	0.517 (−0.121 to 1.155)	.266	1.637 (40)	.11
Duration of mobile phone use	−0.108 (−0.498 to 0.282)	−.089	−0.561 (40)	.58
Average daily internet use	0.045 (−0.228 to 0.317)	.053	0.332 (40)	.74
AR and VR experience (no)	0.583 (0.055 to 1.111)	.357	2.232 (40)	.03

[a]MAR: mobile augmented reality.

[b]AR: augmented reality.

[c]VR: virtual reality.

XSL•FO
RenderX

Discussion

Principal Findings

The study results indicated that participants preferred MAR counseling more than traditional counseling approaches during SDM, particularly regarding trust and participant engagement. This effect may be attributed to participants gaining more disease-specific knowledge from MAR, which, in turn, led to increased engagement [9]. Paired samples *t* tests revealed significantly higher scores for MAR counseling, demonstrating that the digital approach had a positive impact in helping participants comprehend the risks and benefits associated with their treatment options. Furthermore, the findings were strengthened by the linear mixed model analysis, which considered sequence and period effects to adjust the impact factors. Neither the sequence nor the period effects reached statistical significance.

These findings are consistent with those of a previous study that used 3D models in patient counseling for renal masses [15]. Although it is possible that the visually rich nature of MAR contributed to patient engagement, the structured questionnaires assessed more than just visual appeal. Notably, items such as question 3 (“Helps me consider the pros and cons of each option”) and question 6 (“Helps me understand the available medical options”), which evaluated the understanding of pros and cons and treatment options, showed significant improvement in the MAR counseling. This suggests that MAR facilitated greater engagement and deeper cognitive processing of medical information. These findings indicate that the observed improvements in SDM were not limited to increased engagement or visual novelty but included measurable enhancement in understanding treatment risks and options. This supports the value of MAR not only as a communication aid but also as a decision-support tool grounded in cognitive comprehension that strengthens SDM processes in clinical practice.

Implementation Considerations

However, several considerations should be noted. MAR counseling demanded more time and a more complex preparation process compared with traditional counseling. The cost-effectiveness of adopting this new counseling method should be explored thoroughly before full implementation.

MAR offers a powerful solution to enhance telemedicine and improve technological literacy. One key advantage is its ability to address the challenge of interpreting complex medical data, such as 2D DICOM images, which can be difficult for patients without specialized anatomical knowledge to understand. MAR overlays 3D visualizations onto the physical world, making it easier for patients and health care providers to comprehend medical images by offering an interactive, spatial representation of anatomy. This interactive experience enhances the clarity of diagnostic information, allowing for more informed SDM.

Moreover, unlike the use of AR smart glasses, which can sometimes cause discomfort such as dizziness or visual fatigue, patients using MAR on mobile devices did not report such issues. The display on a mobile device offers flexibility in positioning and control, reduces strain on the eyes, and prevents the sensory overload that can occur with head-mounted displays. As a result, patients were more comfortable engaging with the technology for extended periods, facilitating a more seamless integration of MAR into telemedicine and clinical workflows. This ease of use further promotes patient empowerment, enabling them to take an active role in understanding their own health data.

Prior literature has reported that 3D models have made telemedicine visits as effective as in-person visits [15]. This finding could be applied to MAR counseling, as the counseling process can be conducted through smartphones or tablet devices. The integration of MAR into telemedicine can potentially enhance the quality of counseling in the future.

Limitations

However, the implementation of MAR consultation also has hidden challenges. Compared with traditional methods, MAR counseling demands a longer preparation time, additional training for clinical staff, and reliance on stable hardware and network infrastructure. Although MAR technology offers the advantage of accessing virtual resources anytime and anywhere, whether in medical education or clinical settings, its effectiveness and usability remain closely tied to technical settings. When users operate MAR systems in environments with limited bandwidth or outdated hardware, they may experience delays, failed file loading, or degraded 3D rendering performance. These limitations may result in reduced acceptance or satisfaction, regardless of the perceived educational or clinical value of the MAR content.

Additionally, the questionnaires used in this study were developed to reflect the unique attributes of MAR, which may limit direct comparison with SDM outcome measures (eg, the 9-item Shared Decision Making Questionnaire or the OPTION [Observing Patient Involvement] scale). Future studies should incorporate validated tools alongside MAR-specific items to improve generalizability and construct validity. Finally, the study was conducted at a single institution with a modest sample size, which may limit the broader applicability of our findings. Although the sample size was statistically adequate based on pilot data, larger multicenter trials are warranted to improve generalizability and explore subgroup effects. Future multicenter studies with larger and more diverse populations, time-matched conditions, and validated instruments are warranted to confirm the impact and feasibility of MAR counseling in thoracic surgery and beyond.

Conclusions

MAR counseling demonstrated a significantly higher impact on SDM compared with traditional counseling for patients undergoing thoracic surgery. This indicates that MAR consultation is effective in supporting patients’ engagement in the decision-making process and enhancing their understanding, participation, and trust. Importantly, MAR counseling demonstrated statistically significant improvements in participants’ perceptions compared to traditional counseling. In contrast, under traditional counseling, patients often find it difficult to effectively organize their questions. Insufficient interaction with physicians may limit patients’ active

XSL•FO
RenderX

participation in the SDM process. MAR consultation scored significantly higher in the same topics, better supported patients, and promoted communication between physicians and patients. However, we recommend conducting a multicenter study with a usability score assessment to validate the generalizability and feasibility of MAR counseling in future research. This will help ensure the widespread applicability and practicality of this approach in a broader context.

Acknowledgments

The authors wish to thank Mifly Company (Taipei, Taiwan) for their technical assistance in integrating and deploying the mobile augmented reality platform used in this study. No financial support or sponsorship was received from the company. This research was supported by the Tri-Service General Hospital Medical Research Project (TSGH-E-112239). No external commercial or nonprofit funding was received for this study. OpenAI's ChatGPT (GPT-4) was used to assist in language refinement, formatting, and drafting of reviewer response letters during manuscript preparation. The authors take full responsibility for the accuracy and integrity of all content.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to institutional restrictions but are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: KHL
Methodology: KHL, DHW, WR
Data curation: YSC
Formal analysis: YCH
Funding acquisition: KHL
Investigation: YSC
Project administration: KHL
Resources: TWH
Software: YCH, WR, DHW
Supervision: KHL, DHW
Validation: TWH
Visualization: YCH (lead), DHW (supporting)
Writing—original draft: YSC (lead), YCH (lead), JL (supporting)
Writing—review and editing: YSC (lead), YCH (lead), JL (supporting), KHL (supporting), TWH (supporting)
All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demonstration of both traditional and mobile augmented reality counseling methods (MP4 format).
[MP4 File (MP4 Video), 16670 KB - mhealth_v13i1e79632_app1.mp4]

Multimedia Appendix 2
Step-by-step demonstration of the mobile augmented reality 3D model creation and deployment (MP4 format).
[MP4 File (MP4 Video), 88583 KB - mhealth_v13i1e79632_app2.mp4]

Multimedia Appendix 3
CONSORT checklist.
[PDF File (Adobe PDF File), 246 KB - mhealth_v13i1e79632_app3.pdf]

References

1. Azuma R, Baillot Y, Behringer R, Feiner S, Julier S, MacIntyre B. Recent advances in augmented reality. IEEE Comput Grap Appl 2001;21(6):34-47. [doi: 10.1109/38.963459]
2. Craig AB. Mobile augmented reality. In: Craig AB, editor. Understanding Augmented Reality. Waltham: Morgan Kaufmann; 2013:209-220.

XSL•FO
RenderX

3. Arjomandi Rad A, Vardanyan R, Thavarajasingam S, Zubarevich A, Van den Eynde J, Sá MP, et al. Extended, virtual and augmented reality in thoracic surgery: a systematic review. Interact Cardiovasc Thorac Surg 2022 Jan 18;34(2):201-211 [FREE Full text] [doi: 10.1093/icvts/ivab241] [Medline: 34542639]
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020 Jan;70(1):7-30 [FREE Full text] [doi: 10.3322/caac.21590] [Medline: 31912902]
5. Cancer registry annual report. Health Promotion Administration, Taiwan. 2021. URL: https://www.hpa.gov.tw/EngPages/List.aspx?nodeid=1070 [accessed 2025-05-29]
6. Lackey A, Donington J. Surgical management of lung cancer. Semin Intervent Radiol 2013 Jun 28;30(2):133-140 [FREE Full text] [doi: 10.1055/s-0033-1342954] [Medline: 24436529]
7. Kessels RP. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [doi: 10.1177/014107680309600504] [Medline: 12724430]
8. Braddock CH, Fihn SD, Levinson W, Jonsen AR, Pearlman RA. How doctors and patients discuss routine clinical decisions: informed decision making in the outpatient setting. J Gen Intern Med 1997 Jun;12(6):339-345. [doi: 10.1007/s11606-006-5081-x]
9. Deber RB, Kraetschmer N, Irvine J. What role do patients wish to play in treatment decision making? Arch Intern Med 1996 Jul 08;156(13):1414-1420. [doi: 10.1001/archinte.1996.00440120070006]
10. Sandman L, Munthe C. Shared decision making, paternalism and patient choice. Health Care Anal 2010 Mar 30;18(1):60-84. [doi: 10.1007/s10728-008-0108-6] [Medline: 19184444]
11. Tucker JD, Cheng Y, Wong B, Gong N, Nie J, Zhu W, Patient-Physician Trust Project Team. Patient-physician mistrust and violence against physicians in Guangdong province, China: a qualitative study. BMJ Open 2015 Oct 06;5(10):e008221 [FREE Full text] [doi: 10.1136/bmjopen-2015-008221] [Medline: 26443652]
12. Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. Patient Educ Couns 2006 Mar;60(3):301-312. [doi: 10.1016/j.pec.2005.06.010] [Medline: 16051459]
13. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
14. Poprzeczny AJ, Stocking K, Showell M, Duffy J. Patient decision aids to facilitate shared decision making in obstetrics and gynecology: a systematic review and meta-analysis. Obstet Gynecol 2020 Feb;135(2):444-451. [doi: 10.1097/AOG.0000000000003664] [Medline: 31923056]
15. McDonald M, Shirk JD. The effect of digital three-dimensional reality models on patient counseling for renal masses. JSLS 2023;27(1):e2022 [FREE Full text] [doi: 10.4293/JSLS.2022.00084] [Medline: 36818764]

Abbreviations

AR: augmented reality
VR: virtual reality
MAR: mobile augmented reality
SDM: shared decision-making

Edited by L Buis; submitted 25.06.25; peer-reviewed by S Lampridis; comments to author 25.07.25; revised version received 19.09.25; accepted 09.10.25; published 13.11.25.

Please cite as:
Chen YS, Hsu YC, Romalee W, Wang DH, Lai J, Huang TW, Lin KH
Effect of Mobile Augmented Reality Counseling on Improving Shared Decision-Making in Thoracic Surgery: Randomized Clinical Crossover Trial
JMIR Mhealth Uhealth 2025;13:e79632
URL: https://mhealth.jmir.org/2025/1/e79632
doi:10.2196/79632
PMID:

©Ying-Shian Chen, Yi-Chen Hsu, Worachate Romalee, Ding-Han Wang, Jennifer Lai, Tsai-Wang Huang, Kuan Hsun Lin. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Preferences and Willingness to Pay for Health App Assessments Among Health Care Stakeholders: Discrete Choice Experiment

Anna-Lena Frey[1], PhD; Simon Leigh[1,2], PhD; Carla Toro[2], PhD; Carme Pratdepàdua Bufill[3], MSc; Charles McCay[4], BA; Tatjana Prenđa Trupec[5], MCS, MBA; Giuseppe D'Avenio[6], MSc; Menno Kok[7], PhD; Antanas Montvila[8], MSc, MD; Philipp Goedecker[9], MA; Petra Hoogendoorn[10], MSc

[1]Organisation for the Review of Care and Health Apps, Daresbury, United Kingdom
[2]Warwick Medical School, University of Warwick, Coventry, United Kingdom
[3]Fundació TIC Salut Social, Barcelona, Spain
[4]Ramsey Systems Ltd, Shrewsbury, United Kingdom
[5]Care and Public Health Research Institute, Maastricht University, Maastricht, The Netherlands
[6]National Center for Innovative Technologies in Public Health, Istituto Superiore di Sanità, Rome, Italy
[7]EIT Health Belgium and The Netherlands, Rotterdam, The Netherlands
[8]Hospital of Lithuanian University of Health Sciences, Kaunas Clinics, Kaunas, Lithuania
[9]COCIR (Comité de Coordination Européen de l'industrie radiologique, électromédicale et informatique des soins de santé), Brussels, Belgium
[10]National eHealth Living Lab, Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

Corresponding Author:
Petra Hoogendoorn, MSc
National eHealth Living Lab
Department of Public Health and Primary Care
Leiden University Medical Center
Albinusdreef 2
Leiden, 2333 ZA
The Netherlands
Phone: 31 654341785
Email: a.p.y.hoogendoorn@lumc.nl

Abstract

Background: The adoption of high-quality health apps has been slow, despite the myriad benefits associated with their use. This is partly due to concerns regarding the effectiveness, safety, and data privacy of such apps. Quality assessments with robust and transparent criteria can address these concerns and, thereby, encourage the use of high-quality apps. However, a major challenge for such assessments is reaching a scale at which a substantial proportion of the more than 350,000 available health apps can be evaluated.

Objective: To support the scaling of health app quality assessments, this study aimed to examine the preferences and willingness to pay for assessments with different value propositions among potential customers.

Methods: We conducted 2 discrete choice experiments: one with 41 health app developers and another with 46 health system representatives (from health care institutions, authorities, and insurers) from across Europe. Mixed logit models were applied to examine the impact of assessment attributes on participants' choices as well as to calculate marginal willingness to pay and predicted assessment uptake.

Results: Among health app developers, the attributes with the largest impact on assessment choices were the associated clinical care uptake (integration into clinical guidelines and reimbursement or procurement) and cost (purchase price). Increased willingness to use assessed apps and app store integration of assessment results had a moderate impact on choices, while required developer time investment and time until assessment results become available made the smallest contribution. Among health system representatives, increased willingness of clinicians and patients to use evaluated apps had the greatest impact on assessment choices, followed by cost. Time until assessment result availability and the percentage of peers recommending the assessment made a moderate contribution, while reassessment frequency had the smallest impact on choices. On average, health app developers were willing to pay an additional €9020 (95% CI €4968-€13,072) if an assessment facilitates guideline integration and procurement or reimbursement (at the time of data collection, €1=US $1.11), while health system representatives were, on average, willing to

XSL•FO
RenderX

pay €7037 (95% CI €4267-€9806) more if an assessment results in a large, rather than a small, increase in willingness to use the evaluated app. The predicted uptake of assessments that offer the preferred values for all attributes was 88.6% among app developers and 91.1% among health system representatives.

Conclusions: These findings indicate that, to maximize uptake and willingness to pay among health app developers, it is advisable for assessments to facilitate or enable clinical guideline integration and reimbursement or procurement for high-scoring apps. Assessment scaling thus requires close collaboration with health authorities, health care institutions, and insurers. Furthermore, if health system organizations are targeted as customers, it is essential to provide evidence for the assessment's impact on patients' and clinicians' willingness to use health apps.

(JMIR Mhealth Uhealth 2025;13:e57474) doi:10.2196/57474

KEYWORDS

health apps; quality assessment; discrete choice experiment; willingness to pay; health system; app developers

Introduction

Background

There is increasing recognition that digital health technologies, including health apps that support health system services, disease prevention, self-management, and treatment, should play a greater role in health care [1]. Health apps complement in-person care and could ease the strain on clinicians and health systems, which is partly due to a shortage of health care professionals [2,3].

A number of recent studies have demonstrated the beneficial effects of health apps on health outcomes and quality of life [4], including in condition areas such as mental health [5,6], diabetes [7-9], hypertension [7], cancer [10,11], and asthma [12]. Moreover, health apps can contribute to cost savings and more efficient use of health system resources that are facing an increasing demand [13,14]. Indeed, past estimates suggest that using health apps to their full potential could have resulted in vast savings across the European Union in 2017 [15].

To attain these benefits, the widespread use of health apps is necessary, yet adoption has been slow [16], with concerns related to the effectiveness, safety, usability, and data privacy of such apps reported as a major barrier to their use among both patients and clinicians [17,18].

Importantly, these apprehensions are justified in some cases [19], and current European regulations are arguably not sufficient to fully alleviate such concerns [20,21]. First, the Medical Device Regulation (MDR) only covers apps that are *specifically intended by the developer ("manufacturer") to be used for medical purposes*. This creates a "loophole" related to the phrasing of an app's intended purpose, incentivizes development of apps with nonmedical purposes, and leaves a substantial number of health apps unregulated by the MDR [22]. For instance, in a sample of 271 health apps examined for the Dutch Ministry of Health, 215 apps (79.3%) did not fall under the MDR [23]. Such apps, as well as apps falling under MDR risk class I, are not required to undergo conformity assessment by a notified body [23]. Second, although the General Data Protection Regulation applies to health apps available in Europe that process personal data, compliance with this regulation is usually not independently evaluated. Indeed, it has been found that a large proportion of health apps on the market do not meet relevant data protection requirements [24,25]. Third, concerns have been raised regarding the quality and appropriateness of the studies that are deemed acceptable under the MDR to demonstrate a product's safety and effectiveness [26,27]. Finally, the only user-facing output of the MDR assessment is a Conformité Européenne (CE; European Conformity) mark, which consists of the letters "CE" being affixed to medical devices, including health apps, if they have been found to comply with the MDR (through self-assessment for class I apps or through evaluation by a notified body for apps with a higher risk classification). This CE mark only aims to indicate conformity with the MDR, without seeking to provide transparency to users unfamiliar with regulatory requirements regarding the domains that have been assessed and how apps compare across them.

Partly as a result of the aforementioned considerations, a number of third-party health app assessment frameworks have been established by organizations such as health authorities, health care institutions, medical professional associations, and for-profit companies to promote the use of high-quality apps across Europe and beyond [28,29]. However, a major challenge for such assessment initiatives is reaching a scale at which they can evaluate a large proportion of the more than 350,000 health apps available on the global market [30] or even of the approximately 100,000 health apps available in Europe [29,31,32]. Such scaling can only be achieved if, alongside an efficient assessment process and a sufficient number of assessors, an appropriate customer group is targeted and customers' expectations regarding assessment costs and benefits are met.

Objectives

To facilitate the scaling of health app quality assessments, this study aimed to gather insights into the preferences and willingness to pay for health app assessments with different value propositions among potential assessment customers across Europe. For this purpose, employees of health app development companies and health system representatives (from health care institutions, health authorities, health technology assessment bodies, and health insurers) were invited to take part in a discrete choice experiment (DCE).

The study was conducted as part of the European Union–funded Label2Enable project, which aims to support the adoption and operation of the CEN ISO/TS 82304-2 health app quality assessment as a trusted standard across Europe [33,34].

XSL•FO
RenderX

However, the study's findings are not specific to the CEN ISO/TS 82304-2 assessment and provide valuable information for the scaling of health app quality evaluation initiatives more generally.

Methods

Overview

DCEs are commonly used in health care contexts to examine stakeholders' preferences [35]. In such experiments, participants are presented with several choice sets consisting of 2 or more alternatives (eg, 2 treatment options) that differ across a range of attributes (eg, cost, effectiveness, and side effects). Participants repeatedly choose between alternatives with different attribute values, allowing researchers to determine participants' preferences as well as the trade-offs that participants are willing to make across attributes.

As part of this study, the DCE methodology was applied to elicit preferences for health app quality assessments with different value propositions from potential customers of third-party assessments. The study was designed and conducted in line with best practices developed by the International Society for Pharmacoeconomics and Outcomes Research Conjoint Analysis Working Group [36].

Selection of Attributes and Levels

To identify relevant assessment attributes for the DCE as well as realistic values for these attributes, interviews were conducted with representatives of digital health assessment organizations from 7 European countries (for details, refer to Table S1 in Multimedia Appendix 1). The interviews, which also served to inform other aspects of the Label2Enable project, covered topics such as assessment scope, assessment trigger, time requirements, costs and funding, benefits of the assessment for health care stakeholders, progress to date, and barriers to scaling.

Subsequently, a video conference discussion was held with representatives of the stakeholder groups targeted by this study to determine the attributes they regarded as most important. Each representative was called on individually to comment on the attributes identified during the interviews and to suggest additional aspects to consider. All attributes that were indicated as being important by any of the representatives were shortlisted. More specifically, 2 shortlists were created (with some overlapping attributes): one for health app developers and another for health system representatives (from health care institutions, health authorities, health technology assessment bodies, and health insurers). Creating 2 attribute lists for use in 2 separate DCEs was deemed necessary due to the differing priorities and objectives of these stakeholder groups. For instance, an important attribute for health app developers was the amount of time they would need to invest to prepare the necessary documentation for the assessment. By contrast, this attribute was not relevant for health system representatives because they are not involved in gathering documentation as part of the assessment process.

The attribute shortlists were further refined through email correspondence with the stakeholder representatives until a consensus agreement on 5 to 6 attributes was reached. This number of attributes was chosen to limit the cognitive load of the choice task with full profiles. The levels for the selected attributes were based on the range of values reported by the assessment organizations during the interviews as well as the values observed during pilot testing of the CEN ISO/TS 82304-2 assessment. In line with recommendations [36], no more than 3 levels were chosen for each attribute. The final lists of attributes and levels for the 2 DCEs, as presented to participants, are shown in Figures 1 and 2.

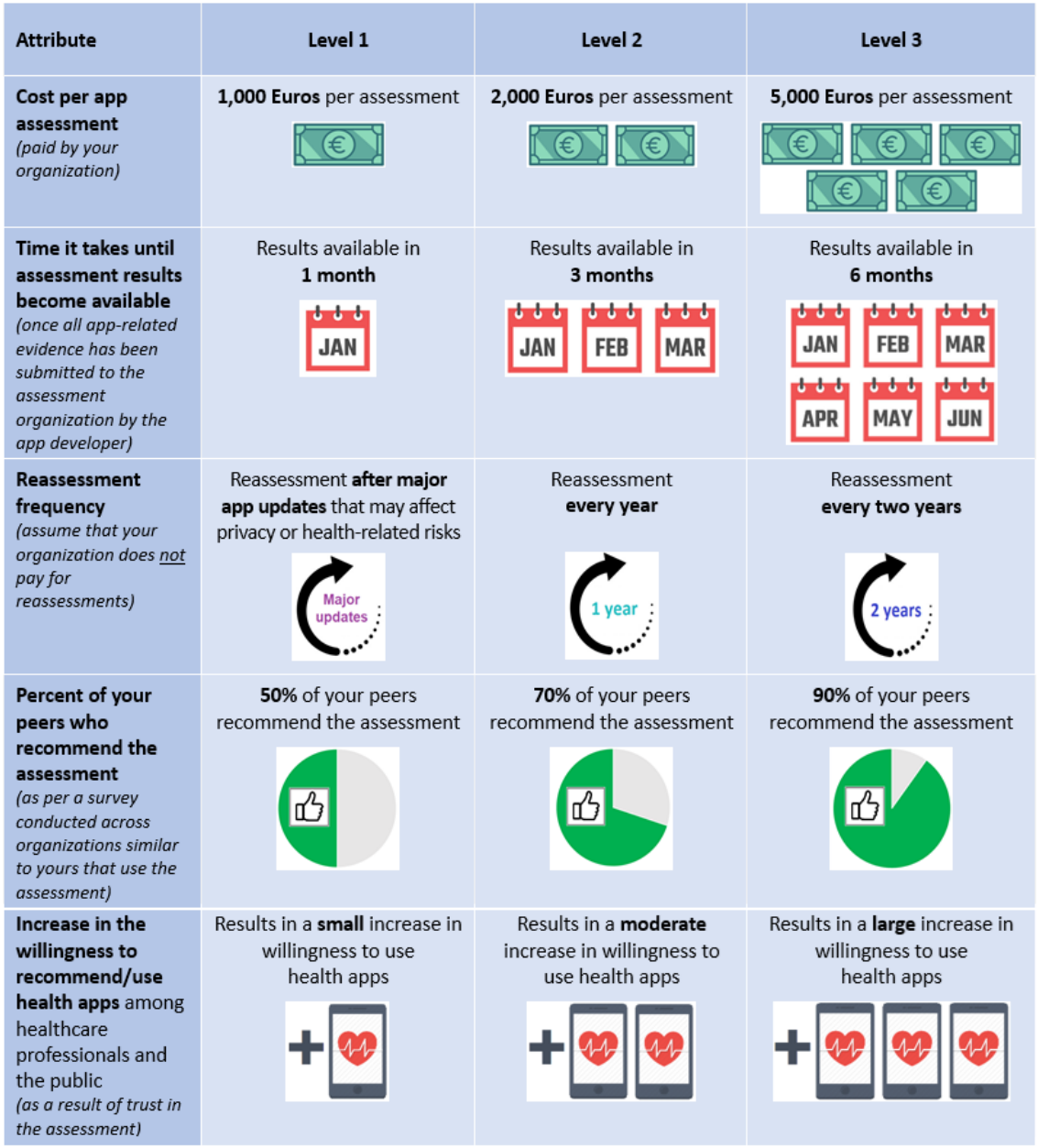
XSL•FO
RenderX

Figure 1. Attributes and levels presented to health app developers during the discrete choice experiment.

Attribute	Level 1	Level 2	Level 3
Cost per app assessment *(paid by your organization)*	**1,000 Euros** per assessment	**2,000 Euros** per assessment	**5,000 Euros** per assessment
Time investment made by your organization *(to collate app-related evidence for the assessment)*	**10 hours** invested by your organization	**20 hours** invested by your organization	**40 hours** invested by your organization
Time it takes until assessment results become available *(once your organization has submitted all app-related evidence to the assessor)*	Results available in **1 month**	Results available in **3 months**	Results available in **6 months**
Impact of the assessment **on clinical care uptake** of your app *(if the results meet stakeholder requirements)*	Does **not** result in either app integration into care **pathway guidelines or** app reimbursement/ **procurement**	Results in app integration into care **pathway guidelines** but **not** in app reimbursement/ **procurement**	Results in app integration into care **pathway guidelines and** app reimbursement/ **procurement**
Integration of assessment results into Google Play and Apple **App stores**	**No** option to prominently **display** assessment **results** in app stores and **no impact on** app search result **rankings**	Option to prominently **display** assessment **results** in app stores **without impact on** app search result **rankings**	Option to prominently **display** assessment **results** in app stores **and prioritisation in** app search result **rankings**
Increase in willingness to use your health app(s) among potential customers *(as a result of trust in the assessment)*	Results in a **small** increase in willingness to use your health app(s)	Results in a **moderate** increase in willingness to use your health app(s)	Results in a **large** increase in willingness to use your health app(s)

XSL•FO
RenderX

Figure 2. Attributes and levels presented to health system representatives during the discrete choice experiment.



Choice Task

For the choice task, a forced-choice elicitation format with 2 alternatives was selected. Specifically, for each choice set, participants were presented with 2 assessments and asked to select the option that they would recommend to their organization. Other approaches, such as presenting >2 alternatives or requiring the ranking of a larger number of options, were deemed to be too cognitively demanding due to the number of included attributes.

After participants had made a selection for a given choice set, they were asked to choose between the assessment they had just selected and an opt-out option (ie, recommending that their organization does not make use of any assessment). Including an opt-out option is beneficial for determining demand [36]. However, if opt-outs are offered as a third option as part of the main choice set, the results can be difficult to interpret. This is because participants may select the opt-out option for reasons unrelated to a true preference, such as to minimize effort and avoid difficult choices [37,38]. The dual response design used in this study has been suggested as a way to address this issue because it requires participants to choose between the offered alternatives on each trial before being given the option to opt out [38].

To ensure that participants could reasonably evaluate choice sets that contained all selected attributes (ie, full profiles), a pilot test was conducted with 15 participants for both DCE versions. The pilot feedback indicated that participants did not find the choice task too complex; thus, full profiles were used.

Experimental Design

The experimental design was generated using Ngene software (version 1.2; ChoiceMetrics Pty Ltd) [39]. For each of the 2 DCEs, a model was specified (refer to the Data Analysis subsection for details) and, through various iterations, the fractional factorial design with the lowest D-error value, and

XSL•FO
RenderX

thus the highest statistical efficiency, was selected [39]. As recommended by the Ngene manual when no previous results are available [39], we used small priors, which indicated the expected sign of the parameters. Specifically, priors for cost, developer time investment, and assessment completion time were assigned a negative sign (−0.00001), and priors for the remaining benefit-related attributes were assigned a positive sign (0.00001).

Each DCE included 18 choice sets generated by Ngene, which, based on the pilot testing, was deemed to yield a suitable survey length without causing fatigue or dropout. In addition, a "rationality test" was incorporated, in which 1 assessment contained presumably preferable values for all attributes compared to the alternative option (ie, a lower level for the first 3 attributes and a higher level for the remaining attributes shown in Figures 1 and 2). The inclusion of such a test is recommended practice to determine whether participants demonstrate rational choice behavior, which is an assumption of the DCE methodology.

Elicitation Method and Data Collection Instrument

The DCEs were presented to participants within a web-based Google Forms survey. At the start of the survey, the following background information was collected: the type, size, and country of the participants' organization; the participants' role in digital health–related decision-making; their previous experience with health app assessment; their views on whether making use of health app assessment would be beneficial for their organization; and their opinion on whether displaying assessment information should be mandatory or voluntary for health app developers.

Health system representatives were additionally asked to rank their preferences for different ways of engaging with health app assessments. Specifically, they ranked the following options from 1 (most preferred) to 5 (least preferred): paying for third-party assessment, paying for access to a library of assessed apps, performing assessments in-house, relying on self-assessments from health app developers, and not making use of assessments. The order in which these options appeared during the ranking task was randomized across participants.

After the background questions, DCE instructions, including a detailed description of each attribute level and an example choice set of 2 assessments, were presented. The DCE was framed by asking participants to imagine that their organization had selected a health app for potential submission to a third-party quality assessment and that they had been asked to make a recommendation regarding which assessment, if any, their organization should purchase. Choice sets were presented in table format, as shown in Figures 1 and 2 but with 2 alternatives.

Participants

The study sample consisted of 41 employees of health app development companies (from hereon referred to as "health app developers" for short) and 46 health system representatives from across Europe. On the basis of power calculations, using the formula recommended by Johnson and Orme [40], which incorporates the number of attribute levels, choice sets, and alternative options, a sample size of 40 participants was required for each group.

The survey was promoted through newsletters of relevant professional organizations, such as the Healthcare Information and Management Systems Society, as well as through social media, professional networking sites, and contacts of the Label2Enable project team.

Data Analysis

Responses to the background questions were summarized using descriptive statistics. For the rankings of assessment options provided by health system representatives, chi-square values were calculated as described by Finch [41]. Owing to a technical issue, information for 1 (2%) of the 41 health app developers was missing regarding their role in decision-making as well as their organization's size and country.

Data from the 2 DCEs were analyzed using 2 separate mixed logit models, with observed assessment choices as the dependent variable and attribute levels of the presented assessments as predictors. In both models, cost (purchase price) and assessment completion time were entered as fixed, continuous variables to allow for the calculation of marginal rates of substitution without complications related to determining mean values [42]. All other attribute levels were entered as random, categorical variables, thus accounting for preference heterogeneity across participants. The categorical variables were dummy coded using level 1 (as indicated in Figures 1 and 2) as the reference category (refer to the paper by Daly et al [43] for a discussion of effects coding vs dummy coding for choice models).

The opt-out decisions, collected after the initial choice between assessments A and B, were incorporated into the analysis as recommended by Diener et al [44]. Specifically, trials on which participants selected the same assessment (A or B) for both the initial question (A vs B) and for the opt-out question (previously selected assessment [A or B] vs no assessment) were treated as 1 choice set with 3 options (A, B, and no assessment). By contrast, trials on which participants selected "no assessment" after initially choosing assessment A or B were treated as 2 choice sets with 2 (A and B) and 3 (A, B, and no assessment) options, respectively. In addition, an alternative specific constant (coded as 1 for "no assessment" and 0 for assessments A and B) was added to the model as a random effect.

Parameter estimation was conducted in Stata (version 18; StataCorp LLC) [45] using 2000 Halton draws to achieve stable coefficients. A sensitivity analysis was performed in which participants who did not select the expected option in the rationality test (health app developers: 3/41, 7%; health system representatives: 5/46, 11%) were excluded. The estimated coefficients for this analysis are presented in Multimedia Appendix 1. In the Results section, the estimates across all participants are reported. This is recommended practice, given that the deletion of *potentially* "irrational" responses may lead to the introduction of biases and the reduction of statistical efficiency and power [46].

Moreover, relative attribute importance was calculated by first subtracting the coefficient of the least preferred level from the coefficient of the most preferred level for each attribute and

XSL•FO
RenderX

then dividing this range for a given attribute by the sum of these ranges across all attributes.

Mean marginal willingness to pay and marginal rates of substitution for assessment completion time were calculated by dividing the estimated model coefficients of the other attributes by the negative coefficients of cost and assessment completion time, respectively. SEs for these values were calculated using the delta method, as instantiated by the *nlcom* function in Stata.

In addition, predicted choice probabilities were examined for the best-, middle-, and worst-case scenarios, with all attributes set to the most preferred, middle, and least preferred levels, respectively, using the methodology and Stata code described by Lancsar et al [42].

Finally, a linear regression analysis was conducted to examine whether responses to the background questions predicted the percentage of times participants opted for "no assessment" during the DCE (refer to Multimedia Appendix 1 for the details and results of this analysis).

Ethical Considerations

The study was reviewed and approved by the Biomedical and Scientific Research Ethics Committee of the University of Warwick (BSREC 82/22-23).

All participants provided informed consent at the start of the survey, after being presented with an information sheet outlining the purpose, methodology, and data-handling arrangements of the study. Participants could withdraw from the study at any point before submitting their responses, in which case none of the survey data were saved. All collected data were fully anonymous. No personally identifiable information was collected at any stage of the survey. Participants were not offered any compensation for completing the survey. This was made clear in the information sheet.

Results

Health App Developer Survey

Participant Characteristics

Health app developers were located across 18 European countries, including France (8/40, 20%), Germany (4/40, 10%), the Netherlands (4/40, 10%), the United Kingdom (4/40, 10%), Denmark (3/40, 8%), and Portugal (3/40, 8%; refer to Table S2 in Multimedia Appendix 1 for the full list of countries).

Most health app developers were based in micro (<10 employees; 11/40, 28%) or small (10-49 employees; 18/40, 45%) companies, rather than medium (50-249 employees; 6/40, 15%), large (250-999 employees; 2/40, 5%), or very large (≥1000 employees; 3/40, 8%) organizations.

Of the apps developed by the participants' companies, 37% (15/41) provided operational support (eg, enabling web-based consultation with health care professionals or offering administrative services), 46% (19/41) supported disease prevention (eg, promoting a healthy lifestyle), 61% (25/41) facilitated self-management (eg, providing disease-specific information or symptom management aid), 66% (27/41) supported health care delivery (eg, assisting with diagnosis, clinical decision-making, treatment, or remote follow-up), 24% (10/41) offered health research support (eg, gathering patient-reported outcomes for clinical trials), and 5% (2/41) had other core functionalities not captured by the aforementioned categories.

With regard to digital health–related decisions, 18% (7/40) of the health app developers reported being the sole or final decision maker, 45% (18/40) indicated being part of a group of decision makers, and 35% (14/40) stated that they provided advice to inform decisions, while 2% (1/40) indicated having no influence on decision-making.

The majority of health app developers (32/41, 78%) had prior experience with health app assessment, which was indicated to include regulatory certification. Of these developers, 47% (15/32) reported that this experience had been mostly positive, while 34% (11/32) and 19% (6/32) stated that their experience had been neutral or mostly negative, respectively.

In addition, across all health app developers, 49% (20/41) strongly agreed and 44% (18/41) agreed that it would be advisable and beneficial for their organization to submit their health apps to a third-party quality assessment in the next 2 years, while 5% (2/41) and 2% (1/41) disagreed and strongly disagreed with this statement, respectively.

Finally, 59% (24/41) of the health app developers stated that, in their opinion, it would be preferable if it was *mandatory* for health app developers to display their app's quality assessment results or a statement that their app has *not* been assessed in app stores and marketing materials, while the remaining 41% (17/41) thought such reporting should be *voluntary*.

Model Estimation Results and Relative Attribute Importance

The mixed logit model estimation results for the health app developer DCE are presented in Table 1. All attributes made a significant contribution to assessment choices, with coefficients associated with levels 2 and 3 differing significantly from those associated with the reference level 1 in all cases, except for the levels of "20 hours developer time investment" and "moderate increase in willingness to use assessed apps."

The estimated coefficients for the sensitivity analysis, excluding participants who did not select the expected option in the rationality test, are presented in Table S3 in Multimedia Appendix 1.

All attribute levels had the expected sign and order, with health app developers, on average, preferring assessments that have a lower cost (purchase price), involve less developer time investment, make results available more quickly (ie, have a shorter assessment completion time), facilitate broader clinical care uptake via clinical guideline integration and procurement or reimbursement, offer full app store integration (ie, display of assessment scores and impact on search result rankings), and lead to a larger increase in willingness to use assessed apps.

Relative attribute importance values indicated that the largest contribution to health app developers' choices was made by the clinical care uptake attribute (40.3%), followed by cost (17.9%).

XSL•FO
RenderX

Increased willingness to use assessed apps (13.8%) and app store integration (13.6%) had a moderate impact on choices, while assessment completion time (9.5%) and developer time investment (4.8%) made the smallest contribution (refer to Figure S1 in Multimedia Appendix 1).

Table 1. Mixed logit model estimation results for health app developers.

	Mean coefficient, (SE)	*P* value	SD (SE)	*P* value
Cost (per €100)	−0.04 (0.01)	<.001	—[a]	—
Assessment completion time (per month)	−0.16 (0.03)	<.001	—	—
Developer time investment (h)				
10 (reference)	0.00 (N/A[b])	—	—	—
20	−0.38 (0.23)	.10	0.52 (0.39)	.18
40	−0.40 (0.19)	.04	0.61 (0.19)	<.001
Impact on clinical care uptake				
None (reference)	0.00 (N/A)	—	—	—
Guideline integration	1.23 (0.20)	<.001	0.63 (0.26)	.02
Guideline integration and procurement or reimbursement	3.33 (0.56)	<.001	1.92 (0.42)	<.001
App store integration				
None (reference)	0.00 (N/A)	—	—	—
Assessment score display	0.46 (0.15)	.003	0.01 (0.11)	.92
Assessment score display and impact on app search result rankings	1.12 (0.20)	<.001	0.74 (0.22)	<.001
Increase in willingness to use assessed apps				
Small (reference)	0.00 (N/A)	—	—	—
Moderate	0.35 (0.26)	.18	0.61 (0.37)	.10
Large	1.14 (0.25)	<.001	0.93 (0.27)	<.001
Alternative specific constant for “no assessment”	0.25 (0.53)	.63	2.14 (0.52)	<.001

[a]Details for the SD (SE and *P* value) are not applicable to fixed effects.

[b]N/A: not applicable (SD, SE, and *P* values are not applicable to reference levels).

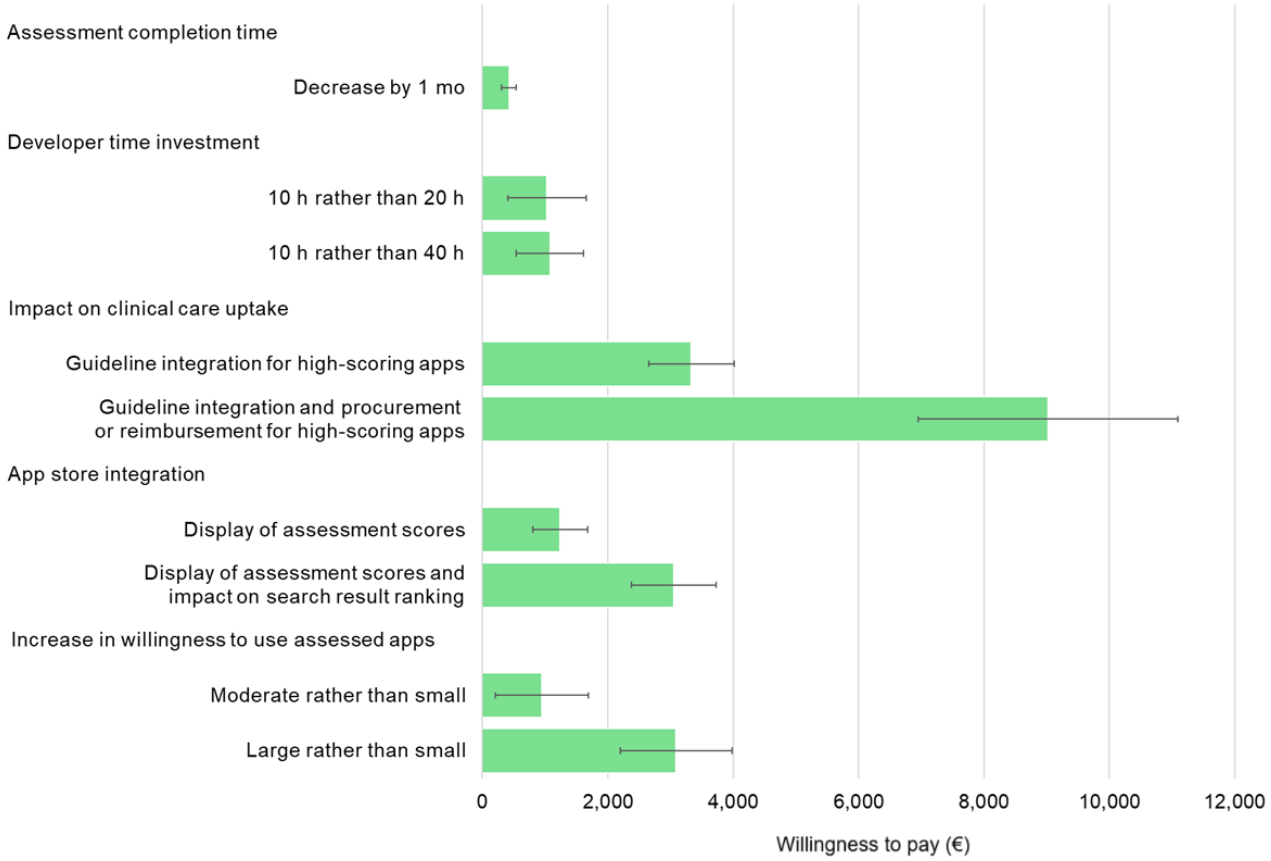
Marginal Willingness to Pay and Marginal Rates of Substitution for Assessment Completion Time

On average, health app developers were willing to pay an additional €9020 (95% CI €4968-€13,072) if an assessment facilitates guideline integration and procurement or reimbursement for high-scoring apps (ie, for apps that meet the requirements of stakeholders involved in clinical care uptake decisions). The marginal willingness to pay was only approximately a third of this amount if an assessment offers full app store integration (€3052, 95% CI €1722-€4381); or leads to a large, rather than a small, increase in willingness to use assessed apps (€3090, 95% CI €1341-€4840). Marginal willingness to pay for receiving the assessment results 1 month sooner was €425 (95% CI €196-€654; refer to Figure 3 and Table S4 in Multimedia Appendix 1 for the full results).

Moreover, health app developers were willing to wait an additional 21 months (95% CI 8-34 mo) for the evaluation results if the assessment facilitates clinical guideline integration and procurement or reimbursement for high-scoring apps. Furthermore, they were willing to wait approximately 7 months longer for assessment results if full app store integration (95% CI 3-11 mo) or a large, rather than small, increase in willingness to use assessed apps (95% CI 2-12 mo) is offered. For a cost reduction of €100, they were willing to wait an extra week (0.2 mo, 95% CI 0.1-0.4 mo) for the assessment results (refer to Table S5 in Multimedia Appendix 1 for all findings).

XSL•FO
RenderX

Figure 3. Health app developers' marginal willingness to pay for different assessment attribute levels (error bars represent SEs). Note that the currency exchange rate at the time of data collection was approximately €1=US $1.11.



Predicted Assessment Uptake

Predicted assessment uptake was examined for the best-, middle-, and worst-case scenarios, with all attribute levels set to the most preferred, middle, and least preferred levels, respectively.

For the best-case scenario, the predicted assessment uptake among health app developers was 88.6% (95% CI 77.9%-96.5%) while for the middle- and worst-case scenarios, the predicted uptake was 49.2% (95% CI 38.9%-62.8%) and 12.4% (95% CI 4.8%-22.1%), respectively.

Health System Representative Survey

Participant Characteristics

Health system representatives worked at the following types of organizations: public health care institutions (18/46, 39%), private health care institutions (10/46, 22%), health authorities (8/46, 17%), health technology assessment bodies (6/46, 13%), private health insurance companies (3/46, 7%), and public health insurance companies (1/46, 2%). These organizations were distributed across 13 European countries, including Spain (14/46, 30%), Belgium (5/46, 11%), the Netherlands (5/46, 11%), the United Kingdom (5/46, 11%), Croatia (4/46, 9%), and Lithuania (3/46, 7%; refer to Table S6 in Multimedia Appendix 1 for the full list of countries).

The organizations' size ranged from very large (≥1000 employees; 15/46, 33%) and large (250-999 employees; 6/46, 13%) to medium (50-249 employees; 12/46, 26%), small (10-49 employees; 10/46, 22%), and micro (<10 employees; 3/46, 7%). The vast majority of organizations (41/46, 89%) were currently using, promoting, reimbursing, or otherwise interacting with digital health technologies.

Most health system representatives reported being part of a group of individuals making digital health–related decisions (25/46, 54%) or providing advice to inform such decisions (14/46, 30%), while 2% (1/46) indicated being the sole or final decision maker related to digital health, and 11% (6/46) stated that they did not contribute to digital health–related decisions at their organization.

The majority of health system representatives (34/46, 74%) had previous experience with health app assessment. Of these, 76% (26/34) reported that this experience had been mostly positive, while 21% (7/34) and 3% (1/34) indicated that their experience had been neutral or mostly negative, respectively.

Furthermore, of all health system representatives, 35% (16/46) strongly agreed and 59% (27/46) agreed that it would be advisable and beneficial for their organization to make use of third-party health app quality assessments in the next 2 years, while 7% (3/46) strongly disagreed with this statement.

Finally, 80% (37/46) of the health system representatives indicated that, in their opinion, it would be preferable if it was *mandatory* for health app developers to display their app's quality assessment results, or a statement that their app has *not* been assessed, in app stores and marketing materials, while the

XSL•FO
RenderX

remaining 20% (9/46) thought that such reporting should be *voluntary*.

Separate response summaries for health care institution employees and representatives of other organization types can be found in Table S7 in Multimedia Appendix 1.

Ranking of Assessment Options

Examining the mean rankings (from 1=most preferred to 5=least preferred) of different ways to engage with health app quality assessments revealed that, on average, health system representatives' most preferred option was to have assessments conducted in-house (mean 2.04, SD 1.10), followed by paying for third-party assessment (mean 2.39, SD 0.98) or access to a library with assessed apps (mean 2.70, SD 1.38). Relying on developer self-assessments (mean 3.35, SD 1.06) or not making use of assessment at all (mean 4.52, SD 1.07) were, on average, the least preferred options. The same pattern of preferences was observed for both health care institution employees and other representatives when the rankings of these groups were considered separately. A chi-square test across all health system representatives indicated that the rankings were not random (χ^2_4=70.2; P<.001). The number of times each option received a given rank is displayed in Table 2.

Table 2. Indication of the frequency with which assessment options were given each rank by health system representatives, with rank 1 indicating the most preferred option and rank 5 the least preferred option (N=46).

Assessment options	Ranks, n (%)				
	1	2	3	4	5
Paid third-party assessment	9 (20)	16 (35)[a]	16 (35)[a]	4 (9)	1 (2)
Paid access to a library with assessed apps	11 (24)	13 (28)[a]	7 (15)	9 (20)	6 (13)
In-house assessment	20 (43)[a]	10 (22)	10 (22)	6 (13)	0 (0)
Developer self-assessment	4 (9)	5 (11)	11 (24)	23 (50)[a]	3 (7)
Not making use of assessment	2 (4)	2 (4)	2 (4)	4 (9)	36 (78)[a]

[a]Highest ranking frequency for each option.

Model Estimation Results and Relative Attribute Importance

The mixed logit model estimation results for the health system representative DCE are shown in Table 3. All attributes made a significant contribution to health system representatives' choices, with coefficients associated with levels 2 and 3 differing significantly from those associated with the reference level 1 in all cases. All attribute levels had the expected sign and order, with health system representatives, on average, preferring assessments that have a lower cost (purchase price); make results available more quickly (ie, have a shorter assessment completion time); offer reassessment after substantial updates, rather than every year or every 2 years; are recommended by a higher proportion of peers; and lead to a larger increase in willingness to use assessed apps.

XSL•FO
RenderX

Table 3. Mixed logit model estimation results for health system representatives.

	Mean coefficient (SE)	*P* value	SD (SE)	*P* value
Cost (per €100)	−0.03 (0.01)	<.001	__[a]	—
Assessment completion time (per mo)	−0.16 (0.03)	<.001	—	—
Reassessment frequency				
After substantial app updates (reference)	0.00 (N/A[b])	—	—	—
Every y	−0.33 (0.17)	.04	0.06 (0.24)	.81
Every 2 y	−0.44 (0.21)	.04	0.66 (0.20)	<.001
Peers recommending the assessment (%)				
50 (reference)	0.00 (N/A)	—	—	—
70	0.79 (0.25)	.002	0.58 (0.51)	.25
90	0.82 (0.18)	<.001	0.68 (0.26)	.01
Increase in willingness to use assessed apps				
Small (reference)	0.00 (N/A)	—	—	—
Moderate	1.22 (0.22)	<.001	0.39 (0.52)	.46
Large	2.14 (0.39)	<.001	1.12 (0.42)	.01
Alternative specific constant for “no assessment”	−1.56 (1.18)	.19	2.49 (0.59)	<.001

[a]Details for the SD (SE and *P* value) are not applicable to fixed effects.

[b]N/A: not applicable (SD, SE, and *P* values are not applicable to reference levels).

The estimated coefficients for the sensitivity analysis, excluding participants who did not select the expected option in the rationality test, are presented in Table S8 in Multimedia Appendix 1.

Relative attribute importance values indicated that increased willingness to use assessed apps had the largest impact on health system representatives’ choices (39.3%), followed by cost (22.3%). Assessment completion time (15.1%) and the percentage of peers recommending the assessment (15.1%) made a moderate contribution to choices, while reassessment frequency had the smallest impact (8.1%; refer to Figure S2 in Multimedia Appendix 1).

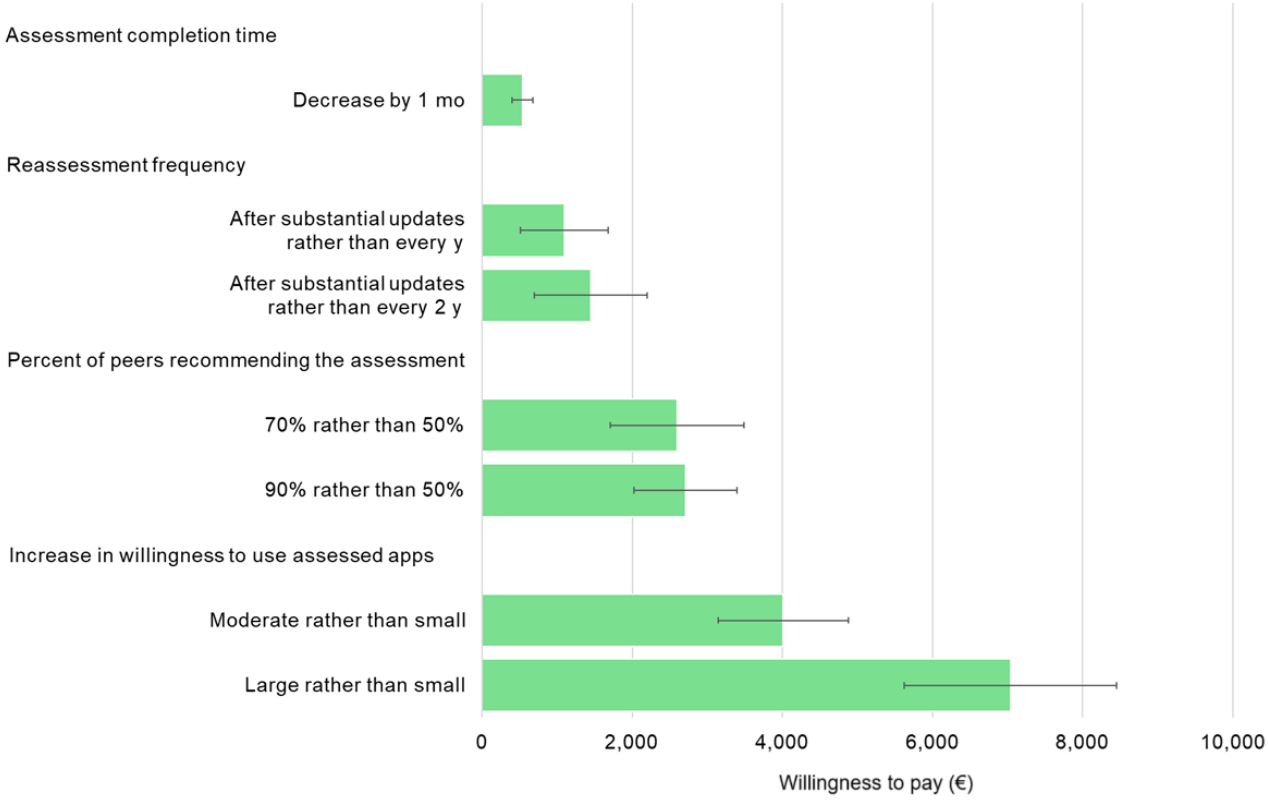
Marginal Willingness to Pay and Marginal Rates of Substitution for Assessment Completion Time

On average, health system representatives were willing to pay an additional €7037 (95% CI €4267-€9806) if an assessment yields a large, rather than a small, increase in willingness to use assessed apps; and €2709 more (95% CI €1355-€4063) if the assessment is recommended by 90%, rather than 50%, of peers. Marginal willingness to pay for app reassessment after substantial updates, rather than every 2 years, was €1448 (95% CI −€25 to €2921); and a decrease in assessment completion time of 1 month was, on average, worth €542 (95% CI €269-€816) to health system representatives (refer to Figure 4 and Table S9 in Multimedia Appendix 1 for the full results).

Furthermore, health system representatives were willing to wait an additional 13 months (95% CI 7-19 mo) for the assessment results if the assessment is associated with a large, rather than a small, increase in willingness to use assessed apps; and a further 5 months (95% CI 2-8 mo) if the assessment is recommended by 90%, rather than 50%, of peers. For a cost reduction of €100, health system representatives were willing to wait an additional week (0.2 mo; 95% CI 0.1-0.3 mo; refer to Table S10 in Multimedia Appendix 1 for the full results).

XSL•FO
RenderX

Figure 4. Health system representatives' marginal willingness to pay for different assessment attribute levels (error bars represent SEs). Note that the currency exchange rate at the time of data collection was approximately €1=US $1.11.



Predicted Assessment Uptake

The predicted assessment uptake was examined for the best-, middle-, and worst-case scenarios, with all attribute levels set to the most preferred, middle, and least preferred levels, respectively.

For the best-case scenario, the predicted assessment uptake among health system representatives was 91.1% (95% CI 68%-99.2%), while for the middle- and worst-case scenarios, the predicted uptake was 78.7% (95% CI 48.7%-94.2%) and 35% (95% CI 11.8%-65.8%), respectively.

Discussion

Principal Findings

The aim of this study was to examine preferences and willingness to pay for health app assessments with different value propositions among potential customers of such assessments, namely health app developers and health system representatives.

The DCEs revealed that, among health app developers, the attributes with the largest impact on assessment choices were clinical care uptake (guideline integration and reimbursement or procurement) and cost (purchase price). Increased willingness to use assessed apps and app store integration had a moderate impact on choices, and assessment completion time and developer time investment made the smallest contribution. Among health system representatives, increased willingness to use assessed apps had the largest impact on participants' choices, followed by cost. Assessment completion time and the percentage of peers recommending the assessment made a moderate contribution to choices, and reassessment frequency had the smallest impact.

On average, health app developers were willing to pay an additional €9020 (95% CI €4968-€13,072) if an assessment facilitates guideline integration and procurement or reimbursement. The marginal willingness to pay was only approximately a third of this amount if an assessment offers full app store integration (assessment score display and impact on app search result rankings; €3052, 95% CI €1722-€4381); or leads to a large, rather than a small, increase in willingness to use assessed apps (€3090; 95% CI €1341-€4840). Health system representatives were, on average, willing to pay €7037 (95% CI €4267-€9806) more if an assessment results in a large, rather than a small, increase in willingness to use assessed apps; and an additional €2709 (95% CI €1355-€4063) if the assessment is recommended by 90%, rather than 50%, of peers.

Attitudes and Preferences Regarding Health App Assessment Attributes

The interest in, and predicted uptake of, health app assessments was high among both health app developers and health system representatives. In both groups, a majority of the participants (health app developers: 38/41, 93%; health system representatives: 43/46, 93%) agreed that it would be advisable and beneficial for their organization to make use of third-party app assessments in the next 2 years. The predicted uptake of assessments that offered the preferred values for all attributes was 88.6% among app developers and 91.1% among health system representatives. This indicates that both groups are viable

XSL•FO
RenderX

target customers for health app assessments if their expectations regarding cost and benefits are met.

In this context, it is worth noting that health system representatives' most preferred option for engaging with app assessments was to have apps evaluated in-house. This may partly be the case because approximately half of the participants (21/46, 46%) in this group worked at large or very large organizations that may have their own product evaluation processes in place. For instance, large health care institutions may contain procurement or IT departments that routinely evaluate products, including digital health technologies. Where such a trusted evaluation process is available in-house, this may be the routine approach for assessing health care products in general and, as such, also the preferred approach for ensuring the quality of used health apps. Nevertheless, making use of third-party assessment was, on average, the second most favored option. Together with the DCE results, this suggests a solid interest in app assessments from health system organizations, as long as a suitable cost-benefit balance is offered.

Specifically, the DCE results indicate that, to maximize uptake among health system organizations, it is important to demonstrate that an assessment substantially increases the public's and health care professionals' willingness to use health apps. This impact could, for instance, be demonstrated by conducting a controlled study in which participants are presented with hypothetical apps that either carry or do not carry a quality assessment label and are asked whether they would use or recommend these apps. A recent study using this approach showed that health care professionals' willingness to recommend health apps was significantly higher for apps that carried a quality label than for those that did not [47]. Besides this direct effect of a quality label, there may also be a more indirect impact of assessment on health care professionals' willingness to use evaluated apps *if* the assessment enables the integration of health apps into clinical guidelines. This may further encourage health app use and recommendation, given that health care professionals rely on clinical guidelines to inform their practice [48,49].

Similarly, assessment may also increase the use of health apps among the public. Yet-to-be-published research from the Label2Enable project shows that nearly 9 in 10 surveyed European citizens think that the government should review and rate apps or pay another organization to do so to help them choose a health app. This indicates a high demand for guidance regarding health app choices among the public. In the future, this research could be extended to further quantify the increase in willingness to use health apps as a result of trust in and visibility, transparency, and clarity of different quality assessments.

Our findings further show that enabling or facilitating the clinical guideline integration and reimbursement or procurement of assessed apps is an important factor for maximizing assessment uptake among health app developers. This requires the close collaboration of assessment providers with health authorities, health care institutions, and health insurers. While reimbursement of health apps is arguably still in its early stages, reimbursement mechanisms have started to emerge across Europe, including through the German digital health app assessment [50], the Belgian National Institute for Health and Disability Insurance assessment [51], and the French "early access to reimbursement for digital devices" program [52]. Our findings underscore that, as expected, such reimbursement programs provide a substantial incentive for developers to submit their apps for associated quality assessments. However, even if reimbursement is offered, assessment scaling can be a challenge [48,53]. Therefore, it is important to also optimize other assessment characteristics, such as cost, developer time investment, and assessment completion time, as well as the visibility of and trust in the assessment. In addition, further efforts are needed to encourage professional medical societies and other relevant bodies to include health apps in clinical recommendations and guidelines [48,49].

Another attribute with a moderate impact on developers' assessment choices was app store integration. Notably, such integration has substantial benefits beyond being an incentive for app developers to opt for quality assessment. Currently, search result rankings in major app stores are influenced by download numbers and user ratings, among other factors [54]. However, research has shown that these metrics are not significantly correlated with health app quality, as determined by in-depth assessment [55]. Given that users rarely consider more than the first 5 to 10 apps listed in the search results before making a choice [56], the current ranking system may indirectly encourage the use of potentially ineffective or unsafe health apps. This could be partially counteracted by incorporating trusted third-party quality assessment scores into search result ranking algorithms, ideally placing a higher weight on such scores than on download numbers and user ratings.

A further attribute that warrants consideration beyond its influence on assessment choices is reassessment frequency. Perhaps somewhat surprisingly, this attribute had the smallest relative impact on health system representatives' assessment choices. A potential contributing factor to this finding may have been uncertainty regarding how frequently reassessment would take place if it was contingent on substantial app updates that affect privacy or health-related risks, which was the reference category for the DCE analysis. It was not possible to specify the reassessment frequency for this scenario because this likely varies widely across apps. Recent data show that the median time since the last update among the highest-ranking health and fitness apps in the Google Play Store is just 14 days [57]. However, updates that impact evaluated components to an extent that warrants reassessment are likely to be substantially less frequent. Nevertheless, due to the fast pace of app development, reassessment at regular intervals or after substantial app changes is important to ensure that key quality information is applicable to the most recent version of the app. If no reassessment takes place or if it is too infrequent or poorly timed, the result could be reliance on outdated information. This, in turn, could potentially lead to the recommendation of health apps that no longer meet quality standards, which could undermine trust in the assessment. Future quantitative and qualitative research could further examine how frequently health apps, on average, provide updates that affect privacy and health risks and to what extent the frequency of such updates may affect preferences for

XSL•FO
RenderX

reassessment, as well as trust in assessment results, among health system representatives and app users.

Willingness to Pay for Health App Assessments

The findings of this study provide useful insights regarding the willingness to pay for health app assessments with varying characteristics across potential customer groups. For instance, the DCE results indicate that health app developers would be willing to pay €6843 more for an assessment with the middle attribute values *compared to an assessment with the least preferred* values, that is, for an assessment requiring 20, rather than 40, hours of time investment (€1076-€1028 = €48); providing assessment results in 3, rather than 6, months (€425 × 3 = €1275); facilitating clinical guideline integration for high-scoring apps (without reimbursement; €3332); displaying assessment results in app stores (without impact on search result ranking; €1240); and resulting in a moderate, rather than a small, increase in willingness to use assessed apps (€948). Similarly, health system representatives would be willing to pay €8587 more for a "middle case" assessment *compared to one with the least preferred attribute values*, that is, for an assessment that provides results in 3, rather than 6, months (€542 × 3 = €1626); is recommended by 70%, rather than 50%, of their peers (€2600); offers reassessment every year, rather than every 2 years (€1448-€1098 = €350); and leads to a moderate, rather than a small, increase in willingness to use assessed apps (€4011).

In this context, it should be noted that the DCEs used in this study only investigated willingness to pay for one-off assessments to avoid dependencies between attributes, such as between cost, time investment, and reassessment frequency. For this reason, the health system DCE instructed participants to assume that reassessment costs would not be covered by their organization. Therefore, a further examination of how reassessment frequency affects willingness to pay is warranted. This is especially important given the fast pace of digital health development, as mentioned previously, and considering the additional cost and resource requirements for assessment organizations that are implementing frequent reassessment.

Moreover, it should be taken into account that the willingness to pay estimates for app developers may not fully capture the necessary trade-offs in resource allocation, especially when it comes to meeting mandatory versus voluntary requirements. Specifically, ensuring that a health app meets regulatory requirements, such as those outlined by the General Data Protection Regulation and MDR, necessitates resource investment; and the more time and money have already been spent on aligning a given app to these mandatory requirements, the less willing or able developers may be to invest additional resources in voluntary quality assessment. This point was raised by several participants in a free-text comment section at the end of the current survey (eg, "We have already spent upwards of 200,000 Euros obtaining our MDR certification - why should we spend more?"). To address these concerns, it is advisable to not only emphasize the additional benefits offered by a given voluntary assessment but also allow for the reuse of documentation from mandatory certifications, where appropriate. This helps to minimize resource requirements and avoid duplicating the MDR related evaluation performed by notified bodies.

Relatedly, it should also be considered how resource requirements, including for covering assessment fees, affect smaller versus larger organizations. As part of this study, both assessment organization interviewees and survey participants voiced concerns that assessment costs or extensive time requirements for app developers could disadvantage smaller companies that do not have the necessary resources to pay for or invest time in the assessment process. This could negatively impact innovation by creating additional barriers for the scaling of start-ups, especially if assessment outcomes, or the lack thereof, have a substantial impact on users' health app choices. Some interviewees suggested that adjusting assessment fees based on company size or charging health system organizations instead of app developers may partially address this issue.

Next Steps

As mentioned in the Introduction section, a major challenge for current health app assessment initiatives, including national programs, is reaching a scale at which they can evaluate a substantial proportion of the large number of health apps on the market [29,31,32]. Such scaling requires both an efficient assessment process with a sufficient number of assessors and a value proposition and pricing that make the assessment attractive to a substantial number of customers, such as health app developers or health system organizations.

This study is part of a larger initiative that aims to establish such a scalable assessment program across Europe, using the quality requirements of CEN ISO/TS 82304-2. In terms of the organizational implementation of this program, it is envisioned that there will be a central governing body, alongside a distributed network of assessment organizations located across Europe, that will carry out the health app evaluations. The intention is to rely, in the first instance, on organizations that already provide digital health assessment services (using various evaluation frameworks) and, therefore, have the necessary resources and structures in place. A pilot study was recently carried out with 5 such organizations to examine their interrater reliability for a draft version of the CEN ISO/TS 82304-2 assessment, which yielded promising results [58].

To enhance the efficiency, quality, affordability, and scalability of CEN ISO/TS 82304-2 assessments and reassessments as well as to avoid a duplication of assessment efforts, automated evaluation will be used, where feasible, and the results from trusted existing assessments will be accepted as evidence for subsets of CEN ISO/TS 82304-2 quality requirements. It is expected that the CEN ISO/TS 82304-2 assessment will be comprehensive enough to provide health care decision makers with the vast majority of the information they require, reducing their assessment efforts to merely examining a small number of context-specific requirements. This expectation is supported by a recent, comparative analysis by the Catalan government body responsible for health app evaluations, which showed that the CEN ISO/TS 82304-2 app assessment handbook covers 93% of the evaluation criteria considered important by the Catalan assessment body [59].

XSL•FO
RenderX

The findings of this study will be used to inform decisions regarding the pricing and value proposition of the CEN ISO/TS 82304-2 assessment. The study showed that both health app developers and health system representatives widely agree that their organizations would benefit from health app assessments and that both groups demonstrate a sufficient willingness to pay for such assessments if their preferences are met. Therefore, both health app companies and health system organizations will likely be targeted as customers of the CEN ISO/TS 82304-2 assessment. In addition, it will be explored how the preferences expressed by these stakeholders as part of this study can be realized to ensure optimal assessment uptake.

The hope is that through the efficient, distributed implementation of the app evaluation and the alignment to customer preferences, the CEN ISO/TS 82304-2 assessment program will eventually be able to provide evaluation results for a substantial proportion of health apps available on the market while using a thorough and consistent assessment approach.

Limitations

A few limitations of this study should be noted. First, participants who took the time to complete the DCE survey were likely already interested in or well informed about the topic of health app assessment. Indeed, this study's results indicate that the majority of participants had previous experience with, and a positive attitude toward, health app assessment. Therefore, the recruited sample may not be entirely representative of all health app developers or health system representatives across Europe. It is possible that willingness to pay and assessment uptake may, in reality, be lower than indicated by these findings (eg, due to a greater negative effect of assessment cost or more opt-out decisions in the larger population compared to this study's sample). Moreover, it is possible that other attributes that were not included in this study may be of importance to some stakeholders, the inclusion of which could result in differing preferences and willingness to pay values. Future research could examine this further by actively recruiting participants with less awareness of, and less positive attitudes toward, digital health and quality assessments.

Second, due to the limited sample size, we were not able to perform subgroup analyses for participants with different characteristics (eg, those with previous positive, negative, or no assessment experience) or for different organization types (eg, for health care institutions, authorities, insurers, or health technology assessment bodies; or for health app developers with varying business models). Future research is warranted to examine whether preferences and willingness to pay for health app assessments differ across these subgroups.

Finally, when interpreting the study findings, it should be kept in mind that the estimates for health app developers and health system representatives cannot be directly compared. This is partly due to the use of different attributes within the DCEs, which was necessary to account for the varying relevance of different attributes across the 2 groups (refer to the Methods section for details).

Conclusions

This study aimed to examine preferences and willingness to pay for health app assessments with different value propositions among potential customers of such assessments. Among both app developers and health system representatives, interest in, and the predicted uptake of, health app assessments was high if the most preferred assessment attribute values were offered. Moreover, for health app developers and health system representatives, the attributes with the largest impact on assessment choices were facilitation of clinical care uptake (guideline integration and reimbursement or procurement) and increased willingness to use assessed apps, respectively. The study's findings provide valuable information, including marginal willingness to pay estimates, that can be used to inform decision-making and support the scaling of health app assessments.

Acknowledgments

The authors are grateful to the following individuals who provided input to the discrete choice experiment design as part of the assessment organization interviews: Vidar Eltoft, Lauren Harkins, Jari Haverinen, Rhodri Joyce, Søren Lange Nielsen, Arve Paulsen, Danny Van Roijen, and Monique Westerlaken. In addition, the authors would like to thank everyone who participated in the study or supported the promotion of the survey. This study was funded by the European Union (grant 101057522). The views and opinions expressed are those of the authors and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

All authors are involved in the Label2Enable project, which supports the implementation of the CEN ISO/TS 82304-2 health app quality assessment framework and related label across Europe. In addition, at the time of this study, ALF and SL were employed by the Organisation for the Review of Care and Health Apps, which provides health app assessments.

Multimedia Appendix 1
Supplementary methods and results.
[PDF File (Adobe PDF File), 450 KB - mhealth_v13i1e57474_app1.pdf]

XSL•FO
RenderX

References

1. Global strategy on digital health 2020-2025. World Health Organization. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2023-11-04]
2. Butcher CJ, Hussain W. Digital healthcare: the future. Future Healthc J 2022 Jul 28;9(2):113-117. [doi: 10.7861/fhj.2022-0046]
3. 2019 global health care outlook shaping the future. Deloitte. URL: https://www2.deloitte.com/tr/en/pages/life-sciences-and-healthcare/articles/2019-global-healthcare-outlook.html [accessed 2023-10-19]
4. Digital health trends 2021: innovation, evidence, regulation, and adoption. IQVIA. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021 [accessed 2023-11-04]
5. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
6. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct 09;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
7. Mao Y, Lin W, Wen J, Chen G. Impact and efficacy of mobile health intervention in the management of diabetes and hypertension: a systematic review and meta-analysis. BMJ Open Diabetes Res Care 2020 Sep;8(1):69 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001225] [Medline: 32988849]
8. Stevens S, Gallagher S, Andrews T, Ashall-Payne L, Humphreys L, Leigh S. The effectiveness of digital health technologies for patients with diabetes mellitus: a systematic review. Front Clin Diabetes Healthc 2022 Oct 24;3:936752 [FREE Full text] [doi: 10.3389/fcdhc.2022.936752] [Medline: 36992773]
9. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297. [doi: 10.2196/12297] [Medline: 30664494]
10. Qin M, Chen B, Sun S, Liu X. Effect of mobile phone app-based interventions on quality of life and psychological symptoms among adult cancer survivors: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Dec 19;24(12):e39799 [FREE Full text] [doi: 10.2196/39799] [Medline: 36534460]
11. Luo X, Chen Y, Chen J, Zhang Y, Li M, Xiong C, et al. Effectiveness of mobile health-based self-management interventions in breast cancer patients: a meta-analysis. Support Care Cancer 2022 Mar;30(3):2853-2876. [doi: 10.1007/s00520-021-06568-0] [Medline: 34561732]
12. Hui CY, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 01;24(3):619-632 [FREE Full text] [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
13. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
14. Jiang X, Ming WK, You JH. The cost-effectiveness of digital health interventions on the management of cardiovascular diseases: systematic review. J Med Internet Res 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: 10.2196/13166] [Medline: 31210136]
15. Socio-economic impact of mHealth: an assessment report for the European Union. PwC. 2013. URL: https://www.gsma.com/iot/wp-content/uploads/2013/06/Socio-economic_impact-of-mHealth_EU_14062013V2.pdf [accessed 2023-10-20]
16. Communication on enabling the digital transformation of health and care in the Digital Single Market, empowering citizens and building a healthier society. European Commission. 2018. URL: https://digital-strategy.ec.europa.eu/en/library/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering [accessed 2023-10-20]
17. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
18. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 05;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]
19. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 05;43:117-134 [FREE Full text] [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
20. Baines R, Hoogendoorn P, Stevens S, Chatterjee A, Ashall-Payne L, Andrews T, et al. Navigating medical device certification: a qualitative exploration of barriers and enablers amongst innovators, notified bodies and other stakeholders. Ther Innov Regul Sci 2023 Mar 04;57(2):238-250 [FREE Full text] [doi: 10.1007/s43441-022-00463-4] [Medline: 36194368]
21. Position on the European health data space. Standing Committee of European Doctors. URL: https://www.cpme.eu/api/documents/adopted/2022/11/cpme.2022-065.FINAL.CPME.position.EHDS.pdf [accessed 2023-11-03]
22. Iqbal JD, Biller-Andorno N. The regulatory gap in digital health and alternative pathways to bridge it. Health Policy Technol 2022 Sep;11(3):100663. [doi: 10.1016/j.hlpt.2022.100663]

XSL•FO
RenderX

23. van Drongelen A, de Bruijn A, Roszek B, Vonk R. Apps under the medical devices legislation. Rijksinstituut voor Volksgezondheid en Milieu. 2019. URL: https://www.rivm.nl/publicaties/apps-under-medical-devices-legislation-apps-onder-medische-hulpmiddelen-wetgeving [accessed 2023-12-16]
24. Fan M, Yu L, Chen S, Zhou H, Luo X, Li S, et al. An empirical evaluation of GDPR compliance violations in android mHealth apps. In: Proceedings of the 31st International Symposium on Software Reliability Engineering. 2020 Presented at: ISSRE '20; October 12-15, 2020; Virtual Event p. 253-264 URL: https://ieeexplore.ieee.org/document/9251060 [doi: 10.1109/issre5003.2020.00032]
25. van Kolfschooten H. The mHealth power paradox: improving data protection in health apps through self-regulation in the European union. SSRN Journal Preprint posted online July 6, 2021 [FREE Full text] [doi: 10.2139/ssrn.3871033]
26. Hulstaert F, Pouppez C, Primus-de Jong C, Harkin K, Neyt M. Gaps in the evidence underpinning high-risk medical devices in Europe at market entry, and potential solutions. Orphanet J Rare Dis 2023 Jul 25;18(1):212 [FREE Full text] [doi: 10.1186/s13023-023-02801-7] [Medline: 37491269]
27. Shatrov K, Blankart CR. After the four-year transition period: is the European Union's medical device regulation of 2017 likely to achieve its main goals? Health Policy 2022 Dec;126(12):1233-1240 [FREE Full text] [doi: 10.1016/j.healthpol.2022.09.012] [Medline: 36202647]
28. D2.1 knowledge tool 1: health apps assessment frameworks. European mHealth Hub. URL: https://mhealth-hub.org/download/d2-1-knowledge-tool-1-health-apps-assessment-frameworks [accessed 2023-10-20]
29. Frey AL, Phillips B, Baines R, McCabe A, Elmes E, Yeardsley-Pierce E, et al. Domain coverage and criteria overlap across digital health technology quality assessments: a systematic review. Health Technol 2024 Dec 20;15(1):15-27. [doi: 10.1007/S12553-024-00934-8]
30. Consumer health apps and digital health tools proliferate, improving quality and health outcomes for patients. IQVIA. URL: https://www.iqvia.com/newsroom/2021/07/consumer-health-apps-and-digital-health-tools-proliferate-improving-quality-and-health-outcomes-for [accessed 2023-10-20]
31. Commission staff working document - impact assessment report accompanying the proposal for a regulation of the European Parliament and of the Council on the European Health Data Space. European Commission. URL: https://health.ec.europa.eu/system/files/2022-05/ehealth_ehds_2022ia_1_en_0.pdf [accessed 2024-04-29]
32. Essén A, Stern AD, Haase CB, Car J, Greaves F, Paparova D, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):31 [FREE Full text] [doi: 10.1038/s41746-022-00573-1] [Medline: 35304561]
33. Hoogendoorn P, Versluis A, van Kampen S, McCay C, Leahy M, Bijlsma M, et al. What makes a quality health app-developing a global research-based health app quality assessment framework for CEN-ISO/TS 82304-2: Delphi study. JMIR Form Res 2023 Jan 23;7:e43905 [FREE Full text] [doi: 10.2196/43905] [Medline: 36538379]
34. About Label2Enable. Label2Enable. URL: https://label2enable.eu/about-the-project [accessed 2023-10-02]
35. Wang Y, Wang Z, Wang Z, Li X, Pang X, Wang S. Application of discrete choice experiment in health care: a bibliometric analysis. Front Public Health 2021;9:673698 [FREE Full text] [doi: 10.3389/fpubh.2021.673698] [Medline: 34150710]
36. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011 Jun;14(4):403-413 [FREE Full text] [doi: 10.1016/j.jval.2010.11.013] [Medline: 21669364]
37. Campbell D, Erdem S. Including opt-out options in discrete choice experiments: issues to consider. Patient 2019 Feb;12(1):1-14. [doi: 10.1007/s40271-018-0324-6] [Medline: 30073482]
38. Veldwijk J, Lambooij MS, de Bekker-Grob EW, Smit HA, de Wit GA. The effect of including an opt-out option in discrete choice experiments. PLoS One 2014;9(11):e111805 [FREE Full text] [doi: 10.1371/journal.pone.0111805] [Medline: 25365169]
39. Ngene 1.2 user manual & reference guide. ChoiceMetrics. 2018. URL: https://www.choice-metrics.com/ [accessed 2024-04-29]
40. Johnson R, Orme B. Getting the most from CBC. Sequim: Sawtooth Software Research Paper Series. 2003. URL: https://sawtoothsoftware.com/resources/technical-papers/getting-the-most-from-cbc [accessed 2024-04-29]
41. Finch H. An introduction to the analysis of ranked response data. Pract Assess Res Eval 2022;27(1):33. [doi: 10.7275/tgkh-qk47]
42. Lancsar E, Fiebig DG, Hole AR. Discrete choice experiments: a guide to model specification, estimation and software. Pharmacoeconomics 2017 Jul;35(7):697-716 [FREE Full text] [doi: 10.1007/s40273-017-0506-4] [Medline: 28374325]
43. Daly A, Dekker T, Hess S. Dummy coding vs effects coding for categorical variables: clarifications and extensions. J Choice Model 2016 Dec;21:36-41. [doi: 10.1016/j.jocm.2016.09.005]
44. Diener C, Orme B, Yardley D. Dual response "none" approaches: theory and practice. In: Proceedings of the 2006 Conference on Sawtooth Software. 2006 Presented at: Sawtooth '06; March 29-31, 2006; Delray Beach, FL p. 157-168 URL: https://tinyurl.com/2fza58em
45. Stata statistical software: release 18. StataCorp. URL: https://www.stata.com/ [accessed 2024-04-29]
46. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ 2006 Aug 13;15(8):797-811. [doi: 10.1002/hec.1104] [Medline: 16615039]

XSL•FO
RenderX

47. Biliunaite I, van Gestel L, Hoogendoorn P, Adriaanse M. Value of a quality label and European healthcare professionals' willingness to recommend health apps: an experimental vignette study. J Health Psychol 2024 Aug 02:13591053241258205 [FREE Full text] [doi: 10.1177/13591053241258205] [Medline: 39096027]
48. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians' attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021 Nov 23;9(11):e33012 [FREE Full text] [doi: 10.2196/33012] [Medline: 34817385]
49. Richter JG, Chehab G, Stachwitz P, Hagen J, Larsen D, Knitza J, et al. One year of digital health applications (DiGA) in Germany - Rheumatologists' perspectives. Front Med (Lausanne) 2022 Oct 25;9:1000668 [FREE Full text] [doi: 10.3389/fmed.2022.1000668] [Medline: 36388899]
50. Digital Health Applications (DiGA). Bundesinstitut für Arzneimittel und Medizinprodukte. URL: https://www.bfarm.de/EN/Medical-devices/Tasks/DiGA-and-DiPA/Digital-Health-Applications/_node.html [accessed 2023-10-24]
51. Mobile medical applications: submit your application. Rijksinstituut voor ziekte- en invaliditeitsverzekering. URL: https://www.riziv.fgov.be/nl/professionals/andere-professionals/fabrikanten-en-verdelers-van-medische-hulpmiddelen/mobiele-medische-toepassingen-uw-aanvraag-indienen [accessed 2023-12-16]
52. Early access to reimbursement for digital devices (PECAN). Guichet National de l'Innovation et des Usages en e-Santé. URL: https://gnius.esante.gouv.fr/en/financing/reimbursement-profiles/early-access-reimbursement-digital-devices-pecan [accessed 2023-10-23]
53. Nicol-Schwarz K. DiGA promised German digital health startups access to 73m patients — but slow insurers and poor adoption hold it back. Sifted EU. URL: https://sifted.eu/articles/diga-promised-german-healthtechs-access-to-73m-patients-but-insurer-roadblocks-and-slow-adoption-are-limiting-its-potential [accessed 2023-12-16]
54. Rotstein G. What are the real app store ranking factors? yellowHEAD. URL: https://www.yellowhead.com/blog/what-are-the-real-app-stores-ranking-factors/ [accessed 2023-10-24]
55. Hyzy M, Bond R, Mulvenna M, Bai L, Frey AL, Carracedo JM, et al. Don't judge a book or health app by its cover: user ratings and downloads are not linked to quality. PLoS One 2024 Mar 4;19(3):e0298977 [FREE Full text] [doi: 10.1371/journal.pone.0298977] [Medline: 38437233]
56. Dogruel L, Joeckel S, Bowman ND. Choosing the right app: an exploratory perspective on heuristic decision processes for smartphone app selection. Mob Media Commun 2014 Dec 29;3(1):125-144. [doi: 10.1177/2050157914557509]
57. Number of days since the highest-ranking mobile apps in the Google Play Store were last updated worldwide as of February 2022, by category. Statista. 2022. URL: https://www.statista.com/statistics/1296548/update-frequency-top-android-apps-by-category/ [accessed 2023-10-24]
58. Frey AL, Matei D, Phillips B, McCabe A, Fuller R, Laibarra B, et al. Testing and iterative improvement of the CEN ISO/TS 82304-2 health app quality assessment: pilot interrater reliability study. JMIR Form Res 2025 Mar 10;9:e64565 [FREE Full text] [doi: 10.2196/64565] [Medline: 40063936]
59. Llebot CB, Hoogendoorn P, Villalobos-Quesada M, Pratdepàdua BC. Integration of CEN ISO/TS 82304-2 for existing health authorities? App assessment frameworks: a comparative case study in Catalonia. JMIR mHealth uHealth (Forthcoming) 2024 [FREE Full text] [doi: 10.2196/67858]

Abbreviations

CE: Conformité Européenne
DCE: discrete choice experiment
MDR: Medical Device Regulation

Edited by L Buis; submitted 17.02.24; peer-reviewed by C Gibson, S Senanayake; comments to author 20.09.24; revised version received 27.09.24; accepted 10.03.25; published 26.05.25.

Please cite as:
Frey AL, Leigh S, Toro C, Pratdepàdua Bufill C, McCay C, Prenđa Trupec T, D'Avenio G, Kok M, Montvila A, Goedecker P, Hoogendoorn P
Preferences and Willingness to Pay for Health App Assessments Among Health Care Stakeholders: Discrete Choice Experiment
JMIR Mhealth Uhealth 2025;13:e57474
URL: https://mhealth.jmir.org/2025/1/e57474
doi:10.2196/57474
PMID:

XSL•FO
RenderX

©Anna-Lena Frey, Simon Leigh, Carla Toro, Carme Pratdepàdua Bufill, Charles McCay, Tatjana Prenđa Trupec, Giuseppe D'Avenio, Menno Kok, Antanas Montvila, Philipp Goedecker, Petra Hoogendoorn. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Economic Evaluation of a Multicomponent mHealth Intervention for Stroke Management in Rural China: Cluster-Randomized Trial With 6-Year Follow-Up

Bolu Yang[1,2*], MSc; Enying Gong[3,4*], PhD, MSc; Xingxing Chen[1,2], MSc; Jie Tan[1,2], MSc; Nicholas Peoples[5], MSc, MD; Yuhan Li[2], MSc; Jiayu Cai[2], BS; Yan Li[6,7], PhD; Brian Oldenburg[8,9], PhD; Chen Chen[1], PhD; Dejin Dong[10], BS; Xiaochen Zhang[2], PhD; Eric Finkelstein[11,12], PhD; Lei Si[13,14*], PhD; Lijing L Yan[1,2,12*], PhD

[1]School of Public Health, Wuhan University, Wuhan, China

[2]Global Health Research Center, Duke Kunshan University, Kunshan, China

[3]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[4]State Key Laboratory of Respiratory Health and Multimorbidity, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[5]Harvard-affiliated Emergency Medicine Residency, Mass General Brigham, Boston, MA, United States

[6]School of Public Health, Shanghai Jiao Tong University, Shanghai, China

[7]Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[8]Baker Heart and Diabetes Institute, Melbourne, Australia

[9]La Trobe University, Melbourne, Australia

[10]Center for Disease Control and Prevention Xingtai, Xingtai, China

[11]Programme in Health Services & Systems Research, Duke-NUS Medical School, Outram, Singapore

[12]Duke Global Health Institute, Duke University, Durham, NC, United States

[13]Translational Health Research Institute, Western Sydney University, Penrith, Australia

[14]School of Health Sciences, Western Sydney University, Campbelltown, Australia

[*]these authors contributed equally

Corresponding Author:
Lijing L Yan, PhD
School of Public Health
Wuhan University
115 Donghu Road
Wuhan, 430071
China
Phone: 86 51236657057
Email: lijing.yan@duke.edu

Abstract

Background: To bridge the gap between clinical guidelines and suboptimal stroke management in rural settings, we conducted an implementation trial using evidence-based, mobile health–enabled strategies to empower primary care providers in rural China. The system-integrated and digital technology–enabled model of care (SINEMA) model was shown to significantly reduce blood pressure and mortality among people with stroke in rural China.

Objective: This study aimed to evaluate the cost-effectiveness of the SINEMA intervention within both the active trial and the post-trial observational periods and its budget impact for potential nationwide scalability.

Methods: In the cluster-randomized implementation trial (the SINEMA trial), 50 villages were randomized to either a 1-year intervention (2017-2018) or usual care, with 1299 patients with stroke followed up until 2022-2023—6 years after the trial baseline. The incremental cost-effectiveness ratios (ICER) for systolic blood pressure reduction and quality-adjusted life year gains were estimated from a health sector perspective. Both probabilistic and deterministic sensitivity analyses were conducted to assess the robustness of the findings. Additionally, a budget impact analysis was performed from a public payer perspective to estimate the per-capita and total costs of national scale-up under 2 scenarios: a standalone intervention and integration into the existing basic public health service system.

XSL•FO
RenderX

Results: The ICER per 1 mmHg systolic blood pressure reduction was $8.4 for the within-trial estimation. The ICER per quality-adjusted life year gained was $837.9 within-trial and $727.9 post-trial, both highly cost-effective relative to any commonly adopted thresholds and robust in sensitivity analyses. The first-year budget impact ranged from $115.6 million to $197.7 million in the 2 scenarios, reducing to $46.6 million to $78.7 million by year 5, with a per-capita cost of $0.03-$0.06.

Conclusions: Our findings demonstrate that the SINEMA intervention was cost-effective during the trial period and remained so throughout the 6-year sustainability observation period. These results highlight the potential of adopting similar health system–integrated, mobile health–enabled strategies to enhance the management of stroke and other chronic diseases in resource-limited settings.

Trial Registration: ClinicalTrials.gov NCT0318585, ClinicalTrials.gov NCT05792618; https://clinicaltrials.gov/study/NCT03185858 and https://clinicaltrials.gov/study/NCT05792618

International Registered Report Identifier (IRRID): RR2-10.3389/fneur.2023.1145562

(JMIR Mhealth Uhealth 2025;13:e75326) doi:10.2196/75326

KEYWORDS

economic evaluation; stroke secondary prevention; budget impact analysis; mHealth; cluster-randomized trial

Introduction

Stroke is a leading cause of mortality and disability, representing a major global health burden with approximately 4.4 million deaths annually [1]. Beyond mortality, stroke survivors incur substantial direct and indirect economic costs, which impose a significant financial burden on individuals, families, and health care systems [2,3]. In rural China, the number of stroke survivors is higher than in urban areas, with a stroke prevalence of approximately 1291.1 cases per 100,000 individuals, and this number continues to rise, encompassing around 5 million survivors [4]. To address this issue, significant evidence exists regarding management and intervention guidelines for stroke survivors in China [5-7]. However, there is a lack of practical implementation to test the effectiveness of these strategies in preventing stroke recurrence and providing essential care, particularly in resource-constrained settings [8]. To address such unmet health gaps, we developed the system-integrated and technology-enabled model of care (SINEMA), a primary care–based intervention with digital health components, to empower primary health care providers for secondary stroke prevention [9]. Through a cluster-randomized implementation trial, we have shown that the SINEMA intervention significantly reduced systolic blood pressure (SBP), stroke recurrence, hospitalization, and disability at 12 months postbaseline [10]. Follow-up data collected in October 2022 and May 2023 highlighted the sustained effectiveness of the intervention, showing significant SBP reductions and a mortality risk reduction of over 20% [11].

Although the effectiveness of the SINEMA model has been demonstrated through a cluster-randomized implementation trial and long-term follow-up [10,11], guidelines emphasize the importance of recording on-site costs to validate the cost-effectiveness of such intervention trials [12]. For example, 2 studies on poststroke management interventions have not only proven their effectiveness but also demonstrated their cost-effectiveness—although both used different telehealth strategies [13,14]. In resource-limited settings, researchers have applied community-based workforce strategies, similar to the SINEMA model, to manage other chronic diseases such as hypertension [15-17]. While these strategies have shown both effectiveness and cost-effectiveness, they did not integrate mobile health (mHealth) technology, and only 2 studies have conducted long-term post-trial follow-up. Previous studies have demonstrated the increasing use of mHealth interventions in primary care settings for managing chronic diseases in low- and middle-income countries (LMICs) with modest evidence of effectiveness in behavioral and clinical outcomes [18], while the evidence regarding how best to manage individuals with stroke history and the cost-effectiveness of mHealth interventions remains limited [19]. Such gaps raise questions about the cost-effectiveness of integrating mHealth technology into the primary care model and the sustainability of such intervention strategies [20-22].

By leveraging comprehensive data from both the within-trial and long-term post-trial follow-up of the SINEMA trial, this study addresses the evidence gap regarding the cost-effectiveness of an mHealth technology–based primary care model for stroke secondary prevention and its sustained outcomes in resource-limited settings. Additionally, we incorporate budget impact analyses to evaluate the feasibility of scaling up similar approaches that integrate mHealth and community-based workforce strategies. Our findings provide valuable insights and practical support for implementing evidence-based chronic disease management strategies in China and other LMICs facing similar challenges.

Methods

SINEMA Trial Design and Participants

This study presents an economic evaluation based on the SINEMA trial, a cluster-randomized implementation trial conducted in a rural region of Hebei Province, China. We conducted both within-trial and post-trial economic evaluations from the health sector perspective, using data from the 1-year trial (ClinicalTrials.gov NCT03185858) and a 6-year observational follow-up postbaseline (ClinicalTrials.gov NCT05792618). Additionally, we performed a budget impact analysis from the public payer perspective.

Detailed information on trial design [9], mHealth intervention design and evaluation [23,24], within-trial results on health

XSL•FO
RenderX

outcomes [10], and long-term follow-up health outcomes [11,25] have been published [9,24]. In brief, 50 villages from 5 townships were randomly assigned in a 1:1 ratio to either the intervention or control arm, with stratification by the township. A total of 1299 participants who had been diagnosed with stroke but were in a clinically stable condition were recruited between June 23 and July 21, 2017, and followed up at 12 months. Village doctors underwent training, were equipped with the SINEMA app, and performed monthly follow-up visits for patients with stroke with support from township physicians and financial incentives. Patients with stroke received these visits either at village clinics or at home and were delivered with daily voice messages containing health education on medication adherence and physical activities. In the control arm, village doctors continued their standard practices as usual, involving general clinical care and the delivery of the basic public health services, which included quarterly follow-up visits for some patients with hypertension or diabetes at no out-of-pocket cost [26].

After concluding the 1-year intervention phase, we proceeded with follow-up data collection to assess the long-term effectiveness of the intervention in October 2022 and May 2023 – approximately 5 and 6 years after the trial baseline. The same methodologies were applied with blinded data collectors.

Study Design of Cost Effectiveness and Budget Impact Analysis

To ensure transparency, we published the study protocol for within-trial economic evaluation [27]. We adhered to the Consolidated Health Economic Evaluation Reporting Standards 2022 guidelines (Table S1 in Multimedia Appendix 2). We describe the previously published within-trial economic evaluation protocol briefly below as well as the post-trial cost-effectiveness analyses and national budget impact analyses in more detail. A CONSORT (Consolidated Standards of Reporting Trails) 2025 checklist is provided in Multimedia Appendix 3.

Assessment of Costs

Costs were categorized and calculated from the use of health care resources and program delivery costs. Discounting was only applied to post-trial economic analysis, where an annual discount rate of 5% was applied. We adjusted all costs to reflect the 2023 US dollar, using the rural consumer price index for conversion [28].

Costs associated with health care resource use include inpatient care, outpatient care, and medication expenses. For the within-trial period, inpatient costs were retrospectively collected from the admission systems of 4 major hospitals at our study sites, which captured about 98% of inpatient care in the region and were verified against patients' self-reported history of inpatient care. These records were categorized into 3 groups: stroke-related inpatient care, cardiovascular disease–related inpatient care (excluding stroke), and other causes of inpatient care (Table S2 in Multimedia Appendix 2). For the base case analysis, all-cause inpatient care costs were included in the analysis. As the data on outpatient costs from medical insurance scheme records was not available, we estimated outpatient costs by using the average cost of outpatient visits excluding the medication costs based on existing literature [29].

Medication costs were valued by multiplying the amount of medication used and the unit price for each type of medication. The questionnaires contained questions regarding the specific subtypes of medication used and the duration and adherence of medication use for 3 major types of medications including antiplatelet, statins and antihypertensive drugs (Multimedia Appendix 1). The unit cost of medication was estimated from the municipal government procurement website [30]. Prices for each specific medication were cross-referenced with local village doctors, and the median values for each category of medications were used to ensure balance at baseline. Outpatient costs, excluding medication costs, were estimated based on existing official Chinese statistics [29]. Resources supporting the SINEMA program—including the development of the SINEMA app, compensation for human resources, and voice message delivery fees—were estimated based on detailed project financial records.

The post-trial analysis relied on patients' self-reported data collected during the 6-year sustainability observation period. Hospitalization costs were estimated using the number of hospital visits recorded in the post-trial period and the average costs associated with different causes of inpatient visits, as determined during the within-trial period. Variables collected during follow-up and the detailed assumptions underlying these estimations are presented in Tables S3 and S4 in Multimedia Appendix 2.

Assessment of Effectiveness and Utility Outcomes

The within-trial measure of effectiveness is the mean reduction in mmHg in SBP from baseline to the end of follow-up at 1 year after randomization in the SINEMA trial. The health state utilities for the SINEMA study participants were estimated from the EQ-5D-5L using the Chinese EQ-5D-5L value set [31]. The quality-adjusted life years (QALYs) gained for both the within-trial period and post-trial period were calculated using the area under the curve approach [32]. For participants who died during this timeframe, their status and date of death were confirmed using the death registry system. A utility score of 0 was then assigned from the date of death onward. For the post-trial period, we similarly calculated the intervening missing utility values by using a linear imputation method and then summed the QALYs on an annual basis (Table S4 in Multimedia Appendix 2).

Economic Evaluation and Statistical Analysis

The incremental cost-effectiveness ratios (ICERs) were computed by dividing the mean difference in costs by the mean difference in effectiveness outcomes. ICERs were represented as additional cost per 1 mmHg reduction in SBP and per QALY gained between the SINEMA intervention and usual care.

We used a mixed multilevel effect model to evaluate effectiveness, with a random intercept for the cluster (village) and a fixed effect for townships to account for the stratified design, baseline SBP or baseline utility value, age, and sex based on the intention-to-treat principle. We used bootstrapping techniques to investigate the uncertainty around the ICERs by

XSL•FO
RenderX

generating 5000 replications to estimate the combined uncertainty of both cost and health outcomes and plotted on the cost-effectiveness plane to represent the uncertainty from 5000 estimates. To understand the cost-effectiveness of the SINEMA intervention across a spectrum of willingness-to-pay (WTP) thresholds, we used cost-effectiveness acceptability curves [33]. The WTP threshold was set as $18,766, equivalent to 1.5 times China's gross domestic product per capita in 2023 [34-36].

We performed both probabilistic sensitivity analysis by implementing nonparametric bootstrapping with 5000 replications and deterministic sensitivity analyses by assessing the impact of varying key input parameters individually. This involved evaluating the implications of adapting the 95% CI for health state utilities, considering 50% cost scenarios in both within-trial and post-trial economic evaluations. Additionally, we considered changes in hospitalization costs and each component of intervention to gauge the robustness of our findings. All statistical analyses were conducted using Stata version 17.

Budget Impact Analysis

We conducted the base-case budget impact analysis from a public payer perspective, quantifying all costs associated with the SINEMA program. These costs include the following: (1) compensation for township physicians, (2) compensation for the local project manager, (3) compensation for village doctors, (4) training costs, (5) voice messaging costs, (6) annual system maintenance and management costs, and (7) intervention development costs. The budget impact analysis assumes that the SINEMA program would be scaled up using existing village doctors and reach all rural households within the first year of implementation.

To quantify total costs, we estimated both unit costs and the number of people expected to receive SINEMA. We first assumed SINEMA as a standalone project as scenario A. Since a large proportion of patients with stroke already receive nationwide services under the national basic public health services, which involves quarterly visits by village doctors to monitor blood pressure, we integrated the SINEMA program into the current public health services for scenario B analysis. The estimation method for costs and the eligible population are detailed in Table S5 in Multimedia Appendix 2. The health service utilization analysis showed that SINEMA led to a significant and consistent increase in the use of antihypertensive and antiplatelet medications compared to usual care. Therefore, these medication costs are included (Table S6 in Multimedia Appendix 2). We assumed that supervision and system coordination costs would decline after the first year, reflecting typical scale-up patterns observed in similar community-based interventions [37,38], where oversight and training expenses declined substantially as programs became integrated into routine service delivery (Table S5 in Multimedia Appendix 2).

Ethical Considerations

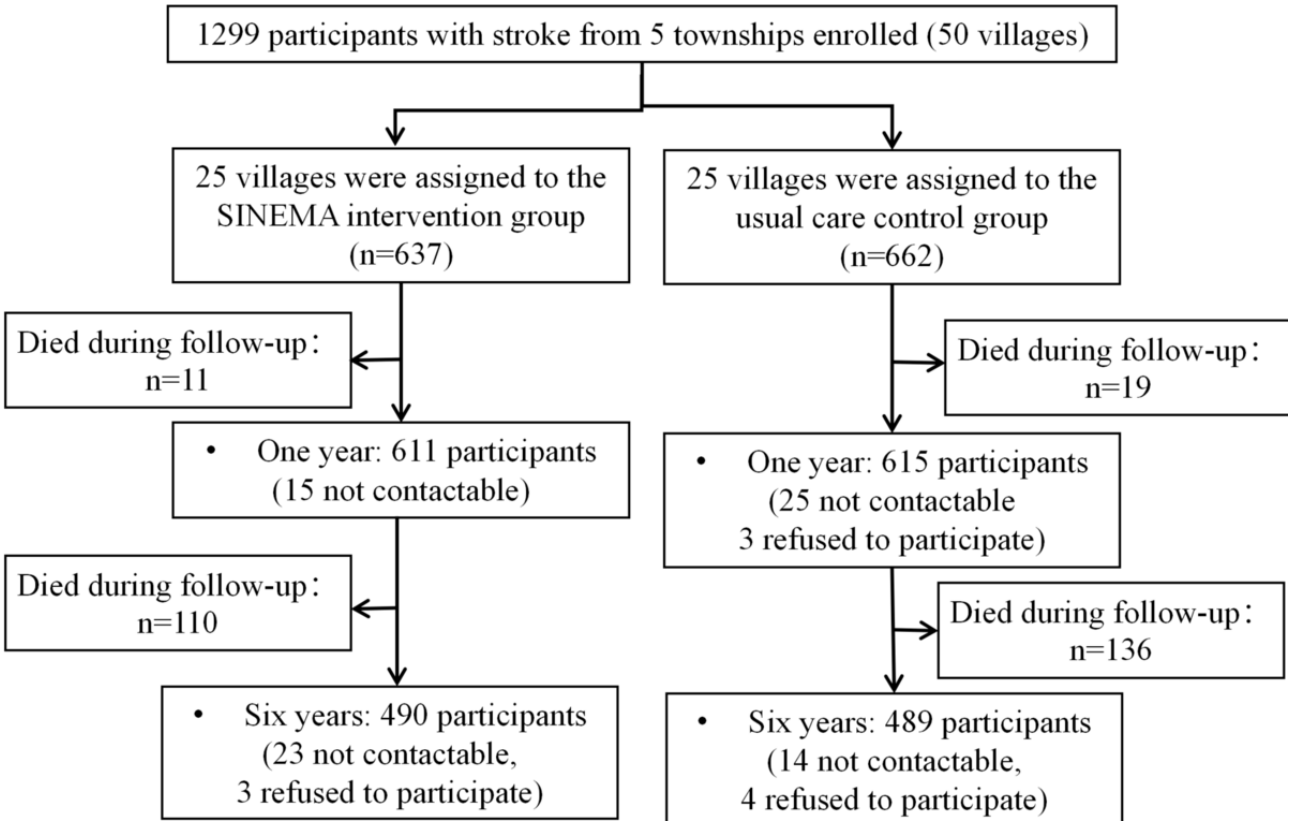
The study protocol was approved by the Institutional Review Boards of Duke University (Pro00082130), Beijing Tiantan Hospital (2015MRC0012), and Duke Kunshan University (China). All participants, including health care providers and patients, provided written informed consent after being informed of the study's purpose, procedures, potential benefits and risks, and data confidentiality safeguards. In addition, cluster-level consent was obtained from township and village opinion leaders. This post-trial follow-up study was approved by the ethical boards of the Chinese Academy of Medical Sciences (CAMS&PUMC-IEC-2022-062 and CAMS&PUMC-IEC-2024-047) and Duke University (Pro00082130-AMD-9.1).

Results

Both groups had comparable baseline characteristics (Table S7 in Multimedia Appendix 2). Participants' average age was 65.6 years (SD 8.2), with 71.2% (n=925) having no or only primary education. Their average SBP was 145.9 mm Hg (SD 22.4), and the mean health state utility score was 0.80 (SD 0.20). By the 1-year follow-up, 30 out of 1299 participants (2%) died, and 43 out of 1299 participants (3%) were lost to follow-up, leaving 611 and 615 participants in the intervention and control groups, respectively, who completed the survey (Figure 1 and Table S8 in Multimedia Appendix 2).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trails) flowchart of the SINEMA trial and the post-trial follow-up.



Within-Trial Cost-Effectiveness Analysis

The total program implementation costs were $32,389.2, averaging $50.8 per participant in the intervention group. During the trial period, the average total cost per person for inpatient visits was slightly higher in the control arm, though the difference was not statistically significant (mean difference: –$32.4, 95% CI –$107.3 to $172.1). The costs for antiplatelets and antihypertensives were significantly higher in the intervention arm, with mean differences of $2.11 (95% CI $1.27-$2.98) and $2.99 (95% CI $0.71-$5.27), respectively, suggesting an increased use of these medications. The total costs, including both intervention-related costs and health service use, were $236.5 per participant in the intervention arm and $213.1 in the control arm, resulting in a mean cost difference of $23.3 (95% CI: –$115.5 to $162.0). The bootstrapped analysis showed a mean difference of –2.8 mmHg (95% CI –4.7 to –1.0) in SBP and a 0.03 (95% CI 0.02-0.04) difference in QALYs between the 2 arms (Tables 1 and 2).

Table 1. Intervention costs during the within-trial period.

Intervention costs	Unit cost ($)	Quantity	Total costs ($)	Total cost per participant ($)
Compensation for township physicians	228.9	5	1144.7	1.8
Compensation for village doctors	414.8	25	10369.9	16.3
Compensation for local project manager	740.7	1	740.7	1.2
Compensation for county specialists	134.7	2	269.3	0.4
Voice messages	0.02	97,000	1548.8	2.4
Printing materials	0.5	700	336.7	0.5
System maintenance and management costs per year	1481.4	1	1481.4	2.3
Intervention development fee	16,497.6	1	16,497.6	25.9
Total	N/A[a]	N/A	32,389.2	50.8

[a]Not applicable.

XSL•FO
RenderX

Table 2. Health service utilization costs and effectiveness during the within-trial period.

Category	Intervention, mean (SD)[a]	Control, mean(SD)[a]	Mean difference (95% CI)
Health service utilization costs ($)[b]			
Outpatient visit (excluding medication costs, per participant)	45.5	45.5	0
All causes of inpatient visit (per participant)[c]	218.3	254.5	–2.4 (–107.3 to 172.1)
Medication costs (individual adherence discounted)[d]			
Antiplatelet	11.0	8.9	2.11 (1.27 to 2.98)
Statin	3.2	3.1	0.09 (–0.67 to 0.50)
Antihypertensive medicines	23.1	20.3	2.99 (0.71 to 5.27)
Average total costs[e]	236.5	213.1	23.3 (–115.5 to 162.0)
Effectiveness			
Change in systolic blood pressure (mmHg)[f]	–7.1 (0.6)	–4.3 (0.7)	–2.8 (–4.7 to –1.0)
Change in health-related quality of life score[f]	0.01 (0.01)	–0.06 (0.01)	0.06 (0.04 to 0.08)
Average quality-adjusted life years gained	0.81 (0.19)	0.79 (0.20)	0.03 (0.02 to 0.04)

[a]Cost values represent mean costs per participant calculated from aggregated data, not individual-level distributions; therefore, SDs are not reported.

[b]Health service utilization costs are reported as total costs per participant.

[c]Inpatient costs represent the mean total costs among participants who had at least one hospitalization.

[d]Medication costs represent the mean total costs among participants who used medication.

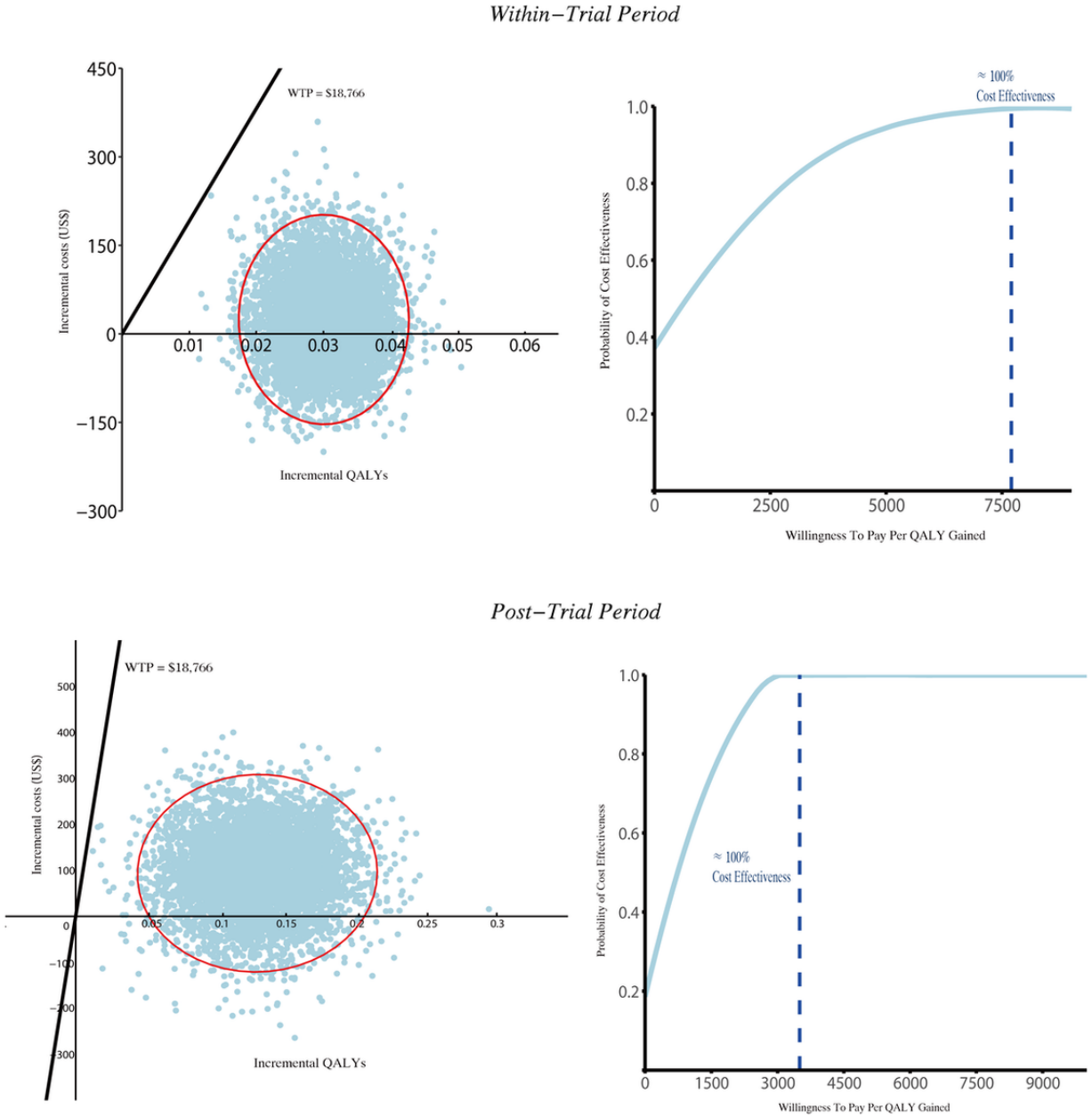
[e]The total cost reflects the mean cost per participant across the entire study population, including those without any hospitalization or medication use.

[f]Change represents the difference from baseline to 1-year follow-up within the trial period.

The ICER was $8.4 per mmHg reduction in SBP (95% CI 4.3-12.6) and $837.9 per QALY gained (95% CI $766.9-$908.9). The bootstrapped estimates of the mean differences in costs and effects were concentrated in the northeast quadrant of the cost-effectiveness plane (Figure 2, top left panel), indicating that the SINEMA intervention had a 100% probability (5000/5000 replications) of being cost-effective based on the predefined WTP threshold of $18,766. The cost-effectiveness acceptability curves for QALYs demonstrated that the probability of SINEMA being cost-effective was 100% at a WTP of approximately $7700 (Figure 2, top right panel).

XSL•FO
RenderX

Figure 2. All currency values are in 2023 US$. ICER: incremental cost-effectiveness ratio. “Within-trial period” refers to the intervention phase (2017–2018). “Post-trial period” refers to the time after the intervention concluded (from 2018 onward). Each blue dot represents an outcome from one of 5,000 iterations. The red circle indicates the 95% confidence interval (CI) of these outcomes. The dashed vertical line marks the probability of cost-effectiveness across willingness-to-pay thresholds, including the estimated threshold at which cost-effectiveness reaches 100%.



The scenario analysis of the ICER to changes in key parameters is presented in Table 3. The base case analysis estimated an ICER of $8.4 (95% CI $4.3-$12.6) per unit reduction in SBP and $837.9 (95% CI $766.9-$908.9) per QALY gained during the trial. Scenario analyses revealed that the ICER ranged from $3.8 to $23.0 per mmHg reduction in SBP and from $298.7 to $1363.9 per QALY gained across different adjustment models. Among intervention-related cost components, variations in the compensation for village doctors had the largest influence on the ICER. Regarding health service utilization costs, changes in annual hospitalization costs contributed the most to variations in the ICER.

XSL•FO
RenderX

Table 3. Sensitivity analyses for calculating incremental cost-effectiveness ratios during the within-trial period.

Additional scenario analysis	Range of values[a]	Cost per mmHg SBP[b] change[c] ($)		Cost per QALY[d] gained[c] ($)	
		Low	High	Low	High
Changes in effectiveness-related parameters					
Systolic blood pressure reduction, mmHg[e]	1.0-4.7	5.2	23.0	N/A[f]	N/A
Difference in QALYs between 2 groups[e]	0.02-0.04	N/A	N/A	603.4	1206.8
Incremental mean medication adherence	0.01-0.04	6.3	8.6	694.1	942.7
Changes in cost-related parameters ($)					
Compensation for township physicians[g]	572.4-1717.1	8.1	8.8	805.5	870.3
Compensation for village doctors[g]	5194.9-15554.9	5.4	11.2	539.4	1114.8
Compensation for local project manager	370.4-1111.1	8.2	8.5	812.7	855.9
Compensation for county specialists	134.65-403.95	8.3	8.4	823.5	837.9
Voice messages	774.4-2323.2	8.1	8.8	805.5	873.9
Printing materials	168.35-505.05	8.4	8.4	834.3	837.9
System maintenance and management costs per year	740.7-2222.1	7.9	8.7	784.0	866.7
Intervention development fee	8248.8-24,746.4	3.8	13.0	381.2	1294.6
Incremental annual medication cost	2.6-7.8	7.5	12.9	692.6	975.1
Incremental annual hospitalization cost	–48.6 to 16.2	4.8	16.7	298.7	1363.9

[a]The range was defined as a 50% decrease or increase from the base case for key parameters, except for systolic blood pressure reduction and difference in QALYs between 2 groups.

[b]SBP: systolic blood pressure.

[c]Calculated based on the 5000 bootstrapping and adjusted for baseline outcome, township, sex, and age.

[d]QALY: quality-adjusted life year.

[e]Range values are the bounds of the 95% CIs from the primary outcome analysis in the SINEMA trial.

[f]Not applicable.

[g]Presented as the average cost for township physicians and village doctors including performance-based incentives.

Post-Trial Cost-Effectiveness Analyses

During the 6-year sustainability observation period, 246 deaths were observed (intervention group: 110/626, 17.3%; control group: 136/643, 20.4%). Of the 1023 surviving participants, 979 (95.7%) completed the survey and 44 (4.3%) were lost to follow-up (Figure 1 and Table S8 in Multimedia Appendix 2).

The total post-trial costs were higher in the intervention group, with a mean difference of $91.5 (95% CI –$67.9 to $250.1). The mean difference in QALYs between the groups was 0.13 (0.06-0.20). This led to a lower ICER of $727.9 (95% CI $626.4-$829.4) in the post-trial period compared to the within-trial period (Table 4). The bootstrapped estimates demonstrated that the probability of the SINEMA intervention being cost-effective remained at 100% (Figure 2, bottom left panel). The cost-effectiveness acceptability curve indicates that the willingness-to-pay threshold for 100% cost-effectiveness is approximately $3,000 (Figure 2, bottom right panel). Scenario analysis revealed that the ICER varied the most when hospitalization costs were adjusted at the second follow-up, but no parameter changes affected the overall cost-effectiveness conclusions (Table 5).

XSL•FO
RenderX

Table 4. Costs and effectiveness during the post-trial period.

Category	Intervention, mean (SD)[a]	Control, mean (SD)[a]	Mean difference (95% CI)
Costs ($)			
Outpatient visit (excluding medication costs, per participant)	124.2	122.7	1.74 (–0.68 to 3.63)
All cause of inpatient visit (per episode)	639.8	564.9	74.80 (–97.96 to 247.49)
Medication costs (individual adherence discounted)			
Antiplatelet	52.7	44.3	8.41 (3.56 to 13.25)
Statin	12.0	11.4	0.53 (–1.25 to 2.31)
Antihypertensive medicines	56.7	48.0	8.69 (1.87 to 15.50)
Average total costs	889.9	798.4	91.50 (–67.87 to 250.80)
Effectiveness			
Change in health-related quality of life score[b]	–0.24 (0.01)	–0.23 (0.01)	–0.02 (–0.06 to 0.02)
Average Quality Adjusted Life Years Gained	2.51 (0.03)	2.41 (0.03)	0.13 (0.06 to 0.20)

[a]Cost values represent mean costs per participant calculated from aggregated data, not individual-level distributions; therefore, SDs are not reported.

[b]Change represents the difference from 1-year to 4.5-year follow-up in the post-trial period.

Table 5. Sensitivity analyses for calculating incremental cost-effectiveness ratios during the post-trial period.

Additional scenario analysis	Range of values[a]	Cost per QALY[b] gained[c] ($)	
		Low	High
Changes in cost-related parameters			
Medication cost at the second follow-up ($)	82.7-160.2	717.6	843.1
Hospitalization cost at the second follow-up ($)	391-994.5	219.2	1328.8
Discount (%)	3-8	511	860.5
Changes in effectiveness-related parameters			
Difference in QALYs between 2 groups[d]	0.07-0.19	488.8	1325.8
Incremental mean medication adherence[d]	0.47-1.41	660.5	839.7

[a]The value range was defined as ±50% of the base case for key parameters, except for the difference in QALYs between the 2 groups and the incremental mean medication adherence.

[b]QALY: quality-adjusted life year.

[c]Calculated based on the 5000 bootstrapping and adjusted for baseline outcome, township, sex, and age.

[d]Low and high values are the bounds of the 95% CIs from the primary outcome analysis in the SINEMA trial.

Budget Impact Analysis

Based on a stroke prevalence of 2.58% [39] and a rural proportion of patients with stroke of 52.9% [29], the eligible population for scaling up the SINEMA intervention to rural China was estimated to be 11,676,010 people. For subsequent years, we estimated the eligible population based on an annual stroke incidence rate of 8.26% [39], assuming a 1% annual national population growth and a 0.16% death rate [29]. This broader reach enabled the amortization of fixed costs such as system development and oversight across a much larger population, resulting in a reduced per capita cost. In scenario A, the program cost was estimated at $16.93 per participant, leading to a total budget impact of $197.72 million in the first year. Since intervention costs were only incurred in the first year, expenditure decreased in subsequent years. By the fifth year, with 4 visits from village doctors per year maintained, the expenditure decreased slightly by 3.2%, reaching $78.07 million (Table 6).

XSL•FO
RenderX

Table 6. Budget impact analysis of the system-integrated and digital technology–enabled model of care (SINEMA) intervention in year 1 and year 5.

Item	Scenario A: stand-alone project		Scenario B: integrated with NBPH[a]	
	Year 1	Year 5	Year 1	Year 5
Eligible population, n	11,676,010	11,695,170	11,676,010	11,695,170
Eligible participants receiving NBPH[b]	N/A[c]	N/A	7,125,158	7,136,851
Eligible participants excluded from NBPH[b]	N/A	N/A	4,550,852	4,558,319
Cost per participant ($)	16.93	6.68	9.90	3.98
Total budget impact[d] ($)	197,726,357	78,067,406	115,585,636	46,592,046
Total population, n	1,411,750,000	1,459,785,833	1,411,750,000	1,459,785,833
Cost per capita[e] ($)	0.14	0.06	0.08	0.03

[a]NBPH: national basic public health.

[b]Eligible participants receiving NBPH are individuals with stroke and hypertension or diabetes living in rural communities. Eligible participants excluded from NBPH are individuals with stroke who do not have hypertension or diabetes.

[c]Not applicable.

[d]Total budget impact is rounded to the nearest $10.

[e]Calculated as the total budget impact divided by the total population.

In scenario B, where the SINEMA intervention was integrated into the existing national basic public health services, we identified that 7,125,158 participants—approximately 61% of the eligible population—had already received these services [39]. The cost per participant in this scenario was reduced to $9.90, which resulted in a total budget impact of $115.59 million in the first year, only slightly over half of the costs in scenario A (Table 6). Results for all years and specific components are shown in Table S9 in Multimedia Appendix 2. Compensation for village doctor visits accounted for the largest portion of the total budget in both scenarios.

Discussion

Principal Results

With a mean follow-up of 6 years of the SINEMA trial, we conducted within-trial and post-trial cost-effectiveness analyses and budget impact analyses for the SINEMA model—an integrated mHealth primary care model for stroke management. The hybrid, type 2, effectiveness-implementation SINEMA trial has previously demonstrated not only the effectiveness of evidence-based strategies but also the feasibility of their implementation strategies [10,40]. Building on these findings, this study further showed that the SINEMA model was cost-effective for controlling blood pressure and improving QALYs compared to usual care, both within the 12-month trial period and during the 5-year post-trial follow-up. The base case analysis estimated an ICER of $8.4 per unit reduction in SBP, $837.9 per QALY gained during the trial, and $727.9 per QALY gained in the post-trial period, much lower than the WTP thresholds. Sensitivity analyses across various scenarios consistently supported these findings, reinforcing the robustness of the results. The budget impact analysis further supported the potential of the SINEMA intervention as a sustainable solution that incorporates both system-level consideration and mHealth empowerment, suitable for national scale-up in managing stroke and potentially for other chronic diseases.

Comparison With Prior Work

Our results demonstrating the high cost-effectiveness of the SINEMA model highlighted the importance of investing in primary health care and mHealth technology for stroke management and improving patient outcomes. To our knowledge, no cost-effectiveness studies on stroke interventions using similar mHealth technology have been conducted in LMICs, aside from one study from the United States and another from India, both of which used telemedicine-based approaches [13,14]. In a US-based study, the ACCESS teleneurology program for acute stroke care reduced unnecessary patient transfers and improved early treatment, resulting in a cost saving of $4241 and a QALY gain of 0.20 per patient over a 90-day horizon, demonstrating economic dominance over standard care [14]. In the Indian study involving 1200 postneurosurgery patients, telemedicine follow-up care was more effective and less costly than routine in-person care, with a cost per utility unit of $40.5 versus $105.4 and an ICER of –$537 [13]. However, the large differences in intervention design and context preclude direct comparison with our results. Moreover, telemedicine adoption in rural China has faced significant barriers, particularly related to digital literacy [41]. The SINEMA model, which includes voice message–based health communication, may help improve digital health acceptance among rural populations. Its cost-effectiveness has been clearly demonstrated, with a cost of $837.9 per QALY gained during the within-trial period—well below the internal cost-effectiveness threshold in China (eg, 1.5 times the 2023 gross domestic product per capita of $18,766).

We also compared our results with other community-based blood pressure–lowering trials in LMICs in rural Asia. For example, the COBRA trial in Pakistan reported an ICER of $23 per mmHg reduction in SBP [42]. The COBIN intervention in

XSL•FO
RenderX

Nepal, though initially cost-effective [37], saw a puzzling reversal of its benefits over time from a 5-mmHg reduction at 1 year to a 5-mmHg increase by 5 years [21]. In contrast, our 5-year post-trial results demonstrated sustained effectiveness of the intervention over time [11]. The ICER in our study was $10.5 per mmHg reduction in SBP, which was lower than that in the COBRA trial. One possible explanation for the higher and sustained long-term cost-effectiveness observed in our study, which was not accounted for in their analyses, could be the integration of mHealth technology, including a provider-facing Android app and patient-oriented voice messages.

Based on our budget impact analysis, scaling up the SINEMA model nationwide is both affordable and impactful. In China, while per capita health expenditure surged from RMB 1090 to RMB 4700 (US $164.9 to US $710.3) [43], spending on specialized public health organizations increased in absolute terms but dropped as a share of total health expenditure, from 8.58% to 5.47% [44]. In our study, the required investment for scaling up the SINEMA intervention represents about 0.15% of the basic national health services expenditure and could reduce hospitalization costs, underscoring the importance of investing in primary care and the value of task-shifting strategies from specialists to primary care providers [45,46]. Our study also explored the scenario of integrating the SINEMA app into the national basic public health service, adding intervention components of voice messaging and frequent SBP electronic records, which is estimated at nearly half the costs of the standard intervention. This finding highlighted the potential for saving costs when leveraging mHealth technology within the existing system. However, potential barriers and challenges, such as regional disparities in health care infrastructure, workforce capacity, and digital literacy, may be raised when considering large-scale implementation, which may lead to additional intervention components and additional costs. Furthermore, given that incentives for village doctors consistently accounted for the largest proportion of overall costs, future studies could further explore the strategies of a risk-based, adaptive delivery model that tailors follow-up intensity based on patients' clinical stability.

Limitations

Unlike most within-trial or modelling studies, this economic evaluation study based on the SINEMA trial has 3 parts—within-trial and post-trial cost-effectiveness analyses and budget impact analyses. We used extensive data on costs and effectiveness at the individual level collected mainly from our trial data, including mortality for both periods, alleviating the need to rely on blood pressure–based parameter estimates from literature and risk prediction models. Our study has some limitations. First, data on outpatient care were unavailable from the trial, requiring us to use external sources. Nonetheless, outpatient costs constitute only a small proportion of total medical costs. Second, while we collected patient-level data for the past year at the end of the active trial in 2018 and in the 5-year post-trial survey (2022-2023), data during the intermediate years (2019-2021) were missing. We extrapolated data for these years from available data. Extensive sensitivity analyses, varying each parameter by ±50%, found that the results were robust, mitigating concerns about missing data. Third, this trial was conducted in a single city in rural China, which may limit its generalizability. Nonetheless, our budget impact analysis explored the potential for national scale-up and highlighted that this intervention aligns well with the existing health system. The mHealth-enabled system, aligned with the current basic public health services framework, has the potential to enhance the management of stroke and other chronic diseases on a larger scale. Lastly, in the budget impact analyses, we lacked cost data from rural areas beyond the trial site and therefore assumed that intervention costs in other rural settings would be comparable to those observed in the trial, which was conducted in a single rural location. Additionally, we projected potential cost reductions in subsequent scale-up years, particularly related to human resources and the mHealth system. While such assumptions are commonly used in similar budget impact analyses, they remain strong and are inherently difficult to validate. Nonetheless, the analyses offer concrete estimates of potential financial implications and serve as a valuable tool for informing policy decisions and guiding resource allocation.

Conclusions

In summary, our findings on the cost-effectiveness of the SINEMA model highlight the value of investing in primary care services and mHealth technology for stroke care in rural China. The budget impact analyses provide evidence for sound policymaking and reveal their significant potential for nationwide scaling up in China. While this study was conducted among patients with stroke in a resource-limited region of China, its high cost-effectiveness highlights the need for future research to explore whether and how similar models—integrating mHealth technologies with the primary care workforce—can be adapted for other chronic conditions and comparable settings in LMICs.

Acknowledgments

We would like to thank the independent International Steering Committee Chair (Yangfeng Wu) and members (Eric Peterson and Craig Anderson) and advisory group members (Alba AmayaBurns, Allan Burns, Ninghua Wang, Xie Bin, Jesse Hao, Jixiang Ma, Jixin Sun, Jianxin Zhang, Jianmin Yao, Jinmei Liu, Qian Long, and Cheng Sun) who have provided great advice in designing and implementing the study. We also thank the Xingtai City Center for Disease Prevention and Control, Nanhe County Center for Disease Prevention and Control, Ren County Center for Disease Prevention and Control, all staff members from township health care centers and village clinics, and all patients who participated in or supported the project. Lei Si can be contacted as a co-corresponding author at l.si@westernsydney.edu.au.

XSL•FO
RenderX

The post-trial follow-up of the SINEMA study was supported by the National Natural Science Foundation of China (grant HH8220122034 to EG), the Nonprofit Central Research Institute Fund of the Chinese Academy of Medical Sciences (grant 2021-RC300-004 to EG), and the National Key R&D Program of China (grant 2023YFC3605000 to LLY and EG). The SINEMA initiative was funded by the United Kingdom Medical Research Council, Economic and Social Research Council, Department for International Development, and Wellcome Trust (grant MR/N015967/1 to LLY). This article's research results were also supported by research funding from the Kunshan Municipal Government (LLY). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

Deidentified individual data can be made available upon written, detailed request directed to the corresponding author. Requests will be reviewed on the basis of scientific merit, ethical review, legal issues, and regulatory requirements.

Authors' Contributions

Corrected:Conceptualization: LLY, LS, EG, BY
Data curation: BY, EG, LS, XC, Yuhan Li
Formal analysis: BY, EG, LS, Yuhan Li
Funding acquisition: LLY, EG
Investigation: JC, Yan Li, BO, CC, NP
Methodology: BY, EG, LS, EF, XZ
Project administration: LLY, LS, EG
Resources: LLY, EG
Supervision: LLY, LS
Validation: Yuhan Li, JC
Visualization: BY, LS
Writing – original draft: BY, EG, LS, LLY
Writing – review and editing: All authors
LS can be contacted as a co-corresponding author at l.si@westernsydney.edu.au.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Trial protocol and long-term follow-up protocol.
[PDF File (Adobe PDF File), 3869 KB - mhealth_v13i1e75326_app1.pdf]

Multimedia Appendix 2
Supplemental tables.
[DOCX File , 65 KB - mhealth_v13i1e75326_app2.docx]

Multimedia Appendix 3
CONSORT checklist.
[PDF File (Adobe PDF File), 124 KB - mhealth_v13i1e75326_app3.pdf]

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820 [FREE Full text] [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
2. Rajsic S, Gothe H, Borba HH, Sroczynski G, Vujicic J, Toell T, et al. Economic burden of stroke: a systematic review on post-stroke care. Eur J Health Econ 2019 Feb;20(1):107-134. [doi: 10.1007/s10198-018-0984-0] [Medline: 29909569]
3. Zhang J, Song S, Zhao Y, Ma G, Jin Y, Zheng Z. Economic burden of comorbid chronic conditions among survivors of stroke in China: 10-year longitudinal study. BMC Health Serv Res 2021 Sep 17;21(1):978 [FREE Full text] [doi: 10.1186/s12913-021-07010-1] [Medline: 34535138]
4. Wang W, Jiang B, Sun H, Ru X, Sun D, Wang L, NESS-China Investigators. Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480 687 adults. Circulation 2017 Feb 21;135(8):759-771. [doi: 10.1161/CIRCULATIONAHA.116.025250] [Medline: 28052979]

XSL•FO
RenderX

5. Notice of the General Office of the National Health Commission on the issuance of the guidelines for the prevention and treatment of stroke in China (2021 edition). National Health Commission of the People's Republic of China. 2021. URL: https://www.nhc.gov.cn/wjw/c100175/202108/df1789a9ba174678b1979bc14117b58f.shtml [accessed 2025-08-18]
6. Zhang T, Zhao J, Li X, Bai Y, Wang B, Qu Y, Chinese Stroke Association Stroke Council Guideline Writing Committee. Chinese Stroke Association guidelines for clinical management of cerebrovascular disorders: executive summary and 2019 update of clinical management of stroke rehabilitation. Stroke Vasc Neurol 2020 Sep;5(3):250-259 [FREE Full text] [doi: 10.1136/svn-2019-000321] [Medline: 32595138]
7. Zhao F, Yue Y, Li L, Lang S, Wang M, Du X, et al. Clinical practice guidelines for post-stroke depression in China. Braz J Psychiatry 2018;40(3):325-334 [FREE Full text] [doi: 10.1590/1516-4446-2017-2343] [Medline: 29412338]
8. Boehme C, Toell T, Lang W, Knoflach M, Kiechl S. Longer term patient management following stroke: a systematic review. Int J Stroke 2021 Oct;16(8):917-926 [FREE Full text] [doi: 10.1177/17474930211016963] [Medline: 33949269]
9. Gong E, Gu W, Sun C, Turner EL, Zhou Y, Li Z, et al. System-integrated technology-enabled model of care to improve the health of stroke patients in rural China: protocol for SINEMA-a cluster-randomized controlled trial. Am Heart J 2019 Jan;207:27-39. [doi: 10.1016/j.ahj.2018.08.015] [Medline: 30408621]
10. Yan LL, Gong E, Gu W, Turner EL, Gallis JA, Zhou Y, et al. Effectiveness of a primary care-based integrated mobile health intervention for stroke management in rural China (SINEMA): a cluster-randomized controlled trial. PLoS Med 2021 Apr;18(4):e1003582 [FREE Full text] [doi: 10.1371/journal.pmed.1003582] [Medline: 33909607]
11. Tan J, Gong E, Gallis JA, Sun S, Chen X, Turner EL, et al. Primary care-based digital health-enabled stroke management intervention: long-term follow-up of a cluster randomized clinical trial. JAMA Netw Open 2024 Dec 02;7(12):e2449561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.49561] [Medline: 39671199]
12. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. Value Health 2015 Mar;18(2):161-172 [FREE Full text] [doi: 10.1016/j.jval.2015.02.001] [Medline: 25773551]
13. Thakar S, Rajagopal N, Mani S, Shyam M, Aryan S, Rao AS, et al. Comparison of telemedicine with in-person care for follow-up after elective neurosurgery: results of a cost-effectiveness analysis of 1200 patients using patient-perceived utility scores. Neurosurg Focus 2018 May;44(5):E17. [doi: 10.3171/2018.2.FOCUS17543] [Medline: 29712532]
14. Whetten J, van der Goes DN, Tran H, Moffett M, Semper C, Yonas H. Cost-effectiveness of access to critical cerebral emergency support services (ACCESS): a neuro-emergent telemedicine consultation program. J Med Econ 2018 Apr;21(4):398-405 [FREE Full text] [doi: 10.1080/13696998.2018.1426591] [Medline: 29316820]
15. Jafar TH, Gandhi M, de Silva HA, Jehan I, Naheed A, Finkelstein EA, COBRA-BPS Study Group. A community-based intervention for managing hypertension in rural South Asia. N Engl J Med 2020 Feb 20;382(8):717-726. [doi: 10.1056/NEJMoa1911965] [Medline: 32074419]
16. Neupane D, McLachlan CS, Mishra SR, Olsen MH, Perry HB, Karki A, et al. Effectiveness of a lifestyle intervention led by female community health volunteers versus usual care in blood pressure reduction (COBIN): an open-label, cluster-randomised trial. Lancet Glob Health 2018 Jan;6(1):e66-e73 [FREE Full text] [doi: 10.1016/S2214-109X(17)30411-4] [Medline: 29241617]
17. Jafar TH, Hatcher J, Poulter N, Islam M, Hashmi S, Qadri Z, Hypertension Research Group. Community-based interventions to promote blood pressure control in a developing country: a cluster randomized trial. Ann Intern Med 2009 Nov 03;151(9):593-601. [doi: 10.7326/0003-4819-151-9-200911030-00004] [Medline: 19884620]
18. Xiong S, Lu H, Peoples N, Duman EK, Najarro A, Ni Z, et al. Digital health interventions for non-communicable disease management in primary health care in low-and middle-income countries. NPJ Digit Med 2023 Feb 01;6(1):12 [FREE Full text] [doi: 10.1038/s41746-023-00764-4] [Medline: 36725977]
19. Allan LP, Beilei L, Cameron J, Olaiya MT, Silvera-Tawil D, Adcock AK, et al. A scoping review of mHealth interventions for secondary prevention of stroke: implications for policy and practice. Stroke 2023 Nov;54(11):2935-2945. [doi: 10.1161/strokeaha.123.043794]
20. Jafar TH, Jehan I, Liang F, Barbier S, Islam M, Bux R, et al. Control of blood pressure and risk attenuation: post trial follow-up of randomized groups. PLoS One 2015;10(11):e0140550 [FREE Full text] [doi: 10.1371/journal.pone.0140550] [Medline: 26540210]
21. Thapa R, Zengin A, Neupane D, Mishra SR, Koirala S, Kallestrup P, et al. Sustainability of a 12-month lifestyle intervention delivered by community health workers in reducing blood pressure in Nepal: 5-year follow-up of the COBIN open-label, cluster randomised trial. Lancet Glob Health 2023 Jul;11(7):e1086-e1095 [FREE Full text] [doi: 10.1016/S2214-109X(23)00214-0] [Medline: 37349035]
22. Norrie J. The importance of long-term follow-up in clinical trials. Lancet Glob Health 2023 Jul;11(7):e995-e996 [FREE Full text] [doi: 10.1016/S2214-109X(23)00244-9] [Medline: 37349045]
23. Wu N, Gong E, Wang B, Gu W, Ding N, Zhang Z, et al. A smart and multifaceted mobile health system for delivering evidence-based secondary prevention of stroke in rural China: design, development, and feasibility study. JMIR Mhealth Uhealth 2019 Jul 19;7(7):e13503 [FREE Full text] [doi: 10.2196/13503] [Medline: 31325288]

XSL•FO
RenderX

24. Gong E, Gu W, Luo E, Tan L, Donovan J, Sun C, et al. Development and local contextualization of mobile health messages for enhancing disease management among community-dwelling stroke patients in rural China: multimethod study. JMIR Mhealth Uhealth 2019 Dec 17;7(12):e15758 [FREE Full text] [doi: 10.2196/15758] [Medline: 31845901]
25. Chen X, Gong E, Tan J, Turner EL, Gallis JA, Sun S, et al. Long-term mortality outcome of a primary care-based mobile health intervention for stroke management: six-year follow-up of a cluster-randomized controlled trial. PLoS Med 2025 Mar;22(3):e1004564 [FREE Full text] [doi: 10.1371/journal.pmed.1004564] [Medline: 40146698]
26. Yip WC, Hsiao WC, Chen W, Hu S, Ma J, Maynard A. Early appraisal of China's huge and complex health-care reforms. Lancet 2012 Mar 03;379(9818):833-842. [doi: 10.1016/S0140-6736(11)61880-1] [Medline: 22386036]
27. Gong E, Yang B, Chen X, Li Y, Li Z, Bettger JP, et al. An economic evaluation of a primary care-based technology-enabled intervention for stroke secondary prevention and management in rural China: a study protocol. Front Neurol 2023;14:1145562 [FREE Full text] [doi: 10.3389/fneur.2023.1145562] [Medline: 37200779]
28. National data. National Bureau of Statistics. URL: https://data.stats.gov.cn/search.htm?s=%E5%86%9C%E6%9D%91CPI [accessed 2025-08-13]
29. China statistical yearbook 2022. National Bureau of Statistics of China. URL: https://www.stats.gov.cn/sj/ndsj/2022/indexeh.htm [accessed 2025-08-13]
30. Hebei Province medical drugs and equipment centralized procurement center. Hebei Provincial Medical Insurance Bureau. URL: https://ylbzj.hebei.gov.cn/cms/category/162 [accessed 2025-08-18]
31. Liu GG, Wu H, Li M, Gao C, Luo N. Chinese time trade-off values for EQ-5D health states. Value Health 2014 Jul;17(5):597-604 [FREE Full text] [doi: 10.1016/j.jval.2014.05.007] [Medline: 25128053]
32. Willan AR, Briggs AH. Statistical Analysis of Cost-Effectiveness Data. Hoboken, NJ: John Wiley; 2006.
33. Bertram MY, Lauer JA, Stenberg K, Edejer TTT. Methods for the economic evaluation of health care interventions for priority setting in the health system: an update from WHO CHOICE. Int J Health Policy Manag 2021 Nov 01;10(11):673-677 [FREE Full text] [doi: 10.34172/ijhpm.2020.244] [Medline: 33619929]
34. Cai D, Shi S, Jiang S, Si L, Wu J, Jiang Y. Estimation of the cost-effective threshold of a quality-adjusted life year in China based on the value of statistical life. Eur J Health Econ 2022 Jun;23(4):607-615 [FREE Full text] [doi: 10.1007/s10198-021-01384-z] [Medline: 34655364]
35. Xu L, Chen M, Angell B, Jiang Y, Howard K, Jan S, et al. Establishing cost-effectiveness threshold in China: a community survey of willingness to pay for a healthylife year. BMJ Glob Health 2024 Jan 09;9(1):e013070 [FREE Full text] [doi: 10.1136/bmjgh-2023-013070] [Medline: 38195152]
36. Liu GG, Guan H, Peng N, Xie S, Wang K, Liu LZ, et al. Key issues of economic evaluations for health technology assessment in China: a nationwide expert survey. Value Health 2024 Nov;27(11):1535-1543 [FREE Full text] [doi: 10.1016/j.jval.2024.06.020] [Medline: 38977183]
37. Krishnan A, Finkelstein EA, Kallestrup P, Karki A, Olsen MH, Neupane D. Cost-effectiveness and budget impact of the community-based management of hypertension in Nepal study (COBIN): a retrospective analysis. Lancet Glob Health 2019 Oct;7(10):e1367-e1374 [FREE Full text] [doi: 10.1016/S2214-109X(19)30338-9] [Medline: 31537367]
38. Finkelstein EA, Krishnan A, Naheed A, Jehan I, de Silva HA, Gandhi M, COBRA-BPS study group. Budget impact and cost-effectiveness analyses of the COBRA-BPS multicomponent hypertension management programme in rural communities in Bangladesh, Pakistan, and Sri Lanka. Lancet Glob Health 2021 May;9(5):e660-e667 [FREE Full text] [doi: 10.1016/S2214-109X(21)00033-4] [Medline: 33751956]
39. Wang Y, Li Z, Gu H, Zhai Y, Zhou Q, Jiang Y, China Stroke Statistics Writing Committee. China stroke statistics: an update on the 2019 report from the National Center for Healthcare Quality Management in Neurological Diseases, China National Clinical Research Center for Neurological Diseases, the Chinese Stroke Association, National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention and Institute for Global Neuroscience and Stroke Collaborations. Stroke Vasc Neurol 2022 Oct;7(5):415-450 [FREE Full text] [doi: 10.1136/svn-2021-001374] [Medline: 35443985]
40. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
41. Du Y, Zhou Q, Cheng W, Zhang Z, Hoelzer S, Liang Y, et al. Factors influencing adoption and use of telemedicine services in rural areas of China: mixed methods study. JMIR Public Health Surveill 2022 Dec 23;8(12):e40771 [FREE Full text] [doi: 10.2196/40771] [Medline: 36563026]
42. Jafar TH, Islam M, Bux R, Poulter N, Hatcher J, Chaturvedi N, Hypertension Research Group. Cost-effectiveness of community-based strategies for blood pressure control in a low-income developing country: findings from a cluster-randomized, factorial-controlled trial. Circulation 2011 Oct 11;124(15):1615-1625 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.039990] [Medline: 21931077]
43. National Health Commission of the People's Republic of China. China Health Statistical Yearbook 2022. Beijing, China: China Union Medical University Press; 2022.
44. Meng Q. Strengthening public health systems in China. Lancet Public Health 2022 Dec;7(12):e987-e988 [FREE Full text] [doi: 10.1016/S2468-2667(22)00206-7] [Medline: 36462519]

XSL•FO
RenderX

45. Anand TN, Joseph LM, Geetha AV, Prabhakaran D, Jeemon P. Task sharing with non-physician health-care workers for management of blood pressure in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Glob Health 2019 Jun;7(6):e761-e771 [FREE Full text] [doi: 10.1016/S2214-109X(19)30077-4] [Medline: 31097278]
46. Naik N, Singh K. Innovations for effective implementation of guideline-based hypertension care in low-income and middle-income countries. Lancet Glob Health 2021 May;9(5):e571-e572 [FREE Full text] [doi: 10.1016/S2214-109X(21)00083-8] [Medline: 33751955]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trails
ICER: incremental cost-effectiveness ratio
LMIC: low- and middle-income country
mHealth: mobile health
QALY: quality-adjusted Life Year
SINEMA: system-integrated and digital technology–enabled model of care
SBP: systolic blood pressure
WTP: willingness-to-pay

Edited by L Buis; submitted 01.04.25; peer-reviewed by X Ding, M Dale; comments to author 13.07.25; revised version received 24.07.25; accepted 08.08.25; published 11.09.25.

Please cite as:
Yang B, Gong E, Chen X, Tan J, Peoples N, Li Y, Cai J, Li Y, Oldenburg B, Chen C, Dong D, Zhang X, Finkelstein E, Si L, Yan LL Economic Evaluation of a Multicomponent mHealth Intervention for Stroke Management in Rural China: Cluster-Randomized Trial With 6-Year Follow-Up
JMIR Mhealth Uhealth 2025;13:e75326
URL: https://mhealth.jmir.org/2025/1/e75326
doi:10.2196/75326
PMID:

©Bolu Yang, Enying Gong, Xingxing Chen, Jie Tan, Nicholas Peoples, Yuhan Li, Jiayu Cai, Yan Li, Brian Oldenburg, Chen Chen, Dejin Dong, Xiaochen Zhang, Eric Finkelstein, Lei Si, Lijing L Yan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Clarifying Blood Indices in Patients With Ovarian Cancer

Chengwei Lian[1,2], BS; Yu Fan[1,2], MSc; Jinke Li[1], PhD

[1]Department of Gynaecology and Obstetrics, West China Second Hospital, Sichuan University, No 20, Sec 3, Renminnan Road, Sichuan 610041, Chengdu, China

[2]Key Laboratory of Birth Defects and Related Diseases of Women and Children, Sichuan University, Ministry of Education, Chengdu, China

Corresponding Author:
Jinke Li, PhD
Department of Gynaecology and Obstetrics, West China Second Hospital, Sichuan University, No 20, Sec 3, Renminnan Road, Sichuan 610041, Chengdu, China

Related Articles:

Comment in: https://mhealth.jmir.org/2025/1/e74931

https://mhealth.jmir.org/2024/1/e56475

(JMIR Mhealth Uhealth 2025;13:e70895) doi:10.2196/70895

KEYWORDS

WeChat; nutrition management; ovarian cancer; chemotherapy; mobile health

This letter is regarding the recent publication "Efficacy of a WeChat-Based, Multidisciplinary, Full-Course Nutritional Management Program on the Nutritional Status of Patients With Ovarian Cancer Undergoing Chemotherapy: Randomized Controlled Trial" [1]. The authors reported a novel approach to managing the nutrition status of patients with ovarian cancer who underwent chemotherapy. With supportive results, it might be a potentially beneficial program for this patient cohort. However, there might be a problem in the description of Figure 7.

The authors described the changes in leukocytes, lymphocytes, neutrophils, and platelets as "inflammation-related blood indices." It might be appropriate for most patients with normal immune status but is not applicable for those undergoing chemotherapy with bone marrow suppression (BMS) [2]. BMS is one of the most common adverse effects in patients undergoing chemotherapy and can be divided into chemotherapy-induced leukopenia, neutropenia, thrombocytopenia, and anemia [3]. The changes in the blood cells above are not only an indication of inflammatory status but also critical to monitor the BMS in patients. It is important to discuss BMS in treatment using chemotherapy separately or combined with others. Additionally, the authors have reported a systematic inflammatory index, which is more accurate, to evaluate the inflammatory status of patients [4]. As for Figure 7, we noticed a significant decrease in leukocytes, dropping almost to the diagnostic criteria of leukopenia [5] in the control group, which suggests more severe BMS in the control group. So we suggest it might be more accurate to describe Figure 7 simply as "the changes in blood cells," and the potential benefit for improving BMS should also be described in this section.

The findings of this study are meaningful and interesting. The authors provided a potential approach to improve the clinical outcomes of patients with ovarian cancer undergoing chemotherapy, especially in China. Additionally, we noticed the intervention could improve BMS for these patients, although it should be described and addressed more accurately in the publication.

Conflicts of Interest

None declared.

References

1. Tian X, Liu Y, Zhang J, et al. Efficacy of a WeChat-based, multidisciplinary, full-course nutritional management program on the nutritional status of patients with ovarian cancer undergoing chemotherapy: randomized controlled trial. JMIR Mhealth Uhealth 2024 Nov 4;12:e56475. [doi: 10.2196/56475] [Medline: 39496160]
2. Markman M, Kennedy A, Webster K, Peterson G, Kulp B, Belinson J. Combination chemotherapy with carboplatin and docetaxel in the treatment of cancers of the ovary and fallopian tube and primary carcinoma of the peritoneum. J Clin Oncol 2001 Apr 1;19(7):1901-1905. [doi: 10.1200/JCO.2001.19.7.1901] [Medline: 11283121]

XSL•FO
RenderX

3. Ten Berg MJ, van den Bemt P, Shantakumar S, et al. Thrombocytopenia in adult cancer patients receiving cytotoxic chemotherapy: results from a retrospective hospital-based cohort study. Drug Saf 2011 Dec 1;34(12):1151-1160. [doi: 10.2165/11594310-000000000-00000] [Medline: 22077503]
4. Hu B, Yang XR, Xu Y, et al. Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. Clin Cancer Res 2014 Dec 1;20(23):6212-6222. [doi: 10.1158/1078-0432.CCR-14-0442] [Medline: 25271081]
5. Weycker D, Li X, Edelsberg J, et al. Risk and consequences of chemotherapy-induced febrile neutropenia in patients with metastatic solid tumors. J Oncol Pract 2015 Jan;11(1):47-54. [doi: 10.1200/JOP.2014.001492] [Medline: 25491042]

Abbreviations

BMS: bone marrow suppression

Edited by T Leung; submitted 05.01.25; this is a non–peer-reviewed article;accepted 24.07.25; published 13.08.25.

Please cite as:
Lian C, Fan Y, Li J
Clarifying Blood Indices in Patients With Ovarian Cancer
JMIR Mhealth Uhealth 2025;13:e70895
URL: https://mhealth.jmir.org/2025/1/e70895
doi:10.2196/70895

© Chengwei Lian, Yu Fan, Jinke Li. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Authors' Reply: Clarifying Blood Indices in Patients With Ovarian Cancer

Jiahua Zhang, MSN

Ward 2, Gynaecological Oncology Centre, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, No.1 Shuaifuyuan, Dongcheng District, Beijing, China

Corresponding Author:
Jiahua Zhang, MSN
Ward 2, Gynaecological Oncology Centre, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, No.1 Shuaifuyuan, Dongcheng District, Beijing, China

Related Articles:

https://mhealth.jmir.org/2024/1/e56475/

https://mhealth.jmir.org/2025/1/e70895/

(JMIR Mhealth Uhealth 2025;13:e74931) doi:10.2196/74931

KEYWORDS

WeChat; nutrition management; ovarian cancer; chemotherapy; mobile health

We agreed with the authors of the letter to the editor [1] that Figure 7 simply describes "the changes in blood cells." Our study [2] only focused on the effect of nutritional status on the inflammatory response of the patients but neglected an important point: the changes in blood cells such as "leukocytes, lymphocytes, neutrophils, and platelets" were also closely related to the occurrence of bone marrow suppression (BMS) in the patients [3]. Figure 7 illustrates that the nutritional intervention program in this study can reduce the occurrence of BMS in patients to a certain extent.

We couldn't agree more with the letter's suggestion. We would like to include the benefits of improved BMS in the Discussion section of our study, which would primarily include the following:

- Improvement in BMS can contribute to the patient's quality of life during chemotherapy [4].
- Reversal of the hematocrit condition corresponds to a reduction in the number or degree of the patient's chemotherapy symptoms: elevated neutrophils lead to lower levels of fatigue, fewer persistent or chronic infections, and less nausea/vomiting [5].
- An increase in hemoglobin improved the patient's appetite and reduced vertigo [6].
- An increase in the number of platelets reduced the incidence of hemorrhage [7].

Conflicts of Interest

None declared.

References

1. Lian C, Fan Y, Li J. Clarifying blood indices in patients with ovarian cancer. JMIR Mhealth Uhealth 2025;13:e70895. [doi: 10.2196/70895]
2. Tian X, Liu Y, Zhang J, et al. Efficacy of a WeChat-Based, multidisciplinary, full-course nutritional management program on the nutritional status of patients with ovarian cancer undergoing chemotherapy: randomized controlled trial. JMIR Mhealth Uhealth 2024 Nov 4;12:e56475. [doi: 10.2196/56475] [Medline: 39496160]
3. Ten Berg MJ, van den Bemt PMLA, Shantakumar S, et al. Thrombocytopenia in adult cancer patients receiving cytotoxic chemotherapy: results from a retrospective hospital-based cohort study. Drug Saf 2011 Dec 1;34(12):1151-1160. [doi: 10.2165/11594310-000000000-00000] [Medline: 22077503]
4. Crawford J, Herndon D, Gmitter K, Weiss J. The impact of myelosuppression on quality of life of patients treated with chemotherapy. Future Oncol 2024;20(21):1515-1530. [doi: 10.2217/fon-2023-0513] [Medline: 38587388]
5. Crawford J, Moore DC, Morrison VA, Dale D. Use of prophylactic pegfilgrastim for chemotherapy-induced neutropenia in the US: a review of adherence to present guidelines for usage. Cancer Treat Res Commun 2021;29:100466. [doi: 10.1016/j.ctarc.2021.100466] [Medline: 34655862]

XSL•FO
RenderX

6. Bryer E, Henry D. Chemotherapy-induced anemia: etiology, pathophysiology, and implications for contemporary practice. Int J Clin Transfus 2018 Oct 23;6:21-31. [doi: 10.2147/IJCTM.S187569]
7. Thrombocytopenia (low platelet count). American Cancer Society. URL: https://www.cancer.org/cancer/managing-cancer/side-effects/low-blood-counts/bleeding.html [accessed 2025-08-06]

Abbreviations

BMS: bone marrow suppression

Edited by TDA Cardoso; submitted 25.03.25; this is a non–peer-reviewed article;accepted 23.07.25; published 13.08.25.

Please cite as:
Zhang J
Authors' Reply: Clarifying Blood Indices in Patients With Ovarian Cancer
JMIR Mhealth Uhealth 2025;13:e74931
URL: https://mhealth.jmir.org/2025/1/e74931
doi:10.2196/74931

© Jiahua Zhang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Authors' Reply: Methodological Considerations in Evaluating Mental Health Apps: Ensuring Reliability and Patient Safety

Seema Mehrotra[1], PhD; Ravikesh Tripathi[1], PhD; Pramita Sengupta[1], PhD; Abhishek Karishiddimath[1], MSc; Angelina Francis[1], MA; Pratiksha Sharma[1], MSc; Paulomi Sudhir[1], PhD; T K Srikanth[2], PhD; Girish Rao[3], MD; Rajesh Sagar[4], MD

[1]Department of Clinical Psychology, National Institute of Mental Health and Neurosciences, Hosur Road, Bengaluru, India

[2]E-Health Research Centre, International Institute of Information Technology Bangalore, Bengaluru, India

[3]Centre for Public Health, Department of Epidemiology, National Institute of Mental Health and Neurosciences, Bengaluru, India

[4]Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, India

Corresponding Author:
Seema Mehrotra, PhD
Department of Clinical Psychology, National Institute of Mental Health and Neurosciences, Hosur Road, Bengaluru, India

Related Articles:

https://mhealth.jmir.org/2025/1/e85329

https://mhealth.jmir.org/2025/1/e79238

(JMIR Mhealth Uhealth 2025;13:e85910) doi:10.2196/85910

KEYWORDS

mental health apps; mHealth; review of apps; smartphone apps; MHapps for indian users; India

We thank Balakrishna [1] for the letter to the editor in response to our article [2] on the systematic review of mental health apps accessible to Indian users and appreciate both their interest and the critical points raised.

As described in our paper, we adopted a multiphased approach to interrater reliability. This included initial training of reviewers; independent rating of 3 apps by 4 primary reviewers, followed by joint discussions with mentors; joint review of 3 additional apps, leading to the creation of an internal guide; independent review of 6 more apps and further mentor discussions; detailed independent review of the remaining apps; and mentor review of a subset of 7 reviewed apps, along with joint rating of all apps flagged as doubtful [2]. We acknowledge the concern regarding the absence of intraclass correlation coefficients due to the lack of dual ratings across all apps. It is a limitation noted in our paper. Given the broad search strategy, with use of 15 mental health–related terms, we had a large set of apps to review even after initial screening (n=792). Moreover, our review extended beyond the Mobile Application Rating Scale to include additional predefined parameters, requiring substantial time and effort. The dynamic nature of app store content further constrained the timeline, limiting the feasibility of dual independent ratings for all apps. Nonetheless, we believe our sequential, collaborative approach helped mitigate rating inconsistencies to a reasonable extent. We appreciate the suggestion to report intraclass correlation coefficients for a random subset, and this exercise is already underway.

Regarding sampling scope, as noted in the Limitations section of our paper, we evaluated only free apps, apps offering free trials, and the freely accessible portions of paid apps. This may have led to an incomplete basis for rating certain apps, and we agree that our findings may not fully represent the broader app ecosystem. However, information regarding involvement of mental health professionals and empirical research was drawn from app store descriptions and "about the app" sections, which were accessible even for paid apps. Our focus on free and partially free apps reflects what is most accessible to users and our findings align with patterns observed in other reviews [3-5]. That said, we acknowledge that ratings on some parameters would have benefited from access to full app content. This review is part of a planned recurring exercise, and in the next phase we will conduct detailed assessments of a random sample of fully paid apps, along with newer entries in this dynamic and evolving space.

Lastly, we concur with the observation that adverse events in clinical trials of mental health apps remain underreported. Ours was a review study, and we hope that our identification of certain instances of unscientific or potentially misleading content will prompt further research into these aspects. Looking ahead, we are developing a web-based platform to guide users and professionals on key cautionary indicators—contextually relevant cues to assess app quality and support informed decision-making.

XSL•FO
RenderX

Acknowledgments

The original study was funded by the Indian Council of Medical Research as part of the funding support for the Centre for Advanced Research in Digital Interventions for Mental Health Care, NIMHANS, Bengaluru.

Conflicts of Interest

None declared.

References

1. Balakrishna H. Methodological considerations in evaluating mental health apps: ensuring reliability and patient safety. JMIR Mhealth Uhealth 2025;13:e85329. [doi: 10.2196/85329]
2. Mehrotra S, Tripathi R, Sengupta P, et al. Evaluating characteristics and quality of mental health apps available in app stores for Indian users: systematic app search and review. JMIR Mhealth Uhealth 2025 Sep 26;13:e79238. [doi: 10.2196/79238] [Medline: 41004798]
3. Singh S, Sharma P, Ghimire P, Shrestha R, Gnanavel S. Assessment of app store description and privacy policy to explore ethical and safety concerns associated with the use of mental health apps for depression. Indian J Psychol Med 2023 Mar;45(2):173-178. [doi: 10.1177/02537176221142046] [Medline: 36925500]
4. Martinengo L, Stona AC, Tudor Car L, Lee J, Griva K, Car J. Education on depression in mental health apps: systematic assessment of characteristics and adherence to evidence-based guidelines. J Med Internet Res 2022 Mar 9;24(3):e28942. [doi: 10.2196/28942] [Medline: 35262489]
5. Leong QY, Sridhar S, Blasiak A, et al. Characteristics of mobile health platforms for depression and anxiety: content analysis through a systematic review of the literature and systematic search of two app stores. J Med Internet Res 2022 Feb 4;24(2):e27388. [doi: 10.2196/27388] [Medline: 35119370]

Edited by S Brini; submitted 15.10.25; this is a non–peer-reviewed article;accepted 20.10.25; published 19.11.25.

Please cite as:
Mehrotra S, Tripathi R, Sengupta P, Karishiddimath A, Francis A, Sharma P, Sudhir P, Srikanth TK, Rao G, Sagar R
Authors' Reply: Methodological Considerations in Evaluating Mental Health Apps: Ensuring Reliability and Patient Safety
JMIR Mhealth Uhealth 2025;13:e85910
URL: https://mhealth.jmir.org/2025/1/e85910
doi:10.2196/85910

© Seema Mehrotra, Ravikesh Tripathi, Pramita Sengupta, Abhishek Karishiddimath, Angelina Francis, Pratiksha Sharma, Paulomi Sudhir, T K Srikanth, Girish Rao, Rajesh Sagar. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Methodological Considerations in Evaluating Mental Health Apps: Ensuring Reliability and Patient Safety

Harikrishnan Balakrishna, MBBS, MBA

Department of Health Services, Family Health Centre, Kilikolloor, Government of Kerala, Mangadu PO, Kollam, India

Corresponding Author:
Harikrishnan Balakrishna, MBBS, MBA
Department of Health Services, Family Health Centre, Kilikolloor, Government of Kerala, Mangadu PO, Kollam, India

Related Articles:

https://mhealth.jmir.org/2025/1/e79238

Comment in: https://mhealth.jmir.org/2025/1/e85910

(JMIR Mhealth Uhealth 2025;13:e85329) doi:10.2196/85329

KEYWORDS

mental health apps; mHealth; review of apps; smartphone apps; MHApps for Indian users; India; mobile phones

We read with great interest and appreciate Mehrotra et al [1] for their systematic review of mental health apps accessible to Indian users. Their assessment of 350 apps provides valuable insight into the current digital mental health landscape in India. Still, several methodological and reporting considerations require clarification. This letter addresses 3 key areas: interrater reliability, sampling scope, and patient safety considerations.

To establish interrater reliability, the Mobile Application Rating Scale (MARS) was developed and validated with independent dual ratings and calculation of intraclass correlation coefficients (ICCs) [2]. Mehrotra et al [1] describe efforts like joint training and expert validation of a small subsample but divide the remaining apps among 4 reviewers for independent evaluation. ICC values for the main ratings were not reported, an approach that deviates from the established MARS methodology [2]. A past study demonstrated that when MARS is applied without strict dual-rater methodology, interrater agreement can be low (Krippendorff α=0.29), indicating high variability among reviewers [3]. Without ICC data, it is difficult to determine whether differences in app quality scores reflect their true variation or rater inconsistency. Reporting ICCs for a random subsample or acknowledging this limitation would improve interpretability for clinical application.

Regarding sampling scope, World Health Organization (WHO) guidance emphasizes transparent reporting and clear description of inclusion criteria for digital health interventions to support accurate interpretation and generalizability [4]. The study assessed only free apps and freely accessible portions of paid apps, yet its conclusions are framed as applying to all “mental health apps available to Indian users.” Premium apps often differ from free ones in professional input, empirical validation, and privacy safeguards. The finding that 65% of apps lacked professional input and only 11% cited empirical research [1] may therefore not represent the entire app ecosystem. This approach limits its generalizability and could mislead policymakers and clinicians relying on evidence-based recommendations. Hence, conclusions should be categorically limited to the free or freemium app marketplace or, in future work, expanded to include paid apps.

Regarding patient safety considerations, WHO guidance emphasizes that systematic evaluations of digital health interventions should transparently identify and report potential harms to support safe implementation and evidence-based decision-making [4]. The authors identified apps with potentially harmful features, such as misleading claims or alarming feedback without crisis support, but did not indicate whether these concerns were addressed or reported in any way. Vulnerable patients experiencing mental health crises could have worsened outcomes when exposed to such inadequately designed assessment tools [5]. Empirical evidence shows that adverse events in consumer-facing mental health apps are underreported and may pose significant risks to users [5]. Highlighting such risks ensures that clinicians, policymakers, and end users are aware of potential harms when translating review findings into practice, without implying that reviewers are responsible for regulatory action.

Clarifying interrater reliability, explicitly defining app sampling scope, and transparently reporting safety-relevant issues strengthen the interpretability and applicability of systematic reviews in digital mental health. Methodological limitations, such as inconsistent rating procedures or selective sampling, reduce confidence in app quality scores and limit generalizability. While patient safety considerations are primarily reporting issues, systematically highlighting potentially harmful apps enhances the practical utility of reviews and supports safe evidence-based decisions.

XSL•FO
RenderX

Acknowledgments

Minor language refinement and reference formatting were assisted by ChatGPT (GPT-5; OpenAI). All factual content, data, and interpretations were authored, verified, and approved entirely by the author.

Conflicts of Interest

None declared.

References

1. Mehrotra S, Tripathi R, Sengupta P, et al. Evaluating characteristics and quality of mental health apps available in app stores for Indian users: systematic app search and review. JMIR Mhealth Uhealth 2025 Sep 26;13:e79238. [doi: 10.2196/79238] [Medline: 41004798]
2. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
3. Miró J, Llorens-Vernet P. Assessing the quality of mobile health-related apps: interrater reliability study of two guides. JMIR Mhealth Uhealth 2021 Apr 19;9(4):e26471. [doi: 10.2196/26471] [Medline: 33871376]
4. Recommendations on Digital Interventions for Health System Strengthening: World Health Organization; 2019. URL: https://www.ncbi.nlm.nih.gov/books/NBK541905/ [accessed 2025-11-05]
5. Taher R, Hsu CW, Hampshire C, et al. The safety of digital mental health interventions: systematic review and recommendations. JMIR Ment Health 2023 Oct 9;10:e47433. [doi: 10.2196/47433] [Medline: 37812471]

Abbreviations

ICC: intraclass correlation coefficient
MARS: Mobile Application Rating Scale
WHO: World Health Organization

Edited by S Brini; submitted 05.10.25; this is a non–peer-reviewed article;accepted 19.10.25; published 19.11.25.

Please cite as:
Balakrishna H
Methodological Considerations in Evaluating Mental Health Apps: Ensuring Reliability and Patient Safety
JMIR Mhealth Uhealth 2025;13:e85329
URL: https://mhealth.jmir.org/2025/1/e85329
doi:10.2196/85329

© Harikrishnan Balakrishna. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing Self-Management Support Apps for Spinal Cord Injury: The Missing Role of Caregivers

Cheila Pessoa[1,2], MSc

[1]Spinal Cord Injury Rehabilitation Unit, Centro de Reabilitação do Norte - Dr. Ferreira Alves (CRN), Valadares, Centro Hospitalar de Vila Nova de Gaia, Avenida Infante Sagres 349, Vila Nova de Gaia, Portugal

[2]Escola Superior de Enfermagem do Porto, Porto, Portugal

Corresponding Author:
Cheila Pessoa, MSc
Spinal Cord Injury Rehabilitation Unit, Centro de Reabilitação do Norte - Dr. Ferreira Alves (CRN), Valadares, Centro Hospitalar de Vila Nova de Gaia, Avenida Infante Sagres 349, Vila Nova de Gaia, Portugal

Related Article:

https://mhealth.jmir.org/2024/1/e53677

(JMIR Mhealth Uhealth 2025;13:e72037) doi:10.2196/72037

KEYWORDS

mobile phone; mHealth; eHealth; telemedicine; telehealth; spinal cord injury; self-management; internet-based intervention; World Wide Web; systematic review; caregiver

I read, with great interest, the recent article by the team from Swiss Paraplegic Research, Nottwil, Switzerland, titled *Self-Management Support Apps for Spinal Cord Injury: Results of a Systematic Search in App Stores and Mobile App Rating Scale Evaluation* [1].

The authors conducted a comprehensive review of available mobile apps aimed at supporting self-management for individuals with spinal cord injury (SCI). Their work provides valuable insights into the current landscape of digital tools in this domain.

As a rehabilitation nurse specializing in SCI at the North Rehabilitation Centre of Portugal, having worked in this field since 2017, I have dedicated my career to researching the rehabilitation process of patients with SCI, particularly the training and empowerment of caregivers [2]. This experience, coupled with my research and app development background, informs my perspective on the study by Bernard et al [1].

I commend the authors for their rigorous systematic search of mobile apps designed to support self-management for individuals with SCI. However, I would like to highlight a critical aspect that may warrant further consideration: the exclusion of caregiver-centered apps. Many patients with SCI, particularly those with cervical injuries, experience severe functional limitations that hinder their ability to independently use self-management apps. As emphasized by Armstrong-Wood et al [3], individuals with high-level SCI may struggle with smartphone accessibility, which underscores the importance of caregiver involvement in the creation of these apps and caregivers' presence during app usage.

A crucial point raised in the study is the role of technological companies in developing self-management apps without direct involvement from SCI specialists or end users. While collaborations with researchers are mentioned, it is essential that these partnerships are more structured and systematic, ensuring that apps meet the real needs of patients with SCI and their caregivers. Without this validation process, there is a risk of developing tools that do not effectively address the complexities of SCI self-management, which calls for a more standardized and research-driven approach to app development.

Furthermore, the study's focus on English-language apps presents a potential limitation. While the authors acknowledge this constraint, restricting searches to English-language apps may exclude valuable resources tailored to non–English-speaking populations, limiting the generalizability of findings.

The integration of validation studies and collaborative research is fundamental to enhancing the effectiveness, usability, and real-world applicability of digital tools in SCI rehabilitation. Expanding this approach will ensure that these tools are not only evidence-based but also practical and adaptable to diverse patient needs.

The study identifies the absence of gamification elements in most reviewed apps, presenting an opportunity for innovation, as interactive learning approaches have demonstrated efficacy in enhancing engagement and knowledge retention [4].

I appreciate the authors' valuable contribution to this evolving field and encourage future research to expand the scope of app evaluations to include caregiver-focused tools, multilingual resources, and usability considerations for individuals with functional impairments. These additions would provide a more comprehensive understanding of digital health solutions in SCI rehabilitation.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of "Self-Management Support Apps for Spinal Cord Injury: Results of a Systematic Search in App Stores and Mobile App Rating Scale Evaluation" declined to respond to this letter.

References

1. Bernard RM, Seijas V, Davis M, et al. Self-management support apps for spinal cord injury: results of a systematic search in app stores and Mobile App Rating Scale evaluation. JMIR Mhealth Uhealth 2024 Dec 19;12:e53677. [doi: 10.2196/53677] [Medline: 39700493]
2. Pessoa C, Fernandes C, Nogueira P. CuizGiver: development of a digital game for the training of caregivers of individuals with spinal cord injury [Article in Portuguese]. Revista Portuguesa de Enfermagem de Reabilitação 2022;5(2). [doi: 10.33194/rper.2022.212]
3. Armstrong-Wood R, Messiou C, Kite A, et al. Smartphone accessibility: understanding the lived experience of users with cervical spinal cord injuries. Disabil Rehabil Assist Technol 2024 May;19(4):1434-1445. [doi: 10.1080/17483107.2023.2192246] [Medline: 37010939]
4. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644. [doi: 10.2196/18644] [Medline: 32773374]

Abbreviations

SCI: spinal cord injury

Edited by D Majekodunmi; submitted 01.02.25; this is a non–peer-reviewed article;accepted 10.03.25; published 14.04.25.

Please cite as:
Pessoa C
Enhancing Self-Management Support Apps for Spinal Cord Injury: The Missing Role of Caregivers
JMIR Mhealth Uhealth 2025;13:e72037
URL: https://mhealth.jmir.org/2025/1/e72037
doi:10.2196/72037

© Cheila Pessoa. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Clinical Efficacy of Multimodal Exercise Telerehabilitation Based on AI for Chronic Nonspecific Low Back Pain: Randomized Controlled Trial

Chongwu Xiao[1,2,3,4,5*], MD; Yijin Zhao[1,2*], MD; Gege Li[1,2*], MS; Zhuodong Zhang[1,2], MS; Siyu Liu[6], BS; Weichao Fan[1,7], MS; Jinjing Hu[1,2], MS; Qiuru Yao[1,7], MS; Chengduan Yang[1,2], PhD; Jihua Zou[1,2,8], MS; Qing Zeng[1,2], MD; Guozhi Huang[1,2], MD

[1]Center for Rehabilitation Medicine, Zhujiang Hospital, Southern Medical University, No.253, Industrial Avenue Middle Guangzhou, Guangzhou, China

[2]School of Rehabilitation Sciences, Southern Medical University, Guangzhou, China

[3]Department of Rehabilitation Medicine, The Second Affiliated Hospital of Guangxi Medical University, Guangxi Medical University, Nanning, China

[4]GuangDong Engineering Technology Research Center of Brain Function Assessment and Neuroregulation Rehabilitation, Guangzhou, China

[5]Institute of Exercise and Rehabilitation Science, Zhujiang Hospital, Southern Medical University, Guangzhou, China

[6]School of Sport Medicine and Physical Therapy, Beijing Sport University, Beijing, China

[7]School of Nursing, Southern Medical University, Guangzhou, China

[8]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China

[*]these authors contributed equally

Corresponding Author:
Guozhi Huang, MD
Center for Rehabilitation Medicine, Zhujiang Hospital, Southern Medical University, No.253, Industrial Avenue Middle Guangzhou, Guangzhou, China

Related Article:

Correction of: https://mhealth.jmir.org/2025/1/e56176

(JMIR Mhealth Uhealth 2025;13:e78188) doi:10.2196/78188

In "Clinical Efficacy of Multimodal Exercise Telerehabilitation Based on AI for Chronic Nonspecific Low Back Pain: Randomized Controlled Trial" (JMIR Mhealth Uhealth 2025;13:e56176) the authors noted one omission.

In the Acknowledgements, the following sentence has been added:

Jihua Zou, Qing Zeng, and Guozhi Huang are corresponding authors and contributed equally to this work.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 28.05.25; this is a non–peer-reviewed article;accepted 29.05.25; published 06.06.25.

Please cite as:
Xiao C, Zhao Y, Li G, Zhang Z, Liu S, Fan W, Hu J, Yao Q, Yang C, Zou J, Zeng Q, Huang G
Correction: Clinical Efficacy of Multimodal Exercise Telerehabilitation Based on AI for Chronic Nonspecific Low Back Pain: Randomized Controlled Trial
JMIR Mhealth Uhealth 2025;13:e78188
URL: https://mhealth.jmir.org/2025/1/e78188
doi:10.2196/78188

© Chongwu Xiao, Yijin Zhao, Gege Li, Zhuodong Zhang, Siyu Liu, Weichao Fan, Jinjing Hu, Qiuru Yao, Chengduan Yang, Jihua Zou, Qing Zeng, Guozhi Huang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.6.2025.

XSL•FO
RenderX

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

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Use and Engagement With Low-Intensity Cognitive Behavioral Therapy Techniques Used Within an App to Support Worry Management: Content Analysis of Log Data

Paul Farrand[1,2*], PhD; Patrick J Raue[3*], PhD; Earlise Ward[4*], PhD; Dean Repper[5*], MSc; Jonathan Baker[6*], MA; Patricia Areán[3*], PhD

[1]Clinical Education, Development and Research, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom

[2]Department of Psychology, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom

[3]AIMS CENTER, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

[4]School of Medicine and Public Health, Carbone Comprehensive Cancer Center, University of Wisconsin-Madison, Madison, WI, United States

[5]Trent PTS, Improving Access to Psychological Therapies, Derby, United Kingdom

[6]Iona Mind Inc, Romford, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Paul Farrand, PhD
Clinical Education, Development and Research
Faculty of Health and Life Sciences
University of Exeter
Sir Henry Wellcome Building for Mood Disorders Reserach
Perry Road
Exeter, EX4 4QG
United Kingdom
Phone: 44 01392725793
Email: p.a.farrand@exeter.ac.uk

Related Article:

Correction of: https://mhealth.jmir.org/2024/1/e47321

(JMIR Mhealth Uhealth 2025;13:e76573) doi:10.2196/76573

In "Use and Engagement With Low-Intensity Cognitive Behavioral Therapy Techniques Used Within an App to Support Worry Management: Content Analysis of Log Data" (JMIR Ment Health 2024;12:e47321), the authors made the following changes.

The authorship was previously published as:

> *Paul Farrand[1,2*], PhD; Patrick J Raue[3*], PhD; Earlise Ward[4*], PhD; Dean Repper[5*], MSc; Patricia Areán[3*], PhD*

The following author, equal contribution tag, ORCID, and associated affiliation have been added in the fifth position of the authorship:

> *Jonathon Baker[6*], MA (ORCID 0009 0000 9923 5042)*
>
> *Iona Mind Inc, Romford, United Kingdom*

The Authors' Contributions was revised from:

> *PF, PA, and PJR conceptualized and designed the project with EW and DR providing theoretical input informing the background of the paper. PF wrote the initial draft of the paper with all authors contributing to the development of the paper, interpretation of the analysis, and editing of the final manuscript. PF was not involved in any part of data analysis.*

To:

> *PF, PA, and PJR conceptualized and designed the project with EW and DR providing theoretical input informing the background of the paper and JB extracting and organising data for analysis, drafting the section describing the app and clarifying any uncertainties regarding description of the app in relation to descriptive log-data. PF wrote the initial draft of the paper with all authors contributing to the development of the paper, interpretation of the analysis, and editing of the final manuscript. PF and JB were not involved in any part of data analysis.*

Additionally, the Conflicts of Interest section previously appeared as:

XSL•FO
RenderX

PF is currently on a paid sabbatical with Iona Mind from the University of Exeter. All other authors have confirmed they have no conflicts of interest to declare.

And will now be changed to:

JB is a full time employee and PF currently on a paid sabbatical with Iona Mind from the University of Exeter. All other authors have confirmed they have no conflicts of interest to declare.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 26.04.25; this is a non–peer-reviewed article;accepted 30.04.25; published 13.05.25.

Please cite as:
Farrand P, Raue PJ, Ward E, Repper D, Baker J, Areán P
Correction: Use and Engagement With Low-Intensity Cognitive Behavioral Therapy Techniques Used Within an App to Support Worry Management: Content Analysis of Log Data
JMIR Mhealth Uhealth 2025;13:e76573
URL: https://mhealth.jmir.org/2025/1/e76573
doi:10.2196/76573
PMID:

©Paul Farrand, Patrick J Raue, Earlise Ward, Dean Repper, Jonathan Baker, Patricia Areán. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Participant Adherence in Repeated-Dose Clinical Studies Using Video-Based Observation: Retrospective Data Analysis

Jihong Song[1], BSN; Sungpil Han[1,2], MD, PhD; Suein Choi[1,2], MD, PhD; Jonghyuk Lim[3], BE; Byeong Yeob Oh[3], BBA; Dongoh Shin[3], MEng; Seunghoon Han[1,2], MD, PhD

[1]Department of Pharmacology, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[2]Department of Clinical Pharmacology and Therapeutics, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[3]CareSquare, Inc., Seoul, Republic of Korea

Corresponding Author:
Seunghoon Han, MD, PhD
Department of Clinical Pharmacology and Therapeutics
Seoul St. Mary's Hospital, College of Medicine
The Catholic University of Korea
222 Banpo-daero, Seocho-gu
Seoul, 06591
Republic of Korea
Phone: 82 10 9484 2415
Fax: 82 2 3147 8357
Email: waystolove@catholic.ac.kr

Related Article:

Correction of: https://mhealth.jmir.org/2025/1/e65668

(JMIR Mhealth Uhealth 2025;13:e79507) doi:10.2196/79507

In "Participant Adherence in Repeated-Dose Clinical Studies Using Video-Based Observation: Retrospective Data Analysis" (JMIR JMIR Mhealth Uhealth 2025;13:e65668), the authors noted three errors.

- The order of authors was incorrectly listed in the originally published article. This has been changed from:

  Seunghoon Han, MD, PhD; Jihong Song, BSN; Sungpil Han, MD, PhD; Suein Choi, MD, PhD; Jonghyuk Lim, BE; Byeong Yeob Oh, BBA; Dongoh Shin, MEng

To:

Jihong Song, BSN; Sungpil Han, MD, PhD; Suein Choi, MD, PhD; Jonghyuk Lim, BE; Byeong Yeob Oh, BBA; Dongoh Shin, MEng; Seunghoon Han, MD, PhD

- In the subsection of "Methods" titled "The System Overview," casing type was changed from:

  (2) Verified Deviated Dosing, (3) Unverified Dosing, (4) Missed Dosing

To:

(2) Verified deviated dosing, (3) Unverified dosing, (4) Missed dosing

- The values in the final column of Table 2, "Allowed time window for SAI (min)," were adjusted to more accurately represent the intended time windows. These values have now been provided in a range rather than a single negative value, and the table will appear as follows:

XSL•FO
RenderX

Table 2. Self-administration of the investigational medicinal product-specific characteristics by cohort.

Cohort	SAI[a] schedule				Dosing notification (min before the planned dosing time)	Dosing reminder (min after the planned dosing time)	Allowed time window for SAI (min)
	SAI day	Number of total SAI days	Number of SAIs per day	Planned dosing time[b]			
1	2-4, 6-8, 10-12, 14-16, 18, 19	14	1	9 AM	–60, –30, 0	15, 30, and 45	–60 to 60
2	1, 2, 6, 7, 9-11	7	1	9 AM	N/A[c]	N/A	–60 to 60
3	1, 2, 4-6, 10, 11	7	1	9 AM	N/A	N/A	–60 to 60
4	1, 2, 12-15, 17-20, 24, 25	12	1	9 AM	N/A	N/A	–60 to 60
5	1-4, 13-17, 21-23	12	1	9 AM	–60, –30, 0	15, 30, and 45	–60 to 60
6	1-32	32	2	Variable[d]	–60, –15, 0	60 and 110	–120 to 120
7	3, 5, 7, 9, 10, 12-14, 16-21, 23-30, 32-45, 47-60, 62-75, 77-90	78	1	9 AM	–30, –5, 0	10, 20, 165, and 665	–180 to 180
8	1-28	28	2	Variable[d]	–60, –10, 0	60, 180, and 350	–120 to 360
9	3-7	5	2[e], 3	8 AM, 2 PM, and 8 PM	–30, 0	N/A	–120 to 120
10	4-8, 11	6	2[e], 3	8 AM, 2 PM, and 8 PM	–30, 0	N/A	–120 to 120
11	5-9, 12	6	2[e], 3	8 AM, 2 PM, and 8 PM	–30, 0	N/A	–120 to 120
12	2-4, 8-96	92	1[e], 2[e], 4	1 AM, 9 AM, 9:10 AM, 5 PM	–10, 0	10	N/A
13	1-89, 92-96	94	2[e], 3	1 AM, 9 AM, 5 PM	–10, 0	10	N/A
14	2, 3, 5, 6, 8, 9, 11	7	1	9 AM	–30, 0	30	–60 to 60

[a]SAI: self-administration of the investigational medicinal product.

[b]The presented data refer to the reference point for the planned dosing time (eg, 9 AM) in each cohort, with minor variability across participants (eg, 9:02 AM, 9:04 AM, 9:06 AM, etc).

[c]N/A: not applicable.

[d]The planned dosing times vary across participants, and no common reference point can be provided.

[e]SAI event and dosing by the investigator on the same day.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 29.06.25; this is a non–peer-reviewed article;accepted 02.07.25; published 30.07.25.

Please cite as:
Song J, Han S, Choi S, Lim J, Oh BY, Shin D, Han S
Correction: Participant Adherence in Repeated-Dose Clinical Studies Using Video-Based Observation: Retrospective Data Analysis
JMIR Mhealth Uhealth 2025;13:e79507
URL: https://mhealth.jmir.org/2025/1/e79507
doi:10.2196/79507
PMID:40736914

XSL•FO
RenderX

©Jihong Song, Sungpil Han, Suein Choi, Jonghyuk Lim, Byeong Yeob Oh, Dongoh Shin, Seunghoon Han. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.07.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: The Role of Environmental Factors in Technology-Assisted Physical Activity Intervention Studies Among Older Adults: Scoping Review

Carl-Philipp Jansen[1,2], PhD; Désirée Nijland[3], PhD; Jean-Michel Oppert[4], PhD; Veysel Alcan[5], PhD; Kirsi E Keskinen[6], PhD; Emmi Matikainen-Tervola[7], PhD; Zada Pajalic[8], PhD; Merja Rantakokko[6,9], PhD; Signe Tomsone[10], PhD; Essi-Mari Tuomola[6], PhD; Erja Portegijs[3], PhD; Erik J Timmermans[11], PhD

[1]Geriatric Center, Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany
[2]Clinic for Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany
[3]Department of Human Movement Sciences, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[4]Department of Nutrition, Pitié-Salpêtrière Hospital (AP-HP), Sorbonne University, Paris, France
[5]Department of Electrical and Electronics Engineering, Engineering Faculty, Tarsus University, Tarsus, Turkey
[6]Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[7]Institute of Rehabilitation, JAMK University of Applied Sciences, Jyväskylä, Finland
[8]Campus Drammen, Faculty of Health and Social Sciences, University of South-Eastern Norway, Oslo, Norway
[9]The Wellbeing Services County of Central Finland, Jyväskylä, Finland
[10]Department of Rehabilitation, Faculty of Health and Sport Sciences, Rīga Stradiņš University, Riga, Latvia
[11]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Corresponding Author:
Carl-Philipp Jansen, PhD
Geriatric Center
Medical Faculty Heidelberg
Heidelberg University
Rohrbacher Straße 149
Heidelberg, 69126
Germany
Phone: 49 062213191760
Email: Carl-Philipp.Jansen@med.uni-heidelberg.de

Related Article:

Correction of: https://mhealth.jmir.org/2025/1/e59570

(JMIR Mhealth Uhealth 2025;13:e83822) doi:10.2196/83822

In "The Role of Environmental Factors in Technology-Assisted Physical Activity Intervention Studies Among Older Adults: Scoping Review" (JMIR Mhealth Uhealth 2025;13(1):e59570) the authors noted one error.

The first sentence of the acknowledgement was an error, stating:

> *"This scoping review is being undertaken as part of the European Cooperation in Science and Technology Action "PhysAgeNet" (CA20104)."*

This sentence has been corrected to:

> *"This publication is based upon work from COST Action CA20104 – Network on evidence-based physical activity in old age (PhsAgeNet), supported by COST (European Cooperation in Science and Technology)."*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article will also be resubmitted to those repositories.

XSL•FO
RenderX

Submitted 09.09.25; this is a non–peer-reviewed article;accepted 10.09.25; published 18.09.25.

Please cite as:
Jansen CP, Nijland D, Oppert JM, Alcan V, Keskinen KE, Matikainen-Tervola E, Pajalic Z, Rantakokko M, Tomsone S, Tuomola EM, Portegijs E, Timmermans EJ
Correction: The Role of Environmental Factors in Technology-Assisted Physical Activity Intervention Studies Among Older Adults: Scoping Review
JMIR Mhealth Uhealth 2025;13:e83822
URL: https://mhealth.jmir.org/2025/1/e83822
doi:10.2196/83822
PMID:

©Carl-Philipp Jansen, Désirée Nijland, Jean-Michel Oppert, Veysel Alcan, Kirsi E Keskinen, Emmi Matikainen-Tervola, Zada Pajalic, Merja Rantakokko, Signe Tomsone, Essi-Mari Tuomola, Erja Portegijs, Erik J Timmermans. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Utility of Digital Phenotyping Based on Wrist Wearables and Smartphones in Psychosis: Observational Study

Zixu Yang[1*], BSc, MSSc; Creighton Heaukulani[2*], PhD; Amelia Sim[3], MBBS, MPH, MRCPsych; Thisum Buddhika[2], MSc; Nur Amirah Abdul Rashid[1], BPSych; Xuancong Wang[2], PhD; Shushan Zheng[3], MBBS, MRCPsych; Yue Feng Quek[1], BSc; Sutapa Basu[3], MBBS, DPM, MD; Kok Wei Lee[3], MBBS, MRCPsych; Charmaine Tang[3], MBBS, MRCPsych, MCI; Swapna Verma[3], MBBS, MD; Robert J T Morris[2,4], PhD; Jimmy Lee[1,3,5], MBBS, MMed, MCI

[1]North Region, Institute of Mental Health, Singapore, Singapore

[2]Ministry of Health Office for Healthcare Transformation, Singapore, Singapore

[3]Department of Psychosis, Institute of Mental Health, Singapore, Singapore

[4]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

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

[*]these authors contributed equally

Corresponding Author:
Jimmy Lee, MBBS, MMed, MCI
North Region, Institute of Mental Health, Singapore, Singapore

Abstract

Background: Digital phenotyping provides insights into an individual's digital behaviors and has potential clinical utility.

Objective: In this observational study, we explored digital biomarkers collected from wrist-wearable devices and smartphones and their associations with clinical symptoms and functioning in patients with schizophrenia.

Methods: We recruited 100 outpatients with schizophrenia spectrum disorder, and we collected various digital data from commercially available wrist wearables and smartphones over a 6-month period. In this report, we analyzed the first week of digital data on heart rate, sleep, and physical activity from the wrist wearables and travel distance, sociability, touchscreen tapping speed, and screen time from the smartphones. We analyzed the relationships between these digital measures and patient baseline measurements of clinical symptoms assessed with the Positive and Negative Syndrome Scale, Brief Negative Symptoms Scale, and Calgary Depression Scale for Schizophrenia, as well as functioning as assessed with the Social and Occupational Functioning Assessment Scale. Linear regression was performed for each digital and clinical measure independently, with the digital measures being treated as predictors.

Results: Digital data were successfully collected from both the wearables and smartphones throughout the study, with 91% of the total possible data successfully collected from the wearables and 82% from the smartphones during the first week of the trial—the period under analysis in this report. Among the clinical outcomes, negative symptoms were associated with the greatest number of digital measures (10 of the 12 studied here), followed by overall measures of psychopathology symptoms, functioning, and positive symptoms, which were each associated with at least 3 digital measures. Cognition and cognitive/disorganization symptoms were each associated with 1 or 2 digital measures.

Conclusions: We found significant associations between nearly all digital measures and a wide range of symptoms and functioning in a community sample of individuals with schizophrenia. These findings provide insights into the digital behaviors of individuals with schizophrenia and highlight the potential of using commercially available wrist wearables and smartphones for passive monitoring in schizophrenia.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-046552

(JMIR Mhealth Uhealth 2025;13:e56185) doi:10.2196/56185

KEYWORDS

schizophrenia; psychosis; digital phenotyping; wrist wearables; mobile phone

Introduction

Schizophrenia is a debilitating condition affecting at least 20 million individuals worldwide [1]. This lifelong condition [2] is frequently associated with impairments in personal, social, occupational, and other important areas of daily function [3]. Despite advancements in pharmacology and psychosocial interventions, recovery rates among patients have not improved

XSL•FO
RenderX

[4], and the high rate of relapse places a heavy burden on patients and their caregivers [5].

In present-day psychiatric care, the clinical status of patients is determined at periodic consultations via clinician interviews and self-reported questionnaires. The episodic nature of these visits and the removal of patients from their natural environment result in limited assessments that do not always reflect the state of patients in different environments and over varying periods of time [6]. Moreover, there is a risk of unreliable self-reporting in patients with schizophrenia, influenced by cognitive impairments, biases, and self-reporting confounders [7].

Digital phenotyping may allow care teams to circumvent these limitations. Digital phenotyping is defined as the moment-by-moment quantification of the individual-level human phenotype in situ using data from smartphones and other personal digital devices [8,9]. Through personal digital devices, insightful and objective longitudinal data, such as on sleep and mobility, can be unobtrusively and continuously collected from patients [10].

There has been growing interest in digital phenotyping in psychiatry [11,12]. People with schizophrenia have shown willingness to use smartphones to monitor their condition [13]. Furthermore, preliminary trials have shown that the long-term use of wearable technology is feasible and acceptable by people with schizophrenia and that paranoia toward devices was not a significant barrier [14-16].

There is also some evidence that wearable and mobile devices can deliver clinically relevant information in real-world contexts. Movement data estimated from a 3-axis accelerometer were negatively correlated with negative symptoms assessed with the Positive and Negative Syndrome Scale (PANSS) in patients with schizophrenia [17]. The number of steps was negatively correlated with PANSS positive-factor, negative-factor, general subscales, and total score, in patients with schizophrenia in an inpatient setting [18]. Sleep duration estimated from mobile phone–based accelerometer data was moderately correlated with subject self-assessment of sleep duration [19]. Geolocation and motor activity inferred from gyroscopes and data from the ambient light sensor, accelerometer, and other sensors in smartphones were used as indices of negative symptoms [20,21]. Data from anonymized call and text message logs may be potential markers for social functioning, and data on screen touches, such as typing speed, may provide novel clinical data about a user's cognitive functioning [9].

Despite this evidence, past studies have noteworthy limitations, such as small sample sizes [17,22], nonclinical samples [23,24], being conducted in closed-ward settings [18], and having used made-for-research wearable devices [22], which would affect the generalizability of the research findings or scaling up to larger populations. To date, few studies have reported digital phenotyping measures collected with both low-cost, commercially available digital wearable devices and smartphones among patients with schizophrenia in a community setting.

The aim of this study was to provide an initial examination of the associations between multimodal digital markers and clinical outcomes in patients with schizophrenia.

Methods

Setting and Participants

The Health Outcomes via Positive Engagement in Schizophrenia (HOPE-S) study is a prospective observational study conducted at the Institute of Mental Health, Singapore, with recently discharged individuals with schizophrenia spectrum disorders. Details of the study have been previously reported [25]. In brief, participants were followed up for up to 6 months.Clinical, cognitive, and functioning data were collected at intervals of every 6 weeks, and digital data from smartphones and wrist wearables were collected continuously over the follow-up period. The clinical, cognitive, and functioning data collected at baseline and the digital data collected in the first week after study enrollment were used for this study. The study aimed to examine the associations between clinical, cognitive, functioning, and digital data to develop prediction algorithms that might detect clinically significant outcomes. We recruited 100 English-speaking patients aged 21 - 65 years between November 2019 and January 2023. One participant requested to be withdrawn from the study after completing the baseline assessment due to objections from family members. Of the remaining participants (n=99), 63 were diagnosed with schizophrenia, 12 with brief psychotic disorder, 11 with schizophreniform disorder, 5 with schizoaffective disorder, 5 with delusional disorder, and 3 with other specified psychotic disorder.

Ethical Considerations

Ethics approval was obtained from the National Healthcare Group's Domain Specific Review Board, Singapore (2019/00720). Written informed consent was obtained from all participants after the study was fully explained. Data were anonymized and encrypted to protect patient privacy. Participants were reimbursed SGD 50 (US $37.19) on completion of the baseline visit for their time, inconvenience, and transportation costs.

Clinical Symptom Severity Measures and Scales

Assessment data from baseline were used for the present analyses. Symptom severity was assessed on the PANSS [26]. The PANSS is a 30-item clinical scale measuring positive symptoms, negative symptoms, and general psychopathology. PANSS factor scores, including positive (PANSS_Pos), negative (PANSS_Neg), cognitive/disorganization (PANSS_Cog/Dis), depression/anxiety (PANSS_Dep/Anx), and hostility (PANSS_Hos), were calculated in accordance with a previously validated factor structure [27].

Negative symptoms were measured with the Brief Negative Symptom Scale (BNSS) [28], a well-established 13-item instrument designed to measure 5 negative symptom domains, including asociality, anhedonia, avolition, blunted affect, and alogia [29].

XSL•FO
RenderX

Depressive symptoms were assessed on the Calgary Depression Scale for Schizophrenia (CDSS) [30], a 9-item scale widely used for assessing depression in schizophrenia that differentiates between depression symptoms and the negative and positive symptoms in schizophrenia.

Neurocognition was assessed with the Brief Assessment of Cognition in Schizophrenia (BACS) [31], a cognitive battery that assesses attention, verbal and working memory, motor and processing speed, verbal fluency, and executive function. A composite *z* score was obtained and used for analysis [31]. The Social and Occupational Functioning Assessment Scale (SOFAS) [32] was used to evaluate the patients' social and occupational functioning.

Digital Data Collection and Preprocessing

Digital data were collected from both wrist wearables and smartphones. Each participant was issued a Fitbit Charge 3 or 4 (Fitbit Inc) at the beginning of the study. Participants were required to wear the Fitbit every day during the study period, including when asleep. The Health Outcomes through Positive Engagement and Self-Empowerment (HOPES) app [33] was installed on a participant's smartphone if the device used the Android operating system. In the circumstance that a participant's phone was not compatible with the HOPES app, we provided a phone for use during the study period. In this case, study coordinators aided in the patient's migration to the study phone, and the patient was instructed to use only the study phone as their primary mobile device during the study. Data collected from the wrist wearables and smartphones were automatically synchronized with the HOPES app and subsequently sent to the research premises for storage and further analysis. Data collected by the Fitbit included data from the heart rate monitor, pedometer (step counts), and sleep sensor, including sleep stages. Data collected from the smartphone included locational mobility features (through GPS), touchscreen tapping speed, and sociability indices from texts and calls (via both the mobile service provider and WhatsApp). Significant steps were taken to ensure data privacy, validate data quality, and engineer better features, which are covered in detail in a publication by Wang et al [33]. The subset of data analyzed in this paper is described in Multimedia Appendix 1. Table 1 summarizes the resulting digital measures used in our analyses.

Table . Digital measures collected by Fitbit wearables and smartphones.

Daily digital measures	Data source	Description
total_sleep_hrs	Wearable	Total hours of sleep, including naps but not including the "wake" stage (see below).
sleep_hrs_wake	Wearable	Total hours of sleep classified as the "wake" sleep stage by the Fitbit, estimated to be the time in bed not sleeping (ie, falling asleep, after waking up, and awaking during the night).
sleep_hrs_light	Wearable	Total hours of sleep classified as the "light" sleep stage by the Fitbit.
sleep_hrs_deep	Wearable	Total hours of sleep classified as the "deep" sleep stage by the Fitbit.
sleep_hrs_REM	Wearable	Total hours of sleep classified as the "REM" (rapid eye movement) sleep stage by the Fitbit.
sleep efficiency	Wearable	Sleep efficiency score from the Fitbit for the main sleep segment.
HR_asleep	Wearable	Average of heart rate samples while asleep, as measured by the Fitbit heart and sleep monitors.
total steps	Wearable	Total steps taken, as measured by the Fitbit.
dist_travelled	Smartphone	Total distance traveled, as computed by GPS processing routines.
mode_intertap_dist	Smartphone	Mode of a rolling 7-day distribution of intertap durations, excluding taps in games (a higher value indicates a slower tapping speed).
total_msg_sent	Smartphone	Total messages (texts and images) sent via mobile service provider and WhatsApp.
screen_time	Smartphone	Total time of smartphone use as represented by the power state logs.

For each daily digital measure, data collected during the 7 days following the baseline visit (not including the data on the day of the baseline visit) were used. The 1-week duration should account for any day-of-week effects in the digital measures. Digital markers were taken as the average of the observations for each digital measure over the week of interest, except for the *mode_intertap_dist* measure, for which the last valid (ie, nonmissing) value was taken since this is already a 7-day rolling window statistic. We required at least 4 valid observations in the week-long observation window to produce a nonnull marker.

XSL•FO
RenderX

Successful data collection rates for the wearables as the fraction of hour-long windows with at least 1 successfully collected heart rate sample in the window of study (ie, the 7 days [168 hours] following the baseline visit) were calculated. A successful data collection rate for the smartphones was calculated similarly using the ambient light sensor. Note that these 2 data sources were continually sampled by each device regardless of user action, so long as the participant was compliant, that is, the wearable was correctly worn, and the phone was powered on with the app running in the background.

Statistical Analysis

Descriptive statistics on clinical measures, digital data measures, and successful data collection rates were analyzed. The successful data collection rates were compared between groups of participants who used study phones and their own phones with the Mann-Whitney *U* test. As an exploratory analysis to discover possible digital markers as factors associated with different clinical measures, a multiple linear regression was performed independently for each pair of digital markers (independent variable) and clinical measure (dependent variable). In each regression model, age was included as a covariate along with an interaction term with the digital measure, which has been reported as a significant factor in relapse behavior, psychosocial functioning, and psychotic symptoms [34,35]. Because a significant difference in digital data collection rates was noticed between participants using a study-provided phone and those using their own phone, we also included a variable indicating whether the participant was using a study-provided phone (along with its interaction with the digital measure) in all models regressing onto a digital measure from the smartphone. To easily compare the relative effect sizes between different measures, both the digital markers and clinical scales were standardized across the population; an intercept term was therefore not included in the models. Due to the particularly skewed distributions of the *distance_travelled* and *total_msg_sent* features, a log(1+x) transform was applied to these features before standardization. As this was an exploratory analysis attempting to uncover trends in the digital markers, multiple testing corrections were not used. A heat map was used to show effect sizes and visualize trends across digital markers and clinical scales.

Results

Sample Characteristics

The demographic and clinical characteristics of the study sample are summarized in Table 2.

XSL•FO
RenderX

Table . Demographic and clinical characteristics of the study sample (n=99).

Characteristics	Values
Sex, n (%)	
Female	57 (57.6)
Male	42 (42.4)
Marital status, n (%)	
Single	85 (85.9)
Married	7 (7.1)
Separated or divorced	7 (7.1)
Ethnicity, n (%)	
Chinese	73 (73.7)
Malay	17 (17.2)
Indian	6 (6.1)
Other	3 (3.0)
Currently on antipsychotics, n (%)	
Yes	97 (98.0)
No	2 (2.0)
Age (years), mean (SD)	29.88 (6.72)
Years of education, mean (SD)	14.07 (2.27)
Duration of illness (years), mean (SD)	2.61 (4.00)
Antipsychotic dose[a] (mg), mean (SD)	484.37 (375.55)
PANSS [b] score, mean (SD)	
Total	51.11 (9.97)
Positive	8.49 (3.82)
Negative	13.20 (4.08)
Cognitive/disorganization	10.55 (2.50)
Depression/anxiety	8.85 (3.43)
Hostility	5.40 (1.89)
BNSS [c] score, mean (SD)	
Total	20.90 (9.30)
Anhedonia	5.51 (2.76)
Asociality	3.34 (1.55)
Avolition	3.30 (1.89)
Blunted	5.77 (3.89)
Alogia	2.82 (2.61)
CDSS[d] total score, mean (SD)	12.31 (3.55)
SOFAS[e] score, mean (SD)	53.96 (11.85)
BACS[f] score, mean (SD)	−1.31 (1.36)

[a]Antipsychotic dose was computed as total daily chlorpromazine equivalent (n=91).

[b]PANSS: Positive and Negative Syndrome Scale.

[c]BNSS: Brief Negative Symptom Scale.

[d]CDSS: Calgary Depression Scale for Schizophrenia.

[e]SOFAS: Social and Occupational Functioning Assessment Scale.

[f]BACS: Brief Assessment of Cognition in Schizophrenia.

XSL•FO
RenderX

Digital Data Collection Rate and Digital Data Profile

In total, 41 and 58 participants were issued a Fitbit Charge 3 and Charge 4, respectively. Thirty-nine participants used a smartphone provided by the study at the baseline visit, while the rest used their own phone. The successful data collection rate in the week following the baseline visit was 91% (SD 22%) for the Fitbit and 82% (SD 32%) for the smartphone. Those using a study-provided phone had a significantly higher successful data completion rate than those using their own phone (mean 91% vs 77%; Mann-Whitney *U* test: *P*<.001). Sensor-specific successful data collection rates are also provided in Multimedia Appendix 2. The sample size of valid digital measures (n=99), along with the means and SDs, are presented in Table 3.

Table . Summary of digital markers.

Visit-level digital markers	Values, mean (SD)
Total_sleep_hrs (n=92)	8.76 (2.13)
Sleep hrs-wake (n=91)	0.83 (0.30)
Sleep hrs-light (n=91)	5.32 (1.71)
Sleep hrs-REM (n=91)	1.68 (0.51)
Sleep hrs-deep (n=91)	1.23 (0.39)
Sleep_efficiency (n=91)	0.93 (0.03)
Total_steps (n=96)	9186.91 (5413.21)
Distance_travelled (km) (n=87)	27.64 (34.59)
HR_asleep (bpm) (n=74)	69.94 (10.06)
Mode_intertap_dist (ms) (n=96)	596.06 (389.08)
Total_msg_sent (n=92)	38.02 (65.16)
Screen_time (hrs) (n=92)	4.74 (3.64)

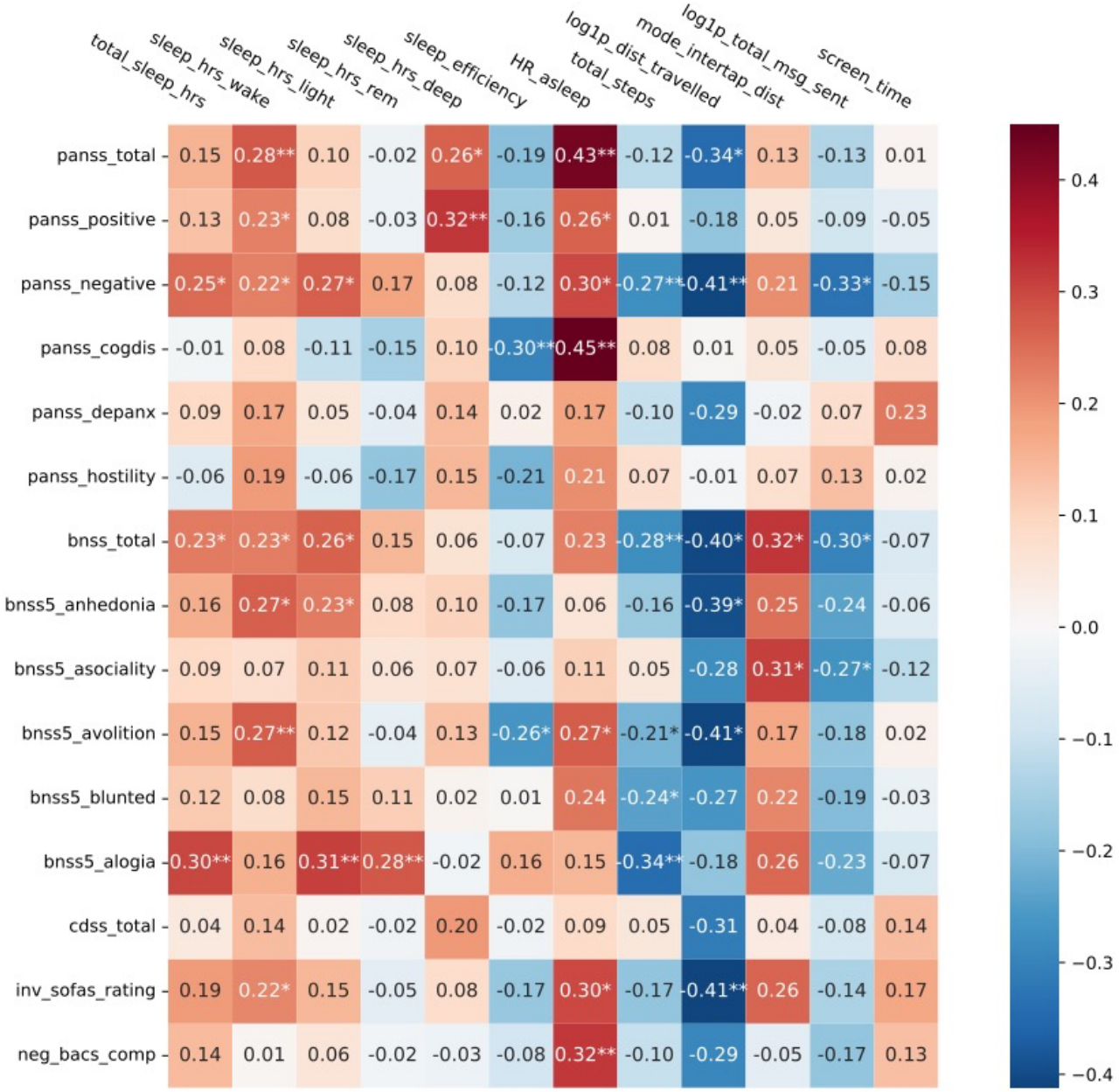
As expected, several digital measures were correlated with one another (such as sleep stage measures). The correlation coefficients between digital measures are provided in Multimedia Appendix 3.

Association Between Clinical Scales and Digital Markers

The linear regression coefficients of the digital measures are visualized for the different clinical measures as a heat map (Figure 1) at the mean age level (29.88 years) and for participants using their own phone. The numbers in the cells are the value of the regression coefficient, which also determines the cell color and intensity.

XSL•FO
RenderX

Figure 1. Regression coefficients between clinical scales and digital markers. Notes: A log(1+x) transformation was applied to *distance_travelled* and *total_msg_sent*. All clinical scales and digital markers were standardized prior to the regression. The Social and Occupational Functioning Assessment Scale (SOFAS) score was inverted (denoted as *inv_soafs_rating*, calculated as 100 minus the SOFAS rating) in the analyses to maintain a consistent direction among the scales and the digital markers, which aids visualization and interpretation. Analogously, the Brief Assessment of Cognition in Schizophrenia composite score was inverted (denoted as *neg_bacs_comp*) by taking the negative of its value. Ordinary least squares: *P_{OLS}<.05, **P_{OLS}<.01.



Starting with the symptoms and their associated digital biomarkers, first, we found that negative symptoms as assessed by both the BNSS and PANSS-negative were associated with the most digital measures (8 of the 12 being studied). Experiential deficits, that is, anhedonia and avolition, tended to be associated with less distance traveled and more hours spent awake in bed; fewer messages sent, a digital marker of sociability, tended to be associated with asociality. Expressive deficits, that is, affect blunting and alogia, tended to be negatively associated with steps taken. Alogia was additionally associated with sleep features such as total sleep hours and hours in light and rapid eye movement (REM) sleep. Second, PANSS-positive and SOFAS scores were associated with at least 3 different digital measures. Third, cognition and cognitive/disorganization symptoms were associated with 1 or 2 digital measures.

Among the digital measures, heart rate measured when asleep was associated with the most clinical measures. Specifically, a higher heart rate was associated with greater severity scores on PANSS-total and its factors (ie, positive, negative, and cognitive/disorganization) and BNSS avolition and asociality, as well as poorer functioning and cognitive performance. Time awake in bed and distance traveled were associated with mostly similar clinical measures, that is, negative symptoms, in particular anhedonia, avolition, and functioning. The only digital

XSL•FO
RenderX

measure not significantly associated with a clinical measure was screen time.

Discussion

Principal Findings

In this study, we examined the associations between digital markers (collected via commercially available wrist-wearable devices and smartphones) and a comprehensive range of clinical assessments in a sample of 99 recently discharged patients with schizophrenia spectrum disorders. Notable patterns of associations for digital measures were observed with psychiatric symptoms, specifically negative symptoms, functioning, and cognitive performance.

Longer total sleep time was positively associated with negative symptom domains; sleep efficiency was negatively associated with BNSS avolition and the PANSS cognitive/disorganization score. These results are in line with previous studies that reported associations between negative symptom severity and sleep disturbances (such as hypersomnia and sleep satisfaction) collected through self-reporting [36,37] and as measured by actigraphy [22,38]. One study also found no correlation between sleep markers collected via actigraphy and negative symptoms assessed via the PANSS-negative symptom subscale [39]. In addition, whether and how each sleep stage is associated with symptoms in schizophrenia remains inconclusive. A previous study reported that REM sleep duration was correlated with the Brief Psychiatric Rating Scale (BPRS) total score but not with the constituent negative symptom score [40]; in contrast, this study found no association between REM sleep duration and PANSS-total score but found a positive association with alogia. Use of nonspecific negative symptom assessment scales such as PANSS and BPRS, variations in sleep instruments, and other clinical factors, such as duration of illness and medication status [41], likely contributed to the mixed results in the literature.

Several types of passive digital phenotyping measures are associated with negative symptoms, such as phone use data (call/text logs, screen time), GPS data (distance traveled), sleep data (total sleep duration), and physical activity (step count). These findings support the use of passively collected digital phenotyping data as objective measures of negative symptoms and functional outcomes. These measures also provide longitudinal data that are usually not available directly to clinicians. Our study found a negative association between physical activity (as measured by step count) and negative symptom severity, particularly BNSS avolition, blunted affect, and alogia. This is in line with previous research that showed associations between decreased activity or sedentary behavior and negative symptoms in schizophrenia [17,18,22,42-44].

Distance traveled, messages sent and touchscreen tapping speed yielded the strongest signals associated with negative symptoms, specifically experiential deficits such as avolition, anhedonia and asociality. These results are consistent with previous studies that reported greater GPS mobility was associated with lower negative symptom severity and better community functioning [20,45]. Social behaviors (such as duration of outgoing calls and number of text messages) have been found to be predictive of psychotic symptoms and relapse in schizophrenia [46,47]. Typing-related data have been reported to be associated with mood symptoms and cognition in people without psychosis [48-50], but no such association was found in our study. The current study appears to be the first to report significant associations between negative symptoms, specifically asociality, and a messaging log feature (ie, the number of messages sent out) as well as touchscreen tapping speed.

We found that increased heart rate while asleep was associated with increases in positive psychopathology and worse cognitive performance. The relationship between increased heart rate and a greater severity of positive symptoms has often been documented [51-53], but only one previous study reported this relationship for heart rate during sleep, supporting our findings [54]. These associations likely indicate altered autonomic functioning, as indexed by increased heart rate, in response to stressors such as manifestations of psychotic symptoms [51,55]. Similarly, altered autonomic functioning may affect cognition. While no studies have reported the association of increased heart rate during sleep with worsening cognition, reduced heart rate variability has been linked to worse cognitive scores in schizophrenia [56].

The successful data collection rates in this study were high—about 91% for the Fitbit devices and 82% for the smartphones, which is comparable to the compliance rates in previous studies [15,19]. The good data collection rates support the potential of wrist wearables and smartphones as longitudinal passive monitoring devices in people with schizophrenia. It should be noted that the enrolled sample was relatively young, with a mean age of 30 years, and this group of people may be more accepting of digital devices. The digital data collection rate from the smartphones was impacted by both the compliance of participants and technical factors, such as proper device functioning and syncing with the mobile app. Some smartphone models automatically suppress all app activity during the night to save battery power, preventing complete data collection from the phone. In addition, some participants were provided with a phone for the study because their personal phone was incompatible with the HOPES app, and it was often the case that the participant's phone was too old. The above 2 points could explain why the data collection rate was higher with study-provided phones, as these (newer) phones could have exhibited better functioning.

Although there is a considerable amount of literature reporting associations between digital markers and clinical symptoms in populations of outpatients with schizophrenia, the majority collected digital data via only smartphones or actigraphy devices. Despite the expanding interest in using commercially available wearables in health care [57], studies using these devices in the medical field are still few. This study contributes such an investigation with a large clinical sample. Notably, the data reported in this study were not affected by Singapore's COVID-19 movement control measures: none of the 1-week digital data collection periods under study overlapped with the movement control period.

XSL•FO
RenderX

Limitations

There are some limitations to this study to be considered. The participants were relatively young (mean age: 30 years) and mildly ill (PANSS-total score: 51) [58] with a short duration of illness (2.6 years), which limits the generalizability of the results. Associations in older, chronic, and more severely ill patients might differ and should be further investigated. Due to the nature of this cross-sectional study, no causal relationships could be inferred. In addition, all participants included in this study were taking psychotropic medications with varying types and dosages, which may have had differential effects on cardiac function [59] and sleep [60], warranting further examination. Participants received financial incentives at timed intervals in the study, which might have positively affected adherence and data acquisition. Similar studies, however, noted that the majority of participants expressed a desire to continue using the wrist wearable without incentive after the study [19]. The longer-term use of wearable devices by individuals with schizophrenia should be further studied. Finally, because the regression analyses were exploratory in nature, and we aimed to detect all potential indicators of symptoms and functioning that could be useful in future studies, corrections for multiple testing were not applied before reporting significant associations.

Despite the limitations of the study, the patterns highlighted in the findings may guide future directions for research. The reported associations suggest that unobtrusively and passively collected digital phenotyping markers hold promise for continuous monitoring of clinical symptoms in people with schizophrenia between clinical visits. Digital phenotyping–enhanced care can improve the identification of patients in need of attention and may allow health care providers to delivery timely interventions [61].

Conclusion

Our study found notable patterns of associations between a range of digital biomarkers and clinically important outcome measures in a community sample of individuals with schizophrenia spectrum disorders. These results support the use of commercial wrist wearables and smartphones in continuous but passive monitoring of the health status of individuals with schizophrenia. Future studies should explore longer-term digital phenotyping observations in naturalistic settings and their utility in predicting clinical events.

Acknowledgments

The authors would like to express their sincere gratitude to the participants for their valuable contributions and to the clinicians who helped in study recruitment and assessments. We would like to thank Dr Gurpreet Rekhi, who trained and supervised the research psychologists in conducting the clinical assessments. The study was supported by the Ministry of Health Office for Healthcare Transformation and the Ministry of Health National Medical Research Council Centre Grant (NMRC/CG/M002/2017_IMH). JL is supported by the Ministry of Health National Medical Research Council Clinician Scientist Award (NMRC/CSAINV17nov005). The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available, because the participants in this study did not give written consent for their data to be shared publicly.

Authors' Contributions

JL, RJTM, ZY, and CH designed the study. RJTM and JL obtained the funding. ZY, NAAR, and TB managed the study. ZY, AS, NAAR, SZ, YFQ, SB, KWL, CT, SV, and JL supported recruitment and data collection. CH, ZY, and XW analyzed the data. CH, ZY, JL, and RJTM interpreted the results. ZY and CH drafted the manuscript. All authors revised and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection and feature processing in the Health Outcomes through Positive Engagement and Self-Empowerment (HOPES) platform.
[DOCX File, 19 KB - mhealth_v13i1e56185_app1.docx]

Multimedia Appendix 2
Sensor-specific successful data collection rates.
[DOCX File, 16 KB - mhealth_v13i1e56185_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Correlations between digital measures.
[DOCX File, 176 KB - mhealth_v13i1e56185_app3.docx]

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov;392(10159):1789-1858. [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Lee J, Remington G. Schizophrenia as a lifelong illness: implications for care. Ann Acad Med Singap 2015 Jan;44(1):26-28. [Medline: 25703494]
3. Bechi M, Bosia M, Spangaro M, et al. Exploring functioning in schizophrenia: predictors of functional capacity and real-world behaviour. Psychiatry Res 2017 May;251:118-124. [doi: 10.1016/j.psychres.2017.02.019] [Medline: 28199909]
4. Jääskeläinen E, Juola P, Hirvonen N, et al. A systematic review and meta-analysis of recovery in schizophrenia. Schizophr Bull 2013 Nov;39(6):1296-1306. [doi: 10.1093/schbul/sbs130] [Medline: 23172003]
5. Pennington M, McCrone P. The cost of relapse in schizophrenia. Pharmacoeconomics 2017 Sep;35(9):921-936. [doi: 10.1007/s40273-017-0515-3] [Medline: 28534254]
6. Cohen AS, Schwartz E, Le T, et al. Validating digital phenotyping technologies for clinical use: the critical importance of "resolution". World Psychiatry 2020 Feb;19(1):114-115. [doi: 10.1002/wps.20703] [Medline: 31922662]
7. Durand D, Strassnig M, Sabbag S, et al. Factors influencing self-assessment of cognition and functioning in schizophrenia: implications for treatment studies. Eur Neuropsychopharmacol 2015 Feb;25(2):185-191. [doi: 10.1016/j.euroneuro.2014.07.008] [Medline: 25104226]
8. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017 Oct 3;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
9. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 5;3(2):e16. [doi: 10.2196/mental.5165] [Medline: 27150677]
10. Torous J, Keshavan M. A new window into psychosis: The rise digital phenotyping, smartphone assessment, and mobile monitoring. Schizophr Res 2018 Jul;197:67-68. [doi: 10.1016/j.schres.2018.01.005] [Medline: 29338959]
11. Griffin B, Saunders KEA. Smartphones and wearables as a method for understanding symptom mechanisms. Front Psychiatry 2019;10:949. [doi: 10.3389/fpsyt.2019.00949] [Medline: 32009990]
12. Seppälä J, De Vita I, Jämsä T, et al. Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: systematic review. JMIR Ment Health 2019 Feb 20;6(2):e9819. [doi: 10.2196/mental.9819] [Medline: 30785404]
13. Beard C, Silverman AL, Forgeard M, Wilmer MT, Torous J, Björgvinsson T. Smartphone, social media, and mental health app use in an acute transdiagnostic psychiatric sample. JMIR Mhealth Uhealth 2019 Jun 7;7(6):e13364. [doi: 10.2196/13364] [Medline: 31199338]
14. Cella M, He Z, Killikelly C, Okruszek Ł, Lewis S, Wykes T. Blending active and passive digital technology methods to improve symptom monitoring in early psychosis. Early Interv Psychiatry 2019 Oct;13(5):1271-1275. [doi: 10.1111/eip.12796] [Medline: 30821079]
15. Meyer N, Kerz M, Folarin A, et al. Capturing rest-activity profiles in schizophrenia using wearable and mobile technologies: development, implementation, feasibility, and acceptability of a remote monitoring platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188. [doi: 10.2196/mhealth.8292] [Medline: 30377146]
16. Kerz M, Folarin A, Meyer N, Begale M, MacCabe J, Dobson RJ. SleepSight: a wearables-based relapse prevention system for schizophrenia. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct: Association for Computing Machinery; 2016:113-116. [doi: 10.1145/2968219.2971419]
17. Cella M, Okruszek Ł, Lawrence M, Zarlenga V, He Z, Wykes T. Using wearable technology to detect the autonomic signature of illness severity in schizophrenia. Schizophr Res 2018 May;195:537-542. [doi: 10.1016/j.schres.2017.09.028] [Medline: 28986005]
18. Shin S, Yeom CW, Shin C, et al. Activity monitoring using a mHealth device and correlations with psychopathology in patients with chronic schizophrenia. Psychiatry Res 2016 Dec 30;246:712-718. [doi: 10.1016/j.psychres.2016.10.059] [Medline: 27836243]
19. Staples P, Torous J, Barnett I, et al. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. NPJ Schizophr 2017 Oct 16;3(1):37. [doi: 10.1038/s41537-017-0038-0] [Medline: 29038553]
20. Raugh IM, James SH, Gonzalez CM, et al. Geolocation as a digital phenotyping measure of negative symptoms and functional outcome. Schizophr Bull 2020 Dec 1;46(6):1596-1607. [doi: 10.1093/schbul/sbaa121] [Medline: 32851401]
21. Strauss GP, Raugh IM, Zhang L, et al. Validation of accelerometry as a digital phenotyping measure of negative symptoms in schizophrenia. Schizophrenia (Heidelb) 2022 Apr 15;8(1):37. [doi: 10.1038/s41537-022-00241-z] [Medline: 35853890]

XSL•FO
RenderX

22. Kluge A, Kirschner M, Hager OM, et al. Combining actigraphy, ecological momentary assessment and neuroimaging to study apathy in patients with schizophrenia. Schizophr Res 2018 May;195:176-182. [doi: 10.1016/j.schres.2017.09.034] [Medline: 29030262]
23. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226. [doi: 10.1037/prj0000130] [Medline: 25844912]
24. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Association for Computing Machinery; 2014:3-14. [doi: 10.1145/2632048.2632054]
25. Abdul Rashid NA, Martanto W, Yang Z, et al. Evaluating the utility of digital phenotyping to predict health outcomes in schizophrenia: protocol for the HOPE-S observational study. BMJ Open 2021 Oct 20;11(10):e046552. [doi: 10.1136/bmjopen-2020-046552] [Medline: 34670760]
26. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophr Bull 1987 Jan 1;13(2):261-276. [doi: 10.1093/schbul/13.2.261]
27. Lim K, Peh OH, Yang Z, et al. Large-scale evaluation of the Positive and Negative Syndrome Scale (PANSS) symptom architecture in schizophrenia. Asian J Psychiatr 2021 Aug;62:102732. [doi: 10.1016/j.ajp.2021.102732] [Medline: 34118560]
28. Kirkpatrick B, Strauss GP, Nguyen L, et al. The Brief Negative Symptom Scale: psychometric properties. Schizophr Bull 2011 Mar;37(2):300-305. [doi: 10.1093/schbul/sbq059] [Medline: 20558531]
29. Ang MS, Rekhi G, Lee J. Validation of the Brief Negative Symptom Scale and its association with functioning. Schizophr Res 2019 Jun;208:97-104. [doi: 10.1016/j.schres.2019.04.005] [Medline: 30987926]
30. Addington D, Addington J, Maticka-Tyndale E, Joyce J. Reliability and validity of a depression rating scale for schizophrenics. Schizophr Res 1992 Mar;6(3):201-208. [doi: 10.1016/0920-9964(92)90003-n] [Medline: 1571313]
31. Keefe RSE, Goldberg TE, Harvey PD, Gold JM, Poe MP, Coughenour L. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. Schizophr Res 2004 Jun 1;68(2-3):283-297. [doi: 10.1016/j.schres.2003.09.011] [Medline: 15099610]
32. Rybarczyk B. Social and Occupational Functioning Assessment Scale (SOFAS). In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology: Springer; 2011:2313. [doi: 10.1007/978-0-387-79948-3_428]
33. Wang X, Vouk N, Heaukulani C, et al. HOPES: an integrative digital phenotyping platform for data collection, monitoring, and machine learning. J Med Internet Res 2021 Mar 15;23(3):e23984. [doi: 10.2196/23984] [Medline: 33720028]
34. Alphs L, Nasrallah HA, Bossie CA, et al. Factors associated with relapse in schizophrenia despite adherence to long-acting injectable antipsychotic therapy. Int Clin Psychopharmacol 2016 Jul;31(4):202-209. [doi: 10.1097/YIC.0000000000000125] [Medline: 26974214]
35. Jeste DV, Maglione JE. Treating older adults with schizophrenia: challenges and opportunities. Schizophr Bull 2013 Sep;39(5):966-968. [doi: 10.1093/schbul/sbt043] [Medline: 23552180]
36. de Crom SAM, Haan LD, Schirmbeck F, GROUP investigators. The association between sleep disturbances and negative symptom severity in patients with non-affective psychotic disorders, unaffected siblings and healthy controls. Psychiatry Res 2021 Mar;297:113728. [doi: 10.1016/j.psychres.2021.113728] [Medline: 33493731]
37. Blanchard JJ, Andrea A, Orth RD, Savage C, Bennett ME. Sleep disturbance and sleep-related impairment in psychotic disorders are related to both positive and negative symptoms. Psychiatry Res 2020 Feb 7;286:112857. [doi: 10.1016/j.psychres.2020.112857] [Medline: 32087449]
38. Docx L, Sabbe B, Provinciael P, Merckx N, Morrens M. Quantitative psychomotor dysfunction in schizophrenia: a loss of drive, impaired movement execution or both? Neuropsychobiology 2013;68(4):221-227. [doi: 10.1159/000355293] [Medline: 24247207]
39. Chung KF, Poon Y, Ng TK, Kan CK. Correlates of sleep irregularity in schizophrenia. Psychiatry Res 2018 Dec;270:705-714. [doi: 10.1016/j.psychres.2018.10.064] [Medline: 30551313]
40. Poulin J, Daoust AM, Forest G, Stip E, Godbout R. Sleep architecture and its clinical correlates in first episode and neuroleptic-naive patients with schizophrenia. Schizophr Res 2003 Jul 1;62(1-2):147-153. [doi: 10.1016/s0920-9964(02)00346-8] [Medline: 12765755]
41. Chan MS, Chung KF, Yung KP, Yeung WF. Sleep in schizophrenia: a systematic review and meta-analysis of polysomnographic findings in case-control studies. Sleep Med Rev 2017 Apr;32:69-84. [doi: 10.1016/j.smrv.2016.03.001] [Medline: 27061476]
42. Abdul Rashid NA, Nurjono M, Lee J. Clinical determinants of physical activity and sedentary behaviour in individuals with schizophrenia. Asian J Psychiatr 2019 Dec;46:62-67. [doi: 10.1016/j.ajp.2019.10.004] [Medline: 31627166]
43. Wichniak A, Skowerska A, Chojnacka-Wójtowicz J, et al. Actigraphic monitoring of activity and rest in schizophrenic patients treated with olanzapine or risperidone. J Psychiatr Res 2011 Oct;45(10):1381-1386. [doi: 10.1016/j.jpsychires.2011.05.009] [Medline: 21679968]

XSL•FO
RenderX

44. Walther S, Ramseyer F, Horn H, Strik W, Tschacher W. Less structured movement patterns predict severity of positive syndrome, excitement, and disorganization. Schizophr Bull 2014 May;40(3):585-591. [doi: 10.1093/schbul/sbt038] [Medline: 23502433]
45. Depp CA, Bashem J, Moore RC, et al. GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. NPJ Digit Med 2019;2:108. [doi: 10.1038/s41746-019-0182-1] [Medline: 31728415]
46. He-Yueya J, Buck B, Campbell A, et al. Assessing the relationship between routine and schizophrenia symptoms with passively sensed measures of behavioral stability. NPJ Schizophr 2020 Nov 23;6(1):35. [doi: 10.1038/s41537-020-00123-2] [Medline: 33230099]
47. Buck B, Scherer E, Brian R, et al. Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. Schizophr Res 2019 Jun;208:167-172. [doi: 10.1016/j.schres.2019.03.014] [Medline: 30940400]
48. Mastoras RE, Iakovakis D, Hadjidimitriou S, et al. Touchscreen typing pattern analysis for remote detection of the depressive tendency. Sci Rep 2019 Sep 16;9(1):13414. [doi: 10.1038/s41598-019-50002-9] [Medline: 31527640]
49. Ross MK, Demos AP, Zulueta J, et al. Naturalistic smartphone keyboard typing reflects processing speed and executive function. Brain Behav 2021 Nov;11(11):e2363. [doi: 10.1002/brb3.2363] [Medline: 34612605]
50. Kapsecker M, Osterlehner S, Jonas SM. Analysis of mobile typing characteristics in the light of cognition. In: 2022 IEEE International Conference on Digital Health (ICDH): IEEE; 2022. [doi: 10.1109/ICDH55609.2022.00022]
51. Bär KJ, Wernich K, Boettger S, et al. Relationship between cardiovagal modulation and psychotic state in patients with paranoid schizophrenia. Psychiatry Res 2008 Jan 15;157(1-3):255-257. [doi: 10.1016/j.psychres.2007.04.021] [Medline: 17919737]
52. Chang HA, Chang CC, Tzeng NS, Kuo TBJ, Lu RB, Huang SY. Cardiac autonomic dysregulation in acute schizophrenia. Acta Neuropsychiatr 2013 Jun;25(3):155-164. [doi: 10.1111/acn.12014] [Medline: 25287469]
53. Chung MS, Yang AC, Lin YC, et al. Association of altered cardiac autonomic function with psychopathology and metabolic profiles in schizophrenia. Psychiatry Res 2013 Dec 30;210(3):710-715. [doi: 10.1016/j.psychres.2013.07.034] [Medline: 23978730]
54. Kalisperakis E, Karantinos T, Lazaridi M, et al. Smartwatch digital phenotypes predict positive and negative symptom variation in a longitudinal monitoring study of patients with psychotic disorders. Front Psychiatry 2023;14:1024965. [doi: 10.3389/fpsyt.2023.1024965] [Medline: 36993926]
55. Stogios N, Gdanski A, Gerretsen P, et al. Autonomic nervous system dysfunction in schizophrenia: impact on cognitive and metabolic health. NPJ Schizophr 2021 Apr 26;7(1):22. [doi: 10.1038/s41537-021-00151-6] [Medline: 33903594]
56. Haigh SM, Walford TP, Brosseau P. Heart rate variability in schizophrenia and autism. Front Psychiatry 2021;12:760396. [doi: 10.3389/fpsyt.2021.760396] [Medline: 34899423]
57. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45. [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
58. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? Schizophr Res 2005 Nov 15;79(2-3):231-238. [doi: 10.1016/j.schres.2005.04.008] [Medline: 15982856]
59. Huhn M, Arndt T, Schneider-Thoma J, Leucht S. Effects of antipsychotics on heart rate in treatment of schizophrenia: a systematic review and meta-analysis. Ther Adv Psychopharmacol 2022;12:20451253221097261. [doi: 10.1177/20451253221097261] [Medline: 35774251]
60. Cohrs S. Sleep disturbances in patients with schizophrenia: impact and effect of antipsychotics. CNS Drugs 2008;22(11):939-962. [doi: 10.2165/00023210-200822110-00004] [Medline: 18840034]
61. Stafford MR, Jackson H, Mayo-Wilson E, Morrison AP, Kendall T. Early interventions to prevent psychosis: systematic review and meta-analysis. BMJ 2013 Jan 18;346:f185. [doi: 10.1136/bmj.f185] [Medline: 23335473]

Abbreviations

BACS: Brief Assessment of Cognition in Schizophrenia
BNSS: Brief Negative Symptom Scale
BPRS: Brief Psychiatric Rating Scale
CDSS: Calgary Depression Scale for Schizophrenia
HOPE-S: Health Outcomes via Positive Engagement in Schizophrenia
HOPES: Health Outcomes through Positive Engagement and Self-Empowerment
PANSS: Positive and Negative Syndrome Scale
REM: rapid eye movement
SOFAS: Social and Occupational Functioning Assessment Scale

XSL•FO
RenderX

Edited by L Buis; submitted 09.01.24; peer-reviewed by L Yang, M Straczkiewicz, P Harvey; revised version received 25.09.24; accepted 02.12.24; published 05.02.25.

Please cite as:
Yang Z, Heaukulani C, Sim A, Buddhika T, Abdul Rashid NA, Wang X, Zheng S, Quek YF, Basu S, Lee KW, Tang C, Verma S, Morris RJT, Lee J
Utility of Digital Phenotyping Based on Wrist Wearables and Smartphones in Psychosis: Observational Study
JMIR Mhealth Uhealth 2025;13:e56185
URL: https://mhealth.jmir.org/2025/1/e56185
doi:10.2196/56185

© Zixu Yang, Creighton Heaukulani, Amelia Sim, Thisum Buddhika, Nur Amirah Abdul Rashid, Xuancong Wang, Shushan Zheng, Yue Feng Quek, Sutapa Basu, Kok Wei Lee, Charmaine Tang, Swapna Verma, Robert J T Morris, Jimmy Lee. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 5.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Measuring Psychological Well-Being and Behaviors Using Smartphone-Based Digital Phenotyping: An Intensive Longitudinal Observational mHealth Pilot Study Embedded in a Prospective Cohort of Women

Li Yi[1,2], PhD; Claudia Trudel-Fitzgerald[3,4,5], PhD; Cindy R Hu[6], PhD; Grete Wilt[6], PhD; Jorge Chavarro[2,7], MD, ScD; Jukka-Pekka Onnela[8], PhD; Francine Grodstein[9], ScD; Laura D Kubzansky[10], PhD; Peter James[1,6,11], ScD

[1]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 401 Park Drive, Suite 401 East, Boston, MA, United States

[10]Department of Social and Behavioral Sciences, Harvard TH Chan School of Public Health, Boston, MA, United States

[11]Department of Public Health Sciences, University of California, Davis School of Medicine, Medical Sciences 1-C, One Shields Ave, Davis, CA, United States

[2]Department of Nutrition, Harvard TH Chan School of Public Health, Boston, MA, United States

[3]Département of Psychology, Université du Québec À Trois-Rivières, Trois-Rivières, QC, Canada

[4]Research Center, Universitaire en Santé Mentale de Montréal, Montréal, QC, Canada

[5]Lee Kum Sheung Center for Health and Happiness, Harvard University, Boston, MA, United States

[6]Department of Environmental Health, Harvard TH Chan School of Public Health, Boston, MA, United States

[7]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

[8]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA, United States

[9]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, United States

Corresponding Author:
Peter James, ScD
Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 401 Park Drive, Suite 401 East, Boston, MA, United States

Abstract

Background: Intensive measures of well-being and behaviors in large epidemiologic cohorts have the potential to enhance health research in these areas. Yet, little is known regarding the feasibility of using mobile technology to collect intensive data in the “natural” environment in the context of ongoing large cohort studies.

Objective: We examined the feasibility of using smartphone digital phenotyping to collect highly resolved psychological and behavioral data from participants in a pilot study with participants in Nurses’ Health Study II, a nationwide prospective cohort of women.

Methods: In this pilot study, an 8-day intensive smartphone protocol was implemented using the “Beiwe” smartphone app. Participants (n=181) completed a baseline survey on day 1 and answered ecological momentary assessment (EMA) surveys twice daily (days 2-8; early afternoon and evening) using their smartphone and provided minute-level accelerometer and GPS data. A feedback survey at the end of Substudy queried participants’ experience with the app and data collection process. We assessed adherence to the protocol by examining completion on EMA surveys and completeness of accelerometer and GPS data at the participant, participant day, and prompt levels.

Results: Our pilot study demonstrated modest overall compliance with smartphone-based surveys: the baseline survey completion rate was high (156/181, 86.2%), but average daily EMA response rates during the 7-day period were lower with 55.6% (SD 3.9%) for early afternoon and 54.7% (SD 3.2%) for evening. We also observed good average daily completeness of smartphone accelerometer (mean 62.0%, SD 4.5%) data and GPS data (mean 57.7%, SD 3.1%). The feedback survey revealed that the participants found “the app easy to use” (median 85.0 on a scale of 1 - 100) and were “willing to repeat similar studies” (median 85.0 on a scale of 1 - 100). Although participants reported feeling their participation was a positive experience (median 64.0 on a scale of 1 - 100), they also identified some important issues, including user fatigue due to repetitive daily surveys.

Conclusions: We observed modest compliance with smartphone surveys and completeness of smartphone passive sensing data in this pilot study compared with similar studies in the past. However, this was not unexpected, given our participants were older (aged 57 - 75 years, with more than 3 decades of follow-up at the time of the substudy) and may encounter more technological

XSL•FO
RenderX

barriers, not to mention that the indication of willingness to participate in such studies again was fairly high. Our findings also highlight that the success and data quality of efforts to obtain daily measures may vary depending on data type and emphasize the need to improve the design of the EMA survey to improve or sustain participant engagement over the study period. Overall, our findings suggest smartphone-based digital phenotyping as a promising technology when embedding in large epidemiological cohorts to collect intensive longitudinal observation data.

(JMIR Mhealth Uhealth 2025;13:e71375) doi:10.2196/71375

KEYWORDS

daily mobility; digital phenotyping; ecological momentary assessments; epidemiological cohort; health behavior; psychological well-being; happiness; smartphone apps and sensors; mobile phone

Introduction

Background

Many dimensions of psychological well-being—both stable traits (eg, optimism) and time-varying states (eg, happiness)—are linked to positive physical health outcomes such as a lower risk of chronic disease and mortality [1-3]. Health behaviors, including diet, physical activity, and sleep, are risk factors for many chronic diseases and may also be influenced by psychological well-being, thus acting as pathways by which psychological well-being affects physical health [4-9]. However, most studies have collected psychological and behavioral data annually or at longer intervals, limiting the ability to examine dynamic relationships between the two, which may be critical for elucidating the mechanisms linking psychological well-being to chronic disease. To better understand the relationships between psychological well-being and health behaviors, it is necessary to collect granular data on both factors on a daily or diurnal basis [9,10].

To address this limitation, emerging research has focused on the use of smartphone-based digital phenotyping, defined as the moment-by-moment quantification of individual human phenotypes via personal smartphone data, to collect temporally intensive longitudinal (eg, minute-by-minute) data [11]. In particular, ecological momentary assessment (EMA) can be combined with accelerometer-enabled smartphones to measure participants' behaviors and experiences repeatedly in real time. This approach facilitates efficient and cost-effective digital phenotyping of psychological well-being and specific health behaviors in a naturalistic setting [12-16]. Consequently, EMA conducted with personal smartphones may facilitate large-scale data collection to study psychological well-being and behavioral health with finer temporal resolution than is possible with standard survey research. Recruiting samples for such research can be difficult and time-consuming, whereas drawing on existing cohorts with participants who have a history of compliance with and commitment to the overall study may provide a more efficient means for conducting such studies. However, embedding smartphone digital phenotyping in large epidemiological cohorts has not yet been sufficiently explored [17], nor has the acceptability and effectiveness of this strategy among populations that have generally agreed to participate in survey research been tested.

Objectives

This study describes the methodology for integrating the Beiwe smartphone digital phenotyping platform into the Nurses' Health Study II (NHSII), a large and long-running US-based cohort of middle-aged and older women (aged 55‐77 years). We collected data on psychological well-being, physical activity, and geolocation from 181 NHSII participants during an 8-day intensive measurement burst that included daily EMA surveys and minute-level accelerometer and GPS sensing. Our aim was to assess the feasibility of using smartphone digital phenotyping to collect highly resolved psychological, behavioral, and environmental data by examining compliance with the survey schedule, completeness of passive sensing data, and participant responses to the end-of-study feedback survey.

Methods

Beiwe Smartphone Substudy Embedded Within the NHSII

NHSII is an ongoing prospective cohort study involving 116,429 female registered nurses aged 25‐42 years at enrollment in 1989. The participants currently reside in all states of the United States. Participants completed a mailed questionnaire asking detailed questions about demographic characteristics, lifestyles, and history of chronic diseases at the time of enrollment. Participants are invited to complete biennial questionnaires to update their medical status and lifestyle factors [18]. The Beiwe Smartphone Substudy of NHSII is a pilot study designed to examine the feasibility of using smartphone apps to collect psychological well-being, physical activity, and geolocation data at a granular level (eg, daily and minute-by-minute) among participants in this ongoing cohort study. The substudy targeted those who have most frequently engaged with the larger NHSII through biennial survey completion and provision of biological specimens [12]. For this pilot study, participants took part in an 8-day intensive measurement burst via the Beiwe app, completing psychological well-being surveys twice daily while the app collected smartphone accelerometer and GPS data at the minute level. At the beginning of the study, the participants were instructed to download the Beiwe app on their smartphones and log on via provided credentials. During the 8-day data collection period, participants were asked to keep their smartphone on them during waking hours, charge it regularly, and connect to Wi-Fi at least once a week to synchronize their data. Trained research assistants provided technical support throughout the study by phone or email. Upon completion of the study, participants were instructed to delete the app, and

XSL•FO
RenderX

their user accounts were deactivated. The pilot study ran from July 21, 2021, to November 21, 2021 (the last day the last participant provided data within 8-day period).

Our goal was to recruit 200 participants for the substudy. To be eligible for an invitation to the pilot study, participants must have been active NHSII participants and not participating in other NHSII substudies. From this pool, we randomly selected 600 individuals to whom we emailed invitations to complete an eligibility screener. The screener included the following additional eligibility criteria: (1) owns an Apple or Android smartphone, (2) lives in the contiguous United States (excluding Hawaii, Alaska, or Puerto Rico), and (3) has access to Wi-Fi at least once a week. Upon confirmation of eligibility, there was an informed consent process; after consenting, they were instructed to download the Beiwe app from the iOS or Android app stores.

Data Collection—Beiwe App

The Beiwe app guided participants through onboarding, during which they gave their consent for the app to assess location data and to send notifications (see Multimedia Appendix 1, eg, screenshots of the app as viewed by the participant). Upon first opening the app (day 1), participants completed a baseline survey and then received 14 EMA surveys over the next 7 days (days 2-8), with 2 surveys administered daily during the same time windows each day. The shorter early afternoon survey (21 questions, including some branched yes or no questions) was sent out randomly between 12 PM and 3 PM; the longer evening survey (63 questions, including some branched yes or no questions) was sent out at 6 PM. The app notified participants when a survey was available for completion and marked it as incomplete if it was not completed before the next survey was sent. For each completed survey, the app recorded the start and end time and computed the duration. Meanwhile, the Beiwe app was programmed to collect raw GPS data every 15 minutes for 90 consecutive seconds and raw accelerometer data at 10 Hz every 30 seconds for 10 consecutive seconds. This approach was intended to minimize battery consumption while maintaining the ability to determine participant movement patterns and behavior throughout the day [19]. Technical support was available throughout the duration of the substudy for participants who encountered issues downloading or using the app. At the conclusion of our study, we downloaded the study data via the "Mano" algorithm in the "Forest" Python library (Python, version 3.9.2; Python Software Foundation).

Smartphone-Based Measures

We assessed our key variables of interest and others that could be relevant when studying linkages between psychological well-being and behavior. We used a standard set of self-report items included on the baseline and EMA surveys administered twice daily (Multimedia Appendix 2), as well as passive sensing of smartphone accelerometer and GPS data.

Psychological Well-Being

In the baseline survey, we assessed *dispositional optimism*, a stable attribute positively associated with higher psychological well-being using the validated Life Orientation Test—Revised [20]. Participants rated their agreement with 3 positive and 3 negative statements on a 5-point Likert scale ranging from "Strongly disagree" (0) to "Strongly agree" (4). The positive statements included the following: (1) In uncertain times, I usually expect the best. (2) I'm always optimistic about my future. (3) Overall, I expect more good things to happen to me than bad. The negative statements included the following: (1) If something can go wrong for me, it will. (2) I hardly ever expect things to go my way. (3) I rarely count on good things happening to me. An overall optimism score was calculated by summing the positive and reverse scores of the negative statements [20].

We assessed time-varying psychological states contemporaneously (in the early afternoon EMA survey) and retrospectively (in the evening EMA survey) using the modified version of the Positive and Negative Affect Schedule—Expanded Form (PANAS-X) [21]. The 10 PANAS-X items designed to capture momentary affect included in the early afternoon EMA surveys were queried using the stem "Right now, I feel…." with response options yes or no scored as 1 or 0, respectively. Four items (happy, determined, enthusiastic, and curious) were summed to derive a *positive affect* scale, and 6 items (sad, afraid, angry, lonely, nervous, and worried) were summed to derive a *negative affect* scale.

A modified PANAS-X designed to capture affect experienced over the course of the day was administered in evening EMA survey and included 11 items queried using the stem "Today; did you feel…." with response options yes (score = 1) or no (score = 0). Five items are summed to derive a scale capturing *positive affect in the day* (happy, determined, enthusiastic, curious, and grateful) and 6 items are summed to derive a scale capturing *negative affect in the day* (sad, afraid, angry, lonely, nervous, and worried). In addition, participants were asked to report when they experienced these feelings (midnight: 5:59 AM; 6-11:59 AM; noon: 5:59 PM; 6-11:59 PM). The evening EMA also included a question on *life satisfaction*: "How satisfied are you with your life at this moment?" (queried on a scale of 1 [very unsatisfied] to 5 [very satisfied]) [22].

Health Behaviors

We obtained participants' *physical activity* (ie, walking steps and walking time) by processing raw accelerometer data with the "Oak" walking recognition method in the "Forest" Python library (Python, version 3.9.2), which estimates gait cadence per second. We have focused on walking because it is the most common form of physical activity performed daily by middle-aged and older-aged women and has been associated with cognitive function and disease outcomes such as cardiovascular disease and cancer [23,24]. Our walking recognition method has been validated against 20 publicly available annotated datasets on walking activity data collected at various body locations (thigh, waist, chest, arm, and wrist). It estimates walking periods with high sensitivity and specificity and is able to account for missing data due to noncontinuous sampling to conserve smartphone battery [25]. Due to the large volume of raw accelerometer data, we implemented an "on-the-cloud" processing approach. Specifically, we fetched hourly level raw data, processed them in temporary cloud computing memory by applying the "Oak" algorithm, and stored

XSL•FO
RenderX

the processed results (minute-level step counts) locally without downloading raw data. This approach substantially reduced the processing time. Details of the validated walking detection algorithm for Beiwe are available in a previous study [25] and through the GitHub link [26]. After identifying the walking periods, we aggregated the second-level data to the minute level, calculating the total number of seconds walked and step counts per minute.

Participants also reported their previous night's sleep and wake times in the evening EMA survey, which we used to determine their *daily sleep duration*. Additionally, they reported their *sleep quality* during the previous night. Multimedia Appendix 2 lists the answer choices for the sleep survey.

Additional Measures

We assessed a number of key covariates relevant to the association between psychological well-being and health behaviors. First, since psychological distress can influence behaviors [13,27] and is related to facets of psychological well-being, we measured depressive symptoms using the Center for Epidemiologic Studies Depression Scale [28] in the baseline survey. Second, we included items to assess strategies participants may use to regulate their affective states (ie, emotion regulation and coping strategies) in the context of stressful events or positive experiences since these may also be related to behaviors [29-32]. These were included on both daily EMA surveys. In the early afternoon EMA survey, we queried the use of strategies to regulate emotions in the context of stressful events occurring in the past hour [33]. Items assessed the nature of the stressful event (eg, argument, financial issues, and traffic) and stress levels ("not at all" or "a little" or "somewhat" or "very"). Participants then rated their levels of agreement on a 5-point Likert scale (strongly disagree to strongly agree) with a list of strategies they might have used to cope with such events (eg, "I controlled my emotions by not showing them" and "I made a plan to make the situation better"). In the evening EMA survey, we presented participants with a list of stressful events and positive experiences and asked whether any had occurred in the past 24 hours (yes or no) and also queried levels of stressful or positive experience on a 4-point Likert scale of 1 (not at all), 2 (a little), 3 (somewhat), and 4 (very) in relation to each event (Multimedia Appendix 2). Additionally, in the evening survey, participants reported which positive and negative emotion they experienced most intensely that day (eg, happy and sad). They were then asked to indicate on a 5-point Likert scale (strongly disagree to strongly agree; Multimedia Appendix 2) the extent to which they used each strategy to cope with such events from a list of strategies presented (eg, "I reveled in the moment and concentrated on how good I felt" and "I made a plan to make the situation better") [33]. They were also asked whether they changed their diet in response to the indicated emotion (ie, increased/decreased/did not change how much I ate) and whether this change was a way to comfort themselves with response options of yes or no. Finally, we inquired about psychotropic medication use in our evening EMA survey, asking whether participants had taken any of the following medications in the prior day: antidepressant medication (for any reason), antianxiety medication, and sleep medication.

Considering that individual daily mobility and environmental contexts may also be important [34], we obtained participants' geolocation every minute by processing raw smartphone GPS data using a Python-based method called "Jasmine" in the "Forest" library (Python, version 3.9.2). This algorithm imputes minute-level geolocation and also calculates daily summary statistics measuring participants' mobility, such as time spent at home and on trips and the number of places visited. Details of the validated GPS data-processing algorithm for Beiwe are available in a previous study [35] and through the GitHub link [26].

End-of-Study Feedback Survey

We invited participants to complete a feedback survey at the conclusion of the study to evaluate their overall experience in participating in the pilot study and their willingness to engage in any future smartphone-based data collection. This survey included items querying their overall experience, any difficulties using the app, and their likelihood of participating in similar future studies and offered the opportunity to provide open-ended feedback on app design and study experience. Feedback survey questions are shown in Multimedia Appendix 2.

Sociodemographic, Lifestyle, and Health Characteristics

To evaluate potential differences between study participants and those in the parent study who did not participate, we also extracted key sociodemographic and health characteristics from the baseline NHSII surveys. These included factors that are not time-varying (race) and time-varying ones from the 2019‐2021 NHSII biennial questionnaire, closest in time to the start of the pilot study in August 2021. These include participants' age at the time of the substudy, race (White/Black/Other; non–time-varying), husband's level of educational attainment (high school or less/more than high school, reported in 1999), percentages married or in a relationship, smoking status (current/former/never), BMI (calculated as weight in kilograms divided by height in meters squared), Alternative Healthy Eating Index (AHEI) diet score (continuous scores 0‐100) [36], and region of residence (Northeast/Midwest/South/West) based on 2010 Census data [37]. The AHEI assigns ratings to foods and nutrients predictive of chronic disease [36]. The AHEI is created by a group of researchers at Harvard TH Chan School of Public Health as an alternative to the US Department of Agriculture's Healthy Eating Index, which measures adherence to the federal Dietary Guidelines for Americans.

Statistical Analysis

Our examination of adherence to the smartphone digital phenotyping data collection protocol had 2 components. For the smartphone EMA surveys, we assessed compliance at the time of receiving each EMA prompt at participant day and participant levels. At the *participant level*, we calculated separate response rates for the baseline survey (day 1) and for the daily early afternoon and evening EMA surveys (days 2-8 only), each defined as the total number of surveys submitted by all participants out of all possible EMA prompts. We also calculated the proportions of participants with 25% (at least 2 out of 7 surveys), 50% (at least 4 out of 7 surveys), and 75% (at least 6 out of 7 surveys) response rates for the early afternoon

XSL•FO
RenderX

and evening EMA surveys separately. At the *participant day* level, we examined the daily response rates for early afternoon and evening EMA surveys separately from days 2 to 8 (defined as the total number of surveys submitted by all participants in days 2, 3, …, 8 divided by total number of participants). In addition, we summarized the length of time it took to complete a given survey (ie, completion duration) in minutes for the baseline survey and for each daily EMA assessment.

To assess the completeness of passive sensing data (accelerometer and GPS), we identified days with high-quality data (ie, valid days) as those with at least 600 valid minutes (10 valid hours) per day, following the established criteria in GPS-based physical activity studies [34]. We also define noncollection days as days without data. In this study, the primary reason for noncollection was attrition (ie, the app was deleted from a participant's phone before the 7-day study period because the participant herself unenrolled from the substudy). Another reason was sensor noncollection, which could be due to a participant forgetting to charge their phone or disabling the GPS or a major update from the operating system causing the Beiwe app to malfunction temporarily, among other reasons. A valid minute was defined as one with at least 1 second of observation time, following guidelines provided by previous Beiwe data studies [25]. We then calculated data completeness at both the *participant day* and *participant* levels. At the *participant day* level, we summarized the valid hours per day for all participants and days and calculated the percentages of valid, nonvalid, and noncollection days, respectively. At the *participant* level, data completeness was defined as the percentage of valid days out of 7 possible days (days 2-8), with proportions calculated for participants with 25% (minimum of 2 valid days), 50% (minimum of 4 valid days), and 75% (minimum of 6 valid days) completeness of accelerometer and GPS data. To test the sensitivity of this measure of completeness to an alternative threshold, we applied a stricter criterion defining a valid day as one with at least 960 valid minutes (16 valid hours), accounting for typical 8-hour sleep periods. Some participants did not delete the app at the conclusion of the study (ie, after day 8) and left the app running for up to 150 days, resulting in the collection of additional EMA surveys and passive sensing data. Their data beyond the study period were not included in the "Smartphone Data Characteristics" section. All statistical analyses were conducted using R (version 4.3.3; R Core Team).

Ethical Considerations

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional review boards of the Brigham and Women's Hospital (protocol numbers 1999P003389). All participants provided informed consent for data collection at the beginning of the pilot study. Participants were not compensated. We encrypted raw data collected by the Beiwe app before sending them to the Amazon Web Services production server. Subsequently, we either securely transferred data to the analytical server (for GPS and survey data) or deleted those data from the temporary storage on the analytics server after processing (for accelerometer data). We protected Amazon Web Services production and analytics servers by 2-factor authentication and Secure Shell Protocol keys. To protect participant identity, we did not enter personal or health-related information through Beiwe.

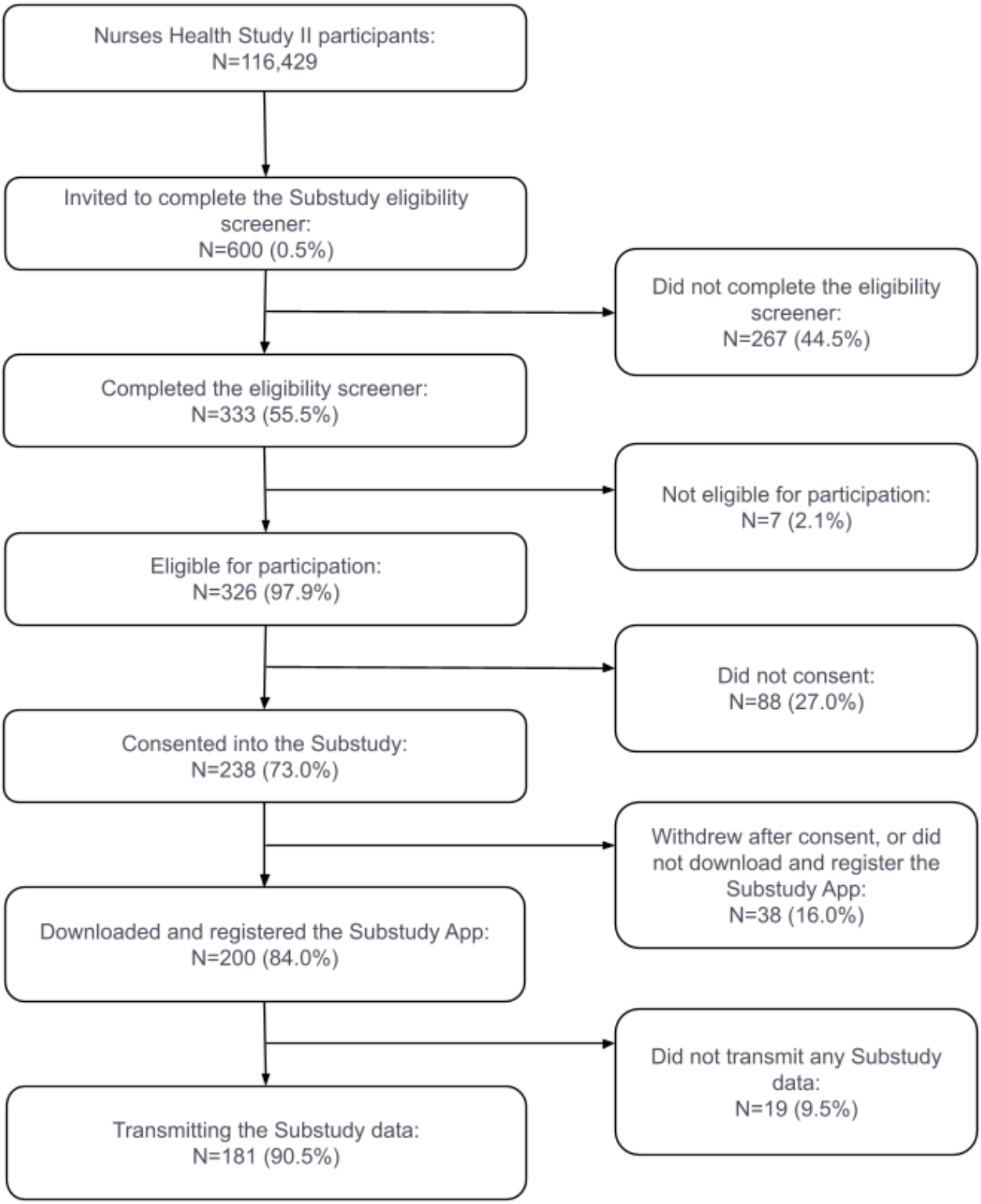
Results

Participant Characteristics

We invited 600 NHSII participants (600/116,429, 0.52% of the larger NHSII cohort) to complete the eligibility screener. Of these 600 participants, 333 (55.5%) completed the screener, 326 (54.3%) were eligible to participate, 238 (39.7%) consented, 200 (33.3%) downloaded and registered the app, and 181 (30.2%) transmitted data. A CONSORT (Consolidated Standards of Reporting Trials) diagram of the participants at different recruitment stages is shown in Figure 1. Table 1 compares the demographics of the Beiwe pilot study participants who transmitted data against those who consented, who were invited, and the active NHSII participants at the beginning of the pilot study (late July 2021). Overall, our sample had a mean age of 67.8 (SD 3.9) years, were all female, and predominantly White (178/181, 98.3%). Most participants (148/181, 81.8%) were married and 58.6% (106/181) were never smokers. Breaking down by Census regions, 11.6% (21/181) of the participants lived in the Northeast, 17.7% (32/181) lived in the West, 26.0% (47/181) lived in the South, and 44.8% (81/181) lived in the Midwest. Most participants (84.5%, 153/181) were iOS users. Across the variables we considered, there were no major differences among the invited, consented, and enrolled participants. Additionally, no major demographic differences were observed between iOS and Android users (Multimedia Appendix 3).

XSL•FO
RenderX

Figure 1. A CONSORT (Consolidated Standards of Reporting Trials) diagram of participation numbers at different recruitment stages of the Beiwe Smartphone Substudy of Nurses' Health Study II.



XSL•FO
RenderX

Table . Characteristics of active Nurses' Health Study II participants at the beginning of the Beiwe smartphone pilot study (August 2021) who were invited, consented, and participated in the study.

Variable	Beiwe Substudy participants—invited (n=600)	Beiwe Substudy participants—consented (n=238)	Beiwe Substudy participants—registered and downloaded the app (n=181)
Age (years), mean (SD)	68.2 (4.1)	68.0 (4.0)	67.8 (3.9)
Race, n (%)			
White	590 (98.3)	235 (98.7)	178 (98.3)
Black	0 (0.0)	0 (0.0)	0 (0.0)
Other/more than 1 race	10 (1.7)	3 (1.3)	3 (1.7)
Husband's highest level of education, n (%)			
High school or less	160 (26.7)	70 (29.4)	51 (28.2)
More than high school	440 (73.3)	168 (70.6)	130 (71.8)
Marital or in a relationship, n (%)			
Married or in a relationship	491 (81.8)	190 (79.8)	148 (81.8)
Smoking status, n (%)			
Current	13 (2.2)	5 (2.1)	5 (2.8)
Former	195 (32.5)	91 (38.2)	70 (38.7)
Never	392 (65.3)	142 (59.7)	106 (58.6)
BMI (kg/m^2), mean (SD)	26.8 (6.2)	27.0 (6.0)	27.0 (6.4)
AHEI score[a], mean (SD)	69.0 (12.7)	70.2 (12.7)	70.0 (12.4)
Region of residence, n (%)			
Northeast	55 (9.2)	29 (12.2)	21 (11.6)
Midwest	259 (43.2)	103 (43.3)	81 (44.8)
South	175 (29.2)	66 (27.7)	47 (26.0)
West	111 (18.5)	40 (16.8)	32 (17.7)

[a]AHEI: Alternative Health Eating Index (ranging from 0 to 110, where a higher score reflects a healthier diet). The AHEI assigns ratings to foods and nutrients predictive of chronic disease (McCullough and Willett [36]). AHEI is created by a group of researchers at Harvard TH Chan School of Public Health as an alternative to the US Department of Agriculture's Healthy Eating Index, which measures adherence to the federal Dietary Guidelines for Americans.

Smartphone Data Characteristics

Data collection was completed in November 2021. In total, we received 40.0 GB of data from 181 participants, including 19.2 GB of accelerometer data, 0.6 GB of GPS data, and 0.01 GB of survey data. Each participant contributed data (at least 1 byte of data within the enrollment period) for an average of 7 out of 8 possible days (SD 2; median 8).

Smartphone-Based Survey Compliance

Each afternoon and evening assessment was sent once a day for 181 participants for 7 days. A total of 181 participants completed 156 baseline surveys (86.2% response) at day 1. From days 2 - 8, they responded to 705 early afternoon EMA prompts out of 1267 prompts sent (181 participants for 7 days; 55.6% response across all participant-days) and 693 evening EMA prompts out of 1267 prompts sent (181 participants for 7 days; 54.7% response across all participant-days). On average, the shorter surveys given with the early afternoon prompts took 1.5 minutes to complete (SD 8.3; median 0.8), while the longer surveys given with the evening prompts took 5.2 minutes (SD 5.3; median 4.3). The response rates to nonbranching items (ie, did not depend upon the answer of "Yes or No" in the previous question) within a submitted survey were 99% for early afternoon EMA (13 nonbranching items out of 21) and 96% for evening EMA surveys (30 nonbranching items out of 63). The daily response rates for the early afternoon and evening EMA surveys fluctuated between 50% and 60% across days 2-8 (mean 55.6%, SD 3.9% for early afternoon; mean 54.7%, SD 3.2% for evening; Figure 2). Each participant completed an average of 4 early afternoon prompts (SD 2; median 4) and 4 evening prompts (SD 2; median 4). Among the 181 participants, 136 (75.1%) completed >25%, 118 (65.2%) completed >50%, and 71 (39.2%) completed >75% of the early afternoon prompts. For evening prompts, 138 (76.2%), 116 (64.1%), and 69 (38.1%) participants completed >25%, >50%, and >75%, respectively (Table 2).

XSL•FO
RenderX

Figure 2. Daily completion with the EMA survey in the early afternoon (A) and evening (B) over the period of the study period (1 week) for the Beiwe Smartphone pilot study in the Nurses' Health Study II (181 surveys sent each day per each EMA survey type). EMA: ecological momentary assessment.

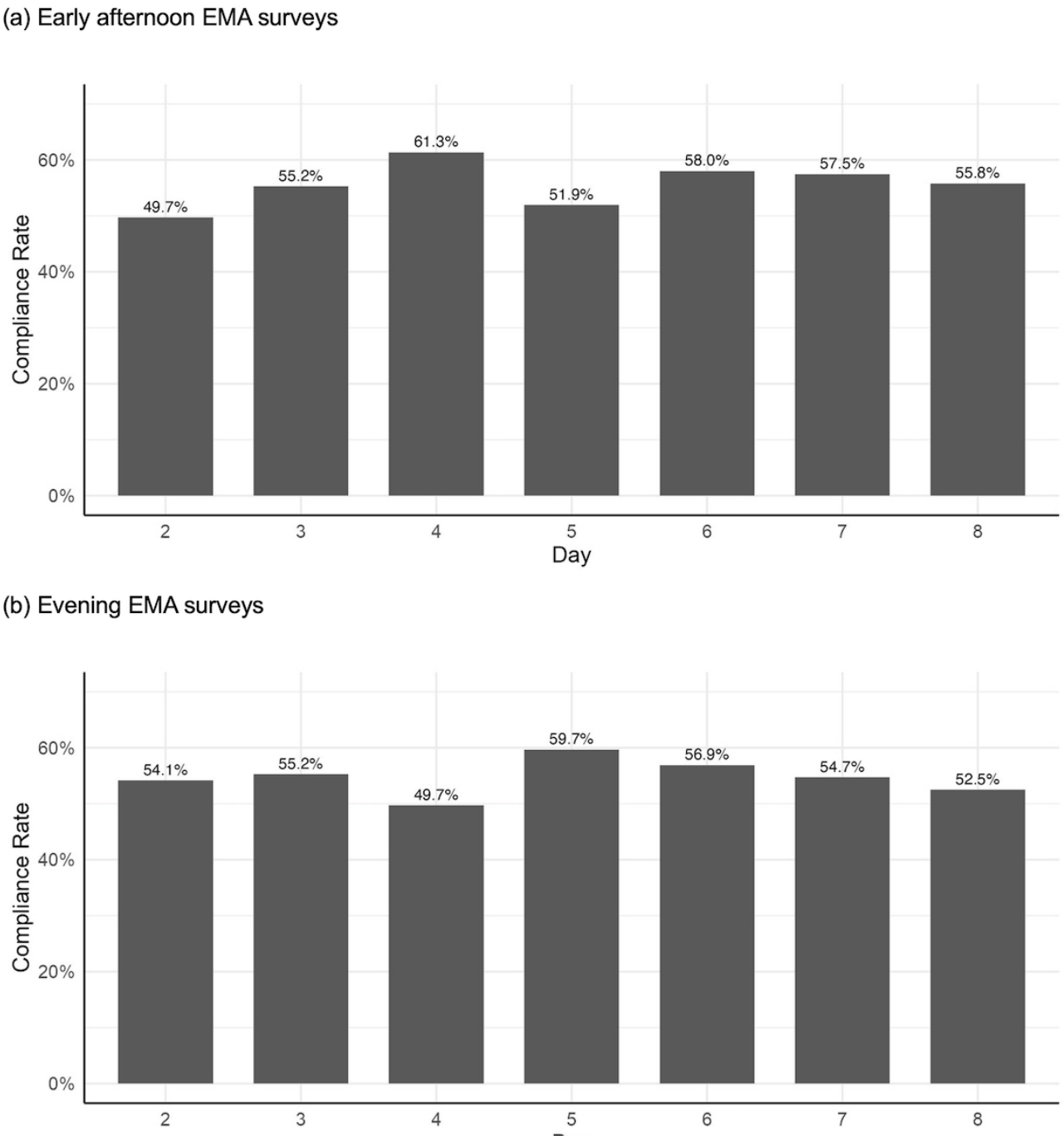


Table . Participant-level compliance of early afternoon and evening ecological momentary assessment surveys over the pilot study period (1 week) (N=181).

Variable	Overall, n (%)	
	Early afternoon survey	Evening survey
Participants completed >25% of prompts sent	136 (75.1)	138 (76.2)
Participants completed >50% of prompts sent	118 (65.2)	116 (64.1)
Participants completed >75% of prompts sent	71 (39.2)	69 (38.1)

Accelerometer and GPS Data Completeness

From 174 out of 181 participants, we collected an average of 11.8 valid hours (SD 5.6, median 12.9) of accelerometer data and 15.2 valid hours (SD 7.0, median 18.0) of GPS data per participant-day in days when any data were collected. Using the criterion of ≥10 valid hours for valid days (Table 3), we obtained 786 valid days (62.04% completeness) of accelerometer

XSL•FO
RenderX

data and 731 valid days (57.70% completeness) of GPS data from 1267 potential data collection days (181 participants on days 2-8).

Table . Completeness of accelerometer and GPS data at the participant-day level over the pilot study period (N=1276[a] total potential data collection days).

Participant day-level data completeness	Overall, n (%)	
	Accelerometer	GPS
Valid days	786 (62.04)	731 (57.70)
Invalid days	293 (23.13)	142 (11,21)
Noncollection days[b]	188 (14.84)	394 (31.10)

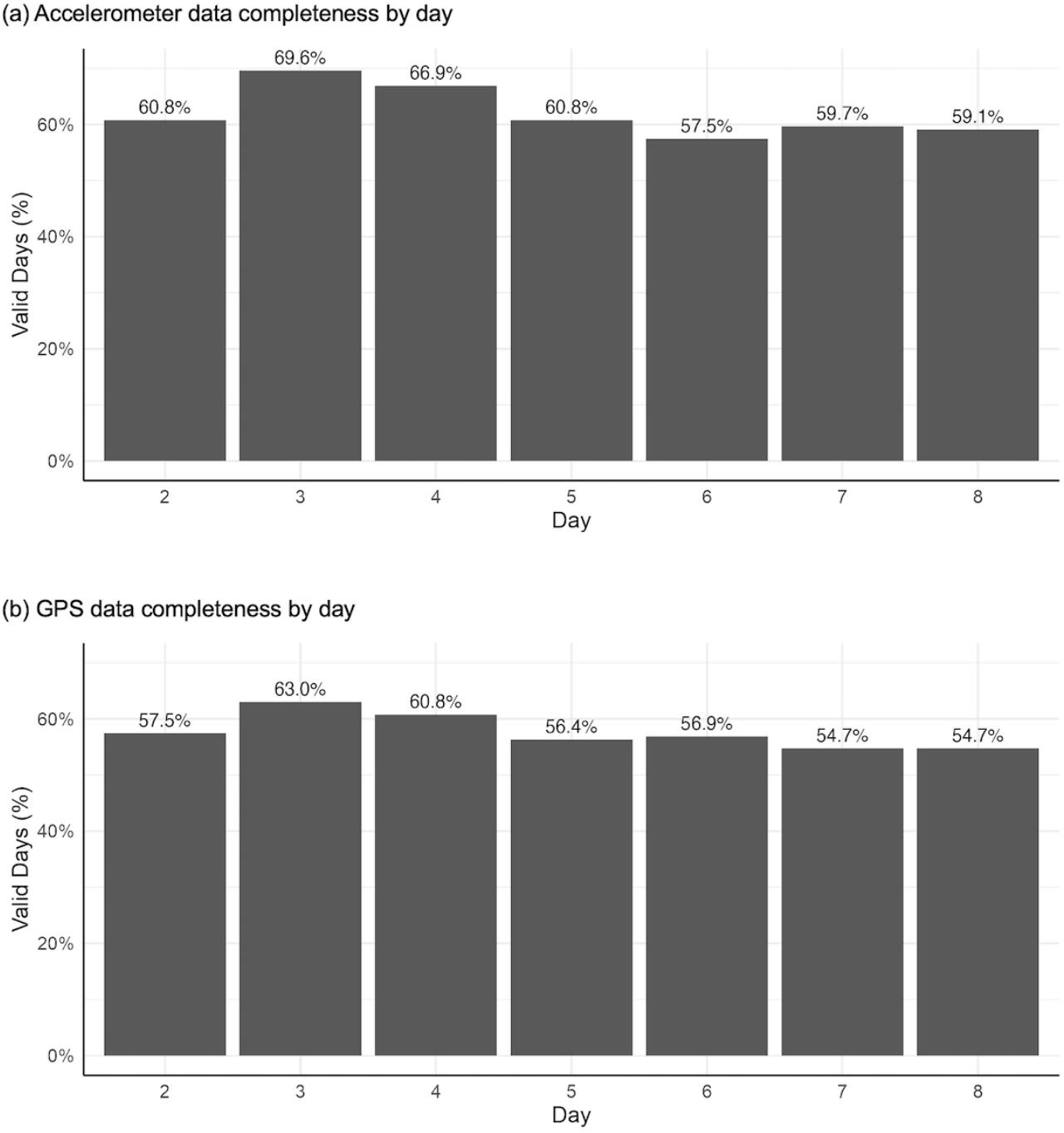
[a]Data included 7 participants who did not contribute any data.

[b]Noncollection days were defined as days without data. In this study, the primary reason for noncollection was attrition (ie, the app was deleted from a participant's phone before the 7-day study period because the participant unenrolled from the substudy). Another reason was sensor noncollection, which could be due to a participant forgetting to charge their phone or disabling the GPS or a major update from the operating system causing the Beiwe app to malfunction temporarily, among other reasons.

Over time, completeness of the accelerometer and GPS data fluctuated between 54.7% and 69.6%, with accelerometer data having slightly higher percentages than GPS data (mean 62.0%, SD 4.5% for accelerometer; mean 57.7%, SD 3.1% for GPS; Figure 3).

XSL•FO
RenderX

Figure 3. Daily completeness of accelerometer (A) and GPS (B) data over the pilot study period (N=1276 total potential data collection days). The statistics for day 1 were not included because the participants started at different times of the day, which meant that the validity criteria were not comparable.



For participant-level completeness, the average valid days for accelerometer and GPS data per participant were 4 for each parameter (SD 3; median 5). Among 181 participants, 145 (80.1%) had 25% valid days of accelerometer data, 112 (61.9%) had 50%, and 79 (43.6%) had 75%, while 125 (69.1%) had 25% valid days of GPS data, 105 (58.0%) had 50%, and 90 (49.7%) had 75% (Table 4).

XSL•FO
RenderX

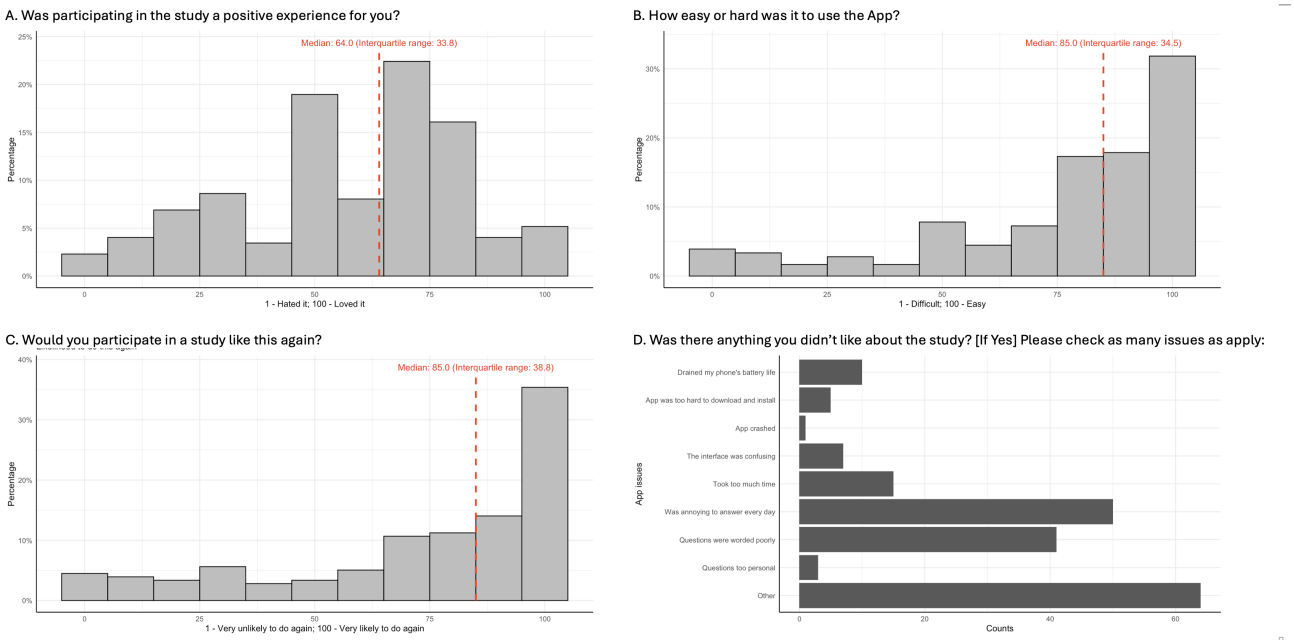
Table . Participant-level data completeness of accelerometer and GPS data over the pilot study period (N=181).

Variable	Overall, n (%)	
	Accelerometer	GPS
Participants (N) who had:		
≥25% data days were valid days	145 (80.1)	125 (69.1)
≥50% data days were valid days	112 (61.9)	105 (58.0)
≥75% data days were valid days	79 (43.6)	90 (49.7)

Feedback Survey

Participants reported an overall relatively "positive experience" (Figure 4A) with data collection (median 64.0 among 174 out of 181 participants, where 1="I hated it" and 100="I loved it") and generally found the Beiwe app easy to use (median 85.0 among 179 out of 181 participants, where 1="difficult" and 100="easy"; Figure 4B). They also indicated a high likelihood of participating again (median 85.0 among 178 out of 181 participants, where 1="Very unlikely" and 100="Very likely"; Figure 4C). Despite this, across a list of potential concerns we provided, 2 main issues were selected with higher frequency (Figure 4D): the Substudy EMA prompt was "annoying to answer every day," and the "questions were worded poorly." Users also reported several major issues of the app, including "App was too hard to download and install" and "The interface was confusing." Additional issues mentioned in the open-ended responses after selecting the "Other" answer included repetitive questions, unclear wording, the need for better notifications such as sound alerts, and lengthy completion times for some tasks (summarized in Multimedia Appendix 4). Users were also asked for open-ended feedback on problems faced during the study (Multimedia Appendix 5 summarized some commonly reported ones). Notably, we found the participants (19/181, 10.5%) who completed the feedback survey but not the baseline surveys mainly had issues including technical difficulties during app installation and registration, inability to reach the study staff for support, and lack of reminders.

Figure 4. An overview of the distribution of answers to the questions in the feedback survey. (A) Whether the study was a positive experience. (B) The ease of use of the app. (C) Likelihood to participate again. (D) Things not liked about the study.



Discussion

Principal Findings

This study examined adherence to a smartphone-based digital phenotyping protocol as part of a pilot study examining willingness to participate in an 8-day intensive measurement protocol to capture psychological well-being, health behaviors, and geolocation data among participants in the NHSII, a nationwide prospective cohort of women in the United States. Our findings indicated high compliance (86%) with the baseline survey and modest compliance (approximately 55% on average) with the daily EMA surveys, with no major differences evident based on the time at which the prompt was delivered (afternoon vs evening) or survey length (evening survey was longer). We also observed good levels of daily completeness (fluctuating between 55% and 70%) of passive sensing data (accelerometer and GPS) using the criterion of 10 valid hours per day. Even with overall modest survey compliance, feedback survey results indicated that participants found the app user-friendly and were willing to engage in similar future studies. However, they also highlighted issues, such as the need for improved survey question wording and user fatigue from repetitive daily surveys. Overall, our evaluation provides valuable insights and lessons

XSL•FO
RenderX

for future studies aimed at incorporating smartphone-based digital phenotyping in large prospective cohorts with participants who more often engage in surveys and periodic assessments than in intensive measurement bursts.

Compared with previous smartphone-based EMA studies that collected data on psychological well-being, our day-level compliance statistics were lower overall. For example, a 2021 systematic review of smartphone-based EMA studies that focused on psychological well-being [38] reported an average compliance rate of 71.6% based on all EMA prompts in each study across 53 studies with a median duration of 7 days. However, these studies often recruited convenience samples and provided compensation to participants. In addition, their participants were often younger (mean 30.8, SD 10.5 years) than our participants (mean 67.8, SD 3.9 years), who may encounter less technological barriers. Furthermore, these studies also represented more of a general population, likely more flexible in survey schedules than our population who were nurses and nursing professionals at baseline. Interestingly, in our study, compliance for the early afternoon and evening EMA surveys varied between 50% and 60% and showed no signs of user fatigue (ie, compliance rates decreased over time). In addition, we found no substantial differences in compliance rates for individual questions in either early afternoon or evening EMA surveys. This is somewhat surprising given that the most common complaint in the feedback survey was that the EMA survey was "boring to answer every day." This may be due in part to conducting this study among a cohort of participants who have been engaged and active in the study for many years.

When we examined survey compliance at the participant level, we found that 32 out of 181 participants were noncompliant (did not submit any early afternoon or evening survey). We therefore speculate that the lower than average compliance with the daily EMA survey may be due to the initial technical difficulties (another common complaint found in the feedback survey) that users encountered in registering and setting up the Beiwe app, which led to early dropout. Therefore, we suggest that future studies ensure timely and adequate technical support during app installation, registration, and throughout the study. This is particularly crucial when the study contains a population that may face higher usability barriers for smartphone apps, such as the older adults in our pilot who may have varying levels of digital literacy [39].

Moreover, according to the feedback survey, participants suggested smartphone EMA survey design ideas to potentially improve their compliance with the schedule, such as daily reminders and notification enhancements with visual or sensory triggers such as icons or sounds, as well as simplifying the layout of the app. We suggest incorporating these functionalities in future studies, particularly in occupational populations with busy work schedules, such as nurses and nursing professionals in our cohort.

Additionally, another 2021 systematic review and meta-analysis [40] found a negative association between the number of items per prompt and EMA compliance in nonclinical studies assessing self-reported health behaviors and psychological constructs in adults. Consequently, our surveys might impose a higher user burden and result in quicker user fatigue than those with fewer questions (eg, asking "how is your current well-being" 3 times a day through established constructs such as Life Orientation Test—Revised and PANAS-X, while most studies previously inquired this through 1 or 2 questions) [41]. Future research should explore strategies to mitigate survey fatigue, such as alternating between truncated and full versions of the EMA survey daily or adopting an intensive longitudinal approach (ie, space out multiple week-long surveys within a year or over several years) [40].

The completeness of our accelerometer and GPS data is slightly lower than that of previous smartphone-based studies that collected similar types of data [17,38,42]. However, this was not unexpected given few studies have simultaneously collected smartphone-based psychological well-being and health behavior data [43-45], which may place higher user burden. From our observations, one of the main reasons for data incompleteness in the passive sensing data was attrition (ie, the app was deleted from a participant's phone before the 7-day study completion date because the participant was unenrolled from the substudy). Besides, to minimize failures in the collection of smartphone passive sensing data, we instructed participants at the start of the data collection period to always carry their smartphones with them. Nonetheless, data noncollection or device noncompliance may still be higher than in similar studies due to the older age of our participants, who may be less likely to carry their smartphones everywhere or keep smartphones on them and charged, compared with the younger populations in these studies [43,44]. This issue was further compounded by our noncontinuous sampling strategy for passive sensing data and the tendency of phone operating systems to disable sensor access when the phone is inactive for extended periods, such as during sleep.

In addition, data collection may also be disrupted by major updates of phone operating system, which may cause the Beiwe app to malfunction temporarily. We recommend that future studies experiment with more advanced passive sensing schemes, such as turning sensors on and off based on the participants' movement status (ie, moving or staying sedentary) rather than a fixed frequency [46], to further conserve batteries while improving the data quality of geolocations and activity tracking.

Furthermore, we found that the completeness of the passive sensing data over the study period was overall higher than EMA survey compliance rate. This indicates that passive sensing data collection may impose a lower burden than smartphone-based surveys and can be extended longer, which is desirable as a 1-week period may not be sufficient for capturing the usual activity patterns of study participants [17,47]. Future studies should consider varying the data collection duration for each data stream instead of maintaining a fixed period for all or reducing the frequency of smartphone surveys (eg, from daily to every other day).

Collecting highly resolved smartphone-based data on psychological well-being and physical activity will facilitate novel investigations of their relationships at finer temporal resolutions, such as daily or diurnal levels. For instance, we can

XSL•FO
RenderX

examine whether a stressful event and strategies used to regulate one's affective states in this context, as reported by the EMA in the last hour, are associated with physical activity levels in the subsequent (or prior) hours. Additionally, we can analyze whether daily positive or negative affect correlates with physical activity on the same or the following day. Furthermore, linking location-based behavior data to environmental factors allows examining the relationship between psychological well-being, health behaviors, and momentary environmental exposures. For example, one could examine whether GPS-based greenspace exposure in the previous hour impacts EMA-reported positive affect. In another example, one can explore momentary associations between GPS-based temperature or $PM_{2.5}$ exposure and smartphone-assessed physical activity [9,48-52].

Strengths and Limitations

Our study has certain limitations. Although our walking recognition algorithm accounts for missing data due to noncontinuous sampling, we may still miss some forms of physical activity due to participants not carrying their smartphones while engaging in these types of activities, which could impact the representativeness of the collected data. Our GPS data may have additional missing data due to environmental conditions (eg, being indoors and heavy tree coverage that can interfere with GPS satellite signals). Additionally, the 1-week data collection period may not capture variability in physical activity and mobility over longer durations, such as seasonal changes [17,53]. The EMA-prompting scheme, administered twice daily, may also fail to capture within-day variations in some experiences, such as positive and negative affect [54]. Furthermore, the NHSII study's homogeneous population—predominantly White, middle-aged and older women who were nurses at the 1989 baseline and have maintained high engagement rates with the cohort over long follow-up periods [18]—may limit the generalizability of our findings regarding device compliance and feasibility of embedding an EMA study within a larger cohort study with participants who generally engage in periodic surveys. Nonetheless, all epidemiologic cohorts focus on compliant participants, and thus our results are informative, especially for other cohorts that may seek to embed smartphone data collection into existing studies.

This study has several strengths. This is one of the first studies to comprehensively examine adherence to an intensive smartphone measurement burst protocol in a large, nationwide prospective cohort of middle-aged and older women. Unlike previous studies, we used multiple metrics to assess compliance with daily EMA surveys over an 8-day period and the completeness of passive sensing data during the same time frame. Our findings offer crucial, detailed insights into the feasibility of and participation in an embedded smartphone-based digital phenotyping protocol into epidemiological cohorts, amidst growing interest in this area [55]. Additionally, the study incorporated feedback survey data, providing subjective and individualized insights that complement objective metrics and help explain trends observed in adherence. This feedback also highlights potential improvements in future study designs and app functionalities.

Conclusions

This pilot study investigated adherence to a smartphone-based data collection protocol embedded within an ongoing cohort study, assessing willingness to participate in an 8-day intensive measurement burst to collect data on psychological well-being, physical activity, and mobility. Our findings suggest modest compliance with smartphone EMA surveys and good completeness of passive sensing data is possible. Adherence remained stable over the 1-week period. Notably, participants expressed a strong interest in participating in similar future studies, according to the feedback survey. These findings have major implications for future research aimed at intensively measuring psychological well-being and physical activity over short periods via smartphones. Smartphone-based digital phenotyping allows researchers to use devices that participants already have and are familiar with for data collection, rather than mailing out wearables, which could be particularly important when seeking to recruit large numbers of participants from established epidemiological cohorts. Overall, these methodologies appear promising for advancing epidemiological research on the psychological and behavioral determinants of health.

Acknowledgments

The authors would like to thank the participants and staff of the Nurses' Health Study II for their valuable contributions. The authors assume full responsibility for analyses and interpretation of these data. This work was supported by the National Institutes of Health (grants R01 HL150119, U01 CA176726, U01 HL145386, and R01 AG85375), Harvard T.H. Chan School of Public Health Dean's Fund for Scientific Advancement Acceleration Award, and the Thomas O. Pyle Fellowship from the Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute.

Data Availability

The datasets generated and analyzed during this study are available in the controlled-access repositories, except for personal identifying data such as names, addresses, smartphone location data, and contact information.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Screenshots of the customized Beiwe app used in the Beiwe Smartphone Substudy of Nurses' Health Study II.
[DOCX File, 218 KB - mhealth_v13i1e71375_app1.docx]

Multimedia Appendix 2
List of survey questions and answer choices delivered by the Beiwe app.
[DOCX File, 27 KB - mhealth_v13i1e71375_app2.docx]

Multimedia Appendix 3
Demographic characteristics of the Beiwe Smartphone Substudy of Nurses' Health Study II cohort by phone operating systems.
[DOCX File, 17 KB - mhealth_v13i1e71375_app3.docx]

Multimedia Appendix 4
A summary of comments participants made to improve the participant experience.
[DOCX File, 22 KB - mhealth_v13i1e71375_app4.docx]

Multimedia Appendix 5
A summary of comments participants made on issues of the app other than the options given.
[DOCX File, 16 KB - mhealth_v13i1e71375_app5.docx]

References

1. Boehm JK, Kubzansky LD. The heart's content: the association between positive psychological well-being and cardiovascular health. Psychol Bull 2012 Jul;138(4):655-691. [doi: 10.1037/a0027448] [Medline: 22506752]
2. James P, Kim ES, Kubzansky LD, Zevon ES, Trudel-Fitzgerald C, Grodstein F. Optimism and healthy aging in women. Am J Prev Med 2019 Jan;56(1):116-124. [doi: 10.1016/j.amepre.2018.07.037] [Medline: 30573140]
3. Martín-María N, Miret M, Caballero FF, et al. The impact of subjective well-being on mortality: a meta-analysis of longitudinal studies in the general population. Psychosom Med 2017 Jun;79(5):565-575. [doi: 10.1097/PSY.0000000000000444] [Medline: 28033196]
4. Trudel-Fitzgerald C, James P, Kim ES, Zevon ES, Grodstein F, Kubzansky LD. Prospective associations of happiness and optimism with lifestyle over up to two decades. Prev Med 2019 Sep;126:105754. [doi: 10.1016/j.ypmed.2019.105754]
5. Boehm JK, Chen Y, Koga H, Mathur MB, Vie LL, Kubzansky LD. Is optimism associated with healthier cardiovascular-related behavior? Meta-analyses of 3 health behaviors. Circ Res 2018 Apr 13;122(8):1119-1134. [doi: 10.1161/CIRCRESAHA.117.310828] [Medline: 29650630]
6. Boehm JK, Qureshi F, Kubzansky LD. Child psychological well-being and adult health behavior and body mass index. Health Psychol 2023 Feb;42(2):73-81. [doi: 10.1037/hea0001261] [Medline: 36595459]
7. Chen R, Del Rosario K, Lockman A, et al. Effects of induced optimism on subjective states, physical activity, and stress reactivity. J Posit Psychol 2023;18(4):592-605. [doi: 10.1080/17439760.2022.2070529] [Medline: 37378047]
8. Kim ES, Kubzansky LD, Soo J, Boehm JK. Maintaining healthy behavior: a prospective study of psychological well-being and physical activity. Ann Behav Med 2017 Jun;51(3):337-347. [doi: 10.1007/s12160-016-9856-y] [Medline: 27822613]
9. Kanning M, Yi L, Yang CH, Niermann C, Fina S. Mental health in urban environments: uncovering the black box of person-place interactions requires interdisciplinary approaches. JMIR Mhealth Uhealth 2023 May 11;11(1):e41345. [doi: 10.2196/41345] [Medline: 37166963]
10. Dunton GF. Sustaining health-protective behaviors such as physical activity and healthy eating. JAMA 2018 Aug 21;320(7):639-640. [doi: 10.1001/jama.2018.6621] [Medline: 29852046]
11. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696. [doi: 10.1038/npp.2016.7]
12. Onnela JP, Dixon C, Griffin K, et al. Beiwe: a data collection platform for high-throughput digital phenotyping. JOSS 2021 Dec 15;6(68):3417. [doi: 10.21105/joss.03417]
13. Liao Y, Chou CP, Huh J, Leventhal A, Dunton G. Examining acute bi-directional relationships between affect, physical feeling states, and physical activity in free-living situations using electronic ecological momentary assessment. J Behav Med 2017 Jun;40(3):445-457. [doi: 10.1007/s10865-016-9808-9] [Medline: 27766481]
14. Liao Y, Schembre SM, O'Connor SG, et al. An electronic ecological momentary assessment study to examine the consumption of high-fat/high-sugar foods, fruits/vegetables, and affective states among women. J Nutr Educ Behav 2018 Jun;50(6):626-631. [doi: 10.1016/j.jneb.2018.02.003] [Medline: 29573964]
15. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]

XSL•FO
RenderX

16. Do B, Mason TB, Yi L, Yang CH, Dunton GF. Momentary associations between stress and physical activity among children using ecological momentary assessment. Psychol Sport Exerc 2021 Jul;55:101935. [doi: 10.1016/j.psychsport.2021.101935] [Medline: 33841048]
17. Yi L, Hart JE, Straczkiewicz M, et al. Measuring environmental and behavioral drivers of chronic diseases using smartphone-based digital phenotyping: intensive longitudinal observational mHealth substudy embedded in 2 prospective cohorts of adults. JMIR Public Health Surveill 2024;10:e55170. [doi: 10.2196/55170]
18. Bao Y, Bertoia ML, Lenart EB, et al. Origin, methods, and evolution of the three nurses' health studies. Am J Public Health 2016 Sep;106(9):1573-1581. [doi: 10.2105/AJPH.2016.303338] [Medline: 27459450]
19. Straczkiewicz M, James P, Onnela JP. A systematic review of smartphone-based human activity recognition methods for health research. NPJ Digit Med 2021 Oct 18;4(1):148. [doi: 10.1038/s41746-021-00514-4] [Medline: 34663863]
20. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. J Pers Soc Psychol 1994;67(6):1063-1078. [doi: 10.1037/0022-3514.67.6.1063]
21. Clark LA, Watson D. E PANAS-x: manual for the positive and negative affect schedule—expanded form. University of Iowa. 1994. [doi: 10.17077/48vt-m4t2]
22. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
23. Del Pozo Cruz B, Ahmadi MN, Lee IM, Stamatakis E. Prospective associations of daily step counts and intensity with cancer and cardiovascular disease incidence and mortality and all-cause mortality. JAMA Intern Med 2022 Nov 1;182(11):1139-1148. [doi: 10.1001/jamainternmed.2022.4000] [Medline: 36094529]
24. Weuve J, Kang JH, Manson JE, Breteler MMB, Ware JH, Grodstein F. Physical activity, including walking, and cognitive function in older women. JAMA 2004 Sep 22;292(12):1454-1461. [doi: 10.1001/jama.292.12.1454] [Medline: 15383516]
25. Straczkiewicz M, Huang EJ, Onnela JP. A "one-size-fits-most" walking recognition method for smartphones, smartwatches, and wearable accelerometers. NPJ Digit Med 2023 Feb 23;6(1):29. [doi: 10.1038/s41746-022-00745-z] [Medline: 36823348]
26. Forest Python library. Onnela Lab. URL: https://github.com/onnela-lab/forest/tree/main/forest [accessed 2025-08-04]
27. Trudel-Fitzgerald C, Tworoger SS, Poole EM, Williams DR, Kubzansky LD. Prospective changes in healthy lifestyle among midlife women: when psychological symptoms get in the way. Am J Prev Med 2016 Sep;51(3):327-335. [doi: 10.1016/j.amepre.2016.04.021] [Medline: 27291076]
28. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]
29. Park CL, Iacocca MO. A stress and coping perspective on health behaviors: theoretical and methodological considerations. Anxiety Stress Coping 2014 Mar 4;27(2):123-137. [doi: 10.1080/10615806.2013.860969]
30. Trudel-Fitzgerald C, Guimond AJ, Kubzansky LD. Emotion regulation and cardiovascular health. In: Handbook of Emotion Regulation: Guilford Press; 2024:465-473.
31. Troy AS, Willroth EC, Shallcross AJ, Giuliani NR, Gross JJ, Mauss IB. Psychological resilience: an affect-regulation framework. Annu Rev Psychol 2023 Jan 18;74(1):547-576. [doi: 10.1146/annurev-psych-020122-041854] [Medline: 36103999]
32. Trudel-Fitzgerald C, Boucher G, Morin C, et al. Coping and emotion regulation: a conceptual and measurement scoping review. Can Psychol 2024 Aug;65(3):149-162. [doi: 10.1037/cap0000377] [Medline: 39371329]
33. Heiy JE, Cheavens JS. Back to basics: a naturalistic assessment of the experience and regulation of emotion. Emotion 2014 Oct;14(5):878-891. [doi: 10.1037/a0037231] [Medline: 24999913]
34. Yi L, Wilson JP, Mason TB, Habre R, Wang S, Dunton GF. Methodologies for assessing contextual exposure to the built environment in physical activity studies: a systematic review. Health Place 2019 Nov;60:102226. [doi: 10.1016/j.healthplace.2019.102226]
35. Liu G, Onnela JP. Bidirectional imputation of spatial GPS trajectories with missingness using sparse online Gaussian Process. J Am Med Inform Assoc 2021 Jul 30;28(8):1777-1784. [doi: 10.1093/jamia/ocab069] [Medline: 34100950]
36. McCullough ML, Willett WC. Evaluating adherence to recommended diets in adults: the Alternate Healthy Eating Index. Public Health Nutr 2006 Feb;9(1A):152-157. [doi: 10.1079/phn2005938] [Medline: 16512963]
37. DeVille NV, Iyer HS, Holland I, et al. Neighborhood socioeconomic status and mortality in the Nurses' Health Study (NHS) and the Nurses' Health Study II (NHSII). Environ Epidemiol 2023 Feb;7(1):e235. [doi: 10.1097/EE9.0000000000000235] [Medline: 36777531]
38. de Vries LP, Baselmans BML, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud 2021;22(5):2361-2408. [doi: 10.1007/s10902-020-00324-7] [Medline: 34720691]
39. Ahmad NA, Mat Ludin AF, Shahar S, Mohd Noah SA, Mohd Tohit N. Willingness, perceived barriers and motivators in adopting mobile applications for health-related interventions among older adults: a scoping review protocol. BMJ Open 2020 Mar 16;10(3):e033870. [doi: 10.1136/bmjopen-2019-033870] [Medline: 32184309]

XSL•FO
RenderX

40. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021 Mar 3;23(3):e17023. [doi: 10.2196/17023] [Medline: 33656451]
41. Stieger S, Götz FM, Gehrig F. Soccer results affect subjective well-being, but only briefly: a smartphone study during the 2014 FIFA World Cup. Front Psychol 2015;6:497. [doi: 10.3389/fpsyg.2015.00497] [Medline: 26029124]
42. Zapata-Lamana R, Lalanza JF, Losilla JM, Parrado E, Capdevila L. mHealth technology for ecological momentary assessment in physical activity research: a systematic review. PeerJ 2020;8:e8848. [doi: 10.7717/peerj.8848] [Medline: 32257648]
43. Triguero-Mas M, Donaire-Gonzalez D, Seto E, et al. Natural outdoor environments and mental health: stress as a possible mechanism. Environ Res 2017 Nov;159:629-638. [doi: 10.1016/j.envres.2017.08.048] [Medline: 28938204]
44. Doherty ST, Lemieux CJ, Canally C. Tracking human activity and well-being in natural environments using wearable sensors and experience sampling. Soc Sci Med 2014 Apr;106:83-92. [doi: 10.1016/j.socscimed.2014.01.048]
45. Yi L, Wang SD, Chu D, Ponnada A, Intille SS, Dunton GF. Examining whether physical activity location choices were associated with weekly physical activity maintenance across 13 months of the COVID-19 pandemic in emerging adults. J Phys Act Health 2022 May 24;19(6):446-455. [doi: 10.1123/jpah.2021-0769]
46. Huang EJ, Onnela JP. Augmented movelet method for activity classification using smartphone gyroscope and accelerometer data. Sensors (Basel) 2020 Jan;20(13):3706. [doi: 10.3390/s20133706]
47. Zenk SN, Matthews SA, Kraft AN, Jones KK. How many days of global positioning system (GPS) monitoring do you need to measure activity space environments in health research? Health Place 2018 May;51:52-60. [doi: 10.1016/j.healthplace.2018.02.004] [Medline: 29549754]
48. Yi L, Xu Y, O'Connor S, et al. GPS-derived environmental exposures during pregnancy and early postpartum–evidence from the MADRES cohort. Science Total Environ 2024 Mar;918:170551. [doi: 10.1016/j.scitotenv.2024.170551]
49. Liu Y, Yi L, Xu Y, et al. Spatial and temporal determinants of particulate matter peak exposures during pregnancy and early postpartum. Environ Adv 2024 Oct;17:100557. [doi: 10.1016/j.envadv.2024.100557] [Medline: 39574825]
50. Yi L, Habre R, Mason TB, et al. Smartphone GPS-based exposure to greenspace and walkability and accelerometer-assessed physical activity during pregnancy and early postpartum-evidence from the MADRES cohort. J Urban Health 2024 Dec;101(6):1128-1142. [doi: 10.1007/s11524-024-00903-6] [Medline: 39145858]
51. Xu Y, Yi L, Cabison J, et al. The impact of GPS-derived activity spaces on personal PM2.5 exposures in the MADRES cohort. Environ Res 2022 Nov;214:114029. [doi: 10.1016/j.envres.2022.114029]
52. Yi L, James P. Invited perspective: how do green- and bluespaces reduce heat-related health risks? Gaining new insights from street-view imagery, deep learning models, and smartphone data. Environ Health Perspect 2025 Apr 14;0. [doi: 10.1289/EHP15400] [Medline: 40228076]
53. Fore R, Hart JE, Choirat C, et al. Embedding mobile health technology into the nurses' health study 3 to study behavioral risk factors for cancer. Cancer Epidemiol Biomarkers Prev 2020 Apr;29(4):736-743. [doi: 10.1158/1055-9965.EPI-19-1386] [Medline: 32098894]
54. Dunton GF, Liao Y, Dzubur E, et al. Investigating within-day and longitudinal effects of maternal stress on children's physical activity, dietary intake, and body composition: protocol for the MATCH study. Contemp Clin Trials 2015 Jul;43:142-154. [doi: 10.1016/j.cct.2015.05.007] [Medline: 25987483]
55. Hystad P, Amram O, Oje F, et al. Bring your own location data: use of google smartphone location history data for environmental health research. Environ Health Perspect 2022 Nov;130(11). [doi: 10.1289/EHP10829]

Abbreviations

AHEI: Alternative Healthy Eating Index
CONSORT: Consolidated Standards of Reporting Trials
EMA: ecological momentary assessment
NHSII: Nurses' Health Study II
PANAS-X: Positive and Negative Affect Schedule—Expanded Form

Edited by L Buis; submitted 16.01.25; peer-reviewed by A Bunevicius, H Maheshwari; revised version received 24.04.25; accepted 04.06.25; published 03.09.25.

Please cite as:
Yi L, Trudel-Fitzgerald C, Hu CR, Wilt G, Chavarro J, Onnela JP, Grodstein F, Kubzansky LD, James P
Measuring Psychological Well-Being and Behaviors Using Smartphone-Based Digital Phenotyping: An Intensive Longitudinal Observational mHealth Pilot Study Embedded in a Prospective Cohort of Women
JMIR Mhealth Uhealth 2025;13:e71375
URL: https://mhealth.jmir.org/2025/1/e71375
doi:10.2196/71375

XSL•FO
RenderX

© Li Yi, Claudia Trudel-Fitzgerald, Cindy R Hu, Grete Wilt, Jorge Chavarro, Jukka-Pekka Onnela, Francine Grodstein, Laura D Kubzansky, Peter James. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Age-Related Patterns in Smartphone Keystroke Dynamics Considering Temporal Variability: Cross-Sectional Study With AI-Based Analysis

Junhyung Moon[1*], PhD; Yu Lim Huh[2*], BS; Hee Young Cho[3,4*], MD, PhD; Chaeyeon Kim[2], BS; Hyeongrae Lee[5], PhD; Dong Pyo Jang[5], PhD; Baek Hwan Cho[1], PhD

[1]Department of Biomedical Informatics, CHA University School of Medicine, CHA University, Seongnam, Republic of Korea

[2]Department of Electronic Engineering, Hanyang University, Seoul, Republic of Korea

[3]Department of Obstetrics and Gynecology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[4]Institute of Reproductive Medicine and Population, Medical Research Center, Seoul National University, Seoul, Republic of Korea

[5]Department of Biomedical Engineering, Hanyang University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Baek Hwan Cho, PhD
Department of Biomedical Informatics, CHA University School of Medicine
CHA University
335, Pangyo-ro, Bundang-gu
Seongnam, 13488
Republic of Korea
Phone: 82 10 8704 5768
Email: baekhwan.cho@cha.ac.kr

Abstract

Background: Keystroke dynamics on smartphones have emerged as a promising form of passive digital biomarker. While previous studies have explored their utility in several diseases and disorders, relatively few have examined how these dynamics change systematically with chronological age in the general population.

Objective: This study aimed to investigate age-related patterns in mobile keystroke dynamics, with a particular focus on temporal variations throughout the day. By identifying behavioral signatures associated with different age groups, we further assess whether artificial intelligence–based models can accurately estimate chronological age using passively collected keystroke data.

Methods: We conducted a field study involving 177 healthy adults in the Republic of Korea, collecting free-living smartphone typing logs over multiple weeks through a custom Android keyboard app (CodeRed Corp). For each keystroke, the app recorded press and release timestamps and key type, from which 43 behavioral features were extracted across categories of speed, frequency, and temporal variability. Weekly feature vectors were constructed at 3 temporal resolutions (6-hour intervals, daily, and weekly). In total, 8 artificial intelligence models, including random forest, TabNet, transformer, and long short-term memory, were trained with participant-wise 10-fold cross-validation. A custom loss function was introduced to reduce intraparticipant prediction variability. Descriptive statistics and ablation studies were conducted to assess behavioral trends and feature contributions.

Results: The study included 177 participants (female: n=115; male: n=62) with a mean age of 28.8 (SD 11.1) years, all residing in the Republic of Korea. On average, data were collected for 25 weeks per participant, resulting in a dataset of more than 2.5 million typing sessions. Descriptive analysis revealed clear age-related differences. Younger participants typed faster and more frequently, while older participants showed slower and more variable typing. The long short-term memory model using the 6-hour interval median features achieved the best age estimation performance (mean absolute error 3.69 years, R^2=0.71). When the customized loss function was applied, the model's performance further improved to a mean absolute error of 3.60, with a reduction in intraparticipant variability in estimated ages by 7.8%. Notably, feature importance analysis suggested that the early morning (midnight to 6 AM) and late evening (6 PM to midnight) periods may carry more age-discriminative keystroke patterns.

Conclusions: Our findings demonstrated that smartphone keystroke dynamics reflect age-sensitive behavioral patterns, particularly when analyzed with fine-grained temporal resolution. While the primary goal was not age estimation per se, the ability to model

XSL•FO
RenderX

these patterns highlights the potential of keystroke dynamics as a passive, unobtrusive behavioral marker for age-related functional characteristics. These insights may inform future applications in digital health, such as age-sensitive personalization or early detection of age-related decline without requiring any active user input.

(JMIR Mhealth Uhealth 2025;13:e80094) doi:10.2196/80094

KEYWORDS

keystroke dynamics; digital biomarkers; age-related behavior; passive sensing; mobile health; temporal variability; artificial intelligence; mobile phone

Introduction

The rapid expansion of mobile health apps has enabled individuals to monitor and manage their health conveniently and continuously. However, these services often rely on users' active engagement, such as logging symptoms or responding to surveys. This dependence on manual input can present a barrier to long-term use. Passive sensing techniques can offer a promising alternative in digital health care, as they enable unobtrusive and continuous health monitoring without requiring explicit user participation [1-4].

Among various passive sensing modalities, digital biomarkers derived from everyday interactions with smartphones have received increasing attention, such as GPS [5,6], biosignals [7,8], and motion sensor data [9,10]. One promising approach involves analyzing mobile keystroke dynamics—patterns such as typing speed, interkey delay, and error frequency during smartphone keyboard usage. These subtle signals are unobtrusively collected and can reflect changes in cognitive and motor function [11-13], offering a valuable source of information for health-related inference.

Keystroke dynamics have been explored in previous work for various health domains like mild cognitive impairment [14], multiple sclerosis [15,16], loneliness [17], bipolar disorder [18], and mental health [13,19]. Specifically, Vesel et al [13] examined how mood, age, and diurnal patterns affect mobile keystroke dynamics using hierarchical mixed-effects models. However, their primary focus was on mood symptomatology—treating keystroke features, such as typing speed and variability, as outcomes influenced by depression severity, age, and time of day. Similarly, Zulueta et al [18] estimated brain age from smartphone typing kinematics and analyzed prediction error in relation to bipolar disorder risk. However, their focus was on diagnostic group differences rather than exploring how keystroke patterns change with age in the general population.

In contrast, our study aimed to conduct an in-depth investigation of age-related variations in keystroke dynamics, with particular attention to temporal keystroke patterns across different age groups. Given that aging is often accompanied by gradual changes in motor and cognitive function, these shifts may be subtly reflected in how individuals type on their smartphones. Importantly, such functional changes are not exclusive to aging—they may also result from sudden health deterioration. For instance, in patients at risk of cognitive impairment, subtle slowing and increasing variability in keystroke patterns may serve as early indicators of decline. Beyond health care, the ability to continuously and passively detect such functional changes could also contribute to security-related monitoring purposes. For instance, keystroke dynamics could reveal anomalous behavior when a compromised account is accessed by someone whose typing rhythm deviates markedly from the legitimate user, thereby serving as a continuous authentication layer.

Methods

Data Collection

We conducted a field study with a total of 177 healthy adults in the Republic of Korea from September 2022 to September 2023. Participants were recruited through online university communities and campus bulletin boards. Inclusion criteria were male and female individuals between 14 and 69 years of age without any mental or physical disabilities, who were Android smartphone users. Exclusion criteria were (1) inability or unwillingness to follow the investigator's instructions; (2) difficulty in communication or lack of cooperation with the study procedures; (3) explicit refusal to participate in the study; (4) previous diagnosis of any psychiatric or physical illness; and (5) for minors (aged 14-19 years), lack of consent from a parent or legal guardian, or cases where parental consent was provided but the participant did not assent to participate. Those interested submitted an online application form that included their age, sex, smartphone operating system, preferred keyboard layout, and contact information. Eligible individuals were contacted to schedule an online meeting, during which they received a detailed explanation of the study and provided informed consent electronically. Following consent, participants were guided through the installation process and instructed to use our custom keyboard app as their default keyboard in daily life. The study was conducted exclusively in smartphone environments, and other platforms, such as desktop keyboards, were not included. Our custom keyboard app, developed for the Android operating system (CodeRed Corp), supported QWERTY layouts for Korean, English, numbers, and special characters. It recorded timestamps for each key-pressing and key-releasing event, along with the type of key (eg, Korean, English, delete, and special character). Per participant, a log file was generated, in which each row contained the timestamp of key press, the timestamp of key release, and the type of the pressed key. These files were automatically and securely transmitted to our research server on a daily basis. Data were collected during participants' natural smartphone usage over a multiple-week period, without task-specific constraints, to ensure ecological validity.

XSL•FO
RenderX

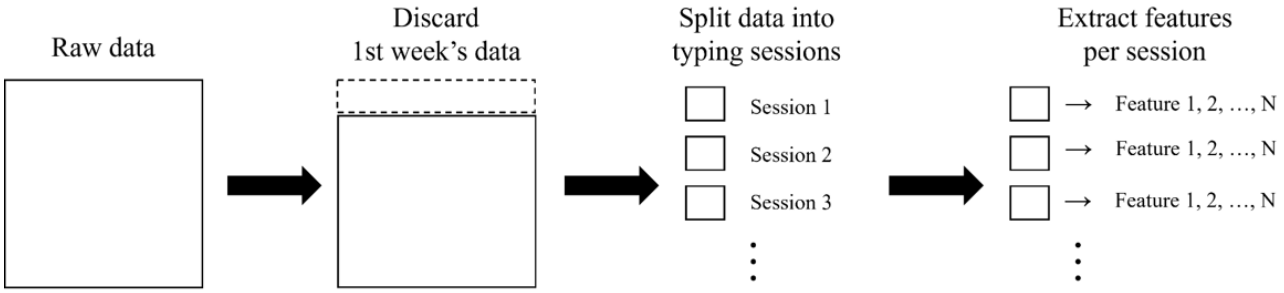
Ethical Considerations

The study was approved by the Institutional Review Board of Hanyang University (HYUIRB-202207-003-04). All participants were provided with detailed information about the study, including its objectives, the procedures for data collection, the devices and mobile apps to be used, the types of data to be collected, and the subsequent processing of these data. They were clearly informed that participation was entirely voluntary and that they could withdraw from the study at any time without penalty. Written informed consent was obtained from all participants before enrollment. A modest financial incentive was offered as compensation for their time and effort. The entire collected dataset was deidentified.

Data Preprocessing and Feature Extraction

As illustrated in Figure 1, we first discarded the data collected from each participant during the first week, since most participants spent the first week as an adaptation week to practice the provided keyboard app. For the remaining data, we excluded the key-press logs whose duration between press and release is longer than 5 seconds. Then, we processed the remaining logs in units of typing sessions. The typing session was determined by referring to the work by Vesel et al [13]: (1) duration between the keyboard activation and deactivation, or (2) if there was no press for more than 8 seconds after the last key release. We excluded the data of the typing sessions where key-pressing activities occurred more than or equal to 10 times per second (ie, interkey intervals ≤100 ms). This conservative threshold was chosen to eliminate abnormal or artifact-prone events, with reference to a large-scale study by Dhakal et al [20]. They analyzed 136,857,600 keystrokes from 168,960 participants who were asked to transcribe 15 English sentences using desktop keyboards. Their results showed that the average interkey interval across all participants was 239 milliseconds, whereas even the fastest typists exhibited an average of 122 milliseconds. Although their study was based on desktop keyboards, these findings provide a reasonable reference point for identifying abnormal keystrokes.

Figure 1. Flow of data preprocessing and feature extraction for each participant's data.



Within each typing session, we calculated speed and frequency-related features (speed and frequency, respectively, in Table 1), based on the existing studies [13,15-18]. The speed feature includes hold time (HT; a duration of each key press), flight time (a duration between release of the previous key and press of the current key), press-to-press latency (PPL), and release-to-release latency. Additionally, we calculated the correction, the HT for the delete key, and pre- and postcorrection, the flight times before and after pressing the delete key, respectively. The frequency features are the number of times any key is pressed and pressing the delete key, respectively (AllKey and DEL). In addition, we calculated various features in order to represent keystroke dynamics in terms of speed and frequency in a more delicate manner. First, we added the features regarding the keyboard layout shift (eg, from Korean QWERTY to English QWERTY): duration of layout shift (dur_shift) and the number of layout shifts (num_shift). Second, we counted the number of pressing any keys and the delete key per second, respectively (AllKeyRate and DELRate). Third, we added 2 features, that is, the duration of typing session (dur_session) and the number of typing sessions (num_session).

XSL•FO
RenderX

Table 1. Keystroke dynamics features (n=43) categorized into speed-related, frequency-related, and temporal variability-related features, which were used as input variables for the analyses in our study.

Category	Value, n	Features
Speed	11	HT[a], FT[b], PPL[c], RRL[d], precorrection, correction, postcorrection, AllKeyRate, DELRate, dur_shift, dur_session
Frequency	4	AllKey, DEL, num_shift, num_session
Speed variability	22	HT_MaxGap, FT_MaxGap, PPL_MaxGap, RRL_MaxGap, precorrection_MaxGap, correction_MaxGap, postcorrection_MaxGap, AllKeyRate_MaxGap, DELRate_MaxGap, dur_shift_MaxGap, dur_session_MaxGap, HT_STD, FT_STD, PPL_STD, RRL_STD, precorrection_STD, correction_STD, postcorrection_STD, AllKeyRate_STD, DELRate_STD, dur_shift_STD, dur_session_STD
Frequency variability	6	AllKey_MaxGap, DEL_MaxGap, num_shift_MaxGap, AllKey_STD, DEL_STD, num_shift_STD

[a]HT: hold time.

[b]FT: flight time.

[c]PPL: press-to-press latency.

[d]RRL: release-to-release latency.

Finally, we designed additional features to capture the temporal variability of speed and frequency features, respectively (speed variability and frequency variability in Table 1). We hypothesized that keystroke dynamics would fluctuate across different times of the day (eg, morning, afternoon, and night), analogous to the diurnal variations observed in physiological indicators [21]. We defined 2 types of variability features. The first, denoted as {Feature_Name}_MaxGap, represents the maximum fluctuation within a given time window (eg, day). It is calculated as the difference between the 95th and 5th percentiles of the feature's values. The other is the SD of each feature's values within a time window ({Feature_Name}_STD).

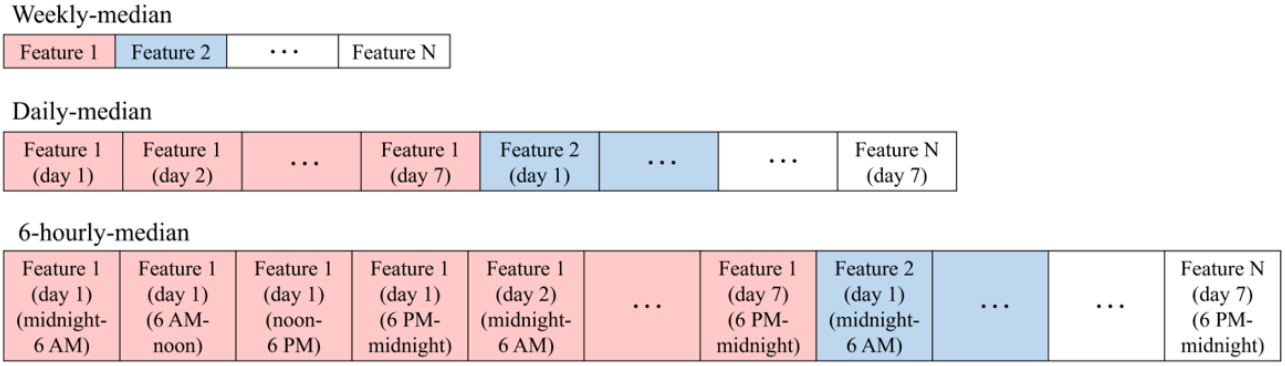
Analysis Through Descriptive and Artificial Intelligence–Based Methods

We analyzed the extracted 43 features, as described in Table 1, across different chronological age groups (younger-than-30-years, 30s, 40s, and 50s). Considering potential options for the analysis time window, such as an hour, a day, and a week, we empirically determined to conduct an analysis on a weekly basis. We believed that a larger window, such as a month, might fail to capture the temporary change in health status, while a smaller window, such as a day, would lead to unnecessarily frequent analyses, introducing redundancy without improving interpretability. For visualization, we followed a three-step aggregation process. First, for each participant, multiple values of each feature were extracted within every 1-hour window. Second, the median value of each feature was computed per participant for each hour. Finally, the average of these per-participant medians was calculated across all participants within the same age group for each hour.

In addition to the descriptive analysis, we investigated the feasibility of artificial intelligence (AI)–based age estimation. We organized the extracted features as weekly feature vectors to serve as inputs for nonsequential and sequential AI models, introduced in the Age Estimation Model Development section. To construct the weekly feature vectors for nonsequential models, we used 3 temporal resolutions: weekly median, daily median, and 6-hour interval median (Figure 2). In the weekly median case, we built a feature vector using the median values of all features for each week. In the daily median case, we organized a feature vector by arranging the median value of each feature per day within each week. Considering the diurnal rhythm mentioned above, we split a day-resolution into hour levels. In the 6-hour interval median case, we constructed a feature vector by arranging the median value of each feature in 6-hour intervals (midnight-6 AM, 6 AM-noon, noon-6 PM, and 6 PM-midnight) for each week. For sequential models, such as a transformer-based and the long short-term memory (LSTM)–based models, the same features were organized as a temporal sequence across days and 6-hour periods, allowing the model to learn temporal dependencies. Accordingly, we used the daily median and 6-hour interval median resolutions only for sequential models, since the sequential information was already abstracted as a median value in the case of the weekly median. Thus, for sequential models, the feature vector using the daily median has a sequence length of 7, while the feature vector using the 6-hour interval median has a sequence length of 28. Given the importance of data order in sequential models, we further adjusted the organized feature vectors to begin with Monday's data and end with Sunday's data.

XSL•FO
RenderX

Figure 2. Examples of weekly feature vectors according to different temporal resolutions.



Age Estimation Model Development

To train and evaluate AI models, we constructed 7 datasets from the original dataset to ensure balanced representation across different age groups. Specifically, we divided the data of participants younger than 30 years into 7 participant-wise subsets (17×1+18×6). This division was based on the average number of participants in the other 3 age groups (30s, 40s, and 50s), which was approximately 17.3. Thus, we used 17 and 18 as the subset sizes for the younger-than-30-years group. Each subset of the younger-than-30-years group was then combined with the full data from the other age groups to form 7 age-balanced datasets. The resulting datasets had mean ages ranging from 38.5 to 38.9 (SD 0.1) years. Sex distribution was also comparable across the datasets, with the number of female participants ranging from 41 to 48 and male participants from 22 to 29. Within each dataset, missing values were imputed using the values from the nearest week. Minimum-maximum normalization was applied within each constructed dataset.

For the age estimation in this study, we used 8 AI models: logistic regression, random forest regressor, extreme gradient boosting (XGBoost) regressor, CatBoost regressor, multilayer perceptron (MLP), and TabNet [22], as well as sequential models based on a transformer and LSTM. Using Python version 3.11.4, we implemented the 8 AI models, which take the constructed weekly feature vectors and then output the estimated age.

For each of the 7 age-balanced datasets, we conducted 10-fold cross-validation while exploring the hyperparameter space to maximize the regression performance. We divided the individual dataset into 10 folds in a participant-wise manner, ensuring that the same participant's data did not appear in both training and test sets. The fold construction was based solely on participant separation without rebalancing age distribution across folds. For logistic regression, we opted for a consistent hyperparameter set, choosing "liblinear" as the solver and "L2" for regularization. For the random forest regressor, XGBoost regressor, and CatBoost regressor, we adjusted the number of estimators and the depth of each estimator, additionally changing the learning rate only for the XGBoost regressor and CatBoost regressor. For MLP, we stacked BatchNorm1d, Linear, ReLU, and Dropout layers using the torch package. Within the architecture, we adjusted the number of layers, the number of nodes in each layer, the number of epochs, and the batch size. We used a dropout rate of 0.3, a weight decay of 0.001, and the ReduceLROnPlateau scheduler for the learning rate. For TabNet, a deep learning model designed for tabular data, we used the pytorch_tabnet package. We adjusted the dimension of decision prediction layers, the dimension of the attentive transformer, the number of decision steps, the number of epochs, and batch size, while using the Adam optimizer and the ReduceLROnPlateau scheduler, from the torch package, for the learning rate. For the transformer-based model, we examined various options for the number of layers, the number of attention heads, the hidden dimension, the learning rate, and the batch size. For the LSTM-based model, we regulated values of the number of layers, the number of nodes in each layer, the learning rate, and the batch size. For the deep learning models, in each fold, we stored the model from the epoch that achieved the best performance on the validation set and used it for evaluation, rather than simply reporting the last epoch.

We hypothesized that if an individual is in good health, the estimated age would remain relatively stable across different weeks. Given that our participants were from a healthy population, we customized the loss function of the AI model to minimize deviations in estimated age within the same participant. The training process was divided into 2 phases. In the first phase, we constructed mini-batches such that each batch contained multiple weekly samples from the same participant. For each batch, we calculated the average L1 loss between the real and estimated ages across the participant's weekly samples. If this average L1 loss exceeded an empirically defined threshold, a penalty term—computed as the average L1 loss multiplied by a weight—was added to the original loss. We experimented with threshold values of 1, 2, and 3, and weight values of 0.2, 0.4, 0.6, 0.8, and 1.0. In the second phase, training continued using the standard L1 loss without the additional penalty. We tuned the number of epochs for each phase to achieve optimal age estimation performance.

Upon completion of model training, we assessed age estimation performance based on mean absolute error (MAE) and R^2. Each metric was then averaged across the results from the 7 age-balanced datasets to obtain the final evaluation scores.

Results

Participant Data Overview

Table 2 shows the characteristics of the dataset of the entire 177 participants, including a sex distribution of 115 females and 62

XSL•FO
RenderX

males. All of them were residents of the Republic of Korea. The participants had a mean age of 28.8 (SD 11.1) years, with a minimum of 19 years and a maximum of 58 years. The majority (n=125) were younger than 30 years old, while the remaining participants were distributed across the following age groups: 30-39 years (n=16), 40-49 years (n=19), and 50-59 years (n=17). On average, data were collected over a 25-week period per participant, ranging from a minimum of 4 weeks to a maximum of 41 weeks. In total, 2,597,692 typing sessions were collected, of which 89,219 (3.4%) were discarded because they contained at least 1 instance of ≥10 consecutive keystrokes occurring within 1 second.

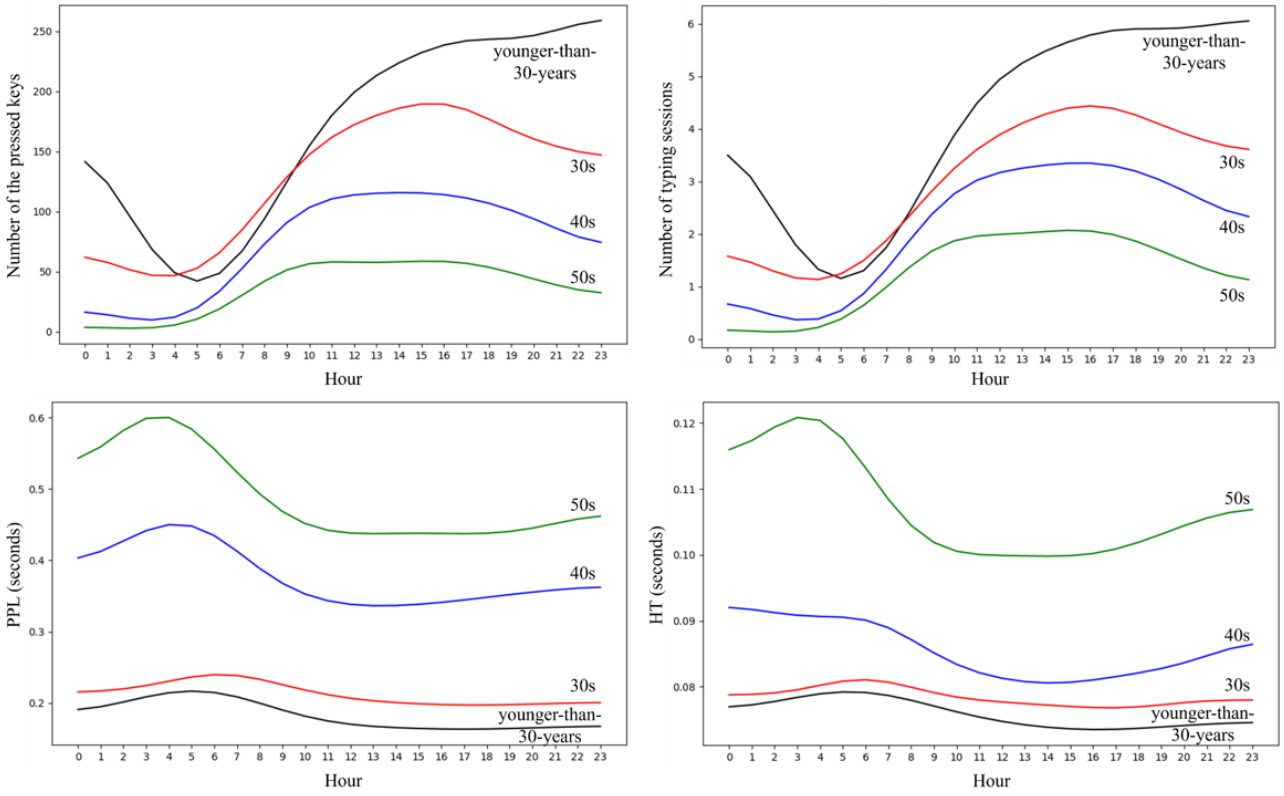
Table 2. Demographic distribution of participants and the duration of keystroke data collection across all participants in our study.

Item	Value
Participants, n	177
Age (y), mean (SD)	28.8 (11.1)
Total group age range (y), n	
<30	125
30-39	16
40-49	19
50-59	17
Sex (female:male), n	115:62
Data collection weeks, mean (SD; minimum-maximum)	25.0 (13.3; 4-41)

Descriptive Analysis Results for Different Age Groups

Figure 3 illustrates the hourly patterns of 4 features—number of pressed keys, number of typing sessions, HT, and PPL—across different age groups (younger than 30 years, 30s, 40s, and 50s). As illustrated in Figure 3, the younger groups (younger than 30 years in black, and 30s in red) consistently exhibited a higher number of pressed keys throughout the day compared with the older groups (40s in blue, and 50s in green). Notably, the younger than 30 years group peaked at 11 PM, whereas the older groups showed a slight decline approaching that hour. The number of typing sessions showed a similar pattern. PPL remained low (approximately 0.2 seconds) for the younger groups, while it was higher for the older groups. HT followed a similar trend to PPL.

Figure 3. Visualization of keystroke dynamics across age groups, showing the hourly distribution of the number of key presses, the number of typing sessions, press-to-press latency, and hold time, highlighting age-related differences over the course of a day. HT: hold time; PPL: press-to-press latency.



XSL•FO
RenderX

Age Estimation Results Using Keystroke Dynamics

Table 3 presents the age estimation results obtained from the 8 AI models used in our study, based on speed and frequency features. These results were produced without applying the customized loss function, as our goal was to first report baseline performance from conventional AI models. Among the nonsequential models, MLP and TabNet demonstrated competitive performance, with MAEs of 3.85 and 4.01, respectively, under the weekly median resolution. These 2 deep learning–based models consistently outperformed traditional machine learning models across all temporal resolutions. Overall, the sequential model LSTM showed improved performance as the temporal resolution became finer. With the 6-hour interval median resolution, LSTM achieved an MAE of 3.69 and an R^2 of 0.71, which is the optimal performance in this experiment. R^2 generally increased with finer temporal resolutions across most of the 8 AI models. Although we also tested the 3-hour interval median resolution, it did not lead to further performance improvement (MAE remained at 3.69).

Table 3. Age estimation results of 8 artificial intelligence models using speed and frequency features and 3 different temporal resolutions (weekly median, daily median, and 6-hour interval median).

	Weekly median		Daily median		6-hour interval median	
AI[a] model	MAE[b]	R^2	MAE	R^2	MAE	R^2
Nonsequential models						
Logistic regression	6.43	0.01	6.77	–0.09	6.81	0.13
Random forest regressor	5.49	0.44	5.34	0.48	5.13	0.53
XGBoost[c] regressor	5.54	0.47	5.38	0.49	5.07	0.55
CatBoost regressor	5.26	0.50	5.28	0.50	5.04	0.55
MLP[d]	3.85	0.69	4.19	0.65	4.36	0.65
TabNet	4.01	0.67	4.24	0.64	4.04	0.65
Sequential models						
Transformer	—[e]	—	4.51	0.59	4.56	0.58
LSTM[f]	—	—	3.86	0.68	3.69	0.71

[a]AI: artificial intelligence.

[b]MAE: mean absolute error.

[c]XGBoost: extreme gradient boosting.

[d]MLP: multilayer perceptron.

[e]Not applicable.

[f]LSTM: long short-term memory.

Impact of Different Combinations of Feature Sets

Using the LSTM model with the 6-hour interval median resolution, which achieved the optimal MAE of 3.69 (Table 3), we conducted an ablation study to examine the effectiveness of different feature set combinations in our age estimation task. For each experiment in Table 4, we regulated hyperparameters of the LSTM model to find the optimal MAE. Overall, the speed feature appeared more influential for age estimation than the frequency feature. Using only the speed feature yielded a comparable MAE of 3.70, in which both speed and frequency features were used (MAE=3.69). In contrast, the inclusion of speed and frequency variability features did not result in noticeable improvements, suggesting their limited effectiveness in this context. When used in isolation, the frequency feature and its variability yielded limited but marginal improvements.

XSL•FO
RenderX

Table 4. Results of the ablation analysis examining the impact of using different combinations of feature sets on age estimation performance.

Number of features	Speed	Frequency	Speed variability	Frequency variability	MAE[a]
15	✓	✓			3.69
43	✓	✓	✓	✓	3.88
11	✓				3.70
33	✓		✓		3.86
4		✓			6.33
10		✓		✓	5.89
22			✓		4.06
6				✓	6.37
28			✓	✓	4.07

[a]MAE: mean absolute error.

Impact of Different Combinations of 6-Hour Periods

Using the LSTM model with the 6-hour interval median resolution, we conducted an additional ablation study to examine the impact of using different 6-hour periods within a single day. For this experiment, we fixed the hyperparameters to those of the optimal case presented in Table 3. As illustrated in Table 5, we conducted experiments using all possible combinations of the four 6-hour periods. Overall, the MAE increased as the total number of hours used for age estimation decreased. When using only one 6-hour period, the MAE ranged from 4.34 to 4.61, indicating relatively poor performance. With 12-hour combinations, the optimal performance (MAE=3.95) was achieved when using data from midnight-6 AM and 6 PM-midnight. A similar trend was observed for the 18-hour combinations, where the lowest MAE (3.87) occurred in 2 combinations that included these time intervals.

Table 5. Results of the ablation analysis examining the impact of using different combinations of 6-hour periods within a day on age estimation performance.

Total duration	Midnight-6 AM	6 AM-noon	Noon-6 PM	6 PM-midnight	MAE[a]
24 h					
1	✓	✓	✓	✓	3.69
18 h					
1		✓	✓	✓	4.02
2	✓		✓	✓	3.91
3	✓	✓		✓	3.87
4	✓	✓	✓		3.87
12 h					
1			✓	✓	4.13
2		✓		✓	4.06
3		✓	✓		4.26
4	✓			✓	3.95
5	✓		✓		3.96
6	✓	✓			4.01
6 h					
1	✓				4.43
2		✓			4.61
3			✓		4.50
4				✓	4.34

[a]MAE: mean absolute error.

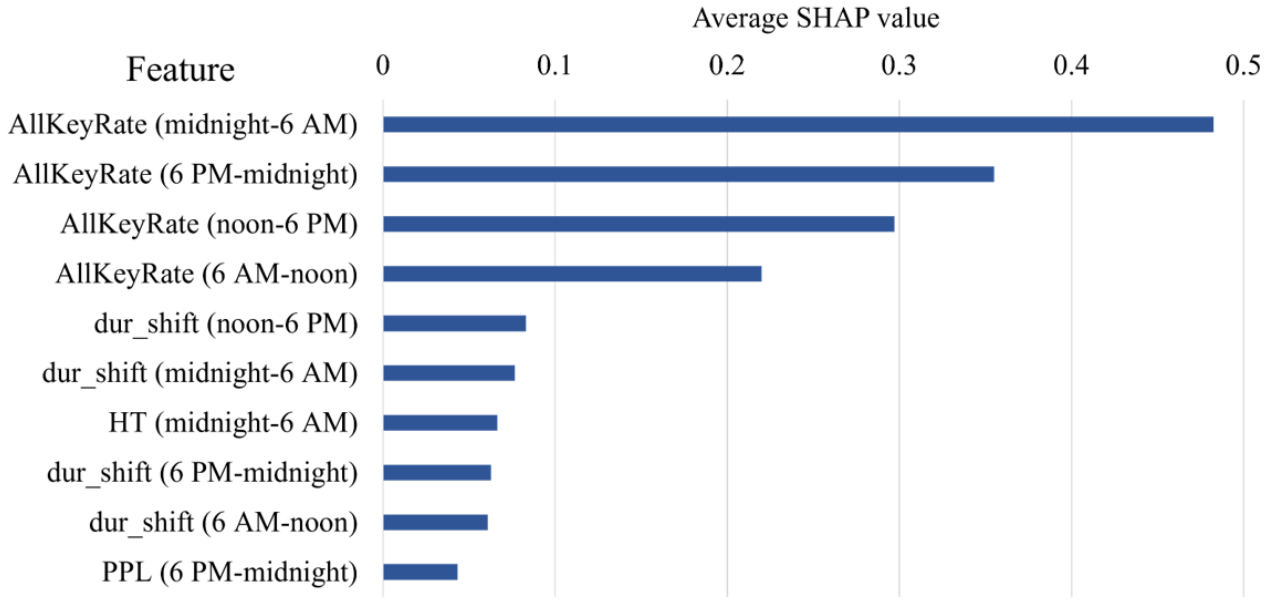
XSL•FO
RenderX

Feature Selection and Optimized Estimation

We applied the previously introduced customized loss function to the LSTM model with the 6-hour interval median resolution to reduce intraparticipant deviation in estimated age. The model was trained from scratch and achieved an MAE of 3.60. To evaluate the effectiveness of the customized loss function, we computed the SD of estimated ages across weeks for each participant and then averaged these values across all participants. The average SD decreased from 1.28 (SD 0.22; without the loss function) to 1.18 (SD 0.14; with the loss function), indicating a 7.8% reduction in intraparticipant variability.

We also applied Shapley Additive Explanations [23] to assess feature importance in the model trained with the customized loss function. As illustrated in Figure 4, the top 10 most influential features out of 60 were predominantly related to keystroke speed and rhythm, such as AllKeyRate, dur_shift, and HT, measured across various 6-hour time periods.

Figure 4. Top 10 most influential features for age estimation, identified via SHAP analysis of the long short-term memory model trained with the customized loss function and 6-hourly-median resolution. HT: hold time; PPL: press-to-press latency; SHAP: Shapley Additive Explanations.



Discussion

Principal Results

This study aimed to investigate whether age-related differences manifest in everyday smartphone keystroke dynamics. Through both descriptive and AI-based analyses, we demonstrated that typing behavior—such as speed, latency, and session frequency—systematically varies across age groups. For instance, younger participants exhibited consistently faster and more frequent typing patterns, particularly during late evening hours, while older individuals showed slower and less variable patterns throughout the day.

Building on these observations, we explored whether such behavioral signals could support accurate AI-driven age estimation. Among the 8 models tested, the LSTM model with the 6-hour interval median resolution achieved the best performance (MAE=3.69, R^2=0.71), highlighting that fine-grained temporal patterns carry meaningful age-related information. While predicted ages that differ from actual ages by only a small margin (eg, within ±5 y) may reflect normal variability, sudden large deviations (eg, 10-20 y) could arise from short-term factors, such as stress, fatigue, or sleep deprivation, as well as from long-term or chronic factors, such as disease onset or persistent cognitive or motor decline. Previous studies have shown that short-term conditions like fatigue or stress can alter keystroke dynamics, increasing interkey latencies and error rates [24,25]. Similarly, chronic neurological or cognitive disorders, including Parkinson disease and mild cognitive impairment, have been associated with measurable changes in typing patterns [26,27]. The magnitude of fluctuation in age predictions is therefore likely to depend on both the timescale and the underlying cause of behavioral change. Future work should not only disentangle short- and long-term contributors but also investigate how the scale of predicted age deviation corresponds to these different contributors. Such insights could enable adaptive age-prediction strategies that dynamically adjust to the likely source of variability, improving both robustness and interpretability in real-world conditions.

Ablation studies further revealed that typing speed features contributed most strongly to model performance, aligning with the behavioral differences observed in the descriptive analysis. Notably, our time-of-day ablation study (Table 5) revealed that the early morning (midnight-6 AM) and late evening (6 PM-midnight) periods contained more informative signals for age prediction. These findings suggest that age-related behavioral differences may be more pronounced during off-peak hours, potentially due to variations in daily routine structures, fatigue, or circadian rhythms.

Taken together, our results indicate that keystroke dynamics encode subtle, age-sensitive behavioral patterns that can be

XSL•FO
RenderX

leveraged for passive, AI-based age estimation—without requiring any explicit input from users.

Comparison With Previous Work

In total, 2 studies are particularly relevant for contextualizing our contributions, that is, studies by Vesel et al [13] and Zulueta et al [18].

Vesel et al [13] applied hierarchical mixed-effects models to examine how mood, age, and diurnal patterns influence keystroke behavior. They found that younger users (20 y) typed substantially faster, less variably, and paused less frequently compared with older users (≥70 y), with effect sizes ranging from 57% to 62%. Importantly, they also identified a significant interaction with diurnal patterns, such that older adults displayed greater slowing and variability in the later hours of the day. Consistent with these findings, our study also observed similar age-related differences in typing speed (Figure 3). However, Vesel et al [13] did not focus on prediction but rather on identifying how typing behavior varies as a function of demographic and clinical factors. They treated age as an independent variable affecting keystroke features, whereas in our study, age was the target variable to be predicted. Moreover, while Vesel et al [13] identified diurnal trends in typing speed and variability, our model leveraged this information to improve prediction through time-resolved feature representation and ablation experiments across 6-hour periods. In this sense, our study operationalizes the temporal behavioral patterns identified by Vesel et al [13] into an actionable, AI-driven predictive framework.

Zulueta et al [18], on the other hand, estimated "brain age" from mobile keystroke kinematics and examined the deviation between predicted and actual age in relation to bipolar disorder risk. They tokenized free-living iOS (BiAffect) typed logs into sessions and then summarized kinematic features per participant. To estimate chronological age, they trained random forest regressors on a 75:25 train-validation split using 10-fold cross-validation with 3 repeats. On the held-out validation set, they reported root mean square errors of 9.5-9.7 years, Breiman pseudo R^2 of 0.42-0.44, and MAEs of 5.5-5.9 years. While they also performed age prediction using smartphone metadata, their goal was to assess brain age as a digital biomarker of psychiatric risk, particularly differentiating prediction error patterns across diagnostic groups (positive vs negative Mood Disorders Questionnaire screens). Their models were evaluated primarily using participant-level aggregates, whereas our study preserved temporal granularity by modeling week-by-week patterns within participants. We also proposed a novel loss function to explicitly reduce intraparticipant variability, which was not addressed in their work. Finally, our best-performing model achieved an MAE of 3.60, substantially lower than the 5.5-5.9 years reported by Zulueta et al [18], possibly indicating the advancement of our method.

In summary, while previous works have laid important theoretical foundations, our study offers a methodological advancement by combining fine-grained temporal feature engineering, intraparticipant variance penalization, and deep learning architectures for high-resolution, week-level age estimation.

Limitations

Despite promising findings, our study has several limitations. First, our dataset was heavily skewed toward younger participants who were younger than 30 years of age. Although we constructed age-balanced subsets to partially mitigate this imbalance, the initial sampling bias may still have influenced model training and evaluation. In particular, residual imbalance across the cross-validation folds could not be fully eliminated, which may have led to performance estimates biased toward younger age groups. Future studies should validate our findings on larger and more demographically diverse populations, including a broader distribution of age groups.

Second, all participants in this study were healthy individuals without known motor or cognitive impairments. While this design choice helped isolate age as a primary factor, it limits the generalizability of our findings to clinical populations. Since keystroke dynamics are known to be affected by neurological and psychiatric conditions, future work should explore how such models perform across varying health statuses.

Third, our study population consisted exclusively of Korean Android users, which may limit the generalizability of our findings. Cultural factors (eg, language-specific typing conventions), linguistic aspects (such as Korean vs alphabet-based input), and operating system–specific characteristics (eg, Android vs iOS) could all affect keystroke dynamics. Furthermore, our study did not consider keystroke dynamics on desktop or other mobile devices (eg, tablet), which may also limit generalizability. To strengthen external validity, future studies should validate the model across multiple platforms and operating systems by including both Android and iOS users, expand recruitment to participants from different countries and linguistic backgrounds, and incorporate broader age groups, including younger adolescents and older adults. Such efforts would allow more comprehensive evaluation of the robustness and applicability of keystroke-based age prediction models.

Fourth, another limitation is that individual and device-specific factors, such as the dominant hand or screen size, were not considered in this study. These factors may systematically influence typing speed, rhythm, and error patterns, thereby introducing variability that is independent of age-related effects. For instance, larger screens may allow for more stable keystrokes, while different input languages and keyboard layouts could alter interkey intervals. Although such factors may add noise to the prediction, they are unlikely to fully explain the observed associations, yet future work should explicitly account for them to improve model validity and generalizability.

Fifth, our study did not evaluate the long-term temporal stability of the proposed models. Although we used multiweek data for training and testing, we did not assess whether age predictions remain consistent over several months. This is especially important for real-world applications where behavioral patterns may shift gradually due to lifestyle changes, seasonality, or device replacement.

XSL•FO
RenderX

Sixth, although our dataset volume was large, deeper statistical analysis revealed that the keystroke features did not follow Gaussian distributions. Specifically, we conducted Shapiro-Wilk normality tests across all features, and none satisfied the assumption of normality (all P values <.001). As a result, Gaussian-based methods, such as linear or quadratic discriminant analysis, would not have been appropriate for this dataset. This supports our choice of nonlinear AI-based approaches, which are better suited to capture the inherently non-Gaussian nature of keystroke dynamics.

Finally, while our analysis of temporal variance features (speed variability and frequency variability) did not lead to meaningful performance gains, we believe this area holds potential. The human body follows natural diurnal rhythms that affect motor and cognitive function across the day. With more refined feature engineering, temporal variability may better capture such rhythms and contribute to improved age estimation in future models.

Conclusions

This study demonstrates that everyday smartphone keystroke dynamics contain measurable behavioral patterns that vary systematically across age groups. By analyzing typing speed, rhythm, and frequency features across different time windows, we showed that younger and older individuals exhibit distinct temporal keystroke signatures—with our results suggesting that early morning and late evening periods may carry more predictive information for age estimation. These findings suggest that keystroke dynamics may serve as meaningful digital behavioral markers of age-related functional differences.

Building on these behavioral insights, we developed AI models to estimate age using passive keystroke data, with the LSTM model and the 6-hour interval median resolution achieving the best performance (MAE=3.69, R^2=0.71). We further introduced a customized loss function that reduced intraparticipant variability, and used Shapley Additive Explanations analysis to identify key typing features contributing to model predictions.

Importantly, this work is not merely an age estimation exercise. Rather, it illustrates how fine-grained, temporally structured passive behavioral data can be used to detect subtle functional differences across individuals. These results suggest potential future applications of mobile keystroke dynamics in digital health, such as aiding the early detection of cognitive or motor decline and supporting age-sensitive personalization of interventions. However, given that our study included only healthy participants, the clinical applicability of these approaches requires further validation in clinical populations.

Acknowledgments

The authors sincerely thank all study participants for their valuable time and commitment to this research.

Data Availability

Data sharing is not applicable to this article, as the data cannot be shared publicly due to restrictions imposed by the ethics approval obtained from the institutional review board.

Funding

This work was supported in part by the National Research Foundation of Korea grant funded by the Korea government (Ministry of Science and ICT; NRF2023R1A2C2003577) and the Advanced Science & Technology Research Agency (ASTRA) Project through the National Research Foundation funded by the Ministry of Science and ICT (RS-2024-00470056).

Authors' Contributions

Conceptualization: JM, BHC
Data acquisition: YLH, HYC, CK, HL, DPJ
Data analysis: JM, YLH, HYC, CK, HL, DPJ, BHC
Funding acquisition: BHC
Investigation: JM, YLH, HYC, CK, HL, DPJ
Project administration: BHC
Revising manuscript: BHC
Supervision: BHC
Visualization: JM, YLH, HYC, CK, HL, DPJ
Writing and revising manuscript: JM, YLH, HYC

Conflicts of Interest

None declared.

References

1. Savvidou F, Tegos SA, Diamantoulakis PD, Karagiannidis GK. Passive radar sensing for human activity recognition: a survey. IEEE Open J Eng Med Biol 2024;5:700-706. [doi: 10.1109/OJEMB.2024.3420747] [Medline: 39184964]

XSL•FO
RenderX

2. Adler DA, Ben-Zeev D, Tseng VW, Kane JM, Brian R, Campbell AT, et al. Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks. JMIR Mhealth Uhealth 2020;8(8):e19962 [FREE Full text] [doi: 10.2196/19962] [Medline: 32865506]
3. Nguyen DT, Zeng Q, Tian X, Chia P, Wu C, Liu Y, et al. Ambient health sensing on passive surfaces using metamaterials. Sci Adv 2024;10(1):eadj6613 [FREE Full text] [doi: 10.1126/sciadv.adj6613] [Medline: 38181071]
4. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and Fitbit data. JMIR Mhealth Uhealth 2019;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]
5. Faurholt-Jepsen M, Busk J, Rohani DA, Frost M, Tønning ML, Bardram JE, et al. Differences in mobility patterns according to machine learning models in patients with bipolar disorder and patients with unipolar disorder. J Affect Disord 2022;306:246-253. [doi: 10.1016/j.jad.2022.03.054] [Medline: 35339568]
6. Palmius N, Tsanas A, Saunders KEA, Bilderbeck AC, Geddes JR, Goodwin GM, et al. Detecting bipolar depression from geographic location data. IEEE Trans Biomed Eng 2017;64(8):1761-1771 [FREE Full text] [doi: 10.1109/TBME.2016.2611862] [Medline: 28113247]
7. Halimeh M, Jackson M, Vieluf S, Loddenkemper T, Meisel C. Explainable AI for wearable seizure logging: impact of data quality, patient age, and antiseizure medication on performance. Seizure 2023;110:99-108 [FREE Full text] [doi: 10.1016/j.seizure.2023.06.002] [Medline: 37336056]
8. Vieluf S, Hasija T, Kuschel M, Reinsberger C, Loddenkemper T. Developing a deep canonical correlation-based technique for seizure prediction. Expert Systems with Applications 2023;234:120986. [doi: 10.1016/j.eswa.2023.120986]
9. Schneider J, Bakštein E, Kolenič M, Vostatek P, Correll CU, Novák D, et al. Motor activity patterns can distinguish between interepisode bipolar disorder patients and healthy controls. CNS Spectr 2022;27(1):82-92. [doi: 10.1017/S1092852920001777] [Medline: 32883376]
10. Wu X, Ma L, Wei P, Shan Y, Chan P, Wang K, et al. Wearable sensor devices can automatically identify the ON-OFF status of patients with Parkinson's disease through an interpretable machine learning model. Front Neurol 2024;15:1387477 [FREE Full text] [doi: 10.3389/fneur.2024.1387477] [Medline: 38751881]
11. Austin D, Jimison H, Hayes T, Mattek N, Kaye J, Pavel M. Measuring motor speed through typing: a surrogate for the finger tapping test. Behav Res Methods 2011;43(4):903-909 [FREE Full text] [doi: 10.3758/s13428-011-0100-1] [Medline: 21494919]
12. Giancardo L, Sánchez-Ferro A, Butterworth I, Mendoza CS, Hooker JM. Psychomotor impairment detection via finger interactions with a computer keyboard during natural typing. Sci Rep 2015;5(1):9678 [FREE Full text] [doi: 10.1038/srep09678] [Medline: 25882641]
13. Vesel C, Rashidisabet H, Zulueta J, Stange JP, Duffecy J, Hussain F, et al. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: a BiAffect iOS study. J Am Med Inform Assoc 2020;27(7):1007-1018 [FREE Full text] [doi: 10.1093/jamia/ocaa057] [Medline: 32467973]
14. Park JH. Discriminant power of smartphone-derived keystroke dynamics for mild cognitive impairment compared to a neuropsychological screening test: cross-sectional study. J Med Internet Res 2024;26:e59247 [FREE Full text] [doi: 10.2196/59247] [Medline: 39475819]
15. Hoeijmakers A, Licitra G, Meijer K, Lam KH, Molenaar P, Strijbis E, et al. Disease severity classification using passively collected smartphone-based keystroke dynamics within multiple sclerosis. Sci Rep 2023;13(1):1871 [FREE Full text] [doi: 10.1038/s41598-023-28990-6] [Medline: 36725975]
16. Lam KH, Twose J, Lissenberg-Witte B, Licitra G, Meijer K, Uitdehaag B, et al. The use of smartphone keystroke dynamics to passively monitor upper limb and cognitive function in multiple sclerosis: longitudinal analysis. J Med Internet Res 2022;24(11):e37614 [FREE Full text] [doi: 10.2196/37614] [Medline: 36342763]
17. Lim S, Kim C, Cho BH, Choi SH, Lee H, Jang DP. Investigation of daily patterns for smartphone keystroke dynamics based on loneliness and social isolation. Biomed Eng Lett 2024;14(2):235-243 [FREE Full text] [doi: 10.1007/s13534-023-00337-0] [Medline: 38374905]
18. Zulueta J, Demos AP, Vesel C, Ross M, Piscitello A, Hussain F, et al. The effects of bipolar disorder risk on a mobile phone keystroke dynamics based biomarker of brain age. Front Psychiatry 2021;12:739022 [FREE Full text] [doi: 10.3389/fpsyt.2021.739022] [Medline: 35002792]
19. Ajilore O, Bark JS, Demos AP, Zulueta J, Stange J, Duffecy J, et al. Assessment of cognitive function in bipolar disorder with passive smartphone keystroke metadata: a BiAffect digital phenotyping study. Front Psychiatry 2025;16:1430303 [FREE Full text] [doi: 10.3389/fpsyt.2025.1430303] [Medline: 40115651]
20. Dhakal V, Feit AM, Kristensson PO, Oulasvirta A. Observations on typing from 136 million keystrokes. United States: ACM; 2018 Presented at: Proceedings of the 2018 CHI conference on human factors in computing systems; 2018 Apr 21; New York, NY, United States. [doi: 10.1145/3173574.3174220]
21. Fagiani F, Marino DD, Romagnoli A, Travelli C, Voltan D, Mannelli LDC, et al. Molecular regulations of circadian rhythm and implications for physiology and diseases. Signal Transduct Target Ther 2022;7(1):41 [FREE Full text] [doi: 10.1038/s41392-022-00899-y] [Medline: 35136018]

XSL•FO
RenderX

22. Arik S, Pfister T. Tabnet: Attentive interpretable tabular learning. In: AAAI.: AAAI Press; 2021 Presented at: Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence; 2021 Feb 02; Palo Alto, CA, USA. [doi: 10.1609/aaai.v35i8.16826]
23. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. USA: Curran Associates Inc; 2017 Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 04; Long Beach, CA, USA. Red Hook, NY, USA URL: https://dl.acm.org/doi/10.5555/3295222.3295230 [doi: 10.5555/3295222.3295230]
24. de Jong M, Bonvanie AM, Jolij J, Lorist MM. Dynamics in typewriting performance reflect mental fatigue during real-life office work. PLoS One 2020;15(10):e0239984 [FREE Full text] [doi: 10.1371/journal.pone.0239984] [Medline: 33022017]
25. Acien A, Morales A, Vera-Rodriguez R, Fierrez J, Mondesire-Crump I, Arroyo-Gallego T. Detection of mental fatigue in the general population: feasibility study of keystroke dynamics as a real-world biomarker. JMIR Biomed Eng 2022;7(2):e41003 [FREE Full text] [doi: 10.2196/41003] [Medline: 38875698]
26. Nguyen TM, Leow AD, Ajilore O. A review on smartphone keystroke dynamics as a digital biomarker for understanding neurocognitive functioning. Brain Sci 2023;13(6):959 [FREE Full text] [doi: 10.3390/brainsci13060959] [Medline: 37371437]
27. Alfalahi H, Khandoker AH, Chowdhury N, Iakovakis D, Dias SB, Chaudhuri KR, et al. Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis. Sci Rep 2022;12(1):7690 [FREE Full text] [doi: 10.1038/s41598-022-11865-7] [Medline: 35546606]

Abbreviations

AI: artificial intelligence
HT: hold time
LSTM: long short-term memory
MAE: mean absolute error
MLP: multilayer perceptron
PPL: press-to-press latency
XGBoost: extreme gradient boosting

Edited by L Buis; submitted 08.07.25; peer-reviewed by D Progonov, JH Park; comments to author 11.08.25; accepted 14.11.25; published 02.12.25.

Please cite as:
Moon J, Huh YL, Cho HY, Kim C, Lee H, Jang DP, Cho BH
Exploring Age-Related Patterns in Smartphone Keystroke Dynamics Considering Temporal Variability: Cross-Sectional Study With AI-Based Analysis
JMIR Mhealth Uhealth 2025;13:e80094
URL: https://mhealth.jmir.org/2025/1/e80094
doi:10.2196/80094
PMID:

©Junhyung Moon, Yu Lim Huh, Hee Young Cho, Chaeyeon Kim, Hyeongrae Lee, Dong Pyo Jang, Baek Hwan Cho. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 02.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Bidirectional Associations Between Physical Activity, Sedentary Behavior, and Daily Symptoms in Patients With Chronic Obstructive Pulmonary Disease: Longitudinal Observational Study

Banchia Palmen[1,2]; Zjala Ebadi[3,4]; Maarten van Herck[2]; Yvonne M J Goërtz[5,6]; Qichen Deng[2]; Melissa S Y Thong[7,8]; Chris Burtin[9]; Jeannette B Peters[4]; Roy T M Sprooten[10]; Erik W M A Bischoff[11]; Emiel F M Wouters[2,10]; Mirjam A G Sprangers[7]; Jan H Vercoulen[3,4]; Sarah Houben-Wilke[2]; Anouk W Vaes[2]; Daisy J A Janssen[12,13]; Martijn A Spruit[2,14]

[1]Institute of Nutrition and Translational Research in Metabolism, Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

[10]Department of Respiratory Medicine, Maastricht University Medical Center, Maastricht, The Netherlands

[11]Department of General Practice, Erasmus MC, Rotterdam, The Netherlands

[12]Department of Health Services Research and Department of Family Medicine, Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

[13]Department of Expertise and Treatment, Proteion, Horn, The Netherlands

[14]Department of Respiratory Medicine, Institute of Nutrition and Translational Research in Metabolism, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

[2]Department of Research and Development, Ciro Centre of Expertise for Chronic Organ Failure, Hornerheide 1, Horn, The Netherlands

[3]Department of Medical Psychology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

[4]Department of Pulmonary Diseases, Radboud University Medical Centre, Nijmegen, The Netherlands

[5]Living Lab in Ageing and Long-Term Care, Maastricht University, Maastricht, The Netherlands

[6]Department of Health Services Research, Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

[7]Amsterdam UMC location University of Amsterdam, Medical Psychology, Amsterdam, The Netherlands

[8]Unit of Cancer Survivorship, Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

[9]REVAL - Rehabilitation Research Center, BIOMED - Biomedical Research Institute, Faculty of Rehabilitation Sciences, Hasselt University, Diepenbeek, Belgium

Corresponding Author:
Banchia Palmen
Institute of Nutrition and Translational Research in Metabolism, Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

Abstract

Background: Questionnaire-based symptom assessment may introduce recall bias and lacks bidirectional exploration. This is particularly relevant, given the unclear direction of the associations between physical activity (PA), sedentary time (ST), and symptoms in patients with chronic obstructive pulmonary disease (COPD). Understanding these associations could inform symptom management strategies and improve patient quality of life.

Objective: This study aimed to investigate the direction of the association between PA, ST, and symptoms in patients with COPD using accelerometry and ecological momentary assessment (EMA).

Methods: A subsample from the FAntasTIGUE study answered 8 randomly timed EMA questionnaires daily for 5 days. Ten symptoms were rated on a 7-point Likert scale: “I feel relaxed, short of breath, energetic, cheerful, insecure, irritated, satisfied, anxious, tired, and mentally fit.” Concurrently, step count and ST were measured using the ActiGraph GT9X Link placed on the right hip. Step count and ST 15 and 30 minutes pre- and post-EMA were used in multilevel models, controlled for pre-EMA steps and ST, and the previous EMA score. Significant confounders were used as covariates, and patient ID was used as random intercept.

Results: Thirty-four patients (19/34, 56% men, mean age 66, SD 7 years; forced expiratory volume in 1 second 52±20% predicted; 1035 EMA responses) were included. Feeling more relaxed was associated with a higher step count 15 minutes post-EMA (β=5.1; 95% CI 0.9 to 10.1; *P*=.046). Conversely, higher step count 15 and 30 minutes pre-EMA was associated with feeling less relaxed (β=-5.2×10^{-4}; 95% CI -9.7×10^{-4} to -7.0×10^{-5}; *P*=.02; and β=-3.2×10^{-4}; 95% CI -5.6×10^{-4} to -7.9×10^{-5}; *P*=.009), more short of breath (β=8.5×10^{-4}; 95% CI 4.7×10^{-4} to 1.2×10^{-3}; *P*<.001; and β=4.6×10^{-4}; 95% CI 2.6×10^{-4} to 6.6×10^{-4};

XSL•FO
RenderX

P<.001), and tired (β=5.1×10^{-4}; 95% CI 7.2×10^{-5} to 9.4×10^{-4}; *P*=.02; and β=2.9×10^{-4}; 95% CI 5.3×10^{-5} to 5.2×10^{-4}; *P*=.02). Higher ST 15 and 30 minutes pre-EMA was associated with feeling more anxious (β=1.7×10^{-4}; 95% CI 1.7×10^{-5} to 3.2×10^{-4}; *P*=.03; and β=8.5×10^{-5}; 95% CI 2.5×10^{-6} to 1.7×10^{-4}; *P*=.04).

Conclusions: A bidirectional association of feeling relaxed with PA was found in patients with COPD. Higher step count was related to feeling more short of breath and tired, whereas higher ST was associated with heightened anxiety.

(JMIR Mhealth Uhealth 2025;13:e65653) doi:10.2196/65653

KEYWORDS

COPD; chronic obstructive pulmonary disease; ecological momentary assessment; accelerometry; physical activity; sedentary behavior; symptom assessment

Introduction

Chronic obstructive pulmonary disease (COPD) is a heterogeneous chronic lung condition characterized by shortness of breath, cough, and/or sputum production. This is, at least partly, due to abnormalities of the airways or alveoli that cause persistent airflow obstruction [1]. The heterogeneity of COPD manifests in variations in disease progression, symptom burden, and response to interventions, requiring personalized care [2]. In addition to respiratory symptoms, patients with COPD frequently experience physical symptoms such as fatigue, alongside psychological symptoms such as depressed mood and anxiety [3,4]. Moreover, they are physically less active and spend more time sitting and lying down compared to healthy controls, contributing to disease progression and poorer health outcomes [5-9].

Both the number of steps per day and sedentary time (ST) are linked to the transient exacerbation of symptoms in patients with COPD [7,10,11]. However, the direction of these associations is unclear [10-13]. For example, while physical activity (PA) may enhance feeling short of breath immediately, during, or after PA, it is also possible that worsening symptoms lead to a decline in PA and increased ST. This short-term bidirectional relationship suggests a potential vicious cycle of physical inactivity and symptom exacerbation. A deeper understanding of these associations is essential for developing targeted interventions aimed at interrupting this cycle and improving symptom management [7,8,14]. Increasing PA and reducing ST have been shown to improve symptoms and health-related quality of life in patients with COPD and may even improve survival [7,8,13,14].

Traditional symptom assessment methods typically involve clinical interviews and questionnaires asking about symptom burden 1 to 2 weeks retrospectively, which can be subject to recall bias [15,16]. This creates a challenge, considering that patients with COPD experience variability in PA, ST, and symptoms within the day and from day to day [17-19]. To address this limitation, ecological momentary assessment (EMA) emerges as a feasible method to assess symptom experience over time. EMA is defined as 'repeated sampling of people's current thoughts, emotions, behavior, physiological states, and context, in their natural environment, typically (but not necessarily) via electronic wearable devices' [15,16,20]. Compared to traditional assessment methods, EMA minimizes recall bias, and ecological validity is maximized because events are monitored as close in time to their occurrence as possible [20-24]. EMA has been validated across multiple study populations, but the use of EMA for capturing symptoms in patients with COPD is still limited, and further examination of its psychometric properties is needed [25-28]. It offers a less cumbersome and cost-effective alternative for assessing multiple symptoms, making it suitable for COPD research [21,22,29]. Combining information collected with EMA and from continuous PA tracking such as accelerometry may reveal the complex bidirectional relationship between PA, ST, and symptoms. The ActiGraph accelerometer (ActiGraph Corp, Pensacola, FL, USA) has been validated for measuring PA and ST in patients with COPD, demonstrating good reliability. Its use in COPD research has been supported because of its sensitivity in detecting variations in PA and ST, making it suitable for real-world monitoring of activity patterns. Limited research is available where the temporal associations between PA, ST, and symptoms in patients with COPD are explored using real-time assessment [29].

Therefore, the main objective of this study was to investigate the direction of the association between PA and ST measured with accelerometry and physical and psychological symptoms measured with EMA in patients with COPD. It is hypothesized that PA would be associated with subsequent increases in symptoms such as breathlessness and fatigue in the short term, while worsening symptoms would predict subsequent reductions in PA and increasing ST. In addition, we aimed to explore whether the strength of these associations differed across the 10 symptoms assessed in this study. By addressing these gaps in knowledge, this study contributes to a more nuanced understanding of how daily activity patterns and symptom fluctuations interact in COPD, with potential implications for personalized intervention strategies.

Methods

Design

These analyses were a part of a multicenter, longitudinal, observational study in patients with COPD investigating physical, systemic, psychological, and behavioral factors associated with precipitation and/or perpetuation of fatigue (FAntasTIGUE study) [30]. The prevalence of fatigue and its associated factors in patients with COPD participating in the FAntasTIGUE study has been published before [3,31].

XSL•FO
RenderX

Ethical Considerations

The study was approved by the Medical Research Ethics Committee United, Nieuwegein, The Netherlands (R17.036/NL60484.100.17) and was registered in the Dutch Trial Register (NTR6933). Written informed consent was obtained from all study participants. Participation was entirely voluntary, and participants could withdraw at any time without consequence. Participants did not receive any financial compensation. Data were deidentified before analysis.

Participants

Patients were recruited between 2018 and 2021 via the pulmonary consultation at the outpatient clinics of the Department of Respiratory Medicine in Maastricht and the Department of Pulmonary Diseases of the Radboud University Medical Center in Nijmegen and via general practitioners (Research Network Family Medicine Maastricht [32] and the Academic General Practitioner Network in the Nijmegen region). In addition, patients with COPD who attended a meeting for patients with chronic lung diseases (cfr. *longpunt*) in Maastricht or Nijmegen and patients recruited from primary and secondary care that previously participated in the Chance Study (NTR3416) and had indicated on the informed consent form that they agreed to be approached for follow-up research were invited to participate in the FAntasTIGUE study. Patients had to meet the following criteria for inclusion: (1) a diagnosis of COPD according to the Global Strategy for the Diagnosis, Management and Prevention of COPD (Global Initiative for Chronic Obstructive Lung Disease [GOLD], grade 1A–4D), (2) no exacerbation-related hospitalization less than 4 weeks preceding enrollment, (3) no use of oral corticosteroids and/or antibiotics less than 4 weeks preceding enrollment, and (4) provided written informed consent.

A subsample of the first 60 patients who were interested in participating in the EMA study had to meet these extra inclusion criteria: (1) access to the internet at home (Wi-Fi) and (2) able to operate an iPod.

Patients lacking sufficient understanding of the Dutch language and/or participating in concurrent intervention studies were excluded.

Measurements

During the baseline assessment, sociodemographic, physical, psychological, and behavioral characteristics (eg, age, sex, BMI, smoking status, marital status, working status, level of education [The Netherlands; low: primary school only and lower vocational education, medium: secondary vocational education, and high: higher vocational education and academic education], Charlson Comorbidity Index [CCI], forced expiratory volume in 1 second [FEV_1] as percentage from the predicted volume, the forced expiratory volume in 1 second to forced vital capacity ratio, GOLD stage based on forced expiratory volume in 1 second % predicted [score of 1-4], occurrence of an exacerbation and exacerbation-related hospitalization in the previous 12 months [33], and the Modified Medical Research Council [mMRC dyspnea scale]) were assessed as described elsewhere [34]. The mMRC dyspnea scale assesses the current severity of dyspnea on a scale of 5 grades (0-4), each describing a level of activity that induces dyspnea, ranging from 0 (“only gets breathless with strenuous exercise”) to 4 (“too breathless to leave the house or breathless when dressing”) [34]. An mMRC score of ≥2 indicates severe dyspnea [1]. The Checklist Individual Strength—subscale subjective fatigue (CIS-Fatigue) was used in which a score of ≥36 points denoted severe fatigue [35]. Anxiety and depression symptoms were assessed using the Hospital Anxiety and Depression Scale (HADS), where a score of >11 points indicates probable symptoms of anxiety/depression [36].

EMA Symptom Assessment

Electronic EMA questionnaires were completed for 5 consecutive days through an iPod Touch (Apple) with a custom-installed app called PsyMate (version 1.1b). During the baseline assessment, patients received oral instructions on how to use the iPod Touch and were given a printed instruction to take home. Patients were told to carry the iPod Touch with them at all times and keep to their usual day/night routine. Patients had to answer the questions immediately after the beep but no longer than 15 minutes after the beep; otherwise, the questionnaire would be skipped. PsyMate was programmed to beep at 8 random time points between 7:30 AM and 10:30 PM. When receiving the beep, patients had to rate 10 symptoms with a 7-point Likert scale, ranging from 1 (“barely”) to 7 (“very”). A score of 1 - to 2 is classified as ‘not’, 3 - to 5 is classified as ‘somewhat’, and 6 - 7 is classified as ‘very’. These symptoms included ‘I feel: relaxed; short of breath; energetic; cheerful; insecure; irritated; satisfied; anxious; tired; and mentally fit.’ On the second day, patients were called to ensure everything was going well and to discuss if any problems occurred.

Physical Activity

During the EMA home monitoring period, patients were asked to wear the ActiGraph GT9X Link 3-axis activity monitor with a sample frequency of 30 Hz for 7 consecutive days. Patients were instructed to wear the ActiGraph on the right hip from the moment they woke up until the moment they went to sleep, keep a normal day/night routine, and remove the ActiGraph if they had to shower or went swimming. During the baseline assessment, the ActiGraph was demonstrated, and printed instructions were provided for patients to take home. Data were recorded in 10-second epochs, and patients’ steps were used to indicate the amount of PA because this is a simple metric and captures the most relevant and problematic daily activity for the majority of patients with COPD [37].

Sedentary Time

Freedson Adult (1998) activity counts were used to categorize ST. ST was tracked during waking hours and was defined as a period in minutes where the activity counts fell into the sedentary range (0-99) with no interruption [38]. For valid PA data, the wear time per day needed to be a minimum of 8 hours with a minimum of 4 days. These 4 days needed to include at least one weekend day and 3 weekdays [37,39].

Data Cleaning

If patients completed less than one-third of the total beeps (ie, 13 of 40 beeps), these patients were excluded using R statistical software [40]. If the time between the first and last answer on

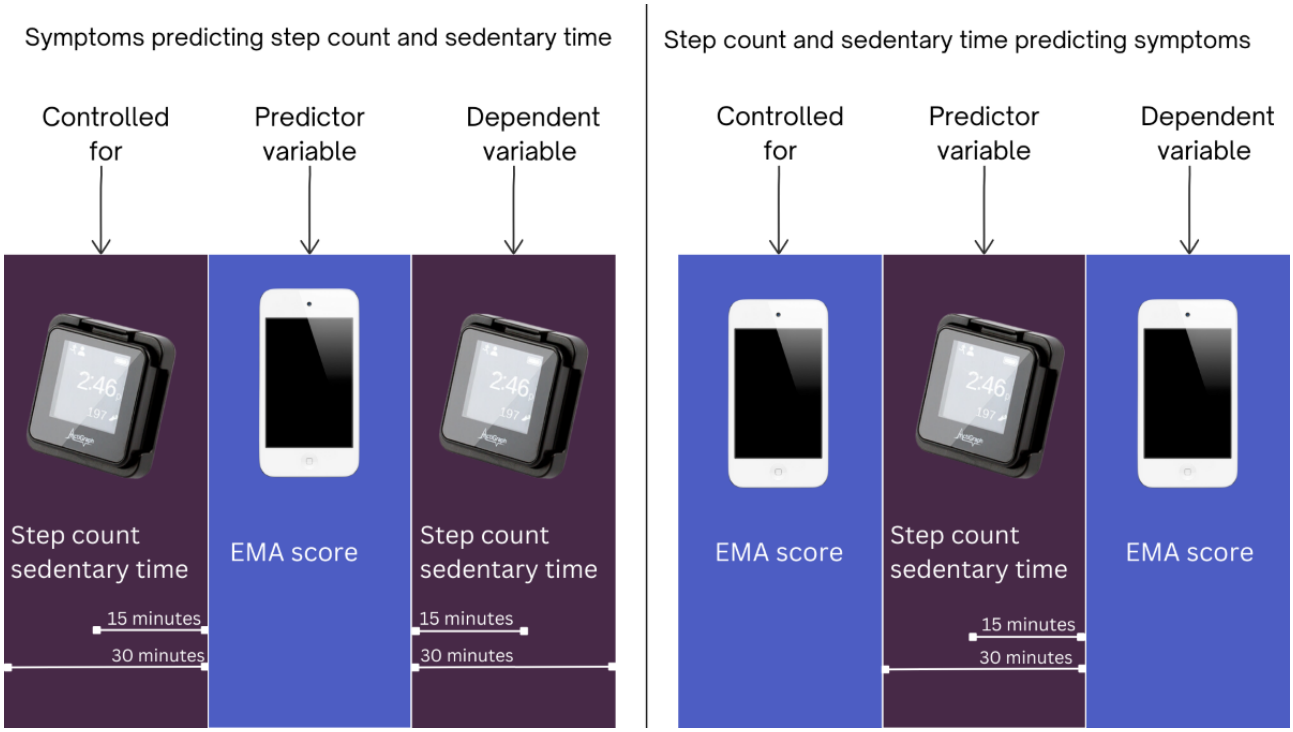
XSL•FO
RenderX

the EMA questionnaire was longer than 10 minutes, this was considered invalid, and the EMA datapoints were excluded. ActiLife6 software was used to check whether PA and ST data of the same 5 EMA home monitoring days were valid, and this was manually aligned [41]. For all the patients from Nijmegen, the ActiGraph started measuring at 9 AM on the first day because of incorrect initialization, but these patients still have the minimum of 8 hours of wear time on this day. To ensure data validity, ActiGraph wear time was verified 15 and 30 minutes before and after each EMA prompt. The EMA datapoint was excluded using MATLAB if the ActiGraph was not worn during the relevant time windows.

Data Analysis

All analyses were performed using SPSS (IBM SPSS Statistics for Windows, version 28.0.1.1; IBM Corp, Armonk, NY, USA). The distribution of data was tested using the Shapiro-Wilk test with a 2-tailed significance of <.05. Descriptive statistics were reported as means and SDs, median and quartiles 1 and 3 (Q1-Q3), percentages, and numbers as appropriate. To examine the direction of the association between steps, ST, and symptoms assessed with EMA, 15- and 30-minute time windows were created before giving the first answer on the EMA questionnaire (pre-EMA) and after giving the last answer (post-EMA), which is in line with previous research [26-28]. The total number of steps and ST within each time window was captured. Data from 11 patients were collected during the COVID-19 pandemic, and patients were measured during all 4 seasons throughout the year, which could potentially influence mood and alter PA and ST patterns. Therefore, based on normality, an independent samples *t* test or Mann-Whitney *U* test was done to compare the average amount of steps and ST during the 5 days between patients measured during COVID-19 and not during COVID-19. Furthermore, as appropriate, a one-way ANOVA or an independent Kruskal-Wallis test was done to compare the average number of steps and ST during the 5 days between the 4 seasons. A priori, a Spearman correlation analysis was run to screen for potential confounders. This was done to determine the relationship between the total number of steps and total ST of the 5 days and between EMA symptom scores and several characteristics, including age, sex, BMI, smoking status, marital status, working status, CCI, GOLD stage, exacerbation history, and exacerbation-related hospitalization in the previous 12 months [8,9,42]. The correlation between time of day with steps, ST, and EMA symptom scores was also tested because previous research showed associations between PA, ST, and time of day [18,43]. Confounders that had a significant correlation (P<.05) were used as covariates in the multilevel models. Patient ID was included as a random intercept, specifying that all observations by the same patient would be correlated with each other. Multilevel linear mixed models were performed using the ‘mixed’ command and restricted maximum likelihood estimations with a significance of <.05. While we hypothesized associations specifically for dyspnea and fatigue based on prior literature and clinical relevance, analyses involving the remaining symptoms were exploratory in nature. These additional analyses aimed to identify potential patterns in the data and generate hypotheses for future research. In multilevel model 1, step count 15 and 30 minutes post-EMA was used as the dependent variable, and the EMA score was used as the predictor (Figure 1). This was controlled for the step count 15 and 30 minutes pre-EMA.

Figure 1. Illustration of the multilevel model methodology. EMA: ecological momentary assessment.



In multilevel model 2, the EMA score was used as the dependent variable, and the step count 15 and 30 minutes pre-EMA was used as predictor. This was controlled for the previous EMA score on the same day, so the first EMA score of a day was not corrected for a previous EMA score. In multilevel model 3, ST 15 and 30 minutes post-EMA was used as the dependent

XSL•FO
RenderX

variable, and the EMA score was used as the predictor. This was controlled for ST 15 and 30 minutes pre-EMA. In multilevel model 4, the EMA score was used as the dependent variable, and the ST 15 and 30 minutes pre-EMA was used as predictor. This was controlled for the previous EMA score. All 10 symptoms were tested in separate multilevel models.

Results

Overview

A total of 60 patients were assessed for eligibility. However, 26 (43%) patients were excluded because of the absence of valid ActiGraph data (n=23) or valid EMA data (n=3). Specifically, a technical error and incorrect initialization of the ActiGraph resulted in 10 (43%) patients lacking any ActiGraph data. In addition, another 10 (43%) patients had ActiGraph data, but they were not valid since seven of these patients had only 1 day of PA measurement, and 3 patients had just 1 day of PA measurement but with less than 4 hours of valid recording. In addition, 3 (13%) patients had a mismatch between the PA and EMA measurement days. Two (67%) patients did not have any EMA data available, and 1 (33%) patient had less than one-third of answered EMA responses. Consequently, 34 patients were included for this analysis.

Participants

As a lot of patients were excluded, the sociodemographic, physical, psychological, and behavioral characteristics of the included and excluded patients were compared. There were no significant differences in any of these characteristics between the included and excluded patients (Table 1). Of the 34 included patients, 19 were men (56%), 32 (94%) had moderate-to-severe COPD, and almost half had at least one exacerbation in the previous 12 months. Severe dyspnea was reported by 13 of 33 patients (39%), and 15 of 34 patients (44%) experienced severe fatigue. Among the patients, 3 of 33 (9%) and 5 of 33 (15%) had probable symptoms of anxiety and depression, respectively. No statistically significant impact of the COVID-19 pandemic or seasonality on both steps and ST was found (all $P>.05$).

XSL•FO
RenderX

Table . Baseline characteristics of the included and excluded patients.

Variable[a]	Included (n=34)	N	Excluded (n=26)	N	*P* value
Age (years), mean (SD)	66 (7)	34	69 (6)	26	.19
Male, n (%)	19 (56)	34	19 (73)	26	.17
BMI[b] (kg/m^2), median (interquartile range)	24.8 (22.3 - 31.5)	34	25.7 (23.7 - 28.2)	26	.71
Current smoker, n (%)	4 (12)	34	6 (23)	26	.18
Married/living together, n (%)	21 (62)	34	18 (69)	26	.89
Currently employed, n (%)	6 (18)	33	1 (4)	24	.11
Level of education, n (%)		34		24	.43
Low	6 (18)		5 (21)		
Medium	15 (44)		9 (38)		
High	13 (38)		10 (42)		
CCI[c] (points), median (interquartile range)	1 (1-2)	34	2 (1-3)	26	.17
FEV_1[d] (% predicted), mean (SD)	52 (20)	34	51 (19)	26	.83
FEV_1/FVC[e] (%), median (interquartile range)	48 (34 - 59)	34	46 (37 - 54)	26	.72
GOLD[f] I/II/III/IV (%)	6/53/29/12	34	8/31/46/15	26	.39
Exacerbation in previous 12 months, n (%)	16 (47)	34	11 (42)	26	.71
Exacerbation-related hospitalization in previous 12 months, n (%)	4 (12)	34	6 (23)	26	.24
mMRC[g] dyspnea (grade), median (interquartile range)	1 (0 - 2)	33	1 (0 - 2)	24	.73
Severe dyspnea, n (%)	13 (39)		11 (46)		
CIS-Fatigue[h] (points), mean (SD)	34 (15)	34	35 (11)	24	.77
Severe fatigue, n (%)	15 (44)		12 (50)		
HADS-anxiety[i] (points), median (interquartile range)	4 (1-7)	33	7 (3-9)	24	.06
Possible symptoms of anxiety (8-10 points), n (%)	4 (12)		6 (25)		
Probable symptoms of anxiety (≥11 points), n (%)	3 (9)		3 (13)		
HADS-depression[i] (points), median (interquartile range)	4 (1-9)	33	4 (2-7)	24	.99

XSL•FO
RenderX

Variable[a]	Included (n=34)	N	Excluded (n=26)	N	*P* value
Possible symptoms of depression (8-10 points), n (%)	4 (12)		4 (17)		
Probable symptoms of depression (≥11 points), n (%)	5 (15)		1 (4)		

[a]Quantitative variables are presented as mean (standard deviation) or median (quartile 1 - quartile 3) for skewed variables.

[b]BMI: Body Mass Index.

[c]CCI: Charlson Comorbidity Index.

[d]FEV1 % predicted: forced expiratory volume in 1 second as percentage from the predicted volume.

[e]FEV1/FVC %: the forced expiratory volume in 1 second to forced vital capacity ratio.

[f]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[g]mMRC: modified Medical Research Council dyspnea Scale.

[h]CIS: Checklist Individual Strength.

[i]HADS: Hospital Anxiety and Depression Scale.

EMA Symptom Assessment

Of the 1054 EMA datapoints, 5 were excluded because the time between the first and last answer of the EMA questionnaire was longer than 10 minutes, and 1 was excluded because it was sent after 10:30 PM. In addition, 13 EMA datapoints were excluded because patients received the first EMA prompt before 9 AM, whereas the ActiGraph started measuring at 9 AM on the first day, so no ActiGraph data were available. Consequently, 1035 EMA responses were analyzed, resulting in a response rate of 76.1% (1035 completed EMA responses of the 1360 possible EMA responses). The mean response rates of the 34 patients per day were 196 of 272 (72.1%), 221 of 272 (81.3%), 213 of 272 (78.3%), 198 of 272 (72.8%), and 207 of 272 (76.1%). The lowest number of total responses during the 5 days was 18 of 40 (n=1), and the highest number of total responses was 38 of 40 (n=3; Table 2;).

The median (Q1-Q3) time of completing the EMA was 1 minute and 1 second (00:48-01:22). During the 5 days, patients generally rated feeling very relaxed, cheerful, and satisfied, while not feeling insecure, irritated, and anxious (Figure 2).

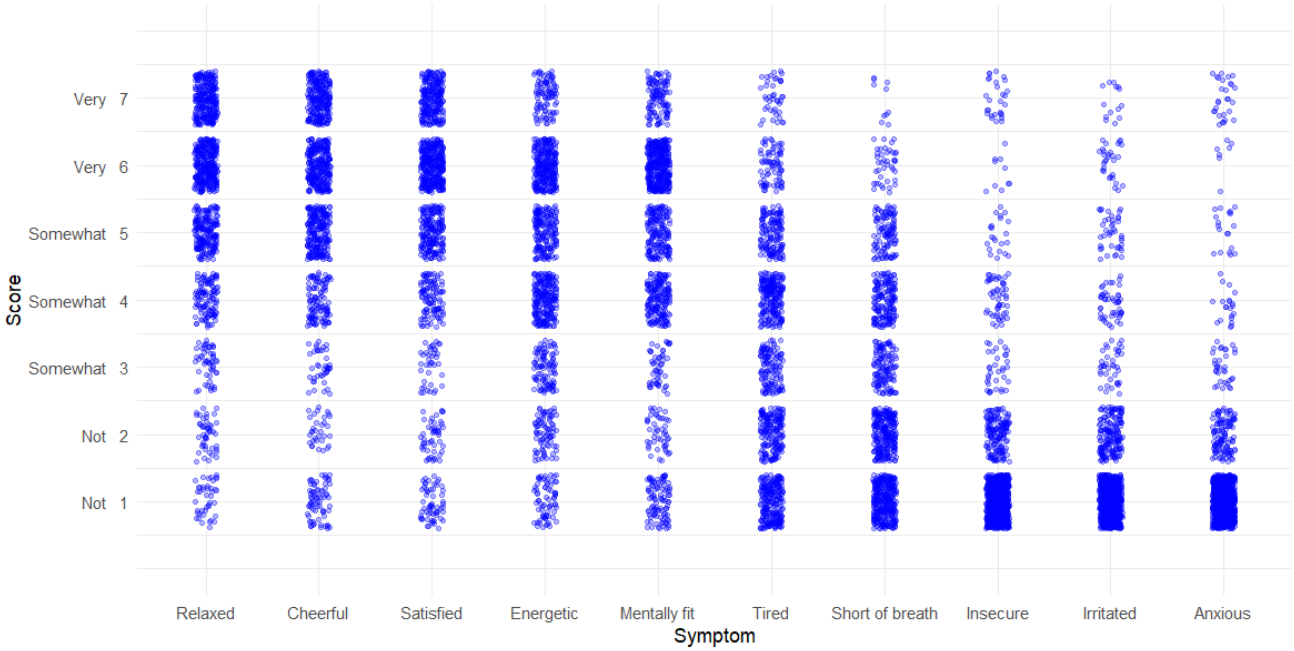
Table . Total EMA[a] responses of the 5 days.

Total EMA responses	Patients, n (%)
18	1 (3)
20	1 (3)
22	1 (3)
23	2 (6)
24	2 (6)
27	2 (6)
28	2 (6)
29	1 (3)
30	3 (9)
31	3 (9)
32	3 (9)
33	1 (3)
34	3 (9)
35	3 (9)
36	2 (6)
37	1 (3)
38	3 (9)

[a]EMA: Ecological Momentary Assessment.

XSL•FO
RenderX

Figure 2. Jittered scatterplot of the EMA-based symptoms scored on a 7-point Likert scale for 5 consecutive days.



Physical Activity

During the 5 days, patients had a median (Q1-Q3) of 19 (3-74) steps pre-EMA and 20 (3-67) steps post-EMA in the 15-minute time windows. During the 30-minute time windows, patients had a median (Q1-Q3) of 55 (10-148) steps pre-EMA and 52 (11-139) steps post-EMA.

In the unadjusted models 1, it was found that when patients reported feeling more energetic and more mentally fit, step count was statistically significantly higher 15 minutes post-EMA, and when patients reported feeling more tired, step count was lower 15 and 30 minutes post-EMA (Table 3).

Table . Associations between symptoms and step count 15 and 30 minutes post–ecological momentary assessment.

Symptom	15-minute coefficient estimate (95% CI)				30-minute coefficient estimate (95% CI)			
	Unadjusted	*P* value	Adjusted[a]	*P* value	Unadjusted	*P* value	Adjusted[a]	*P* value
Relaxed	4.9 (−0.2 to 9.2)	.06	5.1 (0.9 to 10.1)	.046	4.9 (−3.8 to 13.5)	.27	5.8 (−3.6 to 15.2)	.23
Short of breath	−2.6 (−7.6 to 2.5)	.32	0.1 (−5.4 to 5.6)	.97	−5.9 (−14.9 to 3.2)	.20	−1.8 (−12.0 to 8.4)	.72
Energetic	5.9 (1.1 to 10.6)	.02	4.8 (−0.3 to 10.0)	.06	6.6 (−1.9 to 15.1)	.13	4.3 (−5.2 to 13.9)	.37
Cheerful	4.1 (−0.7 to 8.9)	.09	4.0 (−1.1 to 9.1)	.12	6.6 (−1.9 to 15.2)	.13	6.4 (−2.9 to 15.8)	.18
Insecure	2.5 (−3.4 to 8.4)	.40	1.7 (−5.3 to 8.7)	.63	6.4 (−4.2 to 17.0)	.23	6.0 (−7.2 to 19.2)	.37
Irritated	3.6 (−1.8 to 9.1)	.19	2.2 (−3.9 to 8.3)	.49	9.1 (−1.0 to 19.1)	.08	7.4 (−4.1 to 18.9)	.21
Satisfied	2.7 (−2.1 to 7.4)	.27	2.8 (−2.2 to 7.8)	.27	−0.3 (−8.9 to 8.3)	.95	−0.9 (−10.2 to 8.4)	.85
Anxious	2.2 (−4.4 to 8.8)	.52	1.4 (−7.0 to 9.7)	.75	4.1 (−7.5 to 15.8)	.48	3.4 (−12.2 to 19.0)	.67
Tired	−5.4 (−9.9 to −0.8)	.02	−3.9 (−8.4 to 0.7)	.10	−9.8 (−18.0 to −1.5)	.02	−7.4 (−15.9 to 1.1)	.09
Mentally fit	5.0 (0.3 to 9.7)	.04	4.7 (−0.3 to 9.7)	.07	7.6 (−0.7 to 15.9)	.07	7.0 (−2.0 to 16.1)	.13

[a]Adjusted for time of day, working status, GOLD stage, exacerbation history, and exacerbation-related hospitalization in the previous 12 months.

The total number of steps taken during the 5 days was positively correlated with being currently employed (r=0.38; P=.03) and negatively correlated with GOLD stage (r=−0.53; P<.001), exacerbation history (r=−0.41; P=.02), and exacerbation-related

XSL•FO
RenderX

hospitalization in the previous 12 months (r=−0.36; P=.04). Furthermore, steps were negatively correlated with time of day (r=−0.12; P<.001). When adjusting for these significant confounders, no significant associations were found in models 1 between symptoms and step count 30 minutes post-EMA. Only feeling relaxed had a positive relationship with step count 15 minutes post-EMA. That is, for a 1-unit increase in feeling relaxed, the expected value of step count 15 minutes post-EMA increases by 5.1 steps.

Feeling short of breath was positively correlated with GOLD stage (r=0.44; P=.009), whereas feeling energetic was negatively correlated with GOLD stage (r=−0.35; P=.046). Feeling cheerful and satisfied was both negatively correlated with being a current smoker (r=−0.38; P=.03; and r=−0.41; P=.02), and feeling irritated was negatively correlated with CCI points (r=−0.34; P=.049). Furthermore, feeling relaxed was positively correlated with time of day (r=0.12; P<.001), whereas feeling irritated was negatively correlated with time of day (r=−0.07; P=.04).

In both the unadjusted and adjusted models 2, it was found that a higher step count during a 15- and 30-minute time window pre-EMA was associated with feeling less relaxed and feeling more short of breath and tired (Table 4).

Table . Associations between step count 15 and 30 minutes pre–ecological momentary assessment and symptoms.

Symptom	15-minute coefficient estimate (95% CI)				30-minute coefficient estimate (95% CI)			
	Unadjusted	*P* value	Adjusted	*P* value	Unadjusted	*P* value	Adjusted	*P* value
Relaxed	-6.4×10^{-4} (-1.1×10^{4} to -2.0×10^{-4})	.005	-5.2×10^{-4} (-9.7×10^{-4} to -7.0×10^{-5})[a]	.02	-3.9×10^{-4} (-6.2×10^{-4} to -1.5×10^{-4})	.001	-3.2×10^{-4} (-5.6×10^{-4} to -7.9×10^{-5})[a]	.009
Short of breath	8.4×10^{-4} (4.6×10^{-4} to 1.2×10^{-3})	<.001	8.5×10^{-4} (4.7×10^{-4} to 1.2×10^{-3})[b]	<.001	4.6×10^{-4} (2.6×10^{-4} to 6.6×10^{-4})	<.001	4.6×10^{-4} (2.6×10^{-4} to 6.6×10^{-4})[b]	<.001
Energetic	2.3×10^{-4} (-1.8×10^{-4} to 6.3×10^{-4})	.27	2.2×10^{-4} (-1.8×10^{-4} to 6.2×10^{-4})[b]	.29	1.6×10^{-4} (-5.0×10^{-5} to 3.7×10^{-4})	.13	1.6×10^{-4} (-5.4×10^{-5} to 3.7×10^{-4})[b]	.14
Cheerful	-5.1×10^{-5} (-4.3×10^{-4} to 3.3×10^{-4})	.79	-4.0×10^{-5} (-4.2×10^{-4} to 3.4×10^{-4})[c]	.84	-4.5×10^{-5} (-2.5×10^{-4} to 1.5×10^{-4})	.66	-4.0×10^{-5} (-2.4×10^{-4} to 1.6×10^{-4})[c]	.69
Insecure	2.3×10^{-4} (-7.2×10^{-5} to 5.3×10^{-4})	.14	N/A[e]		8.8×10^{-5} (-7.2×10^{-5} to 2.5×10^{-4})	.28	N/A	
Irritated	3.0×10^{-4} (-7.5×10^{-5} to 6.8×10^{-4})	.12	2.6×10^{-4} (1.2×10^{-4} to 6.4×10^{-4})[a,d]	.18	1.0×10^{-4} (-9.3×10^{-5} to 3.0×10^{-4})	.30	8.0×10^{-5} (-1.2×10^{-4} to 2.8×10^{-4})[a,d]	.44
Satisfied	-3.6×10^{-4} (-7.7×10^{-4} to 5.5×10^{-5})	.09	-3.4×10^{-4} (-7.5×10^{-4} to 6.8×10^{-5})[c]	.10	-1.5×10^{-4} (-3.7×10^{-4} to 6.7×10^{-5})	.17	-1.5×10^{-4} (-3.6×10^{-4} to 7.3×10^{-5})[c]	.19
Anxious	1.3×10^{-4} (-1.3×10^{-4} to 3.9×10^{-4})	.33	N/A		5.4×10^{-5} (-8.5×10^{-5} to 1.9×10^{-4})	.45	N/A	
Tired	5.1×10^{-4} (7.2×10^{5} to 9.4×10^{-4})	.02	N/A		2.9×10^{-4} (5.3×10^{-5} to 5.2×10^{-4})	.02	N/A	
Mentally fit	-1.2×10^{-4} (-4.9×10^{-4} to 2.4×10^{-4})	.51	N/A		-6.0×10^{-5} (-2.5×10^{-4} to 1.3×10^{-4})	.54	N/A	

[a]Adjusted for time of day.

[b]Adjusted for Global Initiative for Chronic Obstructive Lung Disease stage.

[c]Adjusted for smoking status.

[d]Adjusted for Charlson Comorbidity Index.

[e]Not applicable.

XSL•FO
RenderX

Sedentary Time

Patients, on average (SD), spent 314 (116) minutes sedentary per patient per day during waking hours. During the 5 days, patients had a median (Q1-Q3) of 9 (3-13) minutes ST pre-EMA and 9 (4-13) minutes post-EMA in the 15-minute time windows. During the 30-minute time windows, patients had a median (Q1-Q3) of 18 (7-24) minutes ST pre-EMA and 18 (9-24) minutes post-EMA.

The total ST during the 5 days was positively correlated with BMI (r=0.36; P=.04) and negatively correlated with being currently employed (r=−0.39; P=.03). In addition, ST was positively correlated with time of day (r=0.24; P<.001). In both the unadjusted and adjusted models 3, no significant associations were found between symptoms and 15 or 30 minutes post-EMA ST (Table 5).

In the unadjusted models 4, it was found that a higher ST during a 15- and 30-minute time window pre-EMA was associated with feeling more anxious (Table 6).

Table . Associations between symptoms and sedentary time 15 and 30 minutes post–ecological momentary assessment.

Symptom	15-minute coefficient estimate (95% CI)				30-minute coefficient estimate (95% CI)			
	Unadjusted	*P* value	Adjusted[a]	*P* value	Unadjusted	*P* value	Adjusted[a]	*P* value
Relaxed	−6.0 (−16.9 to 4.9)	.28	−8.8 (−20.4 to 2.8)	.14	−2.6 (−23.4 to 18.1)	.80	−6.4 (−28.5 to 15.8)	.57
Short of breath	−0.9 (−13.0 to 11.1)	.88	−1.3 (−14.3 to 11.6)	.84	−7.1 (−30.1 to 16.0)	.55	−5.9 (−30.8 to 19.0)	.64
Energetic	4.0 (−7.3 to 15.3)	.49	5.0 (−7.0 to 16.9)	.42	5.6 (−16.1 to 27.3)	.61	5.1 (−17.9 to 28.0)	.67
Cheerful	−3.5 (−14.8 to 7.9)	.55	−3.1 (−15.3 to 9.2)	.62	−5.0 (−26.9 to 16.9)	.65	−6.3 (−29.9 to 17.2)	.60
Insecure	−4.9 (−18.9 to 9.1)	.49	−5.5 (−22.0 to 11.0)	.52	−2.8 (−30.5 to 24.9)	.84	0.8 (−32.1 to 33.6)	.96
Irritated	6.3 (−6.4 to 18.9)	.33	8.9 (−5.1 to 22.9)	.21	8.6 (−15.9 to 33.2)	.49	14.8 (−12.5 to 42.0)	.29
Satisfied	−5.9 (−17.0 to 5.2)	.29	−6.9 (−18.7 to 5.0)	.26	−7.4 (−28.7 to 13.9)	.50	−10.0 (−32.8 to 12.7)	.39
Anxious	7.3 (−8.7 to 23.2)	.37	11.6 (−8.6 to 31.8)	.26	13.5 (−17.4 to 44.4)	.39	26.1 (-12.6 to 64.8)	.19
Tired	−3.2 (−13.7 to 7.4)	.55	−4.4 (−15.6 to 6.8)	.44	−6.2 (−26.5 to 14.1)	.55	−5.9 (−27.3 to 15.6)	.59
Mentally fit	−3.1 (−14.6 to 8.4)	.59	−2.6 (−14.9 to 9.6)	.67	−1.1 (−23.2 to 21.0)	.92	−1.7 (−25.2 to 21.9)	.89

[a]Adjusted for time of day, BMI, and working status.

XSL•FO
RenderX

Table . Associations between sedentary time 15 and 30 minutes pre–ecological momentary assessment and symptoms.

Symptom	15-minute coefficient estimate (95% CI)				30-minute coefficient estimate (95% CI)			
	Unadjusted	*P* value	Adjusted	*P* value	Unadjusted	*P* value	Adjusted	*P* value
Relaxed	1.9×10^{-4} (-6.6×10^{-5} to 4.4×10^{-4})	.15	1.5×10^{-4} (-1.1×10^{-4} to 4.0×10^{-4})[a]	.26	9.4×10^{-5} (-4.5×10^{-5} to 2.3×10^{-4})	.18	6.7×10^{-5} (7.2×10^{-5} to 2.1×10^{-4})[a]	.34
Short of breath	-1.3×10^{-4} (-3.4×10^{-4} to 9.3×10^{-5})	.26	-1.3×10^{-4} (-3.5×10^{-4} to 8.8×10^{-5})[b]	.24	-9.2×10^{-5} (-2.1×10^{-4} to 2.7×10^{-5})	.13	-9.4×10^{-5} (-2.1×10^{-4} to 2.4×10^{-5})[b]	.12
Energetic	-4.9×10^{-5} (-1.8×10^{-4} to 2.8×10^{-4})	.67	5.1×10^{-5} (-1.8×10^{-4} to 2.8×10^{-4})[b]	.66	1.5×10^{-5} (-1.4×10^{-4} to 1.1×10^{-4})	.81	-1.3×10^{-5} (-1.4×10^{-4} to 1.1×10^{-4})[b]	.83
Cheerful	1.2×10^{-4} (-6.9×10^{-5} to 3.4×10^{-4})	.28	1.1×10^{-4} (-1.0×10^{-4} to 3.3×10^{-4})[c]	.30	3.9×10^{-5} (-7.9×10^{-5} to 1.6×10^{-4})	.52	3.5×10^{-5} (-8.2×10^{-5} to 1.5×10^{-4})[c]	.56
Insecure	1.4×10^{-4} (-3.3×10^{-5} to 3.1×10^{-4})	.11	N/A[e]		7.6×10^{-5} (-1.5×10^{-5} to 1.7×10^{-4})	.10	N/A	
Irritated	4.6×10^{-5} (-1.7×10^{-4} to 2.6×10^{-4})	.69	6.0×10^{-5} (-1.5×10^{-4} to 2.7×10^{-4})[a,d]	.58	4.8×10^{-5} (-6.7×10^{-5} to 1.6×10^{-4})	.41	6.7×10^{-5} (-5.9×10^{-5} to 1.7×10^{-4})[a,d]	.34
Satisfied	1.2×10^{-4} (-1.2×10^{-4} to 3.5×10^{-4})	.34	1.1×10^{-4} (-1.3×10^{-4} to 3.5×10^{-4})[c]	.36	6.0×10^{-5} (-7.3×10^{-5} to 1.8×10^{-4})	.39	5.2×10^{-5} (-7.6×10^{-5} to 1.8×10^{-4})[c]	.42
Anxious	1.7×10^{-4} (1.7×10^{-5} to 3.2×10^{-4})	.03	N/A		8.5×10^{-5} (2.5×10^{-6} to 1.7×10^{-4})	.04	N/A	
Tired	-1.9×10^{-4} (-4.4×10^{-4} to 5.6×10^{-5})	.13	N/A		-8.0×10^{-5} (-2.2×10^{-4} to 5.5×10^{-5})	.25	N/A	
Mentally fit	3.7×10^{-5} (-1.7×10^{-4} to 2.5×10^{-4})	.73	N/A		1.4×10^{-6} (-1.2×10^{-4} to 1.1×10^{-4})	.99	N/A	

[a]Adjusted for time of day.

[b]Adjusted for Global Initiative for Chronic Obstructive Lung Disease stage.

[c]Adjusted for smoking status.

[d]Adjusted for Charlson Comorbidity Index.

[e]Not applicable.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to assess the direction of the association between PA and ST measured with accelerometry and physical and psychological symptoms measured with EMA in patients with mild-to-very severe COPD. This study showed that feeling relaxed was positively associated with an increased step count shortly after completing the EMA. A higher step count before EMA completion was associated with feeling less relaxed, more short of breath, and tired. These findings suggest a potential bidirectional relationship between PA and psychological well-being. Furthermore, our results indicate that ST preceding EMA completion was associated with increased feelings of anxiety, underscoring the potential adverse effect of prolonged ST on psychological well-being in patients with COPD.

Steps

In this study, feeling more relaxed was the sole predictor for the number of steps 15 minutes post-EMA, after adjusting for relevant confounders. While this effect size may appear modest,

XSL•FO
RenderX

it should be interpreted in the context of the overall low step counts observed in this population. Given the median step counts of 19 to 20 steps per 15-minute window, even small increases may reflect meaningful changes in short-term activity patterns. Even minor increases in PA can have positive effects on symptoms, functional status, and quality of life in patients with COPD [13,14]. In contrast, a previous study in healthy older adults found no significant associations between feeling calm/relaxed and subsequent PA, nor between positive (happy and cheerful) and negative (anxious, stress, sad, and angry) affect and subsequent PA 15 and 30 minutes post-EMA [44]. Feeling more energetic was associated with increased PA for both 15- and 30-minute time windows in low-active adults [27]. In this study, feeling more energetic was associated with a higher step count 15 minutes post-EMA only in the unadjusted model. Contradicting results are found regarding the relationship between fatigue and PA. One study found no relationship between fatigue and PA in the 15 and 30 minutes post-EMA in low-active adults [27]. Conversely, other studies involving healthy older adults and patients receiving peritoneal dialysis demonstrated that higher fatigue severity and diminished energy were associated with reduced subsequent PA in the 15, 30, 60, and 120 minutes post-EMA [28,45]. These results suggest that higher fatigue results in lower probability of being physically active and in fewer minutes of subsequent PA [28,45]. In this study, feeling more tired was associated with a lower step count 15 and 30 minutes post-EMA only in the unadjusted models. Confounders such as employment status, GOLD stage, exacerbation history, and exacerbation-related hospitalizations might exert a detrimental influence on patients' energy and tiredness, potentially disrupting the relationship with engaging in PA [46,47]. Prior studies showed a positive relationship between improved mental state and engaging in PA in patients with COPD [48]. In this study, feeling more mentally fit was associated with a higher step count 15 minutes post-EMA only in the unadjusted model.

In this study, a higher step count 15 and 30 minutes pre-EMA was associated with feeling less relaxed, more short of breath, and tired. Previously, Hevel et al found no significant associations between step count and subsequent positive (happy and cheerful) and negative (anxious, stressed, sad, and angry) affect in healthy older adults [44]. In contrast, previous studies in nonactive adults and patients receiving peritoneal dialysis found that more PA minutes 15 and 30 minutes pre-EMA were associated with feeling more energetic, positive mood, and lower fatigue severity [27,28,44]. These studies generally demonstrate positive psychological effects after PA. Our findings deviate from these results, potentially because patients with COPD have different experiences with performing PA [37,49]. Physical barriers limit the ability of patients with COPD to engage in PA [8]. Patients frequently report that dyspnea, fatigue, and PA-induced coughing are distressing and discourage being physically active, potentially creating a vicious circle where symptoms discourage activity and inactivity exacerbates symptoms [7,8,14,49]. Consequently, increasing PA may exacerbate symptoms, leading to reduced feelings of relaxation and feeling more short of breath and tired [37].

Sedentary Time

In this study, no significant associations were found between symptoms and subsequent ST. The relationship between symptoms and ST using EMA has been understudied. Prior studies suggest that psychological symptoms such as anxiety, depressed mood, and dyspnea may generally contribute to increased ST in patients with COPD [9,50]. However, these effects may not be directly observed or exist in short-term analysis.

In this study, higher ST was linked to increased feelings of anxiety for both 15- and 30-minute time windows. Only a very small proportion of the study population had probable symptoms of anxiety according to the HADS. This underscores the significant impact of ST on anxiety, aligning with previous research [7].

Methodological Considerations

The current findings need to be interpreted in the light of the number of comparisons that were made in this study. Nonetheless, multiple findings in the same direction, rather than a single statistically significant result, suggest that these are not due to chance alone. Moreover, 'Bonferroni adjustments are at best unnecessary and, at worst, deleterious to sound statistical inference' [51]. This study revealed that EMA through a mobile app was feasible for symptom assessment in technically proficient patients with COPD. Only 3 of 34 patients had insufficient EMA data, and the average response rate (76.1%) is consistent with other feasibility studies exploring the use of EMA in patients with COPD and other clinical populations [26,28,52]. Random time intervals for the EMA assessment prevented patients from anticipating and adjusting their behavior. However, by using time windows surrounding the randomized time intervals of the EMA prompts, instead of examining naturally occurring beginning and ends of taking steps and ST, only partial PA sessions may have been captured. Although data showed a clear variability in step count day to day, we do not expect this to impact the findings as we examined associations between changes in PA, ST, and changes in EMA. The ActiGraph, generally valid and responsive for PA measurements in patients with COPD, had a technical error that invalidated data from 23 patients, reducing the study sample size [37]. The relatively small study sample size and the patients being technically proficient limit the generalizability of our findings. Furthermore, the analyses for symptoms beyond dyspnea and fatigue were exploratory in nature. As these associations were not hypothesized a priori, findings for symptoms such as mental fitness or cheerfulness should be interpreted with caution. Although 1035 EMA scores were related to PA and ST measurements, a larger sample of patients with COPD should be included in future research to test the generalizability and replicability of our findings, including longitudinal follow-up.

Clinical Implications and Future Directions

Future research is necessary to examine the complex associations between PA, ST, and symptoms in patients with COPD, as many external factors could have exerted an influence on patients' ability to perform PA, such as weather and social

XSL•FO
RenderX

context [9,37,49,53]. This study presents findings that contradict previous research, particularly regarding the effects of PA on mood and energy. While it is plausible that patients with COPD experience PA differently due to physical limitations, the study does not provide definitive answers about the underlying reasons for these discrepancies. The absence of a deeper exploration into these mechanisms highlights the need for further research. Future studies should aim to investigate potential explanations, including physiological, psychological, and contextual factors, to better understand how patients with COPD respond to PA and what drives these differences [27].

Conclusions

To conclude, a bidirectional association of feeling relaxed with engaging in PA was observed. More PA pre-EMA seemed to result in adverse physical states, including feeling more short of breath and tired. Furthermore, a longer ST pre-EMA seemed to result in an increased feeling of anxiety. Overall, these findings emphasize the complex relationships between PA, ST, and symptoms in the daily lives of patients with COPD.

Acknowledgments

This project is supported by grant 4.1.16.085 of Lung Foundation Netherlands, Leusden, the Netherlands; AstraZeneca Netherlands; Boehringer Ingelheim Netherlands; and Stichting Astma Bestrijding, Amsterdam, The Netherlands.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to participant confidentiality and privacy concerns, and they are not available for sharing.

Authors' Contributions

BP acts as the guarantor of the content of the manuscript. MVH, YMJG, ZE, MSYT, CB, JBP, EFMW, MAGS, JHV, DJAJ, and MAS designed the study. MVH, YMJG, ZE, JBP, RTMS, EWMAB, JHV, and AWV were responsible for data collection. BP and QD were responsible for data analysis. BP, QD, SH-W, DJAJ, and MAS drafted the manuscript. All authors critically reviewed and revised the manuscript and approved the final version.

Conflicts of Interest

MVH and YMJG report research grant 4.1.16.085 from Lung Foundation Netherlands. MSYT reports research grant from Lung Foundation Netherlands. RTMS reports payment from GSK for lectures and from PENTAX Medical. EWMAB reports research grants from the Dutch Research Council (NWO) and the Netherlands Organization for Health Research and Development (ZonMw). DJAJ reports research grants from Dutch Foundation for Asthma Prevention and the Netherlands Respiratory Society (award), payment from Chiesi, Abbot, and AstraZeneca, and unpaid participation in the Better B steering committee and Wolfson Palliative Care Research Centre Advisory Board. MAS reports research grants from Lung Foundation Netherlands, Dutch Foundation for Asthma Prevention, AstraZeneca, TEVA, Boehringer Ingelheim, and Chiesi. BP, ZE, QD, CB, JBP, EW, MAGS, JHV, SH-W, and AWV declare that they have no conflict of interests.

References

1. Venkatesan P. GOLD COPD report: 2023 update. Lancet Respir Med 2023 Jan;11(1):18. [doi: 10.1016/S2213-2600(22)00494-5] [Medline: 36462509]
2. Cravo A, Attar D, Freeman D, Holmes S, Ip L, Singh SJ. The importance of self-management in the context of personalized care in COPD. Int J Chron Obstruct Pulmon Dis 2022;17:231-243. [doi: 10.2147/COPD.S343108] [Medline: 35095272]
3. Goërtz YMJ, Spruit MA, Van 't Hul AJ, et al. Fatigue is highly prevalent in patients with COPD and correlates poorly with the degree of airflow limitation. Ther Adv Respir Dis 2019;13:1753466619878128. [doi: 10.1177/1753466619878128] [Medline: 31558115]
4. Janssen DJA, Spruit MA, Leue C, et al. Symptoms of anxiety and depression in COPD patients entering pulmonary rehabilitation. Chron Respir Dis 2010 Aug;7(3):147-157. [doi: 10.1177/1479972310369285] [Medline: 20688892]
5. Andersson M, Stridsman C, Rönmark E, Lindberg A, Emtner M. Physical activity and fatigue in chronic obstructive pulmonary disease - a population based study. Respir Med 2015 Aug;109(8):1048-1057. [doi: 10.1016/j.rmed.2015.05.007] [Medline: 26070272]
6. Pitta F, Troosters T, Spruit MA, Probst VS, Decramer M, Gosselink R. Characteristics of physical activities in daily life in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2005 May 1;171(9):972-977. [doi: 10.1164/rccm.200407-855OC] [Medline: 15665324]
7. Tashiro H, Takahashi K. Clinical impacts of interventions for physical activity and sedentary behavior on patients with chronic obstructive pulmonary disease. J Clin Med 2023 Feb 17;12(4):1631. [doi: 10.3390/jcm12041631] [Medline: 36836165]

XSL•FO
RenderX

8. Esteban C, Antón-Ladislao A, Aramburu A, et al. Physical activity and sedentary behaviour in patients admitted with COPD: associated factors. Respir Med Res 2023 Nov;84:101052. [doi: 10.1016/j.resmer.2023.101052] [Medline: 37897880]
9. Harding S, Richardson A, Glynn A, Hodgson L. Influencing factors of sedentary behaviour in people with chronic obstructive pulmonary disease: a systematic review. BMJ Open Respir Res 2024 May 24;11(1):e002261. [doi: 10.1136/bmjresp-2023-002261] [Medline: 38789283]
10. Lee SH, Kim KU, Lee H, Kim YS, Lee MK, Park HK. Factors associated with low-level physical activity in elderly patients with chronic obstructive pulmonary disease. Korean J Intern Med 2018 Jan;33(1):130-137. [doi: 10.3904/kjim.2016.090] [Medline: 28602061]
11. Gimeno-Santos E, Frei A, Steurer-Stey C, et al. Determinants and outcomes of physical activity in patients with COPD: a systematic review. Thorax 2014 Aug;69(8):731-739. [doi: 10.1136/thoraxjnl-2013-204763] [Medline: 24558112]
12. Hanania NA, Müllerova H, Locantore NW, et al. Determinants of depression in the ECLIPSE chronic obstructive pulmonary disease cohort. Am J Respir Crit Care Med 2011 Mar 1;183(5):604-611. [doi: 10.1164/rccm.201003-0472OC] [Medline: 20889909]
13. Xiang X, Huang L, Fang Y, Cai S, Zhang M. Physical activity and chronic obstructive pulmonary disease: a scoping review. BMC Pulm Med 2022 Aug 5;22(1):301. [doi: 10.1186/s12890-022-02099-4] [Medline: 35932050]
14. Ma X, Zhang Q, Gao X, Sun M. Association between physical activity, sedentary behavior and sleep disorders in chronic obstructive pulmonary disease patients: a cross-sectional study. Int J Chron Obstruct Pulmon Dis 2025;20:1175-1187. [doi: 10.2147/COPD.S487616] [Medline: 40297842]
15. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
16. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016 Apr;75:116-123. [doi: 10.1016/j.jpsychires.2016.01.011] [Medline: 26851494]
17. Wu M, Wang Z, Li M, Li K. Daily symptom variability in patients with stable COPD: a narrative review. West J Nurs Res 2018 Oct;40(10):1543-1561. [doi: 10.1177/0193945917705132] [Medline: 28459177]
18. Mesquita R, Spina G, Pitta F, et al. Physical activity patterns and clusters in 1001 patients with COPD. Chron Respir Dis 2017 Aug;14(3):256-269. [doi: 10.1177/1479972316687207] [Medline: 28774199]
19. Weingärtner V, Bausewein C, Higginson IJ, et al. Characterizing episodic breathlessness in patients with advanced disease. J Palliat Med 2013 Oct;16(10):1275-1279. [doi: 10.1089/jpm.2013.0087] [Medline: 24053592]
20. Stone AA, Shiffman S. Ecological momentary assessment (Ema) in behavioral medicine. Ann Behav Med 1994 Sep 1;16(3):199-202. [doi: 10.1093/abm/16.3.199]
21. Miravitlles M, Worth H, Soler Cataluña JJ, et al. Observational study to characterise 24-hour COPD symptoms and their relationship with patient-reported outcomes: results from the ASSESS study. Respir Res 2014 Oct 21;15(1):122. [doi: 10.1186/s12931-014-0122-1] [Medline: 25331383]
22. Verhagen SJW, Hasmi L, Drukker M, van Os J, Delespaul P. Use of the experience sampling method in the context of clinical trials. Evid Based Ment Health 2016 Aug;19(3):86-89. [doi: 10.1136/ebmental-2016-102418] [Medline: 27443678]
23. Lee TY, Chen CH, Chen IM, et al. Dynamic bidirectional associations between global positioning system mobility and ecological momentary assessment of mood symptoms in mood disorders: prospective cohort study. J Med Internet Res 2024 Dec 6;26:e55635. [doi: 10.2196/55635] [Medline: 39642364]
24. Paolillo EW, Tang B, Depp CA, et al. Temporal associations between social activity and mood, fatigue, and pain in older adults with HIV: an ecological momentary assessment study. JMIR Ment Health 2018 May 14;5(2):e38. [doi: 10.2196/mental.9802] [Medline: 29759960]
25. Stinson L, Liu Y, Dallery J. Ecological momentary assessment: a systematic review of validity research. Perspect Behav Sci 2022 Jun;45(2):469-493. [doi: 10.1007/s40614-022-00339-w] [Medline: 35719870]
26. Davis TJ, Hevel DJ, Dunton GF, Maher JP. Bidirectional associations between physical activity and pain among older adults: an ecological momentary assessment study. J Aging Phys Act 2023 Apr 1;31(2):240-248. [doi: 10.1123/japa.2022-0014] [Medline: 35995422]
27. Liao Y, Chou CP, Huh J, Leventhal A, Dunton G. Examining acute bi-directional relationships between affect, physical feeling states, and physical activity in free-living situations using electronic ecological momentary assessment. J Behav Med 2017 Jun;40(3):445-457. [doi: 10.1007/s10865-016-9808-9] [Medline: 27766481]
28. Tarca B, Jesudason S, Wycherley TP, et al. Ecological momentary assessment to explore fatigue, mood, and physical activity levels in people receiving peritoneal dialysis. Kidney Int Rep 2024 Mar;9(3):601-610. [doi: 10.1016/j.ekir.2023.12.024] [Medline: 38481501]
29. Miller SN, Nichols M, Teufel Ii RJ, Silverman EP, Walentynowicz M. Use of ecological momentary assessment to measure dyspnea in COPD. Int J Chron Obstruct Pulmon Dis 2024;19:841-849. [doi: 10.2147/COPD.S447660] [Medline: 38566847]
30. Goërtz YMJ, Looijmans M, Prins JB, et al. Fatigue in patients with chronic obstructive pulmonary disease: protocol of the Dutch multicentre, longitudinal, observational FAntasTIGUE study. BMJ Open 2018 Apr 10;8(4):e021745. [doi: 10.1136/bmjopen-2018-021745] [Medline: 29643168]

XSL•FO
RenderX

31. Van Herck M, Goërtz YMJ, Ebadi Z, et al. Associated factors of fatigue in patients with COPD: results from the FAntasTIGUE study. ERJ Open Res 2025:00857-02024. [doi: 10.1183/23120541.00857-2024]
32. Metsemakers JF, Höppener P, Knottnerus JA, Kocken RJ, Limonard CB. Computerized health information in the Netherlands: a registration network of family practices. Br J Gen Pract 1992 Mar;42(356):102-106. [Medline: 1493025]
33. Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. Chest 2000 May;117(5 Suppl 2):398S-401S. [doi: 10.1378/chest.117.5_suppl_2.398s] [Medline: 10843984]
34. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. Thorax 1999 Jul;54(7):581-586. [doi: 10.1136/thx.54.7.581] [Medline: 10377201]
35. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994 Jul;38(5):383-392. [doi: 10.1016/0022-3999(94)90099-x] [Medline: 7965927]
36. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
37. Demeyer H, Mohan D, Burtin C, et al. Objectively measured physical activity in patients with COPD: recommendations from an International Task Force on Physical Activity. J COPD F 2021;8(4):528-550. [doi: 10.15326/jcopdf.2021.0213]
38. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]
39. Bertoche MP, Furlanetto KC, Hirata RP, et al. Assessment of sedentary behaviour in individuals with COPD: how many days are necessary? ERJ Open Res 2023 Jul;9(4):00732-2022. [doi: 10.1183/23120541.00732-2022] [Medline: 37650084]
40. Palmier-Claus JE, Myin-Germeys I, Barkus E, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011 Jan;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x] [Medline: 20712828]
41. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364. [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
42. Spearman C. The proof and measurement of association between two things. Am J Psychol 1904 Jan;15(1):72. [doi: 10.2307/1412159]
43. van Buul AR, Kasteleyn MJ, Chavannes NH, Taube C. Physical activity in the morning and afternoon is lower in patients with chronic obstructive pulmonary disease with morning symptoms. Respir Res 2018 Mar 27;19(1):49. [doi: 10.1186/s12931-018-0749-4] [Medline: 29587841]
44. Hevel DJ, Dunton GF, Maher JP. Acute bidirectional relations between affect, physical feeling states, and activity-related behaviors among older adults: an ecological momentary assessment study. Ann Behav Med 2021 Feb 12;55(1):41-54. [doi: 10.1093/abm/kaaa027] [Medline: 32441738]
45. Maes IHL, Delespaul PAEG, Peters ML, et al. Measuring health-related quality of life by experiences: the experience sampling method. Value Health 2015 Jan;18(1):44-51. [doi: 10.1016/j.jval.2014.10.003] [Medline: 25595233]
46. López-López L, Torres-Sánchez I, Romero-Fernández R, Granados-Santiago M, Rodríguez-Torres J, Valenza MC. Impact of previous physical activity levels on symptomatology, functionality, and strength during an acute exacerbation in COPD patients. Healthcare (Basel) 2018 Nov 29;6(4):139. [doi: 10.3390/healthcare6040139] [Medline: 30501112]
47. Gudmundsson G, Gislason T, Janson C, et al. Depression, anxiety and health status after hospitalisation for COPD: a multicentre study in the Nordic countries. Respir Med 2006 Jan;100(1):87-93. [doi: 10.1016/j.rmed.2005.04.003] [Medline: 15893921]
48. Tian J, Zhou F, Zhang XG, et al. Experience of physical activity in patients with COPD: a systematic review and qualitative meta-synthesis. Geriatr Nurs 2022;47:211-219. [doi: 10.1016/j.gerinurse.2022.07.013] [Medline: 35940039]
49. Delgado-Ortiz L, Polhemus A, Keogh A, et al. Listening to the patients' voice: a conceptual framework of the walking experience. Age Ageing 2023 Jan 8;52(1):afac233. [doi: 10.1093/ageing/afac233] [Medline: 36729471]
50. Wshah A, Selzler AM, Hill K, Brooks D, Goldstein R. Determinants of sedentary behaviour ini ndividuals with COPD: a qualitative exploration guided by the theoretical domains framework. COPD: Journal of Chronic Obstructive Pulmonary Disease 2020 Jan 2;17(1):65-73. [doi: 10.1080/15412555.2019.1708883]
51. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998 Apr 18;316(7139):1236-1238. [doi: 10.1136/bmj.316.7139.1236] [Medline: 9553006]
52. Sandberg J, Sundh J, Anderberg P, et al. Comparing recalled versus experienced symptoms of breathlessness ratings: an ecological assessment study using mobile phone technology. Respirology 2022 Oct;27(10):874-881. [doi: 10.1111/resp.14313] [Medline: 35697350]
53. Hevel DJ, Drollette ES, Dunton GF, Maher JP. Social and physical context moderates older adults' affective responses to sedentary behavior: an ecological momentary assessment study. J Gerontol B Psychol Sci Soc Sci 2021 Nov 15;76(10):1983-1992. [doi: 10.1093/geronb/gbab036] [Medline: 33656523]

Abbreviations

COPD: chronic obstructive pulmonary disease

XSL•FO
RenderX

EMA: ecological momentary assessment
FEV_1: forced expiratory volume in 1 second
mMRC: Modified Medical Research Council
PA: physical activity
ST: sedentary time

Edited by L Buis; submitted 21.08.24; peer-reviewed by H Demeyer, M Giurgiu, S Lau; revised version received 15.07.25; accepted 09.09.25; published 22.10.25.

Please cite as:
Palmen B, Ebadi Z, van Herck M, Goërtz YMJ, Deng Q, Thong MSY, Burtin C, Peters JB, Sprooten RTM, Bischoff EWMA, Wouters EFM, Sprangers MAG, Vercoulen JH, Houben-Wilke S, Vaes AW, Janssen DJA, Spruit MA
Bidirectional Associations Between Physical Activity, Sedentary Behavior, and Daily Symptoms in Patients With Chronic Obstructive Pulmonary Disease: Longitudinal Observational Study
JMIR Mhealth Uhealth 2025;13:e65653
URL: https://mhealth.jmir.org/2025/1/e65653
doi:10.2196/65653

© Banchia Palmen, Zjala Ebadi, Maarten van Herck, Yvonne M J Goërtz, Qichen Deng, Melissa S Y Thong, Chris Burtin, Jeannette B Peters, Roy T M Sprooten, Erik W M A Bischoff, Emiel F M Wouters, Mirjam A G Sprangers, Jan H Vercoulen, Sarah Houben-Wilke, Anouk W Vaes, Daisy J A Janssen, Martijn A Spruit. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impacts of Environmental Distractions and Interruptions on Unsupervised Digital Cognitive Assessments in Older Adults: Cognitive Ecological Momentary Assessment Study

Matthew Welhaf[1], PhD; Hannah Wilks[2], MA; Andrew J Aschenbrenner[1], PhD; Samhita Katteri[2], BA; John C Morris[1], PhD; Jason J Hassenstab[1], PhD

[1]Department of Neurology, School of Medicine, Washington University in St. Louis, 660 South Euclid Avenue, St Louis, MO, United States
[2]Department of Psychological & Brain Sciences, Washington University in St. Louis, St Louis, MO, United States

Corresponding Author:
Matthew Welhaf, PhD
Department of Neurology, School of Medicine, Washington University in St. Louis, 660 South Euclid Avenue, St Louis, MO, United States

Abstract

Background: Unsupervised cognitive assessments are becoming commonly used in studies of aging and neurodegenerative diseases. As assessments are completed in everyday environments and without a proctor, there are concerns about how common distractions may impact performance and whether these distractions may differentially impact those experiencing the earliest symptoms of dementia.

Objective: We examined the impact of self-reported interruptions, testing location, and social context during testing on remote cognitive assessments in older adults.

Methods: Participants from the Ambulatory Research in Cognition smartphone study were classified as cognitively normal (n=380) or as having very mild dementia (n=37). Participants completed daily tests of processing speed, working memory, and associative memory. At each assessment, participants were asked for their current location and social surroundings, which was used to quantify whether participants were either at home (or not) and by themselves (or not). After each assessment session, participants were asked if they experienced any interruptions. Mixed-effect modeling tested the interactions between location, social context, and clinical status. Additional analyses were conducted by removing sessions where participants reported that they were interrupted at any point during the testing period.

Results: Across all participants, momentary effects of environmental distractions were minimal. Specifically, when tests were completed in the presence of others, participants exhibited slightly increased variability in processing speed (*P*=.04). However, these momentary effects of environmental distractions were dependent upon cognitive status (*P*=.009). Cognitively normal older adults had better visuospatial working memory performance when they completed tests at home compared to when they completed tests away from home (*P*=.001). However, older adults with very mild dementia showed no effect of testing location on the same task (*P*=.36). Conversely, cognitively normal older adults did not differ in their processing speed at either testing location (*P*=.88). Older adults with very mild dementia were slightly faster when not at home (*P*=.04). Social context only impacted variability in processing speed for participants with very mild dementia (*P*=.04). When considering tests completed in the most distracting environments (away from home and in the presence of others), those with very mild dementia showed larger differences only on the visuospatial working memory measure. Additional analyses demonstrated that after removing sessions in which participants self-reported experiencing an interruption (1194/9633, 12.4% of all assessments), these small effects of environmental distractions on cognition remained, but were more apparent in those with very mild dementia.

Conclusions: Social context and location of unsupervised remote cognitive testing have small impacts on performance, but these impacts were not consistent across cognitive domains and were mostly limited to participants demonstrating the earliest symptoms of dementia. Remote cognitive testing provides valid and reliable data in older adults, but care should be taken to allow participants to report distractions that may occur during testing.

(JMIR Mhealth Uhealth 2025;13:e71578) doi:10.2196/71578

KEYWORDS

ecological momentary assessment; environmental factors; cognition; Alzheimer disease; smartphone-based testing; mobile phone

XSL•FO
RenderX

Introduction

It is becoming increasingly common to use remote cognitive assessments to track fluctuations in cognitive performance in older adults [1-6]. These tools are especially useful for measuring subtle fluctuations in cognition during the preclinical stage of Alzheimer disease (AD) [3,4,7-15]. However, because remote testing is typically done in a participant's home, usually without the supervision of a test proctor or clinician, it is hard to know the extent to which the surrounding environment is controlled and distraction-free. Distractions occur in approximately 7% - 10% of at-home cognitive testing sessions [8,16]. Such distractions can lead to worse performance on remote cognitive assessments because they might pull attention away from the task, which could lead to participants making more mistakes and having slower and more variable reaction times (RTs).

Ecological momentary assessment (EMA) is a powerful sampling method that leverages the variability in environmental factors to assess psychological processes in participants' everyday lives. EMA has been used to provide a more naturalistic and sensitive assessment of cognition (and other psychological factors) [2,17] and may provide a more reliable sampling of a participant's cognitive performance that matches their true cognitive ability [15,17]. EMA studies often consider external factors that might impact cognitive performance, such as time of day [18], mood or affect [19], current and recent stressors [20,21], everyday activities such as chores [7,22], and recent physical activities [23]. However, because EMA studies collect data during participants' everyday lives, it is possible that distractions introduced by the environment are more prominent and impact performance more when compared to in-clinic or even remotely proctored assessments. A recent review [24] suggests that testing location or social context can have some meaningful impact on performance on remote cognitive tests (both proctored and unsupervised), but the literature has been limited to assessments in a single age group or in a single cognitive domain, so the generalizability of the findings is unknown. Supporting this claim, a recent empirical study found that during remote cognitive assessments, participants who provided data for normative purposes performed better on measures of processing speed when they completed tests at home and also performed better on measures of recognition and working memory when they completed tests alone [25].

Depending on study goals, environmental distractions may be considered confounds that need to be controlled for in EMA studies of cognition. In studies of aging and neurodegenerative disease, researchers are interested in capturing subtle changes in cognition related to changes in neurological functioning, so understanding the effects of different sources of distraction is critical. Recent work has addressed the potential confounding effect of contextual factors such as environmental and social distractions on remote cognitive performance among cognitively impaired and cognitively unimpaired older adults [26]. Cerino et al [26] reported that there were small effects of environmental and social distractions that were evident on measures of visual short-term memory binding and spatial working memory. Although the effects were not large, they were still significant across all participants, even those with mild cognitive impairment (MCI), suggesting that environmental and social distractions can affect cognition beyond disease-related impairments.

It is important to consider the extent to which those showing the very earliest symptoms of dementia are affected by these distractors compared to cognitively normal individuals. That is, subtle deficits may become more apparent for those experiencing the very earliest symptoms of dementia when tests are completed "in the wild," away from home, or when around others. The goal of this study was to examine the impact of environmental and social distractions during remote cognitive testing in a well-characterized cohort of older adults enrolled in a study of normal aging and dementia. We hypothesized that symptomatic participants would be more impacted by environmental distractions compared to cognitively normal participants, especially when tests were taken in the least optimal settings (ie, both away from home and in the presence of others). Consistent with previous work [26], we operationalized testing location as assessments completed at home (versus not), and social environment as assessments that were completed alone versus in the presence of others.

Methods

Participants

Participants were recruited from studies of aging and dementia at the Knight Alzheimer Disease Research Center at Washington University School of Medicine in St. Louis. The Ambulatory Research in Cognition (ARC) study began in March of 2020. ARC is a custom-built smartphone app that assesses cognition with repeated brief assessments up to 4 times per day for one week [13]. Participants were invited to enroll after completing their annual cognitive and clinical assessment and began ARC testing shortly after enrolling. As data collection is still ongoing, we used a data freeze from June 2024 for the current analyses.

For the current analyses, participants were included if they (1) completed at least 10 sessions during their baseline ARC testing to ensure adequate engagement in the study and enough observations in different environments for comparisons and (2) had a clinical assessment (see below) within a year of starting their ARC testing to ensure an accurate classification of cognitive status.

Clinical Assessments

Clinical status was assessed using the Clinical Dementia Rating (CDR) [27]. The CDR rates cognitive and functional performance on a 5-point scale across 6 domains (memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care). Participants and a collateral source (a close family member or friend) both underwent semistructured interviews to determine CDR scores, with CDRs of 0 indicating cognitive normality, 0.5 indicating very mild dementia, 1 indicating mild dementia, 2 indicating moderate dementia, and 3 indicating severe dementia. For this study, we grouped participants as either cognitively normal (CDR 0) or diagnosed with very mild dementia (CDR 0.5). The

XSL•FO
RenderX

parent study was designed to characterize the transition from cognitive normality to the very first signs of clinical symptoms of AD; therefore, most participants were cognitively normal (CDR 0), with the final sample consisting of 385 cognitively normal participants and 37 participants with a CDR of 0.5. Of the 37 participants with CDR 0.5, a total of 26 (70%) were clinically diagnosed with AD, and the rest with a dementia of different etiology, including Parkinson disease, Lewy body dementia, or uncertain etiology. CDR assessments were completed 7.64 (SD 7.78) weeks from their ARC assessment.

Smartphone Assessment

Methods for the procedures of ARC are detailed elsewhere, so are only briefly described here (for full details on the ARC platform, see studies by Nicosia et al [13] and Wilks et al [18]). Participants completed all ARC assessments on a smartphone (either their own or a study-provided device). ARC assessments were sent to participants using the native iOS (Apple Inc) or Android (Google LLC) notification systems pseudorandomly. Participants were instructed to complete the assessment as soon as possible within a 2-hour window. Participants completed 3 cognitive tasks (described below) up to 4 times per day over the course of a week. Figure 1 provides screenshots of each task presented on ARC.

Figure 1. Examples of grids, prices, and symbols tests from Ambulatory Research in Cognition (ARC).

Symbols

During this processing speed task, participants were randomly shown 3 pairs of abstract shapes and were asked to select which of 2 possible responses matched 1 of the 3 targets [17]. At each assessment, participants completed 12 trials lasting approximately 20 - 60 seconds. Consistent with our previous work [13,18,28], we analyzed symbols' performance in 2 ways. The first was the median RT of correct trials, which we use to capture differences in processing speed. We also calculated a measure of RT variability, the coefficient of variation (CoV). For both median RT and RT CoV, higher scores reflect poorer performance (ie, slower responding or greater variability).

Prices

This associate memory task presented subjects with 2 phases: a learning phase and a recognition phase. During learning, participants were instructed to learn and remember 10 item-price pairs that were presented on screen for 3 seconds per pair. The items were everyday shopping items (eg, foods or household supplies), and prices were randomly assigned 3-digit prices. During the recognition phase, participants were presented with an item they encountered during learning and 2 prices and were asked to choose which price was originally paired with the item. Each administration of this test requires approximately 60 seconds. The primary dependent measure for this task was the error rate during recognition, such that higher scores indicated poorer recognition.

Grids

In this spatial working memory task, participants were presented with images of three everyday objects (key, smartphone, and pen) on a 5x5 grid. Participants were told to remember the location of each item on the grid. After the encoding phase, participants were given a distractor task (locating Fs among a grid of Es). Following the distractor task, participants were asked to recall the location of each item on a blank grid. Participants completed 2 trials at each session (total time 30 - 40 s). The primary dependent measure for this task was the Euclidean distance for each item. Higher scores indicated that participants placed the items further away from the initially encoded location and suggested worse performance [17].

Current Environment and Social Context

At each ARC session, participants answered several questions about their experiences or situations at the time of the notification. The first question, which we used to operationalize testing location, asked, "Where are you right now?" Participants selected one of the following response options: my home, my work, school, another person's home, vehicle, outside, or other. We categorized participants' responses into either home (if they answered "my home") or away (all other responses).

The second question, which we used to operationalize social context, asked, "Who is with you right now?" Participants could select all that apply from the following options: nobody, partner or spouse, family, friends, coworkers, pet, and other. As above, we categorized participants' responses into either testing alone

XSL•FO
RenderX

(if they answered "nobody") or not alone (all other responses). These environment and social context categorizations were consistent with a previous study [26]. As displayed in Table 1, most participants completed their tests at home and alone, with few reports in other contexts. As there was little variation within each category, we chose not to treat these as separate variables in analyses. Additionally, there is some debate in the field of EMA on whether "pets" should be counted as alone or not alone. We considered responses of "pets" as not alone, as it is likely that pets might also be a source of distraction for participants during remote cognitive assessments (eg, a dog barking at a passing car might be a salient distraction). The results were nearly identical when including "pet" responses as alone.

Table . Demographics of sample by CDR[a] status. Gender and race were self-reported.

Demographic variables	CDR 0 (n=380)	CDR=0.5 (n=37)	Statistic	*P* value
Age (years), mean (SD)	75.42 (5.93)	76.57 (6.32)	1.07 (42.76)[b]	.29
Gender, n (%)	226 (56.8)	14 (37.8)	5.17 (1)[c]	.02
Education (years), mean (SD)	16.49 (2.42)	16.08 (2.19)	1.08(44.90)[b]	.28
Race, n (%)				
White	322 (84.7)	33 (89.1)		
Black or African American	61 (15.1)	4 (10.8)		
Other	2 (< 1)	0 (0)		
APOE ε4[d] status, n (%)			8.80 (1)[c]	.003
ε4 carrier	129 (33.5)	22 (59.5)		
ε4 noncarrier	256 (66.5)	15 (40.5)		
Sessions complete, mean (SD)	23.22 (4.29)	21.92 (5.40)	1.42 (40.49)[b]	.16
Prices error rate, mean (SD)	0.29 (0.09)	0.37 (0.08)	–5.40 (45.43)[b]	<.001
Grids Euclidean distance, mean (SD)	0.74 (0.26)	0.95 (0.26)	–4.72 (43.23)[b]	<.001
Symbols median RT[e], mean (SD)	3.25 (1.08)	3.98 (1.27)	–3.39 (41.09)[b]	.002
Symbols RT CoV[f], mean (SD)	0.34 (0.08)	0.39 (0.10)	–2.91 (40.21)[b]	.006
Proportion completed at home, mean (SD)	0.80 (0.19)	0.86 (0.15)	–2.19 (48.76)[b]	.03
Proportion completed alone, mean (SD)	0.49 (0.32)	0.43 (0.32)	1.16 (42.93)[b]	.25

[a]CDR: Clinical Dementia Rating.

[b] *t (df)*

[c]Chi-square (*df*)

[d]*APOE* ε4*:* apolipoprotein ε4.

[e]RT: reaction time.

[f]CoV: coefficient of variation.

Self-Reported Interruptions

Upon completion of each session, participants were asked if they experienced any interruptions during the testing period. This item was used in additional analyses (described below) to account for sessions where participants were acutely aware of potential distractions.

Analytic Plan

Our primary analyses addressed how the environmental factors affect the relationship between CDR status and performance in remote unsupervised testing. In each linear mixed effects model, the outcome measure was either symbols median RT, symbols CoV, prices error rate, or grids Euclidean distance. We included fixed effects of age at ARC testing, years of education, self-reported gender (male or female), and number of completed sessions (to account for practice effects). We also included a fixed effect term for the proportion of sessions completed at home and the proportion of sessions completed alone to account for between-person differences in the testing location or context. That is, some people might complete more ARC assessments at home, and some might complete more assessments alone

XSL•FO
RenderX

than others. To account for within-person variance, we included a term for the reported location (ie, home or not home; coded 0 and 1, respectively) and social context (ie, alone or not alone; coded 0 and 1, respectively) at each ARC session. We included all 2-way interactions between CDR status, within-person location, and within-person social context, and the higher-order 3-way interaction as fixed effects. These terms allowed us to test the differential impact that environmental distractions had on both cognitively normal older adults and those with very mild dementia. All models included a random intercept for the participant.

In our second set of analyses, we removed sessions where, upon completion of each assessment session, participants reported being interrupted. We tested the hypothesis that if participants were aware of their distractions and reported them, then removing these sessions should greatly reduce or eliminate any effects of location or social context on cognitive performance. Conversely, if location and social context effects are still apparent, then we would demonstrate that environmental distractions impact remote cognitive performance beyond instances in which participants are overtly interrupted.

All mixed-effect models were conducted using R software (version 4.4.0; R Foundation for Statistical Computing) including the *lmer* package [29] and *P* values for mixed-effect models were estimated using the *lmerTest* [30] package. Follow-up comparisons were conducted using *emmeans* [31].

Ethical Considerations

All participants provided informed consent, and all procedures were approved by the Human Research Protections Office at Washington University in St. Louis (IRB# 201907164), and the research was conducted in accordance with the Helsinki Declaration. Participants' data were fully anonymized during the data collection process, and researchers cannot access any identifiable participant information without requesting access to the data via a formal data request to the Knight Alzheimer Disease Research Center Administrative Core. As described in Nicosia et al [13], participants were paid US $0.50 per completed assessment session. Participants also received bonus payments for completing all 4 sessions on any given day (US $1.00 per occurrence, max of US $7.00), completing at least 2 assessments per day for 7 days (US $6.00), and completing at least 21 assessments over 7 days (US $5.00). The maximum compensation possible for one 7-day assessment visit was US $32.00.

Results

Participant Demographics

The final sample consisted of 417 participants (see Table 1). Table 2 provides demographic and ARC task information on the full sample, grouped by clinical status. Participants were highly educated, most self-identified their race as White, and were a majority of females. Adherence was high with participants completing, on average, 23.11 (SD 4.41) out of a possible 28 sessions (82.54% adherence).

Table . Mean (SD) and percentage of testing locations and social contexts reported by participants for both cognitively normal and those with very mild dementia across the week of testing (maximum of 28 sessions).

Question	CDR[a] 0 (n=380)		CDR=0.5 (n=37)	
Where are you right now?	Mean (SD)	Total observations, n (%)	Mean (SD)	Total observations, n (%)
My home	18.77 (5.57)	7131 (81)	19.03 (6.09)	704 (86)
My work	0.50 (1.67)	190 (2)	0.00 (0.00)	0 (0)
School	0.02 (0.17)	9 (<1)	0.00 (0.00)	0 (0)
Another person's home	0.82 (2.16)	311 (4)	0.46 (1.30)	17 (2)
Vehicle	1.20 (1.54)	455 (5)	1.19 (1.37)	44 (6)
Outside	0.60 (1.15)	229 (3)	0.35 (0.63)	13 (2)
Other	1.25 (2.12)	476 (5)	0.89 (2.11)	33 (4)
Who is with you right now?				
Nobody	11.33 (7.69)	4306 (49)	9.38 (7.65)	347 (43)
Partner or spouse	5.42 (6.57)	2060 (23)	7.62 (7.55)	282 (34)
Family	1.62 (3.59)	616 (7)	1.24 (3.68)	46 (6)
Friends	0.61 (1.30)	232 (3)	0.16 (0.37)	6 (1)
Coworkers	0.12 (0.64)	44 (<1)	0.03 (0.16)	1 (<1)
Pet	2.01 (4.94)	764 (8)	1.35 (4.13)	50 (7)
Other	0.49 (1.48)	188 (2)	0.22 (0.63)	8 (1)
Multiple	1.55 (3.70)	590 (7)	1.92 (3.85)	71 (9)

[a]CDR: Clinical Dementia Rating.

XSL•FO
RenderX

As expected, participants with very mild dementia performed worse on the ARC assessments when aggregated across the week of testing. Further, there was a small difference in the proportion of tests completed at home, such that individuals with very mild dementia completed a greater proportion of tests at home compared to cognitively normal participants (n=704, 86% vs n=7131, 81%, respectively). There was no difference between cognitively normal participants and those with very mild dementia in the proportion of tests completed alone. Table 1 provides more detailed information regarding the response frequencies to both the location and social surrounding questions for both cognitively normal participants and those with very mild dementia.

Effects of Testing Location and Social Context

Consistent with our other publications in this sample [13,18,28,32], older age at the ARC assessment predicted worse performance on all ARC measures (Table S1 in Multimedia Appendix 1). Male participants made more errors than females on the prices test but were slightly better at the grids test. More years of education were also associated with better performance on prices and grids (but not symbols). Participants who completed more sessions performed better on grids and had faster processing speed and lower variability over time on symbols indicating evidence of practice effects on these tasks.

As seen in Table 3, neither the proportion of tests completed at home (*ß*'s=–0.131 to 0.072, *P*'s>.31) nor the proportion of tests completed alone (*ß*'s=–0.309 to 0.020, *P*'s>.06) was related to performance on any of the ARC tasks. When considering the session-level effects of testing location, there were no significant effects of testing location on ARC performance. Participants showed similar performance on all ARC measures regardless of testing location: prices (M_{home} 0.327, SE 0.007 vs $M_{\text{not home}}$ 0.354, SE 0.011, t_{9399}=–2.773, *P*=.006, Cohen *d*=–0.06, 95% CI –0.10 to –0.02), grids (M_{home} 0.816, SE 0.02 vs $M_{\text{not home}}$ 0.813, SE 0.03, t_{9217}=–0.118, *P*=.91, *d* < –0.01, 95% CI –0.04 to 0.04), symbols processing speed (M_{home} 3.52, SE 0.09 vs $M_{\text{not home}}$ 3.43, SE 0.10, t_{9399}=1.925, *P*=.054, *d*=0.04, 95% CI 0.00 to 0.08), symbols variability in processing speed (M_{home} 0.361, SE 0.007 vs $M_{\text{not home}}$ 0.372, SE 0.01, t_{9366}=–1.362, *P*=.17, *d*=–0.03, 95% CI –0.07 to 0.01). The main effect on prices performance is significant when calculating the simple effect; however, caution should be used when interpreting this effect, as it is likely driven by the interaction between testing location and social context.

Table . Mixed effect models of the interaction between CDR[a] status, testing location, social context on ARC[b] performance (N=417). Exact *P* values are presented in the text. Models adjusted for covariates, including age, years of education, self-reported gender, and number of completed sessions. Full model output is provided Table S1 in Multimedia Appendix 1.

Predictors	Prices error rate, estimate (SE)	Grids Euclidean distance, estimate (SE)	Symbols median RT[c], estimate (SE)	Symbols RT CoV[d], estimate (SE)
Home proportion[e]	–0.015 (0.024)	0.072 (0.072)	–0.131 (0.286)	–0.001 (0.022)
Alone proportion[f]	0.001 (0.014)	0.020 (0.041)	–0.309 (0.163)	0.016 (0.013)
Home[g]	0.004 (0.007)	0.017 (0.019)	–0.032 (0.033)	0.003 (0.006)
Alone[h]	0.006 (0.005)	0.011 (0.012)	–0.022 (0.021)	0.008[i] (0.004)
CDR status x home	0.046 (0.033)	–0.218[i] (0.087)	–0.384[i] (0.152)	0.021 (0.027)
CDR status x alone	0.004 (0.015)	0.038 (0.040)	–0.049 (0.071)	0.032[j] (0.012)
Home or not x alone	0.024[j] (0.009)	0.048[i] (0.024)	0.070 (0.042)	0.002 (0.008)
CDR status x home x alone	–0.045 (0.039)	0.260[i] (0.102)	0.401[i] (0.178)	–0.015 (0.032)

[a]CDR: Clinical Dementia Rating.

[b]ARC: Ambulatory Research in Cognition.

[c]RT: reaction time.

[d]CoV: coefficient of variation.

[e]Home proportion: proportion of sessions completed at home.

[f]Alone proportion: proportion of sessions completed alone.

[g]Home: session completed at home (coded 0) or away from home (coded 1).

[h]Alone: session completed alone (coded 0) or in the presence of others (coded 1).

[i]*P<.05.

[j]**P<.01.

This was also largely the case for session-level effects of social context. Participants showed similar performance, regardless of social context, on prices (M_{alone} 0.336, SE 0.01 vs $M_{\text{not alone}}$ 0.345, SE 0.009, t_{9504}=–0.871, *P*=.38, *d*=–0.02, 95% CI –0.06 to 0.02]), grids (M_{alone} 0.755, SE 0.03 vs $M_{\text{not alone}}$ 0.874, SE 0.03, t_{9333}=–4.398, *P*<.001, *d*=–0.09, 95% CI –0.13 to –0.05), and symbols processing speed (M_{alone} 3.43, SE 0.10 vs $M_{\text{not alone}}$ 3.52, SE 0.09, t_{9223}=–1.878, *P*=.06, *d*=–0.04, 95% CI –0.08 to

XSL•FO
RenderX

0.00). Again, this seemingly large effect is likely driven by higher-order interactions, and so caution should be used when interpreting this effect.

However, there was a significant effect on the variability measure from the symbols task, such that when participants completed sessions in the presence of others (mean 0.377, SE 0.008), they demonstrated increased variability in RTs on the symbols task compared to when they completed sessions alone (mean 0.356, SE 0.009; t_{9469}=–2.563, P=.04, d=–0.05, 95% CI –0.09 to –0.01).

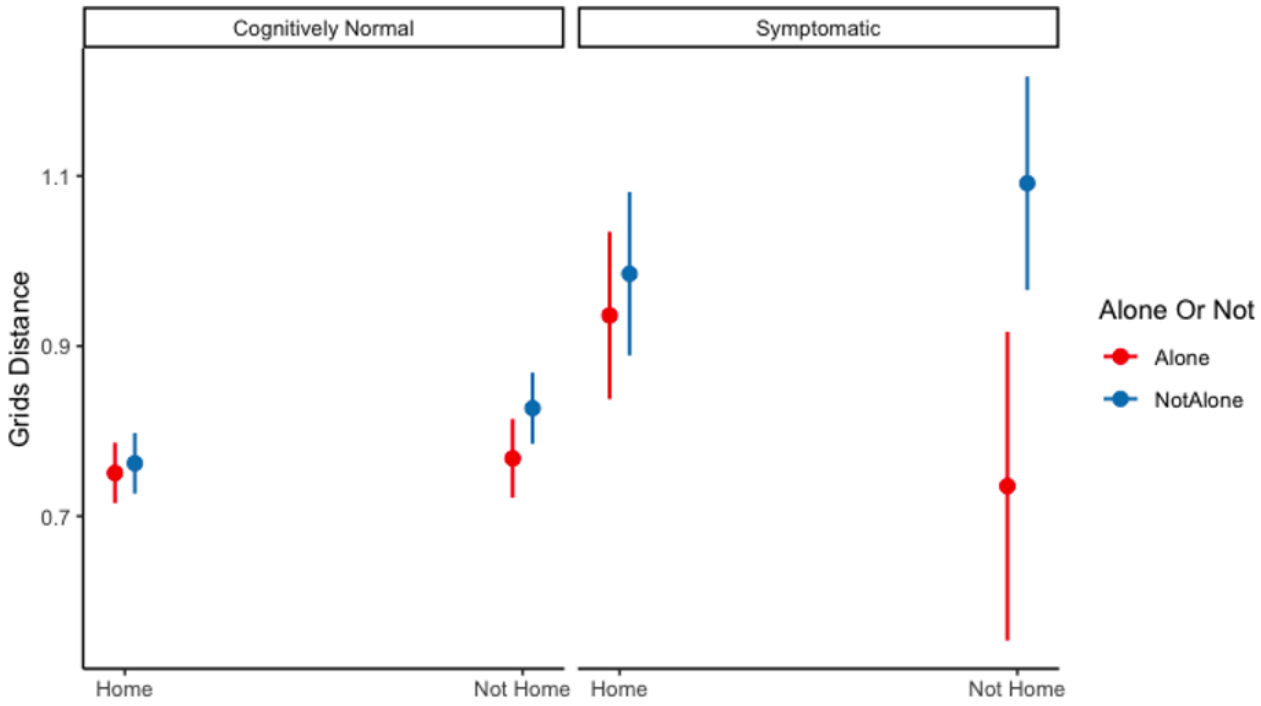
Performance on 3 ARC measures was significantly impacted by both CDR status and testing location or social context (ie, significant 2-way interactions; Table 3). For grid performance, cognitively normal older adults performed better when they completed tests at home compared to when they completed tests away from home (M_{home} 0.714, SE 0.01 vs $M_{not\ home}$ 0.755, SE 0.02, t_{9036}=–3.317, P=.001). However, older adults with very mild dementia did not differ in their grids performance in different testing environments (M_{home} 0.918, SE 0.04 vs $M_{not\ home}$ 0.871, SE 0.06, t_{9036}=–3.317, P=.001). For symbols processing speed, cognitively normal older adults did not differ in either testing location (M_{home} 3.22, SE 0.05 vs $M_{not\ home}$ 3.23, SE 0.06, t_{9152}=–0.157, P=.88). However, older adults with very mild dementia showed a significant, albeit small, difference in processing speed based on their testing location (M_{home} 3.81, SE 0.17) vs $M_{not\ home}$ 3.63, SE 0.19, t_{9254}=0.925, P=.36).

As for social context, there was a significant interaction with CDR status on RT variability on the symbols task. When cognitively normal participants completed the test alone, they were significantly less variable compared to when they completed the tests in the presence of others (M_{alone} 0.337, SE 0.004 vs $M_{not\ alone}$ 0.346, SE 0.05, t_{9261}=–4.650, P=.02). This effect was larger for participants with very mild dementia who completed tests (M_{alone} 0.375, SE 0.02 vs $M_{not\ alone}$ 0.409, SE 0.01, t_{9510}=–2.085, P=.04).

Finally, across all participants, there was evidence that the combination of taking tests away from home and in the presence of others significantly impacted performance on prices and grids. Specifically, when tests were taken in the most suboptimal locations (ie, away from home and in the presence of others), participants, on average, made significantly more errors on prices (mean 0.359, SE 0.01) compared to when they completed the test in “optimal” contexts (ie, at home and alone; mean 0.323, SE 0.008; t_{9535}=–3.218, P=.001). This was also evident for grids as participants were, on average, worse when completing both away from home and in the presence of others (mean 0.917, SE 0.03) compared to when they completed tests at home and alone (mean 0.801, SE 0.02; t_{9170}=2.099, P=.04).

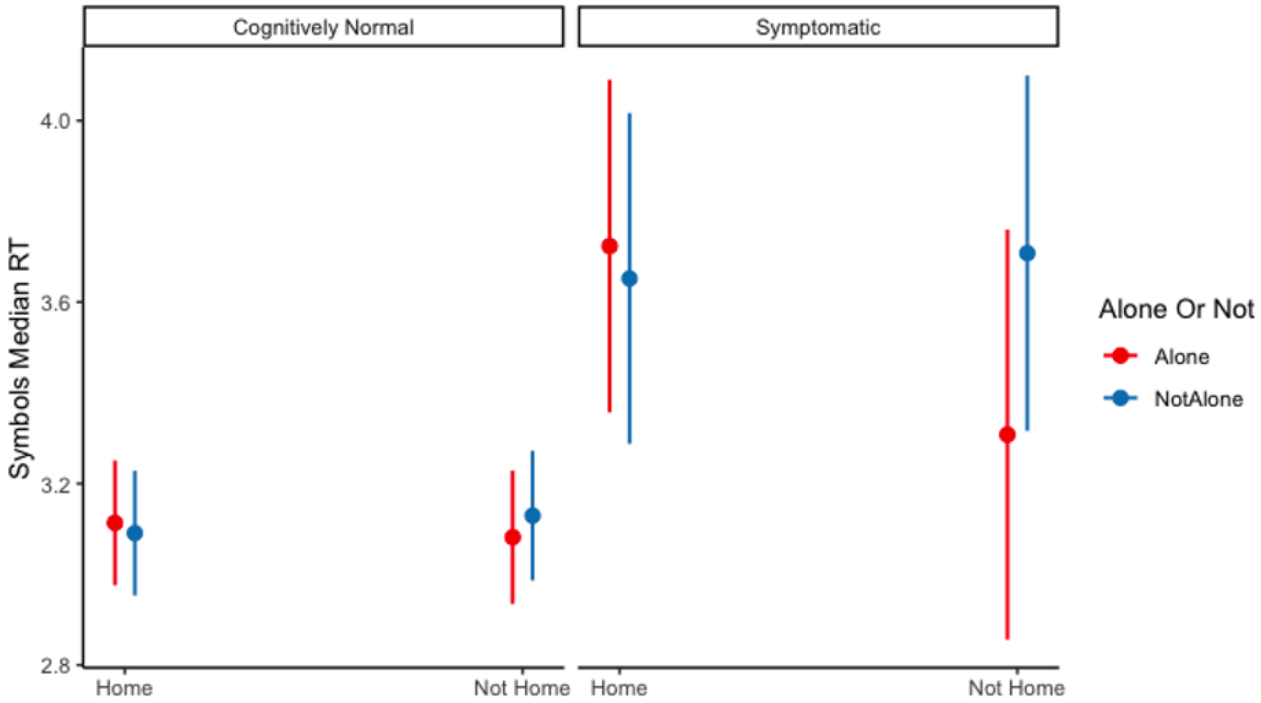
These lower-order interactions were qualified by 3-way interactions on grids (Figure 2) and symbols median RT (Figure 3). Decomposing the interaction suggested that when those with very mild dementia completed tests both away from home and in the presence of others, they had an average grids score of 1.049 (SE 0.06), which was significantly higher than when they completed tests at home and alone (mean 0.894, SE 0.05; t_{9399}=–2.754, P=.006, d=–0.05, 95% CI –0.09 to –0.01]). For processing speed, however, the difference between the optimal (mean 3.85 s, SE 0.18) and suboptimal (mean 3.83 s, SE 0.19) environments was not statistically significant (P=.88).

Figure 2. Estimated effects of testing location and social context on grids for cognitively normal (CDR 0; n=380) and symptomatic (CDR 0.5; n=37) participants. Error bars represent the 95% CI. CDR: Clinical Dementia Rating.



XSL•FO
RenderX

Figure 3. Estimated effects of testing location and social context on symbol median RT for cognitively normal (CDR 0; n=380) and symptomatic (CDR 0.5; n=37) participants. Error bars represent the 95% CI. CDR: Clinical Dementia Rating; RT: reaction time.



Removing Interrupted Sessions

Interruptions were reported on 1194 of 9633 completed sessions (approximately 12% of all sessions). Removing these sessions resulted in an overall similar pattern of effects (Table 4), albeit an attenuated effect. The previously described 3-way interactions between CDR status, testing location, and social context were still evident on the grids and symbols median RT. Thus, even after removing sessions during which participants reported experiencing an overt interruption, the effects of environmental distractions were still apparent, especially for those with very mild dementia.

XSL•FO
RenderX

Table . Mixed effect models of the interaction between CDR[a] status, testing location, and social context on ARC[b] performance after excluding sessions that were interrupted (N=417). Models adjusted for covariates, including age, years of education, self-reported gender, and number of completed sessions. Full model output is provided in Table S2 of Multimedia Appendix 1.

Predictors	Prices error rate, estimate (SE)	Grids Euclidean distance, estimate (SE)	Symbols median RT[c], estimate (SE)	Symbols RT CoV[d], estimate (SE)
Home proportion[e]	–0.015 (0.024)	0.051 (0.072)	–0.087 (0.269)	0.001 (0.020)
Alone proportion[f]	0.005 (0.014)	0.019 (0.042)	–0.286 (0.152)	0.022 (0.012)
Home[g]	0.003 (0.008)	0.009 (0.020)	–0.062 (0.034)	–0.000 (0.006)
Alone[h]	0.006 (0.005)	0.008 (0.013)	–0.037 (0.022)	0.009[i] (0.004)
CDR status x home	0.055 (0.033)	–0.243[j] (0.086)	–0.404[j] (0.148)	0.025 (0.026)
CDR status x alone	0.002 (0.016)	0.029 (0.041)	–0.061 (0.072)	0.034[j] (0.012)
Home or not x alone	0.022[i] (0.010)	0.034 (0.026)	0.094[i] (0.045)	0.001 (0.008)
CDR status x home x alone	–0.057 (0.040)	0.287[j] (0.102)	0.380[i] (0.176)	–0.030 (0.031)

[a]CDR: Clinical Dementia Rating.

[b]ARC: Ambulatory Research in Cognition.

[c]RT: reaction time.

[d]CoV: coefficient of variation.

[e]Home proportion: proportion of sessions completed at home.

[f]Alone proportion: proportion of sessions completed alone.

[g]Home: session completed at home (coded 0) or away from home (coded 1).

[h]Alone: session completed alone (coded 0) or in the presence of others (coded 1).

[i]*P<.05

[j]**P<.01

Discussion

Principal Findings

The overall results of this study illustrate the importance of considering environmental factors in remote studies of cognition where participants may complete assessments in different locations and social contexts. The effect sizes of distractions on cognition were generally small in magnitude and likely would not invalidate remote testing, but distractions may be most important to consider in populations where cognitive impairment is evident.

Across participants, in-the-moment effects of testing location and social context were evident on measures of associative memory and visual working memory but not processing speed. When participants completed tests in the most distracting environments (both away from home and around others), they performed worse. However, for cognitively normal older adults, the impacts of these environmental factors were minimal, if at all apparent. On the other hand, participants with very mild dementia were the most sensitive to distractions, specifically when completing tests away from their homes and in the presence of others. These findings indicate that an inability to maintain performance in the face of distractions may be a novel indicator of very early neurodegenerative disease and emphasize the need for remote studies to carefully consider testing environments. While a strength of EMA is that it can allow for the measurement of cognition in a variety of environments and situations, the potential for distractions should at least be evaluated and, if necessary, included in statistical models. This is particularly true for studies enrolling individuals with very mild dementia symptoms.

The descriptive results from this study revealed that older adult participants engaged in remote cognitive assessments outside of their homes during 81% of assessments and in the company of others during 44% of assessments, both of which are environments that are likely to produce distractions. Our estimates of tests completed in these environments closely match that of Cerino et al [26], who found that older adults in their study reported testing at home in 78% of assessments and testing alone for 44%. Interestingly, Cerino et al [26] reported that participants with MCI completed slightly more tests at home compared to cognitively normal participants (P=.08). Results from our sample replicated these patterns as we found a significant difference between the frequency of tests completed at home between cognitively normal participants and those with a CDR of 0.5, indicating the presence of a very mild dementia. A CDR of 0.5 can be considered very similar to a classification of MCI [33,34]. This suggests that older adults with cognitive difficulties may spend more time at home or may avoid taking tests outside of the home, perhaps due to the impacts of environmental distractions.

Participants in our study have the option to report if they experienced a significant interruption during testing. Interrupted sessions were not uncommon for our participants, as nearly 12% of sessions were reported as having some form of interruption. Interestingly, the effects of environment, while

XSL•FO
RenderX

reduced, remained significant after removing sessions where participants specifically flagged interruptions. Thus, aspects of environmental context that may not rise to the level of an overt interruption may have deleterious effects on cognition, and these effects appear to be magnified in those with dementia symptoms.

Implications

The results of this study have several implications for EMA studies of cognition in older adults at risk for AD. First, the descriptive results showed that a large proportion of assessments are completed at home. Thus, it is possible that we are not fully sampling participants' everyday contexts as well as we can or that participants are more likely to only take tests when they are at home. Due to the age of our cohort, most of our participants are retired and therefore may spend most of their time at home. Although overall adherence was high in this study, it is also possible that our participants chose not to complete tests when they were away from home, which may have led to an underestimation of the effects we observed. In other words, it might appear that older adults enrolled in EMA studies prefer home environments over completing tests "in the wild." While EMA has a clear strength of being able to capture cognition (and other processes) in different environments, future studies should be aware that EMA studies of cognition in older adults might have limited sampling outside the home or in different social contexts because participants tend to complete assessments in minimally distracting and more controlled environments.

A second important implication of this work stems from our findings regarding the impact of environmental factors on cognition. Our participants frequently completed assessments in environments that could be considered distracting. Additionally, participants reported overt interruptions much more frequently than anticipated, during 12% of testing occasions. However, even after accounting for testing location, social context, and overt interruptions, the effects of environmental distractions had very little impact on performance for cognitively normal older adults. However, for those experiencing the earliest symptoms of dementia—in what could be considered the MCI stage, the effects of distractions were stronger, although still at a small effect size. EMA studies should routinely sample for location and social context and allow participants to report interruptions so that these can be included in statistical models. This is particularly important for studies that enroll participants with cognitive impairment.

These findings also have implications for understanding how distractions and interruptions impact functioning during everyday cognitively demanding tasks. Everyday tasks such as driving, managing money and medications, preparing meals, and general housekeeping require several different processes and are important for maintaining functional independence [35,36]. Being interrupted or distracted during these activities might lead to severe outcomes such as accidents or taking the wrong medicine. Our results suggest that techniques or interventions that teach individuals how to minimize distractions, especially during cognitively demanding tasks, might be useful to minimize adverse outcomes.

Limitations and Future Directions

While the results of this study are intriguing and provide insight into the impact of environmental distractions on unsupervised cognitive assessments in older adults, it is worth noting some limitations. The sample used in this study consisted of older adults who are largely retired and therefore, may spend more time at home (which was reflected in the high proportion of tests completed at home, roughly 80%). This also likely contributed to the large standard errors for our estimates, especially among older adults with very mild dementia. Examining these effects in a cohort of younger participants, who are likely still employed and dealing with more day-to-day activities that require them to be out of the house, may more effectively examine the impacts of environmental distractions.

This study also relied on self-reports of testing location and social context, which may not capture objective levels of environmental distractions. Future research could consider incorporating continuous passive monitoring of environmental factors to better contextualize the effects. For example, GPS tracking could be used to detect differences in cognitive performance as a function of participants' distance from home. Likewise, passive monitoring of noise levels from participants' testing devices could be used to objectively quantify distractions in their environment. Additionally, our question regarding "interruptions" was not open-ended, and so we could not capture differences in the types of interruptions participants experienced. For example, notifications from other applications that popped up during the testing may differ in their distraction and impact on performance compared to interruptions from a family member calling for the participant. Future research should consider gathering more detailed information about the types of distractions or interruptions experienced by participants. Finally, this study focused only on the baseline assessment from ARC and how participants' self-reported environmental factors were related to cognitive performance. Future work could investigate if there are longitudinal changes in participants' environmental contexts (frequency of testing at home or alone) and how that is related to cognitive changes over time.

Conclusions

Unsupervised EMA cognitive assessments are a unique data collection method that more closely represents how participants use cognition in their daily lives, but tests are often completed in distracting environments. Our results suggest that effects of environmental distractions, such as overt interruptions during testing and the location and social context in which tests are completed can have small impacts on cognition, and these effects were most apparent for older adults showing early symptoms of dementia. The effect sizes attributable to distractions in cognitively normal participants were very small and are not likely to threaten the validity or reliability of cognitive testing acquired in natural environments on mobile devices; thus, our results continue to support the validity of unsupervised ambulatory cognitive assessments.

XSL•FO
RenderX

Acknowledgments

The authors acknowledge all faculty and staff at the Knight Alzheimer Disease Research Center for their assistance with this project. We thank the research participants in this study for dedicating their time to advancing research on Alzheimer disease. We thank Dr. Dave Balota for comments on an earlier draft of this paper. MW and HW were supported by a National Institute on Aging (NIA) T32 Training Grant (AG000030-47). AJA was supported by the NIA (K01AG071847). JCM was supported by NIA (P30AG066444, P01AG003991, and P01AG026276). JJH is supported by the NIA (R01AG057840). The funding sources had no part in the design of this study or interpretation of the results. The authors attest that there was no use of generative artificial intelligence (AI) technology in the generation of text, figures, or other informational content of this paper.

Data Availability

To access the materials and data analyzed for this project, a formal data request should be submitted to the Knight Alzheimer Disease Research Center Administrative Core. Requesters should indicate the study for which they would like to obtain data. Details and data-sharing policies can be found online [37].

Authors' Contributions

MW, HW, and JJH conceptualized this study. MW, AJA, and JJH handled the methodology. AJA and JJH acquired software and administered this project. MW and HW validated the data. MW worked on the formal analysis, writing of the original draft, and visualization. MW, HW, AJA, and JJH investigated and the curated data. JCM and JJH obtained resources and acquired the funding. HW, AJA, SK, JCM, and JJH reviewed and edited the writing. JJH supervised this study.

Conflicts of Interest

Neither JCM nor his family owns stock or has equity interest (outside of mutual funds or other externally directed accounts) in any pharmaceutical or biotechnology company. All other authors have no disclosures to report.

Multimedia Appendix 1
Complete model outputs for linear effects models presented in main text.
[DOCX File, 20 KB - mhealth_v13i1e71578_app1.docx]

References

1. Atkins AS, Kraus MS, Welch M, et al. Remote self-administration of digital cognitive tests using the brief assessment of cognition: feasibility, reliability, and sensitivity to subjective cognitive decline. Front Psychiatry 2022;13:910896. [doi: 10.3389/fpsyt.2022.910896] [Medline: 36090378]
2. Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. Innov Aging 2019 Jan;3(1):igy038. [doi: 10.1093/geroni/igy038] [Medline: 30619948]
3. Myers JR, Glenn JM, Madero EN, et al. Asynchronous remote assessment for cognitive impairment: reliability verification of the neurotrack cognitive battery. JMIR Form Res 2022 Feb 18;6(2):e34237. [doi: 10.2196/34237] [Medline: 35179511]
4. Öhman F, Berron D, Papp KV, et al. Unsupervised mobile app-based cognitive testing in a population-based study of older adults born 1944. Front Digit Health 2022;4:933265. [doi: 10.3389/fdgth.2022.933265] [Medline: 36426215]
5. Schweitzer P, Husky M, Allard M, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. Int J Methods Psychiatr Res 2017 Sep;26(3):e1521. [doi: 10.1002/mpr.1521] [Medline: 27539327]
6. Thompson LI, Harrington KD, Roque N, et al. A highly feasible, reliable, and fully remote protocol for mobile app‐based cognitive assessment in cognitively healthy older adults. Alz Dem Diag Ass Dis Mo 2022 Jan;14(1):e12283. [doi: 10.1002/dad2.12283]
7. Allard M, Husky M, Catheline G, et al. Mobile technologies in the early detection of cognitive decline. In: Herholz K, editor. PLoS ONE 2014;9(12):e112197. [doi: 10.1371/journal.pone.0112197] [Medline: 25536290]
8. Berron D, Olsson E, Andersson F, et al. Remote and unsupervised digital memory assessments can reliably detect cognitive impairment in Alzheimer's disease. Alzheimers Dement 2024 Jul;20(7):4775-4791. [doi: 10.1002/alz.13919] [Medline: 38867417]
9. Berron D, Glanz W, Clark L, et al. A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. NPJ Digit Med 2024 Mar 26;7(1):79. [doi: 10.1038/s41746-024-00999-9] [Medline: 38532080]
10. Butler J, Watermeyer TJ, Matterson E, Harper EG, Parra-Rodriguez M. The development and validation of a digital biomarker for remote assessment of Alzheimer's diseases risk. Digit Health 2024;10:20552076241228416. [doi: 10.1177/20552076241228416] [Medline: 38269369]
11. Geddes MR, O'Connell ME, Fisk JD, et al. Remote cognitive and behavioral assessment: report of the Alzheimer Society of Canada Task Force on dementia care best practices for COVID‐19. Alzheimers Dement Diagn Assess Dis Monit 2020 Jan;12(1):e12111. [doi: 10.1002/dad2.12111]

XSL•FO
RenderX

12. Moore RC, Ackerman RA, Russell MT, et al. Feasibility and validity of ecological momentary cognitive testing among older adults with mild cognitive impairment. Front Digit Health 2022;4:946685. [doi: 10.3389/fdgth.2022.946685] [Medline: 35990012]
13. Nicosia J, Aschenbrenner AJ, Balota DA, et al. Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. J Int Neuropsychol Soc 2023 Jun;29(5):459-471. [doi: 10.1017/S135561772200042X] [Medline: 36062528]
14. Papp KV, Samaroo A, Chou H, et al. Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. Alz Dem Diag Ass Dis Mo 2021 Jan;13(1):e12243. [doi: 10.1002/dad2.12243]
15. Staffaroni AM, Tsoy E, Taylor J, Boxer AL, Possin KL. Digital cognitive assessments for dementia. Pr Neurol 2020;2020:24-45. [Medline: 33927583]
16. Madero EN, Anderson J, Bott NT, et al. Environmental distractions during unsupervised remote digital cognitive assessment. J Prev Alzheimers Dis 2021;8(3):263-266. [doi: 10.14283/jpad.2021.9] [Medline: 34101782]
17. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan;25(1):14-30. [doi: 10.1177/1073191116643164] [Medline: 27084835]
18. Wilks H, Aschenbrenner AJ, Gordon BA, et al. Sharper in the morning: cognitive time of day effects revealed with high-frequency smartphone testing. J Clin Exp Neuropsychol 2021 Oct;43(8):825-837. [doi: 10.1080/13803395.2021.2009447] [Medline: 35037593]
19. Salthouse TA, Berish DE. Correlates of within-person (across-occasion) variability in reaction time. Neuropsychology 2005 Jan;19(1):77-87. [doi: 10.1037/0894-4105.19.1.77] [Medline: 15656765]
20. Hyun J, Sliwinski MJ, Smyth JM. Waking up on the wrong side of the bed: the effects of stress anticipation on working memory in daily life. J Gerontol Ser B 2019 Jan 1;74(1):38-46. [doi: 10.1093/geronb/gby042]
21. Sliwinski MJ, Smyth JM, Hofer SM, Stawski RS. Intraindividual coupling of daily stress and cognition. Psychol Aging 2006 Sep;21(3):545-557. [doi: 10.1037/0882-7974.21.3.545] [Medline: 16953716]
22. Campbell LM, Paolillo EW, Heaton A, et al. Daily activities related to mobile cognitive performance in middle-aged and older adults: an ecological momentary cognitive assessment study. JMIR mHealth uHealth 2020 Sep 24;8(9):e19579. [doi: 10.2196/19579] [Medline: 32969829]
23. Kekäläinen T, Luchetti M, Terracciano A, et al. Physical activity and cognitive function: moment-to-moment and day-to-day associations. Int J Behav Nutr Phys Act 2023 Nov 22;20(1):137. [doi: 10.1186/s12966-023-01536-9] [Medline: 37993862]
24. Weizenbaum E, Torous J, Fulford D. Cognition in context: understanding the everyday predictors of cognitive performance in a new era of measurement. JMIR mHealth uHealth 2020 Jul 23;8(7):e14328. [doi: 10.2196/14328] [Medline: 32706680]
25. Paolillo EW, Bomyea J, Depp CA, Henneghan AM, Raj A, Moore RC. Characterizing performance on a suite of English-language NeuroUX mobile cognitive tests in a US adult sample: ecological momentary cognitive testing study. J Med Internet Res 2024 Nov 25;26:e51978. [doi: 10.2196/51978] [Medline: 39586088]
26. Cerino ES, Katz MJ, Wang C, et al. Variability in cognitive performance on mobile devices is sensitive to mild cognitive impairment: results from the Einstein aging study. Front Digit Health 2021;3:758031. [doi: 10.3389/fdgth.2021.758031] [Medline: 34927132]
27. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology (ECronicon) 1993 Nov;43(11):2412-2414. [doi: 10.1212/wnl.43.11.2412-a] [Medline: 8232972]
28. Welhaf MS, Wilks H, Aschenbrenner AJ, et al. Naturalistic assessment of reaction time variability in older adults at risk for Alzheimer's disease. J Int Neuropsychol Soc 2024 Jun;30(5):428-438. [doi: 10.1017/S1355617723011475] [Medline: 38282413]
29. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
30. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
31. Lenth R. Emmeans: estimated marginal means, aka least-squares means. GitHub. 2025. URL: https://rvlenth.github.io/emmeans/ [accessed 2025-08-21]
32. Aschenbrenner AJ, Hassenstab J, Morris JC, Cruchaga C, Jackson JJ. Relationships between hourly cognitive variability and risk of Alzheimer's disease revealed with mixed-effects location scale models. Neuropsychology 2024 Jan;38(1):69-80. [doi: 10.1037/neu0000905] [Medline: 37079810]
33. Gross AL, Hassenstab JJ, Johnson SC, et al. A classification algorithm for predicting progression from normal cognition to mild cognitive impairment across five cohorts: the preclinical AD consortium. Alz Dem Diag Ass Dis Mo 2017 Jan;8(1):147-155. [doi: 10.1016/j.dadm.2017.05.003]
34. Morris JC, Storandt M, Miller JP, et al. Mild cognitive impairment represents early-stage Alzheimer disease. Arch Neurol 2001 Mar;58(3):397-405. [doi: 10.1001/archneur.58.3.397] [Medline: 11255443]
35. Fong TG, Gleason LJ, et al. The successful aging after elective surgery functional measures working group. PM R 2015;7(7):727-735. [doi: 10.1016/j.pmrj.2015.01.018]

XSL•FO
RenderX

36. Lawson L, Mc Ardle R, Wilson S, Beswick E, Karimi R, Slight SP. Digital endpoints for assessing instrumental activities of daily living in mild cognitive impairment: systematic review. J Med Internet Res 2023 Jul 25;25:e45658. [doi: 10.2196/45658] [Medline: 37490331]
37. Data request form. WashU Medicine Knight Alzheimer Disease Research Center. URL: https://knightadrc.wustl.edu/data-request-form/ [accessed 2025-08-28]

Abbreviations

AD: Alzheimer disease
ARC: Ambulatory Research in Cognition
CDR: Clinical Dementia Rating
CoV: coefficient of variation
EMA: ecological momentary assessment
Knight ADRC: Knight Alzheimer Disease Research Center
MCI: mild cognitive impairment
RT: reaction time

Edited by L Buis; submitted 17.02.25; peer-reviewed by J Sun, R Moore; revised version received 19.06.25; accepted 05.08.25; published 08.09.25.

Please cite as:
Welhaf M, Wilks H, Aschenbrenner AJ, Katteri S, Morris JC, Hassenstab JJ
Impacts of Environmental Distractions and Interruptions on Unsupervised Digital Cognitive Assessments in Older Adults: Cognitive Ecological Momentary Assessment Study
JMIR Mhealth Uhealth 2025;13:e71578
URL: https://mhealth.jmir.org/2025/1/e71578
doi:10.2196/71578

© Matthew Welhaf, Hannah Wilks, Andrew J Aschenbrenner, Samhita Katteri, John C Morris, Jason J Hassenstab. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Understanding the Relationship Between Ecological Momentary Assessment Methods, Sensed Behavior, and Responsiveness: Cross-Study Analysis

Diane Cook[1], PhD; Aiden Walker[1], BS; Bryan Minor[1], PhD; Catherine Luna[1], MS; Sarah Tomaszewski Farias[2], PhD; Lisa Wiese[3], PhD; Raven Weaver[1], PhD; Maureen Schmitter-Edgecombe[1], PhD

[1]Department of Psychology, College of Arts and Sciences, Washington State University, 3160 Folsom Blvd, Sacramento, WA, United States

[2]Department of Neurology, UC Davis Medical Center, University of California at Davis, Davis, CA, United States

[3]Christine E Lynn College of Nursing, Florida Atlantic University, Boca Raton, FL, United States

Corresponding Author:
Diane Cook, PhD
Department of Psychology, College of Arts and Sciences, Washington State University, 3160 Folsom Blvd, Sacramento, WA, United States

Abstract

Background: Ecological momentary assessment (EMA) offers an effective method to collect frequent, real-time data on an individual's well-being. However, challenges exist in response consistency, completeness, and accuracy.

Objective: This study examines EMA response patterns and their relationship with sensed behavior for data collected from diverse studies. We hypothesize that EMA response rate (RR) will vary with prompt time of day, number of questions, and behavior context. In addition, we postulate that response quality will decrease over the study duration and that relationships will exist between EMA responses, participant demographics, behavior context, and study purpose.

Methods: Data from 454 participants in 9 clinical studies were analyzed, comprising 146,753 EMA mobile prompts over study durations ranging from 2 weeks to 16 months. Concurrently, sensor data were collected using smartwatch or smart home sensors. Digital markers, such as activity level, time spent at home, and proximity to activity transitions (change points), were extracted to provide context for the EMA responses. All studies used the same data collection software and EMA interface but varied in participant groups, study length, and the number of EMA questions and tasks. We analyzed RR, completeness, quality, alignment with sensor-observed behavior, impact of study design, and ability to model the series of responses.

Results: The average RR was 79.95%. Of those prompts that received a response, the proportion of fully completed response and task sessions was 88.37%. Participants were most responsive in the evening (82.31%) and on weekdays (80.43%), although results varied by study demographics. While overall RRs were similar for weekday and weekend prompts, older adults were more responsive during the week (an increase of 0.27), whereas younger adults responded less during the week (a decrease of 3.25). RR was negatively correlated with the number of EMA questions (r=−0.433, P<.001). Additional correlations were observed between RR and sensor-detected activity level (r=0.045, P<.001), time spent at home (r=0.174, P<.001), and proximity to change points (r=0.124, P<.001). Response quality showed a decline over time, with careless responses increasing by 0.022 (P<.001) and response variance decreasing by 0.363 (P<.001). The within-study dynamic time warping distance between response sequences averaged 14.141 (SD 11.957), compared with the 33.246 (SD 4.971) between-study average distance. ARIMA (Autoregressive Integrated Moving Average) models fit the aggregated time series with high log-likelihood values, indicating strong model fit with low complexity.

Conclusions: EMA response patterns are significantly influenced by participant demographics and study parameters. Tailoring EMA prompt strategies to specific participant characteristics can improve RRs and quality. Findings from this analysis suggest that timing EMA prompts close to detected activity transitions and minimizing the duration of EMA interactions may improve RR. Similarly, strategies such as gamification may be introduced to maintain participant engagement and retain response variance.

(JMIR Mhealth Uhealth 2025;13:e57018) doi:10.2196/57018

KEYWORDS

ecological momentary assessment; smart home; smartwatch; cognitive assessment; well-being; monitoring; monitoring behavior; machine learning; artificial intelligence; app; wearables; sensor; effectiveness; accuracy

XSL•FO
RenderX

Introduction

Background

Ecological momentary assessment (EMA), also known as experience sampling or in situ (situated) self-reporting, is a powerful method for capturing real-time insights into an individual's health and well-being. Leveraging the convenience and ubiquity of mobile devices, EMA has been particularly effective in longitudinally monitoring conditions such as depression and mental well-being [1,2], mobility [3], physical activity [4], and fatigue [5].

The strength of EMA lies in its ability to minimize recall bias [6,7] and provide more fine-grained longitudinal data compared with traditional observation methods or retrospective reporting [8,9]. Participants can answer questions without traveling to a research lab, reducing a burden that is already great for individuals with Alzheimer disease and related dementias. Data collected through EMA reveal both within-day and between-day fluctuations in health and well-being [10], laying a foundation for timely, in-the-moment interventions. Furthermore, these responses provide a rich source of patient-reported outcome measures. When combined with objective, sensor-based data representing digital phenotypes, EMA enables the comparison of digital behavior markers with self-reported data.

Despite these advantages, EMA implementation faces challenges, especially in the variability, completeness, and accuracy of participant responses to prompts. Factors such as distraction, self-awareness, boredom, time of day, and interruption burden [11] can impact participant responses. Addressing these issues is essential for maintaining the integrity of research findings. Furthermore, the design of notification strategies may dramatically impact response compliance and quality [12,13].

In this paper, we conduct an examination of EMA response completeness and quality across a spectrum of settings. By leveraging data from multiple studies, we enhance the robustness and generalizability of our conclusions. This broad approach also allows us to consider variables previously unexplored in EMA research. By correlating EMA responses with sensor-derived biomarkers, we will better understand the contextual factors influencing participant engagement with EMA prompts. The findings from this study will inform refinements in EMA methodology and guide future research to improve the reliability of EMA-derived data and the effectiveness of just-in-time interventions.

Previous Work

In-the-moment sampling builds a foundation for designing effective just-in-time interventions [14]. McDevitt-Murphy et al [15] used such a technique to combat anxiety by prompting individuals to write about their worries and initiate physical activity. Similarly, Mair et al [16] found such interventions improved physical activity in older adults, while Perski et al [17] constructed a technology-driven in-the-moment intervention to reduce substance abuse and promote healthy behaviors. Brenner and Ben-Zeen [18] found that letting individuals see the difference between their predicted future EMA reports of mood and actual EMA responses improved the effectiveness of their intervention.

EMA benefits must be weighed with the cost of an increased interruption load. Frequent requests for self-report often lower responsiveness [13]. Many factors have been considered that impact responsiveness. These include the impact of EMA prompts while a person is performing physical activities [19,20] or activities that require concentration [21]. Jeong et al [22] and Rintala et al [23] studied the relationship between a person's location (eg, at home vs away from home) and responsiveness, while Pizza et al [24] and Ziesemer et al [25] found that engagement in social activity was an important consideration. Several studies consider whether prompt time of day influences likelihood of response [12,26-30]. Aminikhanghahi et al [31] found that prompting during transitions between activities improved participant responsiveness. Recently, investigators observed a relationship between the complexity of an EMA interaction and responsiveness [32]. They responded by shortening the questions [33], simplifying the interface [21], and making use of voice input [34].

Researchers also observed that the quality of responses decreases over time [35], becoming bistable or skewed [36], more habitual [37], and less variable [38-40]. Such trends threaten the reliability of research results. For example, Verbeij et al [41] observed decreased convergent validity between reported and objective social media use over 3 weeks of reporting [41].

These findings allow researchers to better design EMA-based studies and control for such factors in their analysis. One difficulty, though, is that some of the findings are contradictory. While Jeong et al [22], McMillan et al [27] and Ziesemer et al [25] reported higher response rates (RRs) when a person was alone, Pizza et al [24] observed no significant impact of social setting and Ponnada et al [26] reported a higher RR when a person was with family or friends. Regarding time of day, some studies found that participants responded more often during afternoons and evenings [26,28,42], while others [43,44] observed more responses early in the day. Khanshan et al [19] found RRs to be higher while participants were active, while Boukhechba et al [29] found higher activity was correlated with missing more responses. These discrepancies may indicate the influence of other population and study parameters. We address the need to consider these factors by performing a multistudy analysis.

Furthermore, Wrzus and Neubauer [45] emphasized the need to employ sensors in the understanding of the interplay between participant behavior, EMA responsiveness, and response quality. Stach et al [46] observe that mobile notification responsiveness is influenced not only by the app delivering the notification but also by user demographics. In our analysis, we leveraged continuously collected sensor data to examine the relationship between EMA responses (as patient-reported outcome measures), digital markers extracted from sensor data (digital phenotyping), and demographics across multiple studies. This comprehensive approach enabled us to uncover patterns and develop recommendations for optimizing EMA methodologies.

XSL•FO
RenderX

Study Goal and Hypotheses

In this study, we performed statistical analysis of EMA responses and sensor data from 9 clinical studies. The overall goal was to provide insights into EMA response patterns that can be used to design and improve future in-the-moment monitoring and intervention studies. We analyzed data from 454 participants to validate the following hypotheses:

First, EMA RR will vary with prompt time of day, number of questions, and sensor-observed activity.

Second, EMA response quality will decrease as a function of the number of days that have elapsed in the study.

Third, both RR and response quality will be impacted by a participant's behavior context, as modeled by markers extracted from mobile data.

Fourth, differences in EMA RRs and response patterns will exist between studies, influenced by participant demographics (ie, participant age) and study purpose (ie, cognitive intervention vs observational).

Methods

EMA Studies

Daily EMA responses were collected between October 2015 and December 2023 for a total of 454 participants across 9 distinct studies. Reporting of the EMA steps is consistent with the STROBE CREMAS (Strengthening the Reporting of Observational Studies in Epidemiology Checklist for Reporting Ecological Momentary Assessment Studies) guidelines [47]. In each study, participants received prompts to answer EMA questions using an in-house app running on a smartwatch or tablet. Participants were trained on how to interact with the app by a research assistant and completed practice prompts and tasks before data collection commenced.

The EMA prompts were randomly distributed within predefined time blocks throughout the day. We adjusted these time blocks when necessary to fit participant schedules. When a prompt was issued, an audio tone was played to alert the participant. The participant would then read the question and select their response from options on a mobile screen. If there was no response within 5 minutes, the prompt was reissued, up to a maximum of 5 attempts (2 attempts for dyad study). If the participant still did not respond, we recorded this as a nonresponse and did not issue further prompts until the next designated time block. Parameters for the studies are summarized in Table 1. In total, these comprise 145,853 EMA prompts with 115,283 sets of responses.

XSL•FO
RenderX

Table . Ecological momentary assessment studies that were included in the analysis.

Study	Population	Sample[a] (n)	Prompts (n)	Questions and tasks: number and topic	RR[b] (%)	Duration
Rural (mc)	Aged 50+ y, cognitive health spectrum, Florida	33	9200	8 (social contact, physical activity, mental activity, environment, and motivation), *n*-back[c]	77.79	2 weeks, day
Function (func)	Aged 50+ y, cognitive health spectrum, Pacific Northwest	26	17,923	11 (social contact, physical activity, mental activity, environment, and activity type), *n*-back, audio[d]	73.85	5 weeks, spaced over 16 months, day+night
Note-book (mn)	Aged 50+ y, cognitive health spectrum, Pacific Northwest	45	17,369	15 (memory strategy, confidence, activity type, environment, and emotion analysis), [*n*-back, audio[d]][e]	79.76	2 weeks, 6 months apart, day
Brain Booster (bb)	Aged 65+ y, subjective cognitive decline, California	184	3207	12 (environment, concentration, and emotion analysis), *n*-back 1x/day	54.35	2 weeks, 6 months apart, day
Compen-satory (cs)	Aged 50+ y, primarily cognitively healthy, Pacific Northwest	89	45,423	11 (activity type, environment, memory strategy, and location), *n*-back, audio[f]	85.09	2 weeks, day
Smart-Home (sh)	Aged 55+ y, cognitively healthy, Pacific Northwest	43	32,065	7 (social contact, physical activity, mental activity, and activity type), *n*-back	78.86	2 weeks, 1 month apart
Dyad (dyad)	Aged 55+ y, cognitively healthy, West Virginia and Virginia	10	5173	8 (dyad interactions and emotion analysis), audio[d]	91.51	2 weeks, day+night
Gsur1 (gsur1)	Students, cognitively healthy, Pacific Northwest	6	3727	12 (social contact, physical activity, mental activity, environment, and activity type), *n*-back	84.52	2 weeks, day
Gsur2 (gsur2)	Students, cognitively healthy, Pacific Northwest	18	12,666	15 (social contact, physical activity, mental activity, environment, activity type, location, and emotion analysis), *n*-back, audio[d]	59.01	2 weeks, day

[a]Number of participants enrolled in the study (through December 2023).

[b]RR: response rate.

[c]*n*-back is a 45-second 1-back shape task.

[d]Audio prompt to describe their day.

[e]Additional tasks introduced late in the study (31.2% of mn participants).

[f]Audio prompt to describe what tasks were assisted by memory strategies that day.

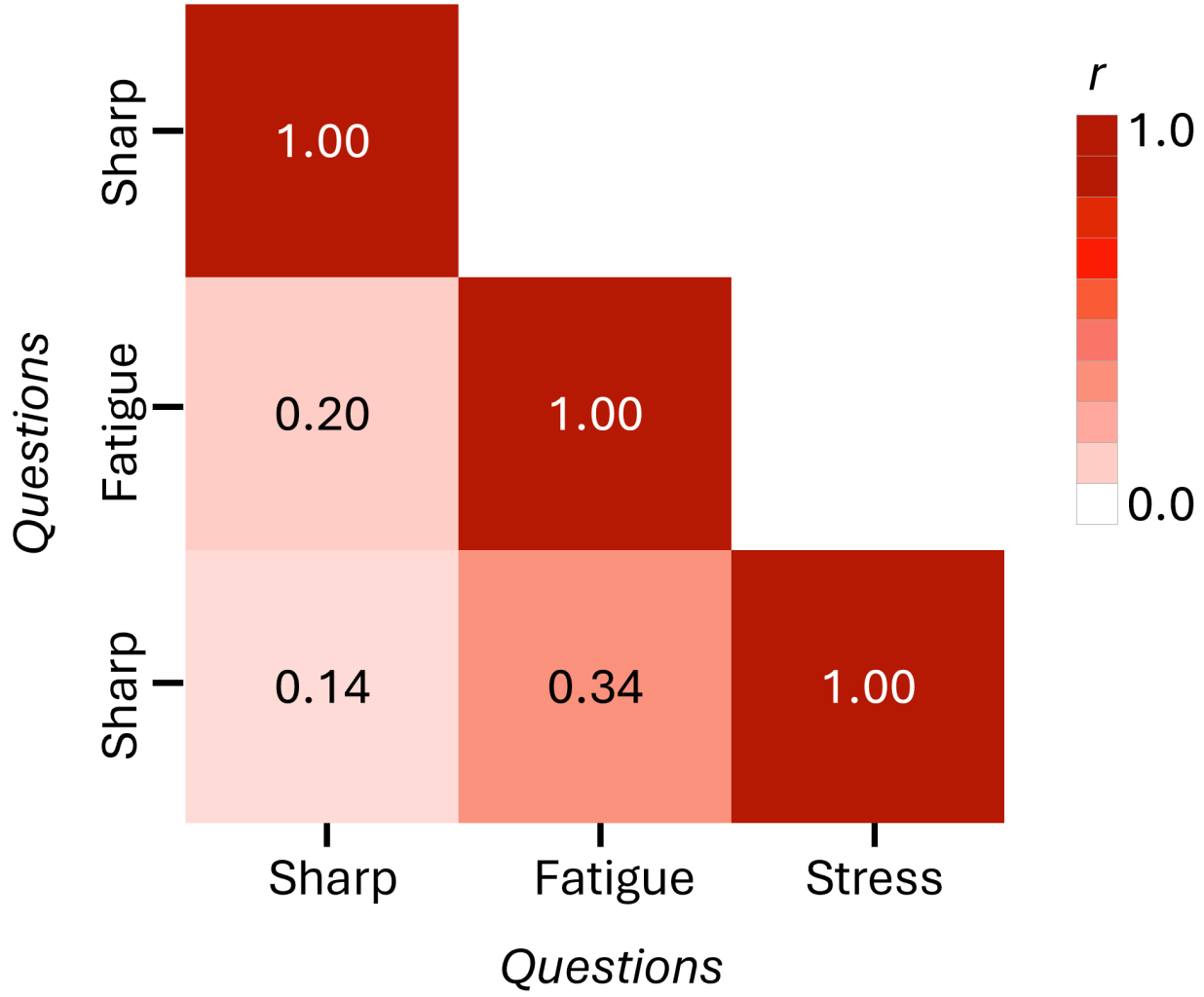
Well-being has been defined multiple ways in the literature [48-51]. Many studies rely on 1-time questionnaires, like the Satisfaction With Life Scale [52], to measure well-being at a single time point. However, well-being is dynamic, fluctuating

XSL•FO
RenderX

across different times and contexts [53]. EMA allows us to capture individuals' daily behaviors within their natural environments alongside real-time self-reports on well-being. In this analysis, we focus on three dimensions of well-being: mental sharpness, physical fatigue, and stress.

Each study included subjective questions that examined (1) Qsharp = the person's mental sharpness, (2) Qfatigue = the person's physical fatigue, and (3) Qstress = the person's level of stress, all self-reported and relevant to the point in time (eg, immediately and up to the previous 2 h) when they received the query. The Likert scales varied between studies, so we normalized the values to a 1 - 5 range. Where needed, we inverted responses for negative questions so that a higher-valued response is more positive in each case (eg, Qsharp=5 indicates feeling greater amounts of mental sharpness; Qfatigue=5 indicates feeling less physically fatigued; Qstress=5 indicates feeling less stressed. Multimedia Appendix 1 plots the relative frequency of each response to the 3 question categories. The studies are ordered primarily by the cognitive health status of participants and secondarily by age. As the plots demonstrate, most of the responses are high but do vary among studies. Notably, the similarity in response between stress and fatigue is higher than other question pairs. This finding is consistent with the correlation that exists between responses to the 3 questions, shown in Figure 1.

Figure 1. Correlation between participant responses to the three question types.



Ethical Considerations

These studies were reviewed and approved by the institutional review boards (IRB protocol 14460) at Washington State University and the University of California at Davis (No. 19784). To participate in any of the studies, participants needed to understand English and sign an informed consent; each person received compensation between US $20 and US $125 for their participation. All data were anonymized before performing analyses.

Statistical Analysis

RR and Quality

We conducted statistical analyses on the data to examine RR, completeness, and quality for each participant and study. Our app records each time a prompt is delivered to the participant. From these records, we calculate RR as the ratio of occurrences when a participant answered at least 1 question to the total number of prompts. The analysis also considers the prompt time of day (morning before 12 PM, afternoon between 12 PM and 5 PM, and evening after 5 PM), day of week, study duration,

XSL•FO
RenderX

and number of questions per prompt. Researchers suggest that user demographics may impact preference for EMA delivery strategies [21]. We examine this factor by considering differences among the studies and between the set of studies involving older adults (aged 50 y or older) and those involving younger adults.

Response completeness reflects the percentage of questions that were answered for the occasions when the participant responded. We excluded nonquestion tasks, such as daily *n*-back shape tests (all but dyad) and audio descriptions of the day's activities (func, mn, cs, dyad, gsur1, and gsur2). Rather than impute missing values, analysis of EMA-reported values only considers completed responses.

Response quality measures habituation that may occur over the duration of the study, resulting in less response variance and state awareness. As reported elsewhere [54-56], we quantified careless responses as the percent responses that fall near (within 0.2) each participant's mode response and also calculated response variance. Changes in response quality may occur over the course of the study as a motivation to provide high-quality responses wanes. Thus, we computed the change in variation and careless responses between the first and second half of the study for each participant.

To gather sensor-derived behavior context, we collected continuous sensor data from Apple watches for 95 participants in the bb study and all participants from the dyad, func, cs, mn, mc, sh, gsur1, and gsur2 studies. In the ihs study, we collected data from ambient sensors installed in participant homes. The smartwatches collected acceleration and rotation readings at 10 Hz and recorded location (latitude, longitude, and altitude) every 5 minutes, or more frequently when movement exceeded a baseline. Data were stored on the watch, encrypted, and downloaded to a password-protected server when the watches were returned. From the raw location data, we defined the participant's home as the location visited most often between 2 AM and 9 AM. We then calculated the distance and bearing from home for each location. Behavior context features were defined as "activity level" (total acceleration) and time "at home" (fraction of time spent at home), calculated for 30 minutes ending at the time the participant provides an EMA response. The data collection app was developed in-house using Swift. Collecting data at these rates ensured 1 full day of data on a single charge.

For studies collecting data only during the day (Table 1), participants wore the watch on their nondominant arm during the day and charged it at night. In studies collecting both day and night data, participants were given 2 watches—one worn during the day and charged at night, and the other worn at night and charged during the day. In the smart home study, EMA responses were reported using a tablet; in all other studies, they were reported via the smartwatch.

In the smart home study, 2‐4 sensors were placed in each room. These were either passive infrared motion sensors with an ambient light sensor or magnetic door sensors with an ambient temperature sensor. Motion sensors detected movement, providing location data within the home. Activity level was estimated by counting motion sensor readings. If an external door was opened and closed, followed by at least 5 minutes of no motion sensor activity, the participant was designated as out of home. The "at home" context quantified the fraction of time spent at home in the 30 minutes before an EMA response.

Another context feature was proximity to a change point. A change point occurs when there is a shift in the underlying process of a time series, such as transitioning from one activity to another. In an earlier work by Aminikhanghahi et al [57], we designed an algorithm to detect these points from continuous sensor data and found that interacting with participants during these times improved RRs and task success. We used the same method here to determine proximity to a change point.

We calculated correlations between the 3 context features (activity level, at home, and change point) and response compliance. Here, a lack of response to a particular prompt is assigned a "compliance" value of 0.0, a complete response is assigned a value of 1.0, and a partial response (some questions answered, some not) is assigned in the range 0.0‐1.0, corresponding to the relative number of questions that were answered.

Study Differences

In this analysis, we selected 9 studies as the basis for comparing EMA delivery strategies and responses. The selection was based on the consistency of the technical data collection tools and hardware, which minimized the influence of these factors on the results. A single software interface was used to deliver EMA prompts across all studies, and the same app collected sensor data in all smartwatch studies, following a protocol similar to that used in the smart home study. Context features were extracted uniformly from the sensor data across datasets.

Participants in all studies were instructed to perform their normal daily routines without changing their behavior to accommodate the data collection. This consistency allowed us to focus on the differences that existed between the settings. Specifically, the participant groups differed across studies in terms of location (4 regions within the United States), age, and cognitive health, supporting generalization of the findings to other EMA studies. The studies also differed in length and number of EMA questions and tasks. These differences represent the focus of our analysis.

We examined differences in EMA response dynamics across the 9 studies. Our approach is twofold; we report results for each study individually and collectively and analyze differences in EMA responses within and between these studies. This technique will allow us to examine the extent to which EMA response patterns differ between studies, using within-study changes as a comparison baseline. We use dynamic time warping (DTW) as a similarity measure because the technique measures similarity between time series that differ in length as well as values. In the case of the Qsharp question, we also visually compare responses over time for studies that focus on cognitive health support for older adults (bb and mn) with observational studies involving cognitively healthy adults (dyad, gsur1, and gsur2).

To deepen the analysis, we fit ARIMA (Autoregressive Integrated Moving Average) models to the aggregated EMA

XSL•FO
RenderX

responses for each question type. The mean response was computed across participants for the first 48 responses, producing a single representative time series. The resulting model parameters indicated how well the shape of responses over time could be modeled.

Results

RR Findings

RR across all participants was 79.95%. This is within 0.76% of the 79.19% average rate Wrzus and Neubauer [45] reported in a survey of 417 EMA studies. Among the prompts that received participant responses, the response completion rate was 88.37%.

Figure 2 plots RR for the studies, ordered from largest to smallest. Response variance is quite different among the studies. Notably, response variance increased as the mean RR decreased. Table 2 shows the overall most popular time of day for responding to EMA prompts. However, these results vary based on participant age. Only considering studies with older adults (age 50 y or older), participants were more responsive in the evening than morning (by 3.95) or afternoon (by 3.58). In contrast, younger adult RR remained fairly uniform throughout the day, with afternoon showing the highest rate. Weekday and weekend prompts yielded very similar RR overall (a difference of 0.04). However, this varied by study and age group. Specifically, while weekend response was higher by 3.25 for younger adults, the rate decreased for older adults by 0.27. These findings highlight the role that participant demographics play in understanding EMA responsiveness; prompt time alone may not be a strong indicator of responsiveness.

Figure 2. Response rate mean and variance. Rates are shown in sorted order; variance is multiplied by 10 to align the plots.

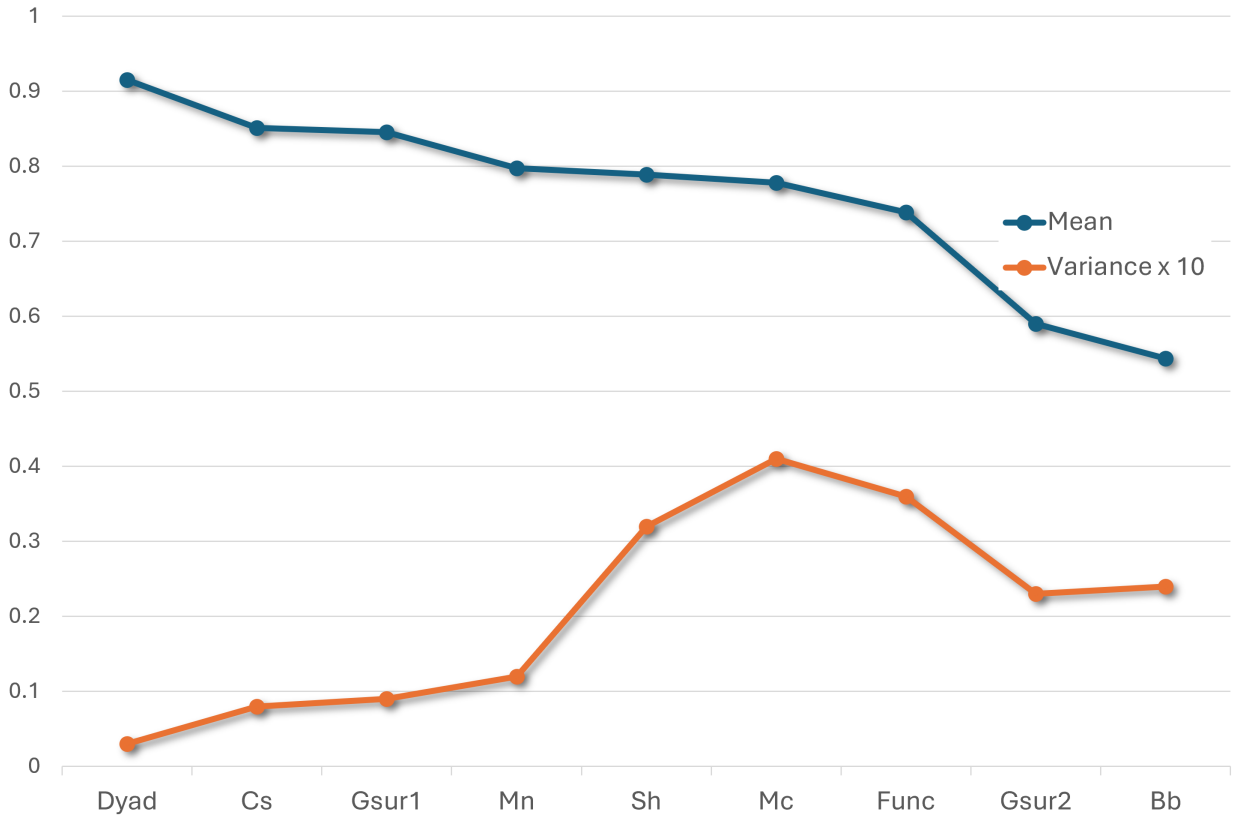


Table . Response rate (%) based on prompt time of day and day of week.

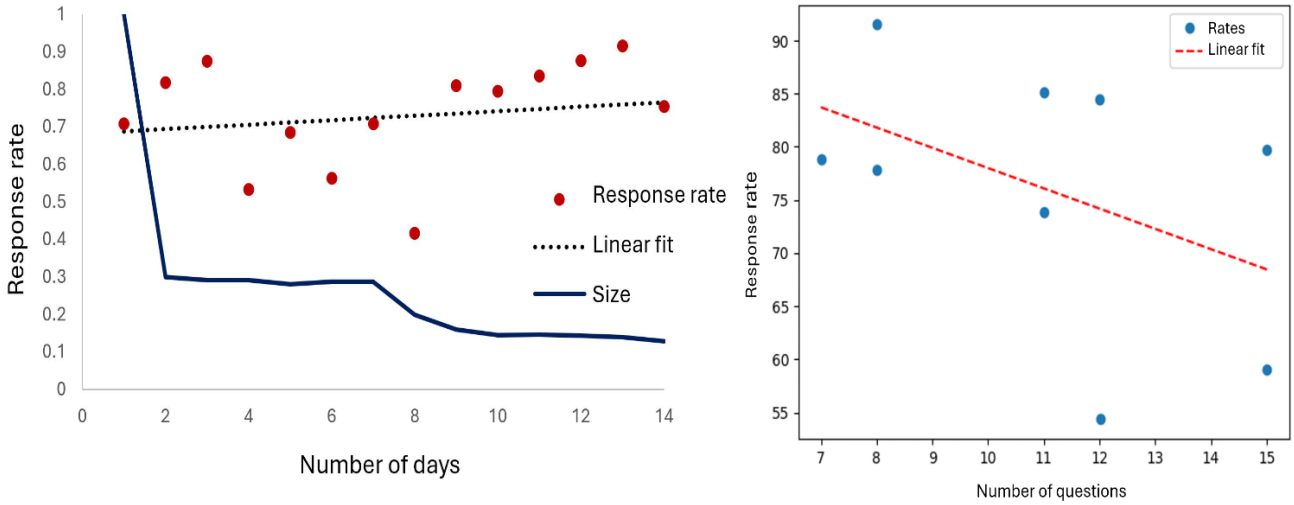
Time		Overall response rate (%)	Response rate among older adults (%)	Response rate among younger adults (%)
Time of day				
	Morning	79.02	80.60	64.86
	Afternoon	79.08	80.97	64.97
	Evening	82.31	84.55	64.45
Day of week				
	Weekday	80.43	82.46	63.93
	Weekend	80.39	82.19	67.18

XSL•FO
RenderX

Figure 3 (left) plots EMA RR as a function of the number of days that elapsed since the beginning of the study for each participant. The linear fit line has a positive slope of 0.078, indicating that the duration does not negatively impact RR. At the same time, the fraction of participants still collecting data is much smaller at the end of this time span. A nonsignificant small positive correlation exists between the study duration and RR (r=0.100, P=.29). This may be due to the EMA prompts becoming more integrated into the participant's routine and their gaining familiarity with the questions and methods of responding.

Figure 3 (right) shows RR as a function of the number of questions asked at the EMA prompt, together with the fitted line. Correlation between the number of questions asked during an EMA prompt and RR is moderate and negative (r=−0.433, P<.001). The line fitting this trend has a slope of −1.90. This finding is consistent with previous work indicating that individuals are less likely to respond when each EMA prompt requires a large time interruption and corresponding task burden [37].

Figure 3. Left: response rate based on number of study days and relative sample size at that time point (fraction of original sample). Right: response rate as a function of the number of questions asked at the EMA prompt.



Response Quality

Table 3 summarizes participant responses and response quality. The average response value across all studies is relatively high, as is the percentage of responses near the mode. There was a noticeable shift toward response uniformity over time, evidenced by a significant increase in carelessness and decrease in variance. Results from individual studies are shown in Multimedia Appendix 2. Although the earlier analysis showed that RR did not decline over the study, the increase in careless responses and decrease in variance indicates a possible increase in habitual responses to EMA questions [37] as participants got bored or felt time pressure to respond to EMA notifications. Some studies inserted a break of 1+ months during data collection. For these studies, we calculated changes within the period before the break and the period after the break. The trends remain the same: carelessness increased and variance decreased over time. These changes are smaller, possibly due to the decreased time (typically 7 d) being examined.

XSL•FO
RenderX

Table . Summary of response averages and quality.

Metric	Value
Qsharp average response (variance)	3.524 (1.072)
Qfatigue average response (variance)	4.157 (0.758)
Qstress average response (variance)	4.432 (0.652)
Careless %	55.71
Interparticipant variance	0.611
Carelessness change	0.22 (*P*<.001)
Variance change	−0.363 (*P*<.001)
Prebreak carelessness change[a]	0.04 (*P*<.001)
Prebreak variance change[a]	−0.082 (*P*<.001)
Postbreak carelessness change[a]	0.09 (*P*<.001)
Postbreak variance change[a]	−0.329 (*P*<.001)

[a]func, mn, bb, and sh studies.

Alignment With Sensor-Observed Behavior

Relationships between RRs and sensor-detected states were also examined. A correlation of r=0.045 (P<.001) exists between RR and activity level over the previous 30 minutes, r=0.174 (P<.001) between RR and time spent at home over the previous 30 minutes, and r=0.124 (P<.001) between RR and nearness to detected activity change points. Multimedia Appendix 3 summarizes the results for individual studies.

RR correlation with activity level is positive but small. This finding, considered together with the discrepancies observed in previous studies [19,29], may indicate that the nature of the person's activity needs to be examined to understand the impact on prompt response. Correlations with at-home time and change points were larger, although still small. Individuals may not want to interrupt their activities, particularly in social situations, to answer lengthy EMA questions. In these contexts, they may ignore the prompt and wait until the next time block. As with our previous study examining RR and change points in smart homes [31], we observe that prompting near a change point improve responsiveness for all studies.

EMA Differences Between Studies

Results from the previous analyses reveal that differences in EMA response compliance and quality may occur as a function of study and participant characteristics. Here, we delve further into EMA response differences between studies. First, we measure the distance (inverse similarity) between EMA time series using DTW. Across all studies and participants, the average within-study DTW distance was 14.141 (SD 11.957) overall (sharp 15.357, highest dyad=36.111, lowest bb=11.486; fatigue 15.014, highest dyad=36.489, lowest bb=10.299; stress 12.051; and highest dyad=33.511, lowest bb=7.069).

In contrast, the average between-study DTW distance was 33.246 (SD 4.971) overall (sharp 44.848, highest dyad/bb=81.448, lowest mc/bb=16.262; fatigue 29.365, highest mn/func=54.927, lowest mc/bb=12.434; stress 25.526, highest mn/func=52.05, lowest mc/bb=8.759). This represents a 135% distance increase over the within-study distances. The large increase indicates that while response variability exists between participants, an even larger difference exists between studies. Table 1 lists some of the parameters that may influence these differences; further investigation may identify more contributing factors.

Plots of the aggregated EMA responses for the 3 question types are shown in Figure 4. The shapes of the aggregated time series are different between the 3 question categories. The Qsharp trend is consistently positive, while there are less noticeable trends for the other questions.

Because 2 of the studies (bb and mn) represent interventions, we examine differences in the response trends between these studies and the 2 observational studies with cognitively healthy older adults (dyad and cs). Aggregated over all 9 studies, the slope of the response fitted line is 0.018, with a y-intercept of 3.181. The fitted line for the intervention studies has a slope of 0.023 (an increase of 0.005) and y-intercept of 2.745 (a decrease of 0.436), while for observational studies with cognitively healthy older adults, the fitted line has a slope of 0.005 (a decrease of 0.013) and y-intercept of 3.449 (an increase of 0.614). Here, we observe that older adults receiving interventions start at a lower self-reported mentally sharp value but indicate that the sharpness increases over the study duration. In comparison, the cognitively healthy adults start at a higher sharp reported value and have a lesser amount of change in reported values during the study.

XSL•FO
RenderX

Figure 4. Aggregated ecological momentary assessment response sequences for (top left) Qsharp, (top middle) Qfatigue, and (top right) Qstress. Aggregated ecological momentary assessment responses to the Qsharp question for (bottom left) cognitive health intervention studies and (bottom right) cognitively healthy adult studies. All Qsharp plots are accompanied by a linear fit line.

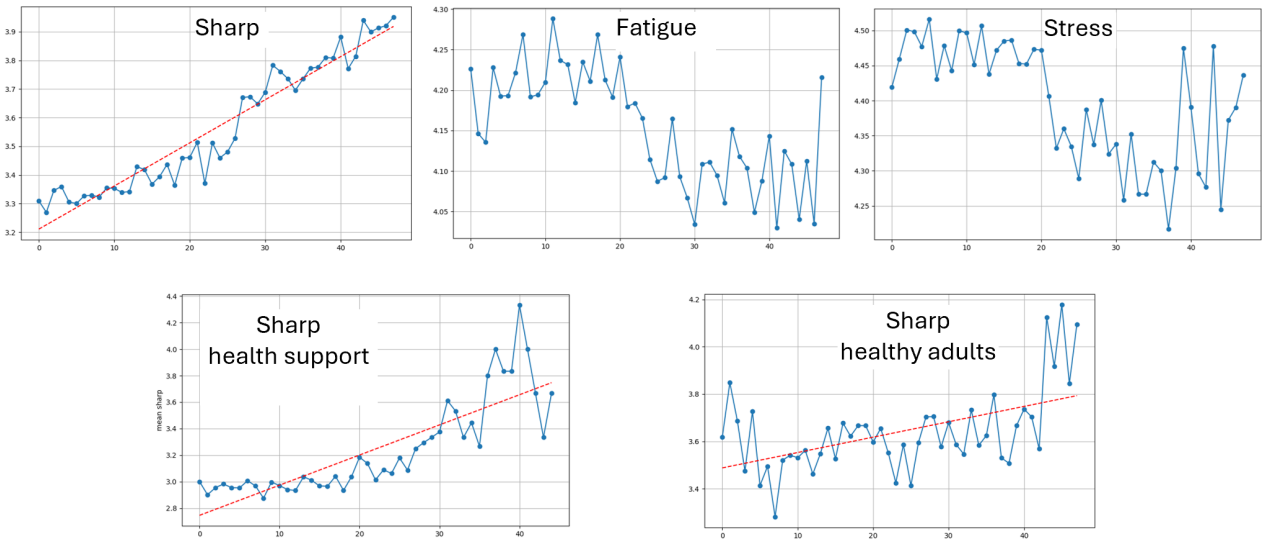


Table 4 presents the ARIMA parameters for the response categories sharp, fatigue, and stress. The log-likelihood values are high for all models, indicating a good fit. Similarly, the negative Akaike Information Criterion and Bayesian Information Criterion values suggest that the models strike a balance between fitting the data well and avoiding unnecessary complexity. The *P* values for the Moving Average of order 1 terms (*P*<.001 for all categories) indicate that errors at lag-1 are significant for predicting the next value in the time series. In addition, the Ljung-Box test results, with values greater than 0.05, suggest that the residuals are not significantly autocorrelated, meaning the models capture relevant patterns beyond mere autocorrelation.

Table . ARIMA (Autoregressive Integrated Moving Average) parameters for response categories.

Metric	Sharp	Fatigue	Stress
Log likelihood (higher is better)	74.19	72.39	64.09
AIC[a] (lower is better)	−142.39	−140.79	−124.17
BIC[b] (lower is better)	−136.84	−137.09	−120.47
MA(1)[c]	−0.63 (*P*<.001)	−0.67 (*P*<.001)	−0.61 (*P*<.001)
Ljung-Box test	0.85	0.21	0.07

[a]AIC: Akaike Information Criterion.

[b]BIC: Bayesian Information Criterion.

[c]MA(1): Moving Average of order 1.

Discussion

Principal Results

Early detection and intervention, crucial for reducing health care costs and enhancing quality of life, depend on timely reports of health status. This cross-study analysis highlights the pivotal role of EMA design in obtaining frequent, accurate EMA responses. In particular, we observed that participant demographics and study parameters markedly influence EMA response values, compliance, and quality. In particular, timing of EMA prompts varied in effectiveness depending on participant age. On the other hand, the negative impact of lengthy EMA interactions was consistent across all of the studies. Supporting eyes-free or hands-free interactions [21,34] may reduce interruption burden. Similarly, employing more frequent but shorter EMA sessions, or µEMA [26], may improve responsiveness. These findings support the first hypothesis and highlight the need to tailor EMA prompt strategies to specific participant and study characteristics.

A consistent observation was an increase in habitual, careless responses over the study duration, supporting the second hypothesis. In this study, 2087 of the 17,939 sessions (11.63%) were started but not completed, so researchers may consider strategies such as randomizing question or task order [58], gamifying the data collection process [8], or allowing selected questions to be skippable [59], as some may hold more interest. Additional factors, such as time pressure from concurrent activities or individual personality traits [55], may also influence responses. Ultimately, these considerations can be included in the study design.

XSL•FO
RenderX

Integrating sensor-based behavioral context is key to improving EMA studies. As hypothesized in the third hypothesis, our analysis revealed a significant relationship between sensor-derived biomarkers and EMA responsiveness. Analyzing these biomarkers could benefit personalized EMA design. In addition, timing EMA prompts close to detecting activity transitions may improve RR and quality.

Finally, our results showed that consistent with the fourth hypothesis, EMA response dynamics varied according to study parameters such as participant age and study type. We observed that response values increased more dramatically for intervention than observational studies, perhaps due to participants viewing their intervention involvement as beneficial, as noted in previous literature [60]. The response series were well fit by statistical models such as ARIMA. Such models provide a basis to compare EMA trajectories between groups and predict future response patterns.

Limitations

This study had several limitations. Although we examined factors like study purpose and participant age, other variables, such as education, race or ethnicity, employment status, environment, and cognitive or physical health, were not examined. In addition, we did not consider the impact of breaks of 1 or more days in data collection. Restarting data collection after a break could affect EMA compliance and response quality.

Another limitation involves the duration of the studies. While some studies spanned 6 months or more, others lasted only 2 weeks. Extending the duration of all studies would enhance the generalizability of our time series findings. The sample sizes also varied considerably, with smaller young adult samples compared with older adult studies. In addition, slight differences in EMA question wording between studies, though normalized, may have influenced responses. Greater uniformity in future studies would provide more consistent data.

A further limitation was the lack of diversity in medical conditions. While we focused on individuals with varying levels of cognitive decline, other physical or psychiatric conditions were not represented. Including a wider range of medical conditions in follow-up studies would provide a more comprehensive understanding of EMA responses across different health conditions.

While the studies were consistent in their data collection hardware and software, technology influences results through factors like accessibility, battery life, ease of use, device familiarity, and extractable digital markers. Future studies can systematically consider alternative technologies.

Finally, this analysis concentrated on a subset of well-being factors. Since well-being encompasses broader emotional, social, and physical domains, future research can include additional dimensions for a more holistic analysis. In particular, a next step can be to evaluate the impact of rural versus urban residence as well as the quality of the build environment on EMA responses.

Conclusions

This study offers insights into the factors that influence EMA responses and lays the groundwork for further exploration. Future studies should broaden the range of participant, study, and interaction characteristics to better inform EMA design strategies. In addition, expanding the range of sensor-based behavioral markers will enhance understanding of how changes in routine and behavior can enrich EMA data collection. Addressing these areas will help refine EMA methodologies, leading to more personalized and effective health monitoring and intervention technologies.

Acknowledgments

This work is supported in part by NIA grants R01AG066748, R01AG065218, and R35AG071451. The work was also supported in part by DOD grant AZ 190055.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bar charts illustrating the distribution of Likert responses for the sharp, fatigue, and stress ecological momentary assessment questions across studies. Higher scores are consistent with better well-being.
[PNG File, 53 KB - mhealth_v13i1e57018_app1.png]

Multimedia Appendix 2
Left: percentage careless responses (within 0.2 of participant mode) and interparticipant variance (multiplied by 100 to normalize scale). Right: difference in percentage near the mode and variance between the first and second half of data collection for each participant.
[PNG File, 468 KB - mhealth_v13i1e57018_app2.png]

Multimedia Appendix 3
Pearson correlation between ecological momentary assessment response and behavior context.
[PNG File, 213 KB - mhealth_v13i1e57018_app3.png]

XSL•FO
RenderX

References

1. Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. JMIR Mhealth Uhealth 2021 Jul 12;9(7):e26540. [doi: 10.2196/26540] [Medline: 34255713]
2. Kraiss JT, Kohlhoff M, Ten Klooster PM. Disentangling between- and within-person associations of psychological distress and mental well-being: an experience sampling study examining the dual continua model of mental health among university students. Curr Psychol 2022 Mar 1;42(20):1-12. [doi: 10.1007/s12144-022-02942-1] [Medline: 35250243]
3. Kheirkhahan M, Nair S, Davoudi A, et al. A smartwatch-based framework for real-time and online assessment and mobility monitoring. J Biomed Inform 2019 Jan;89:29-40. [doi: 10.1016/j.jbi.2018.11.003] [Medline: 30414474]
4. Knell G, Gabriel KP, Businelle MS, Shuval K, Wetter DW, Kendzor DE. Ecological momentary assessment of physical activity: validation study. J Med Internet Res 2017 Jul 18;19(7):e253. [doi: 10.2196/jmir.7602] [Medline: 28720556]
5. Rao C, Di Lascio E, Demanse D, Marshall N, Sopala M, De Luca V. Association of digital measures and self-reported fatigue: a remote observational study in healthy participants and participants with chronic inflammatory rheumatic disease. Front Digit Health 2023;5:1099456. [doi: 10.3389/fdgth.2023.1099456] [Medline: 37426890]
6. Younis EMG, Kanjo E, Chamberlain A. Designing and evaluating mobile self-reporting techniques: crowdsourcing for citizen science. Pers Ubiquit Comput 2019 Apr;23(2):329-338. [doi: 10.1007/s00779-019-01207-2]
7. de Brito JN, Loth KA, Tate A, Berge JM. Associations between parent self-reported and accelerometer-measured physical activity and sedentary time in children: ecological momentary assessment study. JMIR Mhealth Uhealth 2020 May 19;8(5):e15458. [doi: 10.2196/15458] [Medline: 32348283]
8. Doherty K, Balaskas A, Doherty G. The design of ecological momentary assessment technologies. Interact Comput 2020 May 17;32(3):257-278. [doi: 10.1093/iwcomp/iwaa019]
9. van Berkel N, Ferreira D, Kostakos V. The experience sampling method on mobile devices. ACM Comput Surv 2018 Nov 30;50(6):1-40. [doi: 10.1145/3123988]
10. Rey B, Lee B, Choe EK, Irani P. Investigating in-situ personal health data queries on smartwatches. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022 Dec 21;6(4):1-19. [doi: 10.1145/3569481]
11. Howard AL, Lamb M. Compliance trends in a 14-week ecological momentary assessment study of undergraduate alcohol drinkers. Assessment 2024 Mar;31(2):277-290. [doi: 10.1177/10731911231159937] [Medline: 36914966]
12. van Berkel N, Goncalves J, Hosio S, Sarsenbayeva Z, Velloso E, Kostakos V. Overcoming compliance bias in self-report studies: a cross-study analysis. Int J Hum Comput Stud 2020 Feb;134:1-12. [doi: 10.1016/j.ijhcs.2019.10.003]
13. Möller A, Kranz M, Schmid B, Roalter L, Diewald S. Investigating self-reporting behavior in long-term studies. Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; Apr 27 to May 2, 2013; Paris, France p. 2931-2940. [doi: 10.1145/2470654.2481406]
14. Balaskas A, Schueller SM, Cox AL, Doherty G. Ecological momentary interventions for mental health: a scoping review. PLoS ONE 2021;16(3):e0248152. [doi: 10.1371/journal.pone.0248152] [Medline: 33705457]
15. McDevitt-Murphy ME, Luciano MT, Zakarian RJ. Use of ecological momentary assessment and intervention in treatment with adults. Focus (Am Psychiatr Publ) 2018;16(4):370-375. [doi: 10.1176/appi.focus.20180017] [Medline: 31191181]
16. Mair JL, Hayes LD, Campbell AK, Buchan DS, Easton C, Sculthorpe N. A personalized smartphone-delivered just-in-time adaptive intervention (JitaBug) to increase physical activity in older adults: mixed methods feasibility study. JMIR Form Res 2022 Apr 7;6(4):e34662. [doi: 10.2196/34662] [Medline: 35389348]
17. Perski O, Hébert ET, Naughton F, Hekler EB, Brown J, Businelle MS. Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. Addiction 2022 May;117(5):1220-1241. [doi: 10.1111/add.15687] [Medline: 34514668]
18. Brenner CJ, Ben-Zeen D. Affective forecasting in schizophrenia: comparing predictions to real-time ecological momentary assessment (EMA) ratings. Psychiatr Rehabil J 2014;37(4):316-320. [doi: 10.1037/pjr0000105] [Medline: 25496200]
19. Khanshan A, Van Gorp P, Nuijten R, Markopoulos P. Assessing the influence of physical activity upon the experience sampling response rate on wrist-worn devices. Int J Environ Res Public Health 2021 Oct 10;18(20):10593. [doi: 10.3390/ijerph182010593] [Medline: 34682339]
20. Mishra V, Caine K, Lowens B, Kotz D, Lord S. Investigating contextual cues as indicators for EMA delivery. Proc ACM Int Conf Ubiquitous Comput 2017 Sep;2017:935-940. [doi: 10.1145/3123024.3124571] [Medline: 29666845]
21. Yan X, Raj S, Huang B, Park S, Newman M. Toward lightweight in-situ self-reporting: an exploratory study of alternative smartwatch interface designs in context. ACM Trans Interactive, Mobile, Wearable Ubiquitous Technol 2020;4(4):1-22. [doi: 10.1145/3432212]
22. Jeong H, Kim H, Kim R, Lee U, Jeong Y. Smartwatch wearing behavior analysis: a longitudinal study. ACM Trans Interactive, Mobile, Wearable Ubiquitous Technol 2017;1(3):1-31. [doi: 10.1145/3131892]
23. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Momentary predictors of compliance in studies using the experience sampling method. Psychiatry Res 2020 Feb 29;286:112896. [doi: 10.1016/j.psychres.2020.112896] [Medline: 32146247]
24. Pizza S, Brown B, McMillan D, Lampinen A. Smartwatch in vivo. Presented at: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-16, 2016; San Jose, CA p. 5456-5469. [doi: 10.1145/2858036.2858522]

XSL•FO
RenderX

25. Ziesemer K, König LM, Boushey CJ, et al. Occurrence of and reasons for "Missing Events" in mobile dietary assessments: results from three event-based ecological momentary assessment studies. JMIR Mhealth Uhealth 2020 Oct 14;8(10):e15430. [doi: 10.2196/15430] [Medline: 33052123]
26. Ponnada A, Li J, Wang SD, et al. Contextual biases in microinteraction ecological momentary assessment (μEMA) non-response. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022 Mar;6(1):1-24. [doi: 10.1145/3517259] [Medline: 39866712]
27. McMillan D, Brown B, Lampinen A, McGregor M, Hoggan E, Pizza S. Situating wearables: smartwatch use in context. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, Colorado, USA p. 3582-3594. [doi: 10.1145/3025453.3025993]
28. Visuri A, Sarsenbayeva Z, van Berkel N, et al. Quantifying sources and types of smartwatch usage sessions. Presented at: CHI '17; May 6-11, 2017; Denver, Colorado, USA p. 3569-3581. [doi: 10.1145/3025453.3025817]
29. Boukhechba M, Cai L, Chow PI, et al. Contextual analysis to understand compliance with smartphone-based ecological momentary assessment. Presented at: PervasiveHealth '18; May 21-24, 2018; New York, NY, USA p. 232-238. [doi: 10.1145/3240925.3240967]
30. Sun J, Rhemtulla M, Vazire S. Eavesdropping on missing data: what are university students doing when they miss experience sampling reports? Pers Soc Psychol Bull 2021 Nov;47(11):1535-1549. [doi: 10.1177/0146167220964639] [Medline: 33342369]
31. Aminikhanghahi S, Schmitter-Edgecombe M, Cook DJ. Context-aware delivery of ecological momentary assessment. IEEE J Biomed Health Inform 2020 Apr;24(4):1206-1214. [doi: 10.1109/JBHI.2019.2937116] [Medline: 31443058]
32. Ponnada A, Haynes C, Maniar D, Manjourides J, Intille S. Microinteraction ecological momentary assessment response rates: effect of microinteractions or the smartwatch? Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep;1(3):92. [doi: 10.1145/3130957] [Medline: 30198012]
33. Ponnada A, Wang S, Chu D, Do B, Dunton G, Intille S. Intensive longitudinal data collection using microinteraction ecological momentary assessment: pilot and preliminary results. JMIR Form Res 2022 Feb 9;6(2):e32772. [doi: 10.2196/32772] [Medline: 35138253]
34. Kim YH, Chou D, Lee B, et al. MyMove: facilitating older adults to collect in-situ activity labels on a smartwatch with speech. Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022; New Orleans, LA, USA. [doi: 10.1145/3491102.3517457]
35. Eisele G, Vachon H, Lafit G, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. Assessment 2022 Mar;29(2):136-151. [doi: 10.1177/1073191120957102] [Medline: 32909448]
36. Haslbeck JMB, Ryan O. Recovering within-person dynamics from psychological time series. Multivariate Behav Res 2022;57(5):735-766. [doi: 10.1080/00273171.2021.1896353] [Medline: 34154483]
37. Eisele G, Vachon H, Lafit G, et al. A mixed-method investigation into measurement reactivity to the experience sampling method: the role of sampling protocol and individual characteristics. Psychol Assess 2023 Jan;35(1):68-81. [doi: 10.1037/pas0001177] [Medline: 36174163]
38. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. In: Flow and the Foundations of Positive Psychology: Springer; 2014:35-54. [doi: 10.1007/978-94-017-9088-8_3]
39. Fuller-Tyszkiewicz M, Skouteris H, Richardson B, Blore J, Holmes M, Mills J. Does the burden of the experience sampling method undermine data quality in state body image research? Body Image 2013 Sep;10(4):607-613. [doi: 10.1016/j.bodyim.2013.06.003]
40. Vachon H, Bourbousson M, Deschamps T, et al. Repeated self-evaluations may involve familiarization: an exploratory study related to ecological momentary assessment designs in patients with major depressive disorder. Psychiatry Res 2016 Nov 30;245:99-104. [doi: 10.1016/j.psychres.2016.08.034] [Medline: 27541343]
41. Verbeij T, Pouwels JL, Beyens I, Valkenburg PM. The accuracy and validity of self-reported social media use measures among adolescents. Comput Hum Behav Rep 2021 Jan;3:100090. [doi: 10.1016/j.chbr.2021.100090]
42. Maher JP, Rebar AL, Dunton GF. Ecological momentary assessment is a feasible and valid methodological tool to measure older adults' physical activity and sedentary behavior. Front Psychol 2018;9:1485. [doi: 10.3389/fpsyg.2018.01485] [Medline: 30158891]
43. Silvia PJ, Kwapil TR, Eddington KM, Brown LH. Missed beeps and missing data: dispositional and situational predictors of nonresponse in experience sampling research. Soc Sci Comput Rev 2013;31(4):471-481. [doi: 10.1177/0894439313479902]
44. van Berkel N, Goncalves J, Koval P, et al. Context-informed scheduling and analysis: improving accuracy of mobile self-reports. Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 51. [doi: 10.1145/3290605.3300281]
45. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023 Apr;30(3):825-846. [doi: 10.1177/10731911211067538] [Medline: 35016567]
46. Stach M, Mulansky L, Reichert M, Pryss R, Beierle F. Call to action: investigating interaction delay in smartphone notifications. Sensors (Basel) 2024 Apr 19;24(8):2612. [doi: 10.3390/s24082612] [Medline: 38676228]

XSL•FO
RenderX

47. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE Checklist for Reporting EMA Studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151. [doi: 10.2196/jmir.4954] [Medline: 27328833]
48. Diener E, Lucas RE, Oishi S. Advances and open questions in the science of subjective well-being. Collabra Psychol 2018;4(1):15. [doi: 10.1525/collabra.115] [Medline: 30637366]
49. Baselmans BML, van de Weijer MP, Abdellaoui A, et al. A genetic investigation of the well-being spectrum. Behav Genet 2019 May;49(3):286-297. [doi: 10.1007/s10519-019-09951-0] [Medline: 30810878]
50. Bajaj B, Pande N. Mediating role of resilience in the impact of mindfulness on life satisfaction and affect as indices of subjective well-being. Pers Individ Dif 2016 Apr;93:63-67. [doi: 10.1016/j.paid.2015.09.005]
51. Nguyen MH, Hargittai E. Digital disconnection, digital inequality, and subjective well-being: a mobile experience sampling study. J Comput Mediat Commun 2023 Nov 8;29(1):1. [doi: 10.1093/jcmc/zmad044]
52. Diener E, Emmons RA, Laresen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985;49:71-75. [doi: 10.1207/s15327752jpa4091_13]
53. Li Z, Yin X, Jiang S, Wang M, Cai T. Psychological mechanism of subjective well-being: a stable trait or situational variability. Soc Indic Res 2014 Sep;118(2):523-534. [doi: 10.1007/s11205-013-0449-x]
54. Jaso BA, Kraus NI, Heller AS. Identification of careless responding in ecological momentary assessment research: from posthoc analyses to real-time data monitoring. Psychol Methods 2022 Dec;27(6):958-981. [doi: 10.1037/met0000312] [Medline: 34582244]
55. Hasselhorn K, Ottenstein C, Lischetzke T. Modeling careless responding in ambulatory assessment studies using multilevel latent class analysis: factors influencing careless responding. Psychol Methods 2023 May 11. [doi: 10.1037/met0000580] [Medline: 37166858]
56. Welling J, Fischer RL, Schinkel-Bielefeld N. Is it possible to identify careless responses with post-hoc analysis in EMA studies? Presented at: UMAP '21: 29th ACM Conference on User Modeling, Adaptation and Personalization; Jun 21-25, 2021; Utrecht, Netherlands p. 150-156. [doi: 10.1145/3450614.3462237]
57. Aminikhanghahi S, Wang T, Cook DJ. Real-time change point detection with application to smart home time series data. IEEE Trans Knowl Data Eng 2019 May;31(5):1010-1023. [doi: 10.1109/tkde.2018.2850347] [Medline: 35903759]
58. Arslan RC, Reitz AK, Driebe JC, Gerlach TM, Penke L. Routinely randomize potential sources of measurement reactivity to estimate and adjust for biases in subjective reports. Psychol Methods 2021 Apr;26(2):175-185. [doi: 10.1037/met0000294] [Medline: 32584065]
59. Gross T, Malzhacker T. The experience sampling method and its tools: a review for developers, study administrators, and participants. Proc ACM Hum-Comput Interact 2023 Jun 14;7(EICS):1-29. [doi: 10.1145/3593234]

Abbreviations

ARIMA: Autoregressive Integrated Moving Average
DTW: dynamic time warping
EMA: ecological momentary assessment
RR: response rate
STROBE CREMAS: Strengthening the Reporting of Observational Studies in Epidemiology Checklist for Reporting Ecological Momentary Assessment Studies

Edited by L Buis; submitted 06.02.24; peer-reviewed by K Yoshiuchi, R Eckhoff, R Pryss; revised version received 10.09.24; accepted 03.03.25; published 10.04.25.

Please cite as:
Cook D, Walker A, Minor B, Luna C, Tomaszewski Farias S, Wiese L, Weaver R, Schmitter-Edgecombe M
Understanding the Relationship Between Ecological Momentary Assessment Methods, Sensed Behavior, and Responsiveness: Cross-Study Analysis
JMIR Mhealth Uhealth 2025;13:e57018
URL: https://mhealth.jmir.org/2025/1/e57018
doi:10.2196/57018

© Diane Cook, Aiden Walker, Bryan Minor, Catherine Luna, Sarah Tomaszewski Farias, Lisa Wiese, Raven Weaver, Maureen Schmitter-Edgecombe. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feature Selection for Physical Activity Prediction Using Ecological Momentary Assessments to Personalize Intervention Timing: Longitudinal Observational Study

Devender Kumar[1], PhD; David Haag[1,2,3], MSc; Jens Blechert[2,4], PhD; Josef Niebauer[1,5], PhD; Jan David Smeddinck[1], PhD

[1]Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria

[2]Department of Psychology, Paris Lodron University of Salzburg, Salzburg, Austria

[3]Digital Health Information Systems, Center for Health & Bioresources, AIT Austrian Institute of Technology GmbH, Graz, Austria

[4]Centre for Cognitive Neuroscience, Paris Lodron University of Salzburg, Salzburg, Austria

[5]University Institute of Sports Medicine, Prevention and Rehabilitation, Paracelsus Medical University, Salzburg, Austria

Corresponding Author:
Devender Kumar, PhD
Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria

Abstract

Background: There has been a surge in the development of apps that aim to improve health, physical activity (PA), and well-being through behavior change. These apps often focus on creating a long-term and sustainable impact on the user. Just-in-time adaptive interventions (JITAIs) that are based on passive sensing of the user's current context (eg, via smartphones and wearables) have been devised to enhance the effectiveness of these apps and foster PA. JITAIs aim to provide personalized support and interventions such as encouraging messages in a context-aware manner. However, the limited range of passive sensing capabilities often make it challenging to determine the timing and context for delivering well-accepted and effective interventions. Ecological momentary assessment (EMA) can provide personal context by directly capturing user assessments (eg, moods and emotions). Thus, EMA might be a useful complement to passive sensing in determining when JITAIs are triggered. However, extensive EMA schedules need to be scrutinized, as they can increase user burden.

Objective: The aim of the study was to use machine learning to balance the feature set size of EMA questions with the prediction accuracy regarding of enacting PA.

Methods: A total of 43 healthy participants (aged 19 - 67 years) completed 4 EMA surveys daily over 3 weeks. These surveys prospectively assessed various states, including both motivational and volitional variables related to PA preparation (eg, intrinsic motivation, self-efficacy, and perceived barriers) alongside stress and mood or emotions. PA enactment was assessed retrospectively via EMA and served as the outcome variable.

Results: The best-performing machine learning models predicted PA engagement with a mean area under the curve score of 0.87 (SD 0.02) in 5-fold cross-validation and 0.87 on the test set. Particularly strong predictors included self-efficacy, stress, planning, and perceived barriers, indicating that a small set of EMA predictors can yield accurate PA prediction for these participants.

Conclusions: A small set of EMA-based features like self-efficacy, stress, planning, and perceived barriers can be enough to predict PA reasonably well and can thus be used to meaningfully tailor JITAIs such as sending well-timed and context-aware support messages.

(JMIR Mhealth Uhealth 2025;13:e57255) doi:10.2196/57255

KEYWORDS

digital health; behavior change; tailoring; personalization; adaptive systems; ecological momentary assessments; sensing; questionnaires; machine learning; feature selection; situated research; physical activity; implementation intentions; barriers; intention-behavior gap; artificial intelligence; AI; well-being; user assessment; survey; self-efficacy; stress; mood; emotions; mobile phone

XSL•FO
RenderX

Introduction

In light of financial, staffing, and other pressures on societal and health care systems linked to aging populations and the increased prevalence of chronic diseases related to sedentary lifestyles and other behavioral patterns [1], a growing number of digital health apps aim to promote positive lifestyle changes [2]. Such apps have the potential to improve health outcomes [3], prevent diseases [4], and enhance the quality of life for individuals [5]. For instance, encouraging physical activity (PA) and heart-healthy habits can help prevent serious health issues like coronary heart disease, diabetes, and cancer [6]. Yet, long-term adherence to PA presents a significant challenge [7]. Despite the well-documented benefits of regular exercise for overall health and disease prevention, many people struggle to maintain consistent PA [7].

To improve the effectiveness and adherence to PA, just-in-time adaptive interventions (JITAIs) are being investigated to tailor personalized and contextualized digital health support, often enabled by mobile health technologies such as wearable sensing devices and smartphones [8,9]. JITAIs offer a promising solution to tackle issues around physical inactivity, enabling effective behavior change and habitualization by providing personalized and timely support to individuals (eg, sending motivational messages to incentivize movement after prolonged periods of inactivity). JITAIs are configured to use real-time data and context awareness to deliver the "right" interventions precisely when they are needed the most [8,9]. JITAIs based on passive sensing have already been applied to various areas including eating disorders [10], mental health conditions, obesity and weight management, PA promotion [11], and smoking cessation [12].

Tailoring JITAIs based on passively sensed contextual factors (PSCFs), such as location, activity type or levels, daily weather conditions, or an individual's heart rate over time, is commonly observed in the literature. However, these passively sensed features face challenges in capturing relevant signals, which enable accurate and reliable characterizations of user contexts. These accurately represent what matters to a user at a given time, for example, informing the decision rules within a JITAI. To establish relevance, accuracy, and reliability, it would be necessary to validate passively sensed feature sets against the subjective user experience, difficult to do in the context of emotions and self-regulation that play a central part in the enactment of health behaviors such as PA [13]. Therefore, tailoring PA-fostering JITAIs based on PSCF comes with a considerable risk of misaligning intervention timing and content with users' current states. Misaligned interventions may annoy users or could hinder engagement leading to drop out from JITAI use [9,14]. In literature, self-report measures of momentary affect and motivation have been shown to be closely related to actual PA [15,16]. This raises the question whether these self-reports—referred to as ecological momentary assessment (EMA)—could directly be used as an addition or alternative to tailoring JITAIs via PSCF.

The EMA can present a valuable option for collecting near-real-time information on the experiences, interests, abilities, needs, behaviors, or other contextual circumstances of individuals in their natural context [17,18]. The advantages of EMA lie in providing rich, context-specific data with reduced memory biases, facilitating a deeper understanding of human behaviors and experiences directly tied to the individual. EMAs are therefore widely used across various fields to gain insights into psychological and behavioral dynamics [18]. Although EMA has shown tremendous potential in capturing individuals' momentary experiences, its use in PA adherence prediction has not been widely explored. Additionally, EMA comes with challenges such as participant burden and reactance [18]. Participant burden refers to the issue that asking too many questions or asking questions too frequently can leave participants annoyed, potentially causing them to ignore EMA prompts or discontinue their participation entirely [19]. Reactance, on the other hand, can be another outcome of overly frequent or extensive inquiries, which might also unduly influence participants' perceptions or behaviors [18]. Compared to EMA deployment in limited-duration study settings, these concerns are even more relevant when EMA is intended to inform JITAI tailoring over prolonged periods of time (ie, months or years in real-world applications). Thus, alongside investigating whether EMA can be used to accurately predict PA engagement, it is critical to identify the most predictive EMA questions that allow for a sufficiently accurate prediction of PA engagement without overburdening the user.

Accordingly, we investigated whether EMA can be used to predict PA engagement and thereby inform the contextualized tailoring of JITAIs. Further, from a practical perspective, we focused on understanding design implications and strategies to balance concerns of EMA fatigue [20,21] with traditional optimization and feature selection techniques in machine learning (ML). The ML techniques can more efficiently uncover hidden patterns and relationships in the data compared to traditional survey validation methods. This work is novel in exploring the viability of EMA for the timewise tailoring of JITAIs—in terms of "when a JITAI should optimally be delivered"—and in preparing such tailoring not directly based on observed variable thresholds but on prediction outcomes. These overarching aims are reflected in the following guiding research questions (RQs):

- RQ1: How accurately can EMAs predict PA engagement or adherence?
- RQ2: Which motivational, emotional, or volitional psychological states captured through EMAs will best predict PA? What is the smallest set of predictors that balances acceptable user burden with practically sufficient prediction accuracy?

Based on the existing EMA literature [22], we expect that EMA can inform behavioral predictions as they closely reflect a participant's perceived state, which can often be related to subsequent behavior. Regarding the constructs that could inform the prediction of PA from EMA, health behavior models, such as the Integrated Behavior Change Model [23], the Health Action Process Approach [24], or the Temporal Self-Regulation Theory [13] alongside our prior work on predictors of PA [15], suggest motivational and volitional qualities of self-regulation are closely related to PA engagement. This includes constructs

XSL•FO
RenderX

such as intrinsic motivation, intention, action planning, and self-efficacy alongside anticipated contextual barriers and momentary affect (eg, current mood or stress). Therefore, we expect these factors to be strong predictors of PA engagement in this study.

Methods

Study Design

The EMA data used in this paper were collected in our prior investigation [15], aimed at understanding the determinants and barriers of PA engagement. Following informed consent and after completing psychometric and demographic (see Table S1 in Multimedia Appendix 1) questionnaires in an web-based survey, participants began the study's EMA phase. During this phase, 4 EMA prompts were sent to the participants on a daily basis at fixed time points (9 AM, 1 PM, 5 PM, and 9 PM) over the course of a 3-week study period. Additionally, participants were also able to report PA independently of those fixed EMA prompts. After the 3 weeks of EMA, another web-based questionnaire was sent to the participants, to assess compliance and reactivity. However, the analyses in this study focus on the EMA data.

Ethical Considerations

The study received ethics approval from the University of Salzburg, Austria (GZ 11/2020). Informed consent was provided by all participants prior to study enrollment and covered all primary and secondary analyses. Study participants were compensated with US $32 and received personalized feedback based on their data after the study concluded [25]. All data in this paper were anonymized.

EMA Measures

While selecting EMA items, we aimed to align with the structure proposed in contemporary models of health behavior, such as the Health Action Process Approach [26], the Integrated Behavior Change Model for PA [23], or the Temporal Self-Regulation Theory [13]. These models propose a 2-step process to explain the implementation of health behaviors such as PA. This process consists of (1) a motivational phase, which leads to the formation of intentions and (2) a volitional phase that bridges the gap between intention and health behavior enactment.

To reflect these 2 steps in our study design, we first sampled candidate predictors proposed by these models as relevant to the motivational phase at each EMA prompt. This included measures of momentary mood (10 items from the Positive and Negative Affect Schedule [27]: happy, relaxed, active, irritated, concerned, depressed, nervous, stressed, energetic, and tired) and stress :2 items from the Perceived Stress Scale (PSS) [28]. The German version by Schneider et al [29] anticipated barriers to PA ("How well would your given circumstances allow you to be physically active at the moment?"; BarrPA), and pain as a barrier to PA ("At the moment, do you have physical complaints that impede physical activity?"). Furthermore, the morning prompt contained items assessing sleep quality ("How good was your sleep?"; SleepQlt), the time of falling asleep ("When did you fall asleep?"; ST), and the waking time ("When did you wake up in the morning?"; WT), which could also be either barriers or resources for PA engagement.

Next, we prospectively asked the participants about their intentions to be physically active ("Do you intend to be physically active in the next 4 hours?," yes or no). Only if they responded with "yes," we further assessed volitional determinants of health behavior enactment such as planning specificity ("How specifically did you plan this physical activity?"; ActPlan), self-efficacy ("How strongly do you believe you can enact your plan under the given circumstances?"; ActPlanSE), and momentary intrinsic motivation ("Independent of the circumstances, how motivated are you right now to be physically active?"). Given the practical purpose around supporting adherence or enactment of planned and intended PA, selecting for positive PA intention provides reasonable grounds.

All of these items were developed or adjusted to fit the specific needs of this study. Except intention and sleep duration, they were measured on a horizontal slider from 0-not at all to 100-very much. Within each EMA prompt, participants finally reported their PA retrospectively, ("Have you been physically active in the last 4 hours?," yes or no), which we used as the primary outcome in our analyses (see Haag et al [15] for cross-validation of these PA self-reports against wearable data). For this EMA sampling, we first used the Smarteater app (internally developed by the University of Salzburg), and in a later recruitment phase, switched to the m-Path platform [30]. However, each participant only used 1 platform, and assessments were identical on both platforms. Therefore, data quality was not impaired due to switching between the two platforms. For the full list of EMA items, please refer to Table S2 in Multimedia Appendix 1.

Study Participants

Participants were healthy individuals who self-reported having no limitations in their ability to perform PA, and did not describe themselves as competitive athletes. As of June 2022, a total of 49 participants were enrolled in the data collection, which was conducted in a phased manner. Of these, 10 participants dropped out before completing all 3 weeks. However, the data from 4 of the 10 dropouts were still used for this investigation since they filled out enough EMA prompts to be included over 1 week. The final sample included 43 participants (31 female and 12 male) between 19 -67 years of age (mean 39.14, SD 15.53) years. In total, 16 reported high, 22 moderate, and 5 low activity levels in the International Physical Activity Questionnaire [31] completed at the beginning of the study. Based on the participants' self-reported height and weight, their BMI ranged from 19.0 to 39.6 (mean 24.4, SD 3.4) kg/m^2, with 2 participants being obese (BMI≥30.0 kg/m^2), 11 overweight (BMI=25.0 - 29.9 kg/m^2), and 29 with normal weight (BMI=18.5 - 24.9 kg/m^2). The BMI for 1 participant was omitted due to being unrealistic (ie, unreasonable self-reported height and weight).

For further details on study setup, procedure, as well as the primary outcomes, please see Haag et al [15].

XSL•FO
RenderX

Data Cleaning and Feature Engineering

Each EMA prompt contained at least 16 questions. Some questions in the EMA prompts included conditional subquestions (eg, planning specificity only when a PA was intended) Questions requiring per-day-level measurements (eg, sleep quality and duration) were limited to 1 morning prompt per day. This approach resulted in 41 distinct questions that were assessed in the EMA. For practical reasons, we only considered the prospective and momentary questions as predictors of retrospectively assessed PA in the subsequent analysis since they are better suited for modeling the likelihood of upcoming PA adherence at runtime. Further details of the EMA questions can be found in Table S2 in Multimedia Appendix 1.

During the data cleaning process, EMA questions with less than a 30% response compliance rate were dropped as data augmentation on such low compliance rates could introduce a bias in the models [32]. EMA prompts for which participants did not report whether they performed the PA were also discarded; this was done retrospectively. For example, if a participant did not report engaging in PA within the last 4 hours at 1 PM, their 9 AM prompt data were excluded from the analysis. The times participants fell asleep and woke up were self-reported once only in the first EMA of the day. Therefore, empty cells for sleep and wake time for EMA prompts 2, 3, and 4 were populated with values from the first EMA of each day. Any remaining missing values for sleep hours, wakeup time, and sleep qualitywhen participants did not respond to the day's first EMA prompt were replaced with their respective median values up to that point in time. The categorical EMA questions with "yes or no" answers were transformed to binary values (1/0). The StandardScaler from the *Scikit-learn* package [33] was used to scale all the continuous variables to unit variance. This standardization of input features can be beneficial for the performance and convergence of ML algorithms.

Model Training and Validation

For modeling, a representative selection of common ML methods, including logistic regression, decision tree, support vector machine [34], k-nearest neighbor (KNN) [35], random forest (RF) [36], and extreme gradient boosting (XGBoost or XGB) [37] was used to compare the model prediction performances. The decision to choose traditional ML models for exploration over neural networks was made, given the relatively small size of the dataset, since neural networks typically require an extensive dataset for training. The data were split into training or test sets (in a ratio of 80:20) using the *Scikit-learn* package [33]. A stratified 5-fold cross-validation method on the training dataset was used to evaluate model performance and hyperparameter tuning [38]. The stratified k-fold cross validation method was chosen to ensure that each fold's class distribution similar to the overall class distribution in the dataset. This is particularly useful when dealing with imbalanced datasets, where one class may have significantly more instances than the others. All models were then evaluated on the test dataset. The GridSearchCV scoring method from *Scikit-learn* [33] was used to find the optimal hyperparameters.

Model prediction performances were compared using the area under the receiver operating characteristic curve, as it is a more appropriate measure compared to simple accuracy, especially for imbalanced datasets [39]. The area under the receiver operating characteristic curve visualizes the performance of the model at various probability thresholds for classification and helps in selecting an appropriate threshold that balances the trade-off between true positives and false positives.

To overcome data imbalance, we attempted upsampling with the synthetic minority oversampling technique (SMOTE) [40] that is implemented in the Python library *imbalanced-learn* [41] on the training data. We trained the models with the upsampled data and compared their internal validation results on the test dataset.

For identifying the top features predicting whether PA took place within a given EAM slot, we used the recursive feature elimination (RFE) technique from *Scikit-learn* [33]. RFE is primarily used in ML to select the most relevant and important features from a given dataset. It recursively trains the model using subsets of features and ranks them based on their contribution to improving prediction outcomes. By iteratively eliminating less important features, RFE provides an optimal subset of features that maximizes the model's performance. After the RFE, we used the Shapley Additive Explanations (SHAP) framework [42] for understanding the feature importance within the best-performing models. The models were built in Python 3.8 using *Scikit-learn* (version 1.2.0) package [33]. The entire training process was conducted on a Windows 11 operating system with 64 GB RAM, a dual-core Intel Core i9 processor, and an Nvidia graphics card (Nvidia RTX A4000) with 16 GB memory.

Results

Overview

Following data cleaning, 23 of 41 EMA questions fulfilled the minimum data compliance requirements and were used as features for building the models and predicting the compliance or noncompliance of PA. In this section, we first outline the performance of the range of selected candidate ML models in predicting PA engagement when all the features were used in training. Thereafter, we will show the results of feature importance in predicting PA engagement using the RFE technique and a SHAP value visualization. We conclude with a practical discussion on selecting an appropriate model or ensemble and feature set, contextualizing the work around more general concerns in model training and selection for dynamic personalization based on EMA.

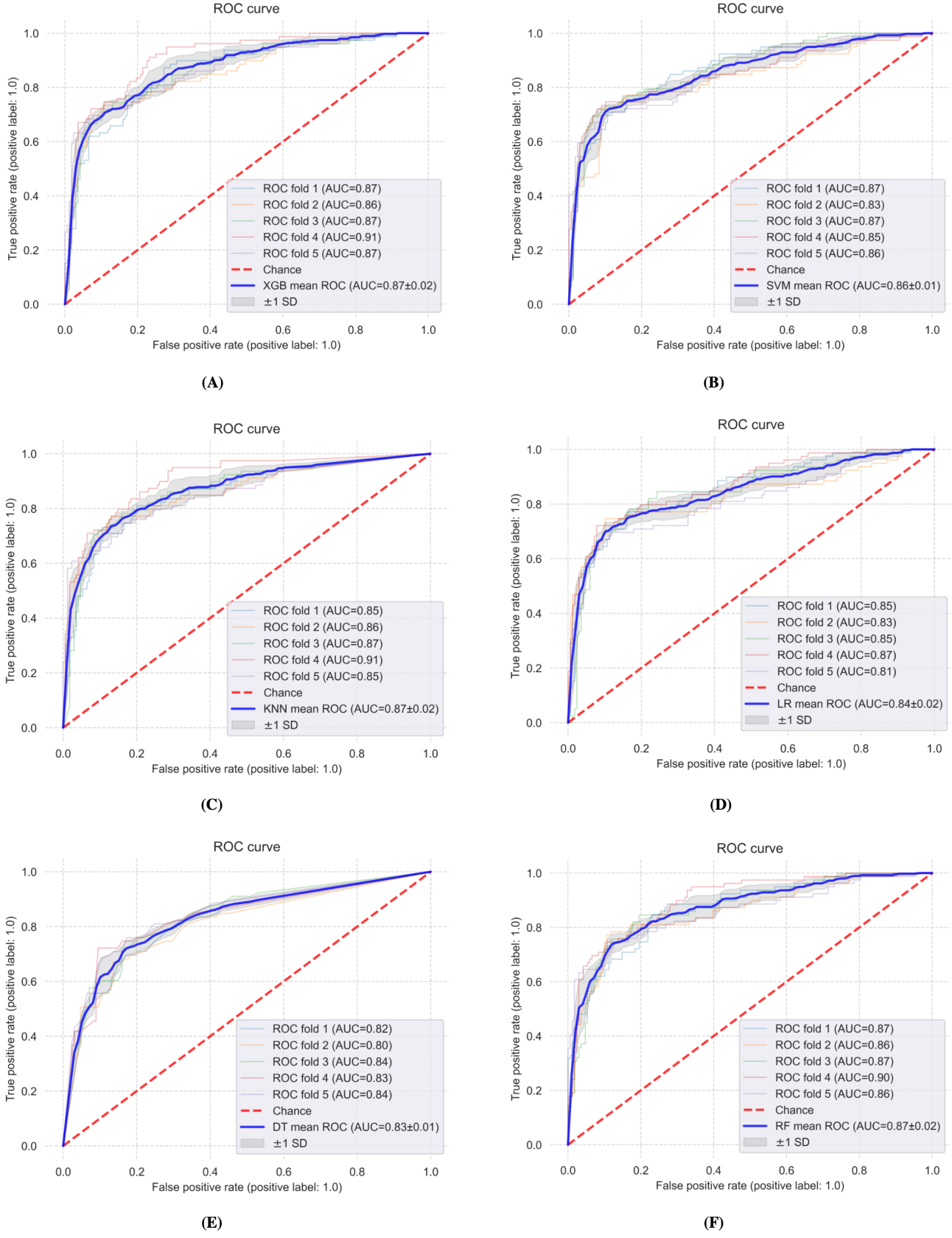
PA Prediction Performance of Various Model Using All Available Features

Figure 1 shows the area under the receiver operating characteristic curve of various models in a 5-fold cross-validation on training data, and Table 1 presents their area under the curve (AUC) scores on training and test sets. XGB, KNN, and RF models have shown the best performance and achieved a mean AUC score of 0.87 (SD 0.02) on 5-fold cross-validation. However, among them, XGB achieves a slightly higher AUC score (0.87) on the unseen test set. We also tested the upsampling technique SMOTE on the training data;

XSL•FO
RenderX

however, as shown in the 2 rightmost columns in Table 1, this did not significantly improve the test set AUC score. Since SMOTE does not improve the performance on the test set, it was not used for further analysis.

Figure 1. (**A**) XGB, (**B**) SVM, (**C**) KNN, (**D**) LR, (**E**) DT, and (**F**) RF show the performance of selected machine learning models in 5-fold cross-validation using all the features. AUC: area under the curve; DT: decision tree; KNN: k-nearest neighbor; LR: logistic regression; RF: random forest; ROC: receiver operating characteristic curve; SVM: support vector machine; XGB: extreme gradient boosting.



XSL•FO
RenderX

Table . Performance of various models on 5-fold cross-validation and test set[a].

Model name	AUC[b] score 5-fold cross-validation, mean (SD)	AUC score on test set	SMOTE AUC score 5-fold cross-validation, mean (SD)	SMOTE AUC score on test set
XGB[c]	0.87 (0.02)	0.87	0.93 (0.01)	0.86
RF[d]	0.87 (0.02)	0.86	0.94 (0.01)	0.86
KNN[e]	0.87 (0.03)	0.86	0.92 (0.01)	0.85
LR[f]	0.84 (0.02)	0.85	0.86 (0.02)	0.83
SVM[g]	0.86 (0.02)	0.86	0.86 (0.03)	0.84
DT[h]	0.83 (0.01)	0.83	0.86 (0.01)	0.81

[a]Contrasted with performance following synthetic minority oversampling technique (SMOTE) [8] of training data.

[b]AUC: area under the curve.

[c]XGB: extreme gradient boosting.

[d]RF: random forest.

[e]KNN: k-nearest neighbor.

[f]LR: logistic regression.

[g]SVM: support vector machine.

[h]DT: decision tree.

Feature Importance

Overview

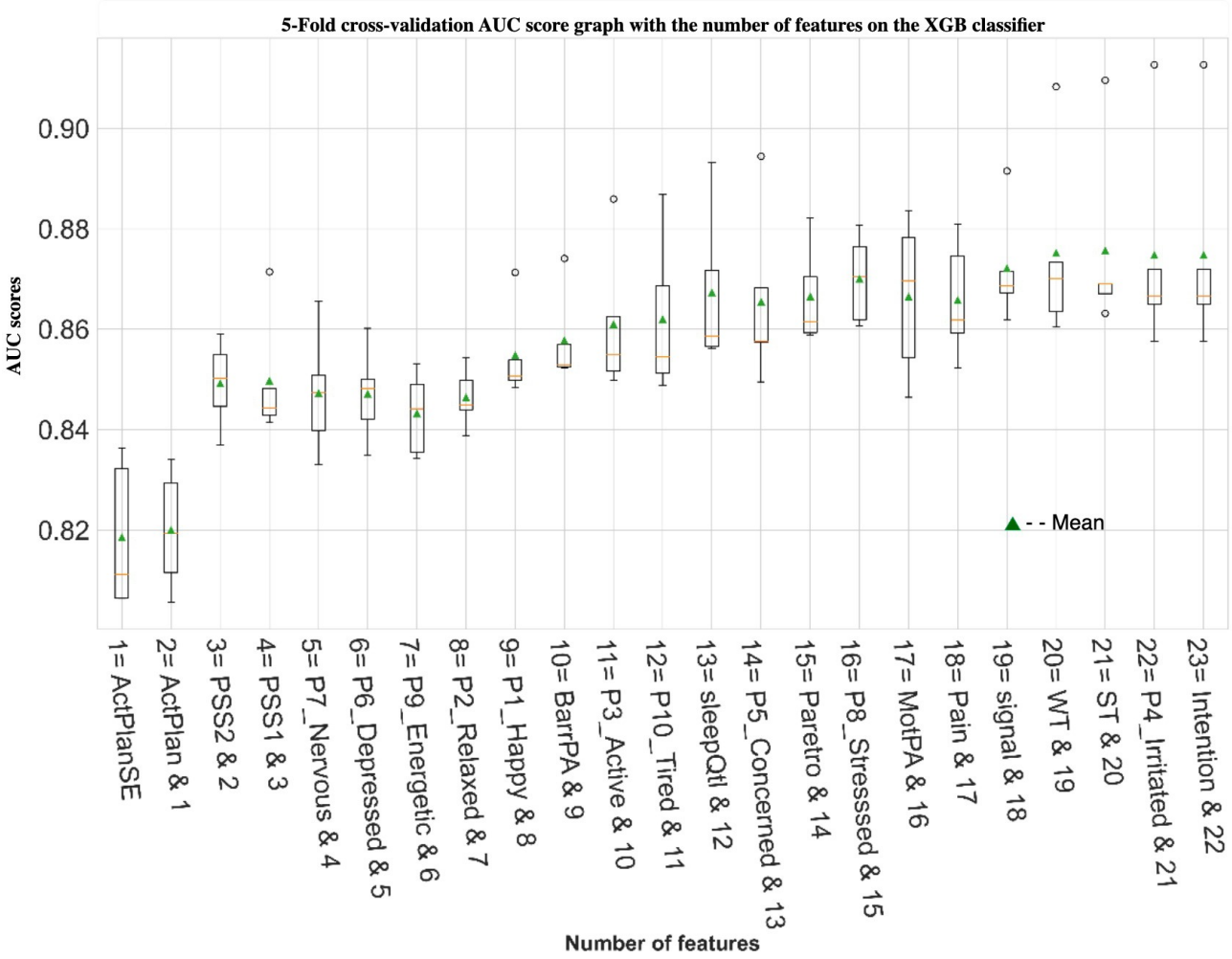
To reduce potential questionnaire fatigue [20] caused by the extensive EMA survey length and finding the smallest but accurate subset of EMA questions for predicting the PA, after finalizing the best-performing model on all 23 features, we focused on understanding the feature importance in predicting PA. Since the XGB model showed slightly better performance on the test dataset (Table 1) compared to RF and KNN, explorations into feature importance were done exclusively for XGB .

Recursive Feature Elimination

Figure 2 shows the selected top-n (n=number 1, 2, 3, etc) features identified by RFE, alongside with a box plot of the corresponding AUC scores for the XGB classifier. As depicted in Figure 2 , for the XGB models, the highest mean AUC score (0.87) is achieved when using 13 features and remains the same even when all 23 features are used. Interestingly, a competitive AUC score of 0.85 is achieved with just 3 features, which combine ActPlan and ActPlanSE as indicators of volitional intention, along with PSS2 as a marker of stress. An increase to an AUC score of 0.86 is only achieved when using 10 or more features.

XSL•FO
RenderX

Figure 2. XGB models with growing feature set sizes: boxplots of AUC scores with a number of features using recursive feature elimination. The numbers (1, 2, 3, etc) indicate the top n-number of features and their names. Feature sets are listed as "additonal_feature & X (identifier of compounded prior features)." AUC: area under the curve; XGB: extreme gradient boosting.



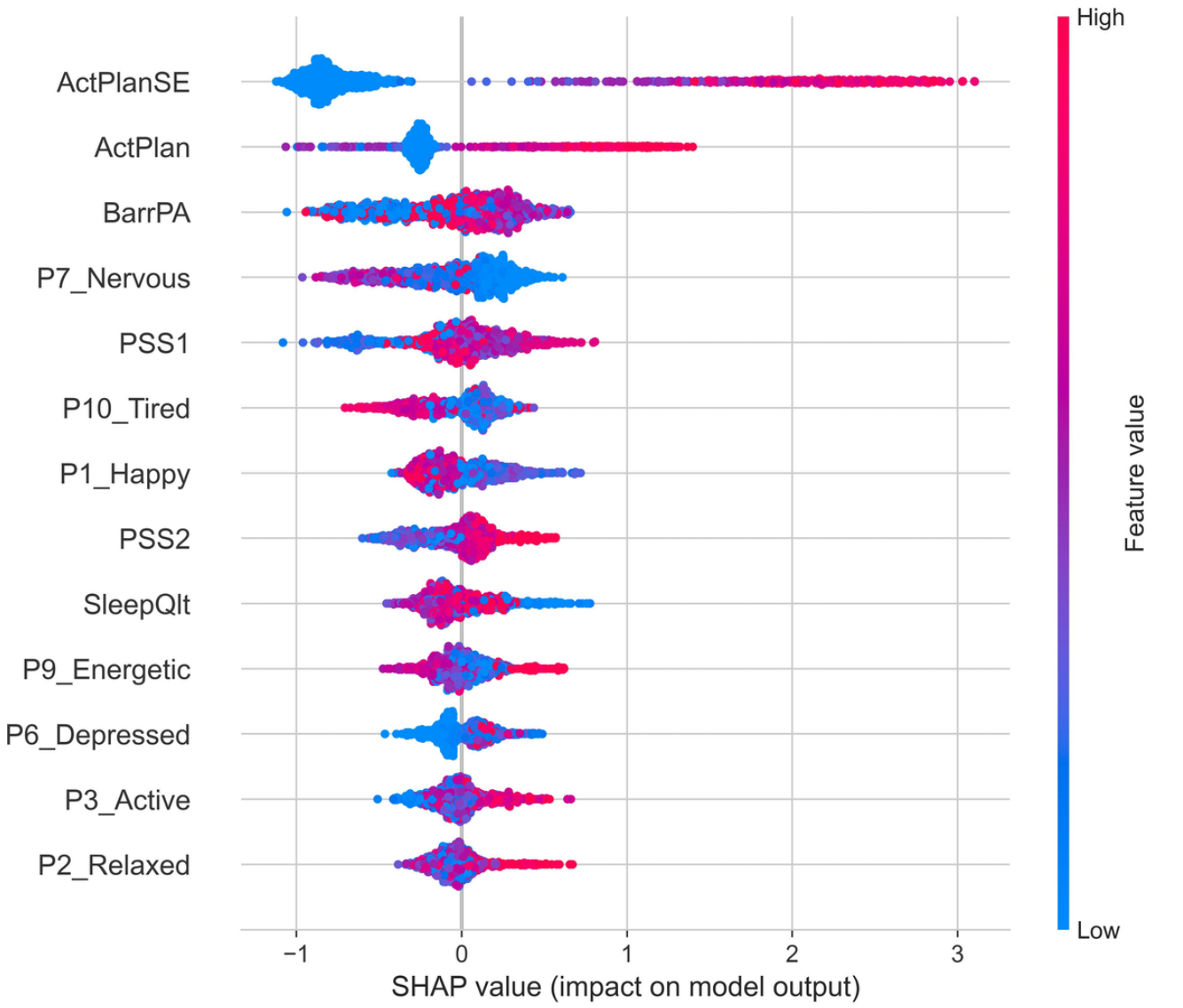
SHAP Value–Based Feature Importance

Figure 3 shows a SHAP value graph of the top 13 features and their contributions in PA prediction by the XGB model. For brevity, we present SHAP values of 13 features, as subsequent features do not improve the AUC score (as apparent in Figure 2). SHAP value graph indicates the correlations through which different features contribute to the model's output for individual predictions.

Similar to the RFE (Figure 2), items ActPlanSE (self-efficacy) and ActPlan (PA planning specificity) remain the top 2 features with the most notable impact on model output (indicated by wide impact score spread combined with decisive directionality on the horizontal axis). The positive SHAP values indicate that the feature increases the PA enactment prediction, while negative values indicate the opposite. The color of the bar represents the feature value (red for high and blue for low). The slight difference in the feature ranking by RFE and SHAP value is due to differences in their feature selection methodology and objective of feature selection. RFE focuses on improving overall model performance metrics, whereas SHAP values aim to provide interpretable explanations for individual prediction by a given feature.

XSL•FO
RenderX

Figure 3. SHAP values graph of top 13 features indicating how each feature is contributing to the extreme gradient boosting model's output. SHAP: Shapley Additive Explanations.



Discussion

Principal Findings

In this study, we explored the viability of predicting PA engagement based on common supervised ML methods with features drawn from EMA data collected in the participants' natural contexts. Conceptually, if PA engagement or nonengagement can be predicted, this can inform the timing of issuing JITAIs. For instance, if negative engagement is predicted, motivational messages of encouragement, or more specific planning or replanning, can be sent. Since EMA requires the active involvement of the user, we also explored which features or questions of the EMA were key determinants for effectively predicting PA engagement. This exploration offers a framework for informing trade-off decisions that can help in designing JITAI systems that take valuable information derived from EMA without overburdening the users.

EMA and PA Engagement Prediction

Regarding RQ1 on how accurately EMAs and ML models can help predict PA engagement, all the ML models (Table 1) produce AUC scores of 80 or higher on test data that were not seen by the models in training. When considering smaller performance differences, both XGB and RF models performed best on both cross-validation (mean 0.87, SD 0.2) and test sets (AUC 0.87 and 0.86, respectively), broadly indicating that the questions used in the current EMA could be a viable option for PA engagement prediction and can complement the tailoring of passive sensing–based JITAIs to foster PA.

While such performance levels would not satisfy strong reliability requirements (eg, in health diagnostics), they can arguably be reasonably used in JITAI settings around fostering long-term PA, where occasional type 1 or type 2 errors do not have grave consequences. Compared to previous work on the same dataset, which used non-ML models to investigate determinants of PA enactment in a theory-driven manner [15], this study adopts a practical approach on the potential of operationalizing EMA for actively driving JITAIs independently of preconditions, such as the PA being explicitly intended. Thereby, favoring model performance over analyses intended to inform theory building for feature composition.

XSL•FO
RenderX

The outcomes of automated feature selection corroborate earlier findings [15] with non-ML modeling, as feature importance analysis indicates that self-efficacy (ActPlanSE) and planning specificity (ActPlan) are the key relevant predictors of PA engagement. Here, self-efficacy represents a participant's confidence in their ability to engage in the intended activity, and planning specificity refers to how specifically they had planned that activity. This result also falls in line with other previous publications indicating a close relationship between PA engagement and these volitional components [43-45]. Further, analysis based on RFE on the XGB model (Figure 2) indicates stress (PSS1 and PSS2) as another potentially relevant factor that can be a qualifier for PA outcomes. This association of stress and PA is also consistent with previous EMA literature [16].

Psychological Considerations and Contextualization in Theory

In light of the dual-process models of health behavior [13,23] that the present data collection was based upon, it is not surprising to see the volitional constructs (ActPlanSE and ActPlan) ending up having greater feature importance. However, even though we found a resemblance to these psychological models, our results are not to be confused with an evaluation of such models. Instead, our findings represent a practical investigation of the ML methodology being applied to determine which ones (within a given set of candidate EMA items) could be used for a runtime system to predict PA adherence and control adaptive interventions accordingly. This implies that the proposed feature selection might be very different in another set of potential predictors and necessitates the interpretation of our findings in light of our specific EMA design.

For example, consider the intention item or feature that was ranked at the very bottom of our RFE (Figure 2). From a theoretical point of view, intention would often be seen as the pivotal point in health behavior engagement [24]. Therefore, it may seem very surprising that intention (Figure 2) is being ranked the least predictive feature in our dataset. This discrepancy originates in the structure of our EMA, with ActPlanSE and ActPlan only being assessed if participants report to have the intention for PA. Therefore, these features would have missing values for each episode where no intention is reported. In these cases, values of the volitional determinants were replaced with zeros. Thus, the binary intention data are encoded in ActPlanSE and ActPlan. In the RFE process, the presented models will have picked up on this characteristic and ranked intention itself very low since it does not contain additional information if ActPlan or ActPlanSE is included. However, this does not imply that we can forgo the assessment of intention since it is only meaningfully possible to assess planning and self-efficacy when an intention is given. In practical terms, this would imply that the precondition of intention being given would be wrapped into the wording for an ActPlan item if it were to be presented without an explicit item on intention preceding it (ie, "In case you are intending to exercise, how specifically have you planned").

Nonetheless, the indicated feature combination including ActPlan and stress items corresponds well with expectations that can be derived from the aforementioned dual-process models of health behavior [13,23,26], representing elements that can be seen to capture aspects of both a motivational phase and a volitional phase.

Trade-Off Between EMA Burden and PA Prediction Accuracy

The potential viability of the sets of EMA questionnaires selected for this study was informed by psychological theories. However, the list of questions as a whole is clearly too burdensome for being of practical use in guiding JITAIs. As indicated in the EMA literature [19-21], questionnaire fatigue or overload is a real concern for practicality. As shown in Figure 2, an AUC of over 0.85 is achieved by just using the top 3 features (ie, ActPlanSE, ActPlan, and PSS2) as compared to the maximal achievable AUC of 0.87 when all the 23 features were used. Moreover, as evidenced by the RFE process, the maximal AUC score of 0.87 is already achieved with 13 features, indicating that beyond these 13 features, the rest other features are not adding additional benefit for predicting PA engagement. Even within these 13 features, after adding an item from the stress questions (PSS1 and PSS2), Positive and Negative Affect Schedule mood questions (P7_nerves, P5_concern, P1_happy, P6_depressed, and P7_energetic) are arguably not adding significant improvements in AUC score to justify their inclusion as frequently asked items in a practical longer-term deployment. The PA-friendly external circumstances (BarrPA) and the absence of barriers and sleep quality (SleepQlt) do minor improvement of the AUC score and could be candidates for inclusion with the EMA for research purposes but would likely not be included in a production system, as the improvements are negligible for practical purpose.

For most practical purposes and keeping the EMA questionnaire burden to the minimum, in settings where an intention for PA is given (eg, supporting the engagement of a PA plan), a deployed JITAI model could be driven by just 3 EMA items (ActPlanSE, ActPlan, and PSS2). If the intention is not given, the first item could be rephrased as indicated earlier.

Design Implications

For the practicality of JITAIs or any intervention intended for fostering PA to be driven by regular EMA in live deployment situations, it is essential to select EMA questions effectively and avoid causing EMA fatigue. In relation to RQ2, about which concepts will best inform PA engagement prediction (as shown in the Feature Importance section), features (EMA questions) related to self-efficacy, planning, and stress have reasonable PA prediction power. These outcomes indicate that the long list of potentially relevant EMA questions used in PA prediction can be effectively cut short to a degree where real-world deployments of JITAIs with decision rules or trigger points [14] being informed by EMAs appear plausible. In addition, the proposed XGB model can be integrated into existing digital health platforms or mobile apps for effective EMA-based JITAIs and fostering PA engagement.

Limitations and Future Work

This study is formative with regard to surveying potentially relevant EMA constructs and measures. It does not include

XSL•FO
RenderX

observations or testing of the app of the derived models in practical deployment, which is a clearly indicated step for future work. The study also included a relatively small number of participants (n=43) and a limited time frame (3 weeks). To ensure the generalizability of the ML model (XGB) for automated physical activity prediction and the customization of JITAIs across diverse demographics, further research with larger sample sizes and longer study durations is needed. Additionally, in this study setup, the EMA frequency was established at 4 times a day. Future work should explore the impact of varying EMA frequencies per day on the predictive capabilities of the model for PA.

Conclusions

This paper presents a formative investigation into EMA's effectiveness in predicting PA implementation and whether EMA can be of practical use in gathering temporal context for JITAI decision-making. The outcomes with common supervised ML models, especially XGB with a mean AUC score of 0.87 (SD 0.02), indicate that EMA can offer relevant PA prediction power and thereby has the potential to complement more common passive sensing JITAI-tailoring approaches. Furthermore, the investigation around finding the right trade-off between EMA question load and PA engagement, prediction-accuracy of ML models indicates that self-efficacy (ActPlanSE) and planning specificity (ActPlan) play the most important role in determining the PA prediction under the assumption that initial intention to perform PA was already given. Three features, ActPlanSE, ActPlan, and PSS2 (stress), allowed for achieving an AUC score of 0.85 as compared to the maximum AUC score of 0.87 when all the 23 EMA questions were used.

Acknowledgments

The authors thank all study participants and all addtional contributors for their time and cooperation. The authors especially thank Sebastian Gruber and Christoph Zeiner for their support during the design and early analysis of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A Detailed list of ecological momentary assessment (EMA) questions used in the study and list of questionnaires included in the pre- and post-EMA surveys.
[DOCX File, 33 KB - mhealth_v13i1e57255_app1.docx]

References

1. Global health and aging. World Health Organization.: National Institute on Aging, National Institutes of Health, US Department of Health and Human Services; 2011. URL: https://www.nia.nih.gov/sites/default/files/2017-06/global_health_aging.pdf [accessed 2024-12-21]
2. Woessner MN, Tacey A, Levinger-Limor A, Parker AG, Levinger P, Levinger I. The evolution of technology and physical inactivity: the good, the bad, and the way forward. Front Public Health 2021;9:655491. [doi: 10.3389/fpubh.2021.655491] [Medline: 34123989]
3. Khan N, Marvel FA, Wang J, Martin SS. Digital health technologies to promote lifestyle change and adherence. Curr Treat Options Cardio Med 2017 Aug;19(8):60. [doi: 10.1007/s11936-017-0560-4]
4. Gray R, Indraratna P, Lovell N, Ooi SY. Digital health technology in the prevention of heart failure and coronary artery disease. Cardiovasc Digit Health J 2022 Dec;3(6 Suppl):S9-S16. [doi: 10.1016/j.cvdhj.2022.09.002] [Medline: 36589760]
5. Victoria-Castro AM, Martin ML, Yamamoto Y, et al. Impact of digital health technology on quality of life in patients with heart failure. JACC Heart Fail 2024 Feb;12(2):336-348. [doi: 10.1016/j.jchf.2023.09.022] [Medline: 37943227]
6. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. Can Med Assoc J 2006 Mar 14;174(6):801-809. [doi: 10.1503/cmaj.051351]
7. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Feb 19;18(4):2023. [doi: 10.3390/ijerph18042023] [Medline: 33669679]
8. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 3;16(1):31. [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
9. Wunsch K, Eckert T, Fiedler J, Woll A. Just-in-time adaptive interventions in mobile physical activity interventions—a synthesis of frameworks and future directions. Eur Health Psychol 2022;22(4):10 [FREE Full text]
10. Ralph-Nearman C, Sandoval-Araujo LE, Karem A, et al. Using machine learning with passive wearable sensors to pilot the detection of eating disorder behaviors in everyday life. Psychol Med 2024 Apr;54(6):1084-1090. [doi: 10.1017/S003329172300288X] [Medline: 37859600]

XSL•FO
RenderX

11. Nahum-Shani I, Smith SN, Tewari A, et al. Just-in-time adaptive interventions (JITAIs): an organizing framework for ongoing health behavior support. Ann Behav Med 2014;52(6):446-462. [doi: 10.1007/s12160-016-9830-8]
12. Naughton F. Delivering "just-in-time" smoking cessation support via mobile phones: current knowledge and future directions. Nicotine Tob Res 2017 Mar;19(3):379-383. [doi: 10.1093/ntr/ntw143] [Medline: 27235703]
13. Hall PA, Fong GT. Temporal self-regulation theory: a model for individual health behavior. Health Psychol Rev 2007 Mar;1(1):6-52. [doi: 10.1080/17437190701492437]
14. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
15. Haag D, Carrozzo E, Pannicke B, Niebauer J, Blechert J. Within-person association of volitional factors and physical activity: insights from an ecological momentary assessment study. Psychol Sport Exerc 2023 Sep;68:102445. [doi: 10.1016/j.psychsport.2023.102445] [Medline: 37665897]
16. Schultchen D, Reichenberger J, Mittl T, et al. Bidirectional relationship of stress and affect with physical activity and healthy eating. Br J Health Psychol 2019 May;24(2):315-333. [doi: 10.1111/bjhp.12355] [Medline: 30672069]
17. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
18. Dunton GF. Ecological momentary assessment in physical activity research. Exerc Sport Sci Rev 2017 Jan;45(1):48-54. [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]
19. Eisele G, Vachon H, Lafit G, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. Assessment 2022 Mar;29(2):136-151. [doi: 10.1177/1073191120957102]
20. Yang YS, Ryu GW, Choi M. Methodological strategies for ecological momentary assessment to evaluate mood and stress in adult patients using mobile phones: systematic review. JMIR Mhealth Uhealth 2019 Apr 1;7(4):e11215. [doi: 10.2196/11215] [Medline: 30932866]
21. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: a real-world feasibility study. J Affect Disord 2020 Sep 1;274:733-741. [doi: 10.1016/j.jad.2020.05.067] [Medline: 32664009]
22. Mikus A, Hoogendoorn M, Rocha A, Gama J, Ruwaard J, Riper H. Predicting short term mood developments among depressed patients using adherence and ecological momentary assessment data. Internet Interv 2018 Jun;12:105-110. [doi: 10.1016/j.invent.2017.10.001] [Medline: 30135774]
23. Hagger MS, Chatzisarantis NLD. An integrated behavior change model for physical activity. Exerc Sport Sci Rev 2014 Apr;42(2):62-69. [doi: 10.1249/JES.0000000000000008] [Medline: 24508739]
24. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors: the health action process approach. Eur Psychol 2008 Jan;13(2):141-151. [doi: 10.1027/1016-9040.13.2.141]
25. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323. [doi: 10.2196/jmir.5870] [Medline: 27986644]
26. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
27. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063]
28. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
29. Schneider EE, Schönfelder S, Domke-Wolf M, Wessa M. Measuring stress in clinical and nonclinical subjects using a German adaptation of the Perceived Stress Scale. Int J Clin Health Psychol 2020;20(2):173-181. [doi: 10.1016/j.ijchp.2020.03.004] [Medline: 32550857]
30. Mestdagh M, Verdonck S, Piot M, et al. m-Path: an easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. Front Digit Health 2023;5:1182175. [doi: 10.3389/fdgth.2023.1182175]
31. Craig CL, Marshall AL, Sjöström M, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
32. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput Surv 2022 Jul 31;54(6):1-35. [doi: 10.1145/3457607]
33. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
34. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol TIST ACM 2011;2(3):1-27. [doi: 10.1145/1961189.1961199]
35. Kramer O. K-nearest neighbors. In: Dimensionality Reduction with Unsupervised Nearest Neighbors: Springer; 2013:13-23. [doi: 10.1007/978-3-642-38652-7_2]

XSL•FO
RenderX

36. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
37. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016; San Francisco, CA, United States p. 785-794. [doi: 10.1145/2939672.2939785]
38. Brownlee J. Imbalanced Classification with Python: Machine Learning Mastery; 2020.
39. Melo F. Receiver operating characteristic (ROC) curve. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H, editors. Encyclopedia of Systems Biology: Springer; 2013:1818-1823. [doi: 10.1007/978-1-4419-9863-7_242]
40. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321-357. [doi: 10.1613/jair.953]
41. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res 2017;18(17):1-5 [FREE Full text]
42. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 2017 Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4-9, 2017; Long Beach, CA, United States p. 4768-4777. [doi: 10.5555/3295222.3295230]
43. Scholz U, Keller R, Perren S. Predicting behavioral intentions and physical exercise: a test of the health action process approach at the intrapersonal level. Health Psychol 2009;28(6):702-708. [doi: 10.1037/a0016088]
44. Scholz U, Schüz B, Ziegelmann JP, Lippke S, Schwarzer R. Beyond behavioural intentions: planning mediates between intentions and physical activity. Br J Health Psychol 2008 Sep;13(Pt 3):479-494. [doi: 10.1348/135910707X216062] [Medline: 17553212]
45. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]

Abbreviations

AUC: area under the curve
EMA: ecological momentary assessment
JITAI: just-in-time adaptive intervention
KNN: k-nearest neighbor
ML: machine learning
PA: physical activity
PSCF: passively sensed contextual factor
PSS: Perceived Stress Scale
RF: random forest
RFE: recursive feature elimination
RQ: research question
SHAP: Shapley Additive Explanations
SMOTE: synthetic minority oversampling technique
XGB: extreme gradient boosting

Edited by D Cox; submitted 13.02.24; peer-reviewed by AA wani, A Hassan, S Mao; revised version received 24.10.24; accepted 12.11.24; published 24.01.25.

Please cite as:
Kumar D, Haag D, Blechert J, Niebauer J, Smeddinck JD
Feature Selection for Physical Activity Prediction Using Ecological Momentary Assessments to Personalize Intervention Timing: Longitudinal Observational Study
JMIR Mhealth Uhealth 2025;13:e57255
URL: https://mhealth.jmir.org/2025/1/e57255
doi:10.2196/57255

© Devender Kumar, David Haag, Jens Blechert, Josef Niebauer, Jan David Smeddinck. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Participant Compliance With Ecological Momentary Assessment in Movement Behavior Research Among Adolescents and Emerging Adults: Systematic Review

Shirlene Wang[1,2], PhD; Chih-Hsiang Yang[3], PhD; Denver Brown[4], PhD; Alan Cheng[5], BHSc; Matthew Y W Kwan[5,6], PhD

[1]Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA, United States

[2]Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[3]Department of Exercise Science and TecHealth Center, University of South Carolina, Columbia, SC, United States

[4]Department of Psychology, The University of Texas at San Antonio, San Antonio, TX, United States

[5]Department of Child and Youth Studies, Brock University, St. Catherines, ON, Canada

[6]Department of Family Medicine, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Shirlene Wang, PhD
Department of Population and Public Health Sciences, University of Southern California
1875 N Soto Street
Los Angeles, CA, 90032
United States
Phone: 1 3125327663
Email: shirlene@northwestern.edu

Abstract

Background: Adolescence through emerging adulthood represents a critical period associated with changes in lifestyle behaviors. Understanding the dynamic relationships between cognitive, social, and environmental contexts is informative for the development of interventions aiming to help youth sustain physical activity and limit sedentary time during this life stage. Ecological momentary assessment (EMA) is an innovative method involving real-time assessment of individuals' experiences and behaviors in their naturalistic or everyday environments; however, EMA compliance can be problematic due to high participant burdens.

Objective: This systematic review synthesized existing evidence pertaining to compliance in EMA studies that investigated wake-time movement behaviors among adolescent and emerging adult populations. Differences in EMA delivery scheme or protocol, EMA platforms, prompting schedules, and compensation methods—all of which can affect participant compliance and overall study quality—were examined.

Methods: An electronic literature search was conducted in PubMed, PsycINFO, and Web of Science databases to select relevant papers that assessed movement behaviors among the population using EMA and reported compliance information for inclusion (n=52) in October 2022. Study quality was assessed using a modified version of the Checklist for Reporting of EMA Studies (CREMAS).

Results: Synthesizing the existing evidence revealed several factors that influence compliance. The platform used for EMA studies could affect compliance and data quality in that studies using smartphones or apps might lessen additional burdens associated with delivering EMAs, yet most studies used web-based formats (n=18, 35%). Study length was not found to affect EMA compliance rates, but the timing and frequency of prompts may be critical factors associated with missingness. For example, studies that only prompted participants once per day had higher compliance (91% vs 77%), but more frequent prompts provided more comprehensive data for researchers at the expense of increased participant burden. Similarly, studies with frequent prompting within the day may provide more representative data but may also be perceived as more burdensome and result in lower compliance. Compensation type did not significantly affect compliance, but additional motivational strategies could be applied to encourage participant response.

Conclusions: Ultimately, researchers should consider the best strategies to limit burdens, balanced against requirements to answer the research question or phenomena being studied. Findings also highlight the need for greater consistency in reporting and more specificity when explaining procedures to understand how EMA compliance could be optimized in studies examining physical activity and sedentary time among youth.

XSL•FO
RenderX

Trial Registration: PROSPERO CRD42021282093; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=282093

(JMIR Mhealth Uhealth 2025;13:e52887) doi:10.2196/52887

KEYWORDS

compliance; ecological momentary assessment; mobile health; adolescents; emerging adults; physical activity; movement behavior; systematic review; cognitive; social; development; youth; literature search; inclusion; data quality; mobile phone

Introduction

The transition from adolescence (aged 10-18 years) to emerging adulthood (aged 18-29 years) represents a critical developmental period [1] that is often associated with significant changes in lifestyle behaviors [2-4]. To establish more effective interventions or programs to help youth (ie, adolescents and emerging adults) promote or sustain more physical activity (PA) and limit sedentary time (ST), there has been growing interest in studying the dynamic relationships between youths' cognitions, social and environmental contexts, and movement behaviors. Research has consistently shown greater PA and less ST being beneficial for several health outcomes for both adolescents and emerging adults including self-reported health status, chronic diseases such as obesity and cardiovascular disease, and indicators of mental health [5,6]. Ecological momentary assessment (EMA) is an innovative method for collecting intensive longitudinal data to better understand the dynamic predictors of these critical health-promoting behaviors [7]. In brief, EMA involves repeated real-time assessment of individuals' experiences and behaviors in their real world and natural environments, which do not require participants to recall extended periods of time, helping to reduce participant recall biases and increase ecological validity [8]. These methods have been implemented in several studies to date, helping to advance knowledge related to real-time correlates of PA and ST among youth. While there has been considerable progress in this field, EMA compliance can, at times, be problematic among the youth population. Our understanding of EMA compliance among movement behavior studies remains poor and thus warrants greater attention to help inform the development of future and more effective EMA protocols.

Data collection efforts involving repeated participant observations such as EMA allow researchers to study the temporal patterns and momentary processes that influence movement behaviors while allowing for the disaggregation of data into between-subjects and within-subjects for analysis [9]. EMA methods have evolved over time, beginning with more traditional methods such as paper and pencil diaries and repeated telephone calls, to the integration with technology using palm computers or SMS text messaging, and now smartphone apps developed for EMA [10]. There are several sampling approaches used in EMA, each suited to different types of research questions and contexts. In signal-contingent sampling, participants are prompted to provide data at random or within predetermined intervals. In contrast, data are only collected after the occurrence of specific events in event-contingent sampling. Given the ubiquity and rapid technological advances of smartphones to reduce some of the participant burden, it is not surprising that EMA research among adolescent and emerging adulthood populations has risen over recent years. Delivering EMA protocols through smartphones is often considered a feasible and acceptable method to collect intensive longitudinal data among youth samples due to their high rates of device use [11,12]. In other words, the utilization of smartphones has the unique potential to improve our understanding of correlates of PA and ST, but poor adherence to the EMA protocol may bias the associations observed, ultimately hindering opportunities to develop effective interventions.

Despite the aid of technology, EMA using smartphones remains a burden-heavy method for data collection for participants. Its repeated self-report nature requiring conscious effort lends itself to potentially substantive missing data as answering prompts in naturalistic settings may not always be feasible. For example, participants may not be able to answer an EMA prompt in a timely manner due to engagement in sport or PA, being prompted in unsafe contexts (eg, while driving), or if the prompt was sent at an inconvenient time (eg, socializing with friends or napping). For these reasons, it is understood that missingness in EMA research is inevitable as an increased burden leads to lower study compliance. However, missing data that results from a low EMA response rate raises questions about the quality of the data being collected. Despite its time-consuming nature, compliance with EMA is essential for the validity or representativeness of daily phenomena that may influence movement behaviors among youth [13].

Compliance is commonly defined as the ratio of the number of measurement occasions participants completed divided by the maximum number of measurement occasions allowed by the protocol of the EMA study [14]. Conducting an EMA study involves multiple decision-making processes to determine the protocol, including but not limited to the EMA assessment tool and platform, compensation, study duration, assessment frequency, sampling scheme, and questionnaire density [15]. Missingness and low compliance can impact the data quality and be susceptible to misinterpretation. For example, systematic missingness could introduce bias as the resulting data does not reflect the true distribution of behaviors. Similarly, the overall sample may not represent the intended population and limit generalizability. Missing data also reduces sample size, which decreases the statistical power of the study or increases the variability of estimates. Finally, if there is missing data at critical times (eg, ignoring prompts during stressful times or missing prompts because they are engaging in PA within a PA study), the trends could be misinterpreted. While there are techniques to impute missing data, imputing EMA data can be challenging due to the inherent dynamic nature of the data and lead to inaccurate results. Given that compliance is critical to understanding the representativeness of the collected data and earlier publications pushing for compliance reporting for EMA

XSL•FO
RenderX

studies [7,16,17] it is crucial to understand the factors and study methodological features that may be influencing EMA compliance.

Indeed, studies have examined the application of EMA in youth. Previous reviews have focused on summarizing or reporting the main characteristics of studies that apply EMA to study mental health or psychopathology in daily life. A paper by Mason et al [18] examined the association between EMA-measured contextual factors and dietary intake in youth. Liao et al [16] conducted a systematic review of common EMA methods of diet and PA research in youth. In a previous systematic review of EMA protocol and compliance in adolescents, Wen et al [12] concluded that compliance was similar for studies longer than 3 weeks compared to studies less than 1 week. Interestingly, another systematic review of EMA studies in the adult population found no identified associations between any features of the EMA protocol and the compliance rate [17]. Given that youth populations may not fully understand what would be required in an EMA study, or can also be generally ambivalent toward research, it is crucial to understand potential methods and practices that may result in greater compliance. The existing reviews included EMA studies that focused on a wide range of health outcomes, which makes it difficult to disentangle the potential impact of EMA protocols for specific behavioral outcomes. Movement behaviors, which can include PA and sedentary behaviors may be particularly unique, as the continual collection of movement behavior data through self-reports or device-based assessments may represent an additional source of participant burden. A more nuanced understanding of EMA compliance in movement behavior studies is needed. In addition, compliance cannot be solely interpreted without considering the quality of EMA design, as poor design features also lead to a greater burden and decreased compliance. Suggested guidelines for reporting EMA study protocol, such as the Checklist for Reporting of EMA Studies (CREMAS) framework, have been adopted by EMA researchers to evaluate the quality and report the design features in a standard manner [16].

Therefore, the purpose of this systematic review was to apply the CREMAS guidelines to (1) synthesize the existing evidence on compliance for EMA studies focusing on wake-time movement behaviors (ie, PA and ST) among adolescent and emerging adult populations, and (2) evaluate the quality of the included EMA studies. We summarize the repeated measure methodologies used when examining PA and ST behaviors and describe any salient factors that may positively or negatively impact EMA compliance.

Methods

Protocol and Registration

Before conducting this systematic review, a protocol was developed and preregistered on PROSPERO (submitted October 21, 2021; registration number CRD42021282093). The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed [19], and items were reported using the PRISMA checklist (Multimedia Appendix 1).

Search Strategy

An electronic literature search was conducted in the PubMed, PsycINFO, and Web of Science databases on October 19, 2022. First, titles and abstracts of papers were searched using the following terms: (“ambulatory assessment” OR “experience sampling” OR “ecological momentary assessment” OR “EMA” OR “ESM” OR “daily diary”) AND (“physical activity” OR “exercise” OR “sedentary” OR “movement behavior” OR “movement-based behavior” OR “posture” OR “sport” OR “sitting” OR “screen” OR “inactivity” OR “leisure activity”) AND (“young adult*” OR “youth” OR “college student*” OR “university student*” OR “emerging adult*” OR “adolescence” OR “adolescent” OR “high school students” or “secondary school students” OR “post secondary students”). The same search strategy was used across all 3 databases.

Inclusion and Exclusion Criteria

We included studies that met the following criteria: (1) published in a peer-reviewed academic journal, (2) full text was available, (3) published in the English language, (4) delivered EMA prompts at least one time each day during the study period (minimum requirement for daily diary studies), (5) provide quantitative data for movement-based activity (ie, PA or ST; to exclude protocol papers), (6) provide quantitative compliance data (% or number of completed) for EMA prompt responses, and (7) the sample consisted of adolescents or young adults (aged 10-29 years). Papers were excluded if (1) the population included a mix of participants older or younger than the target age range (ie, <10 or >29 years of age), (2) the study design was an intervention, (3) the paper reported on multiple studies that used EMA, or (4) the paper was not the first study to publish using a particular data set (ie, an EMA study was only counted once when multiple publications came from the same EMA project; the first publication was included in these instances so each study protocol was only represented once).

Study Selection

All records found from the database searches were imported into Covidence systematic review software (Veritas Health Innovation). Duplicates were removed before titles and abstracts were screened independently by 2 reviewers (SW and AC) for initial inclusion. All studies retained after the title and abstract screening were retrieved as full papers and further assessed for final inclusion by 2 reviewers independently. Disagreements at both screening stages were resolved by discussion during review team meetings.

Data Extraction and Synthesis

For all studies that met our inclusion criteria, the lead author (SW) extracted the following data into an Excel (Microsoft Corp) data entry form: (1) authors and year of publication; (2) sample characteristics (ie, sample size, country, mean age of sample); (3) EMA device or platform; (4) types of movement activity that was assessed, including whether it was self-reported or device-measured; (5) coassessment of diet; (6) study duration; (7) prompting schedule (event or time or signal contingent); (8) prompt frequency; (9) number of items for each prompt; (10) prompt times; (11) compensation or types of incentive; and (12) reported details about compliance (rate, inclusion criteria, and

XSL•FO
RenderX

missing data relationships). Multimedia Appendix 2 [20-68] presents the extracted data. In addition, a random sample of 20% of the included papers was independently extracted by a second review team member (AC) for validation (interrater reliability >80%). The compliance rate was calculated for studies that reported the total number of surveys completed or the average number of surveys completed by dividing by the total number of expected delivered surveys as reported by the prompting scheme in the paper's methods. Given the population, many studies had different prompting scheduled for school days versus weekend days. Therefore, we considered the maximum or most burdensome day when calculating the density of prompts (number of prompts per day).

A quality score for each study was calculated using a modified version of the Checklist for Reporting of EMA Studies (CREMAS) [16]. The checklist was used to assess if descriptions were provided for EMA technology, EMA protocol training, study duration, prompting scheme, prompting frequency, study attrition, prompt latency, missing data, and limitations of EMA. Each item was scored with 0, 0.5, or 1 (representing No, Somewhat, Yes). Scores ranged from 0 to 9, with greater scores representing higher quality. Scores can be found in Multimedia Appendix 2 [20-68].

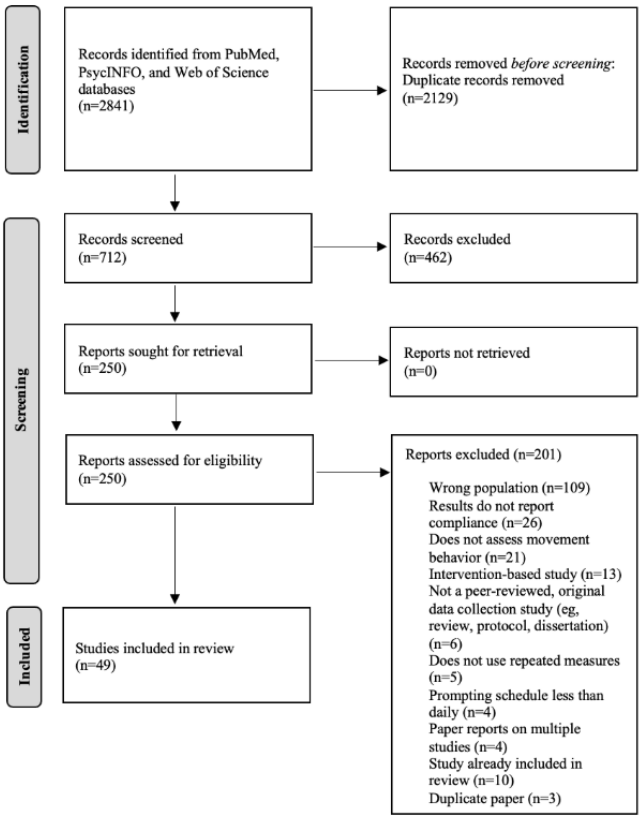
Given the heterogeneity in the samples and methods in the studies that met the inclusion criteria, further analysis through conducting a meta-analysis was not appropriate. We proceeded with our objective of conducting a narrative synthesis, and studies were reviewed and compared on a variety of characteristics, including PA/ST assessment method, EMA platform, number of study days, prompting schedule, prompting frequency, number of prompts per day, compensation, and the comeasurement of diet. Differences in compliance between groups in each category were analyzed using Kruskal-Wallis tests. Additionally, potential interactions between significant factors were presented in contingency tables as many of the factors we examined could coexist with each other to compound burden (EMA delivery scheme × study duration × prompt frequency). For example, very intensive prompting (multiple times per day) may only be tolerable for a short period of time in comparison to end-of-day prompts completed 1 time per day.

Results

Overview

Our systematic search conducted at 2 time points (ie, November 2021 and October 2022) captured a total of 2841 papers, of which 2129 papers were duplicates and therefore removed. A total of 712 remaining papers were initially screened. After title and abstract screening, an additional 462 papers were deemed irrelevant, with the remaining 250 papers proceeding for full-text screening. Following the full-text review, 201 papers were found to have not met the eligibility criteria, resulting in 49 papers retained for data extraction and included in this review [20-68]. A PRISMA diagram for selecting the final papers for synthesizing is presented in Figure 1.

Figure 1. Adapted PRISMA flow diagram for identification and selection of studies. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



XSL•FO
RenderX

Study Quality

Table 1 presents the quality scores of the final papers. The average CREMAS score was 6.84 (SD 1.26). All studies reported (yes or somewhat) the study duration, prompting scheme, and prompting frequency. However, studies most infrequently reported on if and how participants were trained on the EMA protocol (n=18, 37% not reported), how missing data were handled (n=28, 57% not reported), attrition (n=26, 53% not reported), and latency (n=42, 86% not reported). These results highlight the continued variability in reporting critical attributes within PA and ST research among adolescents and emerging adults when applying EMA methodologies.

Table 1. Study (n=49) quality scores based on the adapted CREMAS[a] tool.

CREMAS item	Yes, n (%)	Somewhat, n (%)	No, n (%)
EMA[b] technology	35 (71)	12 (25)	2 (4)
EMA[b] training	23 (47)	8 (16)	18 (37)
Study duration	48 (98)	1 (2)	0 (0)
Prompting scheme	47 (96)	2 (4)	0 (0)
Frequency	48 (98)	1 (2)	0 (0)
Attrition	11 (22)	12 (25)	26 (53)
Latency	3 (6)	4 (8)	42 (86)
Missing data	19 (39)	2 (4)	28 (57)
Limitations of EMA[b]	26 (53)	10 (20)	13 (27)

[a]CREMAS: Checklist for Reporting of EMA Studies.

[b]EMA: ecological momentary assessment.

Descriptive Statistics

PA was the primary outcome in 40 (82%) of the 49 studies. More studies used self-reports of PA and ST (n=29, 59%) than device-based assessments (n=20, 41%) to measure movement behaviors. For compensation, 9 (18%) studies provided some form of course credit, 12 (24%) studies provided only monetary compensation, and 5 (10%) studies provided both options. In contrast, 1 study explicitly stated that no compensation was provided for participation, and another 22 (45%) of the 49 studies did not report how participants were compensated.

Although most studies were conducted in the United States (n=36, 73%) among nonclinical samples (n=46, 94%), there was substantial heterogeneity in the EMA study methodologies. For example, the sample size of these studies ranged from 8 to 1493 participants (mean 179.77, SD 272.43), study durations from 1 to 30 days (including 4 studies that used multiple waves of EMA data collection periods [29,48,53,61]), and a variety of prompting schedules being applied. A total of 12 (24%) studies consisted of daily dairies, whereby participants were asked to respond to a single EMA prompt once each day [27-29,31,35,39-41,54,59,62,66]. Another 36 (73%) studies involved multiple prompts each day, including 4 studies that used event-contingent prompting [20,36,44,65]. These 4 studies required participants to self-report eating episodes [65], or the prompts were triggered by their current location [44] or current activity [20,36]. Among the 32 studies with only a time-based sampling component (predetermined sets of prompts being administered), prompting schedules included random prompts during predetermined time intervals (n=30, 61%; eg, one random prompt for each 2-hour interval), prompts at a fixed schedule (n=8, 16%; eg, every 30 min during waking hours), or prompts at a personalized time (n=1, 2%; eg, participant's blood glucose check schedule). One study did not report the prompting scheme. A total of 13 (27%) of the 49 studies did not report the length of the EMA surveys. Of the studies that were reported, the number of items per prompt ranged from 1 to 45 (mean 11.78, SD 8.80). There was a negative but nonsignificant correlation between the length of the survey and compliance rate (r=–0.10; P=.57).

Compliance Results

Overall, compliance data were extracted for 48 studies, as we could not extract the compliance rate for 1 study that reported the average number of event-contingent prompts responded to [64]. The compliance rate was calculated for studies that reported the total number of surveys completed or the average number of surveys completed by dividing by the total number of expected delivered surveys as reported by the prompting scheme in the paper's methods. Importantly, it should be noted that 13 (27%) of the 48 studies reported dropping participants from their analyses—which invariably improved calculations of compliance. Decision rules to exclude participants in final analyses varied, with various thresholds of minimum compliance ranging from 30% to 80% missing data or removal of participant data if compliance rates were beyond 2-3 standard deviations. It may also be possible that the remaining 35 studies also removed participant data without being specified. Among the 48 studies with reported compliance, the mean compliance was 80.63% (range 40.2%-99.3%). The mean compliance, however, was higher among the daily diary studies (91.03%; range 88.3%-96%) than studies with multiple prompts each day (77.16%; range 40%-95%).

Overall, mean compliance among the daily diary studies was relatively high and consistent across studies. Compliance was greater than 88% in all studies, with study durations ranging

XSL•FO
RenderX

from 6 to 30 days of data collection. Among the 12 daily diary studies, 10 studies reported using web-based platforms, including 1 study that also used paper and pencil reporting. One study was phone-based, and 1 study did not report the modality of EMA collection. All studies were published after 2013, but 9 of these daily diary studies were published within 5 years of this conducted review.

The mean compliance among studies with multiple prompts per day was more variable when compared to daily diary studies, with reported compliance as low as 40% to as high as 94%. In examining prompt type and duration of EMA studies, there were no distinct patterns to compliance based on whether studies applied fixed or random prompting scheduled duration of the EMA studies. There is more variability when examining the interactions between prompt schedules and study duration (range 71.45%-90%), but with a caveat that some include only a small number of studies (eg, only 2 studies that used a fixed prompting schedule for less than 1 week).

With regard to prompt schedules, among these multiple EMA prompt studies, there was a substantial number of studies (n=8) that included very high numbers of EMA surveys being delivered per day (range 17-68). Interestingly, the mean compliance of these studies was higher (82.10%; range 50.20%-99.30%) when compared to studies (n=27) that included 3 to 9 prompts per day (76.43%; range 40%-89.30%). Two studies did not specify the number of prompts they sent each day. Studies with the most intensive EMA protocols (ie, ≥17 prompts per day) had compliance of greater than 82%, apart from 1 study with 50% compliance. The highest compliance rate (99%) was reported by Kanning et al [50], who intensively delivered EMA prompts every 45 minutes for 1 day to study the real-time association of PA with affect. Three studies used extra devices to send electronic EMA surveys, while 4 studies used paper and pencil EMA methods.

Upon closer examination of studies prompting participants between 3 and 9 times each day, 19 of the 27 studies were found to have had compliance above 75%. Interestingly, 2 studies used paper and pencil as the modality of EMA collection; however, both studies had compliance between 76% and 77%. Studies with lower compliance varied in sample size (ie, 20 to 284 participants), study duration (ie, 4 to 30 days), prompting schedules, forms of compensation, and year of publication (ie, 2013-2022).

Factors Related to Compliance

Due to the small sample size of studies in various categories of synthesizing, nonparametric Kruskal-Wallis tests were performed to determine the effects of the factors on median compliance rates. These results are presented in Table 2. The results indicated nonsignificant differences for the movement behavior assessment method (χ^2_2=0.2; *P*=.88); number of items per prompt (χ^2_3=1.2; *P*=.75); number of days (χ^2_3=1.1; *P*=.77); or compensation structure (χ^2_4=3.9; *P*=.42). There was a statistically significant difference between compliance based on prompting frequency (χ^2_3=15.2; *P*<.001) and the binned number of prompts per day (χ^2_3=1.2; *P*<.001). Furthermore, a statistically significant difference was observed for compliance based on the prompting schedule (χ^2_4=9.9; *P*=.04) and compliance based on the EMA platform (χ^2_5= 9.5; *P*=.04). Additional linear regression analyses did not find a statistically significant relationship between the number of study days and the number of prompts per day (coded as continuous variables) with compliance. There was no statistically significant difference in mean compliance between the studies that only used objective assessments (n=15; 81.57%) versus self-reported PA (n=33; 80.20%; t_{46}=–0.11; *P*=.46). There was no statistically significant difference in compliance between the studies that had PA as the primary outcome (40/48, 81.22%) versus not (8/48, 77.66%; t_{46}=–0.23; *P*=.41).

XSL•FO
RenderX

Table 2. Comparisons of compliance by EMA[a] study design factors.

Factor	Value, n	Median compliance	Rank sum	Chi-square (*df*)	*P* value
PA[b]/ST[c] assessment				0.2 (2)	.88
Device assessed	15	81.57	381.5		
Self-report	29	80.25	688.5		
Both	4	84.70	106		
EMA platform				11.4 (5)	.04
App on personal device	8	75.21	132.5		
Extra device	13	79.29	278		
Paper and pencil	7	74.60	127.5		
Telephone call	2	80.00	53		
Web-based	16	86.15	503		
NR[d]	2	91.33	82		
Days				1.1 (3)	.77
<7	9	78.37	226.5		
7-14	21	81.27	527		
14-28	13	83.10	331		
>28	5	75.56	91.5		
Prompting schedule				9.5 (4)	.04
Event	2	79.25	30		
Event + interval	2	81.60	40		
Random	9	80.10	178		
Interval	27	83.95	805		
Random + interval	8	70.14	122		
Prompting frequency				15.2 (3)	<.001
Daily	12	91.03	456.5		
Event + multiple	3	80.17	51		
Multiple	32	76.72	646		
NR[d]	1	82.5	22.5		
Number of prompts per day				17.7 (3)	<.001
1	12	91.03	394.5		
2-4	7	77.23	119		
4-10	19	77	306.5		
>10	4	80.65	83		
Number of items per prompt				1.2 (3)	.75
0-10	16	79.54	301		
11-20	15	78.33	293		
21-30	4	82.95	86		
41-50	2	73.51	23		
Compensation				3.9 (4)	.42
Course credit	9	81.34	235.5		
Monetary	12	80.34	250.5		
Both	5	80.90	112.5		

XSL•FO
RenderX

Factor	Value, n	Median compliance	Rank sum	Chi-square (*df*)	*P* value
None	1	99.30	48		
NR[d]	21	79.53	529.5		

[a]EMA: ecological momentary assessment.

[b]PA: physical activity.

[c]ST: sedentary time.

[d]NR: not reported.

Discussion

Principal Findings

This systematic review addressed a current knowledge gap by synthesizing the body of literature that has used EMA methodologies to investigate adolescents' and emerging adults' movement behaviors for the purpose of identifying protocol-related factors that influence EMA compliance. Only 1 previous study has specifically examined factors related to compliance among youth. Heron et al [11] conducted a systematic review to provide best practices by identifying common approaches and challenges and providing reporting guidelines for compliance; however, do not specifically identify factors besides recommendations for staff to complete compliance check-in and implement compliance-based incentives. While there is substantial heterogeneity among the included EMA studies, the pooled compliance rate was high and comparable to that reported in previous reviews. Unlike the few findings in previous literature, this review identified several specific EMA design factors based on the CREMAS checklist (ie, delivery scheme, study duration, and prompt frequency), including factors that interact to influence participants' compliance in responding to EMA surveys. It should be noted, however, that the disparities in compliance attributable to EMA design factors were rather small as this population surprisingly responded consistently to prompts. This being said, there were no major discrepancies based on the factors we analyzed that led to considerable missingness.

Associations Between PA Assessment Tool and Compliance

A total of 20 studies measured PA using an accelerometer which may have reduced self-report biases and recall errors attributable to self-reported instruments [69] but potentially increased the burden associated with the study overall. We observed no significant difference in EMA compliance based on the PA assessment method. This suggests that compliance rates to EMA are not better when participants must wear a separate device (but do not have to self-report their PA levels through EMA). Despite the potential additional burden of wearing an accelerometer (uncomfortable, unfashionable, or inconvenient), participants did not reduce response rates to EMA.

Associations Between EMA Duration, Prompting Schedule, and Compliance

Overall, there was no difference in compliance between schedules of different study lengths. Most studies lasted between 7 and 14 days (n=21). Our findings suggest that the timing and frequency of EMA prompts may be a more critical factor in EMA design when attempting to optimize compliance. Logically, participants are more likely to respond if prompts align with their schedule or if they occur at predictable intervals. Therefore, the frequency of EMA prompts must be balanced to minimize the burden while providing researchers with sufficient representative EMA data for analysis. At the same time, longer study durations (eg, ≥14 days) may provide more representative data on participants' behavior and experiences but can be perceived as more burdensome and result in noncompliance. One potential explanation could be that participants have incorporated EMA into their routines, and the behavior of responding has become a habit. It is important to acknowledge, however, that while participants may initially be able to handle the prompting schedule, longer EMA studies may be subject to temporal influences like schedule changes (eg, summer break for adolescents and transition to another work schedule for emerging adults) that could threaten compliance rate.

Ultimately, researchers should consider the minimum required number of days or prompts per day necessary to answer their research question based on the nature of the phenomena being studied. How frequently are movement behaviors or the predictors of the behavior expected to change? These considerations need to be weighed against participants' daily schedules and routines. For example, some studies involving adolescents intentionally did not prompt participants during school hours, which stands to be missed when this population tends to have opportunities to be physically active (eg, physical education, clubs, and athletics) [70]. Researchers could take this further and use adaptive sampling techniques to adjust the frequency based on contextual factors such as weekday school attendance or prompting participants between structured class times (ie, recess time or transitions between classes). In addition, most studies used only an interval-based design (n=27). These studies had the highest compliance, which supports the assumption that receiving prompts at consistent intervals is less disruptive than prompts at random and unpredictable times. Additional consideration may include the use of passive-sensing (eg, wearable monitoring) in conjunction with EMA or active-sensing, requiring participants to respond to prompts being delivered.

Associations Between the EMA Platform and Compliance

We found that the platform used to deliver the EMA study or protocol could significantly impact compliance and data quality. In our sample of studies, most studies (n=16) used a web-based format (Qualtrics survey) to administer EMA. However, we did not disentangle how the link was provided to participants. For example, some researchers may have emailed the link, whereas

XSL•FO
RenderX

others sent an SMS text message to participants. Both options still capitalize on the convenience of mobile devices. While carrying an extra device (n=13) may have been initially novel to youth, it may be considered burdensome now, especially in studies that require participants to wear an accelerometer to capture movement behavior. Due to the increased prevalence of smartphone ownership in adolescents, it is expected that EMA researchers will adopt smartphones or smartphone apps to lessen the additional burdens associated with delivering EMAs. Within smartphone apps, it is worth considering additional factors that may influence compliance, such as the user interface or design of the platform or the use of gamification aimed at increasing engagement [71]. Intensive self-monitoring is becoming more acceptable or feasible, especially in studying health and mental health, and there is evidence of youth and emerging adults using self-care health-tracking apps in daily life with high engagement [72]. But beyond wearable devices, the adoption of self-monitoring motivating factors is still in the early stages of PA research [70].

Associations Between EMA Response Frequency and Prompts Per Day and Compliance

The adoption and maintenance of movement behaviors in youth are complicated as there is a combination of factors with different possible temporalities. Due to the feasibility and to support more advanced theories, more studies have included multiple prompts per day (ie, 12 daily diary studies vs 36 studies that prompted numerous times per day). However, we observed that the daily diary studies had higher compliance. While more frequent prompts provide more comprehensive data for researchers, this also increases participants' burdens. Frequency should depend on the research question, the population being studied, and the EMA platform being used.

Associations Between Participant Compensation Type and Compliance

Most included studies did not adequately report their compensation mode or protocol (n=21). Based on the studies that did report compensation type, there was no difference between providing monetary compensation or course credit. With monetary compensation, there are also different levels and the potential need to determine a rate per unit of burden, which is beyond our analyses. Beyond incentives, researchers should consider additional motivational strategies to encourage compliance with more burdensome schedules. For example, participants could be more compliant if the research question feels relevant and reminded of the importance of the research they are participating in. Alternatively, participants may prefer reporting on positive health behaviors, depending on whether the study participants were able to or interested in reviewing their data and the relationships found. Due to self-presentation bias, researchers must emphasize that all data are valuable to prevent participants from not responding if they did not engage in positive health-enhancing behaviors such as PA. Importantly, aesthetically pleasing apps that minimize usability issues will likely be critical factors that help maximize compliance in EMA studies examining PA and ST among youth [69].

Combinations of Different EMA Features and Compliance

In addition to studying the independent impact of EMA design parameters, this study provided unique contributions to the literature by inspecting the combinations of different EMA design features and their associations with compliance among youth. Based on the studies included, we found that daily diaries consistently provided an average of at least 90% compliance regardless of study duration. However, it should be noted that even the more burdensome delivery schemes with multiple EMA prompts per day with 7-14 days of study duration had relatively higher compliance rates, primarily when the prompts were delivered at random times. As the daily EMA prompts increase to 9 prompts or more, the predictable fixed interval yields the highest compliance rate. The caveat is that only 1 EMA study in this category provided the highest compliance rate (99%), but the study consisted of only 1 day of intensive prompting. There was no noticeable interaction between EMA prompting density and study duration, except that longer EMA studies with multiple random prompts per day (<9) had relatively lower compliance (75%).

Strengths and Limitations of This Review

This review follows a preregistered search strategy, evaluates compliance to EMA by individual and design factors, and assesses an understudied age range. While this review addresses a critical gap in our current knowledge, several limitations exist. First, we did not account for sample size in our analyses. Second, other relevant EMA studies may not be identified in this review based on our preregistered search criteria. This may be particularly important with the high likelihood of some publication biases, with potential studies with low participant compliance remaining unpublished. Third, and similarly, many of the studies excluded or dropped participants' data with lower compliance from analyses, which may bias our findings to reveal higher study compliance. Nevertheless, EMA studies often have selection bias such that participants who join anticipate being able to be compliant. Our results are still useful as we can still learn from which features are most relevant and predictive of compliance among all these high compliance studies. Fourth, our analyses are based on what was reported in the papers—authors were not contacted to provide additional information when reporting was inadequate. Fifth, the outcome of this study was compliance which we used as a proxy for participant or study burden. Many studies have directly measured burden through usability scales or qualitative studies. Moreover, we did not extract the constructs assessed by the studies through EMA. It is possible that some constructs are more cognitively difficult and cause more burden leading to lower compliance due to participants avoiding completing them. One of our exclusion criteria removed papers that reported on multiple studies; while this simplified data analysis, this may have reduced the information our review gathered on compliance. We excluded papers that reported on multiple studies to maintain consistency in the studies analyzed. Given we were assessing the quality of reporting, authors may have needed to reduce the details of study methods in their papers to meet journal word counts, which obscure insights into compliance trends. Additionally, if the second study described

XSL•FO
RenderX

in the paper is an improved iteration of the first study, this contaminates the pooled analyses. Finally, following the CREMAS guidelines, we only extracted data if studies reported on various factors outlined in the checklist, which may exclude extracting potentially relevant details such as what participant training entailed or what attrition rates were.

Recommendations and Future Directions

This systematic review closely followed each reporting criterion within CREMAS. Additionally, the study explores interactions between some of the more salient factors likely to impact EMA compliance in PA and ST research. Although there was no precise combination of factors that were indicative of higher compliance, especially among delivery schemes with multiple EMAs each day, it should be reiterated that the research design needs to fit the research question and the phenomena of interest—and researchers should think about how to minimize participant burden wherever possible. Importantly, as EMA research continues to increase rapidly in this field, authors must report as much as possible about the design of the study and compensation so future researchers can replicate successful designs.

Beyond what was examined in this review, other factors must be further explored as factors associated with EMA compliance. The first is related to participant training. We did not extract the quality or quantity of EMA training or levels of participant contact in each study—although this was not reported in most studies. Especially with the adoption of mobile devices, researchers must be able to support participants in the case of technical difficulties. Additionally, reporting on difficulties encountered and their impact on compliance would be beneficial, particularly as more studies adopt passive sensing devices to assess PA and ST. Second, more specificity in terms of participant recruitment would also be helpful. How participants were recruited for EMA studies may be particularly important. There may be certain response biases that exist, as participants may not fully understand the high degree of participant burden in community sampling (eg, going into places and spaces to recruit) when compared to participants who were required to reach out to a study team or make an effort to go into a lab or receive in-person training for the study. Moreover, the length of the EMA survey may play a role. We found that the studies with the most EMA prompts per day had high compliance rates, but this may be due to short surveys. The length of time it takes to complete the surveys (number of questions per survey or total completion time) may be a metric for future study. Additionally, greater transparency is needed in reporting, as many papers only report on the number of items that were analyzed when many more items could have been collected. Additionally, this paper was not able to further examine if the amount within the compensation structure matters, as most papers did not disclose the structure. As this may be an important factor, papers should report this information for future analysis. Preregistration on platforms such as the Open Science Framework may help to address this shortcoming. Finally, while the papers all reported on compliance, less than half the studies reported on missing data relationships by assessing if data were missing at random or related to demographic factors. More attention and reporting should be encouraged around patterns of missingness (sporadic, complete attrition after a given point, only during specific windows). Similarly, as compensation is often used as a technique to encourage compliance, the amount of compensation or details on the compensation structure should be reported for future researchers to specifically examine this factor.

Conclusions

As EMA methodology becomes more prevalent in movement behavior research among youth, it is imperative to understand the potential impact of specific EMA protocol features on study compliance to reduce participant burden and enhance data quality. While this intensive data collection effort involving repeated observations of the participants enables researchers to study the temporal patterns and momentary processes that influence PA and ST, compliance is critical for data quality and validity of the inferences that can be made. Although overall findings did indicate relatively high EMA compliance for PA and ST studies among adolescents and emerging adults, there were some differences based on delivery schemes, study duration, and prompt frequencies. Greater consistency in reporting and more specificity when explaining procedures are required to understand further how EMA compliance could be optimized among this population.

Acknowledgments

SW received funding from the National Science Foundation (award number 2150617) and was supported by the National Institutes of Health (T32CA193193). MYWK is supported by an Ontario Government Early Researchers Award and an SSHRC Tier 2 Canada Research Chair. We would like to acknowledge all the researchers who published the articles that contributed to this review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 100 KB - mhealth_v13i1e52887_app1.pdf]

Multimedia Appendix 2

XSL•FO
RenderX

Extracted data.
[XLSX File (Microsoft Excel File), 89 KB - mhealth_v13i1e52887_app2.xlsx]

References

1. Arnett JJ. Emerging adulthood: a theory of development from the late teens through the twenties. Am Psychol 2000;55(5):469-480. [doi: 10.1037/0003-066x.55.5.469]
2. Kwan MY, Cairney J, Faulkner GE, Pullenayegum EE. Physical activity and other health-risk behaviors during the transition into early adulthood: a longitudinal cohort study. Am J Prev Med 2012;42(1):14-20. [doi: 10.1016/j.amepre.2011.08.026] [Medline: 22176841]
3. Schulenberg JE, Maggs JL. A developmental perspective on alcohol use and heavy drinking during adolescence and the transition to young adulthood. J Stud Alcohol Suppl 2002(14):54-70. [doi: 10.15288/jsas.2002.s14.54] [Medline: 12022730]
4. Winpenny EM, van Sluijs EMF, White M, Klepp KI, Wold B, Lien N. Changes in diet through adolescence and early adulthood: longitudinal trajectories and association with key life transitions. Int J Behav Nutr Phys Act 2018;15(1):86 [FREE Full text] [doi: 10.1186/s12966-018-0719-8] [Medline: 30200990]
5. Zhang T, Lu G, Wu XY. Associations between physical activity, sedentary behaviour and self-rated health among the general population of children and adolescents: a systematic review and meta-analysis. BMC Public Health 2020;20(1):1343 [FREE Full text] [doi: 10.1186/s12889-020-09447-1] [Medline: 32883275]
6. Groves CI, Huong C, Porter CD, Summerville B, Swafford I, Witham B, et al. Associations between 24-h movement behaviors and indicators of mental health and well-being across the lifespan: a systematic review. JASSB 2024;3(1):9. [doi: 10.31234/osf.io/p9zhy]
7. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
8. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
9. Dunton GF. Ecological momentary assessment in physical activity research. Exerc Sport Sci Rev 2017;45(1):48-54 [FREE Full text] [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]
10. Smyth JM, Heron KE. Health psychology. In: Handbook of Research Methods for Studying Daily Life. New York, NY, US: The Guilford Press; 2012.
11. Heron KE, Everhart RS, McHale SM, Smyth JM. Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: a systematic review and recommendations. J Pediatr Psychol 2017;42(10):1087-1107. [doi: 10.1093/jpepsy/jsx078] [Medline: 28475765]
12. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
13. Courvoisier DS, Eid M, Lischetzke T. Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. Psychol Assess 2012;24(3):713-720. [doi: 10.1037/a0026733] [Medline: 22250597]
14. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. J Med Internet Res 2019;21(12):e14475 [FREE Full text] [doi: 10.2196/14475] [Medline: 31808748]
15. Hektner J, Schmidt J, Csikszentmihalyi M. Experience Sampling Method. United Kingdom: SAGE Publications, Inc; 2007.
16. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
17. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021;23(3):e17023 [FREE Full text] [doi: 10.2196/17023] [Medline: 33656451]
18. Mason TB, Do B, Wang S, Dunton GF. Ecological momentary assessment of eating and dietary intake behaviors in children and adolescents: a systematic review of the literature. Appetite 2020;144:104465 [FREE Full text] [doi: 10.1016/j.appet.2019.104465] [Medline: 31541670]
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339(jul21 1):b2535-b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
20. Koch ED, Tost H, Braun U, Gan G, Giurgiu M, Reinhard I, et al. Relationships between incidental physical activity, exercise, and sports with subsequent mood in adolescents. Scand J Med Sci Sports 2020;30(11):2234-2250. [doi: 10.1111/sms.13774] [Medline: 33448493]
21. Sufrinko AM, Howie EK, Charek DB, Elbin RJ, Collins MW, Kontos AP. Mobile ecological momentary assessment of postconcussion symptoms and recovery outcomes. J Head Trauma Rehabil 2019;34(6):E40-E48. [doi: 10.1097/htr.0000000000000474]

XSL•FO
RenderX

22. Bernstein EE, Curtiss JE, Wu GWY, Barreira PJ, McNally RJ. Exercise and emotion dynamics: an experience sampling study. Emotion 2019;19(4):637-644. [doi: 10.1037/emo0000462] [Medline: 29999384]
23. Maher JP, Harduk M, Hevel DJ, Adams WM, McGuirt JT. Momentary physical activity co-occurs with healthy and unhealthy dietary intake in African American college freshmen. Nutrients 2020;12(5):1360 [FREE Full text] [doi: 10.3390/nu12051360] [Medline: 32397433]
24. Raudsepp L, Vink K. Longitudinal associations between sedentary behavior and depressive symptoms in adolescent girls followed 6 years. J Phys Act Health 2019;16(3):191-196. [doi: 10.1123/jpah.2018-0123] [Medline: 30744487]
25. Gilchrist JD, Pila E, Lucibello KM, Sabiston CM, Conroy DE. Body surveillance and affective judgments of physical activity in daily life. Body Image 2021;36:127-133. [doi: 10.1016/j.bodyim.2020.11.004] [Medline: 33321272]
26. MacIntyre RI, Heron KE, Braitman AL, Arigo D. An ecological momentary assessment of self-improvement and self-evaluation body comparisons: associations with college women's body dissatisfaction and exercise. Body Image 2020;33:264-277 [FREE Full text] [doi: 10.1016/j.bodyim.2020.04.002] [Medline: 32473545]
27. Riley KE, Park CL, Laurenceau JP. A daily diary study of rumination and health behaviors: modeling moderators and mediators. Ann Behav Med 2019;53(8):743-755 [FREE Full text] [doi: 10.1093/abm/kay081] [Medline: 30358802]
28. Dalton ED. Exercise-related coping beliefs predict physical activity levels in response to naturally occurring stress: a daily diary study of college students. J Am Coll Health 2022;70(2):411-419. [doi: 10.1080/07448481.2020.1752218] [Medline: 32343219]
29. Riley KE, Tigershtrom A, Park CL, Lauranceau JP. Predictors of health behaviors during a college national championship sports event: a daily diary study. J Am Coll Health 2021;69(8):928-936 [FREE Full text] [doi: 10.1080/07448481.2020.1719110] [Medline: 32813633]
30. Raudsepp L, Riso EM. Longitudinal associations between sedentary behavior of adolescent girls, their mothers, and best friends. Pediatr Exerc Sci 2017;29(3):419-426. [doi: 10.1123/pes.2016-0255] [Medline: 28486024]
31. Conroy DE, Maher JP, Elavsky S, Hyde AL, Doerksen SE. Sedentary behavior as a daily process regulated by habits and intentions. Health Psychol 2013;32(11):1149-1157 [FREE Full text] [doi: 10.1037/a0031629] [Medline: 23477579]
32. Rusby JC, Westling E, Crowley R, Light JM. Psychosocial correlates of physical and sedentary activities of early adolescent youth. Health Educ Behav 2014;41(1):42-51 [FREE Full text] [doi: 10.1177/1090198113485753] [Medline: 23640122]
33. Smith KE, Mason TB, O'Connor SM, Wang S, Dzubur E, Crosby RD, et al. Bi-directional associations between real-time affect and physical activity in weight-discordant siblings. J Pediatr Psychol 2021;46(4):443-453 [FREE Full text] [doi: 10.1093/jpepsy/jsaa121] [Medline: 33313910]
34. Kracht CL, Beyl RA, Maher JP, Katzmarzyk PT, Staiano AE. Adolescents' sedentary time, affect, and contextual factors: an ecological momentary assessment study. Int J Behav Nutr Phys Act 2021;18(1):53 [FREE Full text] [doi: 10.1186/s12966-021-01121-y] [Medline: 33858416]
35. Cho D, Armeli S, Weinstock J, Tennen H. Daily- and person-level associations between physical activity and alcohol use among college students. Emerg Adulthood 2020;8(5):428-434 [FREE Full text] [doi: 10.1177/2167696818809760] [Medline: 34350067]
36. Reichert M, Tost H, Reinhard I, Schlotz W, Zipf A, Salize HJ, et al. Exercise versus nonexercise activity. Med Sci Sports Exercise 2017;49(4):763-773. [doi: 10.1249/mss.0000000000001149]
37. Bourke M, Hilland TA, Craike M. Daily physical activity and satisfaction with life in adolescents: an ecological momentary assessment study exploring direct associations and the mediating role of core affect. J Happiness Stud 2021;23(3):949-968. [doi: 10.1007/s10902-021-00431-z]
38. Gorely T, Atkin AJ, Biddle SJ, Marshall SJ. Family circumstance, sedentary behaviour and physical activity in adolescents living in England: project STIL. Int J Behav Nutr Phys Act 2009;6:33 [FREE Full text] [doi: 10.1186/1479-5868-6-33] [Medline: 19519913]
39. Kelly SM, Updegraff JA. Substituting activities mediates the effect of cognitive flexibility on physical activity: a daily diary study. J Behav Med 2017;40(4):669-674. [doi: 10.1007/s10865-017-9839-x] [Medline: 28255751]
40. Ratcliff MB, Zeller MH, Inge TH, Hrovat KB, Modi AC. Feasibility of ecological momentary assessment to characterize adolescent postoperative diet and activity patterns after weight loss surgery. Surg Obes Relat Dis 2014;10(4):705-710 [FREE Full text] [doi: 10.1016/j.soard.2014.01.034] [Medline: 24837564]
41. Maher JP, Doerksen SE, Elavsky S, Conroy DE. Daily satisfaction with life is regulated by both physical activity and sedentary behavior. J Sport Exerc Psychol 2014;36(2):166-178. [doi: 10.1123/jsep.2013-0185] [Medline: 24686953]
42. Andorko ND, Rakhshan-Rouhakhtar P, Hinkle C, Mittal VA, McAllister M, DeVylder J, et al. Assessing validity of retrospective recall of physical activity in individuals with psychosis-like experiences. Psychiatry Res 2019;273:211-217 [FREE Full text] [doi: 10.1016/j.psychres.2019.01.029] [Medline: 30658204]
43. Sala M, Brosof LC, Rosenfield D, Fernandez KC, Levinson CA. Stress is associated with exercise differently among individuals with higher and lower eating disorder symptoms: an ecological momentary assessment study. Int J Eat Disord 2017;50(12):1413-1420 [FREE Full text] [doi: 10.1002/eat.22799] [Medline: 29098699]
44. Dunton GF, Dzubur E, Intille S. Feasibility and performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity. J Med Internet Res 2016;18(6):e106 [FREE Full text] [doi: 10.2196/jmir.5398] [Medline: 27251313]

XSL•FO
RenderX

45. Dunton GF, Whalen CK, Jamner LD, Floro JN. Mapping the social and physical contexts of physical activity across adolescence using ecological momentary assessment. Ann Behav Med 2007;34(2):144-153. [doi: 10.1007/bf02872669]
46. Brannon EE, Cushing CC, Crick CJ, Mitchell TB. The promise of wearable sensors and ecological momentary assessment measures for dynamical systems modeling in adolescents: a feasibility and acceptability study. Transl Behav Med 2016;6(4):558-565 [FREE Full text] [doi: 10.1007/s13142-016-0442-4] [Medline: 27678501]
47. Fitzsimmons-Craft EE, Ciao AC, Accurso EC. A naturalistic examination of social comparisons and disordered eating thoughts, urges, and behaviors in college women. Int J Eat Disord 2016;49(2):141-150 [FREE Full text] [doi: 10.1002/eat.22486] [Medline: 26610301]
48. Bickham DS, Blood EA, Walls CE, Shrier LA, Rich M. Characteristics of screen media use associated with higher BMI in young adolescents. Pediatrics 2013;131(5):935-941 [FREE Full text] [doi: 10.1542/peds.2012-1197] [Medline: 23569098]
49. Bohnert A, Burdette K, Dugas L, Travers L, Randall E, Richards M, et al. Multimethod analyses of discretionary time use and health behaviors among urban low-income african-American adolescents: a pilot study. J Dev Behav Pediatr 2013;34(8):589-598. [doi: 10.1097/DBP.0b013e3182a5d297] [Medline: 24131882]
50. Kanning M, Ebner-Priemer U, Brand R. Autonomous regulation mode moderates the effect of actual physical activity on affective states: an ambulant assessment approach to the role of self-determination. J Sport Exerc Psychol 2012;34(2):260-269. [doi: 10.1123/jsep.34.2.260] [Medline: 22605366]
51. Kirshnit CE, Ham M, Richards MH. The sporting life: athletic activities during early adolescence. J Youth Adolescence 1989;18(6):601-615. [doi: 10.1007/bf02139076]
52. Rouse PC, Biddle SJH. An ecological momentary assessment of the physical activity and sedentary behaviour patterns of university students. Health Educ J 2010;69(1):116-125. [doi: 10.1177/0017896910363145]
53. Yang C, Conroy DE. Momentary negative affect is lower during mindful movement than while sitting: an experience sampling study. Psychol Sport Exercise 2018;37:109-116. [doi: 10.1016/j.psychsport.2018.05.003]
54. Rebar AL, Dimmock JA, Rhodes RE, Jackson B. A daily diary approach to investigate the effect of ego depletion on intentions and next day behavior. Psychol Sport Exercise 2018;39:38-44. [doi: 10.1016/j.psychsport.2018.07.010]
55. Bejarano CM, Cushing CC, Crick CJ. Does context predict psychological states and activity? An ecological momentary assessment pilot study of adolescents. Psychol Sport Exercise 2019;41:146-152. [doi: 10.1016/j.psychsport.2018.05.008]
56. Bourke M, Hilland TA, Craike M. Contextual influences on the within-person association between physical activity and affect in adolescents: an ecological momentary assessment study. J Behav Med 2021;44(3):296-309. [doi: 10.1007/s10865-020-00197-4] [Medline: 33387176]
57. Biddle SJH, Marshall SJ, Gorely T, Cameron N. Temporal and environmental patterns of sedentary and active behaviors during adolescents' leisure time. Int J Behav Med 2009;16(3):278-286. [doi: 10.1007/s12529-008-9028-y] [Medline: 19238558]
58. Cushing CC, Bejarano CM, Mitchell TB, Noser AE, Crick CJ. Individual differences in negative affectivity and physical activity in adolescents: an ecological momentary assessment study. J Child Fam Stud 2018;27(9):2772-2779. [doi: 10.1007/s10826-018-1128-y]
59. Armstrong-Carter E, Telzer EH. Bidirectional spillover across days between family assistance and physical health experiences during adolescence. J Fam Psychol 2021;35(7):875-885 [FREE Full text] [doi: 10.1037/fam0000836] [Medline: 33705177]
60. Rost EA, Glasgow TE, Calderwood C. Active today, replenished tomorrow? How daily physical activity diminishes next-morning depletion. Appl Psychol Health Well Being 2021;13(1):219-238. [doi: 10.1111/aphw.12229] [Medline: 32956557]
61. King C, Heo J, Lee JW, Hji-Avgoustis S, Lee S. Subjective well-being, activity types, and social context in undergraduate students' daily experiences: an experience sampling study. Coll Stud J 2020;54(1):106-116 [FREE Full text]
62. Arigo D, Pasko K, Mogle JA. Daily relations between social perceptions and physical activity among college women. Psychol Sport Exerc 2020;47:101528 [FREE Full text] [doi: 10.1016/j.psychsport.2019.04.018] [Medline: 32831642]
63. Pannicke B, Reichenberger J, Schultchen D, Pollatos O, Blechert J. Affect improvements and measurement concordance between a subjective and an accelerometric estimate of physical activity. Eur J Health Psychol 2020;27(2):66-75. [doi: 10.1027/2512-8442/a000050]
64. Doyle C, Werner E, Feng T, Lee S, Altemus M, Isler JR, et al. Dev Psychobiol 2015;57(5):607-625 [FREE Full text] [doi: 10.1002/dev.21317] [Medline: 25945698]
65. Ranzenhofer LM, Solhjoo S, Crosby RD, Kim BH, Korn R, Koorathota S, et al. Autonomic indices and loss-of-control eating in adolescents: an ecological momentary assessment study. Psychol Med 2022;53(10):4742-4750. [doi: 10.1017/s0033291722001684]
66. Hamilton JL, Hutchinson E, Evankovich MR, Ladouceur CD, Silk JS. Daily and average associations of physical activity, social media use, and sleep among adolescent girls during the COVID-19 pandemic. J Sleep Res 2023;32(1):e13611. [doi: 10.1111/jsr.13611] [Medline: 35535484]
67. Li YM, Hachenberger J, Lemola S. The role of the context of physical activity for its association with affective well-being: an experience sampling study in young adults. Int J Environ Res Public Health 2022;19(17):10468 [FREE Full text] [doi: 10.3390/ijerph191710468] [Medline: 36078182]

XSL•FO
RenderX

68. Aurora P, Disabato DJ, Coifman KG. Positive affect predicts engagement in healthy behaviors within a day, but not across days. Motiv Emot 2022;46(2):211-225. [doi: 10.1007/s11031-021-09924-z]
69. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000;71 Suppl 2(sup2):1-14. [doi: 10.1080/02701367.2000.11082780] [Medline: 25680007]
70. Kwan MYW, Dutta P, Bray SR, Brown DMY, Cairney J, Dunton GF, et al. Methods and design for the ADAPT study: application of integrated approaches to understanding physical activity during the transition to emerging adulthood. BMC Public Health 2020;20(1):426 [FREE Full text] [doi: 10.1186/s12889-020-08484-0] [Medline: 32234011]
71. Doherty K, Balaskas A, Doherty G. The design of ecological momentary assessment technologies. Interact Comput 2020;32(3):257-278. [doi: 10.1093/iwcomp/iwaa019]
72. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]

Abbreviations

CREMAS: Checklist for Reporting of EMA Studies
EMA: ecological momentary assessment
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ST: sedentary time

Edited by L Buis; submitted 18.09.23; peer-reviewed by J Lai, S Jia; comments to author 06.11.23; revised version received 30.07.24; accepted 16.01.25; published 11.02.25.

Please cite as:
Wang S, Yang CH, Brown D, Cheng A, Kwan MYW
Participant Compliance With Ecological Momentary Assessment in Movement Behavior Research Among Adolescents and Emerging Adults: Systematic Review
JMIR Mhealth Uhealth 2025;13:e52887
URL: https://mhealth.jmir.org/2025/1/e52887
doi:10.2196/52887
PMID:

©Shirlene Wang, Chih-Hsiang Yang, Denver Brown, Alan Cheng, Matthew Y W Kwan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing Social Interaction and Loneliness and Their Association With Frailty Among Older Adults With Subjective Cognitive Decline or Mild Cognitive Impairment: Ecological Momentary Assessment Approach

Bada Kang[1,2], PhD; Dahye Hong[1,3], MPH; Seolah Yoon[1,3], BSN; Chaeeun Kang[4]; Jennifer Ivy Kim[1], PhD

[1]Mo-Im Kim Nursing Research Institute, Yonsei University College of Nursing, Seoul, Republic of Korea

[2]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

[3]College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea

[4]Department of Nursing, Yonsei University College of Nursing, Seoul, Republic of Korea

Corresponding Author:
Jennifer Ivy Kim, PhD
Mo-Im Kim Nursing Research Institute
Yonsei University College of Nursing
50-1, Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 10 3279 2750
Email: redyivy@gmail.com

Abstract

Background: Frail older adults are at greater risk of adverse health-related outcomes such as falls, disability, and mortality. Mild behavioral impairment (MBI), which is characterized by neurobehavioral symptoms in individuals without dementia, is a crucial factor in identifying at-risk groups and implementing early interventions for frail older adults. However, the specific role of social functioning, which encompasses social interaction and loneliness levels, in relation to frailty within this group remains unclear.

Objective: This study investigated the association between frailty status, social interaction frequency, and loneliness levels among older adults with subjective cognitive decline (SCD) or mild cognitive impairment (MCI) while adjusting for MBI symptoms in 2 contexts: the presence and severity of MBI symptoms.

Methods: Older adults with SCD or MCI were recruited from an outpatient clinic specializing in the early diagnosis and care management of dementia at a community health center, as well as from a community service center in Seoul, South Korea. Using an ecological momentary assessment approach, participants reported their daily social interaction frequency and loneliness level via a mobile app, 4 times daily for 2 weeks. Frailty status, the outcome variable, was assessed using the Korean version of the frailty phenotype questionnaire. Additionally, MBI symptoms were assessed using the 34-item MBI-Checklist covering 5 domains. Multinomial logistic regression analyses were performed to investigate the association between frailty status (robust, prefrail, and frail), and the independent variables, adjusting for the presence or severity of MBI symptoms.

Results: Among the 101 participants analyzed, 29.7% (n=30) of participants were classified as prefrail, and 12.8% (n=13) of participants were classified as frail. Higher average daily social interaction scores were consistently associated with lower odds of a frail status compared to a robust status. This was evident in the models adjusted for both the global presence (relative risk ratio [RRR] 0.18, *P*=.02) and global severity (RRR 0.20, *P*=.02) of MBI symptoms.

Conclusions: Frequent social interaction was inversely associated with frail status in older adults with SCD or MCI, even after adjusting for the presence and severity of MBI symptoms. These findings highlight the potential of social functioning as a modifiable factor for addressing frailty among older adults at risk of cognitive and functional decline. Future prospective studies using real-time measurements are needed to refine these findings and further investigate additional risk factors and functional outcomes in this group.

(JMIR Mhealth Uhealth 2025;13:e64853) doi:10.2196/64853

XSL•FO
RenderX

KEYWORDS

geriatric; older; elderly; ageing; association; correlation; cognitive impairment; ecological momentary assessment; frailty; mild behavioral impairment; dementia; Alzheimer; isolation; lonely; social; interaction; self-reported; psychogeriatrics

Introduction

The demographic shift toward an increasingly aging population presents significant public health challenges, with a prominent focus on aging-related adverse health outcomes such as frailty. Frailty is a multidimensional syndrome characterized by a decline in physiological reserve across multiple systems, leading to increased vulnerability to stressors and adverse outcomes. These outcomes include physical limitations, such as impaired mobility, muscle weakness, fatigue, cognitive decline, social isolation, and psychological distress [1-3]. Frailty manifests through physical, cognitive, and psychological components, often resulting in increased risks of falls, disability, hospitalization, and mortality [1,2,4]. Physically, frailty impairs an individual's ability to perform daily activities, while cognitively, it may exacerbate the risks of memory and attention deficits [1-6]. Social and psychological factors also play a role, as frailty is often associated with social isolation and depressive symptoms [3,7]. Frailty and prefrailty incidences were estimated at 43 and 151 new cases per 1000 person-years, respectively, indicating that approximately 1 in 6 community-dwelling older adults may be frail [8]. Frailty is more common in women, older individuals, those with a lower socioeconomic status, and individuals with comorbid chronic diseases and disabilities [1,8-10].

Subjective cognitive decline (SCD) refers to a self-perceived decline in cognitive performance without objective evidence of cognitive impairment [11,12]. Mild cognitive impairment (MCI) represents a state between normal cognition and dementia, characterized by a cognitive decline more pronounced than expected for an individual's age and educational level, but without significant interference in daily activities [13,14]. The co-occurrence of frailty and early-stage cognitive decline has been well documented, with studies showing an increased likelihood of frailty in older adults with SCD or MCI. Two systematic reviews and meta-analyses have offered solid evidence of this relationship [15,16]. Specifically, a cross-sectional study of 2386 individuals with SCD revealed that they had a higher likelihood of being prefrail or frail, compared with individuals with normal cognition, even after controlling for factors such as sociodemographic characteristics, physical functionality, psychosocial aspects, and biochemical influences [12]. The Victoria Longitudinal Study, which followed 632 individuals without dementia, also showed that increased frailty was associated with a more rapid memory decline [5]. Therefore, it is essential to identify individuals with SCD or MCI who are at risk of becoming prefrail or frail, and to implement preventative interventions that focus on modifiable lifestyle and health factors.

Although not all cases of frailty and early cognitive impairment are reversible, evidence supports the potential reversibility of both conditions, including SCD and MCI, through the early detection of at-risk groups and the application of tailored interventions [2,15,17,18]. In addition, factors such as age, sex, and activities of daily living are associated with frailty and cognitive decline [19-22], and mild behavioral impairment (MBI) is also correlated with higher levels of frailty and an increased risk of cognitive decline [2,3]. MBI, which emerges later in life, is a neurobehavioral syndrome characterized by sustained and impactful neuropsychiatric symptoms across 5 domains: decreased drive and motivation, affective dysregulation, impulse dyscontrol, social inappropriateness, and abnormal perception or thought content [23]. MBI is considered a precursor to, or concurrent with, MCI, affecting older adults with either normal cognition or SCD [23]. Significant associations have been observed between certain MBI domains and increased frailty risk in individuals with MCI [3]. Additionally, frailty was associated with both the presence and severity of MBI, especially in men [2]. These findings underscore the importance of assessing MBI to detect risks before the onset of frailty and dementia and warrant the application of interventions for its potential modification.

Psychosocial functioning, which encompasses how individuals engage with society and adapt to environmental challenges, is also an influential factor associated with frailty [24]. Infrequent social interactions, which reflect the quantitative aspects of an individual's limited contact with others [7], and loneliness or perceived social isolation are important factors in this context [25]. Although a definitive theory connecting limited social interaction, loneliness, and frailty has yet to be established, various studies have indicated that reducing social isolation and loneliness in later life can significantly improve overall well-being, underscoring the need for collective consideration of how these factors impact physical health [7]. The English Longitudinal Study of Aging demonstrated that social isolation and loneliness increased the risk of frailty in older adults [26]. Furthermore, given the significant impact of social isolation and loneliness, addressing even moderate levels of these factors is important, as they are associated with an increased risk of worsening frailty [27]. Understanding these psychosocial factors is vital in the broader aging context, as they affect not only physical health but also psychological well-being, thereby potentially influencing older adults' overall quality of life and health care needs [7]. Therefore, the interplay between psychosocial factors and health outcomes underscores the importance of holistic approaches in developing geriatric and preventive strategies.

The ecological momentary assessment (EMA) provides a dynamic and immediate record of personal experiences. It is particularly effective for older adults with cognitive decline, offers a comprehensive view of their everyday social lives, and yields more nuanced insights than conventional global measures [28]. By capturing patterns of daily social interaction or loneliness levels, the EMA addresses restrictions on time or recall bias [29]. However, the relationship between frailty, social interaction, and loneliness levels, particularly in the context of MBI symptomatology, remains underexplored, while few studies have used EMA to examine the real-time social functional status

XSL•FO
RenderX

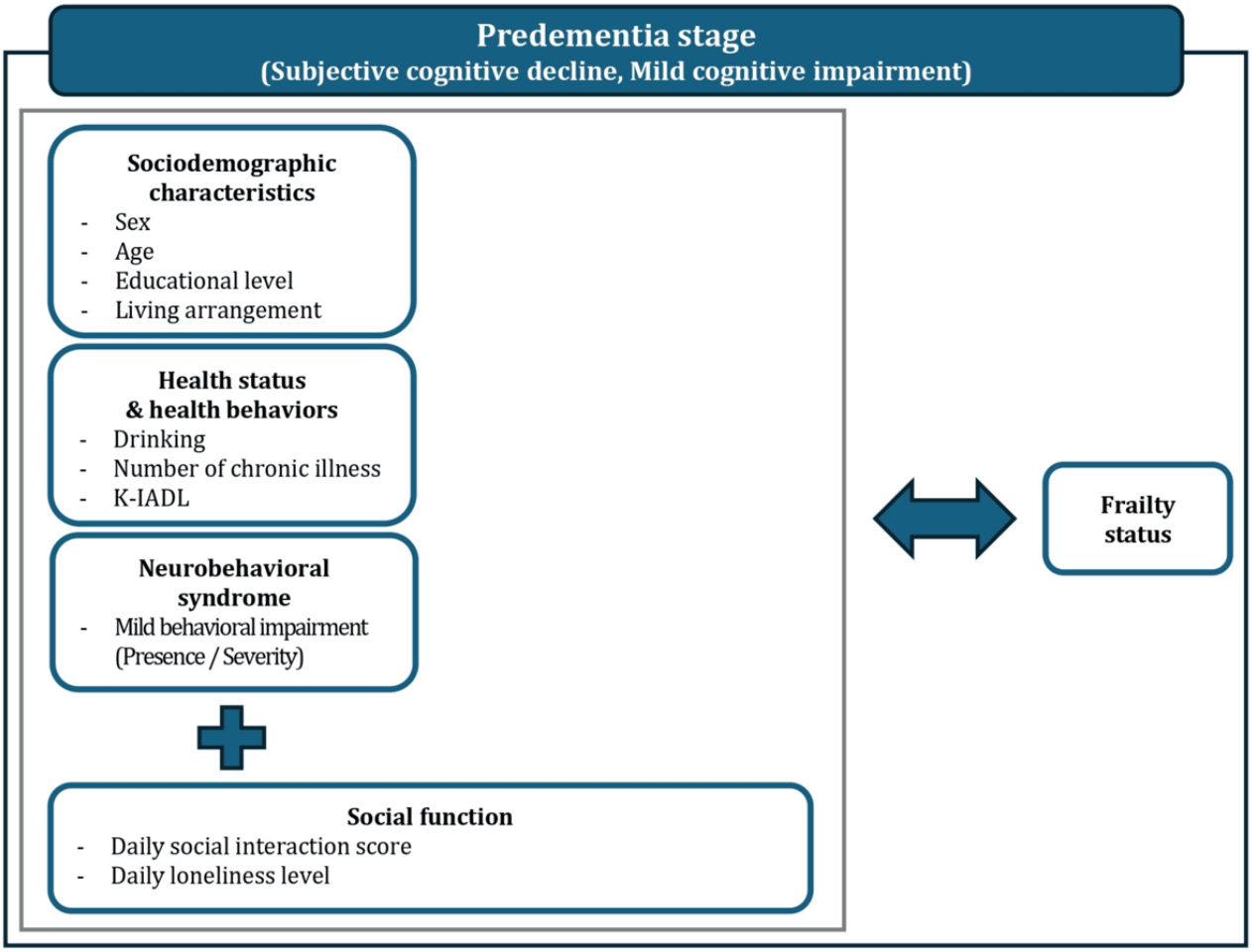
among older adults with cognitive decline. Therefore, this study aimed to investigate the cross-sectional relationships between social interaction frequency, loneliness levels, and frailty status among older adults with SCD or MCI in two contexts: (1) adjusting for the presence of MBI symptoms and (2) adjusting for the severity of MBI symptoms. We hypothesized that, similar to MBI symptoms, both the frequency of social interactions and the level of loneliness would be significantly correlated with frailty. This exploration is anticipated to yield deeper insights into psychosocial factors related to frailty among individuals with SCD or MCI.

Methods

Design

Our study represents the first wave of a 3-year prospective longitudinal study aimed at developing functional prediction models based on behavioral and psychosocial indicators for older adults with SCD or MCI. The study protocol was published elsewhere [30]. The following is a detailed explanation of the variables considered in this study. We constructed the conceptual framework of our study, based on the social-ecological model for older adults [31] and the successful aging model proposed by Rowe and Kahn [32], as shown in Figure 1.

Figure 1. The proposed conceptual framework illustrates the relationship between frailty status, social interaction, and loneliness levels, while adjusting for the presence and severity of mild behavioral impairment symptoms among older adults with subjective cognitive decline or mild cognitive impairment. Based on the social-ecological model for older adults and the successful aging model proposed by Rowe and Kahn, individual determinants include sociodemographic characteristics, health status, and health behaviors. Social function, including daily social interaction frequency and daily loneliness level, is also considered. K-IADL: Korean Instrumental Activities of Daily Living scale.



Participants and Setting

For the larger 3-year longitudinal study, we initially estimated the required sample size using R (generalized linear mixed model, effect size=0.30, k=2, power=0.80, α=0.05, df=30), yielding a minimum of 106 participants. The study protocol has been published elsewhere [30]. The effect size was selected based on prior studies examining the relationship between frailty and social-behavioral factors, which reported moderate effect sizes in similar aging populations [33,34]. Based on these findings, we adopted an effect size of 0.30 as a conservative estimate, to reflect a moderate expected effect while ensuring robustness in our calculation. To account for potential attrition, we adjusted the sample size to 127 participants, anticipating a 20% dropout rate. This adjustment aligns with established guidelines that emphasize careful assessment of attrition rates in aging populations, particularly given the health-related challenges that contribute to higher dropout rates [35,36].

XSL•FO
RenderX

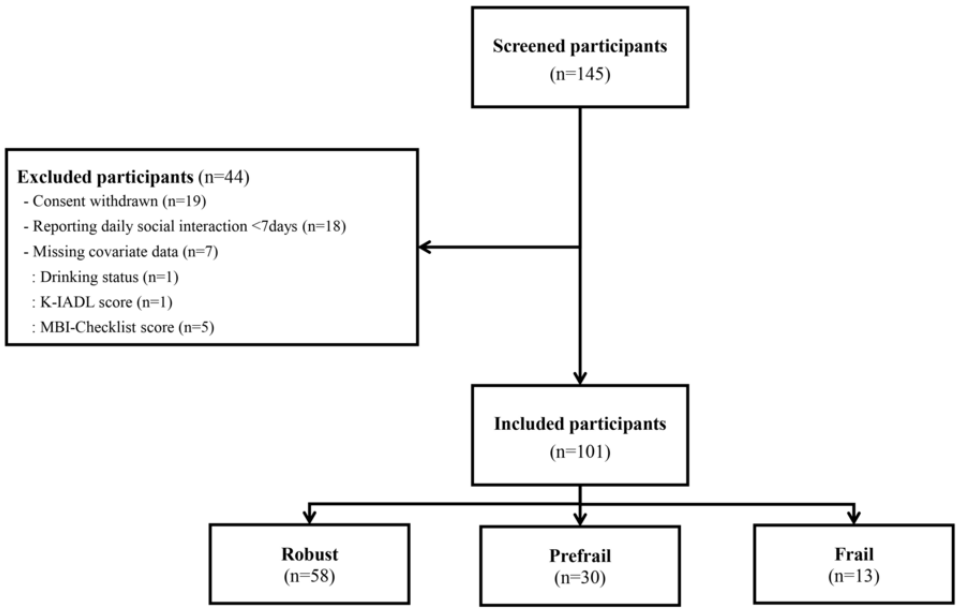
We initially recruited 145 individuals from a dementia relief center and a community service center in Seoul, South Korea (SCD: n=91; MCI: n=54) between October 2022 and November 2023. The dementia relief center houses an outpatient clinic that specializes in early diagnosis and care management of dementia, whereas the community service center serves as a place for older adults to engage in recreational and wellness programs, irrespective of a dementia diagnosis. Specific criteria were applied to recruit study participants from each center. The common inclusion criteria for the study participants were as follows: (1) older than 65 years, (2) ability to use a smartphone, (3) ability to respond to momentary questionnaires via a mobile app, and (4) ability to provide written consent for participation.

Individuals who met the inclusion criteria were selected for the SCD group. Participants were either (1) individuals from the dementia relief center who, according to a nurse's report, had neither been diagnosed with MCI nor dementia; or (2) individuals recruited from the community service center through convenience sampling, who self-reported having no history of MCI or dementia diagnosis and scored above 24 on the Korean Mini-Mental State Examination, second edition. Further, we confirmed SCD in individuals from both the dementia relief center and the community service center by ensuring that they provided an affirmative response to the question, "Do you think your memory has gotten worse, compared with the previous year?"

The MCI group consisted exclusively of participants clinically diagnosed with MCI by medical doctors at the dementia relief center. This group was screened to include only individuals who scored above 18 on the Korean Mini-Mental State Examination, second edition. Participants were excluded if they were (1) illiterate; (2) diagnosed with neurological diseases such as epilepsy, stroke, Parkinson disease, or other types of brain damage; (3) diagnosed with psychiatric diseases such as schizophrenia or bipolar disorder; and (4) undergoing critical illness treatments such as chemotherapy, had severe cardiovascular disease, or had abused substances (including narcotics or alcohol) within the last 3 years.

After excluding individuals who withdrew their consent (n=19), those who reported EMA responses for less than a week or reported no EMA responses per day (n=18), and those with missing covariate data (n=7), 101 individuals were included in the final sample. The participant selection process is illustrated in Figure 2.

Figure 2. Sequential steps involved in the recruitment and selection of study participants. We recruited 145 individuals aged ≥65 years from a dementia relief center or a community service center in Seoul, South Korea. The figure depicts the subdivision of these individuals into 3 groups: 58 with robust, 30 with prefrail, and 13 frail, based on specific inclusion and exclusion criteria. K-IADL: Korean Instrumental Activities of Daily Living scale; MBI: mild behavioral impairment.



Measures

Frailty

Frailty status was the primary outcome variable in this study. Frailty was assessed using the Korean version of the frailty phenotype questionnaire (FPQ), which is specifically designed to screen for the Fried Frailty Phenotype in community-dwelling older adults [37]. While the Fried Frailty Phenotype requires direct measurements of gait speed, grip strength, and physical activity, the FPQ has been validated as an effective proxy for these assessments. Among Korean older adults, the FPQ demonstrated satisfactory diagnostic accuracy for the Fried Frailty Phenotype, with an area under the curve of 0.89, a high sensitivity of 81.7%, and a specificity of 82.5% [37].

The FPQ is particularly well-suited for epidemiological research due to its low resource requirements and ease of administration,

XSL•FO
RenderX

making it a practical screening tool for community-dwelling older adults where detailed clinical assessments may not be feasible. The FPQ comprises 5 questions addressing the key aspects of frailty: fatigue (exhaustion), resistance (weakness), ambulation (slowness), inactivity, and weight loss. Scoring from 0 to 5, where 0 indicates robustness, 1-2 indicates a prefrail status, and ≥ 3 indicates a frail status.

Daily Social Interaction Frequency and Level of Loneliness

Daily social interaction frequency and level of loneliness were the main independent variables. Daily social interaction frequency was measured to reflect quantifiable aspects of social contact, while the level of loneliness was assessed to capture one's perceived view of the consequences of social contact [7]. While the level of loneliness is often measured with a single directional question about one's loneliness status, a solid assessment method for social interaction has yet to be established [7]. Despite the complexity and heterogeneity of its indicators, social isolation can be quantitatively assessed by aggregating various factors into a composite index [7]. Building on a prior study that determined social activity frequency by inquiring about the frequency of participant engagement in socially interactive activities among older adults without clinical signs of dementia rated on a 5-point scale [38], this study quantitatively assessed the daily frequency of social interaction.

Given the cognitive challenges faced by our participants with SCD or MCI, simplifying the data collection process was essential to reduce participant burden and ensure consistent and reliable responses. For this reason, we did not differentiate between types of social interactions (eg, in-person, phone-based), as the study's primary goal was to explore the overall relationship between social engagement and frailty in older adults with cognitive decline.

The EMA method was particularly well-suited for this purpose because it allowed us to collect real-time data on both the frequency of social interactions and participants' perceived loneliness. This approach minimizes recall bias, particularly relevant for individuals with cognitive difficulties. A mobile EMA app was used. With smartphone ownership rates among older Korean adults in 2023 reaching 97.2% for those aged 60-69 years and 72.9% for those aged 70 years and older [39], the feasibility of this method was significantly enhanced.

The EMA method facilitated real-time data collection through a mobile phone app that set off alarms 4 times daily (upon waking and 3 times during mealtimes) over 2 weeks. Each alarm prompted participants to answer 2 questions. The first question assessed the frequency of social interactions since upon waking or since the last mealtime: "How many times have you had a social encounter?" Participants recorded the number of face-to-face meetings, phone calls, or video calls lasting more than 5 minutes using a 5-point Likert scale (0 for "no contact," 1 for "once," 2 for "two times," 3 for "three times," and 4 for "four or more times"). The second question assessed loneliness: "At this moment, how lonely do you feel?" using a similar 5-point Likert scale (0="very lonely" to 4="not lonely at all"). Both questions were asked 4 times a day at set intervals (upon waking and after breakfast, lunch, and dinner). Figure 3 shows the captured screens from the mobile app. The research staff provided follow-up contacts for troubleshooting as needed. Participants received gifts worth US $50 after completing the 2-week data collection period.

XSL•FO
RenderX

Figure 3. The mobile app screens captured 2 questions that assessed daily social interaction frequency and loneliness levels. Daily social interaction frequency was measured by asking, “How many times did you interact with others, including face-to-face meetings, phone calls, or video calls lasting more than five minutes?” Responses were categorized into 5 options: no contact, once, twice, 3 times, and 4 or more times. Loneliness levels were measured by asking, “At this moment, how lonely do you feel?” Responses were also categorized into 5 options: very lonely, lonely, neutral, not lonely, and not lonely at all. Both questions were administered 4 times daily at specific intervals (upon waking and after breakfast, lunch, and dinner).

We implemented 4 strategies to ensure accurate EMA data collection. First, we provided comprehensive training sessions, including repeated practice, to ensure participants could use the mobile app independently. Second, we provided detailed instructional manuals for completing the response process (Multimedia Appendices 1-3). Although the EMA questions are simple, we addressed potential challenges given that the participants were individuals with SCD or MCI. Third, we established a helpdesk for assistance with any app-related issues during the follow-up period. Last, we offered direct support as needed, with written contact information provided to ensure participants could reach researchers via phone or in person. These measures aimed to enhance participants’ understanding and proficiency, contributing to more accurate data collection. Approximately 15 participants requested assistance during the study, mainly due to issues with mobile EMA alarms not triggering due to internet connectivity issues or difficulties with smartphone lock functions. Researchers provided phone guidance and, when necessary, conducted in-person visits to resolve the issues and offered additional training.

XSL•FO
RenderX

The quantity of social interaction and the level of loneliness were indicated as daily mean scores based on previous studies [28,40,41]. Data processing involved calculating the maximum daily social interaction frequency as the final daily score. For instance, if a participant reported varying frequencies of social encounters throughout day 1, the highest frequency reported that day was recorded as the final score. The average daily social interaction score was then determined by summing the scores and dividing them by the total number of valid responses. Similarly, the average daily loneliness score was calculated by adding the total loneliness scores recorded each day (with at least one response) and dividing them by the total number of valid response days.

Presence and Severity of MBI Symptoms

The presence and severity of MBI symptoms were included as key covariates. The Mild Behavioral Impairment Checklist (MBI-C) was used to assess the onset of sustained and impactful neurocognitive symptoms in the predementia stage [23]. The MBI-C systematically evaluates symptoms across the five domains of MBI, persisting for at least 6 months: (1) decreased motivation, (2) affective dysregulation, (3) impulse dyscontrol, (4) social inappropriateness, and (5) abnormal perception or thought content. Each symptom was initially assessed dichotomously (yes/no), followed by a severity rating of mild (1), moderate (2), or severe (3) [23].

In this study, the Korean version of the MBI-C was used to evaluate the presence and severity of MBI symptoms. The MBI-C, a brief screening tool previously validated, was implemented to detect MBI in line with established criteria [3]. The MBI-C has been translated into various languages including Korean. Psychometric evaluation of the Korean version demonstrated significant correlations with the Neuropsychiatric Inventory in individuals with amnestic MCI (r=0.25, P<.01) and non-amnestic MCI (r=0.36, P<.05) [42,43]. In this study, Cronbach α for the MBI-C was 0.89. For presence assessment, we used a dichotomous method for both domain-specific and global MBI symptoms, with a cut-off score > 0 indicating the presence of MBI symptoms [2]. Severity was assessed as either (1) domain-specific, where the sum of scores from each individual domain was calculated; or (2) global, involving the cumulative sum of scores across all domains [2].

Covariates

Frailty encompasses numerous biological, physiological, and environmental changes that typically occur during the aging process [21]. Therefore, we considered sociodemographic and health-related characteristics, health behaviors, and functional limitations as covariates. The demographic characteristics of participants included sex (male/female), age group (<70, 70-74, 75-79, and ≥80 years), educational level (<9/≥9 years), and living arrangement (living alone/residing with others).

For health-related characteristics, we considered the total number of chronic illnesses diagnosed, including hypertension, diabetes, cardiovascular diseases, cerebrovascular diseases, gastrointestinal diseases, chronic respiratory diseases, chronic kidney diseases, and musculoskeletal conditions. Participants' cognitive status was dichotomously categorized as SCD or MCI based on their responses, which were recorded to evaluate the inclusion criteria. Health behavior aspects included drinking status (active drinker/nondrinker). Owing to the limited sample size of former or current smokers (n=2), smoking status was excluded from the statistical analysis.

As demonstrated in previous studies, dependence on managing higher levels of functional performance is associated with frailty [21,22]. Therefore, functional limitations were assessed using the Korean Instrumental Activities of Daily Living (K-IADL) scale, which measures the ability to perform complex daily tasks such as cooking, grooming, housekeeping, shopping, financial management, transportation, laundry, short-distance walking, making phone calls, and managing medications [44]. Each item on the K-IADL scale is rated from 0 to 4, with higher scores indicating greater dependency. The K-IADL's validation, conducted with older adults aged older than 65 years, demonstrated strong internal consistency (Cronbach α=0.94), high interrater reliability (κ value=0.81-0.95), and test-retest reliability correlation (≥ 0.70) [44]. In this study, high K-IADL scores indicated diminished functional capabilities and were analyzed as continuous variables.

Statistical Analysis

Descriptive statistics were used to outline the general characteristics of the participants and compare differences between the robust, prefrail, and frail groups. Categorical variables are presented as counts and percentages, while continuous variables are summarized as means with SDs.

To assess the normality of continuous variables, we performed the Shapiro-Wilk test. If the data were normally distributed and variance was equal, analysis of variance was applied to compare differences between groups. When normality was not met or variance was unequal, the Kruskal-Wallis test was used. A chi-square test was used to assess differences in categorical variables between groups, with the Fisher exact test applied when the expected cell count was less than 5.

We conducted multinomial logistic regression analyses to examine the association between frailty status, social interaction frequency, and loneliness levels among older adults with MCI or SCD. Frailty status was the primary outcome variable and was categorized into 3 groups: robust, prefrail, and frail. Social interaction frequency and daily loneliness levels were the independent variables, assessed using EMA.

To account for potential confounding effects, MBI symptoms were included as a key covariate. Additionally, other confounders, including sex, age group, educational level, living arrangement, number of chronic illnesses, cognitive status, health behaviors, and functional limitations, were selected based on their established relevance to frailty and social engagement in previous studies [2,3,19-22].

Separate multinomial logistic regression models were performed to adjust for MBI symptom presence and severity. All independent variables were simultaneously entered into the model to ensure consistency and comparability across analyses. Multicollinearity was assessed using the variance inflation factor to confirm that the included variables were not highly correlated. Model fit was evaluated using the generalized

XSL•FO
RenderX

Hosmer-Lemeshow test and model specification was checked for validity. Statistical significance was set at $P<.05$ and all analyses were performed using Stata software (version 16; StataCorp).

Ethical Considerations

This study was approved by the Institutional Review Board of Severance Hospital, Yonsei University Health System (number 4-2022-0637). In accordance with ethical guidelines, all participants provided their written informed consent prior to their inclusion in the study. To ensure ethical integrity, individuals with severe cognitive decline who were assessed as incapable of making informed decisions regarding participation were excluded from the study.

Results

Participants' General Characteristics

Table 1 presents the general characteristics of participants. A total of 101 participants were included in the analysis; 58 (57.4%) were categorized as robust, 30 (29.7%) as prefrail, and 13 (12.9%) as frail. Significant differences were observed between the groups in age and global MBI symptom severity. The mean age and SDs for each group were robust (74.9, SD 5.1 years), prefrail (75.7, SD 6.5 years), and frail (82.0, SD 4.9 years), with corresponding age ranges of 65-87, 65-89, and 73-89 years, respectively. For global MBI severity, the median score for all participants was 3 (IQR 1-9.5). The robust group had the lowest median severity score (2, IQR 1-4.5), while the frail group exhibited the highest severity (9, IQR 3-23.5).

XSL•FO
RenderX

Table 1. Baseline characteristics of participants by frailty status.

Variables	Total (N=101)	Robust (n=58)	Prefrail (n=30)	Frail (n=13)	*P* value
Sex, n (%)					.47
Men	42 (41.6)	27 (46.5)	10 (33.3)	5 (38.5)	
Women	59 (58.4)	31 (53.5)	20 (66.7)	8 (61.5)	
Age (years)					
Range	65-89	65-87	65-89	73-89	—[a]
Mean (SD)	76.1 (5.9)	74.9 (5.1)	75.7 (6.5)	82.0 (4.9)	<.001
Educational level , n (%)					.06
<9 years	30 (29.7)	12 (20.7)	12 (40.0)	6 (46.2)	
≥9 years	71 (70.3)	46 (79.3)	18 (60.0)	7 (53.8)	
Living arrangement[b], n (%)					.22
Living alone	19 (18.8)	8 (13.8)	9 (30.0)	2 (15.4)	
Residing with others	82 (81.2)	50 (86.2)	21 (70.0)	11 (84.6)	
Cognitive status, n (%)					.06
SCD[c]	67 (66.3)	44 (75.9)	16 (53.3)	7 (53.9)	
MCI[d]	34 (33.7)	14 (24.1)	14 (46.7)	6 (46.1)	
Drinking status[b], n (%)					.46
Active drinker	25 (24.8)	17 (29.3)	5 (16.7)	3 (23.1)	
Nondrinker	76 (75.2)	41 (70.7)	25 (83.3)	10 (76.9)	
Number of chronic illnesses, mean (SD)	2.01 (1.11)	1.96 (1.07)	1.93 (1.14)	2.46 (1.19)	.31
K-IADL[e,f]					
Range	1-17	1-16	8-17	9-16	—
Median (IQR)	10 (10-10)	10 (10-10)	10 (10-10)	10 (10-12.5)	—
Mean (SD)	10.4 (1.9)	10.4 (2.1)	10.4 (1.7)	11.0 (1.9)	.64
Global MBI[g] symptom presence, n (%)	81 (80.2)	45 (77.5)	24 (80.0)	12 (92.3)	.64
Global MBI[g] symptom severity[f]					
Range	0-41	0-25	0-40	0-41	—
Median (IQR)	3 (1-9.5)	2 (1-4.5)	8.5 (1-12.3)	9 (3-23.5)	—
Mean (SD)	6.9 (9.0)	4.0 (5.1)	9.2 (10.1)	14.5 (13.8)	.002

[a]Not applicable.

[b]Fisher exact test was applied.

[c]SCD: subjective cognitive decline.

[d]MCI: mild cognitive impairment

[e]K-IADL: Korean Instrumental Activities of Daily Living scale.

[f]Kruskal-Wallis test was applied.

[g]MBI: mild behavioral impairment.

Table 2 displays a descriptive summary of social interaction frequency and loneliness levels for all participants, stratified by frailty status. We summarized the average maximum daily social interaction score by calculating the average of each participant's daily maximum scores across the study period, resulting in a single, average score per participant. No significant differences were observed in the mean average maximum daily social interaction scores among the groups: robust (3.0, SD 0.8), prefrail (2.9, SD 0.9), and frail (2.4, SD 0.9). Similarly, no significant differences were found in the mean average maximum daily loneliness levels, with scores of 3.4 (SD 0.5) for the robust group, 3.3 (SD 0.6) for the prefrail group, and 3.1 (SD 0.9) for the frail group.

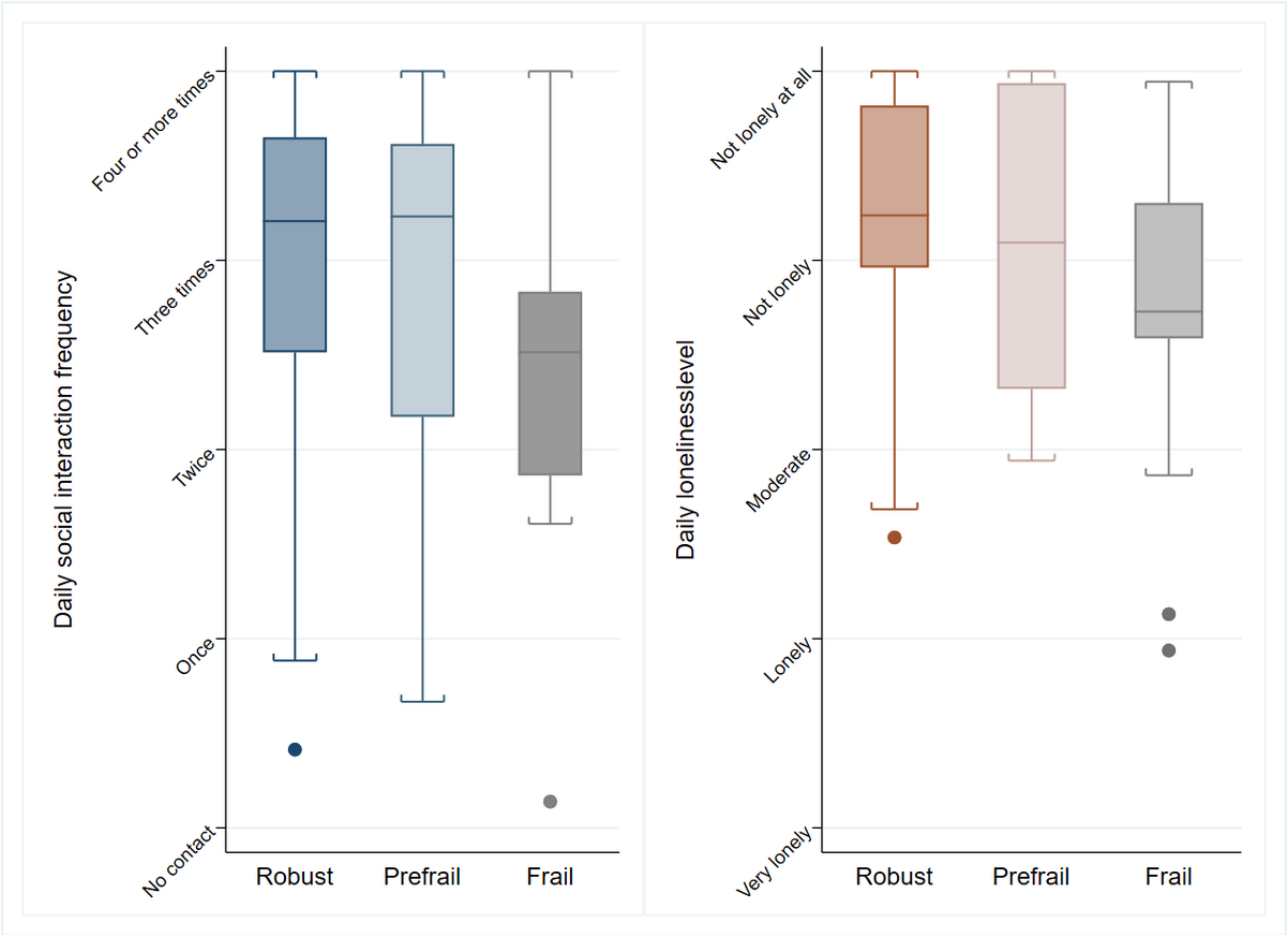
XSL•FO
RenderX

Table 2. Descriptive statistics of daily social interaction frequency and loneliness levels by frailty status.

	Total (N=101)	Robust (n=58)	Prefrail (n=30)	Frail (n=13)	*P* value
Social interaction frequency per day					
Range	0-4	0-4	0-4	0-4	
Mean (SD)	2.9 (0.8)	3.0 (0.8)	2.9 (0.9)	2.4 (0.9)	.08
Loneliness level per day					
Range	0-4	1-4	0-4	0-4	
Mean (SD)	3.4 (0.6)	3.6 (0.5)	3.3 (0.6)	3.1 (0.9)	.06

Figure 4 illustrates the variations of the average daily social interaction scores and average daily loneliness levels categorized by frailty status. The robust group is characterized by the highest median values in daily social interactions and the lowest loneliness levels compared with prefrail and frail groups. In contrast, the frail group demonstrates the lowest median social interaction frequency and the highest loneliness levels. Furthermore, the density distribution of mean maximum daily social interaction scores and mean maximum daily loneliness level by frailty status is illustrated in Multimedia Appendices 4 and 5. While the overall distribution is right skewed across all participants, the frail group exhibits a relatively higher density in the lower frequency of social interactions compared with the robust or prefrail group (Multimedia Appendix 4). In addition, the frail group shows a relatively lower density of participants experiencing lower loneliness levels compared with robust and prefrail groups, even though the overall distribution is right-skewed (Multimedia Appendix 5).

Figure 4. Distribution of average daily social interaction score and average daily loneliness level by frailty status. This figure displays the median levels and range variability of both average daily social interaction score and average daily loneliness levels among participants, categorized into robust, prefrail, and frail groups.



XSL•FO
RenderX

Primary Outcomes

Association Between Daily Social Interaction Frequency, Daily Loneliness Level, and Frailty Adjusting for the Presence of Global MBI Symptoms

Table 3 illustrates the relationship between frailty, daily social interaction frequency, and daily loneliness level after adjusting for the presence of global MBI symptoms. After adjusting for the presence of global MBI symptoms, neither the average maximum daily social interaction score (relative risk ratio [RRR] 0.74, *P*=.42) nor the average maximum daily loneliness level (RRR 0.89, *P*=.76) were significantly associated with prefrail status. Similarly, no significant relationship was observed between the average maximum daily loneliness level and frail status (RRR 0.27, *P*=.10). However, participants with higher average maximum daily social interaction scores were significantly less likely to be frail (RRR 0.18, *P*=.02). Sociodemographic and health-related characteristics also showed significant associations with frailty. Participants with SCD were less likely to be prefrail compared with those with MCI (RRR 0.31, *P*=.03), while older age was associated with a higher likelihood of being frail (RRR 1.56, *P*=.001). Figure 5 displays the average predictive marginal effects of sociodemographic characteristics, health status, global MBI symptom presence, and social functioning on the probability of being classified as robust, prefrail, or frail.

Table 3. Multinomial logistic regression model for the association between frailty, global MBI-symptom presence, daily social interaction frequency, and level of loneliness (N=101) (Wald χ222=52.04).

Variables	Prefrail (base outcome: robust)		Frail (base outcome: robust)	
	RRR[a] (95% CI)	*P* value	RRR (95% CI)	*P* value
Women (Ref: men)	2.18 (0.57-8.26)	.25	3.54 (0.42-29.72)	.24
Age (years)	1.01 (0.92-1.12)	.79	1.56 (1.20-2.01)	.001
Educational level ≥ 9 years (Ref: <9 years)	0.46 (0.16-1.33)	.15	0.43 (0.06-2.89)	.38
Living alone (Ref: Residing with others)	2.44 (0.68-8.68)	.17	0.11 (0.003-3.27)	.20
SCD[b] (Ref: MCI[c])	0.31 (0.11-0.87)	.03	0.35 (0.06-2.17)	.26
Active drinker (Ref: Nondrinker)	0.62 (0.15-2.51)	.50	9.10 (0.61-135.11)	.11
Number of chronic illnesses	0.78 (0.48-1.26)	.31	2.04 (0.73-5.66)	.17
K-IADL[d]	1.13 (0.86-1.49)	.37	1.38 (0.85-2.25)	.18
Global MBI[e]-symptom presence	0.89 (0.25-3.15)	.85	0.36 (0.01-8.91)	.53
Average maximum daily social interaction score	0.74 (0.36-1.53)	.42	0.18 (0.05-0.72)	.02
Average maximum daily loneliness level	0.89 (0.41-1.92)	.76	0.27 (0.05-1.31)	.10

[a]RRR: relative risk ratio.

[b]SCD: subjective cognitive decline.

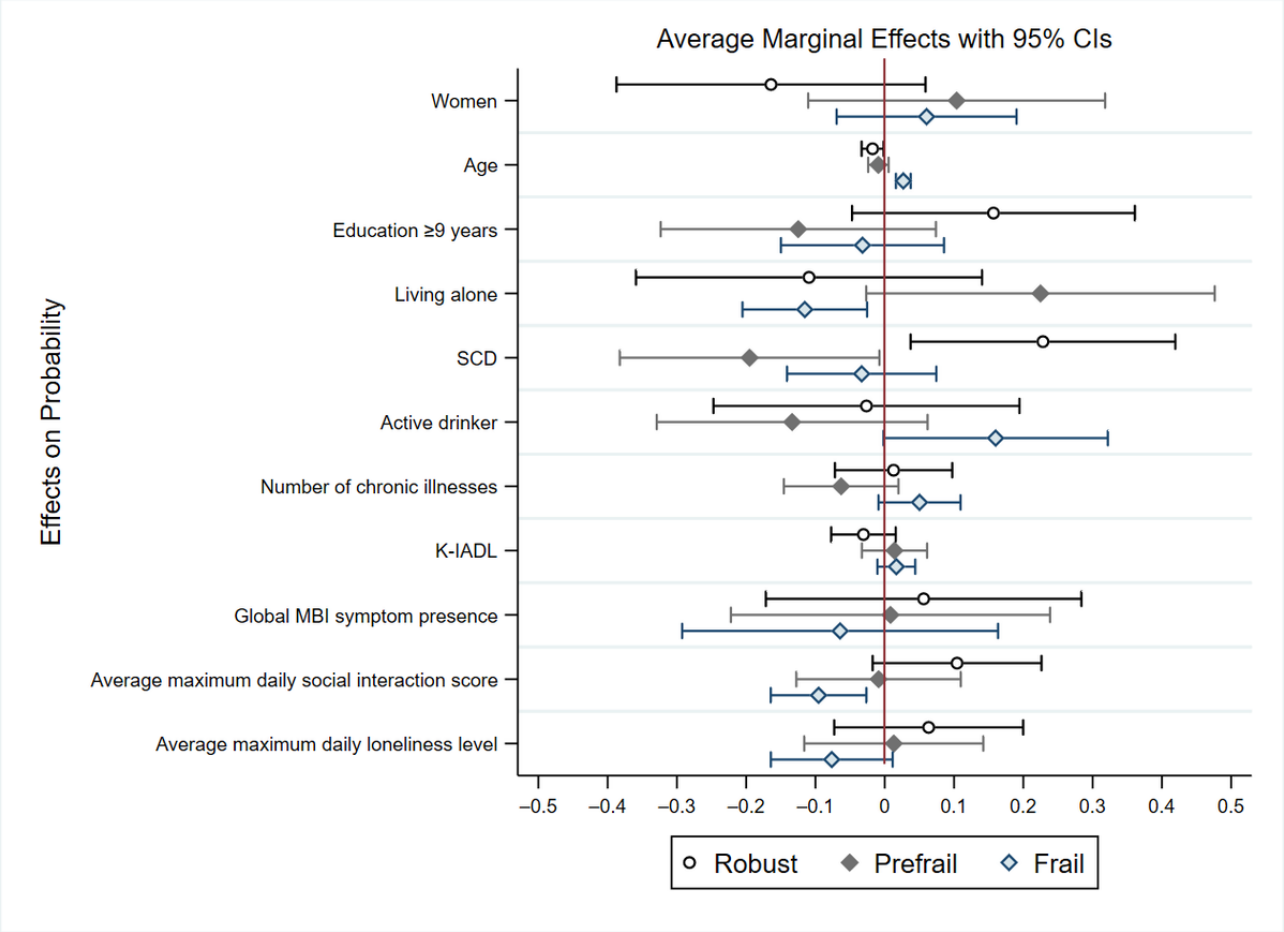
[c]MCI: mild cognitive impairment.

[d]K-IADL: Korean Instrumental Activities of Daily Living scale.

[e]MBI: mild behavioral impairment.

XSL•FO
RenderX

Figure 5. This figure shows the average predictive marginal effects of sociodemographic characteristics, health status, global MBI symptom presence, and social functioning on the probability of being classified as robust, prefrail, or frail. The predictive margins were calculated using multinomial logistic regression, with error bars representing 95% CIs for each effect. K-IADL: Korean Instrumental Activities of Daily Living scale; MBI: mild behavioral impairment; SCD: subjective cognitive decline.



Association Between Daily Social Interaction Frequency, Daily Loneliness Level, and Frailty Adjusting for the Severity of Global MBI Symptoms

Table 4 presents the associations between frailty, daily social interaction frequency, and daily loneliness levels after adjusting for the severity of global MBI symptoms. When adjusting for the severity of global MBI symptoms, no significant associations were observed between prefrail status and either the average maximum daily social interaction score (RRR 0.76, *P*=.48) or the average maximum daily loneliness level (RRR 1.19, *P*=.69). Likewise, no significant relationship was identified between frail status and average maximum daily loneliness level (RRR 0.27, *P*=.14). However, a significant inverse relationship was found between the average maximum daily social interaction scores and frail status (RRR 0.20, *P*=.02). Additional sociodemographic and health-related factors were associated with prefrail status. Participants with SCD were less likely to be prefrail compared with those with MCI (RRR 0.33, *P*=.04). Moreover, participants with more severe MBI symptoms were more likely to be prefrail (RRR 1.09, *P*=.04), and older age remained significantly associated with being frail (RRR 1.53, *P*=.002). Figure 6 presents the average predictive marginal effects of sociodemographic characteristics, health status, global MBI symptom severity, and social functioning on the probability of being classified as robust, prefrail, or frail.

XSL•FO
RenderX

Table 4. Regression models for the association between frailty, global MBIa-symptom severity, daily social interaction frequency, and level of loneliness (N=101; Wald χ222=56.59).

Variables	Prefrail (Base outcome: Robust)		Frail (Base outcome: Robust)	
	RRR[b] (95% CI)	*P* value	RRR (95% CI)	*P* value
Women (Ref: men)	2.22 (0.56-8.71)	.25	2.93 (0.38-22.34)	.30
Age (years)	0.97 (0.88-1.10)	.63	1.53 (1.16-2.02)	.002
Educational level ≥9 years (Ref: <9 years)	0.52 (0.17-1.59)	.26	0.54 (0.08-3.65)	.52
Living alone (Ref: Residing with others)	2.58 (0.69-9.69)	.16	0.08 (0.002-2.57)	.16
SCD[c] (Ref: MCI[d])	0.33 (0.11-0.97)	.04	0.44 (0.07-2.57)	.36
Active drinker (Ref: Nondrinker)	0.65 (0.15-2.68)	.55	10.29 (0.61-174.13)	.11
Number of chronic illnesses	0.72 (0.44-1.18)	.20	2.14 (0.74-6.12)	.15
K-IADL[e]	1.10 (0.84-1.44)	.49	1.31 (0.81-2.13)	.27
Global MBI[f]-symptom severity	1.09 (1.01-1.18)	.04	1.06 (0.95-1.17)	.32
Average maximum daily social interaction score	0.76 (0.36-1.61)	.48	0.20 (0.05-0.78)	.02
Average maximum daily loneliness level	1.19 (0.51-2.78)	.69	0.27 (0.05-1.56)	.14

[a]MBI: mild behavioral impairment.

[b]RRR: relative risk ratio.

[c]SCD: subjective cognitive decline.

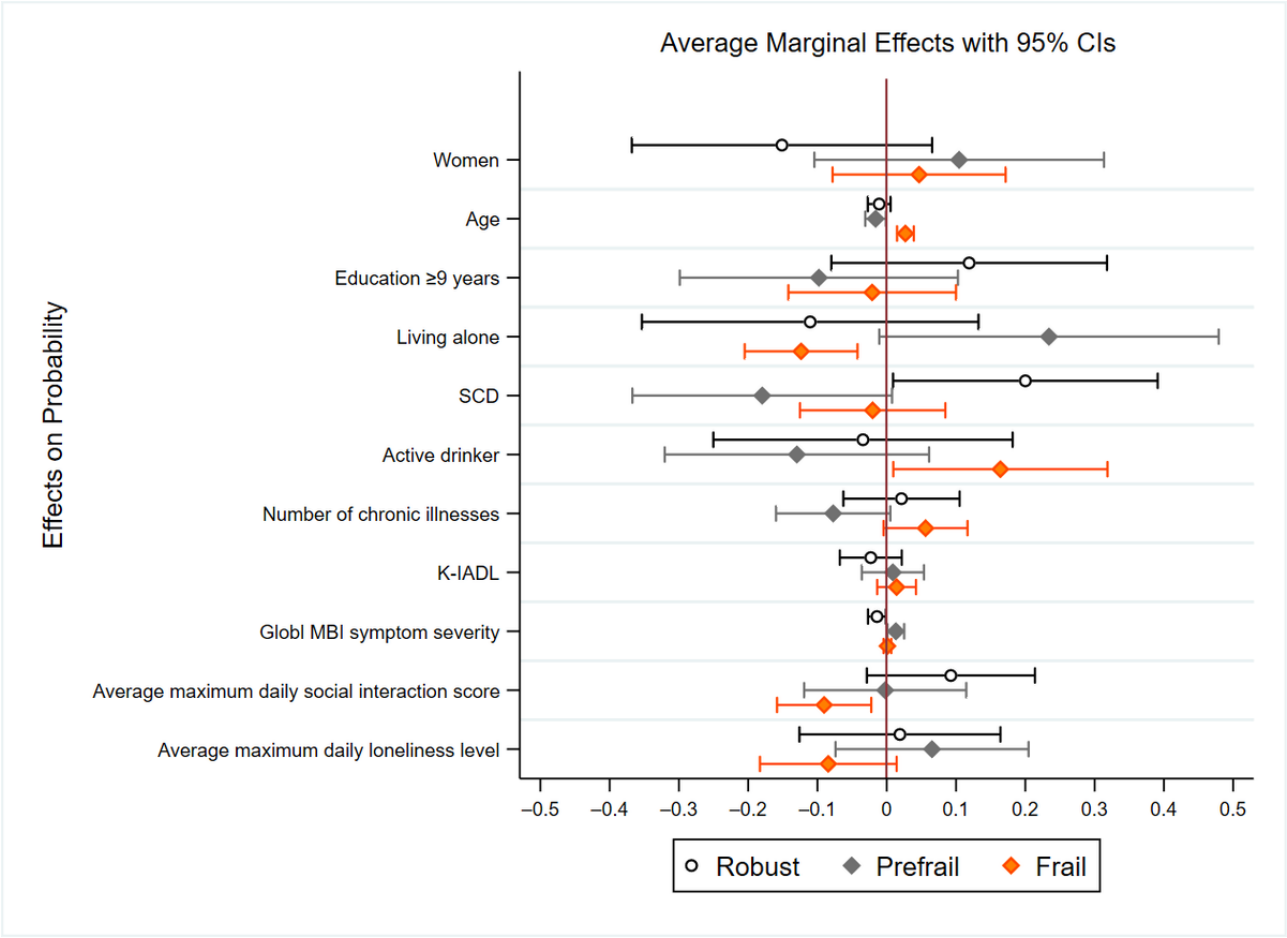
[d]MCI: mild cognitive impairment.

[e]K-IADL: Korean Instrumental Activities of Daily Living scale.

[f]MBI: mild behavioral impairment.

XSL•FO
RenderX

Figure 6. This figure shows the average predictive marginal effects of sociodemographic characteristics, health status, global MBI symptom severity, and social functioning on the probability of being classified as robust, prefrail, or frail. The predictive margins were calculated using multinomial logistic regression, with error bars representing 95% CIs for each effect. K-IADL: Korean Instrumental Activities of Daily Living scale; MBI: mild behavioral impairment; SCD: subjective cognitive decline.



Discussion

Principal Findings

This study demonstrated a significant association between frailty and the frequency of daily social interaction in a cohort of 101 individuals at risk of dementia in South Korea. Specifically, higher average daily social interaction scores were significantly associated with lower relative risk ratios of being frail, even after adjusting for the presence and severity of global MBI symptoms. These findings underscore the potential role of daily social interaction frequency as a modifiable behavioral factor in individuals at the early stages of dementia and frailty risk.

Our study found that frail older adults exhibited more severe MBI symptoms than robust or prefrail groups. This observation aligns with prior research suggesting a potential link between frailty and the severity of MBI symptoms [2]. Furthermore, our findings indicate that frail older adults had lower average maximum daily social interaction frequency and higher loneliness levels compared with robust or prefrail groups. Unlike previous studies that assessed social interaction or loneliness at a single time point [7], our study used a 2-week tracking period. This approach also builds on previous research exploring how frailty in older adults, characterized by decreased mobility and difficulty with daily activities, may contribute to reduced social engagement, increased loneliness, and further functional decline [7].

Our findings also align with existing evidence showing that both the presence and severity of global MBI symptoms are associated with frailty. This underscores the importance of assessing both cognitive and behavioral impairments when identifying individuals at risk of dementia [2]. However, given the relatively small sample size of frail individuals in this study, these findings should be interpreted with caution. Nonetheless, our results support prior research linking neuropsychiatric symptoms, such as psychosis, not only to an elevated risk of dementia but also frailty, which is often associated with accelerated cognitive decline [45]. To strengthen these observations, future research should consider analyzing larger cohorts with detailed assessments of domain-specific MBI symptoms and frailty. Such studies could yield more robust evidence, further elucidating the relationship between specific MBI domains and frailty and reinforcing their potential as targets for early intervention in individuals at risk of dementia and functional decline.

Our findings are consistent with those of a previous study that used data from the Korean Frailty and Aging Cohort Study, which found that less frequent contact with acquaintances was associated with a higher prevalence of frailty [46]. Similar patterns have been observed worldwide. For instance, a 1-year

XSL•FO
RenderX

prospective cohort study in Japan demonstrated that greater social isolation, measured using the Lubben Social Network Scale, increased the risk of developing prefrail status [47]. Among Italian community-dwelling older adults, a significant relationship was found between frailty and higher levels of social isolation, measured using the Friendship Scale, and frailty [48]. Although the precise mechanisms through which social connections influence frailty are not yet fully understood, growing epidemiological evidence suggests that social interactions can affect health outcomes through psychosocial, behavioral, and biological pathways. A lack of social interactions, for instance, can amplify the sympathetic nervous system and the hypothalamus-pituitary-adrenal axis, thus leading to a chronic mild inflammatory state, which is characteristic of frailty syndrome [49-51]. Limited social interactions can also increase stress-related hormones such as glucocorticoids, impacting immune functions and potentially affecting the biological aspects of frailty [52]. Furthermore, frequent social interactions can motivate frail older adults to engage in cognitively and physically stimulating activities, providing daily support and health management resources [52,53].

Overall, our findings underscore the importance of social interaction correlates to frailty among older adults. This issue is particularly relevant in East Asian societies, where traditional family structures and intergenerational support are highly valued [54]. However, the rapid demographic shift toward an aging population and declining family size in developed East Asian countries, such as South Korea, Japan, and China, poses challenges to maintaining adequate social contact for older adults [54,55]. This phenomenon may increase the risk of frailty in older adults who lack sufficient social engagement due to changing family dynamics [56]. Furthermore, in developed countries, where health care systems are more established, promoting social functioning could be a cost-effective, nonpharmacological intervention to reduce frailty and its associated health care burden [57,58]. These findings suggest that public health policies aimed at enhancing social interactions, particularly for those at risk of cognitive decline, could reduce frailty and promote healthy aging in both East Asian and developed countries more broadly.

Our study did not establish a significant correlation between daily loneliness and frailty, in contrast to previous research. This discrepancy may be attributed to 2 primary reasons. First, our method of measuring loneliness was based on a single direct question covering a 2-week period. While we aimed to reflect on the participants' subjective perceptions of their social interactions, a more in-depth, long-term study may be necessary to capture the full complexity of loneliness. A systematic review and meta-analysis of the relationship between loneliness and physical frailty in community-dwelling older adults indicated that baseline loneliness could predict worsening frailty, although longitudinal evidence is limited [59]. Therefore, future studies should focus on the cumulative effects of comprehensive potential factors prevalent in older adults, such as decreased physical activity, poor sleep quality, and malnutrition [59-61]. Additionally, these studies should examine these factors within the context of social vulnerability and frailty, while controlling for the effects of social networks [61].

Second, unlike the frequency of social interactions, loneliness is a subjective feeling, and perceptions of loneliness can vary significantly among individuals [7]. Typically, loneliness is seen as a negative response to the gap between actual and desired relationships [62]. However, the threshold for feeling lonely often depends on sociocultural context or personal standards [63,64]. Therefore, using more nuanced, context-specific questionnaires could lead to a deeper understanding of loneliness [7]. Nonetheless, given that our study participants were older adults at risk for dementia, using numerous detailed questionnaires might not be the most feasible approach. An alternative method could involve analyzing patterns among participants who provide similar responses to brief, directional questions, and then prospectively studying their perceptions of loneliness in response to specific events in a real-time setting. This approach could yield valuable insights into subjective experiences of loneliness in this population.

Comparison With Prior Work

A key strength of our study lies in its detailed exploration of the relationship between social interaction and frailty status among older adults with SCD or MCI. By categorizing frailty into robust, prefrail, and frail groups, our findings build upon and extend previous research, indicating that older adults with higher levels of daily social interaction are less likely to exhibit frailty, even after accounting for cognitive and behavioral symptoms. While these findings align with prior research suggesting the protective effects of social interaction on health, our study adds nuance by examining these relationships specifically within older adults at risk of cognitive decline.

Our results support existing evidence that frail older adults exhibit lower social interaction frequency and higher loneliness levels compared with their robust and prefrail counterparts. These findings underscore the importance of social interaction as a critical behavioral dimension for health outcomes in older adults at risk of both cognitive and functional decline. While causation cannot be established, our findings suggest that fostering social connections may mitigate frailty risks and promote better health outcomes.

What distinguishes our study is its use of real-time data collection methods, which allowed us to capture day-to-day variations in social functioning with greater accuracy [40]. This approach offers a more granular understanding of how social interaction relates to frailty and cognitive status, building on prior retrospective studies. By applying real-time assessment within the context of frailty and cognitive decline, this study strengthens the evidence base for social interaction as a modifiable factor and highlights the need for future research to explore causative pathways.

Although the immediate practical application of these findings in clinical settings may be limited, this study contributes to the growing understanding of the relationship between social function, frailty, and cognitive decline. It highlights the potential for improving social engagement interventions that may reduce physical challenges targeted to older adults with SCD or MCI. Future research should focus on identifying the specific types, patterns, and contexts of social interaction that are most effective in reducing frailty and promoting health. Longitudinal and

XSL•FO
RenderX

interventional studies will be crucial in establishing causation and designing tailored interventions to enhance social functioning in these at-risk populations.

Limitations

This study had several limitations. First, the sample size of 101 participants, particularly the small sample size of the frail group, may limit the generalizability of the findings. A larger and more diverse sample could yield more robust results and allow for detailed analyses across different subgroups. Second, the cross-sectional design of the study enabled the identification of associations but did not establish causation. Prospective longitudinal studies are necessary to determine the directionality of the relationships among frailty, MBI symptoms, and social functioning. Additionally, Alzheimer disease pathology, hormonal changes, and chronic inflammation are potential mechanistic links between frailty and cognitive decline. While this study focuses on social interaction, physiological dimensions such as physical activity, metabolic health, and nutritional status were not assessed to minimize the participants' measurement burden during the 2-week survey period. The primary aim of our study was to demonstrate how capturing temporal patterns in social functioning could inform daily life modifications to support health. Future research should integrate complementary physiological measures to enhance understanding and inform more comprehensive intervention strategies in longitudinal or experimental studies to better understand the directionality of these relationships. For example, studies using matching designs over extended timespans would be essential. Third, this study relied on self-reported data for loneliness levels and the tools used to measure daily social interaction frequency and loneliness levels, which may be susceptible to social desirability bias, which could influence how participants report their feelings and circumstances. Future analyses could incorporate digital devices, such as actigraphy, to log social interaction frequency and patterns, thereby integrating supplementary objective measures. Additionally, qualitative aspects of loneliness, such as its depth and contextual nuances, may require more comprehensive assessment methodologies. Tools like daily diaries or note-taking functions are advisable for future EMA approaches to capture more detailed information on the types and quality of social interactions. Moreover, framing questions with both positive and negative perspectives could mitigate response patterns where participants consistently answer in one direction, enhancing the robustness of findings in larger studies. Fourth, the use of a self-reported frailty assessment via the FPQ may introduce recall bias or inaccuracies in participants' self-perceptions. While the FPQ is a validated and practical screening tool for community-dwelling older adults, it may lack the granularity of more detailed clinical assessments. Future studies could consider integrating tools such as the Clinical Frailty Scale to capture a broader spectrum of frailty characteristics and enhance precision in frailty classification. Finally, while our study provided step-by-step guidance and a helpdesk for participants to navigate the EMA app, applying the EMA approach via mobile devices to older adults less familiar with mobile technology requires more comprehensive and detailed guidance. Ensuring that participants can use the mobile app independently is an essential step for successfully replicating this study methodology.

Conclusions

Our study suggests that daily social interaction is associated with frailty status among older adults with SCD or MCI, irrespective of the presence or severity of MBI symptoms. This finding builds on previous research, highlighting the positive potential of improving social interaction to reduce frailty risk among older adults with cognitive decline. To deepen understanding and enhance timely intervention strategies, future studies should explore the qualitative aspects of social interactions and frailty in greater depth. While the immediate clinical application of these findings may be limited, the use of real-time assessment in social functioning offers a promising path for further investigation. Integrating real-time social data with objective measures could provide a more personalized and comprehensive approach to care.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT; number 2022R1C1C2003706), the Basic Science Research Program through the NRF, funded by the Ministry of Education (number 2020R1A6A1A03041989), and the Brain Korea 21 Four Projects funded by the National Research Foundation.

Data Availability

The datasets used or analyzed in this study are available from the first author upon reasonable request.

Authors' Contributions

BK supervised, conceptualized, and designed the study framework. JK designed the study framework, analyzed the data, and drafted and edited the manuscript. DH, SY, and CK recruited the participants and collected the data. All authors contributed collectively to the data interpretation. The final draft of the manuscript has been thoroughly reviewed and approved by all the authors.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
The timing of the four daily alarms.
[PNG File , 2199 KB - mhealth_v13i1e64853_app1.png]

Multimedia Appendix 2
First step to respond to the application.
[PNG File , 1816 KB - mhealth_v13i1e64853_app2.png]

Multimedia Appendix 3
Second step to respond to the application.
[PNG File , 2234 KB - mhealth_v13i1e64853_app3.png]

Multimedia Appendix 4
Distribution of mean maximum daily social interaction scores by frailty status, based on multiple daily scores for each participant. The black histogram represents robust participants, the grey histogram represents prefrail group, and the blue histogram represents frail group. Participants recorded the number of face-to-face meetings, phone calls, or video calls lasting more than five minutes using a five-point Likert scale (0 for “no contact,” 1 for “once,” 2 for “two times,” 3 for “three times,” and 4 for “four or more times”).
[PNG File , 20 KB - mhealth_v13i1e64853_app4.png]

Multimedia Appendix 5
Distribution of mean maximum daily loneliness levels by frailty status, based on multiple daily scores for each participant. The black histogram represents robust participants, the grey histogram represents prefrail group, and the red histogram represents frail group. Participants assessed loneliness using a five-point Likert scale (0 = “very lonely” to 4 = “not lonely at all”), with higher scores indicating lower levels of loneliness.
[PNG File , 31 KB - mhealth_v13i1e64853_app5.png]

References

1. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. Lancet 2019;394(10206):1365-1375. [doi: 10.1016/s0140-6736(19)31786-6]
2. Guan DX, Rockwood K, Smith EE, Ismail Z. Sex moderates the association between frailty and mild behavioral impairment. J Prev Alzheimers Dis 2022;9(4):692-700. [doi: 10.14283/jpad.2022.61] [Medline: 36281673]
3. Fan S, Liang X, Yun T, Pei Z, Hu B, Ismail Z, et al. Mild behavioral impairment is related to frailty in non-dementia older adults: a cross-sectional study. BMC Geriatr 2020;20(1):510 [FREE Full text] [doi: 10.1186/s12877-020-01903-2] [Medline: 33246409]
4. Markle-Reid M, Browne G. Conceptualizations of frailty in relation to older adults. J Adv Nurs 2003;44(1):58-68. [doi: 10.1046/j.1365-2648.2003.02767.x] [Medline: 12956670]
5. Thibeau S, McDermott K, McFall GP, Rockwood K, Dixon RA. Frailty effects on non-demented cognitive trajectories are moderated by sex and alzheimer's genetic risk. Alzheimers Res Ther 2019;11(1):55 [FREE Full text] [doi: 10.1186/s13195-019-0509-9] [Medline: 31221191]
6. Wallace LMK, Theou O, Godin J, Andrew MK, Bennett DA, Rockwood K. Investigation of frailty as a moderator of the relationship between neuropathology and dementia in alzheimer's disease: a cross-sectional analysis of data from the rush memory and aging project. Lancet Neurol 2019;18(2):177-184. [doi: 10.1016/s1474-4422(18)30371-5]
7. Hajek A, Riedel-Heller SG. Loneliness and Social Isolation in Old Age: Correlates and Implications. London: Taylor & Francis; 2023.
8. Ofori-Asenso R, Chin KL, Mazidi M, Zomer E, Ilomaki J, Zullo AR, et al. Global incidence of frailty and prefrailty among community-dwelling older adults: a systematic review and meta-analysis. JAMA Netw Open 2019;2(8):e198398 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.8398] [Medline: 31373653]
9. Buckinx F, Rolland Y, Reginster JY, Ricour C, Petermans J, Bruyère O. Burden of frailty in the elderly population: perspectives for a public health challenge. Arch Public Health 2015;73(1):19 [FREE Full text] [doi: 10.1186/s13690-015-0068-x] [Medline: 25866625]
10. Won CW, Lee S, Kim J, Chon D, Kim S, Kim CO, et al. Korean frailty and aging cohort study (KFACS): cohort profile. BMJ Open 2020;10(4):e035573 [FREE Full text] [doi: 10.1136/bmjopen-2019-035573] [Medline: 32327477]

XSL•FO
RenderX

11. Jessen F, Amariglio RE, Buckley RF, van der Flier WM, Han Y, Molinuevo JL, et al. The characterisation of subjective cognitive decline. Lancet Neurol 2020;19(3):271-278 [FREE Full text] [doi: 10.1016/S1474-4422(19)30368-0] [Medline: 31958406]
12. Hsieh TJ, Chang HY, Wu IC, Chen CC, Tsai HJ, Chiu YF, et al. Independent association between subjective cognitive decline and frailty in the elderly. PLoS One 2018;13(8):e0201351 [FREE Full text] [doi: 10.1371/journal.pone.0201351] [Medline: 30071051]
13. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. Lancet 2006;367(9518):1262-1270. [doi: 10.1016/S0140-6736(06)68542-5] [Medline: 16631882]
14. Panza F, Lozupone M, Solfrizzi V, Sardone R, Dibello V, Di Lena L, et al. Different cognitive frailty models and health- and cognitive-related outcomes in older age: from epidemiology to prevention. J Alzheimers Dis 2018;62(3):993-1012 [FREE Full text] [doi: 10.3233/JAD-170963] [Medline: 29562543]
15. Borges MK, Canevelli M, Cesari M, Aprahamian I. Frailty as a predictor of cognitive disorders: a systematic review and meta-analysis. Front Med (Lausanne) 2019;6:26 [FREE Full text] [doi: 10.3389/fmed.2019.00026] [Medline: 30838210]
16. Grande G, Haaksma ML, Rizzuto D, Melis RJF, Marengoni A, Onder G, et al. Co-occurrence of cognitive impairment and physical frailty, and incidence of dementia: systematic review and meta-analysis. Neurosci Biobehav Rev 2019;107:96-103 [FREE Full text] [doi: 10.1016/j.neubiorev.2019.09.001] [Medline: 31491474]
17. Ruan Q, Yu Z, Chen M, Bao Z, Li J, He W. Cognitive frailty, a novel target for the prevention of elderly dependency. Ageing Res Rev 2015;20:1-10. [doi: 10.1016/j.arr.2014.12.004] [Medline: 25555677]
18. Solfrizzi V, Scafato E, Lozupone M, Seripa D, Giannini M, Sardone R, Italian Longitudinal Study on Aging Working Group. Additive role of a potentially reversible cognitive frailty model and inflammatory state on the risk of disability: the Italian longitudinal study on aging. Am J Geriatr Psychiatry 2017;25(11):1236-1248. [doi: 10.1016/j.jagp.2017.05.018] [Medline: 28689645]
19. Panza F, Solfrizzi V, Frisardi V, Maggi S, Sancarlo D, Adante F, et al. Different models of frailty in predementia and dementia syndromes. J Nutr Health Aging 2011;15(8):711-719 [FREE Full text] [doi: 10.1007/s12603-011-0126-1] [Medline: 21968870]
20. Mulero J, Zafrilla P, Martinez-Cacha A. Oxidative stress, frailty and cognitive decline. J Nutr Health Aging 2011;15(9):756-760 [FREE Full text] [doi: 10.1007/s12603-011-0130-5] [Medline: 22089224]
21. Nourhashémi F, Andrieu S, Gillette-Guyonnet S, Vellas B, Albarède JL, Grandjean H. Instrumental activities of daily living as a potential marker of frailty: a study of 7364 community-dwelling elderly women (the EPIDOS study). J Gerontol A Biol Sci Med Sci 2001;56(7):M448-M453. [doi: 10.1093/gerona/56.7.m448] [Medline: 11445604]
22. Nguyen TTH, Nguyen AT, Vu THT, Dau NT, Nguyen PQ, Nguyen TX, et al. Association of frailty status and functional disability among community-dwelling people aged 80 and older in Vietnam. Biomed Res Int 2021;2021:7109452 [FREE Full text] [doi: 10.1155/2021/7109452] [Medline: 34337049]
23. Ismail Z, Agüera-Ortiz L, Brodaty H, Cieslak A, Cummings J, Fischer CE, NPS Professional Interest Area of the International Society of to Advance Alzheimer's ResearchTreatment (NPS-PIA of ISTAART). The mild behavioral impairment checklist (MBI-C): a rating scale for neuropsychiatric symptoms in pre-dementia populations. J Alzheimers Dis 2017;56(3):929-938 [FREE Full text] [doi: 10.3233/JAD-160979] [Medline: 28059789]
24. Freitag S, Schmidt S. Psychosocial correlates of frailty in older adults. Geriatrics (Basel) 2016;1(4):26 [FREE Full text] [doi: 10.3390/geriatrics1040026] [Medline: 31022819]
25. Cacioppo JT, Fowler JH, Christakis NA. Alone in the crowd: the structure and spread of loneliness in a large social network. J Pers Soc Psychol 2009;97(6):977-991 [FREE Full text] [doi: 10.1037/a0016076] [Medline: 19968414]
26. Davies K, Maharani A, Chandola T, Todd C, Pendleton N. The longitudinal relationship between loneliness, social isolation, and frailty in older adults in England: a prospective analysis. Lancet Healthy Longev 2021;2(2):e70-e77. [doi: 10.1016/s2666-7568(20)30038-6]
27. Jarach CM, Tettamanti M, Nobili A, D'avanzo B. Social isolation and loneliness as related to progression and reversion of frailty in the survey of health aging retirement in Europe (SHARE). Age Ageing 2021;50(1):258-262 [FREE Full text] [doi: 10.1093/ageing/afaa168] [Medline: 32915990]
28. Zhaoyang R, Sliwinski MJ, Martire LM, Katz MJ, Scott SB. Features of daily social interactions that discriminate between older adults with and without mild cognitive impairment. J Gerontol B Psychol Sci Soc Sci 2021;79(4):gbab019 [FREE Full text] [doi: 10.1093/geronb/gbab019] [Medline: 33528558]
29. van Berkel N, Goncalves J, Lovén L, Ferreira D, Hosio S, Kostakos V. Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. Int J Hum Comput Stud 2019;125:118-128. [doi: 10.1016/j.ijhcs.2018.12.002]
30. Kang B, Ma J, Jeong I, Yoon S, Kim JI, Heo S, et al. Behavioral marker-based predictive modeling of functional status for older adults with subjective cognitive decline and mild cognitive impairment: study protocol. Digit Health 2024;10:20552076241269555 [FREE Full text] [doi: 10.1177/20552076241269555] [Medline: 39193313]
31. Boulton ER, Horne M, Todd C. Multiple influences on participating in physical activity in older age: developing a social ecological approach. Health Expect 2018;21(1):239-248 [FREE Full text] [doi: 10.1111/hex.12608] [Medline: 28768065]

XSL•FO
RenderX

32. Rowe JW, Kahn RL. Successful aging. Gerontologist 1997;37(4):433-440. [doi: 10.1093/geront/37.4.433] [Medline: 9279031]
33. Chen LK, Hwang AC, Lee WJ, Peng LN, Lin MH, Neil DL, Taiwan Health Promotion Intervention Study for Elders research group. Efficacy of multidomain interventions to improve physical frailty, depression and cognition: data from cluster-randomized controlled trials. J Cachexia Sarcopenia Muscle 2020;11(3):650-662 [FREE Full text] [doi: 10.1002/jcsm.12534] [Medline: 32134208]
34. Zhang H, Hao M, Hu Z, Jiang S, Li Y, Wang X, et al. Social frailty and its association with cognitive trajectories in older adults: a prospective cohort study. Alzheimers Res Ther 2025;17(1):50 [FREE Full text] [doi: 10.1186/s13195-025-01687-w] [Medline: 39972471]
35. Fewtrell MS, Kennedy K, Singhal A, Martin RM, Ness A, Hadders-Algra M, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch Dis Child 2008;93(6):458-461 [FREE Full text] [doi: 10.1136/adc.2007.127316] [Medline: 18495909]
36. Beller J, Geyer S, Epping J. Health and study dropout: health aspects differentially predict attrition. BMC Med Res Methodol 2022;22(1):31 [FREE Full text] [doi: 10.1186/s12874-022-01508-w] [Medline: 35094681]
37. Kim S, Kim M, Jung HW, Won CW. Development of a frailty phenotype questionnaire for use in screening community-dwelling older adults. J Am Med Dir Assoc 2020;21(5):660-664. [doi: 10.1016/j.jamda.2019.08.028] [Medline: 31672563]
38. Krueger KR, Wilson RS, Kamenetsky JM, Barnes LL, Bienias JL, Bennett DA. Social engagement and cognitive function in old age. Exp Aging Res 2009;35(1):45-60 [FREE Full text] [doi: 10.1080/03610730802545028] [Medline: 19173101]
39. Jung YC, Kim YH, Oh YS. Korea Media Panel Survey 2023. Jincheon-gun, Chungcheongbuk-do, Republic of Korea: Korea Information Soceity Development Institute; Dec 01, 2023:1-666.
40. Liu H, Xie QW, Lou VWQ. Everyday social interactions and intra-individual variability in affect: a systematic review and meta-analysis of ecological momentary assessment studies. Motiv Emot 2018;43(2):339-353. [doi: 10.1007/s11031-018-9735-x]
41. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 2019;7(10):e14149 [FREE Full text] [doi: 10.2196/14149] [Medline: 31621642]
42. Lin RS, Yu DS, Chau PH, Li PW, Ismail Z. Reliability and validity of the traditional Chinese version of the mild behavioral impairment - checklist among persons with mild cognitive impairment - a validation study. J Geriatr Psychiatry Neurol 2023;36(1):26-38. [doi: 10.1177/08919887221093363] [Medline: 35430911]
43. Kang Y, Chin J, Han N, Park J, Yeom J, Yang S, et al. P2-292: mild behavioral impairment (MBI) in MCI, SCD, and normal elderly: a pilot study for validation of the mild behavioral impairment checklist (MBI-C). Alzheimers Dement 2018;14(7S_Part_15):P793. [doi: 10.1016/j.jalz.2018.06.981]
44. Won C, Rho Y, SunWoo D, Lee Y. The validity and reliability of Korean instrumental activities of daily living(K-IADL) scale. Ann Geriatr Med Res 2002 Dec 31;6(4):273-280 [FREE Full text]
45. Ruthirakuhan M, Ismail Z, Herrmann N, Gallagher D, Lanctôt KL. Mild behavioral impairment is associated with progression to Alzheimer's disease: a clinicopathological study. Alzheimers Dement 2022;18(11):2199-2208 [FREE Full text] [doi: 10.1002/alz.12519] [Medline: 35103400]
46. Chon D, Lee Y, Kim J, Lee KE. The association between frequency of social contact and frailty in older people: Korean frailty and aging cohort study (KFACS). J Korean Med Sci 2018;33(51):e332 [FREE Full text] [doi: 10.3346/jkms.2018.33.e332] [Medline: 30546284]
47. Uno C, Okada K, Matsushita E, Satake S, Kuzuya M. Friendship-related social isolation is a potential risk factor for the transition from robust to prefrailty among healthy older adults: a 1-year follow-up study. Eur Geriatr Med 2021;12(2):285-293. [doi: 10.1007/s41999-020-00422-y] [Medline: 33128225]
48. Mulasso A, Roppolo M, Giannotta F, Rabaglietti E. Associations of frailty and psychosocial factors with autonomy in daily activities: a cross-sectional study in Italian community-dwelling older adults. Clin Interv Aging 2016;11:37-45 [FREE Full text] [doi: 10.2147/CIA.S95162] [Medline: 26811675]
49. Wang J, Maxwell CA, Yu F. Biological processes and biomarkers related to frailty in older adults: a state-of-the-science literature review. Biol Res Nurs 2019;21(1):80-106. [doi: 10.1177/1099800418798047] [Medline: 30198309]
50. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001;56(3):M146-M157. [doi: 10.1093/gerona/56.3.M146]
51. Holt-Lunstad J. The major health implications of social connection. Curr Dir Psychol Sci 2021;30(3):251-259. [doi: 10.1177/0963721421999630]
52. Devita M, Marescalco E, Bordignon A, Trevisan C, Sergi G, Coin A. Social interaction level modulates the impact of frailty on cognitive decline: a longitudinal study. Aging Ment Health 2024;28(4):652-657. [doi: 10.1080/13607863.2023.2247349] [Medline: 37602456]
53. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the lancet commission. Lancet 2020;396(10248):413-446. [doi: 10.1016/s0140-6736(20)30367-6]

XSL•FO
RenderX

54. Nakagawa T, Cho J, Yeung DY. Successful aging in East Asia: comparison Among China, Korea, and Japan. J Gerontol B Psychol Sci Soc Sci 2021;76(Suppl 1):S17-S26 [FREE Full text] [doi: 10.1093/geronb/gbaa042] [Medline: 32324214]
55. Alavi K, Sutan R, Shahar S, Manaf MRA, Jaafar MH, Abdul Maulud KN, et al. Connecting the dots between social care and healthcare for the sustainability development of older adult in Asia: a scoping review. Sustainability 2022;14(5):2573. [doi: 10.3390/su14052573]
56. Newmyer L, Verdery AM, Wang H, Margolis R. Population aging, demographic metabolism, and the rising tide of late middle age to older adult loneliness around the world. Popul Dev Rev 2022;48(3):829-862 [FREE Full text] [doi: 10.1111/padr.12506] [Medline: 37649654]
57. Seino S, Nishi M, Murayama H, Narita M, Yokoyama Y, Nofuji Y, et al. Effects of a multifactorial intervention comprising resistance exercise, nutritional and psychosocial programs on frailty and functional health in community-dwelling older adults: a randomized, controlled, cross-over trial. Geriatr Gerontol Int 2017;17(11):2034-2045. [doi: 10.1111/ggi.13016] [Medline: 28393440]
58. Lestari SK, Eriksson M, de Luna X, Malmberg G, Ng N. Frailty and types of social relationships among older adults in 17 European countries: a latent class analysis. Arch Gerontol Geriatr 2022;101:104705 [FREE Full text] [doi: 10.1016/j.archger.2022.104705] [Medline: 35461166]
59. Kojima G, Taniguchi Y, Aoyama R, Tanabe M. Associations between loneliness and physical frailty in community-dwelling older adults: a systematic review and meta-analysis. Ageing Res Rev 2022;81:101705. [doi: 10.1016/j.arr.2022.101705] [Medline: 35932978]
60. Ong AD, Uchino BN, Wethington E. Loneliness and health in older adults: a mini-review and synthesis. Gerontology 2016;62(4):443-449 [FREE Full text] [doi: 10.1159/000441651] [Medline: 26539997]
61. Kojima G, Taniguchi Y, Kitamura A, Fujiwara Y. Is living alone a risk factor of frailty? a systematic review and meta-analysis. Ageing Res Rev 2020;59:101048. [doi: 10.1016/j.arr.2020.101048] [Medline: 32173535]
62. Russell DW, Cutrona CE, McRae C, Gomez M. Is loneliness the same as being alone? J Psychol 2012;146(1-2):7-22. [doi: 10.1080/00223980.2011.589414] [Medline: 22303609]
63. Rapolienė G, Aartsen M. Lonely societies: low trust societies? further explanations for national variations in loneliness among older Europeans. Eur J Ageing 2022;19(3):485-494 [FREE Full text] [doi: 10.1007/s10433-021-00649-z] [Medline: 36052198]
64. Swader CS. Loneliness in Europe: personal and societal individualism-collectivism and their connection to social isolation. Soc Forces 2018;97(3):1307-1336. [doi: 10.1093/sf/soy088]

Abbreviations

EMA: ecological momentary assessment
FPQ: frailty phenotype questionnaire
K-IADL: Korean Instrumental Activities of Daily Living scale
MBI: mild behavioral impairment
MBI-C: Mild Behavioral Impairment Checklist
MCI: mild cognitive impairment
RRR: relative risk ratio
SCD: subjective cognitive decline

Edited by L Buis; submitted 30.07.24; peer-reviewed by V Rashedi, M Schraefel; comments to author 10.09.24; revised version received 25.10.24; accepted 10.04.25; published 22.04.25.

Please cite as:
Kang B, Hong D, Yoon S, Kang C, Kim JI
Assessing Social Interaction and Loneliness and Their Association With Frailty Among Older Adults With Subjective Cognitive Decline or Mild Cognitive Impairment: Ecological Momentary Assessment Approach
JMIR Mhealth Uhealth 2025;13:e64853
URL: https://mhealth.jmir.org/2025/1/e64853
doi:10.2196/64853
PMID:40210431

©Bada Kang, Dahye Hong, Seolah Yoon, Chaeeun Kang, Jennifer Ivy Kim. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health-Promoting Effects and Everyday Experiences With a Mental Health App Using Ecological Momentary Assessments and AI-Based Ecological Momentary Interventions Among Young People: Qualitative Interview and Focus Group Study

Selina Hiller[1,2], MSc; Christian Götzl[1,3,4], MA; Christian Rauschenberg[5,6], MSc, PhD; Janik Fechtelpeter[7,8,9,10], MSc; Georgia Koppe[7,8,9], PhD; Eva Wierzba[5], MSc; Julia Sauter[1]; Sina Dietrich[1]; Daniel Durstewitz[7,10,11], PhD; Ulrich Reininghaus[5,6,12,13], Dipl-Psych, MSc, PhD; Silvia Krumm[1,14], MA, PhD

[1]Department of Psychiatry II, University of Ulm and BKH Guenzburg, Guenzburg, Germany

[2]Technical University of Munich, School of Medicine and Health, Department of Psychiatry and Psychotherapy, TUM University Hospital, Munich, Germany

[3]Department of Forensic Psychiatry and Psychotherapy, University of Ulm and BKH Guenzburg, Ulm, Germany

[4]Department of Psychosomatic Medicine and Psychotherapy, University of Ulm, Ulm, Germany

[5]Department of Public Mental Health, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[6]German Center for Mental Health (DZPG), partner site Mannheim-Heidelberg-Ulm, Mannheim, Germany

[7]Interdisciplinary Center for Scientific Computing, Faculty of Mathematics and Computer Science, Heidelberg University, Heidelberg, Germany

[8]Hector Institute for AI in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany

[9]Department for Psychiatry and Psychotherapy, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany

[10]Department of Theoretical Neuroscience, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany

[11]Faculty of Physics and Astronomy, Heidelberg University, Heidelberg, Germany

[12]Centre for Epidemiology and Public Health, Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[13]ESRC Centre for Society and Mental Health, King's College London, London, United Kingdom

[14]Department of Psychiatry and Psychotherapy, Leipzig University, Medical Faculty, Leipzig, Germany

Corresponding Author:
Selina Hiller, MSc
Technical University of Munich, School of Medicine and Health, Department of Psychiatry and Psychotherapy, TUM University Hospital
Ismaningerstraße 22
Munich, 81675
Germany
Phone: 49 89 4140 ext 6430
Email: selina.hiller@tum.de

Abstract

Background: Considering the high prevalence of mental health conditions among young people and the technological advancements of artificial intelligence (AI)–based approaches in health services, mobile health (mHealth) apps for mental health are a promising way for low-threshold and large-scale mental health promotion, prevention, and intervention strategies, especially for young people. However, insufficient evidence on health-promoting effects and deficient user-centric designs emphasize the necessity for participatory methods in the interventions' development processes.

Objective: This study aimed to explore young people's everyday experiences using an AI-based mHealth app for mental health promotion based on ecological momentary assessments and ecological momentary interventions. Our analysis of qualitative data focused on exploring young people's use patterns in daily life and mental health–promoting effects.

Methods: We conducted problem-centered interviews and focus groups with a subsample of 27 young people aged 14 to 25 years, who were among the participants of 2 microrandomized trials testing and evaluating an AI-based mHealth app (AI4U

XSL•FO
RenderX

training). Our study used a participatory approach, with "co- and peer researchers" from the dialogue population actively engaged in research processes and data analysis. Structural content analysis guided the qualitative analysis.

Results: Participants reported enhanced emotional self-awareness and regulation in daily life through the ecological momentary assessments and ecological momentary interventions. Young people appreciated the AI4U training for managing emotions and stress. They had no trust issues regarding disclosing their mental health via the AI4U training in daily life. Some faced challenges integrating it into their daily routines and highlighted the value of autonomy in use decision-making processes.

Conclusions: Our findings reveal that young people benefited from enhanced emotional awareness and management through the use of the AI4U training, appreciating its anonymity for facilitating emotional disclosure. The results suggest that enhanced self-directed use may improve daily life integration, although participants noted that they sometimes avoided using the AI4U training during distress despite recognizing its potential benefits. These findings indicate the importance of balancing directed use and autonomy in digital interventions to harmonize compliance with effectiveness in daily life. We highlight the importance of participatory research for tailored digital mental health solutions.

(JMIR Mhealth Uhealth 2025;13:e65106) doi:10.2196/65106

KEYWORDS

mobile health app; ecological momentary assessment; ecological momentary intervention; just-in-time adaptive intervention; digital training; mental health promotion; artificial intelligence; participatory research methods; user experiences; cocreation; young people; youth; adolescent; mHealth; interview; focus group; participatory approach; self-awareness; digital intervention; digital health

Introduction

Digital Mental Health Interventions for Young People

According to the Global Burden of Disease study, almost 14 million young people in Europe aged 10 to 24 years in 2019 met the diagnostic criteria for a mental health condition [1]. The onset of most mental health conditions has been found to occur in youth or early adulthood, and there is an evident link between occurrence in early life and persistence or exacerbation in adulthood, highlighting the importance of early interventions [2-5]. Research calls for holistic and multi- and transdisciplinary approaches for mental health promotion (eg, improving resilience and strengthening protective factors) and more tailored interventions for young people [2]. There is evidence indicating that existing services are underused by young people due to social factors such as stigma, embarrassment, lack of trust in mental health professionals, and structural barriers related to the availability and accessibility of proper support [2,4,6-8]. The desire for self-reliance and autonomy grows from puberty into early adulthood and constitutes a further major barrier in seeking informal or professional help for mental health problems among young people [9].

Technological innovations provide new forms of efficient, large-scale, and low-threshold health services. With the growing accessibility to technology, smartphones have become powerful tools, making mobile approaches to mental health care and promotion services increasingly popular [10-13]. The widespread use of smartphones among young people makes them an attractive platform for mobile health (mHealth) apps or digital interventions aimed at promoting mental health through self-help features [14-23]. Evidence-based mHealth apps have the potential to target a wide range of users, promoting mental health and preventing mental health conditions in both clinical and nonclinical populations [10]. Most currently available reimbursable digital health applications in Germany meet the evaluation criteria for patient-relevant clinical benefits [24]. Digital self-monitoring of symptoms or behavior (eg, via ecological momentary assessments [EMA]) allows for the collection of ecological valid data on mental health outcomes and precise clinical assessment [11,21,25-27]. On the basis of EMA data, real-time and context-dependent personalized intervention suggestions (eg, ecological momentary interventions [EMI] or, synonymously, just-in-time adaptive interventions) for improving mental health can be provided to the user and subsequently evaluated [28-32]. Interventions that are based on real-time data may further increase user adherence and engagement and improve effects on mental health [10].

Research emphasizes the importance of personalization in mHealth apps as it simplifies the process for users to find suitable interventions and can improve the effectivity of the selected interventions [10]. Modern artificial intelligence (AI) methods pave the way for new and targeted forms of mHealth services [33-35]. AI is used for a variety of tasks, including risk prediction and prognosis, diagnostic classification, and personalization of interventions, among others [34,36]. In the context of personalizing mHealth apps, time-series models such as recurrent neural networks have been proposed to build a dynamic and predictive model of individual behavior based on which interventions can be optimally selected [32,35]. This allows for a personalized and adaptive allocation of EMI.

However, time expenditure and the potential of mHealth apps as an additional stressor in daily life have been recently identified as major barriers to daily mHealth app use among young people [37]. In addition, studies still face high attrition rates and low engagement from young participants, such as in the study by Peuters et al [15], in which young people reported neglecting several features of the mHealth app, mainly due to time constraints. Interviews with homeless young people using an EMA-based mHealth app revealed that repetitive and predictable questions and answer options quickly led to boredom among the participants [38].

Public health literature increasingly calls for research to guarantee active engagement with dialogue populations in the

XSL•FO
RenderX

research process to enhance the acceptance and implementation of health promotion and prevention measures [39-44]. This approach ensures the inclusion and empowerment of affected individuals, including marginalized groups, in research processes that impact them and establishes transformative knowledge through the integration of new perspectives. Exploratory qualitative research and participatory approaches are essential to inquire about the existing gaps in understanding how mHealth app use can be optimized. Digital services, especially for young digital natives, require a collaboration on equal terms among and active participation by health care professionals, developers, and individuals from the dialogue population, and successful coproduction especially requires the integration of potential users early in the development process [11,17,45,46]. An in-depth understanding of expectations or experiences of people regarding digital mental health care approaches is crucial for the development of effective and trustworthy solutions [47].

However, qualitative studies on people's real-world experiences with mHealth apps based on EMA have predominantly centered on specific groups, such as patients with bipolar disorder [48], patients with depression [49], or homeless young people [38], or on family research [50]. One study conducted interviews with college students on their experience with an EMA-based mHealth app aimed at improving emotional well-being [51]. However, the participation period was rather short, averaging 10 days, and the evaluation focused on methodological and design-related research questions rather than on individual effects on mental health. Qualitative data on the use of EMI tend to be lacking. One study investigated the effects of a mobile meditation intervention through participant interviews [52]. However, Xie et al [52] targeted adults diagnosed with clinically elevated levels of depression or anxiety. To our knowledge, no existing qualitative study using principles of participatory research has explored the subjective effects of an AI-based mHealth app on the mental health and emotional resilience in daily life of healthy young people.

Study Objectives

The objective of this qualitative study was to explore the perspectives of young people (aged 14-25 years) on the use of an AI-based digital training administered via a smartphone-based app for personalized mental health promotion in youth based on EMA and EMI (AI4U training). We were particularly interested in their views regarding use and integration of the mHealth app into everyday life and the subjective effects on their mental health, especially health-promoting effects. The following research questions were addressed:

1. What are young people's subjectively perceived health-related effects of the AI-based mHealth app?
2. How do young people perceive the use of an AI-based mHealth app designed to promote mental health in everyday life based on EMA and EMI?

Methods

Study Design

Study Background

This qualitative study formed part of the "AI4U—artificial intelligence for personalized digital mental health promotion in youth" living laboratory. Living laboratories strive to foster collaboration between science and society and engage target or dialogue populations and stakeholders from a diverse range of relevant sectors to explore and study innovative models of action such as innovations in health services [53]. The living laboratory AI4U was dedicated to the participatory development, optimization, and evaluation of an AI-based mHealth app for personalized mental health promotion in youth (AI4U training). The living laboratory involved the direct participation of relevant stakeholders, users from the dialogue population, and an interdisciplinary research group embedded in a transdisciplinary infrastructure to ensure ongoing quality management, sustainability appraisal, and dissemination. Optimization and initial evaluation of the AI4U training involved 3 separate consecutive microrandomized trials (MRTs) in line with the multiphase optimization strategy [54]. In the MRTs, participants aged 14 to 25 years were recruited and asked to complete the 40-day AI4U training using an mHealth app, with a target sample size of 60 participants for each MRT. In the 3 within-subject MRTs, participants were repeatedly randomized using a 1:1 ratio to either AI-based allocation of EMI components based on EMA ratings or to a random allocation of EMI components at a maximum of 210 decision points per participant to examine the effect of the AI algorithm on proximal outcomes in youth. EMA ratings consisted of questions on momentary (positive and negative) affect, activities (eg, physical activity), and perceived contexts (Multimedia Appendix 1). Further details on the living laboratory AI4U can be found in previous papers [55-57].

This study focused on qualitative data collected as part of the first 2 MRTs (MRT 1 and MRT 2) to allow for adaptation and optimization of the mHealth app based on subjective experiences. Due to the participatory approach of the AI4U living laboratory, young people from the study's dialogue population were involved early in the study. A female researcher (SH) was employed as a "peer researcher." This ensured the active participation of young people in many aspects of the decision-making processes (including study design, implementation, and dissemination). "Co-researchers" (JS, SD, and 2 others), that is, young people of the same age as the dialogue population with no pertinent qualitative research experience, were included in project-related research processes by taking part in "participatory research workshops". Similar to research workshops held with qualitative researchers, these workshops aimed to increase the quality of data analysis through joint interaction [58,59]. Co-researchers met on a regular basis with the research team (SH and CG) to participate in qualitative research. To enable adequate participation, co-researchers received training on core competencies in qualitative research (eg, interviewing and data analysis) and were involved in various stages of the research processes throughout the study. Key

XSL•FO
RenderX

elements were the participatory development of interview topic guides, joint moderation of focus groups (FGs), and subsequent discussion of preliminary results [56]. Similar participatory health research has shown that participatory evaluation processes with nonscientific co-researchers from a project's dialogue population promotes the generation of transformative knowledge and facilitates shared decision-making in qualitative research, particularly between academic researchers and co-researchers [60].

Interventions

The AI4U training is a digital training that consists of four innovative forms of digital mental health promotion, which users can apply directly in their everyday lives: (1) EMA [25,26], (2) EMI [29,31,61], (3) AI [62-64], and (4) a web-based dashboard for digital monitoring and feedback. The AI4U training was delivered over a 40-day intervention period in addition to public mental health care as usual available through the German health and social care system. The digital training consisted of at least one face-to-face session with a coach and an EMI administered through a smartphone-based app for adaptive real-time and real-world transfer. The sessions with the coach provided guidance on the content and structure of the training and were complemented by optional sessions with counselors for young people receiving care from educational counseling services. The AI4U training consisted of a 10-day introductory period with self-monitoring using EMA (8 per day) and a 30-day training period with self-monitoring using 6 EMA per day. EMA entries could be viewed on a dedicated digital dashboard using a web browser. The AI4U training intended to improve resilience, fostering a compassionate self-image and strengthening self-care skills in everyday life (Multimedia Appendix 1). Following each MRT, the digital training was optimized further based on feedback from participants at the end of the study and data from qualitative interviews.

Sample

We recruited participants for the qualitative process evaluation consecutively among the young people who had used the AI4U training for the previous 40 days. For this, we used a purposive sampling strategy to select participants with relevant insights into the use and impact of the AI4U training. This method allowed us to specifically target individuals who met predefined criteria, such as various age groups or users with various degrees of adherence. In addition, we aimed for a balanced gender distribution and inclusion of young people from varying socioeconomic and educational backgrounds (Multimedia Appendix 2). We contacted participants via email to inform them of the purpose of our interviews and the possibility of study participation.

Data Collection

We used the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [65] to ensure a clear and thorough description of the qualitative research methodology (Multimedia Appendix 3). Problem-centered interviews (PCIs) were held between October 2022 and November 2023. The FGs were held between December 2022 and July 2023. All interviews were conducted remotely using a video-based online tool. Experienced researchers (SH and CG) conducted the PCIs. In line with our participatory approach, one of our co-researchers (JS and SD) moderated the FGs. They signed a confidentiality agreement a priori and were provided with a guide for remote FGs (Multimedia Appendix 4). To ensure the quality of data collection, we practiced the moderation beforehand in a participatory research workshop, and co-researchers were supported by one of our research assistants with adequate scientific expertise (SH and CG).

For the PCIs and FGs, we used semistructured topic guides (Multimedia Appendix 5). The topic guide for the FGs was developed together with our co-researchers during our participatory research workshops. Both interview methods covered questions on (1) the use of the AI4U training in everyday life, (2) the evaluation of the EMA and EMI, and (3) trust factors in the AI4U training. Although the interview topic guides largely overlapped, the PCIs focused on the subjective health-promoting effects of the AI4U training, whereas these aspects were not a primary discussion point in the FGs. We used Microsoft PowerPoint slides (Microsoft Corp) to visualize the EMI and the dashboard interface to facilitate feedback on the key content of the AI4U training. After each FG, there was a short discussion between the co-researcher and the co-moderator to reflect on first impressions or discuss potential field notes taken.

Data Analysis

The PCIs and FGs were held in German, audio recorded, and transcribed verbatim. For the transcription, we used f4 automatic speech recognition (Dr. Dresing & Pehl GmbH). Transcripts were subsequently pseudonymized and revised by student assistants. SH conducted data analysis using structural qualitative content analysis [66,67]. After reading of and familiarization with the data material, particularly interesting or important text passages were highlighted. In accordance with the semistructured interview topic guides, we created thematic categories and corresponding descriptions of coding rules for each of the categories. We then started with open coding of data material, facilitating the exploration of previously undisclosed themes. This process was followed by a subsequent categorization of coded text passages according to the thematic categories previously defined. If emerging codes were not suitable to any of the previously defined categories, we inductively added new thematic categories or codes. We coded transcripts using the MAXQDA 2020 software (VERBI GmbH) [68]. Validation procedures within the data analysis process encompassed multiple stages. After reviewing the initial coding tree within our research team (SH, CG, and SK), we modified and structured it accordingly. The coding trees and passages from the FG transcripts were subsequently thoroughly examined and validated with the co-researchers (JS, SD, and 2 others) in 2 distinct participatory research workshops. We marked the codes and subcodes that emerged through cocreation and validation in the participatory research workshop in the coding trees of the MAXQDA files accordingly (Multimedia Appendix 6). In a third participatory research workshop, we examined all the text passages previously highlighted as interesting and discussed and interpreted their potential meaning together.

XSL•FO
RenderX

Ethical Considerations

This study was funded by the Ministry of Science, Research, and Arts of the State of Baden-Württemberg, Germany (study period from January 2021 to June 2024), and approved by the Medical Ethics Review Committee II of Heidelberg University (Mannheim Medical Faculty; reference 2022-550). All participants provided written informed consent. For underage participants, we also obtained informed consent from parents or legal guardians. Participation in the study was voluntary. Consent could be withdrawn at any time without giving reasons. Participants were covered by liability and commuting accident insurance for the duration of the study. Participant data were analyzed and exchanged exclusively using study pseudonyms. The pseudonymized data were securely stored and processed in accordance with the General Data Protection Regulation. Participants received a compensation of €20 (US $22.40) in the form of a voucher redeemable at various stores across Germany.

Results

Sample Characteristics

Overall, 114 young people completed the 40-day AI4U training during the first 2 MRTs (N=57 in MRT 1 and N=57 in MRT 2). Of those 114 young people, we included 27 (23.7%; 13/57, 23% from MRT 1 and 14/57, 25% from MRT 2) in our qualitative study distributed across 16 PCIs and 3 FGs with 3 to 5 participants each (Multimedia Appendix 2). In the PCIs, the youngest participant was aged 14 years, and the oldest was aged 25 years (mean age 20.38, SD 3.78 years), whereas in the FGs, the youngest participant was aged 18 years and the oldest was aged 25 years (mean age 21.36, SD 1.84 years). Most participants had obtained their high school diploma or were already pursuing higher academic education (20/27, 74%). The average duration of the PCIs was 44 minutes, and the average duration of the FGs was 1 hour and 11 minutes. In the PCIs, gender distribution was even (8/16, 50% male individuals and 8/16, 50% female individuals). For MRT 1, we conducted 2 FGs divided by gender and with an equal number of participants (3 participants in each FG). The third FG with participants of MRT 2 was composed of 5 female individuals only.

Overview

The following sections report on the subjective perspectives of participants on the effects of the AI4U training on their mental health, safety and trust considerations, the integration and role of the AI4U training in everyday life, and the value of self-directed use behavior. As the AI4U training was primarily administered through a smartphone-based app, participants used the terms *AI4U training* and *mHealth app* interchangeably in the interviews. Content within parentheses in the following sections reflects the relative frequency, and content within square brackets reflects attribution of participant statements made during the PCIs or FGs. *P* stands for *participant*. Multimedia Appendix 2 provides the participant characteristics.

Subjectively Experienced Effects Following the AI4U Training

Mood Monitoring via EMA Seemed to Increase the Mental Health Awareness of Young People

Most participants (PCIs: 14/16; FGs: 7/11) reported that the awareness of their acute mental health status via EMA led to improved self-consciousness of emotions and subsequent self-reflection. One person especially appreciated the following:

> *...that it made you very aware of your own emotions. And just that, I found, often helped in some way, because then I frequently...had to think about how I was feeling at the moment. And I really liked that. Um, yes, just to become more aware of yourself and your own feelings.* [PCI_P27; page 103]

Participants mentioned that EMA were especially helpful to understand the reasons or contextual factors behind their emotions, that is, why they were feeling distressed (PCIs: 6/16; participant FG2_P1):

> *How do I feel, why do I feel this way? It makes you a bit more reflective. Understanding the reasons behind your emotions, why you behave like that, contributes to feeling better. I do feel that I learned something about myself through this.* [PCI_P16; page 134]

One interviewee argued that understanding contextual triggers enabled her to classify emotions on a more “objective level” (PCI_P16; page 60), whereas another participant (PCI_P13) highlighted that EMA helped reconsider recent emotional reactions as less severe upon further reflection.

Several young people (PCIs: 7/16; participant FG3_P3) highlighted that they aimed to sustain their mental health awareness over the long term, reflected more consistently on their daily experiences, and focused on the positive aspects in their everyday life even after the study period:

> *But I have to say, after I returned the [study phone], I had this feeling of emptiness [laughs]. Simply because I somehow mentally got used to these [EMA]. So, whenever something happened during the day, I immediately thought, “Okay, this has contributed to me being like a four out of seven in terms of my disappointment or something.” So, it definitely became ingrained in my mind.* [FG3_P3; page 55]

Some participants (PCIs: 6/16; FGs: 3/11) reported being interested in their long-term behavior or emotional trajectory and accessed the dashboard rather at the end of the study period. They described the dashboard and its visualizations of EMA responses as interesting but with rather low value relevant to their mental health promotion (PCIs: 6/16; FGs: 3/11):

> *[The dashboard]...was and is not a surprise because these are things that I have learned about myself for a long time. That's why I simply believe that it wasn't so surprising to see it, because I've repeatedly thought about these things before.* [PCI_P26; page 182]

XSL•FO
RenderX

However, some participants (PCIs: 5/16; FGs: 4/11) reported that they would have likely used the dashboard more frequently if it had been integrated into the app and not only accessible via a browser on a laptop.

EMI Appeared to Improve Emotion Regulation and Health-Promoting Routines in the Daily Life of Young People

According to a number of participants (PCIs: 7/16), EMI assisted in regulating emotions more effectively:

> *...I just felt mentally calmer and maybe you didn't see the problem as the central problem, but everything was put into perspective a bit and it, um, yes, it mentally brought me down a bit and calmed me down.* [PCI_P22; page 186]

Through improved emotion regulation, several participants (PCIs: 3/16) mentioned learning that emotions, especially negative ones, were transient. For 2 of the 27 of the participants (PCI_P12 and PCI_P22), this health-related knowledge was retained in the long term. One participant reported that the EMI "Emotions as a Wave" (Multimedia Appendix 1) was particularly helpful:

> *I've always seen emotions as something that keeps coming back and stays, not something that fades away. But "Emotions as a Wave" is something that has changed my perspective on emotions itself because...it taught me that emotions can also pass and that resulted in me becoming a calmer person, actually.* [PCI_P12; page 100]

The breathing intervention was used as a health-promoting routine for relaxation and composure throughout the day by several individuals (PCIs: 5/16; participants FG3_P3 and FG3_P4) even after completing the AI4U training. Many participants (PCIs: 10/16; FGs: 6/11) reported that the "Moments of Joy" and "Positive Data Log" interventions stimulated positive affect by helping them focus on the enjoyable aspects of a day and diverting their attention from the negative ones. Another participant described the "Moments of Joy" intervention as helpful in becoming more "self-efficacious" (PCI_P14; page 188), realizing the extent to which she could also influence her own happiness or satisfaction. Another 2 of the 27 participants (PCI_P14 and PCI_P28) reported that they became aware of the importance of breaks in their daily lives:

> *...Whenever I have school assignments, I let them influence me a lot. I knew that before, but I've learned a bit that they shouldn't completely fill my daily life, that I should still give myself more time. And the app helped me in setting that up somehow.* [PCI_P28; page 212]

When comparing EMA with EMI, a few participants (PCI_P17, PCI_P18, and FG2_P2) emphasized the particular benefit from EMA, whereas one described the EMI as "nice to have" but not necessary (FG2_P2; page 122). However, others (PCIs: 3/16) argued that perception of emotions via EMA was their first step toward change. Some participants (PCIs: 3/16) noted that EMA alone would be insufficient to develop new health-promoting routines. One participant reasoned that awareness of their own emotions enabled them to better manage them through the proposed EMI:

> *[The EMA have] caused me to simply become more aware of how I feel. And I think that led to better understanding these feelings and, as a result, either changing or accepting them. So, yes, just being able to handle them better...And especially the [EMI] could often make a difference there. For me, it was really often with nervousness that the [EMI] helped, for example.* [PCI_P27; page 211]

Young People's Trust Factors on Data Disclosure During Use of an mHealth App

Transparent Data Handling Seems to Be a Crucial Factor for Young People's Confidence in Data Protection

As long as young people were confident that their data would not be misused, they reported that data disclosure via smartphones was not a use barrier (PCIs: 3/16; FGs: 3/11). A total of 4 of the participants (PCI_P15, PCI_P28, FG1_P2, and FG2_P2) stressed that transparency about the use of their data was crucial. Another 2 participants (FG1_P2 and FG1_P1) found it acceptable for the app to collect data only based on voluntary input and expressed concern about the collection of information that was not self-entered. Some participants (PCIs: 4/16; participant FG2_P3) felt more secure using the AI4U training because it was developed by professionals and used within an official study. In comparison, mHealth apps from an app store were deemed as less trustworthy by 2 participants (PCI_P28 and FG2_P3).

Some argued that, given the vast amount of personal information on the internet, data disclosure is something one "can't prevent anyway" (PCI_P22 [page 233] and PCI_P21 [page 48]), and for others, it "doesn't matter now anyway" (PCI_P21 [page 48] and FG2_P1 [page 97]). In total, 2 participants (PCI_P17 and PCI_P26) emphasized that one is used to providing personal information on their smartphone in everyday life. Therefore, it did not feel as if one was conveying information to someone else:

> *I didn't really feel like I was sharing anything, but rather that I was doing it for myself...I mean, I'm used to handling a smartphone and I always write my notes on the iPad. So, it wasn't strange for me to type something on a device.* [PCI_P26; page 63]

In total, 2 participants reported not perceiving the AI4U training as "invasive," particularly when compared to algorithms on social media (FG3_P3 [pages 259 and 263] and FG3_P4 [page 287]). Some (PCIs: 3/16) noted that the app could probably not make any meaningful use of the personal data anyway. One interviewee emphasized her autonomy and personal agency in data sharing:

> *I'm generally not the type of person to think, "Oh God, data privacy," and that something private will be stolen from me, because I still feel like I'm an independent person. If an AI knows more about my emotions, I don't know how it could influence me in*

XSL•FO
RenderX

> *the long run. I just don't see the risk there.* [PCI_P26; page 67]

Anonymity as a Necessary Requirement for Sharing Emotions

One participant noted that one had to "get used to" to being asked about their mental health or described it as an unfamiliar "change" (PCI_P15; page 4). For some (PCIs: 4/16; participant FG1_P3), disclosing emotions via the EMA felt unusual at first:

> *I first had to get used to the idea that there's now an app, to which I should somehow describe my feelings, in some tasks that was the case. That was definitely something to get used to, but...somehow it subsided a bit, I would say. You just really got used to the fact that there's no human there. Yeah, but it was a bit strange at the beginning...* [FG1_P3; pages 94-116]

Many participants (PCIs: 9/16) argued that they developed trust in the AI4U training because they did not perceive it as a real person or a human stranger. According to several participants (PCIs: 4/16), this significantly facilitated the expression of emotions:

> *So [it] definitely made it easier [to trust the app]. I believe I would have had a lot more reservations, if I had known that it was a human responding to me, especially because it would be a person I don't know. And an AI is still an AI, even if your subconscious mind sometimes forgets that, but I think it definitely made it easier for me to talk about it and to overcome that barrier.* [PCI_P27; pages 249-251]

Young people reported feeling that their data were more securely handled than with humans and highlighted the barriers to sharing emotions during face-to-face communication as a reason (PCIs: 6/16; participants FG1_P3 and FG2_P1). Reported concerns of some young people (PCIs: 3/16; participant FG2_TN1) focused, for instance, on the fear of nonacceptance, judgment, and intimate disclosure typical in interactions with humans. According to some of them, the AI4U training was "just" an app that, unlike friends, could not misuse the data about their mental health:

> *On one hand, I felt safe, so to speak, because I know that an AI can't really do much with [my information]. If I tell a person about it, I'm exposing myself to the person, and I don't know what the person does with the information. And that wasn't the case with the AI.* [PCI_P16; page 36]

Certain participants (PCIs: 6/16) reported that an mHealth app allows for a certain emotional distance and anonymity as a prerequisite for being open about their own emotions:

> *...you don't see the direct reaction of people. The app accepts that. You can input everything and I believe you can be more honest with such an app, than if a person would ask me when I'm really feeling bad in a situation, "How are you?" I would probably still say, "Yeah, I'm fine," but with the app, well, nobody really gets to know...It provides a bit more anonymity, I would say.* [PCI_P14; page 42]

The AI4U training was sometimes described as a "neutral diary" and catch basin for emotions that provided feedback based on objective data, not subjective assessment, for instance, compared to feedback from a therapist (PCIs: 4/16). Another participant (PCI_P12) identified this neutral feedback as stabilizing and more reliable and accurate, in contrast to human feedback.

Integration of AI4U Training Into Young People's Everyday Lives

Role of the mHealth App in the Daily Lives of Young People

Given that some participants (PCIs: 6/16) reported having a strict daily routine and everyday stress, they mentioned that they expected the AI4U training to assist them by helping reduce daily stress and monitor their mood. Some participants (PCIs: 6/16; FG3_P4) emphasized that they would not have paused their daily routine without the prompt from the AI4U training:

> *The app allowed me, so to speak, to take breaks, because it prescribed them to me. So, I think that when such an impulse comes from outside, telling you "now do it," you're more likely to do it, otherwise you always find an excuse...* [PCI_P14; page 164]

One participant highlighted that this heightened the awareness of her daily life structure and the importance of self-care within it:

> *In daily life, nothing reminds me...to check on my well-being. It doesn't mean that my everyday life is very bad, but it's about the fact that my everyday life is sometimes so hectic that I forget to take care of my well-being, my mental state...We're so focused on our tasks that we forget about ourselves...and the app reminded one...to pay attention of what one should actually take care of, because the basic mindset shapes your everyday life. If you're not doing well, you won't perform at school, for example.* [PCI_P12; pages 34-50]

Several young people (PCIs: 6/16; participants FG3_P3 and FG3_P2) pointed out that they preferred using the AI4U training when experiencing discomfort in daily life. One participant remarked that she completed the EMA more frequently when experiencing negative emotions:

> *I feel like I tended to do them more when I wasn't feeling well...And when I was feeling good, I thought, "Oh well, I don't need to enter it now, I'm doing fine."* [PCI_P23; page 102]

In total, 2 participants (PCI_P27 and FG3_P3) considered it less necessary when the app suggested interventions when they were feeling well. Another 2 participants (PCI_P17 and PCI_P22) reported that, when they were in a positive mood, they clicked through the EMA more quickly and superficially. One participant reflected on the AI4U training as rather an option to be accessed on an ad hoc basis in times of acute need:

> *I do think it could be a permanent companion, since one doesn't have to use it every day, if you realize, today I don't need it, today I'm fine.* [PCI_P14; page 172]

XSL•FO
RenderX

Without an urgent need, young people noted that they prioritized social interactions, especially if their daily life rhythm was quite strict, for example, due to work or educational duties (PCIs: 11/16; participant FG3_P1):

> *...I prefer to go to yoga or something similar and think about [my emotions] there, rather than having to fill out on my phone six times a day how I feel, where it just fell short. Eventually, I already knew the questions by heart and just clicked through them quickly.* [FG3_P1; page 295]

One participant (PCI_P29) mentioned refusing the AI4U training in acute situations of stress or negative feelings, preferring suggestions when he calmed down moments later. During stressful times, 2 participants (PCI_P23 and PCI_P15) were skeptical about repeatedly entering their emotions into EMA or EMI. However, another participant (FG2_P2) reflected that he neglected the app's suggestions in challenging situations even when they could have been helpful.

Some participants (PCIs: 7/16) noted that the health-promoting routines they had learned during the AI4U training were practiced less frequently after the study period. Others (PCIs: 3/16) observed that, without a reminder, it was more difficult to maintain enough self-discipline for engaging in health-promoting routines. According to 2 participants (PCI_P14 and PCI_P23), habits were difficult to change, especially with increased stress and demands in everyday life. A few participants reported being less inclined to disrupt their routines (PCIs: 3/16):

> *I just think my daily routine is quite pre-structured with things that I can't change fundamentally. So, I have to go to school when I have to go to school...I wouldn't say that [the AI4U training] has restructured anything in that regard.* [PCI_P28; page 204]

In total, 2 participants (PCI_P24 and PCI_P25) felt that they might not be in as great need of mental health training compared to young people with actual mental health conditions:

> *It felt more for, uhm, sick people...who have serious problems in everyday life...I believe that it is much better received by them than it is by us (mhm), because it sometimes just felt like it wasn't tailored exactly to us. So, what I expected was not what it turned out to be.* [PCI_P24; page 27]

Experiences With and Challenges Integrating the AI4U Training Into Daily Life

Overall feedback on daily life integration of the AI4U training was mixed. Most participants argued that app use was feasible when being at home or alone or when having a flexible daily rhythm (PCIs: 11/16; participants FG1_TN2 and FG2_TN2), and more challenging when they were on the go, in public, at work, or at school (PCIs: 10/16; FGs: 8/11). According to some participants (PCIs: 7/16), EMA or EMI were skipped when they spent time with their friends or family. A few participants (PCIs: 3/16; FGs: 4/11) reported feeling uncomfortable when they had to justify smartphone use to friends who perceived it as impolite, or when the alarm disturbed others. A total of 2 participants (PCI_P15 and PCI_P28) expressed discomfort or concerns about others witnessing them and judging their engagement with the intervention:

> *I somewhat refrained from EMI, especially at school or in public, as I was concerned about what my classmates might think if suddenly I close my eyes and sat at my desk.* [PCI_P15; page 18]

Some participants (PCIs: 5/16; FG2_P2) indicated that use significantly depended on an individual's daily rhythm and was linked to the time of day and adaptability of daily duties. Several participants (PCIs: 8/16; FG2_P1) noted that a self-determined daily routine made it easier to integrate the AI4U training. Some (PCIs: 6/16; FGs: 3/11) reported that, when time was limited or the environment was unsuitable, EMA or EMI were either rushed through or skipped completely. A few participants (PCIs: 3/16) observed that, as the AI4U training became part of their daily routine, its use became less demanding.

Some participants (PCIs: 6/16; FGs: 3/11) reported EMA as less time-consuming and easier to fit into their daily lives compared to EMI. However, they reported that it was dependent on the perceived complexity of the EMI. For instance, a few (PCIs: 3/16; FGs: 3/11) perceived the "Breathing Intervention," "Moments of Joy," and "Positive Data Log" interventions (Multimedia Appendix 1) of the AI4U training appealing due to their simplicity, the lack of mental strain, and the feasibility to integrate them into daily life. More complex interventions such as "Emotional Compass" were often considered too challenging to apply in a daily life context, resulting in limited impact (PCIs: 9/16; FGs: 5/11). Others argued that EMI required increased mental effort and, consequently, more time (PCIs: 3/16; FGs: 5/11):

> *I declined EMI more often than EMA, because I found them more challenging to integrate into everyday life, especially while being on the go.* [PCI_P27; page 179]

Some participants agreed that the alarm of the EMA prompts was sometimes distracting (PCIs: 6/16; FGs: 6/11) and 6 EMA per day were too demanding to be conducted regularly as requested (PCIs: 3/16; FGs: 5/11):

> *I probably wouldn't have needed [the EMA] for myself so frequently. Maybe it would have been enough for me two times a day, in the morning or evening, or perhaps midday, and then an EMI in the evening. That would probably have been the best approach for me.* [PCI_P22; page 178]

One participant (PCI_P28) mentioned that she felt bad declining prompts, whereas others reported that they felt "pressured" (PCI_P23; page 318) or "confined" by the app (FG2_P2; page 118). As a consequence, a number of participants (PCIs: 6/16; FG3_P4) described that they sometimes felt forced to conduct the AI4U training. A few (PCIs: 4/16; FGs: 4/11) sometimes perceived the AI4U training as an additional stressor:

> *It was difficult to integrate the app into my daily life in a way that made sense, that I can take away what they wanted to provide me, like doing EMI and such. For me, it was more of a stress factor (laughs) that I should take a break and do the EMI...* [FG3_P1; page 43]

XSL•FO
RenderX

As reported by some (PCIs: 8/16), it was burdensome to pause or interrupt ongoing activities in their daily lives. One participant noted having felt distracted by the alarms:

> *I also found it a bit distracting...When you're trying to focus, you're repeatedly pulled out of that focus to do [the EMA or EMI]. I found it a bit challenging to sit down in the library and start doing breathing exercises. So, I have to say, integrating it into my daily routine was relatively difficult and, I must admit, a bit tiring.* [PCI_P24; page 19]

Young People's Preferences Regarding Flexible and Self-Directed AI4U Training Use

Young People Demand Autonomy in Daily AI4U Training Use Behavior

From the perspective of some individuals, the temporal postponement of the EMA and EMI (PCIs: 5/16; participants FG1_P3 and FG2_P3) or muting the alarms (participants PCI_P27, FG1_P2, and FG1_P3) should be more flexible. Several young people (PCIs: 7/16; FGs: 7/11) expressed a desire for more self-determination in use behavior, such as having a stored calendar with personalized availability and preferred use times. Others (FGs: 5/11) suggested options for self-selection, for instance, via 2 variants of an EMI or a shortened version that they could choose from.

A few participants (PCIs: 9/16; FG1_P1 and FG3_P4) argued that increased self-determination could improve daily life integration and reduce the pressure, which some individuals experienced during the study period:

> *I believe it would be better for me personally to use it, if I had more control over inputting information myself...But given the frequency with which the app is now, with the EMA almost five times a day, I just don't think that's feasible for me.* [PCI_P23; page 302]

One participant (PCI_P23) remarked that the AI4U training had to fit into her daily routine without altering her habits or making her feel compelled to do so. Another participant emphasized that denying a prompt was always an option, and therefore, she had less of a problem with the app's suggestions:

> *I mean, we always had the option to do the exercises ourselves...You always have the option to simply decline. Then you can still choose your own exercise.* [FG3_P3; page 249]

Young People Seem to Value Control in Content Selection and Decision-Making Processes

A few participants (PCIs: 3/16; FG3_P3) mentioned that they did not want to rely entirely on the app's decision-making processes and expressed some doubts regarding the app's ability to consistently capture their emotions accurately:

> *I think it's quite individual. I also don't believe that, to be honest, it is possible in the long run for [the app] to adapt to everyone, to all [personality] types. I can't quite imagine that right now.* [PCI_P25; page 166]

Consequently, certain participants demanded an active role in influencing the degree of personalization, for instance, through a feedback function on the individual perceived usefulness of the EMI (participants PCI_P15 and PCI_P17; FGs: 3/11). In total, 3 participants from the FGs highlighted the importance of being able to deviate from the app's content suggestions (FGs: 3/11). One participant emphasized the value of the interplay between self-directed use and app suggestions:

> *I would find it quite good if there was perhaps an app that still allows for some personal control despite AI. Because of course, the AI is hugely advantageous as it suggests something and maybe recognizes patterns that one might not have noticed, and providing EMI accordingly, but perhaps an additional option would be good, where you could say, "No, I don't feel like doing this EMI, I would prefer to choose another one." Sometimes you don't really know which exercise is appropriate, and then the AI might have an advantage because it has already collected data.* [FG1_P1; page 124]

Some of our participants (PCIs: 4/16; FGs: 3/11) underscored that meaningful change can only occur through individual initiative. They emphasized a certain degree of willingness to change and openness to self-reflection as prerequisites for long-lasting change in health-promoting routines:

> *I believe that an app can indeed accomplish something like that. However, as a human being, you must have the ability. You have to be able to receive information or something and work with it...The app can definitely [help], if you look at the dashboard and are a reflective person with a lot of self-awareness who can work very well with oneself, then definitely.* [PCI_P12; page 88]

This was also highlighted by another participant, who emphasized that the AI4U training only set prompts, whereas actual processing and change had to be self-motivated:

> *No, I managed to [strengthen my skills and handle my emotions better in everyday life] on my own, but the app mainly helps in getting started with it and...become more aware of your emotions...So, before I can learn to deal with my emotions, I have to perceive them first. And the app helped with that. It didn't help me so much with processing them, but processing mainly happens through the perception and then acceptance of my emotions.* [PCI_P17; page 214]

Another participant stated that he did not want to be controlled by an mHealth app and did not want an app to dictate a journey toward self-optimization:

> *If, for instance, it was an app that somehow had the message, "You have to become better or whatever," then I would have dropped out of the study...I don't want an app to dictate to me that things have to be better or whatever.* [PCI_P11; pages 186-188]

XSL•FO
RenderX

Discussion

Self-Reflection and Regulation of Negative Affect Improved During the AI4U Training

Our qualitative results point toward the positive impact of sampling emotions (via EMA) and real-time mobile interventions based on moment and life context (via EMI) on the mental health of young people. There was broad consensus across our participants that the EMA were helpful for self-reflection on emotions and their respective triggers. Previous qualitative studies have reported similar enhanced self-awareness among young people after engaging in EMA for up to 3 weeks [18,38,50,51]. Participants reported that the EMI embedded in the AI4U training improved emotion regulation and helped relativize negative emotions and unwind from daily challenges. Aligned with this, an existing mHealth app that provides digital cognitive reappraisal strategies to young people has resulted in reduced impulsivity among users [18]. Other mHealth apps have elicited similar positive effects on participants' emotion regulation and reduction in stress or anxiety levels [12]. Our findings suggest that EMA may foster emotional awareness, which, in turn, seems to enhance the subjective effectiveness of EMI in managing and adapting to those emotions.

Young People's Trust in Disclosing Emotions on Digital Devices Is Driven by Anonymity

People's trust is crucial for the willingness to share data and can be shaped by the perceived usefulness, sensitivity, and anonymity of personal health data [69]. As our participants tested the AI4U training as part of a research study and trusted the confidential data processing, they did not report any concerns regarding data security. For adequate trust in a digital intervention, qualitative studies highlight young people's demand for transparent data handling underlying evidence-based approaches, credibility of the developers, and young people's direct involvement through cocreation in the development process [46,70].

However, trust to disclose emotions was also driven by our participants' perception of the mHealth app as nonhuman. According to them, the AI4U training provided a sense of emotional distance and anonymity compared to face-to-face communication and significantly facilitated the unveiling of emotions via EMA. Previous qualitative research has identified that anonymity on digital applications reduces young people's fear of stigma associated with the disclosure of mental health issues [46]. There seems to be no significant difference in young people's willingness to share data about mental health online compared to their willingness to disclose information about alcohol consumption, physical activity, or academic performance [71]. Our participants valued the AI4U training's neutrality, especially when compared to the risk of judgment or potential misuse of data on mental health when interacting with friends, family, or mental health professionals. Our findings align with those of existing studies and suggest that discussing and sharing mental health issues on digital devices from scientific app providers is not a barrier for young people.

Young People's Flexible Engagement With mHealth Apps While Balancing Everyday Commitments

Some participants highlighted the prioritization of personal activities over AI4U training use when feeling well and resilient within their routines. In line with this, previous studies have identified a rather situational use pattern where an mHealth app is used only when immediate needs arise or in preparation for expected challenges [37,52]. In contrast, a few participants avoided using the AI4U training when experiencing extreme negative emotions. A previous study also identified high levels of overall negative affect among young people as negative contributors to their compliance rates with subsequent EMA prompts [72]. Our results suggest that there might be within-participant ambiguities in young people's views on the training's role in daily life. Some of our participants also acknowledged that it probably would have been most beneficial to engage with the mHealth app and follow the prompts during moments of emotional distress or psychological discomfort. Participants recognized that, without the AI4U training, they would not have allocated time for mental health–promoting activities in their daily life and noted that the AI4U training helped them become more aware of the need to take breaks, which they found difficult to implement without the training. In fact, it is tempting to speculate that EMI tailored to current needs, emotional state, and contextual factors may, in part, need to be perceived as disruptive to existing habits and daily life structures to nudge individuals and initiate change in psychological processes and mechanisms. In other words, timely interventions might be essential for achieving the intended health-promoting effects. To address this, future development should further strive to strike the right balance among minimizing perceived interference, optimizing acceptance, and ensuring that interventions are accessible and used during crucial moments when they can have the most impact.

Our findings suggest that young people tended to integrate the AI4U training into their existing daily routines and preferred to approach it in a flexible and nonbinding manner, that is, engaging with EMA or EMI depending on their personal capacity. Our participants perceived the AI4U training more as a "tool" used while maintaining user agency. Nonnegotiable commitments such as work or academic duties were central in their daily lives, and participants reported being rather reluctant to pause any of their ongoing activities to use the AI4U training. Time constraints and lack of alignment with other daily activities have previously been identified as major barriers to the use of mental health services among young people [6]. In line with previous research, young people tended to answer EMA questions rather superficially when being occupied with other tasks [51], and an increased stress level in young people is a predictor of noncompletion of subsequent EMA prompts [72]. According to some of our participants, it was difficult to integrate the AI4U training into their daily routines, and some felt stressed by the frequent alarms. However, the AI4U training already allowed for muting notifications, the interactions were rather brief (approximately 1 minute for EMA), and not every prompt included an EMI. Our participants were informed about the mute function during prestudy briefing sessions. Similarly, our research team emphasized during those briefing sessions

XSL•FO
RenderX

that the training's intention was not to interfere with daily life. Nevertheless, it seems that this did not result in all participants developing a relaxed attitude toward responding to the EMA or EMI or muting the notifications.

In another study, college students received up to 19 EMA per day but did not report feeling burdened by completing them [51]. This study used context- and signal-dependent prompts to avoid interrupting participants during specific levels of smartphone activity, for instance, if the data showed that they were in the process of sending an email. This approach may have contributed to young people feeling less overwhelmed. mHealth apps with gamification elements, a comprehensive backstory narrative, mini games, or performance levels have demonstrated lower attrition rates and enhanced resilience effects in young people compared to less gamified but equally evidence-based mHealth apps [73]. Image-based EMA prompts or prompts via wearable devices could also reduce use burden and enhance compliance among young people [51]. Future mHealth app development requires further participatory co-design to promote a continuous engagement of young people in mental health–promoting approaches.

Healthy young people may have found that their existing routines were effective and already helped them navigate their day-to-day activities, leaving little reason to change them. In contrast, those who are psychologically distressed may find that their coping mechanisms or routines hinder their daily lives. This experience might lead to greater openness toward the necessity of modifying these routines, for instance, through interventions such as EMI. In a previous cluster-randomized controlled trial, the absence of immediate and significant mental health–promoting effects among a nonclinical sample of healthy young people was attributed to their low baseline distress levels [19]. Thus, it is important to note that healthy people might report more disadvantages because they benefit less from the training.

A Balance Between Self-Determination and Guidance in Daily App Use Behavior Stimulates Compliance

Some of our participants suggested that a greater self-directed use behavior could improve daily life integration, and some highlighted the value of an active involvement in the personalization process. In their development process into adulthood, young people pursue increased autonomy and independence, which also influences their mental health help-seeking behavior in daily life [9]. Their desire for shared decision-making approaches is in line with existing research on attitudes of young people toward mHealth apps. Previously, young people have stressed that an app should never replace human judgment and demanded an active role in steering AI decision-making processes or intervention and content selection [55,56,74]. This was introduced as the concept of "personalized personalization" [55]. Interviews with young people on other mHealth apps have shown that not the quantity but, rather, the timing of EMA throughout the day is challenging, and preset times of use would enhance adherence [38]. In the AI4U training, participants were allowed to set their activity times in advance and had the option to reschedule or trigger an EMA or EMI themselves if necessary. However, it seems that this flexibility was not clear to everyone. On the basis of those results, the option to choose a different EMI is now suggested more proactively in newer versions of the AI4U training. In addition, there is an increased emphasis on individual decision-making power in use behavior. It is important to note that, given the MRT design, only half of the EMI assignments in the AI4U training were AI based. Thus, this does not allow us to clearly draw conclusions about the quality of the AI decision-making processes.

A recent study cautions against increased customization by young people themselves given that "participants may express their desire for customizability in an interview situation, but in reality, they may respond better to clearer design, short interactions with the app, and easy access" [37]. This aligns with feedback from several participants, who reported difficulties in maintaining health-promoting routines learned during the study without the app, resulting in less frequent implementation or complete nonimplementation of these practices. In addition, some participants highlighted that excessive freedom of choice can be overwhelming, suggesting that AI can provide valuable support in this context.

The results point toward the important interplay between AI and human decision-making for an optimal user experience among young people. The challenge for future development processes of mHealth apps is to balance self-determination with targeted external guidance in use and decision-making to achieve the highest health-promoting outcome.

Limitations

A major limitation of this study is the homogeneity of our study sample. Most of the interview participants (20/27, 74%) had obtained their high school diploma or were pursuing higher education. Participants in the PCIs and FGs were mainly students from disciplines including medicine, social work, business, and psychology. Thus, the recruitment strategy of this qualitative study reached predominantly young people who already actively engaged with mental or physical health and social behaviors in their everyday lives or academic education. Most students probably experienced higher levels of academic stress, and their strict daily schedules might have biased our results on daily life integration of the AI4U training. For students of such demanding academic fields, a high level of self-reliance is to be expected, which might also have influenced our results on demanded autonomy. Furthermore, we were unable to recruit young people from diverse migration backgrounds. Almost all the participants of this qualitative study (26/27, 96%) were born in Germany. Currently, there is limited research on the effects of mHealth apps on marginalized individuals from lower socioeconomic or with migration backgrounds [12,17,45]. Limited access to mHealth apps and a lack of alignment between digital interventions and the needs of marginalized individuals potentially limits their health-promoting effects and negatively influences compliance rates [75-77]. In addition, we recruited only a small number of participants, and the gender distribution was not evenly balanced. Due to low inclusion of young men in MRT 2, conducting a second FG with male individuals was not possible. Efforts to address data saturation by increasing the number of male participants in the PCIs did not yield the

XSL•FO
RenderX

intended balance. A recent review on mHealth apps indicates a comparable skewness toward more female participants [17]. A potential explanation for this might be that help-seeking behavior is often associated with weakness [6,78]. There is a need for better understanding why particularly young men underuse health promotion services and what it takes to better engage this subgroup. One study, for instance, used a participatory approach to address gender differences in help seeking and mental health needs [79]. By adding gamification elements centering on sports (ie, football), they aimed for a more preference-based development process of an mHealth app particularly targeted at young men.

Some young people perceived the AI4U training as an additional stressor, with some participants highlighting that this stress was intensified by their desire to meet the requirements for adequate study participation. Young people also perceived the use of a second study smartphone for the AI4U training as burdensome. Hence, we cannot ascertain whether the stress stemmed from participation in the study or the app format itself. Due to the MRT study design, participants were unable to determine whether an EMI was based on random or AI-based assignment. Therefore, statements made in the interviews and FGs about the quality of the suggestions for EMI and the AI features must be interpreted with caution. We conducted the interviews and FGs with a subsample of all AI4U participants and only recruited participants for our qualitative study among the participants of MRT 1 and MRT 2. Therefore, generalizability of individual statements to the entire study population across all MRTs is limited.

Conclusions

The approach of this study allowed for an in-depth understanding of young people's experiences and views on the use of an AI-based mHealth app (AI4U training). Participatory approaches throughout the qualitative research process enabled the shared decision-making and empowerment of young people. The major contribution of our study is that, unlike previous qualitative studies that interviewed young people on their attitudes and preferences on mHealth apps or health promotion services [70,74], we elicited insights on real-world user experiences and subjectively experienced mental health–promoting effects. mHealth apps provide young people with an anonymous environment in which they can open up but that simultaneously allows for the maintenance of control over health-promoting actions. According to young people, an AI-based mHealth app can improve emotional resilience, self-awareness, and health-promoting routines in daily life. Our study also highlights some challenges with integrating an mHealth app into the structured routines of healthy young people. Therefore, flexibility and adaptability seem to enhance the reach, effectiveness, and scalability of an mHealth app.

Acknowledgments

Please note that the authors SD and JS were not officially affiliated with the Department of Psychiatry II, University of Ulm and BKH Guenzburg, Guenzburg, Germany as part of an ongoing working contract. Rather, they participated in our study as co-researchers, i.e. study participants. With that, they were an external part of the research team at the Department of Psychiatry II, University of Ulm and BKH Guenzburg, Germany and were compensated per participation in the participatory research workshops. The authors would like to extend their sincere appreciation for Benedikt Pfeiffer and Tanja Brackman, who constantly attended the recurring participatory research workshops. The authors would like to thank their team member and research associate Sonja Schneiderat, who also works as a school social worker and supported the critical analysis of the findings based on her practical experience. They are also thankful for the support of their student research assistants (Marla Kordes and Beatrice Müller), who assisted in the transcription processes and literature research. The authors' sincere thanks especially go to all participants of the problem-centered interviews and focus groups, without whom this study would not have been possible. This work was funded by a living laboratory grant from the Federal Ministry of Science, Research, and Arts of Baden-Württemberg, Germany, to UR, GK, SK, and DD (grant 31–7547.223–7/3/2); a Deutsche Forschungsgemeinschaft (German Research Foundation) Heisenberg professorship to UR (grant 389624707); and a grant from the Hector II Foundation awarded to JF. The sponsors had no role in the collection, analysis, and interpretation of the data; the writing of the report; or the decision to submit the paper for publication.

Authors' Contributions

UR, CR, SK, DD, GK, JF, EW, CG, and SH designed the study. SH and CG conducted the qualitative data collection. SH analyzed the qualitative data. The co-researchers (JS and SD) regularly engaged in participatory research workshops and helped SH prepare data material and validate the results of the qualitative analysis. SH wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
AI4U training components.
[DOCX File , 21 KB - mhealth_v13i1e65106_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Study participant overview.
[DOCX File , 25 KB - mhealth_v13i1e65106_app2.docx]

Multimedia Appendix 3
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 97 KB - mhealth_v13i1e65106_app3.pdf]

Multimedia Appendix 4
AI4U script for the remote focus groups for the co-researchers.
[DOCX File , 18 KB - mhealth_v13i1e65106_app4.docx]

Multimedia Appendix 5
Interview topic guide (problem-centered interviews and focus groups).
[DOCX File , 58 KB - mhealth_v13i1e65106_app5.docx]

Multimedia Appendix 6
Coding categories (problem-centered interviews and focus groups).
[DOCX File , 88 KB - mhealth_v13i1e65106_app6.docx]

References

1. Castelpietra G, Knudsen AK, Agardh EE, Armocida B, Beghi M, Iburg KM, et al. The burden of mental disorders, substance use disorders and self-harm among young people in Europe, 1990-2019: findings from the Global Burden of Disease Study 2019. Lancet Reg Health Eur 2022 May;16:100341 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100341] [Medline: 35392452]
2. Colizzi M, Lasalvia A, Ruggeri M. Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care? Int J Ment Health Syst 2020;14:23 [FREE Full text] [doi: 10.1186/s13033-020-00356-9] [Medline: 32226481]
3. Schlack R, Peerenboom N, Neuperdt L, Junker S, Beyer AK. The effects of mental health problems in childhood and adolescence in young adults: results of the KiGGS cohort. J Health Monit 2021 Dec;6(4):3-19 [FREE Full text] [doi: 10.25646/8863] [Medline: 35146318]
4. Otto C, Reiss F, Voss C, Wüstner A, Meyrose A, Hölling H, et al. Mental health and well-being from childhood to adulthood: design, methods and results of the 11-year follow-up of the BELLA study. Eur Child Adolesc Psychiatry 2021 Oct;30(10):1559-1577 [FREE Full text] [doi: 10.1007/s00787-020-01630-4] [Medline: 32918625]
5. Uhlhaas PJ, Davey CG, Mehta UM, Shah J, Torous J, Allen NB, et al. Towards a youth mental health paradigm: a perspective and roadmap. Mol Psychiatry 2023 Aug;28(8):3171-3181 [FREE Full text] [doi: 10.1038/s41380-023-02202-z] [Medline: 37580524]
6. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021 Feb;30(2):183-211. [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
7. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
8. Stunden C, Zasada J, VanHeerwaarden N, Hollenberg E, Abi-Jaoudé A, Chaim G, et al. Help-seeking behaviors of transition-aged youth for mental health concerns: qualitative study. J Med Internet Res 2020 Oct 05;22(10):e18514 [FREE Full text] [doi: 10.2196/18514] [Medline: 33016882]
9. Ishikawa A, Rickwood D, Bariola E, Bhullar N. Autonomy versus support: self-reliance and help-seeking for mental health problems in young people. Soc Psychiatry Psychiatr Epidemiol 2023 Mar;58(3):489-499 [FREE Full text] [doi: 10.1007/s00127-022-02361-4] [Medline: 36112162]
10. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
11. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. Psychol Psychother 2019 Jun;92(2):277-297. [doi: 10.1111/papt.12222] [Medline: 30924316]
12. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 08;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]

XSL•FO
RenderX

13. Renfrew ME, Morton DP, Morton JK, Hinze JS, Beamish PJ, Przybylko G, et al. A web- and mobile app-based mental health promotion intervention comparing email, short message service, and videoconferencing support for a healthy cohort: randomized comparative study. J Med Internet Res 2020 Jan 06;22(1):e15592 [FREE Full text] [doi: 10.2196/15592] [Medline: 31904578]
14. Bell IH, Thompson A, Valentine L, Adams S, Alvarez-Jimenez M, Nicholas J. Ownership, use of, and interest in digital mental health technologies among clinicians and young people across a spectrum of clinical care needs: cross-sectional survey. JMIR Ment Health 2022 May 11;9(5):e30716 [FREE Full text] [doi: 10.2196/30716] [Medline: 35544295]
15. Peuters C, Maenhout L, Cardon G, De Paepe A, DeSmet A, Lauwerier E, et al. A mobile healthy lifestyle intervention to promote mental health in adolescence: a mixed-methods evaluation. BMC Public Health 2024 Jan 02;24(1):44 [FREE Full text] [doi: 10.1186/s12889-023-17260-9] [Medline: 38166797]
16. Ong WY, Sündermann O. Efficacy of the mental health app "Intellect" to improve body image and self-compassion in young adults: a randomized controlled trial with a 4-week follow-up. JMIR Mhealth Uhealth 2022 Nov 21;10(11):e41800 [FREE Full text] [doi: 10.2196/41800] [Medline: 36409542]
17. Ding X, Wuerth K, Sakakibara B, Schmidt J, Parde N, Holsti L, et al. Understanding mobile health and youth mental health: scoping review. JMIR Mhealth Uhealth 2023 Jun 16;11:e44951 [FREE Full text] [doi: 10.2196/44951] [Medline: 37220197]
18. Dubad M, Elahi F, Marwaha S. The clinical impacts of mobile mood-monitoring in young people with mental health problems: the MeMO study. Front Psychiatry 2021;12:687270 [FREE Full text] [doi: 10.3389/fpsyt.2021.687270] [Medline: 34393850]
19. Kenny R, Fitzgerald A, Segurado R, Dooley B. Is there an app for that? A cluster randomised controlled trial of a mobile app-based mental health intervention. Health Informatics J 2020 Sep;26(3):1538-1559 [FREE Full text] [doi: 10.1177/1460458219884195] [Medline: 31702409]
20. Litke SG, Resnikoff A, Anil A, Montgomery M, Matta R, Huh-Yoo J, et al. Mobile technologies for supporting mental health in youths: scoping review of effectiveness, limitations, and inclusivity. JMIR Ment Health 2023 Aug 23;10:e46949 [FREE Full text] [doi: 10.2196/46949] [Medline: 37610818]
21. Marciano L, Saboor S. Reinventing mental health care in youth through mobile approaches: current status and future steps. Front Psychol 2023;14:1126015 [FREE Full text] [doi: 10.3389/fpsyg.2023.1126015] [Medline: 36968730]
22. Rauschenberg C, Schick A, Hirjak D, Seidler A, Paetzold I, Apfelbacher C, et al. Evidence synthesis of digital interventions to mitigate the negative impact of the COVID-19 pandemic on public mental health: rapid meta-review. J Med Internet Res 2021 Mar 10;23(3):e23365 [FREE Full text] [doi: 10.2196/23365] [Medline: 33606657]
23. Rauschenberg C, Schick A, Goetzl C, Roehr S, Riedel-Heller SG, Koppe G, et al. Social isolation, mental health, and use of digital interventions in youth during the COVID-19 pandemic: a nationally representative survey. Eur Psychiatry 2021 Mar 09;64(1):e20 [FREE Full text] [doi: 10.1192/j.eurpsy.2021.17] [Medline: 33686930]
24. Wolff LL, Rapp M, Mocek A. Critical evaluation of permanently listed psychosocial digital health applications into the directory for reimbursable digital health applications of the BfArM. [Kritische Prüfung der Evidenz dauerhaft aufgenommener psychosozialer digitaler Gesundheitsanwendungen in das Verzeichnis für erstattungsfähige digitale Gesundheitsanwendungen des BfArM]. Psychiatr Prax 2023 Mar;50(2):67-79. [doi: 10.1055/a-1875-3635] [Medline: 36170874]
25. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
26. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry 2018 Jun;17(2):123-132 [FREE Full text] [doi: 10.1002/wps.20513] [Medline: 29856567]
27. Kramer I, Simons CJ, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. World Psychiatry 2014 Feb;13(1):68-77 [FREE Full text] [doi: 10.1002/wps.20090] [Medline: 24497255]
28. Rauschenberg C, Hirjak D, Ganslandt T, Schulte-Strathaus JC, Schick A, Meyer-Lindenberg A, et al. Digital forms of service delivery for personalized crisis resolution and home treatment. [Digitale Versorgungsformen zur Personalisierung der stationsäquivalenten Behandlung]. Nervenarzt 2022 Mar;93(3):279-287 [FREE Full text] [doi: 10.1007/s00115-021-01100-5] [Medline: 33730181]
29. Reininghaus U. Ecological momentary interventions in psychiatry: The momentum for change in daily social context. [Ambulatorische Interventionen in der Psychiatrie: das Momentum für Veränderung im alltäglichen sozialen Kontext]. Psychiatr Prax 2018 Mar;45(2):59-61. [doi: 10.1055/s-0044-101986] [Medline: 29495051]
30. Reininghaus U, Paetzold I, Rauschenberg C, Hirjak D, Banaschewski T, Meyer-Lindenberg A, et al. Effects of a novel, transdiagnostic ecological momentary intervention for prevention, and early intervention of severe mental disorder in youth (EMIcompass): findings from an exploratory randomized controlled trial. Schizophr Bull 2023 May 03;49(3):592-604 [FREE Full text] [doi: 10.1093/schbul/sbac212] [Medline: 36738168]
31. Schulte-Strathaus JC, Rauschenberg C, Baumeister H, Reininghaus U. Ecological momentary interventions in public mental health provision. In: Montag C, Baumeister H, editors. Digital Phenotyping and Mobile Sensing. Cham, Switzerland: Springer; 2023:427-439.

XSL•FO
RenderX

32. Fechtelpeter J, Rauschenberg C, Jalalabadi H, Boecking B, van Amelsvoort T, Reininghaus U, et al. A control theoretic approach to evaluate and inform ecological momentary interventions. Int J Methods Psychiatr Res 2024 Dec;33(4):e70001. [doi: 10.1002/mpr.70001] [Medline: 39436927]
33. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence-empowered mHealth: scoping review. JMIR Mhealth Uhealth 2022 Jun 09;10(6):e35053 [FREE Full text] [doi: 10.2196/35053] [Medline: 35679107]
34. Durstewitz D, Koppe G, Meyer-Lindenberg A. Deep neural networks in psychiatry. Mol Psychiatry 2019 Nov;24(11):1583-1598. [doi: 10.1038/s41380-019-0365-9] [Medline: 30770893]
35. Koppe G, Guloksuz S, Reininghaus U, Durstewitz D. Recurrent neural networks in mobile sampling and intervention. Schizophr Bull 2019 Mar 07;45(2):272-276 [FREE Full text] [doi: 10.1093/schbul/sby171] [Medline: 30496527]
36. Milne-Ives M, Selby E, Inkster B, Lam C, Meinert E. Artificial intelligence and machine learning in mobile apps for mental health: a scoping review. PLOS Digit Health 2022 Aug;1(8):e0000079 [FREE Full text] [doi: 10.1371/journal.pdig.0000079] [Medline: 36812623]
37. Ribanszki R, Saez Fonseca JA, Barnby JM, Jano K, Osmani F, Almasi S, et al. Preferences for digital smartphone mental health apps among adolescents: qualitative interview study. JMIR Form Res 2021 Aug 27;5(8):e14004 [FREE Full text] [doi: 10.2196/14004] [Medline: 34128814]
38. Acorda D, Businelle M, Santa Maria D. Perceived impacts, acceptability, and recommendations for ecological momentary assessment among youth experiencing homelessness: qualitative study. JMIR Form Res 2021 Apr 06;5(4):e21638 [FREE Full text] [doi: 10.2196/21638] [Medline: 33821805]
39. Harting J, Kruithof K, Ruijter L, Stronks K. Participatory research in health promotion: a critical review and illustration of rationales. Health Promot Int 2022 Jun 23;37(Supplement_2):ii7-i20 [FREE Full text] [doi: 10.1093/heapro/daac016] [Medline: 35748283]
40. Bergold J, Thomas S. Participatory research methods: a methodological approach in motion. [Partizipative Forschungsmethoden: ein methodischer Ansatz in Bewegung]. Forum Qual Soc Res 2012;13(1). [doi: 10.17169/fqs-13.1.1801]
41. Stock C, Dias S, Dietrich T, Frahsa A, Keygnaert I. Editorial: how can we co-create solutions in health promotion with users and stakeholders? Front Public Health 2021;9:773907 [FREE Full text] [doi: 10.3389/fpubh.2021.773907] [Medline: 34957026]
42. Bemme D, Béhague D. Theorising the social in mental health research and action: a call for more inclusivity and accountability. Soc Psychiatry Psychiatr Epidemiol 2024 Mar;59(3):403-408. [doi: 10.1007/s00127-024-02632-2] [Medline: 38407626]
43. Fisher HL. The public mental health revolution must privilege lived experience voices and create alliances with affected communities. World Psychiatry 2024 Feb;23(1):2-3 [FREE Full text] [doi: 10.1002/wps.21149] [Medline: 38214647]
44. Reininghaus U, Reinhold AS, Priebe S, Rauschenberg C, Fleck L, Schick A, et al. Toward equitable interventions in public mental health: a review. JAMA Psychiatry 2024 Dec 01;81(12):1270-1275. [doi: 10.1001/jamapsychiatry.2024.3206] [Medline: 39412787]
45. Piers R, Williams JM, Sharpe H. Review: can digital mental health interventions bridge the 'digital divide' for socioeconomically and digitally marginalised youth? A systematic review. Child Adolesc Ment Health 2023 Feb;28(1):90-104. [doi: 10.1111/camh.12620] [Medline: 36510375]
46. Ito-Jaeger S, Perez Vallejos E, Logathasan S, Curran T, Crawford P. Young people's trust in cocreated web-based resources to promote mental health literacy: focus group study. JMIR Ment Health 2023 Jan 09;10:e38346 [FREE Full text] [doi: 10.2196/38346] [Medline: 36622752]
47. Hafner J, Schönfeld S, Tokgöz P, Hrynyschyn R, Schlubach A, Dockweiler C. Acceptance of digital health interventions in light to moderate depressive disorders - a qualitative analysis from the perspective of affected persons, relatives and health professionals. Akzeptanz von digitalen gesundheitstechnologien bei leichter bis mittelgradiger depression – eine qualitative analyse aus der perspektive von betroffenen, angehörigen und leistungserbringern. Psychiatr Prax 2022 Sep;49(6):313-321. [doi: 10.1055/a-1472-5388] [Medline: 34015851]
48. Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: a qualitative study of patient and clinician experiences. Int J Bipolar Disord 2020 Dec 01;8(1):38 [FREE Full text] [doi: 10.1186/s40345-020-00201-5] [Medline: 33258015]
49. Folkersma W, Veerman V, Ornée DA, Oldehinkel AJ, Alma MA, Bastiaansen JA. Patients' experience of an ecological momentary intervention involving self-monitoring and personalized feedback for depression. Internet Interv 2021 Dec;26:100436 [FREE Full text] [doi: 10.1016/j.invent.2021.100436] [Medline: 34430220]
50. Swendeman D, Sumstine S, Brink A, Mindry D, Medich M, Russell M. Smartphone self-monitoring by young adolescents and parents to assess and improve family functioning: qualitative feasibility study. JMIR Form Res 2020 Jun 23;4(6):e15777 [FREE Full text] [doi: 10.2196/15777] [Medline: 32574148]
51. Chan L, Swain VD, Kelley C, de Barbaro K, Abowd GD, Wilcox L. Students' experiences with ecological momentary assessment tools to report on emotional well-being. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-20. [doi: 10.1145/3191735]

XSL•FO
RenderX

52. Xie Q, Dyer RL, Lam SU, Frye C, Dahl CJ, Quanbeck A, et al. Understanding the implementation of informal meditation practice in a smartphone-based intervention: a qualitative analysis. Mindfulness (N Y) 2024 Feb;15(2):479-490. [doi: 10.1007/s12671-024-02304-x] [Medline: 39070019]
53. Zipfel N, Horreh B, Hulshof CTJ, de Boer AG, van der Burg-Vermeulen SJ. The relationship between the living lab approach and successful implementation of healthcare innovations: an integrative review. BMJ Open 2022 Jun 28;12(6):e058630. [doi: 10.1136/bmjopen-2021-058630] [Medline: 35768105]
54. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland: Springer; 2018.
55. Götzl C, Hiller S, Rauschenberg C, Schick A, Fechtelpeter J, Fischer Abaigar U, et al. Artificial intelligence-informed mobile mental health apps for young people: a mixed-methods approach on users' and stakeholders' perspectives. Child Adolesc Psychiatry Ment Health 2022 Nov 17;16(1):86 [FREE Full text] [doi: 10.1186/s13034-022-00522-6] [Medline: 36397097]
56. Hiller S, Rauschenberg C, Götzl C, Fechtelpeter J, Koppe G, Durstewitz D, et al. Gemeinsam gestalten?! Wie junge menschen, praxis und wissenschaft bei der entwicklung einer smartphone-app zusammenfinden. unsere jugend 2023 Feb 01;75(2):77-91 [FREE Full text] [doi: 10.2378/uj2023.art11d]
57. Rauschenberg C, Schirmbeck F, Fechtelpeter J, Wierzba E, Hiller S, Kahr K, et al. Effects of AI4U training, a machine learning-based, adaptive ecological momentary intervention for personalized mental health promotion in youth: findings from a micro-randomized trial. PsyArXiv Preprints. Preprint posted online on October 14, 2024 2025 [FREE Full text]
58. Gramespacher E, Albert K, Hunger I, Lüsebrink I. Research workshops - basis for qualitative research. [Forschungswerkstätten - Basis für qualitative Forschung]. Leipziger Sportwissenschaftliche Beiträge 2009;50(1):98-119 [FREE Full text]
59. Reichertz J. Gemeinsam Interpretieren: Die Gruppeninterpretation als Kommunikativer Prozess. Berlin, Germany: Springer VS; 2013.
60. Schaefer I, Bär G, Contributors of the Research Project ElfE. The analysis of qualitative data with peer researchers: an example from a participatory health research project. [Die Auswertung qualitativer Daten mit Peerforschenden: Ein Anwendungsbeispiel aus der partizipativen Gesundheitsforschung]. Forum Qual Soc Res 2019;20(3) [FREE Full text] [doi: 10.17169/fqs-20.3.3350]
61. Myin-Germeys I, Klippel A, Steinhart H, Reininghaus U. Ecological momentary interventions in psychiatry. Curr Opin Psychiatry 2016 Jul;29(4):258-263. [doi: 10.1097/YCO.0000000000000255] [Medline: 27153125]
62. Hess F, Monfared Z, Brenner M, Durstewitz D. Generalized teacher forcing for learning chaotic dynamics. arXiv Preprint posted online on June 7, 2023 [FREE Full text] [doi: 10.48550/arXiv.2306.04406]
63. Mikhaeil JM, Monfared Z, Durstewitz D. On the difficulty of learning chaotic dynamics with RNNs. arXiv Preprint posted online on October 14, 2021 [FREE Full text] [doi: 10.48550/arXiv.2110.07238]
64. Brenner M, Hess F, Koppe G, Durstewitz D. Integrating multimodal data for joint generative modeling of complex dynamics. In: Proceedings of the 41st International Conference on Machine Learning. 2024 Presented at: ICML 2024; July 21-27, 2024; Vienna, Austria. [doi: 10.48550/arXiv.2212.07892]
65. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
66. Kuckartz U. Qualitative Text Analysis: A Guide to Methods, Practice & Using Software. Thousand Oaks, CA: SAGE Publications; 2014.
67. Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Weinheim, Germany: Beltz Juventa; 2022.
68. VERBI software. MAXQDA. URL: https://www.maxqda.com/about [accessed 2025-04-10]
69. Helou S, Abou-Khalil V, El Helou E, Kiyono K. Factors related to personal health data sharing: data usefulness, sensitivity and anonymity. Stud Health Technol Inform 2021 May 27;281:1051-1055. [doi: 10.3233/SHTI210345] [Medline: 34042839]
70. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I, NeurOx Young People's Advisory Group. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. Biomed Inform Insights 2019;11:1178222619829083 [FREE Full text] [doi: 10.1177/1178222619829083] [Medline: 30858710]
71. Herbert C, Marschin V, Erb B, Meißner D, Aufheimer M, Bösch C. Are you willing to self-disclose for science? Effects of privacy awareness and trust in privacy on self-disclosure of personal and health data in online scientific studies-an experimental study. Front Big Data 2021;4:763196 [FREE Full text] [doi: 10.3389/fdata.2021.763196] [Medline: 35005619]
72. Murray AL, Brown R, Zhu X, Speyer LG, Yang Y, Xiao Z, et al. Prompt-level predictors of compliance in an ecological momentary assessment study of young adults' mental health. J Affect Disord 2023 Feb 01;322:125-131 [FREE Full text] [doi: 10.1016/j.jad.2022.11.014] [Medline: 36372127]
73. Litvin S, Saunders R, Jefferies P, Seely H, Pössel P, Lüttke S. The impact of a gamified mobile mental health app (eQuoo) on resilience and mental health in a student population: large-scale randomized controlled trial. JMIR Ment Health 2023 Jul 21;10:e47285 [FREE Full text] [doi: 10.2196/47285] [Medline: 37477955]

XSL•FO
RenderX

74. Garrido S, Cheers D, Boydell K, Nguyen QV, Schubert E, Dunne L, et al. Young people's response to six smartphone apps for anxiety and depression: focus group study. JMIR Ment Health 2019 Oct 02;6(10):e14385 [FREE Full text] [doi: 10.2196/14385] [Medline: 31579023]
75. Rubeis G, Ketteler D. Who benefits from the app? Internet- and mobile-based interventions (IMIs) and the tension between autonomy and patient well-being. Wem nützt die app? Internet- und mobilgestützte interventionen (IMIs) im spannungsfeld von autonomie und patientenwohl. Psychother Psychosom Med Psychol 2020 Nov;70(11):467-474. [doi: 10.1055/a-1104-5459] [Medline: 32069513]
76. Müller LR, Gossmann K, Schmid RF, Rosner R, Unterhitzenberger J. A pilot study on ecological momentary assessment in asylum-seeking children and adolescents resettled to Germany: investigating compliance, post-migration factors, and the relation between daily mood, sleep patterns, and mental health. PLoS One 2021;16(2):e0246069 [FREE Full text] [doi: 10.1371/journal.pone.0246069] [Medline: 33524043]
77. Köhnen M, Dreier M, Freuck J, Härter M, Dirmaier J. Acceptance and use of the e-mental health website - www.psychenet.de. Akzeptanz und nutzung einer website mit gesundheitsinformationen zu psychischen erkrankungen? www.psychenet.de. Psychiatr Prax 2022 May;49(4):205-212. [doi: 10.1055/a-1502-0961] [Medline: 34102698]
78. Pattyn E, Verhaeghe M, Bracke P. The gender gap in mental health service use. Soc Psychiatry Psychiatr Epidemiol 2015 Jul;50(7):1089-1095. [doi: 10.1007/s00127-015-1038-x] [Medline: 25788391]
79. Cheng VW, Davenport TA, Johnson D, Vella K, Mitchell J, Hickie IB. An app that incorporates gamification, mini-games, and social connection to improve men's mental health and well-being (MindMax): participatory design process. JMIR Ment Health 2018 Nov 19;5(4):e11068 [FREE Full text] [doi: 10.2196/11068] [Medline: 30455165]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research
EMA: ecological momentary assessments
EMI: ecological momentary interventions
FG: focus group
mHealth: mobile health
MRT: microrandomized trial
PCI: problem-centered interview

Edited by L Buis; submitted 06.08.24; peer-reviewed by L McCann, J Schobel, PH Liao; comments to author 17.02.25; revised version received 21.02.25; accepted 11.03.25; published 29.04.25.

Please cite as:
Hiller S, Götzl C, Rauschenberg C, Fechtelpeter J, Koppe G, Wierzba E, Sauter J, Dietrich S, Durstewitz D, Reininghaus U, Krumm S
Health-Promoting Effects and Everyday Experiences With a Mental Health App Using Ecological Momentary Assessments and AI-Based Ecological Momentary Interventions Among Young People: Qualitative Interview and Focus Group Study
JMIR Mhealth Uhealth 2025;13:e65106
URL: https://mhealth.jmir.org/2025/1/e65106
doi:10.2196/65106
PMID:

©Selina Hiller, Christian Götzl, Christian Rauschenberg, Janik Fechtelpeter, Georgia Koppe, Eva Wierzba, Julia Sauter, Sina Dietrich, Daniel Durstewitz, Ulrich Reininghaus, Silvia Krumm. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Longitudinal Ecological Momentary Assessment Completion: Results From 12 Months of Burst Sampling in the TIME Study

Tyler Prochnow[1], PhD; Wei-Lin Wang[2], PhD; Shirlene Wang[2], PhD; Jixin Li[3], MA; Alexander J Rothman[4], PhD; Stephen S Intille[3], PhD; Donald Hedeker[5], PhD; Genevieve F Dunton[2], PhD

[1]School of Public Health, Texas A&M University, College Station, TX, United States

[2]Population and Public Health Sciences & Psychology, University of Southern California, Los Angeles, CA, United States

[3]Khoury College of Computer Sciences and the Bouvé College of Health Sciences, Northeastern University, Boston, MA, United States

[4]College of Liberal Arts, University of Minnesota, Minneapolis, MN, United States

[5]Department of Public Health Sciences, University of Chicago, Chicago, IL, United States

Corresponding Author:
Tyler Prochnow, PhD
School of Public Health
Texas A&M University
711 TAMU
College Station, TX, 77843
United States
Phone: 1 2629450275
Email: tprochnow@tamu.edu

Abstract

Background: Ecological momentary assessment (EMA) is a valuable method for capturing real-time data on behaviors and experiences in naturalistic settings. However, maintaining participant engagement in longitudinal (ie, multiburst) EMA studies remains challenging, particularly when collecting intensive data over extended periods. Understanding factors affecting completion rates is essential for designing more effective EMA protocols and interpreting results accurately.

Objective: This study investigated factors influencing EMA completion rates in a 12-month intensive longitudinal study among young adults in the United States, examining both time-varying factors and stable individual characteristics.

Methods: Young adults (N=246, ages 18-29 years) participated in the Temporal Influences on Movement and Exercise (TIME) study, responding to smartphone-based EMA prompts during biweekly measurement bursts (4-day periods of intensive sampling), with continuous passive data collection via smartwatches. Each burst included signal-contingent prompts delivered approximately once per hour during waking hours, resulting in an average of 12.1 (SD 1.3) prompts per day. Multilevel logistic regression models examined the effects of time-varying temporal factors (time of day, day of week, season, and time in study), contextual factors (phone screen status, phone usage, and location), behavioral factors (sleep duration, physical activity levels, and travel status), and psychological factors (momentary affect and stress) on prompt completion. Models also included time-invariant demographic characteristics (sex, race, ethnicity, education, and employment) and tested interactions between time in study and other predictors.

Results: Mean completion rate was 77% (SD 13%). Hispanic participants showed lower odds of completion compared to non-Hispanic participants (odds ratio [OR] 0.79, 95% CI 0.63-0.99; *P*=.04) and employed participants were less likely to complete prompts than unemployed participants (OR 0.75, 95% CI 0.61-0.92; *P*<.01). Having the phone screen on at prompt delivery substantially increased completion odds (OR 3.39, 95% CI 2.81-4.09; *P*<.001), while being away from home reduced completion likelihood, with particularly low odds when at sports facilities (OR 0.58, 95% CI 0.47-0.74; *P*<.001) or restaurants and shops (OR 0.61, 95% CI 0.51-0.72; *P*<.001). Short sleep duration the previous night (OR 0.92, 95% CI 0.87-0.99; *P*=.02) and traveling status (OR 0.78, 95% CI 0.75-0.82; *P*<.001) were associated with lower completion odds. Higher momentary stress levels predicted lower completion of subsequent prompts (OR 0.85, 95% CI 0.78-0.93; *P*<.001). Completion odds declined over the 12-month study (OR 0.95, 95% CI 0.94-0.96; *P*<.001), with significant interactions between time in study and various predictors, indicating changing patterns of engagement over time.

XSL•FO
RenderX

Conclusions: Findings highlight the dynamic nature of EMA engagement in longitudinal multiburst studies and underscore the importance of considering time-varying and time-invariant factors in study design and analysis. This study provides valuable insights for researchers designing intensive longitudinal studies in behavioral science and digital health. Potential strategies for optimizing EMA protocols could include tailoring prompt schedules to individual contexts and developing adaptive sampling techniques.

(JMIR Mhealth Uhealth 2025;13:e67117) doi:10.2196/67117

KEYWORDS

ecological momentary assessment; longitudinal study; compliance; mobile health; young adults

Introduction

Traditional cross-sectional and retrospective self-report measures are prone to substantial limitations in studying health behaviors and psychological processes [1]. Cross-sectional designs only provide a snapshot, lacking insight into within-person variations over time and context [2,3]. Retrospective surveys require participants to summarize or recall experiences over lengthy periods, introducing recall biases that distort reports [2,3]. For example, physical activity levels likely vary within days and across days of the week in systematic ways obscured in generalized retrospective reports [3,4]. Aggregated retrospective summaries also fail to capture specific within-day or day-to-day fluctuations in behaviors, feelings, thoughts, and exposures that may be highly predictive of outcomes [3]. Traditional designs lacking temporal specificity severely limit understanding of the microtemporal relationships between psychological factors and behavioral choices hypothesized by many health behavior theories and models [1-3,5]. Intensive longitudinal designs using ecological momentary assessment (EMA) are needed to overcome these limitations and capture predictive within-day and day-to-day variations [1-3].

EMA overcomes many limitations of traditional designs by using repeated natural sampling of current behaviors, experiences, and contexts [1-3]. EMA prompts participants to provide real-time self-reports multiple times per day in natural environments, reducing recall biases [1-3]. Capturing microtemporal fluctuations via intensive longitudinal EMA enables examination of within-day and day-to-day variations in behaviors and putative causal factors, unavailable in traditional designs [1-3]. For example, EMA allows modeling of physical activity as a function of rapidly changing variables such as location, affect, and stress measured at the same momentary time scale [6-9]. EMA's temporal precision permits study of directionality and time-varying effect magnitudes [10]. EMA's ecological validity also facilitates idiographic analysis of dynamic relationships among variables unfolding uniquely for each person [10,11]. Thus, intensive longitudinal EMA provides new opportunities to understand microtemporal influences on health behaviors.

However, EMA is prone to missing data due to noncompletion with intensive sampling protocols [12,13]. This can bias results if missing data are systematic [12,13]. For example, individuals may be less likely to comply when under stress or engaging in unhealthy behaviors, distorting conclusions [12,13]. It is critical to maximize completion, which remains a challenge in long-term studies [14]. Completion rates vary widely in EMA research, ranging from 42% to 99% with a mean of 82% [12]. Lower completion is associated with minority ethnicity, low socioeconomic status, and mental health symptoms [12]. Additionally, time-varying factors such as affect, social context, and time of day have been found to predict EMA compliance [15]. Completion also decays over time, with gradual drop-offs or sudden declines [13]. This complicates separating true change from artifactual patterns. It is important to note that most EMA studies are relatively short-term, typically lasting from a few days to several weeks, with some extending to a few months [12]. Longer-term studies, particularly those extending beyond 6 months, are relatively rare in EMA literature. This limited duration of typical EMA studies restricts our understanding of long-term patterns in compliance and the sustainability of intensive longitudinal designs. Systematically missed EMA prompts reduce statistical power and can preclude multilevel modeling approaches common for EMA data [1,12,14,16]. Thus, optimizing and tracking completion is essential for valid EMA studies. Understanding factors influencing completion is critical for interpreting findings and guiding effective retention strategies. Further, most prior research on EMA completion has focused on single measurement burst or limited measurement burst designs [1,3,12]. Consequently, there is a significant gap in our understanding of EMA compliance in multimeasurement burst designs with frequent bursts, particularly regarding changes in completion rates over extended periods and interactions between time-varying and time-invariant factors across time [1,3,12]. This lack of knowledge presents a critical challenge for researchers aiming to implement long-term, intensive EMA protocols, as it limits our ability to anticipate and mitigate compliance issues in such designs.

The Temporal Influences on Movement and Exercise (TIME) study combines 12 months of smartwatch data collection and EMA in young adults [16]. Young adulthood is a high-risk period for declining physical activity and deteriorating sleep patterns [17,18]. The TIME study's intensive EMA protocol also enables examining how psychological and contextual variables relate to completion. Specifically, the length of follow-up (across 12 months), in combination with passive data collection (accelerometer data collection), and frequent intensive EMA "bursts" represent advances in EMA research. For example, the study can evaluate whether EMA completion rates decay more steeply when participants report higher stress levels or more negative affect. Additionally, it can examine if factors such as sleep quality, physical activity levels, or social interactions interact with time to predict changes in compliance. Findings from these analyses may point to the need for adaptive EMA protocols that tailor prompt frequency, timing, or content

XSL•FO
RenderX

based on detected psychological states, behaviors, or contextual factors, potentially improving long-term engagement and data quality. The 12-month duration also permits evaluating long-term patterns such as seasonal variations. Thus, the TIME study provides a unique opportunity to elucidate participation and engagement factors important for future intensive longitudinal research.

Therefore, this study aims to evaluate the TIME study's EMA completion rates and predictors. Specifically, the purpose of this study was to examine whether time-varying temporal, contextual, behavioral, and psychological factors (eg, time of day, day of week, screen status, location, days enrolled in study, time elapsed from last active phone use, sleep duration the prior night, current day activity level, current day average stress, previous hour physical activity, previous hour affect state, and previous hour stress level) and time-invariant personal characteristics (eg, sex at birth, race, ethnicity, education, and employment status) were associated with EMA completion rates. Investigating these patterns through an unprecedented, year-long intensive protocol provides critical methodological insights into the feasibility, sustainability, and data quality of long-term EMA studies. These findings can guide interpretations of results from longitudinal multiburst EMA studies, as systematically missed EMA prompts may distort conclusions by underrepresenting critical moments or obscuring true temporal relationships between variables. Identifying factors linked to completion can inform refinements to future EMA protocols and retention strategies to enhance data quality and representativeness. By understanding predictors of nonresponse, researchers can develop targeted strategies to reduce systematic missingness, such as implementing adaptive sampling techniques or personalized engagement approaches. These refinements aim to preserve statistical power by maximizing response rates, particularly during critical periods or states prone to missingness, ultimately providing more valid and reliable results for understanding real-time health behaviors and experiences. This study represents an important step in elucidating the sustainability and validity of long-term intensive EMA for studying health behaviors, offering empirical evidence to support best practices for designing and implementing longitudinal multiburst EMA studies that balance data quality with participant burden and engagement.

Methods

Design

The TIME study used a 12-month intensive longitudinal multiburst design using personal smartphones to continuously monitor young adults' health behaviors and associated psychological processes. As detailed in the study protocol [16], participants serve as their own controls in a within-subject case-crossover framework, capturing dynamic fluctuations in behaviors and predictors over time and context. Passive sensing via smartwatch and phone sensors provides continuous measurement of activity behaviors and contextual factors such as time and location hypothesized to reactively influence those behaviors [16]. Intermittently prompted EMAs on smartphones gather self-reports on reflective cognitive factors such as intentions and self-control. This innovative combination of continuous passive monitoring and bursting EMA surveys provides rich temporal data on factors influencing physical activity, sleep, and sedentary time. The extensive 12-month duration enabled studying maintenance as well as adoption of health behaviors.

Ethical Considerations

The study was approved by the institutional review board at the University of Southern California (USC; HS-18-00605). The study was performed in accordance with the ethical standards as laid down in the Declaration of Helsinki and its later amendments. All participants provided informed consent to have their deidentified data published in journals. Participants were compensated up to US $100 for each 4-week period (US $1260 total). This included US $10 for each EMA burst period they completed, with at least 8 prompts per day. In addition, if a participant answered more than 11 EMA prompts on a given burst period day, the participant receives a US $5 bonus per day.

Participants

The study recruited young adults across the United States (N=246). Inclusion and exclusion criteria were assessed by self-report during the screening process. Inclusion criteria for the study were as follows: (1) aged 18-29 years living in the United States, (2) intention to engage in recommended levels of moderate to vigorous physical activity (≥150 minutes/week moderate or ≥75 minutes/week vigorous intensity) within the next 12 months, (3) use an Android-based smartphone as their only primary personal mobile device with no intention to switch to a non-Android phone, (4) able to speak and read English, and (5) plan to reside in a home with Wi-Fi connectivity during the study period. Exclusion criteria were (1) physical or cognitive disabilities that prevent participation; (2) health issues that limit physical activity; (3) any diagnosed sleep disorders; (4) unable to wear a smartwatch or answer EMA surveys at home, work, school, or another location where a substantial amount of time is spent; (5) spends more than 3 hours/day on a typical weekday or weekend day driving; (6) owns an Android phone version 6.0 (or older), or if the app will not function on the phone due to other technical issues; (7) currently owns and wears a smartwatch; (8) uses a pay-as-you-go data plan or data plan with less than 2 gigabytes of data per month; or (9) currently pregnant. Participants were recruited regardless of baseline activity level. A variety of recruitment methods were used to recruit socioeconomically and racially or ethnically diverse young adults, including (1) sending emails to individuals enrolled in a USC longitudinal cohort study of young adults, (2) referrals from existing participants (word of mouth), and (3) contacting participants identified using ResearchMatch, a national health volunteer registry [19].

Of 246 participants fully enrolled in the TIME study, 3 were excluded due to missing key demographic information, specifically sex at birth and age, reducing the subject-level sample size to 243. Within this cohort, 29 participants were further removed from the analysis due to missing phone data, including screen state or phone usage, thereby decreasing the sample size to 214. Additionally, missing accelerometer data

XSL•FO
RenderX

from the watch led to the exclusion of 5 participants, culminating in a sample of 209. Lastly, missing EMA data necessitated the removal of 2 participants, resulting in a final analytical sample of 207. See Multimedia Appendix 1 for a more detailed overview of participant exclusions.

Protocol

Data were collected between March 2020 and August 2022. Due to health and safety concerns arising from the COVID-19 pandemic, all study procedures were conducted remotely. Potential participants filled out a web-based interest form to screen eligibility (n=1202). For eligible and interested potential participants, a videoconference orientation and consent session over Zoom (Zoom Communications) was then scheduled (n=645). This session involved reviewing all parts of the study, obtaining informed consent, and downloading the custom TIME study smartphone app onto the participant's smartphone (n=332). During the orientation session, participants received instructions on how to use the study app to complete EMA surveys. During the following week, individuals participated in their first 4-day EMA measurement burst period (further described below), during which the TIME app triggered surveys once per hour during the participant's waking hours. Participants who successfully completed at least 8 surveys per day during this first EMA measurement burst period were fully enrolled in the study and mailed a smartwatch (N=246). If completion was below 8 surveys/day for the first measurement burst period, participants were unenrolled from the study. Among the fully enrolled cohort (N=246), 228 participants remained active in the study during the first 3 months. By month 6, the number of participants had decreased to 186, and only 123 participants completed the full-year study. The overall attrition rate was 50%.

EMA data were collected using the custom TIME app developed for Android smartwatches and smartphones. While more information about the app can be found in the study protocol paper [16], the app consists of an auditory prompting with a smartphone notification, followed by a welcome screen, followed by EMA items with one item per screen, and lastly a thank you screen. The app was downloaded directly to a participant's personal Android phone from the Google Play Store but was only accessible to authorized study participants. Once the participant received the smartwatch by mail, the TIME app was downloaded to the watch paired with the smartphone. EMA surveys were prompted in four ways: (1) daily sleep-wake time surveys to tailor prompt timing to participants' schedules, (2) end-of-day surveys each evening, (3) 4-day measurement burst periods every 2 weeks with signal-contingent prompting approximately every hour, and (4) context-sensitive surveys based on passive sensing of location and activity. Participants also completed web-based questionnaires at baseline, 6 months, and 12 months, and an exit interview. This study focused on the completion of the 4-day measurement burst EMA prompts.

EMA measurement bursts were scheduled every 2 weeks, resulting in up to 26 burst periods throughout the year-long study, totaling 104 days of intensive data collection per participant. The burst periods were randomly scheduled but structured to always include 2 weekdays and 2 weekend days, with a minimum of 7 days between each burst. Each burst encompassed 4 consecutive days of data collection, with signal-contingent prompts occurring multiple times daily. These prompts were randomly triggered approximately once per hour during participants' waking hours, specifically between 10 and 50 minutes past each hour to prevent consecutive prompts from occurring too close together. Participants received a notification via the TIME app 1 day before each burst period was set to begin. This notification provided participants with a one-time option to delay the start of the burst by 2 days if needed.

To encourage ongoing participation and attentiveness, the researchers implemented several engagement strategies, including (1) texting participants when their EMA completion declined, (2) providing financial incentives based EMA completion, (3) allowing participants with high completion (ie, wearing the smartwatch for more than 23 hours/day on 24 days/month, answering more than 24 of the end-of-day EMA smartphone prompts/month, answering more than 11 EMA smartphone prompts/day during measurement bursts, and answering more than 50% of the micro-EMA questions prompted on the smartwatch) to keep their smartwatch at the end of the study, (4) sending quarterly participant newsletters, and (5) providing a report-back of individual data to participants who completed the study. Participants were oriented to the study protocol and given a number and instructed to text the study staff with any questions, concerns, or technical issues. On a weekly basis, staff review data uploaded to the study server and contact participants by email or SMS text messages, which is another avenue to address technical issues. After completing each EMA prompt, participants were shown a humorous and lighthearted "thank you" message. A bank of 250 unique messages was used to maintain novelty throughout the study. On a weekly basis, staff review data uploaded to the study server and contact participants by email or SMS text message in the case of missing data to encourage compliance and address technical issues. Additionally, 20% of the EMA prompts included attention check questions. These questions were designed to be both entertaining and unambiguous, allowing researchers to assess whether participants were paying attention and responding thoughtfully to the EMA questions. More information regarding the attention check questions as well as a qualitative evaluation of participant engagement can be found elsewhere [20].

Activity Monitoring

Raw accelerometer data (using the embedded triaxial accelerometer) were collected on the phone and watch at ~50 Hz continuously. After data collection, the raw accelerometer data from the smartwatch were processed using the monitor-independent movement summary (MIMS) unit algorithm [21]. The MIMS-unit algorithm computes a summary of total movement within an epoch and has been used to identify percentile cutoffs for physical activity among young adults at a population scale [22]. MIMS-unit values were calculated for 1-second epochs for all the raw acceleration data collected from the wrists.

XSL•FO
RenderX

Location Sensing

Location data (latitude and longitude) were gathered once per minute using the built-in GPS features on the phone. By the end of each day, the location data were converted to location clusters, which were then used for self-reported location labeling. First, at midnight, the app clustered the entire day's location data using the DBSCAN (density-based spatial clustering of applications with noise) algorithm [23]. Then, for each day, the clustered points were combined with the previous day's cluster points and were clustered again using DBSCAN to generate location clusters for the participant, but this time considering all clusters identified since the start date. A cluster was formed when there were at least 5 stay points at that location; a stay point was defined as spending at least 5 minutes continuously in one 10 m radius (4th decimal precision for longitude/latitude). The global clusters were saved as polygons of bounding points (using the concave hull [24]).

In real time, the app also checked if the current location was within an already identified cluster for at least 3 minutes; if this condition was satisfied, the phone presented a location question to gather semantic labels. If participants answered this survey, then the label was saved for the cluster. The question was presented again for this cluster until the labeled cluster was *confirmed* (when 80%+ of the labels for that location were the same, and at least 4 labels were reported). Once the label was confirmed for a cluster, no location question was prompted for this cluster again until an additional 60 days had passed, after which the question was presented again. If the location was again confirmed, another 60 days would pass without questions about the cluster; otherwise, more questions were asked until a label was again confirmed.

Measures

Table 1 provides an overview of the measures used for analysis in this study, divided by the type of variable.

Table 1. Variables used in the analysis of ecological momentary assessment prompt completion.

Type	Variable	Collection method
Time-varying temporal	• Time of the day • Type of day (weekday/weekend) • Months enrolled in the study • Season (Reference=Summer)	• EMA[a]
Time-varying contextual	• Active phone screen state (Reference=Phone not actively being used at time of prompt) • Phone usage (time since last phone use) • Physical location	• EMA
Time-varying behavioral	• Sleep duration the prior night • Previous 10-minute physical activity (as an ordinal and continuous variable) • Daily status (eg, sick or traveling)	• Accelerometer • EMA
Time-varying psychological	• Affect state on previous prompt • Stress level on previous prompt	• EMA
Time-invariant demographics	• Sex at birth • Age • Race • Ethnicity • Education • Employment status	• Questionnaire

[a]EMA: ecological momentary assessment.

EMA Completion

The primary outcome of interest is prompt completion (ie, 1=completed and 0=not completed) as defined by responding and completing the full survey within the 10 minutes of the initial prompt. If a response is not provided, up to 2 reprompts will be provided at 5-minute intervals. After 10 minutes, the EMA survey becomes inaccessible and is recorded as not completed. The completion rate was measured at the prompt level and represents the availability of data.

Time-Varying Factors

Temporal

Temporal covariates were measured to examine associations with prompt completion.

- Time of the Day:
  Time of the day was grouped into 4 categories: "morning" (8 AM to 12 noon), "afternoon" (12 noon to 4 PM), "evening" (4 PM to 8 PM), and "night" (8 PM to midnight). Time of the day was then dummy coded with "morning" as the reference category.
- Type of Day:

XSL•FO
RenderX

Each day was coded as a weekday or a weekend day (Saturday and Sunday) as a categorical variable. The type of day was then dummy coded with weekday as reference.

- Months Enrolled in the Study:
  In this study, the number of months in the study was used as a predictor for nonresponse.
- Season:
  Season of the year (summer, spring, fall, and winter) was used to better understand the role of seasonality on prompt response, with summer serving as the reference category.

Contextual

Time-varying contextual covariates were measured to examine associations with prompt completion.

- Active Phone Screen State:
  For each survey, a variable was created to display whether a prompt occurred when the phone was in active use with the screen on. Phone screen status was dummy coded with screen off as reference (screen off=0 and screen on=1).
- Time Since Last Phone Use:
  The time difference between a given prompt and the last active phone screen locking event (ie, when the active use session ends) was calculated as time since last phone use. However, a ceiling of 60 minutes was set, based on the prior work, to avoid long time gaps due to sleep time or the device being off. If the prompt occurred during active phone use, then the time difference was logged as 0 minutes.
- Location:
  Participants provided semantic location labels. The question asked, “Where are you right now?” with 21 answer options: “home,” “work,” “school/college,” “park/playground,” “sports field/court/golf course,” “gym/health club/fitness center,” “friend’s/romantic partner’s place,” “family member’s place,” “restaurant/bar/café,” “store/shopping venue,” “church/place of worship,” “movie theatre/entertainment venue,” “beach/pool,” “transit center/bus stop,” “medical clinic/hospital,” “in car/vehicle/train,” “salon/barber/spa,” “library/museum,” “gas station/convenience store,” “parking lot/structure,” “hotel/motel,” and “other.” First, all location points within a 5-minute interval were gathered, and then the point closest in time to the prompt was selected. Finally, among all semantic location labels reported for this cluster, the one with the largest count was picked. Only the most frequently occurring labels were included: “home of participant,” “home of relatives or friends,” “work or school,” “shop or restaurant,” and “indoor or outdoor sport place.” Finally, all other location labels that did not occur frequently were tagged as “other” locations. The location dummy variable was coded with “home” as a reference category.

Behavioral

Time-varying behavioral covariates were also explored.

- Sleep Duration the Prior Night:
  Sleep duration was calculated as the time difference between the retrospective wake-up time on the day of the survey and the retrospective bedtime on the preceding night.
- Daily Status:
  Daily status was meant to code whether the participant noted that they were sick or traveling on that day as these states can affect responsiveness to EMA prompts.
- Previous 10 Minutes Physical Activity:
  MIMS in the previous 10 minutes were calculated to determine proximal activity predictors. These data were analyzed in 2 different forms to examine the relationship between recent physical activity and EMA completion. First, we treated MIMS as an ordinal categorical variable to test for potential threshold effects and nonlinear relationships between activity intensity and EMA completion. We established activity intensity categories based on the distribution of MIMS values in our sample and aligned with conceptually meaningful activity intensity levels: “missing data”: no accelerometer data available; “less than 10 MIMS per minute”: representing primarily sedentary behavior and light physical activity; “10-15 MIMS per minute”: representing moderate-intensity physical activity; “greater than 15 MIMS per minute”: representing vigorous-intensity physical activity. These cut-points were determined through exploratory analysis of our dataset, examining the distribution of MIMS values during known activity types reported by participants, and selected to create meaningful distinctions between different activity intensities relevant to our research questions [21,25]. Second, we also analyzed MIMS as a continuous variable to examine the potential linear relationship between activity intensity and EMA completion probability. This dual analytical approach allowed us to determine whether physical activity's effect on EMA completion follows a threshold pattern (better captured by distinct categories) or a dose-response relationship (better captured by a continuous measure). Additionally, using both approaches serves as a sensitivity analysis to ensure our findings are robust across different analytical methods.

Psychological

Time-varying psychological factors were also explored.

- Previous Prompt Positive Affect State:
  Positive affect was measured using a set of questions asking participants to rate their current mood, such as happy, energetic, and relaxed, on a scale from 1 (not at all) to 5 (extremely), and we created a composite score out of these 3 items for the positive affect at each EMA prompt.
- Previous Prompt Stress Level:
  Stress was assessed using a single-item measure that asked participants to rate their current stress level on a scale from 1 (not at all stressed) to 5 (extremely stressed) at each EMA prompt.

Time-Invariant Covariates

Time-invariant covariates were recorded using a self-reported questionnaire at the beginning of the study. Sex at birth (binary variable: Female=1 Ref=Male), race (categorical variable), education (ordinal variable), and employment status (binary variable: 1=employed, 0=other) were all used in this study to examine completion as they have been previously used in previous studies.

XSL•FO
RenderX

Data Analysis

Data analysis was conducted using multilevel logistic regression models to examine predictors of EMA prompt completion (yes/no) over the 12-month study period. This approach was chosen to account for the nested structure of the data, with multiple observations (level 1) clustered within participants (level 2). Five separate models were constructed to investigate different sets of predictors. Separate models were used due to the large number of predictors and the potential for multicollinearity among certain variables. This approach allowed us to examine specific sets of predictors and interactions in detail while maintaining model stability and interpretability. Additionally, using separate models enabled us to explore different aspects of EMA completion behavior (eg, contextual factors and psychological states) without overfitting the data or compromising statistical power. The use of multiple models also provided means to cross-validate findings and assess the robustness of predictor effects across different model specifications. Model 1 focused on demographics (sex, age, race, ethnicity, education, and employment status), time in study (both between- and within-subject effects), and date and time variables (season, day of week, and time of day). Model 2a examined demographics, time in study, phone status (screen on/off and usage in prior hour), physical context (location), and physical activity (categorized into levels based on MIMS units). Model 2b was similar to 2a but treated physical activity as a continuous variable. Model 3a investigated demographics, time in study, sleep duration (categorized as short, normal, or long), daily status (sick or traveling), and lagged positive affect. Model 3b was similar to 3a but replaced positive affect with stress. All models included both between-subject and within-subject effects for time-varying predictors, representing their deviations from the grand and subject mean, respectively. This allows for the separation of between-person differences and within-person fluctuations in the predictors. Interaction terms between time spent in the study and other predictors (eg, sex, phone status, affect, and stress) were also included in each model to examine how the influence of various factors on EMA completion might change over the course of the study. We examined interactions between time spent in the study (both between-subject and within-subject effects) and a range of factors. These included demographic characteristics such as participant sex, contextual factors like phone screen status before reci and phone usage in the prior hour, and behavioral measures such as physical activity levels. Additionally, we investigated how time spent in the study might moderate the effects of psychological factors on EMA completion, focusing on interactions with positive affect and stress levels. All of these interaction terms were included to assess whether the influence of these factors on EMA completion rates changed as participants progressed through the year-long study period. By incorporating these interactions, we aimed to capture the dynamic nature of factors influencing EMA compliance over time and to identify any temporal patterns in participant engagement with the study protocol. The outcome variable in all models was the binary completion status of each EMA prompt (1=completed and 0=not completed). Odds ratios (ORs) with 95% CIs were calculated for each predictor, representing the change in odds of EMA completion associated with a one-unit increase in the predictor.

Results

Descriptive Results

Among the final analytical sample (n=207), 55.56% (n=115) identified as female, 43.48% (n=90) as White, and 30.43% (n=63) identified as Hispanic. In terms of education and employment, 45.41% (n=94) had a college degree or higher, and 59.9% (n=124) were fully employed. The mean age of the participants was 23.45 (SD 3.12) years.

Prompt Completion Descriptive Statistics

This intensive design resulted in an average of 12.1 (SD 1.3) prompts per participant per day, and 939.7 (SD 444.7) prompts per participant over the course of the whole survey (up to 1 year). On average, each participant had 51 (SD 30.5) days nested within 15.5 (SD 8.4) bursts. The overall mean completion rate for participants in the analytical sample was 0.77 (SD 0.13), indicating 77% of prompts were completed.

Factors Predicting Prompt EMA Completion

Multilevel logistic regression analysis was conducted to examine time-varying and time-invariant predictors of EMA prompt completion over the 12-month study period. All models can be found in Table 2.

XSL•FO
RenderX

Table 2. Multilevel logistic regression models predicting EMA completion rates in a 12-month longitudinal study.

	Model 1		Model 2a		Model 2b		Model 3a		Model 3b	
Predictor	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Intercept	3.80 (2.51-5.75)	<.001	3.40 (1.66-6.98)	<.01	4.57 (1.54-13.54)	<.01	3.78 (1.20-11.91)	.02	5.84 (2.95-11.56)	<.001
Time-invariant demographic										
Sex (Female=1; reference: male)	0.78 (0.48-1.26)	.30	0.78 (0.63-0.96)	.02	0.73 (0.58-0.92)	<.01	0.84 (0.69-1.02)	.08	0.86 (0.72-1.03)	.13
Age (Centering to 25 years)	0.97 (0.93-1.01)	.17	0.97 (0.93-1.01)	.14	0.97 (0.93-1.02)	.26	0.98 (0.94-1.02)	.26	0.98 (0.94-1.02)	.26
Race: African American (reference: White)	1.10 (0.78-1.55)	.59	0.88 (0.60-1.29)	.50	0.85 (0.57-1.28)	.43	1.09 (0.77-1.55)	.61	1.06 (0.75-1.50)	.73
Race: Asian (reference: White)	0.74 (0.58-0.95)	.02	0.74 (0.56-0.98)	.03	0.73 (0.54-0.99)	.04	0.72 (0.56-0.92)	<.01	0.73 (0.57-0.93)	.01
Race: Other (reference: White)	0.88 (0.68-1.14)	.34	0.78 (0.59-1.04)	.09	0.72 (0.53-0.97)	.03	0.86 (0.67-1.11)	.26	0.87 (0.68-1.12)	.29
Hispanic ethnicity (reference: Non-Hispanic)	0.79 (0.63-0.99)	.04	0.78 (0.61-1.00)	.05	0.76 (0.58-0.99)	.04	0.79 (0.64-0.99)	.04	0.77 (0.63-0.94)	.02
Education: College and above (reference: less than college level)	1.22 (0.97-1.54)	.09	1.39 (1.08-1.79)	.01	1.35 (1.03-1.77)	.03	1.22 (0.96-1.53)	<.01	1.20 (0.96-1.52)	.11
Employ status: Employed (reference: other)	0.75 (0.61-0.92)	<.01	0.72 (0.57-0.91)	<.01	0.67 (0.52-0.86)	<.01	0.77 (0.63-0.95)	.02	0.77 (0.63-0.94)	.01
Time-varying temporal										
Month-stayed in study (BS[b])	1.07 (0.99-1.15)	.07	1.16 (1.01-1.33)	.03	1.12 (0.92-1.36)	.26	1.11 (0.89-1.38)	.36	1.09 (0.96-1.23)	.21
Month-stayed in study (WS[c])	0.95 (0.94-0.96)	<.001	0.99 (0.97-1.01)	.22	0.97 (0.95-0.99)	.01	0.93 (0.91-0.95)	<.001	0.95 (0.93-0.97)	<.001
Season: spring (reference: summer)	1.05 (1.00-1.10)	.04	__[d]	—	—	—	—	—	—	—
Season: fall (reference: summer)	1.01 (0.97-1.06)	.60	—	—	—	—	—	—	—	—
Season: winter (reference: summer)	1.06 (1.01-1.11)	.01	—	—	—	—	—	—	—	—
Day of week: weekend (reference: weekday)	1.00 (0.97-1.03)	.98	—	—	—	—	—	—	—	—
Time of day: afternoon (reference: morning)	1.01 (0.97-1.06)	.63	—	—	—	—	—	—	—	—
Time of day: evening (reference: morning)	0.98 (0.94-1.03)	.54	—	—	—	—	—	—	—	—
Time of day: night (reference: morning)	0.76 (0.62-0.93)	<.01	—	—	—	—	—	—	—	—
Time-varying contextual										
Screen on before answering the survey (reference: screen off)	—	—	3.39 (2.81-4.09)	<.001	3.85 (3.64-4.08)	<.001	—	—	—	—
Phone usage in prior hour (BS)	—	—	1.02 (0.72-1.45)	.89	0.76 (0.64-0.90)	<.01	—	—	—	—
Phone usage in prior hour (WS)	—	—	0.80 (0.77-0.83)	<.001	0.82 (0.81-0.83)	<.001	—	—	—	—
Missing location (reference: home)	—	—	0.53 (0.51-0.56)	<.001	0.54 (0.51-0.57)	<.001	—	—	—	—

XSL•FO
RenderX

Predictor	Model 1		Model 2a		Model 2b		Model 3a		Model 3b	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Home of relatives or friends (reference: home)	—	—	0.84 (0.77-0.92)	<.001	0.83 (0.74-0.92)	<.001	—	—	—	—
Work or school (reference: home)	—	—	0.70 (0.65-0.75)	<.001	0.69 (0.63-0.75)	<.001	—	—	—	—
Shop or restaurant (reference: home)	—	—	0.61 (0.51-0.72)	<.001	0.62 (0.51-0.75)	<.001	—	—	—	—
Indoor or outdoor sport place (reference: home)	—	—	0.58 (0.47-0.72)	<.001	0.58 (0.46-0.75)	<.001	—	—	—	—
Others (reference: home)	—	—	0.60 (0.48-0.74)	<.001	0.53 (0.41-0.67)	<.001	—	—	—	—
Time-varying behavioral										
MIMS[e] in prior 10 minutes: Missing	—	—	1.30 (1.23-1.38)	<.001	—	—	—	—	—	—
MIMS in prior 10 minutes: Less than 10 MIMS per minute	—	—	1.35 (1.28-1.43)	<.001	—	—	—	—	—	—
MIMS in prior 10 minutes: Equal or more than 10 and Less than 15 MIMS per minute	—	—	1.30 (1.22-1.39)	<.001	—	—	—	—	—	—
MIMS in prior 10 minutes: Equal or more than 15 and Less than 20 MIMS per minute	—	—	1.16 (1.08-1.25)	<.001	—	—	—	—	—	—
MIMS in prior 10 minutes (BS)	—	—	—	—	1.29 (0.53-3.18)	.57	—	—	—	—
MIMS in prior 10 minutes (WS)	—	—	—	—	0.91 (0.82-1.01)	.06	—	—	—	—
Sleep short (Less than 6 hours; reference: sleep between 6 and 11 hours)	—	—	—	—	—	—	0.92 (0.87-0.99)	.02	0.92 (0.86-0.98)	.02
Sleep long (More than 11 hours; reference: sleep between 6 and 11 hours)	—	—	—	—	—	—	1.01 (0.95-1.08)	.74	1.01 (0.95-1.08)	.72
No sleep data (reference: sleep between 6 and 11 hours	—	—	—	—	—	—	0.96 (0.91-1.01)	.08	0.96 (0.91-1.01)	.08
Sick (reference: nonsick)	—	—	—	—	—	—	1.04 (0.96-1.12)	.36	1.04 (0.96-1.13)	.36
Traveling (reference: nontraveling)	—	—	—	—	—	—	0.78 (0.75-0.82)	<.001	0.78 (0.75-0.81)	<.001
Time-varying psychological										
Positive affect (BS)	—	—	—	—	—	—	1.07 (0.73-1.58)	.73	—	—
Positive affect (WS)	—	—	—	—	—	—	0.97 (0.87-1.09)	.61	—	—
Stressed (BS)	—	—	—	—	—	—	—	—	0.89 (0.66-1.22)	.48
Stressed (WS)	—	—	—	—	—	—	—	—	0.85 (0.78-0.93)	<.001
Interactions										
Month-stayed in study (BS) × sex	1.02 (0.92-1.13)	.72	—	—	—	—	—	—	—	—

XSL•FO
RenderX

Predictor	Model 1		Model 2a		Model 2b		Model 3a		Model 3b	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Month-stayed in study (WS) × sex	0.99 (0.98-1.00)	<.01	—	—	—	—	—	—	—	—
Month-stayed in study (BS) × screen on	—	—	1.02 (0.99-1.06)	.23	—	—	—	—	—	—
Month-stayed in study (WS) × screen on	—	—	0.97 (0.95-0.99)	<.001	—	—	—	—	—	—
Month-stayed in study (BS) × phone usage in prior hour (BS)	—	—	0.94 (0.87-1.01)	.08	—	—	—	—	—	—
Month-stayed in study (WS) × phone usage in prior hour (BS)	—	—	0.98 (0.97-0.99)	<.001	—	—	—	—	—	—
Month-stayed in study (BS) × phone usage in prior hour (WS)	—	—	1.00 (0.99-1.01)	.37	—	—	—	—	—	—
Month-stayed in study (WS) × phone usage in prior hour (WS)	—	—	1.00 (1.00-1.01)	<.01	—	—	—	—	—	—
Month-stayed in study (BS) × MIMS in prior 10 minutes (BS)	—	—	—	—	0.98 (0.82-1.17)	.86	—	—	—	—
Month-stayed in study (WS) × MIMS in prior 10 minutes (BS)	—	—	—	—	0.99 (0.97-1.01)	.25	—	—	—	—
Month-stayed in study (BS) × MIMS in prior 10 minutes (WS)	—	—	—	—	1.00 (0.98-1.02)	.83	—	—	—	—
Month-stayed in study (WS) × MIMS in prior 10 minutes (WS)	—	—	—	—	1.00 (0.99-1.01)	.73	—	—	—	—
Month-stayed in study (BS) × positive affect (BS)	—	—	—	—	—	—	0.99 (0.92-1.07)	.75	—	—
Month-stayed in study (WS) × positive affect (BS)	—	—	—	—	—	—	1.01 (0.99-1.01)	.18	—	—
Month-stayed in study (BS) × positive affect (WS)	—	—	—	—	—	—	1.01 (0.99-1.03)	.23	—	—
Month-stayed in study (WS) × positive affect (WS)	—	—	—	—	—	—	1.00 (0.99-1.01)	.82	—	—
Month-stayed in study (BS) × stressed (BS)	—	—	—	—	—	—	—	—	0.99 (0.93-1.05)	.76
Month-stayed in study (WS) × stressed (BS)	—	—	—	—	—	—	—	—	1.00 (0.99-1.01)	.45
Month-stayed in study (BS) × stressed (WS)	—	—	—	—	—	—	—	—	1.02 (1.01-1.04)	<.01
Month-stayed in study (WS) × stressed (WS)	—	—	—	—	—	—	—	—	1.00 (1.00-1.01)	.76

[a]OR: odds ratio.

[b]BS: between subject.

[c]WS: within subject.

[d]Not applicable.

[e]MIMS: monitor-independent movement summary.

Time-Invariant Demographic Factors

Several time-invariant demographic factors were found to significantly predict EMA completion odds across multiple models. In Model 2a, sex at birth played a role, with females showing lower odds of completion compared to males (OR 0.78, 95% CI 0.63-0.96; *P*=.02). While no significant effect of age was observed in any model, race and ethnicity were important predictors. Model 1 showed that Asian participants had lower odds of completion compared to White participants

XSL•FO
RenderX

(OR 0.74, 95% CI 0.58-0.95; *P*=.02), and Hispanic participants showed lower odds of completion compared to non-Hispanic participants (OR 0.79, 95% CI 0.63-0.99; *P*=.04). Education level emerged as a positive predictor in Model 2a, with higher education levels (college and above) associated with increased odds of EMA completion (OR 1.39, 95% CI 1.03-1.77; *P*=.01). In contrast, employment status was negatively associated with completion across models, as fully employed individuals showed lower odds of completion than those who were not (OR 0.75, 95% CI 0.61-0.92, *P*>.01 in Model 1).

Time-Varying Temporal Factors

Multiple regression analyses revealed several significant time-varying temporal factors influencing EMA completion odds. In Model 1, time of day played a crucial role, with nighttime associated with lower odds of completion compared to morning hours (OR 0.76, 95% CI 0.62-0.93; *P*>.01). Interestingly, no significant difference was found between weekdays and weekends across all models, suggesting that the day of the week did not statistically impact completion rates. Seasonal effects were observed in Model 1, with higher odds of completion in spring (OR 1.05, 95% CI 1.00-1.10; *P*=.04) and winter (OR 1.06, 95% CI 1.01-1.11; *P*=.01) compared to summer. The duration of participation in the study showed a significant within-subject decrease in completion over time (OR 0.95, 95% CI 0.94-0.96, *P*<.001 in Model 1), indicating declining completion odds as participants progressed through the study.

Time-Varying Contextual Factors

Models 2a and 2b highlighted the importance of phone-related factors for EMA completion. Having the phone screen on before answering the survey significantly increased the odds of completion (OR 3.39, 95% CI 2.81-4.09, *P*<.001 in both models), while higher within-subject phone usage in the prior hour decreased completion odds (OR 0.80, 95% CI 0.77-0.83, *P*<.001 in Model 2a; OR 0.82, 95% CI 0.81-0.83, *P*<.001 in Model 2b). Physical location was also a significant predictor in these models, with being outside the home generally associated with lower completion odds. For example, compared to being at home, participants were less likely to complete EMAs when at work or school (OR 0.70, 95% CI 0.65-0.75, *P*<.001 in Model 2a).

Time-Varying Behavioral Factors

The study found significant associations between behavioral factors and EMA completion odds across different models. In Models 3a and 3b, sleep duration from the prior night emerged as an important predictor, with short sleep duration (less than 6 hours) associated with lower completion odds (OR 0.92, 95% CI 0.87-0.99; *P*=.02) compared to normal sleep duration. Model 2a examined physical activity levels in the 10 minutes preceding the EMA prompt, finding that lower activity levels were generally associated with higher completion odds compared to high activity levels (ORs ranging from 1.16 to 1.35, all *P*<.001). However, when physical activity was used as a continuous variable (Model 2b), no significant effects were observed. Models 3a and 3b found that traveling was associated with decreased completion odds (OR 0.78, 95% CI 0.75-0.82, *P*<.001 in both models) although no significant effect was found for being sick.

Time-Varying Psychological Factors

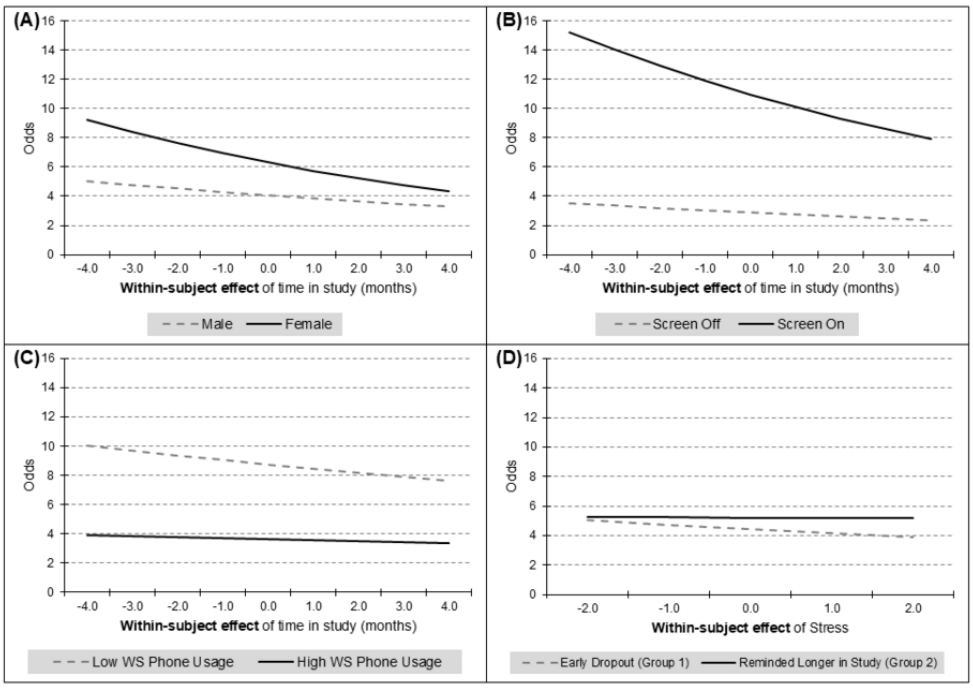
Affective states and stress levels from the previous prompt showed significant associations with completion odds in Models 3a and 3b. Model 3a indicated that neither between-subject nor within-subject positive affect was statistically associated with completion odds, suggesting that the experience of positive emotions at the prior EMA prompt may not influence subsequent EMA completion behavior. However, Model 3b showed that within-subject stress levels had a significant negative association with odds of completing the subsequent EMA prompt (OR 0.85, 95% CI 0.78-0.93; *P*<.001), and it implied that participants were less likely to complete EMA surveys when experiencing elevated stress during the prior EMA prompt.

Interactions

The analysis revealed several significant interaction effects across different models, providing insights into how the influence of various factors on EMA completion odds changed over the course of the study. In Model 1, an interaction effect was observed between participant sex and within-subject time spent in the study, suggesting that females may experience a slightly faster decline in survey completion odds as the study progresses, relative to males (OR 0.99, 95% CI 0.98-1.00; *P*>.01; see Figure 1A). Model 2a showed interaction effects between screen status and within-subject time spent in the study (OR 0.97, 95% CI 0.95-0.99; *P*<.001), indicating that while having the phone screen on substantially increased completion odds initially, this positive effect diminished over time in the study (see Figure 1B). Similarly, the interaction between phone usage in the prior hour and within-subject time in the study (OR 0.98, 95% CI 0.97-0.99; *P*<.001) suggests that the negative impact of recent phone usage on completion odds became less pronounced as participants progressed through the study (see Figure 1C). Additionally, Model 3b revealed interactions between psychological factors and total time a participant stayed in the study (between-subject time effect). Model 3b revealed a significant interaction between within-subject stress levels and the duration of total time participants remained in the study (OR 1.02, 95% CI 1.01-1.04; *P*>.001). While the main effect of within-subject momentary stress was negatively associated with EMA completion, participants who remained in the study longer exhibited a less pronounced negative impact on survey completion behavior compared to those who dropped out at an earlier stage (see Figure 1D).

XSL•FO
RenderX

Figure 1. Interaction effects of key predictors and study duration on EMA completion. These plots illustrate the interaction effects between key predictors and study duration on momentary EMA completion odds. (A) Interaction effect between sex and within-subject time duration in study (ie, the deviation from each participant's average month-stayed in the study) on completion odds as analyzed in Model 1. (B) Interaction effect between phone screen state and within-subject time duration in study on completion odds as examined in Model 2a. (C) Interaction effect between within-subject phone usage (ie, the deviation score from each participant's average phone usage) and within-subject time duration in study on completion odds, also in Model 2a. (D) Interaction effect between within-subject momentary stress (ie, the deviation score from each participant's average stress) and between-subject study duration (ie, early dropout vs reminded longer in study) on completion odds as detailed in Model 3b. EMA: ecological momentary assessment; WS: within subject.



Discussion

Principal Findings

This study examined factors influencing EMA completion in a 12-month intensive mutiburst longitudinal study among young adults. Our findings reveal a complex interplay of time-varying temporal, contextual, behavioral, and psychological factors and time-invariant individual characteristics that affect EMA completion over time.

Our findings revealed an interesting pattern regarding phone-related factors. While having the phone screen on at the moment of prompt delivery substantially increased the odds of EMA completion, greater phone usage in the prior hour predicted lower completion odds. This seemingly contradictory result may reflect different aspects of device engagement. Having the screen on at prompt delivery represents immediate accessibility and attention to the device, facilitating a quick response. In contrast, higher prior usage may indicate engagement with competing activities on the phone (eg, social media, games, and work tasks) that could reduce willingness to interrupt these activities to complete a survey. Additionally, higher prior usage might indicate a pattern of more fragmented attention or task-switching, making participants less likely to commit to completing the multiquestion EMA. This pattern suggests that EMA designs might benefit from leveraging moments when the device is already in use, while being mindful that excessive prior engagement with the device might create competing demands on user attention.

The observed seasonal effects on completion rates, with higher odds of completion in spring and winter compared to summer, likely reflect behavioral patterns related to indoor or outdoor activity. Research has shown that physical activity levels vary seasonally, which could affect availability for completing assessments [26]. During spring and winter, participants may spend more time indoors, where their phones are more readily accessible and there are fewer competing activities. In contrast, summer months often involve more outdoor activities, travel, and social events where participants may be physically separated from their devices or less willing to interrupt activities to respond to prompts. These seasonal variations could also be attributed to changes in daily routines, such as academic schedules for students or shifts in work patterns [27]. Seasonal changes in mood, particularly seasonal affective disorder, might impact participants' willingness to respond to EMA prompts [28]. Environmental factors like daylight hours and weather conditions have also been found to influence smartphone usage patterns [29], potentially affecting EMA completion rates. Further, the lower completion odds observed outside the home environment, particularly at work or school, highlight the impact of daily routines and competing demands on participants' ability or willingness to respond to prompts. These findings align with previous studies that have identified location and daily activities as important factors in EMA completion [1,7]. Understanding these influences could help researchers design more adaptive EMA protocols that account for predictable fluctuations in participant engagement throughout the year.

XSL•FO
RenderX

Behavioral and psychological states also significantly influenced the odds of EMA completion. Short sleep duration on the previous night and higher levels of physical activity in the preceding 10 minutes were associated with lower completion odds, suggesting that fatigue and engagement in other activities may impede EMA participation. These findings are consistent with previous research indicating that sleep and activity patterns can affect EMA completion [13]. Specifically, studies have shown that poor sleep quality and shorter sleep duration are associated with lower EMA compliance rates [13,30]. This may be due to increased fatigue or cognitive impairment following insufficient sleep, which can reduce participants' motivation or ability to respond to prompts. Regarding physical activity, higher levels of activity, especially during vigorous exercise, have been associated with decreased likelihood of responding to EMA prompts [7,31]. This could be due to the physical unavailability to interact with devices during exercise or the disruption of attention to prompts during activity engagement. Momentary stress was negatively associated with EMA completion rates. The negative association with stress aligns with previous research suggesting that heightened stress can reduce compliance with study protocols [30]. These results underscore the complex relationship between momentary psychological states and EMA engagement, highlighting the need for further research in this area.

Individual characteristics, including sex, race, ethnicity, education, and employment status, were found to be significant predictors of EMA completion. These findings highlight concerns about accessibility and representation when designing and implementing longitudinal multiburst EMA studies, as they may influence participants' ability or willingness to engage with the protocol over time. The observed differences in completion rates across demographic groups align with previous research indicating disparities in EMA compliance [12]. Specifically, studies have found lower compliance rates among racial and ethnic minorities [12,13] and individuals with lower socioeconomic status [32]. Gender differences have also been observed, with some studies reporting lower compliance among males [30], although findings on gender disparities are mixed and may depend on the study context and population.

The decline in completion rates over the course of the study, as indicated by the significant within-subject effect of time, is a common challenge in longitudinal EMA research [1,12]. This decline may reflect participant fatigue, waning motivation, or changes in the perceived value of participation over time [13,33]. The interaction effects observed between time in study and other predictors provide valuable insights into how the influence of various factors on EMA completion evolves throughout a long-term study. For example, the diminishing effect of phone screen status on completion rates over time suggests that participants may become less responsive to this cue as the study progresses [34]. Similarly, the changing relationship between time in study and psychological factor (stress) indicates that the impact of momentary states on compliance may shift as participants become more accustomed to the EMA protocol [30]. The observed interactions between time spent in the study and various predictors, including sex and psychological states, further emphasize the dynamic nature of EMA engagement over extended study periods. These interactions suggest that the factors influencing compliance may shift over time, potentially due to changes in participant motivation, study fatigue, or evolving life circumstances [35].

While within-subject effects of time reveal important patterns in individual trajectories of EMA completion, the between-subject time in study interactions warrant separate consideration. These interactions provide insights into differences between participants who remain in the study for longer periods versus those who drop out earlier. The interaction with stress levels suggests that participants who remained in the study longer were better able to maintain compliance despite experiencing stress. These findings highlight the complex relationship between participant retention, EMA completion, and psychological factors, emphasizing the need to consider both short-term fluctuations and long-term engagement patterns when interpreting EMA data [1,36].

Limitations

Several limitations should be considered when interpreting these results. First, the study sample was limited to young adults in the United States, potentially limiting generalizability to other age groups or cultural contexts. Second, the use of Android smartphones exclusively may have introduced selection bias and limits the applicability of findings to users of other mobile platforms. Third, the COVID-19 pandemic occurred during the study period, potentially influencing participants' daily routines and psychological states in ways that may not be representative of typical conditions. Additionally, while our models accounted for a wide range of factors, unmeasured variables may still have influenced EMA completion odds. These could include individual differences in personality traits (eg, conscientiousness), technology literacy, or specific life events not captured by our assessments. The use of single-item measures for some psychological constructs (eg, affect and stress) may not fully capture the complexity of these states. Finally, due to the large number of observations per person, this study had statistical power to detect very small effects. While statistically significant, these small effects may not be substantially or clinically meaningful, and their practical implications should be interpreted cautiously. Finally, given the large number of predictors tested in our models, it is important to note that some of the findings may be spurious due to multiple testing. While we have confidence in our overall results, individual predictor effects, especially those with borderline significance, should be interpreted with caution and validated in future studies.

Implications

These findings have important implications for the design and implementation of future EMA studies. Researchers should consider tailoring prompt schedules to individual participants' routines and contexts to maximize completion rates, as suggested by Nahum-Shani et al [10]. However, it is crucial to recognize that overly tailored designs may yield less representative data if prompts are only delivered during convenient times. This could potentially bias the sample toward certain contexts or psychological states, undermining the ecological validity of the data. Therefore, a balance must be struck between optimizing

XSL•FO
RenderX

completion rates and maintaining the representativeness of the collected data. The development of adaptive sampling strategies that account for participants' current state and context could help maintain engagement over extended study periods, but these strategies should be designed to ensure adequate sampling across various contexts and states. The development of adaptive sampling strategies that account for participants' current state and context could help maintain engagement over extended study periods [37]. The observed demographic differences in completion rates suggest that targeted strategies may be necessary to ensure equitable representation in EMA data collection. These strategies may include providing additional support or incentives for groups with lower completion rates or developing culturally sensitive approaches to EMA implementation.

The strong influence of contextual factors such as phone use on EMA completion odds highlights the potential for leveraging smartphone features to enhance EMA engagement. For example, researchers might explore the use of smartwatch-based prompts or other wearable devices to increase the accessibility of EMA surveys [34]. The findings regarding psychological states and completion rates suggest that researchers should carefully consider the timing and frequency of prompts, potentially incorporating measures of participant burden or adjusting protocols based on reported stress levels.

Conclusions

This study provides valuable insights into the factors influencing EMA completion in a year-long multiburst intensive longitudinal study. The findings highlight the complex and dynamic nature of participant engagement with EMA protocols, emphasizing the need for thoughtful study design and analysis to maximize data quality and representativeness. The observed interactions between time in study and various predictors underscore the importance of considering temporal dynamics in EMA research, particularly for long-term studies. Future research should continue to explore strategies for optimizing EMA methodologies, such as adaptive sampling techniques and the integration of passive sensing data, to enhance our understanding of real-time experiences and behaviors in naturalistic settings. By addressing the challenges identified in this study, researchers can work toward developing more effective and sustainable EMA protocols, ultimately improving the quality and utility of intensive longitudinal data in behavioral science research.

Acknowledgments

This research was supported by the National Institutes of Health (NIH/NHLBI U01HL146327). Additional support for the author's effort was provided by the NIH through a K01 Career Development Award (grant 1K01HL169818). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart describing participant exclusions.
[DOCX File , 46 KB - mhealth_v13i1e67117_app1.docx]

References

1. Burke LE, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]
2. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
3. Dunton GF. Ecological momentary assessment in physical activity research. Exerc Sport Sci Rev 2017;45(1):48-54 [FREE Full text] [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]
4. Romanzini CLP, Romanzini M, Batista MB, Barbosa CCL, Shigaki GB, Dunton G, et al. Methodology used in ecological momentary assessment studies about sedentary behavior in children, adolescents, and adults: systematic review using the checklist for reporting ecological momentary assessment studies. J Med Internet Res 2019;21(5):e11967 [FREE Full text] [doi: 10.2196/11967] [Medline: 31094349]
5. Dunton GF. Sustaining health-protective behaviors such as physical activity and healthy eating. JAMA 2018;320(7):639-640 [FREE Full text] [doi: 10.1001/jama.2018.6621] [Medline: 29852046]
6. Dunton GF, Wang SD, Do B, Courtney J. Early effects of the COVID-19 pandemic on physical activity locations and behaviors in adults living in the United States. Prev Med Rep 2020;20:101241 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101241] [Medline: 33173751]
7. Dunton GF, Liao Y, Intille S, Huh J, Leventhal A. Momentary assessment of contextual influences on affective response during physical activity. Health Psychol 2015;34(12):1145-1153 [FREE Full text] [doi: 10.1037/hea0000223] [Medline: 26053885]

XSL•FO
RenderX

8. Dunton GF, Huh J, Leventhal AM, Riggs N, Hedeker D, Spruijt-Metz D, et al. Momentary assessment of affect, physical feeling states, and physical activity in children. Health Psychol 2014;33(3):255-263 [FREE Full text] [doi: 10.1037/a0032640] [Medline: 23668846]
9. Dunton GF, Kawabata K, Intille S, Wolch J, Pentz MA. Assessing the social and physical contexts of children's leisure-time physical activity: an ecological momentary assessment study. Am J Health Promot 2012;26(3):135-142. [doi: 10.4278/ajhp.100211-QUAN-43] [Medline: 22208410]
10. Nahum-Shani I, Smith S, Spring B, Collins L, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing Health behavior support. Ann Behav Med 2018;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
11. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
12. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017;19(4):e132. [doi: 10.2196/jmir.6641] [Medline: 28446418]
13. Courvoisier DS, Eid M, Lischetzke T. Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. Psychol Assess 2012;24(3):713-720. [doi: 10.1037/a0026733] [Medline: 22250597]
14. Degroote L, DeSmet A, De Bourdeaudhuij I, Van Dyck D, Crombez G. Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2020;17(1):35 [FREE Full text] [doi: 10.1186/s12966-020-00932-9] [Medline: 32151251]
15. Dzubur E, Huh J, Maher J, Intille S, Dunton G. Response patterns and intra-dyadic factors related to compliance with ecological momentary assessment among mothers and children. Transl Behav Med 2018;8(2):233-242 [FREE Full text] [doi: 10.1093/tbm/ibx002] [Medline: 29381785]
16. Wang S, Intille S, Ponnada A, Do B, Rothman A, Dunton G. Investigating microtemporal processes underlying health behavior adoption and maintenance: protocol for an intensive longitudinal observational study. JMIR Res Protoc 2022;11(7):e36666 [FREE Full text] [doi: 10.2196/36666] [Medline: 35834296]
17. Hayley AC, Skogen JC, Øverland S, Wold B, Williams LJ, Kennedy GA, et al. Trajectories and stability of self-reported short sleep duration from adolescence to adulthood. J Sleep Res 2015;24(6):621-628. [doi: 10.1111/jsr.12316] [Medline: 26172979]
18. Corder K, Winpenny E, Love R, Brown HE, White M, Sluijs EV. Change in physical activity from adolescence to early adulthood: a systematic review and meta-analysis of longitudinal cohort studies. Br J Sports Med 2019;53(8):496-503 [FREE Full text] [doi: 10.1136/bjsports-2016-097330] [Medline: 28739834]
19. Harris PA, Scott KW, Lebo L, Hassan N, Lightner C, Pulley J. ResearchMatch: a national registry to recruit volunteers for clinical research. Acad Med 2012;87(1):66-73 [FREE Full text] [doi: 10.1097/ACM.0b013e31823ab7d2] [Medline: 22104055]
20. Wang SD, Hatzinger L, Morales J, Hewus M, Intille S, Dunton GF. Burden and inattentive responding in a 12-month intensive longitudinal study: interview study among young adults. JMIR Form Res 2024;8:e52165 [FREE Full text] [doi: 10.2196/52165] [Medline: 39093606]
21. John D, Tang Q, Albinali F, Intille S. An open-source monitor-independent movement summary for accelerometer data processing. J Meas Phys Behav 2019;2(4):268-281 [FREE Full text] [doi: 10.1123/jmpb.2018-0068] [Medline: 34308270]
22. Belcher B, Wolff-Hughes D, Dooley E, Staudenmayer J, Berrigan D, Eberhardt M, et al. US population-referenced percentiles for wrist-worn accelerometer-derived activity. Med Sci Sports Exerc 2021;53(11):2455-2464 [FREE Full text] [doi: 10.1249/MSS.0000000000002726] [Medline: 34115727]
23. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Washington, DC: AAAI Press; 1996:226-231.
24. Moreira A, Santos MY. Concave hull: a k-nearest neighbours approach for the computation of the region occupied by a set of points. In: Proceedings of the Second International Conference on Computer Graphics Theory and Applications (VISIGRAPP 2007) - Volume 2: GRAPP. Setúbal, Portugal: SciTePress; 2007:61-68.
25. Qin X, Zhu W, Zhu L, Liu J, Liao J. Linking MIMS with ActiGraph count: an equating study. Meas Phys Educ Exerc 2022;27(2):97-104. [doi: 10.1080/1091367x.2022.2072688]
26. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. Public Health 2007;121(12):909-922. [doi: 10.1016/j.puhe.2007.04.009] [Medline: 17920646]
27. Robbins ML, Kubiak T. Ecological momentary assessment in behavioral medicine. In: The Handbook of Behavioral Medicine. Hoboken, NJ: Wiley Blackwell; 2014:429-446.
28. Grimaldi S, Partonen T, Haukka J, Aromaa A, Lönnqvist J. Seasonal vegetative and affective symptoms in the Finnish general population: testing the dual vulnerability and latitude effect hypotheses. Nord J Psychiatry 2009;63(5):397-404. [doi: 10.1080/08039480902878729] [Medline: 19363741]

XSL•FO
RenderX

29. Christensen MA, Bettencourt L, Kaye L, Moturu ST, Nguyen KT, Olgin JE, et al. Direct measurements of smartphone screen-time: relationships with demographics and sleep. PLoS One 2016;11(11):e0165331 [FREE Full text] [doi: 10.1371/journal.pone.0165331] [Medline: 27829040]
30. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. J Med Internet Res 2019;21(2):e11398 [FREE Full text] [doi: 10.2196/11398] [Medline: 30720437]
31. Bruening M, van Woerden I, Todd M, Brennhofer S, Laska MN, Dunton G. A mobile ecological momentary assessment tool (devilSPARC) for nutrition and physical activity behaviors in college students: a validation study. J Med Internet Res 2016;18(7):e209 [FREE Full text] [doi: 10.2196/jmir.5969] [Medline: 27465701]
32. Snippe E, Simons CJP, Hartmann JA, Menne-Lothmann C, Kramer I, Booij SH, et al. Change in daily life behaviors and depression: within-person and between-person associations. Health Psychol 2016;35(5):433-441 [FREE Full text] [doi: 10.1037/hea0000312] [Medline: 26690641]
33. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
34. Ponnada A, Wang S, Chu D, Do B, Dunton G, Intille S. Intensive longitudinal data collection using microinteraction ecological momentary assessment: pilot and preliminary results. JMIR Form Res 2022;6(2):e32772 [FREE Full text] [doi: 10.2196/32772] [Medline: 35138253]
35. Stone AA, Shiffman S, Atienza A, Nebeling L. The Science of Real-Time Data Capture: Self-Reports in Health Research. New York, NY: Oxford University Press; 2007.
36. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York, NY: Guilford Press; 2013.
37. Intille S, Haynes C, Maniar D, Ponnada A, Manjourides J. µEMA: microinteraction-based ecological momentary assessment (EMA) using a smartwatch. In: UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, NY: Association for Computing Machinery; 2016:1124-1128.

Abbreviations

DBSCAN: density-based spatial clustering of applications with noise
EMA: ecological momentary assessment
MIMS: monitor-independent movement summary
OR: odds ratio
TIME: Temporal Influences on Movement and Exercise
USC: University of Southern California

Edited by L Buis; submitted 03.10.24; peer-reviewed by J Courtney, R Pryss; comments to author 18.03.25; revised version received 09.05.25; accepted 31.07.25; published 22.10.25.

Please cite as:
Prochnow T, Wang WL, Wang S, Li J, Rothman AJ, Intille SS, Hedeker D, Dunton GF
Understanding Longitudinal Ecological Momentary Assessment Completion: Results From 12 Months of Burst Sampling in the TIME Study
JMIR Mhealth Uhealth 2025;13:e67117
URL: https://mhealth.jmir.org/2025/1/e67117
doi:10.2196/67117
PMID:

©Tyler Prochnow, Wei-Lin Wang, Shirlene Wang, Jixin Li, Alexander J Rothman, Stephen S Intille, Donald Hedeker, Genevieve F Dunton. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Care Workers' Experience With a Psychological Self-Monitoring App During the COVID-19 Pandemic: Mixed Methods Study

Lydia Khaldoun[1,2], BSc; Francois Bellemare[1,3], BSc; Christine Genest[1,4], PhD; Nicolas Bergeron[5,6], MD; Steve Geoffrion[1,2], PhD

[1]Centre de recherche de l'Institut universitaire en santé mentale de Montréal, Montreal, QC, Canada
[2]École de psychoéducation, Faculté des arts et des sciences, Université de Montréal, Montreal, QC, Canada
[3]Département de psychologie, Faculté des arts et des sciences, Université de Montréal, Montreal, QC, Canada
[4]Faculté des sciences infirmières de l'Université de Montréal, Université de Montréal, Montreal, QC, Canada
[5]Centre de recherche du Centre hospitalier de l'Université de Montréal, Montreal, QC, Canada
[6]Département de psychiatrie et d'addictologie, Faculté de médecine, Université de Montréal, Montreal, QC, Canada

Corresponding Author:
Lydia Khaldoun, BSc
Centre de recherche de l'Institut universitaire en santé mentale de Montréal
7331, rue Hochelaga
Montreal, QC, H1N 3V2
Canada
Phone: 1 514 251 4015 ext 3937
Email: lydia.khaldoun@umontreal.ca

Abstract

Background: Health care workers (HCWs) are at risk of experiencing psychological distress, particularly during the COVID-19 pandemic. Psychological self-monitoring apps may contribute to reducing symptoms of depression, anxiety, and trauma exposure by enhancing emotional self-awareness. This study focused on how a basic psychological self-monitoring app was experienced by HCWs during the COVID-19 pandemic in Quebec by exploring users' experience and factors contributing to their adherence.

Objective: This study aimed to explore HCWs' experiences with a psychological self-monitoring app, including if their satisfaction with the app, their perception of its contribution to self-awareness, and their experience of distress influenced their adherence to the app.

Methods: HCWs in Quebec were invited to respond weekly to questions about their well-being via a mobile app. A convergent mixed methods design was used. Sample data (N=424) were collected from the app, a postparticipation questionnaire was administered, and 30 semistructured interviews were conducted. Correlations and hierarchical multiple regression models were conducted to examine possible factors influencing participants' adherence, and a thematic analysis was used to further explore their experience.

Results: Over a 12-week-period, mean adherence to the psychological self-monitoring app was 74.5% (SD 29.4%) and mean satisfaction was 80% (SD 20%). Most participants perceived that the app contributed moderately (165/418, 39.5%) or a lot (140/418, 33.5%) to enhancing their self-awareness. The significant regression model ($F_{5,401}$=6.59; P<.001) suggested that around 7.6% of adherence variation could be explained by satisfaction (β=.16; t_{401}=3.14; P=.002) and the app's perceived contribution to self-awareness (β=.15; t_{401}=2.88; P=.004). Biological sex (369/419, 88.1% female and 50/419, 11.9% male), age (mean 40.8, SD 9.9 y), and the experience of psychological distress at least once in 12 weeks (228/420, 54.3%) were not statistically significant predictors of adherence. Emergent themes from the 30 interviews highlighted participants' experiences. Psychological self-monitoring was seen as an introspective practice, with reports of enhanced self-awareness and self-care practices. Interviewees generally considered the app as practical, but it did not suit everyone's preferences. Potential app enhancements were provided by the participants.

Conclusions: A simple psychological self-monitoring app could be an interesting tool for HCWs who wish to improve their self-awareness and prevent psychological distress, particularly in health crises such as pandemics.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2025;13:e70412) doi:10.2196/70412

KEYWORDS

psychological self-monitoring; self-monitoring app; adherence; user experience; health care workers; well-being

Introduction

Background

Workers in health and social services networks (health care workers; HCWs), are exposed to various stressful occupational situations [1,2]. They are reported to be prone to psychological distress [3,4], characterized by painful mental and physical symptoms often assessed through self-reported measures of posttraumatic stress (PTS), anxiety, and depression [5]. Furthermore, times of crisis such as the COVID-19 pandemic can be particularly distressful for HCWs [6]. A recent meta-analysis reported that the prevalence of depression and anxiety among HCWs was 37% and 39%, respectively [7], while another found that 47% of the HCWs experienced burnout, 38% experienced anxiety, 34% experienced depression, 30% experienced acute stress disorder, and 26% experienced PTS disorder [8].

Exploring and assessing efficient ways to support HCWs is crucial. On the basis of the National Institute for Health and Care Excellence recommendations [9], a continuous monitoring of various reactions could be an effective strategy for identifying individuals at risk and preventing the development of mental health disorders. Hence, psychological self-monitoring can be defined as the examination of one's personal condition through one's thoughts, emotions, behaviors, and consequences.

Studies suggest that psychological self-monitoring may promote emotional self-awareness, which is the ability to recognize and describe one's mood [10]. For example, in a qualitative study (N=21) tracking mood and physical activity in participants with bipolar disorder, half of the participants reported that psychological self-monitoring helped them recognize their emotional states and link them to factors such as sleep or exercise, thereby promoting behavioral changes [11]. Thereon, researchers have found that emotional self-awareness plays a mediating role in the relationship between psychological self-monitoring and psychological distress reduction [12-16]. Subsequently, awareness of a state of stress would lead the person to use coping strategies to adapt and cope with a situation or an environment [17]. Emotional self-awareness has been identified as an important factor for the well-being of prospective HCWs [18]. Hence, fostering emotional self-awareness with psychological self-monitoring might reduce psychological distress among HCWs, thereby improving their well-being.

To facilitate psychological self-monitoring, using digital technology is a promising solution, particularly during a pandemic [19,20]. Reviews have shown that mobile devices can improve users' psychological well-being even without professional guidance [21-24]. Notably, apps that reduce psychological distress symptoms often invite users to practice psychological self-monitoring [12,25-28]. However, those apps offer other components and features, such as behavioral interventions and educational modules. There is limited understanding of psychological self-monitoring's isolated influence on mental health management, and no studies previously focused on HCWs.

In addition, a major issue raised by researchers is the actual use of a digital tool. In a systematic review of available digital self-help interventions, indications of sustained use, such as the completion of all assessments or the use of the apps for more than 6 weeks, varied from 0.5% to 28.6% [29]. In this regard, adherence is the term often used to describe how much a technology is used [30] and is mostly measured by the number of logins, completed modules, or completed activities [31]. Measures of adherence have been positively associated with positive psychological outcomes in digital interventions [31]. A younger age and a biologically female sex were found to positively influence adherence [32,33]; however, research is needed to identify other factors [34].

Exploring users' experience could provide a holistic perspective on their interaction with the psychological self-monitoring app. User experience reflects the perceptions and responses resulting from the use of a digital tool [35]. It can be investigated by analyzing induced feelings, judgments, and behaviors [36], and both qualitative and quantitative approaches can be used.

In parallel, the expectation-confirmation model suggests that app users form expectations, and their satisfaction is determined by how well those expectations match their actual experience [37]. In this context, satisfaction could be conceptualized as an overall evaluation of user experience with the psychological self-monitoring app [38]. Considering that satisfied users are more likely to sustain app use [32,37,39,40], a positive association could be hypothesized between adherence and satisfaction with a psychological self-monitoring app. According to the same model, perceived usefulness is proposed to be another key factor influencing adherence [37], a finding supported by reviews [32,33,41]. Indeed, users who feel that a psychological self-monitoring app positively influences their psychological state or their emotional self-awareness may perceive the app as useful, and thus, have higher adherence. However, the experience of psychological distress may or may not influence adherence, as conflicting results are found in the literature [34,42-47]. In a psychological self-monitoring app, a higher adherence could possibly be expected from users who experienced psychological distress, as they might have felt a greater need to monitor their psychological state.

Objectives

In short, studies suggest that a psychological self-monitoring app could help users become more aware of their emotional state and could lead to improvements in well-being management. Further exploration is needed to discover the user experience with a simple psychological self-monitoring app and potential factors associated with its adherence. This study investigated Canadian HCWs' use of a basic psychological self-monitoring

XSL•FO
RenderX

app during the first 2 COVID-19 waves in Quebec [48-50]. The first objective was to examine whether adherence to a basic psychological self-monitoring app was associated with participants' satisfaction, perceived app contribution to enhance emotional self-awareness, and the presence of psychological distress, when adjusting for age and biological sex. It was hypothesized that adherence would be higher whenever participants (1) had an increased satisfaction, (2) had an increased perception of the app's contribution to emotional self-awareness, and (3) experienced psychological distress during their self-monitoring. The second objective was to explore user experience through thematic analysis of qualitative data.

Methods

Study Design

Considering the exploratory aims of this research, a convergent mixed methods design was adopted [51]. Quantitative data were collected from a psychological self-monitoring app and an online questionnaire. Qualitative data were collected from individual interviews.

Study Population

Participants were Canadian HCWs mostly working in urban areas of Quebec's integrated health and social services centers. They had participated in a program implemented during the COVID-19 pandemic to monitor HCWs' psychological distress with an app. Inclusion criteria were having access to a smartphone and to Wi-Fi and working during the participation period in 1 of the 8 following integrated health and social services centers: University of Montreal Hospital Centre, Est de Montréal, Centre-Sud de Montréal, de-la-Capitale, Montérégie, Laval, Des Îles, and Estrie-CHUS. There were no exclusion criteria as all HCWs in these centers were invited to participate. Convenience sampling was used and participation to each phase of the project was voluntary. For the first objective, 424 users (n=369, 87% women; n=50, 11.8% men; and n=5, 1.2% missing data) accepted to fill out a postparticipation questionnaire. Participants' age ranged from 18 to 68 (mean 40.8, SD 9.9) years and their years of work experience ranged from 0 to 43 (mean 13.6, SD 9.6) years. For the second objective, the analyzed data were collected from a subsample of 25 women and 5 men who were randomly selected for an interview. Their age ranged between 24 and 57 (mean 43, SD 9) years, with 0 to 30 years of work experience (mean 15, SD 9 years).

Ethical Considerations

Ethical approval for the research project involving the self-monitoring app was granted by the University of Montreal Hospital Centre research ethics committee (reference: MP-02-2021-8963). Approval for this secondary analysis was granted by the University of Montreal's research ethics committee in education and psychology (reference: 2023-4981). Informed consent was obtained from all individual participants included in the study. While participants did not receive monetary compensation, they were eligible to win 1 of 40 Can $50 (US $36.35) prepaid cards. All study data were deidentified and identifying details were removed from verbatims to ensure confidentiality.

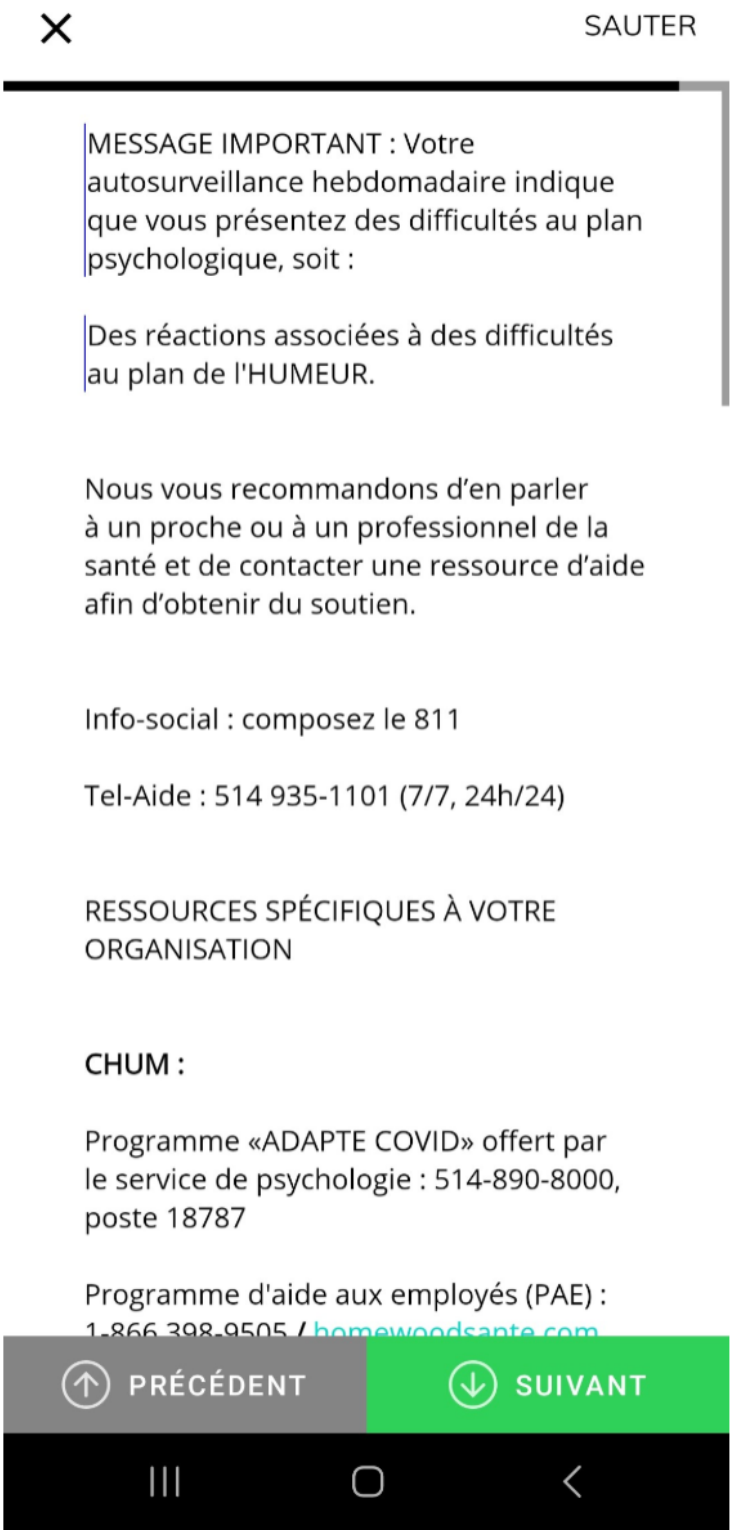
Procedure

Between May 2020 and February 2021, 831 HCWs voluntarily participated in a psychological distress self-monitoring program. Following methodological aspects of the ecological momentary assessment [52], participants were asked to answer a series of questions on an app each week over a 12-week period. They installed the Avicenna app, formally known as Ethica, a platform that enables researchers to collect data provided by their participants [53]. The app sent weekly notification prompting self-monitoring with subsequent reminders every 12 hours. Data collection was confidential.

On the app, the questions assessed several elements related to COVID-19 exposure, perceived availability of social support, and perceived quality of life (Multimedia Appendix 1). In addition, anxiety, PTS, and depression symptoms were measured by validated questionnaires (see the Instruments and Measures section). When a participant's self-reported responses reached the significant threshold of one or more of these 3 validated questionnaires, the participant received a message indicating the presence of significant distress, followed by a list of available resources (Figure 1). If no significant distress was detected, only the list of resources was displayed. In the event of suicidal ideations, the regional helpline's contacts details were displayed on the next page. In addition, an email was automatically sent to the project coordinator, who ensured a follow-up call with the participant.

XSL•FO
RenderX

Figure 1. Screenshot of the psychological self-monitoring app in case of distress.



After the longitudinal data collection phase, between February and April 2021, users were invited to complete a postparticipation questionnaire hosted on the SurveyMonkey site [54]. Multiple-choice questions were asked to assess user experience with the psychological self-monitoring app, including perceived satisfaction and perceived contribution of the app to self-awareness (Multimedia Appendix 2). Formal psychometric validation was not conducted before deployment. However, to support content validity and face validity, items were designed based on existing frameworks for evaluating digital tools and were reviewed by the multidisciplinary researchers for clarity and relevance. Out of the 194 respondents interested in a telephone interview to further explore their experience, 30 were randomly selected. Interviews were individual, led by the research coordinator and lasted 15 minutes on average. They were transcribed by members of the research team, who ensured confidentiality by removing any data that could identify specific individuals.

XSL•FO
RenderX

Instruments and Measures

Outcome: Adherence to the Psychological Self-Monitoring App

Adherence was measured using each participant's completion rate. Completion rate was calculated by considering the frequency of completed weekly monitoring over 12 weeks. An adherence rate of 100% meant that the participant used the app 12 times, as intended.

Association 1: Satisfaction

Satisfaction was measured in the postparticipation questionnaire, which was specifically developed for this study. Participants rated their satisfaction with the app used for psychological self-monitoring (eg, ease of installing the app and user-friendliness of the app) on a scale between 0 and 10, which represented "extremely dissatisfied" and "very satisfied," respectively.

Association 2: Perceived Contribution of the App to Self-Awareness

Perceived contribution to self-awareness was also measured in the postparticipation questionnaire. Participants rated the extent to which psychological self-monitoring on the app raised awareness or prompted reflection about their well-being. Responses varied according to a 4-point Likert scale, corresponding to 0="not at all," 1="a little," 2="moderately," and 3="a lot."

Association 3: Experience of Psychological Distress

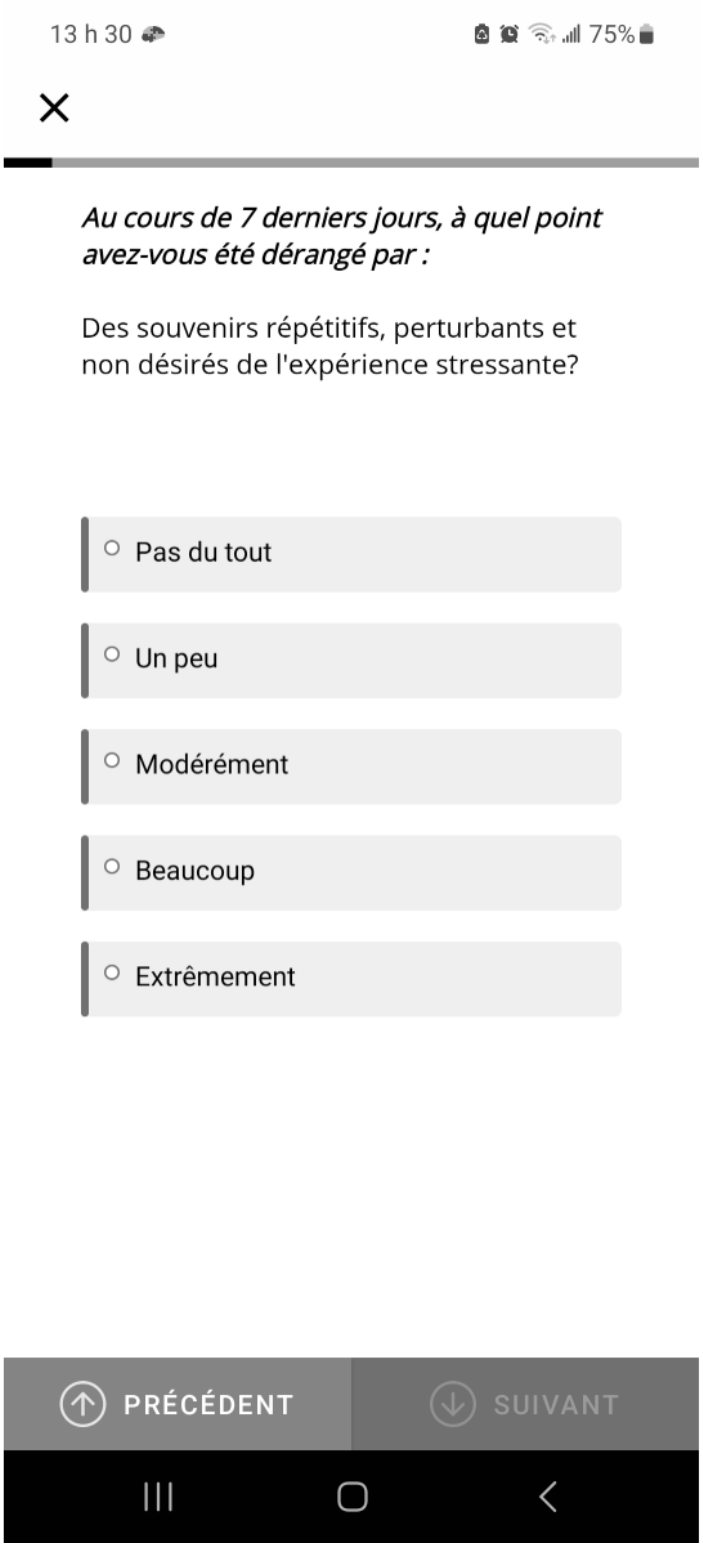
Psychological distress was measured by considering 3 validated psychometric tests that participants weekly filled in the psychological self-monitoring app. During psychological self-monitoring, if at least one score reached the threshold for at least one of these questionnaires, significant psychological distress was identified. This variable was coded either as 0="absent" or 1="present."

The Generalized Anxiety Disorder-7 [55] is a 7-item questionnaire assessing general anxiety symptoms. Participants indicated on a 4-point Likert scale (0="not at all," 1="several days," 2="more than half the days," and 3="nearly every day") how often they have been bothered over the past week by each symptom. An example of an item is "feeling nervous, anxious or on edge." The total score ranges from 0 to 21 and was calculated by adding up the scores for all 7 items. A total score of ≥10 (cut-off) indicated moderate or severe generalized anxiety symptoms. In this study, participants filled the French version of the questionnaire weekly. It has been validated with an excellent internal consistency (Cronbach α=0.90) [56].

The PTSD Checklist for DSM-5 [57] is a self-report questionnaire assessing the severity of PTS symptoms. Participants indicated how bothered they were over the past week by each symptom on a 5-point Likert scale (0="not at all," 1="a little bit," 2="moderately," 3="quite a bit," and 4="extremely"). An example of an item is "repeated, disturbing, and unwanted memories of the stressful experience" (Figure 2). A total score was calculated by adding up the scores for all items and the cut-off was set at 13. In this study, participants filled an abbreviated (8 items) French version, which was previously validated [58,59] with a sensitivity of 0.9, a specificity of 0.23, and a Cronbach α of 0.87.

XSL•FO
RenderX

Figure 2. Screenshot of the psychological self-monitoring app assessing posttraumatic stress symptoms.



The Patient Health Questionnaire-9 (PHQ-9) [60] is a 9-item questionnaire assessing depressive symptoms. Participants indicated how much they have been bothered over the past week by each symptom on a 4-point Likert scale (0="not at all," 1="several days," 2="more than half the days," and 3="nearly every day"). An example of an item is "little interest or pleasure in doing things." A total score was calculated by adding up the scores for all items, and the cut-off was set at 10. In this study, participants filled the French version, which was previously validated with a Cronbach α of 0.86 [61,62].

User Experience

To address the second objective, individual telephone interviews were conducted. These semistructured interviews followed a grid (Multimedia Appendix 3). Its development was inspired by the Mobile App Rating Scale, a reliable tool for classifying and assessing the quality of mobile health apps [63]. It was

XSL•FO
RenderX

adjusted to align with the research objectives, exploring subjects such as experience and satisfaction with the app, psychological self-monitoring, emotional self-awareness, coping strategies, and app use.

Data Analysis

Statistical Analysis

SPSS Statistics (version 29.0.1.0; IBM Corp) was used to carry out the analysis. Missing data were not statistically addressed considering they composed less than 5% of each examined variable. Statistics were computed based on available data for each variable.

First, descriptive statistics were used to examine frequency, average, and SD for each variable. Next, bivariate correlations explored relationships among biological sex, age, app adherence, satisfaction, perceived contribution to self-awareness, and psychological distress, with $P<.05$ considered statistically significant. Correlation strengths were classified as small ($r\approx0.10$), medium ($r\approx0.30$), and large ($r\geq0.50$) [64]. Finally, a hierarchical multiple regression was conducted to further examine whether satisfaction, perceived contribution to self-awareness, and psychological distress influenced adherence to the psychological self-monitoring app. Confounding variables were participants' biological sex and age.

During the review processes, nonparametric tests were conducted to examine whether there were significant differences between app users who did not participate in either the questionnaire or the interview, those who only participated in the questionnaire, and those who participated in both the questionnaire and the interview. Subsequently, inconsistencies were identified in the coding of psychological distress, specifically regarding the cut-off score used for the PHQ-9. While a cut-off score of ≥10 is widely accepted for identifying major depression [60], some data had been coded using a threshold of ≥9. To address this issue, a consistent cut-off score of ≥10 was used for PHQ-9, a rigorous reprocessing and verification of all study variables were done, and all analyses were carefully rerun to ensure the accuracy and reliability of results.

Thematic Analysis

An inductive and progressive thematic analysis was conducted using Braun and Clake approach [65] to analyze data from 30 interviews. Analysis began with a familiarization with the material, followed by generating initial codes for significant excerpts aligned with the study's aims. These codes were reviewed to ensure they reflected participants' statements. Next, codes were compiled into a separate file for sorting into potential themes, grouping salient units into progressively named and refined themes. Patterns within each theme were organized into subthemes, and an interjudge agreement was sought between first and second author to enhance clarity. Results were documented in a table with central themes, subthemes, codes, and verbatim excerpts, culminating in a thematic map summarizing the findings. In this paper, fictive names were attributed to participants.

Results

Sample Analysis

Out of the 831 app users, 407 (49%) did not complete the postparticipation questionnaire, while 424 (51%) did, including 30 (3.6%) participants who also took part in an interview. The study sample consisted of users who completed the questionnaire, including those interviewed. Descriptive statistics for this study sample are presented in Table 1. To assess potential selection bias, users who did not complete the postparticipation questionnaire were compared with the study sample (Table S1 in Multimedia Appendix 4). Results of Mann-Whitney *U* and chi-square tests indicated no significant differences in biological sex ($P=.58$) or psychological distress ($P=.77$). However, study sample was on average older ($U=74{,}310$; $P=.02$; $r=-0.08$) and had significantly higher adherence scores ($U=54{,}513$; $P<.001$; $r=0.33$) than nonrespondents. In addition, participants who were interviewed were compared to those who only completed the questionnaire (Table S2 in Multimedia Appendix 4). Results of Mann-Whitney *U* and chi-square tests indicated no significant differences in age ($P=.15$), biological sex ($P=.41$), satisfaction ($P=.56$), psychological distress ($P=.96$), or perceived contribution to self-awareness ($P=.25$). However, interviewees had significantly higher adherence scores ($U=3994$; $P=.002$; $r=0.14$) than those who only completed the questionnaire.

XSL•FO
RenderX

Table 1. Characteristics of questionnaire respondents and interview participants.

Characteristics	Questionnaire respondents (n=424)	Interview participants (n=30)
Age (y), mean (SD)	40.8 (9.9)	43 (9)
Satisfaction (0 to 10), mean (SD)	8.0 (2.0)	8 (2)
Adherence (%), mean (SD)	74.5 (29.4)	88 (23)
Biological sex, n (%)		
Female	369 (88.1)	25 (83.3)
Male	50 (11.9)	5 (16.7)
Perceived contribution to self-awareness, n (%)		
Not at all	20 (4.8)	0 (0)
A little	93 (22.2)	6 (20)
Moderately	165 (39.5)	9 (30)
A lot	140 (33.5)	14 (46.7)
Psychological distress, n (%)		
Present	228 (54.3)	16 (53.3)
Absent	192 (45.7)	14 (46.7)

Correlation and Regression Analysis

For the study sample, a correlation analysis was conducted to examine relationships among the research variables (Table 2). Adherence was found to have small but significant positive associations with both satisfaction (r=0.19; P<.001) and perceived contribution to self-awareness (r=0.20; P<.001). In addition, a moderate positive correlation was observed between satisfaction and perceived contribution to self-awareness (r=0.37; P<.001). Regarding psychological distress, small negative correlations emerged with age (r=−0.18; P<.001) and satisfaction (r=−0.11; P=.03), suggesting that participants who experienced psychological distress tended to be younger and reported lower satisfaction with the app. No significant association was found between psychological distress and adherence (P=.08).

XSL•FO
RenderX

Table 2. Correlation analysis (Pearson r and two-tailed *P* value) among research variables.

Variable	BS[a]	Age	PD[b]	Adherence	Satisfaction	PCSA[c]
BS						
r	1	0.013	−0.011	−0.020	−0.027	−0.001
P value	—[d]	.80	.83	.69	.58	.99
Age						
r	0.013	1	−0.175	−0.049	0.056	0.085
P value	.80	—	<.001	.32	.26	.09
PD						
r	−0.011	−0.175	1	0.086	−0.107	0.001
P value	.83	<.001	—	.08	.03	.98
Adherence						
r	−0.020	−0.049	0.086	1	0.191	0.201
P value	.69	.32	.08	—	<.001	<.001
Satisfaction						
r	−0.027	0.056	−0.107	0.191	1	0.365
P value	.58	.26	.03	<.001	—	<.001
PCSA						
r	−0.001	0.085	0.001	0.201	0.365	1
P value	.99	.09	.98	<.001	<.001	—

[a]BS: biological sex.

[b]PD: psychological distress.

[c]PCSA: perceived contribution to self-awareness.

[d]Not applicable.

A hierarchical multiple regression analysis was conducted predicting adherence with age and biological sex entered first, satisfaction with the app, perceived contribution to self-awareness, and presence of psychological distress entered second (Table 3). The first model with age and biological sex was not statistically significant ($F_{2,404}$=0.79; *P*=.46). The second model, which included all predictors, was statistically significant ($F_{5,401}$=6.59; *P*<.001), accounting for 7.6% of the variance in adherence. Specifically, satisfaction (β=.16; t_{401}=3.14; *P*=.002) and perceived contribution to self-awareness (=0.15; t_{401}=2.89; *P*=.004) were significantly associated with adherence, indicating that individuals with higher satisfaction or perceived contribution to self-awareness scores had higher adherence. However, the psychological distress was not a statistically significant predictor of adherence (*P*=.07).

Table 3. Multiple hierarchical regressions analysis of adherence.

Models	*B*	β (SE; 95% CI)	*t* test (*df*)	*P* value	R^2
Model 1					.004
Biological sex	−1.76	−.02 (4.53; −10.66 to 7.14)	−0.39 (404)	.77	
Age	−0.17	−.06 (0.15; −0.46 to 0.12)	−1.18 (404)	.24	
Model 2					.076
Satisfaction	2.49	.16 (0.79; 0.93 to 4.05)	3.14 (401)	.002	
PCSA[a]	4.99	.15 (1.73; 1.58 to 8.40)	2.89 (401)	.004	
PD[b]	5.25	.09 (2.87; −0.39 to 10.88)	1.83 (401)	.07	

[a]PCSA: perceived contribution to self-awareness.

[b]PD: psychological distress.

XSL•FO
RenderX

User Experience Themes

Overview

Interviewees' characteristics are provided in Table 1. From the thematic analysis, 3 main themes emerged (Textbox 1). The first theme regrouped instances in which participants described psychological self-monitoring as an introspective practice, imbued with reports of self-awareness and self-care. The second theme characterized the app as practical but not a one-size-fits-all, and presented how different app properties influenced user experience. The last theme regrouped potential enhancements that the participants mentioned to improve their experience.

Textbox 1. Thematic analysis of user experience interviews.

Introspective practice

- Self-awareness: Taking time to pause; focusing; acknowledging; reacting to alerts
- Self-care: Being more attentive; talking to support system; consulting primary care physician; practicing self-compassion; problem-solving

Simple app but not a one-size-fits-all

- Practical use: Ease of use; time efficiency; portability
- Kinds of questions: Clarity; repetitiveness; range
- Presence of notifications: Incentive; bothersome
- Display of resources: Relevant; rarely used

Potential enhancements

- Feedback: Weekly overall assessment; relativity to other users; progress tracking
- Writing space: For reflection; for context
- Beyond self-monitoring: Psychological tools; discussion forum; maintaining simplicity

Introspective Practice

Participants reported a rather unanimous view of psychological self-monitoring. Most participants viewed completing the weekly questionnaires as a moment for reflection and self-awareness, allowing them to pause and consider their well-being. This opportune time allowed them to take a step back to focus on their state, allowing its recognition. Some participants added that psychological self-monitoring enhanced awareness of psychological fluctuations over the weeks:

> *When I was answering the questions, it was [helpful] because I had the time to ask myself the question and to question myself: am I stressed? Are there things that get to me more than before?... It seemed to make me aware of what was really going on around me, and it gave me an indicator of "am I more stressed than usual? Is it getting to me? How do I see things around me? Are things going well, or not so well?"* [Sam]

> *I found that [using the app] had benefits, in the sense that it helped us to make certain realizations, that it helped us to take a moment to stop when it's not always something we think about doing.* [Maria]

Half of the participants in this subsample received a message alerting of a significant psychological distress, leading to varied reactions. Some participants were not surprised by the warning. Feeling it aligned with their internal state, they considered the psychological self-monitoring app reflected their experience:

> *[The distress warning] didn't surprise me because I know myself well and I'm really familiar with mental health problems, so I was able to see what was going on.* [Maria]

> *It just echoes something I was feeling, but having an external message perhaps helped me to confirm that feeling a bit. Logically, it's not just me imagining things that aren't going well and I do have the right to say to myself that there's something wrong and that it's not just me imagining things.... In any case, this message confirms that I have to be "easy" on myself. That's what it's saying.* [Mark]

Interestingly, a participant with lasting distress reported her frustration to be repeatedly "told again" about it and recommended less frequent self-monitoring:

> *I don't remember how many weeks it lasted..., but I knew I wasn't doing well so I was still going to be told that I wasn't doing well. [The distress warning] is still going to come out....You know, I don't need to be told again, I know it's about the same level. But I think, for someone who doesn't know or who wants to assess or who's in between, I think it's important to do it often.* [Nicole]

However, others were caught off guard. Initial reactions of surprise, irritation, anger, or denial were reported in the interviews, as demonstrated in the subsequent quotes. Acknowledging the distress warning as a reflection of their questionnaire responses was reported as an important step toward self-awareness and self-care practices:

> *[Receiving the distress warning] was still a surprise, it really annoyed me, I always had warnings...Except it was because I was in denial!...it's just that it was*

XSL•FO
RenderX

> *spitting all my distress and confusion and despair right back in my face....at first, I was genuinely irritated. I was getting tired of it; it was getting on my nerves. Later, at some point during the summer when I was filling in the damn questions, I was like "why do I feel this way? why do I feel like bawling?" Then, at some point, I was answering the questions and I was so uncomfortable I had tears. Then, I was like "OK, let's stop burying our heads in the sand!"* [Jude]

> *I was actually surprised, because I wasn't expecting [to receive the distress warning]. I didn't really think anything was wrong.... It really made me react. It got to me....I found [the psychological self-monitoring app] helpful in the end. It was a help. Maybe it prevented me from falling into depression or whatever.* [Riley]

For some participants, the distress warning was seen as a "wake-up call" (Victoria) or "a kick in the butt" (Charlie) that prompted them to take actions, such as consulting their physician (Charlie). Other actions taken after receiving the warning included being more attentive to one's condition, being kinder to oneself, seeking and adopting concrete solutions for individual stressors, etc:

> *[The distress warning] led me to call my psychiatrist and talk about it...When it asked the famous question "Would you be better off dead?" I answered "yes." But I answered the question. I was honest with myself. I was surprised by my message, but I tell myself it's a reflection of my answers.* [Megan]

However, participants who did not receive the distress warning also reported adopting changes prompted by the psychological self-monitoring practice. Actions included reducing media exposure; talking with family, friends, and colleagues about their mental health; asking for help from their manager; and seeking work-life balance. This mobilization was often propelled by their reaction to their responses when filling out the questionnaires:

> *When I say more negative things in the app, I try to verbalize them a little more with my family...or I try to change my way of seeing things too...I try to put things into perspective a little, and tell myself that how I felt was normal.* [Sophia]

A small proportion of participants specified that they might have adopted those changes without psychological self-monitoring. However, they presumed that they would have delayed action or waited for their condition to worsen before acting:

> *[Self-monitoring] actually made it a bit easier to take the decision to say "I've got this job to do, but it won't be at all costs," to give myself limits. [Without the app], maybe it would have taken longer to set up, maybe it would have come later. Maybe it would have taken on proportions that would have been a bit more troublesome.* [Mark]

> *I saw that things were starting to go downhill more and more, so I stopped listening to the press briefings....I even deleted my [social media account] just to give you an idea....Otherwise, I started doing more yoga, to switch things up while I waited for the pool to reopen. [Without the app], maybe it would have taken longer for me to realize what was going on.* [Julia]

Interestingly, about half of the interviewees viewed psychological self-monitoring itself as a self-care practice, considering the app as an effective psychological self-help tool. As demonstrated in the following quotes, some even integrated it into their routine, and it became a habit to the point of waiting for the weekly notification:

> *I thought that [psychological self-monitoring] was a moment for me, I always did it at the same time, after reading the newspaper on Saturday, and it was a bit of a retrospective of my week, part of my little ritual of taking care of myself, after my newspaper, my coffee, and then I wondered how I was doing.* [Mika]

> *It's so useful. I found it interesting because every week, you were practically waiting for the app to tell you "you have a new survey to do." It allowed me to keep up to date with where I was...I thought it was a great, relevant tool, [especially] for someone like me who has generalized anxiety disorder.* [Megan]

> *It allows me, even now [post-participation], to take more time, to self-monitor, to look at how I'm doing, to stop and say to myself "well, is this good? Or am I going the wrong way?"* [David]

A minority of participants reported that psychological self-monitoring did not have a significant impact on the way they managed their well-being. Notably, they considered that they were already managing well on their own or that no particular actions were needed:

> *I'm not sure [the app influenced my actions], maybe like 10%-5%....I'm quickly looking for solutions and thinking about how I can think differently.* [Clara]

> *[When I got the distress warning] I wasn't really surprised...But I didn't feel that I necessarily needed help, you know, I read a lot about that and maybe that's why I didn't seek specific help, I didn't feel that it was that problematic in relation to work.* [Patricia]

A Practical App but Not a One-Size-Fits-All

Overall, participants found the psychological self-monitoring app practical, which seems to have contributed to their satisfaction. They appreciated that the app was simple, easy to use, and not time-consuming (approximately 5 minutes on average). In addition, the app's accessibility on smartphones contributed to its portability and practicality:

> *I often did it at work between two things. It paged me, I did it straight away. I found it practical. It's good because it doesn't take long to do.* [Charlie]

> *[The app] was friendly, easy, I liked it, I had a reminder, "oh yes, I have to answer." Again, it's fast.*

XSL•FO
RenderX

> *The fact that it's an app, we all agree, it's on a different level compared to the paper questionnaires, we can agree that it was a lot easier.* [Sara]

Few participants reported technical issues, such as glitches and the absence of notifications, which negatively affected their experience and adherence. Very few contacted the research coordinator for help:

> *At a certain point, it just sort of fell through. I don't know why, I had some sort of technical problem and it didn't work. And after that, I kind of lost interest and disengaged.* [William]

How participants perceived the questions seemed to have influenced their experience and overall satisfaction. First, most of the participants found the questions clear and direct. The use of Likert scales made psychological self-monitoring easy and fast, with very few participants reporting that the questionnaires were long and that it was sometimes hard to choose between 2 scales:

> *The highlights [of the app]? The questions. I mean, they were clear. It wasn't complicated.* [Julia]
>
> *What I liked about the app was that it was easy to use. The questions were simple. Sometimes I'd get stuck between two answers. I said to myself that it's normal because they're not development questions....In general, the questions were quite explicit.* [Frederick]

Second, users had to answer a wide range of questions. Some participants reported finding them extensive because personal and professional aspects were included:

> *[The questions] were so broad, it seemed to cover every area of mental health you could experience. I was really positively surprised.... [My expectations were] exceeded even, because it was all subjects that I didn't expect that I might have difficulties with in my life,...it's things that you don't think about, it's so banal, but it's true that it was affected by COVID.* [Sophia]
>
> *I liked the app and the questionnaires because there was a lot of professional and personal portions to see what affected me the most.* [Sabrina]

However, a significant portion of participants found them too centered on either psychological symptoms or COVID-19. For instance, a participant expressed less interest in questions about COVID-19 exposure from patients. Three participants (David, Mika, and Isabelle) explicitly mentioned the desire to skip the questions they perceived as less relevant to them, such as PTS symptoms assessed in the PTSD Checklist for DSM-5:

> *I work in a hospital, but I'm in a field where I'm not close to patients. So there were questions about [exposure to COVID-19] where I felt a little less concerned, even though yes, the pandemic affected my work very, very directly...It's as if there were questions, I would have liked to skip sometimes, because I thought "well, it's not for me."* [Axel]
>
> *[As improvements,] if it would have been possible to skip the PTS thing since it never applied to me. I could have eliminated a third of the questions because, unless you've been through a trauma while you're answering, I mean, it's quite surprising that this aspect changes when you haven't been through a trauma.* [Isabelle]

Nevertheless, few participants mentioned a sense of comfort knowing that the app was tailored for HCWs' well-being, particularly during times of crisis. They invited their colleagues to participate in the research project, showed interest in using the psychological self-monitoring app for more than 12 weeks, or suggested to make such apps available for all HCWs:

> *I found [the app] recomforting. It was like someone was...thinking about taking care of my well-being.* [Remy]
>
> *It went really well. It was very positive and I'd be happy to use it again if there were ever another study like that with this app.* [Riley]
>
> *I think that [the psychological self-monitoring app] is something that needs to be deployed among HCW, because I don't think we realize how much more distressed we [actually] are...I could even be a spokesperson and say, "Look, it bothered me so much at the beginning..." I found it tiring, but it was salutary for me. I wasn't really expecting it, but that's what it was.* [Jude]

The repetitiveness of the weekly questions generated divided opinions. On one side, it was redundant, and some participants admitted paying less attention or being more passive when filling out the questionnaires. On the other side, it was predictable. Some participants appreciated knowing what would be asked of them in advance and revealed paying more attention to their state during the week. Regardless, the repetitiveness allowed some to identify trends and changes over the 12-week period:

> *Since it was always the same questions from week to week, I found that it made things easier once you got used to the questions.... I'm someone who's introspective and self-observant, so in the week between answering the questionnaires, I paid a bit more attention....I knew what to pay more attention to.* [Maria]
>
> *I found the questionnaire relatively long, and at a certain point, you come to know the questions and generally, yes, there's variability in some places, but there are other questions that haven't varied all that much.* [Mika]
>
> *I thought the questions would change over time. But I found that they were always pretty much the same questions. At some point, not that you want to, but it becomes automatic when you answer the question. You think less about it because it's the same questions....But still, sometimes yes it was the same question, but sometimes I didn't give the same answer either, and it made me think. You know, what did I*

XSL•FO
RenderX

> *do this week that made me more tired, or more stressed? It gave me time to stop and think.* [David]

The notification reminding participants to complete the psychological self-monitoring seemed to have impacted their satisfaction and adherence. As expected, notifications served as an incentive; participants who did not receive any notification reported either forgetting to complete their monitoring or losing interest in the app:

> *The good thing was that even if sometimes I forgot to take part in the study during the day...the app would remind me that I hadn't completed it. I found that interesting. I found it practical.* [Frederick]

At the same time, a few participants expressed dissatisfaction with the timing of notifications. Participants found that those sent at inconvenient hours, such as during sleep, were bothersome and preferred more control over notification timing:

> *The only annoyance I've had with the app is the notifications. I turned off the app's notifications because it was ringing at midnight or 5am on days off. I was a bit at a loss.* [Audrey]

Regarding help resources displayed upon the completion of the questionnaires, most participants reported not using them despite finding them relevant. Furthermore, a minority of participants reported not noticing those resources, and a few were confident they were not displayed, possibly due to technical issues:

> *I took [screenshot of the resources] and I really took it as a helping hand, saying "we're really here, whatever happens..."But it was really circumstances that made me turn more to my parents, to my partner, to friends.* [Sabrina]

Potential Enhancements

When discussing their expectations and recommendations for the psychological self-monitoring app, participants mentioned several desired features not currently offered. First, many expressed a desire for feedback after completing the questionnaires, suggesting that a form of result, such as an overall assessment, would enhance their experience. Some suggested a score or a graphic to help them track their progress through the weeks. In the following extract, a participant mentioned her expectation to see an overall assessment of other users:

> *A comment on what more could have been done, for example, is to have a view of the longitudinal evolution. From one week to the next, is it "better"? Is it less good? In other words, having feedbacks.* [Mark]
>
> *What I also expected was to have an overall picture, my overall picture at the end, over the course of the weeks. Like a small graph, it's going well, it's not going well, something like that. Then, I also expected to have an overall portrait of the people who participated in general so I could have a comparison too. Are we all in the same state of mind? Or is it just that...you know, sometimes you feel a bit alone out there, wondering if you're...normal.* [Alex]
>
> *As improvements, a visual or an awareness for people to see, let's say, the percentage of COVID or the percentage of how people actually assess their stress. It's always the same questions but you can't really see the results...see the survey pulse. It would have been fun to see how people felt and whether we're part of the people who are on average more stressed, less stressed, etc.* [Sam]

Second, a few participants mentioned an interest for a writing space within the app. Some envisioned using it to provide context for their answers, as they felt certain responses were influenced by unrelated events and wished to add comments. Others wanted it as a journaling space to reflect on their feelings and thoughts while completing the questionnaires:

> *It's easy with answer choices, but each time, I had a comment that I would have added. I don't know how to say this...In addition to the answer choices, I would have liked if there was a little box, 200 words, because in the end I was often left wanting more, thinking "yes but."* [Sara]
>
> *Perhaps I was expecting to have spaces for answers and comments, as if to clarify my thoughts? It wasn't in the app.* [Axel]

Finally, some participants expressed a desire for features that extend beyond self-monitoring. For example, a few participants suggested integrating psychological tools within the app, such as the informational sheets sent by email and tailored coping strategies. One participant [William] also mentioned the possibility of chatting with other users or counselors:

> *Maybe the app could suggest things, maybe that's why there's no colour, why it's more neutral.* [Clara]
>
> *[To enhance the app] if, say, I'd had a bereavement because of COVID, if I'd been told, on the basis of what I'd answered in the questionnaire, "we can see that you've experienced such and such a thing this week, try such and such a thing."* [Axel]
>
> *Perhaps there could have been tools for sharing with others who are in the same situation as us....You know, when you're told "be careful, it's red," maybe there could be exchange groups or chat rooms with other people like us, with counselors. It could have been something great.* [William]

However, many participants appreciated the app's simplicity and recommended maintaining that aspect:

> *[To enhance the app], I don't really know, because it's important that it's not too long. I'm not sure that adding things is relevant.* [Patricia]
>
> *No [suggestions], I'd keep it that way. What I liked was that it was short. That too was appreciated. It didn't take an hour. I liked that, but no, I don't see what could be different.* [Emily]

XSL•FO
RenderX

Discussion

Principal Findings

Overview

This mixed methods study aimed to explore Canadian HCWs' experience with a psychological self-monitoring app. To examine whenever satisfaction, perceived contribution to self-awareness, and experience of psychological distress influenced users' adherence, a multiple hierarchical regression analysis was performed on quantitative data while adjusting for age and biological sex. To further explore user experience, thematic analysis was performed on qualitative data.

Over 12 weeks, the average adherence rate to the psychological self-monitoring app was 74.5%. This can be characterized as high considering that only 0.5% to 28.6% of research participants using self-help apps or programs for depression and anxiety sustain their use after 6 weeks [29,66]. Statistical analysis suggested that approximately 7.6% of adherence's variation could be explained by satisfaction with the app and perceived contribution to self-awareness. Thematic analysis of qualitative data offered a more profound exploration of participants' experience. Mixed methods findings are discussed following the key areas Chan and Honey [67] identified in their integrative review on users' perceptions: satisfaction, ease of use, helpfulness, technical and perceived issues, and improvements. Then, the specific case of psychological distress is discussed.

Satisfaction

The high satisfaction rate of 80% was unexpected given that the psychological self-monitoring app was primarily intended for data collection. Interface design is known to significantly impact satisfaction and adherence [68], with visual appeal [69] and personalization [70] enhancing user experience. The Avicenna app used for psychological self-monitoring offers limited design options, but the qualitative component of this study may help explain the discrepancy between our results and prior research. Interviewees clarified the unexpectedly high satisfaction rate, emphasizing the app's practicality, simplicity, and time efficiency. Unlike in the study by Alqahtani and Orji [71], no one criticized the basic interface. This suggests that users are more likely to adhere to a psychological self-monitoring app they perceive as convenient, consistent with findings from other studies [41,72]. Thus, even a very basic psychological self-monitoring app can offer a satisfactory user experience.

It could also be explained by the way satisfaction was conceptualized. In this study, it was defined broadly as an overall reflection of users' appreciation of their experience with the app. When measured, satisfaction was explained to participants with nonexhaustive examples ("eg, ease of installation, friendliness, etc"). Due to this large definition, participants may have prioritized assessing ease of use over design. As noted by Ng et al [73], satisfaction often overlaps with other user engagement indicators such as usability and acceptability, and further research is needed to clarify its definition in the context of mental health apps.

Ease of Use

Furthermore, interviews revealed that a key facilitator of app's adherence and satisfactory experience was its ease of use. Psychological self-monitoring required users to answer questions that were perceived as clear, straightforward, and requiring minimal efforts. This could corroborate with other studies suggesting that users prefer apps that avoid cognitive overload [33]. Given that participants were HCWs in a health crisis, a simple and intuitive app may be suitable for this population, providing a straightforward experience under stressful conditions.

Helpfulness

In parallel, satisfaction rates may have been influenced by users' perceptions of the app's usefulness or helpfulness. This study hypothesized that the usefulness of a psychological self-monitoring app lies in its ability to enhance emotional self-awareness and found that perceived contribution to self-awareness was positively associated with both satisfaction and adherence, corroborating previous reviews [32,41]. Notably, 305 out of 424 respondents (71.9%) perceived that the app contributed at least moderately to improving their emotional self-awareness. This is illustrated with how interviewees characterized their experience as introspective. Completing questionnaires about COVID-19 exposure and its impacts on psychological state seemed to have heightened participants' attention to their well-being. This finding is significant for HCWs, who are expected to prioritize their patients' needs over their own, particularly in time of health crisis.

This result is particularly interesting for interviewees experiencing psychological distress. Some were surprised when the app alerted of significant distress, suggesting varying levels of emotional self-awareness before monitoring. Those surprised likely had lower initial emotional self-awareness and found the app valuable for enhancing it, possibly leading to higher perceived contribution to self-awareness, satisfaction, and adherence. Conversely, participants already aware of their distress may not have felt that the app significantly improved their self-awareness, likely impacting the perceived contribution to self-awareness, satisfaction, and adherence. Overall, initial emotional self-awareness might influence user experience and adherence, indicating the need for further research in this area.

In addition, some interviewees integrated the use of the app into their weekly routine. Developing a habit of psychological self-monitoring can not only enhance users' adherence [33,72] but may also persist after discontinuing app use [74]. Longitudinal studies are required to assess the potential long-lasting effects of psychological self-monitoring on users' practices.

Furthermore, many participants reported taking steps to maintain or enhance their well-being, regardless of whether they were experiencing distress. This reinforces the idea that self-awareness is crucial for adopting coping strategies [17] and that it can mediate the positive association between psychological self-monitoring and well-being [12-16]. Notably, some interviewees hinted that psychological self-monitoring may have facilitated earlier emotional self-awareness and

XSL•FO
RenderX

mobilization, potentially preventing psychological distress among HCWs.

Technical and Perceived Issues

The most frequently reported barrier to app use was the absence of reminders or their inconvenient timing. Reports indicating that reminders positively impact adherence, while their absence negatively affects it, align with a systematic review [70]. In addition, few interviewees were bothered by reminders displayed at inconvenient times. This corroborates previous findings suggesting detrimental effects of reminders sent at inconvenient times or places on app adherence [33].

Some interviewees reported finding the psychological self-monitoring questions redundant or too centered on a certain aspect. Those criticisms may hint at users' diverse monitoring needs. The app did not offer the option to customize questions, limiting users' ability to tailor their psychological self-monitoring experience to their individual preferences. Considering the critical role of app personalization [70,75], allowing users to select the questionnaires they find most pertinent to monitor may foster a more relevant experience. Thereon, the app admittedly did not adopt a user-centered approach. The fact that some HCWs expressed comfort, knowing the app was deployed out of concern for their well-being, suggests that users may value apps specifically tailored to them.

Improvements

Reported suggestions to improve the app consisted of adding features to enhance psychological self-monitoring and supposedly self-awareness. A common recommendation involved incorporating feedback mechanisms, such as displaying results or summaries within the app. Many interviewees expected some form of feedback or visual outcome. This desire aligns with existing evidence highlighting the perceived usefulness [66,76] and effectiveness [77] of feedback in mental health apps, as well as users' interest in tracking progress or trends in their behavior through visual or numerical summaries [33]. More specifically, providing users with their scores on assessment tools and allowing comparisons with aggregated results from other users could further enhance self-awareness and engagement, as suggested in studies of apps targeting anxiety and depression [75,78,79].

In addition, few interviewees suggested integrating a writing space either to add contextual information or to journal about their experience. Thereon, a study suggested that a mood-tracking and journaling app can enhance introspection [80], and another study found that journaling was associated with improvements in psychological well-being of users with a predisposition for self-reflection [81].

While a proportion of interviewees liked the app for its simplicity, some suggested including more complex features and components within the app, namely educational content and discussion forums. These recommendations suggest unmet needs for psychological information and social connections, which can negatively impact user experience and adherence. Indeed, health information and social opportunities were found to increase users' intention to engage with well-being apps [33]. This suggests that an app offering psychological self-monitoring as its only feature may not be suitable for users who have needs beyond an active monitoring of their psychological well-being. Integrating tools to help users manage psychological well-being, along with personalized strategy recommendations based on individual distress profiles, could improve the perceived relevance of the app. Prior research has shown that tailoring interventions to users' specific profiles, rather than offering a wide range of generic strategies, tends to increase engagement and perceived usefulness [82].

As previously reported, participants also mentioned a desire to choose the questions they are asked, the timing they receive the reminders, and the frequency of their self-monitoring. As hinted in other studies [71,75,83], this result highlights the importance of not overlooking personalization opportunities, even in basic self-monitoring apps. Promoting user interaction with the app by offering options to customize certain components may help reduce perceptions of repetitiveness or inconvenience and increase alignment with individual preferences, ultimately supporting greater adherence [70].

Experience of Psychological Distress

Correlation analysis suggested an association between age and psychological distress, with younger participants reporting more distress than older participants. This result aligns with the research by Park et al [84], in which younger adults reported greater distress and less social support, mindfulness, and emotion regulation skills than older adults during the pandemic. In the same vein, an observational study used socioemotional selectivity theory, suggesting that older adults are motivated to optimize their emotional experiences in the limited life that they have left [85], to explain similar result in their 29-wave longitudinal survey [86]. These findings underscore the vulnerability of younger HCWs and the importance of ensuring that support strategies are accessible and relevant to their specific needs.

Although the presence of psychological distress was negatively associated with satisfaction and that satisfaction was positively associated with adherence, the correlation between psychological distress and adherence was not statistically significant (P=.08), suggesting a potential trend that warrants further investigation. Previous studies found conflicting results regarding psychological distress and adherence to mental health apps [34,42-47]. The hypothesis that users in distress would adhere more to a psychological self-monitoring app as they might feel a greater need to thoroughly monitor their state was infirmed. The expectation-confirmation model [37] and the technology acceptance model [78] are 2 complementary models that may help explain these results.

According to the first one, users' expectations and needs play a crucial role in their satisfaction and subsequent adherence. Some participants were satisfied with the app's simple features, as it met or exceeded their expectations, positively influencing satisfaction and adherence. Others might have had higher expectations and needs that the basic psychological self-monitoring app did not fulfill, leading to lower satisfaction and reduced adherence. This could be particularly the case of distressed users, who might have expected features and

XSL•FO
RenderX

components in the app that could help alleviate their distress, such as the recommended inclusion of psychoeducational materials and personalized strategies. This would align with prior research showing that psychological self-monitoring apps are most effective at reducing psychological distress when they include additional components, such as behavioral interventions and educational modules [12,25-28]. Hence, the app may have not met the heightened expectations of distressed users, leading to a lower satisfaction. Nevertheless, it seems that it was not detrimental to the point of affecting adherence. Indeed, some distressed users may have had other expectations that were met and that supported their adherence, despite an overall lower satisfaction. These insights highlight the importance of considering users' various expectations and to clearly communicate to users what they can expect from an app to prevent deception.

The technology acceptance model further complements this perspective by emphasizing how 2 key factors, perceived ease of use and perceived usefulness, mediate the relation between external factors (eg, user characteristics and system features) and adherence [78]. In this study, although the app may have been perceived as easy to use, its perceived usefulness among participants experiencing psychological distress may have been limited, particularly those who were self-aware. Once again, these users were likely seeking concrete tools or strategies to manage their distress, which the app did not provide. As such, self-monitoring alone may not have been perceived as helpful by individuals in significant distress, possibly leading to reduced satisfaction and early discontinuation of use due to unmet needs. However, adherence may be influenced by other factors, such as users' intrinsic motivation, which might not be directly impacted by psychological distress.

Moreover, results suggest that users, whenever in distress or not, can experience positive outcome from a basic psychological self-monitoring app. While psychological distress can be associated with less satisfaction, it does not necessarily deter users from adhering to the app or perceiving it as beneficial. This suggests that even if the app does not fully meet the expectations and needs of users, certain aspects may still provide value, leading to continued use and perceived usefulness. Future research, notably randomized controlled trails, is needed to rigorously assess the impact of minimalist psychological self-monitoring apps on multiple well-being outcomes.

Clinical Implications

These findings have clinical implications, suggesting that psychological self-monitoring could benefit individuals struggling with self-awareness by enhancing emotional self-awareness and encouraging self-care, and facilitating early detection of distress to prevent further aggravation. Organizations could propose psychological self-monitoring programs to support well-being, particularly in professions at high risk of burnout and turnover. Although this study focused on Canadian HCWs' experience during the COVID-19 pandemic, research, including randomized controlled trials, is needed to assess psychological self-monitoring app effects. However, findings suggest that the app helped some users open up to their support system about their mental health, including their managers and colleagues. Mental health clinicians could incorporate psychological self-monitoring apps into their practice as part of a comprehensive plan, using collected data to explore clients' symptoms and coping strategies. Clinicians should focus on offering personalized feedback and educational materials.

Limitations and Future Studies

Several limitations should be considered when interpreting the present findings, and external validity must be approached with caution. First, participation bias and convenience sampling may have influenced both the quantitative and qualitative results. Although all 831 app users were solicited to fill the postparticipation questionnaire, only 424 (51%) responded, and 149 (17.9%) agreed to be interviewed. Comparative analyses between participants and nonparticipants were conducted posteriori and suggested that higher participation in the research study was associated with greater adherence, with interviewees showing the highest adherence. This trend may indicate that more engaged or satisfied users were overrepresented, potentially leading to an overestimation of adherence and other positive outcomes such as user experience. While the mixed methods approach is a strength, a sequential design could have provided deeper insights, particularly for users with lower adherence, satisfaction, or those experiencing psychological distress.

Second, regarding sample characteristics, a predominance of biologically female participants was observed, representing 88.1% (369/419) of the quantitative sample and 83% (25/30) of the qualitative sample. However, this heterogeneity is consistent with the health care workforce population in Quebec where a similar sex disproportion has been reported in 2023 [87].

Third, the generalizability of the findings to other regions should be considered cautiously. Crisis support resources offered in the app were specific to users' institutions and pandemic-related initiatives. Nevertheless, the psychological distress levels observed in the sample are comparable to those reported in meta-analyses of HCWs' mental health during the COVID-19 pandemic, and the instruments used to assess distress were consistent with those used in international studies [7,8,48,49]. It is therefore reasonable to believe that the findings could be similar for other HCWs in industrialized countries during the COVID-19 pandemic.

Finally, psychological self-monitoring was assessed through self-reports collected on a weekly basis, introducing a potential recall bias. To enhance accuracy, ecological momentary assessment typically consists of 6 assessments per day [88]. However, data suggest that users may have different preferences and offering a customizable self-monitoring frequency could represent a more acceptable alternative, potentially improving user experience and adherence.

This study findings reinforce the idea that a one-size-fits-all approach limits the potential of apps and digital interventions, even when users perceive them as easy to use and helpful. Notably, satisfaction and perceived contribution to self-awareness accounted for less than 8% of the variation in

XSL•FO
RenderX

adherence. In the context of psychological self-monitoring for HCWs, additional app characteristics and features need to be explored to better meet users' diverse needs. Future research should incorporate available design recommendations [68,89] while paying close attention to the specific needs and preferences reported by HCWs in this study. More broadly, the COVID-19 pandemic has highlighted the urgent need for proactive programs to support HCWs during crises. Although robust evidence for the effectiveness of psychosocial interventions remains limited [90], the findings suggest potential benefits of basic psychological self-monitoring apps. Further research is warranted to evaluate their impact on psychological well-being and to identify optimal design features that maximize adherence, satisfaction, and mental health outcomes.

Conclusions

A simple psychological self-monitoring app could be an interesting tool for individuals who want to improve their emotional self-awareness and proactively adopt self-care practices. This study presented how HCWs in a health crisis may benefit from such apps. Some of the adherence to the app could be explained by how satisfied users were and how usefully they thought the app contributed to enhancing their self-awareness. Complementary features, such as access to feedback and psychological tools, may enhance the user experience. Simplicity, ease of use, tailoring, and customization options should be considered when developing psychological self-monitoring apps.

Acknowledgments

The authors would like to acknowledge all workers in health and social services networks for their part during the COVID-19 pandemic. The authors would like to thank all the participants for taking the time to participate in this research. This work was funded by the Fonds de Recherche du Québec—Ministère de l'Économie et Innovation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaires on COVID-19 exposure, social support, and quality of life.
[DOCX File , 25 KB - mhealth_v13i1e70412_app1.docx]

Multimedia Appendix 2
Postparticipation questionnaire.
[DOCX File , 24 KB - mhealth_v13i1e70412_app2.docx]

Multimedia Appendix 3
Interview grid-participation questionnaire.
[DOCX File , 25 KB - mhealth_v13i1e70412_app3.docx]

Multimedia Appendix 4
Characteristics of the participants in different phases of the research.
[DOCX File , 18 KB - mhealth_v13i1e70412_app4.docx]

References

1. Elder E, Johnston AN, Wallis M, Greenslade JH, Crilly J. Emergency clinician perceptions of occupational stressors and coping strategies: a multi-site study. Int Emerg Nurs 2019 Jul;45:17-24. [doi: 10.1016/j.ienj.2019.03.006] [Medline: 31053392]
2. Khalid I, Khalid TJ, Qabajah MR, Barnard AG, Qushmaq IA. Healthcare workers emotions, perceived stressors and coping strategies during a MERS-CoV outbreak. Clin Med Res 2016 Mar;14(1):7-14 [FREE Full text] [doi: 10.3121/cmr.2016.1303] [Medline: 26847480]
3. Cabarkapa S, Nadjidai SE, Murgier J, Ng CH. The psychological impact of COVID-19 and other viral epidemics on frontline healthcare workers and ways to address it: a rapid systematic review. Brain Behav Immun Health 2020 Oct;8:100144 [FREE Full text] [doi: 10.1016/j.bbih.2020.100144] [Medline: 32959031]
4. Luo M, Guo L, Yu M, Jiang W, Wang H. The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public - a systematic review and meta-analysis. Psychiatry Res 2020 Sep;291:113190. [doi: 10.1016/j.psychres.2020.113190] [Medline: 32563745]
5. APA Dictionary of Psychology. American Psychological Association. URL: https://dictionary.apa.org/psychological-distress [accessed 2024-10-24]

XSL•FO
RenderX

6. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
7. Narapaka PK, Katikala K, Ponnala VR, Meenakshi S, Viswas G, Singh M, et al. Psychological health among healthcare professionals during COVID-19 pandemic: an updated meta-analysis. Indian J Psychiatry 2024 May;66(5):413-420 [FREE Full text] [doi: 10.4103/indianjpsychiatry.indianjpsychiatry_173_23] [Medline: 38919571]
8. Huang J, Huang ZT, Sun XC, Chen TT, Wu XT. Mental health status and related factors influencing healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. PLoS One 2024;19(1):e0289454 [FREE Full text] [doi: 10.1371/journal.pone.0289454] [Medline: 38241316]
9. Post-traumatic stress disorder. National Institute for Health and Care Excellence. 2018. URL: https://www.nice.org.uk/guidance/ng116 [accessed 2021-03-21]
10. Swinkels A, Giuliano TA. The measurement and conceptualization of mood awareness: monitoring and labeling one's mood states. Pers Soc Psychol Bull 1995 Sep 01;21(9):934-949. [doi: 10.1177/0146167295219008]
11. Saunders KE, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes JR, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. Eur Psychiatry 2017 Mar;41:115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
12. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
13. Incecik E, Taylor RW, Valentini B, Hatch SL, Geddes JR, Cleare AJ, et al. Online mood monitoring in treatment-resistant depression: qualitative study of patients' perspectives in the NHS. BJPsych Bull 2020 Apr;44(2):47-52 [FREE Full text] [doi: 10.1192/bjb.2019.92] [Medline: 31992374]
14. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
15. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. J Am Med Inform Assoc 2016 May;23(3):477-484. [doi: 10.1093/jamia/ocv165] [Medline: 26911822]
16. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health 2016 Nov 23;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]
17. Lazarus R, Folkman F. Stress, Appraisal, and Coping. Cham, Switzerland: Springer; 1984.
18. Ouliaris C. The importance of self-awareness: musings of a medical student. Australas Psychiatry 2019 Jun;27(3):267-269. [doi: 10.1177/1039856219839479] [Medline: 30931586]
19. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. Lancet Psychiatry 2020 Sep;7(9):813-824. [doi: 10.1016/S2215-0366(20)30307-2] [Medline: 32682460]
20. Zhong S, Yang X, Pan Z, Fan Y, Chen Y, Yu X, et al. The usability, feasibility, acceptability, and efficacy of digital mental health services in the COVID-19 pandemic: scoping review, systematic review, and meta-analysis. JMIR Public Health Surveill 2023 Feb 13;9:e43730 [FREE Full text] [doi: 10.2196/43730] [Medline: 36634261]
21. Kampmann IL, Emmelkamp PM, Morina N. Meta-analysis of technology-assisted interventions for social anxiety disorder. J Anxiety Disord 2016 Aug;42:71-84. [doi: 10.1016/j.janxdis.2016.06.007] [Medline: 27376634]
22. Karatzias T, Shevlin M, Murphy J, McBride O, Ben-Ezra M, Bentall RP, et al. Posttraumatic stress symptoms and associated comorbidity during the COVID-19 pandemic in Ireland: a population-based study. J Trauma Stress 2020 Aug;33(4):365-370 [FREE Full text] [doi: 10.1002/jts.22565] [Medline: 32662129]
23. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
24. Ménard F, Ouellet-Morin I, Guay S. Efficacité des applications mobiles et des messages textes comme intervention en cybersanté mentale pour les 3 blessures de stress opérationnel les plus fréquentes chez le personnel de la sécurité publique : une recension-cadre. Santé mentale au Québec 2021;46(1):251-275 [FREE Full text] [doi: 10.7202/1081518ar]
25. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review. Psychiatr Serv 2017 Nov 01;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]
26. Dewey D, McDonald MK, Brown WJ, Boyd SJ, Bunnell BE, Schuldberg D. The impact of ecological momentary assessment on posttraumatic stress symptom trajectory. Psychiatry Res 2015 Dec 15;230(2):300-303. [doi: 10.1016/j.psychres.2015.09.009] [Medline: 26381184]
27. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

XSL•FO
RenderX

28. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
29. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
30. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
31. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
32. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
33. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
34. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
35. ISO 9241-210: ergonomics of human-system interaction part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2023-03-02]
36. Mahlke S. User experience of interaction with technical systems. Technische Universität Berlin. 2008 Mar 4. URL: https://depositonce.tu-berlin.de/items/d55d37d1-5b5d-4674-af14-d3ef7c59aefd [accessed 2025-07-10]
37. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351-370. [doi: 10.2307/3250921]
38. Ware JEJ, Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. Eval Program Plann 1983;6(3-4):247-263. [doi: 10.1016/0149-7189(83)90005-8] [Medline: 10267253]
39. Kim B. The diffusion of mobile data services and applications: exploring the role of habit and its antecedents. Telecommun Policy 2012 Feb;36(1):69-81. [doi: 10.1016/j.telpol.2011.11.011]
40. Limayem M, Hirt SG, Cheung CM. How habit limits the predictive power of intention: the case of information systems continuance. MIS Q 2007;31(4):705-737. [doi: 10.2307/25148817]
41. Renfrew ME, Morton DP, Northcote M, Morton JK, Hinze JS, Przybylko G. Participant perceptions of facilitators and barriers to adherence in a digital mental health intervention for a nonclinical cohort: content analysis. J Med Internet Res 2021 Apr 14;23(4):e25358 [FREE Full text] [doi: 10.2196/25358] [Medline: 33851925]
42. Calear AL, Christensen H, Mackinnon A, Griffiths KM. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
43. Cillessen L, van de Ven MO, Compen FR, Bisseling EM, van der Lee ML, Speckens AE. Predictors and effects of usage of an online mindfulness intervention for distressed cancer patients: usability study. J Med Internet Res 2020 Oct 02;22(10):e17526 [FREE Full text] [doi: 10.2196/17526] [Medline: 33006567]
44. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
45. Fuhr K, Schröder J, Berger T, Moritz S, Meyer B, Lutz W, et al. The association between adherence and outcome in an internet intervention for depression. J Affect Disord 2018 Mar 15;229:443-449. [doi: 10.1016/j.jad.2017.12.028] [Medline: 29331706]
46. Gulliver A, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Batterham PJ. Predictors of acceptability and engagement in a self-guided online program for depression and anxiety. Internet Interv 2021 Sep;25:100400 [FREE Full text] [doi: 10.1016/j.invent.2021.100400] [Medline: 34026569]
47. Mohr DC, Kwasny MJ, Meyerhoff J, Graham AK, Lattie EG. The effect of depression and anxiety symptom severity on clinical outcomes and app use in digital mental health treatments: meta-regression of three trials. Behav Res Ther 2021 Dec;147:103972 [FREE Full text] [doi: 10.1016/j.brat.2021.103972] [Medline: 34600398]
48. Dufour MM, Bergeron N, Rabasa A, Guay S, Geoffrion S. Assessment of psychological distress in health-care workers during and after the first wave of COVID-19: a Canadian longitudinal study. Can J Psychiatry 2021 Sep;66(9):807-814 [FREE Full text] [doi: 10.1177/07067437211025217] [Medline: 34124958]
49. Dufour MM, Bergeron N, Guay S, Geoffrion S. Assessment of psychological distress in health care workers during the first two waves of COVID-19: a follow-up of a Canadian longitudinal study. Chronic Stress (Thousand Oaks) 2022;6:24705470221108144 [FREE Full text] [doi: 10.1177/24705470221108144] [Medline: 35757571]

XSL•FO
RenderX

50. Rapisarda F, Bergeron N, Dufour MM, Guay S, Geoffrion S. Longitudinal assessment and determinants of short-term and longer-term psychological distress in a sample of healthcare workers during the COVID-19 pandemic in Quebec, Canada. Front Psychiatry 2023 May 18;14:1112184 [FREE Full text] [doi: 10.3389/fpsyt.2023.1112184] [Medline: 37275978]
51. Creswell JW, Clark VL. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: SAGE Publications; 2007.
52. Stone AA, Shiffman S. Ecological momentary assessment (Ema) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
53. Patient-centric clinical trial and health research software platform. Avicenna. URL: https://avicennaresearch.com/ [accessed 2024-10-04]
54. SurveyMonkey homepage. SurveyMonkey. URL: https://www.surveymonkey.com/ [accessed 2025-07-10]
55. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
56. Micoulaud-Franchi JA, Lagarde S, Barkate G, Dufournet B, Besancon C, Trébuchon-Da Fonseca A, et al. Rapid detection of generalized anxiety disorder and major depression in epilepsy: validation of the GAD-7 as a complementary tool to the NDDI-E in a French sample. Epilepsy Behav 2016 Apr;57(Pt A):211-216 [FREE Full text] [doi: 10.1016/j.yebeh.2016.02.015] [Medline: 26994447]
57. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD checklist for DSM-5 (PCL-5). U.S. Department of Veterans Affairs. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2025-07-10]
58. Ashbaugh AR, Houle-Johnson S, Herbert C, El-Hage W, Brunet A. Psychometric validation of the English and French versions of the posttraumatic stress disorder checklist for DSM-5 (PCL-5). PLoS One 2016;11(10):e0161645 [FREE Full text] [doi: 10.1371/journal.pone.0161645] [Medline: 27723815]
59. Price M, Szafranski DD, van Stolk-Cooke K, Gros DF. Investigation of abbreviated 4 and 8 item versions of the PTSD checklist 5. Psychiatry Res 2016 May 30;239:124-130. [doi: 10.1016/j.psychres.2016.03.014] [Medline: 27137973]
60. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
61. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
62. Carballeira Y, Dumont P, Borgacci S, Rentsch D, de Tonnac N, Archinard M, et al. Criterion validity of the French version of Patient Health Questionnaire (PHQ) in a hospital department of internal medicine. Psychol Psychother 2007 Mar;80(Pt 1):69-77. [doi: 10.1348/147608306X103641] [Medline: 17346381]
63. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
64. Cohen J. The test that a proportion is .50 and the sign test. In: Statistical Power Analysis for the Behavioral Sciences. Milton Park, UK: Routledge; 1988.
65. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
66. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
67. Chan AH, Honey ML. User perceptions of mobile digital apps for mental health: acceptability and usability - an integrative review. J Psychiatr Ment Health Nurs 2022 Feb;29(1):147-168. [doi: 10.1111/jpm.12744] [Medline: 33604946]
68. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
69. Perrig SA, Ueffing D, Opwis K, Brühlmann F. Smartphone app aesthetics influence users' experience and performance. Front Psychol 2023;14:1113842 [FREE Full text] [doi: 10.3389/fpsyg.2023.1113842] [Medline: 37388645]
70. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
71. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]
72. Tam C, Santos D, Oliveira T. Exploring the influential factors of continuance intention to use mobile apps: extending the expectation confirmation model. Inf Syst Front 2018 May 31;22(1):243-257. [doi: 10.1007/s10796-018-9864-5]
73. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
74. Clark M, Southerton C, Driller M. Digital self-tracking, habits and the myth of discontinuance: it doesn't just 'stop'. New Media Soc 2022 Mar 18;26(4):2168-2188. [doi: 10.1177/14614448221083992]

XSL•FO
RenderX

75. Garrido S, Cheers D, Boydell K, Nguyen QV, Schubert E, Dunne L, et al. Young people's response to six smartphone apps for anxiety and depression: focus group study. JMIR Ment Health 2019 Oct 02;6(10):e14385 [FREE Full text] [doi: 10.2196/14385] [Medline: 31579023]
76. Fiol-deRoque MA, Serrano-Ripoll MJ, Mira-Martínez S, Pastor-Moreno G, Sitges C, García-Buades ME, et al. Process evaluation of PsyCovidApp, a digital tool for mobile devices aimed at protecting the mental health of healthcare professionals during the COVID-19 pandemic: a mixed method study. Front Psychol 2024;15:1378372 [FREE Full text] [doi: 10.3389/fpsyg.2024.1378372] [Medline: 38577118]
77. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
78. Davis FD, Granić A. The Technology Acceptance Model: 30 Years of TAM. Cham, Switzerland: Springer; 2024.
79. Radomski AD, Bagnell A, Curtis S, Hartling L, Newton AS. Examining the usage, user experience, and perceived impact of an internet-based cognitive behavioral therapy program for adolescents with anxiety: randomized controlled trial. JMIR Ment Health 2020 Feb 07;7(2):e15795 [FREE Full text] [doi: 10.2196/15795] [Medline: 32022692]
80. Russel C, Waworuntu A, Surya D, Zebua JN, Sandy R. Development of an android-based journaling and mood tracking mobile application for improving mental health using firebase application programming interface. Int J Sci Rep 2023 Mar 24;9(4):107-112. [doi: 10.18203/issn.2454-2156.IntJSciRep20230716]
81. MacIsaac A, Mushquash AR, Wekerle C. Writing yourself well: dispositional self-reflection moderates the effect of a smartphone app-based journaling intervention on psychological wellbeing across time. Behav Change 2022 Nov 29;40(4):297-313. [doi: 10.1017/bec.2022.24]
82. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020;3:85 [FREE Full text] [doi: 10.1038/s41746-020-0293-8] [Medline: 32566763]
83. Balaskas A, Schueller SM, Cox AL, Doherty G. Understanding users' perspectives on mobile apps for anxiety management. Front Digit Health 2022;4:854263 [FREE Full text] [doi: 10.3389/fdgth.2022.854263] [Medline: 36120712]
84. Park C, Kim D, Russell B, Fendrich M. Wisdom-related emotion regulation resources explain middle-aged and older adults' COVID-19-related distress. Innov Aging 2023;7(Supplement_1):121. [doi: 10.1093/geroni/igad104.0394]
85. Carstensen LL, Isaacowitz DM, Charles ST. Taking time seriously. A theory of socioemotional selectivity. Am Psychol 1999 Mar;54(3):165-181. [doi: 10.1037//0003-066x.54.3.165] [Medline: 10199217]
86. Best R, Strough J, Bruine de Bruin W. Age differences in psychological distress during the COVID-19 pandemic: March 2020 - June 2021. Front Psychol 2023 Feb 6;14:1101353 [FREE Full text] [doi: 10.3389/fpsyg.2023.1101353] [Medline: 36814666]
87. Emplois dans certaines catégories professionnelles. Institut de la Statistique du Québec. URL: https://statistique.quebec.ca/vitrine/egalite/dimensions-egalite/travail/emplois-certaines-categories-professionnelles [accessed 2025-04-06]
88. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
89. Punchoojit L, Hongwarittorrn N. Usability studies on mobile user interface design patterns: a systematic literature review. Adv Hum Comput Interact 2017;2017:1-22. [doi: 10.1155/2017/6787504]
90. Hoedl M, Osmancevic S, Thonhofer N, Reiter L, Schoberer D. Psychosocial interventions for healthcare workers during the COVID-19 pandemic: rapid review and meta-analysis. Wien Med Wochenschr 2024 Mar;174(3-4):79-86 [FREE Full text] [doi: 10.1007/s10354-023-01013-9] [Medline: 37191933]

Abbreviations

HCW: health care worker
PHQ-9: Patient Health Questionnaire-9
PTS: posttraumatic stress

Edited by M DeJonckheere; submitted 22.12.24; peer-reviewed by T Barros Pontes e Silva, Y Wei; comments to author 09.03.25; revised version received 29.04.25; accepted 17.06.25; published 07.08.25.

Please cite as:
Khaldoun L, Bellemare F, Genest C, Bergeron N, Geoffrion S
Health Care Workers' Experience With a Psychological Self-Monitoring App During the COVID-19 Pandemic: Mixed Methods Study
JMIR Mhealth Uhealth 2025;13:e70412
URL: https://mhealth.jmir.org/2025/1/e70412
doi:10.2196/70412
PMID:

XSL•FO
RenderX

©Lydia Khaldoun, Francois Bellemare, Christine Genest, Nicolas Bergeron, Steve Geoffrion. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 07.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone-Based Care Platform Versus Traditional Care in Primary Knee Arthroplasty in the United States: Cost Analysis

Jess H Lonner[1], MD; Ashwini Naidu-Helm[2], MBA; David Van Andel[3], MS, PT; Mike B Anderson[3], MSc; Richard Ditto[2], MS; Roberta E Redfern[3], MPH, PhD; Jared Foran[4], MD

[1]Rothman Orthopaedics, Rothman Orthopaedic Institute, 825 Old Lancaster Rd, Suite 100, 140, Philadelphia, PA, United States

[2]Health Economic and Market Access, Zimmer Biomet, Warsaw, IN, United States

[3]Clinical Affairs, Zimmer Biomet, Warsaw, IN, United States

[4]Panorama Orthopedics and Spine Center, Golden, CO, United States

Corresponding Author:
Jess H Lonner, MD
Rothman Orthopaedics, Rothman Orthopaedic Institute, 825 Old Lancaster Rd, Suite 100, 140, Philadelphia, PA, United States

Abstract

Cost savings were achieved with the use of a smartphone-based care management platform, considering several health care resources following knee arthroplasty procedures without negatively impacting clinical outcomes.

Trial registration: ClinicalTrials.gov NCT03737149; https://clinicaltrials.gov/study/NCT03737149

(JMIR Mhealth Uhealth 2025;13:e46047) doi:10.2196/46047

KEYWORDS

telerehabilitation; telehealth; telemedicine; rehabilitation; physiotherapy; mobile health; knee arthroplasty

Introduction

The number of knee arthroplasty procedures performed in the United States has increased annually, accompanied by rises in health care resource use and costs [1]. Risk-sharing and bundled payment plans have been implemented with the goal of improving care and controlling costs through the sharing of financial responsibility for a 90-day joint replacement episode-of-care (EOC) [2]. Approximately 36% of the EOC cost has been attributed to postdischarge services [3]. In-person outpatient physiotherapy (PT) has been the standard of care (SoC) postoperatively but is associated with significant cost and may be overused following uncomplicated knee arthroplasty [4]. Telerehabilitation and mobile health delivery of PT has gained popularity in recent years [5]. The aim of this study was to compare costs of a smartphone-based care management platform (sbCMP) with traditional care in adult patients undergoing knee arthroplasty.

Methods

Study Design and Analysis

Patients undergoing primary total or partial knee arthroplasty were randomized (block randomization, block size of 4) to receive institutional SoC (control group) or sbCMP (treatment group) using the mymobility app (Zimmer Biomet), as previously described [6]. The 90-day health care intervention collected information on PT visits, manipulations under anesthesia, non-SoC physician visits, all-cause readmissions, emergency department (ED) visits, and urgent care visits.

Costs associated with the use of each resource were estimated from multiple sources [4,7,8]. The cost of the sbCMP was estimated based on the maximum cost per patient. Costs were multiplied across number of uses; the average cost per patient was calculated using the entire cohort as the denominator and compared between groups by a 2-tailed student *t* test. Costs in the noncrossover treatment group (patients who did not receive traditional PT in the original study) were also calculated. One-way and two-way deterministic sensitivity analyses were performed.

Ethical Considerations

The multicenter randomized controlled trial (ClinicalTrials.gov NCT03737149) received central Institutional Review Board approval (20182103). All patients voluntarily provided written informed consent upon enrollment, with the opportunity to withdraw at any time. Smartwatches were provided (approximately US $329 in value). Data were deidentified prior to analyses.

Results

Preoperative patient characteristics as well as baseline and postsurgical clinical outcomes were similar between the two groups [6]. The estimated costs associated with health care resources are presented in Table 1. The number of events was lower in the treatment group in each category except for

XSL•FO
RenderX

non-SoC physician office visits. Patients using the sbCMP had significantly fewer ED visits and readmissions. The average cost per patient was approximately US $732 lower in the treatment group after including the cost of the sbCMP. The reduction in the number of in-person PT visits accounted for the bulk of the potential savings in the cohort, with costs approximately US $400 less in the treatment group. The average number of PT visits in the control and sbCMP groups were 9.75 (SD 3.98) and 5.40 (SD 5.51), respectively. Considering only noncrossovers, average costs were further reduced to approximately US $186 per patient through 90 days (Table 2). Sensitivity analyses demonstrated cost savings across all scenarios (Multimedia Appendix 1).

Table . Estimated costs associated with resource use as a function of treatment group.

Variable	Cost per unit (US $)[a]	Control arm (n=244)			Treatment arm (n=208)			*P* value
		Frequency, n	Total cost per group (US $)	Average cost per patient (US $)	Frequency, n	Total cost per group (US $)	Average cost per patient (US $)	
Physiotherapy[b]	126[c] [4]	1736	218,736	896.46	799	100,674	484.01	<.001
Readmission	9723 [7]	16	155,568	637.57	5	48,615	233.73	.055
ED visit[d]	519 [8]	16	8304	34.03	5	2595	12.48	.03
MUA	1549	10	15,490	63.48	4	6196	29.79	.20
Physician office visit	27	67	1809	7.41	77	2079	10.00	.18
Urgent care visit	100	3	300	1.23	2	200	0.96	.79
Smartphone-based care platform	137	0	0	0.00	208	28,496	137.00	—
Total			400,207	1640.19		188,855	907.96	.001
Cost reduction			—	—		211,352	732.24	

[a]Reference citations indicated the studies from which the costs were derived.

[b]Estimates based on categorical collection of physiotherapy visits used. Categories included 1 - 3 visits, 4 - 6 visits, 7 - 9 visits, 10 - 12 visits, and 13 or more visits. Given the largest category did not include an upper bound, the lowest number of visits in each category was applied, providing conservative estimates in both groups.

[c]Weighted mean accounting for location (home, outpatient, or both).

[d]Represents only emergency department visits that did not result in admission. Readmissions associated with an emergency department visit are categorized only as readmission, without impacting emergency department visit rate.

[e]ED: emergency department.

[f]MUA: manipulations under anesthesia.

Table . Resource use and estimated costs within the treatment noncrossover group, where in-person physiotherapy was not used.

Variable	Treatment group noncrossover (n=61), n	Total cost per group (US $)	Average cost per patient (US $)
Physiotherapy	0	0	0
Readmission	0	0	0
ED[a] visit	1	519	8.51
MUA[b]	1	1549	25.39
Physician office visit	34	918	15.05
Urgent care visit	0	0	0
Smartphone-based care platform	61	8357	137
Total		11,343	185.95

[a]ED: emergency department.

[b]MUA: manipulations under anesthesia.

XSL•FO
RenderX

Discussion

This analysis demonstrated cost savings for patients using the sbCMP for self-directed rehabilitation following primary knee arthroplasty. Most savings were derived from the reduction in PT, the most common resource used. Cost savings were also achieved with the sbCMP for all health care resources except physician office visits. It is possible that this increase in office visits is the result of improved communication between patients and their care team via real-time messaging. This may have prompted additional unexpected office visits for postoperative concerns, which may have prevented more costly ED visits or readmissions. Alternately, this could be due to a patient need for feedback and assurance not received during in-person PT.

To our knowledge, this is the first cost comparison analysis of this care management platform compared with traditional care. There are few studies of other telerehabilitation programs compared with traditional care in this context. One study reported reduced outpatient PT costs with increased use of web-based PT [9] and another reported overall EOC cost savings with telerehabilitation and acknowledged improved outcomes, without attributing costs to the type of event [10].

A key limitation of this analysis is the uncertainty regarding the generalizability of the list price for the sbCMP used in the analysis. Additionally, the cost data may not be generalizable. Postacute care use was not accounted for given it is not expected to differ between treatment modalities.

In conclusion, the sbCMP has a potential for robust cost savings without negatively impacting postsurgical outcomes in patients undergoing knee arthroplasty.

Acknowledgments

The study was funded by Zimmer Biomet.

Conflicts of Interest

JHL receives royalties and research support from Zimmer Biomet and Smith and Nephew and is a paid consultant. JHL also receives stock or stock options and research support from Force Therapeutics and is a paid consultant. ANH, DVA, MBA, RD, and RER are employees of Zimmer Biomet. JF receives stock or stock options from Cast Orthopedics, is on the editorial or governing board of the Journal of Arthroplasty, and is a paid consultant for Zimmer Biomet.

Multimedia Appendix 1
One- and two-way sensitivity analyses.
[DOCX File, 15 KB - mhealth_v13i1e46047_app1.docx]

References

1. Sloan M, Premkumar A, Sheth NP. Projected volume of primary total joint arthroplasty in the U.S., 2014 to 2030. J Bone Joint Surg Am 2018 Sep 5;100(17):1455-1460. [doi: 10.2106/JBJS.17.01617] [Medline: 30180053]
2. Comprehensive care for joint replacement model. Centers for Medicare & Medicaid Services. URL: https://innovation.cms.gov/innovation-models/cjr [accessed 2023-01-27]
3. Bozic KJ, Ward L, Vail TP, Maze M. Bundled payments in total joint arthroplasty: targeting opportunities for quality improvement and cost reduction. Clin Orthop Relat Res 2014 Jan;472(1):188-193. [doi: 10.1007/s11999-013-3034-3] [Medline: 23649225]
4. Yayac M, Moltz R, Pivec R, Lonner JH, Courtney PM, Austin MS. Formal physical therapy following total hip and knee arthroplasty incurs additional cost without improving outcomes. J Arthroplasty 2020 Oct;35(10):2779-2785. [doi: 10.1016/j.arth.2020.04.023] [Medline: 32674941]
5. Cottrell MA, Galea OA, O'Leary SP, Hill AJ, Russell TG. Real-time telerehabilitation for the treatment of musculoskeletal conditions is effective and comparable to standard practice: a systematic review and meta-analysis. Clin Rehabil 2017 May;31(5):625-638. [doi: 10.1177/0269215516645148] [Medline: 27141087]
6. Crawford DA, Duwelius PJ, Sneller MA, et al. 2021 Mark Coventry Award: use of a smartphone-based care platform after primary partial and total knee arthroplasty: a prospective randomized controlled trial. Bone Joint J 2021 Jun;103-B(6 Supple A):3-12. [doi: 10.1302/0301-620X.103B6.BJJ-2020-2352.R1] [Medline: 34053272]
7. Phillips JLH, Rondon AJ, Vannello C, Fillingham YA, Austin MS, Courtney PM. How much does a readmission cost the bundle following primary hip and knee arthroplasty? J Arthroplasty 2019 May;34(5):819-823. [doi: 10.1016/j.arth.2019.01.029] [Medline: 30755375]
8. Sibia US, Mandelblatt AE, Callanan MA, MacDonald JH, King PJ. Incidence, risk factors, and costs for hospital returns after total joint arthroplasties. J Arthroplasty 2017 Feb;32(2):381-385. [doi: 10.1016/j.arth.2016.08.003] [Medline: 27597429]
9. Zachwieja E, Theosmy EG, Yacovelli SJ, Beatty EW, McGrath ME, Lonner JH. Web-based self-directed exercise program is cost-effective compared to formal physical therapy after primary total knee arthroplasty. J Arthroplasty 2020 Sep;35(9):2335-2341. [doi: 10.1016/j.arth.2020.04.061] [Medline: 32423757]

XSL•FO
RenderX

10. Prvu Bettger J, Green CL, Holmes DN, et al. Effects of virtual exercise rehabilitation in-home therapy compared with traditional care after total knee arthroplasty: VERITAS, a randomized controlled trial. J Bone Joint Surg Am 2020 Jan 15;102(2):101-109. [doi: 10.2106/JBJS.19.00695] [Medline: 31743238]

Abbreviations

ED: emergency department
EOC: episode-of-care
PT: physiotherapy
sbCMP: smartphone-based care management platform
SoC: standard of care

Edited by L Buis; submitted 05.06.23; peer-reviewed by DZ Nin, EL Woon; revised version received 06.12.24; accepted 20.12.24; published 03.02.25.

Please cite as:
Lonner JH, Naidu-Helm A, Van Andel D, Anderson MB, Ditto R, Redfern RE, Foran J
Smartphone-Based Care Platform Versus Traditional Care in Primary Knee Arthroplasty in the United States: Cost Analysis
JMIR Mhealth Uhealth 2025;13:e46047
URL: https://mhealth.jmir.org/2025/1/e46047
doi:10.2196/46047

© Jess H Lonner, Ashwini Naidu-Helm, David Van Andel, Mike B Anderson, Richard Ditto, Roberta E Redfern, Jared Foran. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX